Prelims

Jonny Saunders
February 26, 2021

Contents

TODO:

1. Abstract

[abstract itself]

[summary of intellecual merit, broader impacts?]

- why haven't we done these experiments already? what's the role of animal models? what's the way forward??? neurophys in animals as speech models, but what *kind* of experiments are likely to help us with this question of what phonemes are.

2. Introduction

2.1 Phonemes are Language Games

> "Consider for example the proceedings that we call "games". [...] For if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. [...] Are they all 'amusing'? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. [...] Look at the parts played by skill and luck; and at the difference between skill in chess and skill in tennis.
>
> And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail. [...] And we extend our concept as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres."
>
> *-Wittgenstein, Philosophical Investigations: 66-67[1]*

Cognitive reality is characterized by its discreteness: rather than a continuous undifferentiated gradient wash of sensation and cognition, we experience objects, concepts, and categories. Speech is a continuous, high-dimensional, high-variability acoustic signal, yet it is perceived as a small number of relatively-discrete phonemes[2]. The acoustic structure of phonemes is a sort of "Family Resemblance"[1] — the truly extravagant variability of speech has thus far defied any simple, definite acoustic parameterization of its phonemes. Instead, individual utterances within a phonetic category vary along high numbers of feature-dimensions, none of which are necessary nor sufficient for a listener to identify it[3, 4].

There are different types of category structure, and what typifies family resemblance structures is 1) multiply defined - category membership is assesed across many imperfect 'features' none of which is necessary nor sufficient, 2) prototypicality - some instances are better 'examples' of a category than others, category membership is not binary, 3) context dependent - which feature is important depends on the features present in the instance and the context in which it is being compared. [5]

2.2 A Very Simple Model...

Category representation theories are intimately related (and occasionally literally isometric to [6]) to theories of the measurement of similarity, which is dominated by geometric models[7]. These models nearly universally presuppose that categories exist in a feature space such that there exist some number of features that describe each instance of an object to be categorized.

To begin perhaps purposely naively, we will formulate a very simple geometric model of perceptual categories:

Suppose that some sensory stimulus $\mathbf{s}$ was composed of some set of physical attributes a_i in the d-dimensional "stimulus space" $\mathbf{S}$ capable of fully representing all stimuli for a given sensory modality (as opposed to a particular set of eg. parameterized stimuli)

$$\mathbf{s} = \{a_0, a_i, \dots a_d \ : a \in \mathbf{S}\} \tag{1}$$

For example, a digital sound is fully defined by the amplitudes of the waveform at each of its samples, or an image is defined as the wavelength and intensity of light at each pixel. Since a_i are arbitrary, $\mathbf{S}$ can represent a set of static attributes, or a set of attributes through time.

The sensory stimulus $\mathbf{s}$ is processed into some percept $\mathbf{p}$ composed of perceptual attributes b_i in the e-dimensional "perceptual space" $\mathbf{P}$

$$\mathbf{p} = \{b_0, b_i, \dots b_e \ : b \in \mathbf{P}\} \tag{2}$$

such that some perceptual computation M maps $\mathbf{S}$ to $\mathbf{P}$.

$$M = f : \mathbf{S} \rightarrow \mathbf{P} \tag{3}$$

$$\mathbf{p} = M(\mathbf{s}) \tag{4}$$

Like $\mathbf{S}$, the form of $\mathbf{P}$ is arbitrary, so while the discussion that follows treats it like a continuously-valued metric space, it could also consist of a collection of binary/discrete properties (like traditional phonetic descriptions like [$\pm$ voiced]), as in, for example [7, 8]

The objective of the observer is to infer the category c_s given $\mathbf{s}$'s representation as $\mathbf{p}$.

$$c_s = max(\{p(c_i|\mathbf{p}) : c_i \in \mathbf{C}\}) \tag{5}$$

The form of the sensory-perceptual mapping M, the perceptual space $\mathbf{P}$ it constructs, and the inference of category identity c_s it supports serve as a loom for a few threads of the speech perception problem scattered across a few disciplines and vocabularies.

2.3 ...and its history

> 💡 - Make sure to refer back to the 3 properties of family resemblance categories and use that to structure this section!!!

A prominent strain of phonetics research in the US, largely associated with the Haskins Labs ([9] and see [10, p. 51]), has characterized the speech perception problem as resolving a set of acoustic "cues" into phonetic identity:

> "Liberman, Cooper, and Pierre Delattre began to study the acoustic speech signal, to determine how it represents the consonants and vowels of spoken words, and to discover the acoustic structure (the 'cues') essential for their identification by listeners. [...] By selectively including and eliminating elements of acoustic structure,l Liberman and his colleagues could determine what bits of structure provided information for the different phonetic properties of spoken words."
>
> -Carol Fowler & Katherine S. Harris in [10, p. 51]

The "cue discovery" paradigm of phonetics research posits that, for the auditory component of phonetic perception, the elements in $\mathbf{P}$ are linear combinations of the features in $\mathbf{S}$ whose manipulation can influence the identity of the perceived phoneme. These features represent familiar phonetic parameterizations like voice onset times or formant frequency ratios. The mapping M that constructs $\mathbf{p}$ is taken to be a fixed, innate feature of the auditory system: "this version of the auditory theory takes the perceived boundary between one phonetic category and another to correspond to a naturally-occurring discontinuity in perception of the relevant acoustic continuum." [11].

The conclusion of cue-based research is summarized neatly by Philip, Robert E. Remez, and Jennifer Pardo with respect to their sinewave synthesis experiments: "Question: Which acoustic elements are essential for the perception of speech? Answer: None[12]." The failure to find a simple parameterization of phonetic categories as acoustic cues motivated an abandonment of an acoustic account of phonetic perception entirely in favor of a motor theory of perception that positied a special, evolved "speech module" that linked the wily acoustics of speech sounds to the action of the articulatory system:

> "For if phonetic categories were acoustic patterns, and if, accordingly, phonetic perception were properly auditory, one should be able to describe quite straightforwardly the acoustic basis for the phonetic category and its associated percept. According to the motor theory, by contrast, one would expect the acoustic signal to serve only as a source of information about the gestures; hence the gestures would properly define the category" [11]

Purely motor theories of speech have been diversely problematized, not least of all by the many demonstrations that animals that conspicuously lack a human articulatory system are capable of phonetic categorization[13, 14, 15]. The acoustic problem of speech perception was simply too difficult to be solved by an evolutionarily plausible auditory system – how could the family resemblance structure of phonetic categories be learned without some explicit, innate knowledge of the acoustic consequences of articulation?[4]

Research on infant acquisition of speech sounds has since demonstrated the profound plasticity of the auditory system and its ability to learn the complex statistical dependencies between the acoustic attributes of speech[16]. A family of models based primarily on the work of Patricia Kuhl and colleagues describe the stimulus space $\mathbf{S}$ as acoustic features based on the "basic cuts" of sensitivity in the auditory system[17]. Infants exploit the statistical regularity and patterns of feature co-uccurance to learn some mapping M that constructs a "warped" perceptual space $\mathbf{P}$ that clusters features in $\mathbf{S}$ into acoustic "prototypes."[16]

Phonetic category identity then consists of some density in $\mathbf{P}$, the center of which is the "ideal" phonetic exemplar most likely to be identified with a particular categoery, and proceeding from this center point one transitions from off-target imperfect examplars to overlapping densities of other phonetic categories. Extensions to the model make this formulation explicit, like Kronrod, Coppess, and Feldman's[18] bayesian model that offers a unified explanation of the strong categorical perception of stop consonants and the weaker categorical perception of vowels. Their model describes phonetic identification as an inference problem that depends on both the acoustic properties of a stimulus and prior knowledge of phonetic categories, defined as some mean and variance in an arbitrary perceptual space.

In this model, the difficulty of the acoustic problem of speech perception carefully described by cue-centric phonetic research is resolved by suggesting the auditory system relies on sharp internal representations of category identity for phonemes that have a large degree of uninformative variaance, like stop consonants.

The degree of arbitrariness is problematic for the model, however. The proposition that there is some stimulus space $\mathbf{P}$ *that supports linearly-separable phonetic categories is emphatically counterevidenced by the 70 years of cue-based research that has attempted to find one (cite violations of gestalt principles from [19] and [3]). These prototype models, without weighting for the informativeness of a particular dimension in context (as opposed to some global weight) would be vulnerable to misidentifying speech when the most dominant cue was made redundant, when in fact human listeners will adapt to using a more informative cue. In fact a lot of the research relies on carefully parameterized speech, so if they considered the cases where those cues failed then such a single-density-based prototype model. Having nonlinear blobby parameterizations of prototypes doesn't really solve the problem either, as you would then just require an additional downstream 'readout' layer that could compute the conditions where a particular dimension*

Their future directions says that identifying and learning the dimensions is of critical importance, we can extend our model by continuing Kronrod's emphasis on the information contained in each perceptual dimension and allow it to vary by context...

2.4 An extension to our model...

Instead of a static perceptual space $\mathbf{P}$ where a given stimulus $\mathbf{s}$ is mapped to a single percept $\mathbf{p}$ (ie. M is injective), we can extend our very simple model by introducing some notion of reweighting perceptual dimensions. Rather than inferring category directly from $\mathbf{P}$ as in eq. 5, the features $b_e \in \mathbf{P}$ are reweighted by some weight vector $\mathbf{w}$ computed as some function W of the representation $\mathbf{p} = M(\mathbf{s})$ and some prior knowledge of the category structure of $\mathbf{C}$

$$\mathbf{w} = W(\mathbf{p}, \mathbf{C}) \tag{6}$$

$$c_s = max\big(\{p(c_i|\mathbf{p} \cdot \mathbf{w}) : c_i \in \mathbf{C}\}\big) \tag{7}$$

Recall that since the features $a \in \mathbf{S}$ are arbitrary, they can include time-varying features, so the weighting function W can, for example, incorporate contextual effects from the recent perceptual past. Category inference being dependent on W has equivalent interpretations in the parlance of artificial neural networks and geometry: as a self-attention mechanism (eg. [20]) giving higher weight to more informative features, or as "collapsing" or "expanding" un/informative dimensions.

2.5 And its implications...

The notion of different perceptual features having different weights or importance depending on the acoustic context and the category structure of the phonemes for a particular language is of course far from new.

cue weighting, different types of cues, contextual and informative [9] !!

A parallel line of thought to the generative models that posit phonetic identity as some positive description of cues or perceptual features are discriminative models that focuses on the features that can be used to tell phonemes apart. A prominent family of discriminative models in phonetics are those that describe a hierarchy of contrastive features[21, 22, 23]. Though they are diverse in their details, in these models M is again typically some fixed feature of the auditory system, and the perceptual space $\mathbf{P}$ that it constructs is some set of high-level descriptions like voicing, frication, or articulator configuration. Typically these features are binary (eg. +/- voiced), rather than continuous.

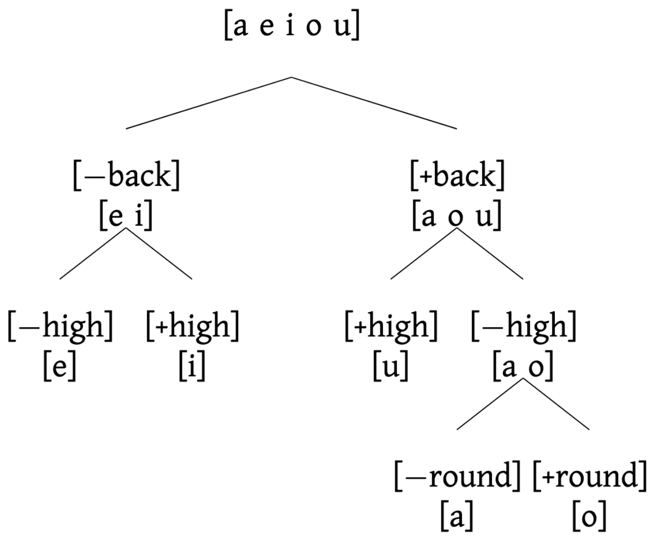

Figure 1: Contrastive hierarchy for Russian Vowels, reproduced from [24] without permission

As an example, consider the proposed contrastive feature hierarchy for russian vowels from [24] (Figure 1). Vowel identification is dominated by the primary contrast of +/- back, and successive constrastive features eliminate candidate phonemes until the true phoneme is identified. W's dependence on $\mathbf{C}$ is exemplified (fix passive voice..) by its treatment of "round": -back vowels [e i] are fully determined by +/- high, so for a percept $\mathbf{p}$ with -back, the weight of "round" should be 0. Put another way, the importance of a given feature is dependent on the phonemes that are left ambiguous without it. Any given feature's importance depends on both the set of available features and the set of available categories (the dependence of W, and thus roughly the "meaning" of $\mathbf{P}$, on $\mathbf{C}$ can also be thought of as the "task demand" on phonetic perception, see for example the discussion of [8] in [25]).

features like +rhotic though don't correspond to anything in the input space tho[26]

expand here on how parameterized stimuli with a single contrast aren't really modeling the problem: like shouldn't it matter how bad the speech sounds sound for claims about natural speech perception? the real question is, during perception, how are the different perceptual axes normalized/selected/weighted; during learning, how does the auditory system learn the space of features? When there is only one feature present the auditory system is performing a qualitatively different task. The use of parameterized stimuli is itself a strong assumption on the nature of the problem that the auditory system is solving. Even parameterizing a family resemblance is so because you assume the weight and salience of different cues. Additionally since there is some "basic cuts" argument to be made about the auditory system and the types of cues that it selects, you're unlikely to hit those if you just use some arbitrary array of stimuli: speech sounds come pre-optimized for mammalian auditory systems (though obvs mice aren't people) a la adaptive dispersion

"I should emphasize, nevertheless, that there is a great deal of evidence that practice, even large amounts of it, does not produce efficient perception of acoustic alphabets. This is clear, not only in the example of the Morse code, but even more convincingly, perhaps, in the repeatedly unsuccessful attempts to find nonspeech sounds that will work well as part of a reading machine for the bling. Many sound alphabets have been given a thorough trial, but none has proved adequate. It must surely give us pause to know that, while sounds are the universal carriers of language, only one set of sounds — those of speech — serves well."[27] Their conclusions are wrong – that this means that speech is special and has its own processing modality – but the observation does indeed point to the joint optimization of a phonetic space over an auditory space as being constitutive of language, and a potent reason to use speech sounds for category learning.

The notion of the informativeness of different featural dimensions has been given its *fullest* treatment in Keith Kluender and Christian Stilp's application of information theory to phonetic perception[28, 29, 30, 31]. They summarize their argument, elegantly as always

> "If one's problem is finding the right fencing to corral a unicorn, then there is really no problem at all. Instead the problem is dissolved upon discovery that unicorns do not exist. !!
>
> Here, we ask the reader to consider the possibility that there are no objects of perception [...]. Like unicorns, they do not exist at all. Instead, there are *objectives* for perception. [...] Perceptual success does not require recovery or representations of the world per se." [28]

They argue that the central operation of sensory systems is to adapt to regularity at multiple scales in order to efficiently extract meaningful information from their environment. !! Rather than a faithful representation of articulatory maneuvers (as in motor theory) or a warped, but still bijective relationship between the acoustic space and perceptual space (as in perceptual warping), they argue that sensory systems discard information that is predictable based on (multiscale) context, and instead represent just the unpredictable, "information-bearing" in an appropriately Shannonistic sense, dimensions.

Though theoretically all configurations of frequencies and amplitudes are possible, naturally produced sounds are strongly constrained by the physics of their production – much of the variation in natural sounds is predictable. Rather than representing the fullness of acoustic variation, the auditory system adapts to redundancies and regularities in sounds to preferentially represent only the unpredictable, informative variation in an "efficient code" [32, 33]. In the case of phonetic perception, where the objective of the listener is to identify the phoneme intended by the speaker rather than perceiving a sound qua sound, the listener attempts to learn auditory features that are maximally informative of phonetic identity[34, 35, 28, 29].

This information-theoretic account provides a mechanism for learning the dimensions of **P** and the form of W. Rather than some a priori, fixed inventory of articulatory/acoustic cues, a listener should learn some set of perceptual features that support the identification of phonemes given the phonemic inventory of their language and the acoustic variability (eg. accent, environment, timbre, etc.) that they are exposed to. Individual listeners do indeed use different combinations of cues with different weights[36] which are stable over time[37]. Rather than learning some category center and spread over some preexisting perceptual feature space, the task of the listener is to learn the feature space itself.

The difference between learning **P** and the operation of W is a matter of timescale: over short timescales, W reweights the features in **P** depending on those features that are contextually informative of phonetic identity. While the observation that individual cues are informative, uninformative, and anti-informative depending on the context of surrounding phonemes is a central feature of argument for a motor theory[4], an information-theoretic view interprets this problem as a reweighting of individual features: /s/ differs from /f/ along different featural axes than /s/ differs from /k/, so /s/ shouldn't necessarily rely on the same inventory of acoustic features in all contexts — and particularly when cues are rendered uninformative, the auditory system should adapt to emphasize those that still are(eg. as in [34] and [38]). Contextual effects on phonetic categorization are of course well known (see [2]). Where perceptual warping accounts cannot explain results where some or all of the typical acoustic features are replaced, like sine-wave speech[39], noise-vocoded speech[40], or joint spectrotemporal degradation[41]; an information-theoretic view argues that listeners will adapt to use any cues that are still present (as in [34]).

The auditory system does *not* seem to operate in an entirely information-maximizing way when identifying phonemes, however. Consider a category structure like that used by Couchman, Coutinho, and Smith (2010, [42]) depicted in figure 2. Each stimulus is composed of four binary features (columns), and stimulus identity is defined by the first feature (0 = category A, 1 = category B). The remaining three features are "epiphenomenal," but stimuli in category B have a greater sum than those in category A. A perfect, information-maximizing observer would learn to only attend to the first dimension, but in speech and many other perceptual categories observers use many, even uninformative dimensions[42, 5] (but see [43]). Non-speech sounds that are strictly uninformative of phonetic identity like pure tones and sweeps can nevertheless strongly influence the perceived phoneme[44, 45], even when the sounds are not immediately adjacent[46]. Such an influence of many, imperfect stimulus dimensions on perception is our signpost to indicate we've arrived back in the bewildering little shire of category structures with family resemblance.

The differing (often implicit) assumptions about the 🙃 ~very complicated model™ ~🙃 characterize the major historical disputes in categorical phonetic perception, but also *<suggest the kinds of experiments that might resolve them>.*

!! expand on each of these: a) Arguably, the careful work of cue theorists led them to motor theories of perception because of a characterization of M as fixed that made the non-invariant acoustic structure of phonetic categories impossible for the auditory system to compute. *but their work was extremely valuable because it explicated the nonlinear nature of acoustic cues and the family resemblance structure of acoustic properties.* !! b) Work in animal models and infant speech perception demonstrated that phonetic categories were indeed learned !! (cite infants can acquire all phonemes), but the use of parametric stimuli led to overly-parsimonious models that don't capture the true scope of the problem. !! need experiments that satisfy the "real problem" (review previous sections and highlight each of the ways the family resemblance structure of phonemes indicates a particular experimental design parameter, but that we need to finish it by adding a neural layer... which we get to in the next section...)

*Integrate this into the discussion about infant speech learning research in previous para - The idea that speech acquisition necessarily involves learning the features that are maximally informative is demonstrated by the ability for infants to discriminate between the phonemes of any language, but during language acquisition become specifically attuned to the phonemes of the language(s) they are taught. Though this is typically discussed as learning the statistical regularities of speech sounds (**need to cite more because claim of typicality**[47][17]), the act of emphasizing the statistical regularity must necessarily mean collapsing those phoentic contrasts that are not present in the language – they aren't informative because no one uses that contrast. indeed they trade off – infants that are better at discriminating the phonemes in their language are worse at discrmiinating those in a non-native language[47] (babies initially can learn all phonemes[17], so they have to learn some feature which necessarily compresses the auditory space[48])*

*and focusing on the acquisition of informative stimulus dimensions fundamentally alters the research question. The problem is the mutual translation/misundertanding of what cues *are* – a lot of neurophys research into language ends up using parameterized speech because we want to create parameters and then* *look for analogies in the brain, either in single neurons or populations. Neuroscientists interpret these cues as 'constitutive' of the phoneme rather than a particular cue describing it (try to find ye old phonetics lit that talks about cue validity as being a problem even in phonetics). This is the pt to turn to 'so instead we need to let the brain reveal its order to us, when presented with a complex array of stimuli, which features does the brain encode and how are they represented???'*

Category A	Category B
0 0 0 0	1 1 1 1
0 1 0 0	1 0 1 1
0 0 1 0	1 1 0 1
0 0 0 1	1 1 1 0

Figure 2: Category structure reproduced from [42] without permission. Each stimulus (row of four digits) is composed of four features (columns). Category identity is determined by the first feature (0 = A, 1 = B), but three other "irrelevant" features are present.

2.6 Neural mechs

Until now our very simple model has been entirely theoretical, describing the general requirements of the computation of phonetic category identity, but the form of any biological computation is necessarily constrained by the substrate of its implementation (roughly, Marr's levels, for a recent discussion see [49]). Though the model could be retained in its current form by recasting **P** as the neural representation of perceptual dimensions from which category $c \in \mathbf{C}$ is inferred, this would require strong assumptions about the form of the neural representation of perceptual dimensions, and in a practical modeling context assumes we have enough information to infer it. To preserve generality at the cost of complexity, we add an additional "layer" to the model,

$$\mathbf{n} = \{n_0, n_i, ...n_{dn} : n \in N^{dm} \subseteq \mathbb{R}^{dn}\} \tag{8}$$

where a neural state **n**, a dn-dimensional instantaneous firing rate of neurons n_i in some neural manifold $\mathbf{N}^{dm}$ of dimension dm embedded within $\mathbb{R}^{dn}$. The manifold embedding N reflects the intrinsic constrants network structure poses on the possible states $\mathbf{n} \in \mathbf{N} \subseteq \mathbb{R}$, but the embedding is arbitrary.

The neural layer is incorporated by modifying equation 6 such that

$$M_n = f(\mathbf{s}, \mathbf{p}) : \mathbf{S} \rightarrow \mathbf{N} \tag{9}$$
$$M_p = f(\mathbf{n}) : \mathbf{N} \rightarrow \mathbf{P} \tag{10}$$

where some sensory input **s** is mapped to some neural state **n**, which supports some percept **p** from which phonetic category is computed. The dependence of M_n on **p** reflects the possibility of top-down influence on the neural representation of a given stimulus. talk about the representation of time in the model. !!

note that what we're doing here is largely accounting for incomplete observation and agnosticism of the implementation of perceptual representation. For example there might be some real perceptual dimension that is not independently represented in the neural space, but is computed "downstream" by some structure that we're not observing. In the case of making a claim on the structure of neural representation (talk about alternatives briefly, that not everything necessarily is represented by the firing rate) and full observation, $\mathbf{N} = \mathbf{P}$ – where $\mathbf{P}$ is then the perceptual space represented by the brain from which category identity is computed. So talk about when we separate vs. when we treat them as the same in following section

more on levels of analysis here? The inextricability of talking about implementation and theory is precisely reflected in the obligation of understanding the ways that the particular system results in the idiosyncracies of the observable behavior – or the degree to which an explanation of the implementation explains and recapitulates the idiosyncracies of the observable behavior is the degree to which it is more or less "correct", in a strict modeling sense. So, precisely for the same reason that we care that our theoretical model accurately describes observable behavior, it is impossible to separate a theoretical model from its implementation – though the temporary illusion is invaluable.

Arguably a computational strategy common to all sensory systems is to exploit regularities in the statistical structure of the natural world to form an efficient sensory representation[50, 51, 32, 31, 52, 53](cite more here bc broad claim). Though !! the task of phonetic perception is a truly monstrous one (*expand more here?*), work since the heyday of motor theories has demonstrated the remarkable ability of the auditory system to perform the fundamental computations of phonetic categorization has given the problem an air of tractability. And though we still are methodologically limited in our ability to study spech perception in humans at the spatiotemporal scales of its computation, work in animal models as well as recent advances in human brain electrophysiology have given some of the first glimpses.

Several features of our model are happily known to be true of neurons in mammalian auditory cortex.

Neurons in primary auditory cortex jointly encode multiple dimensions of sound[51]. In ferrets presented with an array of stimuli that varied by pitch, timbre, and azimuth[54], more A1 neurons were observed to be sensitive to two or three dimensions (36% and 29%, respectively) than a single dimension (23%). In a subset of neurons, these responses were temporally complex such that the dimensions could be partially recovered by separating sustained from onset responses[55]. Similar results have been observed in marmosets (combined sensitivity to amplitude modulation, frequency modulation, etc. [56]) and in studies that estimated the dimensionality of receptive fields from complex stimuli like dynamic ripples in cats[57]. This is perhaps unsurprising, as cortical neurons being sensitive to multiple dimensions of a stimulus is a trivial reformulation of the well-known hierarchical processing throughout the auditory system (for a review, see [58]): cortical neurons representing "higher order" properties of a stimulus necessarily implies sensitivity to multiple features of the stimulus (provided a generously-enough low-level description of the stimulus feature space).

Maciello and colleagues recently argued that joint, rather than independent encoding of multiple stimulus dimensions is computationally advantageous[59]. Though sensitivity to multiple features makes response patterns ambiguous with respect to the value of any individual dimension, joint encoding provides more information about all represented dimensions to a downstream decoder. If it is the case that joint encoding is constitutive of auditory representations, and individual stimulus or perceptual dimensions are never (or rarely) represented independently, behavior that reflects sensitivity to family resemblance structure rather than optimal rule-based categorization is parsimonious. If all features are estimated simultaneously, influence of "nontarget" dimensions becomes unsurprising.

Auditory cortical neurons adapt to predictable acoustic statistics in order to represent more informative stimulus dimensions at both short and long timescales.

A rich body of research has described the many conditions that auditory representations are modulated by context (for a review, see [60]) at timescales as short as hundreds of milliseconds[61, 62]. Processes like forward masking (cite), stimulus-specific adaptatation (SSA, cite), and suppression of background noise all reflect the general principle that auditory representations adapt to predictable acoustic statistics (cites here) in order to form robust, invariant representations of auditory objects[63] by emphasizing the maximally informative dimensions[57].

Adaptation to noise or stimulus statistics can be characterized as a short-term 'reweighting' of features through processes like synaptic depression[64, 65] or microcircuit interactions[66, 67]. In tasks based on simple parametric sounds, representations of task-relevant stimuli are enhanced on the order of minutes[68]. Animals trained on multiple tasks had neurons that adapted their receptive fields to facilitate the different task demands[69] and reward structures[70]. David and Shamma (2013[71]) argue that short-term integration of auditory context could also be a substrate for representing and comparing auditory features that occur through time.

The auditory system is also plastic on longer timescales to represent the dimensions of sound that are maximally informative to the demands placed on it. Rats trained using a single set of stimuli had differential enhancement of sensitivity to frequency or intensity depending on which they were trained to attend to[72]. Bieszczad and Weinberger observed that such enhancement correlated with the strength of a learned memory trace[73].

speech-specific stuff

The Superior Temporal Gyrus (STG) in humans, or secondary parabelt regions in some other species, of auditory cortex is the primary candidate for representation of higher-order auditory features used in speech perception. Damage to the left posterior Superior Temporal Gyrus, containing BA 22 "Wernicke's area," has long been associated with receptive aphasia, but a variety of human and animal studies have given further insight on the character of speech processing within the STG.

A series of studies from Edward Chang and colleagues recording electrophysiological activity in human temporal lobe using high-density multi-electrode arrays have contributed greatly to our understanding of the encoding of speech sounds, particularly in the superior temporal gyrus (STG) (<- redundancy supreme here)[74]. !!

Recordings of high-gamma (70-150Hz) power show individual electrode sites in middle to posterior STG are selective to acoustically similar groups of phonemes (eg. obstruent vs. sonorant selectivity, plosive vs. fricative selectivity, etc.) in humans passively listening to natural speech samples[75]. These phonetic sensitivites were reflective of sensitivity to multiple complex acoustic features that are correlated within phonetic categories and that "maximiz[e] vowel discriminability in the neural domain."[75]. Lower frequency (<50Hz) macrocortigraphy recordings also show that subpopulations of pSTG neurons carry information that allows discrimination of consonant-vowel token category analogously to behavioral categorization[76]. !!

In the anterior STG (aSTG), individual sorted units recorded from one person demonstrated complex, speech-specific reponses when one subject was presented with a wide array of sounds[77]. Many (66 of 141) units demonstrated selectivity to one or a few words that was invariant across speaker. Speech selectivity was only partially explained by a linear combination of acoustic features (linear spectrograms and MFCCs), and did not (over-)generalize to noise-vocoded speech, time-reversed speech. Unit responses to individual phonemes also differed by the recent phonetic past, all together suggesting that some units in aSTG are selective to the fine spectrotemporal structure of speech sounds at single-to-few phoneme timescales [77].

Though acoustic response profiles are spatially heterogeneous across the STG and between individuals[75, 78], there does appear to be some functional distinction between anterior and posterior STG with respect to speech sound processing. In macroelectrode recordings in humans listening to natural sentences, pSTG electrodes selectively track phrase-level onsets, while aSTG electrodes have more sustained responses through a phrase. The dissociation between onset and sustained responses was not reflective of the discontinuous vs. continuous nature of consonants and vowels, as selectivity to groups of phonemes (vowels, plosives, nasals, etc.) was mixed in both anterior and posterior STG [78]. Information useful for discrimination of phonetic identity in the pSTG develops and reaches a peak 100-150ms or so after speech sound onset[75, 76], and neural state space projections onto axes representing the activity of neurons sensitive to sound onset or sustained sound show a reliable sweep between posterior and anterior STG on the order of seconds. summarize description of temporal processing distributed across multiple regions that potentially reflects different parts of the information being reflected in different... codes.

Animal research of neural mechanisms of speech sound processing is quite sparse, and so our understanding is relatively coarse and by analogy from more general auditory research. Speech training in rats evokes a complex set of changes to acoustic response properties in several auditory cortical fields loosely analogous to secondary cortical areas in humans[79]. Neurons in the anterior auditory field (AAF) and A1 were more responsive to the initial consonant in consonant-vowel (/CV/) pairs in trained vs. control rats (27% and 57% more spiking activity, respectively). Additionally, the proportion of neurons that were responsive to 2kHz tones (the spectral peak in the speech tokens used) increased by 65% in AAF and 38% in A1 after speech training compared to control rats. In contrast, in response to vowels VAF and PAF were less responsive following speech training (42% and 30% fewer spikes, respectively, vs. controls). In neurons that had similar frequency tuning, responses to consonants were more correlated in AAF and VAF, and responses to vowels were less correlated in AAF, A1, and VAF after speech training (vs. controls)[79].

These results[79] may not establish definitive roles for secondary auditory fields in rodent auditory cortex, but in sum do suggest that speech training induces long-lasting plasticity in auditory cortex, and suggests that processing may be distinct for different acoustic features in anterior vs. posterior fields as in humans. Mice trained to discriminate speech sounds were returned to chance following lesions of auditory cortex[80], indicating its necessity. Task-specific plasticity[81] and contribution to processing task-relevant auditory stimulus categories[82] has been previously demonstrated in AAF, which is thought to operate as a parallel processing system, with response latencies comparable to or lower than A1 in cats[83] and mice[84]. PAF is a secondary auditory cortical area and thought to be downstream from both A1 and AAF[85, 86]. Though their functional specialization of computational role might not be equivalent in humans, it is parsimonious to assume that primary and secondary auditory cortical areas in nonhuman mammalian auditory systems process acoustic information in such a way that supports the recognition of phonetic identity.

Talk about the categorical decisionmaking process downstream, implications for role of nonauditory, frontal, etc. zones that actually do the integration with syntactical, semantic information. Differentiate that we're concerned about the derivation of the acoustic level, the perceptual dimensions that facilitate, but may not constitute the identity of a phoneme. !!

2.7 uh is there a name for the conclusion of an introduction because i need to make a section break to write it lmao !!

In lightly constraining the constitution of **N**, loosely the neural "representation" of phonetic information, the human and animal results hint at the dissociation between **N** and **P** in our model — en passant to *the statement of the research problem.*

Suppose that one dimension $b_{vot} \in \mathbf{P}$ is the voice onset time, which dissociates voiced from unvoiced consonants (eg. /b/ vs. /p/) as the time between the onset of phonation and the occlusion of the stop. Further suppose a neural system analogous to the temporal landmark model suggested by [78] where the high-energy plosive of the occlusion is "encoded" by the activity of some region analogous to the phrase-onset sensitivity of pSTG, and the sonorant, spectral quality of the voicing is encoded by another region. In this scheme, some downstream region (really need to give a name to the "readout" part of the model) infers VOT by comparing the relative timing of gross spiking activity between these two regions. In this hopelessly naïve instantiation of our model, the dimension b_{vot} is some real-valued (though not necessarily linear) value from negative to positive voice onset times. Such a dimension is not present in **N** as characterized by the n-dimensional space of, say, instantaneous firing rate of n neurons, requiring M_p, the map- !!

ping between them.

The dissociation of the descriptions of **N** and **P** thus, in our model, defines the research problem:

1) We characterize the problem the brain faces in auditory phonetic perception is to learn some perceptual space **P** of maximally informative perceptual dimensions that supports the identification of received phonemes by flexibly adapting to the information present in the phoneme.

2) Understanding the neural mechanisms of auditory phonetic perception is describing the way **P** is implemented by some neural state manifold **N** in such a way that the difference between **N** and **P** is minimized.

It is not necessarily the case that we should expect to find neurons, or even collections of neurons, whose time-averaged firing rate is the literal measurement of the perceptual dimensions used to compute phonetic identity. We also don't expect to be able to estimate the full manifold of all neurons that are involved with the process, so there will ultimately always be some gap between **P** and **N**. Roughly, kept independent, **P** is the level of "representation" — the basis from which the brain derives its use of phonetic information (though we dont' characterize **P** as the unique source of information, as information is represented at multiple scales (syntactic, semantic) and is bidirectional (predictive as well as receptive)) — while **N** is the level of "implementation". Finding a "neural representation" of phonemes is thus describing the implementation of **P** by **N**, ergo constructing them in a such a way that their difference is minimized.

This distinction may read as trivial, but it precludes a majority of the common methodological kinks of contemporary cognitive neuroscience. The implicit assumption of "decoding"-based analysis strategies is that neural representation is encoded in the language of time-averaged firing rate, and that the accuracy of some (usually uninterrogated) classification algorithm on the timeseries of firing rates (or BOLD level, or EEG bandpass amplitude, etc.) is reflective of the presence or absence of category information in the data. The same assumption is made in the case of so-called "Representational Similarity Analysis," and any number of other analytical ruts that uncritically characterize the geometry of the brain and the perceptual reality it supports as euclidean spaces with the axes of whatever recording methodology is handy for the dataset.

In both, the geometry of the perceptual space is also typically uninterrogated, where the parameters that were used to synthesize the stimuli, or the category labels imposed by the researcher are analyzed as if they were faithfully represented by the brain. This, despite the creation of non-isomorphic representations of physical phenomena being the entire goal of efficient perception (see [30, 87]) — if representation operated like an isomorphism then perceptual learning would be entirely unnecessary.

Rather than assuming the perceptual structure of phonemes by prespecifying cues and synthesizing sounds, or assuming the representational language of the brain to be time-averaged firing rate, we take the role of empirical geometers and attempt to preserve as much of the natural complexity of the problem and derive both from the data.

3. Specific Aims

3.1 Scraps

- Segmenting strategies [88]
- Scrambled vs. unscrambled sounds? (cites 12, 18, and 25 in [89])
- inferring perception-action loops from data [90]
- complementary roles of cell types and manifold dynamics [91]
- LFADS for sequential autoencoders [92]
- modeling auditory waveform with kernels [32]
- brain is actually a dynamic system and need to model the manifold [93] becasue the same brain region does multiple things at the same time with the manifold lol [94]
- ?time constant of auditory sensitivity in STG neurons?
- The natural analog of the philosophical problem of universals in the conditioning paradigm is stimulus generalization [95]
- Neural nets for estimating nonlinear STRFs, se [51]
- extracting maximally informative features [35]
- creating superstimuli [96]
- estimating nonlinear STRF[97]
- remember to return to shepard and 2nd order isomor-

phism stuff

The history of this question includes Shepard and Tversky's multidimensional scaling and its criticisms, and also extends through Shepherds' "second-order isomorphisms" (cite representation is representation of similarity)

arguably the cue-theorists arrived at the wrong conclusions was because of their belief about the innateness of the auditory-perceptual mapping: it must have been genetic, so therefore language is parsimoniously some special module, etc. etc. Research based on synthesized parameters based on cues then carry that error further by not representing the full scope of the problem. like how they eventually discarded the notion of cues (definitely need more detail in that story about specific examples of how cues are conflicting in different contexts) was because they considered their interaction with other cue dimensions. If we instead take the info-theoretic perspective seriously then learning a phoneme should be the act of learning the maximally informative dimensions. since we see individual differences in cue weighting within individuals, we would also expect people's dimensions to be different... but if there is only one or a few carefully parameterized dimensions of variation present in the stimulus set, of course they'll learn those, so we need to instead use a stimulus set that preserves as much of the natural variation within category as possible and allow the animals to learn the contrastive dimensions themselves. using only two categories is of course a simplification, but it still mimics at least the nature of the learning problem in qualitative form, and also [evidence that infants learn stop consonant boundaries early and they are primary and near-universal across languages indicating that they are sorta self-stable system where the big featural distinction of being stops makes it so they are like a 'submodule' within a phonetic set.]

parameterized vs natural speech is actually reflective of a much larger positivist/naturalist philosophical divide – they presuppose by testing a parameter of category membership, but postiive evidence is not evidence that parameter is actually constitutive of the category itself – for example if you had two categories "games" and "cars," "weight" might be a reasonably good way to assign category membership, but it is not at all the only, or even the most salient difference between those categories. Like i feel like I'm crazy sometimes because shouldn't the fact that synthesized speech sounds sound bad be a problem? They might have all the theoretical justification in the world but the fact that they so badly imitate what even a plausible phoneme would sound like should be like a red flag for the generalizability of the conclusions that can be drawn from them.

theoretical problems with simplified stimuli - low-dimensional and linearly-separable stimulus spaces are fundamentally different than the high complexity of naturalistic stimuli... for all we know the computations are just straight up not comparable! [98]

3.2 behavior

If the objective of the listener is to understand, ie. to be able to parse the speech sounds made by their interlocuter, then how is that different than that of the mouse, which is to get water? They are identical when water is only given when knowledge is demonstrated, but that is impossible when the chance of false positive is 50%. more importantly how that intersects with passive learning/non-rewarded phoeme studies.

reasons for speech stimuli: category complexity depends on the density of the space. the competition for desire for rich vocabulary of phonemes with limited articulatory palette means that we need to fit a shitload of acoustic complexity into an extremely small temporal window with a small amount of potential variation. So yeah parameterized mouse calls might work but that's like a feature of the density of the communication space, but they also have extremely subtle cues in their environment that they need to parse... so speech sounds are good because they're not species-specific but also because they're stimuli that we know have a potential subjective categorization structure but one that is sufficiently complex. speech sounds also take advantage of the innate contours of the auditory system,

trying a fresh rewrite: q: why use natural speech rather than some other synthesized, complex, high-dimensional acoustic stimulus? a: though the question is about auditory category learning in general, the auditory system is not some lockean tabula rasa because natural law dictates that auditory reality isn't some equiprobable playground where all sounds are possible. the auditory system evolved to be better able to learn certain acoustic contrasts compared to others because the fact that some contrasts are more informative than others is written into the very sinew of natural law (cite patricia kuhl's 'basic cuts' argument, tony zador 'critique of pure learning'). it is also not sufficient to identify one or a few of these 'natural auditory-perceptual gradients' and synthesize stimuli along them: the problem that languages have been solving for <many> years is how to pack many contrasts that are all mutually intelligible at rapid timescales (low … resolution?) across those gradients. Close phonetic contrasts are thus complex stimuli optimized to be discriminable by the mammalian auditory system in a dense category-space, making the reliance on the family resemblance-type structure (rather than a simple rule-based solution) that typifies phonetic identification and other complex category processing necessary

The requirement for doing it online is because what you're doing is doing a much more efficient exploration of the massive massive stimulus space– theoretically if you freaking play a billion phonemes of infinite variation you will just be grid searching all the same space that you would by presenting it online. Sooooo if we can't make online stimulus modulation work, then we just need to make sure we have sufficient samples to tile the space. Importantly though, since we're not necesarily trying to explain *speech as such, but rather then learning of some general auditory categories, the degree to which our stimuli (and thus our estimates of perceptual dimension) only really affects the degree to which we simulate the problem of speech. What could be degraded? well, it could be the case that we use too few stimuli to have a sufficiently complex categorization in the first place, but that's pretty unlikely because of the extreme variability of speech across vowel contexts, let alone speakers. Fitting after training, or like even online fitting, or even just like testing their responses to generated stimuli afterwards is totally valid as a test of the validity of the dimensions.*

3.3 imaging

3.4 analysis & modeling

Neuroscientists sorta blithely assume what the features of a stimulus are, from the seemingly harmless and physically based – frequency, direction, angle, etc. – to the absurd – rsa et al. But these dimensions rarely behave like 'real' perceptual dimensions [99] – the transformation is actually the critical part.

assuming feature dimensions is always a bad assumption – eg what features have the metric structure that measure similarity/dissimilarity of rectangles? [99]

Lots of people already talking about this, but even criticisms sorta treat perceptual dimensions as a given, and it is the brain's fault that it doesn't represent them. [100]

4. Significance & Broader Impacts

5. Notes

5.1 Bailey & Summerfield - 1980

A perceptual system in which the information for phonetic perception was a set of cues would have to incorporate three kinds of knowledge if it were to function successfully. It would have to know, first, which aspects of the acoustic signal are cues and which are not; second, it would need to possess a sensitivity to the pattern of cooccurrence of cues for each phone in its perceptual repertoire; third, it would need to appreciate the proper temporal coordination of the cues within each pattern. There is no reason, in principle, why a device could not be built to perceive phonetic identity from a substrate of acoustic cues, provided it was endowed with an articulatory representation sufficient to embody these three kinds of knowledge. However, we doubt that such a system could evolve in the natural world. For a species to acquire a knowledge of articulatory constraints, it would be necessary first that information specifying those constraints be available for the species, and second that the species possess a prior sensitivity to that information. The knowledge that a particular set of cues combine to indicate the presence of a given phone could be acquired in either of two ways. The identity of the phone could be specified independently of the set of acoustic cues, but this would hardly solve the problem and would preempt the need to evolve a sensitivity to the cues. Alternatively, the signal could specify directly both the identity of the cues and their temporal coordination, but then information in the signal that specified the coherence of its elements would, isomorphically, specify the articulatory event from which that coherence derived. However, the presence of this information about articulation in the signal, and a predisposition to register it on the part of the perceiver, would obviate the need for any internalized articulatory referent to mediate the acoustic-phonetic translation.

These considerations lead us to question the validity of equating the operational and functional definitions of an acoustic cue. A cue was defined operationally as a physical parameter of a speech signal whose manipulation systematically changes the phonetic interpretation of the signal. Although it is clear that perceptual sensitivity must exist to the consequences of manipulating a cue, it is not necessary to suppose that the cue is registered in perception as a discrete functional element.[4]

6. meta

6.1 to-read

- revisit the tversky lit and check Danielle's cites for more
- the long-term imaging/ephys papes
- [59]
- [101]
- [102]
- [90]
- [103]
- [94]
- [104]
- [105]
- [106]
- [107]
- [108]
- [109]
- [110]
- [111]
- [112]
- [113] - methods
- [114] - methods
- [88] - methods

6.2 bookmarks

- [91] - p6

6.3 scraps - neuro

- auditory processing as domain-general and domain-specific across multiple timescales [89]
- abrupt transitions, at least in neural data [101]
- multimodal representations and preserved neural manifold dynamics across inference tasks in M1 [94]

- timescales of processing expand across auditory hierarchy (and more generally have different timescales of integration and lags) [89] and are lateralized [115]
- contributions from basal ganglia in reward learning for acoustic dimentions [116]
- this bif ol review [117]
- neurons that process auditory information at phonetic timescales are relatively insensitive to spectral quality [89]
- find where this go -> Indeed different people have different cue weightings that are more or less adaptive[118]
- emergence of invariant reps in secondary auditory cortex[119]
- vocalization sensitive neurons in anterior left acx with different projection patterns from/to L6 that are experience dependent. (cfos[115])

6.4 scraps -

7. References

References

[1] Ludwig Wittgenstein. *Philosophical Investigations*. Oxford: Basil Blackwell, 1968. 250 pp. ISBN: 978-0-631-11900-5 (cit. on p. 3).

[2] Lori L. Holt and Andrew J. Lotto. "Speech Perception as Categorization". In: *Attention, Perception, & Psychophysics* 72.5 (July 1, 2010), pp. 1218–1227. ISSN: 1943-393X. DOI: 10.3758/APP.72.5.1218. URL: https://doi.org/10.3758/APP.72.5.1218 (visited on 01/15/2021) (cit. on pp. 3, 7).

[3] Leigh Lisker. "Rapid versus Rabid: A Catalogue of Acoustic Features That May Cue the Distinction". In: *The Journal of the Acoustical Society of America* 62.S1 (1977), S77. ISSN: 00014966. DOI: 10.1121/1.2016377 (cit. on pp. 3, 5).

[4] P J Bailey and Q Summerfield. "Information in Speech: Observations on the Perception of [s]-Stop Clusters". In: *Journal of experimental psychology. Human perception and performance* 6.3 (Aug. 1980), pp. 536–563. ISSN: 0096-1523. DOI: 10.1037/0096-1523.6.3.536. pmid: 6447767 (cit. on pp. 3, 4, 7, 14).

[5] Eleanor Rosch and Carolyn B Mervis. "Family Resemblances: Studies in the Internal Structure of Categories". In: *Cognitive Psychology* 7.4 (Oct. 1, 1975), pp. 573–605. ISSN: 0010-0285. DOI: 10.1016/0010-0285(75)90024-9. URL: http://www.sciencedirect.com/science/article/pii/0010028575900249 (visited on 01/12/2021) (cit. on pp. 3, 7).

[6] S Edelman. "Representation Is Representation of Similarities." In: *The Behavioral and brain sciences* 21.4 (Aug. 1998), 449–67, discussion 467–98. ISSN: 0140-525X. pmid: 10097019 (cit. on p. 3).

[7] Amos Tversky. "Features of Similarity". In: 84.4 (1977) (cit. on pp. 3, 4).

[8] Michael D. Lee and Daniel J. Navarro. "Extending the ALCOVE Model of Category Learning to Featural Stimulus Domains". In: *Psychonomic Bulletin & Review* 9.1 (Mar. 1, 2002), pp. 43–58. ISSN: 1531-5320. DOI: 10.3758/BF03196256. URL: https://doi.org/10.3758/BF03196256 (visited on 02/15/2021) (cit. on pp. 4, 6).

[9] Jessamyn Schertz and Emily J. Clare. "Phonetic Cue Weighting in Perception and Production". In: *WIREs Cognitive Science* 11.2 (2020), e1521. ISSN: 1939-5086. DOI: 10.1002/wcs.1521. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1521 (visited on 01/26/2021) (cit. on pp. 4, 5).

[10] John J Ohala et al., eds. *A Guide to the History of the Phonetic Sciences in the United States*. 1999. URL: https://escholarship.org/uc/item/6mr8317x#article_main (visited on 01/27/2021) (cit. on p. 4).

[11] Alvin M. Liberman and Ignatius G. Mattingly. "The Motor Theory of Speech Perception Revised". In: *Cognition* 21.1 (1985), pp. 1–36. ISSN: 00100277. DOI: 10.1016/0010-0277(85)90021-6 (cit. on p. 4).

[12] *Haskins Laboratories*. Aug. 9, 2020. URL: https://web.archive.org/web/20200809223413/http://www.haskins.yale.edu/featured/sws/sws.html (visited on 01/26/2021) (cit. on p. 4).

[13] Kathy M. Carbonell and Andrew J. Lotto. "Speech Is Not Special… Again." In: *Frontiers in psychology* 5 (June June 3, 2014), p. 427. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.00427. pmid: 24917830. URL: http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00427/abstract (visited on 02/02/2017) (cit. on p. 4).

[14] AJ Lotto, KR Kluender, and LL Holt. "Animal Models of Speech Perception Phenomena". In: *Chicago Linguistic Society* (1997). URL: https://www.researchgate.net/profile/Keith_Kluender/publication/237280984_(from_K._Singer_R._Eggert__G._Anderson_(Eds.)_Chicago_Linguistic_Society_Volume_33_(Chicago_Linguistic_Society_Chicago)._pp._357-367_(1997).)_Animal_models_of_speech_perception_phen (visited on 02/02/2017) (cit. on p. 4).

[15] Keith R. Kluender. "Contributions of Nonhuman Animal Models to Understanding Human Speech Perception". In: *The Journal of the Acoustical Society of America* 107.5 (May 2000), pp. 2835–2835. ISSN: 0001-4966. DOI: 10.1121/1.429153. URL: http://asa.scitation.org/doi/10.1121/1.429153 (visited on 02/02/2017) (cit. on p. 4).

[16] Patricia K. Kuhl. "A New View of Language Acquisition". In: *Proceedings of the National Academy of Sciences* 97.22 (Oct. 24, 2000), pp. 11850–11857. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.97.22.11850. pmid: 11050219. URL: https://www.pnas.org/content/97/22/11850 (visited on 07/28/2019) (cit. on p. 4).

[17] Patricia K. Kuhl. "Early Language Acquisition: Cracking the Speech Code". In: *Nature Reviews Neuroscience* 5.11 (11 Nov. 2004), pp. 831–843. ISSN: 1471-0048. DOI: 10.1038/nrn1533. URL: https://www.nature.com/articles/nrn1533 (visited on 01/15/2021) (cit. on pp. 4, 7).

[18] Yakov Kronrod, Emily Coppess, and Naomi H. Feldman. "A Unified Account of Categorical Effects in Phonetic Perception". In: *Psychonomic Bulletin & Review* 23.6 (Dec. 24, 2016), pp. 1681–1712. ISSN: 1069-9384. DOI: 10.3758/s13423-016-1049-y. URL: http://link.springer.com/10.3758/s13423-016-1049-y (visited on 01/20/2017) (cit. on p. 5).

[19] Robert E. Remez et al. "On the Perceptual Organization of Speech". In: *Psychological Review* 101.1 (Jan. 1994), pp. 129–156. ISSN: 0033-295X. DOI: 10.1037/0033-295X.101.1.129. pmid: 8121955 (cit. on p. 5).

[20] Ashish Vaswani et al. *Attention Is All You Need*. Dec. 5, 2017. arXiv: 1706.03762 [cs]. URL: http://arxiv.org/abs/1706.03762 (visited on 01/28/2021) (cit. on p. 5).

[21] B Elan Dresher. "The Contrastive Hierarchy in Phonology". In: *Contrast in phonology: theory, perception, acquisition* 13 (2008), p. 11. ISSN: 1718-3510. DOI: 10.1017/CBO9780511642005 (cit. on p. 5).

[22] G.N. Clements. "Feature Organization". In: *Encyclopedia of Language & Linguistics*. Elsevier, 2006, pp. 433–440. ISBN: 978-0-08-044854-1. DOI: 10.1016/B0-08-044854-2/00055-9. URL: https://linkinghub.elsevier.com/retrieve/pii/B0080448542000559 (visited on 01/30/2021) (cit. on p. 5).

[23] Morris Halle, Bert Vaux, and Andrew Wolfe. "On Feature Spreading and the Representation of Place of Articulation". In: *Linguistic Inquiry* 31.3 (July 1, 2000), pp. 387–444. ISSN: 0024-3892. DOI: 10.1162/002438900554398. URL: https://doi.org/10.1162/002438900554398 (visited on 01/29/2021) (cit. on p. 5).

[24] Pavel Iosad. "Vowel Reduction in Russian: No Phonetics in Phonology". In: (2012). DOI: 10.1017/S0022226712000102 (cit. on p. 6).

[25] Danielle J. Navarro. "Between the Devil and the Deep Blue Sea: Tensions Between Scientific Judgement and Statistical Model Selection". In: *Computational Brain & Behavior* 2.1 (Mar. 1, 2019), pp. 28–34. ISSN: 2522-087X. DOI: 10.1007/s42113-018-0019-z. URL: https://doi.org/10.1007/s42113-018-0019-z (visited on 02/15/2021) (cit. on p. 6).

[26] Mona Lindau. "The Story of /r/". In: *The Journal of the Acoustical Society of America* 67.S1 (Apr. 1, 1980), S27–S27. ISSN: 0001-4966. DOI: 10.1121/1.2018134. URL: https://asa.scitation.org/doi/abs/10.1121/1.2018134 (visited on 01/28/2021) (cit. on p. 6).

[27] A. M. Liberman. "Some Characteristics of Perception in the Speech Mode". In: *Research Publications - Association for Research in Nervous and Mental Disease* 48 (1970), pp. 238–254. ISSN: 0091-7443. pmid: 5458835 (cit. on p. 6).

[28] Keith R. Kluender, Christian E. Stilp, and Fernando Llanos Lucas. "Long-Standing Problems in Speech Perception Dissolve within an Information-Theoretic Perspective". In: *Attention, Perception, & Psychophysics* 81.4 (May 1, 2019), pp. 861–883. ISSN: 1943-393X. DOI: 10.3758/s13414-019-01702-x. URL: https://doi.org/10.3758/s13414-019-01702-x (visited on 07/27/2019) (cit. on pp. 6, 7).

[29] Keith R. Kluender, Christian E. Stilp, and Michael Kiefte. "Perception of Vowel Sounds Within a Biologically Realistic Model of Efficient Coding". In: *Vowel Inherent Spectral Change*. Ed. by Geoffrey Stewart Morrison and Peter F. Assmann. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 117–151. ISBN: 978-3-642-14208-6 978-3-642-14209-3. DOI: 10.1007/978-3-642-14209-3_6. URL: http://link.springer.com/10.1007/978-3-642-14209-3_6 (visited on 11/09/2018) (cit. on pp. 6, 7).

[30] Christian E. Stilp and Keith R. Kluender. "Efficient Coding and Statistically Optimal Weighting of Covariance among Acoustic Attributes in Novel Sounds". In: *PLoS ONE* 7.1 (Jan. 23, 2012). Ed. by David S. Vicario, e30845. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0030845. URL: https://dx.plos.org/10.1371/journal.pone.0030845 (visited on 11/09/2018) (cit. on pp. 6, 11).

[31] C. E. Stilp, T. T. Rogers, and K. R. Kluender. "Rapid Efficient Coding of Correlated Complex Acoustic Properties". In: *Proceedings of the National Academy of Sciences* 107.50 (Dec. 14, 2010), pp. 21914–21919. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1009020107. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1009020107 (visited on 11/09/2018) (cit. on pp. 6, 8).

[32] Evan C. Smith and Michael S. Lewicki. "Efficient Auditory Coding". In: *Nature* 439.7079 (Feb. 23, 2006), pp. 978–982. ISSN: 1476-4687. DOI: 10.1038/nature04485. pmid: 16495999 (cit. on pp. 7, 8, 11).

[33] Maria N Geffen et al. "Auditory Perception of Self-Similarity in Water Sounds." In: *Front Integr Neurosci* 5 (May 2011), p. 15. ISSN: 1662-5145. DOI: 10.3389/fnint.2011.00015. pmid: 21617734. URL: http://dx.doi.org/10.3389/fnint.2011.00015 (cit. on p. 7).

[34] Michael Kiefte and Keith R. Kluender. "Absorption of Reliable Spectral Characteristics in Auditory Perception". In: *The Journal of the Acoustical Society of America* 123.1 (Jan. 1, 2008), pp. 366–376. ISSN: 0001-4966. DOI: 10.1121/1.2804951. URL: https://asa.scitation.org/doi/10.1121/1.2804951 (visited on 01/21/2021) (cit. on p. 7).

[35] Shi Tong Liu et al. "Optimal Features for Auditory Categorization". In: *Nature Communications* 10.1 (1 Mar. 21, 2019), p. 1302. ISSN: 2041-1723. DOI: 10.1038/s41467-019-09115-y. URL: https://www.nature.com/articles/s41467-019-09115-y (visited on 02/01/2021) (cit. on pp. 7, 11).

[36] Paul Iverson and Patricia K. Kuhl. "Influences of Phonetic Identification and Category Goodness on American Listeners' Perception of /r/ and /l/". In: *The Journal of the Acoustical Society of America* 99.2 (Feb. 1, 1996), pp. 1130–1140. ISSN: 0001-4966. DOI: 10.1121/1.415234. URL: https://asa.scitation.org/doi/abs/10.1121/1.415234 (visited on 01/20/2021) (cit. on p. 7).

[37] Pamela Souza et al. "Reliability and Repeatability of the Speech Cue Profile". In: *Journal of speech, language, and hearing research: JSLHR* 61.8 (Aug. 8, 2018), pp. 2126–2137. ISSN: 1558-9102. DOI: 10.1044/2018_JSLHR-H-17-0341. pmid: 30073277 (cit. on p. 7).

[38] Dave F. Kleinschmidt and T. Florian Jaeger. "Robust Speech Perception: Recognize the Familiar, Generalize to the Similar, and Adapt to the Novel." In: *Psychological Review* 122.2 (Apr. 2015), pp. 148–203. ISSN: 1939-1471, 0033-295X. DOI: 10.1037/a0038695. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0038695 (visited on 02/28/2019) (cit. on p. 7).

[39] R. E. Remez et al. "Speech Perception without Traditional Speech Cues". In: *Science* 212.4497 (May 22, 1981), pp. 947–949. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.7233191. pmid: 7233191. URL: https://science.sciencemag.org/content/212/4497/947 (visited on 01/26/2021) (cit. on p. 7).

[40] Matthew H. Davis et al. "Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence from the Comprehension of Noise-Vocoded Sentences". In: *Journal of Experimental Psychology. General* 134.2 (May 2005), pp. 222–241. ISSN: 0096-3445. DOI: 10.1037/0096-3445.134.2.222. pmid: 15869347 (cit. on p. 7).

[41] Taffeta M. Elliott and Frédéric E. Theunissen. "The Modulation Transfer Function for Speech Intelligibility". In: *PLOS Computational Biology* 5.3 (Mar. 6, 2009), e1000302. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000302. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000302 (visited on 02/01/2021) (cit. on p. 7).

[42] Justin J. Couchman, Mariana V. C. Coutinho, and J. David Smith. "Rules and Resemblance: Their Changing Balance in the Category Learning of Humans (Homo Sapiens) and Monkeys (Macaca Mulatta)". In: *Journal of experimental psychology. Animal behavior processes* 36.2 (Apr. 2010), pp. 172–183. ISSN: 0097-7403. DOI: 10.1037/a0016748. pmid: 20384398. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2890302/ (visited on 01/12/2021) (cit. on pp. 7, 8).

[43] Stephen E. G. Lea and A. J. Wills. "Use of Multiple Dimensions in Learned Discriminations". In: *Comparative Cognition & Behavior Reviews* 3 (2008). ISSN: 19114745. DOI: 10.3819/ccbr.2008.30007. URL: http://comparative-cognition-and-behavior-reviews.org/2008/vol3_lea_wills/ (visited on 01/13/2021) (cit. on p. 7).

[44] L. L. Holt, A. J. Lotto, and K. R. Kluender. "Neighboring Spectral Content Influences Vowel Identification". In: *The Journal of the Acoustical Society of America* 108.2 (Aug. 2000), pp. 710–722. ISSN: 0001-4966. DOI: 10.1121/1.429604. pmid: 10955638 (cit. on p. 7).

[45] Lori L. Holt. "The Mean Matters: Effects of Statistically Defined Nonspeech Spectral Distributions on Speech Categorization". In: *The Journal of the Acoustical Society of America* 120 (5 Pt 1 Nov. 2006), pp. 2801–2817. ISSN: 0001-4966. DOI: 10.1121/1.2354071. pmid: 17091133 (cit. on p. 7).

[46] Lori L. Holt. "Temporally Nonadjacent Nonlinguistic Sounds Affect Speech Categorization". In: *Psychological Science* 16.4 (Apr. 2005), pp. 305–312. ISSN: 0956-7976. DOI: 10.1111/j.0956-7976.2005.01532.x. pmid: 15828978 (cit. on p. 7).

[47] Patricia K Kuhl et al. "Phonetic Learning as a Pathway to Language: New Data and Native Language Magnet Theory Expanded (NLM-e)". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1493 (Mar. 12, 2008), pp. 979–1000. DOI: 10.1098/rstb.2007.2154. URL: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2007.2154 (visited on 01/15/2021) (cit. on p. 7).

[48] *Foreign-Language Experience in Infancy: Effects of Short-Term Exposure and Social Interaction on Phonetic Learning | PNAS*. URL: https://www.pnas.org/content/100/15/9096 (visited on 01/16/2021) (cit. on p. 7).

[49] Iris van Rooij and Giosuè Baggio. "Theory before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science". In: (Feb. 28, 2020). DOI: 10.31234/osf.io/7qbpr. URL: https://psyarxiv.com/7qbpr/ (visited on 02/15/2021) (cit. on p. 8).

[50] Patricia K. Kuhl. "Brain Mechanisms in Early Language Acquisition". In: *Neuron* 67.5 (Sept. 9, 2010), pp. 713–727. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2010.08.038. URL: http://www.sciencedirect.com/science/article/pii/S0896627310006811 (visited on 01/15/2021) (cit. on p. 8).

[51] Andrew J. King, Sundeep Teki, and Ben D. B. Willmore. "Recent Advances in Understanding the Auditory Cortex". In: *F1000Research* 7 (2018). ISSN: 2046-1402. DOI: 10.12688/f1000research.15580.1. pmid: 30345008 (cit. on pp. 8, 9, 11).

[52] Jennifer K. Schiavo and Robert C. Froemke. "Capacities and Neural Mechanisms for Auditory Statistical Learning across Species". In: *Hearing Research*. Annual Reviews 2019 376 (May 1, 2019), pp. 97–110. ISSN: 0378-5955. DOI: 10.1016/j.heares.2019.02.002. URL: http://www.sciencedirect.com/science/article/pii/S0378595518304441 (visited on 08/20/2019) (cit. on p. 8).

[53] H B Barlow. "Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology?" In: *Perception* 1.4 (Dec. 1, 1972), pp. 371–394. ISSN: 0301-0066. DOI: 10.1068/p010371. URL: https://doi.org/10.1068/p010371 (visited on 02/05/2021) (cit. on p. 8).

[54] Jennifer K. Bizley et al. "Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex". In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 29.7 (Feb. 18, 2009), pp. 2064–2075. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.4755-08.2009. pmid: 19228960 (cit. on p. 9).

[55] Kerry M. M. Walker et al. "Multiplexed and Robust Representations of Sound Features in Auditory Cortex". In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31.41 (Oct. 12, 2011), pp. 14565–14576. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.2074-11.2011. pmid: 21994373 (cit. on p. 9).

[56] Xiaoqin Wang et al. "Sustained Firing in Auditory Cortex Evoked by Preferred Stimuli." In: *Nature* 435.7040 (2005), pp. 341–6. ISSN: 1476-4687. DOI: 10.1038/nature03565. pmid: 15902257 (cit. on p. 9).

[57] Craig A. Atencio and Tatyana O. Sharpee. "Multidimensional Receptive Field Processing by Cat Primary Auditory Cortical Neurons". In: *Neuroscience* 359 (Sept. 17, 2017), pp. 130–141. ISSN: 1873-7544. DOI: 10.1016/j.neuroscience.2017.07.003. pmid: 28694174 (cit. on p. 9).

[58] Tatyana O Sharpee, Craig A Atencio, and Christoph E Schreiner. "Hierarchical Representations in the Auditory Cortex". In: *Current Opinion in Neurobiology*. Networks, Circuits and Computation 21.5 (Oct. 1, 2011), pp. 761–767. ISSN: 0959-4388. DOI: 10.1016/j.conb.2011.05.027. URL: http://www.sciencedirect.com/science/article/pii/S095943881100095X (visited on 02/04/2021) (cit. on p. 9).

[59] Matthew V. Macellaio et al. "Why Sensory Neurons Are Tuned to Multiple Stimulus Features". In: *bioRxiv* (Dec. 30, 2020), p. 2020.12.29.424235. DOI: 10.1101/2020.12.29.424235. URL: https://www.biorxiv.org/content/10.1101/2020.12.29.424235v1 (visited on 01/09/2021) (cit. on pp. 9, 14).

[60] C Angeloni and MN Geffen. "Contextual Modulation of Sound Processing in the Auditory Cortex". In: *Current Opinion in Neurobiology*. Neurobiology of Behavior 49 (Apr. 1, 2018), pp. 8–15. ISSN: 0959-4388. DOI: 10.1016/j.conb.2017.10.012. URL: https://www.sciencedirect.com/science/article/pii/S0959438817302325 (visited on 02/05/2021) (cit. on p. 9).

[61] Isabel Dean et al. "Rapid Neural Adaptation to Sound Level Statistics". In: *Journal of Neuroscience* 28.25 (June 18, 2008), pp. 6430–6438. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0470-08.2008. pmid: 18562614. URL: https://www.jneurosci.org/content/28/25/6430 (visited on 02/05/2021) (cit. on p. 9).

[62] Neil C. Rabinowitz et al. “Contrast Gain Control in Auditory Cortex”. In: *Neuron* 70.6 (June 23, 2011), pp. 1178–1191. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2011.04.030. pmid: 21689603 (cit. on p. 9).

[63] Neil C. Rabinowitz et al. “Constructing Noise-Invariant Representations of Sound in the Auditory Pathway”. In: *PLoS biology* 11.11 (Nov. 2013), e1001710. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1001710. pmid: 24265596 (cit. on p. 9).

[64] Nima Mesgarani et al. “Mechanisms of Noise Robust Representation of Speech in Primary Auditory Cortex”. In: *Proceedings of the National Academy of Sciences of the United States of America* 111.18 (May 6, 2014), pp. 6792–6797. ISSN: 1091-6490. DOI: 10.1073/pnas.1318017111. pmid: 24753585 (cit. on p. 9).

[65] Stephen V. David et al. “Rapid Synaptic Depression Explains Nonlinear Modulation of Spectro-Temporal Tuning in Primary Auditory Cortex by Natural Stimuli”. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 29.11 (Mar. 18, 2009), pp. 3374–3386. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.5249-08.2009. pmid: 19295144 (cit. on p. 9).

[66] Ryan G Natan et al. “Complementary Control of Sensory Adaptation by Two Types of Cortical Interneurons”. In: *eLife* 4 (Oct. 13, 2015). Ed. by Andrew J King, e09868. ISSN: 2050-084X. DOI: 10.7554/eLife.09868. URL: https://doi.org/10.7554/eLife.09868 (visited on 02/04/2021) (cit. on p. 9).

[67] Ryan G. Natan, Winnie Rao, and Maria N. Geffen. “Cortical Interneurons Differentially Shape Frequency Tuning Following Adaptation”. In: *Cell Reports* 21.4 (Oct. 24, 2017), pp. 878–890. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2017.10.012. URL: http://www.sciencedirect.com/science/article/pii/S2211124717314298 (visited on 02/04/2021) (cit. on p. 9).

[68] Jonathan Fritz et al. “Rapid Task-Related Plasticity of Spectrotemporal Receptive Fields in Primary Auditory Cortex”. In: *Nature Neuroscience* 6.11 (Nov. 2003), pp. 1216–1223. ISSN: 1097-6256. DOI: 10.1038/nn1141. pmid: 14583754 (cit. on p. 9).

[69] Jonathan Fritz, Mounya Elhilali, and Shihab Shamma. “Active Listening: Task-Dependent Plasticity of Spectrotemporal Receptive Fields in Primary Auditory Cortex”. In: *Hearing Research* 206.1-2 (Aug. 2005), pp. 159–176. ISSN: 0378-5955. DOI: 10.1016/j.heares.2005.01.015. pmid: 16081006 (cit. on p. 9).

[70] Stephen V. David, Jonathan B. Fritz, and Shihab A. Shamma. “Task Reward Structure Shapes Rapid Receptive Field Plasticity in Auditory Cortex”. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.6 (Feb. 7, 2012), pp. 2144–2149. ISSN: 1091-6490. DOI: 10.1073/pnas.1117717109. pmid: 22308415 (cit. on p. 9).

[71] Stephen V. David and Shihab A. Shamma. “Integration over Multiple Timescales in Primary Auditory Cortex”. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 33.49 (Dec. 4, 2013), pp. 19154–19166. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.2270-13.2013. pmid: 24305812 (cit. on p. 9).

[72] D. B. Polley. “Perceptual Learning Directs Auditory Cortical Map Reorganization through Top-Down Influences”. In: *Journal of Neuroscience* 26.18 (2006), pp. 4970–4982. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.3771-05.2006. pmid: 16672673. URL: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3771-05.2006 (cit. on p. 9).

[73] Kasia M. Bieszczad and Norman M. Weinberger. “Representational Gain in Cortical Area Underlies Increase of Memory Strength”. In: *Proceedings of the National Academy of Sciences of the United States of America* 107.8 (Feb. 23, 2010), pp. 3793–3798. ISSN: 1091-6490. DOI: 10.1073/pnas.1000159107. pmid: 20133679 (cit. on p. 9).

[74] Han Gyol Yi, Matthew K. Leonard, and Edward F. Chang. “The Encoding of Speech Sounds in the Superior Temporal Gyrus”. In: *Neuron* 102.6 (June 19, 2019), pp. 1096–1110. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2019.04.023. pmid: 31220442. URL: https://www.cell.com/neuron/abstract/S0896-6273(19)30380-0 (visited on 07/28/2019) (cit. on p. 9).

[75] N. Mesgarani et al. “Phonetic Feature Encoding in Human Superior Temporal Gyrus”. In: *Science* 343.6174 (Feb. 28, 2014), pp. 1006–1010. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1245994. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1245994 (visited on 11/09/2018) (cit. on pp. 9, 10).

[76] Edward F. Chang et al. “Categorical Speech Representation in Human Superior Temporal Gyrus”. In: *Nature Neuroscience* 13.11 (11 Nov. 2010), pp. 1428–1432. ISSN: 1546-1726. DOI: 10.1038/nn.2641. URL: https://www.nature.com/articles/nn.2641 (visited on 01/21/2021) (cit. on pp. 9, 10).

[77] Alexander M. Chan et al. "Speech-Specific Tuning of Neurons in Human Superior Temporal Gyrus". In: *Cerebral Cortex* 24.10 (Oct. 1, 2014), pp. 2679–2693. ISSN: 1047-3211. DOI: 10.1093/cercor/bht127. URL: https://doi.org/10.1093/cercor/bht127 (visited on 01/17/2021) (cit. on pp. 9, 10).

[78] Liberty S. Hamilton, Erik Edwards, and Edward F. Chang. "A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus". In: *Current Biology* 28.12 (June 18, 2018), 1860–1871.e4. ISSN: 0960-9822. DOI: 10.1016/j.cub.2018.04.033. URL: http://www.sciencedirect.com/science/article/pii/S0960982218304615 (visited on 06/21/2019) (cit. on p. 10).

[79] Crystal T. Engineer et al. "Speech Training Alters Consonant and Vowel Responses in Multiple Auditory Cortex Fields". In: *Behavioural Brain Research* 287 (July 1, 2015), pp. 256–264. ISSN: 0166-4328. DOI: 10.1016/j.bbr.2015.03.044. URL: http://www.sciencedirect.com/science/article/pii/S0166432815002090 (visited on 02/03/2021) (cit. on p. 10).

[80] Jonny L. Saunders and Michael Wehr. "Mice Can Learn Phonetic Categories". In: *The Journal of the Acoustical Society of America* 145.3 (Mar. 1, 2019), pp. 1168–1177. ISSN: 0001-4966. DOI: 10.1121/1.5091776. URL: https://asa.scitation.org/doi/abs/10.1121/1.5091776 (visited on 08/12/2019) (cit. on p. 10).

[81] H. Takahashi et al. "Learning-Stage-Dependent, Field-Specific, Map Plasticity in the Rat Auditory Cortex during Appetitive Operant Conditioning". In: *Neuroscience* 199 (Dec. 29, 2011), pp. 243–258. ISSN: 0306-4522. DOI: 10.1016/j.neuroscience.2011.09.046. URL: https://www.sciencedirect.com/science/article/pii/S0306452211011316 (visited on 02/23/2021) (cit. on p. 10).

[82] Zhiyue Shi et al. "Anterior Auditory Field Is Needed for Sound Categorization in Fear Conditioning Task of Adult Rat". In: *Frontiers in Neuroscience* 13 (2019). ISSN: 1662-453X. DOI: 10.3389/fnins.2019.01374. URL: https://www.frontiersin.org/articles/10.3389/fnins.2019.01374/full (visited on 02/22/2021) (cit. on p. 10).

[83] Andres Carrasco and Stephen G. Lomber. "Neuronal Activation Times to Simple, Complex, and Natural Sounds in Cat Primary and Nonprimary Auditory Cortex". In: *Journal of Neurophysiology* 106.3 (June 8, 2011), pp. 1166–1178. ISSN: 0022-3077. DOI: 10.1152/jn.00940.2010. URL: https://journals.physiology.org/doi/full/10.1152/jn.00940.2010 (visited on 02/23/2021) (cit. on p. 10).

[84] Jennifer F. Linden et al. "Spectrotemporal Structure of Receptive Fields in Areas AI and AAF of Mouse Auditory Cortex". In: *Journal of Neurophysiology* 90.4 (Oct. 1, 2003), pp. 2660–2675. ISSN: 0022-3077. DOI: 10.1152/jn.00751.2002. URL: https://journals.physiology.org/doi/full/10.1152/jn.00751.2002 (visited on 02/23/2021) (cit. on p. 10).

[85] Pritesh K. Pandya et al. "Spectral and Temporal Processing in Rat Posterior Auditory Cortex". In: *Cerebral Cortex* 18.2 (Feb. 1, 2008), pp. 301–314. ISSN: 1047-3211. DOI: 10.1093/cercor/bhm055. URL: https://doi.org/10.1093/cercor/bhm055 (visited on 02/25/2021) (cit. on p. 10).

[86] Andres Carrasco and Stephen G. Lomber. "Evidence for Hierarchical Processing in Cat Auditory Cortex: Nonreciprocal Influence of Primary Auditory Cortex on the Posterior Auditory Field". In: *Journal of Neuroscience* 29.45 (Nov. 11, 2009), pp. 14323–14333. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2905-09.2009. pmid: 19906979. URL: https://www.jneurosci.org/content/29/45/14323 (visited on 02/25/2021) (cit. on p. 10).

[87] Xiaoqin Wang. "Neural Coding Strategies in Auditory Cortex". In: *Hearing Research*. Auditory Cortex 2006 - The Listening Brain 229.1 (July 1, 2007), pp. 81–93. ISSN: 0378-5955. DOI: 10.1016/j.heares.2007.01.019. URL: http://www.sciencedirect.com/science/article/pii/S0378595507000366 (visited on 01/20/2021) (cit. on p. 11).

[88] Zoe C. Ashwood et al. *Mice Alternate between Discrete Strategies during Perceptual Decision-Making*. preprint. Neuroscience, Oct. 21, 2020. DOI: 10.1101/2020.10.19.346353. URL: http://biorxiv.org/lookup/doi/10.1101/2020.10.19.346353 (visited on 01/09/2021) (cit. on pp. 11, 14).

[89] Sam V. Norman-Haignere et al. "Hierarchical Integration across Multiple Timescales in Human Auditory Cortex". In: *bioRxiv* (Oct. 1, 2020), p. 2020.09.30.321687. DOI: 10.1101/2020.09.30.321687. URL: https://www.biorxiv.org/content/10.1101/2020.09.30.321687v1 (visited on 01/09/2021) (cit. on pp. 11, 14, 15).

[90] Fernando E. Rosas et al. *Causal Blankets: Theory and Algorithmic Framework*. Sept. 29, 2020. arXiv: 2008.12568 [nlin, q-bio]. URL: http://arxiv.org/abs/2008.12568 (visited on 01/09/2021) (cit. on pp. 11, 14).

[91] Alexis Dubreuil et al. "Complementary Roles of Dimensionality and Population Structure in Neural Computations". In: *bioRxiv* (July 4, 2020), p. 2020.07.03.185942. DOI: 10.1101/2020.07.03.185942. URL: https://www.biorxiv.org/content/10.1101/2020.07.03.185942v1 (visited on 01/09/2021) (cit. on pp. 11, 14).

[92] Chethan Pandarinath et al. "Inferring Single-Trial Neural Population Dynamics Using Sequential Auto-Encoders". In: *Nature Methods* 15.10 (10 Oct. 2018), pp. 805–815. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0109-9. URL: https://www.nature.com/articles/s41592-018-0109-9 (visited on 01/17/2021) (cit. on p. 11).

[93] Björn Brembs. "The Brain as a Dynamically Active Organ". In: *Biochemical and Biophysical Research Communications* (Dec. 11, 2020). ISSN: 0006-291X. DOI: 10.1016/j.bbrc.2020.12.011. URL: http://www.sciencedirect.com/science/article/pii/S0006291X20321872 (visited on 01/17/2021) (cit. on p. 11).

[94] Juan A. Gallego et al. "Cortical Population Activity within a Preserved Neural Manifold Underlies Multiple Motor Behaviors". In: *Nature Communications* 9.1 (1 Oct. 12, 2018), p. 4233. ISSN: 2041-1723. DOI: 10.1038/s41467-018-06560-z. URL: https://www.nature.com/articles/s41467-018-06560-z (visited on 01/09/2021) (cit. on pp. 11, 14).

[95] Eleanor Rosch. "Wittgenstein and Categorization Research in Cognitive Psychology". In: *Meaning and the Growth of Understanding: Wittgenstein's Significance for Developmental Psychology*. Ed. by Michael Chapman and Roger A. Dixon. Berlin, Heidelberg: Springer, 1987, pp. 151–166. ISBN: 978-3-642-83023-5. DOI: 10.1007/978-3-642-83023-5_9. URL: https://doi.org/10.1007/978-3-642-83023-5_9 (visited on 01/12/2021) (cit. on p. 11).

[96] R. C. deCharms, D. T. Blake, and M. M. Merzenich. "Optimizing Sound Features for Cortical Neurons". In: *Science (New York, N.Y.)* 280.5368 (May 29, 1998), pp. 1439–1443. ISSN: 0036-8075. DOI: 10.1126/science.280.5368.1439. pmid: 9603734 (cit. on p. 11).

[97] Misha B. Ahrens, Jennifer F. Linden, and Maneesh Sahani. "Nonlinearities and Contextual Influences in Auditory Cortical Responses Modeled with Multilinear Spectrotemporal Methods". In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 28.8 (Feb. 20, 2008), pp. 1929–1942. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.3377-07.2008. pmid: 18287509 (cit. on p. 11).

[98] Friedrich Schuessler et al. *The Interplay between Randomness and Structure during Learning in RNNs*. Oct. 25, 2020. arXiv: 2006.11036 [q-bio]. URL: http://arxiv.org/abs/2006.11036 (visited on 01/09/2021) (cit. on p. 12).

[99] David H Krantz and Amos Tversky. "Similarity of Rectangles: An Analysis of Subjective Dimensions". In: *Journal of Mathematical Psychology* 12.1 (Feb. 1975), pp. 4–34. ISSN: 00222496. DOI: 10.1016/0022-2496(75)90047-4. URL: https://linkinghub.elsevier.com/retrieve/pii/0022249675900474 (visited on 02/28/2019) (cit. on p. 13).

[100] Erin Goddard et al. "Interpreting the Dimensions of Neural Feature Representations Revealed by Dimensionality Reduction". In: *NeuroImage*. New Advances in Encoding and Decoding of Brain Signals 180 (Oct. 15, 2018), pp. 41–67. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2017.06.068. URL: http://www.sciencedirect.com/science/article/pii/S1053811917305396 (visited on 01/14/2021) (cit. on p. 13).

[101] Daniel Durstewitz et al. "Abrupt Transitions between Prefrontal Neural Ensemble States Accompany Behavioral Transitions during Rule Learning". In: *Neuron* 66.3 (May 13, 2010), pp. 438–448. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2010.03.029. URL: http://www.sciencedirect.com/science/article/pii/S0896627310002321 (visited on 01/09/2021) (cit. on p. 14).

[102] Matthew G. Perich, Juan A. Gallego, and Lee E. Miller. "A Neural Population Mechanism for Rapid Learning". In: *Neuron* 100.4 (Nov. 21, 2018), 964–976.e7. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2018.09.030. pmid: 30344047. URL: https://www.cell.com/neuron/abstract/S0896-6273(18)30832-8 (visited on 01/09/2021) (cit. on p. 14).

[103] Mark R. Saddler, Ray Gonzalez, and Josh H. McDermott. *Deep Neural Network Models Reveal Interplay of Peripheral Coding and Stimulus Statistics in Pitch Perception*. preprint. Animal Behavior and Cognition, Nov. 20, 2020. DOI: 10.1101/2020.11.19.389999. URL: http://biorxiv.org/lookup/doi/10.1101/2020.11.19.389999 (visited on 01/09/2021) (cit. on p. 14).

[104] Francesca Mastrogiuseppe and Srdjan Ostojic. "Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks". In: *Neuron* 99.3 (Aug. 8, 2018), 609–623.e29. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2018.07.003. URL: http://www.sciencedirect.com/science/article/pii/S0896627318305439 (visited on 01/09/2021) (cit. on p. 14).

[105] Greta Kaufeld et al. “Linguistic Structure and Meaning Organize Neural Oscillations into a Content-Specific Hierarchy”. In: *Journal of Neuroscience* 40.49 (Dec. 2, 2020), pp. 9467–9475. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0302-20.2020. pmid: 33097640. URL: https://www.jneurosci.org/content/40/49/9467 (visited on 01/09/2021) (cit. on p. 14).

[106] Manuel Beiran et al. *Shaping Dynamics with Multiple Populations in Low-Rank Recurrent Networks*. Nov. 17, 2020. arXiv: 2007.02062 [q-bio]. URL: http://arxiv.org/abs/2007.02062 (visited on 01/09/2021) (cit. on p. 14).

[107] Ines Hipolito et al. *Markov Blankets in the Brain*. June 4, 2020. arXiv: 2006.02741 [physics, q-bio]. URL: http://arxiv.org/abs/2006.02741 (visited on 01/09/2021) (cit. on p. 14).

[108] Philip R. L. Parker et al. “Movement-Related Signals in Sensory Areas: Roles in Natural Behavior”. In: *Trends in Neurosciences* 43.8 (Aug. 1, 2020), pp. 581–595. ISSN: 0166-2236, 1878-108X. DOI: 10.1016/j.tins.2020.05.005. pmid: 32580899. URL: https://www.cell.com/trends/neurosciences/abstract/S0166-2236(20)30123-5 (visited on 01/09/2021) (cit. on p. 14).

[109] Matthew Warburton et al. “Getting Stuck in a Rut as an Emergent Feature of a Dynamic Decision-Making System”. In: *bioRxiv* (June 3, 2020), p. 2020.06.02.127860. DOI: 10.1101/2020.06.02.127860. URL: https://www.biorxiv.org/content/10.1101/2020.06.02.127860v1 (visited on 01/09/2021) (cit. on p. 14).

[110] Sukbin Lim et al. “Inferring Learning Rules from Distributions of Firing Rates in Cortical Neurons”. In: *Nature Neuroscience* 18.12 (12 Dec. 2015), pp. 1804–1810. ISSN: 1546-1726. DOI: 10.1038/nn.4158. URL: https://www.nature.com/articles/nn.4158 (visited on 01/09/2021) (cit. on p. 14).

[111] Federico Battiston et al. *Networks beyond Pairwise Interactions: Structure and Dynamics*. June 2, 2020. arXiv: 2006.01764 [cond-mat, physics:nlin, physics:physics, q-bio]. URL: http://arxiv.org/abs/2006.01764 (visited on 01/09/2021) (cit. on p. 14).

[112] Hiroyuki K. Kato et al. “Dynamic Sensory Representations in the Olfactory Bulb: Modulation by Wakefulness and Experience”. In: *Neuron* 76.5 (Dec. 6, 2012), pp. 962–975. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2012.09.037. pmid: 23217744. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3523713/ (visited on 01/09/2021) (cit. on p. 14).

[113] Juan A. Gallego et al. “Neural Manifolds for the Control of Movement”. In: *Neuron* 94.5 (June 7, 2017), pp. 978–984. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2017.05.025. pmid: 28595054 (cit. on p. 14).

[114] Tony Hyun Kim et al. “Long-Term Optical Access to an Estimated One Million Neurons in the Live Mouse Cortex”. In: *Cell Reports* 17.12 (Dec. 20, 2016), pp. 3385–3394. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2016.12.004. pmid: 28009304. URL: https://www.cell.com/cell-reports/abstract/S2211-1247(16)31676-X (visited on 01/09/2021) (cit. on p. 14).

[115] Robert B. Levy et al. “Circuit Asymmetries Underlie Functional Lateralization in the Mouse Auditory Cortex”. In: *Nature Communications* 10.1 (1 June 25, 2019), p. 2783. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10690-3. URL: https://www.nature.com/articles/s41467-019-10690-3 (visited on 01/06/2021) (cit. on p. 15).

[116] Sung-Joo Lim, Julie A. Fiez, and Lori L. Holt. “How May the Basal Ganglia Contribute to Auditory Categorization and Speech Perception?” In: *Frontiers in Neuroscience* 8 (2014). ISSN: 1662-453X. DOI: 10.3389/fnins.2014.00230. URL: https://www.frontiersin.org/articles/10.3389/fnins.2014.00230/full (visited on 01/16/2021) (cit. on p. 15).

[117] Josef P Rauschecker and Sophie K Scott. “Maps and Streams in the Auditory Cortex: Nonhuman Primates Illuminate Human Speech Processing”. In: *Nature neuroscience* 12.6 (June 2009), pp. 718–724. ISSN: 1097-6256. DOI: 10.1038/nn.2331. pmid: 19471271. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2846110/ (visited on 02/02/2021) (cit. on p. 15).

[118] Meghan Clayards. “Differences in Cue Weights for Speech Perception Are Correlated for Individuals within and across Contrasts”. In: *The Journal of the Acoustical Society of America* 144.3 (Sept. 1, 2018), EL172–EL177. ISSN: 0001-4966. DOI: 10.1121/1.5052025. URL: https://asa.scitation.org/doi/10.1121/1.5052025 (visited on 01/22/2019) (cit. on p. 15).

[119] Isaac M. Carruthers et al. "Emergence of Invariant Representation of Vocalizations in the Auditory Cortex". In: *Journal of Neurophysiology* 114.5 (Aug. 26, 2015), pp. 2726–2740. ISSN: 0022-3077. DOI: 10.1152/jn.00095.2015. URL: https://journals.physiology.org/doi/full/10.1152/jn.00095.2015 (visited on 02/04/2021) (cit. on p. 15).