A single-stage computational model of phoneme category acquisition

Ewan Dunbar[1], Brian Dillon[1], and William Idsardi[1,2]
[1]Department of Linguistics, [2]Neuroscience and Cognitive Science Program, University of Maryland

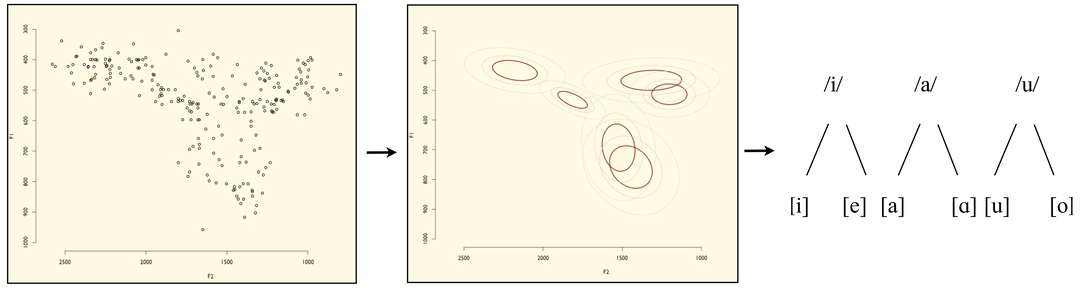
Phonological Acquisition

- Consensus in the theoretical and experimental literature implies **two-stage** category learning, in which a stage of **phone** learning is distinguished from a stage of **phoneme** learning
- Could we construct a **single-stage model** of **simultaneous** process and category information which **eliminates the phone level**?

The Two-Stage Model

Learning of **categories** and **processes** are usually treated as separate procedures. First, perform **phonetic-level** categorization to get **phones** [1,2]. Second, identify **processes** relating phones to build **phonemes** [3].

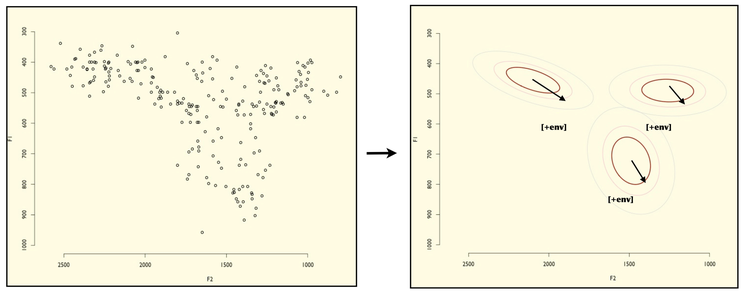
Acoustics → **Phones** → **Phonemes**



The Single-Stage Model

If **allophonic relations** and **phoneme clusters** are represented in the **same (acoustic) space**, they can be estimated jointly in a single-stage acquisition model:

Acoustics → **Phonemes and Processes**

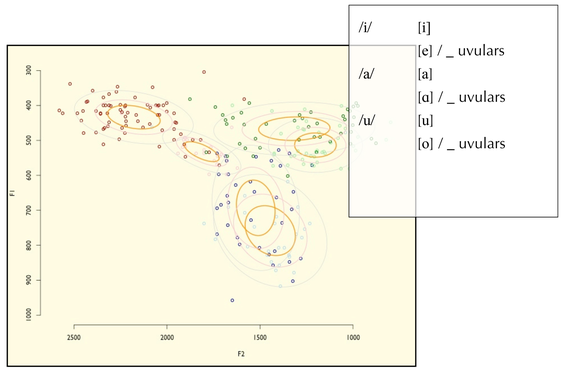


Here processes are estimated by considering average effect of [uvular] context across all vowel tokens of a given phoneme category.

This provides a proof-of-principle that transforms aid the recovery of phonemic categories; future work will focus on joint estimation of processes and categories.

Case Study: Inuktitut

Inuktitut has three vowel phonemes, but six vowel phones:



Clustering Component

Model the Inuktitut vowel space as multiple overlapping Gaussians in F1 x F2 space [1,2].

$$f(x_i|\Psi) = \sum_{k=1}^{K} \pi_k \phi(x_i|\theta_k)$$

$$\Psi = (\pi', \theta')'$$

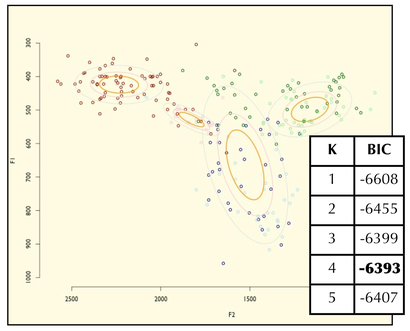
Model assigns a probability to a data set (likelihood); fit by maximizing likelihood.

Current Simulations

Two sets of vowel training data were constructed: **raw acoustic values** (uncorrected <F1,F2> points) versus **transformed acoustic values** (the **average lowering effect of uvulars on each phoneme** was **subtracted** from all tokens of that phoneme in uvular context).

Inuktitut data collected by Derek Denis and Mark Pollard from a female speaker from Cape Dorset (Kinngait). Formant values labeled by hand <F1,F2> for 239 vowel tokens. Simulations conducted in R using the MClust package [4-6]. Mixtures of Gaussians fit using EM, optimized on BIC. Additional penalty of 2K numerical parameters added to transformed BICs.

Results : Raw Data

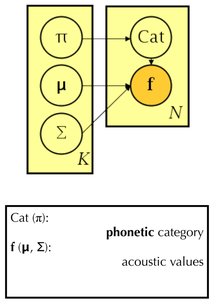


K	BIC
1	-6608
2	-6455
3	-6399
4	**-6393**
5	-6407

Discussion

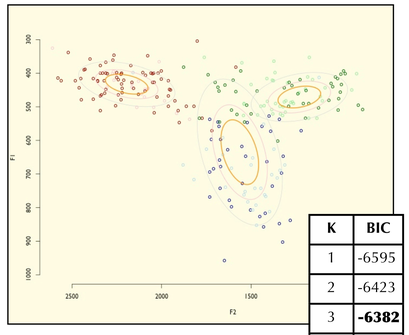
- Raw data leads to a **four vowel solution**
- Clusters are **neither phonetic nor phonemic**
- Second-stage techniques have empirical and logical problems [7]

Model



Cat (π): **phonetic** category
f (μ, Σ): acoustic values

Results : Transformed Data

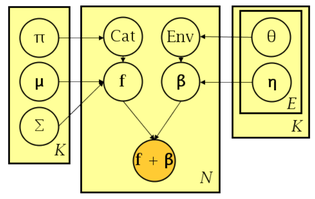


K	BIC
1	-6595
2	-6423
3	**-6382**
4	-6417

Discussion

- Transformed data leads to a **three vowel solution**
- **Phonemic categorization** is achieved in initial clustering
- Allophonic processes are represented as **subsymbolic shifts in acoustic space**
- Can be straightforwardly extended to hierarchical models
- BIC for transformed model is **higher** than for raw model; three-vowel solution is **objectively better** unless the hypothesis space is restricted

Model



Cat (π): **phonemic** category
f (μ, Σ): latent acoustic values
Env (θ): phonemic **environment**
β (η): environment-dependent **acoustic transform**

Conclusions

- A two-stage model **fails to construct Inuktitut phonology**
- Single-stage models acquire **phonemes** by relying on **sub-symbolic allophonic processes**
- Future work will include inference over transforms and compare larger classes of models; objective model comparion metrics like BIC allow us to **directly compare the fit of different models**; in this case, it favors the phonemic single-stage solution

References

[1] Vallabha,et al. 2007. PNAS 104:13273–13278 [2] deBoer and Kuhl.2001. JASA 110(5): 2703 [3] Peperkamp, et al. 2007. Cognition 101:B31-B41; [4] Fraley and Raftery, 2002. J. Am. Stat. Ass. 97:611-631; [5] Fraley andRaftery, 2009. http://CRAN.R-project.org/package=mclust [6] McLachlan, 1987. J. Roy. Stat. Soc. 36(3):318-324 [7] Dunbar, 2009. MOT Phonology Workshop.

Acknowledgments

Thanks to Derek Denis, Mark Pollard, and Alana Johns for Inuktitut data. Thanks to Jordan Boyd-Graber, Elan Dresher, Jeff Lidz, Mark Liberman, and Colin Phillips for discussion of the issues involved. This work was supported by NSF IGERT DGE—0801465 to the University of Maryland.