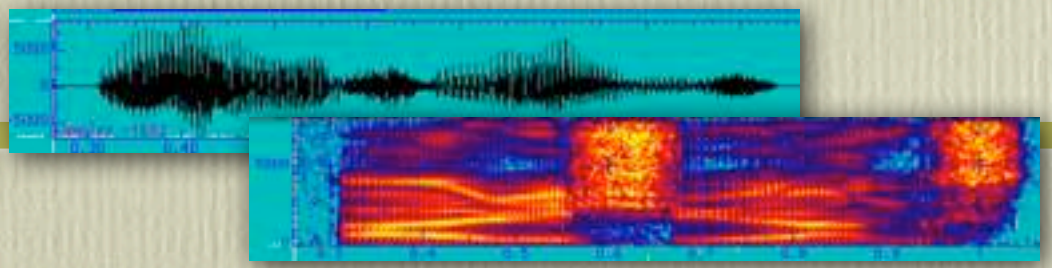
人文社会系研究科基礎文化研究専攻言語学専門分野

音響音声学

(Topics in Acoustic Phonetics)



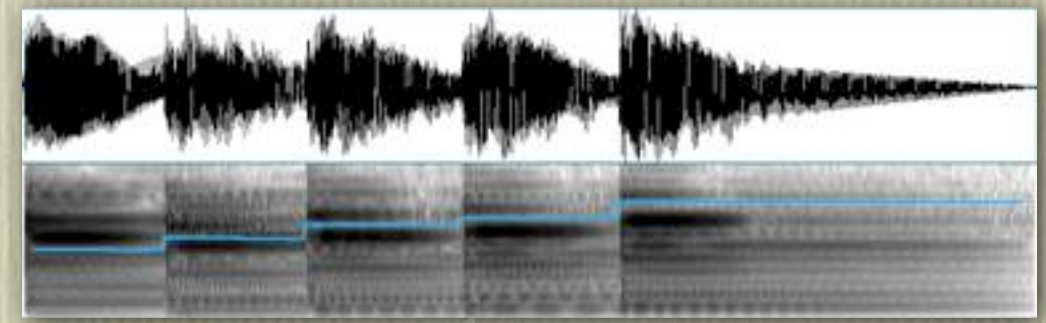
峯松 信明

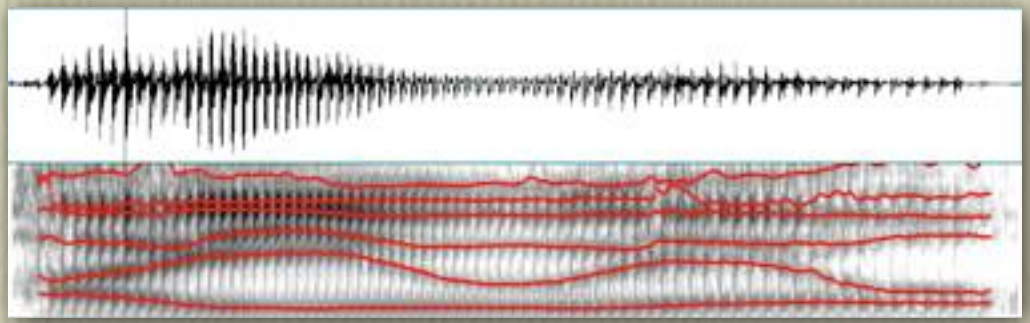
工学系研究科電気系工学専攻

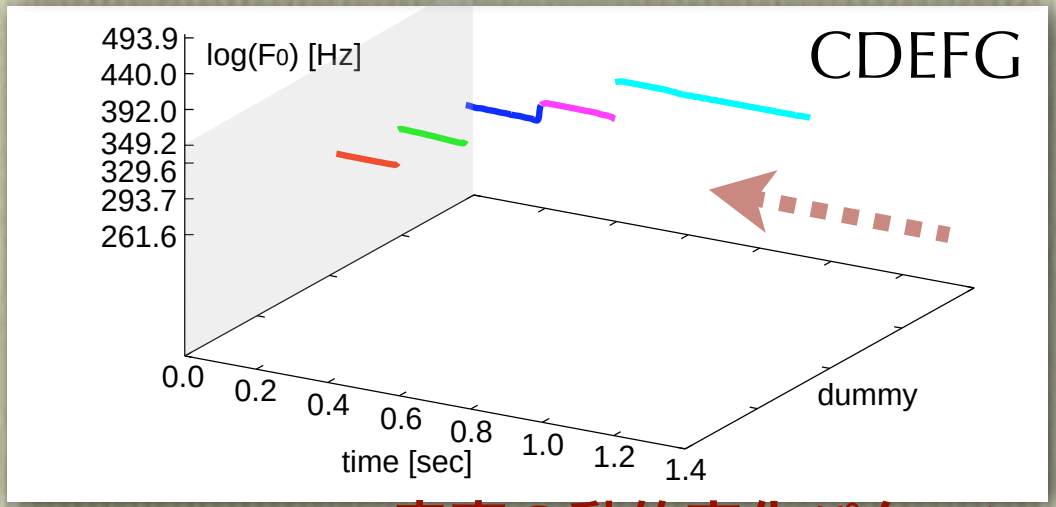
本発表の流れ

- **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？

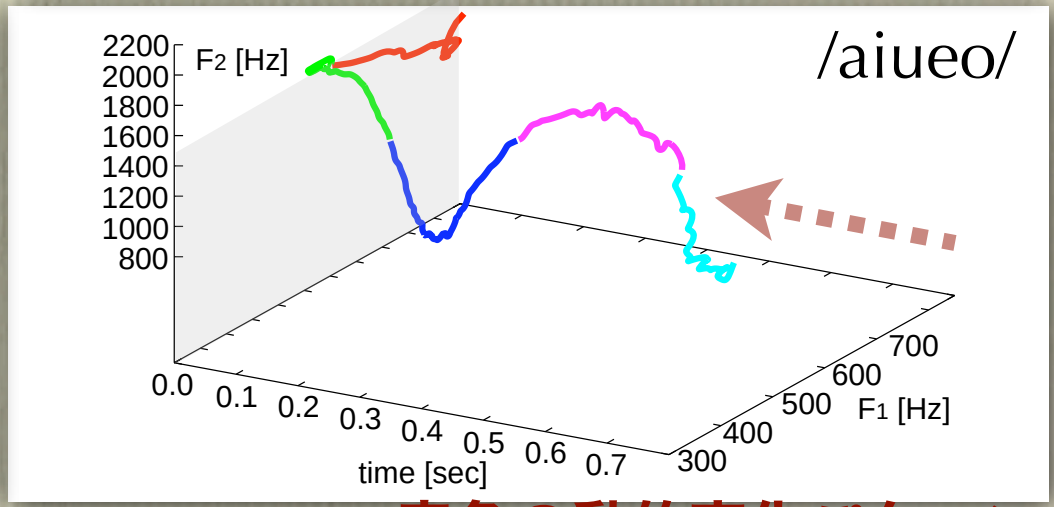
音声の構造的表象／音色の相対音感



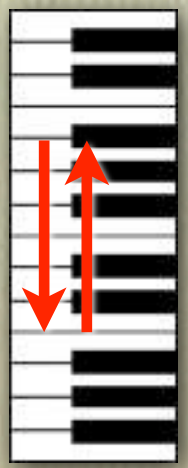




音高の動的変化パターン

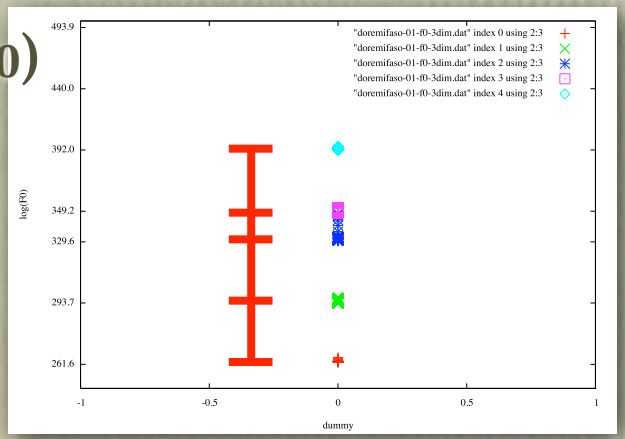


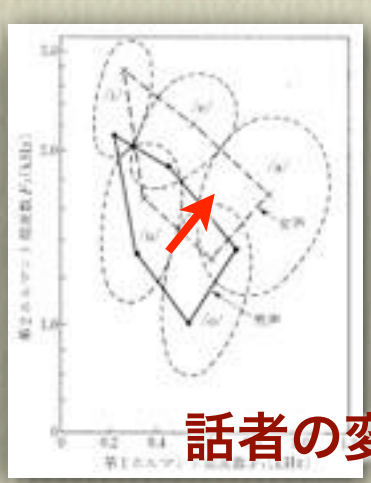
音色の動的変化パターン



調の変化

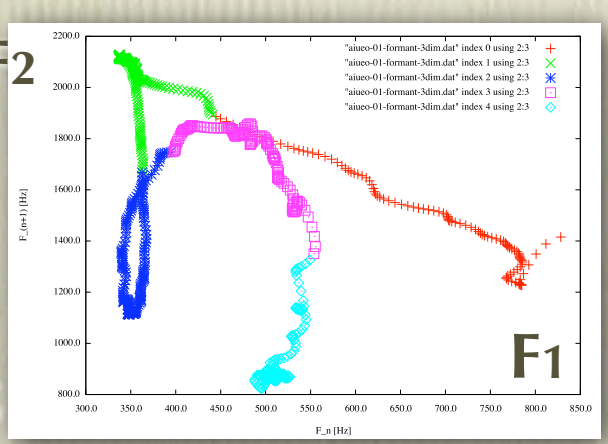
log(F0)





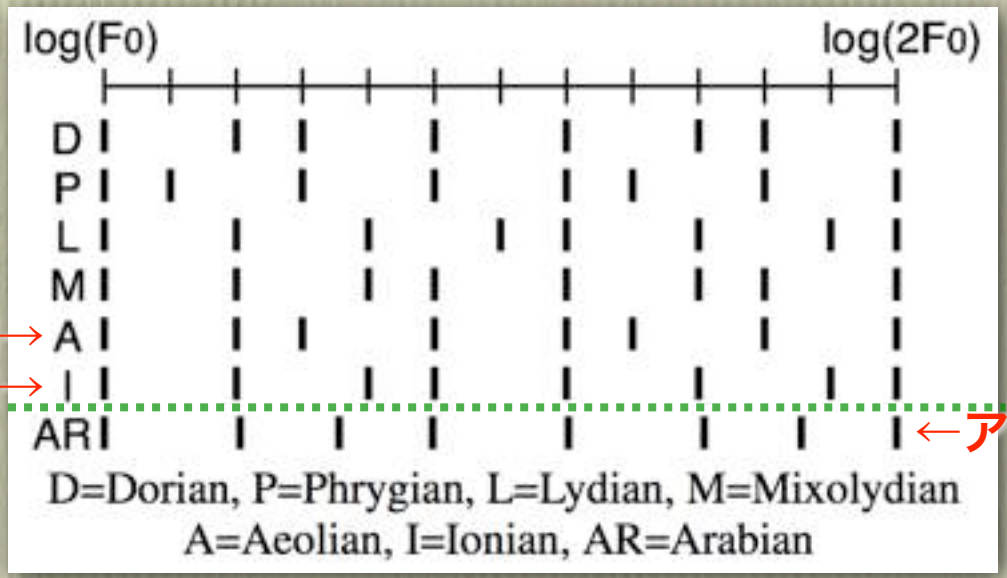
話者の変化

F2



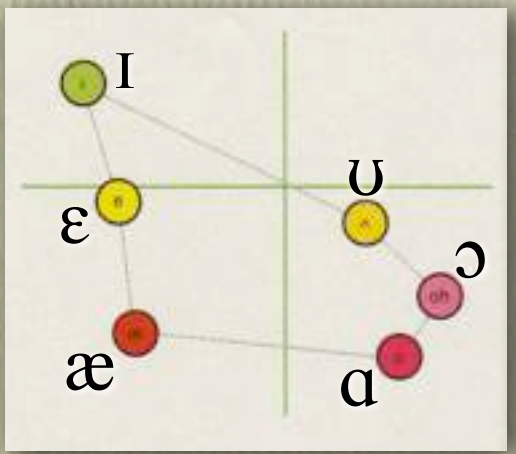
音声の構造的表象／音色の相対音感

音楽における調不変の音配置とその変種

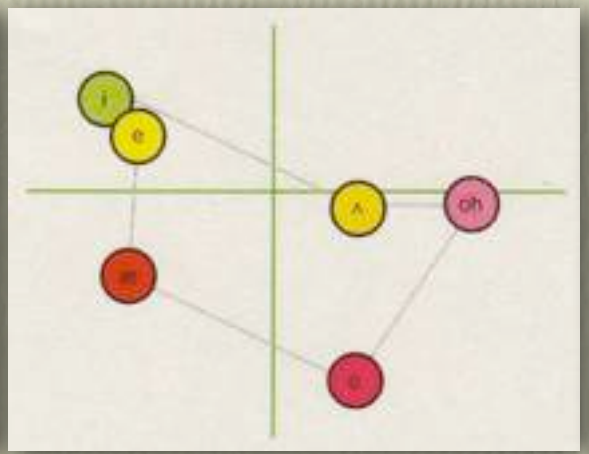


- 西洋音楽＝5全音＋2半音
- 種々の配置＝教会音楽
- 民族音楽には半音以外の配置

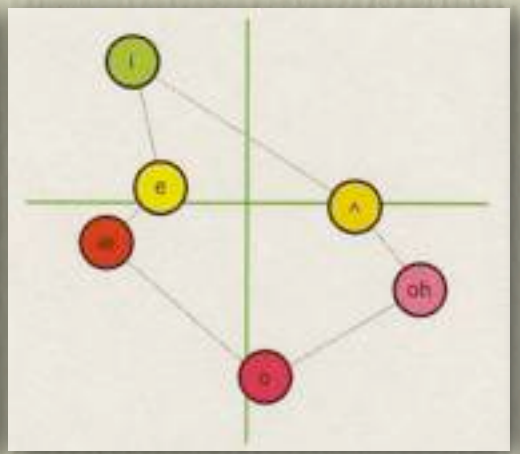
音声における話者不変の音配置とその変種＝欧米の方言



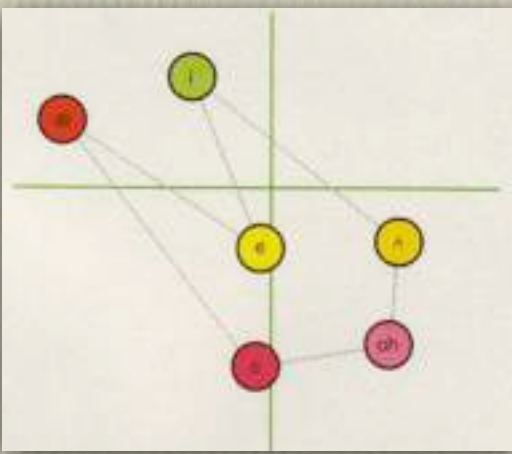
Williamsport, PA



Chicago, IL



Ann Arbor, MI



Rochester, NY

話者がコロコロ変わる音声の知覚

話者性が時間軸に沿って変化する音声

- もし全体的表象が使用されていれば，同定率は低下するはず。

音声刺激の作成

- HMM合成（男性アナウンサー7名／ATR503文）
- メルケプストラム（0〜24次元），7状態5分布
- 無意味8モーラ列（F0=LHHHLLLL，4型）
  - 促音，撥音，拗音，濁音，半濁音などのモーラは使用せず。全43種類
- 話者性変化のタイミング
  - 8/4/2/1モーラ，1音素，1分布（5人／音素）

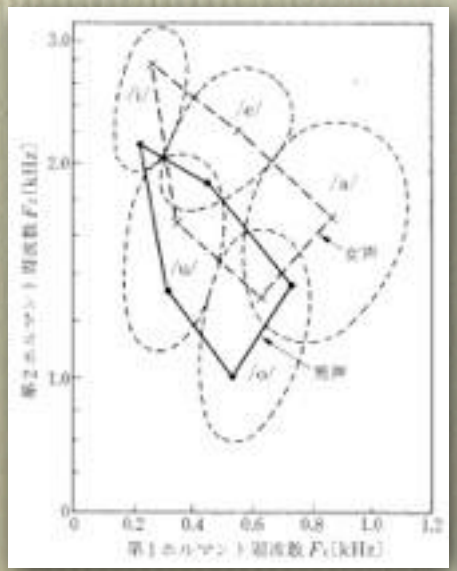
音声の構造的表象／音色の相対音感

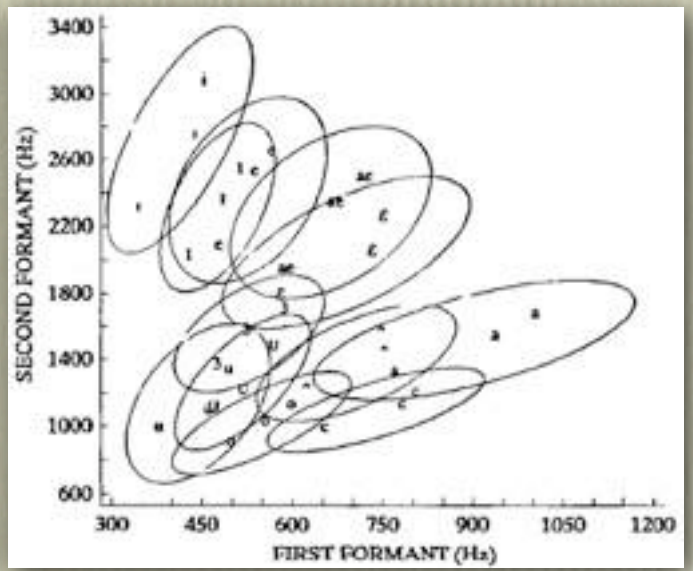
言語化困難な相対音感者（ラーラ音感者）

- 次に示すメロディーの３番目の音を覚えて下さい。その後，別のメロディーを提示します。同じ音が出て来たら挙手しなさい。
  - メロディーをシンボル列に変換できないので，困難な問いとなる。

言語化困難な音声の相対音感者（幼児的な成人？）

- 次に示す発声の３番目の音を覚えて下さい。その後，別の発声を提示します。同じ音が出て来たら挙手しなさい。
  - 発声をシンボル列（音韻列）に変換できなければ，困難な問いとなる



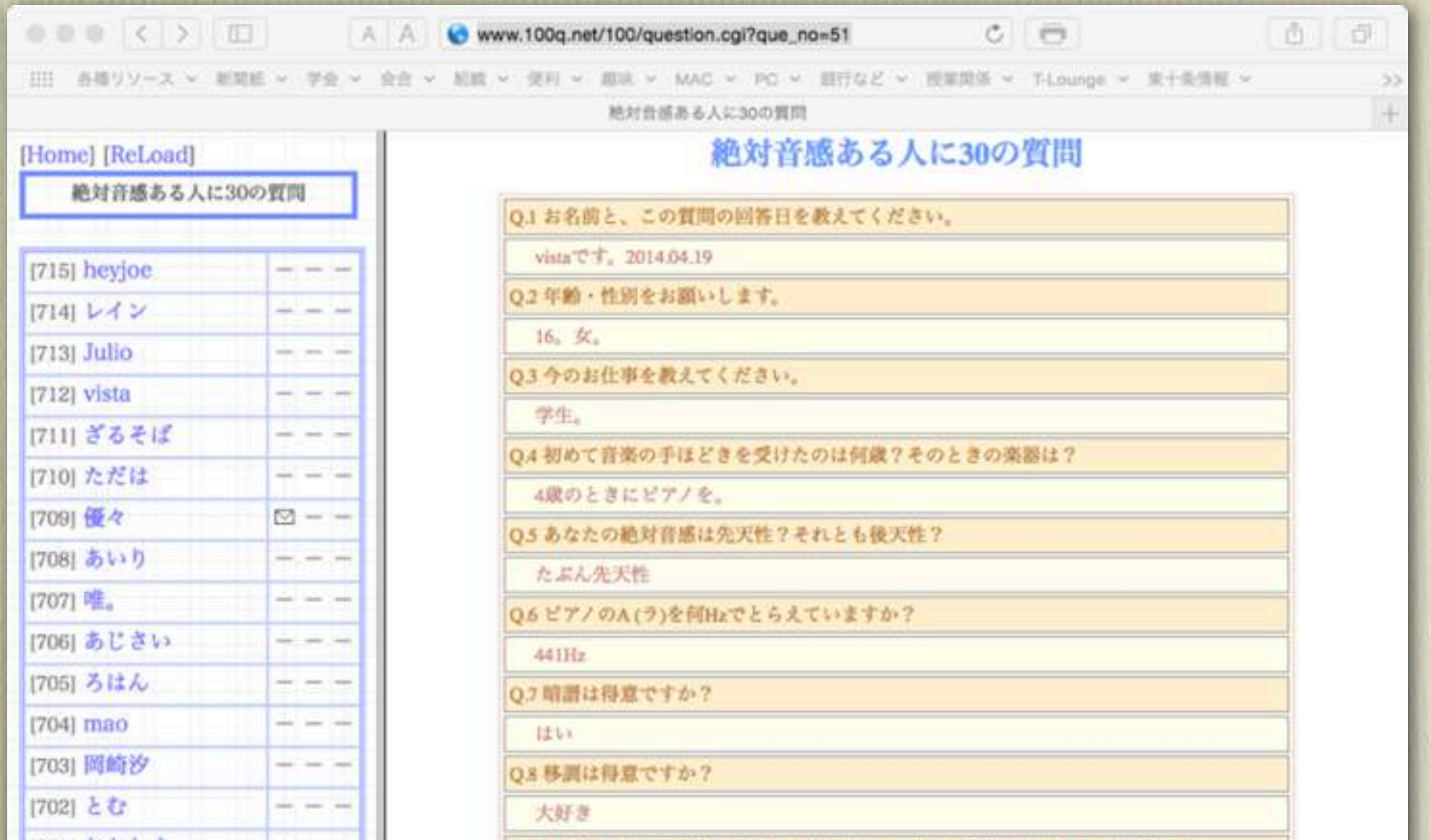


英語圏には十分な教育を受けているが，読み書きに苦労する人が多く存在しなければならない？

興味深いサイト

絶対音感ある人に30の質問

- http://www.100q.net/100/question.cgi?que_no=51

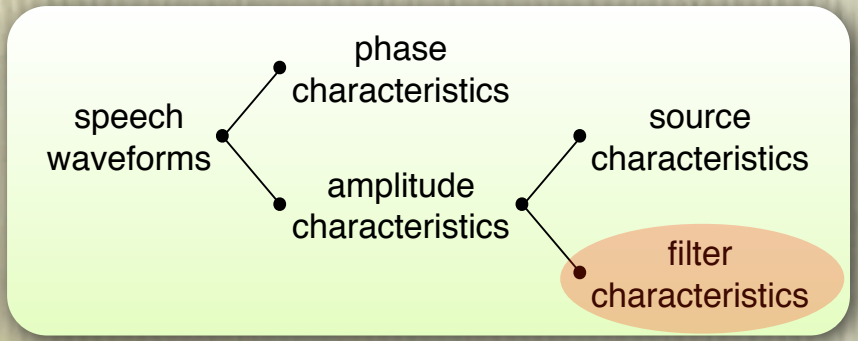


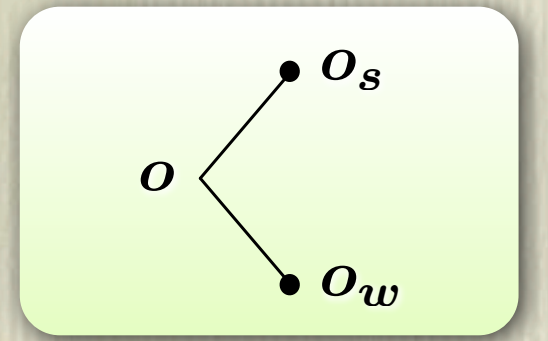
音声模倣とその技術的実装

まねだ聖子・松田聖子・神田沙也加

学習話者そっくりの声色で読み上げる技術＝音声合成

- Blizzard Challengeでは学習話者の個人性の再現も採点対象[7]
- 黒柳徹子を使えば，黒柳徹子の声になる。





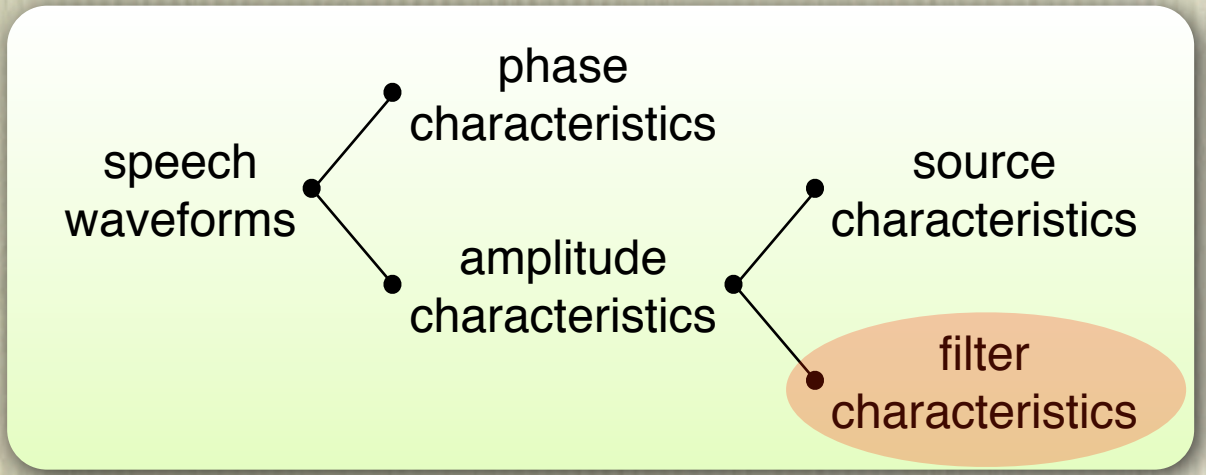
他者の音声の模倣＝声帯模写となる方々

（重度）自閉症者に見られる音声模倣＝声帯模写

- 七色の声を持つ中村メイ子の声をそっくり真似る[8]
- 相手そっくりの声を模倣する[11]
- 車，電車，などの音響音の模倣[12]
- 移調してしまうと，その曲だと認識してくれない[8]
- 母親の声は理解できるが，それ以外は難しい[13]

「言語＋非言語」が同居したままの音声の捉え方

- 音声コミュニケーションに困難を抱える場合が多い[18]



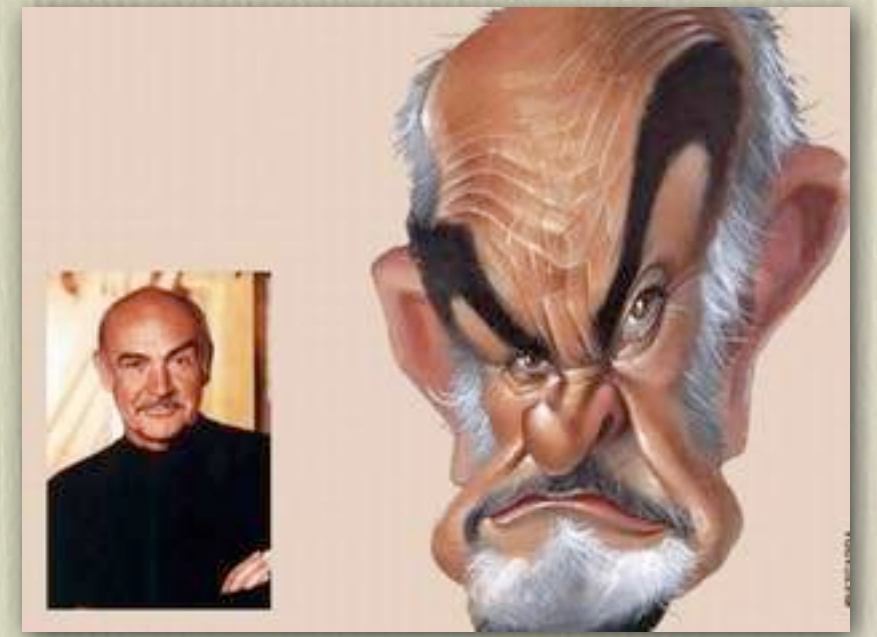
とある自閉症者の訴え

とある自閉症者（アスペルガー症候群）の手記

- 「発達障害当事者研究」（綾屋紗月，熊谷晋一郎著） [9]
- 「外国語の発音練習」「カラオケ」が難しい。
  - どうしても，先生／職業歌手の声帯模写をしてしまう。
- みんなの真似は真似じゃない。だって，声色違うじゃない。
  - 「自分の声でいいんだよ」と言われるけど。
  - 「そもそも，私の声って何なの？いつの私の声のこと言ってるの？」

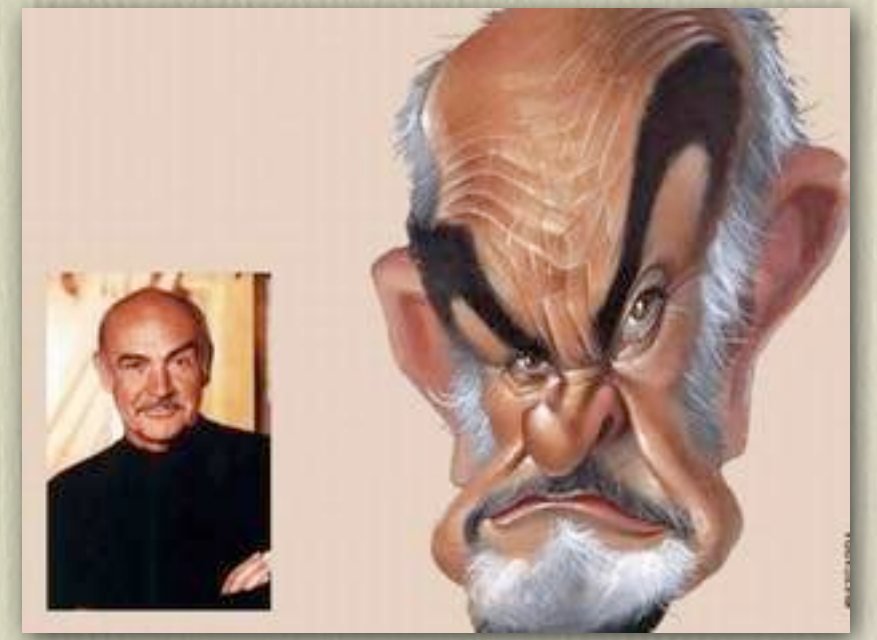
発達的視点から考える技術的欠損

- **ものまね歌合戦って，面白いですか？**
  - 何が面白いのか，さっぱり理解できません。
- **この似顔絵，似てるって分かりますか？**
  - 分かりません。こっちの方が似てると思いますが。
- **綾屋さんの言語活動の主メディア**
  - 手話と文字言語



発達的視点から考える技術的欠損

- **ものまね歌合戦って，面白いですか？**
  - 何が面白いのか，さっぱり理解できません。
- **この似顔絵，似てるって分かりますか？**
  - 分かりません。こっちの方が似てると思いますが。
- **綾屋さんの言語活動の主メディア**
  - 手話と文字言語



発達的視点から考える技術的欠損

- **ものまね歌合戦って，面白いですか？**
  - 何が面白いのか，さっぱり理解できません。
- **この似顔絵，似てるって分かりますか？**
  - 分かりません。こっちの方が似てると思いますが。
- **綾屋さんの言語活動の主メディア**
  - 手話と文字言語

自閉症の方々に見られる症状

とある web より

自閉症の特徴の強みと弱み

強み→① 具体的なことをよく理解し、記憶する。
② 目で見て認知したり記憶する視覚的な認識・記憶力がいい。
③ 決まったパターンのくり返しに強い。
④ 好きなことへの集中力。

弱み→① 曖昧なこと、抽象的なことに弱い。
(一つひとつの情報はキャッチしていても、それらの相互関係がつかみにくい。
目に見えないこと、経験していないことを想像することが難しい。)
② 時間の見通しをたてるのが苦手。
(物事の終わりがわかりにくい。いつもの流れが変更されると、わからなくなる。)
③ 状況を認識すること。
(人の表情、しぐさ雰囲気などが理解しにくく、人の感情がわかりにくい。
怒られているのに嬉しがったり、ほめられているのに知らん顔など・・・。)
④ 話し言葉への理解、自分からのコミュニケーションが難しい。
(言葉が出てもオウム返しになるなど。)
⑤ 感覚刺激に対して特異な反応をする。
(感覚刺激に対して過敏だったり鈍感だったりする。感覚刺激が一度にたくさん入りすぎてしまう。特定の感覚刺激に苦痛を感じる。)

幼児の言語獲得と音声模倣

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[7]
- 動物学的には非常に稀な行為。霊長類では人間だけ[8]
- 他の動物では小鳥，クジラ，イルカくらいか[10]

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[9]
  - 車，ドア，椅子，犬，猫，音を真似る。人の声も音でしかない。
  - 良い九官鳥を聞くと，飼い主が分かる。
- 幼児の音声模倣
  - 動物学的には奇妙な模倣行為[10]
  - 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。



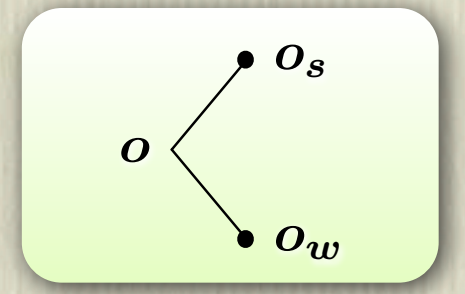

自閉症・・絶対的記憶・・動物・・？？

動物の情報処理と自閉症者の情報処理

- Dr. Temple Grandin（アスペルガー＆動物学者）
- 「動物感覚」[17]
  - 動物と自閉症者の情報処理的類似性を主張
  - 局所的／具体的／実体的　←→　全体的／抽象的／概念的

（定型発達を遂げた）人間が有する特異的な能力？

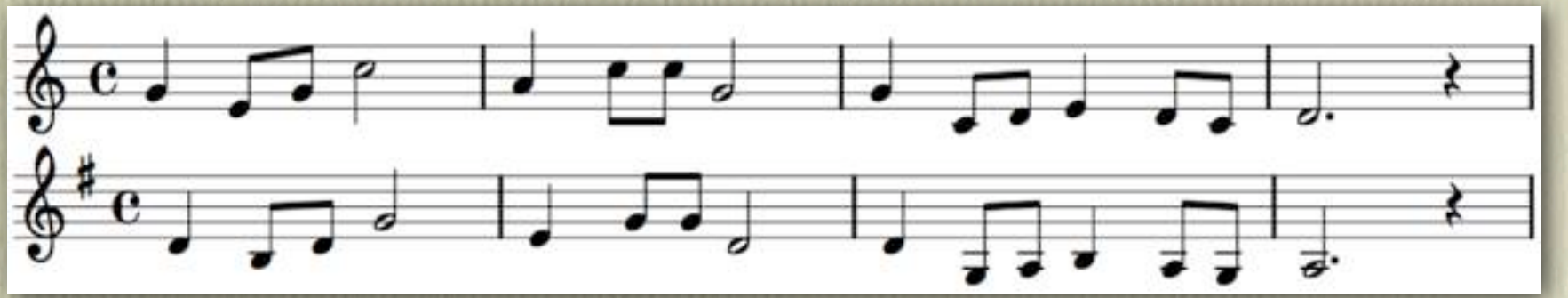
- 音を用いた情報伝達において，情報同一性は如何に確保できる？
  - 動物，重度自閉症者，現在の音声認識（情報分離が困難）
    - 情報（メッセージ）の同一性 ＝ 音響的特徴（o）の同一性
  - 定型発達を遂げた人間
    - 情報の同一性を確保するために，音の同一性は必要でなくなった種。
    - では，音のどの側面は同一なのか？
      - 語全体の語形・音形・ゲシュタルト



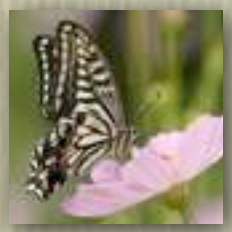
生物が獲得した静的バイアス除去術

音高の恒常的・不変的認知はどこまで遡れるのか？[6]

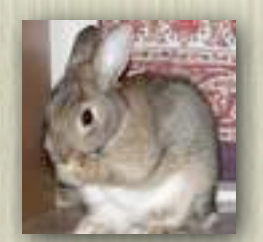
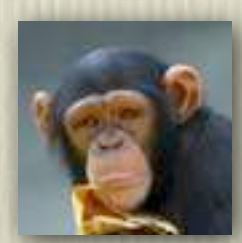
1
2



1 = 2



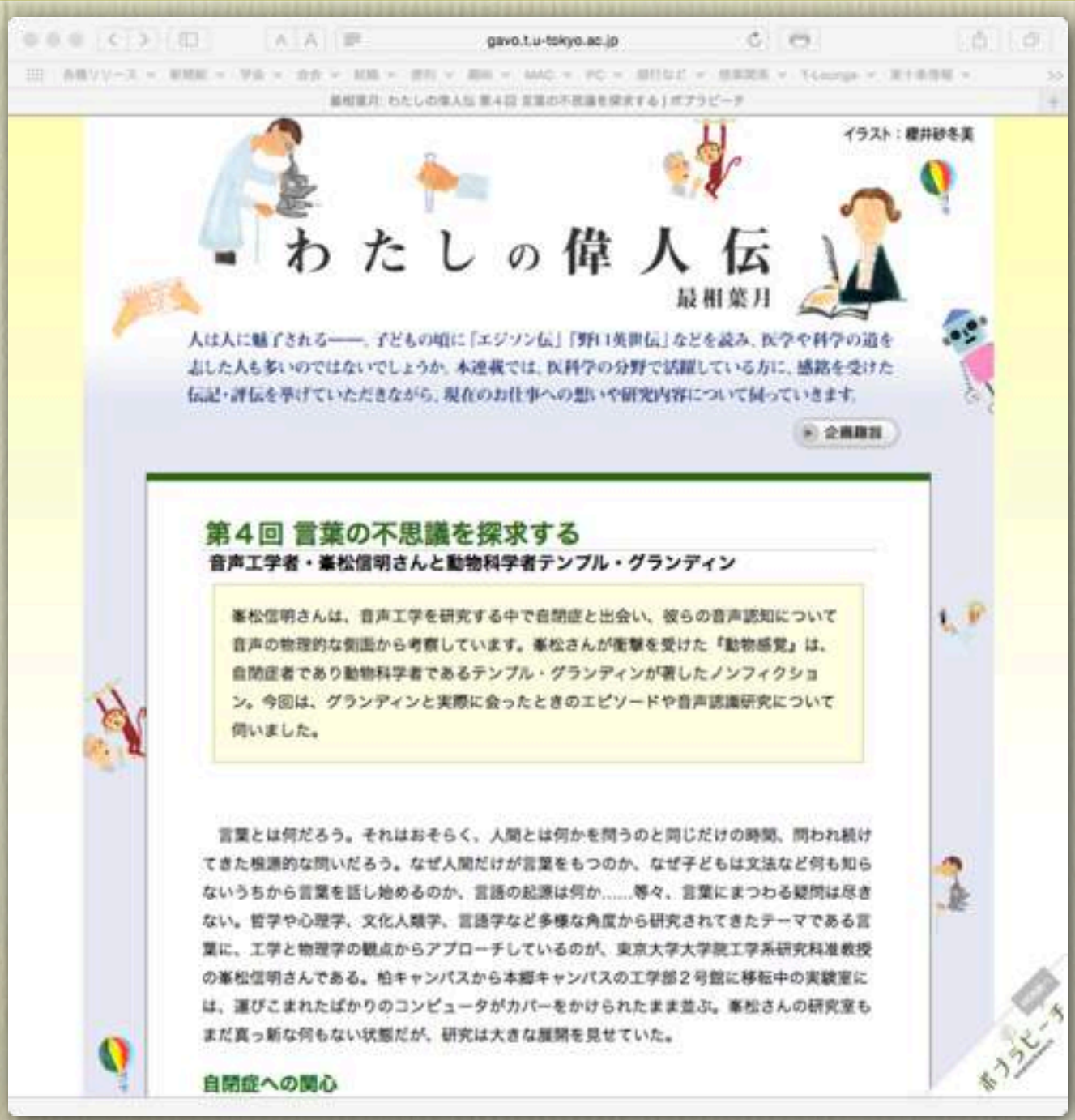




彼女と会ってきました

第六章　言葉の不思議を探究する　・・・・・・119

音声工学者・峯松信明と動物科学者テンプル・グランディンの自閉症報告

こそっと隠してあります



本当の音声の絶対音感者？

とある自閉症児が書いた本

どこにいてもどんなときでも、僕がわかる言葉は、お母さんだけです。

僕は、どうして今まで言葉が理解できないのか、わかりませんでした。他のみんなが指示されたことにすぐに反応できて、その通りに動けることが不思議でした。

僕には聞こえないのです。

音は聞こえているけれど、意味になって頭の中に入ってこないのです。話しているのが本人だとわかれば、慣れれば言っていることはわかります。でも、同じ人でも場所や状況が違うと、その人だということがわからないのです。

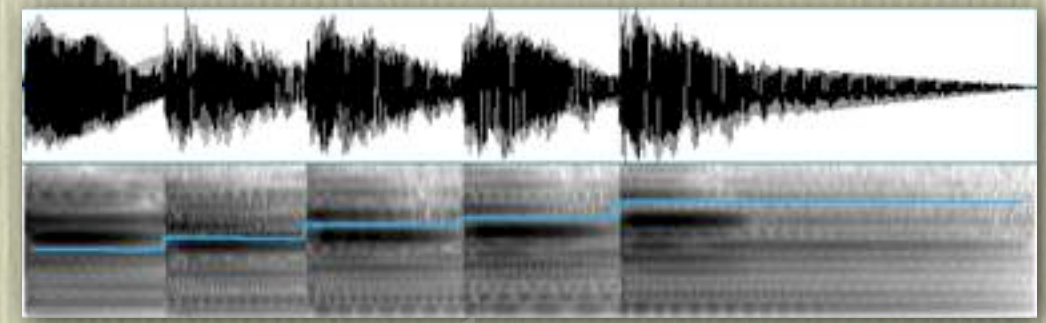
僕はお母さんの言うことならすべてわかります。それは、第1に安心感、第2に言葉のリズムや高低が良くわかっていること、第3に話の予測がつきやすいためでしょう。

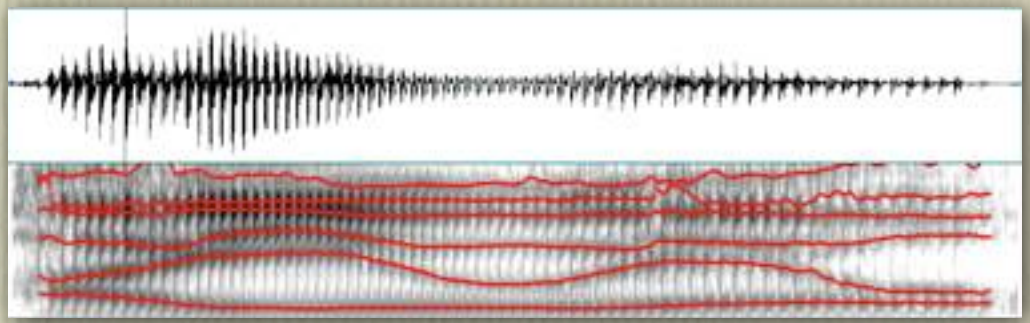
http://www.nhk.or.jp/school-blog/300/195393.html

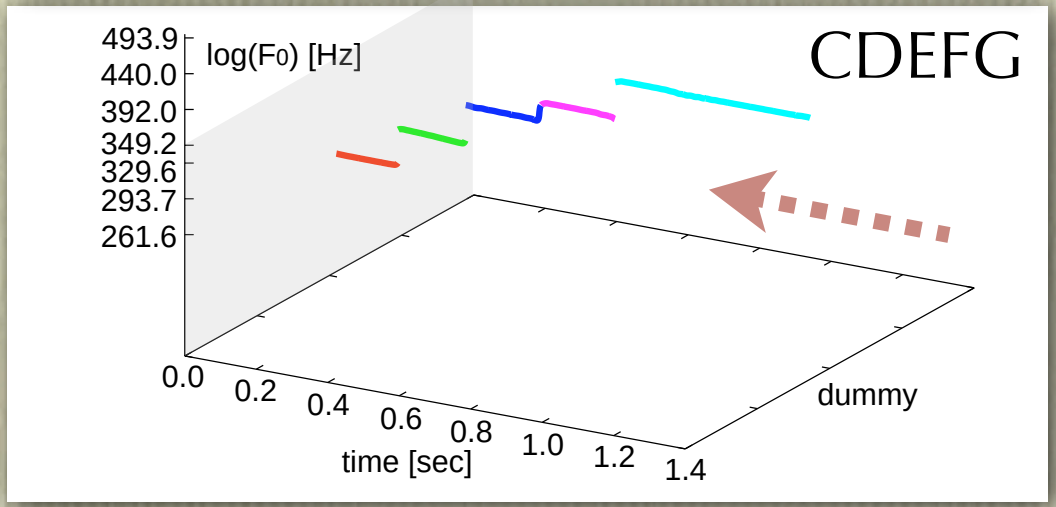
本発表の流れ

- **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？

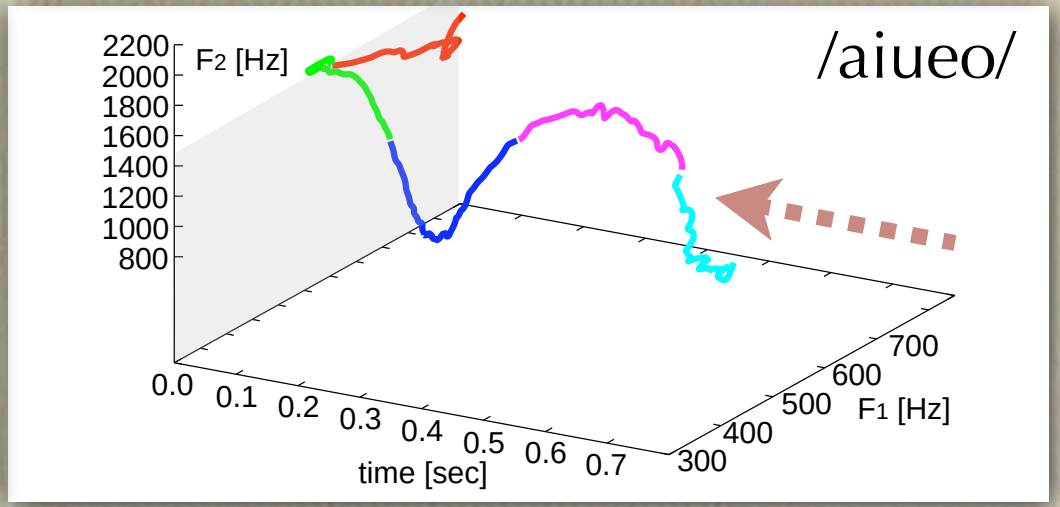
音高の相対音感／音色の相対音感



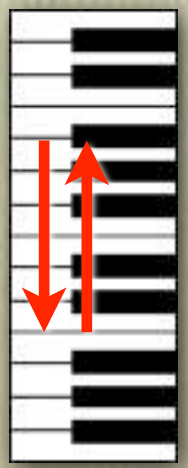




音高の動的変化パターン

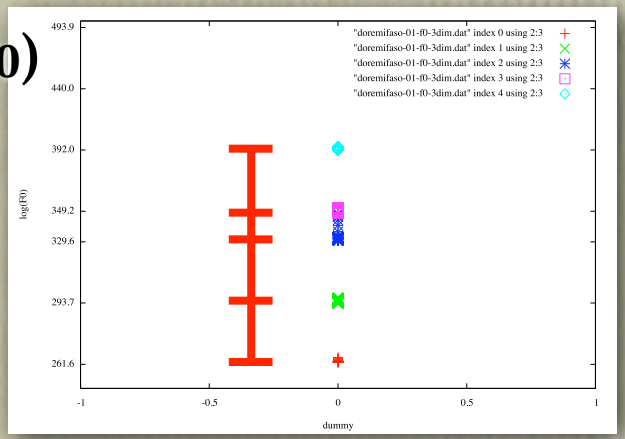


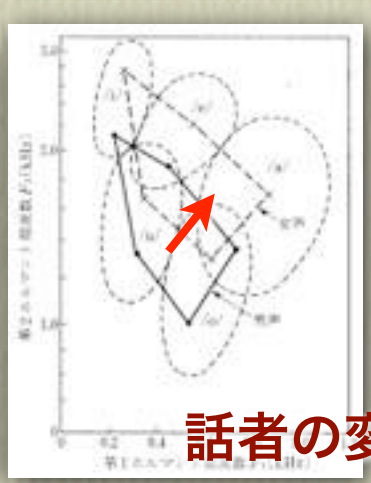
音色の動的変化パターン



調の変化

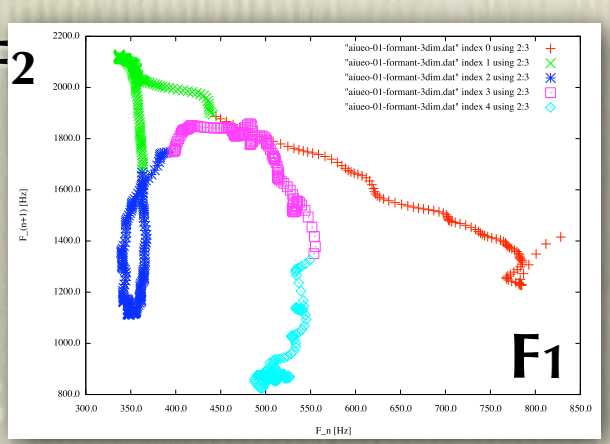
log(F0)





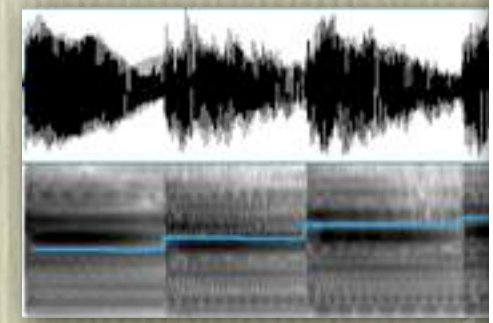
話者の変化

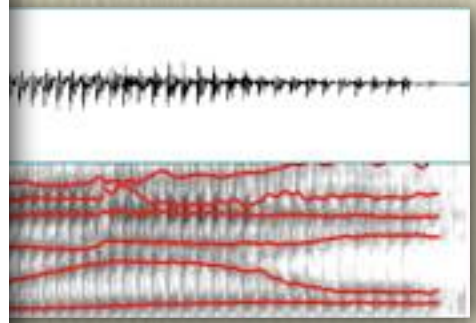
F2

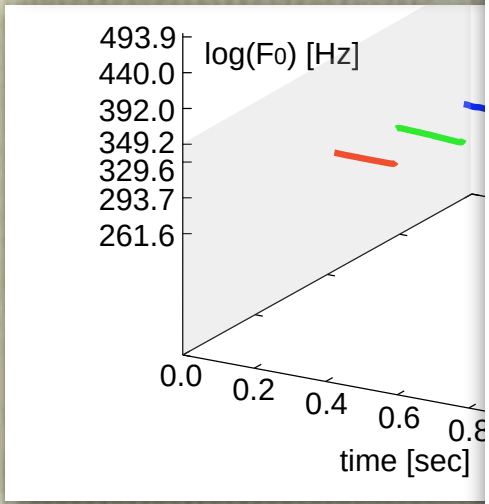


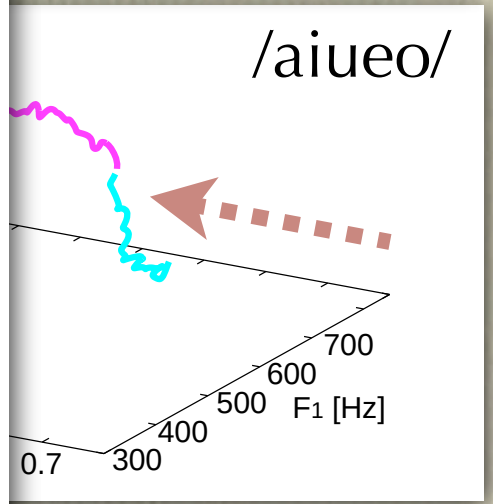
F1

音高の相対音感／音色の相対音感





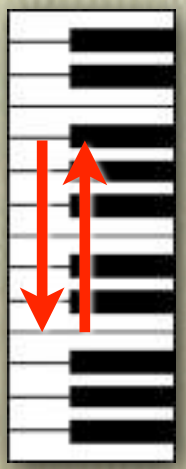


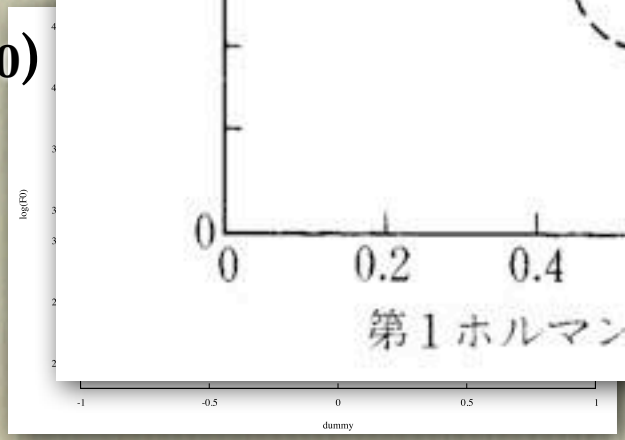


音高

動的変化パターン

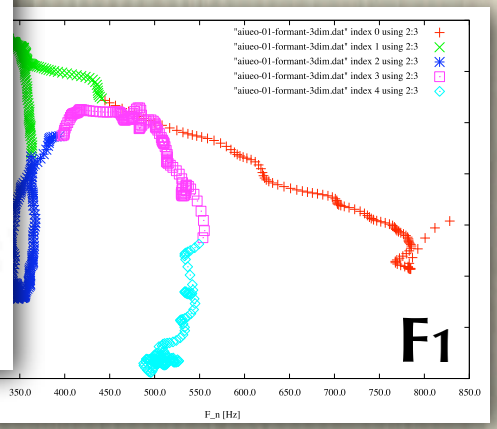
log(F0)





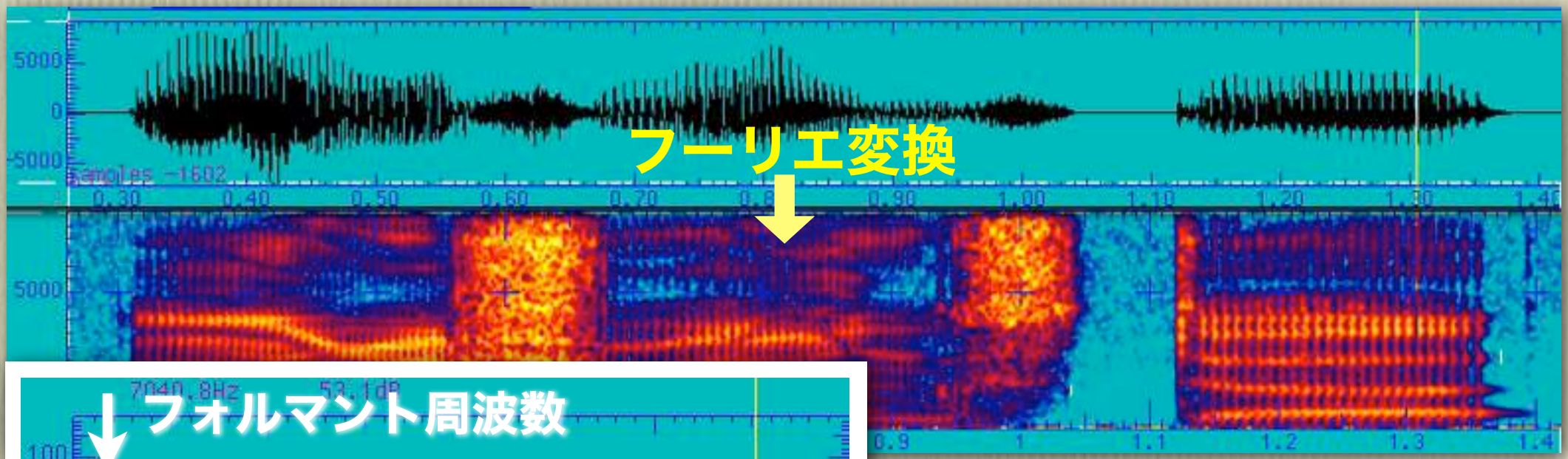
調の変化

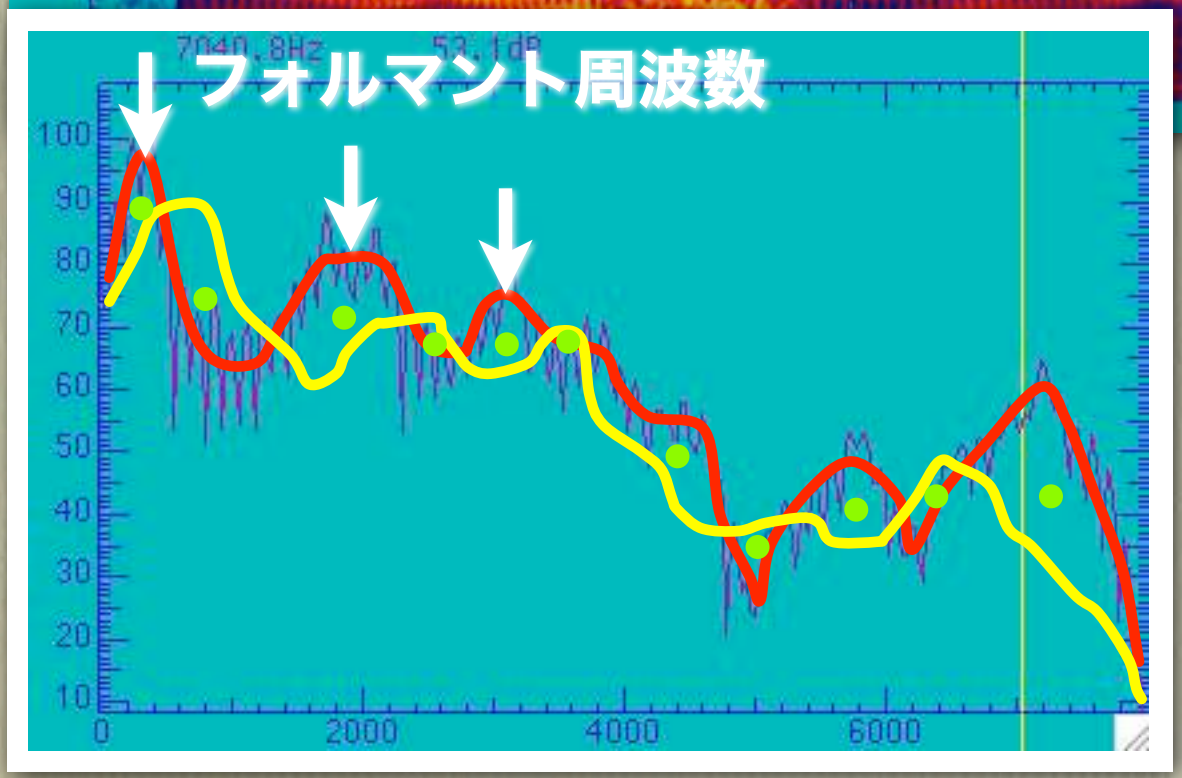
話者の変化



スペクトル包絡の効率的なベクトル表現

音声波形→スペクトラム→ケプストラム





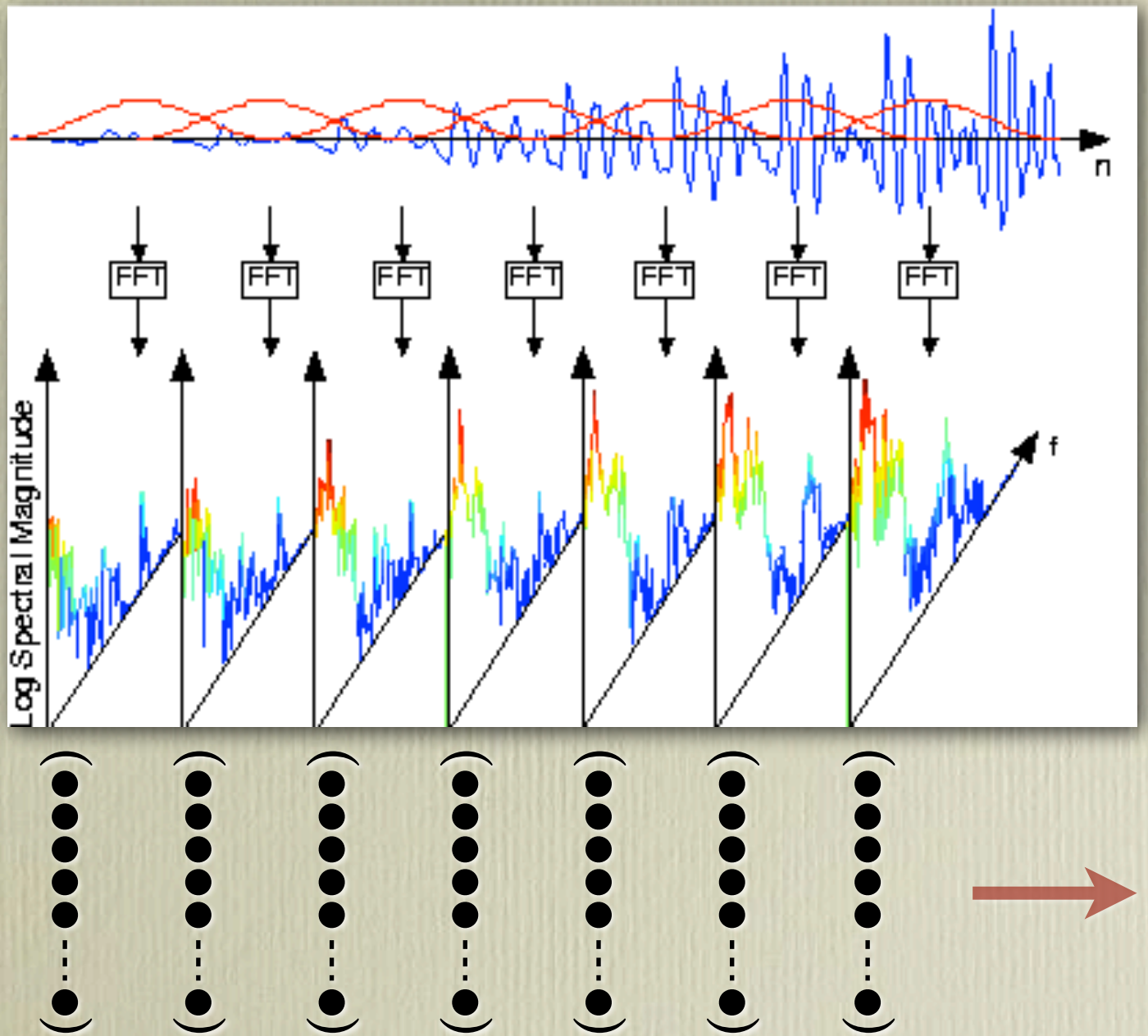
逆フーリエ変換

$$\rightarrow \quad (c_0, c_1, \cdots, c_N)$$

$$\rightarrow \quad \sum_{i=0}^{N} (c_i - c'_i)^2$$

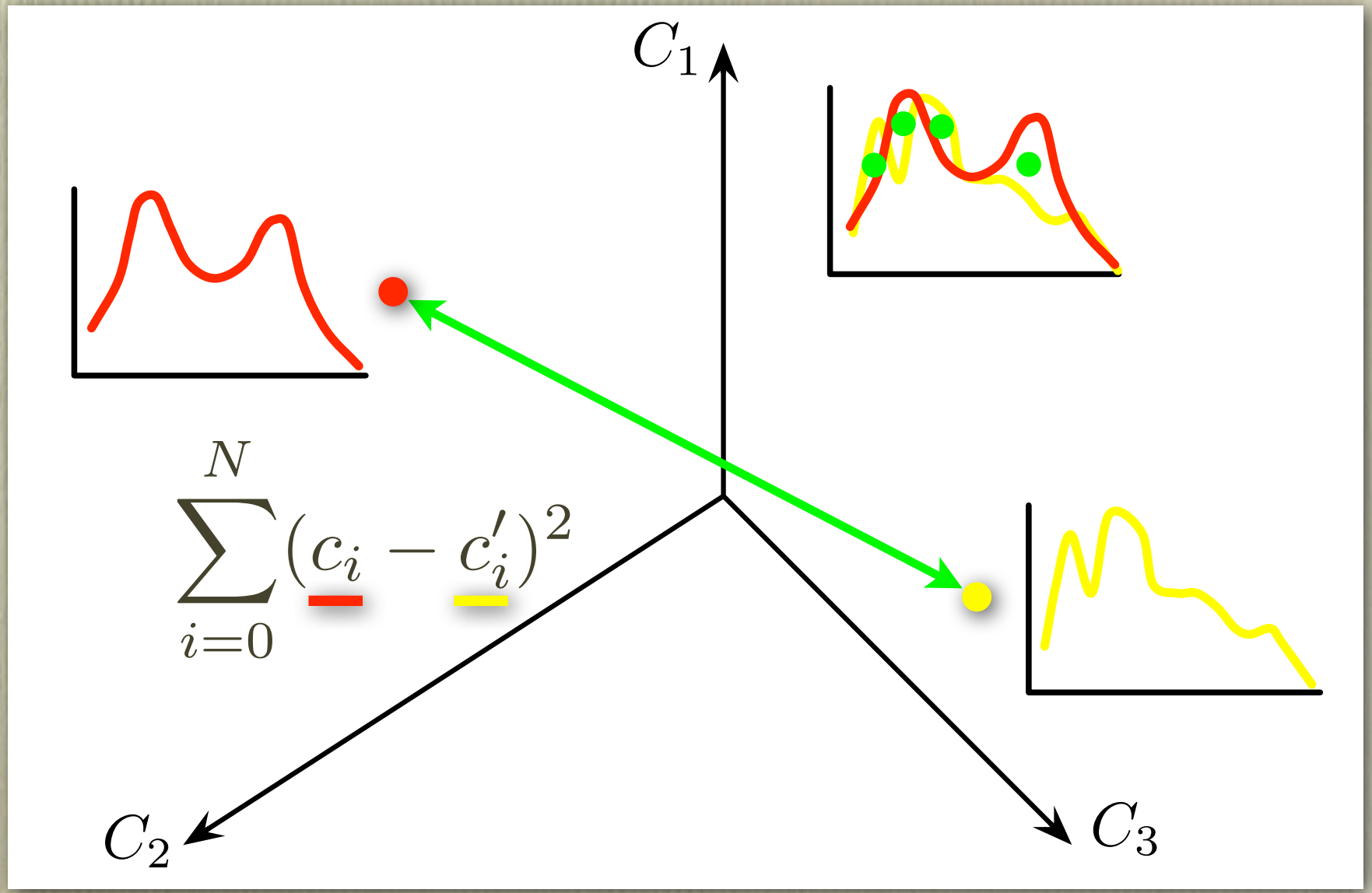
スペクトル包絡の効率的なベクトル表現

- 音声波形→スペクトラム→ケプストラム



スペクトル包絡の効率的なベクトル表現

ケプストラム空間における「点」と「点間距離」

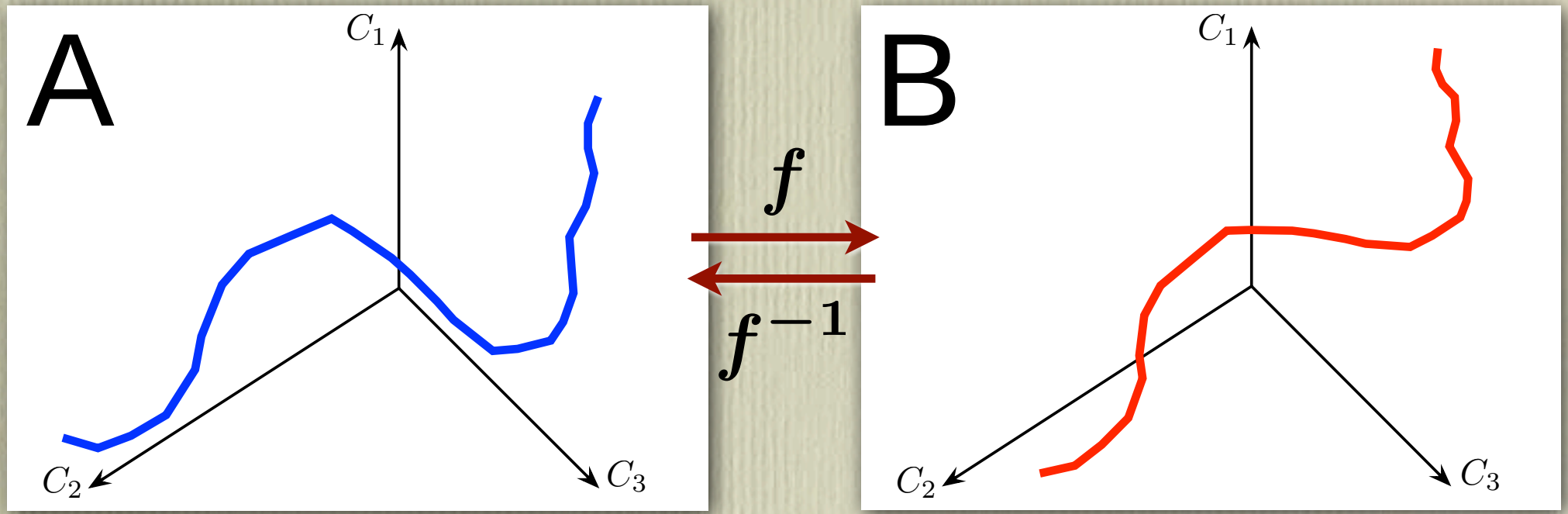


点＝スペクトル包絡，点間距離＝スペクトル間差異

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 ⟷ 話者Ｂの音響空間



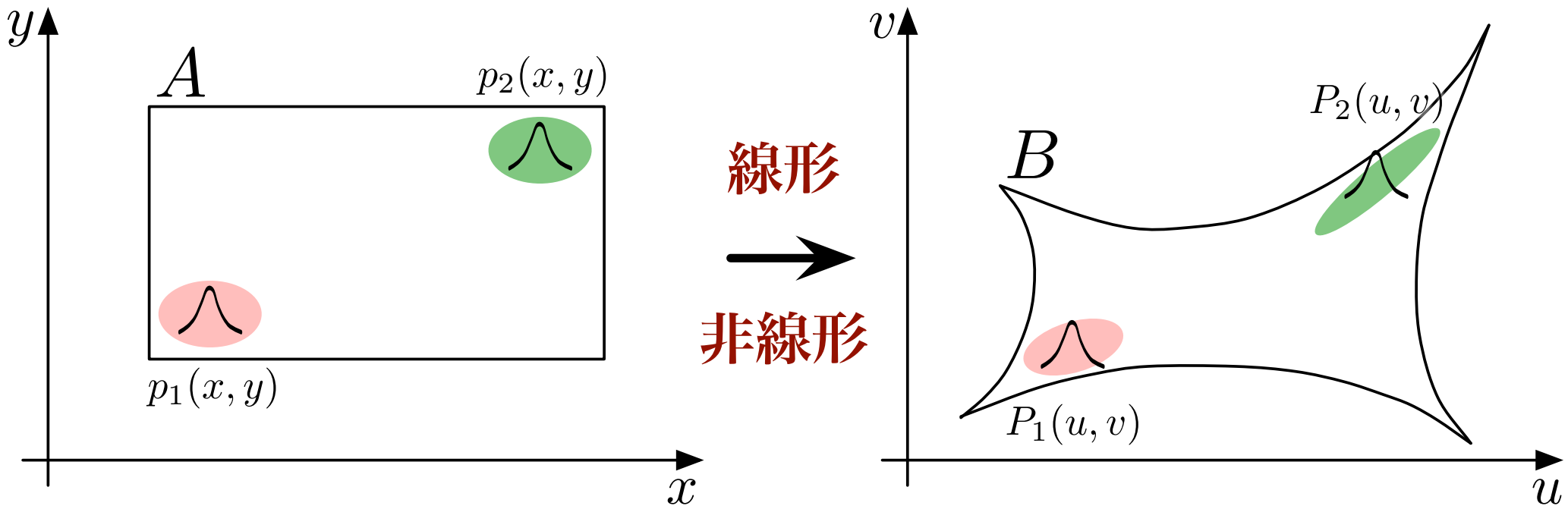
話者Ａの声を65億全ての話者の声へ変形する

- 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

二人の話者空間（一対一対応）における不変音響量

- 音響事象を点ではなく，分布として表現する。
  - 空間Aの分布 p は空間Bの分布 P へと写像される。
  - p と P は異なる物理特性を持つ（［あ］と［あ］）
  - 両空間において不変な物理量はどこに？不変コントラストは存在する？



変換不変な音響量の数学的探求

変数変換と積分

- 一変数：$x = x(t) \quad (x_1 = x(t_1), x_2 = x(t_2))$

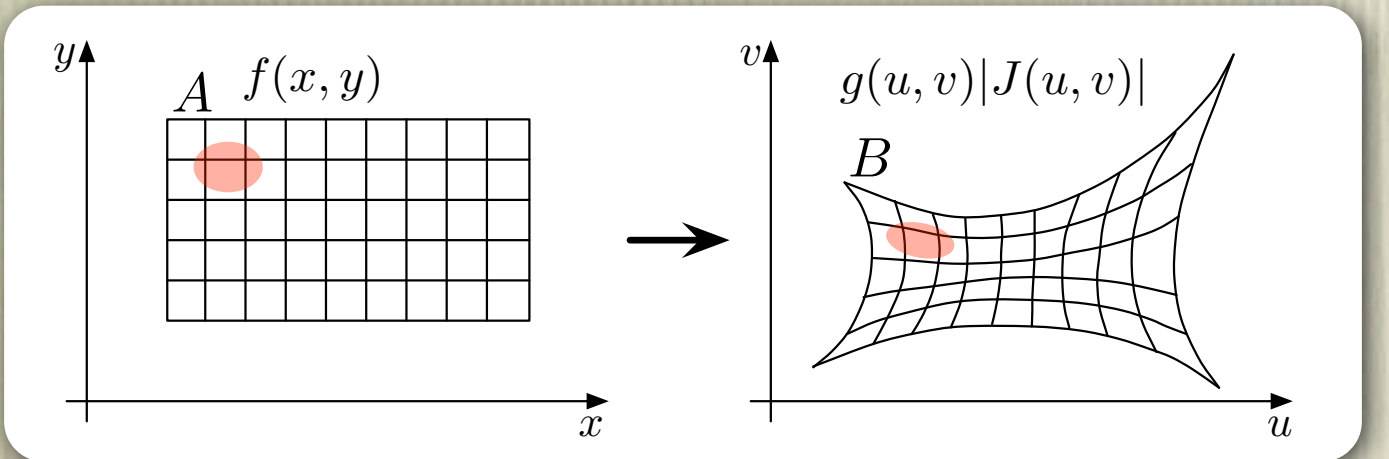
$$\int_{x_1}^{x_2} f(x)dx = \int_{t_1}^{t_2} f(x(t))\frac{dx(t)}{dt}dt = \int_{t_1}^{t_2} g(t)x'(t)dt$$

- 二変数：$x = x(u,v),\ y = y(u,v)$

$$\begin{aligned} x &= 3u + 2v - 5 \\ y &= 4u + 5v + 3 \end{aligned}$$

$$\iint_A f(x,y)dxdy = \iint_B f(x(u,v), y(u,v))|J(u,v)|dudv$$

$$= \iint_B g(u,v)|J(u,v)|dudv \qquad J(u,v) \equiv \frac{\partial(x,y)}{\partial(u,v)} \equiv \det \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$



変換不変な音響量の数学的探求

変数変換と確率密度分布関数

- 一変数：$x = x(t) \quad (x_1 = x(t_1), x_2 = x(t_2))$

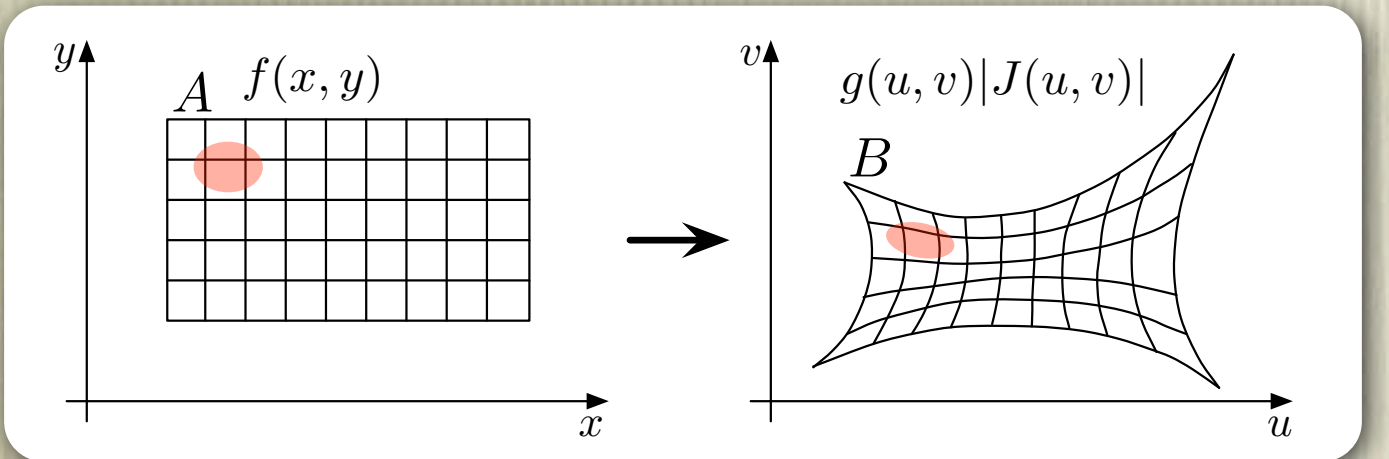
$$1.0 = \int_{x_1}^{x_2} p(x)dx = \int_{t_1}^{t_2} p(x(t))\frac{dx(t)}{dt}dt = \int_{t_1}^{t_2} q(t)x'(t)dt$$

- 二変数：$x = x(u,v),\ y = y(u,v)$

$$\begin{aligned} x &= 3u + 2v - 5 \\ y &= 4u + 5v + 3 \end{aligned}$$

$$1.0 = \iint_A f(x,y)dxdy = \iint_B f(x(u,v), y(u,v))|J(u,v)|dudv$$

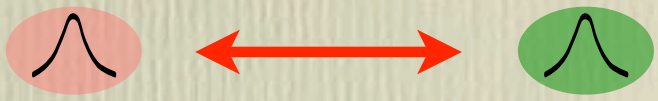
$$= \iint_B g(u,v)|J(u,v)|dudv \qquad J(u,v) \equiv \frac{\partial(x,y)}{\partial(u,v)} \equiv \det \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$



変換不変な音響量の数学的探求

バタチャリヤ距離（＝分布間距離尺度の一つ）

- 座標変換による式の変形 $x = x(u,v),\ y = y(u,v)$
- $BD(p_1(x,y), p_2(x,y))$



$$= -\log \iint \sqrt{p_1(x,y)p_2(x,y)}dxdy$$

$$= -\log \iint \sqrt{q_1(u,v)q_2(u,v)}|J(u,v)|dxdy$$

$$= -\log \iint \sqrt{q_1(u,v)|J(u,v)| \cdot q_2(u,v)|J(u,v)|}dudv$$

$$= -\log \iint \sqrt{P_1(u,v)P_2(u,v)}dudv$$

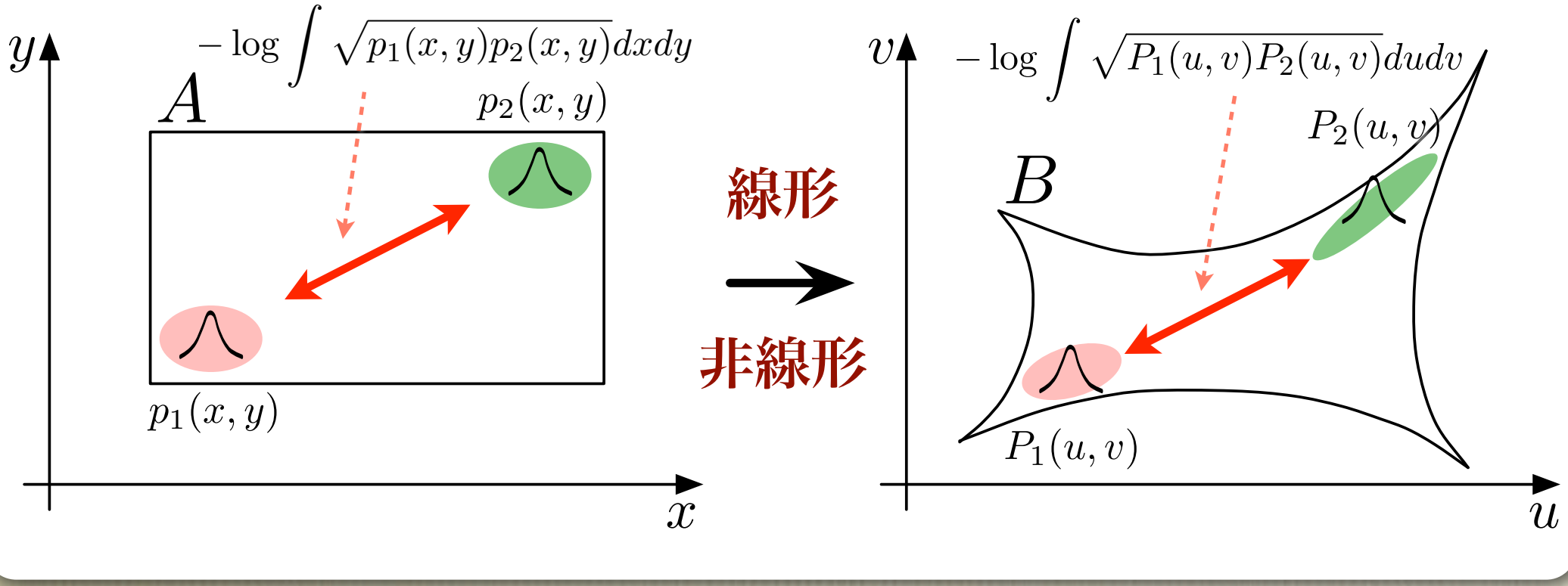
$$= BD(P_1(u,v), P_2(u,v))$$

$q_1(u,v) = p_1(x(u,v), y(u,v)),$ $\qquad J = \text{Jacobian}$

変換不変な音響量の数学的探求

二人の話者空間（一対一対応）における不変音響量

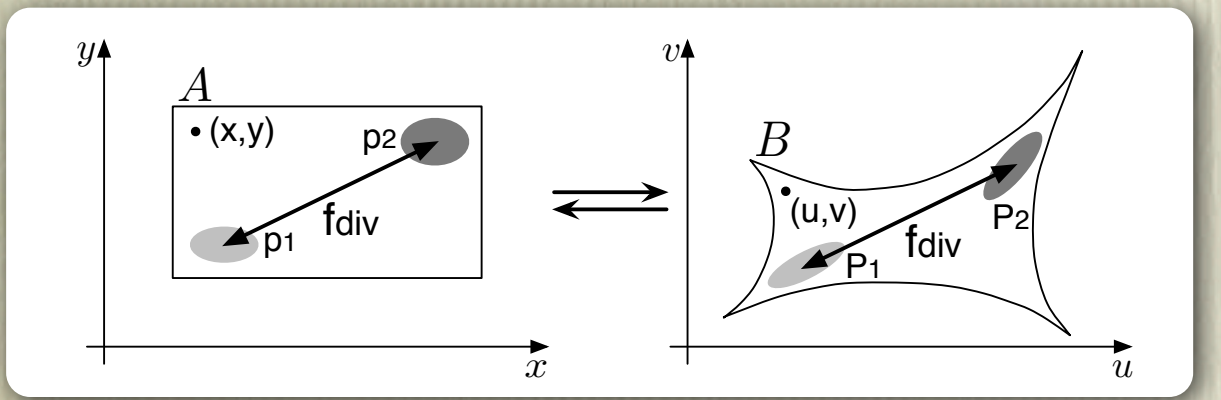
- 音響事象を点ではなく，分布として表現する。
  - 空間Aの分布 p は空間Bの分布 P へと写像される。
  - p と P は異なる物理特性を持つ（［あ］と［あ］）
  - 両空間において不変な物理量はどこに？不変コントラストは存在する？
  - 各事象は可変，しかし，少なくともバタチャリヤ距離は不変。



変換不変な音響量の数学的探求

変換不変量の一般式はあるのか？

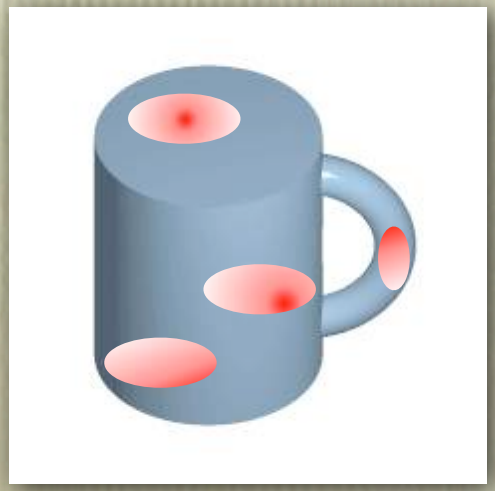
- f-divergence 不変性の十分性
  - $f_{div}(p_1, p_2) = \int p_2(\boldsymbol{x}) g\left(\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\right) d\boldsymbol{x}$
  - $g(t) = t\log(t) \rightarrow f_{div} = \mathrm{KL-div.} \qquad g(t) = \sqrt{t} \rightarrow -\log(f_{div}) = \mathrm{BD}$
  - $f_{div}(p_1, p_2) = f_{div}(P_1, P_2)$
- f-divergence 不変性の必要性
  - $\int M(p_1(\boldsymbol{x}), p_2(\boldsymbol{x}))d\boldsymbol{x}$ が如何なる可逆＆連続の変換に対しても不変
  - この場合, $M = p_2(\boldsymbol{x}) g\left(\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\right)$ であることが必要。

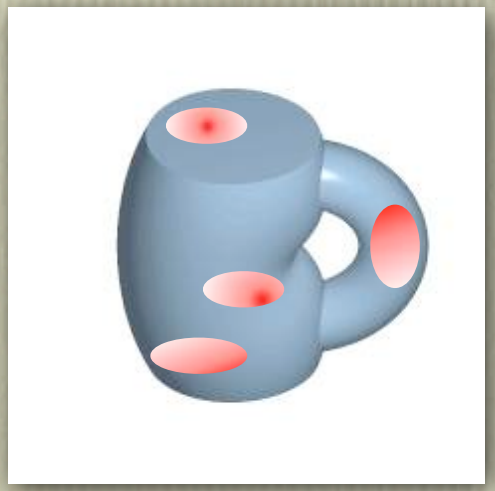


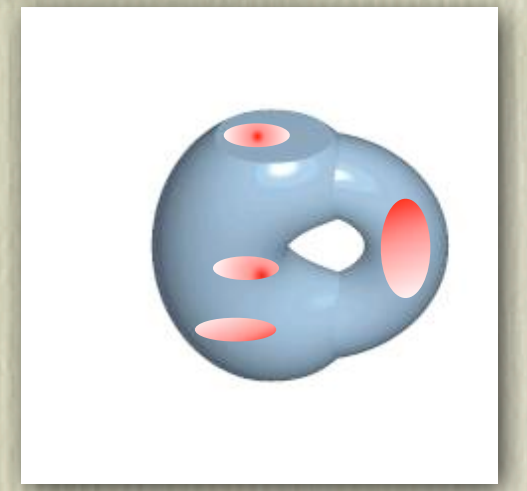
変換不変な音響量の数学的探求

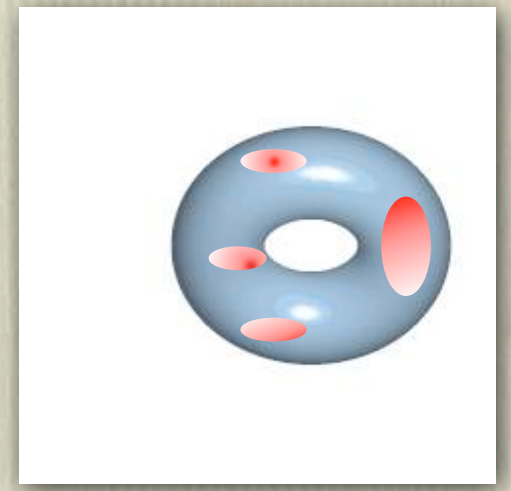
位相幾何学（トポロジー）における不変量

- 連続かつ可逆な任意の変形を施しても不変なる幾何学的性質





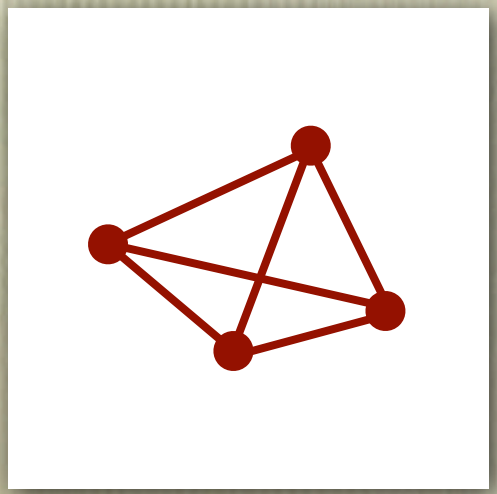


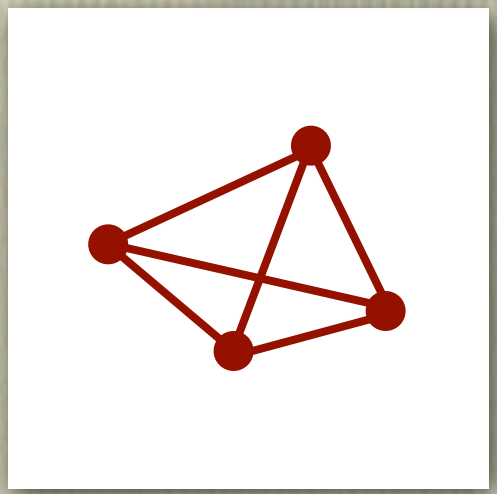


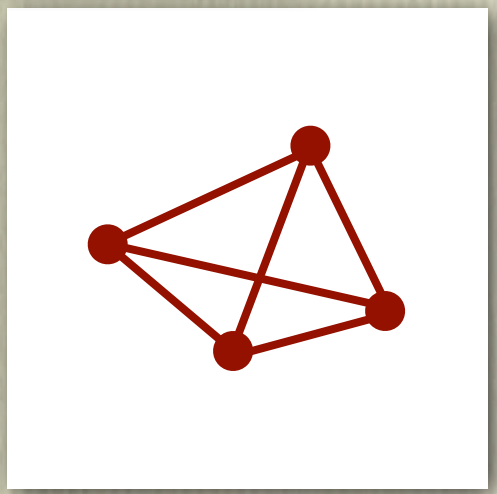
変換不変な音響量の数学的探求

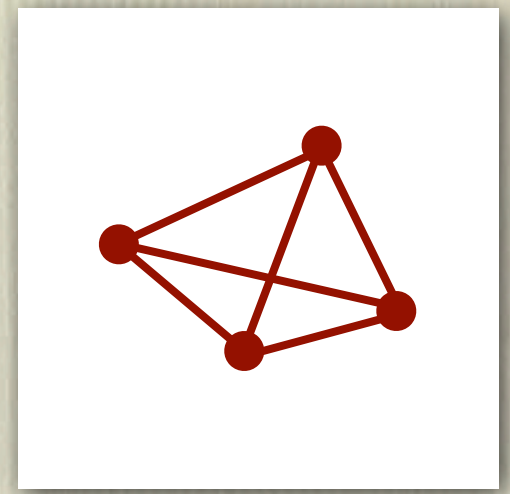
位相幾何学（トポロジー）における不変量

- 連続かつ可逆な任意の変形を施しても不変なる幾何学的性質



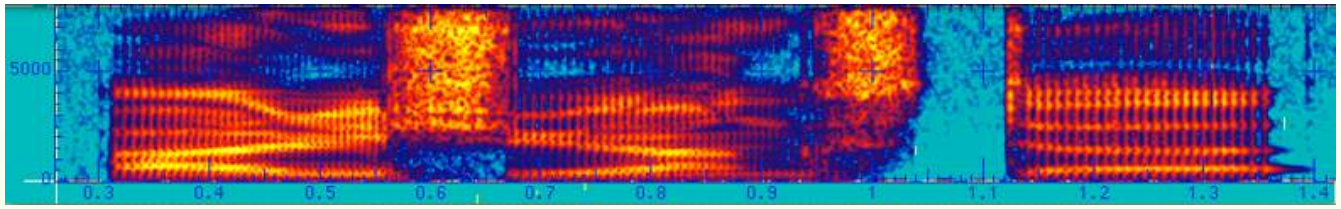




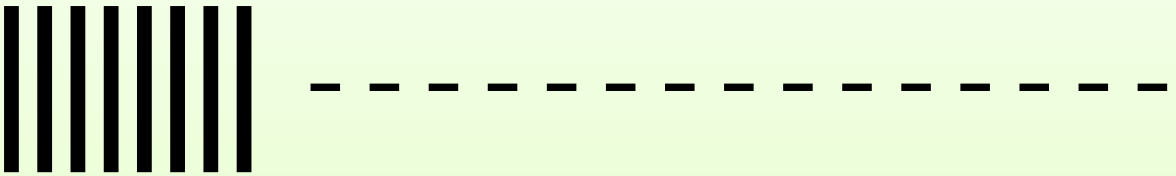


分布間距離群としての音声表象

頑健に話者不変な音声表象＝構造的・全体的表象

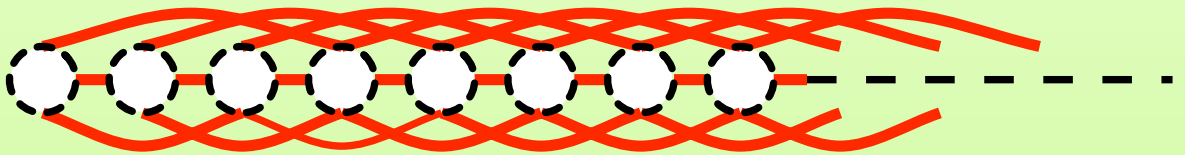


Sequence of spectrum slices

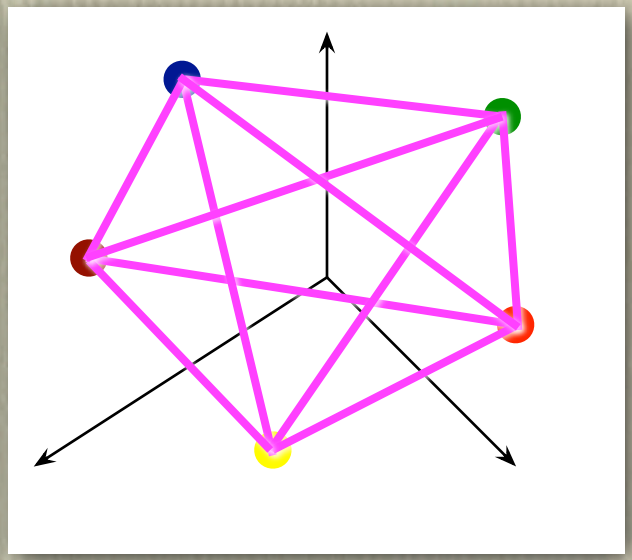


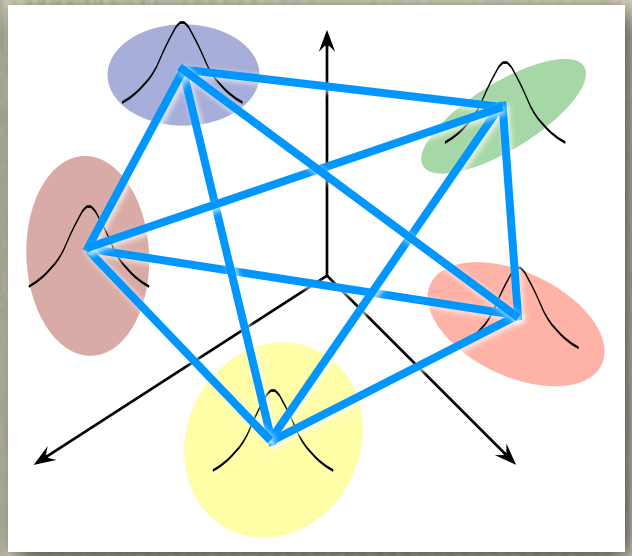
Sequence of cepstrum vectors

Sequence of distributions



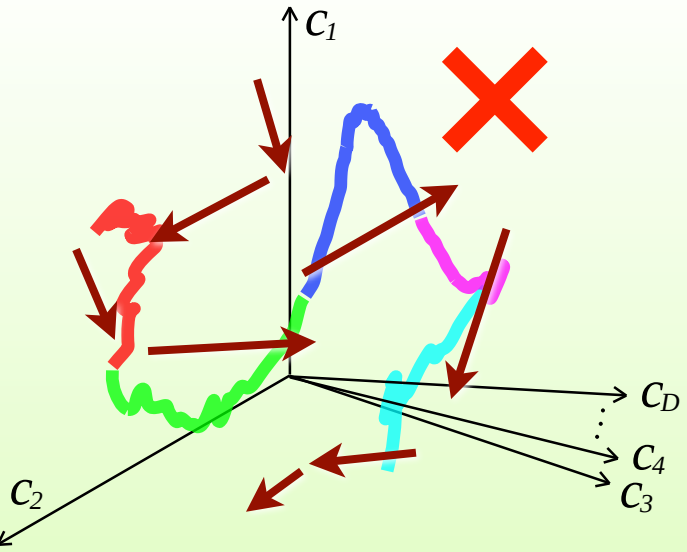
Structuralization by interrelating temporally-distant events

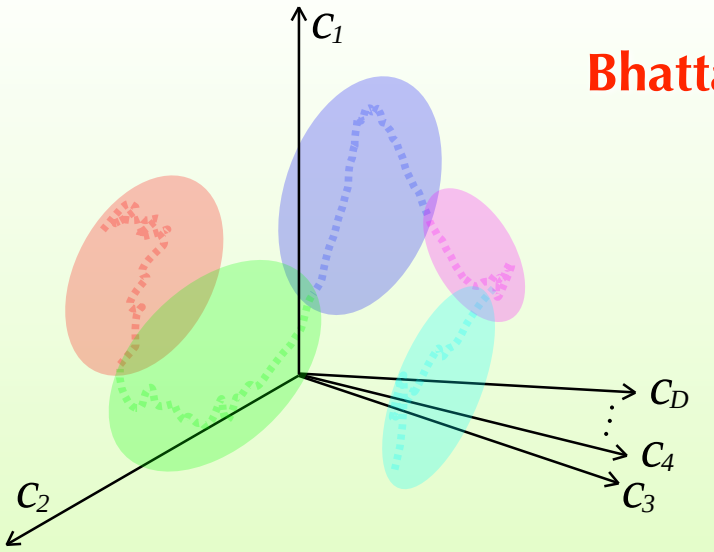


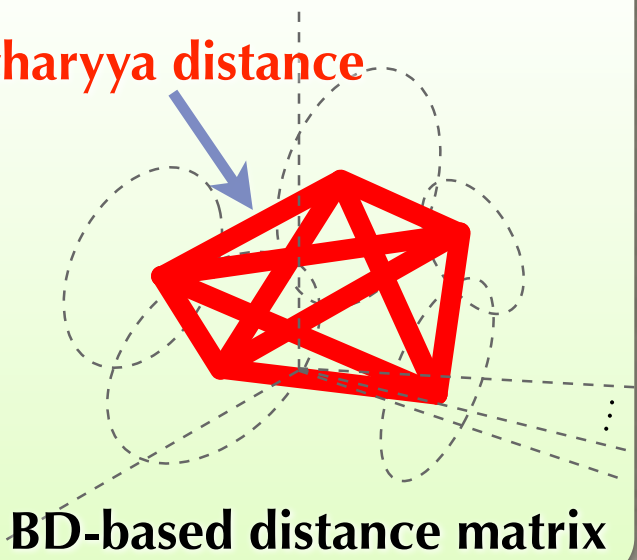


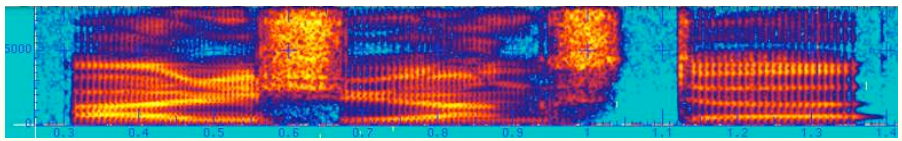
分布間距離群としての音声表象

ケプトラム系列 → 分布系列 → 距離行列









spectrogram (spectrum slice sequence)

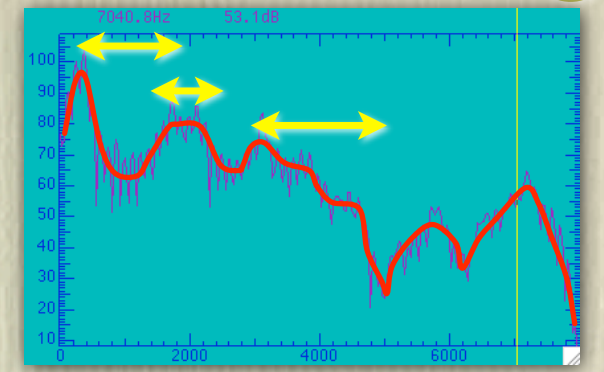
cepstrum vector sequence

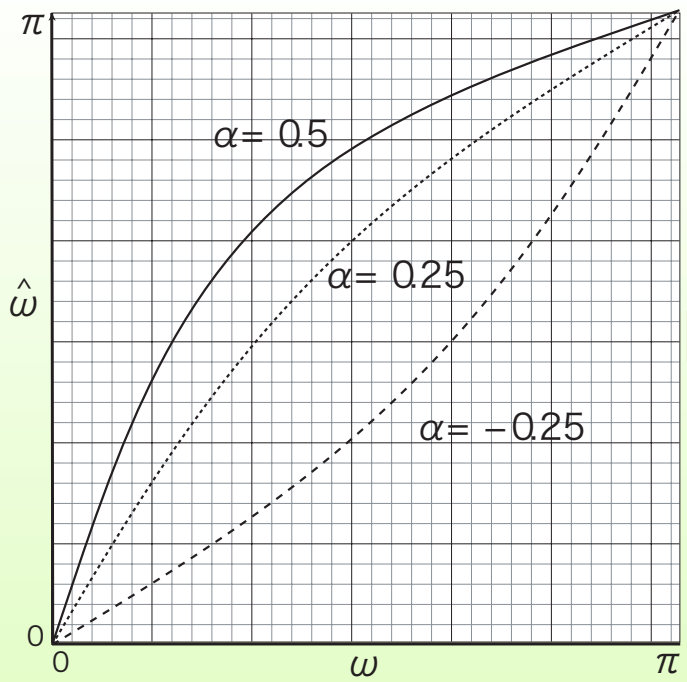
distribution sequence

声道長の変化＝行列 A の掛け算

具体的な行列 A の実装は？

- 声道長が伸びる＝フォルマントがより低く
- 声道長が縮む　＝フォルマントがより高く
- → スペクトルに対する周波数ウォーピング





$$\hat{\boldsymbol{c}} = (\hat{c}_1\ \hat{c}_2\ \hat{c}_3\ \hat{c}_4 \cdots)^t$$

$$\boldsymbol{A} = \begin{pmatrix} 1-\alpha^2 & 2\alpha-2\alpha^3 & \cdots & \cdots \\ -\alpha+\alpha^3 & 1-4\alpha^2+3\alpha^4 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$\boldsymbol{c} = (c_1\ c_2\ c_3\ c_4 \cdots)^t.$$

$$a_{ij} = \frac{1}{(j-1)!} \sum_{m=\max(0,j-i)}^{j} \binom{j}{m} \times \frac{(m+i-1)!}{(m+i-j)!} (-1)^m \alpha^{(2m+i-j)}$$

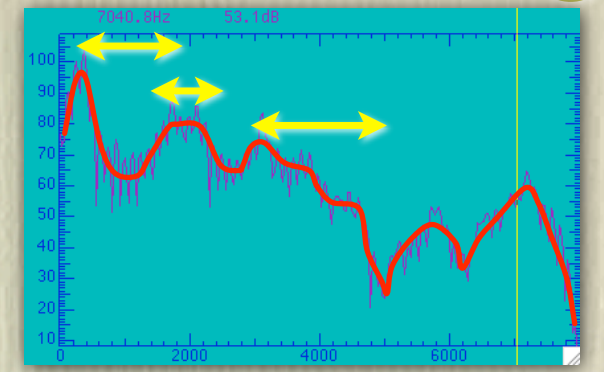
$$\hat{z}^{-1} = \frac{z^{-1}-\alpha}{1-\alpha z^{-1}},\quad z = e^{j\omega},\quad \hat{z} = e^{j\hat{\omega}}$$

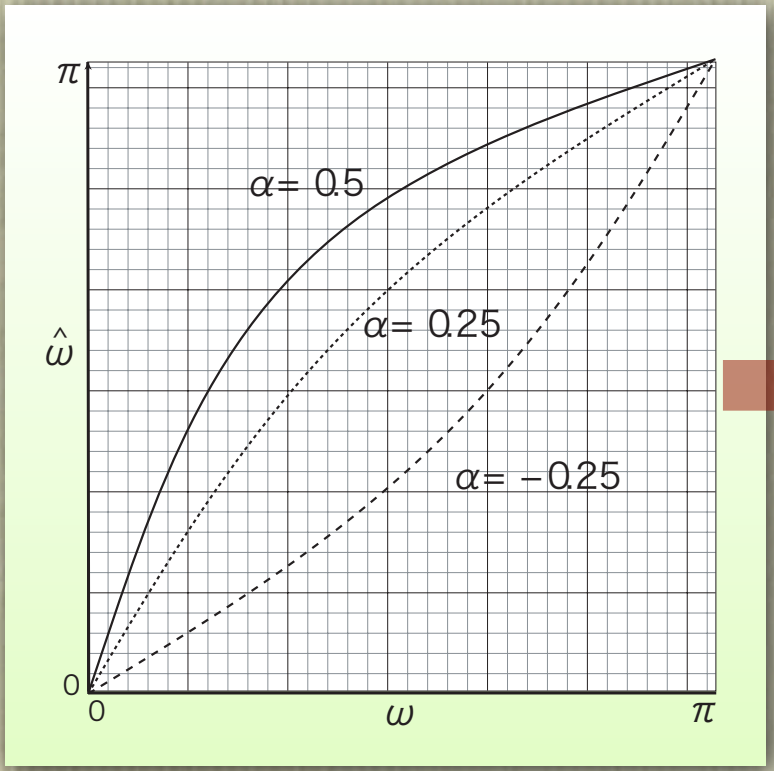
$$\boldsymbol{c'} = \boldsymbol{Ac}$$

声道長の変化＝行列 A の掛け算

具体的な行列 A の実装は？

- 声道長が伸びる＝フォルマントがより低く
- 声道長が縮む　＝フォルマントがより高く
- → スペクトルに対する周波数ウォーピング





$$\hat{\boldsymbol{c}} = (\hat{c}_1\ \hat{c}_2\ \hat{c}_3\ \hat{c}_4\ \cdots)^t$$

$$\boldsymbol{A} = \begin{pmatrix} 1-\alpha^2 & 2\alpha-2\alpha^3 & \cdots & \cdots \\ -\alpha+\alpha^3 & 1-4\alpha^2+3\alpha^4 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$\boldsymbol{c} = (c_1\ c_2\ c_3\ c_4\ \cdots)^t.$$

$$a_{ij} = \frac{1}{(j-1)!} \sum_{m=\max(0,j-i)}^{j} \binom{j}{m} \times \frac{(m+i-1)!}{(m+i-j)!} (-1)^m \alpha^{(2m+i-j)}$$

$$\hat{z}^{-1} = \frac{z^{-1}-\alpha}{1-\alpha z^{-1}},\quad z = e^{j\omega},\quad \hat{z} = e^{j\hat{\omega}}$$

$$\boldsymbol{c'} = \boldsymbol{Ac}$$

行列 A の幾何学的性質

$$\begin{pmatrix} \hat{c}_1 \\ \hat{c}_2 \end{pmatrix} = \begin{pmatrix} 1-\alpha^2 & 2\alpha-2\alpha^3 \\ -\alpha+\alpha^3 & 1-4\alpha^2+3\alpha^4 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

$$\boldsymbol{T} = \boldsymbol{R} + \boldsymbol{O}$$

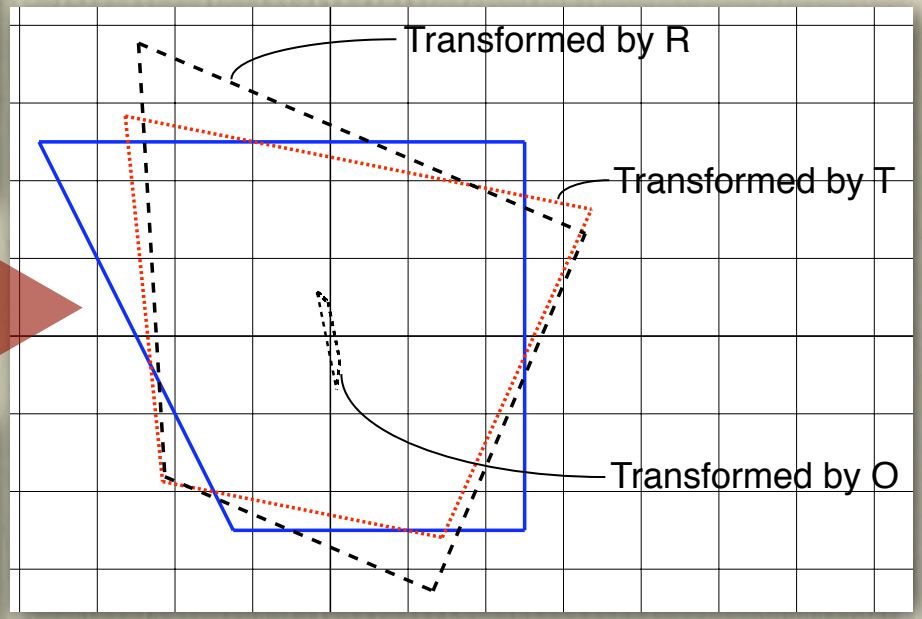
$$\boldsymbol{R} = \begin{pmatrix} 1-2\alpha^2 & 2\alpha(1-\frac{1}{2}\alpha^2) \\ -2\alpha(1-\frac{1}{2}\alpha^2) & 1-2\alpha^2 \end{pmatrix}$$

$$\boldsymbol{O} = \begin{pmatrix} \alpha^2 & -\alpha^3 \\ -\alpha & -2\alpha^2+3\alpha^4 \end{pmatrix}.$$

$$\boldsymbol{A} = \begin{pmatrix} 1-\alpha^2 & 2\alpha-2\alpha^3 & \cdots & \cdots \\ -\alpha+\alpha^3 & 1-4\alpha^2+3\alpha^4 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$\boldsymbol{R} \simeq \begin{pmatrix} 1-2\alpha^2 & 2\alpha\sqrt{1-\alpha^2} \\ -2\alpha\sqrt{1-\alpha^2} & 1-2\alpha^2 \end{pmatrix}$$

$$= \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{pmatrix} \quad (\alpha = \sin\theta)$$



N次元ではどうなる？

行列 Aの幾何学的性質

N次元空間における回転行列とは？

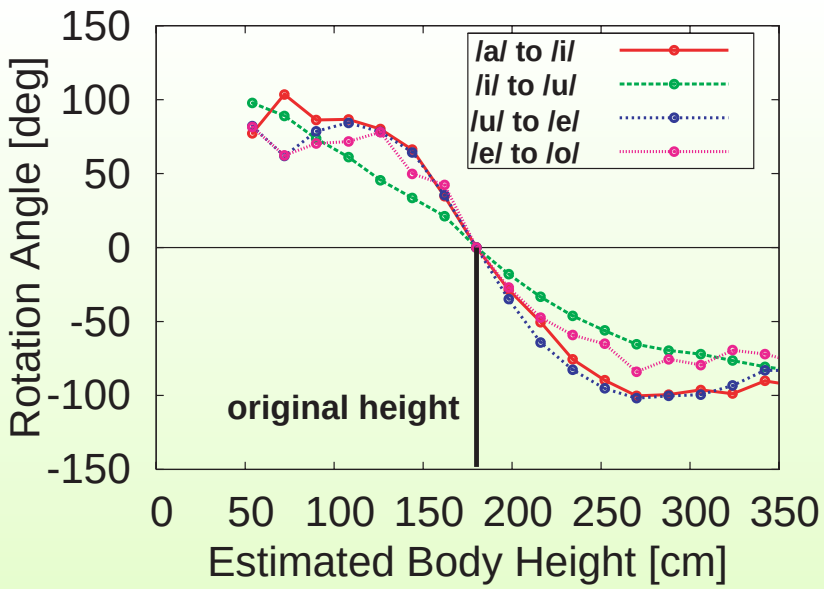
$$R^t R = R R^t = I$$

$$\det R = +1.$$

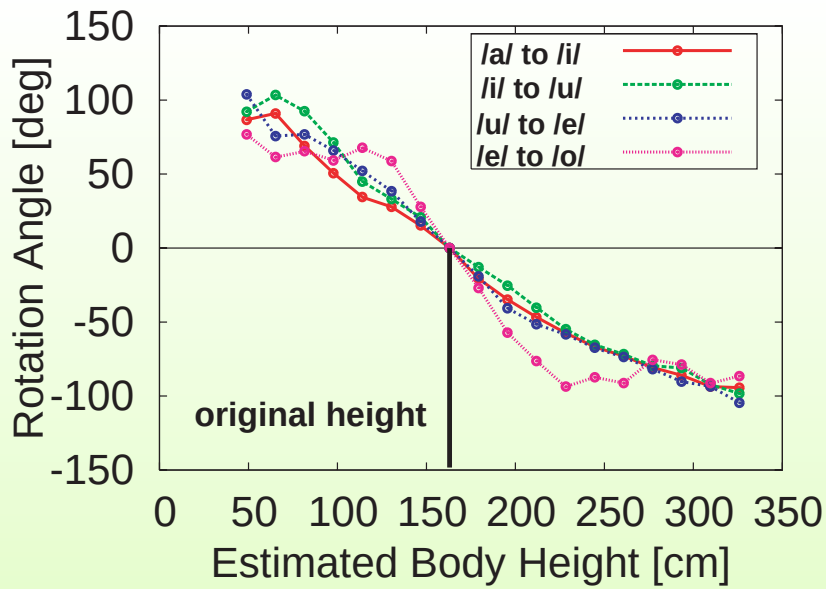
近似的に成立！

$$a_{ij} = \frac{1}{(j-1)!} \sum_{m=\max(0,j-i)}^{j} \binom{j}{m} \times \frac{(m+i-1)!}{(m+i-j)!} (-1)^m \alpha^{(2m+i-j)}$$

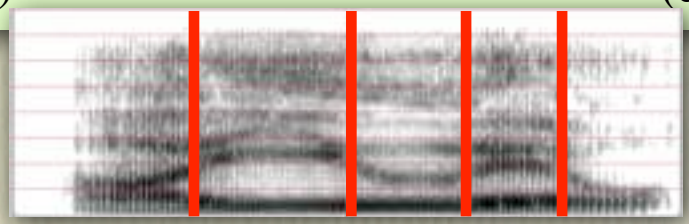
周波数ウォーピングはケプストラムを回転させる！



(a):MFCC (male)

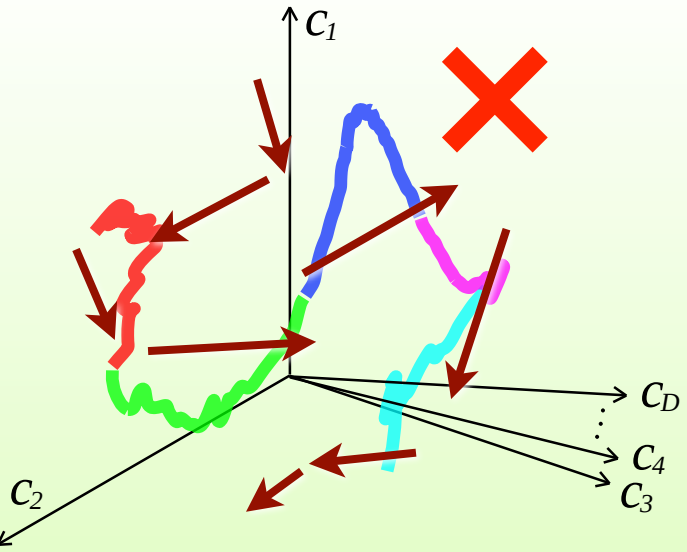


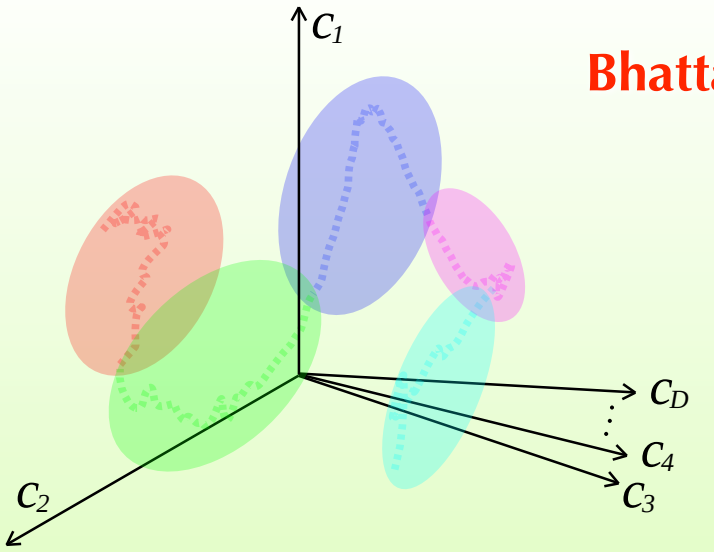
(d):MFCC (female)

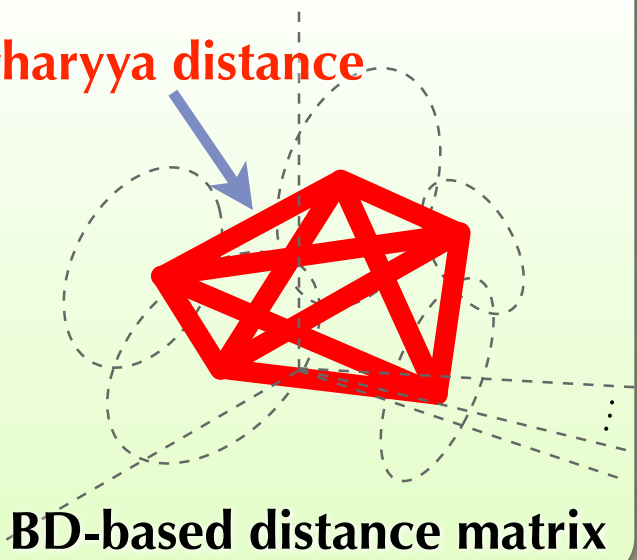


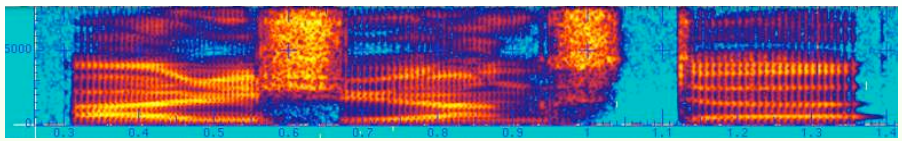
分布間距離群としての音声表象

ケプトラム系列 → 分布系列 → 距離行列









spectrogram (spectrum slice sequence)

cepstrum vector sequence

distribution sequence

音高の相対音感／音色の相対音感

音高＝基本周波数＝１次元の量

- 音の高さ＝直感的に理解しやすい。
  - 単旋律のメロディーを聴いて，音高の動きの様子は把握しやすい。
  - 鼻歌聞かせて「メロディーを描いて」と言えば，指で描ける。
  - 言葉としても「高い↔低い」の対義語で事足りる。
  - 一次元の量だから，その動きの様子を「視覚的に」捉えやすいから？

音色＝周波数軸のエネルギー分布＝多次元の量

- 音の音色＝何か「もやもや」していて掴みどころがない。
  - 「あいうえお」と聞いて，音色の動きの様子を把握できる？
  - その動きを「描いて」と言われても，どう描くべきかすら分らない。
  - 言葉としても「太い↔細い」「しぶい↔若い」など色々。
  - 多次元の量だから，その動きの様子は「視覚的に」捉えられない？
    - 四次元をありありと感覚できる数学者なら捉えられる？
- **隣接音だけでなく，離れた音とのコントラストも必要**

面白い事実

Dyslexia であることの利点

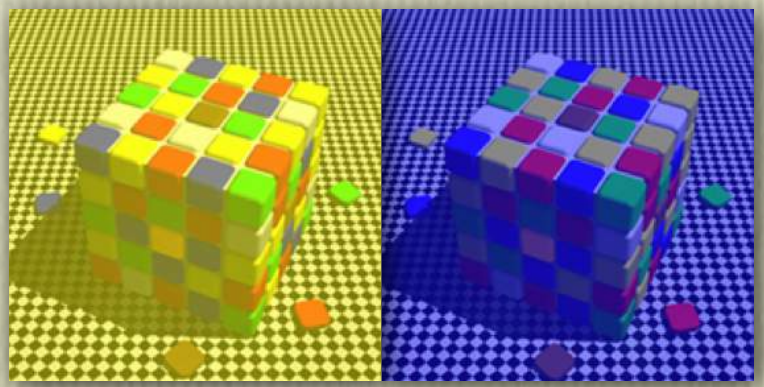
- 空間把握能力
  - 空間における物体の形，大きさ，動き，位置，位置関係，及びそれらの相互関係を把握する能力
- つながりを把握する能力
  - 異なる事物や概念，出来事の相互関係を見抜く力。様々な領域のアプローチやテクニックを使い，物事を様々な視点から見る力
- 物語を作る能力
  - 過去の個人的経験の心的場面を繋ぎ合わせて，過去・現在をリアルに思い出したり，未来をリアルに描く力
- 未来を予測する能力
  - エピソードのシミュレーションを使い，過去や未来の状態を正確に予測する力

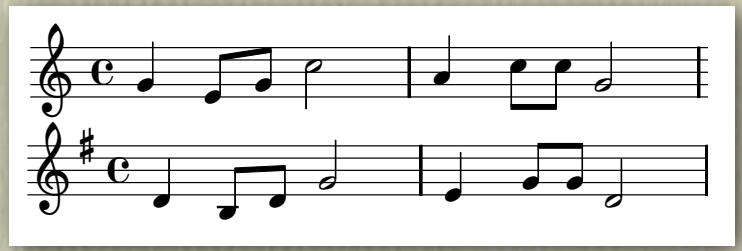
http://ondyslexia.blogspot.jp/2013/03/blog-post_20.html

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

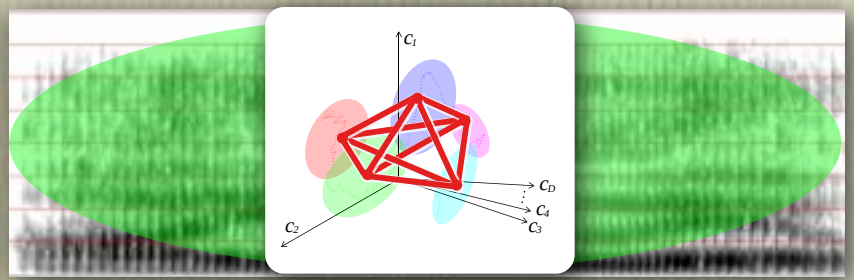
- コントラスト情報に基づく処理が重要
- コントラスト群から成る全体的パターン処理が要素同定を可能

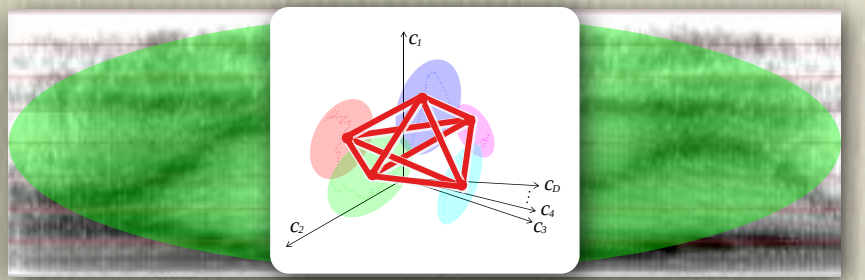




音色の恒常・不変的認知

- 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



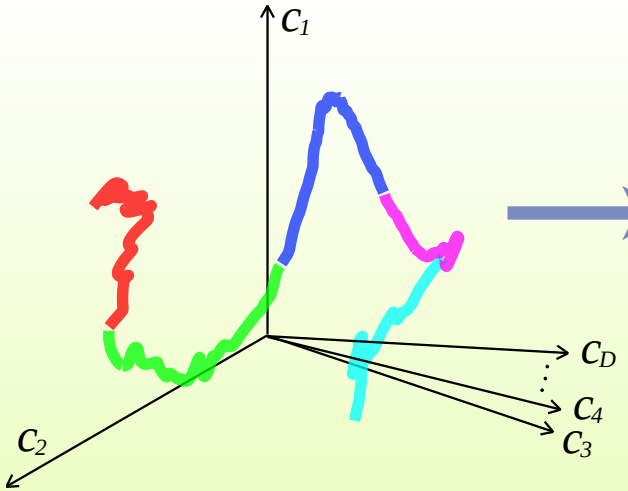


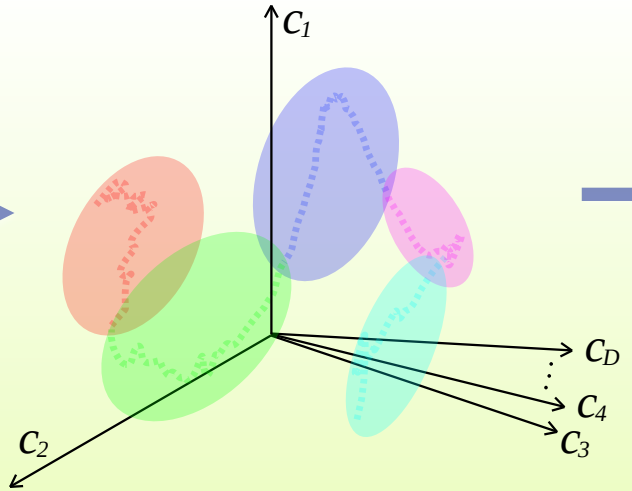
本発表の流れ

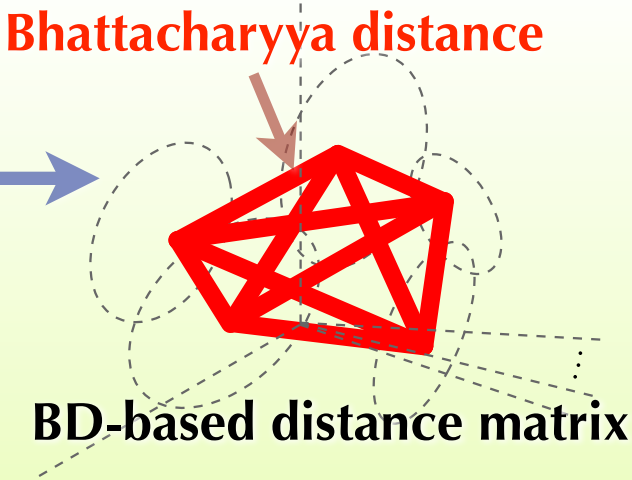
- **刺激の物理的多様性とその認知的不変性**
  - 見え／色み／音高の多様性と自然・進化が編み出した解決方法
- **音声の物理的多様性とその認知的不変性**
  - 音色の多様性と工学者が編み出した解決方法
- **音声の構造的表象とそれに関する様々な考察**
  - 常識を覆すことで，違和感の解消を試みてみる。
- **音声の構造的表象と数学的表現と技術的実装**
  - 体格・性別に不変な音声波形・スペクトルの表現とは？
- **音声の構造的表象を用いた音声アプリケーション**
  - 音声認識，音声合成，発音分析，etc
- **音声の構造的表象の言語学的妥当性**
  - 何故，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

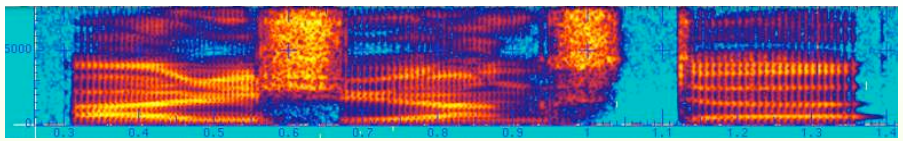
音声の構造的表象の工学的・実験的検証

f-div. (BD)に基づく一発声の構造化









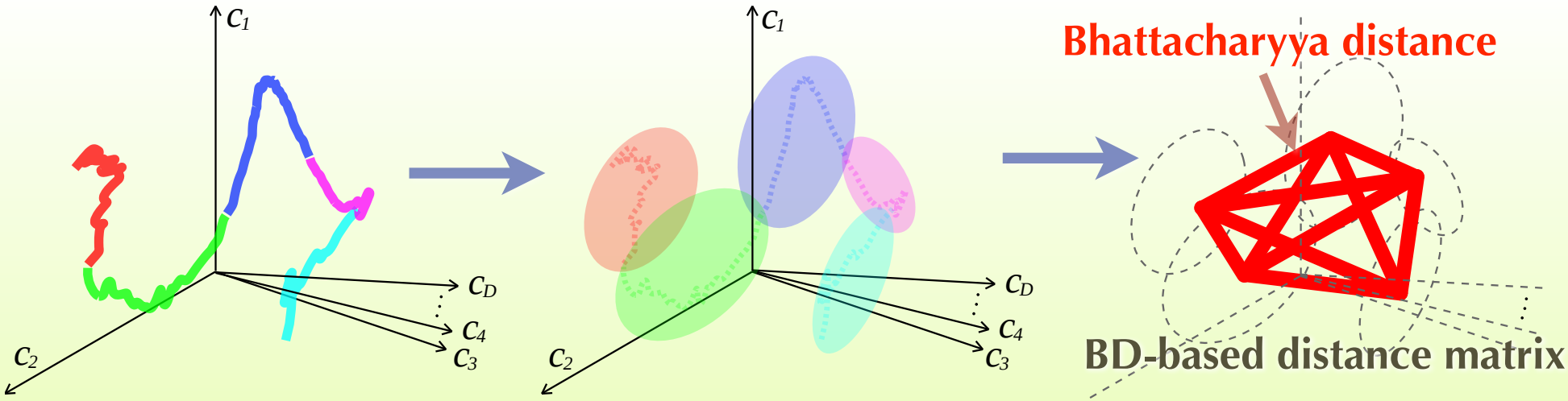
spectrogram (spectrum slice sequence)

cepstrum vector sequence

distribution sequence

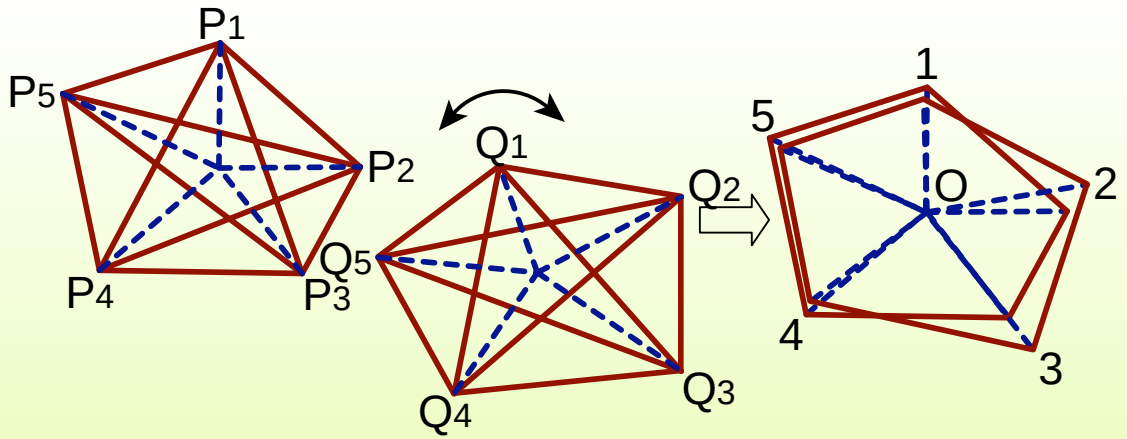
音声の構造的表象の工学的・実験的検証

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

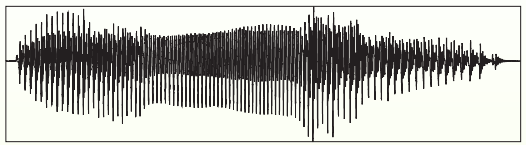
- ２距離行列間のユークリッド距離



- 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



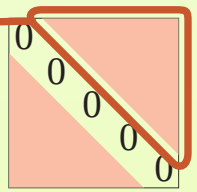
Cepstrum vector sequence

Cepstrum distribution sequence (HMM)

Distances of distributions

Structure (distance matrix)

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



$=$

Statistical structure model

Word 1

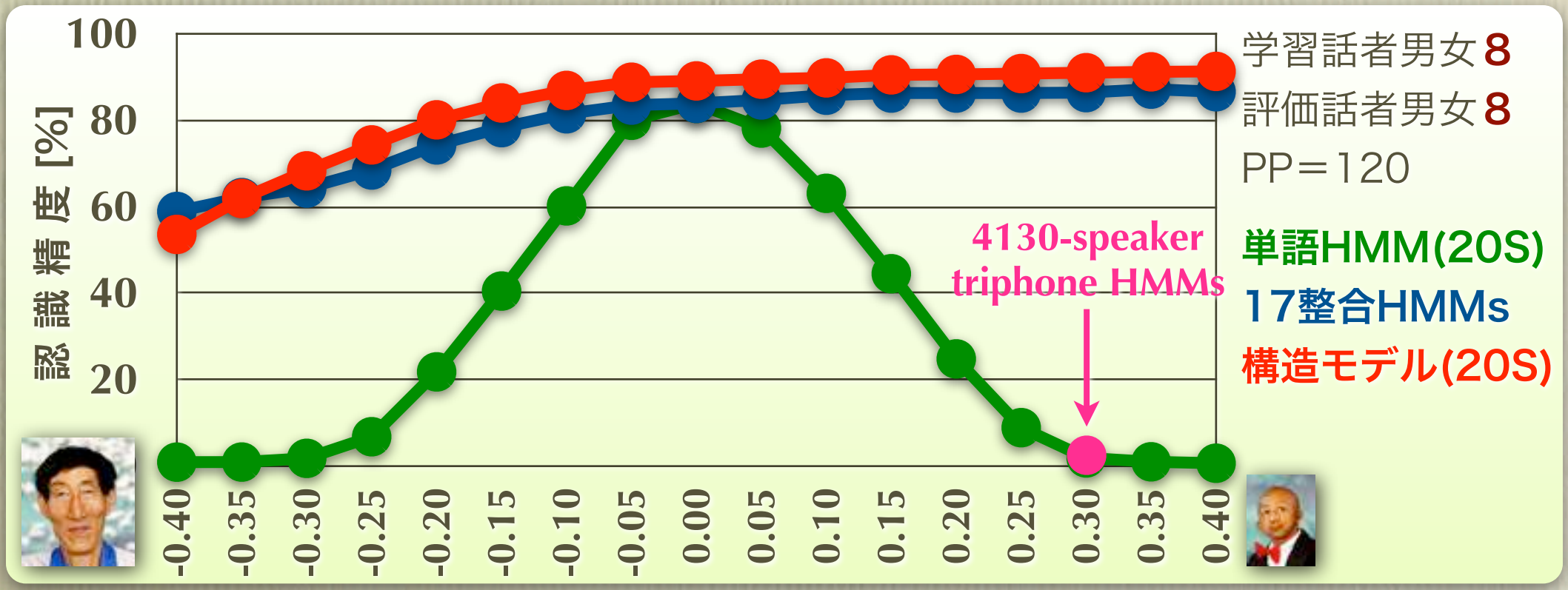
Word 2

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

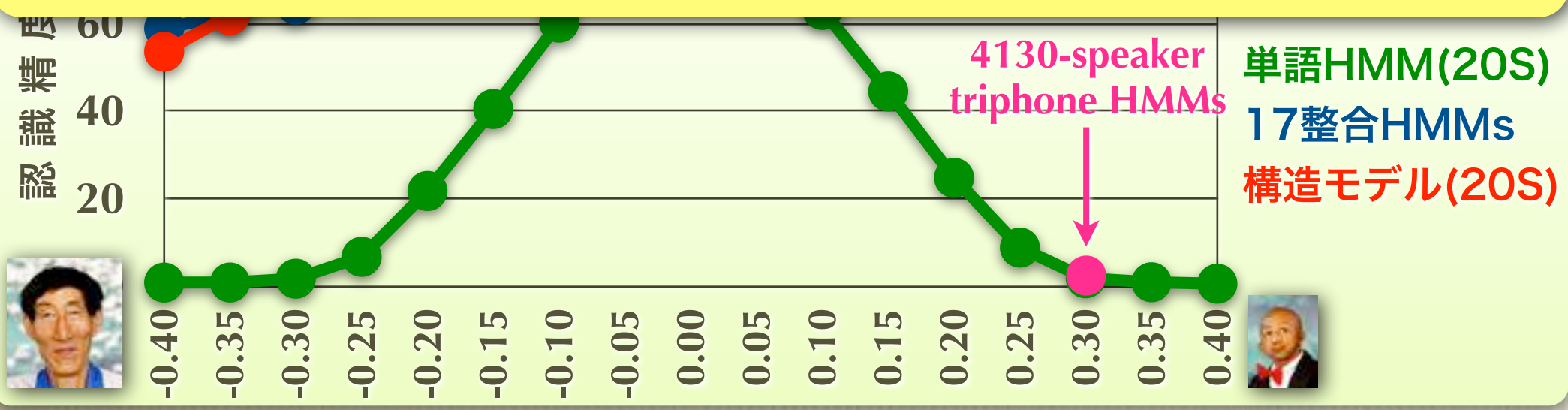


音声の構造的表象の工学的・実験的検証

孤立単語音声の認識実験

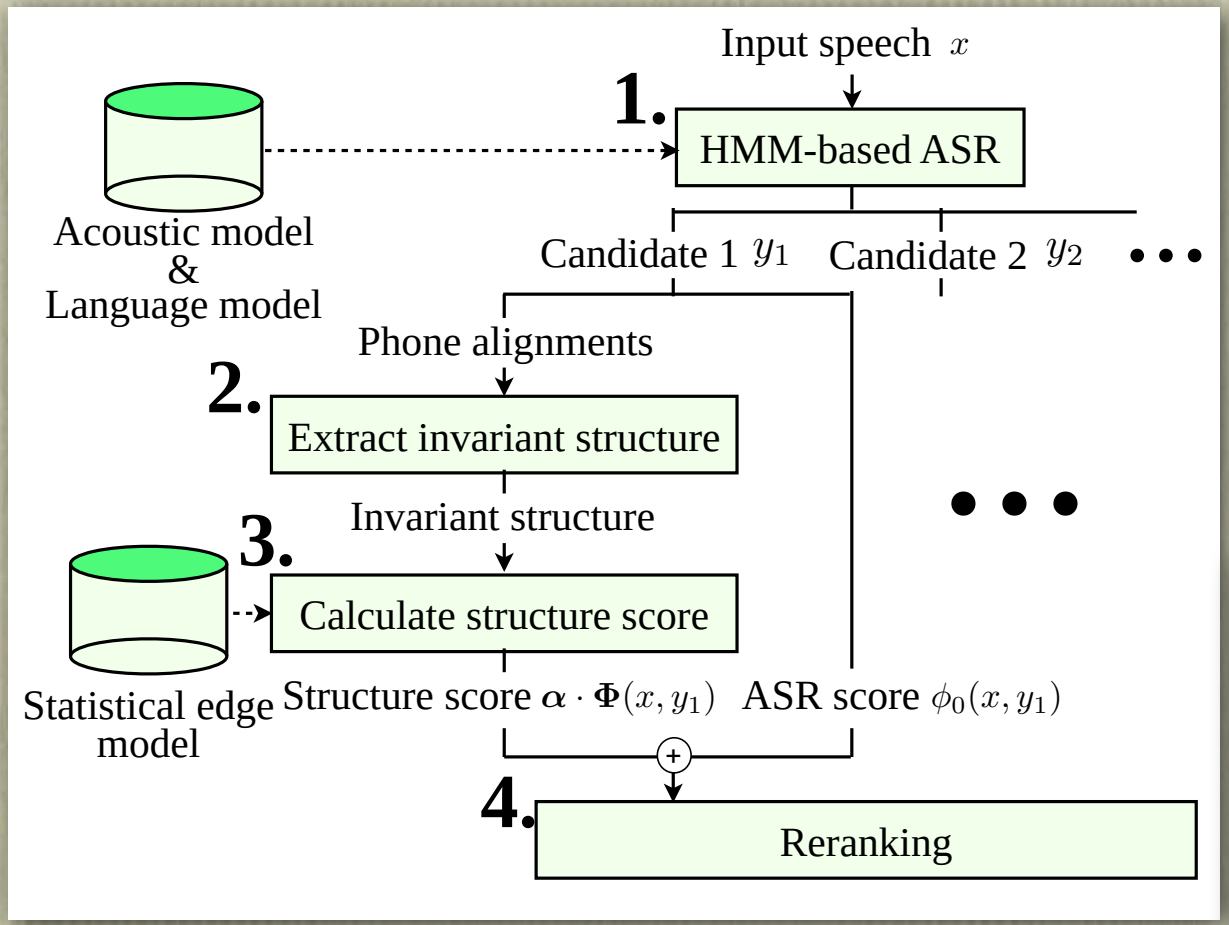
- 二つの問題とその解決
  - 強すぎる不変性→マルチストリーム構造化による都合のよい不変性へ
  - 高すぎる次元数→線形判別分析（LDA）による次元数削減
- 孤立単語認識実験による提案手法の評価

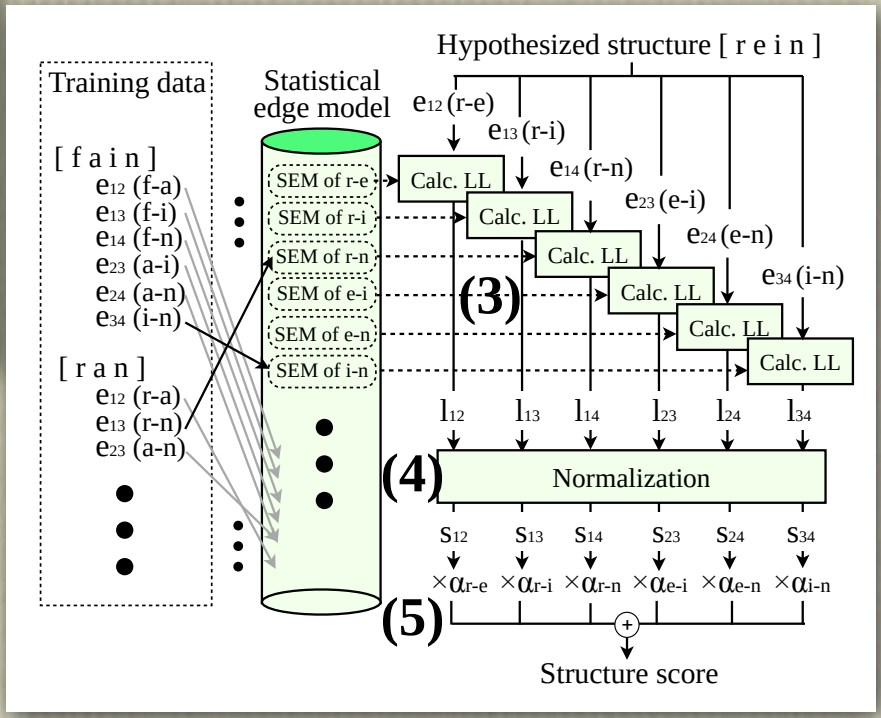
孤立提示された音を音韻同定する能力は音声言語運用には不要なのかもしれない



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





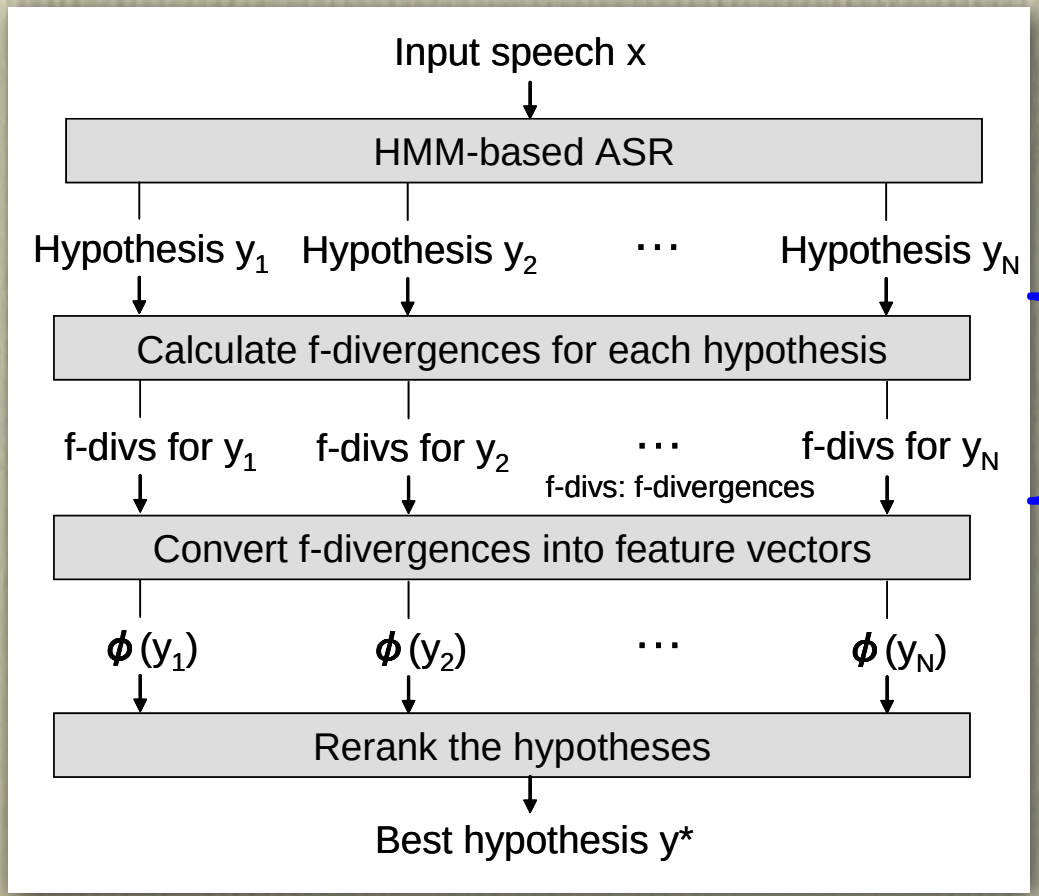
Application of structures to ASR

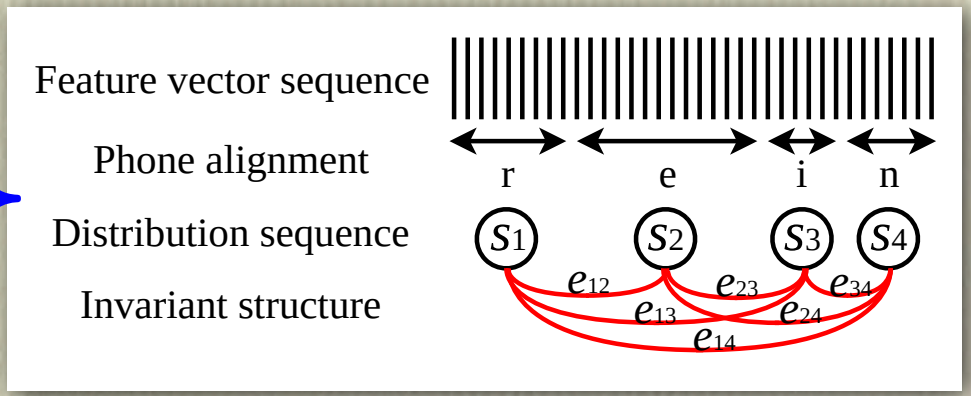
Application to more realistic ASR tasks [Suzuki+'15]

- Digits recognition and LVCSR (dictation)

Use of structural features in discriminative reranking

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

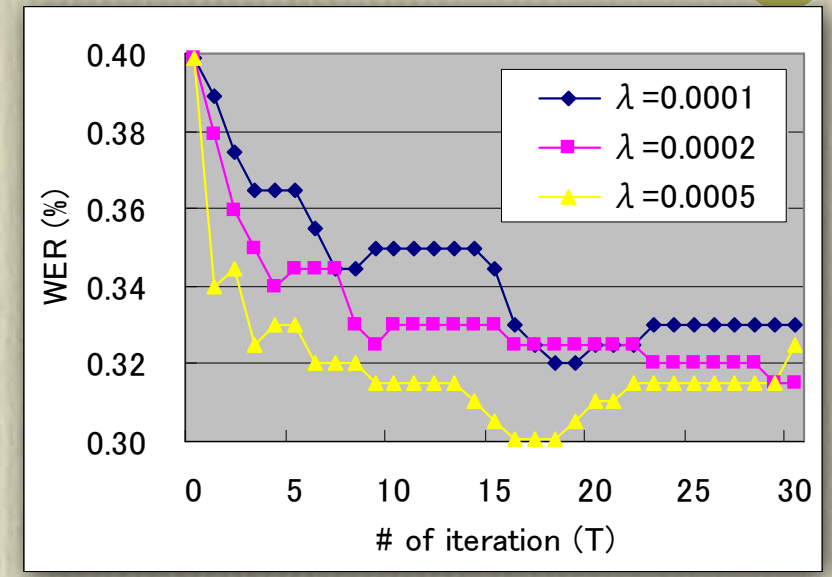




Application of structures to ASR

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.

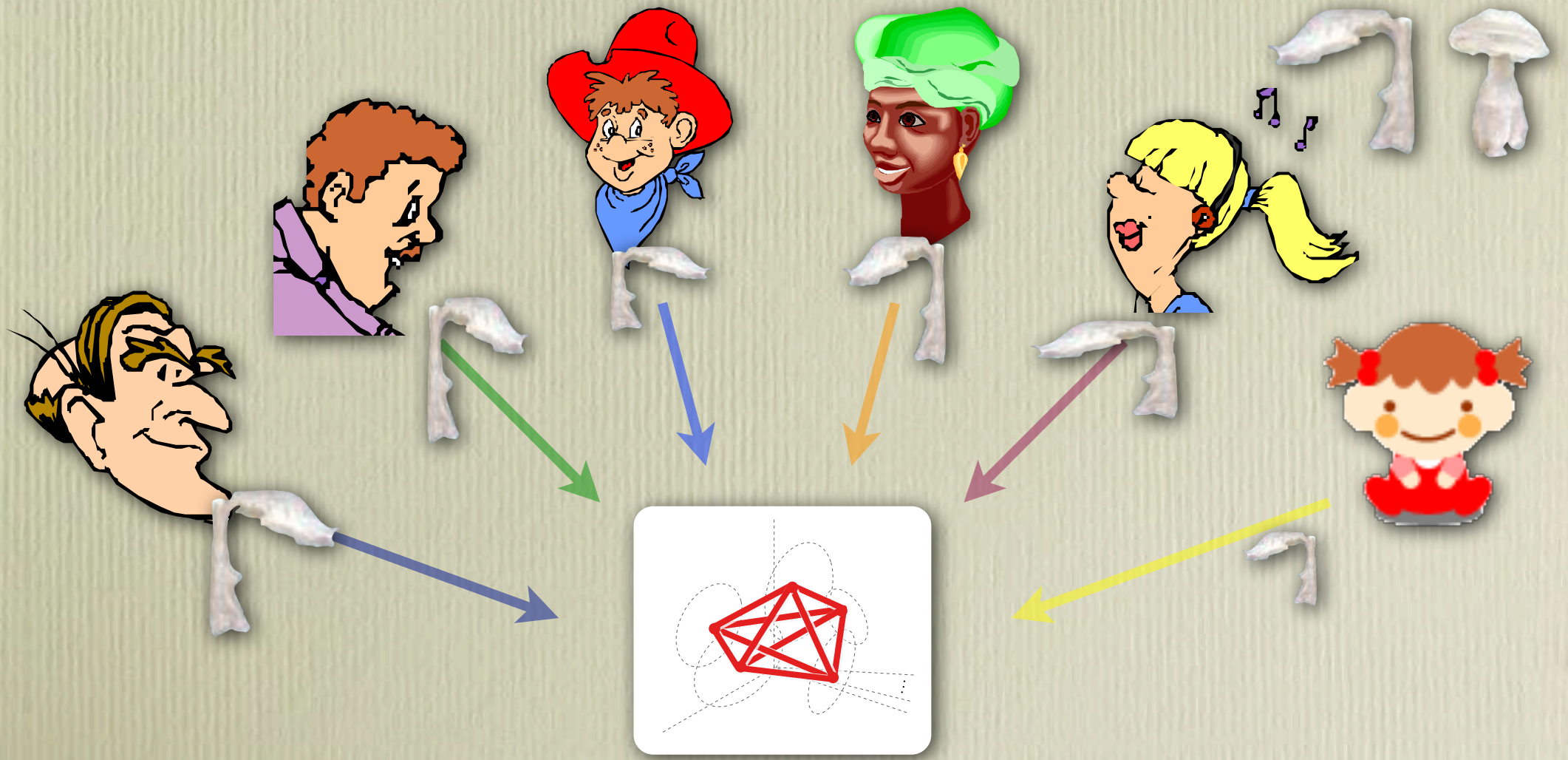
Table 6: CERs of the LVCSR experiment.

Baseline	Proposed	Relative improvement
2.67%	2.53%	5.24%

構造表象からの音声生成

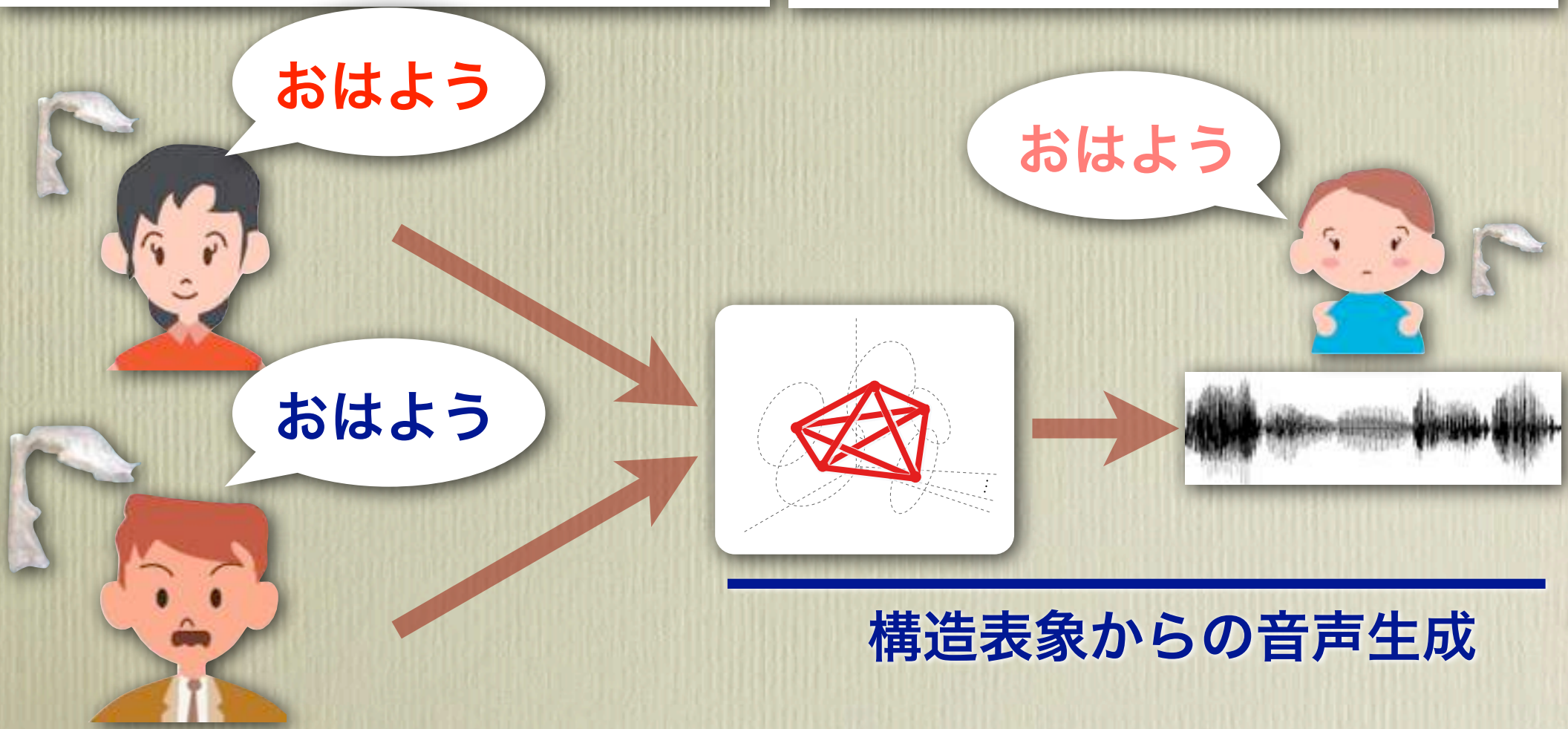
構造＝音声から話者の身体情報を取り除いた抽象表象

- 声道の長さ・弱母音発声時の声道形状
- これが個人によって異なるから，声を聞けば話者が同定できる



構造表象からの音声生成

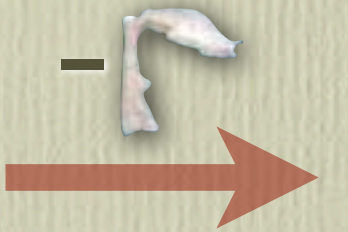
幼児の音声模倣行為の情報論的解釈

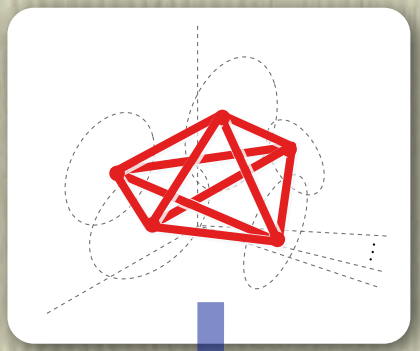


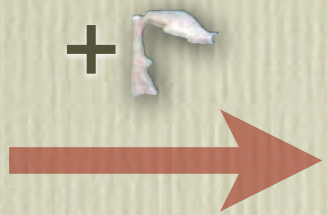
構造表象からの音声生成

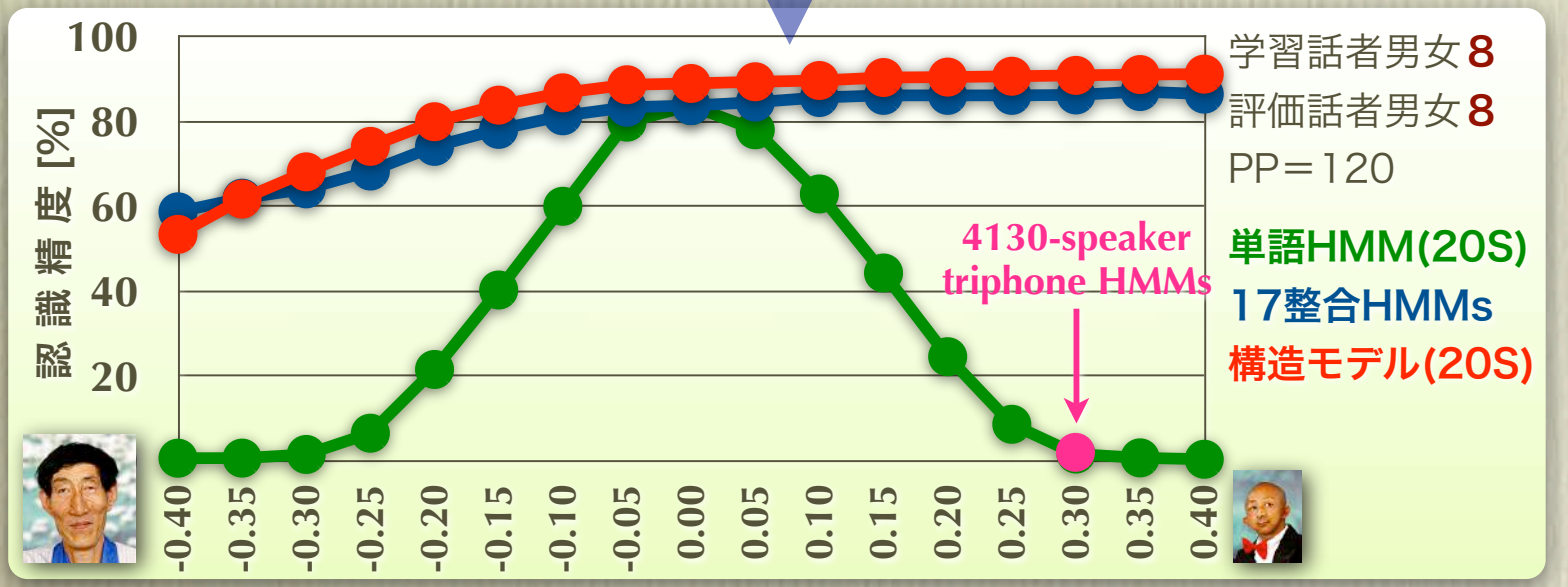
嫁さんと娘の音声を使ったデモンストレーション

- 真似るべき対象の構造：千恵（母親）提供
- 音の実体の初期条件　：礼佳（３歳時）提供









構造表象からの音声生成

- 身体楽器を戻す＝初期条件を与える
- その楽器で発声＝構造制約条件を満たす音色運動の生成