Introduction We hear spoken words as a series of discrete sounds, or phonemes. Yet speech is continuous, and the articulatory maneuvers required to produce one phoneme powerfully influence those needed to produce the next – they are "coarticulated." For example, the phoneme /o/ is perceived to be identical in the words "bone" and "tone," but the acoustic structure of each /o/ is dramatically different. Coarticulation introduces strong spectral and temporal nonlinearities in speech sounds such that there is no set of acoustic properties unique to a phoneme or necessary for its perception (*1*). As a result, linguists are unsure how phonemes are recognized. Two competing neurolinguistic theories have attempted to explain phonetic processing: A "generative" theory posits that phonemes are loose acoustic templates, and phoneme identity is computed by a hierarchy of nonlinear feature-detecting neurons that combine basic acoustic features that match the template (*2*). A "discriminative" theory posits that phonemes have no positive description, and we instead only learn rules about the acoustic features that tell them apart (*1*). Phoneme identity is then computed as a path through a cortical network's possible states, with inhibitory interneurons preventing transitions to forbidden states according to those rules (as in (*3*)). These models however, like all neurocomputational models of speech perception, are unconstrained by neurobiological data. *I propose to develop phonetic discrimination in mice as a model for speech perception. I will then directly test the predictions made by competing neurolinguistic models*

Project Description **Aim 1:** Establish mice as a model for speech perception. **Aim 2:** Identify candidate cortical regions and processing mechanisms with widefield calcium imaging.

Aim 1: Mice have several powerful methodological advantages over humans as a model for phonetic research. First, their lack of semantic understanding eliminates an intractable confound in humans in which syntactic and semantic context can be used to "fill in" phonetic information. Second, their lack of speech production isolates acoustic theories of speech perception from motor theories that frame perception as "simulating" the motor pattern of the speaker. Third, a widening array of optogenetic, electrophysiological, and imaging tools available in mice overcome the lack of spatiotemporal resolution that limits neurophysiological research in humans. I have successfully demonstrated that mice are capable of learning and generalizing consonant identity across vowel contexts, speakers, and genders. As of this writing, 12 mice have learned to discriminate between consonant-vowel pairs beginning with either /b/ or /g/ (eg. "bo", "gih") in a two-alternative forced choice task. They are capable of generalizing from a training set of 20 recordings to a superset of over 200 with only a slight loss in accuracy, indicating that they have learned the phoneme itself rather than overlearned the training set. To our knowledge, this is the first time mice have been shown to be capable of learning phonetic categories, and, indeed, categories of non-species-specific natural sounds in general. I will continue to validate the model by investigating the acoustic features mice use to discriminate phonemes, and compare these features to those used by humans. We are collaborating with Dr. Kaori Idemaru, a phonetician who will aid in this process.

Aim 2: The brain region or regions that compute phoneme identity remain unidentified. Prior human research has used tools lacking the spatiotemporal resolution to provide a definitive location, and animal research has used naïve, often anaesthetized animals, limiting the ability to link candidate regions to perception. I will identify candidate regions by imaging nearly the entire temporal lobe using transgenic mice expressing the calcium indicator GCaMP6 as they are performing a head-fixed version of the task in Aim 1. I have already begun pilot imaging in naïve mice. Because the two models make different predictions about the activity in a phoneme processing region, testing model predictions and identifying candidate regions will be a joint process. The generative model predicts that a phoneme class should have an invariant pattern of *spatial* activity, because the same feature-detecting neurons should be activated by a phoneme regardless of variation in vowel context, speaker, etc. The discriminative model predicts invariant patterns of *temporal* activity within a region. Because this model only describes what a phoneme is *not*, variations of a phoneme can evoke different paths through network states as long as they don't enter forbidden states. The presence and location of either of these forms of invariance will be assessed with standard methods, by fitting neural activity with a support vector machine or hidden Markov model, respectively. I will collaborate with Dr. Yashar Ahmadian, a theoretical neuroscientist, in order to build and validate these models. Awake imaging during behavior gives two additional experimental tools: imaging across the training process allows us to investigate how phoneme perception is learned by observing changes in background and evoked activity. Comparing activity evoked during correct and incorrect trials allows us to investigate what causes errors in perception, and ensure our models are not fit to spurious data from incorrect trials.

Future Directions: After a putative neurocomputational mechanism is found with widefield imaging, I will proceed to investigate its cellular basis. These experiments would couple tetrode or whole-cell electrophysiological recordings with viral projection tracing to relate the stimulus-evoked responses of a single cell to those of its presynaptic afferents. Electrophysiology and viral tracing are routine techniques in our lab.

Intellectual Merit Constraining linguistic models of speech perception will transform deep questions in several fields. 1) The fact that mice are capable of perceiving phonetic categories already provides strong evidence against a long-held belief that the neural mechanisms of speech perception are unique to humans. 2) Languages have specific subsets of phonemes that are thought to be maximally discriminable. Better understanding of the neural mechanisms underlying that discrimination will allow linguists to assess the neural basis for deep phonetic and syntactic constraints that they impose. 3) Infants must learn to segment individual phonemes from a continuous stream of speech, but it is unclear how. Imaging as mice learn would provide the first mammalian model for infant speech acquisition. 4) Progress in understanding the processing of natural sounds has been slow, in part because it is unclear what features of the sound are being extracted. This model allows neurophysiological data to be linked to behavioral perceptual reports, and the long history of phonetic research informs experiments and allows results to be translated to humans. 5) Computation in the brain is inherently dynamic, but models of dynamic computation are in their infancy. The inherently temporal, nonlinear nature of speech will provide a strong basis for developing the dynamic models necessary to improve our basic understanding of cortical computation in general.

Broader Impact Speech is fundamental to human society. A better mechanistic understanding of speech perception will have at least two broad societal impacts. 1) Currently, the best speech recognition algorithms are based on biologically-inspired artificial recurrent neural networks, but their accuracy still doesn't approach that of a two-year old. Directly mimicking speech processing in the brain could dramatically improve their accuracy, in turn profoundly benefiting consumer technology and medical technology serving those with physical disabilities. 2) Processing phonemes not present in one's native language – for example the phonemes /r/ and /l/ for native Japanese speakers – is cognitively and socially burdensome. Some artificial speech synthesis techniques improve their discriminability (*4*), and if the mechanism of discrimination was better understood, it should be possible to make real-time translators that emphasize certain acoustic features to enhance phonetic discriminability.

1. Kluender KR, et al. *Vowel Inherent Spectral Change* (2013), pp.117-151.
2. Rauschecker JP, Scott SK. **Nat. Neurosci.** (2009) 12:718-724
3. Rutishauser U, et al. **PLoS Comput. Biol.** (2015) 11: e1004039
4. Miyawaki K, et al. **Percept. Psychophys.** (1975) 18:331-340