

性差・年齢差を越えた 音声のモデル化とその応用

峯松 信明

大学院工学系研究科
工学部電子情報工学科



情報・システム工学概論のキーワード

音声を用いた コミュニケーション

抽象化 モデル化 情報技術

情報・システム工学概論のキーワード



本講義の流れ

抽象化・モデル化と音声コミュニケーション技術

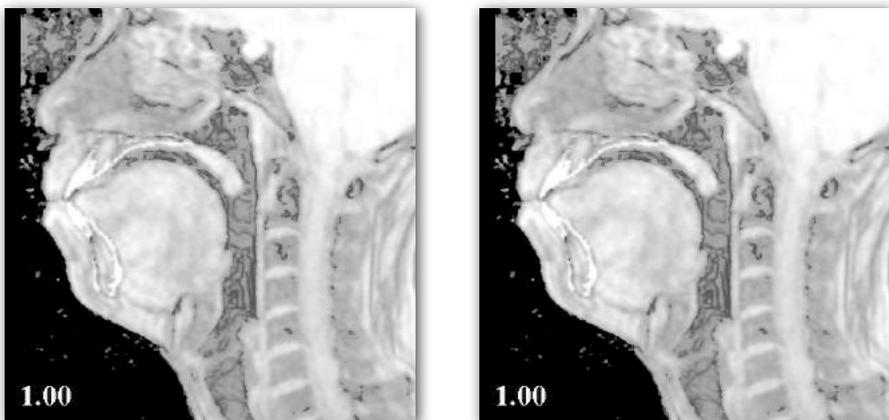
音声の物理学・基礎編

- 音声の生成からその物理的（音響的）側面まで
- 人間が受け取るメディア情報の多様性と不変性
 - 視・聴・触・味・嗅、多様に変形しても同じだと分かる。
- 多様性の「そぎ落とし」と内在する不変性
 - その数学的な回答とそれに基づく音声のモデル化
- 多様性に打ち勝つ音声技術の構築と応用
 - 音声認識， 外国語発音分析を例にとって
- 「そぎ落とし」が気付かせてくれる異分野との接点
 - あれと、あれは、同じ原理なのか？
 - 惑星とリンゴは同じである。音声とXXは同じである。

口の中の隙間を変える！？

でかいのは牛タンだけじゃない！人タンだって！

- しかもこんなにダイナミックに動いちゃいます！



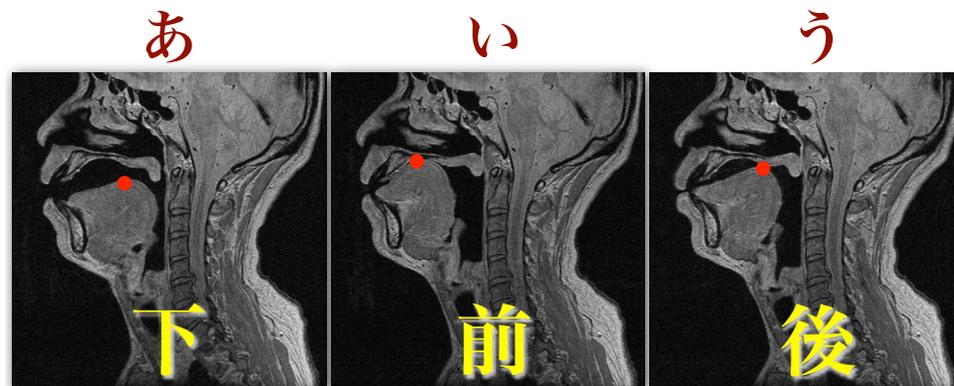
(株)ATR人間情報科学研究所提供

- 「あご」の動き → 「べ口」の動き → 「口の中のすき間」の変形

音声の生成 ～母音～

音声ってどうやって生まれるんだろう？

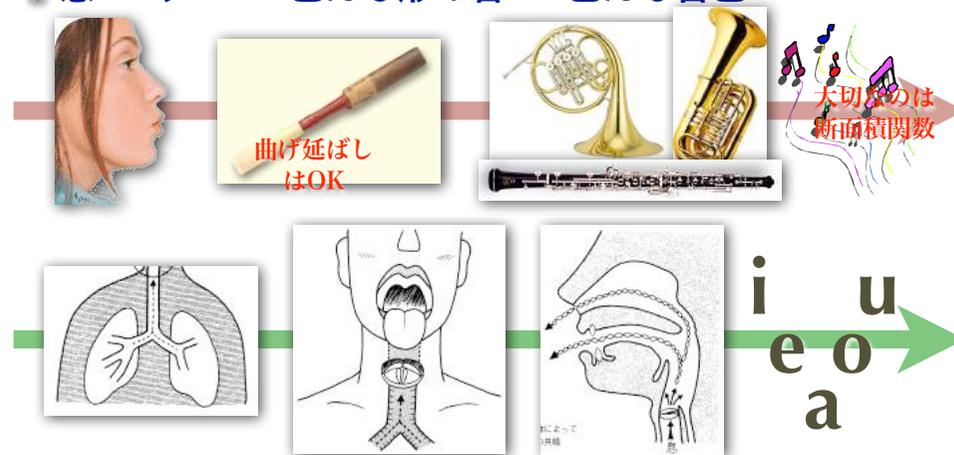
- 「あいうえお」って言ってごらん？何が動いてる？
 - 「口の開き方」つまり「あご」が動いてるのは、分かるよね。
 - 「べ口」が大きく動いているってのは分かるかな？
 - 「べ口」の位置を変えるために「あご」を動かしている、とも言える。



(株)ATR人間情報科学研究所提供

お口は楽器、色々な音をかなでます

息 → ブー → 色々な形の管 → 色々な音色



管の形の違い = 出てくる音の音色の違い

お口は楽器、色んな音をかなでます

息 → ブー → 色んな形の管 → 色んな音色

曲げ延ばしはOK

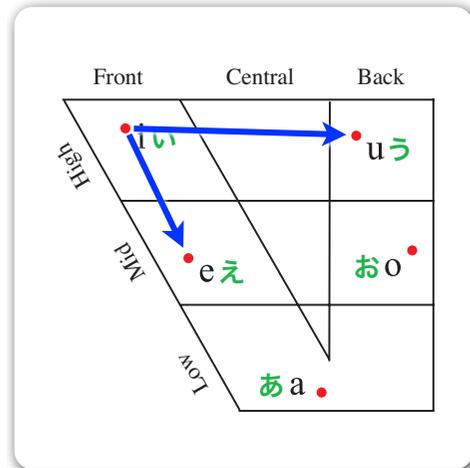
大切なのは断面積関数

図 6-1 (a) 横断線での声道の断面積関数

でも我々の楽器はグニャグニャ形が変わる

日本語とアメリカ英語の母音図

あ=下, い=前, う=後

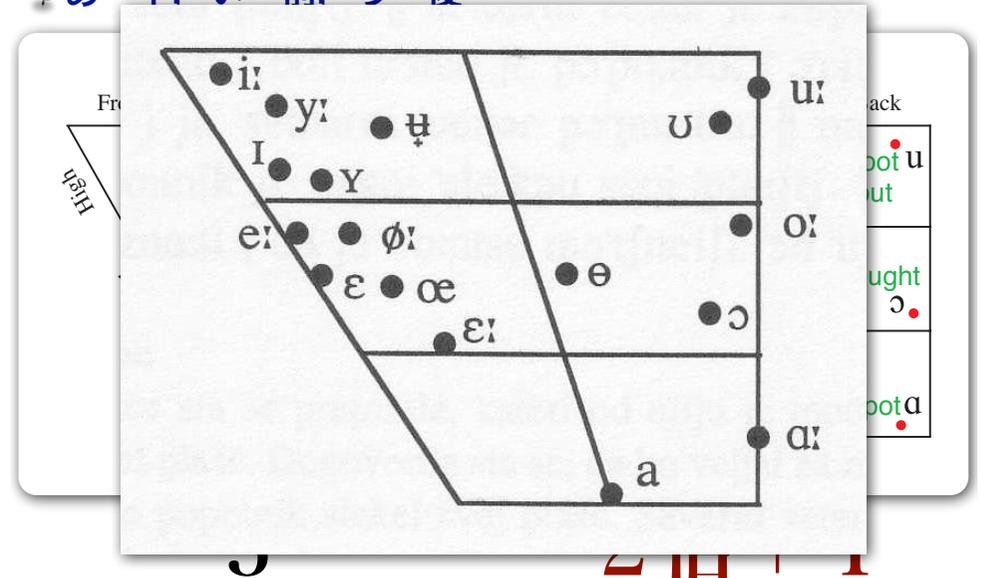


5

声帯音源+喉 = ブザー+管

日本語とアメリカ英語の母音図

あ=下, い=前, う=後

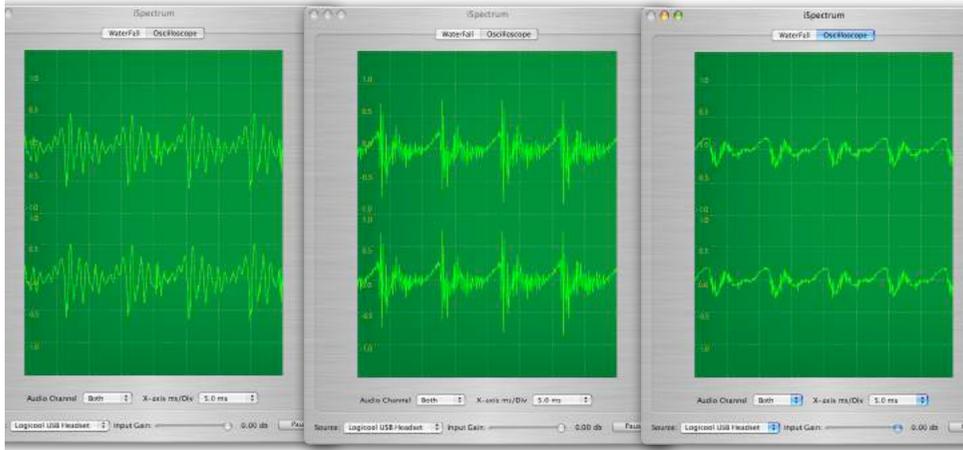


2倍+1

空気振動としての音声 ～波～

じゃ、その空気粒子振動としての声（音）を見る！

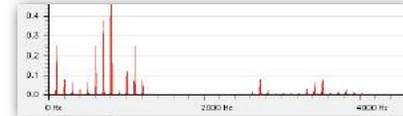
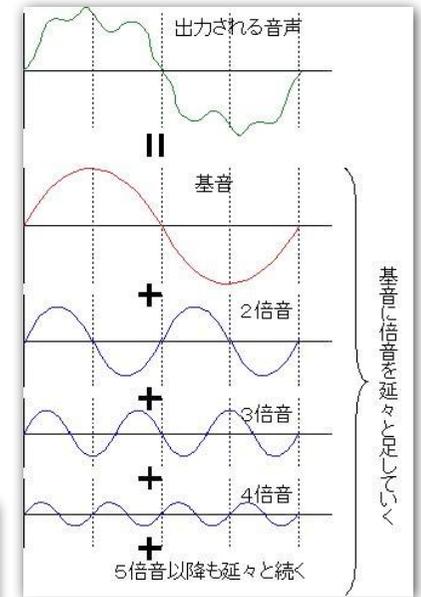
- 波形表示ソフトで見た峯松の「あ」「い」「う」
- 周期的な波形の連続→音の「高さ」を感覚できる



波＝基本的な波の足し合わせ

分解と合成（フーリエ変換）

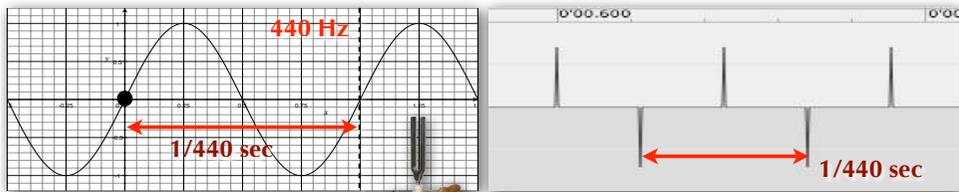
- 波＝基本音＋2倍音＋3倍音＋・・・
- 周波数：振動回数／秒
- 波：これらに適切な重みをかけて足しあわされた結果
- どの周波数の波は強く、どの周波数の波は弱いのか？
- どの周波数のエネルギーは強く、どの周波数のエネルギーは弱いのか？
- 横軸周波数、縦軸エネルギー強度としてグラフを書くと・・・
- それがスペクトル



波とそのスペクトルの例

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

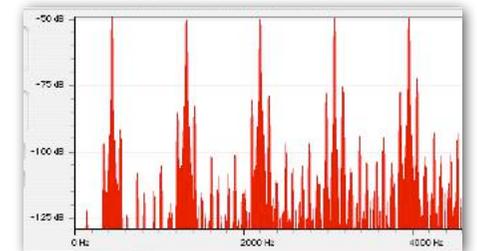
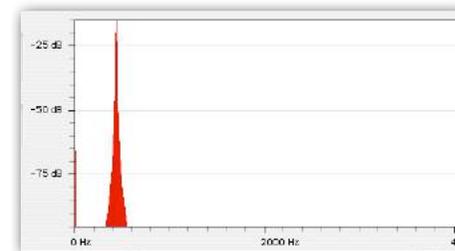
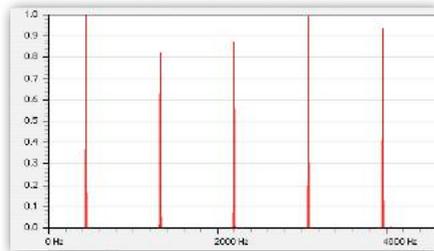
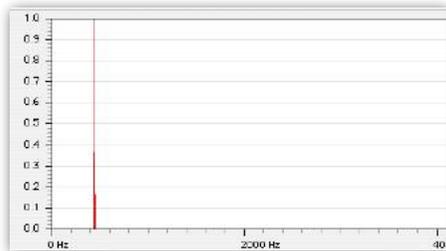
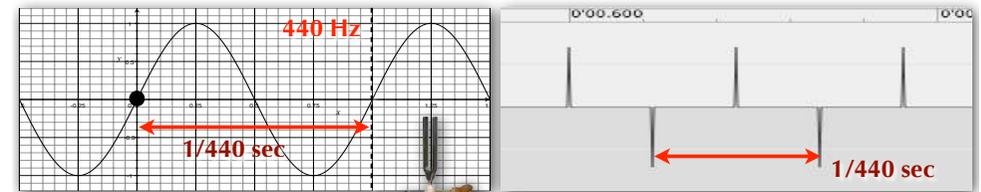
● w_1 基本波＋ w_2 第二高調波＋ w_3 第三高調波＋ w_4 第四高調波＋



スペクトル→対数スペクトル

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

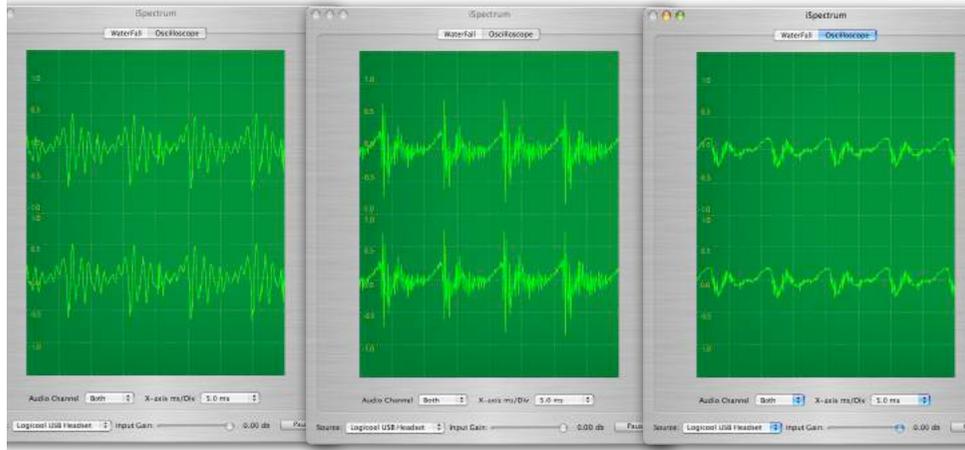
● w_1 基本波＋ w_2 第二高調波＋ w_3 第三高調波＋ w_4 第四高調波＋



空気振動としての音声 ～波～

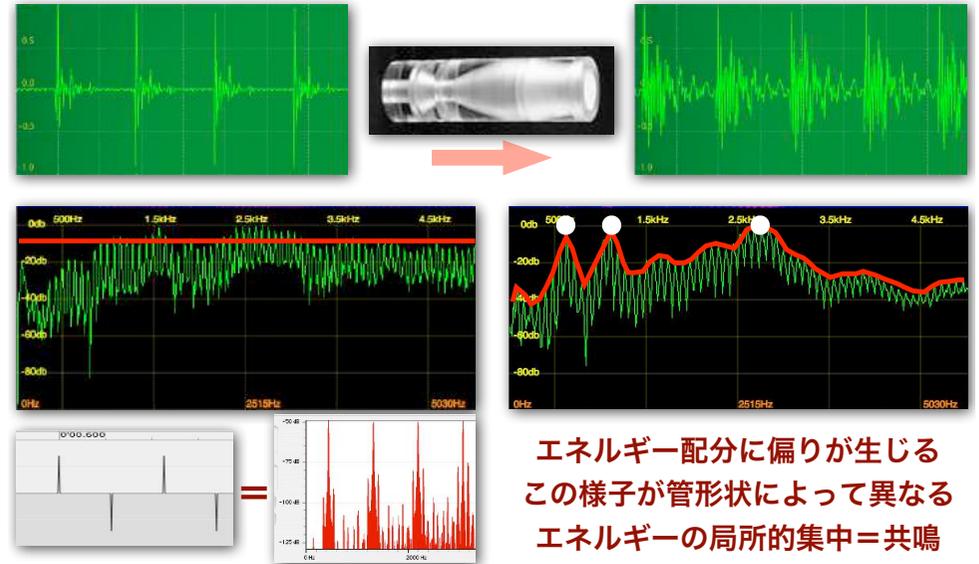
じゃ、その空気粒子振動としての声（音）を見る！

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



音声の対数スペクトル

ブザー（パルス列）→管→「あ～」



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

共振・共鳴現象が生まれる仕組み

母音は疎密波, , でも、ちょっと特殊な疎密波

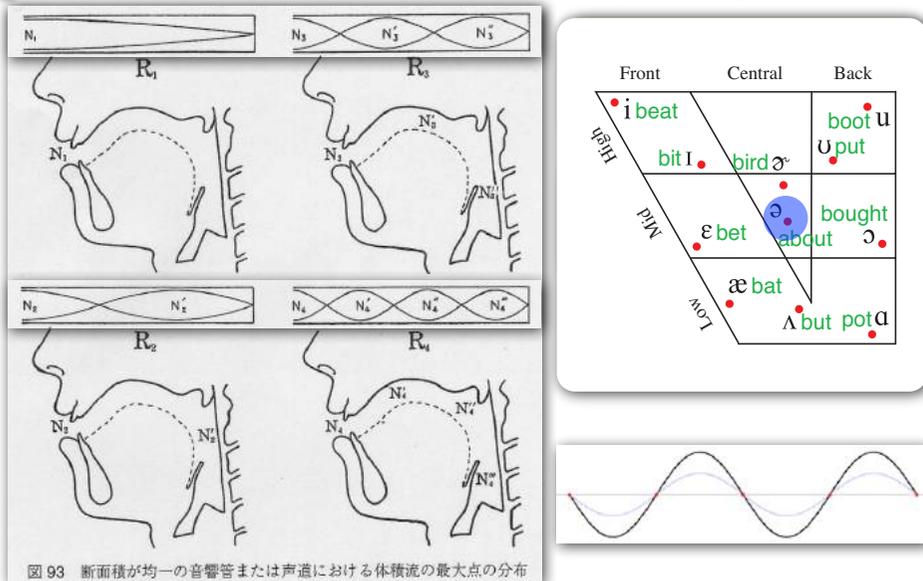
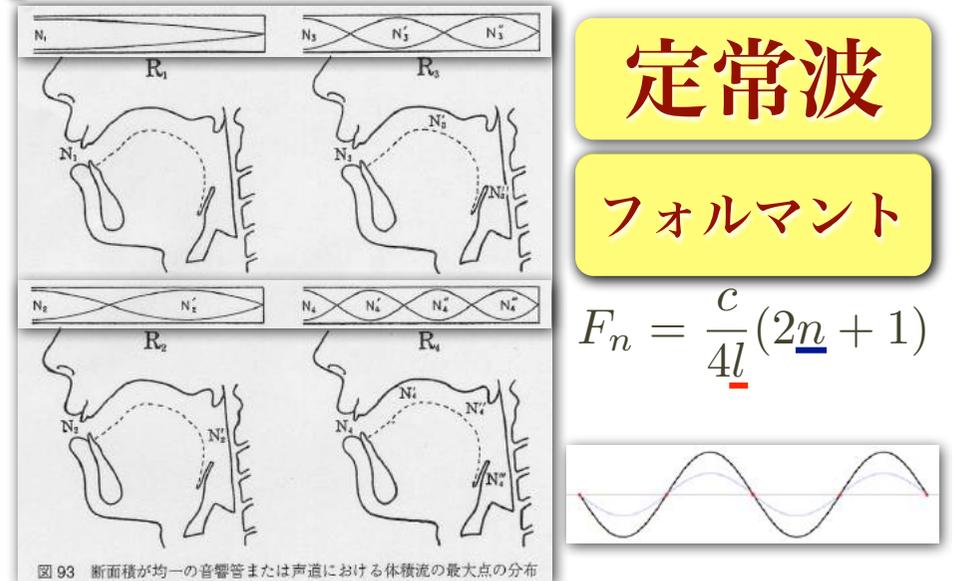


図 93 断面積が均一の音響管または声道における体積流の最大点の分布

共振・共鳴現象が生まれる仕組み

母音は疎密波, , でも、ちょっと特殊な疎密波



$$F_n = \frac{c}{4l}(2n + 1)$$

図 93 断面積が均一の音響管または声道における体積流の最大点の分布

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

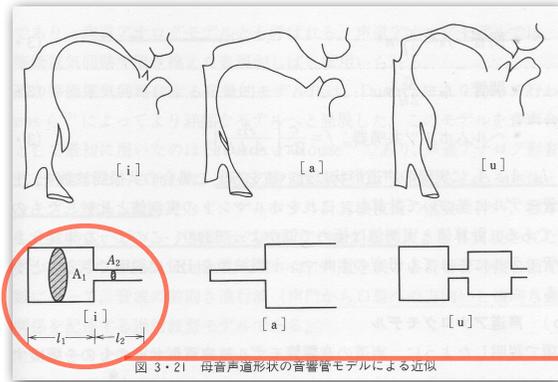


図 3・21 母音声道形状の音響管モデルによる近似

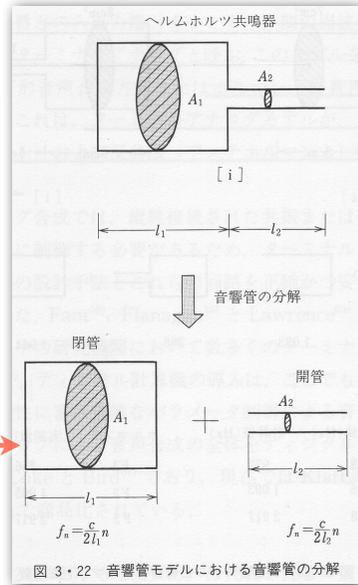


図 3・22 音響管モデルにおける音響管の分解

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

共振周波数と基本周波数

周波数(frequency) = 頻度(frequency)

一秒間当たり（その出来事が）何回起こるのか？

共振周波数：喉の形状によって決まる → 音色

喉の形状が決める周波数の定常波だけが残る、他は減衰する。

500 Hz = 1秒間に500回振動する定常波

第一フォルマント・第二フォルマント・・・

各フォルマント周波数の間隔は色々

基本周波数：声帯の振動頻度によって決まる → ピッチ

声帯の締め具合によって声帯の振動回数は変わる。

150 Hz = 1秒間に150回声帯が振動する。

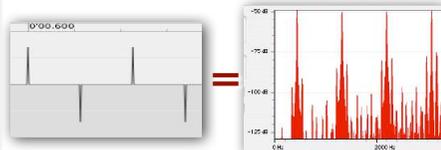
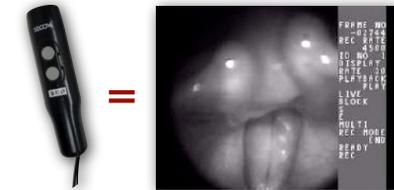
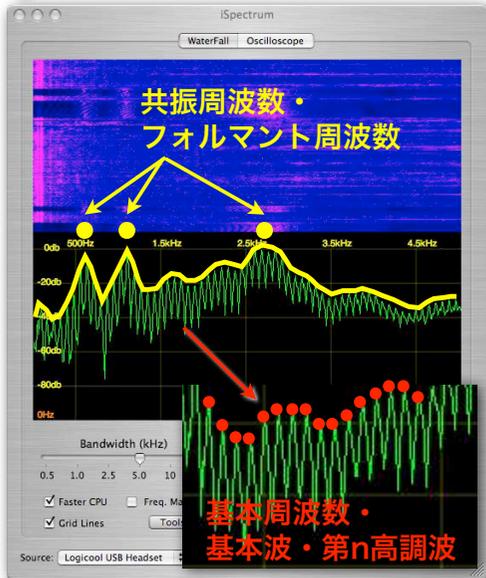
これは、150回振動する波を生むのでは？それは喉で減衰するの？

基本波・第二高調波・第三高調波・・・

高調波は一定の間隔で並ぶ。

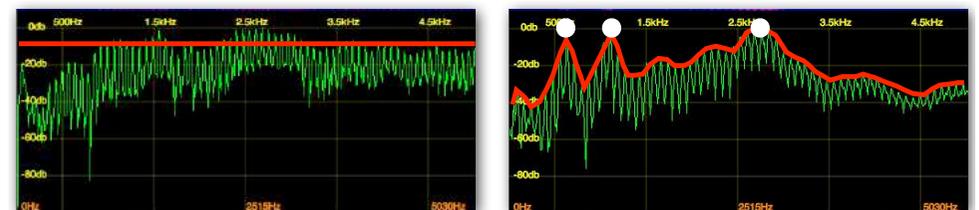
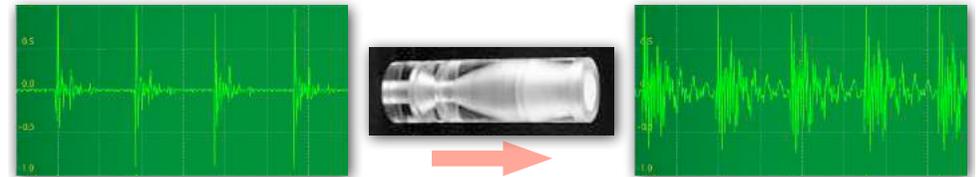
共振周波数と基本周波数

リアルタイム音声分析ソフトで声を「見て」みる



音声の対数スペクトル

ブザー（パルス列） → 管 → 「あ～」

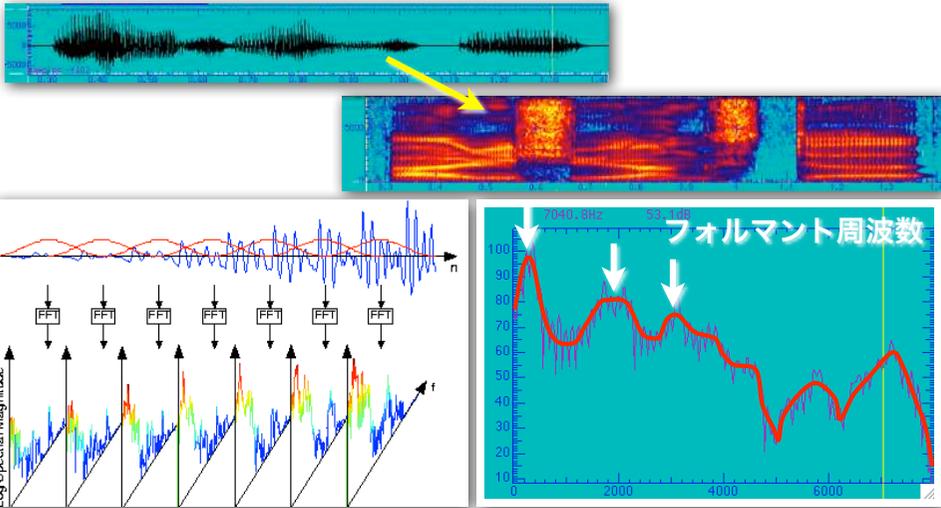


エネルギー配分に偏りが生じる
この様子が管形状によって異なる
エネルギーの局所的集中 = 共鳴

連続的に変化する音声の音響分析

音声波形からスペクトル系列へ

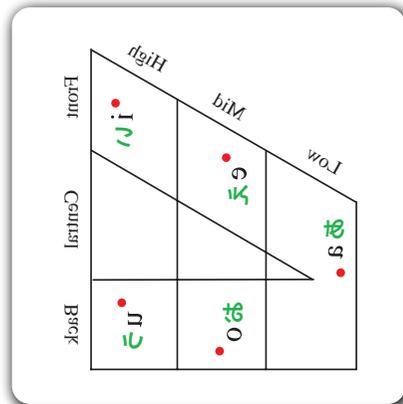
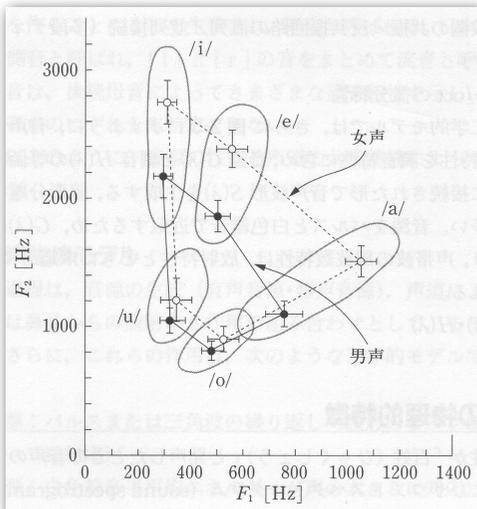
窓掛け→短時間フーリエ変換→対数パワースペクトル系列



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



耳から脳へ 音楽の通り道

耳から脳へ 音楽の通り道

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

1つの楽器も、複数の周波数が複雑に組み合わさったものだ

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

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

基底膜の振動は場所によって異なる周波数に対応している

聴神経線維はそれぞれの周波数に対応した神経線維になっている

聴覚野

質問 1

Q: 管を短くするとどんな声になるのか?

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

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

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

図 3-21 母音声道形状の音響管モデルによる近似

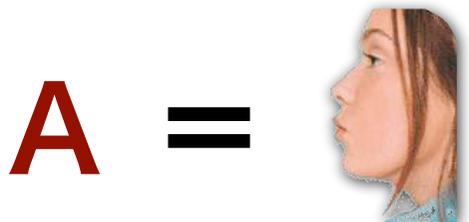
図 3-22 音響管モデルにおける音響管の分解

A → **B** → **C**

質問2

Q: 管を「あ」から何の形に変えたのか当てなさい。

この人の「あ」の管の形から変えていきます。



Replay

本講義の流れ

抽象化・モデル化と音声コミュニケーション技術

音声の物理学・基礎編

音声の生成からその物理的（音響的）側面まで

人間が受け取るメディア情報の多様性と不変性

視・聴・触・味・嗅, 多様に変形しても同じだと分かる。

多様性の「そぎ落とし」と内在する不変性

その数学的な回答とそれに基づく音声のモデル化

多様性に打ち勝つ音声技術の構築と応用

音声認識, 外国語発音分析を例にとって

「そぎ落とし」が気付かせてくれる異分野との接点

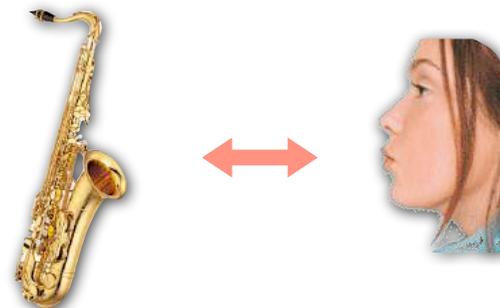
あれと、あれは、同じ原理なのか？

惑星とリンゴは同じである。音声とXXは同じである。

質問2

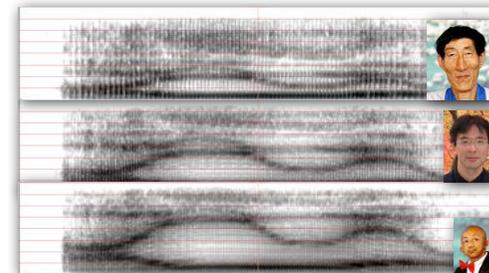
Q: 管を「あ」から何の形に変えたのか当てなさい。

この人の「あ」の管の形から変えていきます。



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

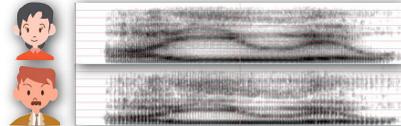
音高の変化

- 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

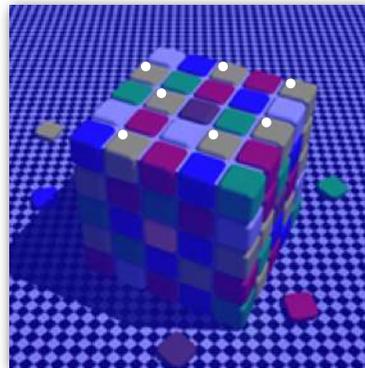
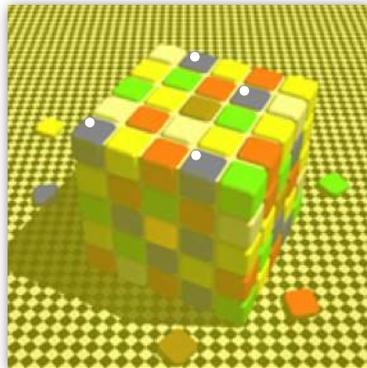
- 男性のおはよう！と女性のおはよう！
- 大人のおはよう！と子供のおはよう！



でも、我々は容易に「同一性」を認知できる

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



- 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

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

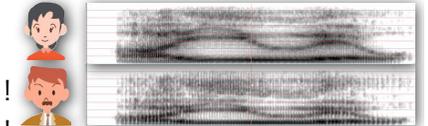


色みの変化

- 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



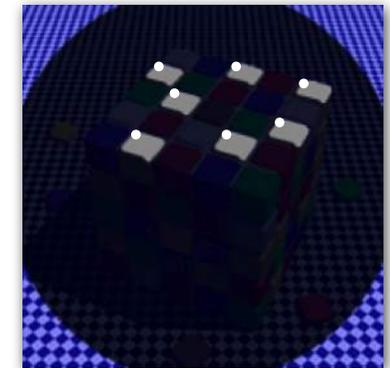
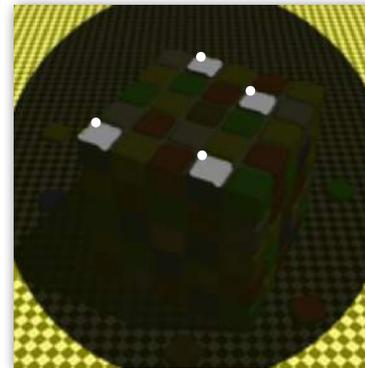
音色の変化

- 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

色みの偏差とその認知的不変性

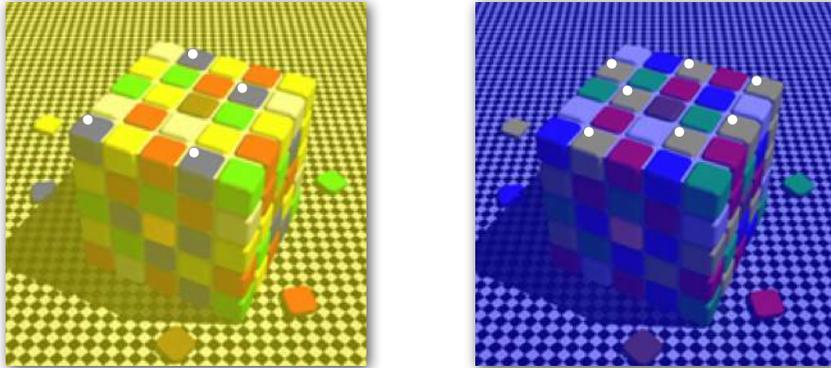
黄・青眼鏡を通して眺めるルービックキューブ



- 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



- 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

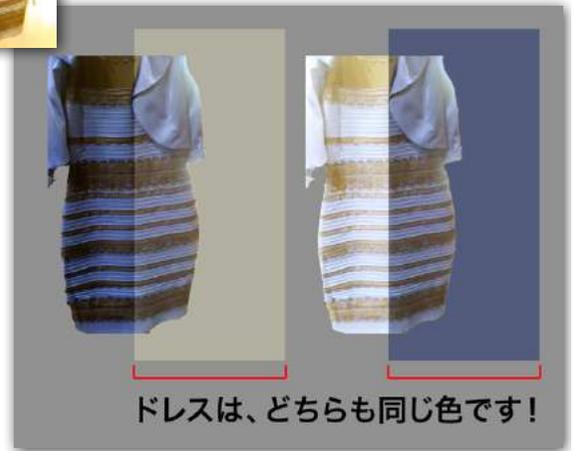
音高の偏差とその認知的不変性

カラオケでキーを上げ下げして曲を聞く



- 絶対音感者（ドレミは音名）
 - 1 = ソーミソドラードドソー, 2 = レーシレソーミーソソラー
- 言語化可能な相対音感者（ドレミは階名）
 - 1 = ソーミソドラードドソー, 2 = ソーミソラードドソー
- 言語化困難な相対音感者（ラーラ音感者）
 - 1 = ラーララーラーラーラー, 2 = ラーララーラーラーラー
- 異なる音を同一と主張し、同一の音を異なると主張する。
- 各音を持つ基本周波数（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

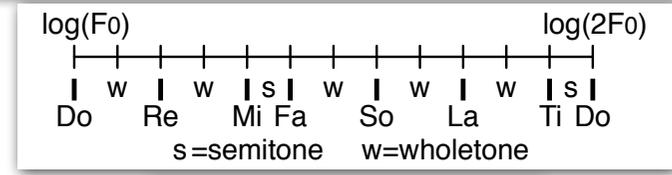
これ、覚えてますか？



ドレスは、どちらも同じ色です！

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



- 各音を持つ基本周波数（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



但し、孤立音の同定は不可能
コンテキストの欠損=コントラストの欠損



各音を持つ基本周波数（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

音高の恒常的・不変的認知はどこまで遡れるのか？

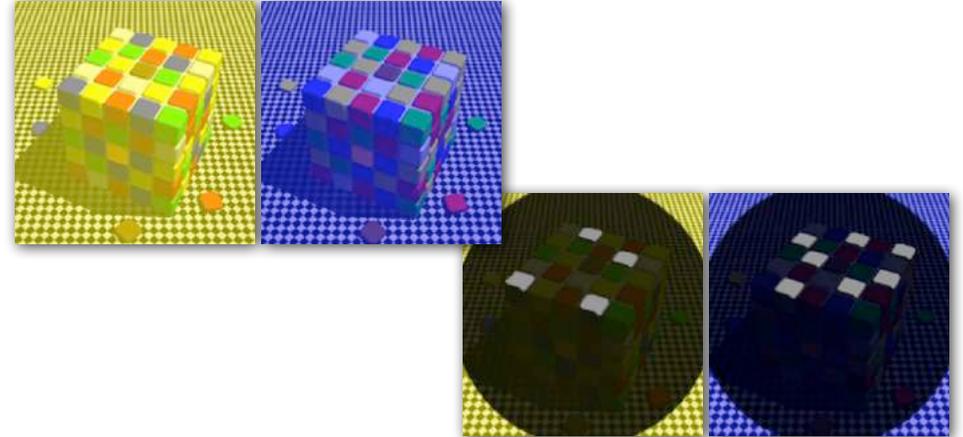


1 = 2



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？



本講義の流れ

抽象化・モデル化と音声コミュニケーション技術

音声の物理学・基礎編

● 音声の生成からその物理的（音響的）側面まで

人間が受け取るメディア情報の多様性と不変性

● 視・聴・触・味・嗅、多様に変形しても同じだと分かる。

● 多様性の「そぎ落とし」と内在する不変性

● その数学的な回答とそれに基づく音声のモデル化

多様性に打ち勝つ音声技術の構築と応用

● 音声認識、外国語発音分析を例にとって

「そぎ落とし」が気付かせてくれる異分野との接点

● あれと、あれは、同じ原理なのか？

● 惑星とリンゴは同じである。音声とXXは同じである。

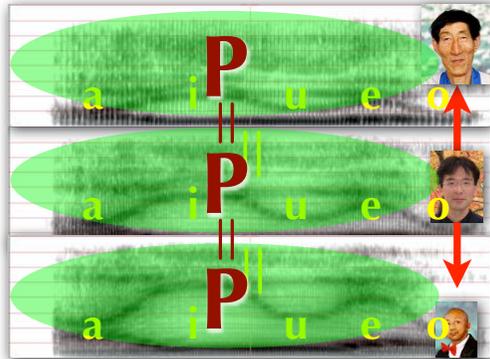
音色の偏差とその認知的不変性

音高の個人差を生み出す要因 ～高い低い～

- 男女の音高偏差 = 声帯の長さ・重さの性差

音色の個人差を生み出す要因 ～太い細い～

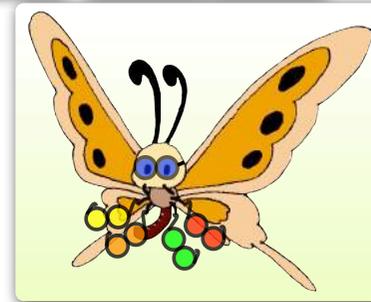
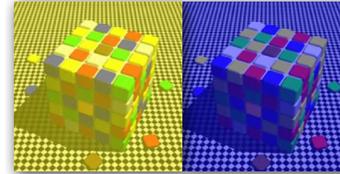
- 男女の音色偏差 = 声道の形状 (主に長さ) の性差



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

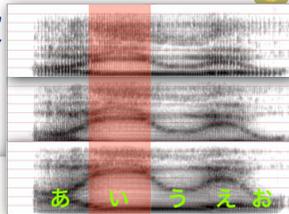
- コントラスト情報に基づく処理が重要
- コントラスト群から成る全体のパターン処理が要素同定を可能



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。



音声模倣の二面性 ～音真似と?真似～

音声模倣 = 親の発声行為を子が積極的に模倣する行為

- これを通して幼児は言語を獲得する。
- 動物学的には非常に稀な行為。霊長類では人間だけ。
- 他の動物では小鳥、クジラ、イルカくらいか。
- 動物の模倣 = 声帯模写, ヒトの音声模倣 ≠ 声帯模写
 - 九官鳥の音声模倣
 - 車, ドア, 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
 - いくら良い子でも, 声から父親を割り出せずにお巡りさんは困る。



音声模倣の二面性 ～音真似と？真似～

親の発声 → 音韻同定 → 音韻列 → 個々の音韻を発声？



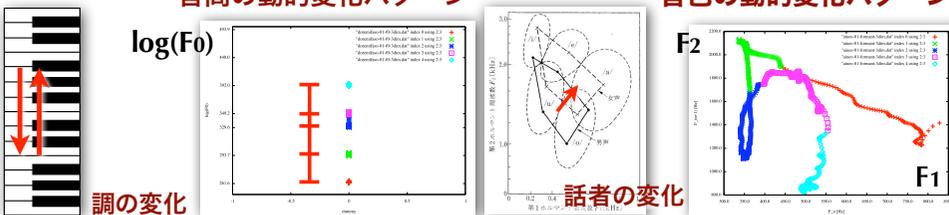
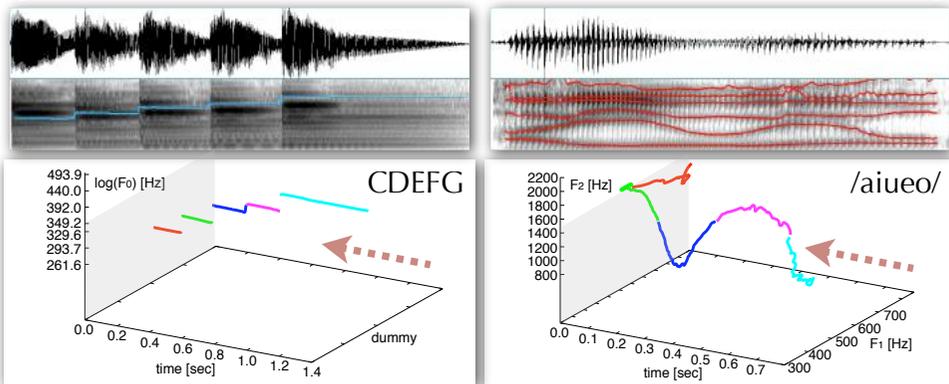
音韻意識（仮名の意識）が希薄／しり取りも出来ない。

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？



音声の構造的表象／音色の相対音感



音色の偏差とその認知的不変性

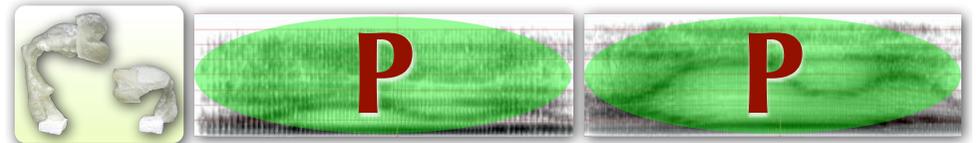
色み・音高の恒常・不変的認知

- 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

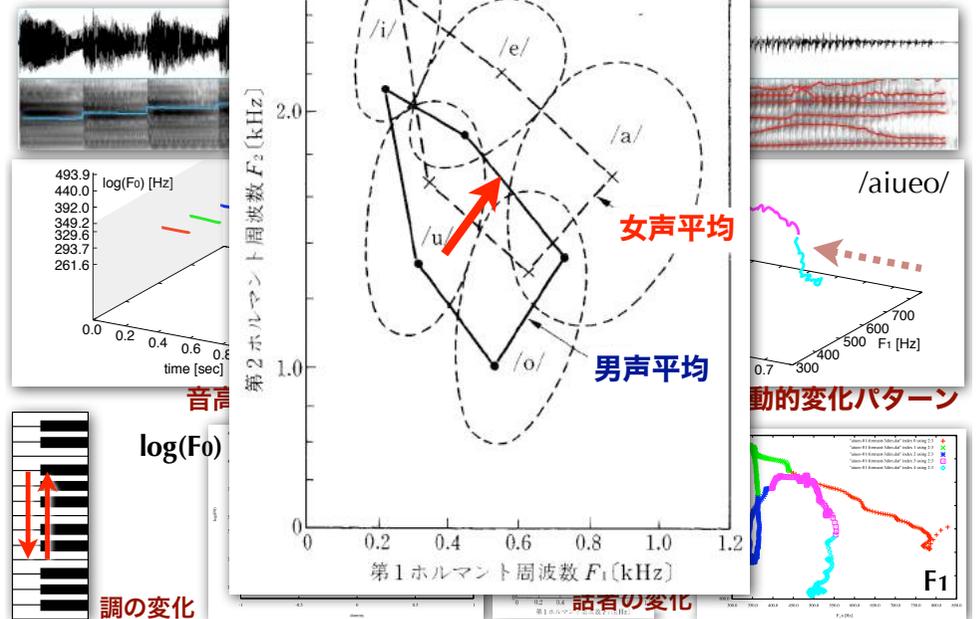


音色の恒常・不変的認知

- 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

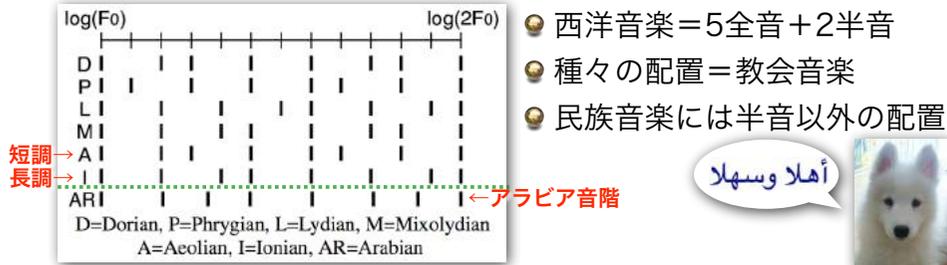


音声の構造的表象（音色の相対音感）



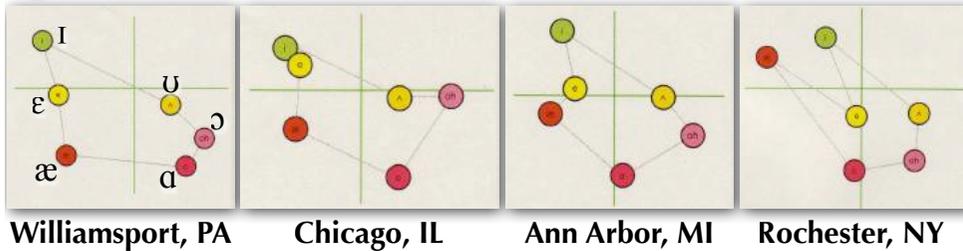
音声の構造的表象 / 音色の相対音感

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



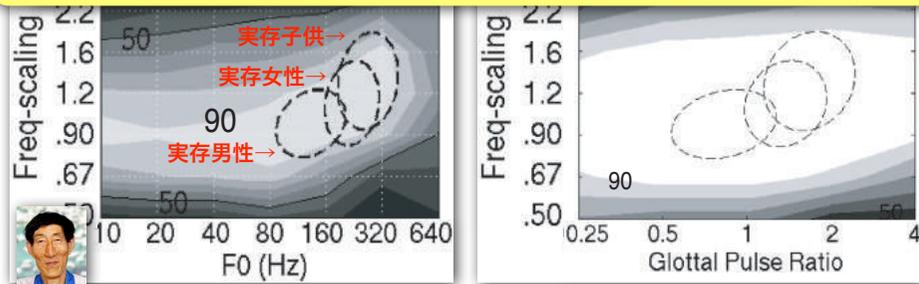
音声の構造的表象 / 音色の相対音感

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



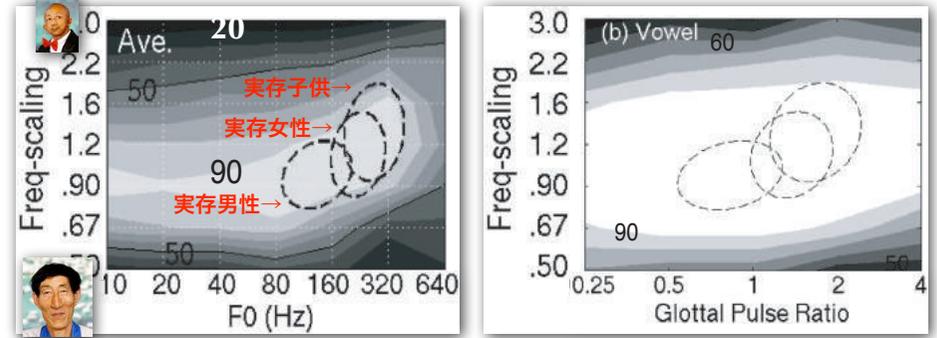
音声の構造的表象 / 音色の相対音感

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。
- 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。



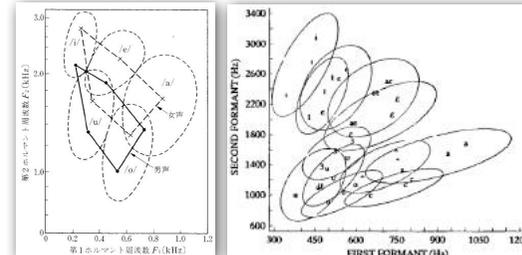
音声の構造的表象 / 音色の相対音感

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、読み書きに苦勞する人が多く存在しなければならない？

音声の構造

言語化困難な相対

次に示すメロディー・メロディーを提示

メロディー・メロディーを提示

言語化

次に示すメロディー・メロディーを提示

メロディー・メロディーを提示



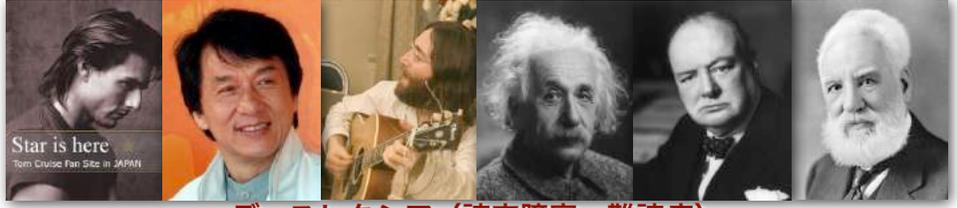
その後、別のメロディーを提示

メロディー・メロディーを提示

言語化

次に示すメロディー・メロディーを提示

メロディー・メロディーを提示



ディスレクシア (読字障害・難読症)

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・・・

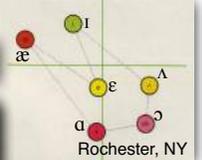
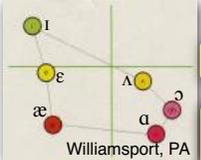
- 一人ずつ引き取られた。
- 彼らは10年後どんな発音をしているのだろうか？



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
幼児が無視するのは無視する機械



情報・システム工学概論のキーワード



情報・システム工学概論のキーワード



賢馬ハンスから学べること



抽象化 モデル化 情報技術

本講義の流れ

抽象化・モデル化と音声コミュニケーション技術

音声の物理学・基礎編

- 音声の生成からその物理的（音響的）側面まで

人間が受け取るメディア情報の多様性と不変性

- 視・聴・触・味・嗅，多様に変形しても同じだと分かる。

多様性の「そぎ落とし」と内在する不変性

- その数学的な回答とそれに基づく音声のモデル化

多様性に打ち勝つ音声技術の構築と応用

- 音声認識，外国語発音分析を例にとって

「そぎ落とし」が気付かせてくれる異分野との接点

- あれと，あれは，同じ原理なのか？

- 惑星とリンゴは同じである。音声とXXは同じである。