Similarity-based phonological generalization

Brandon Prickett - bprickett@umass.edu - University of Massachusetts Amherst

Society for Computation in Linguistics 2018

1) Abstraction-based Generalization

- Phonological rules can generalize to novel segments.
  - ⇒ For example, when asked to add the possessive morpheme /-z/ to the name *Bach*, English speakers generalize a voicing assimilation process to the consonant /x/ (this is often called the *Bach Test*; Halle 1978).
- One explanation for this is that speakers represent the process using simple, abstract feature bundles:
  - ⇒ For example: "[+voice] becomes [-voice] after [-voice]"
  - ⇒ I'll call this hypothesis ***abstraction-based*** generalization.
- GMECCS (Moreton et al. 2017) is a maximum entropy (MaxEnt) phonotactic model that uses abstraction-based generalization.
  - ⇒ GMECCS uses a gradient descent learning update, like most MaxEnt models.
  - ⇒ The constraint set is every possible combination of relevant features (and their corresponding values): e.g. *[+voice], *[-voice], *[+voice, +continuant], *[-voice, +continuant], *[-voice, -continuant], etc…
  - ⇒ Crucially, there are simple, abstract constraints such as *[-voice], which cause generalization to novel sounds.

2) Similarity-based Generalization

- An alternative theory is that the generalization seen in phonology is due to "online calculations of similarity" (Cristia et al. 2013:279).
  - ⇒ I'll call this ***similarity-based*** generalization.
- Sim-Gen is a MaxEnt learner that differs from GMECCS in two ways:
  - ⇒ Its constraint set only has single-segment constraints:

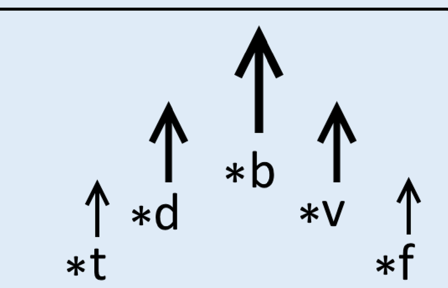
*[+voice, +labial, +continuant]	*v
*[-voice, +labial, +continuant]	*f
*[-voice, -labial, +continuant]	*s
*[+voice, -labial, +continuant]	*z
etc...	

  - ⇒ Each constraint's weight updates *leak* onto featurally similar constraints:

$$\Delta w_i = \eta[E_{empircal}[c_i] - E_{model}[c_i]]$$

$$\delta w_j = \theta\left[\frac{\Delta w_i}{\mathrm{dist}(c_i, c_j)}\right]$$

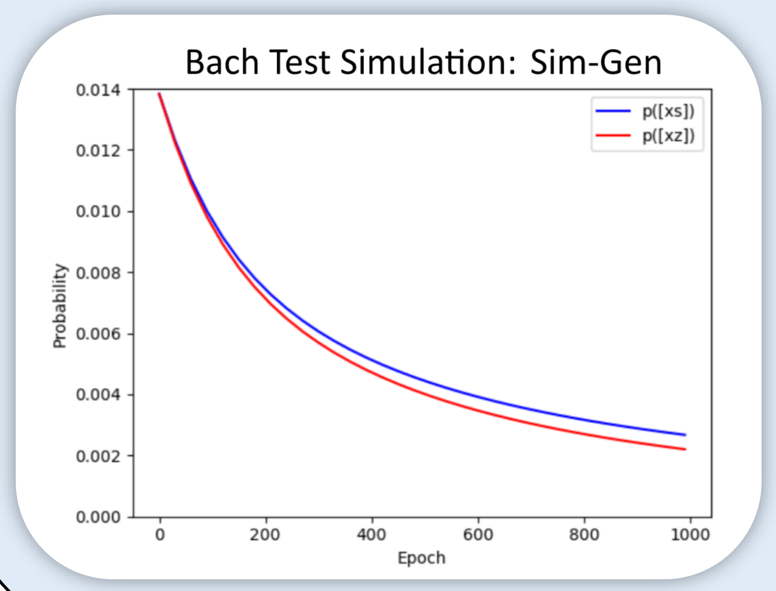
$$dist(c_i, c_j) = |Features\ c_i\ \&\ c_j\ differ\ in|$$



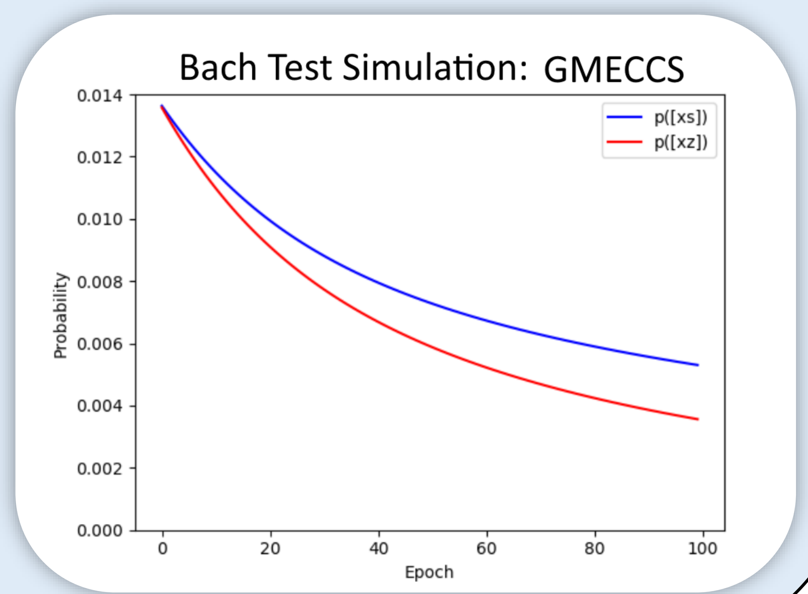
- Because of these differences, Sim-Gen's generalization is similarity-based.

3) Simulating the Bach Test

- To see if similarity-based generalization predicts the kind of behavior observed in the Bach Test, I trained GMECCS and Sim-Gen on an analogous phonotactic pattern.
  - ⇒ Clusters made up of the segments {d, t, z, s, k, and g} all agreed in voicing.
  - ⇒ Both models also had constraints that referred to [x] and [ɣ], though these sounds were unattested in the learning data.

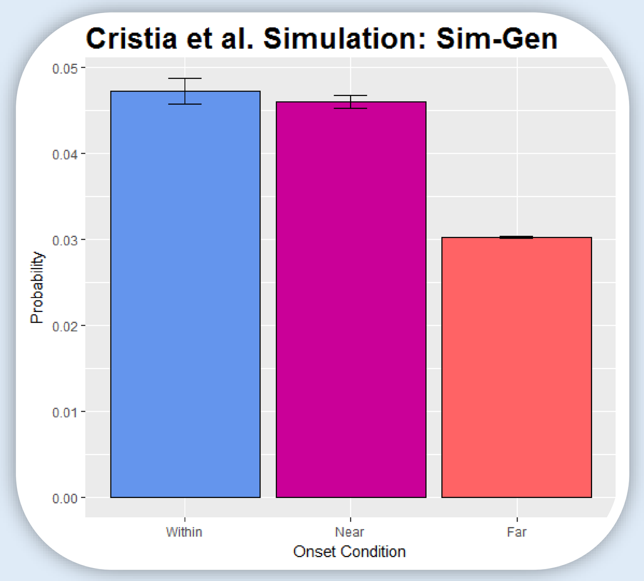


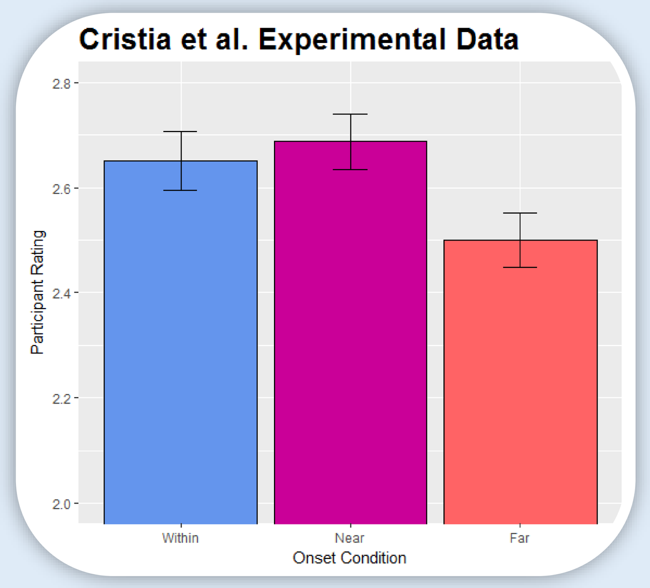
- Both models had an η of 0.05, and Sim-Gen had a θ of 0.5.
- These results show that both models can capture the kind of generalization observed by Halle (1978).
  - ⇒ i.e. both learners eventually assign a higher probability to [xs] than [xz].

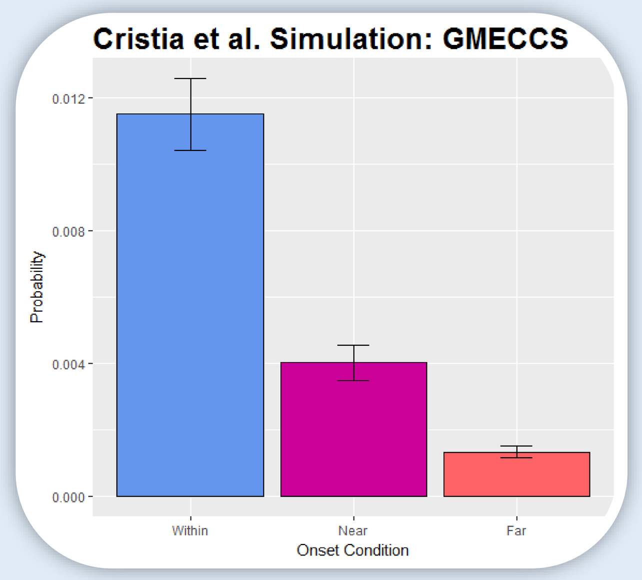


4) Simulating Cristia et al. (2013)

- Cristia et al. (2013) taught participants an onset restriction (e.g. all onsets had to be [+voice]) and withheld three kinds of segments from training:
  - ⇒ ***Within***: within the legal onsets' feature bundle. ***e.g. [b]*** if legal onsets were [+voice]
  - ⇒ ***Near***: phonetically similar to the legal onsets. ***e.g. [k]*** if legal onsets were [+voice]
  - ⇒ ***Far***: phonetically dissimilar to the legal onsets. ***e.g. [p]*** if legal onsets were [+voice]
  - ⇒ I simulated this experiment using GMECCS and Sim-Gen (η=0.05, θ=0.5). The figures below show the average probabilities each learner assigned to the various segment categories, averaged across all experimental conditions (probabilities are taken from each learner's epoch that correlated most with the human data).







  - ⇒ Sim-Gen assigns equal probabilities to the Within and Near conditions (matching the experimental data), while GMECCS does not.

5) Discussion

- While both the similarity-based and abstraction-based learners were able to model the Bach Test, only the similarity-based learner was able to predict the results of Cristia et al. (2013).
- Future work:
  - ⇒ Are there better similarity metrics? So far, I've tried similarity based on natural class membership (Frisch et al. 2004), articulatory phonetics (Mielke 2011), and lexical context (Mikolov et al. 2013) to no avail.
  - ⇒ Sim-Gen captures the same complexity bias effects that Moreton et al (2017) found with GMECCS. Are there any biases that Sim-Gen might predict that other MaxEnt models don't?

References

Cristia, A., J. Mielke, R. Daland, & S. Peperkamp (2013). Similarity in the generalization of implicitly learned sound patterns. *Laboratory Phonology, 4(2)*, 259-285.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory, 22(1)*, 179-228.

Halle, M. (1978). Knowledge unlearned and untaught: what speakers know about the sounds of their language. In: M. Halle, J. Bresnan & G.A. Miller (eds.), *Linguistic theory and psychological reality*. MIT Press, Cambridge, MA, 294-303.

Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua, 122(2)*, 145-163.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Moreton, E., J. Pater, & K. Pertsova (2017). Phonological concept learning. *Cognitive science, 41(1)*, 4-69.