The learnability of tones from the speech signal

Kristine M. Yu

Department of Linguistics
University of Maryland College Park
University of Massachusetts Amherst

NECPhon, Yale University
October 15, 2011

Overview

1. **What is the setting of the learning problem for learning phonological categories?**

Overview

1. **What is the setting of the learning problem for learning phonological categories?**
2. **What structure might there be in the hypothesis space for learning phonological categories?**

Overview

1. **What is the setting of the learning problem for learning phonological categories?**
2. **What structure might there be in the hypothesis space for learning phonological categories?**

▶ Model system: lexical tones in tonal languages

Overview

1. **What is the setting of the learning problem for learning phonological categories?**
2. **What structure might there be in the hypothesis space for learning phonological categories?**

▶ Model system: lexical tones in tonal languages
▶ Methods:
   0. Theoretical inquiry
   1. Cross linguistic fieldwork
   2. Psychological experiments
   3. Computational modeling

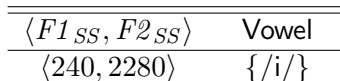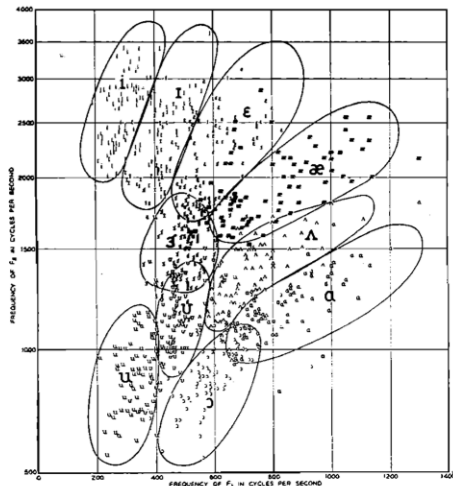# The target of learning: a vowel map in 2-D formant space

Figure: Peterson and Barney (1952): An
English vowel map in $\langle F1_{SS}, F2_{SS} \rangle$ space

The target of learning: a vowel map in 2-D formant space

$\langle F1_{SS}, F2_{SS}\rangle$	Vowel
$\langle 240, 2280\rangle$	$\{/i/\}$

Figure: Peterson and Barney (1952): An English vowel map in $\langle F1_{SS}, F2_{SS}\rangle$ space

The target of learning: a vowel map in 2-D formant space

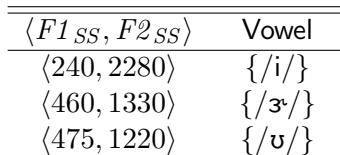| $\langle F1_{SS}, F2_{SS}\rangle$	Vowel
$\langle 240, 2280\rangle$	$\{/i/\}$
$\langle 460, 1330\rangle$	$\{/ɝ/\}$

Figure: Peterson and Barney (1952): An
English vowel map in $\langle F1_{SS}, F2_{SS}\rangle$ space

The target of learning: a vowel map in 2-D formant space

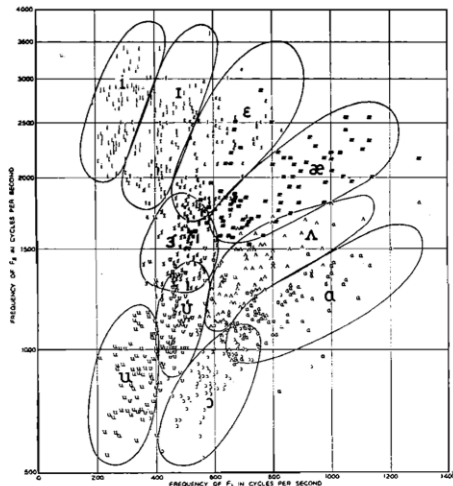| $\langle F1_{SS}, F2_{SS} \rangle$	Vowel
$\langle 240, 2280 \rangle$	$\{/i/\}$
$\langle 460, 1330 \rangle$	$\{/ɝ/\}$
$\langle 475, 1220 \rangle$	$\{/ʊ/\}$

Figure: Peterson and Barney (1952): An English vowel map in $\langle F1_{SS}, F2_{SS} \rangle$ space
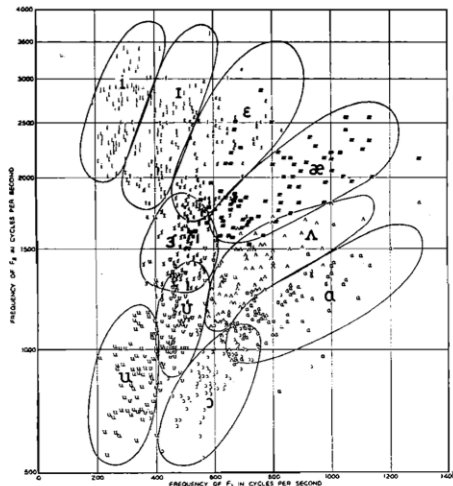
The target of learning: a vowel map in 2-D formant space

$\langle F1_{SS}, F2_{SS}\rangle$	Vowel
$\langle 240, 2280\rangle$	$\{/i/\}$
$\langle 460, 1330\rangle$	$\{/ɝ/\}$
$\langle 475, 1220\rangle$	$\{/ʊ/\}$
$\langle 686, 1028\rangle$	$\{/ɑ, ɔ/\}$

Figure: Peterson and Barney (1952): An
English vowel map in $\langle F1_{SS}, F2_{SS}\rangle$ space

A strategy for characterizing the learning problem
Characterizing tonal maps
Defining tonal maps
The learnability of tonal maps

The target of learning: a vowel map in 2-D formant space

$\langle F1_{SS}, F2_{SS} \rangle$	Vowel
$\langle 240, 2280 \rangle$	$\{/i/\}$
$\langle 460, 1330 \rangle$	$\{/ɝ/\}$
$\langle 475, 1220 \rangle$	$\{/ʊ/\}$
$\langle 686, 1028 \rangle$	$\{/ɑ, ɔ/\}$
$\langle 400, 3500 \rangle$	$\{/i/\}$

Figure: Peterson and Barney (1952): An English vowel map in $\langle F1_{SS}, F2_{SS} \rangle$ space

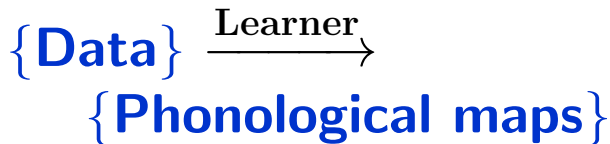# The target of learning: a vowel map in 2-D formant space

$\langle F1_{SS}, F2_{SS} \rangle$	Vowel
$\langle 240, 2280 \rangle$	$\{/i/\}$
$\langle 460, 1330 \rangle$	$\{/ɝ/\}$
$\langle 475, 1220 \rangle$	$\{/ʊ/\}$
$\langle 686, 1028 \rangle$	$\{/ɑ, ɔ/\}$
$\langle 400, 3500 \rangle$	$\{/i/\}$
⋮	⋮

Figure: Peterson and Barney (1952): An English vowel map in $\langle F1_{SS}, F2_{SS} \rangle$ space

The target of learning: what are tones?

$$\{\textbf{Data}\} \xrightarrow{\textbf{Learner}}$$
$$\{\textbf{Phonological maps}\}$$

A strategy for characterizing the learning problem
Characterizing tonal maps
Defining tonal maps
The learnability of tonal maps

The target of learning: what are tones?

$$\{\textbf{Phonetic data}\} \xrightarrow{\textbf{Learner}}$$
$$\{\textbf{Phonological maps}\}$$

Restriction for this project: **"pure speech"** situation—refer only to acoustic information (methodological abstraction)

The target of learning: what are tones?

$$\{\textbf{Phonetic data}\} \xrightarrow{\textbf{Learner}}$$
$$\{\textbf{Tonal maps}\}$$

Restriction for this project: **"pure speech"** situation—refer only to acoustic information (methodological abstraction)

Defining phonological maps

Phonological maps:
{sequences of phonetic parameter vectors} $\rightarrow$ {sets of phonological categories}

- ▶ **Generalization** from finite sample to infinite set in learning

Defining phonological maps

Phonological maps:
{sequences of phonetic parameter vectors} $\rightarrow$
{sets of phonological categories}

- **Generalization** from finite sample to infinite set in learning
- Connected regions contain **too many points to be enumerated**

Defining phonological maps

Phonological maps:
{sequences of phonetic parameter vectors} →
{sets of phonological categories}

- **Generalization** from finite sample to infinite set in learning
- Connected regions contain **too many points to be enumerated**
- **Ambiguity** ⇒ **probabilistic distribution** of phonological categories over phonetic spaces (Pierrehumbert 2003)

$\langle F1_{SS} = 686, F2_{SS} = 1028 \rangle \mapsto \{/ɑ, ɔ/\}$

Defining phonological maps

Phonological maps:
{sequences of phonetic parameter vectors} $\rightarrow$
$P_1 \times P_2 \times \cdots \times P_c$

- **Generalization** from finite sample to infinite set in learning
- Connected regions contain **too many points to be enumerated**
- **Ambiguity** $\Rightarrow$ **probabilistic distribution** of phonological categories over phonetic spaces (Pierrehumbert 2003)

$$\langle F1_{SS} = 686, F2_{SS} = 1028\rangle \mapsto \{p(/\mathrm{a}/) = 0.45,\ p(/\mathrm{ɔ}/) = 0.55\}$$

Characterizing phonological maps

Key questions:

Characterizing phonological maps

Key questions:

1. What kinds of **phonological categories** are to be represented in the range of the map? (Here: **phonemes**, by stipulation)

Characterizing phonological maps

Key questions:

1. What kinds of **phonological categories** are to be represented in the range of the map? (Here: **phonemes**, by stipulation)

2. What is the **phonetic parameter space**—the space of phonetic parameters—for the phonological categories?

Characterizing phonological maps

Key questions:

1. What kinds of **phonological categories** are to be represented in the range of the map? (Here: **phonemes**, by stipulation)

2. What is the **phonetic parameter space**—the space of phonetic parameters—for the phonological categories?

3. What are **properties of the distribution** of the phonological categories over the phonetic parameter space?

Methodological abstraction: which parameters?

Reality: **Probabilistic distribution** of phonological categories over phonetic spaces

Methodological abstraction: which parameters?

Reality: **Probabilistic distribution** of phonological categories over phonetic spaces

- ▶ **Model**: **partition** of set of phonological categories over phonetic spaces
  - ▶ Tonal identification (humans), hard classification algorithms (machines)

Methodological abstraction: which parameters?

Reality: **Probabilistic distribution** of phonological categories over phonetic spaces

- ▶ **Model**: **partition** of set of phonological categories over phonetic spaces
  - ▶ Tonal identification (humans), hard classification algorithms (machines)
- ▶ Example: A two tone tonal inventory, e.g. $\{H, L\}$

$p(x|\omega_i)$

ω_1 ω_2

0.4

0.3

0.2

0.1

x
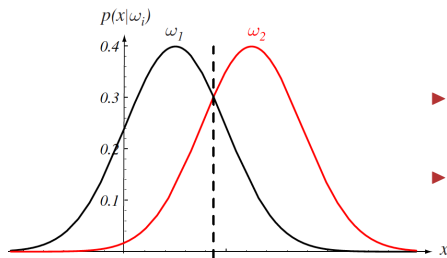
Duda, Hart and Stark (2001)

- ▶ Probability distribution $p(x|\omega)$ over x, x = mean fundamental frequency (f0)
- ▶ Two classes: $\omega_1 = L$, $\omega_2 = H$

Phonological maps are non recursively-enumerable

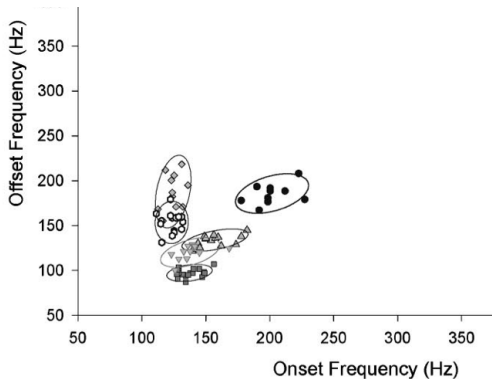## Phonological maps are defined over **real-valued parameters**
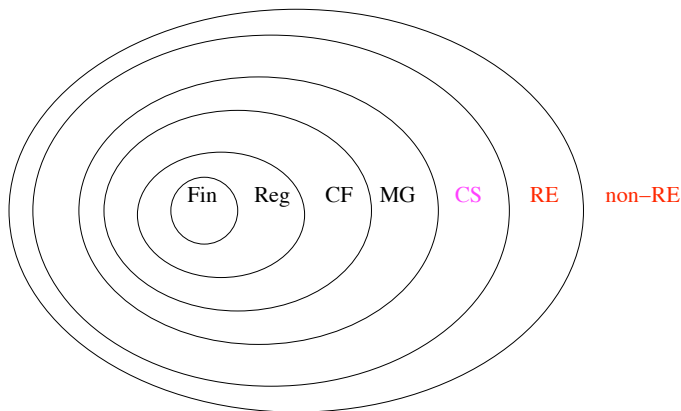
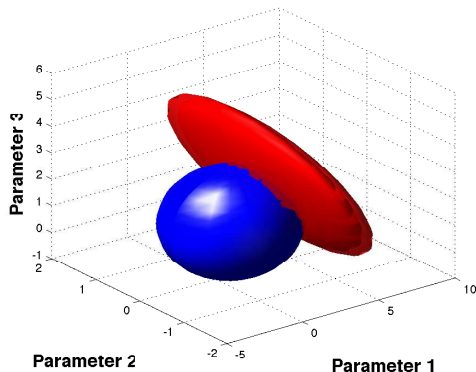Figure: The Chomsky hierarchy of formal languages

Can we characterize tonal maps as being feasibly learnable?

► In phonetic space: each **parameter** defines a **dimension** and can take a real value

Figure: Map in a **2-D** parameter space

Can we characterize tonal maps as being feasibly learnable?

Figure: Map in a **3-D** parameter space

▶ In phonetic space: each **parameter** defines a **dimension** and can take a real value

▶ Potentially an **infinite number of parameters**, each with a potentially infinite range of possible values

Structure permits feasible learning even in infinite spaces

But comfort from the finiteness of the space of possible grammars is tenuous indeed. *For a grammatical theory with an infinite number of possible grammars might be well structured, permitting informed search that converges quickly to the correct grammar—even though uninformed, exhaustive search is infeasible. And it is of little value that exhaustive search is guaranteed to terminate eventually when the space of possible grammars is finite, if the number of grammars is astronomical.* **In fact, a well-structured theory admitting an infinity of grammars could well be feasibly learnable, while a poorly structured theory admitting a finite, but very large, number of possible grammars might not**.
(Tesar and Smolensky 2000: 3)

A strategy for characterizing the learning problem
Characterizing tonal maps
Defining tonal maps
The learnability of tonal maps

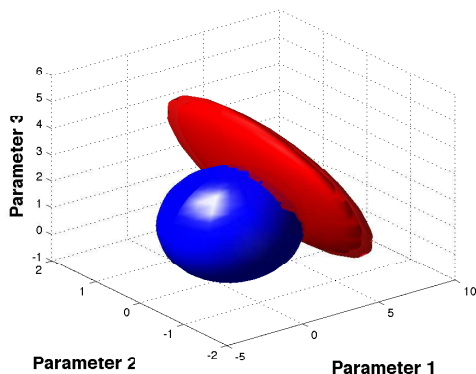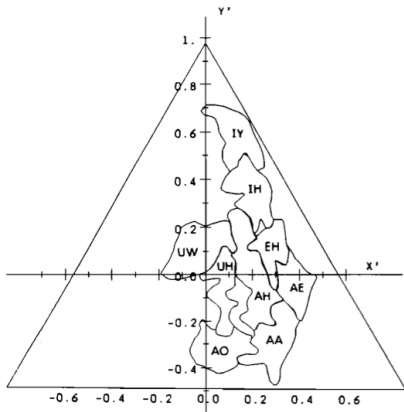# Can we characterize tonal maps as being feasibly learnable?

Figure: Map in a **3-D** parameter space

▶ In phonetic space: each **parameter** defines a **dimension** and can take a real value

▶ Potentially an **infinite number of parameters**, each with a potentially infinite range of possible values

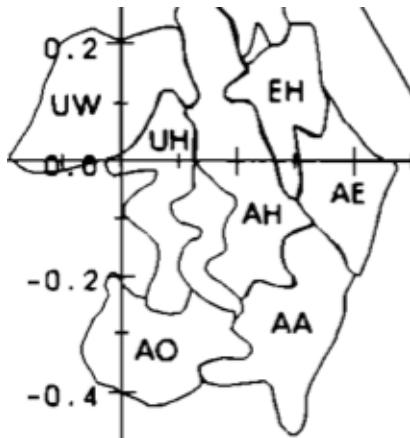# Can we characterize tonal maps as being feasibly learnable?

▶ In phonetic space: each **parameter** defines a **dimension** and can take a real value

▶ Potentially an **infinite number of parameters**, each with a potentially infinite range of possible values

▶ Complex **shapes/distributions** can make maps in even 2-D spaces not feasibly learnable

Figure: Scary map in a **2-D** parameter space (Miller 1989)

Can we characterize tonal maps as being feasibly learnable?

- ▶ In phonetic space: each **parameter** defines a **dimension** and can take a real value
- ▶ Potentially an **infinite number of parameters**, each with a potentially infinite range of possible values
- ▶ Complex **shapes/distributions** can make maps in even 2-D spaces not feasibly learnable

Figure: Scary map in a **2-D** parameter space (Miller 1989)

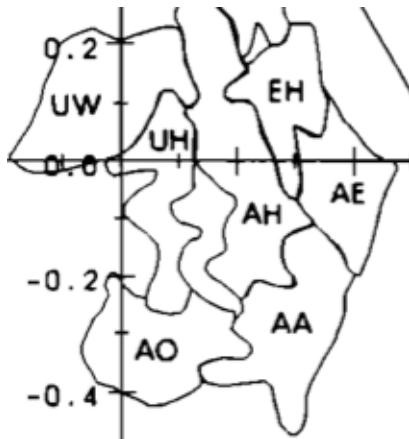# Can we characterize tonal maps as being feasibly learnable?

Figure: Scary map in a **2-D** parameter space
(Miller 1989)

► In phonetic space: each
  parameter defines a **dimension**
  and can take a real value

► Potentially an **infinite number
  of parameters**, each with a
  potentially infinite range of
  possible values

► Complex **shapes/distributions**
  can make maps in even 2-D
  spaces not feasibly learnable

⇒ **there must be restrictive
  structure in the hypothesis
  space**

Characterizing structure in the hypothesis space

1. Any characterization of structure is **conditioned on the parameter space** in which the tonal maps are defined

   ⇒ Need to do phonetic studies of **relevant phonetic parameters** for defining tonal maps

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

Characterizing structure in the hypothesis space

1. Any characterization of structure is **conditioned on the parameter space** in which the tonal maps are defined

   ⇒ Need to do phonetic studies of **relevant phonetic parameters** for defining tonal maps

2. Need a way to diagnose **feasible learnability** from characterized structure

   Mathematical complexity metric: **Vapnik-Chervonenkis (VC) dimension** (Vapnik 1998, Vapnik and Chervonenkis 1971)
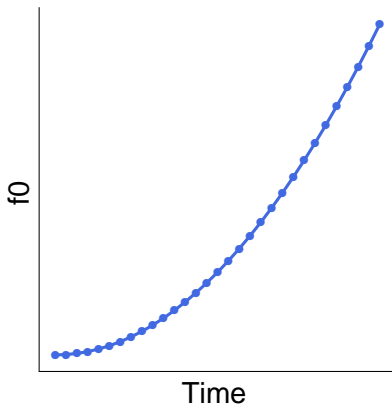
Cross-linguistic tonal language sample

Language	Area	Tonal inventory
Bole	Nigeria	˥, ˩ (H,L)
Mandarin	Beijing	˥, ˩˥, ˨˩˦, ˥˩
Cantonese	Hong Kong	˥, ˧˥, ˧, ˨˩, ˩˧, ˨
Hmong	Laos/Thailand	˥, ˧, ˨, ˥˩, ˨˩, ˩, ˧˩

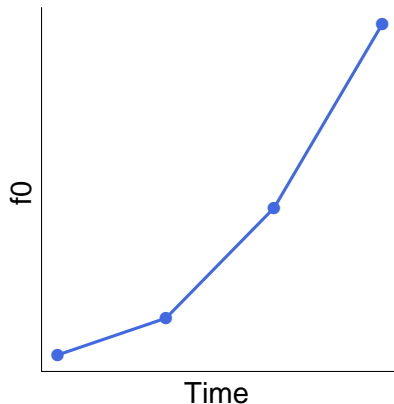- Languages chosen for diversity in **level/contour** distinctions and **voice quality** contrasts
- Multiple speakers (6M/6F for all but Bole (3M/2F))
- All legal **bitone** combinations recorded sentence-medially

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

Temporal resolution: how many samples? (I)

Dense sampling

Coarse sampling

Each sampled point could contribute to **complexity** in tonal map!

Temporal resolution: how many samples? (II)

Dense sampling

- ▶ Gauthier et al. (2007): **30 samples/syllable (1 sample/6 ms)**
- ▶ Automatic speech recognition: **1 sample/10 ms**

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

Temporal resolution: how many samples? (II)

Dense sampling

- ▶ Gauthier et al. (2007): **30 samples/syllable (1 sample/6 ms)**
- ▶ Automatic speech recognition: **1 sample/10 ms**

Coarse sampling

- ▶ Linguistics: Chao (1933, 1968), International Phonetic Alphabet
  ꜒,꜓,꜔,꜕, **3 samples/tone**

Temporal resolution: how many samples? (II)

Dense sampling

- ▶ Gauthier et al. (2007): **30 samples/syllable (1 sample/6 ms)**
- ▶ Automatic speech recognition: **1 sample/10 ms**

Coarse sampling

- ▶ Linguistics: Chao (1933, 1968), International Phonetic Alphabet
  ˥,˦,˧,˨, **3 samples/tone**
- ▶ Automatic speech recognition
  - ▸ **3 - 5 samples/tone**: Qian et al. (2007): Cantonese; Wang and
    Levow (2008), Zhou et al. (2008): Mandarin
  - ▸ Tian et al. (2004): Higher tonal ID accuracy with **4
    samples/tone** than 1 sample/10 ms (Mandarin)

Temporal resolution: how many samples? (II)

Dense sampling

▶ Gauthier et al. (2007): **30 samples/syllable (1 sample/6 ms)**

▶ Automatic speech recognition: **1 sample/10 ms**

Coarse sampling

▶ Linguistics: Chao (1933, 1968), International Phonetic Alphabet
˥,˦,˧,˨, **3 samples/tone**

▶ Automatic speech recognition

  ▸ **3 - 5 samples/tone**: Qian et al. (2007): Cantonese; Wang and
    Levow (2008), Zhou et al. (2008): Mandarin
  ▸ Tian et al. (2004): Higher tonal ID accuracy with **4
    samples/tone** than 1 sample/10 ms (Mandarin)

**Hypothesis: Good tonal category separability can be maintained
under coarse temporal sampling of phonetic parameters.**

Human perception experiments: stimuli

▶ Cantonese tritones: nonce 3-syllable phrases built from syllables in the lexicon

▶ First and third syllables held fixed:
$< wai\dashv, \{wai\urcorner, wai\lrcorner, wai\dashv, wai\lrcorner, wai\lrcorner, wai\dashv\}, mat\dashv > \triangleright$

Tritone	Gloss
$< wai\dashv, wai\urcorner, mat\dashv >$	fear power clean
$< wai\dashv, wai\lrcorner, mat\dashv >$	fear appoint clean
$< wai\dashv, wai\dashv, mat\dashv >$	fear fear clean
$< wai\dashv, wai\lrcorner, mat\dashv >$	fear surround clean
$< wai\dashv, wai\lrcorner, mat\dashv >$	fear great clean
$< wai\dashv, wai\dashv, mat\dashv >$	fear stomach clean

Human perception experiments: stimuli

▶ Cantonese tritones: nonce 3-syllable phrases built from syllables in the lexicon

▶ First and third syllables held fixed:
$< wai\dashv, \{wai\urcorner, \wedge, \urcorner, \lrcorner, \lrcorner, \dashv\}, mat\dashv > \triangleright$

Tritone	Gloss
$< wai\dashv, wai\urcorner, mat\dashv >$	fear power clean
$< wai\dashv, wai\wedge, mat\dashv >$	fear appoint clean
$< wai\dashv, wai\urcorner, mat\dashv >$	fear fear clean
$< wai\dashv, wai\lrcorner, mat\dashv >$	fear surround clean
$< wai\dashv, wai\wedge, mat\dashv >$	fear great clean
$< wai\dashv, wai\dashv, mat\dashv >$	fear stomach clean

▶ Syllables identified with orthographic characters

▶ Some characters may be more frequent than others:
$\wedge > \lrcorner > \urcorner >> \wedge > \urcorner, \dashv$ (based on corpus count of Mandarin cognates, Da (2004)

Human perception experiment

- **Stimuli**: Cantonese tritones,
  $< wai\dashv, \{wai\urcorner, \diagup, \dashv, \diagdown, \diagup, \dashv\}, mat\dashv >$ from 5 speakers (3M, 2F)

Human perception experiment

▶ **Stimuli**: Cantonese tritones,
$< wai\dashv, \{wai\daleth, \mathcal{A}, \dashv, \downarrow, \mathcal{A}, \dashv\}, mat\dashv >$ from 5 speakers (3M, 2F)

▶ **Methodological inspiration**: Multiple phoneme restoration in interrupted speech (Warren 1970)

▶ **Manipulated variable**: SAMPLING RESOLUTION
(2, 3, 5, 7 samples/syllable, intact)

Human perception experiment

▶ **Stimuli**: Cantonese tritones,
$< wai\dashv, \{wai\urcorner, \lrcorner, \dashv, \lrcorner, \lrcorner, \dashv\}, mat\dashv >$ from 5 speakers (3M, 2F)

▶ **Methodological inspiration**: Multiple phoneme restoration in interrupted speech (Warren 1970)

▶ **Manipulated variable**: SAMPLING RESOLUTION
(2, 3, 5, 7 samples/syllable, intact)

▶ **Task**: 6-alternative forced choice orthographic identification of second tone in tritone

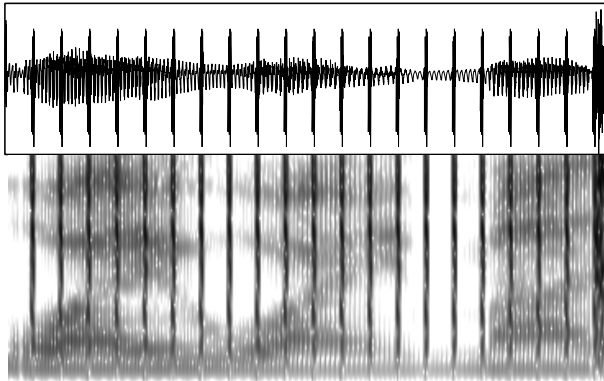▶ **Participants**: 39 native Cantonese speakers, tested in Hong Kong and Los Angeles

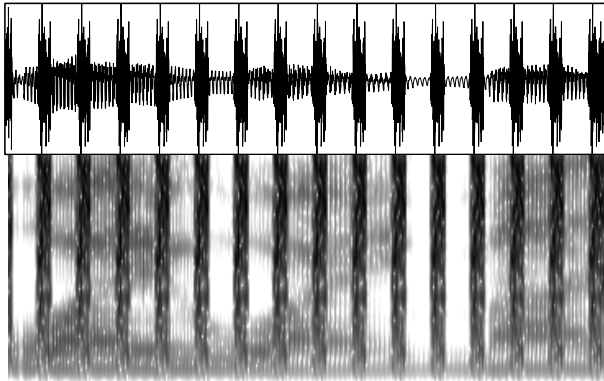# Stimuli example: waveform/spectrogram

▷ [**Intact** tritone]
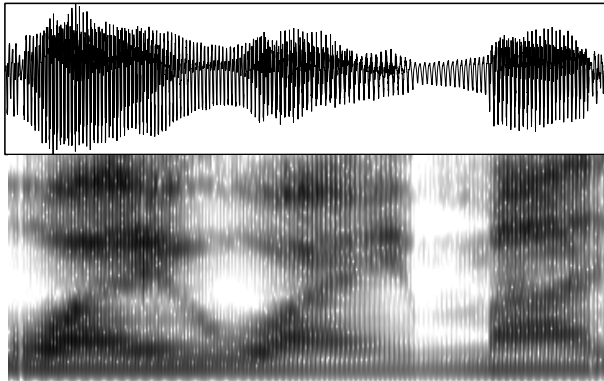
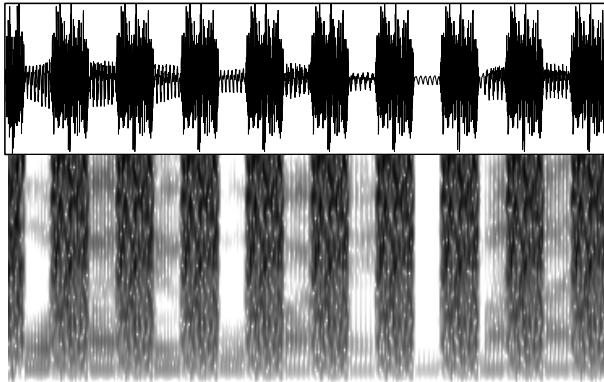# Stimuli example: waveform/spectrogram

▷ [**7 samples** per syllable]

Stimuli example: waveform/spectrogram
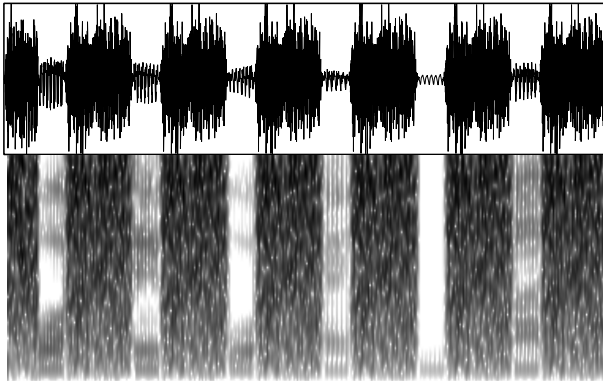
▷ [**5 samples** per syllable]

Stimuli example: waveform/spectrogram

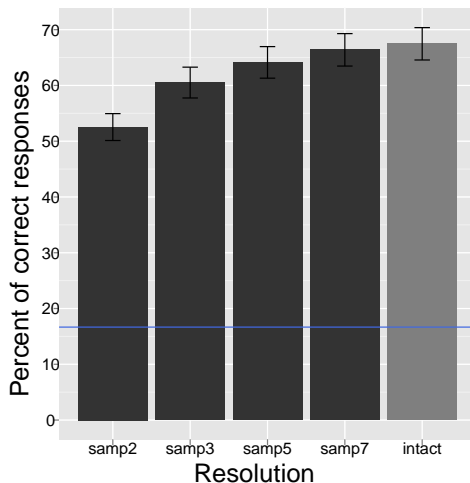▷ [**3 samples** per syllable]

Stimuli example: waveform/spectrogram

▷ [**2 samples** per syllable]

Tonal ID accuracy maintained with coarse resolution

**Tonal ID accuracy well above chance
even down to 2 samples/syllable!**

Resolution	Percent correct
samp2	52.54 (2.41)
samp3	60.51 (2.76)
samp5	64.13 (2.83)
samp7	66.38 (2.91)
intact	67.46 (2.90)

Computational modeling for insight into experiment

- ▶ What were listeners listening to?
- ▶ Effects of particular task/stimuli?

Computational modeling for insight into experiment

- ▶ What were listeners listening to?
- ▶ Effects of particular task/stimuli?

**Computational modeling allows explicit and tradeable assumptions.**

Computational modeling for insight into experiment

- ▶ What were listeners listening to?
- ▶ Effects of particular task/stimuli?

Computational modeling allows explicit and tradeable assumptions.

- ▶ Assume: mean f0 values extracted from each sample, for 2-7 samples per syllable
    - ▶ Extracted using implementation of RAPT pitch tracker (Talkin 1995)
- ▶ Assume: no lexical bias
    - ▶ Uniform prior (all tonal categories equally likely)
- ▶ Ask: How accurate is tonal identification by machine?

Computational modeling: parameterization of data
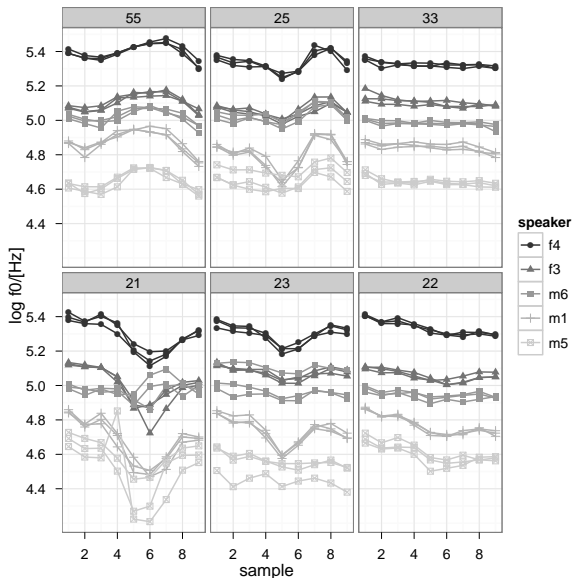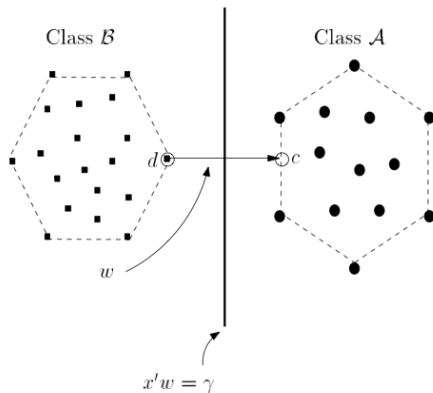
Computational modeling: parameterization of data
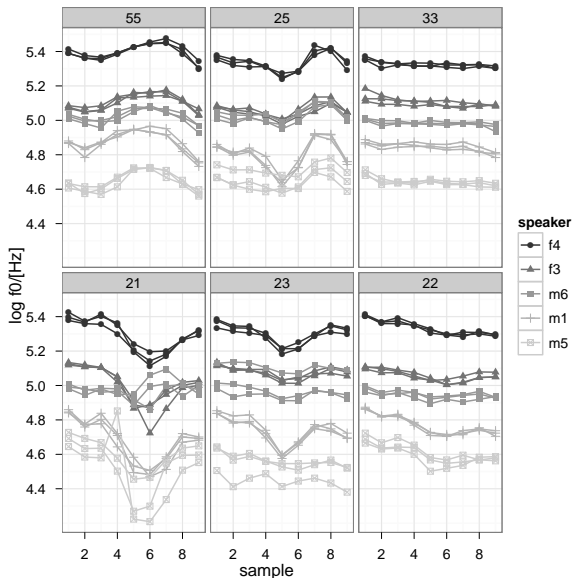
▶ Standardized data: per-speaker z-scores for log transformed f0 values (Levow 2006)

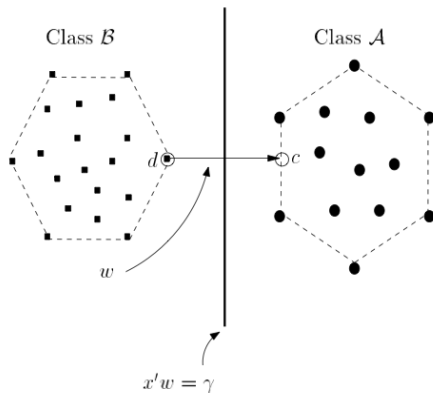# Computational modeling: support vector machines

Class $\mathcal{B}$

Class $\mathcal{A}$

d

c

w

$x'w = \gamma$

Bennett and Bredensteiner (2000), Vapnik (1995)

1. Given labeled **training data**, e.g. $<< 200, 210, 224 >, \daleth >$

Computational modeling: support vector machines

Class $\mathcal{B}$

Class $\mathcal{A}$

d

c

w

$x'w = \gamma$

Bennett and Bredensteiner (2000), Vapnik (1995)

1. Given labeled **training data**, e.g. $<< 200, 210, 224 >, \rceil >$

2. Draw convex hull around data from a given category

3. Find **separating hyperplane** maximizing margin between convex hulls

Computational modeling: support vector machines

Class $\mathcal{B}$

Class $\mathcal{A}$

d

c

w

$x'w = \gamma$

Bennett and Bredensteiner (2000), Vapnik (1995)

1. Given labeled **training data**, e.g. $<< 200, 210, 224 >, \daleth >$

2. Draw convex hull around data from a given category

3. Find **separating hyperplane** maximizing margin between convex hulls

4. Use separating hyperplane to classify **test data** (unseen data): train on 4 speakers, test on 5th, average results

Support vector machine classification results

▶ SVM classification accuracy ≈75% for all conditions

Support vector machine classification results

- ▶ SVM classification accuracy ≈75% for all conditions
- ▶ Accuracy with as few as 6 real values not statistically different from accuracy with 69

Support vector machine classification results

▶ SVM classification accuracy ≈75% for all conditions

▶ Accuracy with as few as 6 real values not statistically different from accuracy with 69

Sufficiency of coarse temporal resolution in humans and machines hints at structure in the class of tonal maps

Linear discriminant analysis for dimensionality reduction
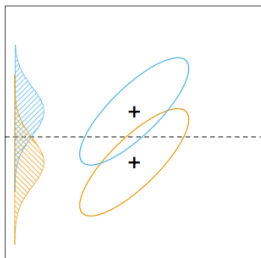
Don't project there!   **Project here!**
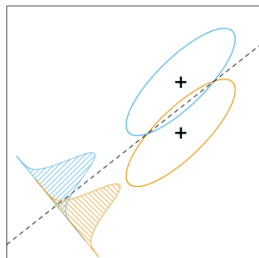
(Hastie, Tibshirani, and Friedman 2009)

- ▶ Project onto axis to maximize ratio of **between-class** to **within-class scatter**
- ▶ **Between-class scatter**: roughly, distance between class means
- ▶ **Within-class scatter**: class variances

Cross-linguistic computational modeling for sampling resolution example: Bole, log f0 values
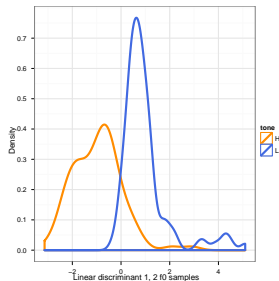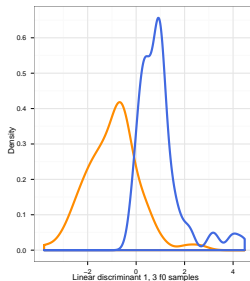
2 log f0 values   3 log f0 values   10 log f0 values
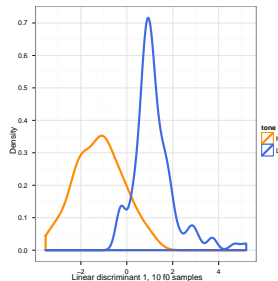
Little difference in overlap between H/L from 2 to 10 f0 samples

Structure in the class of tonal maps

What do tonal maps in the studied languages indicate about potential structure in the class of tonal maps in natural language?

Structure in the class of tonal maps

What do tonal maps in the studied languages indicate about potential structure in the class of tonal maps in natural language?

Tonal concepts in low-dimensional spaces for single speakers for languages studied are near-linearly separable

Mandarin single speaker space: log f0, 3 values

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space
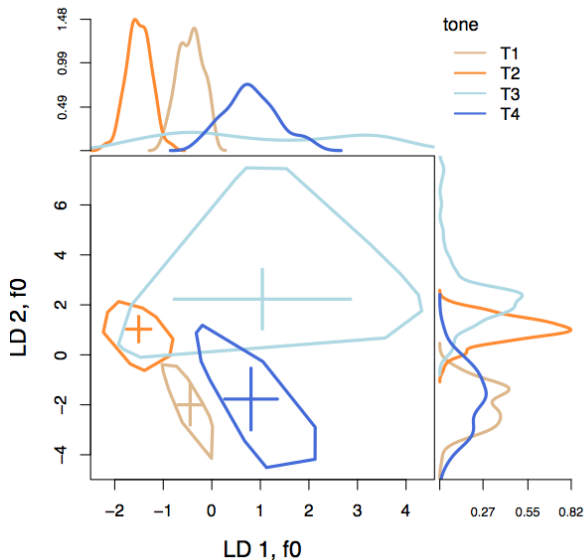
Cantonese single speaker space: log f0, Δf0, 2 values each

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space
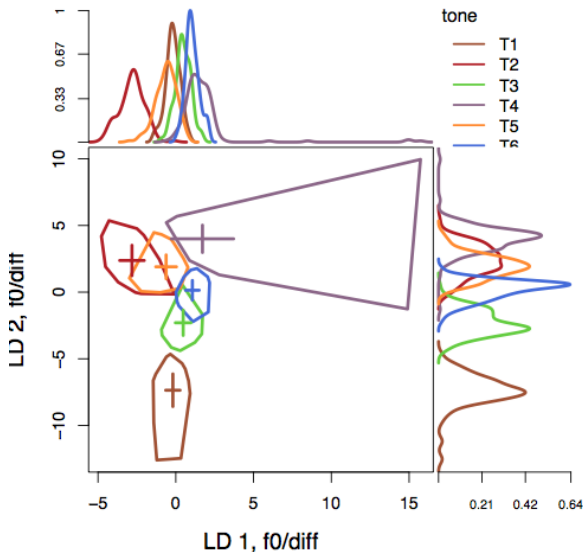
White Hmong single speaker space: log f0, 10 values

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space
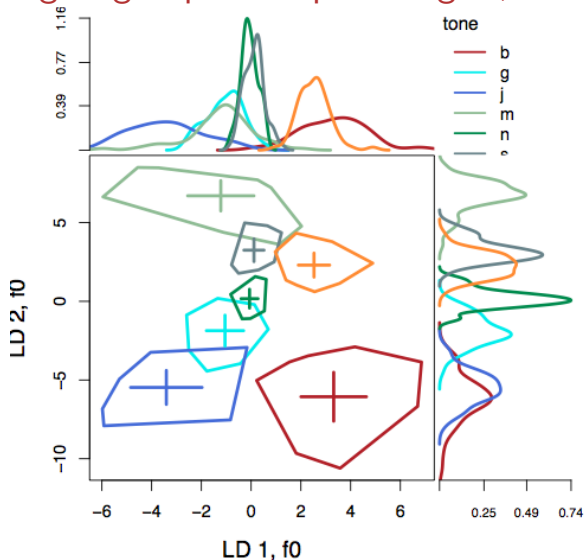
VC dimension: definition by example — **rays in** $\mathbb{R}$
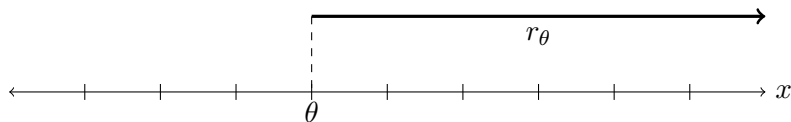
$$r_\theta = \{x \in \mathbb{R} | \theta \leq x\}$$

$$r_\theta = \begin{cases} 1 & \text{if } \theta \leq x \\ 0 & \text{otherwise} \end{cases}$$

$$r_\infty = \{\} \quad \forall x \in \mathbb{R} \quad (\text{empty ray})$$

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

if $\{S \cap T | T \in \mathcal{T}\} = \wp(S)$, then S is **shattered** by $\mathcal{T}$

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

if $\{S \cap T | T \in \mathcal{T}\} = \wp(S)$, then S is **shattered** by $\mathcal{T}$

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

$$\text{if } \{S \cap T | T \in \mathcal{T}\} = \wp(S), \text{ then } S \text{ is \textbf{shattered} by } \mathcal{T}$$

S	$|S|$	$\wp(S)$	T for $T \cap S$	Shattered?
$\{\}$	0	$\{\}$	r_∞	Yes

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

$$\text{if } \{S \cap T | T \in \mathcal{T}\} = \wp(S), \text{ then } S \text{ is } \textbf{shattered} \text{ by } \mathcal{T}$$

S	$|S|$	$\wp(S)$	T for $T \cap S$	Shattered?
$\{\}$	0	$\{\}$	r_∞	Yes
$\{1\}$	1	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	Yes

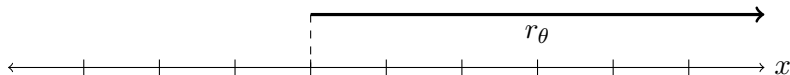# VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

if $\{S \cap T | T \in \mathcal{T}\} = \wp(S)$, then S is **shattered** by $\mathcal{T}$

S	$|S|$	$\wp(S)$	T for $T \cap S$	Shattered?
$\{\}$	0	$\{\}$	r_∞	Yes
$\{1\}$	1	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	Yes

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

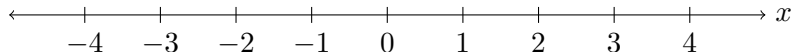if $\{S \cap T | T \in \mathcal{T}\} = \wp(S)$, then S is **shattered** by $\mathcal{T}$

S	$|S|$	$\wp(S)$	T for $T \cap S$	Shattered?
$\{\}$	0	$\{\}$	r_∞	Yes
$\{1\}$	1	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	Yes
$\{0, 1\}$	2	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

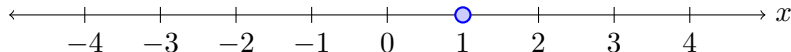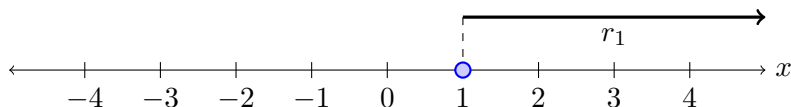if $\{S \cap T | T \in \mathcal{T}\} = \wp(S)$, then S is **shattered** by $\mathcal{T}$

S	$|S|$	$\wp(S)$	T for $T \cap S$	Shattered?
$\{\}$	0	$\{\}$	r_∞	Yes
$\{1\}$	1	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	Yes
$\{0, 1\}$	2	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	
		$\{0, 1\}$	$r_{\theta \leq 0}$	

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

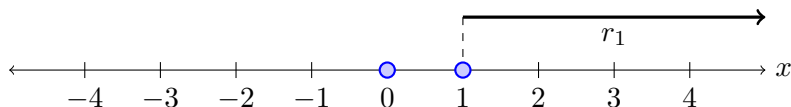if $\{S \cap T | T \in \mathcal{T}\} = \wp(S)$, then S is **shattered** by $\mathcal{T}$

S	$|S|$	$\wp(S)$	T for $T \cap S$	Shattered?
$\{\}$	0	$\{\}$	r_∞	Yes
$\{1\}$	1	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	Yes
$\{0, 1\}$	2	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	
		$\{0, 1\}$	$r_{\theta \leq 0}$	
		$\{0\}$	??	**No!**

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

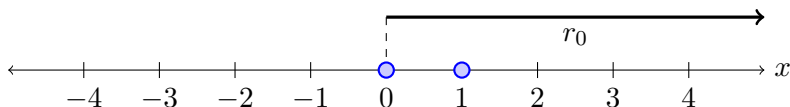if $\{S \cap T | T \in \mathcal{T}\} = \wp(S)$, then S is **shattered** by $\mathcal{T}$

S	$|S|$	$\wp(S)$	T for $T \cap S$	Shattered?
$\{\}$	0	$\{\}$	r_∞	Yes
$\{1\}$	1	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	Yes
$\{0,1\}$	2	$\{\}, \{1\}$	$r_\infty, r_{\theta \leq 1}$	
		$\{0,1\}$	$r_{\theta \leq 0}$	
		$\{0\}$	??	**No!**

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

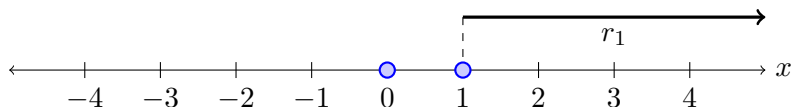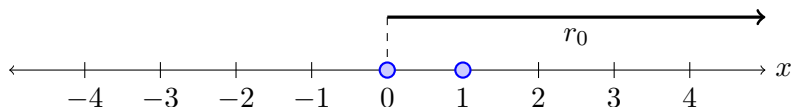if $\{S \cap T | T \in \mathcal{T}\} = \wp(S)$, then S is **shattered** by $\mathcal{T}$

$$VC(\mathcal{T}) = max\{|S| : S \text{ is shattered by } \mathcal{T}\} = 1$$

VC dimension: definition by example — **rays in** $\mathbb{R}$

Given sample $S \subseteq \mathbb{R}$, class of tonal maps $\mathcal{T}$

$$\boxed{\text{if } \{S \cap T | T \in \mathcal{T}\} = \wp(S), \text{ then } S \text{ is } \textbf{shattered} \text{ by } \mathcal{T}}$$

What if $\mathcal{T}$ consisted of the union of a finite number of intervals on $\mathbb{R}$?
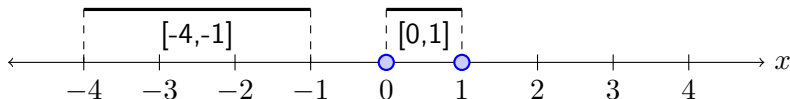
$$\boxed{VC(\mathcal{T}) = max\{|S| : S \text{ is shattered by } \mathcal{T}\} \text{ \textbf{infinite}}}$$

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

VC dimension and feasible learnability

Finite VC dimension is a criterion for feasible learnability

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

VC dimension and feasible learnability

Finite VC dimension is a criterion for feasible learnability

- VC dim of **ellipsoids** in $\mathbb{R}^d : (d^2 + 3d)/2$ (Akama et al. 2011)

VC dimension and feasible learnability

Finite VC dimension is a criterion for feasible learnability

- VC dim of **ellipsoids** in $\mathbb{R}^d$: $(d^2 + 3d)/2$ (Akama et al. 2011)
- VC dim of **arbitrary convex polygons** in $\mathbb{R}^d$ $\forall d$ is infinite (Blumer et al. 1989)

VC dimension and feasible learnability

Finite VC dimension is a criterion for feasible learnability

- ▶ VC dim of **ellipsoids** in $\mathbb{R}^d$: $(d^2 + 3d)/2$ (Akama et al. 2011)
- ▶ VC dim of **arbitrary convex polygons** in $\mathbb{R}^d$ $\forall d$ is infinite (Blumer et al. 1989)
- ▶ VC dimension is applicable to real and discrete spaces

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

VC dimension and feasible learnability

Finite VC dimension is a criterion for feasible learnability

- ► VC dim of **ellipsoids** in $\mathbb{R}^d$: $(d^2 + 3d)/2$ (Akama et al. 2011)
- ► VC dim of **arbitrary convex polygons** in $\mathbb{R}^d$ $\forall d$ is infinite (Blumer et al. 1989)
- ► VC dimension is applicable to real and discrete spaces
- ► VC dimension of constraint ranking/weighting hypothesis spaces for OT and HG is finite (Riggle 2009, Bane et al. 2010)

The VC dimension of linear half spaces is finite

Figure 3.2 A set of three points that is shatterable by half-spaces in $\mathcal{R}_c^2$

Figure: VC dimension of linear half spaces in $\mathbb{R}^2$ (Heinz and Riggle 2011), relevant for VC dim of harmonic grammar (Pater 2008, Potts et al. 2010)

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

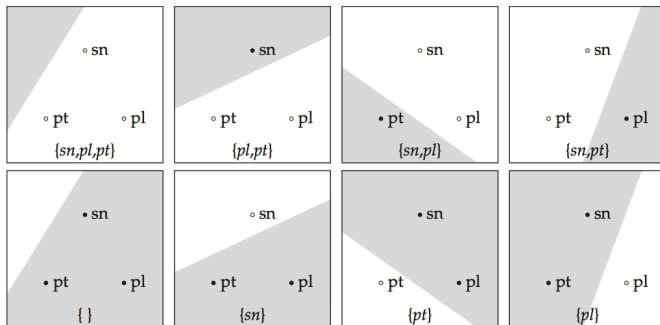# The VC dimension of linear half spaces is finite
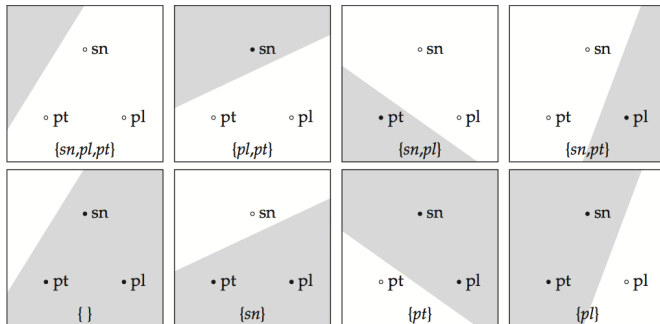
Figure 3.2 A set of three points that is shatterable by half-spaces in $\mathscr{R}^2_c$

Figure: VC dimension of linear half spaces in $\mathbb{R}^2$ (Heinz and Riggle 2011), relevant for VC dim of harmonic grammar (Pater 2008, Potts et al. 2010)

The hypothesis space of any linear learning algorithm is feasibly learnable

Conclusions

Some points:

▶ There is **structure** in the potentially high-dimensional definition of **phonological maps**

Conclusions

Some points:

- ▶ There is **structure** in the potentially high-dimensional definition of **phonological maps**
- ▶ To study phonological category learning, we need to understand how the **hypothesis space** is structured

Conclusions

Some points:

- ▶ There is **structure** in the potentially high-dimensional definition of **phonological maps**
- ▶ To study phonological category learning, we need to understand how the **hypothesis space** is structured
- ▶ To characterize structure in the hypothesis space, we need to understand what **phonetic parameters** are involved

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

Conclusions

Is the class of tonal maps in natural language feasibly learnable?

Conclusions

Is the class of tonal maps in natural language feasibly learnable?

▶ Sufficiency of coarse temporal resolution consistent with structure in tonal maps

Conclusions

Is the class of tonal maps in natural language feasibly learnable?

- ▶ Sufficiency of coarse temporal resolution consistent with structure in tonal maps
- ▶ Studied tonal maps appear to have nearly linearly separable concepts in small parameter spaces

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

Conclusions

Is the class of tonal maps in natural language feasibly learnable?

- ▶ Sufficiency of coarse temporal resolution consistent with structure in tonal maps

- ▶ Studied tonal maps appear to have nearly linearly separable concepts in small parameter spaces

- ▶ Hypothesis spaces with finite VC dimension are feasibly learnable

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

Conclusions

Is the class of tonal maps in natural language feasibly learnable?

- ▶ Sufficiency of coarse temporal resolution consistent with structure in tonal maps
- ▶ Studied tonal maps appear to have nearly linearly separable concepts in small parameter spaces
- ▶ Hypothesis spaces with finite VC dimension are feasibly learnable
- ▶ **We can study the learnability of classes of grammars and phonological maps in a unified way**

A strategy for characterizing the learning problem
Characterizing tonal maps
Temporal resolution and parameter spaces
Learnability and structure in the hypothesis space

Acknowledgments

- ▶ For help with recordings, linguistic consultation:
    - ▶ Alhaji Maina Gimba and Russell Schuh (Bole)
    - ▶ Jianjing Kuang (Beijing Mandarin)
    - ▶ Cindy Chan, Vincie Ho, Hiu Wai Lam, Shing Yin Li, Cedric Loke (Cantonese)
    - ▶ Chou Khang and Phong Yang, CSU Fresno Department of Linguistics (Hmong)
- ▶ For help with perception experiments, data processing:
    - ▶ Hiu Wai Lam, Prairie Lam; Cindy Chan, Samantha Chan, Chris Fung, Shing Yin Li, Cedric Loke, Antonio Sou, Grace Tsai, Joanna Wang
- ▶ For invaluable discussion: Edward Stabler and Megha Sundara; Abeer Alwan, Robert Daland, Bruce Hayes, Sun-Ah Jun, Patricia Keating, John Kingston, Jody Kreiman, Mark Liberman, Russell Schuh, Colin Wilson, and Kie Zuraw; U. Maryland PFNA group
- ▶ This work was supported by a NSF graduate fellowship, NSF grant BCS-0720304, and a UCLA Linguistics Department Ladefoged scholarship and Summer Graduate Research Fellowship