Phonetic Category Acquisition

Modeling ~~Phonology~~

Naomi Feldman
University of Maryland

IPAM: Probabilistic Models of Cognition
University of California, Los Angeles
July 14, 2011

Listen Carefully!

Hindi:

A B X

Listen Carefully!

Hindi:

A B X

Answer: A

dental vs. retroflex contrast (A=dental, B=retroflex)

Listen Carefully!

Salish:

A B X

Listen Carefully!

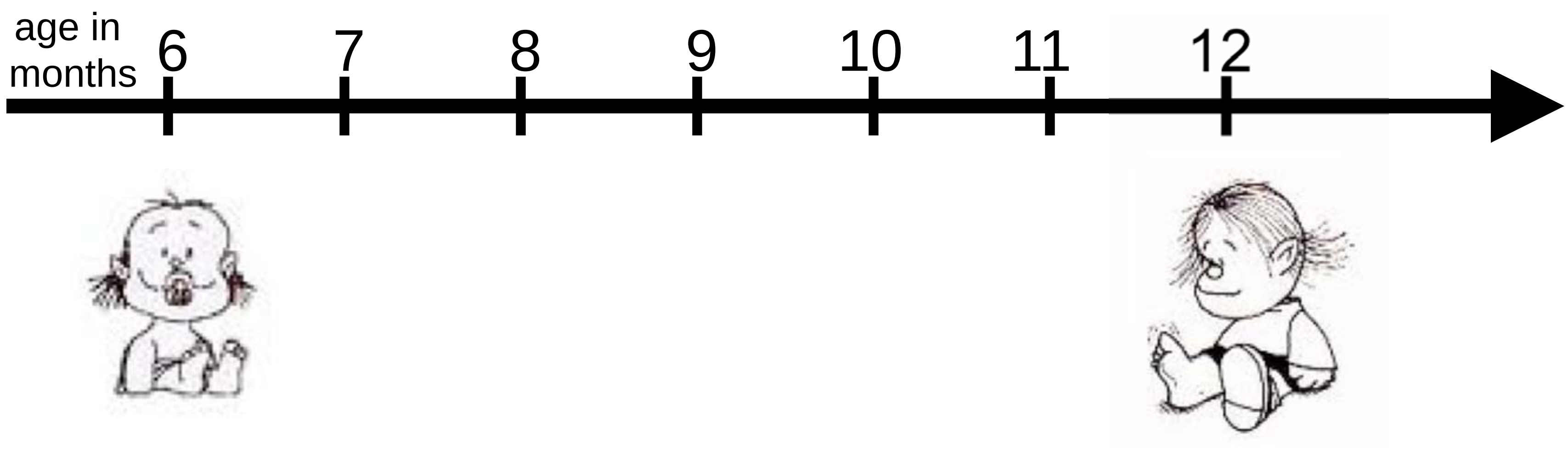
Salish:

A B X

Answer: A

velar vs. uvular contrast (A=velar, B=uvular)

Learning Sound Categories

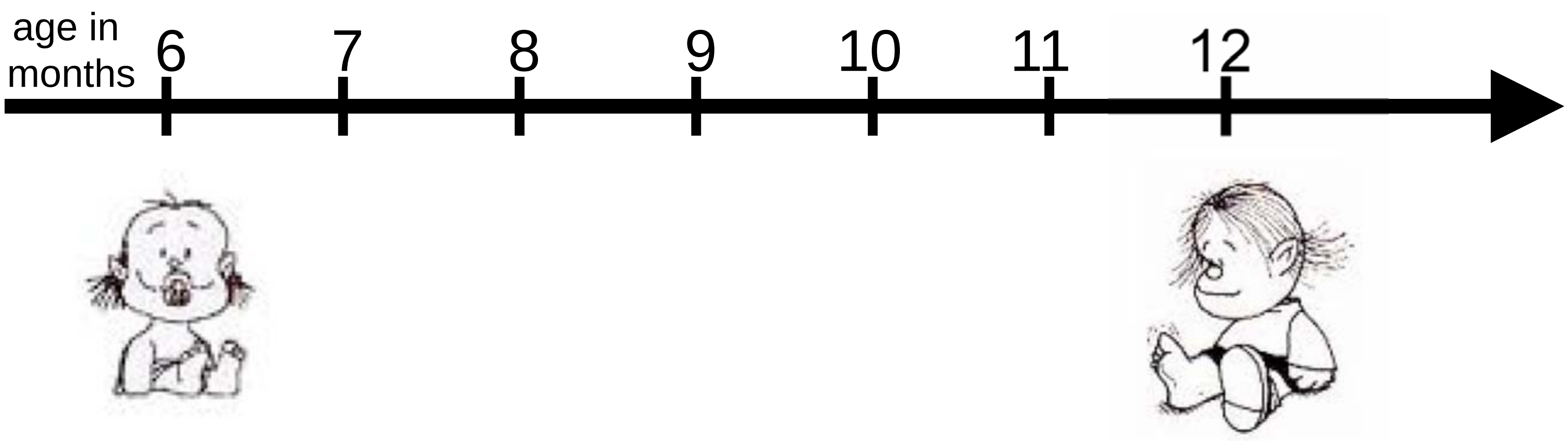


Learning Sound Categories

6-8 months:
discriminate non-native consonant contrasts

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

(Werker & Tees, 1984)



Learning Sound Categories

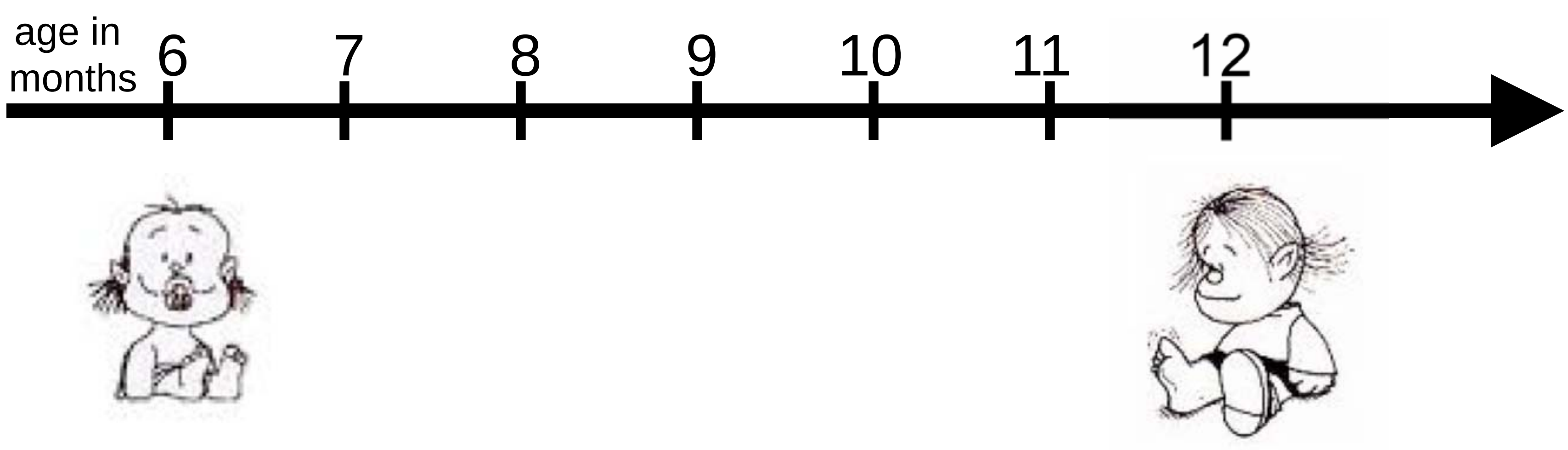
6 months: some language-specific perception of vowels

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?

An Inference Problem

Learner recovering linguistic structure

Hypotheses: possible linguistic analyses
Data: corpus (language input)

$$p(h \mid d) \propto p(d \mid h)\,p(h)$$

An Inference Problem

Learner recovering linguistic structure

Hypotheses: possible linguistic analyses
Data: corpus (language input)

$$p(h \mid d) \propto p(d \mid h)p(h)$$

What types of hypotheses should learners consider?

Outline

- Distributional learning
- Lexical-distributional learning
- Learning English vowels
- Dealing with systematic variability

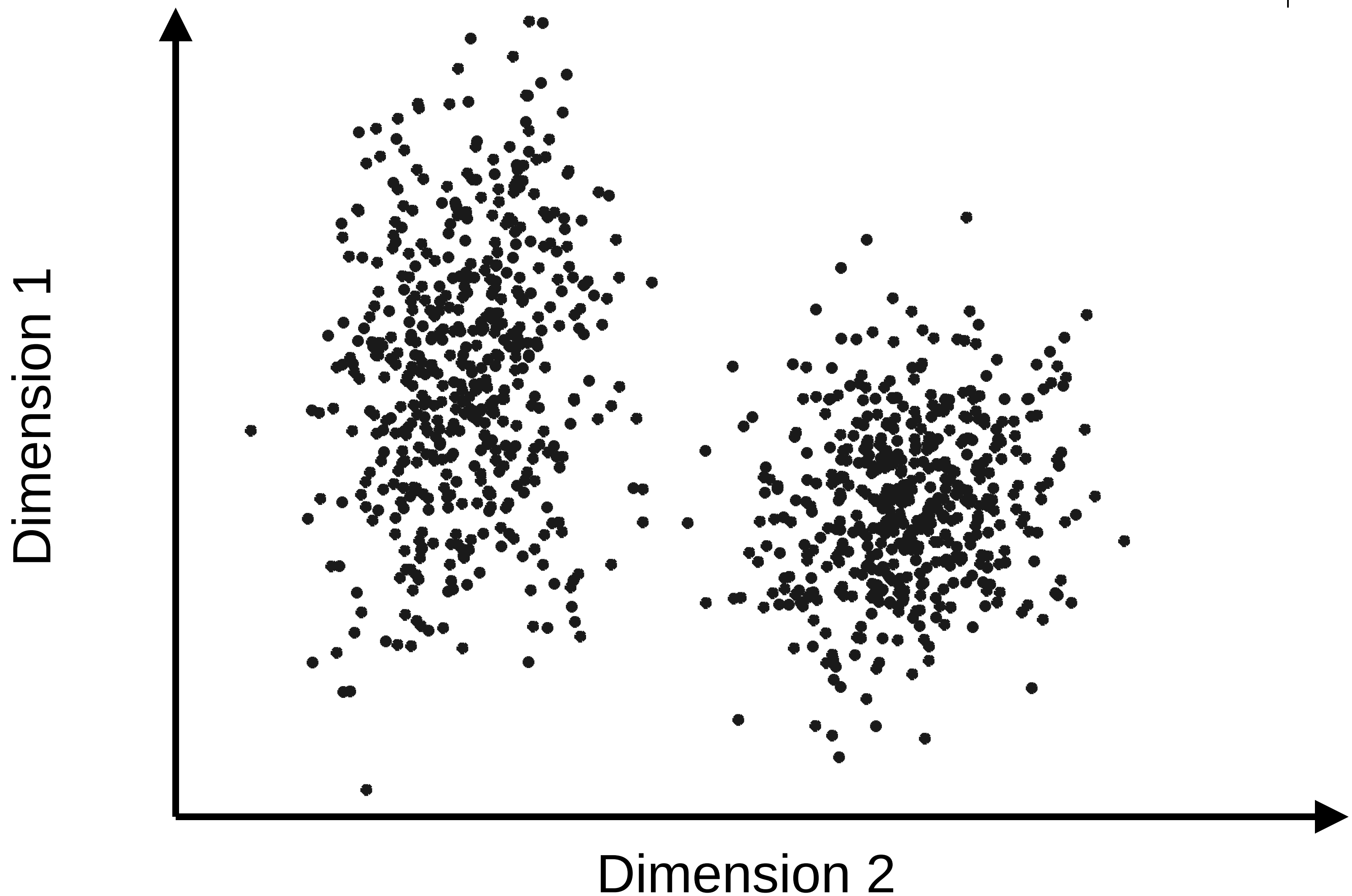
(Joint work with Tom Griffiths, James Morgan, Sharon Goldwater)

Outline

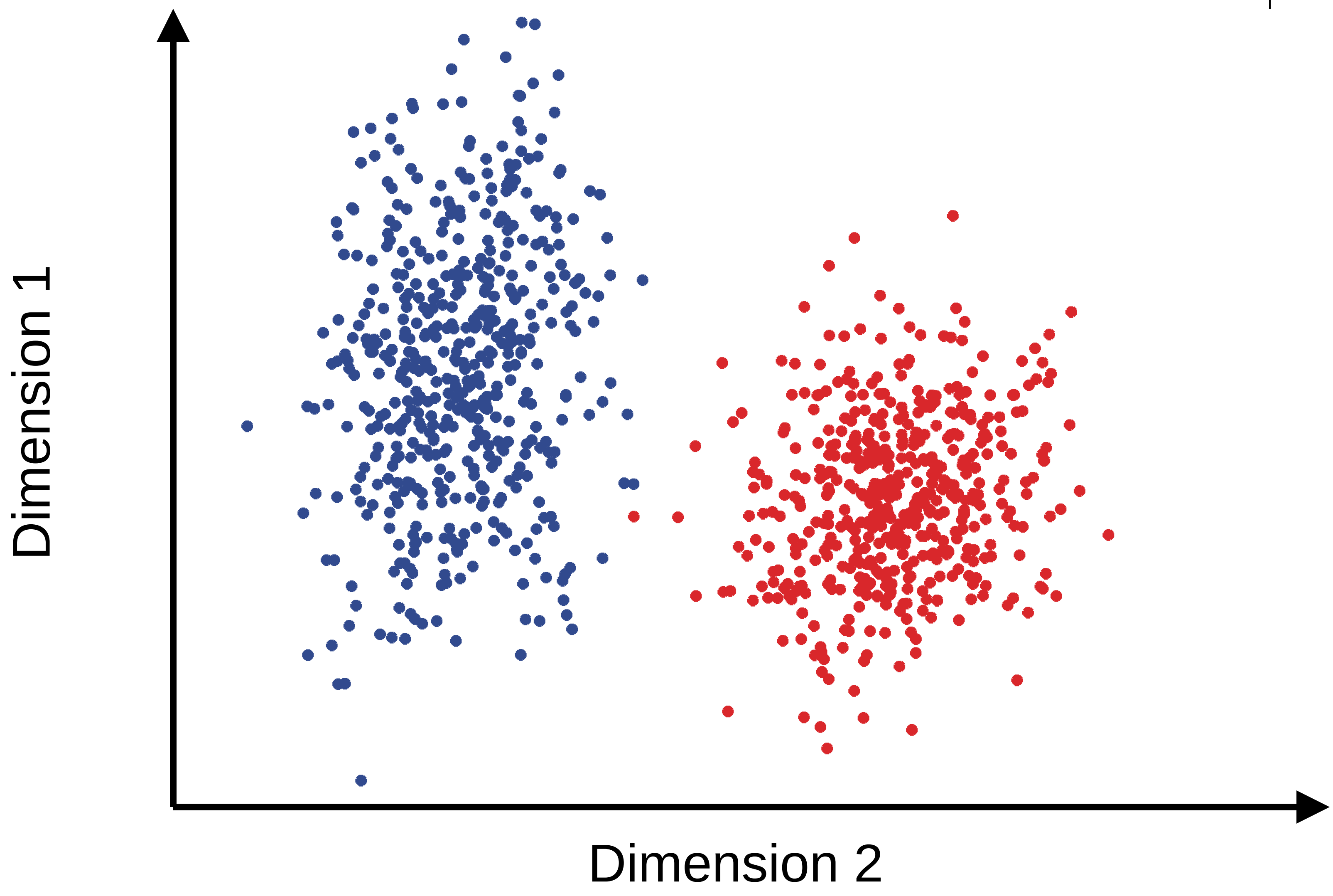
- Distributional learning
- Lexical-distributional learning
- Learning English vowels
- Dealing with systematic variability

(Joint work with Tom Griffiths, James Morgan, Sharon Goldwater)

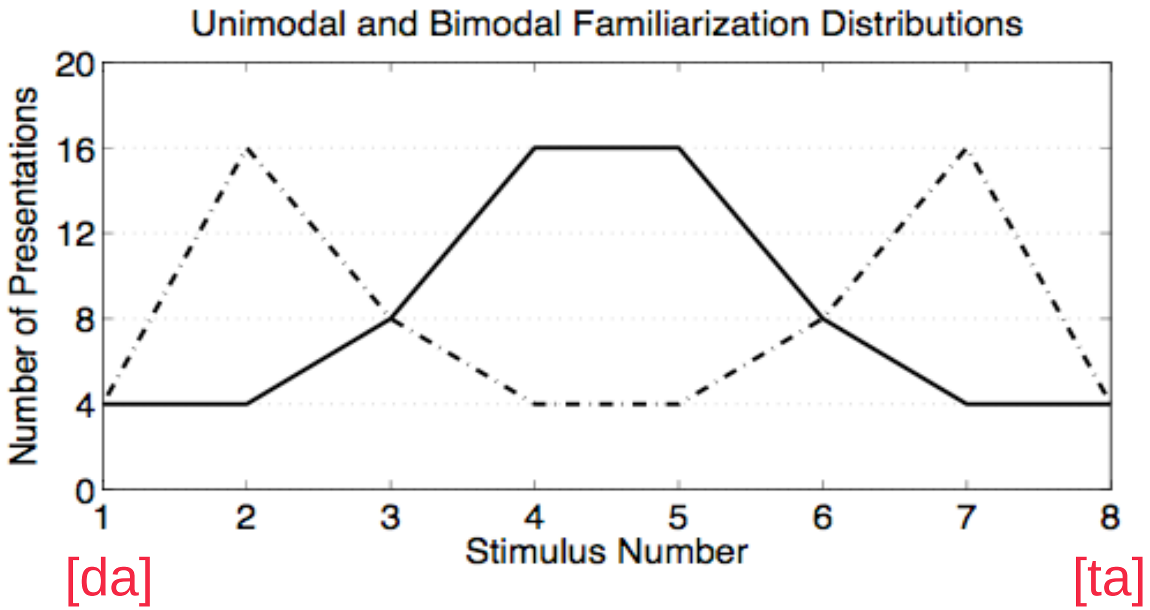
Distributional Learning



Distributional Learning



Distributional Learning



Bimodal group: good discrimination between endpoints
Unimodal group: poor discrimination between endpoints

(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

<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

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

$\downarrow$

Corpus

A Generative Model

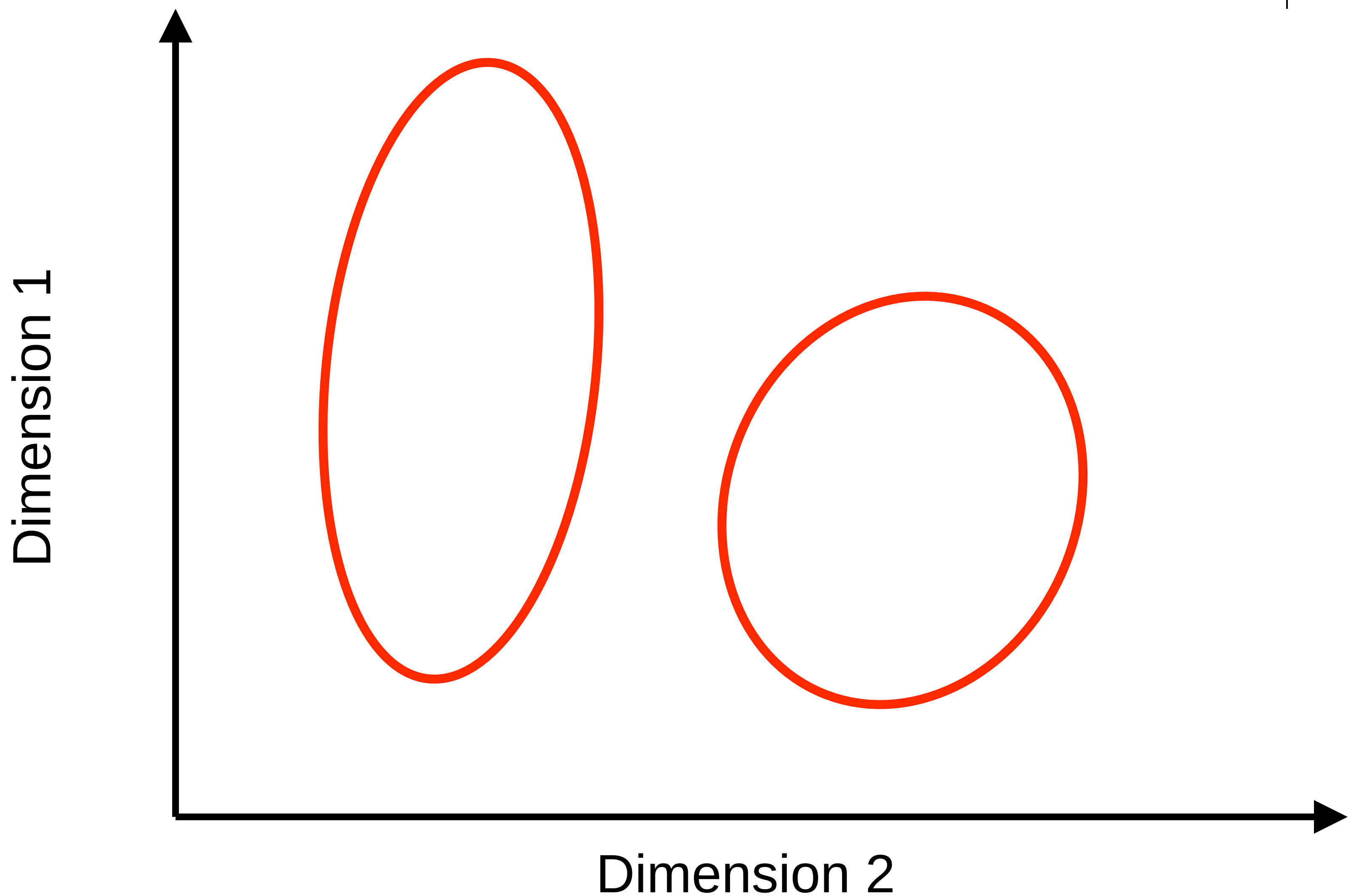
<u>To create a corpus</u>

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

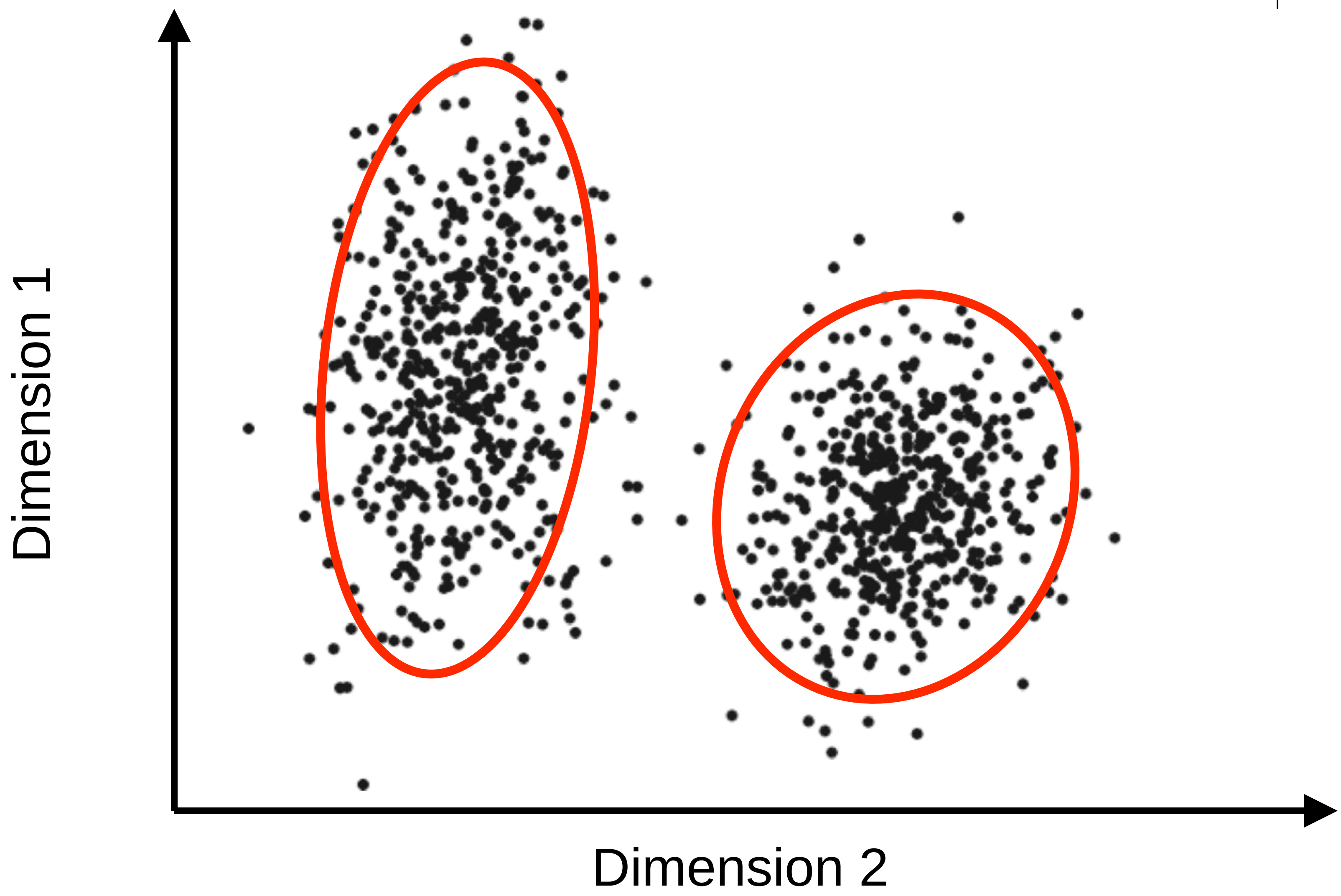
Phonetic Categories

Corpus

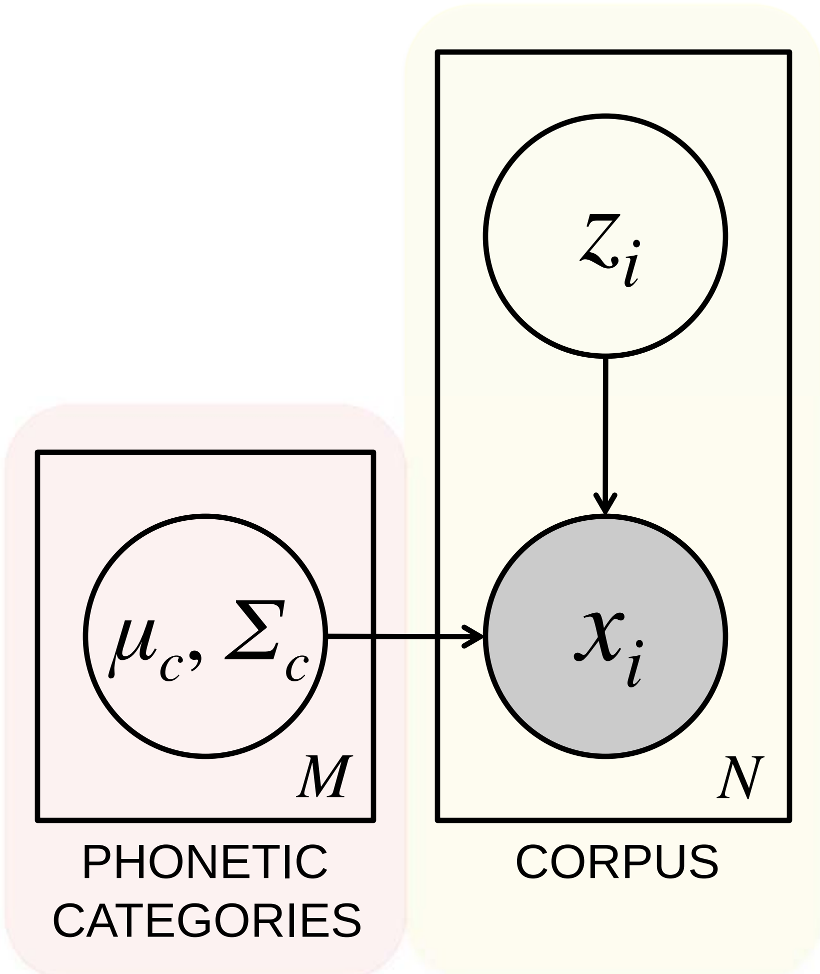
A Generative Model



A Generative Model



A Generative Model

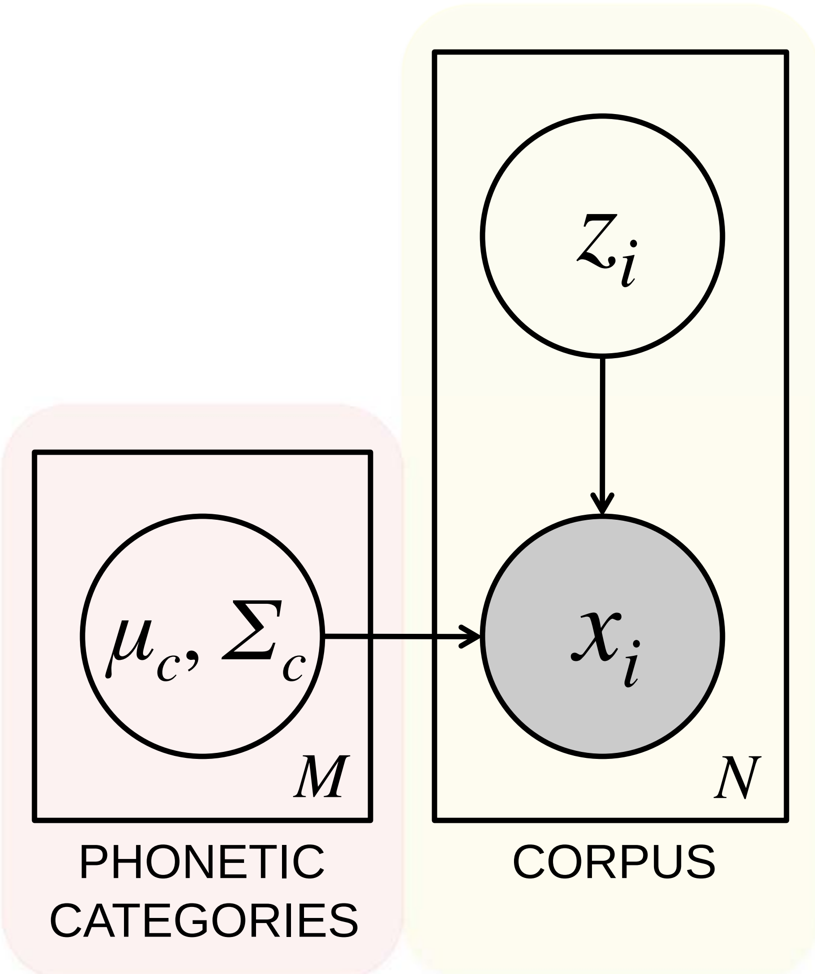


μ_c, Σ_c : parameters of category c

z_i : category of sound i

x_i : acoustics of sound i

A Generative Model



μ_c, Σ_c : parameters of category c

z_i : category of sound i

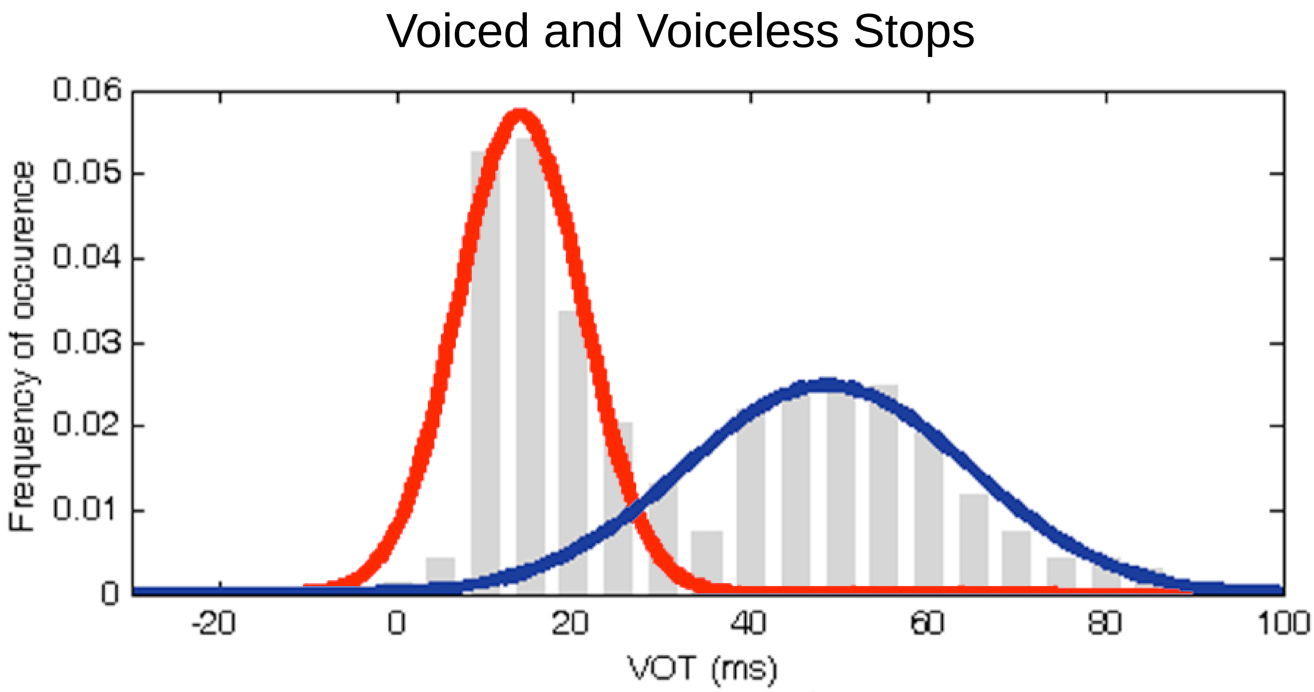
x_i : acoustics of sound i

Need to infer hidden variables:

- Parameters for each category
- Category label for each point

Can use Expectation Maximization, Gibbs sampling, online gradient descent, etc.

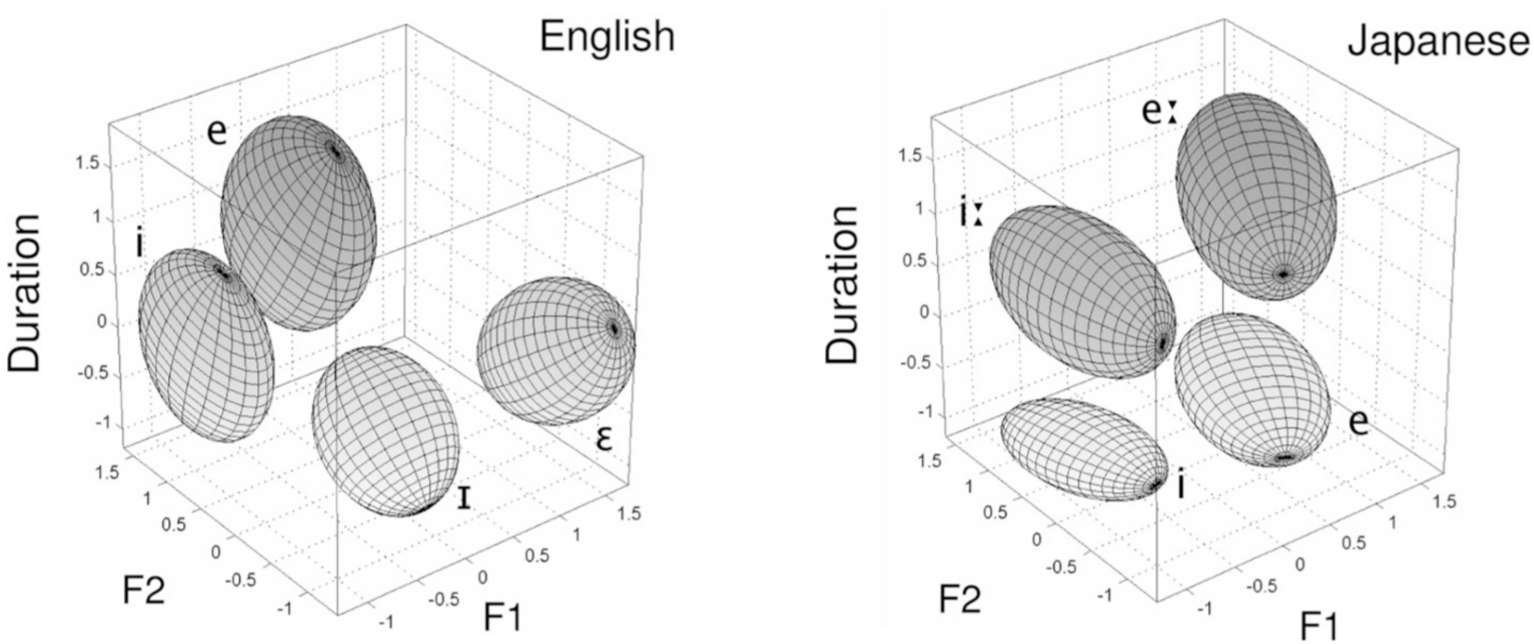
Distributional Learning



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

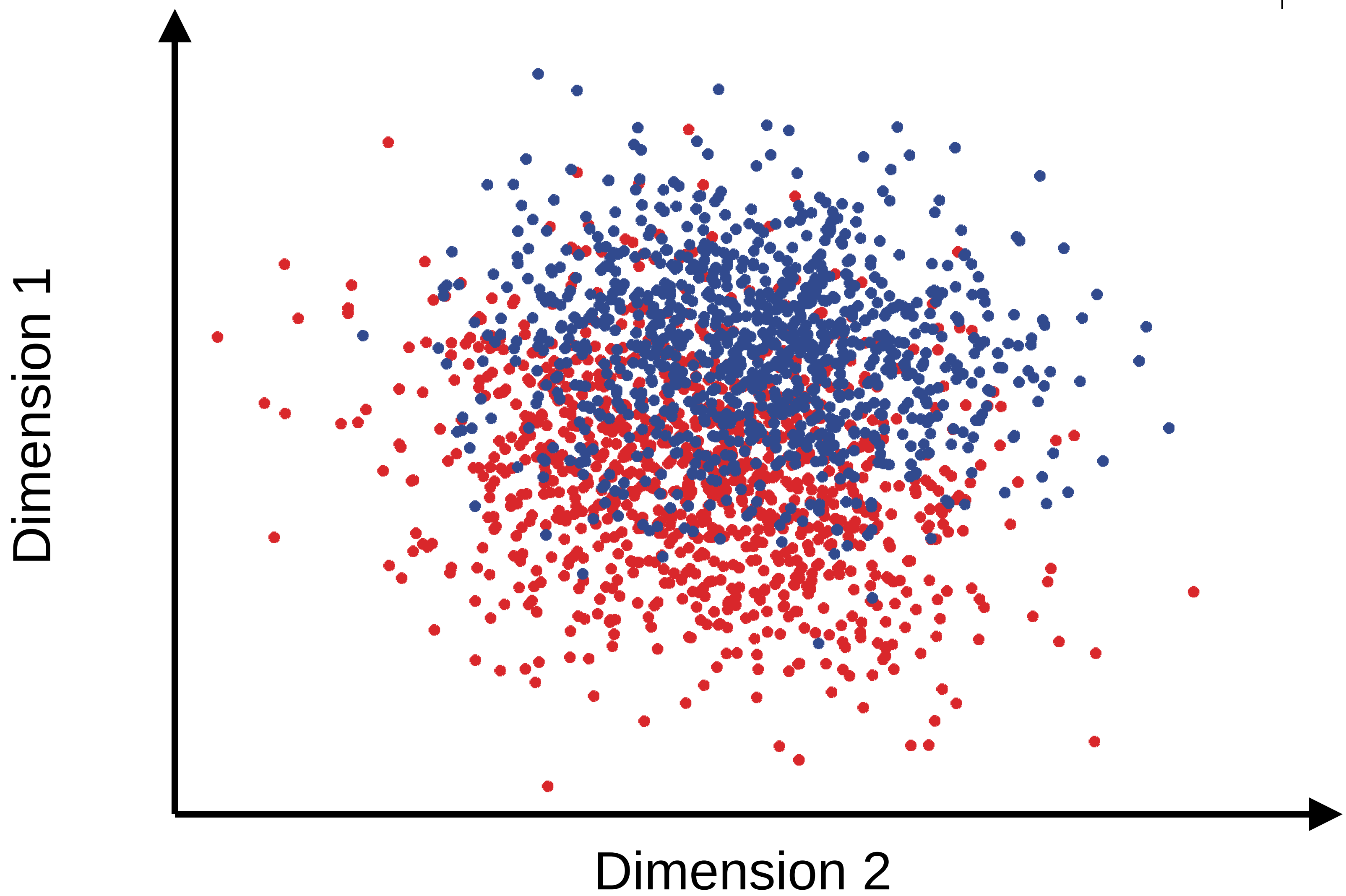
Distributional Learning

Vowel Categories (Single Speakers)

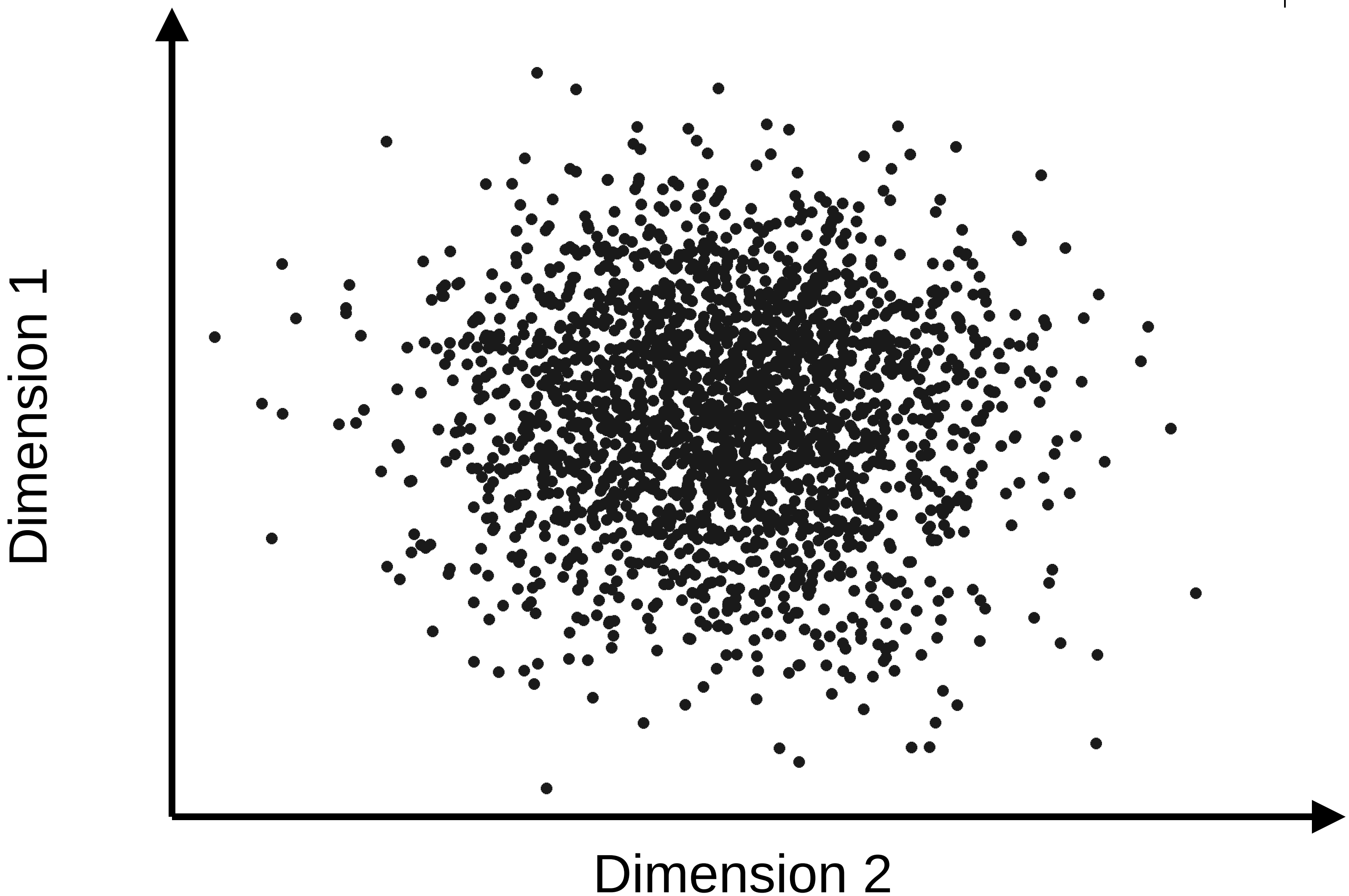


(Vallabha, McClelland, Pons, Werker, & Amano, 2007)

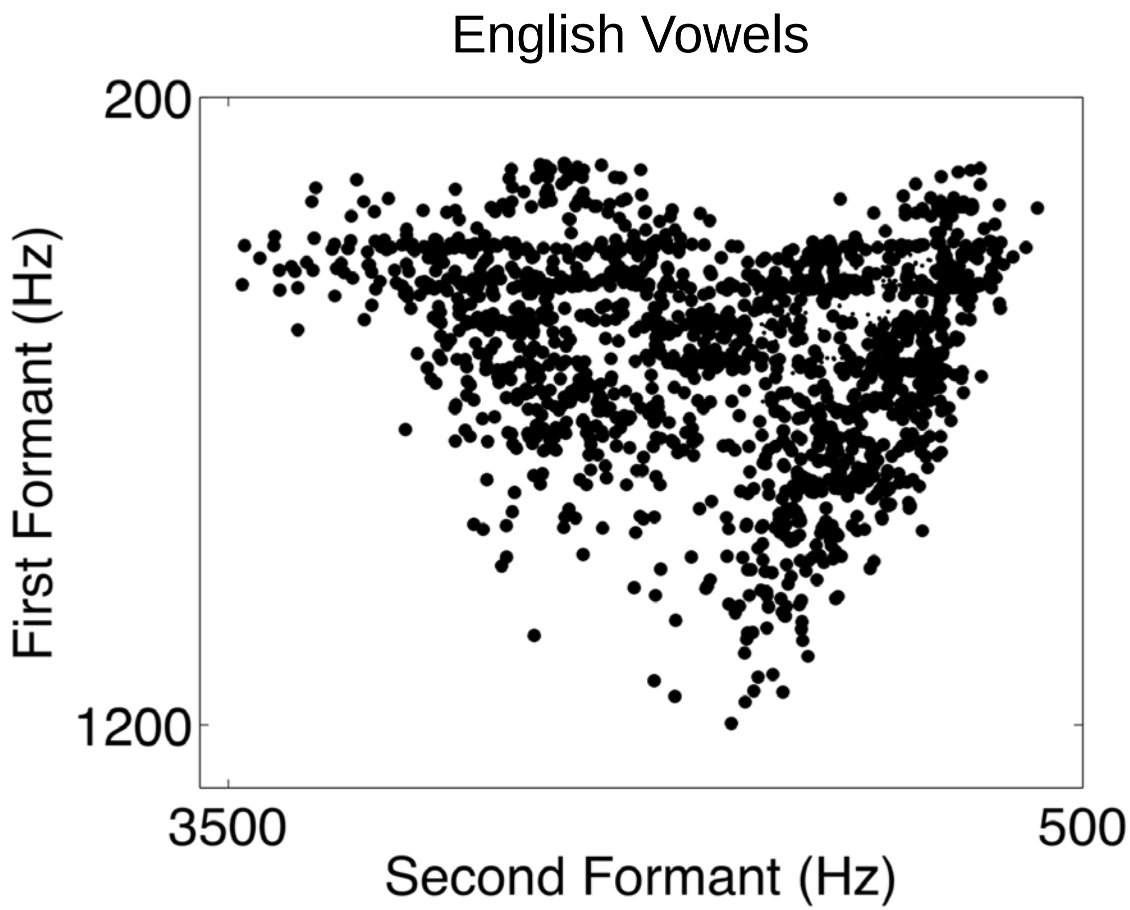
Overlapping Categories



Overlapping Categories

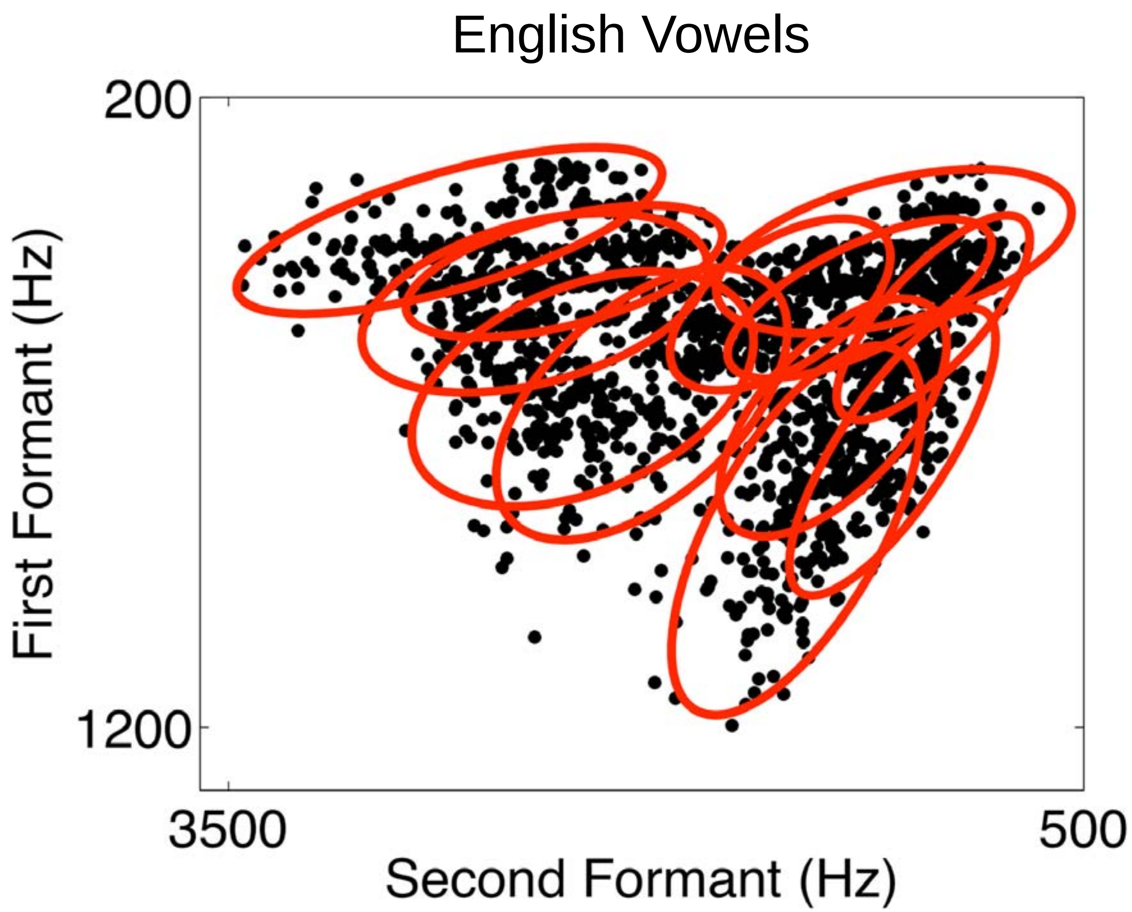


A Difficult Problem



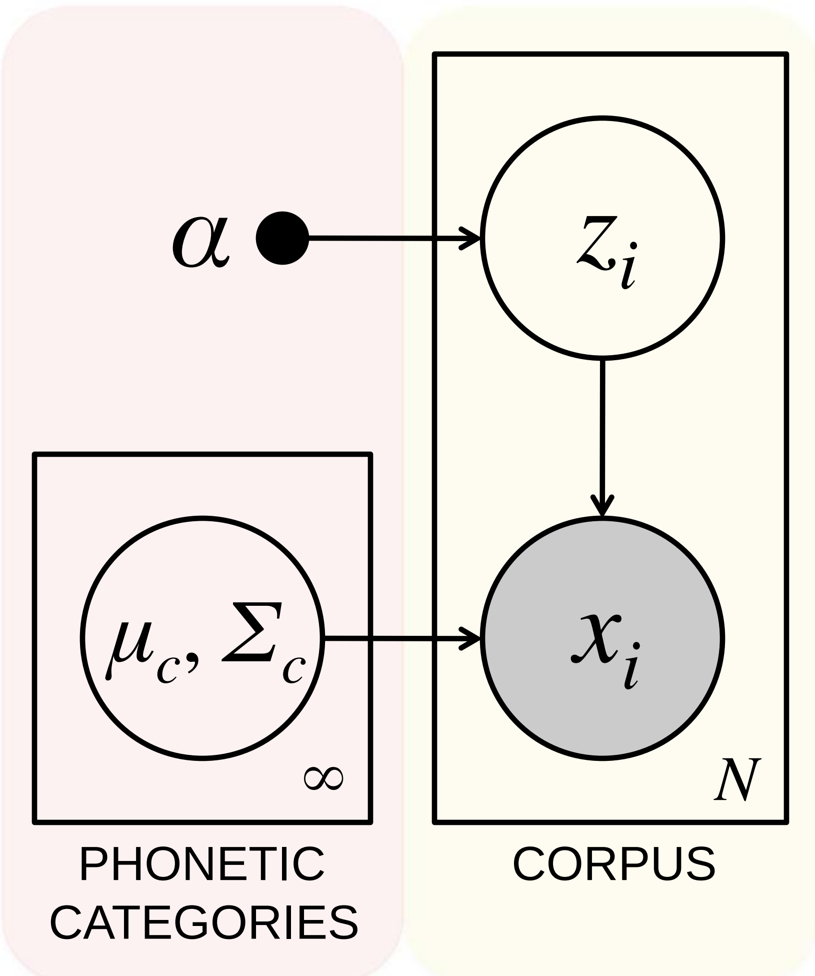
(Hillenbrand, Getty, Clark, & Wheeler, 1995)

A Difficult Problem



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

A Fancier Generative Model



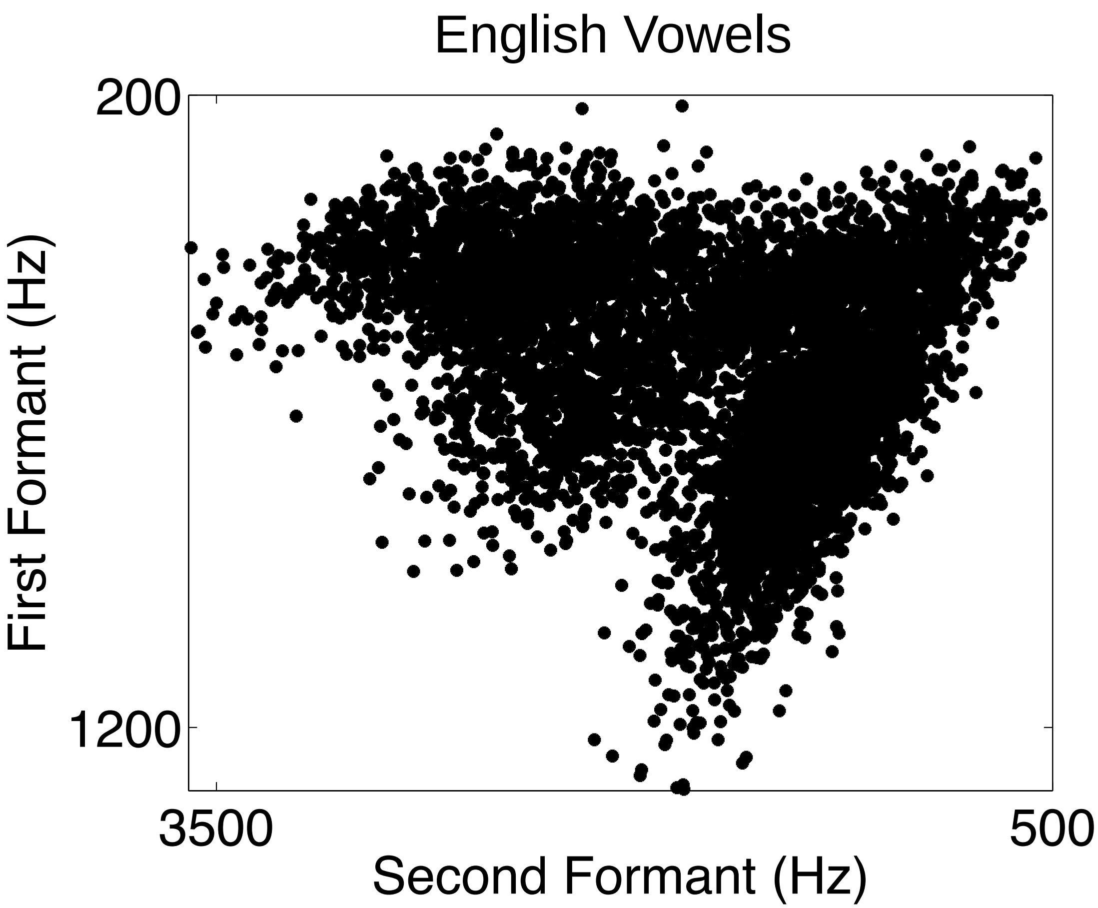
μ_c, Σ_c : parameters of category c

z_i : category of sound i

x_i : acoustics of sound i

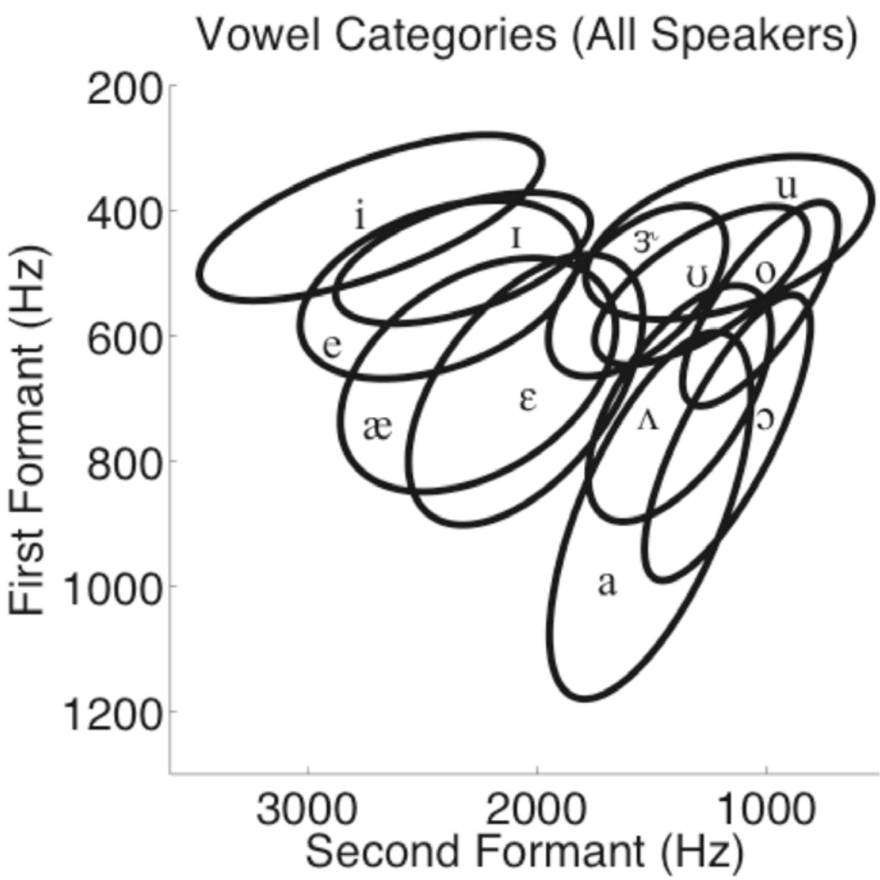
α : concentration parameter

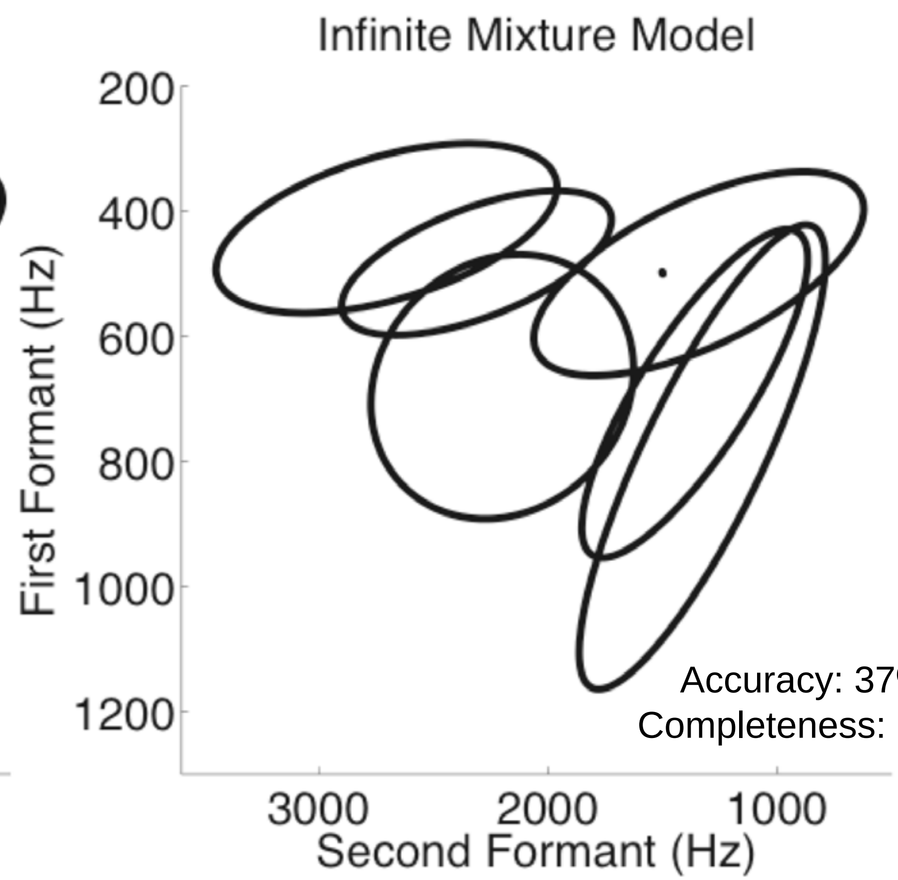
Training Corpus



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

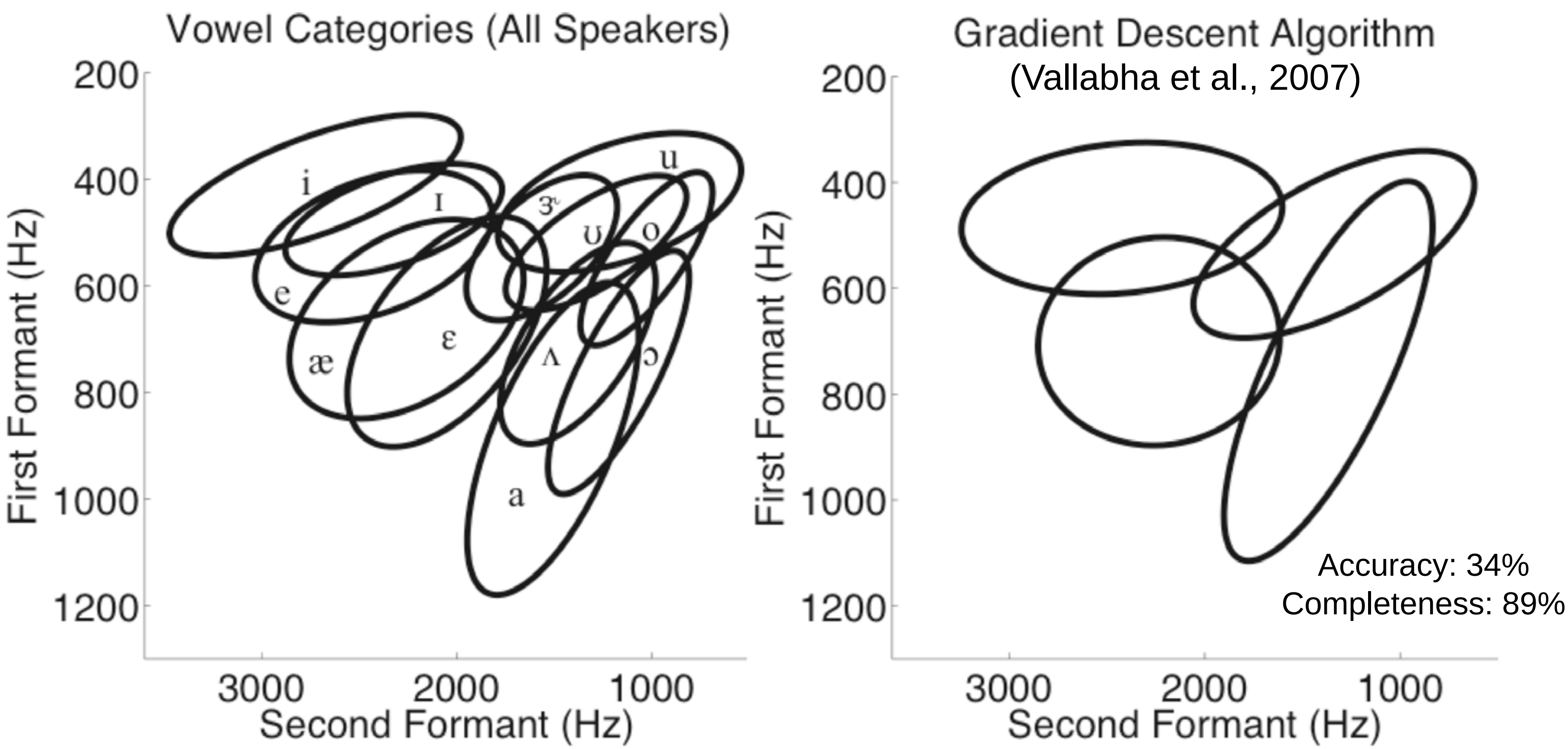
Distributional Learning





Accuracy: 37%
Completeness: 57%

Distributional Learning



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

$\downarrow$

Corpus

Outline

- Distributional learning
- Lexical-distributional learning
- Learning English vowels
- Dealing with systematic variability

(Joint work with Tom Griffiths, James Morgan, Sharon Goldwater)

Word Segmentation Task

Familiarization:

Test:

familiar unfamiliar

“Success”: Difference in looking times between familiar and unfamiliar words in fluent speech

Word Learning

bike

BIKE

bike

bike

bike

bike

bike

bike

bike

BIKE

Word Categorization

bike

BIKE

bike

bike

bike

bike

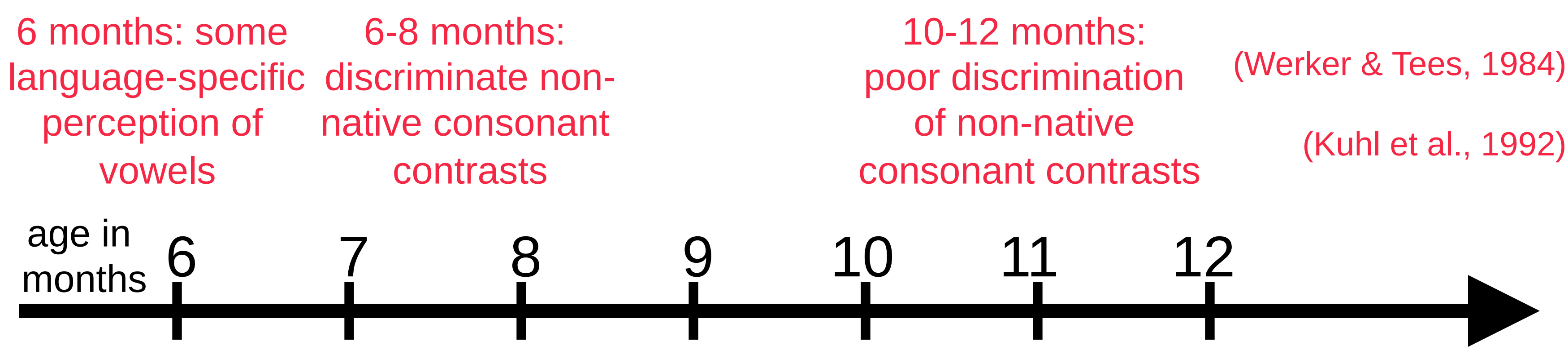
bike

bike

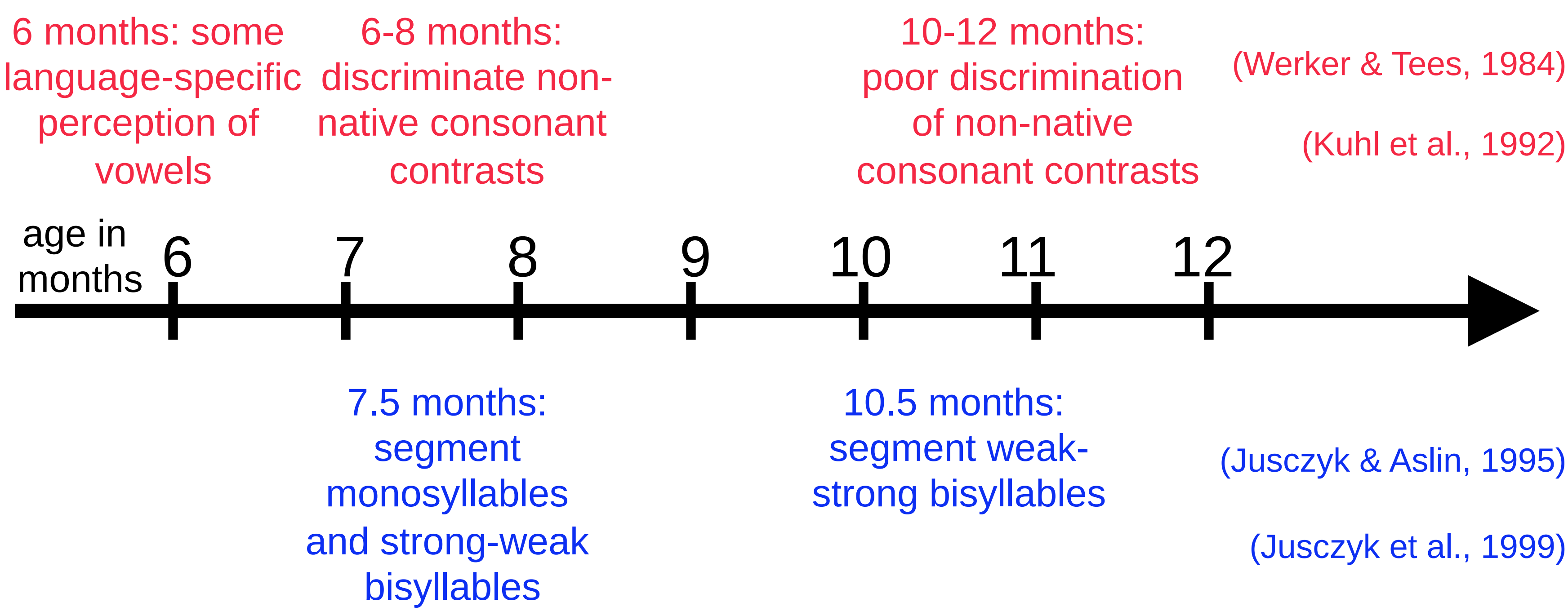
bike

BIKE

Phonetic Category Learning



Phonetic Category Learning



Phonetic Category Learning

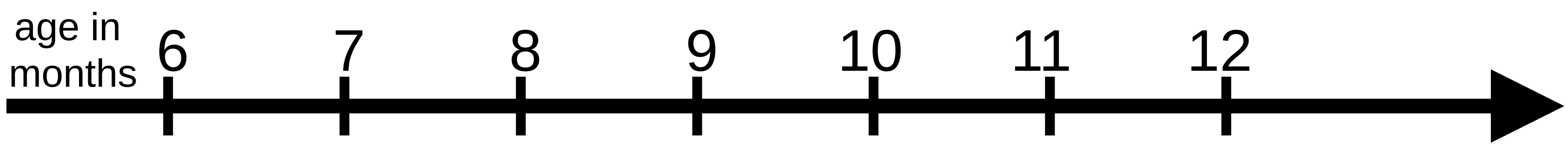
6 months: some language-specific perception of vowels

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)

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 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

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

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

$\downarrow$

Lexicon

$\downarrow$

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

<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

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

<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
   - Generate an acoustic value each phonetic category contained in that lexical item

Phonetic Categories

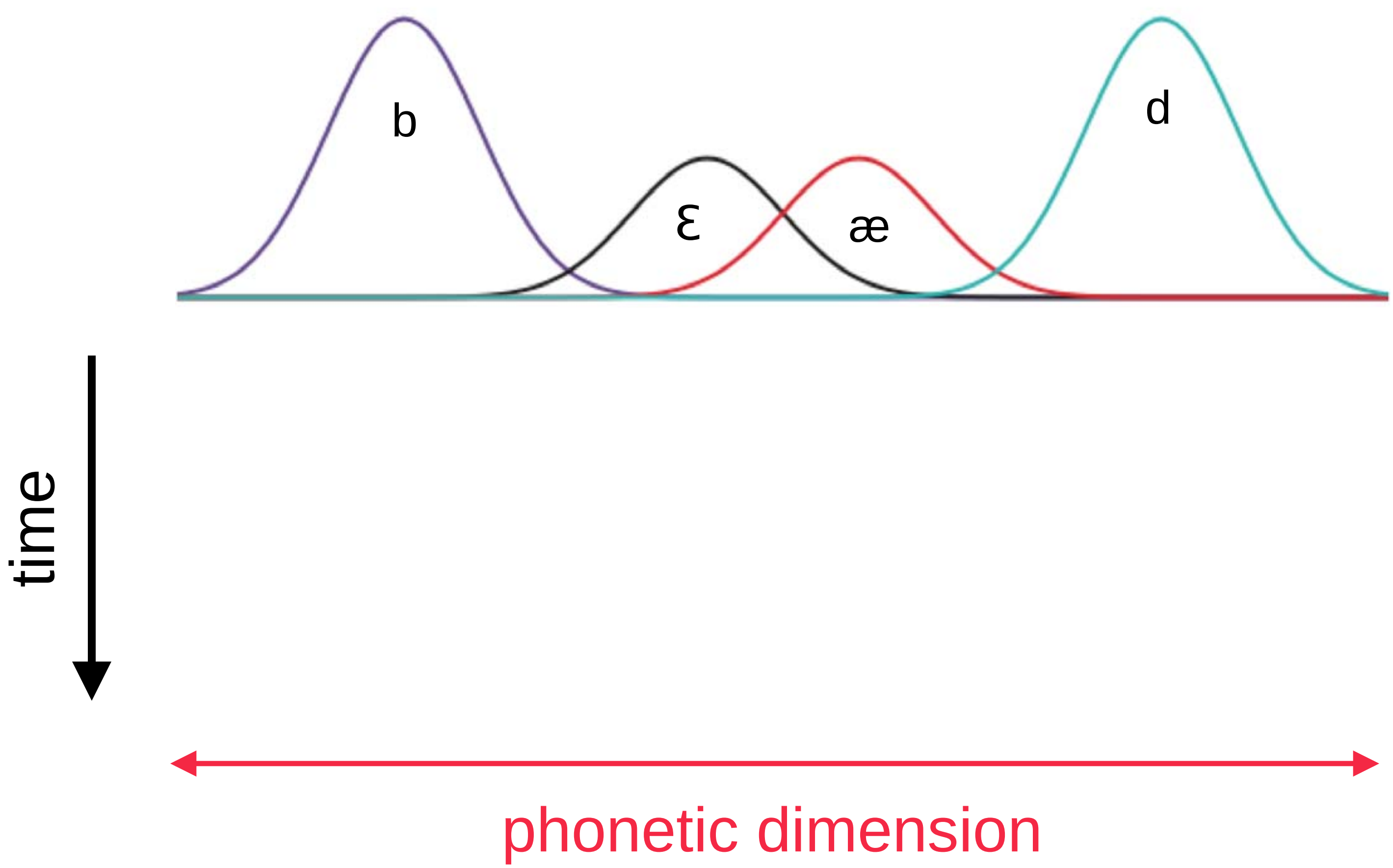
↓

Lexicon

↓

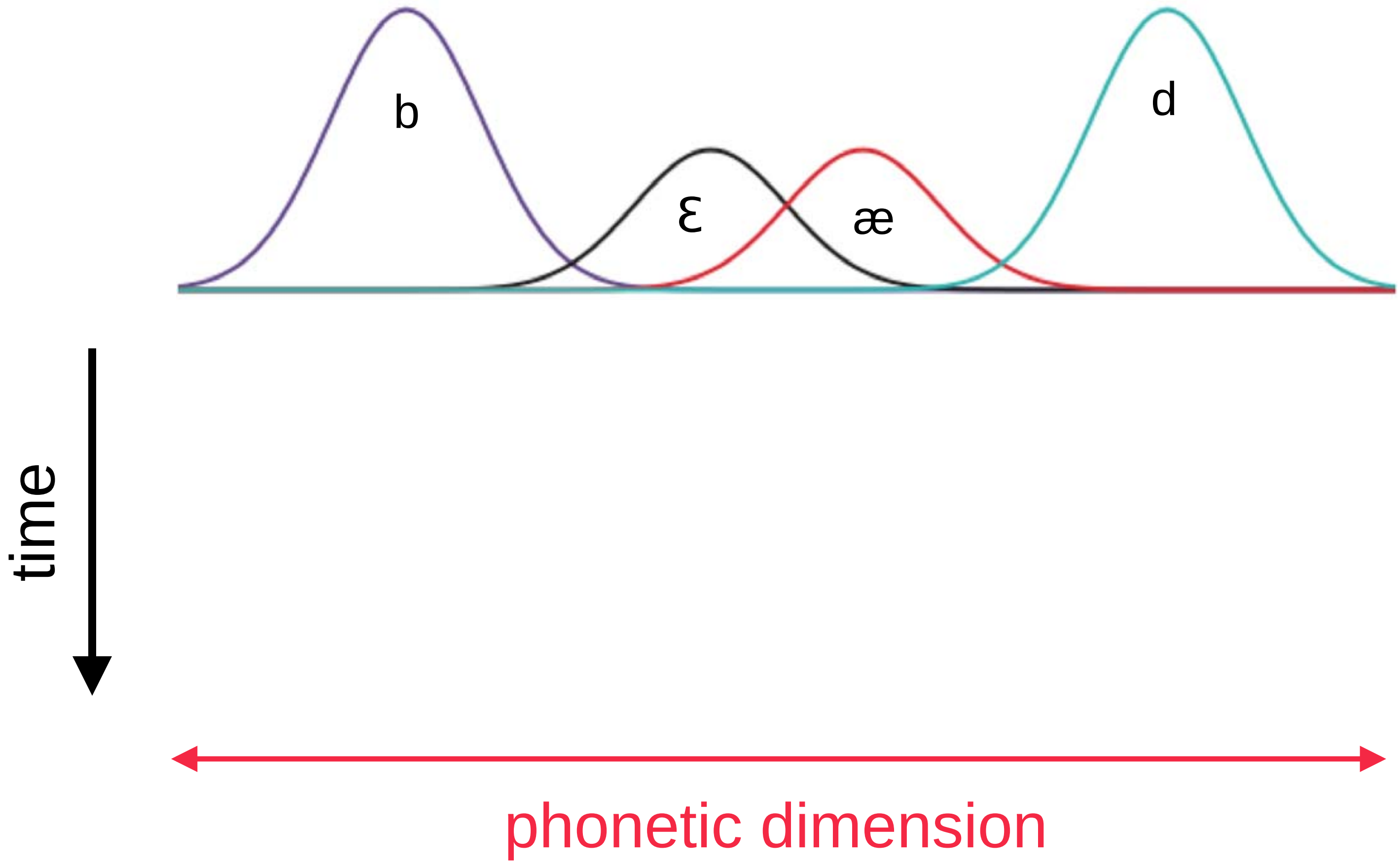
Corpus

A Better Generative Model



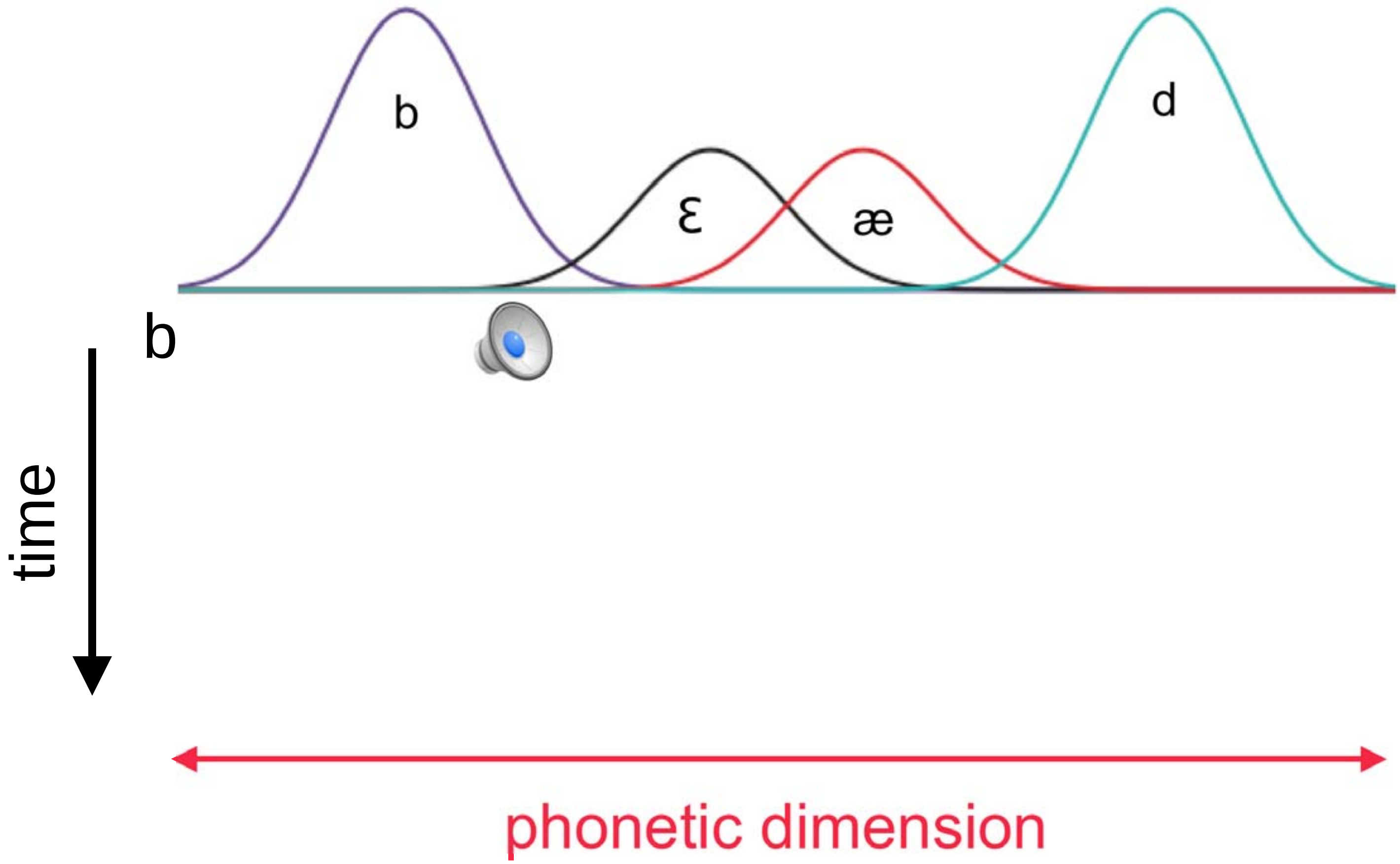
A Better Generative Model

Lexicon: /bæd/, /bɛd/
'bad' 'bed'



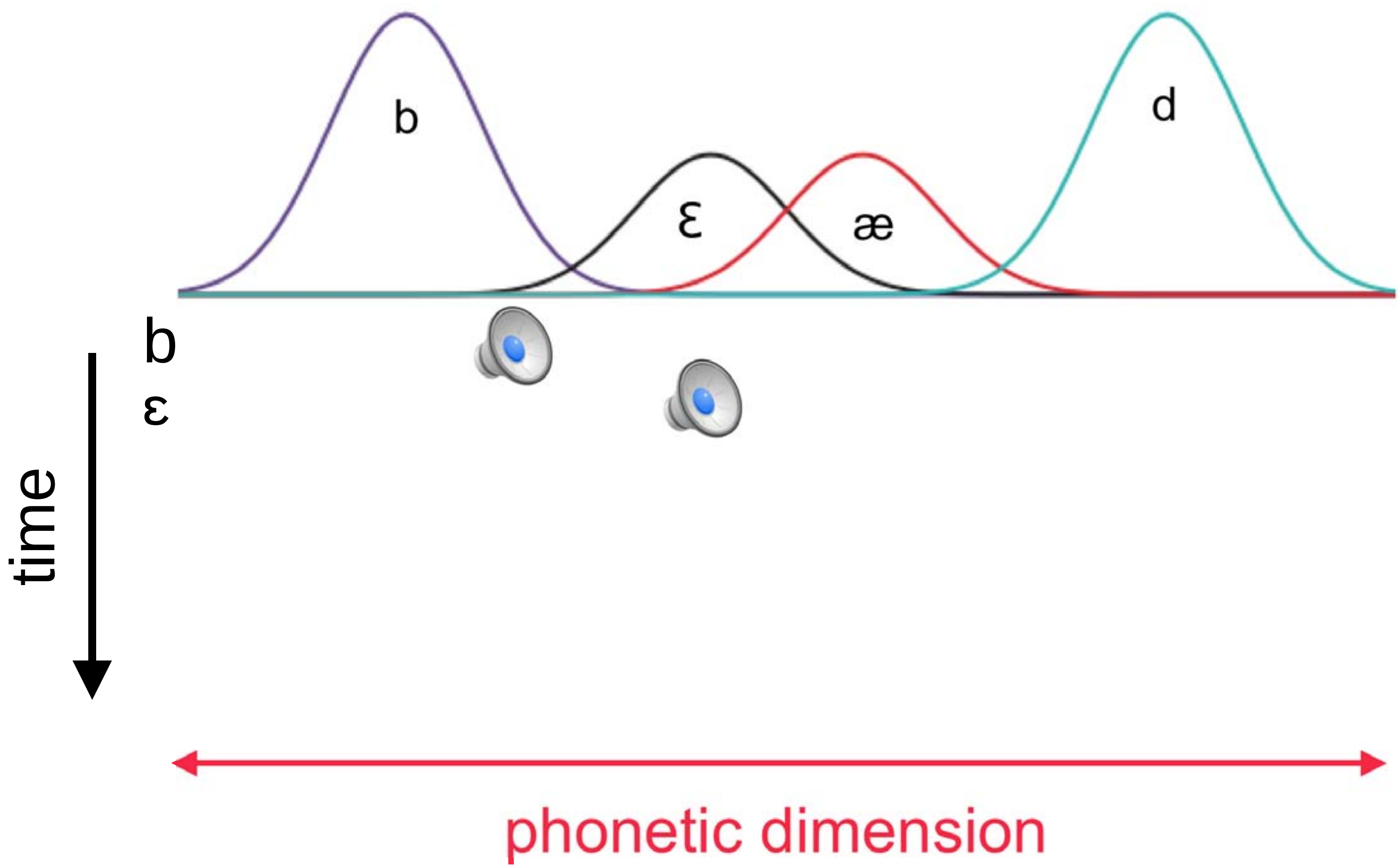
A Better Generative Model

Lexicon: /bæd/, /bɛd/
‘bad’ ‘bed’



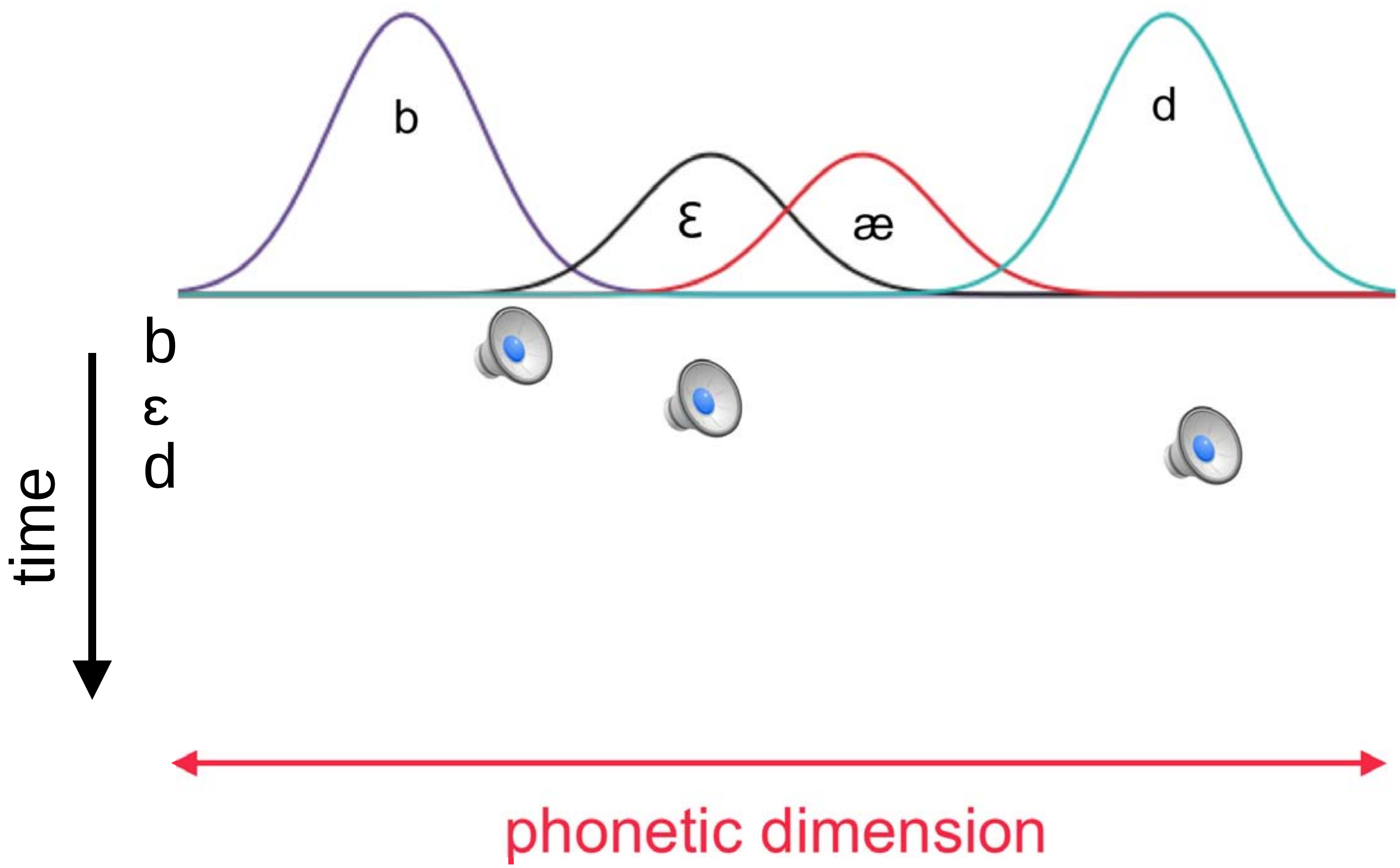
A Better Generative Model

Lexicon: /bæd/, /bɛd/
‘bad’ ‘bed’



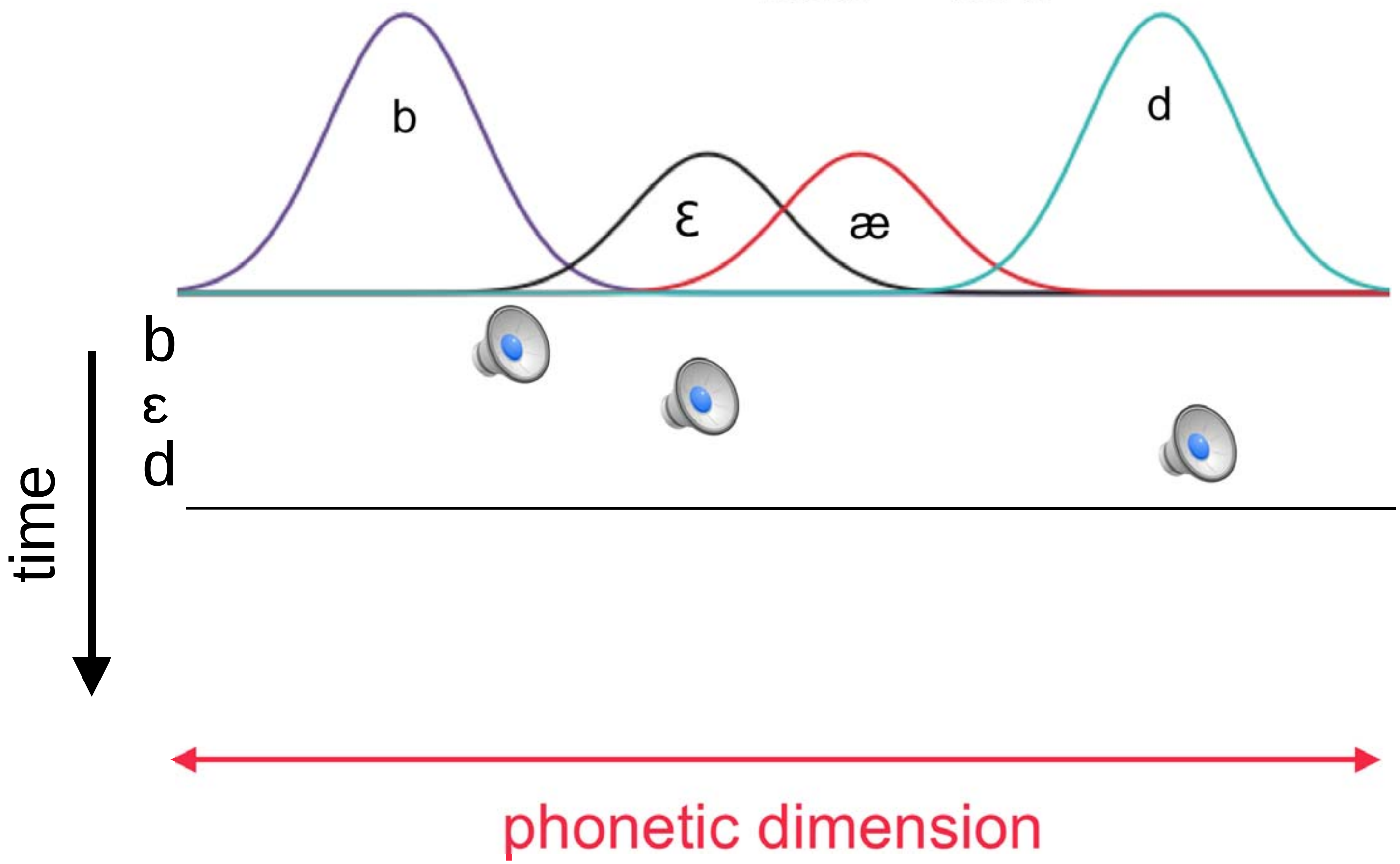
A Better Generative Model

Lexicon: /bæd/, /bɛd/
‘bad’ ‘bed’



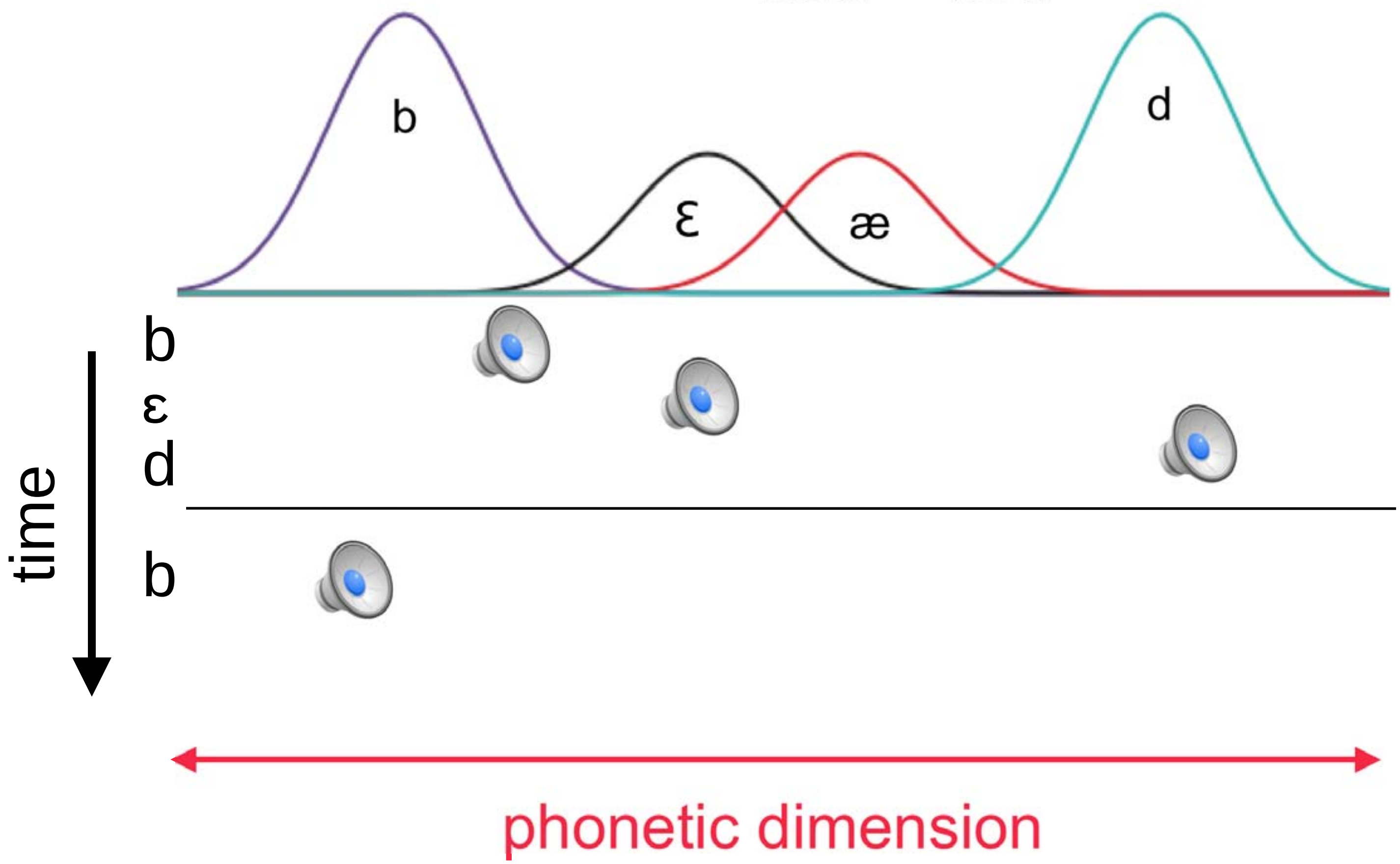
A Better Generative Model

Lexicon: /bæd/, /bɛd/
'bad' 'bed'



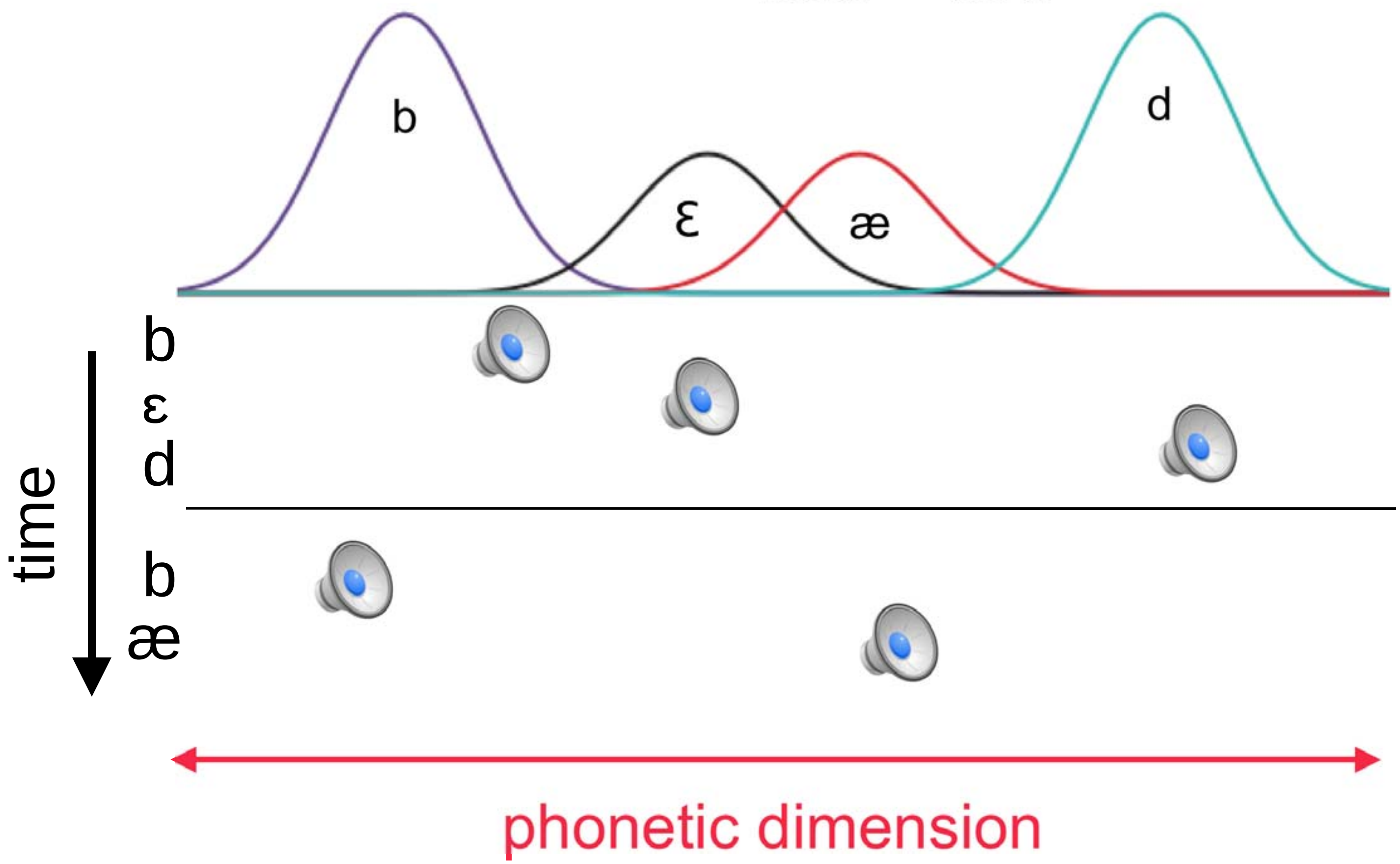
A Better Generative Model

Lexicon: /bæd/, /bɛd/
'bad' 'bed'



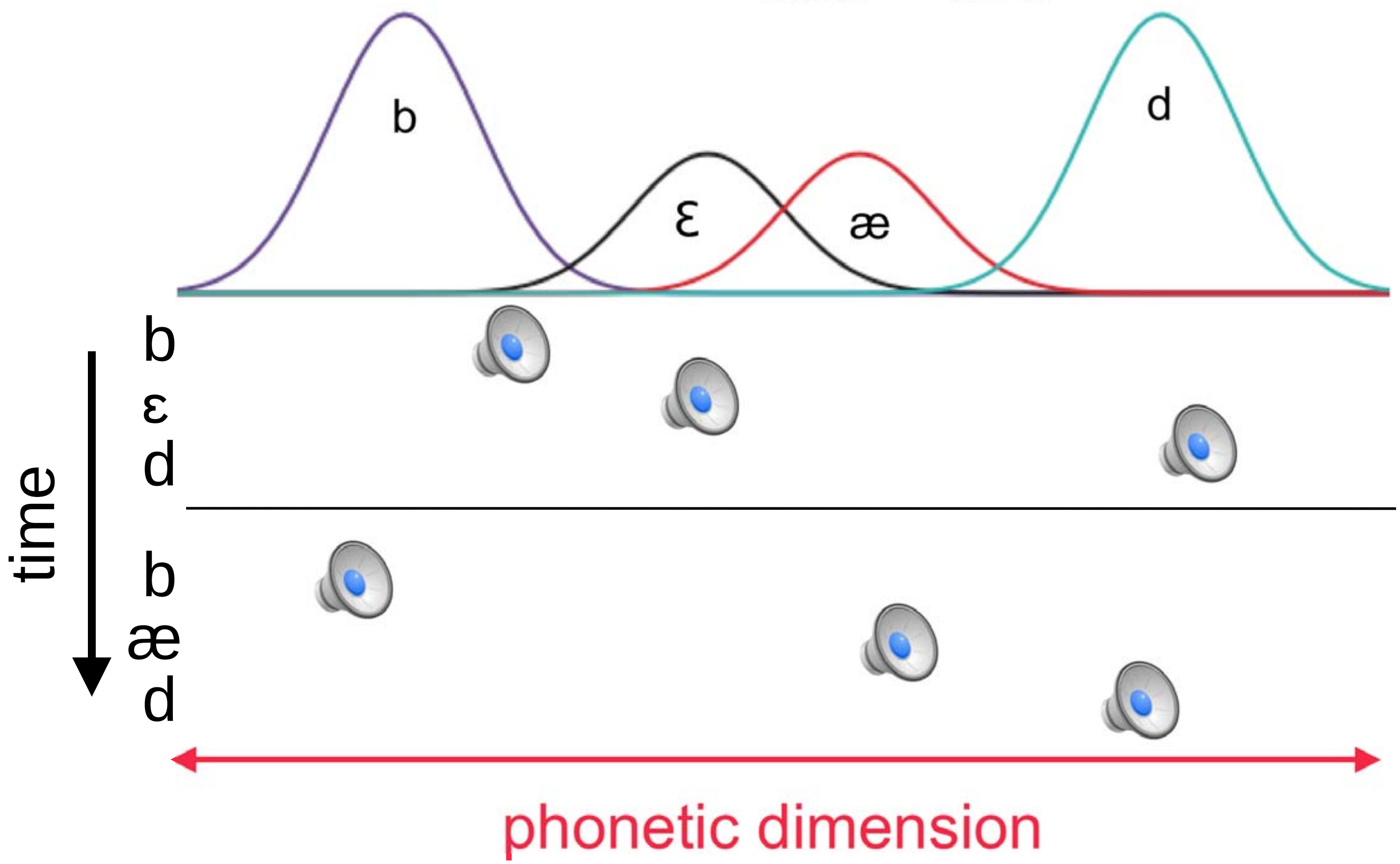
A Better Generative Model

Lexicon: /bæd/, /bɛd/
'bad' 'bed'

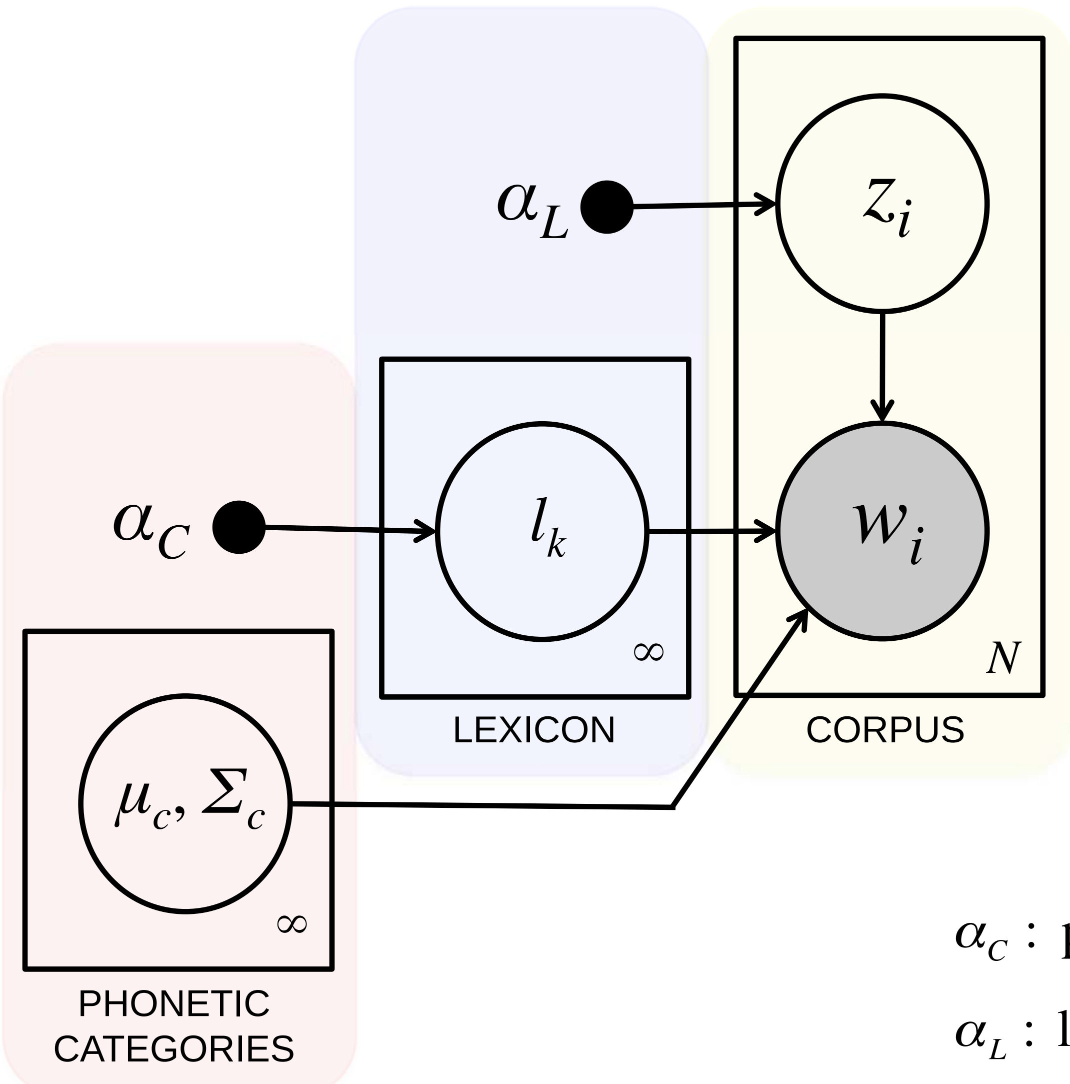


A Better Generative Model

Lexicon: /bæd/, /bɛd/
'bad' 'bed'



A Better Generative Model



μ_c, Σ_c : parameters of category c

l_k : form of lexical item k

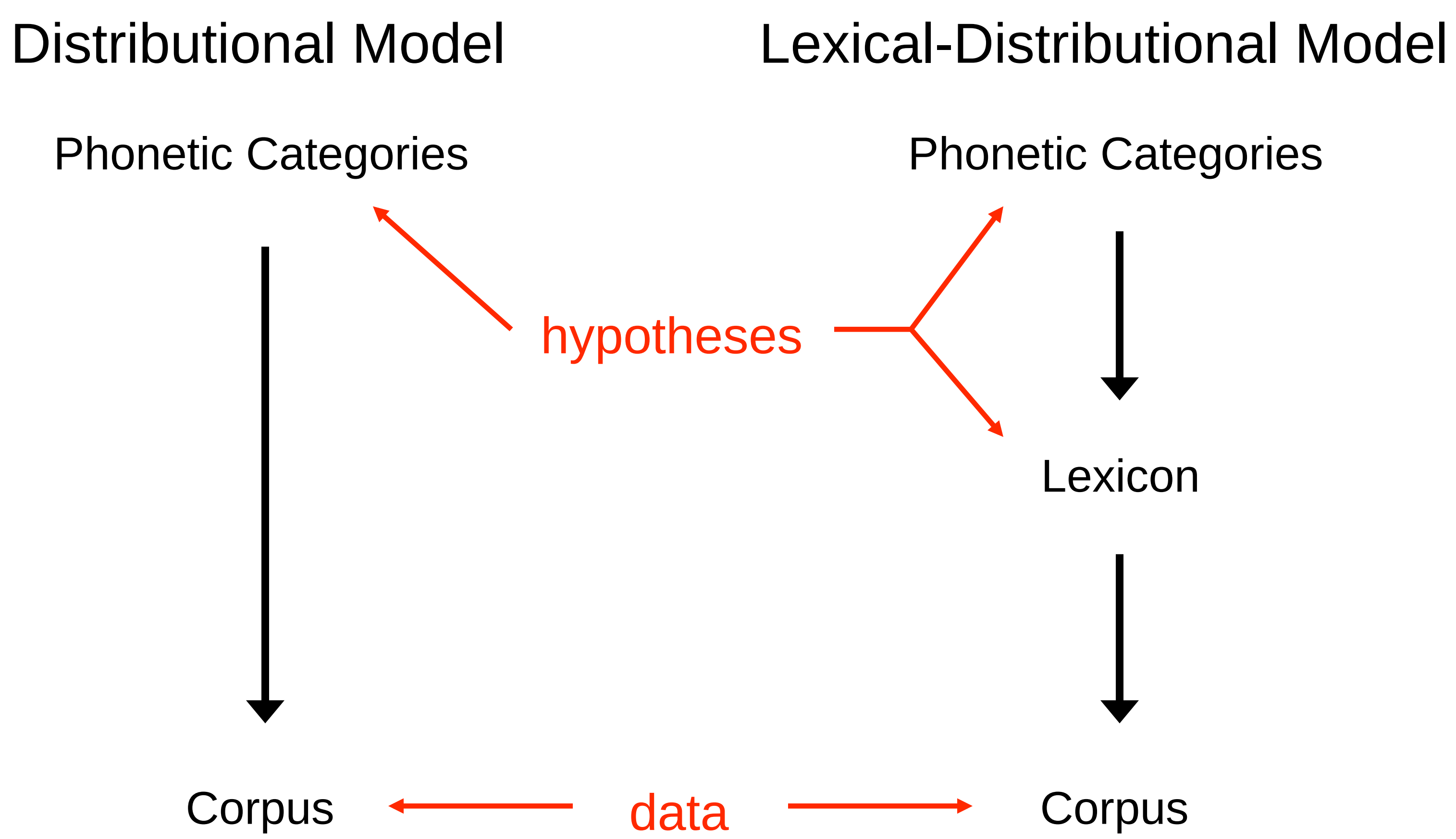
z_i : category of word i

w_i : acoustics of word i

α_C : phonetic concentration parameter

α_L : lexical concentration parameter

Models of Category Learning



Models of Category Learning

Distributional

- Assume sounds are generated independently of their neighbors
- Infer category parameters
- Phonetic categories characterize the types of variability found among sounds in the corpus

Lexical-Distributional

- Assume sounds are generated as parts of words
- Infer category parameters and forms of lexical items
- Phonetic categories are overhypotheses about the types of variability seen in lexical items

Qualitative Behavior

Compare lexical-distributional model's behavior on two lexicons

- Informative lexicon: 'add', 'ebb'
- Minimal pair lexicon: 'add, 'Ed', 'ab', 'ebb'

Qualitative Behavior

Compare lexical-distributional model’s behavior on two lexicons

- Informative lexicon: ‘add’, ‘ebb’
- Minimal pair lexicon: ‘add, ‘Ed’, ‘ab’, ‘ebb’

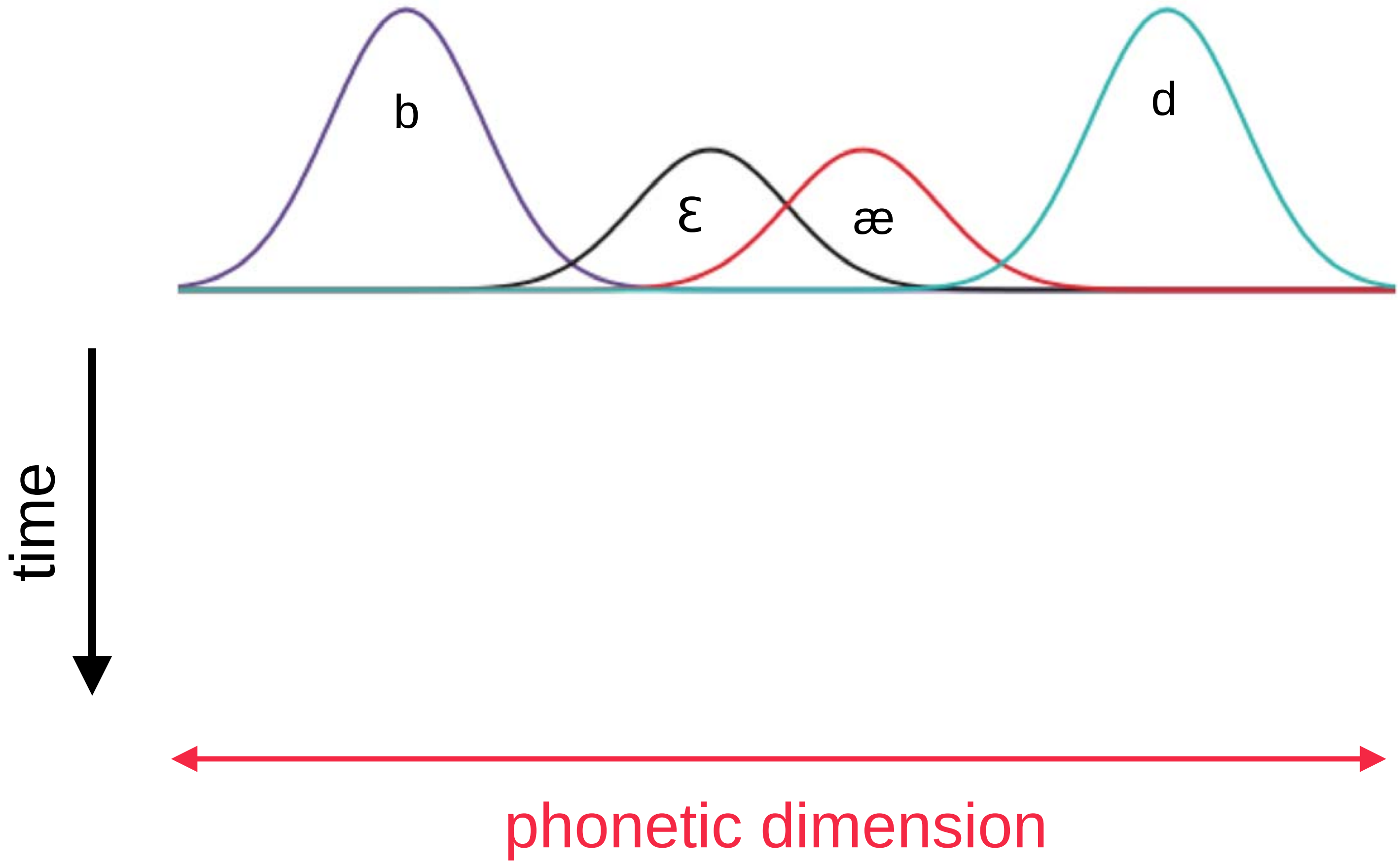
Minimal pairs:

add vs. Ed

Typically taken as evidence
that sounds are different

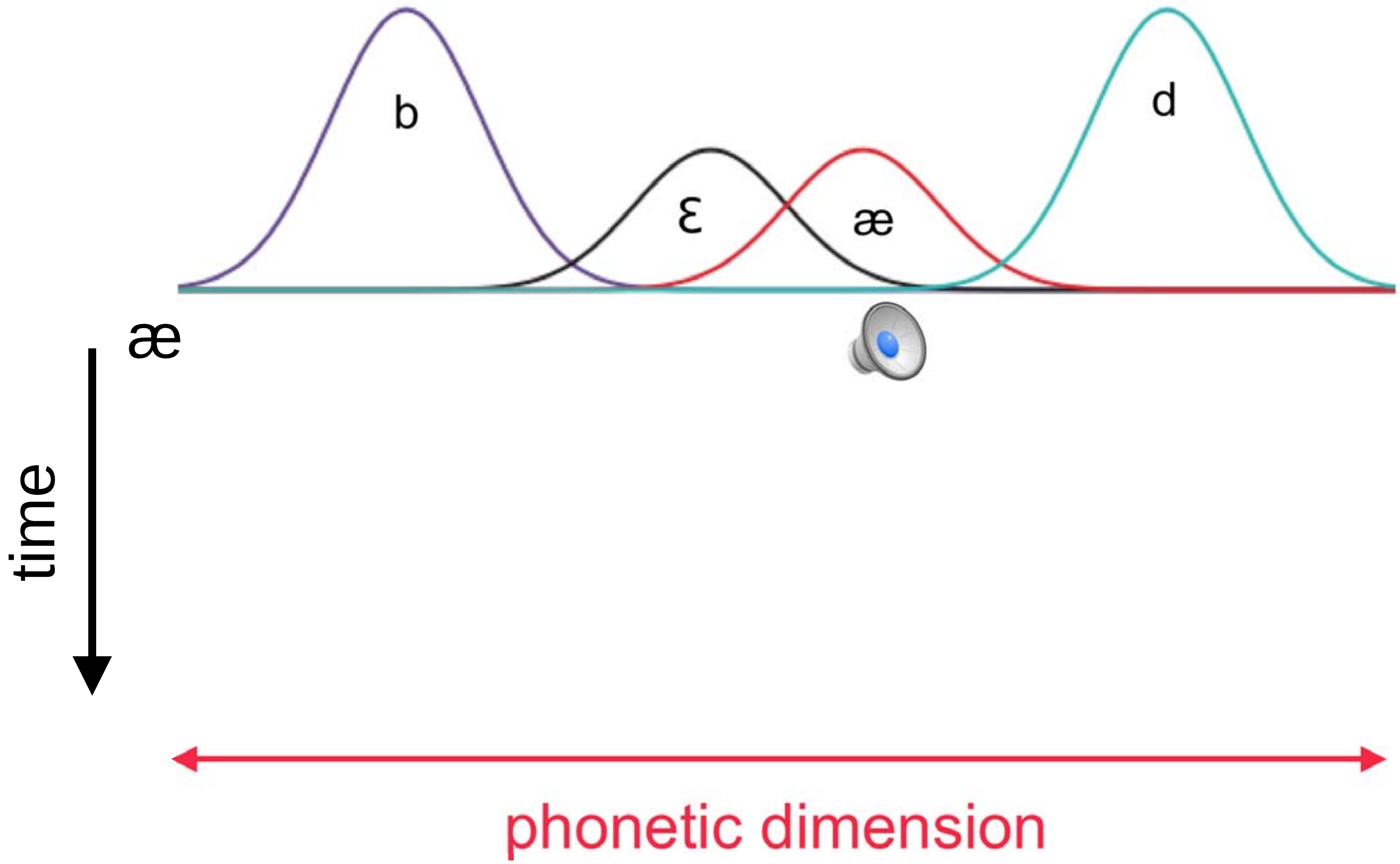
Informative Lexicon

Lexicon: /æd/, /ɛb/
‘add’ ‘ebb’



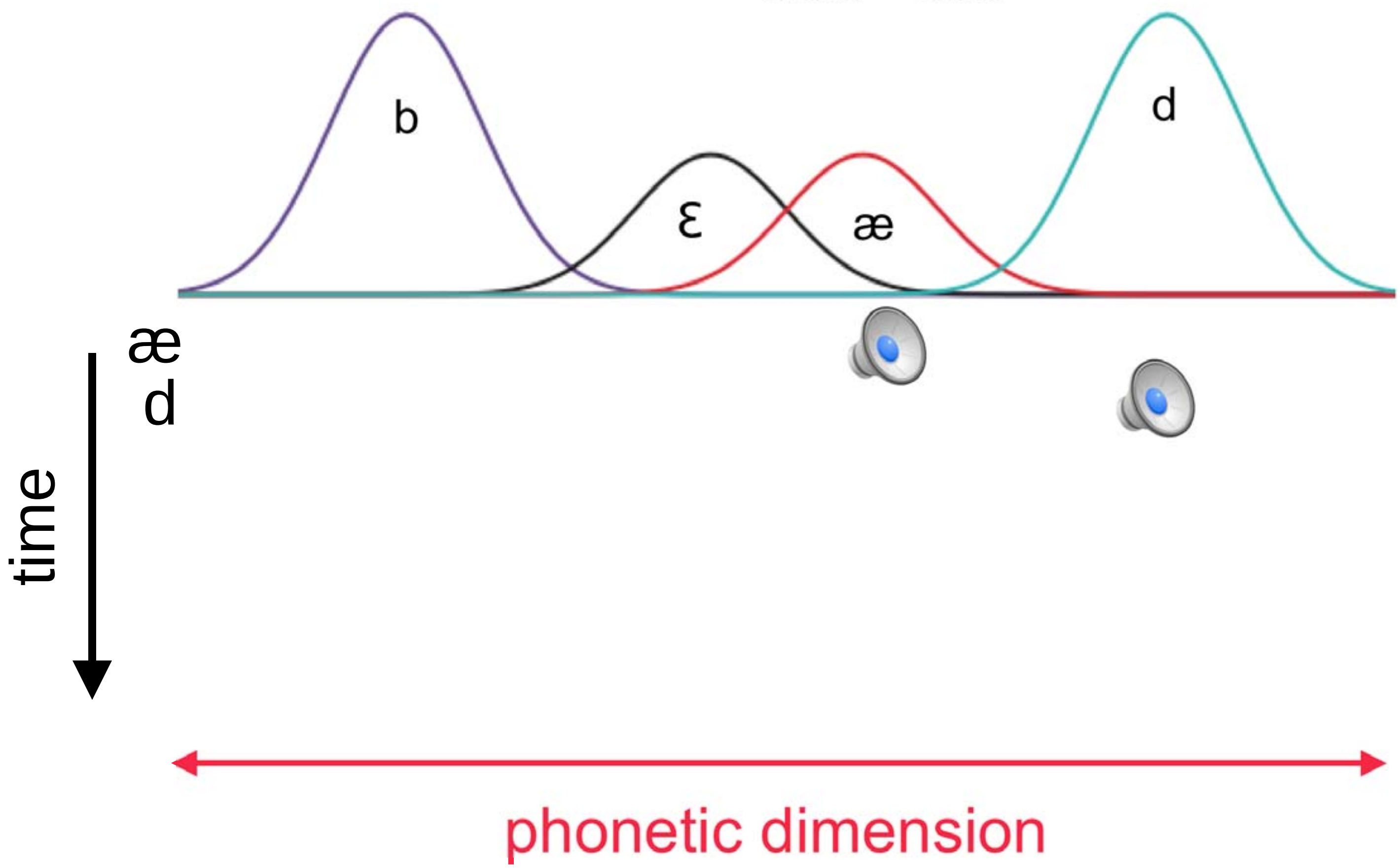
Informative Lexicon

Lexicon: /æd/, /ɛb/
'add' 'ebb'



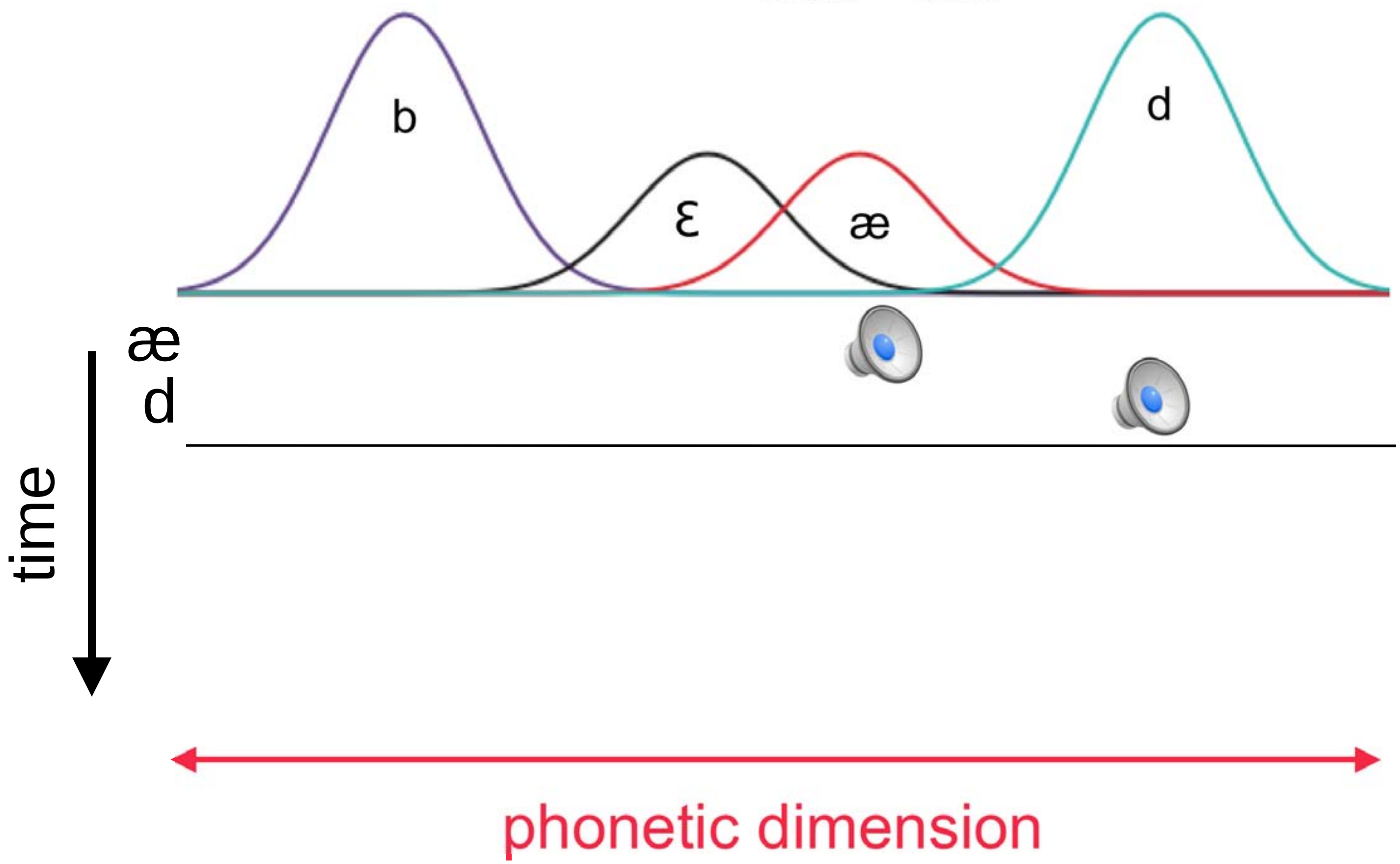
Informative Lexicon

Lexicon: /æd/, /ɛb/
‘add’ ‘ebb’

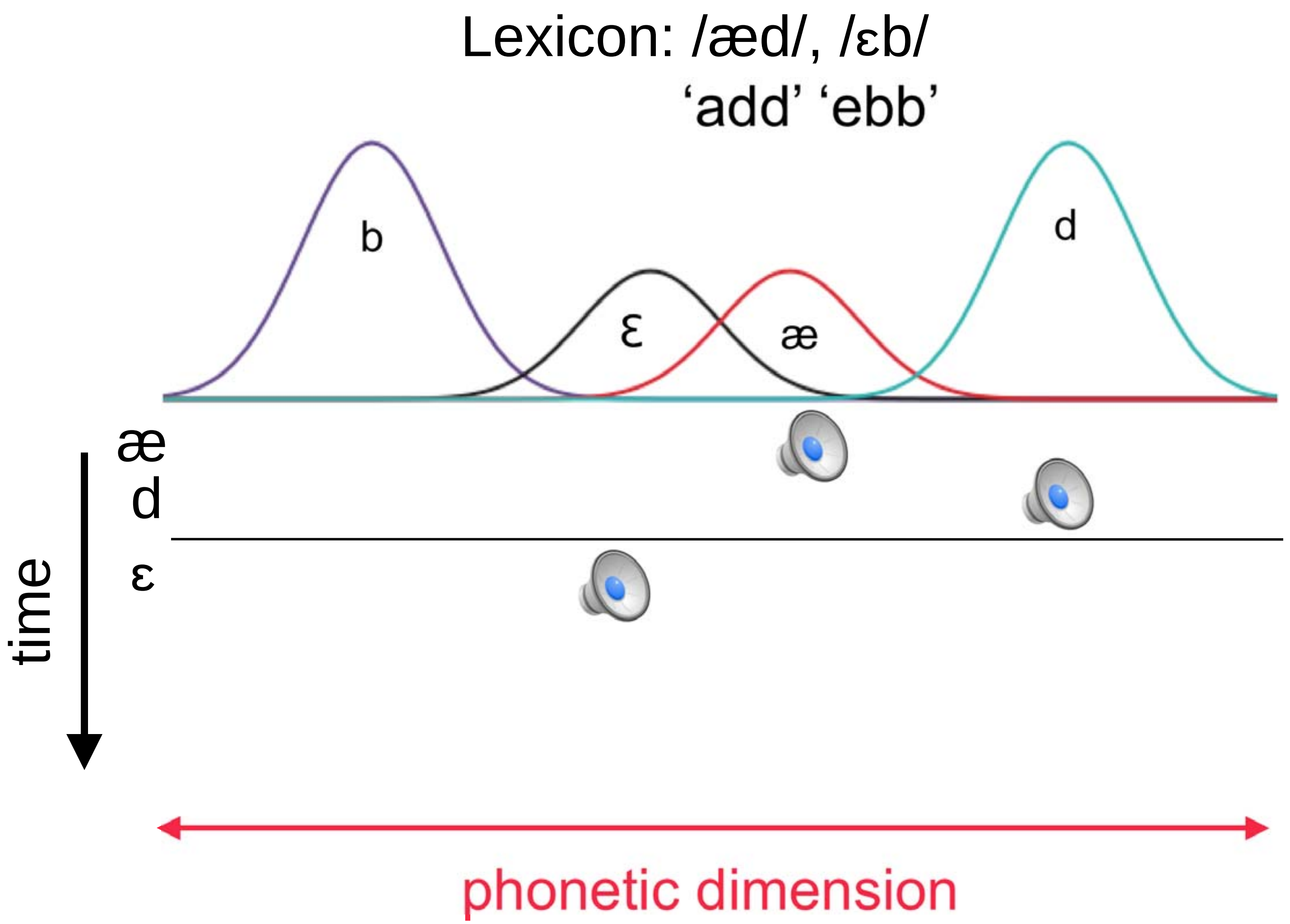


Informative Lexicon

Lexicon: /æd/, /ɛb/
‘add’ ‘ebb’

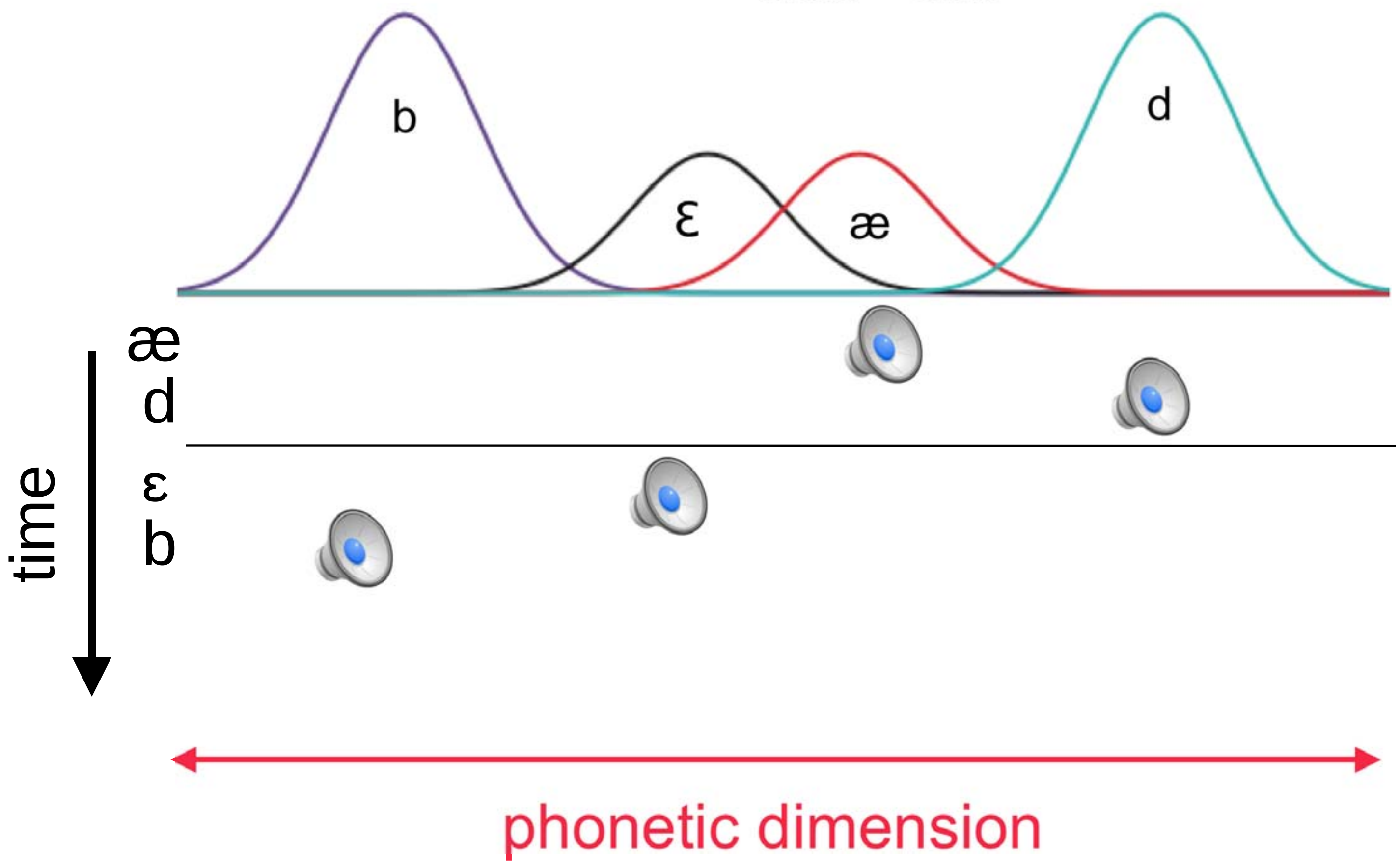


Informative Lexicon



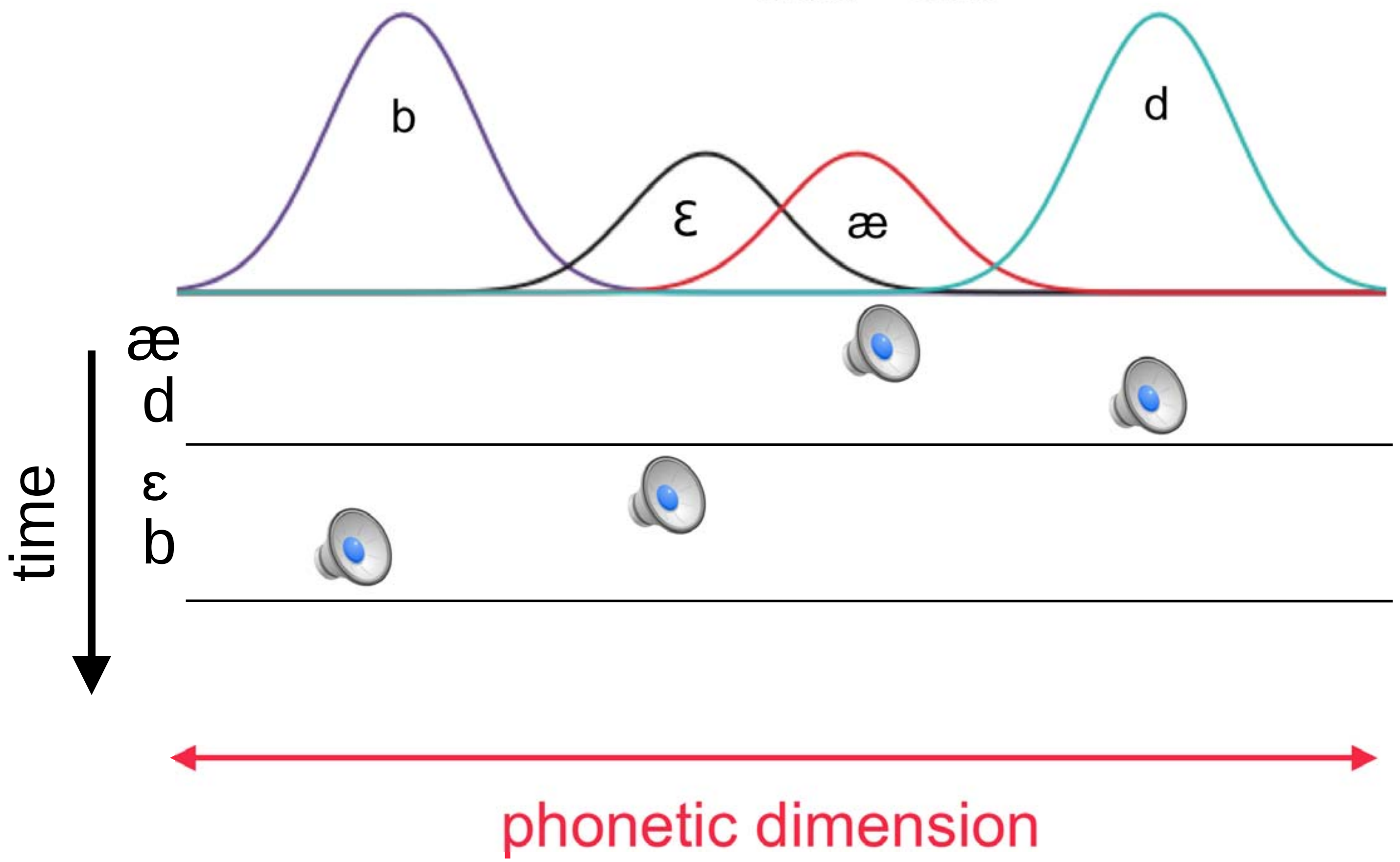
Informative Lexicon

Lexicon: /æd/, /ɛb/
'add' 'ebb'



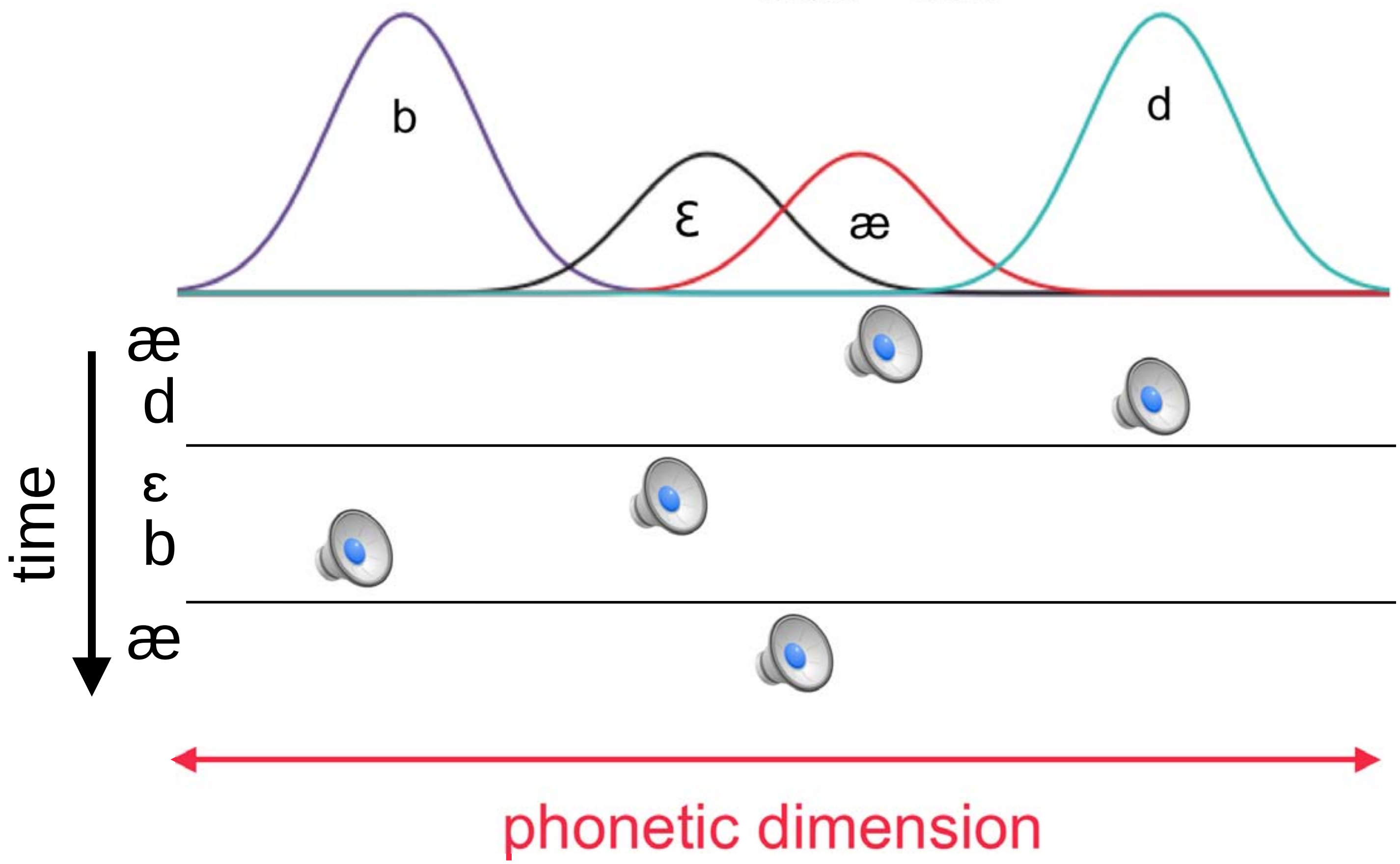
Informative Lexicon

Lexicon: /æd/, /ɛb/
'add' 'ebb'



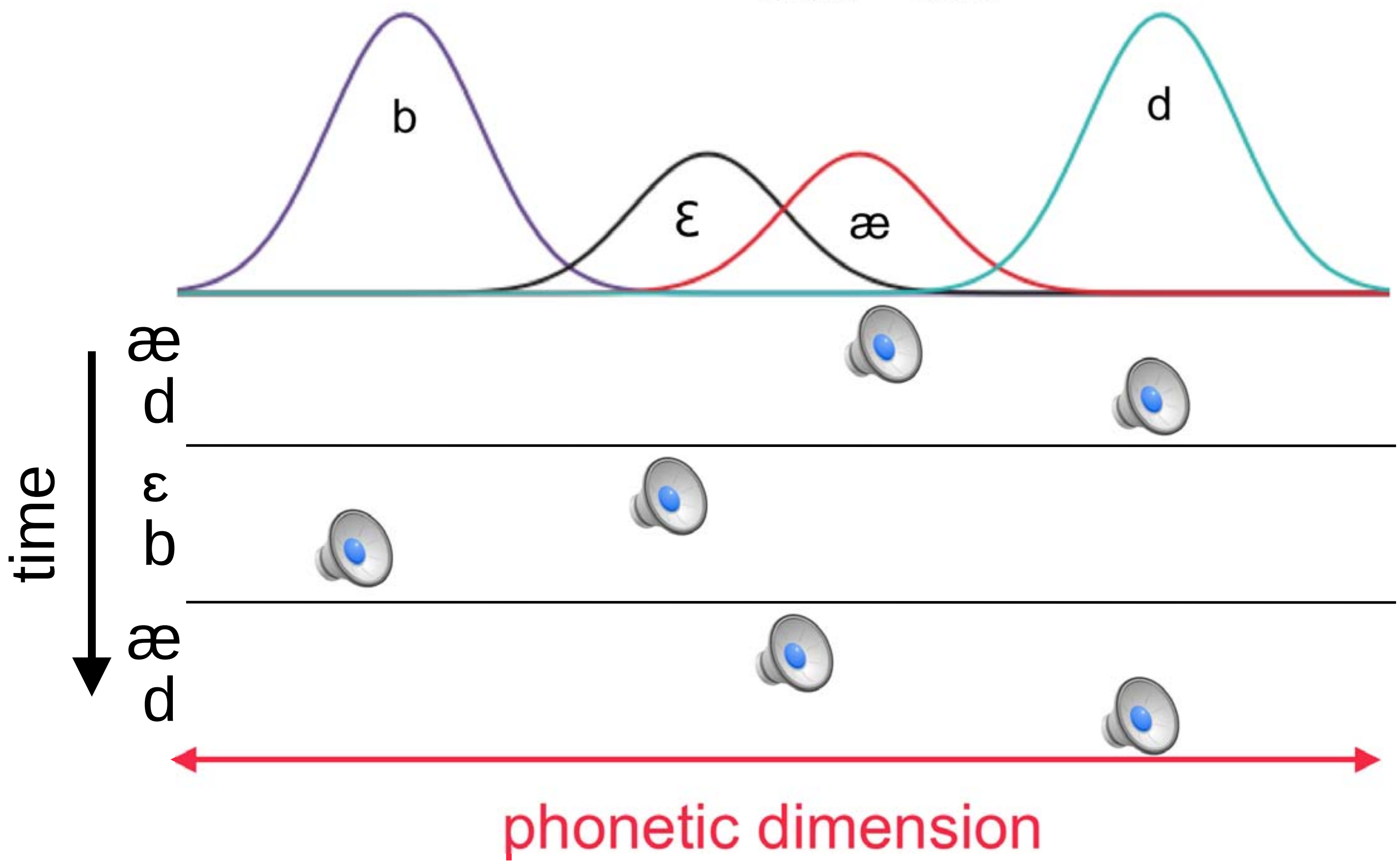
Informative Lexicon

Lexicon: /æd/, /ɛb/
'add' 'ebb'

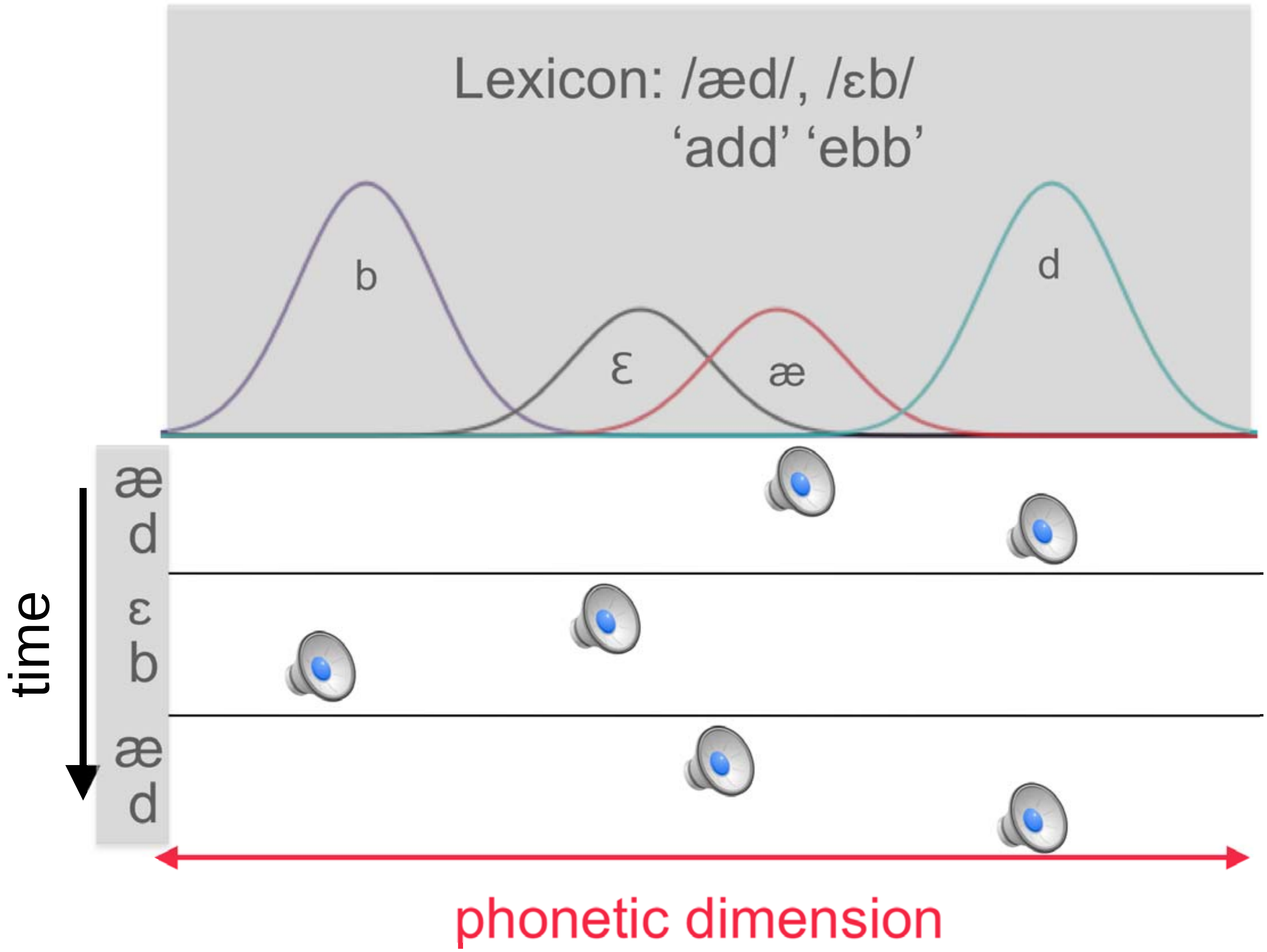


Informative Lexicon

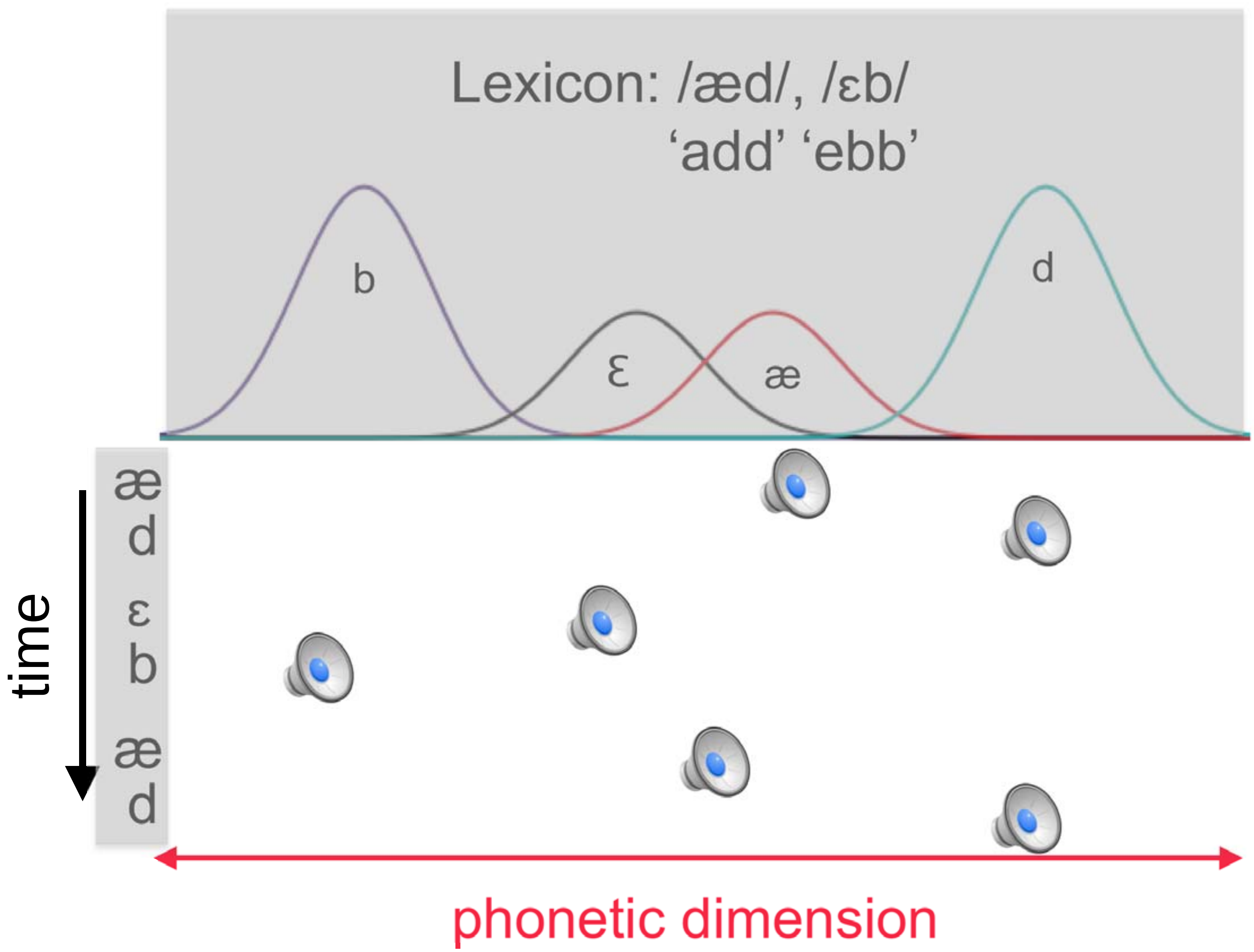
Lexicon: /æd/, /ɛb/
'add' 'ebb'



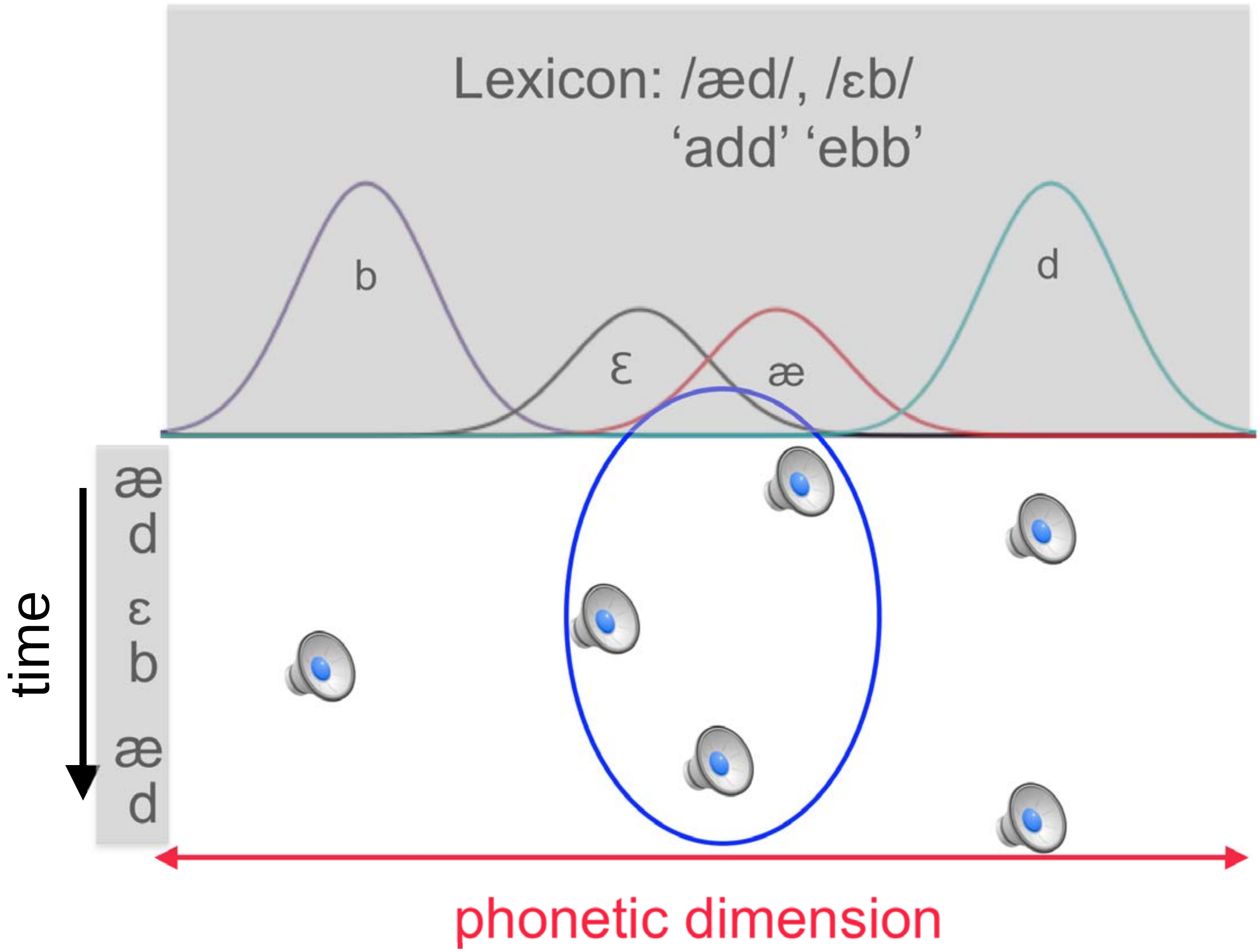
Informative Lexicon



Distributional Model

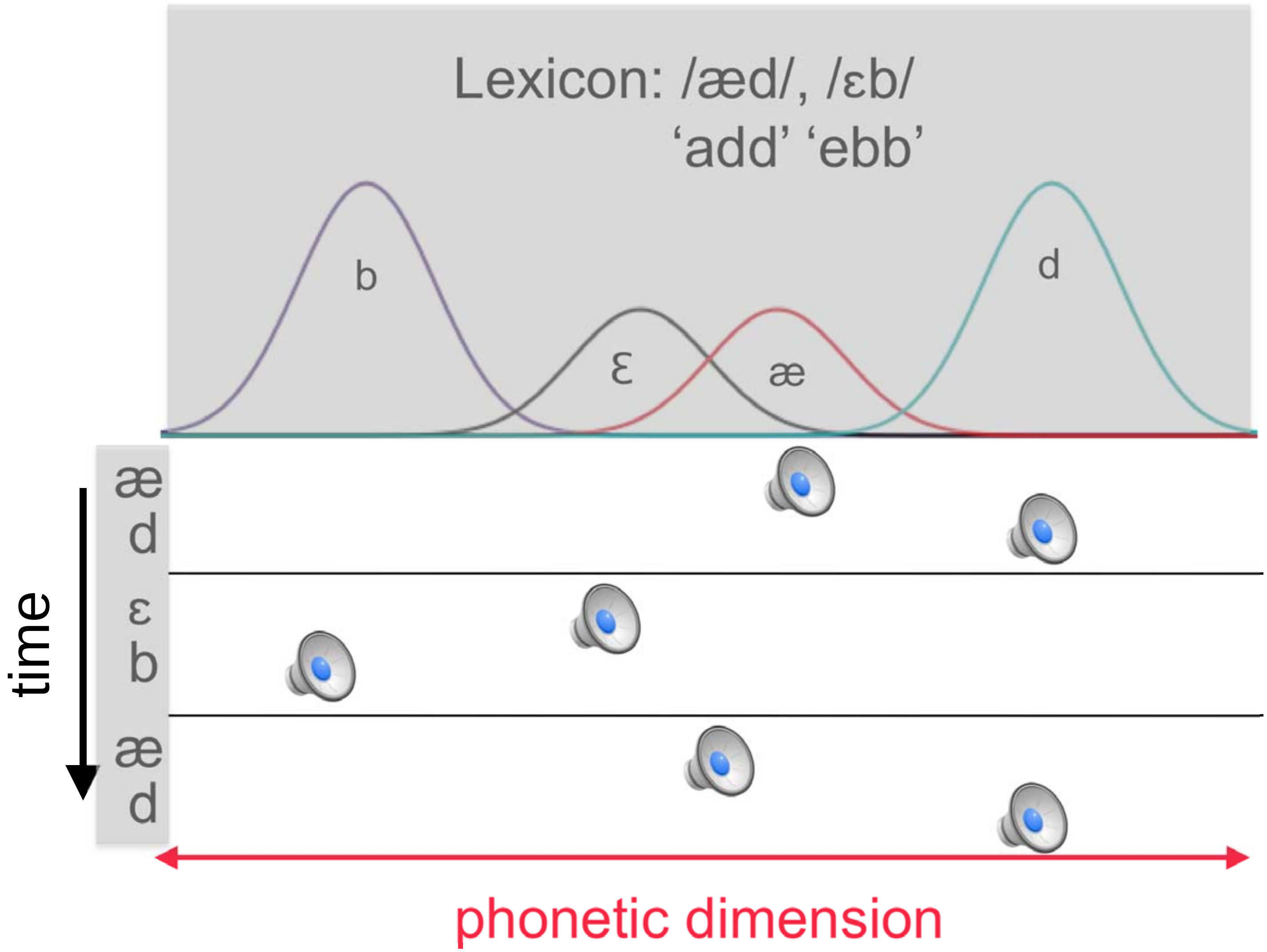


Distributional Model

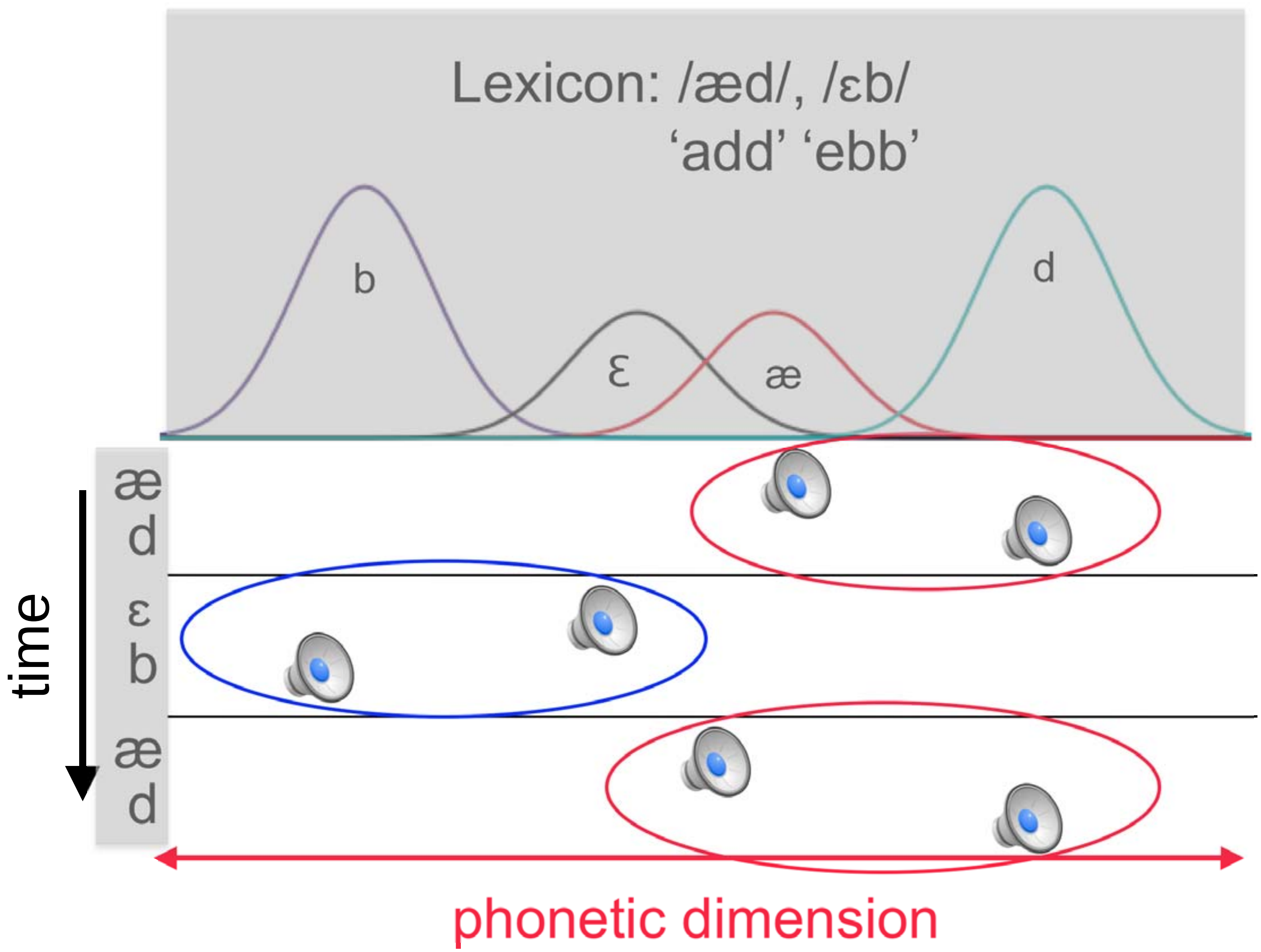


Distributional Model

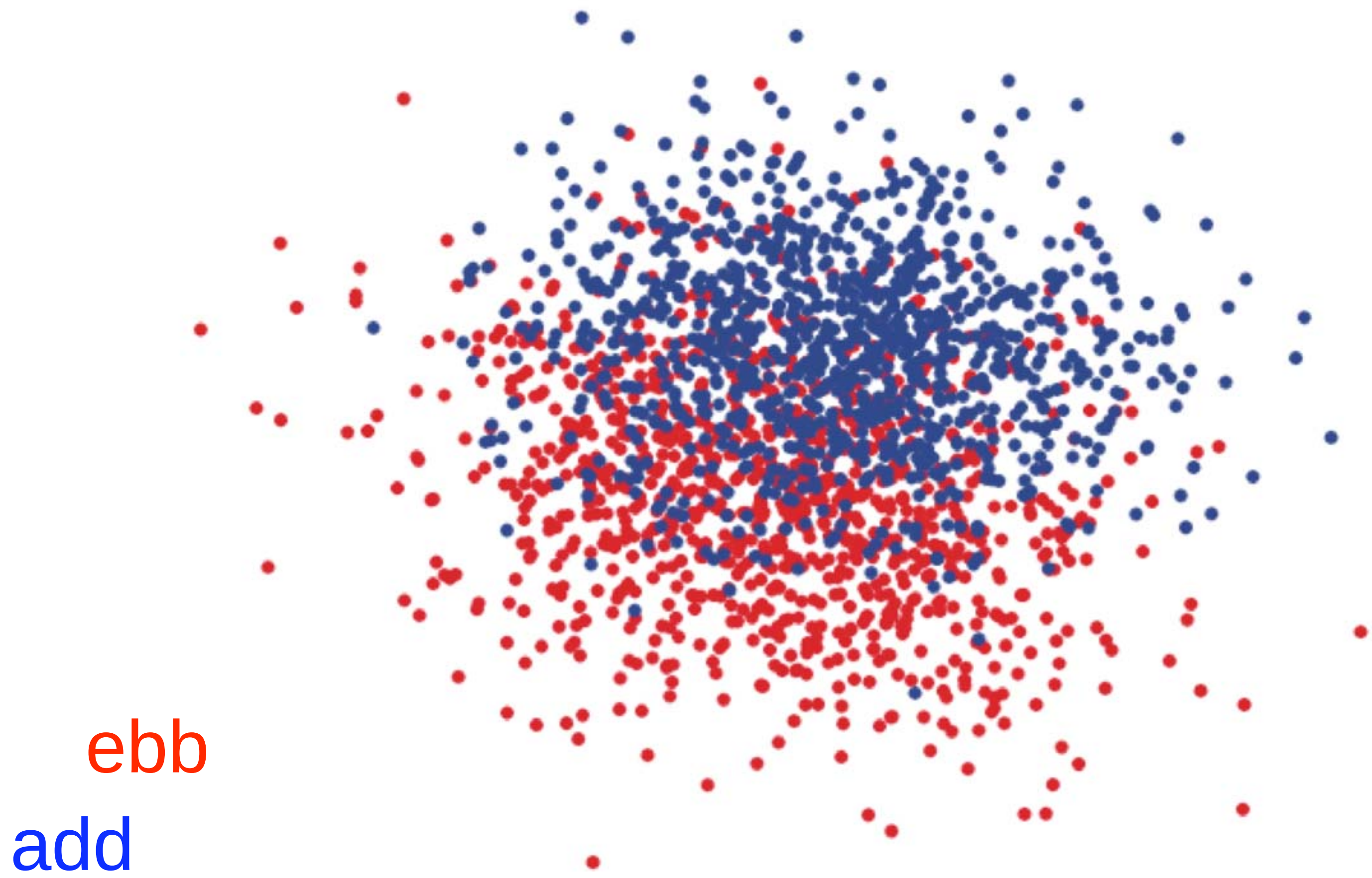
Lexical-Distributional Model



Lexical-Distributional Model

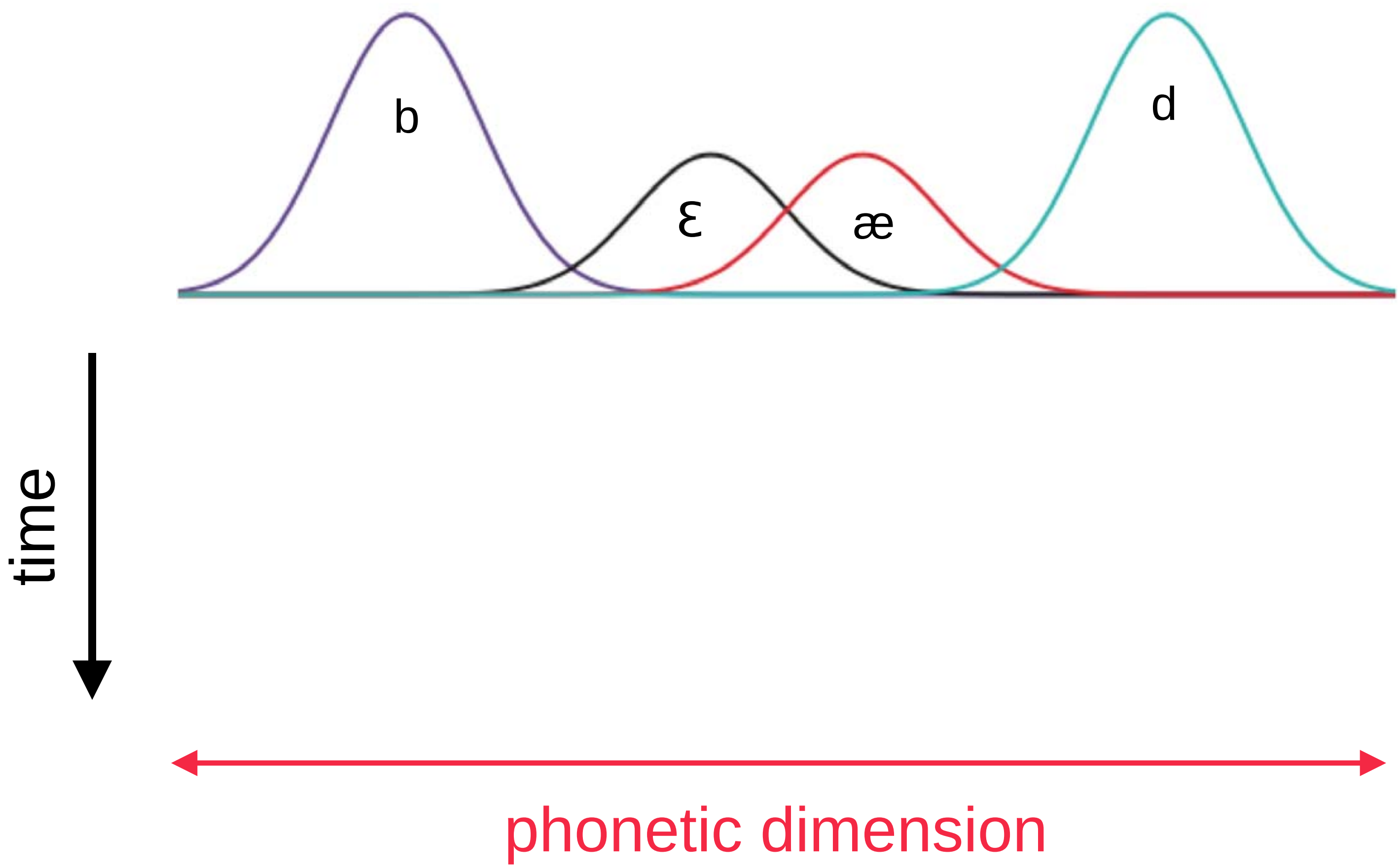


Lexical-Distributional Model



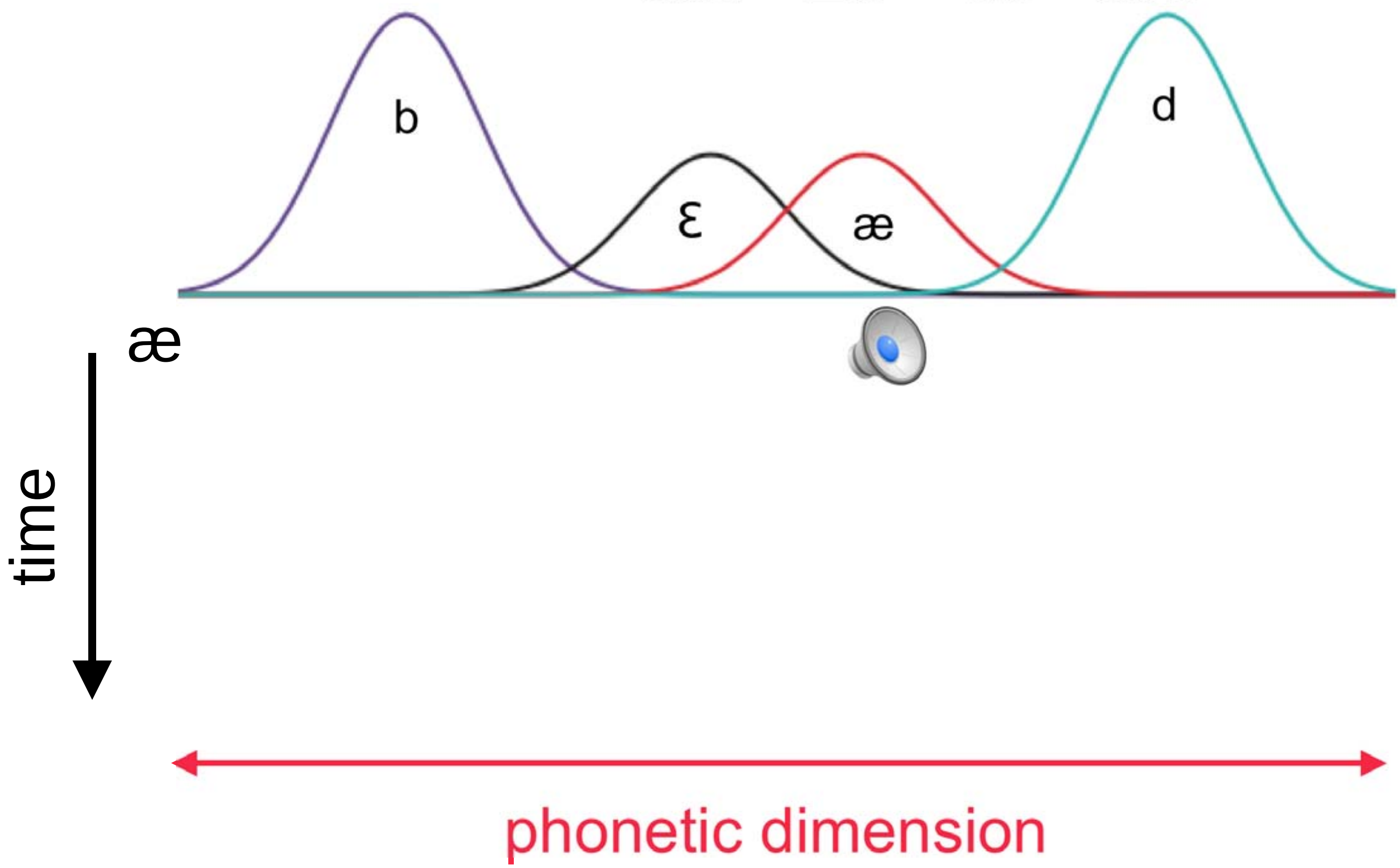
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æb/, /ɛb/
‘add’ ‘Ed’ ‘ab’ ‘ebb’



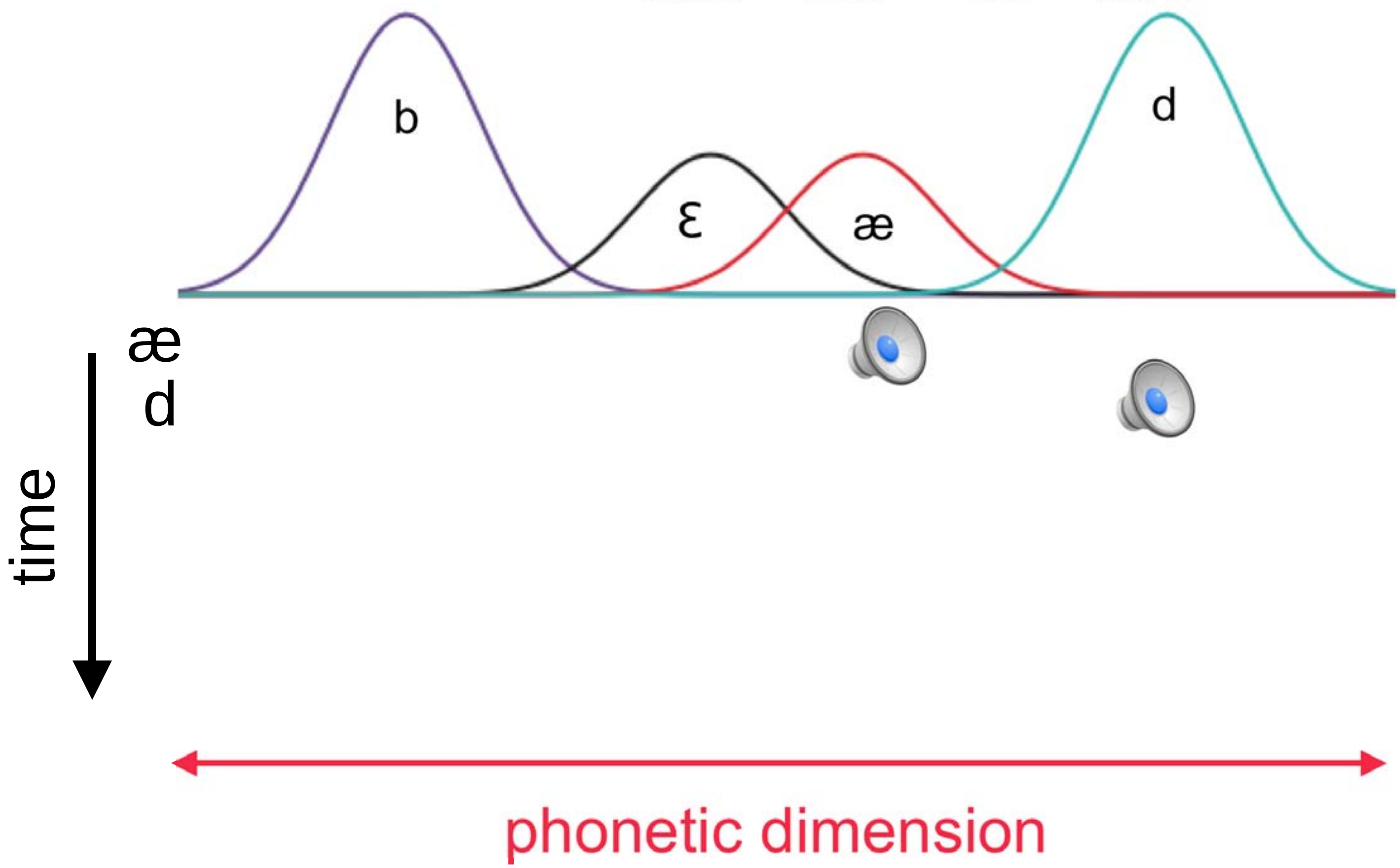
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æb/, /ɛb/
‘add’ ‘Ed’ ‘ab’ ‘ebb’



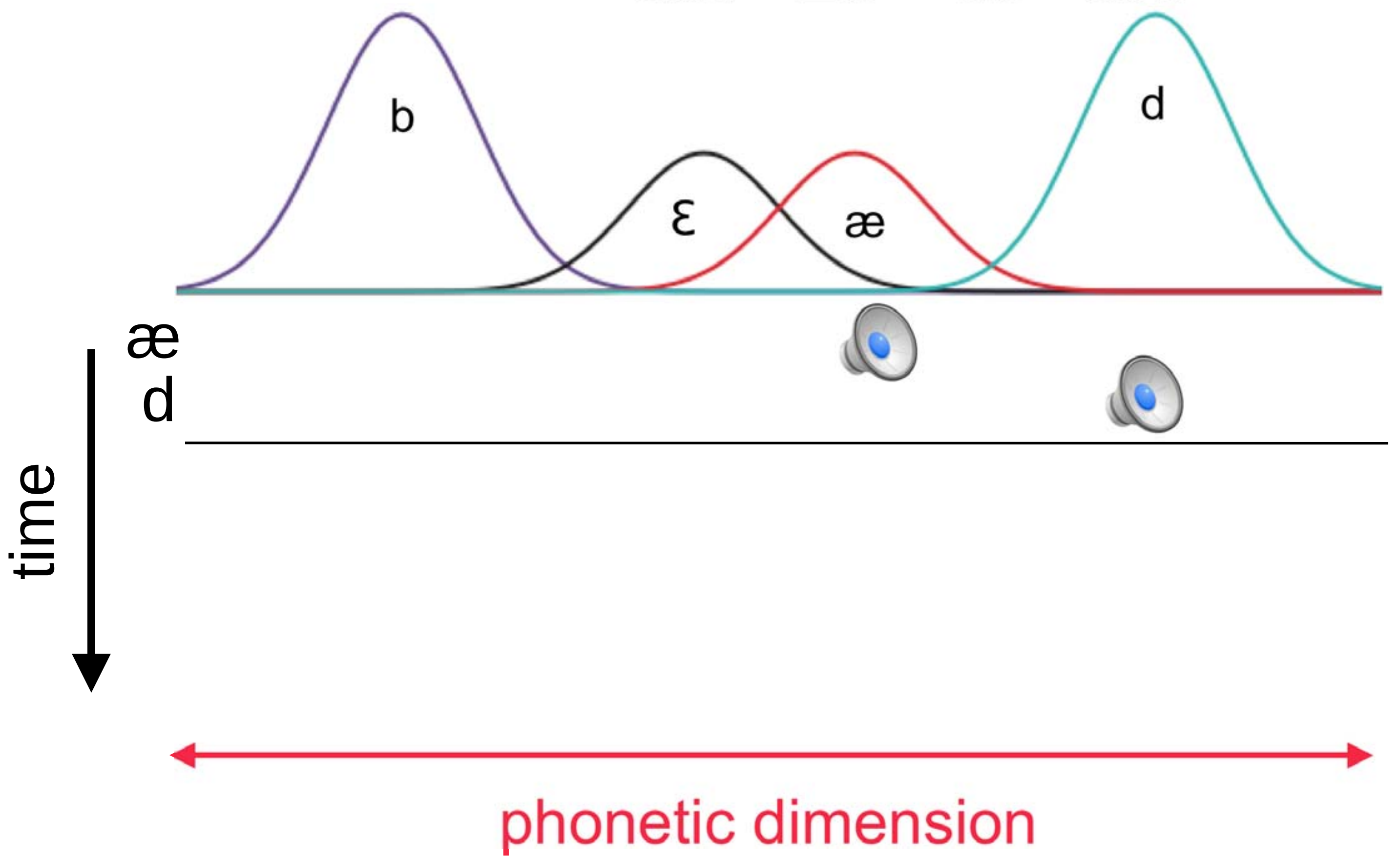
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æb/, /ɛb/
‘add’ ‘Ed’ ‘ab’ ‘ebb’



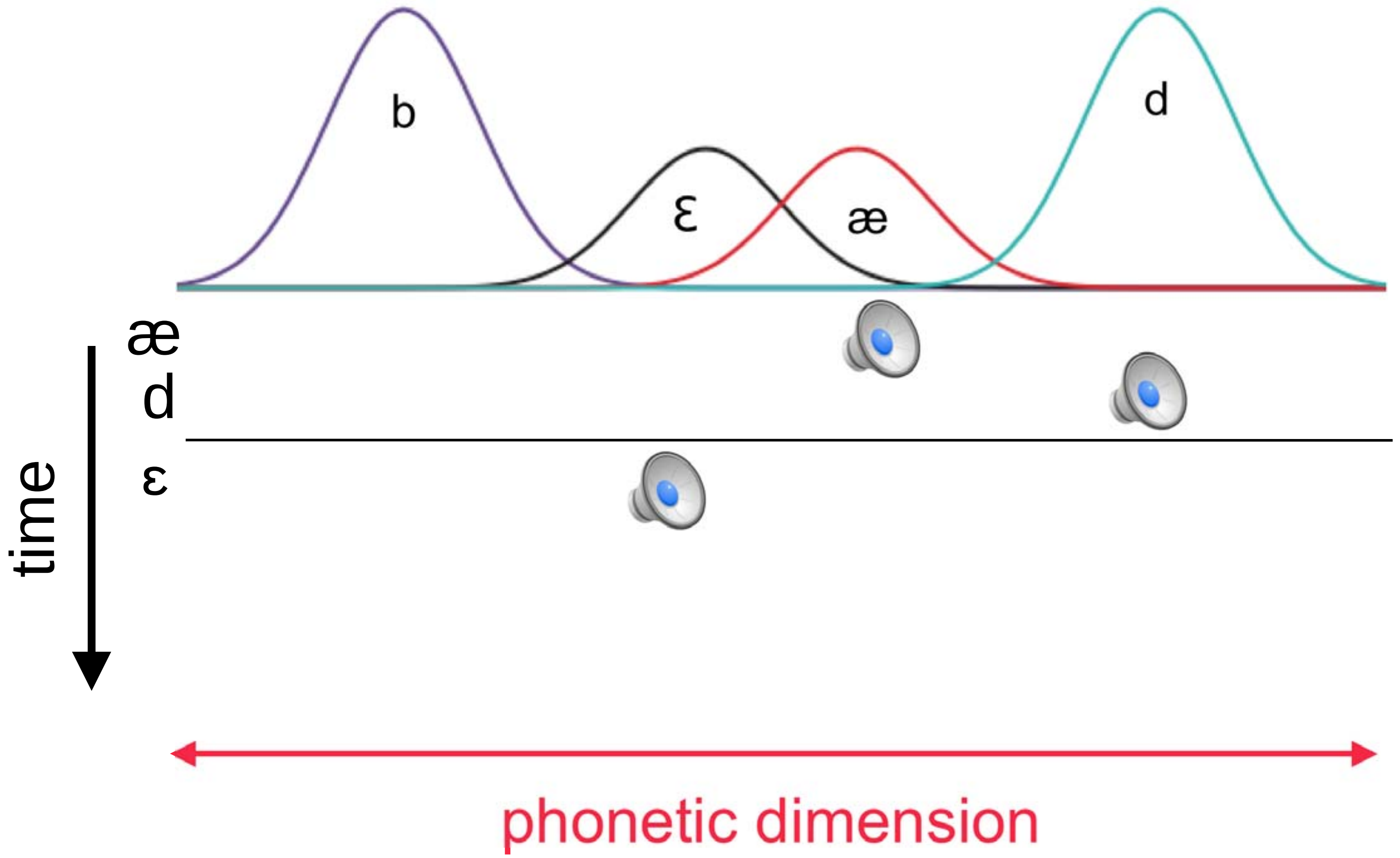
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æb/, /ɛb/
‘add’ ‘Ed’ ‘ab’ ‘ebb’



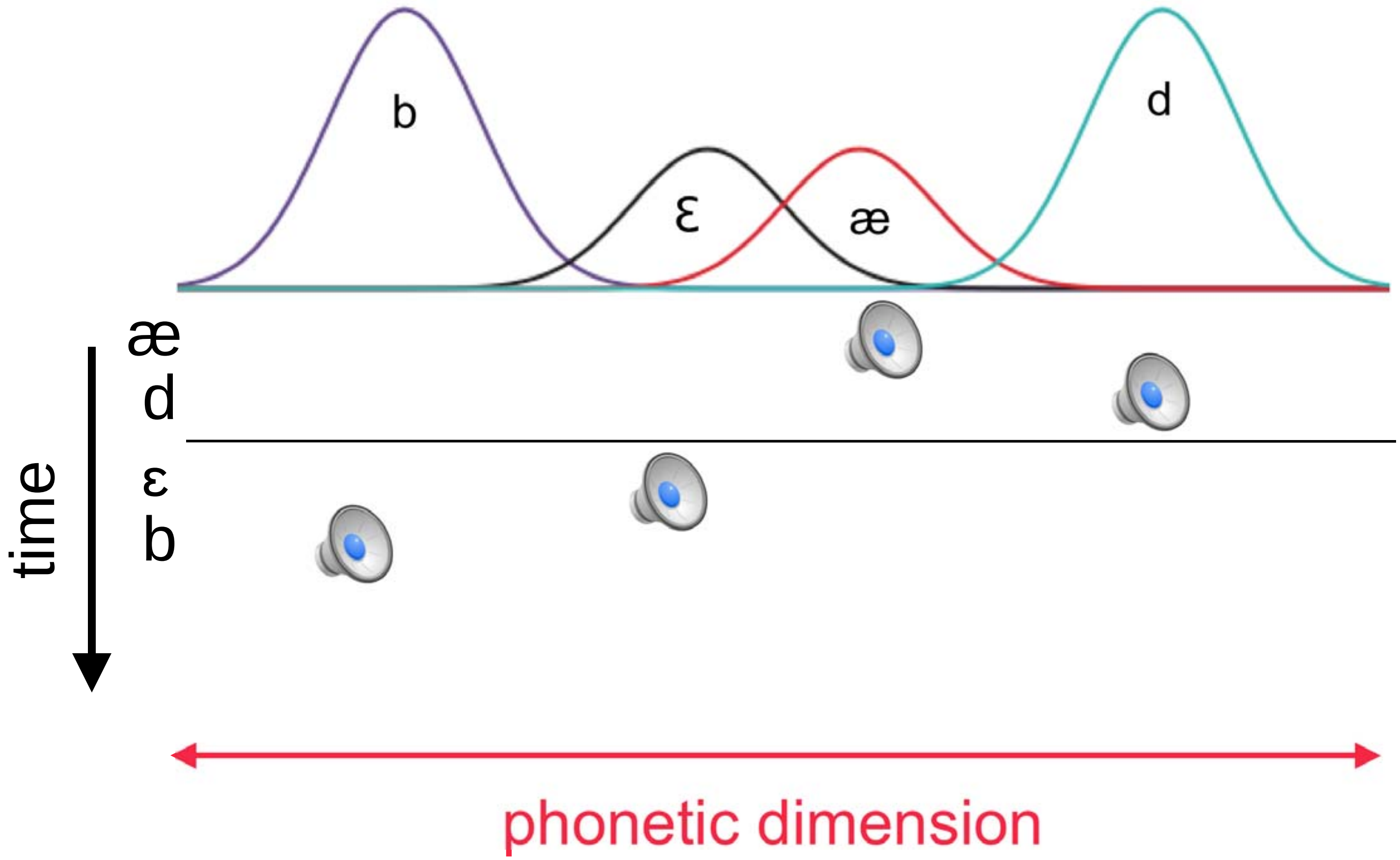
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æb/, /ɛb/
'add' 'Ed' 'ab' 'ebb'



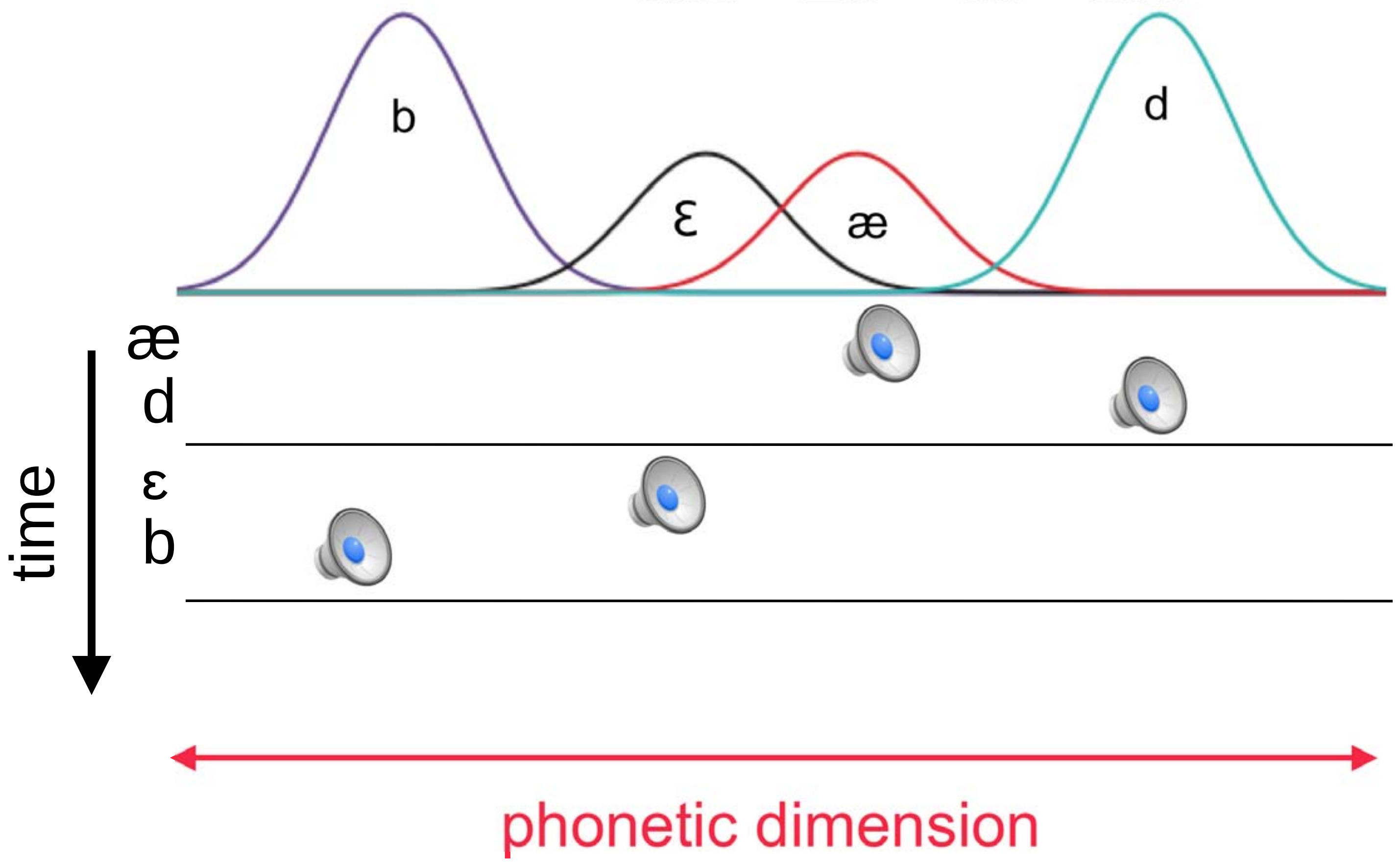
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æb/, /ɛb/
'add' 'Ed' 'ab' 'ebb'



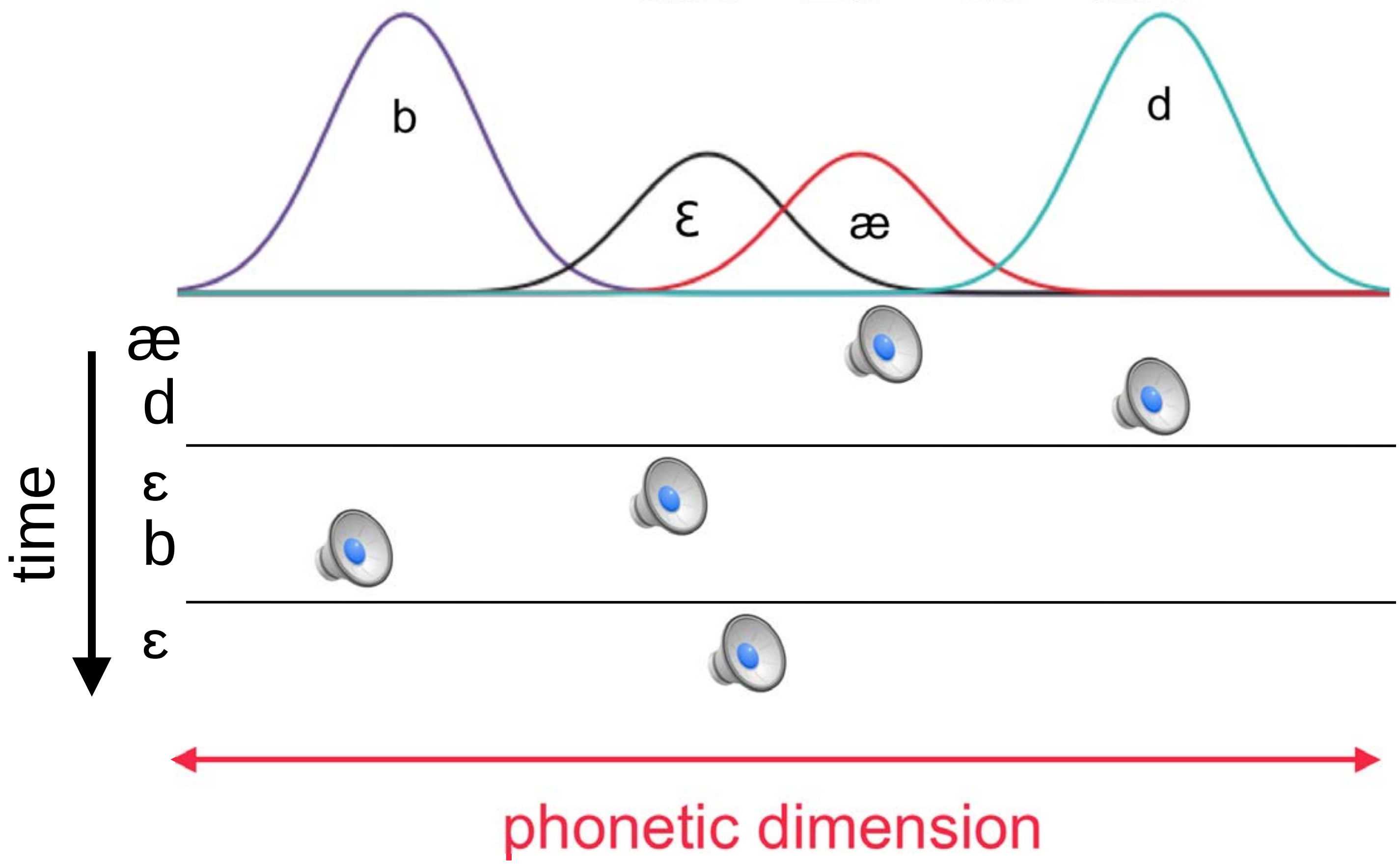
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æb/, /ɛb/
‘add’ ‘Ed’ ‘ab’ ‘ebb’



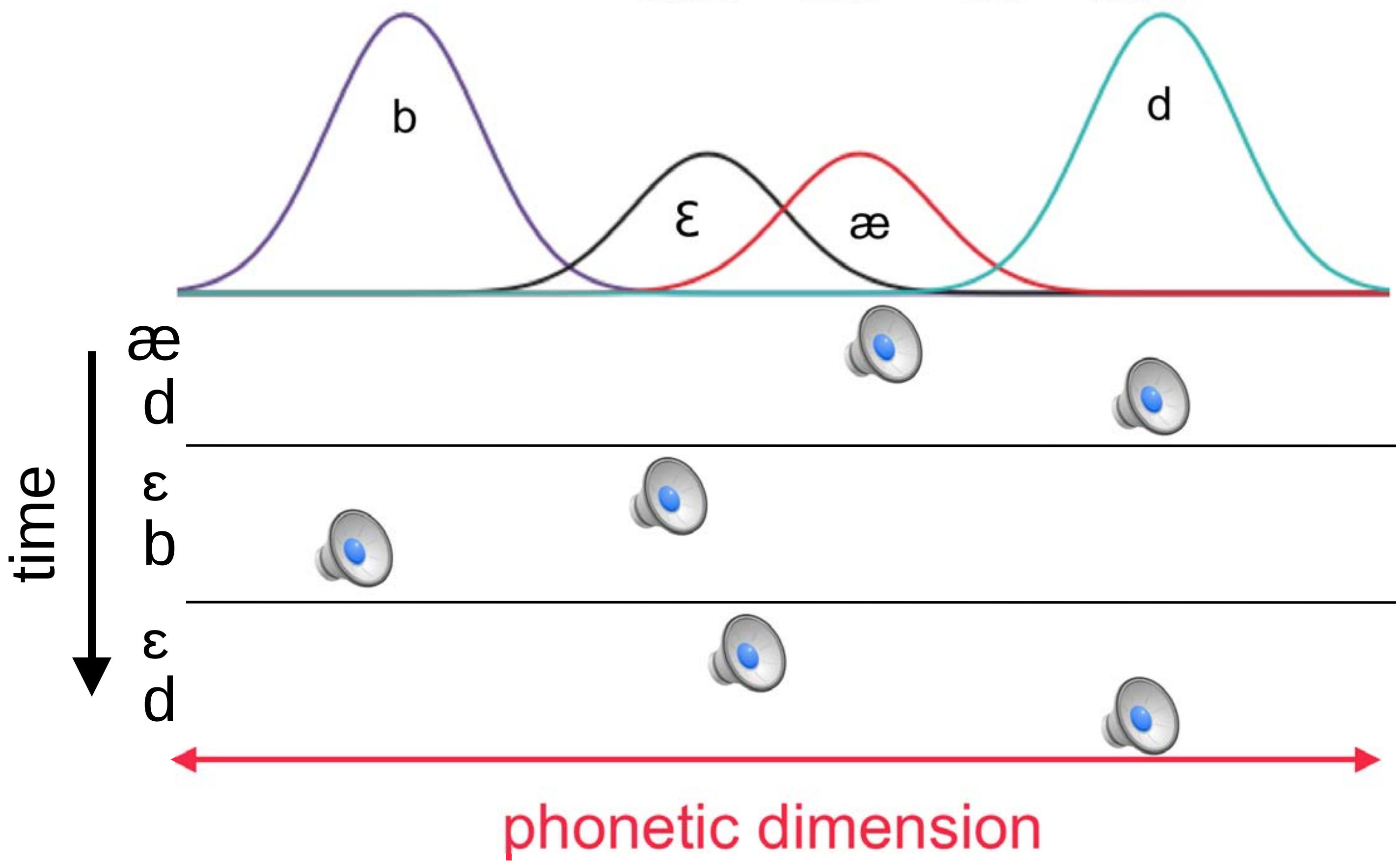
Minimal Pair Lexicon

Lexicon: /æd/, /ɛd/, /æb/, /ɛb/
'add' 'Ed' 'ab' 'ebb'

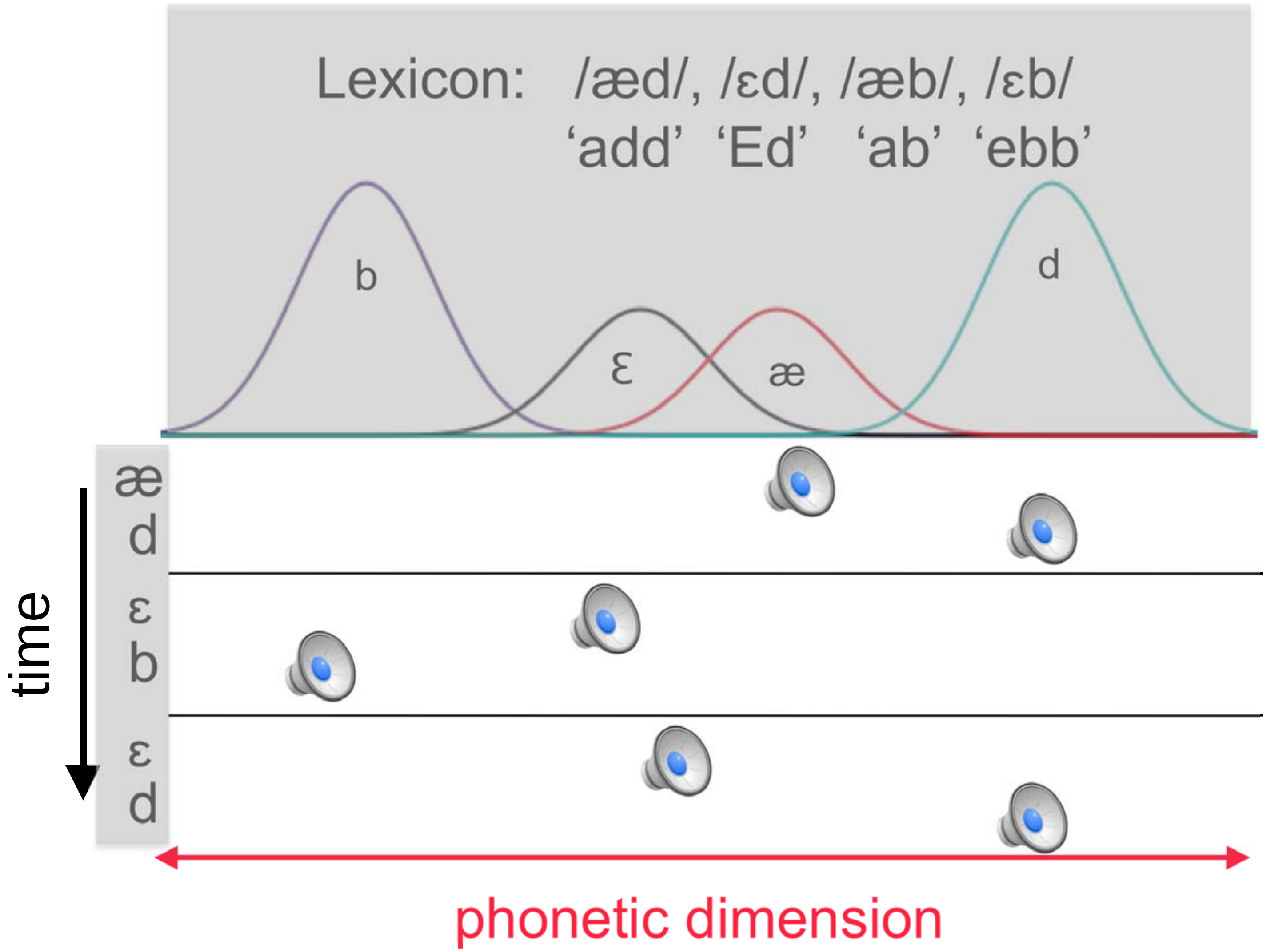


Minimal Pair Lexicon

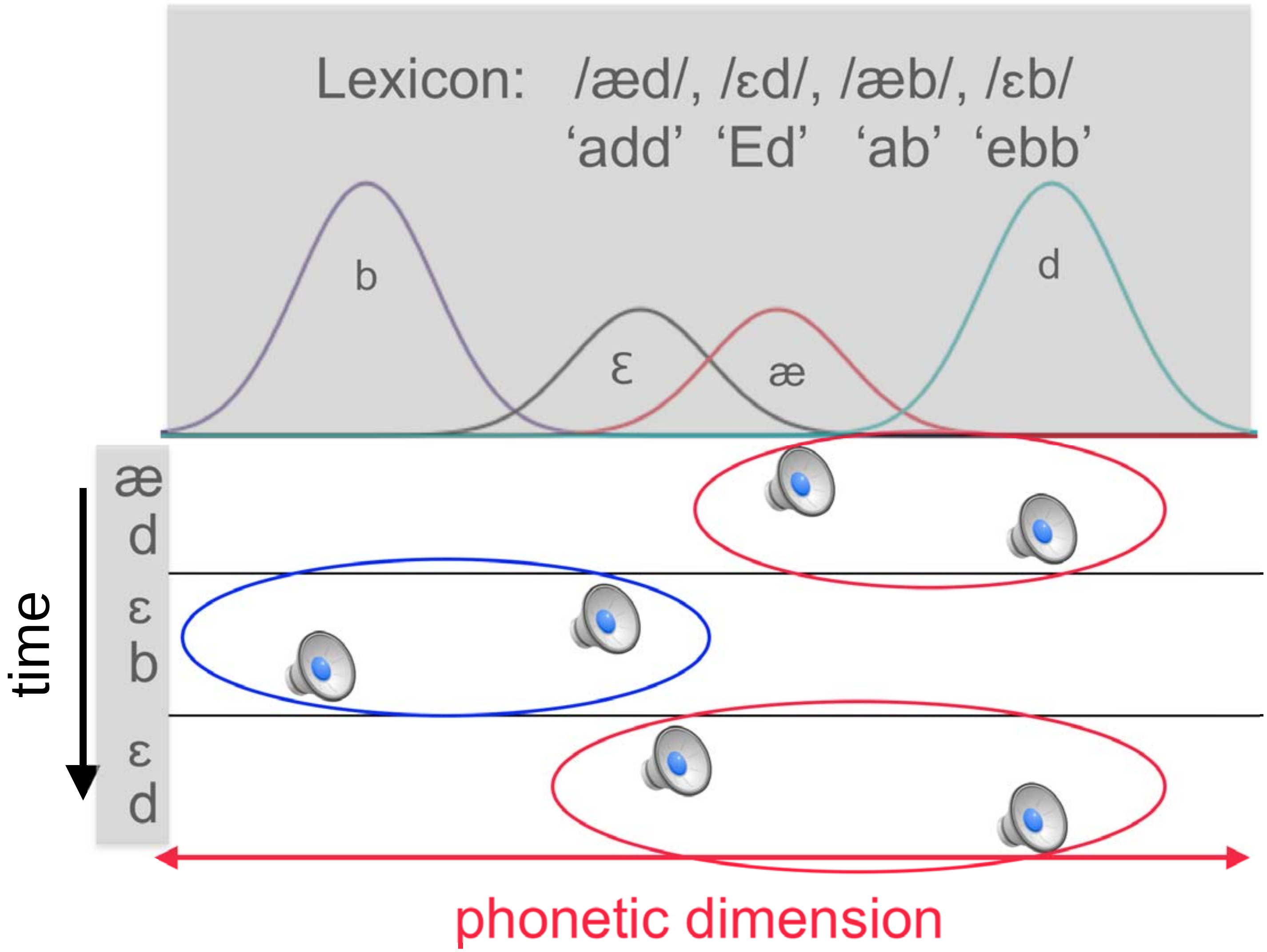
Lexicon: /æd/, /ɛd/, /æb/, /ɛb/
'add' 'Ed' 'ab' 'ebb'



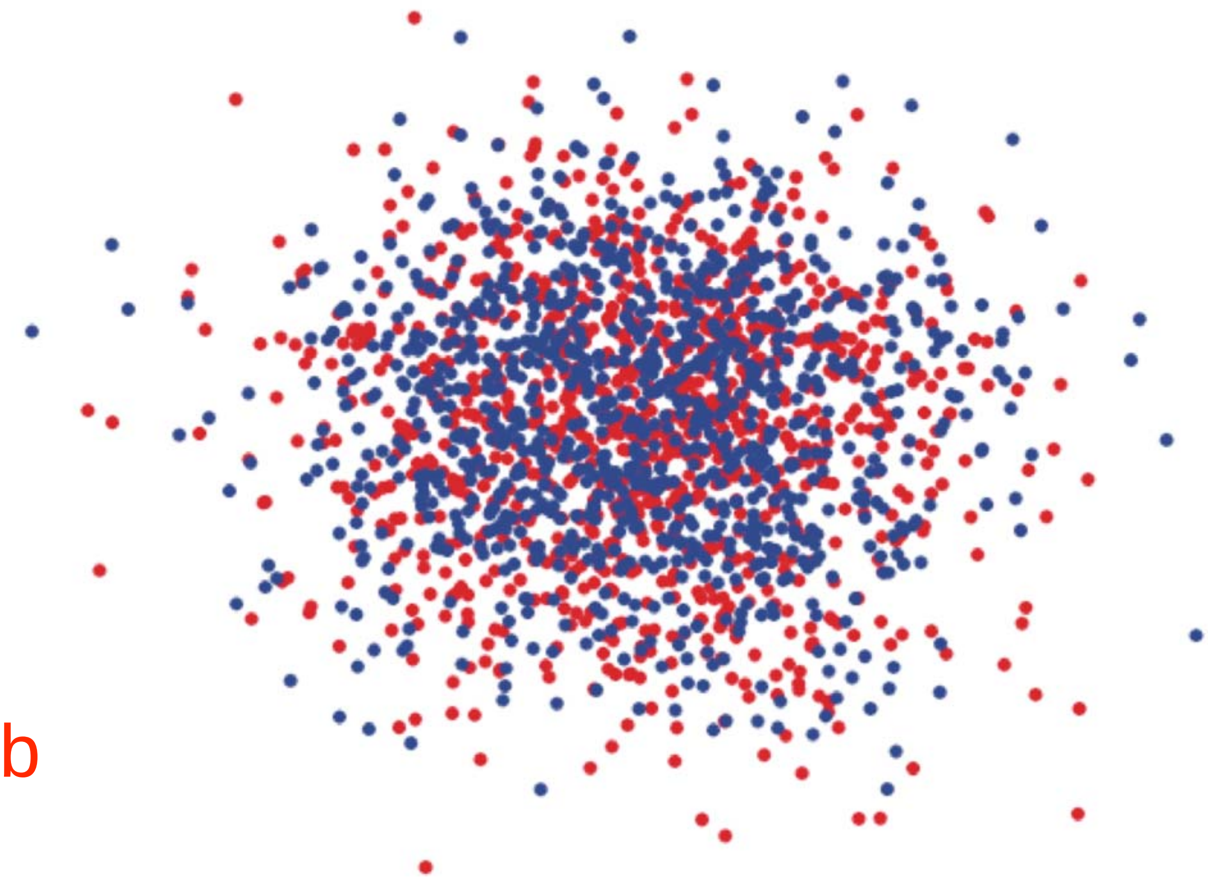
Minimal Pair Lexicon



Minimal Pair Lexicon



Minimal Pair Lexicon



ab

Ed ebb

add

Lexical-Distributional Learning

- If the lexicon contains disambiguating information, the learner should use this information to disambiguate overlapping categories

Lexical-Distributional Learning

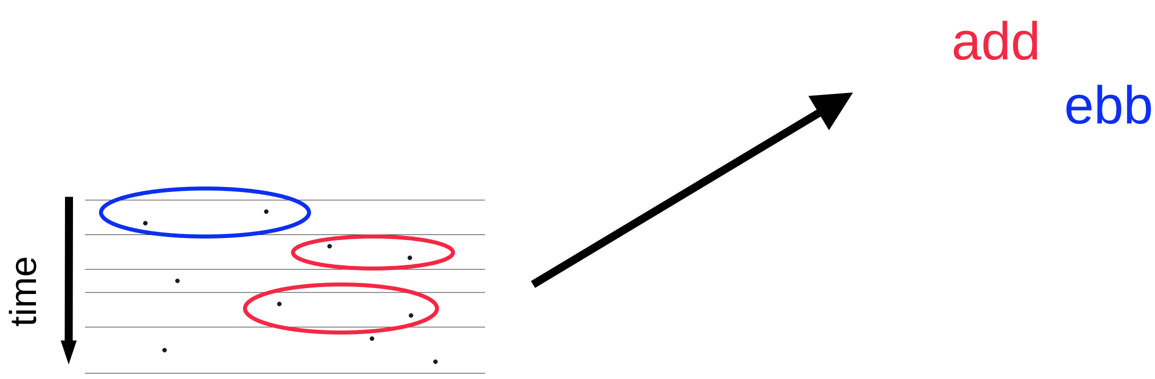
- If the lexicon contains disambiguating information, the learner should use this information to disambiguate overlapping categories

- Learner uses each level of structure to constrain the other:

Lexical-Distributional Learning

- If the lexicon contains disambiguating information, the learner should use this information to disambiguate overlapping categories

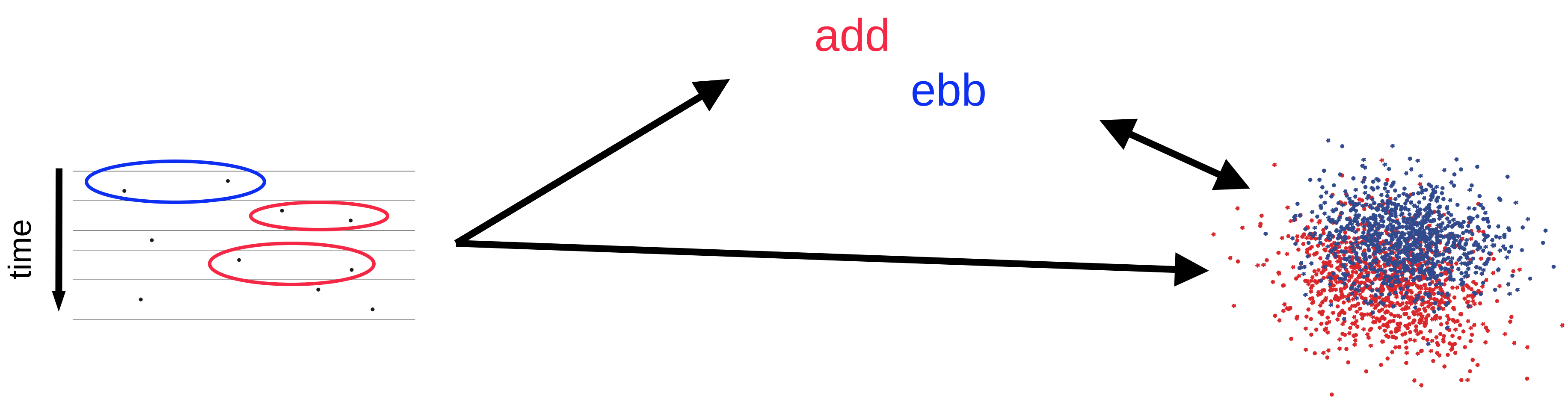
- Learner uses each level of structure to constrain the other:
  - Distributional information helps determine which words are tokens of the same lexical item



Lexical-Distributional Learning

- If the lexicon contains disambiguating information, the learner should use this information to disambiguate overlapping categories

- Learner uses each level of structure to constrain the other:
  - Distributional information helps determine which words are tokens of the same lexical item
  - Lexical information helps determine which sounds are part of the same phonetic category.



Empirical Evidence

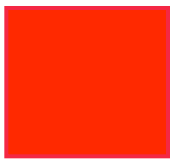
Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)

(Thiessen, 2007)

Empirical Evidence

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

Habituation:



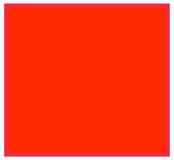
“daw”

(Thiessen, 2007)

Empirical Evidence

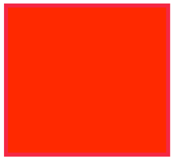
Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)

Habituation:



“daw”

Switch trial:



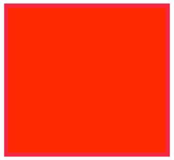
“taw”

(Thiessen, 2007)

Empirical Evidence

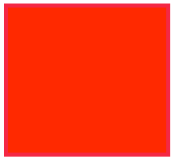
Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)

Habituation:



“daw”

Switch trial:



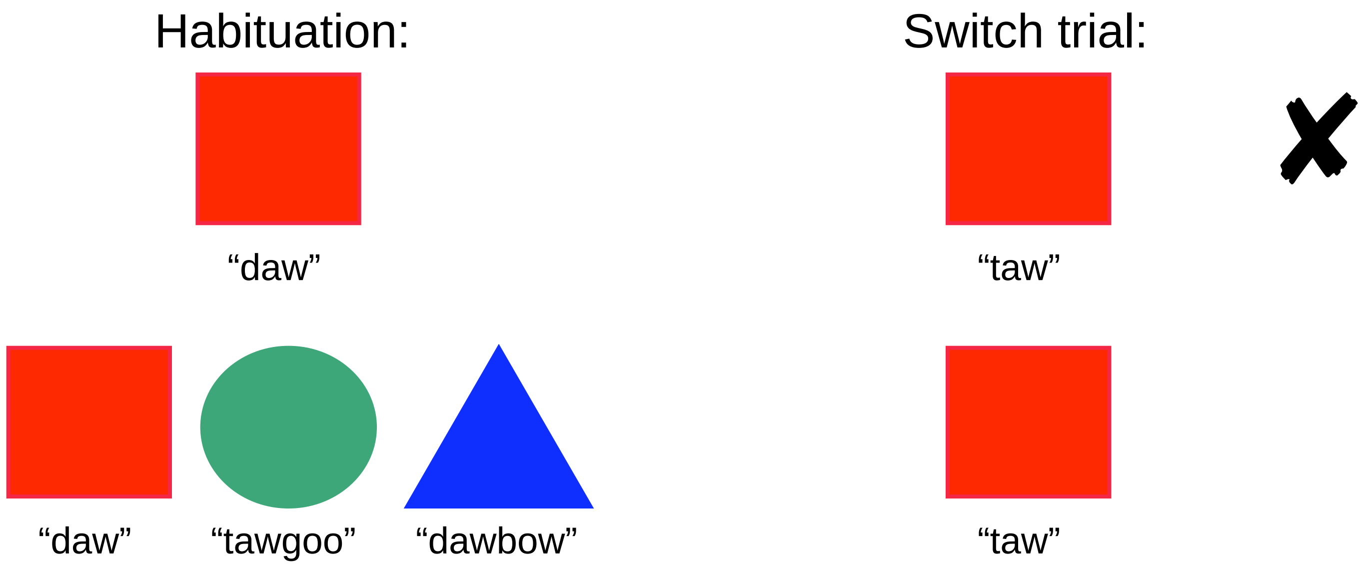
✗

“taw”

(Thiessen, 2007)

Empirical Evidence

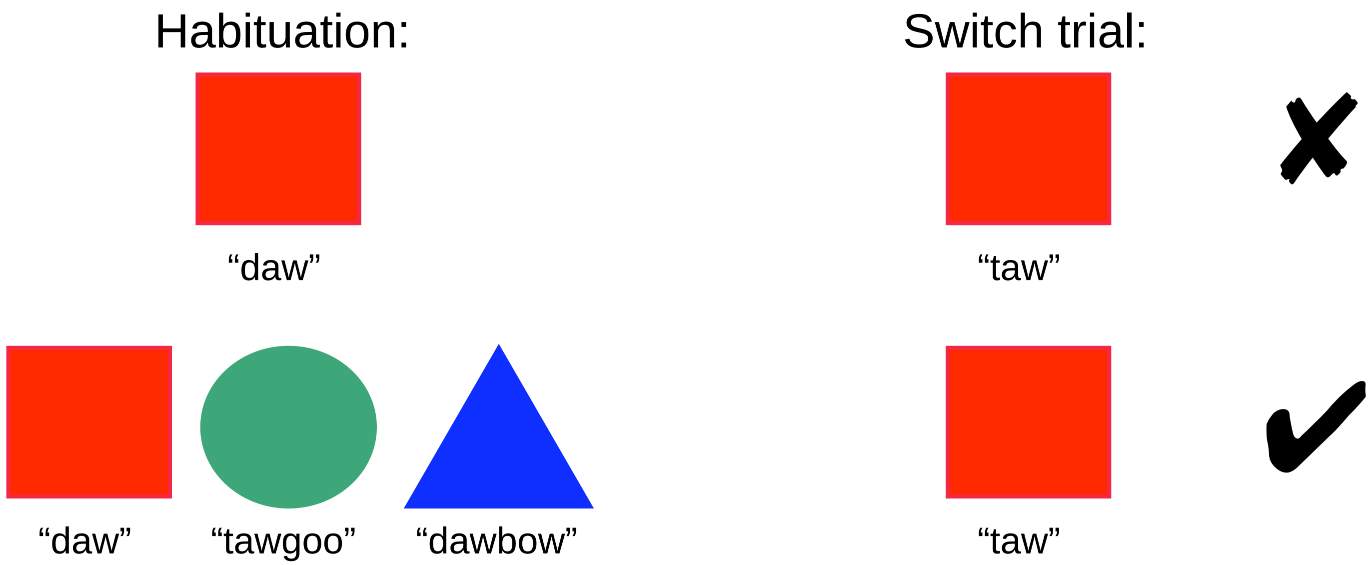
Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)



(Thiessen, 2007)

Empirical Evidence

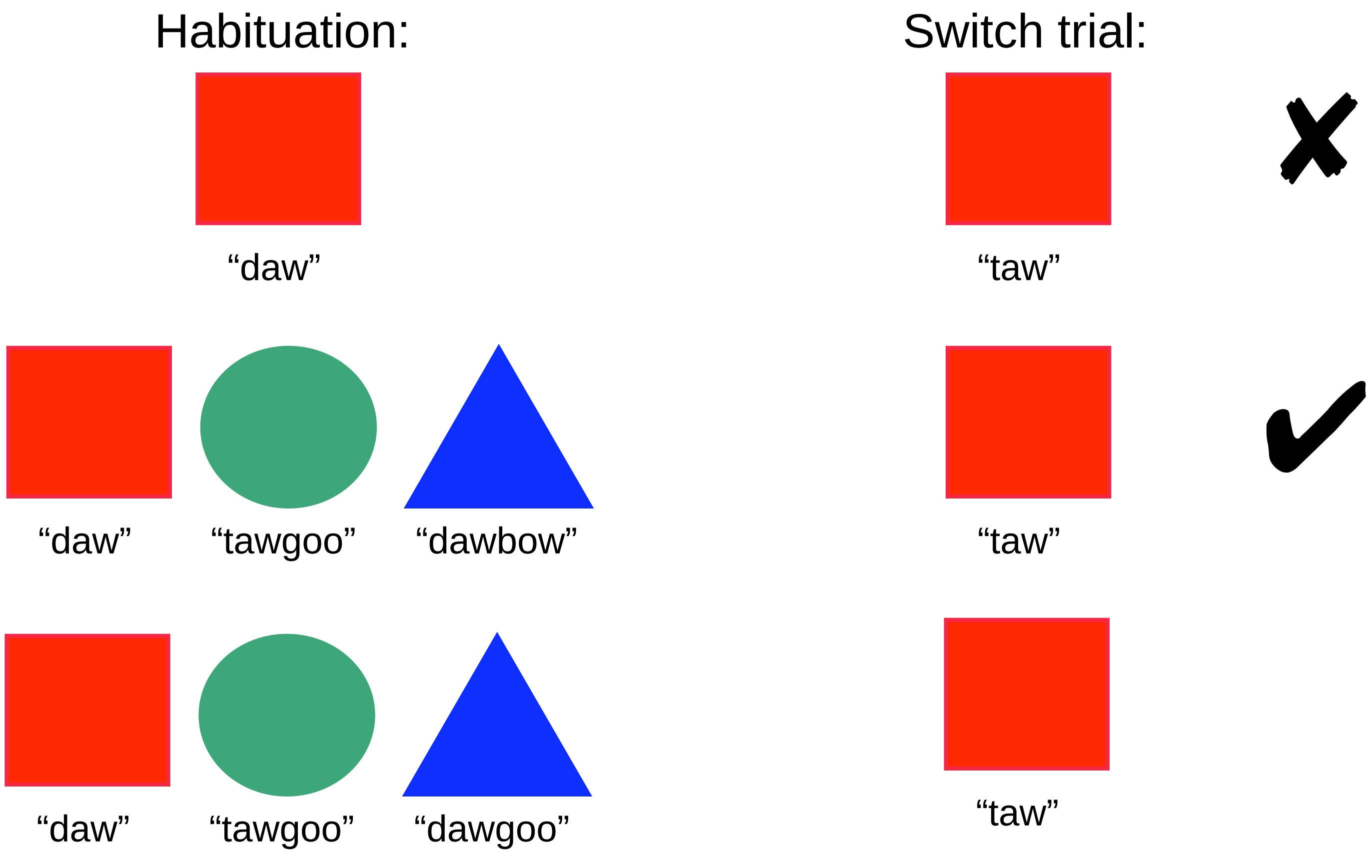
Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)



(Thiessen, 2007)

Empirical Evidence

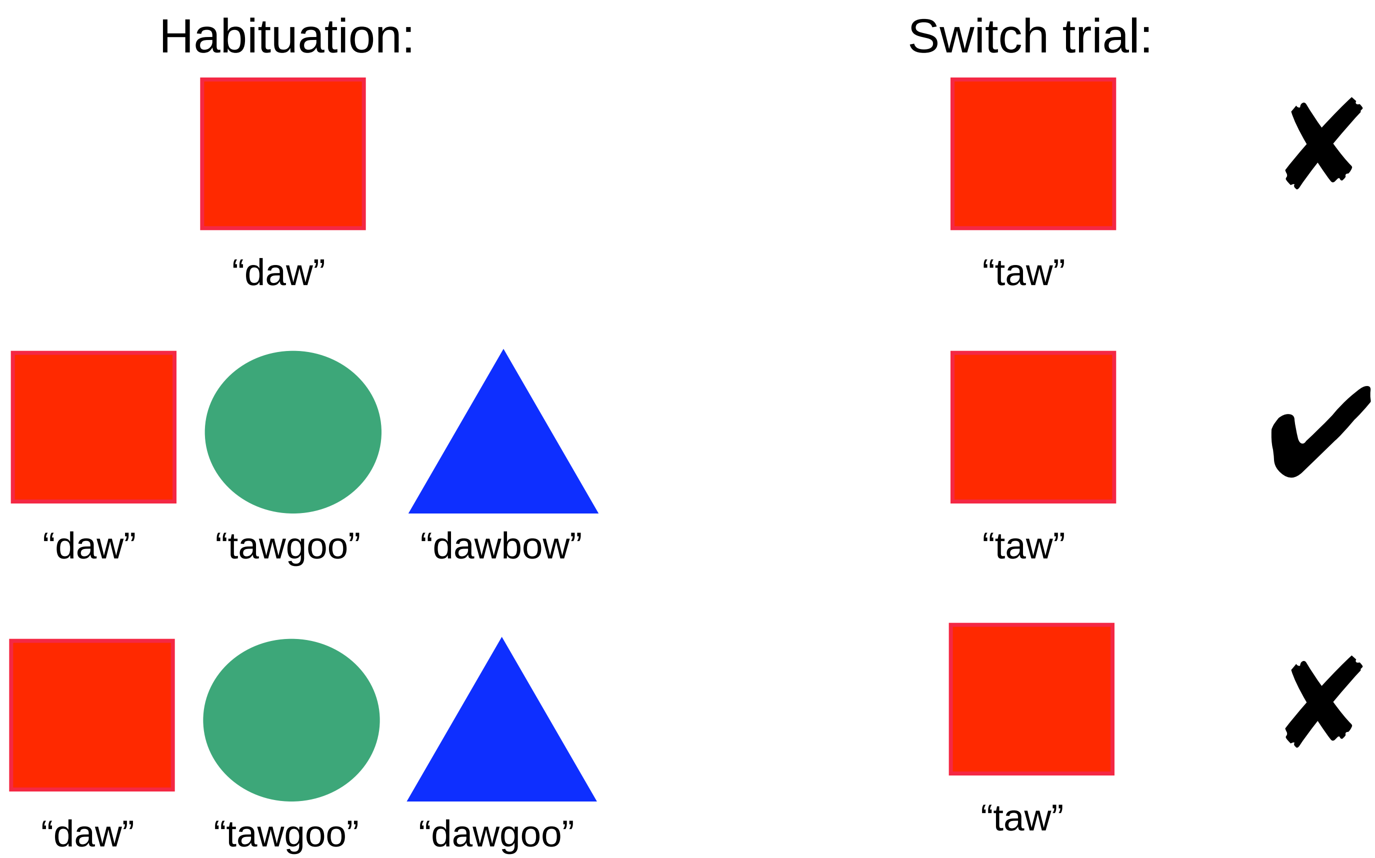
Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)



(Thiessen, 2007)

Empirical Evidence

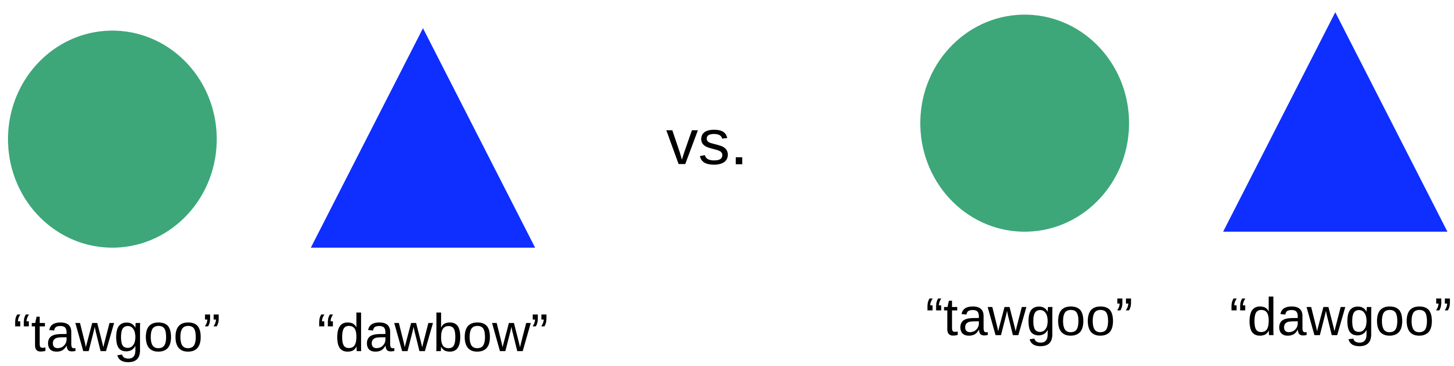
Evidence from 15-month-olds in “switch” task (Stager & Werker, 1997)



(Thiessen, 2007)

Empirical Evidence

- 15-month-olds show better discrimination when lexicon provides disambiguating information (Thiessen, 2007)



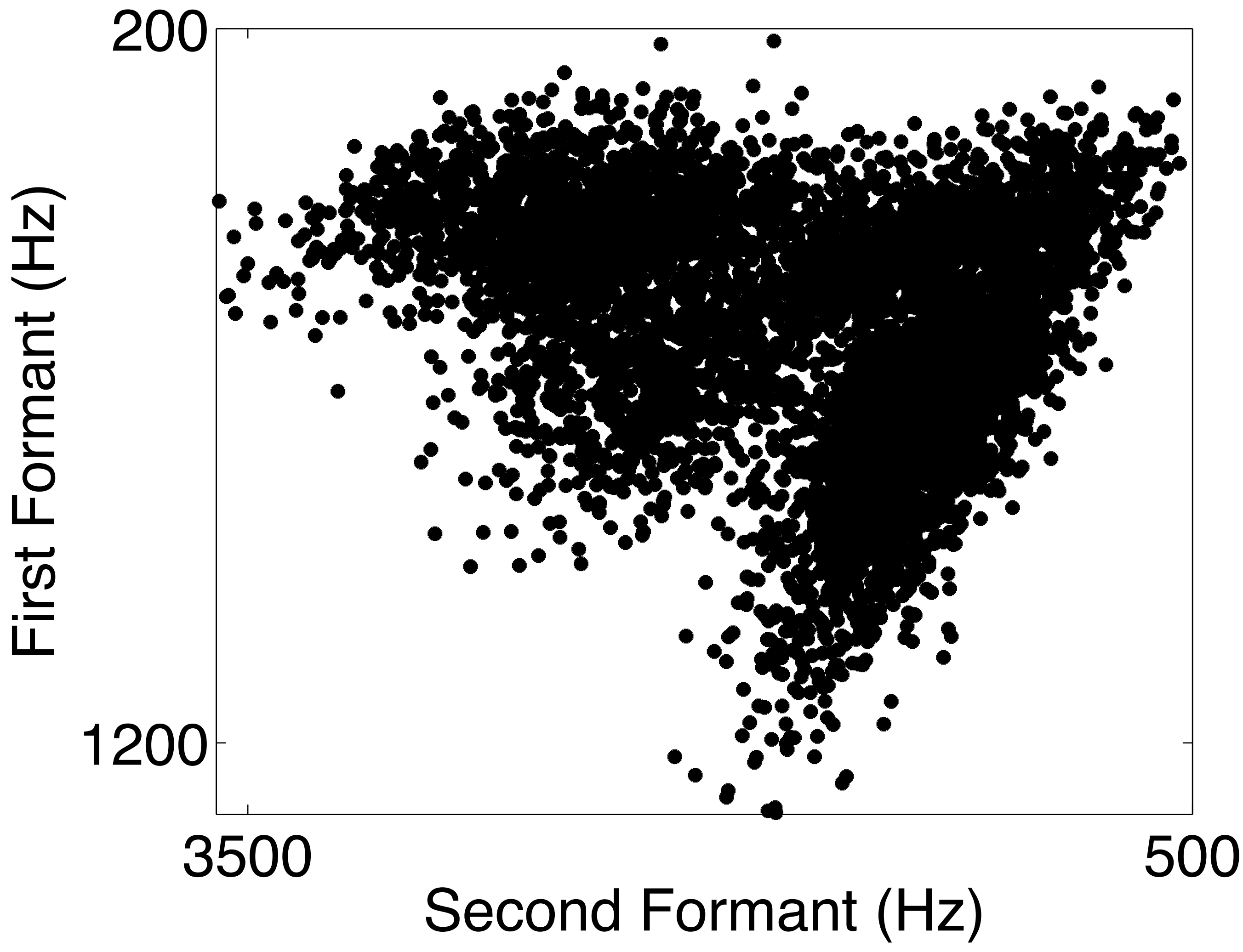
- Adults show similar behavior in a non-referential task when learning about vowel categories (Feldman, Myers, White, Griffiths, & Morgan, 2011)

Outline

- Distributional learning
- Lexical-distributional learning
- Learning English vowels
- Dealing with systematic variability

(Joint work with Tom Griffiths, James Morgan, Sharon Goldwater)

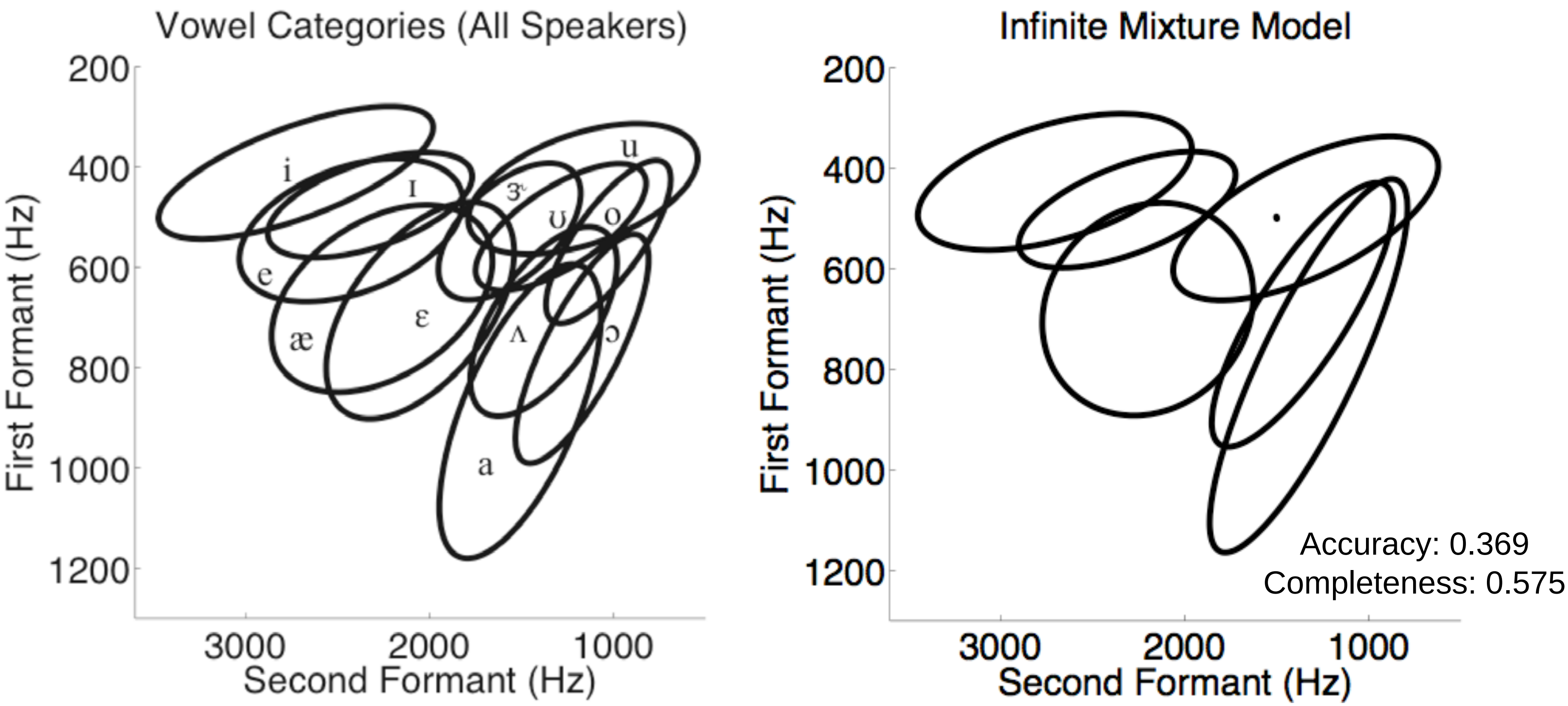
Simulations



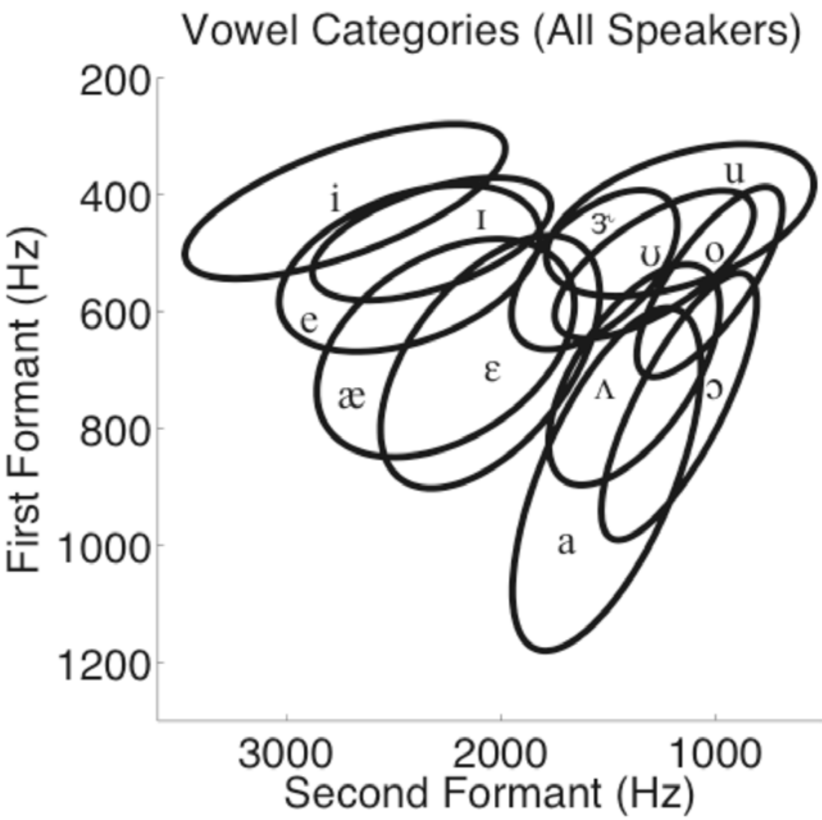
Simulations

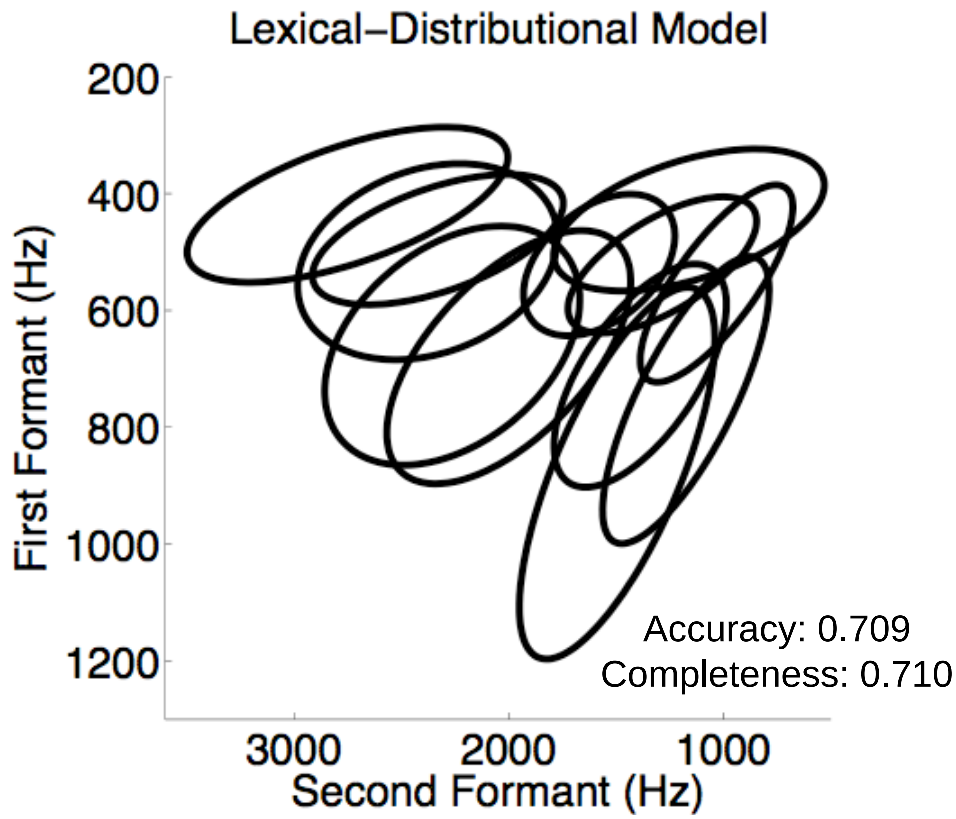
- Lexicon from CHILDES Parental Corpus (Li & Shirai, 2000)
  - Orthographic forms phonematized using Carnegie Mellon Pronouncing Dictionary
  - Lexical items sampled according to corpus frequency
- Corpus of 5000 word tokens, comprising 6,409 vowel tokens and 8,917 consonant tokens
- Acoustic values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

Distributional Model

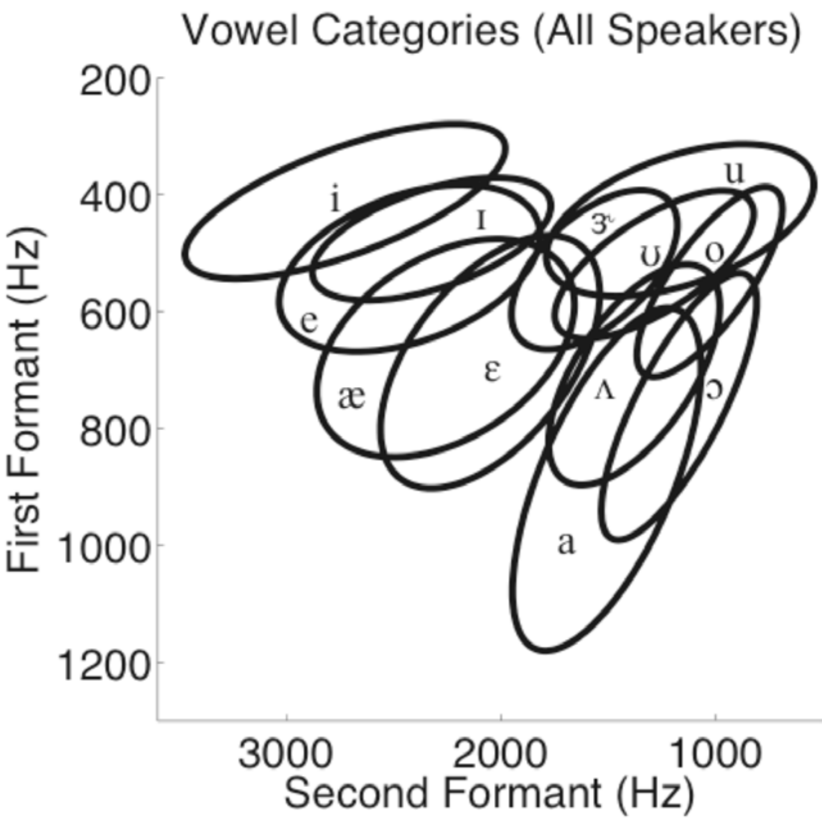


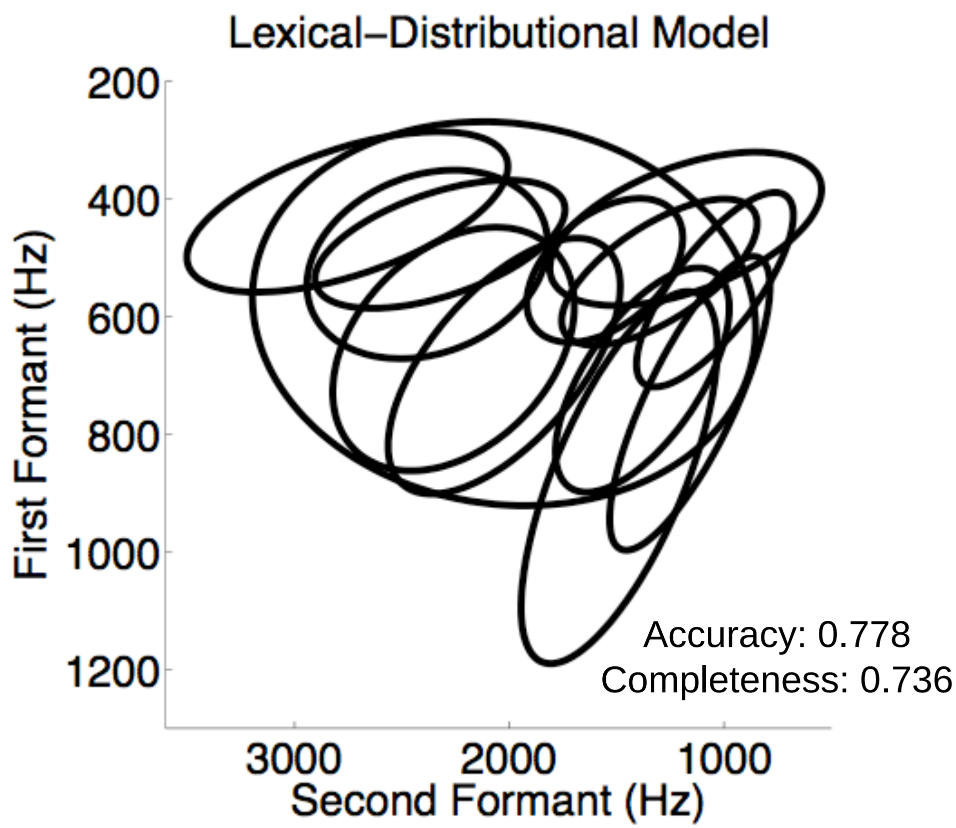
Lexical-Distributional Model





Lexical-Distributional Model





Benefit of Using Words

F-Score

Distributional	Lexical-Distributional (α_L=10)	Lexical-Distributional (α_L=10,000)
0.45	0.76	0.74

Number of Categories

	Lexical-Distributional Model						Distributional Model

Number of Categories

α_C	Lexical-Distributional Model						Distributional Model
0.1							
1							
10							

Number of Categories

α_C	Lexical-Distributional Model						Distributional Model
	α_L=1	α_L=10	α_L=100	α_L=1,000	α_L=10,000		
0.1							
1							
10							

Number of Categories

α_C	Lexical-Distributional Model						Distributional Model
	α_L=1	α_L=10	α_L=100	α_L=1,000	α_L=10,000		
0.1							6
1							6
10							7

Number of Phonetic Categories (gold standard = 12)

Number of Categories

α_C	Lexical-Distributional Model						Distributional Model
	α_L=1	α_L=10	α_L=100	α_L=1,000	α_L=10,000		
0.1	14	13	13	12	12		6
1	14	14	13	12	12		6
10	14	13	13	12	12		7

Number of Phonetic Categories (gold standard = 12)

Number of Categories

α_C	Lexical-Distributional Model						Distributional Model
	α_L=1	α_L=10	α_L=100	α_L=1,000	α_L=10,000		
0.1	14	13	13	12	12		6
1	14	14	13	12	12		6
10	14	13	13	12	12		7

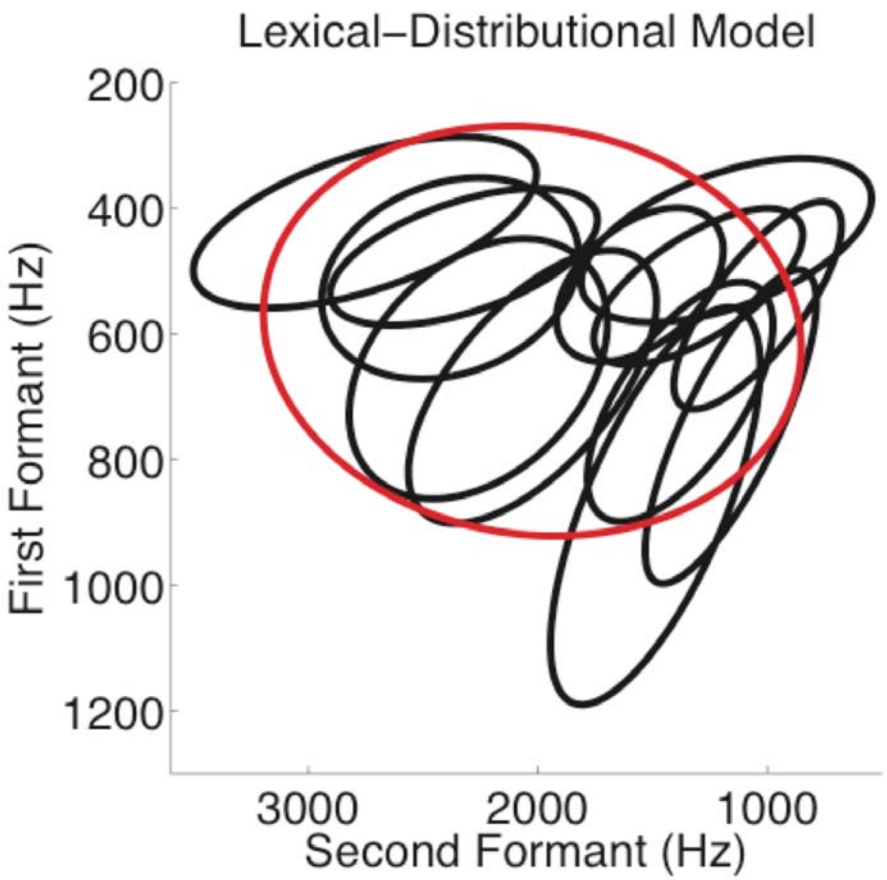
Number of Phonetic Categories (gold standard = 12)
Number of Lexical Items (gold standard = 1019)

Number of Categories

α_C	Lexical-Distributional Model						Distributional Model
	α_L=1	α_L=10	α_L=100	α_L=1,000	α_L=10,000		
0.1	14 900	13 916	13 969	12 1145	12 1601		6
1	14 899	14 912	13 968	12 1138	12 1605		6
10	14 900	13 926	13 958	12 1164	12 1602		7

Number of Phonetic Categories (gold standard = 12)
Number of Lexical Items (gold standard = 1019)

Lexical-Distributional Model



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

Minimal Pairs

- Phonologists use minimal pairs to identify contrastive categories
- Minimal pairs make it *more* difficult to distinguish between phonemes if no meanings are known: items in the pair could be the same word
- Model can overcome minimal pair problem with certain parameter values, but children may use other strategies

More Interactions in Learning?

- Phonotactics
  - Sensitivity to phonotactics at 9 months could make a learner more willing to accept multiple lexical items with a common consonant frame (Jusczyk et al., 1994)

- Semantics
  - Semantic information may help pull apart minimal pairs (Yeung & Werker, 2009; but see Thiessen, 2007)
  - Semantic information may help a learner recognize redundant lexical items

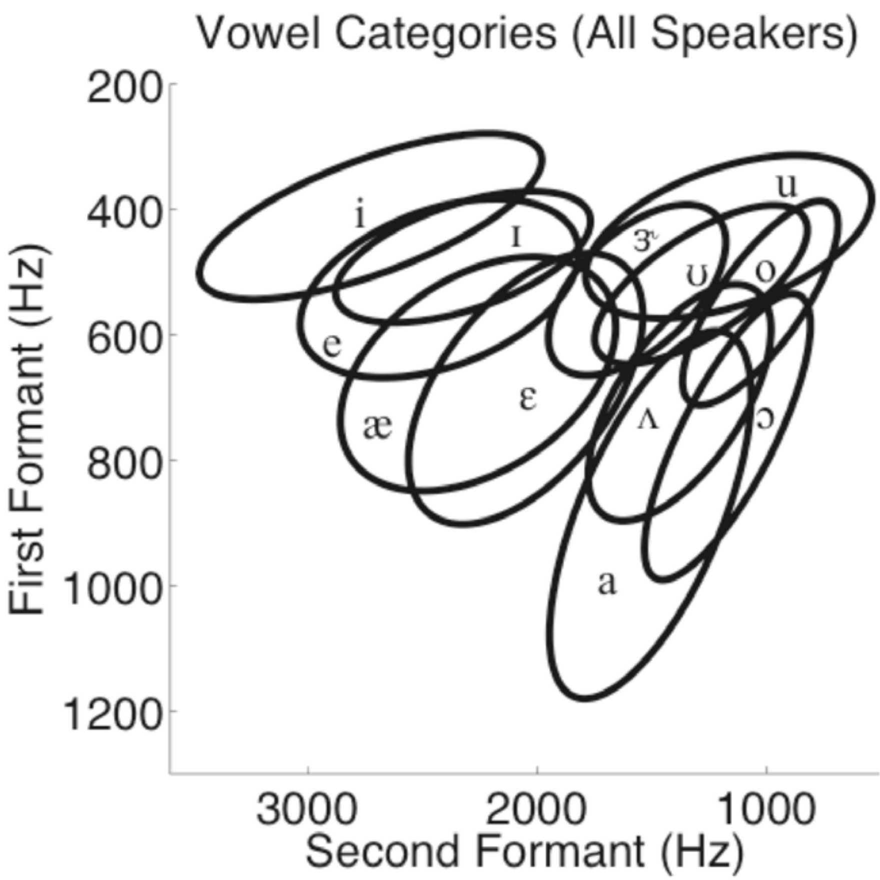
Simulations

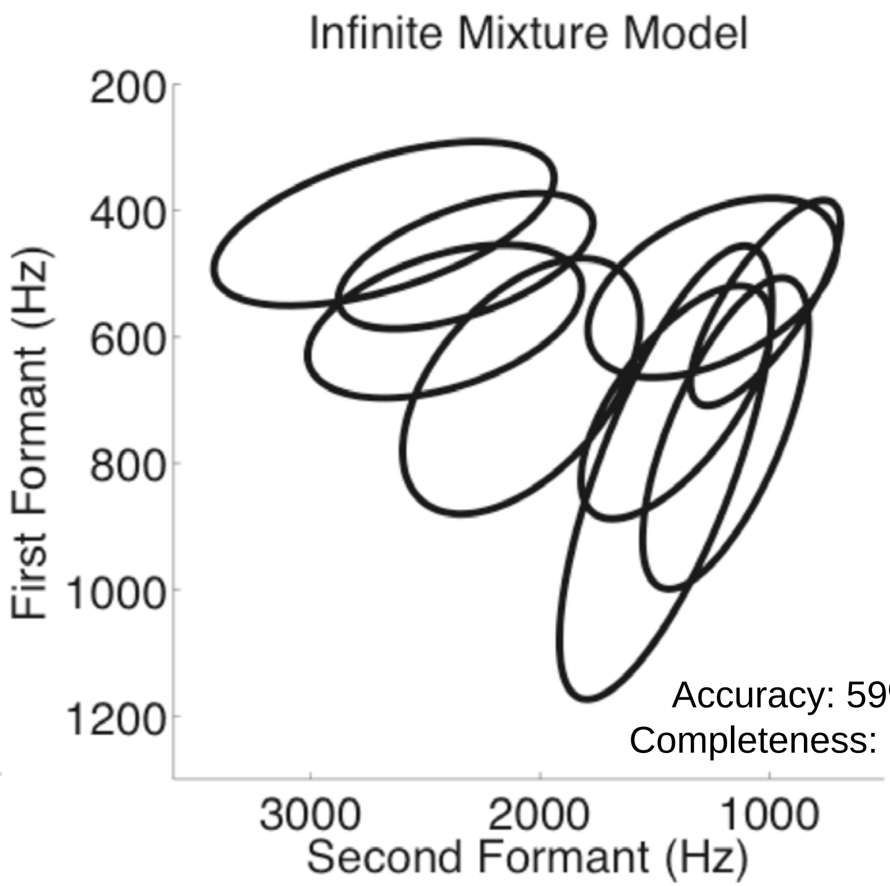
- Lexicon generated from the model
  - Words composed only of vowels
  - Structure of the lexicon matches the learner's expectations

- Corpus of 5000 word tokens, comprising 22,397 vowel tokens

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

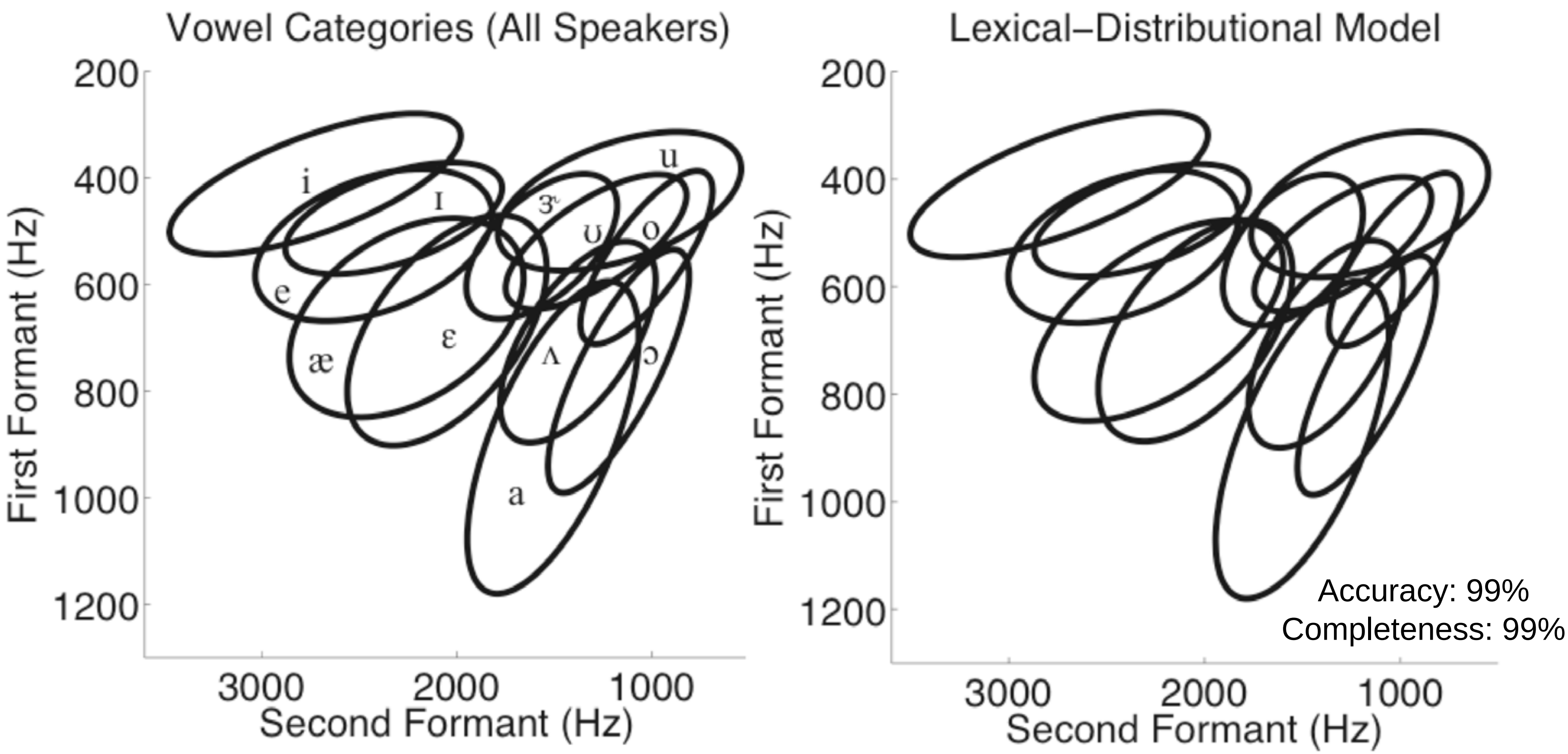
Distributional Model





Accuracy: 59%
Completeness: 62%

Lexical-Distributional Model



Summary

- Using information from words can help disambiguate overlapping categories, even if the forms in the lexicon are not given explicitly to the learner
- Qualitative behavior mimics human data
- Interactive learning poses different challenges than learning each domain in isolation
  - Disambiguating overlapping categories is difficult in isolation
  - Similar-sounding words are difficult for interactive learner

Outline

- Distributional learning
- Lexical-distributional learning
- Learning English vowels
- Dealing with systematic variability

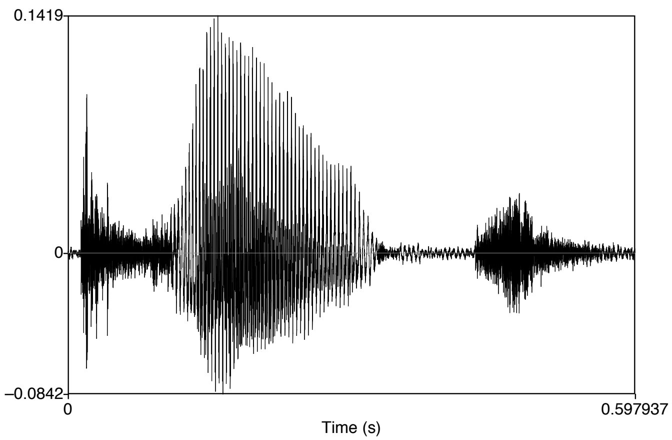
Work by Ewan Dunbar, Brian Dillon, & Bill Idsardi
More information: http://ling.umd.edu/~emd/ or emd@umd.edu

Phonological Alternations

Lexical-distributional model assumes a single Gaussian distribution for a phonetic category, regardless of context

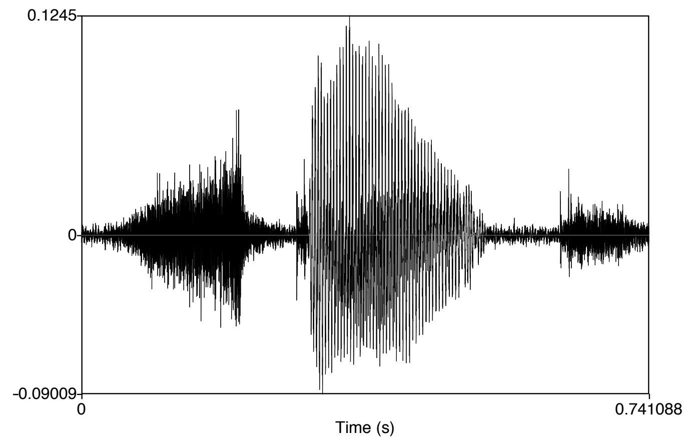
What about phonological alternations?

Phonological Alternations



“Kate”

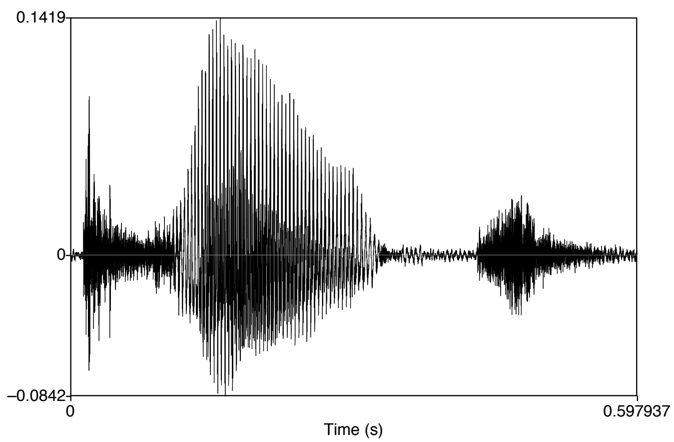
[k^h] at the beginning
of a stressed syllable



“skate”

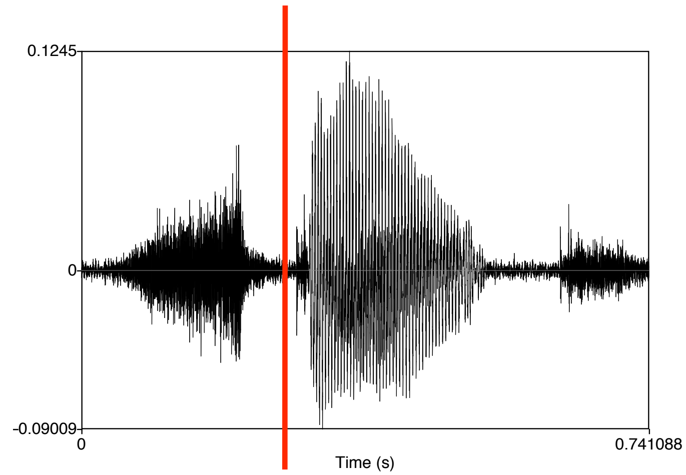
[k] in an ‘sk’ cluster

Phonological Alternations



“Kate”

[k^h] at the beginning of a stressed syllable



“skate”

[k] in an ‘sk’ cluster

Phonological Alternations

[k] and [k^h] are allophones of the same phoneme

- Complementary distribution: [k] and [k^h] appear in different phonological contexts
- No minimal pairs involving [k] and [k^h]
- Speakers and listeners think of [k] and [k^h] as “the same sound”

Typically characterized by a rule:

$k \rightarrow k^h$ at the beginning of a stressed syllable

Learning Phonemes: Option 1

Two stages:

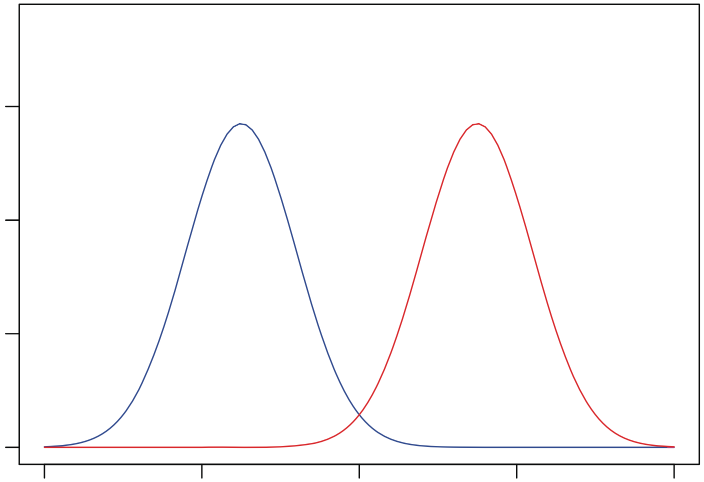
1. Learn separate phonetic categories for [k] and [k^h]

2. In a separate learning process, notice that the [k] and [k^h] occur in complementary distribution, and infer that they are allophones of a single phoneme

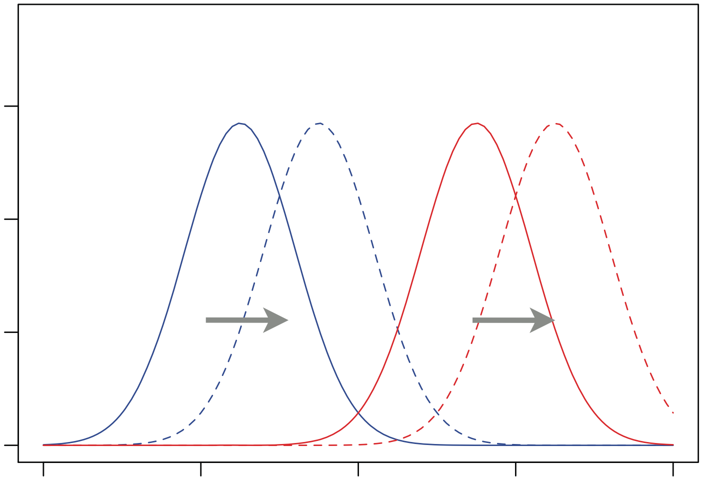
Learning Phonemes: Option 2

Give up the assumption that sound categories are Gaussian distributions

categories are Gaussians



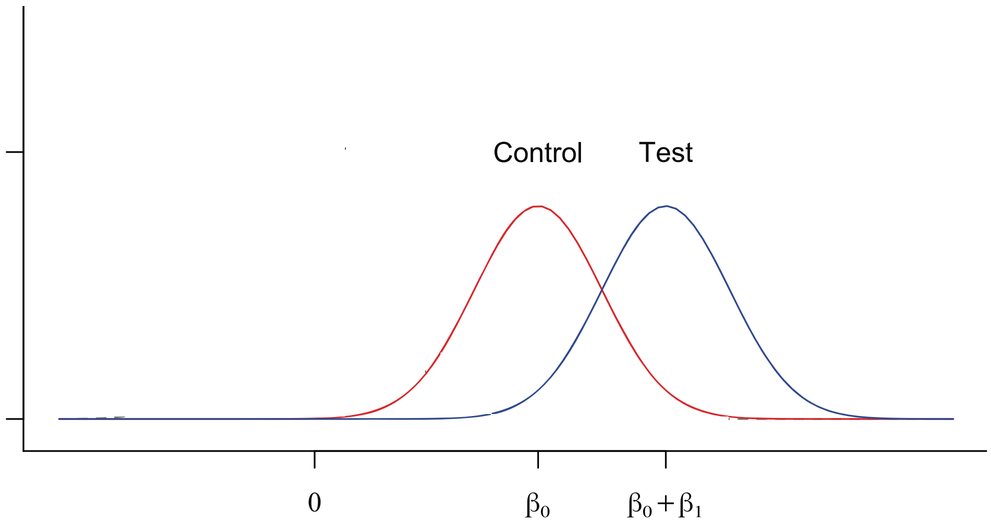
categories are linear models



(Dunbar, Dillon, & Idsardi, in preparation)

Linear Models

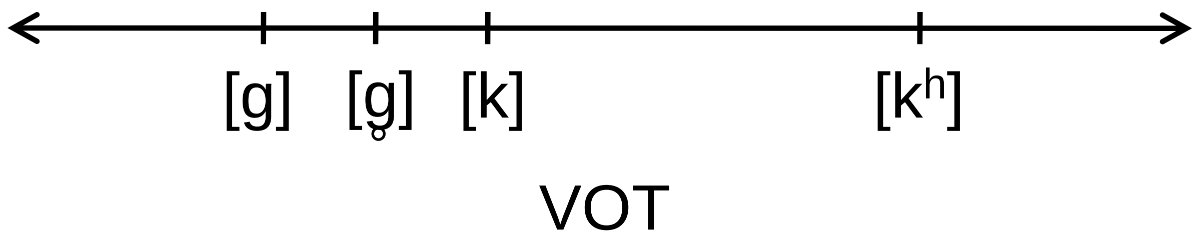
$$Y_i \sim N(\boxed{\boxed{\beta_0} + \beta_1} X_{1i})$$



t-test/ANOVA

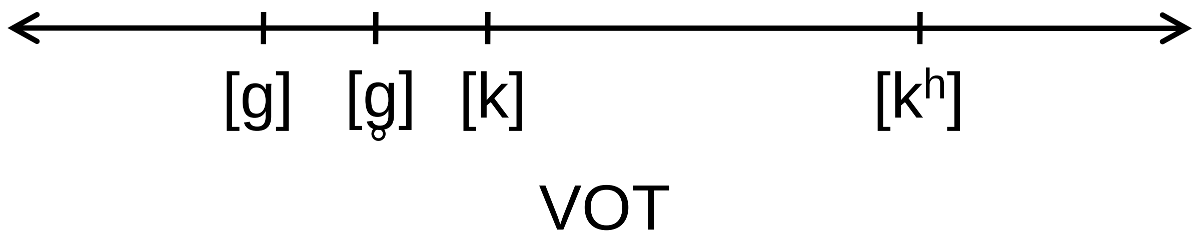
(Dunbar, Dillon, & Idsardi, in preparation)

Mixture of Linear Models

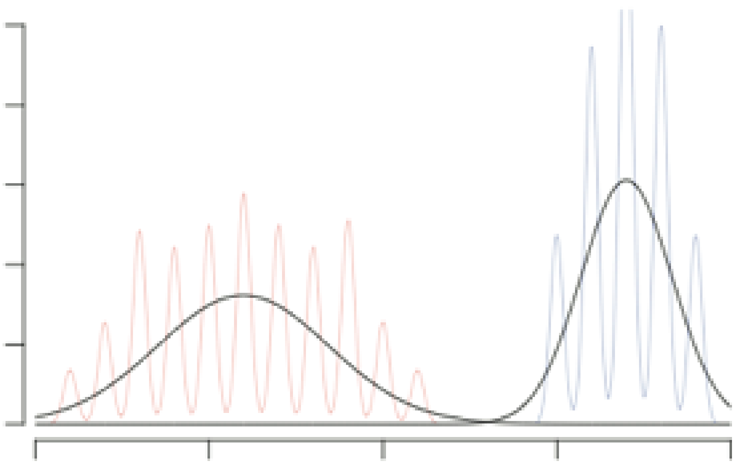


(Dunbar, Dillon, & Idsardi, in preparation)

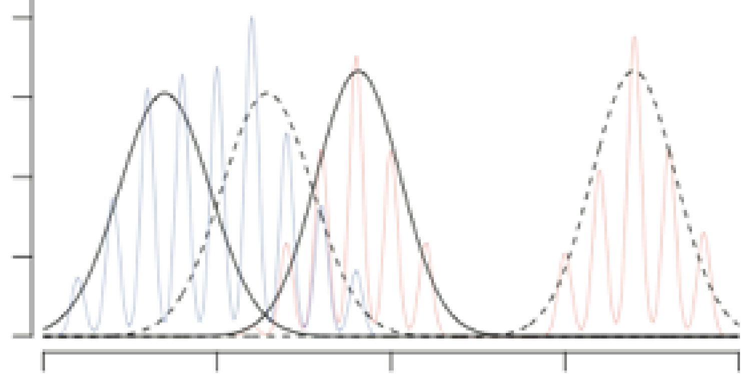
Mixture of Linear Models



categories are Gaussians



categories are linear models



(Dunbar, Dillon, & Idsardi, in preparation)

Mixture of Linear Models

Inuktitut: Vowels change before uvular consonants

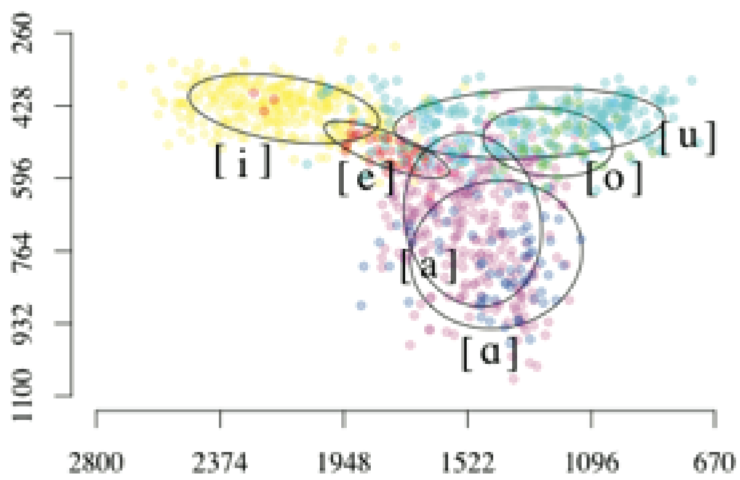
i → e

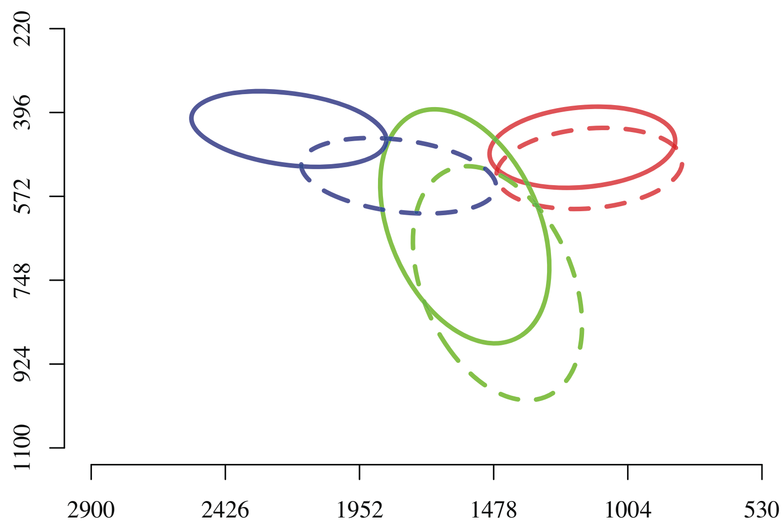
u → o

a → ɑ

e.g., /aŋijuk/ → [aŋijuk] "older sibling"

/aŋijuq/ → [aŋijoq] "big"





(Dunbar, Dillon, & Idsardi, in preparation)

How are sound categories learned?

An Inference Problem

Learner recovering linguistic structure

Hypotheses: possible linguistic analyses
Data: corpus (language input)

$$p(h \mid d) \propto p(d \mid h)p(h)$$

An Inference Problem

Learner recovering linguistic structure

Hypotheses: possible linguistic analyses
Data: corpus (language input)

$$p(h \mid d) \propto p(d \mid h)p(h)$$

What types of hypotheses should learners consider?

An Inference Problem

Distributional Model

Phonetic Categories

↓

Corpus

Lexical-Distributional Model

Phonetic Categories

↓

Lexicon

↓

Corpus

Mixture of Linear Models

Phonemes

↓

Corpus

?

Phonetic Category Learning

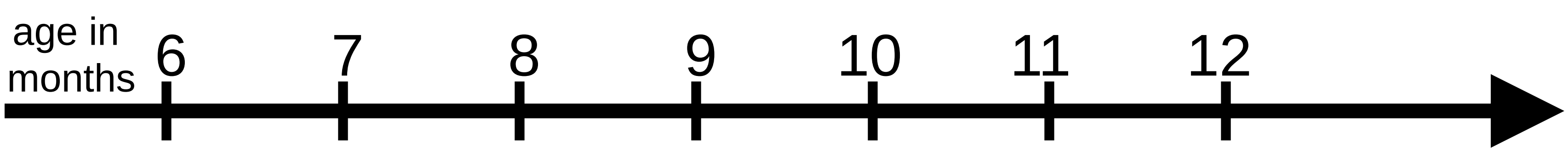
6 months: some language-specific perception of vowels

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)



Phonetic Category Learning

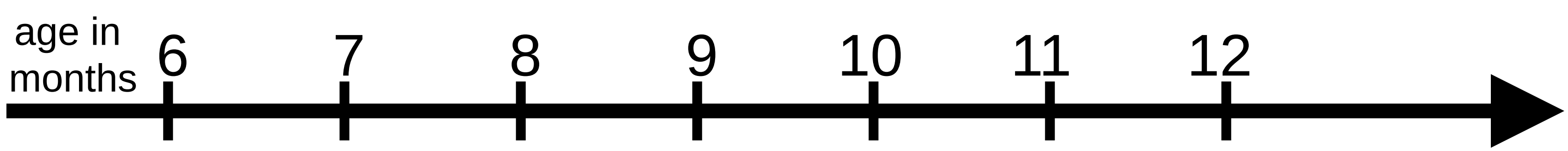
6 months: some language-specific perception of vowels

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)

Acknowledgements

Word model joint work with Tom Griffiths, James Morgan, Sharon Goldwater; allophone model by Ewan Dunbar, Brian Dillon, Bill Idsardi

Thanks to:

Sheila Blumstein, Adam Darlow, Mark Johnson, Joseph Williams

Metcalf Infant Research Lab (Brown)

Computational Cognitive Science Lab (Berkeley)

Computational Modeling Reading Group (Brown)

NSF grants BCS-0924821, BCS-0631518

NIH grant HD032005

AFOSR grant FA9550-07-1-0351