

# Temporal Integration in the McGurk Effect

Virginie van Wassenhove<sup>1</sup>; Ken W. Grant<sup>2</sup>; David Poeppel<sup>1</sup>

<sup>1</sup> Cognitive Neuroscience of Language Laboratory  
Biology Department  
University of Maryland,  
College Park, MD, USA

<sup>2</sup> Auditory-Visual Speech Recognition Laboratory  
Army Audiology and Speech Center  
Walter Reed Army Medical Center  
Washington, DC, USA

## INTRODUCTION

### The McGurk Effect

The McGurk effect first reported by McGurk and McDonald (1976) is used to examine multisensory integration in speech. In its "fusion" component, the illusion emerges when a participant is presented with an auditory bilabial (e.g. /pa/) dubbed onto a visual velar (e.g. viseme /ka/). Under these conditions participants consistently report hearing an alveolar /ta/, a virtual percept resulting from the AV fusion.

### A Case of Multisensory Integration

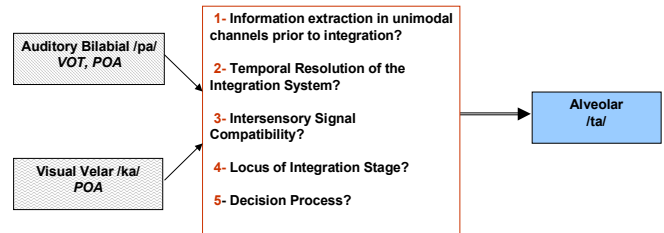
Sources of information originating from a common event (e.g. AV utterance) must share cues to ensure multisensory information binding at the integration stage. Spatial and temporal coincidence prevail as fundamental constraints on the integration process – i.e. signals in close temporal and spatial proximity are more readily bound into a perceptual unit.

Thus, *large timing discrepancies between sensory modalities should intuitively reduce the probability for bimodal information to be fused as a single event.*

### Sources of Information in Bimodal Speech:

- Place of Articulation (POA): primarily provided by the visual modality ("visemes") but also present in the auditory signal (F2/F3 formants transitions).
- Voicing (VOT): entirely provided by the auditory signal.

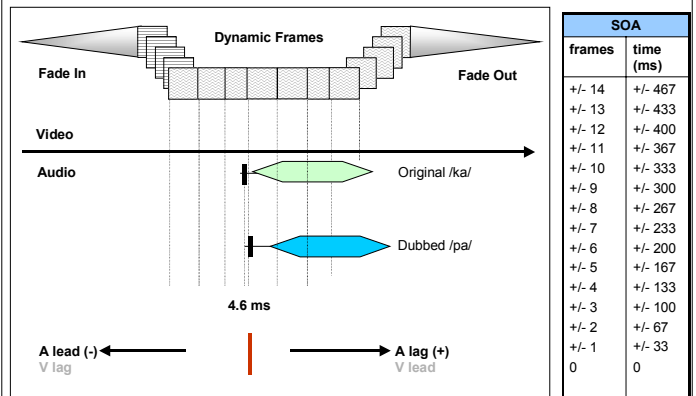
### Issues in Bimodal Speech:



## WORKING HYPOTHESES PSYCHOPHYSICS

- Based on previous research (Massaro *et al.*, 1996, Munhall *et al.*, 1996), desynchronization of AV signal is predicted to lead to fusion rate decrement. However, no temporal boundaries have yet been defined.
- Inter-individual differences in fusion rate are expected based upon the intrinsic population variability.

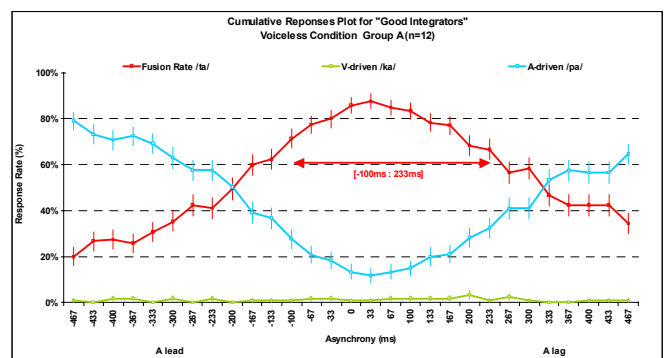
## STIMULI ALIGNMENTS



## PSYCHOPHYSICS

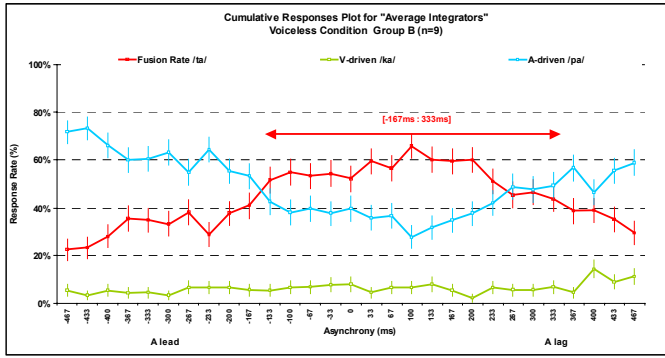
- AV pairs tested:**
  - Congruent Pair: AV /ta/
  - Incongruent Pair: A/pa/ dubbed onto V/ka/ ( $A_p V_k$ )
- 29 AV timings:** [-467ms ; +467ms] in increments of 33.33ms (+/- 1 frame increment)
- 2 Tasks:**
  - Identification Task (3 AFC):**  $A_p V_k$  KA, PA, TA ?
  - Temporal Judgment Task (2 AFC):**
    - $A_p V_k$  Simultaneous,
    - $A_t V_t$  Successive?

## IDENTIFICATION TASK - Group A (n=12)



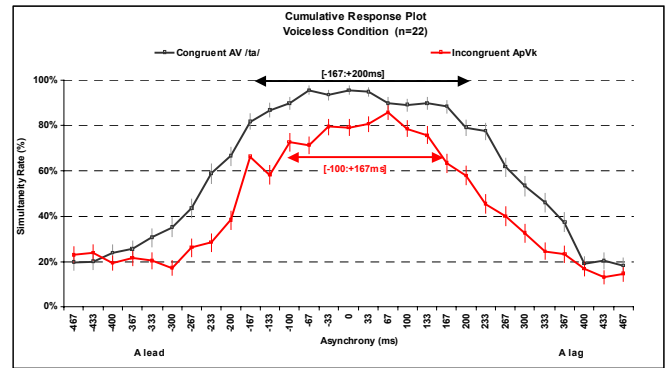
- Grouping based on synchrony condition: fusion rate >70%
- TWI determined with 90% confidence,  $p < 0.1$

## IDENTIFICATION TASK - Group B (n=9)



- Grouping based on synchrony condition: 40% < fusion rate < 70%
- TWI determined with 90% confidence,  $p < 0.1$

## TEMPORAL JUDGMENT TASK (n=22)



- TWI determined with 90% confidence,  $p < 0.1$

## PSYCHOPHYSICS FINDINGS

- An optimal fusion rate remains robust within  $[-100\text{ms}; +233\text{ms}]$  of asynchrony for "good integrators". The **temporal window of integration (TWI)** hereby defined suggests a resolving power of the AV integration process constrained by the auditory processing window of 250ms proposed by Poeppel (2001).
- This **TWI is off-centered**, favoring visual leads for higher fusion rate. Similar results were found in connected speech (Grant, 2001).
- **Inter-individual differences** reliably separate 2 groups of participants according to their optimal fusion rate. This differentiation may originate from the individual's dominant - or preferred - modality (Giard and Peronnet, 1999) or more specifically from the individual's lip-reading ability.
- **"Real" and "illusory" /ta/ can be differentiated** on the basis of the subjective simultaneity task, suggesting that unimodal information is preserved beyond the integration stage or that temporal matching across modalities occurs prior to integration.

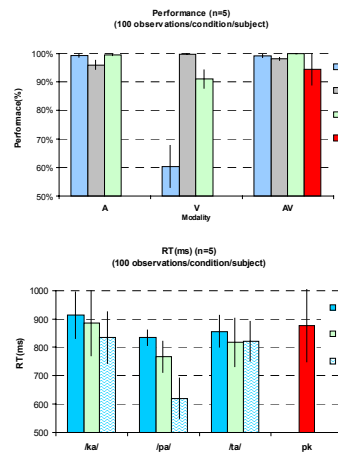
## WORKING HYPOTHESES ELECTROPHYSIOLOGY (EEG)

- Behaviorally, a facilitation effect is expected such that reaction time to AV stimuli be faster than to unimodal conditions.
- In addition to earlier studies showing the involvement of auditory cortices in AV integration (see Calvert, 2001 for review) the temporal window of integration established in the psychophysics suggests that early AV interactions might impact the N1/P2 complex when compared to the auditory condition alone.
- The electrophysiological profile for a congruent /ta/ should significantly differ from that of an illusory /ta/ in light of the differential profile of simultaneity judgment to temporal asynchronies.

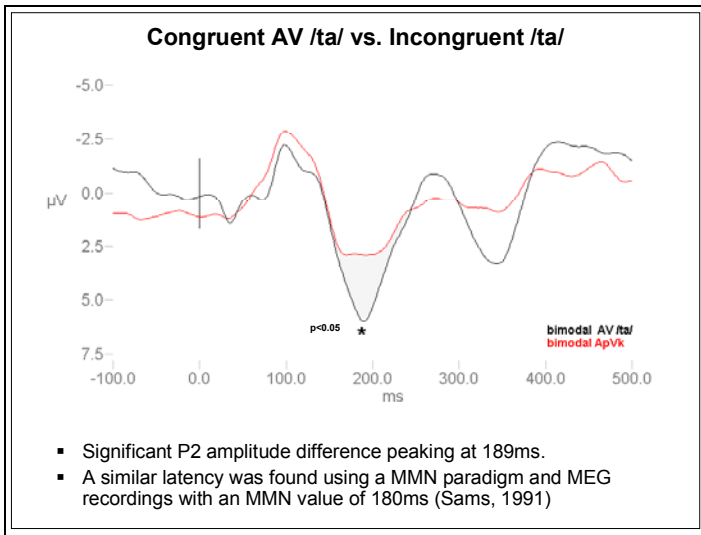
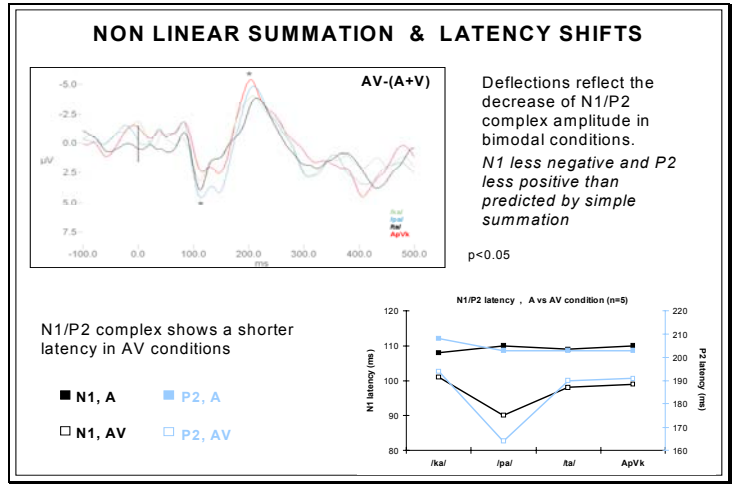
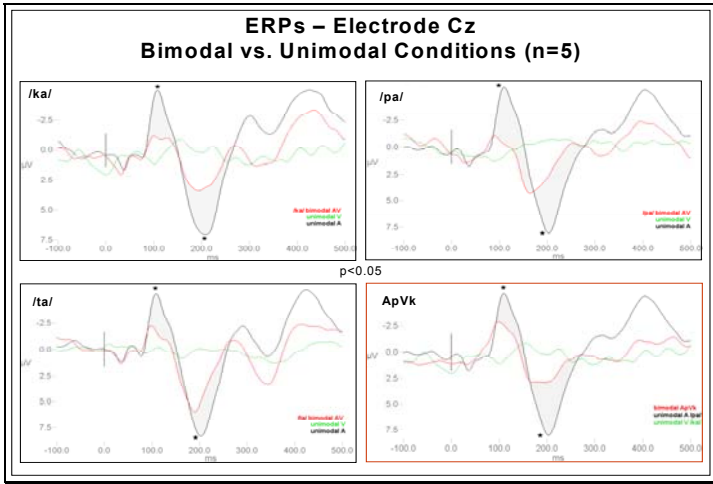
## ELECTROPHYSIOLOGY (EEG)

- **Stimuli:**  
Unimodal Block (A, V): /ka/, /pa/, /ta/  
Bimodal Block (AV): /ka/, /pa/, /ta/ and incongruent  $A_pV_k$   
Each stimulus presented 100 times.
- **Task (3AFC):**  
Identify A, V or AV token as /ka/, /pa/ or /ta/
- **Recording Settings:**  
32 channels Electro Cap (ref. left and right mastoids)  
A/D Rate : 1kHz  
AC recording, Band Pass filtering: 1Hz – 100Hz  
Gain: 1000
- **ERPs Analysis:**  
Only correct epoch considered  
Baseline corrected 500 ms prior to stimulus onset  
Zero-phase-shift (2 passes) Butterworth low-pass filter 30Hz, 24dB roll-off  
Artifact rejection threshold : +/- 50uV

## BEHAVIORAL RESULTS (n=5)



- Ceiling Performances except for visual /ka/.
- Participants belong to "good integrators" category (fusion rate >90%).
- RT facilitation not significant.
- Bilabial /pa/ is much faster in V condition alone.



### CONCLUSIONS

- Early interactions between auditory and visual inputs are reflected in the N1/P2 complex.
- The latency and amplitude of P2 significantly (p<0.05) differ across AV utterances, consistent with a syllabic encoding time in the order of 250 ms and in agreement with the hypothesized TWI found in psychophysics.
- A significant P2 amplitude difference between “real” /ta/ and “illusory” /ta/ was shown, consistent with previous findings by Sams (1991) - although no MMN paradigm was used.

### FURTHER DIRECTIONS

- Visual interactions with early auditory processing are consistent with psychophysical results showing that a high fusion rate can be elicited up to 100ms of auditory lead.
- The notion of “supra-additivity” introduced by Stein and Meredith (1993) is not observed in the ERPs. Supra-additivity of multisensory neurons in adjacent polysensory areas (Tpt, pSTP) might interfere with early auditory processing, such that the recorded N1/P2 complex now reflects both early auditory and multimodal processing.

### REFERENCES

- Calvert, G.A. (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex* 11, 1110-1123.
- Giard, M.H. and Peronnet, F. (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience* 11(5) 473-490.
- Grant, K.W. and Greenberg, S. (2001) Speech intelligibility derived from asynchronous processing of auditory-visual information. *Proceedings of the Workshop on Audio-Visual Speech Processing (AVSP-2001)*
- McGurk, H. and McDonald, J. (1976) Hearing Lips and Seeing Voices. *Nature* 264, 746-747.
- Massaro, D.W., Cohen, M.M. and Smeele, P.M. (1996) Perception of asynchronous and conflicting visual and auditory speech. *JASA* 100(3), 1777-1786.
- Munhall, K., Gribble, P., Sacco, L. and Ward, M. (1996) Temporal constraints on the McGurk effect. *Perception and Psychophysics* 58 (3).
- Poeppel, D. (2001) Pure word deafness and the bilateral processing of the speech code. *Cognitive Science* 25, 679-693.
- Sams, M. and Aulanko, R. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters* 127, 141-147.
- Stein, B.E. and Meredith, M.A. (1993) *The merging of the Senses*. MIT Press.
- Teder-Sälejär, W.A., McDonald, J.J., Di Russo, F., Hillyard, S.A. (2002) An analysis of audio-visual integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*.

**Research Supported by: NIDCD DC 0463801**

**Contact: Virginie van Wassenhove, vvw@glue.umd.edu**