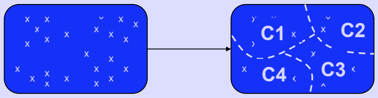
Psych 215L:
Language Acquisition

Lecture 5
Phonemes & Phonology

Speech Perception: Computational Problem

Divide sounds into contrastive categories



Dillon, Dunbar, & Idsardi 2012

Phonetic vs. phonemic categories & their relationship to phonology

"…the general approach to phonological category formation as perceptually driven statistical inference has led to the view that the categorization acquired by the learner is in some sense isomorphic to all and only the distinctions present in the acoustics."

"…it is common for phonological theories to distinguish between phones and phonemes. Phonemes are language-specific, abstract categories used for the purposes of memory encoding in the lexicon. A single phoneme, however, may comprise a set of distinct pronunciations (or phones) that reflect its phonological environment…"

Dillon, Dunbar, & Idsardi 2012

Example of the phone/phoneme distinction

/t/ = phoneme
[t] = phone/phonetic category, as in "sit" (unreleased "t" word-finally)
[tʰ] = phone/phonetic category, as in "top" (aspirated "t" before vowel)
etc.

Example phonological rule

Add morphology –ing: sit + -ing = sitting
Pronunciation: /sɪt/ + /ɪŋ/ = /sɪɾɪŋ/

Rule: /t/ ➡ /ɾ/ / stressed vowel __ unstressed vowel

Dillon, Dunbar, & Idsardi 2012

Categories relevant for word-learning & word manipulation

= not phonetic categories, but phonemic categories

"…if the goal of phonological acquisition is to discover the categories used in lexical storage, then phonetic categories are not the desired end state of phonological acquisition…"

Dillon, Dunbar, & Idsardi 2012

Two-stage acquisition process

"…disconnect between the statistically-induced phonetic categories and the phonemic categories that are the target of acquisition has led to an implicit two-stage view of phonological learning. That is, learners first learn phones using statistical interference over acoustic input, and then build phonemes and phonological systems by identifying relations between these phonetic categories…a second stage of acquisition that subsequently builds the relevant phonemic categories from the phonetic categories…"

One-stage could be better (phones + phonemes simultaneously)

"…the need for a close relationship between phonetic and phonological learning has been noted by a number of researchers investigating the acquisition of phonological systems…"

Dillon, Dunbar, & Idsardi 2012

If we have a two-stage acquisition process, we need to figure out how the second stage works

"…it is necessary to develop an explicit theory of how to group phones into phonemes…Ideas about this procedure are implicit in much of theoretical linguistics…Peperkamp, Le Calvez, Nadal & Dupoux (2006) have proposed solving this problem by comparing the sequence-level distributions of pairs of phones: that is, for each pair of phones *p1, p2,* they examined the probability distribution over phones adjacent to *p1* as versus *p2.* They proposed that phones with the most dissimilar context distributions are more likely to be variants of the same phoneme, with the probability distributions reflecting a generalization of the traditional notion of complementary distribution…"

Dillon, Dunbar, & Idsardi 2012

Complementary distribution

Ex: [t] vs. $[t^h]$

p1 contexts: c1, c2, c3, c4, c5, c6, c7
Ex: "sit", "spot", "lift", "pat", "crest", "loot", "but",

p2 contexts: c8, c9
Ex: "stop", "step"

p1 and p2 would be treated as allophones of the same phoneme because they're phonetically similar, but don't seem to appear in the same places in words. (Therefore, they're not important for distinguishing word meaning, but do affect pronunciation.)

Dillon, Dunbar, & Idsardi 2012

The second stage may not be so easy if the first stage isn't perfect.

"…the success of a two-stage approach to phonological learning crucially depends on the accuracy achieved in the first stage. Errors made in the phone acquisition stage could in principle impair the ability of a second-stage mechanism to extract the correct phonology…"

So maybe we should really look at this one-stage combination idea

"…we explore the feasibility of a single-stage approach to phonological categorization…all theories of phonological acquisition must address this mapping from acoustics to discrete categories. Because of this fact, asserting the feasibility of a single-stage approach amounts to asserting the possibility of folding the acquisition of processes and phoneme level categories into the initial mapping from acoustics to linguistic categories."

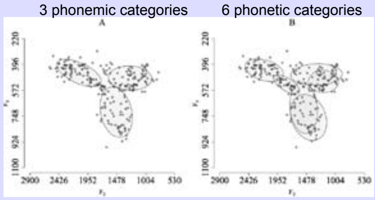
Dillon, Dunbar, & Idsardi 2012

Case study: Inuktitut vowels (3 phonemes)

Allophonic variation, based on presence of following uvular consonant

"Figure 1. Plots of Inuktitut vowels, both grouping (panel A) and splitting (panel B) predictable allophones, in F2 × F1 (backness by height) space. The ellipses mark a 66% confidence region for Gaussians estimated by maximum likelihood on the points from the indicated category."

3 phonemic categories 6 phonetic categories



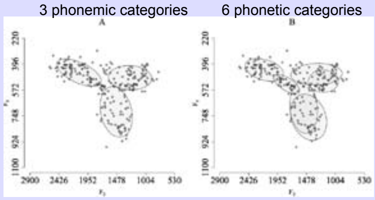
Dillon, Dunbar, & Idsardi 2012

Two-stage goals

(1) Recover phonetic categories correctly

(2) Map correct phonetic categories to phonemic categories

3 phonemic categories 6 phonetic categories



Dillon, Dunbar, & Idsardi 2012

Data set: 239 vowel tokens elicited from native Inuktitut speaker

Model input set: 1000-12000 data points sampled from data set

Modeling:

- Bayesian estimator: point estimate taken from sample of posterior distribution
- input modeled as a mixture of Gaussian distributions, given a Dirichlet Process prior ("infinite mixture of Gaussians")

"This represents a particular way of stating formally that the hypothesis space is all possible Gaussian mixture models, including models with different numbers of categories, along with a particular way of weighting different mixture models (a Dirichlet process in this context is essentially a certain prior probability distribution over mixture models)…because the learner by hypothesis needs to estimate the number of categories justified by the data…**"**

Dillon, Dunbar, & Idsardi 2012

Implementing the learning (Bayesian inference): Details

"A posterior sample from a Dirichlet process mixture of Gaussians was drawn using a Gibbs sampler with component parameters drawn from a Normal-Inverse-Wishart distribution with fixed inverse scale matrix and degrees of freedom parameter, and with location parameter M and inverse scale parameter ω; M was itself sampled from a normal distribution centered at zero, and ω from an inverse Gamma distribution; α was sampled from a Gamma distribution. (See Escobar & West, 1994; West, 1995; and Neal, 2000, for the basic details of the algorithm). To fit each model, a sample of 500 points was drawn from the Gibbs sampler at a lag of 10 after 1200 burn-in samples. The sample with the highest joint posterior density was used as a point estimate. Hyperparameters were tuned to ensure that they were appropriate to find between one and seven categories on the raw data from which the training corpus was sampled.*"*

Dillon, Dunbar, & Idsardi 2012

Assessing model performance: Ideal benchmark

"…we first constructed ideal sets of Gaussian phonetic and phonemic categories using the maximum likelihood estimators for each phoneme (sample means and sample covariances), for each different data set used to train the model…. Using these Gaussians as category models, we classified the data sets from which the Gaussians were constructed using a Bayes-optimal decision rule, labelling a point according to the mixture component with the highest posterior probability given that point."

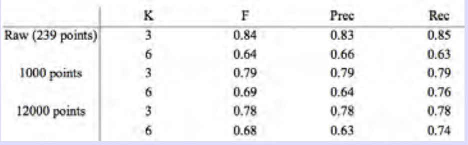
Pairwise comparison metrics

"*Pairwise* refers to the fact that the statistics are constructed by examining every pair of data points and asking whether the two are in the same class (according to either the fitted model or the ideal model). Pairwise statistics are used in clustering evaluation to avoid the issue of constructing a mapping between the model's categories and the true categories; they are still meaningful even if the model finds the wrong number of categories."

Dillon, Dunbar, & Idsardi 2012

Assessing model performance: From idealized categories

"…These values represent the highest possible pairwise *F-scores*…for comparisons between the ideal models' predictions and the data. Two different versions of the true classification are evaluated with this baseline: a three-category phonemic solution (phoneme labels, *K* = 3) and a six-category phonetic solution (phoneme labels plus an indicator for a following uvular, *K* = 6)."



	K	F	Prec	Rec
Raw (239 points)	3	0.84	0.83	0.85
	6	0.64	0.66	0.63
1000 points	3	0.79	0.79	0.79
	6	0.69	0.64	0.76
12000 points	3	0.78	0.78	0.78
	6	0.68	0.63	0.74

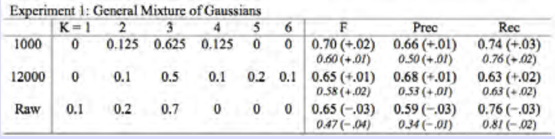
Point: Real life data don't precisely correspond to the assumption that data are generated by a mixture of Gaussian categories, so this shows the best this kind of assumption can do with these data.

Dillon, Dunbar, & Idsardi 2012

Assessing model performance

10-fold cross validation from model

"Left of table shows distribution over number of resulting categories, and right of table shows pairwise scores at test... In parentheses is the difference from scores on training data. Comparisons to both 3- and 6-category (italicized) classifications are shown..."



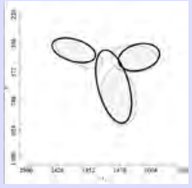
Experiment 1: General Mixture of Gaussians

	K = 1	2	3	4	5	6	F	Prec	Rec
1000	0	0.125	0.625	0.125	0	0	0.70 (+.02) *0.60 (+.01)*	0.66 (+.01) *0.50 (+.01)*	0.74 (+.03) *0.76 (+.02)*
12000	0	0.1	0.5	0.1	0.2	0.1	0.65 (+.01) *0.58 (+.02)*	0.68 (+.01) *0.53 (+.01)*	0.63 (+.02) *0.63 (+.02)*
Raw	0.1	0.2	0.7	0	0	0	0.65 (−.03) *0.47 (−.04)*	0.59 (−.03) *0.34 (−.01)*	0.76 (−.03) *0.81 (−.02)*

Dillon, Dunbar, & Idsardi 2012

Assessing model performance

"…the MOG model is capable of finding three-category solutions which are not unlike the phonemes of Inuktitut; this is seen in the classification scores for the 1000-point models: the *F* scores are reasonably close to the *F* scores for the ideal models (compare Table 1), and are reasonably well-balanced between precision and recall…"



sample 3 categories found (phonemic categories)

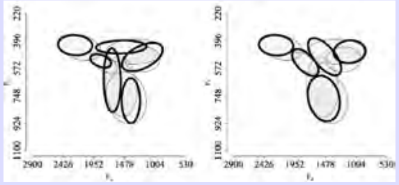
pretty good match to ideal three categories

Dillon, Dunbar, & Idsardi 2012

Assessing model performance

"More fine-grained phonetic solutions become apparent as the number of data points increases…the likelihood term, all other things being equal, prefers larger numbers of categories (the mixture model with the highest possible likelihood would generally be obtained with as many categories as data points)."

Sample 6 and 5 categories found (phonetic categories…or are they?)



Dillon, Dunbar, & Idsardi 2012

Learner biases play an important role

"The results of Experiment 1 suggest that a learner assuming Gaussian categories (a simplifying assumption shared with previous research) could come to a phonemic analysis of Inuktitut vowels with the appropriate biases, but also that an analysis with phone-like categories could also be found with the appropriate bias. The role of bias is important here, as these solutions are not "in the data": the learning outcome depends on the specific bias implied by the model and its hyperparameter settings, in conjunction with the amount of data the model is given. In particular, as the number of data points increases, models tend to prefer a greater number of mixture components."

Dillon, Dunbar, & Idsardi 2012

About those three-category solutions that looked pretty good…

"…such a model would require a second stage wherein learners rediscover the systematic relationships between particular contexts and the pronunciations of these categories. Importantly, the systematic relationship between a phoneme and its retracted allophone in Inuktitut forms an active piece of knowledge that speakers must acquire: even in novel words, speakers adapt the pronunication of the phoneme to its phonological environments."

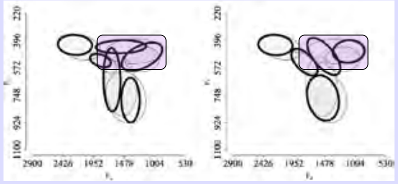
Example of pronunciation knowledge that would have to be rediscovered

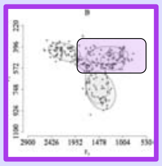
/titirauti/ is pronounced [titeraute] in contexts where the next consonant is uvular [q].

Dillon, Dunbar, & Idsardi 2012

And about those five-category and six-category solutions…

"…Visual inspection of the resulting five- and six-category models suggests that these models would not provide adequate input to a second stage of learning based on a complementary distribution test...it should be the case that the acquired categories align with the phonetic categories of the target system…"

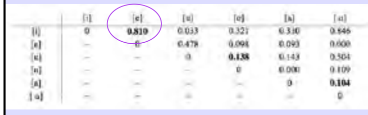




Dillon, Dunbar, & Idsardi 2012

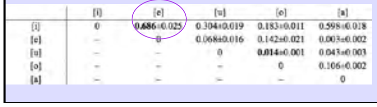
And about those five-category and six-category solutions…

"…the SKLD for [i]-[e] is consistently among the highest values found, suggesting that complementarity-based metrics for discovering phonemic identity could readily recover the relation between these two phones given this MOG…"



	[i]	[e]	[u]	[o]	[a]	[ɑ]
[i]	0	**0.810**	0.033	0.321	0.330	0.646
[e]	–	0	0.478	0.098	0.093	0.600
[u]	–	–	0	**0.138**	0.143	0.504
[o]	–	–	–	0	0.000	0.109
[a]	–	–	–	–	0	**0.104**
[ɑ]	–	–	–	–	–	0

6-category SKL divergence



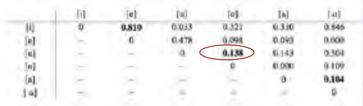
	[i]	[e]	[u]	[o]	[a]
[i]	0	**0.686**±0.025	0.304±0.019	0.183±0.011	0.598±0.018
[e]	–	0	0.068±0.016	0.142±0.021	0.003±0.002
[u]	–	–	0	**0.014**±0.001	0.043±0.003
[o]	–	–	–	0	0.106±0.002
[a]	–	–	–	–	0

5-category SKL divergence

Dillon, Dunbar, & Idsardi 2012

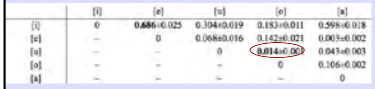
And about those five-category and six-category solutions…

"…However, [o]-[u] consistently had some of the lowest SKLD scores. This is consistent with the visual observation that the models did not correctly identify [o]-[u], instead splitting the /u/ phoneme in an inappropriate way…unlikely that the five- or six- category MOG solutions found for this data could provide input to a second stage of learning…"



	[i]	[e]	[u]	[o]	[a]	[ɑ]
[i]	0	**0.810**	0.033	0.321	0.330	0.646
[e]	–	0	0.478	0.098	0.093	0.600
[u]	–	–	0	**0.138**	0.143	0.504
[o]	–	–	–	0	0.000	0.109
[a]	–	–	–	–	0	**0.104**
[ɑ]	–	–	–	–	–	0

6-category SKL divergence



	[i]	[e]	[u]	[o]	[a]
[i]	0	**0.686**±0.025	0.304±0.019	0.183±0.011	0.598±0.018
[e]	–	0	0.068±0.016	0.142±0.021	0.003±0.002
[u]	–	–	0	**0.014**±0.001	0.043±0.003
[o]	–	–	–	0	0.106±0.002
[a]	–	–	–	–	0

5-category SKL divergence

Dillon, Dunbar, & Idsardi 2012

So let's look at that single-stage learning model

"…we develop a model that takes a substantially different approach to solving the same problem by factoring out predictable acoustic variation that arises due to the grammatical rules of the language in the acoustic space, rather than waiting to discover them based on strings of discrete categories…"

Dillon, Dunbar, & Idsardi 2012

So let's look at that single-stage learning model

Basic idea: Check utility of factoring out this info

"…we manually remove the phonetic effect due to following uvulars from all vowel tokens occurring in that context. We then train a MOG model on the resulting transformed data to demonstrate the usefulness of factoring out such transforms…"

Psychological plausibility: How could children learn to factor out this info

"…we take up the question of how these transforms are acquired…"

Big picture idea: Learning two levels at once…sort of

"…The phonetic category model that results from this procedure is one in which finding phones becomes irrelevant, because the uvular retraction rule has already been handled at the phonetic level…"

Dillon, Dunbar, & Idsardi 2012

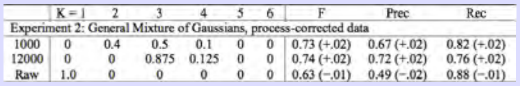
Utility check: Factoring out the acoustic effect of uvulars

"…The mean F1, F2 value for all the points which occurred before a uvular was computed (F+u); the mean F1, F2 value for all the points which did not occur before a uvular was computed (F-u); and the points which occurred before a uvular were corrected for the effect of the following uvular consonant by subtracting (F+u − F-u) from the formant value. This correction was calculated once for all three vowel phoneme categories, so that all pre-uvular points had the same vector subtracted, regardless of whether they were /i/, /a/, or /u/ tokens…"

Dillon, Dunbar, & Idsardi 2012

Utility check: Factoring out the acoustic effect of uvulars

"…It can be seen that across both small and large training sets, 3-phoneme solutions are the most common solution reached by the model. The distribution of the phonemes in a three-category solution, as in Experiment 1 (see Fig. 2), line up closely with the target phonemic categories…this model approximates a listener that can make use of the context in which a segment occurred to adapt its acoustic models (as humans do: see, for example, Nearey 1990; Whalen, Best & Irwin, 1997), thus making some regions of uncertainty less ambiguous, and making better phonemic category models available to the learner…"



	K = 1	2	3	4	5	6	F	Prec	Rec
Experiment 2: General Mixture of Gaussians, process-corrected data									
1000	0	0.4	0.5	0.1	0	0	0.73 (+.02)	0.67 (+.02)	0.82 (+.02)
12000	0	0	0.875	0.125	0	0	0.74 (+.02)	0.72 (+.02)	0.76 (+.02)
Raw	1.0	0	0	0	0	0	0.63 (−.01)	0.49 (−.02)	0.88 (−.01)

Dillon, Dunbar, & Idsardi 2012

Psychological plausibility check: Learning the uvular rule

Learner knows to look for rules relating phonetic categories

"…we model the learner as searching for a set of sets of subcategories, where the subcategories within a set are related by some rule. In other words, each phoneme is defined by a set (in this case a pair) of Gaussians, one for the pre-uvular realizations of that phoneme, and another for realizations of that phoneme in other contexts…"

Dillon, Dunbar, & Idsardi 2012

Psychological plausibility check: Learning the uvular rule

Learner knows to look for rules relating phonetic (sub-)categories

"…we model the learner as searching for a set of sets of subcategories, where the subcategories within a set are related by some rule. In other words, each phoneme is defined by a set (in this case a pair) of Gaussians, one for the pre-uvular realizations of that phoneme, and another for realizations of that phoneme in other contexts…"

Learner knows phonetic (sub-)categories must appear in complementary distribution

"An additional constraint we impose in our model is that the data points which are attributed to the two Gaussians need to be in complementary distribution: one Gaussian models the points appearing in a conditioning environment, and the other models the points appearing elsewhere…"

Dillon, Dunbar, & Idsardi 2012

Psychological plausibility check: Learning the uvular rule

Learner knows that a phonetic (sub-)category should be homogeneous

"…we add an additional constraint of homogeneity of variance. This means that, for the allophonic sub-clusters making up each phoneme, the covariance matrix of the Gaussian (which defines its size, shape, and orientation) must be the same. This should be familiar because it is exactly the constraint that defines a linear model in statistics…The learner must construct a set of categories, each of which is a linear model predicting the phonetic values for the set of segments being categorized (in this case, vowels) from this discrete indicator variable. Because it is a linear model, the learner therefore finds, for each category, an intercept (overall category mean F1 and F2), and an effect of conditioning environment (a shift in phonetic space), in addition to estimating variance. In this way, the model can thus be said to simultaneously discover a set of phoneme categories and a set of associated phonetic rules…"

Dillon, Dunbar, & Idsardi 2012

Psychological plausibility check: Learning the uvular rule

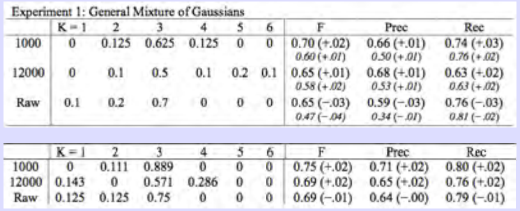
Input data: Learner heeds uvular consonants in particular

"…we annotated the data points with a vector of indicator variables marking the presence or absence of a following uvular consonant…"

Dillon, Dunbar, & Idsardi 2012

Psychological plausibility check: Learning the uvular rule

"…overall, phoneme classification performance is better than in Experiment 1. In particular, when the classification scores for either of the datasets are compared to the corresponding classification scores from Experiment 1, they are seen to be higher…"



Experiment 1: General Mixture of Gaussians

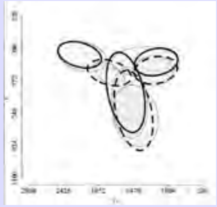
	K = 1	2	3	4	5	6	F	Prec	Rec
1000	0	0.125	0.625	0.125	0	0	0.70 (+.02) *0.60 (+.01)*	0.66 (+.01) *0.50 (+.01)*	0.74 (+.03) *0.76 (+.02)*
12000	0	0.1	0.5	0.1	0.2	0.1	0.65 (+.01) *0.58 (+.02)*	0.68 (+.01) *0.53 (+.01)*	0.63 (+.02) *0.63 (+.02)*
Raw	0.1	0.2	0.7	0	0	0	0.65 (−.03) *0.47 (−.04)*	0.59 (−.03) *0.34 (−.01)*	0.76 (−.03) *0.81 (−.02)*

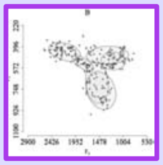
	K = 1	2	3	4	5	6	F	Prec	Rec
1000	0	0.111	0.889	0	0	0	0.75 (+.02)	0.71 (+.02)	0.80 (+.02)
12000	0.143	0	0.571	0.286	0	0	0.69 (+.02)	0.65 (+.02)	0.76 (+.02)
Raw	0.125	0.125	0.75	0	0	0	0.69 (−.01)	0.64 (−.00)	0.79 (−.01)

Dillon, Dunbar, & Idsardi 2012

Psychological plausibility check: Learning the uvular rule

"…when the model does find three categories, its classification performance is better than that of the three- category MOG models… appears to better approximate the true structure of the phonemic categories, rather than idiosyncracies of the training data…"





Dillon, Dunbar, & Idsardi 2012

Big picture contribution

"…regressing out predictable effects of phonological context can improve classification (Nearey 1990), as well as providing greater separation of acoustic clusters (Cole et al, 2010). The models we described here extend this research by examining the impact of these techniques for the problem of language acquisition…"

Impacts for infants

"…the single-stage approach returns a much more deployable set of phonological knowledge: a set of phonemes, and the processes that relate them to their allophones…Together with the phoneme categories, this knowledge gives the language user all the knowledge necessary to produce an appropriate vowel token given a phonological environment."

Dillon, Dunbar, & Idsardi 2012

Open question: Knowing about the effect of uvulars

"…One important issue for the current model concerns the discovery of potential conditioning environments. Although much of the model operated in an unsupervised fashion, the model did not need to determine which tokens were in a uvular context…a more complete model could possibly incorporate predictors for all possible conditioning environments…"

Open question: knowing to condition on consonant phonemes

"…If the learner has access to some feature parse of the consonants [i.e., uvular] before they have identified the consonant categories in her language, then this information could potentially serve as predictors or conditioning environments in a mixture of linear models…"