Hearing lips and seeing voices in the brain: motor mechanisms of speech perception

Jeremy I. Skipper[1, 2], Virginie van Wassenhove[3], Howard C. Nusbaum[1], & Steven L. Small[1, 2]

[1]*Department of Psychology,* [2]*Department of Neurology; The University of Chicago, Chicago, Illinois, USA*

[3]*Linguistics Department; University of Maryland, College Park, Maryland, USA*

Introduction

- How does visual information about place of articulation affect the neural mechanisms that mediate audiovisual (AV) speech perception?
- Motor theories of speech perception propose that activity in the motor system mediates recognition of acoustic patterns (e.g., Stevens & Halle, 1967) .
- Does fMRI activation from an incongruent AV stimulus, (i.e., the McGurk fusion effect; McGurk & MacDonald, 1976) differ from congruent AV counterparts in cortical areas that subserve processing in both speech perception and production?
- If speech production areas are involved in decoding observed mouth movements, these areas might work harder (i.e., be more active) to process an AV pairing in which acoustics and visual information disagree.

Methods

- 21 right handed native English speakers; fMRI at 3 Tesla (GE Medical Systems, Milwaukee, WI) with spiral acquisition (TR/TE 1.5/24 ms, 71 degree flip angle, 24 cm FOV, effective resolution 3.8x3.75x3.75mm).
- **Table 1** describes the imaging paradigm:
- Stimuli were 1.5 s long and randomly presented in rapid event-related design with variable ISI.
- Participants "passively" (i.e., no overt motor task to produce activation confounded with hypotheses) listened to and viewed AV /pa/, /ka/, and /ta/ and an incongruent AV stimulus (acoustic /p/ with visual /k/).
- The incongruent AV stimulus (/pk/) is perceived as /ta/ out of the scanner (van Wassenhove et al., 2001, Submitted).
- After scanning, participants indicated which stimulus was heard most frequently in the prior "passive" run: consistent perception of /pk/ as /pa/, /ka/, or /ta/ should result in hearing that stimulus twice as often as others. Second, another AV condition (aAV) was run using 3AFC as /pa/, /ka/, or /ta/ responses. These tasks provided a behavioral measure of perception of /pk/ and recognition of /pa/, /ka/, or /ta/ in the (noisy) scanning environment.
- In a final scan, participants spoke /pa/, /ka/, or /ta/ in 12-second blocks to image brain regions active in producing syllables.
- Individual participant's images were analyzed using AFNI (Cox, 1996) and SUMA programs (Saad et al., 2004). Deconvolution/regression analysis (Ward, 2001) revealed significant voxels for individuals. After cortical surfaces were reconstructed (i.e., flattened; Fischl et al., 1999a) and registered to a template of average curvature (Fischl et al., 1999b), surface-based ANOVA and conjunction analysis (i.e., revealing brain activation common to tasks and excluding activation absent in one or more tasks; Price & Friston, 1997) revealed significant activation across individuals.
- Frontal regions of interest (ROIs; see **Figure 1**) were created with both Freesurfer's automated parcellation (Fischl et al., 2003) and hand drawn regions and used to perform analyses, including discriminant analysis (Haxby et al., 2001) in those regions overlapping speech production.
- Unless otherwise stated, all imaging data is significant at $p < .05$ corrected using a false discovery rate algorithm (FDR; Genovese et al., 2002).

Results

1. 13 participants perceived the McGurk effect (/ta/), 4 perceived /ka/, and 4 perceived /ka/ or /ta/ when presented /pk/. Unless otherwise noted only the /ta/ group is discussed (see **Table 2**).
2. Congruent AV stimuli activated regions involved in speech production: there was extensive overlap between perceiving AV stimuli (excluding /pk/) and speaking the syllables (see **Figure 2**).
3. Activation associated with congruent AV perception was similar to that elicited by /pk/ perception:
   a. All AV (including /pk/) stimuli showed a distributed (see **Figure 3a**) but overlapping (see **Figure 3b and 3c**) pattern of activation in motor areas.
   b. Consistent with these results, and contrary to our prediction, /pk/ did not produce greater activation when linearly contrasted with the other AV stimuli in motor regions (e.g. 3(/pk/) ≠ /pa/+/ka/+/ta/).
4. /pk/ did, however, produce greater activation compared to the individual AV stimuli:
   a. Increases were in all ROIs: regions overlapping speech production (see **Figures 1 and 2**)
   b. Discriminant analysis showed that the distribution of activity for /pk/ cannot be distinguished (in inferior frontal, premotor, and primary motor cortex) from other AV stimuli, despite that /pa/, /ka/, or /ta/ could easily be distinguished. Even in those regions where /pk/ was distinguishable, it was so only at a reduced probability relative to the other stimuli.
   c. Results suggest that /pk/, does not just increase brain activity involved in speech motor areas, but results in a distributed pattern of activity that is similar to one of the other stimuli.
5. Preliminary results suggest that /pk/ activation is most strongly correlated with activation produced by /ta/ in right motor areas.
   a. /pk/ was most correlated with /ta/ in 6/10 of the analyzed frontal regions including primary motor cortex (see **Figure 4**).
   b. In left motor areas, /pk/ was correlated with /ka/ in two areas and with /pa/ in two areas.
   c. These patterns of activation seem to change over time in meaningful ways before "settling" (see **Figure 5a**).
   d. For the group of participants who perceived the /pk/ stimulus as /ka/, frontal regions are most correlated with /ka/ (see **Figure 5b**).
6. Though our analyses have focused on motor areas involved in speech production:
   a. Many strong effects of /pk/ were in visual regions including both the calcarine sulcus and fusiform gyrus (see also **Figure 3c**): preliminary results suggest that early /pk/ activation in the calcarine sulcus is most strongly correlated with AV /ka/ (the visual portion of the /pk/ stimulus) but as time progresses becomes increasingly more correlated with AV /ta/ (the participant's perception).
   b. Posterior STG did not show increased activation for /pk/, nor could it be distinguished from the other syllables with discriminant analysis, and activation correlated with /ta/ over time (see **Figure 6a and 6b**). The supramarginal gyrus, however, did show increased activation and could be distinguished though activation patterns were most strongly correlated with /ta/.

Discussion

- Results show that the motor system plays a significant role in audiovisual speech perception.
- Results also demonstrate that motor system activity predicts participants' perception.
- This is evidence for a model like analysis-by-synthesis (Stevens & Halle, 1967) in which motor activity simulates speech production to generate internal speech representations used for recognition.
- Results also demonstrate that /pk/ affects processing in visual cortex (e.g., V1) in such a way as to suggest that novel audio - visual pairings may "re-tune" visual cortex.
- These results extend those of an earlier fMRI study of the McGurk effect showing that premotor activity plays a significant role in perception and that effects in visual cortex may be earlier than thought (Jones & Callan, 2003).

References

Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers in Biomedical Research, 29*(3), 162-73.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., & Dale, A. M. (2003). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex, 14*(1), 11-22.

Fischl, B., Sereno, M.I., & Dale, A.M. (1999b). Cortical surface-based analysis II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage, 9*:195-207.

Fischl, B., Sereno, M.I., Tootell, R.B.H., & Dale, A.M. (1999b). High-resolution inter-subject averaging and a coordinate system for the cortical surface. *Human Brain Mapping, 8*, 272-284.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15*(4), 870-878.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science 293*(5539), 2425-2430.

Jones, J. A. & Callan, D. E. (2003). Brain activation during audiovisual speech perception: An fMRI study of the "McGurk effect". *NeuroReport, 14*, 1129-1133.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(588), 746-748.

Price, C. J. & Friston, K. J. (1997). Cognitive conjunction: a new approach to brain activation experiments. *Neuroimage, 5*(4 Pt 1), 261-270.

Saad, Z.S., Reynolds, R.C., Argall, B., Japee, S., & Cox, R.W. (2004). SUMA: An Interface For Surface-Based Intra- And Inter-Subject Analysis With AFNI. Proceedings of the 2004 IEEE International Symposium on Biomedical Imaging, in press.

Stevens, K. N. & Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed), *Models for the perception of speech and visual form*. Cambridge, MA: MIT Press

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). *Timing of auditory-visual integration in the McGurk effect.* Paper presented at the Proceedings of the annual meeting of the Society for Neuroscience, San Diego, CA.

Ward, B. D. (2001). *Deconvolution Analysis of FMRI Time Series Data* (Technical Report): Milwaukee, Wisconsin: Biophysics Research Institute, Medical College of Wisconsin.

(pdf: http://www.home.uchicago.edu/~skipper/papers/poster.html)

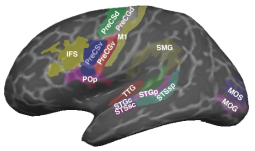

Figure 1: Guide to the location of brain regions on a flattened hemisphere. Frontal regions shown are the same used in ROI analyses. Abbreviations: IFS, inferior frontal sulcus; LH, left hemisphere; M1, primary motor cortex; MOG, middle occipital gyrus; MOS, middle occipital sulcus; PreCGd, dorsal precentral gyrus; PreCGv, ventral precentral gyrus; PreCSd, dorsal precentral sulcus; PreCSv, ventral precentral sulcus; POp, pars opercularis of the inferior frontal gyrus; RH, right hemisphere; SMG, supramarginal gyrus; STGa, central superior temporal gyrus; STGp, posterior superior temporal gyrus; STSsc, central portion of the upper bank of the superior temporal sulcus; STSsp, posterior portion of the upper bank of the superior temporal sulcus; TTG, transverse temporal gyrus.

Functional Scan #	Condition	Stimuli Presented	Task
Randomly Presented	Audiovisual (AV)	/pk/, /pa/, /ka/, and /ta/	Watch and Listen to Speaker
Randomly Presented	Audio-alone (A)	/pa/, /ka/, and /ta/	Listen to Speaker
Randomly Presented	Video-alone (V)	/pa/, /ka/, and /ta/	Watch Speaker
4	Audiovisual (AV)	/pa/, /ka/, and /ta/	Watch and Listen to Speaker
Frequency Judgment: "Which stimulus did you hear most frequently: /PA/, /KA/, or /TA/?"			
5	"Active" Audiovisual (aAV)	/pk/, /pa/, /ka/, and /ta/	Watch and Listen to Speaker: classify syllable as /PA/, /KA/, or /TA/ with 3 Alternative Forced Choice (3AFC)
6	Speaking (S)	/pa/, /ka/, and /ta/	Read: /PA/, /KA/, or /TA/ and repeat out loud

Table 1: fMRI scanning sequence and task descriptions (/pk/ = acoustic /p/ with visual /k/)

	3AFC Response During Functional Scan # 5 (%)							Frequency Judgment (%)		
Group	/pa/=/pa/	/pa/=/ka/	/pa/=/ta/	/pk/=/pa/	/pk/=/ka/	/pk/=/ta/	/pk/=/ta/	/pa/	/ka/	/ta/
PA	0	0	0	0	0	0	0	0	0	0
TA (N=13)	97	82	96	2	15	83	23	15	62	
KA (N=4)	100	96	92	4	76	20	0	50	50	
KA/TA (N=4)	100	94	96	2	47	51	25	75	0	
Mean	99	91	95	3	46	51	16	47	37	

Table 2: 13/21 participants perceived the McGurk effect, /ta/, when presented /pk/. Responses when asked to classify /pk/, /pa/, /ka/, and /ta/ as /pa/, /ka/, or /ta/ using 3AFC are sorted by group as indicated by hierarchical clustering. The latter resulted in a /ta/ (i.e., the classic McGurk effect) and /ka/ (i.e., visual capture) group and a group who responded with equal likelihood /ta/ and /ka/.

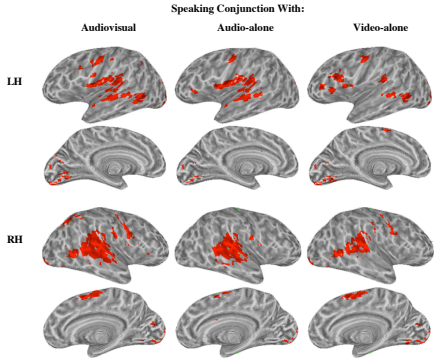

Figure 2: Perception of congruent AV stimuli overlapped speech production areas. Conjunction of activation for audiovisual (AV), audio-alone (A), and visual-alone (V) conditions with speaking (S) showing regions used to produce and perceive syllables. Overlapping frontal regions in the AV condition are the insula, superior frontal gyrus, PreCGd, PreCGv, PreCSd, PreCSv, IFS, POp, and M1 (see **Figure 1** for abbreviations). Note the relative lack of overlap between S and A but not S and V.

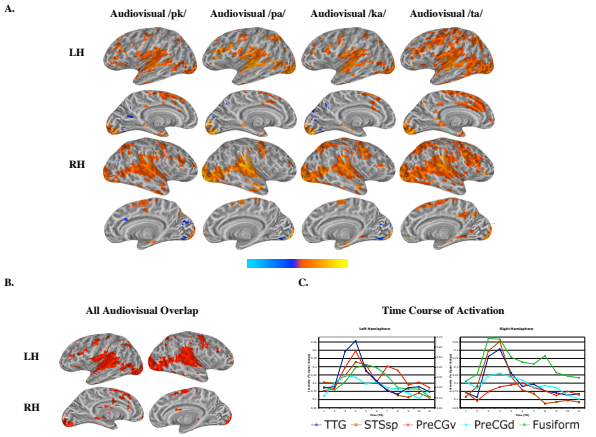

Figure 3: Distribution of activation associated with AV stimuli was similar to that elicited by /pk/. A. Activation for AV /pk/, /pa/, /ka/, and /ta/ showing the distributed nature of processing AV stimuli ($p<.05$ corrected; cool colors (blue) and warm colors (yellow) indicate negative and positive activation respectively). **B.** Conjunction of activation for AV /pk/, /pa/, /ka/, and /ta/ highlighting overlapping regions. **C.** Time course of activation for AV stimuli: though statistical significance has not yet been determined, qualitatively the relative time course for motor (and visual) activation (relative to activation in primary auditory cortex) is longer.

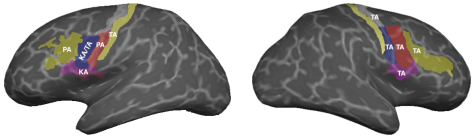

Figure 4. /pk/ was most strongly correlated with /ta/ in 6/10 of the analyzed frontal regions: Pairwise correlations were performed on mean corrected coefficients from active voxels across AV stimuli between /pk/ and the other AV stimuli collapsed over time. Shown in each region is the strongest correlation between /pk/ and /pa/, /ka/, or /ta/ as determined by Friedman test and a posthoc nonparametric tests for differences between means (Nemenyi). /ta/ consistently showed the greatest rank in PreCGd, PreCGv, PreCSv, IFS, POp, and M1 (see **Figure 1** for abbreviations).

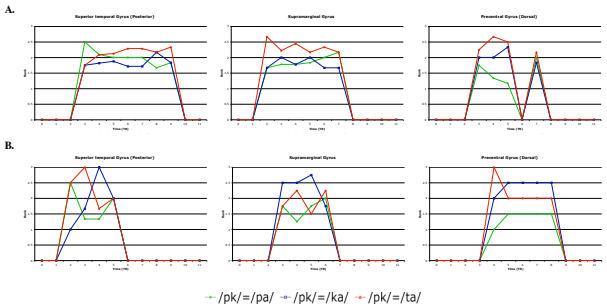

Figure 5. /pk/ was most strongly correlated with /ta/ (see Figure 6) but this pattern changed over time in meaningful ways before "settling": Analysis as described in **Figure 4** but applied to each time point after stimulus onset. The strongest correlation (as determined by Friedman rank) was significantly different from one of the other correlations (as determined by Nemenyi test) in all shown timepoints (with exceptions in A.: PreCGd TRs 6 and 7 and B.: STGp TR 5). **A.** Three exemplary regions in the /ta/ group (see **Table 2**). Note how the STGp pattern of correlation is initially strongest between /pk/ and /pa/ (consistent with the auditory portion of /pk/) but over time becomes more correlated with /ta/ (the participant's perception). **B.** Three exemplary regions in the /ka/ group (see **Table 2**): Note how the STGp pattern of correlation is initially strongest between /pk/ and /pa/Vta/ but over time becomes more correlated with /ka/ (the participant's perception and the visual portion of the stimulus).

Acknowledgements

Bernadette Brogan, E Chen, Shahrina Chowdhury, Fred Dick, Petr Hlustik, Robert Lyons, Daniel Margoliash, Xander Meadow, Mom, David Poeppel, Ziad Saad, Ana Solodkin, Ryan Walsh, and Nicholas Wymbs

NIDCD: Language and Speech Dr. Judith Cooper, Program Officer)

Work supported by NIH RO1-DC03378