Modality-Independent Cortical Encoding of Phonological Primes

Spoken communication relies on transmission of a continuous and analog signal that carries the message from the speaker to the listener. Yet, the consensus is that speech processing, i.e. the perception and articulation of speech, requires analyses in terms of algebraic units referred to as distinctive features (Chomsky & Halle, 1968; Hale & Reiss, 2008). Individual sounds in words, such as the *k^h-sound* in 'cat', are decomposed into bundles of distinctive features. An obvious associated issue concerns whether these features encode articulatory/motor information or whether they represent perceptual/acoustic information? Emerging neurobiological evidence cautions against a purely acoustic (Guenther et al., 2006), articulatory (Holt & Lotto, 2008), or even a simple mirror-neuron driven hypothesis (Hickok et al., 2011). Rather, Hickok et al. (2011) explicitly argue that speech processing requires an integration of acoustic/perceptual and articulatory/motor information in the brain, primarily carried out by the *Spt,* a brain area situated in the Sylvian fissure at the parietal-temporal boundary.

Crucially, Spt activates both during passive perception of speech sounds and during sub-vocal/covert articulation. Sub-vocal activation implies that Spt is not being driven by overt auditory feedback, suggesting rather that it is involved in sensorimotor integration (Hickok et al. 2011). Further, Poeppel and Idsardi (2011) report different sub-regional patterns of activity associated with the sensory and motor phases, suggesting distinct neuronal subpopulations for each phase . More broadly, while Spt is not speech-specific, activating reliably for perception of tonal melodies and tasks involving humming, speech induced activity in Spt is highly correlated with pars opercularis in Broca's region. Activity in Spt is also reported to be motor-effector selective (Pa & Hickok, 2008), with noticeably more robust activity when motor tasks involve the vocal apparatus as opposed to manual effectors. It is conveniently located in between networks of auditory (superior temporal sulcus) and motor (pars opercularis, premotor cortex) regions, and diffusion tensor imaging studies suggest that Spt and posterior sector of Broca's region are densely connected at the anatomic level. It is, thus, both anatomically and functionally well positioned to support sensory-motor integration of the type required for computing algebraic primes from substantive information. Given these findings a crucial question for any approach to the phonology and phonetics interface is concerned with the nature of the neural wetware and processes transducing vibrations in the ears to abstractions in the brain.

Substance-free phonology (Reiss, 2017; Reiss & Volenec, 2022) appeals to the conjunctive neural representation hypothesis (Stein & Meredith, 1993) to posit that features are realized by the functional connectivity among unique sub-populations of neurons with selective responses to specific spectro-temporal receptive fields (STRF). Crucially, the identity of features is argued to be tied to the unique place, rate and frequency coding of the responding neurons, and are likely stable across participants and languages. Mesgarani et al.(2014) report that subjects implanted with multi-electrode arrays exhibit systematic selectivity of electrodes to phonetic properties specific to classes of speech sounds (e.g., stops vs. fricatives) during perception. Although this is positive evidence for the representation of acoustic-phonetic information in the brain, note that while classes of sounds discussed by Mesgarani et al. (2014) (e.g., fricatives, stops) are defined *in terms of* phonological features, any given class is (usually) defined by a set of (multiple) valued features (e.g., fricatives might be defined as [+Consonantal, +Continuant …]). Thus, a population of neurons responding to fricative sounds is, in fact, being activated by multiple different features. Localization of substantively defined classes (plosives vs. nasals) is different from localizing individual features. The latter localizes the computational primitives of the theory in the brain, while the former merely helps identify higher-order constructs of broader granularity. Here we report findings from an ongoing functional magnetic resonance imaging study designed to probe the neural substrates of individual features (e.g. [±Voice], or [Voice][1]), with a focus on identifying the regions of the auditory cortex responsible for primal phonological computations.

To this end our design systematically contrasts speech sounds in terms of single distinctive features (e.g. *sue* vs. *zoo* differ only in terms of [±Voice] in the initial position) in order to probe the hypothesis that individual features are realized as stable patterns of neural activation maps. Further, in order to test the modality independence of phonological primes we scan participants in three modes – overt production, passive perception and silent/inner speech – and employ cognitive subtraction (Friston et al. 1996) to establish the stability of primal maps across task-modes. Pilot study was conducted using a block-related design, with three runs for three task-modes. All runs were conducted on the same day, with five minutes break in between. In each mode participants were exposed to eight blocks of twenty-five stimuli tokens each, with four blocks representing voiced and voiceless words respectively. In order to avoid VOT complications we restricted our target consonants to English /s/ and /z/ in the initial position only. We also restricted stimuli to a single minimal-pair (sue vs. zoo) to ensure similar frequency effects of the lexical items. A whole-brain univariate analysis of the pilot data using two regressors (voiced and voiceless) and two contrast conditions (voiced >> voiceless, voiceless >> voiced) showed stable activation in the right-anterior prefrontal cortex (right-AntPFC) for production mode, and in the right-AntPFC and right dorso-lateral prefrontal cortex (right-dlPFC) for the silent mode. In the perception mode we noticed severe effects of repitition suppression effects, with no voxels passing statistical threshold. While activity in the right-dlPFC in the silent mode is expected given the region's role in internal simulation

[1] While significant differences of opinion exist within phonological theory, and indeed the substance-free program, regarding the binary-ness of features, the study discussed here is agnostic to this particular issue.

of anticipated speech-motor movements (Tourville et al., 2011), the lack of any significant activation in the established phonological regions of interest (RoI) – inferior frontal gyrus (IFG), superior temporal sulcus (STS), superior temporal gyrus (STG) and Brodmann Areas 44, 6 etc – and suppression effects in perception necessitated the modification of both scanning and analyses parameters. Besides repition suppression effects we hypothesized that lack of activation in the perception mode is likely to be a confound stemming from scanner noises triggering additional hemodynamic response functions besides the task-relevant stimuli. Likewise, we hypothesized that the univariate nature of our analysis is likely to mask distributed effects that do not reach statistical significance individually (Haxby et al., 2001).

In the second stage of the study we are utilizing a modified event-related design for stimuli presentation combined with a gradient echo EPI scanning parameter that significantly reduces both repetition suppression effects stemming from repeated exposure to stimuli as well as interference from scanner-noise in perception tasks by presenting stimuli in periods of silence between data acquisition pulses. Functional data collected from this modified design is then used to train a Support Vector Machine (SVM) to classify voiced and voiceless conditions in each task mode, followed by application of the trained model to functional data from predefined RoIs, which consisted of the IFG, STS, STG, BA 44, 6 and area SPT. Initial results indicate that the SVM trained on voxel intensity values obtained from the RoIs appear to be relatively robust against noise and variability in the pilot fMRI data, being able to draw an optimal decision boundary that maximally separated the conditions (including in the perception mode). Currently we are applying this same methodology to the functional data collected using the modified design, and this talk aims to provide a comparative analyses of both data-sets to illustrate that (a) primes of phonological computation consistently invoke stable functionally connected maps, with area SPT playing a crucial role across task-modes, and (b) functional localization of such neural substrates are confounded by both methodological and analytical choices. A crucial aspect of our study is its focus on the modality-independent nature of phonological primes, and our comparative analyses illustrate the advantages of adopting SVM classification algorithms to detect such modality-independent effects in cognitive tasks.

References:

Chomsky, N., & Halle, M. (1968). The sound pattern of English.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *Neuroimage*, *4*(2), 97-104.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301.

Hale, M., & Reiss, C. (2008). *The Phonological Enterprise*. OUP Oxford.

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.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407–422.

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, *17*(1), 42–46.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010.

Poeppel, D., & Idsardi, W. (2011). Recognizing words from speech: The perception-action-memory loop. *Lexical Representation: A Multidisciplinary Approach*, *17*, 171.

Reiss, C. (2017). Substance free phonology. In *The Routledge handbook of phonological theory* (pp. 425-452). Routledge.

Reiss, C., & Volenec, V. (2022). Conquer primal fear: Phonological features are innate and substance-free. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *67*(4), 581-610.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. MIT press.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. Language and Cognitive Processes, 26(7), 952-981.