A Direct Connection from Auditory to Visual Word Form Areas Bridges the Auditory and Written Speech Networks: Evidence from fMRI and MEG Connectivity

Lillian Chang[1], Philip Cho[1], Srikanth Damera[1], Kimberly Jin[1], Kathleen M. Gates[2]., Lynne E. Bernstein[3], Maximilian Riesenhuber[1]

[1]Georgetown University - Department of Neuroscience, [2]University of North Carolina at Chapel Hill - Department of Psychology, [3]George Washington University - Department of Speech, Language & Hearing Sciences

Introduction

It is well known that the "visual word form area" (VWFA), located in the ventrolateral occipital temporal cortex (vOTC), contains a lexicon for written words[1-4]. This derives from the idea that dedicated representations are learned for commonly encountered visual objects of written language— i.e., real words (RWs). However, it is unclear how the VWFA develops in the vOTC.

We recently provided evidence for the existence of an auditory lexicon in the auditory anteroventral stream[5]: a so-called "auditory word form area" (AWFA) paralleling the VWFA for written words[6]. Given that written words are learned after auditory speech has been learned for most individuals, we hypothesize that the VWFA's development is guided by its corresponding area in the auditory modality where lexical representations have already been learned for spoken words.

We will test the following novel hypotheses using fMRI and MEG: i) the VWFA should respond to auditory words, but the converse does not need to be true (i.e., the AWFA responding to written words); ii) the VWFA directly connects to the AWFA, and not indirectly via a higher-level processing region like the inferior frontal cortex (IFC), which is associated with both reading and auditory speech processing.

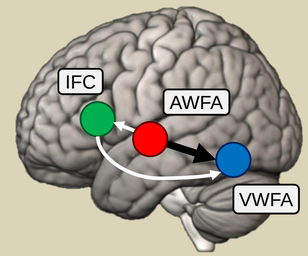


Fig 1. Our hypothesized direct connection from AWFA→VWFA (black arrow) vs. the alternative hypothesis of an indirect connection from AWFA→IFC→VWFA (white arrows).

fMRI Connectivity Analyses using Individual-subject ROIs

Experimental Design

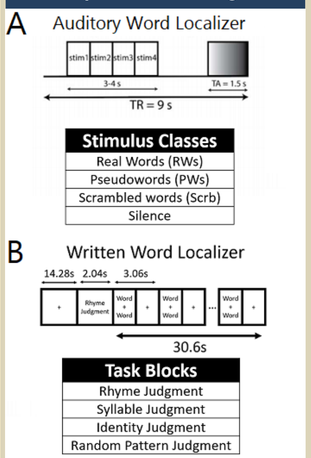


Fig 2. fMRI scans acquired from a 3T Siemens Prisma scanner with 64-channel head coil. 20 right-handed English speakers (13 Females, ages 18-34). **A.** Slow (TR=9s) event-related (ER) paradigm used in the auditory localizer scan. **B.** Block design paradigm used for the written word localizer scan. TR=2s. **C.** Rapid (TR=3.36s) ER paradigm used for the RA scans with RW stimuli. Runs from this scan were used for functional connectivity analyses.

Defining Individual-subject ROIs

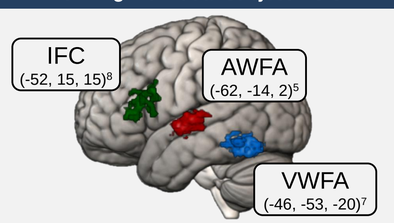


Fig 3. iROIs used in the fMRI connectivity analyses (N=20 per region). iROIs projected to the surface for visualization. Literature coordinates (in MNI) were used as reference locations when choosing clusters for iROIs.

- Individual regions-of-interest (iROIs) were defined using each functional localizer based on published methods[5,8] using SPM12.
  - Written word localizer: VWFA and IFC.
  - Auditory word localizer: AWFA and IFC.
- We prioritized IFC defined using the auditory word localizer. For N=2 without one, we substituted with the written word localizer-defined IFC.
- Cluster sizes (mean±std): VWFA: 28.55±5.3; AWFA: 29.15±4.7; IFC: 26.1±5.3.

Cross-modal Activation Differences

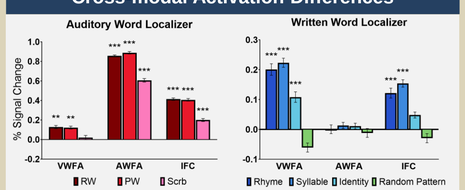
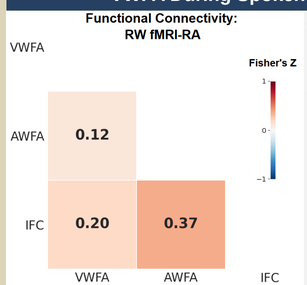


Fig 4. Activation profiles. Error bars indicate within-subject SEM. Right-tailed one-sample t-tests. *p<0.05, **p<0.01, ***p<0.001.

- Spoken words activate the VWFA (RW: p=0.0074, PW: p=0.0047). However, the AWFA shows no activation for written words (all conditions p>0.05).

Significant Functional Connectivity Between AWFA & VWFA During Spoken Word Processing

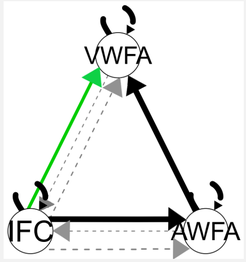


- Connectivity between all ROIs are significantly different from zero (p<0.001).
- Path comparisons:
  - VWFA↔AWFA vs. VWFA↔IFC: p = 0.0563
  - VWFA↔AWFA vs. AWFA↔IFC: p<0.0001
  - VWFA↔IFC vs. AWFA↔IFC: p = 0.0131

Fig 5. Mean functional connectivity. Bolded values indicate significant two-tailed one-sample t-tests (p<0.05).

GIMME Elucidates Direction of Information Flow

Group Iterative Multiple Models Estimation (GIMME) is a model-building algorithm that identifies and estimates time series network models using a structural equation modeling framework and autoregressive techniques. GIMME reliably detects directed, dynamic relations/edges across ROIs/nodes[9,10]. A major strength of GIMME is the option to cluster individuals into subgroups based on similarities in functional connectivity patterns.



- GIMME confirms our AWFA→VWFA hypothesis and elucidates path directions at the group, subgroup, and individual level.
  - CFI>0.97
  - RMSEA and SRMR were predominantly < 0.06.

Fig 6. Effective connectivity results using GIMME. Black = group path; Green = subgroup path; Grey = individual path; solid = lag-0 path; dashed = lag-1 path; circle path = autoregressive).

Confirmatory MEG Analyses from the MOUS Database Support fMRI Results

Experimental Design

To improve on the slow temporal resolution of fMRI and further investigate the neural dynamics between the AWFA, VWFA and IFC, we reanalyzed a large open science data set, the MOUS dataset[11], which offered high temporal resolution MEG data for 102 subjects listening to auditory words.

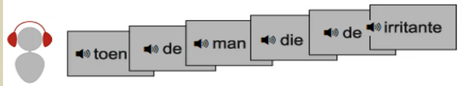


Fig 7: Auditory language task from the Mother of All Unification Studies (MOUS) database. MEG data were collected with a 275-channel axial gradiometer system (CTF) with a sampling frequency of 1,200 Hz. 95 right-handed Dutch speakers (subset of the original 102 MEG subjects) were used for analyses. Figure adapted from Schoffelen et al. (2019)[11].

Another poster from our group also reanalyzed the MOUS dataset to investigate written and spoken word representations!
Nov 12, 8am-12pm
WCC Halls A-C UU20
(PSTR113.18)

MEG Analyses Using Anatomically-restricted ROIs

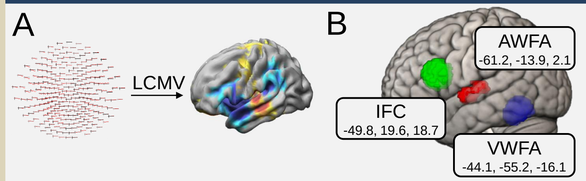


Fig 8. MEG Analyses **A.** Sensor layout and time-courses (left), and snapshot of source-reconstructed dynamics following LCMV beamforming (right). **B.** Leveraging from the iROIs defined using functional localizers, we created anatomically-restricted ROIs (aROIs) using a data-driven method to calculate the optimal ROI size to encompass the variability in iROI locations (e.g. the spread of iROIs shown in Fig 3).

- **Auditory word onset times** were determined using the Montreal ForcedAligner.
- After obtaining the sensor averages, we used linearly constrained minimum (LCMV) **beamforming** to localize neuronal activations. Source models for each individual were obtained by inversely-warping a template source model (5mm dipole spacing) from MNI coordinates to the subject's anatomical MRI.
- **Coherence** between ROIs was calculated at 2.5 ms intervals from -0.2s to 0.5s around the stimulus. Time window was determined by 1/f.

MEG Activation and Coherence Analyses Further Support an AWFA→VWFA Pathway

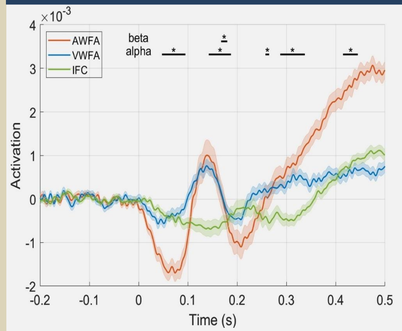


Fig 9. MEG Results. Beamforming activations locked on spoken word onset. Black asterisks and lines indicate significant AWFA↔VWFA coherence (p<0.025).

- Significant coherence between AWFA↔VWFA starting at 40ms in the alpha and 170ms in the beta band frequencies (p<0.025).
- No significant coherence between IFC↔AWFA or IFC↔VWFA (both p>0.05).
- Activation latencies also support a direct interaction from AWFA (12ms) to VWFA (26ms) rather than an indirect route via the IFC, which responded at a later latency (44ms).

Summary

- Using individual-subject ROIs, we found differential cross-modal activation in the VWFA and AWFA—spoken words activate the VWFA, but written words do not activate the AWFA—and significant functional connectivity between the visual and auditory word form areas during spoken word processing.
- An effective connectivity analysis using GIMME further elucidated that there is a direct connection from AWFA to VWFA.
- Coherence analyses of the MOUS MEG dataset identified significant coherence between VWFA and AWFA as early as 40ms after auditory word onset.
  - In contrast, no evidence was found for coherence between VWFA↔IFC or the AWFA↔IFC, further arguing against the alternate hypothesis that the IFC acts as an interface point between the AWFA and VWFA.
- Our results from fMRI activation, functional connectivity, effective connectivity, and supplementing MEG activation and coherence analyses from a separate large database thus support a model in which the AWFA provides direct input to the VWFA, resulting in selectivity of the VWFA to auditory words. This direct input from the AWFA to the VWFA region could serve to guide the learning of the written word lexicon in the VWFA during reading acquisition.

Acknowledgements

- Research supported by the National Science Foundation (BCS-1756313) and Georgetown University.
- Dr. Ian DeWitt for the auditory scrambling code used in the auditory word localizer.

References

1. Glezer, L. S., Jiang, X., & Riesenhuber, M. (2009). Evidence for highly selective neuronal tuning to whole words in the "visual word form area". *Neuron*, *62*(2), 199-204.
2. Glezer, L. S., Kim, J., Rule, J., Jiang, X., & Riesenhuber, M. (2015). Adding words to the brain's visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. *Journal of Neuroscience*, *35*(12), 4965-4972.
3. Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage, 21*(3), 946-953.
4. Lochy, A., Jacques, C., Maillard, L., Colnat-Coulbois, S., Rossion, B., & Jonas, J. (2018). Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. *Proceedings of the National Academy of Sciences, 115*(32), E7595-E7604.
5. Damera, S. R., Chang, L., Nikolov, P. P., Mattei, J. A., Banerjee, S., Glezer, L. S., ... & Riesenhuber, M. (2023). Evidence for a Spoken Word Lexicon in the Auditory Ventral Stream. *Neurobiology of Language, 4*(3), 420-434.
6. DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences, 109*(8), E505-E514.
7. Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *Neuroimage, 22*(1), 466-476.
8. Glezer, L. S., Eden, G., Jiang, X., Luetje, M., Napoliello, E., Kim, J., & Riesenhuber, M. (2016). Uncovering phonological and orthographic selectivity across the reading network using fMRI-RA. *Neuroimage*, *138*, 248-256.
9. Gates, K. M. & Molenaar, P. C. Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. NeuroImage 63, 310–319 (2012).
10. Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K. & Guiskewicz, K. Unsupervised Classification During Time-Series Model Building. Multivar Behav Res 52, 129–148 (2017).
11. Schoffelen, J. M., Oostenveld, R., Lam, N. H., Uddén, J., Hultén, A., & Hagoort, P. (2019). A 204-subject multimodal neuroimaging dataset to study language processing. *Scientific data*, *6*(1), 17.