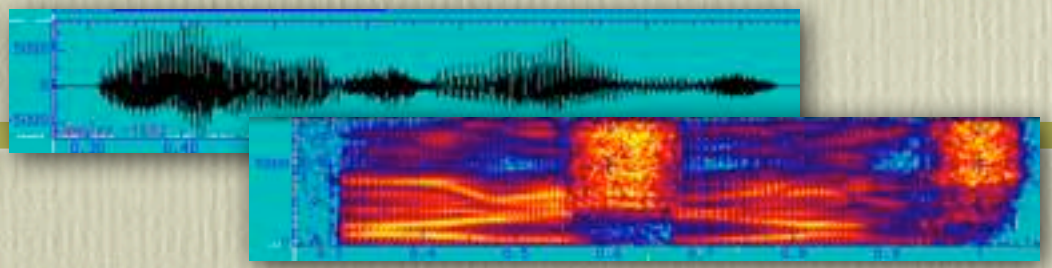
人文社会系研究科基礎文化研究専攻言語学専門分野

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

教科書のコラム記事より

コーヒーブレイク

スペクトル包絡に安住していてよいのだろうか？

音声認識にしろ，音声合成にしろ，スペクトル包絡が基本的な音声特徴量として用いられている。人間の聴覚は音声の位相成分に鈍感との知見より，位相スペクトルを削除して振幅スペクトルを抽出し，さらに，音素情報は音声の音高成分とは独立であるため，ピッチハーモニクスを削除してスペクトル包絡特性を抽出している。しかし，包絡特性には音素情報と話者情報とが同居している。音声認識は音声中の音素情報（テキスト情報）を抽出することが目的だが，スペクトル包絡から話者情報を削除した特徴量を定義することはできないのだろうか。

不特性話者音響モデルは話者独立モデルとも呼ばれるが，この独立性は話者性を削除して得られるのではなく，多数の話者からデータを集めることで，話者性を分布の中に隠すことで得られるモデルである。位相やピッチは物理的に削除して独立性を担保するが，話者性は隠すことで独立性を担保している。

幼児の言語獲得は他者の音声活動を模倣することが必要となる（音声模倣）。この場合，話者性までを模倣しようとはしない。声帯模写はしない。話者の違いを超えた模倣をしている。音声模倣をする動物は小鳥，クジラ，イルカなどがあるが，彼らは基本的に音響的な模倣をする（なお，動物は相対音感を持っていないため，移調前後のメロディの同一性の認識が難しい。異なる音は異なるものと

教科書のコラム記事より

して扱っているのだろう)。動物とは異なり，人が他者の発声をコピーして言葉を獲得するとき，話者性には鈍感なのである。しかし技術はそうなっていない。ある話者の音声サンプルを用いて音声合成装置をつくれば，当然その人の声を出力する装置ができる。機能的には声帯模写装置と呼ぶべきである。

近年，音声工学の分野でもコンテスト形式の研究発表会が多くなってきた。音声合成の場合，blizzard challenge と呼ばれるワークショップが毎年行われている。ここでは合成音の品質を評価する際に，言葉としての自然さ以外に，学習話者の個人性が適切に再現されているか否かも評価対象となる。やはり今の音声合成技術は，声帯模写技術と呼んだほうが適切のように思われる。

言葉を真似る模倣行為が声帯模写的になってしまう場合がある。発達障害の一種，自閉症の中で見られる症状である。このような場合，音声言語の獲得はしばしば難しくなる。音の獲得 $\neq$ 言語の獲得なのだろう。人間と動物に見られる音情報処理の差異，典型的な言語発達が容易/困難な場合に見られる音情報処理の差異を概観すると，話者性を削除して音声を表現する技術の構築が待たれる。確率論は，集めてしまえば消したい要因が消せることを保証するが（$P(a) = \sum_b P(a, b)$），これはあまりにもナイーブすぎないか。話者性をそぎ落とす一手法として**音声の構造的表象**が提案されている。興味のある読者は文献35) を参照されたい。

母音＝定常波　～気柱の共鳴現象～

複雑な管になっても原理は同じ

- 定常波の共振周波数を求めて

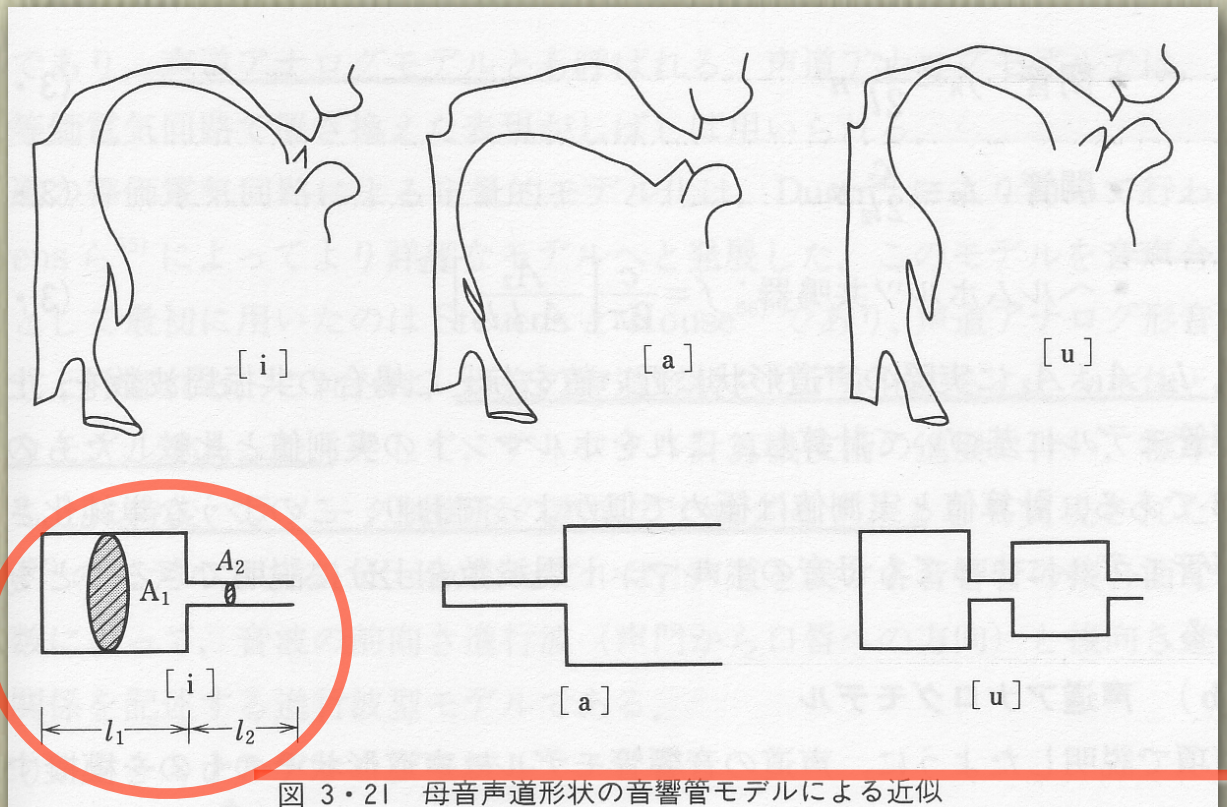


図 3・21　母音声道形状の音響管モデルによる近似

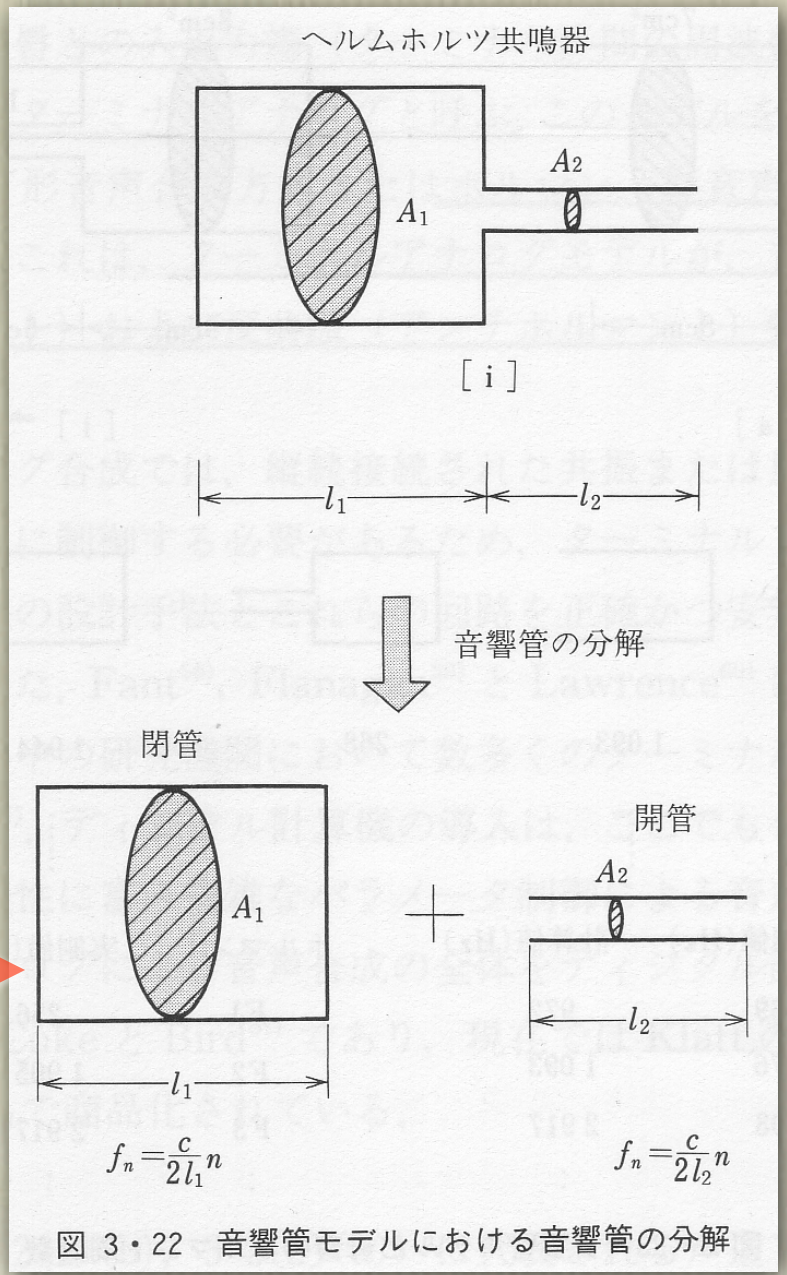


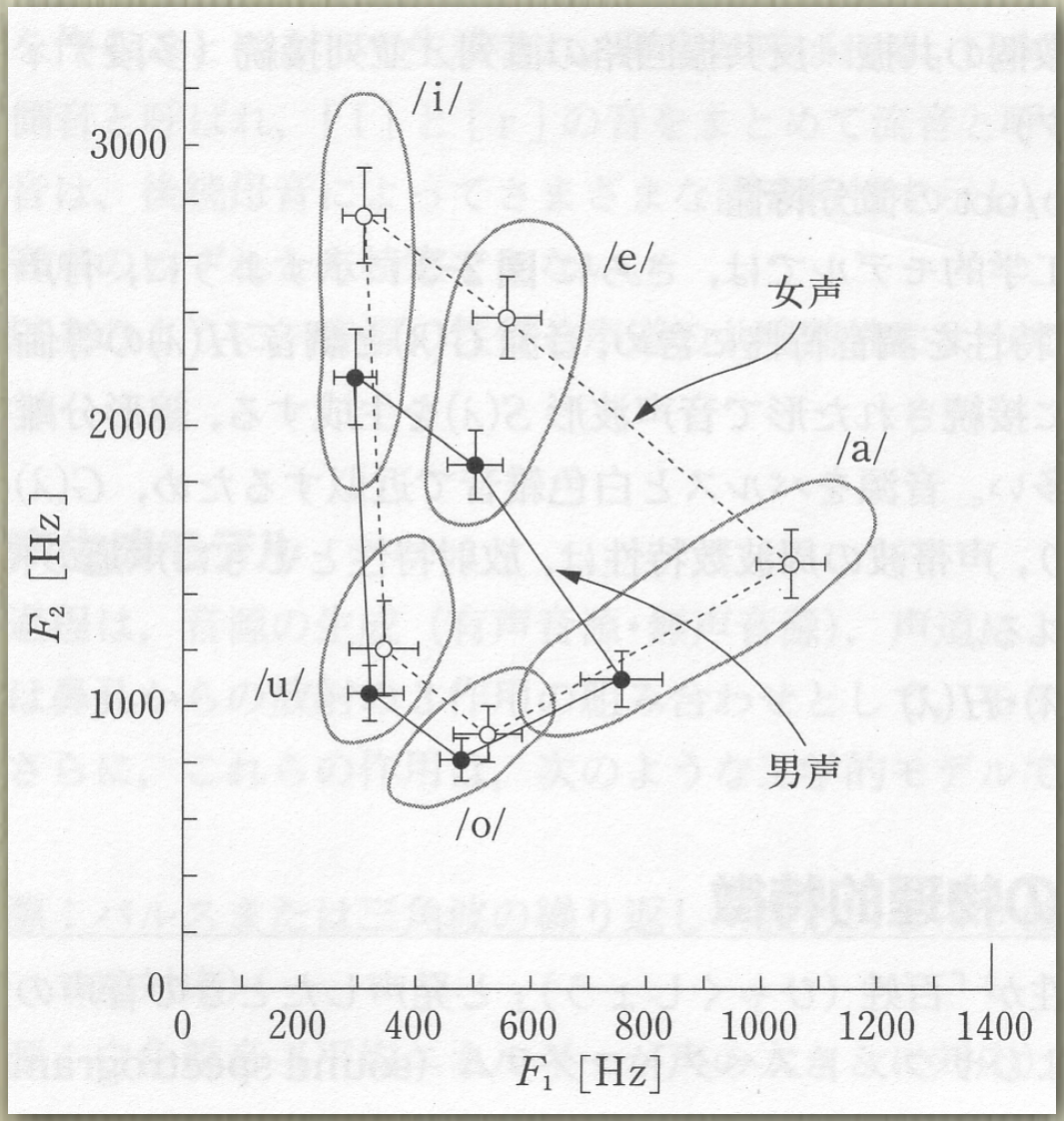
図 3・22　音響管モデルにおける音響管の分解

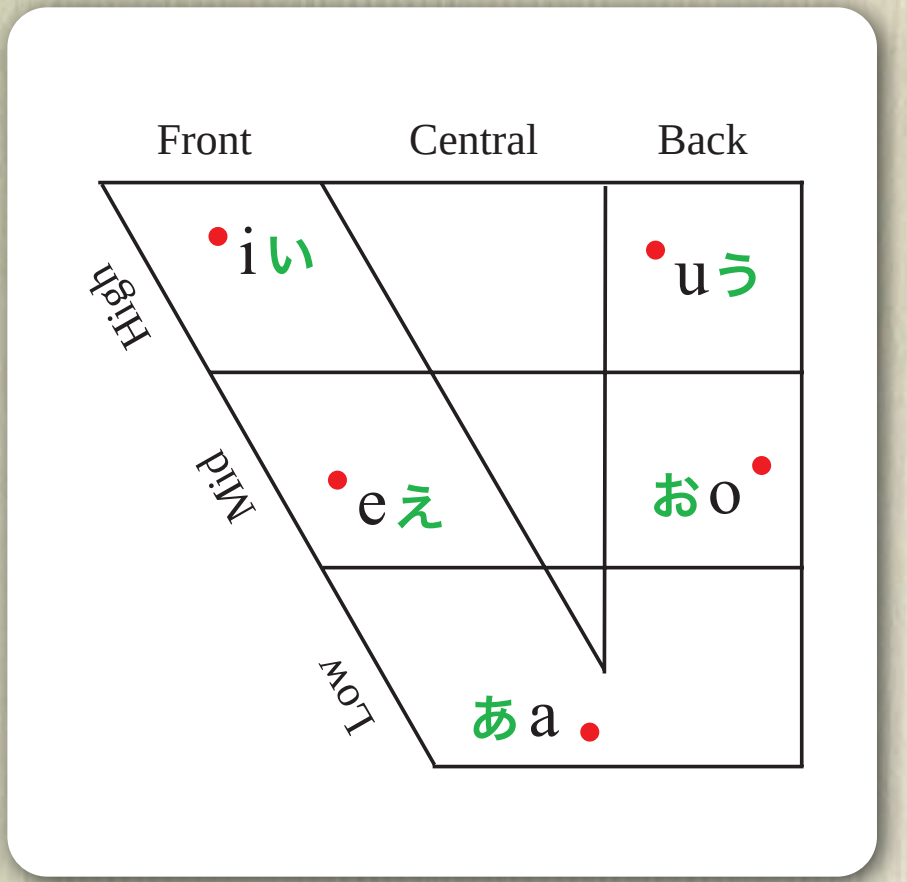
$$f_n = \frac{c}{2l_1}n \qquad f_n = \frac{c}{2l_2}n \qquad f = \frac{c}{2\pi}\left[\frac{A_2}{A_1 l_1 l_2}\right]^{1/2}$$

話者間における母音の差異

形の違い＝長さの違い＝共振周波数の違い

- 「あ」の一部＝「お」の一部





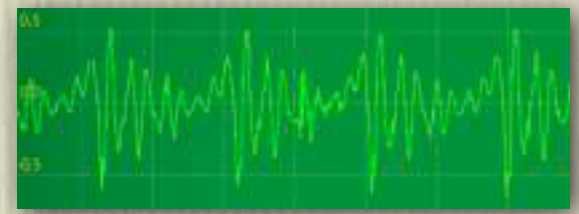
音声が運ぶ様々な情報

言語的情報

- 何を話したのか？
  - 狭義の言語的情報，語彙，音素
- どのように話したのか？
  - パラ言語的情報，意図，発話スタイル，感情

非言語的情報

- 意図的制御は困難であり，不可避的に付与されてしまう情報
  - 話者性，年齢，性別，体格，健康状態
  - マイク，伝送特性，部屋の音響特性



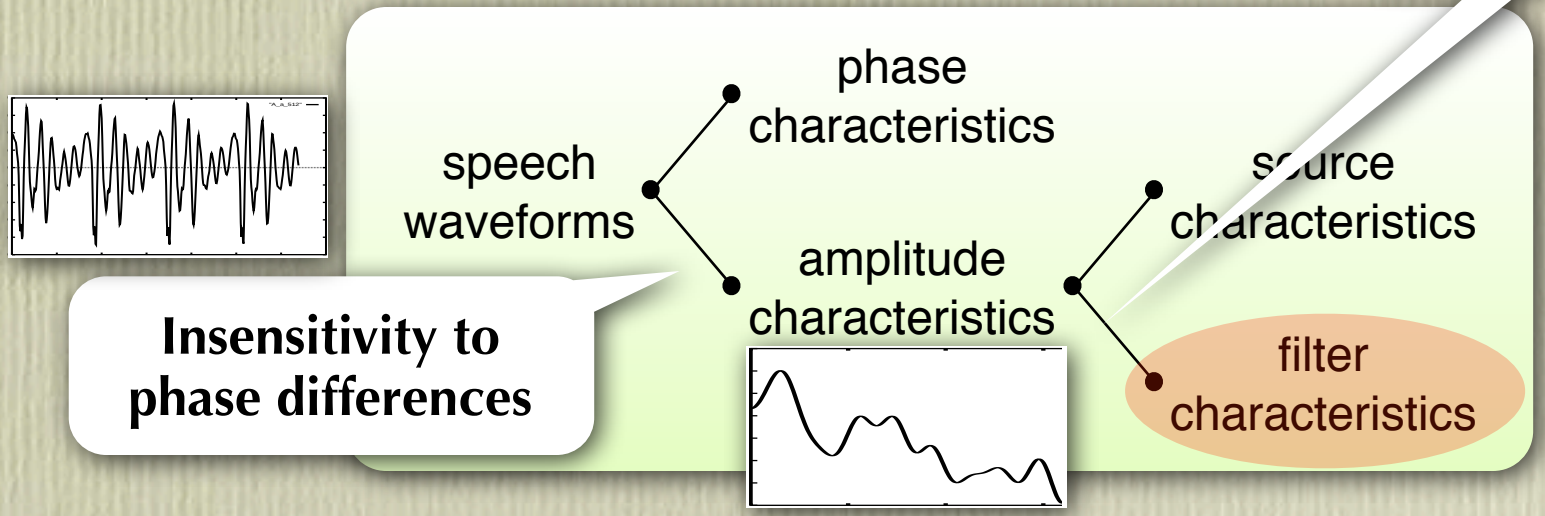
音声信号：一次元の数値列

- 多様な情報を適切に反映しつつ数値列を生成：音声合成
- 数値列から様々な情報を的確に抽出：音声認識・理解
  - 計算機に**音声コミュニケーション**能力を与えるためには？

人間のような

その情報を運ぶ媒体・音響特徴量

二段階の分離に基づく特徴量抽出



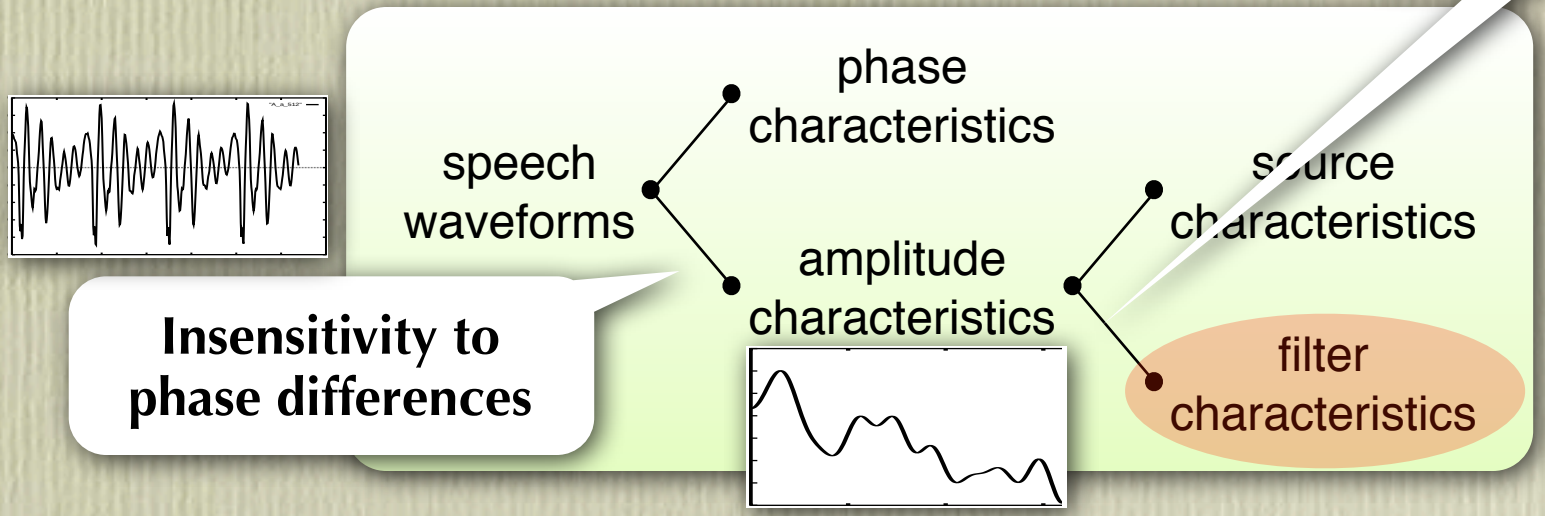
- スペクトル包絡(o)は何を運ぶのか？　　言・パラ言・非言

oの中のある特定の情報のみに着眼したい。

- 当該情報に対応しない特徴量を揃える　　**特徴量正規化**
- 当該情報に対応しないモデルパラメータを調節　　**モデル適応**
- 確率定義に従って着目しない情報を分布に隠す　　**統計的モデル**
- ラベル情報を使って識別的な特徴量へ変換　　**識別的変換**
- 当該情報に直接対応する特徴量を探求する　　**不変量**

その情報を運ぶ媒体・音響特徴量

二段階の分離に基づく特徴量抽出



- スペクトル包絡(o)は何を運ぶのか？　　言・パラ言・非言

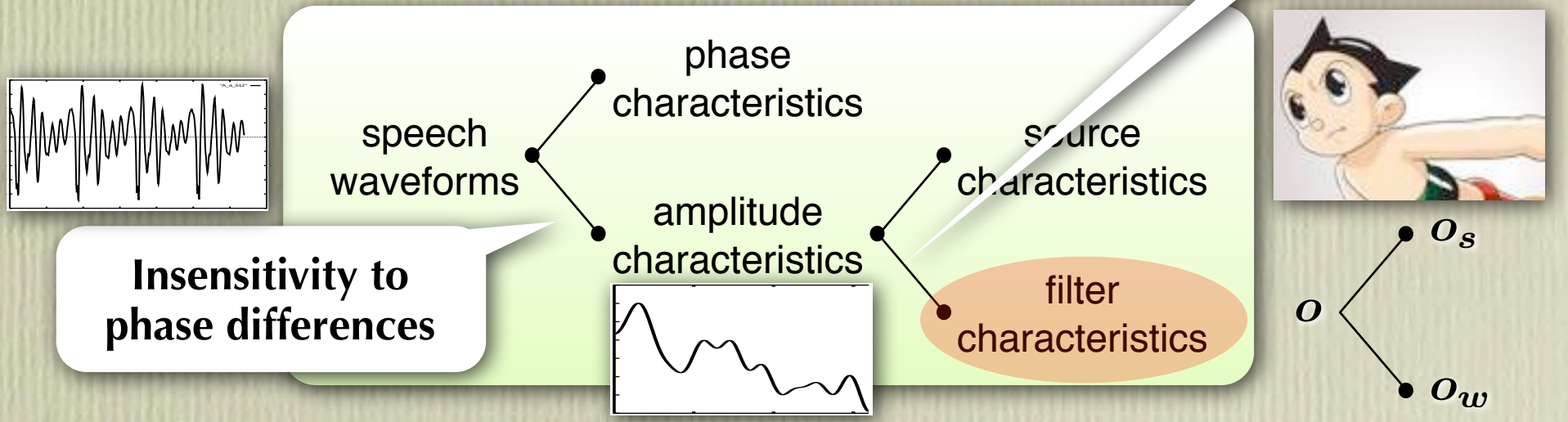
二つの音響モデル $P(o|w)$ と $P(o|s)$

s = speaker
w = word

- 不特定話者の単語音響モデル
  - $P(o|w) = \sum_s P(o,s|w) = \sum_s P(o|w,s)P(s|w) \sim \sum_s P(o|w,s)P(s)$
- テキスト非依存の話者音響モデル
  - $P(o|s) = \sum_w P(o,w|s) = \sum_w P(o|w,s)P(w|s) \sim \sum_w P(o|w,s)P(w)$
  - 集めてしまえば「確率の定義」が見たくないものを隠してくれる。

その情報を運ぶ媒体・音響特徴量

二段階の分離に基づく特徴量抽出



- スペクトル包絡(o)は何を運ぶのか？　　言・パラ言・非言

真の音声の統計的モデル　～波形の統計的モデル～

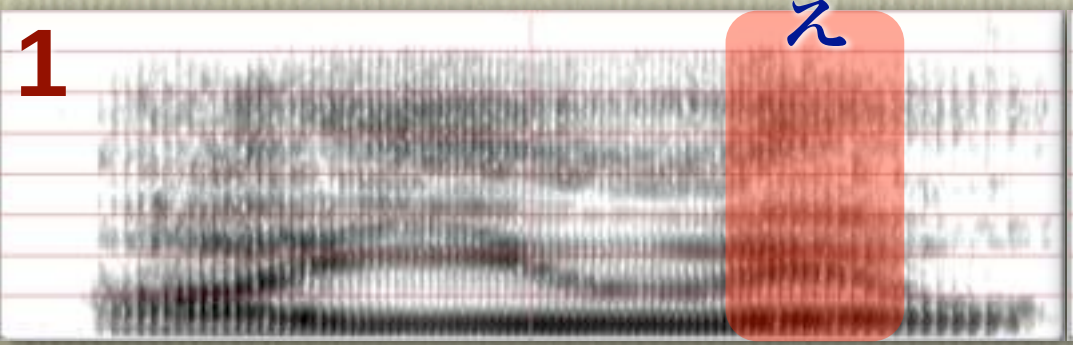
- 不特定話者・不特定基本周波数・不特定位相の音響モデル
  - 見たくないものは全て「確率の定義」で集めて隠してしまおう。
  - $P(o|w) \approx \sum_{s,h,p} P(o|w,s,h,p)P(s)P(h)P(p)$
    - s：speaker, h：harmonics, p：phase
  - 一般的な解決策：各手法の組み合わせ
    - 最終的に性能を最大化する組み合わせを追求する。

音声物理の多様性と音声知覚の不変性

身長（喉の長さ）の違いと声の違い

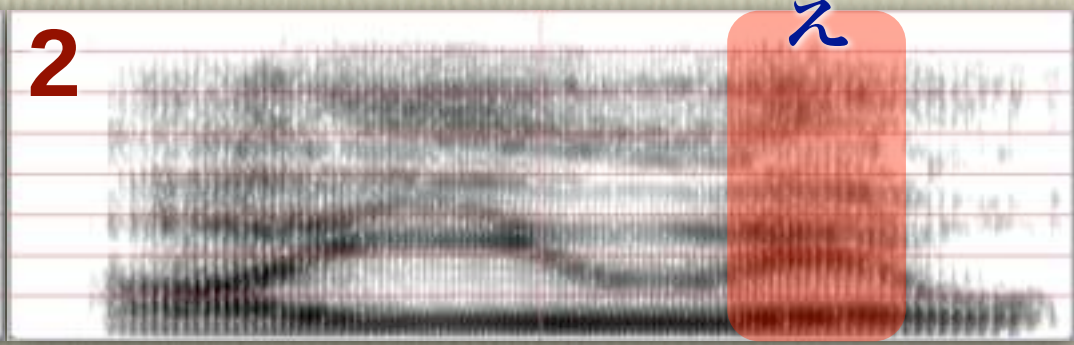
- 分析合成と呼ばれる技術（STRAIGHT）を用いて音声を変形
- いろんな身長の男声を生成／但しオリジナルは167cm
- どう聞こえますか？

1



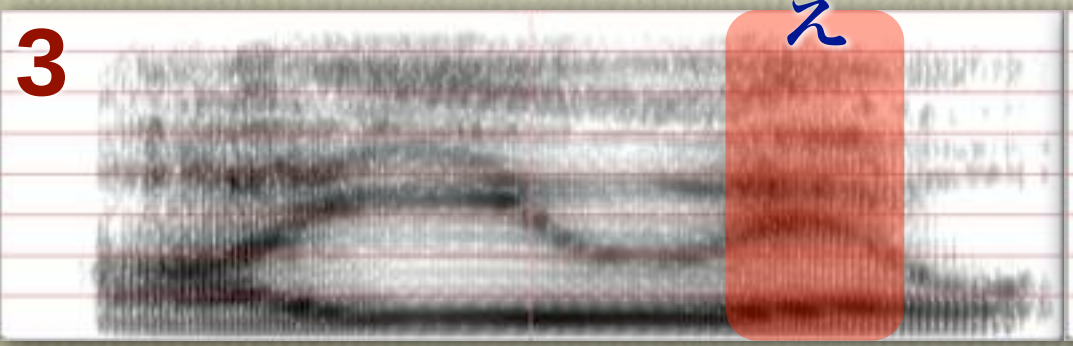
200 cm = 6.56 feet

2



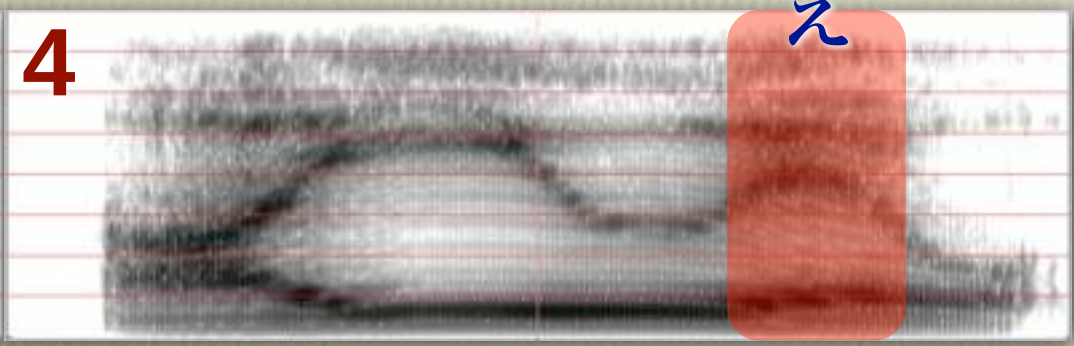
167 cm = 5.48 feet

3



117 cm = 3.84 feet

4

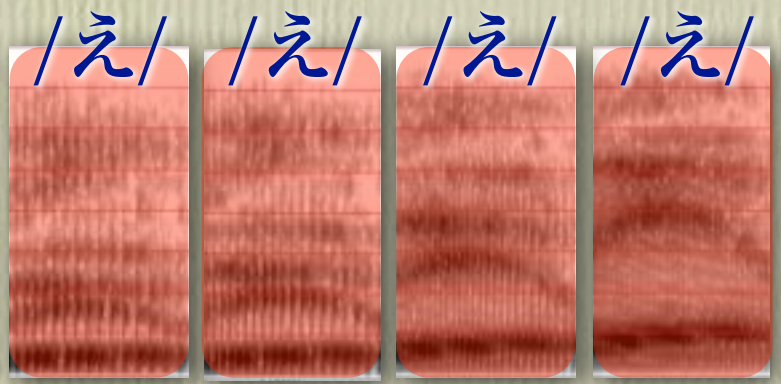


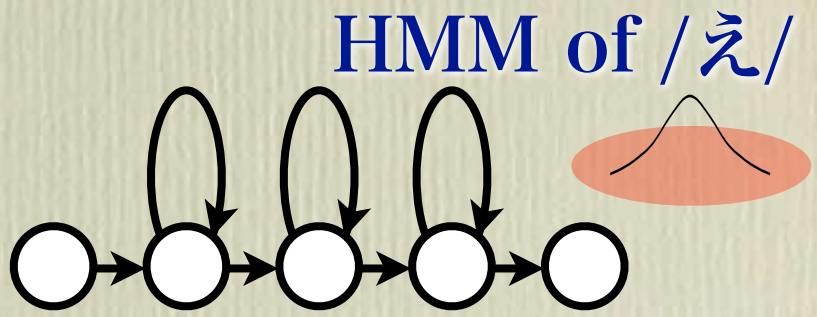
83 cm = 2.72 feet

音声工学（科学）のアプローチ

音声認識 ＝ 音声→テキスト変換

- しかし，個々の音韻の音的実体は様々な音となる
  - 性別，年齢，マイク，部屋，伝送系などなど
  - IBM の偉業：35万人の音声を収録





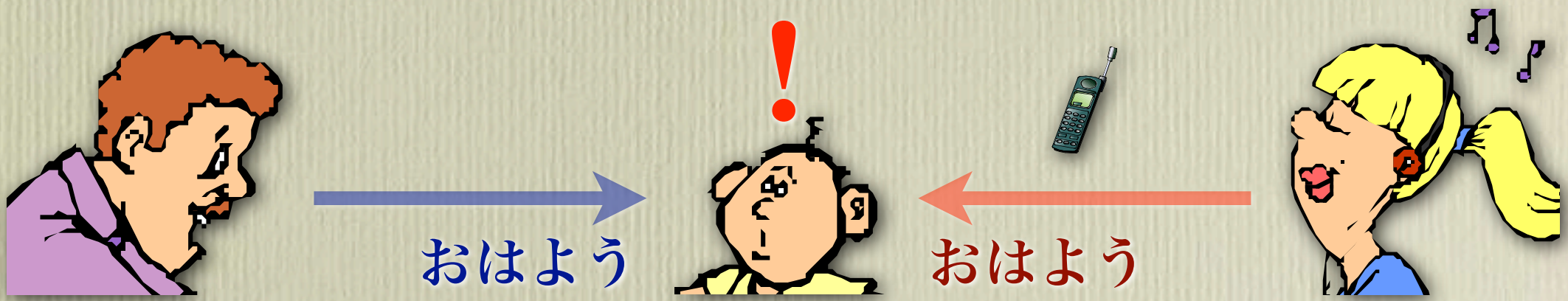
$$P(o|w) = \sum_s P(o,s|w) = \sum_s P(o|w,s)P(s|w) \sim \sum_s P(o|w,s)P(s)$$

この両者の一体何が「同一」なのか？

音を集めることは本当に必要なのか？

幼児の音声音響的環境　～Another PoS～

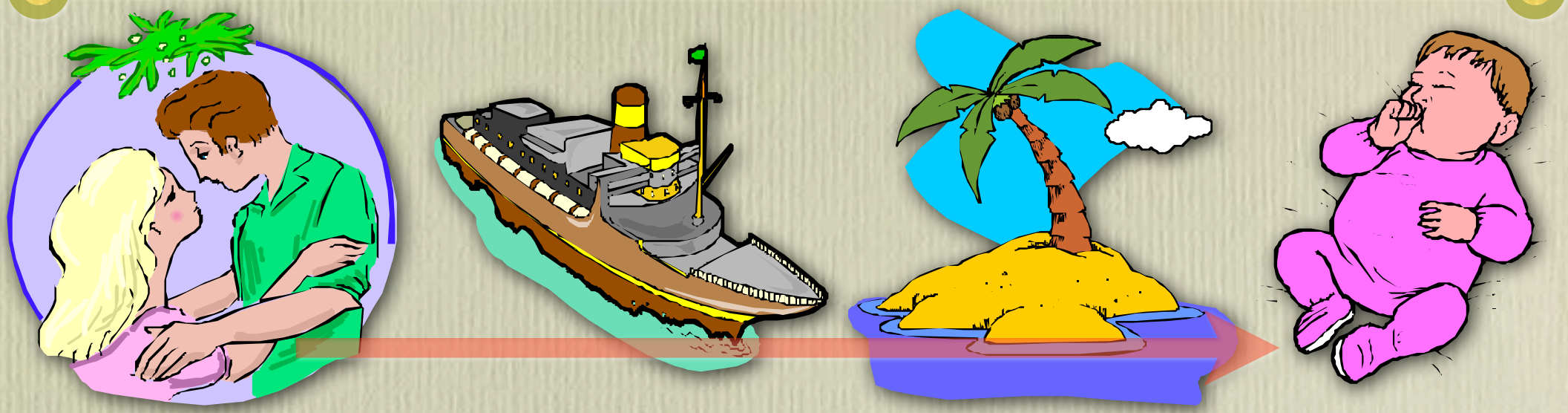
- 大部分は母親と父親の音声（日本の場合は母親ばかり？）
- 話し出せば，人の聞く声は半分は自分の声
- 極端に偏った音声コーパスに基づく超頑健な音声情報処理

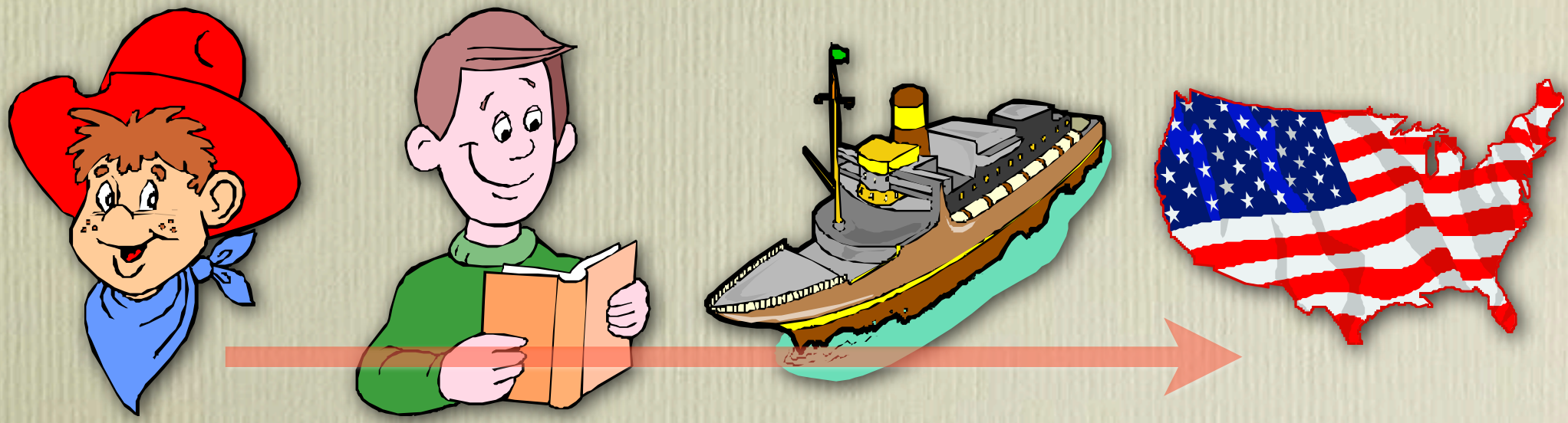


音声物理の多様性と音声知覚の不変性

- 集めずに知覚できる同一性とは一体何なのか？

意地悪な思考実験





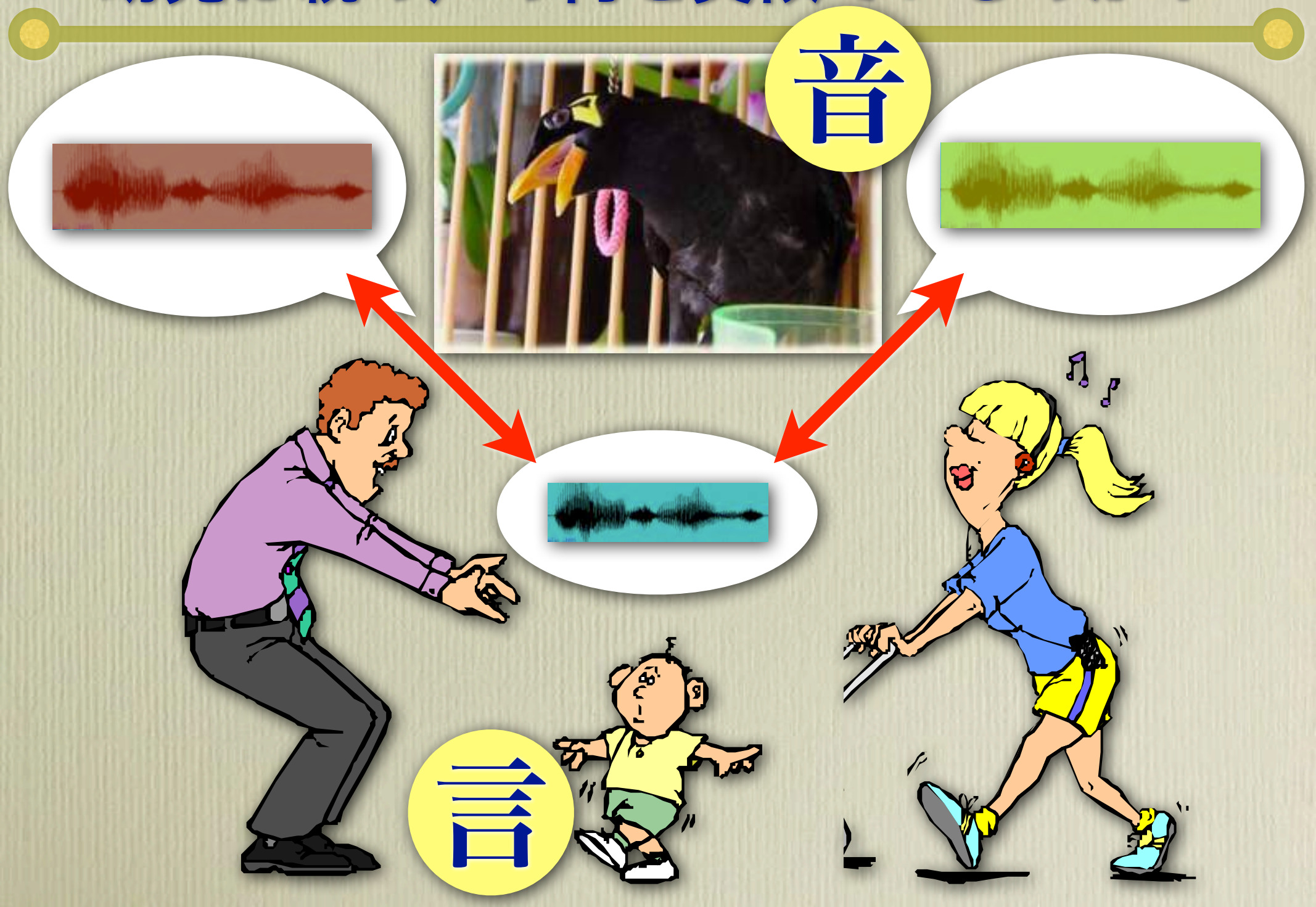
子育て奮闘中のお母様方

20／20

音声認識研究者

0／N

幼児は親の声の何を真似ているのか？



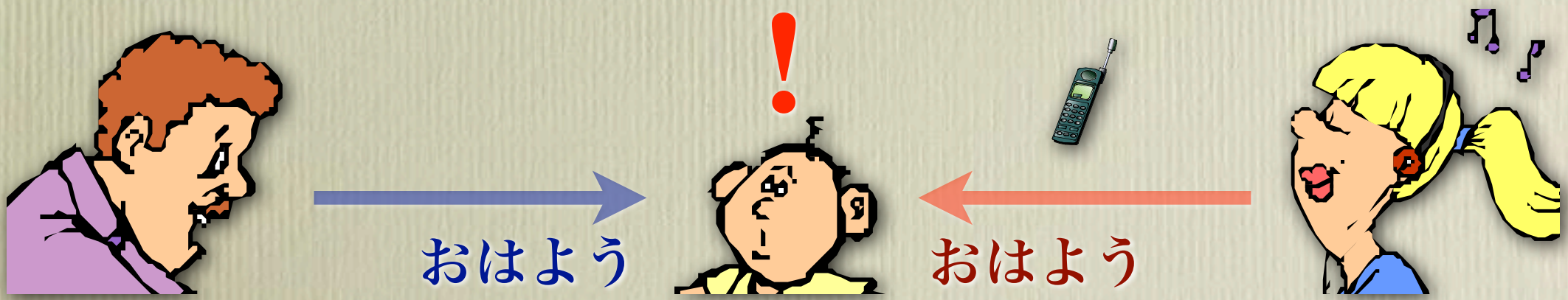
要素から？全体から？

発達心理学の主張　～全体から入る音声処理～

- 音シンボル（音韻）の意識の定着は小学校入学以降
- 単語音声を要素に分割することが困難でも，音声活動を始める
- 「子供は語全体の音形・ゲシュタルトを獲得してから，語を構成する個々の分節音の獲得へと進む（加藤'03, 早川'06）」
  - 語（発話）全体の音形を音にしたものが音声

幼児の音声音響的環境　～Another PoS～

- 大部分は母親と父親の音声／人の聞く声は半分は自分の声
- 極端に偏った音声コーパスに基づく超頑健な音声情報処理



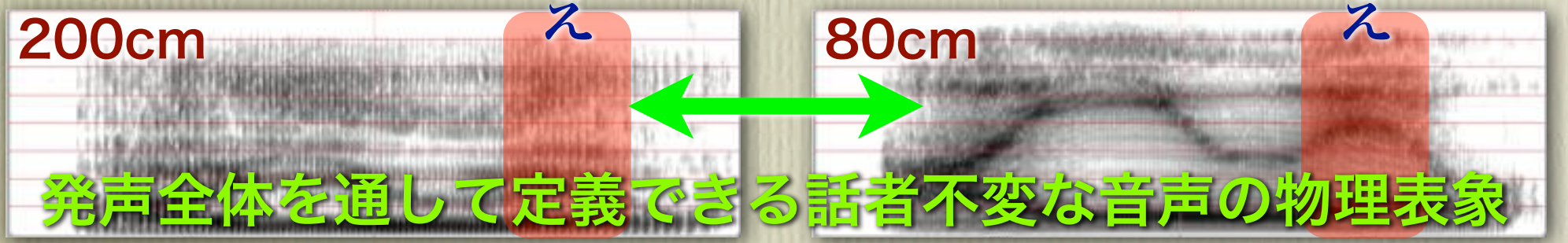
？二つの問題？

音声物理の多様性と音声知覚の不変性→another PoS

- [え] は年齢・性別・マイクなどの非言語的要因によって変化
  - 音声認識（IBM ViaVoice）＝ 35万人の話者から [え] を集める
  - 幼児 ＝ 大半は母親／父親，そして，自分の声

九官鳥の音声模倣と幼児の音声模倣

- 九官鳥は声（音）を真似る
  - 良い九官鳥は聞けば飼い主が分かる。「声そのもの」が模倣対象
- 幼児は親の声の何を真似ているのか？
  - 声を真似ようとはしていない。電話声になるようになんかしない。
  - 個々の音をモーラ同定して，それを一つずつ再生する？　それは困難
    - 語全体の音形・語ゲシュタルトをまず獲得。分節音はその後。



ある種の違和感

音声言語工学の構築してきた技術

- 波形素片／スペクトル素片に対するDBの構築
  - 波形接続型音声合成／HMM音声合成（学習用話者の声の合成）
- 数理統計的手法に基づく音響モデルのパラメータ推定
  - 数千～数十万人の音声を用いた音響モデルによる不特定話者音声認識
- 圧倒的な計算速度の向上と大規模クラスターの構成
  - 様々な環境別に構築したシステム群によるパラレルデコーディング

ふと，我が子を見てみる・・・

- 母親の声を一番よく聞いて日本語を獲得したはず・・・
  - でも，母親の声の模倣（声帯模写）なんて一度もしていない。
- この子の聞く声の多くは，母親，自分，そして，父親・・・
  - 昨晩，お婆ちゃんに初めて声を聞かせた。電話で。で，会話してた。
- 音声って，非常にモロい物理現象なんだよな・・・
  - でも，何故か，そのメディアを使うのが一番楽なんだよな・・・

教科書のコラム記事より

コーヒーブレイク

スペクトル包絡に安住していてよいのだろうか？

音声認識にしろ，音声合成にしろ，スペクトル包絡が基本的な音声特徴量として用いられている。人間の聴覚は音声の位相成分に鈍感との知見より，位相スペクトルを削除して振幅スペクトルを抽出し，さらに，音素情報は音声の音高成分とは独立であるため，ピッチハーモニクスを削除してスペクトル包絡特性を抽出している。しかし，包絡特性には音素情報と話者情報とが同居している。音声認識は音声中の音素情報（テキスト情報）を抽出することが目的だが，スペクトル包絡から話者情報を削除した特徴量を定義することはできないのだろうか。

不特性話者音響モデルは話者独立モデルとも呼ばれるが，この独立性は話者性を削除して得られるのではなく，多数の話者からデータを集めることで，話者性を分布の中に隠すことで得られるモデルである。位相やピッチは物理的に削除して独立性を担保するが，話者性は隠すことで独立性を担保している。

幼児の言語獲得は他者の音声活動を模倣することが必要となる（音声模倣）。この場合，話者性までを模倣しようとはしない。声帯模写はしない。話者の違いを超えた模倣をしている。音声模倣をする動物は小鳥，クジラ，イルカなどがあるが，彼らは基本的に音響的な模倣をする（なお，動物は相対音感を持っていないため，移調前後のメロディの同一性の認識が難しい。異なる音は異なるものと

教科書のコラム記事より

して扱っているのだろう)。動物とは異なり，人が他者の発声をコピーして言葉を獲得するとき，話者性には鈍感なのである。しかし技術はそうなっていない。ある話者の音声サンプルを用いて音声合成装置をつくれば，当然その人の声を出力する装置ができる。機能的には声帯模写装置と呼ぶべきである。

近年，音声工学の分野でもコンテスト形式の研究発表会が多くなってきた。音声合成の場合，blizzard challenge と呼ばれるワークショップが毎年行われている。ここでは合成音の品質を評価する際に，言葉としての自然さ以外に，学習話者の個人性が適切に再現されているか否かも評価対象となる。やはり今の音声合成技術は，声帯模写技術と呼んだほうが適切のように思われる。

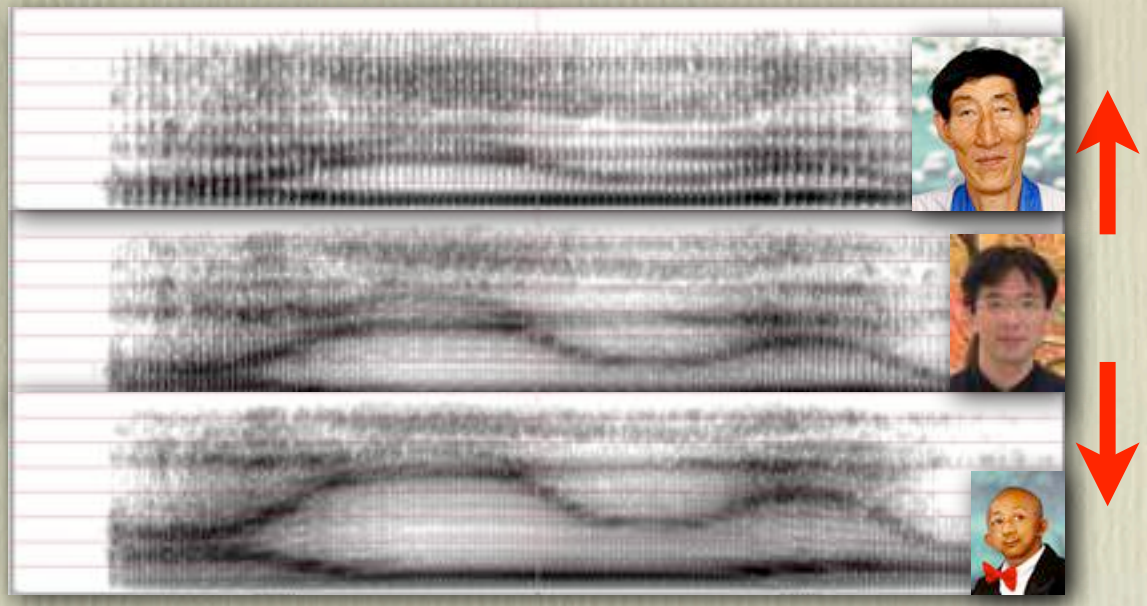
言葉を真似る模倣行為が声帯模写的になってしまう場合がある。発達障害の一種，自閉症の中で見られる症状である。このような場合，音声言語の獲得はしばしば難しくなる。音の獲得 $\neq$ 言語の獲得なのだろう。人間と動物に見られる音情報処理の差異，典型的な言語発達が容易/困難な場合に見られる音情報処理の差異を概観すると，話者性を削除して音声を表現する技術の構築が待たれる。確率論は，集めてしまえば消したい要因が消せることを保証するが（$P(a) = \sum_b P(a, b)$），これはあまりにもナイーブすぎないか。話者性をそぎ落とす一手法として**音声の構造的表象**が提案されている。興味のある読者は文献35) を参照されたい。

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？

年齢・性別・体格による音声の変形

巨人と小人の会話は，何故成立するのか？

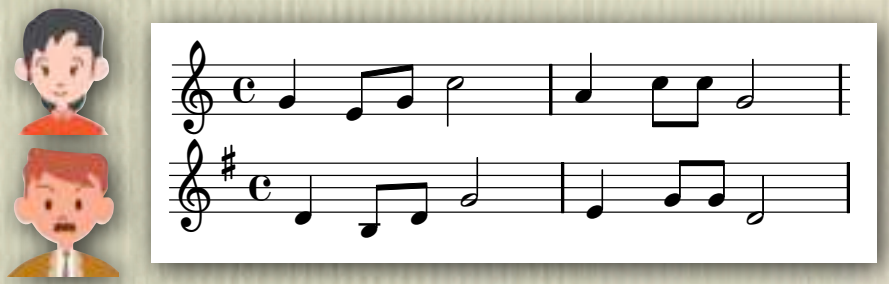


刺激の物理的多様性とその認知的不変性

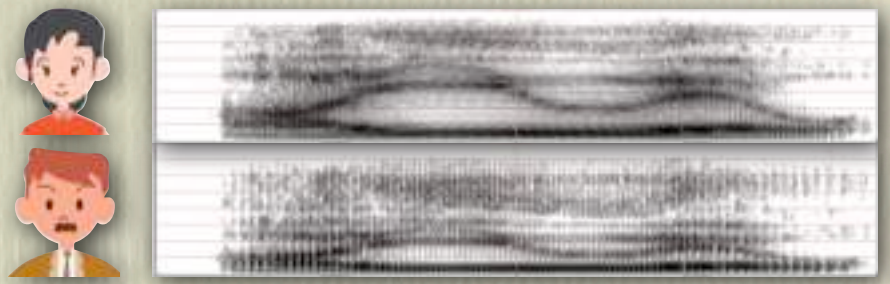
感覚受容器が受け取る情報は容易に変貌する

- 見えの変化
  - 視点を変えて見た犬
  - 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

刺激の物理的多様性とその認知的不変性

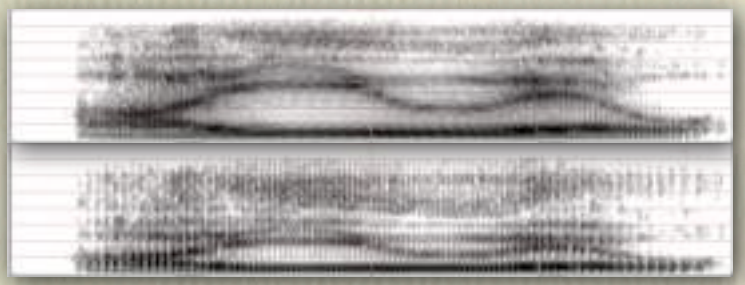
感覚受容器が受け取る情報は容易に変貌する

- 見えの変化
  - 視点を変えて見た犬
  - 対象との距離を変えて見た像

- 色みの変化

静的偏差による刺激変形と
刺激変形に不変な認知様式

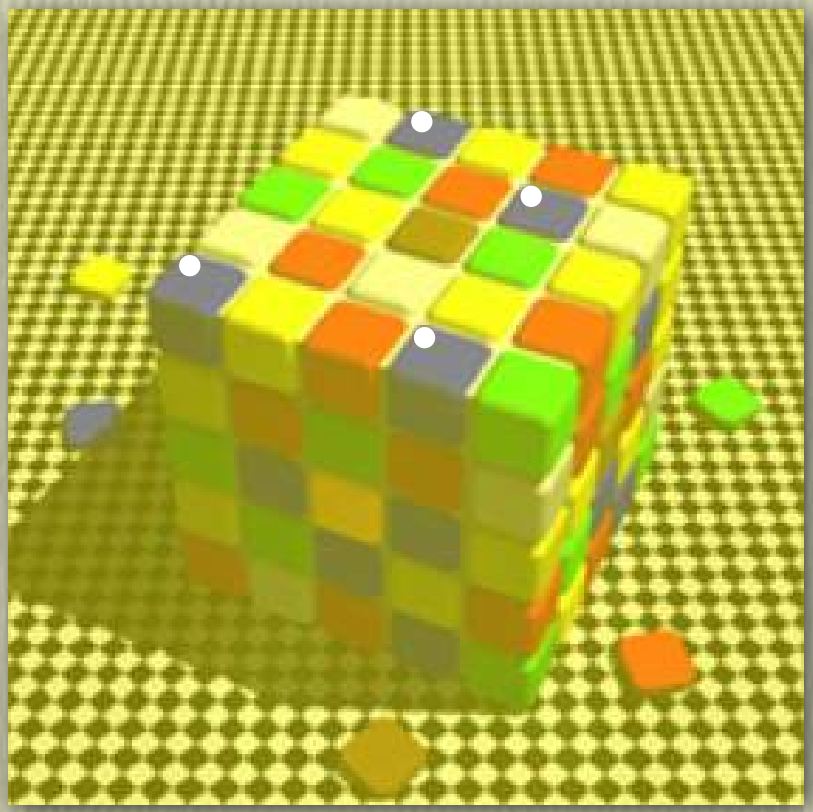
  - カラオケでのキーの上げ下げ
- 音色の変化
  - 男性のおはよう！と女性のおはよう！
  - 大人のおはよう！と子供のおはよう！

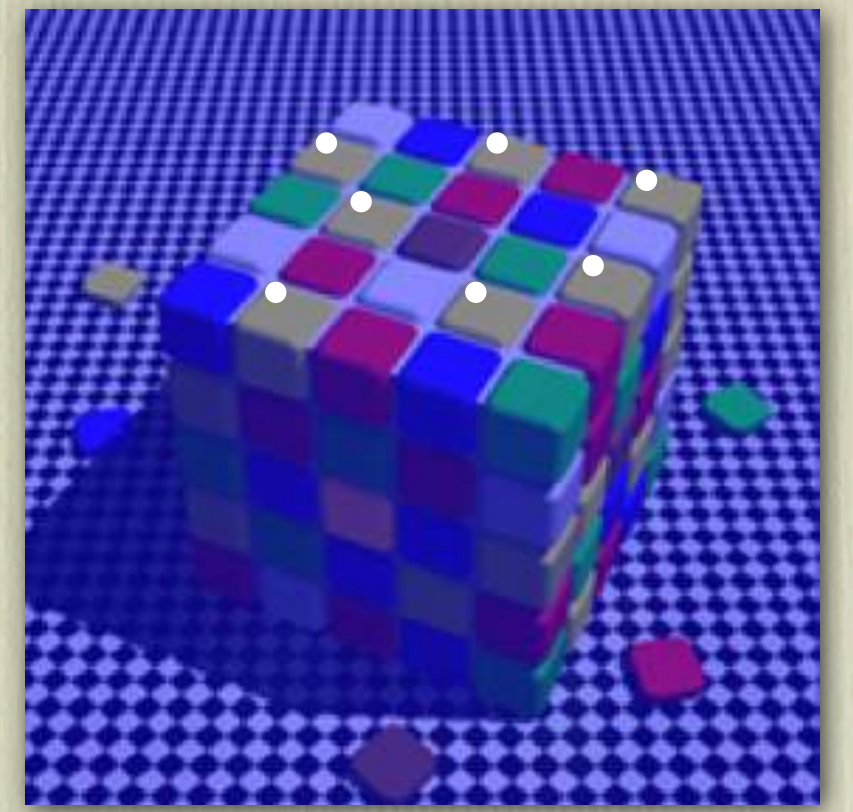


でも，我々は容易に「同一性」を認知できる

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[1,2]

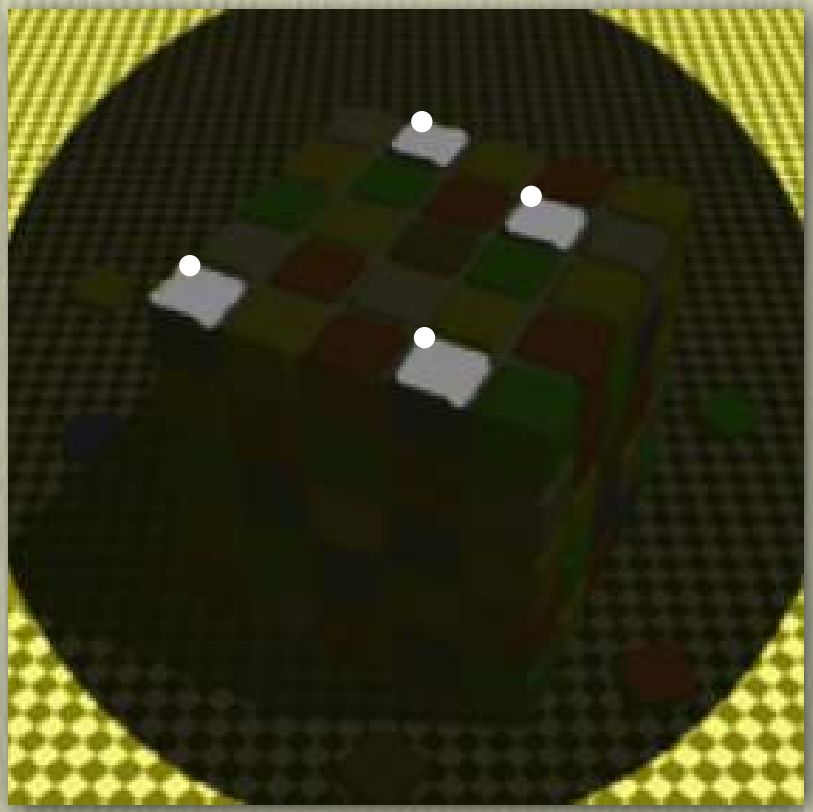


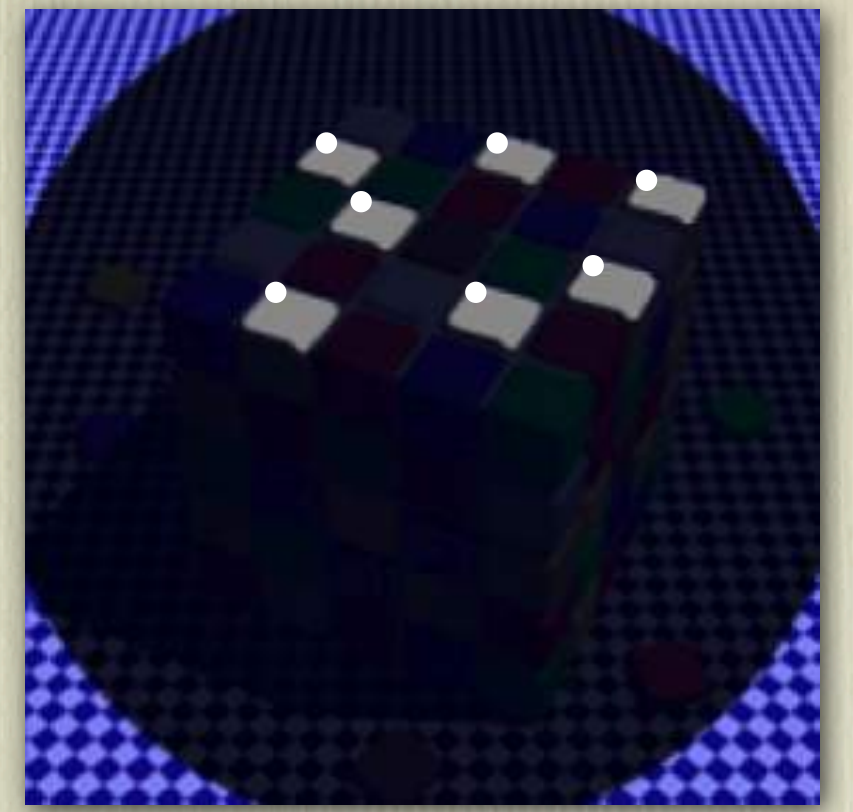


- 両者が同一のキューブであることは容易に認知可能
- 異なる色を同一と主張し，同一の色を異なると主張する。
- 各パッチが持つ波長（絶対量）だけではなく，**各パッチが他のパッチ群とどのようなコントラストを持つのか**，が非常に重要

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[1,2]



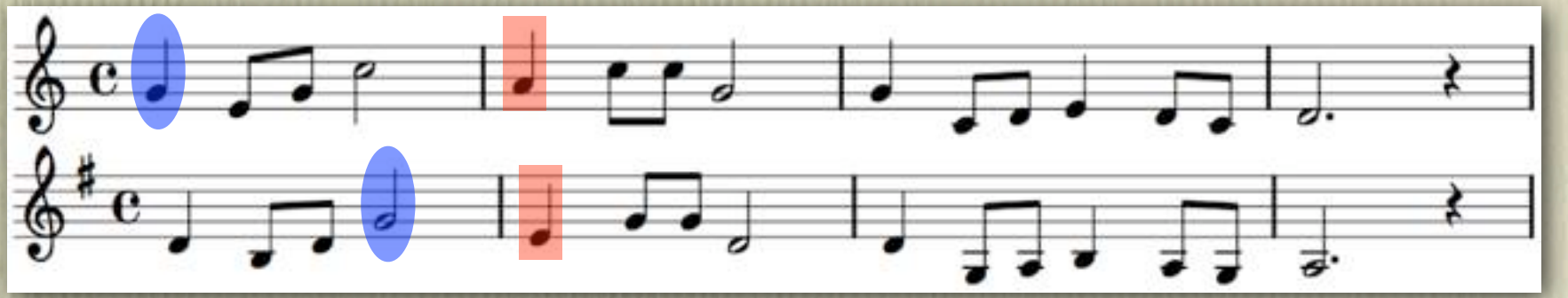


- 両者が同一のキューブであることは容易に認知可能
- 異なる色を同一と主張し，同一の色を異なると主張する。
- 各パッチが持つ波長（絶対量）だけではなく，**各パッチが他のパッチ群とどのようなコントラストを持つのか**，が非常に重要

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[3,4]

1
2



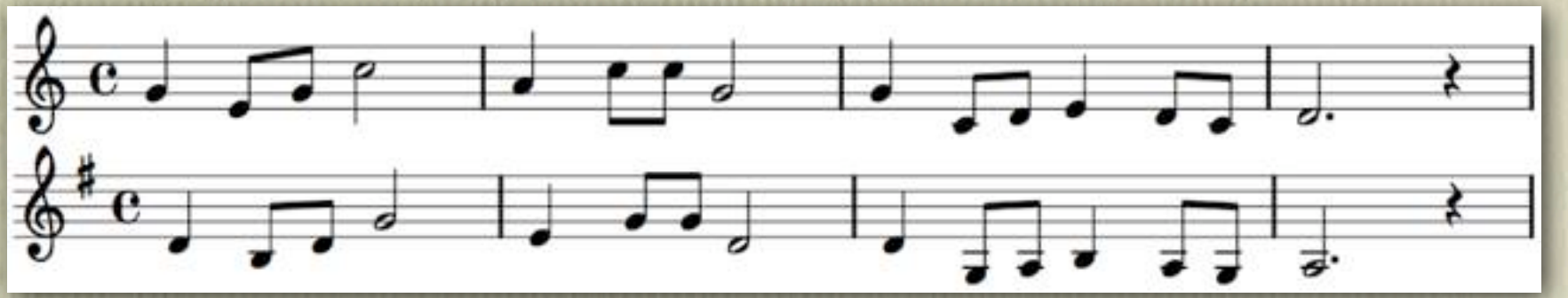
- 絶対音感者（ドレミは音名）
  - １＝ソーミソドーラードドソー，２＝レーシレソーミーソソレー
- 言語化可能な相対音感者（ドレミは階名）
  - １＝ソーミソドーラードドソー，２＝ソーミソドーラードドソー
- 言語化困難な相対音感者（ラーラ音感者）
  - １＝ラーラララーラーラララー，２＝ラーラララーラーラララー
- 異なる音を同一と主張し，同一の音を異なると主張する。
- 各音が持つ基本周波数（絶対量）ではなく，**各音が他の音群とどのようなコントラストを持つのか**，のみによって決定

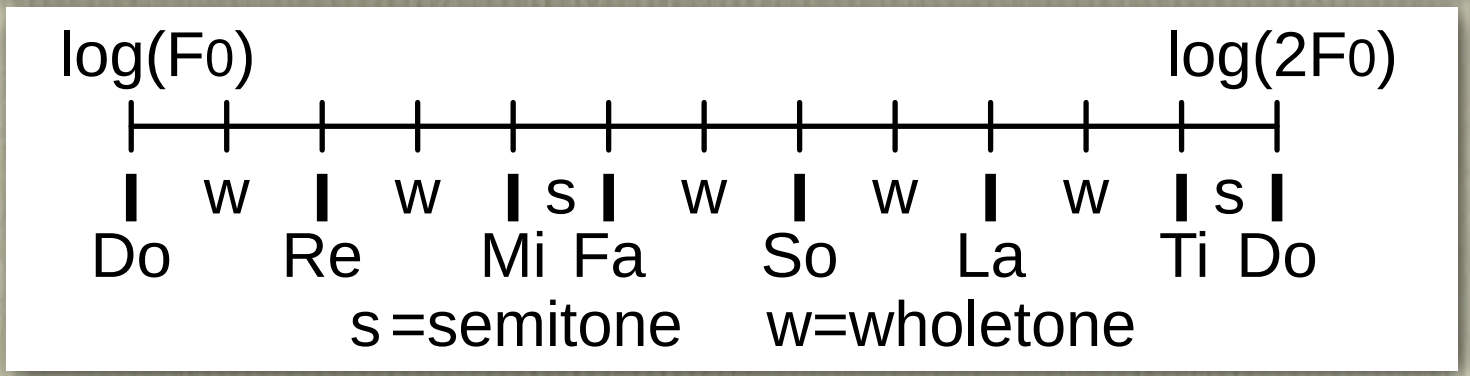
音高の偏差とその認知的不変性

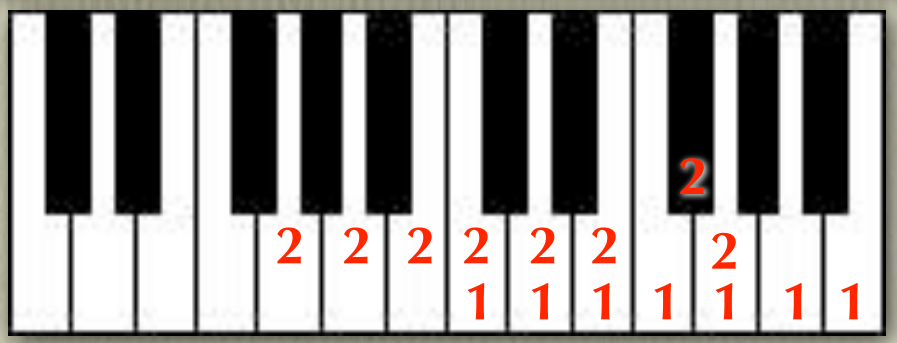
カラオケでキーを上げ下げして曲を聞く[3,4]

1

2



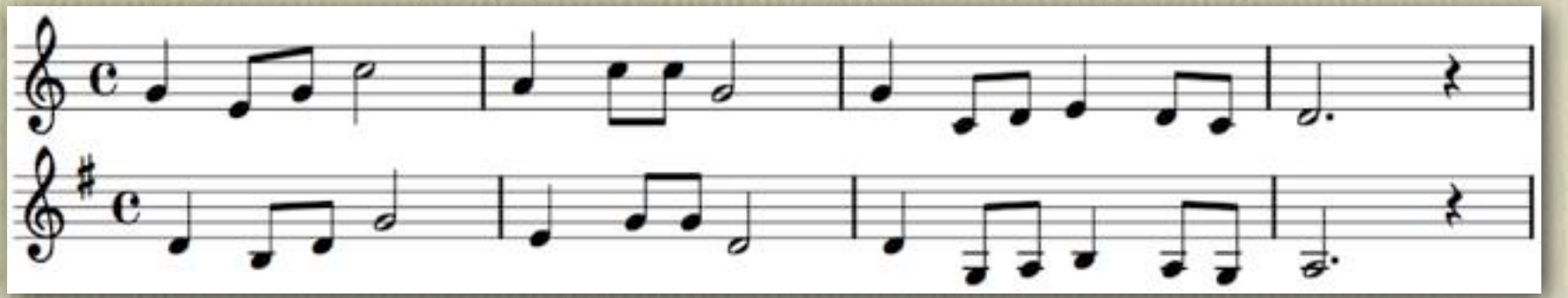




- 各音が持つ基本周波数（絶対量）ではなく，**各音が他の音群とどのようなコントラストを持つのか**，のみによって決定

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[3,4]



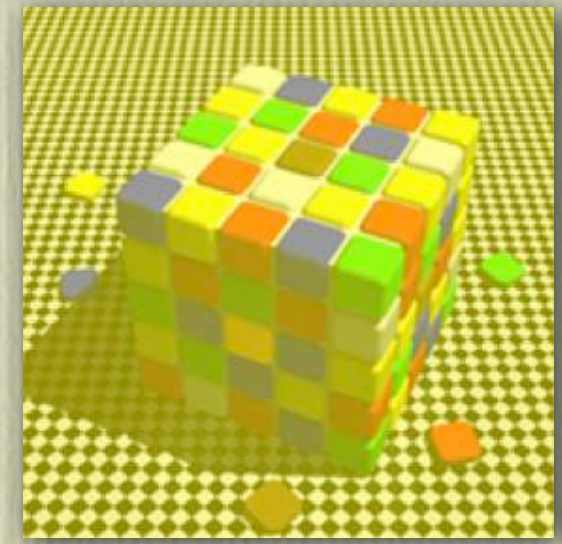
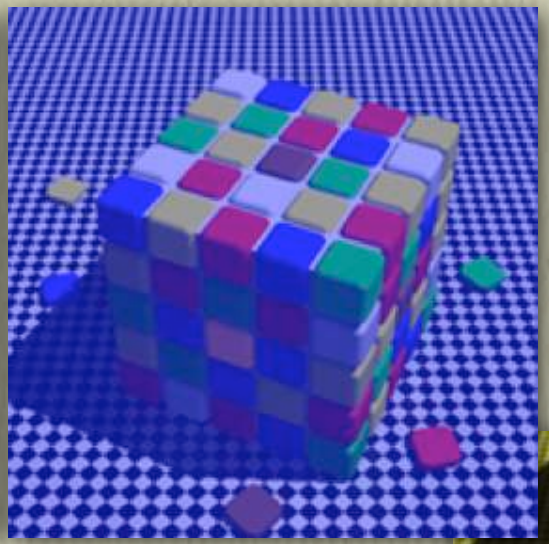
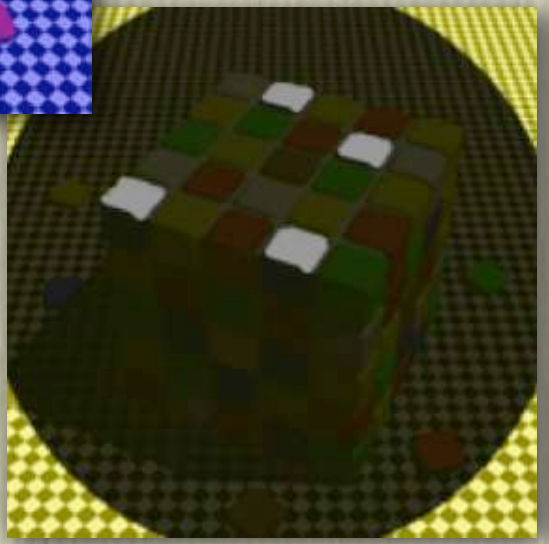
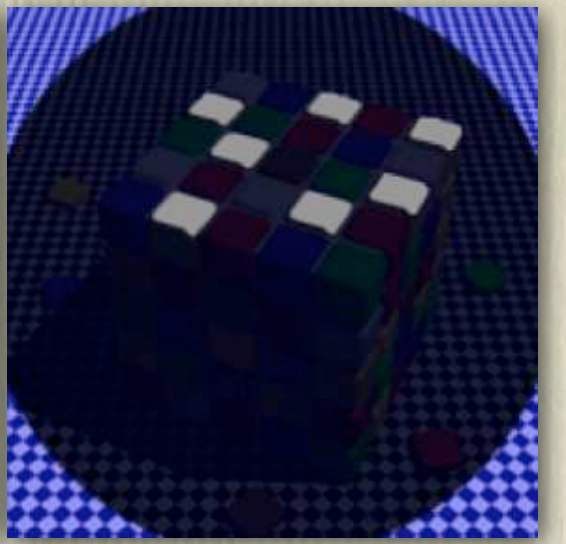
log(F0) log(2F0)

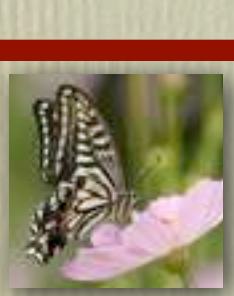
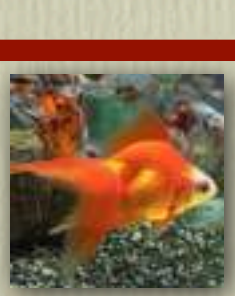
但し，孤立音の同定は不可能
そこにはコントラストが無いから

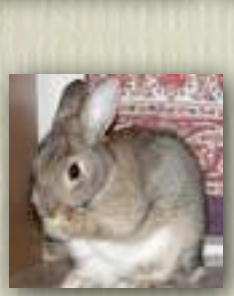
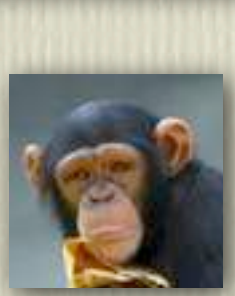
- 各音が持つ基本周波数（絶対量）ではなく，**各音が他の音群とどのようなコントラストを持つのか**，のみによって決定

生物が獲得した静的バイアス除去術

色の恒常的・不変的認知はどこまで遡れるのか？[5]

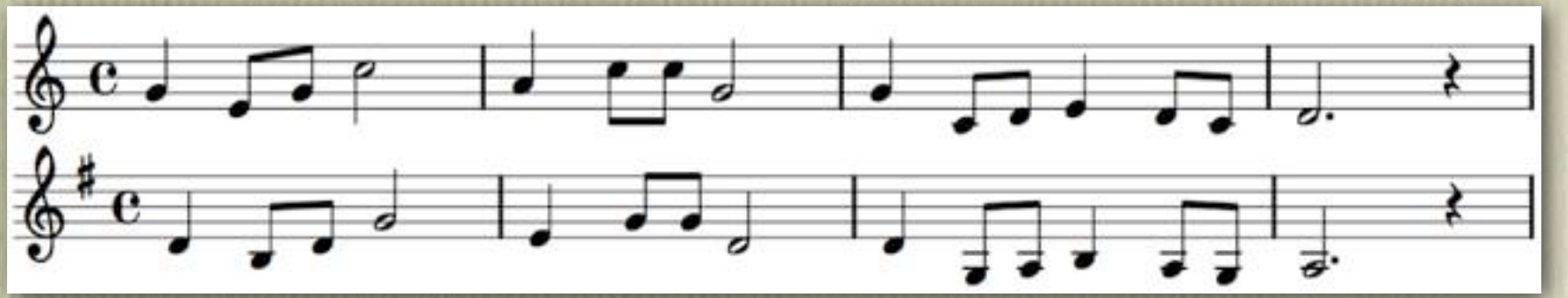




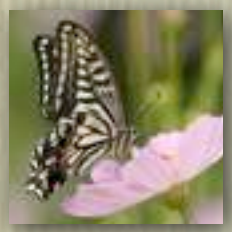

生物が獲得した静的バイアス除去術

音高の恒常的・不変的認知はどこまで遡れるのか？[6]

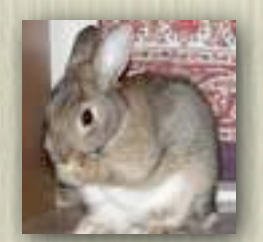
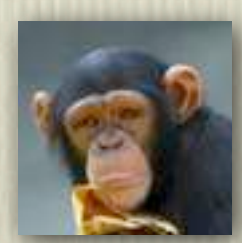
1
2



1 = 2

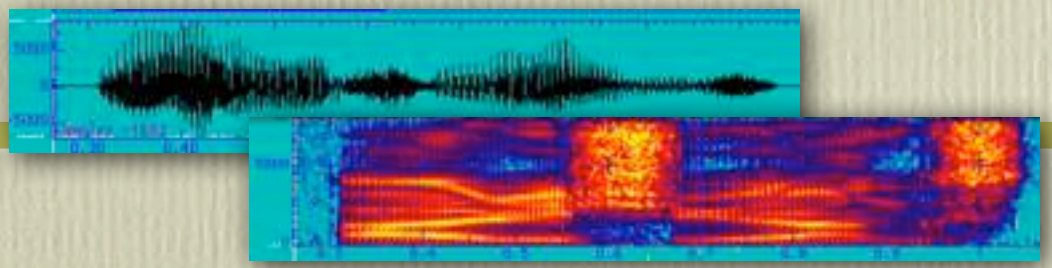



人文社会系研究科基礎文化研究専攻言語学専門分野

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

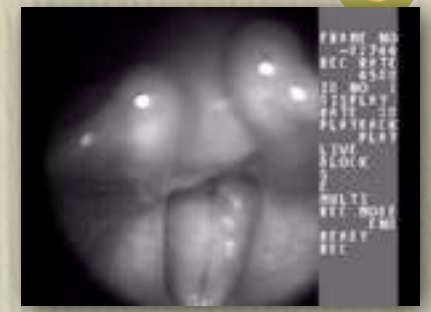
本発表の流れ

- **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？

音色の偏差とその認知的不変性

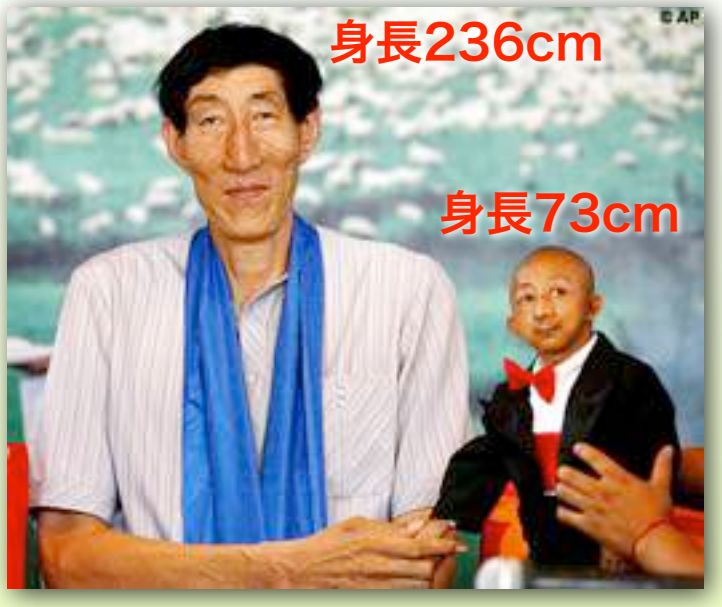
音高の静的偏差を生み出す要因

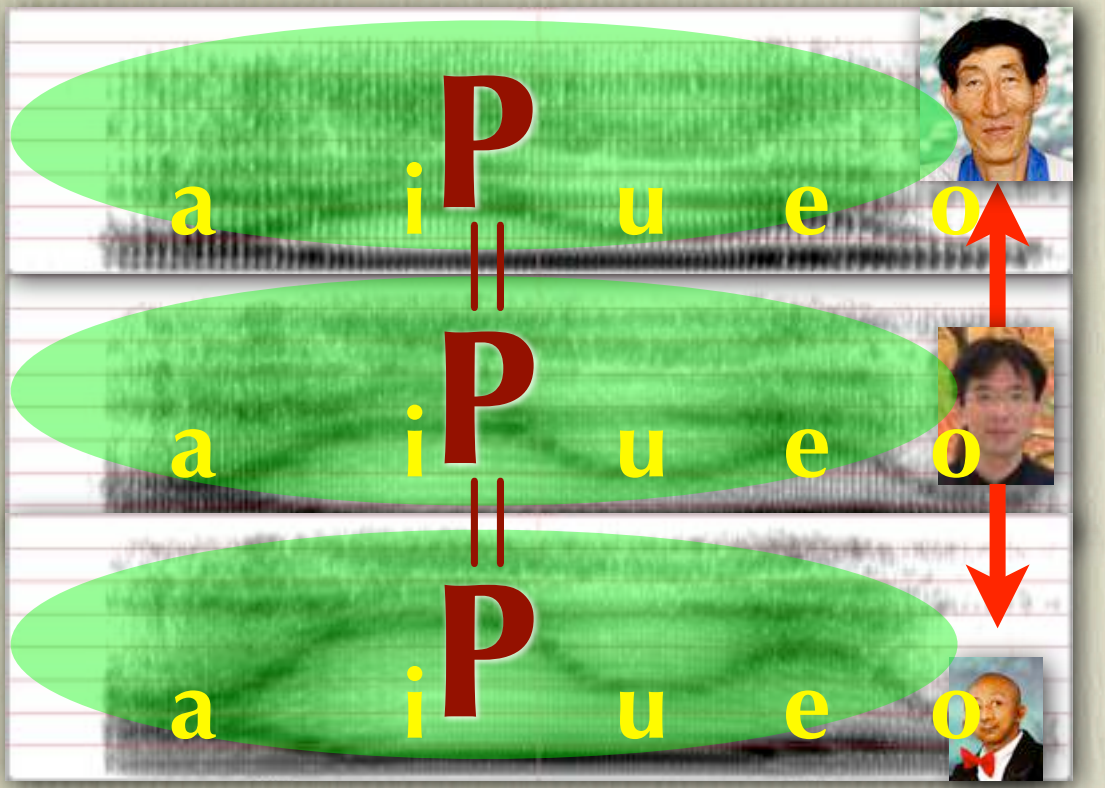
- 男女の音高偏差＝声帯の長さ・重さの性差



音色の静的偏差を生み出す要因

- 男女の音色偏差＝声道の形状（主に長さ）の性差

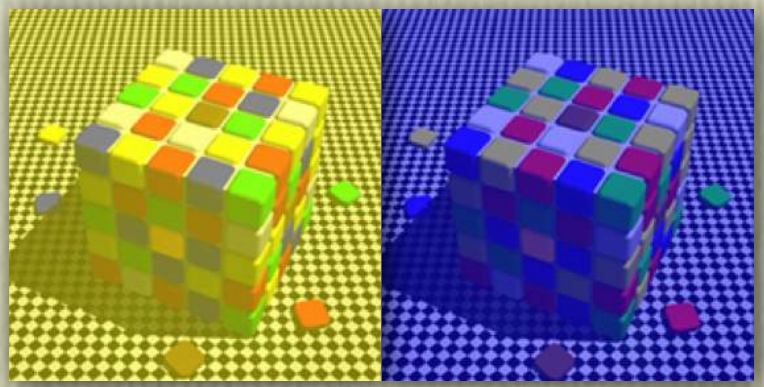


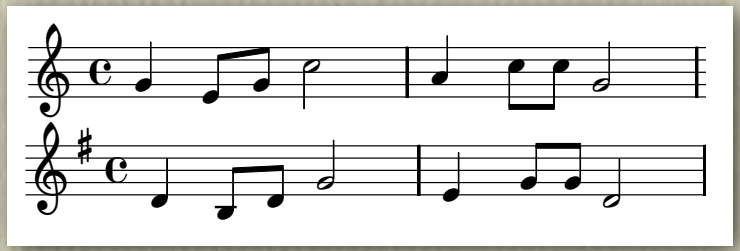


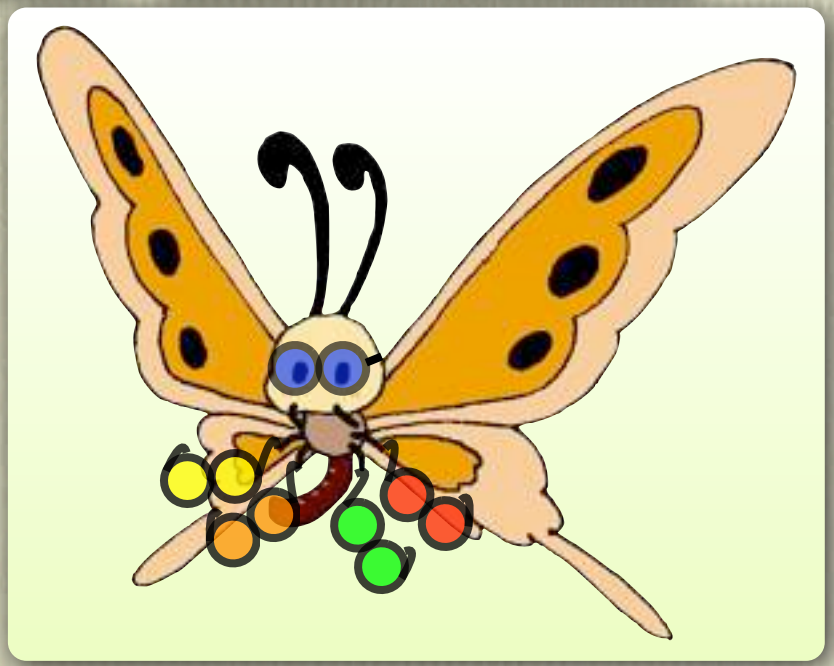
音色の偏差とその認知的不変性

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
- 要素同定ではなく，コントラスト群から成る全体的パターン処理



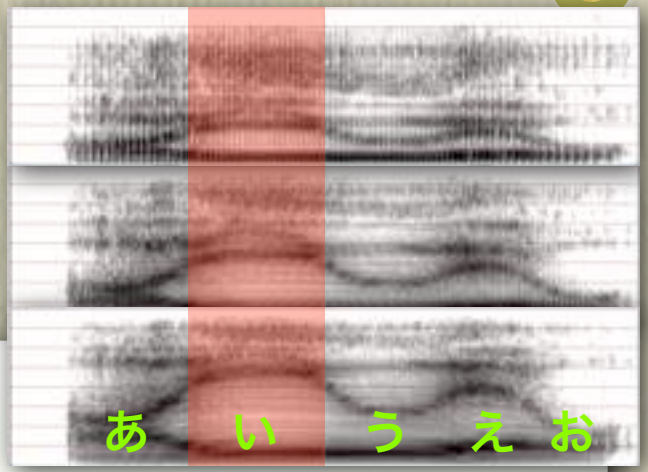




音色の偏差とその認知的不変性

音色の偏差に対する工学的な常套手段

- 音声ストリームを要素列として表象し,
- 個々の要素の統計モデルを作る。

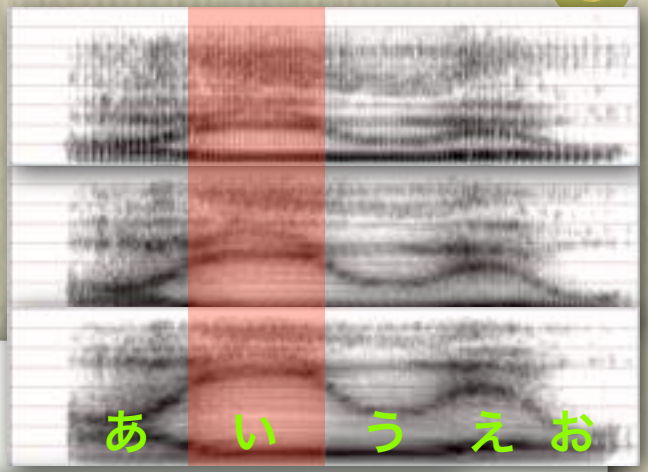


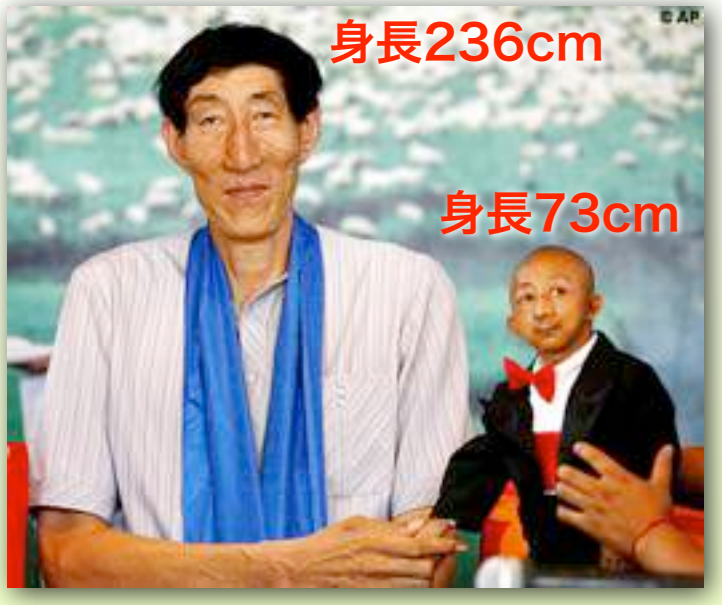
数千〜数十万

音色の偏差とその認知的不変性

音色の偏差に対する工学的な常套手段

- 音声ストリームを要素列として表象し,
- 個々の要素の統計モデルを作る。





数千　数千万

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

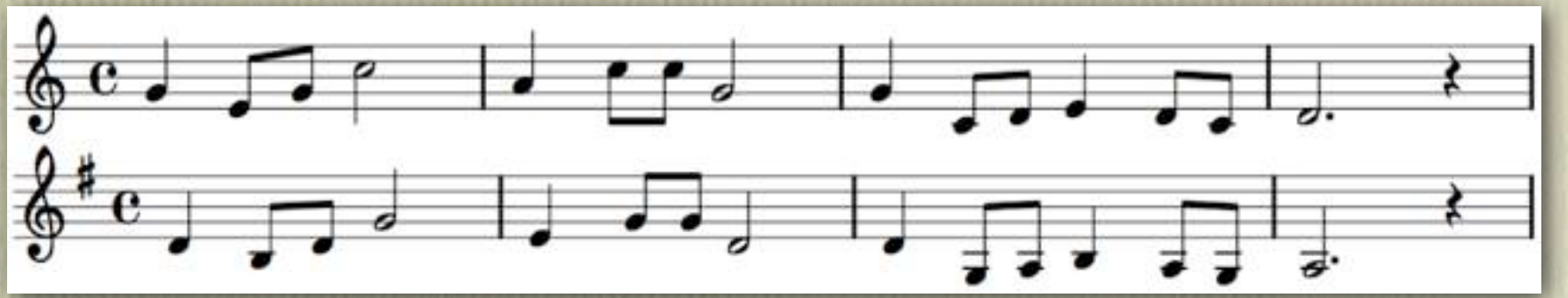
- 九官鳥の音声模倣[9]
  - 車，ドア，椅子，犬，猫，音を真似る。人の声も音でしかない。
  - 良い九官鳥を聞くと，飼い主が分かる。
- 幼児の音声模倣
  - 動物学的には奇妙な模倣行為[10]
  - いくら良い子でも，声から父親を割り出せずにお巡りさんは困る。



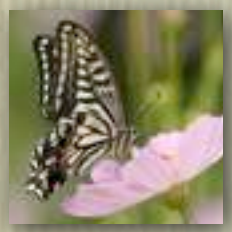

生物が獲得した静的バイアス除去術

音高の恒常的・不変的認知はどこまで遡れるのか？[6]

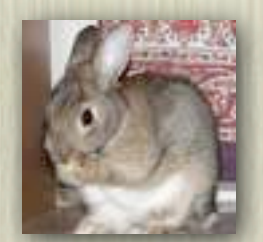
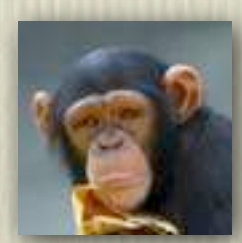
1
2



1 = 2



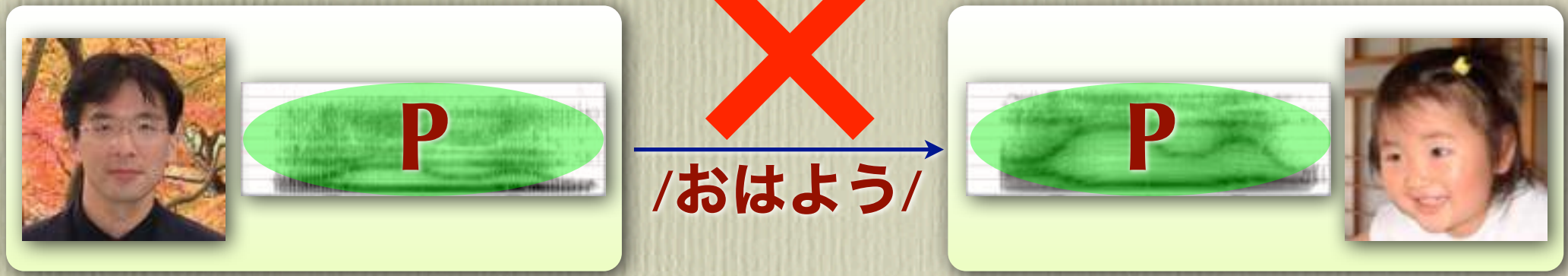




声帯模写と非声帯模写

- 松田聖子・まねだ聖子・神田沙也加

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（仮名の意識）が希薄／しり取りも出来ない。

発達心理学からの回答

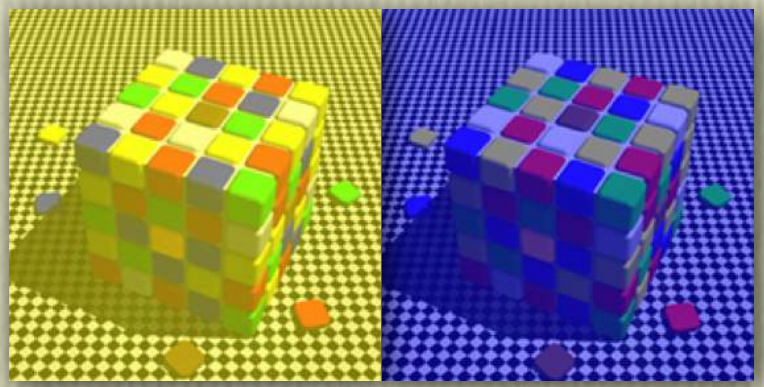
- 幼児は語全体の語形・音形・枠組み・ゲシュタルトを獲得し，その後，個々の分節音（音韻・仮名）を獲得する
- 語ゲシュタルトには話者の情報は含まれない。話者不変量
  - if not，幼児は動物のように音声模倣をすることになる。
- 語ゲシュタルトの物理的・音響的定義は何か？
  - 親の声と幼児の声の「物理的な共通項」は何か？

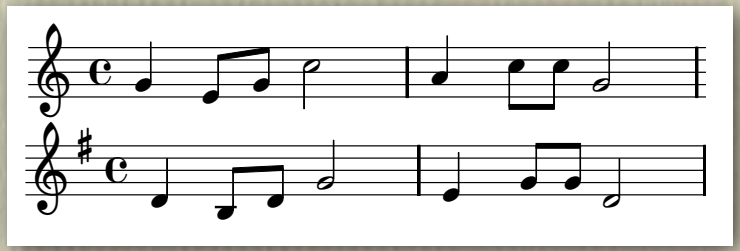
？

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

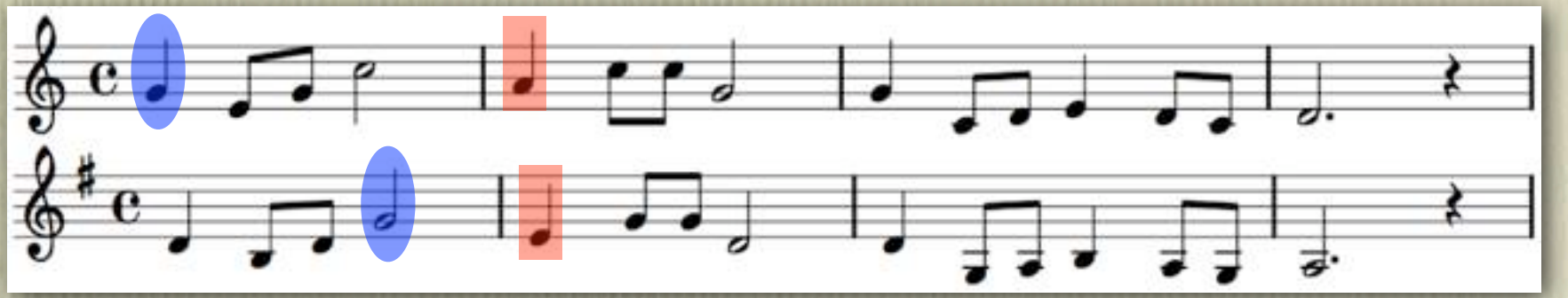
本発表の流れ

- **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[3,4]

1
2



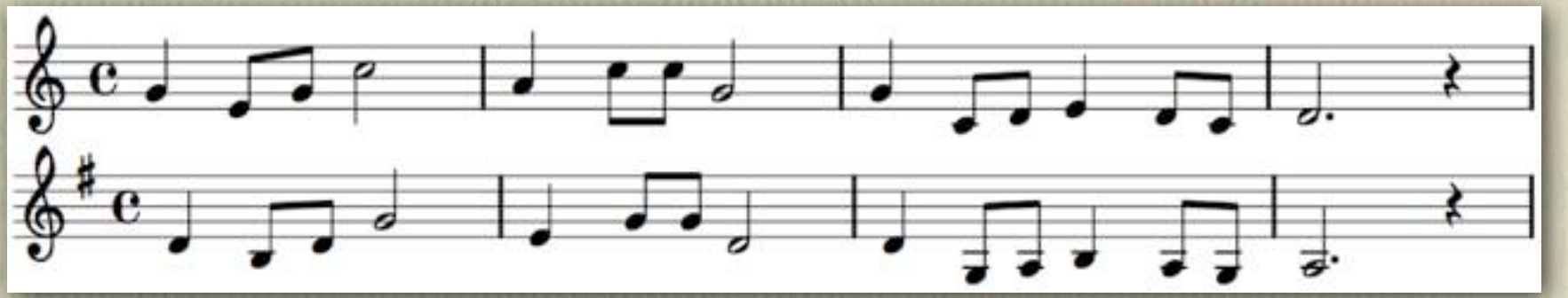
- 絶対音感者（ドレミは音名）
  - １＝ソーミソドーラードドソー，２＝レーシレソーミーソソレー
- 言語化可能な相対音感者（ドレミは階名）
  - １＝ソーミソドーラードドソー，２＝ソーミソドーラードドソー
- 言語化困難な相対音感者（ラーラ音感者）
  - １＝ラーラララーラーラララー，２＝ラーラララーラーラララー
- 異なる音を同一と主張し，同一の音を異なると主張する。
- 各音が持つ基本周波数（絶対量）ではなく，**各音が他の音群とどのようなコントラストを持つのか**，のみによって決定

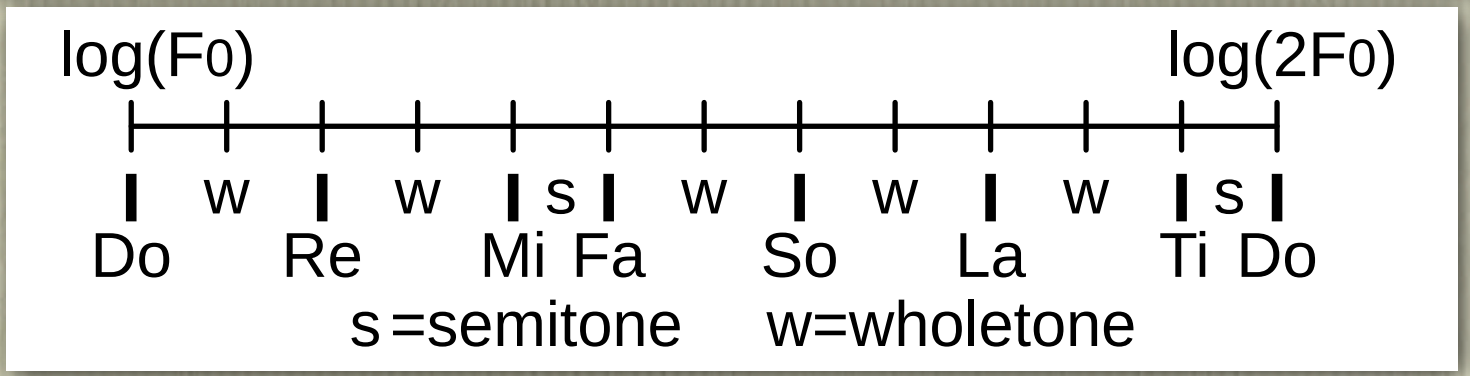
音高の偏差とその認知的不変性

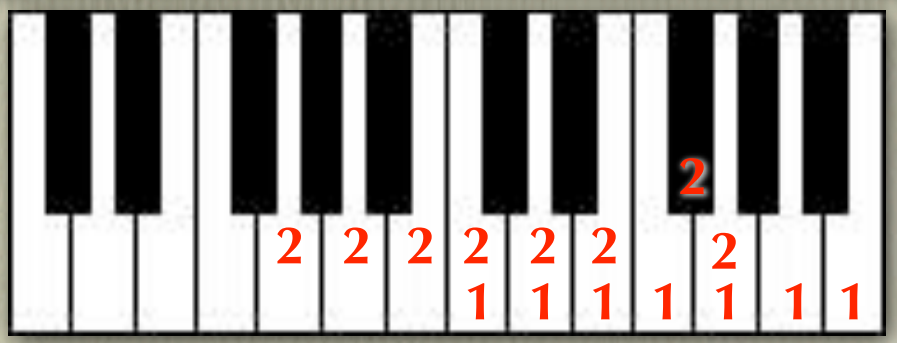
カラオケでキーを上げ下げして曲を聞く[3,4]

1

2



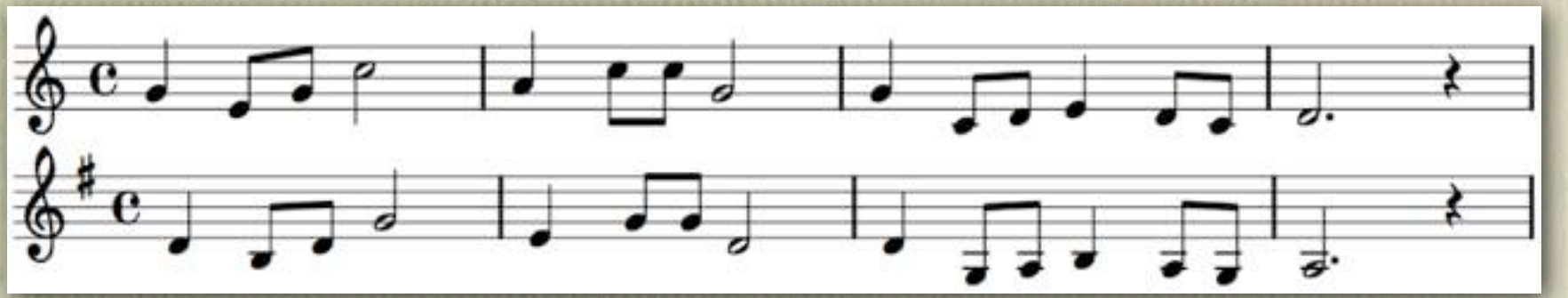




- 各音が持つ基本周波数（絶対量）ではなく，**各音が他の音群とどのようなコントラストを持つのか**，のみによって決定

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[3,4]

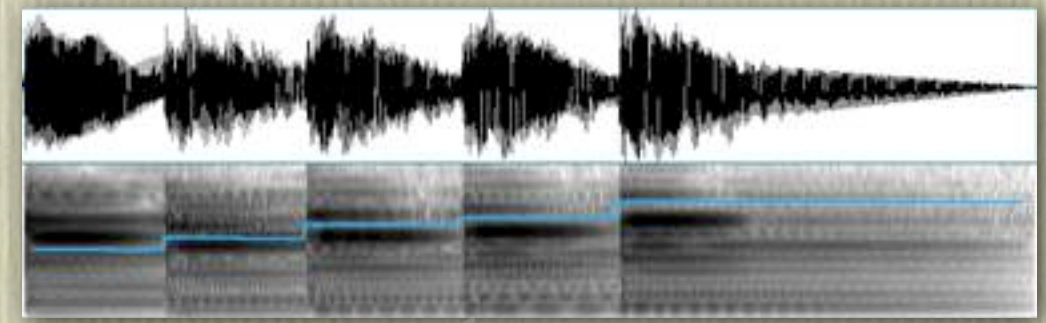


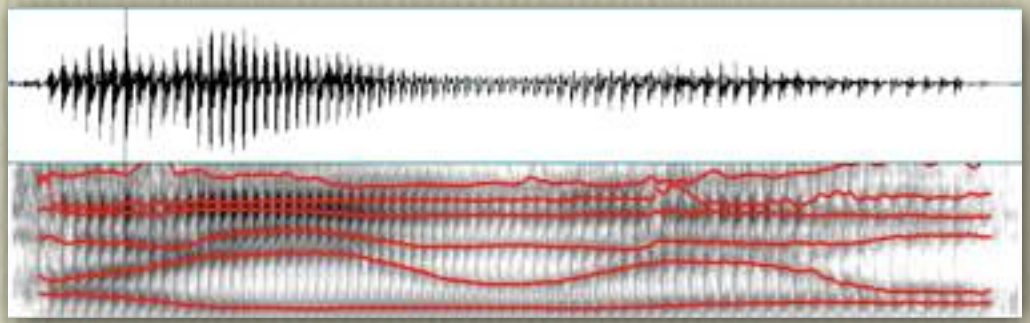
log(F0) log(2F0)

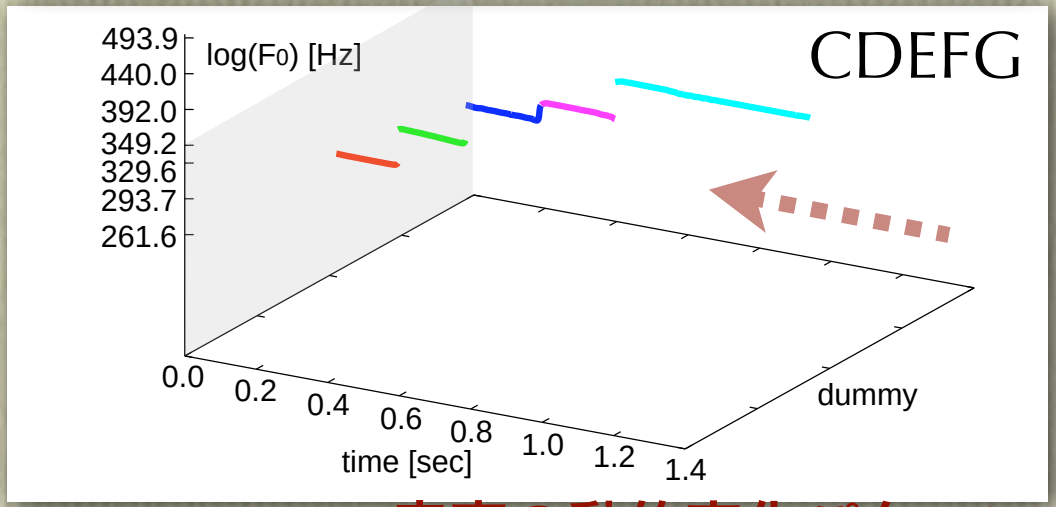
但し，孤立音の同定は不可能
そこにはコントラストが無いから

- 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

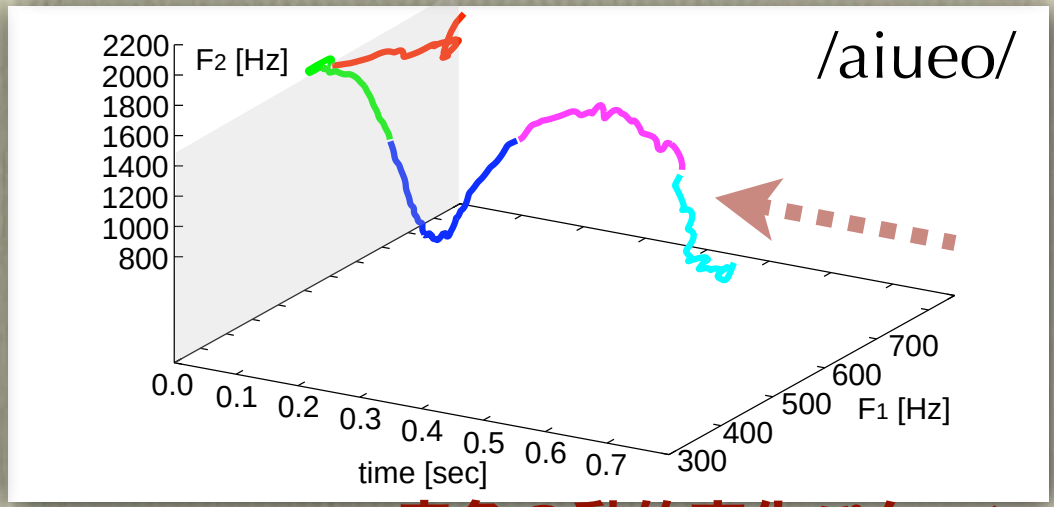
音声の構造的表象／音色の相対音感



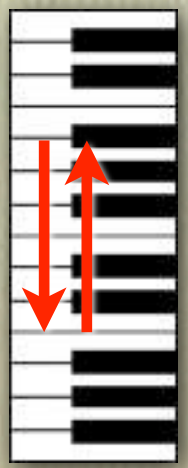




音高の動的変化パターン

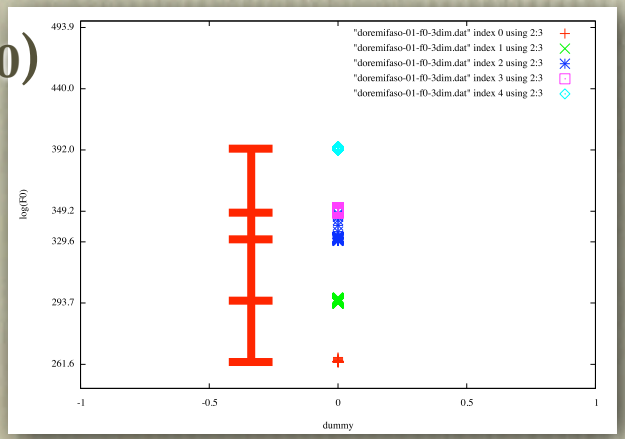


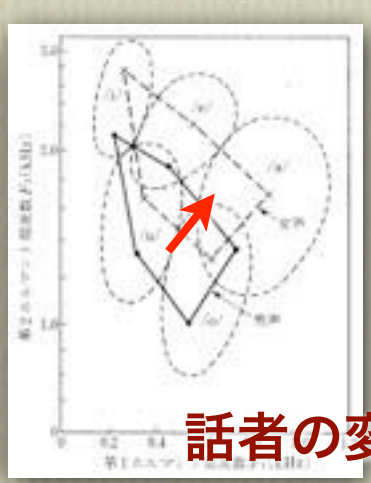
音色の動的変化パターン



調の変化

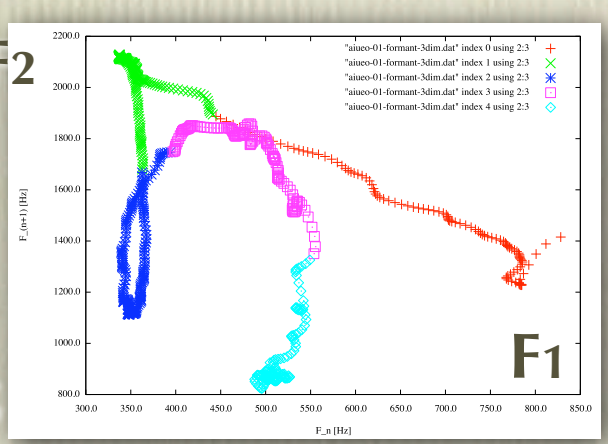
log(F0)



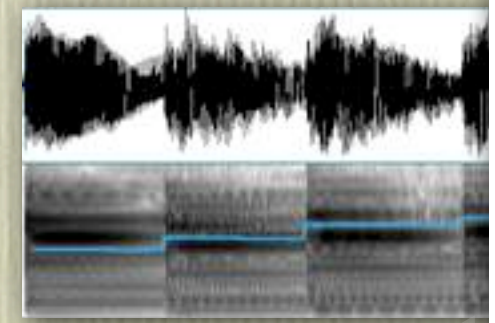


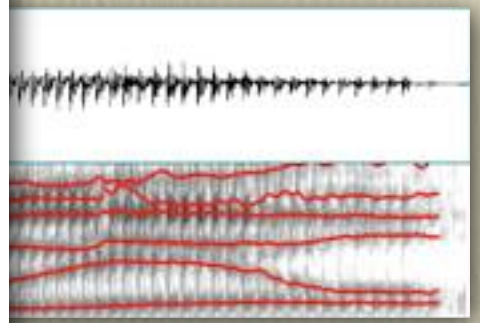
話者の変化

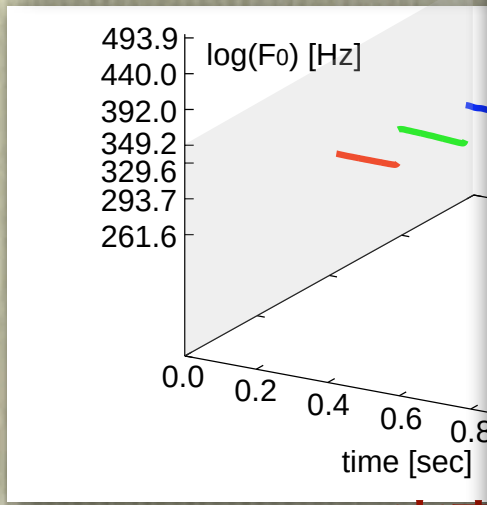
F2



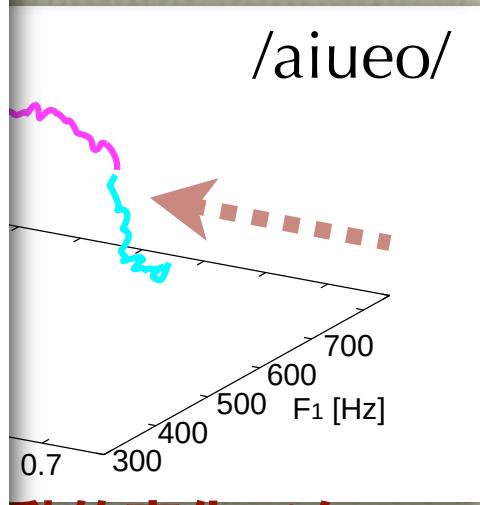
音声の横造的声色（音色）の相対音感



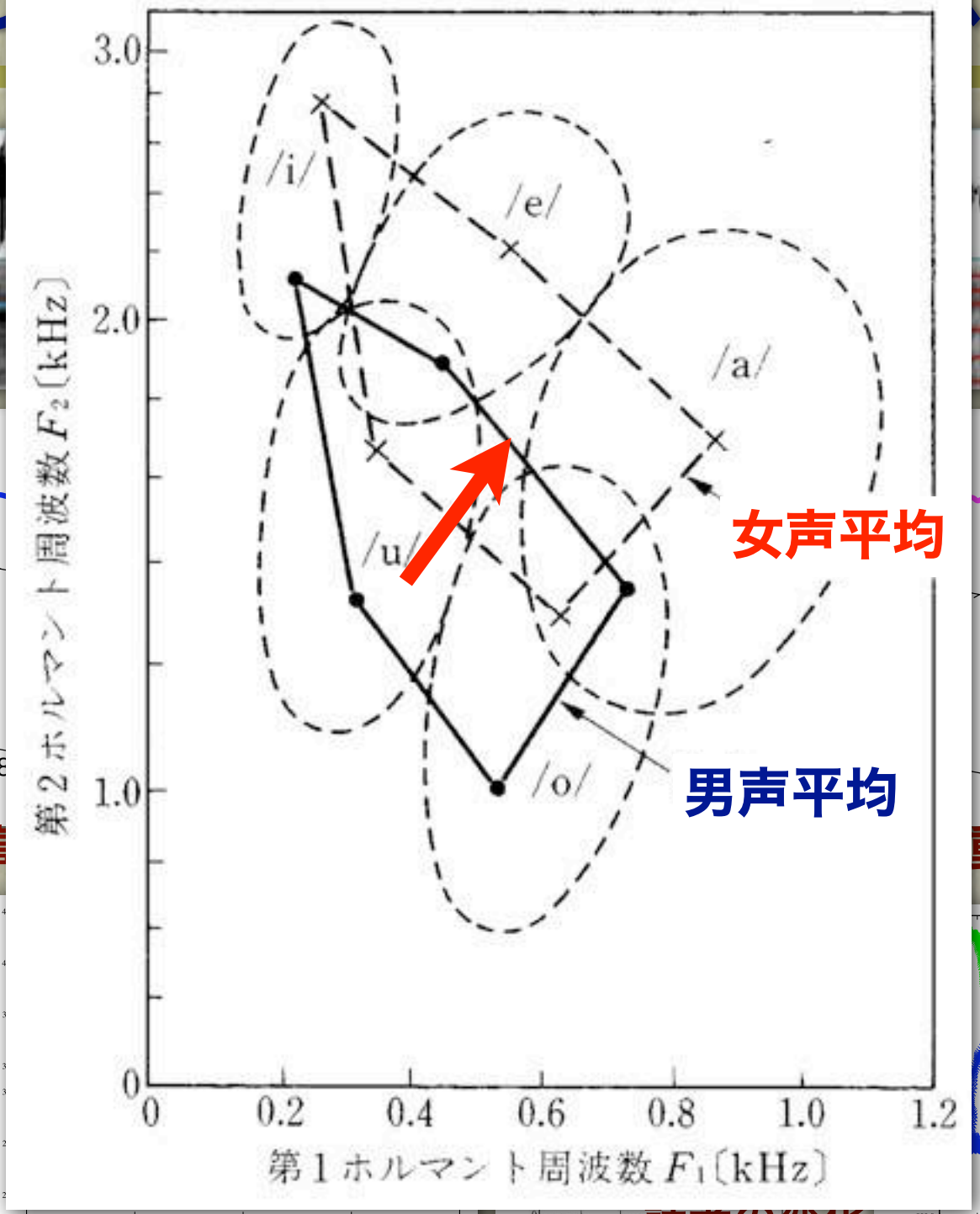


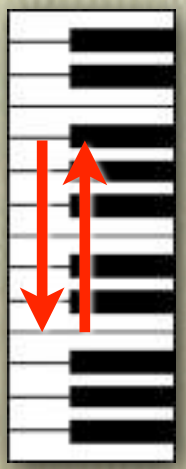


音高



動的変化パターン

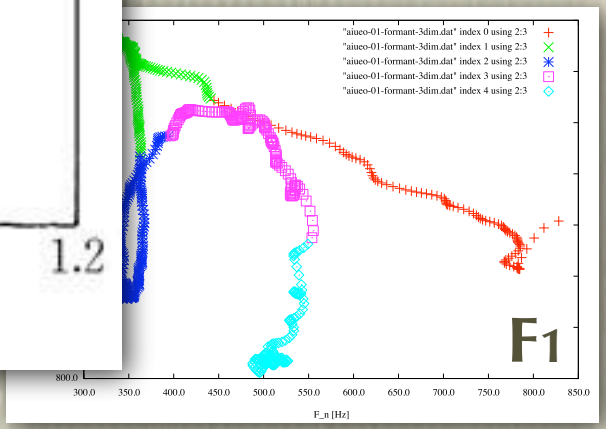




調の変化

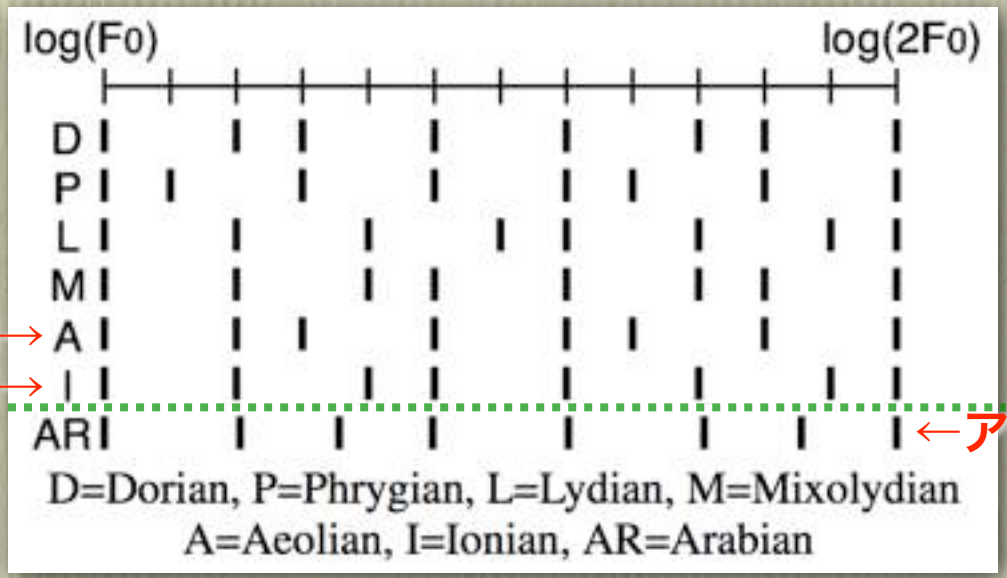
log(F0)

話者の変化



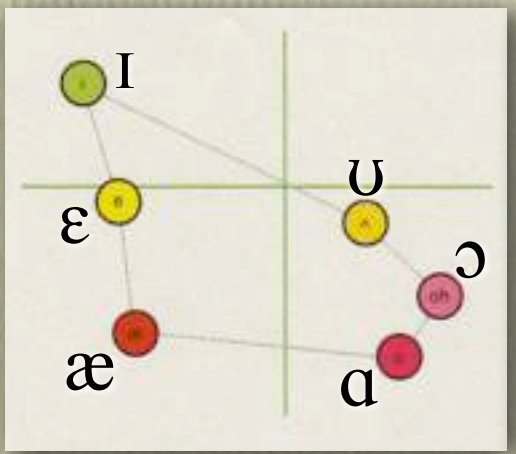
音声の構造的表象／音色の相対音感

音楽における調不変の音配置とその変種

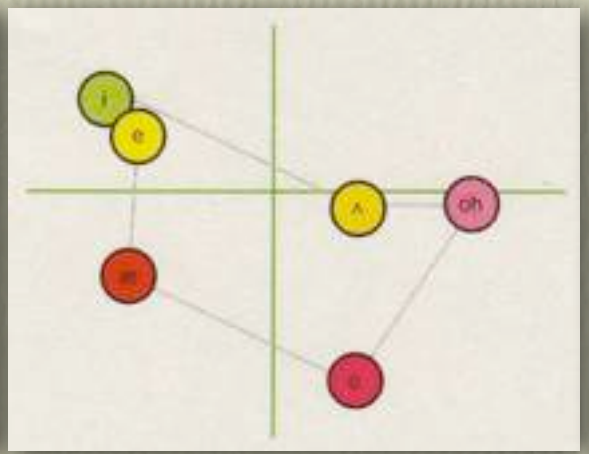


- 西洋音楽＝5全音＋2半音
- 種々の配置＝教会音楽
- 民族音楽には半音以外の配置

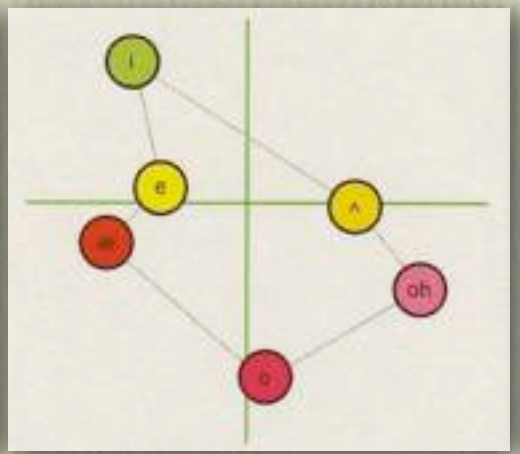
音声における話者不変の音配置とその変種＝欧米の方言



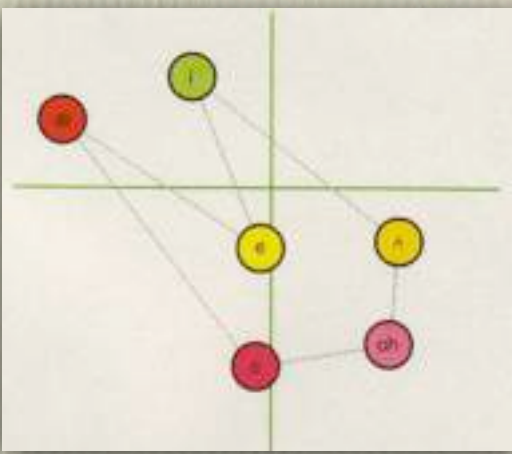
Williamsport, PA



Chicago, IL



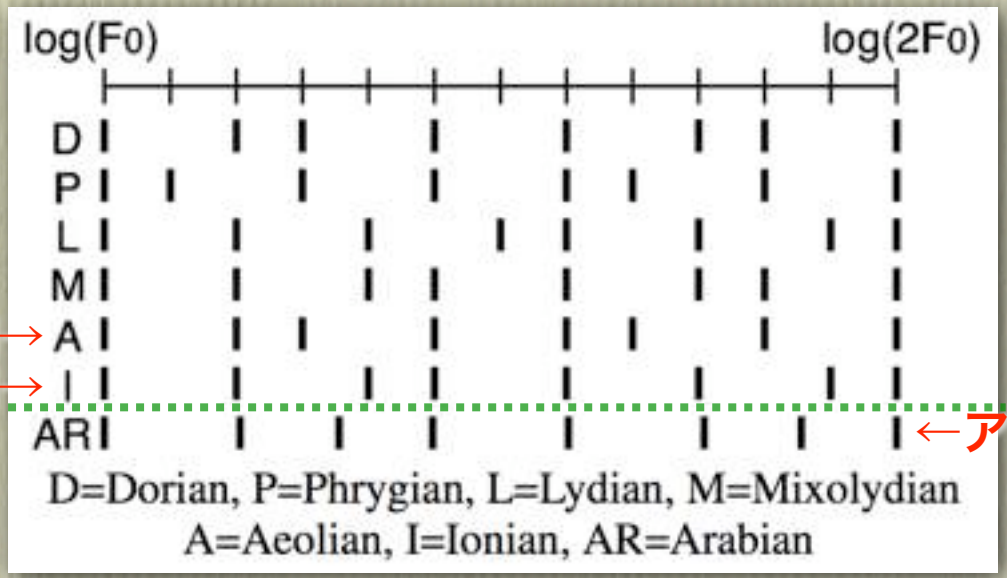
Ann Arbor, MI



Rochester, NY

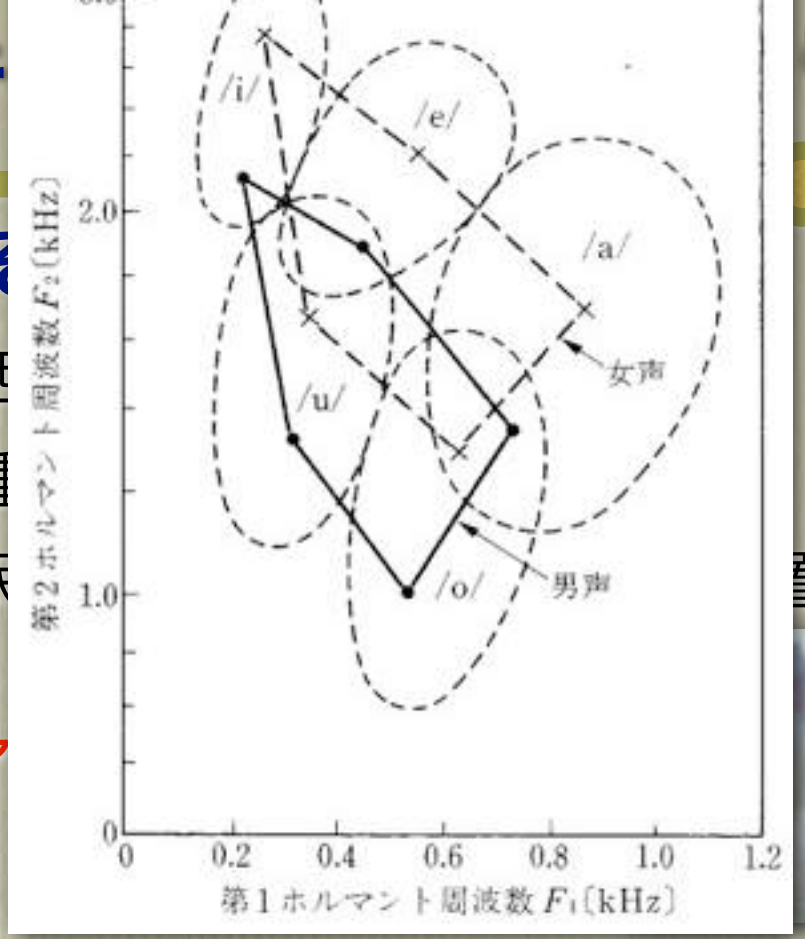
音声の構造的表象／音

音楽における調不変の音配置とそ

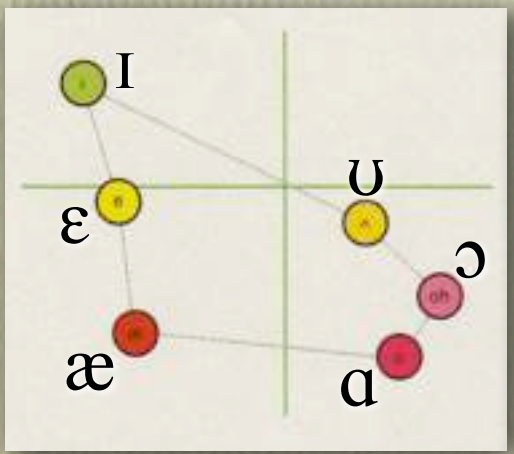


- 西
- 種
- 民

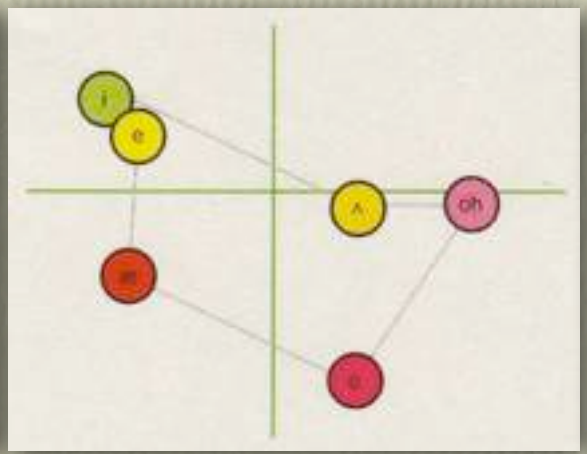
置



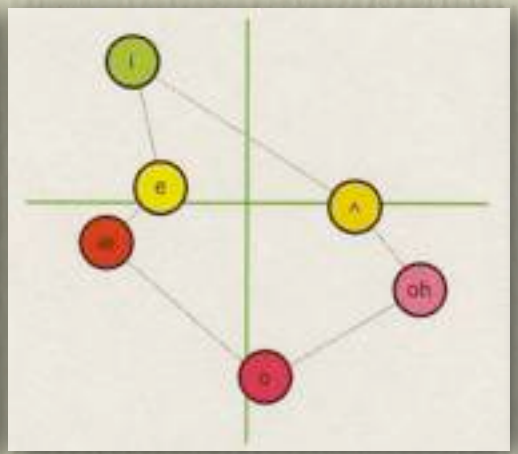
音声における話者不変の音配置とその変種＝欧米の方言



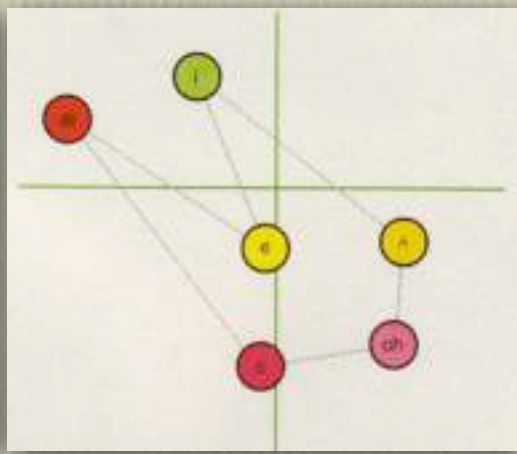
Williamsport, PA



Chicago, IL



Ann Arbor, MI



Rochester, NY

相対音感者による転調部分のドレミ同定

曲の途中で調が変わるとドレミ同定はどうなる？

- 絶対音感者　～音名としてのドレミを使って書き起こす～
  - 何ら問題無く，ドレミ同定する。
  - 時には転調したことに気付かないことすらある。
- 相対音感者　～階名としてのドレミを使って書き起こす～
  - 転調すると，とたんに「ドレミ」同定ができなくなる。
    - いわゆるパニクる。その後しばらくして，まだ出来るようになる。
    - アラビア犬のワルツ→ドレミが聞こえる所と聞こえない所がある。

発話の途中で話者が変わるとモーラ同定はどうなる？

- 話者が変わろうが，同定率が落ちない人
  - 音色の絶対音感者（音色の絶対的特性に基づいて判断する）？
- 話者が変わると，同定率が落ちる人
  - 音色の相対音感者（音色の相対的特性に基づいて判断する）？
  - 話者変化による音色の変化を音韻の変化と捉える人がいる？

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）

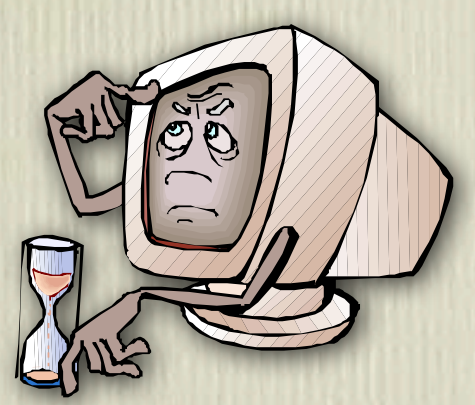
話者がコロコロ変わる音声の知覚

実験手順

- 話者性変化間隔6種類 × 25音声 = 150音声刺激
- Web上の音声ファイルをクリックにてヘッドフォン提示
- 聴取回数 = 2回
- 8モーラであること，一部モーラの欠落は事前教示

三種類の被験者

- 音声研究に従事する大学院生（合成音評価実験経験有）5名
- 法学部大学生（合成音評価実験経験無）3名
- HVite＋CSRC不特定話者音響モデル＋8モーラ認識文法



話者がコロコロ変わる音声の知覚

実験結果に対する四つの予測

- 前後音との関係に基づく聴取 ＝ 音声をより大きな単位で知覚
  - 話者性変化頻度の上昇と共にモーラ同定率は劣化
- 音声をより小さな単位で知覚 ＝ 分析的な聴取
  - 話者性変化頻度によらず一定のモーラ同定率
- 音色の相対音感の存在を本当に知らない唯一の実体
  - 不特定話者音響モデル ＝ 一定のモーラ同定率
- 話者性頻度が極めて大 ＝ 話者性の差の表出・知覚が困難
  - 話者性変化頻度の極端な上昇によってモーラ同定率は向上

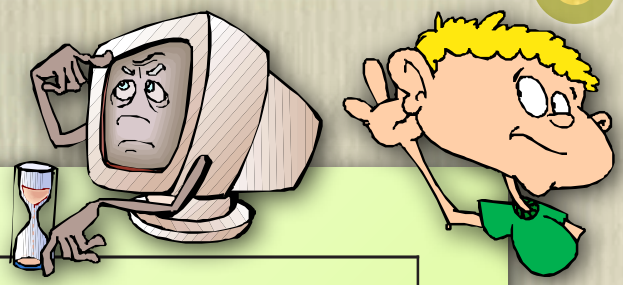
不特定話者音声の例

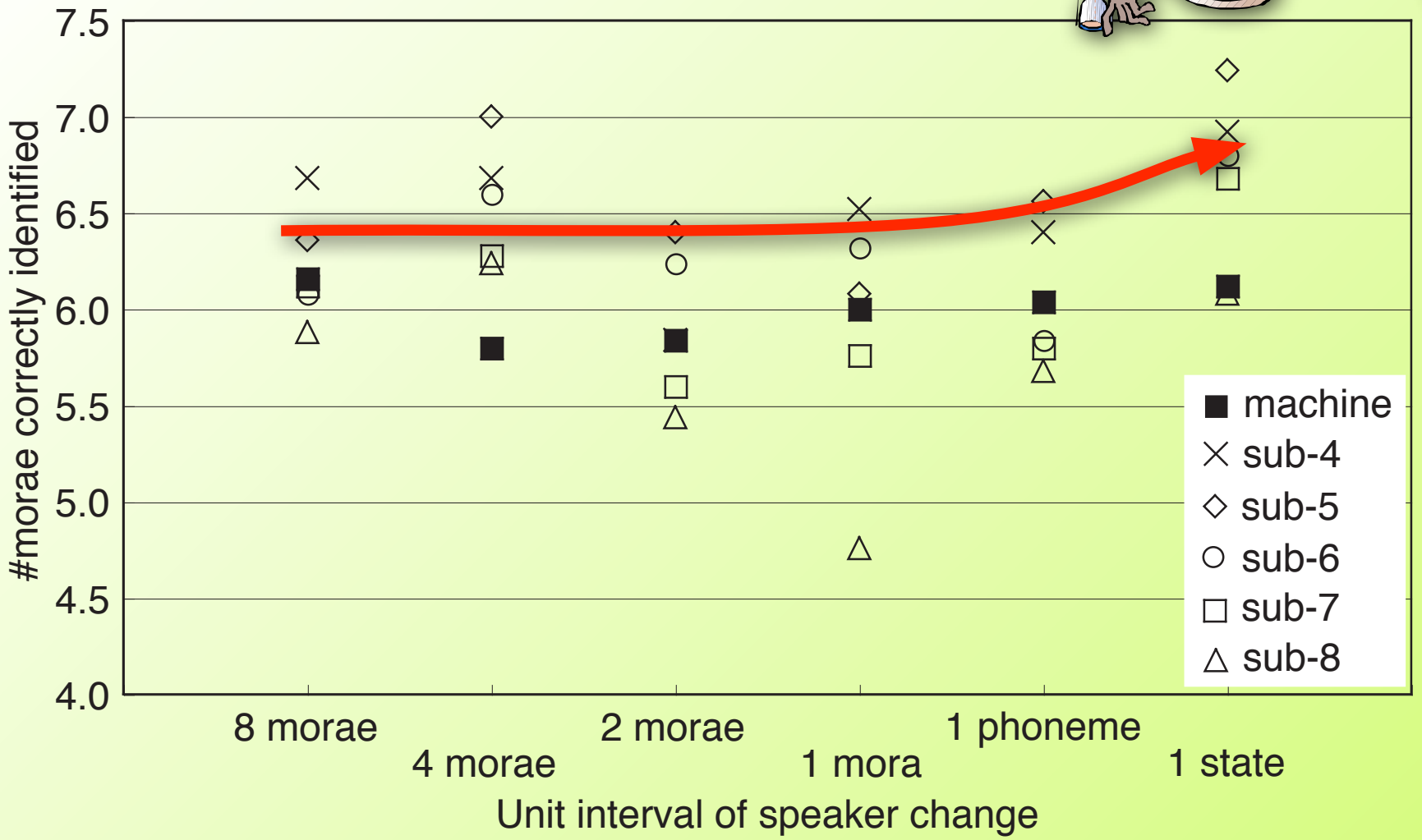
- 1モーラ単位での不特定話者音声
  - 話者性の変化を音韻の変化として認知する可能性

レカルツヌチオキ

話者がコロコロ変わる音声の知覚

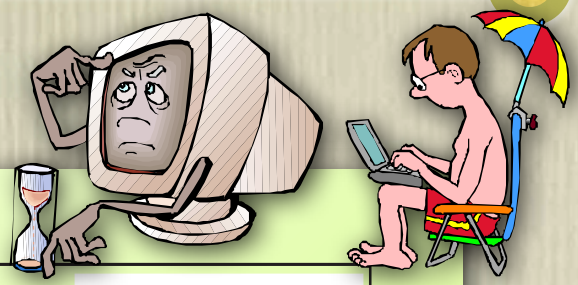
実験結果（計算機と音声研究者）

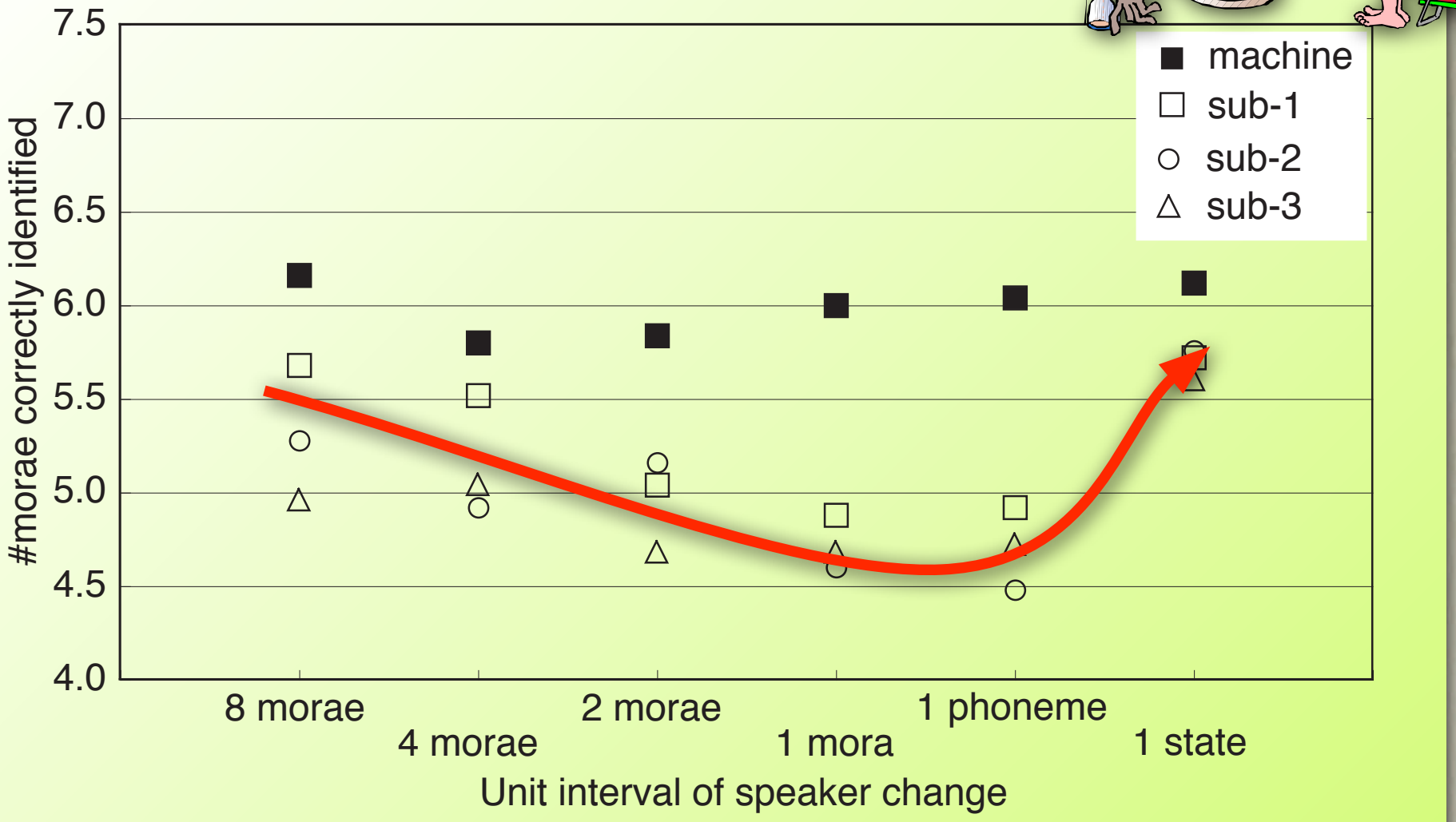




話者がコロコロ変わる音声の知覚

実験結果（計算機と法学部学生）





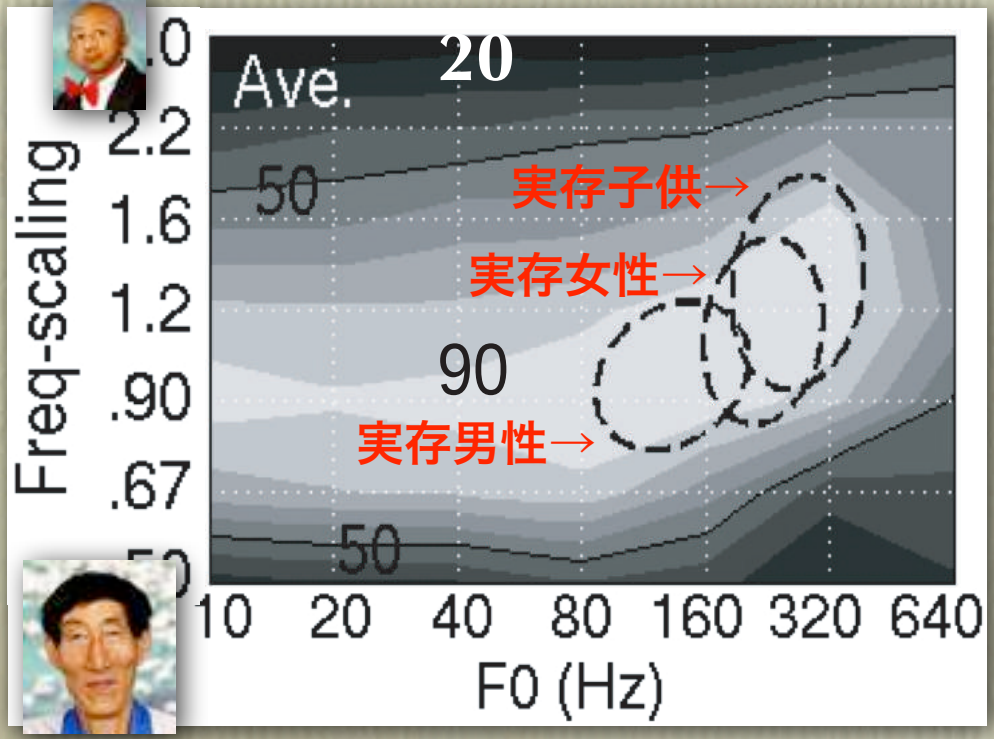
音声の構造的表象／音色の相対音感

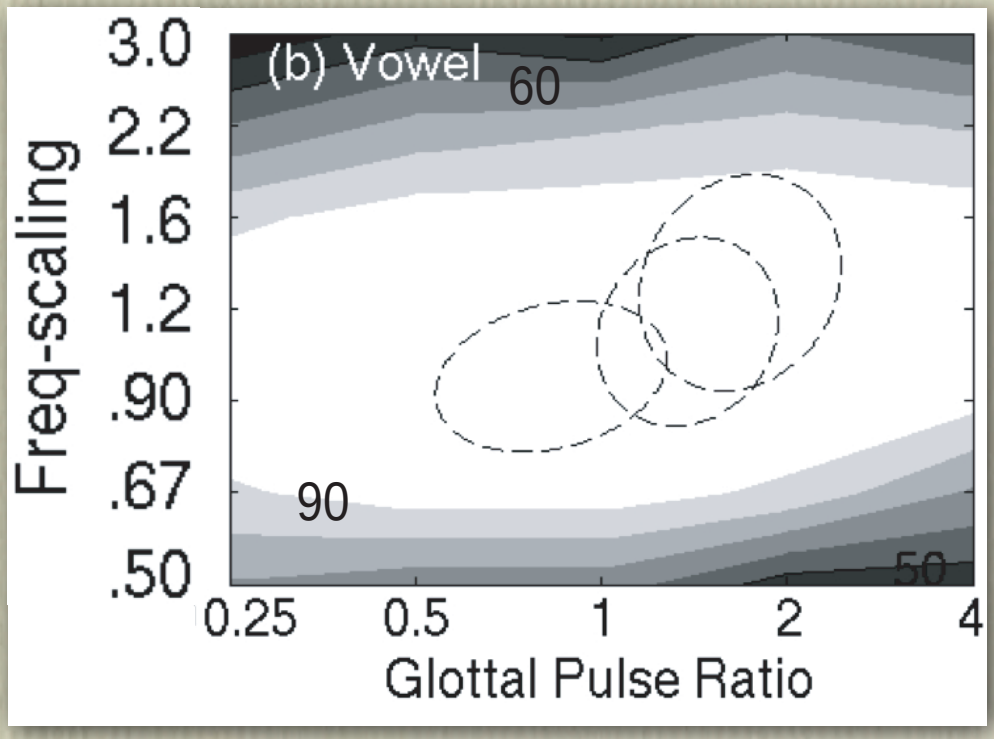
言語化できる相対音感者が出来ないこと

- 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

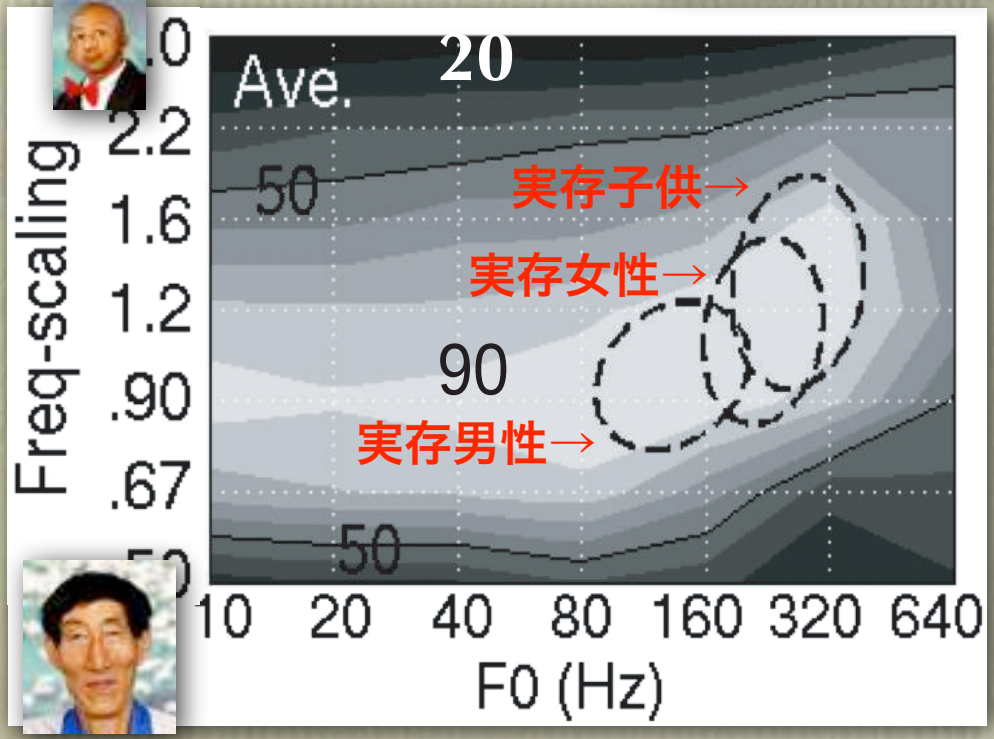
- 孤立母音の同定は困難になる[18]
- でも，無意味語でよいので単語音声にすると書き起こせる[19]

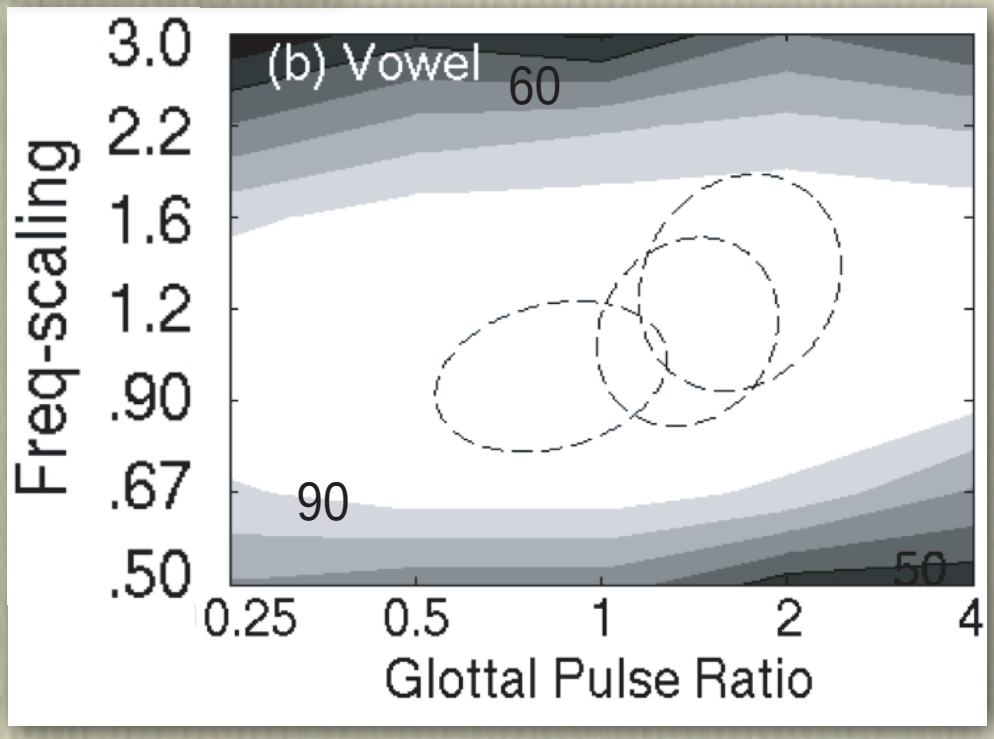




音声の構造的表象／音色の相対音感

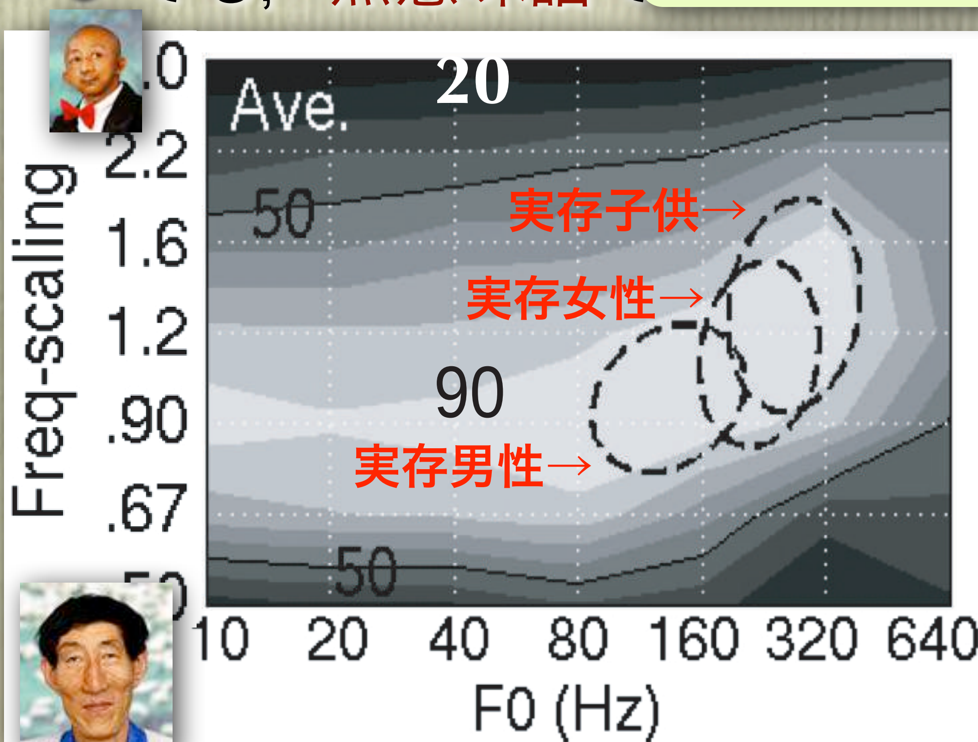
- 言語化できる相対
  - 孤立的に提示され　　とは出来ない。
  - 孤立的に提示され　　人などいるのか？
- 巨人＆小人の音声　　単語同定実験
  - 孤立母音の同定は
  - でも，無意味語で　　と書き起こせる[19]

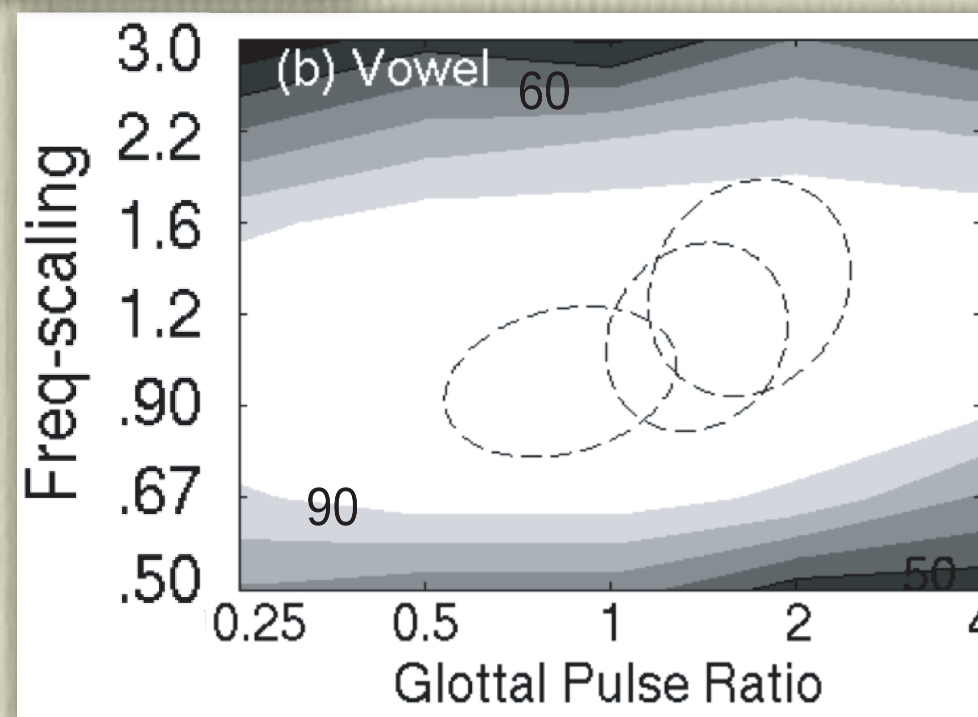




音声の構造… …の相対音感

- 言語化できる相対… …こと
  - 孤立的に提示され… …とは出来ない。
  - 孤立的に提示され… …人などいるのか？
- 巨人＆小人の音声… …単語同定実験
  - 孤立母音の同定は…
  - でも，無意味語で… …と書き起こせる[19]

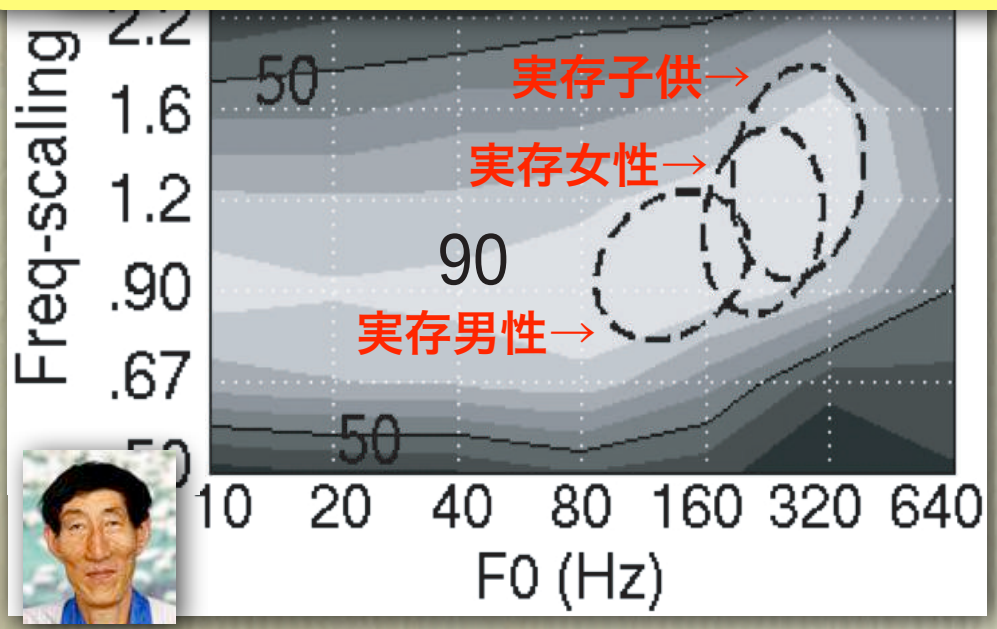


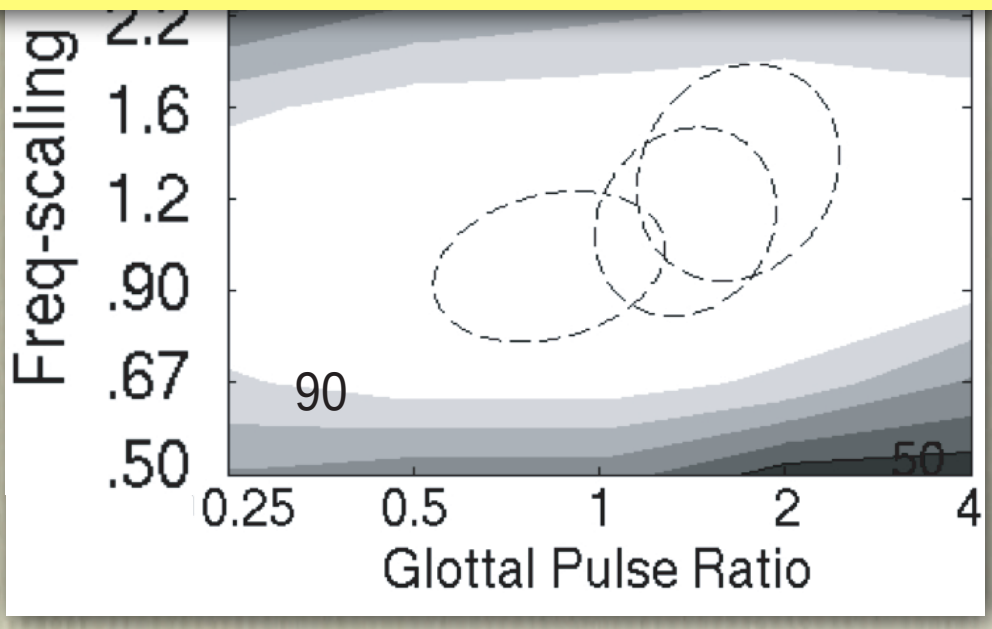


音声の構造的表象／音色の相対音感

- 言語化できる相対音感
  - 孤立的に提示され　　とは出来ない。
  - 孤立的に提示され　　人などいるのか？
- 巨人＆小人の音声　　単語同定実験

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





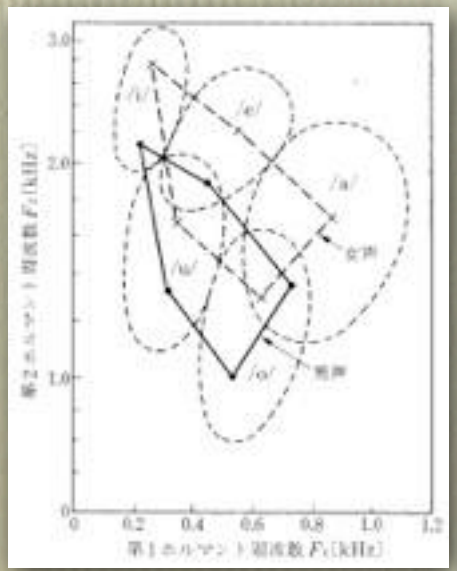
音声の構造的表象／音色の相対音感

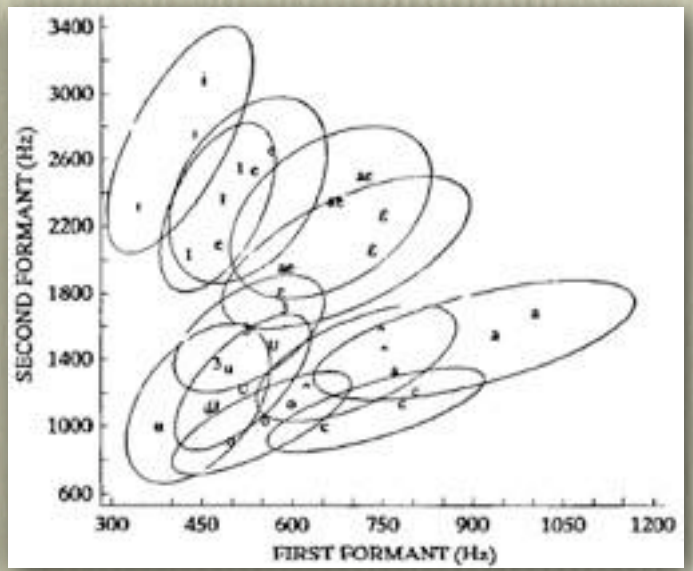
言語化困難な相対音感者（ラーラ音感者）

- 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？

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？）

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

ディスレクシア（読字障害・難読症）

音声の構造 の相対音感

言語化困難な相対 感者）

- 次に示すメロディ て下さい。その後，別の メロディーを提示 来たら挙手しなさい。
  - メロディーをシン ，困難な問いとなる。

言語化困難な音声 児的な成人？）

- 次に示す発話の３ い。その後，別の発話を 提示します。同じ なさい。
  - 発話をシンボル列 れば，困難な問いとなる

ディスレクシア（読字障害・難読症）

とある作文

「あ」という声を聞いて母音「あ」と同定する能力は音声言語運用に必要か？

――音声認識研究からの一つの提言――

第4章

話し言葉の音声

峯松 信明

はじめに ～何、この変なタイトル？～

タイトルを見て、多くの読者が首を傾げていることだろう。しかし、十一頁の本記事を読み終えた時に、ほぼ全ての読者に私の意図は通じるもの、と考えている。そう。「「あ」という声を聞いて、それを有限個の音カテゴリの一つとしての母音「あ」であると同定する能力は、音声言語運用の必要条件ではない。」との主張を本稿では展開する（文献1）（文献2）。

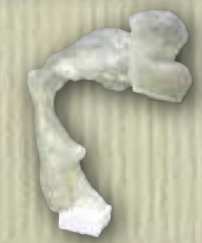
そんな馬鹿な、と思われるかもしれない。こんな実験を考えてみよう。身長300㎝の巨人と50㎝の小人に孤立母音を発声してもらう。通常音声学の教科書には、F_1・F_2の母音図が出ている（図1参照）。複数の男性／女性のサンプルから、凡そ男性の各母音はこの領域、女性の各母音はこの領域にある、といった図である。フォルマント周波数（共鳴周波数）は声道長に依存するため、身長が50㎝、300㎝という架空の大人を想定した場合、彼らの母音は、通常知られている領域の外に存在する。そのような母音でも、現在の音声分析・再合成技術を使えば非常に高品質な音声として生成できる。さて、聞いたことのない母音音声を孤立提示されて、読者は同定できるだろうか？

文献(5)によれば、これは困難なタスクであることが分かる。しかし、その巨人、小人が無意味モーラ列を単

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？

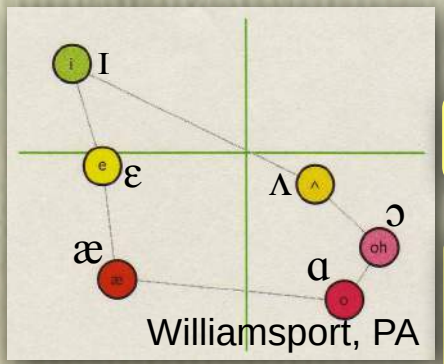


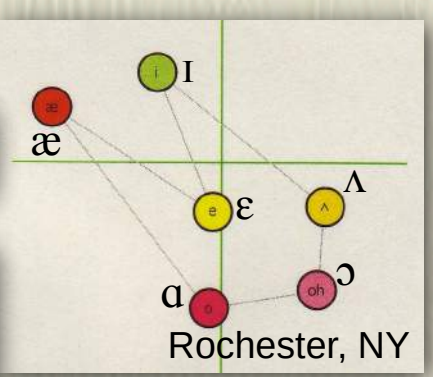
声道形状の性差＝音色の差異

? ?

方言差＝音色の差異

幼児が学ぶものを学ぶ機械
幼児が無視するものは無視する機械





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（仮名の意識）が希薄／しり取りも出来ない。

発達心理学からの回答

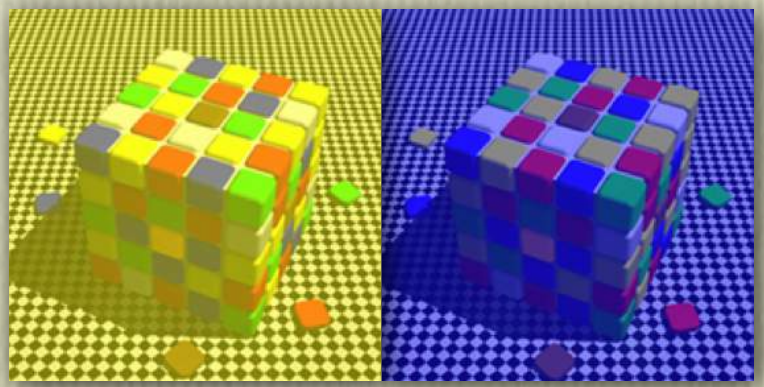
- 幼児は語全体の語形・音形・枠組み・ゲシュタルトを獲得し，その後，個々の分節音（音韻・仮名）を獲得する
- 語ゲシュタルトには話者の情報は含まれない。話者不変量
  - if not，幼児は動物のように音声模倣をすることになる。
- 語ゲシュタルトの物理的・音響的定義は何か？
  - 親の声と幼児の声の「物理的な共通項」は何か？

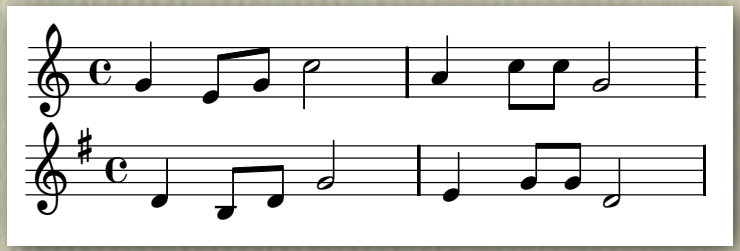
？

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



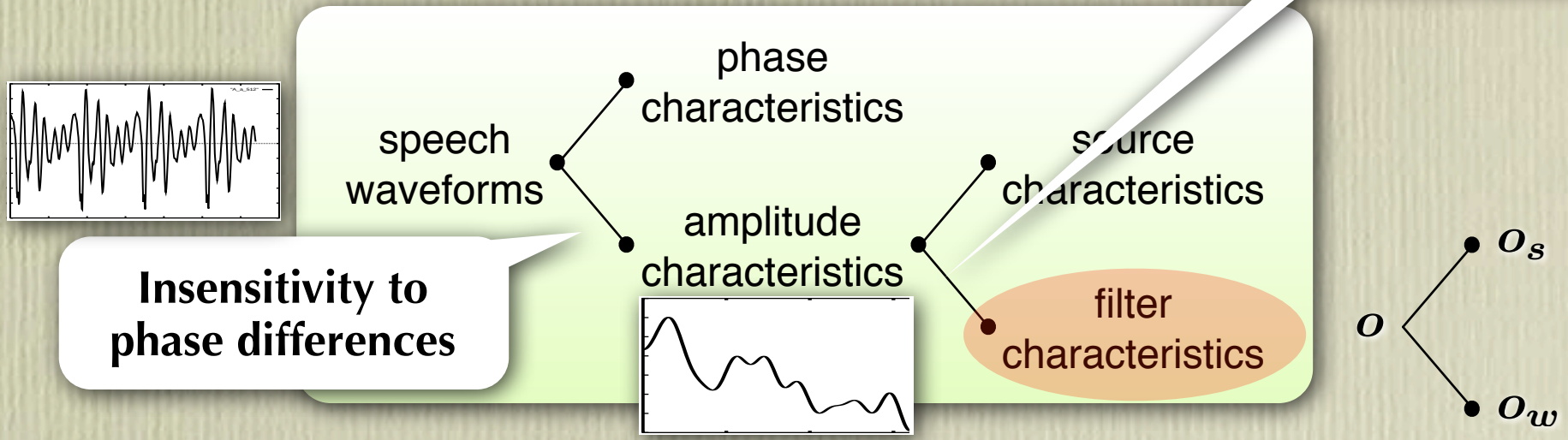


音色の恒常・不変的認知

- 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

その情報を運ぶ媒体・音響特徴量

二段階の分離に基づく特徴量抽出



- スペクトル包絡(o)は何を運ぶのか？　　言・パラ言・非言

真の音声の統計的モデル　～波形の統計的モデル～

- 不特定話者・不特定基本周波数・不特定位相の音響モデル
  - 見たくないものは全て「確率の定義」で集めて隠してしまおう。
  - $P(o|w) \approx \sum_{s,h,p} P(o|w,s,h,p)P(s)P(h)P(p)$
    - s：speaker, h：harmonics, p：phase
  - 一般的な解決策：各手法の組み合わせ
    - 最終的に性能を最大化する組み合わせを追求する。