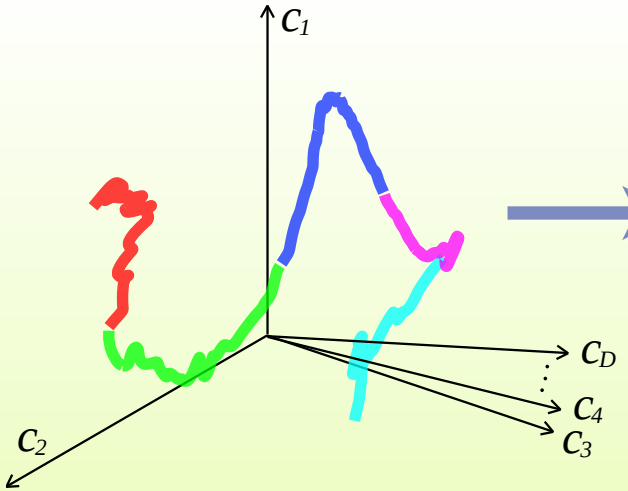
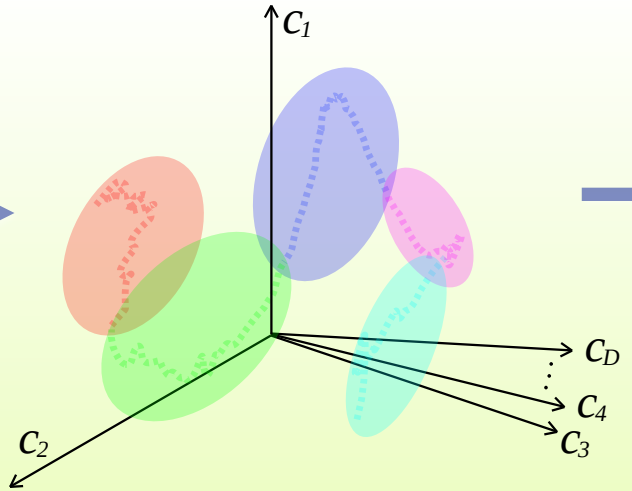
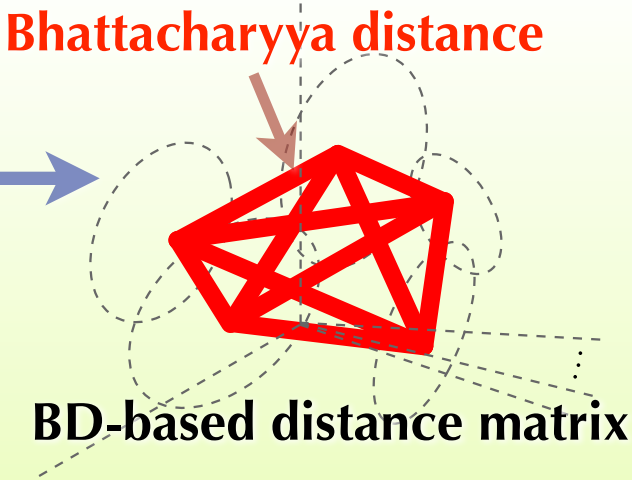
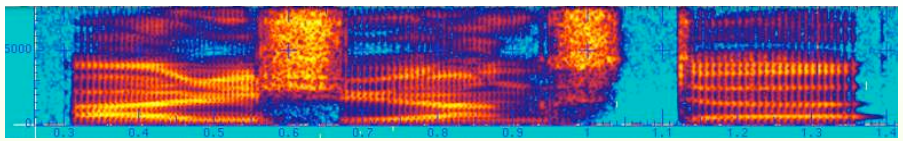
音声の構造的表象の工学的・実験的検証

f-div. (BD)に基づく一発声の構造化









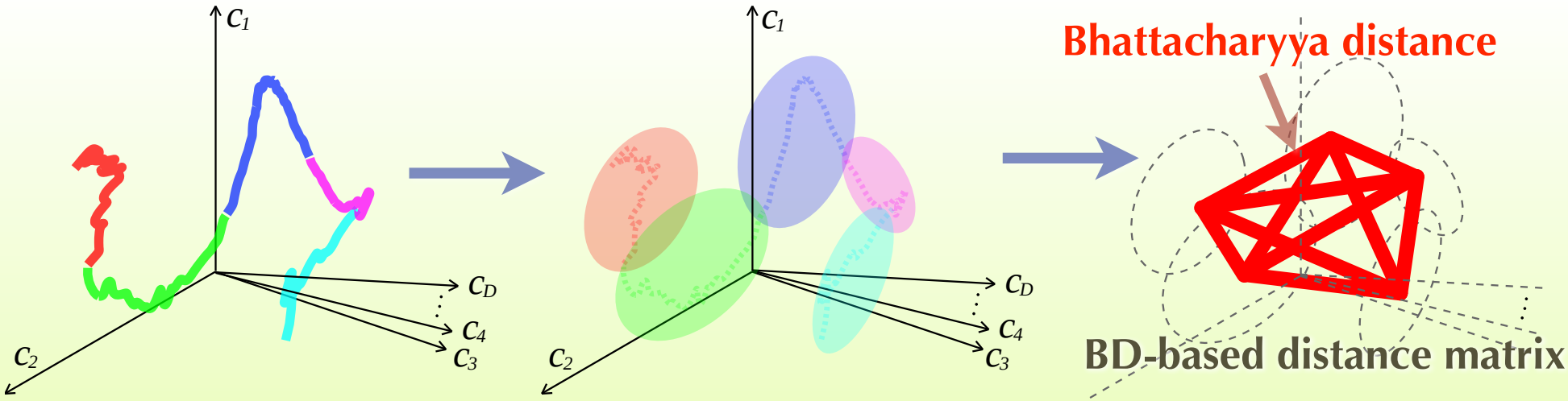
spectrogram (spectrum slice sequence)

cepstrum vector sequence

distribution sequence

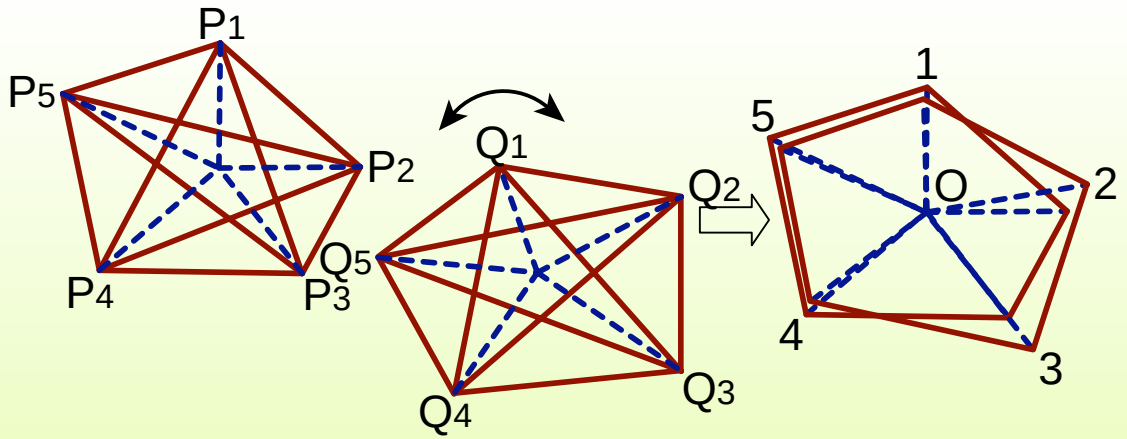
音声の構造的表象の工学的・実験的検証

f-div. (BD)に基づく一発声の構造化



２発声（＝２距離行列）間の音響照合

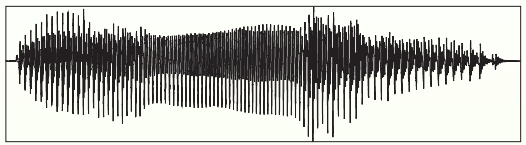
- ２距離行列間のユークリッド距離



- 回転：声道長差異
- シフト：マイク差異
- 話者適応・環境適応後のスコアが適応処理無しで算出
- 話者性を削除した音声表象

音声の構造的表象の工学的・実験的検証

Speech signal



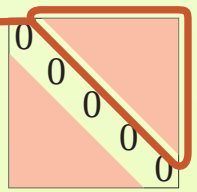
Cepstrum vector sequence

Cepstrum distribution sequence (HMM)

Distances of distributions

Structure (distance matrix)

$s = (s_1, s_2, \ldots) =$



$=$

Statistical structure model

Word 1

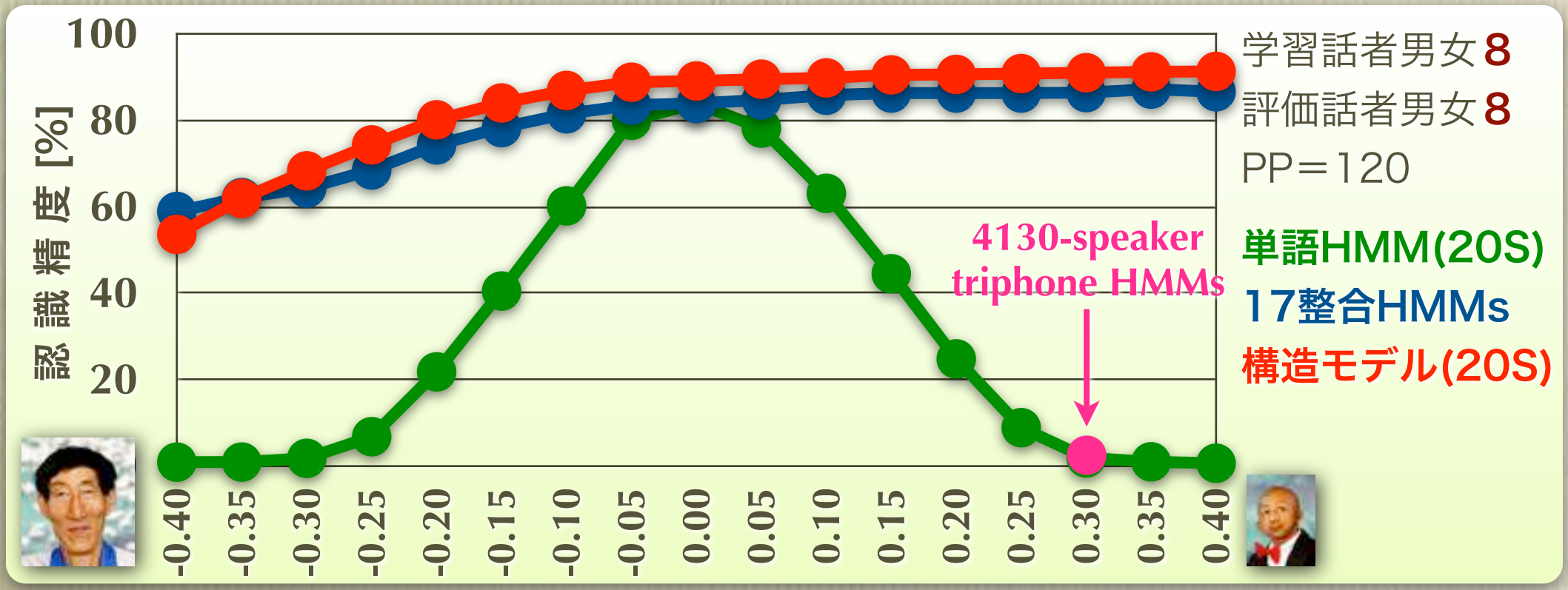
Word 2

Word N

音声の構造的表象の工学的・実験的検証

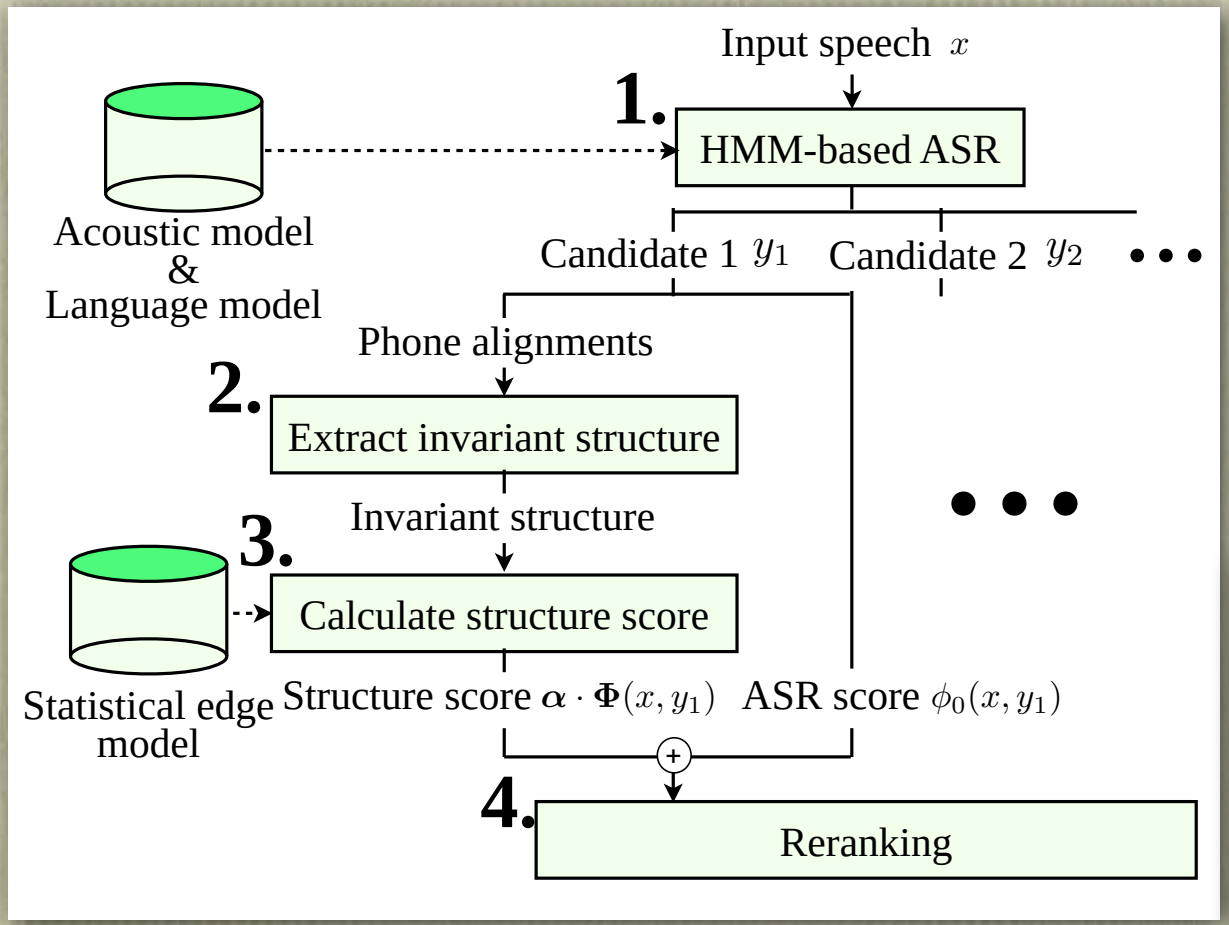
孤立単語音声の認識実験

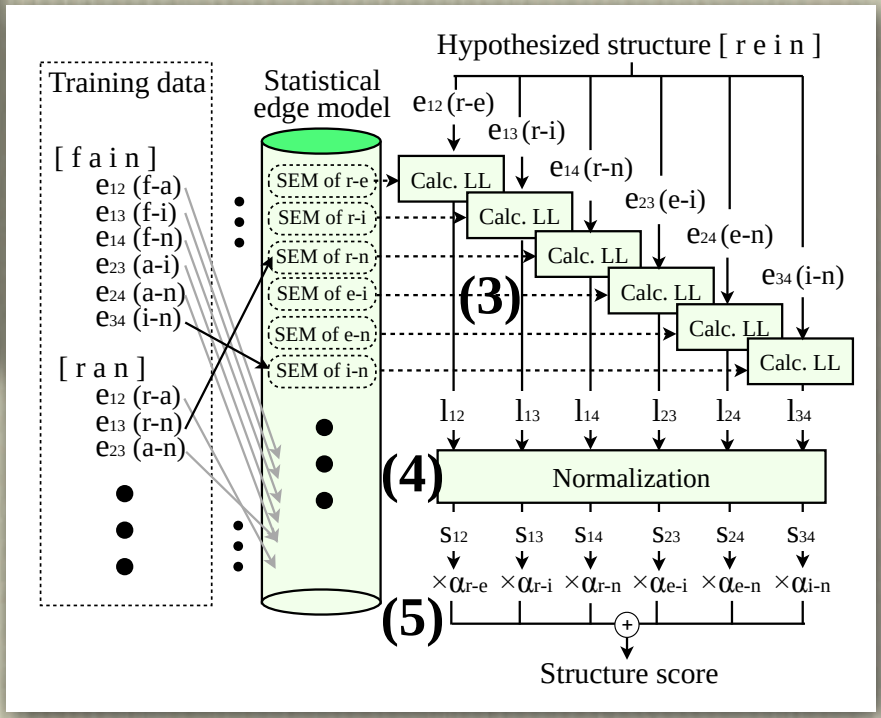
- 二つの問題とその解決
  - 強すぎる不変性→マルチストリーム構造化による都合のよい不変性へ
  - 高すぎる次元数→線形判別分析（LDA）による次元数削減
- 孤立単語認識実験による提案手法の評価
  - 日本語五母音を並び替えて作成される120単語の孤立単語認識



大語彙連続音声認識への応用

構造表象を複数仮説のリランキング処理に応用





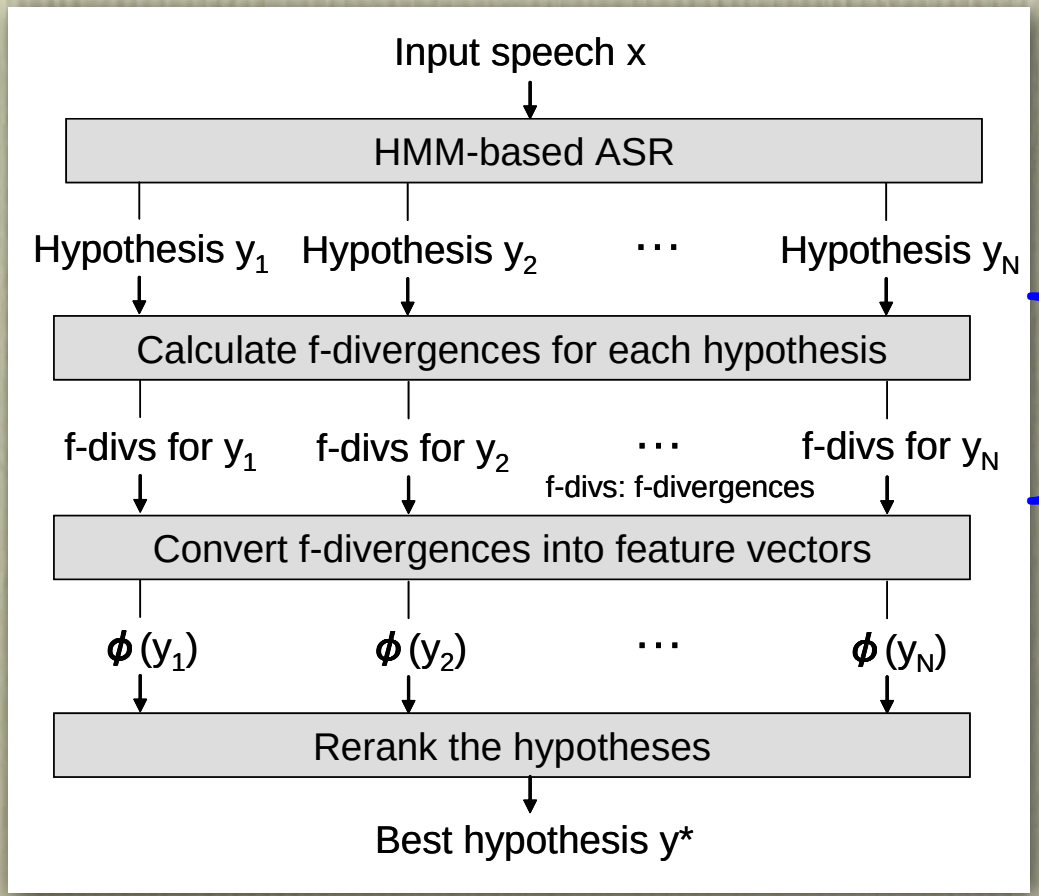
Application of structures to ASR

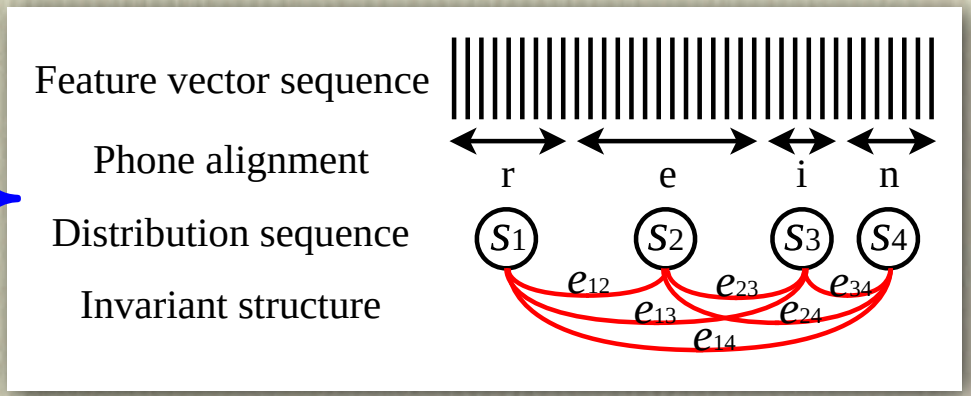
Application to more realistic ASR tasks [Suzuki+'15]

- Digits recognition and LVCSR (dictation)

Use of structural features in discriminative reranking

- Str. scores and ASR scores are combined with average perceptron.

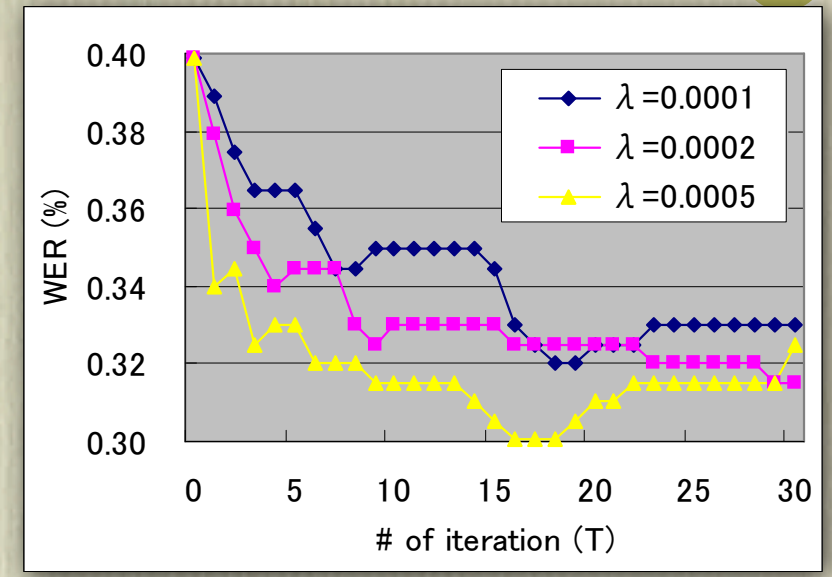




Application of structures to ASR

Continuous digits recognition

- Language = Japanese
- Baseline = GMM-HMM ASR
- Reranking = averaged perceptron
- Error reduction rate = 30%



Large vocabulary continuous speech recognition

- Language = Japanese
- Baseline = DNN-HMM ASR
- Reranking = averaged perceptron
- Error reduction rate = 5%

Many errors are due to a large number of homonyms in Japanese.

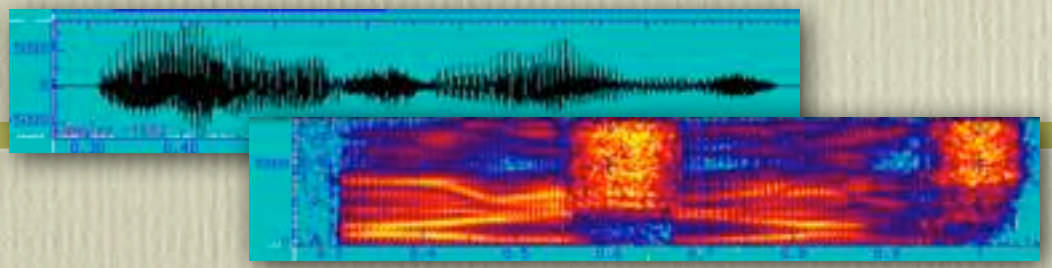
Table 6: CERs of the LVCSR experiment.

Baseline	Proposed	Relative improvement
2.67%	2.53%	5.24%

人文社会系研究科基礎文化研究専攻言語学専門分野

音響音声学

(Topics in Acoustic Phonetics)



峯松 信明

工学系研究科電気系工学専攻

本発表の流れ

- **刺激の物理的多様性とその認知的不変性**
  - 見え／色み／音高の多様性と自然・進化が編み出した解決方法
- **音声の物理的多様性とその認知的不変性**
  - 音色の多様性と工学者が編み出した解決方法
- **音声の構造的表象とそれに関する様々な考察**
  - 常識を覆すことで，違和感の解消を試みてみる。
- **音声の構造的表象と数学的表現と技術的実装**
  - 体格・性別に不変な音声波形・スペクトルの表現とは？
- **音声の構造的表象を用いた音声アプリケーション**
  - 音声認識，音声合成，発音分析，etc
- **音声の構造的表象の言語学的妥当性**
  - 何故，こうしてこなかったのか？　観測技術の功罪？

そもそもの出発点はここ

学習者の英語発音を評価する技術の構築

- 発音の習熟度を定量的にスコア化する。
- 個々の発声における，不適切な発音部位を検出する。

発音評価技術構築における根本的な問題

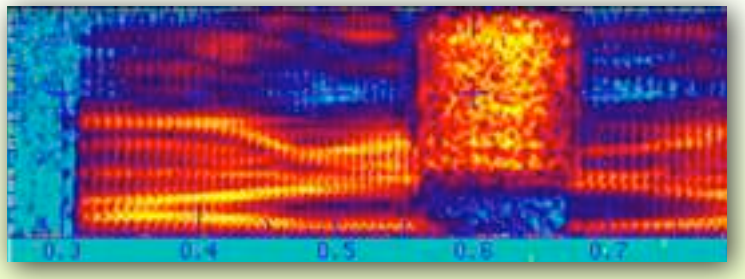
- 教師音声データベースから個々の音素の音響モデルを構築する。
- 学習者は小学生かもしれない。
- 音韻の違い＝話者の違い＝管形状の違い＝音色の違い

学習者にあった音響モデルの構築

- 話者適応技術
  - 大人用モデル／子供用モデル
- 発音評定？声帯模写評定？

峯松の声から峯松を消したい

- そんなことができるのか？



$- \quad = \quad ?$

- 個々の発声における，不適切な発音部位を検出する。

発音評価技術構築における根本的な問題

- 教師音声データベースから個々の音素の音響モデルを構築する。
- 学習者は小学生かもしれない。
- 音韻の違い＝話者の違い＝管形状の違い＝音色の違い

学習者にあった音響モデルの構築

- 話者適応技術
  - 大人用モデル／子供用モデル
- 発音評定？声帯模写評定？

峯松の声から峯松を消したい

- そんなことができるのか？

で，こういうのができました。

峯松の普通の英語発音はどちらに近いのか？

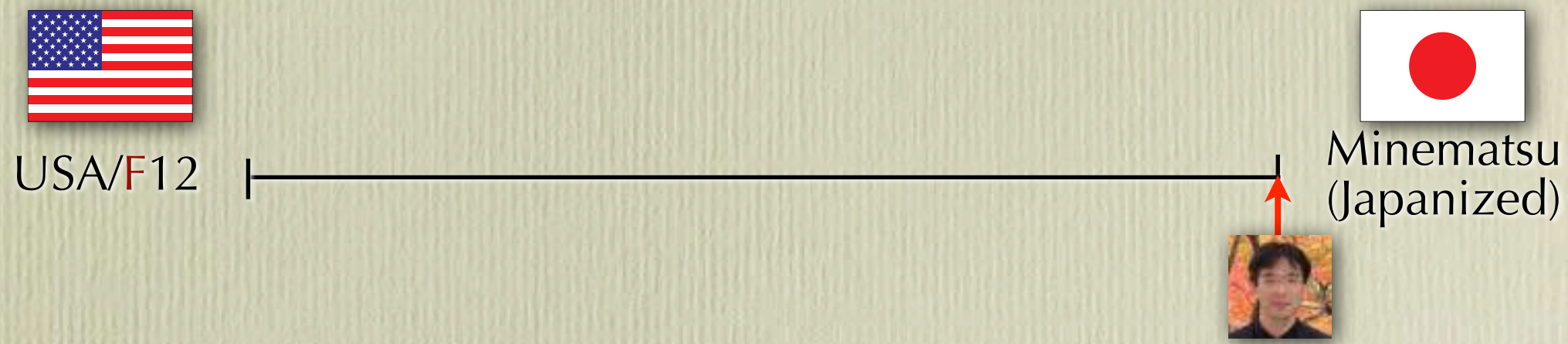
speaker	USA/F12	×	Minematsu	○	Minematsu
gender	female	×	male	○	male
age	??	×	36	○	36
mic	Sennheiser	×	cheap mic	○	cheap mic
room	recording room	×	living room	○	living room
AD	SONY DAT	×	PowerBook	○	PowerBook
proficiency	perfect	△	good	×	Japanized

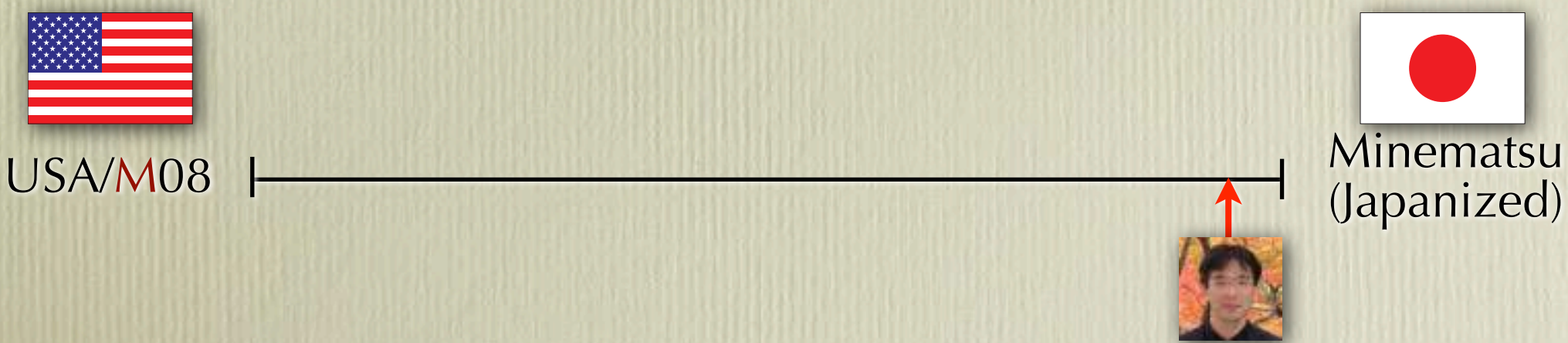
Those answers will be straight forward if you think them through carefully first.

で，こういうのができました。

母語話者発音と日本語的峯松発音の音響モデルを利用

- 各モデルから峯松英語発声が観測される確率 $P(o|M)$





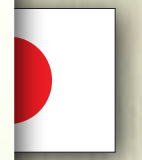
で，こういうのができました。

母語話者発音と日本語的峯松発音の音響モデルを利用

- 峯松発音が，所望の音素列の読み上げである確率 $P(M|o)$

$$
\begin{aligned}
P(M|o) &= P(p_1, ..., p_N|o) \\
&= \frac{P(o|p_1, ..., p_N)P(p_1, ..., p_N)}{\sum_{p_i} P(o|p_1, ..., p_N)P(p_1, ..., p_N)} \\
&\approx \frac{P(o|p_1, ..., p_N)}{\sum_{p_i} P(o|p_1, ..., p_N)} \\
&\approx \frac{P(o|p_1, ..., p_N)}{\max_{p_i} P(o|p_1, ..., p_N)} \\
&= \frac{P(o|M)}{\max_M P(o|M)} \\
&= \text{GOP (Goodness Of Pronunciation)}
\end{aligned}
$$

USA



matsu
ized)

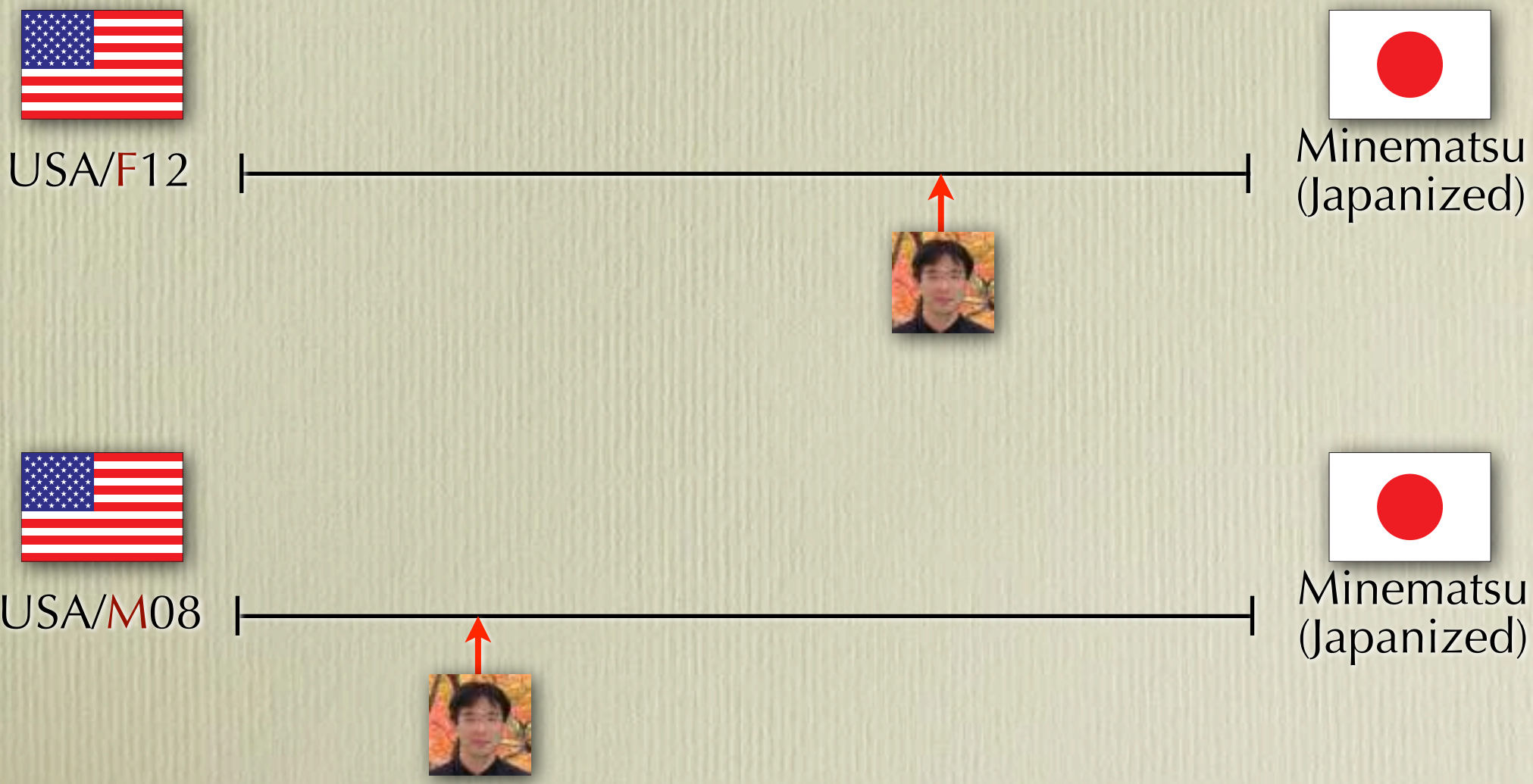
USA

matsu
ized)

で，こういうのができました。

母語話者発音と日本語的峯松発音の音響モデルを利用

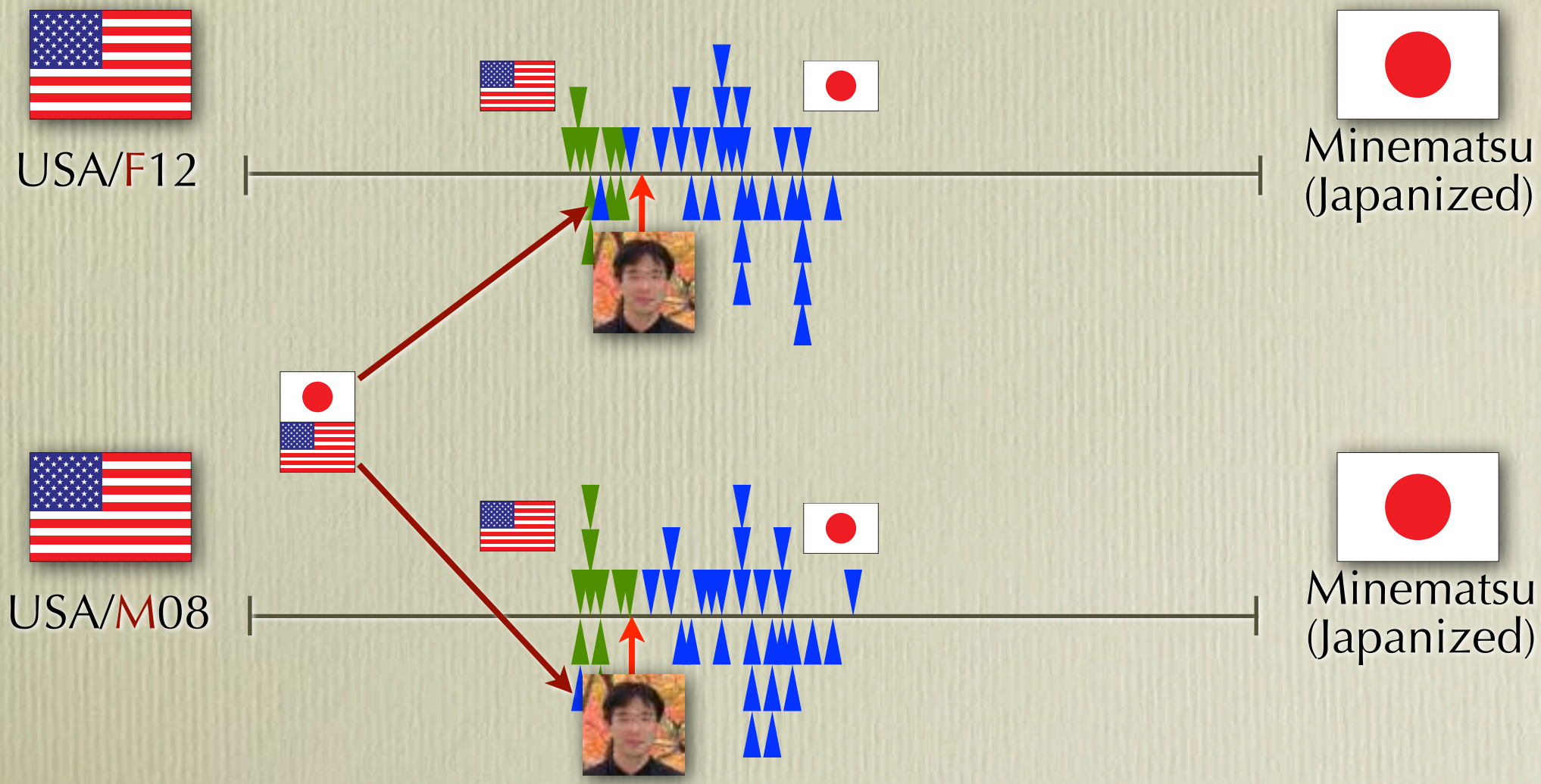
- 峯松発声が，所望の音素列の読み上げである確率 $P(M|o)$



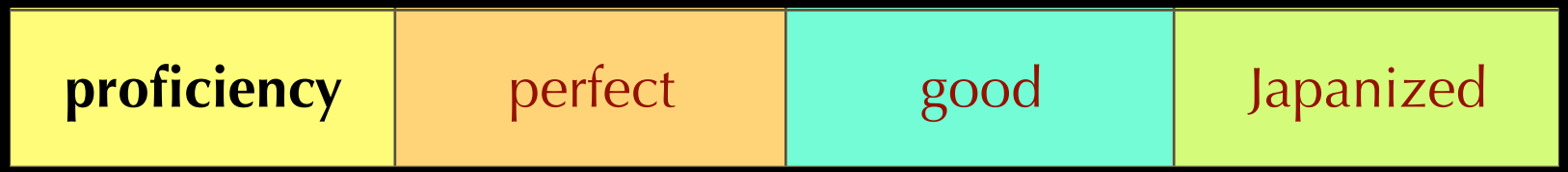
で，こういうのができました。

母語話者発音と日本語的峯松発音の構造モデルを利用

- 母語発音構造，日本語的峯松発音構造と，峯松発音構造を比較

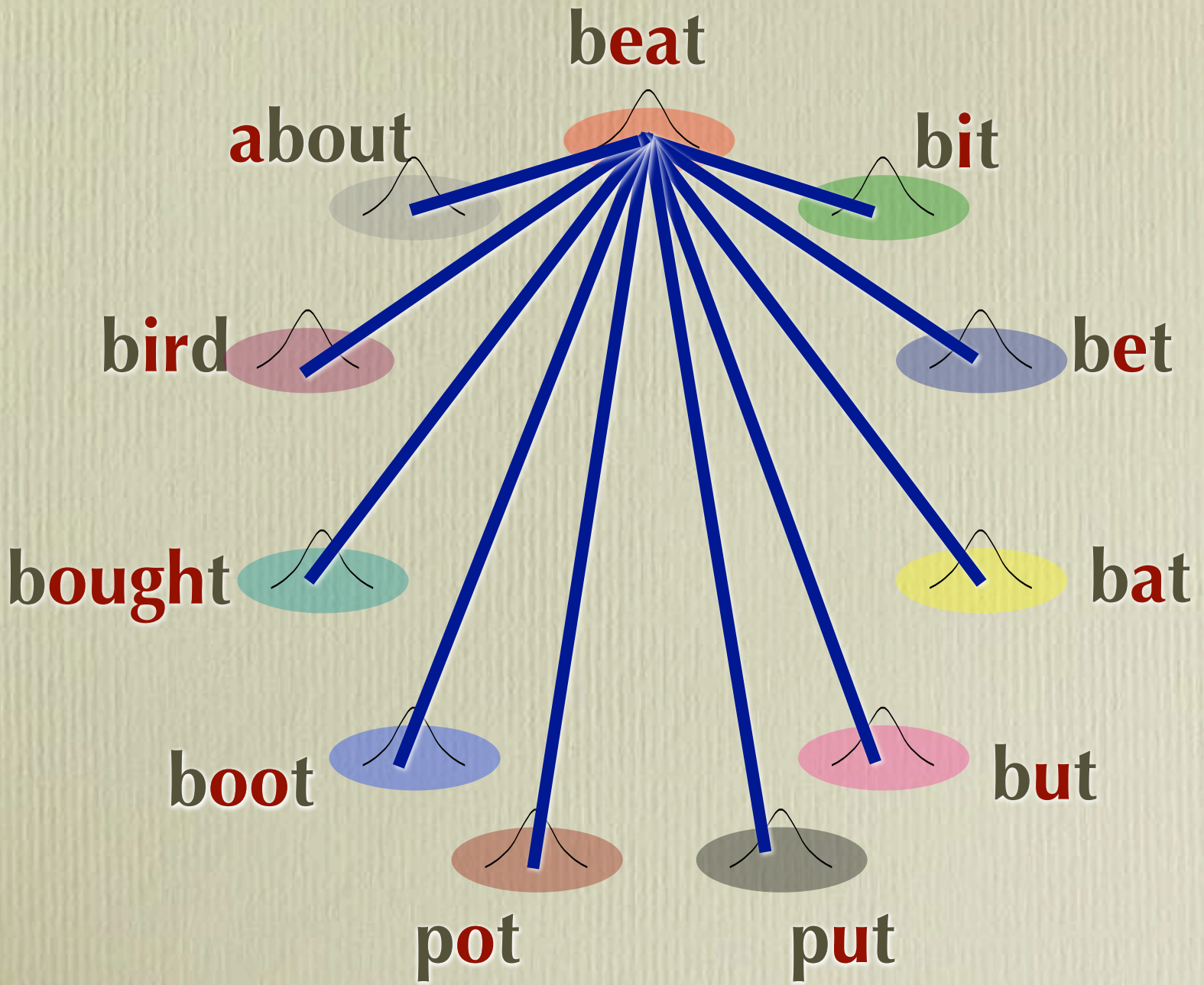


speaker	USA/F12	Minematsu	Minematsu
gender	female	male	male
age	??	36	36
mic	Sennheiser	cheap mic	cheap mic
room	recording room	living room	living room
AD	SONY DAT	PowerBook	PowerBook
proficiency	perfect	good	Japanized


proficiency
perfect
good
Japanized

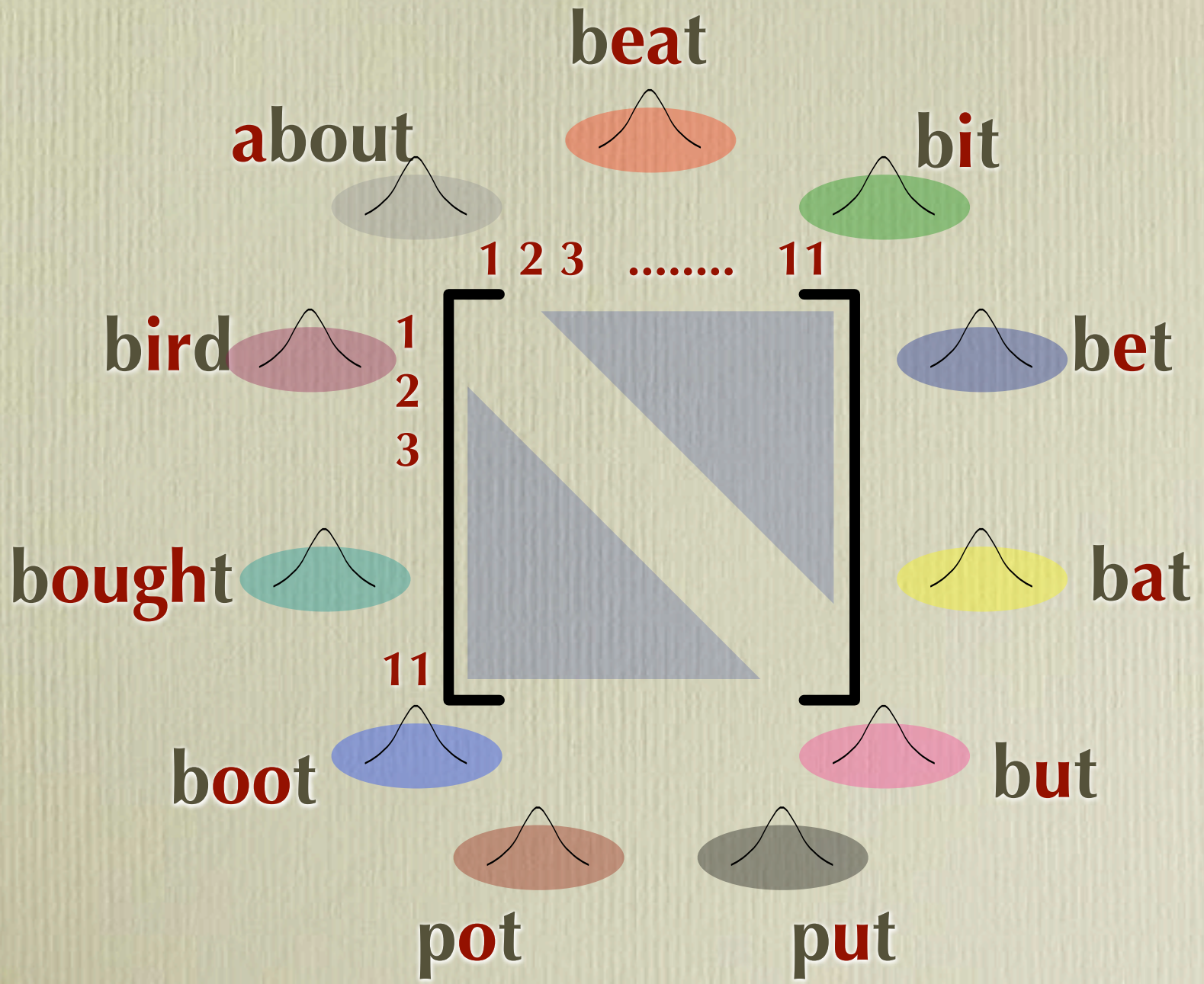
英語母音の発音トレーニング

- 個々の母音ではなく，母音群の体系を評価対象



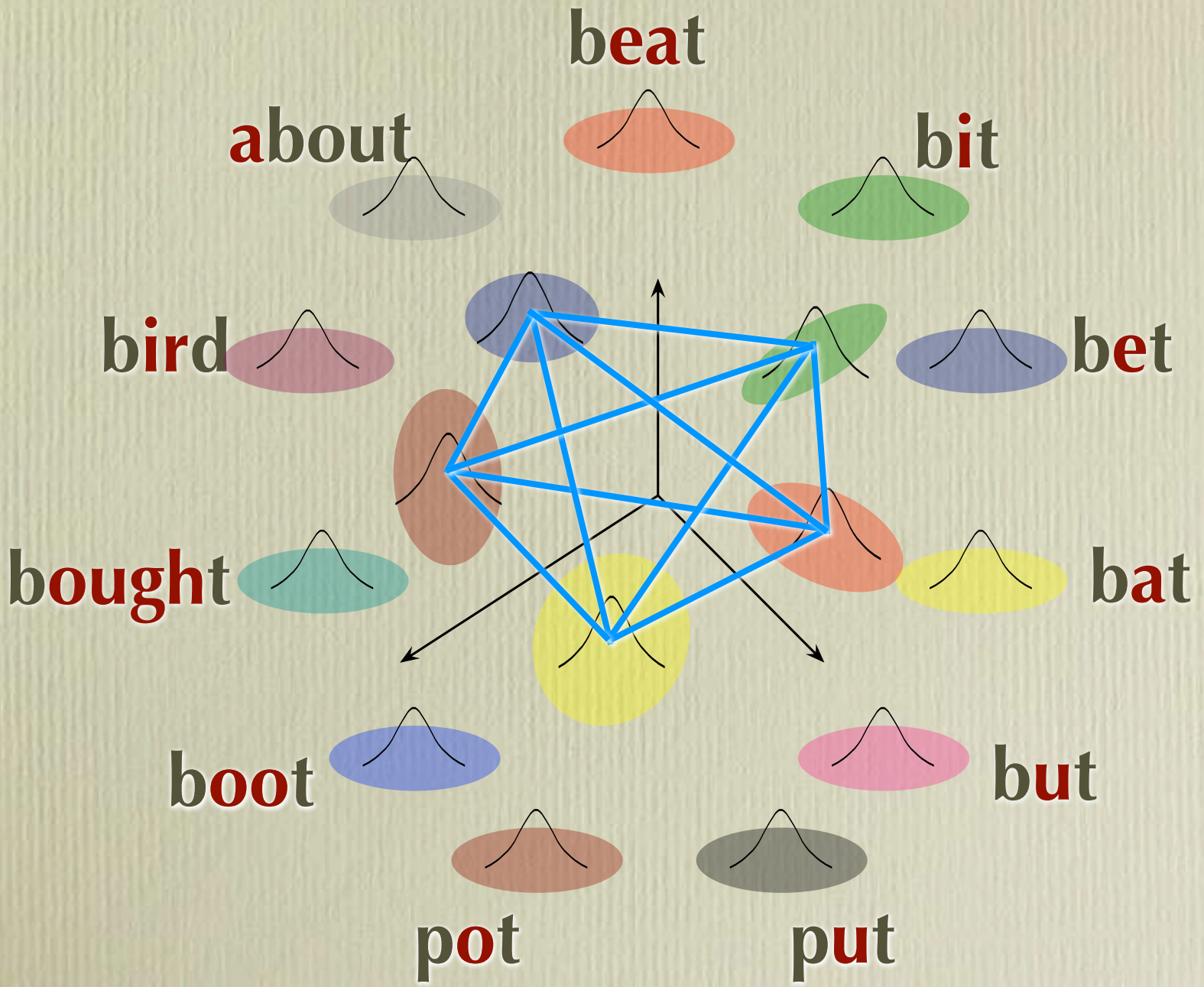
英語母音の発音トレーニング

- 個々の母音ではなく，母音群の体系を評価対象



英語母音の発音トレーニング

- 個々の母音ではなく，母音群の体系を評価対象

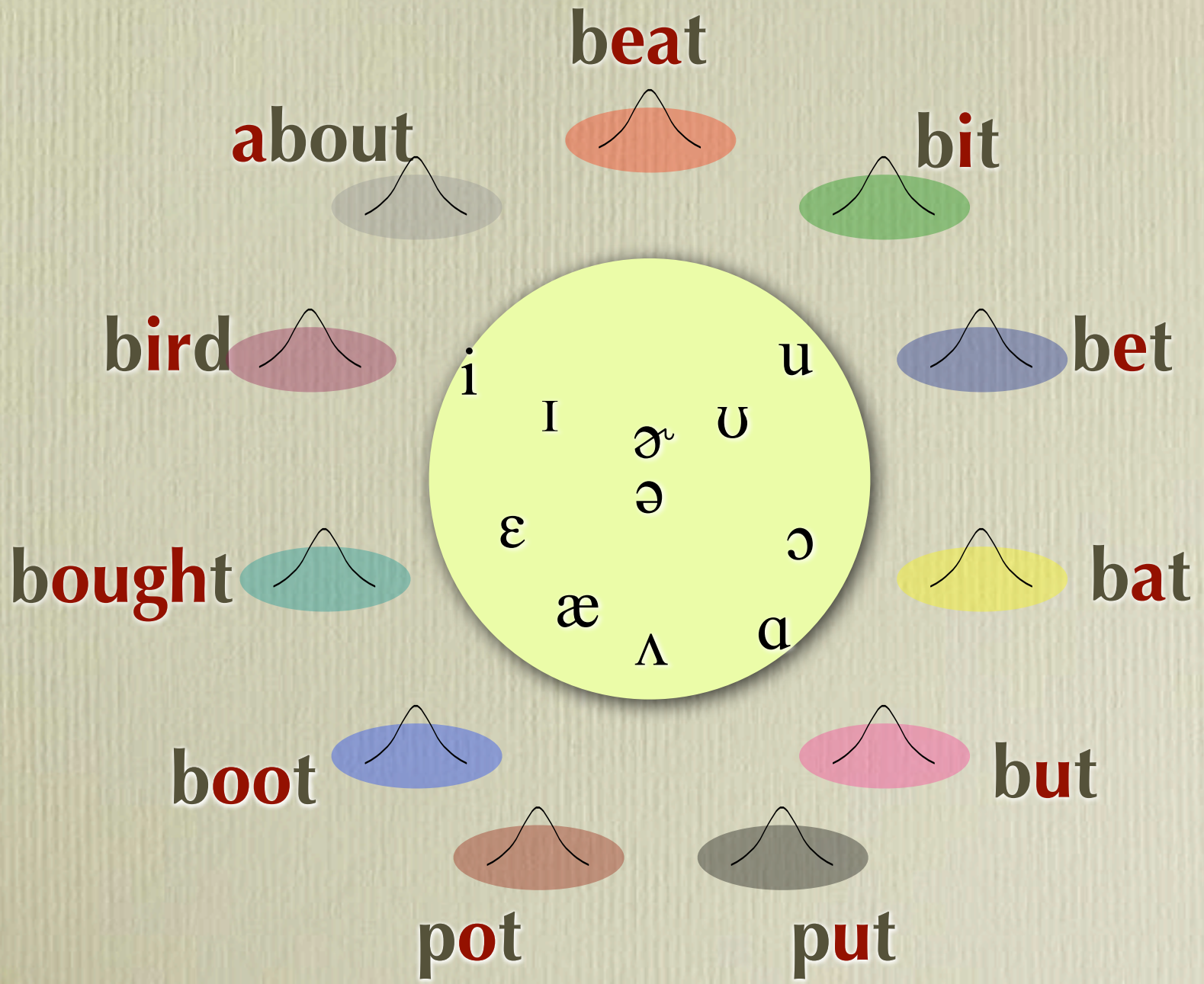


英語母音の発音トレーニング

- 個々の母音ではなく，母音群の体系を評価対象

英語母音の発音トレーニング

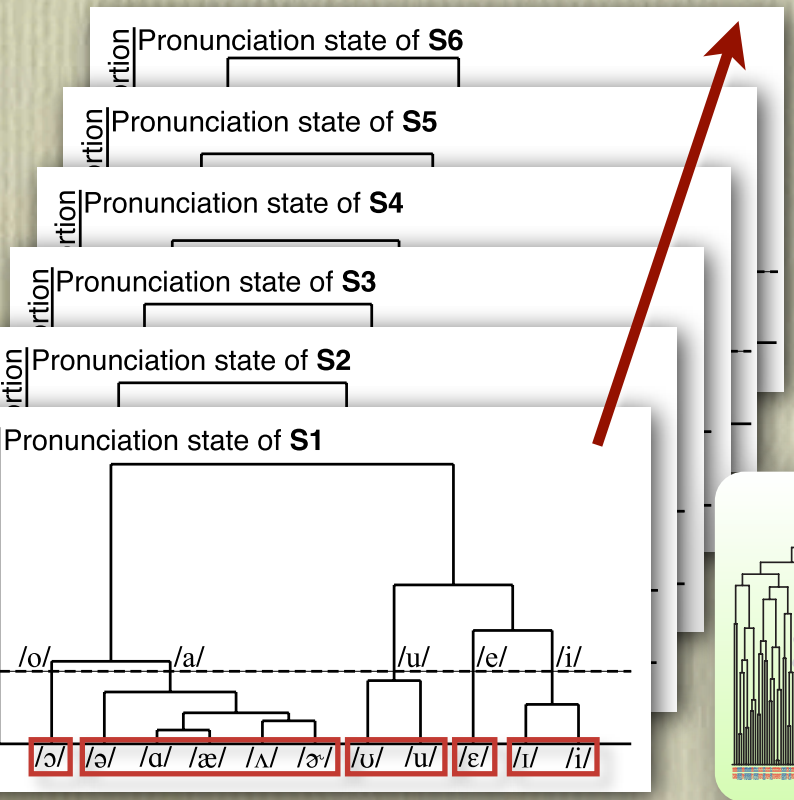
- 個々の母音ではなく，母音群の体系を評価対象

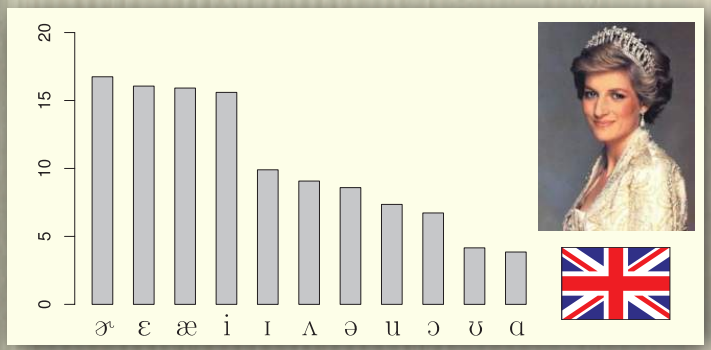


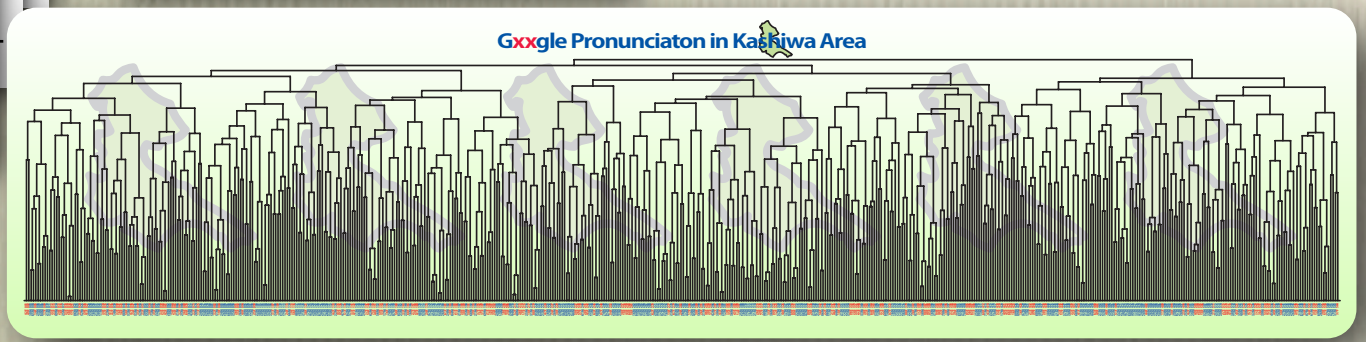
母音体系に見る発音習熟度

発音の構造表象に基づいて構築した技術・機能

- 学習者の母音体系がどのような状態にあるのかを記録する
- お手本の母音構造と比較して，矯正対象の母音を教示する
- 複数の教師からお手本としたい教師（相手）を選ぶ
- 年齢，性別などには影響されない学習者発音の分類







学習者の発音状態の記録とその遷移

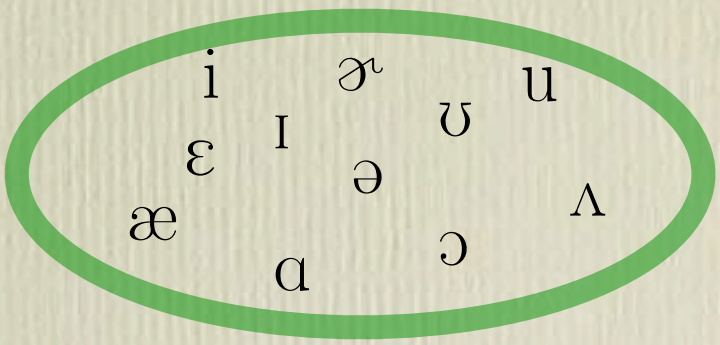
分析対象の音声資料

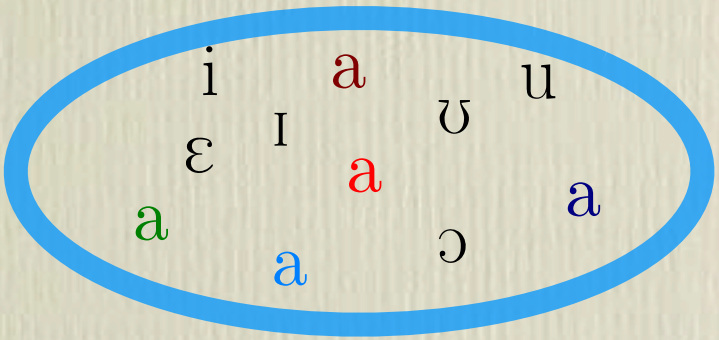
- 英語劇経験者による英語＆日本語単語音声
  - beat, bit, bet, bat, but, pot, bought, boot, put, bird, about　x 1
  - ばと，びと，ぶと，べと，ぼと　x 5

日本人学習者の母音発音の模擬的生成

- 幾つかの米語母音と日本語母音の置き換え

a	ɑ, æ, ʌ, ə, ɚ
i	ɪ, i
u	ʊ, u
e	ɛ
o	ɔ

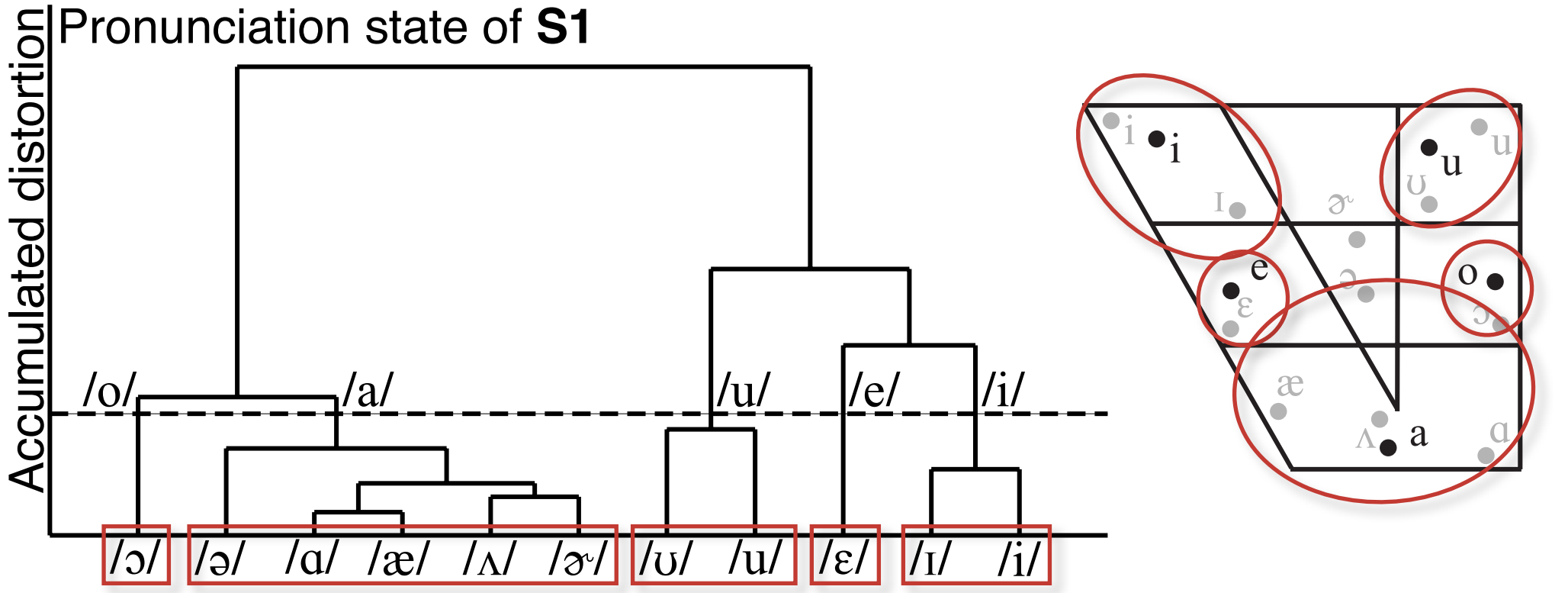




- 「あ」的な母音の置換
- 「い」的な母音の置換
- ：

学習者の発音状態の記録とその遷移

S1からS6への遷移の様子の記述

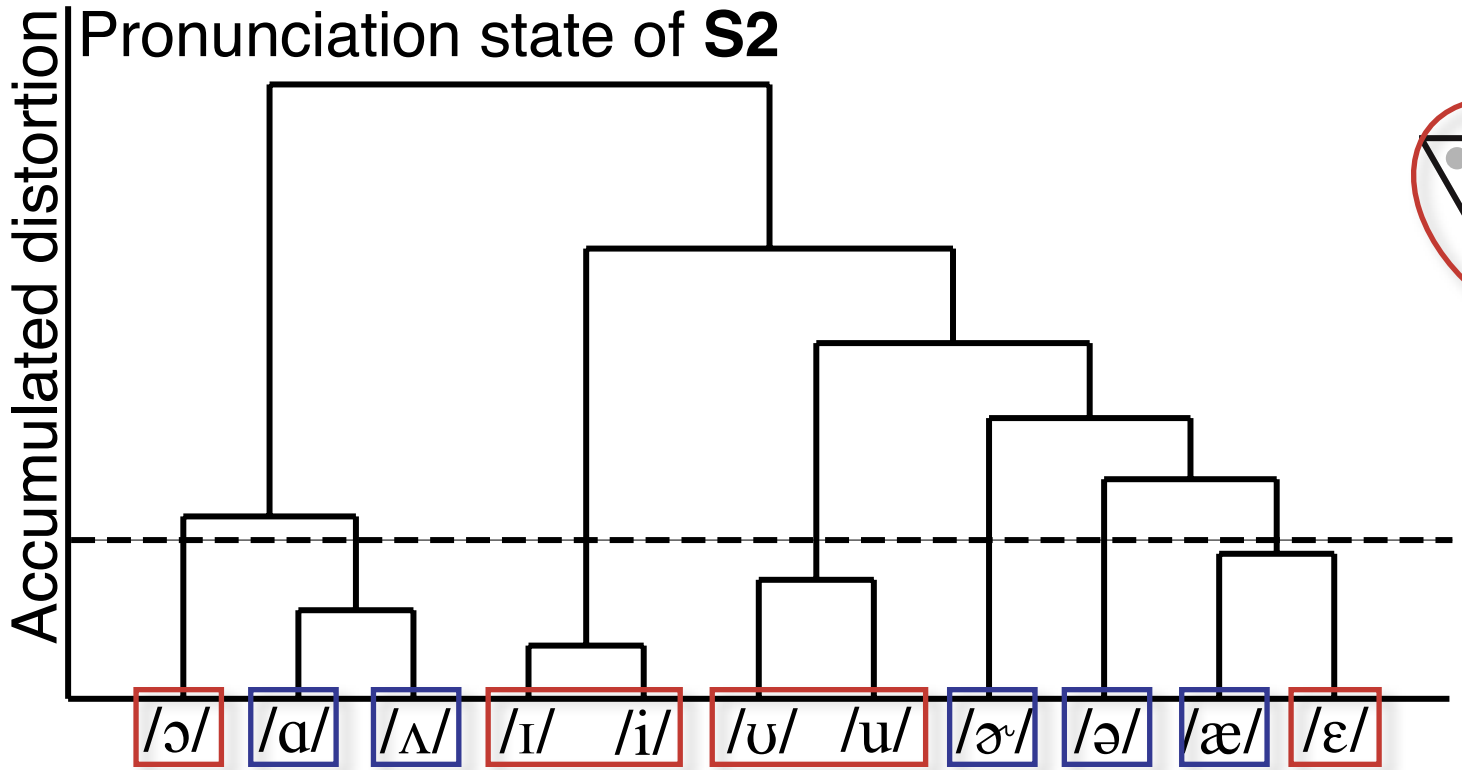


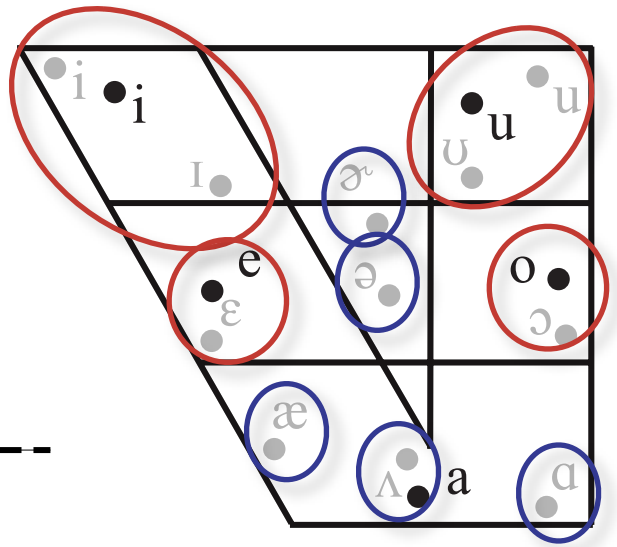
S1 → S2 → S3 → S4 → S5 → S6

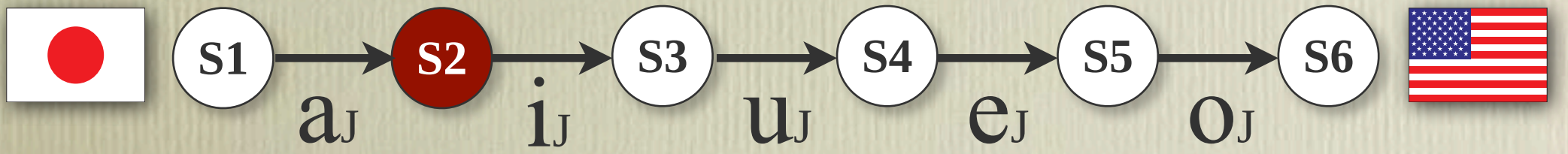
a_J i_J u_J e_J o_J

学習者の発音状態の記録とその遷移

S1からS6への遷移の様子の記述

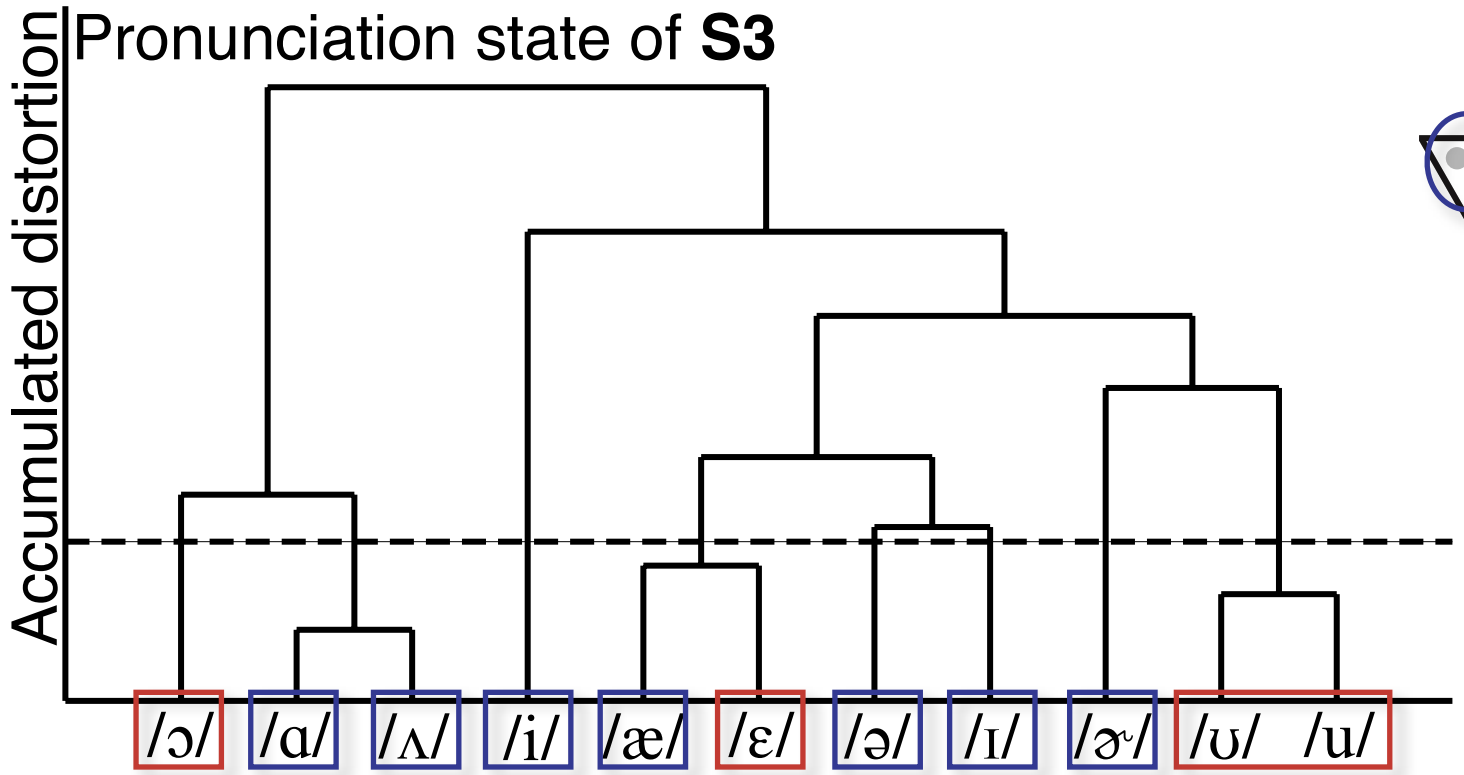


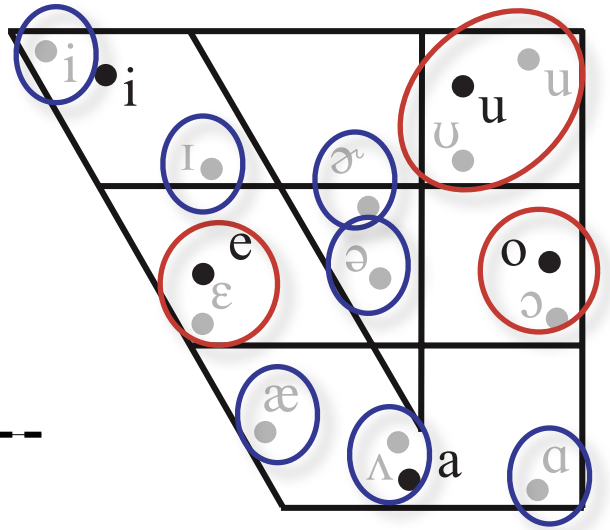




学習者の発音状態の記録とその遷移

S1からS6への遷移の様子の記述



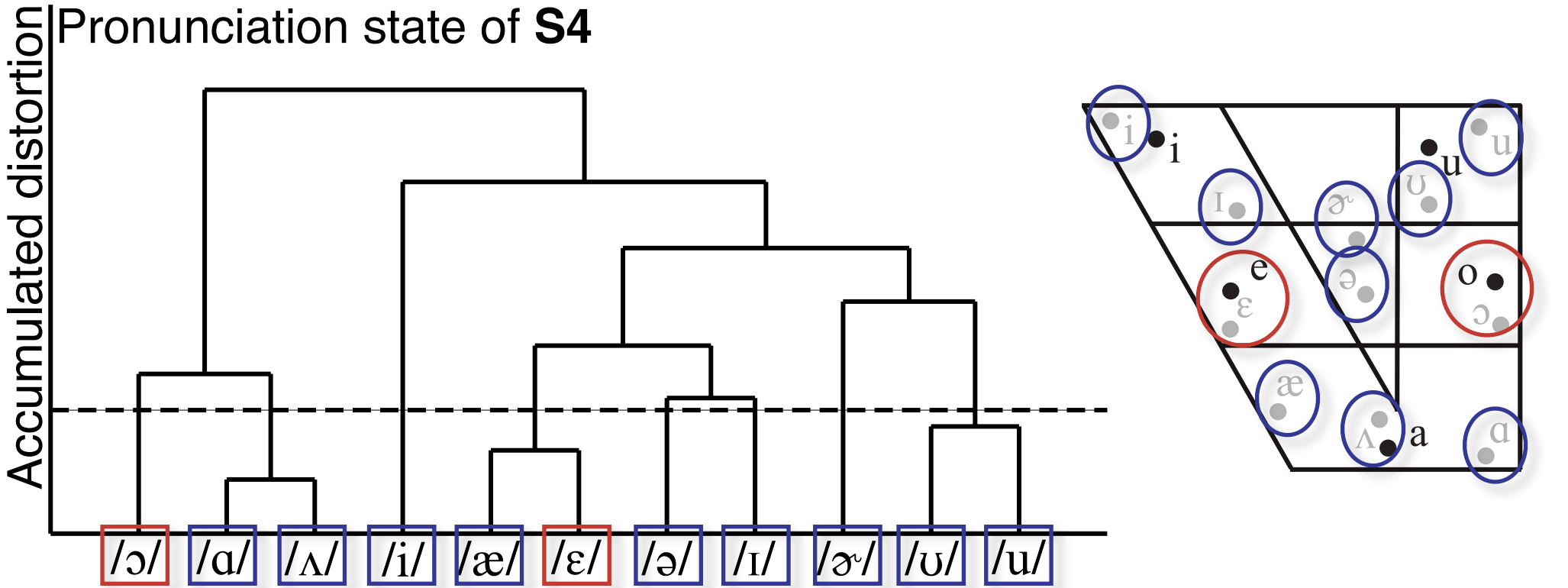


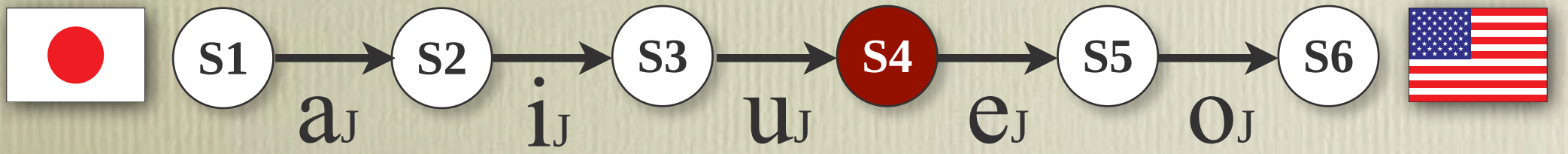
S1 → S2 → S3 → S4 → S5 → S6

a_J i_J u_J e_J o_J

学習者の発音状態の記録とその遷移

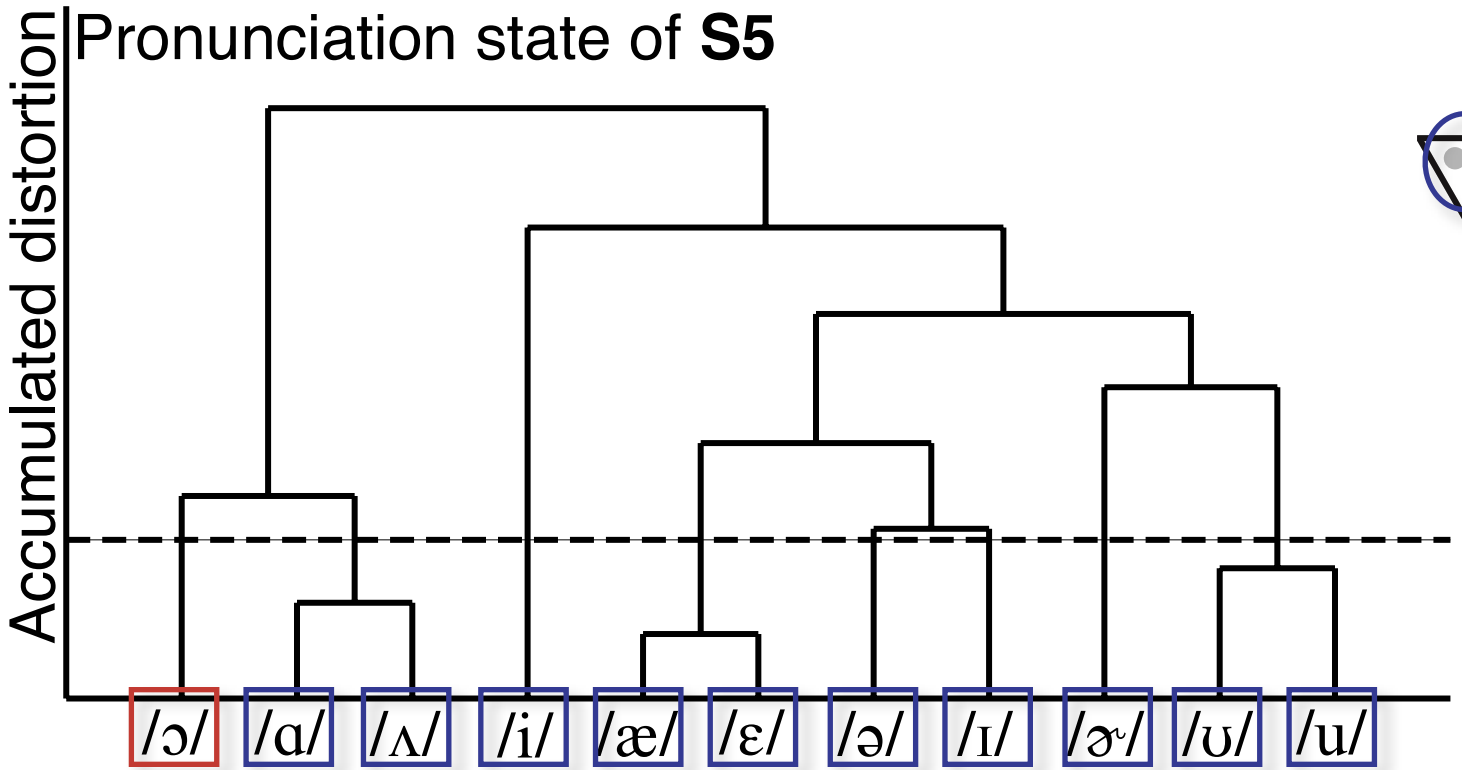
S1からS6への遷移の様子の記述

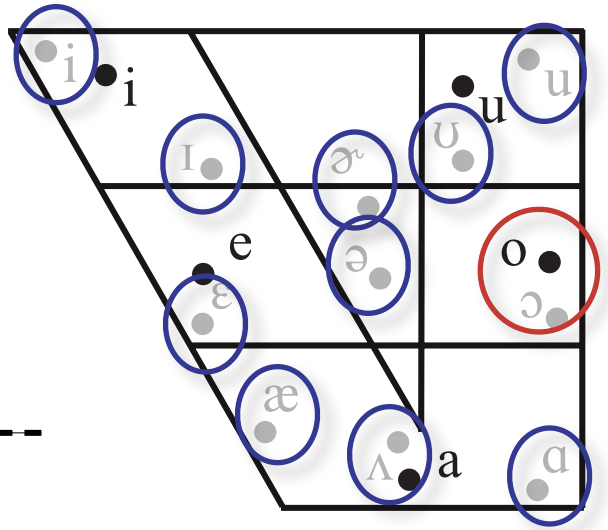


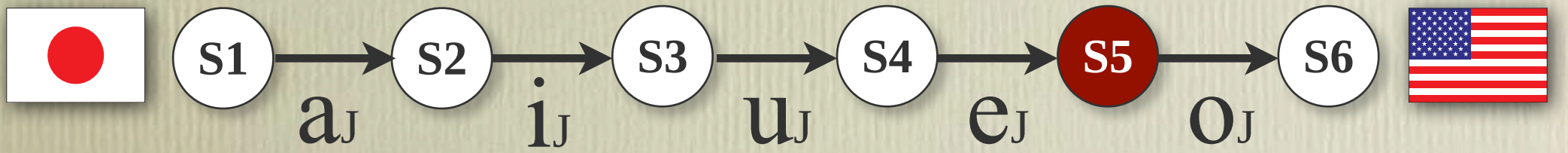


学習者の発音状態の記録とその遷移

S1からS6への遷移の様子の記述

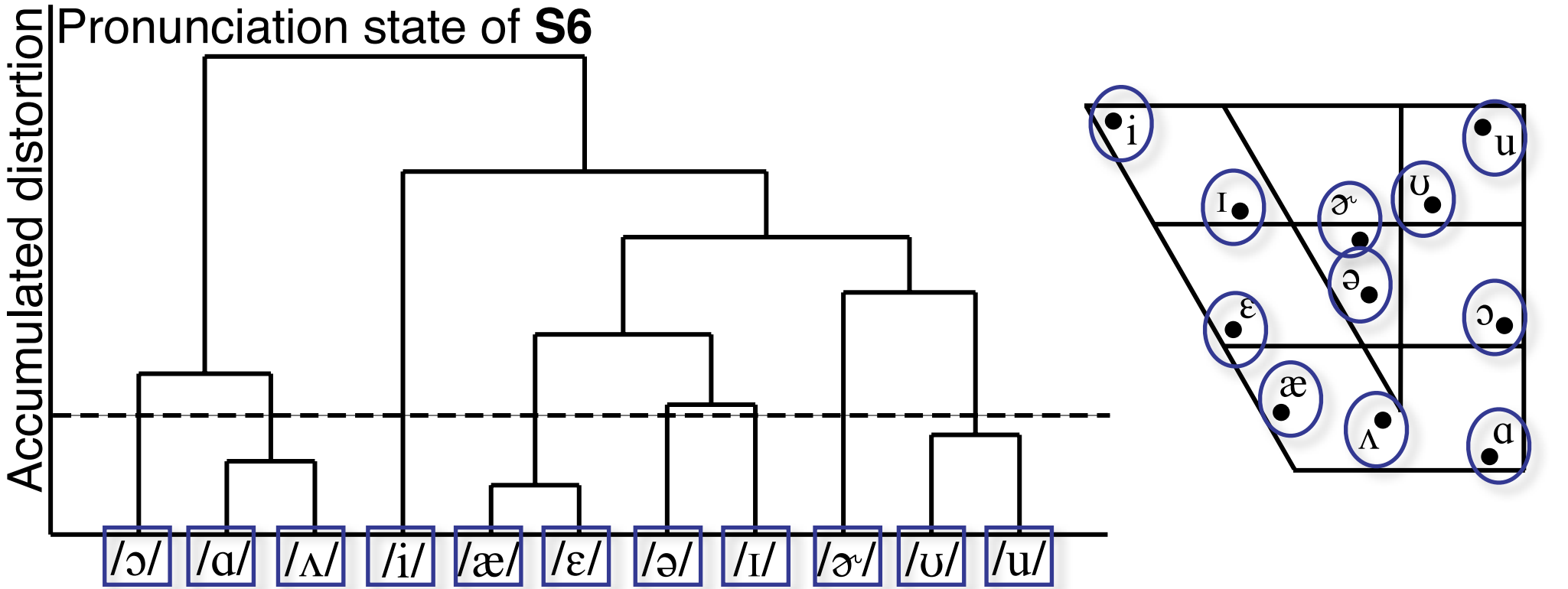


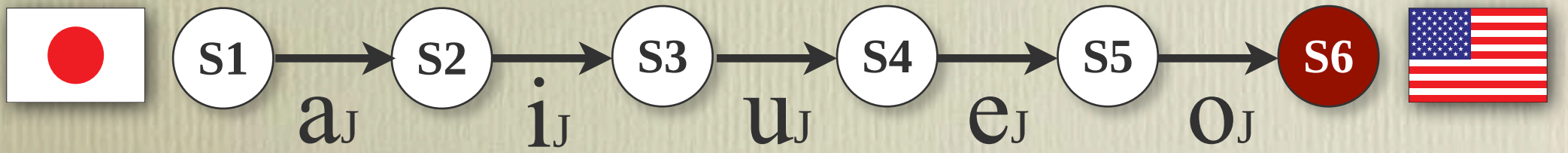




学習者の発音状態の記録とその遷移

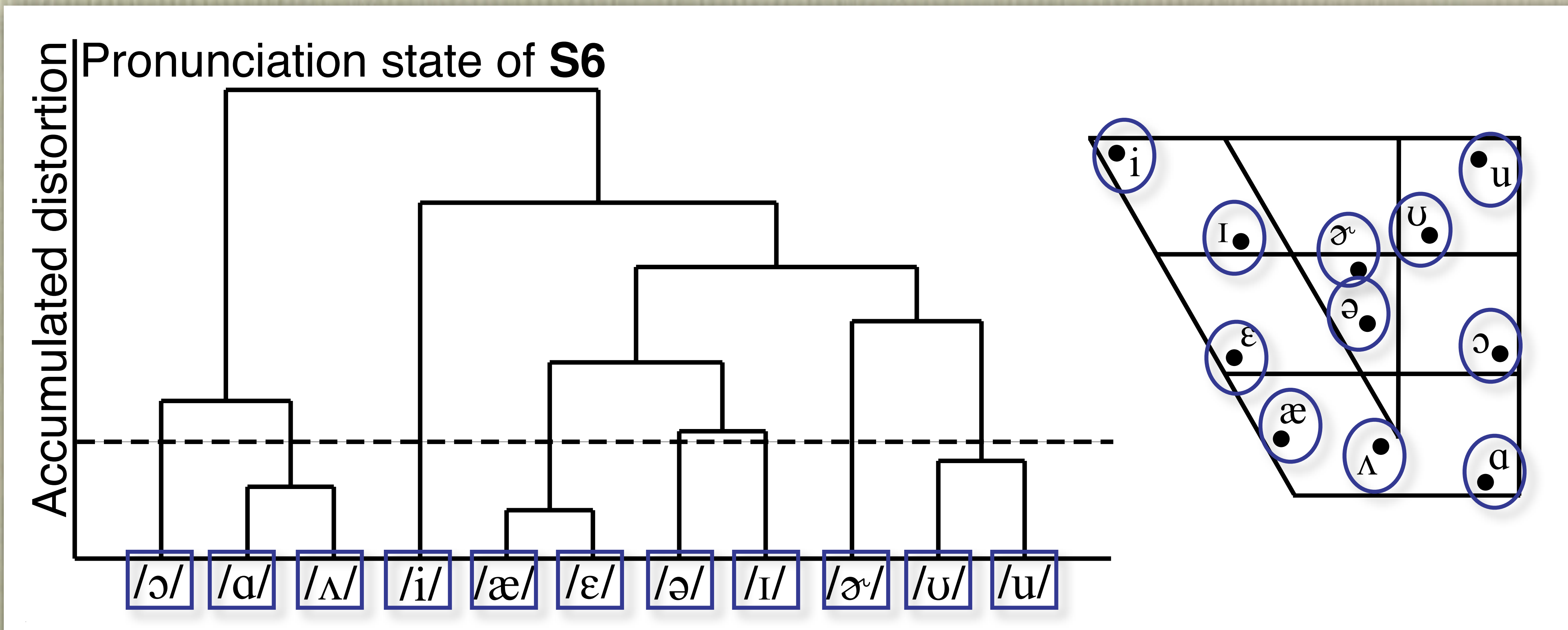
S1からS6への遷移の様子の記述





学習者の発音状態の記録とその遷移

S1からS6への遷移の様子の記述

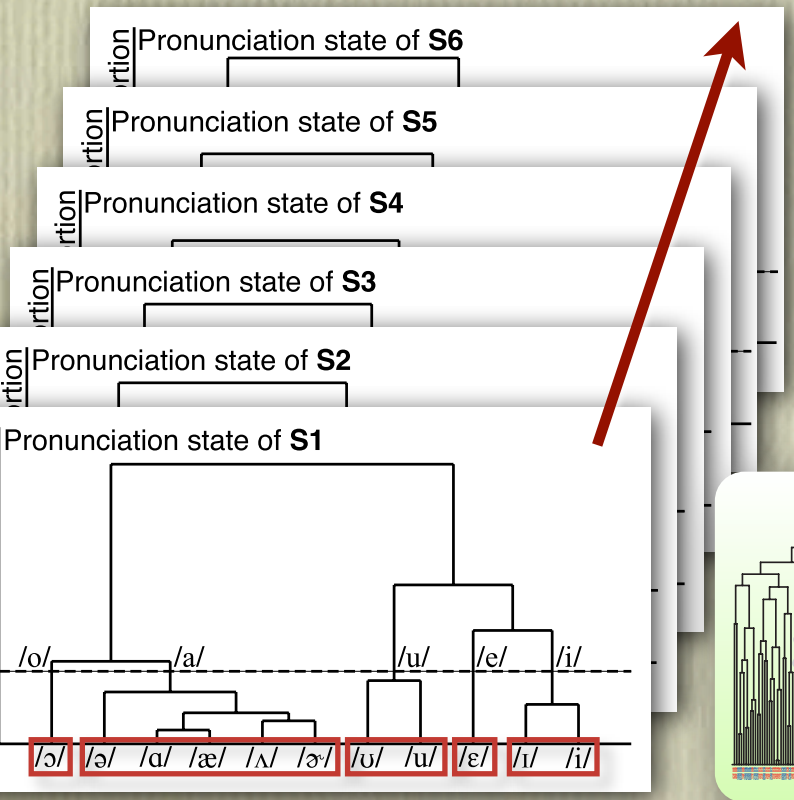


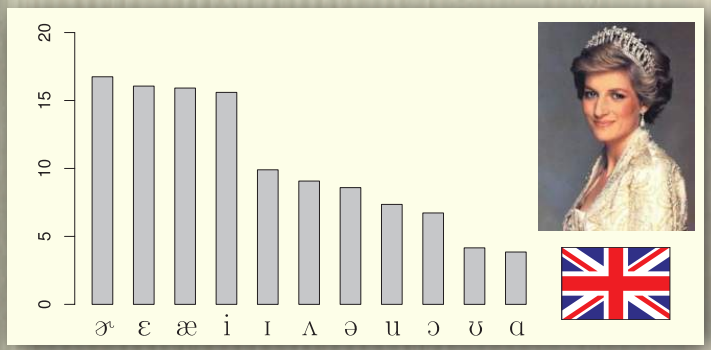
Gooooooooooooooooooooooooal!!

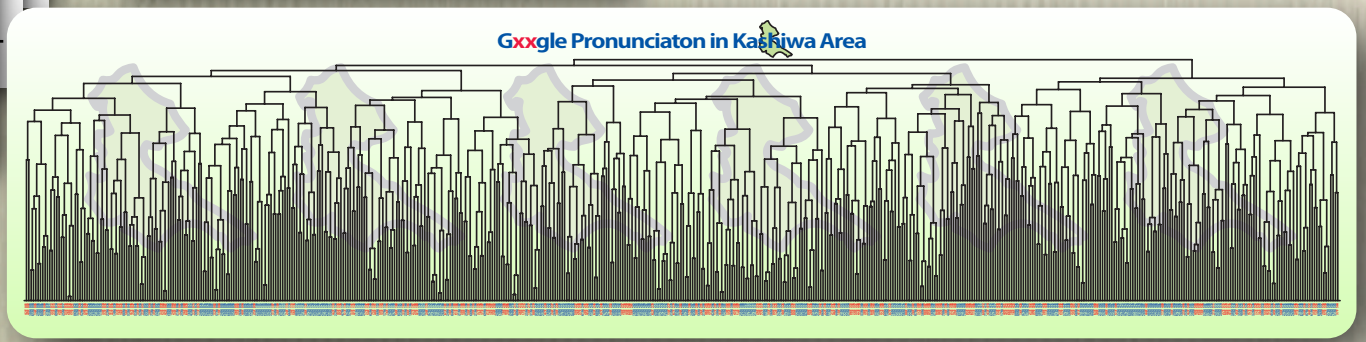
母音体系に見る発音習熟度

発音の構造表象に基づいて構築した技術・機能

- 学習者の母音体系がどのような状態にあるのかを記録する
- お手本の母音構造と比較して，矯正対象の母音を教示する
- 複数の教師からお手本としたい教師（相手）を選ぶ
- 年齢，性別などには影響されない学習者発音の分類

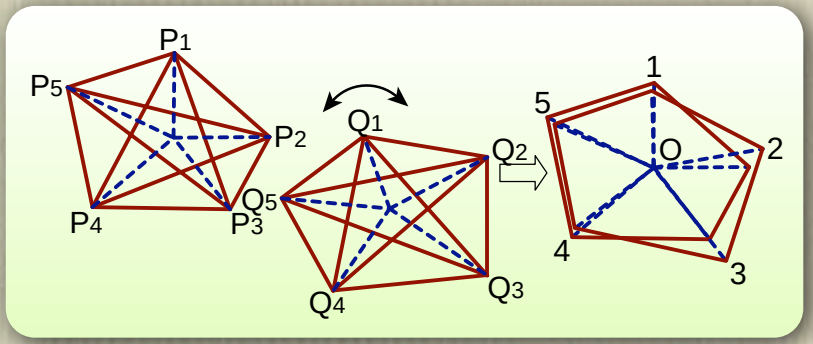






優先的に矯正すべき母音の教示

全体的構造歪みと局所的構造歪み



- 全体的構造歪み→発音習熟度の推定

$$TD(S,T) = \sqrt{\frac{1}{M}\sum_{i<j}(S_{ij} - T_{ij})^2}$$

- 局所的構造歪み→矯正すべき発音部位の推定

$$LD(S,T,i) = \sum_{j=1}^{M} |S_{ij} - T_{ij}|$$

  - LD ＝ 大 ↔ 当該母音の矯正度 ＝ 大

$$S = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & & \\ & 0 & & & \\ & & 0 & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix} \end{array} \quad \longleftrightarrow \quad T = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & & & & \\ & 0 & & & \\ & & 0 & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix} \end{array}$$

優先的に矯正すべき母音の教示

S8：米語，S1-S7：日本語混入米語

- 教師と学習者の母音距離行列のみから，優先的矯正対象を推定

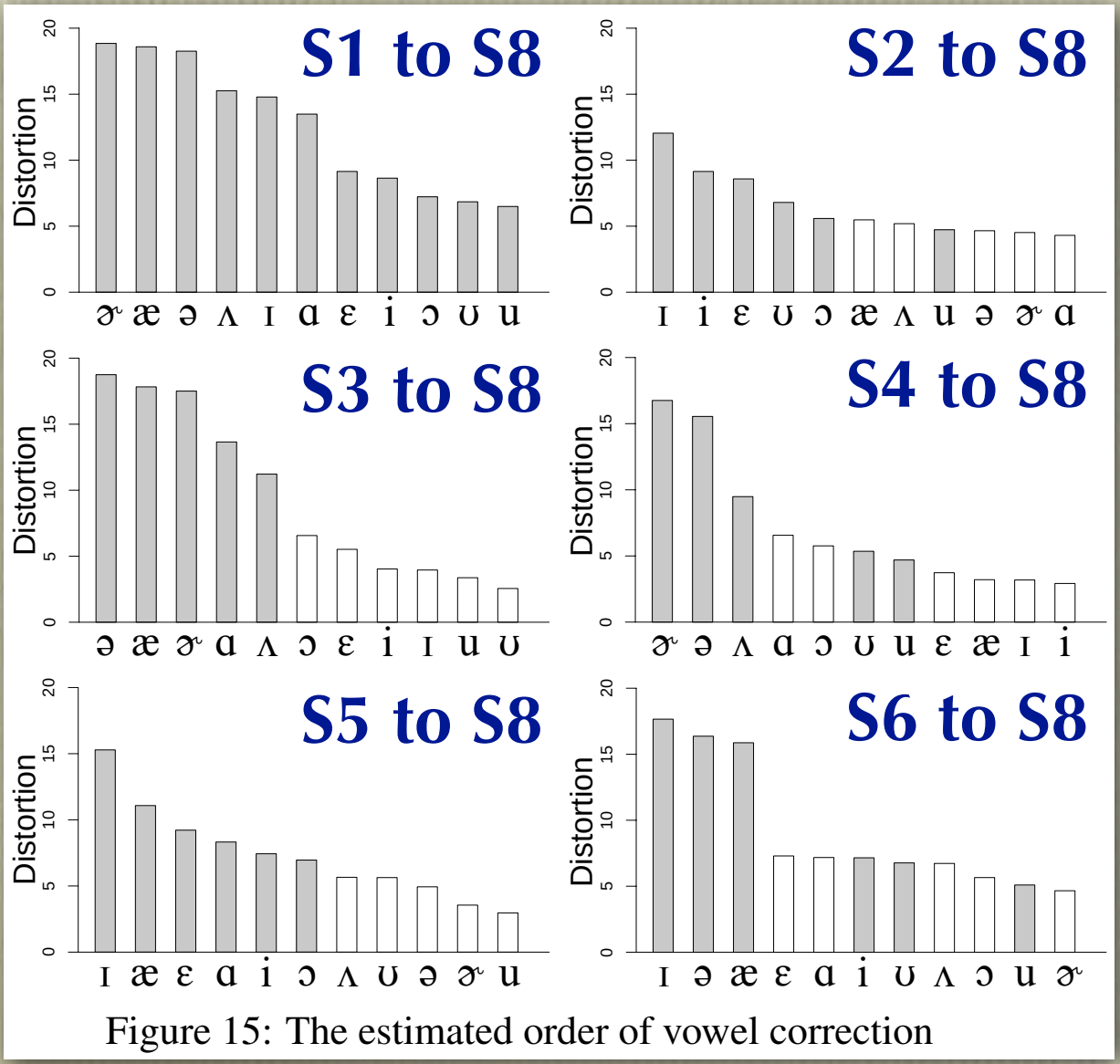
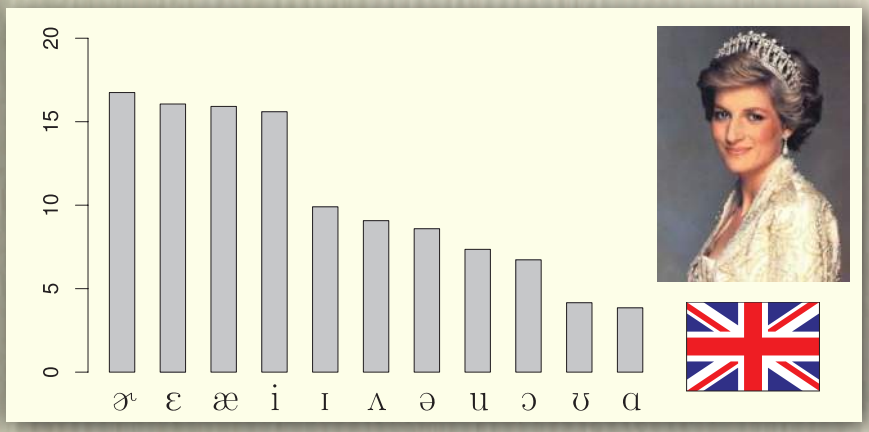


Figure 15: The estimated order of vowel correction

	ɑ	æ	ʌ	ə	ɚ	ɪ	i	ʊ	u	ɛ	ɔ
S1	J	J	J	J	J	J	J	J	J	J	J
S2	E	E	E	E	E	J	J	J	J	J	J
S3	J	J	J	J	J	E	E	E	E	E	E
S4	E	E	J	J	J	E	E	J	J	E	E
S5	J	J	E	E	E	J	J	E	E	J	J
S6	E	J	E	J	E	J	J	J	J	E	E
S7	J	E	J	E	J	E	E	E	E	J	J
S8	E	E	E	E	E	E	E	E	E	E	E

：日本語母音で置換

：置換無し

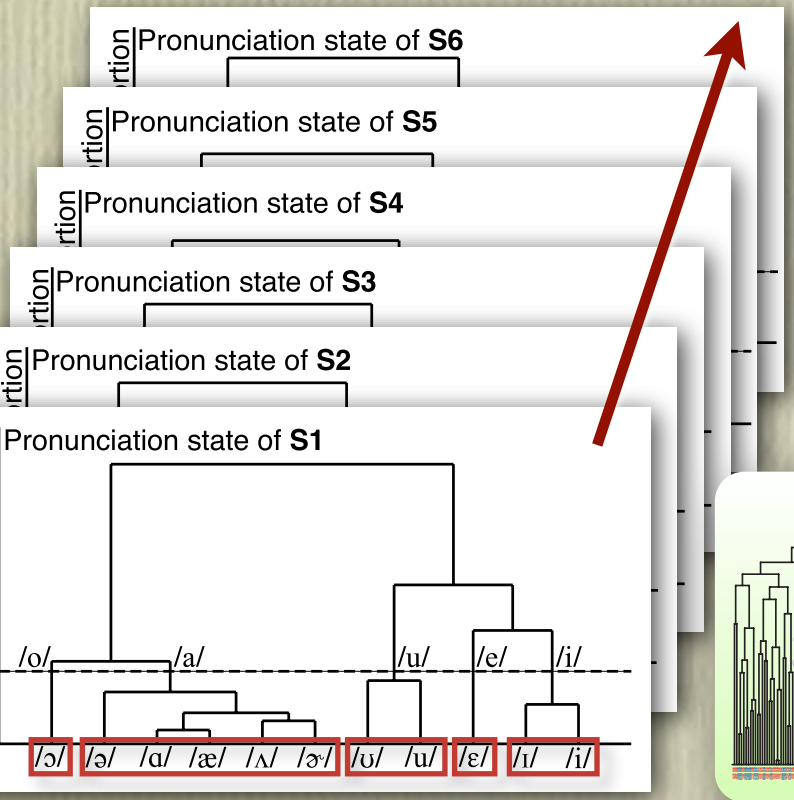


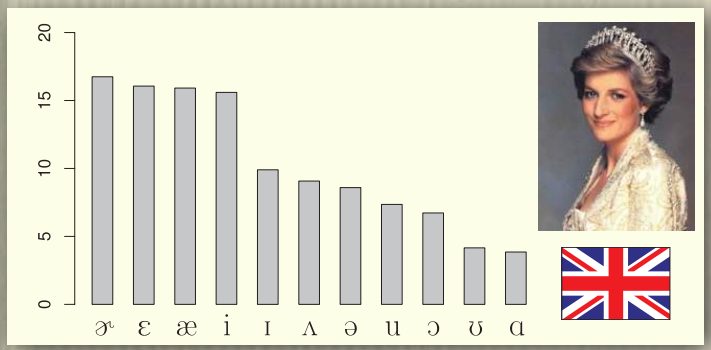
お手本となる相手を選ぶインタフェース

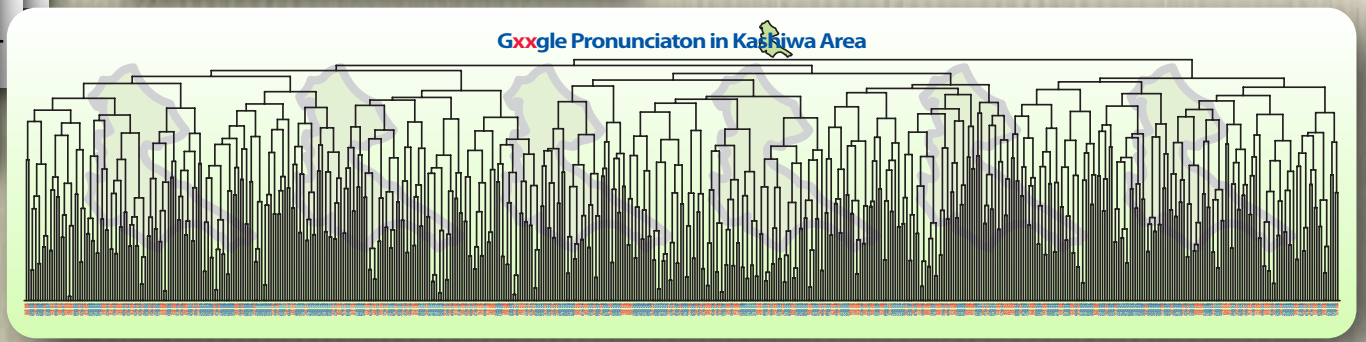
母音体系に見る発音習熟度

発音の構造表象に基づいて構築した技術・機能

- 学習者の母音体系がどのような状態にあるのかを記録する
- お手本の母音構造と比較して，矯正対象の母音を教示する
- 複数の教師からお手本としたい教師（相手）を選ぶ
- 年齢，性別などには影響されない学習者発音の分類







２通りの学習者分類

12人の帰国子女による96人の学習者シミュレーション

- 米語11母音＋日本語５母音を含む単語音声収録
  - beat, bet, bat, but, ... × 1回
  - ばと, びと, ぶと, ...　　× 5回
- 米語母音の一部を日本語母音で置換
- どの米語母音を置換するのか？ ８通りの置換パターン
  - 12人の帰国子女 x ８通りの置換パターン ＝ 96人の学習者
  - [A〜L] x [1〜8] ＝ 96個の 11x11 距離行列

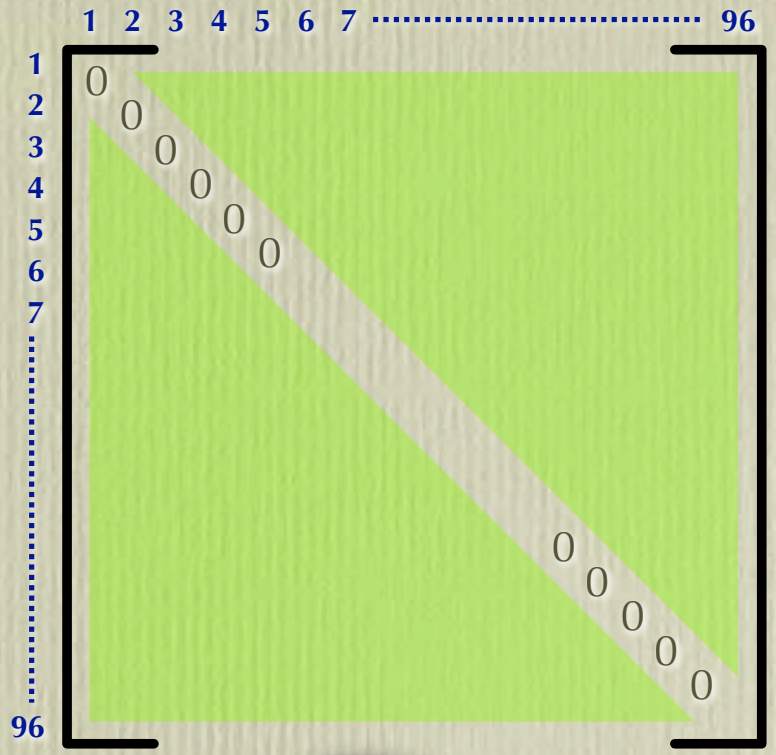
あ	ɑ, æ, ʌ, ə, ɚ
い	ɪ, i
う	ʊ, u
え	ɛ
お	ɔ

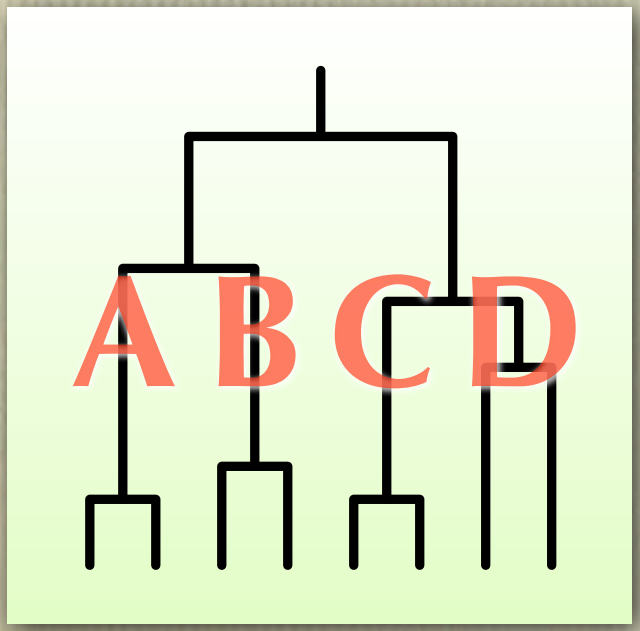
	ɑ	æ	ʌ	ə	ɚ	ɪ	i	ʊ	u	ɛ	ɔ
S1	J	J	J	J	J	J	J	J	J	J	J
S2	E	E	E	E	E	J	J	J	J	J	J
S3	J	J	J	J	J	E	E	E	E	E	E
S4	E	E	J	J	J	E	E	J	J	E	E
S5	J	J	E	E	E	J	J	E	E	J	J
S6	E	J	E	J	E	J	J	J	J	E	E
S7	J	E	J	E	J	E	E	E	E	J	J
S8	E	E	E	E	E	E	E	E	E	E	E

２通りの学習者分類

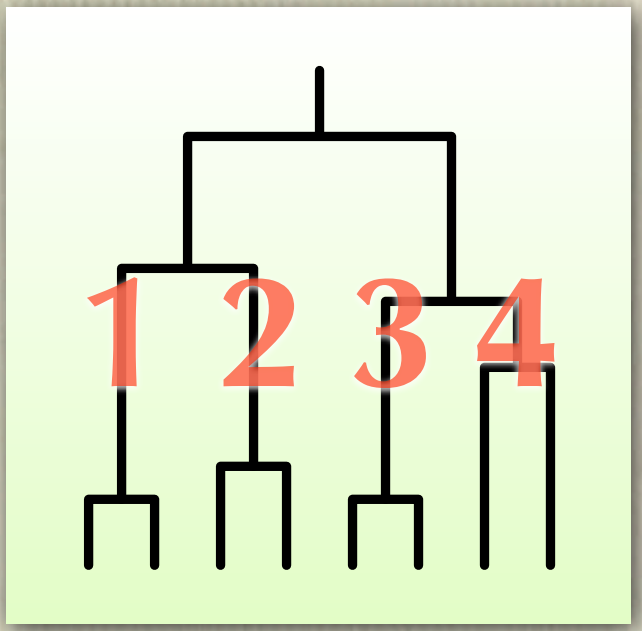
96 x 96 の距離行列に基づく学習者分類

- 話者：A～L
- 発音：1～8





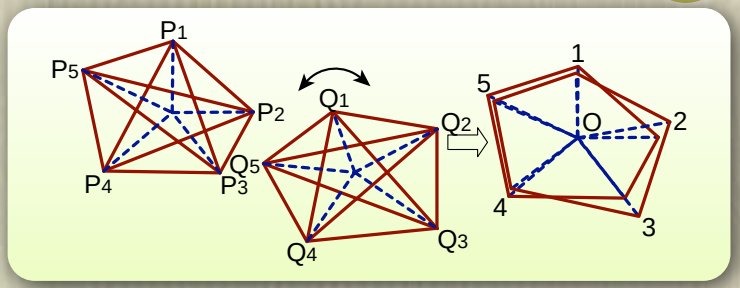
?

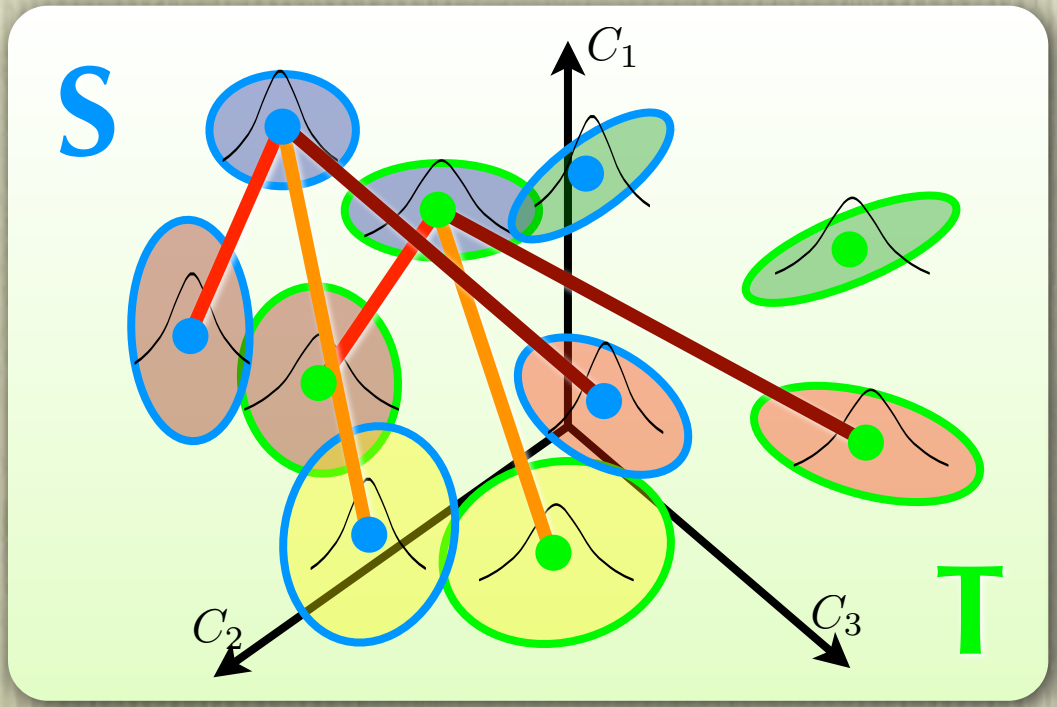


２通りの学習者分類

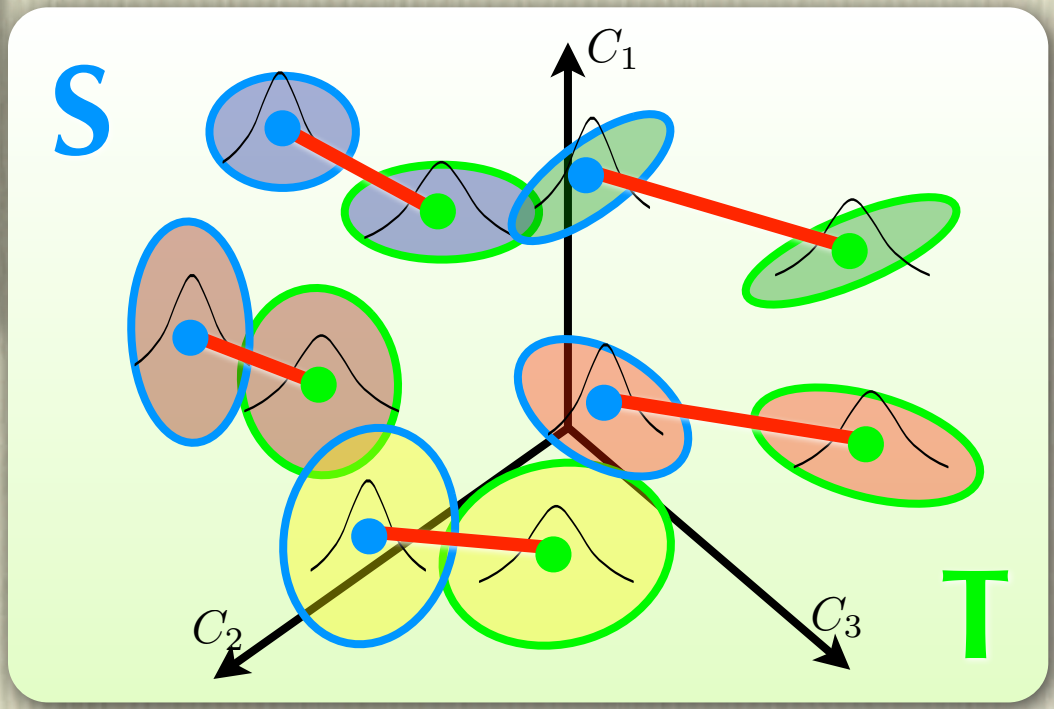
２通りの構造間距離

- コントラストの比較に基づく構造間距離
- 音の実体の比較に基づく構造間距離





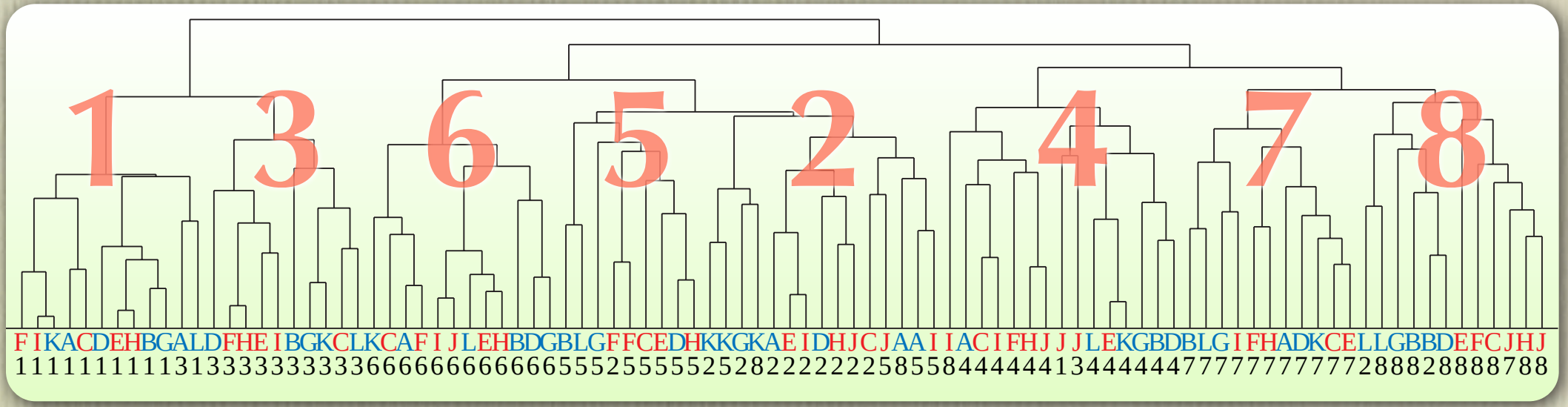
$$\sqrt{\frac{1}{M}\sum_{i<j}(S_{ij}-T_{ij})^2}$$



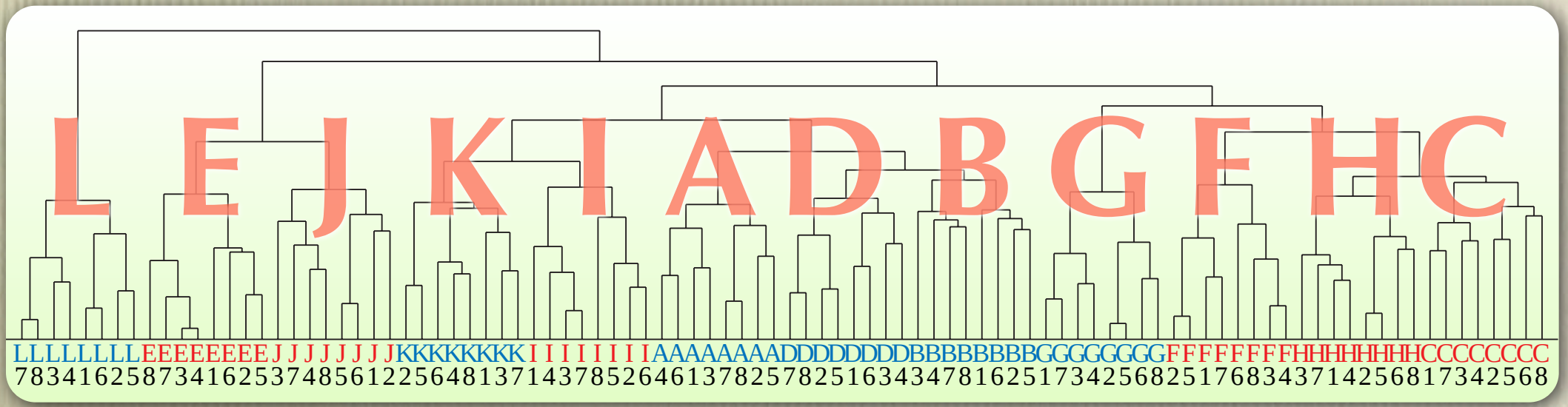
$$\sqrt{\frac{1}{M}\sum_{i} BD(v_i^S, v_i^T)}$$

２通りの学習者分類

コントラストの比較に基づく構造間距離

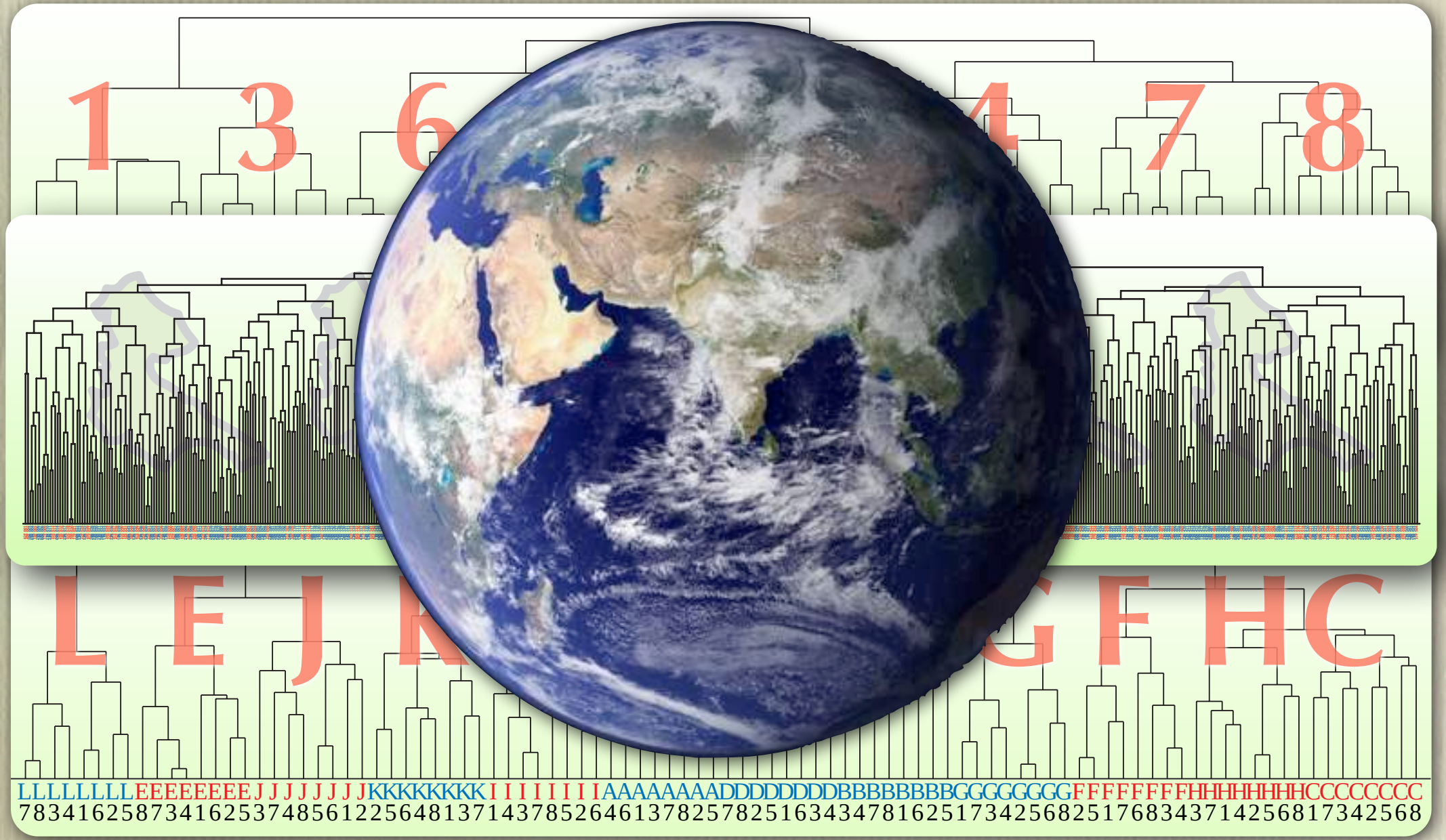


音の実体の比較に基づく構造間距離



２通りの学習者分類

コントラストの比較に基づく構造間距離



The current state of English

It is the only language used for global communication.

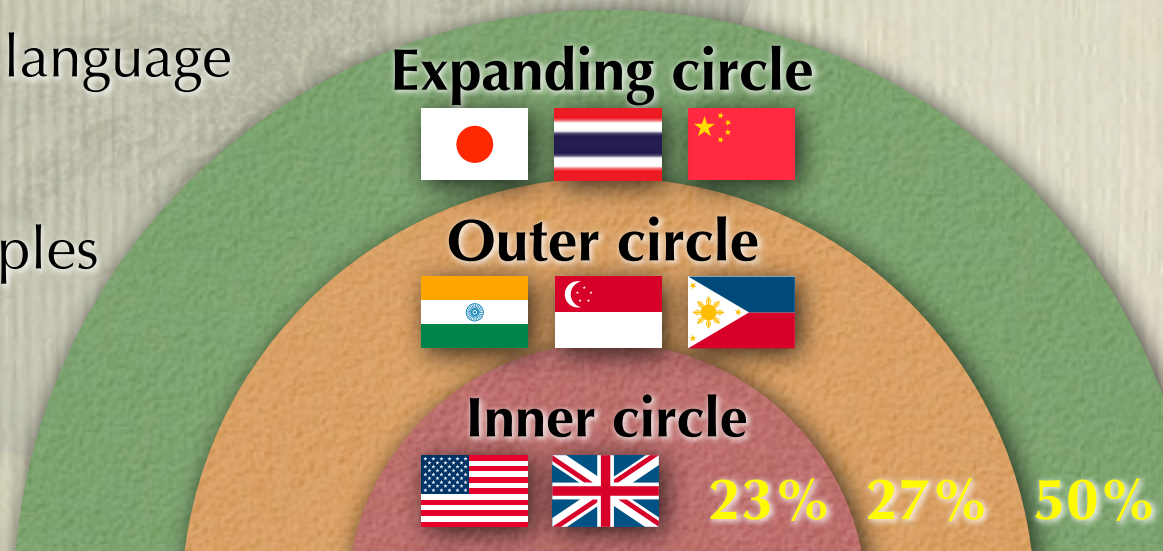
- About **1.5 billion** users on earth

It has the largest diversity in its form.

- Internationalization of a thing inevitably alters its form.
- English is not exceptional.
  - Syntax, pragmatics, lexical choice, spelling, **pronunciation**, etc

World Englishes (WE)

- Three circles model [Kachru1992]
  - E as native / official / foreign language
- No standard pronunciation
  - AE and BE are just two examples of **accented** Engilshes.



Pronunciation diversity of WE

Is English a useful tool or a troublesome tool?

- A useful tool for global communication
  - The same language can be shared by all.
- A troublesome tool for global communication
  - Its pronunciation diversity can cause miscommunications.

http://academicaffairs.ucdavis.edu/diversity/

Diversity of pronunciation in WE

What is the minimal unit and how many units?

Country?

↓

Region / State / Prefecture?

↓

City / Town / Village?

↓

Individual!

[Kachru 1992]

Diversity of pronunciation in WE

What is the minimal unit and how many units?

Country?

↓

Region / State / Prefecture?

↓

City / Town / Village?

↓

Individual!

↓

1.5 billions!

[Kachru 1992]

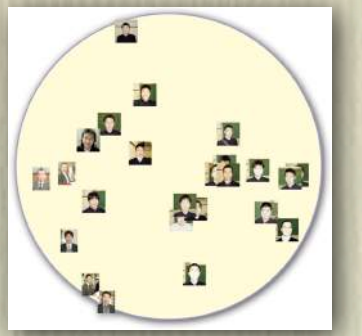
Huge pron. diversity in World Englishes

1) native language, 2) official langua

Speaker-basis pronunciation clustering

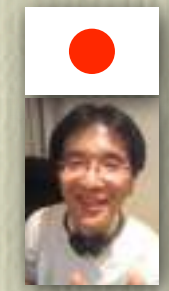
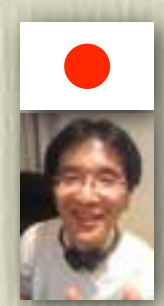
Requires a speaker-basis pronunciation distance matrix

$$\begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ d_{31} & & & \\ \vdots & & & \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{bmatrix}$$



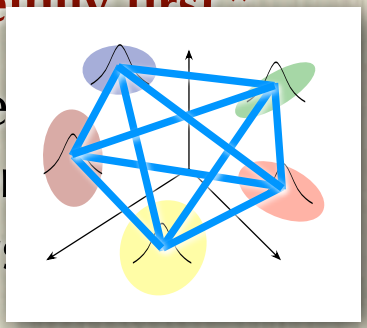
What is technically challenging?

- To which is Minematsu's natural pronunciation closer?

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

- Pronunciation distance = phonetic distance between speake
  ≠ acoustic distance between speake
  ≠ spectral distance between speaker



Pron. clustering using real data of WE

Speech Accent Archive (SAA) [Weinberger'13]

- A common paragraph read by about 1.8K international speakers
  - The paragraph is designed to achieve high phonemic coverage of AE.
- Speech samples and their narrow IPA transcripts are provided.

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.

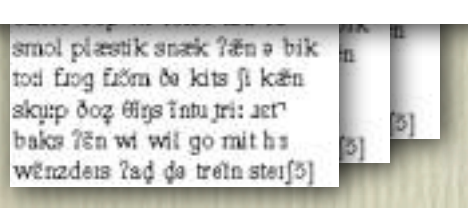
[plis kol stelə ask hɨ tu bɹiŋ
di: θiŋks wiθ hɨ fɹʊm ðə stoʊ
siks spʊnz of fɹɛʃ snoʊ piːz
faɪv θɪkˀ θɪkʰ slæps oʊv blu
ʧiz ?æn mɛbi ə snæk fɔ hɨz
bɹʌðə bɑp wi ?oːlsoʊ nid ?ə
smɔl plæstɪk sneɪk ?æn ə bɪg
toʊ fɹɑg fɹʊm ðə kɪds ʃi kæn
skuːp ðoʊz θiŋz ɪntu θɹiː ɹɛdˀ
bæɡz ?æn wi wil go mit hɨz
wɛnzdeɪ ?aɾ ðə tɹeɪn steɪʃ͡ə]

Pron. clustering using real data of WE

Speech Accent Archive (SAA) [Weinberger'13]

- A common paragraph read by about 1.8K international speakers

[pʰlis kɔl stɛːlʌ ɑskˀ ɜ tə bɹĩĩŋ ðiz θĩŋz̩ wɪf hɜ fɹʌ̃mŋ ðə stɔɹ siks spunz̩ əy̩ fɹɪʃ sn̩oʊ piːs faɪf θɪk slæb̩s əv blu ʧiːz ɛn mă̈ɪbi ɜ snækˀ foɻ hɜ bɹɑɾə̆ ʔə brʌðə bɑp wi ɔlʸsŏ nid ə smɔlʸ plæstɪk sneɪk ɛn ə bɪk tʊi fɹɔg̃ fɛ̆ ðə kidz̩ ʃi kĕn skøp ðiz θĩŋs ɪntu fɹi ɹɛd bægz̩ ɛn wi wɪl goʊ miːd ɜ̆ wĕnz̩d̩eɪ ɛt d̪ə tɹeɪn steɪʃən]



Pron. … of WE

Speech Ac…

- A commo… …nal speakers
  - The para… …verage of AE.
- Speech sa… …provided.

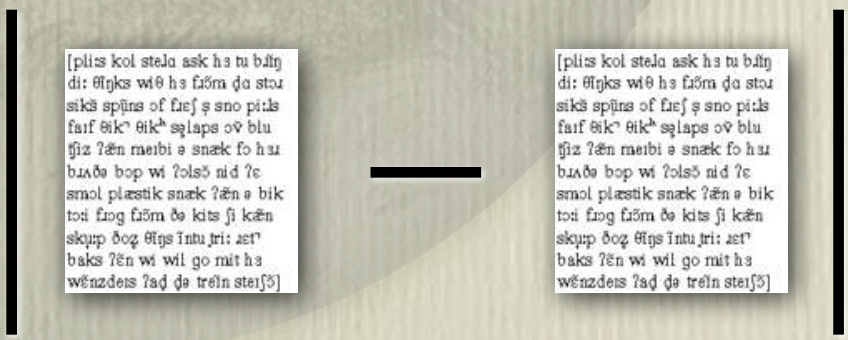
Please call Stella. Ask … the store: Six spoons o… cheese, and maybe a s… small plastic snake an… scoop these things into… Wednesday at the train

Vowels and Consonants used in Acoustic Analysis					
1. i	2. ɪ̈	3. iː	4. i̯	5. ï	6. ĩ
7. y	8. ɪ̯	9. ɪ	10. ɪː	11. ɪ̯	12. ɪ̃
13. e	14. ë	15. ẽ	16. ɛ	17. ɛ̈	18. ɛ̃
19. æ	20. æ	21. æː	22. æ̃	23. a	24. ã
25. ɨ	26. ɨ̯	27. ɨ̃	28. ʉ	29. ʉ̯	30. ɞ
31. ɜ	32. ɜ	33. ɐ	34. ɐ̃	35. ʉ̃	36. ɵ
37. ɵ̃	38. ə	39. ə̆	40. ə̯	41. ə̃	42. ə̯
43. ɯ	44. ɯ̈	45. ɯ̃	46. u	47. ŭ	48. uː
49. ü	50. ũ	51. ũː	52. ʊ	53. ɤ	54. o
55. ö	56. õ	57. ʌ	58. ʌ̃	59. ɔ	60. ɔː
61. ɔ̈	62. ɔ̃	63. ɑ	64. ɑː	65. ä	66. ã
67. p	68. pʰ	69. p̄	70. b	71. b̄	72. b̯
73. ɸ	74. β	75. β̯	76. β̃	77. f	78.v
79. ɣ	80. ʋ	81.m	82. m̯	83. m̩	84.n
85. n̯	86. n̩	87. ɳ	88. ɲ	89. ŋ	90. ɴ
91.t	92. tʰ	93. t̪	94. t̩	95. t’	96. t̄
97. d	98. d̯	99. d̄	100. d̩	101. s	102. s̪
103. sʲ	104. z	105. z̪	106. ɹ	107. ɻ	108. ɻ̩
109. r	110. ɾ	111. ɾ̩	112. l	113. l̩	114. lʴ
115. θ	116. ð	117. ɕ	118. ʑ	119. ʑ̩	120. ʃ
121. ʒ	122. ç	123. j	124. j̩	125. k	126. kʰ
127. k̪	128. k’	129. k̪ʰ	130. k̄	131. g	132. g̩
133. ḡ	134. g̊	135. x	136. ɣ̩	137. ɣ	138. ɯ̩
139. ʔ	140. h	141. ɦ	142. w	143. ɥ	144. pɸ
145. tθ	146. dð	147. ts	148. dz	149. tɕ	150. dʑ
151. tʃ	152. dʒ	153. kx			

Pron. clustering only based on SAA

N speakers

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & N \end{array} \\ \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & \dots & d_{2N} \\ d_{31} & & & \\ \vdots & & & \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{bmatrix} \end{array}$$

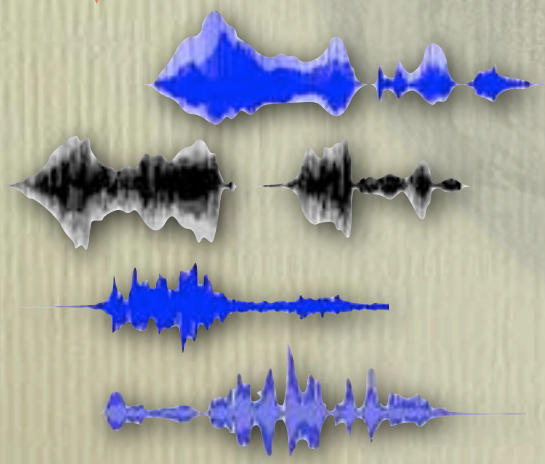


[Miller et al.'95, Bailey et al.'05, Wieling et al.'12]

Pron. clustering only based on SAA

N speakers

$$\{d_{mn}\} \approx \{p_{mn}\} \ ?$$

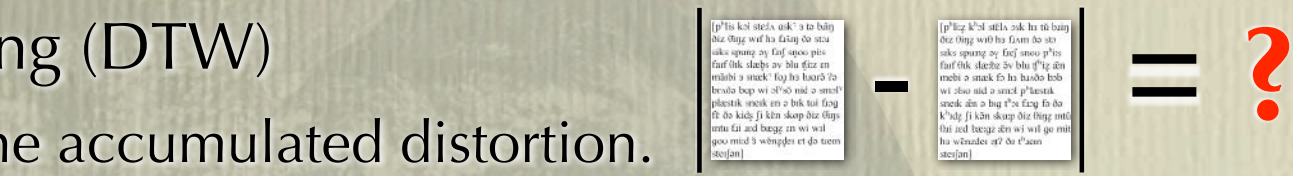


Pron. Structure Analysis

IPA-based reference pron. distance

Optimal alignment bet. two transcripts [Shen et al.,'13]

- Dynamic Time Warping (DTW)
  - DTW can minimize the accumulated distortion.

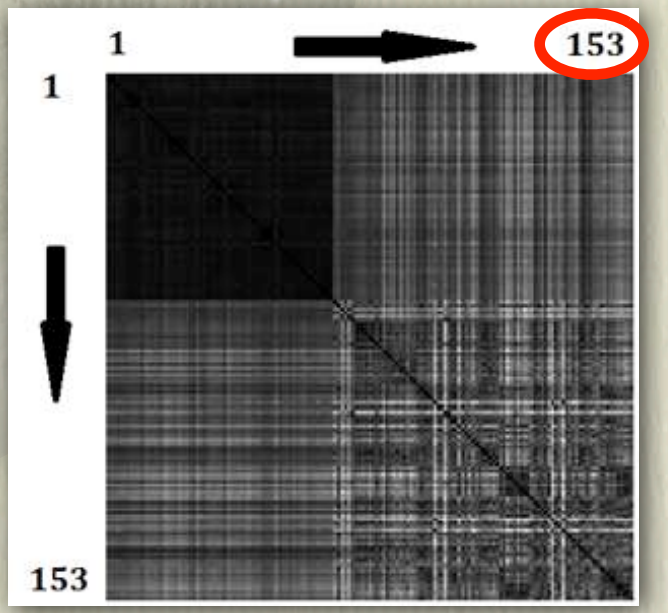


```
p_h l i z   k A l     s t E l a
 |  | | |   | | |     | | | | |
p_h l i s   k O l_G   s t E l V
```

```
b l @ u
| | /
b l u
```

```
? O l s o
\ \ \ \
A s o
```

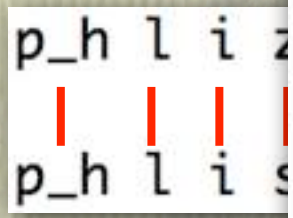
  - Similar to edit-distance-based alignment of transcripts [Wieling et. al,'12]
- DTW requires a distance matrix of all the 153 IPA symbols used.
  - 20 productions for each by a phonetician
  - HMM is built for each symbol (SD-HMM)
    - HMM = Hidden Markov Model
  - Acoustic distance is obtained from each HMM (phone) pair.

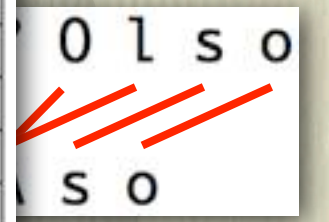


IPA-b…tance

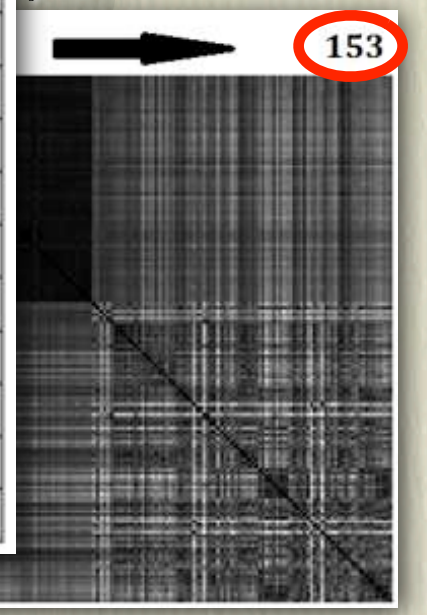
Optimal al… …et al.,'13]

- Dynamic T…
  - DTW can…





  - Similar to… …Wieling et. al,'12]
- DTW requ… …ymbols used.
  - 20 produ…
  - HMM is b…
    - HMM …
  - Acoustic … each HM…

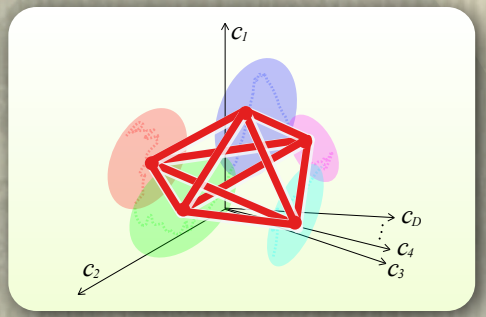


Vowels and Consonants used in Acoustic Analysis					
1. i	2. ɪ̈	3. iː	4. i̯	5. ï	6. ĩ
7. y	8. ɪ̯	9. ɪ	10. ɪː	11. ɪ̯	12. ĩ
13. e	14. ë	15. ẽ	16. ɛ	17. ë	18. ɛ̃
19. æ	20. æ	21. æː	22. æ̃	23. a	24. ã
25. i̯	26. i̯	27. ĩ	28. ʉ	29. ʉ̯	30. ɚ
31. ɜ	32. ɝ	33. ɐ	34. ɐ̃	35. ʉ̃	36. ɵ
37. ɵ̃	38. ə	39. ə̃	40. ə̯	41. ə̃	42. ə̯
43. ɯ	44. ɯ̈	45. ɯ̃	46. u	47. ŭ	48. uː
49. ü	50. ũ	51. ũː	52. ʊ	53. ɤ	54. o
55. ö	56. õ	57. ʌ	58. ʌ̃	59. ɔ	60. ɔː
61. ɔ̃	62. ɔ̃	63. ɑ	64. ɑː	65. ɑ̈	66. ɑ̃
67. p	68. pʰ	69. p̄	70. b	71. b̄	72. b̯
73. ɸ	74. β	75. β̞	76. β̃	77. f	78.v
79. ɣ	80. ʋ	81.m	82. m̥	83. m̩	84.n
85. n̥	86. n̪	87. ɳ	88. n̩	89. ŋ	90. ɴ
91.t	92. tʰ	93. t̪	94. t̩	95. t'	96. t̄
97. d	98. d̪	99. d̄	100. d̩	101. s	102. s̪
103. sʲ	104. z	105. ʐ	106. ɹ	107. ɻ	108. ɻ
109. r	110. ɾ	111. ɾ	112. l	113. l̩	114. lˠ
115. θ	116. ð	117. ɕ	118. ʑ	119. ʑ	120. ʃ
121. ʒ	122. ç	123. j	124. j̄	125. k	126. kʰ
127. k̟	128. k'	129. k̟ʰ	130. k̄	131. g	132. g
133. ḡ	134. g̊	135. x	136. ɣ	137. ɣ	138. ɯ̟
139. ʔ	140. h	141. ɦ	142. w	143. ɥ	144. pɸ
145. tθ	146. dð	147. ts	148. dz	149. tɕ	150. dʑ
151. tʃ	152. dʒ	153. kx			

Pron. distance calculation using structure

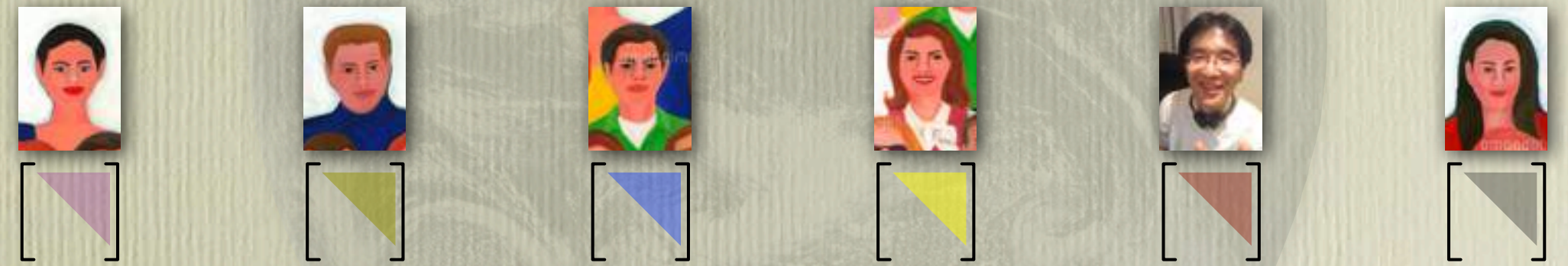
A common paragraph to pron. structure

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 .....



221

221



IPA-based distance

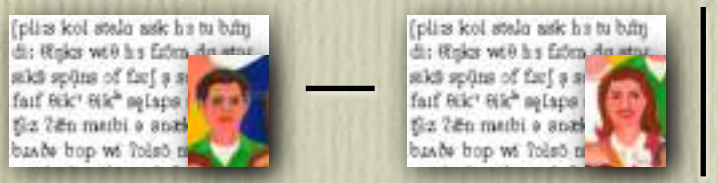
1.5 billions

1.5 billions

Pron. clustering using real data of WE

Use of IPA transcripts to prepare reference distances

- DTW-based calculation of the reference distance bet. transcripts



Prediction of the ref. distances using pron. structures

- SVR-based supervised prediction using structures as input features

Use of phonemic transcripts to calculate distances

- Corresponds to calculate pron. distances somewhat coarsely.

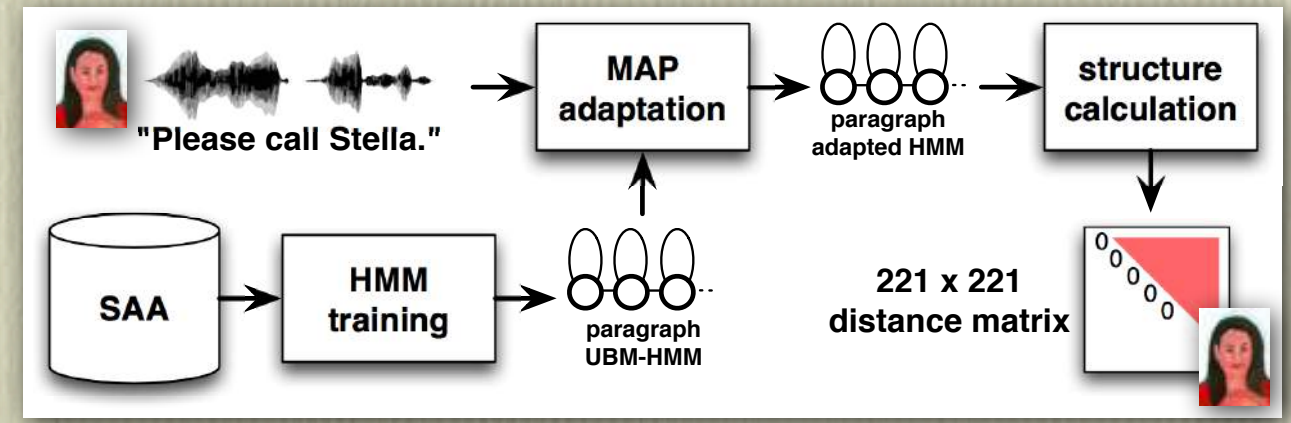
/ pəliːz̩ kəlʸ stɪlə ǽsk hɜɹ tʷǒ bɹɪ̃ːŋ / #symbols = 153

[p ah l iy z k ao l s t ih l ah ae s k hh ah r t ow b r ih ng] #symbols = 39

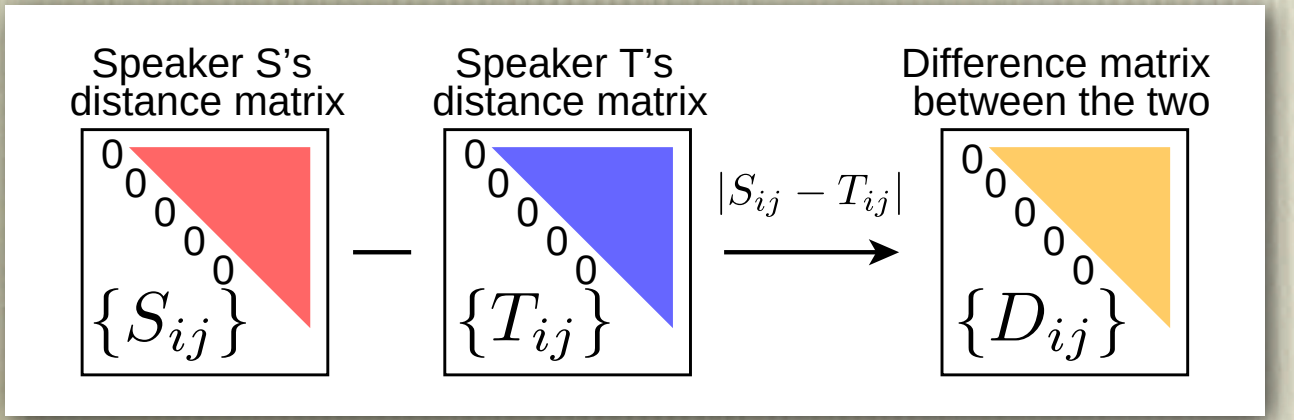
Pron. clustering using real data of WE

SVR-based prediction of IPA distances [Kasahara'14]

- Pronunciation structure extraction from an SAA sample



- Differential features from two pronunciation structures

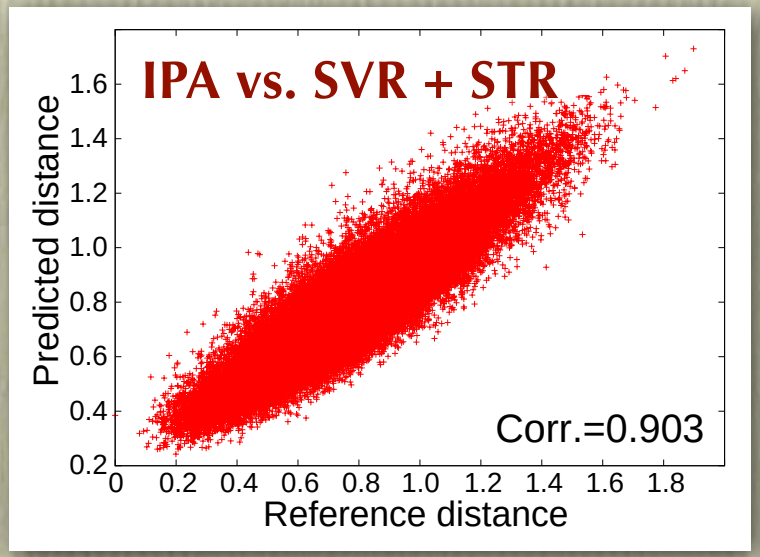


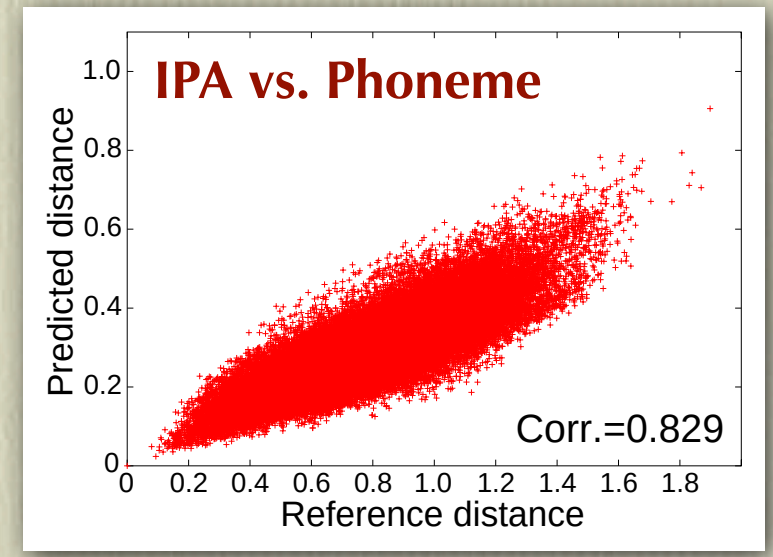
Pron. clustering using real data of WE

Three kinds of distances

- IPA-based distances
- SVR-based predicted distances
- Phoneme-based distances

Assessment in terms of correlation



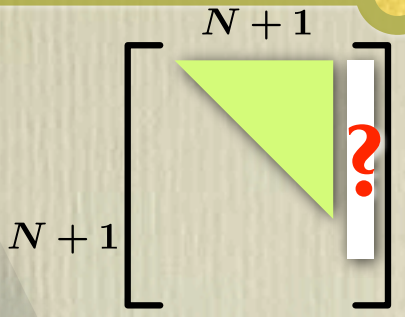


- IPA phone : minimum unit of sounds perceived by phoneticians
- Phoneme : minimum unit of sounds perceived by general listeners
- Our method can estimate distances better than general listeners.

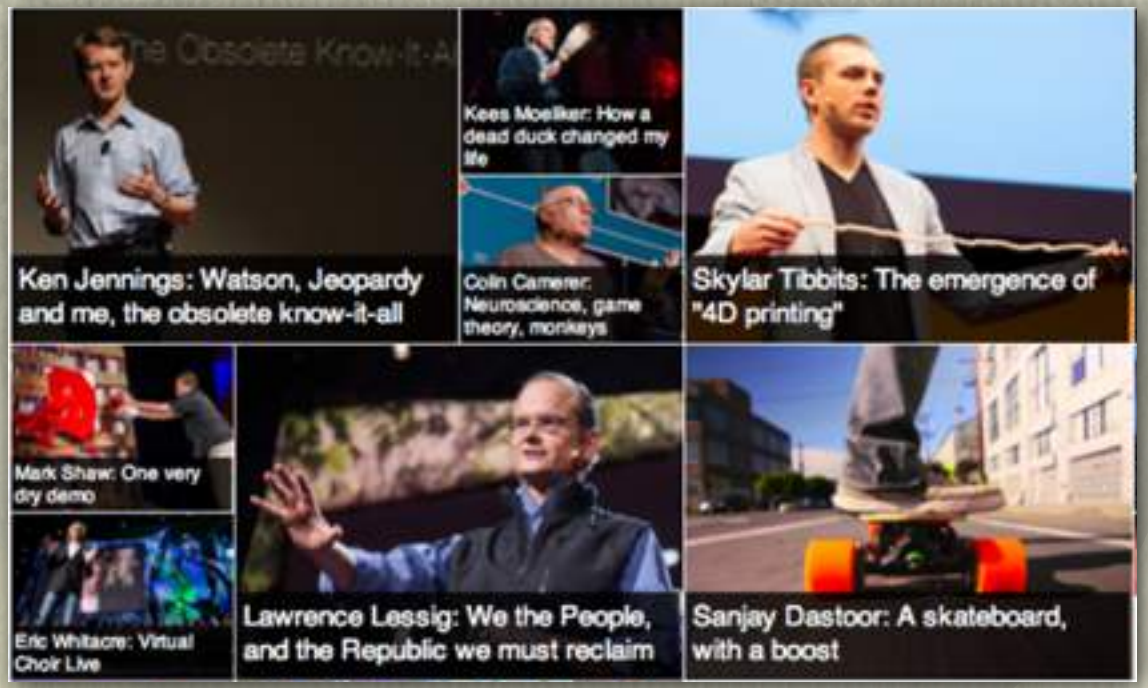
Application of speaker-pair-open prediction

TED talks browser from your viewpoint

- If TED talkers provide their SAA readings....
- If these readings are transcribed by phoneticians....



Visualization of pronunciation diversity [Kawase et al.,'14]

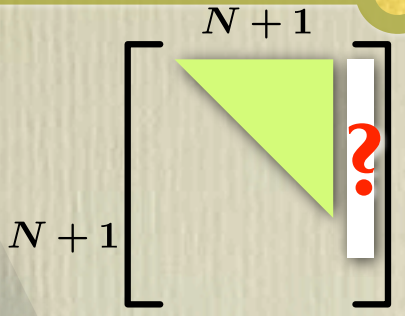


TED

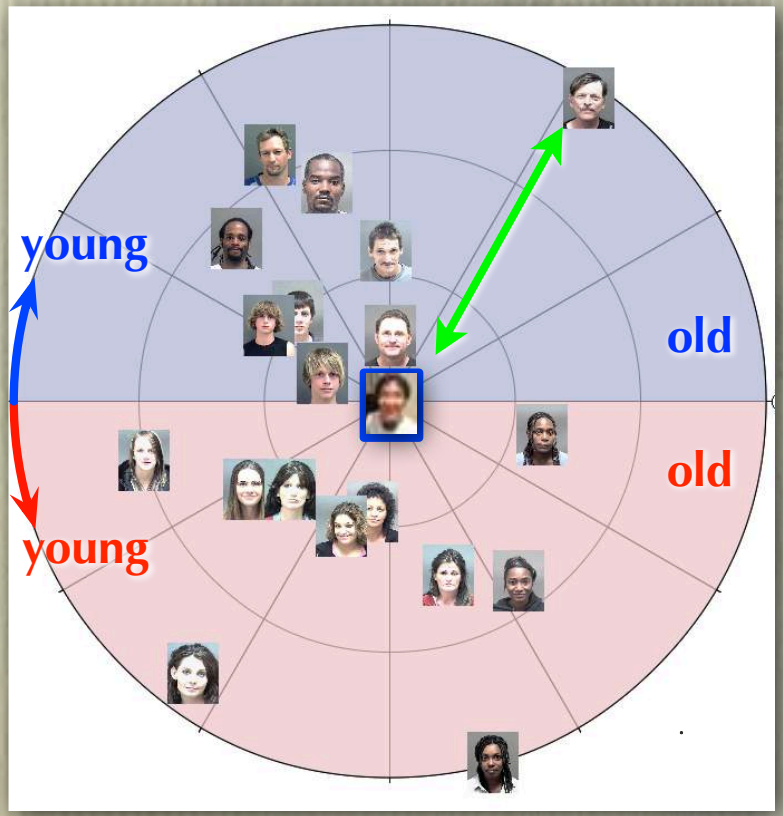
Application of speaker-pair-open prediction

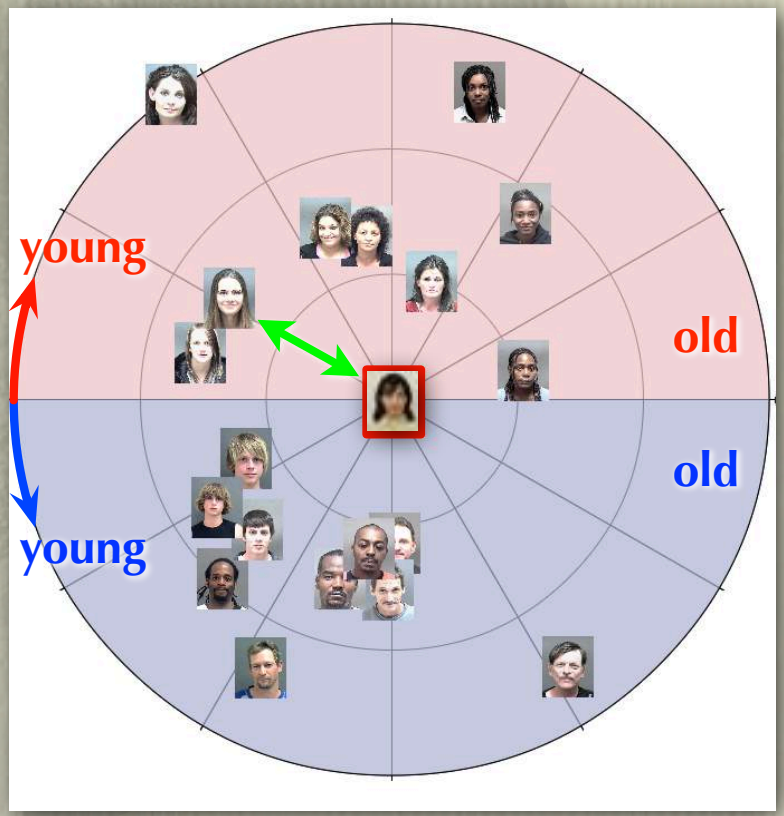
TED talks browser from your viewpoint

- If TED talkers provide their SAA readings....
- If these readings are transcribed by phoneticians....



Visualization of pronunciation diversity [Kawase *et al.*,'14]

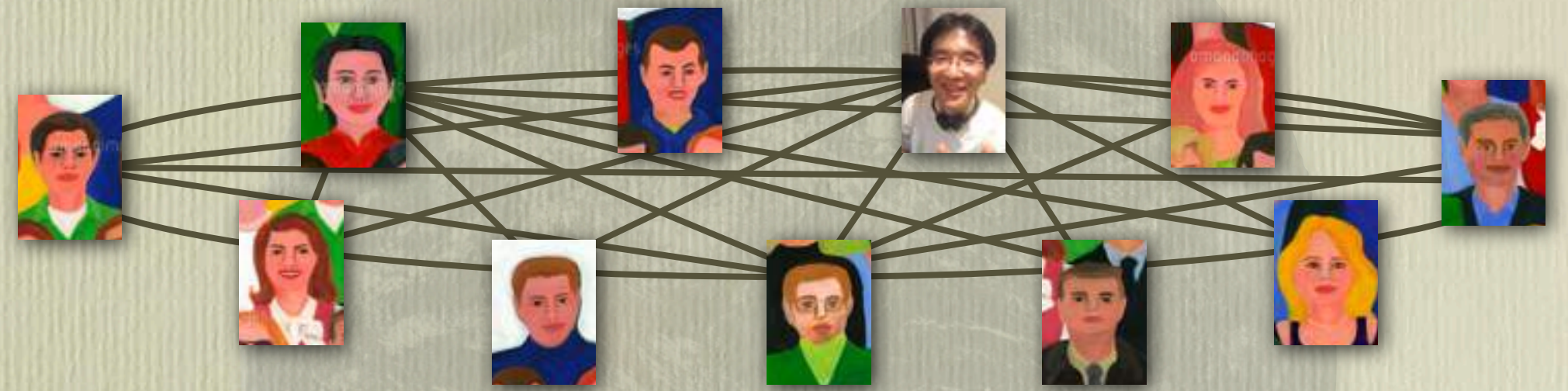




Y. Kawase, et al., "Visualization of pronunciation diversity of World Englishes from a speaker's self-centered viewpoint"

Possible application of spk-open prediction

Individual-based and really global map of WE pron.



Can be used for WE communication facilitator

- Easy access to speakers with pronunciation similar to yours

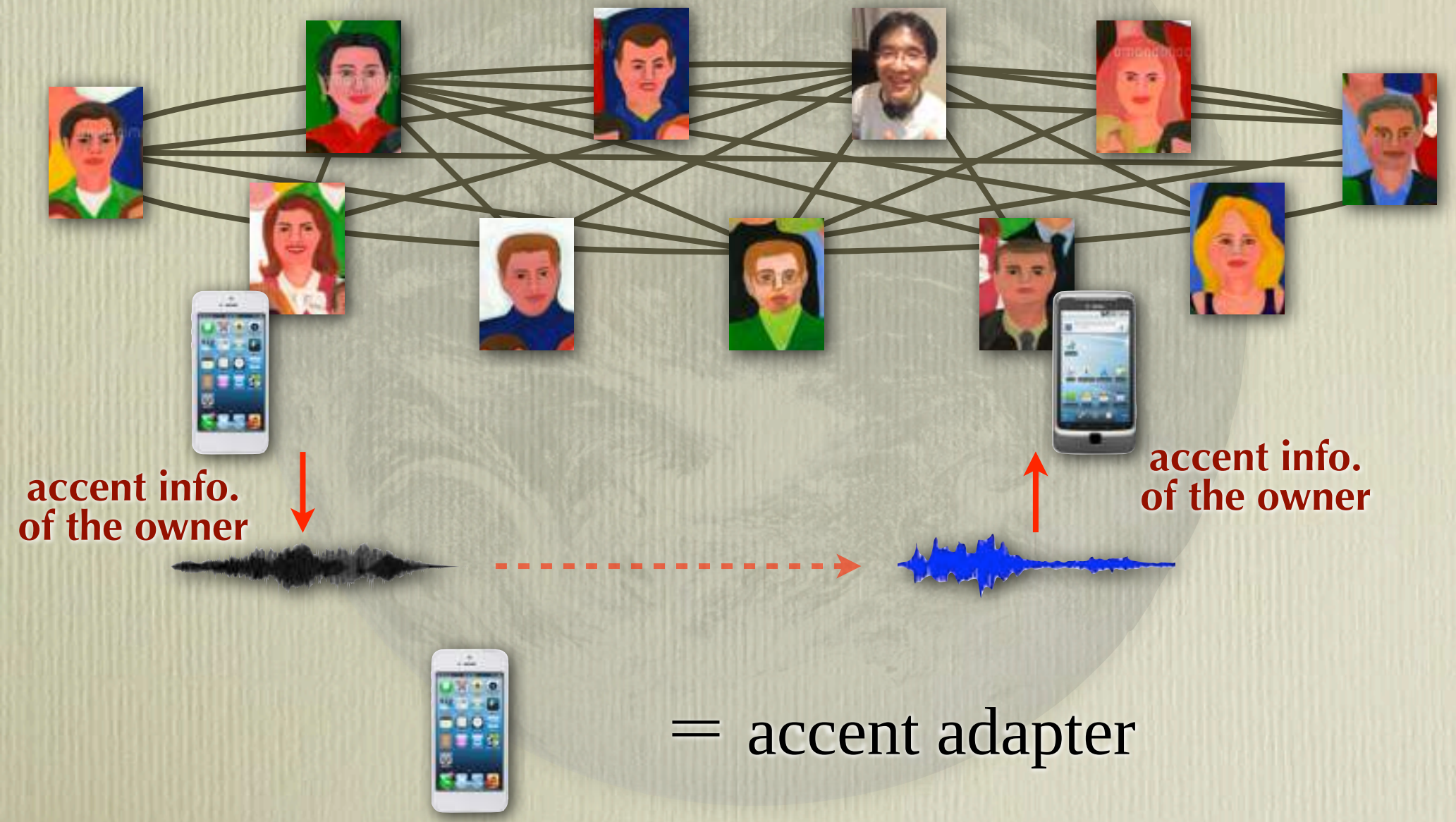
at a restaurant

at a hotel

at a ticket office

Possible application of spk-open prediction

Adaptation of others' accents to a listener's own accent



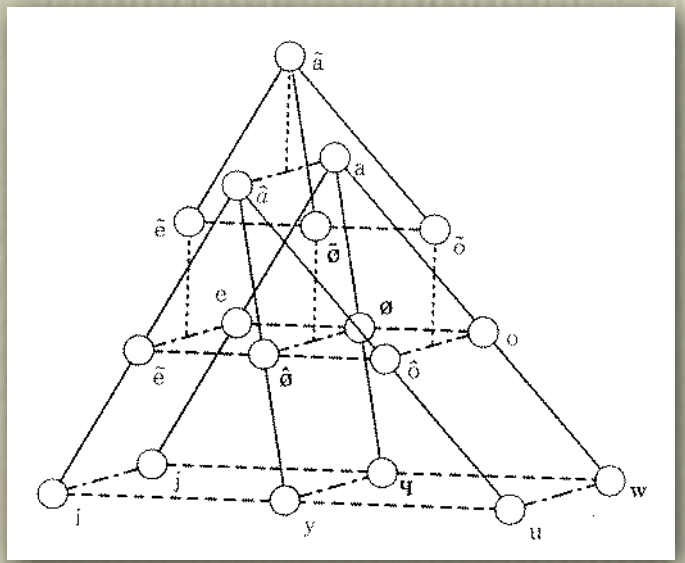
古典的音韻論に見られる主張

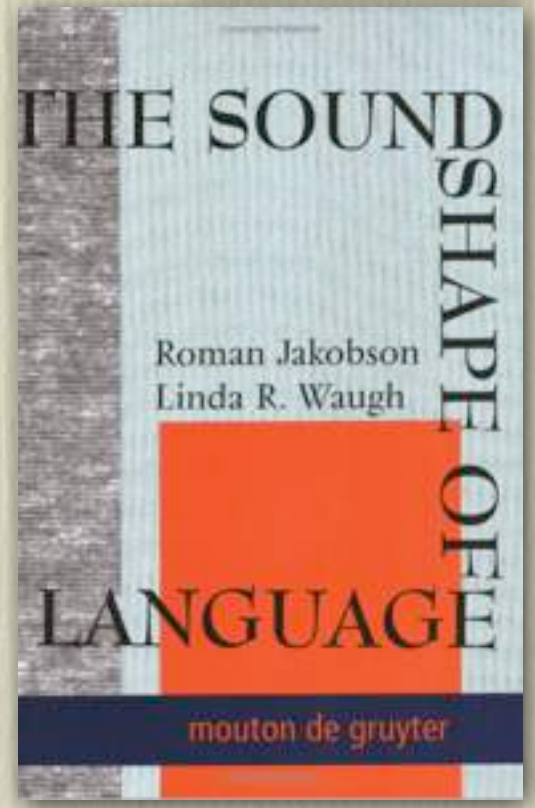
Roman Jakobson (1896-1982)

- The sound shape of language (1949)

Physiologically identical sounds may possess different values in conformity with the whole sound system, i.e. with their relations to the other sounds.

We have to put aside the accidental properties of individual sounds and substitute a general expression that is the common denominator of these variables.

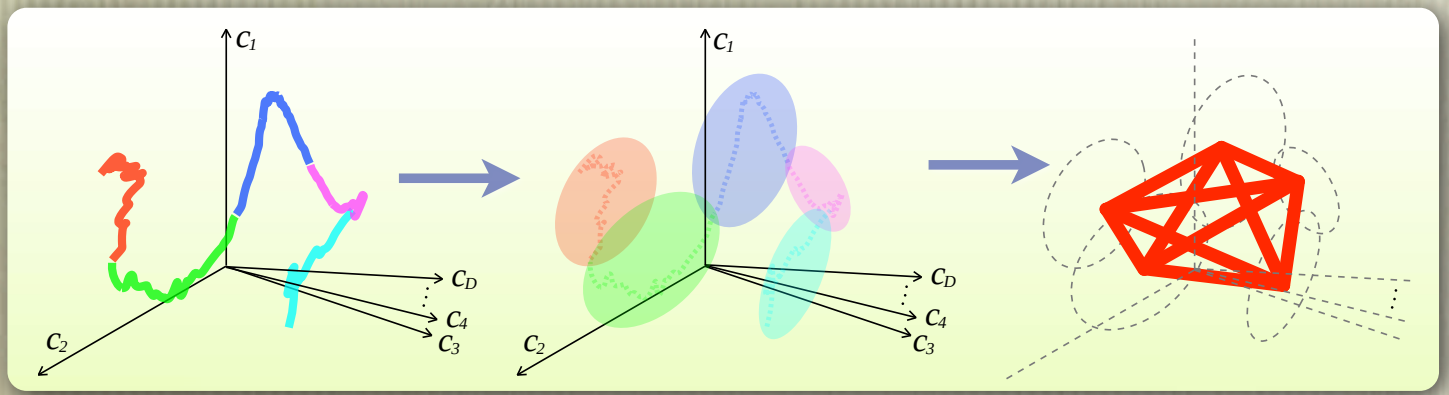


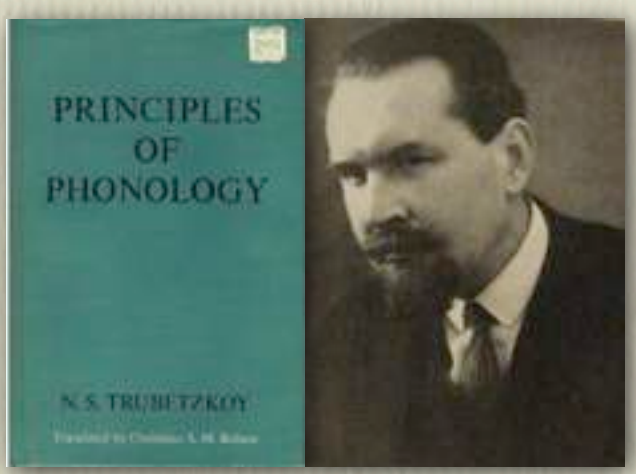


古典的音韻論に見られる主張

Nikolay Trubetskoy (1890-1938)

- The Principle of Phonology (1939)
- The phonemes should not be considered as building blocks out of which individual words are assembled. Each word is a phonic entity, a Gestalt, and is also recognized as such by the hearer.
- As a Gestalt, each word contains something more than sum of its constituents (phonemes), namely, the principle of unity that holds the phoneme sequence together and lends individuality to a word.

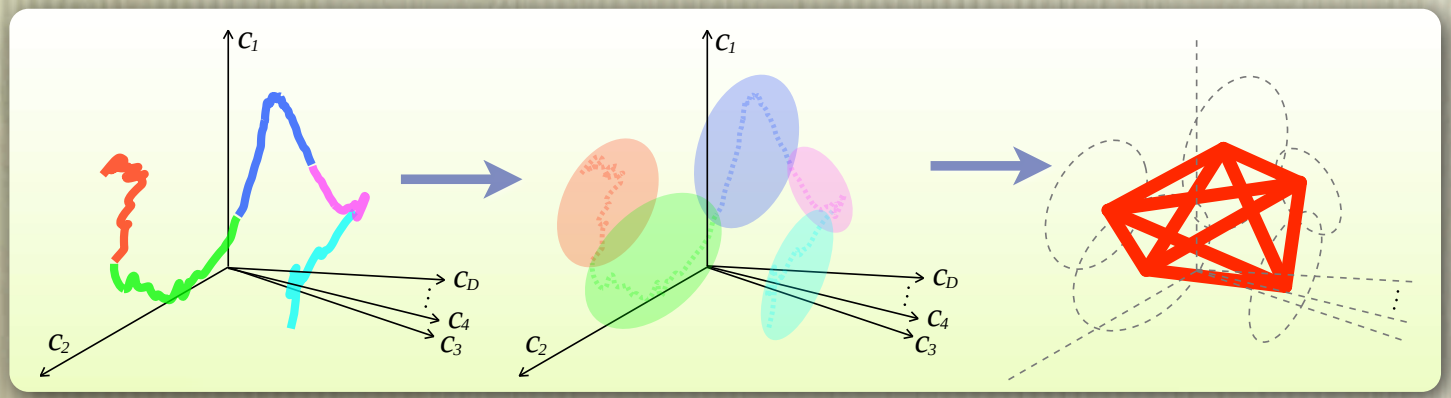


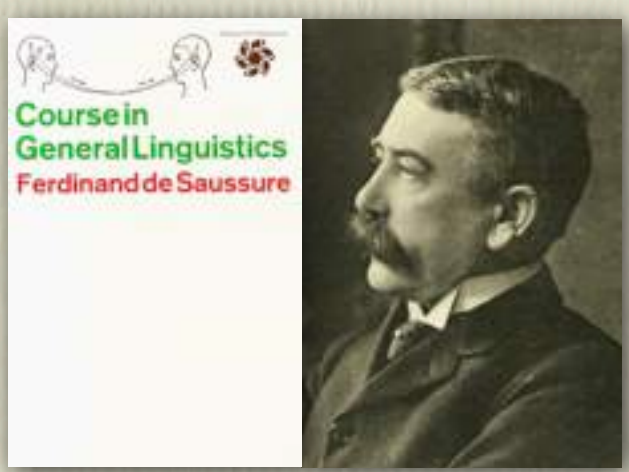


古典的音韻論に見られる主張

Ferdinand de Saussure (1857-1913)

- Father of modern linguistics
- Course in General Linguistics (1916)
- What defines a linguistic element, conceptual or phonic, is the relation in which it stands to the other elements in the linguistic system.
- The important thing in the word is not the sound alone but the phonic differences that make it possible to distinguish this word from the others.
- Language is a system of only conceptual differences and phonic differences.

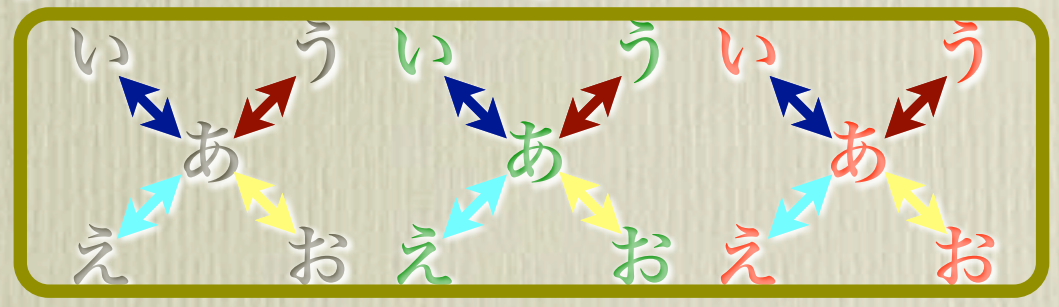




「あ」って何だろう？

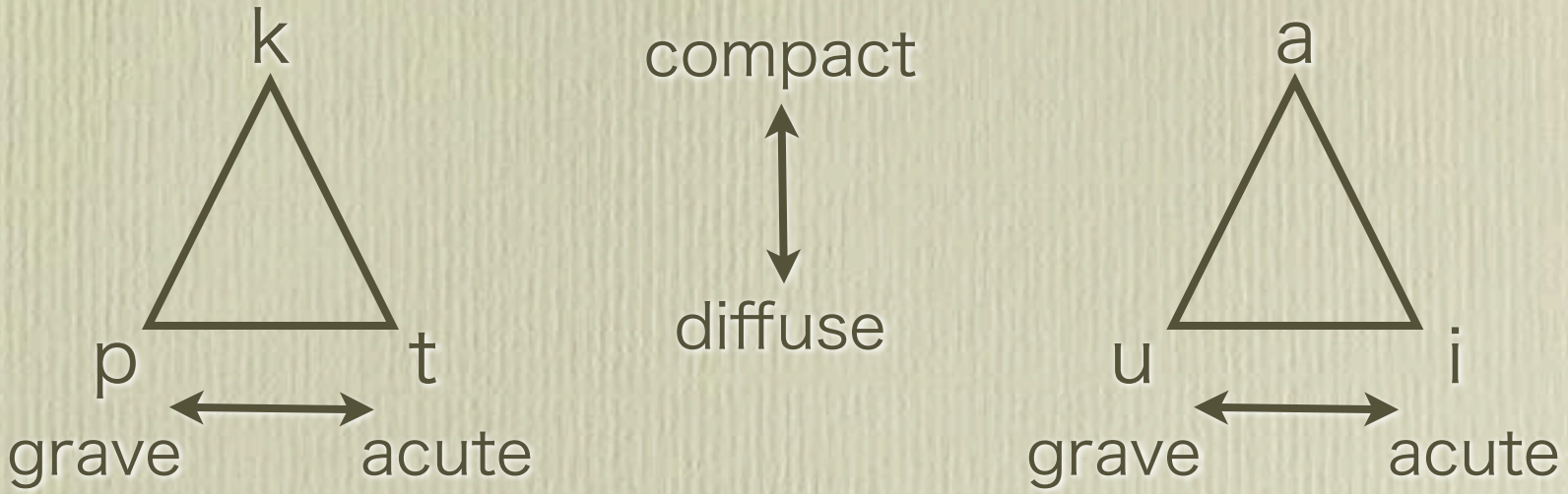
二つの見方, Element first or system first?

- An introduction to descriptive linguistics
  - Written by Gleason, H. A, 1961
  - A phoneme is a class of sounds that: (1) are phonetically similar and (2) show certain characteristic patterns of distribution in the language or dialect under consideration.
  - A phoneme is one element in the sound system of a language having a characteristic set of interrelations with each of the other elements in that system.
  - The phoneme cannot, therefore, be acoustically defined. The phoneme is a feature of language structure. That is, it is an abstraction from the psychological and acoustical patterns that enables a linguist to describe the observed repetitions of things that seem to function within the system as identical in spite of obvious differences. The phonemes of a language are a set of abstractions.



ヤコブソンの弁別素性

弁別素性を用いた音的差異・音体系の記述



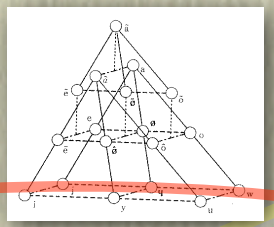
弁別素性を用いた音素の記述

	p	b	m	f	v	θ	ð	t	d	n	s	z	l	ɾ	ʃ	ʒ	ʧ	ʤ	j	ɹ	k	g	ŋ	w	ʔ	h
Back	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+
High	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-
Coronal	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
Anterior	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
Labial	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
Continuant	-	-	-	+	+	+	+	-	-	-	+	+	+	-	+	+	-	-	+	+	-	-	-	+	-	+
Lateral	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Nasal	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
Sonorant	-	-	+	-	-	-	-	-	-	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+	-	-
Strident	-	-	-	+	+	-	-	-	-	-	+	+	-	-	+	+	+	+	-	-	-	-	-	-	-	-
Voiced	-	+	+	-	+	-	+	-	+	+	-	+	+	+	-	+	-	+	+	+	-	+	+	+	-	-

二つの構造主義とその変遷

レヴィ・ストロース
(1905-)
構造人類学・神話学

構造主義思想



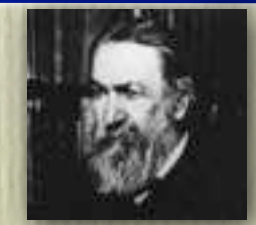
ローマン・ヤコブソン
(1896-1982)
構造音韻論

アメリカ構造言語学

ヨーロッパ構造言語学

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学

1800 1900 2000



エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

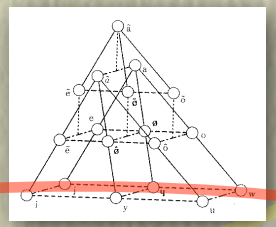
ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

コネクショニスト

複雑系

二つの構造主義とその変遷

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学



ローマン・ヤコブソン
(1896-1982)
構造音韻論

レヴィ・ストロース
(1905-)
構造人類学・神話学

構造主義思想

アメリカ構造言語学

ヨーロッパ構造言語学

1800 1900 2000

フェルディナン・ド・ソシュール(1857-1913)

「言語が含むのは，言語体系に先立って存在する観念でも音でもなく，ただこの体系から生じる観念的差異と音的差異だけである」

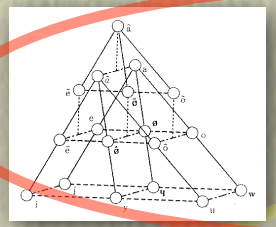
意味の世界が，各言語において，どのように領域分割しているのか？

音の世界が，各話者において，どのように領域分割しているのか？

要素の定義は，他者との差異を通して初めて可能になる。

複雑系

二つの構造主義とその変遷



レヴィ・ストロース
(1905-)
構造人類学・神話学

ローマン・ヤコブソン
(1896-1982)
構造音韻論

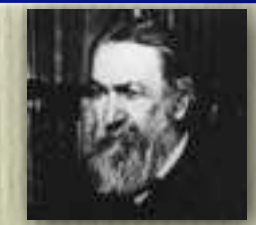
構造主義思想

アメリカ構造言語学

ヨーロッパ構造言語学

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学

1800 1900 2000



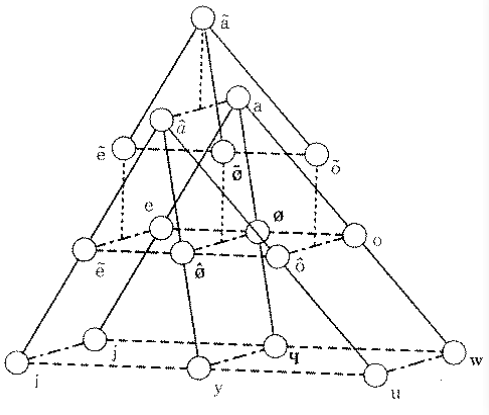
エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

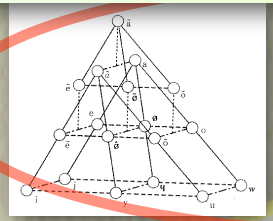
ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

コネクショニスト

複雑系



	p	b	m	f	v	θ	ð	t	d	n	s	z	l	ɾ	ʃ	ʒ	ʧ	ʤ	j	ɹ	k	g	ŋ	w	ʔ	h
Back	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+
High	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-
Coronal	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
Anterior	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
Labial	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
Continuant	-	-	-	+	+	+	+	-	-	-	+	+	+	-	+	+	-	-	+	+	-	-	-	+	-	+
Lateral	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Nasal	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
Sonorant	-	-	+	-	-	-	-	-	-	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+	-	-
Strident	-	-	-	+	+	-	-	-	-	-	+	+	-	-	+	+	+	+	-	-	-	-	-	-	-	-
Voiced	-	+	+	-	+	-	+	-	+	+	-	+	+	+	-	+	-	+	+	+	-	+	+	+	-	-



ローマン・ヤコブソン
(1896-1982)
構造音韻論

アメリカ構造言語学

ヨーロッパ構造言語学

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学

1800 1900 2000

ローマン・ヤコブソン(1896-1913)

「二音素間の差異を幾つかの弁別素性を用いて表現」

音素群の幾何学構造

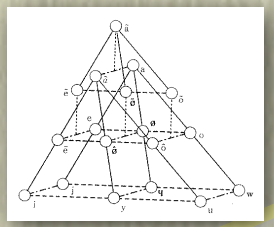
やがて音素そのものを素性の束として定義するが，これは勇み足か？

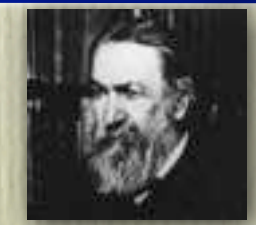
生態学的認識論・アフォーダンス

コネクショニスト

複雑系

二つの構造主義とその変遷





レヴィ・ストロース
(1905-)
構造人類学・神話学

構造主義思想

ローマン・ヤコブソン
(1896-1982)
構造音韻論

アメリカ構造言語学

ヨーロッパ構造言語学

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学

1800

1900

2000

エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

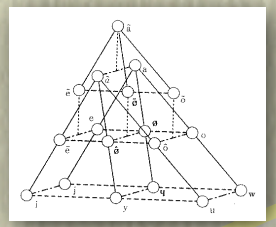
ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

コネクショニスト

複雑系

二つの構造主義とその変遷

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学



ローマン・ヤコブソン
(1896-1982)
構造音韻論

レヴィ・ストロース
(1905-)
構造人類学・神話学

構造主義思想

アメリカ構造言語学

ヨーロッパ構造言語学

1800 1900 2000

レヴィ・ストロース(1905-)

「未開社会の親族構造の中に不変的な数学的構造を見いだした。また，抽象化を通して種々の神話の中に不変的な構造を見いだした」

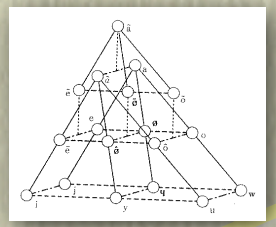
構造＝ある変換操作によって変化しない特性・関係

構造と変換は表裏一体であり，どの変換に着眼するかで構造の定義も異なる。

コネクショニスト

複雑系

二つの構造主義とその変遷



レヴィ・ストロース
(1905-)
構造人類学・神話学

構造主義思想

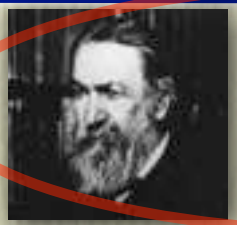
ローマン・ヤコブソン
(1896-1982)
構造音韻論

アメリカ構造言語学

ヨーロッパ構造言語学

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学

1800 1900 2000



エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

コネクショニスト

複雑系

二つの構造主義とその変遷

エルンスト・マッハ(1838-1916)

「物体の質量は，その物体の周りの全ての物体との関係で決る。他に何もない空間の中では，ある物体の質量には，何の意味もない」

全ての自然現象は感性的諸要素の関数的相互依存関係に基づく複合体

全ての観測結果は人間の感覚量でしかない。空間・時間感覚も人間の感覚の一つ。

その意味において，ユークリッド空間よりも非ユークリッド空間の方が本質的。

エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

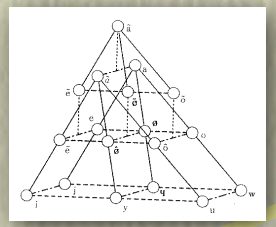
クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

コネクショニスト

複雑系

二つの構造主義とその変遷



レヴィ・ストロース
(1905-)
構造人類学・神話学

構造主義思想

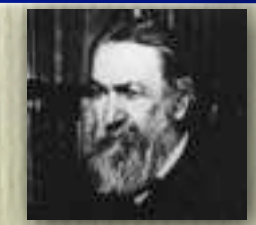
ローマン・ヤコブソン
(1896-1982)
構造音韻論

アメリカ構造言語学

ヨーロッパ構造言語学

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学

1800 1900 2000



エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

コネクショニスト

複雑系

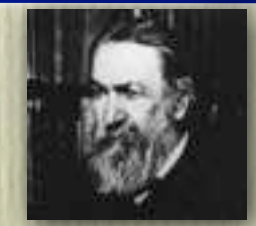
二つの構造主義とその変遷

クリスチャン・フォン・エーレンフェルス(1859-1932)

「知覚は，対象の個別刺激を統合して起こるものではなく，それ以前に全体的な枠組の中で認識が起こる」

ゲシュタルト＝全体＝「部分の単純な総和」以上のもの

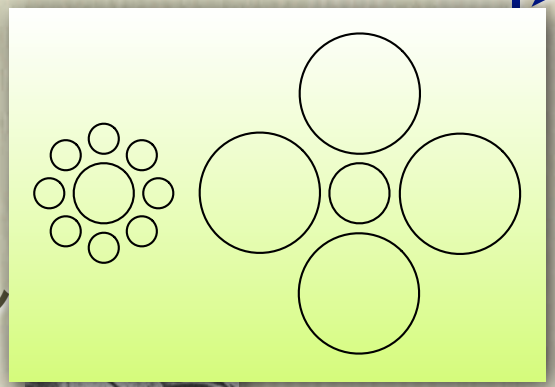
移調性と不変性：ゲシュタルトにある種の変形を施しても，ゲシュタルトは不変（メロディーと転調など）



エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

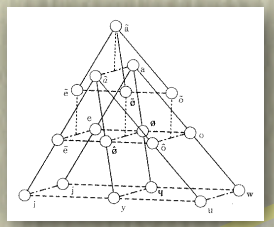
ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス



コネクショニスト

複雑系

二つの構造主義とその変遷



レヴィ・ストロース
(1905-)
構造人類学・神話学

構造主義思想

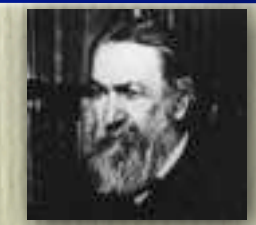
ローマン・ヤコブソン
(1896-1982)
構造音韻論

アメリカ構造言語学

ヨーロッパ構造言語学

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学

1800 1900 2000



エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

コネクショニスト

複雑系

二つの構造主義とその変遷

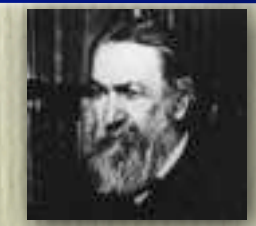
ジェームス・ギブソン(1904-1979)

「視覚は『要素刺激感覚＋そのまとめ上げ』ではなく，『対象の動き（変形）の中に見られる不変的特性』に視覚の本質がある」

変形と不変　⇔　移調性と不変性　⇔　変換と構造

面性の知覚：光が作る**差異の構造**に基づいて行なわれる（**光の差異が成す不変項**）。

情報は人間の頭の中で作り出すのではなく，環境そのものの中に存在している。



エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

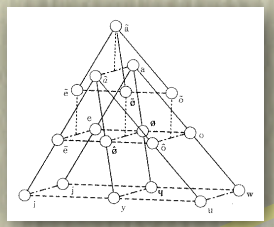
コネクショニスト

複雑系

二つの構造主義とその変遷

レヴィ・ストロース
(1905-)
構造人類学・神話学

構造主義思想



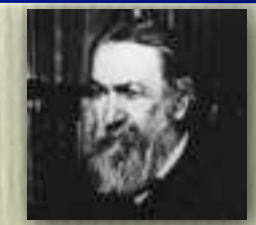
ローマン・ヤコブソン
(1896-1982)
構造音韻論

アメリカ構造言語学

ヨーロッパ構造言語学

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学

1800 1900 2000



エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

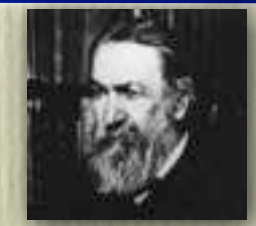
コネクショニスト

複雑系

二つの構造主義とその変遷

エルンスト・マッハ(1838-1916)

「私達は個々の要素についての研究から始めなければならないが，自然は要素と共に始まった訳ではない。私達が圧倒的な全体系から時々目をそらし，個別的な点に集中できるのは確かに幸福なことである。しかしさしあたり無視されていたことを，改めて補修し修正しながら研究することを怠ってはならない。」（要素還元主義への警鐘）



エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

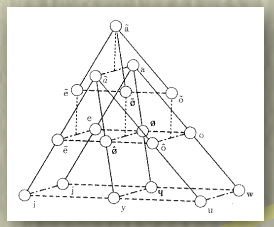
コネクショニスト

複雑系

要素感覚を集めるのではなく，要素間の関係・差異を統合することで
着眼する変換・変形操作に対して不変なる構造を見いだすこと。

レヴィ・ストロース
(1905-)
構造人類学・神話学

構造主義思想



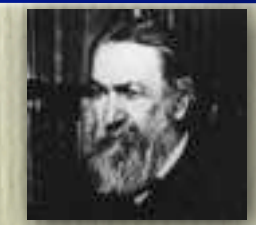
ローマン・ヤコブソン
(1896-1982)
構造音韻論

アメリカ構造言語学

ヨーロッパ構造言語学

フェルディナン・ド・ソシュール
(1857-1913)
一般言語学

1800 1900 2000



エルンスト・マッハ
(1838-1916)
物理学・音響学・心理学

クリスチャン・フォン・エーレンフェルス
(1859-1932)
ゲシュタルト心理学

ジェームス・ギブソン
(1904-1979)
生態学的認識論・アフォーダンス

コネクショニスト

複雑系