ACQUISITION OF FOREIGN PHONETIC CONTRASTS MEDIATED BY WHITE MATTER CONNECTIVITY

Francis C. K. Wong[a], *Bharath Chandrasekaran*[b], *Kyla Garibaldi*[a] *& Patrick C. M. Wong*[a,c]

[a]Communication Neural Systems Research Group, the Roxelyn and Richard Pepper, Department of Communication Sciences and Disorders, Northwestern University, USA;
[b]Department of Communication Sciences and Disorders, the University of Texas at Austin, USA;
[c]Hugh Knowles Center for Clinical and Basic Science in Hearing and Its Disorders, Northwestern University, USA
francis@northwestern.edu; bchandra@mail.utexas.edu;
kylagaribaldi2013@u.northwestern.edu; pwong@northwestern.edu

ABSTRACT

Brain imaging studies have shown that both the frontal and the parietal-temporal cortical areas are engaged during speech comprehension. These areas have been theorized to be components (nodes) of a language network. However, research on the intriguing role played by white matter connectivity (connections between nodes) is relatively sparse. In this study, we investigated such neuroanatomical bases that underlie the interplay between phonetic processing and semantic processing. We trained participants with a sound-to-word learning paradigm in which they learned to use a foreign phonetic contrast for signaling word meaning. Using diffusion tensor imaging (DTI), we found that white matter integrity in the left parietal-temporal region positively correlated with performance in sound-to-word learning.

Keywords: diffusion tensor imaging, white matter, MRI, speech learning

1. INTRODUCTION

Speech comprehension requires an integration of acoustic-phonetic inputs into one's semantic system. Brain imaging studies have shown that both the frontal and the parietal-temporal areas are engaged in this process [5, 17]. Among neurocognitive models of auditory language processing [7, 10, 11, 13], Hickok and Poeppel's dual stream model [11] has made explicit predictions about the anatomical bases of the temporal-frontal pathways in the proposed language network, a dorsal stream that maps sound to articulatory representation and a ventral stream that maps sound to meaning.

With the recent advance in brain imaging tools that examine white matter structures, pathways connecting cortical areas, it is now possible to look for precise anatomical correlates of these processing streams as well as to study their functional roles in speech perception and comprehension. The studies of Flöel, et al. [6], Glasser and Rilling [9] and Saur, et al. [14] are examples of such in which techniques from diffusion tensor imaging (DTI) were used.

In DTI, white matter integrity is reflected by a measurement called fractional anisotropy (FA). FA quantifies the overall *directionality of water diffusion* in a neural tissue. The presence of densely packed axons that align with one another would hinder water diffusion in a direction perpendicular to the fiber bundles, hence, making water diffusion more *anisotropic*, i.e. diffusion with a preferential direction [1].

In the present study, we aimed to investigate white matter structures that mediate the interplay between phonetic processing and semantic processing. We trained native English speakers to use pitch patterns, acoustically signaled by changes in fundamental frequency, to contrast meanings. For example, participants learned to associate '*pesh*' presented with a falling pitch pattern with a picture of a table, and '*pesh*' presented with a rising pitch pattern with a picture of a pencil. To successfully learn these sound-picture pairings, participants had to be sensitive not just to the segmental features that came naturally to them based on their native language, but also to the novel supra-segmental features that were not used in their native language. Using DTI, we identified brain regions where white matter integrity mediated learning success.

2. MATERIALS AND METHODS

2.1. Participants

Twenty right-handed [12] native speakers of American English (8 males and 12 females; mean age = 25.9, s.d. = 4.79) participated in the study. All participants passed a bilateral pure-tone audiometric screening at 25 dB HL across octaves from 500 to 4000 Hz and provided their written consents prior to participation. None of the participants had prior exposure to tone languages. They were evaluated with a nonverbal IQ test (Test of Nonverbal Intelligence; TONI-3; [3]). Their mean standard score was 119.35 (s.d. = 11.17).

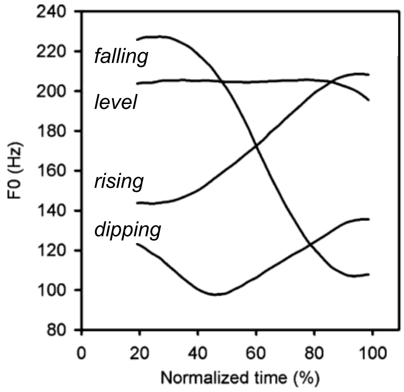
2.2. Sound-to-word learning program

The sound-to-word learning program employed in this study, and the construction of the materials, had been detailed in our previous study [4]. Participants underwent a 9-day training in which they learned to associate speech stimuli with pictures of objects presented on a computer screen.

2.3. Stimuli

Training stimuli consisted of words that were constructed based on six English monosyllabic pseudowords (‘*pesh*’, ‘*dree*’, ‘*nuck*’, ‘*vece*’, ‘*fute*’ and ‘*ner*’) each superimposed, using the Pitch Synchronous Overlap and Add (PSOLA) method implemented in the software Praat (http://www.praat.org), with four pitch patterns that resembled the Mandarin Chinese level, rising, dipping and falling tones, Figure 1. Hence 24 words were constructed and each of them was paired with a unique picture designating the meaning of the word. In total, 96 tokens were created (24 words, 2 male talkers and 2 female talkers). See our previous study [4] for a full description.

Figure 1: An exemplar pitch pattern from a male talker.



2.4. Training and testing procedures

Each session of training lasted for about 30 minutes, there was no more than a two-day gap between sessions, and no more than one training session per day. Each session was divided into a training phase and a Word Identification Test. In the training phase, words of the same base syllable were presented in the same block, so that words in a block were minimally contrasted by pitch. In each block, each sound-picture pairing was presented with an inter-trial interval of three seconds. At the end of each block, participants were tested on the words they had just learned: each sound was played and then participants had to select the correct picture from four choices. If the participant selected a wrong picture, the correct picture would be displayed briefly before the presentation of the next test item.

In the Word Identification Test, participants were presented with all the speech tokens one at a time in a pseudorandom order, and were asked to identify, untimed, each token by selecting a picture from 24 possible choices. This procedure was repeated once for each of the four talkers. No feedback was given during the test.

2.5. Image acquisition

Diffusion-weighted images were acquired on a 3-Tesla Siemens Verio with an 8-channel head coil. The diffusion weighted volumes were achieved with 64 diffusion encoding directions [1] with an isotropic voxel size of 2 mm^3. For anatomical images, they were acquired using an MPRAGE sequence[2] with an isotropic voxel size of 1 mm^3.

2.6. Image analysis

The preprocessing of the diffusion data was carried out using diffusion toolbox (FDT) in FSL [16]. The DTI data were first corrected for eddy currents and head motion. After that a diffusion tensor model was fitted at each voxel to compute, among other measures, fractional anisotropy (FA).

The FA maps created were then further processed using the track-based spatial statistics (TBSS) routine [15] where each individual FA map was aligned to the FMRIB58_FA template using the nonlinear registration tool FNIRT in FSL. These aligned FA maps were averaged to create a mean FA map and a thinning algorithm was applied to create a mean FA skeleton that represents the centers of all fiber bundles common to all participants. After that, each participant’s aligned FA map was projected onto the skeleton such that an alignment-invariant track representation of FA values was achieved for each participant.

3. RESULTS

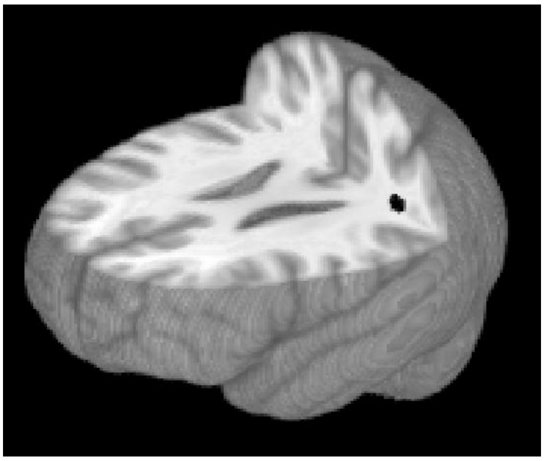
3.1. Sound-to-word learning performance

By the end of the sound-to-word learning program, on average a word ID score of 0.696 (s.d. = 0.249) was achieved. The final session word ID score correlated positively with IQ, Pearson's $r = 0.484$ ($p < 0.05$).

3.2. White matter integrity predicts learning success

Since IQ was a reliable predictor of learning success, it was included in the voxel-wise multiple regression analysis to estimate the unique contribution of white matter integrity to learning success. Specifically, both FA and IQ were used as the predictors in the regression model with final session word ID score as the criterion variable. Only one cluster, in the left parietal-temporal region (highlighted in Figure 2, MNI peak: −34, −51, 23), survived thresholding (uncorrected $p < 0.001$ and cluster size $> 20\ \text{mm}^3$) where FA reliably predicted word ID score ($\beta = 0.763$, $t = 7.328$, adjusted R-square = 0.794).

Figure 2: White matter integrity in the left parietal-temporal region (highlighted in black) reliably predicted sound-to-word learning success. The white matter skeleton was shown in grey.



4. CONCLUSION

In this study we report that white matter integrity, as measured by FA, in the left parietal-temporal region predicts sound-to-word learning success. Anisotropy of water diffusion in white matter has been argued to be due to the dense packing of intact axons that restrict water diffusion perpendicular to the axons [1]. The positive correlation between FA and generalization score therefore suggests that the denser the packing of the white matter in that region, i.e. the greater the connectivity in the corresponding neural pathway, the more successful one will be.

Our preliminary analysis using the technique of probabilistic tractography, where voxels in this region were used as the tracking seed, revealed an extreme capsule system that subserves the function of mapping sound to meaning. This ventral sound-to-meaning pathway is composed of an MdLF branch running along the STG, an ILF branch running along the MTG and the EmC, a long-distance fiber connecting the parietal-temporal region with the inferior frontal region. These findings converge with those of Saur, et al. [14] and collectively provide evidence for the importance of the ventral pathway [18], but not the dorsal pathway [2, 8], in speech perception and comprehension.

5. ACKNOWLEDGEMENTS

This research is supported by the National Science Foundation (Grant No. BCS-0719666) and the National Institutes of Health (Grant Nos. R01DC008333 and R21DC009652).

6. REFERENCES

[1] Beaulieu, C. 2002. The basis of anisotropic water diffusion in the nervous system - A technical review. *NMR in Biomedicine* 15(7-8), 435-55.

[2] Brauer, J., Anwander, A., Friederici, A.D. 2011. Neuroanatomical prerequisites for language functions in the maturing brain. *Cerebral Cortex* 21, 459-466.

[3] Brown, L., Sherbenou, R.J., Johnsen, S.K. 1997. *TONI-3: Test of Nonverbal Intelligence* (3rd ed.). Austin, Texas: Pro-Ed.

[4] Chandrasekaran, B., Sampath, P.D., Wong, P.C.M. 2010. Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America* 128(1), 456.

[5] Duffau, H. 2008. The anatomo-functional connectivity of language revisited. New insights provided by electrostimulation and tractography. *Neuropsychologia* 46(4), 927-34.

[6] Flöel, A., Vries, M.H., de Scholz, J., Breitenstein, C., Johansen-Berg, H. 2009. White matter integrity in the vicinity of Broca's area predicts grammar learning success. *NeuroImage* 47(4), 1974-81.

[7] Friederici, A.D. 2002. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6(2), 78-84.

[8] Friederici, A.D. 2009. Allocating functions to fiber tracts: facing its indirectness. *Trends in Cognitive Sciences* 13(9), 370-371.

[9] Glasser, M.F., Rilling, J.K. 2008. DTI tractography of the human brain's language pathways. *Cerebral Cortex* 18(11), 2471-2482.

[10] Hagoort, P. 2005. On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences* 9(9), 416-423.

[11] Hickok, G., Poeppel, D. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8(5), 393-402.

[12] Oldfield, R.C. 1971. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9(1), 97-113.

[13] Price, C.J. 2000. The anatomy of language: contributions from functional neuroimaging. *Journal of Anatomy* 197(3), 335-359.

[14] Saur, D., Kreher, B.W., Schnell, S., et al. 2008. Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America* 105(46), 18035-18040.

[15] Smith, S.M., Jenkinson, M., Johansen-Berg, H., et al. 2006. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage* 31(4), 1487-1505.

[16] Smith, S.M., Jenkinson, M., Woolrich, M.W., et al. 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23, S208-219.

[17] Vigneau, M., Beaucousin, V., Hervé, P.Y., et al. 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *NeuroImage* 30(4), 1414-1432.

[18] Weiller, C., Musso, M., Rijntjes, M., Saur, D. 2009. Please don't underestimate the ventral pathway in language. *Trends in Cognitive Sciences* 13(9), 369-370.

---

[1] Acquisition parameters: b-value = 1000 s/mm^2; TR = 9.8 s; TE = 96 ms; FoV = 256 mm × 256 mm, plus one reference volume without diffusion weighting (b-value = 0 s/mm^2).

[2] Magnetization-prepared rapid acquisition gradient echo (160 slices; slice thickness = 1 mm; TR = 2.3 s; TE = 3.39 ms; flip angle = 9 degree; FOV = 256 mm × 256 mm).