Using Bayesian Approaches to Study Human Sound and Word Learning

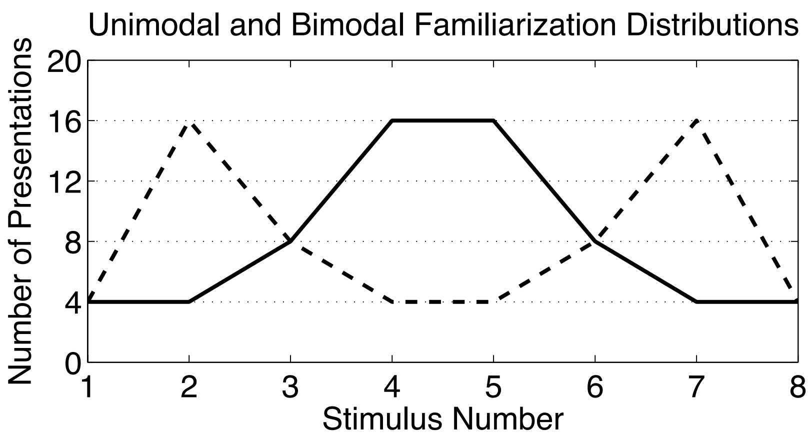
Naomi Feldman
University of Maryland

CLSP Miniworkshop:
Zero Resource Speech Technologies
and Models of Early Language Acquisition
July 17, 2012

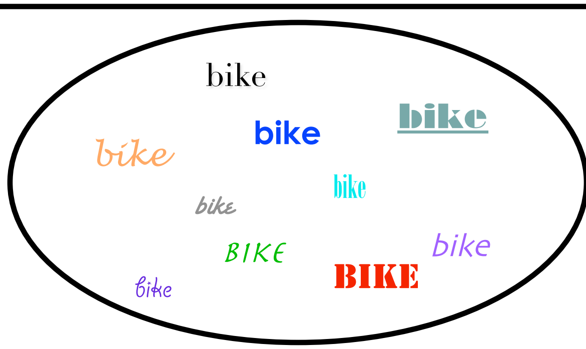
Joint work with Erin Bennett, Sharon Goldwater, Tom Griffiths, Yakov Kronrod, James Morgan, Emily Myers, Katherine White

Language Acquisition

Infants have a machine learning problem to solve…



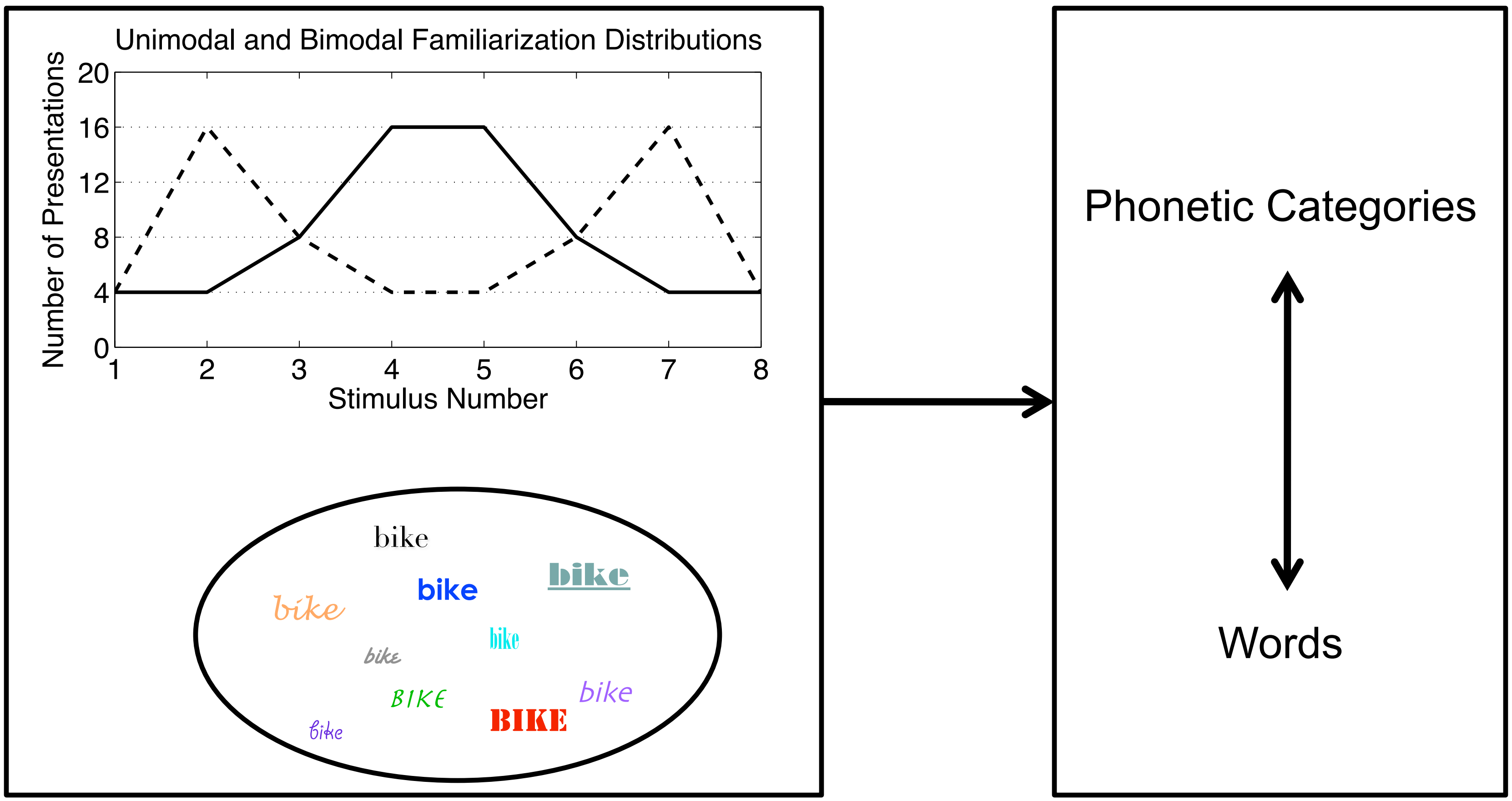
Phonetic Categories



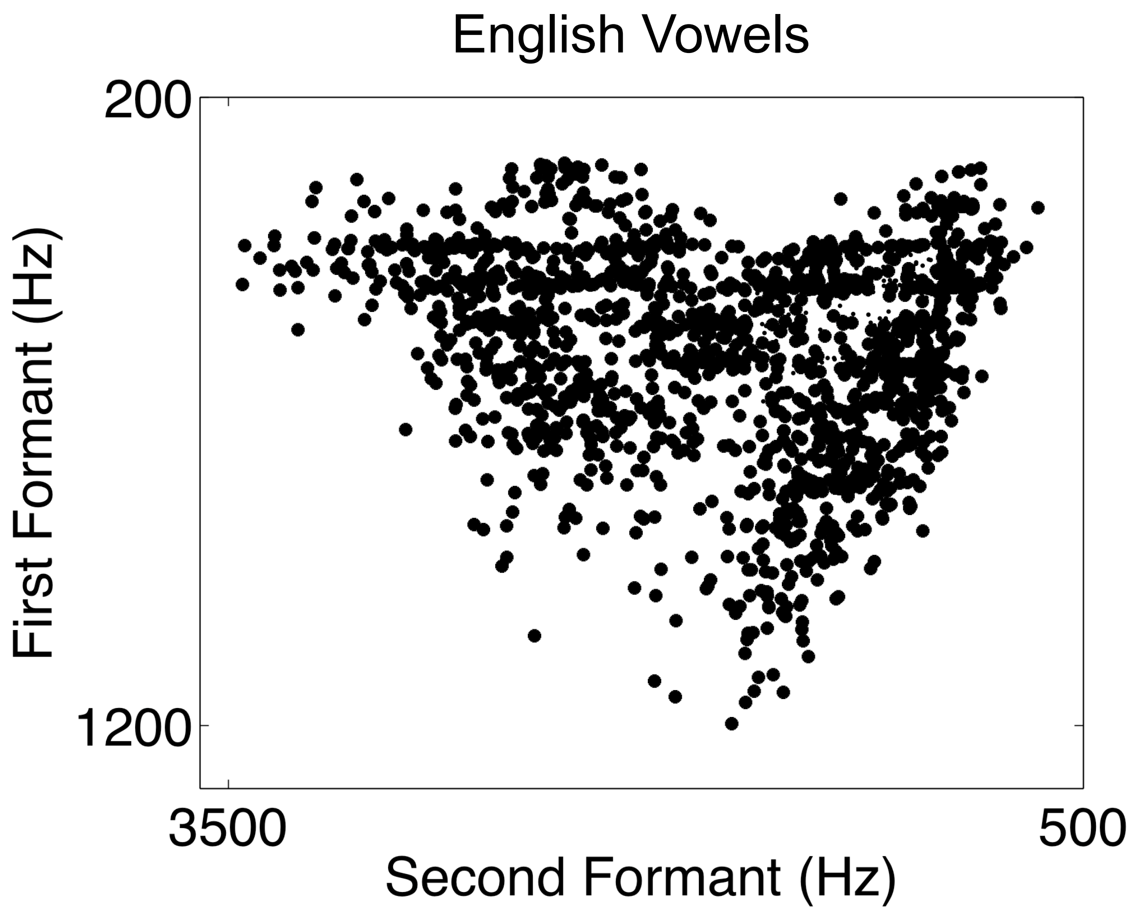
Words

Language Acquisition

Infants have a machine learning problem to solve…

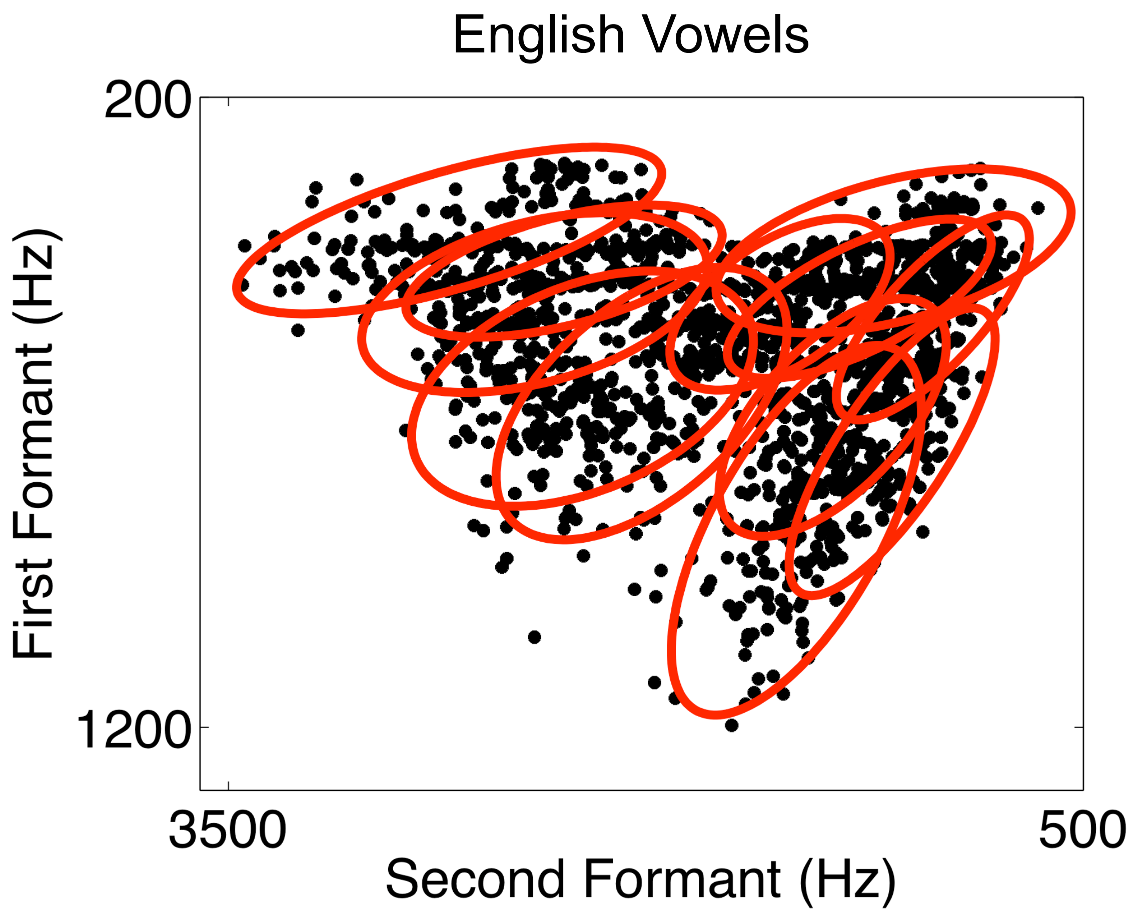


A Difficult Problem



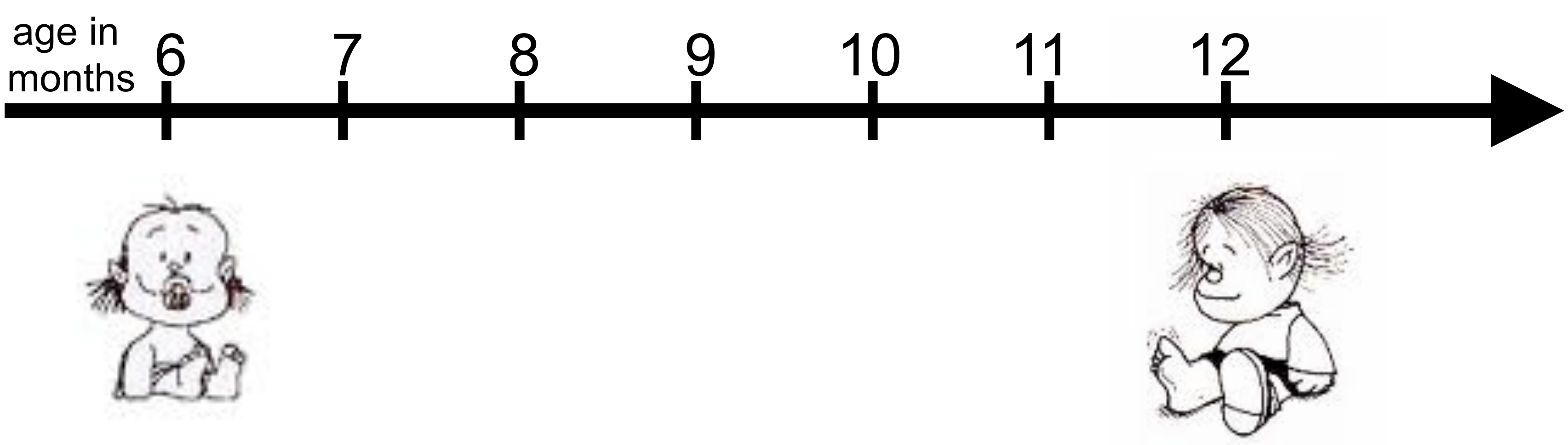
(Hillenbrand, Getty, Clark, & Wheeler, 1995)

A Difficult Problem



(Hillenbrand, Getty, Clark, & Wheeler, 1995)

Phonetic Category Acquisition

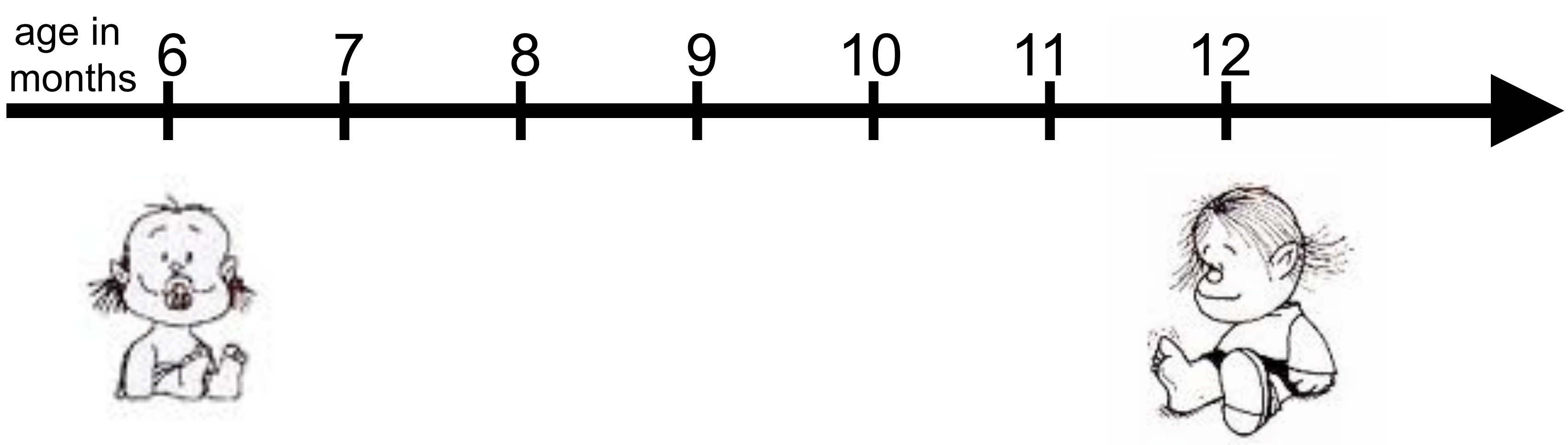


Phonetic Category Acquisition

6-8 months:
discriminate non-native consonant contrasts

10-12 months:
poor discrimination of non-native consonant contrasts

(Werker & Tees, 1984)



Phonetic Category Acquisition

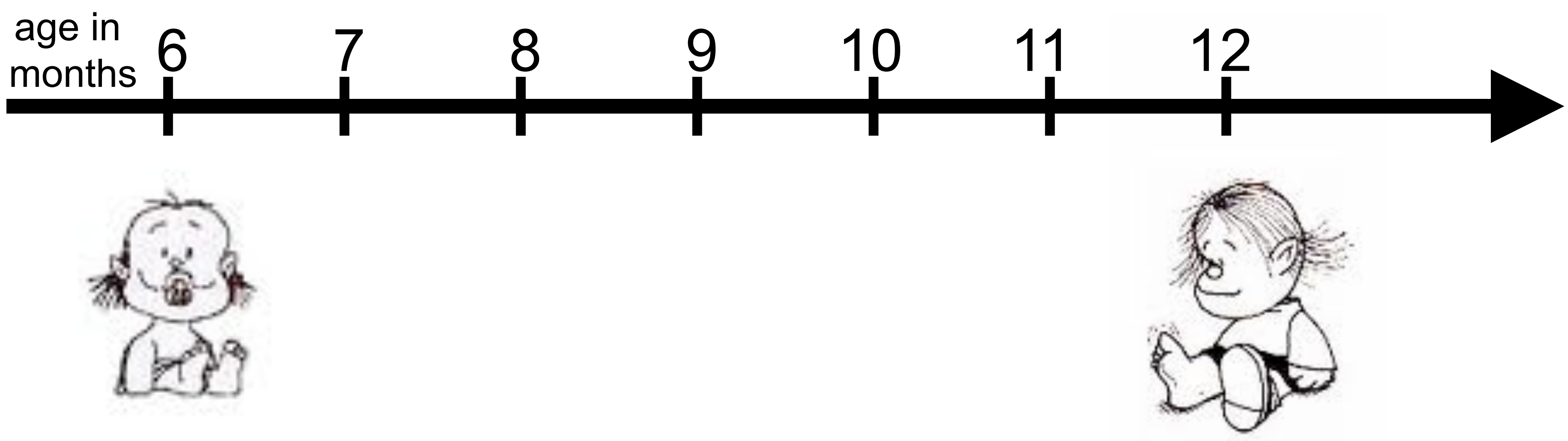
6 months: some language-specific perception of vowels

6-8 months: discriminate non-native consonant contrasts

10-12 months: poor discrimination of non-native consonant contrasts

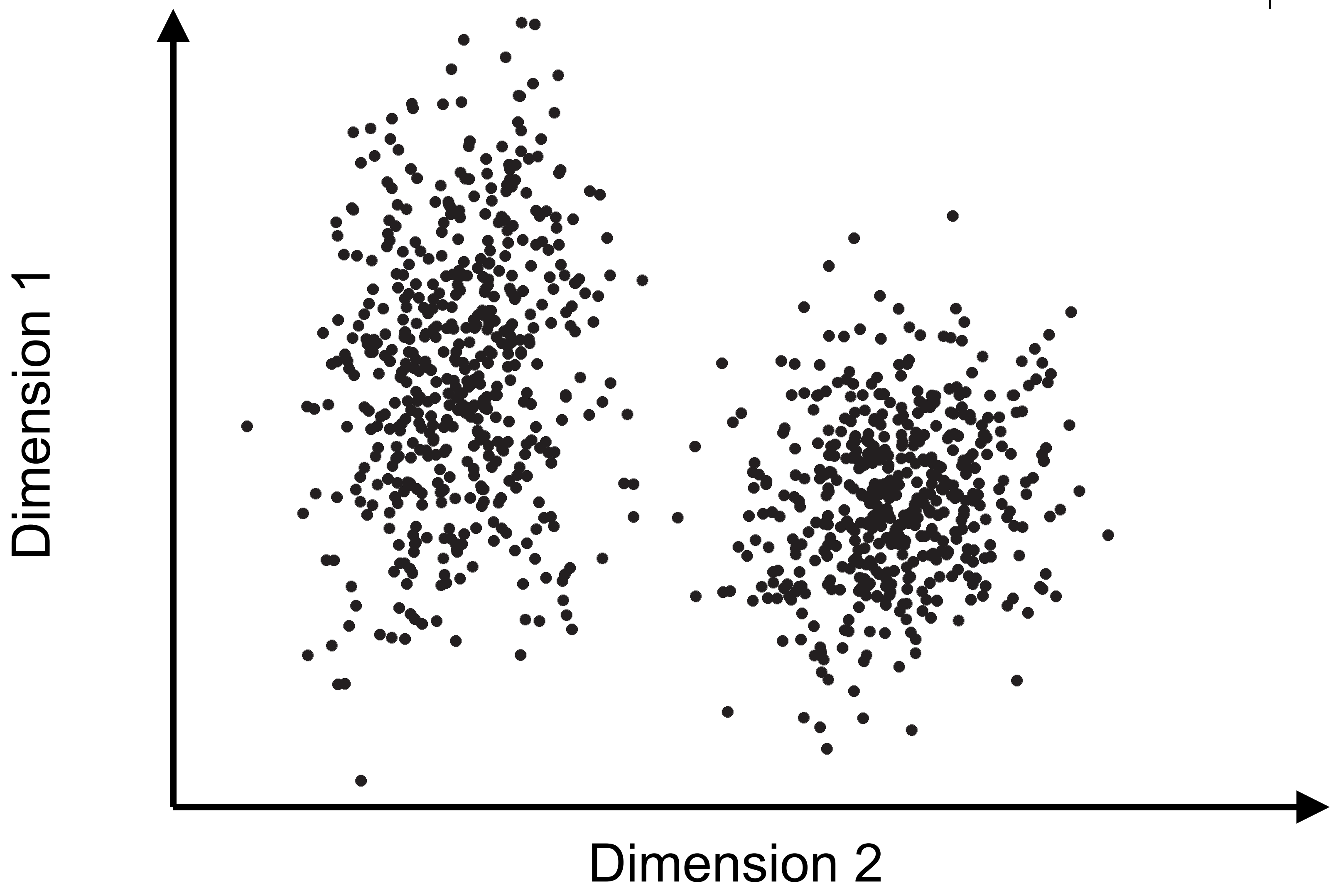
(Werker & Tees, 1984)

(Kuhl et al., 1992)

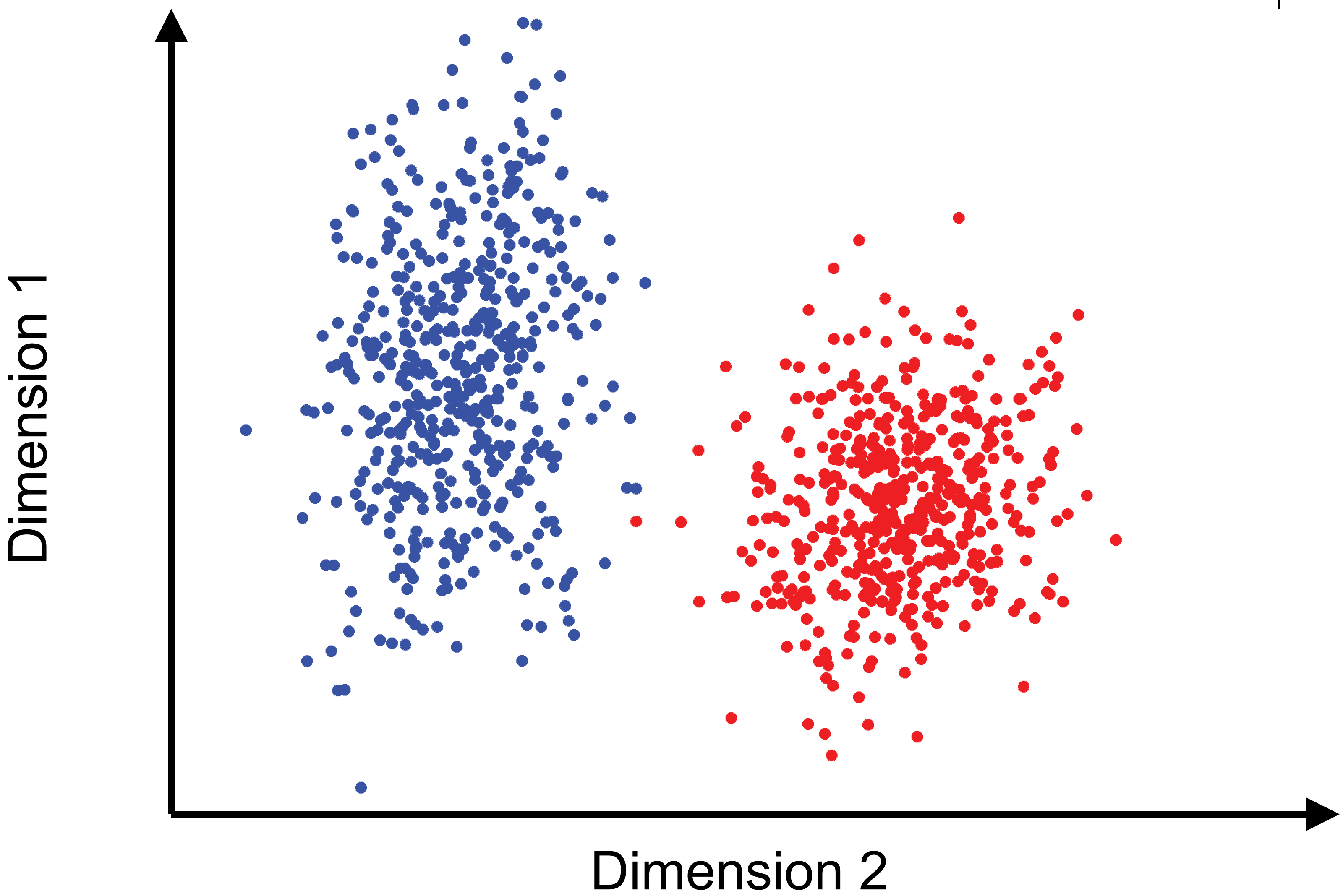


How are sound categories learned?

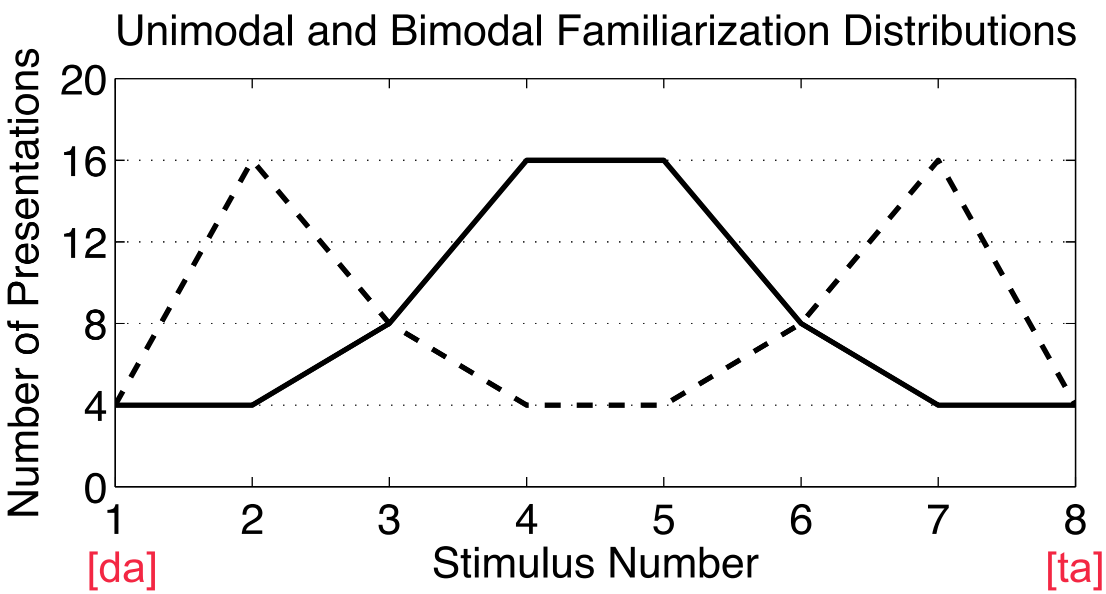
A Clustering Problem



A Clustering Problem

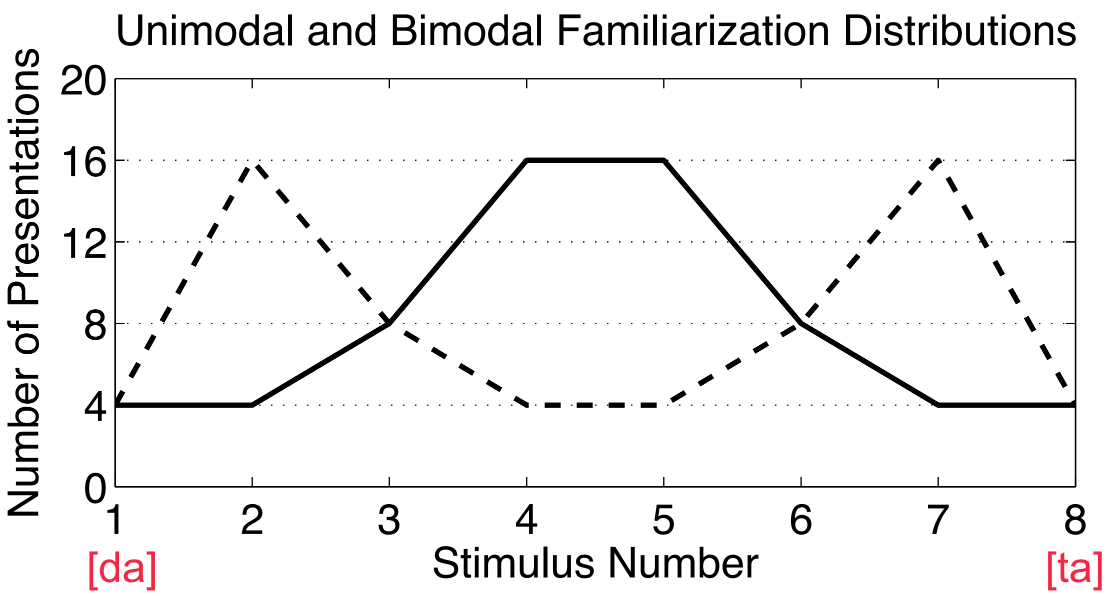


Distributional Learning



(Maye, Werker, & Gerken, 2002)

Distributional Learning



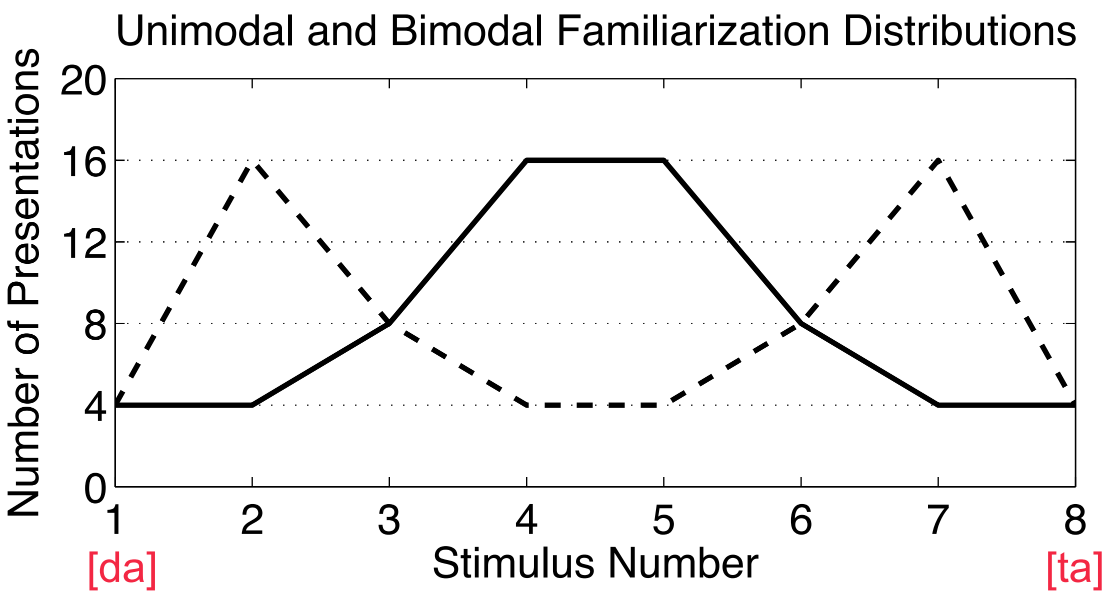
Test stimuli

Alternating trials (da_1-ta_8-da_1-ta_8)

Non-alternating trials (da_3-da_3-da_3-da_3, ta_6-ta_6-ta_6-ta_6)

(Maye, Werker, & Gerken, 2002)

Distributional Learning



Test stimuli

Alternating trials (da_1-ta_8-da_1-ta_8)

Non-alternating trials (da_3-da_3-da_3-da_3, ta_6-ta_6-ta_6-ta_6)

Bimodal ✔ Unimodal ✘

(Maye, Werker, & Gerken, 2002)

A Generative Model

To create a corpus

Phonetic Categories

↓

Corpus

A Generative Model

To create a corpus

1. Generate a phonetic category inventory

Phonetic Categories

↓

Corpus

A Generative Model

<u>To create a corpus</u>

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category

Phonetic Categories

↓

Corpus

A Generative Model

To create a corpus

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a corpus

Phonetic Categories

↓

Corpus

A Generative Model

<u>To create a corpus</u>

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a corpus
   - For each sound, sample a phonetic category according to its frequency

Phonetic Categories

↓

Corpus

A Generative Model

<u>To create a corpus</u>

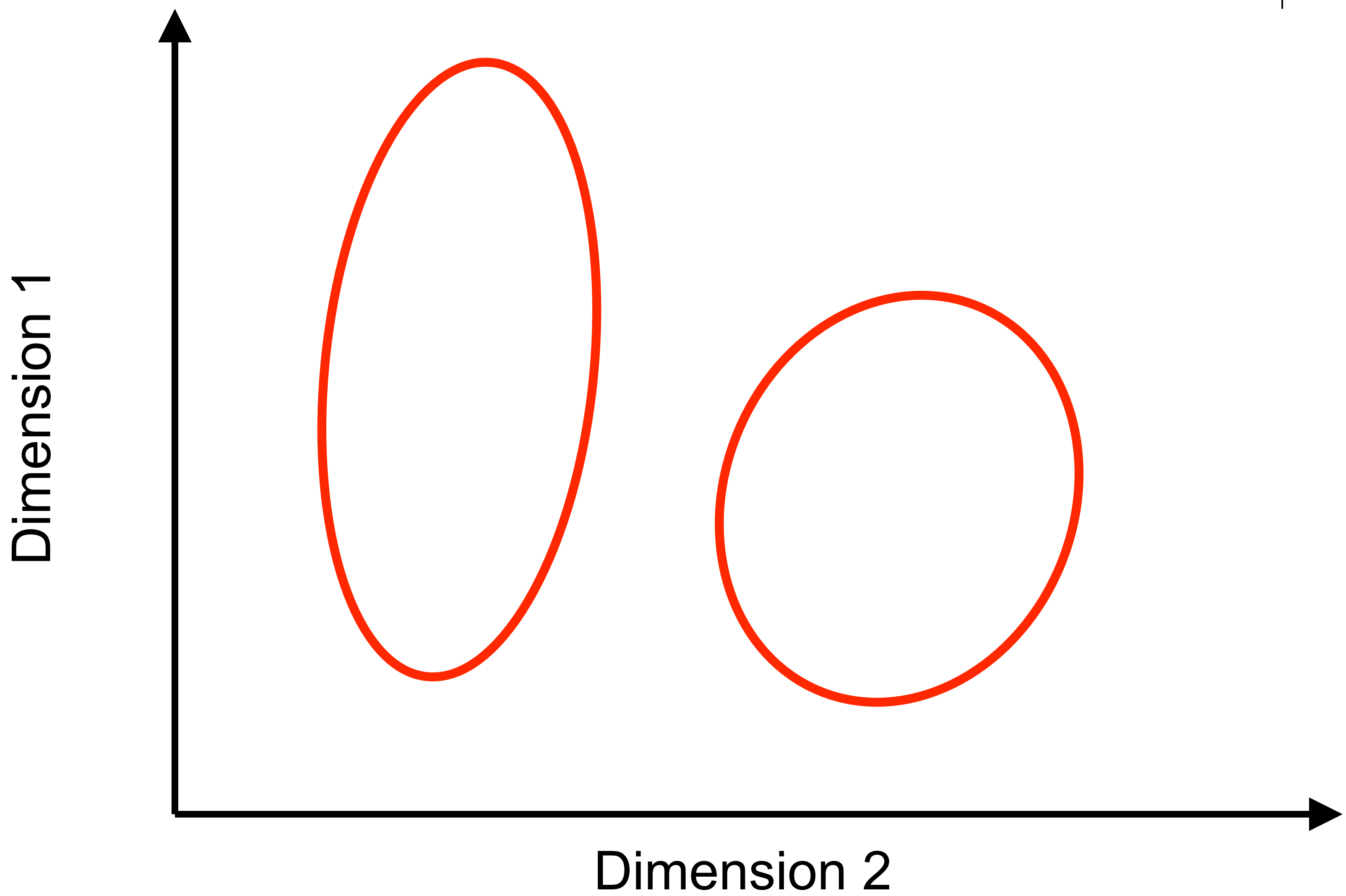
1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a corpus
   - For each sound, sample a phonetic category according to its frequency
   - Generate an acoustic value from the Gaussian distribution associated with that category

Phonetic Categories

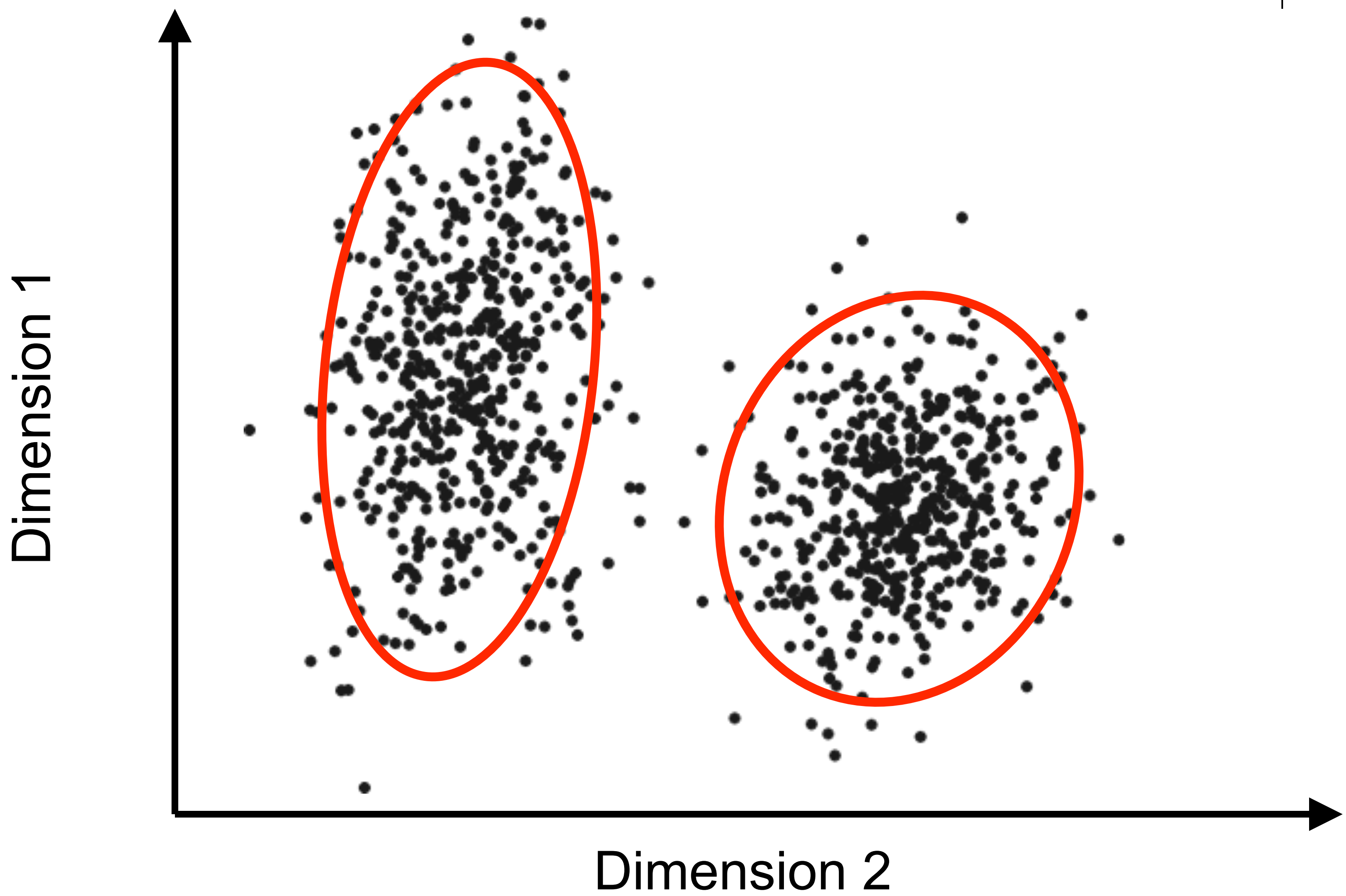
↓

Corpus

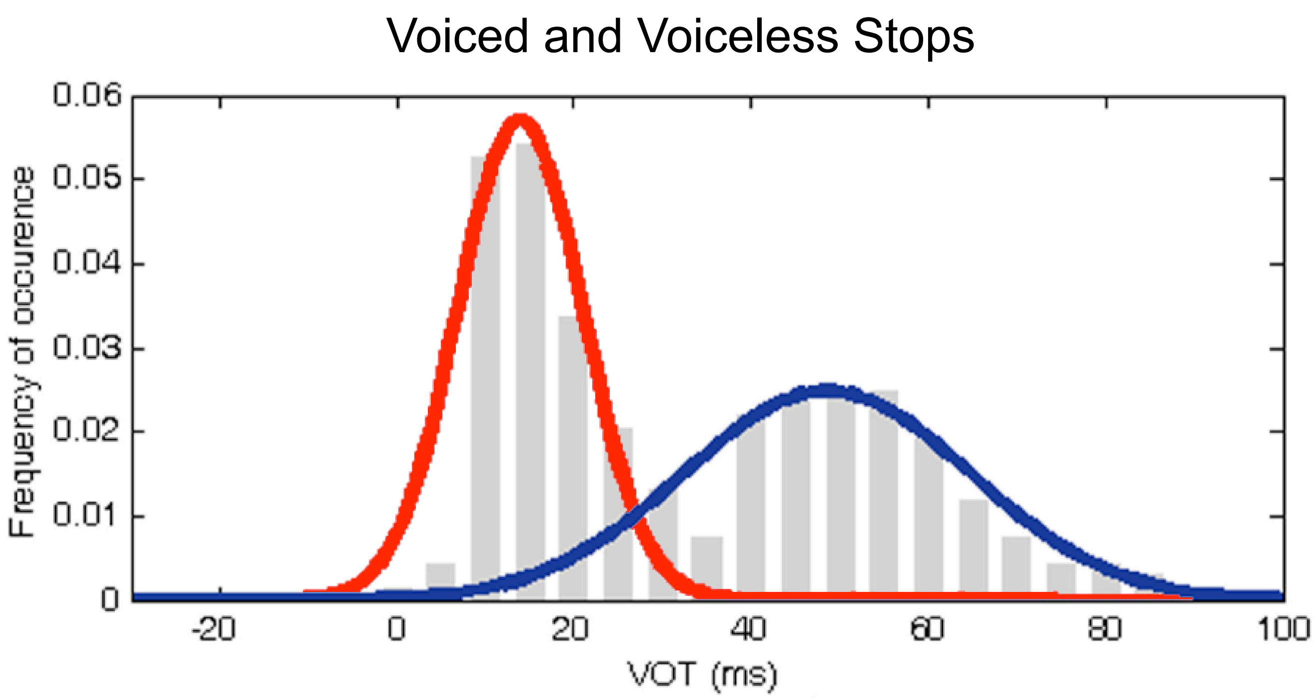
A Generative Model



A Generative Model



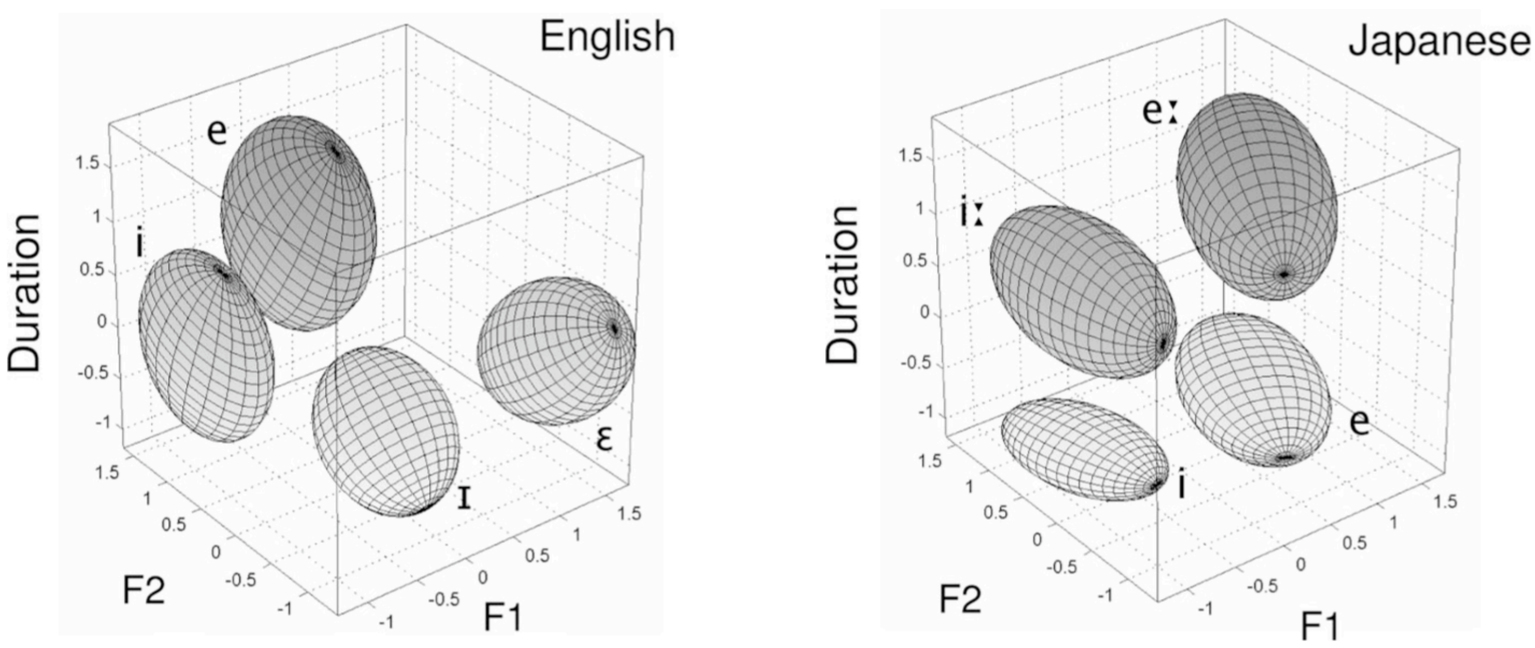
Distributional Learning



(Toscano & McMurray, 2008; McMurray, Aslin, & Toscano, 2009)

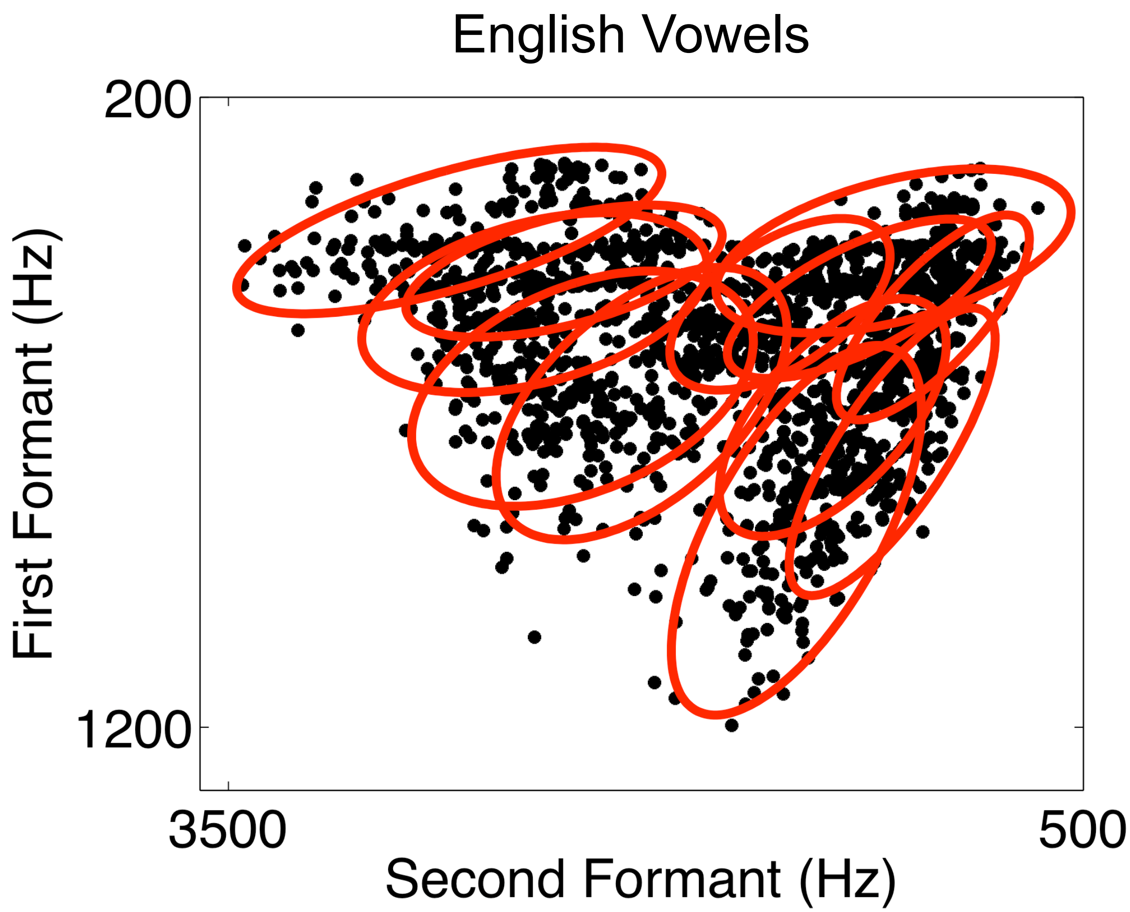
Distributional Learning

Vowel Categories (Single Speakers)



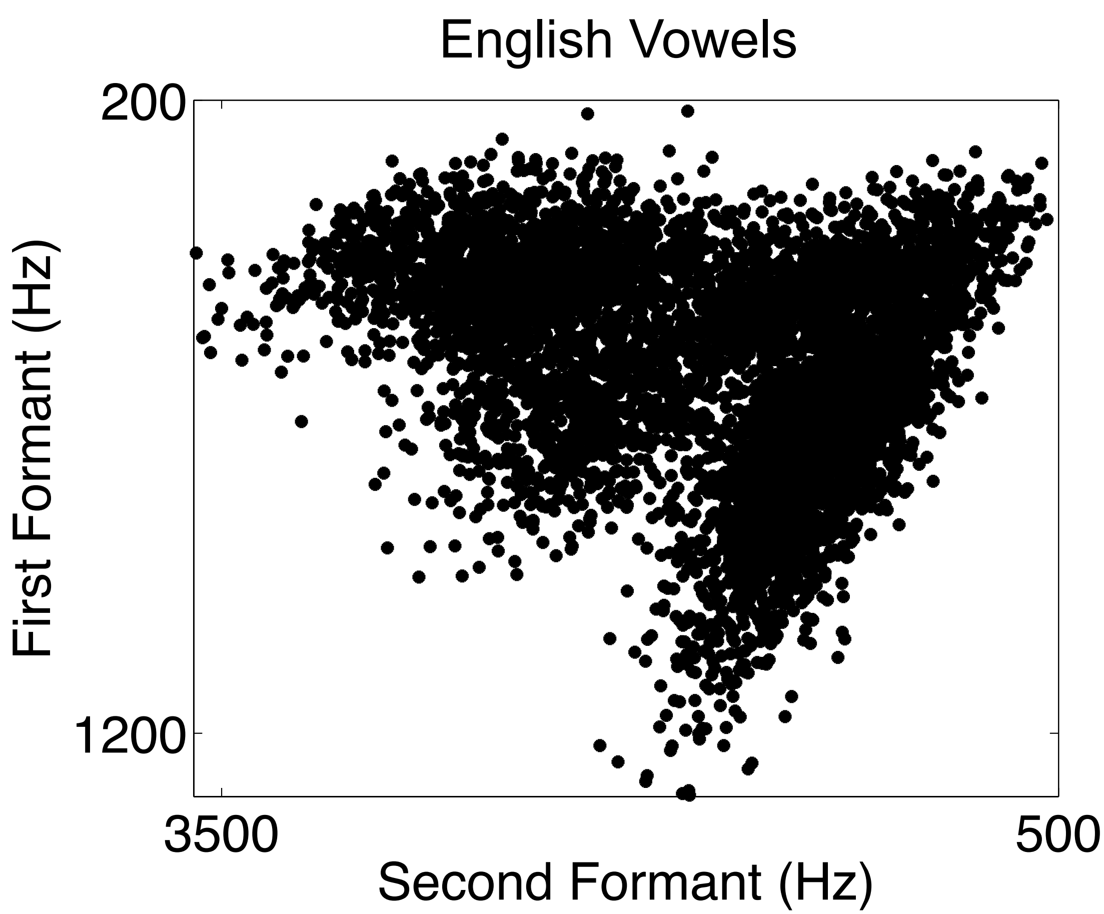
(Vallabha, McClelland, Pons, Werker, & Amano, 2007)

A Difficult Problem



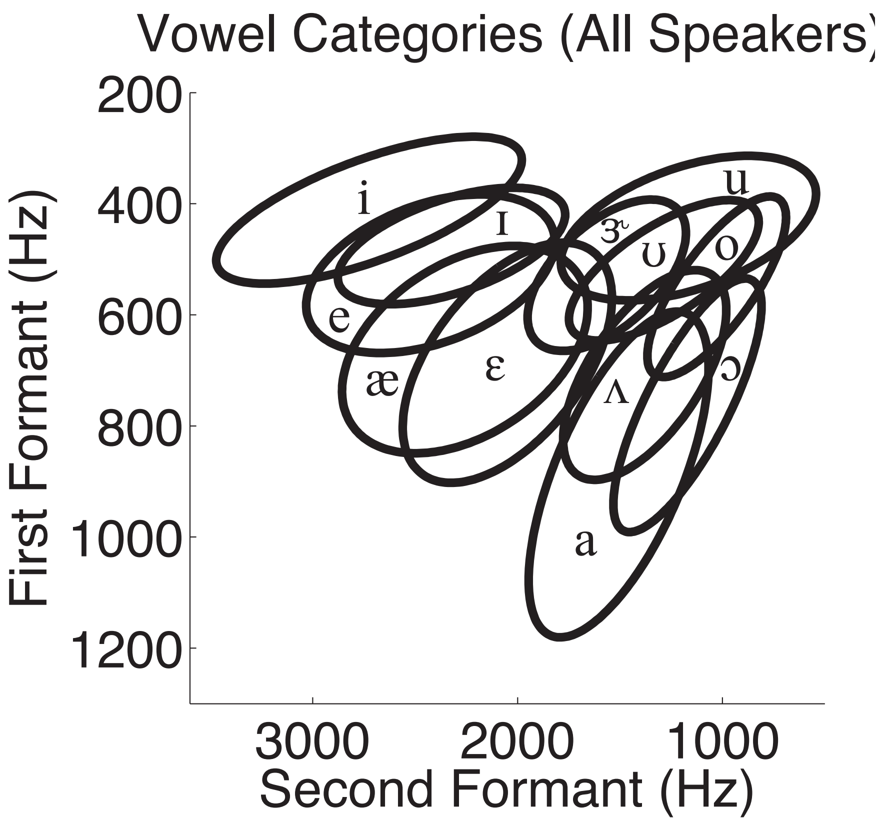
(Hillenbrand, Getty, Clark, & Wheeler, 1995)

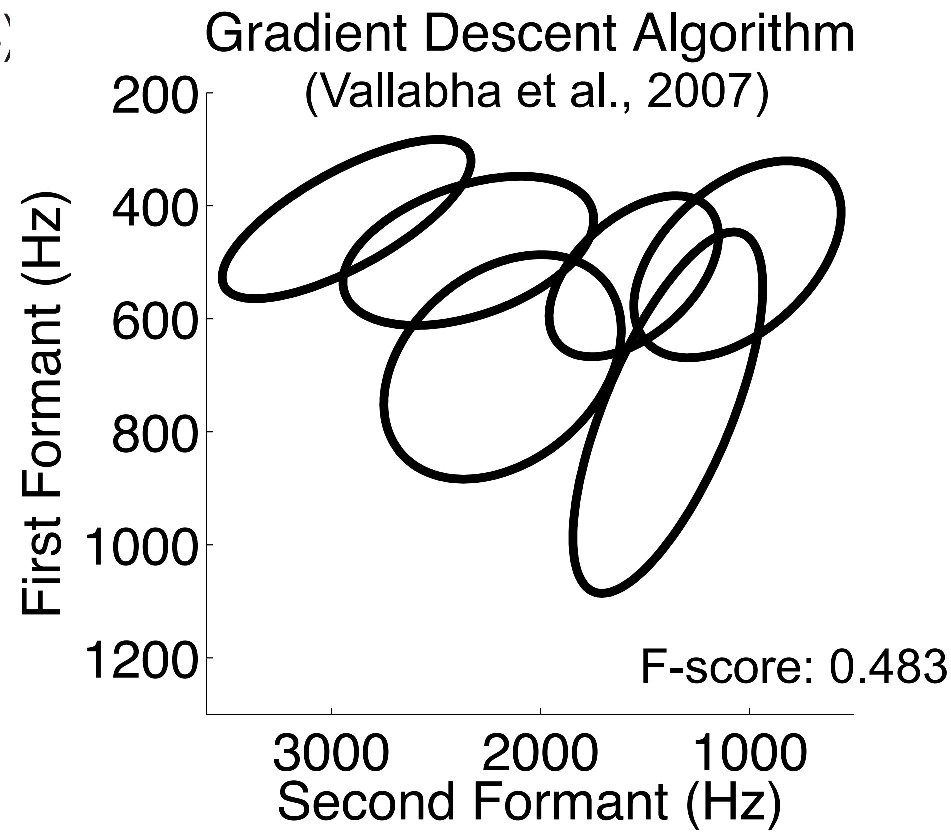
Training Corpus



Corpus of 6,409 vowel tokens generated from Gaussian categories from Hillenbrand et al. (1995); frequencies match corpus frequencies

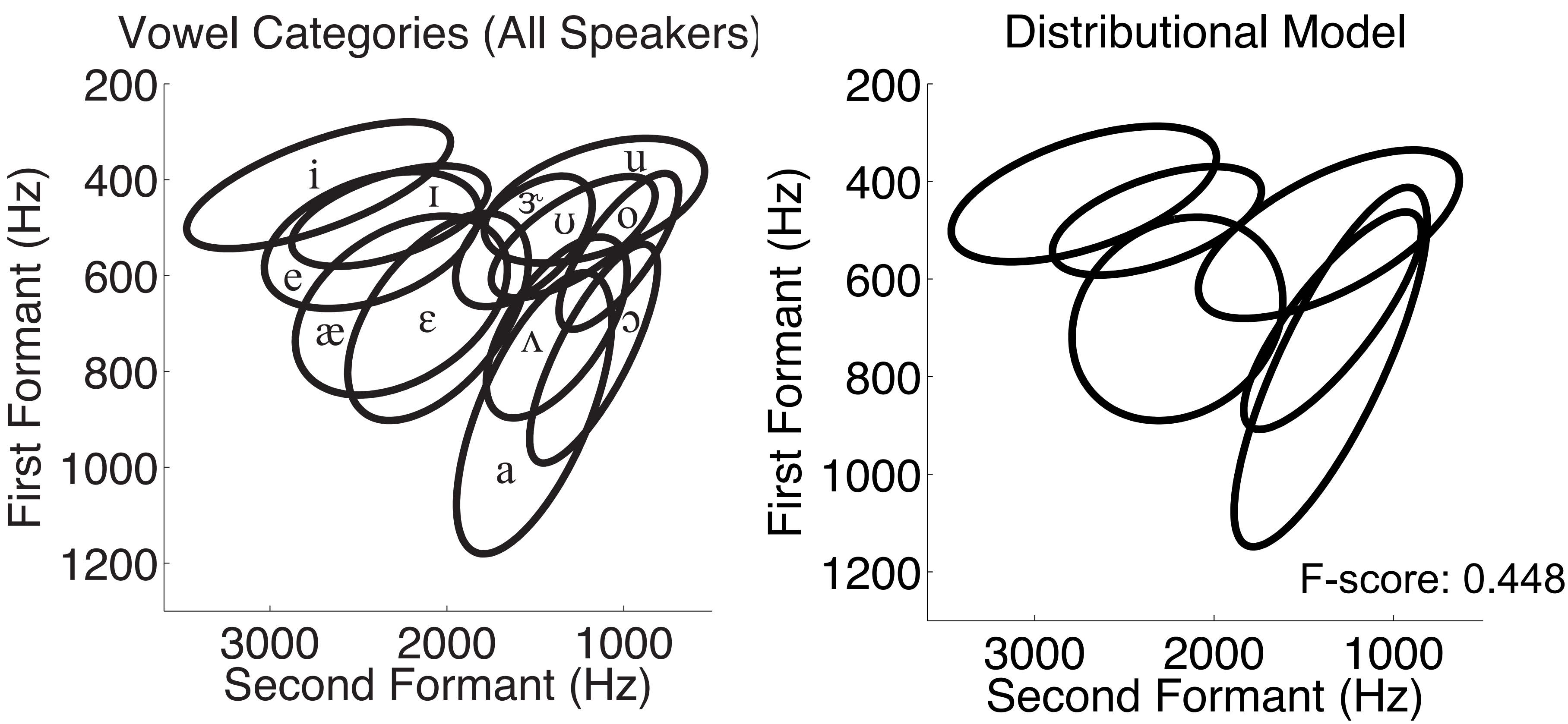
Distributional Learning





(Feldman, Griffiths, Goldwater, & Morgan, in prep; see also Dillon, Dunbar, & Idsardi, in press)

Distributional Learning



(Feldman, Griffiths, Goldwater, & Morgan, in prep; see also Dillon, Dunbar, & Idsardi, in press)

A Generative Model

<u>To create a corpus</u>

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a corpus
   - For each sound, sample a phonetic category according to its frequency
   - Generate an acoustic value from the Gaussian distribution associated with that category

Phonetic Categories

↓

Corpus

Hypothesis

Infants use top-down constraints
from words when acquiring
phonetic categories

(see also Swingley & Aslin, 2007; Swingley, 2009; Thiessen, 2007, 2011)

Word Segmentation Tasks

Familiarization:

Test:

A B

Word Learning

bike BIKE

bike bike

bike bike

bike

bike

bike

BIKE

Word Learning

bike

BIKE

bike

bike

bike

bike

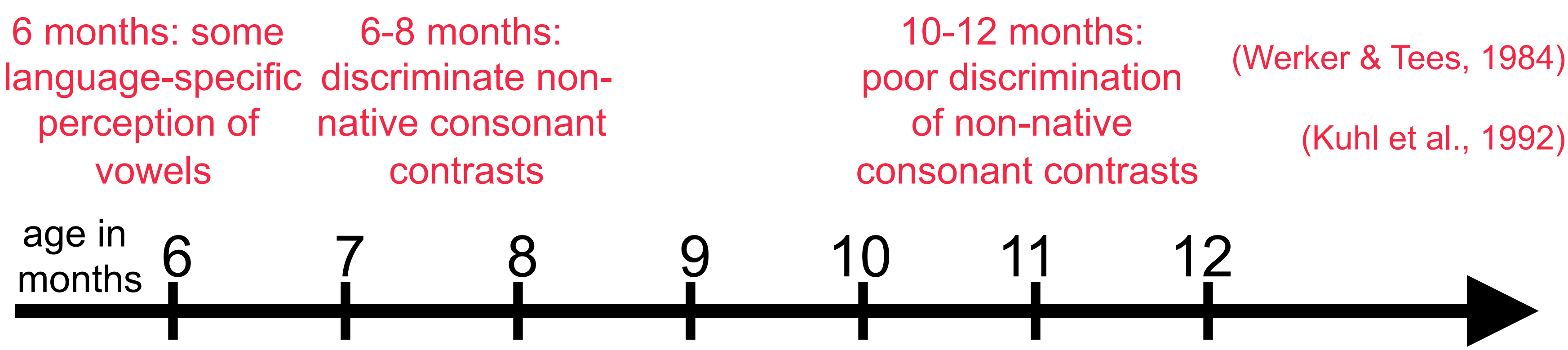
bike

bike

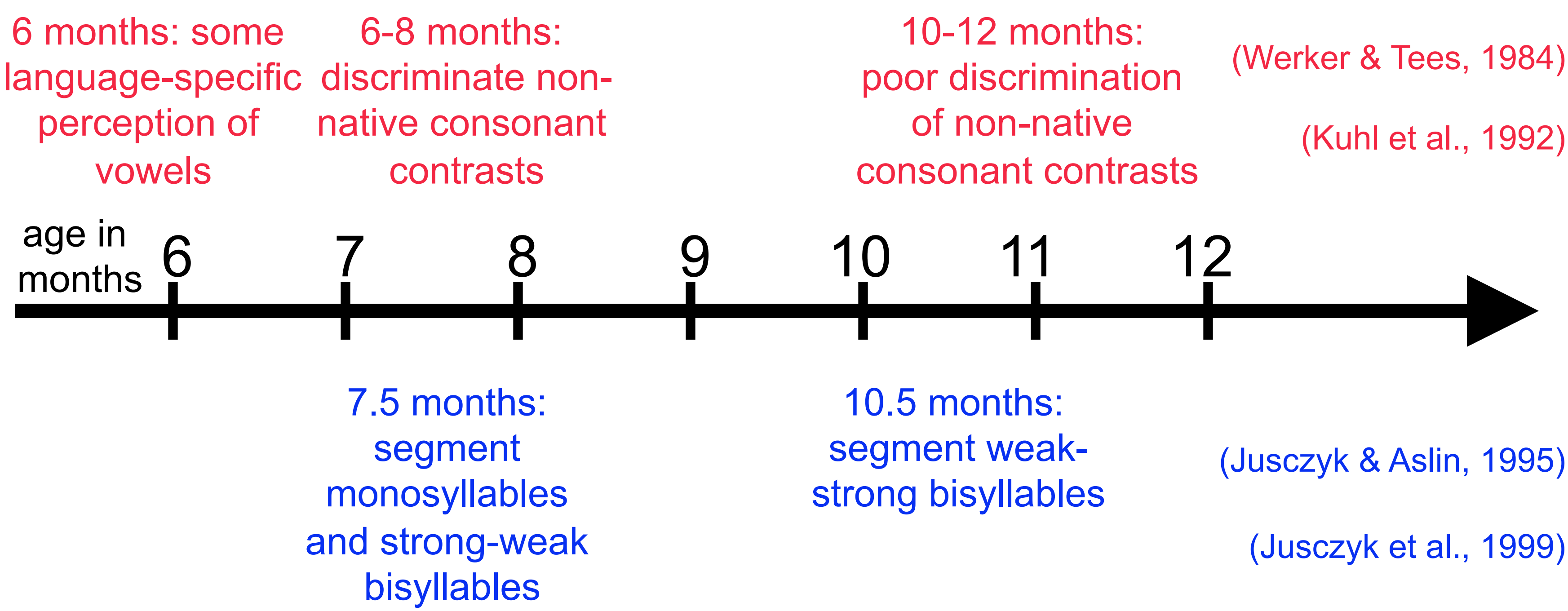
bike

BIKE

Timecourse of Acquisition



Timecourse of Acquisition



Timecourse of Acquisition

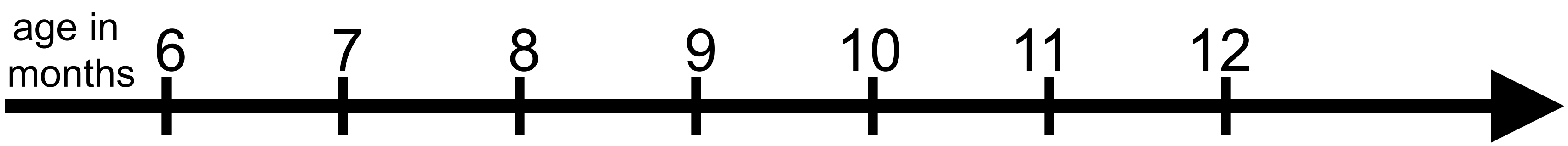
6 months: some language-specific perception of vowels

6-8 months: discriminate non-native consonant contrasts

10-12 months: poor discrimination of non-native consonant contrasts

(Werker & Tees, 1984)

(Kuhl et al., 1992)



6 months: segment words next to “Mommy” or baby’s own name

7.5 months: segment monosyllables and strong-weak bisyllables

10.5 months: segment weak-strong bisyllables

(Jusczyk & Aslin, 1995)

(Jusczyk et al., 1999)

(Bortfeld et al., 2005)

Hypothesis

Infants use top-down constraints from words when acquiring phonetic categories

1. Formalize a model that can simultaneously learn sounds and words

2. Show that infants are sensitive to words in ways the model would predict

Hypothesis

Infants use top-down constraints from words when acquiring phonetic categories

1. Formalize a model that can simultaneously learn sounds and words

2. Show that infants are sensitive to words in ways the model would predict

A Better Generative Model

To create a corpus

Phonetic Categories

↓

Lexicon

↓

Corpus

A Better Generative Model

<u>To create a corpus</u>

1. Generate a phonetic category inventory

Phonetic Categories

↓

Lexicon

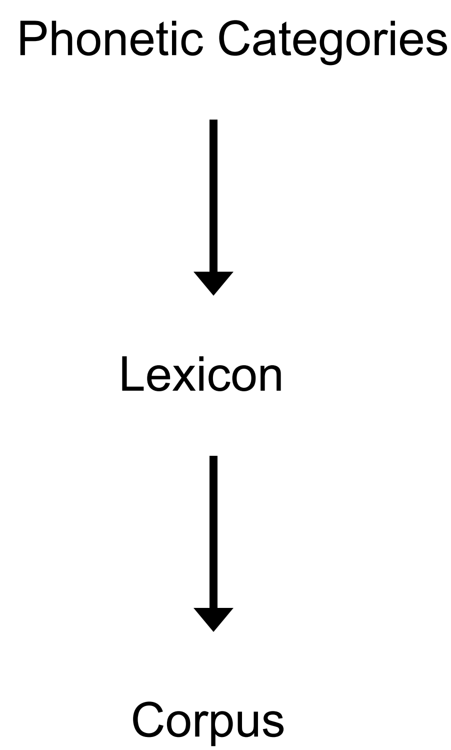
↓

Corpus

A Better Generative Model

<u>To create a corpus</u>

1. Generate a phonetic category inventory
    - Sample a mean, covariance, and frequency of occurrence for each Gaussian category



A Better Generative Model

To create a corpus

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a lexicon

Phonetic Categories

↓

Lexicon

↓

Corpus

A Better Generative Model

<u>To create a corpus</u>

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a lexicon
   - Sample a length and frequency of occurrence for each lexical item

Phonetic Categories

↓

Lexicon

↓

Corpus

A Better Generative Model

<u>To create a corpus</u>

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a lexicon
   - Sample a length and frequency of occurrence for each lexical item
   - For each phoneme slot, sample a phonetic category from the phonetic category inventory

Phonetic Categories

↓

Lexicon

↓

Corpus

A Better Generative Model

To create a corpus

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a lexicon
   - Sample a length and frequency of occurrence for each lexical item
   - For each phoneme slot, sample a phonetic category from the phonetic category inventory
3. Generate a corpus

Phonetic Categories

↓

Lexicon

↓

Corpus

A Better Generative Model

<u>To create a corpus</u>

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a lexicon
   - Sample a length and frequency of occurrence for each lexical item
   - For each phoneme slot, sample a phonetic category from the phonetic category inventory
3. Generate a corpus
   - For each word, sample a lexical item according to its frequency

Phonetic Categories

↓

Lexicon

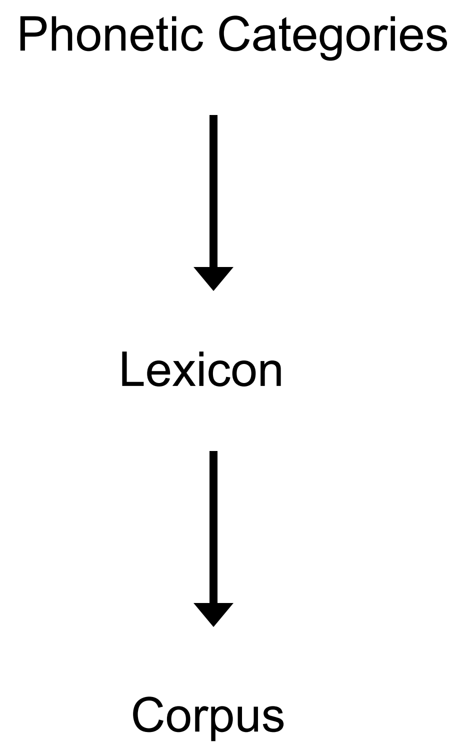
↓

Corpus

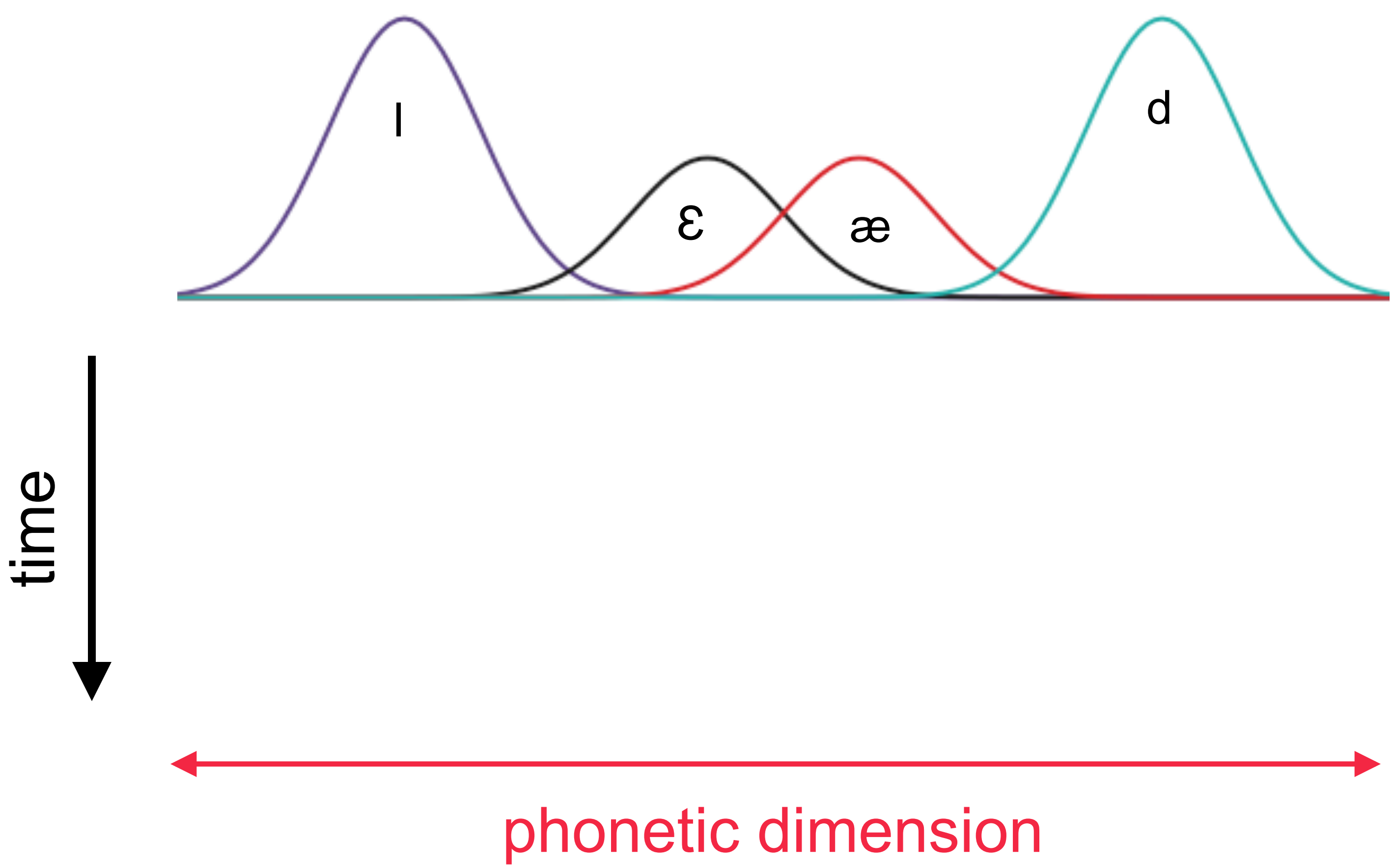
A Better Generative Model

To create a corpus

1. Generate a phonetic category inventory
   - Sample a mean, covariance, and frequency of occurrence for each Gaussian category
2. Generate a lexicon
   - Sample a length and frequency of occurrence for each lexical item
   - For each phoneme slot, sample a phonetic category from the phonetic category inventory
3. Generate a corpus
   - For each word, sample a lexical item according to its frequency
   - Generate an acoustic value each phonetic category contained in that lexical item

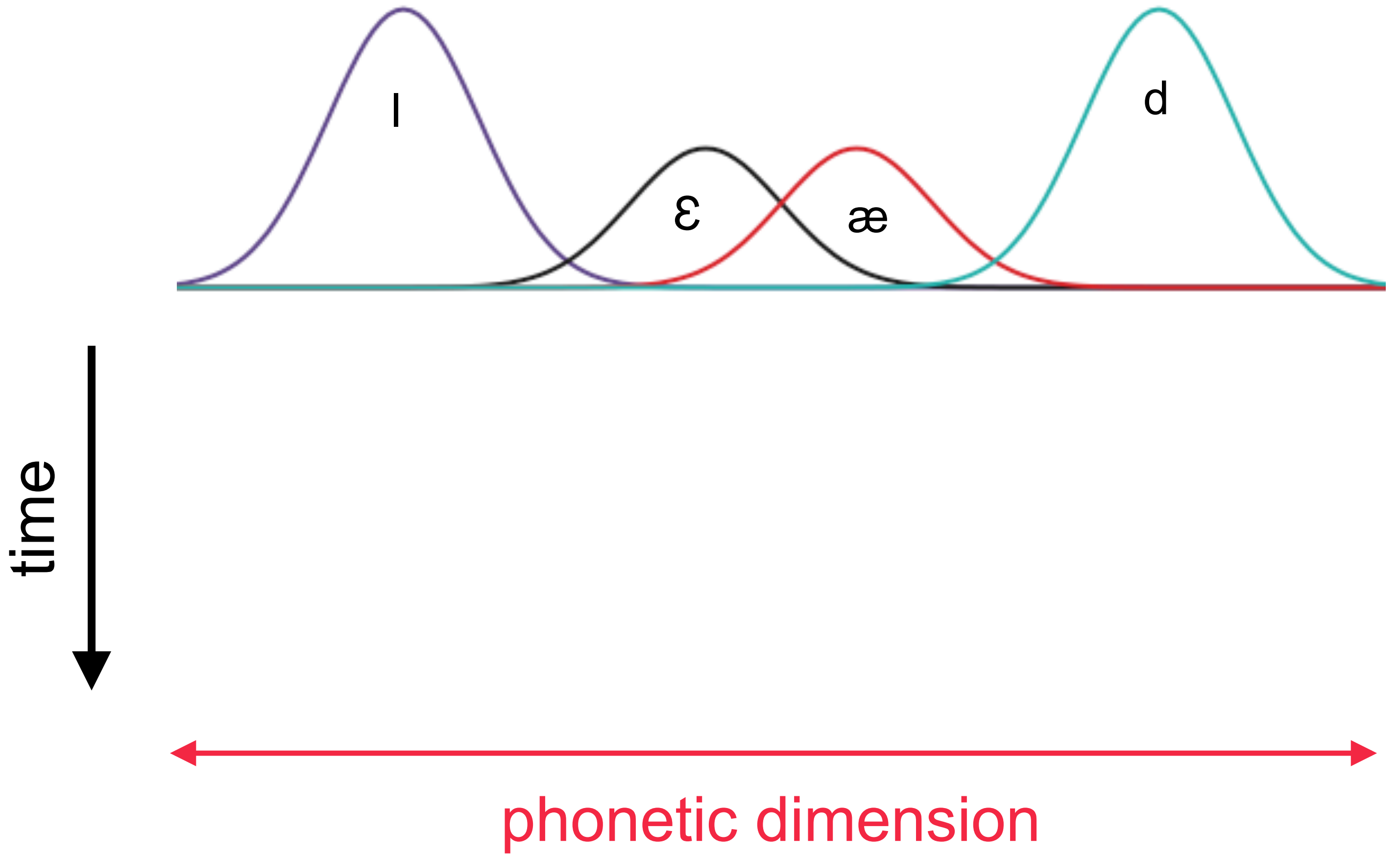


A Better Generative Model

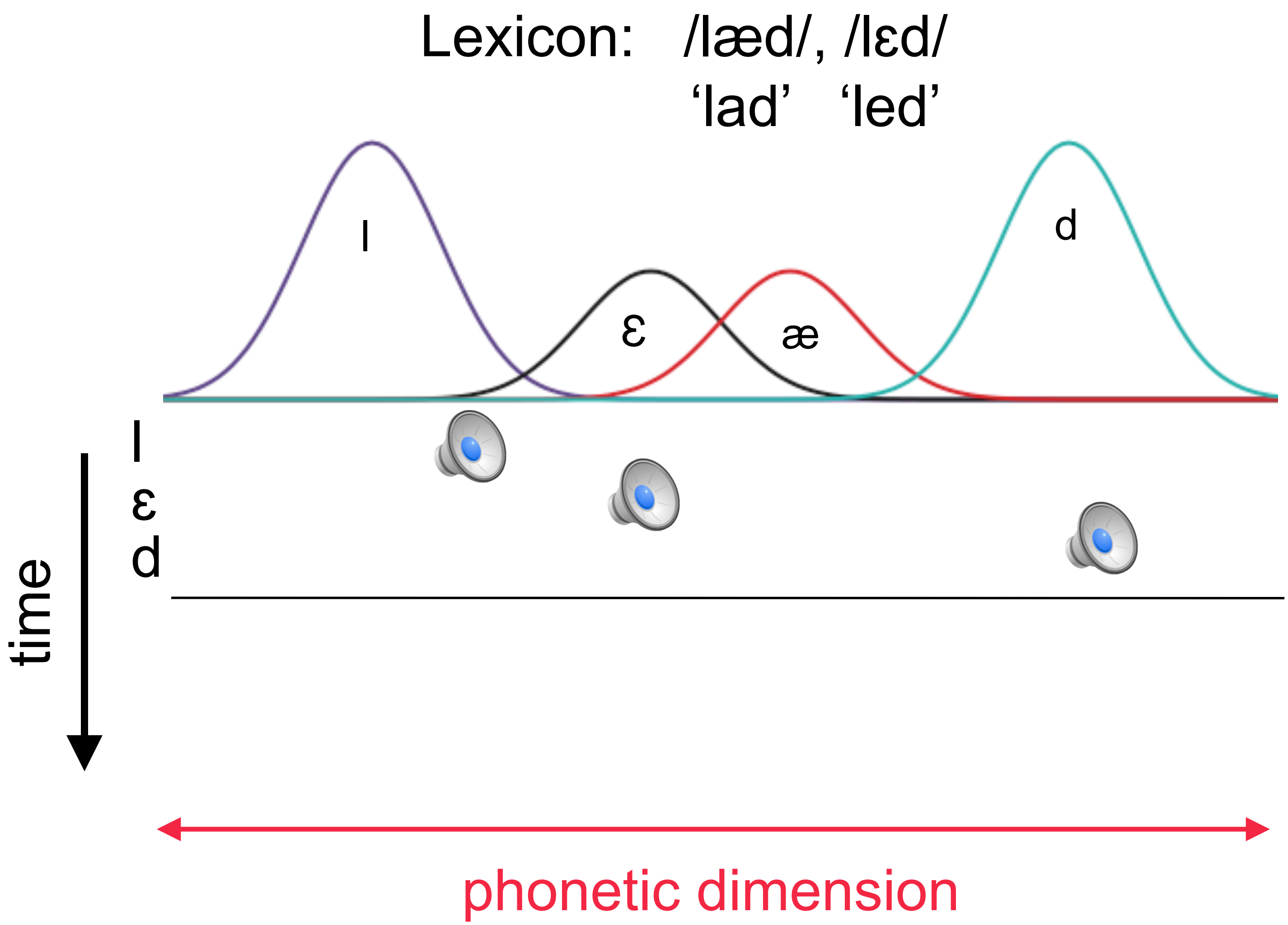


A Better Generative Model

Lexicon: /læd/, /lɛd/
'lad' 'led'

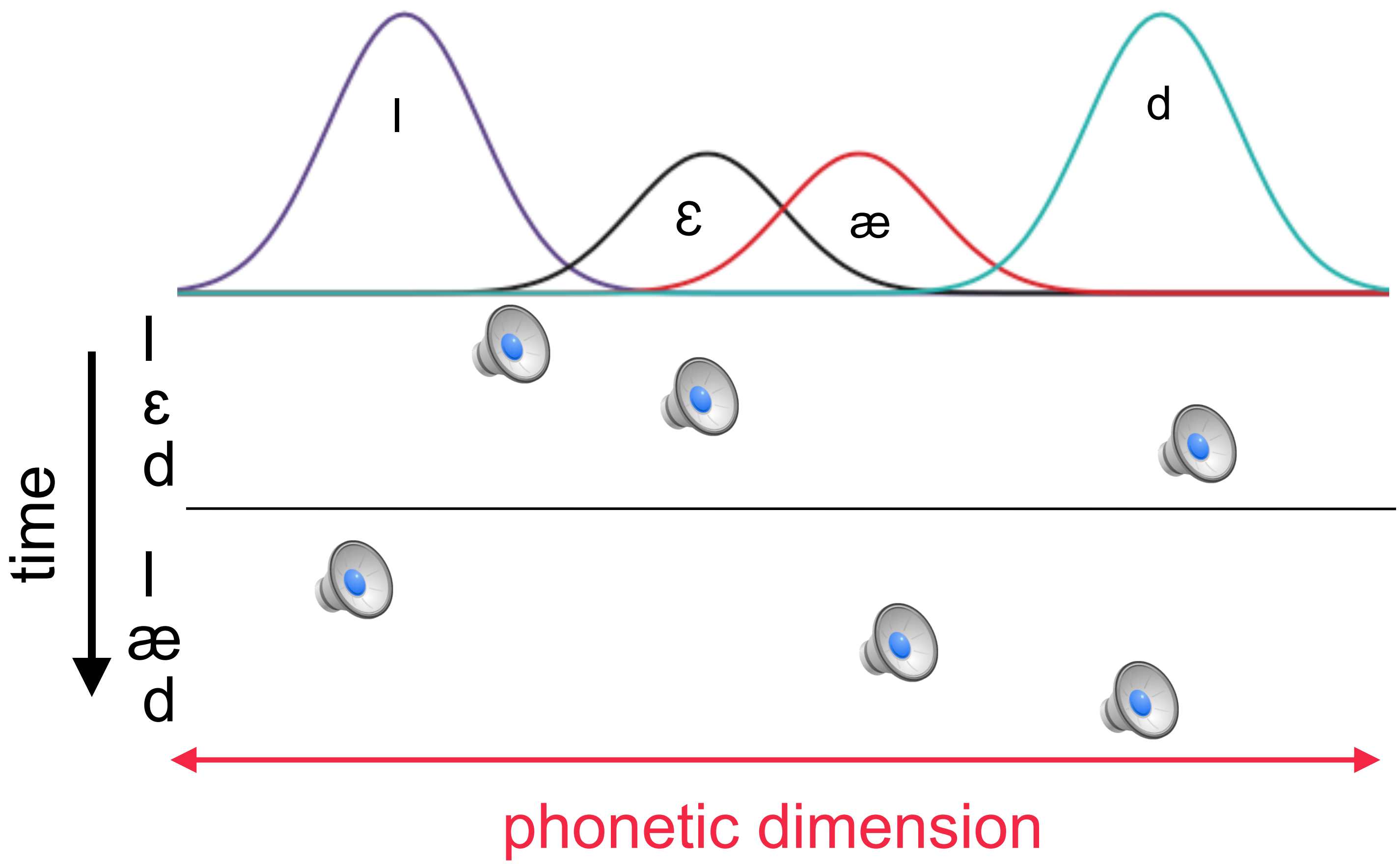


A Better Generative Model



A Better Generative Model

Lexicon: /læd/, /lɛd/
'lad' 'led'



Qualitative Behavior

Compare lexical-distributional model's behavior on two lexicons

- Non-minimal pair lexicon: 'add', 'el'
- Minimal pair lexicon: 'add, 'Ed', 'Al', 'el'

Qualitative Behavior

Compare lexical-distributional model's behavior on two lexicons

- Non-minimal pair lexicon: 'add', 'el'
- Minimal pair lexicon: 'add, 'Ed', 'Al', 'el'

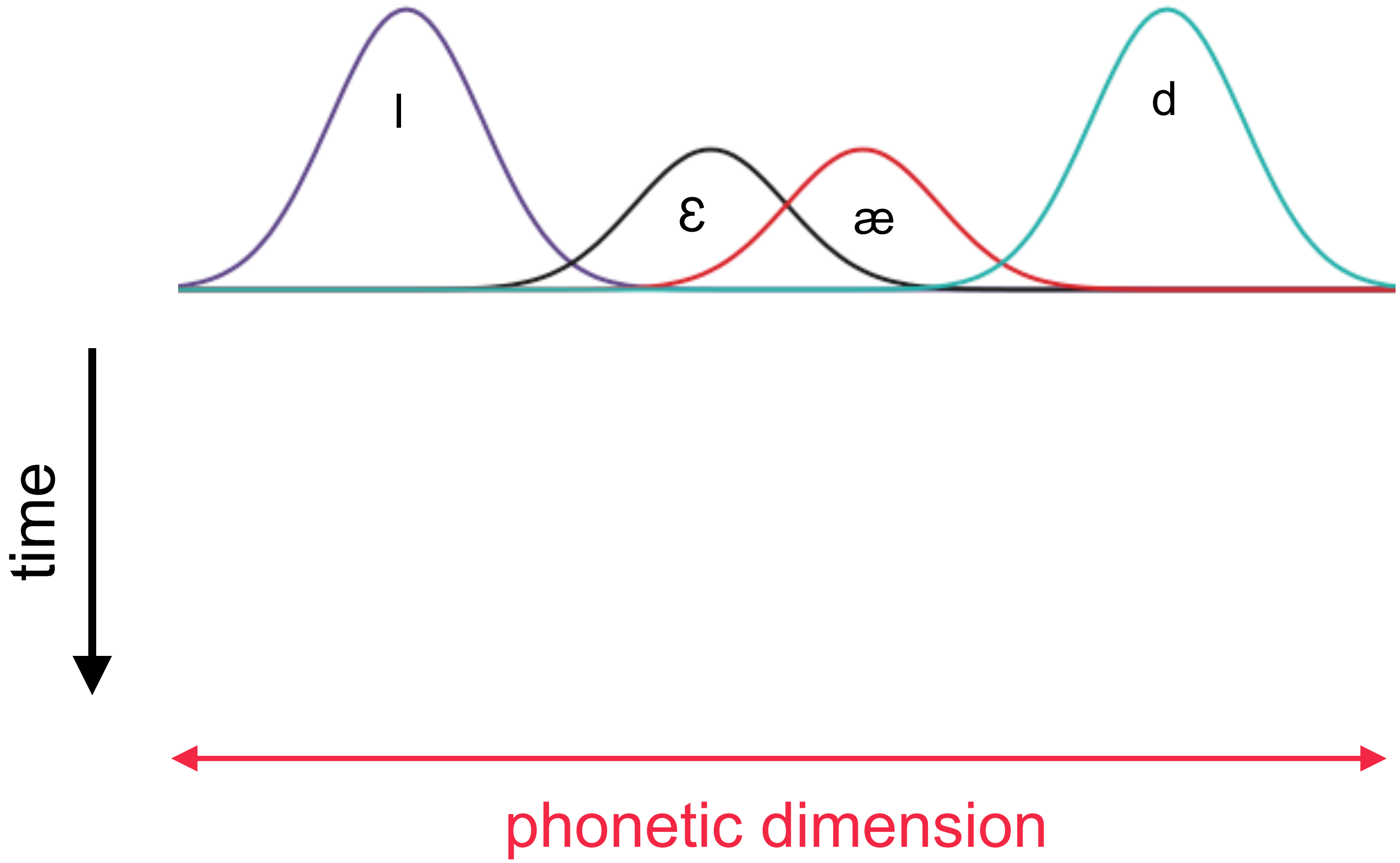
Minimal pairs:

add vs. Ed

Typically taken as evidence
that sounds are different

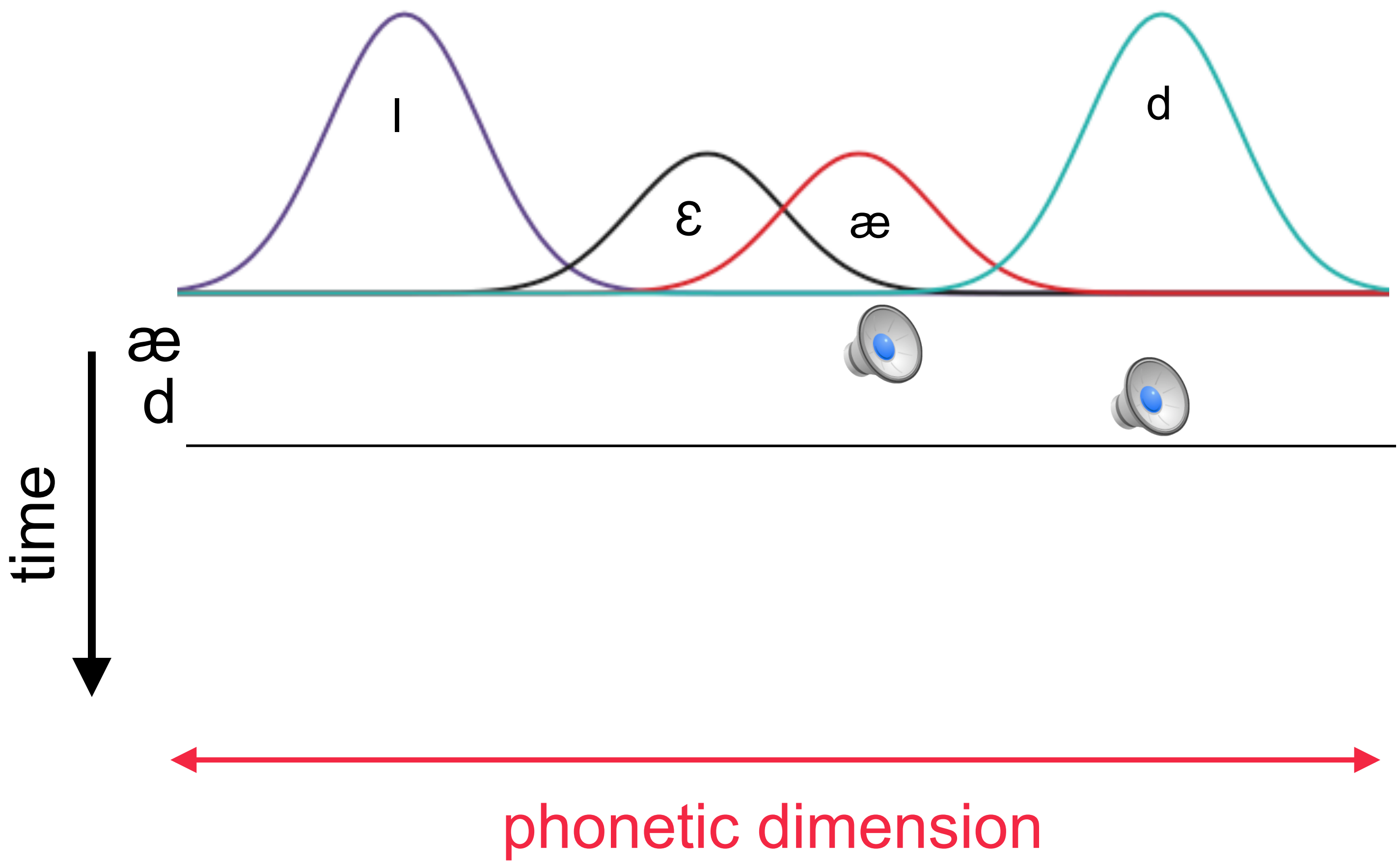
Non-Minimal Pair Lexicon

Lexicon: /æd/, /ɛl/
‘add’ ‘el’



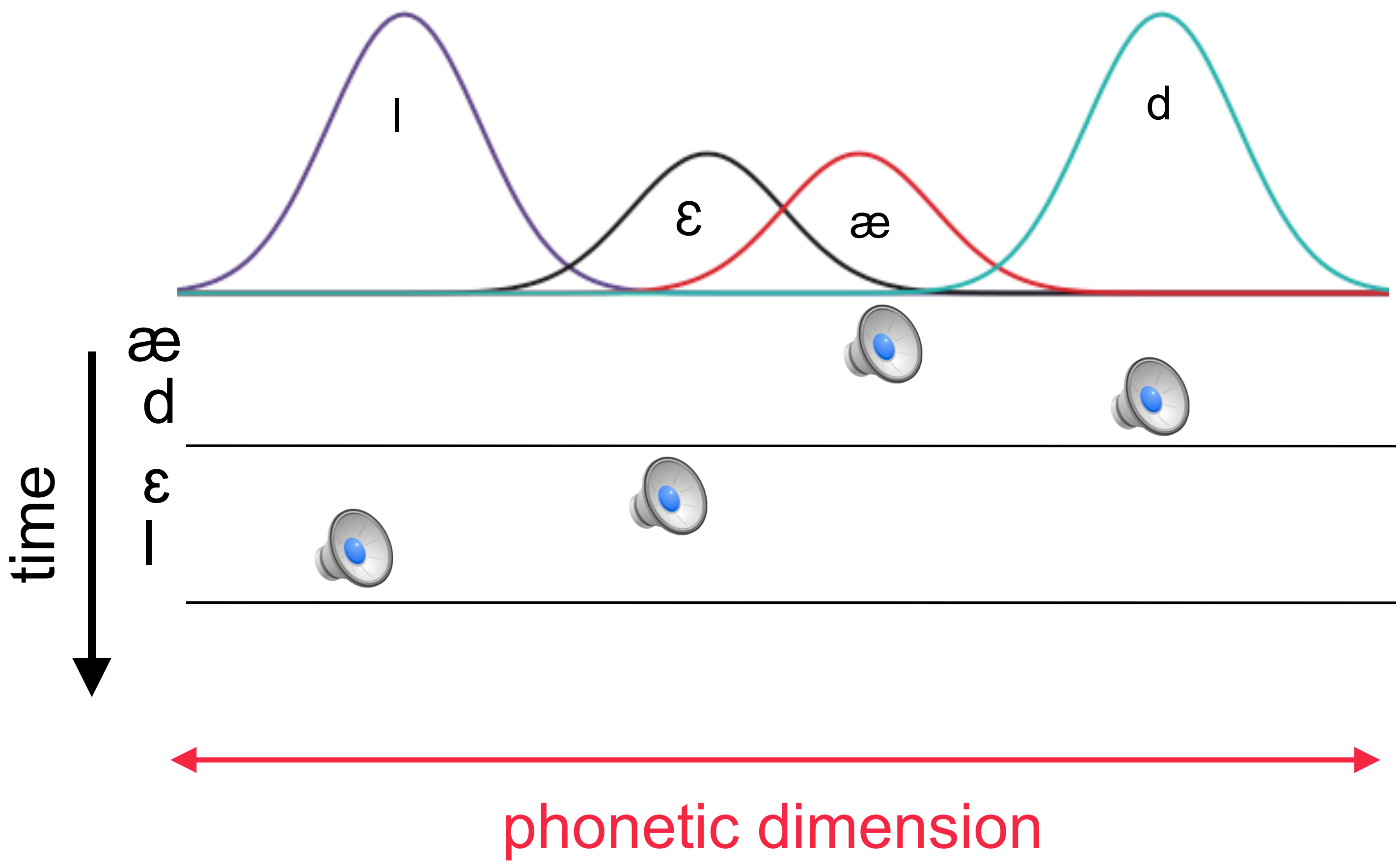
Non-Minimal Pair Lexicon

Lexicon: /æd/, /ɛl/
'add' 'el'



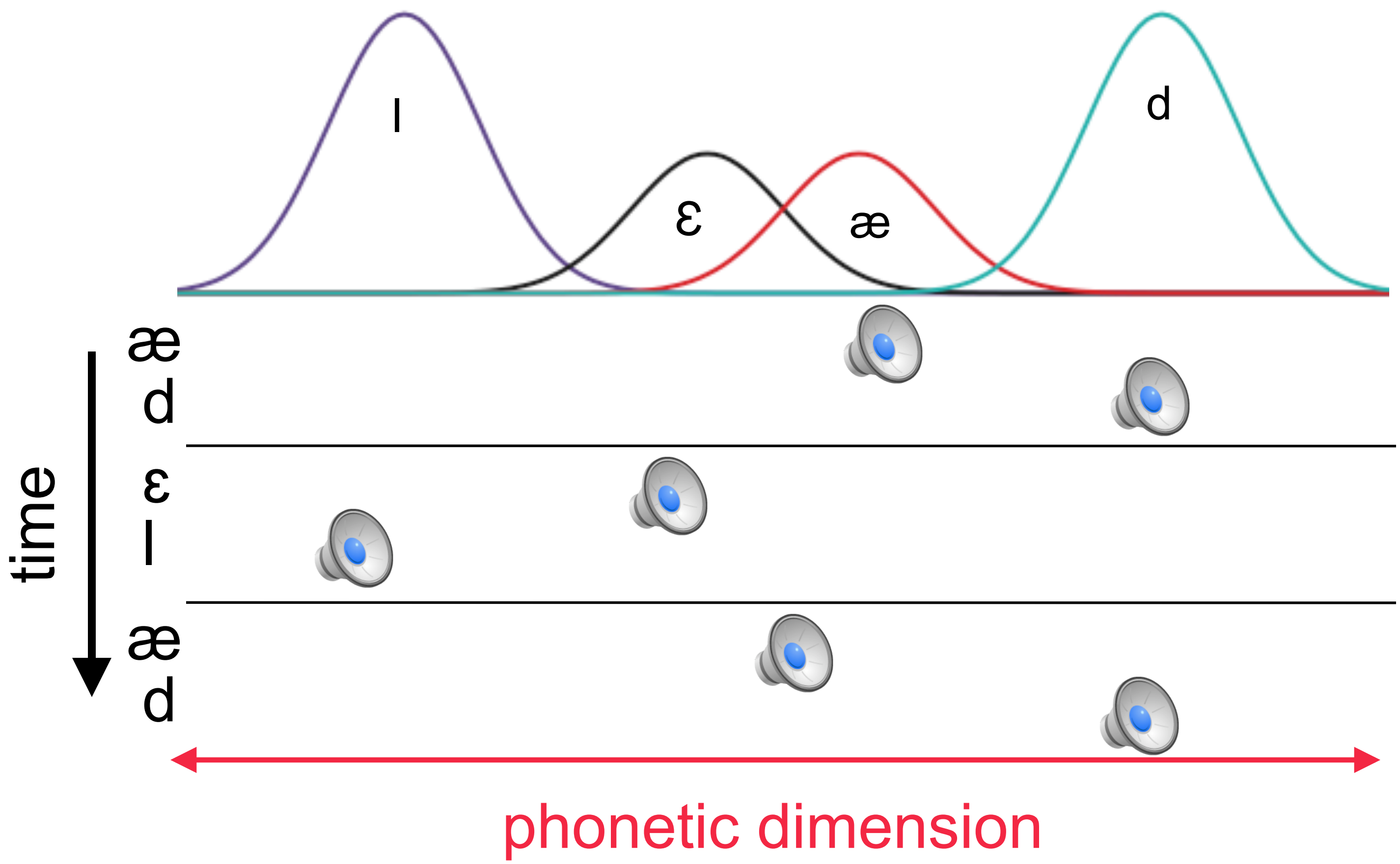
Non-Minimal Pair Lexicon

Lexicon: /æd/, /ɛl/
'add' 'el'

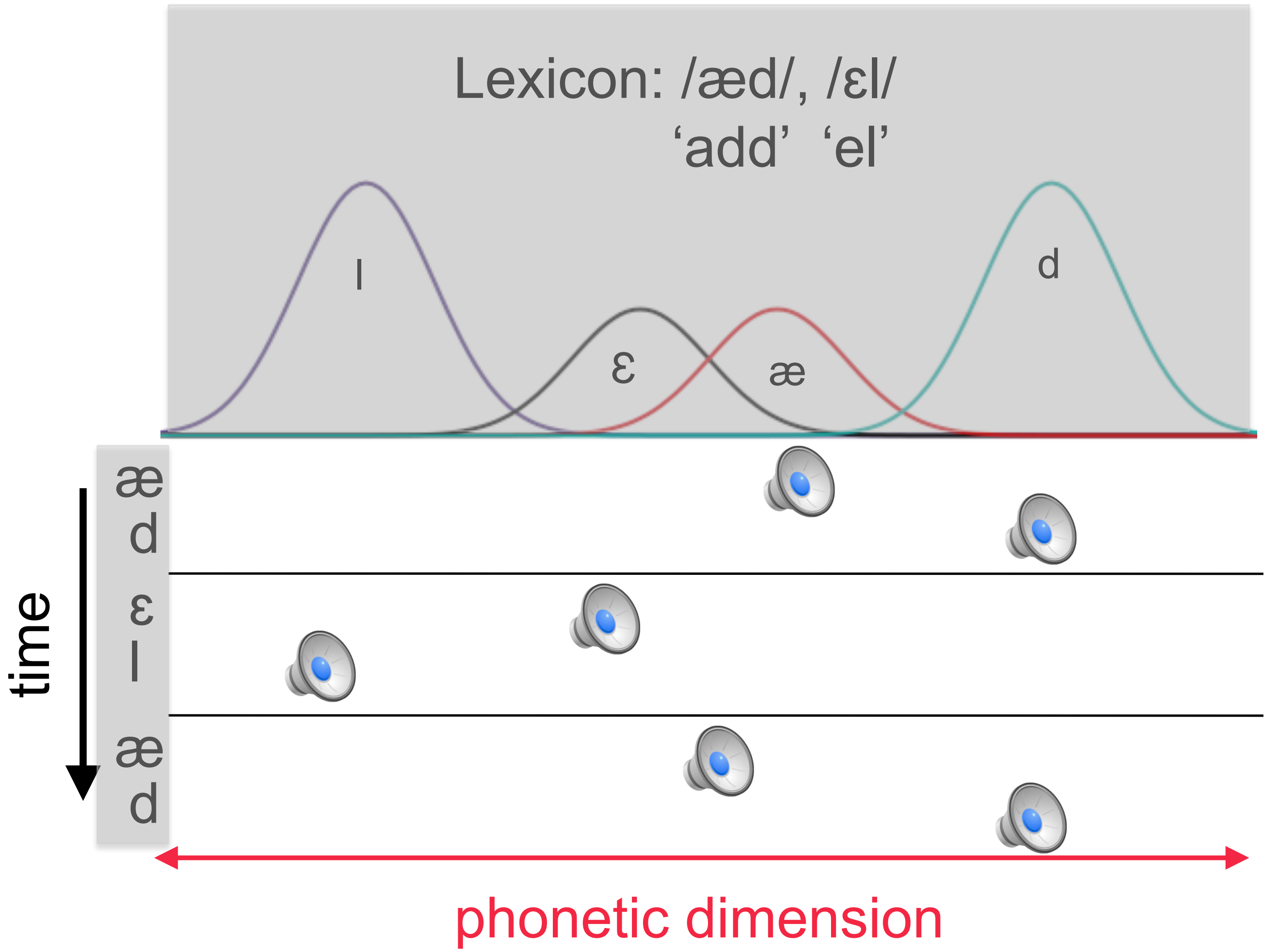


Non-Minimal Pair Lexicon

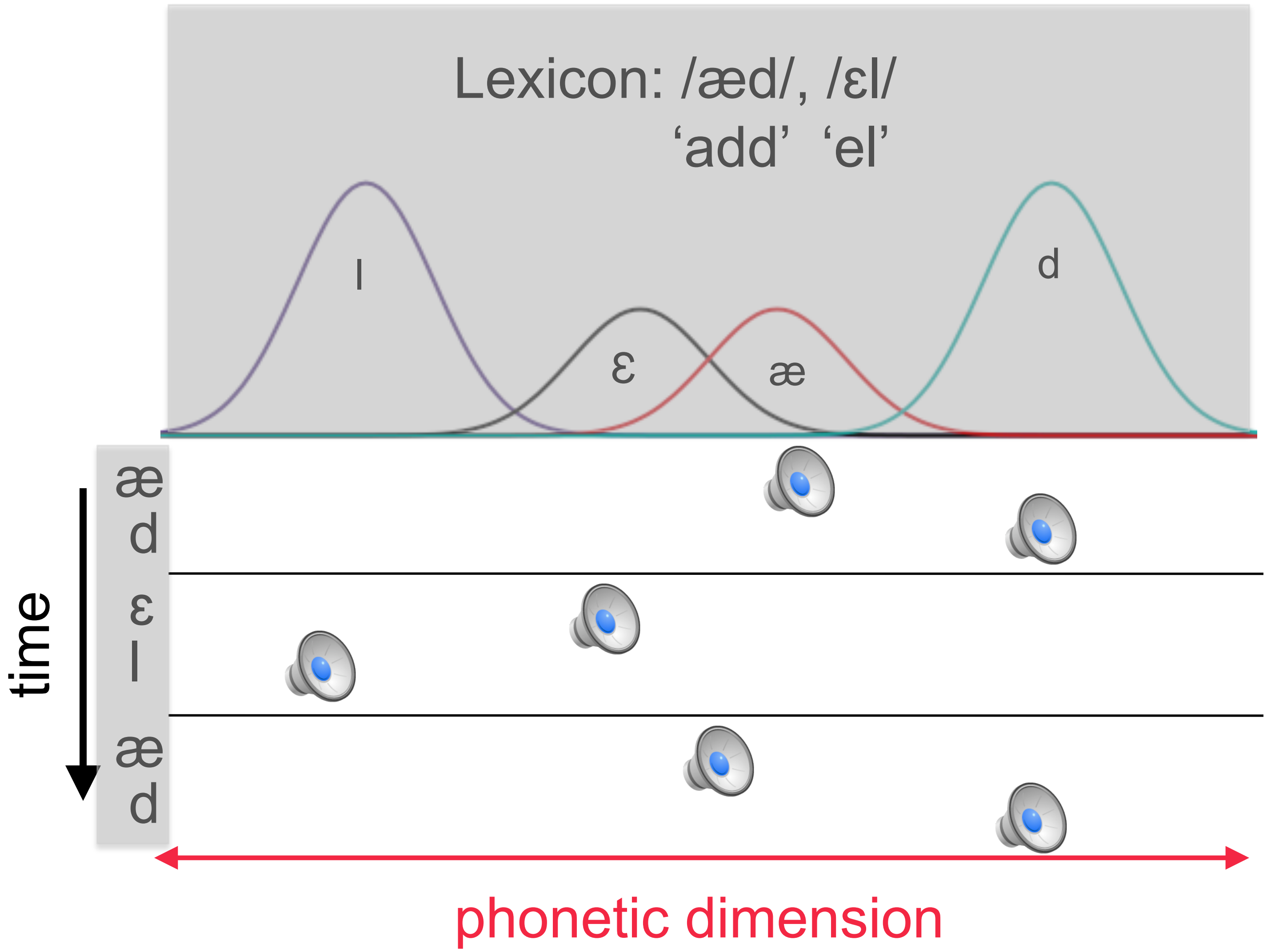
Lexicon: /æd/, /ɛl/
'add' 'el'



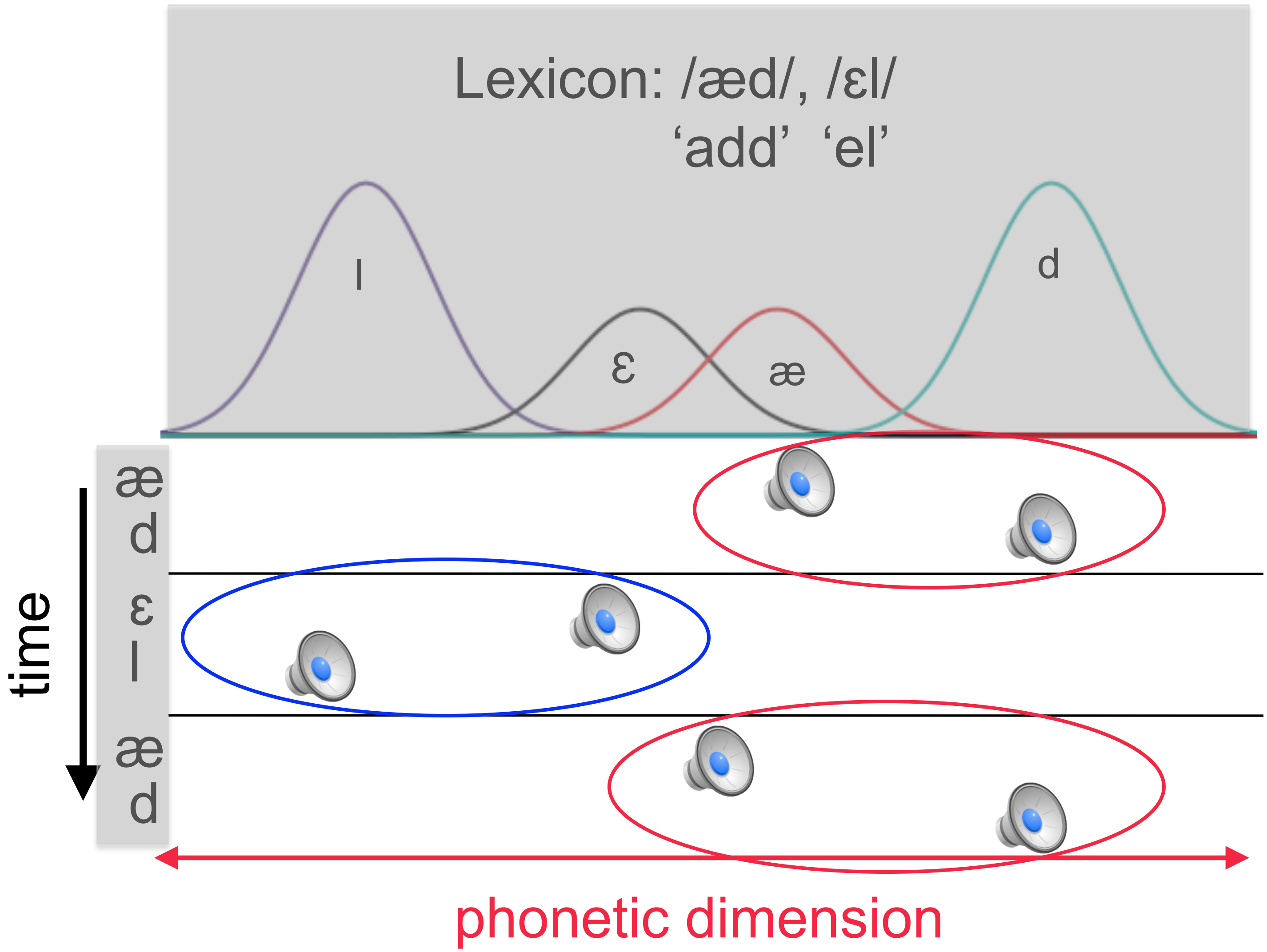
Non-Minimal Pair Lexicon



Non-Minimal Pair Lexicon



Non-Minimal Pair Lexicon



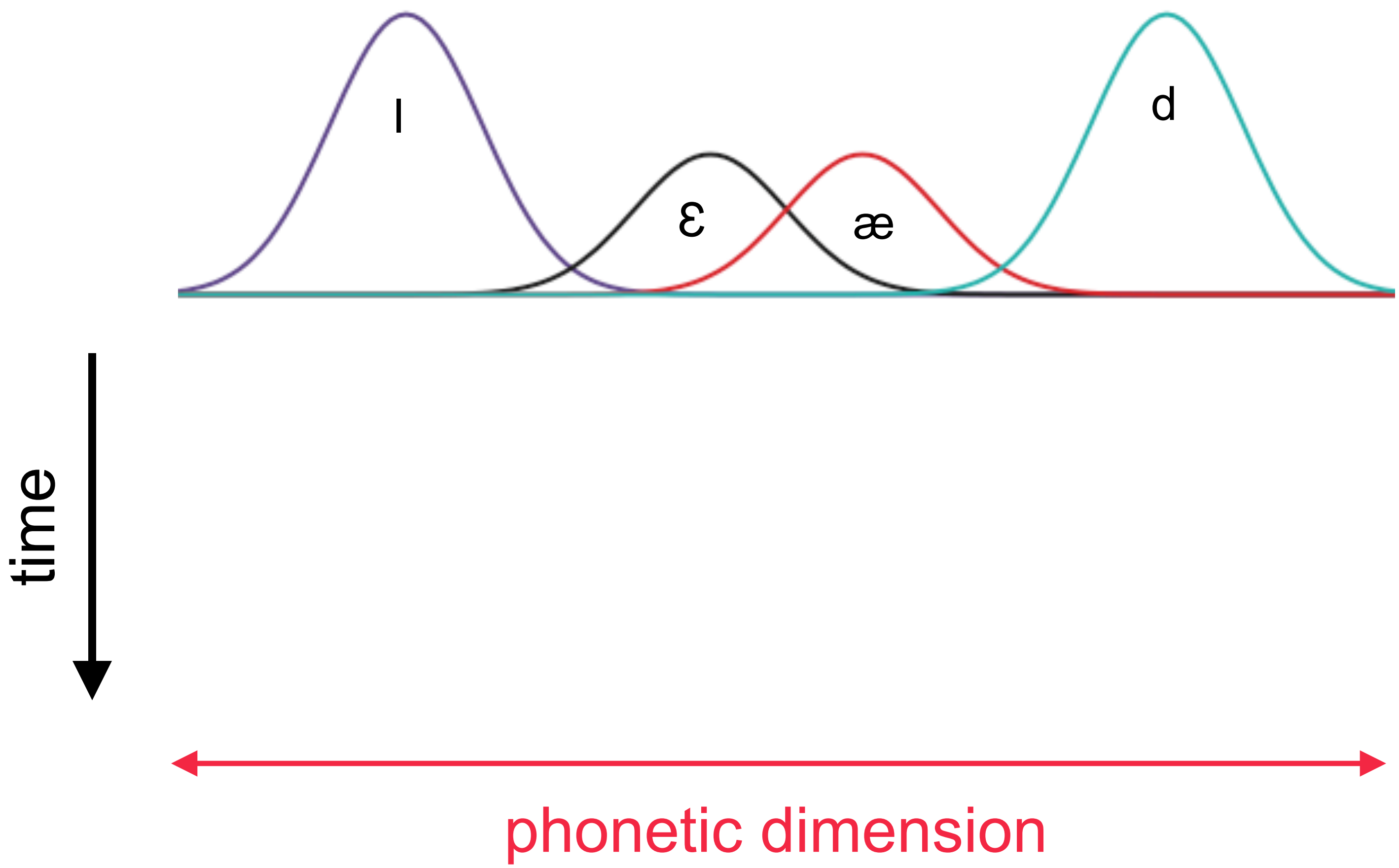
Non-Minimal Pair Lexicon

el

add

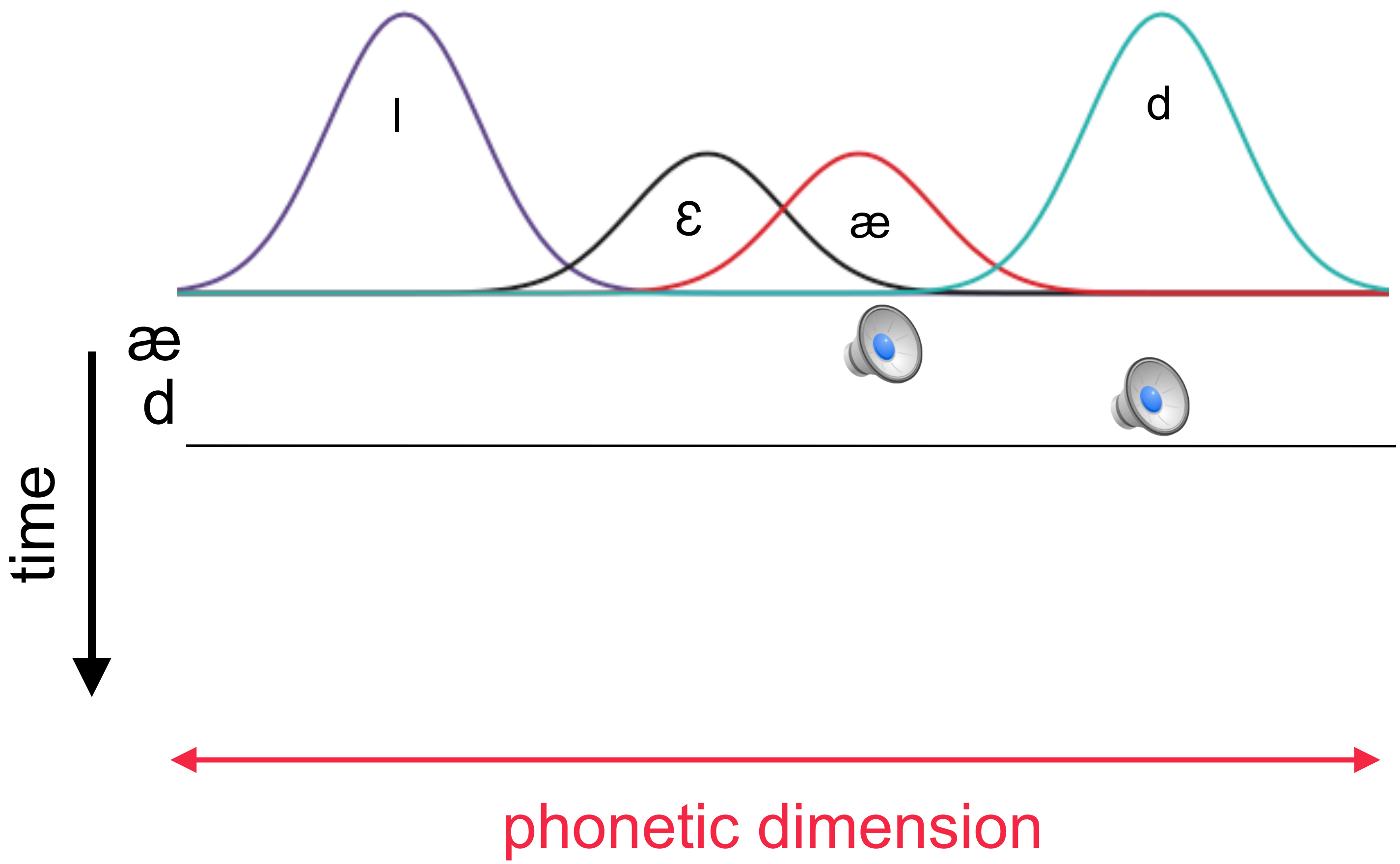
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æl/, /ɛl/
'add' 'Ed' 'Al' 'el'



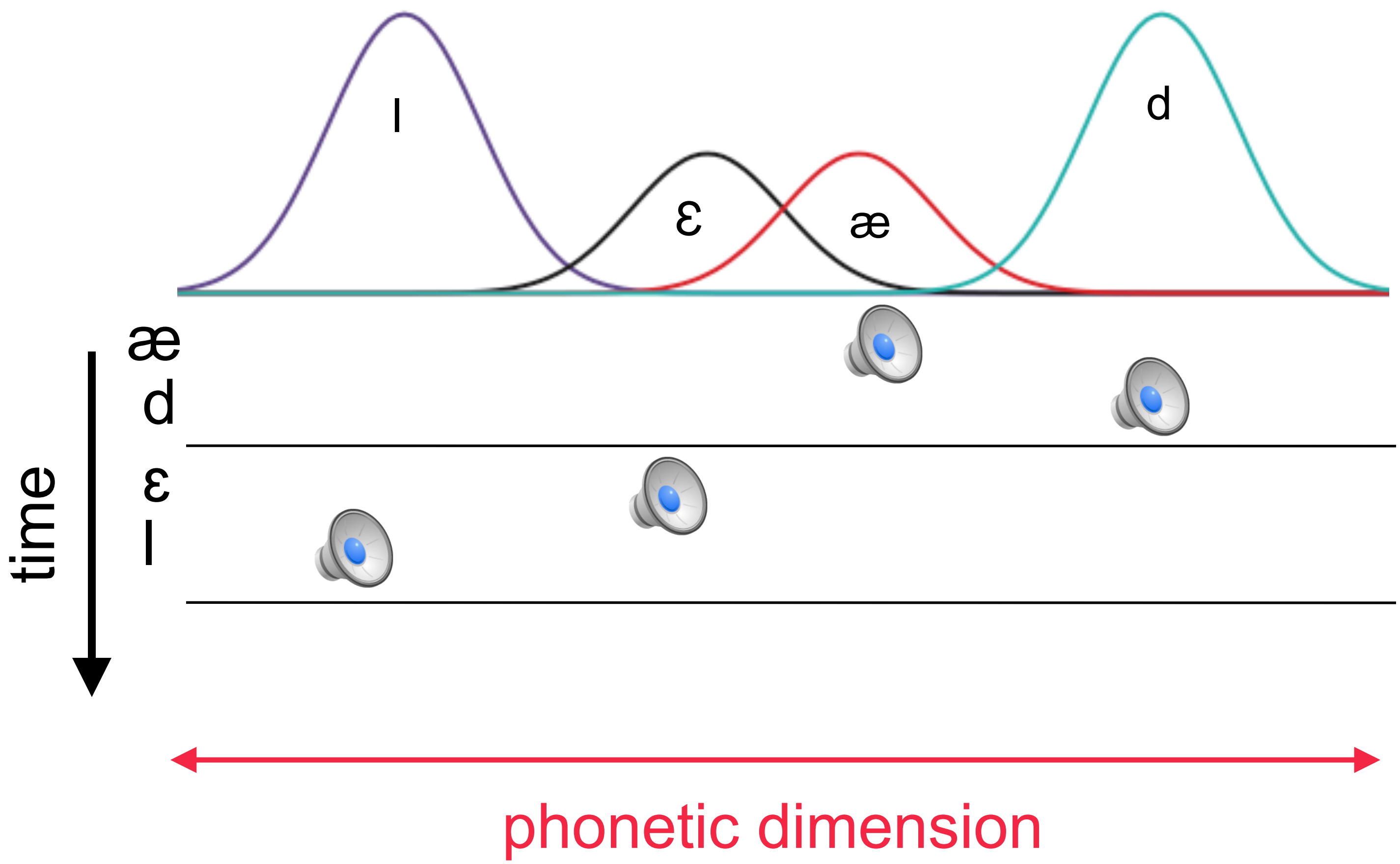
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æl/, /ɛl/
'add' 'Ed' 'Al' 'el'



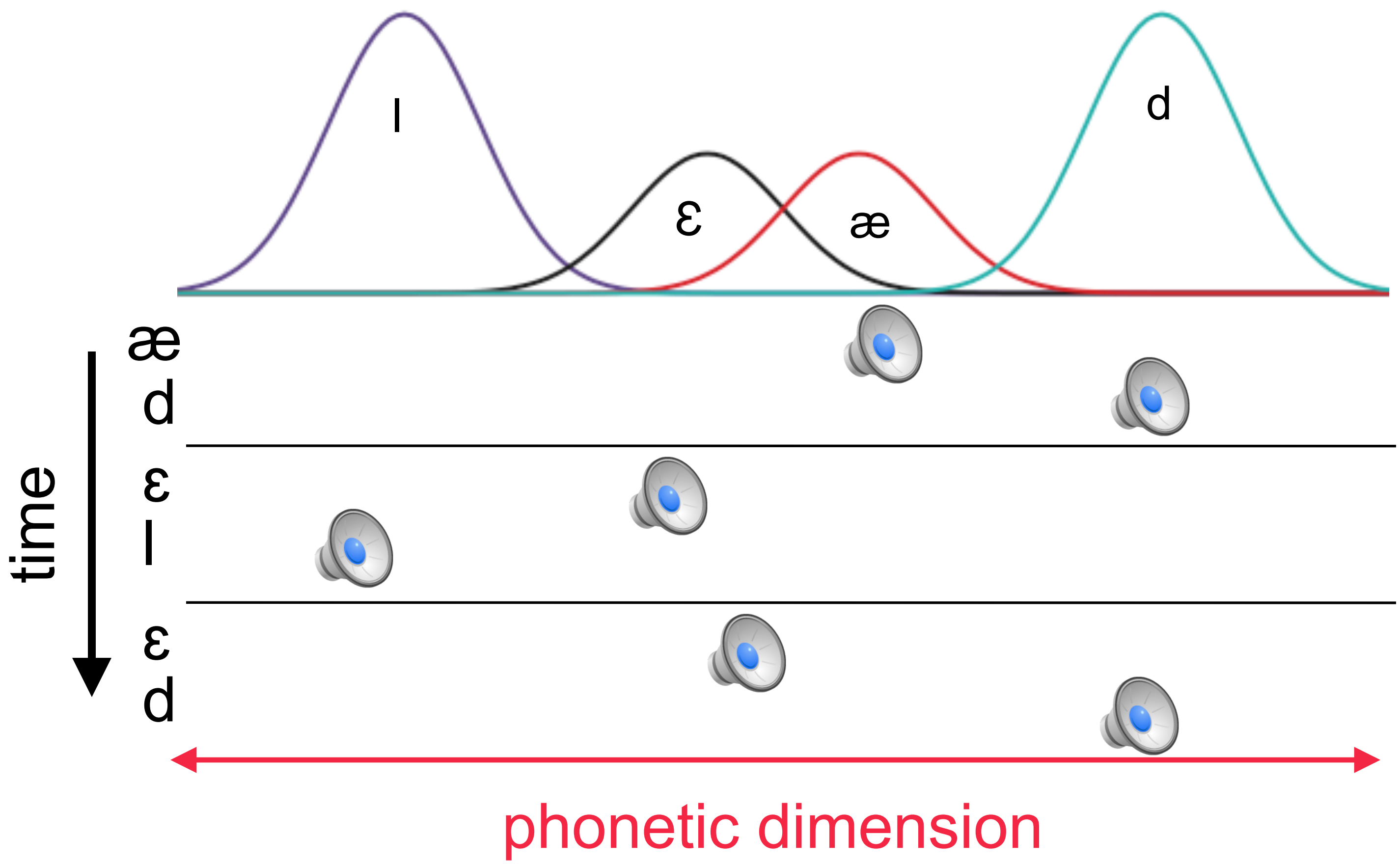
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æl/, /ɛl/
‘add’ ‘Ed’ ‘Al’ ‘el’

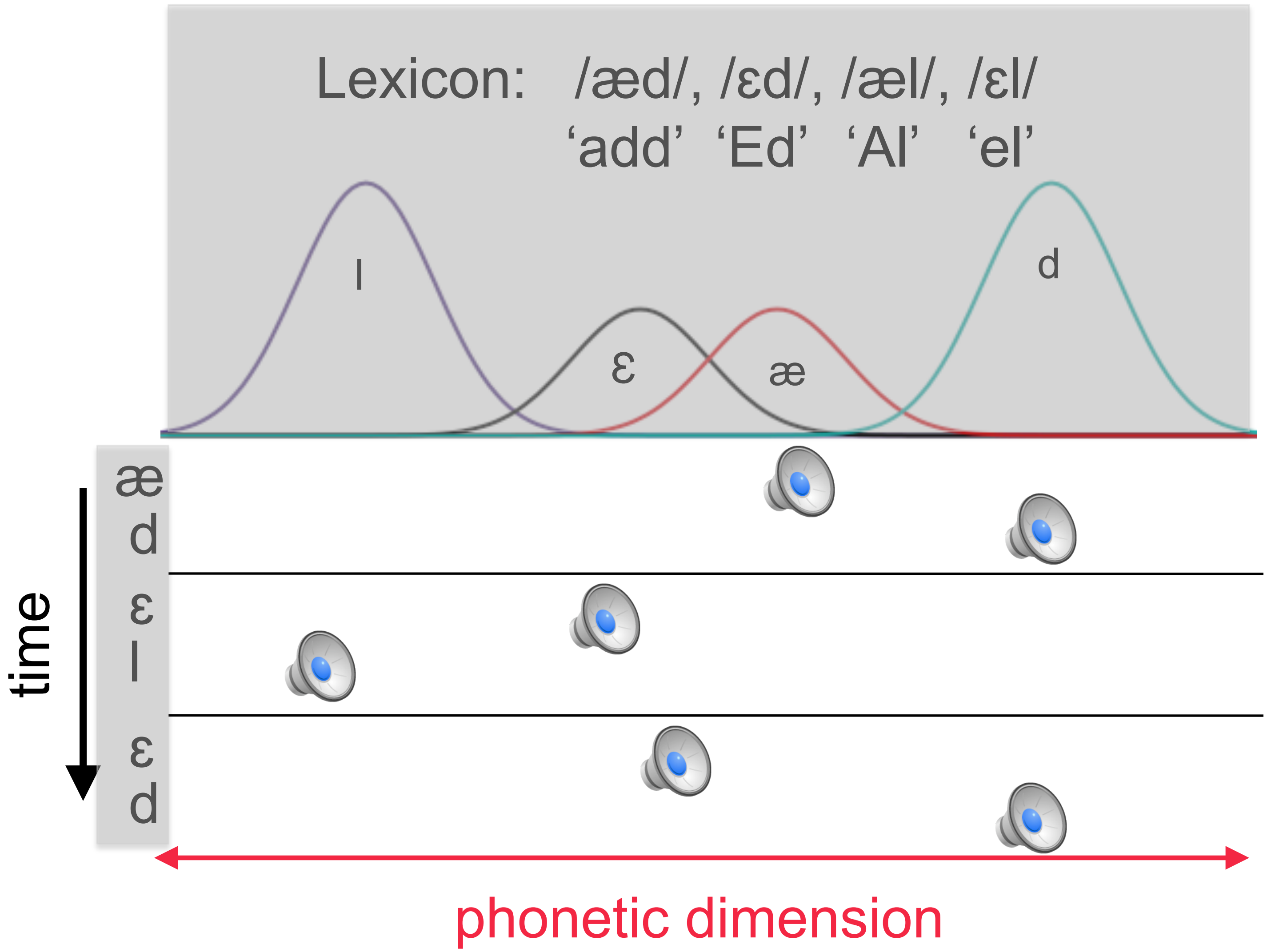


Minimal Pair Lexicon

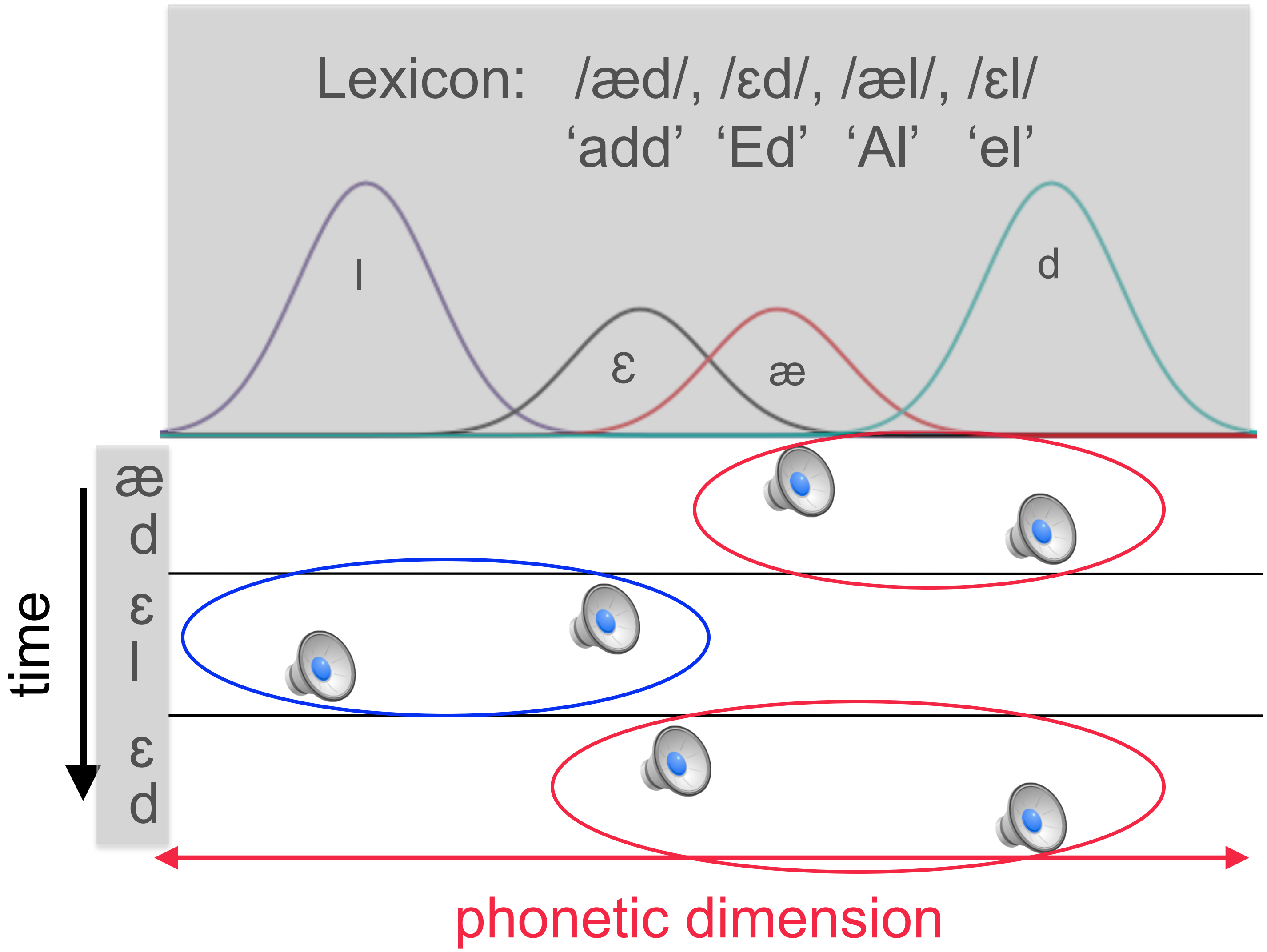
Lexicon: /æd/, /ɛd/, /æl/, /ɛl/
‘add’ ‘Ed’ ‘Al’ ‘el’



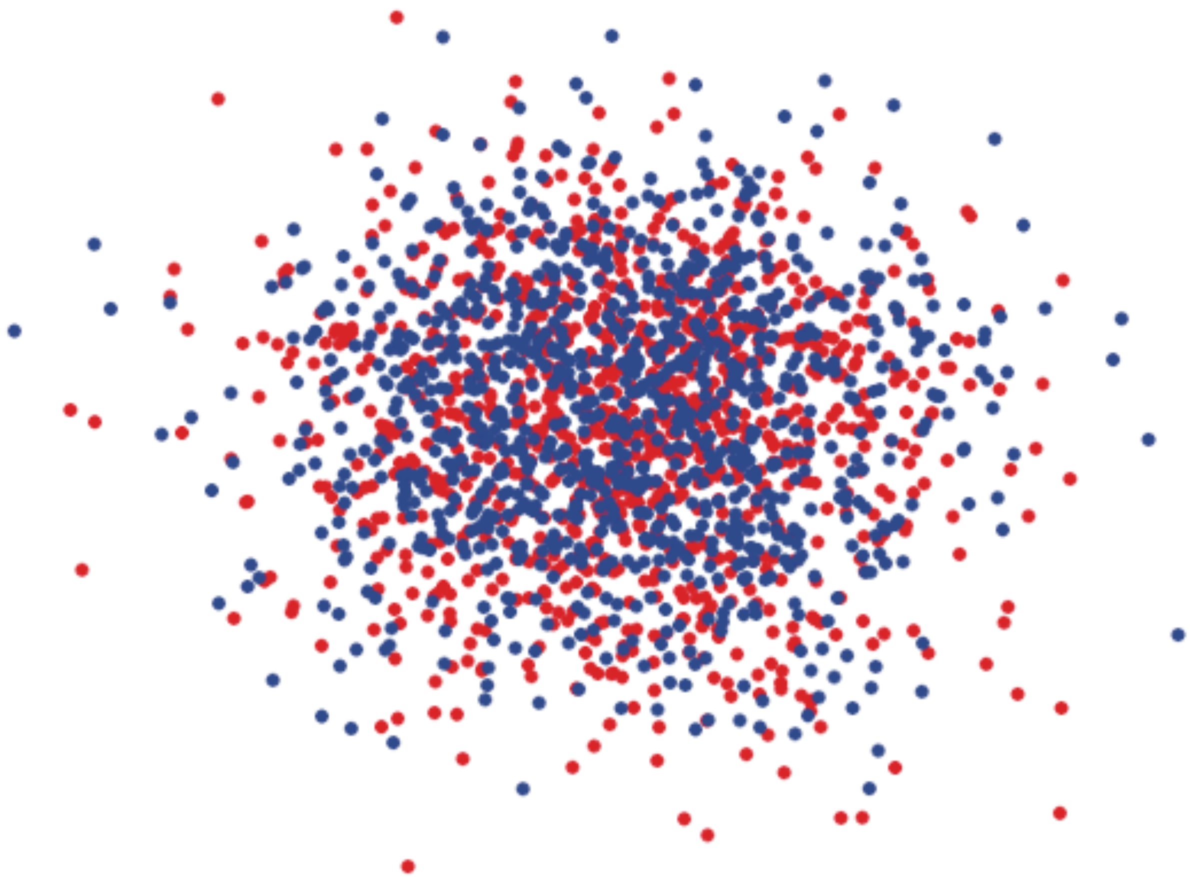
Minimal Pair Lexicon



Minimal Pair Lexicon



Minimal Pair Lexicon

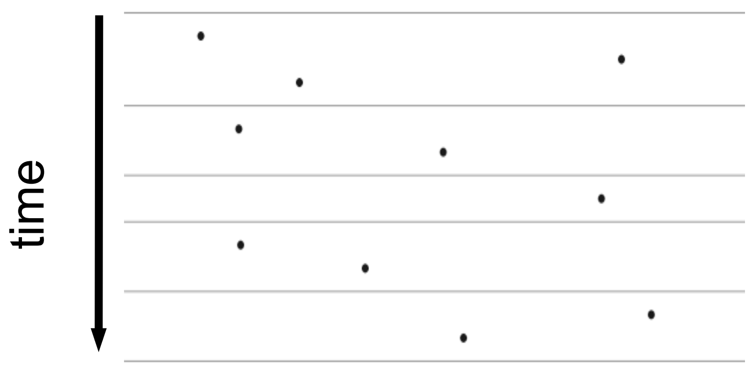


Al

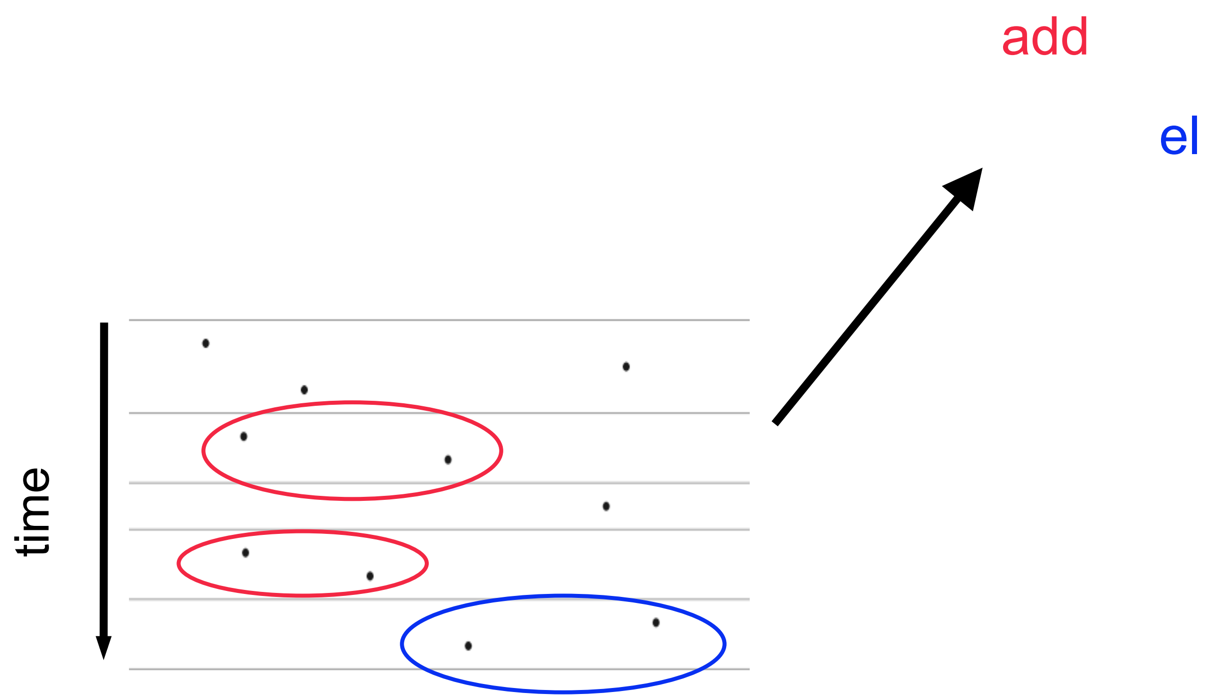
Ed el

add

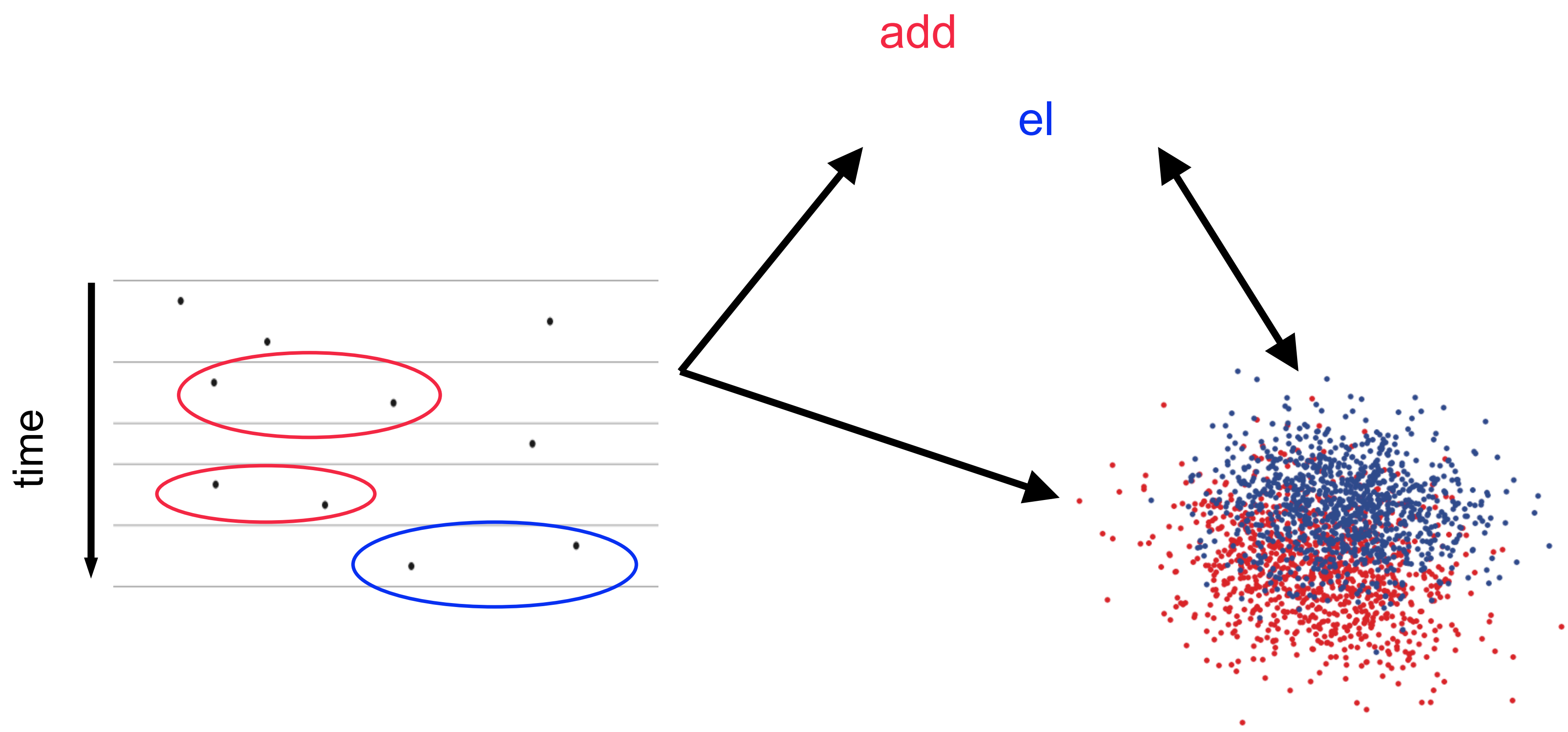
Lexical-Distributional Learning



Lexical-Distributional Learning



Lexical-Distributional Learning



Two Generative Models

Distributional Model

Phonetic Categories

↓

Corpus

Lexical-Distributional Model

Phonetic Categories

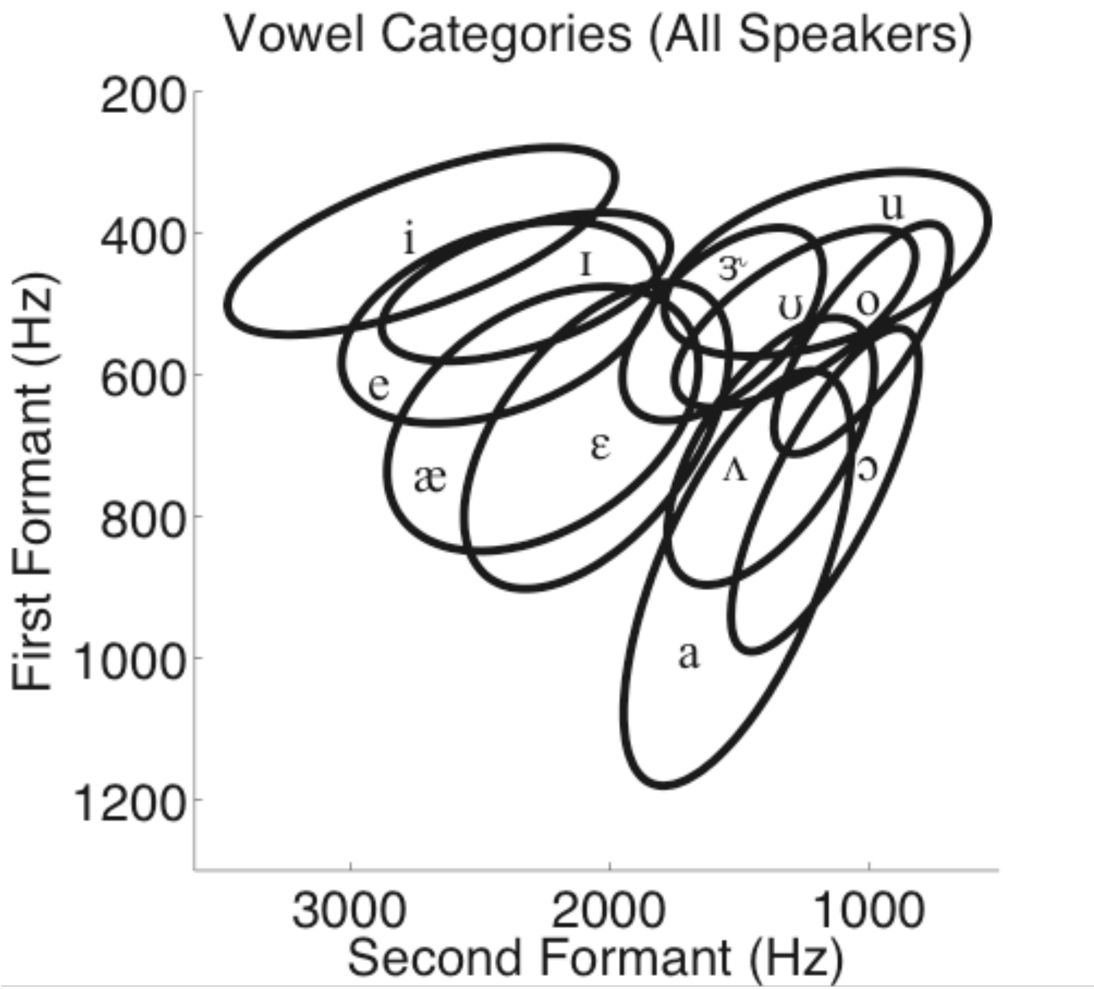
↓

Lexicon

↓

Corpus

Words Can Help Ideal Learners



(Hillenbrand et al., 1995)

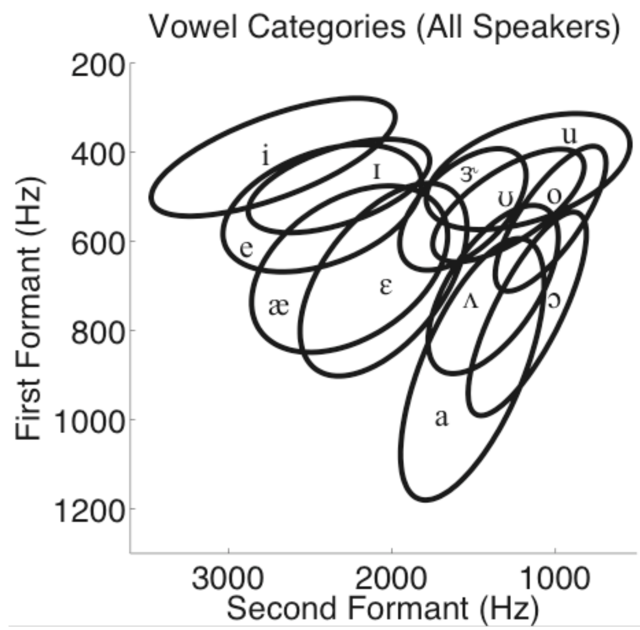
Words Can Help Ideal Learners

- Created corpus of 5000 word tokens from CHILDES Parental Corpus (Li & Shirai, 2000)
  - Orthographic forms phonematized using Carnegie Mellon Pronuncing Dictionary
  - Words sampled according to corpus frequency
  - 6,409 vowel and 8,917 consonant tokens

Words Can Help Ideal Learners

- Created corpus of 5000 word tokens from CHILDES Parental Corpus (Li & Shirai, 2000)
  - Orthographic forms phonematized using Carnegie Mellon Pronuncing Dictionary
  - Words sampled according to corpus frequency
  - 6,409 vowel and 8,917 consonant tokens

- Acoustic values for vowels sampled based on Hillenbrand et al. (1995) data
  - Means, covariance matrices computed from speakers' productions
  - Speech sounds generated from Gaussians

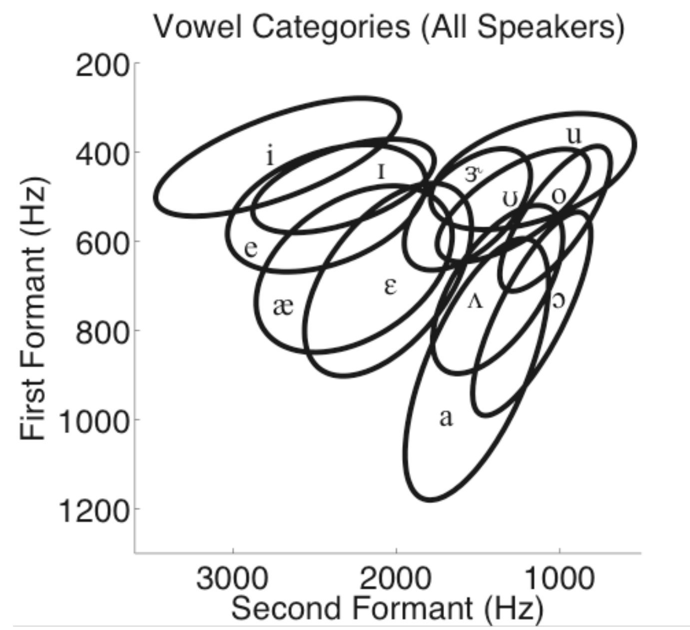


Words Can Help Ideal Learners

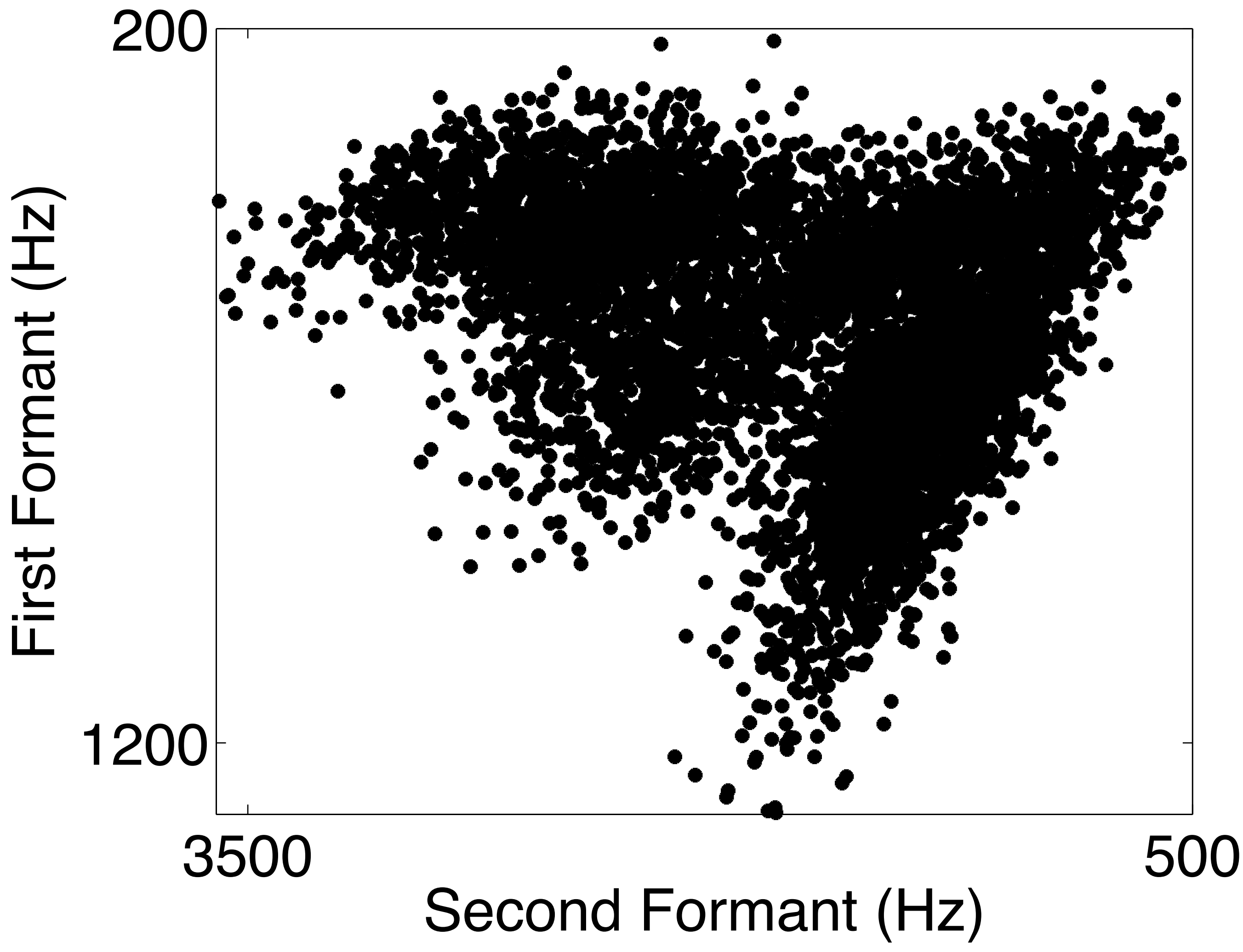
- Created corpus of 5000 word tokens from CHILDES Parental Corpus (Li & Shirai, 2000)
  - Orthographic forms phonematized using Carnegie Mellon Pronuncing Dictionary
  - Words sampled according to corpus frequency
  - 6,409 vowel and 8,917 consonant tokens
- Acoustic values for vowels sampled based on Hillenbrand et al. (1995) data
  - Means, covariance matrices computed from speakers' productions
  - Speech sounds generated from Gaussians

'school' represented as

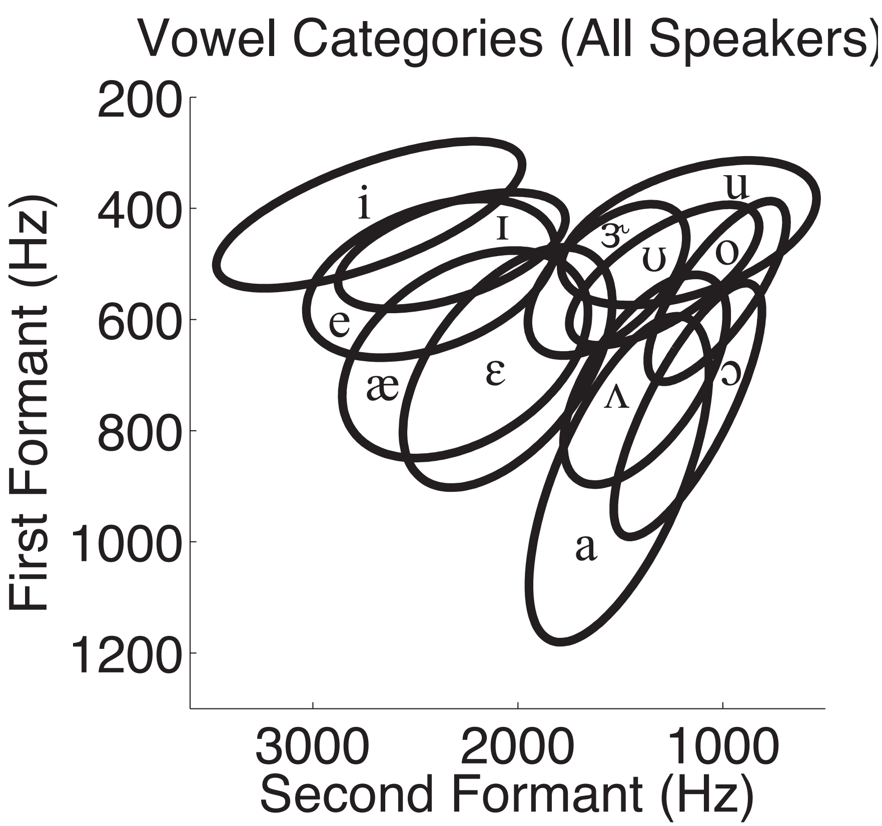
$$\mathrm{s\ k} \begin{bmatrix} 423 \\ 1152 \end{bmatrix} \mathrm{l}$$

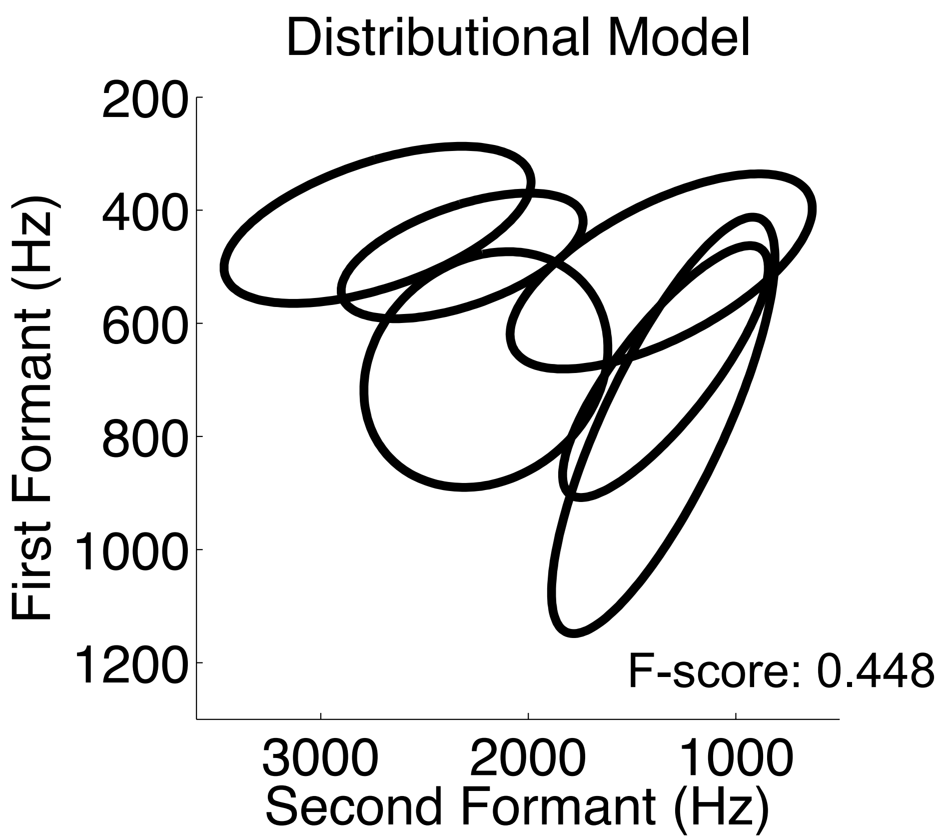


Words Can Help Ideal Learners



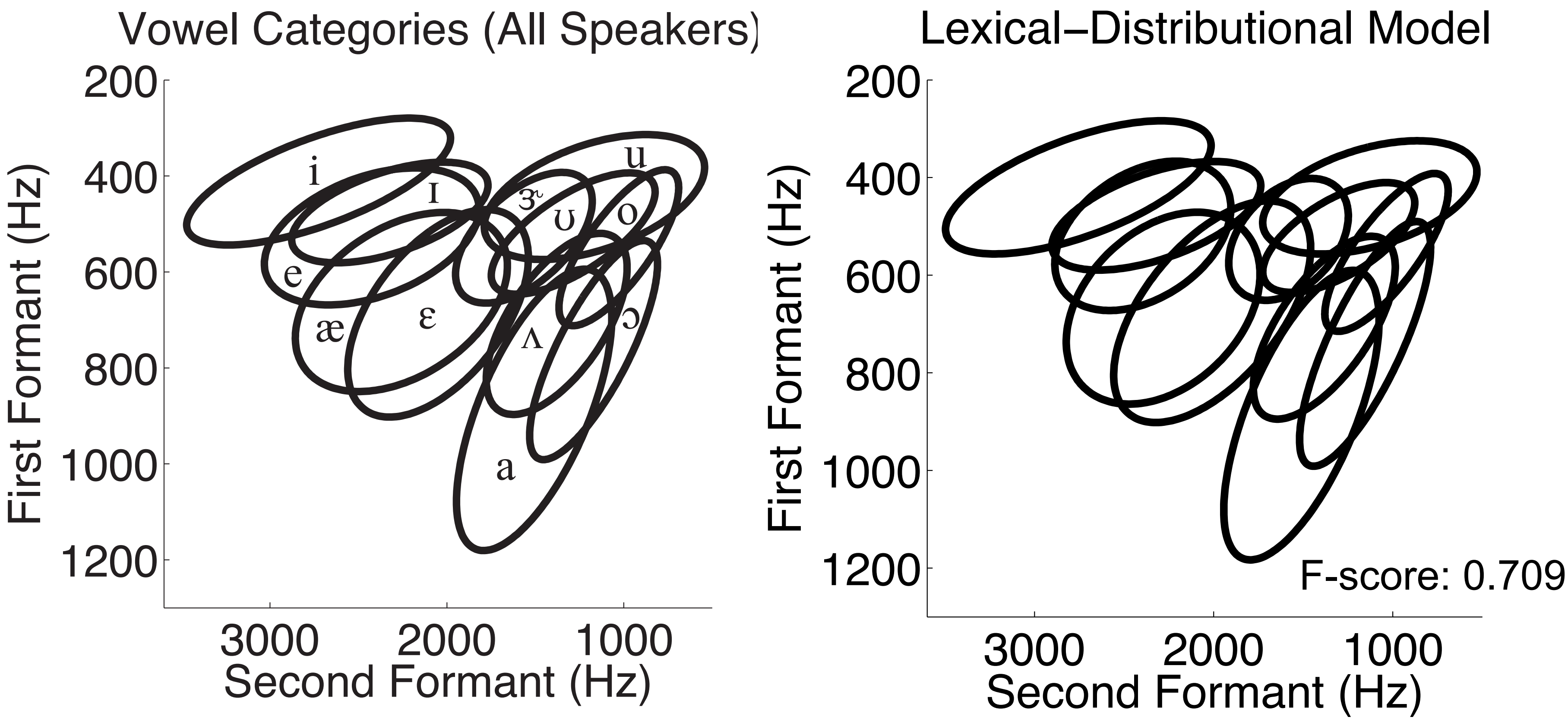
Results: Distributional





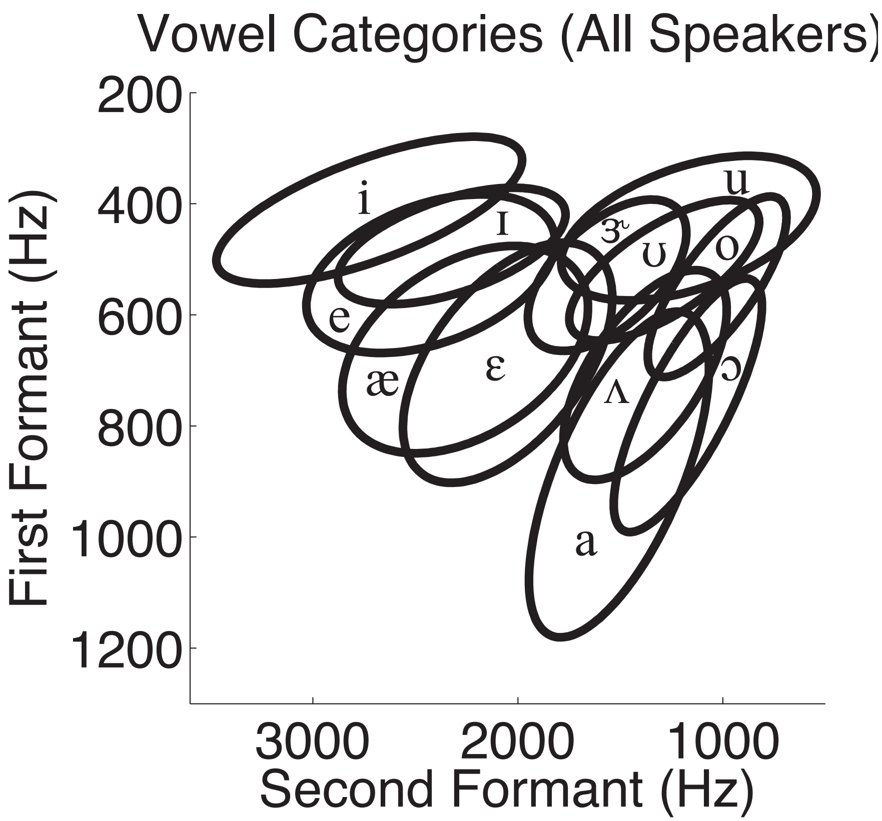
(Feldman, Griffiths, Goldwater, & Morgan, in prep)

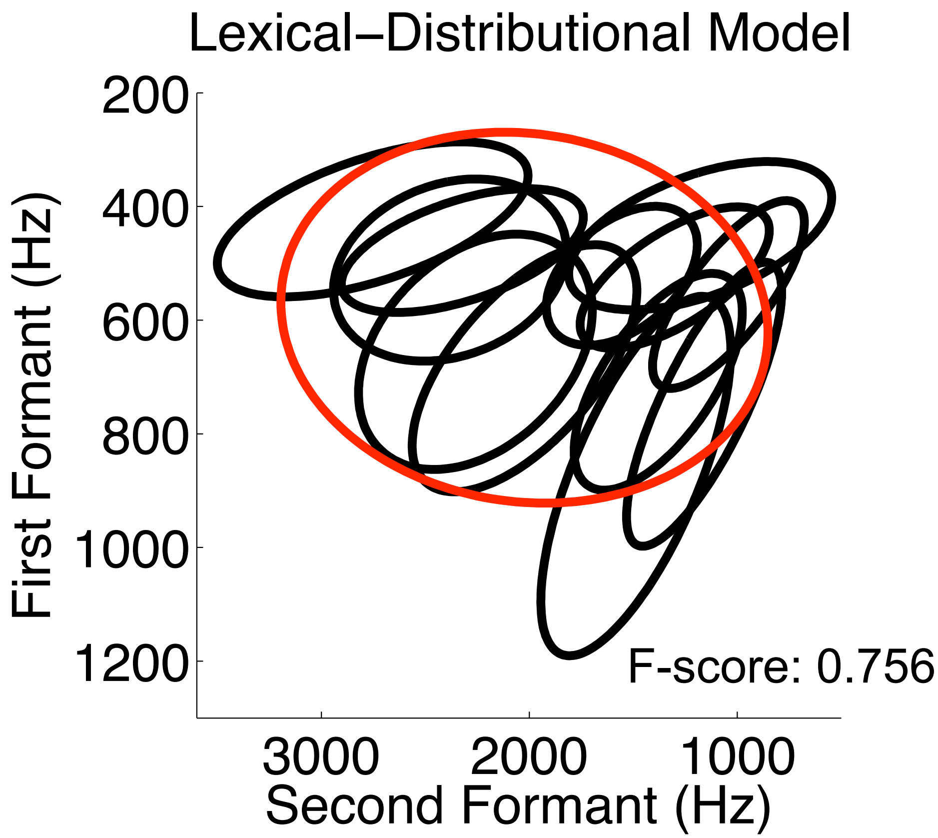
Results: Lexical-Distributional



(Feldman, Griffiths, Goldwater, & Morgan, in prep)

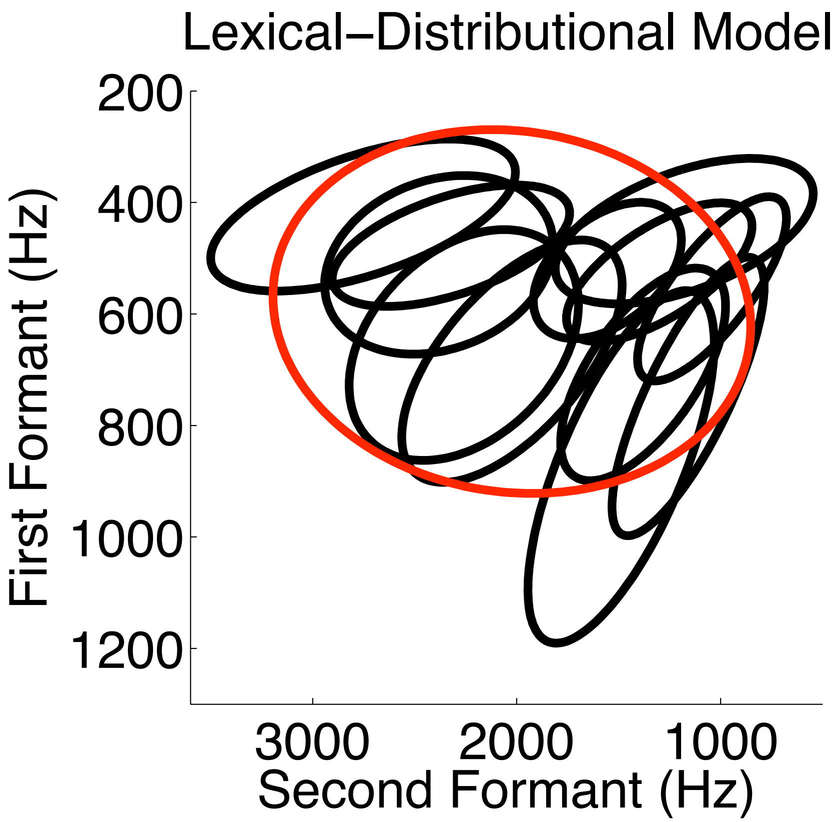
Results: Lexical-Distributional





(Feldman, Griffiths, Goldwater, & Morgan, in prep)

Mistakes from Minimal Pairs



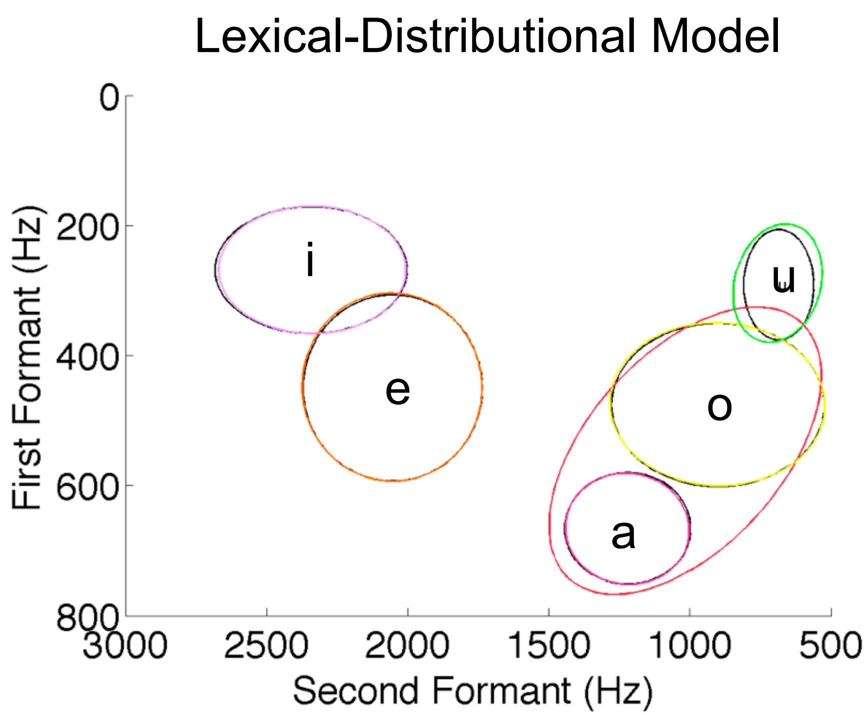
Extra category includes:

- find, found
- think, thank
- will, we'll, well
- give, gave
- made, mad, mid
- big, bag
- way, we

as well as lexical items that were not minimal pairs

(Feldman, Griffiths, Goldwater, & Morgan, in prep)

Mistakes from Minimal Pairs



- Tested on Spanish corpus
  - Categories based on production data
  - Words frequencies from CHILDES
- Model confused by morphological patterns

(Bennett, Kronrod, & Feldman, in progress)

Word and Sound Learning

- A developing lexicon can help an ideal learner separate overlapping categories
- Non-minimal pairs in the lexicon are critical for learning
- Predicts that children should be sensitive to word-level information (non-minimal pairs) during phonetic learning

Hypothesis

Infants use top-down constraints from words when acquiring phonetic categories

1. Formalize a model that can simultaneously learn sounds and words

2. Show that infants are sensitive to words in ways the model would predict

Data from Children

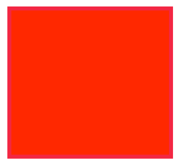
Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)

(Thiessen, 2011)

Data from Children

Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)

Habituation:



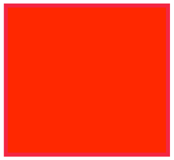
“daw”

(Thiessen, 2011)

Data from Children

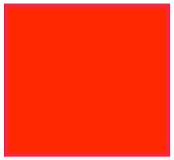
Evidence from 15-month-olds in "switch" task (Stager & Werker, 1997)

Habituation:



"daw"

Switch trial:



"taw"

(Thiessen, 2011)

Data from Children

Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)

Habituation:

“daw”

Switch trial:

“taw”

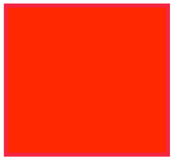
✗

(Thiessen, 2011)

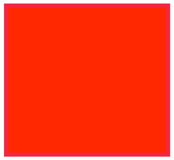
Data from Children

Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)

Familiarization: Habituation: Switch trial:



“daw”



“taw”

✗

“tawgoo”

“dawbow”

(Thiessen, 2011)

Data from Children

Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)

Familiarization:	Habituation:	Switch trial:	
	“daw”	“taw”	✗
“tawgoo” “dawbow”	“daw”	“taw”	

(Thiessen, 2011)

Data from Children

Evidence from 15-month-olds in "switch" task (Stager & Werker, 1997)

Familiarization:	Habituation:	Switch trial:	
	"daw"	"taw"	✗
"tawgoo" "dawbow"	"daw"	"taw"	✔

(Thiessen, 2011)

Data from Children

Evidence from 15-month-olds in "switch" task (Stager & Werker, 1997)

Familiarization:	Habituation:	Switch trial:	
	"daw"	"taw"	✗
"tawgoo" "dawbow"	"daw"	"taw"	✓
"tawgoo" "dawgoo"	"daw"	"taw"	

(Thiessen, 2011)

Data from Children

Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)

Familiarization:	Habituation:	Switch trial:	
	“daw”	“taw”	✗
“tawgoo” “dawbow”	“daw”	“taw”	✓
“tawgoo” “dawgoo”	“daw”	“taw”	✗

(Thiessen, 2011)

Are infants sensitive to word-level cues during the time when they are first learning phonetic categories?

Data from Children

Test 8-month-olds in a non-referential task

Non-Minimal Pair	**Minimal Pair**
gutah, gutah	gutah, gutaw
litaw, litaw	litah, litaw

(Feldman, Myers, White, Griffiths, & Morgan, submitted)

Data from Children

Test 8-month-olds in a non-referential task

Non-Minimal Pair	**Minimal Pair**
gutah, gutah	gutah, gutaw
litaw, litaw	litah, litaw

Test stimuli
Alternating trials (tah-taw-tah-taw)
Non-alternating trials (tah-tah-tah-tah, taw-taw-taw-taw)

(Feldman, Myers, White, Griffiths, & Morgan, submitted)

Data from Children

Test 8-month-olds in a non-referential task

Non-Minimal Pair	Minimal Pair
gutah, gutah	gutah, gutaw
litaw, litaw	litah, litaw

Test stimuli
Alternating trials (tah-taw-tah-taw)
Non-alternating trials (tah-tah-tah-tah, taw-taw-taw-taw)

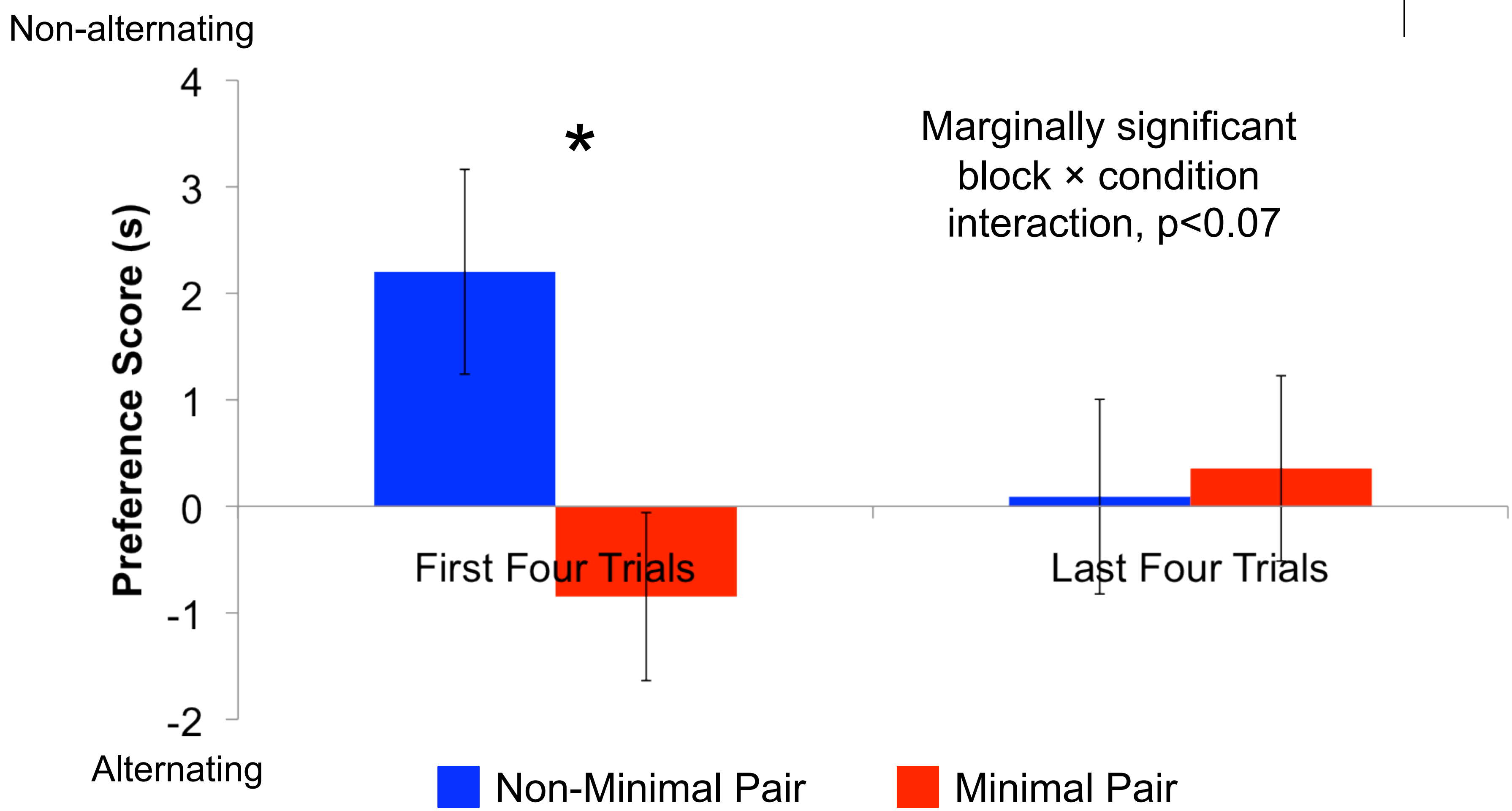
Preference Score: Non-alternating - Alternating

(Feldman, Myers, White, Griffiths, & Morgan, submitted)

Data from Children

(Feldman, Myers, White, Griffiths, & Morgan, submitted)

Data from Children



(Feldman, Myers, White, Griffiths, & Morgan, submitted)

Data from Children

Test 8-month-olds in a non-referential task

Non-Minimal Pair	**Minimal Pair**
gutah, gutah	gutah, gutaw
litaw, litaw	litah, litaw

Test stimuli

Alternating trials (tah-taw-tah-taw)

Non-alternating trials (tah-tah-tah-tah, taw-taw-taw-taw)

Non-Minimal Pair ✔ Minimal Pair ✘

(Feldman, Myers, White, Griffiths, & Morgan, submitted)

Hypothesis

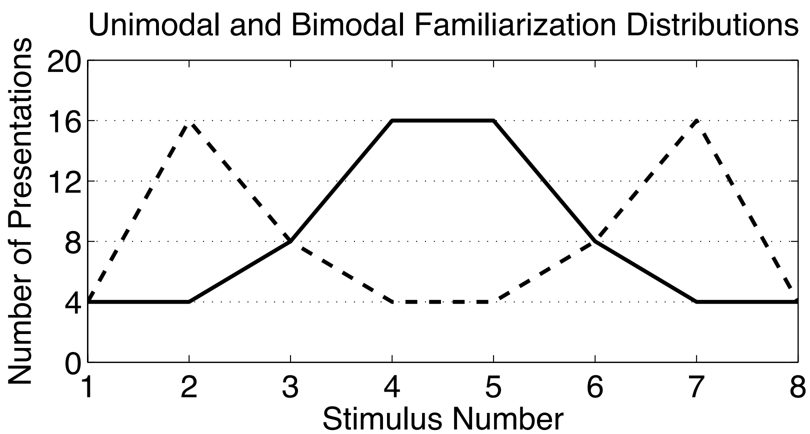
Infants use top-down constraints from words when acquiring phonetic categories

1. Formalize a model that can simultaneously learn sounds and words

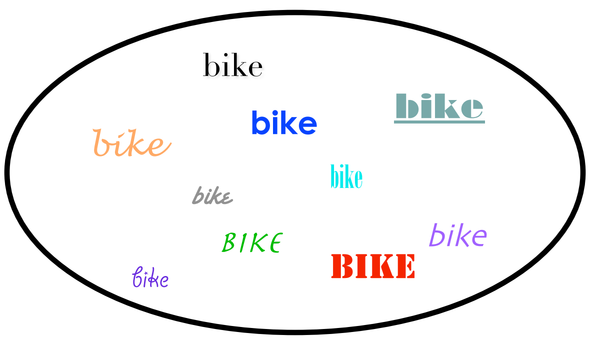
2. Show that infants are sensitive to words in ways the model would predict

Language Acquisition

Infants have a machine learning problem to solve…



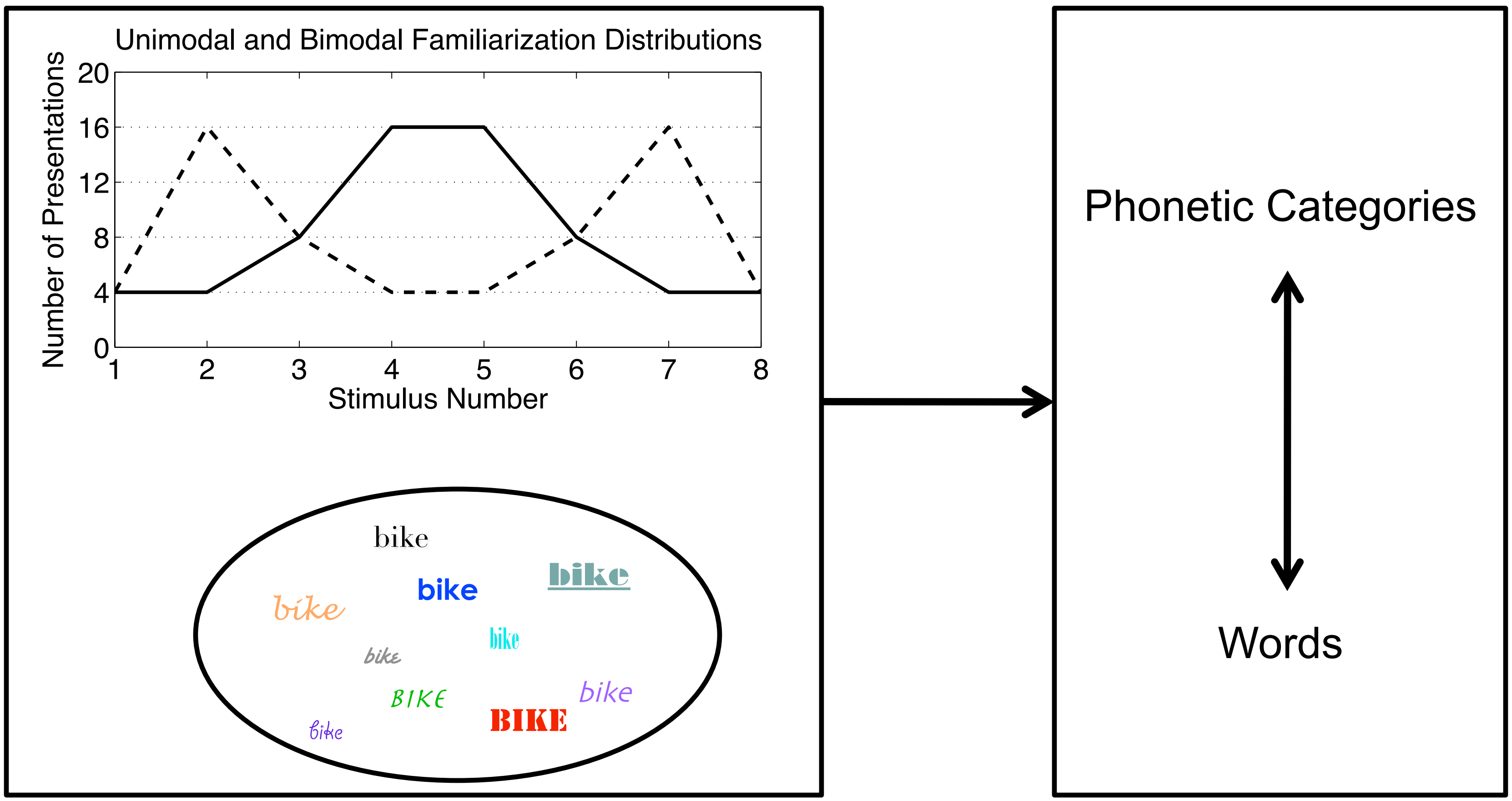
Phonetic Categories



Words

Language Acquisition

Infants have a machine learning problem to solve…



Scaling Up to Real Speech

Just defining the phonetic emission model more realistically won't work

- No way of accounting for predictable phonological variability
- Can't even do speaker normalization (not interchangeable; every sound in a word is uttered by the same speaker)
- What can we draw from this work?

Phonetic Categories

↓

Lexicon

↓

Corpus

The Right Generative Model

Distributional Model

Phonetic Categories

↓

Corpus

Lexical-Distributional Model

Phonetic Categories

↓

Lexicon

↓

Corpus

The Right Generative Model

Distributional Model

Phonetic Categories

↓

Corpus

Lexical-Distributional Model

Phonetic Categories

↓

Lexicon

↓

Corpus

Mixture of Linear Models

Phonemes

↓

Corpus

(Dillon, Dunbar, & Idsardi, in press)

The Right Generative Model

Distributional Model

Phonetic Categories

↓

Corpus

Lexical-Distributional Model

Phonetic Categories

↓

Lexicon

↓

Corpus

Mixture of Linear Models

Phonemes

↓

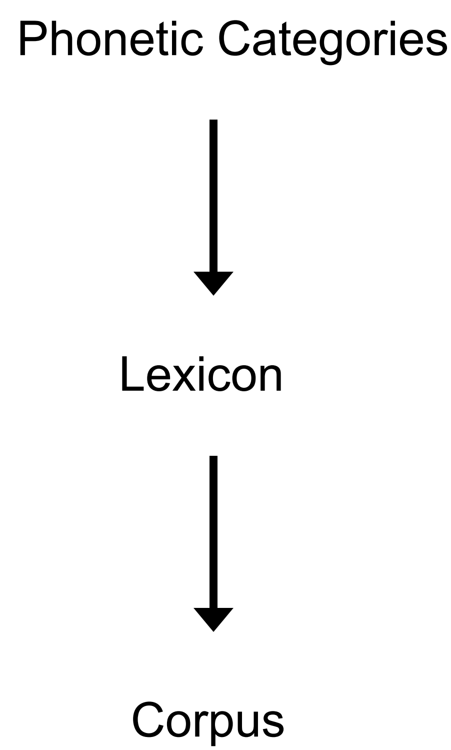
Corpus

?

(Dillon, Dunbar, & Idsardi, in press)

The Right Generative Model

- Phonetic categories are part of the prior distribution over lexical items
  - Can be learned by computing statistics over the developing lexicon
  - Knowledge of phonetic categories allows rapid learning of new words, without re-estimating parameters each time
- Many other parts of this prior distribution
  - Anything that can best be learned by computing statistics over the lexicon
  - Phonological alternations, phonotactics, morphology



Evidence from Bilinguals

- Catalan-Spanish bilinguals' vowel discrimination shows a surprising pattern (Bosch & Sebastián-Gallés, 2003; Sebastián-Gallés & Bosch, 2009):

Contrast	4 months	8 months	12 months
[e]-[ɛ] (Catalan only)	✔	✘	✔
[o]-[u] (both languages)	✔	✘	✔

- S&B (2009) suggested 8-month-olds might be confused because of high number of cognates e.g., 'boat' [barko]~[barku]

- Preliminary results: Need a way to represent parallels across languages

Erin Bennett
Yakov Kronrod

Prior Distribution Over Words

- What is the form of this prior distribution?
- Can we learn the form of the prior distribution, rather than specifying it in advance?
- How can we define a prior distribution that represents parallels across multiple languages?

Acknowledgments

Lexical-distributional model
Joint work with Tom Griffiths, Sharon Goldwater, James Morgan

Infant experiment
Joint work with Emily Myers, Katherine White, Tom Griffiths, James Morgan

Bilingual learning
Joint work with Erin Bennett, Yakov Kronrod

This work was funded in part by:
NSF grants BCS-0924821, BCS-0631518
NIH grant HD032005
AFOSR grant FA9550-07-1-0351