

グローバル時代の英語発音と その科学的な分析方法

峯松 信明

東京大学大学院情報理工学系研究科



講演者自己紹介



氏名：峯松信明

所属：東京大学大学院情報理工学系研究科

：電子情報学専攻・准教授

専門：音声科学・工学（語学支援応用を含む）

経歴：高校時代に一時期英語教師を目指すも理科系に

：駒場時代は2年間，舞台の上で過ごした@ESS

：主役@1年目，演出@2年目

：1989年学部4年夏から音声研究を開始

：1995年博士（工学）取得@東大工学系

：1996年～@豊橋技術科学大学情報工学系

：2000年～@東京大学大学院情報理工学系研究科

本日のメニュー

音声の物理学とその基礎

- 声が生まれる仕組み／音色，ピッチ，とその物理的対象
- 声の多様性（性別，年齢，体格によって変わる様々な声色）

声の共通項（構造的表象）とその発音評定への応用

- 発音学習とは教師の声の何を真似る作業を指すのか？
 - 幼児は親の声の何を真似て言葉を獲得するのか？／音色の相対感覚
 - 声帯模写と発音学習（音声模倣）の違いに立脚した基盤技術構築
- 学習者はどのような発音を目指すべきか？
 - 母語話者のような発音 vs. 伝わる発音
- 構造表象に基づく発音診断ソフトの開発

音声の構造的表象を通して考える発音教育

- 見えないもの，意識できないものを捉えるには？

まとめ

音声の生成 ～母音～

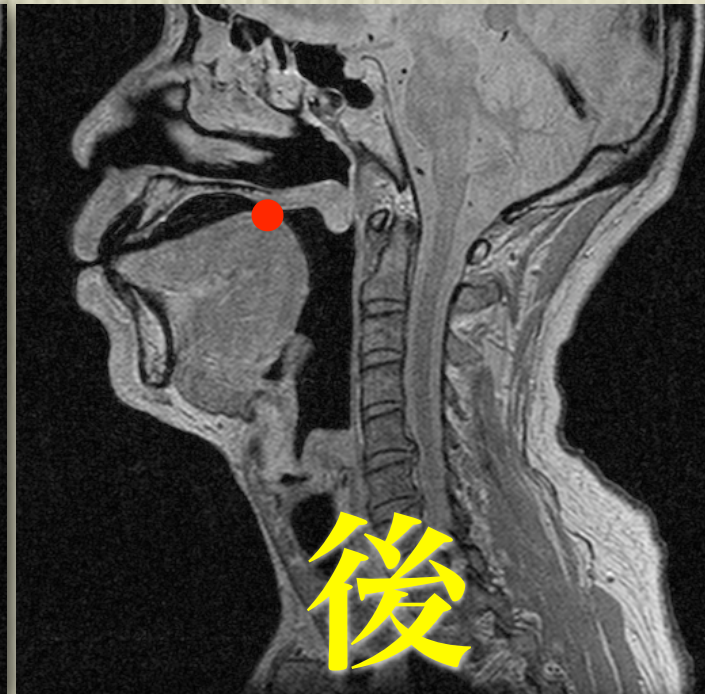
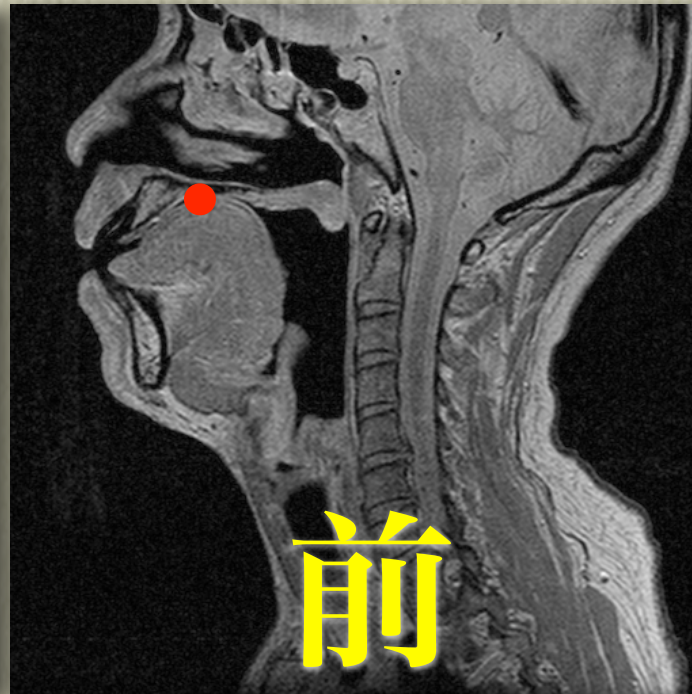
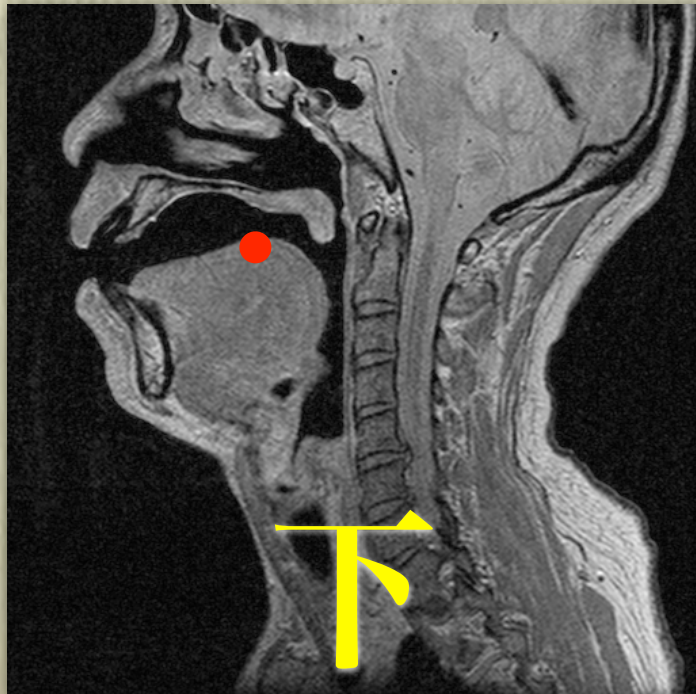
音声ってどうやって生まれるんだろう？

- 「あいうえお」と言ってみる。何が動いているだろうか？
 - 「口の開き方」つまり「あご」が動いてるのは、外から見ても分かる。
 - 口の中では「ベロ」の位置が大きく動いているのも分かる。
 - 「ベロ」の位置を変えるために「あご」を動かしている？

あ

い

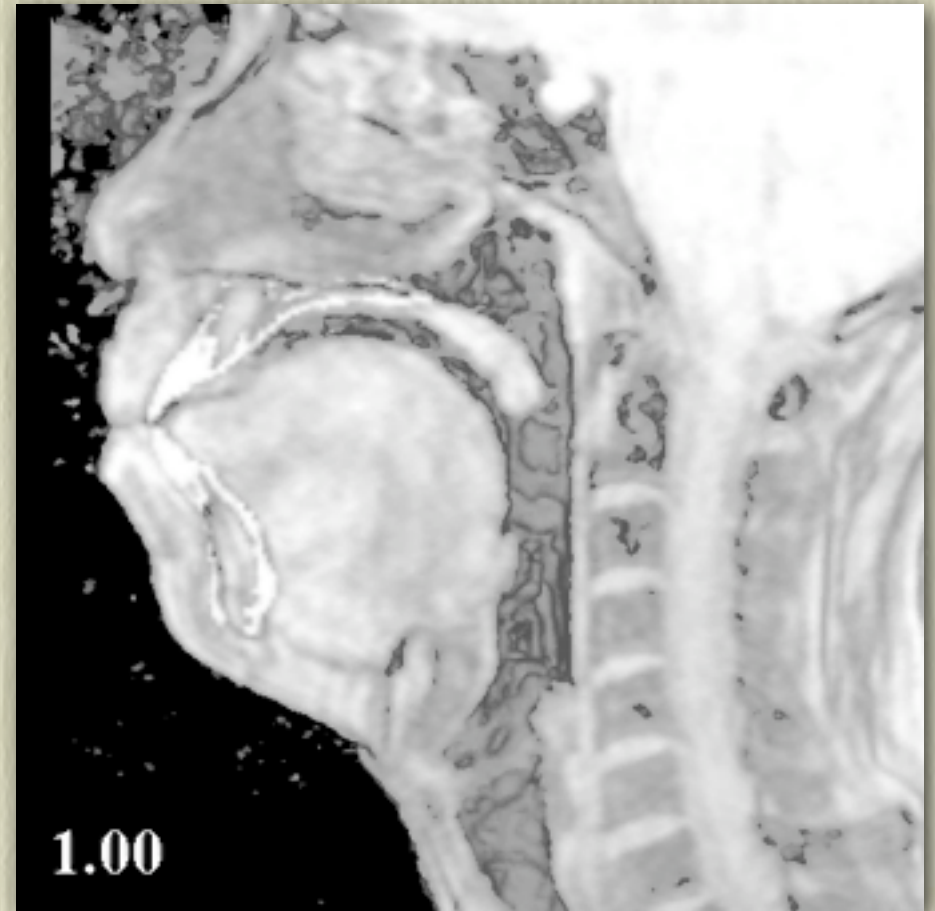
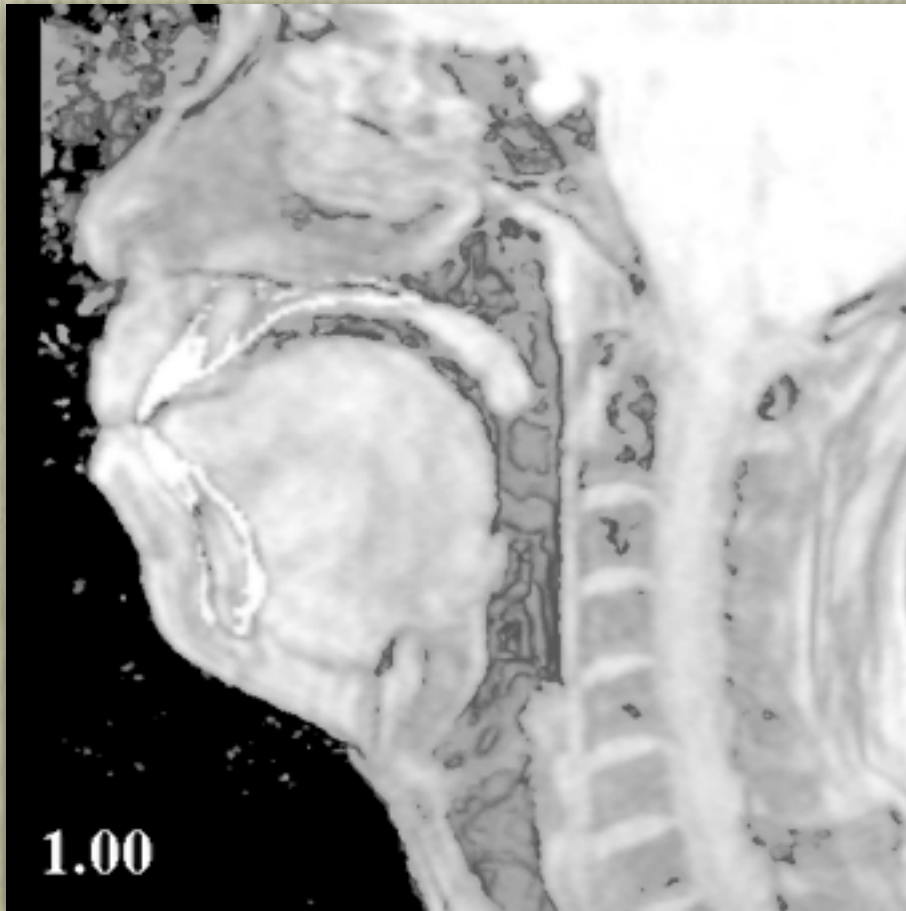
う



口の中の隙間を変える!?

でかいのは牛タンだけじゃない！人タンだって！

● しかもこんなにダイナミックに動いちゃいます！

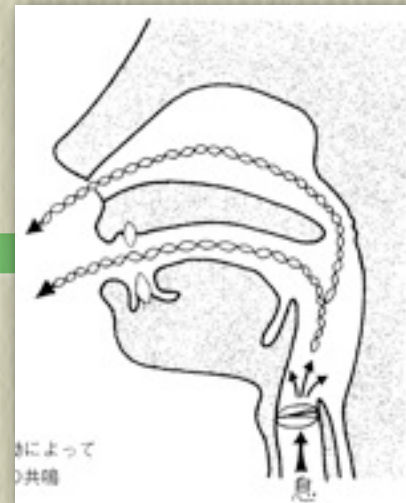
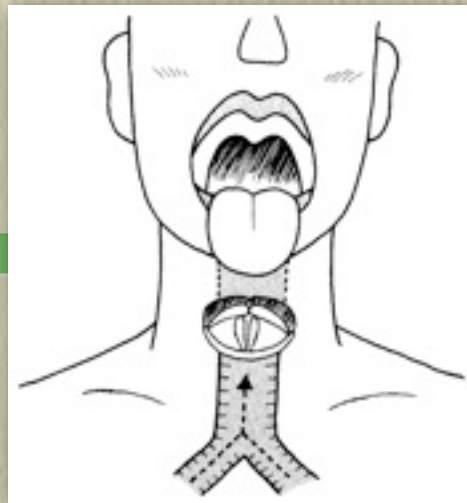
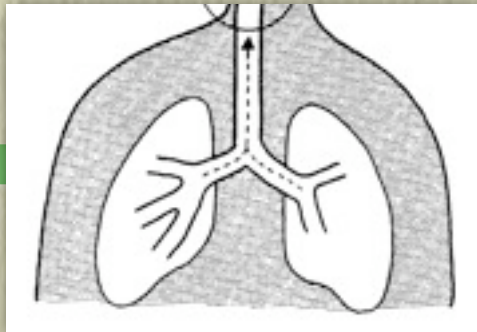
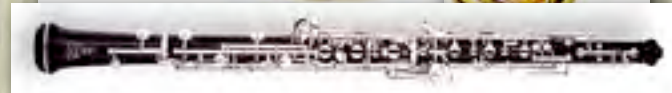


(株)ATR人間情報科学研究所提供

● 「あご」の動き→「ベロ」の動き→「口の中のすき間」の変形

お口は楽器、 色んな音を奏でます

息 → ブー → 色んな形の管 → 色んな音色

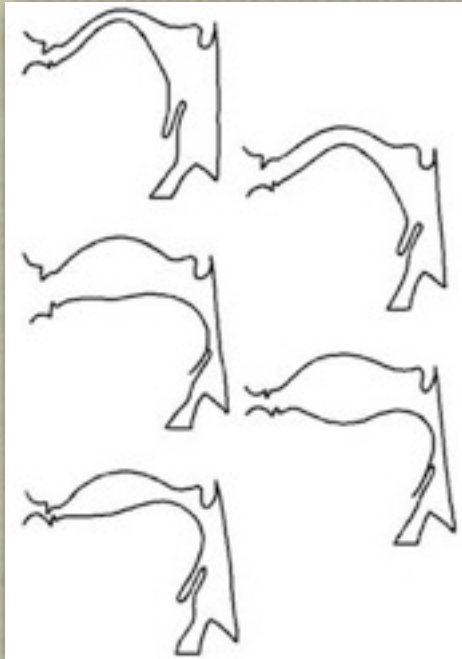
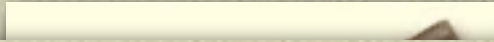


i u
e o
a

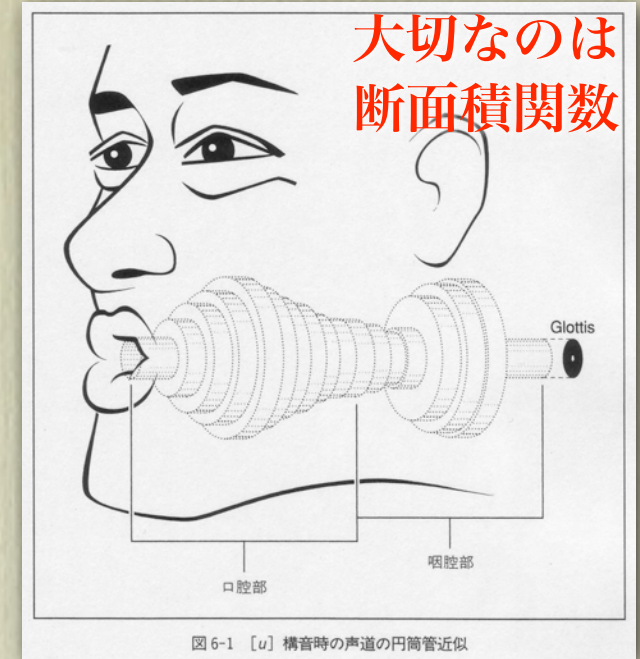
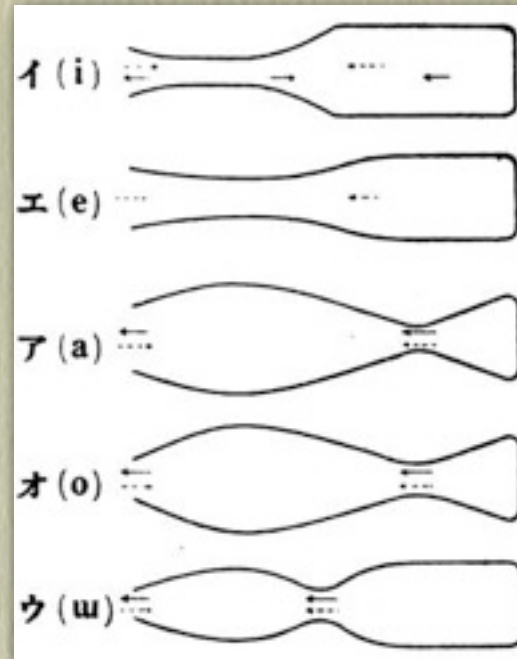
管の形の違い＝出てくる音の音色の違い

お口は楽器、色んな音を奏でます

息 → ブー → 色んな形の管 → 色んな音色

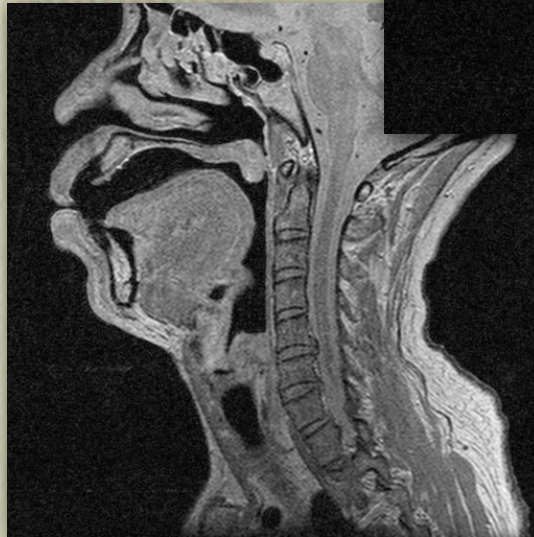
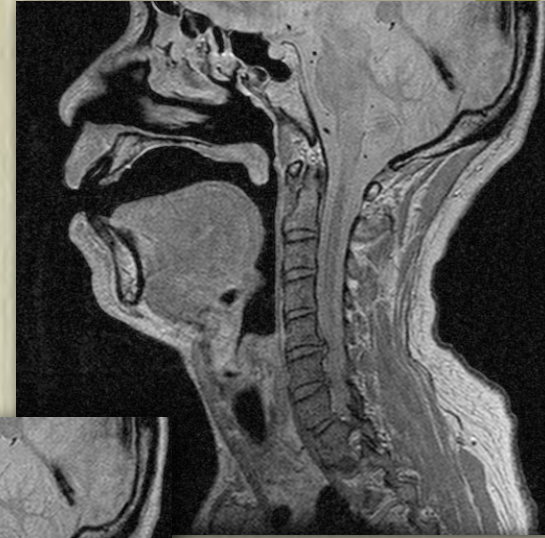
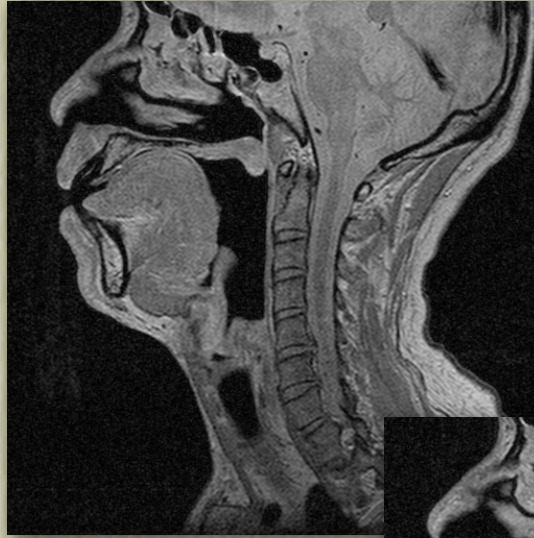
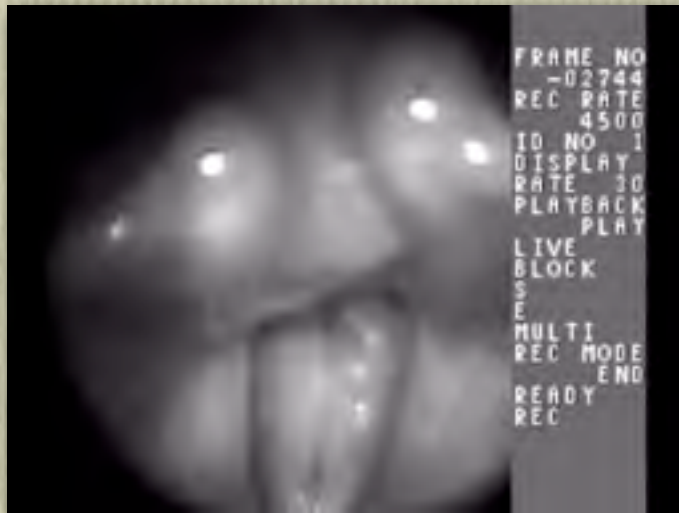


曲げ延ばし
はOK



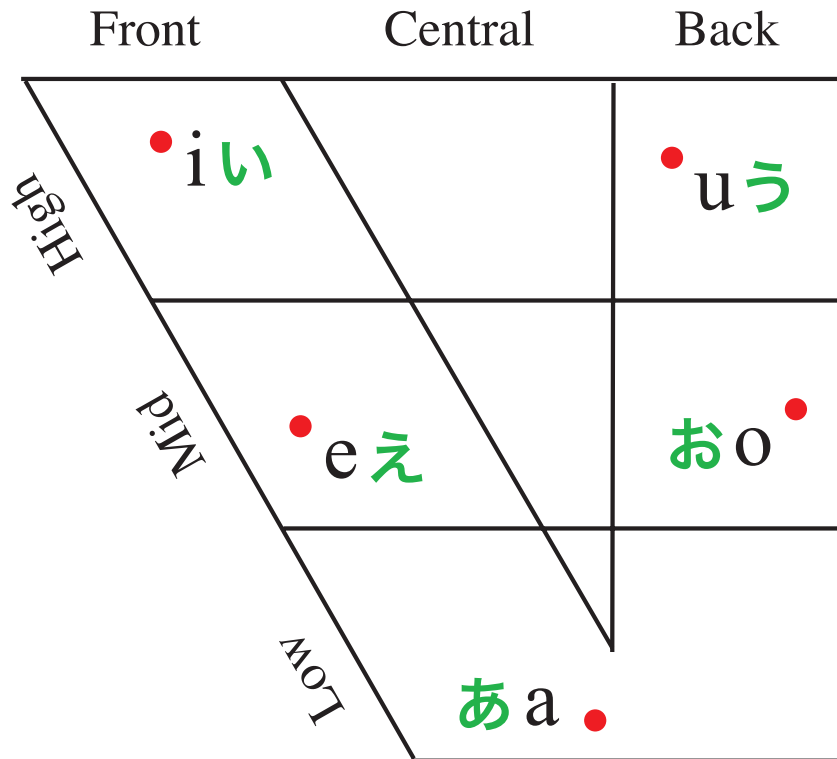
でも我々の楽器はグニャグニャ形が変わる

声帯音源 + 喉 = ブザー + 管

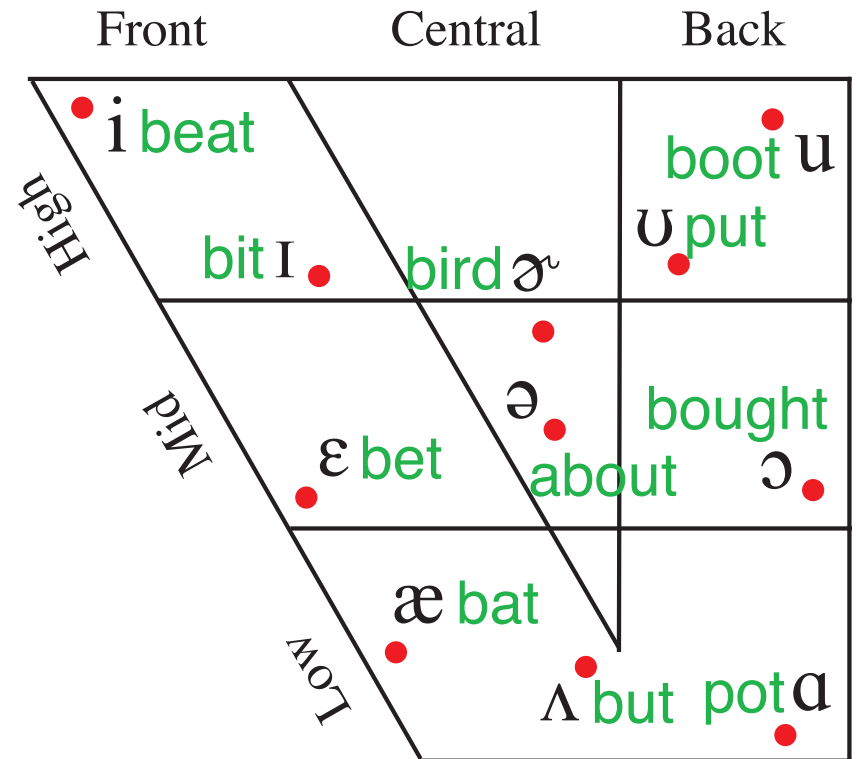


日本語とアメリカ英語の母音図

あ=下, い=前, う=後



5

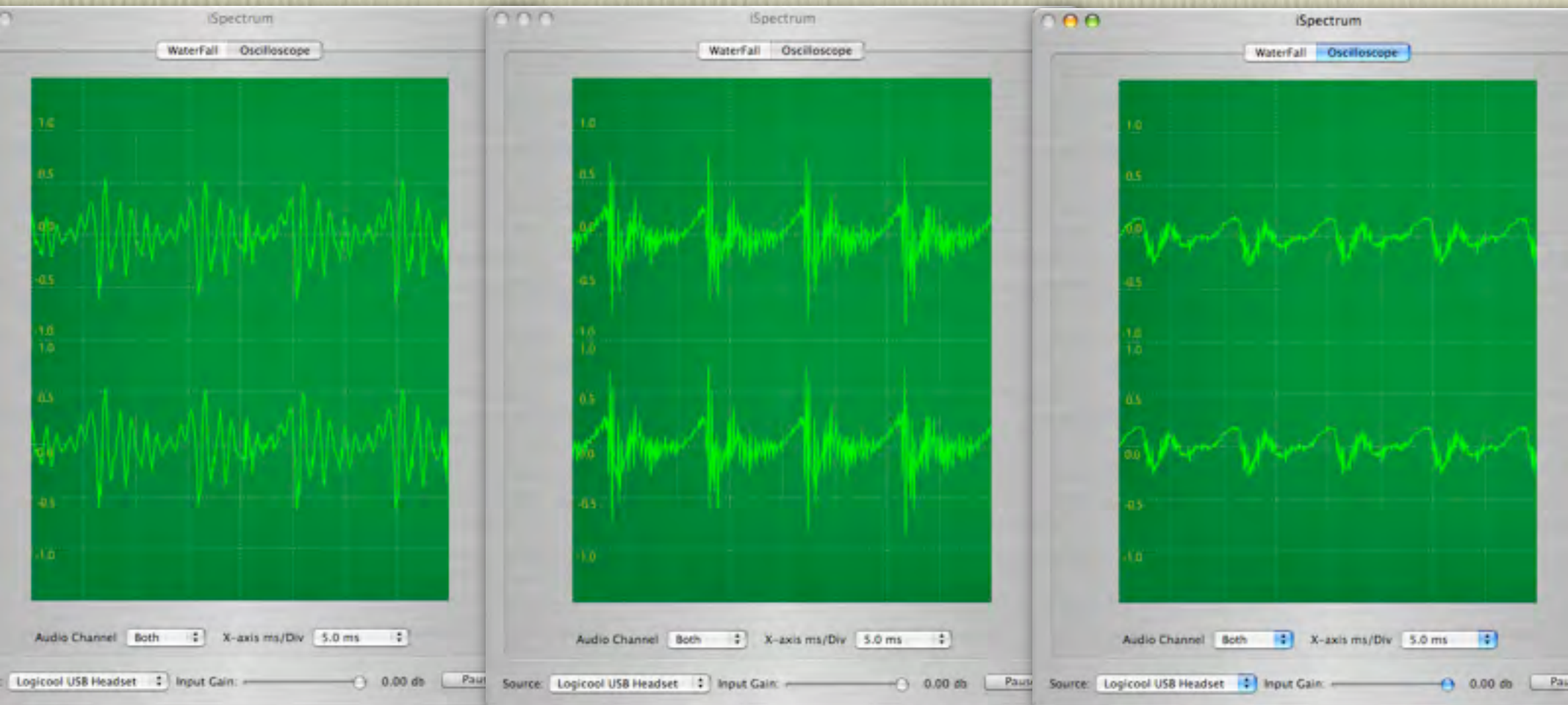


2倍+1

空気振動としての音声 ～波～

じゃ、その空気粒子振動としての声（音）を見る！

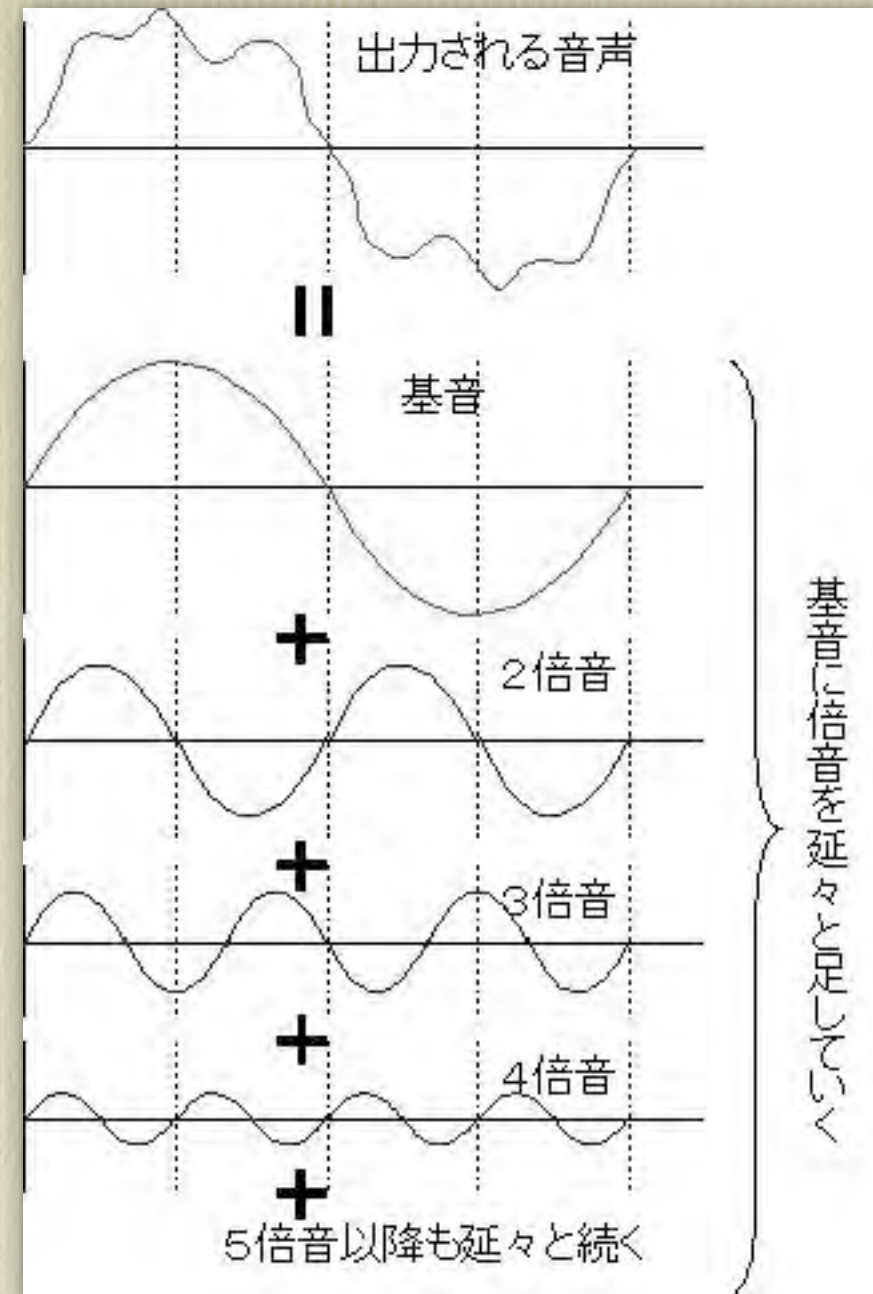
- 波形表示ソフトで見た峯松の「あ」「い」「う」
- 周期的な波形の連続→音の「高さ」を感覚できる



波＝基本的な波の足し合わせ

分解と合成（フーリエ変換）

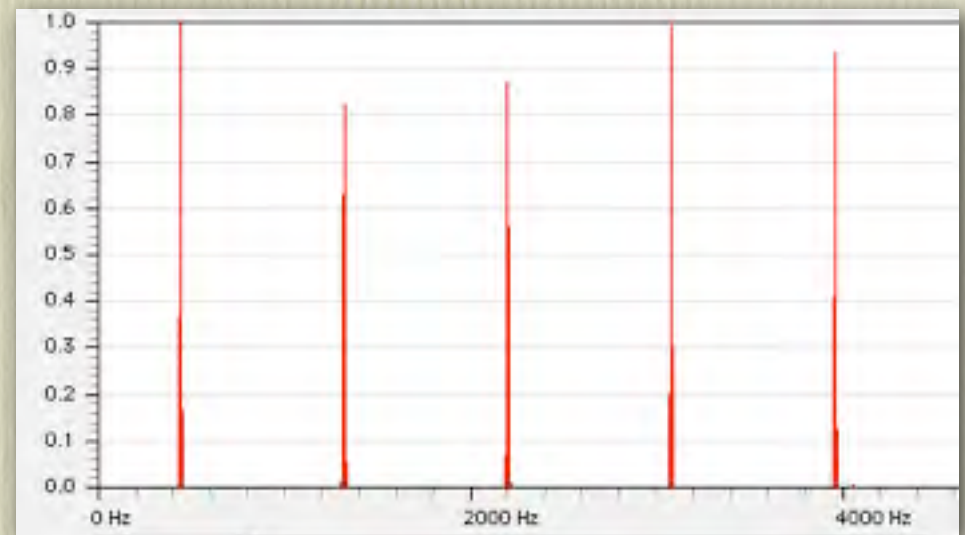
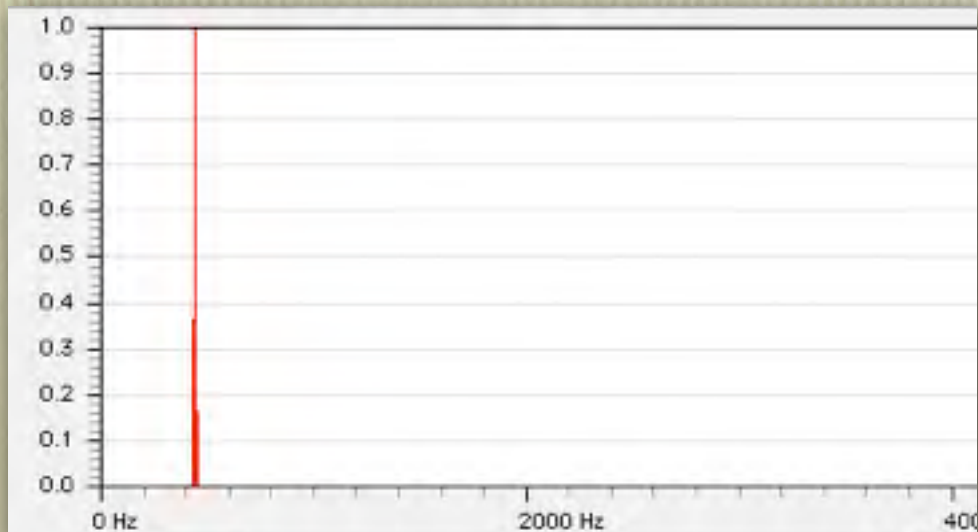
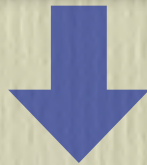
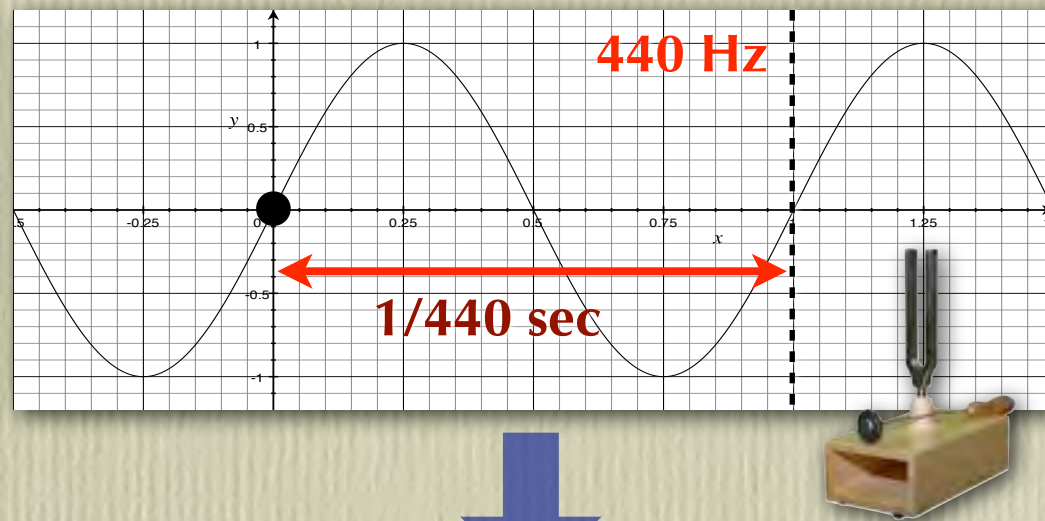
- 波＝基本音＋2倍音＋3倍音＋・・・
- 周波数：振動回数／秒
- 波：これらに適切な重みをかけて足しあわされた結果
- どの周波数の波は強く、どの周波数の波は弱いのか？
- どの周波数のエネルギーは強く、どの周波数のエネルギーは弱いのか？
- 横軸周波数、縦軸エネルギー強度としてグラフを書くと・・・
- それが**スペクトル**



波とそのスペクトルの例

波形 = w_1 基本音 + w_2 2倍音 + w_3 3倍音 + w_4 4倍音 +

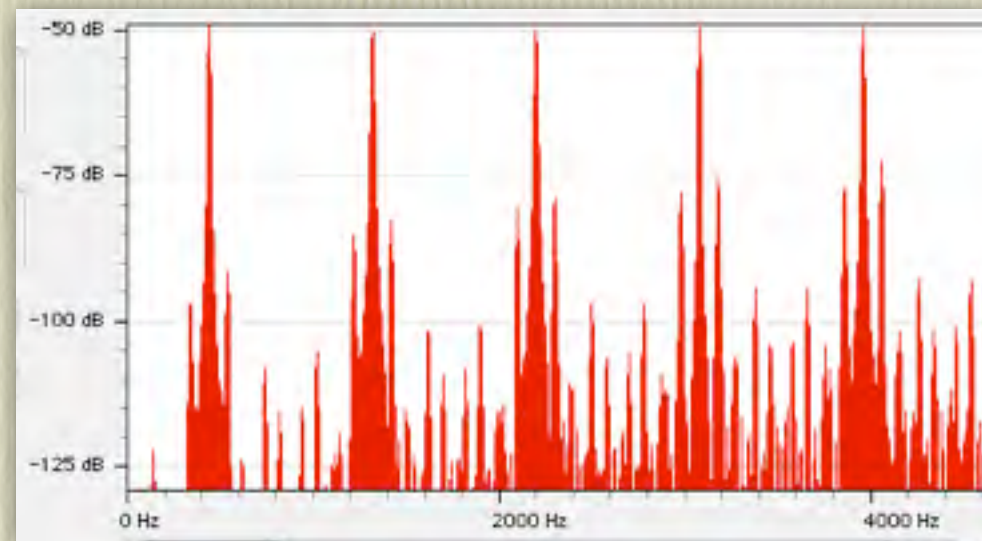
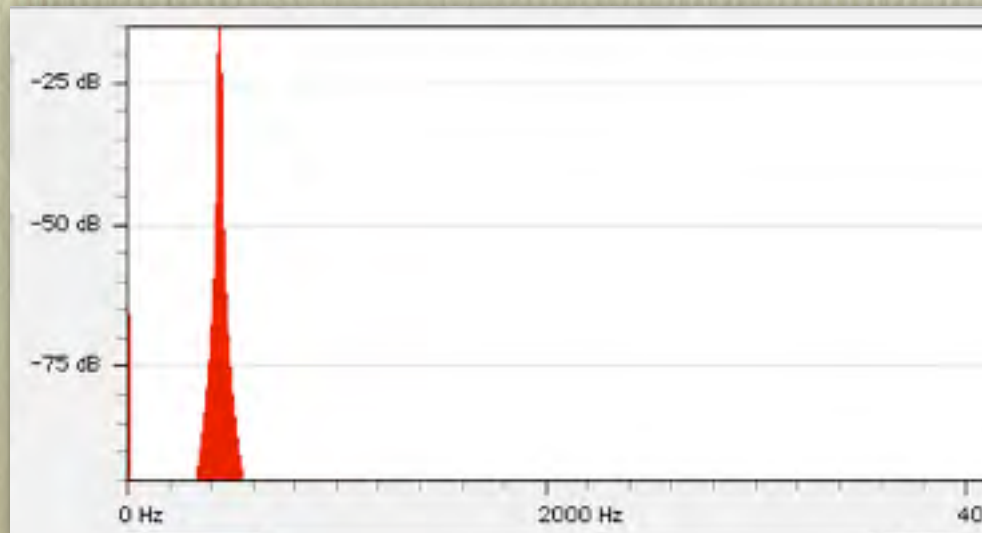
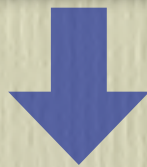
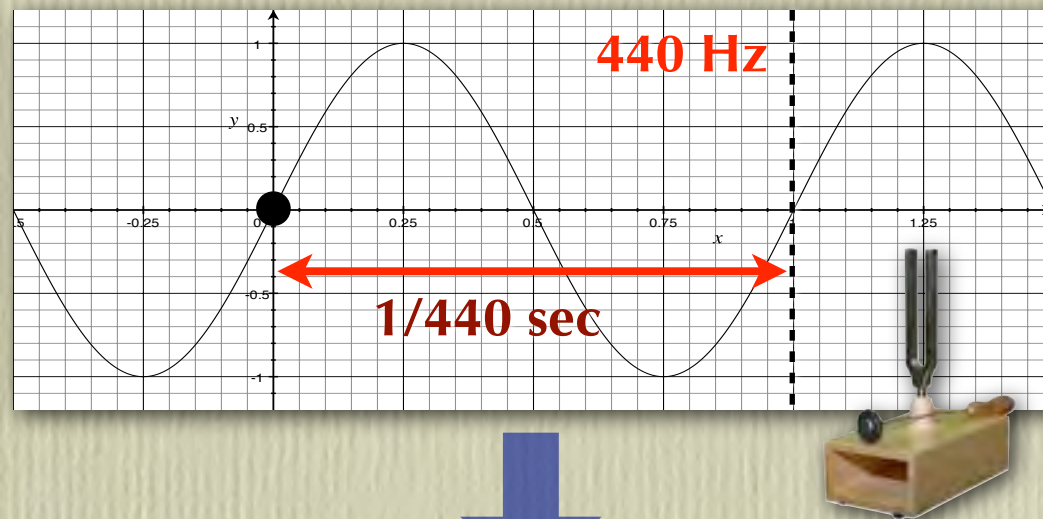
w_1 基本波 + w_2 第二高調波 + w_3 第三高調波 + w_4 第四高調波 +



スペクトル→対数スペクトル

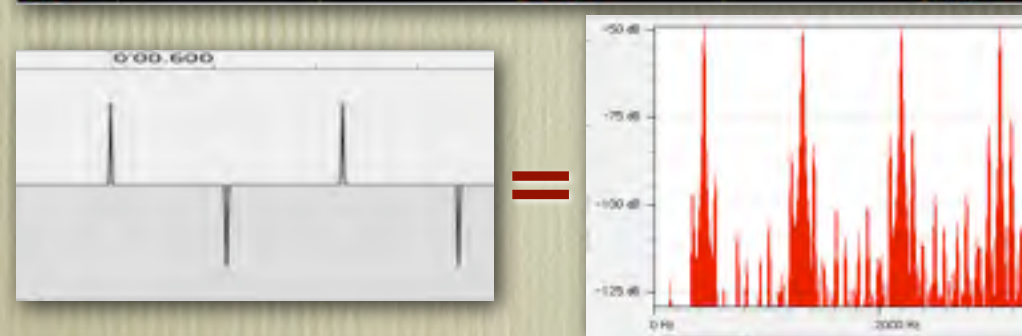
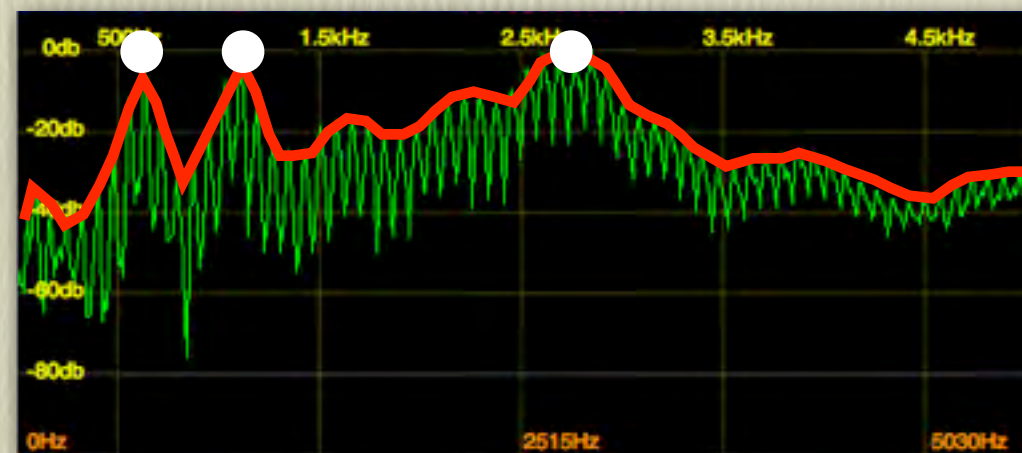
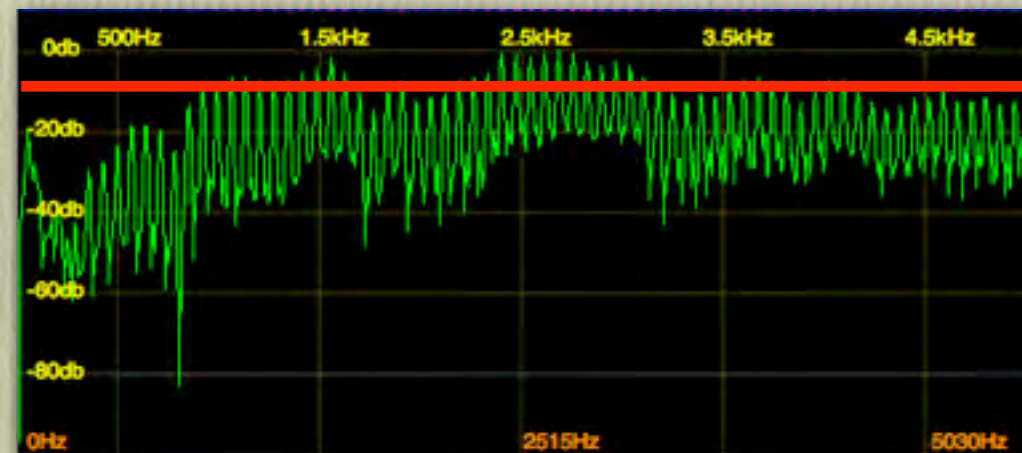
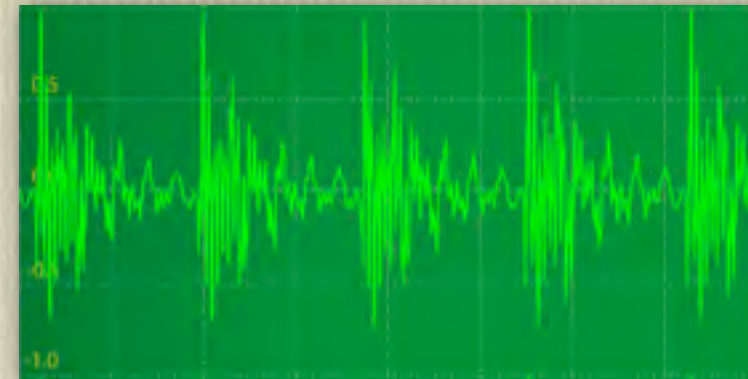
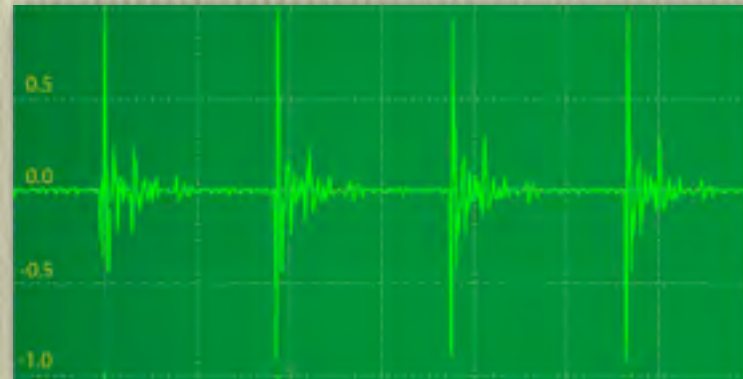
波形 = w_1 基本音 + w_2 2倍音 + w_3 3倍音 + w_4 4倍音 +

w_1 基本波 + w_2 第二高調波 + w_3 第三高調波 + w_4 第四高調波 +



音声の対数スペクトル

ブザー（パルス列）→管→「あ〜」

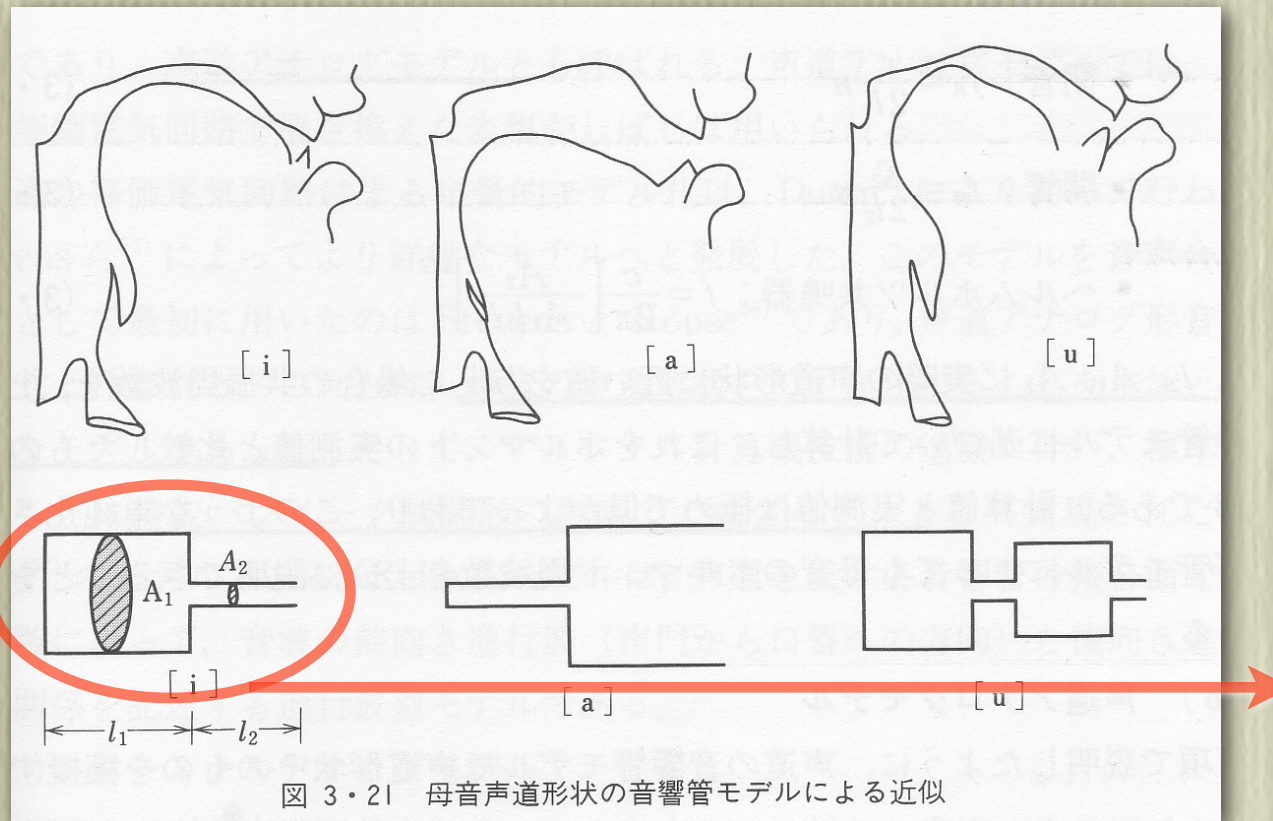


エネルギー配分に偏りが生じる
この様子が管形状によって異なる
エネルギーの局所的集中＝共鳴

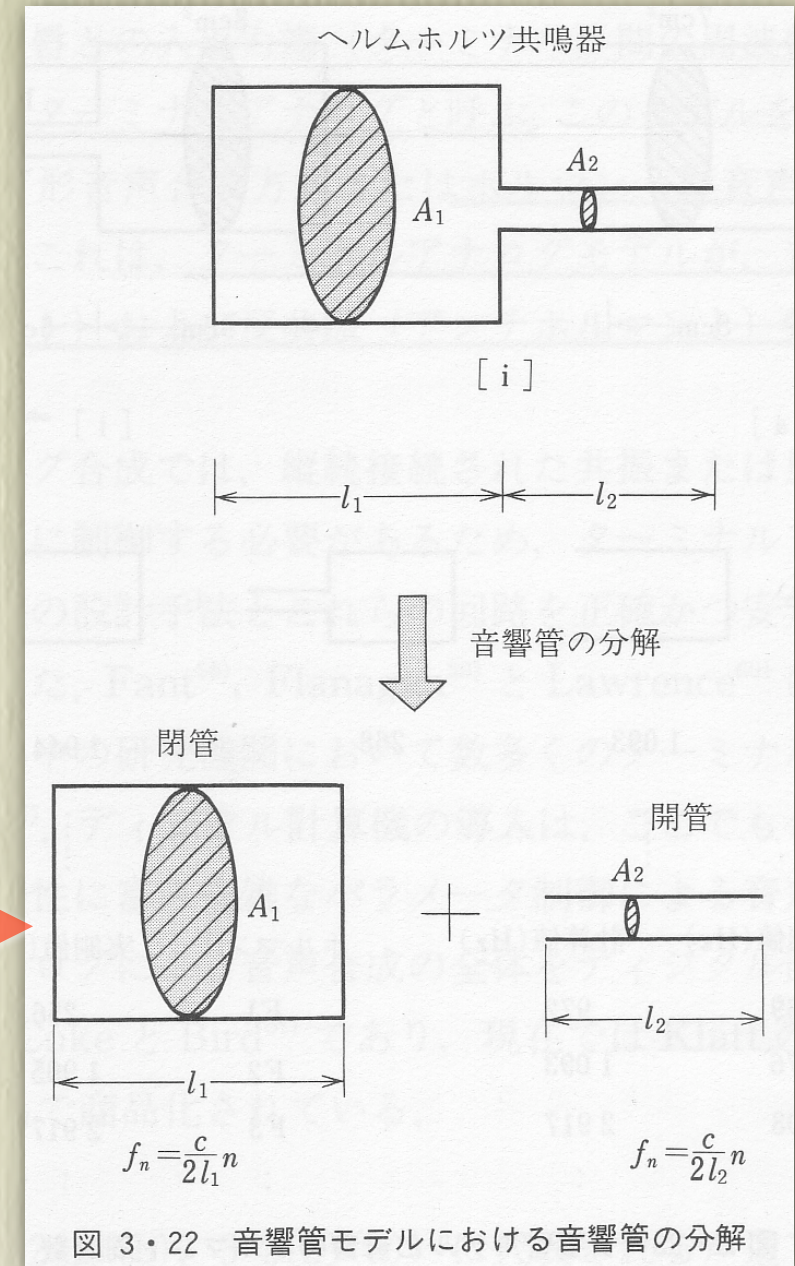
エネルギーの局所的集中＝共鳴

管の形状が共鳴の様子を決める

- 音声の共振（共鳴）周波数を求めて
- = フォルマント周波数

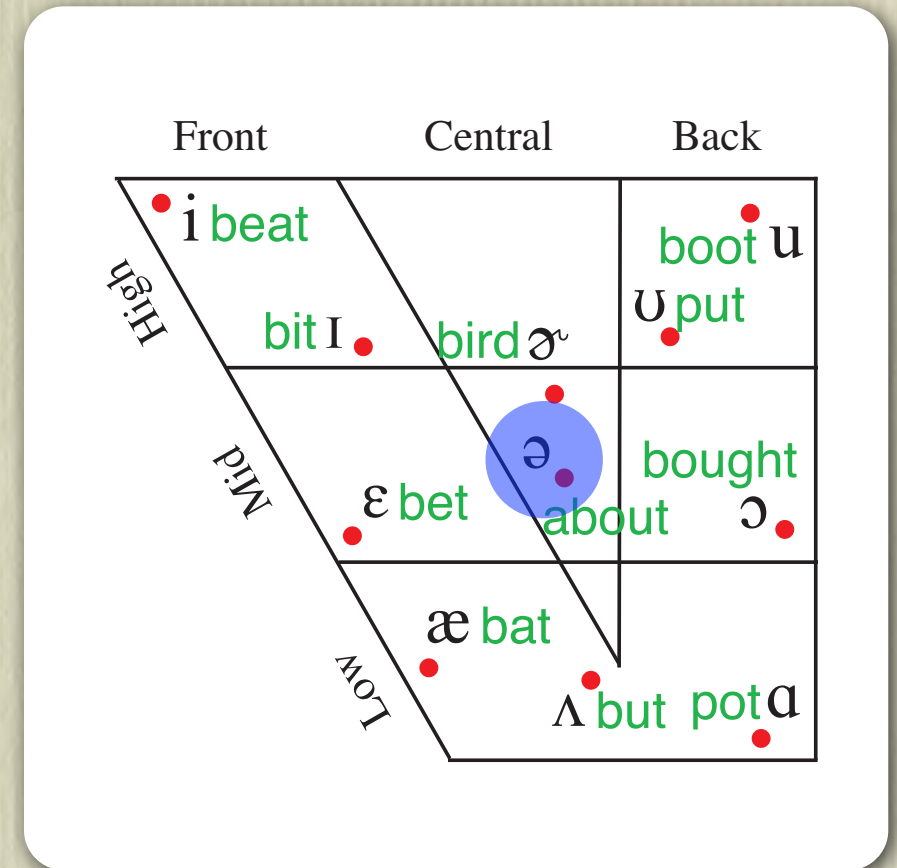
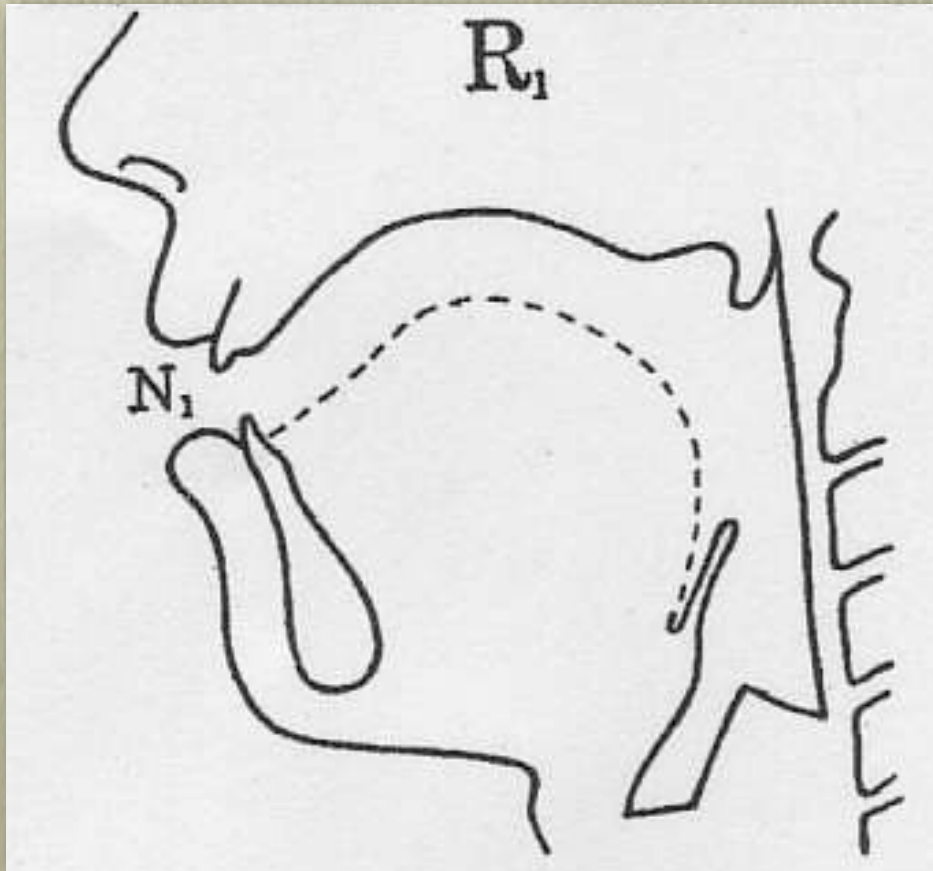


$$f_n = \frac{c}{2l_1} n \quad f_n = \frac{c}{2l_2} n \quad f = \frac{c}{2\pi} \left[\frac{A_2}{A_1 l_1 l_2} \right]^{1/2}$$



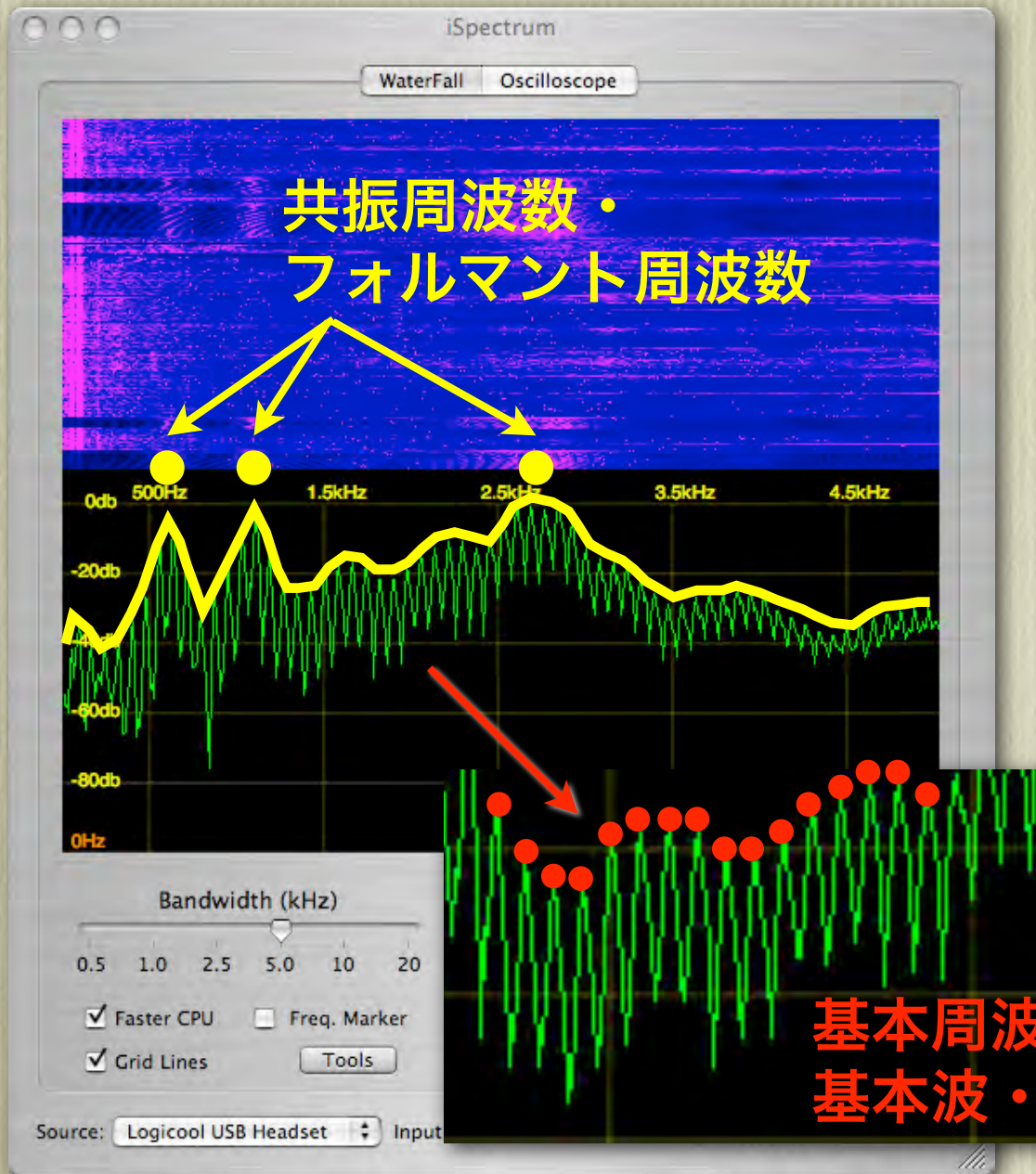
ホームポジション=schwa

喉が断面積一定の管になることはあるのか？



音色とピッチ＝共振周波数と基本周波数

音色＝エネルギー分布，ピッチ＝櫛の歯の間隔

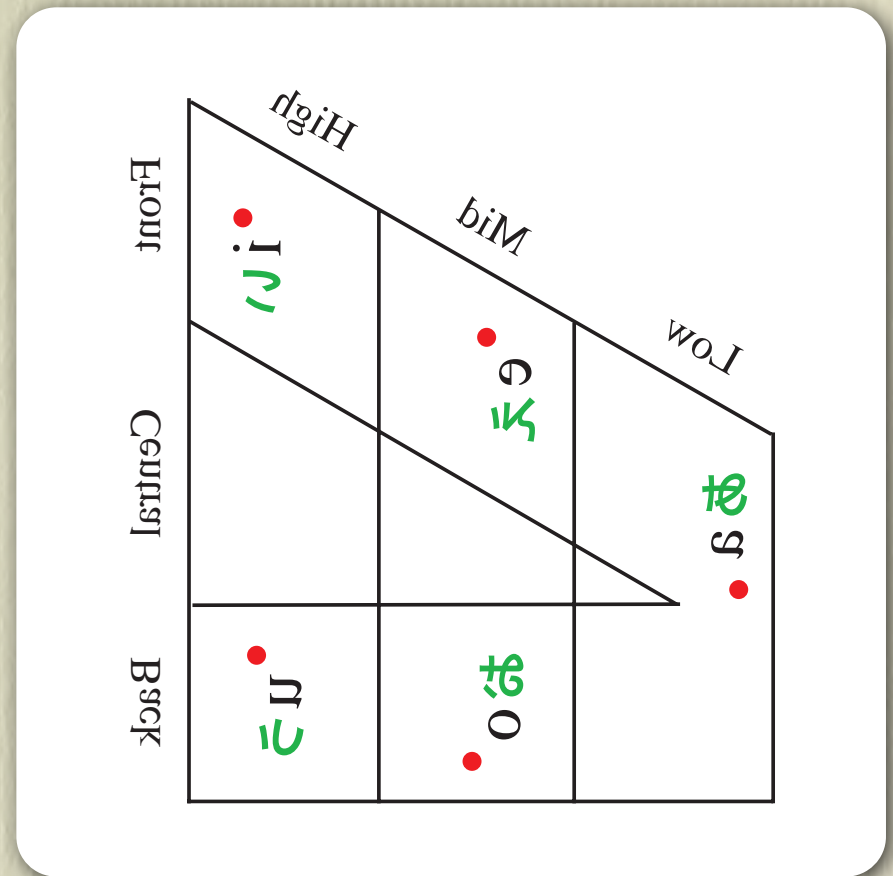
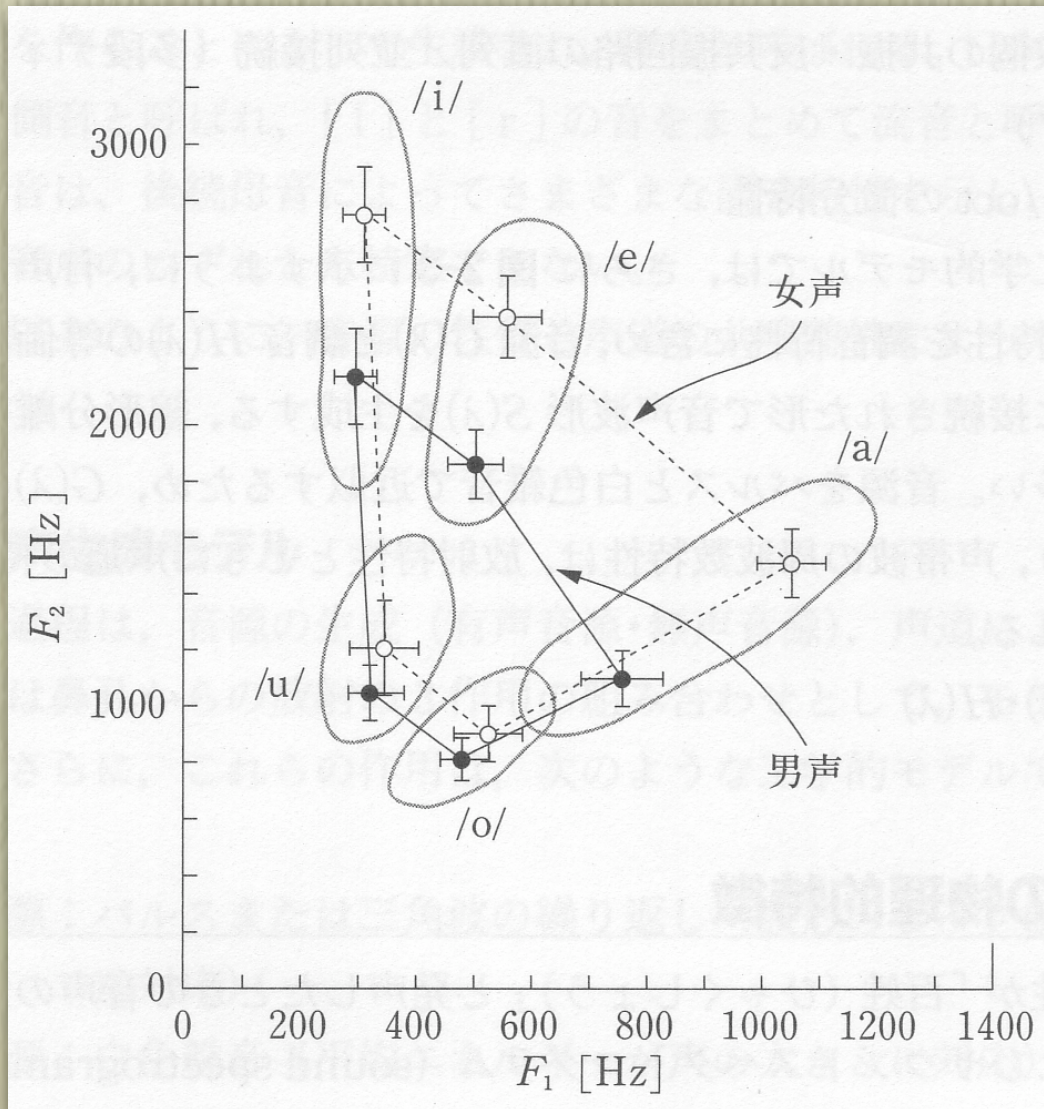


基本周波数・
基本波・第n高調波

母音の違いと話者の違い

「母音の違い」も「話者の違い」も共振周波数の違い

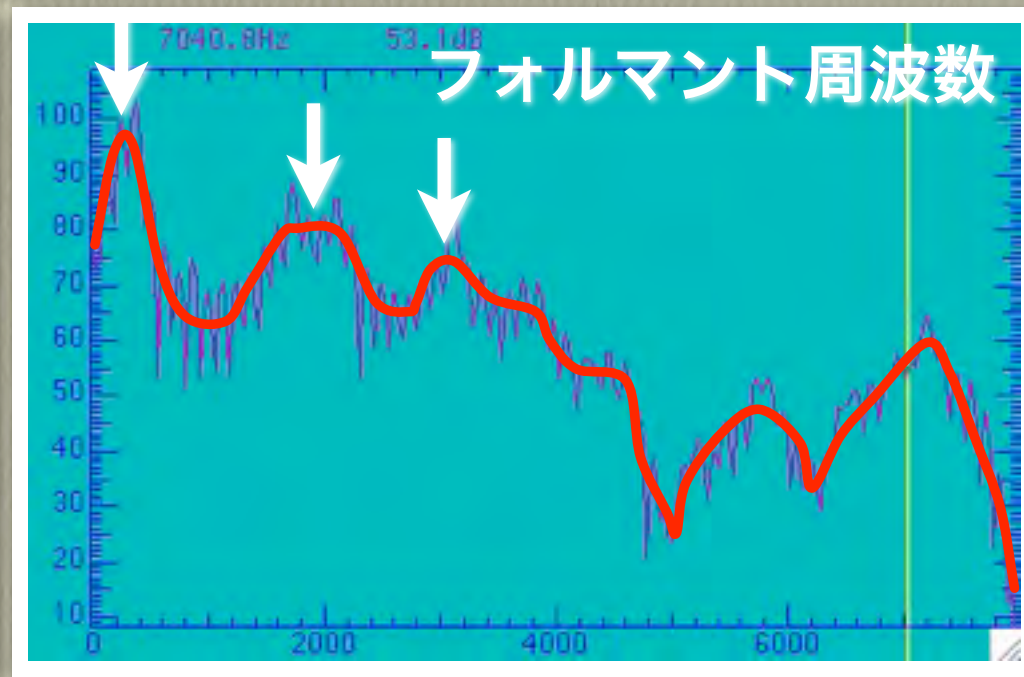
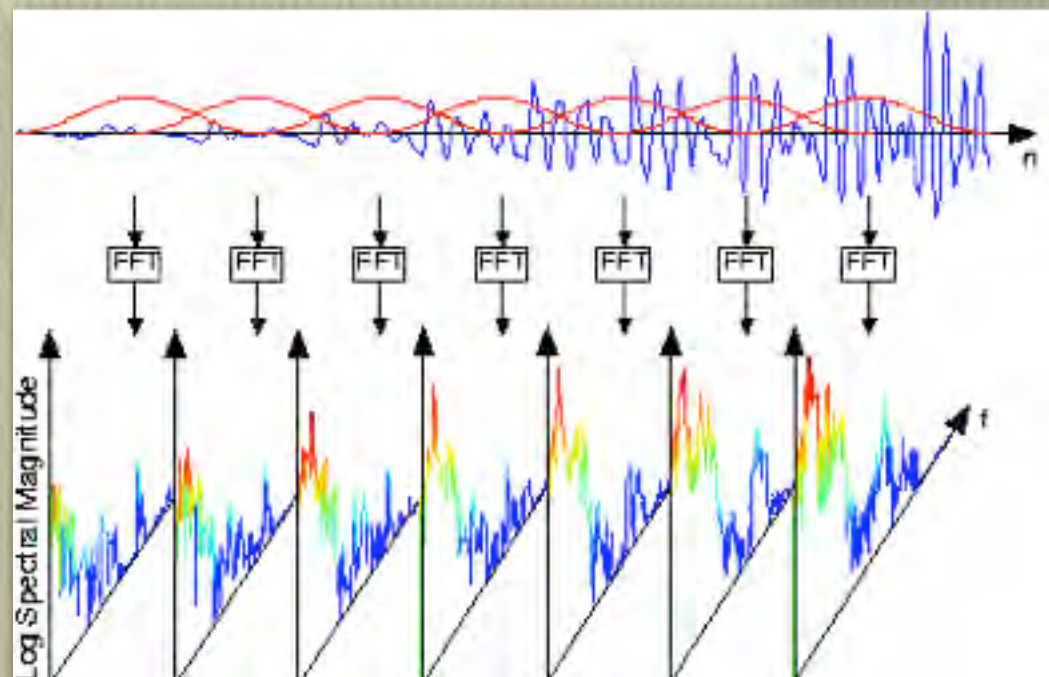
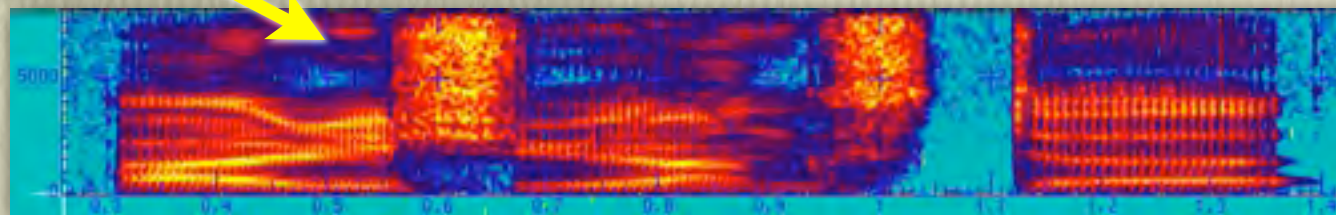
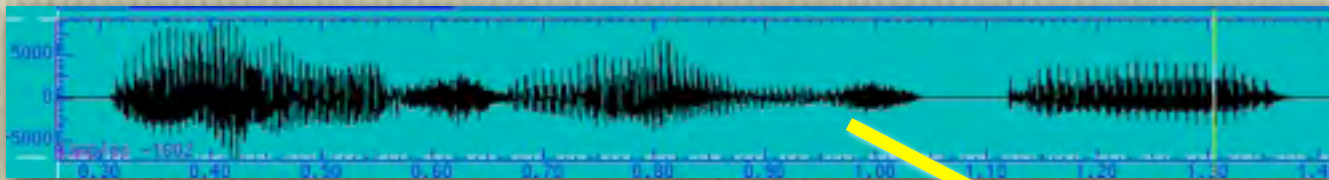
● 日本語 5 母音の第一・第二フォルマント周波数



連続的に変化する音声の音響分析

音声波形からスペクトル系列へ

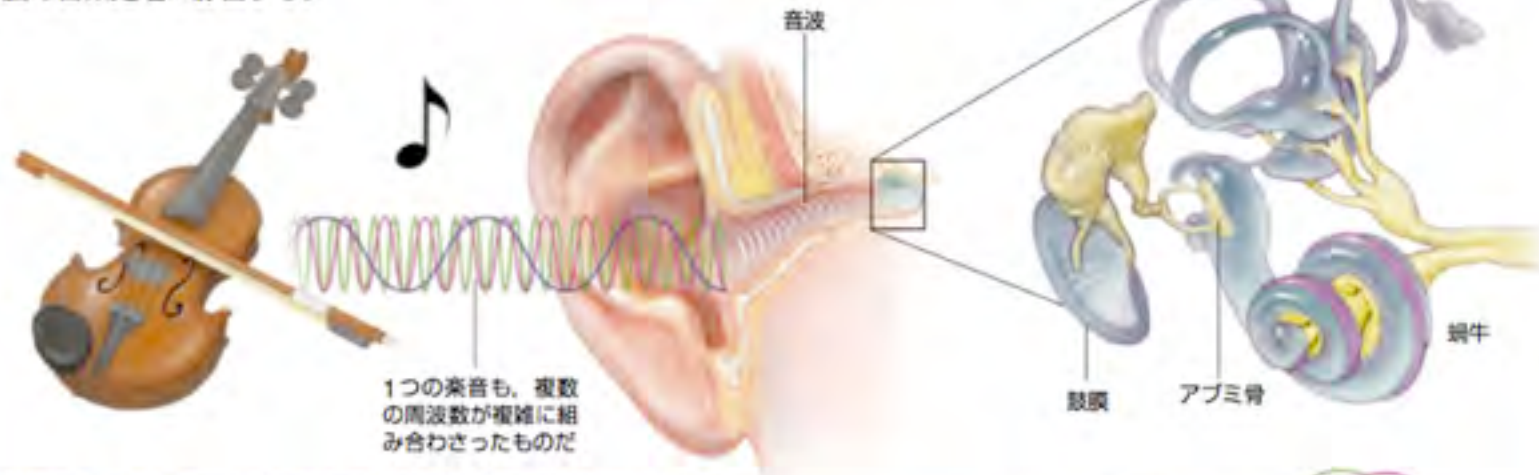
窓掛け→短時間フーリエ変換→対数パワースペクトル系列



耳から脳へ ～音楽の通り道～

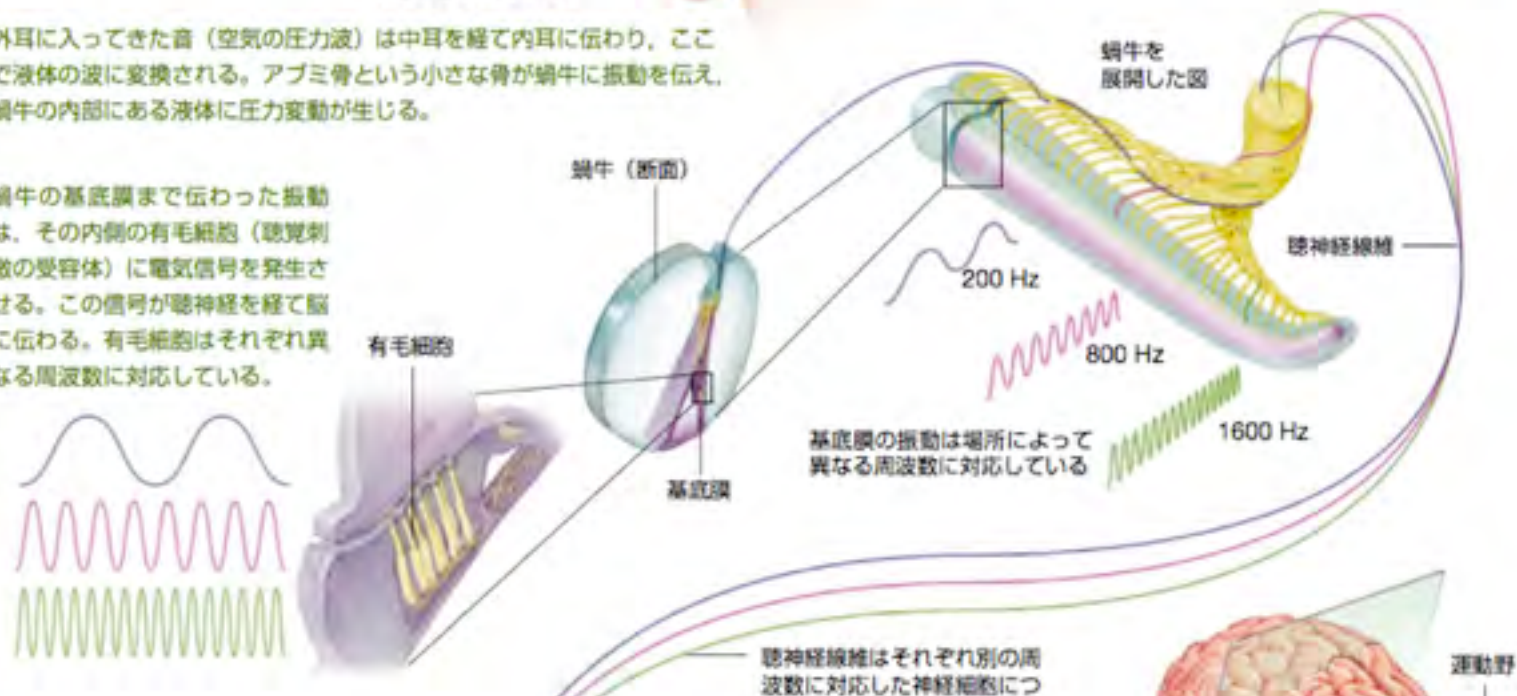
耳から脳へ 音楽の通り道

音楽を聴くと、脳では聴覚野だけでなく、さまざまな領域が反応する。通常なら音とは別の処理に関係している領域も加わり、視覚や触覚、感情といった体験すべてが脳の音楽処理に影響する。



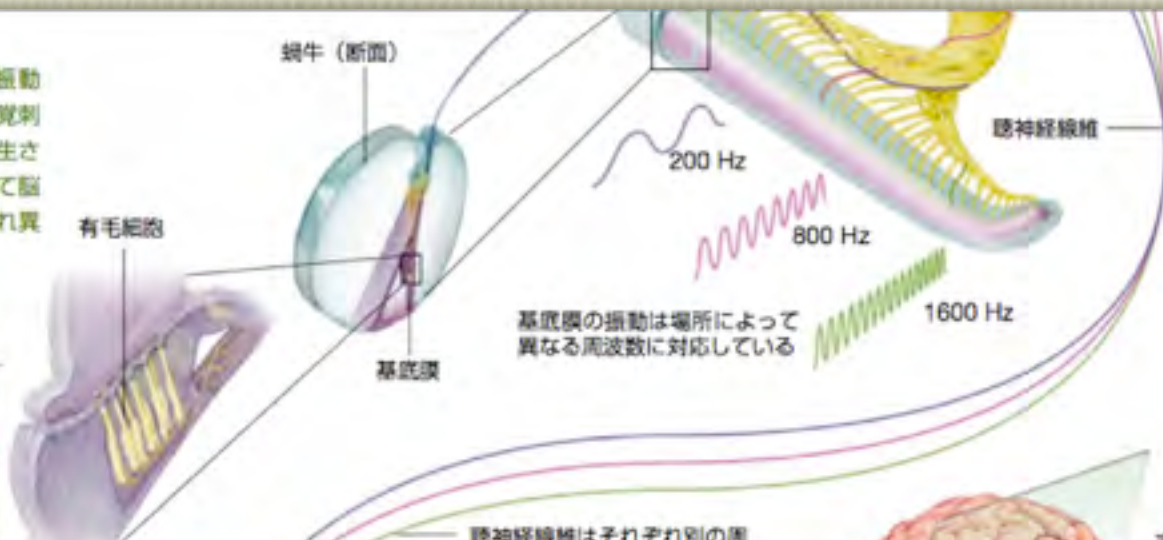
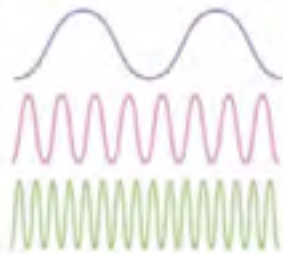
外耳に入ってきた音（空気の圧力波）は中耳を経て内耳に伝わり、ここで液体の波に変換される。アブミ骨という小さな骨が蝸牛に振動を伝え、蝸牛の内部にある液体に圧力変動が生じる。

蝸牛の基底膜まで伝わった振動は、その内側の有毛細胞（聴覚刺激の受容体）に電気信号を発生させる。この信号が聴神経を経て脳に伝わる。有毛細胞はそれぞれ異なる周波数に対応している。



耳から脳へ ～音楽の通り道～

蝸牛の基底膜まで伝わった振動は、その内側の有毛細胞（聴覚刺激の受容体）に電気信号を発生させる。この信号が聴神経を経て脳に伝わる。有毛細胞はそれぞれ異なる周波数に対応している。

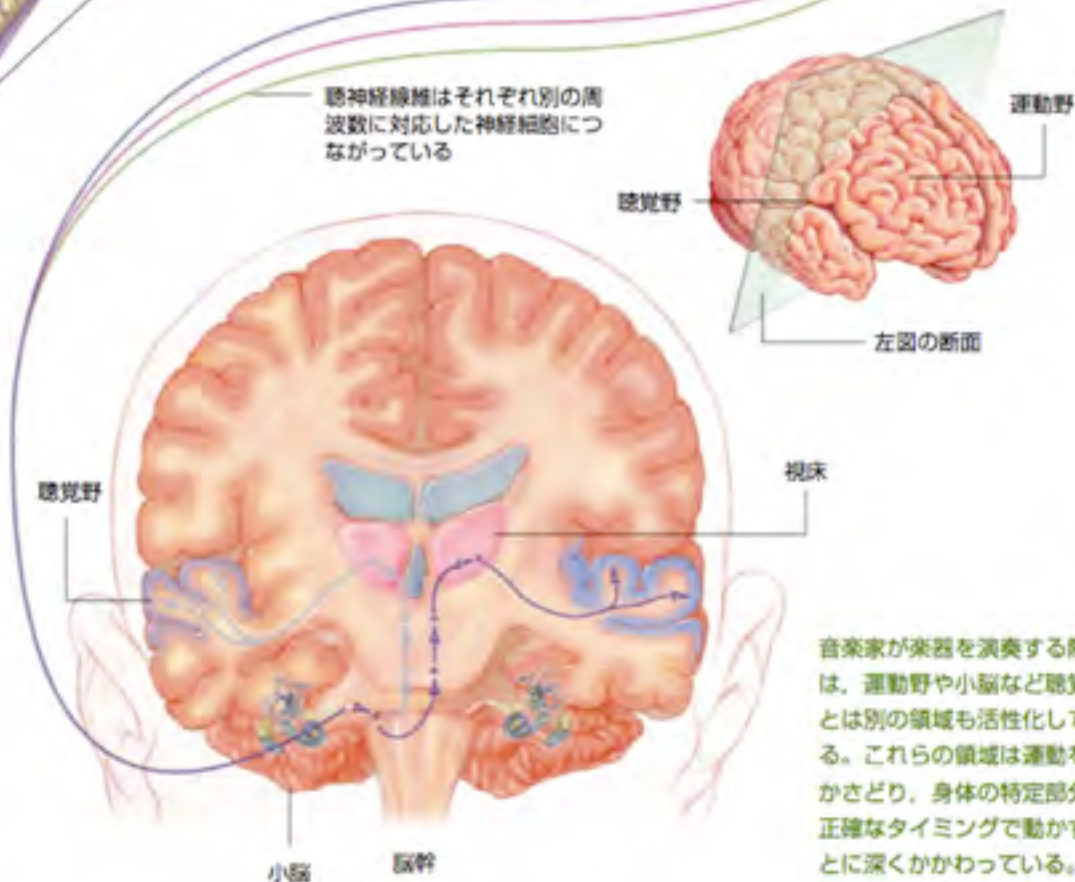


基底膜の振動は場所によって異なる周波数に対応している

聴神経線維はそれぞれ別の周波数に対応した神経細胞につながっている

脳による音楽の処理は階層的な面もあるし、分散的な処理による部分もある。聴覚野のうち、一次聴覚野という部分が音楽知覚の最初の段階にかかわっている。耳など下位聴覚系（大脳新皮質以外の聴覚システム）からの信号が視床を経由して一次聴覚野に伝わり、ここで音程（周波数）や音調曲線（音の高低変化のパターン）といった旋律の基礎をなす要素が認識される。

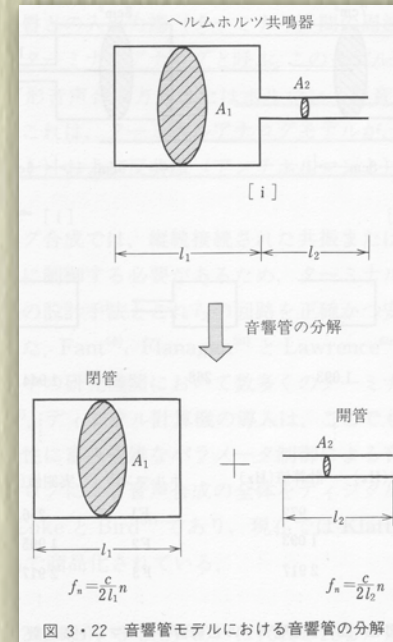
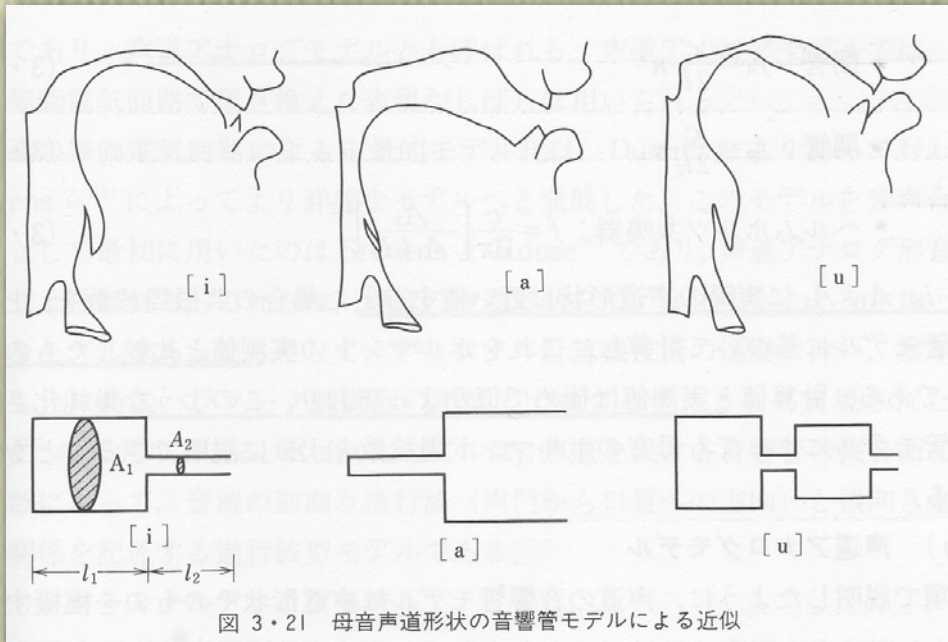
一次聴覚野は経験に応じて“再調整”され、重要な音や楽音に対してはより多くの細胞が最大限に反応するようになる。学習によるこの調整は、大脳新皮質の二次聴覚野や聴覚連合野など、高次の関連領域での処理にも影響を及ぼす。これらの領域は和声や旋律、リズムなど、より複雑な要素を処理していると考えられる。



音楽家が楽器を演奏する際には、運動野や小脳など聴覚野とは別の領域も活性化している。これらの領域は運動をつかさどり、身体の特定期間を正確なタイミングで動かすことに深くかかわっている。

質問 1

Q：管を短くするとどんな声になるのか？



$$f_n = \frac{c}{2\underline{l_1}} n$$

$$f_n = \frac{c}{2\underline{l_2}} n$$

$$f = \frac{c}{2\pi} \left[\frac{A_2}{\underline{A_1 l_1 l_2}} \right]^{1/2}$$

A → B → C

質問 2

Q：管を「あ」から何の形に変えたのか当てなさい。

● この人の「あ」の管の形から変えていきます。

A =



A → B B → A

Replay

本日のメニュー

音声の物理学とその基礎

- 声が生まれる仕組み／音色，ピッチ，とその物理的対象
- 声の多様性（性別，年齢，体格によって変わる様々な声色）

声の共通項（構造的表象）とその発音評定への応用

- 発音学習とは教師の声の何を真似る作業を指すのか？
 - 幼児は親の声の何を真似て言葉を獲得するのか？／音色の相対感覚
 - 声帯模写と発音学習（音声模倣）の違いに立脚した基盤技術構築
- 学習者はどのような発音を目指すべきか？
 - 母語話者のような発音 vs. 伝わる発音
- 構造表象に基づく発音診断ソフトの開発

音声の構造的表象を通して考える発音教育

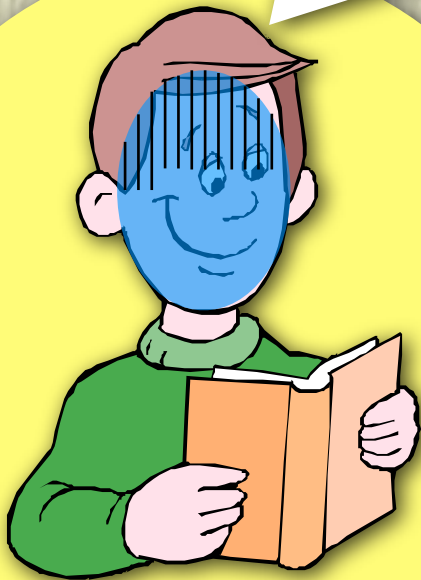
- 見えないもの，意識できないものを捉えるには？

まとめ

発音学習と声帯模写

ベッキーの発音に憧れる男子中学生の例

いや、あの、物真似タレント
目指してるんじゃないんですけど



まずは女声を出すトレーニング
から始めましょう

声の二面性 ～個人性と音韻性～

男子中学生の「あ」とベッキーの「あ」



男子中学生の「あ」とその学生の「い」

- 両方とも声の「音色」の変化でしかない

- 物理的には声の「同一の側面」が異なる方向に変化しただけ

調音（構音）音声学的に考えると・・・

- 男子中学生とベッキーの違い：喉のデフォルト形状の違い

- 男子中学生：より長い，ベッキー：より短い

- ある話者の「あ」と「い」の違い：喉の形状の違い

170cm

S：はろー，はろー

T：はろー，じゃないのよ，Hello よ。

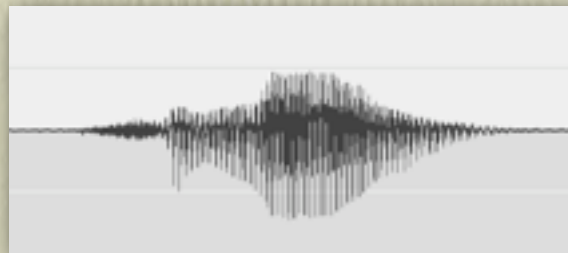
S：はろー，はろー，Hello，Hello・・・

150cm

そんなの簡単じゃん！



[helou]



音声模倣の二面性 ～音真似と？真似～

音声模倣＝親の発声行為を子が積極的に模倣する行為

- これを通して幼児は言語を獲得する[Juszyk'00]。
- 動物学的には非常に稀な行為。霊長類では人間だけ[Gruhn'06]。
- 他の動物では小鳥，クジラ，イルカくらいか[岡ノ谷'08]。

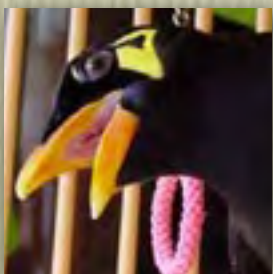
動物の模倣＝声帯模写， ヒトの音声模倣≠声帯模写

● 九官鳥の音声模倣

- 車，ドア，犬，猫，音を真似る。人の声も音でしかない[宮本'95]。
- 良い九官鳥を聞くと，飼い主が分かる[宮本'95]。

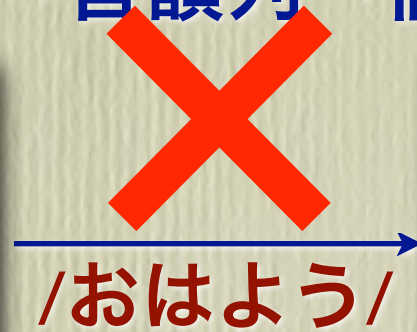
● 幼児の音声模倣

- 動物学的には**奇妙な**模倣行為[岡ノ谷'08]
- いくら良い子でも，声から父親を割り出せずにお巡りさんは困る。



音声模倣の二面性 ～音真似と？真似～

親の発声 → 音韻同定 → 音韻列 → 個々の音韻を発声



- 音韻意識が希薄／しり取りも出来ない[原'03]。

発達心理学からの回答

- 幼児は語全体の音形[加藤'03]， ゲシュタルト[早川'06]， Related spectrum pattern[Lieberman'80]を獲得し， その後， 音韻を獲得
- 語ゲシュタルトには話者の情報は含まれない。話者不変量
 - if not, 幼児は動物のように音声模倣をすることになる。
- 語ゲシュタルトの物理的・音響的定義は何か？
 - 親の声と幼児の声の「物理的な共通項」は何か？



本日のメニュー

音声の物理学とその基礎

- 声が生まれる仕組み／音色，ピッチ，とその物理的対象
- 声の多様性（性別，年齢，体格によって変わる様々な声色）

声の共通項（構造的表象）とその発音評定への応用

- 発音学習とは教師の声の何を真似る作業を指すのか？
 - 幼児は親の声の何を真似て言葉を獲得するのか？／音色の相対感覚
 - 声帯模写と発音学習（音声模倣）の違いに立脚した技術構築
- 学習者はどのような発音を目指すべきか？
 - 母語話者のような発音 vs. 伝わる発音
- 構造表象に基づく発音診断ソフトの開発

音声の構造的表象を通して考える発音教育

- 見えないもの，意識できないものを捉えるには？

まとめ

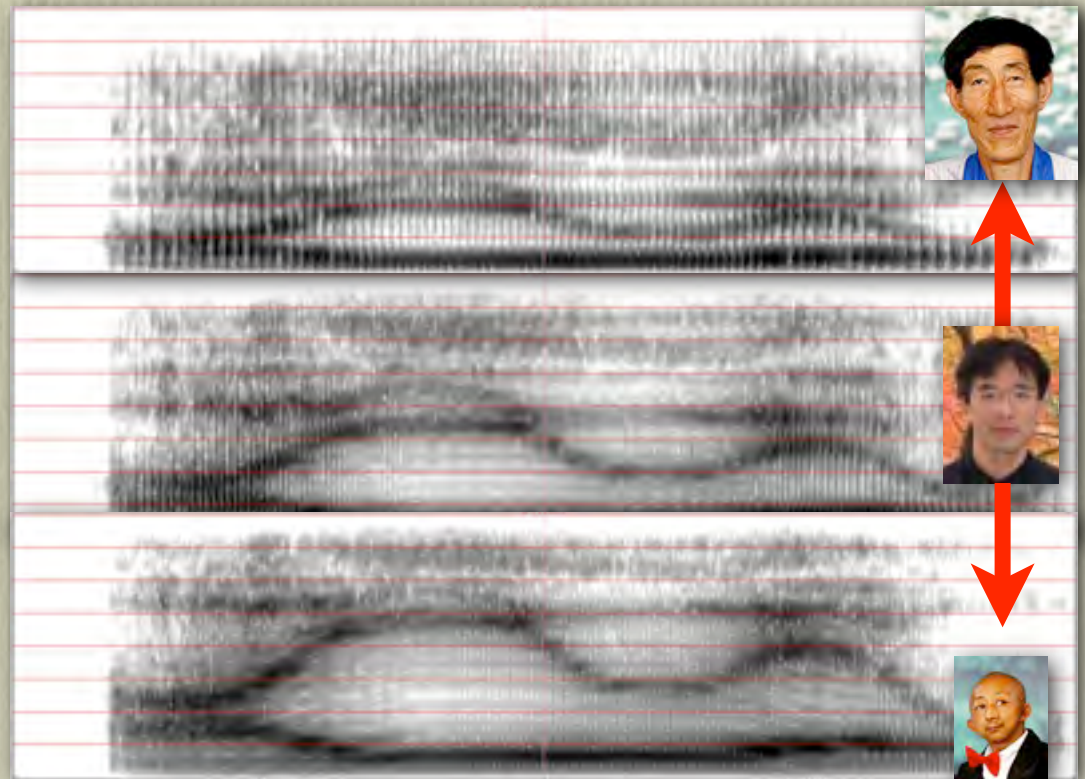
体格・性別・年齢を超えたパターン

音高の個人差を生み出す要因 ～高い低い～

● 男女の音高偏差＝声帯の長さ・重さの性差

音色の個人差を生み出す要因 ～太い細い～

● 男女の音色偏差＝声道の形状（主に長さ）の性差



世界一小さな女子高校生

頭でっかちではない赤ん坊？



Jyoti Amge
World's Smallest Girl

current.com

刺激の物理的多様性とその認知的不変性

感覚受容器が受け取る情報は容易に変貌する

● 見えの変化

- 視点を变えて見た犬
- 対象との距離を変えて見た像

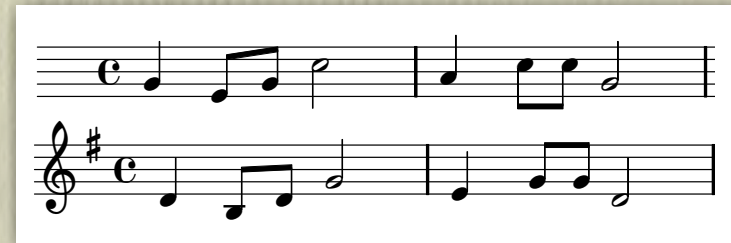
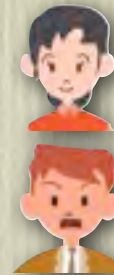


● 色みの変化

- 朝日の花と夕焼け空の花
- 異なる色眼鏡を通して見た像

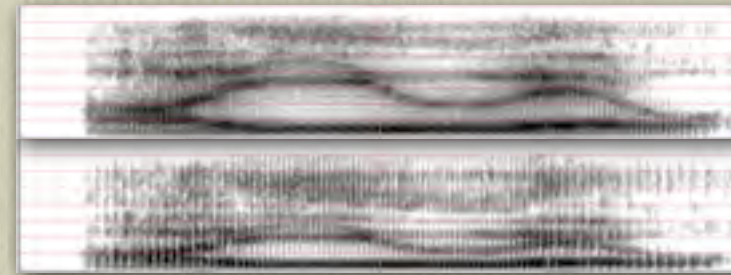
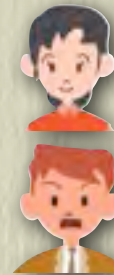
● 音高の変化

- 男性のハミングと女性のハミング
- カラオケでのキーの上げ下げ



● 音色の変化

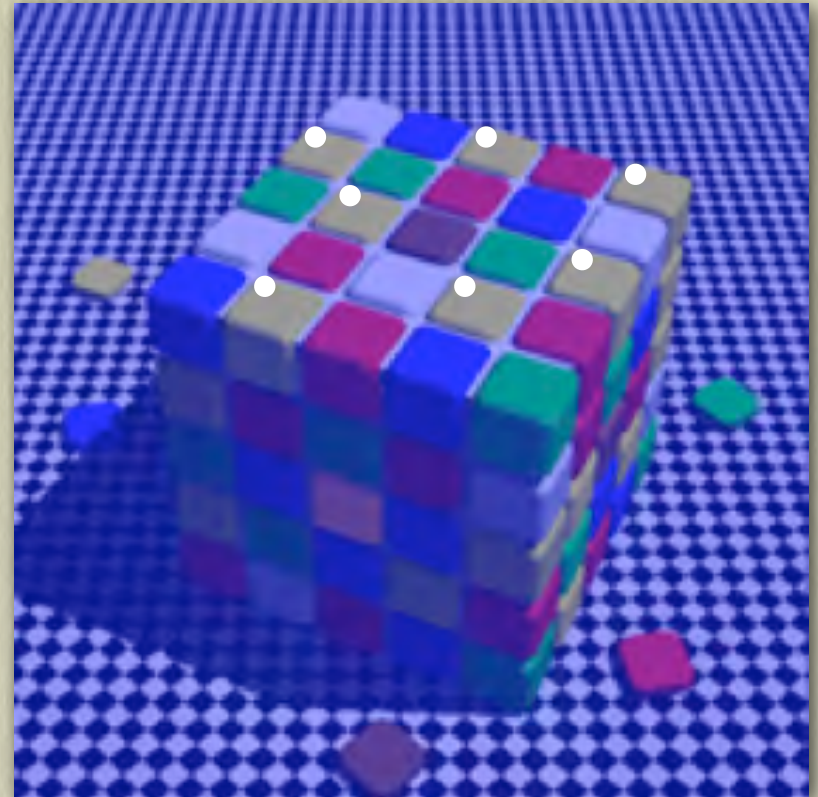
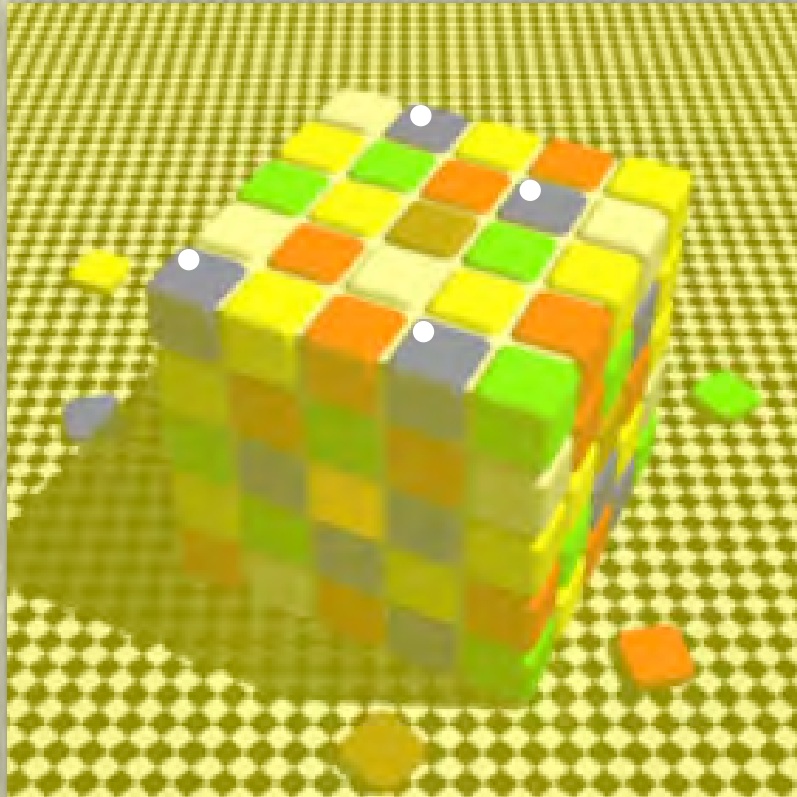
- 男性のおはよう！と女性のおはよう！
- 大人のおはよう！と子供のおはよう！



でも、我々は容易に「同一性」を認知できる

色みの偏差とその認知的不変性

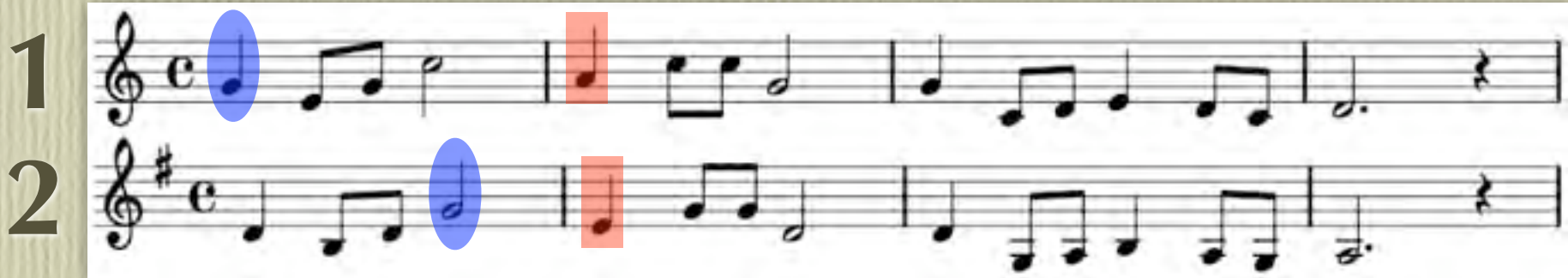
黄・青眼鏡を通して眺めるルービックキューブ [Lotto'99]



- 両者が同一のキューブであることは容易に認知可能
- 異なる色を同一と主張し，同一の色を異なると主張する。
- 各パッチが持つ波長（絶対量）だけではなく，**各パッチが他のパッチ群とどのようなコントラストを持つのか**，が非常に重要

音高の偏差とその認知的不変性

カラオケでキーを上げ下げして曲を聞く [東川'05]



● 絶対音感者（ドレミは**音名**）

● 1 = ソーミソドーラードドソー, 2 = レーシレソーミーソソレー

● 言語化可能な相対音感者（ドレミは**階名**）

● 1 = **ソ**ーミソド**ーラ**ードドソー, 2 = ソーミソ**ド**ー**ラ**ードドソー

● 言語化困難な相対音感者（**ラーラ**音感者）

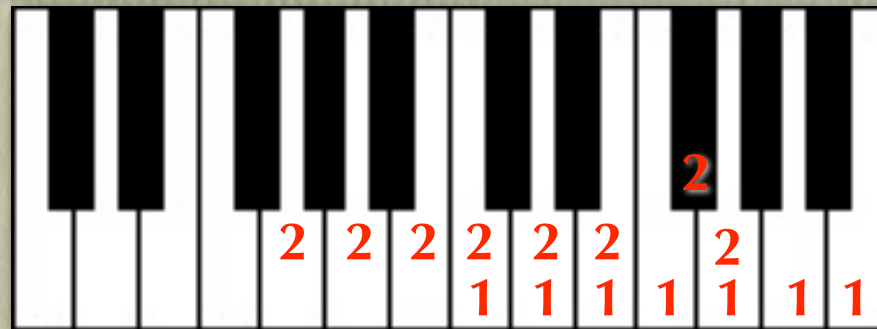
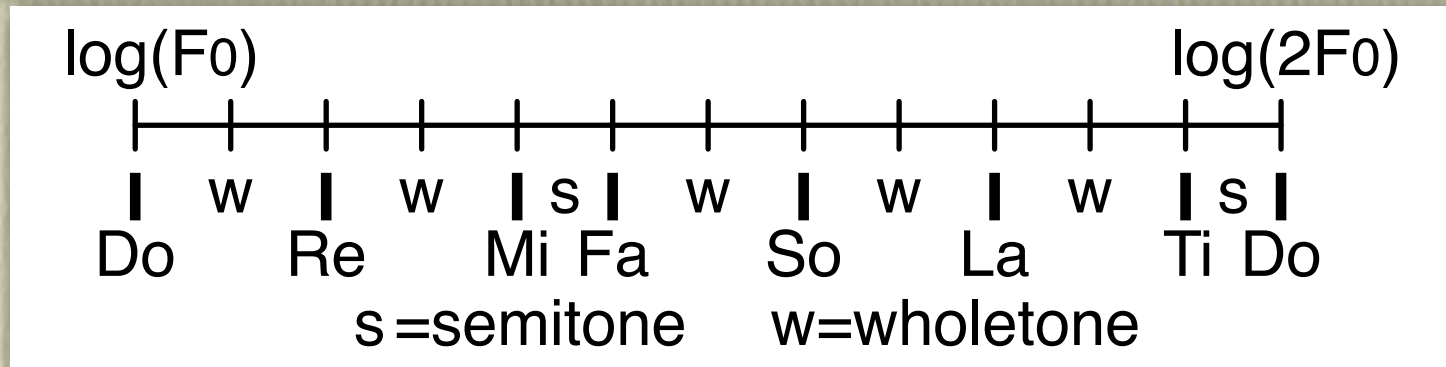
● 1 = ラーラララーラーラララー, 2 = ラーラララーラーラララー

● 異なる音を同一と主張し, 同一の音を異なると主張する。

● 各音が持つ基本周波数（絶対量）ではなく, **各音が他の音群とどのようなコントラストを持つのか**, のみによって決定

音高の偏差とその認知的不変性

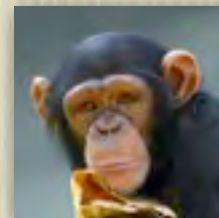
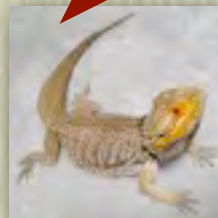
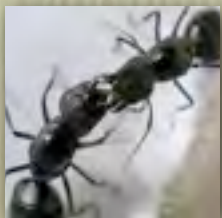
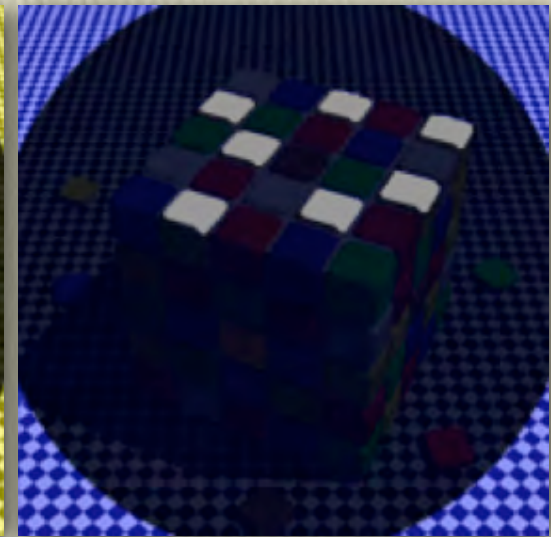
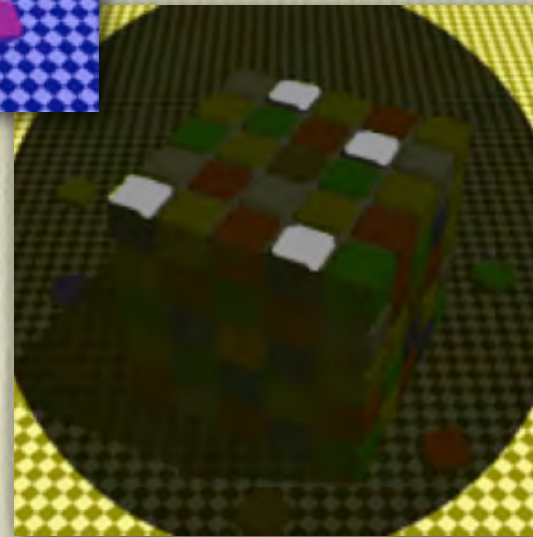
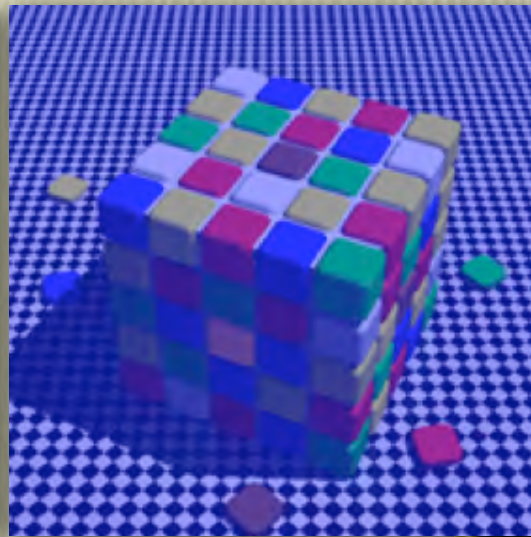
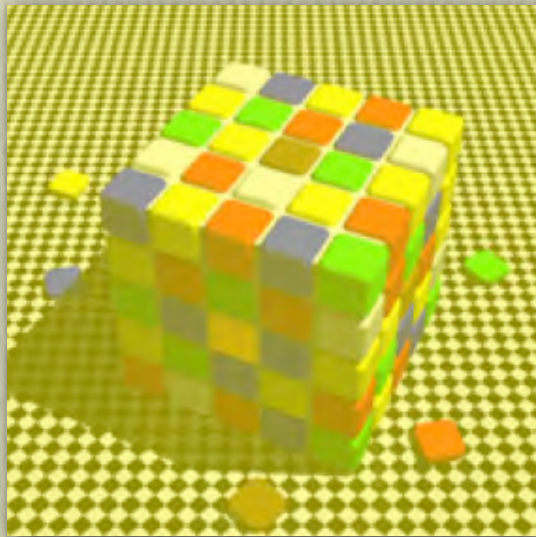
カラオケでキーを上げ下げして曲を聞く [東川'05]



🔔 各音が持つ基本周波数（絶対量）ではなく、**各音が他の音群とどのようなコントラストを持つのか**、のみにによって決定

生物が獲得した静的バイアス除去術

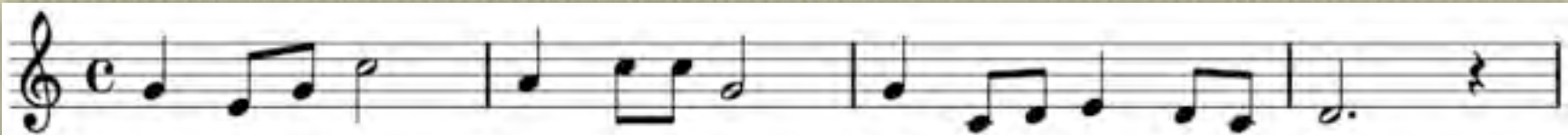
色の恒常的知覚はどこまで遡れるのか？ [Briscoe'01]



生物が獲得した静的バイアス除去術

音高の恒常的知覚はどこまで遡れるのか？ [Hauser'03]

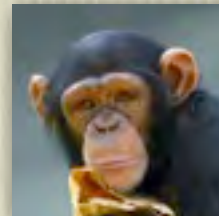
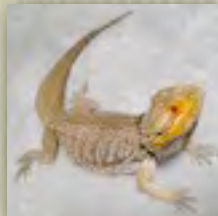
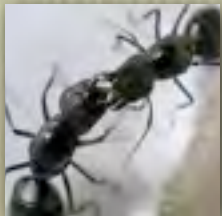
1



2



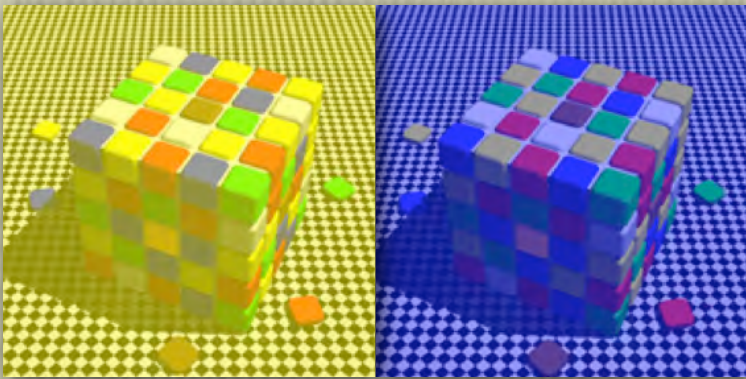
1 = 2



音色の偏差とその認知的不変性

色み・音高の恒常・不変的認知

- コントラスト情報に基づく処理が重要
- コントラスト群から成る全体的パターン処理が要素同定を可能に

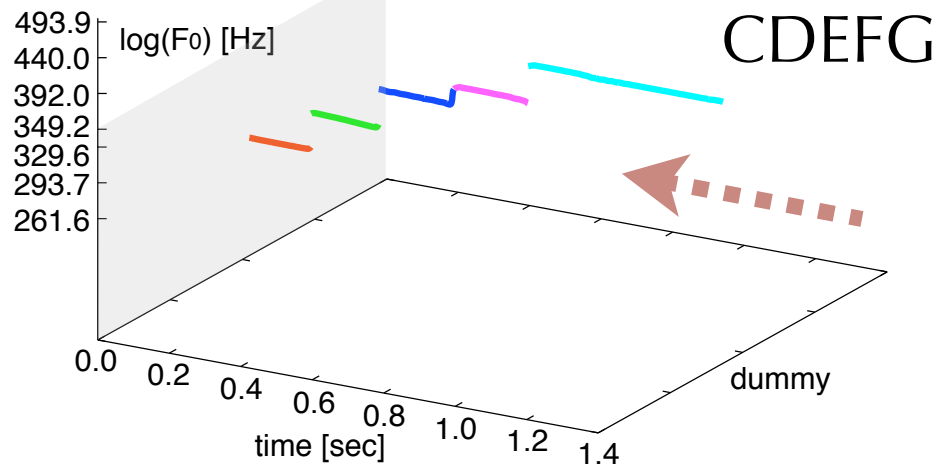
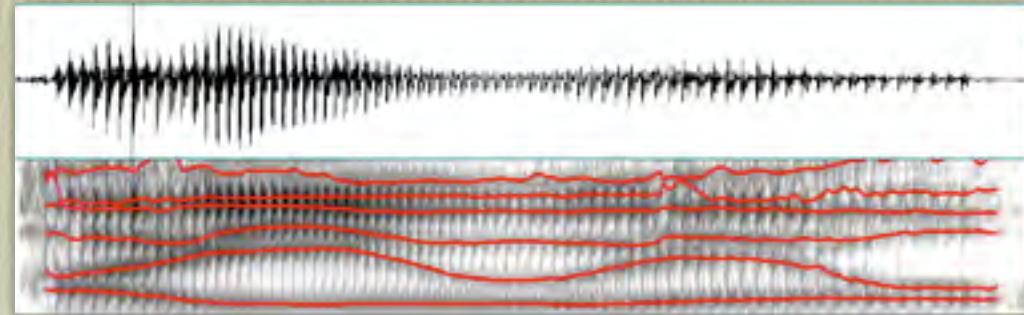
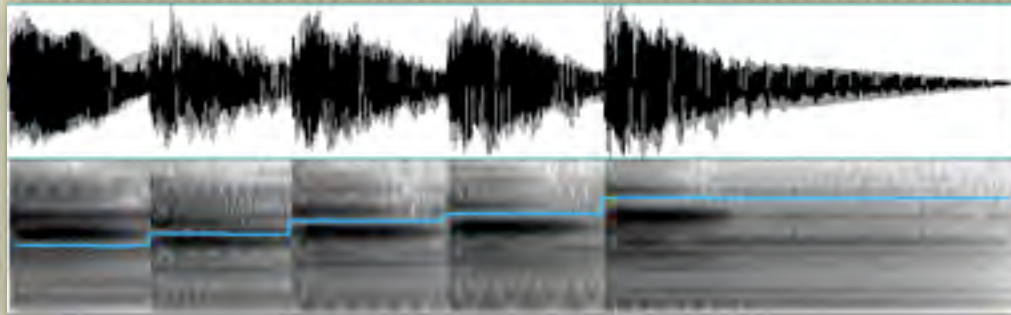


音色の恒常・不変的認知

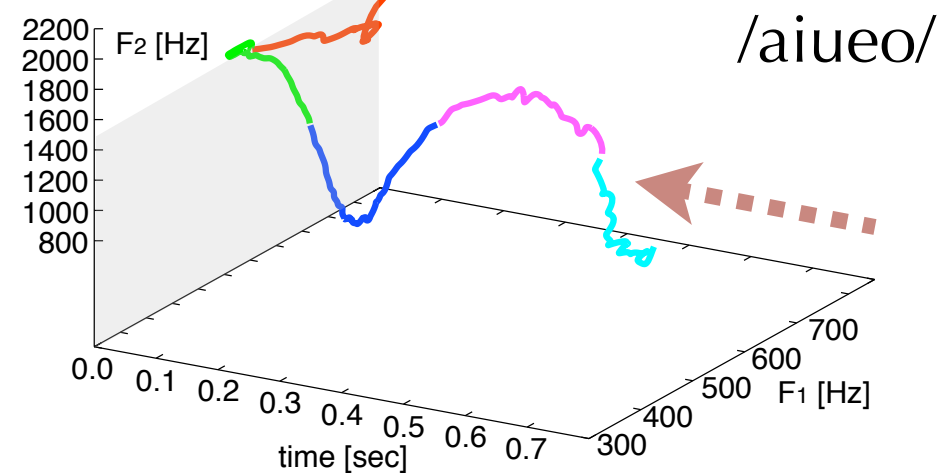
- コントラスト情報に基づく処理が重要
- コントラスト群から成る全体的パターン処理が要素同定を可能に



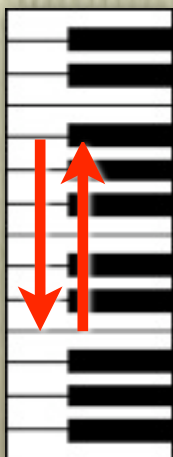
音声の構造的表象／音色の相対感覚



音高の動的変化パターン

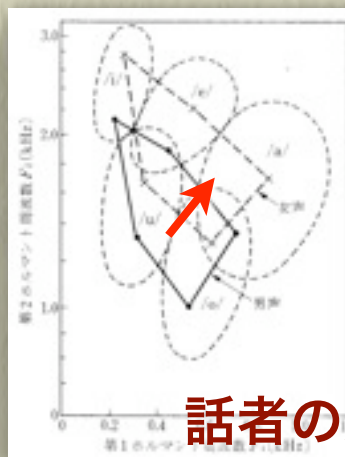
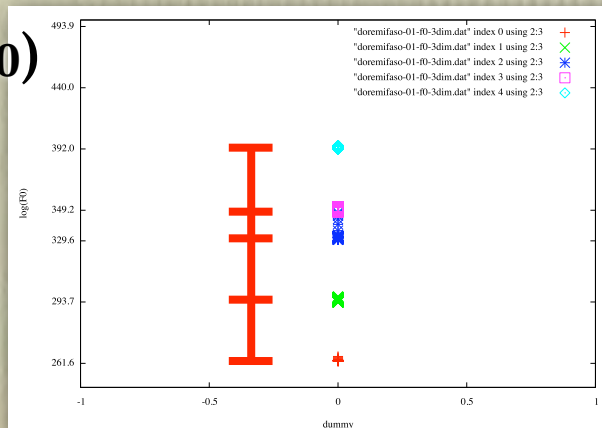


音色の動的変化パターン

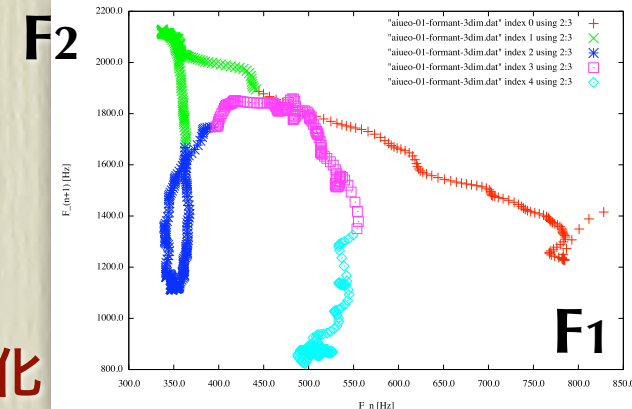


調の変化

$\log(F_0)$

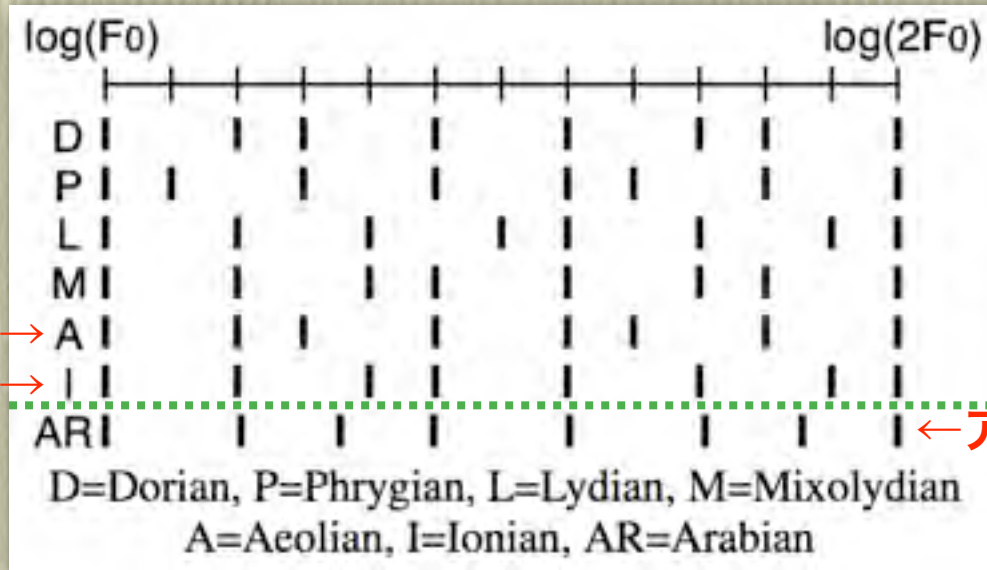


話者の変化



音声の構造的表象／音色の相対感覚

音楽における調不変の音配置とその変種

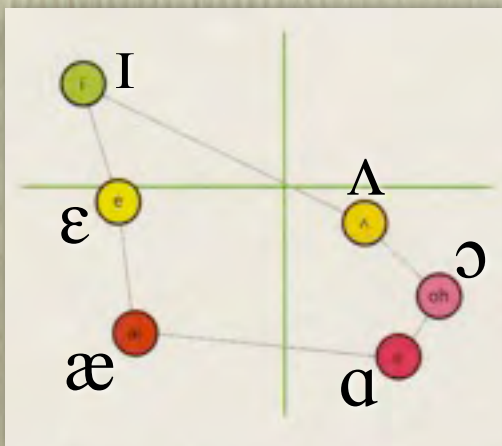


- 西洋音楽 = 5全音 + 2半音
- 種々の配置 = 教会音楽
- 民族音楽には半音以外の配置

أهلاً وسهلاً



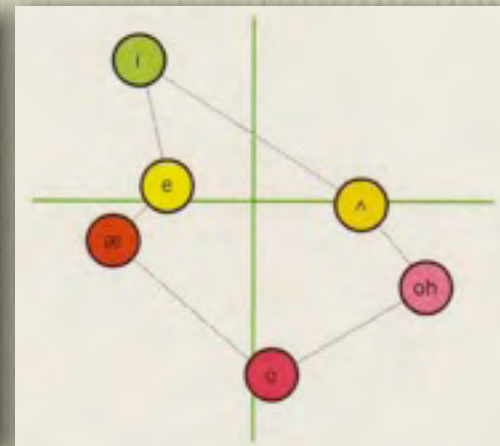
音声における話者不変の音配置とその変種 = 欧米の方言



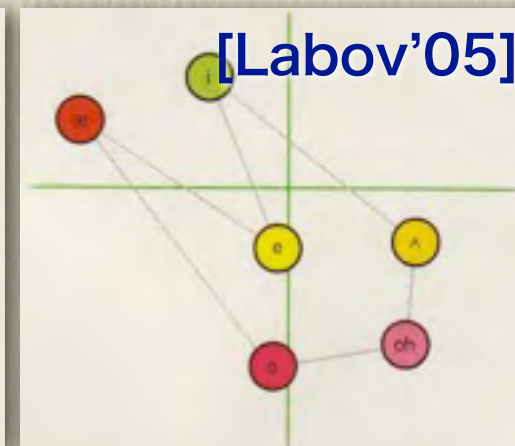
Williamsport, PA



Chicago, IL



Ann Arbor, MI



Rochester, NY

[Labov'05]

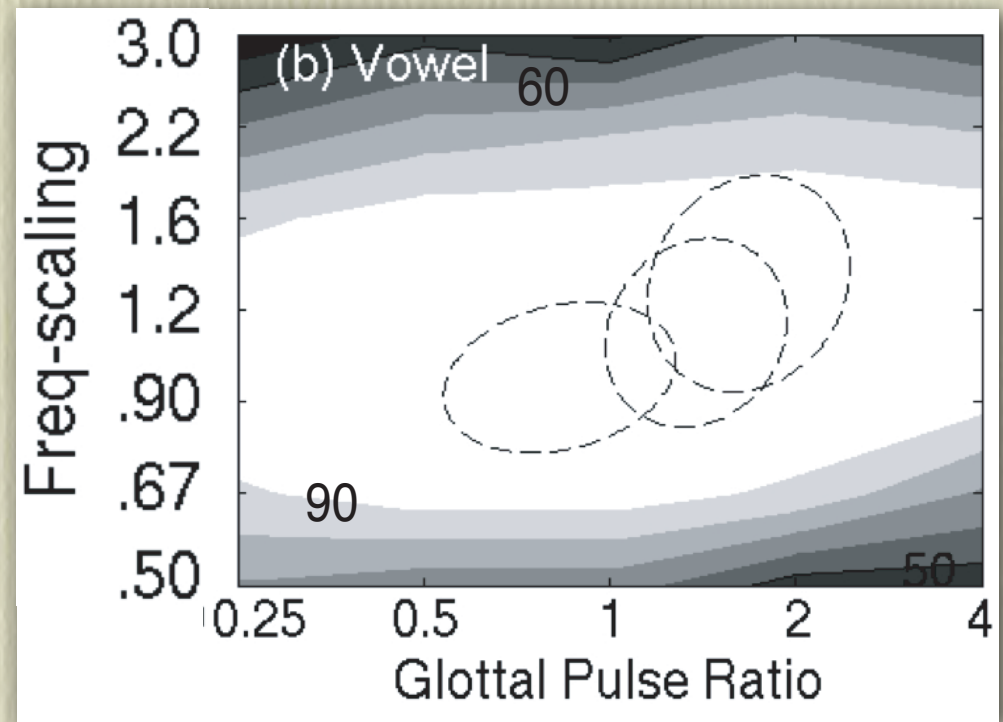
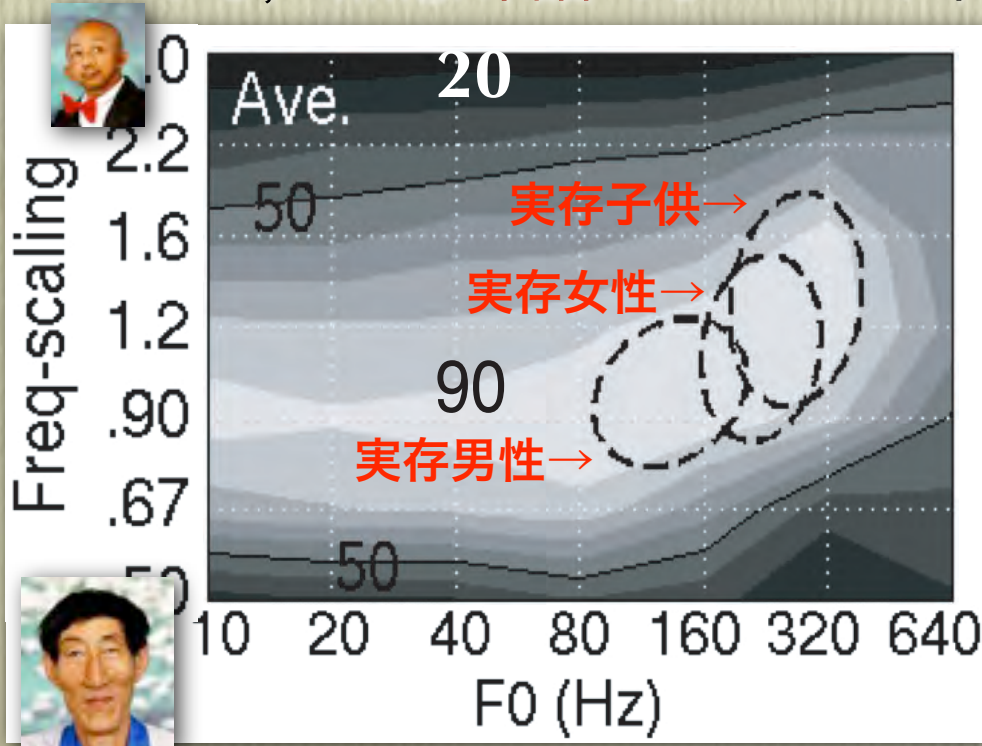
音声の構造的表象／音色の相対感覚

言語化できる相対音感者が出来ないこと

- 孤立的に提示された音をドレミ同定することは出来ない。
- 孤立的に提示された音を母音同定できない人などいるのか？

巨人&小人の音声を使った母音同定・単語同定実験

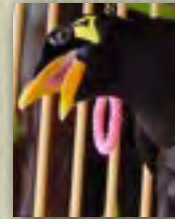
- 孤立母音の同定は困難になる[青木'04]。
- でも、無意味語でよいので単語音声だと書き起こせる[林'07]。



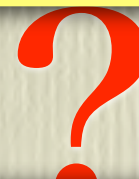
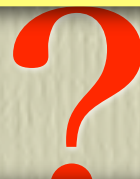
興味深い思考実験を一つ

一卵性双生児が生まれた直後に両親が離婚した・・・

- 一人ずつ引き取られた。
- 彼らは10年後どんな発音をしているのだろうか？



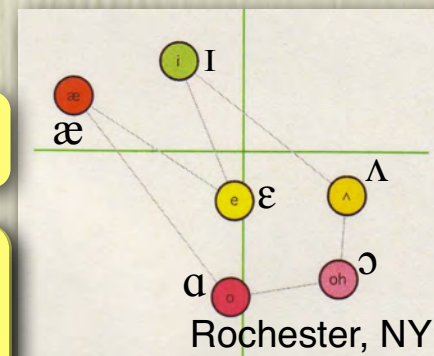
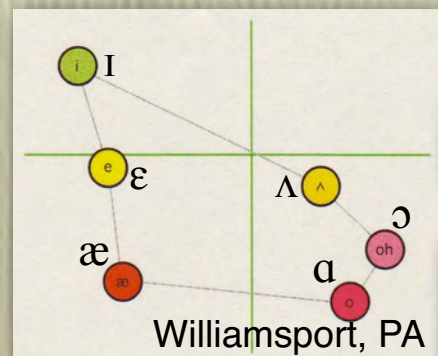
声道形状の性差 = 音色の差異



方言差 = 音色の差異

発音評定技術 =

学習者・教師間での音体系比較



本日のメニュー

音声の物理学とその基礎

- 声が生まれる仕組み／音色，ピッチ，とその物理的対象
- 声の多様性（性別，年齢，体格によって変わる様々な声色）

● 声の共通項（構造的表象）とその発音評定への応用

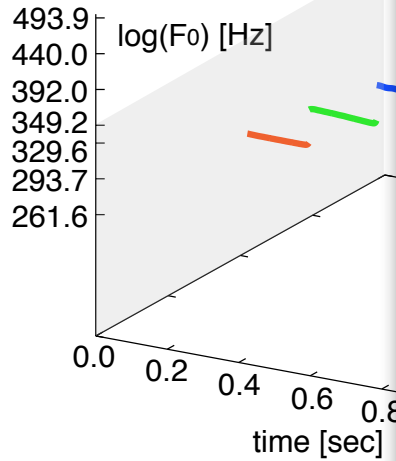
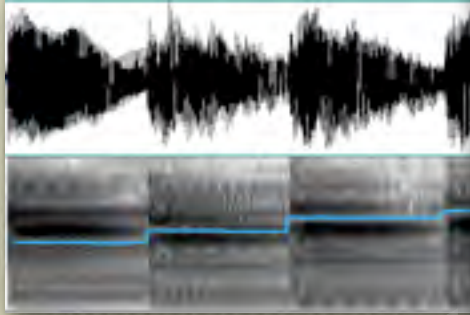
- 発音学習とは教師の声の何を真似る作業を指すのか？
 - 幼児は親の声の何を真似て言葉を獲得するのか？／音色の相対感覚
 - 声帯模写と発音学習（音声模倣）の違いに立脚した基盤技術構築
- 学習者はどのような発音を目指すべきか？
 - 母語話者のような発音 vs. 伝わる発音
- 構造表象に基づく発音診断ソフトの開発

音声の構造的表象を通して考える発音教育

- 見えないもの，意識できないものを捉えるには？

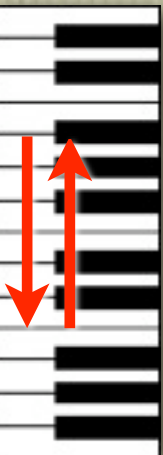
まとめ

音声の構造的な特徴 / 文脈の相対音感

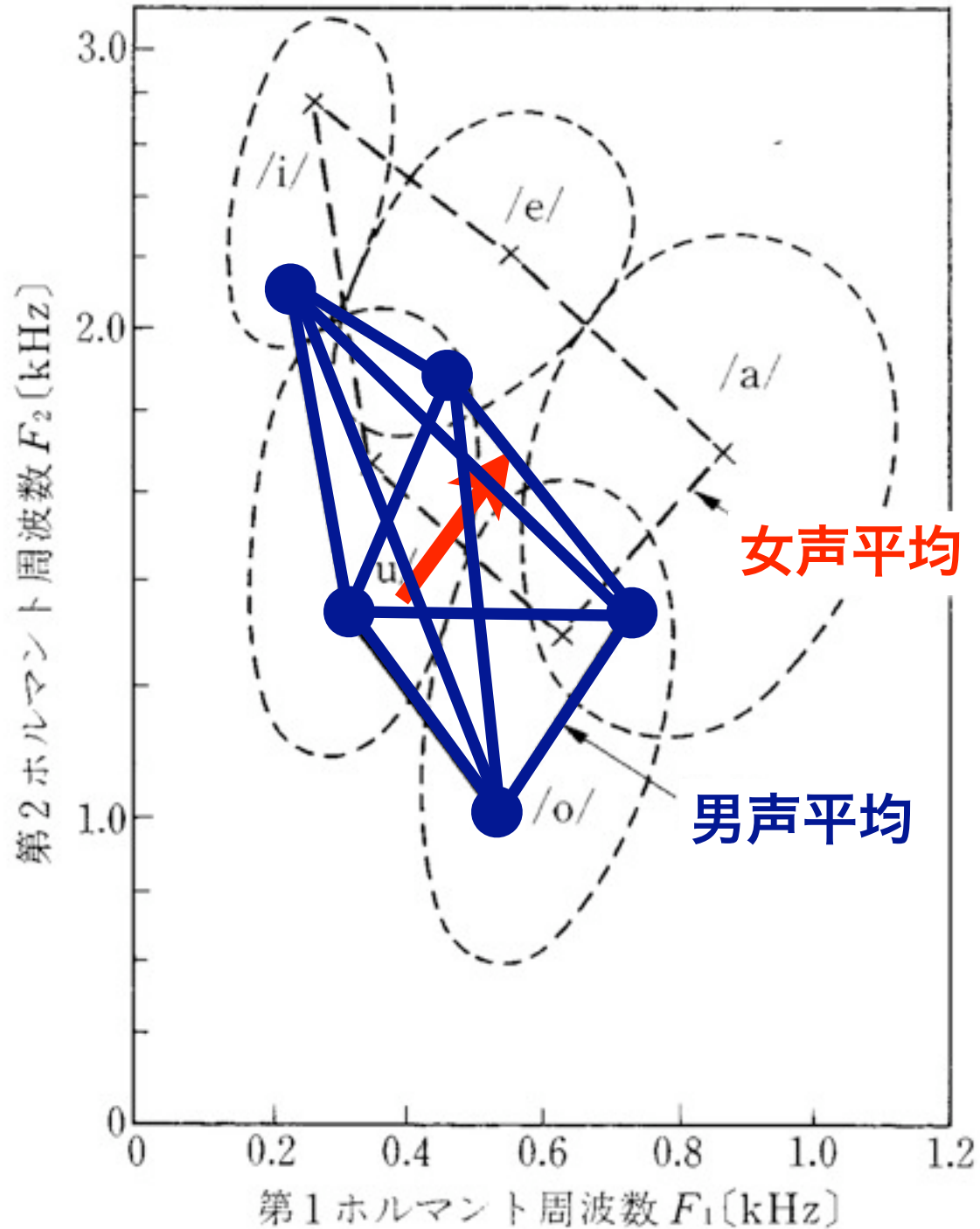


音高

$\log(F_0)$

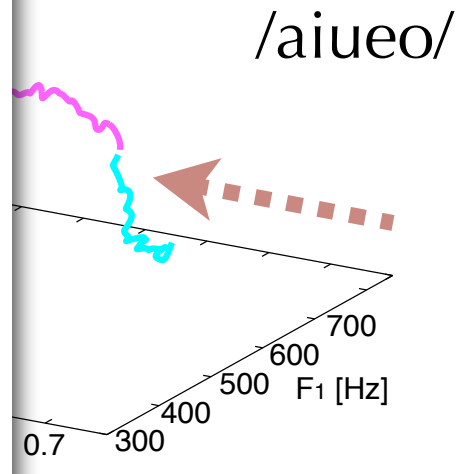
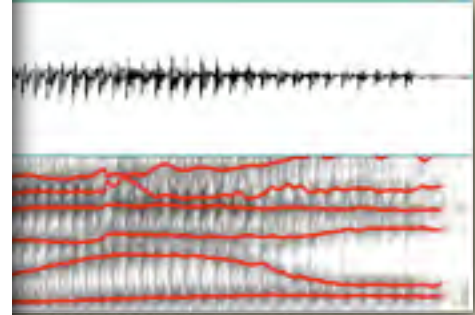


調の変化

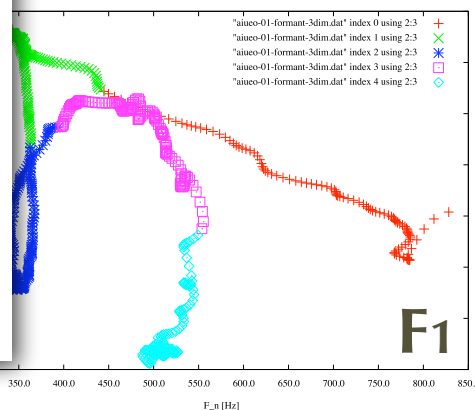


第2ホルマント周波数 F_2 [kHz]

第1ホルマント周波数 F_1 [kHz]



動的変化パターン

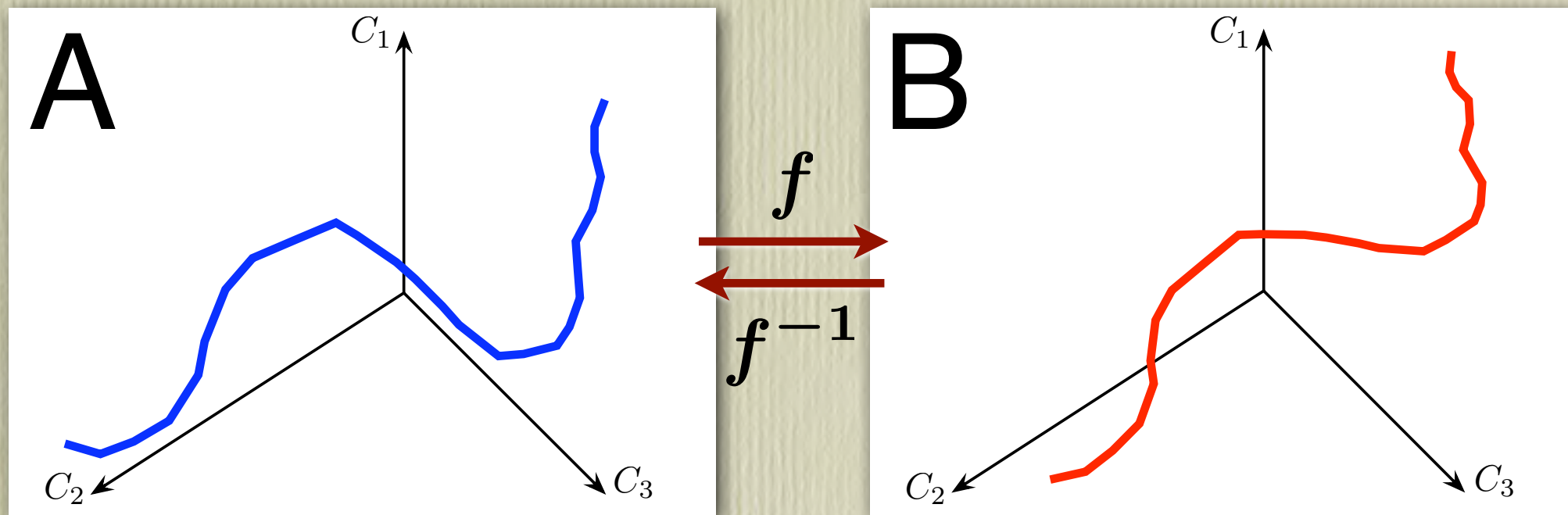


話者の変化

変換不変な音響量の数学的探求

話者の違い＝空間写像（話者・声質変換）

話者 A の音響空間 $\longleftrightarrow$ 話者 B の音響空間



話者 A の声を65億全ての話者の声へ変形する

65億 x 65億 の写像関数が定義可能

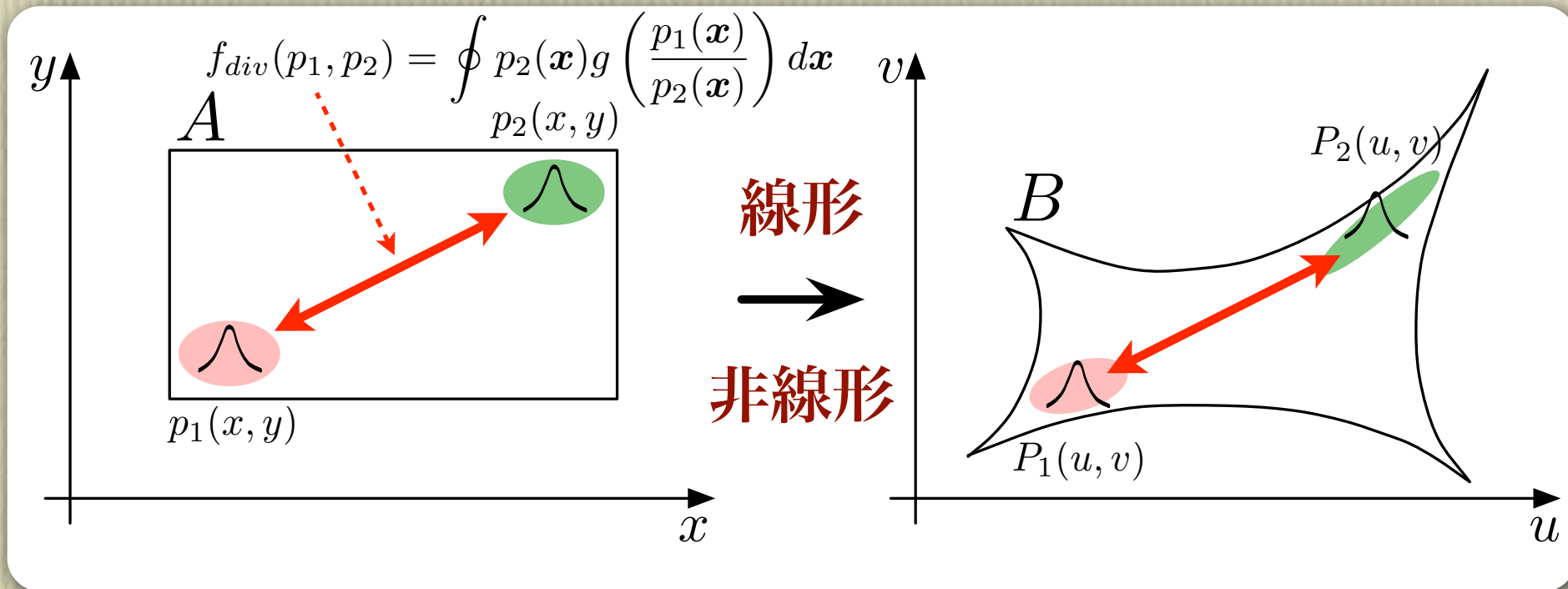
話者不変のコントラスト＝写像不変のコントラスト

任意の写像に対して不変なるコントラスト量は存在するのか？

変換不変な音響量の数学的探求

話者の違い＝空間写像（話者・声質変換）

話者 A の音響空間 $\longleftrightarrow$ 話者 B の音響空間



話者 A の声を65億全ての話者の声へ変形する

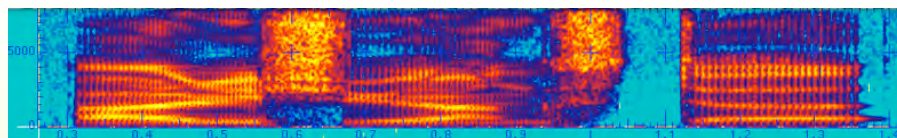
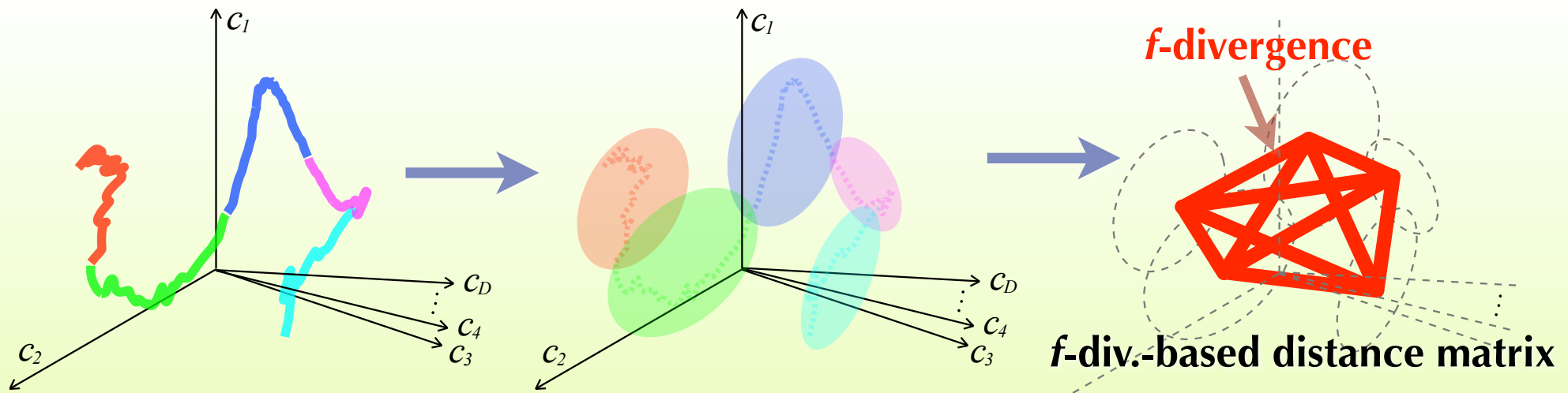
65億 x 65億 の写像関数が定義可能

話者不変のコントラスト＝写像不変のコントラスト

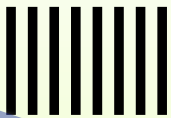
任意の写像に対して不変なるコントラスト量は存在するのか？

体格・性別・年齢を超えたパターン

f-div. に基づく一発声の構造化



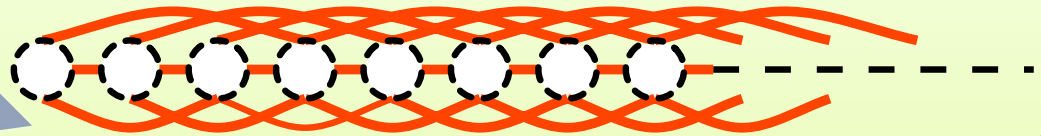
spectrogram (spectrum slice sequence)



cepstrum vector sequence



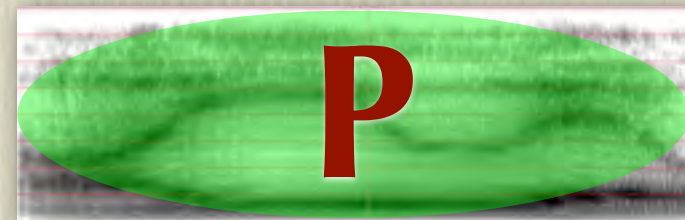
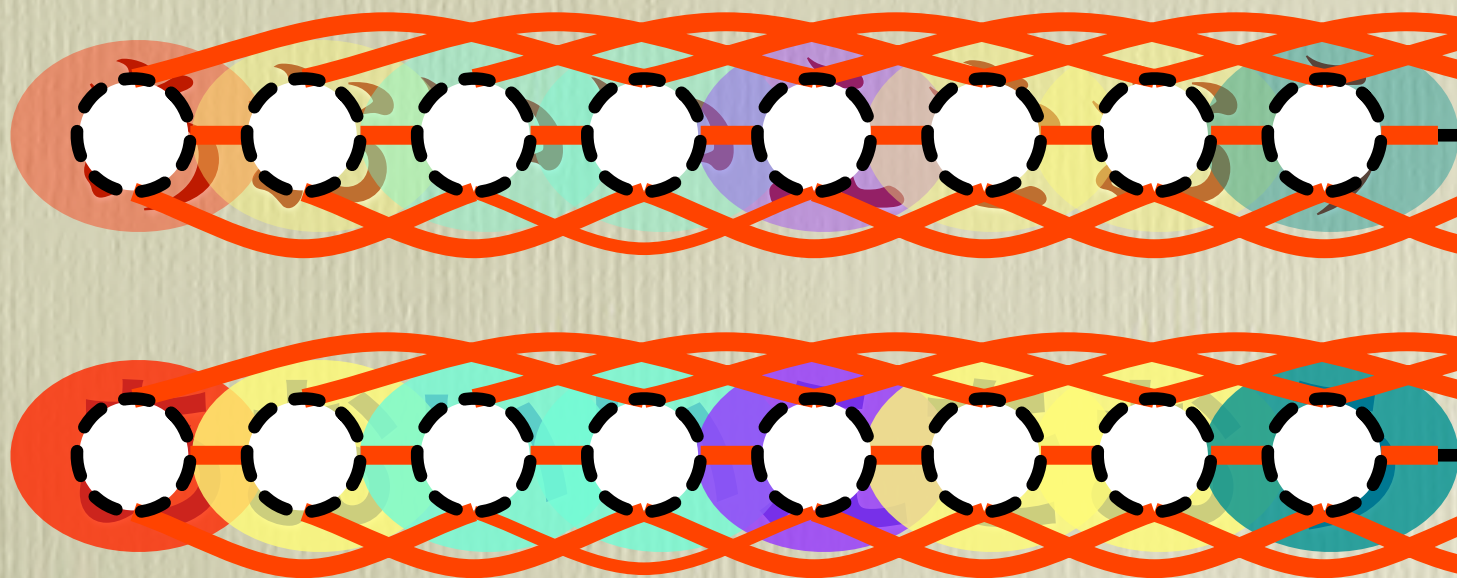
distribution sequence



音色の動きをモデル化する？

喉が音色を作る／声帯が音高を作る

- 喉形状を変えると音色が色々変わる
 - 「あおいえをおう」と「あおいえをおう」
- 声帯の締め具合を変えると音高が色々変わる



ちょっと面白いデモ

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.

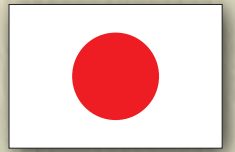
ちょっと面白いデモ

母語話者発音と日本語的峯松発音の音響モデルを利用

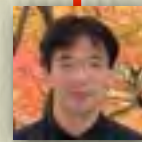
● 各モデルから峯松英語発声が観測される確率 $P(o|M)$



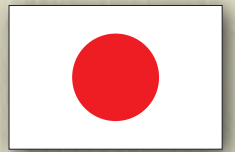
USA/F12



Minematsu
(Japanized)



USA/M08



Minematsu
(Japanized)



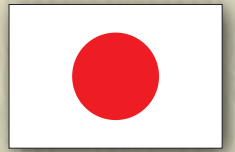
ちょっと面白いデモ

母語話者発音と日本語的峯松発音の音響モデルを利用

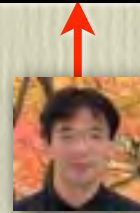
● 峯松発声が、所望の音素列の読み上げである確率 $P(M|o)$



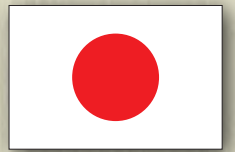
USA/F12



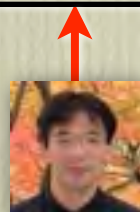
Minematsu
(Japanized)



USA/M08



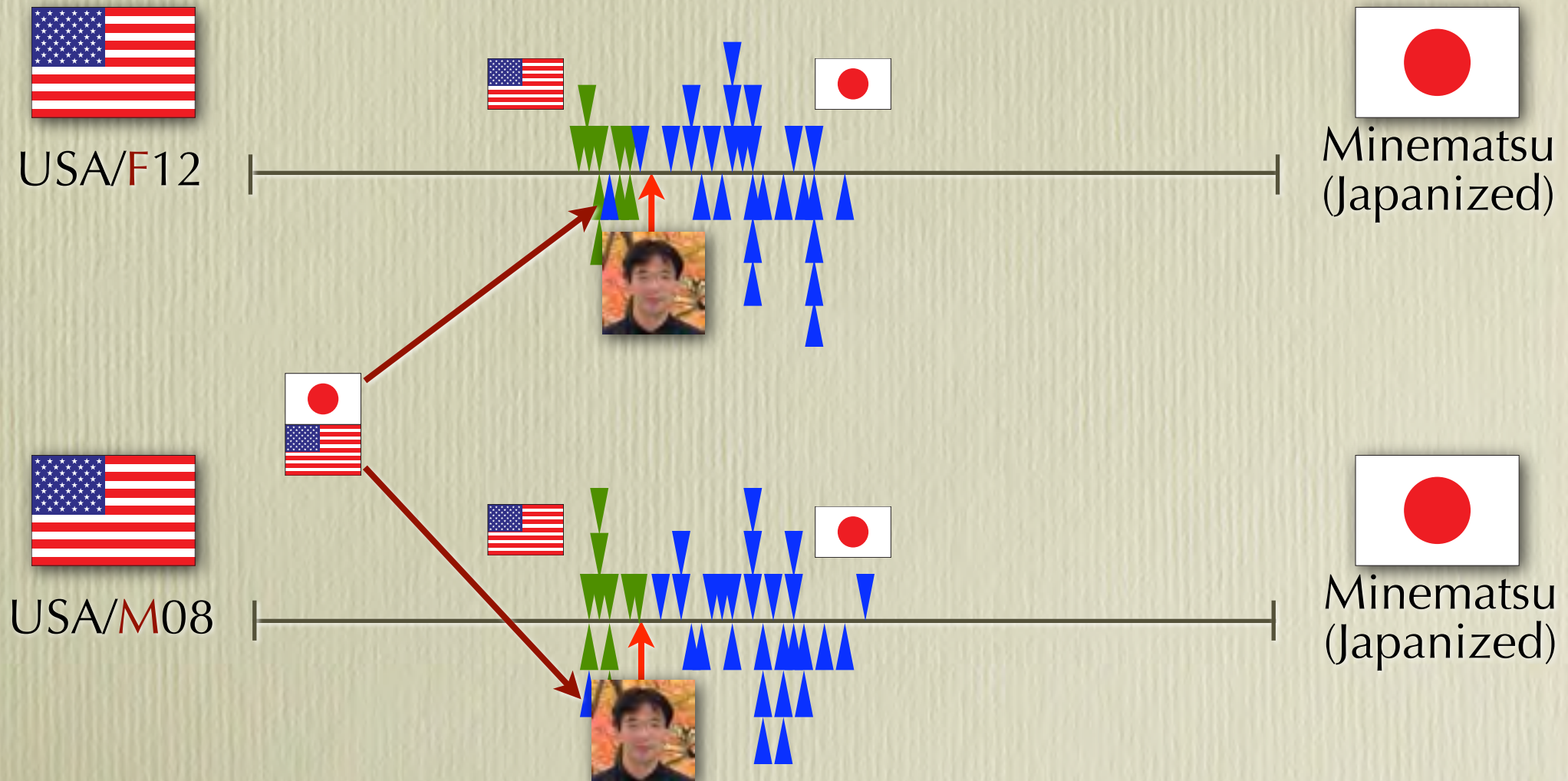
Minematsu
(Japanized)



ちょっと面白いデモ

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



proficiency

perfect

good

Japanized

とある業者の小学生用CALL

母語話者小学生の音声データベースの構築

Benesse®

68W1KD-P
おうちのかたへ 保存版

通じる英語を、毎日そだてる。
BE-GO

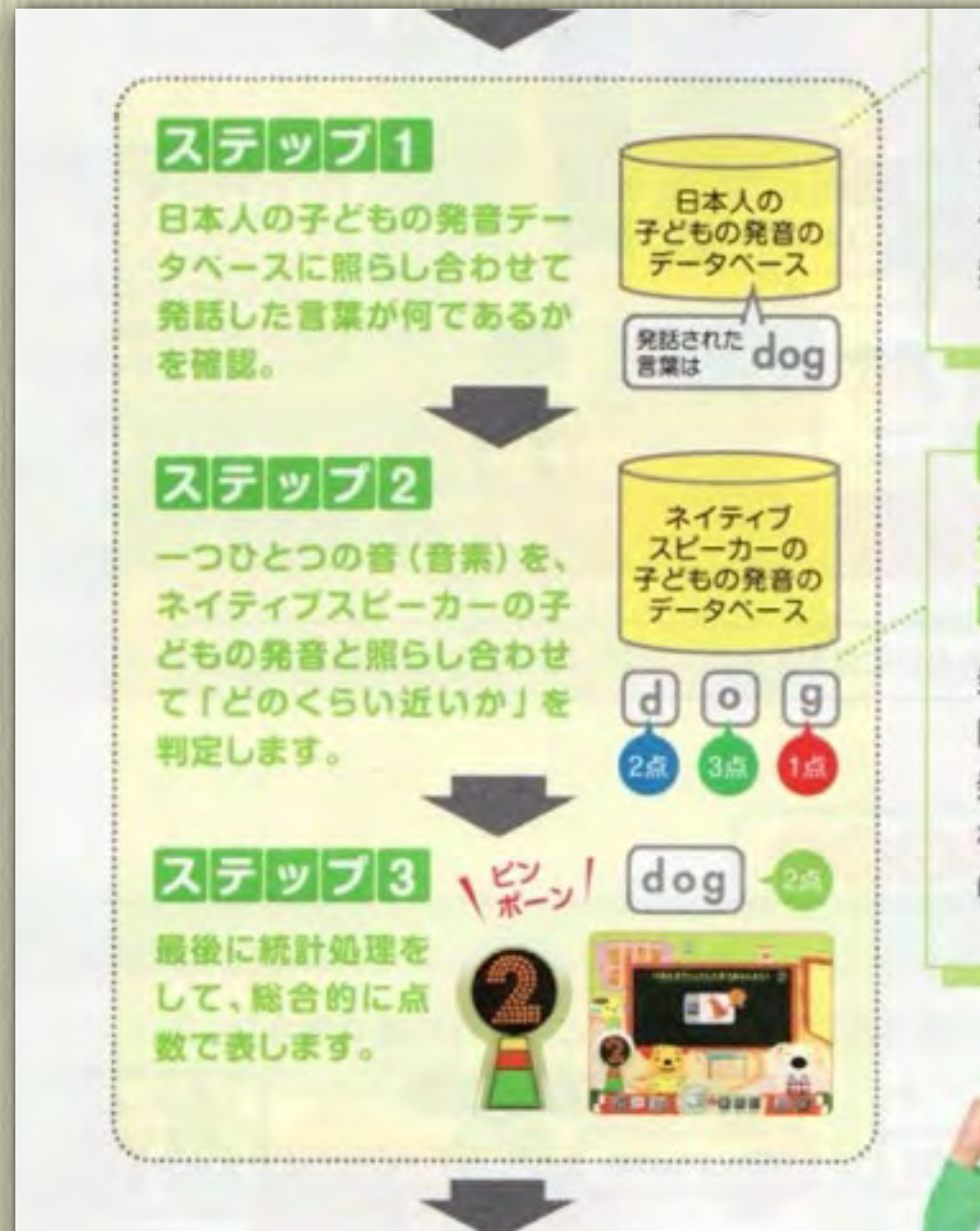
英語の耳を磨く
通じる発音
べーごの英語学習
学習プログラム
お届けする教材

「将来に向けて役立つ英語力」のために、今だからこそ育てる
「英語の耳」と「通じる発音の基礎力」を
しっかり身につける
BE-GO First Friends

ネイティブスピーカーの英語を、ゲーム形式で楽しく・たくさん「聴いて話す」学習で、
発音練習に合わせて力を積み上げる「新しい英語学習」です！

この冊子で
ご紹介している
ゲームの一部を
体験していただけます。
お子さまと一緒に
ぜひお試しください！

無料
おためし
CD-ROM
入り



本日のメニュー

音声の物理学とその基礎

- 声が生まれる仕組み／音色，ピッチ，とその物理的対象
- 声の多様性（性別，年齢，体格によって変わる様々な声色）

● 声の共通項（構造的表象）とその発音評定への応用

- 発音学習とは教師の声の何を真似る作業を指すのか？
 - 幼児は親の声の何を真似て言葉を獲得するのか？／音色の相対感覚
 - 声帯模写と発音学習（音声模倣）の違いに立脚した基盤技術構築
- 学習者はどのような発音を目指すべきか？
 - 母語話者のような発音 vs. 伝わる発音
- 構造表象に基づく発音診断ソフトの開発

音声の構造的表象を通して考える発音教育

- 見えないもの，意識できないものを捉えるには？

まとめ

母語話者の発音 vs. 伝わる発音



どのような発音を身につけたいのか？

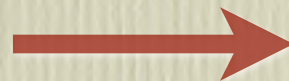
目指すべき発音は多種多様

- ハリウッドスターを目指す日本人女優
- 外交官，国連事務官，海外在駐員
- 頻繁に国際会議にて発表する研究者
- ドイツ出張が日常茶飯事の日本人ビジネスマン
- 海外旅行で買い物ができればよい観光客
- :



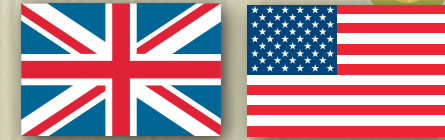
目標とする発音（モデル話者）は学習者が決めるべき

- あの人のような発音を身につけたい。
- あの人レベルになれば，まあ，仕事上問題無い。



学習対象言語は英語？米語？

英語CALL・米語CALLシステム

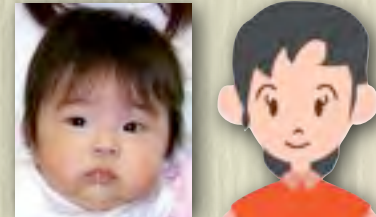


- 英国で子供の声を集める／米国で子供の声を集める

どの地域の英語（発音）を学習対象とすべきか？

- これは正しい問いの立て方だろうか？
- そもそも「母国語の獲得」＝「その国の言語の獲得」なのか？

母国語＝mother tongue



- 国や地域が言語を獲得させる訳では無い。
- 母親（父親）とのコミュニケーションが言語を獲得させる。
 - 「母国語の獲得」＝「お母さん語」「母親の発音」の獲得
 - 母親の発音の癖（訛り）も獲得することになる

結局「誰の」発音を真似るのか，真似たいのか？

- 真似ると言っても，声帯模写ではない。
- 年齢・性別・体格を越えた音パターンの獲得

2011年より始まる小学校での英語教育

小学校 5, 6 年生を対象とした「話す・聞く」教育

- 240万人の子どもたちが一斉に英語を学習し始める。
- 英語教師の補充はどうなっているのか？
 - 8万人の英語教師が必要
 - 文科省「学級担任に授業の中核となって生徒に接してもらう」
- 文字は使わない・教えない「話す・聞く」教育を
 - 学級担任は果たして話せるのか？ 聞けるのか？
- 自習型副教材の導入はあるのか？
 - マルチメディアドリル？ 音声技術の導入？

実はこの子も・・・

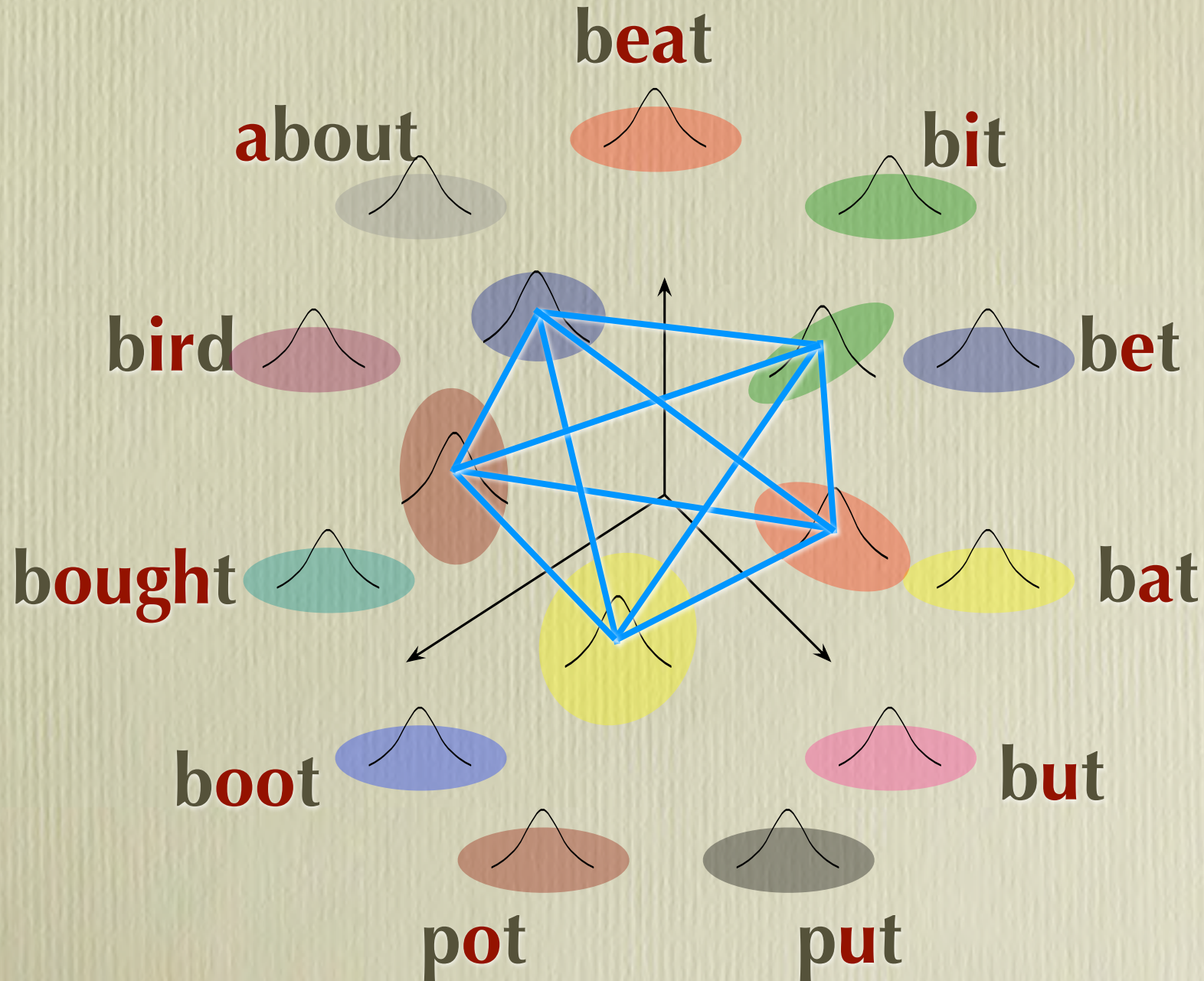


5



英語・米語母音の発音トレーニング

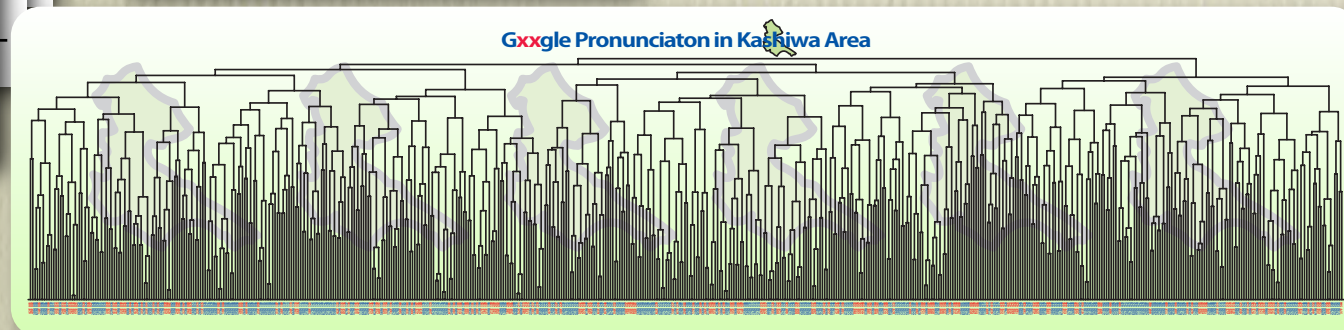
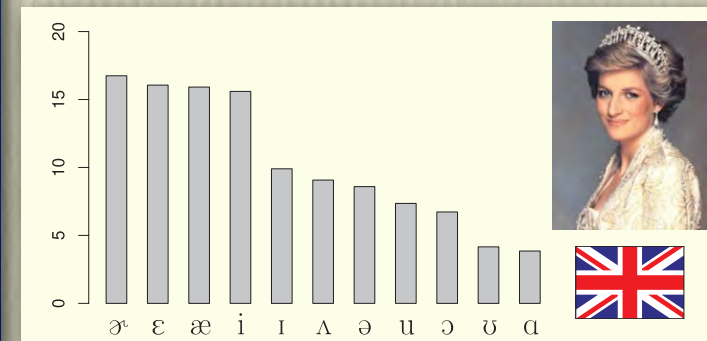
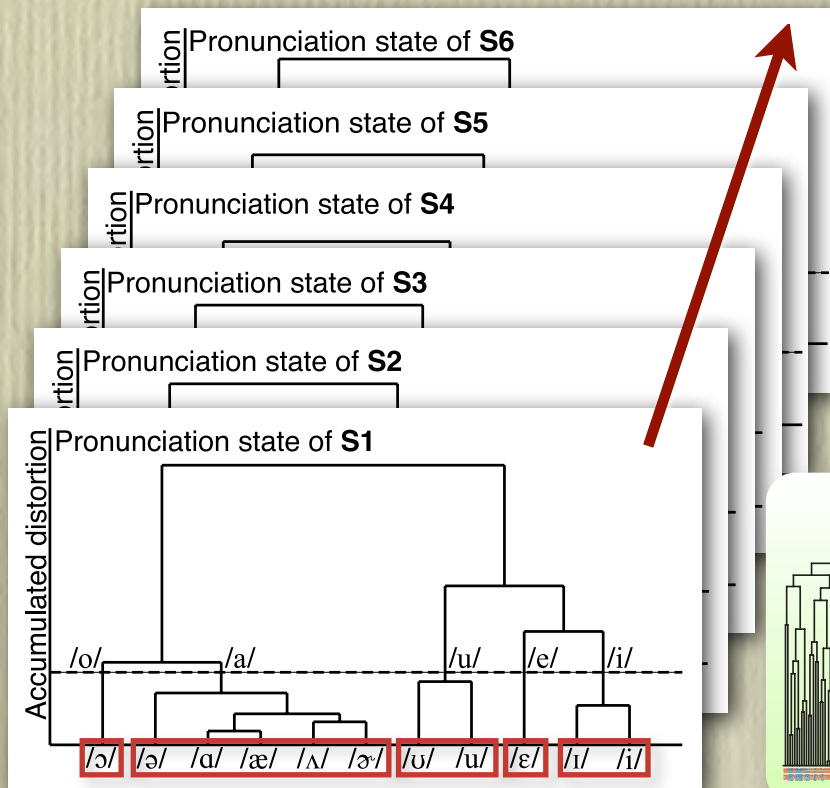
個々の母音ではなく，母音群の体系を評価対象



英語・米語母音体系を対象としたCALL

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

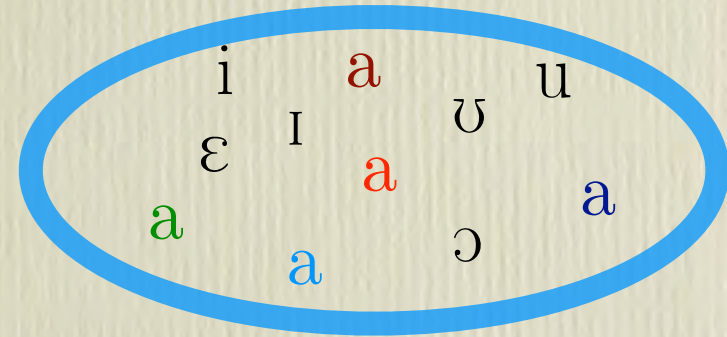
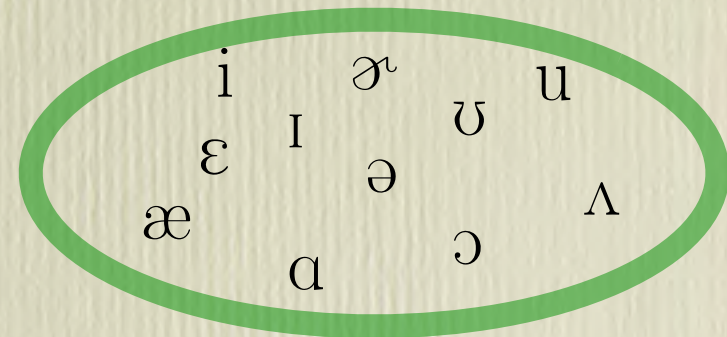
ばと, びと, ぶと, べと, ぼと x 5

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

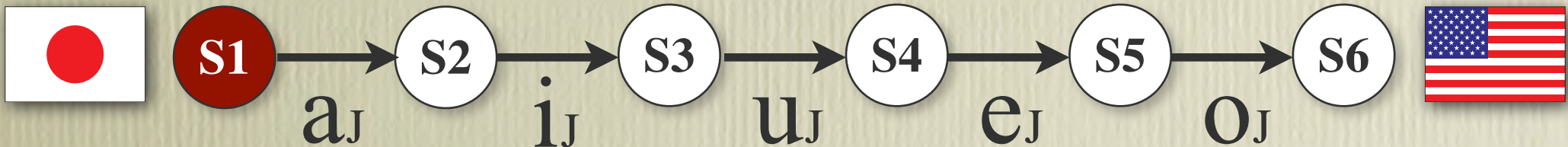
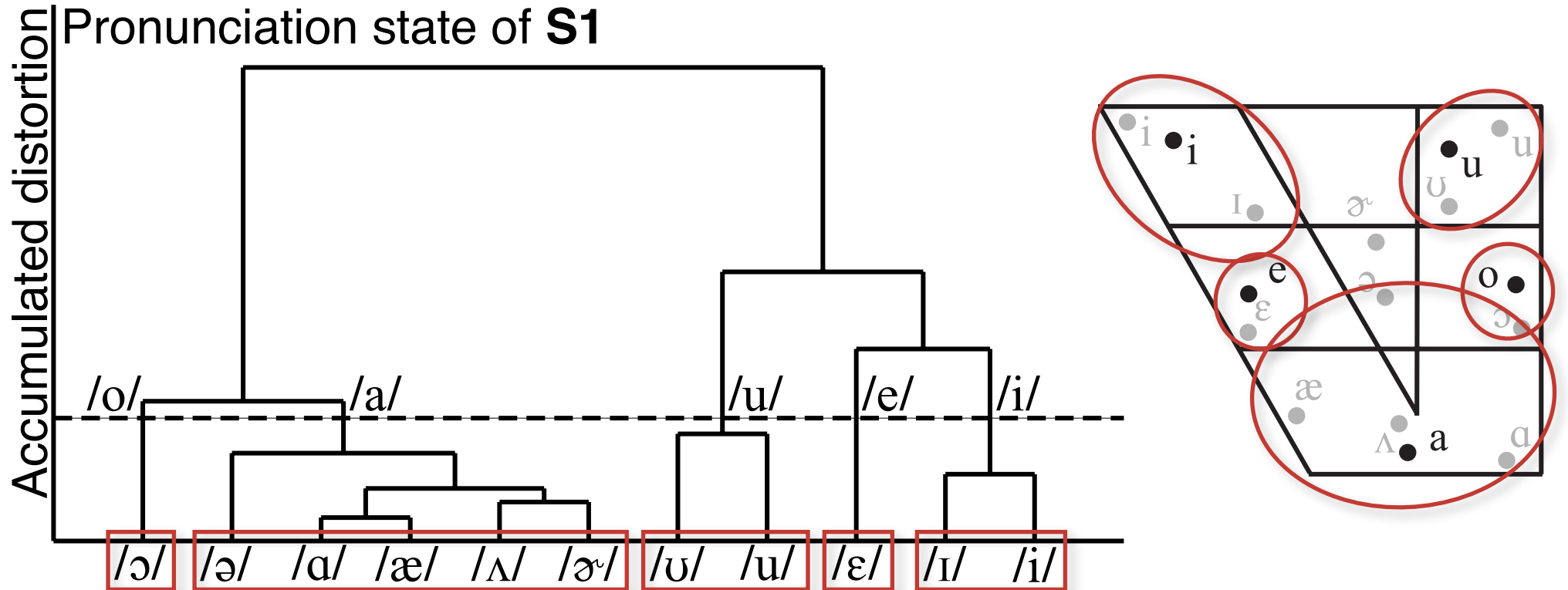
a	ɑ, æ, ʌ, ə, ɜ̃
i	ɪ, i
u	ʊ, u
e	ɛ
o	ɔ

- 「あ」 的な母音の置換
- 「い」 的な母音の置換
- :



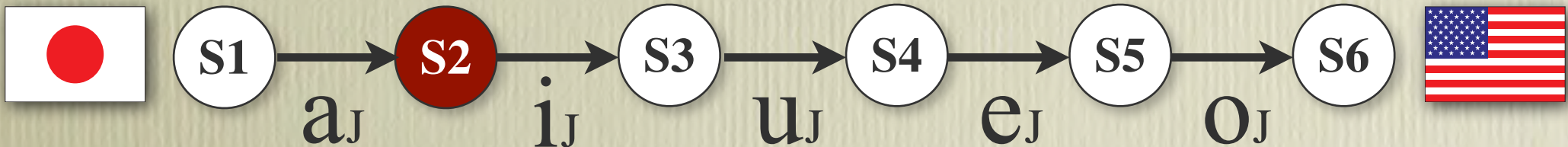
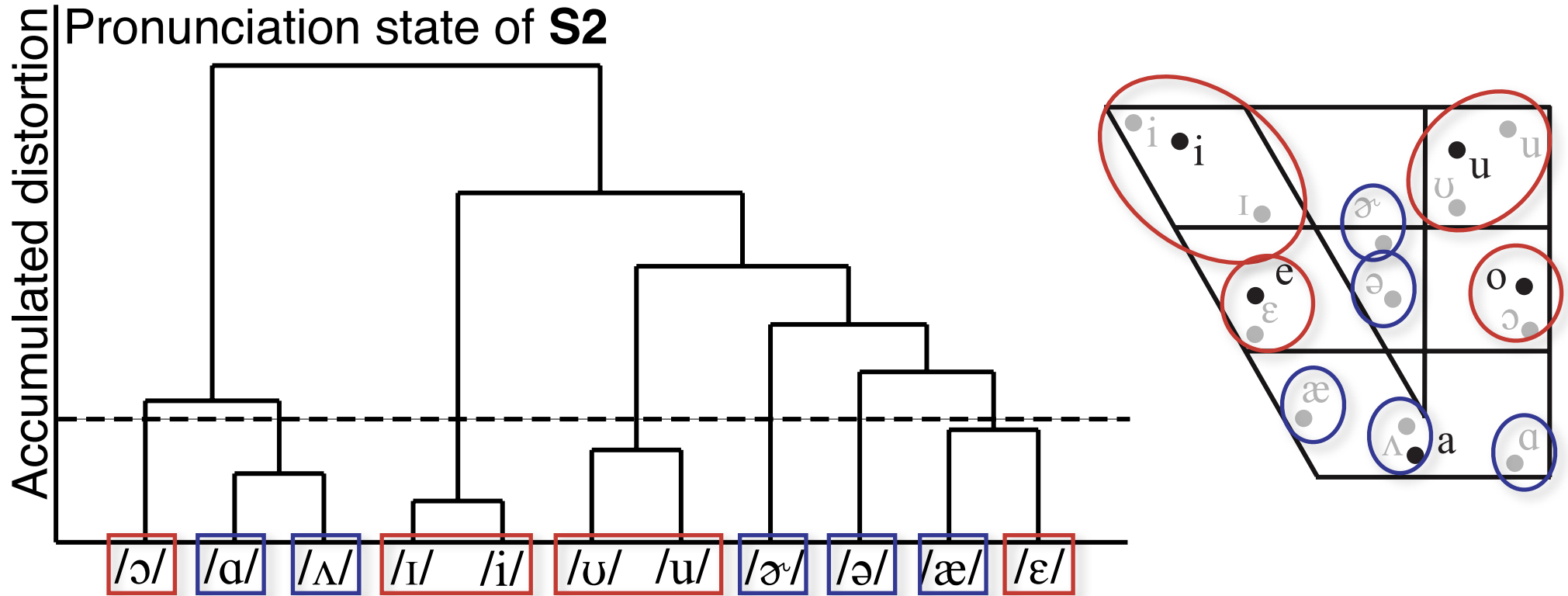
学習者の発音状態の記録とその遷移

S1からS6への遷移の様子の記事



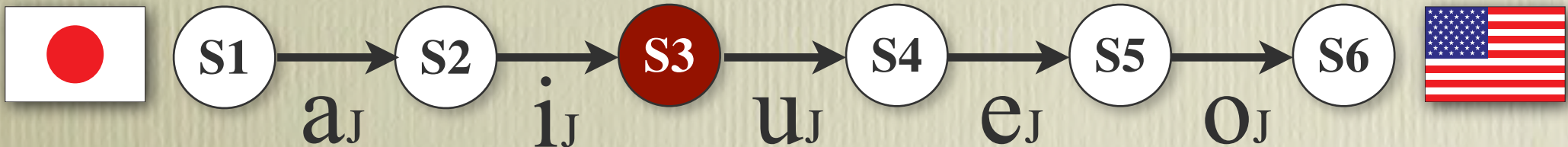
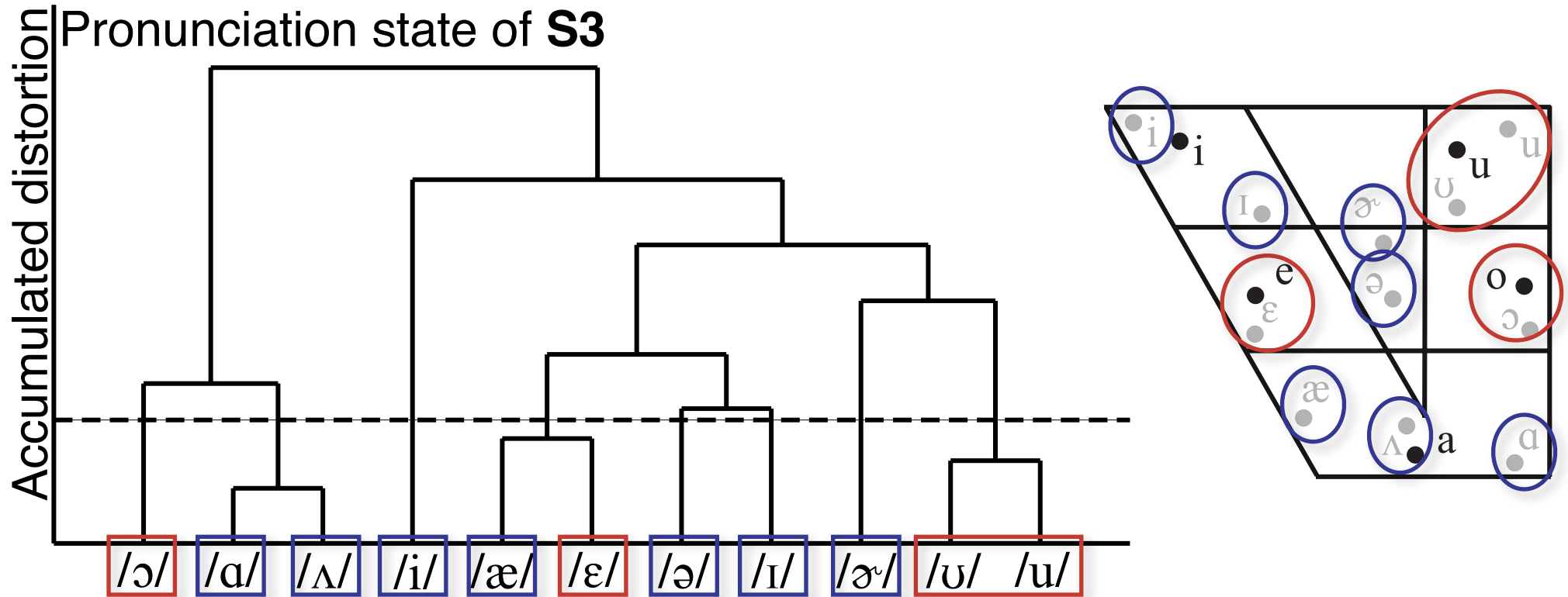
学習者の発音状態の記録とその遷移

S1からS6への遷移の様子の記事



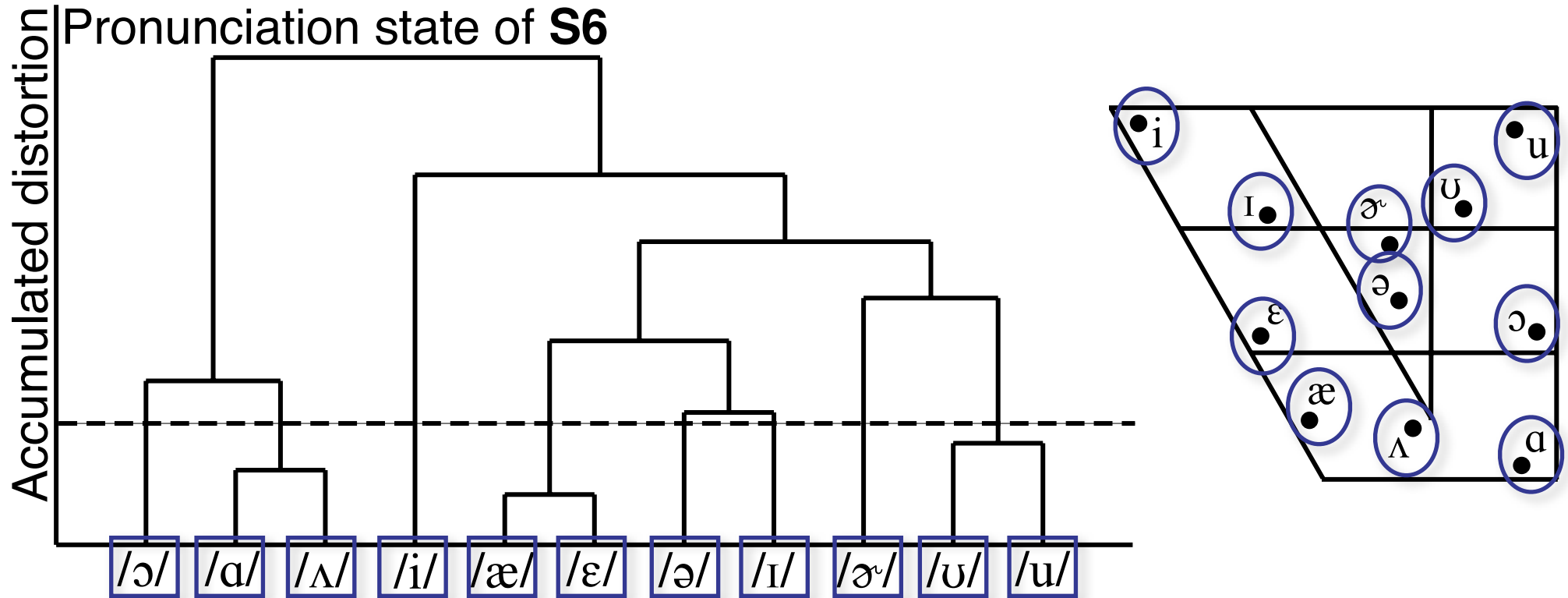
学習者の発音状態の記録とその遷移

S1からS6への遷移の様子の記事



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事



S1

a_J

S2

i_J

S3

u_J

S4

e_J

S5

o_J

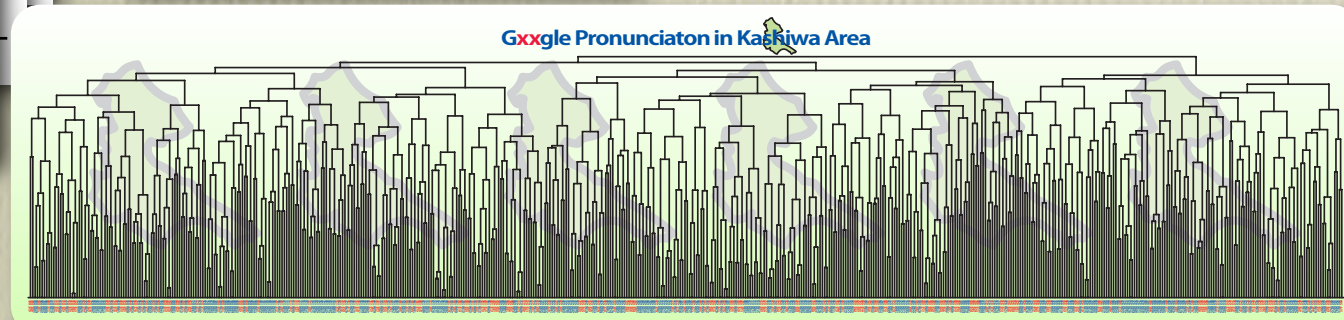
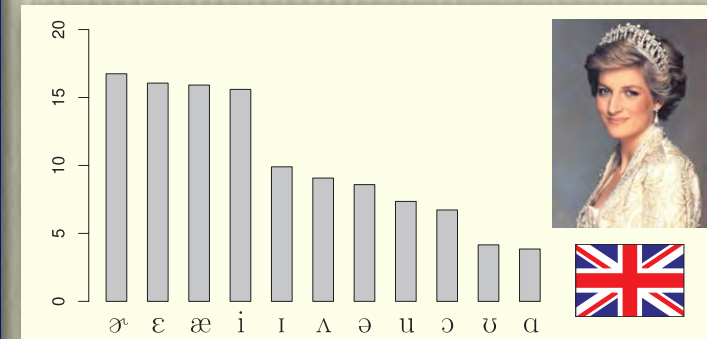
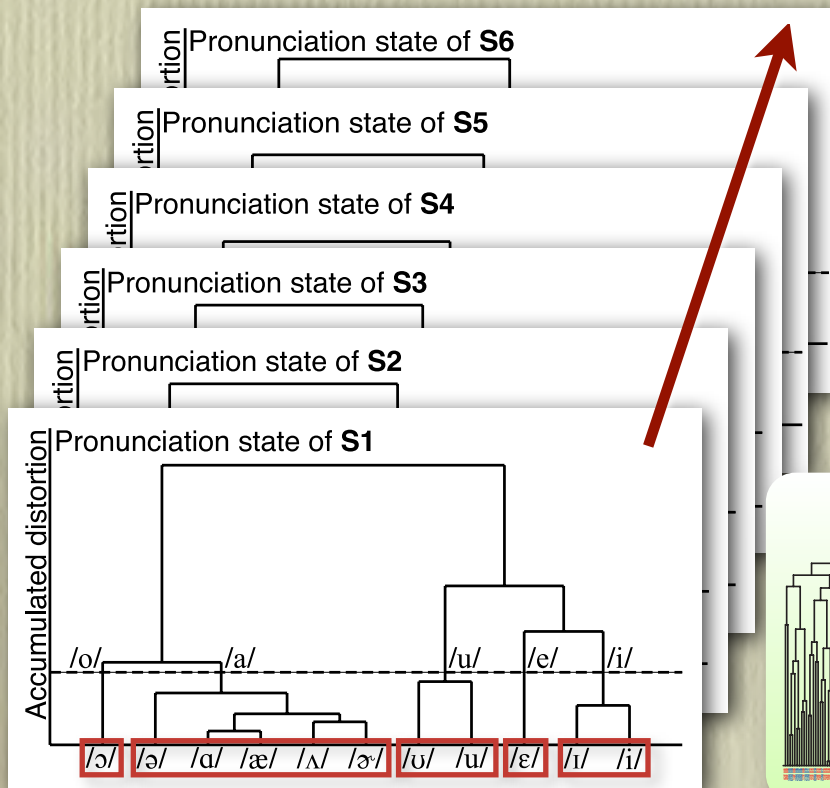
S6



英語・米語母音体系を対象としたCALL

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

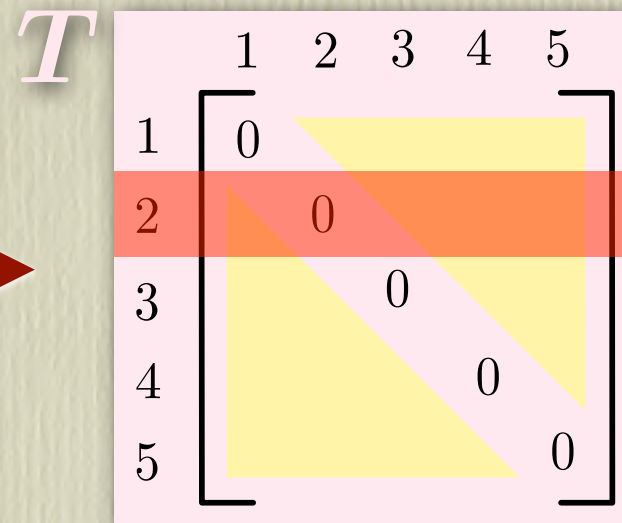
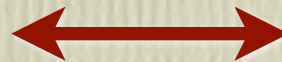
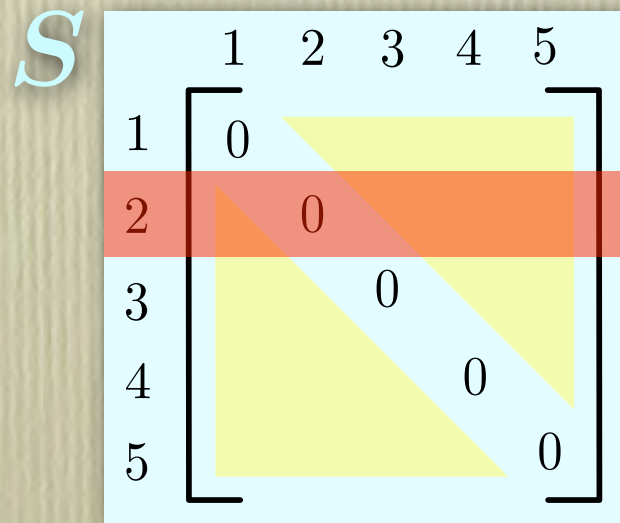
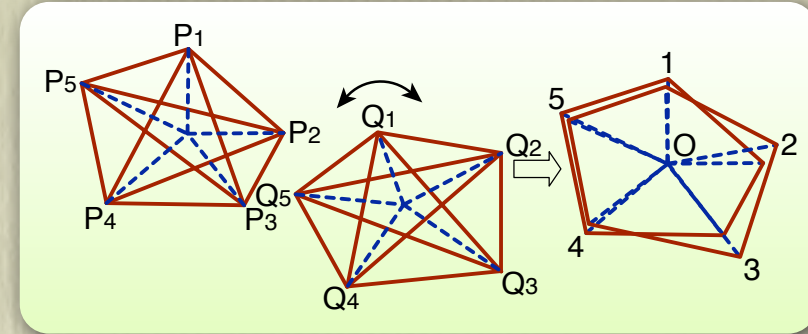
- 全体的構造歪み→発音習熟度の推定

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

- 局所的構造歪み→矯正すべき発音部位の推定

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

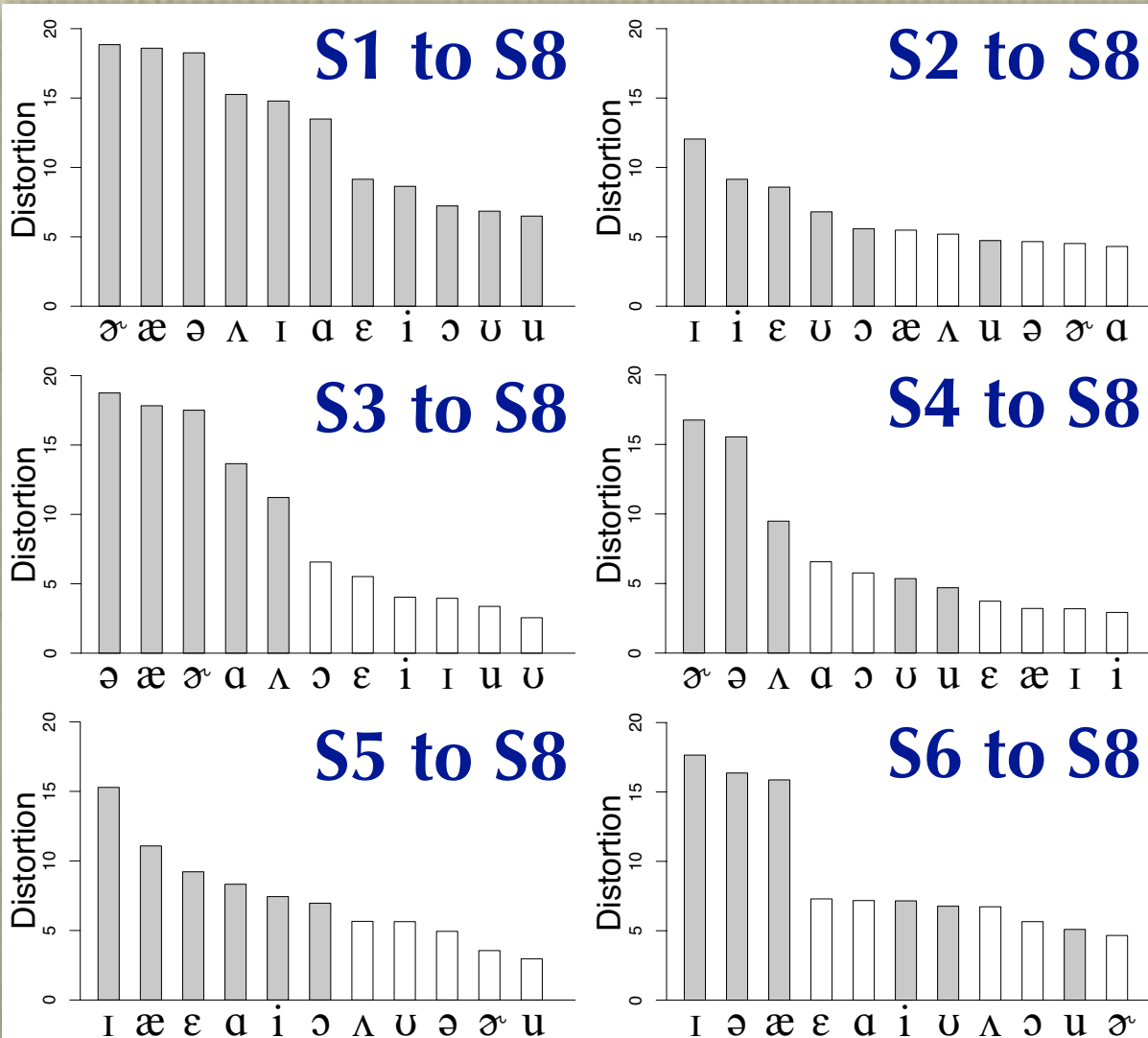
- LD = 大 ↔ 当該母音の矯正度 = 大



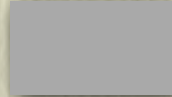
優先的に矯正すべき母音の教示

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



	ɑ	æ	ʌ	ə	ɜ	ɪ	ɪ	ʊ	u	ε	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

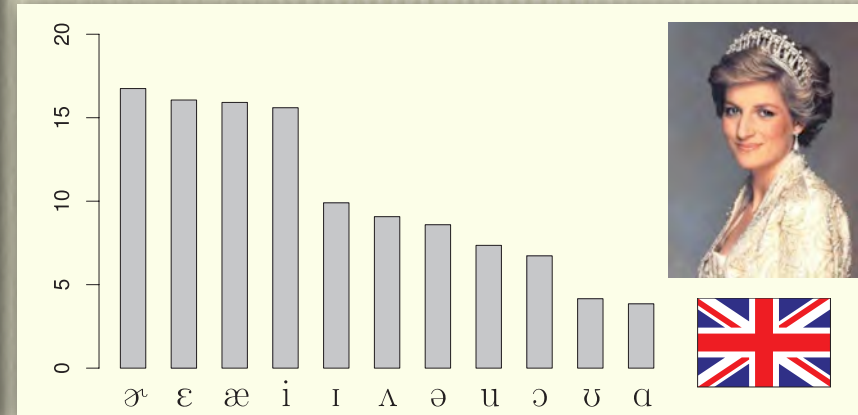


Figure 15: The estimated order of vowel correction

お手本となる相手を選ぶインタフェース

Model

Select two teachers.

OK

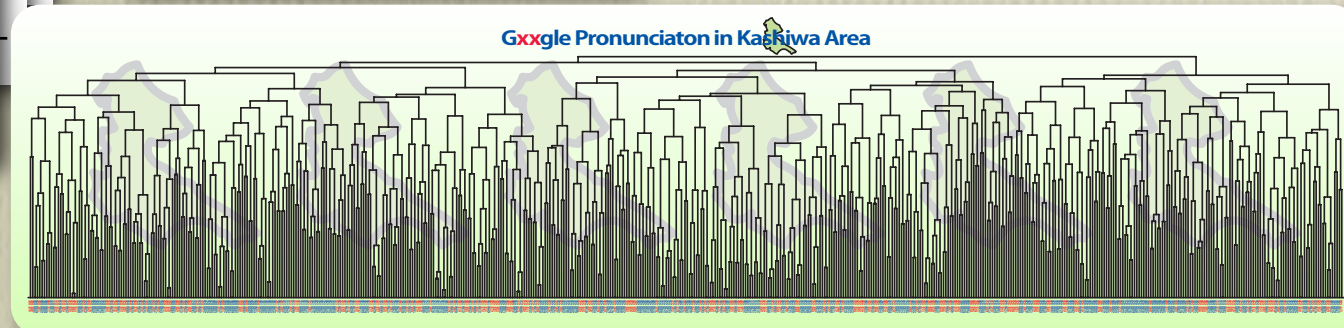
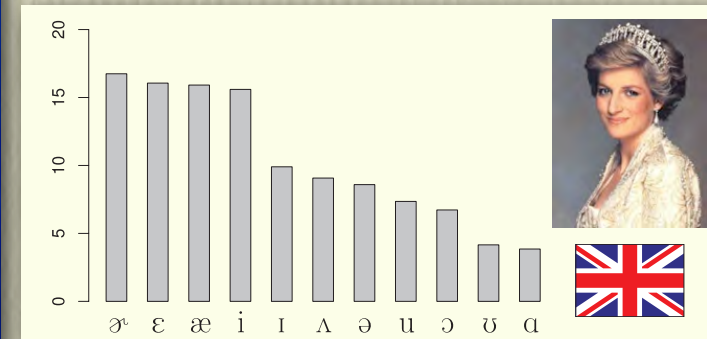
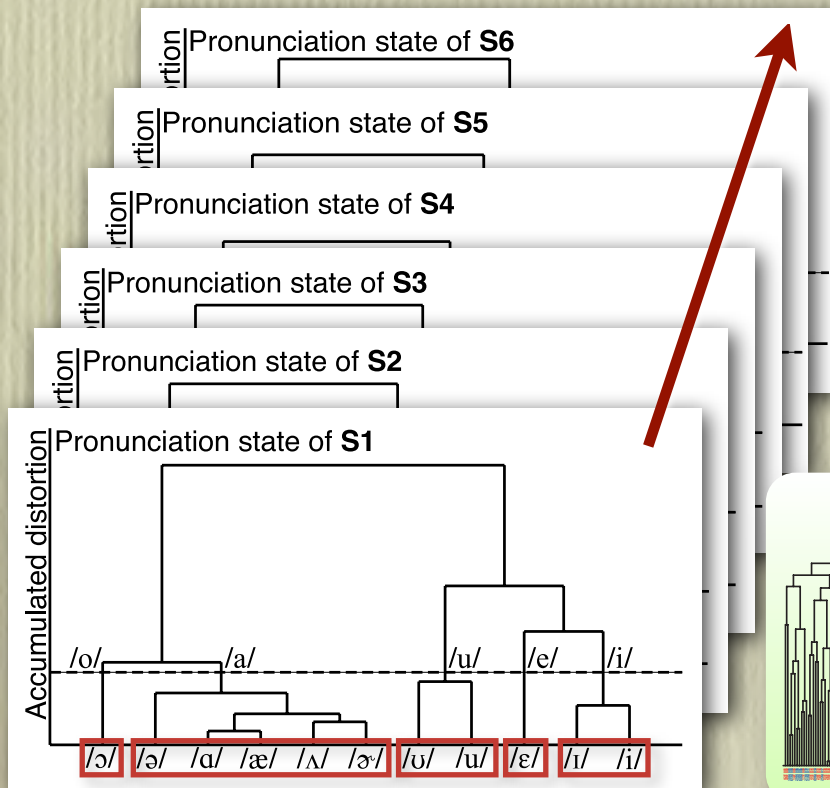
LONGMAN PRONUNCIATION DICTIONARY J.C. Wells

ルミナス 英和辞典 小池 浩一

英語・米語母音体系を対象としたCALL

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

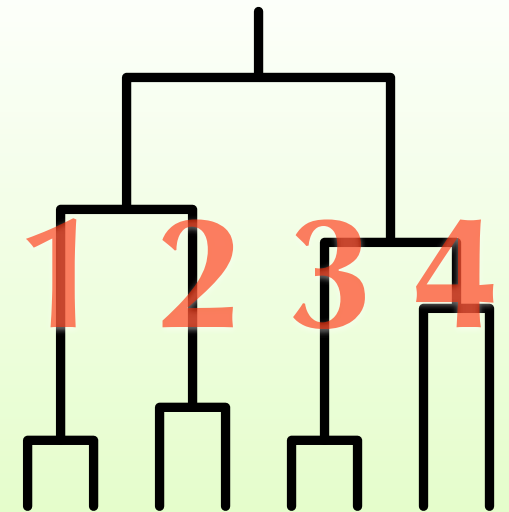
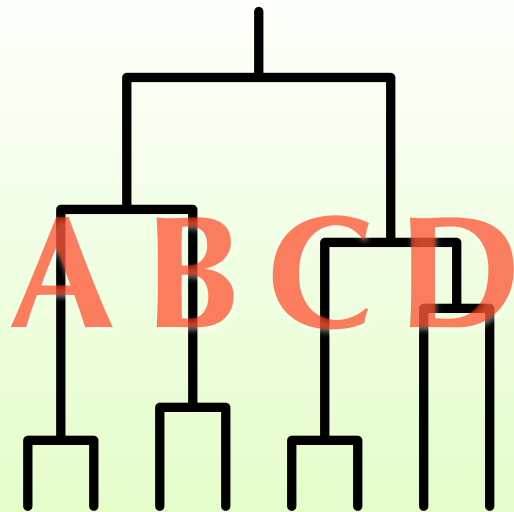
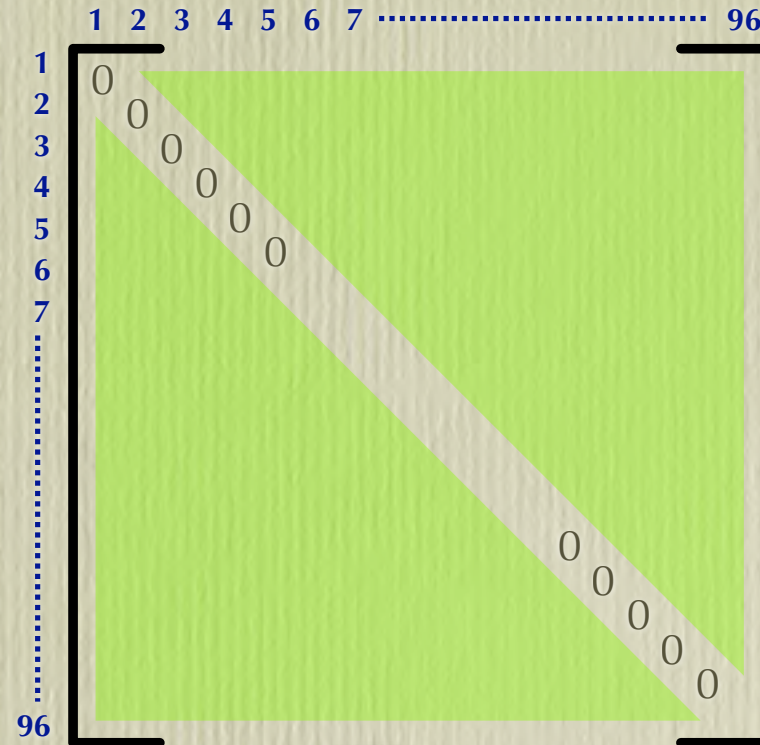
- 米語 11 母音 + 日本語 5 母音を含む単語音声収録
 - beat, bet, bat, but, × 1 回
 - ばと, びと, ぶと, × 5 回
- 米語母音の一部を日本語母音で置換
- どの米語母音を置換するのか？ 8 通りの置換パターン
 - 12人の帰国子女 × 8通りの置換パターン = 96人の学習者
 - [A~L] × [1~8] = 96個の 11x11 距離行列

あ	ɑ, æ, ʌ, ə, ø
い	ɪ, i
う	ʊ, u
え	ɛ
お	ɔ

[illegible]

2通りの学習者分類

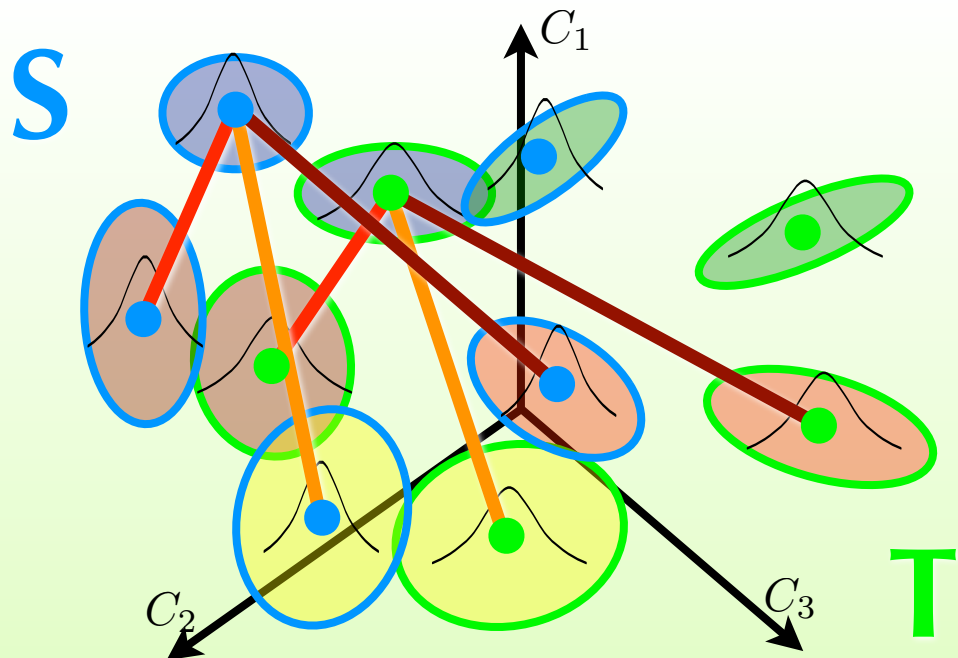
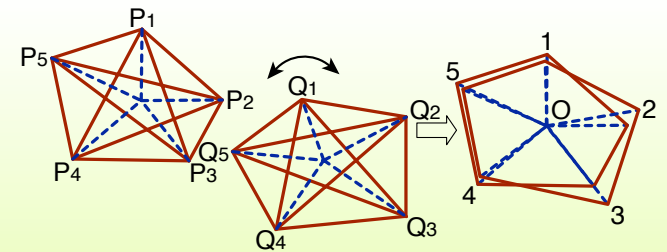
96 x 96 の距離行列に基づく学習者分類



2通りの学習者分類

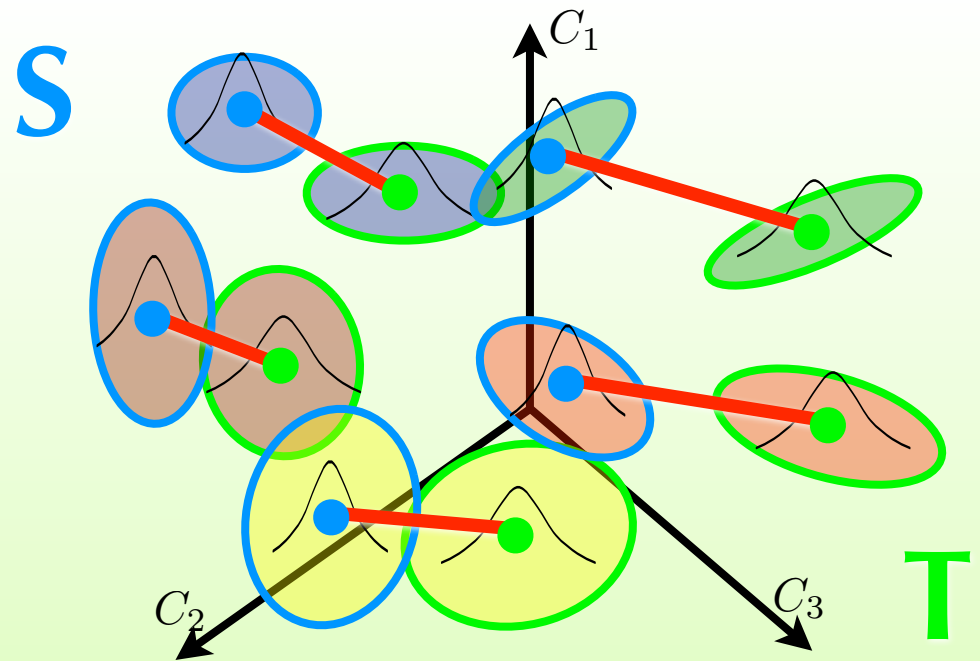
2通りの構造間距離

- 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



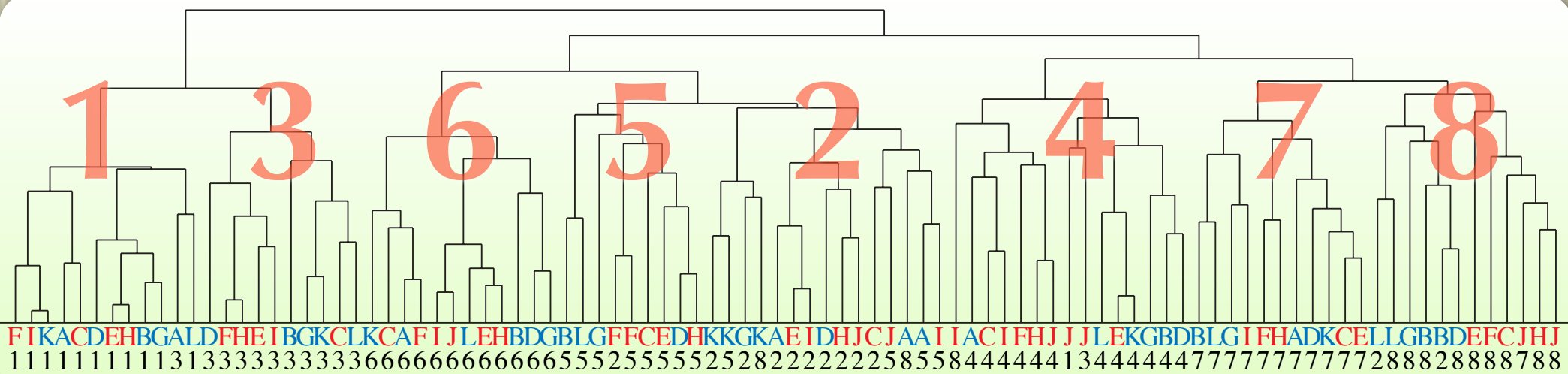
$$\sqrt{\frac{1}{M} \sum_{i < j} (S_{ij} - T_{ij})^2}$$

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

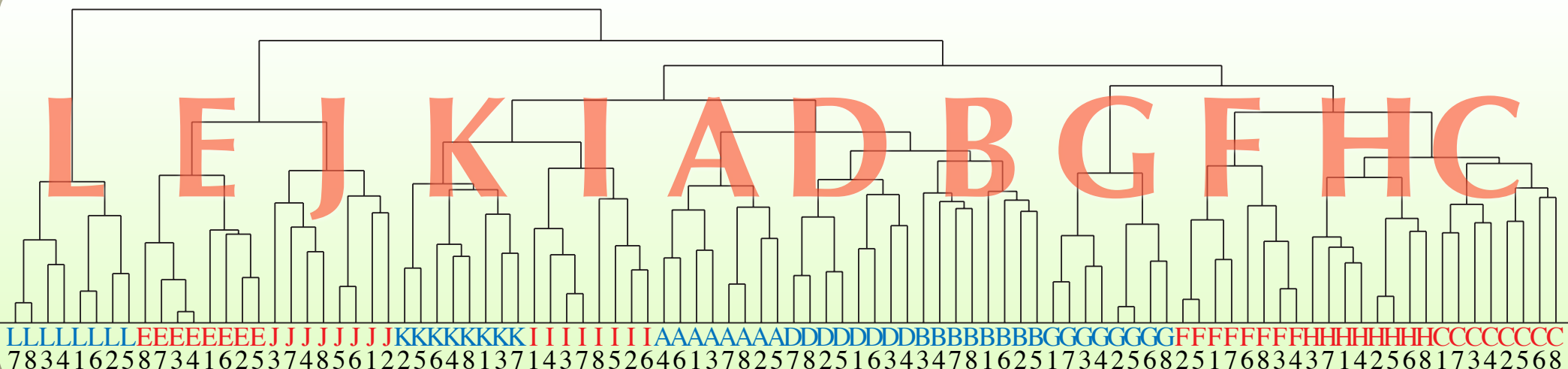


2通りの学習者分類

コントラストの比較に基づく構造間距離



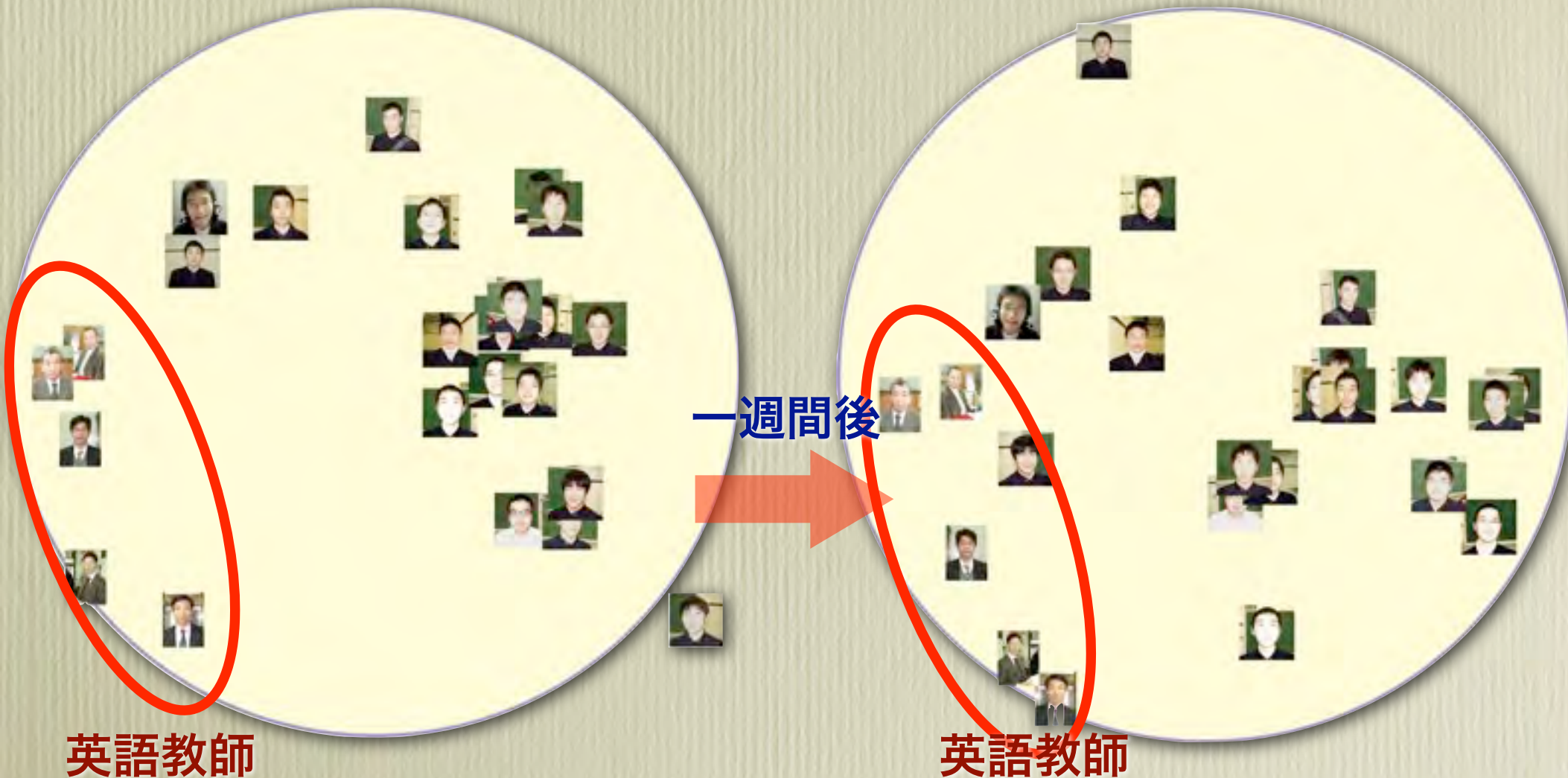
音の実体の比較に基づく構造間距離



クラスマップとファミリーマップ

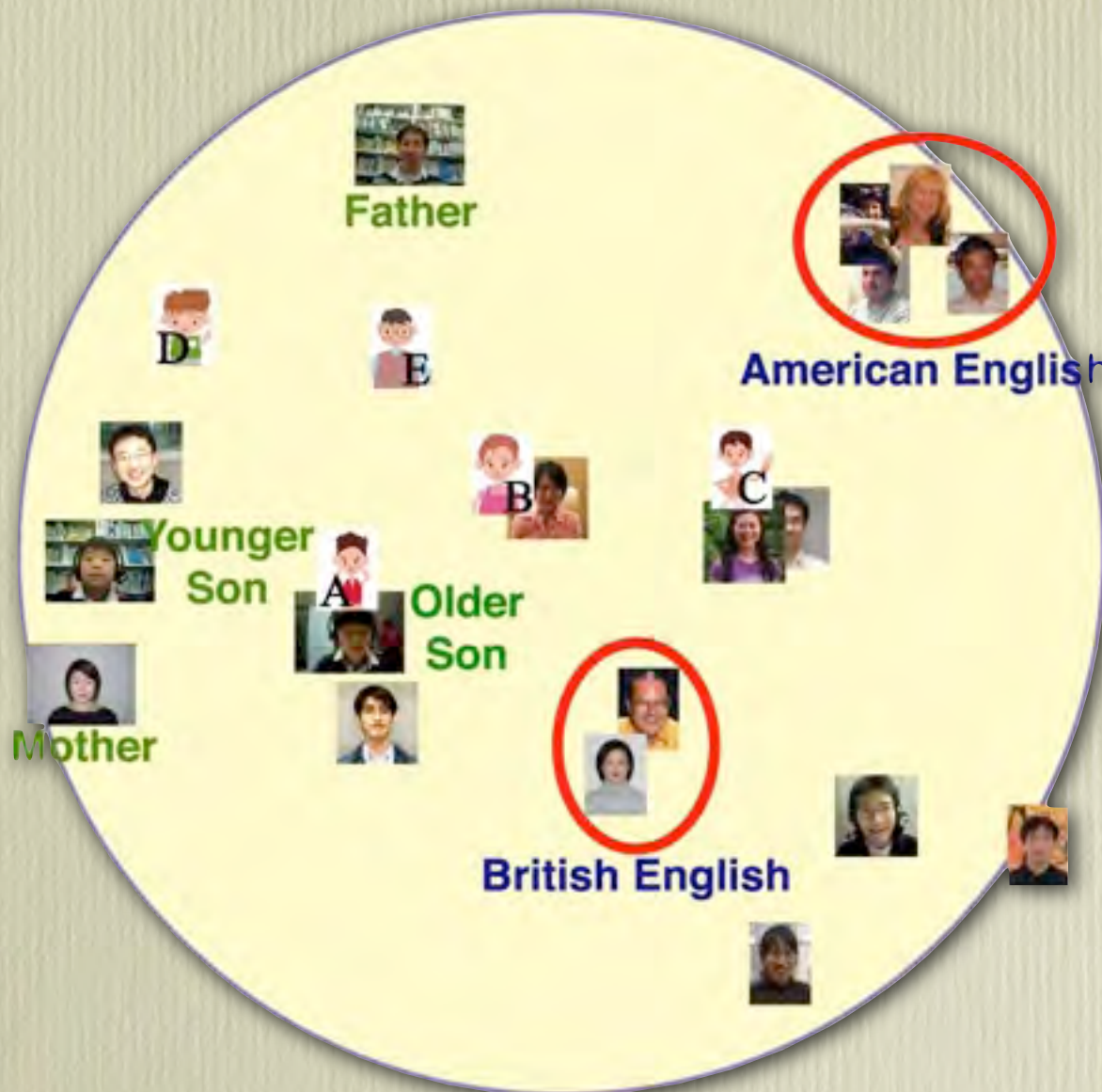
中学生20名と英語教師を分類

- 一週間の発音練習の前後で収録&分類



クラスマップとファミリーマップ

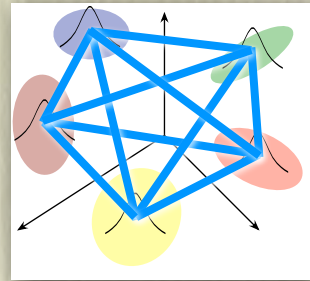
ご家族の皆さん＋英語教師＋その他を分類



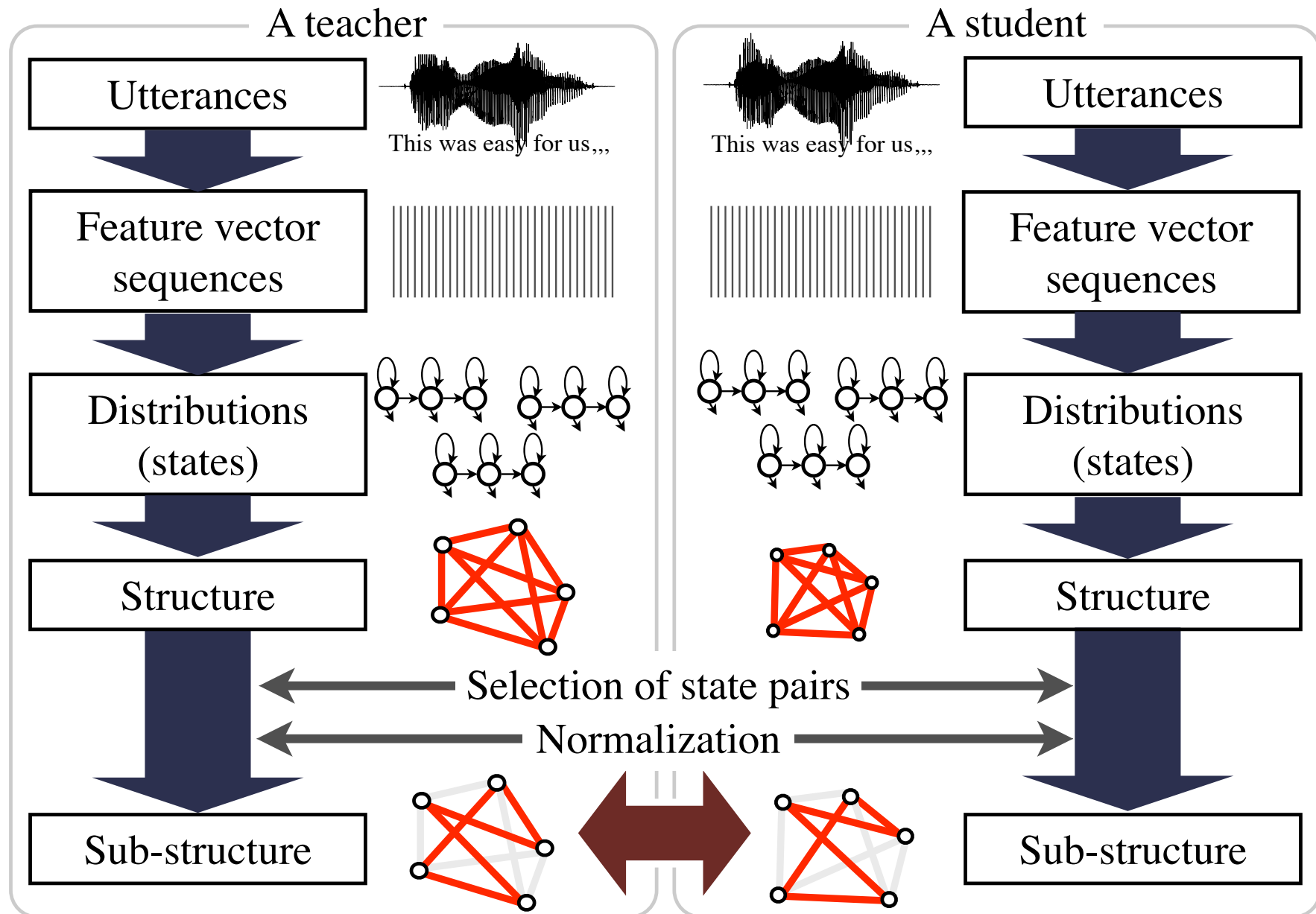
英語全音素群の体系に基づく発音評定

実験条件

- ERJ (English Read by Japanese) データベース
 - 200名の大学生 (男女100名ずつ) による読み上げ文音声DB
 - セット6 (26名) を使用 / それ以外のセットはシステム開発に利用
- 音素体系の音響的構造 (形態) の推定
 - セット6 = 各話者約70文の発声データ
 - 43個の英語音素について, 音響モデル (HMM) を学習
 - 音素間の距離 (f-divergence) を計測 & 距離行列化 (音韻体系化)
- 母語話者英語教師の距離行列 (音韻体系) との比較
 - 構造: 男性教師1名 / GOP: 20名の教師 (DB中の全教師音声)
- 広く使われている従来の技術との比較実験
 - GOP (Goodness Of Pronunciation)
 - 学習者発声に対して不特定話者音響モデルを用いて計算される, 意図された音素列に対する事後確率 (Posterior Probability : $P(M|O)$)



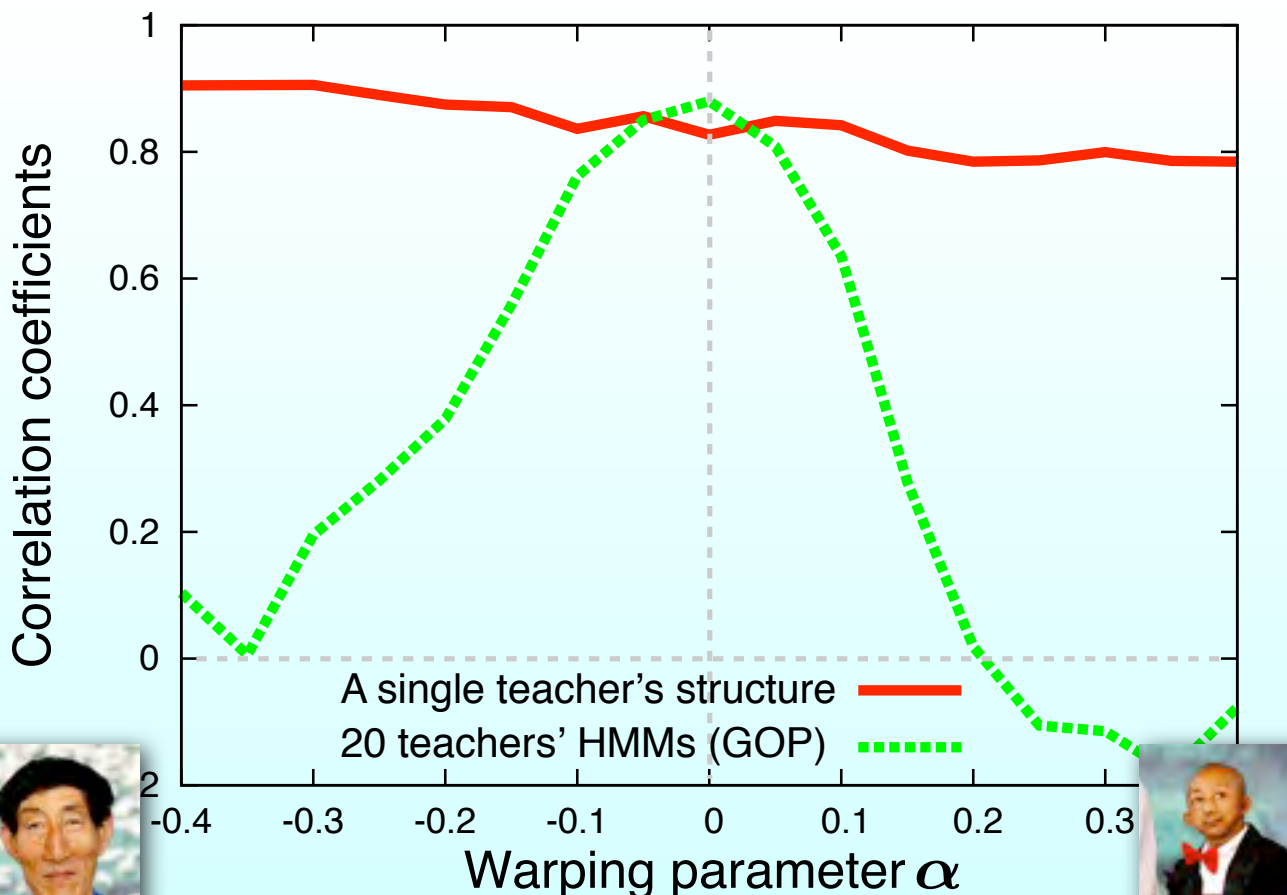
英語全音素群の体系に基づく発音評定



英語全音素群の体系に基づく発音評定

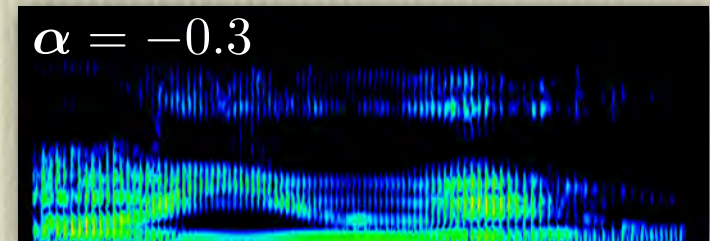
実験結果

- 学習者発音に対するスコア付け ← 教師によるスコアとの相関
 - GOP : 20名の英語教師の音声より作成した発音の音響モデル
 - 音韻体系表象 : 1名の英語教師の音声より作成した発音の音響モデル

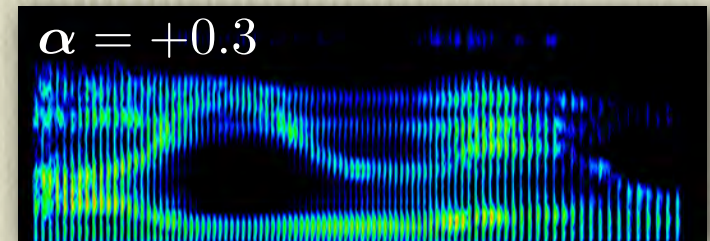


● 声色変換の例

$\alpha = -0.3$



$\alpha = +0.3$



こういうのを作ってみたい・・

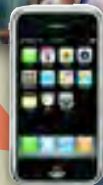
フレーズの発声を用いたモデル話者・学習者間の比較

- 「一発声→構造（形態）パターン」への変換
- パターン間の比較を通して行なう矯正箇所・教示の生成

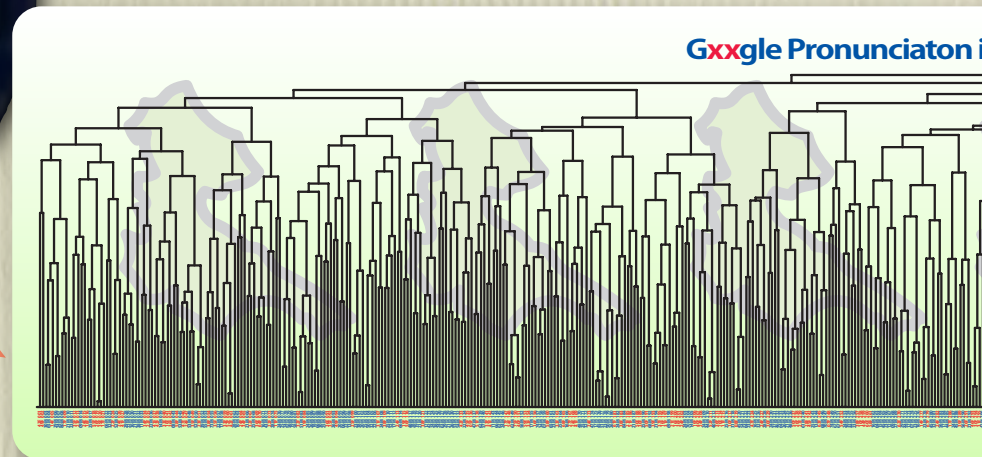
Phrase of the week!!



“You’ve got a friend in me.”



World Englishes
Individual Englishes



本日のメニュー

音声の物理学とその基礎

- 声が生まれる仕組み／音色，ピッチ，とその物理的対象
- 声の多様性（性別，年齢，体格によって変わる様々な声色）

声の共通項（構造的表象）とその発音評定への応用

- 発音学習とは教師の声の何を真似る作業を指すのか？
 - 幼児は親の声の何を真似て言葉を獲得するのか？／音色の相対感覚
 - 声帯模写と発音学習（音声模倣）の違いに立脚した基盤技術構築
- 学習者はどのような発音を目指すべきか？
 - 母語話者のような発音 vs. 伝わる発音
- 構造表象に基づく発音診断ソフトの開発

音声の構造的表象を通して考える発音教育

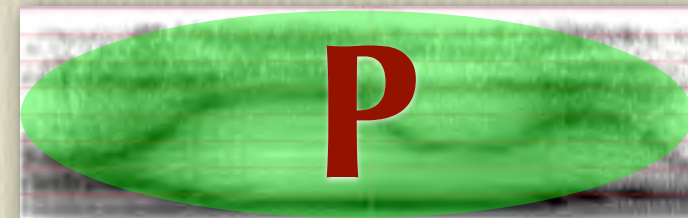
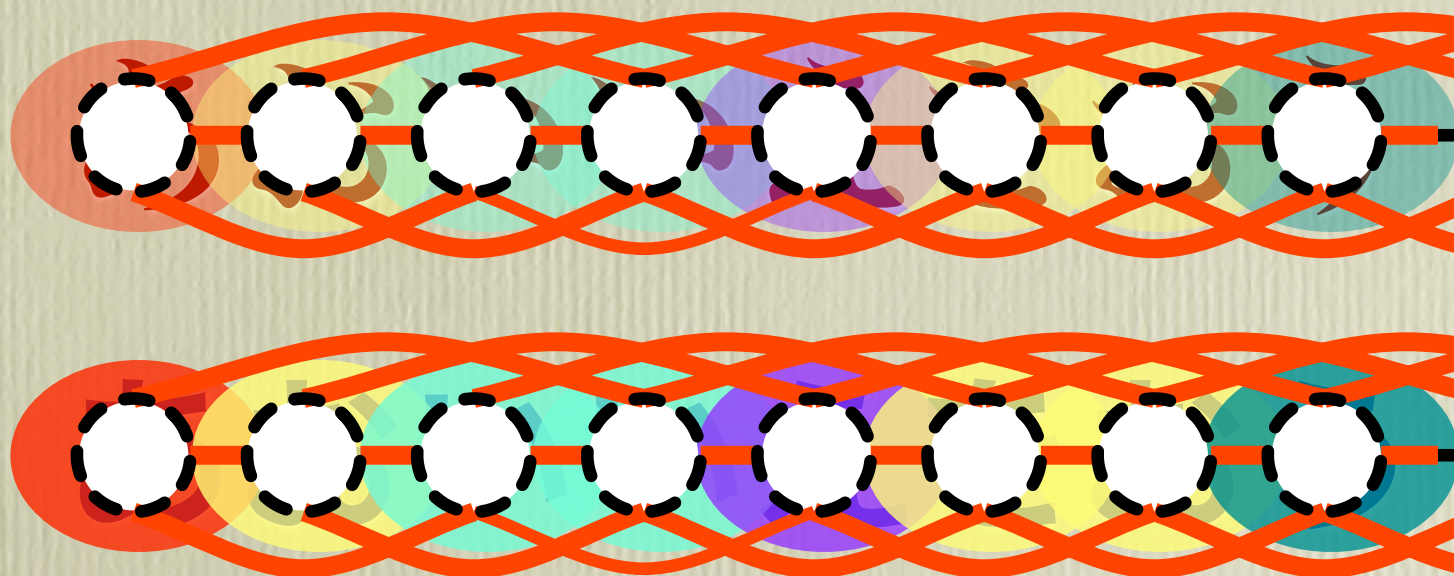
- 見えないもの，意識できないものを捉えるには？

まとめ

音色の動きをモデル化する？

喉が音色を作る／声帯が音高を作る

- 喉形状を変えると音色が色々変わる
 - 「あおいえをおう」と「あおいえをおう」
- 声帯の締め具合を変えると音高が色々変わる



まとめ

音声の物理学とその基礎

声の共通項を求めて ～音声の構造的表象～

- 発音学習とは教師の声の何を真似る作業を指すのか？
- 幼児は親の声の何を真似て言葉を獲得するのか？／共通パターン
- 一般的な知覚的恒常性から考える音声知覚／音色の相対感覚

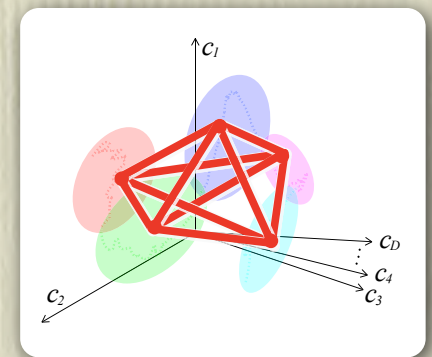
声の共通項の数学的定義とそれに基づく発音分析

- 年齢・性別・体格による差異を取り去った発音の骨格抽出
- 発音習熟度推定, 発音矯正箇所の特徴, 発音分類 . . .

学習者はどのような発音を目指すべきか？

- 英語？米語？いや, 誰の発音が目標なのか？
- World Englishes から Individual Englishes へ

音声の構造的表象を通して考える発音教育



Call for participation

Second Language Studies: Acquisition, Learning, Education and Technology

http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/

Q= iOS 4

新聞+ 学会+ 会合+ 組織+ 便利+ 趣味+ MAC+ ニュース (7729)+ PC+ Mac+ T-Lounge+ 旅行など+ 検索範囲+

INTERSPEECH 2010 Satellite Workshop on
**"Second Language Studies:
Acquisition, Learning, Education and Technology"**
Co-organized by [AESOP](#), [SLaTE](#), [NICT](#) & [LASS](#)
September 22-24, 2010,
[The International Conference Center,](#)
[Waseda University](#), Tokyo, Japan.
[AESOP](#): Asian English Speech cOrpus Project
[SLaTE](#): the ISCA SIG on Speech and Language Technology in Education
[NICT](#): National Institute of Information and Communications Technology
[LASS](#): Language And Speech Science of Waseda University

!!! THANK YOU FOR YOUR CONTRIBUTIONS. WE'VE RECEIVED 65 SUBMISSIONS !!!

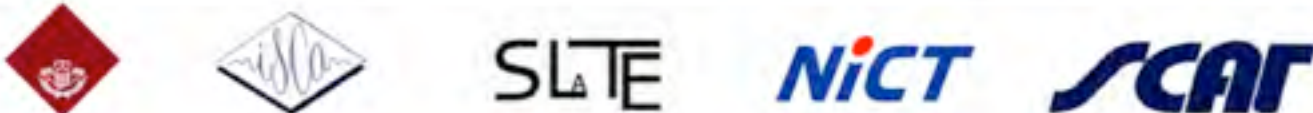
An open symposium on ["Primary School English Education in Asia"](#) will be held one day before the workshop. Admission to the symposium is free.

Call for participation

Primary School English Education in Asia

<http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/symposium.html> Q= iOS 4

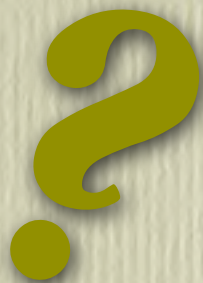
新聞+ 学会+ 金会+ 組織+ 便利+ 趣味+ MAC+ ニュース (7729)+ PC+ Mac+ T-Lounge+ 旅行など+ 授業英語+



**An open symposium on
"Primary School English Education in Asia"**
Co-organized by [AESOP](#), [SLaTE](#), [NICT](#) & [LASS](#)
September 21, 2010,
[Ono Auditorium](#),
[Waseda University](#), Tokyo, Japan.

Aim of the symposium:
English education will be implemented in every public primary school in Japan in 2011, and it has been a part of the core curriculum for several years in China, Taiwan, and Korea. Earlier introduction of English lessons does not simply mean that junior high school English education has been shifted to primary schools. Some unique methodologies are applied by individual teachers. In this symposium, leading teachers and educators are invited from China, Taiwan, Korea and Japan to give talks on various aspects of English education based on their teaching experiences. Their talks may include the history of primary school English education in their regions, unique methodologies developed for young children and the effectiveness of these methodologies, how young children acquire English skills of hearing, speaking, reading, and writing, teacher training methodologies, particularly those related to primary school English education and aimed at teachers with little or no previous experience in teaching English, general problems that teachers are currently facing, and future plans.

Invited speakers:



ご静聴, 有り難うございました

