LI 308: Computational Models of Sound Change

James Kirby and Morgan Sonderegger

27 July 2015

Pierrehumbert model explorations

- Selected results

Varying ε : 1 category

- Trajectory of mean: increasingly noisy
- Trajectory of variance: increasingly steep

- Long-term, qualitative outcome: Invariant

Varying window size: 1 category

- Larger window:
  - Variance increases more slowly
  - Mean `` ``
- Window size parametrizes timescale of simulation
- Qualitative outcome: invariant

Varying n_{trench} : 1 category

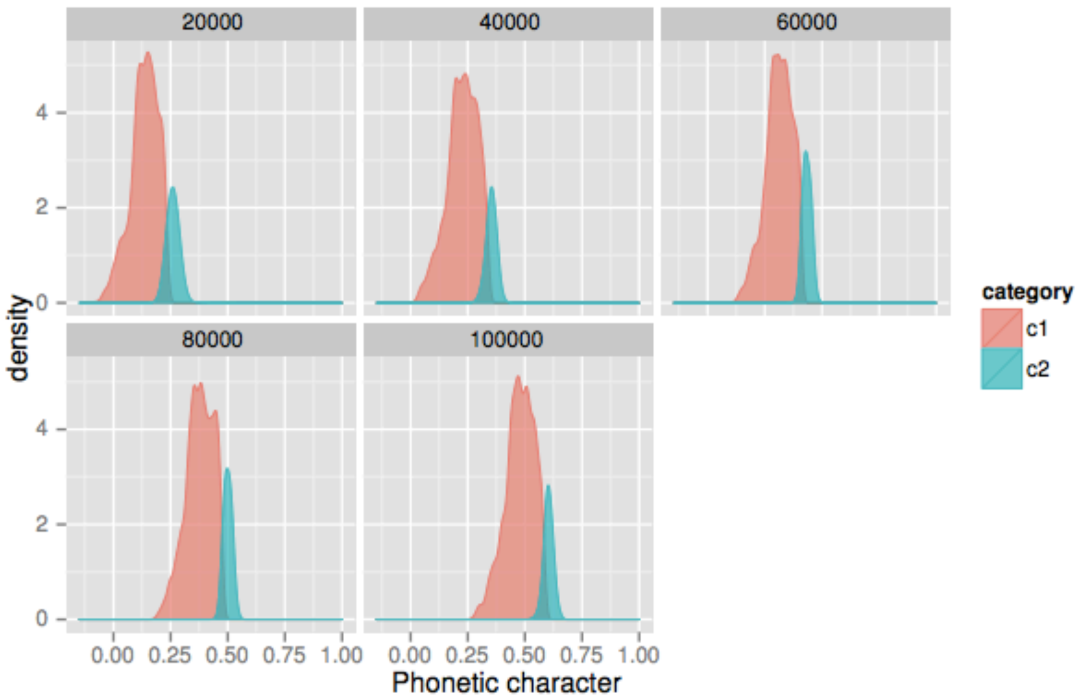
- n_{trench} small enough relative to ε : variance <u>never</u> stabilizes
  - Outweighed by production noise
  - Surprising!

- Larger n_{trench} : variance stabilizes
  - Stable value decreases

- Qualitative outcome: varies

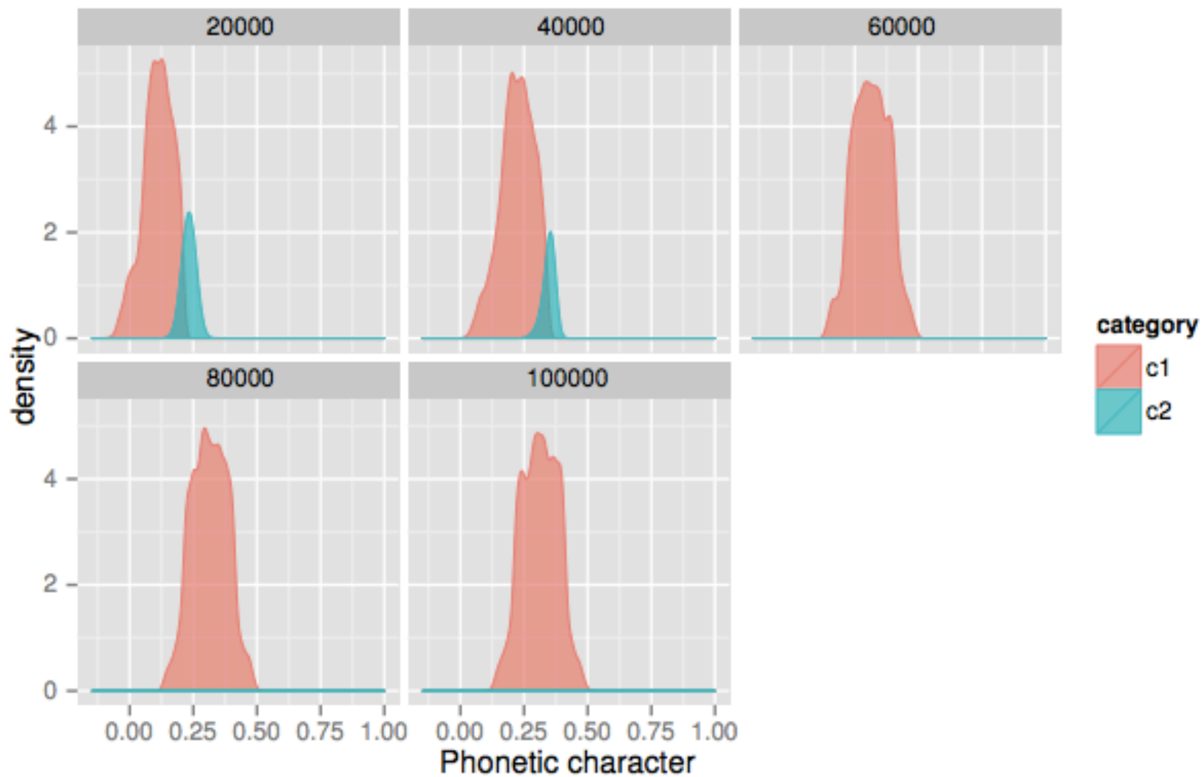
Varying n_{trench}, λ : 2 categories

- n_{trench} = 750, λ =-0.07



- Entrenchment (+ category overlap effect) balances out lenition

- n_{trench} = 500, λ =-0.07



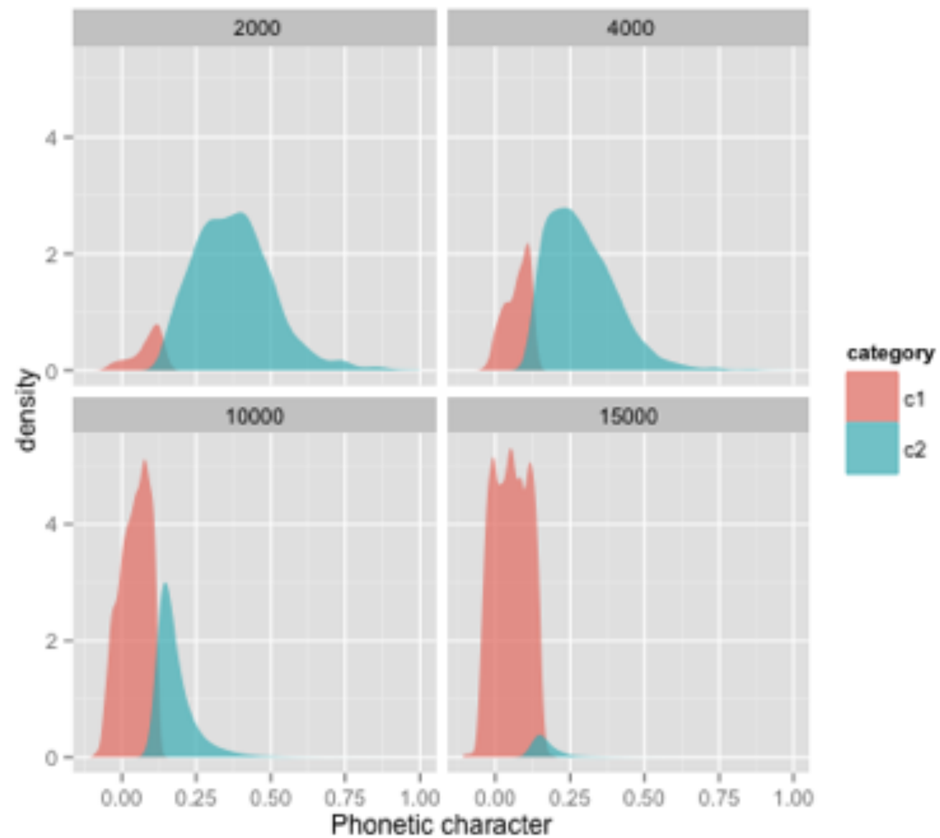
- Lenition overcomes entrenchment (+category overlap effect)

- Qualitative outcome: Varies
  - Neutralization if lenition strong enough relative to entrenchment

Varying p, λ : 2 categories

- λ for c_1 <u>and</u> c_2: λ_1, λ_2
- Pierrehumbert case: λ_2=-0.1, λ_1=0, entrenchment, p=0.75
- Varying p: No effect! c_2 always neutralizes.

λ_2=-0.1, λ_1=0, p=0.05



- Lenition values determine neutralization, not p
  - Category frequency ≈ doesn't matter
  - Surprising (why?)

- Qualitative outcome: varies

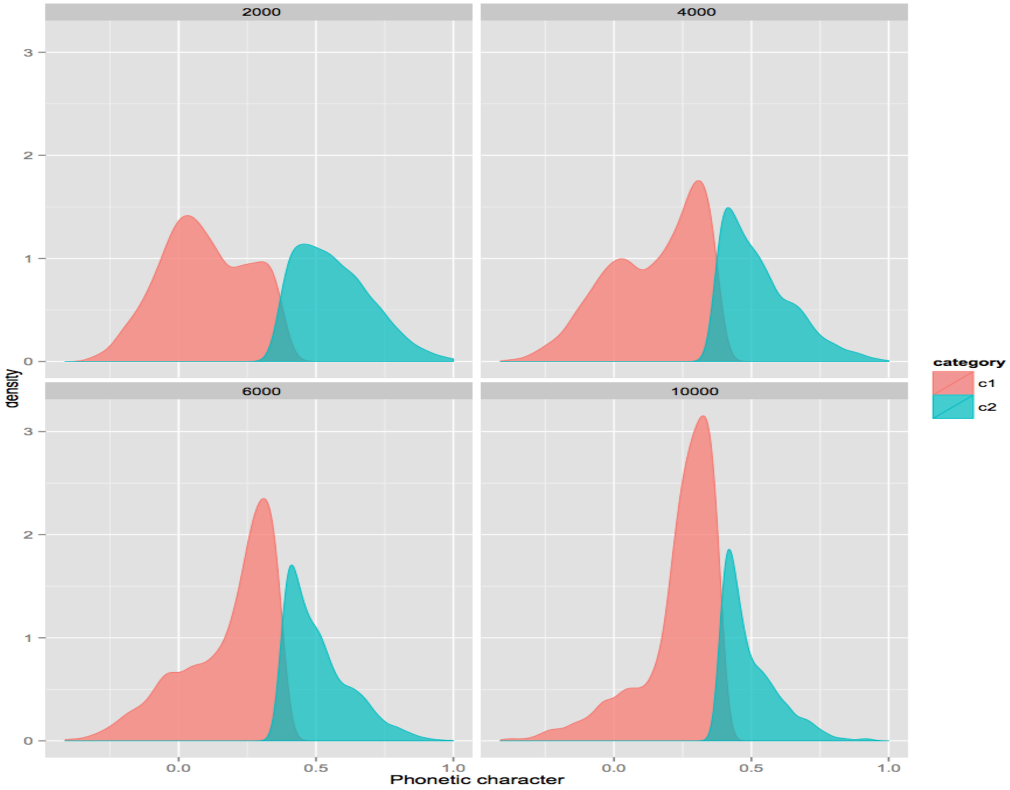
Summary: parameter space

- Some parameters (varied alone) affect aspects of category evolution, but no/little effect on eventual outcome
  - Ex: window size, noise

- Some change the outcome a lot
  - n_{trench}
  - $\lambda_{1,}\ \lambda_2$, interaction with p
  - n_{trench} interaction with $\lambda_{1,}\ \lambda_2$

Note for Lab 2

- Make sure to run simulations sufficiently long
  - (increase *t* if necessary)



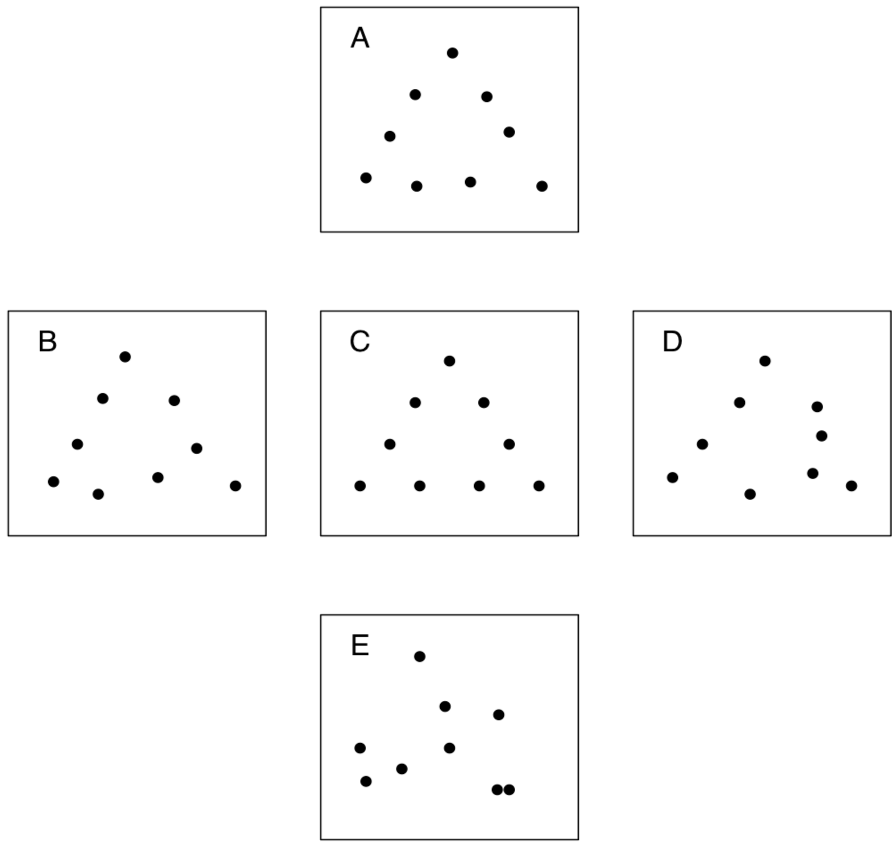
Categorization as typicality

How can we explain typicality?

- Exemplar theories
  - Category = multiple instances
  - No independent representation
  - Typicality: "best exemplar"

- Prototype theory
  - Categories represented by a prototype
  – Typicality: function of similarity to the prototype

Exemplar theories



Store every member ("exemplar") of the category

Formalizing exemplar theories

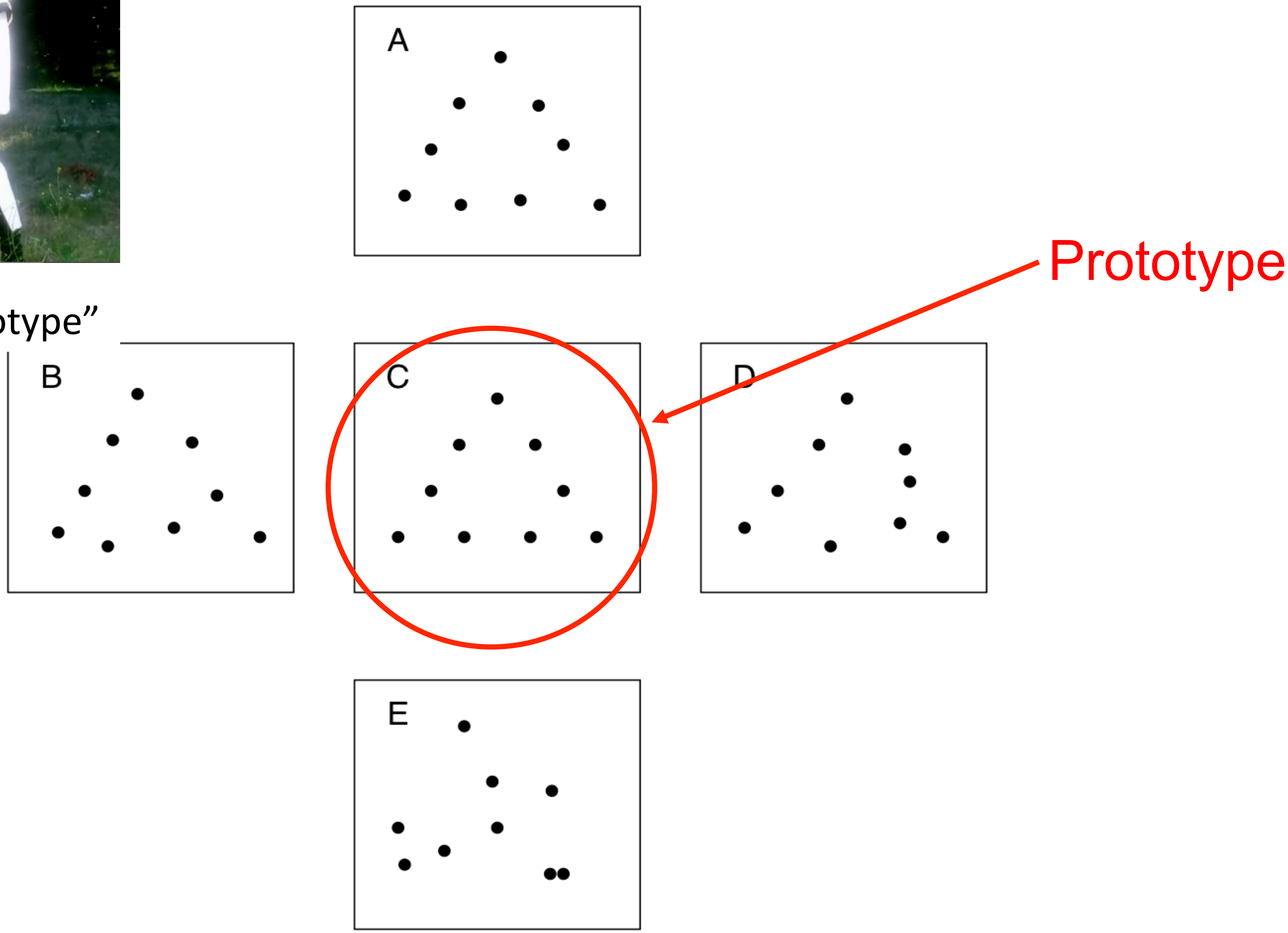
- Multiple things stored per category
  - Pierrehumbert: All members
  - Wedel: *n* most recent members

- **Representation:** e.g. a set of stored exemplars $y_1, y_2, \ldots, y_n$, each with a weight

- **Distance**: Window function, or generalized context model (Nosofsky, 1986)

Prototypes

Outkast, “Prototype”



(after Posner & Keele, 1968)

Formalizing prototype theories

- Single thing stored per category
  - Often: most frequent or “typical” member

- **Representation**: e.g. average of all inputs so far

- **Distance:** e.g. Euclidean

$$d(x, A) = \left[\sum_i (x_i - \eta_{A,i})^2 \right]^{1/2}$$

  - de Boer: Weighted Euclidean in (F1, F2’) space

Prototypes vs. exemplars

- Both approaches seem reasonable...?
- Both have been applied in speech
  (prototypes: Samuel 1982, Kuhl 1991; exemplars: Pisoni 1992, etc.)
- Can be hard to tell predictions apart, in practice
  - (why?)

Background: Vowel inventory typology

- UPSID: 451 inventories
  (Maddieson, 1984; Maddieson & Precoda, 1990)

- Certain regions of formant space preferred (e.g. point vowels)

- n = 5-7 vowels : preferred
  - N = 1-2: vertical

- Central vowels
  - largely only introduced at n=6, n=9+
  - central vowels ⇒ peripheral vowels

More cross-linguistic tendencies: Schwartz et al. 1997

Explanations for patterns

- Structural ("innate")
  (Clements, 2003; Dresher, 2003; Rice, 1999; Government Phonology…)
  - Preferred feature, segment values
  - Preferred feature, segment inventories
    (e.g. symmetry, contrastive hierarchy..)

- Functional
  - Quantal Theory (Stevens 1972, 1989)
  - Versions of dispersion
    (Liljencrants & Lindblom, 1972; Flemming, 1995 et seq; others)
  - Combo (e.g. Carré 1994)

Dispersion

- Idea: Vowels are optimally dispersed in some space
  - Usually: Acoustic
- Liljencrants & Lindblom (1972)
  - (Day 1)

de Boer (1999, 2000)

- Previous work:
  1. Assumes innate structures, or biases
  2. Global optimization of inventories
  3. Considers inventories of individuals (not population-level "inventory")
- Questions:
  - How does population-level inventory result from individual actions/learning?
  - Can we get self-organizing behavior without assuming (1) or (2)?

Model assumptions: High level

- No <u>global</u> optimization
  (c.f. Lijencrants & Lindblom; Flemming)

- Update based on imitations only
  (no talking to self)

- No "looking inside partner's head"
  - Don't get partner's <u>intended category</u>
    (c.f. Wedel, any "supervised learning")

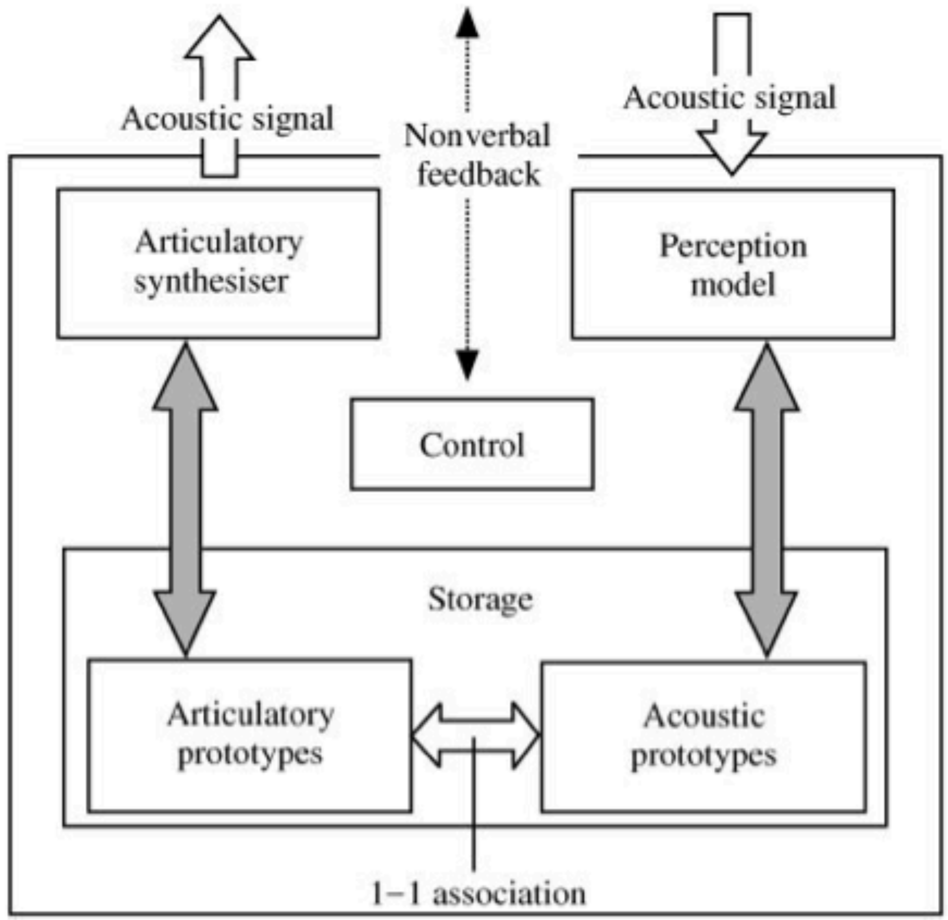
Goals, non-goals

- Not modeling acquisition
  - Children learn vowel systems differently

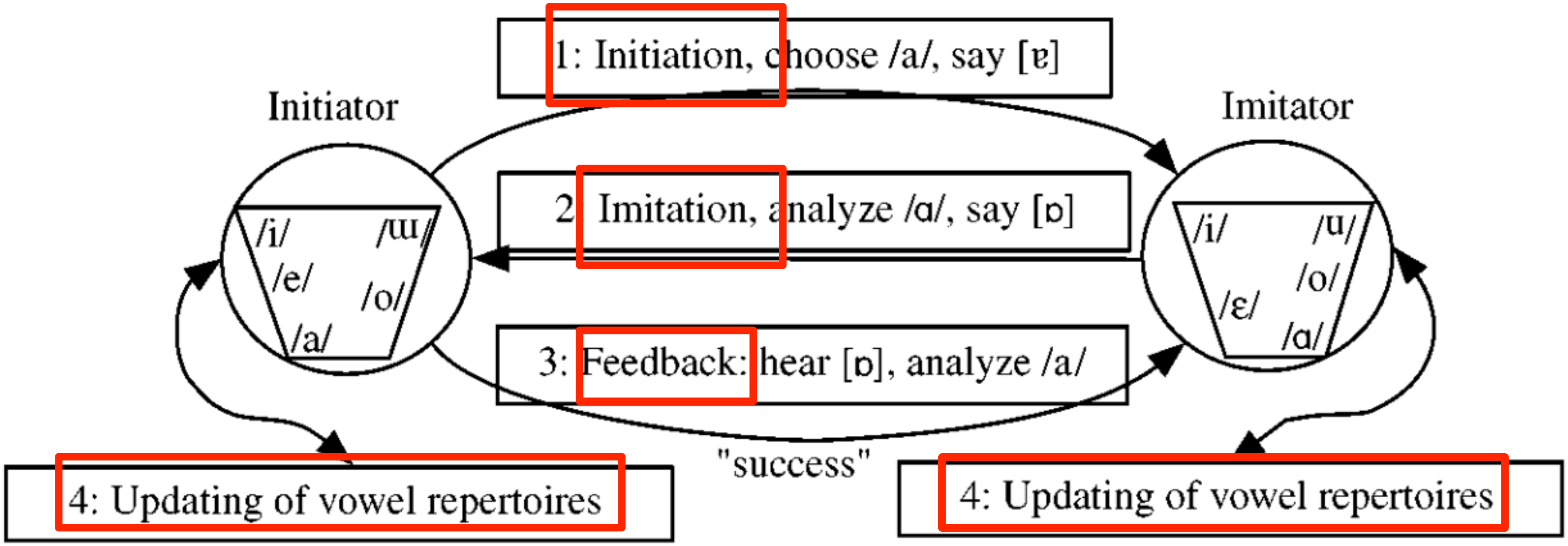
- Not modeling historical evolution
  - That's more complex (454)

- Baseline
  - "whether self-organization can explain the universal tendencies of human vowel inventories." (453)

Agent architecture



Imitation game



Changing the vowel system

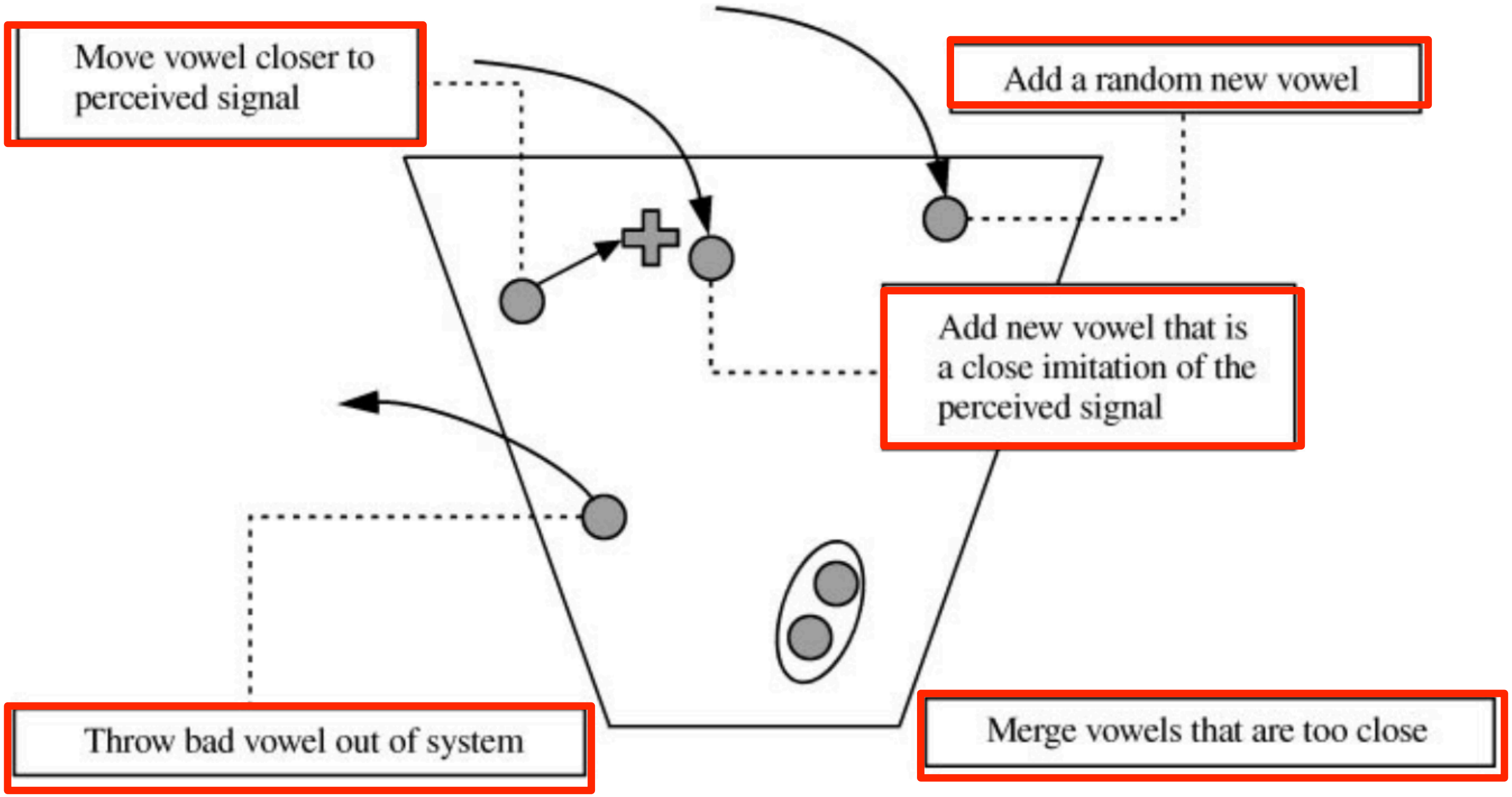


Figure 3. Changes an agent can make to its vowel system. Circles indicate vowels in the agent's repertoire (both articlatory and acoustic aspects) while the cross indicates the position (in acoustic space) of the signal the agent just perceived.

Simulation parameters

Notation	Description
ψ_{ac}	Maximal noise added to formants
λ	Relative weight of F_2' with respect to F_1
n_{agents}	number of agents
n_{its}	number of iterations
ϵ_{artic}	articulatory step size for shifting prototype

Varied in coming simulations (next slides)

- (See extensions handout)

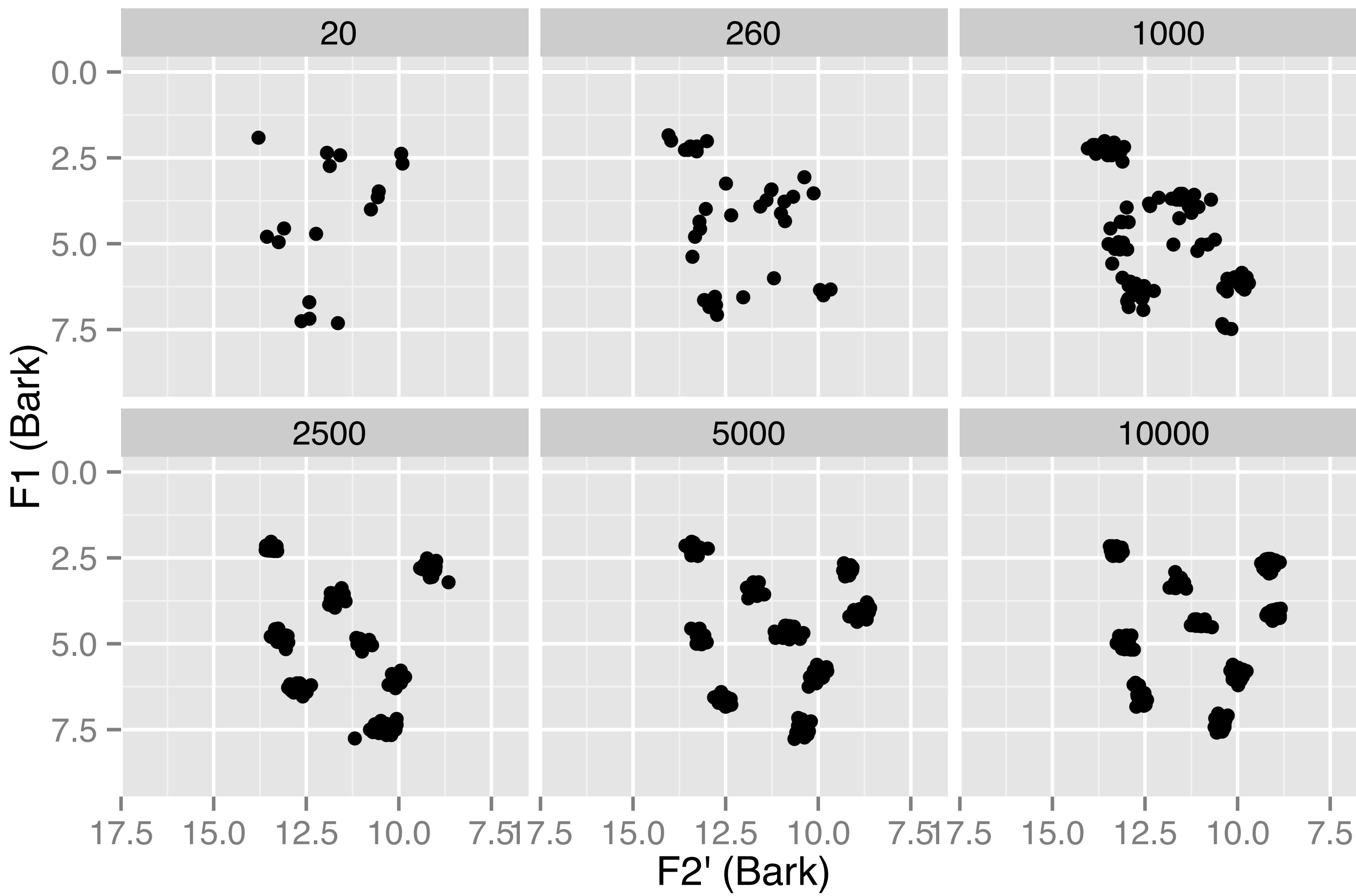
Simulation parameters

$\theta_{discard}$	success/use threshold below which vowels are discarded
$\theta_{acoustMerge}$	(weighted) Euclidean distance in $F1/F2'$ space under which vowels are merged
$\theta_{articMerge}$	Euclidean distance in articulatory space under which vowels are merged
$n_{minUses}$	minimum number of times a vowel must be used to consider discarding
$p_{addition}$	probability with which an agent adds a random new vowel
$p_{cleanUp}$	probability with which an agent cleans up (discards and merges) in each round

Not varied in coming simulations (next slides)

Sample run

- n_{its}=10000, n_{agents}=20, ψ=0.1, ε_{artic}=0.03, λ = 0.3



Evaluation measures

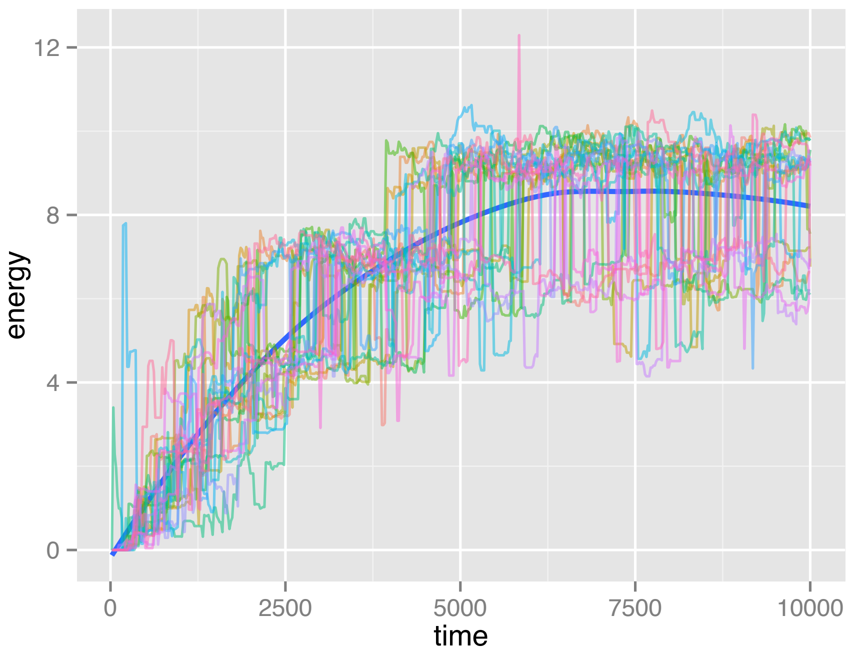
- Success rate (1 agent): $SR = \sum_{v} \frac{s_v}{u_v}$

Evaluation measures

- Inventory energy: $$E = \sum_{(v_1, v_2)\,pairs} \frac{1}{D(v_1, v_2)^2}$$

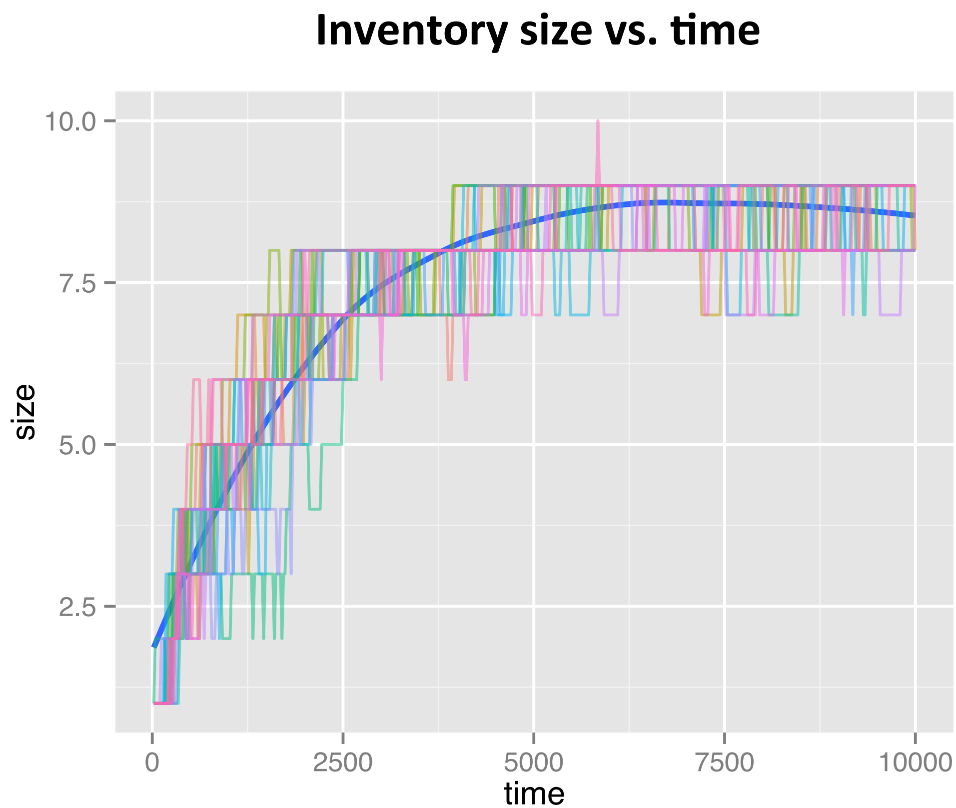
- Motivation:
  "…minimizing the energy function has been shown to result in realistic vowel systems"
  (e.g. Liljencrants & Lindblom, 1972)

Inventory energy vs. time



Evaluation measures

- Inventory size



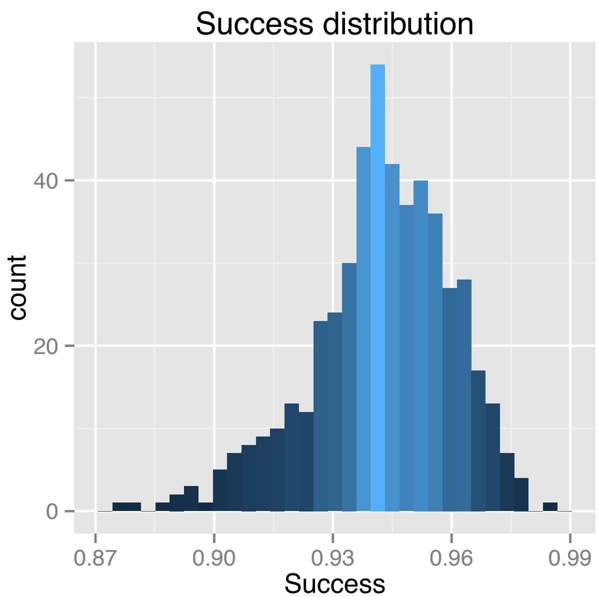
- Success prob, energy, inventory size all stabilize around t=5000

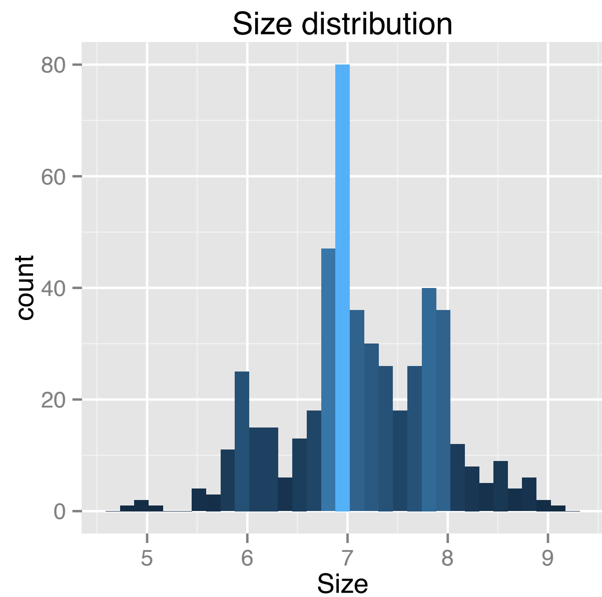
Q: why are inventory size and energy trajectories similar?

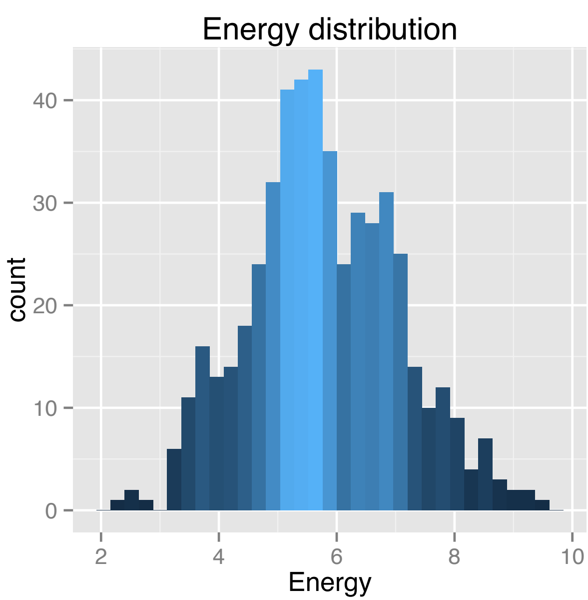
- Finding 1: Vowel systems emerge (for some parameter settings)
  - Individuals converge on similar systems of vowels
  - Relatively stable over time

- How variable is the resulting system, for fixed parameter settings?
  - Important for any simulation where evolution is stochastic
    - (e.g. stochastic GLA, agent-based simulations, …)

1. Run 500 times
   - n_{its}=5000, n_{agents}=20, ψ=0.1, $\varepsilon_{\text{artic}}$=0.03, λ = 0.3
2. Examine histograms of evaluation measures
   - Mean over all agents' inventories after n_{its} interactions







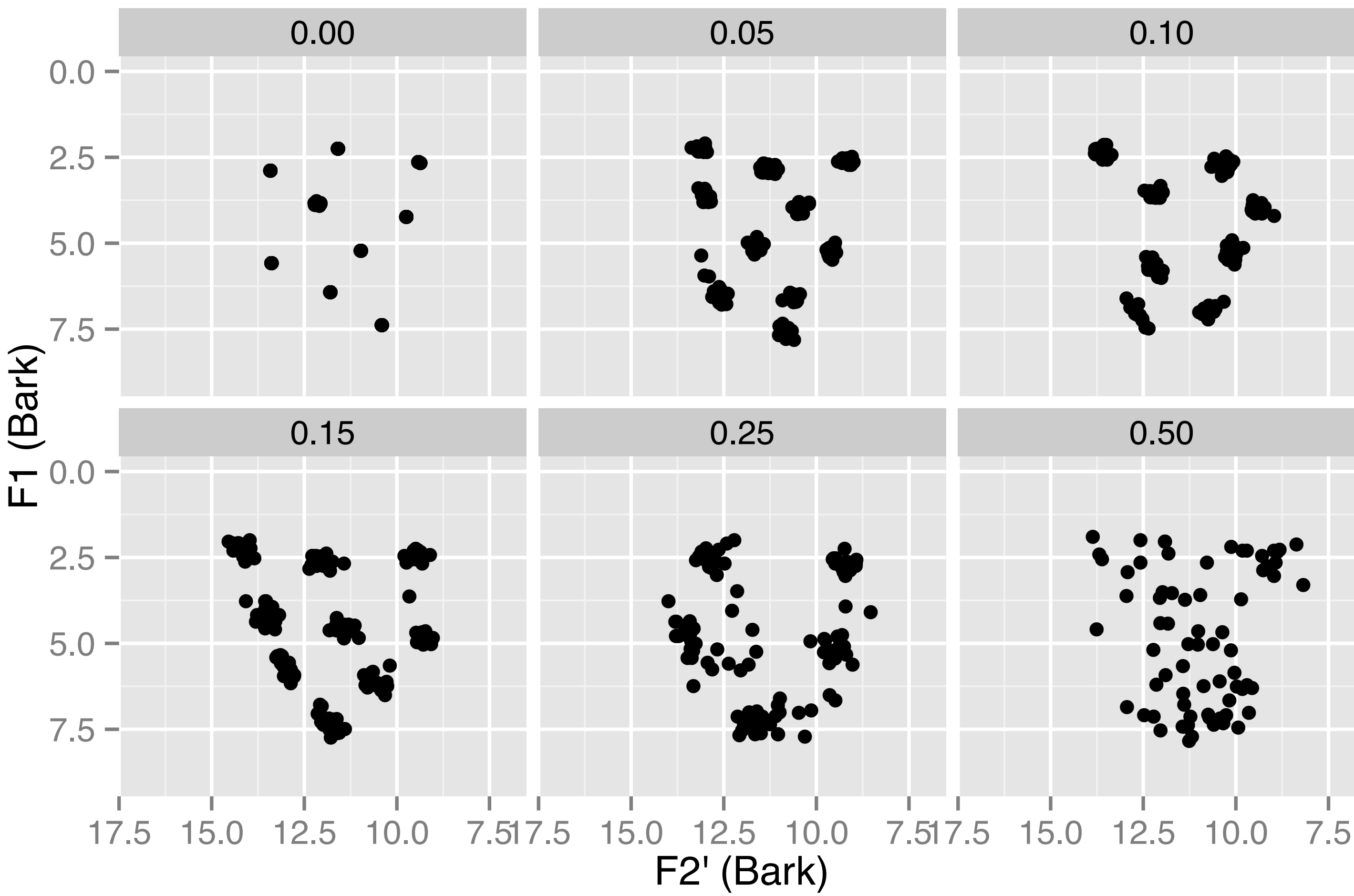
- Size: 7.1 ± 0.8
- Energy: 5.8 ± 1.3
- Success: 0.94 ± 0.02

Varying parameters

- "The question remains whether these results are due to fine-tuning of parameters or whether they are an inevitable result of the interactions between the agents." (456)

- Sensitivity analysis
  - Good modeling practice

Varying the amount of noise



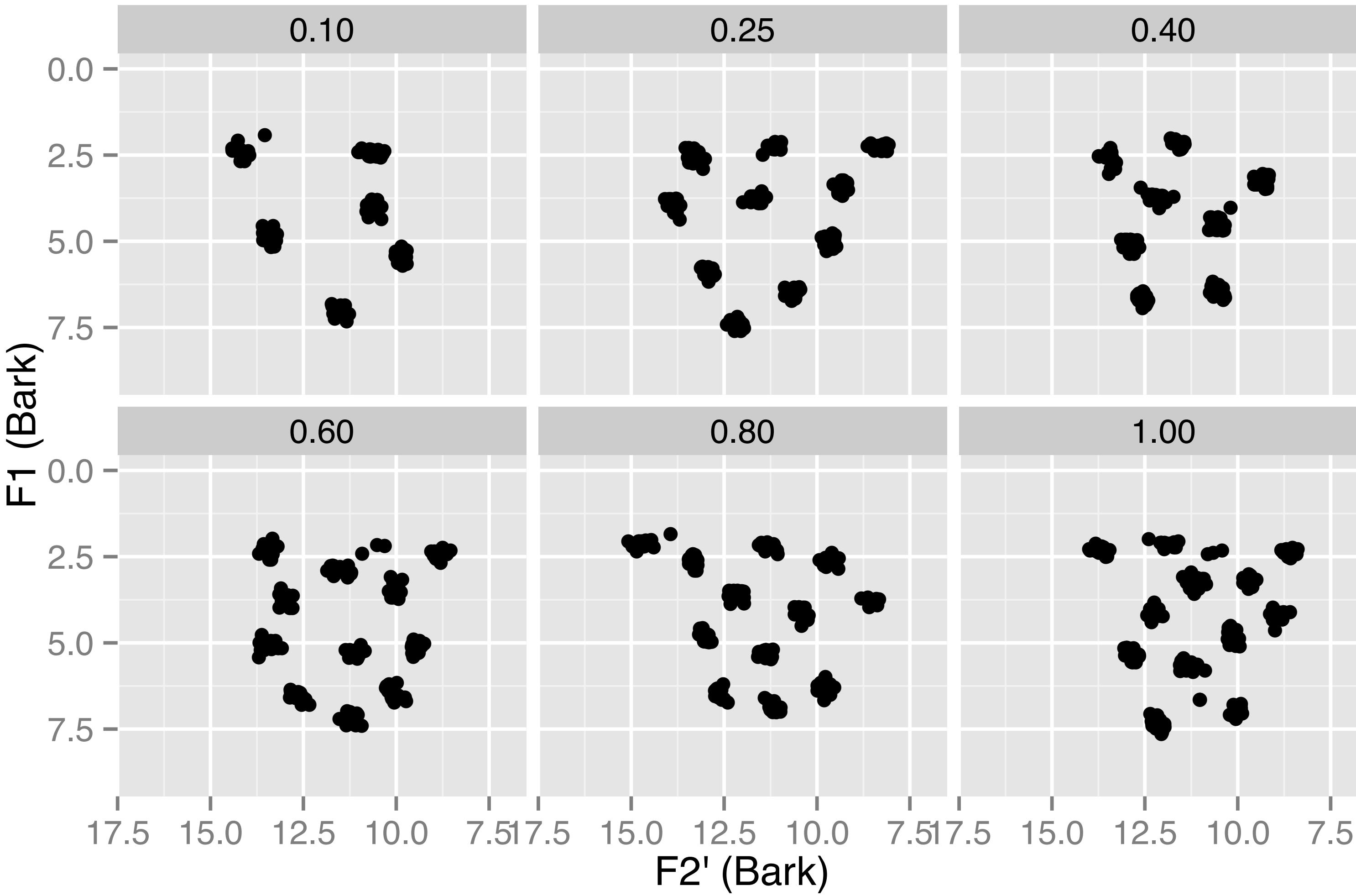
ψ = 0, 0.05, 0.10, 0.15, 0.25, 0.50

- Increased ψ:
  - Success rate: Decreases (0.98-0.74) *
  - Number of vowels: Decreases (9.00-3.60)
  - System coherence across agents: Decreases
  - Energy: Decreases

- Why?
- Which ones look "realistic"?

* Opposite pattern from de Boer. Interaction with another parameter we've set differently? (Or a bug?)

Varying λ



- Increased λ:
  - Success rate: approx. constant
  - Size, energy: Increases
  - More vowel height dimensions, sort of more backness dimensions

- Why?
- Which ones look realistic?

Varying n_{agents}

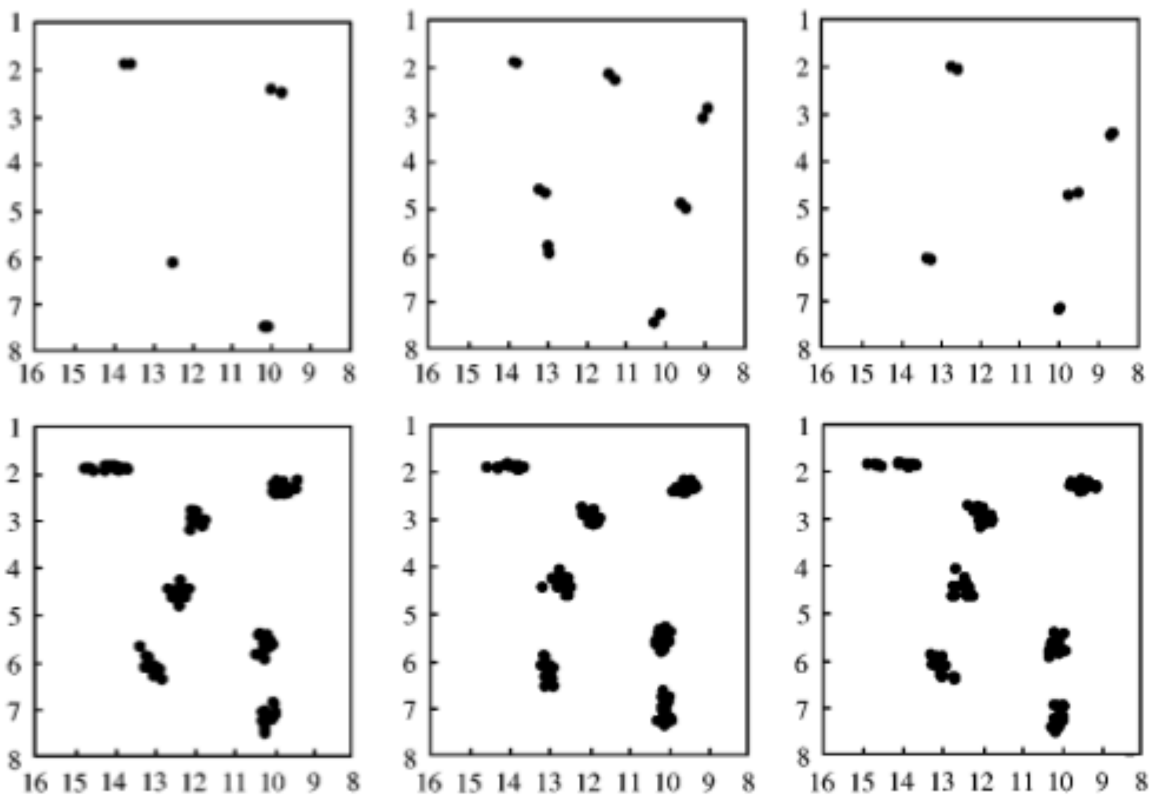
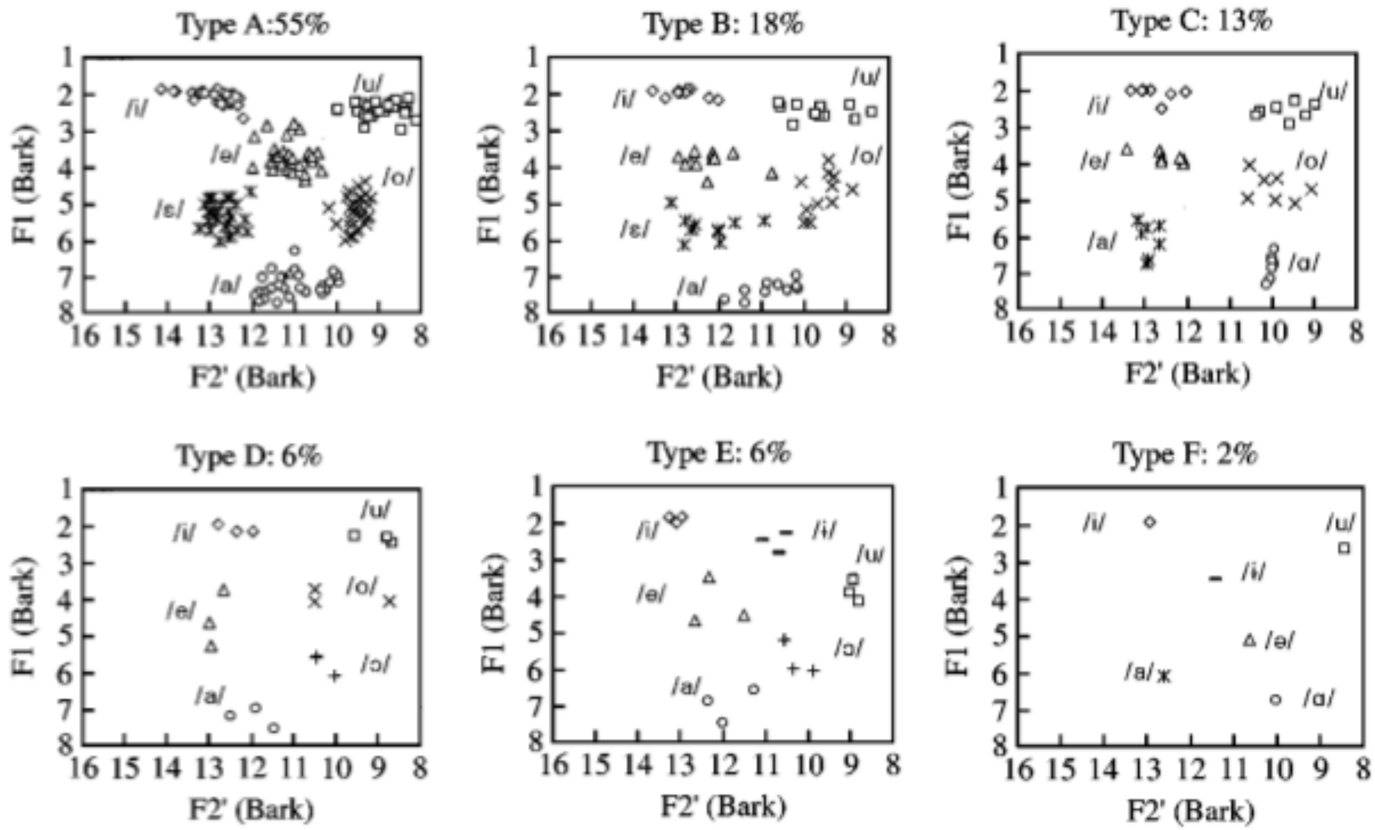


Figure 9. Populations of 2 (top) and 20 (bottom) agents shown at 250, 350 and 450 games per agent from left to right (F'_2 in Bark on x-axis and F_1 in Bark on y-axis). Note the relative instability over time of the vowel system in the small population.

- Increased n_{agents}:
  - Success rate: approx. constant
  - Size, energy: increase (a little)
  - Agents converge, but prototypes move around less

- Why?

Comparisons to empirical data



- Empirical: 43% Type A, 20% Type B, 7% Type C
- Sensitivity to parameters?

- Other predictions within same n:
  - n=3 : Predicts too many "vertical" systems
  - n=4, 5, 7: Very good
  - n=8+ : Worse (c.f. Liljencrants & Lindblom)

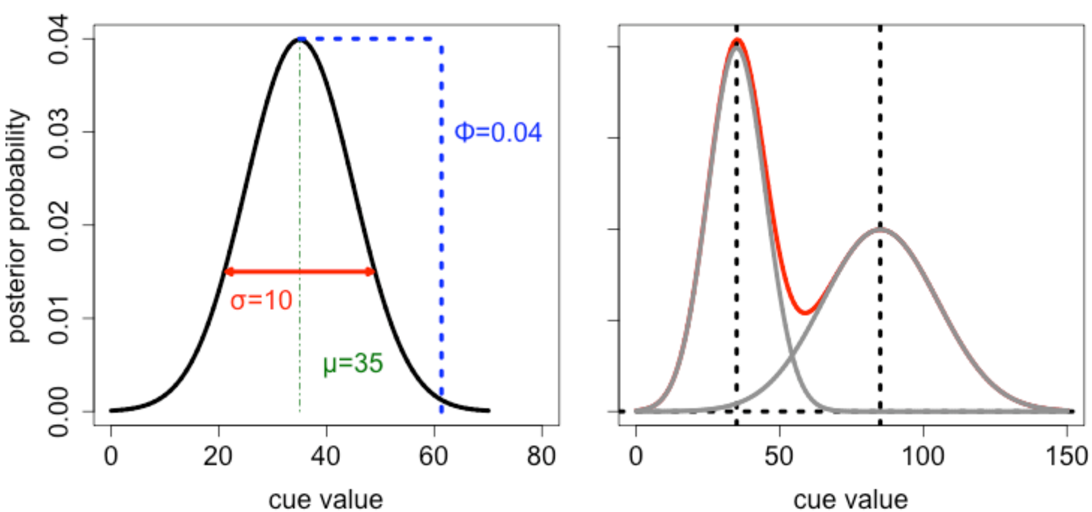
- Predictions of relative frequency of different n:
  - Not so clear how to calculate
  - Under one method (Fig 11), reasonable fit, but lower mean N predicted than observed.

What have we learned?

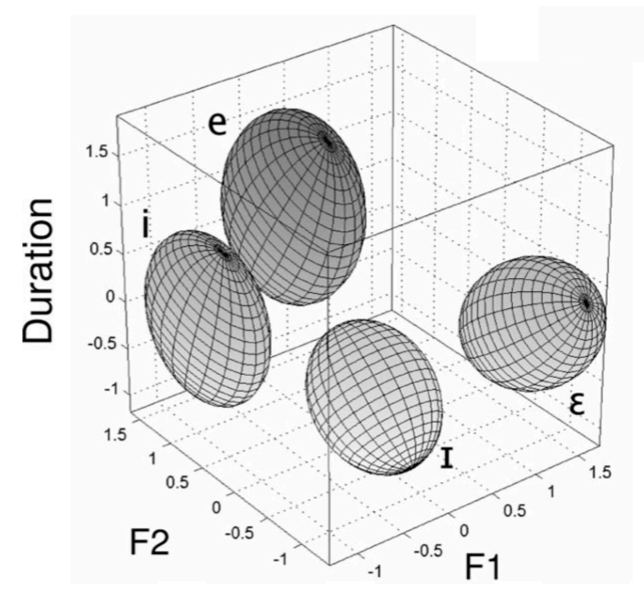
- Existence proofs?
- Explicitness / implementation?
- Counterintuitive results?
- Qualitative predictions?
- Baseline?

Multidimensionality

- de Boer's model: reduces vowels to 2 dimension; considers Euclidean distance
- What is λ in a higher-dimensional space?

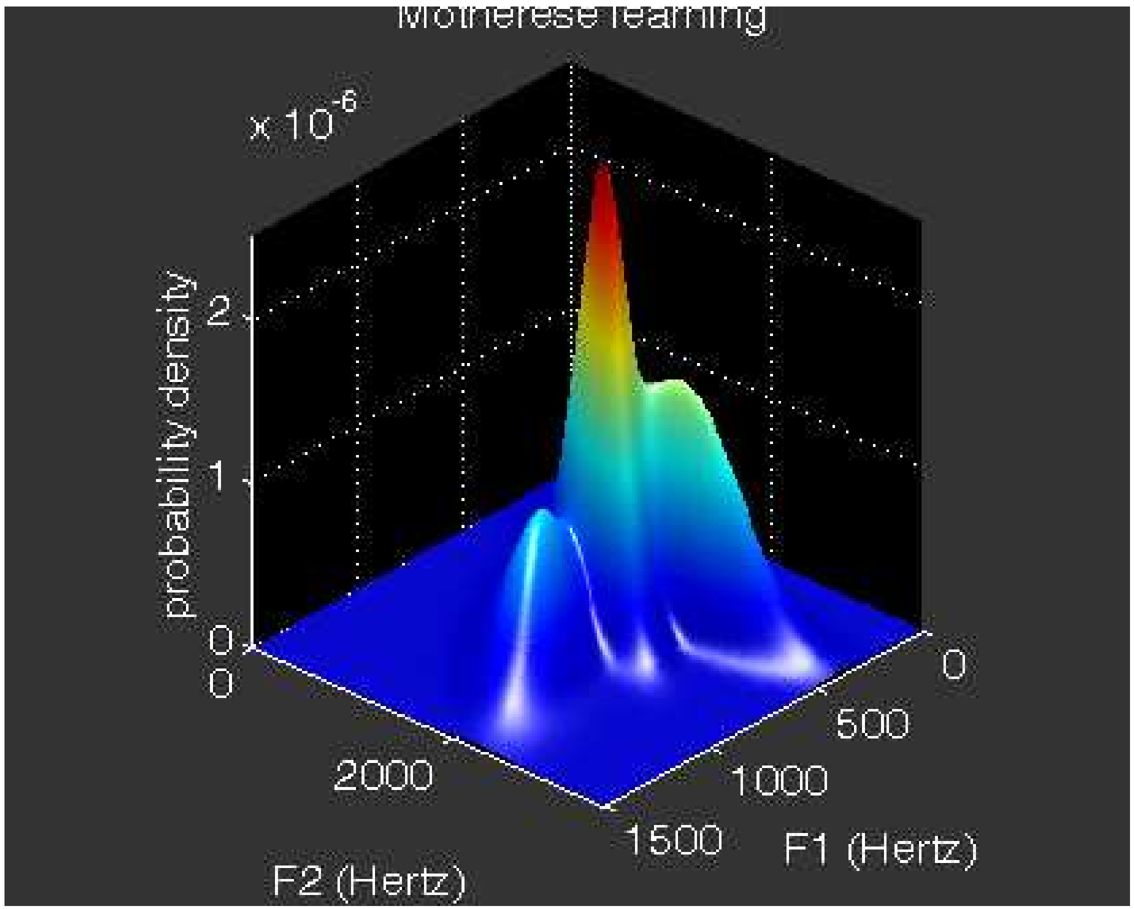


(Clayards, 2008; Kirby, 2013)



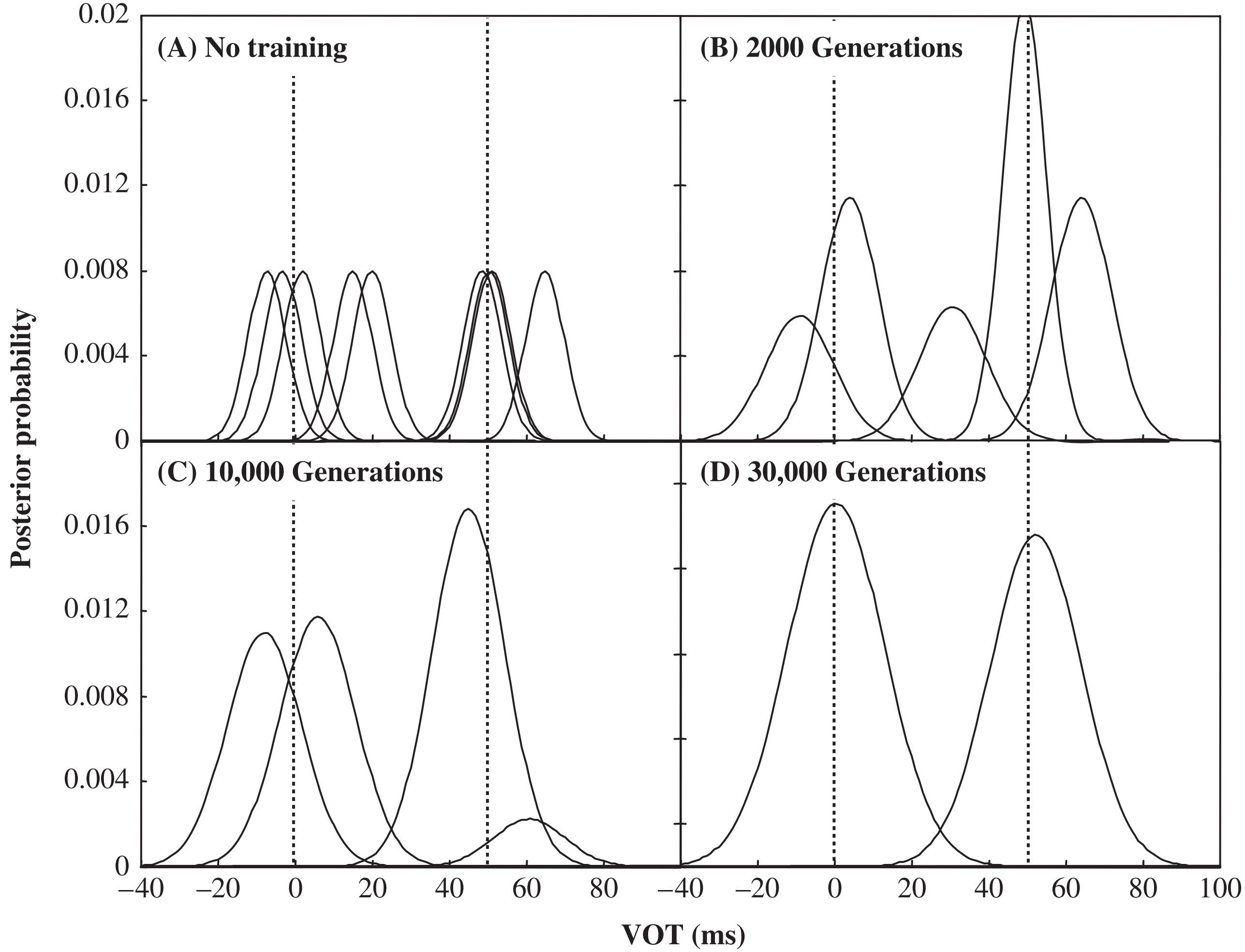
(Vallabha et al., 2007)

Where do categories come from?



(de Boer & Kuhl, 2003)

Where do categories come from?



(McMurray et al., 2009; cf. Vallabha et al., 2007; Kirby, 2011; Dillon et al., 2013; all of unsupervised ASR…)

Extra

Articulatory representation

- [height, back, round] $\in [0,1]^3$

- No noise in production

Vowel	*p*	*h*	*r*
[a]	0	0	0
[ɶ]*	0	0	1
[ɐ]	0.5	0	0
[ɐ̹]*	0.5	0	1
[ɑ]	1	0	0
[ɒ]	1	0	1
[e]	0	0.5	0
[ø]	0	0.5	1
[ə]	0.5	0.5	0
[e]	0.5	0.5	1
[ɤ]	1	0.5	0
[o]	1	0.5	1
[i]	0	1	0
[y]	0	1	1
[ɨ]	0.5	1	0
[ʉ]	0.5	1	1
[ɯ]	1	1	0
[u]	1	1	1

Articulatory synthesizer

- Step 1: Articulation → F1-F4

$$
\begin{aligned}
F_1 = &((-392 + 392r)h^2 + (596 - 668r)h + (-146 + 166r))p^2 \\
&+ ((348 - 348r)h^2 + (-494 + 606r)h + (141 - 175r))p \\
&+ ((340 - 72r)h^2 + (-796 + 108r)h + (708 - 38r))
\end{aligned}
$$

$$
\begin{aligned}
F_2 = &((-1200 + 1208r)h^2 + (1320 - 1328r)h + (118 - 158r))p^2 \\
&+ ((1864 - 1488r)h^2 + (-2644 + 1510r)h + (-561 + 221r))p \\
&+ ((-670 + 490r)h^2 + (1355 - 697r)h + (1517 - 117r))
\end{aligned}
$$

…

- Step 2: Add noise

drawn uniformly from

$[-\frac{\psi_{ac}}{2}, \frac{\psi_{ac}}{2}]$ parameter

$$F_i = F_i(1 + v_i)$$

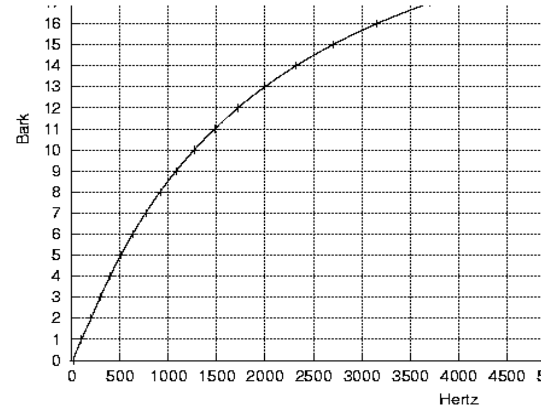
Prototypes

- Vowel *v* stored as prototype (no exemplars!)
  1. Articulatory aspect: [h, b, r]
  2. Acoustic aspect: F1-F4, synthesized

- Also:
  – Success count: s_v
  – Use count: u_v

Perception

1. (F1, F2, F3, F4): Hz → bark



2. (F1, F2, F3, F4) → (F1, F2')
   - F2': Effective second-formant frequency
     (Mantakas, Schwartz, Escudier, 1986)
   - Roughly: Weight formant peaks closer than critical distance *c* (3.5 bark)

- Formants of higher frequency are "blurred"

Perceptual distance

$$D = \sqrt{(F_1^a - F_1^b)^2 + \lambda(F_2^{a'} - F_2^{b'})^2}$$

relative weight of higher formants
w.r.t. F1

- de Boer: λ = 0.3 (cites previous work)
- Interpretation: How much vowel space "stretched" horizontally vs. vertically

Model details: Imitation game

- *nAgents* : # of agents
- *nIts* : # of interactions

1. initialize agents with empty inventories

2. do *nIts* interactions:
   - pick two agents at random
   - agents interact

TABLE III. Basic organization of the imitation game

	Imitator	Imitator	
1	If (V = Ø) Add random vowel to V Pick random v vowel v from V $u_v \leftarrow u_v + 1$ *Produce signal* A_1: $A_1 \leftarrow ac_v + noise$		
		Receive signal A_1 If ($V = \emptyset$) **Find phoneme** (v_{news}, A_1) $V \leftarrow V \cup v_{new}$ Cacalculate v_{rec}: $v_{rec} \in V \wedge \neg\exists v_2 : (v_2 \in V \wedge D(A_1, ac_{v2}) < D(A_1, ac_{vrec}))$ Produce signal A_2: $A_2 \leftarrow ac_{vrec} + noise$	2

3	Receive signal A_2.
	Calculate v_{rec}:
	$v_{rec} \in V \wedge \neg\exists v_2 : (v_2 \in V \wedge D(A_2, ac_{v2}) < D(A_2, ac_{vrec}))$
	If ($v_{rec} = v$)
	Send nonverbal feedback:
	success.
	$s_v \leftarrow s_v + 1$
	Else
	Send nonverbal feedback:
	failure.		

		Receive nonverbal f **Update** V **according** **feedback signal.**	
5	**Do other updates of** V.	**Do other updates of**	5

Step 1: Agent 1 (production)

Step 2: Agent 2 (perception, production)

Step 3: Agent 1 (perception, send feedback)

Step 4: Agent 2 (receive feedback; update vowel system)

Step 5: Agents 1 & 2 (update vowel system)

Shift closer (v, A)	**Find phoneme** (v_{new}, A)	**Update according to feedback signal**
$v_{best} \leftarrow v$	$ar_v \leftarrow (0.5, 0.5, 0.5)$	$v_{vrec} \leftarrow u_{vrec} + 1$
For (all six neighbors v_{neigh} of v) do:	$ac_v \leftarrow S(ar_v)$	If (feedback signal = *success*)
If ($D(ac_{vneigh}, A) <$	$s_v \leftarrow 0$	**Shift closer** (v_{rec}, A_1)
$D(ac_{vrec}, A)$)		
$v_{best} \leftarrow v_{neigh}$	$u_v \leftarrow 0$	$s_{vrec} \leftarrow s_{vrec} + 1$
$v \leftarrow v_{best}$	Do	Else
	$v_{new} \leftarrow v$	If ($u_{vrec}/s_{vrec} > threshold$)
	Shift closer (v_{new}, A)	**Find phoneme**
		(v_{new}, A_1)
	Until ($v = v_{new}$)	$V \leftarrow V \cup v_{new}$
		Else
		Shift closer (v_{rec}, A_1)

ε_{artic} away, 6 directions

Shift closer (v, A)	**Find phoneme** (v_{new}, A)	**Update according to feedback signal**
$v_{best} \leftarrow v$	$ar_v \leftarrow (0.5, 0.5, 0.5)$	$v_{vrec} \leftarrow u_{vrec} + 1$
For (all six neighbors v_{neigh} of v) do:	$ac_v \leftarrow S(ar_v)$	If (feedback signal $=$ *success*)
If ($D(ac_{vneigh}, A) <$	$s_v \leftarrow 0$	**Shift closer** (v_{rec}, A_1)
$D(ac_{vrec}, A)$)		0.5
$v_{best} \leftarrow v_{neigh}$	$u_v \leftarrow 0$	$s_{vrec} \leftarrow s_{vrec} + 1$
$v \leftarrow v_{best}$	Do	Else
	$v_{new} \leftarrow v$	If ($u_{vrec}/s_{vrec} >$ *threshold*)
	Shift closer (v_{new}, A)	**Find phoneme** (v_{new}, A_1)
	Until ($v = v_{new}$)	$V \leftarrow V \cup v_{new}$
		Else
		Shift closer (v_{rec}, A_1)

Step 5

Merge (v_1, v_2, V)
If $(s_{v1}/u_{v1} < s_{v2}/u_{v2})$
  $s_{v2} \leftarrow s_{v2} + s_{v1}$
  $u_{v2} \leftarrow u_{v2} + u_{v1}$
  $V \leftarrow V - v_1$
Else
  $s_{v1} \leftarrow s_{v1} + s_{v2}$
  $u_{v1} \leftarrow u_{v1} + u_{v2}$
$V \leftarrow V - v_2$

Do other updates of V
For $(\forall v \in V)$//Remove bad vowels
  if $(s_v/u_v < \textit{throwaway threshold} \wedge u_v > \textit{min. uses})$
    $V \leftarrow V - v$
For $(\forall v_1 \in V)$//Merging of vowels
  For $(\forall v_2 : (v_2 \in V \wedge v_2 \neq v_1))$
    If $(D(ac_{v1}, ac_{v2}) < \textit{acoustic merge threshold})$
      Merge (v_1, v_2, V)
    If (Euclidean distance between av_{v1} and $av_{v2} < \textit{articulatory merge threshold}$)
      Merge(v_1, v_2, V)
Add new vowel to V with small probability.

Step 5

0.7 5

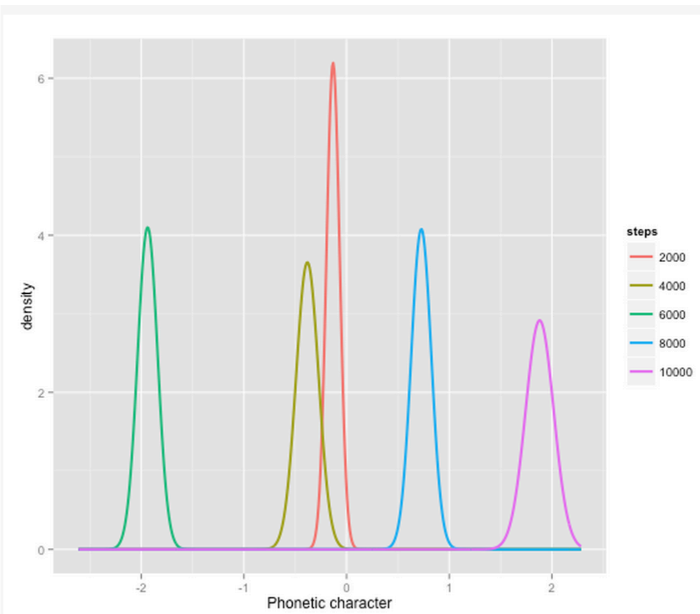
Merge (v_1, v_2, V)
If $(s_{v1}/u_{v1} < s_{v2}/u_{v2})$
$s_{v2} \leftarrow s_{v2} + s_{v1}$
$u_{v2} \leftarrow u_{v2} + u_{v1}$
$V \leftarrow V - v_1$
Else
$s_{v1} \leftarrow s_{v1} + s_{v2}$
$u_{v1} \leftarrow u_{v1} + u_{v2}$
$V \leftarrow V - v_2$

Do other updates of V
For $(\forall v \in V)$//Remove bad vowels
if $(s_v/u_v <$ *throwaway threshold* $\wedge\ u_v >$ *min. uses*$)$
$V \leftarrow V - v$
For $(\forall v_1 \in V)$//Merging of vowels
For $(\forall v_2{:}(v_2 \in V \wedge v_2 \neq v_1))$??
If $(D(ac_{v1}, ac_{v2}) <$ *acoustic merge threshold*$)$ 0.17
Merge (v_1, v_2, V)
If (Euclidean distance between av_{v1} and $av_{v2} <$ *articulatory merge threshold*)
Merge(v_1, v_2, V)
Add new vowel to V with small probability. 0.01

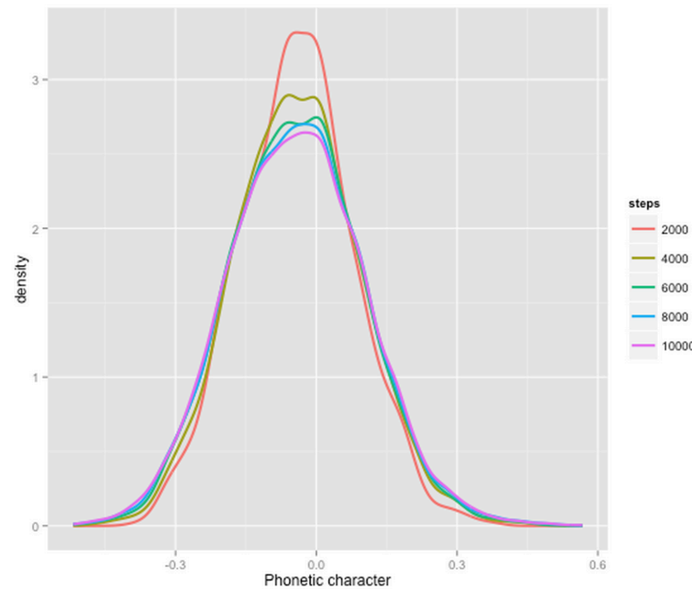
Varying τ

- Lower τ ⇒ more randomness

0.1



10000



- Need high enough decay time for stable category
- Qualitative outcome: similar?