Neural networks supporting auditory-visual speech:
Evidence from invasive neural recordings in humans

EunSeon Ahn, John Plass, Adrian Rakochi, William Stacey, David Brang
University of Michigan

UNIVERSITY OF MICHIGAN

INTRODUCTION

Late integration models posit that modality specific stimuli are processed in a domain-specific manner in their respective brain regions before their information is passed along and integrated in AV areas.[1,2]

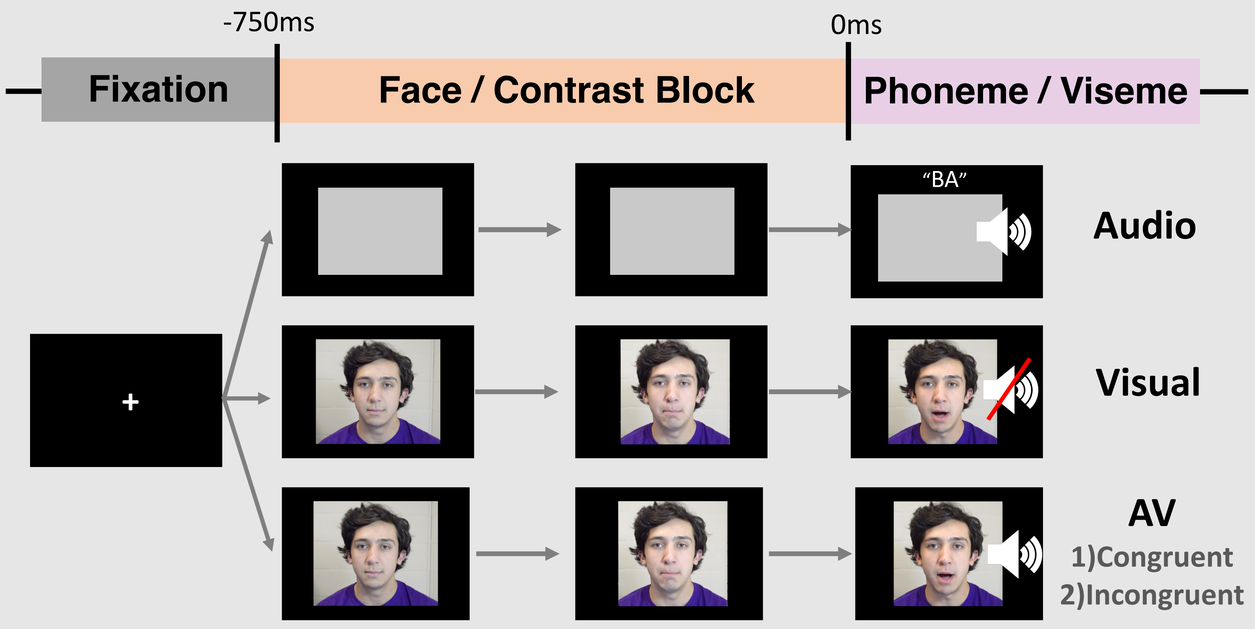
Early integration models, in contrast, suggest that auditory-visual integration occurs in parallel with modality specific processes.

Unlike fMRI, electrocorticographic (ECoG) recordings provide high spatial and temporal resolution well-suited for testing such models.[3]

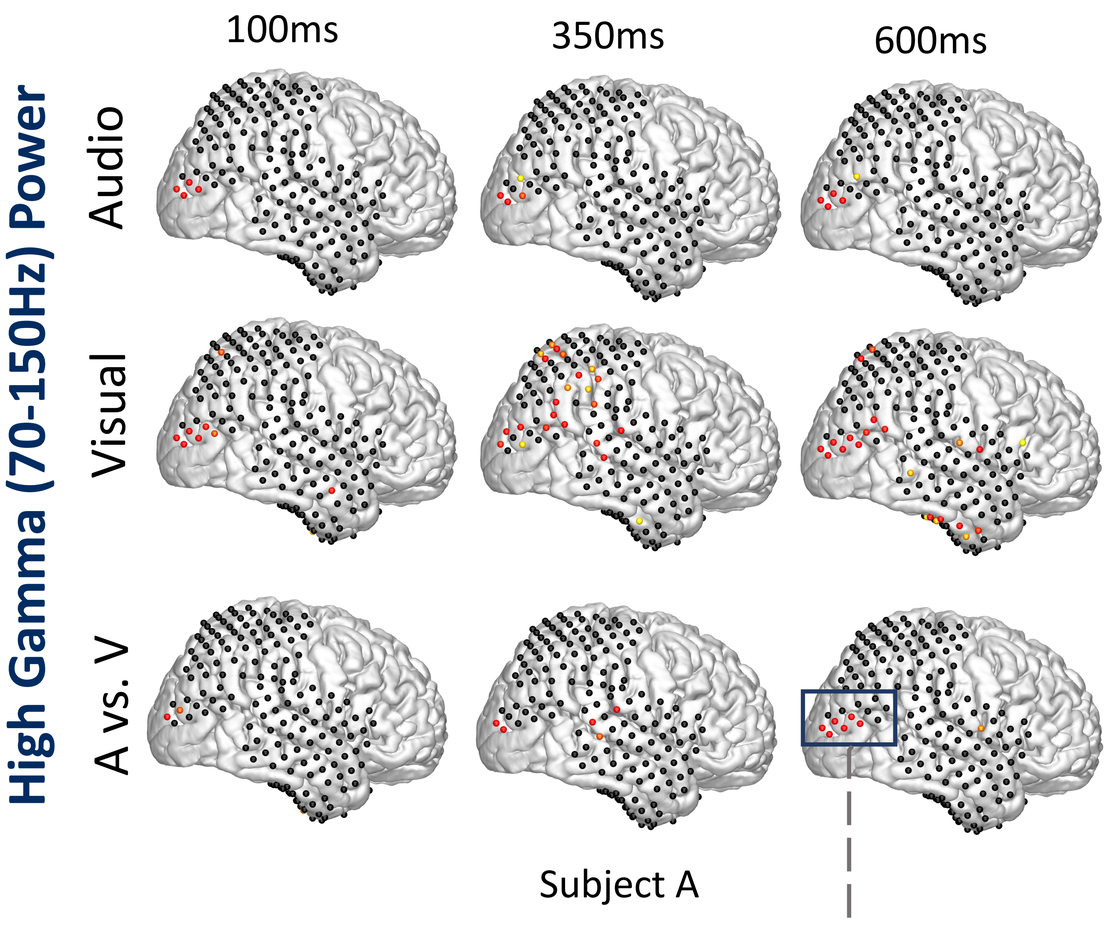
Using ECoG, we looked for evidence that would provide insight into the distribution and timing of auditory-visual speech integration.

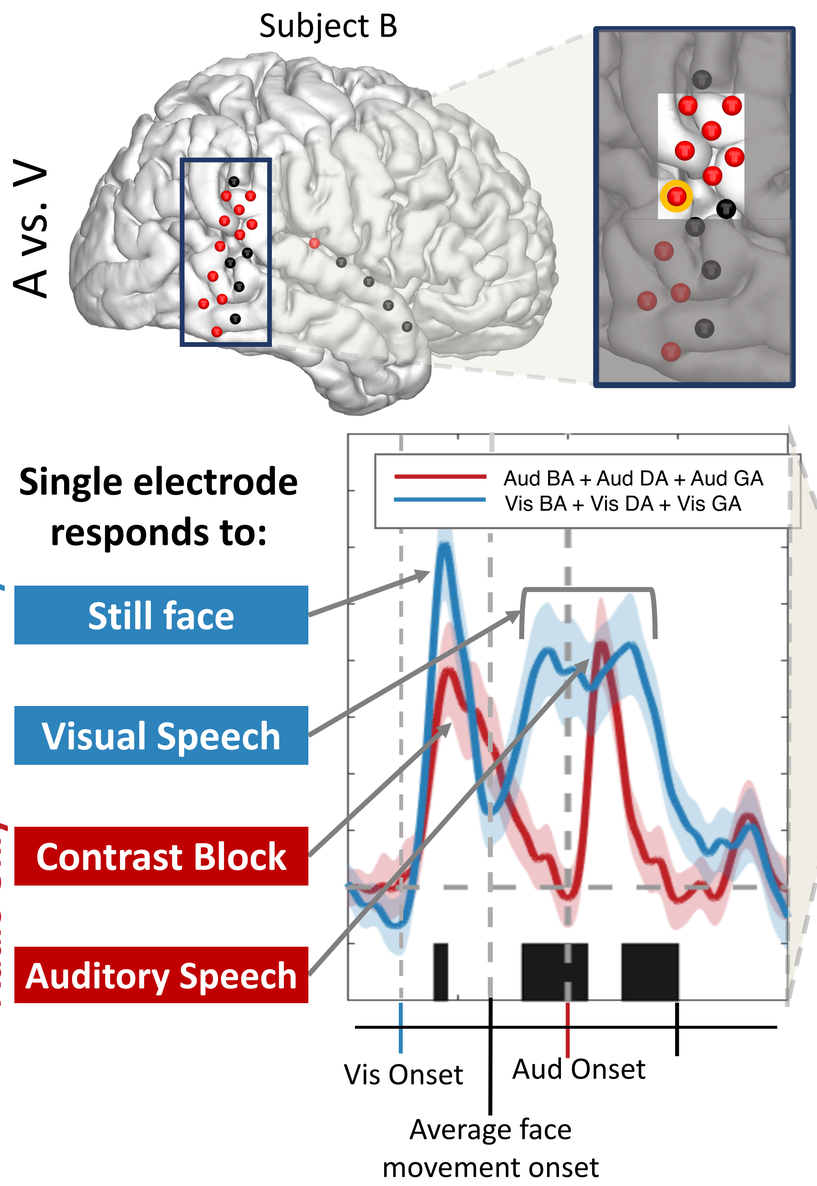
METHODS

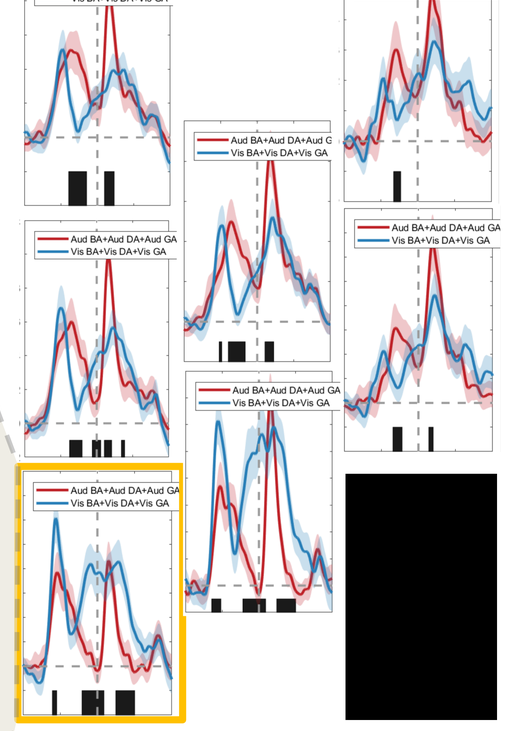
Intractable epilepsy electrocorticography (ECoG) patients



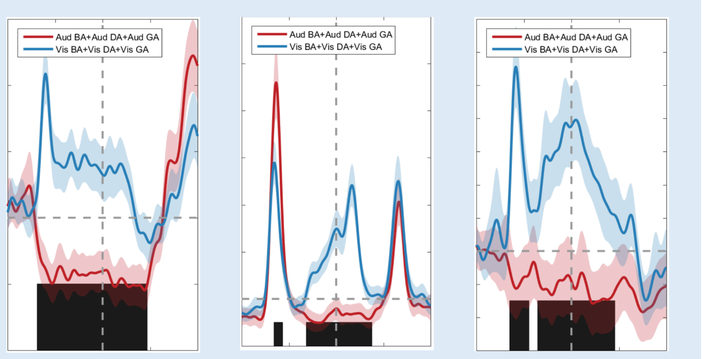
AFTER FACE / CONTRAST BLOCK (BEFORE PHONEME ONSET)



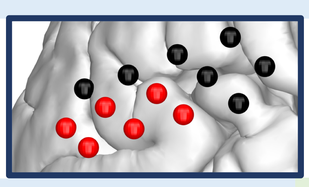


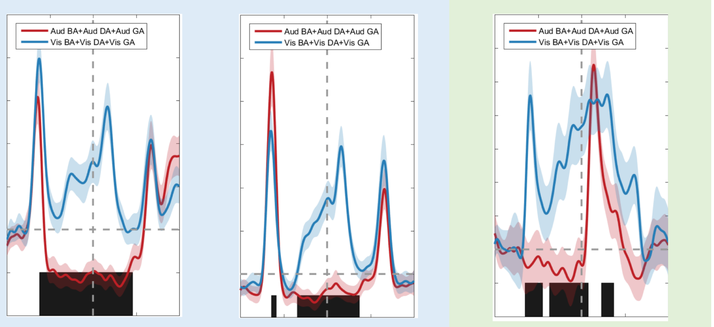


- This "multisensory" electrode in pSTS region is activated by the initial presentation of both visual and auditory stimuli (separately), suggesting non-modality specific processing of unisensory stimuli



Visual Only
Audio Only





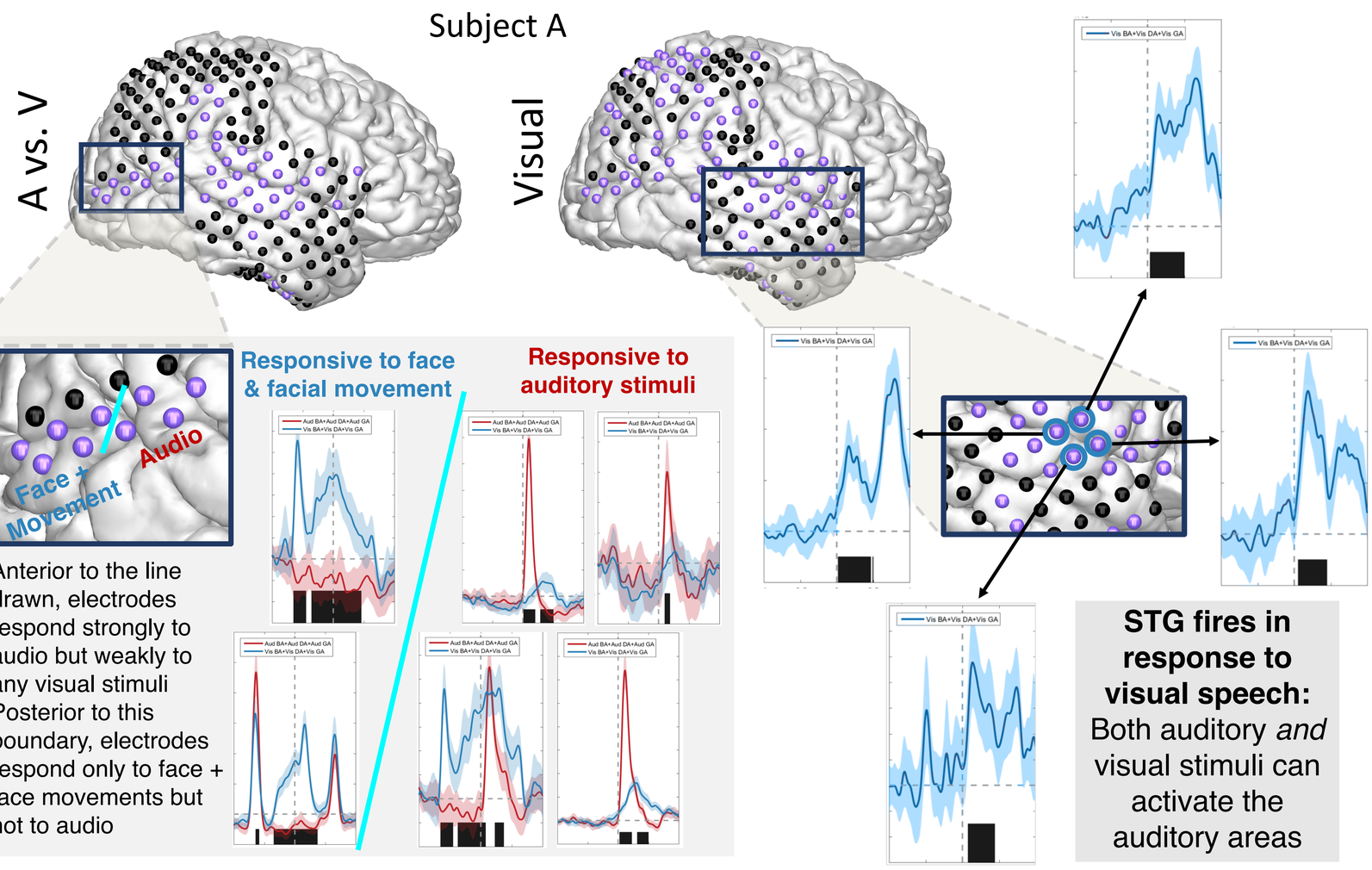
Different patterns of activity for contrast block vs. face (+ movements)

- Sustained activity for faces + face movement
- Responds to onset and offset of contrast block

"Multisensory" Electrode

- Activated by face + face movement
- Activated by audio

AFTER PHONEME / VISEME



- Anterior to the line drawn, electrodes respond strongly to audio but weakly to any visual stimuli
- Posterior to this boundary, electrodes respond only to face + face movements but not to audio

STG fires in response to visual speech:
Both auditory *and* visual stimuli can activate the auditory areas

DISCUSSION

Data suggest:

1.) AV processing occurs in parallel with the initial processing of the stimuli

2.) Distinct audio and visual information processing is not restricted to their respective modality specific brain areas.[4,5]

3.) There is a functional boundary near/in pSTS (which in previous literature has shown a distinction between noisy vs. clean AV speech[6]) to audio-alone vs. visual-alone speech.

⬇

AV processing and integration likely occurs in a parallel, non-serial manner much earlier in the processing hierarchy than is suggested by the late integration model.

REFERENCES

1. Grant KW, Walden BE, Seitz PF. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. Journal of the Acoustical Society of America. 1998;103:2677–2690.
2. Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *68*, 169–181.
3. Parvizi, J., & Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. *Nature neuroscience*, 1.
4. Zweig, L. J., Grabowecky, M., Suzuki, S., Towle, V., Tao, J., Wu, S., & Brang, D. (2016). Silent lip reading generates speech signals in auditory cortex. *Journal of Vision*, *16*(12), 463-463.
5. Brang, D., Towle, V. L., Suzuki, S., Hillyard, S. A., Di Tusa, S., Dai, Z., ... & Grabowecky, M. (2015). Peripheral sounds rapidly activate visual cortex: evidence from electrocorticography. *Journal of neurophysiology*, *114*(5), 3023-3028.
6. Ozker, M., Yoshor, D., & Beauchamp, M. S. (2018). Converging Evidence From Electrocorticography and BOLD fMRI for a Sharp Functional Boundary in Superior Temporal Gyrus Related to Multisensory Speech Processing. *Frontiers in human neuroscience*, *12*.

This study was supported by NIDCD R00 DC013828.