

音声インタフェースを備えた ロボットの作成

～音声インタフェースの構築～

峯松 信明

東京大学工学部電子情報工学科



本実験の目的と流れ

本実験の目的

- 音声インタフェースを備えたロボットの構築を通して下記を習得
 - Cプログラミング, 信号処理, 音声認識, マイコンプログラミング, 無線通信, ロボット (ハードウェア) 製作などなど
 - 実験報告書執筆やプレゼンテーションのお作法

本実験の流れ

- 第1日～5日：音声インタフェースの構築
 - 各自で課題を行う。班を構成することは基本的に無い。
- 第6日～10日：ロボットの製作と音声インタフェースとの統合
 - 四人一組で班を構成する。各班で一台ロボットを製作する。
 - ロボットと音声インタフェースの統合時には、役割分担した方が良い？
- 第11日：参加全班で障害物競走（予定）
- 第12日：参加全班で発表会

およその最終イメージ

音声認識＋無線通信＋ロボット

- 孤立単語音声認識を走らせる。入力された単語音声を、**指定した単語セット（辞書）内の何れかの単語として同定する。**
- 音声信号をテキストへ変換する。「右」「進め」「回れ」など
- 認識**単語**を**ロボットコマンド**に翻訳して Bluetooth で送る。
- 単語とコマンドは同一である必要はない。
- 受信したコマンドに従って動作するようにロボットを製作する。
- 認識できる単語もロボットコマンドも自由に定義できる。
- 最終的な目的に応じて、種々の**ロボット要素動作**を自分で考える。



Bluetooth



音声信号処理の基礎

峯松 信明

東京大学工学部電子情報工学科



音声の生成 ～母音～

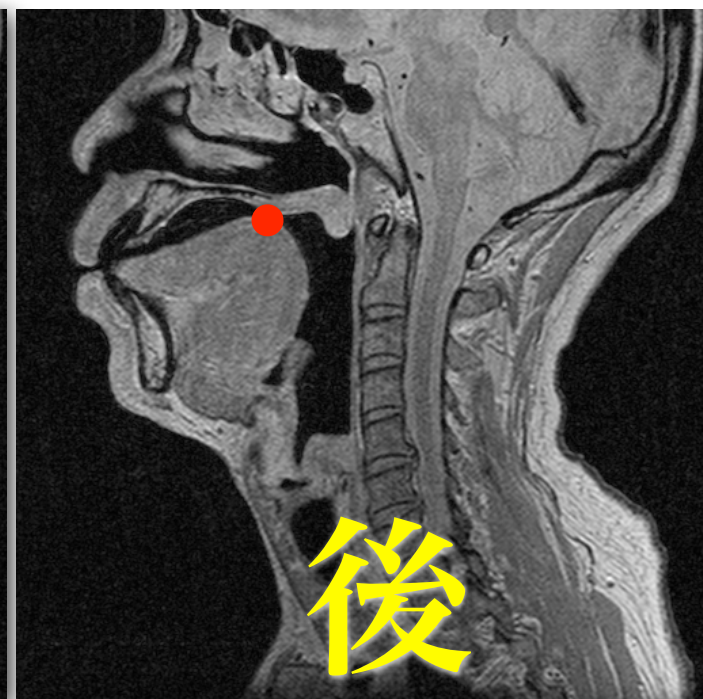
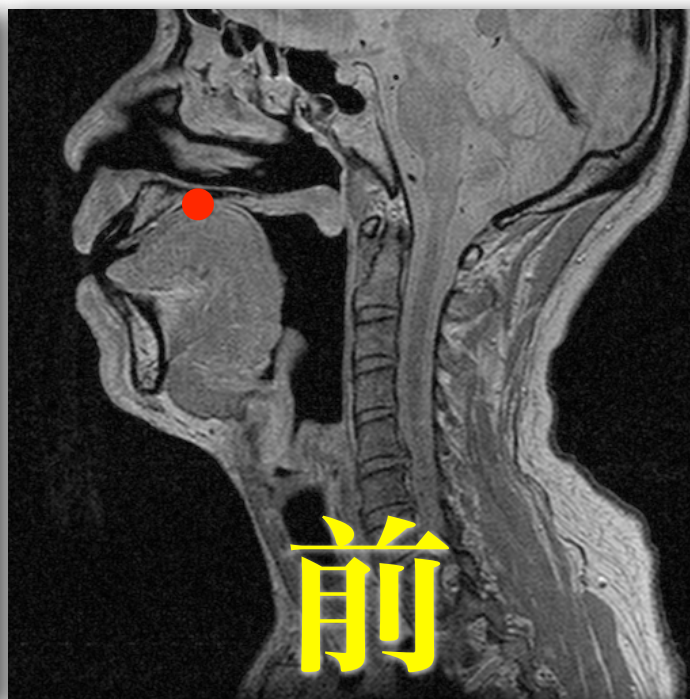
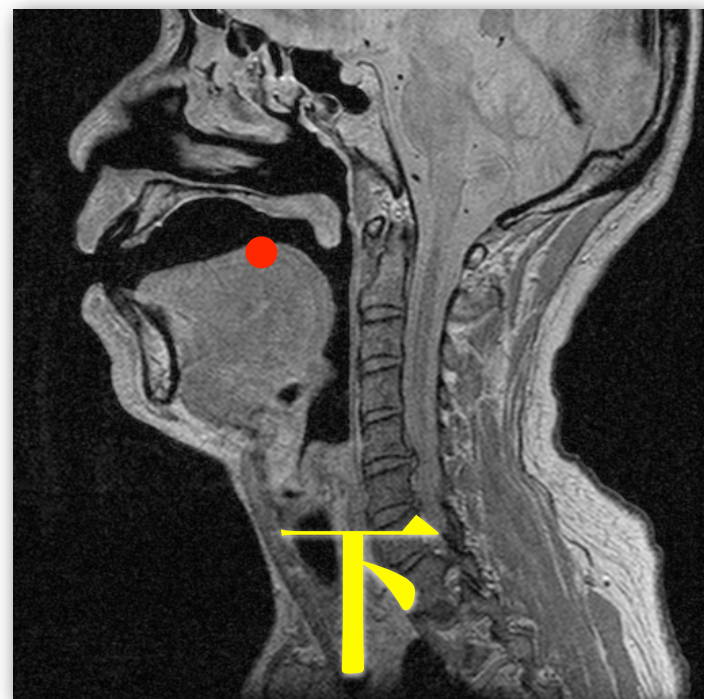
音声ってどうやって生まれるんだろう？

- 「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。
 - 「ベロ」が大きく動いているってのは分かるかな？
 - 「ベロ」の位置を変えるために「あご」を動かしている、とも言える。

あ

い

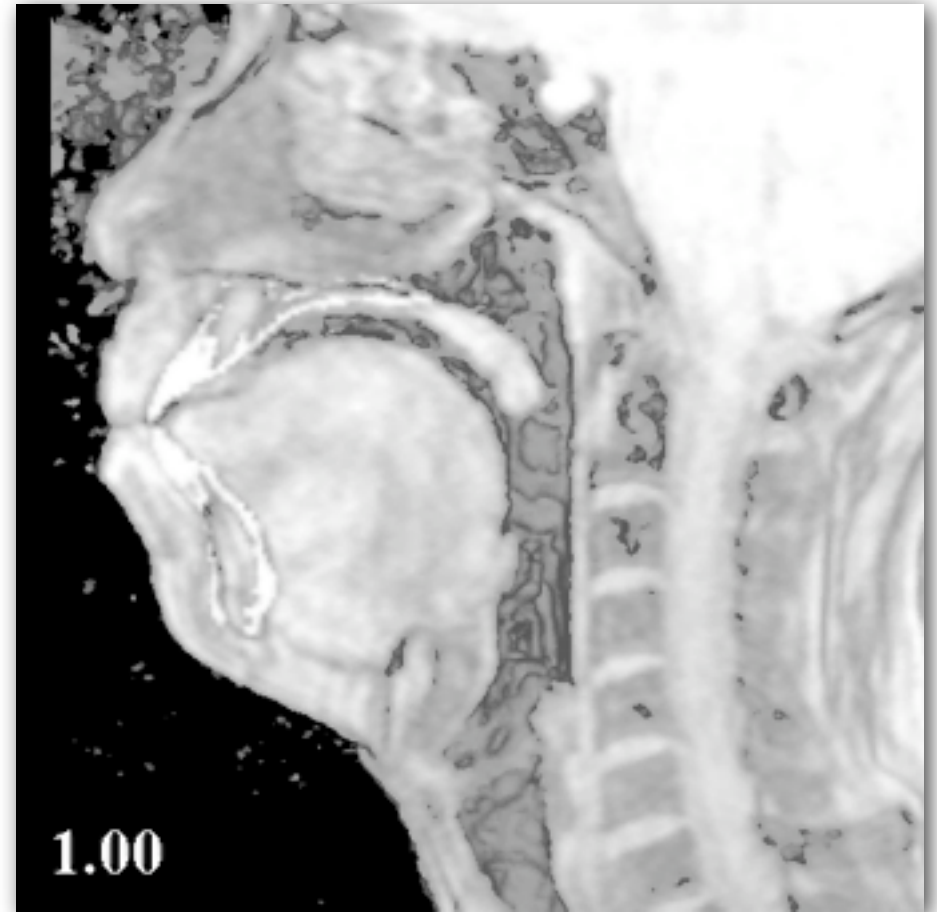
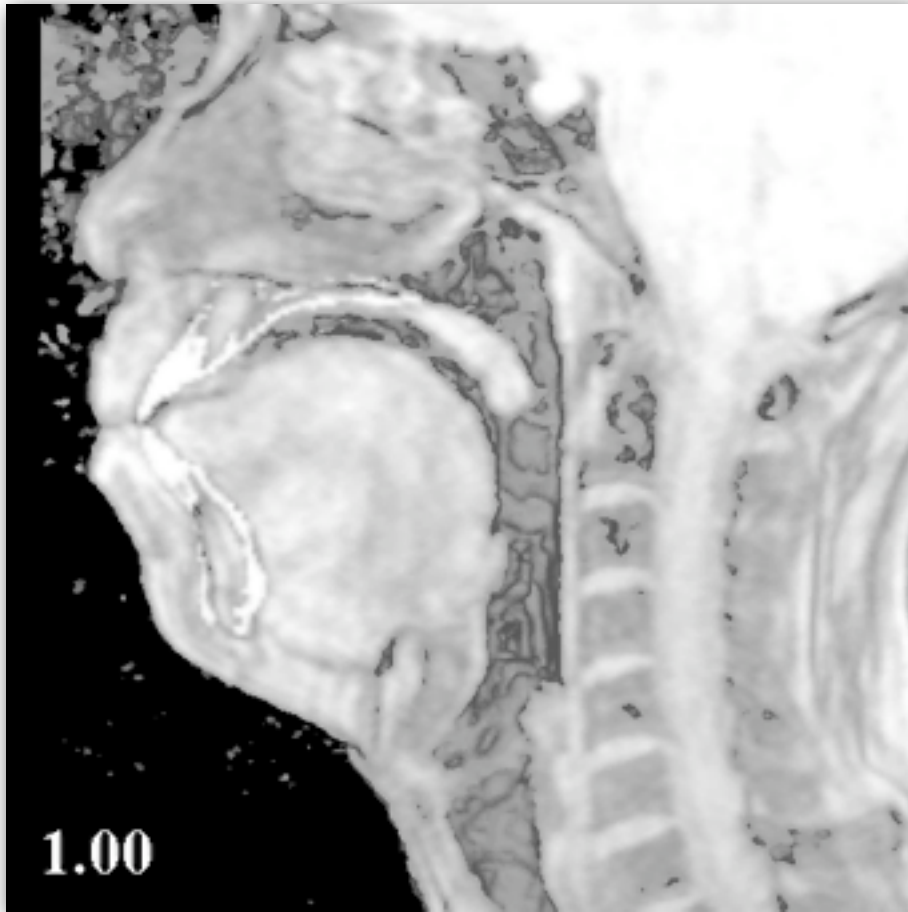
う



口の中の隙間を変える!?

でかいのは牛タンだけじゃない！人タンだって！

● しかもこんなにダイナミックに動いちゃいます！

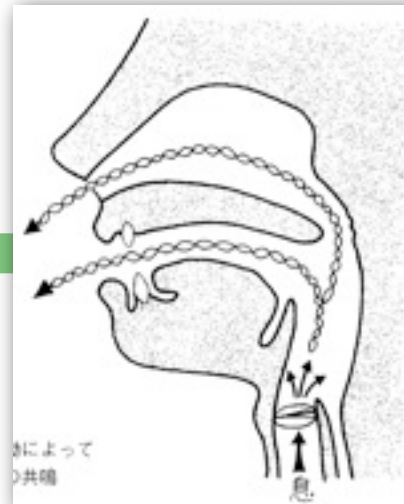
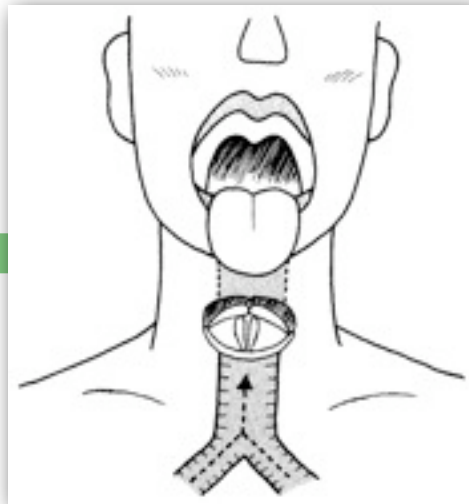
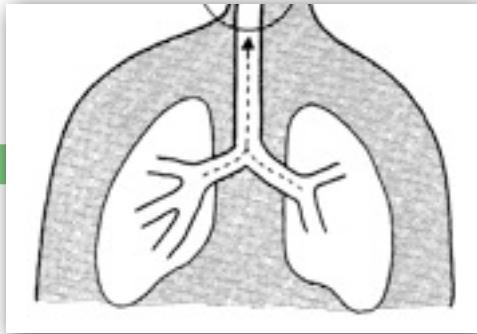
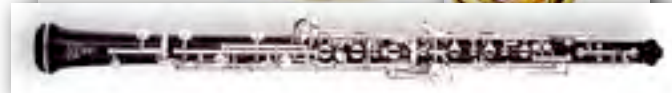


(株)ATR人間情報科学研究所提供

● 「あご」の動き→「ベロ」の動き→「口の中のすき間」の変形

お口は楽器, 色々な音をかなでます

息 → ブー → 色々な形の管 → 色々な音色



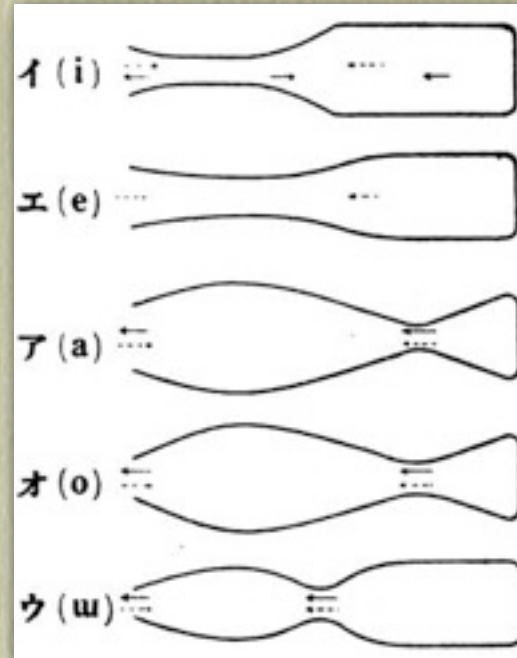
i u
e o
a

管の形の違い = 出てくる音の音色の違い

お口は楽器、 色々な音をかなでます

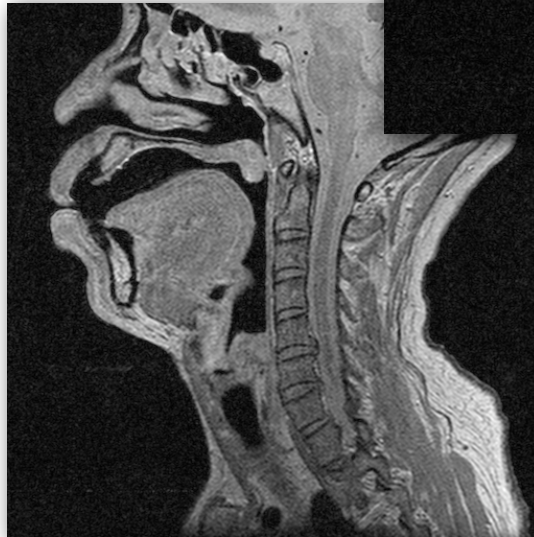
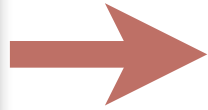
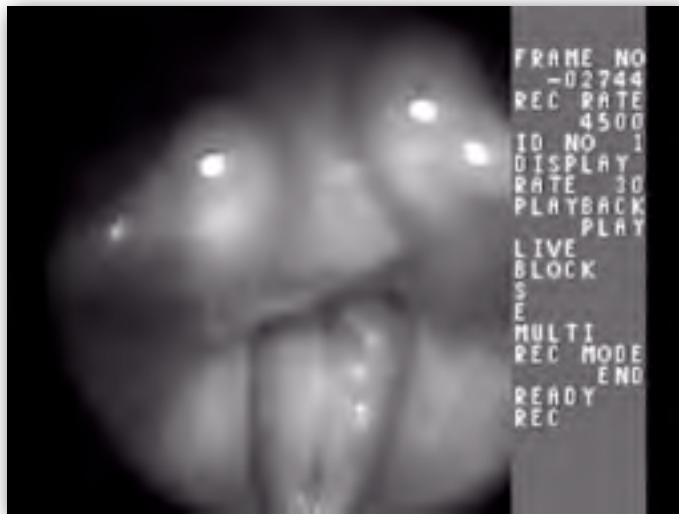
息 → ブー → 色々な形の管 → 色々な音色

曲げ延ばし
はOK

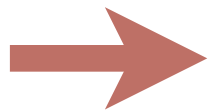


でも我々の楽器はグニャグニャ形が変わる

声帯音源+喉 = ブザー+管

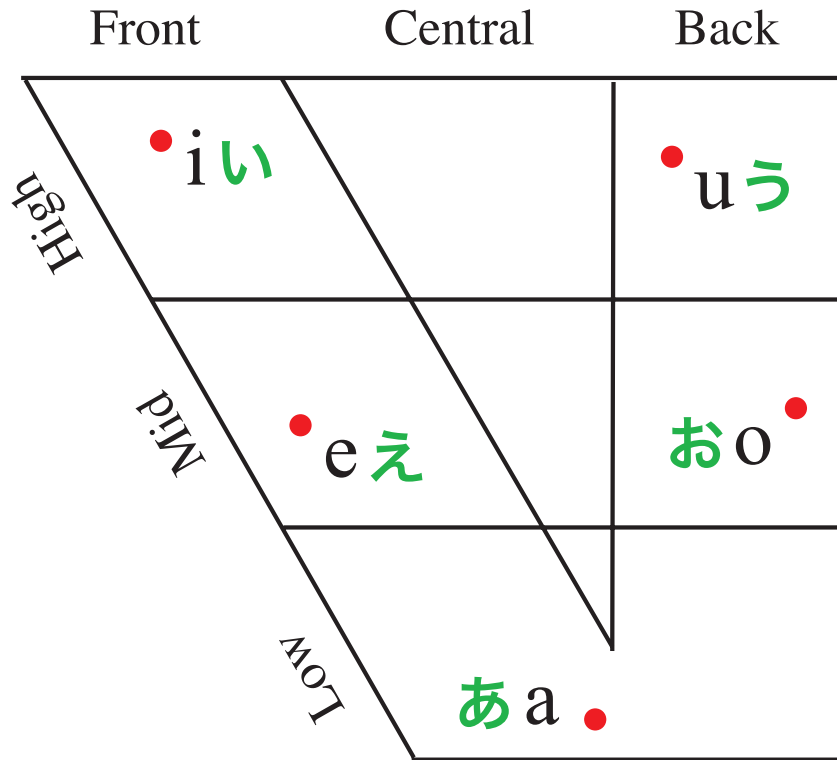


声帯音源 + 喉 = ブザー + 管

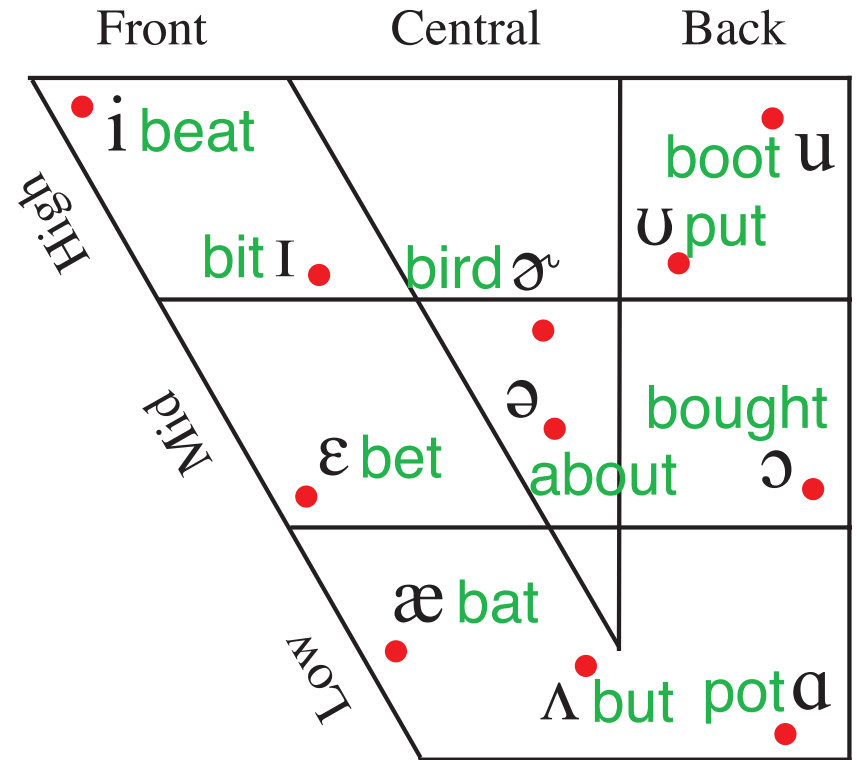


日本語とアメリカ英語の母音図

あ=下, い=前, う=後



5

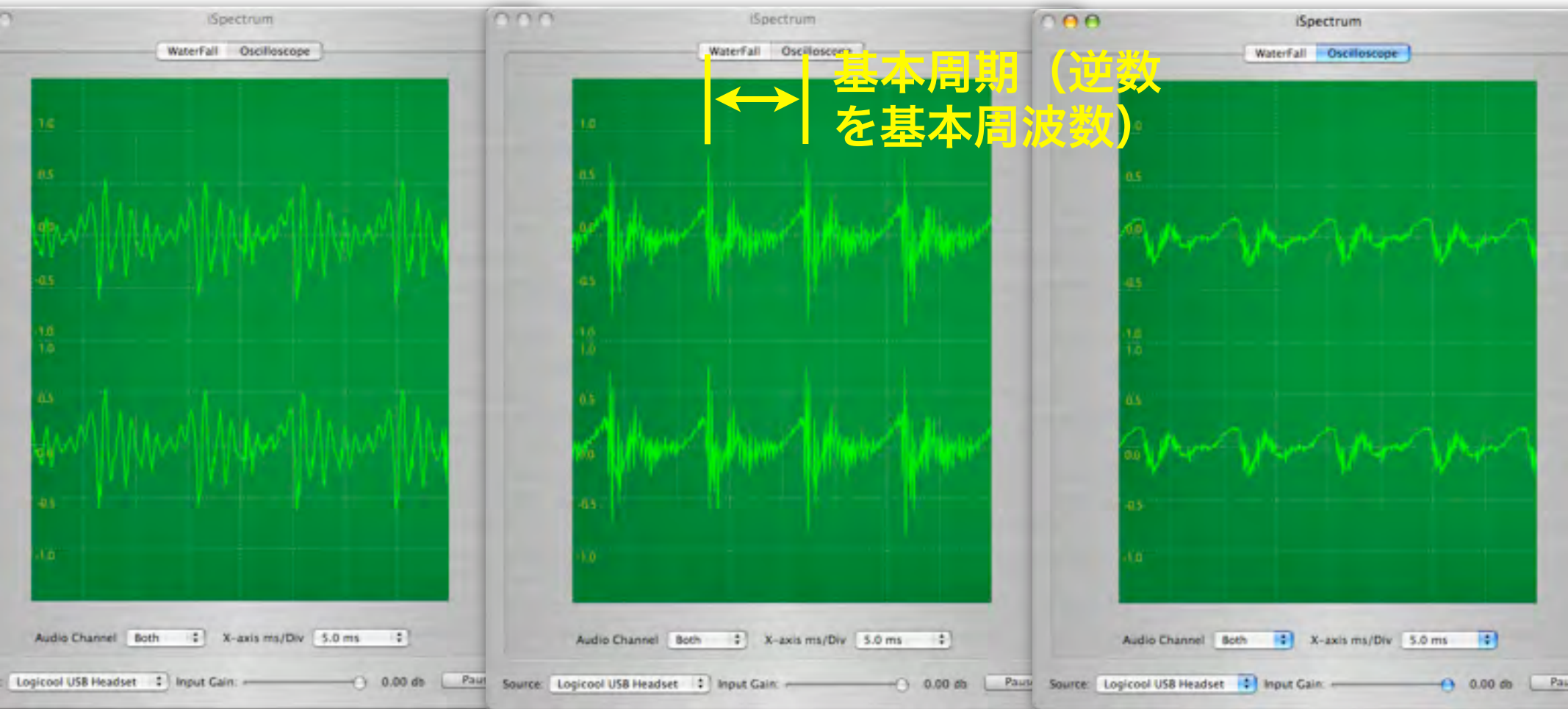


2倍+1

空気振動としての音声 ～波～

じゃ、その空気粒子振動としての声（音）を見る！

iSpectrum (on Macintosh) で見た峯松の「あ」「い」「う」

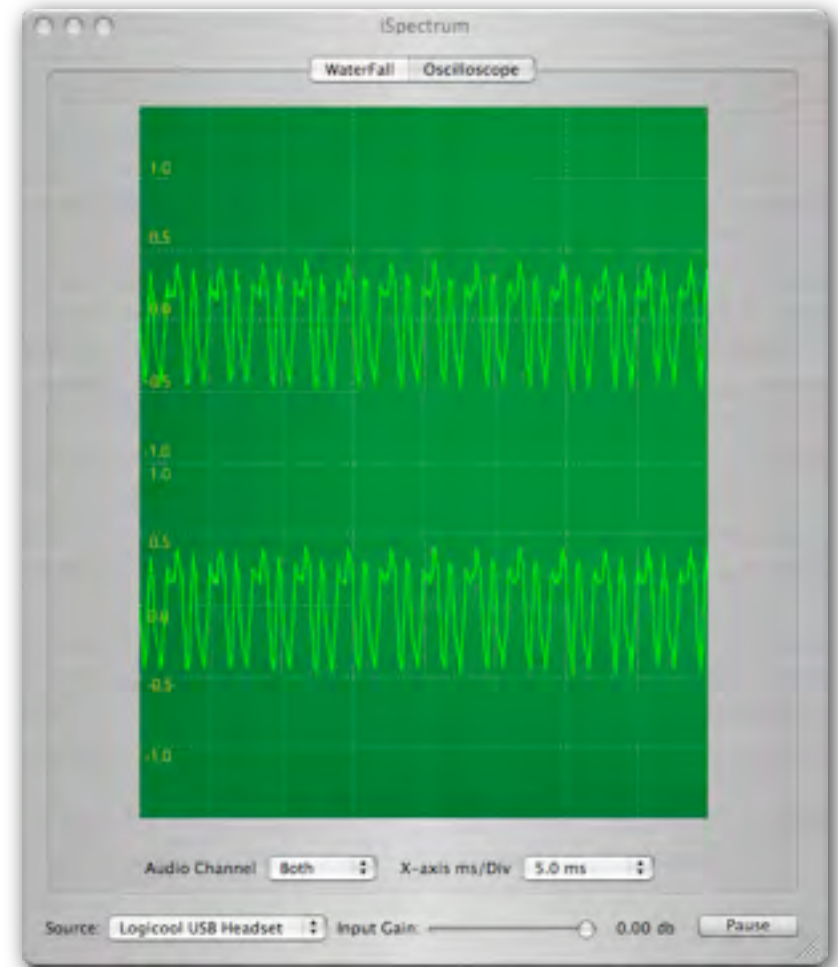
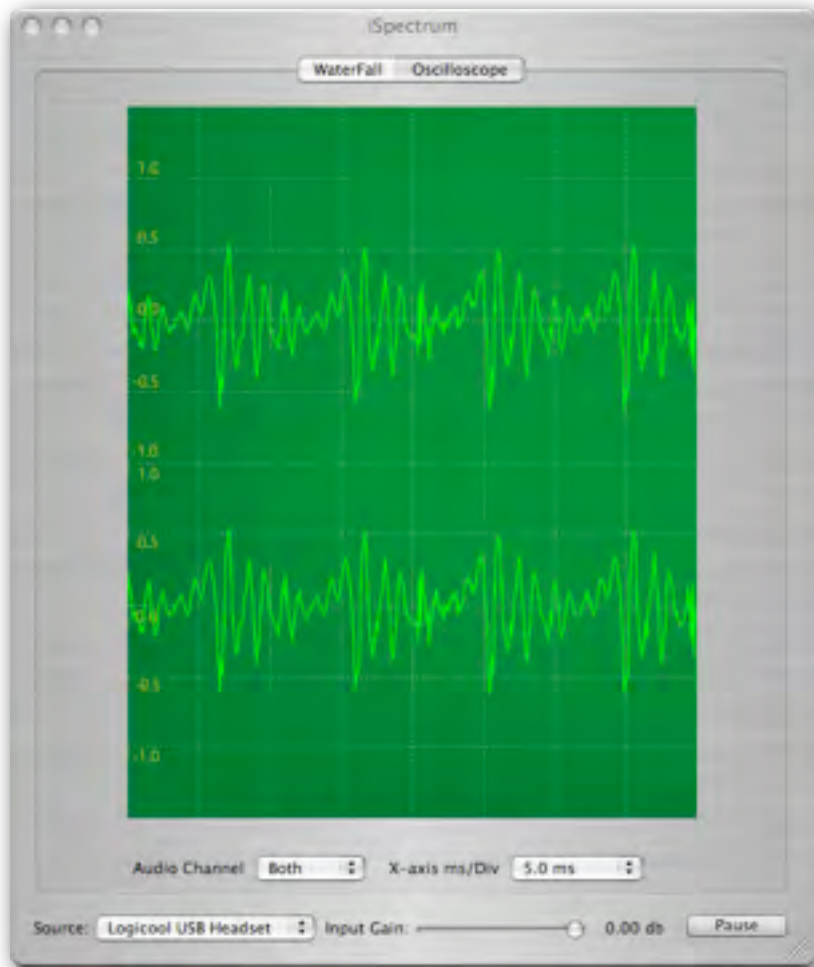


気体の中の音の速度と液体の中での音の速度はどちらが早い？

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！

🎧 iSpectrum で見た峯松の「低いあ」「高いあ」



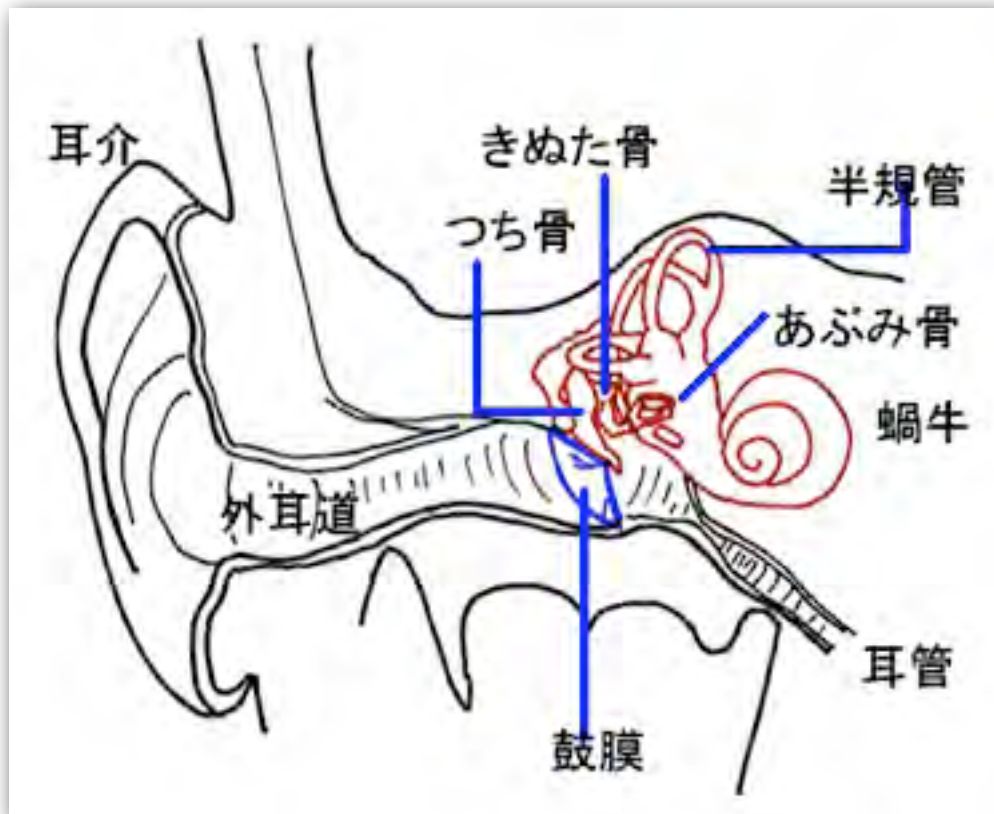
🎧 空気粒子の振動＝音，じゃ、空気粒子の移動は何になる？

空気振動としての音声 ～波～

音＝空気粒子の振動が全方向に伝わる縦波・疎密波

空気振動によって「モノ」は動くか

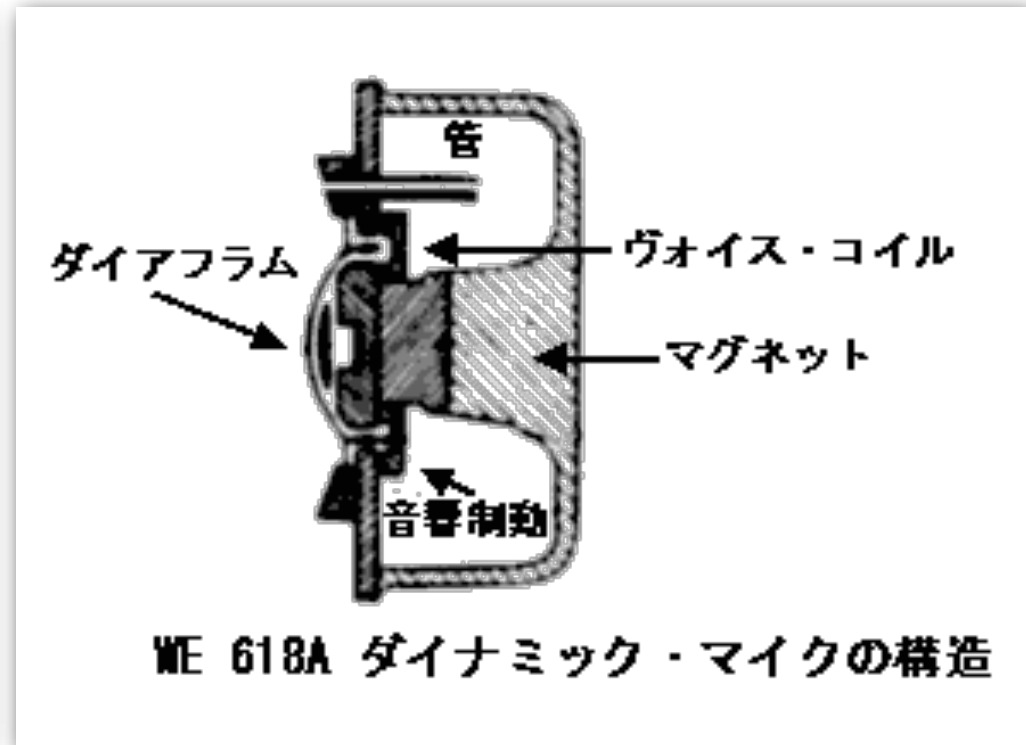
- 「あーー」と手に向かって話してみる（声）
- 「ふーっ」と手に向かって吹きかけてみる（風）
- 空気振動によって「モノ」は振動しない？いやする！



空気振動としての音声 ～波～

空気振動によって振動する導体があれば・・・

その振動を磁場の中で行なわせれば・・・



電気が流れる!!

空気振動としての音声 ～波～

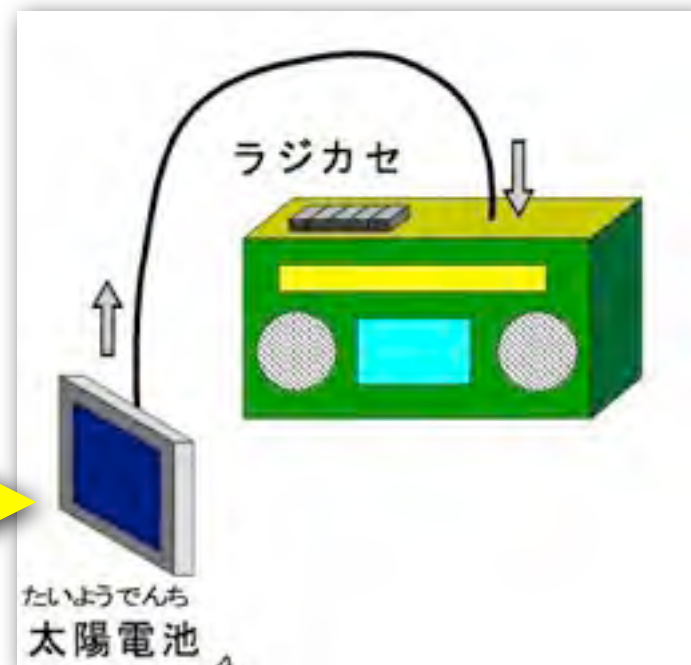
振動（運動）エネルギー→電気エネルギー



音声のマイク録音＝立派な交流発電

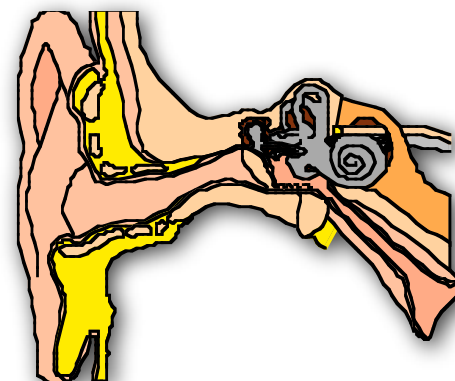
情報の本質はどこに・・・？

メディアは空気である必要は一切無い



必要なのは（ある特性を担った）振動

- 空気分子の音響的振動／アルミ箔の機械的振動
- 光強度の振動／電子の電氣的振動
- 最後は空気分子の音響的振動へ
 - ヒトの感覚器が捉えられるのは空気振動だけ



空気振動としての音声 ～波～

母音は疎密波, , , でも, ちょっと特殊な疎密波

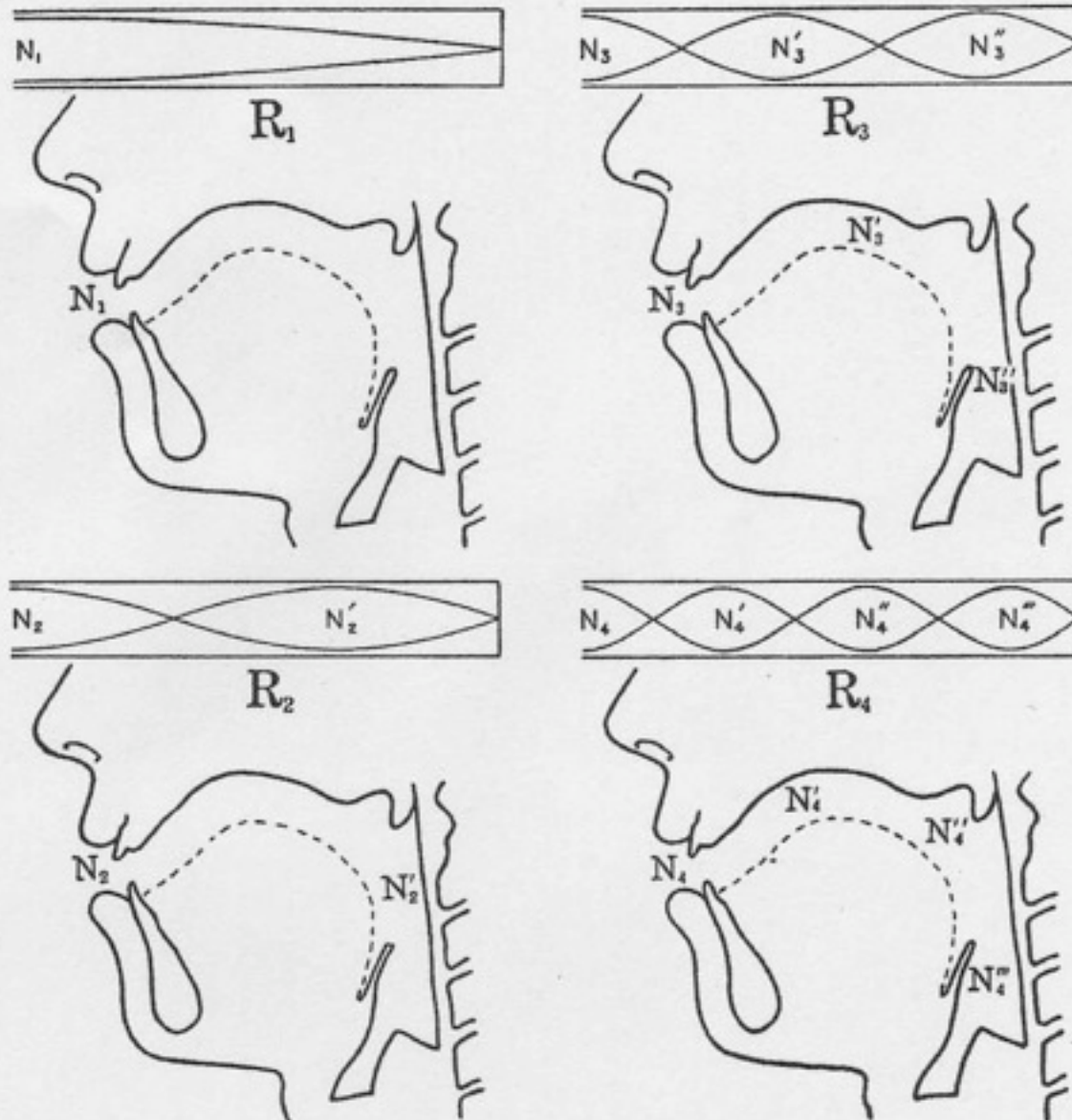


図 93 断面積が均一の音響管または声道における体積流の最大点の分布

定常波

共振周波数

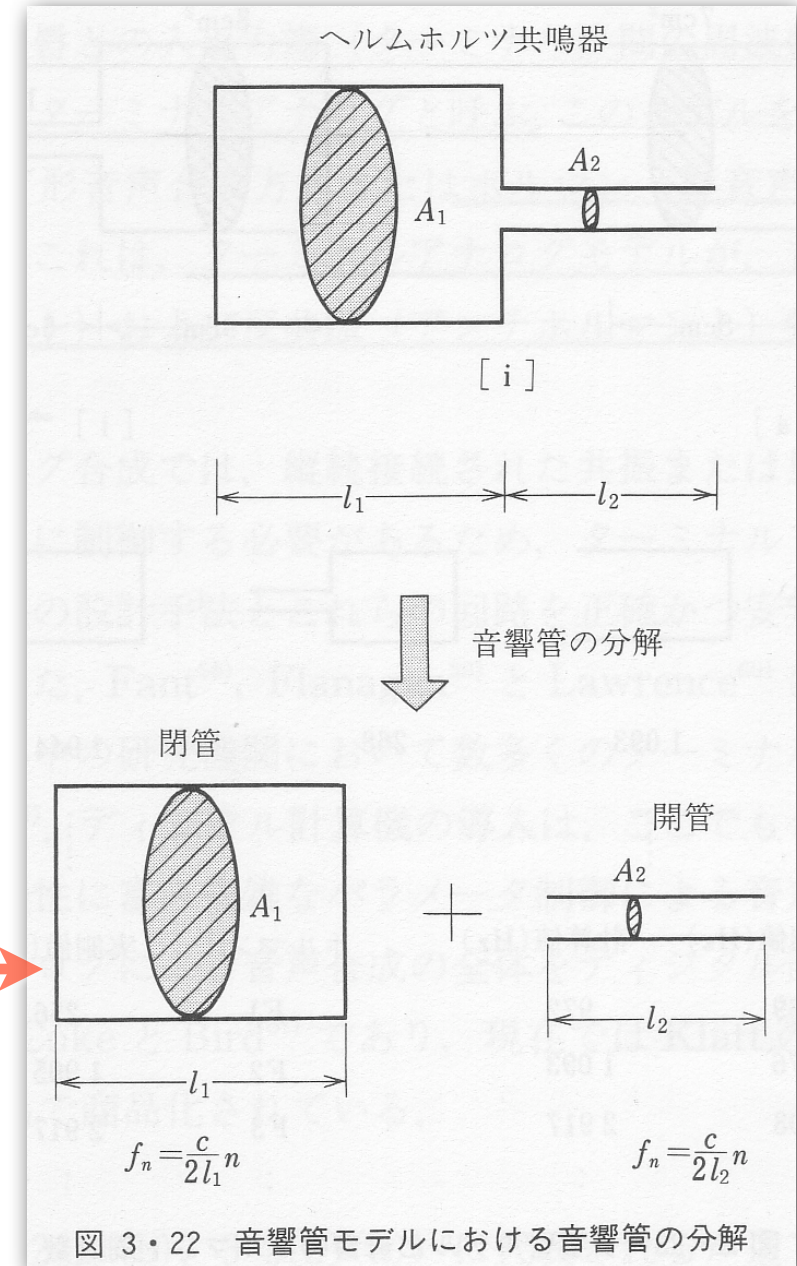
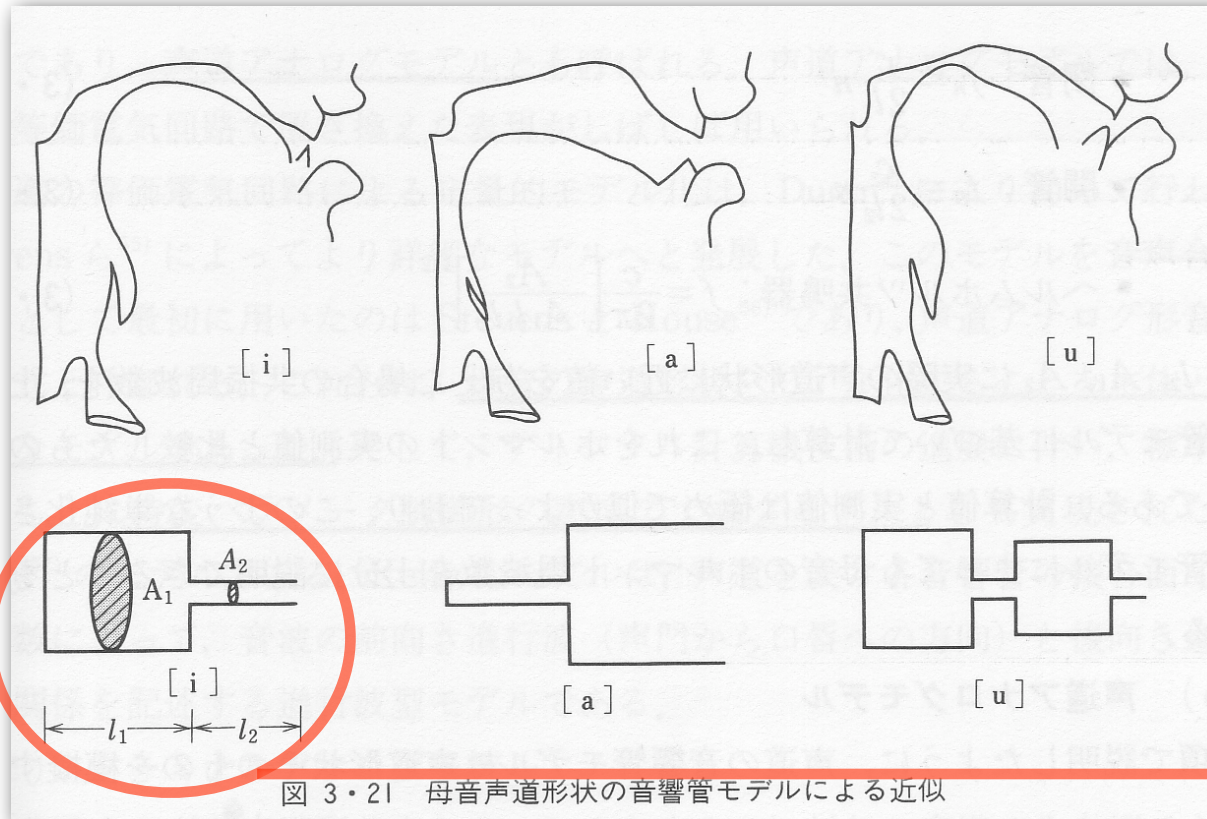
$$F_n = \frac{c}{4l} (2\underline{n} + 1)$$



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

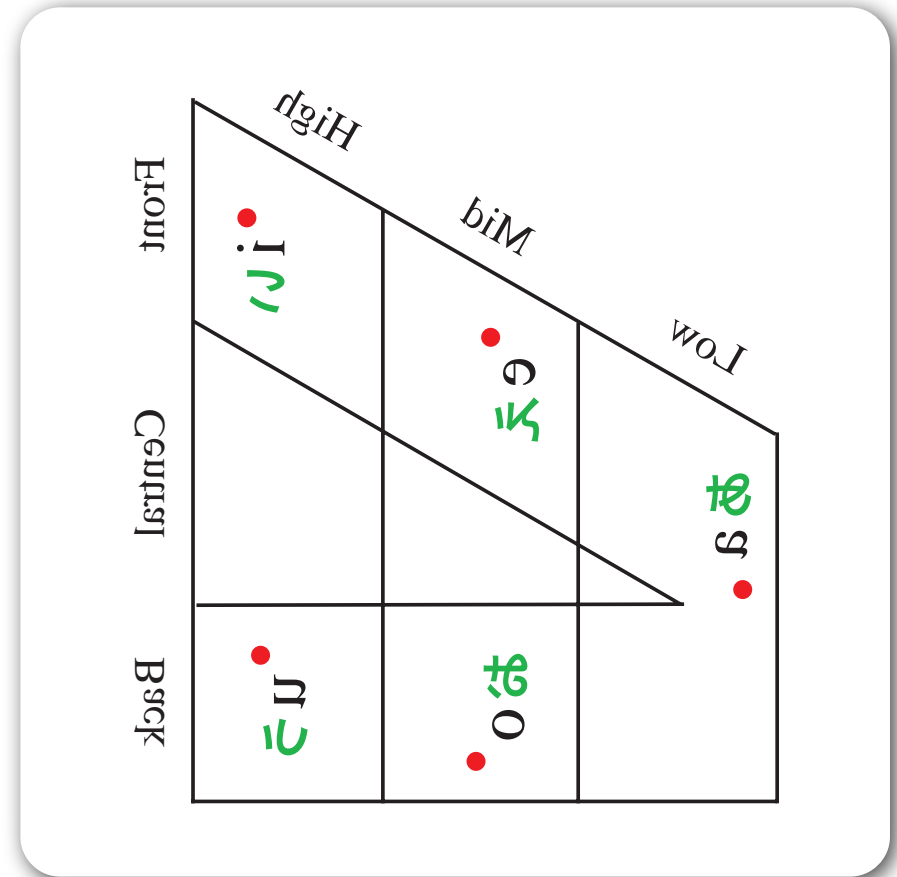
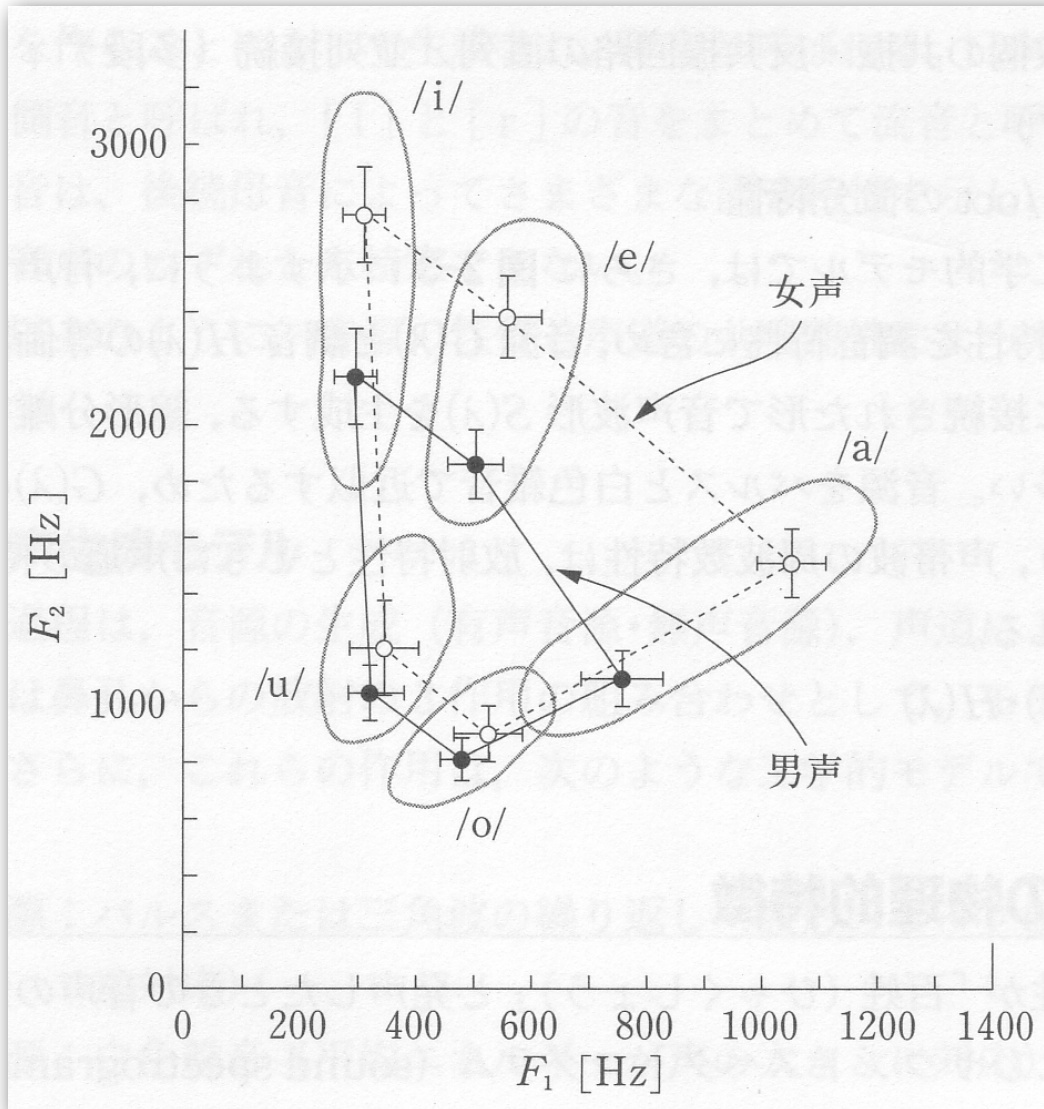


$$f_n = \frac{c}{2l_1} n \quad f_n = \frac{c}{2l_2} n \quad f = \frac{c}{2\pi} \left[\frac{A_2}{A_1 l_1 l_2} \right]^{1/2}$$

母音＝定常波 ～気柱の共鳴現象～

形の違い＝長さの違い＝共鳴周波数の違い

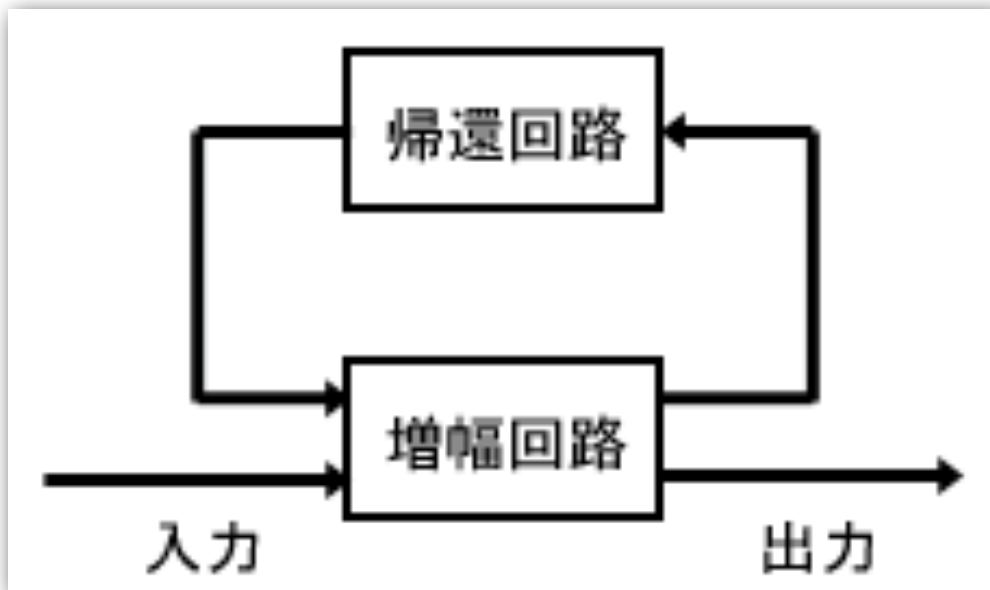
「あ」の一部＝「お」の一部



母音生成とLCR共振回路

母音発声とかけて、交流発振器と解く。その心は？

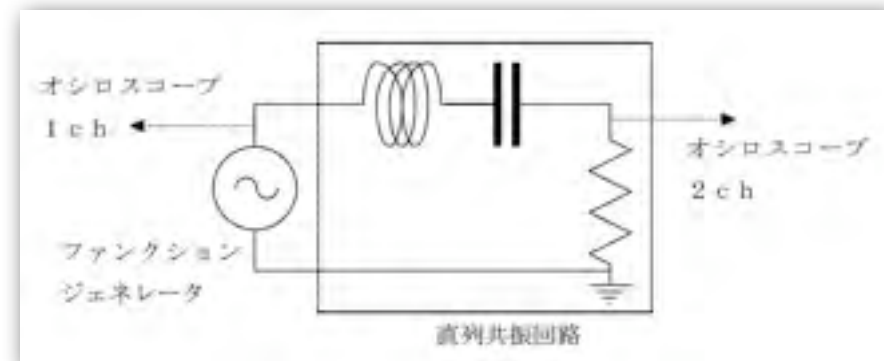
● どちらも共振／共鳴現象でしょう。



LCR帰還回路に周波数選択性を持たせると、ある特性の波だけを生成するようになる。

共振周波数の変化は何によってもたらされるのか？

- 電子振動の場合：交流発振器
 - 帰還回路のLCRの値
- 空気粒子振動の場合：母音発声
 - 口のサイズ／形

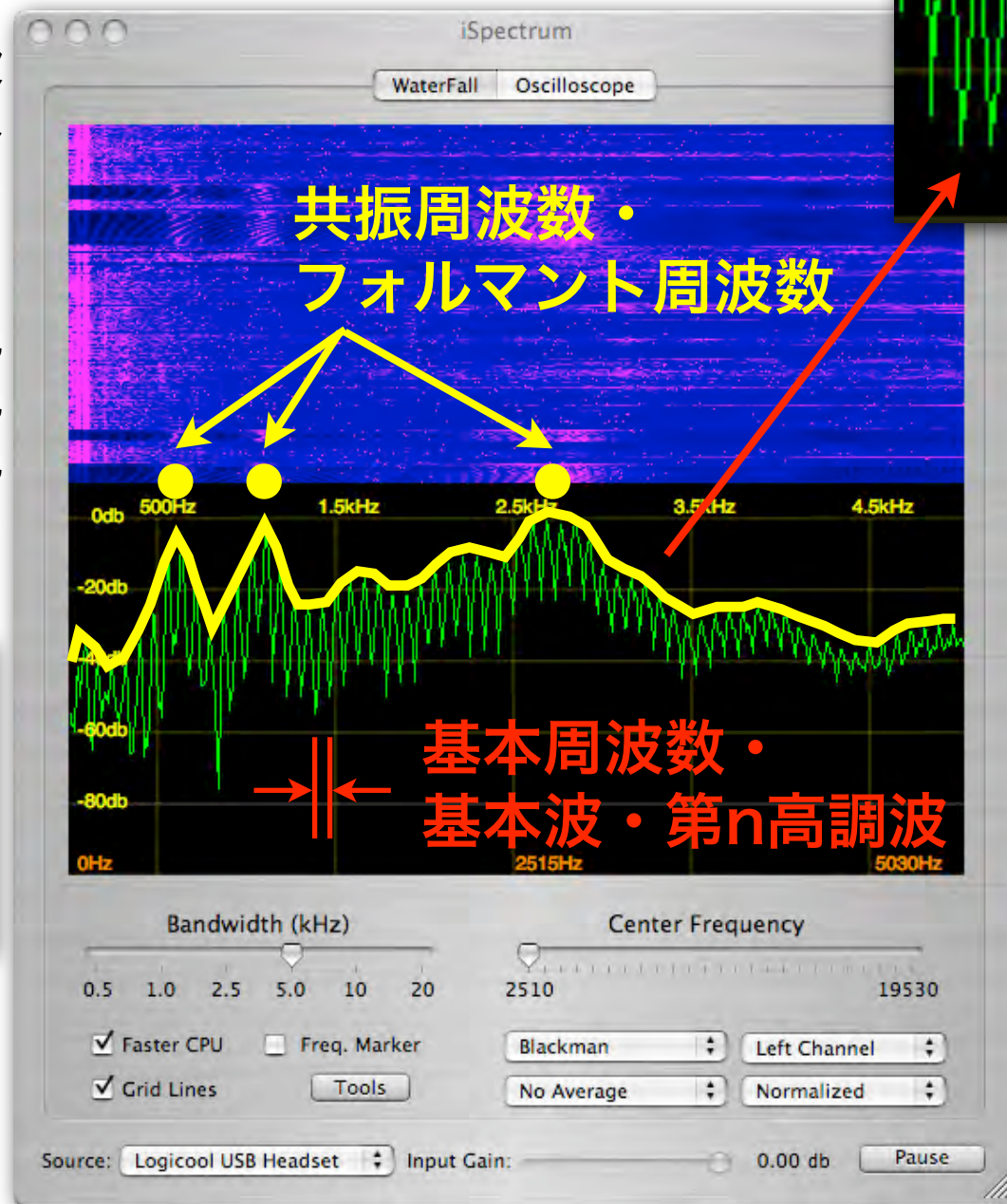


音声の音響分析

リアルタイム FFT アナライザー

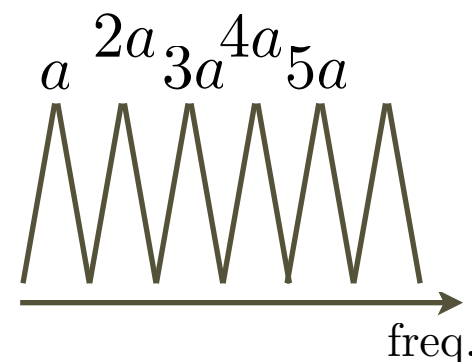
喉の形を変えると共振周波数が変わる。つまり、エネルギー分布の様子（パワースペクトル）が変わる。

これを、音響用語では音色と呼ぶ。楽器の違いは音色の違い、母音の違いも音色の違いである。話者の違いもまた、音色の違いである



共振周波数・
フォルマント周波数

基本周波数・
基本波・第n高調波



厳密には「音高 = a 」であって、ピークの間隔ではない。調波構造が無くても音高は感覚できる。



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。

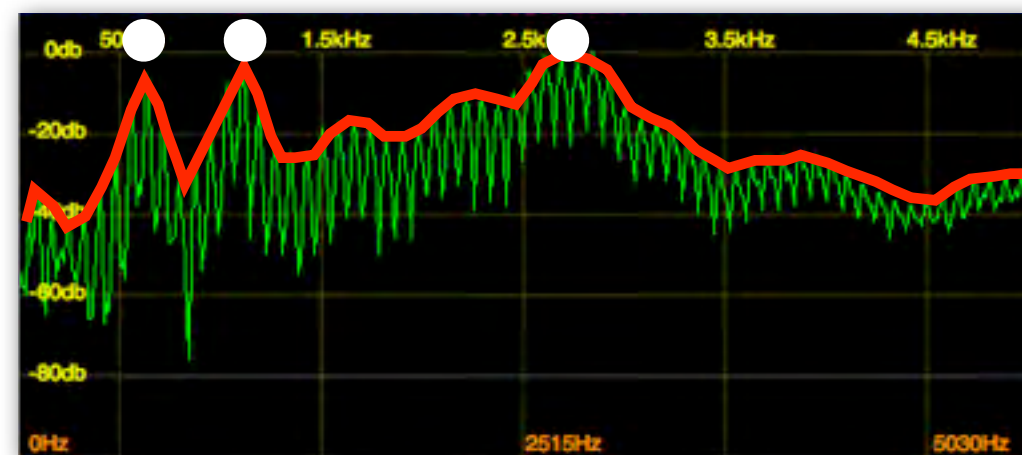
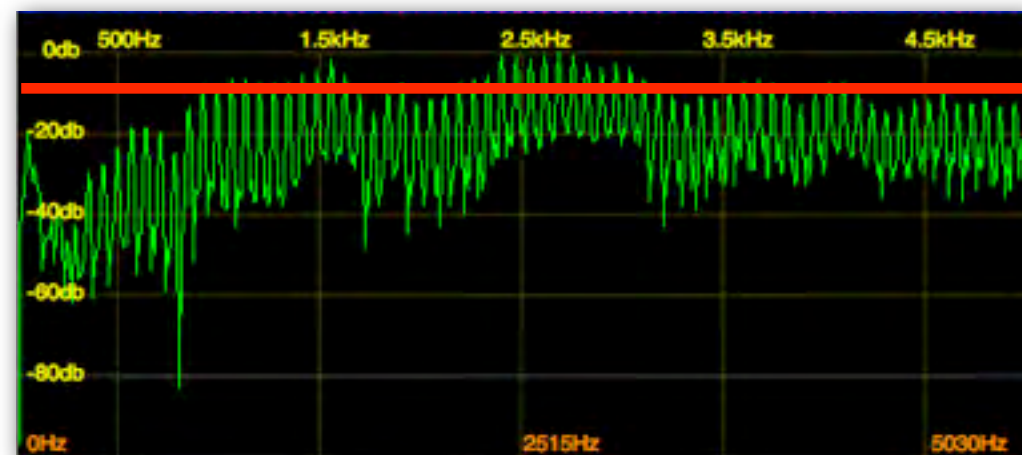
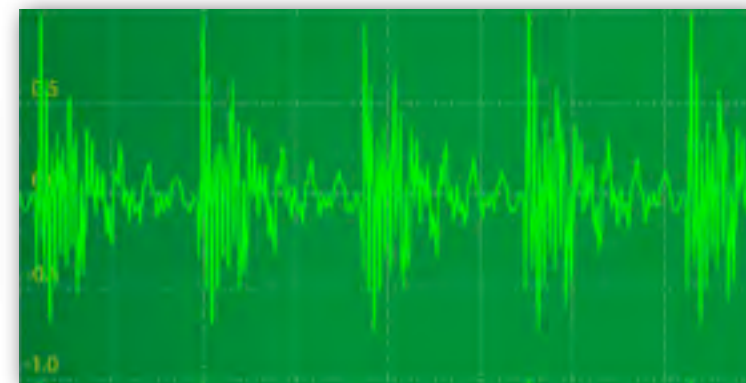
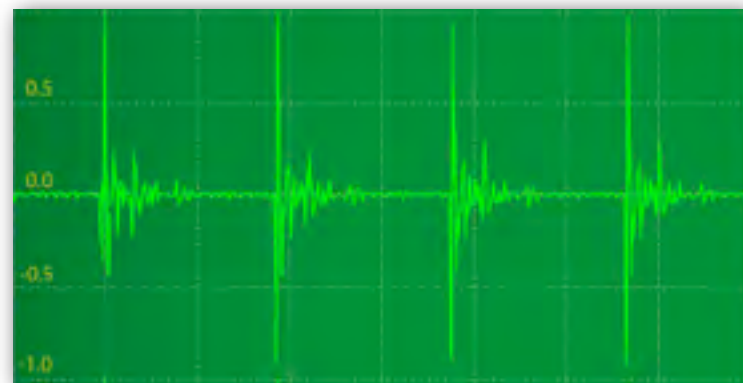
● 150回振動する波を生むのでは？ それは喉の共振で減衰するの？

● 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

声帯音源＋喉 ＝ ブザー＋管

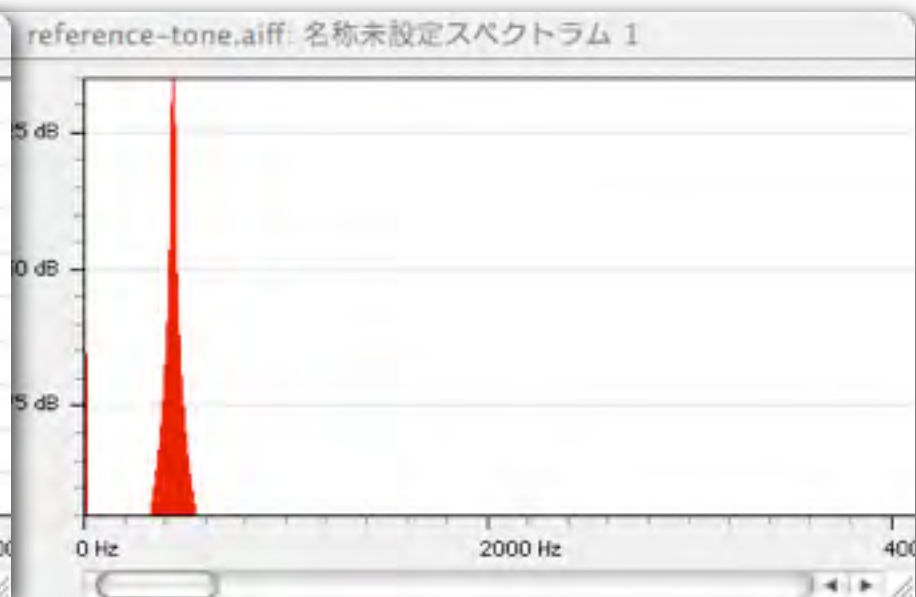
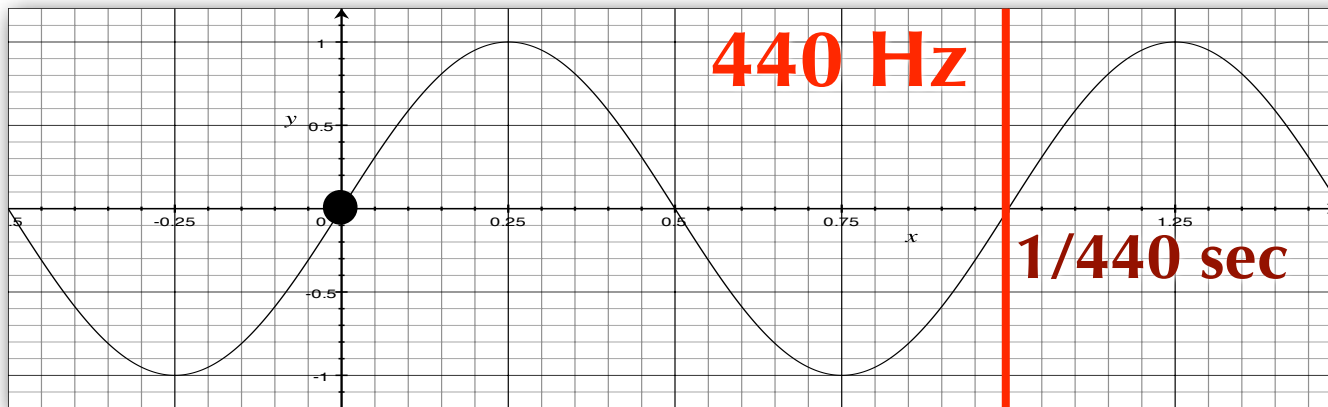
パルス列（ブザー）→管→定常波（あ～）



- 本当のパルス列ならばスペクトルは平坦になる。
- 例えば 100Hz, 200Hz, 300Hz,,, に等しい強さでサイン波ができる。
- それが**管の形が要求する周波数のみ**生き残り，他は減衰する。

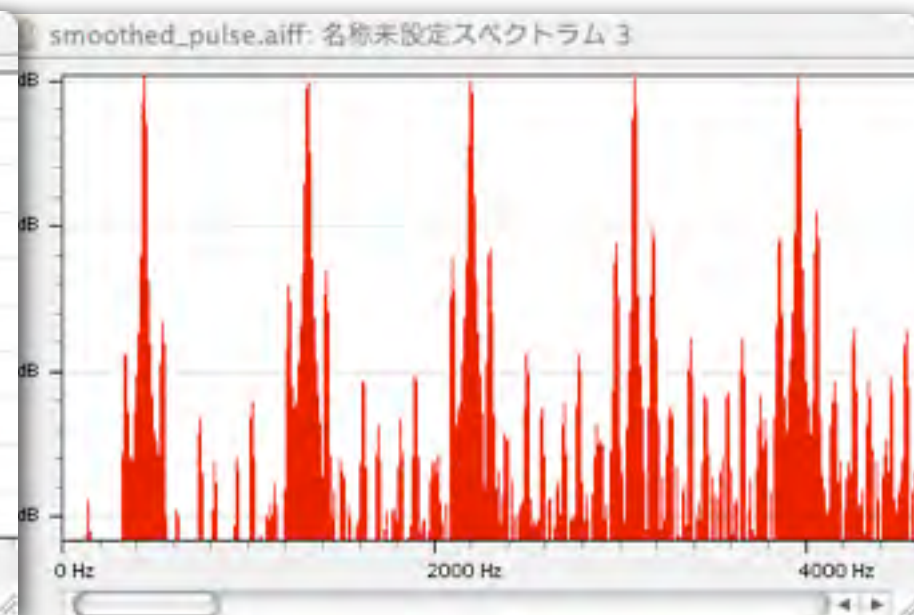
音の「高さ」をもっと詳しく

基本周波数 = $1 / \text{基本周期}$



音の「高さ」をもっと詳しく

これがパルス列的な波形になると・・・

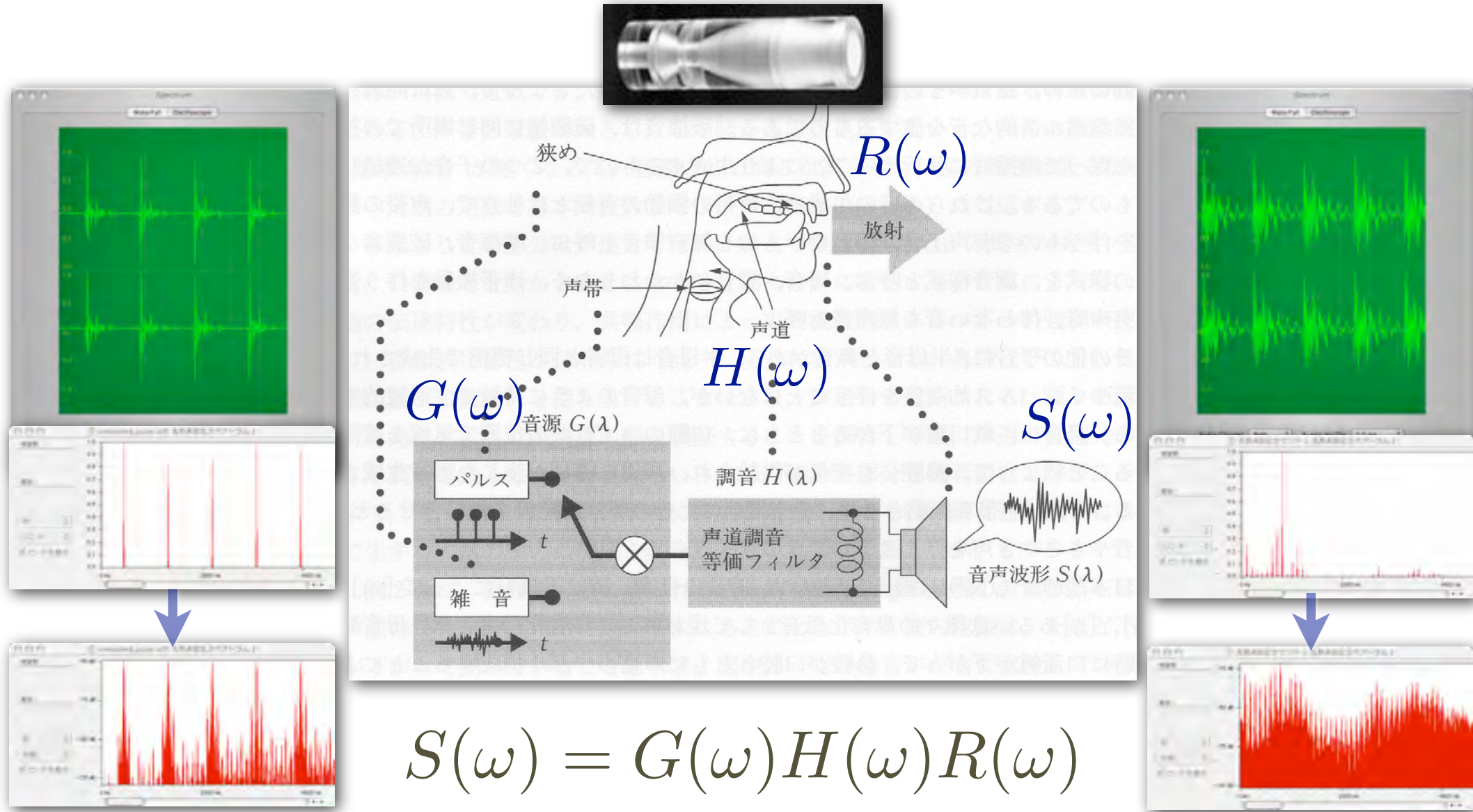


パルス列 = 440, 440x2, 440x3.... Hz のサイン波の重ね合わせ

音声の音響分析

音声（母音）生成のソース・フィルタモデル

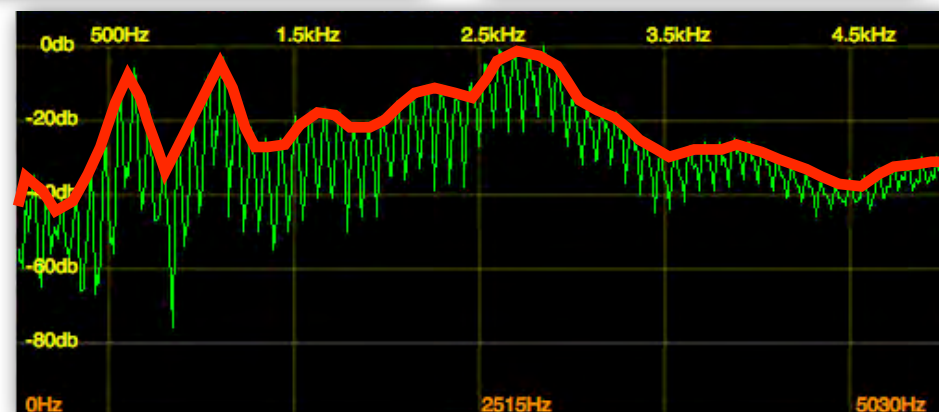
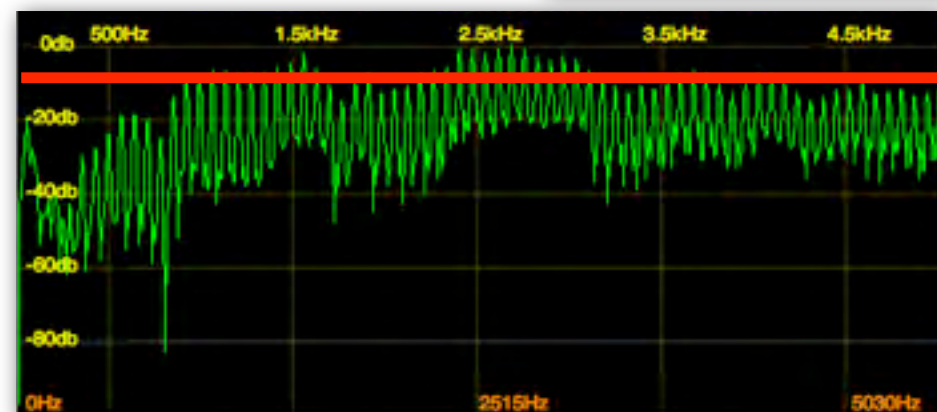
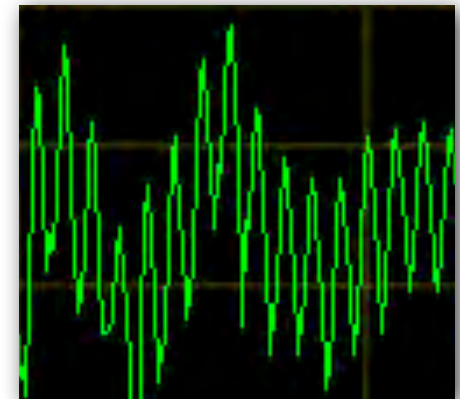
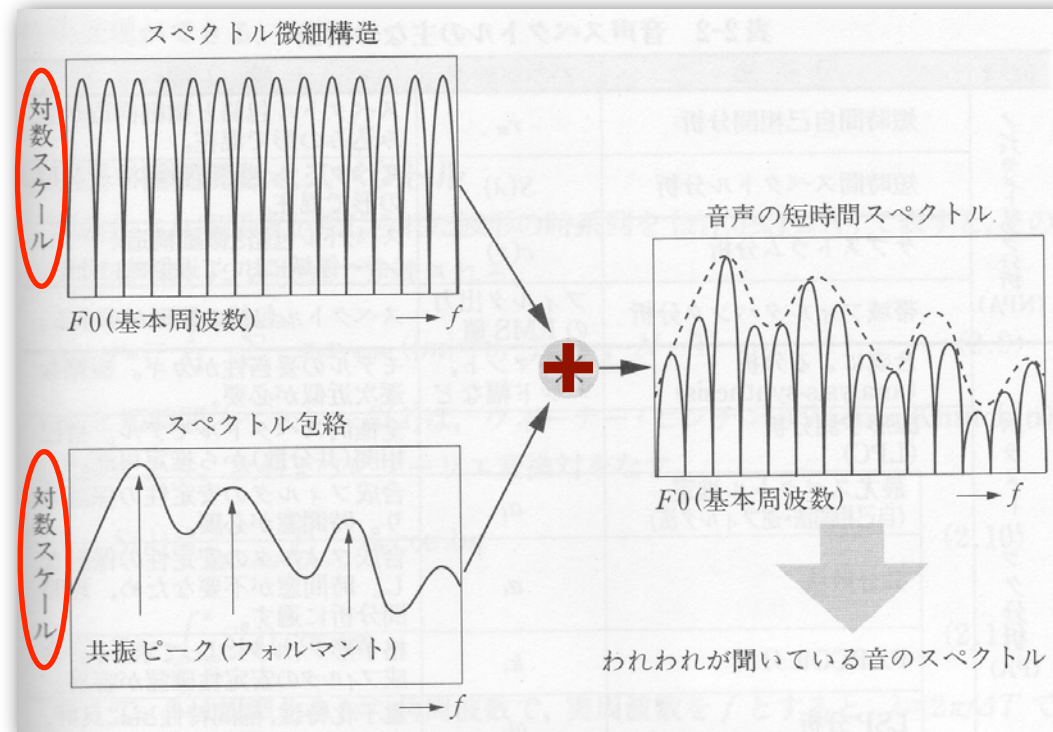
● 音声→音源の特性＋声道フィルタの特性



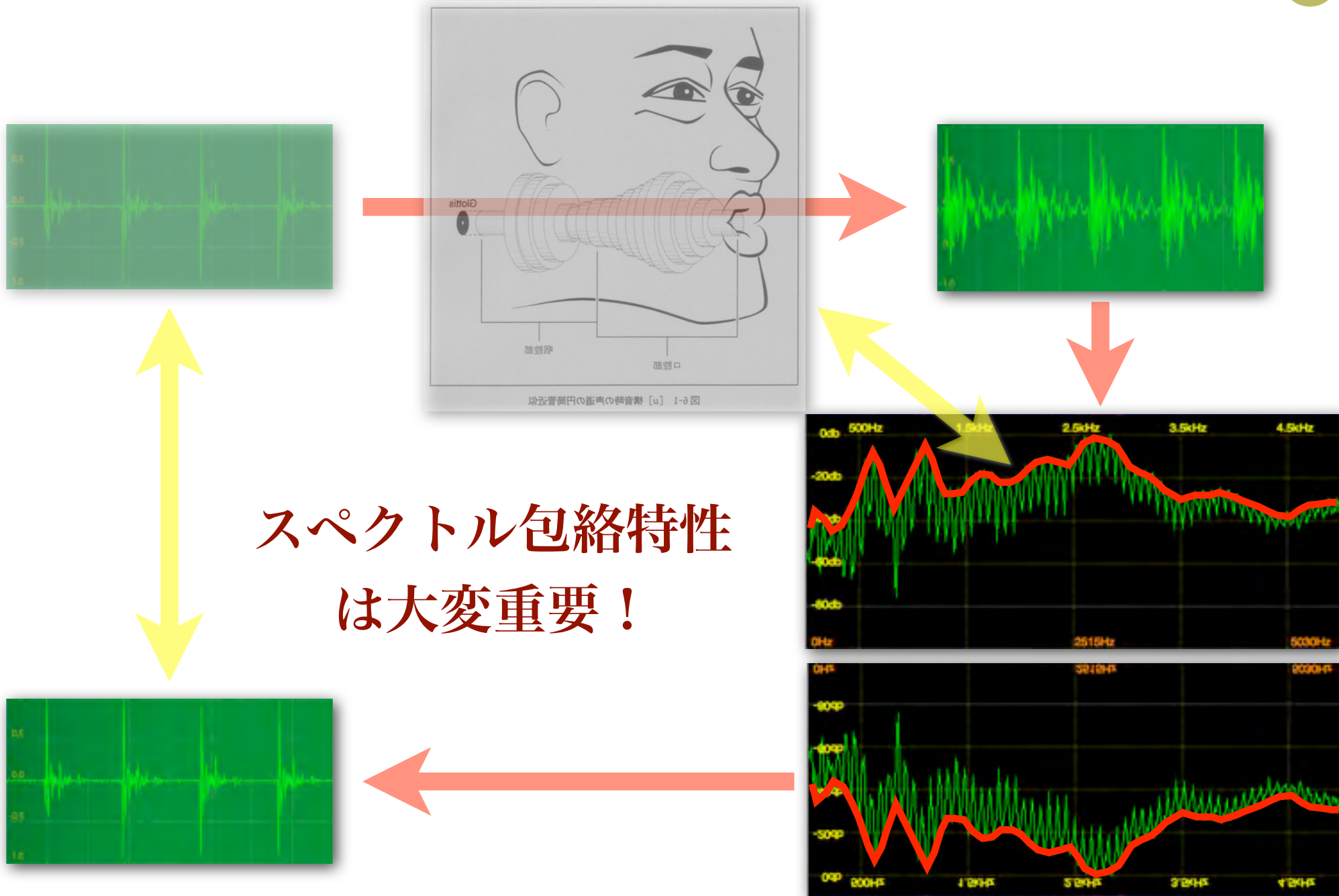
音声の音響分析

音声（母音）生成のソース・フィルタモデル

● 対数スペクトル特性においては、音源特性 + 声道特性



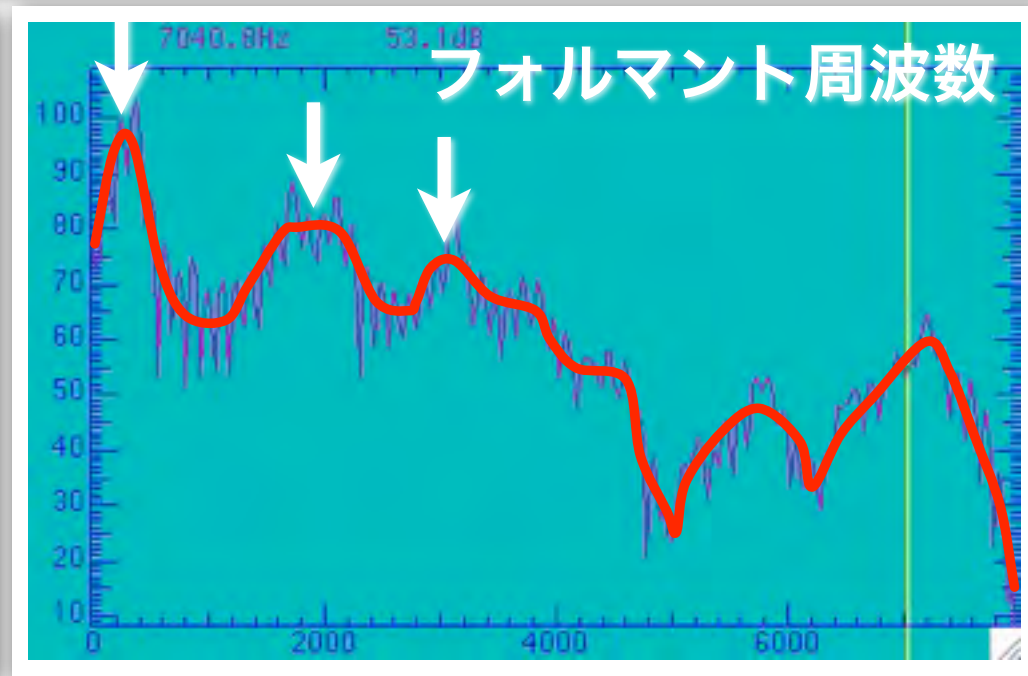
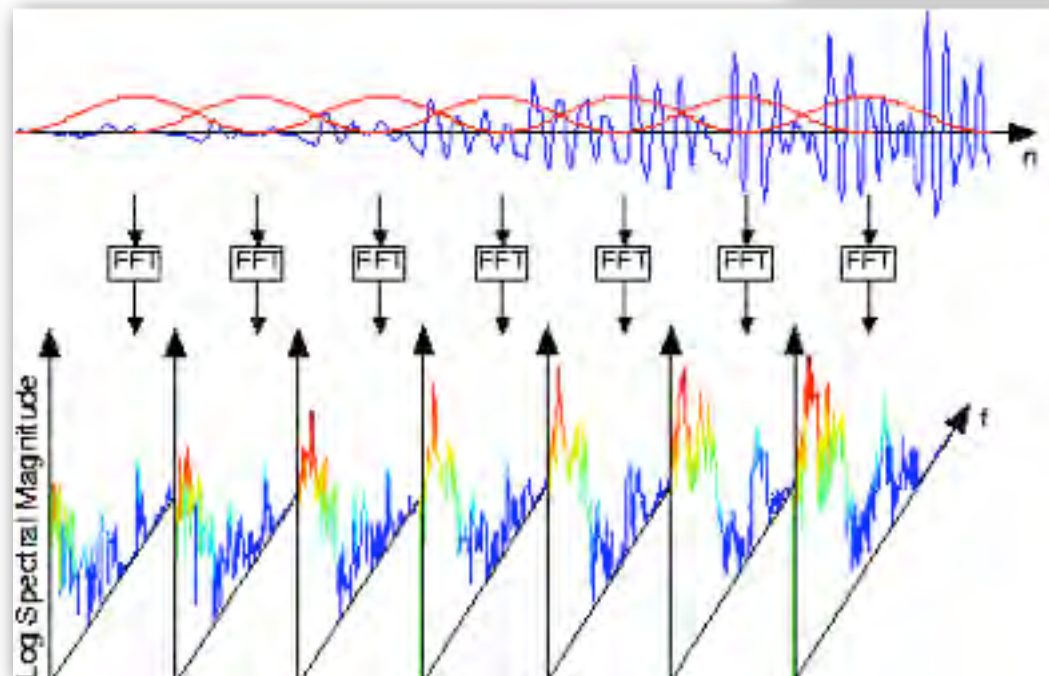
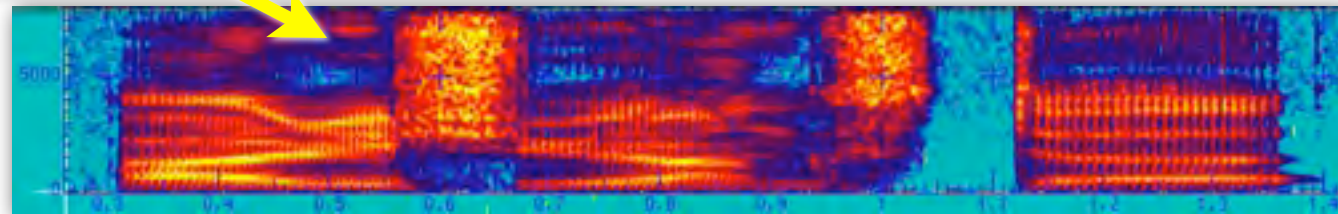
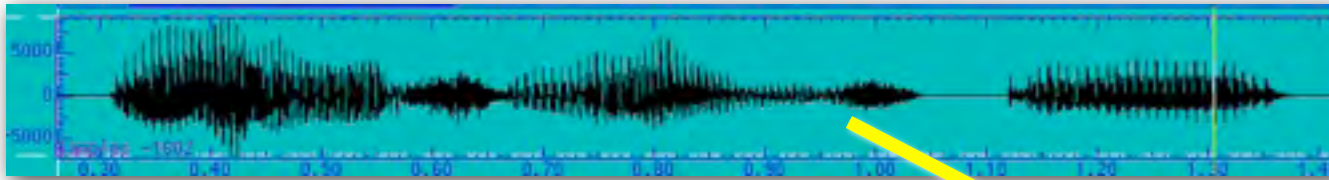
音声波形を使って推定できるもの



音声の音響分析

音声波形からスペクトルへ

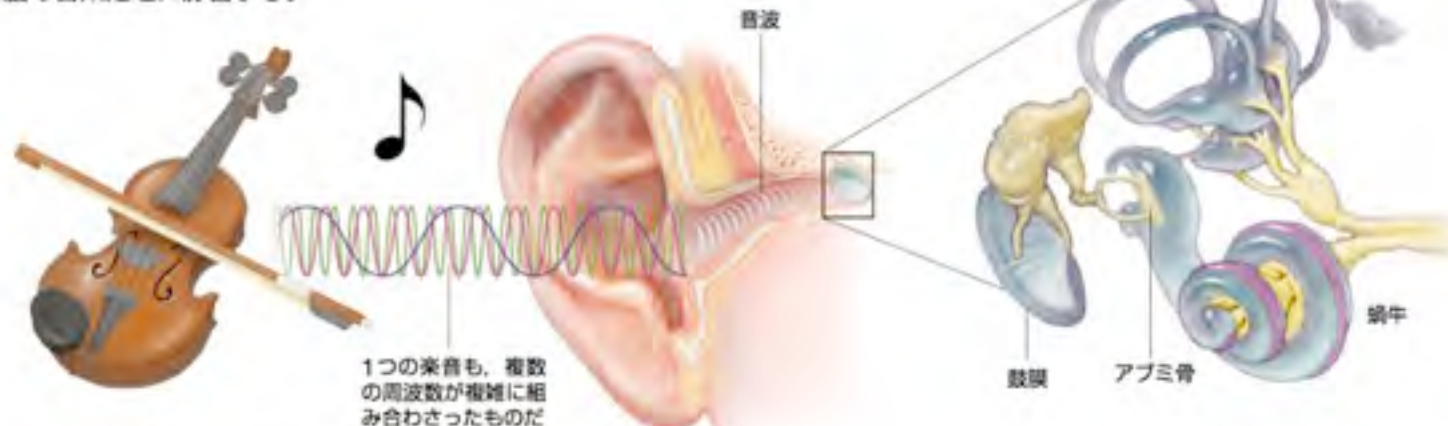
窓掛け→短時間フーリエ変換→対数パワースペクトル包絡



耳から脳へ 音楽の通り道

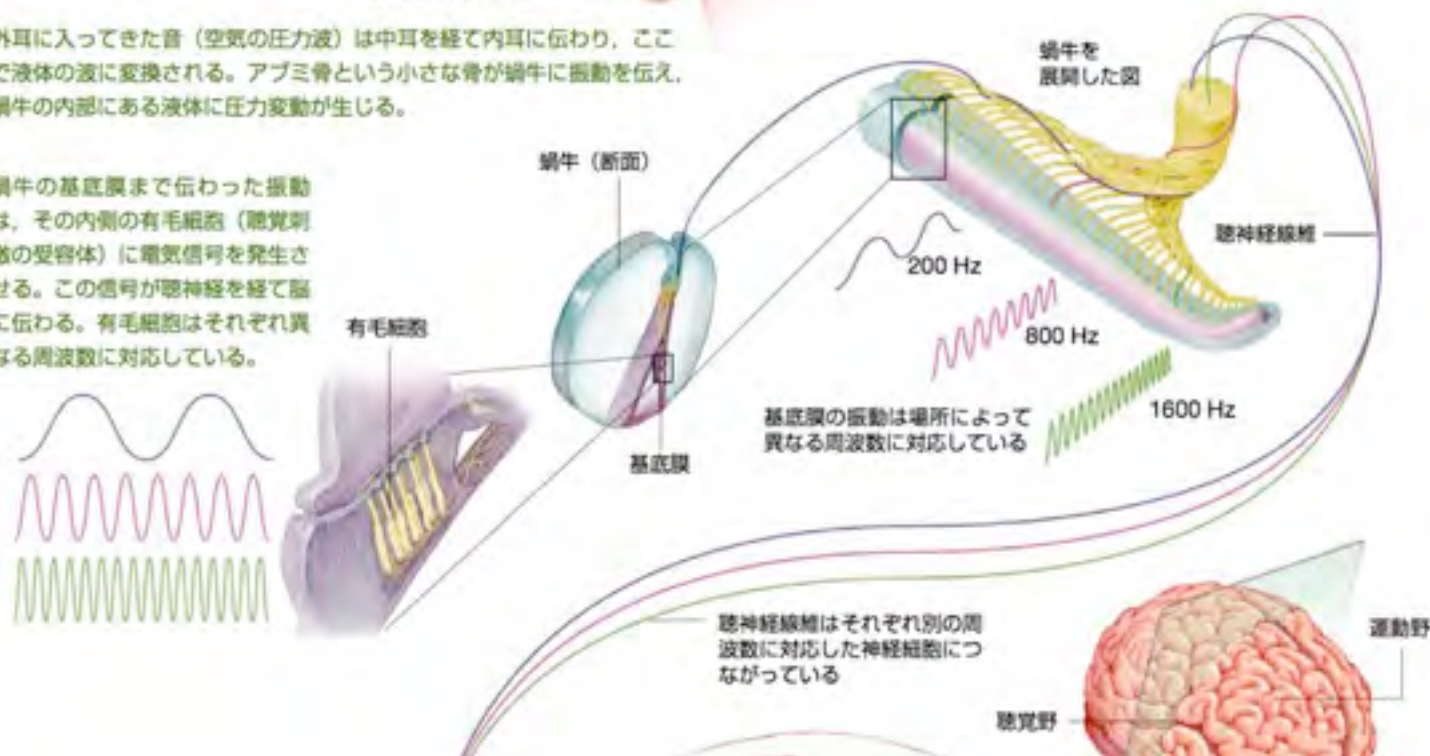
耳から脳へ 音楽の通り道

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



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

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



耳＝フーリエ分析器

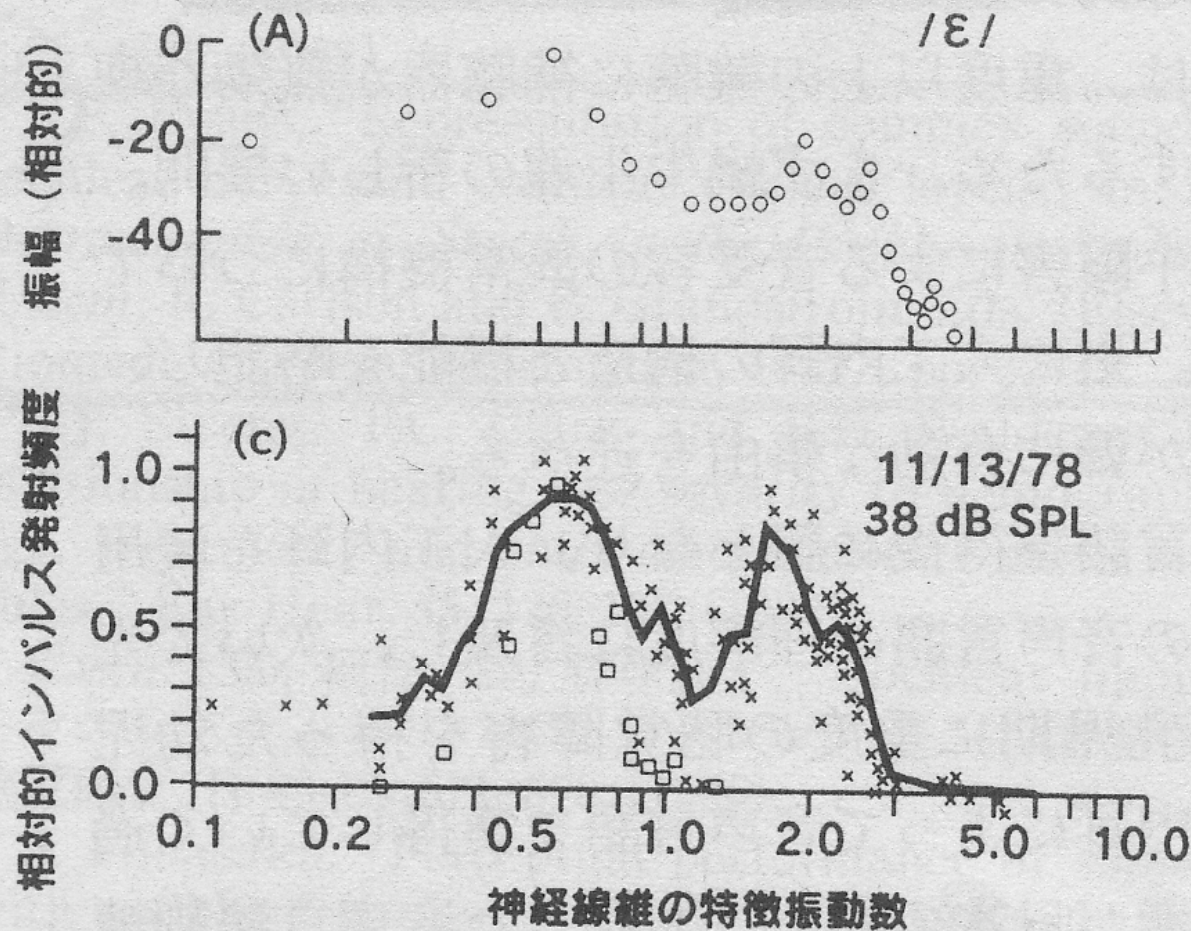


図-3 合成母音/ε/に対する聴神経の反応

正常蝸牛は時々刻々と入力音のフーリエ解析をしている。この機能は蝸牛障害で失われる。A：/ε/の振幅スペクトル，C：様々な特徴振動数の神経繊維の/ε/に対する相対的インパルス発射頻度（文献5）より引用）

ちょっと耳テスト

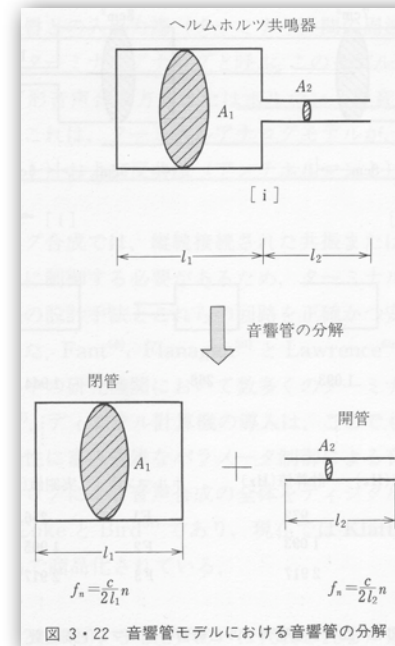
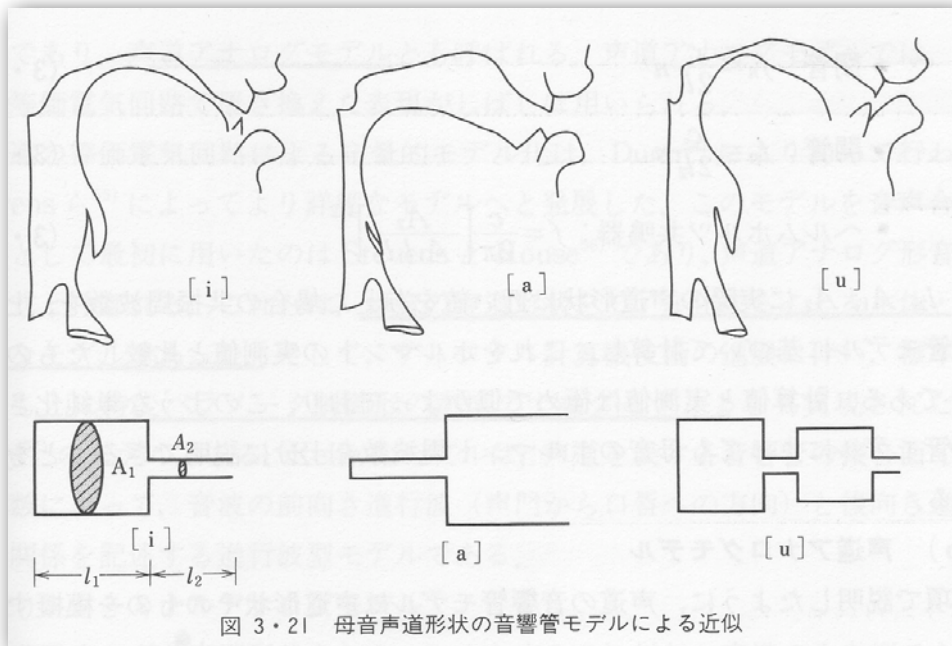
峯松 信明

東京大学工学部電子情報工学科



質問 1

Q：管を短くするとどんな声になるのか？



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

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

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

A → B → C

リアルな喉モデル



質問2

Q：管を「あ」から何の形に変えたのか当てなさい。

● この人の「あ」の管の形から変えていきます。

A

=



A



B

B



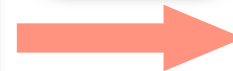
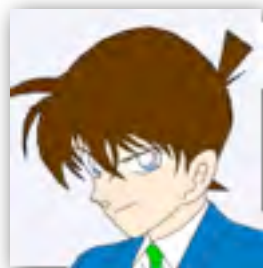
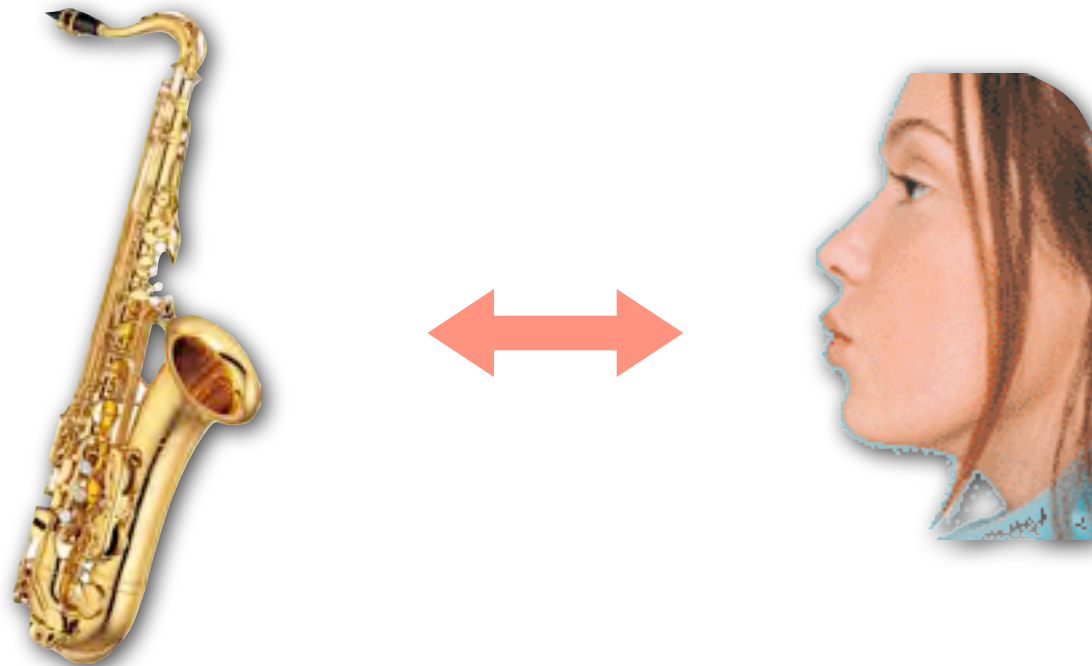
A

Replay

質問 2

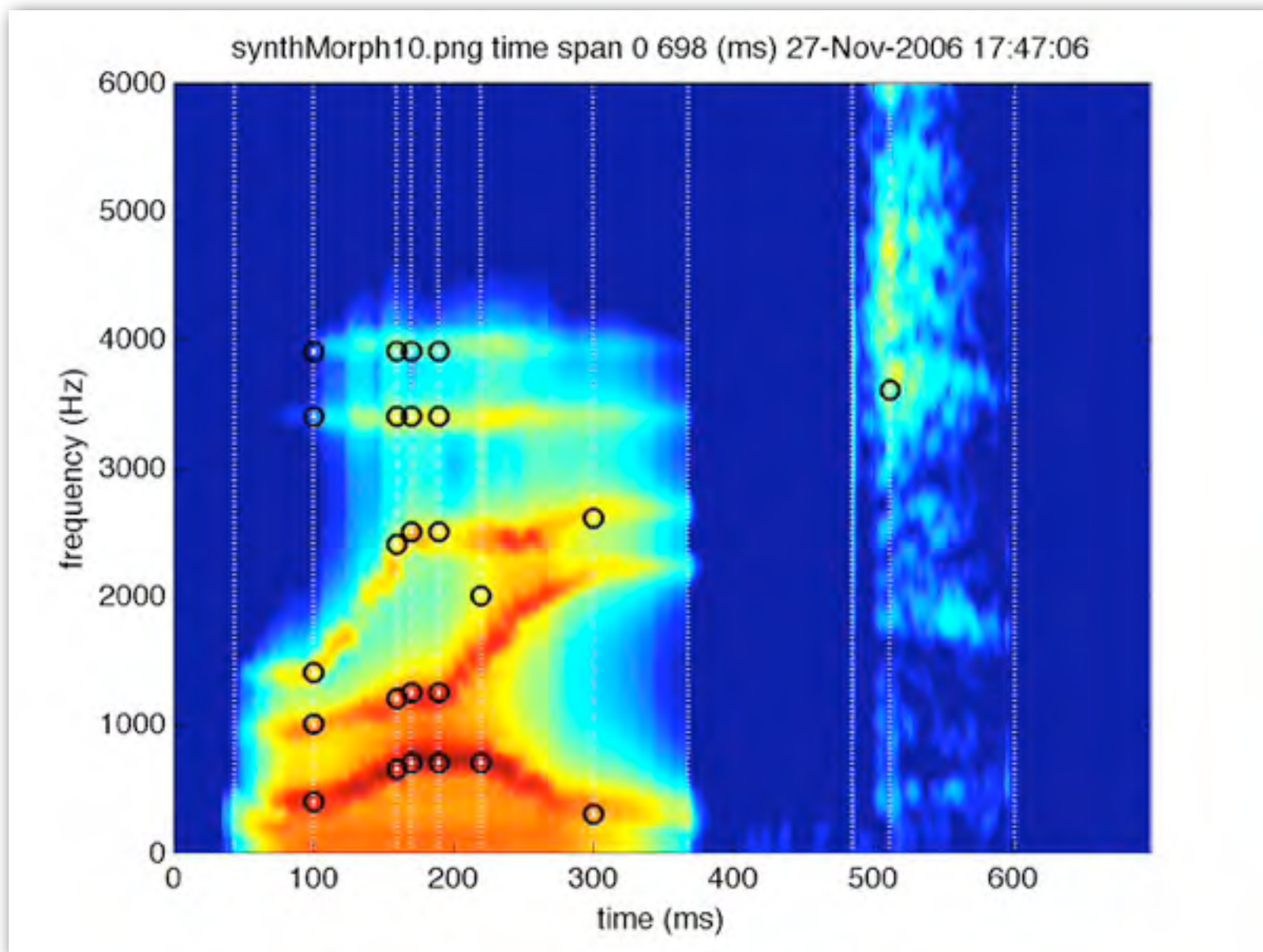
Q：管を「あ」から何の形に変えたのか当てなさい。

● この人の「あ」の管の形から変えていきます。



音声モーフィング (変形)

音韻の種類を連続的に変えてみる。 $r \rightleftharpoons l$



音声モーフィング（変形）

普通の声歌声に変えてみる。

● 分析再合成技術

- 音声を音色・ピッチ・パワー・長さの各要素にばらして、各々を適宜変形・修正して、再度統合して音声にする。

● 歌声音声に対する分析再合成

- 女声の持続母音/a/（440Hz, 原音声）
 - 再合成音声（変換無し）
 - 再合成音声（ソプラノ化）
 - 再合成音声（アルト化）
- 男声による旋律/a/（原音声）
 - 再合成音声（変換無し, バリトンとして使用）
 - 再合成音声（テノール化）
 - 再合成音声（バス化）
- 出来上がった5声による演奏（元は2音声）

音声モーフィング (変形)

怒らせたり悲しませたり楽しませたり・・・

