

Prelims

Jonny Saunders
January 20, 2021

Contents

1	Abstract	3
2	Introduction	3
2.1	Outline	3
2.2	Phonemes are Language Games	3
2.3	Learning to play	5
2.4	<some of that neural theories of phonetic processing	5
2.5	models	6
2.6	scraps	7
3	Methods	7
3.1	Scraps	7
3.2	behavior	7
3.3	imaging	8
3.4	analysis & modeling	8
4	Specific Aims	8
5	Significance & Broader Impacts	8
6	Notes	8
6.1	Bailey & Summerfield - 1980	8
7	meta	9
7.1	to-read	9
7.2	bookmarks	10

1. Abstract

[abstract itself]

[summary of intellectual 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 Outline

- Language Games
 - Category structure of phonemes
 - family resemblances in cognitive lit - why is phonetic identification a family resemblance structure?
 - general statement of geometry of perceptual spaces
 - family resemblances that differ in many senses really are defined by contrast between them – which sense lets me distinguish the objects in question? put another way, selecting or rotating axes depending on short-term acoustic statistics.
 - update to idea of perceptual geometry to include contextual dependence and reweighting
 - Primary object is to learn the axes – rather than learning some prebuilt structure that 'exists in the world'; brain has to figure out what parts of the world are informative, which is necessarily in context
- Learning to play
 - Infant speech learning, statistical regularity, perceptual warping
 - general statement about the normalization of redundancy and adaptation to statistical regularity as a fundamental part of the auditory system (or maybe a brief allusion to it and discuss in more detail in neurophys section)
 - But perceptual systems don't find 'totally optimal' solutions as might be predicted from simpler experiments...
 - Representativeness means that there will be con-

tributions from all the feature axes, even when they're irrelevant in the particular context.

- Neurophys

2.2 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[63]

Cognitive reality is characterized by its discreteness: rather than a continuous undifferentiated gradient wash of sensation and cognition, we experience objects, concepts, and thoughts. Speech is a continuous, high-dimensional, high-variability acoustic signal, yet it is perceived as a small number of relatively-discrete phonemes[21]. The acoustic structure of phonemes is a sort of “Family Resemblance”[63] — 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 [39].

Evidence supporting the family resemblance argument

- phonetic contrasts are multiply defined [39, 2]
- Cognitive categorization learning mostly operates as family resemblances that have incomplete/nonplatonic feature sets that unite them ([50][49] [10])
- tversky talked about this in terms of set theory and resemblance [57] [59]
- People can switch discriminatory dimension depending on which is predictable vs. informative – people use whatever cue is present and informative [25]
- animals use family resemblance of multiple features even when there is a single dimension that is perfectly informative of category membership [35, 10]

Basic description of geometry of phonetic space

Category representation theories are intimately related (and occasionally literally isometric to [14]) to theories of the measurement of similarity, which is dominated by geometric models [58]. 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.

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)

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 [29] – 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? [29]

actually 'warping' perceptual space relative to acoustic space is already a really common idea in phonetics lit [22] and is a sorta trivial reformulation of the idea that the auditory system is learning to represent the maximally informative dimensions of the stimulus, so a perceptual warping is just a reflection of the condensation of representation of within-category variation (ie. not being represented/generalized over/compressed/whatever you want to call it) and a maximization of representation of the between-category variation. Accounts of exemplars and stimulus geometry are complementary here: saying that perceptual space is clustered near exemplars and sparser away from them is the same thing as

saying they are embedded in a space whose dimensions that maximize inter-category discriminability. Put another way, instances where there is not a clear exemplar to 'warp' perceptual space (as in the 'low-r' group in [22]) could also correspond to the absence of a clear perceptual dimension structure within the presented stimulus space: maybe those listeners discriminability feature dimensions don't feature F3 prominently, and in instances where clear exemplars warp the perceptual space, those dimensions are emphasized by increasing the weight of existing feature dimensions, or the perceptual space is 'rotated' to emphasize them.

Context effects are reweighting stimulus dimensions, necessarily dependent on the comparisons/dimensions present

the discussion of the failure of a fully-specified, prototype-style geometric model with a stable mean has been more fully articulated and argued in phonetics in the context of a contrastive hierarchy [11] – indeed this is echoed in [30] indirectly, in the sense that some cues are more informative than others and thus since vowels have big overarching feature descriptions they are more clearly separated, rely less on the internal featural model, and have a more continuous perception within the category. The problem with the contrastive hierarchy is that it presupposes unproblematically the featural dimensions (“+nasal” is an unambiguous, nonprobabilistic description), and if instead the featural dimensions are probabilistic, analogue, etc. then fully specified minimal pairs and contrastive hierarchies are not in conflict, but instead reflect the degree of information that is contained within a particular cue. Indeed [30] specifically say that how the phonetic representations is of paramount importance in their future directions. Basically what we're doing here is recasting this in a geometric lens that emphasizes the space.

relate back to similarity discussion at the beginning: in family resemblance categories, 'which' dimension you select depends on the possible other phonemes in the space, or the possible variants in acoustic features that could conflict depending on the particularity of the speaker's voice/etc.

The debate over cues being real is a potent one in linguistics and phonology “It is argued that it is inappropriate to ascribe a psychological status to cues whose only reality is their operational role as physical parameters whose manipulation can change the phonetic interpretation of a signal” [2] except that one doesn't need to appeal to literally trying to recover the articulatory event, instead you can appeal to trying to recover the acoustic consequences of the articulatory event – because these particular acoustic attributes covary, as they must given the integrated nature of the articulatory system, one can use that covariation as a hint for which of the cues should be infor-

mative at that time – eg. some “seemingly phonetically unrelated cues” (but not necessarily) can indicate which contrasts are valuable at the given time: or, what do the cues “mean” given the totality of the acoustic context.[6.1](#)

... so since the goal is not to prespecify cues/contrasts, recast the problem in terms of learning the dimensions of the stimulus space.

2.3 Learning to play

💡 - Start with demonstration from tensorflow.js playground that learns a simple 2-d nonlinearity by becoming sensitive to the product of the two ‘base’ dimensions. it’s a projection to a geometry that allows them to be discriminable. that’s the basic idea.

In learning to identify the phonemes present in a given language, one must learn how the particular acoustic features of a phonetic class are similar to other members of the class and different than members of a different class. Such features can be based on formant transitions, timing and duration of silent gaps, frication, etc. and any number of combinations of “raw” acoustic information into higher-order descriptions. Unless sensitivity, or more generally, “representation” of such higher-order features is innate, the task of learning learning phonemes is not that of learning “where” each phoneme is clustered in some pre-existing phonetic-feature space, but instead that of *learning which features are maximally informative to identify the phonemes.*[\[28\]](#) – (note that we are agnostic to implementation here, not saying maximally informative dimensions and citing kluender in order to uncritically endorse their information-theoretic framework (though we will later critically suggest it), we could just as easily learn a positive, generative model of phoentic categories. The argument is that needing to learn the features themselves, and perhaps tautologically, the features that are learned are the ones that are capable of supporting phonetic identification. also the contribution from basic acoustic features/structure of acoustic reality is real, see Kuhl’s ‘basic cuts’ argument [\[34\]](#))

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**[\[31\]](#)[\[34\]](#)), 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 discrminating those in a non-native language[\[31\]](#)

(babies initially can learn all phonemes[\[34\]](#), so they have to learn some feature which necessarily compresses the auditory space[\[16\]](#))

Arguably this is the central function of all sensory systems system - to exploit regularities in the statistical structure of sensory input to form a maximally efficient representation, *begin here again with* [\[33\]](#).

*and focusing on the acquisition of informative stimulus dimensions fundamentally alters the research question. The problem is the mutual translation/misunderstanding 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???’*

2.4 <some of that neural theories of phonetic processing

💡 - literally follow the bouncing ball along the bottom of your screen to this freaking paper [\[27\]](#)

start with description of why this should be preposterously difficult for the auditory system, but that normalizing over statistical regularities is the bread and butter of the auditory system, and if that translates into reweighting a nonlinear basis set of phonetic perception then the problem becomes a whole lot more tractable. (mebs revisit why [\[2\]](#) thought it would be impossible without resorting to motor theories of language)

speech categorization is a big neurolinguistic prob[\[64\]](#)

Rather than being indeependent ‘levels,’ algorithm and implementation have to be the same thing – the way that phonetic dimensions are implemented in the brain strongly constrains the possible types of dimensions that can be learned – eg. people have tried to explain how context can be incorporated in a ton of different ways.

It’s all about the left anterior superior temporal gyrus[\[64\]](#). Specifically, neurons in STG encode higher-order acoustic properties that correspond to those present in categories of

speech sounds (eg. frication vs. sonority, formant band combinations). Tuning isn't 'clean' – neighboring cells have dramatically different tuning, and all reflect some sort of complex spectrotemporal sensitivity (firing to specific speech sounds, but none to tones/simple sounds) (left aSTG)[7], and are very heterogeneous between people. combined with animal lit about developed sensitivity, it's probably the case that people learn their own basis sets for feature detection in secondary auditory cortical areas. Indeed different people have different cue weightings that are more or less adaptive[9]

*get putative mouse "analogue" from crystal engineer's papers
vocalization sensitive neurons in anterior left acx with different
projection patterns from/to L6 that are experience dependent.
(cfos[36])*

Reciprocal connections with straitum could facilitate the plasticity in cortex b/c dopaminergic projections responsive to reward [15]

Auditory system makes efficient codes that collapse uninformative variability[54, 55], and learns the statistical structure inherent in acoustic reality [52] and phonetic production specifically[32] – responses to sound become "non-isomorphic" to the acoustic features in the sound [56, 61] as dimensions that are more informative than raw acoustic features are computed. *not* representing the sound precisely is more efficient than representing it directly because then you can take advantage of the *informative* elements of the sound rather than the ones that are spandrels of the physics of the acoustic generator. —

Lucky for us... learning features is like exactly what deep neural networks do, and is a sorta trivial extension of another way of viewing populations: response profiles of neurons...

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. [19]

Brain does indeed learn and use multiple stimulus dimensions rather than computing stimulus dimensions independently — so behavioral results from family resemblance experiments actually should be expected[40]

This also merges us with kluender/stilp's work on efficient coding, removing unnecessary stimulus dimensions (reread/cite 'longstanding problems disappear in information theoretic framework')

—

Also, multidimensionally tuned neurons are like already there lol

—

so multidimensionally tuned neurons, family resemblance

data, and the highly-correlated spectral characteristics of sound all suggest that phonemes need to be interrogated in their natural complexity,

—

We've even seen neurons remap their receptive fields to represent maximally informative dimensions[47]

- auditory processing as domain-general and domain-specific across multiple timescales [43]
- why are auditory neurons potentially sensitive to multiple stimulus features/how does that contribute to generalizable ill-defined categories? [40]
- abrupt transitions, at least in neural data [13]
- other reward-learning regions like RSC [42]
- multimodal representations and preserved neural manifold dynamics across inference tasks in M1 [17]
- timescales of processing expand across auditory hierarchy (and more generally have different timescales of integration and lags) [43] and are lateralized [36]
- categorical representation of phonemes in STG, smooth gradients in F2 onset make discrete changes in linear readouts of "neural representation" [8]
- freaking the multiplexing of stimulus dimensions also exists in auditory cortical neurons doiiii [60, 5]
- contributions from basal ganglia in reward learning for acoustic dimensions [38]

probs w/ discriminatory models: how is the comparison done? eg. you could start learning features by just comparing every x thing with y thing, but then you would have to hold some representation of each in order to compare.

2.5 models

computational. models that have attempted to explain phonetic processing??? is this its own section or what?

zoo of processing models and discussion of bayesian generative models [30]. categorical effects are from large amount of 'noise' variance, or variance on uninformative dimensions. if it's the case that there are many dimensions that have imperfect, sometimes conflicting information, then that would be reflected in categorical perception. Their discussion asks the question what effects coarticulation might have on the meaning of tau, and this is a potential one – it could be the case that since the category structure is a family resemblance, and as such only a few of the cues are informative at a particular time, then

relationship between generative and discriminative models here... the means by which these features are learned is ultimately the question of implementation that grounds these orbiting ideas. How do family resemblances work? why is it possible that there are categories that operate without logical structure? why is it that we will use all the dimensions of a problem even when there is an optimal, low-dimension solution (contrast with techniques like SVM that without regularization inevitably converge on a 'one true feature' that can perfectly distinguish states). what are phonemes is a question of how are they implemented.

2.6 scraps

- Short description of phonetic acoustics, why they're games
- General statement on importance of understanding neural implementation of a game-recognition system
- 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 positive 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! [53]

levels of analysis:

phonetic perception has paradoxes at several levels of analysis that are not mutually discrete.

ontic/algorithmic: what *are* phonemes? are they positive descriptions of combinations of features, or negative descriptions of forbidden spectrotemporal state transitions?

implementation: to some degree the methodological and theoretical disagreements between the feature-detection and population-computation models of phonetic perception mirror the single-cell/multicellular computation dichotomy described in the introduction of [12].

- speed of processing vs. variability within category
- neurons that process auditory information at phonetic timescales are relatively insensitive to spectral quality [43]

3. Methods

3.1 Scraps

- Segmenting strategies [1]
- Scrambled vs. unscrambled sounds? (cites 12, 18, and 25 in [43])
- inferring perception-action loops from data [48]
- complementary roles of cell types and manifold dynamics [12]
- LFADS for sequential autoencoders [44]
- modeling auditory waveform with kernels [54]
- brain is actually a dynamic system and need to model the manifold [6] because the same brain region does multi-

ple things at the same time with the manifold lol [17]

- ?time constant of auditory sensitivity in STG neurons?
- The natural analog of the philosophical problem of universals in the conditioning paradigm is stimulus generalization [49]
- Neural nets for estimating nonlinear STRFs, see [27]

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

3.3 imaging

3.4 analysis & modeling

4. Specific Aims

5. Significance & Broader Impacts

6. Notes

6.1 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.[2]

7. meta

7.1 to-read

- revisit the tversky lit and check Danielle's cites for more
- the long-term imaging/ephys papes
- [40]
- [13]
- [46]
- [48]
- [51]
- [17]
- [41]
- [24]
- [4]
- [20]
- [45]
- [62]
- [37]
- [3]
- [23]
- [18] - methods
- [26] - methods

- [1] - methods

7.2 bookmarks

- [12] - p6

8. References

References

- [1] 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](https://doi.org/10.1101/2020.10.19.346353). URL: <http://biorxiv.org/lookup/doi/10.1101/2020.10.19.346353> (visited on 01/09/2021).
- [2] 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](https://doi.org/10.1037/0096-1523.6.3.536). PMID: 6447767.
- [3] Federico Battiston et al. *Networks beyond Pairwise Interactions: Structure and Dynamics*. June 2, 2020. arXiv: [2006.01764](https://arxiv.org/abs/2006.01764) [cond-mat, physics:nlin, physics:physics, q-bio]. URL: <http://arxiv.org/abs/2006.01764> (visited on 01/09/2021).
- [4] Manuel Beiran et al. *Shaping Dynamics with Multiple Populations in Low-Rank Recurrent Networks*. Nov. 17, 2020. arXiv: [2007.02062](https://arxiv.org/abs/2007.02062) [q-bio]. URL: <http://arxiv.org/abs/2007.02062> (visited on 01/09/2021).
- [5] 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](https://doi.org/10.1523/JNEUROSCI.4755-08.2009). PMID: 19228960.
- [6] 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](https://doi.org/10.1016/j.bbrc.2020.12.011). URL: <http://www.sciencedirect.com/science/article/pii/S0006291X20321872> (visited on 01/17/2021).
- [7] 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](https://doi.org/10.1093/cercor/bht127). URL: <https://doi.org/10.1093/cercor/bht127> (visited on 01/17/2021).
- [8] 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](https://doi.org/10.1038/nn.2641). URL: <https://www.nature.com/articles/nn.2641> (visited on 01/21/2021).
- [9] 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](https://doi.org/10.1121/1.5052025). URL: <https://asa.scitation.org/doi/10.1121/1.5052025> (visited on 01/22/2019).
- [10] 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](https://doi.org/10.1037/a0016748). PMID: 20384398. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2890302/> (visited on 01/12/2021).
- [11] 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](https://doi.org/10.1017/CBO9780511642005).
- [12] 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](https://doi.org/10.1101/2020.07.03.185942). URL: <https://www.biorxiv.org/content/10.1101/2020.07.03.185942v1> (visited on 01/09/2021).
- [13] 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](https://doi.org/10.1016/j.neuron.2010.03.029). URL: <http://www.sciencedirect.com/science/article/pii/S0896627310002321> (visited on 01/09/2021).
- [14] 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.

- [15] Gangyi Feng, Han Gyi Yi, and Bharath Chandrasekaran. “The Role of the Human Auditory Corticostriatal Network in Speech Learning”. In: *Cerebral Cortex* (New York, N.Y.: 1991) (Dec. 7, 2018). ISSN: 1460-2199. DOI: [10.1093/cercor/bhy289](https://doi.org/10.1093/cercor/bhy289). pmid: 30535138.
- [16] *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).
- [17] 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](https://doi.org/10.1038/s41467-018-06560-z). URL: <https://www.nature.com/articles/s41467-018-06560-z> (visited on 01/09/2021).
- [18] 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](https://doi.org/10.1016/j.neuron.2017.05.025). pmid: 28595054.
- [19] 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](https://doi.org/10.1016/j.neuroimage.2017.06.068). URL: <http://www.sciencedirect.com/science/article/pii/S1053811917305396> (visited on 01/14/2021).
- [20] 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).
- [21] 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](https://doi.org/10.3758/APP.72.5.1218). URL: <https://doi.org/10.3758/APP.72.5.1218> (visited on 01/15/2021).
- [22] 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](https://doi.org/10.1121/1.415234). URL: <https://asa.scitation.org/doi/abs/10.1121/1.415234> (visited on 01/20/2021).
- [23] 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](https://doi.org/10.1016/j.neuron.2012.09.037). pmid: 23217744. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3523713/> (visited on 01/09/2021).
- [24] 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](https://doi.org/10.1523/JNEUROSCI.0302-20.2020). pmid: 33097640. URL: <https://www.jneurosci.org/content/40/49/9467> (visited on 01/09/2021).
- [25] Michael Kieffe 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](https://doi.org/10.1121/1.2804951). URL: <https://asa.scitation.org/doi/10.1121/1.2804951> (visited on 01/21/2021).
- [26] 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](https://doi.org/10.1016/j.celrep.2016.12.004). pmid: 28009304. URL: [https://www.cell.com/cell-reports/abstract/S2211-1247\(16\)31676-X](https://www.cell.com/cell-reports/abstract/S2211-1247(16)31676-X) (visited on 01/09/2021).
- [27] 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](https://doi.org/10.12688/f1000research.15580.1). pmid: 30345008.
- [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](https://doi.org/10.3758/s13414-019-01702-x). URL: <https://doi.org/10.3758/s13414-019-01702-x> (visited on 07/27/2019).
- [29] 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](https://doi.org/10.1016/0022-2496(75)90047-4). URL: <https://linkinghub.elsevier.com/retrieve/pii/0022249675900474> (visited on 02/28/2019).

- [30] 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](https://doi.org/10.3758/s13423-016-1049-y). URL: <http://link.springer.com/10.3758/s13423-016-1049-y> (visited on 01/20/2017).
- [31] 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](https://doi.org/10.1098/rstb.2007.2154). URL: <https://royalsocietypublishing.org/doi/full/10.1098/rstb.2007.2154> (visited on 01/15/2021).
- [32] 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](https://doi.org/10.1073/pnas.97.22.11850). PMID: 11050219. URL: <https://www.pnas.org/content/97/22/11850> (visited on 07/28/2019).
- [33] 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](https://doi.org/10.1016/j.neuron.2010.08.038). URL: <http://www.sciencedirect.com/science/article/pii/S0896627310006811> (visited on 01/15/2021).
- [34] 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](https://doi.org/10.1038/nrn1533). URL: <https://www.nature.com/articles/nrn1533> (visited on 01/15/2021).
- [35] 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](https://doi.org/10.3819/ccbr.2008.30007). URL: http://comparative-cognition-and-behavior-reviews.org/2008/vol3_lea_wills/ (visited on 01/13/2021).
- [36] 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](https://doi.org/10.1038/s41467-019-10690-3). URL: <https://www.nature.com/articles/s41467-019-10690-3> (visited on 01/06/2021).
- [37] 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](https://doi.org/10.1038/nn.4158). URL: <https://www.nature.com/articles/nn.4158> (visited on 01/09/2021).
- [38] 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](https://doi.org/10.3389/fnins.2014.00230). URL: <https://www.frontiersin.org/articles/10.3389/fnins.2014.00230/full> (visited on 01/16/2021).
- [39] Leigh Lisker. “Rapid versus Rabad: 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](https://doi.org/10.1121/1.2016377).
- [40] 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](https://doi.org/10.1101/2020.12.29.424235). URL: <https://www.biorxiv.org/content/10.1101/2020.12.29.424235v1> (visited on 01/09/2021).
- [41] 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](https://doi.org/10.1016/j.neuron.2018.07.003). URL: <http://www.sciencedirect.com/science/article/pii/S0896627318305439> (visited on 01/09/2021).
- [42] Adam M. P. Miller, William Mau, and David M. Smith. “Retrosplenial Cortical Representations of Space and Future Goal Locations Develop with Learning”. In: *Current Biology* 29.12 (June 17, 2019), 2083–2090.e4. ISSN: 0960-9822. DOI: [10.1016/j.cub.2019.05.034](https://doi.org/10.1016/j.cub.2019.05.034). URL: <http://www.sciencedirect.com/science/article/pii/S0960982219306037> (visited on 01/09/2021).
- [43] 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](https://doi.org/10.1101/2020.09.30.321687). URL: <https://www.biorxiv.org/content/10.1101/2020.09.30.321687v1> (visited on 01/09/2021).
- [44] 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](https://doi.org/10.1038/s41592-018-0109-9). URL: <https://www.nature.com/articles/s41592-018-0109-9> (visited on 01/17/2021).

- [45] 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](https://doi.org/10.1016/j.tins.2020.05.005). PMID: [32580899](https://pubmed.ncbi.nlm.nih.gov/32580899/). URL: [https://www.cell.com/trends/neurosciences/abstract/S0166-2236\(20\)30123-5](https://www.cell.com/trends/neurosciences/abstract/S0166-2236(20)30123-5) (visited on 01/09/2021).
- [46] 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](https://doi.org/10.1016/j.neuron.2018.09.030). PMID: [30344047](https://pubmed.ncbi.nlm.nih.gov/30344047/). URL: [https://www.cell.com/neuron/abstract/S0896-6273\(18\)30832-8](https://www.cell.com/neuron/abstract/S0896-6273(18)30832-8) (visited on 01/09/2021).
- [47] 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](https://doi.org/10.1523/JNEUROSCI.3771-05.2006). PMID: [16672673](https://pubmed.ncbi.nlm.nih.gov/16672673/). URL: <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3771-05.2006>.
- [48] Fernando E. Rosas et al. *Causal Blankets: Theory and Algorithmic Framework*. Sept. 29, 2020. arXiv: [2008.12568](https://arxiv.org/abs/2008.12568) [nlin, q-bio]. URL: <http://arxiv.org/abs/2008.12568> (visited on 01/09/2021).
- [49] 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](https://doi.org/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).
- [50] 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](https://doi.org/10.1016/0010-0285(75)90024-9). URL: <http://www.sciencedirect.com/science/article/pii/0010028575900249> (visited on 01/12/2021).
- [51] 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](https://doi.org/10.1101/2020.11.19.389999). URL: <http://biorxiv.org/lookup/doi/10.1101/2020.11.19.389999> (visited on 01/09/2021).
- [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](https://doi.org/10.1016/j.heares.2019.02.002). URL: <http://www.sciencedirect.com/science/article/pii/S0378595518304441> (visited on 08/20/2019).
- [53] Friedrich Schuessler et al. *The Interplay between Randomness and Structure during Learning in RNNs*. Oct. 25, 2020. arXiv: [2006.11036](https://arxiv.org/abs/2006.11036) [q-bio]. URL: <http://arxiv.org/abs/2006.11036> (visited on 01/09/2021).
- [54] 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](https://doi.org/10.1038/nature04485). PMID: [16495999](https://pubmed.ncbi.nlm.nih.gov/16495999/).
- [55] 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](https://doi.org/10.1073/pnas.1009020107). URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.1009020107> (visited on 11/09/2018).
- [56] 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](https://doi.org/10.1371/journal.pone.0030845). URL: <https://dx.plos.org/10.1371/journal.pone.0030845> (visited on 11/09/2018).
- [57] A. Tversky and Itamar Gati. *Studies of Similarity*. undefined. 1978. URL: [/paper/Studies-of-similarity-Tversky-Gati/93ad5669cc4d8300de0d5d1f9e2c0ed2479d9596](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=93ad5669cc4d8300de0d5d1f9e2c0ed2479d9596) (visited on 01/12/2021).
- [58] Amos Tversky. “Features of Similarity”. In: 84.4 (1977).
- [59] Amos Tversky and David H. Krantz. “The Dimensional Representation and the Metric Structure of Similarity Data”. In: *Journal of Mathematical Psychology* 7.3 (Oct. 1970), pp. 572–596. ISSN: 0022-2496. DOI: [10.1016/0022-2496\(70\)90041-6](https://doi.org/10.1016/0022-2496(70)90041-6). URL: <http://linkinghub.elsevier.com/retrieve/pii/0022249670900416> (visited on 11/22/2017).
- [60] 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](https://doi.org/10.1523/JNEUROSCI.2074-11.2011). PMID: [21994373](https://pubmed.ncbi.nlm.nih.gov/21994373/).

- [61] 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](https://doi.org/10.1016/j.heares.2007.01.019). URL: <http://www.sciencedirect.com/science/article/pii/S0378595507000366> (visited on 01/20/2021).
- [62] 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](https://doi.org/10.1101/2020.06.02.127860). URL: <https://www.biorxiv.org/content/10.1101/2020.06.02.127860v1> (visited on 01/09/2021).
- [63] Ludwig Wittgenstein. *Philosophical Investigations*. Oxford: Basil Blackwell, 1968. 250 pp. ISBN: 978-0-631-11900-5.
- [64] Han Gyo 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](https://doi.org/10.1016/j.neuron.2019.04.023). pmid: [31220442](https://pubmed.ncbi.nlm.nih.gov/31220442/). URL: [https://www.cell.com/neuron/abstract/S0896-6273\(19\)30380-0](https://www.cell.com/neuron/abstract/S0896-6273(19)30380-0) (visited on 07/28/2019).