

# 話者を単位とした世界諸英語の クラスタリングと地図化

峯松信明, 塩澤史野, 齋藤大輔 (東大)

mailto: [mine@gavo.t.u-tokyo.ac.jp](mailto:mine@gavo.t.u-tokyo.ac.jp)

## 多様な発音・世界諸英語

### まずは幾つか例を

一人で色々な英語を喋ってみる (YouTube より)



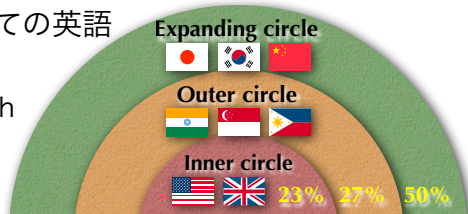
### Kachruの 3 circles model [Kachru'92]

日本語／中国語と異なり, 英語には共通発音 (普通語) がない

母国語／公用語／外国語としての英語

23%/27%/50%

AE も BE も accented English



## 本発表の流れ

### はじめに

受け入れなければならない発音の多様性 (世界諸英語)

発音多様性の最小単位は?

### 世界諸英語コーパスを使ったクラスタリング

クラスタリングに必要なもの

Speech Accent Archive

### 話者を単位とした世界諸英語発音クラスタリング

年齢・性別に頑健な発音分析法

IPA 距離の自動予測実験

一つの応用可能性

### まとめと今後の課題

## 外国語／方言訛りの最小単位は?



[Kachru 1992]

# 本発表の流れ

## はじめに

- 受け入れなければならない発音の多様性 (世界諸英語)
- 発音多様性の最小単位は?

## 世界諸英語コーパスを使ったクラスタリング

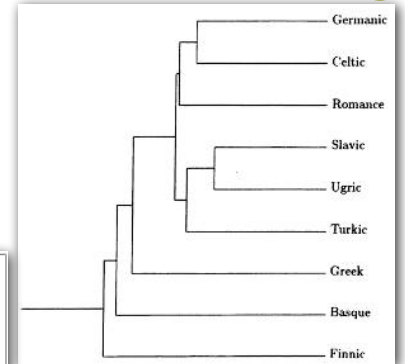
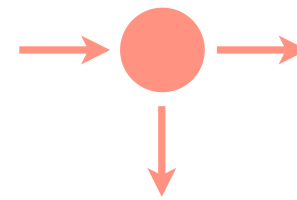
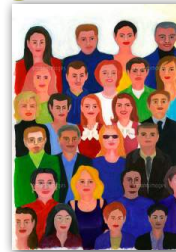
- クラスタリングに必要なもの
- Speech Accent Archive

## 話者を単位とした世界諸英語発音クラスタリング

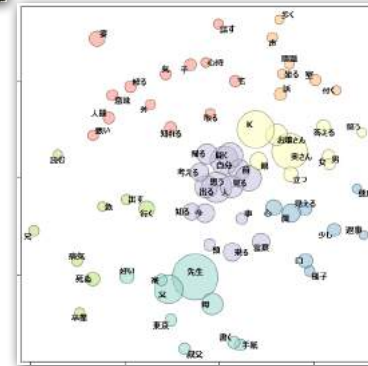
- 年齢・性別に頑健な発音分析法
- IPA 距離の自動予測実験
- 一つの応用可能性

## まとめと今後の課題

# クラスタリングをするために必要なもの



Dendrogram



2-dim scatter plot (Multi-dimensional scaling)

# N人の話者の発音クラスタリング

N speakers



	1	2	...	N
1	$d_{11}$	$d_{12}$	...	$d_{1N}$
2	$d_{21}$	$d_{22}$	...	$d_{2N}$
3	$d_{31}$			
:	:			
N	$d_{N1}$	$d_{N2}$	...	$d_{NN}$

$$\left| \text{Lower Triangle} \right| = (500 \times 500 - 500) / 2 = 124,750$$

How to calculate pron. distance automatically?

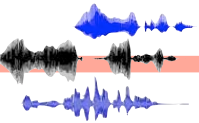
# Speech Accent Archive

## 特定パラグラフを読ませた世界諸英語音声コーパス

- 米語の音素 (対) の coverage を考えたパラグラフ設計
- 発話者参加型のコーパス構築 (現在約1,800名)
- IPA narrow transcription の提供 (約150種類のIPA記号)
- <http://accent.gmu.edu>

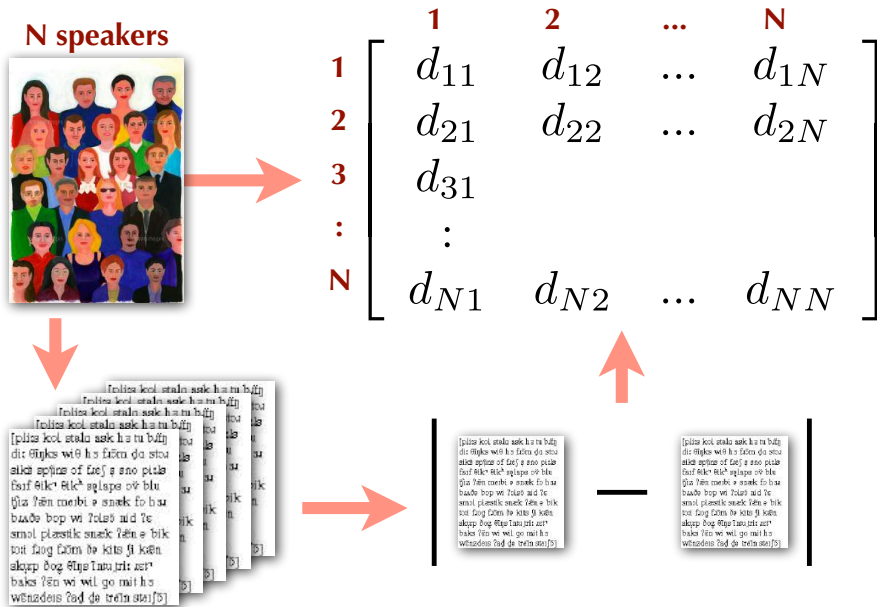


Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

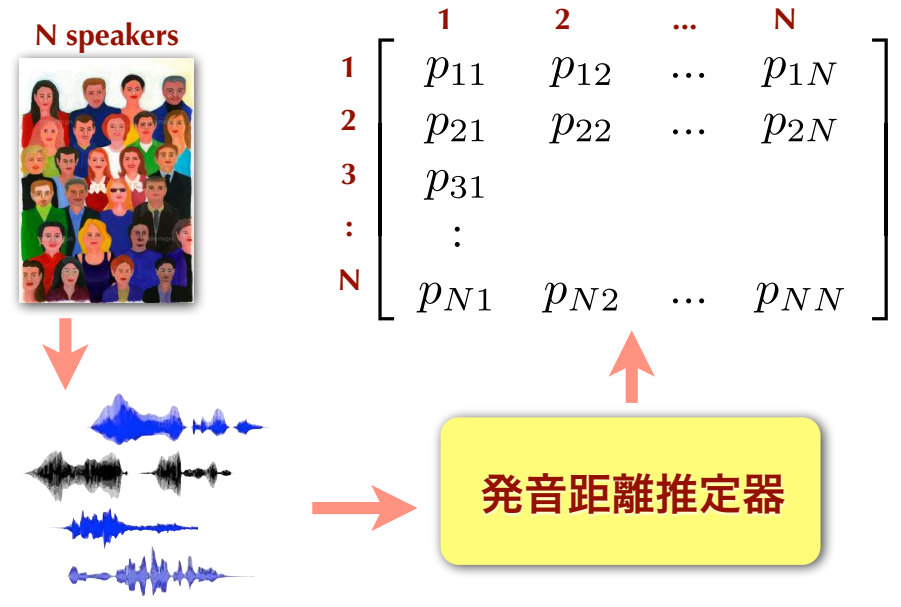


[pɪˈsɪs kəʊt stɔːl əsɪk hɪr frɒm ðə stɔːr]  
[pɪˈsɪs kəʊt stɔːl əsɪk hɪr frɒm ðə stɔːr]  
[pɪˈsɪs kəʊt stɔːl əsɪk hɪr frɒm ðə stɔːr]  
[sɪks spʊːnz əv ˈfrɛʃ ˈsnoʊ piːz] ə fɪv θɪk slæb ɒv bluː tʃiːz  
[ənd məbeɪ ə snæk fɔː hɜː bɒðə bɒb]  
[wiː əlsoʊ niːd ə smɔːl plæstɪk sneɪk ənd ə bɪg tɔɪ frɒg fɔː ðə kɪdz]  
[ʃiː kæn skɒp ðɪz θɪŋz ɪntə θriː rɛd bægz]  
[ænd wiː wɪl go miːt hɜː wɛdnədeɪ fɔːd ɔː ðə treɪn stɔːʃɪn]

# 話者を単位とした世界英語発音分類



# 話者を単位とした世界英語発音分類



## 二話者の発音距離の推定問題

何が難しいのか？ ～発音距離と音響距離～



“Those answers will be straightforward if you think them through carefully first.”

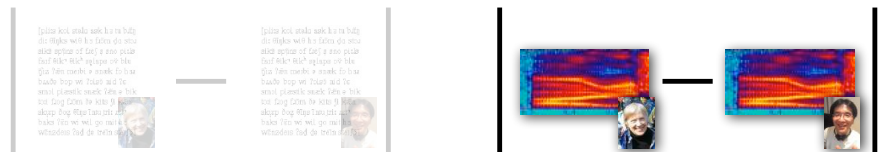


## 二話者の発音距離の推定問題

何が難しいのか？ ～発音距離と音響距離～



“Those answers will be straightforward if you think them through carefully first.”



# 本発表の流れ

## はじめに

- 受け入れなければならない発音の多様性 (世界諸英語)
- 発音多様性の最小単位は？

## 世界諸英語コーパスを使ったクラスタリング

- クラスタリングに必要なもの
- Speech Accent Archive

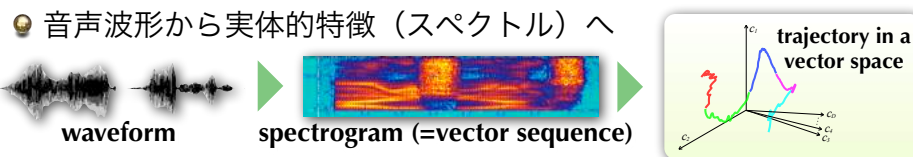
## 話者を単位とした世界諸英語発音クラスタリング

- 年齢・性別に頑健な発音分析法
- IPA 距離の自動予測実験
- 一つの応用可能性

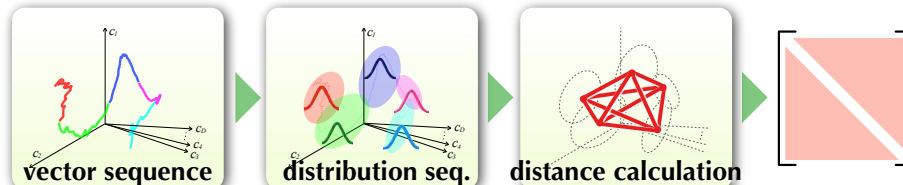
## まとめと今後の課題

# 発音構造解析 [峯松'06]

## 一つの発声から一つの構造 (距離行列) を計測する



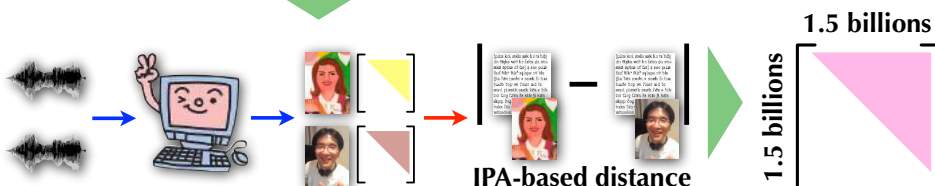
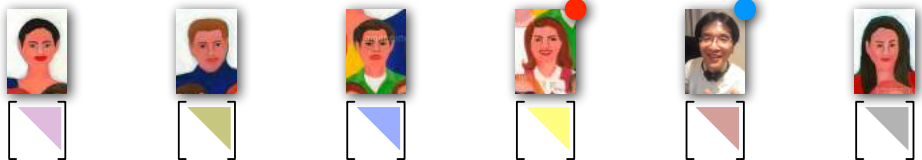
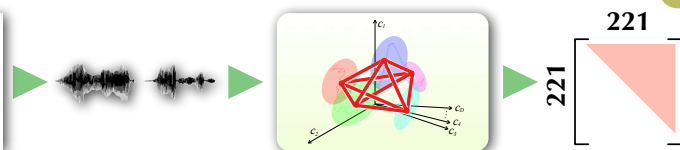
## 実体的特徴から関係的・構造的特徴へ



- “Please call Stella. Ask her ...” → 221 x 221 距離行列
- 特徴ベクトル → DNN → 事後確率ベクトル
  - DNN: 着目する音に対して推定される各音素らしさ
  - 構造解析: 着目する音と発声中の各音との距離

# 発音構造解析に基づく発音距離推定

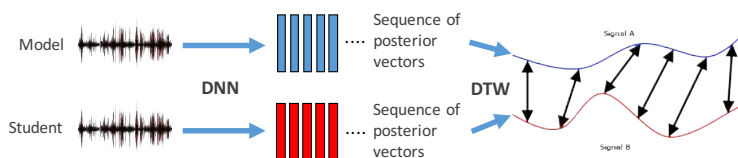
Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack .....



# IPA距離はどうやって求める？

## DTW (Dynamic Time Warping)

- 文字列と文字列 (二つの IPA 書き起こし) 間の DTW
- IPAシンボル種類数 = 153
  - 任意のシンボルと任意のシンボルの間の差異
  - 音声学者1人から集めた 153 シンボルの声から計測



# 予測実験の各種条件

構造解析結果を使って予測に使う素性(差業列)を計算

$$\begin{bmatrix} \text{yellow triangle} \\ \{S_{nm}\} \end{bmatrix} - \begin{bmatrix} \text{red triangle} \\ \{T_{nm}\} \end{bmatrix} \rightarrow \begin{bmatrix} \text{blue triangle} \\ \{D_{nm}\} \end{bmatrix} \quad D_{nm} = |S_{nm} - T_{nm}|$$

#elements = 221 x 220 / 2 = 24K

## 予測 (回帰) 手法

サポートベクター回帰



## 予測実験の3つのモード

- Speaker-open mode
- Speaker-pair-open mode
- Speaker-open and speaker-pair-open mode

# 実験結果

IPA距離と自動予測距離との相関[Sato+'15]

mode	spk-open	spk-pair-open	both
corr.	<b>0.50</b>	<b>0.87</b>	<b>0.77</b>

## この相関は高いのか低いのか?

- IPA書き起こしを米語音素書き起こしに半自動で変換
- IPA距離と「音素書き起こし間距離」との相関
  - 音素書き起こし間距離 = 書き起こし間のDTW距離
  - IPAシンボル数 = 153, 音素数 = 約40
  - 音素書き起こしは, 一般人ができる書き起こし
  - Corr. = **0.75**

training {T <sub>i</sub> }	testing {X <sub>i</sub> }
T <sub>1</sub> - T <sub>2</sub>	X <sub>1</sub> - T <sub>1</sub>
T <sub>1</sub> - T <sub>3</sub>	X <sub>1</sub> - T <sub>2</sub>
T <sub>4</sub> - T <sub>7</sub>	X <sub>1</sub> - T <sub>3</sub>
T <sub>5</sub> - T <sub>9</sub>	X <sub>2</sub> - T <sub>8</sub>
⋮	⋮

# 予測実験の3つのモード

Speaker-open と Speaker-pair-open

speaker-open		speaker-pair-open	
training	testing	training	testing
A - B	D - H	A - B	A - C
B - C	Y - D	B - C	B - D
B - F	G - X	B - F	C - F
Z - A	M - J	Z - A	Z - B
⋮	⋮	⋮	⋮

Speakers are *not* shared. Speaker pairs are *not* shared.

Speakers are shared. Speaker pairs are *not* shared.

## Speaker-open + speaker-pair-open

training {T <sub>i</sub> }	testing {X <sub>i</sub> }	
T <sub>1</sub> - T <sub>2</sub>	X <sub>1</sub> - T <sub>1</sub>	← speaker-open
T <sub>1</sub> - T <sub>3</sub>	X <sub>1</sub> - T <sub>2</sub>	
T <sub>4</sub> - T <sub>7</sub>	X <sub>1</sub> - T <sub>3</sub>	← speaker-pair-open
T <sub>5</sub> - T <sub>9</sub>	X <sub>2</sub> - T <sub>8</sub>	
⋮	⋮	

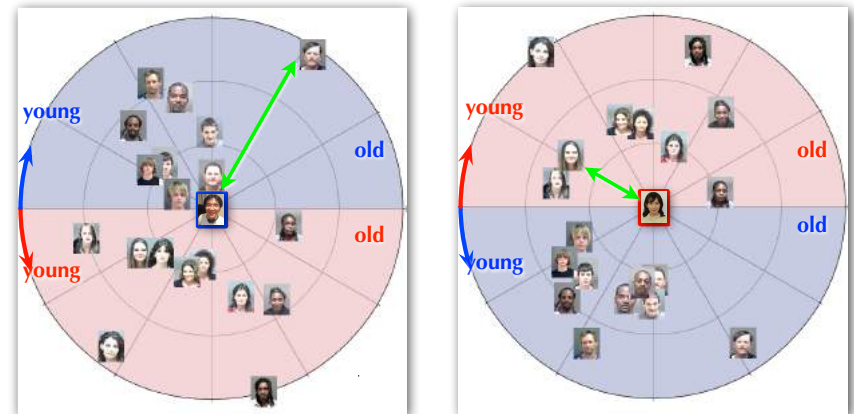
Speakers are shared only partially. Speaker pairs are *not* shared.

# 一つの応用可能性

世界諸英語話者の中に自身を位置づける

- 自分を中心に世界諸英語を眺める
- 自己中心的な世界諸英語ブラウザー[Kawase+'14]

training {T <sub>i</sub> }	testing {X <sub>i</sub> }
T <sub>1</sub> - T <sub>2</sub>	X <sub>1</sub> - T <sub>1</sub>
T <sub>1</sub> - T <sub>3</sub>	X <sub>1</sub> - T <sub>2</sub>
T <sub>4</sub> - T <sub>7</sub>	X <sub>1</sub> - T <sub>3</sub>
T <sub>5</sub> - T <sub>9</sub>	X <sub>2</sub> - T <sub>8</sub>
⋮	⋮



## 本発表の流れ

### はじめに

- 受け入れなければならない発音の多様性（世界諸英語）
- 発音多様性の最小単位は？

### 世界諸英語コーパスを使ったクラスタリング

- クラスタリングに必要なもの
- Speech Accent Archive

### 話者を単位とした世界諸英語発音クラスタリング

- 年齢・性別に頑健な発音分析法
- IPA 距離の自動予測実験
- 一つの応用可能性

### まとめと今後の課題

## まとめと今後の課題

- 避けて通れない英語発音の多様性問題
- クラスタリングに必要なこと・SAAコーパス
- 発音構造解析とIPA距離予測問題
- 世界諸英語ブラウザ
- 今後の課題

- Speaker-open mode での精度向上
- パラグラフ依存性は回避できるのか？
- 韻律的差異は導入できないか？
- 英語教育実践の場への導入の可能性



## 二つ告知させてください

### 峯松・齋藤研オープンラボ

- 毎年五月祭の時に、オープンラボを行なっています。
- 今年は**5月20日, 21日(土, 日)**です。是非お越しください。

### 人文系大学院生向け「音響音声学1・2」

- 通年・毎週水曜2限 (10:25~12:10) 希望者はまずメールを!!  
メアドはスライド1枚目
- 授業 web : [goo.gl/Ly9ZpH](http://goo.gl/Ly9ZpH) シラバスは下記
- 毎年、外部の人が数名（無料で）受講しています
  - 語学教師, 言語聴覚士, 声優, ボイストレーナー, 言語障害児の母親, 保育士, など音声に興味はあるが物理を学んだことがない方々
  - 「共同研究上必要」と言えば, 事務方もOKするはず・・・

本授業では高校で物理を履修しなかった学生を対象に、音声の物理的・音響的側面について分かり易く解説する。音声は音、即ち、空気（酸素・窒素・二酸化炭素など）の振動現象でしかない。しかし、その振動現象を鼓膜が捉えると、言語メッセージ、意図、感情、更には話者の健康状態など、様々な情報を我々は知覚できる。一体、空気振動のどこにこれらの豊富な情報が隠れているのだろうか？

音響音声学（1）では、音の基礎物理から始め、音声を音響的に眺めるために必要な基礎知識を提供すると共に、音刺激に対するインタフェースである聴覚の処理についても学ぶ。



ご清聴, 有り難うございました

