It's About Time - Presentation in Honor of Ira Hirsh

Ken W. Grant, Van Summers, Brian Walden, Mary Cord
Walter Reed Army Medical Center
Army Audiology and Speech Center
Washington, DC

Steven Greenberg
International Computer Science Institute
Berkeley, CA

David Poeppel, Virginie van Wassenhove
Department of Linguistics and Biology
Neuroscience and Cognitive Science
University of Maryland
College Park, MD

Some Colleagues at CID (1976, 1981-1985)

Boettcher
Bohne
Clark
Davis
DeFilippo
Divenyi
Eldredge
Engebretson
Erber
Formby
Gagne
Geers
Gibian
Gilkey
Hirsh
Kuhl
Lauter
Long
Miller
Mills
Monsen
Moog
Pascoe
Pastore
Popelka
Ronken
Sachs
Scott
Sinex
Singh
Skinner
Spiegel
Watson
Weisenberger
Zurek
et al

CID Research Environment (1976-1985)

Acoustic Events and Sequences - Hirsh 1974

- Levels of timing perception - Three ranges
  - simultaneity or fused events - 1-15 ms
  - Gestalt properties or figural aspects - 20-100 ms
  - separate events or item-by-item analyses - up to 500 ms
- Asymmetry of temporal windows
  - not usually seen for trained observers
  - A proceeds V - 60 ms
  - V proceeds A - 100 ms (Hirsh and Fraisse, 1964)

Spectro-Temporal Properties for Speech

Determine precisely the temporal window of integration (TWI) for Auditory and Auditory-Visual speech input.

Spectral-Temporal Integration

- Time-Course of Spectral-Temporal Integration
  - ***Within Modality (A)***
    - 4 narrow spectral slits
    - delay two middle bands with respect to outer bands
  - ***Across Modality (AV)***
    - 2 narrow spectral slits (2 outer bands) plus speechreading
    - incongruent versus congruent AV speech tokens
    - delay audio with respect to video

AUDIO-ALONE EXPERIMENTS

Auditory Tasks

TIMIT Sentences

- Word Recognition (all words)
  - Audio slits 1 + 4
  - Audio slits 2 + 3 presented at various temporal asynchronies

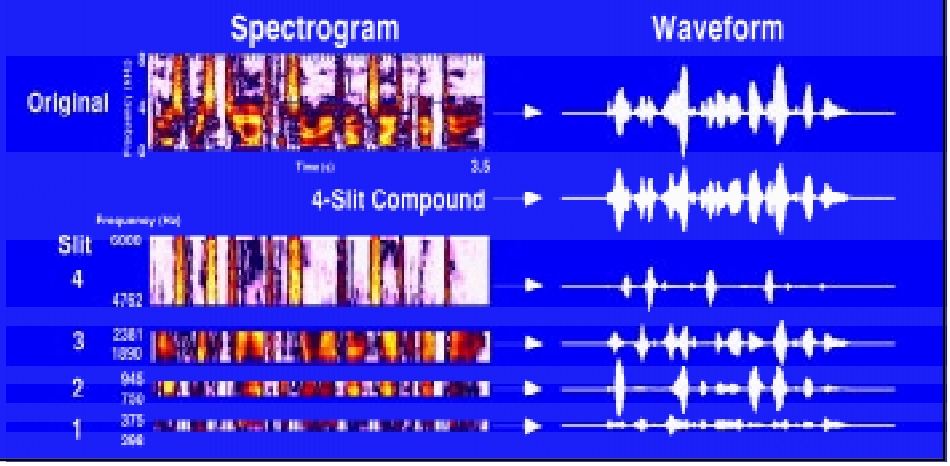
IEEE Sentences

- Synchrony Discrimination
  - 2IFC adaptive tracking

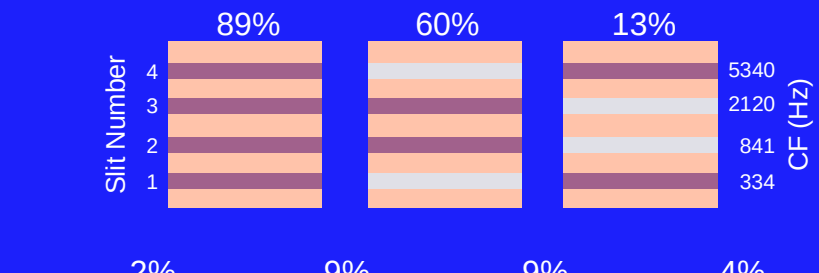
Audio (Alone) Spectral Slit Paradigm

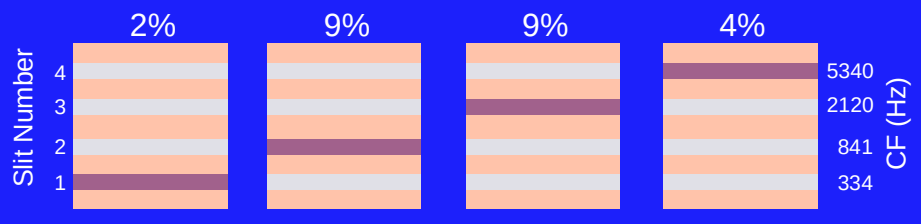
The edge of each slit was separated from its <u>nearest</u> neighbor by an octave

Can listeners decode spoken sentences using just four narrow (1/3 octave) channels ("slits") distributed across the spectrum? – YES (cf. next slide)



Word Intelligibility - Single and Multiple Slits

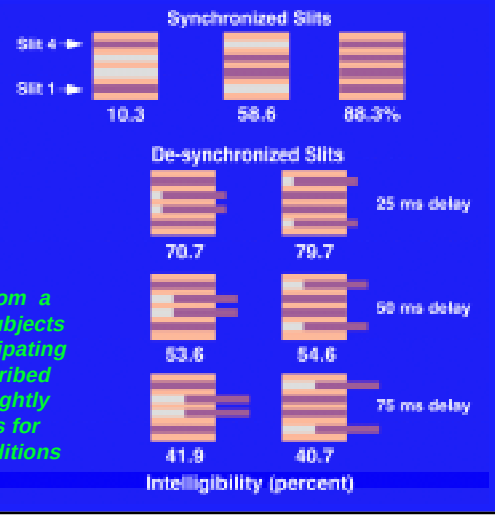




Slit Asynchrony Affects Intelligibility

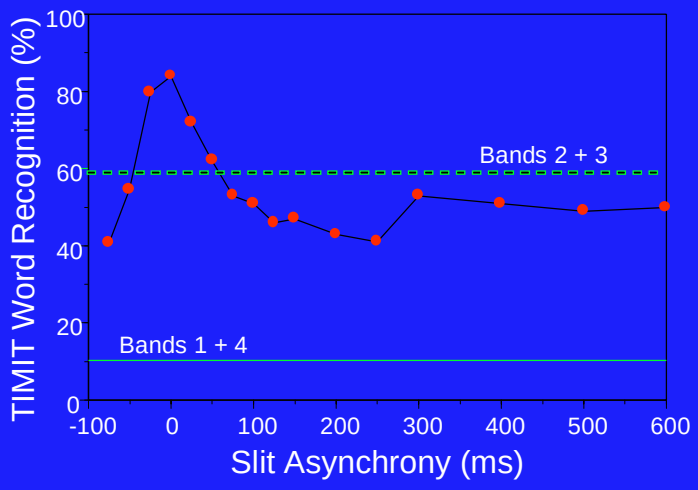
Desynchronizing the slits by more than 25 ms results in a significant decline in intelligibility

The effect of asynchrony on intelligibility is relatively symmetrical



These data are from a different set of subjects than those participating in the study described earlier - hence slightly different numbers for the baseline conditions

TIMIT Sentence Recognition - % Words Correct



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

AUDIO-VISUAL EXPERIMENTS

Auditory-Visual Tasks

IEEE Sentences

- Recognition of key words
  - Audio slits 1 + 4
  - Video presented at various temporal asynchronies
- Synchrony discrimination
  - 2IFC adaptive tracking

CV Syllables

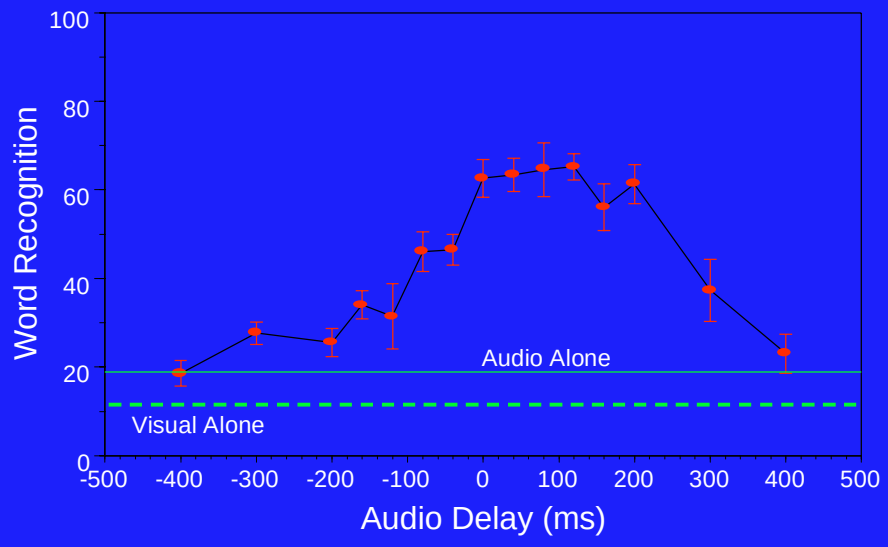
- Recognition of McGurk pairs
  - Audio /pa/, /ba/, /ta/, /da/
  - Video /ka/, /ga/, /ta/, /da/
- Synchrony identification and discrimination
  - Yes/No single interval simultaneity judgments
  - 2IFC adaptive tracking

Auditory-Visual Asynchrony - Paradigm

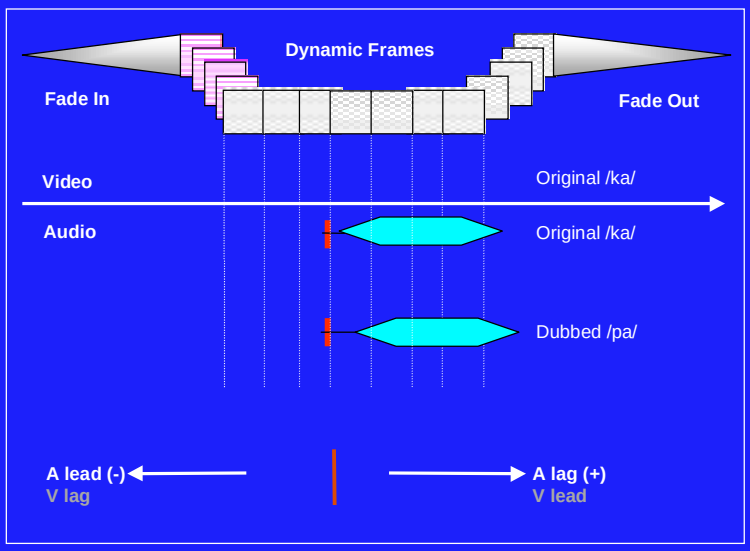
Video of spoken (Harvard/IEEE) sentences, presented in tandem with a sparse spectral representation (low- and high-frequency slits) of the same material

Baseline Condition
SYNCHRONOUS A/V

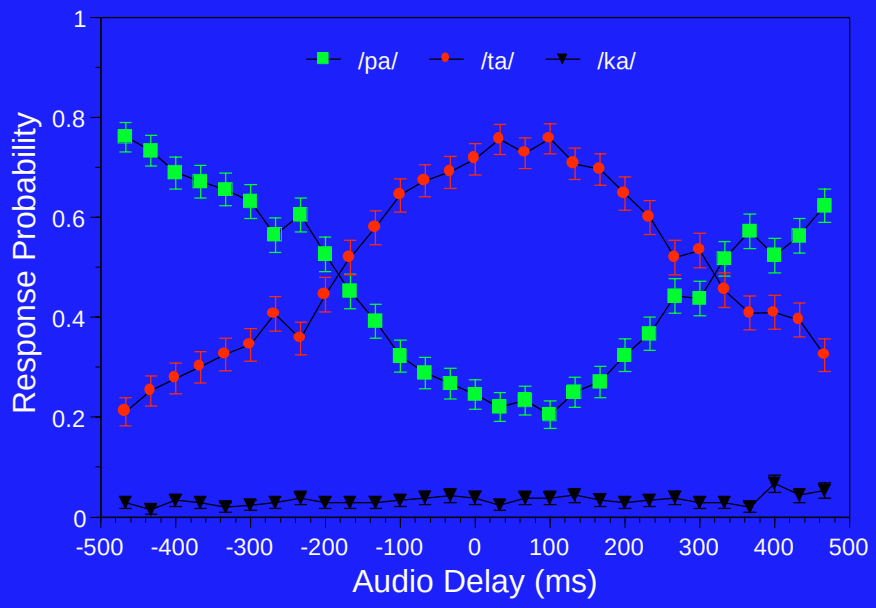
IEEE Sentence Recognition - % Words Correct

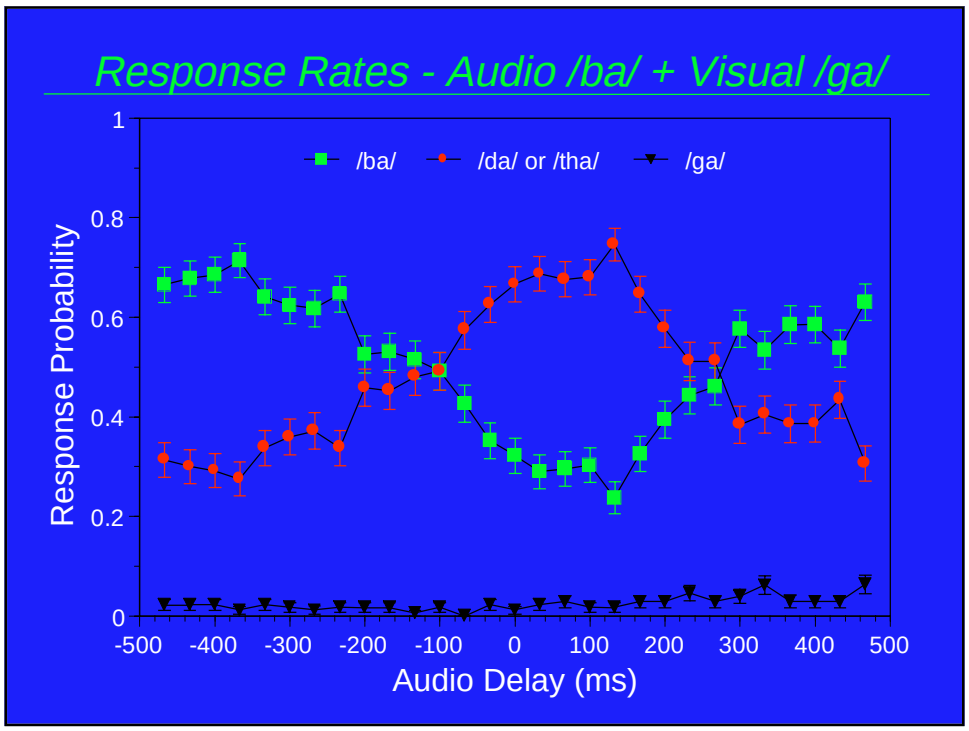


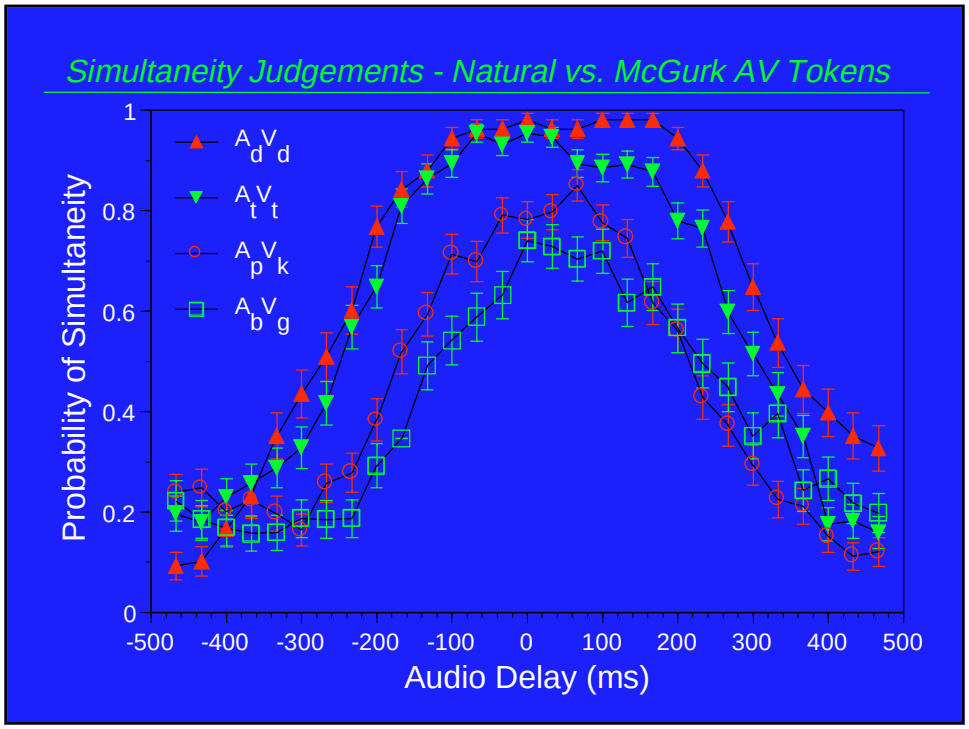
STIMULUS ALIGNMENTS



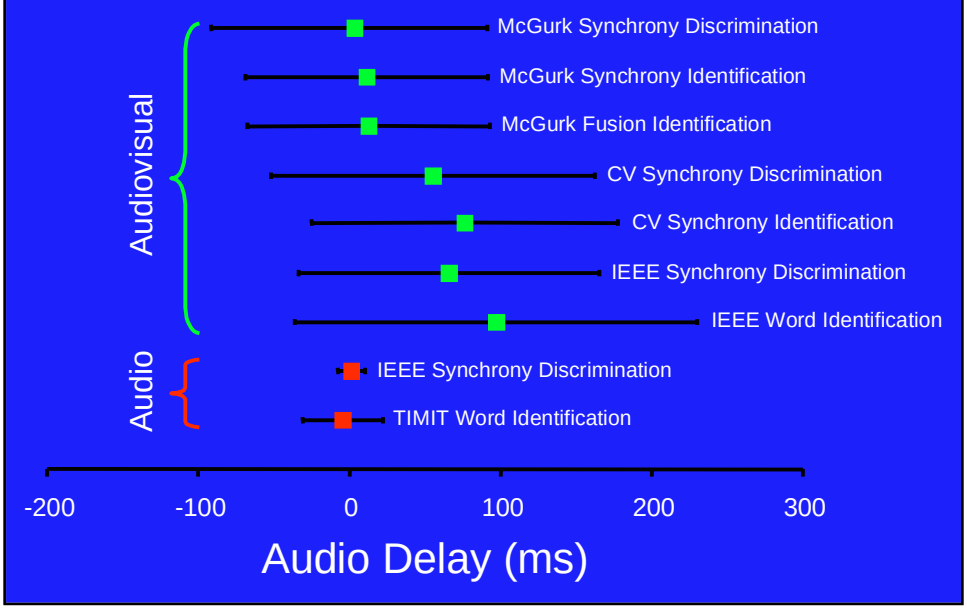
Response Rates - Audio /pa/ + Visual /ka/




Response Rates - Audio /ba/ + Visual /ga/
/ba/
/da/ or /tha/
/ga/
Response Probability
1
0.8
0.6
0.4
0.2
0
-500
-400
-300
-200
-100
0
100
200
300
400
500
Audio Delay (ms)


Simultaneity Judgements - Natural vs. McGurk AV Tokens
A_dV_d
A_tV_t
A_pV_k
A_bV_g
Probability of Simultaneity
1
0.8
0.6
0.4
0.2
0
-500
-400
-300
-200
-100
0
100
200
300
400
500
Audio Delay (ms)

Temporal Window of Integration



Conclusions

Within Modality (A)

- TWI is symmetrical
- TWI roughly 50 ms or less (phoneme?)

Across Modality (AV)

- TWI is highly asymmetrical favoring visual leads
- TWI is roughly 160-220 ms (syllable?)
- TWI for Incongruent CV's (McGurk Stimuli) is not as wide as TWI for natural congruent CV's

Conclusions

Two Temporal Windows of Integration

- Short (< 50 ms) when acoustic signal is first
- Long (≈ 200 ms) when visual signal is first

These results are consistent with Hirsh's view of different perceptual phenomena associated with distinct ranges of time intervals:

< 50 ms for within channel integration of acoustic elements

> 150 ms for cross channel integration of multimodal events