LI 308: Computational Models of Sound Change

James Kirby & Morgan Sonderegger

23 July 2015

Pierrehumbert (2001)

Frequency effects and their relation to lenition (weakening)

- Frequency often correlated with synchronic reduction (Hooper, 1976; Bybee, 2000; Jurafsky et al., 2002; Munson and Solomon, 2004...)
  - *memory* $>$ [mɛmɹi] vs. *mammary* $\not>$ [mæmɹi] (Hooper, 1976)
  - *told* $>$ [tol] vs. *meant* $\not>$ [mɛn] (Hooper, 1976)

- Frequency also correlated with rates of diachronic change (e.g. Philips, 1984, 1986; Bybee, 2001, 2002; but cf. Labov, 1981, 1994)
  - (low-frequency) *nude* */njud/ $>$ [nuːd], not [njuːd], while
  - (high-frequency) *new* */nju/ $>$ [njuː], not [nuː]

Pierrehumbert (2001)

- Goal: an existence proof by constructing ‘a formal architecture which is capable of capturing these regularities’
- Should also predict that ‘some outcomes are possible and others are not’

Pierrehumbert (2001)

‘By examining the consequences of the perception-production loop over time, we provide a formal framework for thinking about the quantitative predictions of usage-based phonology, as proposed by Bybee. We **derive** the finding that leniting historical changes are more advanced in frequent words than in rarer ones. Calculations are presented which reveal the interaction of production noise, lenition and entrenchment. A realistic treatment is also provided for the time course of a phonological merger which originates from lenition of a marked category.’

Pierrehumbert (2001)

‘By examining the consequences of the perception-production loop over time, we provide a formal framework for thinking about the quantitative predictions of usage-based phonology, as proposed by Bybee. We derive the finding that leniting historical changes are more advanced in frequent words than in rarer ones. Calculations are presented which **reveal** the interaction of production noise, lenition and entrenchment. A realistic treatment is also provided for the time course of a phonological merger which originates from lenition of a marked category.’

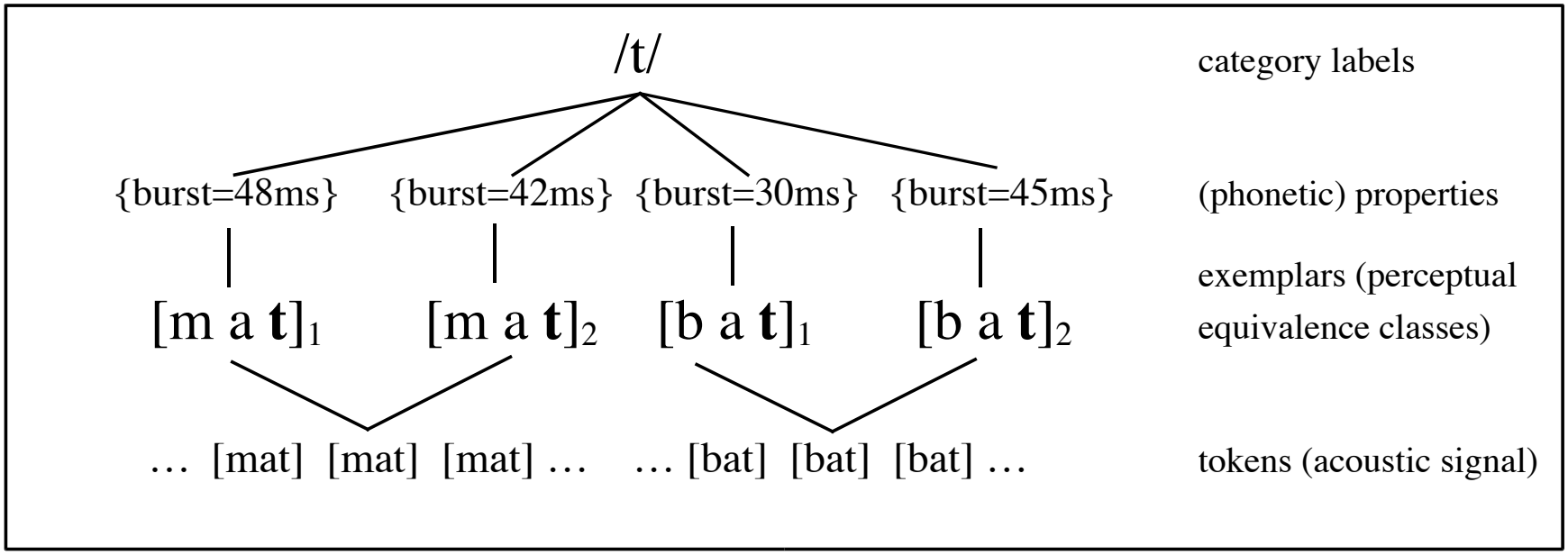
Pierrehumbert (2001)

‘By examining the consequences of the perception-production loop over time, we provide a formal framework for thinking about the quantitative predictions of usage-based phonology, as proposed by Bybee. We derive the finding that leniting historical changes are more advanced in frequent words than in rarer ones. Calculations are presented which reveal the interaction of production noise, lenition and entrenchment. A **realistic treatment** is also provided for the time course of a phonological merger which originates from lenition of a marked category.’

Elements of exemplar ‘theory’

- Memory is episodic, highly detailed
- System is a map from points in multidimensional phonetic space to category labels...
- ...independent of where those labels might come from

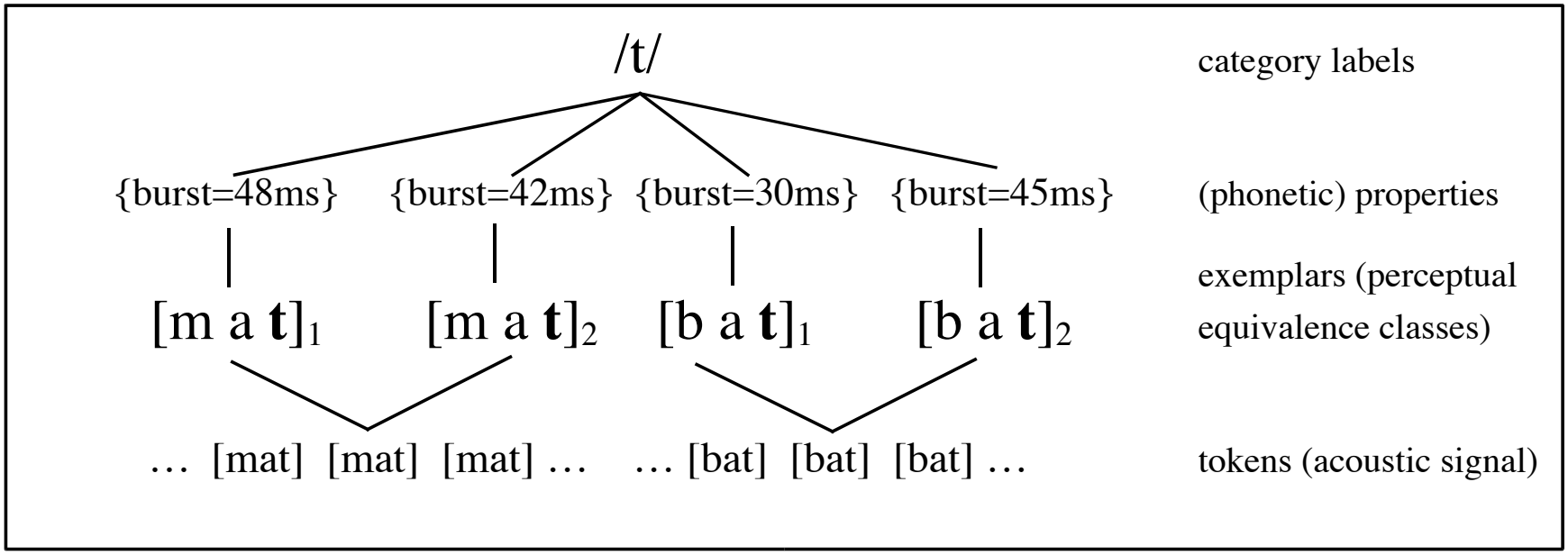
Exemplar storage: Units (knowledge state)



- An exemplar list $E(L)$ is a list of exemplars associated with label L:

$$E(L) = e_1^L, \ldots, e_n^L$$

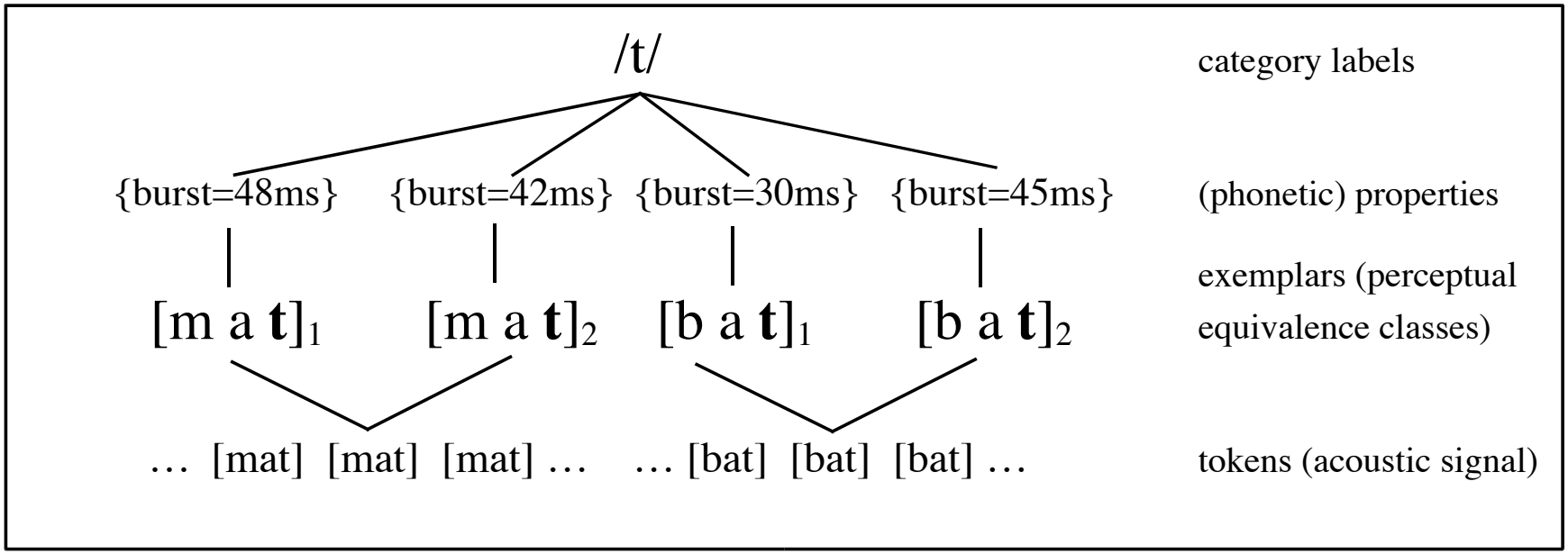
Exemplar storage: Units (knowledge state)



- Individual exemplars represented as (property, label) pairs

$$e_1^L = \{burst, 48\,ms\}, e_2^L = \{burst, 42\,ms\}, \ldots$$

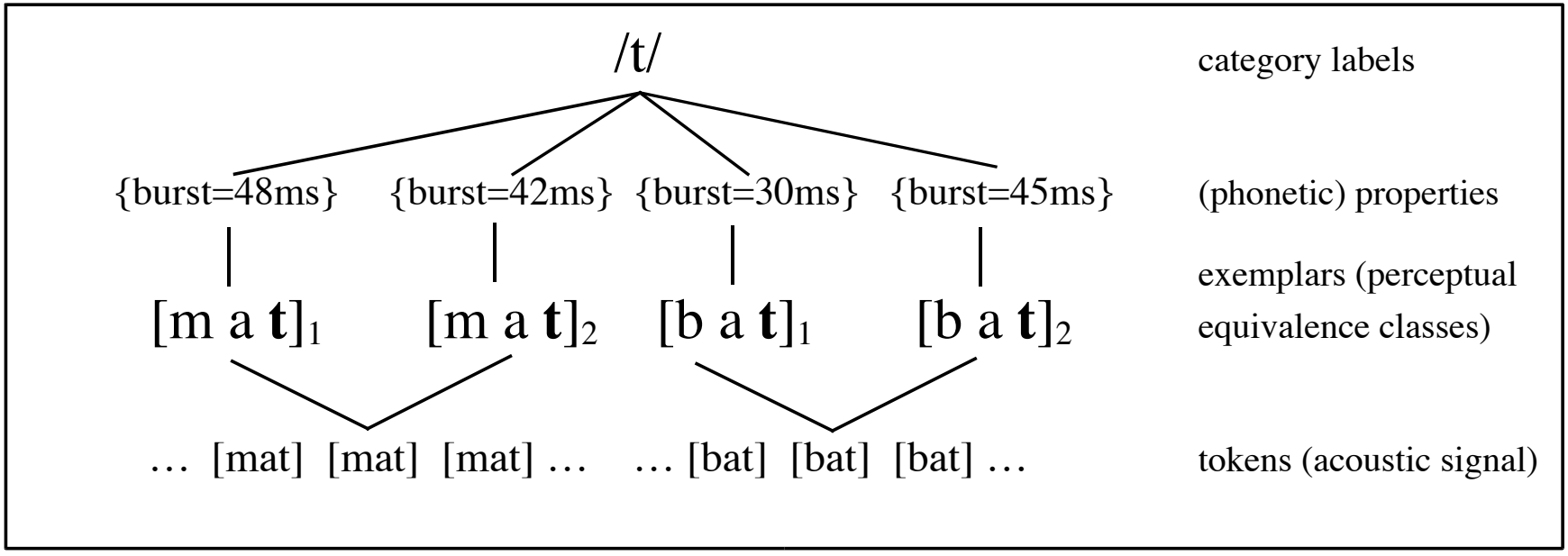
Exemplar storage: Units (knowledge state)



- Individual exemplars represented as (property, label) pairs

$$e_1^L = \begin{bmatrix} burst, & 48\,ms \\ ons_{f0}, & 122\,Hz \\ \dots & \dots \end{bmatrix}, e_2^L = \begin{bmatrix} burst, & 42\,ms \\ ons_{f0}, & 113\,Hz \\ \dots & \dots \end{bmatrix}, \dots$$

Exemplar storage: Constraints

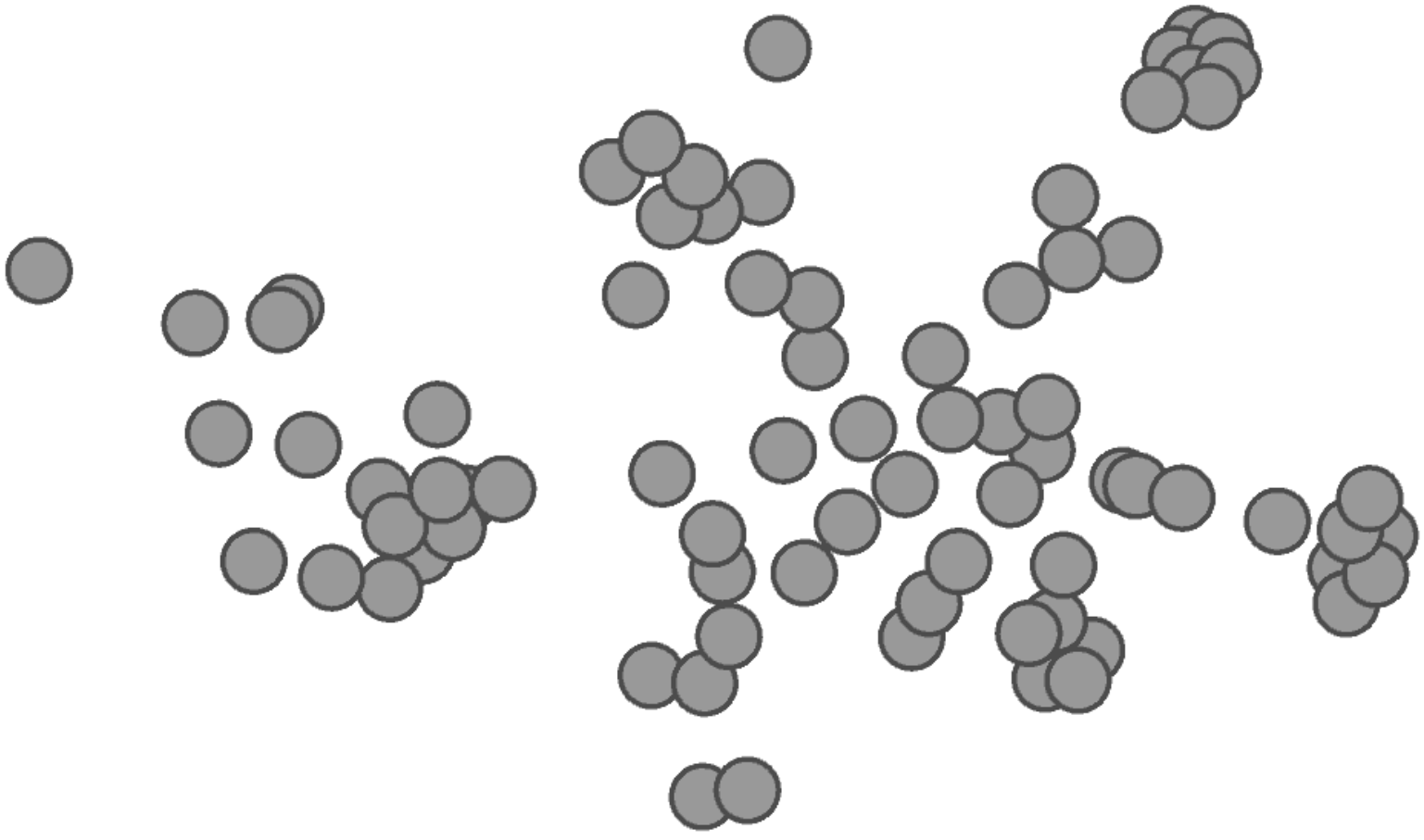


- Memory limitations → memories decay
- Perceptual limitations → granular representations

Model 1: Single category

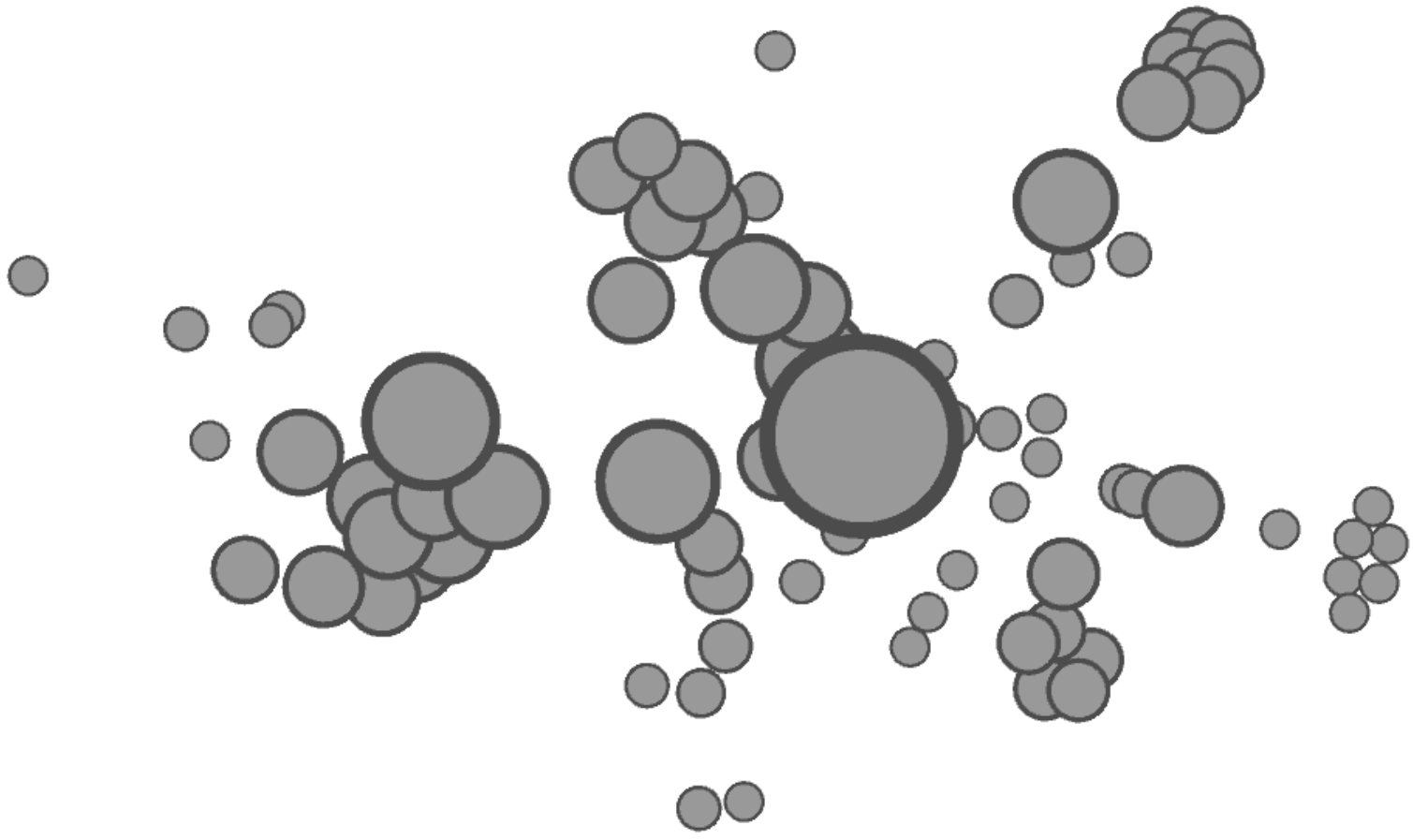
- Single speaker/hearer, talking to themself
- Entails simple(st) network structure
- Knowledge state as detailed above
- What are the assumptions about communication and learning?

Model 1: Production



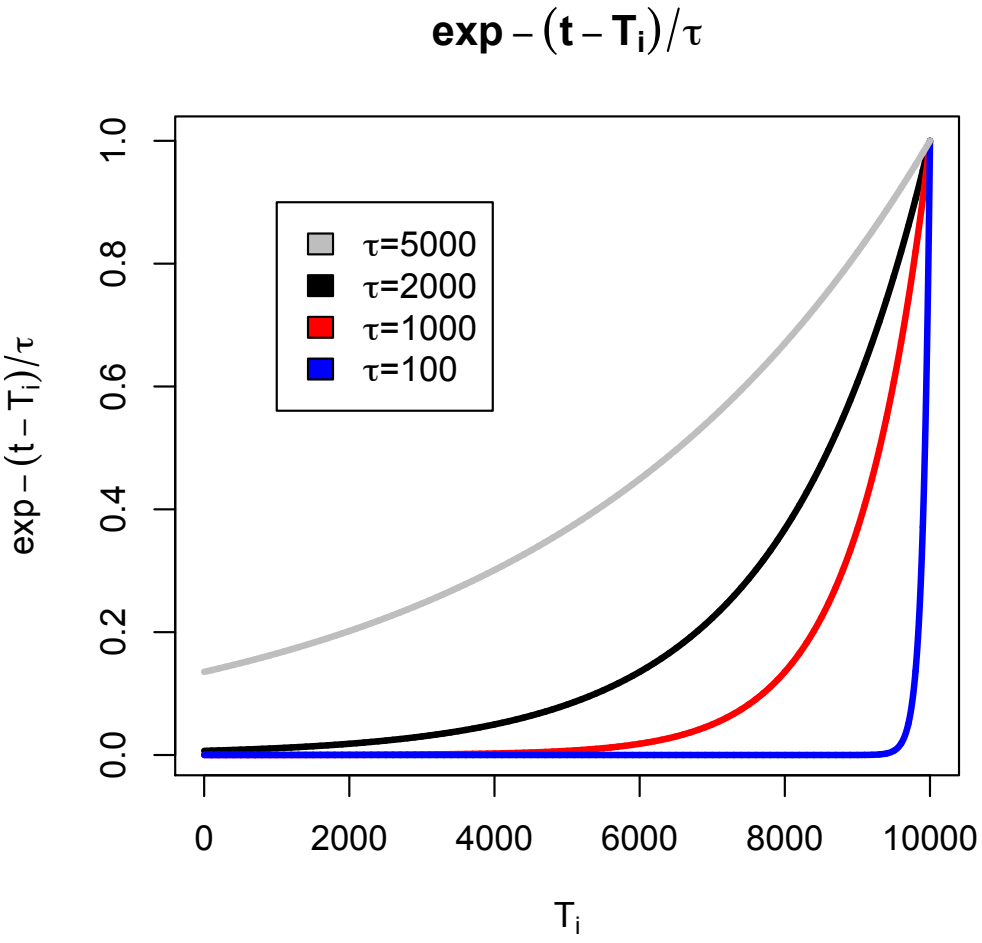
...if target were selected with uniform probability

Model 1: Production



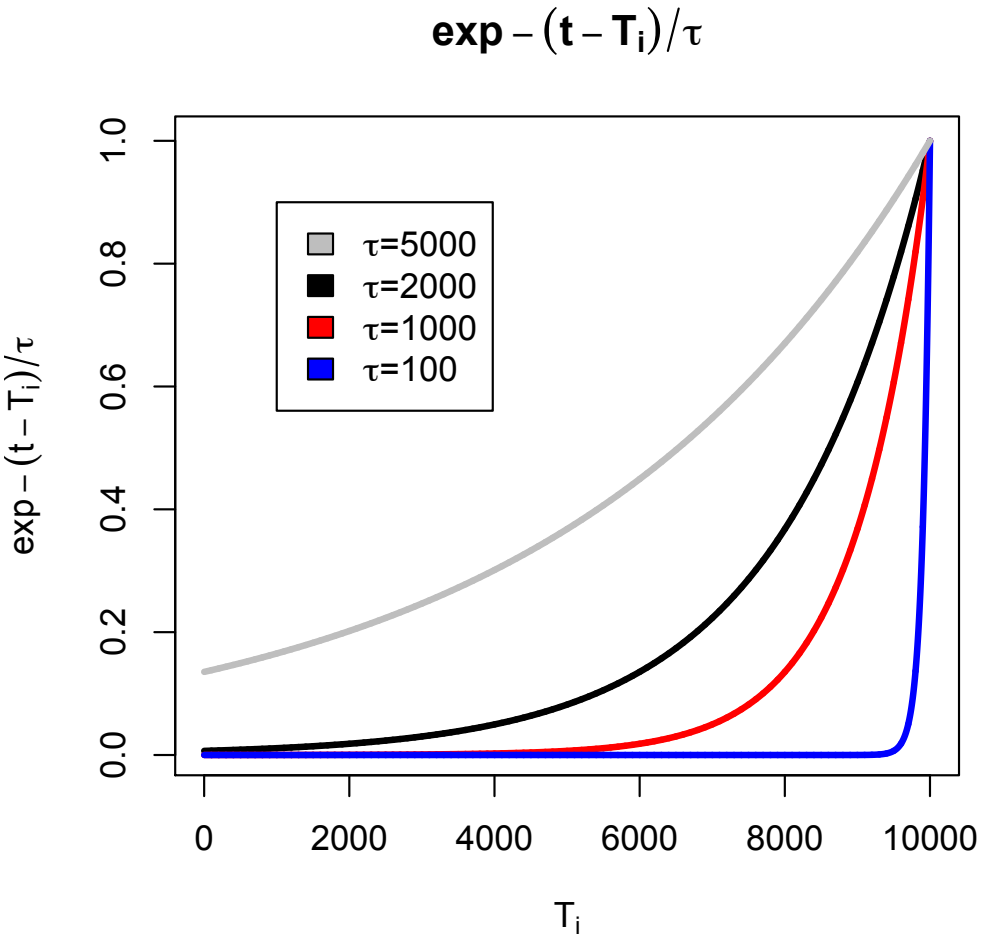
...target selected weighted by recency: $\exp(-\frac{t-T_i}{\tau})$

Model 1: Production



- Need some way model recency, based on:
  - current time
  - time exemplar was stored
  - memory decay weight
- Called resting activation level

Model 1: Production



- Resting activation level a function of time and τ:

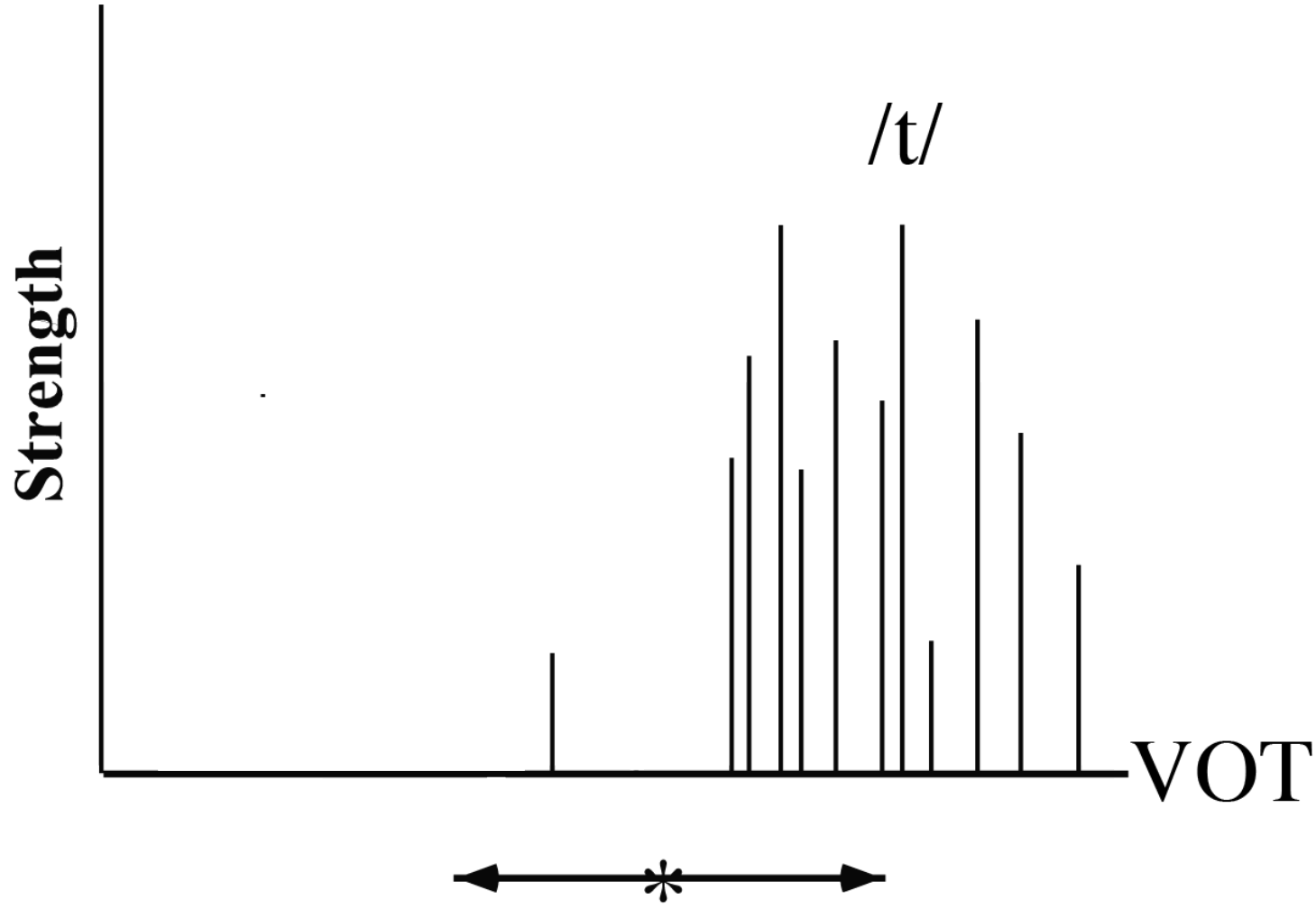
$$\exp(-\frac{t-T_i}{\tau})$$

- Bounded between 0 and 1
- As numerator gets bigger, weight gets smaller

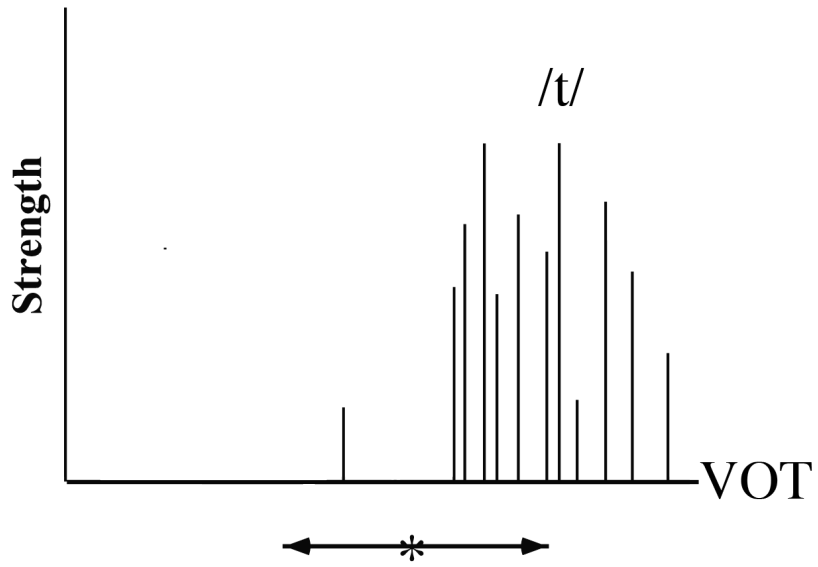
Model 1: Production

- Noise: 'random deviations from the phonetic target due to noise in the motor control and execution'
- Modify production target $x = x_{target} + \epsilon$
- ϵ : random number chosen from a uniform distribution over some fixed range (e.g., from $-.1$ to $.1$)
- (No one seems to have ever dealt with the problem of how to choose an empirically grounded ϵ...)

Model 1: Perception



Model 1: Perception

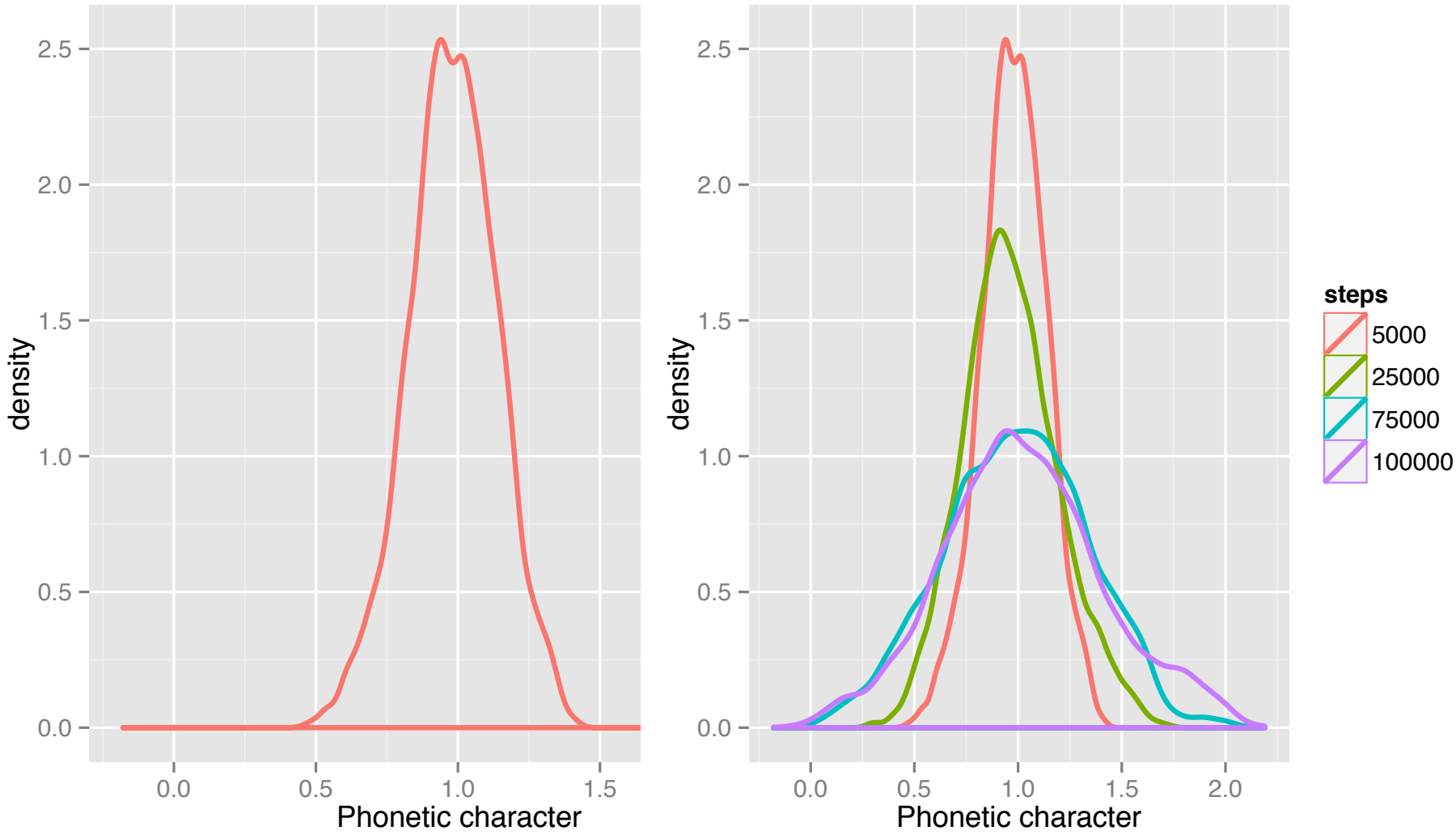


$$W(x - e_i^L) = \begin{cases} 1 & \text{if } |x - e_i^L| < 0.05 \\ 0 & \text{otherwise} \end{cases}$$

$$score(L, x) = \sum_{i=1}^{n} W(x - e_i^L)$$

If score $\neq 0$, x is retained. (learning!)

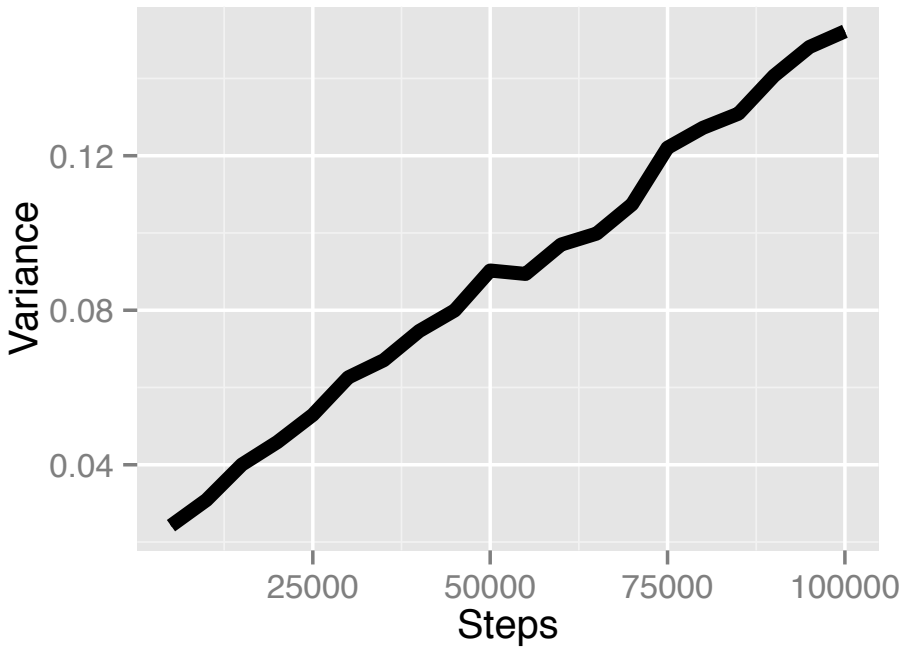
Model 1: Results



cf. Example 1.1

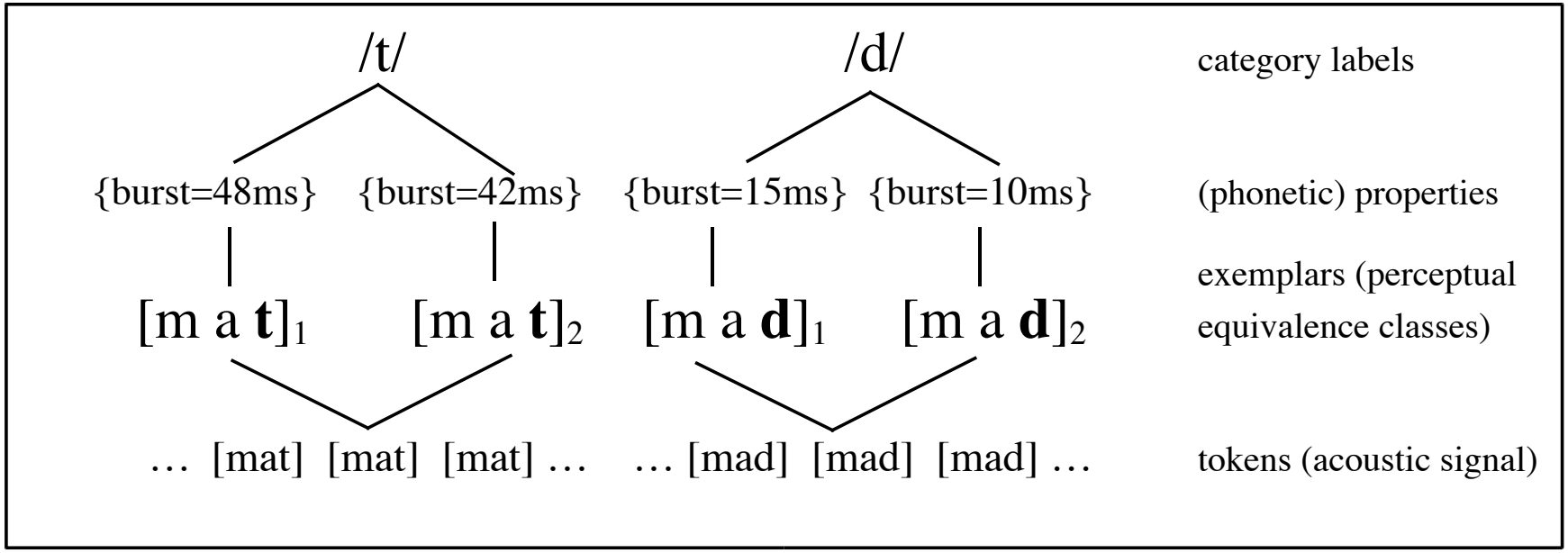
Model 1: Results

What does P conclude from this?



- ‘The total representation of the category is strengthened as more and more memories are stored’
- ‘the variance of the distribution ... increases with usage’

Model 1a: Two categories

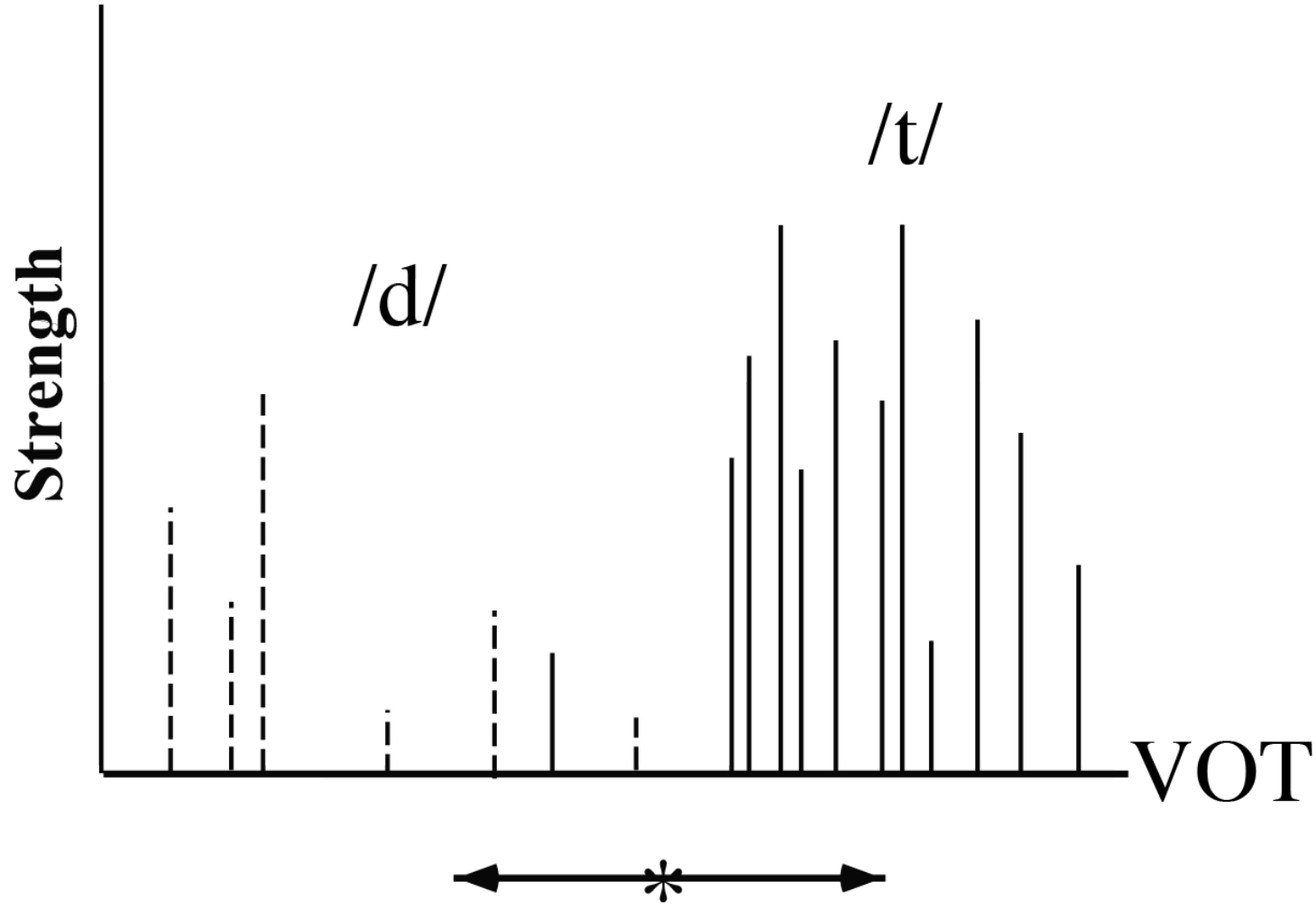


- How do we extend this model to more than one category?

Model 1a: Production

- Now that there are two categories A and B, which one does the speaker utter an instance of at any given timestep?
- Each list seeded w/ a single value; then randomly produce a token x of A or B with probability p and $1 - p$, respectively
- p is fixed/static: e.g. phrase-final voiced obstruents (B) are *a priori* less likely than phrase-final voiceless obstruents (A)

Model 1a: Perception



Model 1a: Perception

- Now there is a meaningful score for each list (category) L:

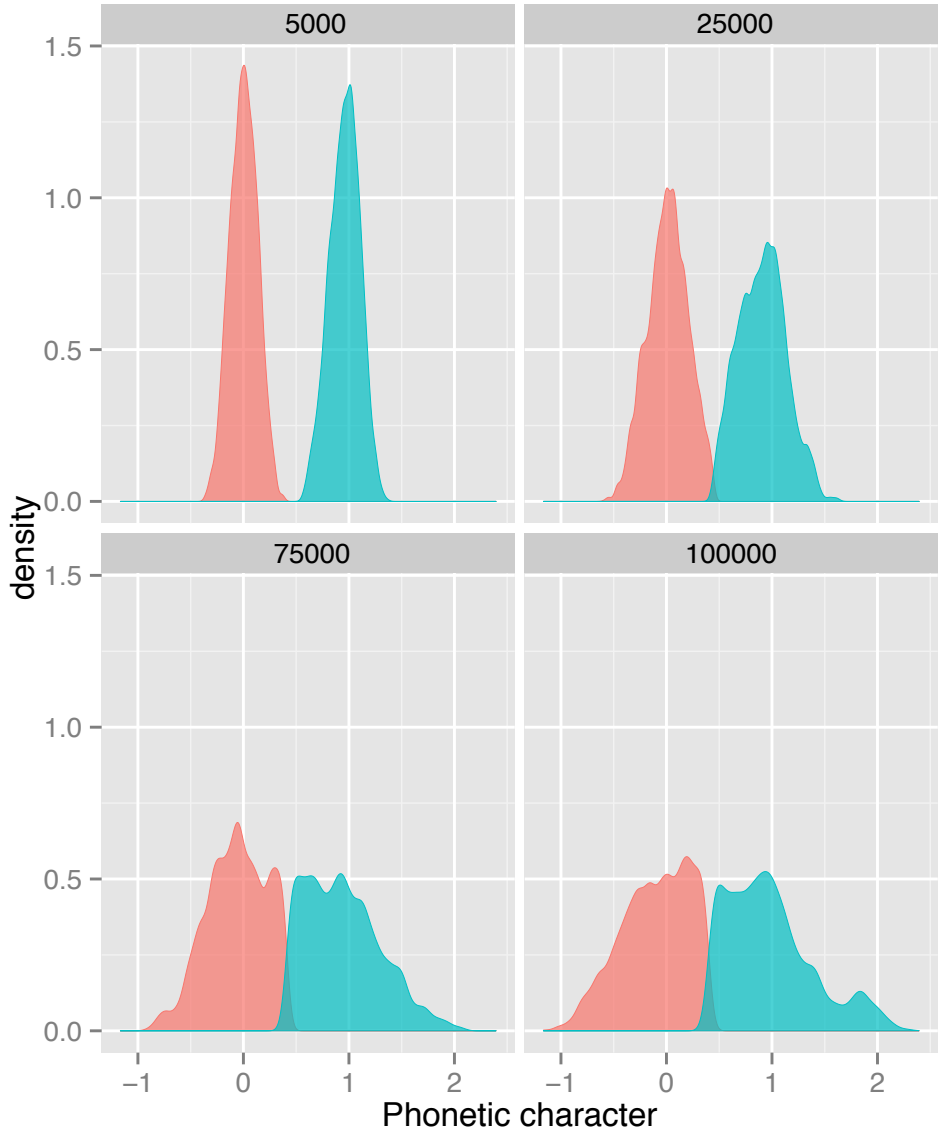
$$score(L, x) = \sum_{i}^{n} \underbrace{W(x - e_i^L)}_{\text{distance}} \overbrace{\exp(-\frac{t - T_i}{\tau})}^{\text{activation level}} \quad (1)$$

- Assign input x to list with the highest score:

$$\arg\max_L score(L_n, x), L \in \{L_1, \ldots, L_n\} \quad (2)$$

- If $L_i = L_j$ for all i, j, discard x.

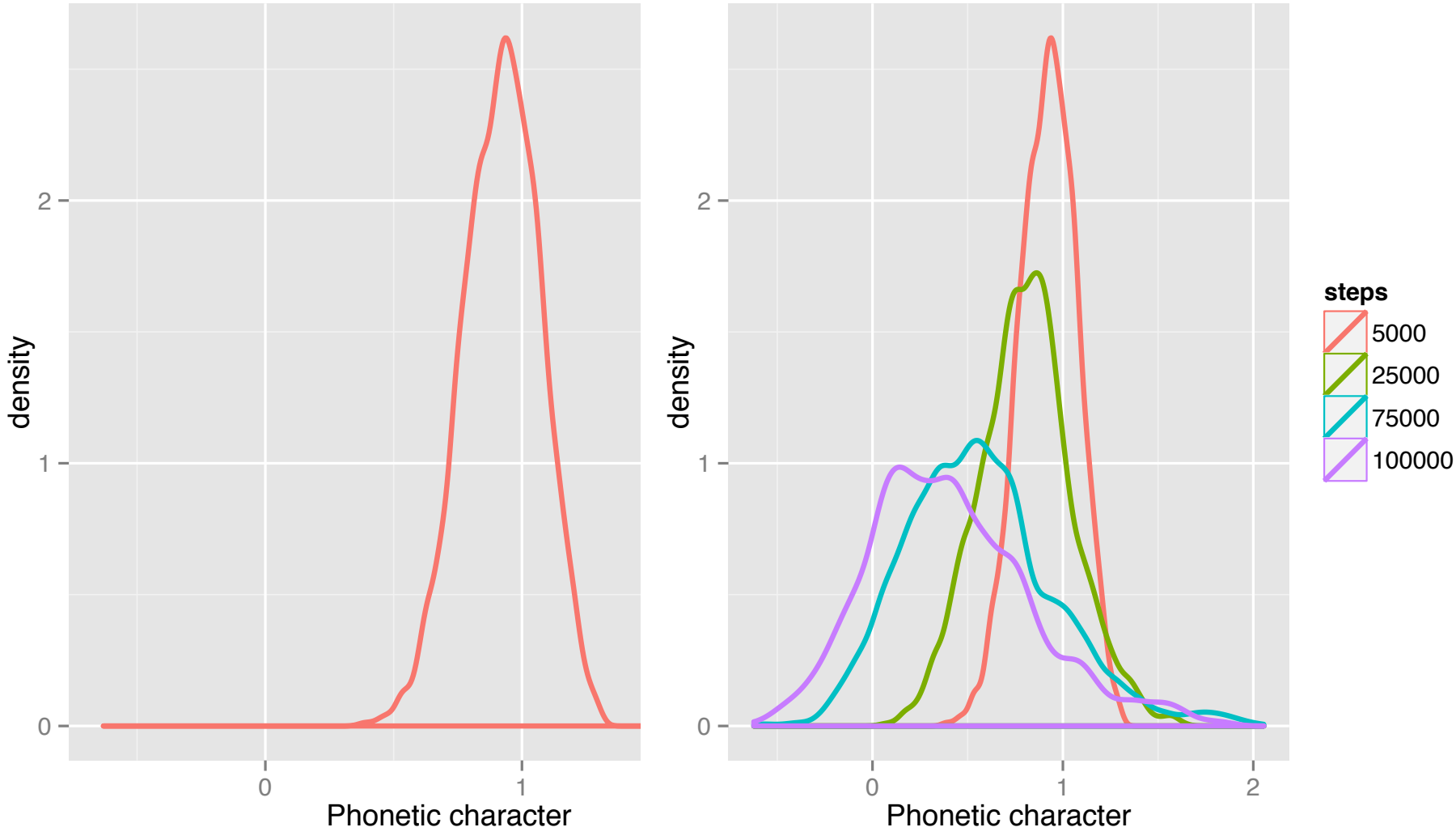
Model 1a: Results (cf. Example 2.1)



Model 2: Bias (lenition)

- Motivation: 'tendency to undershoot ... in order to save effort and speed up communication' (Lindblom, 1984)
- Drives changes like schwa reduction and /t/-deletion?
- Production target is now $x + \epsilon + \lambda$
- λ is a constant (static) parameter (e.g., -0.1)

Model 2: Results



Model 2: Discussion

- ‘We derive the finding that leniting historical changes are more advanced in frequent words than in rarer ones....’
- In what sense does this model ‘derive’ this?
- What is P’s synchronic interpretation of this result? What additional assumptions does it make or rely on?

Model 3: Lenition + Entrenchment

- **Problem**: variance appears to increase without bound
- Aggravated by lenition bias, but...
- Phonetic variability *decreases* with practice (Lee et al 1999)
- 'There is no combination of parameter settings for the model which allows a category to fill out after being seeded by a single example, without simultaneously predicting that the spreading out will go on indefinitely'

Model 3: Lenition + Entrenchment

- **Solution**: entrenchment
- Production involves **averaging over multiple exemplars**
- Pick n_{trench} closest exemplars to target, where 'close' is the inverse of the weighted distance used in perceptual scoring:

$$d_i = |x_{target} - e_i^L| \exp(\frac{t - T_i}{\tau}) \quad (3)$$

$$\omega(T_i, t) = \exp(\frac{t - T_i}{\tau}) \quad (4)$$

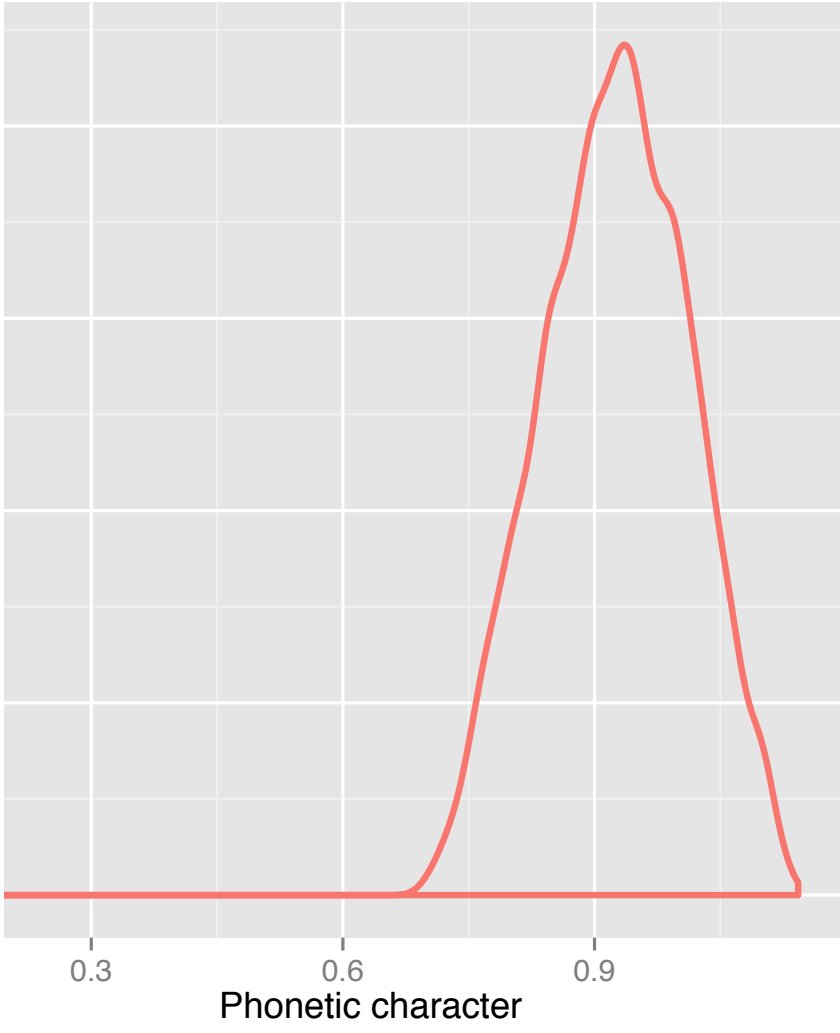
Model 3: Lenition + Entrenchment

- **Solution**: entrenchment
- Production involves **averaging over multiple exemplars**
- Pick n_{trench} closest exemplars to target, where 'close' is the inverse of the weighted distance used in perceptual scoring:

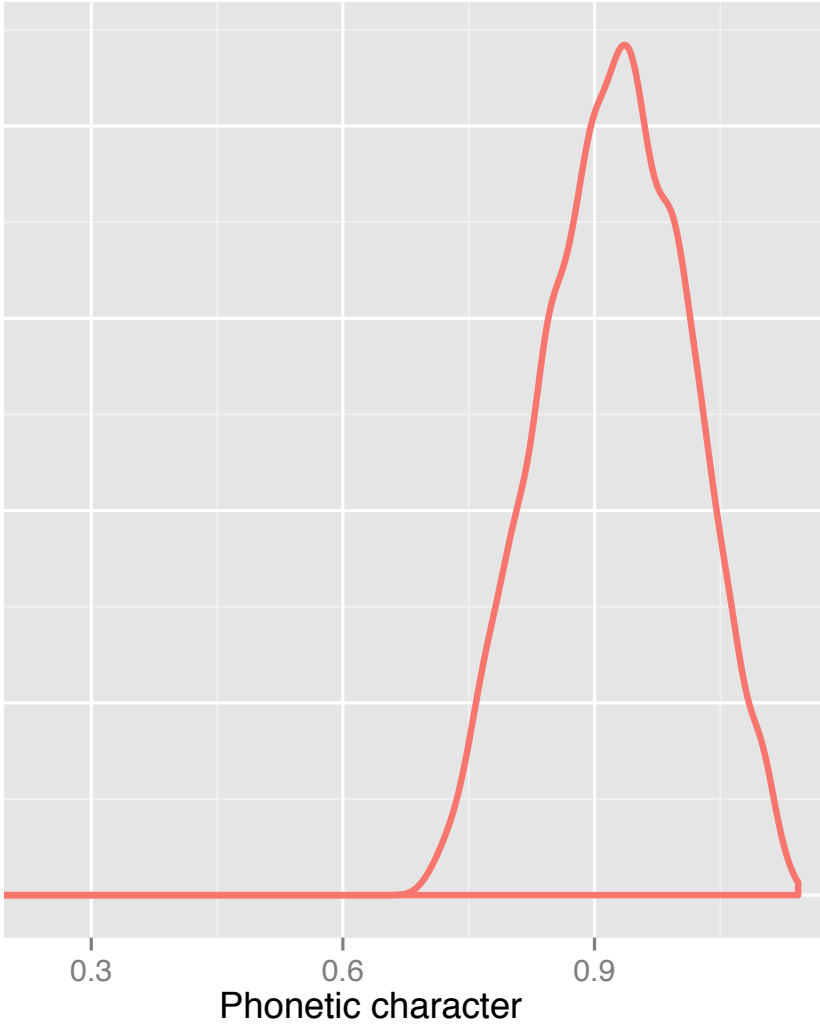
$$d_i = |x_{target} - e_i^L| \, \omega(T_i, t) \tag{5}$$

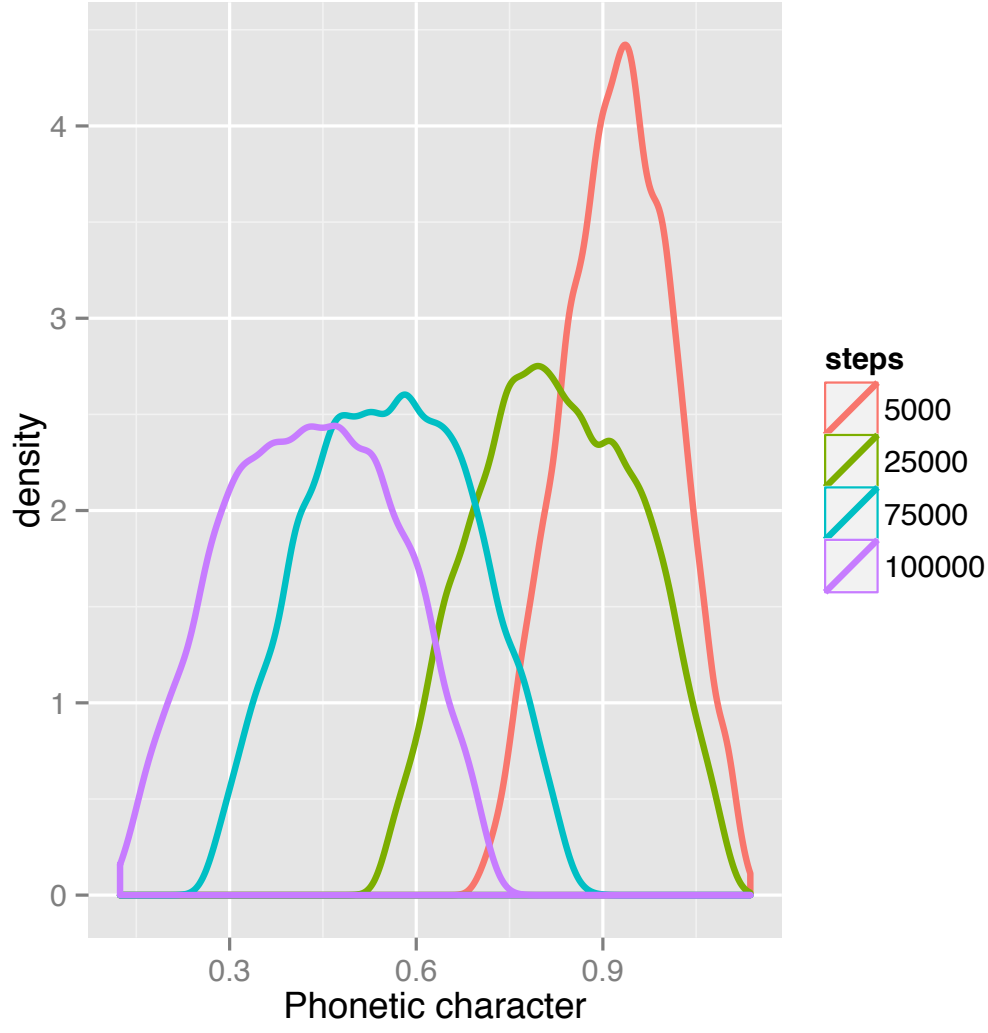
$$x = \frac{\sum x_i \omega(T_i, t)}{\sum \omega(T_i, t)} \text{ for } n_{trench} \text{ closest exemplars} \tag{6}$$

Model 3: Results

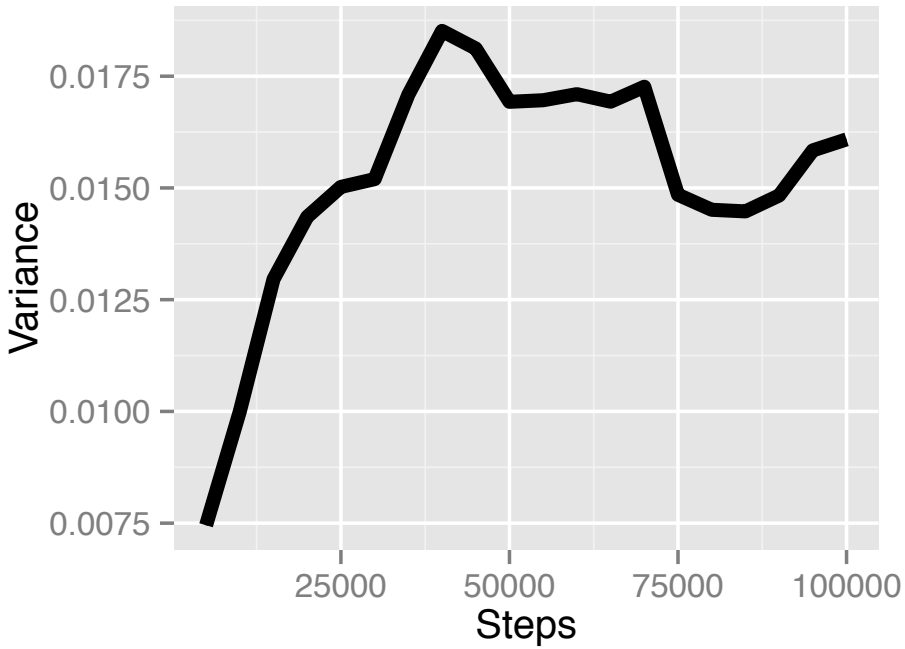


Model 3: Results



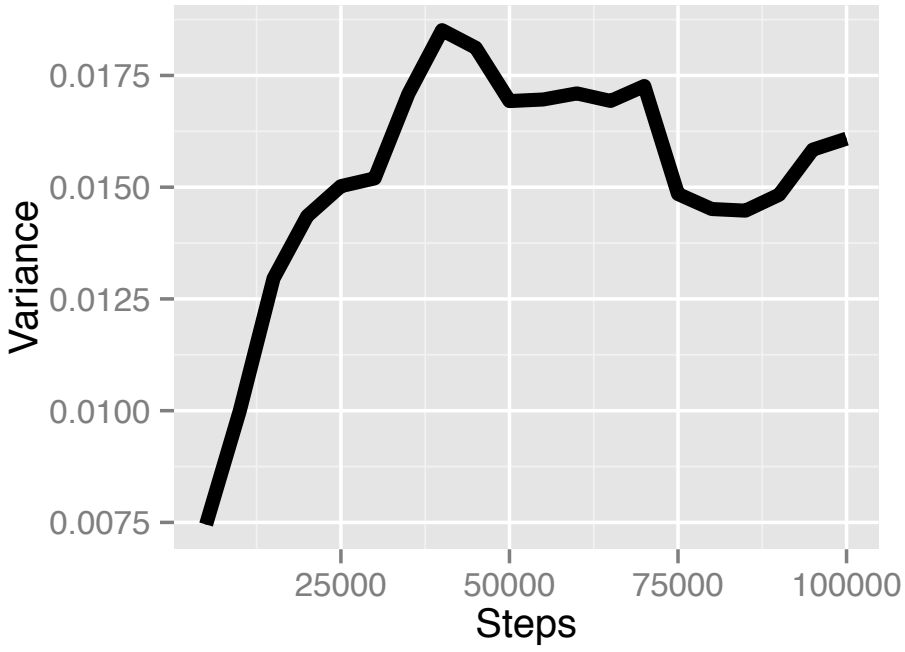


Model 3: Conclusions



- ‘With the particular parameter settings selected here, the spreading effects arising from production noise and lenition and the anti-diffusive effect of entrenchment have essentially cancelled out in determining the variance.’

Model 3: Conclusions



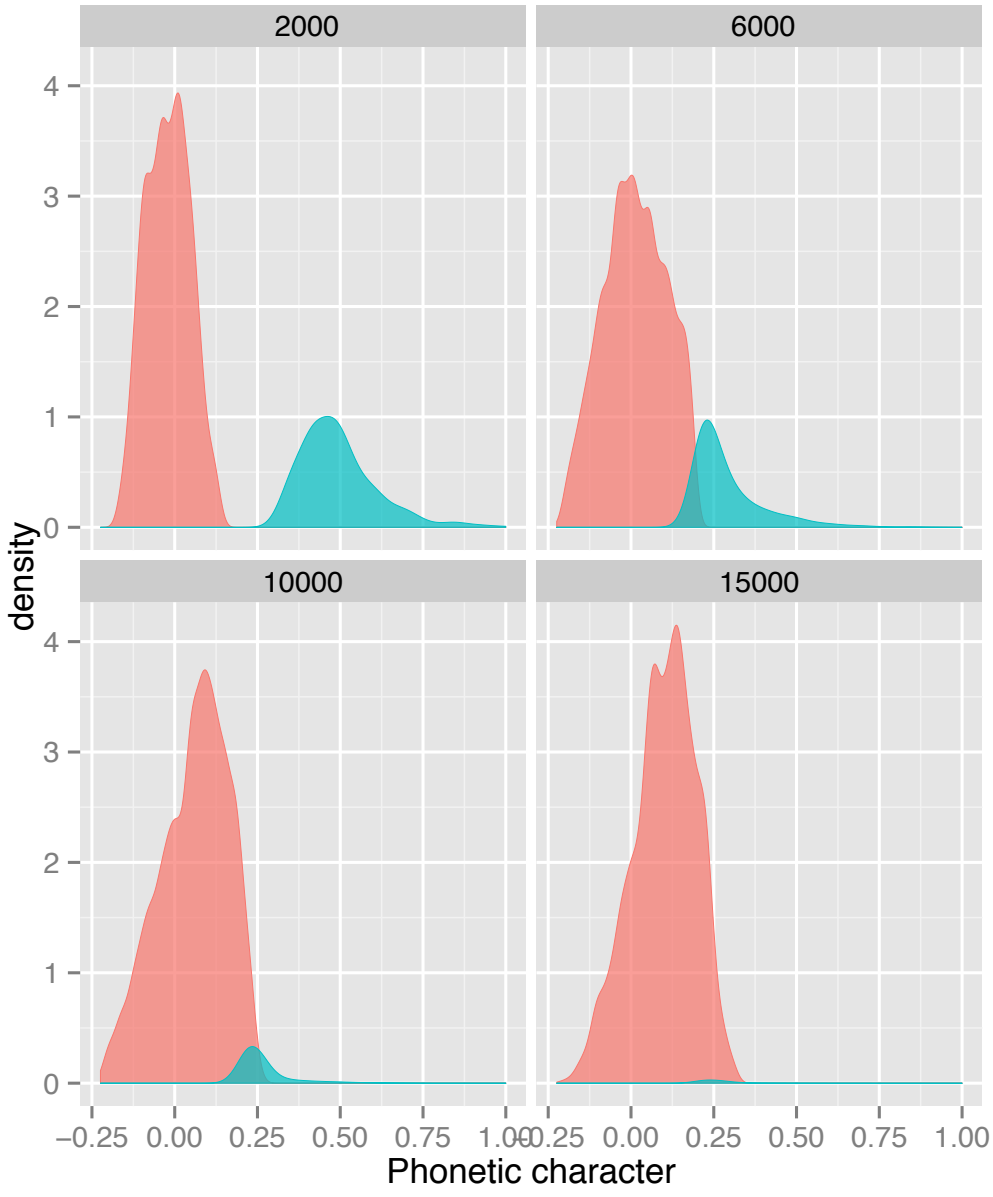
- In other words: variance stabilizes at some point
- For what range of parameter settings will this property obtain?

- (Extra credit: why does Wedel's model not seem to need this?)

Model 4: Neutralization

- **Intuition:** lenition bias in a region of phonetic space shared by multiple categories will eventually result in stable merger
- 'We consider the case of a marked phonological category competing with an unmarked one....we take the unmarked category to be more frequent than the marked one...'
- P shows us lenition *plus* entrenchment *plus* unequal frequency
- (What about e.g. just unequal frequencies, full stop?)

Model 4: Results



- Is this a ‘realistic treatment of the time course of phonological merger’?
- If it is, what should the time course of change look like? (in individuals/in communities)

Assumptions/limitations

- **Category labels.** Here: given. Where do they come from?
- **Words vs. phonemes.** What level of granularity is being modeled here? How do these levels relate to one another?
- **Unidimensionality of phonetic character.** What would be required to move to a multidimensional representation?
- **Network/social structure.** Realism of modelling single speaker-hearer?
- ...

What have we learned?

- (Non-)existence proof(s)?
- Explicitness?
- Counterintuitive results?
- Qualitative predictions?
- Baseline?

Lab: extensions and explorations

- **Frequency and lenition.** What is necessary for neutralization to obtain: differences in frequency (p), lenition, or both?
- **Entrenchment: one category.** Does the variance always stabilize with nonzero entrenchment? Evaluate the claim that stable variance decreases as n_{trench} increases.
- **Entrenchment: two categories.** Vary the amount of entrenchment in the neutralization simulation. What happens?
- **Window size.** How does changing the window size impact the evolution of categories in the two category cases?

Lab: extensions and explorations

- **Changing τ.** How does changing τ effect the evolution of the mean and variance of a category over time?
- **Noise.** Is there an effect of increasing/decreasing the amount of noise (i.e., changing endpoints of the distribution from which noise (ϵ) is drawn)?
- **Lenition.** What happens in the neutralization scenario as the amount of lenition λ is varied?
- **Advanced topics.** 3 categories? Alternative scoring function?

Post guidelines

- What did you try?
- What did you find? (pictures optional)
- Can you give an intuitive explanation for why you got the result that you did?
- What empirical/linguistic prediction(s) does your finding make?

Post guidelines

- 1 member of each group: please **email us group list** ASAP
- Posts due **5pm Sunday 26 July**