

I485F 知覚情報処理特論

1. イントロダクション (ヒトの知覚処理)

2023年4月15日(土)

人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

Email: unoki@jaist.ac.jp



1. 知覚

■ 知覚 (Perception)

ヒトが外界からの刺激を感じ取り、意味づけすること。

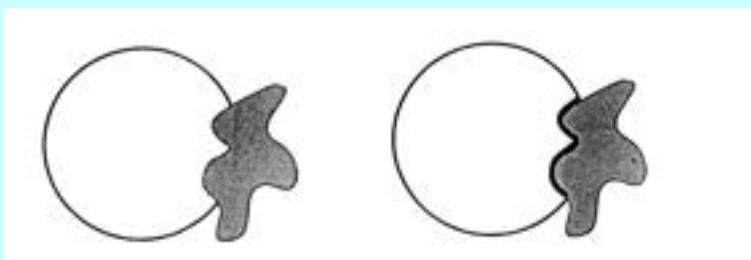
視覚、聴覚、嗅覚、味覚、体性感覚、平衡感覚などの感覚情報をもとに、「熱い」「重い」「固い」などという自覚的な体験として再構成する処理のこと。

■本講義の目的

ヒトの知覚系に関する生理学・心理学上の処理過程においてその機構と特性について学ぶ。

特に、聴覚系の生理とモデルならびに聴知覚とモデルについて学び、現在までに提案されている聴覚モデル・シミュレータの事例を取り上げて議論する。

Visual examples



2. 五感

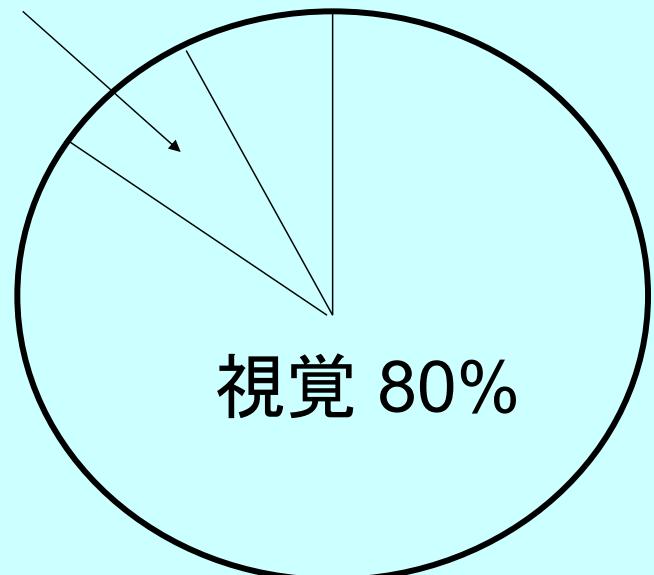
感覚器官

- 視覚 目
- 聴覚 耳
- 触覚 皮膚
- 味覚 舌
- 嗅覚 鼻

ヒトが外界から得る情報の割合

聴覚 10%

ホント？



■ モダリティ

感覚器官を使って外界を
知覚する手段のこと

3. 視覚

■ 一般に人間が外界から得る情報の8割程度は視覚によるもの（本当？）

■ 視覚が捉える情報：光（電磁波）

可視光：波長 $0.4\sim0.8\mu\text{m}$

放射線：光より波長が短いもの

電波：光より波長が長いもの

■ 光の強度、波長（色）にかかる時空間的な状態を検知

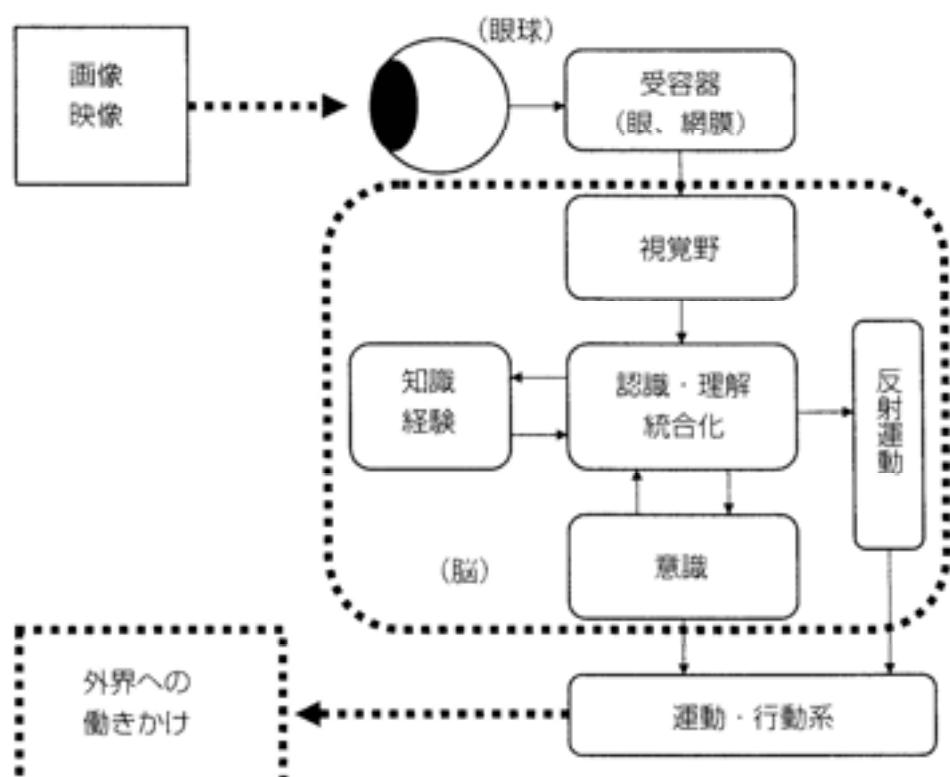
視覚的情報の要素

形、大きさ、位置、配置、
距離、奥行き、方向、構
造、テクスチャー、対称
性、隠蔽、相互干渉

出現時刻、消失時刻、
長さ（時間的）、周
期性、規則性、ラン
ダム性

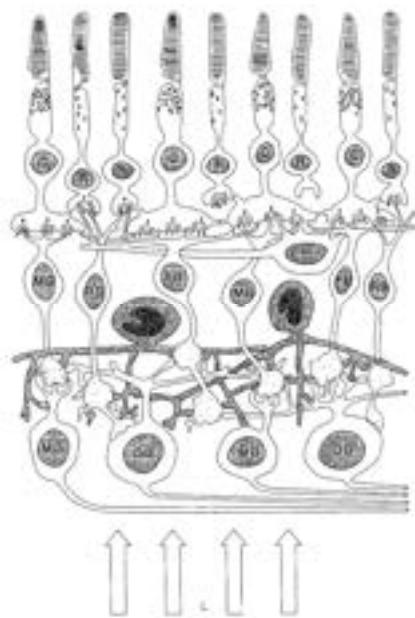
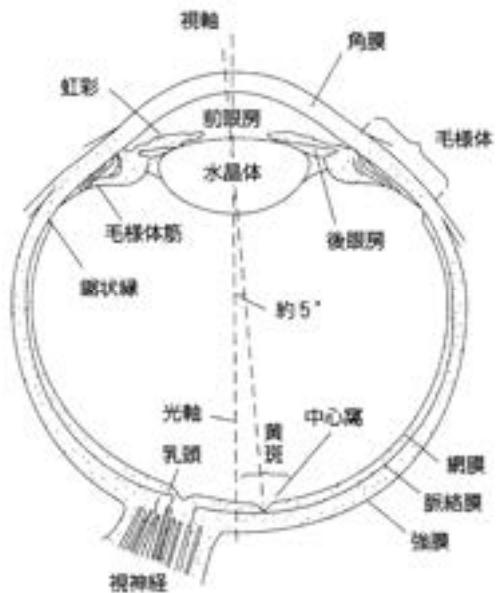


視覚のメカニズム



眼球の構造

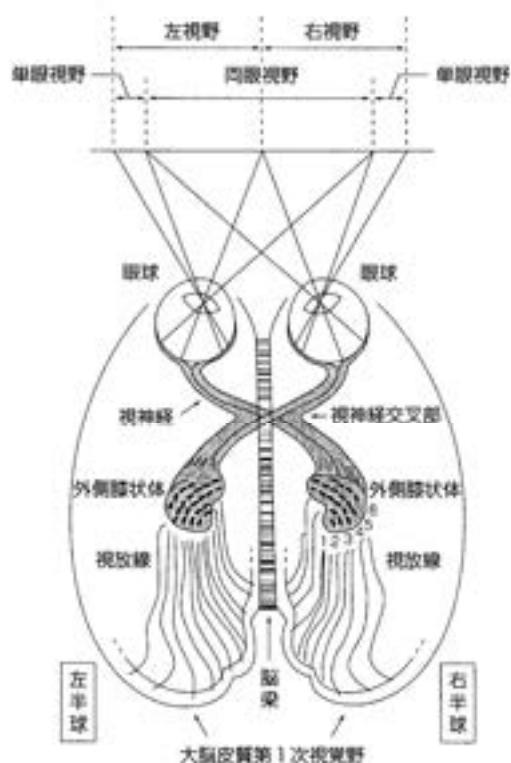
▼眼の構造図



R: 棒状体
C: 膜状体
H: 水平細胞
MB(Rod型),
PB(Rod型),
FB(Fan型): 双極細胞
MG, DG: 神經節細胞
A: アマクリン細胞
S: 先

▲網膜の断面の概念図

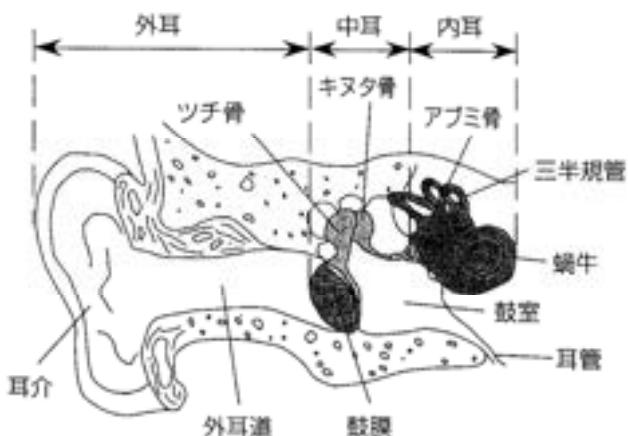
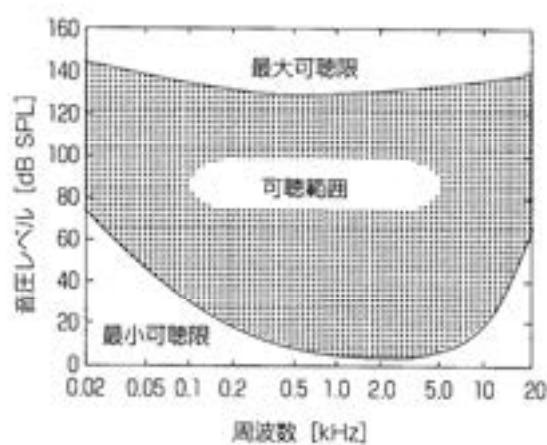
神経経路と視野



1, 3, 5: 交差性神経入力, 2, 4, 6: 非交差性神経入力。
右半面の視野は左脳に、左半面の視野は右脳に投射される。

4. 聴覚

- 一般に人間が聴覚から得られる情報は、視覚が五感全体の8割程度であるのに対し、1割程度であるといわれている（本当？）
- 聴覚が捉える情報：音の圧力変化



聴覚末梢系のメカニズム

音の伝達路

空気の
圧力振動

外耳道

耳小骨

鼓膜

機械振動

リンパ液内の振動

前庭階

鼓室階

内有毛細胞・
外有毛細胞の
毛のたわみ

周波数分解

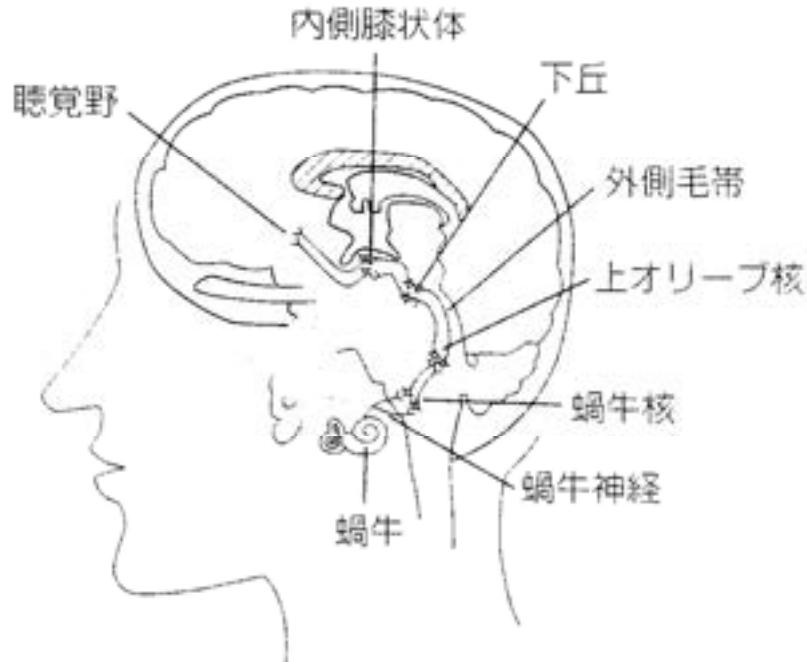
基底板
神経よ

電気信号
(パルス)

内耳

中耳

聴覚中枢系の伝達経路



5. 触覚

■触覚

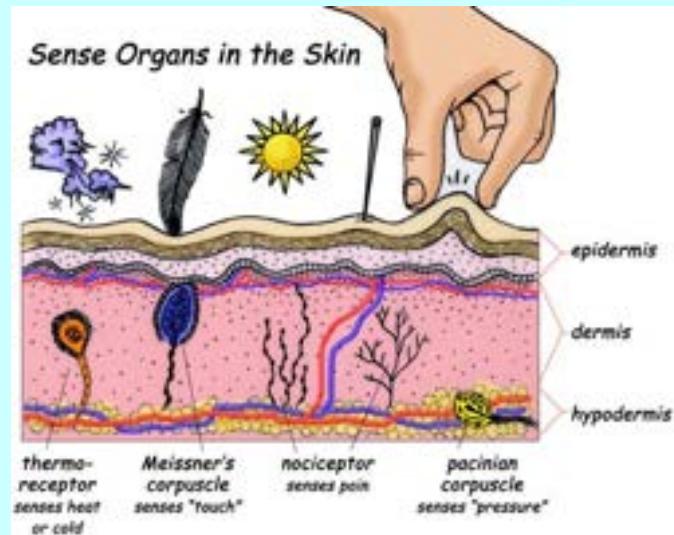
主に皮膚に存在する受容細胞によって受容され、
体表面に生起すると知られている感覚のこと
を指す。深部感覚などとあわせて体性感覚と呼
ばれることが多い。

接触覚、圧覚、振動覚、温度覚、痛覚

深部感覚（筋肉、腱、関節など）

■皮膚感覺

- 触圧覚（モノに触れたり、圧迫を受けるときの感覚）
- 振動覚（モノの振るえを感じる感覚）
- 温度覚（冷たさ・暖かさを感じる感覚）
- 痛覚（痛みを感じる感覚）

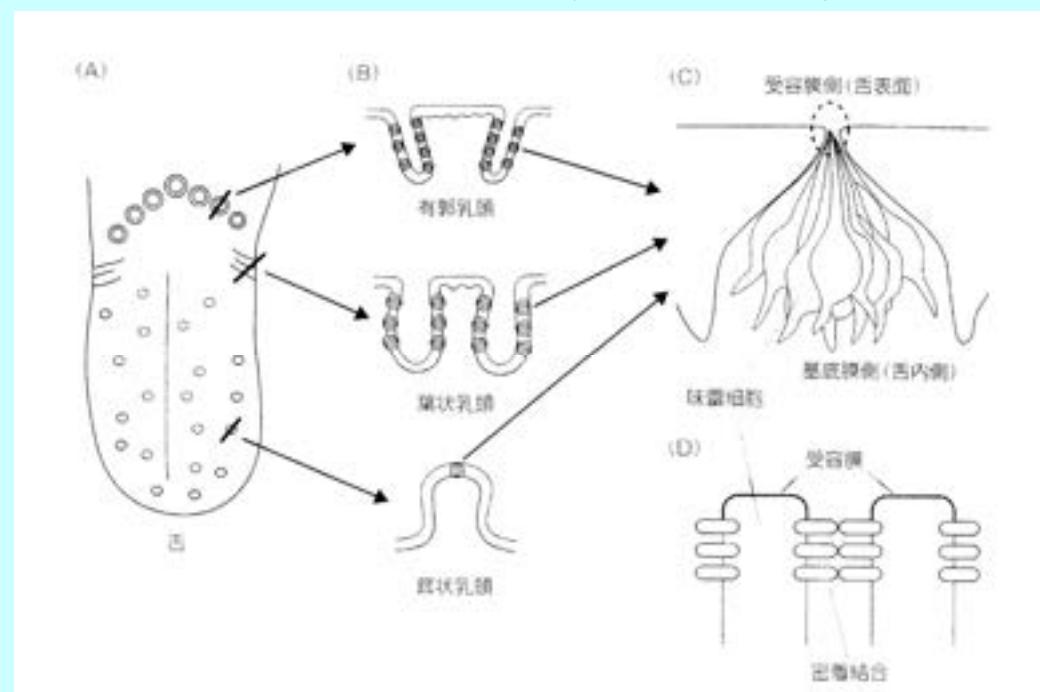


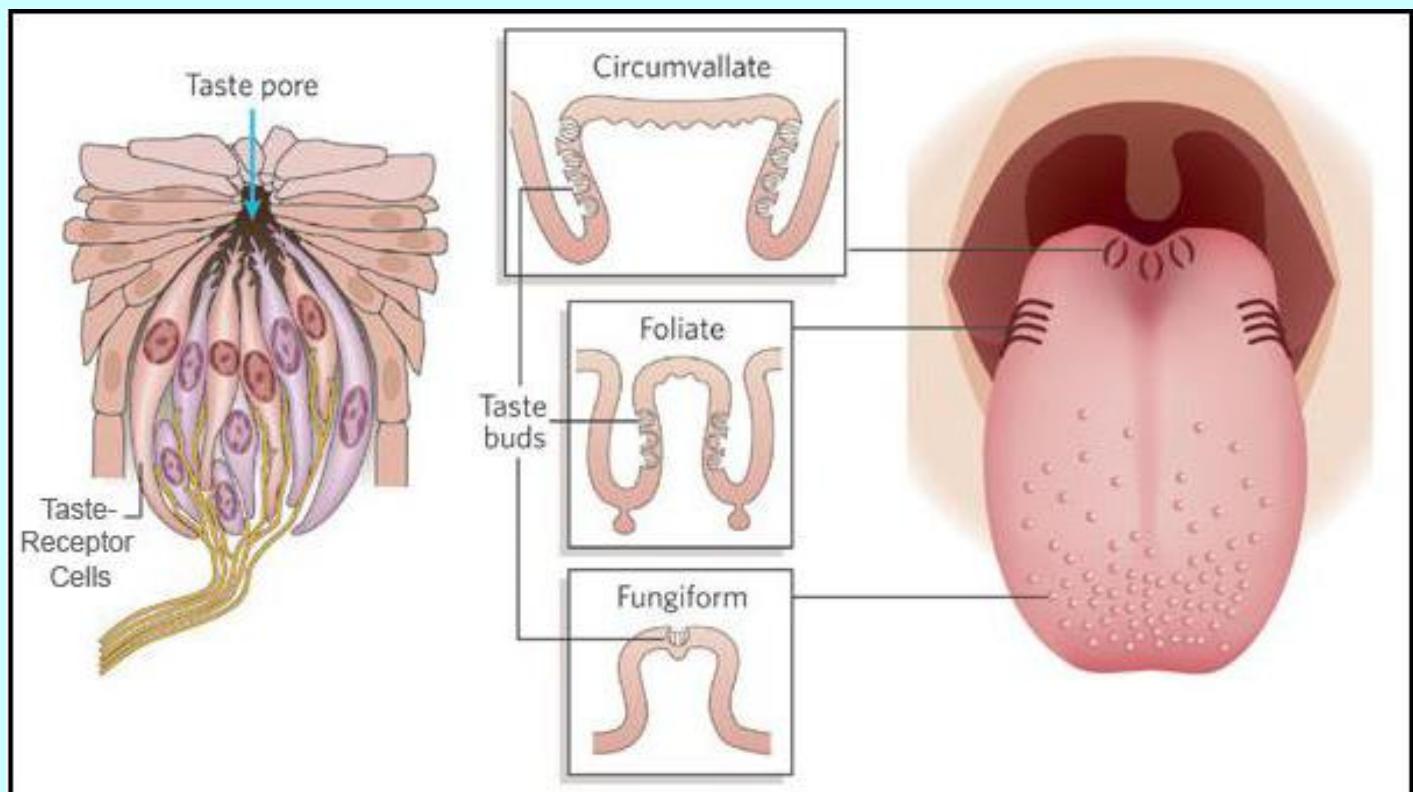
6. 味覚

■味覚

口にする物の化学的特性に応じて認識される感覚

- 塩味
- 酸味
- 甘味
- 苦味
- 旨味

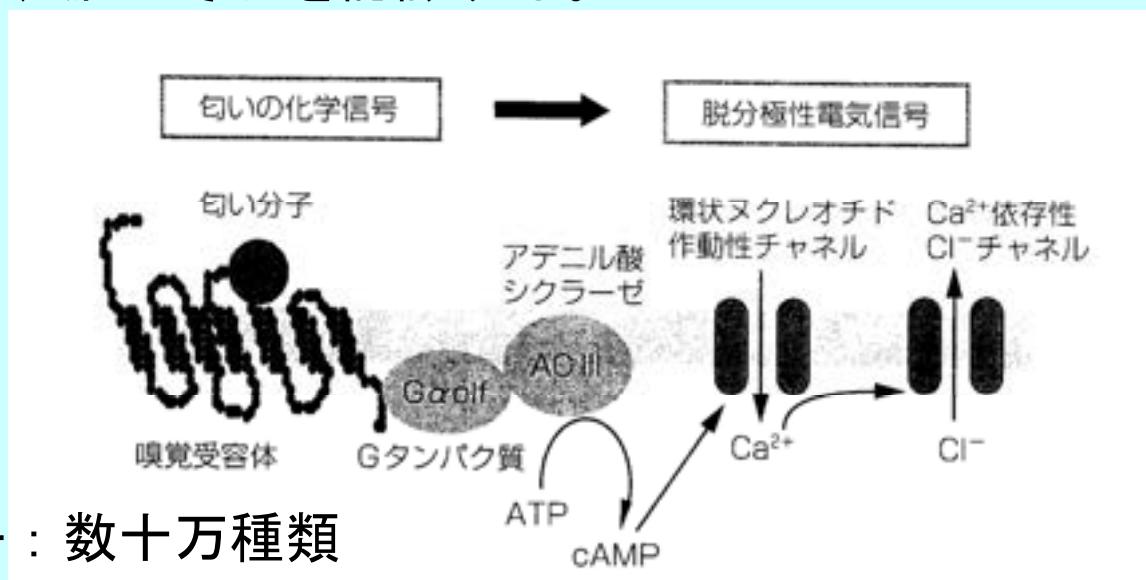




7. 嗅覚

■ 嗅覚

化学物質を受容器で受け取ることで生じる感覚のこと
ヒトにおいては鼻腔の奥にある嗅細胞により電気信号に変換し、脳でそれを認識する。



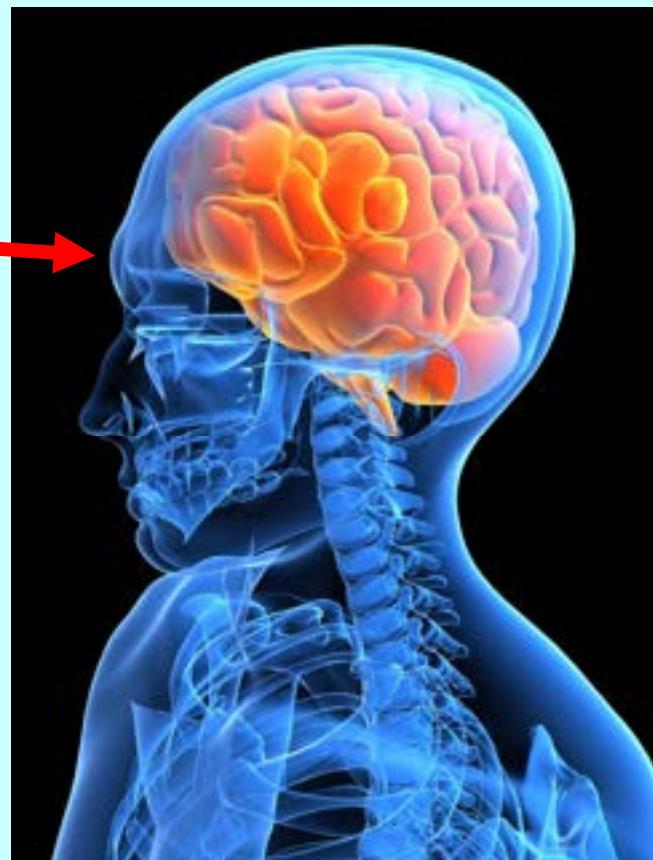
■匂い分子：数十万種類

8. マルチモダリティ

どのように知覚して
いるのか？

五感における
マルチモダリティ

外界の刺激を
あたえると



リバースエンジニアリング

- 人工視覚用デバイス
- 全方位視覚センサ
- 聴覚センサ
- 嗅覚センサ
- 触覚センサ
- 味覚センサ
- ▪ ▪ ▪

ヒトの知覚特性を利用した信号処理

I485F 知覚情報処理特論

2. 聴覚信号処理 (聴覚のメカニズム)

2023年4月15日(土)

人間情報学研究領域

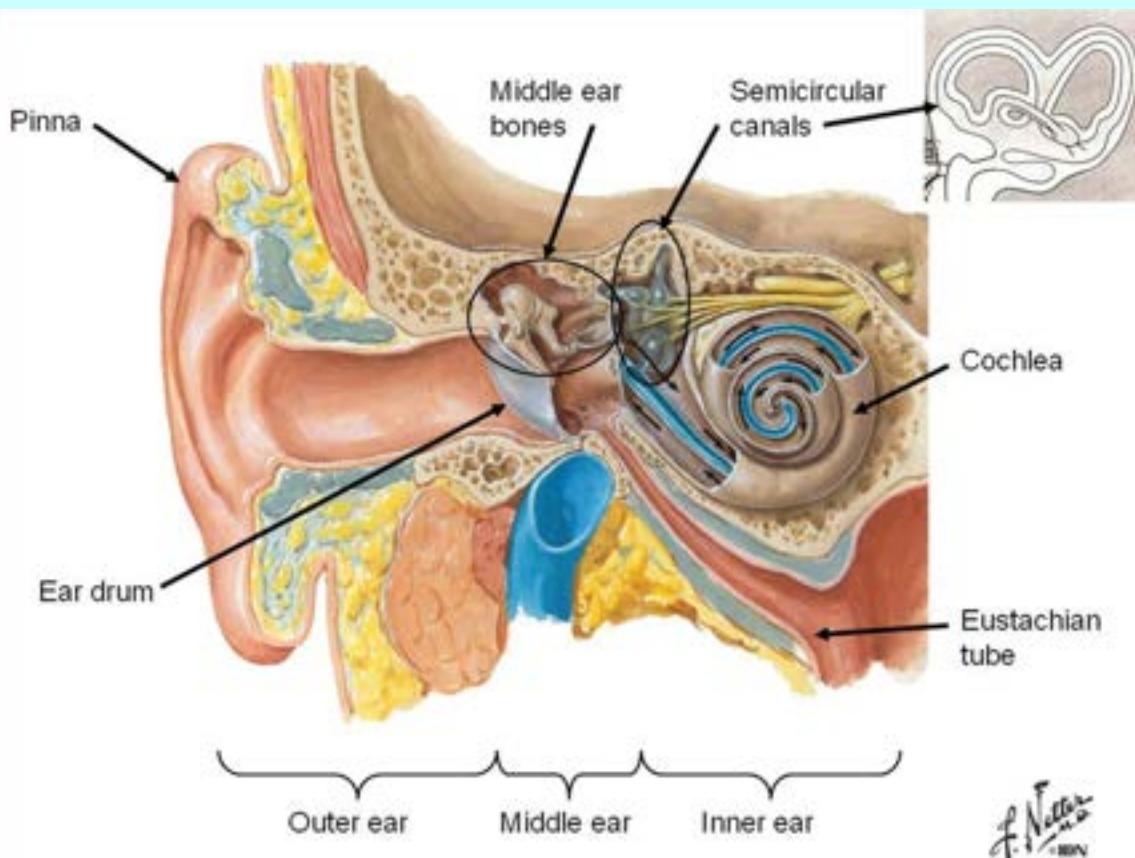
鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

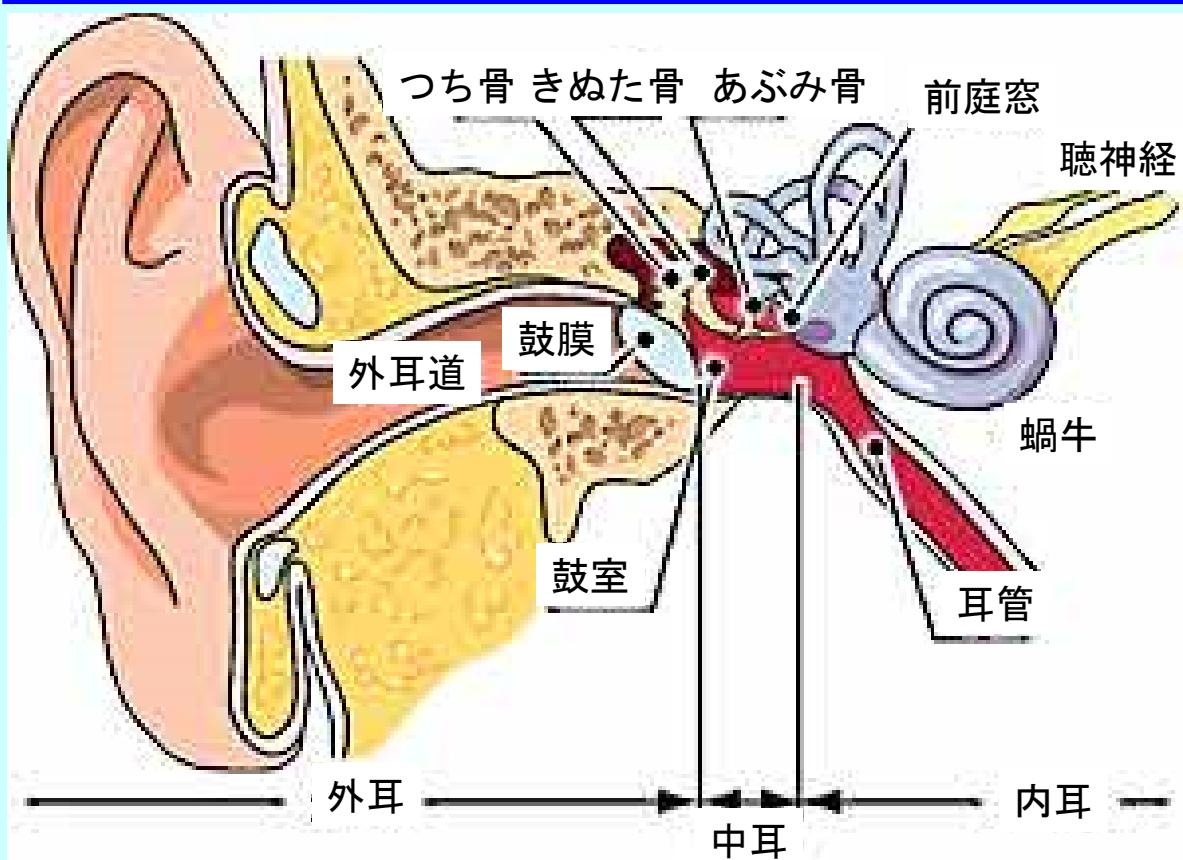
Email: unoki@jaist.ac.jp



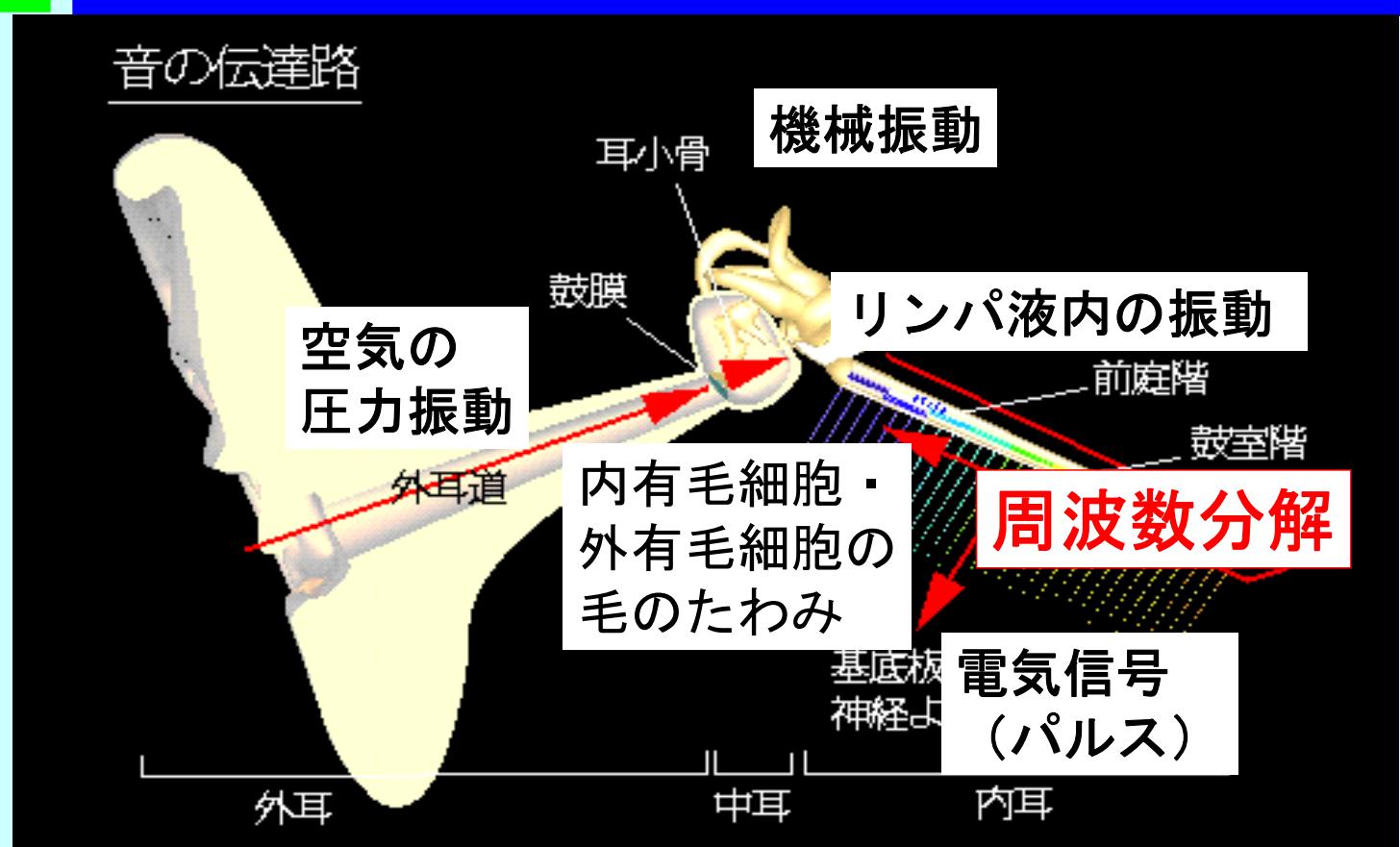
1. 聴覚系の基本的な構造と機能



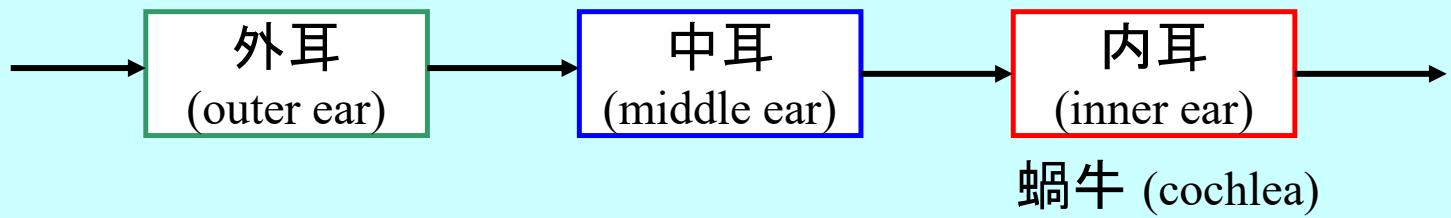
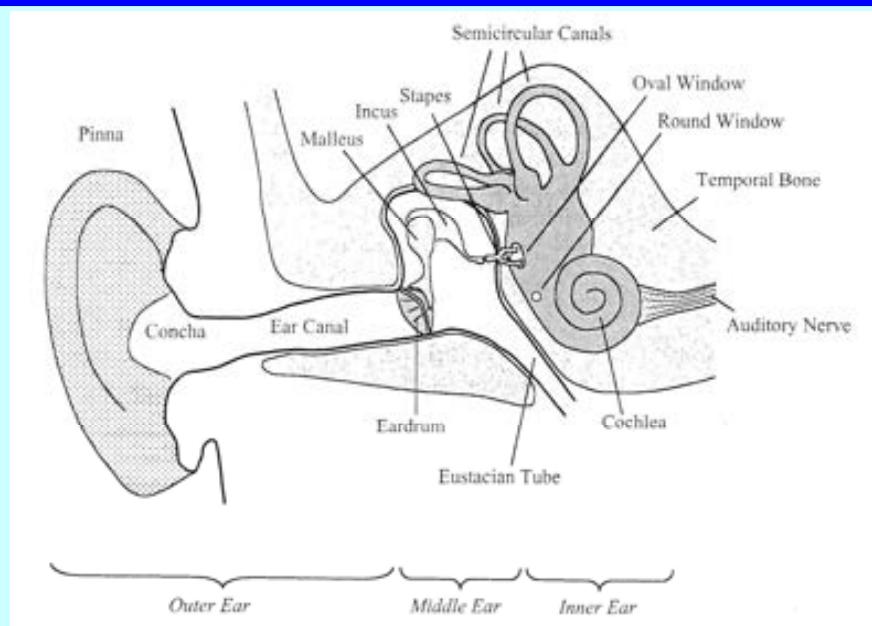
聴覚の構造



伝達経路の概要



2. 聴覚の機能の細分化

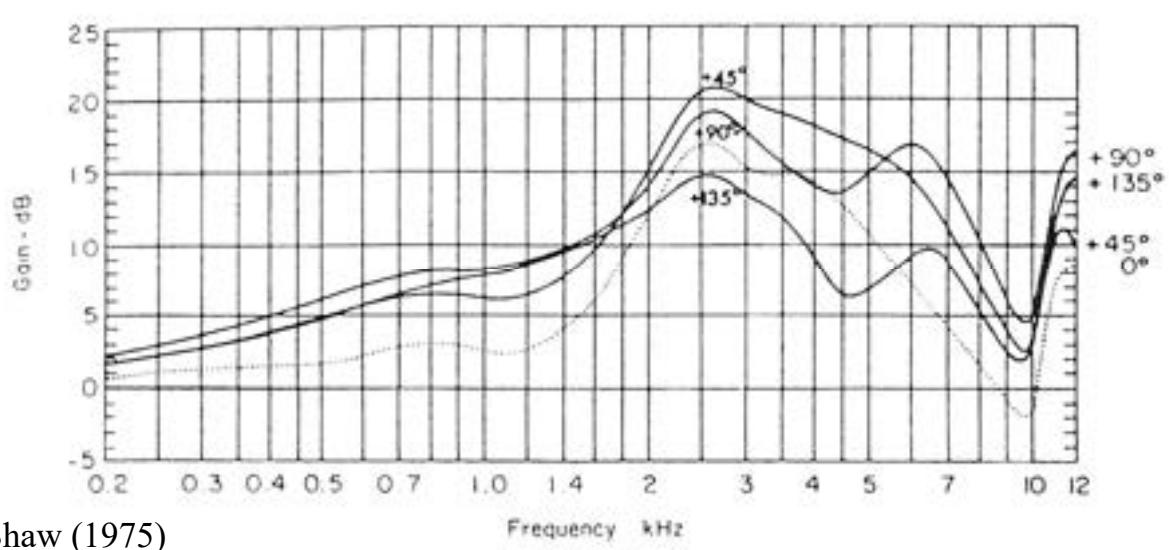


2.1 外耳

■ 構成 耳介 (pinna)

外耳道 (auditory canal)

■ 機能 高域強調→空間知覚で重要



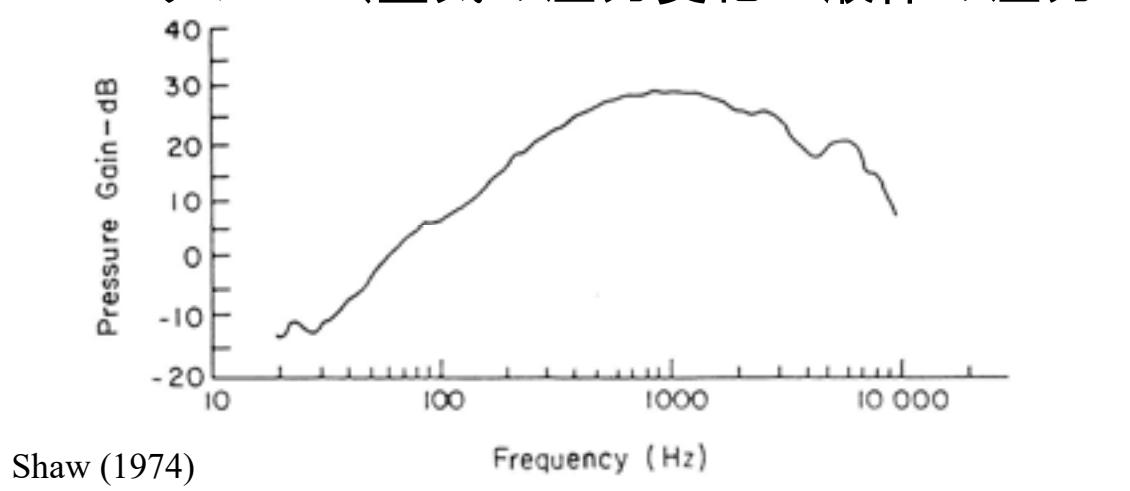
2.2 中耳

■構成 耳小骨 (ossicles; 人間の中で一番小さい骨)

- つち骨 (malleus), きぬた骨 (incus), あぶみ骨 (stapes)

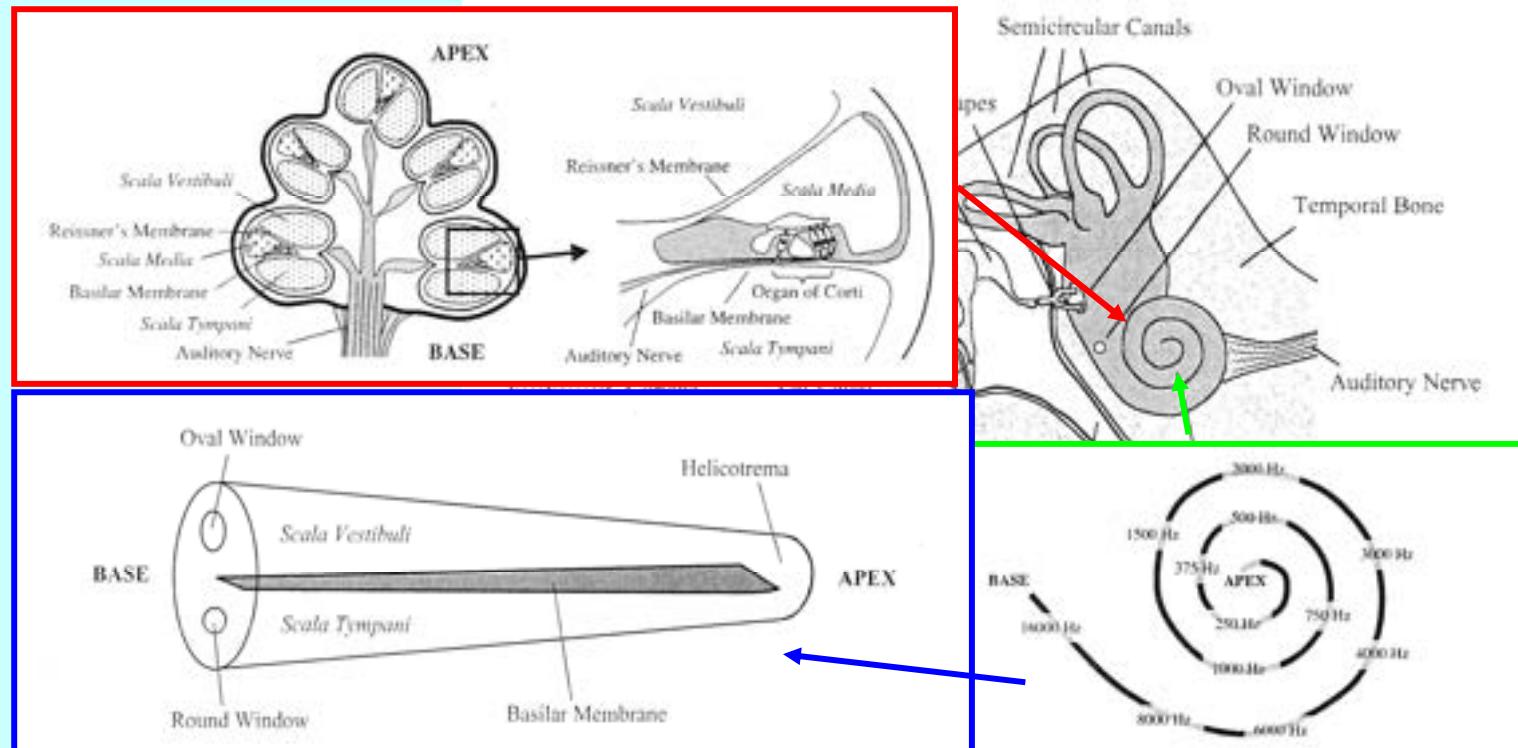
■機能 インピーダンスマッチング

- 音響インピーダンス (空気の圧力変化→液体の圧力変化)



2.3 内耳

■蝸牛 (cochlear)



蝸牛の部位

■ 二つの膜で分割

- ライスネル膜 (Reissner's membrane)

- 基底膜 (basilar membrane)

■ 二つの室で構成

始点 蝸牛底 (base)

- 前庭階 (scala vestibuli)

- 鼓室階 (scala tympani)

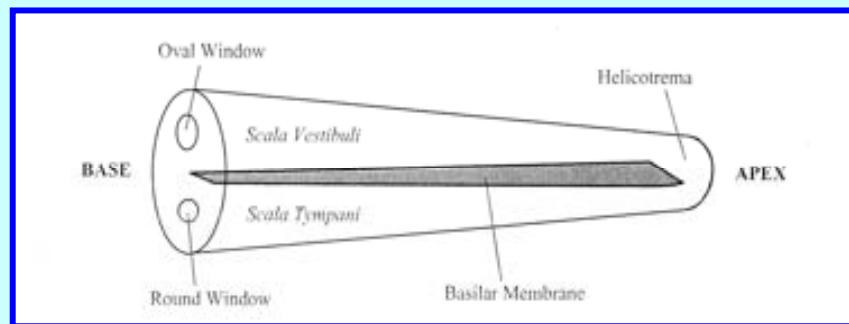
先端 蝸牛頂 (apex)

■ 入口

- 前庭窓 (oval window)

■ 開口部

- 蝸牛窓 (round window)



蝸牛の働き

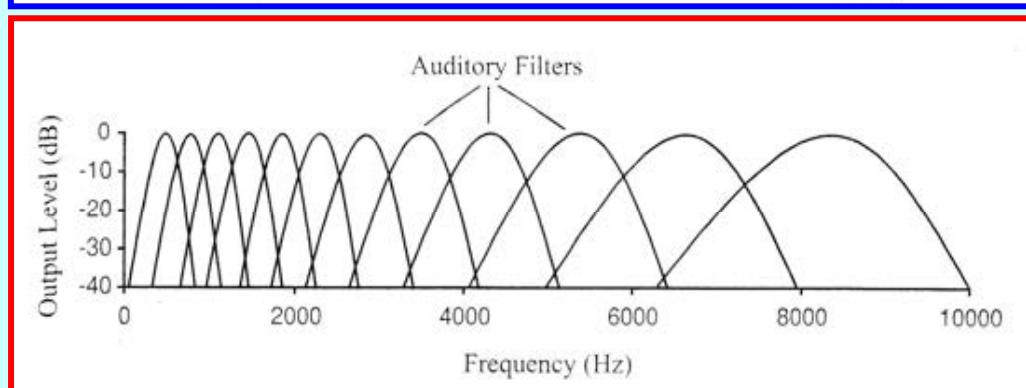
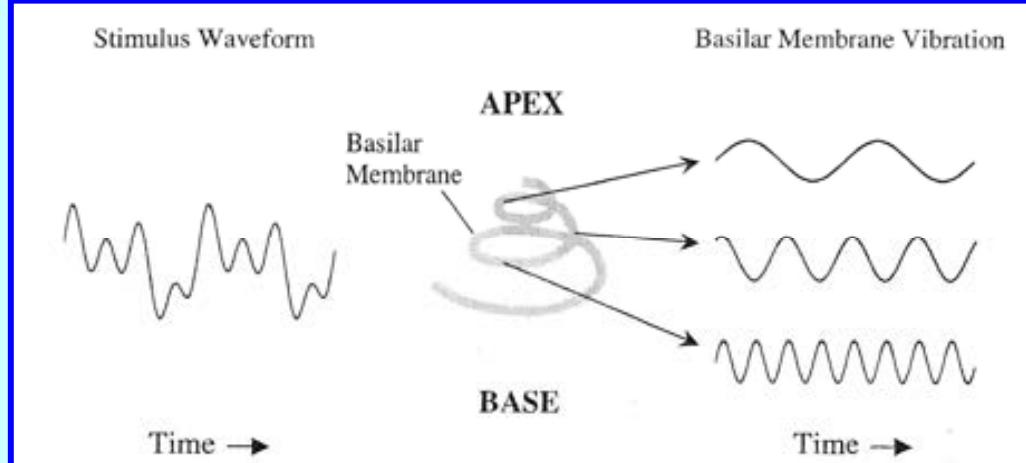
■ 周波数分解

基底膜

(basilar membrane)

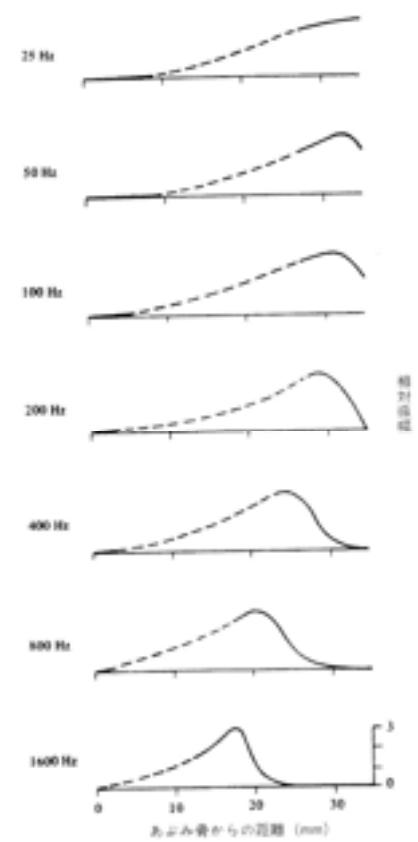
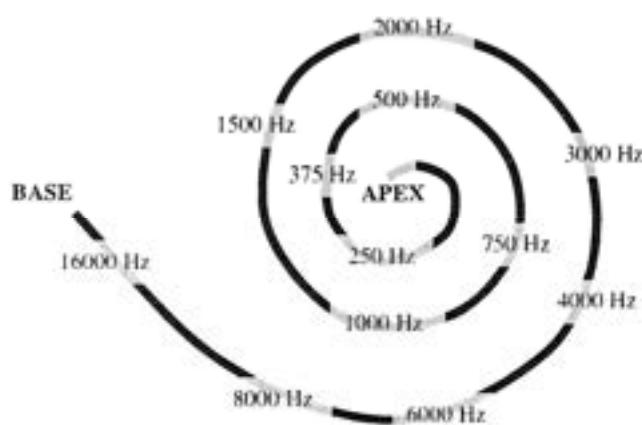
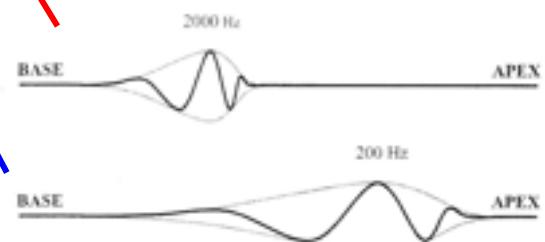
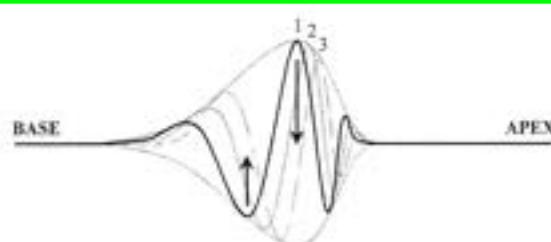
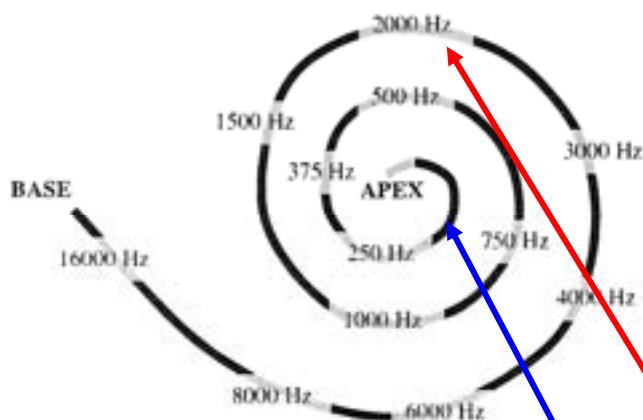
場所の情報

→周波数の情報



基底膜の働き

■ 基底膜位置の変動 → 周波数応答



基底膜振動
(basilar membrane motion)
←二つの階の圧力変動により
生じる
周波数に応じて遅延が見られる

基底膜振動パターンの共振特性

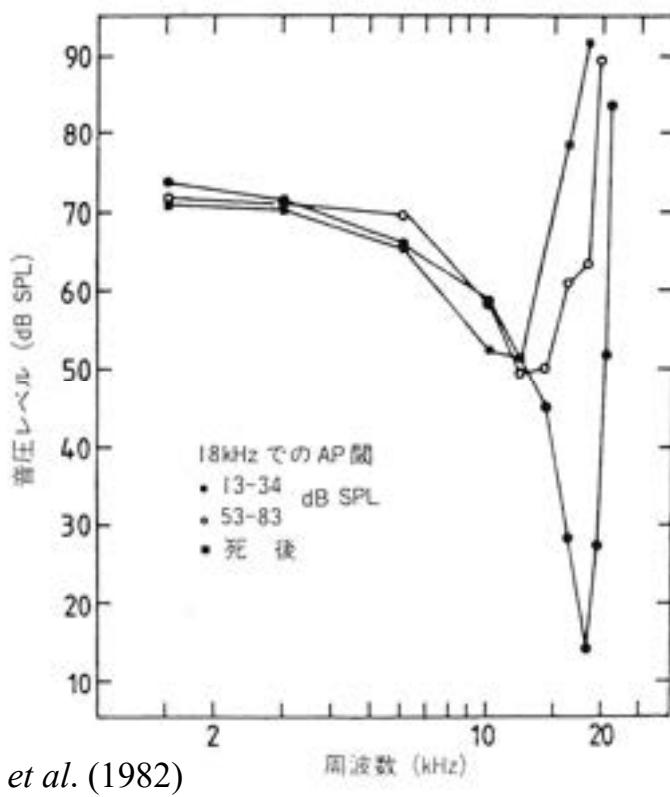
■ 同調特性

(tuning curve)

同調の度合い

→ Q値 ($Q_{10\text{dB}}$)

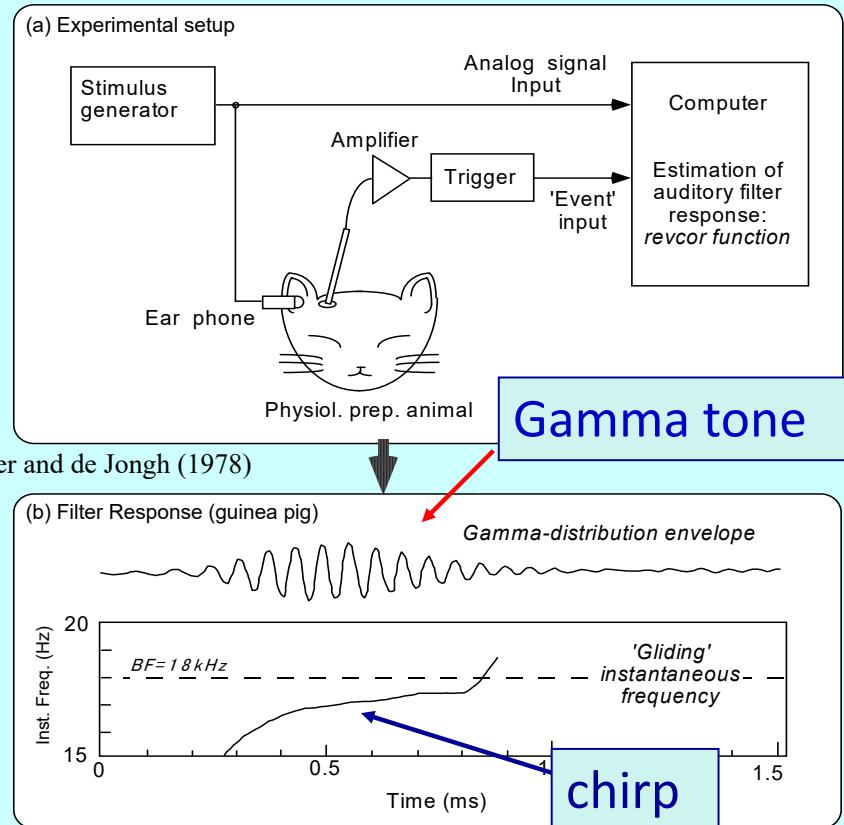
3~10

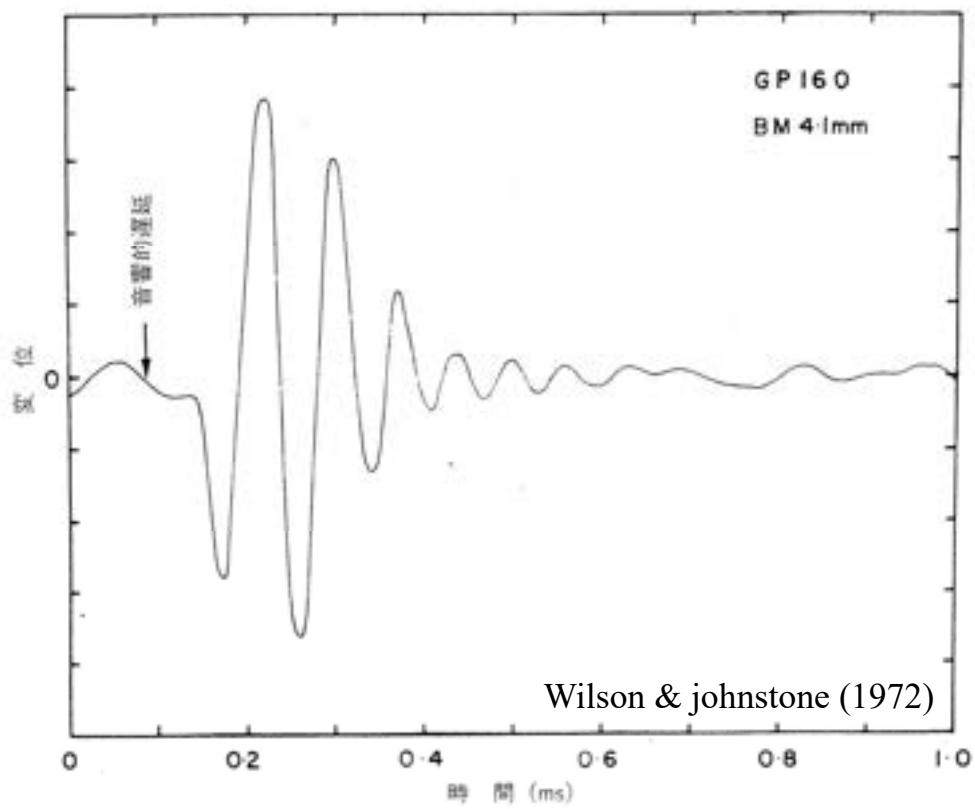


基底膜のインパルス応答関数

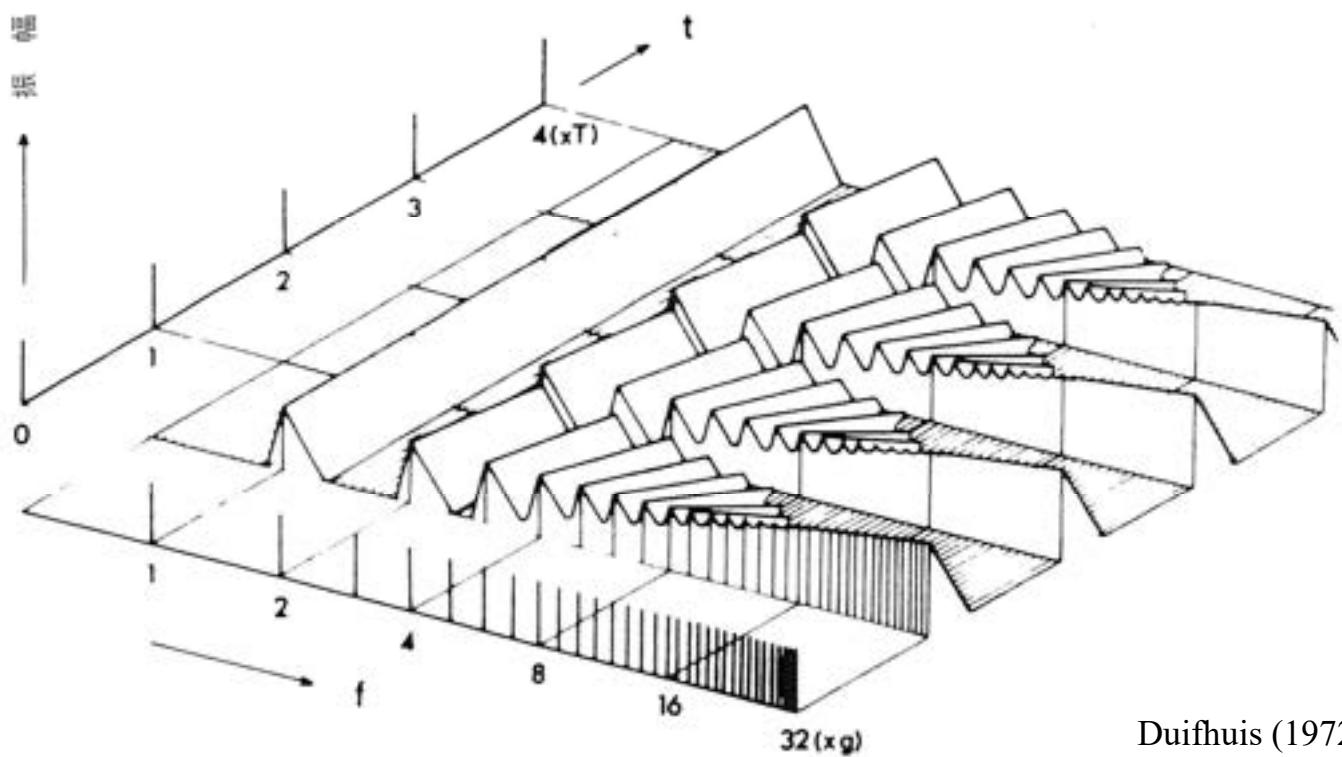
■ 測定法

- メスバウアー効果
を利用するもの
- レーザー干渉法



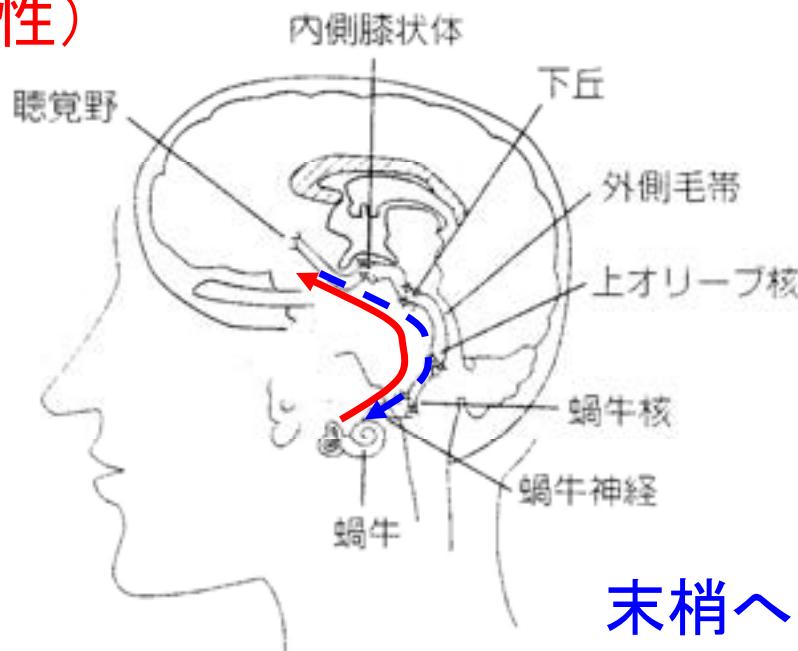


基底膜応答の図式化



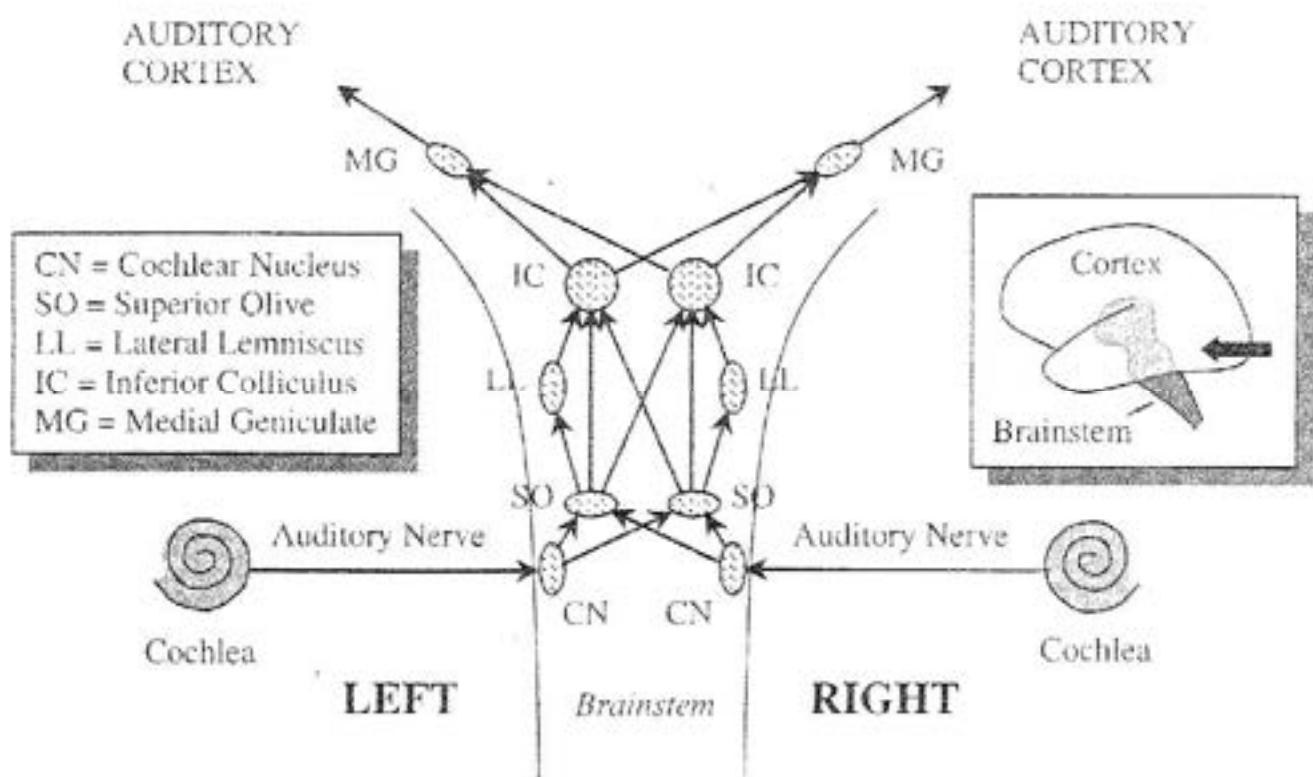
3. 聴覚中枢系

聴覚野への伝達経路 (求心性)

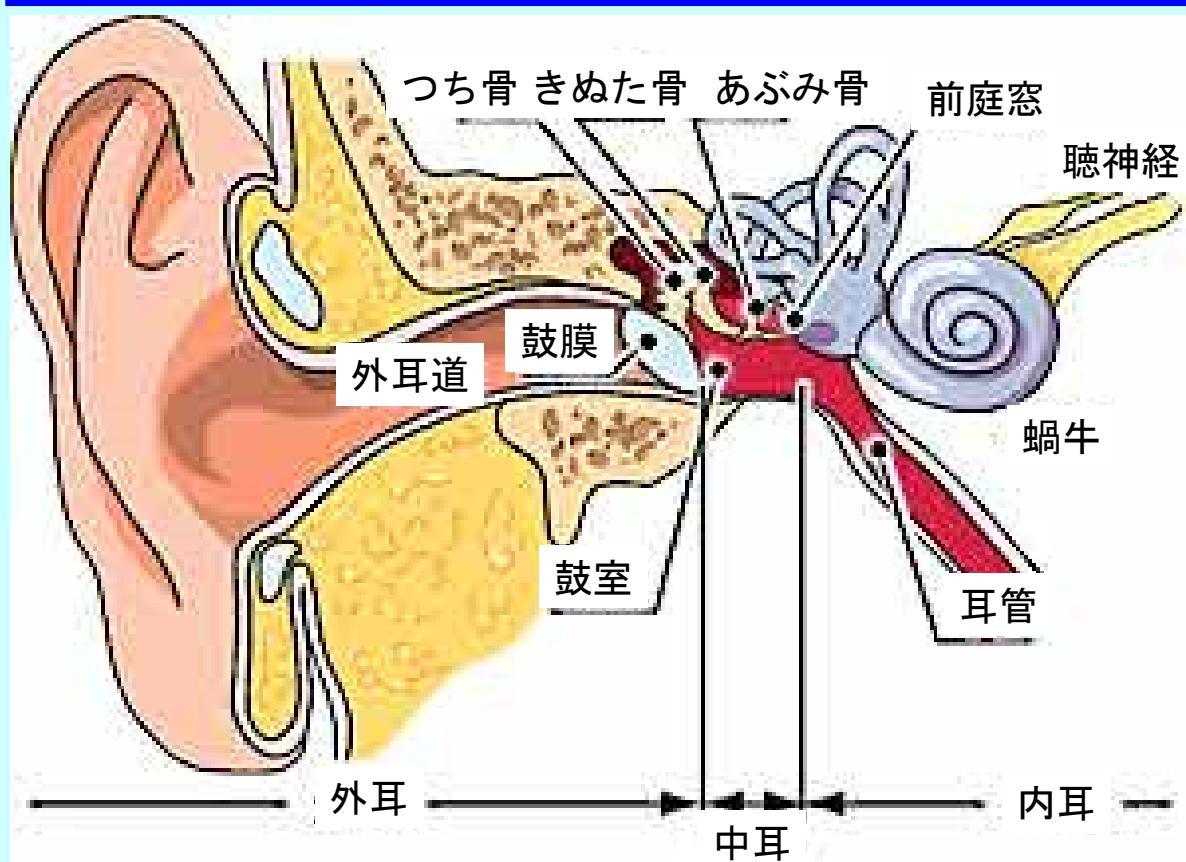


末梢への伝達経路 (遠心性)

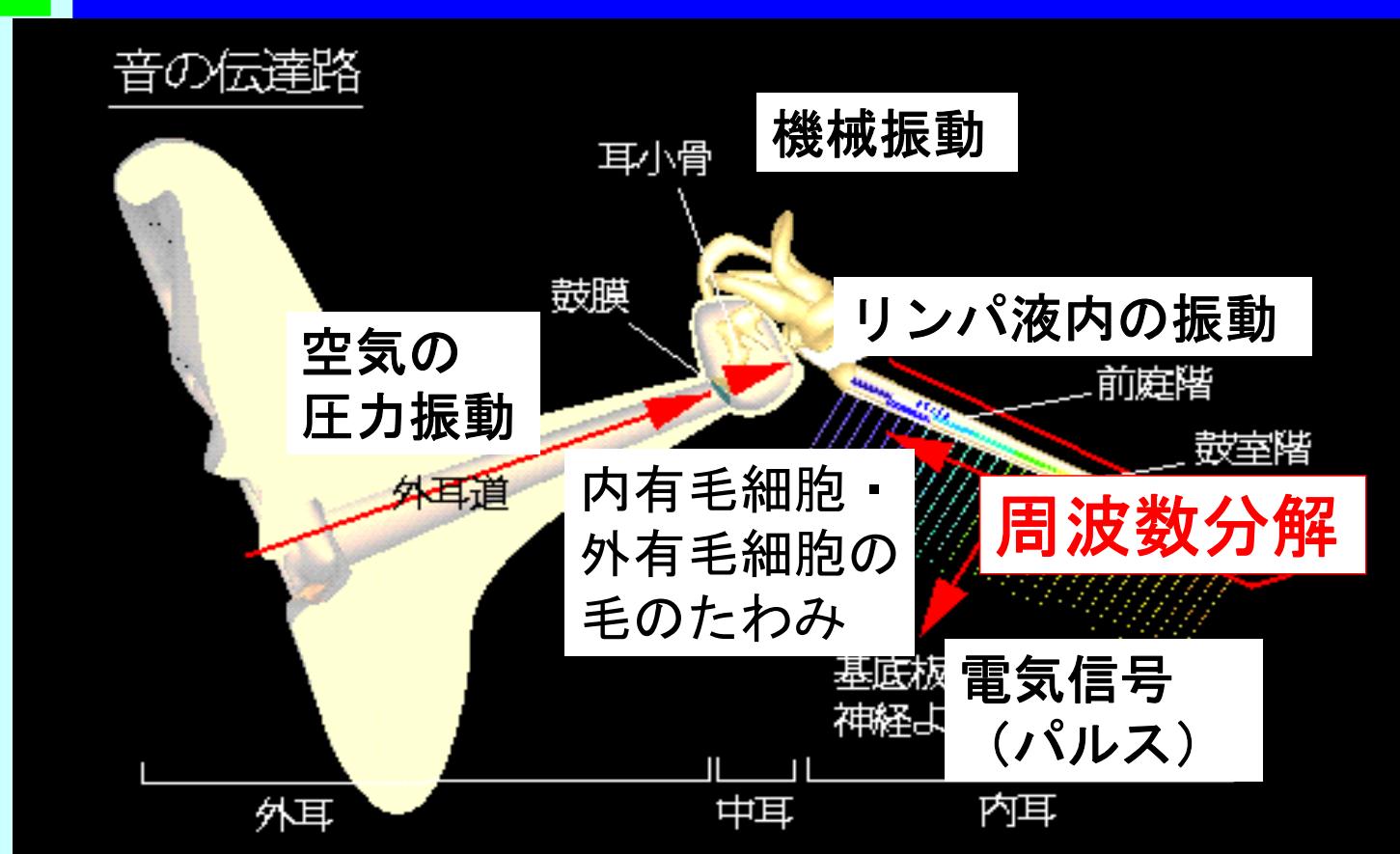
Auditory pathways

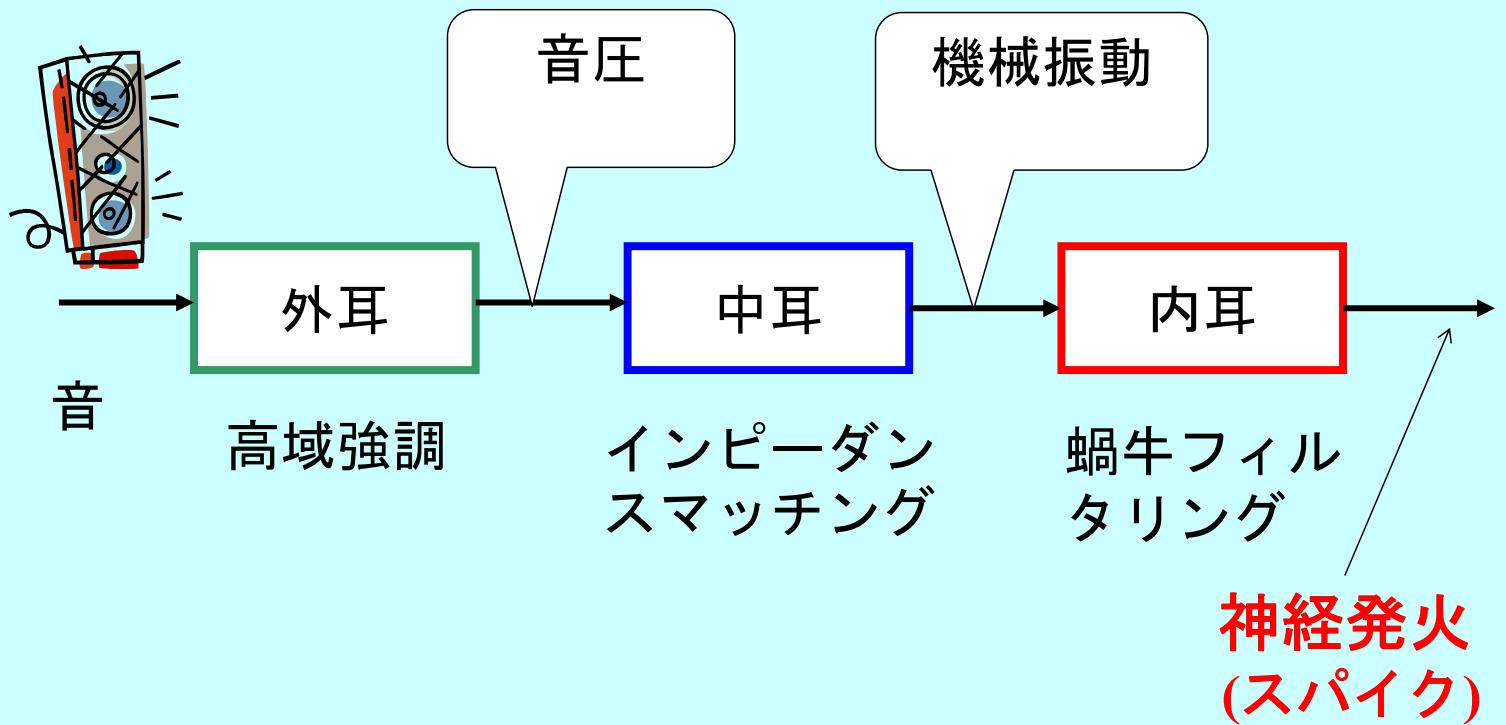


4. 聴覚の機能的なしくみ

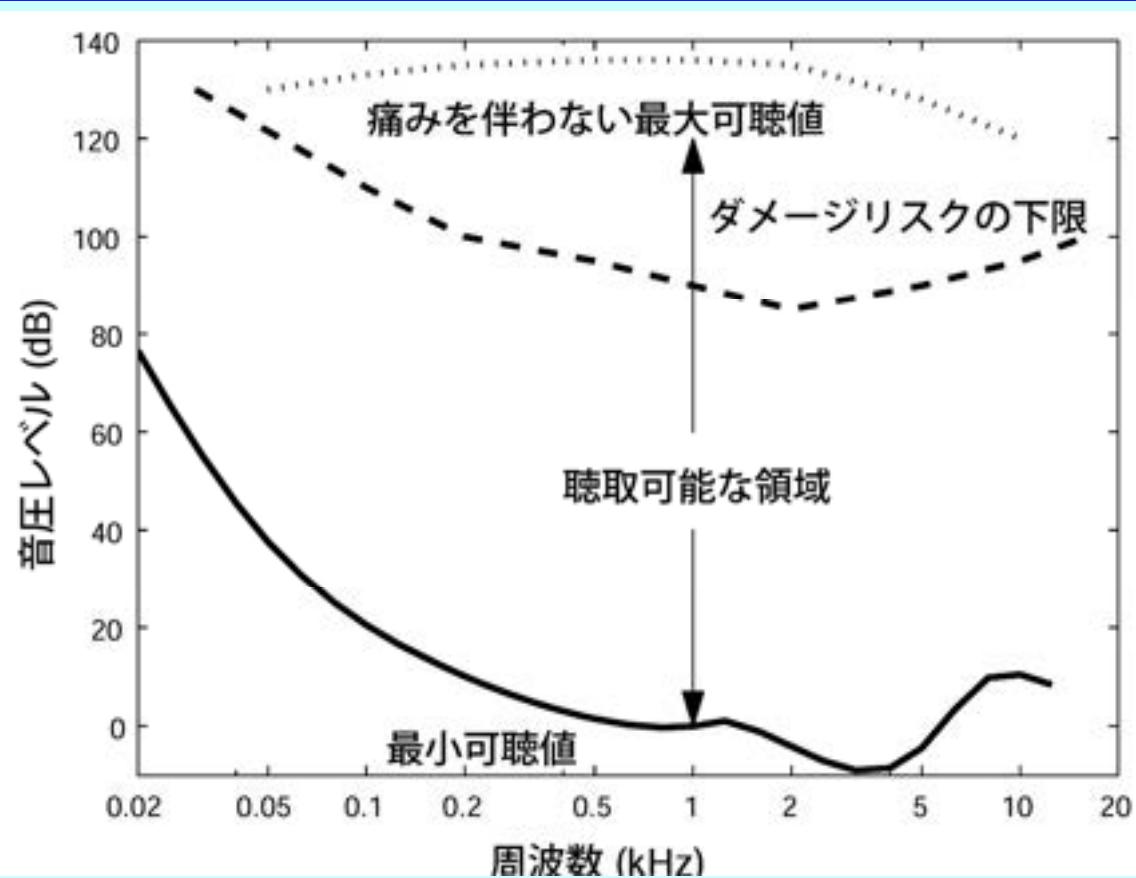


どうやって音を聴いているんだろう…？？？





可聴周波数帯域



デモ YouTube

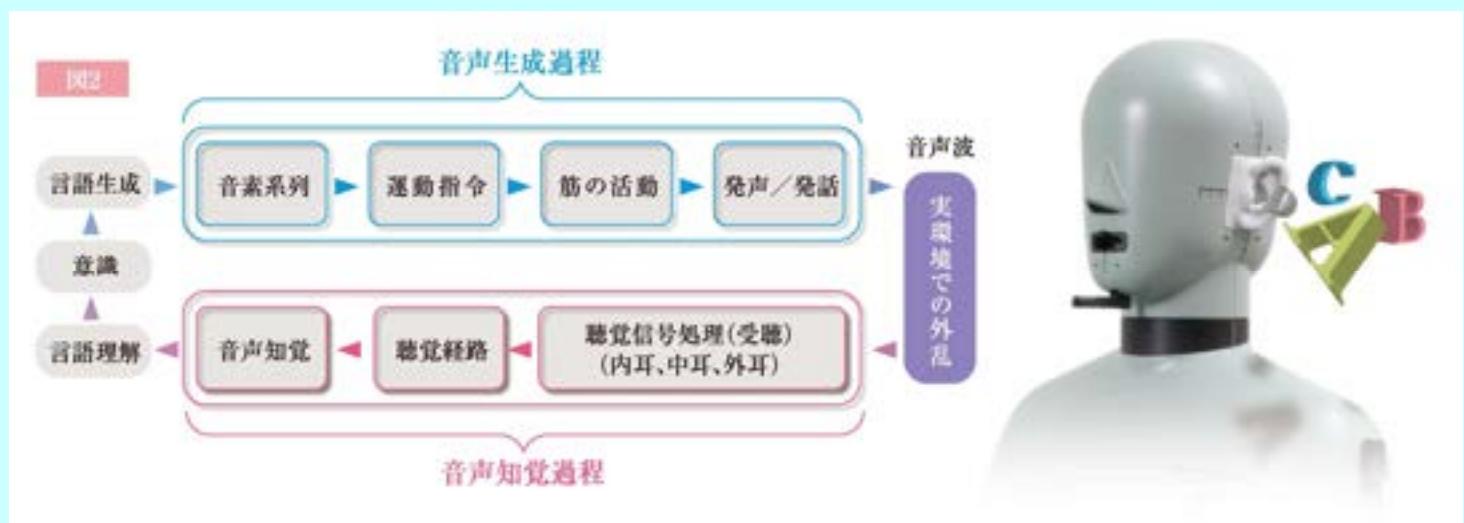
■ Auditory System:

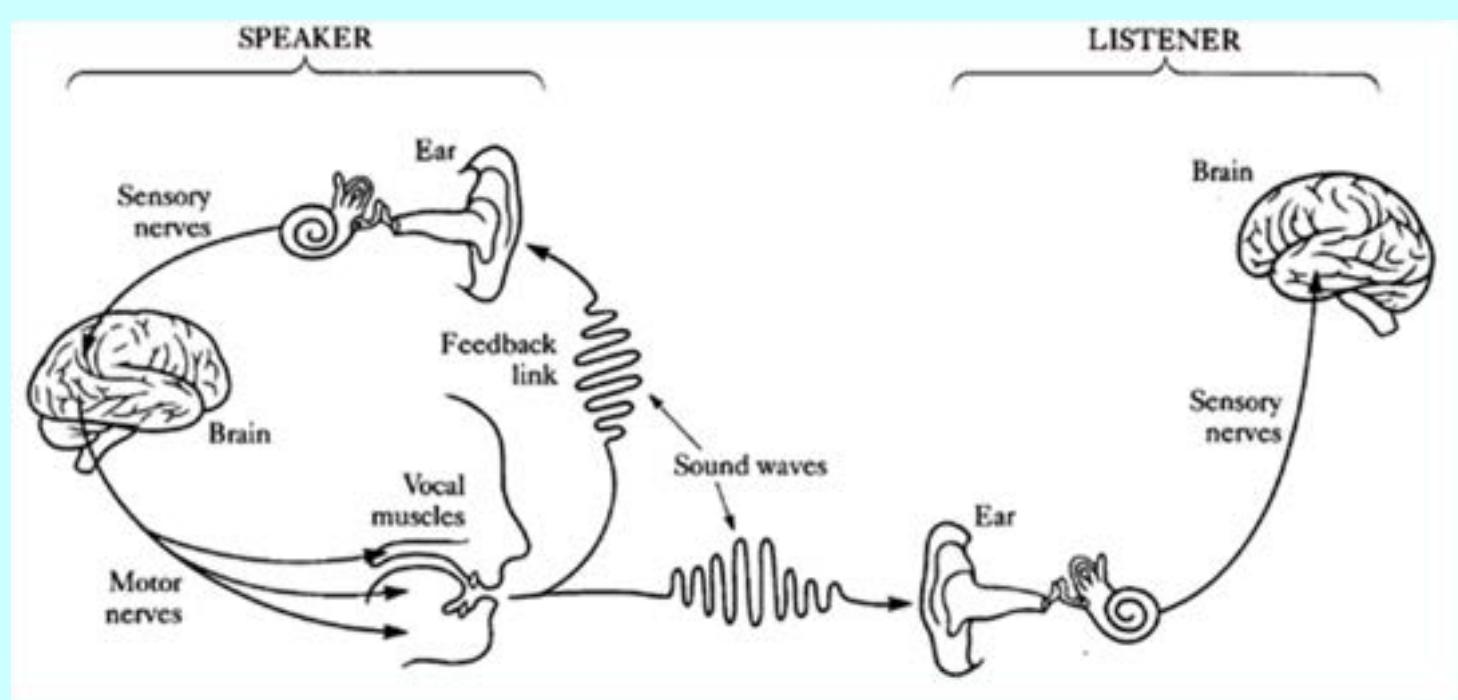
<http://www.youtube.com/watch?v=IGSK1iPSJ7I&feature=related>

■ Auditory Transduction:

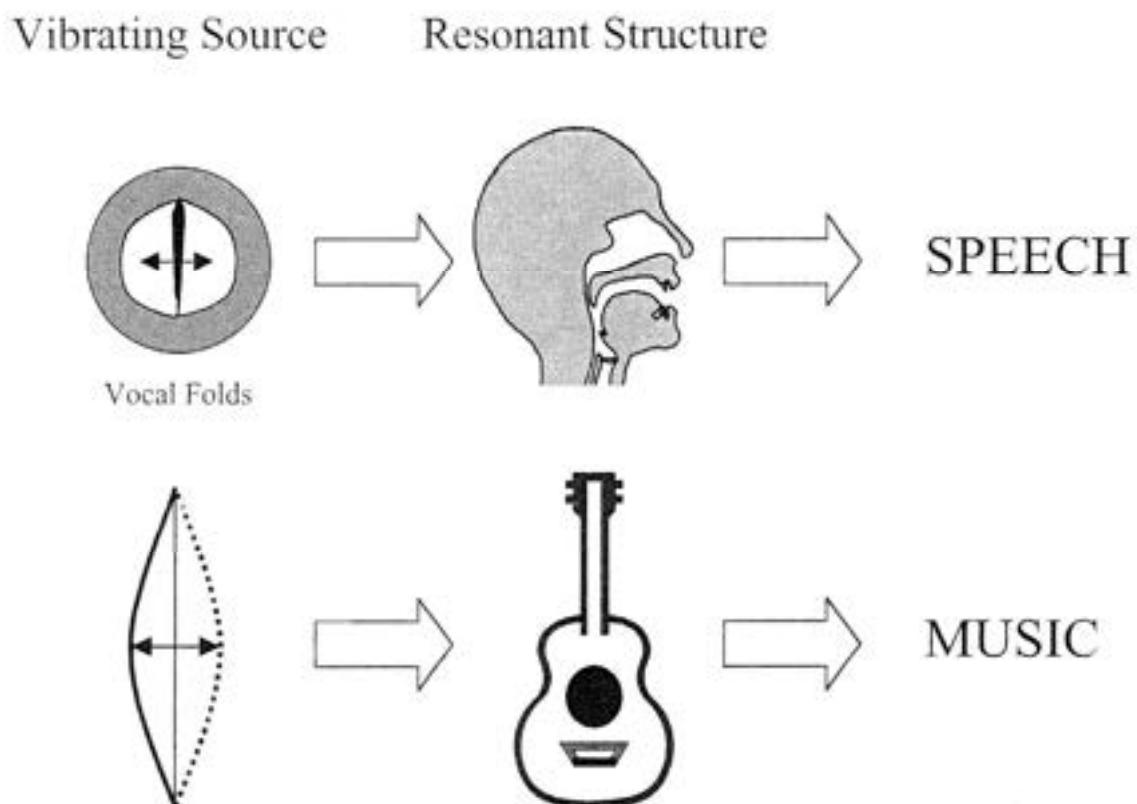
<http://www.youtube.com/watch?feature=endscreen&NR=1&v=PeTriGTENoC>

5. 音声生成と音声知覚





音源の特性と音声発話



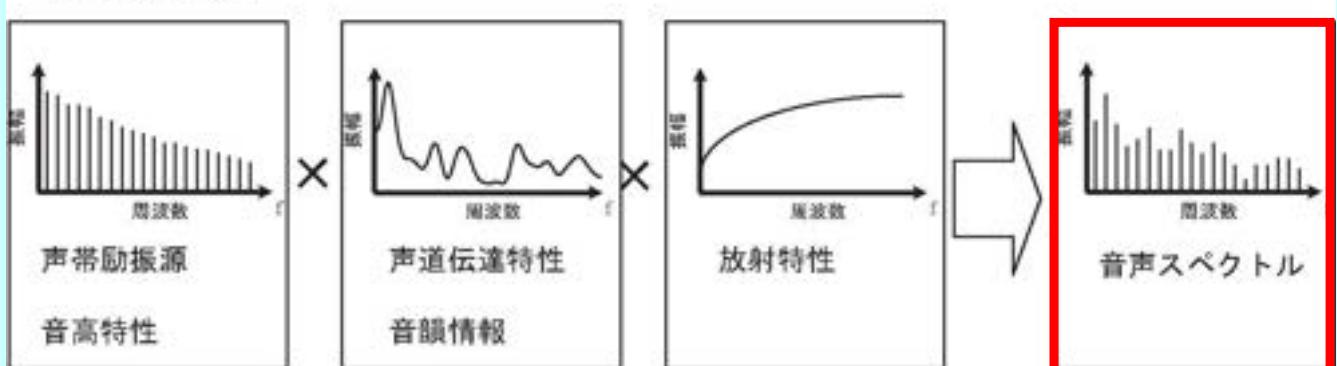
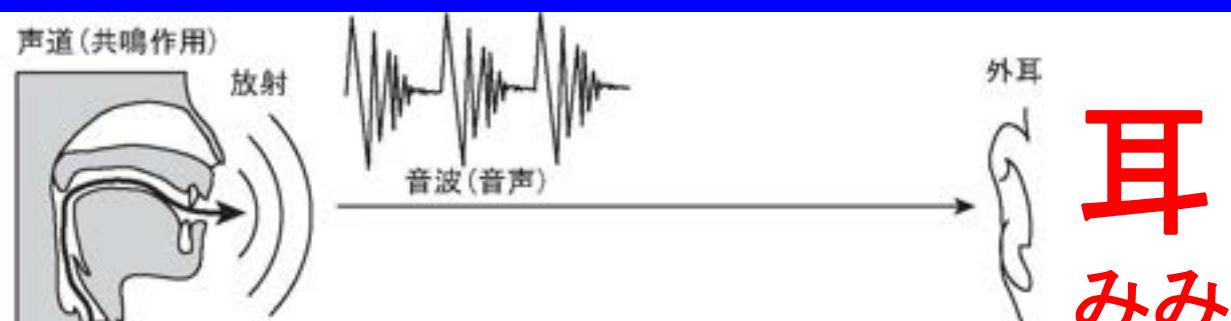
発声のしくみ



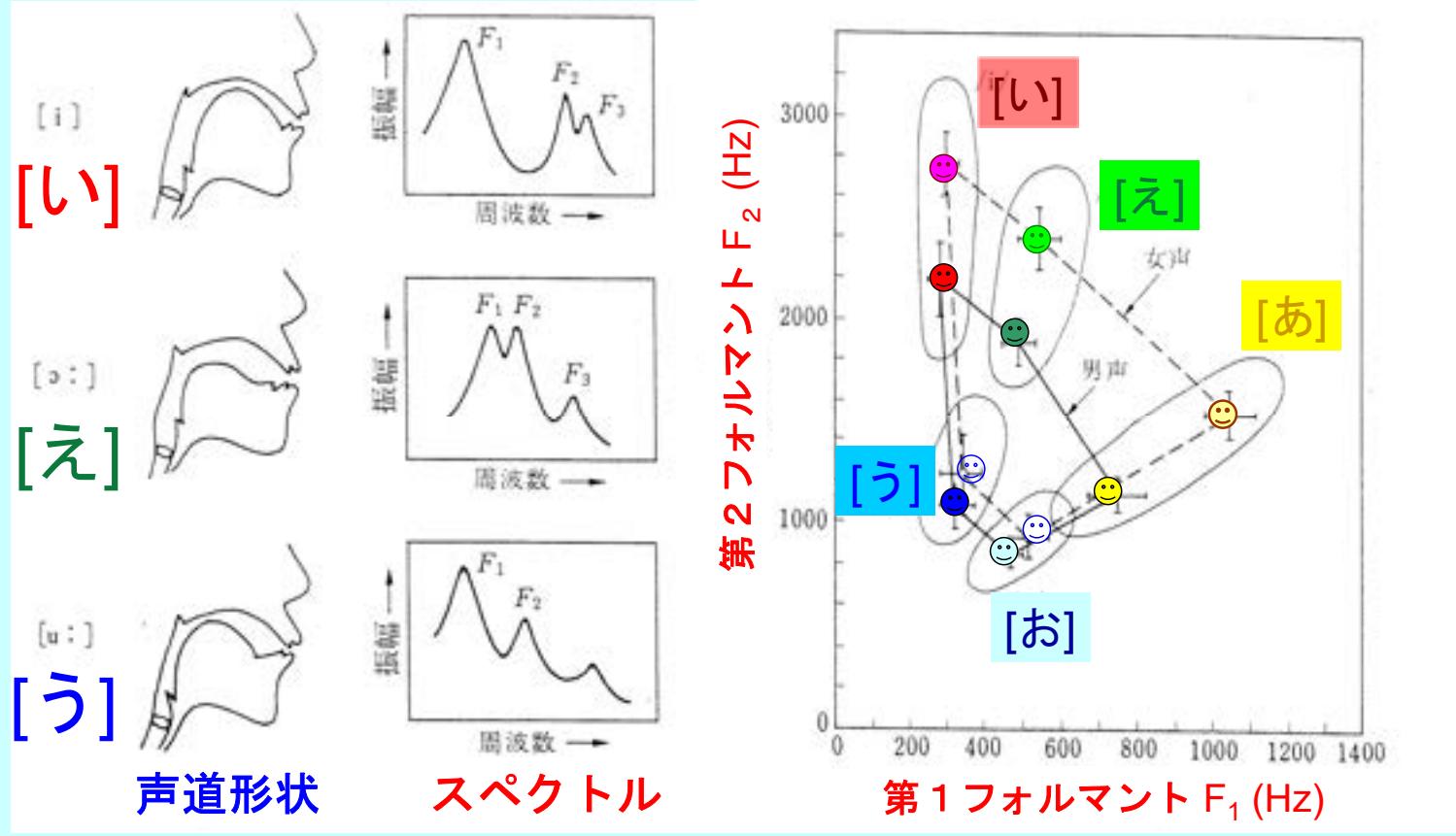
- 肺に溜めた空気を押し出し
- 空気が声帯を振動させて母音を発生させて
- 口腔および鼻腔で特定の周波数を共鳴または減衰させて
- 唇、舌、歯などで子音を付加した空気の振動を大気中に伝播させる

音声の伝播 ⇒ 波が耳へ伝わる

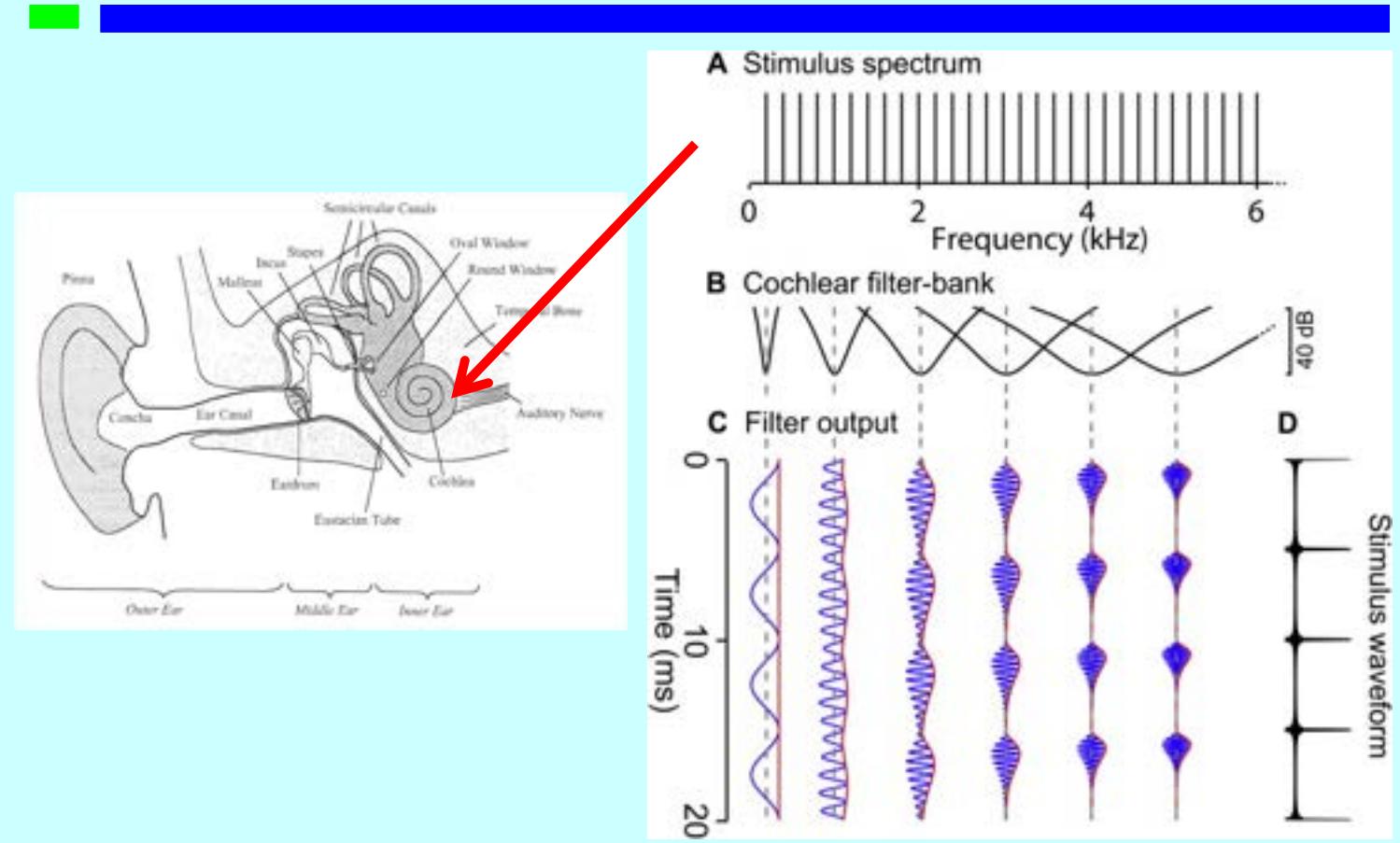
声
こえ



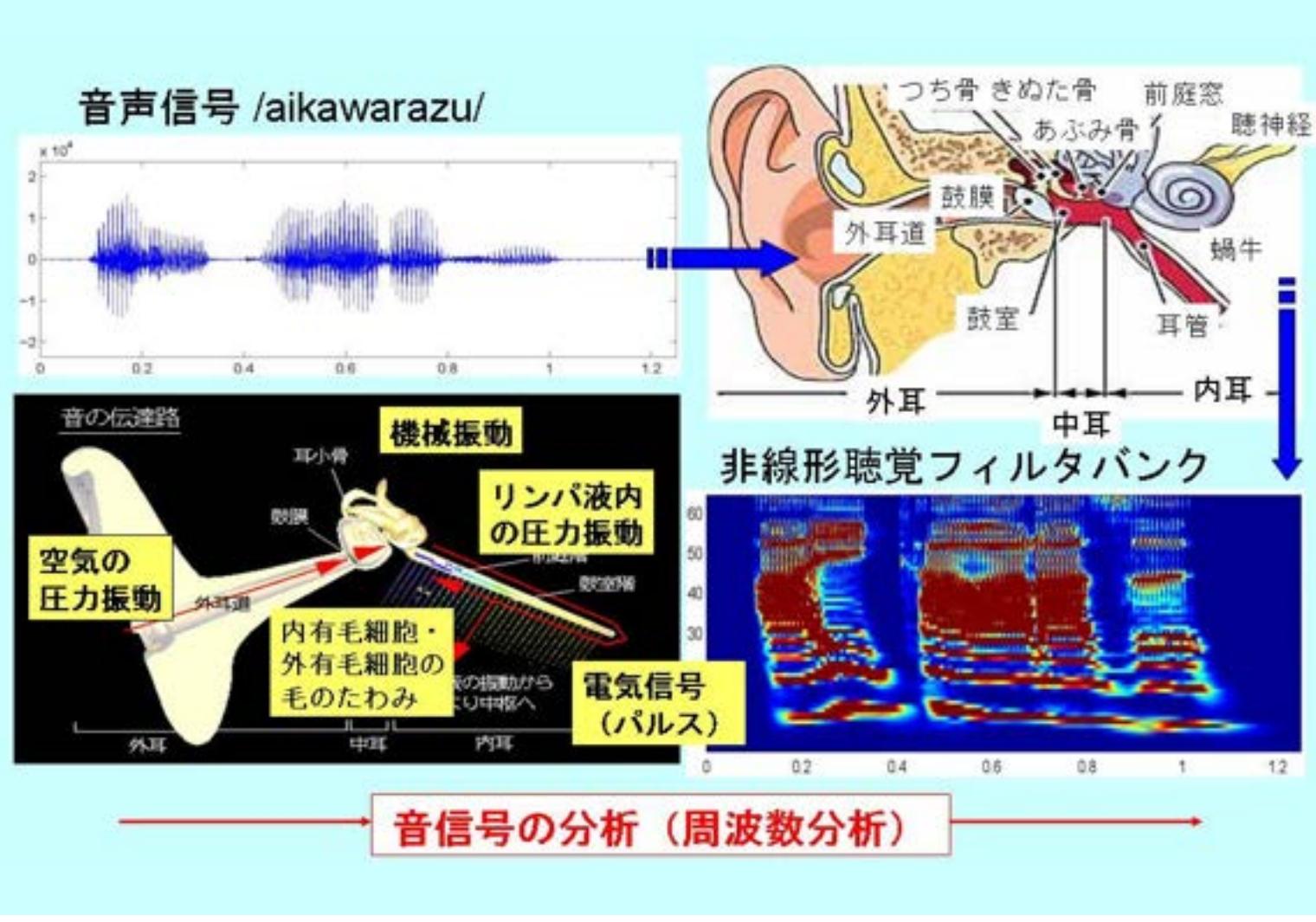
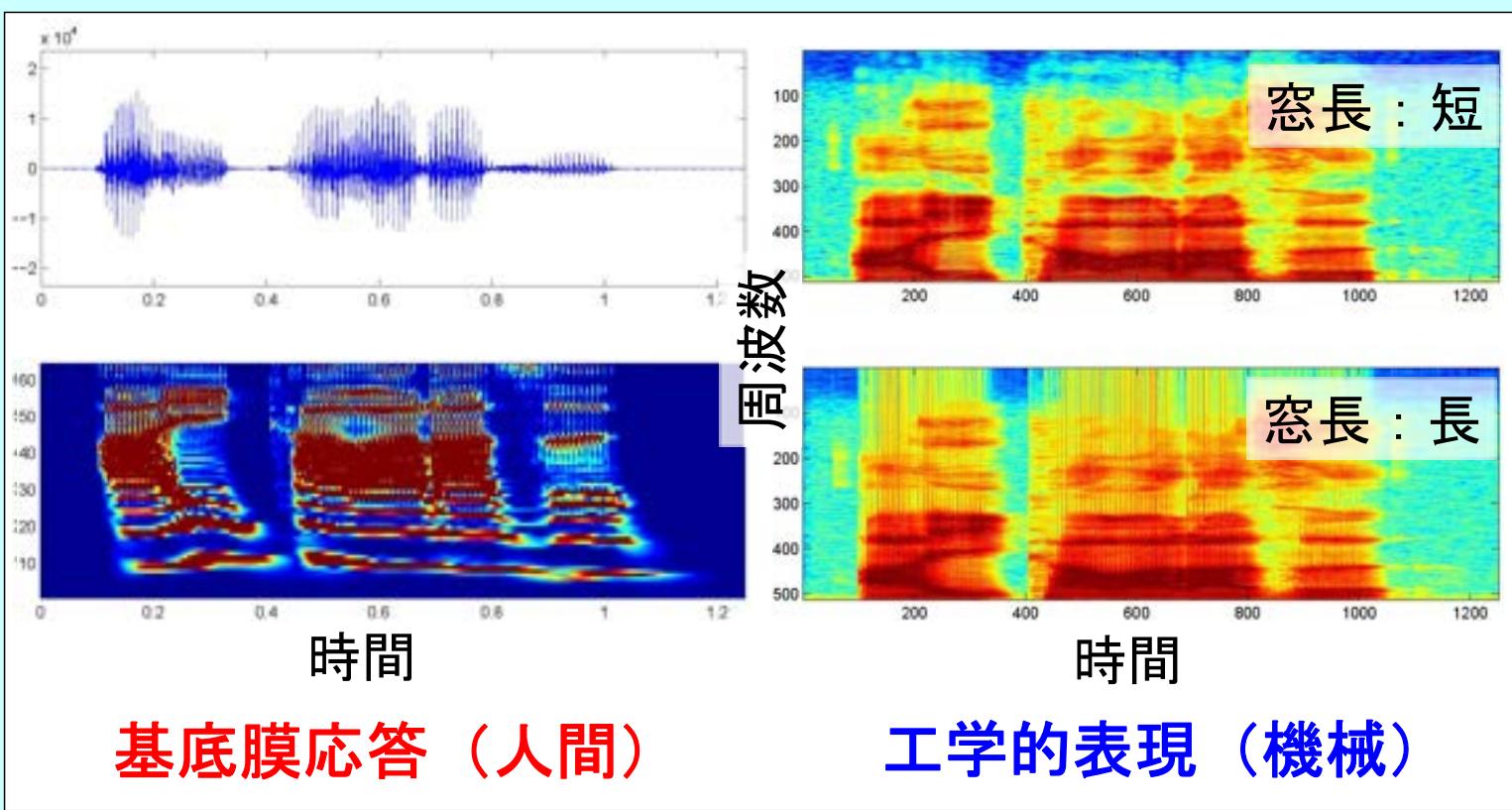
声道形状とスペクトルの関係



音声知覚



6. 音の情報表現（人間 vs. 計算機）



I485F 知覚情報処理特論

3. 音の物理表現と信号処理

2023年4月16日(日)

人間情報学研究領域

鶴木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

Email: unoki@jaist.ac.jp

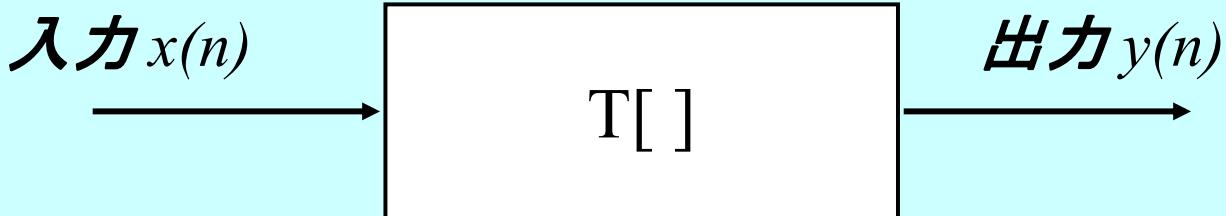


1. 導入

■ 質問1: 線形と非線形の違いは何か?

システム (処理)

$$y(n) = T[x(n)]$$



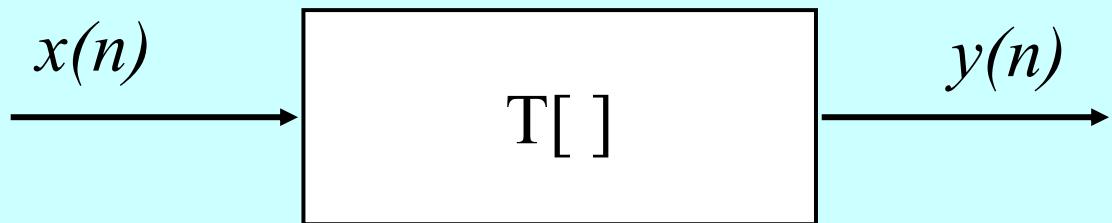
質問2: どのが線形処理でどのが非線形処理?

Fourier変換, Laplace変換, wavelet変換, z変換,
音響システム, 聴覚

2. 線形性の概念

■ 線形性の定義

$$y(n) = T[x(n)]$$



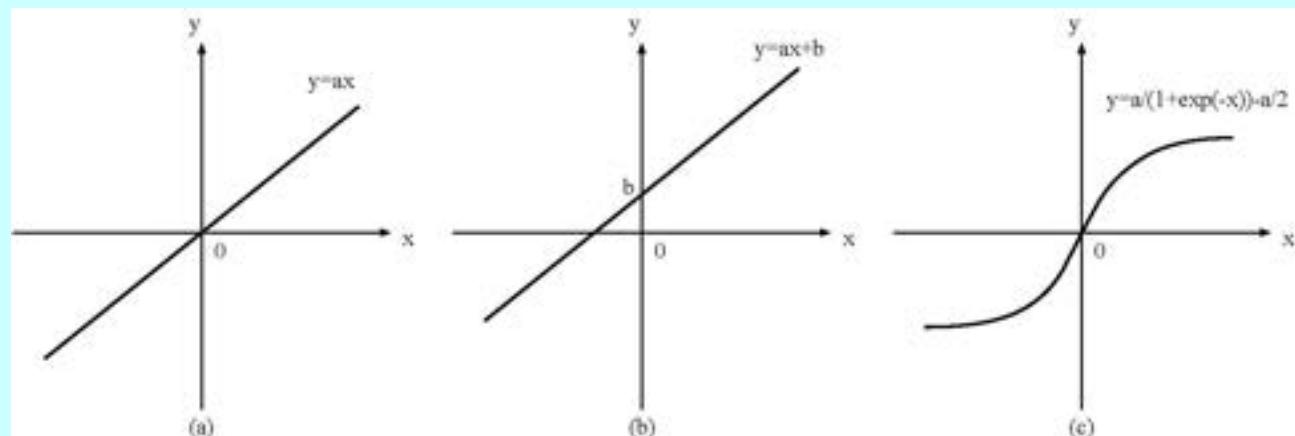
$$\begin{aligned} T[ax_1(n)] + T[bx_2(n)] &= aT[x_1(n)] + bT[x_2(n)] \\ &= ay_1(n) + by_2(n) \end{aligned}$$

■ 必要十分条件

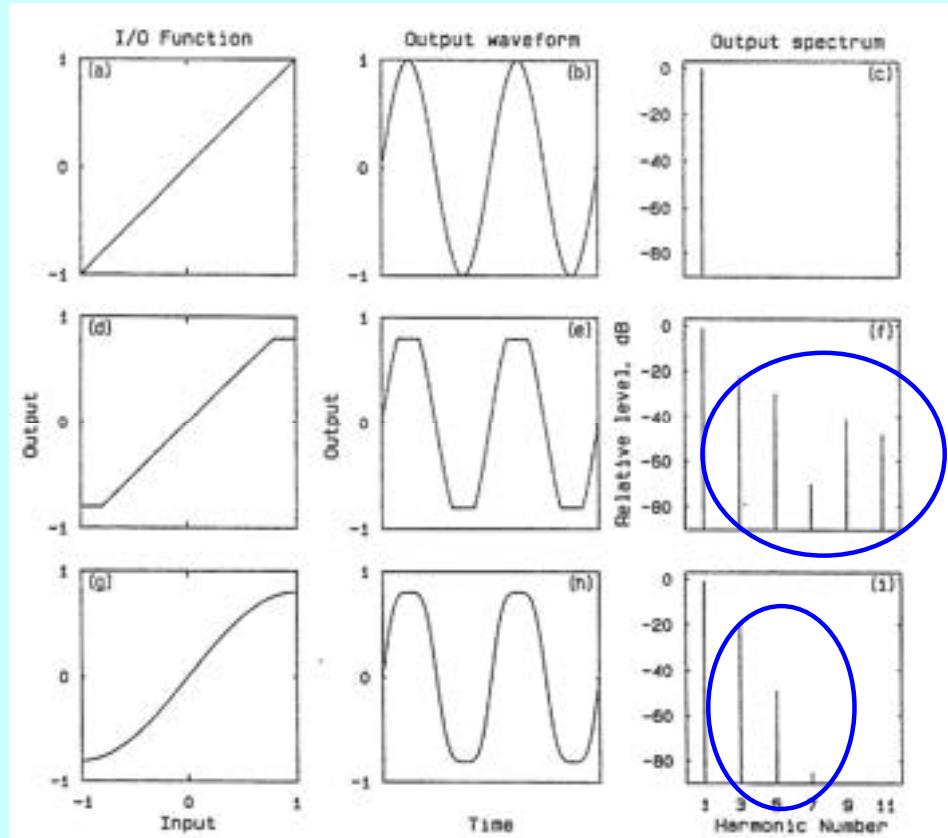
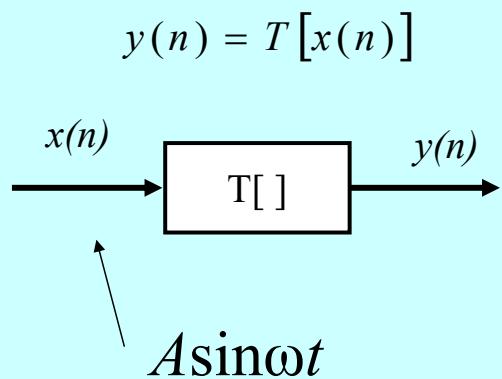
- 加法性 (Additivity)
- 同時性 (Homogeneity)
- 重ね合わせ (Superposition)

それ以外は
非線形

■ 質問: どれが線形性?

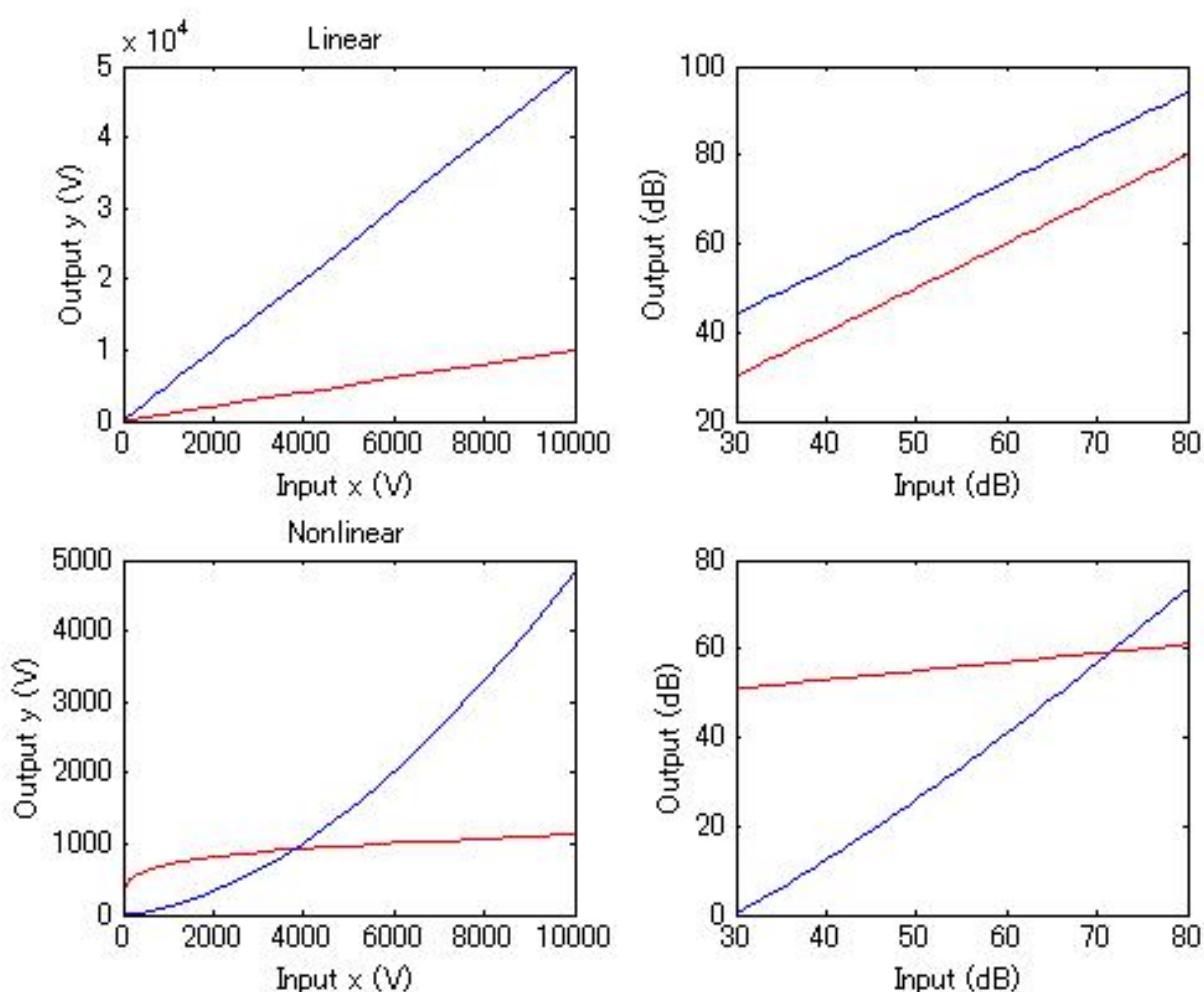


例題



質問:

なぜ調波性が
みられるのか?



3. 音

■ “音 (sound)”

- 音波（物理的なもの）
- 音波によって引き起こされた感覚（心理的なもの）

■ 聴覚心理学 (Psychoacoustics)

- 音の物理的特性とそれが引き起こす感覚との対応

■ 聴覚における知覚の側面

- 聴覚系の解剖学的知見
- 聴覚系の生理学的知見

⇒ 聴覚心理学（行動学的な研究）からの解説

身の回りの音



■ 音声 (speech or non-speech) ?



■ 音声 (speech sound/voice)



⇒人間の発声する音

4. 音の物理的性質

4.1 音の性質

■ 音波・・・空気の疎密波

■ 最も単純な音 正弦波 (sine wave)

■ 正弦波信号

■ 振幅 (amplitude)

$$\rightarrow A$$

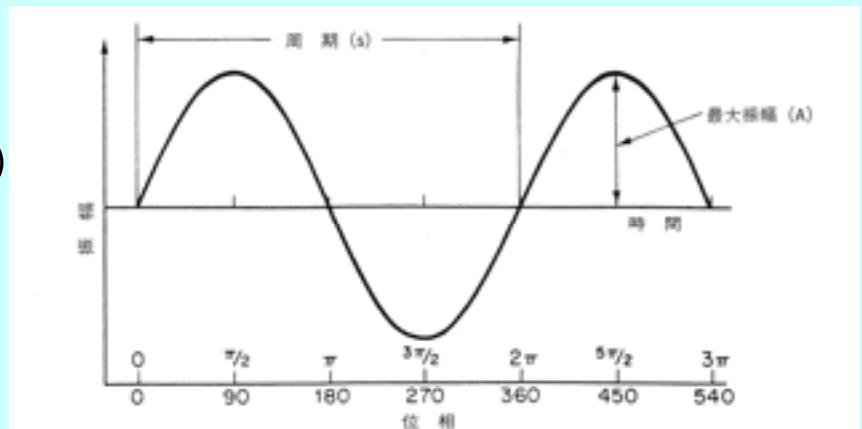
■ 周波数 (frequency)

$$\rightarrow f$$

■ 位相 (phase)

$$\rightarrow \theta$$

$$x(t) = A \sin(2\pi ft + \theta)$$



4. 音の物理的性質

4.2 Fourier分析とスペクトル表現

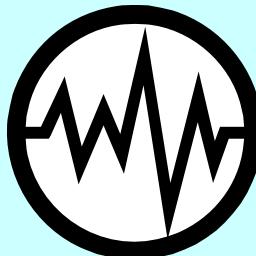
■ 離散Fourier級数展開

$$\left\{ \begin{array}{l} c_k = \int_{-T/2}^{T/2} \tilde{x}_a(t) e^{-jk\omega_0 t} dt \\ \tilde{x}_a(t) = \frac{1}{T} \sum_{k=-\infty}^{\infty} c_k e^{jk\omega_0 t} \end{array} \right.$$

■ 離散Fourier積分

$$\left\{ \begin{array}{l} \tilde{X}(k) = \sum_{n=0}^{N-1} \tilde{x}(n) W_N^{kn} \\ \tilde{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}(k) W_N^{-kn} \end{array} \right.$$

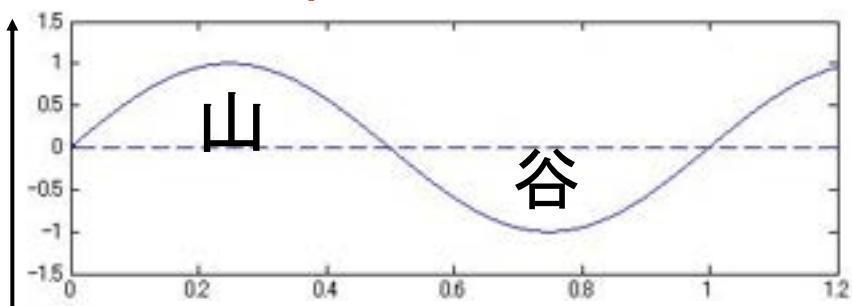
音の振動・・・周波数 (Hz)



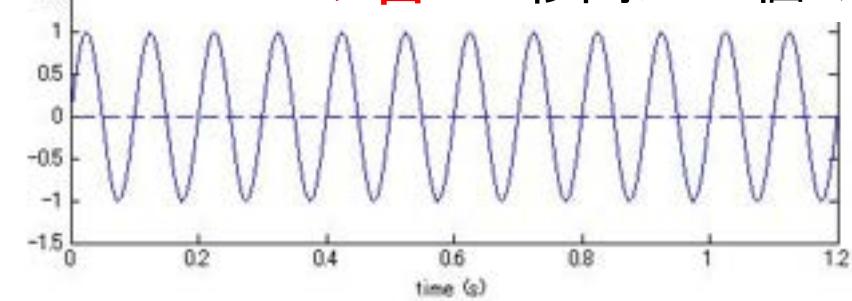
“波”

波の大きさ

1 Hzの音 \Rightarrow 1秒間に1個の波

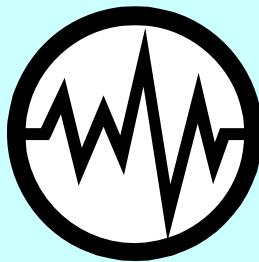


10 Hzの音 \Rightarrow 1秒間に10個の波

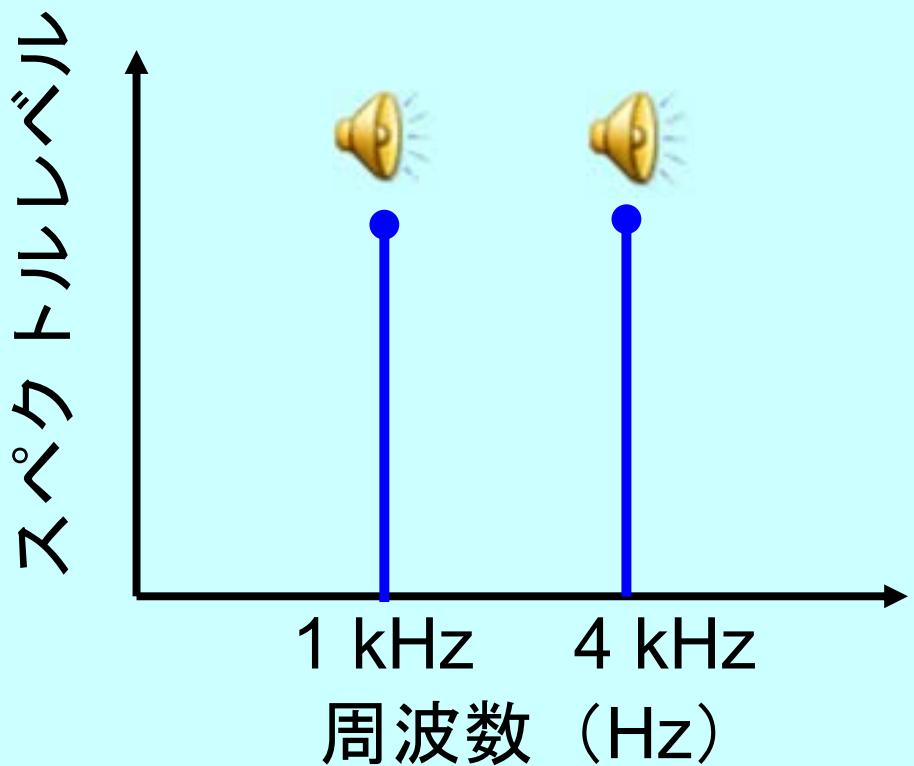


時間 (秒)

音の振動成分・・・スペクトル



“波”



波の分解・合成

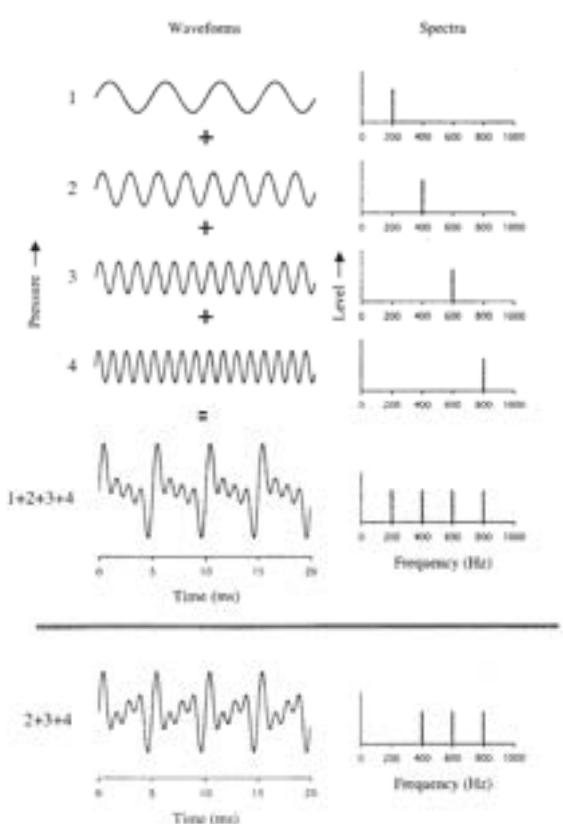
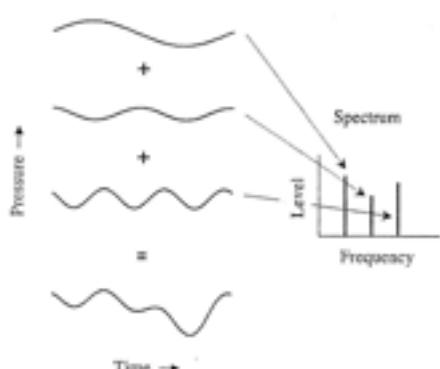
■ Fourierの考え方

■ 周期的な波 =

直流分

+ 正弦波の重ね合わせ

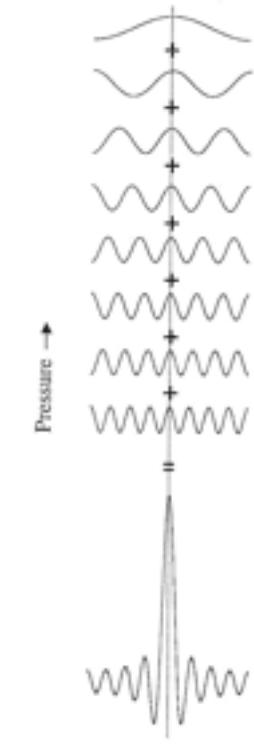
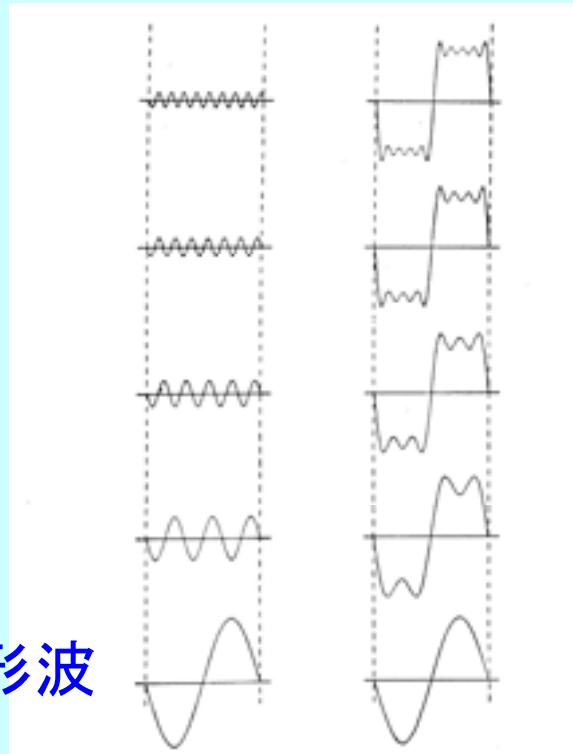
+ 余弦波の重ね合わせ



矩形波／インパルス性信号の分解

■ Fourier級数展開

矩形波



インパルス性信号

信号とスペクトルの関係①

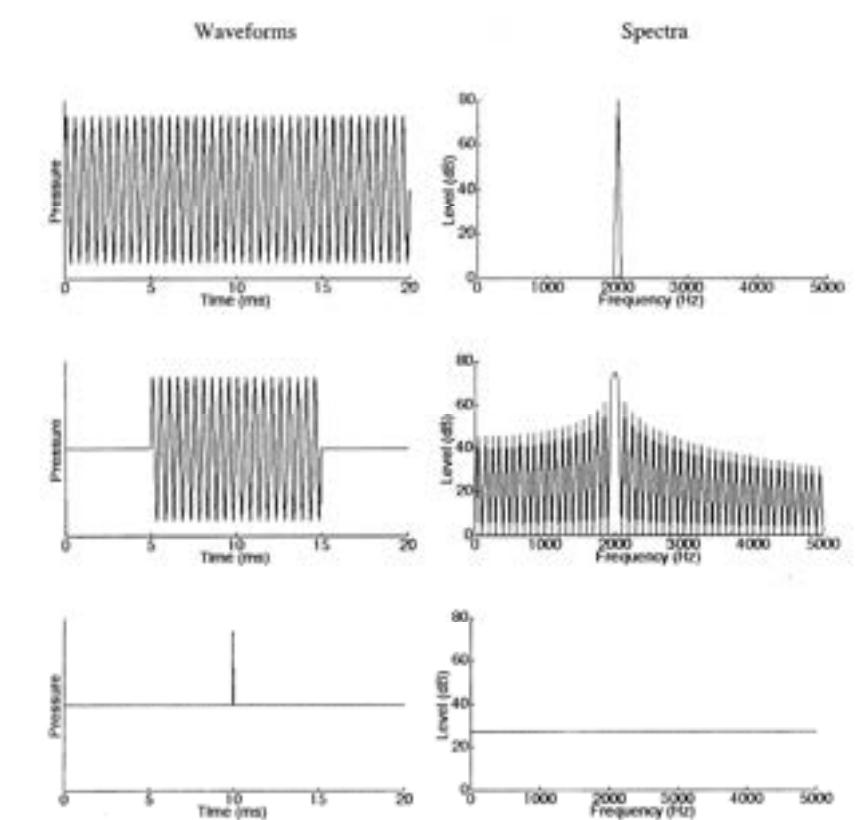
■ 正弦波の持続時間

を短くしていくと・・・

■ 正弦波

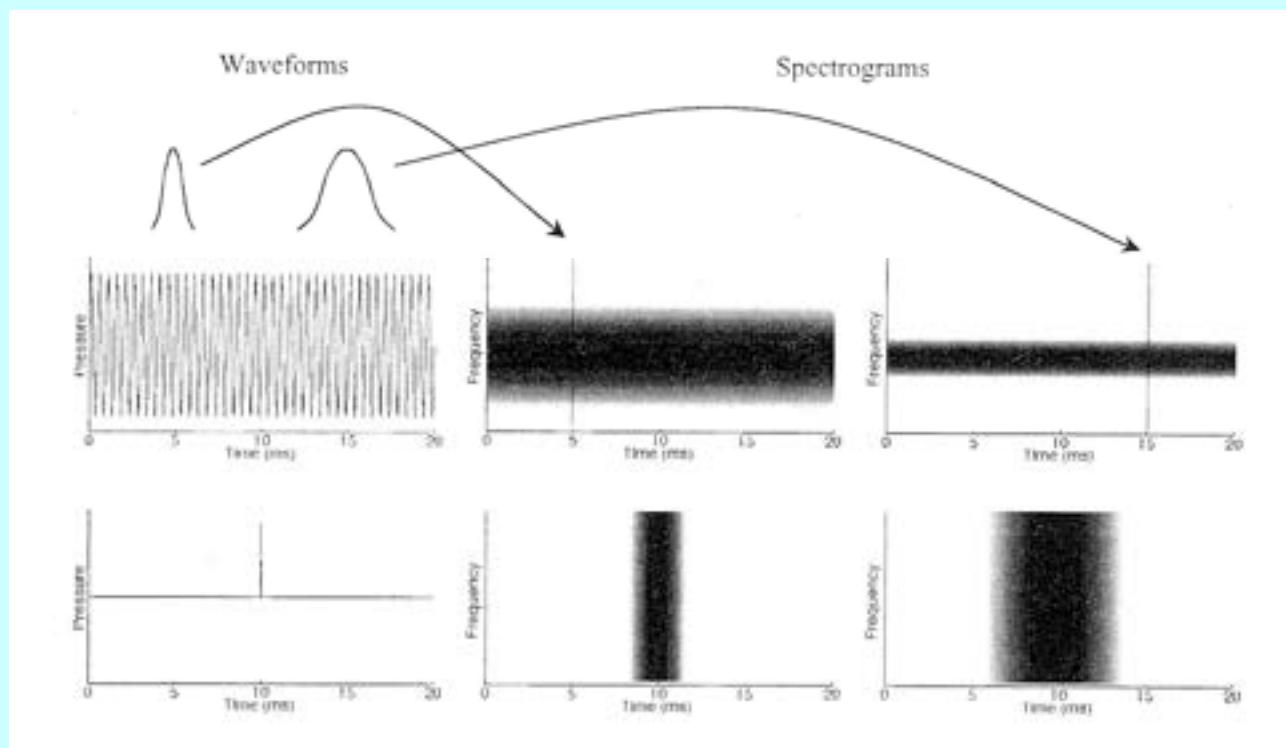
■ トーンバースト

■ インパルス



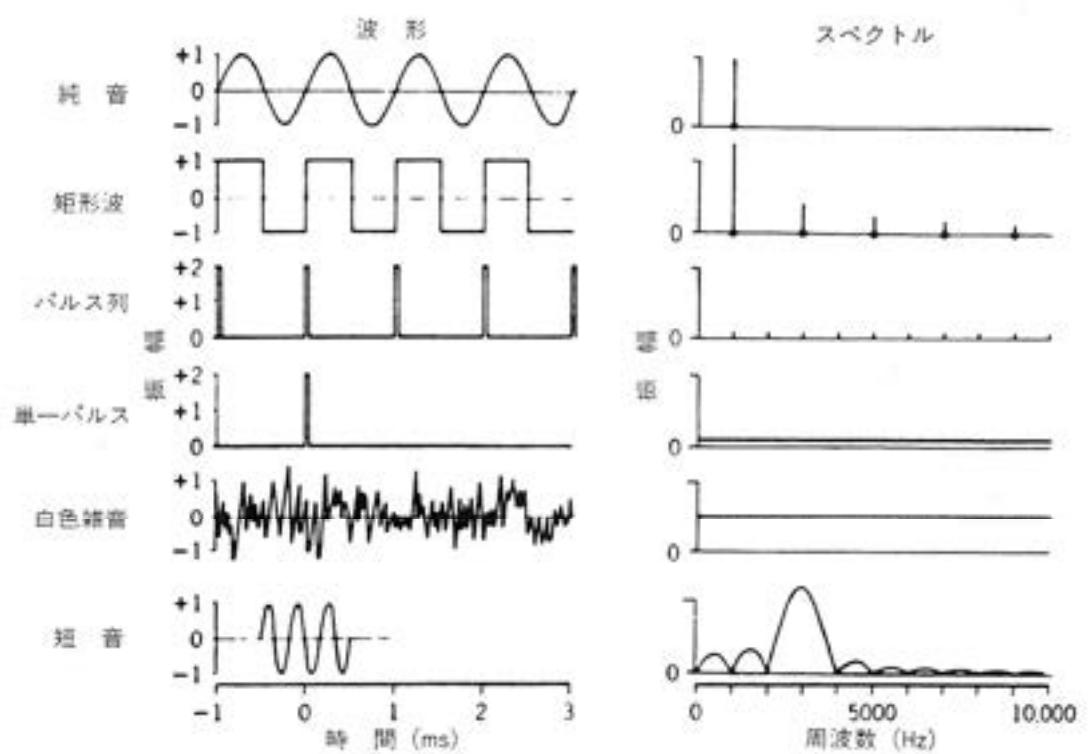
信号とスペクトルの関係②

■ Fourier分析の窓長の影響



信号とスペクトルの関係③

■ 様々な信号の場合



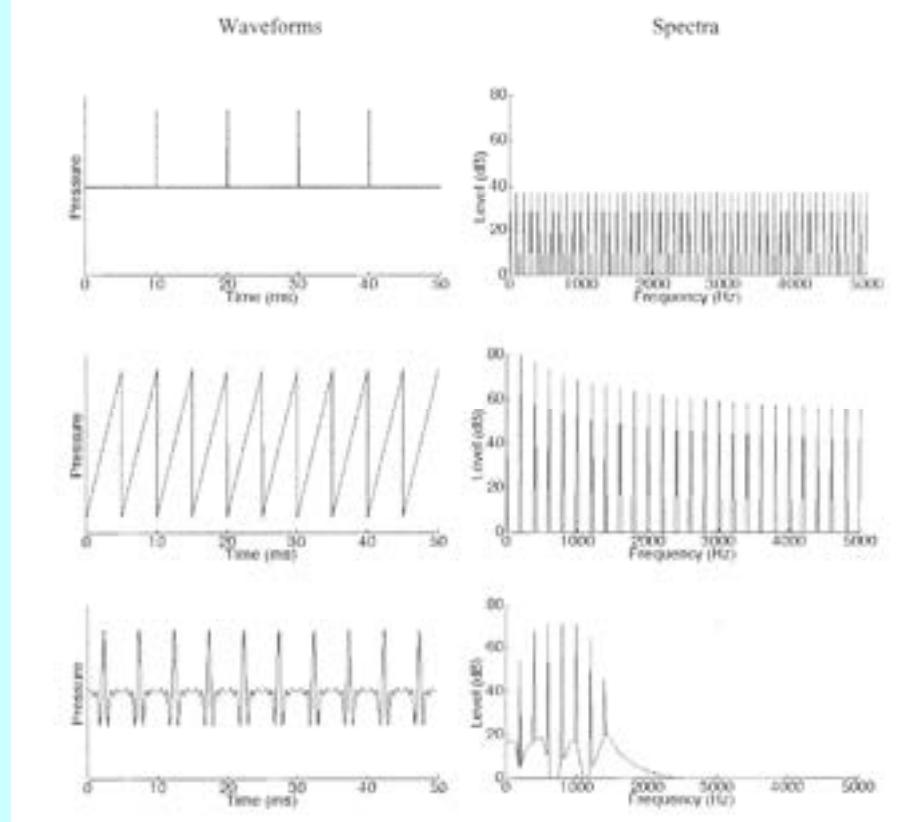
信号とスペクトルの関係④

■ 周期信号の場合

■ パルス列

■ 三角波列

■ 周期信号

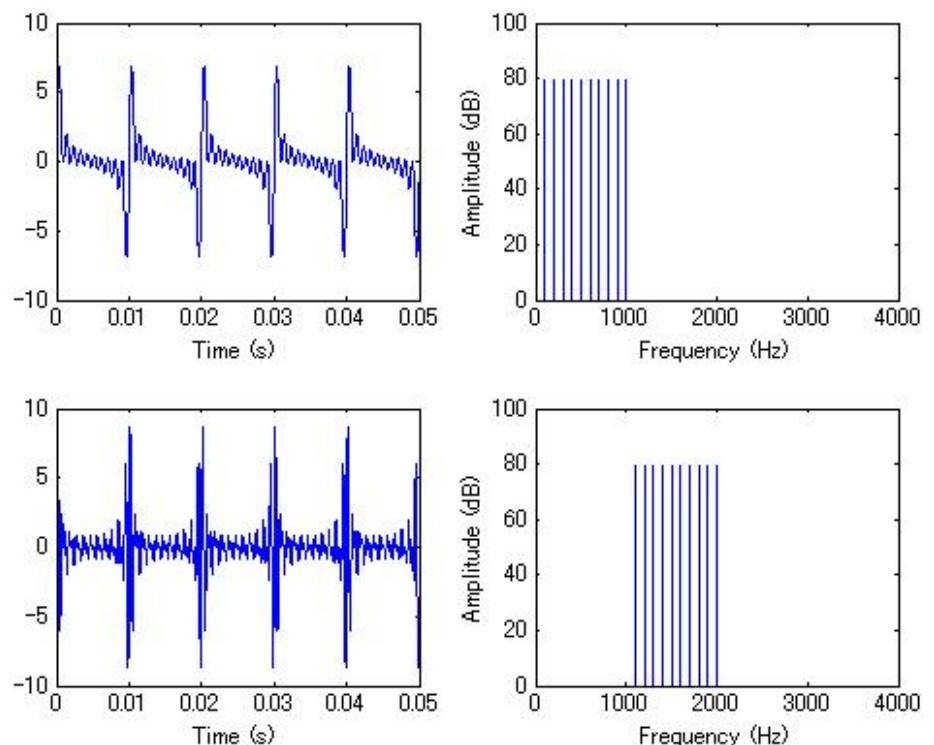


信号とスペクトルの関係⑤

■ 調波信号の場合

■ 低域のみ

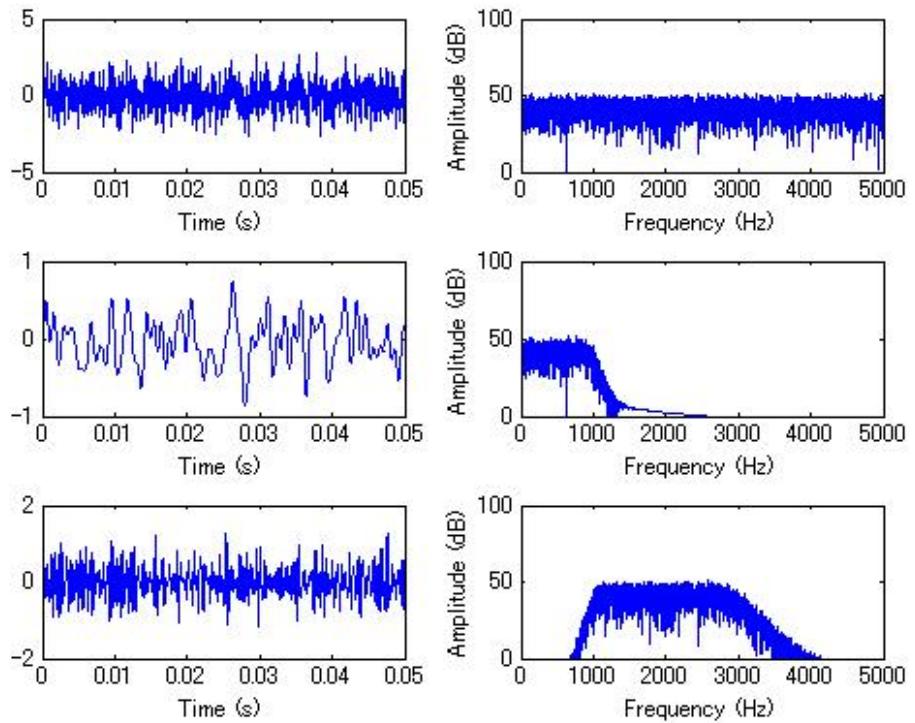
■ 高域のみ



信号とスペクトルの関係⑥

■ 白色雑音の場合

■ 白色雑音



■ 低域成分のみの雑音

■ 帯域制限雑音

4. 音の物理的性質

4.3 音のレベルの測定

■ 音の強度 (intensity)

音場中で単位面積を1秒間に通過する音のエネルギー

強さの基準 : $10^{-12} \text{ W/m}^2 = 2 \times 10^{-5} \text{ N/m}^2 = 20 \mu\text{Pa}$

■ レベル (level) dB

$$\text{デシベル値} = 10 \log_{10}(I_o/I_i)$$

入力の強度 I_i , 出力の強度 I_o

(例) 60 dB SPL → 0 dBの基準レベルより 60 dB 上

10^{-6} W/m^2 の強度をもつ

強さの比が 10 倍 (1/10 倍) → 10 dB 増加 (減少)

強さの比が 2 倍 (1/2 倍) → 3 dB 増加 (減少)

■ レベル (level) dB

$$\text{デシベル値} = 20 \log_{10}(P_o/P_i)$$

入力の圧力 P_i , 出力の圧力 P_o

音圧レベルの単位 dB SPL (Sound Pressure Level)

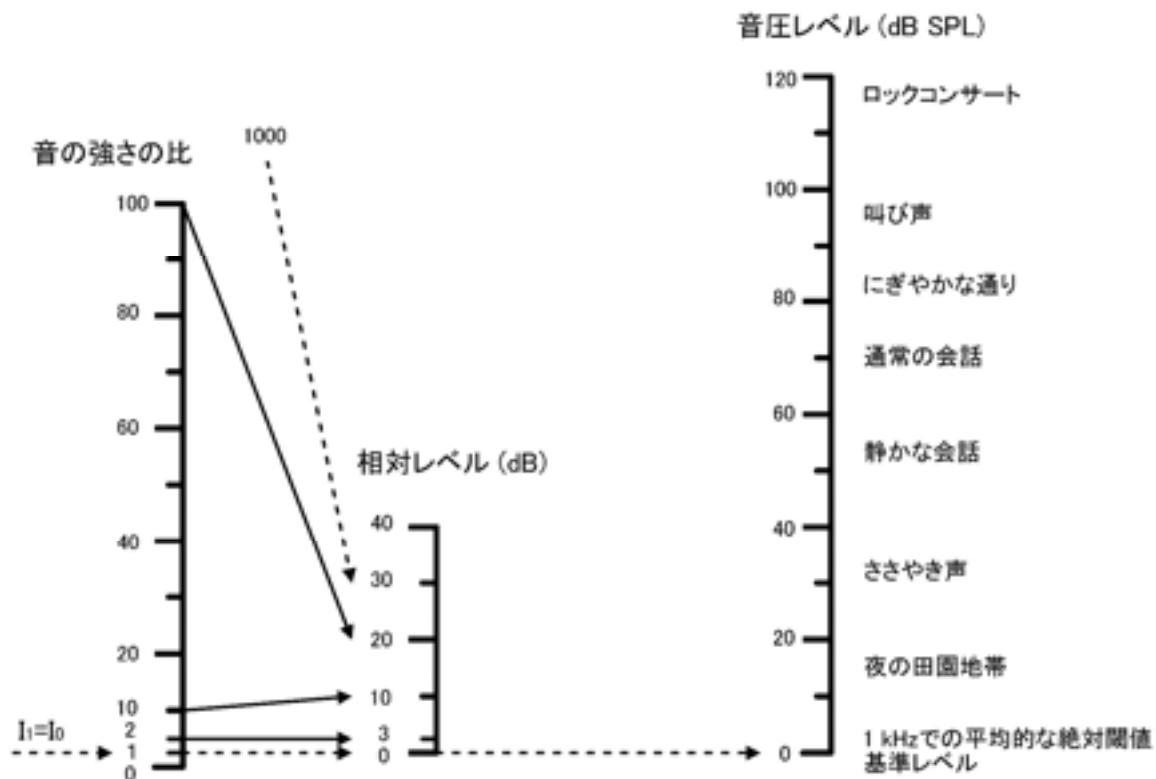
感覚レベルの単位 dB SL (Sensational Level)

■ 実効値 → 音圧の 2 乗平均値

root-mean-square value: RMS value

■ エネルギー密度 1 Hz幅の帯域内エネルギー

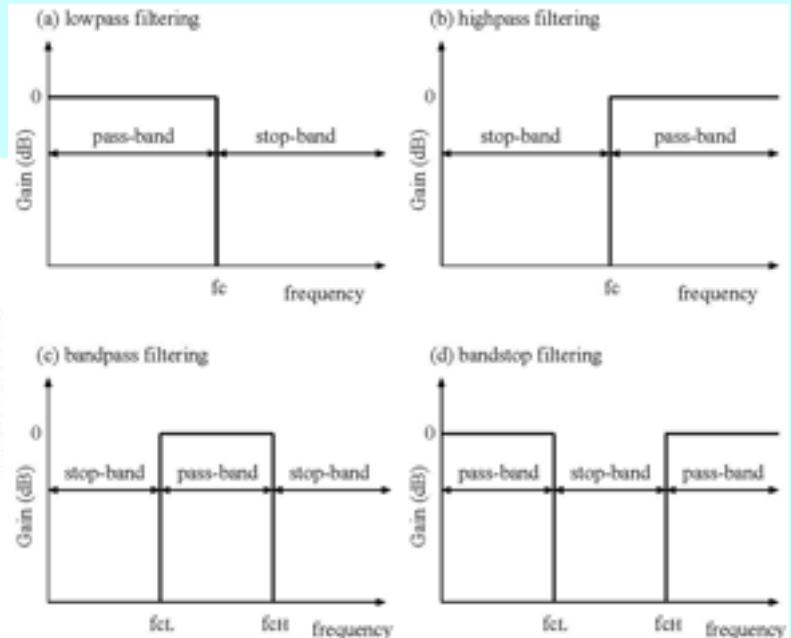
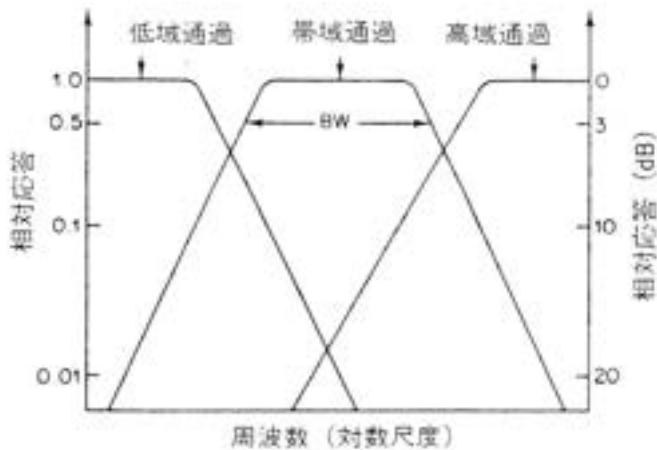
デシベル値の関係



5. フィルタとその特性

■ フィルタ (filter)

- 低域通過フィルタ
- 帯域通過フィルタ
- 高域通過フィルタ



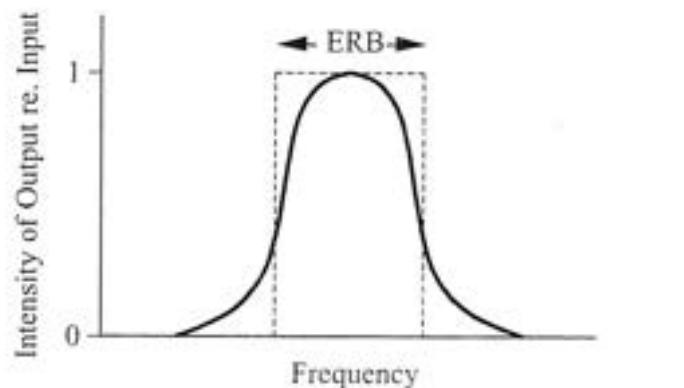
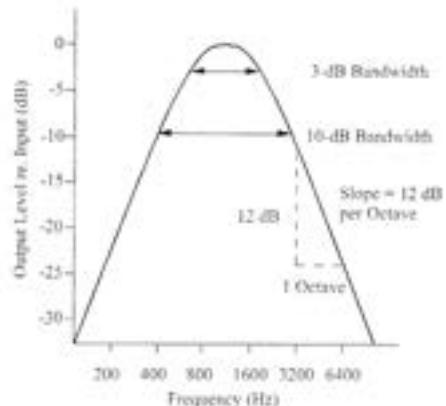
フィルタの特性

■ 帯域幅 (bandwidth)

- 3 dB帯域幅 (10 dB帯域幅)
- 等価矩形帯域幅

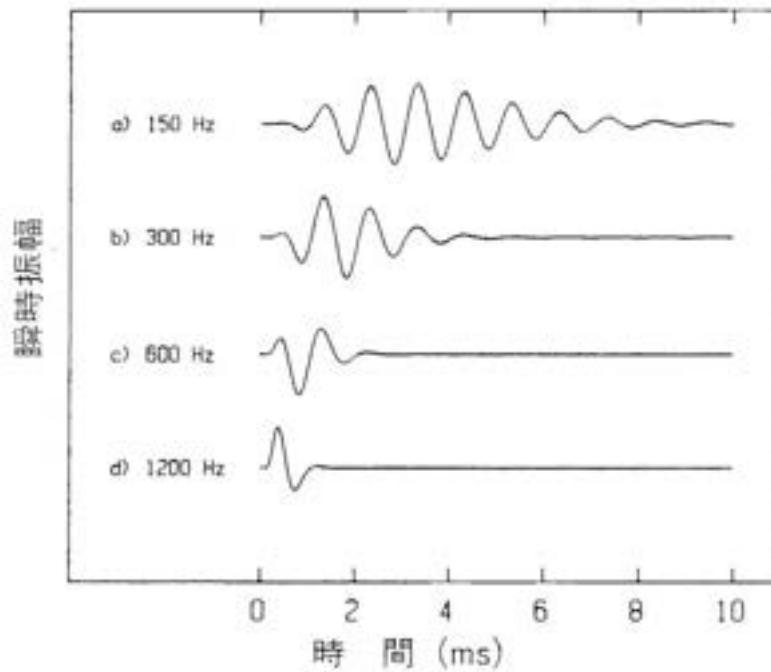
■ フィルタのQ値 (フィルタの鋭さの度合い)

- $Q = \text{中心周波数} / \text{帯域幅}$

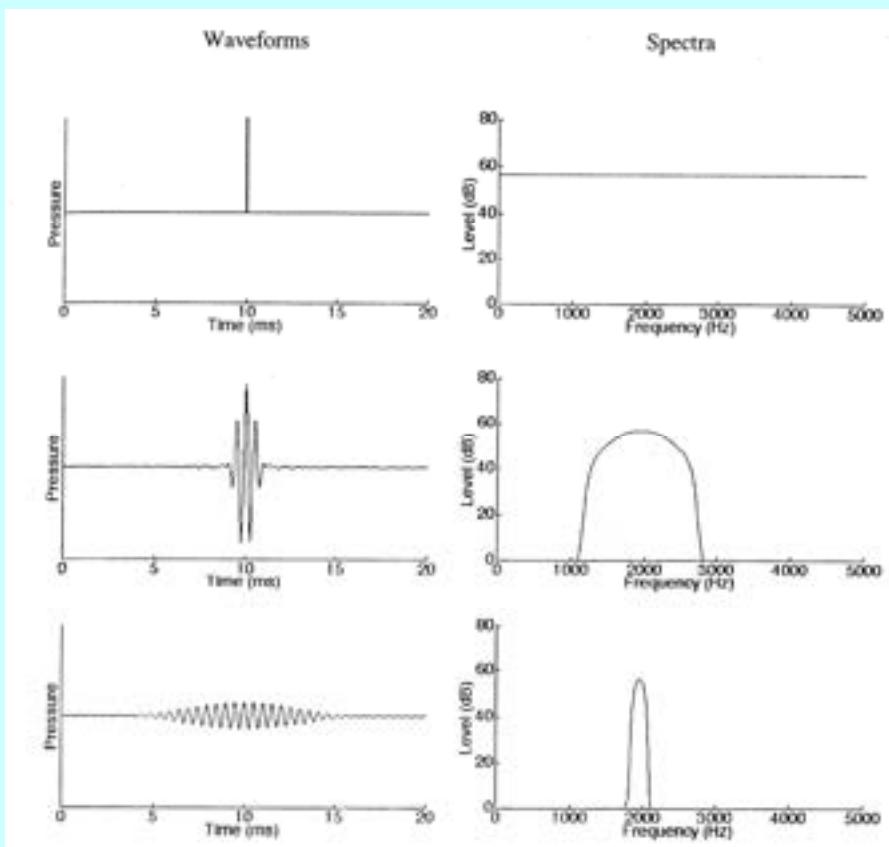


フィルタの応答

■ インパルス応答 (impulse response)

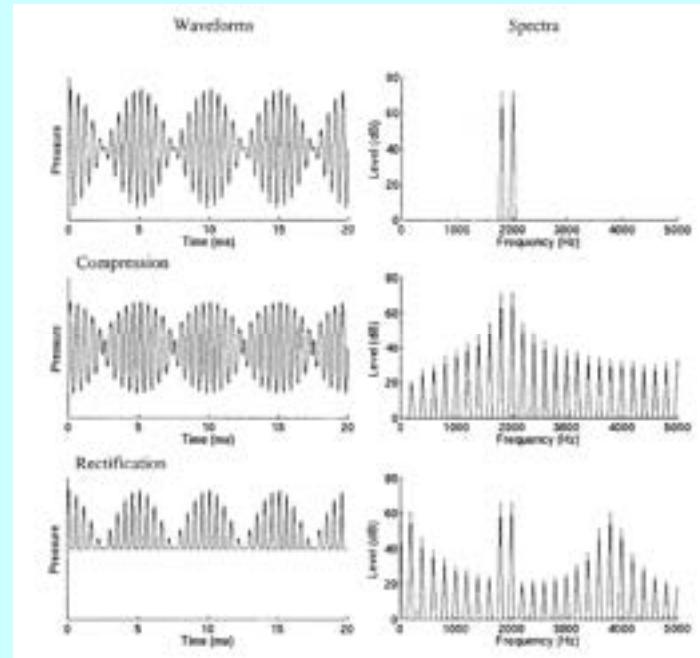
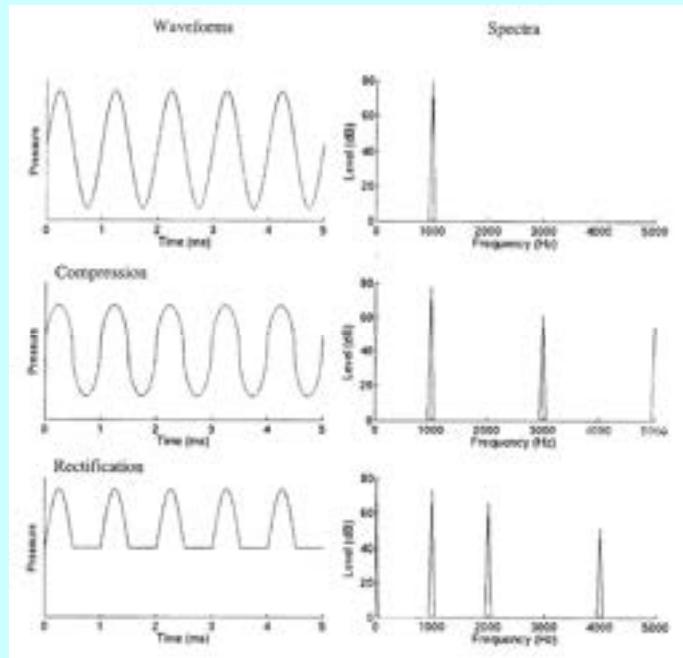


フィルタ帯域幅とインパルス応答



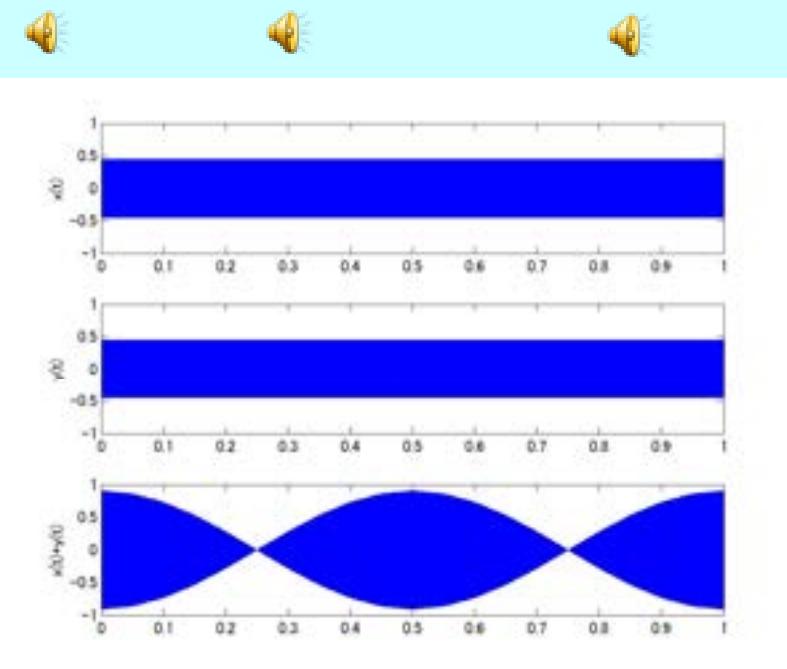
6. 波形操作

- 圧縮 (compression) の効果
- 半波整流 (rectification) の効果



うなり

- うなり (beats)
- (例) 1000 Hzと1002 Hzの音の和

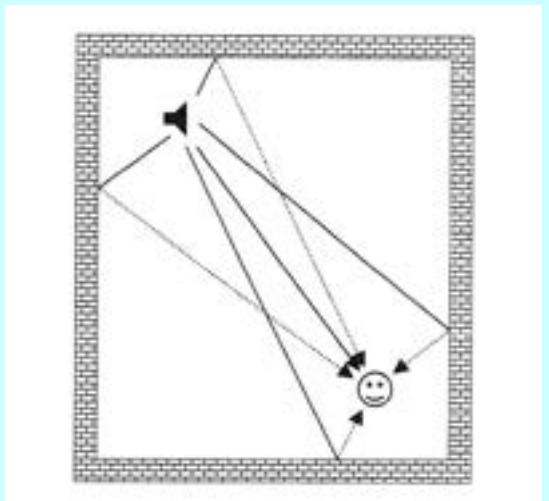
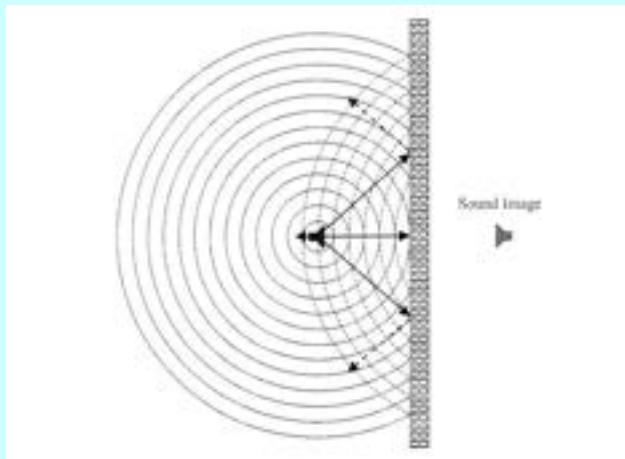


7. 音の伝播

■ 音の反射 (reflection)

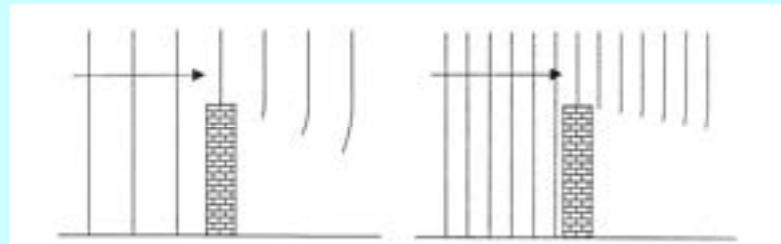
- 直接反射 (echo)
- 残響 (reverberation)

■ 音の回折 (diffraction)

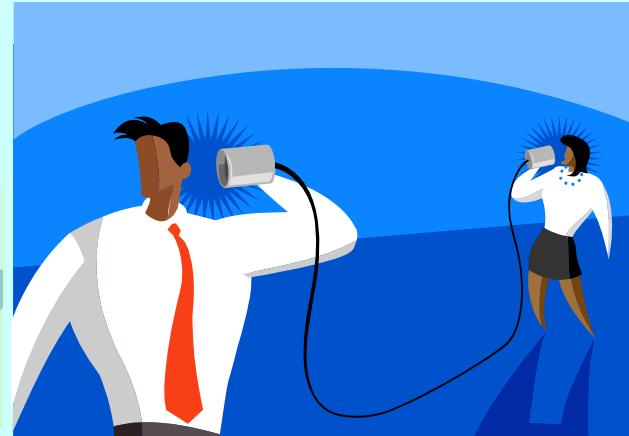
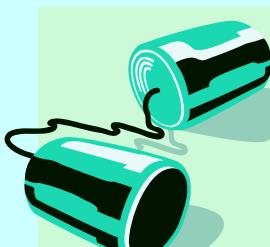


低い周波数音

高い周波数音



音声の伝播（空气中／媒質）



骨を伝わる音（骨伝導音）



骨を伝わるぜ！

I485F 知覚情報処理特論

4. 音の分析合成技術

2023年4月16日(日)

人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

Email: unoki@jaist.ac.jp



0. 音

■ “音 (sound)”

- 音波（物理的なもの）
- 音波によって引き起こされた感覚（心理的なもの）

■ 気圧

- $100,000 \text{ (N/m}^2\text{)}$

■ 圧力の変化

- → 音波

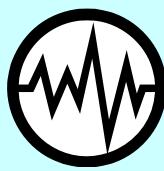
■ 音速

- 330 m/s (740 miles per hour)

1. 心理物理学

- 心理物理学 (psychophysics) とは
 - 心理現象を科学的に研究する方法の一つ
 - ISO5492による定義：刺激とその対応する感覚的反応との関係についての研究
- 聴知覚に関する心理物理学 (psychoacoustics)
 - 音響特徴と心理量の関係を調べるもの
 - 音響特徴を操作するためのツール
⇒ 音の分析合成系

Sound
(Stimulus)



Behavioral
Responses

- ◆ Audible
- ◆ Judgment

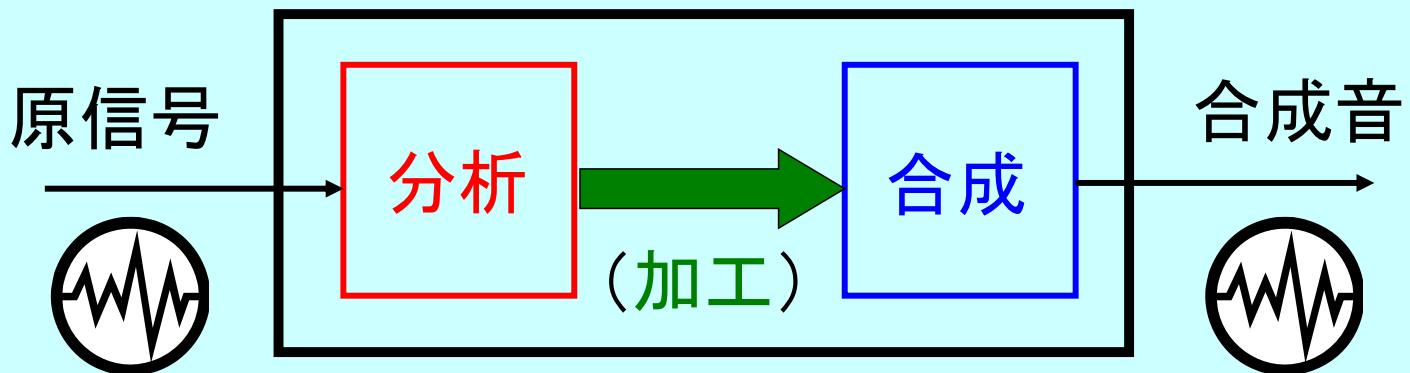
Acoustic

↔

Psycho

2. 音の分析・合成の考え方

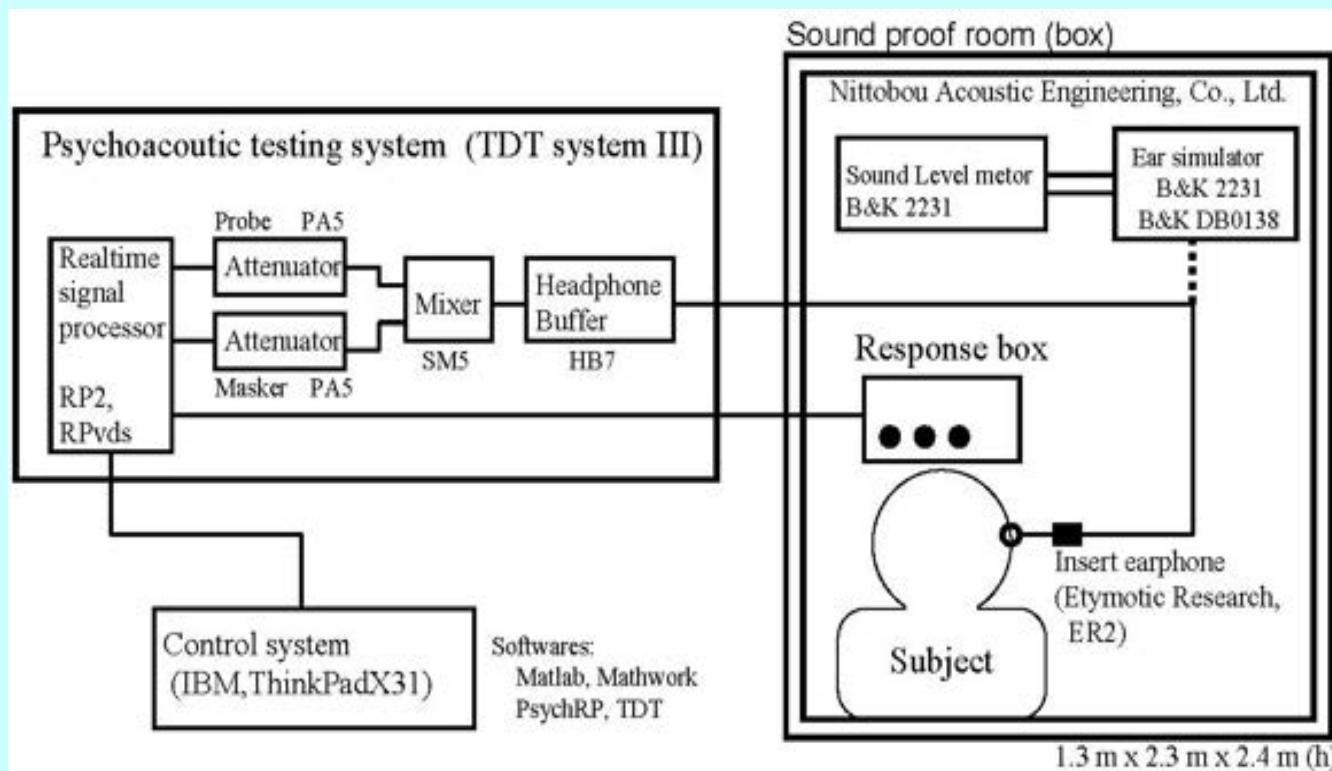
- 分析 (analysis)
- 合成 (synthesis)



信号を分析 → (加工) → 合成して
元の信号に戻すこと

分析合成系の利用例

心理物理実験：音を呈示してその聴知覚反応を調べること



3. 分析手法

■ ノンパラメトリック分析

分析対象の信号に対して、特定のモデルを仮定せずに分析を行う手法。万能であるが抽出すべき特徴（パラメータ）は非常に多くなる（情報量も多くなる）。

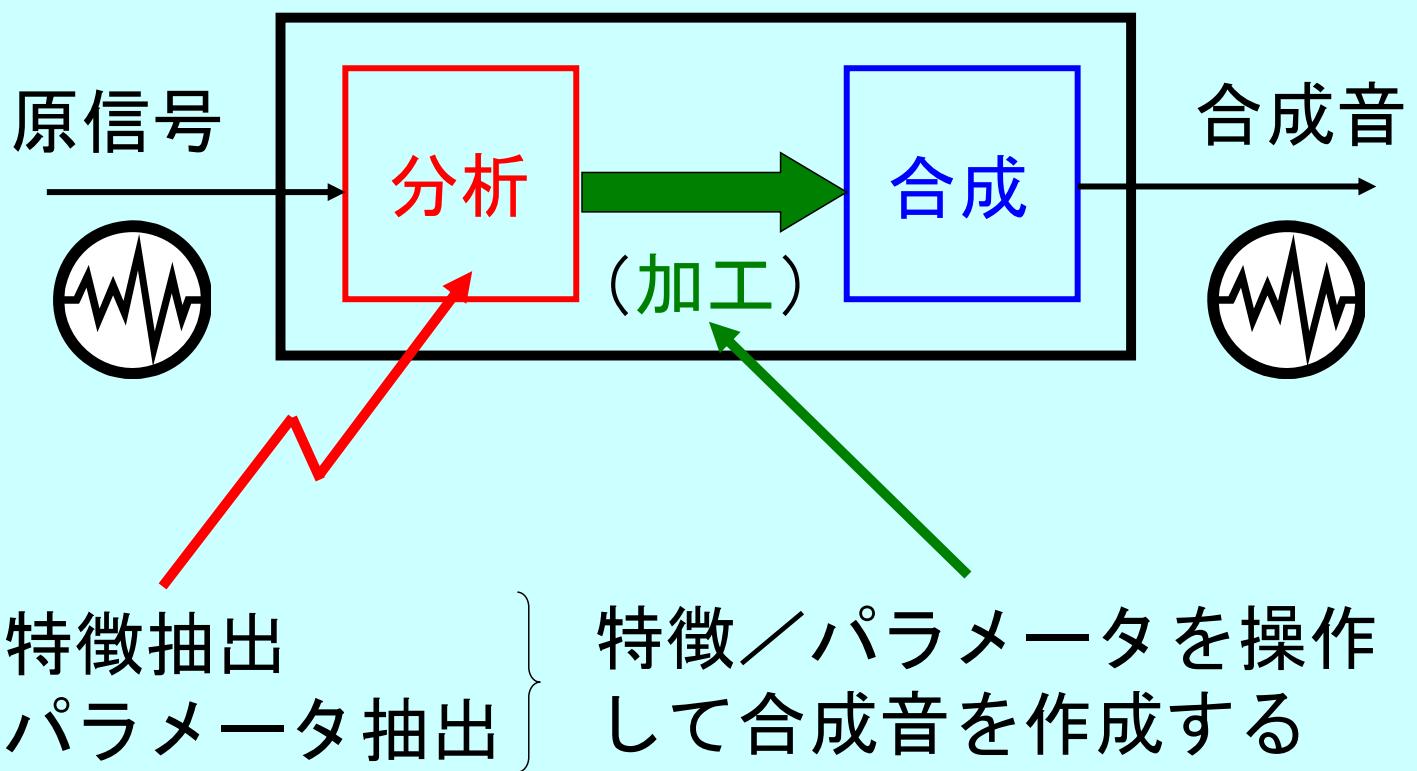
例えば $Y(z) = H(z)X(z) + N(z)$

■ パラメトリック分析

分析対象の信号に対して、特定のモデル化を行い、そのモデルを表現する特徴パラメータを抽出する。音声をよく表現するモデルを用意できれば、能率的な分析が可能（情報量も少なくなる）。

例えば、 $Y = f(a, b, c, \dots)$

ノンパラメトリック／パラメトリック分析



ノンパラメトリック分析

■ 短時間自己相関分析

スペクトル包絡と微細構造が畳み込みの形で表現

■ 短時間スペクトル分析

スペクトル包絡と微細構造が積の形で表現

■ ケプストラム分析

スペクトル包絡と微細構造がケフレンシー領域で分離

■ 帯域通過フィルタバンク分析

スペクトル包絡の概形

時間的な振幅包絡線情報とキャリア情報の積で表現

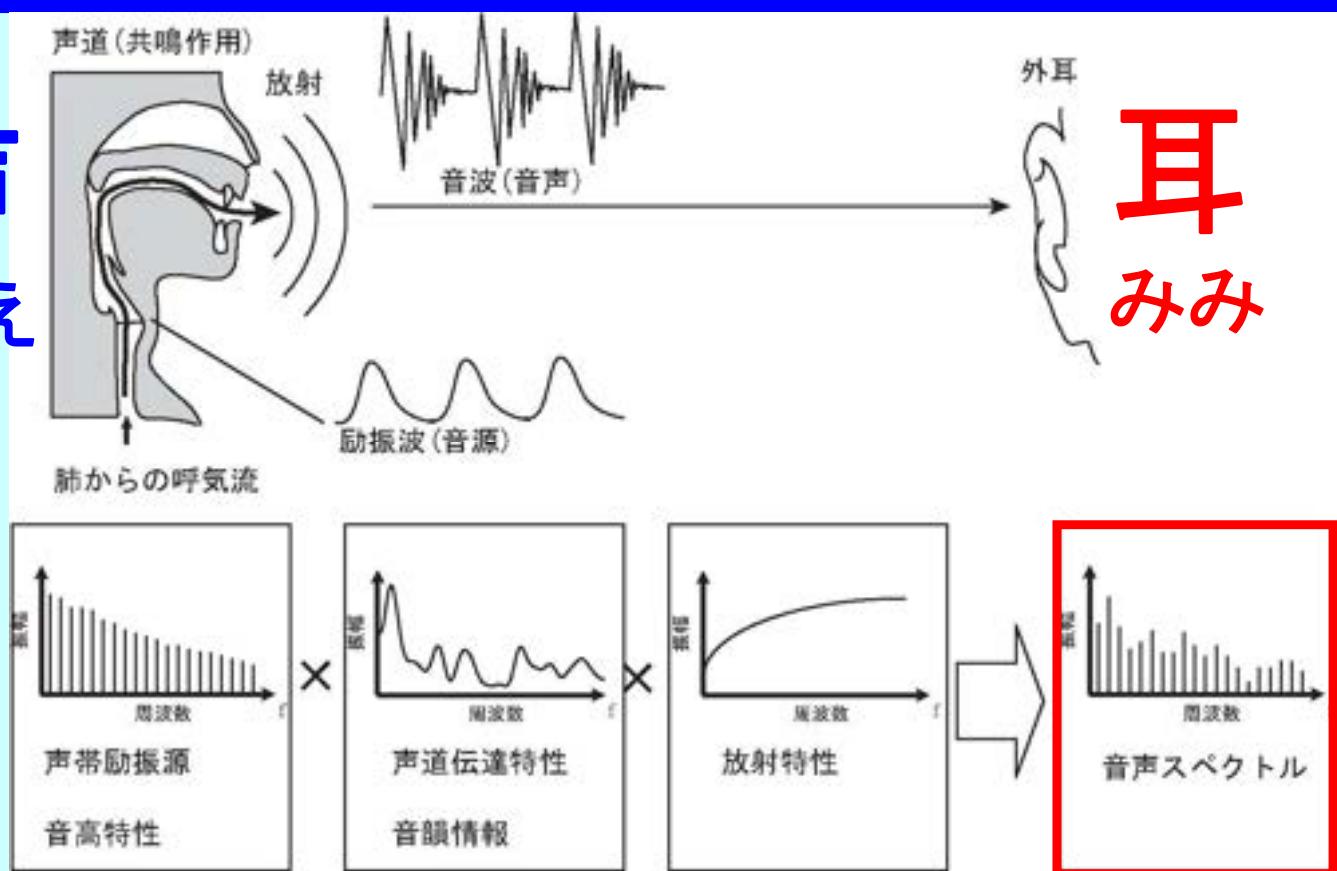
■ 零交叉数分析

ベースバンド周波数の表現

ソースフィルタモデル

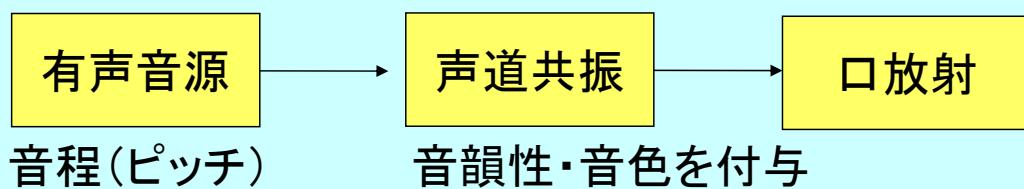
声
こえ

耳
みみ

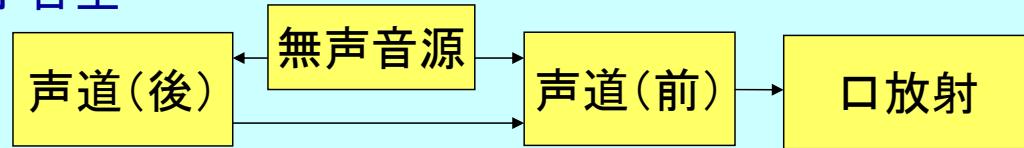


音声の構成

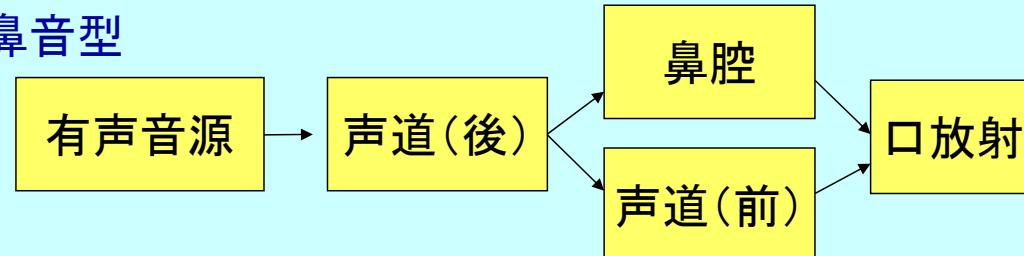
母音型



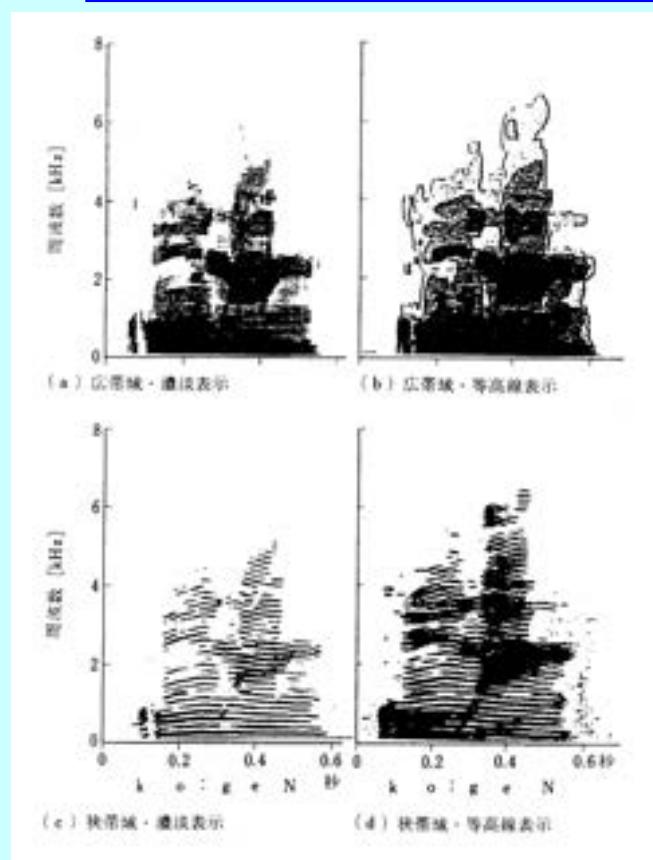
子音型



鼻音型



短時間スペクトル分析



■ Fourier変換に基づく分析

■ 短時間Fourier変換

信号 : $x(t)$

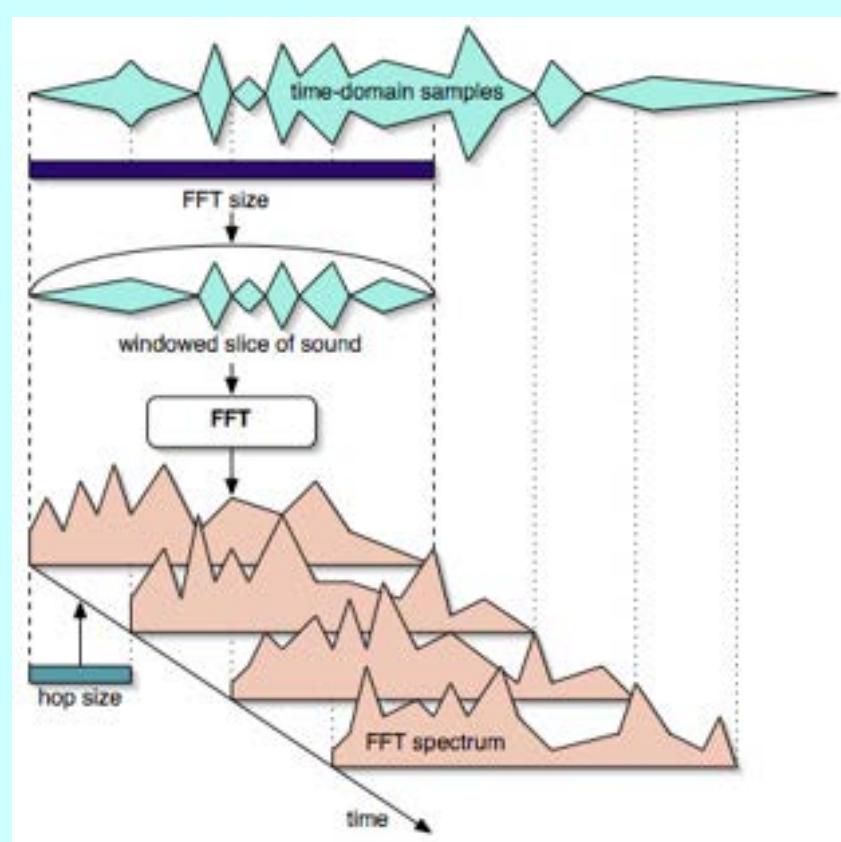
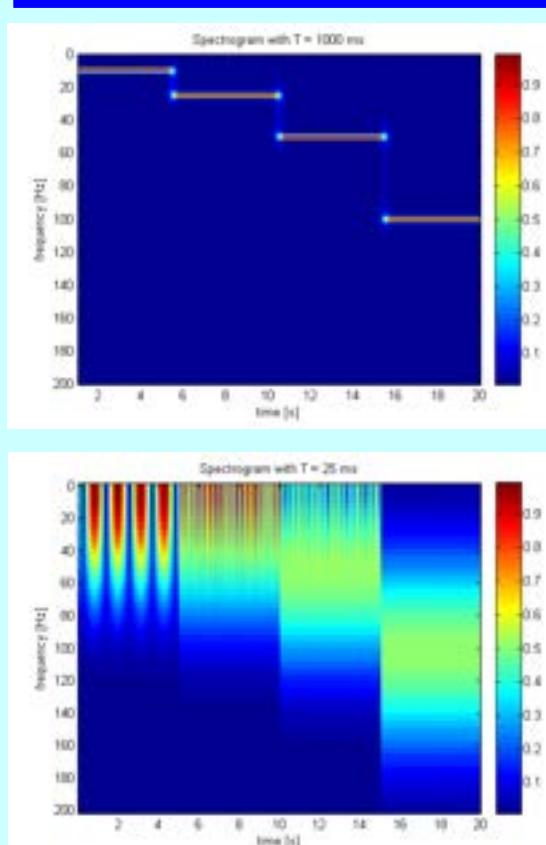
窓関数: $w(t)$

$$S(\omega, \tau) = \int x(\tau)w(t-\tau)\exp(-j\omega t)dt$$

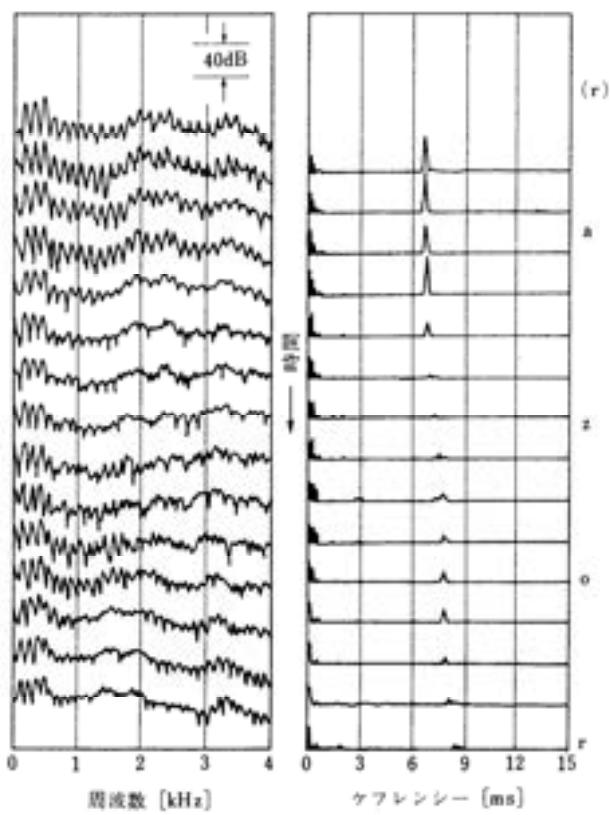
■ 窓関数の長さ（窓長）により見るものが異なる

■ 広い場合 \Rightarrow 微細構造

■ 狹い場合 \Rightarrow スペクトル包絡



ケプストラム分析



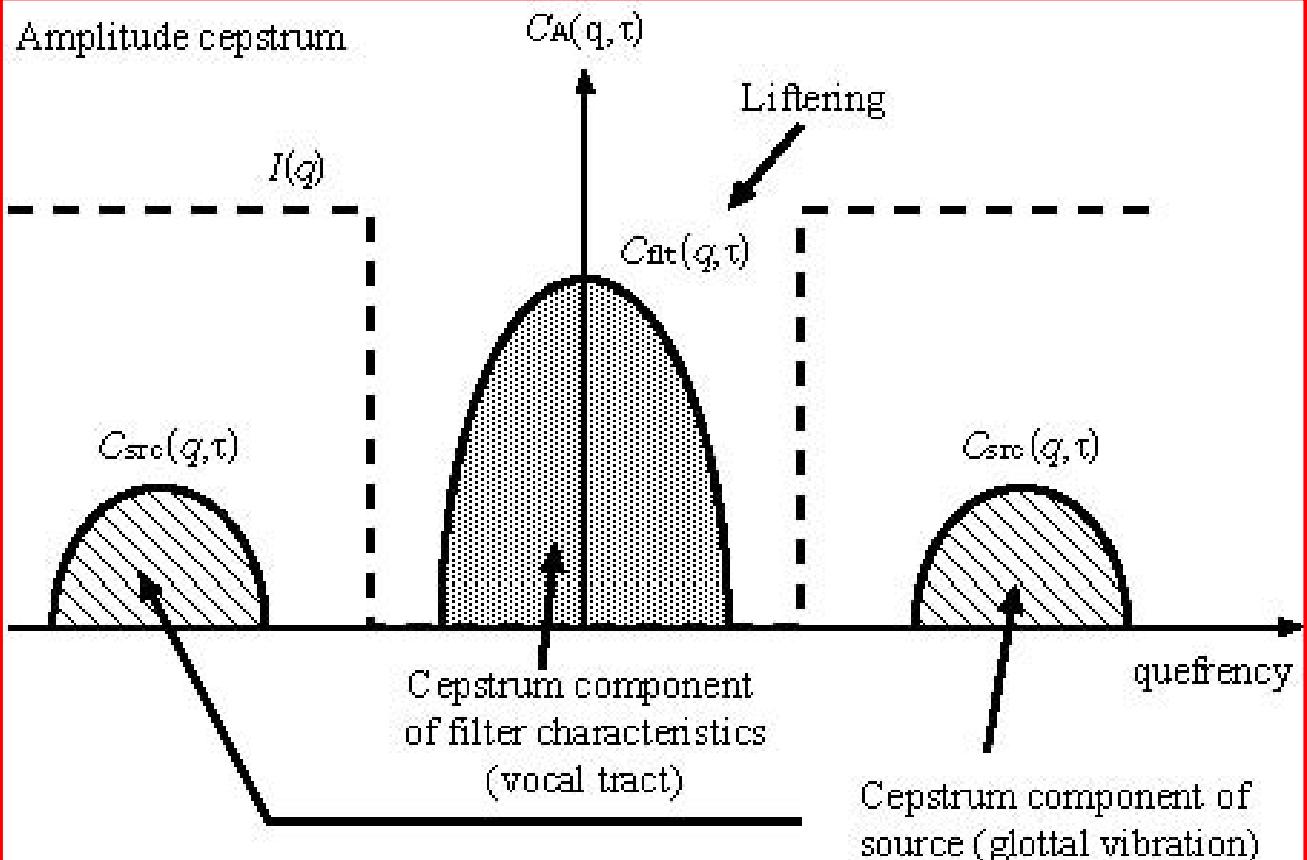
- ケプストラム (cepstrum)
準同型分析
- 複素ケプストラム : $c(q)$
- 実ケプストラム : $C(q)$

$$c(q) = \int \log[S(\omega)] \exp(j\omega q) d\omega$$

$$C(q) = \int \log|S(\omega)| \exp(j\omega q) d\omega$$

スペクトル包絡情報と微細構造の分離

音源フィルタモデル：リフタリング

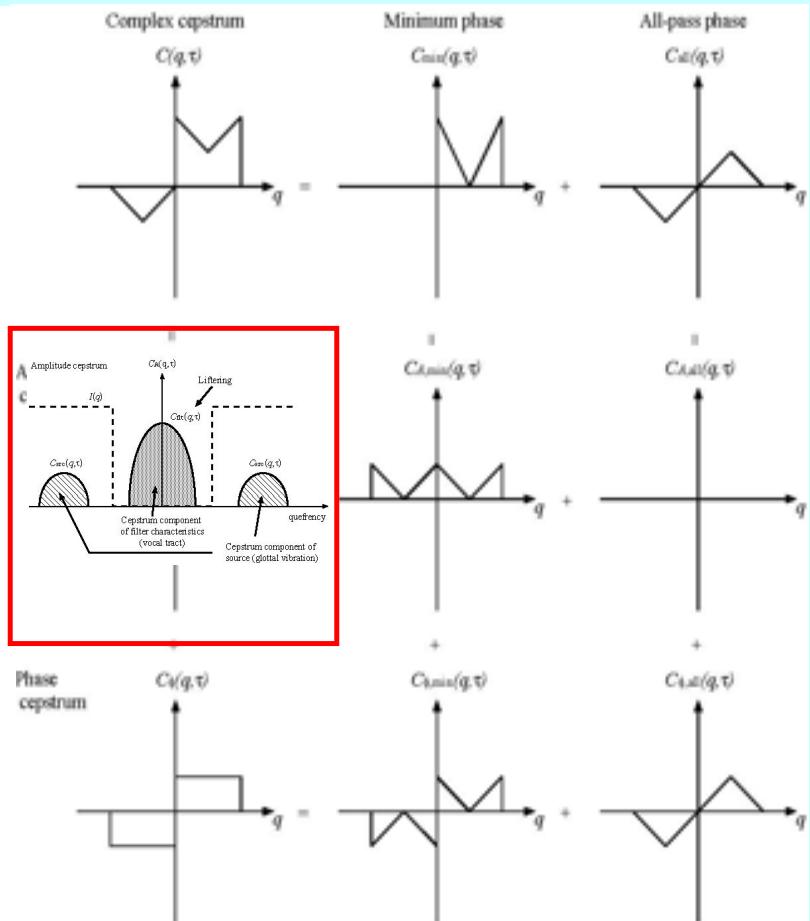


複素ケプストラム分析

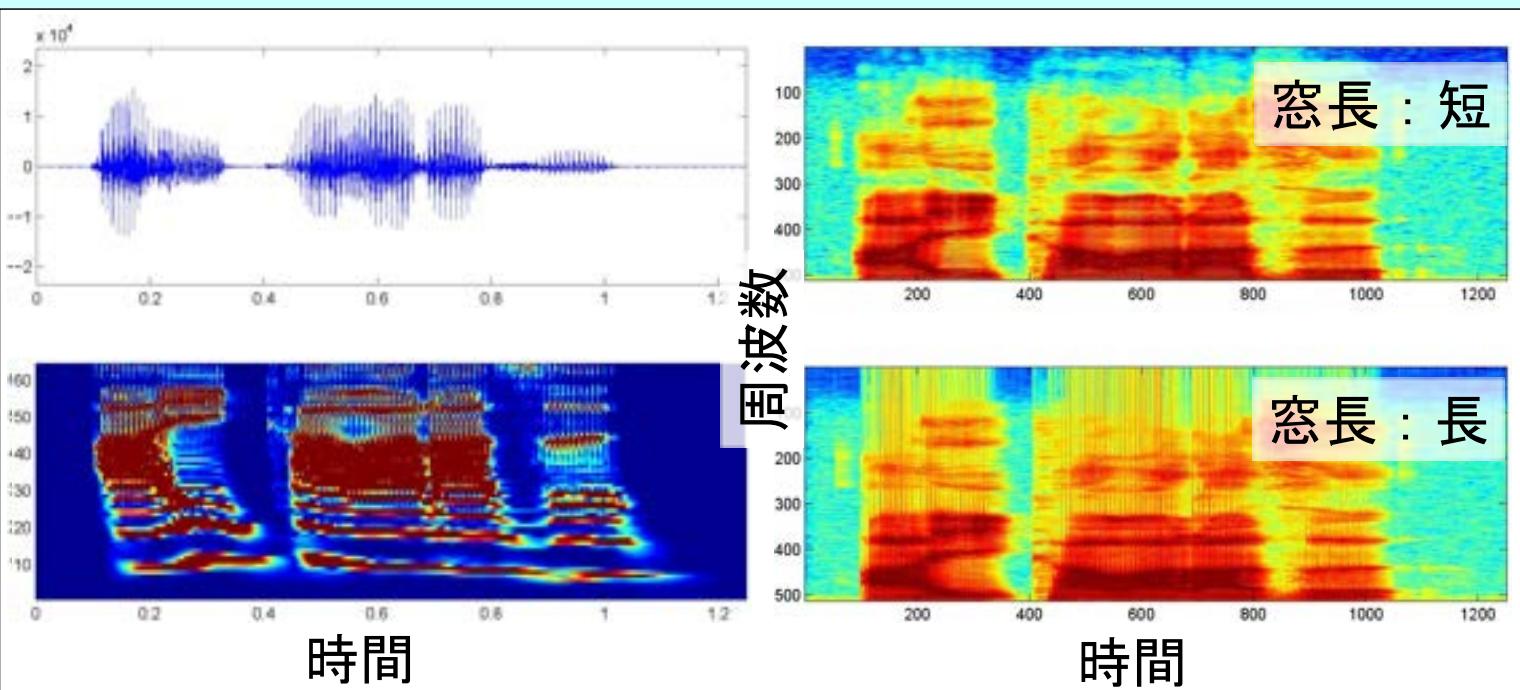
$$\begin{aligned}
 x(t, \tau) &= x_{\min}(t, \tau) * x_{\text{all}}(t, \tau) \\
 (\text{Periodic}) &\quad (\text{Minimum-Phase Component}) \quad (\text{All-Pass Component}) \\
 \\
 X(\omega, \tau) &= X_{\min}(\omega, \tau) \times X_{\text{all}}(\omega, \tau) \\
 (\text{Complex}) &\quad (\text{Complex}) \quad (\text{Complex}) \\
 || \\
 |X(\omega, \tau)| &= |X_{\min}(\omega, \tau)| \times |X_{\text{all}}(\omega, \tau)| \\
 (\text{Real}) &\quad (\text{Real}) \quad (\text{Real}) \\
 \times \\
 e^{j\phi(\omega, \tau)} &= e^{j\phi_{\min}(\omega, \tau)} \times e^{j\phi_{\text{all}}(\omega, \tau)} \\
 (\text{Complex}) &\quad (\text{Complex}) \quad (\text{Complex})
 \end{aligned}
 \quad (\text{Time domain})$$

$$\begin{aligned}
 &\stackrel{\mathcal{F}}{\Downarrow} \log \stackrel{\mathcal{F}^{-1}}{\Uparrow} \\
 \log X(\omega, \tau) &= \log X_{\min}(\omega, \tau) + \log X_{\text{all}}(\omega, \tau) \\
 (\text{Complex}) &\quad (\text{Complex}) \quad (\text{Complex}) \\
 || \\
 \log |X(\omega, \tau)| &= \log |X_{\min}(\omega, \tau)| + \log |X_{\text{all}}(\omega, \tau)| \\
 (\text{Real}) &\quad (\text{Real}) \quad (\text{Real}) \\
 + \\
 j\phi(\omega, \tau) &= j\phi_{\min}(\omega, \tau) + j\phi_{\text{all}}(\omega, \tau) \\
 (\text{Imaginary}) &\quad (\text{Imaginary}) \quad (\text{Imaginary})
 \end{aligned}
 \quad (\text{Frequency domain})$$

$$\begin{aligned}
 &\stackrel{\mathcal{F}^{-1}}{\Downarrow} \stackrel{\mathcal{F}}{\Uparrow} \\
 C(\omega, \tau) &= C_{\min}(\omega, \tau) + C_{\text{all}}(\omega, \tau) \\
 (\text{Asymmetric}) &\quad (\text{Asymmetric}) \quad (\text{Asymmetric}) \\
 || \\
 C_A(\omega, \tau) &= C_{A,\min}(\omega, \tau) + C_{A,\text{all}}(\omega, \tau) \\
 (\text{Even func.}) &\quad (\text{Even func.}) \quad (\text{Even func.}) \\
 + \\
 C_S(\omega, \tau) &= C_{S,\min}(\omega, \tau) + C_{S,\text{all}}(\omega, \tau) \\
 (\text{Odd func.}) &\quad (\text{Odd func.}) \quad (\text{Odd func.})
 \end{aligned}
 \quad (\text{Quefrency (time) domain})$$



フィルタバンク



基底膜応答（人間）

工学的表現（機械）

パラメトリック分析

■ 合成による分析 (AbS: Analysis-by-Synthesis)

モデルの精密化が可能

■ 線形予測分析 (LPC: Linear Predictive Coding)

全極型スペクトルで表現

- 最尤スペクトル推定
- 共分散法
- PARCOR分析
- LSP分析

安定性の保証あり

安定性の保証なし

格子フィルタの利用

量子化／補間特性が良好

4. 分析合成系

■ ボコーダ (Vocoder: Voice coderの略称)

H. Dudley (1939)の発明

■ 狙い

- 音声波形に含まれる冗長な成分を捨て、聴覚的に重要な情報のみを正確に備えたパラメータを抽出する
- 特徴パラメータの効率的な符号化を行う
- 符号化された特徴パラメータから、元の音声にできるだけ忠実な音声（明瞭性・自然性のよい音声）を復元する

代表的な分析合成系

- チャネルボコーダ H. Dudley (1939)
- フォルマントボコーダ W. A. Munson (1950)
- パターンマッチングボコーダ C. P. Smith (1957)
- 相関ボコーダ M. R. Schroeder (1959)
- 位相ボコーダ J. L. Flanagan (1966)
- 最尤ボコーダ 板倉&斎藤 (1968)
- ホモモルフィックボコーダ A. V. Oppenheim (1969)
- 線形予測ボコーダ 板倉&斎藤 (1969)
- LSPボコーダ B. S. Atal (1971)
- STRAIGHT 河原 (1999)

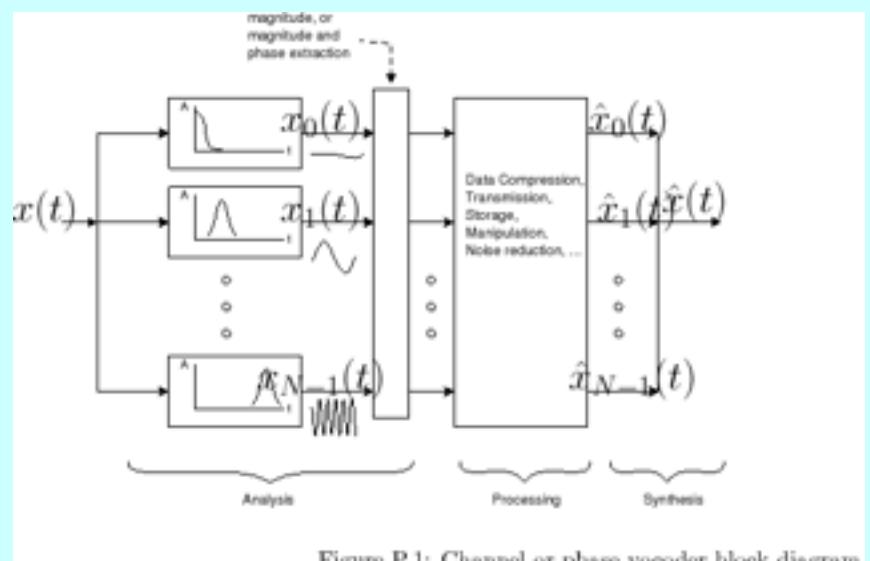
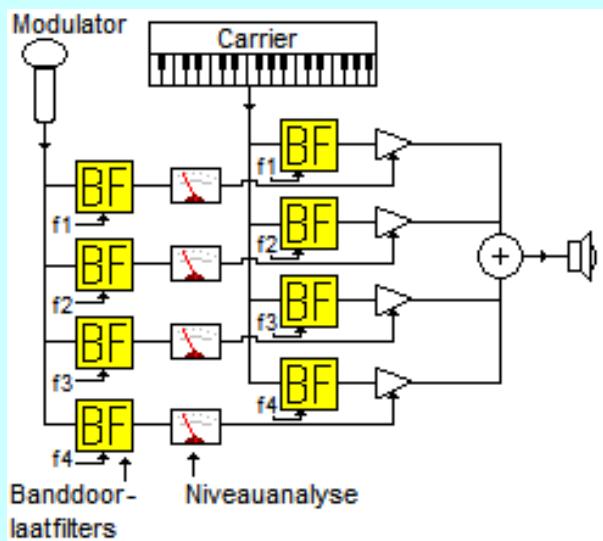


Figure P.1: Channel or phase vocoder block diagram.

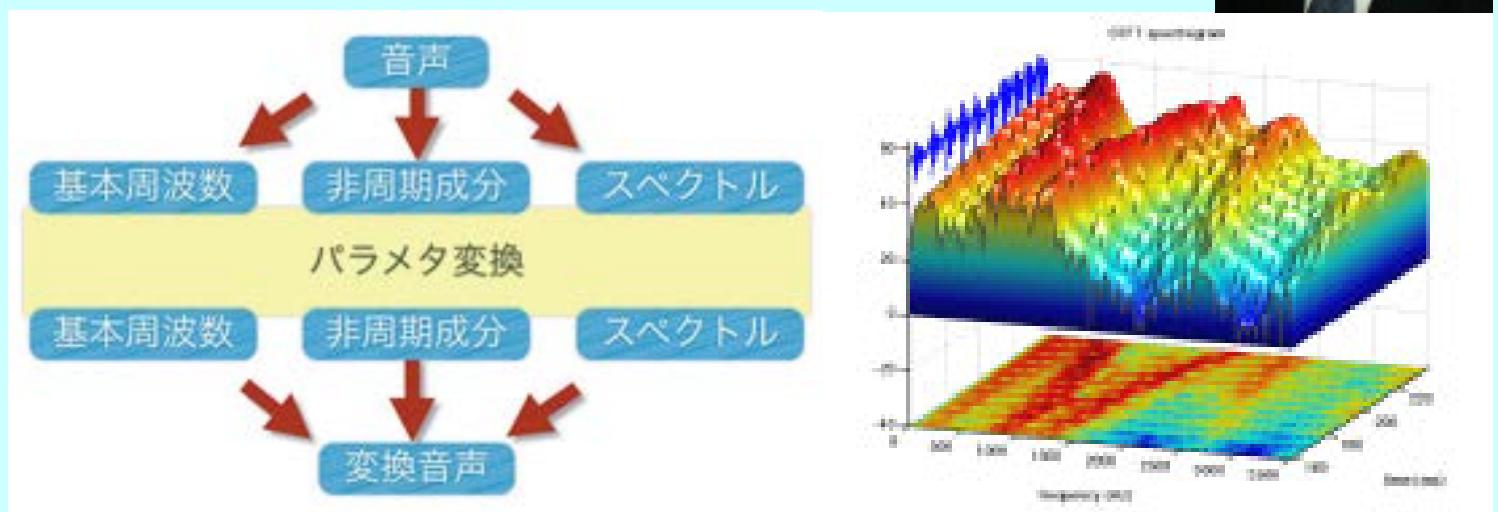
STRAIGHT

■ STRAIGHT

Speech Transformation and Representation using Adaptive Interpolation of
weiGHTed spectrum

http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index_j.html

<http://www.crestmuse.jp/cmstraight/>



- STRAIGHTは、音声のピッチや声質、体の大きさなどの様々な属性を操作するためのツールです。自然音声とほとんど区別できない高い品質を実現するために、STRAIGHTは、聴覚モデルと信号処理の最新の成果を取り入れて進化し続けています。2007年に森勢君が発明したTANDEMに基づいて、STRAIGHTが根本的に再構築されてTANDEM-STRAIGHTとなりました。（先行して公開している [e-SocietyによるSTRAIGHTのC言語版](#)につづいて、2009年1月からTANDEM-STRAIGHTのMatlab版、2009年11月からMatlab非依存のアプリケーションが利用できるようになりました。これらの入手に関しては、[大学等の窓口（河原@和歌山大学）](#)と、[企業等の窓口（山本@関西TLO）](#)とが用意されていますので、そちらにお問い合わせ下さい。）
- STRAIGHTは、音声知覚の研究用のツールとして開発されました。1930年代から本格化した膨大な音声知覚研究の蓄積を活かして、その延長上に新しい研究を積み上げができるように、音源とフィルタという概念に基づいた構造とっています。
- STRAIGHTは、[ATR](#)人間情報通信研究所で発明され、聴覚の計算論の研究の中核的な技術として[科学技術振興事業団](#)のCRESTの一である『[聴覚脳プロジェクト](#)』の下で研究が進められてきました。現在も、[CrestMuseプロジェクト](#)をはじめとして数多くの研究プロジェクトの基盤技術として利用されています。

Singing voice conversion

Interspeech2007@Antwerp, Belgium
Synthesis of Singing Challenge Competition

“Vocal Conversion from Speaking Voice to Singing Voice Using STRAIGHT” by T. Saitou, M. Goto, M. Unoki, and M. Akagi

Audio files: Original Male Voice 🔊 Synthesized Male Singing 🔊
Original Female Voice 🔊 Synthesized Female Singing 🔊

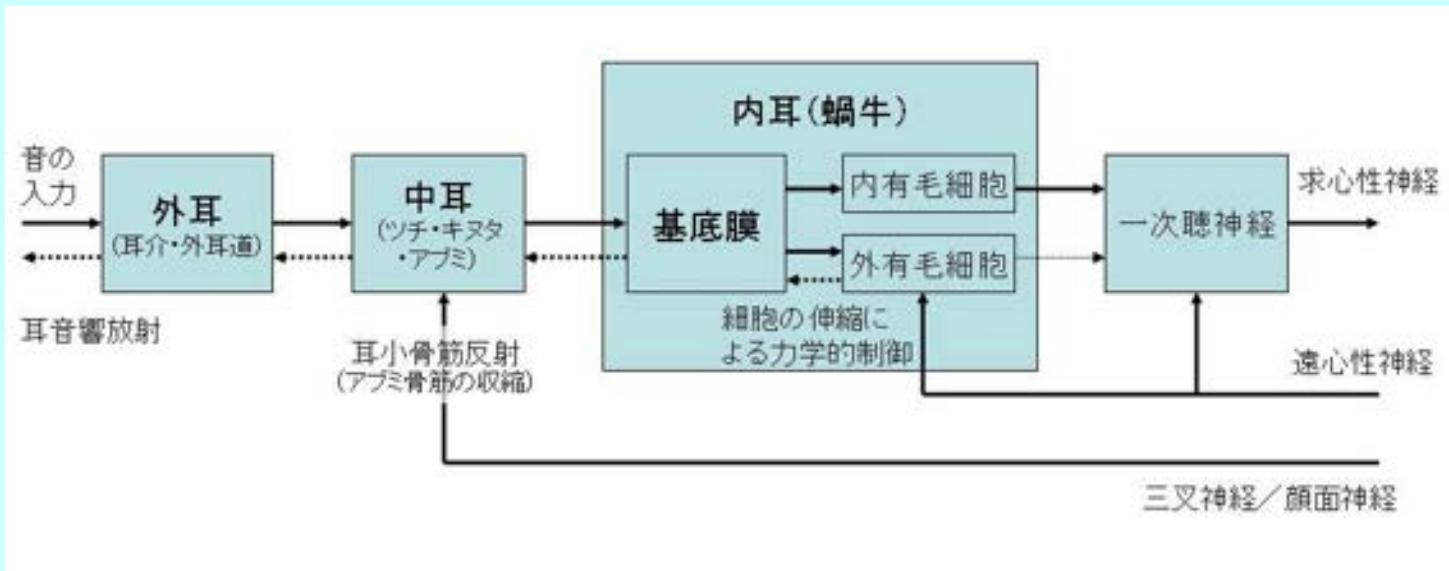
The Synthesizer Song: Original Male Voice 🔊 Baritone 🔊
Original Female Voice 🔊 Alto 🔊
Synthesized Baritone + Alto 🔊

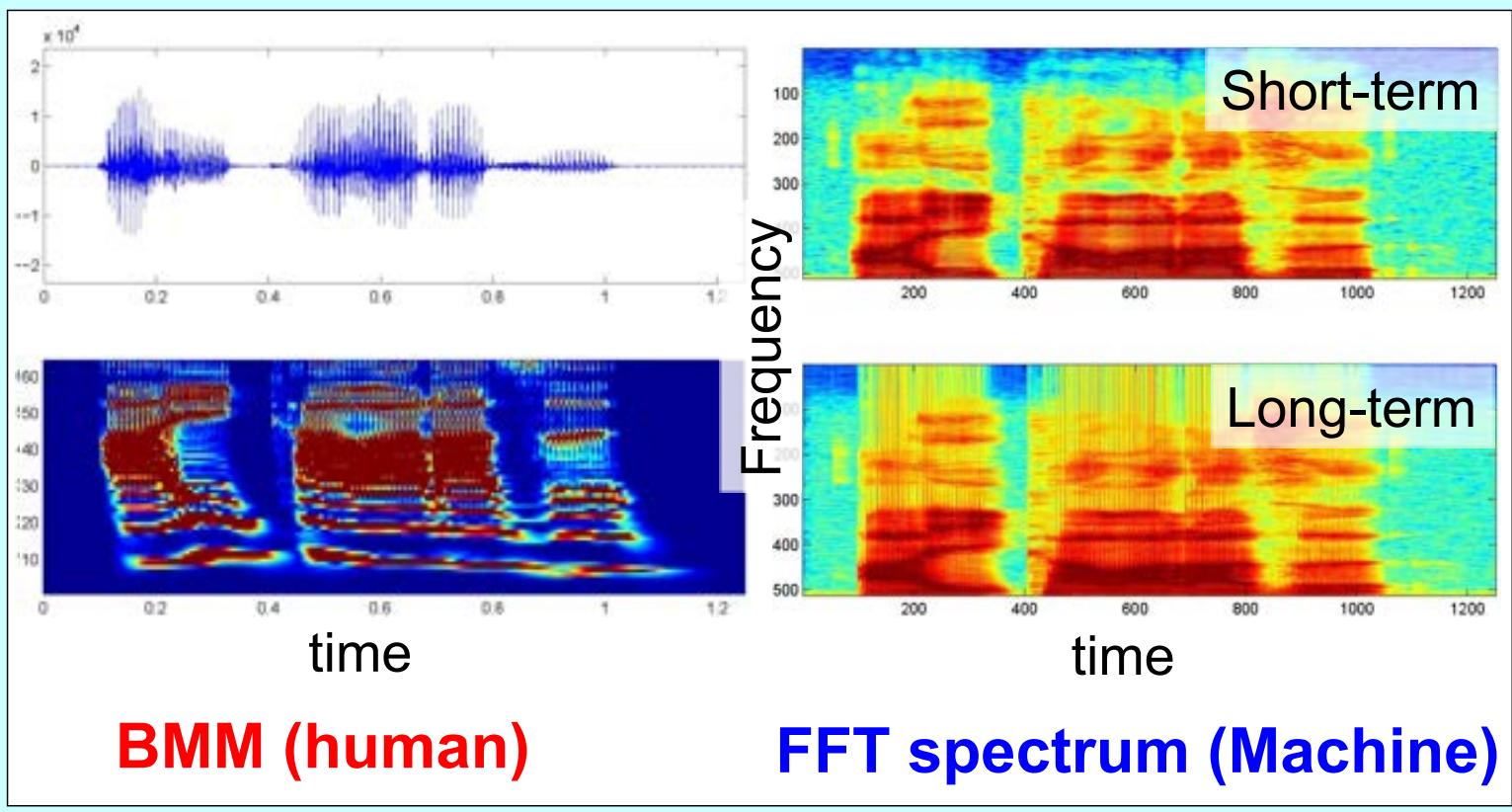
Voting result: Voice source: 2.6, Articulation: 2.8, Expression: 2.3, Overall judgment: 2.4 AVERAGE: 2.5 (1st place)

http://www.interspeech2007.org/Technical/synthesis_of_singing_challenge.php

5. 聴覚フィルタバンク

- 聴覚末梢系 = 周波数分析器と等価（非線形処理）
- 聴覚特性を考慮した信号分析合成
⇒ 聴覚フィルタバンク





I485F 知覚情報処理特論

5. 聴覚末梢系の生理とモデル

2023年5月6日(土)

人間情報学研究領域

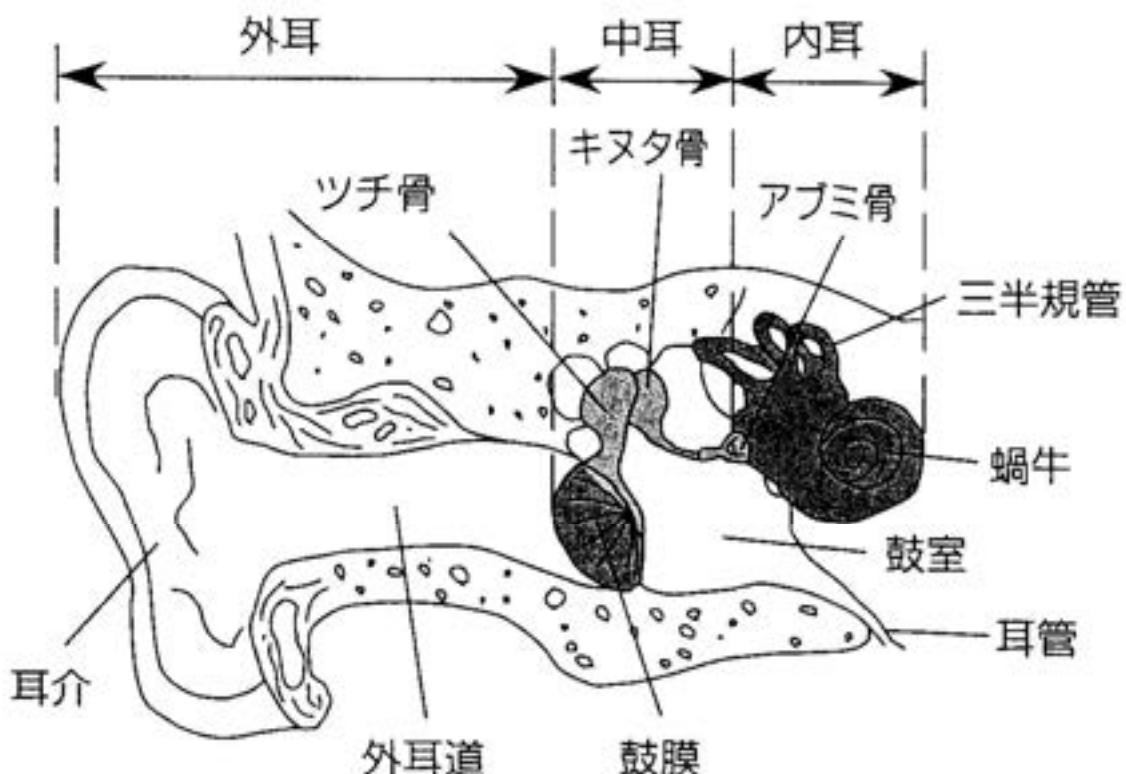
鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

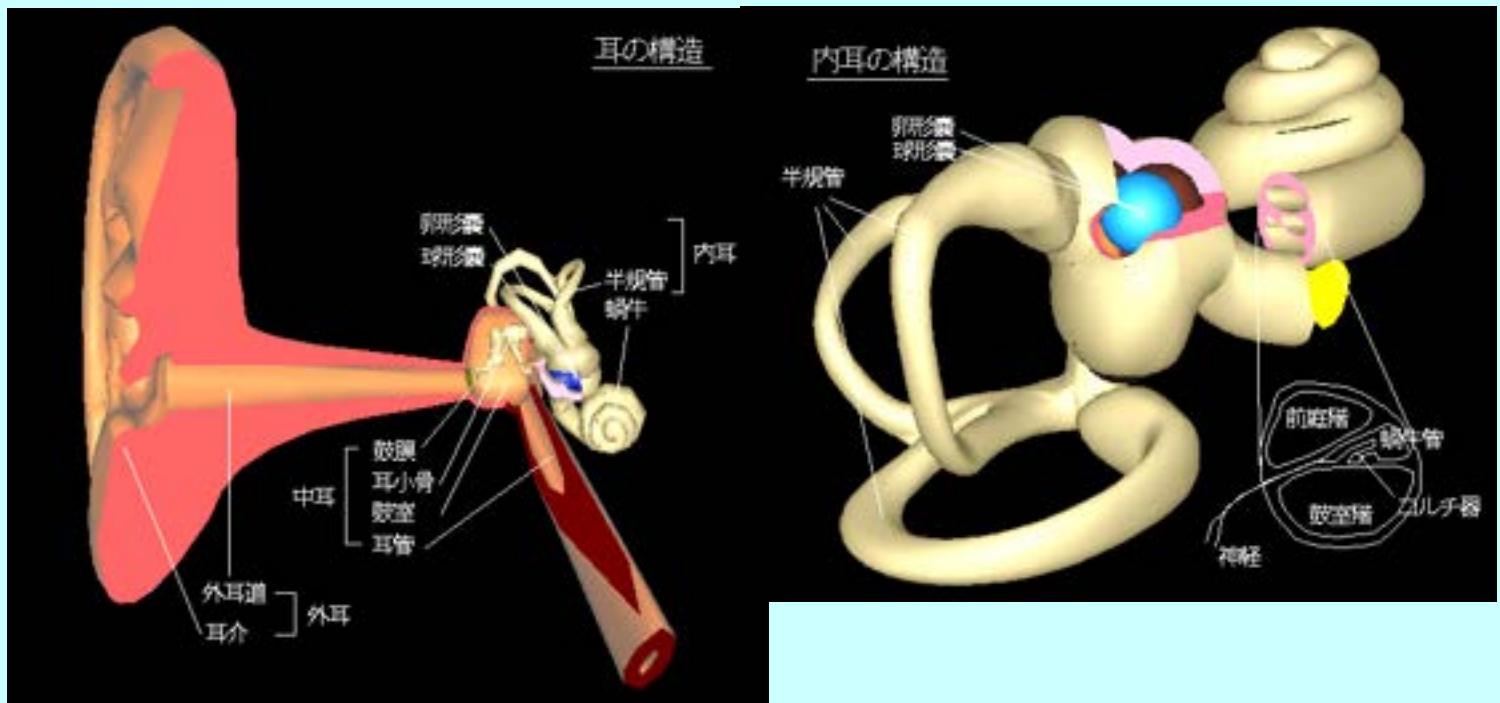
Email: unoki@jaist.ac.jp



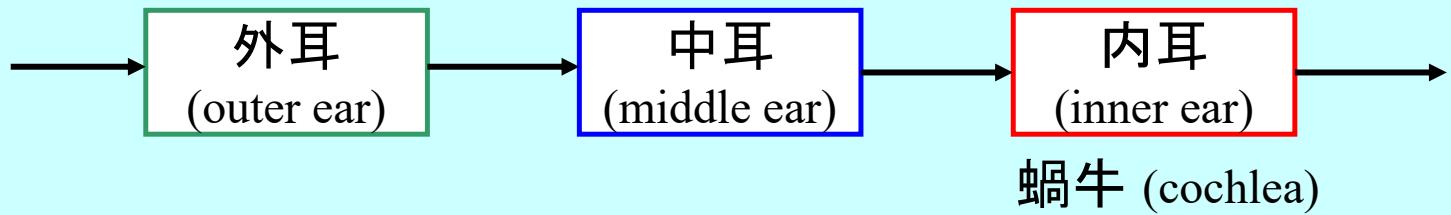
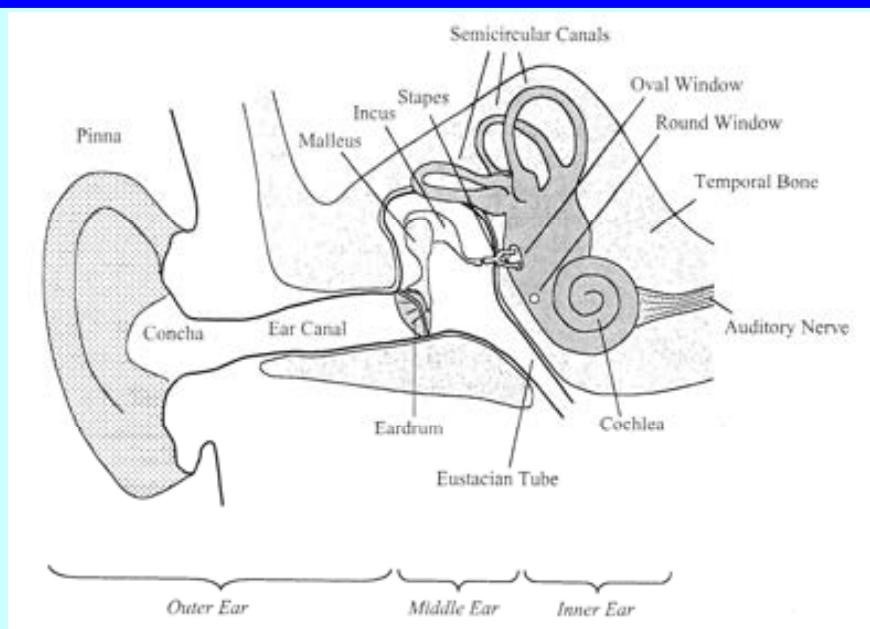
1. 聴覚系の基本的な構造と機能



聴覚の構造



2. 聴覚の機能の細分化



2.1 外耳

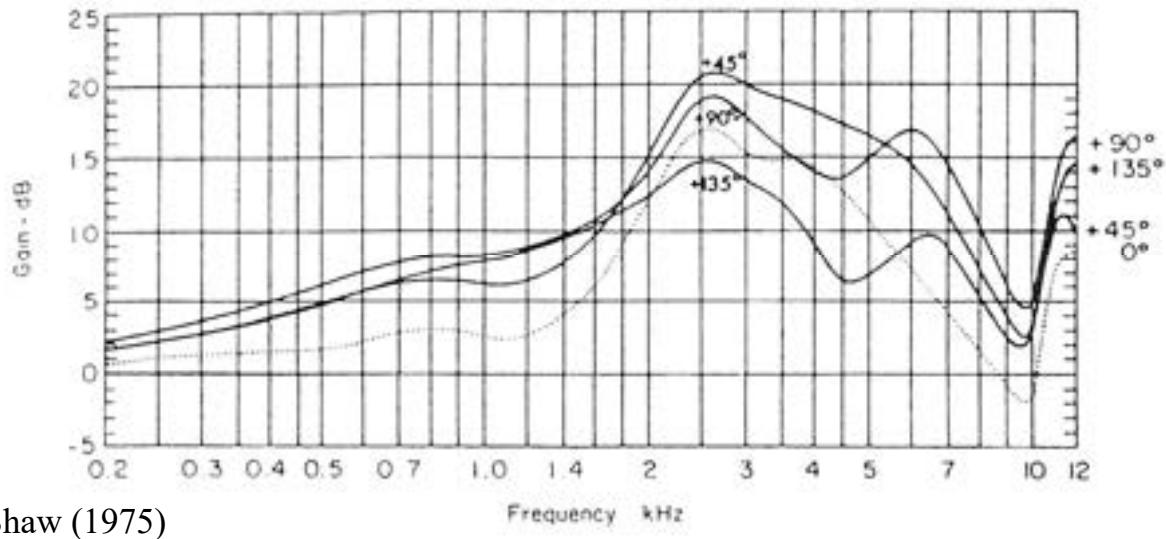
■構成

耳介 (pinna)

外耳道 (auditory canal)

■機能

高域強調→空間知覚で重要



2.2 中耳

■構成

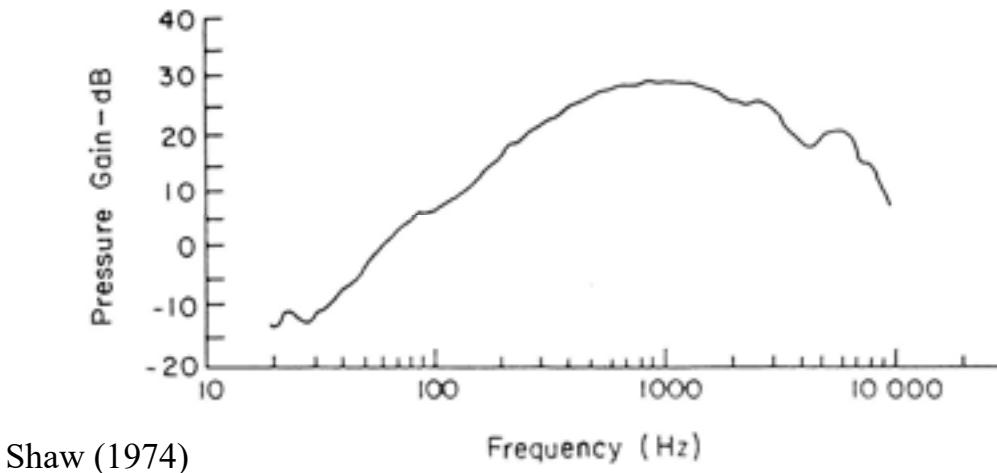
耳小骨 (ossicles; 人間の中で一番小さい骨)

■ つち骨 (malleus), きぬた骨 (incus), あぶみ骨 (stapes)

■機能

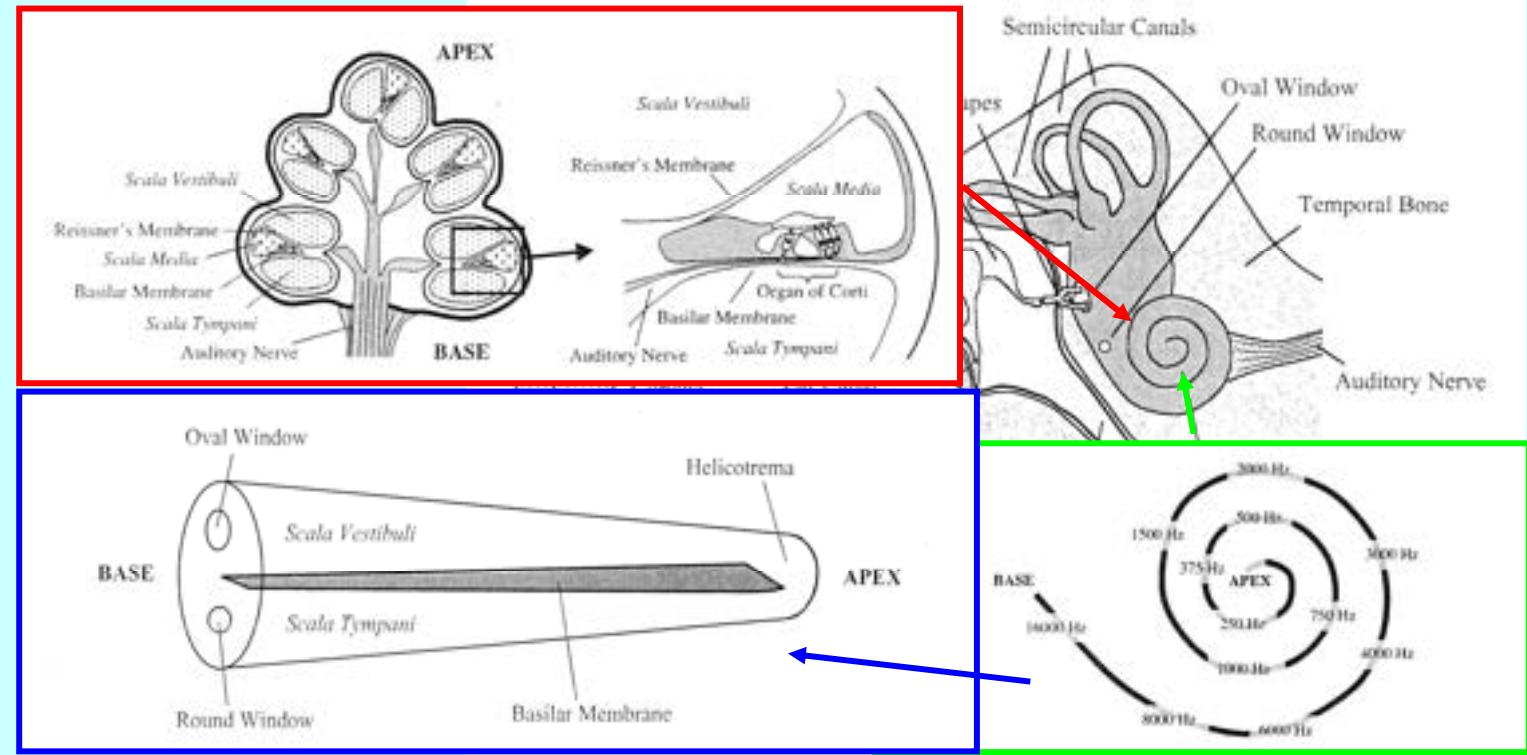
インピーダンスマッチング

■ 音響インピーダンス (空気の圧力変化→液体の圧力変化)



2.3 内耳

■ 蝸牛 (cochlear)



蝸牛の部位

■ 二つの膜で分割

- ライスネル膜 (Reissner's membrane)
- 基底膜 (basilar membrane)

■ 二つの室で構成

- 前庭階 (scale vestibuli)
- 鼓室階 (scala tympani)

始点 蝸牛底 (base)

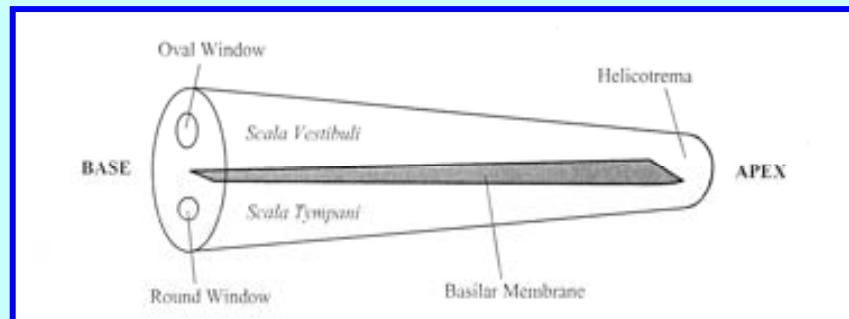
先端 蝸牛頂 (apex)

■ 入口

- 前庭窓 (oval window)

■ 開口部

- 蝸牛窓 (round window)

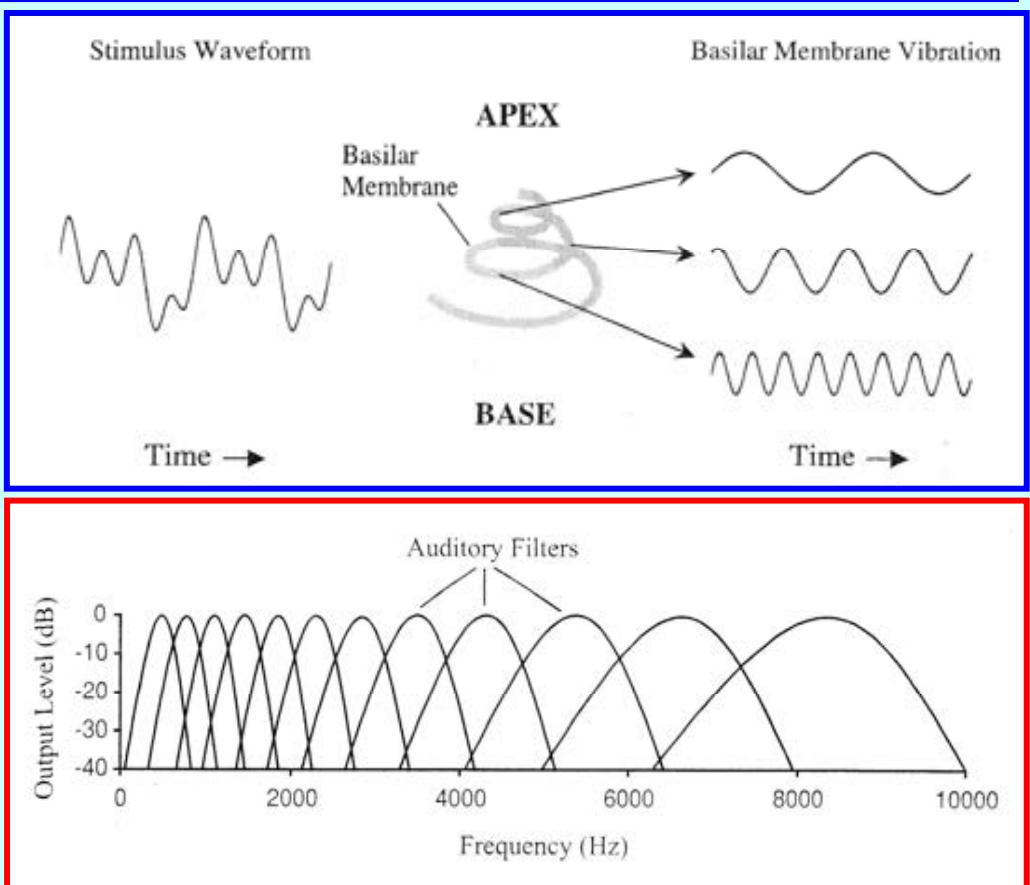


蝸牛の働き

■ 周波数分解

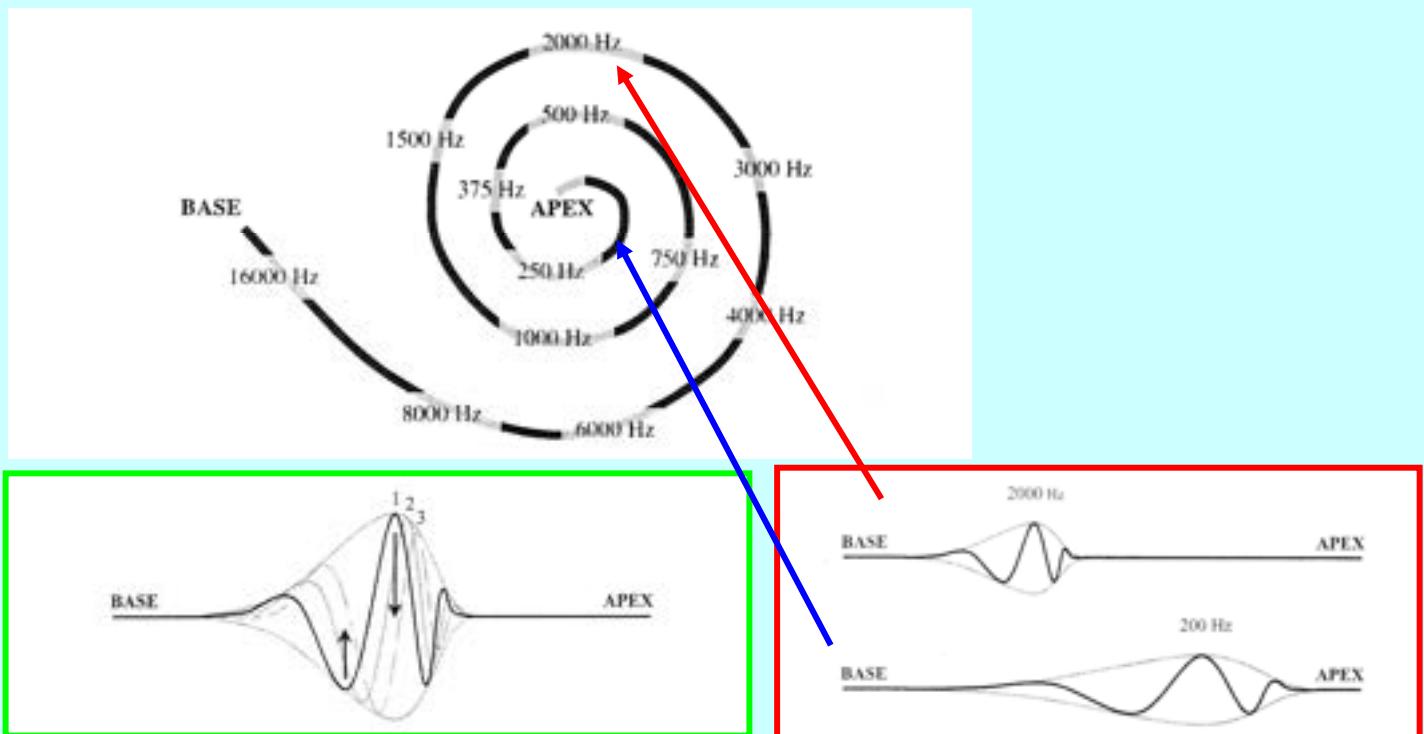
基底膜
(basilar membrane)

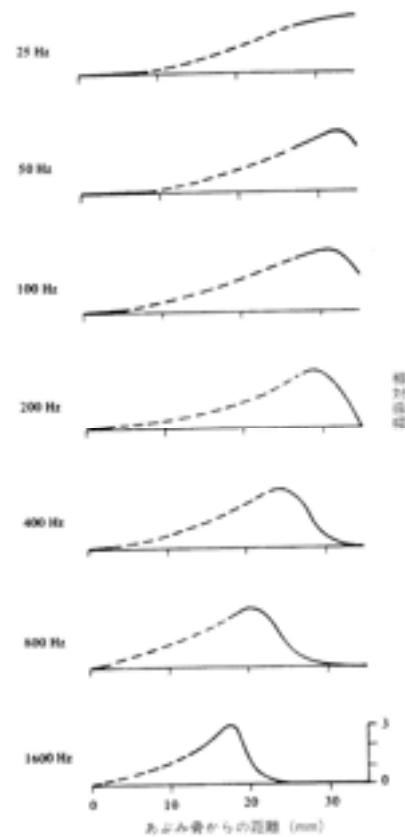
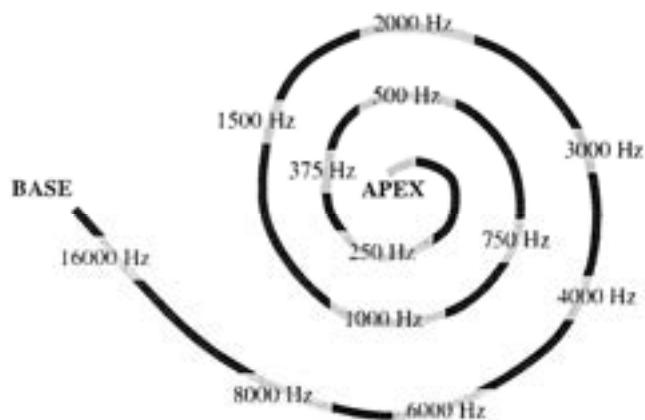
場所の情報
→周波数の情報



基底膜の働き

■ 基底膜位置の変動 → 周波数応答



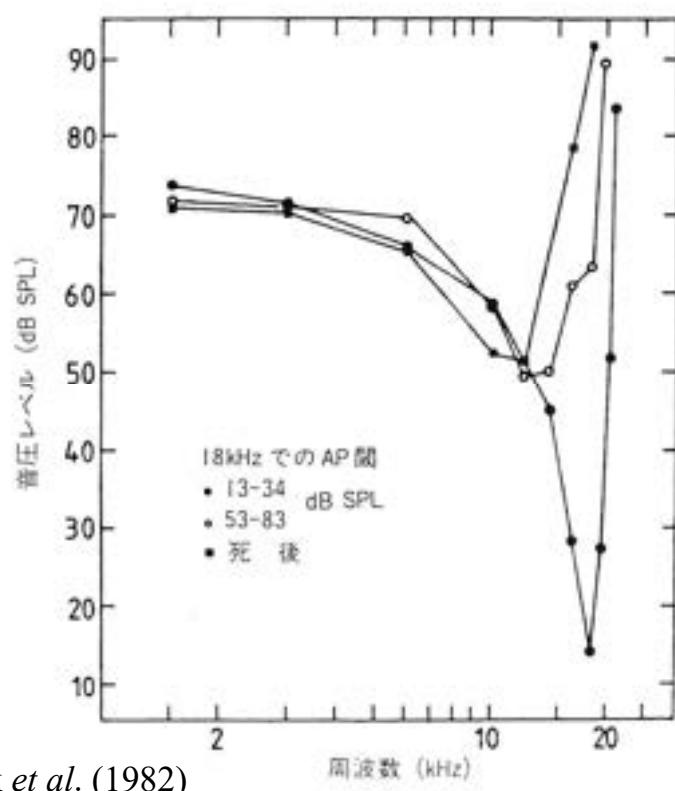


基底膜振動
(basilar membrane motion)
 ←二つの階の圧力変動により
 生じる
 周波数に応じて遅延が見られる

基底膜振動パターンの共振特性

■ **同調特性**
 (tuning curve)

同調の度合い
 → Q値 (Q_{10dB})
 3~10

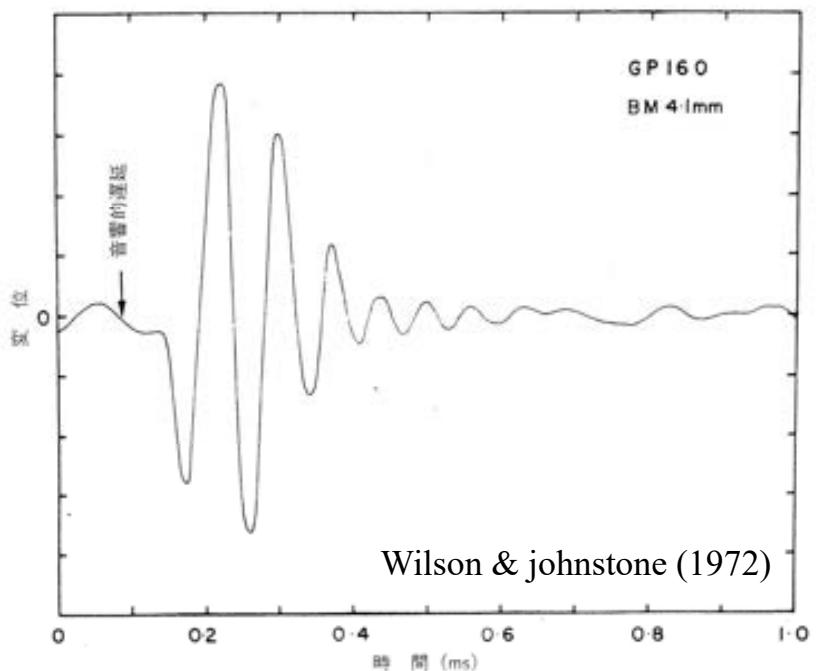


Sellick et al. (1982)

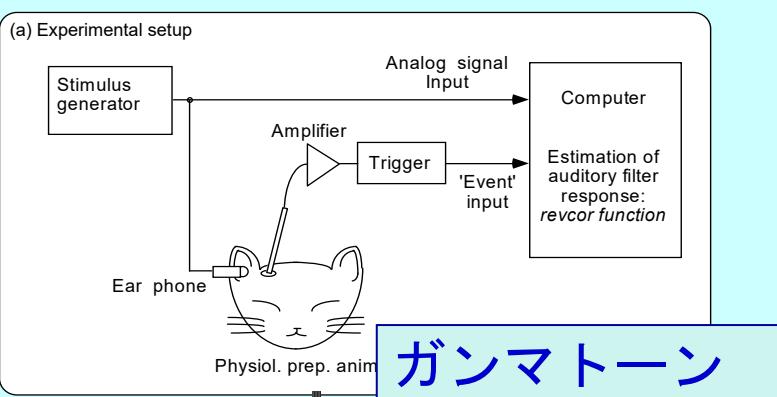
基底膜のインパルス応答関数

■ 基底膜の測定方法

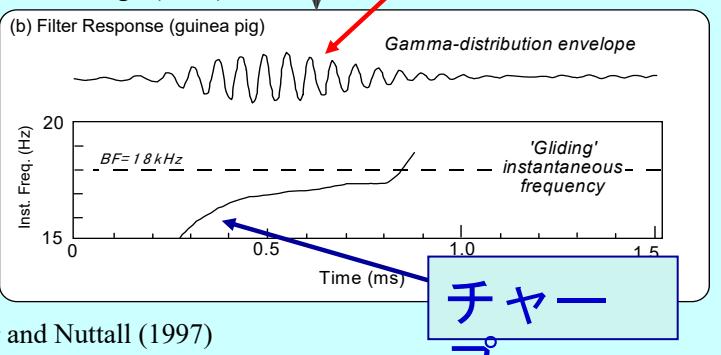
- メスバウアー効果
を利用するもの
- レーザー干渉法



蝸牛(基底膜)に関する生理学的知見

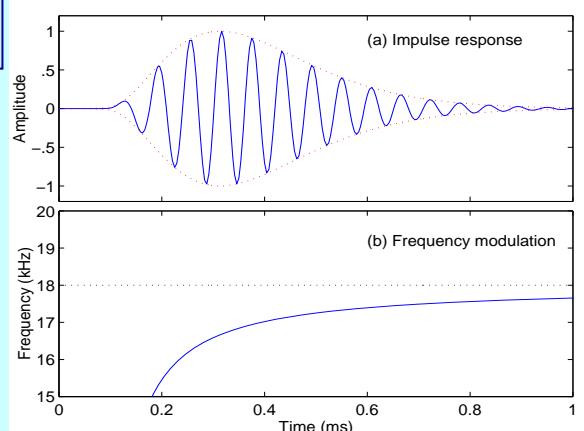


de Boer and de Jongh (1978)



de Boer and Nuttall (1997)

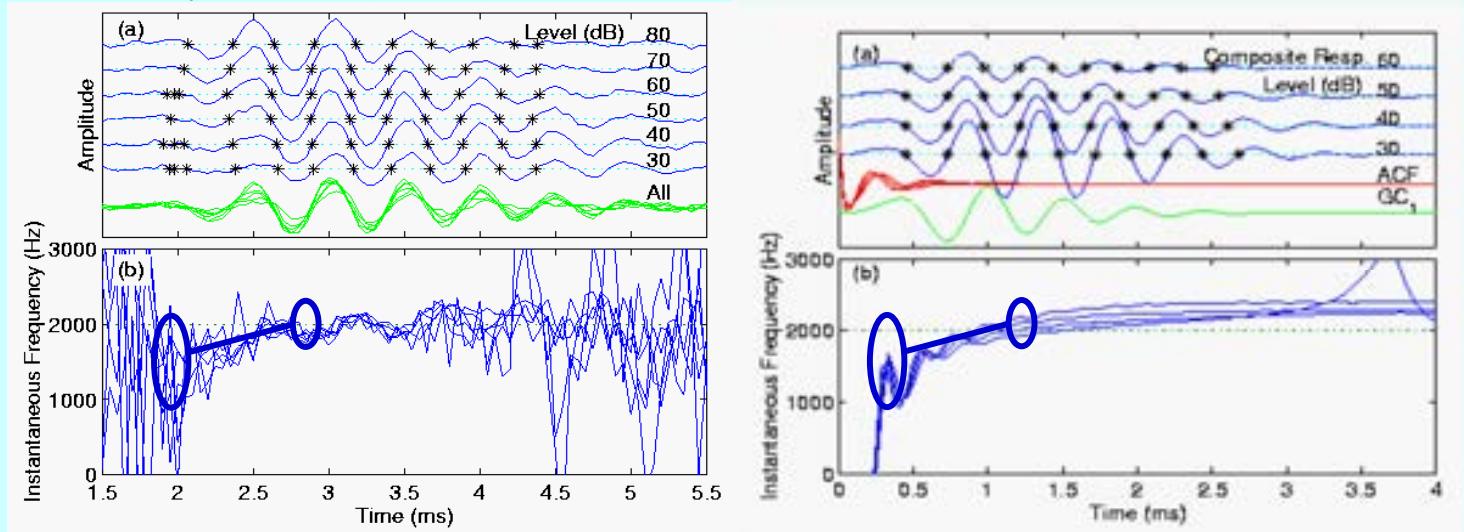
ガンマ+チャープ



生理学的データとの対応

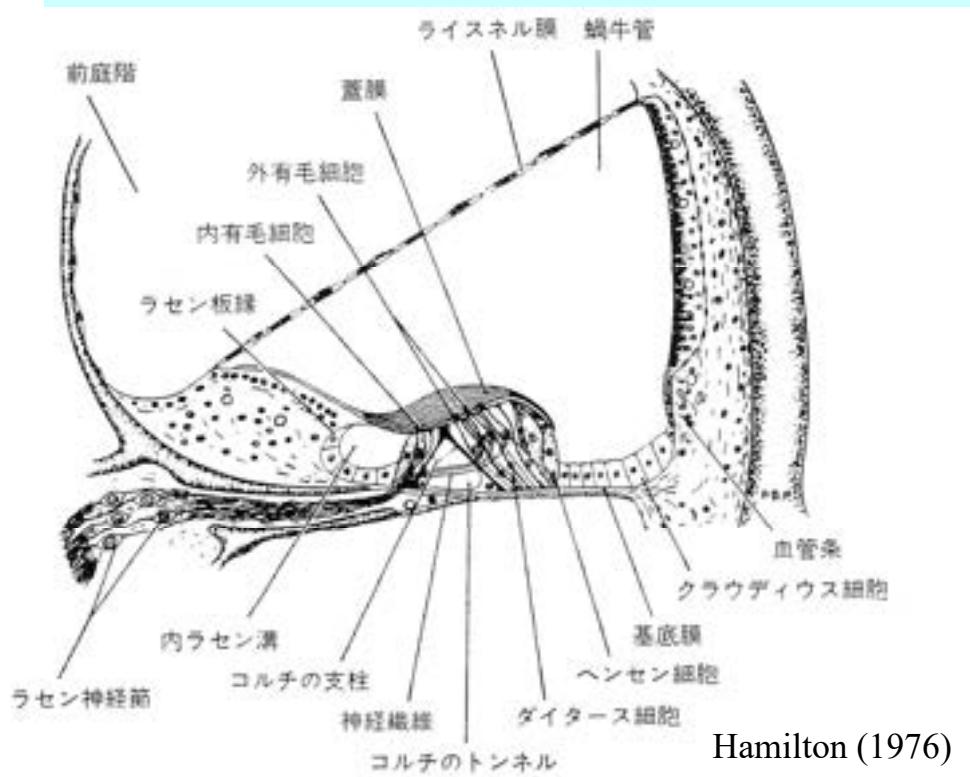
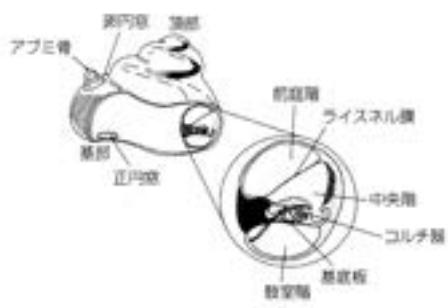
Carney *et al.* (1999)

圧縮型ガンマチャーブ



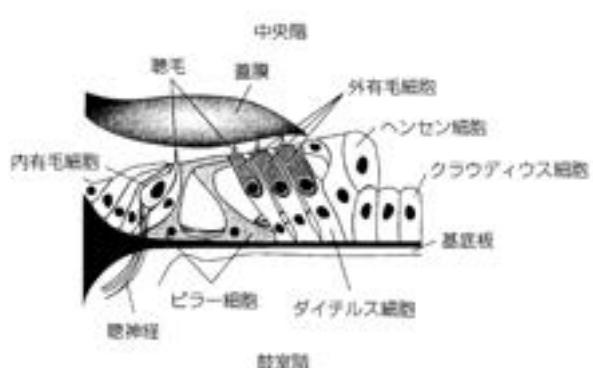
- ゼロ交差点の一致(レベルに依存しない)
- 周波数移行(チャーブ)の変化

2.4 変換過程

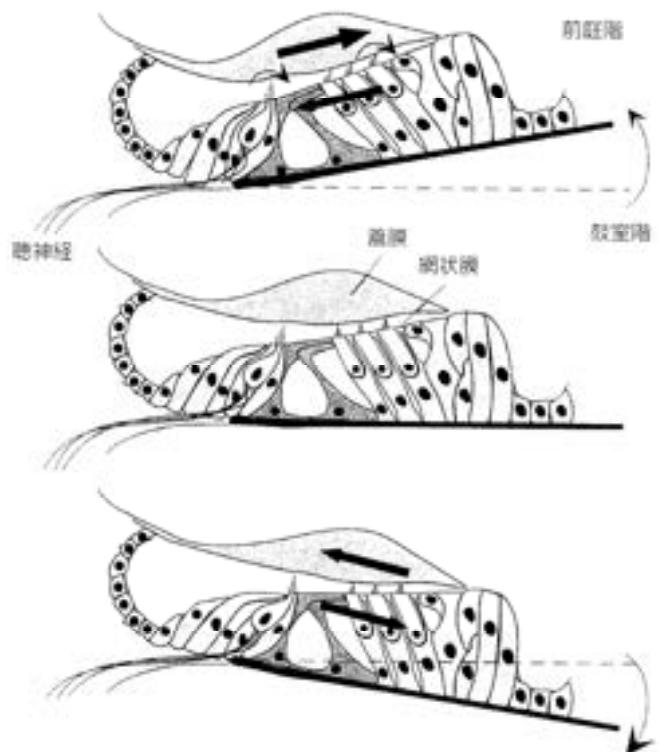


Hamilton (1976)

コルチ器



コルチ器は基底板上に位置し、その内部には内有毛細胞と外有毛細胞の2種類の感覚細胞が存在する。



① 有毛細胞

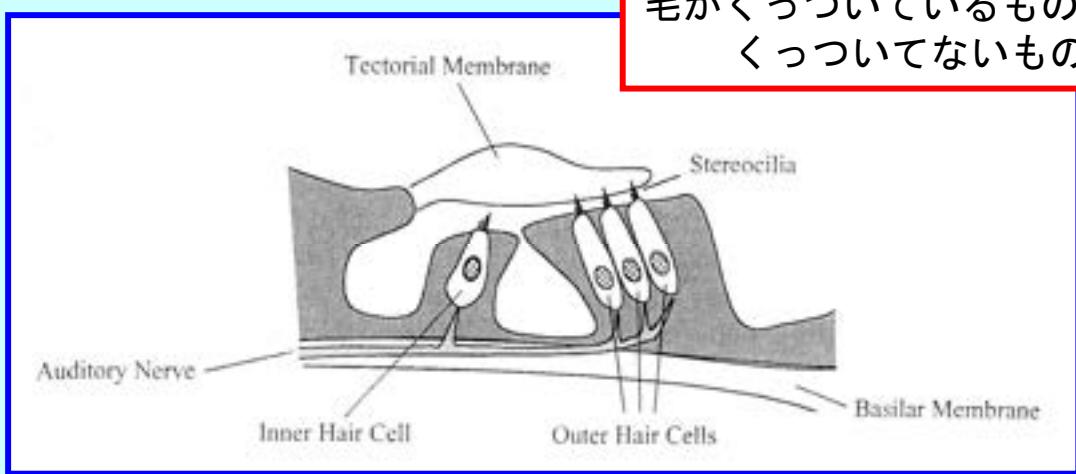
■ コルチ器 (organ of Corti)

■ 有毛細胞 (hair cell)

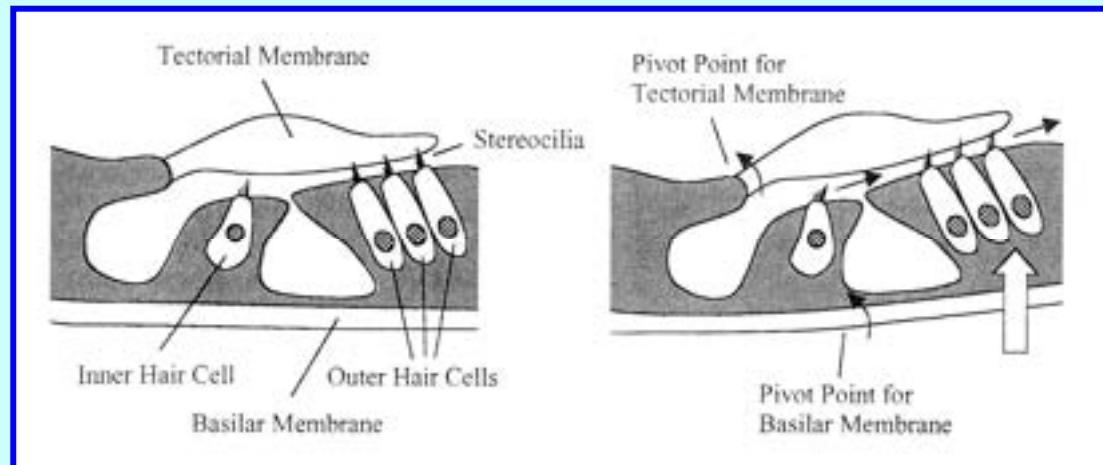
- 内有毛細胞 (inner hair cell: IHC) 3500個程度
- 外有毛細胞 (outer hair cell: OHC) 25000個程度

■ 蓋膜 (tectorial membrane)

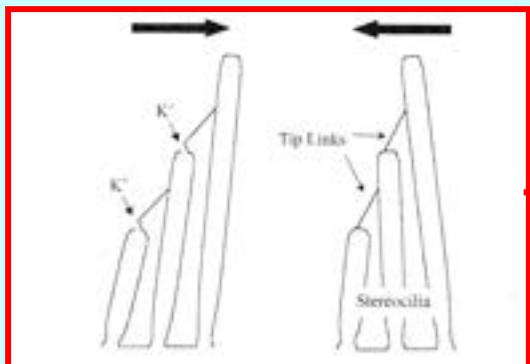
毛がくっついているもの → OHC
くっついてないもの → IHC



有毛細胞のはたらき



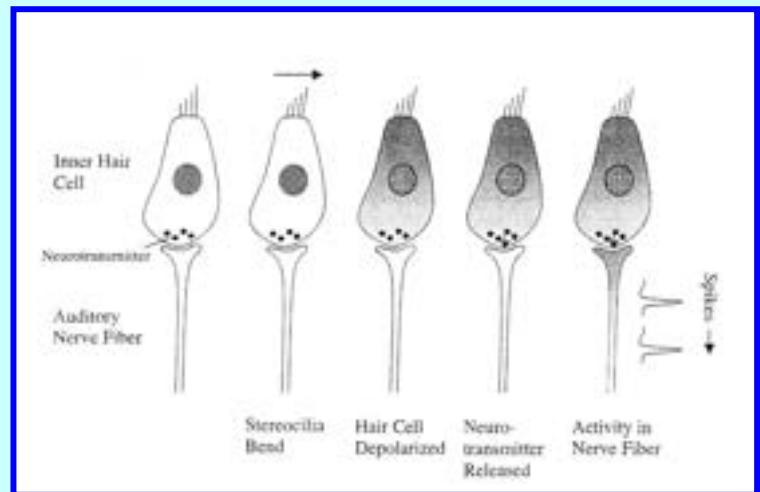
OHCは
増幅の働き



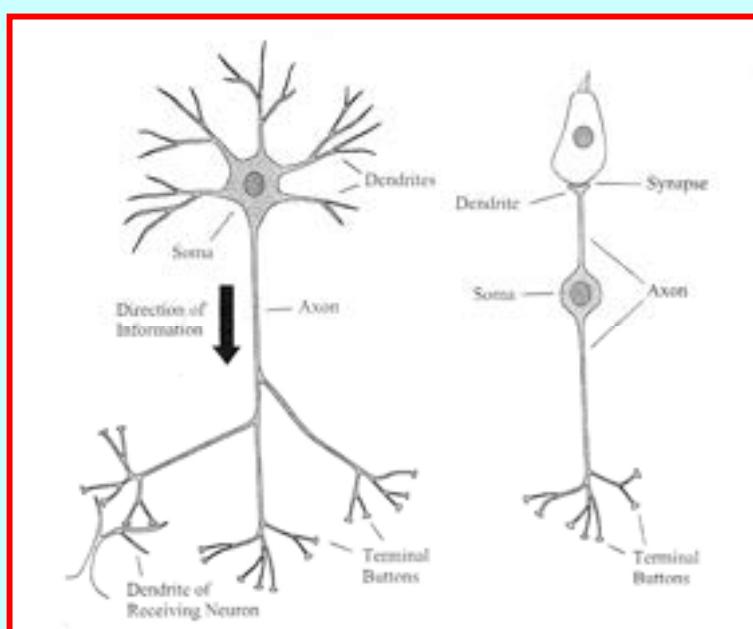
毛のたわみの変化
→ 細胞の興奮
→ 聴神経の活動
(神経発火)

②聴神経

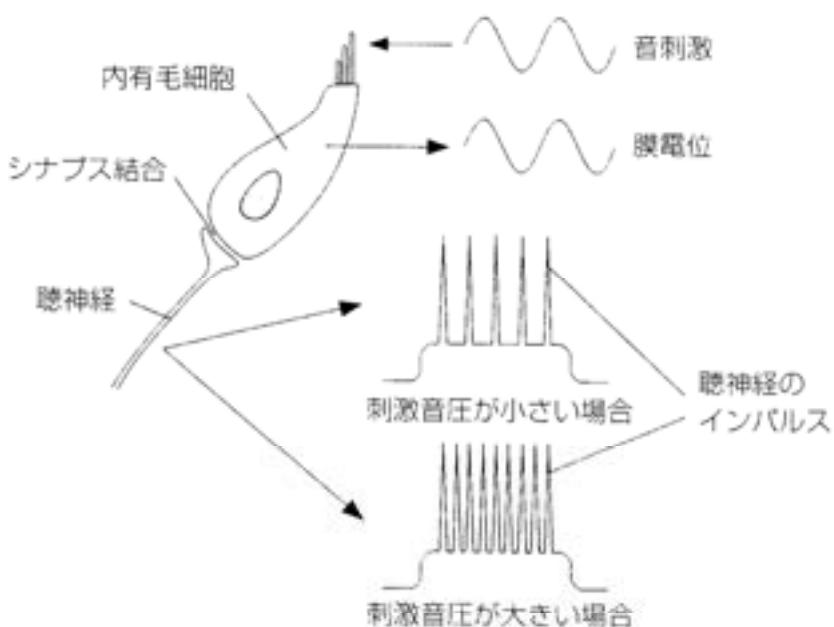
IHCのはたらき→聴神経のはたらき



神経発火



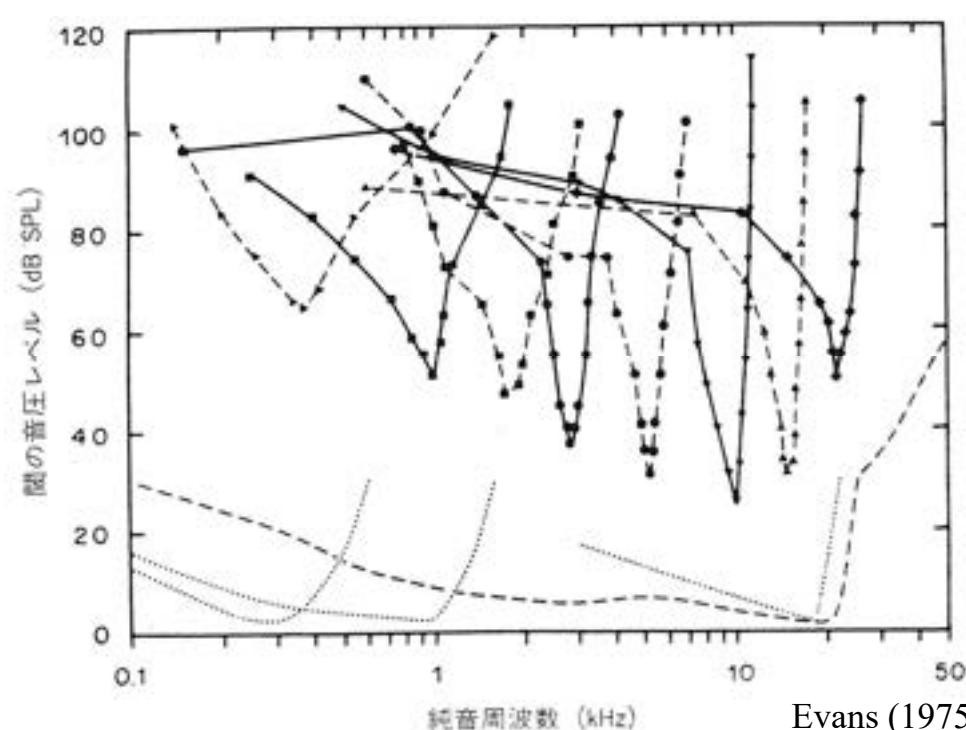
電気信号への変換



内有毛細胞の膜電位の変化は、音刺激と同様の波形を示すが、神経の活動電位はインパルス状である。このインパルスの発火頻度は入力音圧が大きいほど高い。

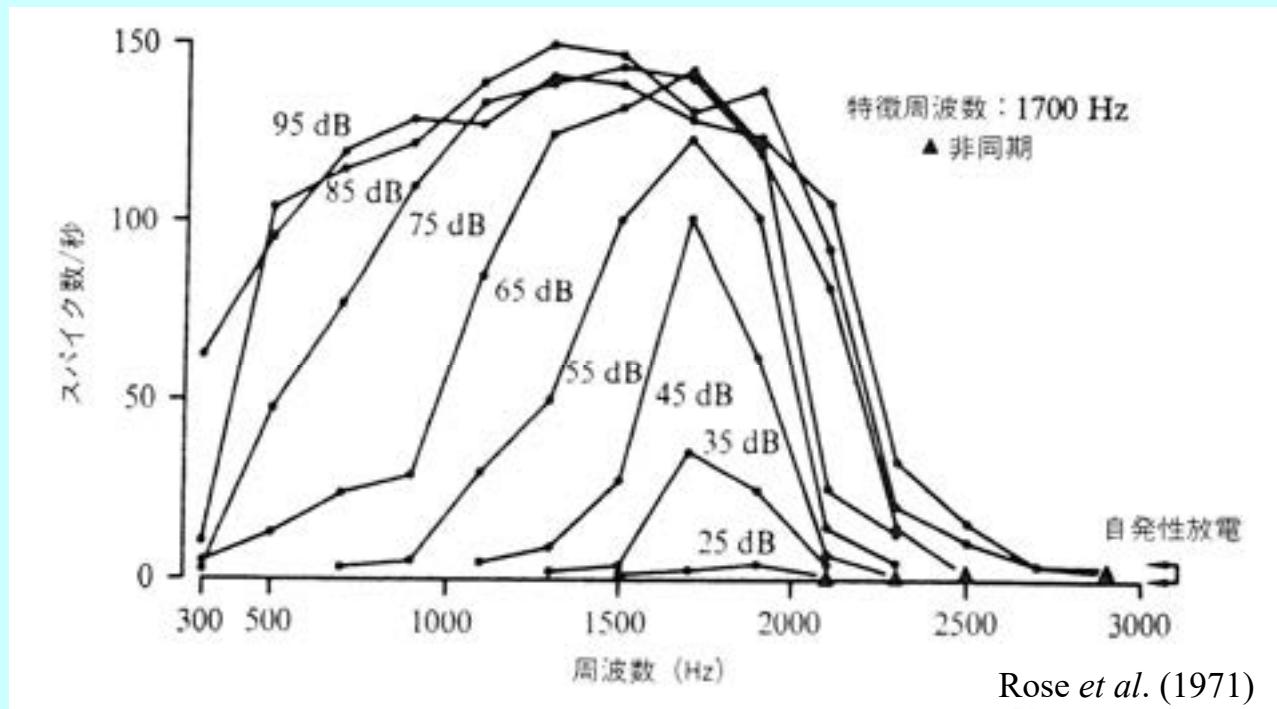
聴神経の同調特性

■ 麻酔したネコの聴神経の複数ニューロンから得た同調特性

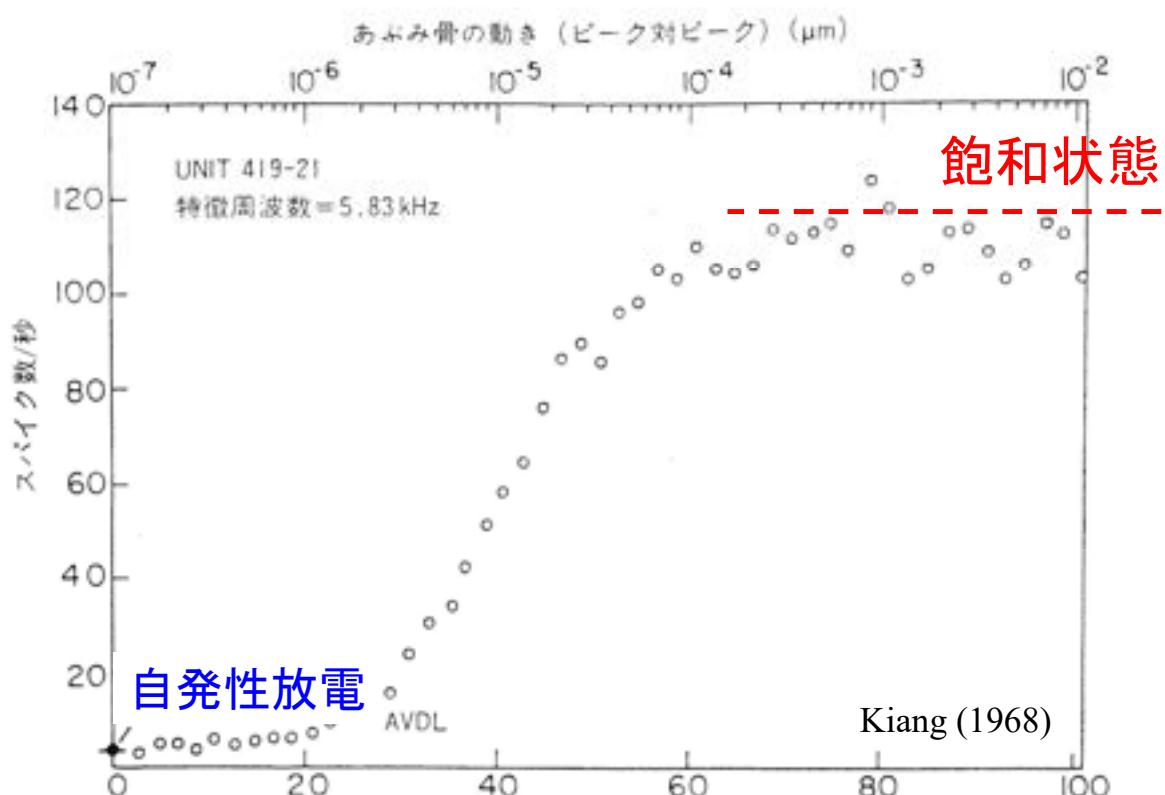


聴神経の等強度曲線

- 麻酔したリスザルの聴神経の単一ニューロンから得た等強度曲線



聴神経の発火とレベルの関係

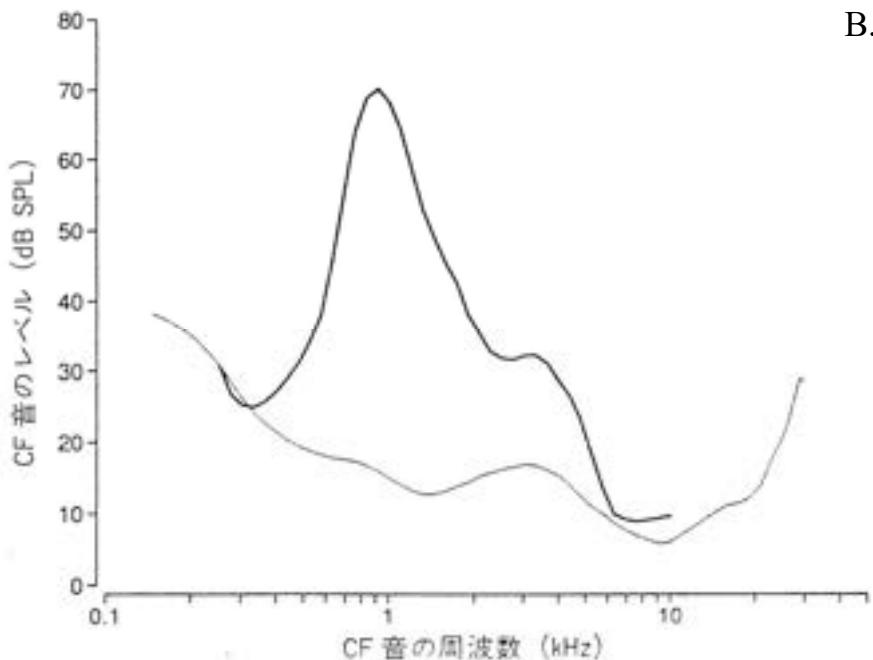


神経興奮パターン (excitation pattern)

■ 特徴周波数 (CF)

神経線維の閾値が最小になるときの周波数

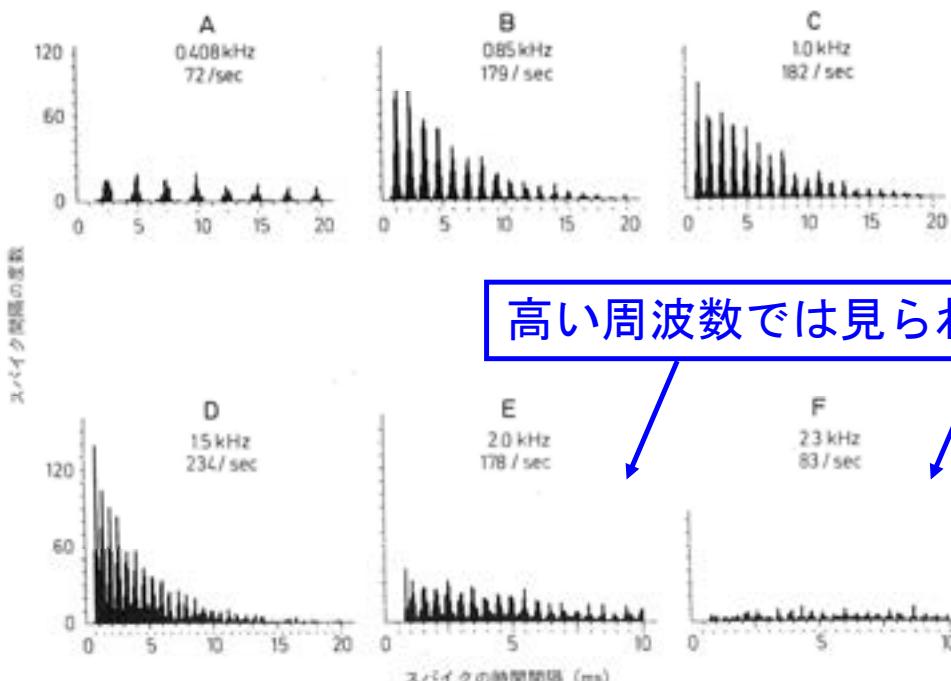
B. Delgutte



位相固定 (phase locking)

■ 発火間の時間間隔が刺激波形の周期の整数倍になる

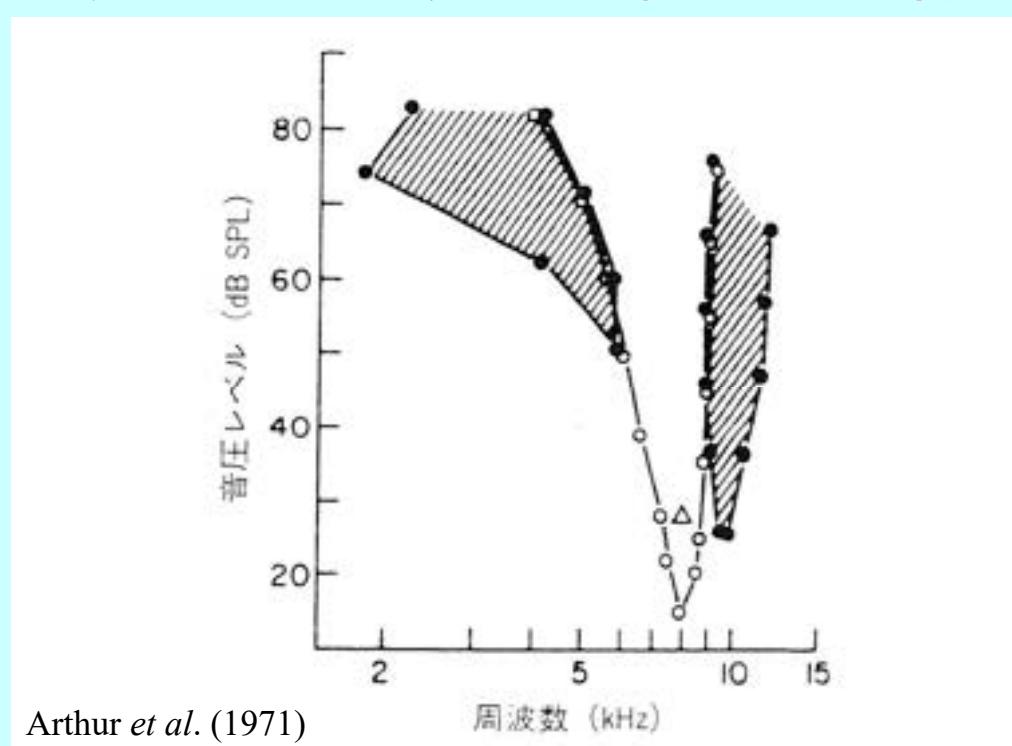
Rose et al. (1968)



高い周波数では見られなくなる

二音抑圧 (two-tone suppression)

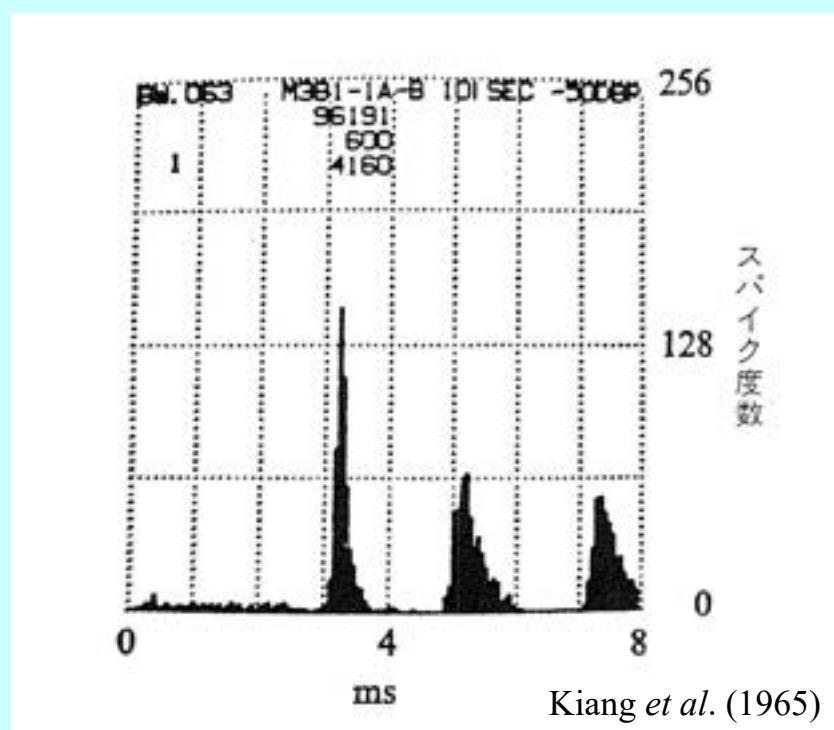
- ある音に対する反応がもう一つの音によって抑圧される現象



複合音に対する位相固定

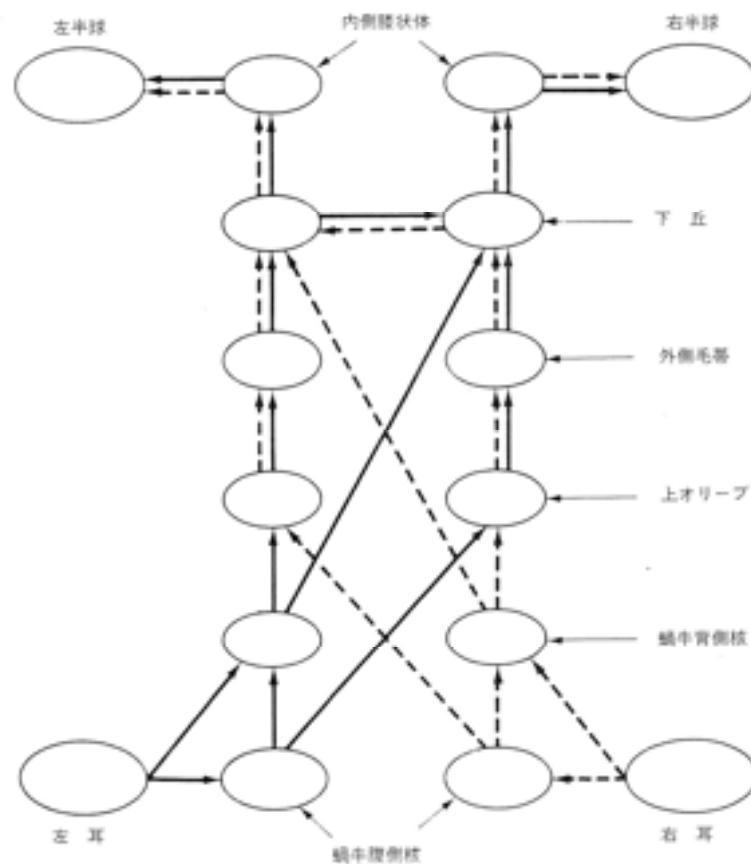
- PSTヒストограм (post-stimulus time histogram)

クリック音に
対する発火頻度

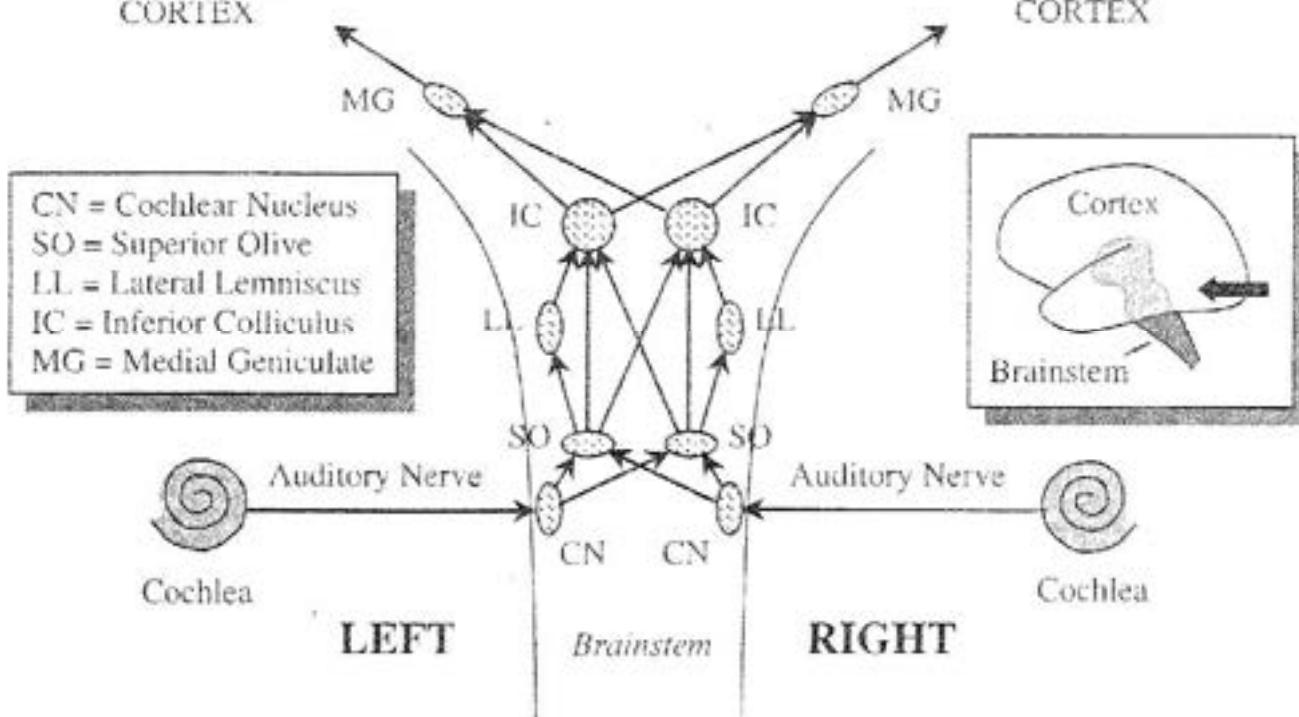


2.5 聴覚高次レベルにおける神経反応

伝達経路



AUDITORY CORTEX AUDITORY CORTEX



I485F 知覚情報処理特論

6. 聽覚系の非線形性 (圧縮と抑圧, 耳音響放射)

2023年5月6日(土)

人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

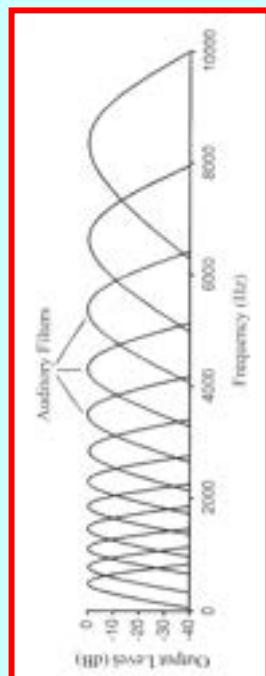
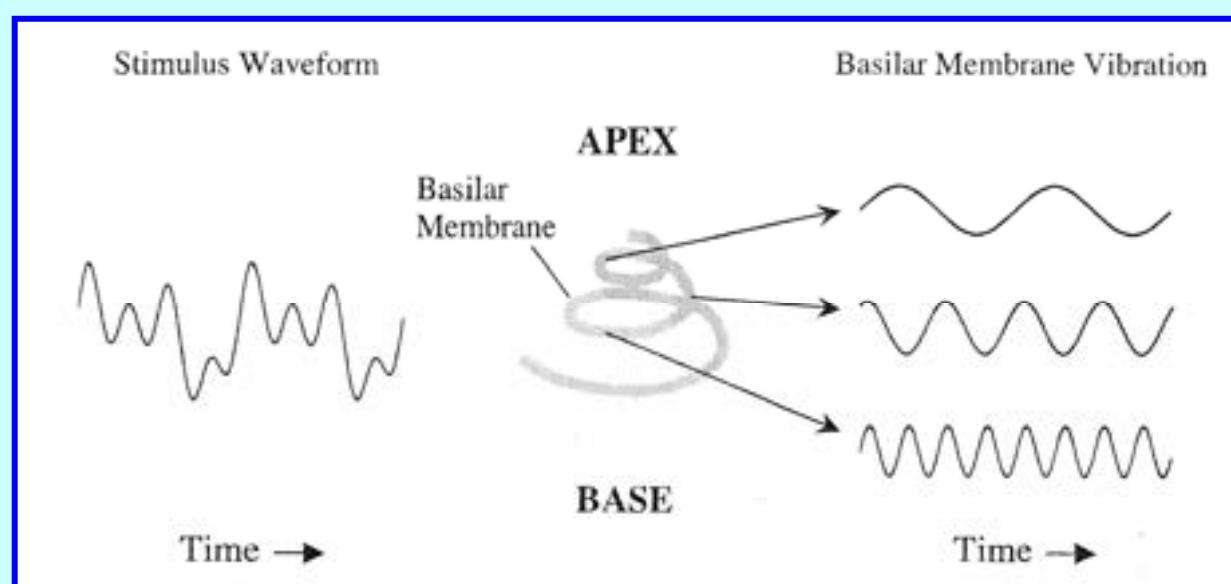
Email: unoki@jaist.ac.jp



1. 周波数選択性

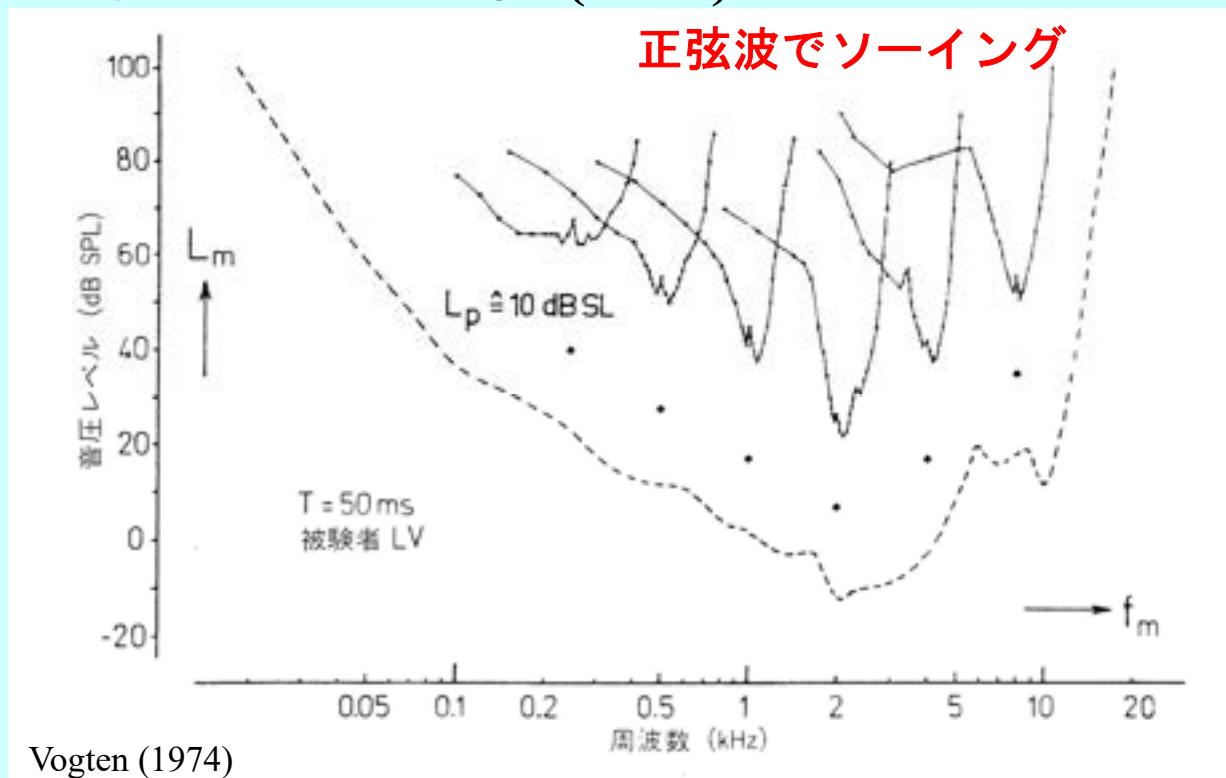
■周波数選択性 (frequency selectivity)

- 複合音中の正弦波成分を分解する能力

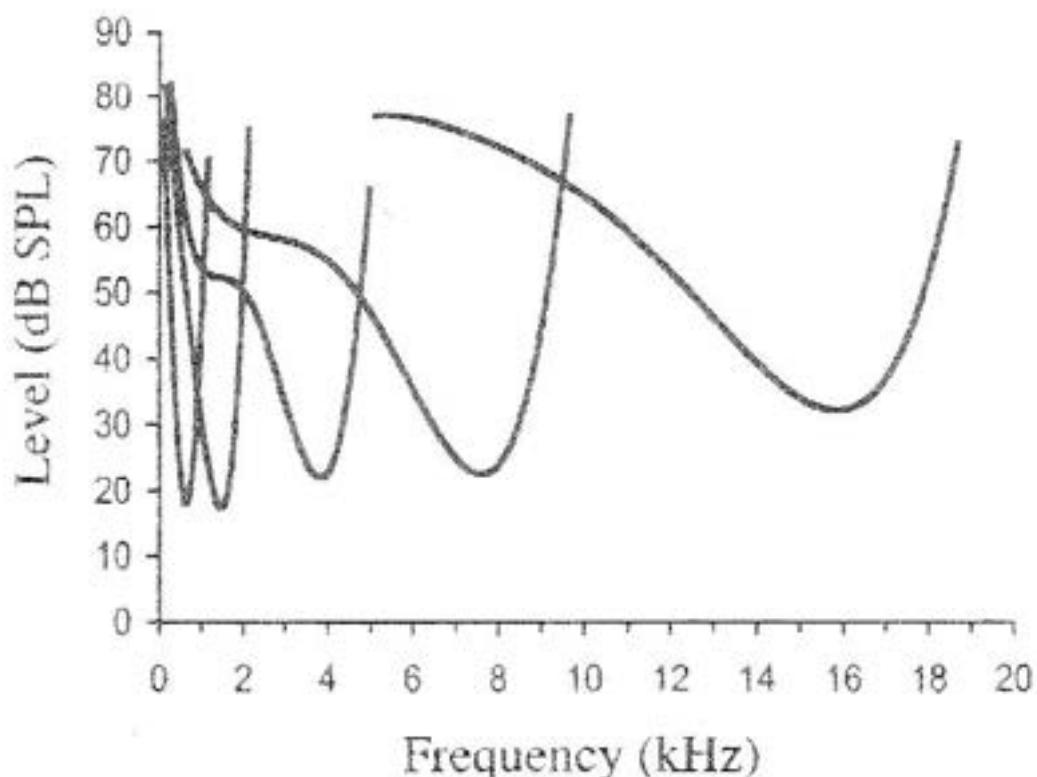


2. フィルタ形状の推定

心理物理的同調曲線 (PTC)

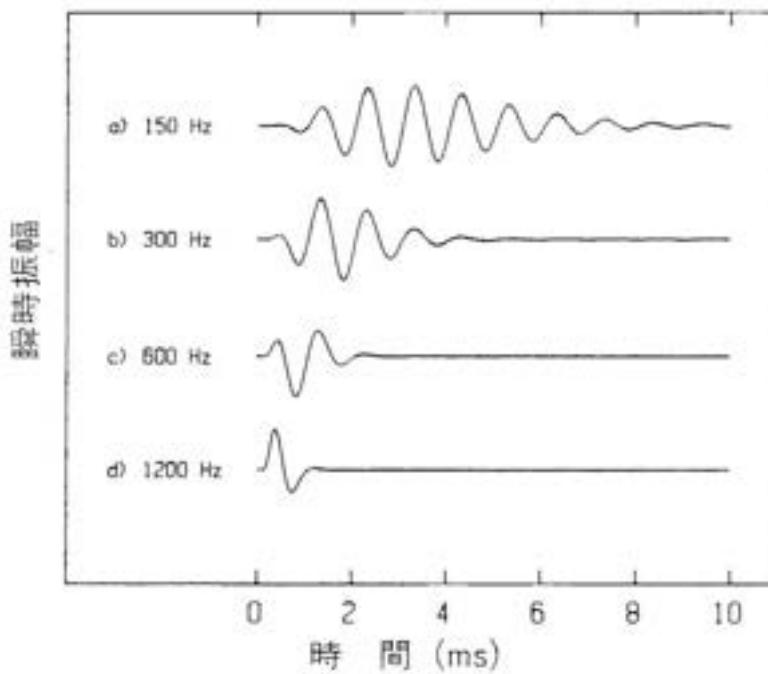


聴神経の発火特性（同調特性）



フィルタの応答

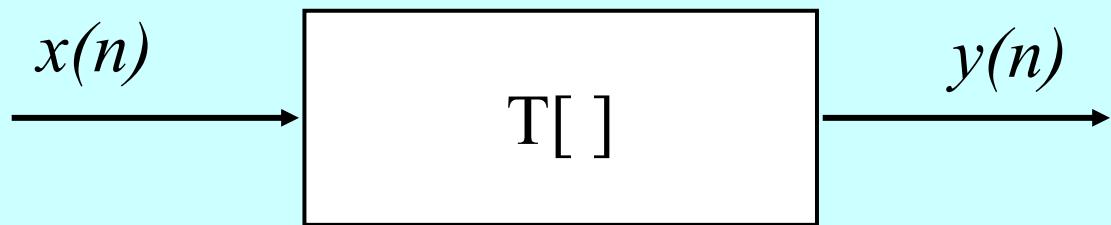
■ インパルス応答 (impulse response)



3. 線形性の概念

■ 線形性の定義

$$y(n) = T[x(n)]$$



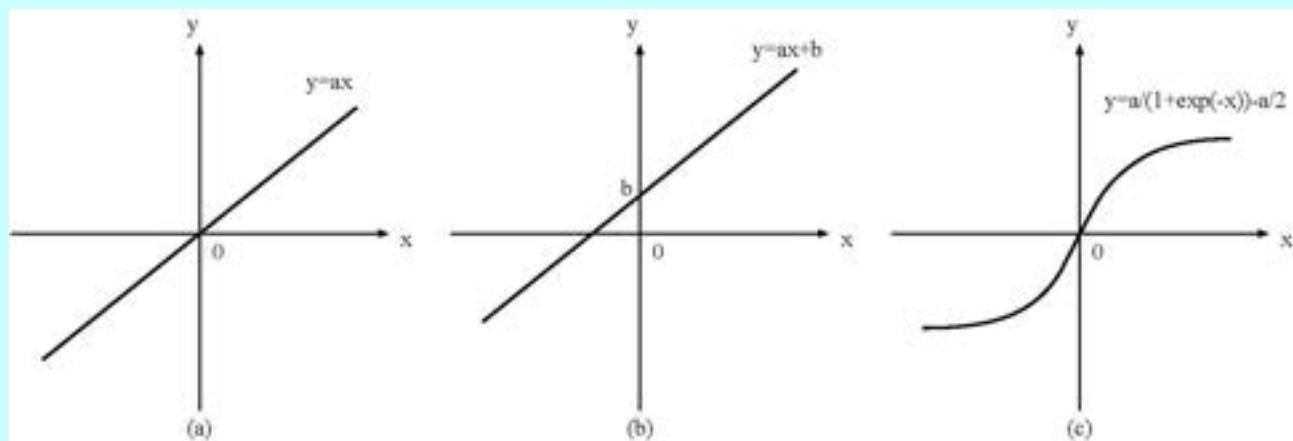
$$\begin{aligned} T[ax_1(n)] + T[bx_2(n)] &= aT[x_1(n)] + bT[x_2(n)] \\ &= ay_1(n) + by_2(n) \end{aligned}$$

■ 線形性の条件

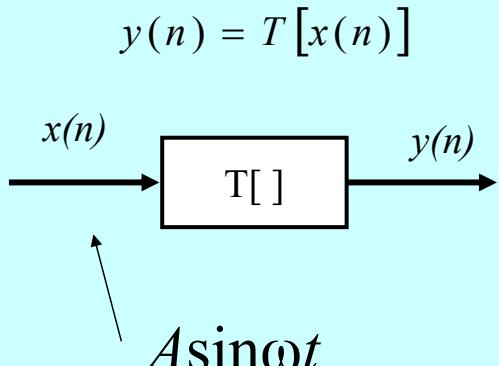
- 加法性 (Additivity property)
- 同次性 (Homogeneity property)
- 重ね合わせ (Superposition property)

これ以外
はすべて
非線形性

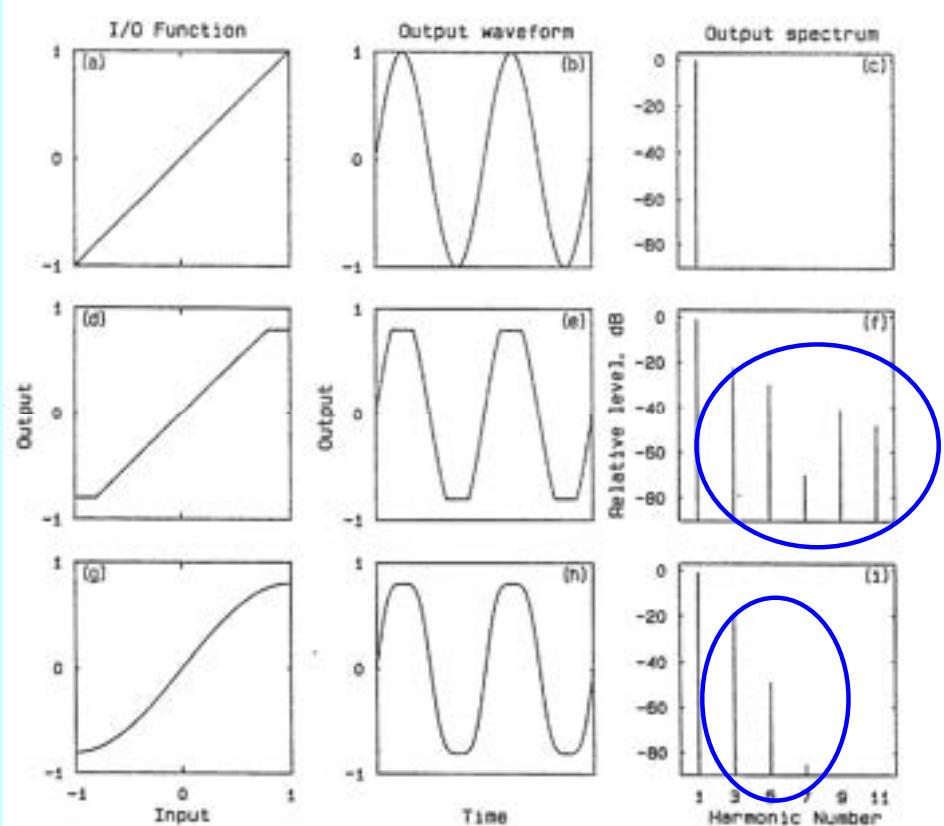
■ どれが線形システムか？



例題

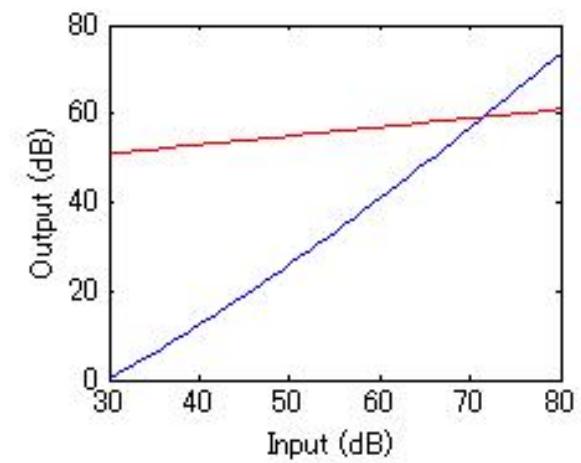
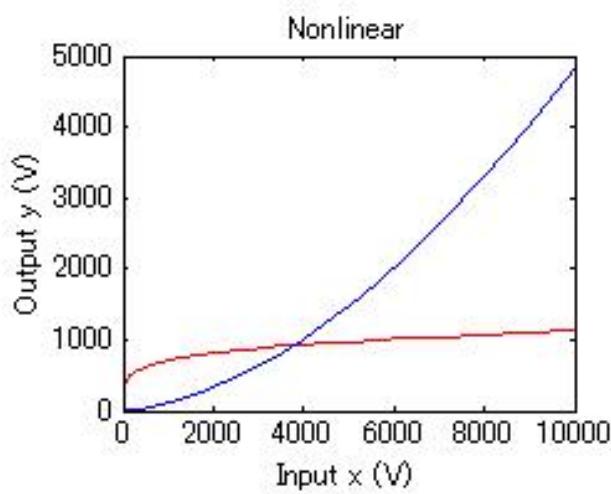
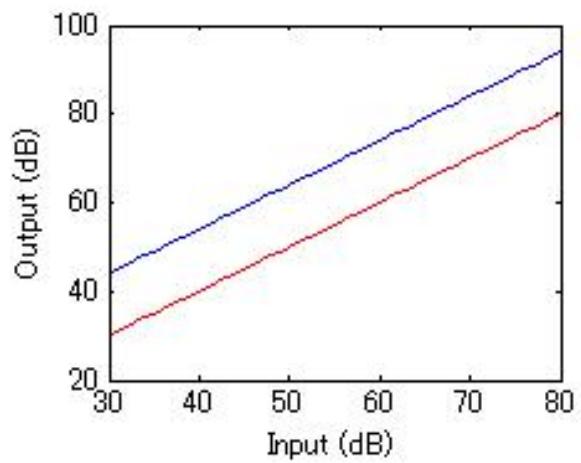
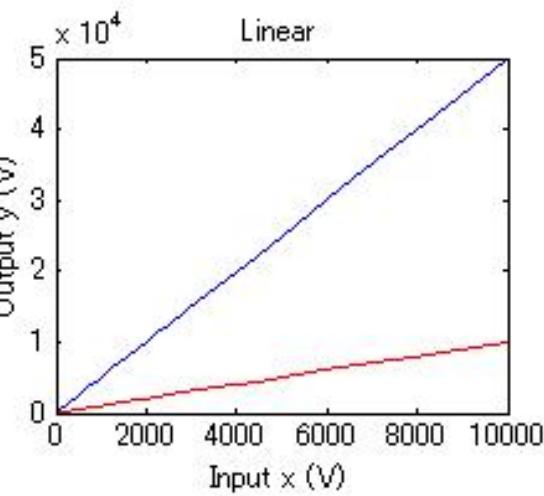


$A\sin\omega t$



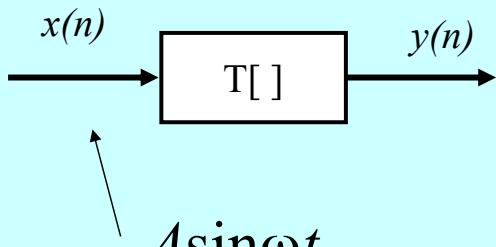
Question:

何故、高調波が
見られるのか？

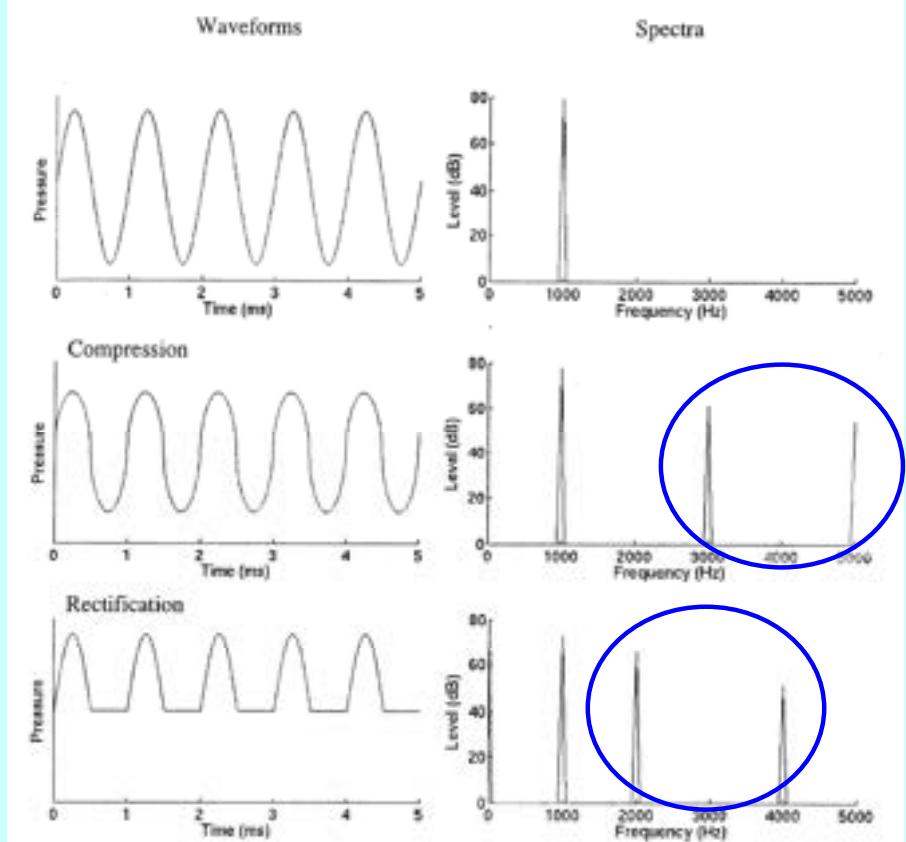


歪積

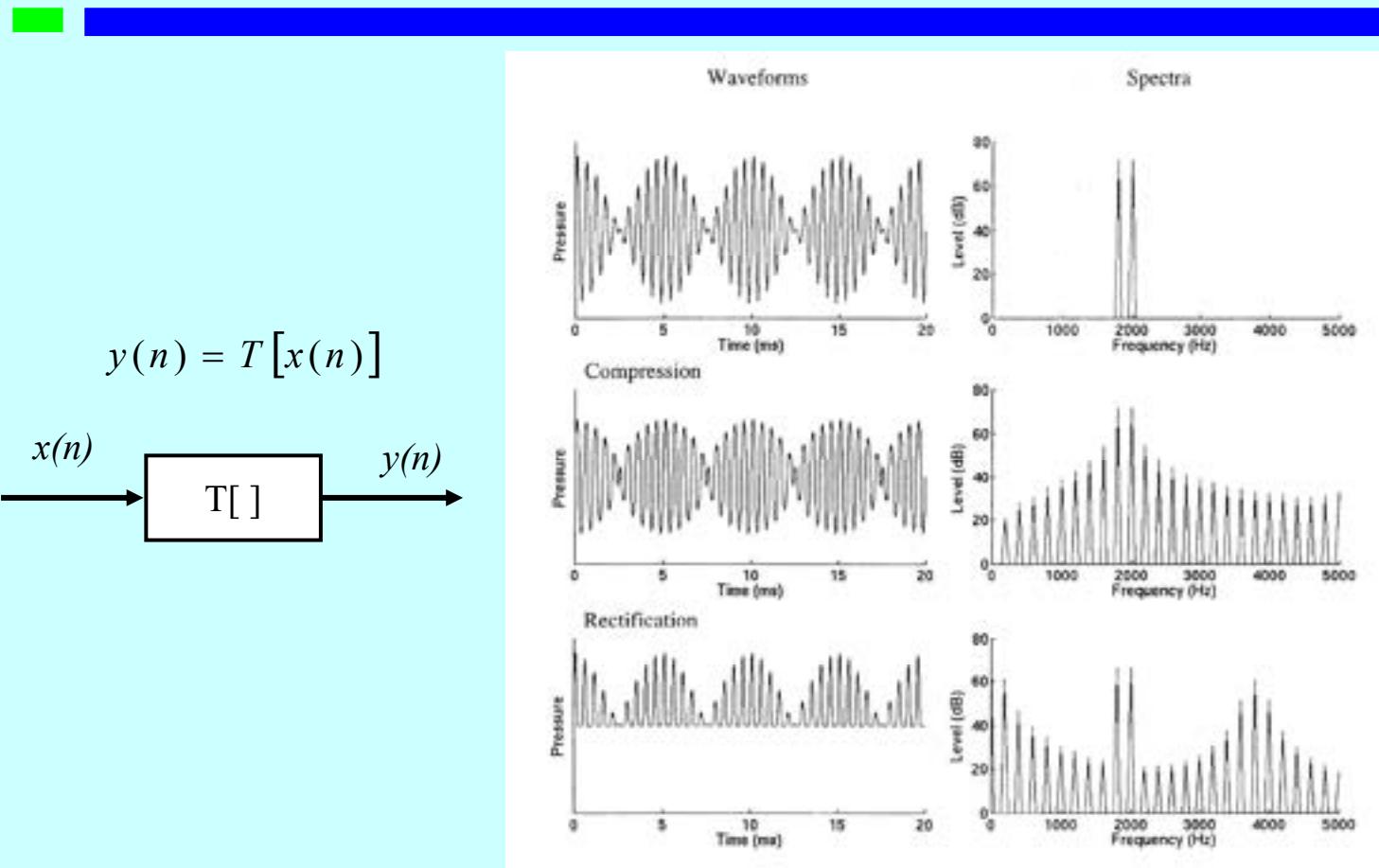
$$y(n) = T[x(n)]$$



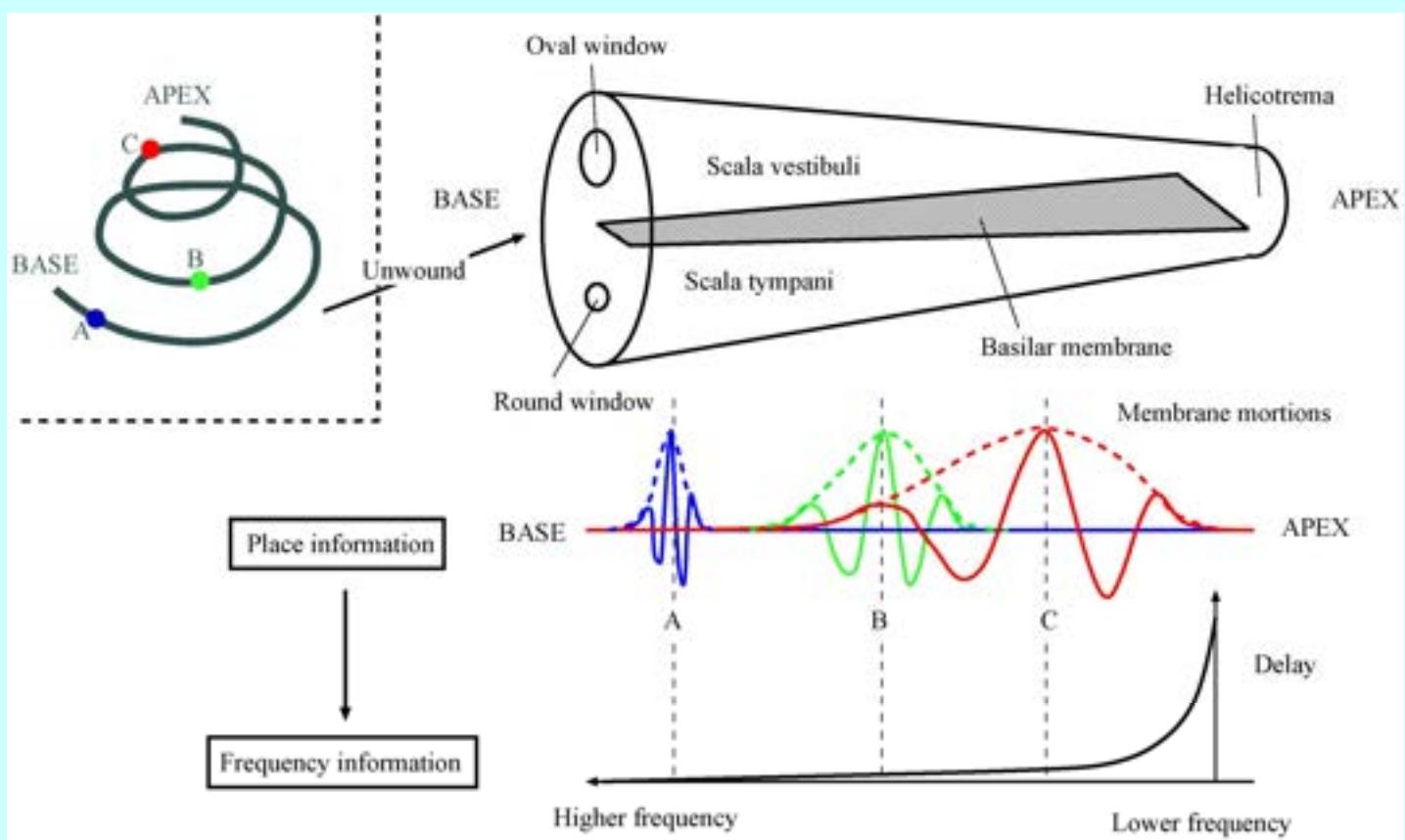
$A \sin \omega t$



圧縮と整流

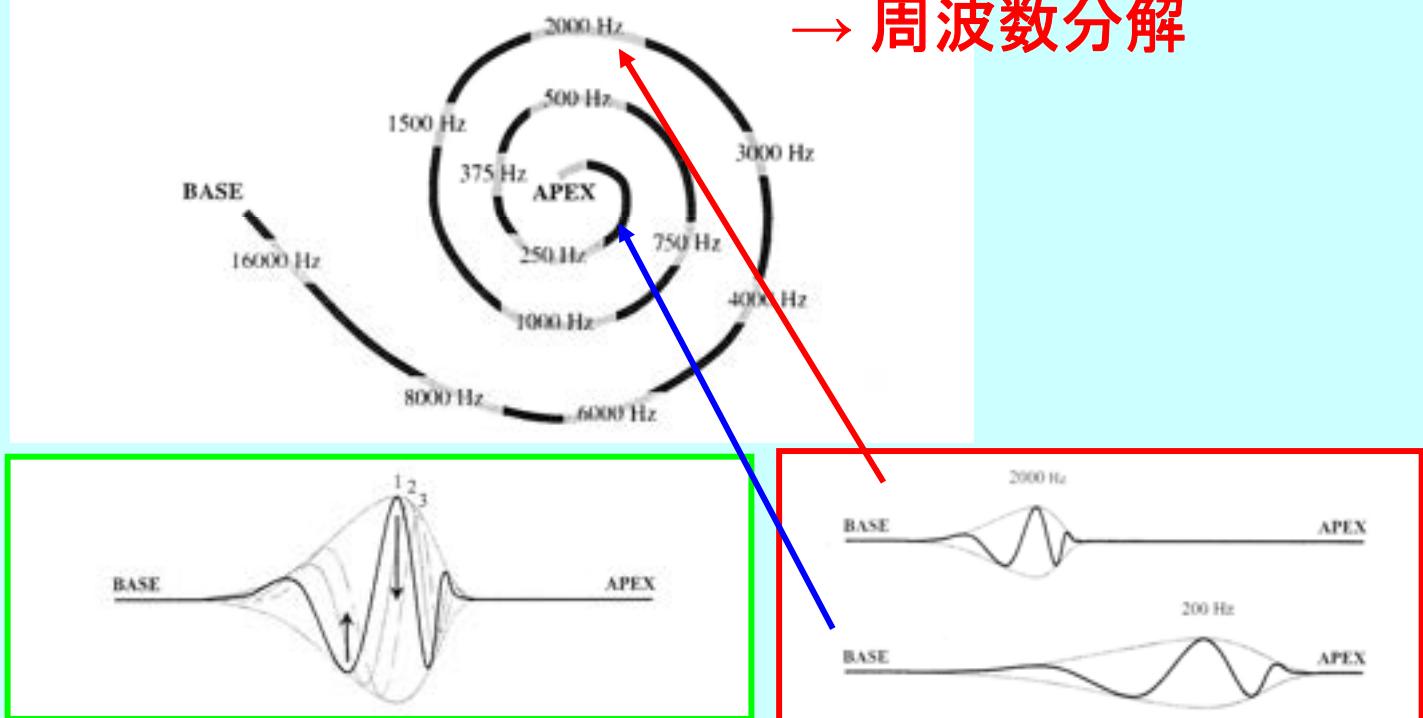


4. 基底膜で観測される周波数選択性

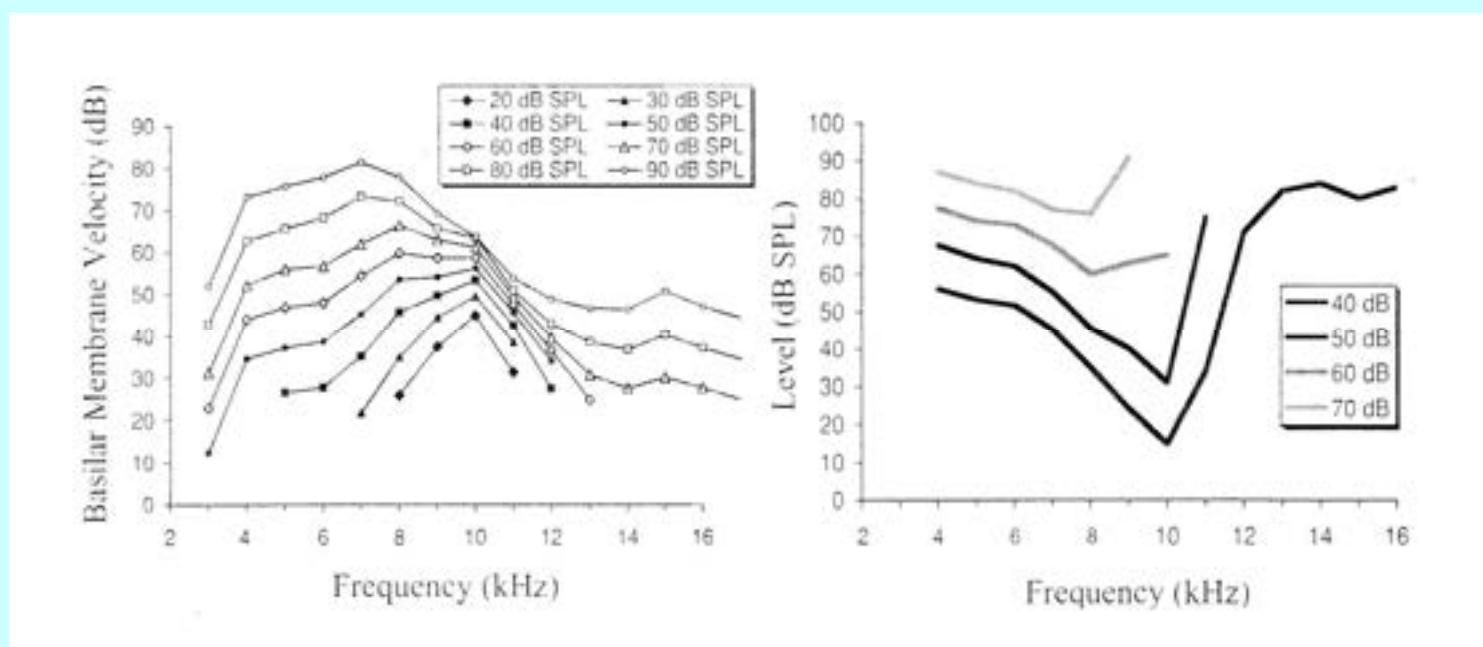


■ 基底膜振動(BMM)

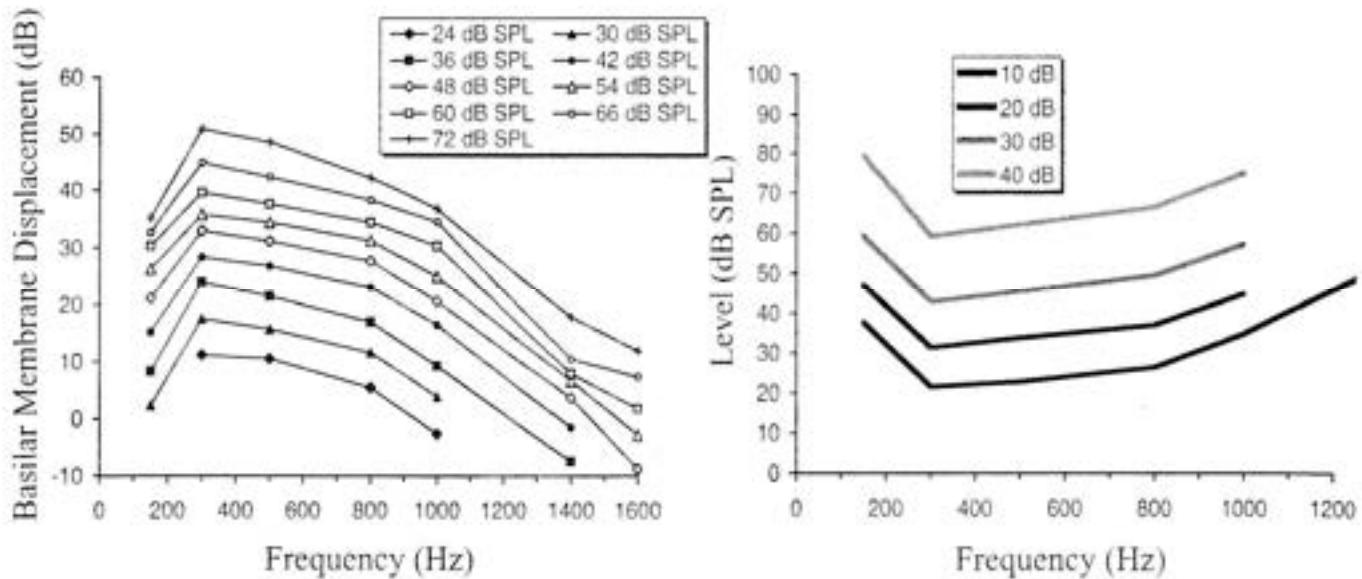
→ 周波数分解



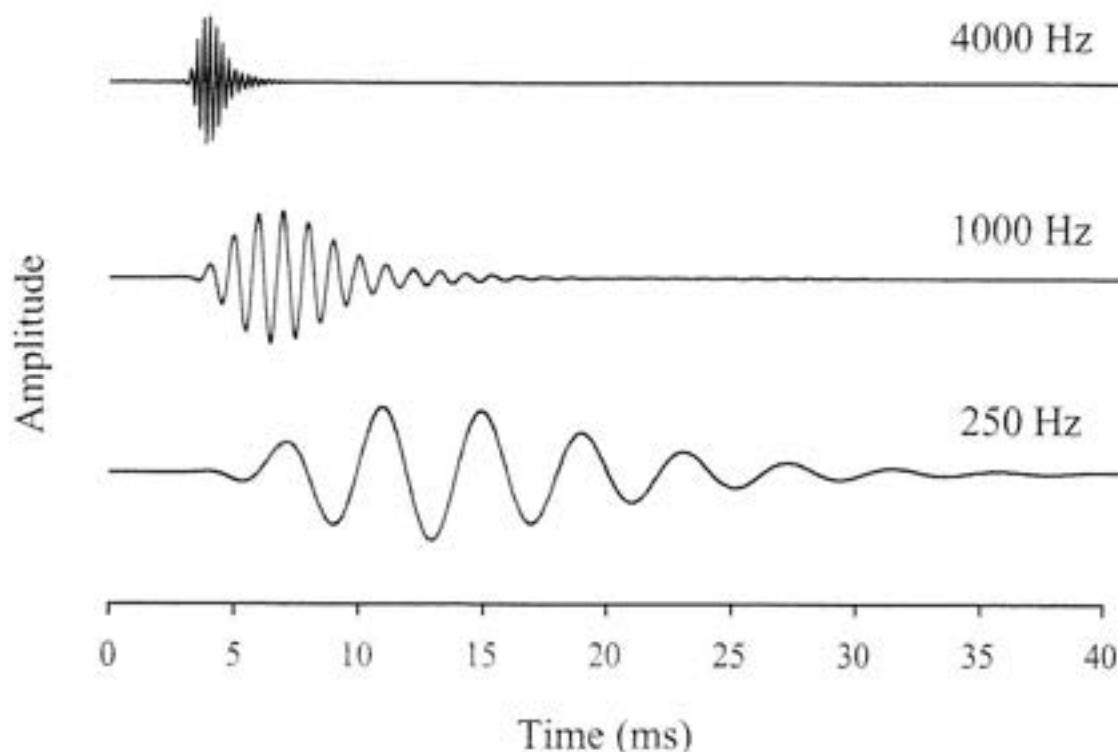
■ 蝸牛底側（高い周波数に対する応答）



■ 蝸牛頂側（低い周波数に対する応答）

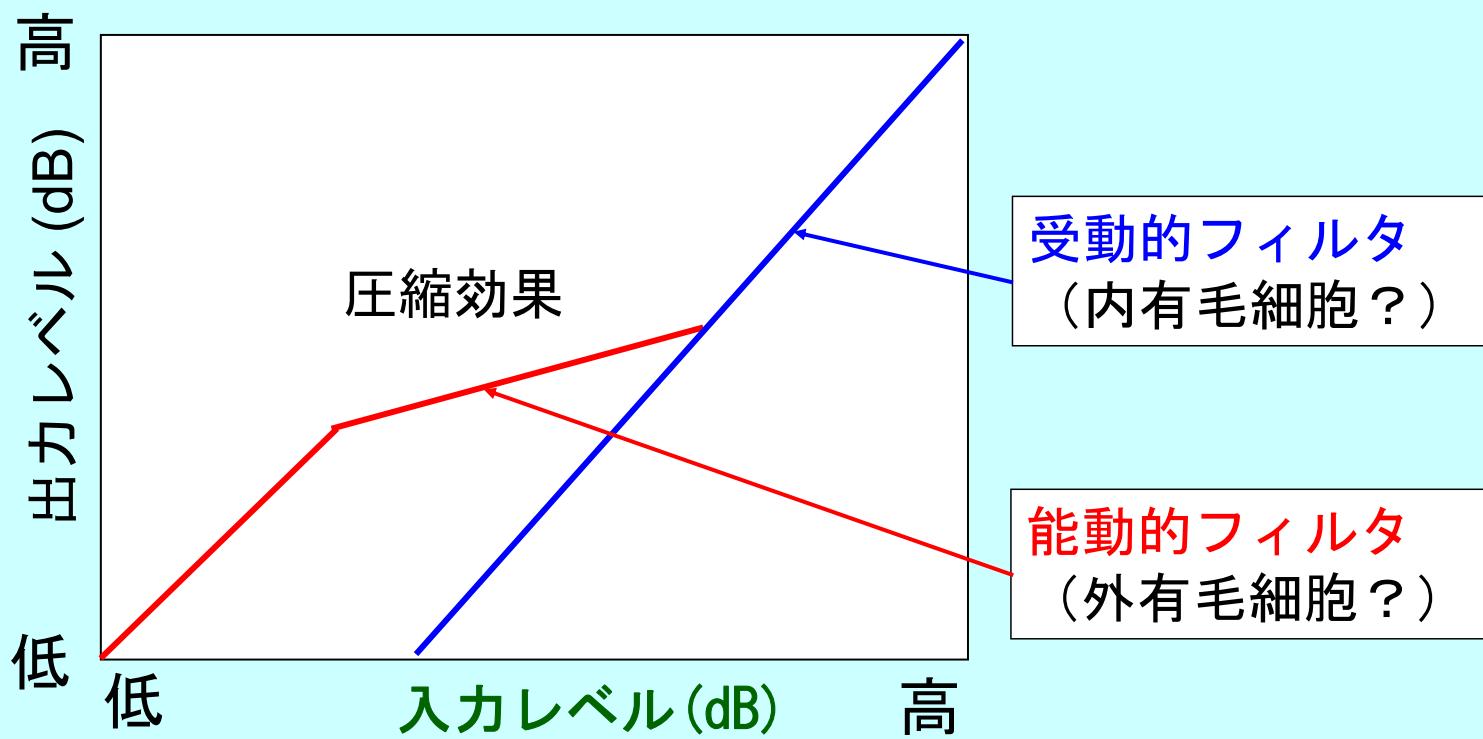


基底膜のインパルス応答関数

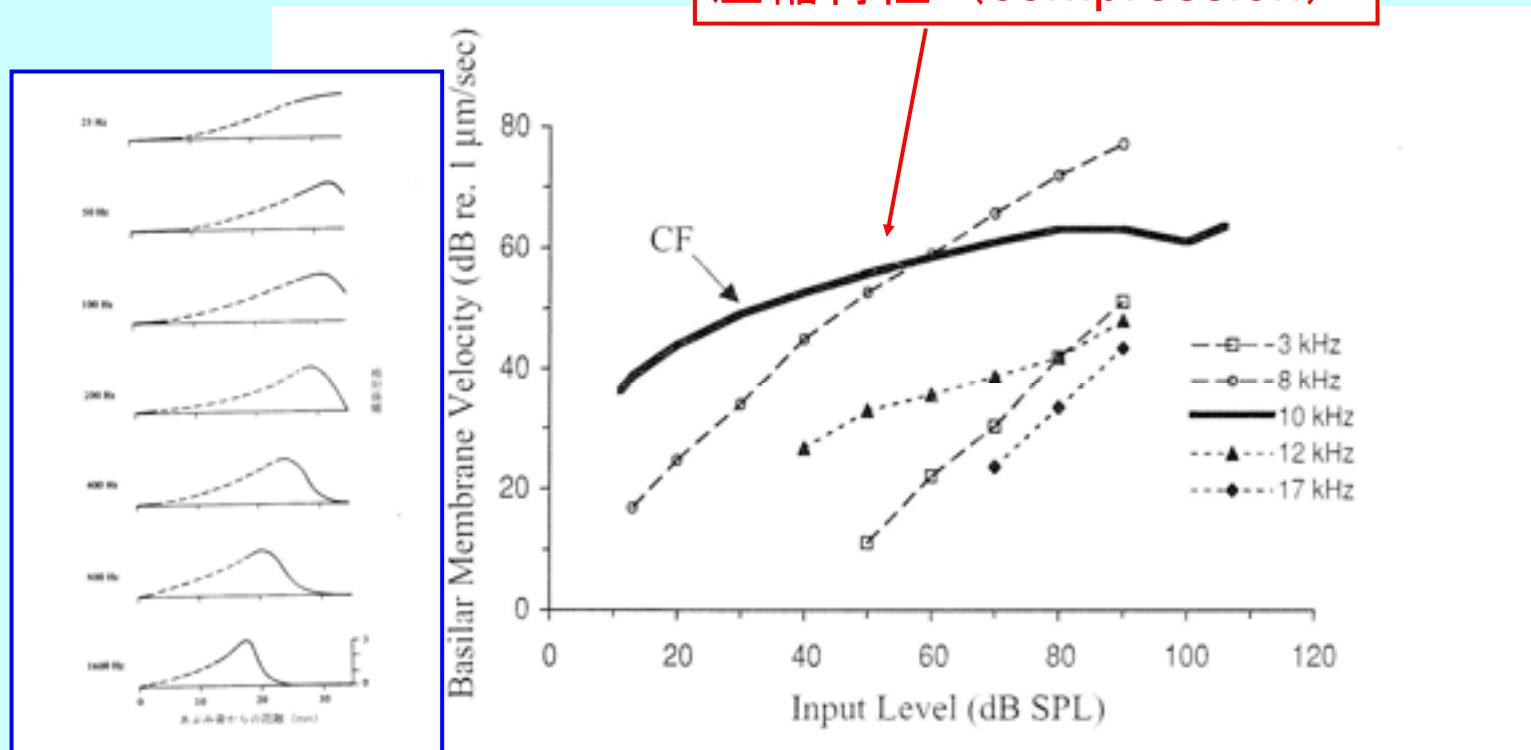


5. 入出力特性

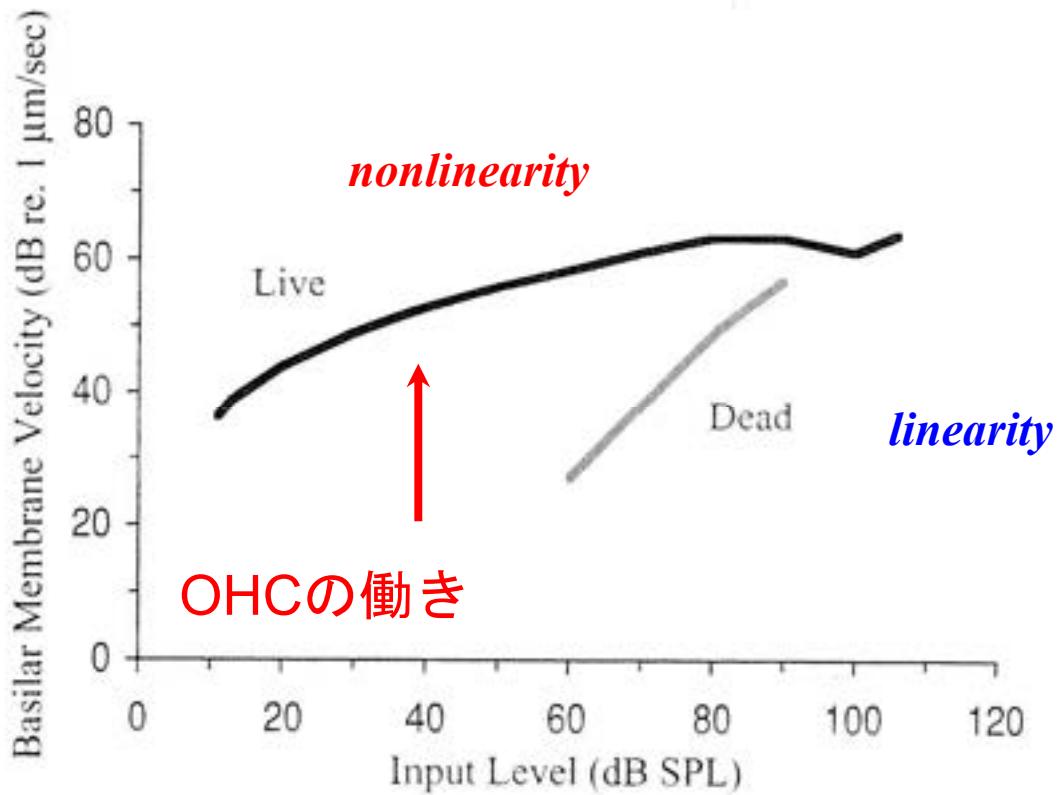
- #### ■ 圧縮特性（いわゆる自動ゲインコントロールの働き）



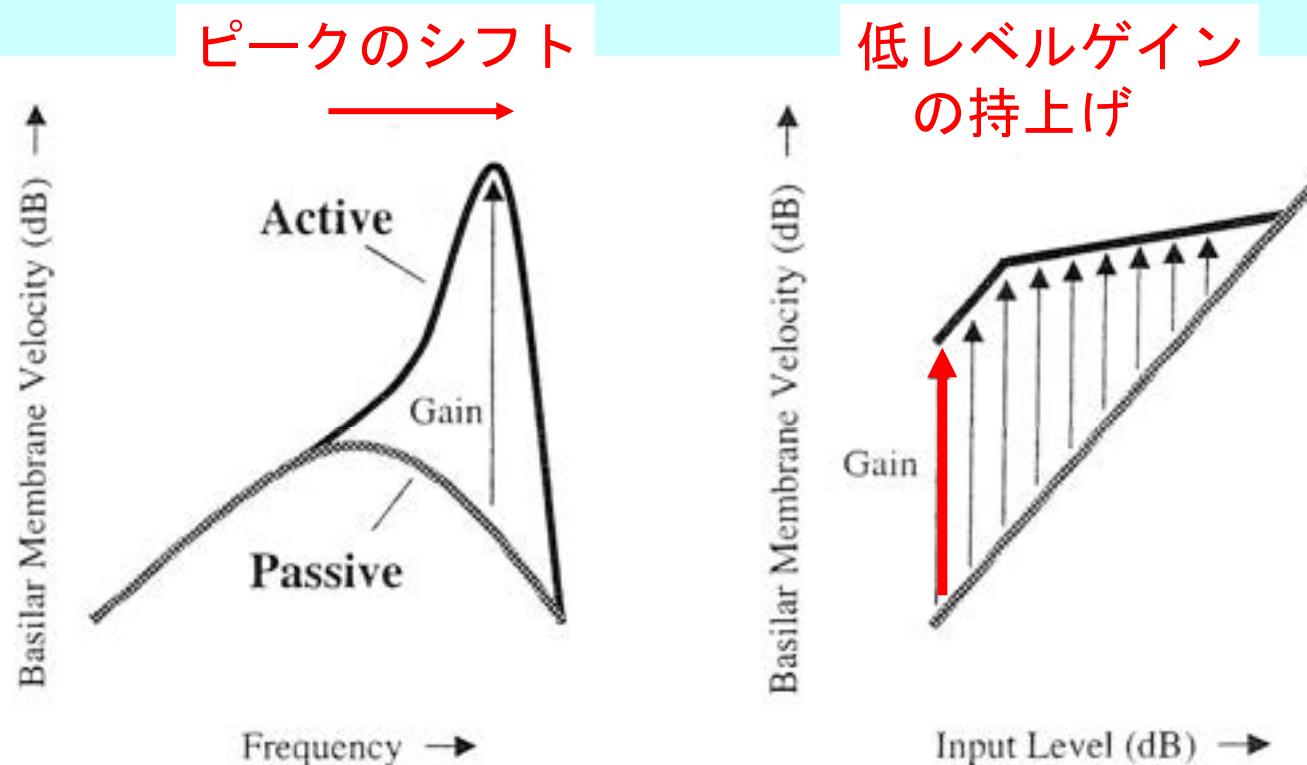
基底膜振動パターンの入出力特性



入出力特性の非線形性

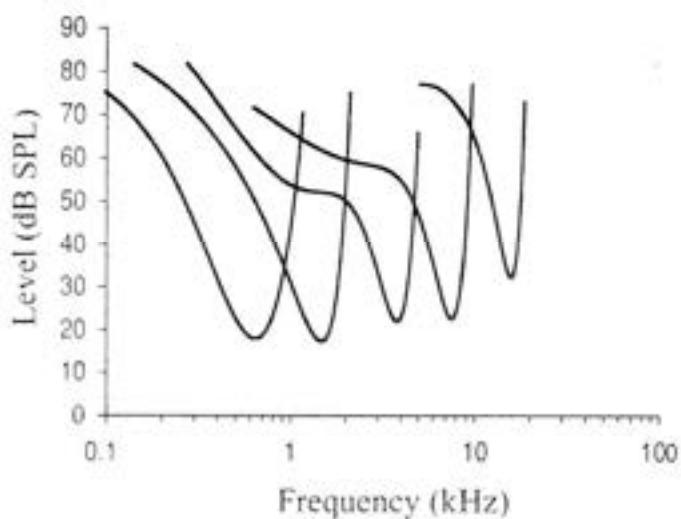
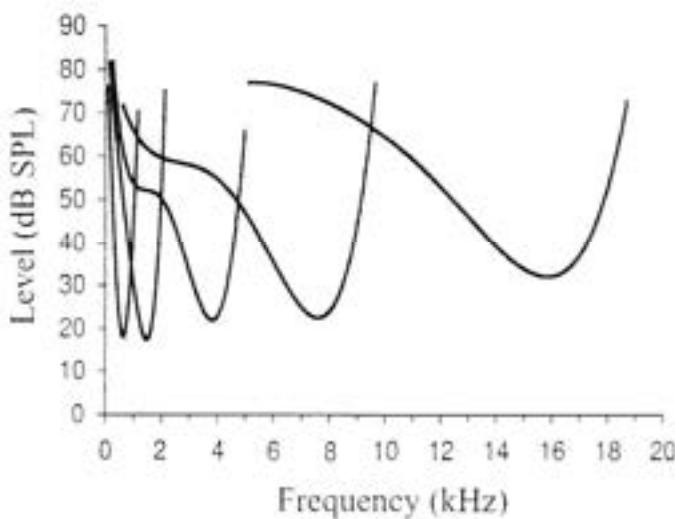


OHCの寄与

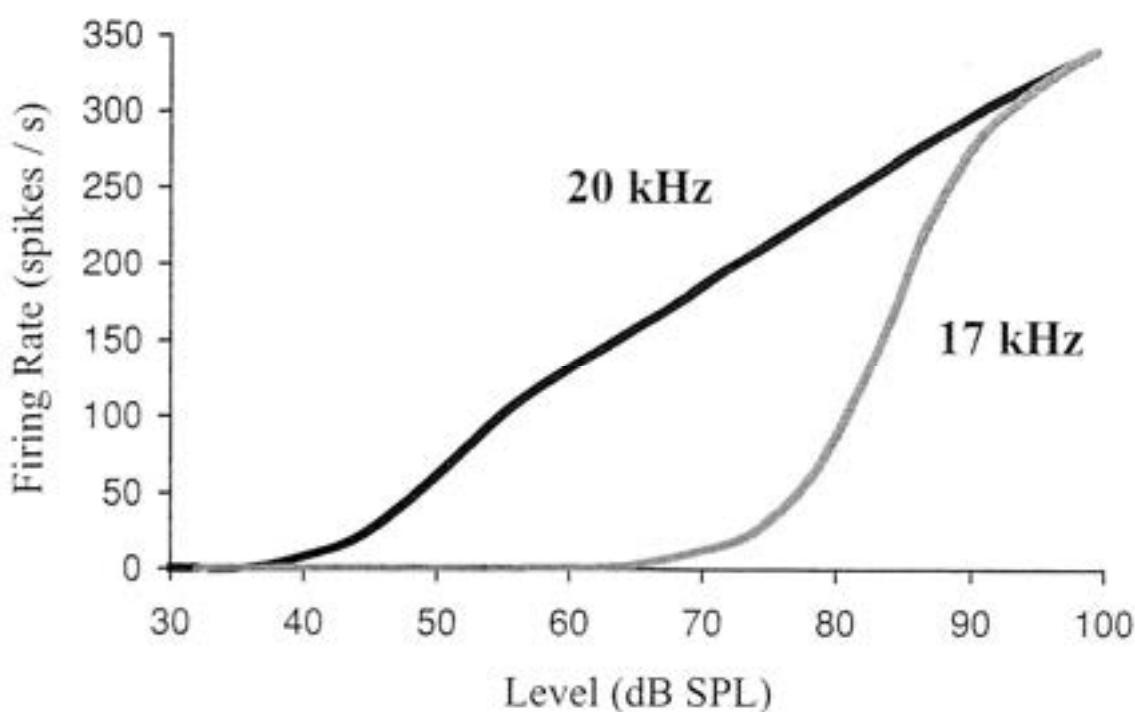


6. 聴神経で観測される周波数選択性

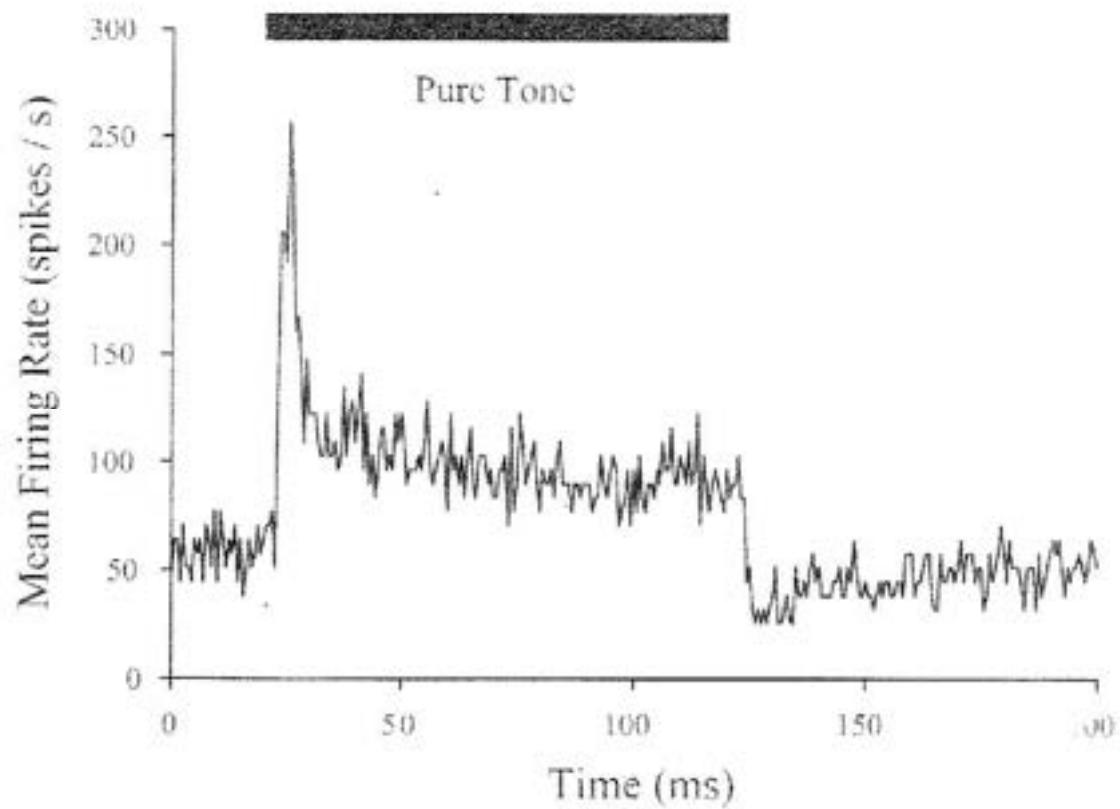
神経発火で見られる同調曲線



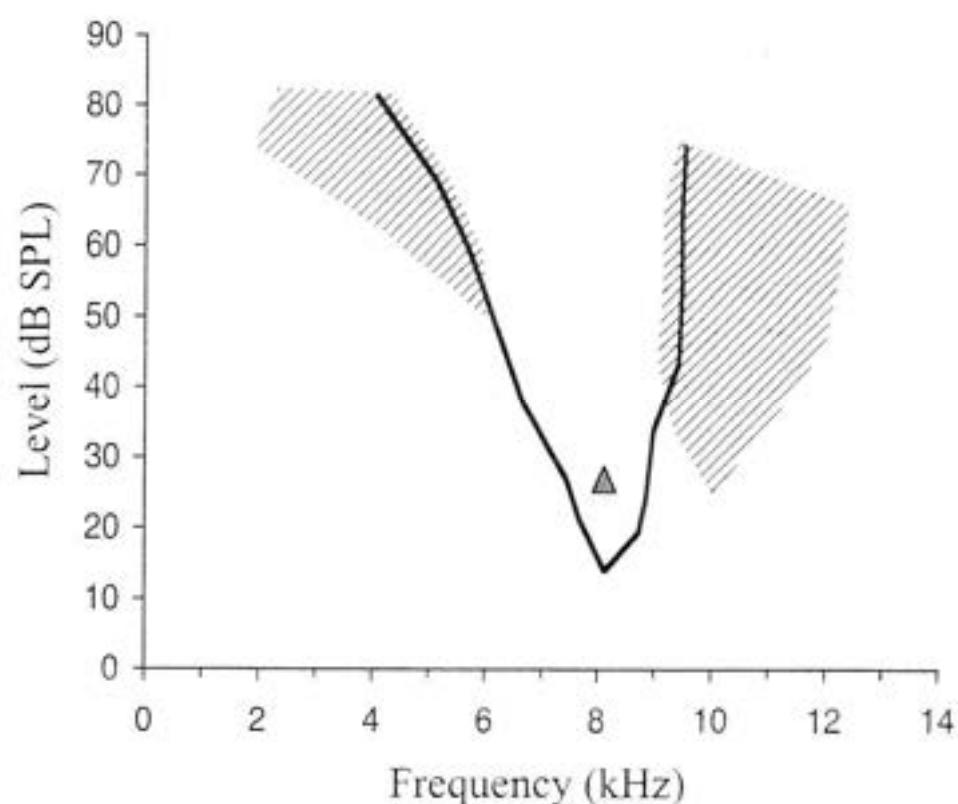
神経発火の割合



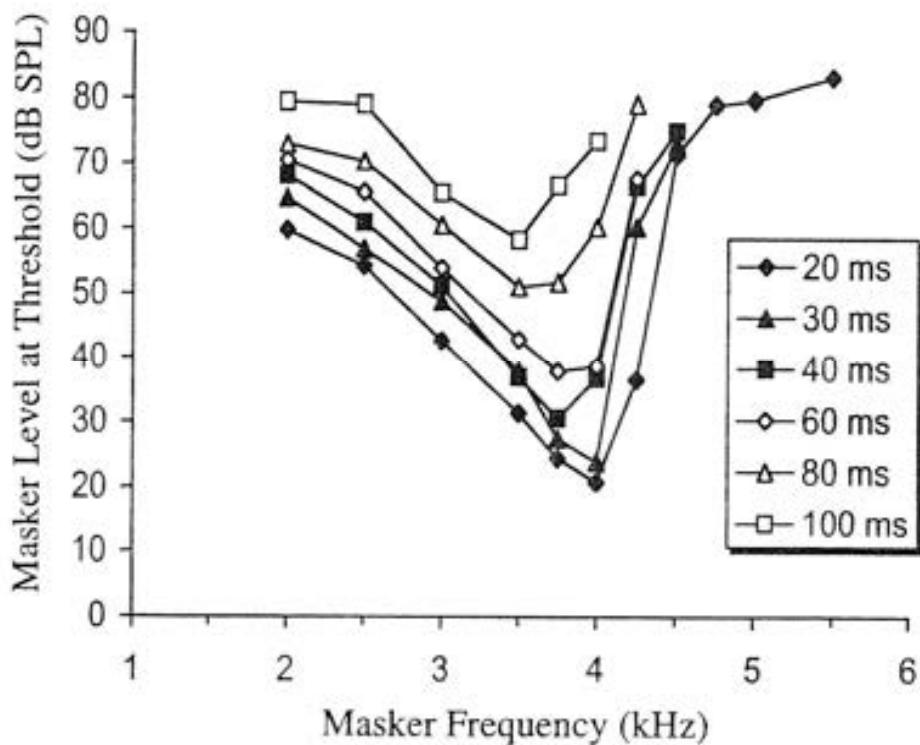
神経発火の割合



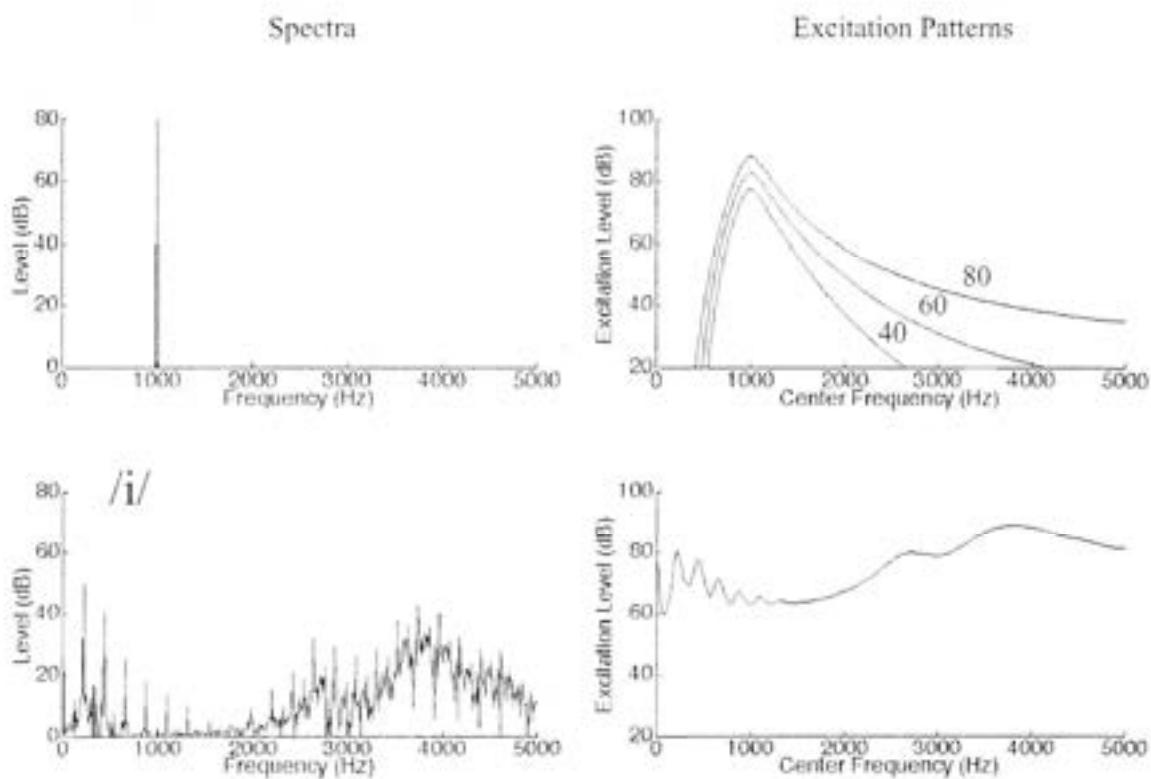
二音抑圧



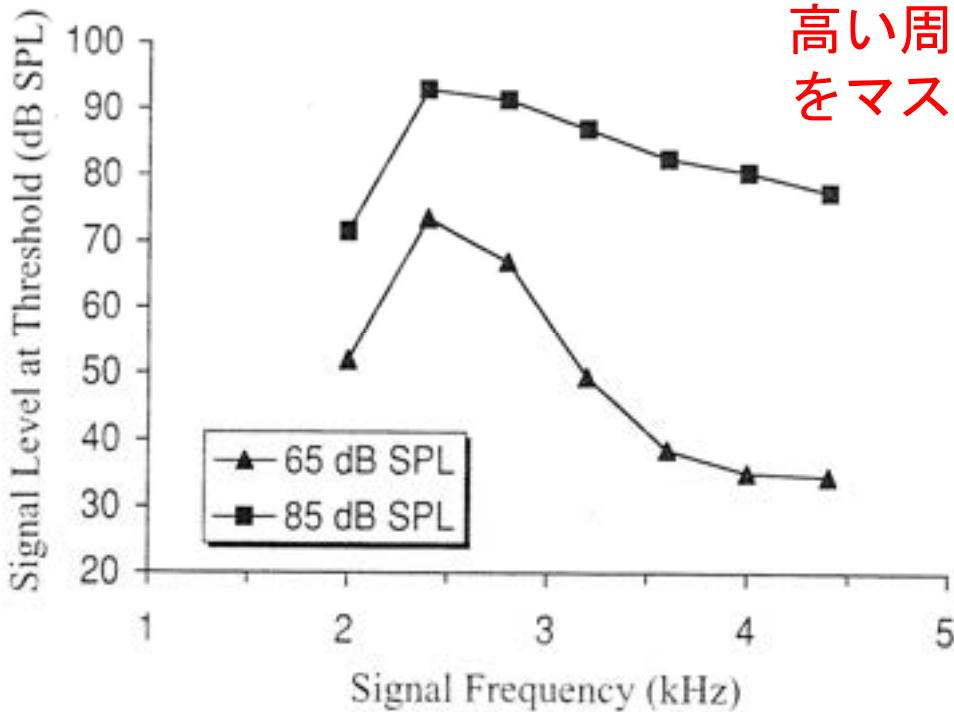
心理物理同調曲線



エキサイテーションパターン



マスキングの上方への拡がり

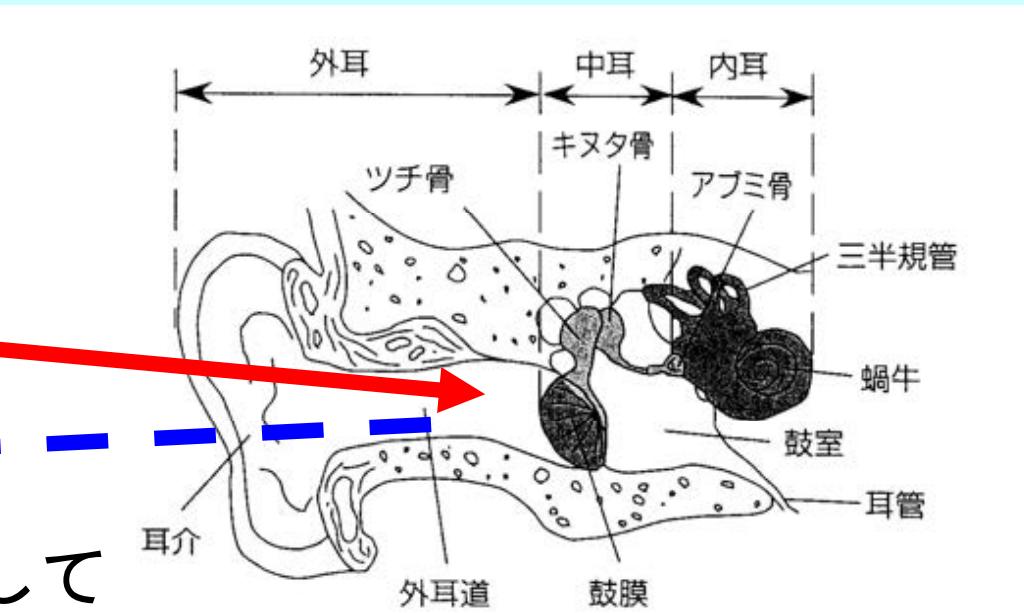


低い周波数成分は
高い周波数の信号
をマスクしやすい

7. 耳音響反射

Oto-acoustic emission

音を入力
非線形歪として
反射音を観測



I485F 知覚情報処理特論

7. 聴覚中枢系の生理とモデル

2023年5月7日(日)

人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

Email: unoki@jaist.ac.jp



0. From Ear to Brain

Peripheral to Auditory cortex (Afferent)

Auditory cortex (AC)

Medial Geniculate (MG)

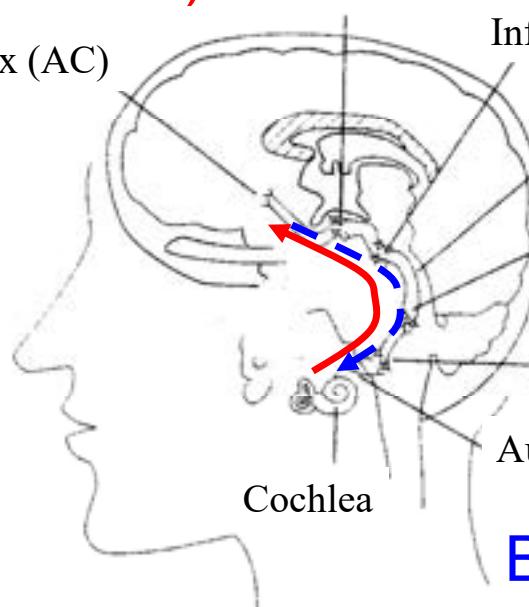
Inferior Colliculus (IC)

Lateral Lemniscus (LL)

Superior Olive (SO)

Cochlear Nucleus (CN)

Auditory nerve (AN)



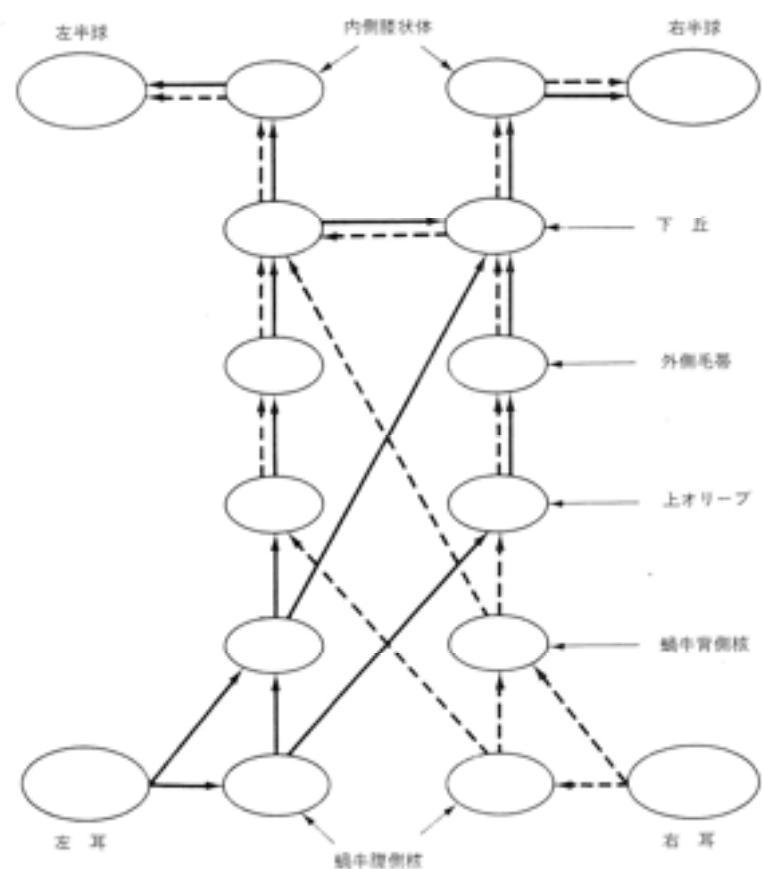
Brain to Peripheral (Efferent)

様々な種の可聴域

■ Rat	0.25-70 kHz
■ Mouse	0.2-70 kHz
■ Guinea pig	0.2-45 kHz
■ 猫	0.125-60 kHz
■ 人間	<i>0.02-20 kHz</i>

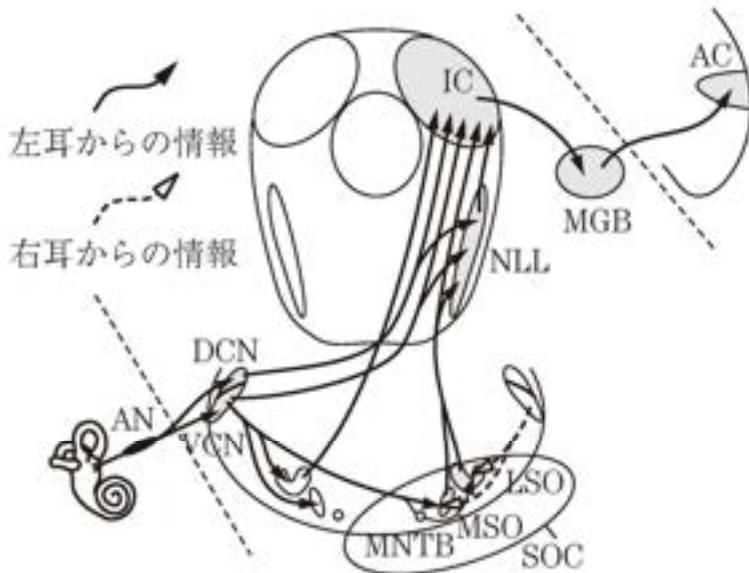
1. 聴覚中枢系

伝達経路



脳：脳幹，小脳，大脳

神経核 (CN)



AN : 聽神經

VCN : 複側蝸牛神經核

DCN : 背側蝸牛神經核

MNTB : 台形体内側核

MSO : 上オリーブ内側核

LSO : 上オリーブ外側核

SOC : 上オリーブ複合体 (MNTB, MSO, LSO から構成)

NLL : 外側毛帶核

IC : 下丘

MGB : 内側膝状体

AC : 聴覚皮質

聴覚情報の伝播経路

2. 蝸牛神経核の応答とモデル

■ 蝸牛神経核 (cochlear nucleus, CN)

■ 腹側蝸牛神経核 (ventral CN, VCN)

■ 背側蝸牛神経核 (dorsal CN, DCN)

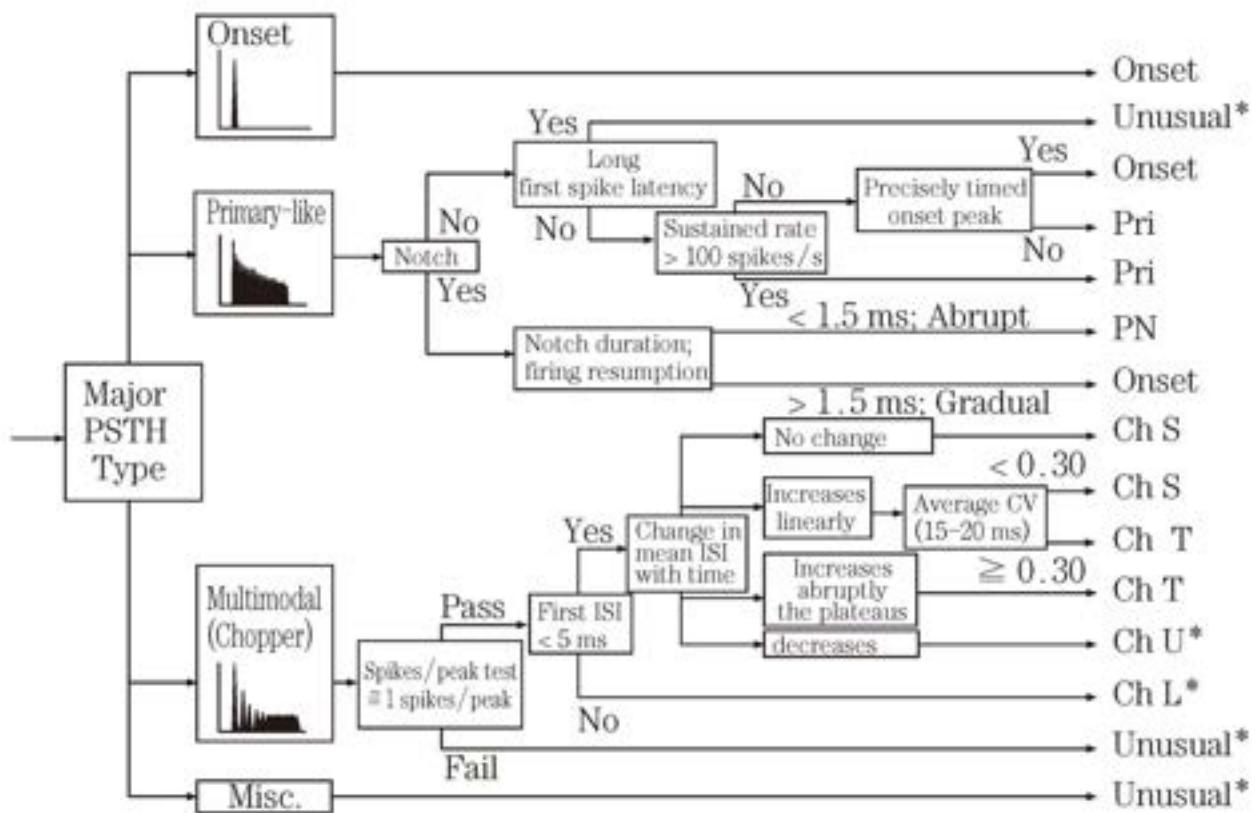
■ 時間応答パターン

■ 音に対するニューロンの反応を時間経過とともに観察したときに現れるパターンのこと

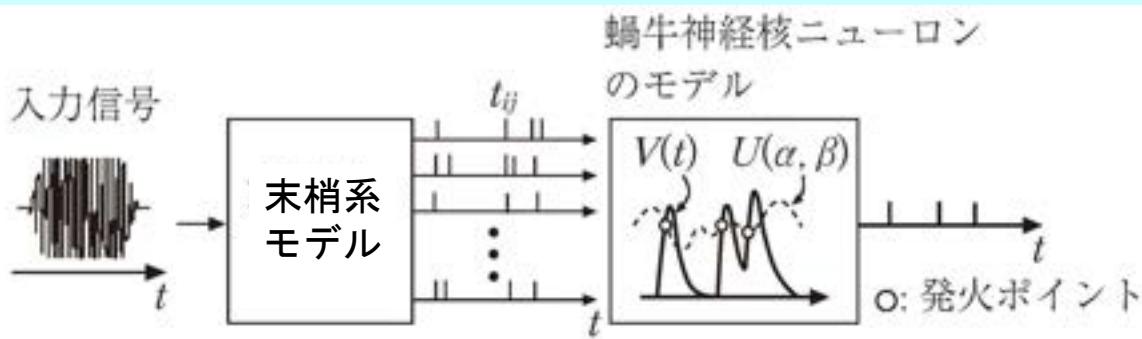
■ 周波数応答パターン

■ 様々な周波数の音を呈示し、ニューロンの反応強度（発火頻度）を周波数の関数として表わしたときに現れるパターンのこと

腹側核の時間応答パターン



■ 単一蝸牛神経核ニューロンモデル

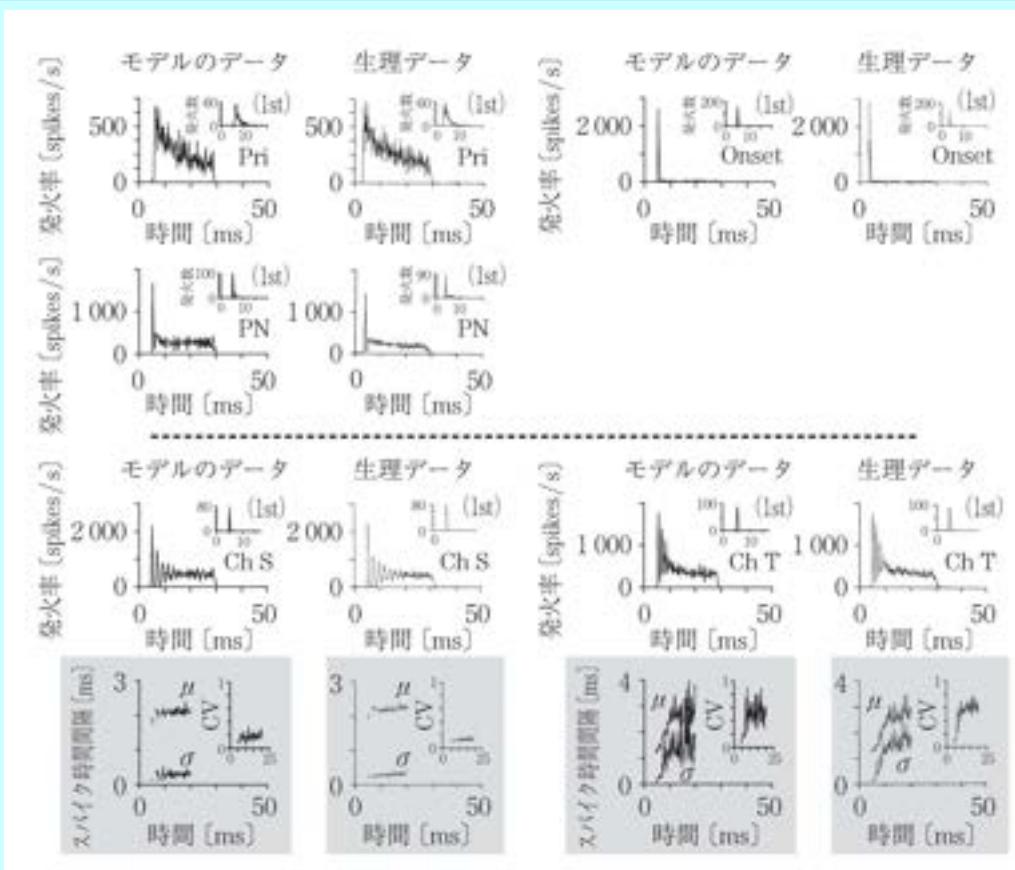


$V(t)$: シナプス後電位

$U(\alpha, \beta)$: 発火の閾値を一様乱数でモデル化

t_r : 不応期

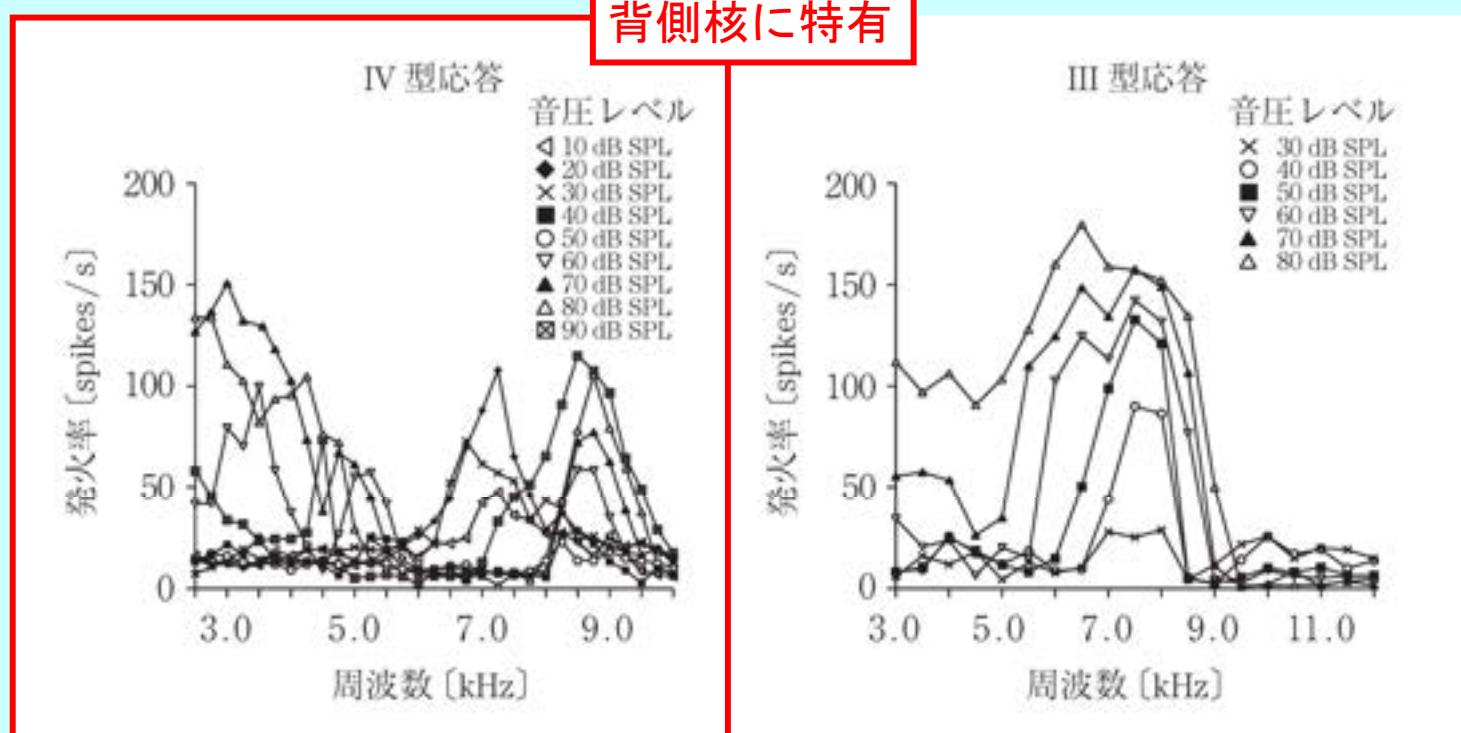
Blackburn & Sachs (1989)の生理データ



背側核の周波数応答パターン

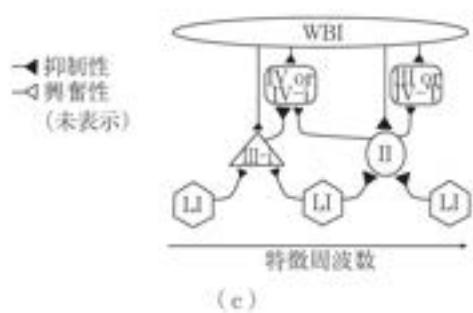
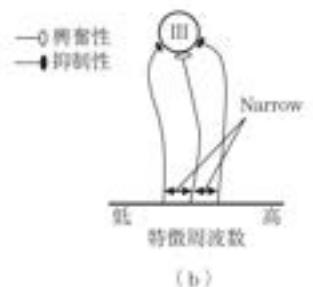
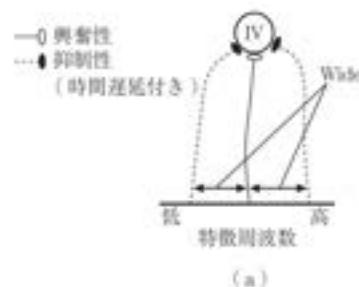
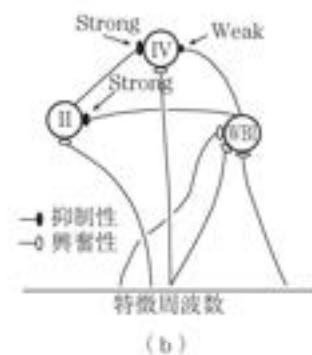
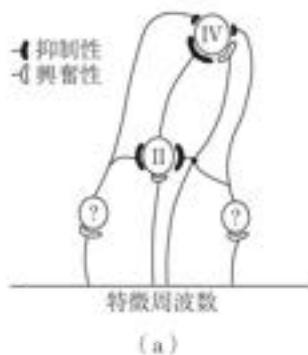
I型, II型：自然発火の有無

背側核に特有



概念モデル

IV型応答を説明する概念モデル

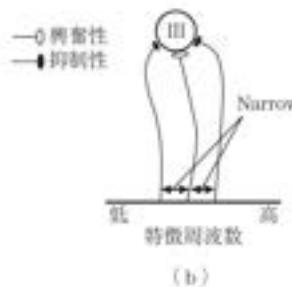
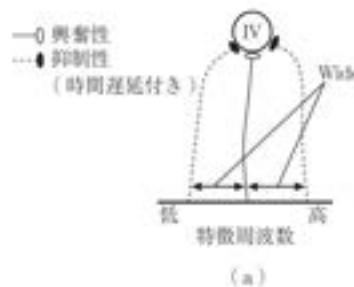
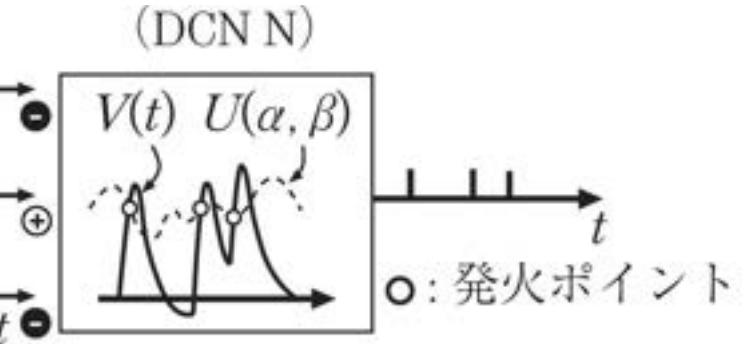


IV型応答を説明する神経回路網モデル

単一背側蝸牛神経核モデル

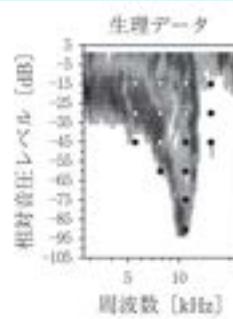
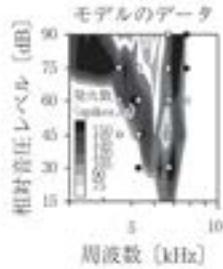
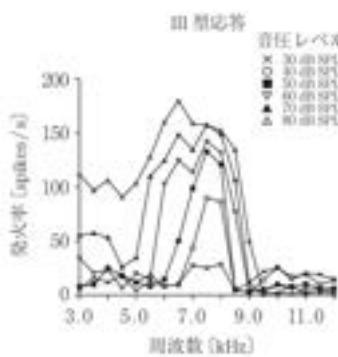
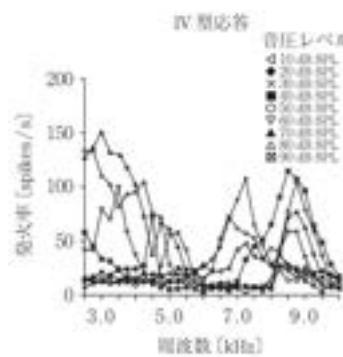


単一背側蝸牛神経核 ニューロンモデル

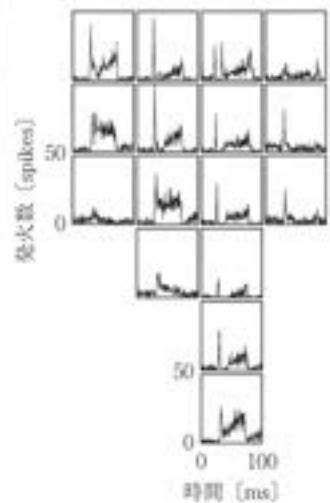
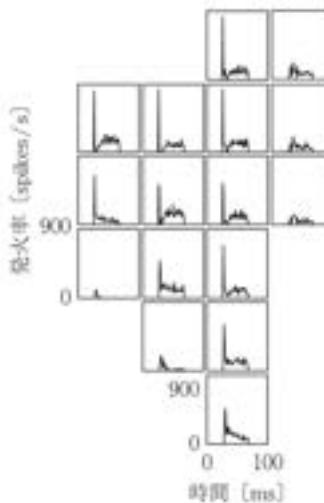
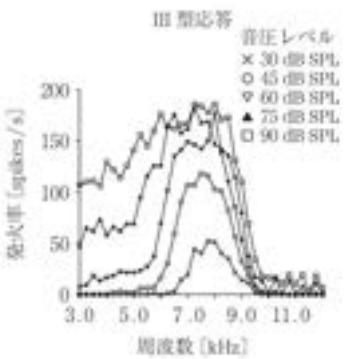
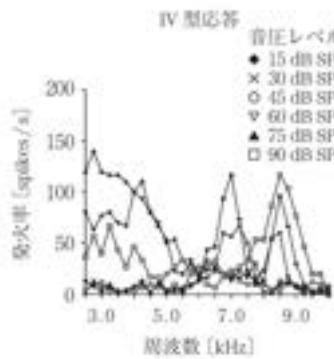


モデルにより再現されたIV型応答と時間応答パターン

生理データ



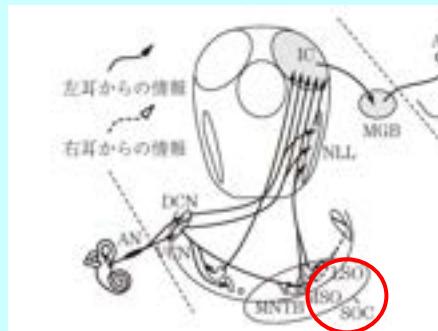
モデルのデータ



3. 上オリーブ複合体の機能

上オリーブ複合体

方向定位に利用されているといわれている



AN: 感覚神経
VCN: 視床脚牛神経核
DCN: 背側脚牛神経核
MNTB: 上部体内側核
MSO: 上オリーブ内側核
LSO: 上オリーブ外側核
SOC: 上オリーブ複合体 (MNTB, MSO, LSO から構成)
NLL: 外側毛浦核
IC: 下丘
MGB: 内側膝状体
AC: 聴覚皮質

音源定位に必要な特徴

両耳間の時間差 (interaural time difference, ITD)

両耳間の音圧差 (interaural level difference, ILD)

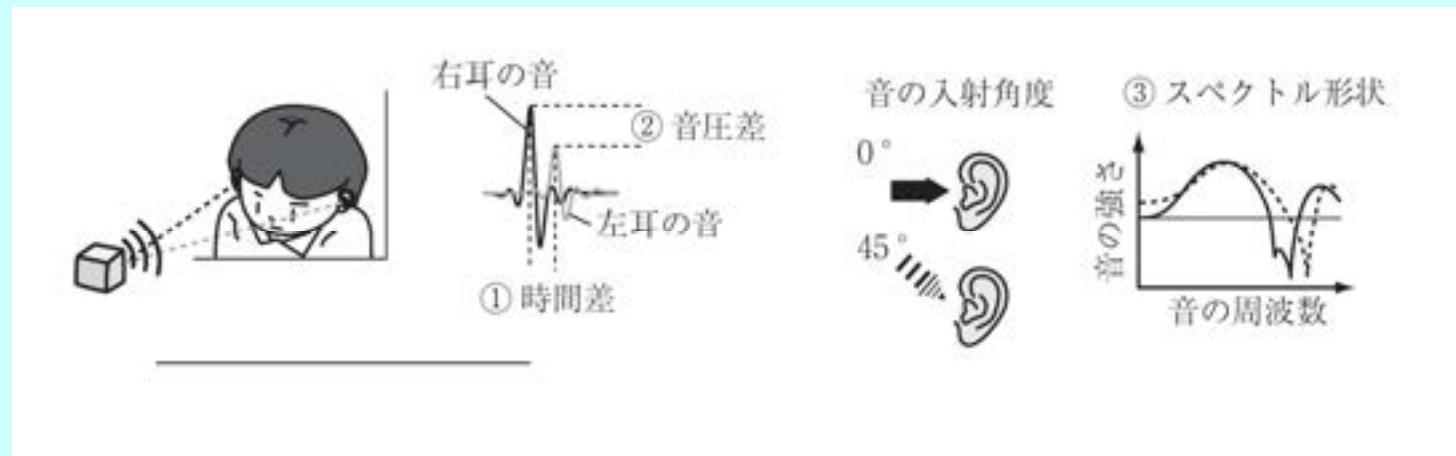
耳介により生じる鼓膜面でのスペクトル形状の違い

■ 両耳間時間差 (ITD)

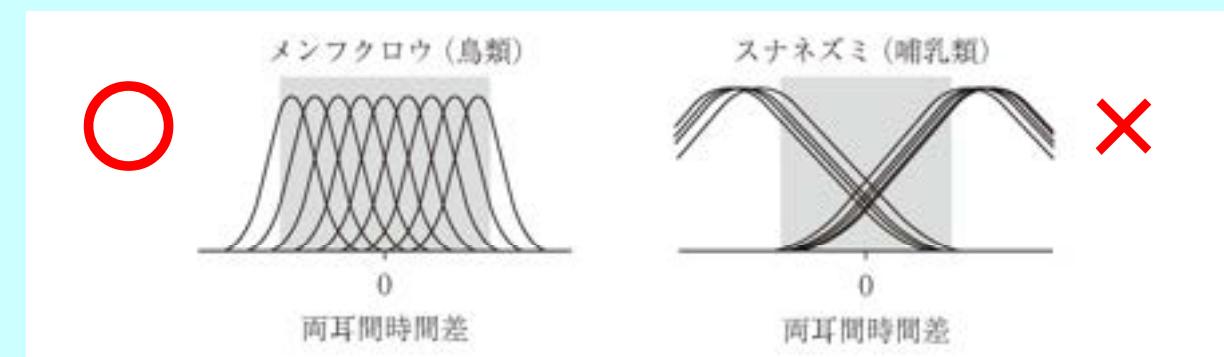
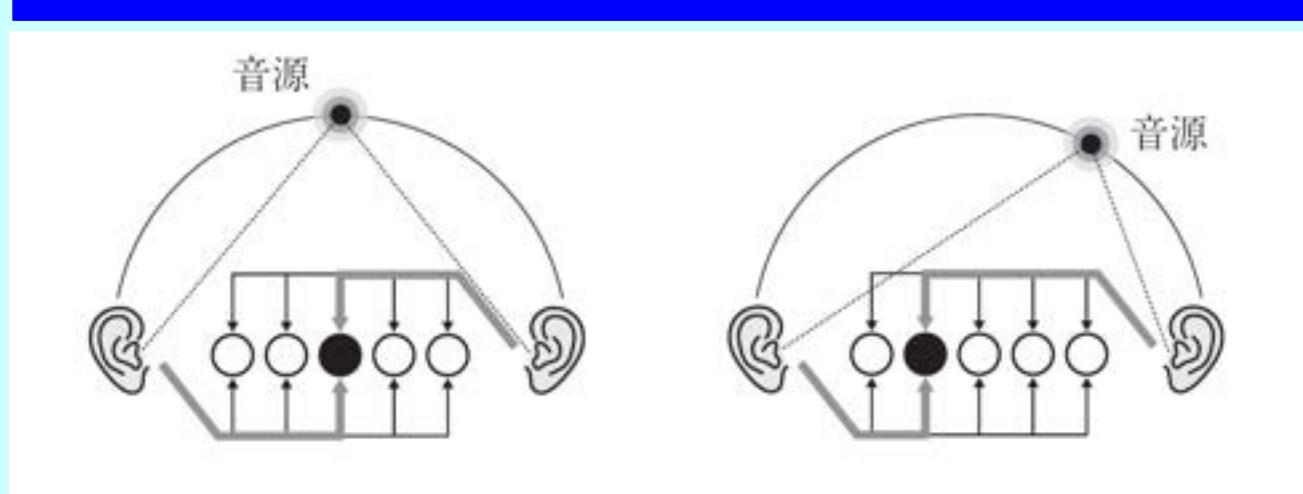
上オリーブ内側核 (medial superior olive, MSO)

■ 両耳間音圧差 (ILD)

上オリーブ外側核 (lateral superior olive, LSO)



Jeffressモデル



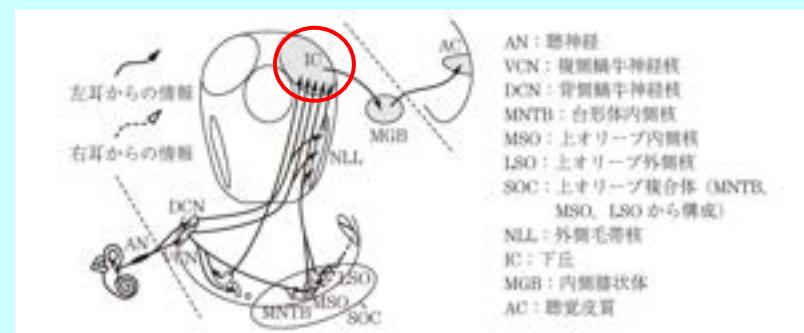
4. 下丘の複雑な時間応答パターン

- 下丘での時間応答パターンは蝸牛神経核でみられたものよりもさらに複雑化
- 蝸牛神経核での表現：5種類

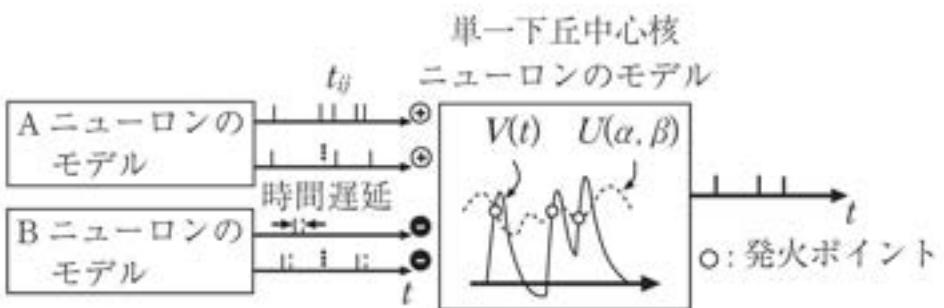
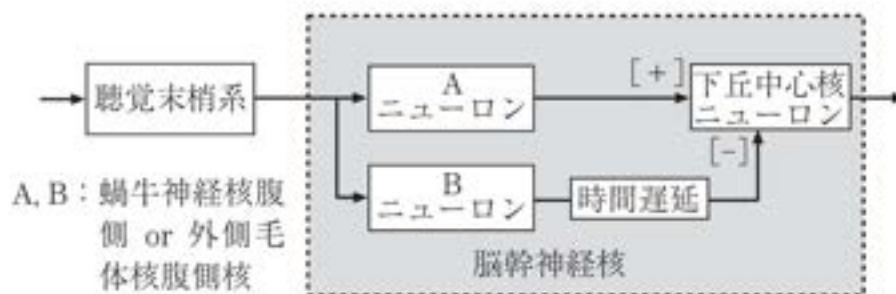


- 下丘中心核での表現：10種類

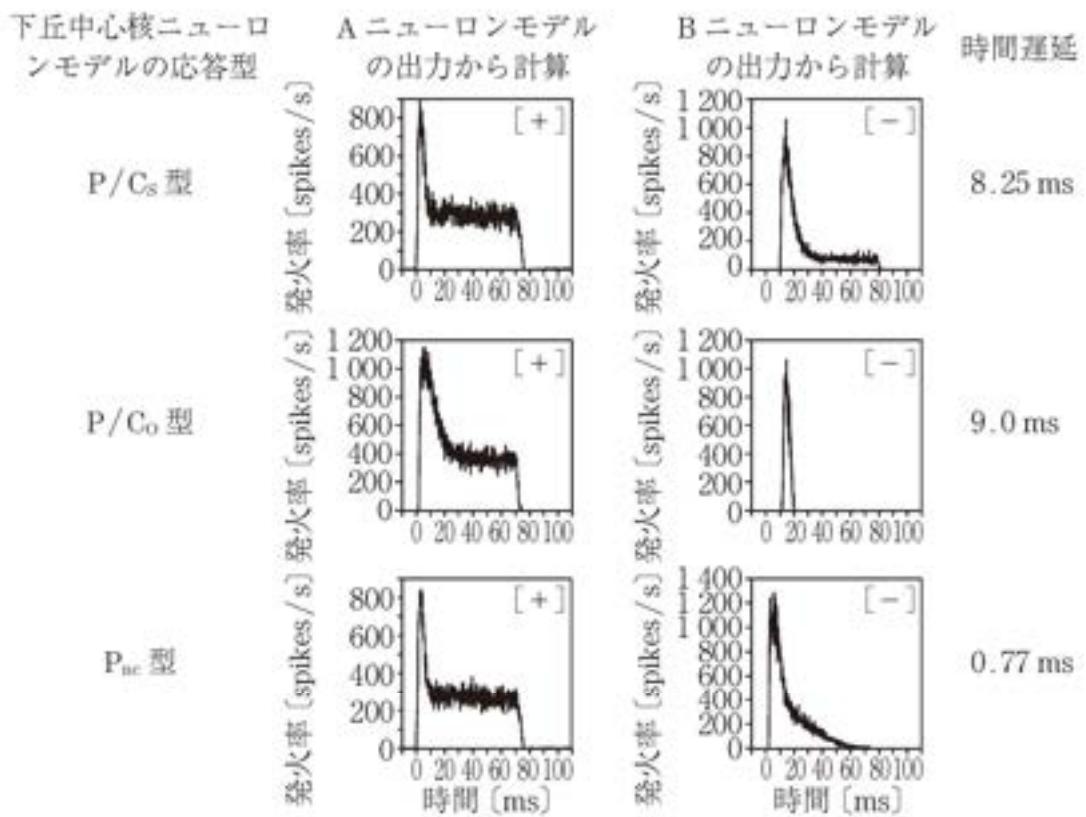
- チョッパー型
- サステインド型
- ポーザー型
- オンセット型
- ····



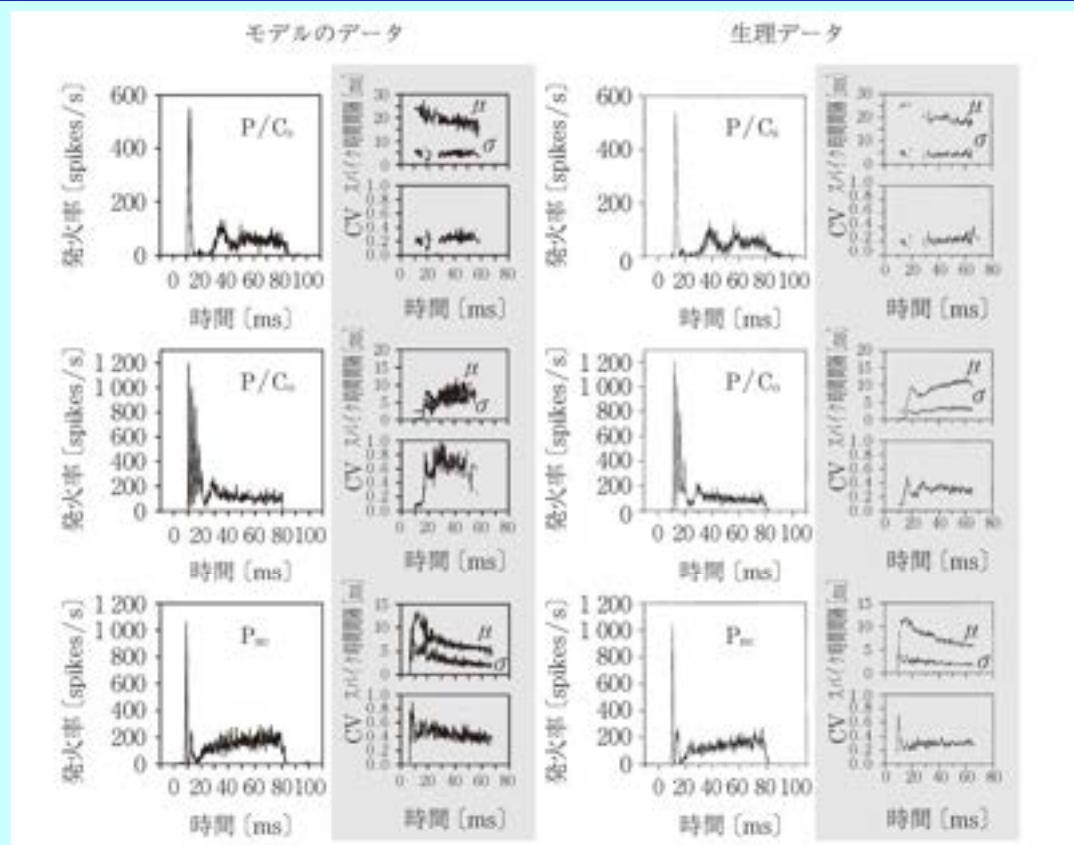
下丘中心核ニューロンモデル



下丘中心核ニューロンモデルへの入力情報の例



モデルによるデータの説明



I485F 知覚情報処理特論

8. マスキング

2023年5月20日(土)

人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

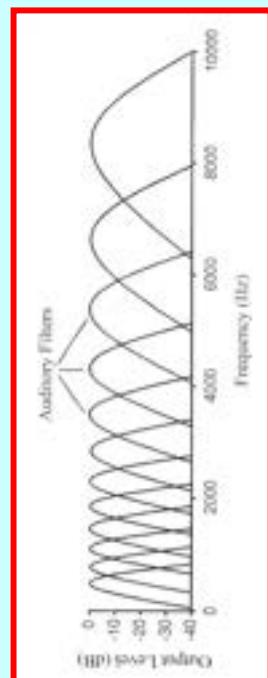
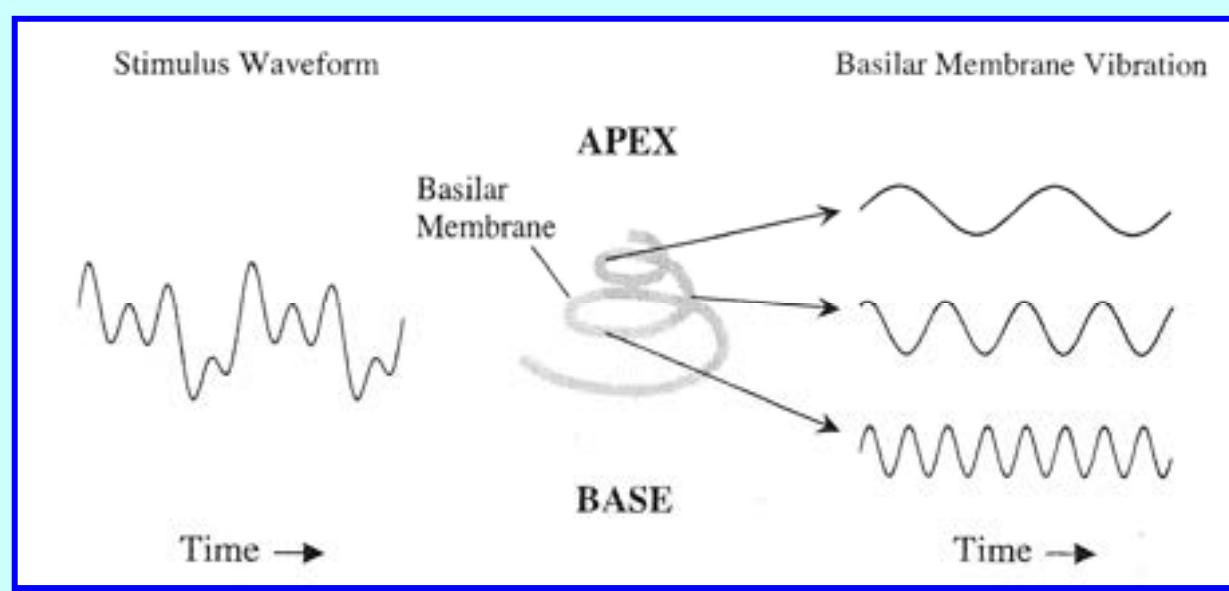
Email: unoki@jaist.ac.jp



1. 周波数選択性

■周波数選択性 (frequency selectivity)

■複合音中の正弦波成分を分解する能力

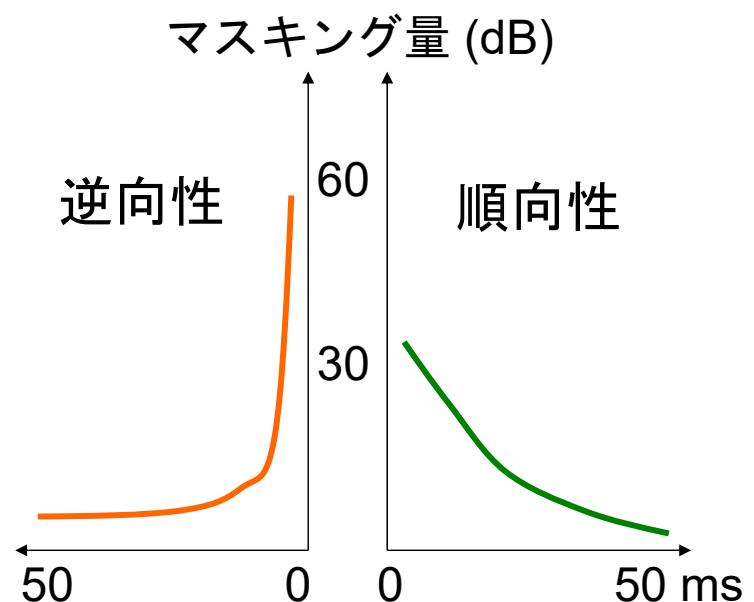
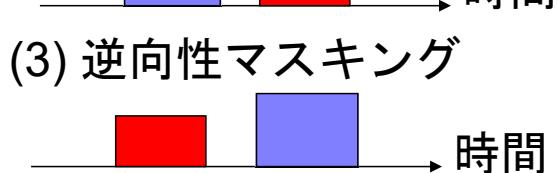
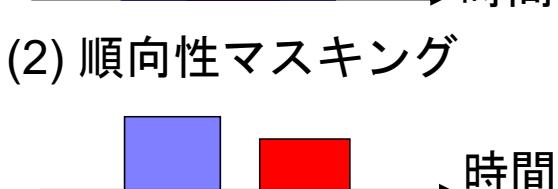
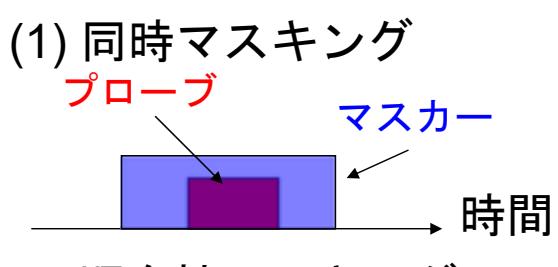


2. マスキング

■ マスキングの定義

- ある音に対する最小可聴値（threshold of audibility）が、他の（マスクする）音の存在によって上昇する現象
- ある音に対する最小可聴値が、他の（マスクする）音の存在によって上昇するときの上昇量

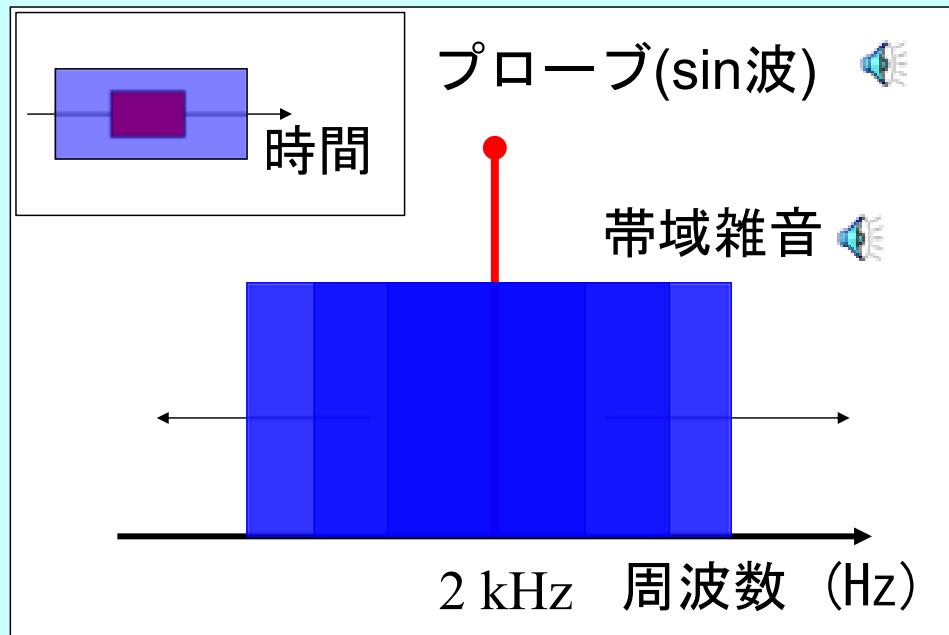
- 複合音中の正弦波信号を分解する能力を調べる手法
- マスキング：聴きたい音（プローブ）が妨害音（マスカ―）によって聴き取り難くなること



3. Fletcherの帯域幅拡張実験

■ マスキング閾値 (masking threshold)

雑音によって音が聴こる・聴こえなくなる境い目



徐々に帯域幅を広げると…

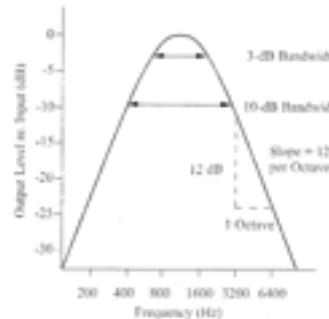


プローブの音圧レベルを上げると…

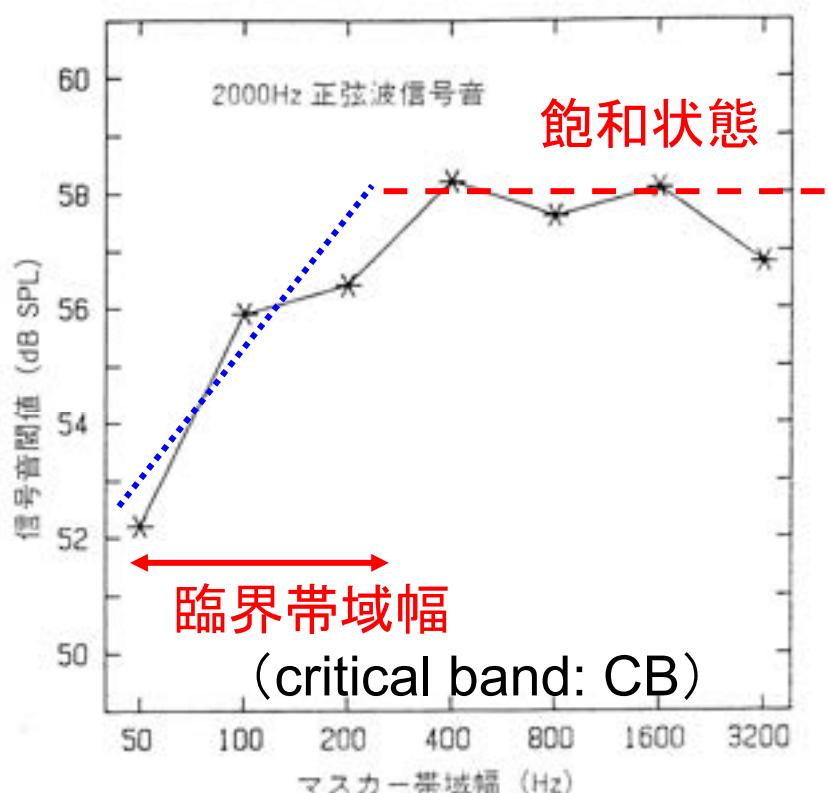


マスキング閾値特性

聴覚フィルタ
(auditory filter)



対称なフィルタ
形状を仮定



Shooneveldt & Moore (1989)

パワースペクトルモデル

■ マスキングのパワースペクトルモデル

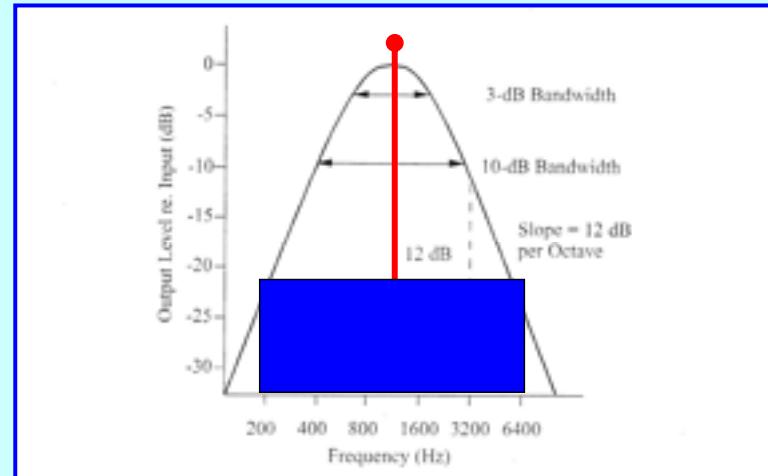
信号の閾値が、聴覚フィルタを通過する雑音の量によって決定される (Patterson & Moore, 1986)

→ フィルタを通過するSN比で決定される

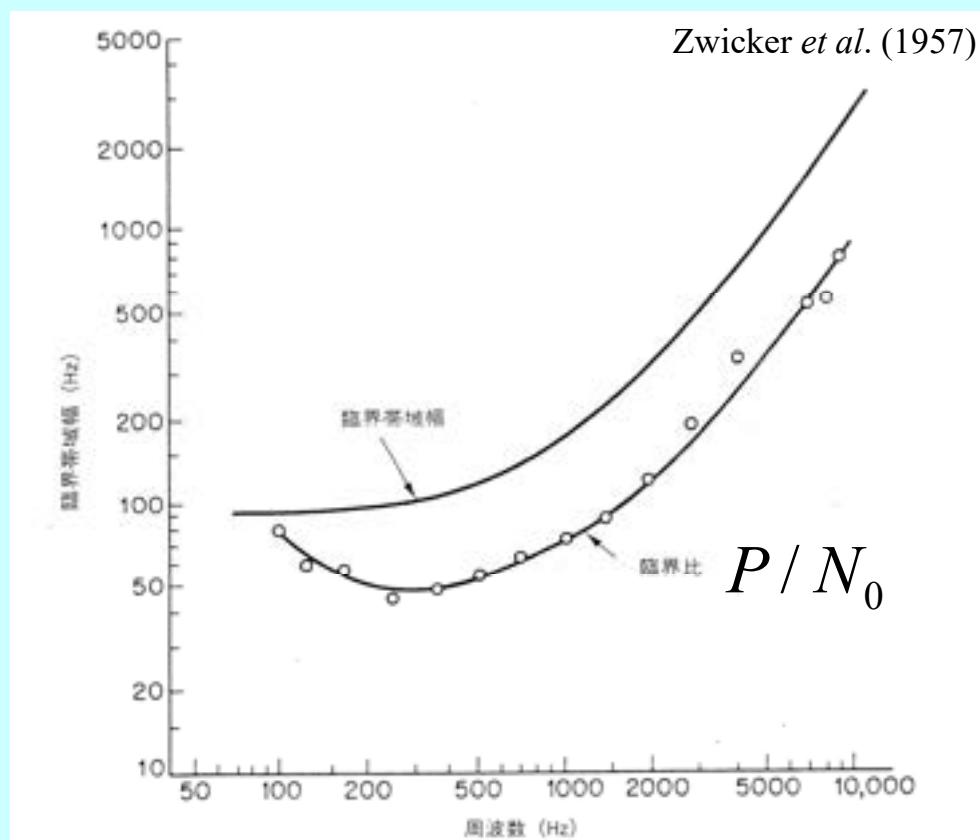
(注意) 長時間スペクトルを仮定

$$W = P / (K \times N_0)$$

聴覚フィルタ
(auditory filter)



臨界帯域幅と臨界帯域比



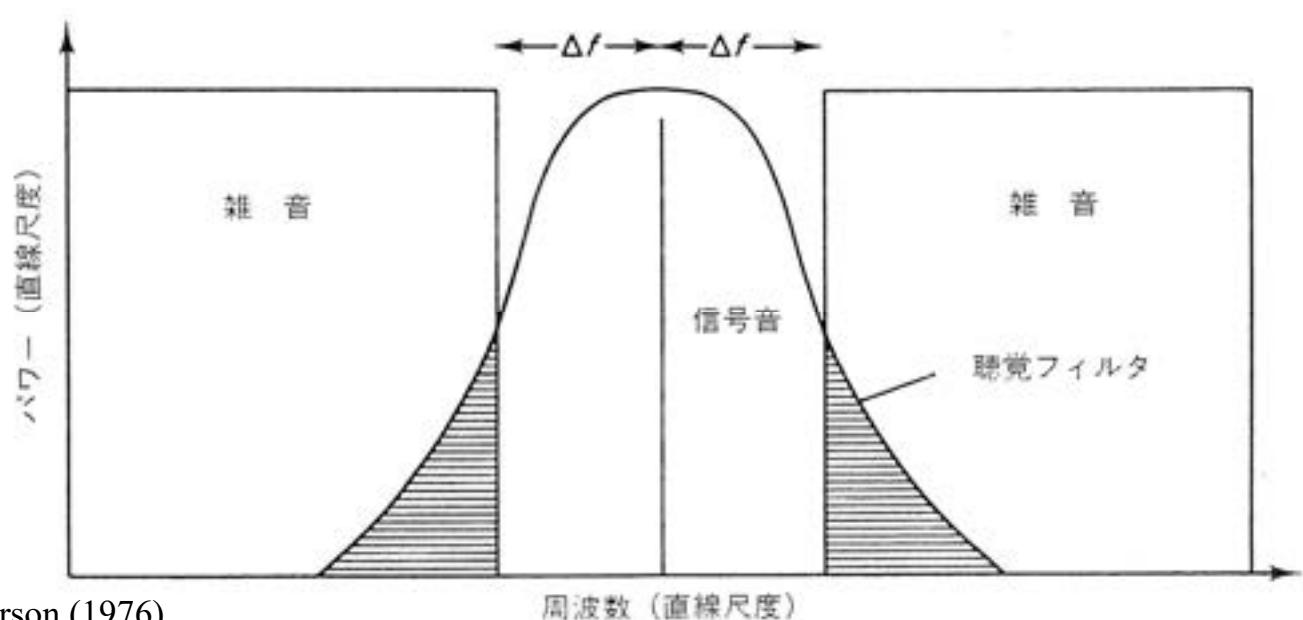
いくつかの疑問点

- そもそも聴覚フィルタ形状は対称なのか？
- 信号周波数とフィルタの中心周波数は一致しているのか？
- SN比最大の考え方は妥当だとして、もしフィルタ形状が非対称であればどうなるのか？
その場合は、信号周波数とフィルタの中心周波数も一致しないのではないか？

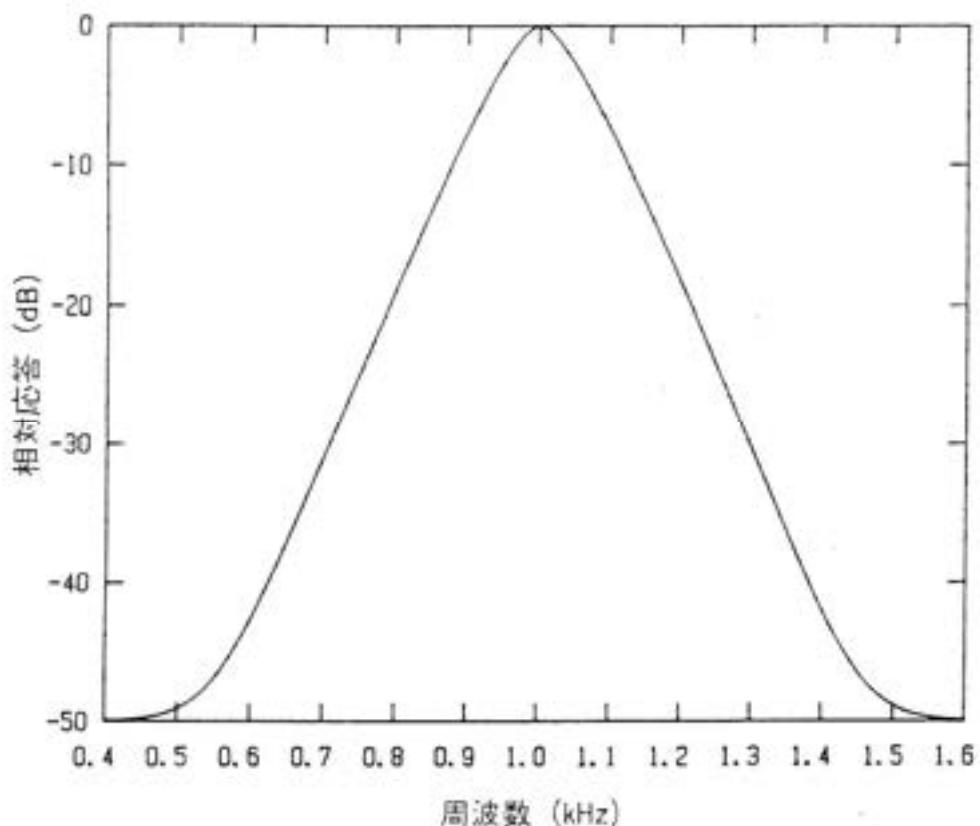
4. ノッチ雑音マスキング法

Notched-noise masking method (Patterson, 1976)

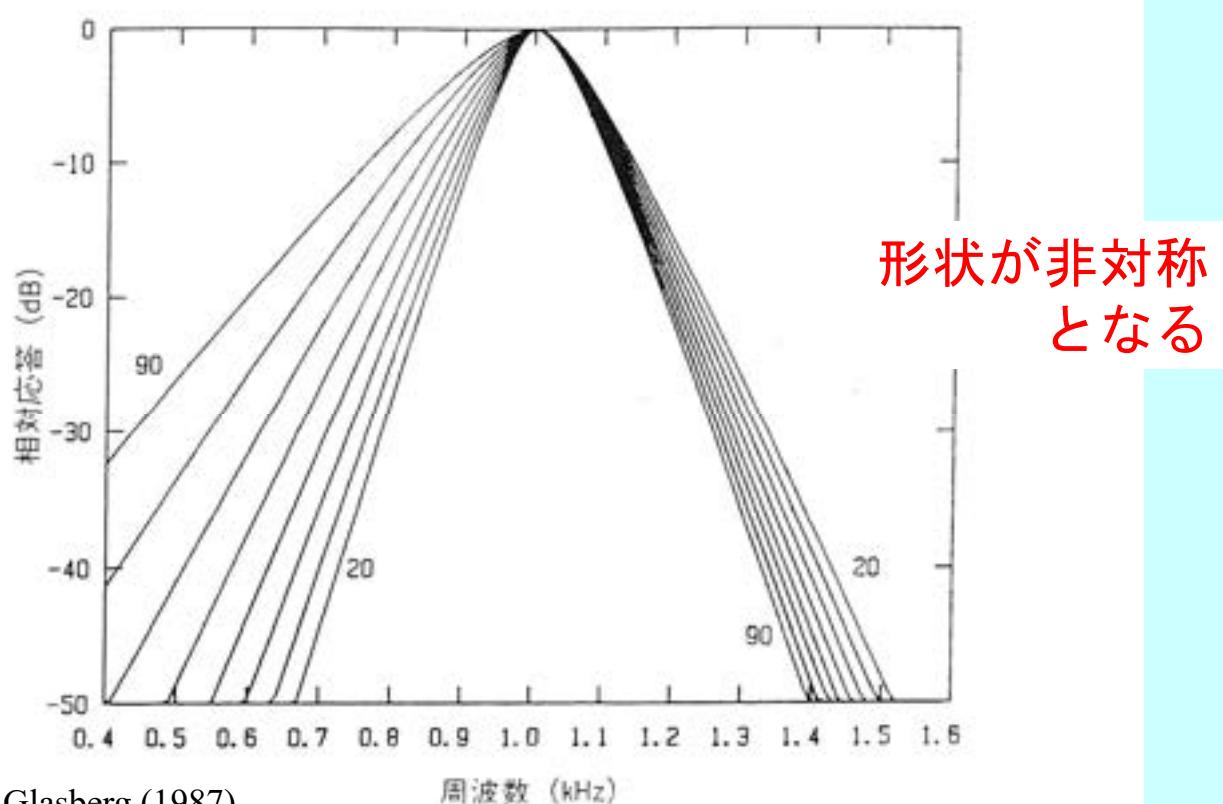
■ 離調聴取 (off-frequency listening)



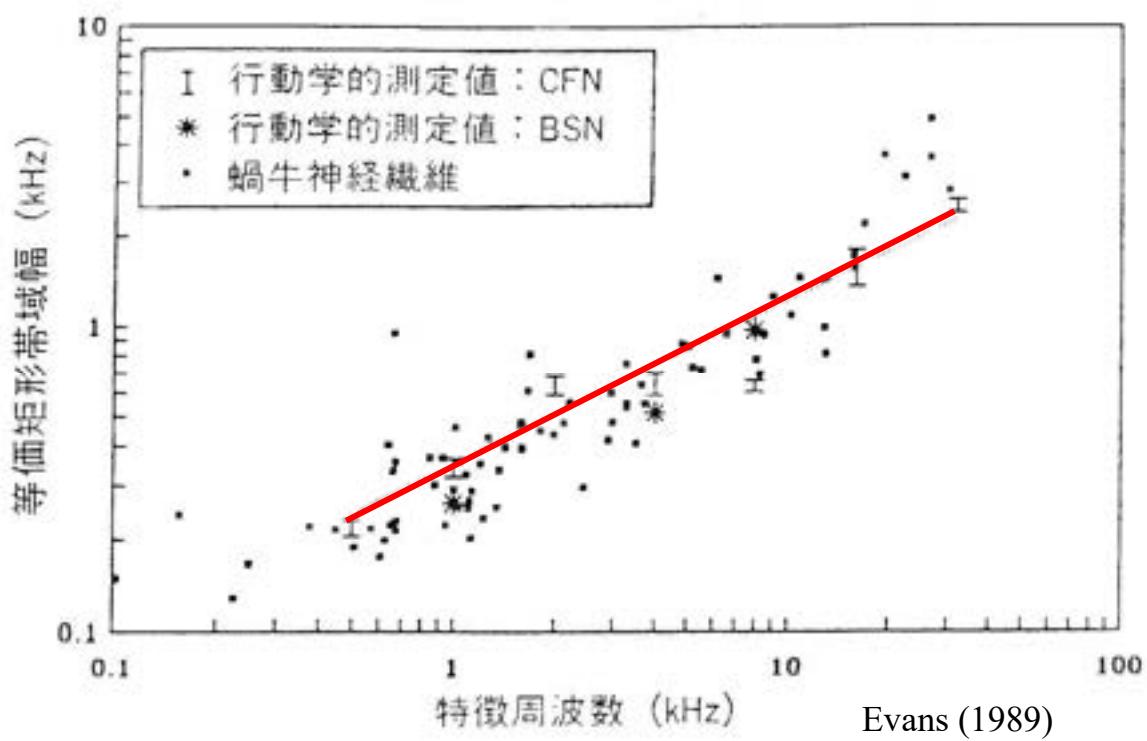
推定された聴覚フィルタ形状



聴覚フィルタ形状のレベル依存性

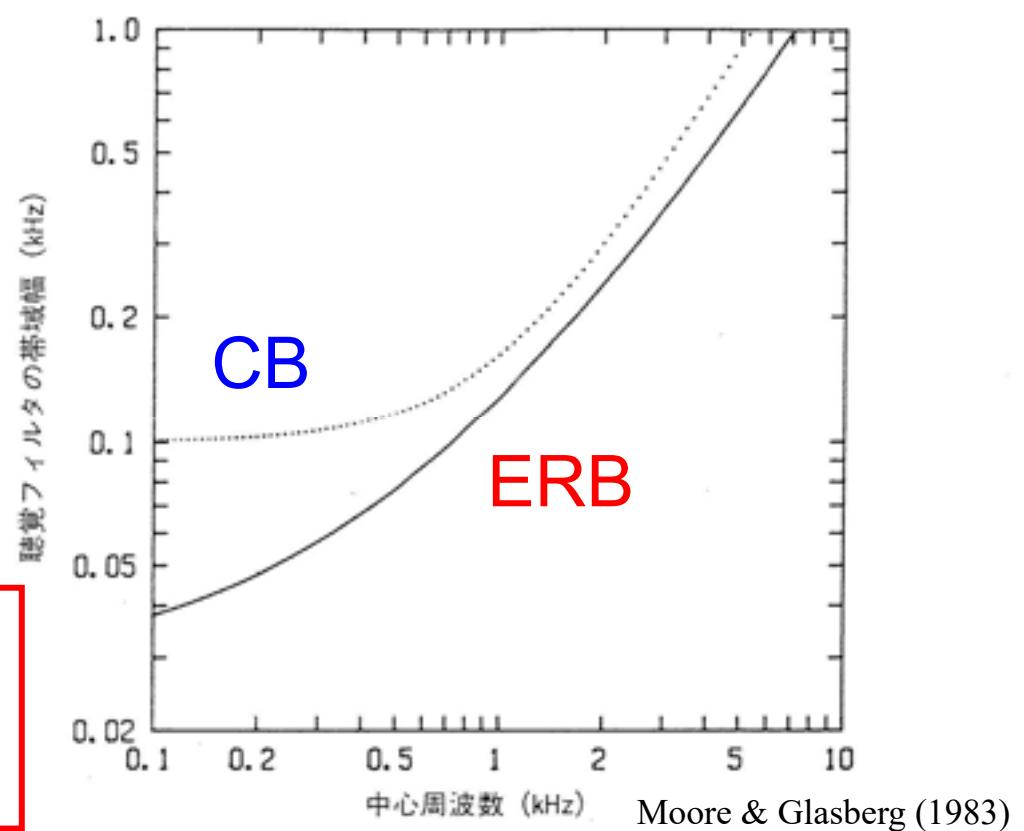


5. 等価矩形帯域幅 (ERB)



1 ERB-rate = 0.9 mm (Greenwood, 1961; Moore, 1986)

フィルタ帯域幅



■ 実験環境の整備



6. マスキングのメカニズム

■ マスキング

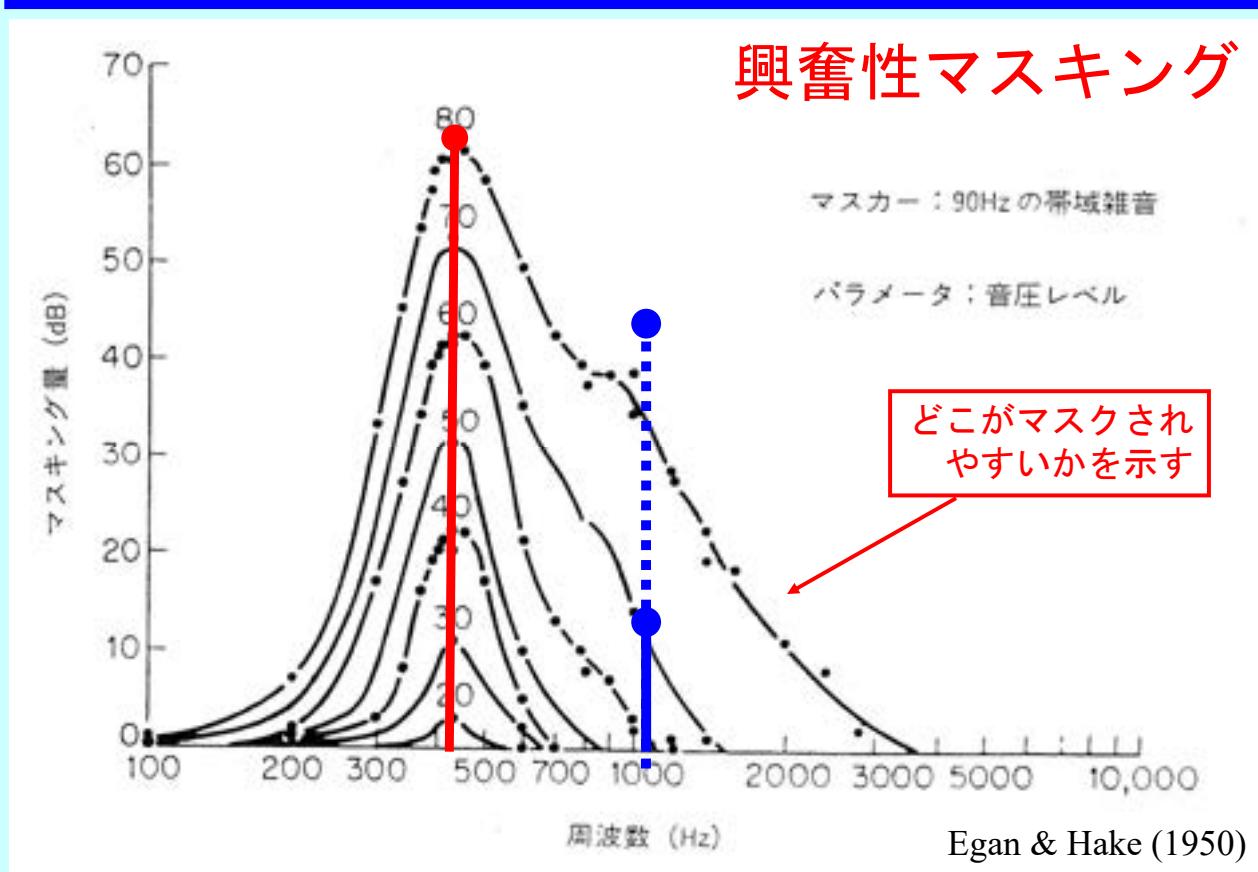
■ 興奮性マスキング

信号によって引き起こされた興奮（神経活動）が他の音によって埋没（swamping）すること

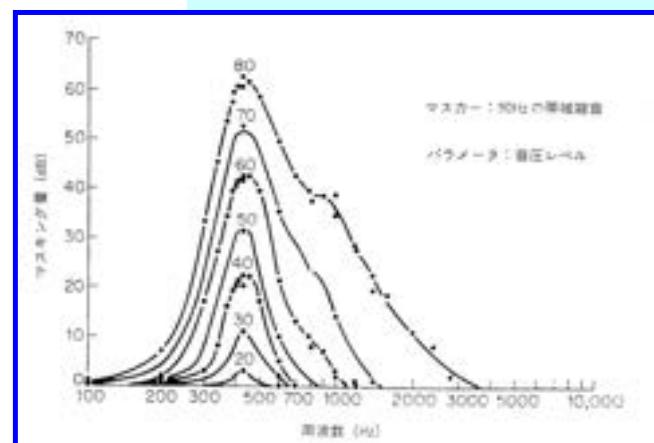
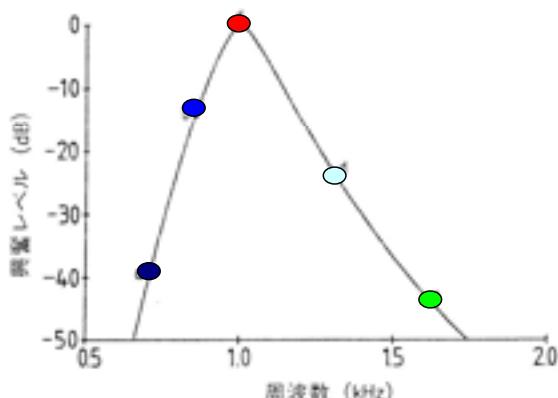
■ 抑圧性マスキング

信号によって引き起こされた興奮が他の音の存在によって抑圧（suppression）されること

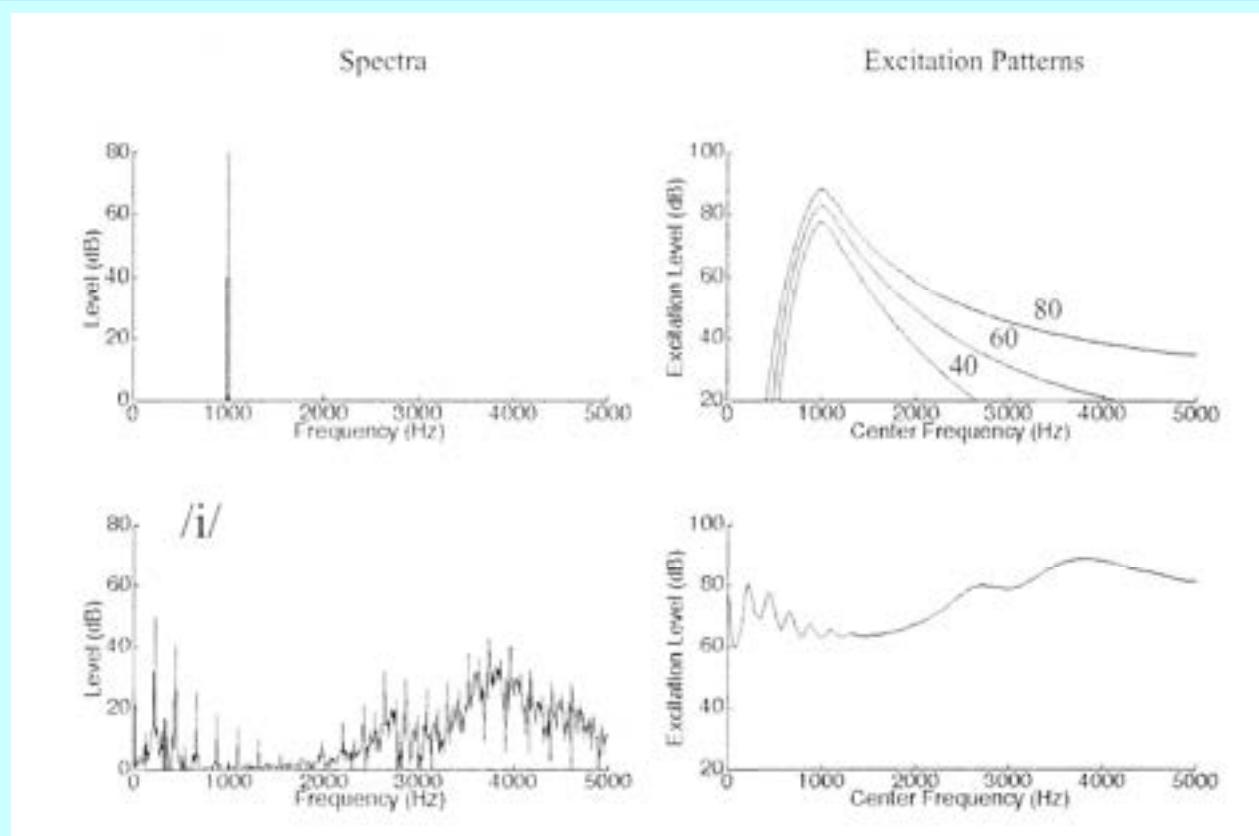
7. マスキングパターン



1 kHzに対する興奮パターン

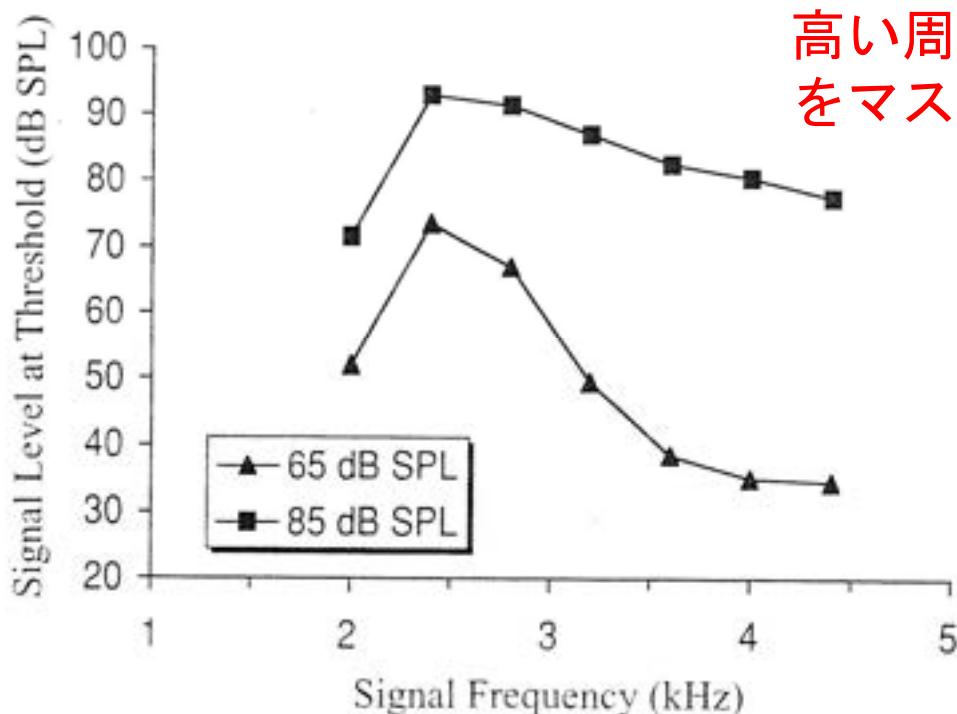


エキサイテーションパターン



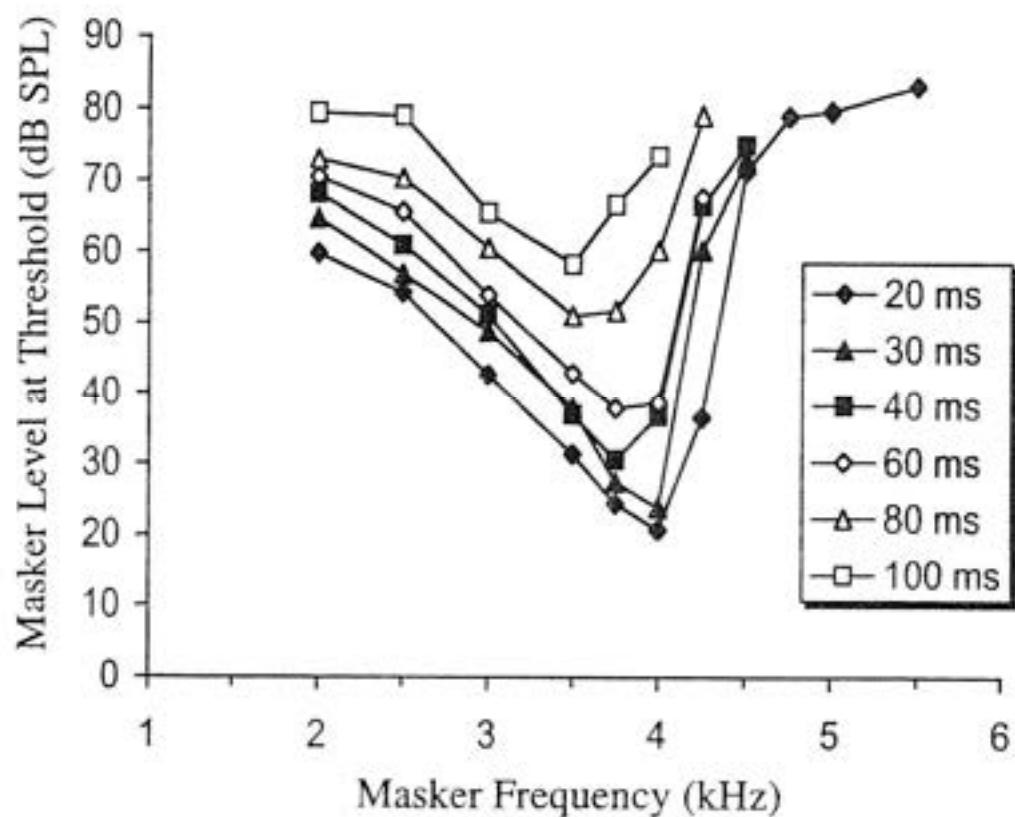
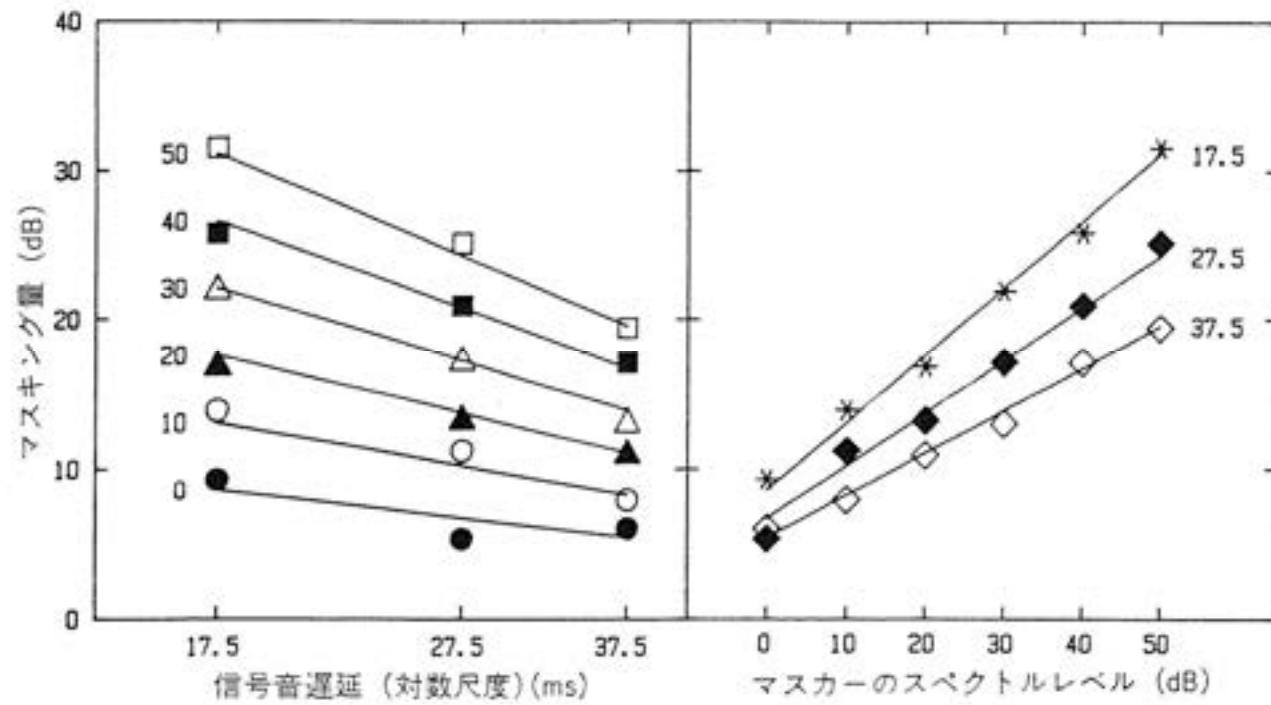
マスキングの上方への拡がり

低い周波数成分は
高い周波数の信号
をマスクしやすい



8. 非同時マスキング

順向性マスキング (forward masking)



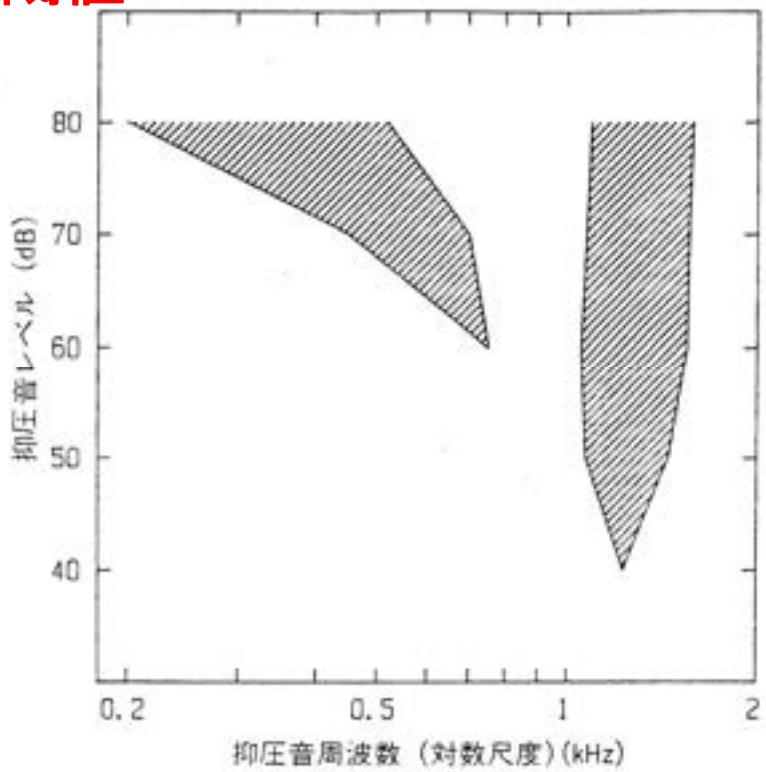
非同時マスキングから得られる抑圧の証拠

■ パルセーション閾値

■ 1 kHz信号

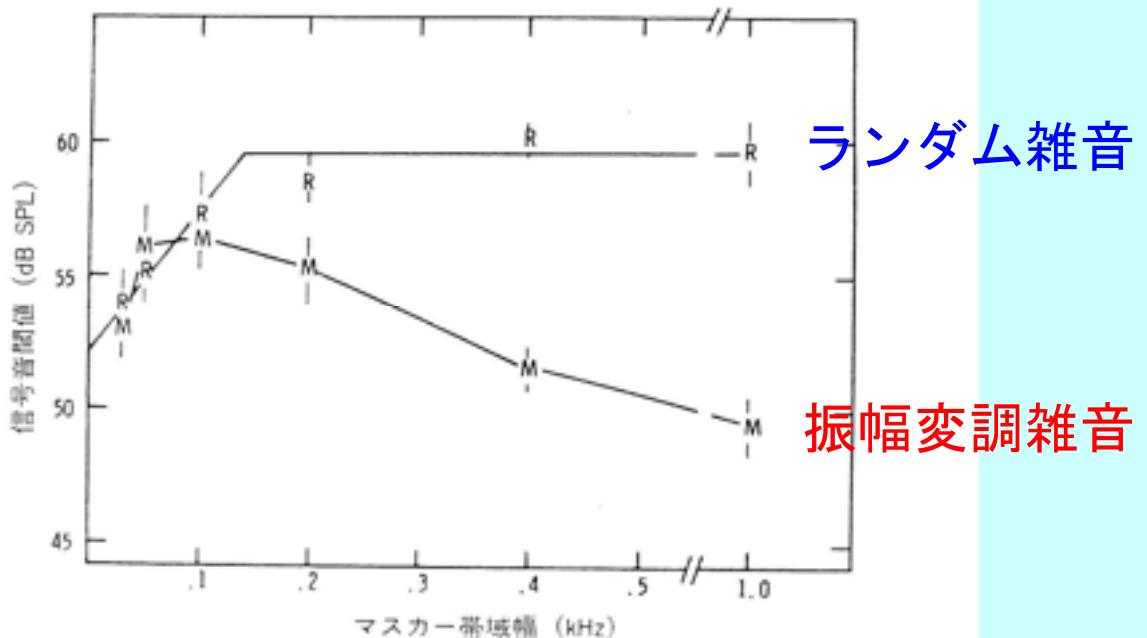
■ 40 dB SPL

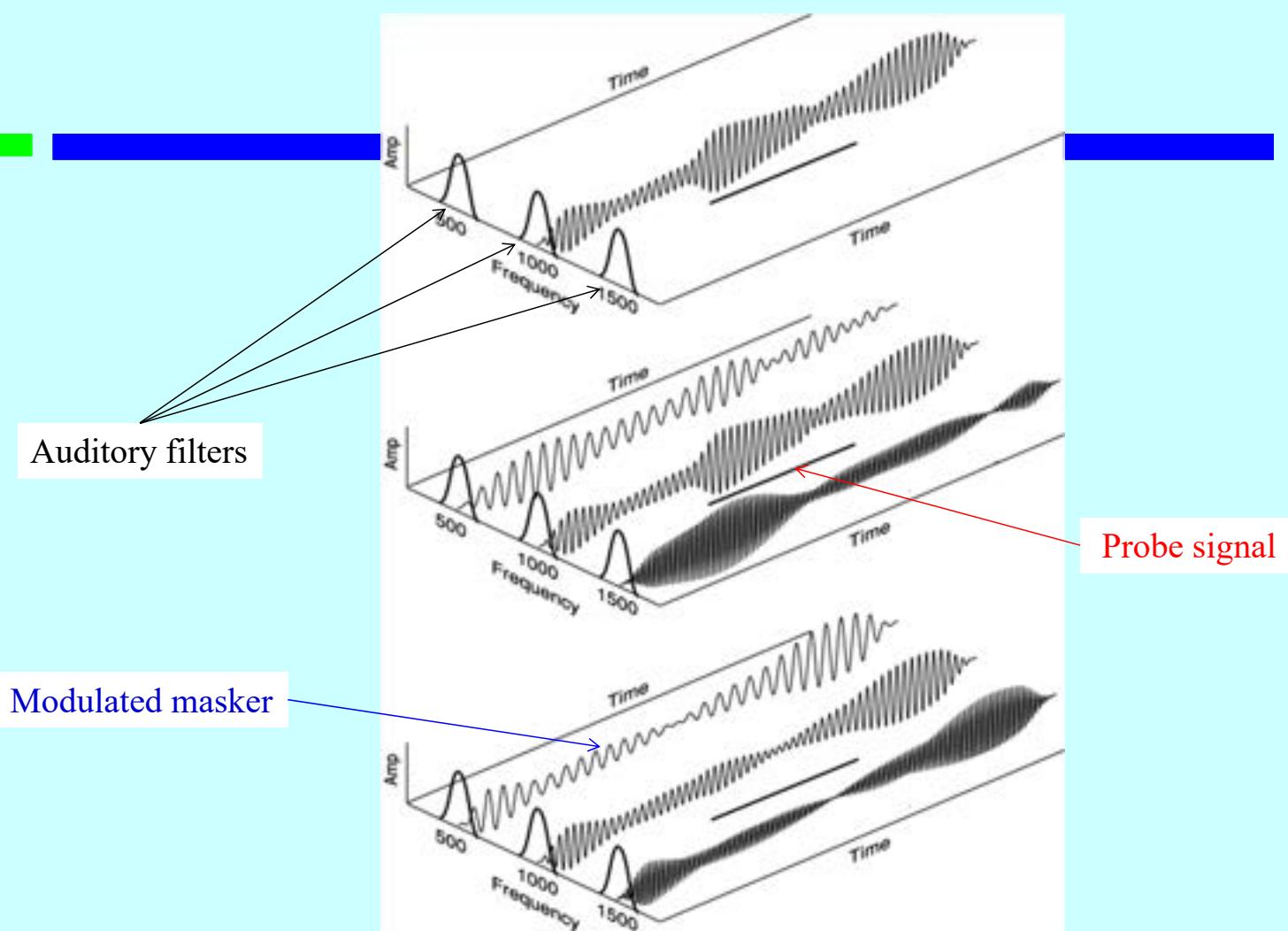
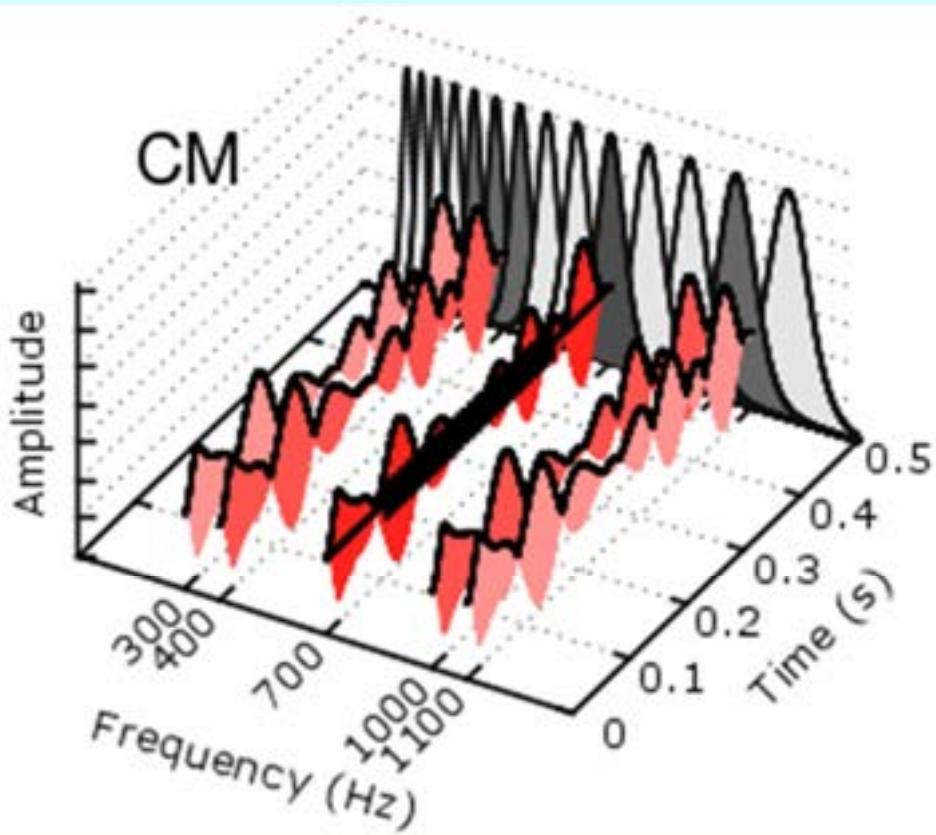
■ 斜線部：閾値
が3 dBダウン
する領域



9. 共変調マスキング解除 (CMR)

■ マスカーの振幅包絡間の手がかり (相関 or 無相関) としたマスキング解除





I485F 知覚情報処理特論

9. 周波数選択性と聴覚フィルタ

2023年5月20日(土)

人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

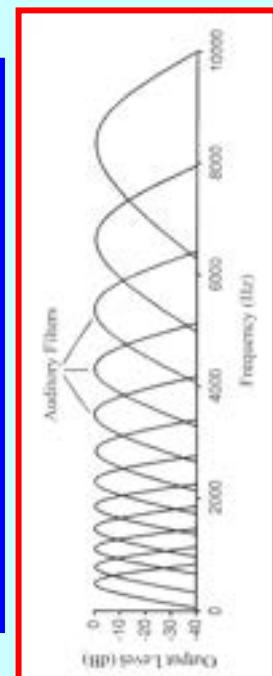
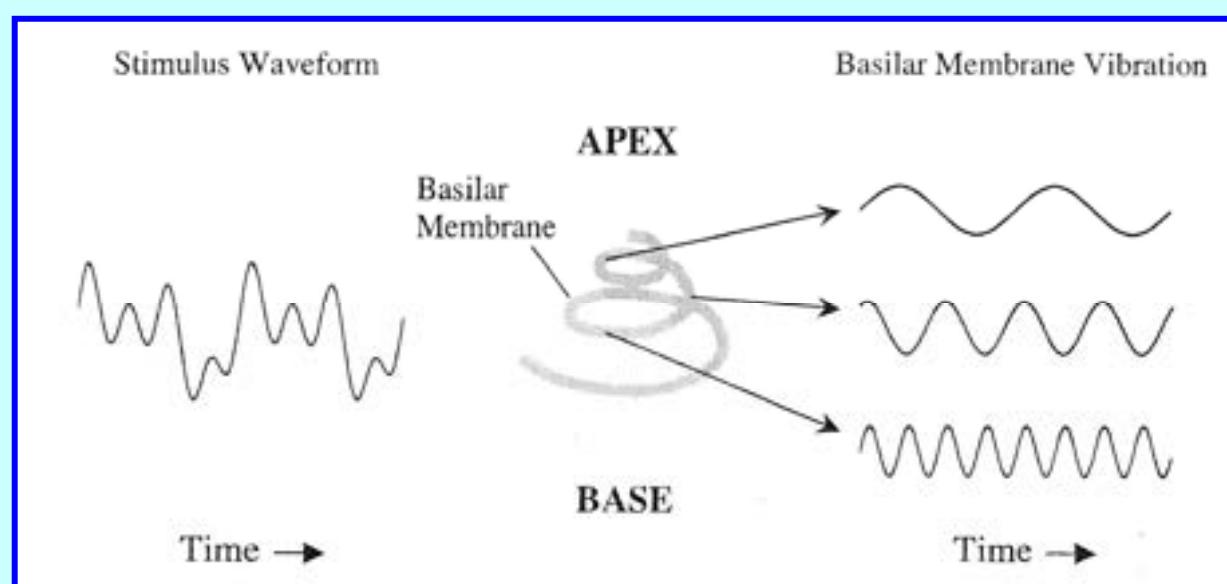
Email: unoki@jaist.ac.jp



1. 周波数選択性

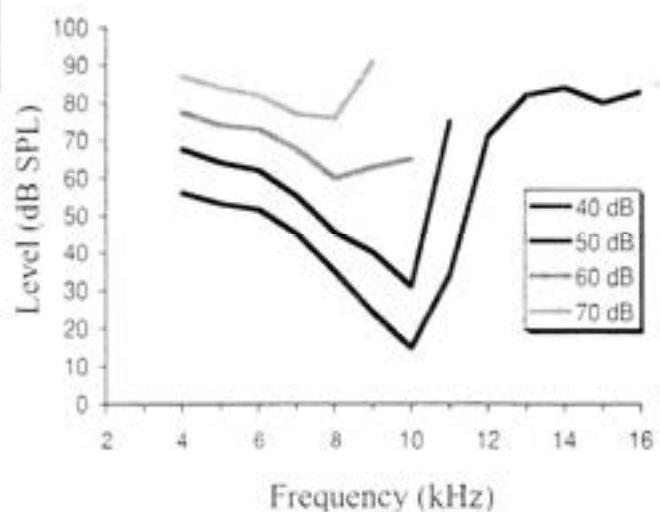
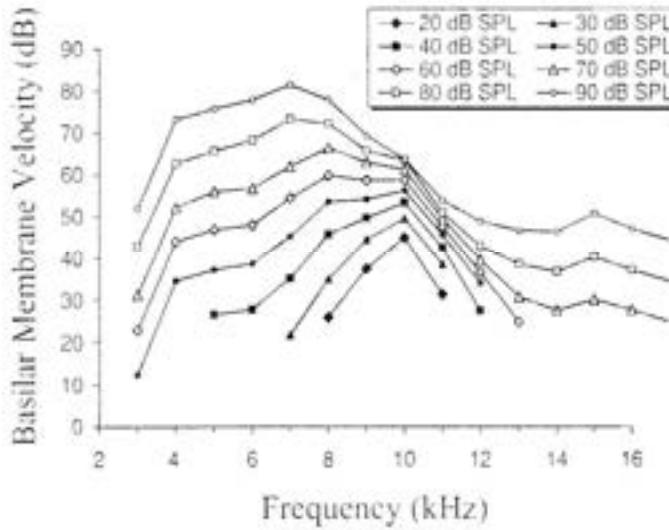
■周波数選択性 (frequency selectivity)

■複合音中の正弦波成分を分解する能力

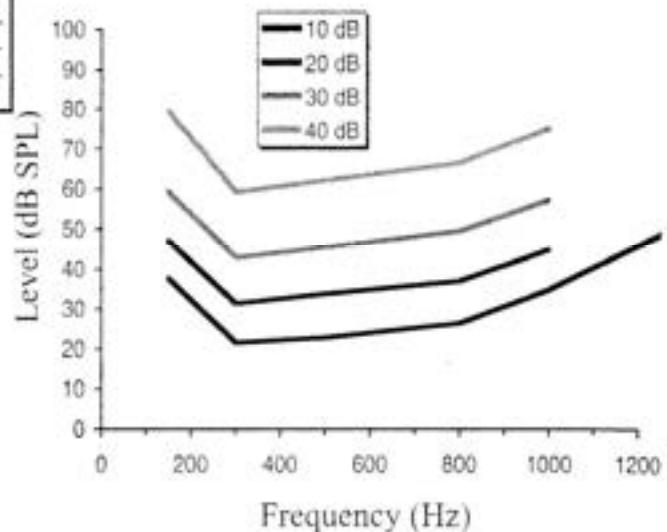
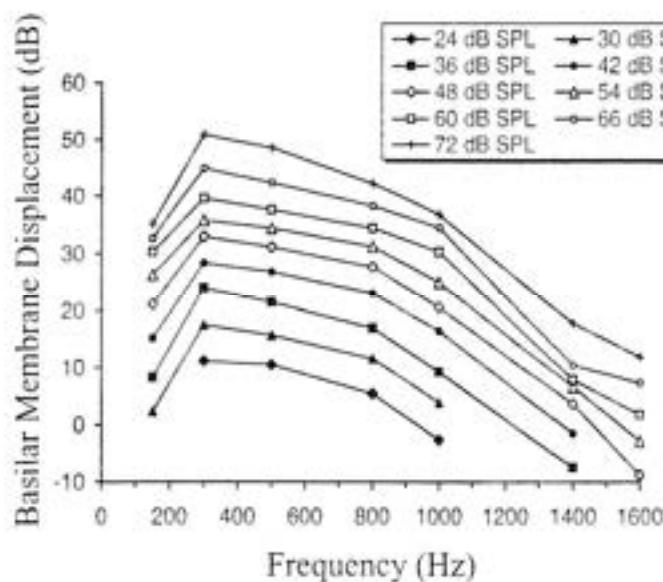


基底膜で観測される周波数選択性

■ 蝸牛底側（高い周波数に対する応答）

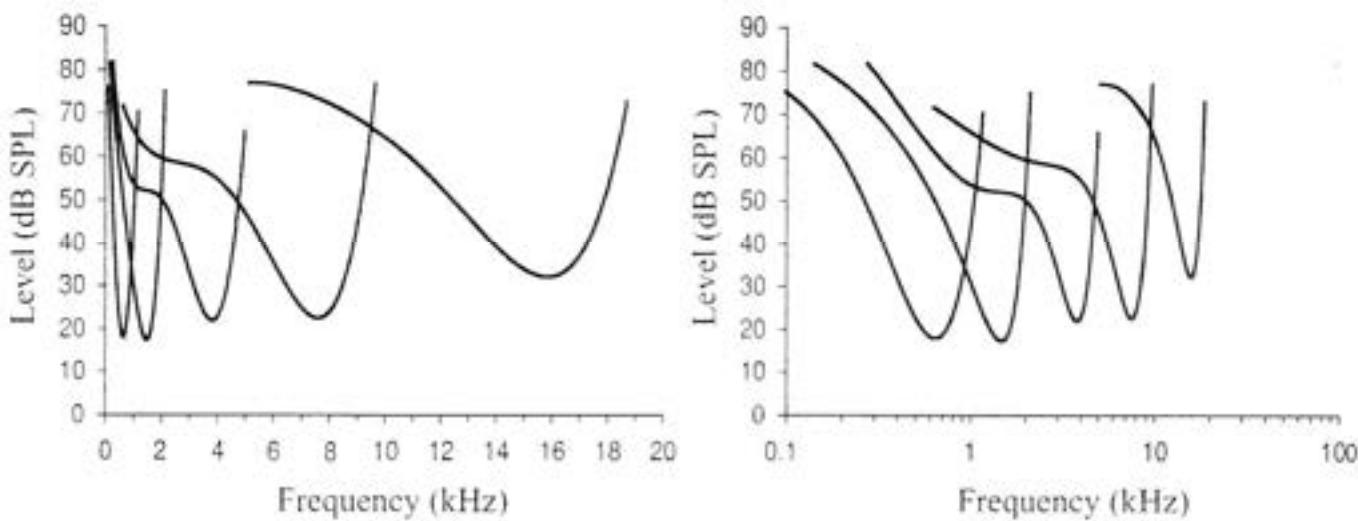


■ 蝸牛頂側（低い周波数に対する応答）



聴神経で観測される周波数選択性

神経発火で見られる同調曲線



2. 周波数選択性を調べる方法

■ 周波数選択性

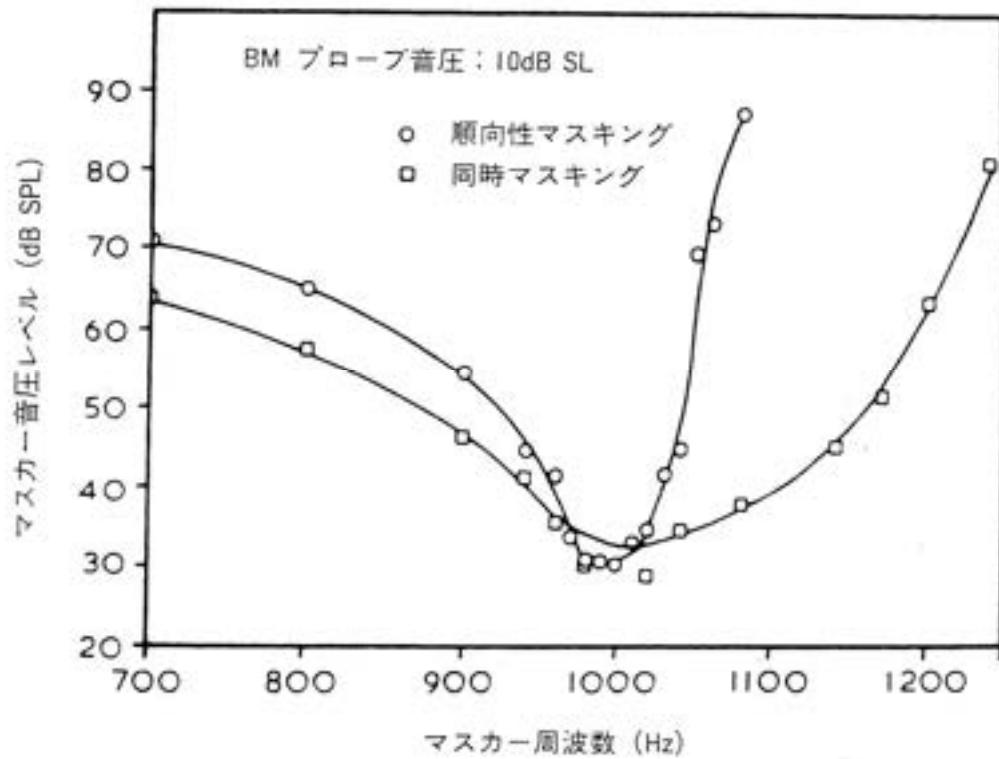
ある音を正弦波成分に分解すること

■ 心理物理的に調べる方法

- 心理物理的同調曲線 (PTC) の測定

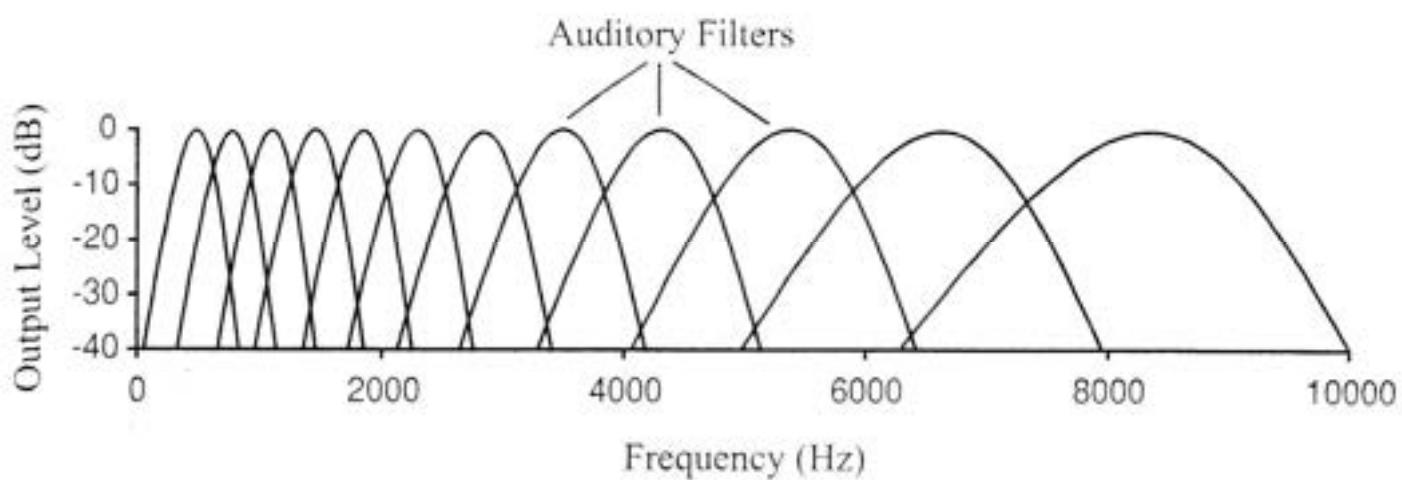
- 聴覚フィルタの推定

心理物理同調曲線（PTC）



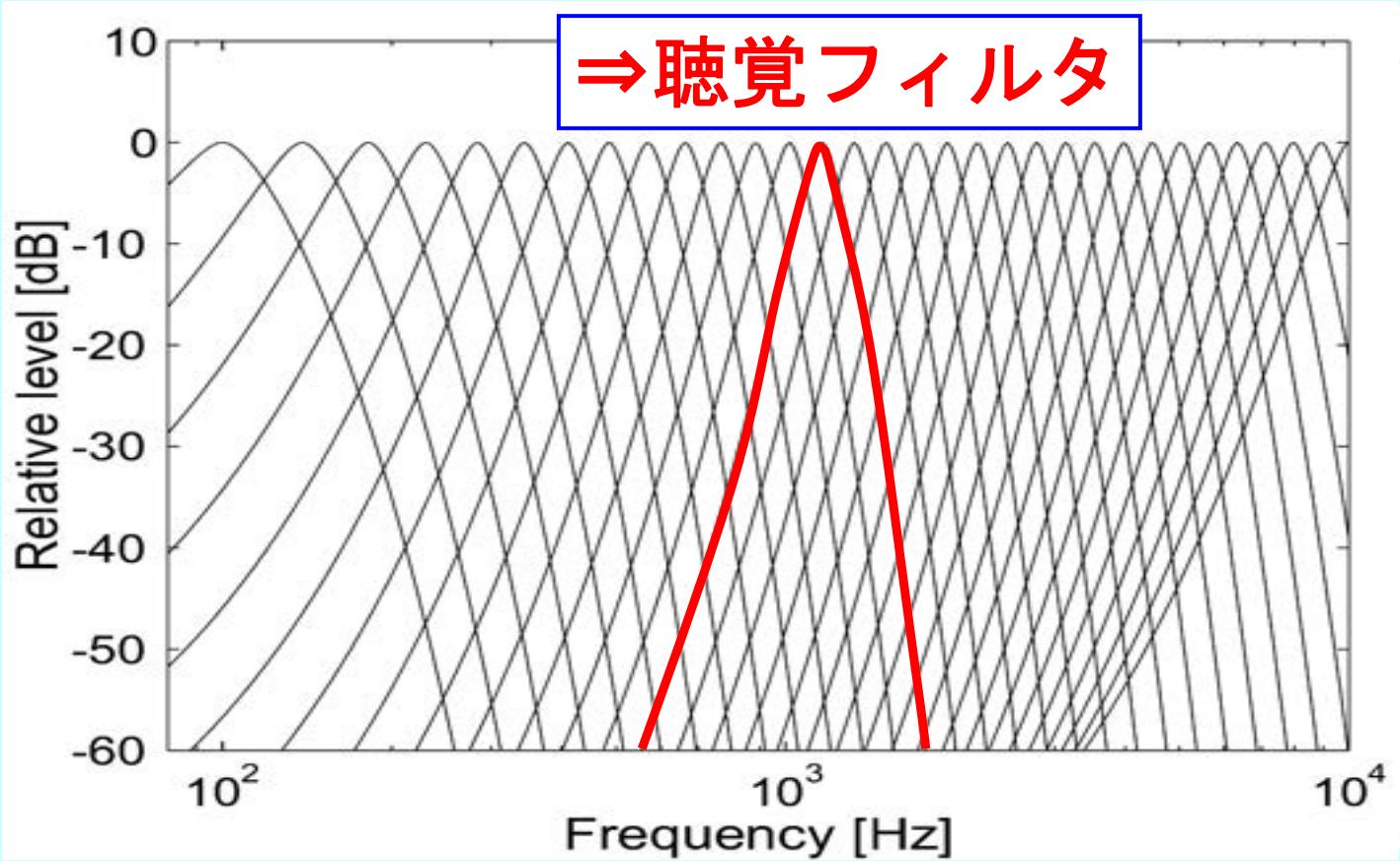
3. 聴覚フィルタバンク

フィルタバンク (filterbank)



非対称、信号レベルに依存
定Qフィルタ

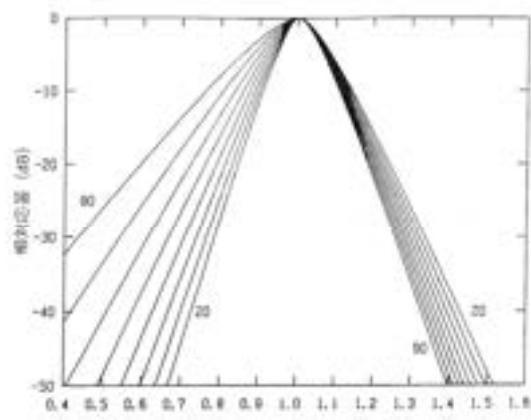
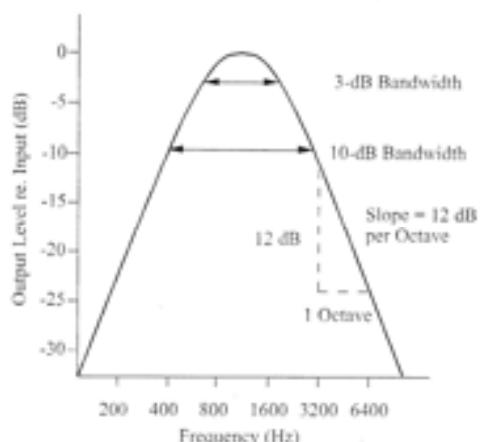
心理物理的同調曲線



4. 聽覚フィルタ

■ “聴覚フィルタ (auditory filter)”

- 周波数選択性に密接に関係→必要な音を正弦波成分に分解する
 - 幅の広がったフィルタ → 周波数分解能（悪）
 - 鋭いフィルタ → 特定の周波数成分だけ通す（良）



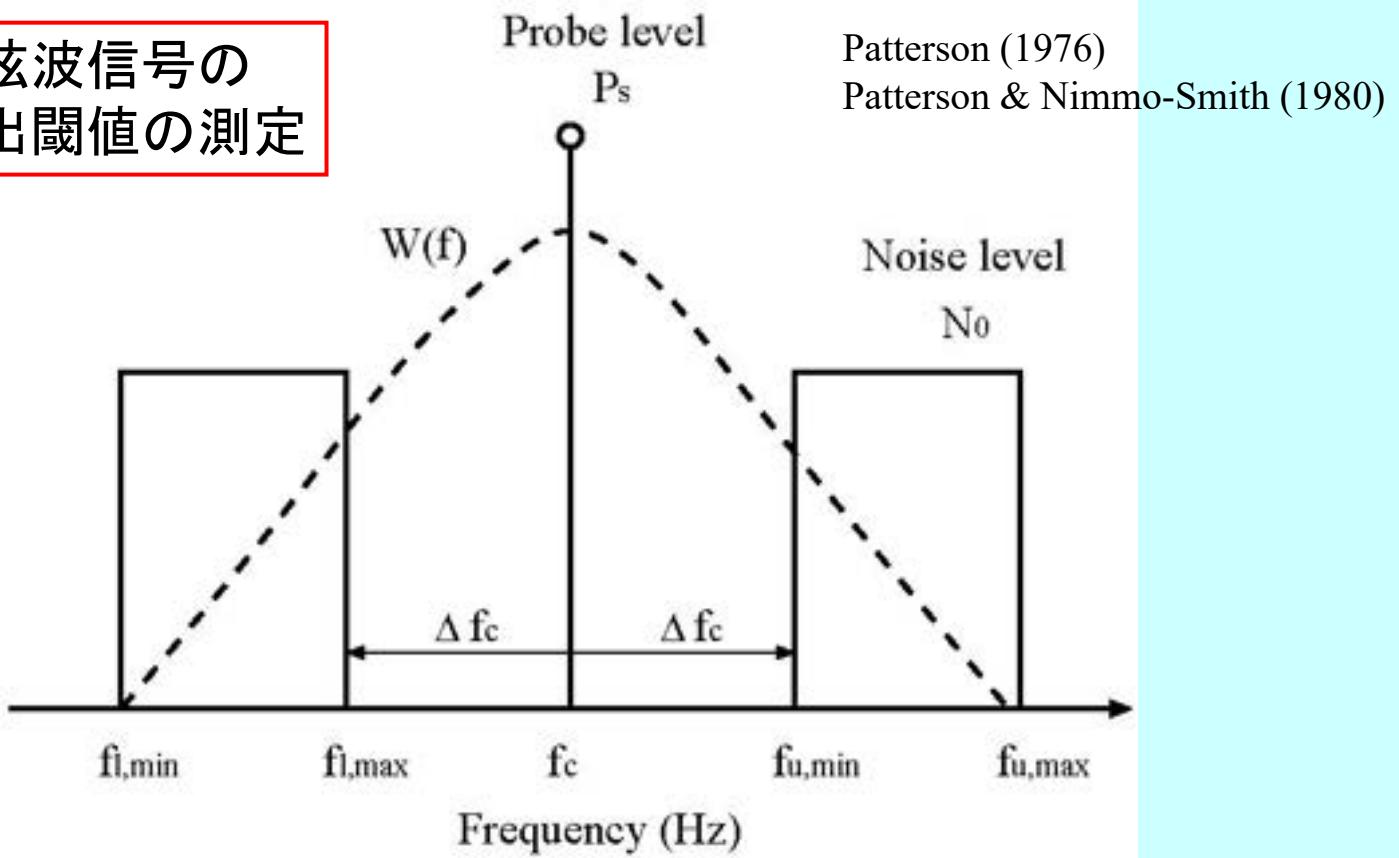
5. 聴覚フィルタの推定【現在】

- 聴覚末梢系における周波数選択性の解明
- 様々なレベル・周波数でのノッチ雑音マスキング実験
 - Baker *et al.* (1998) 250, 500, 1k, 2k, 3k, 4k, 6kHz
 - Glasberg and Moore (2000) 250, 500, 1k, 2k, 4kHz
- ⇒ roex filterでの適合
 - フィルタ形状：レベルに関して変化
 - ピーク周波数でのゲイン：刺激レベルの増加に伴い減少
 - 圧縮の度合
 - 250 Hz ~ 1000 Hz : 信号周波数の増加と共に増加
 - 1000 Hz ~ : おおよそ一定

- 聴覚フィルタバンクの構築に向けて
 - 圧縮型Gammachirp フィルタの適合
 - Patterson, Unoki, & Irino (2003)
 - Roex/Gammachirp フィルタの系統的検討
 - Unoki, Irino, Moore, Glasberg, & Patterson (2006)
- アプローチ
 - 様々な信号周波数に対するノッチ雑音マスキング閾値の測定
 - 測定結果からの聴覚フィルタ形状の推定
 - 聴覚フィルタの非線形性の比較検討

Unoki, Ito, Ishimoto, & Tan (2006)の報告

正弦波信号の検出閾値の測定

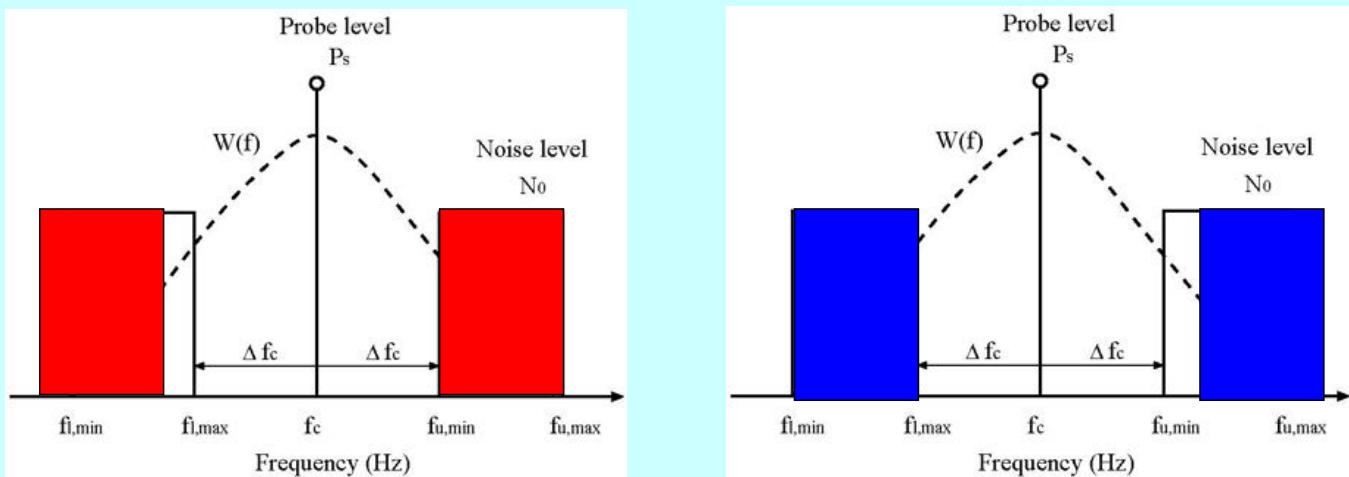


刺激生成・刺激形状

- 信号周波数（可聴域をカバーする7種類）
 - f_c : 125, 250, 500, 1000, 2000, 4000, 6000 Hz
- プローブレベル P_s : 可変
- マスカーレベル N_0 : 固定
 - 125 Hz: $N_0 = 37.3, 47.3, 57.3 \text{ dB/Hz}$
 - 250, 500, 1000, 2000 Hz: $N_0 = 27.3, 37.3, 47.3 \text{ dB/Hz}$
 - 4000, 6000 Hz: $N_0 = 17.3, 27.3, 37.3 \text{ dB/Hz}$
- ノッチ幅の移動 : スライド形式
- 雑音帯域幅（片側ずつ）: $0.4f_c$
- UEN (uniformly exciting noise) を利用しない

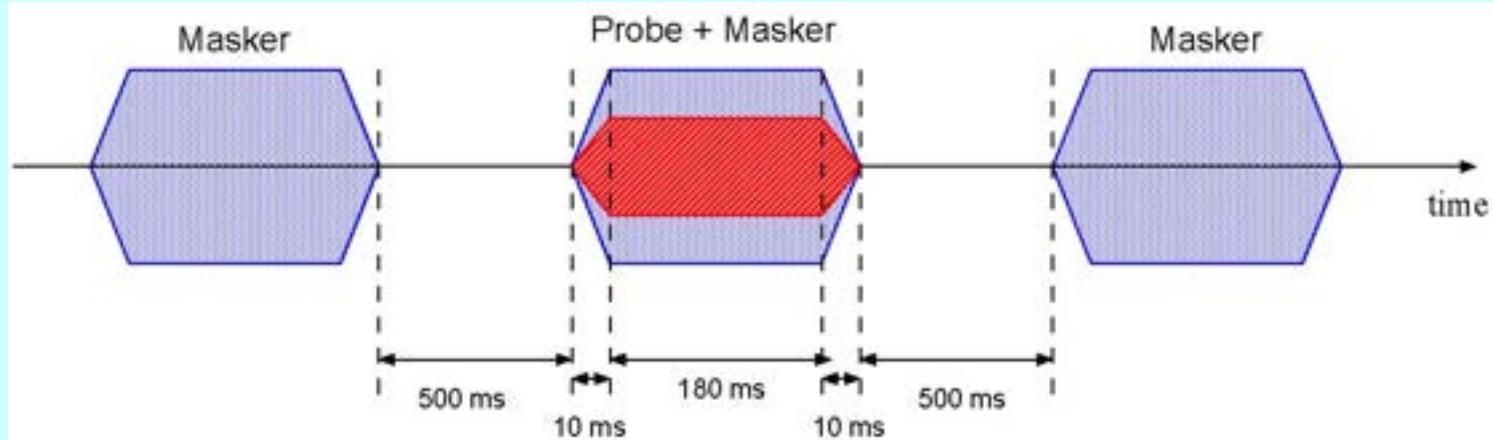
ノッチの対称・非対称条件

- 対称の場合 (○) $\Delta f_c / f_c = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$
- 非対称の場合 $\Delta f_c / f_c = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$
- 低域側に非対称 (◀) 高域側に非対称 (▶) :
- 低域側が0.2だけ広い場合 高域側が0.2だけ広い場合



- 3区間 ($\Rightarrow 3\text{AFC}$)
- プローブの配置：ランダム
- 過渡部：cos ramp
- 雑音作成：Linux上 (Matlab ver. 6.5.1, Mathworks)
- 刺激作成・制御

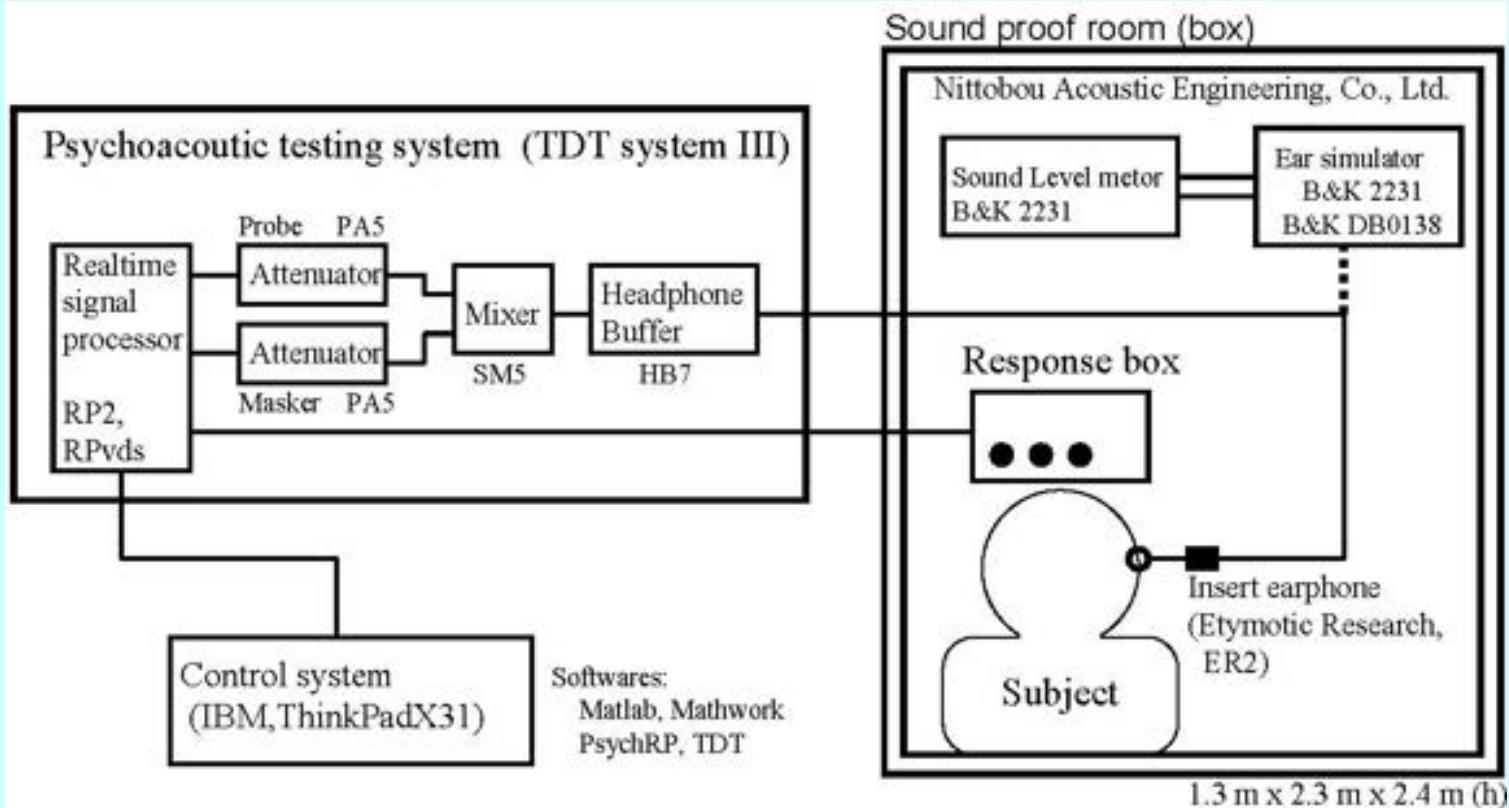
TDT RP2 (RPvds, サンプリング周波数 24.41 kHz), PsychRP



被験者

- 男性5名
 - MU, YI, YY, AH, MT
- 年齢：23歳～35歳
- 聴力レベル：両耳とも10 dB HL以内
 - RION AA-72Bオージオメータ
 - 測定周波数範囲：125 Hz – 8 kHz
- 呈示：単耳（特性のよい方の耳）
- 絶対閾値
 - dB HL -> dB SPLへ変換
- 訓練あり：2時間以上

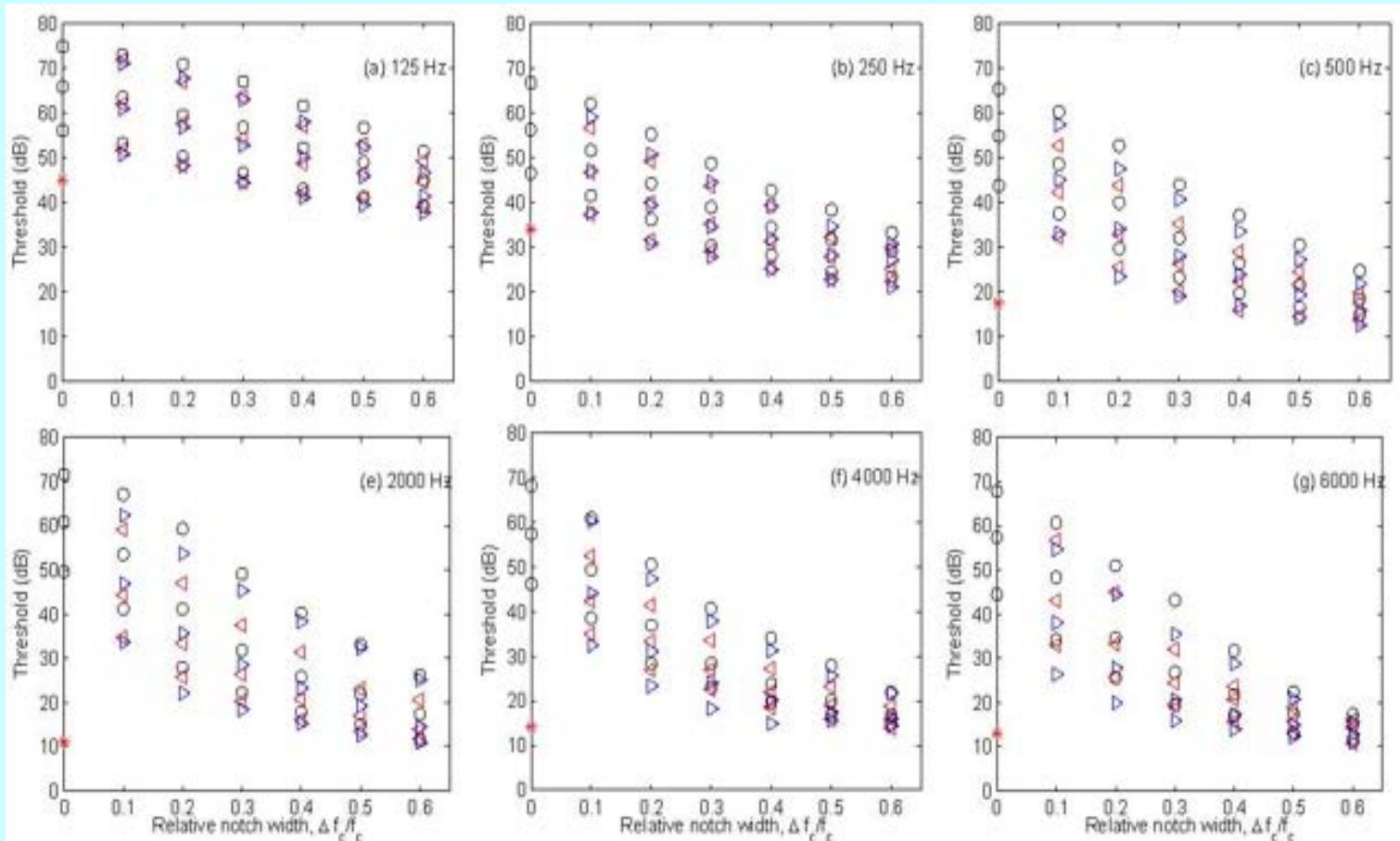
実験システム



実験手続き

- 3肢強制選択 (3AFC) を利用して閾値を測定
- three-down one-up 手続き (79.4% tracking, Levitt, 1970)
- 応答BOXを利用して回答
- 正解のフィードバックあり (LEDの点灯)
- しきい値の算出
 - 12回の転換点を利用
 - 最初の4点 : 5 dB 刻み
 - 後ろ8点 : 2 dB刻み
 - 後ろ8点の転換点の平均値 ⇒ マスキング閾値
- 2回測定 (8 dB以上の差がある場合、取り直し)

実験結果



測定結果のまとめ

■ 測定された閾値の結果

■ 絶対閾値より上に位置

■ 対称・非対称のノッチ配置の条件

⇒ 聴覚フィルタ形状が対称ではない

■ マスカーレベルが高いとき

低域側非対称の閾値 < 高域側非対称での閾値

⇒ フィルタ形状が低域側に非対称である可能性

■ マスカーレベルが低いとき

両者の差が見られない

⇒ フィルタ形状が対称である可能性

聴覚フィルタ形状の推定

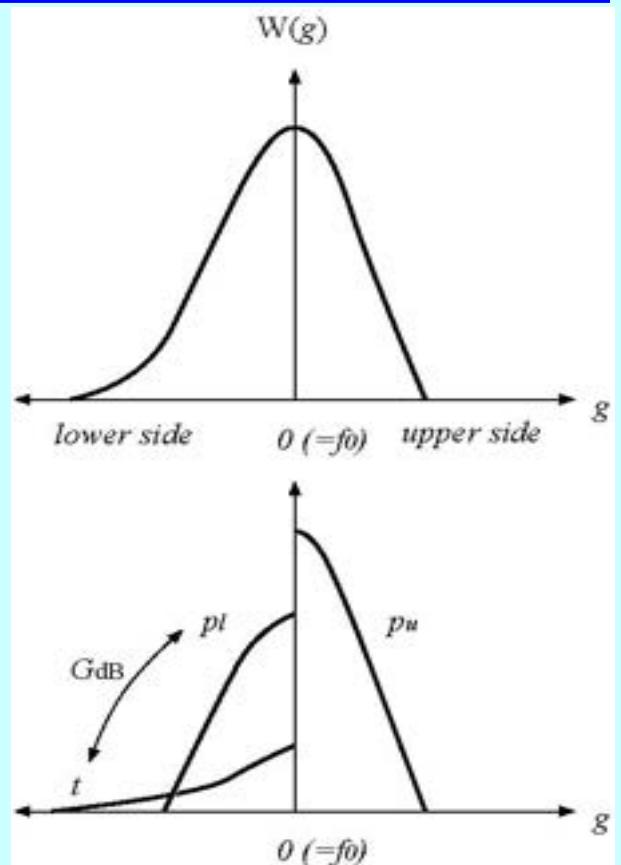
Roex (Rounded Exponential) filter
(Glasberg & Moore, 2000)

$$W(g) = \begin{cases} (1-w)(1+p_l g) \exp(-p_l g) \\ + w(1+tg) \exp(-tg), & f \leq f_c \\ (1+p_u g) \exp(-p_u g), & f > f_c \end{cases}$$

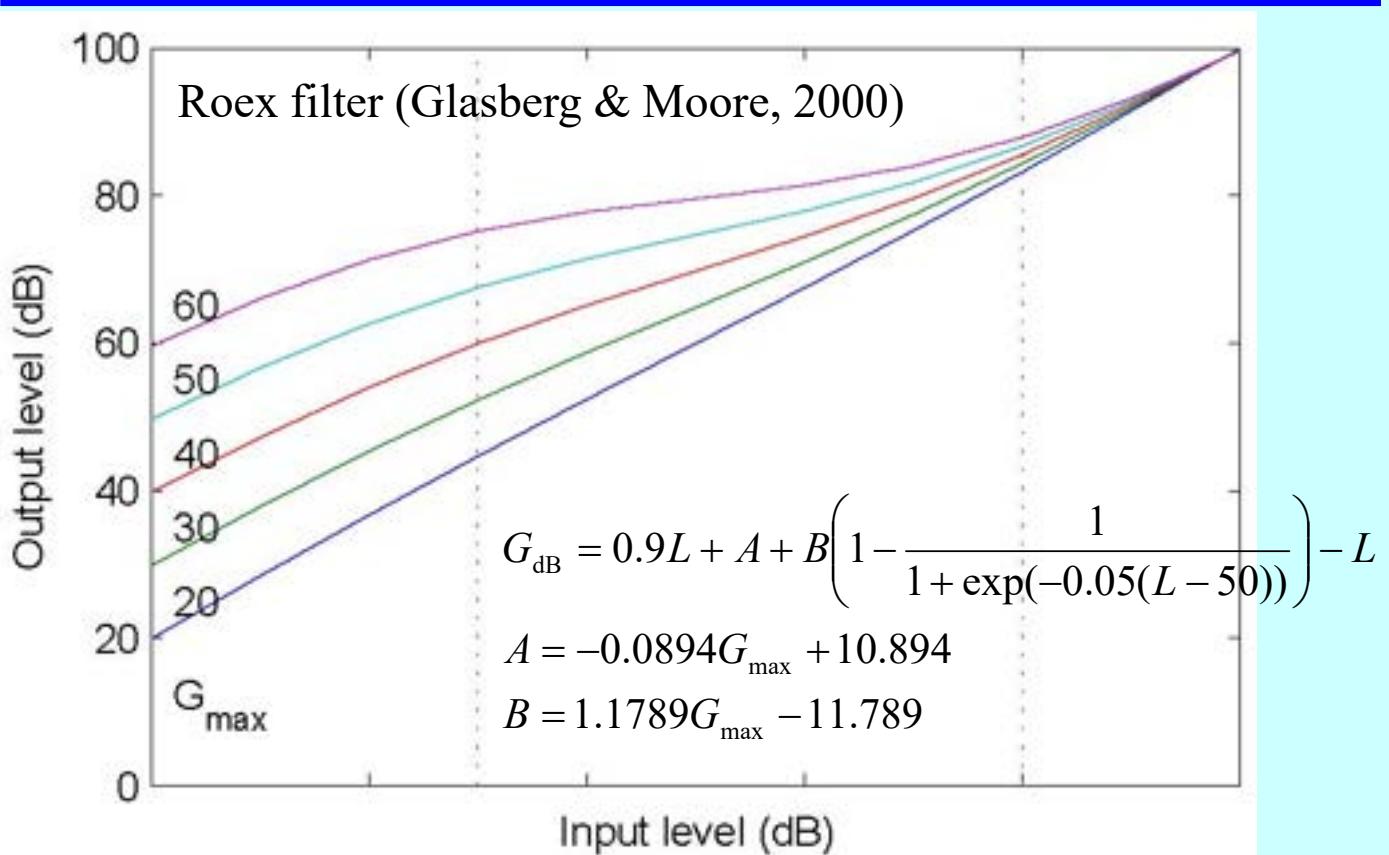
$$g = |f - f_c| / f_c$$

$$w = 1/(G_{\text{lin}} + 1)$$

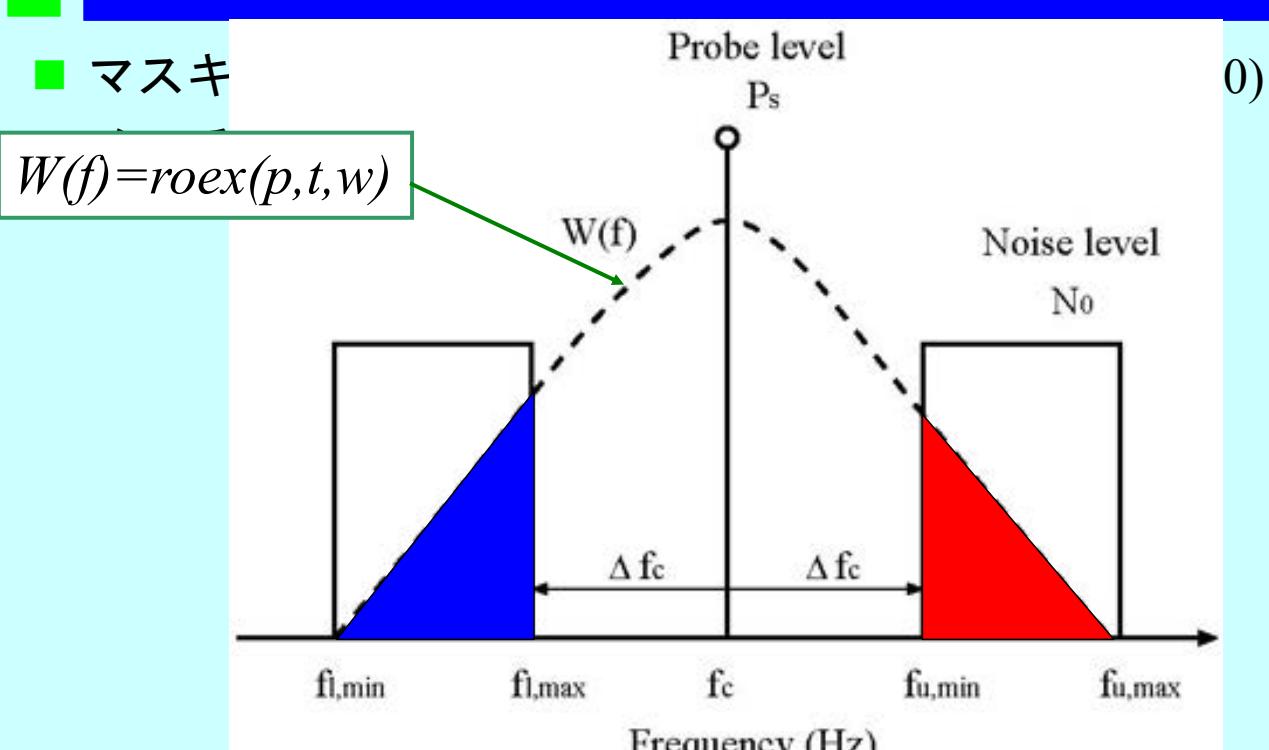
$$G_{\text{lin}} = 10^{G_{\text{dB}}/10}$$



入出力特性



マスキングモデル



$$P_s = K + N_0 + 10 \log_{10} \left\{ \int_{f_{l,\min}}^{f_{l,\max}} W(f) df + \int_{f_{u,\min}}^{f_{u,\max}} W(f) df \right\}$$

■ 適合パラメータ

- フィルタパラメータ: p_b, t, G_{\max}, p_u

- 検出効率 K , 絶対しきい値 Abs

■ 個別適合: 各周波数ごとに適合

■ 補正: MidEar補正 (Moore *et al.*, 1997)

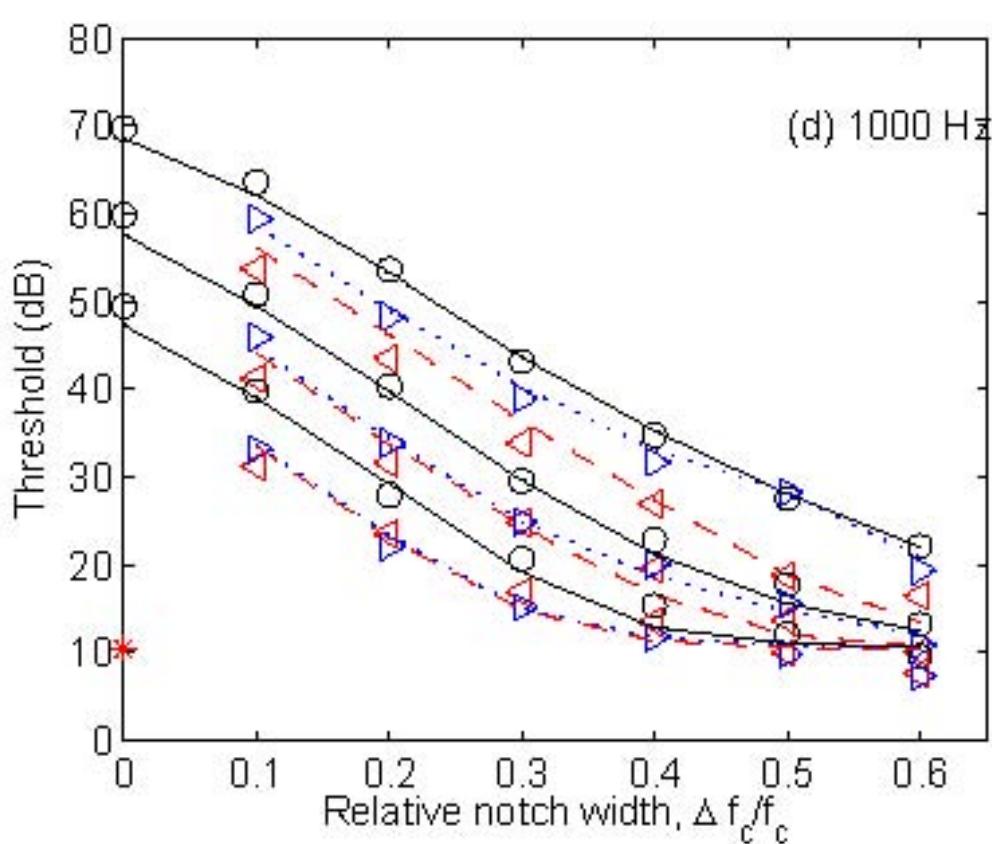
■ 最適化: PolyFit (Polynomial fitting) 法

- マスキングしきい値と推定値の実効誤差 \Rightarrow 最小

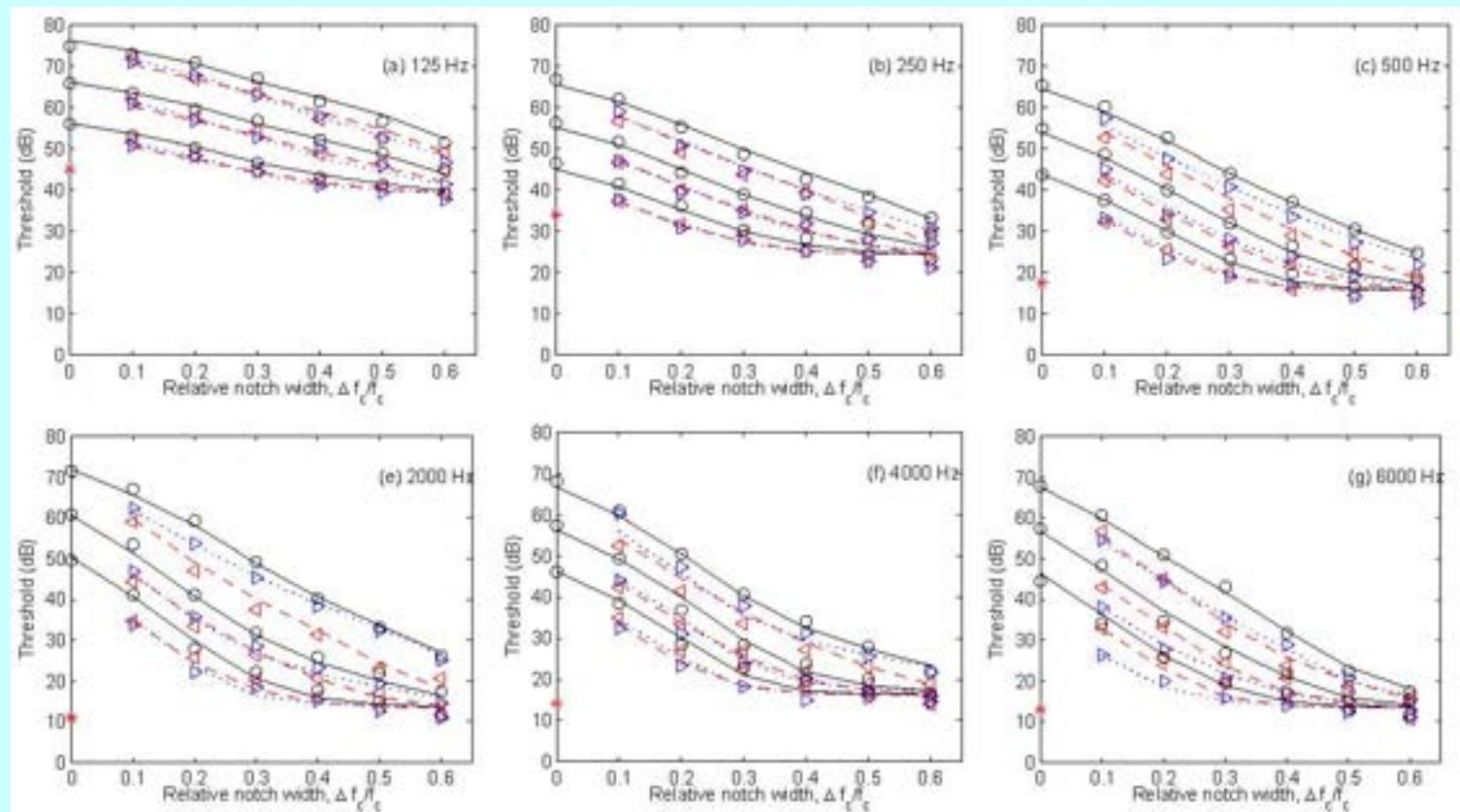
■ 離調聴取を考慮

- ピーク周波数に関するSNR最大のフィルタ位置を選択

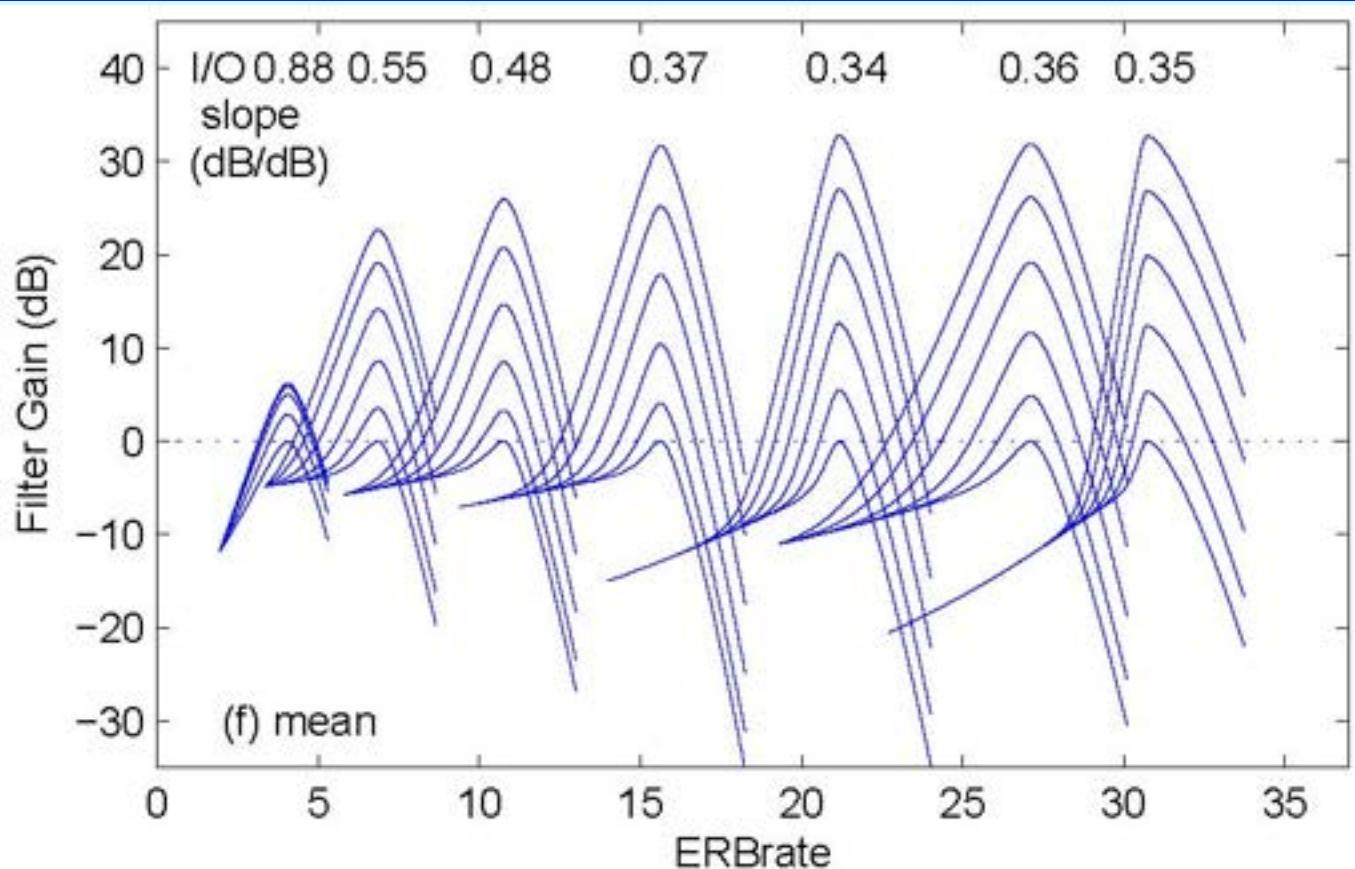
マスキング閾値と推定値



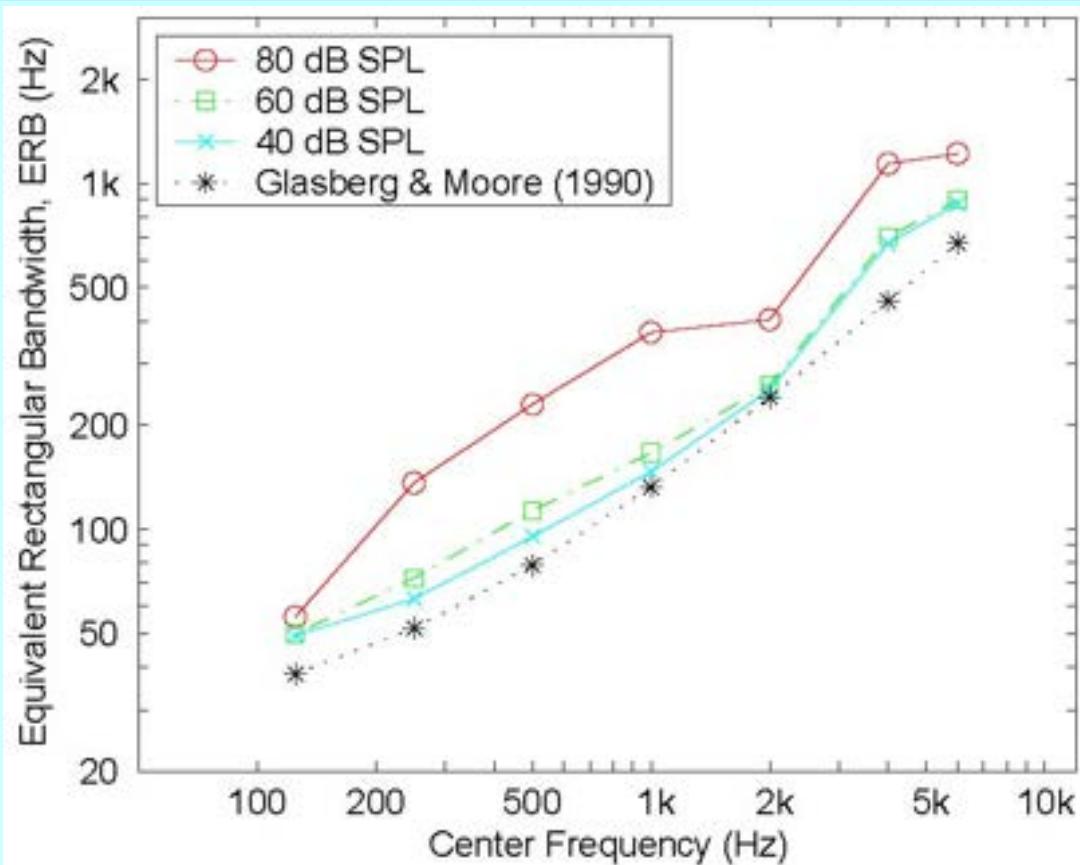
マスキング閾値と推定値



フィルタ形状（平均データ）

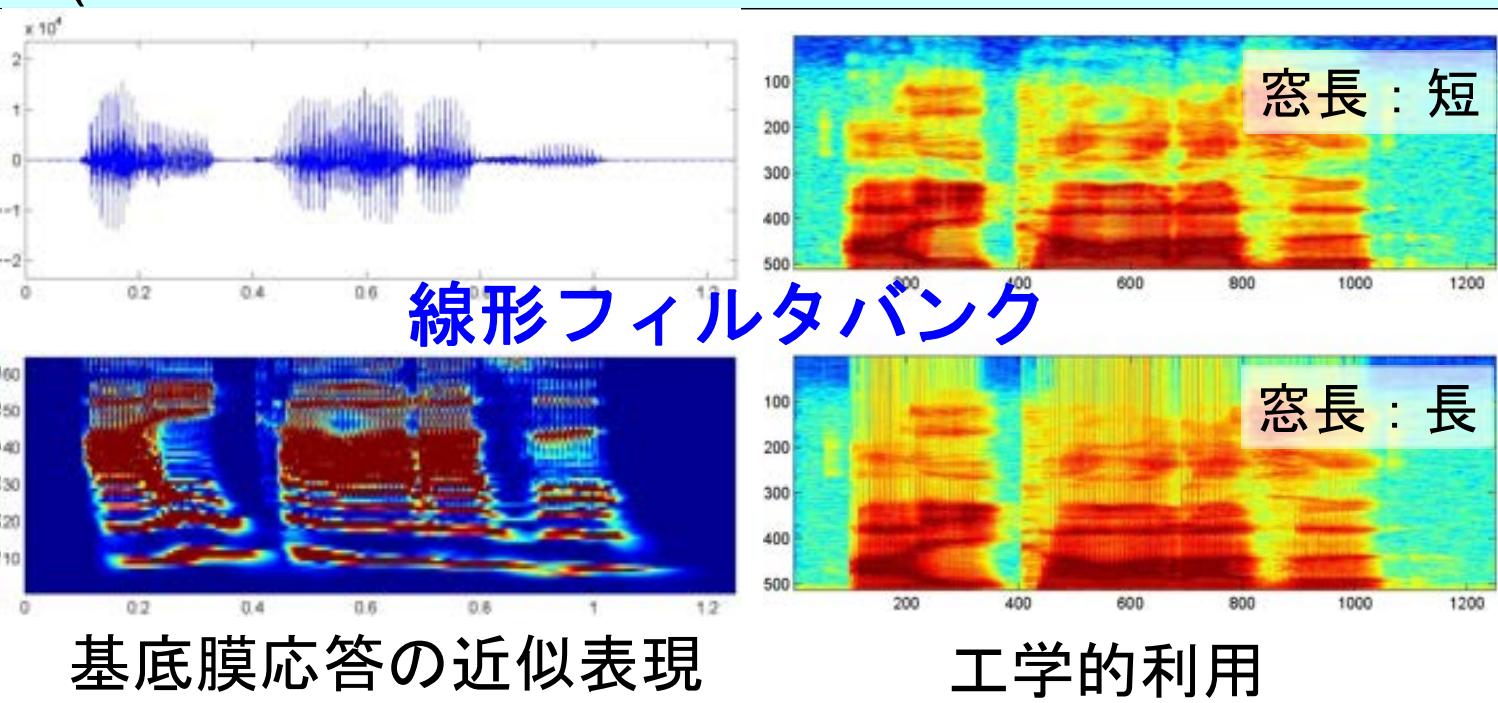


フィルタ帯域幅 (ERB)



6. 聴覚フィルタバンクの実現に向けて

定Qフィルタバンク vs. 短時間フーリエ変換
(ウェーブレット変換) (窓関数による切取り)



I485F 知覚情報処理特論

10. 音の大きさの知覚とモデル

2023年5月20日(土)

人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>
Email: unoki@jaist.ac.jp



1. 音の知覚



■ 三つの属性:

- “音の大きさ” ← “音の強さ”
- “音の高さ” ← “周期” or “調波”
- “音色” ← その他いろいろ

■ “音” の知覚 (perception) の三つの属性

■ 音の大きさ ⇒ ラウドネス (loudness)

音を「小さい」から「大きい」に伸びた尺度上に順序づけられる音の強さに関する聴覚知覚の属性

■ 音の高さ ⇒ ピッチ (pitch)

音を音階上に順序づけることができるような聴覚知覚の属性

■ その他 (音色) ⇒ 音色 (timbre)

同じラウドネスで同じピッチをもつ二つの音を異なる音であると断言できるような聴覚知覚の属性

2. 最小可聴値

■ 最小可聴値 (threshold of audibility)

ある条件のもとで聞き取ることができる最も低いレベルのこと。聴覚閾値あるいは絶対閾値ともいう。

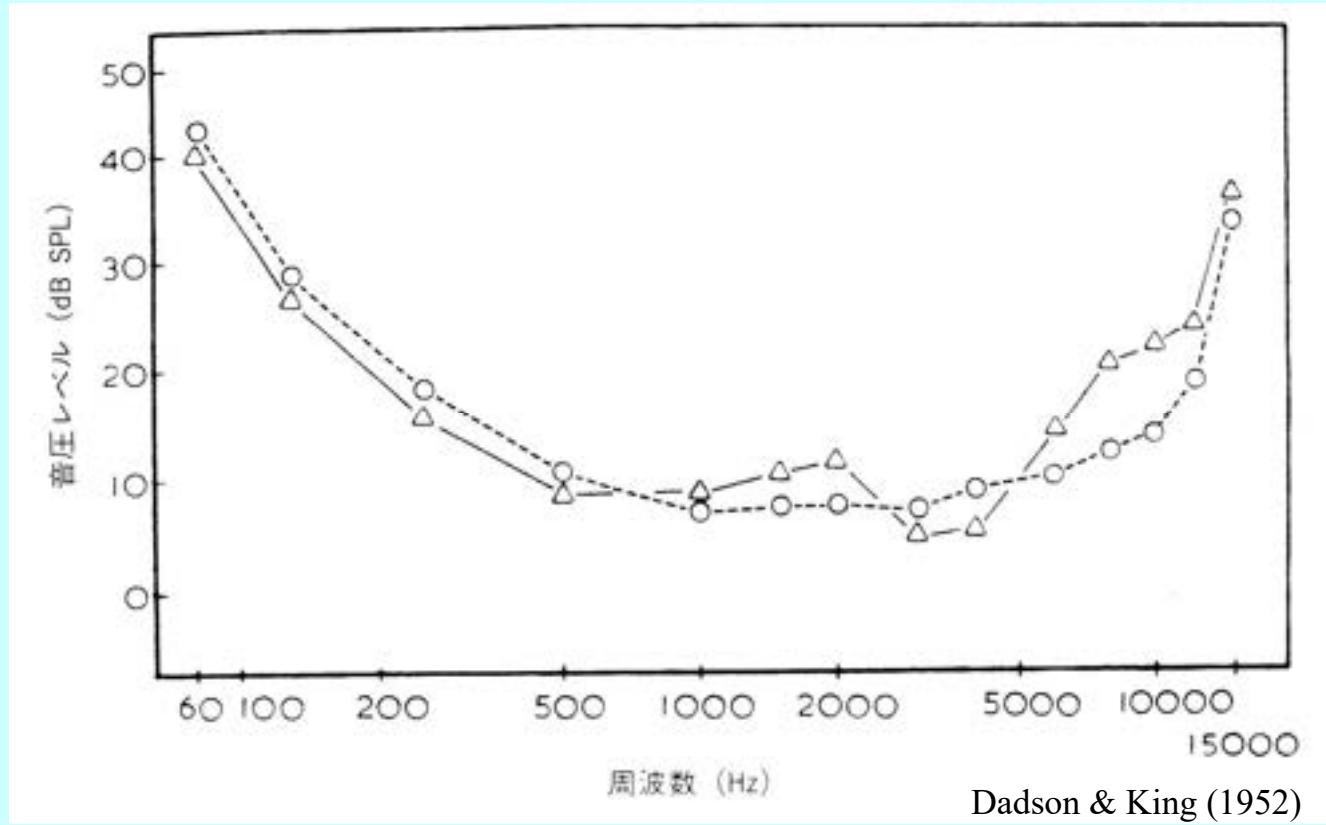
■ 最小可聴音圧 (MAP: minimum audible pressure)

プローブマイクを利用して、外耳道の入口付近あるいは鼓膜前で測定。最小の可聴レベル。

■ 最小可聴音場 (MAF: minimum audible field)

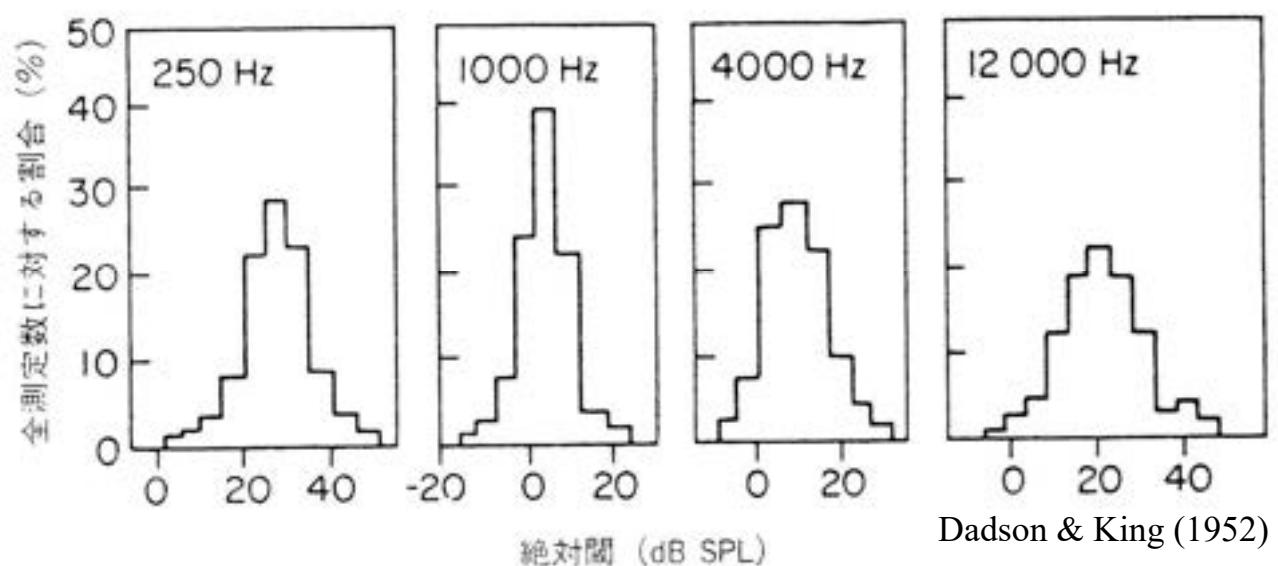
自由音場（無響室）内でスピーカーにより、測定する最小の可聴レベル。頭の中心を想定するため、「レベル」ではなく「野 or 域」という表現になる。

最小可聴音圧 (MAP)



Dadson & King (1952)

絶対閾値の分布



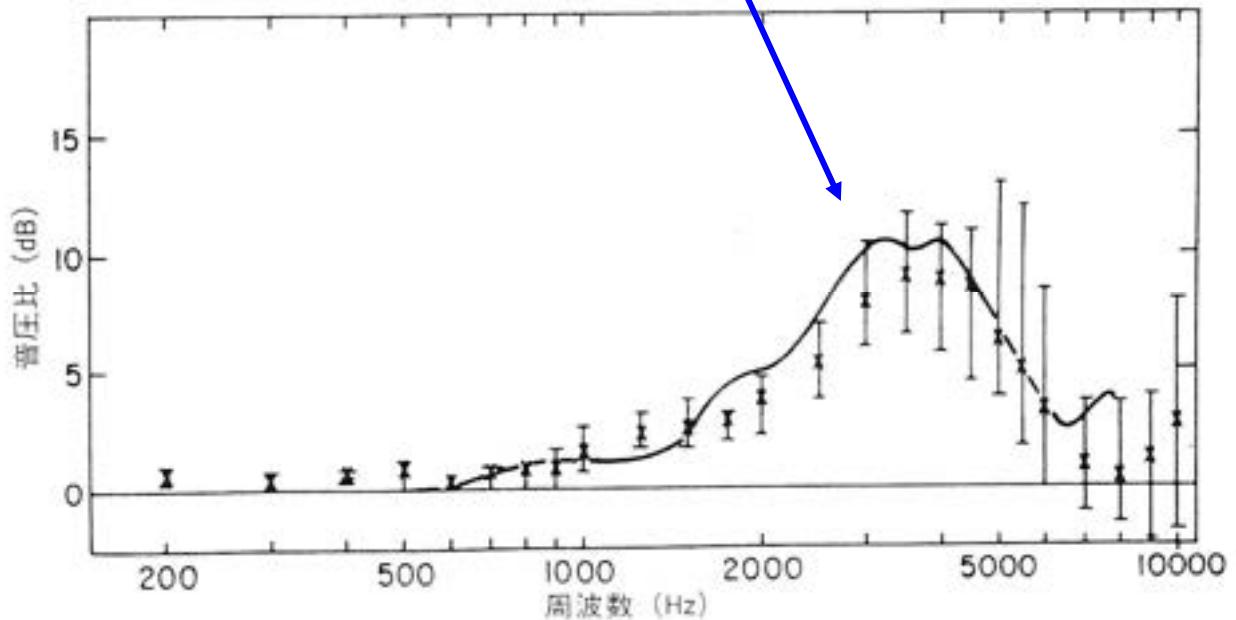
Dadson & King (1952)

4種類の周波数, 198例

周波数に関する絶対閾値の変化⇒聴力曲線

外耳道入口と鼓膜との間の音圧比

外耳による強調作用

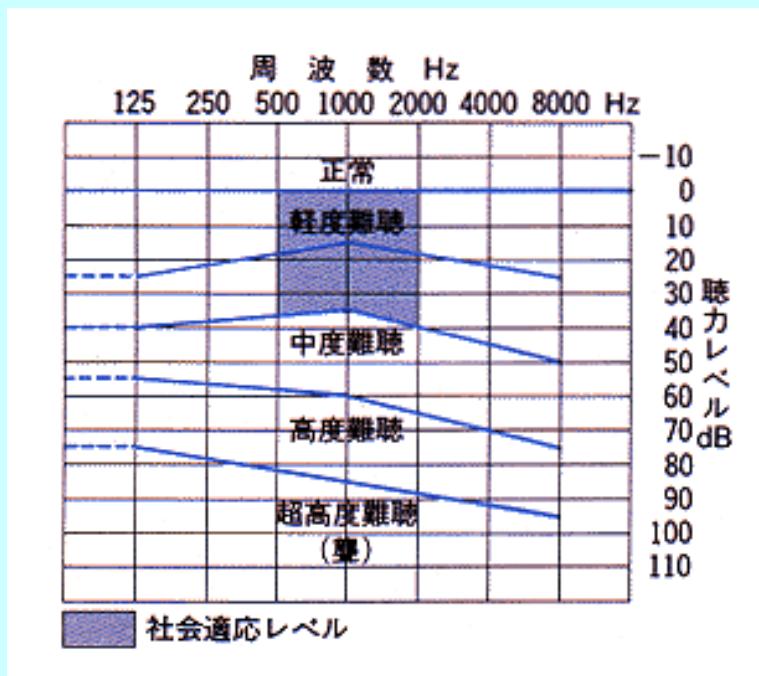


Djupesland & Zwislocki (1972)

オージオグラム

■ オージオグラム (audiogram)

聴力曲線（閾値の上昇を下向きに書いたもの）



単位

dB HL (hearing level)

ヨーロッパ、日本

dB HTL (hearing threshold level)

アメリカ

3. ラウドネス

■ ラウドネス (loudness)

音が小さい、大きいという感覚的な量感を表わす言葉

■ 単位 : sone (ソーン)

■ 周波数1 kHz, 音圧レベル40 dBの純音のラウドネス
= 1 sone

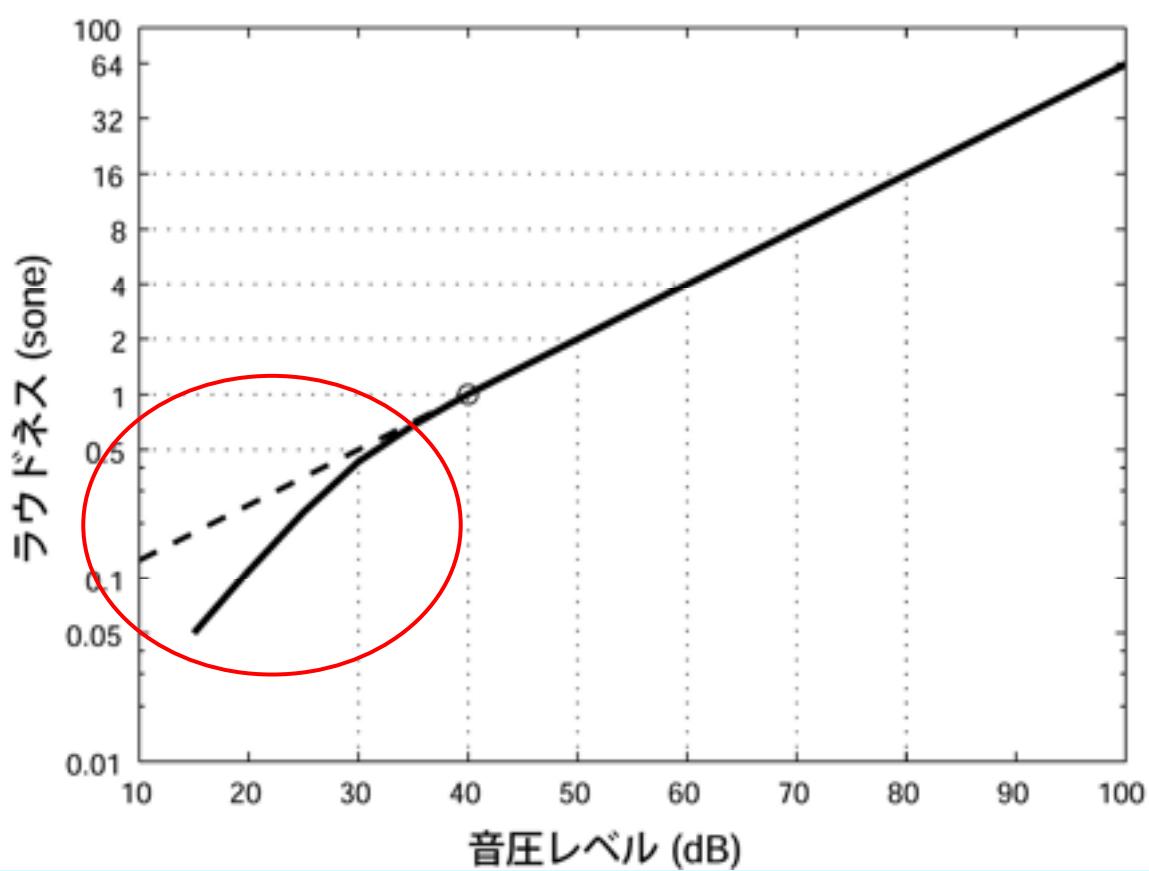
■ スティーブンスのべき乗則

ラウドネス N , 音の強さ I , 音圧 P

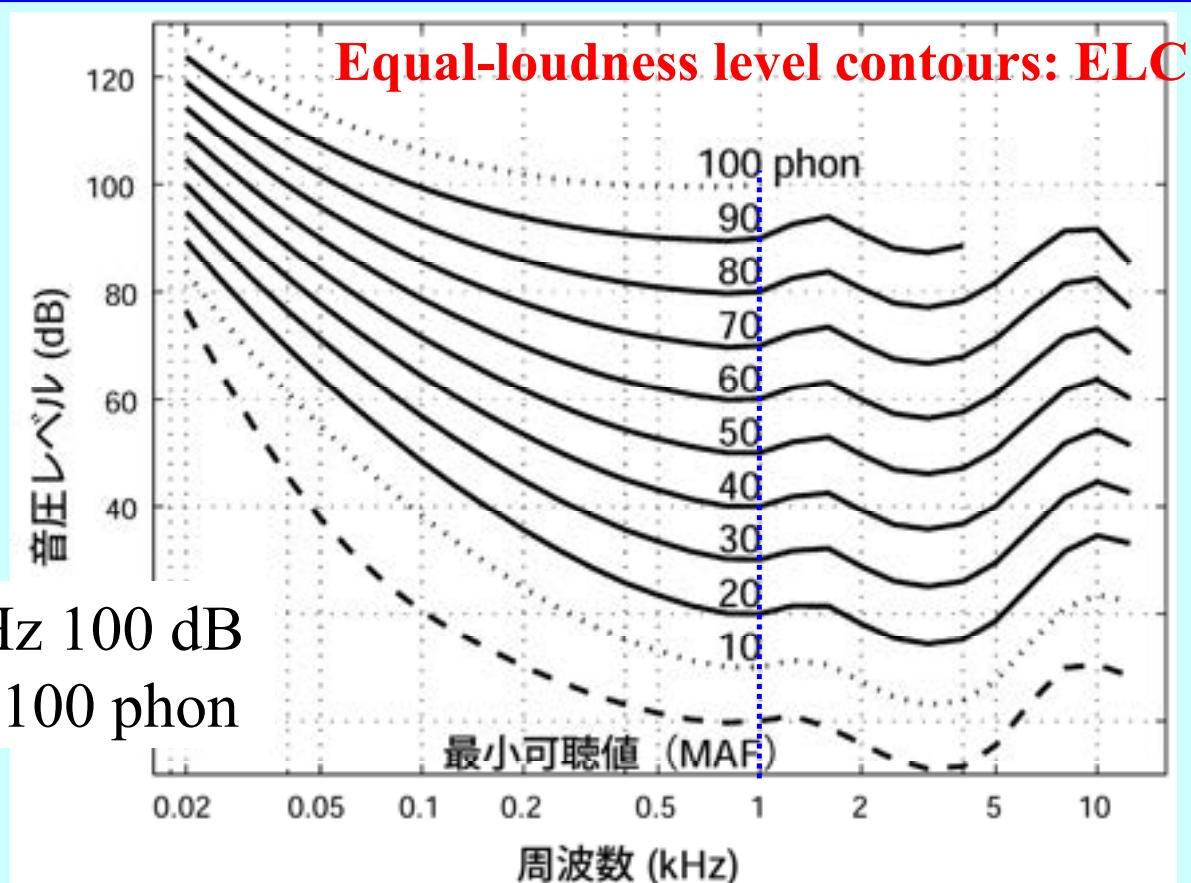
$$N = kI^\alpha = k'P^{2\alpha}$$

スティーブンスのべき乗則に基づくラウドネスの成長則

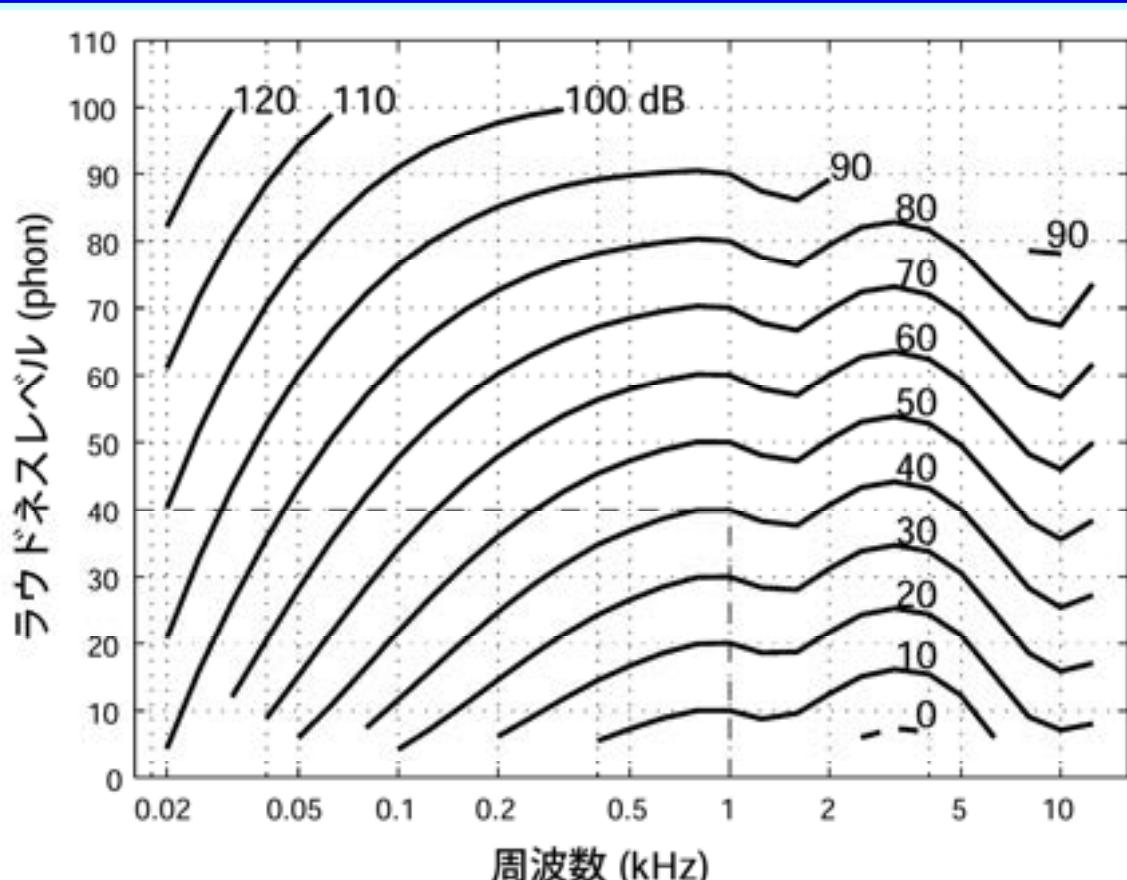
ラウドネス曲線



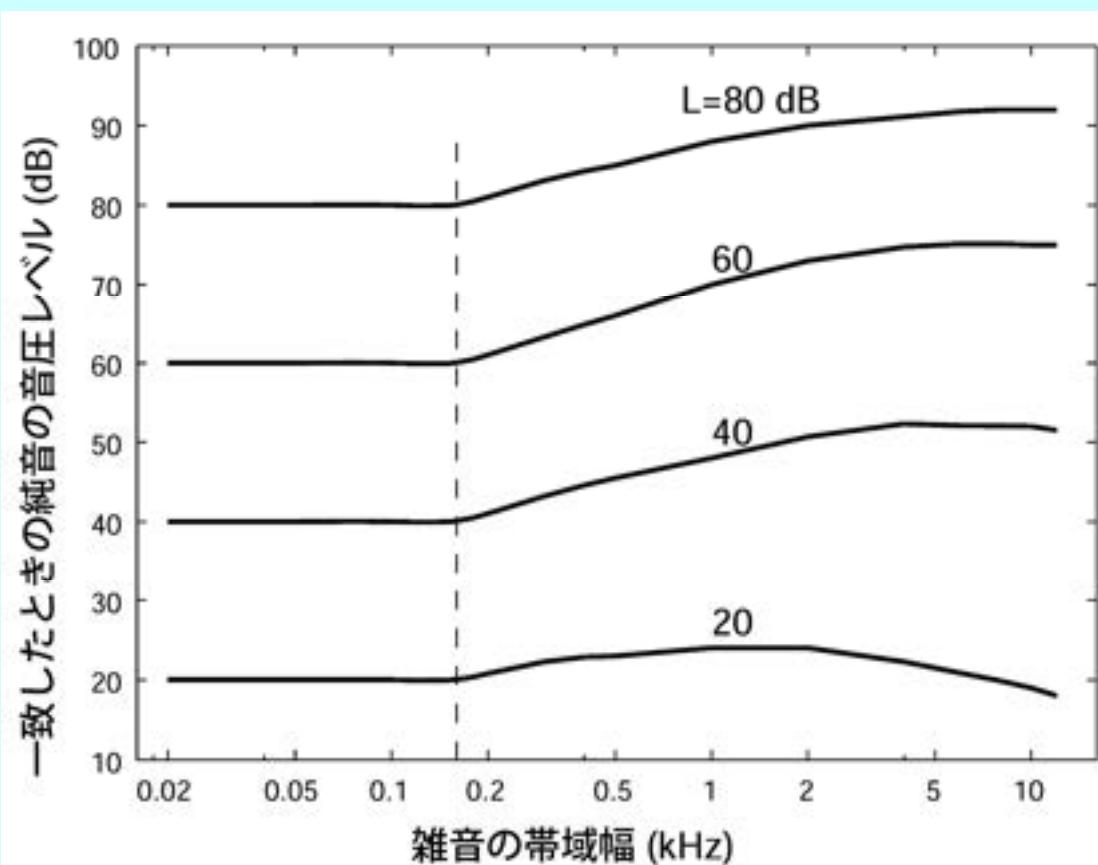
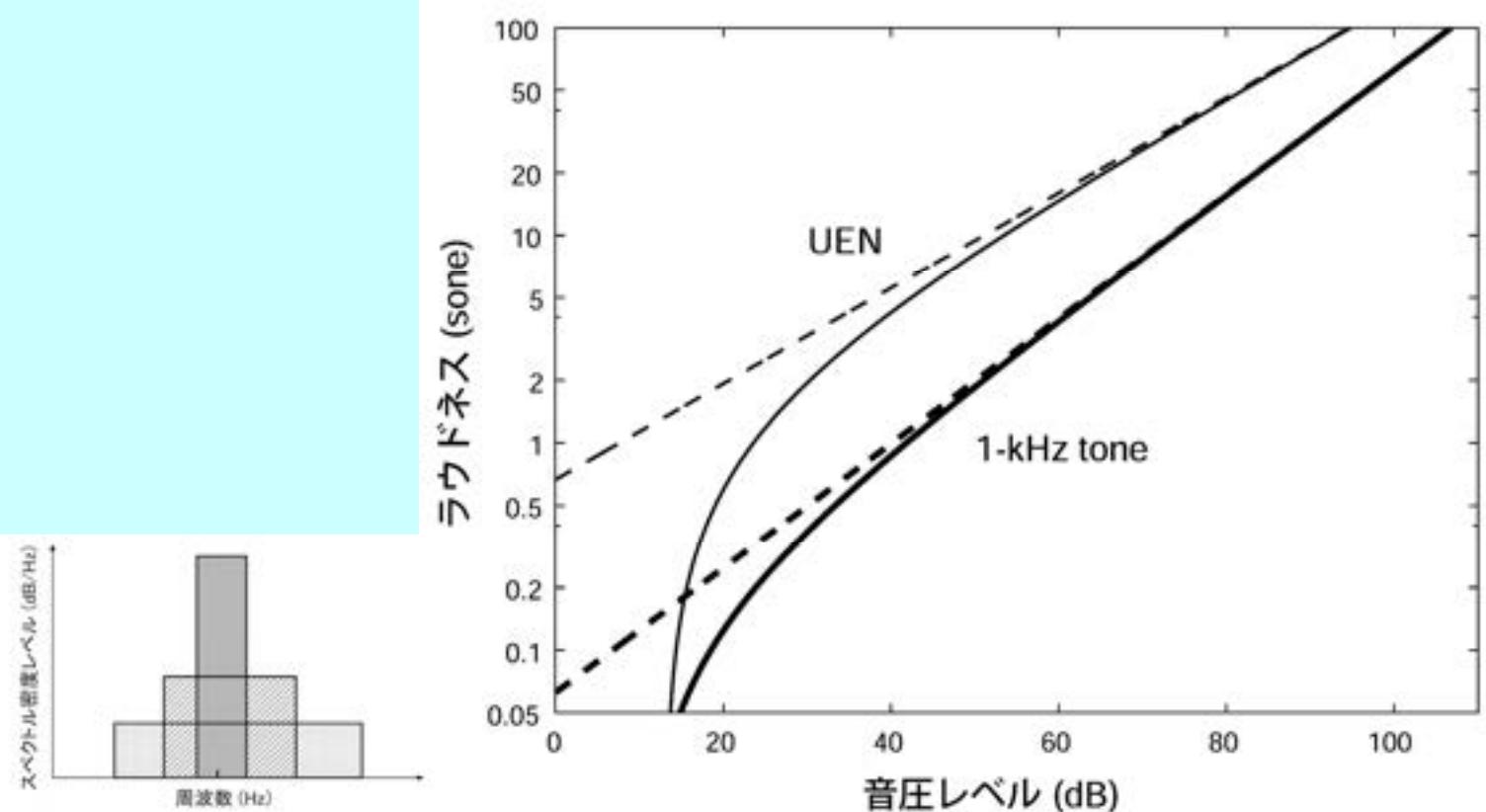
4. 音の等ラウドネスレベル曲線



音圧レベルに対するラウドネスレベル

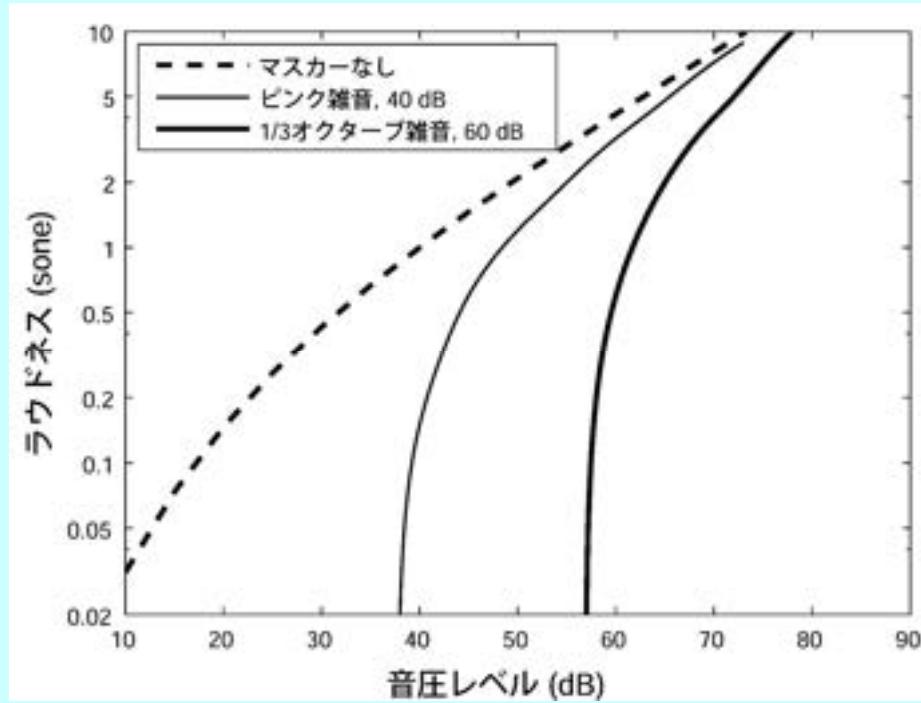


ラウドネス関数

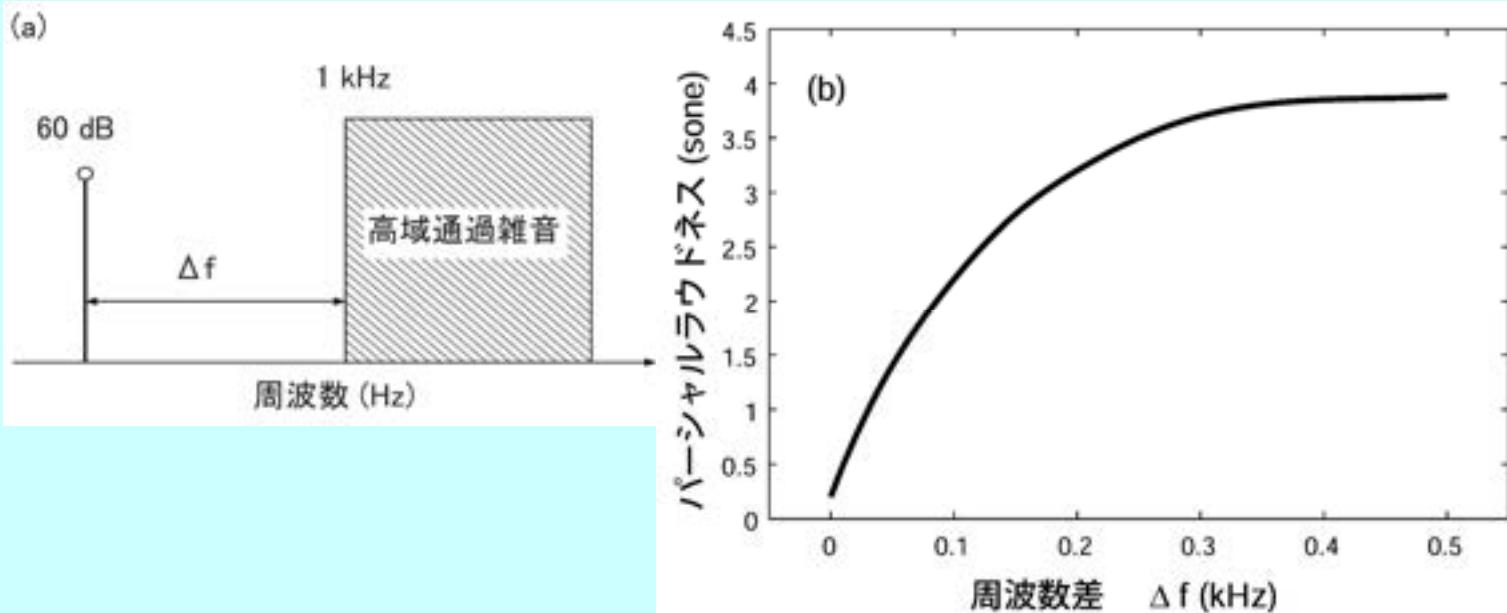


5. パーシャルラウドネス

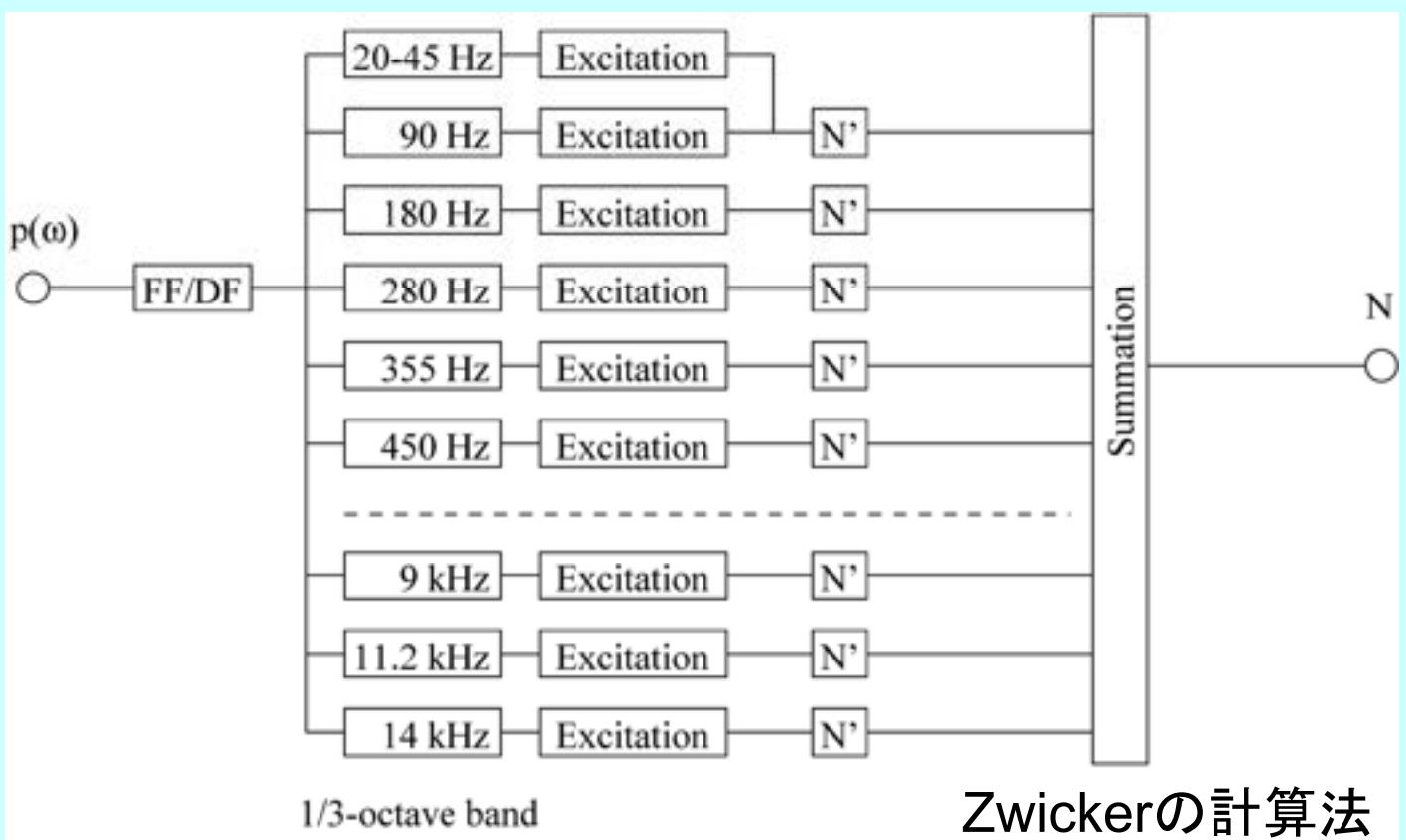
- ある音のラウドネスの知覚は、マスキング効果を及ぼす音の有無や種類によって変化する



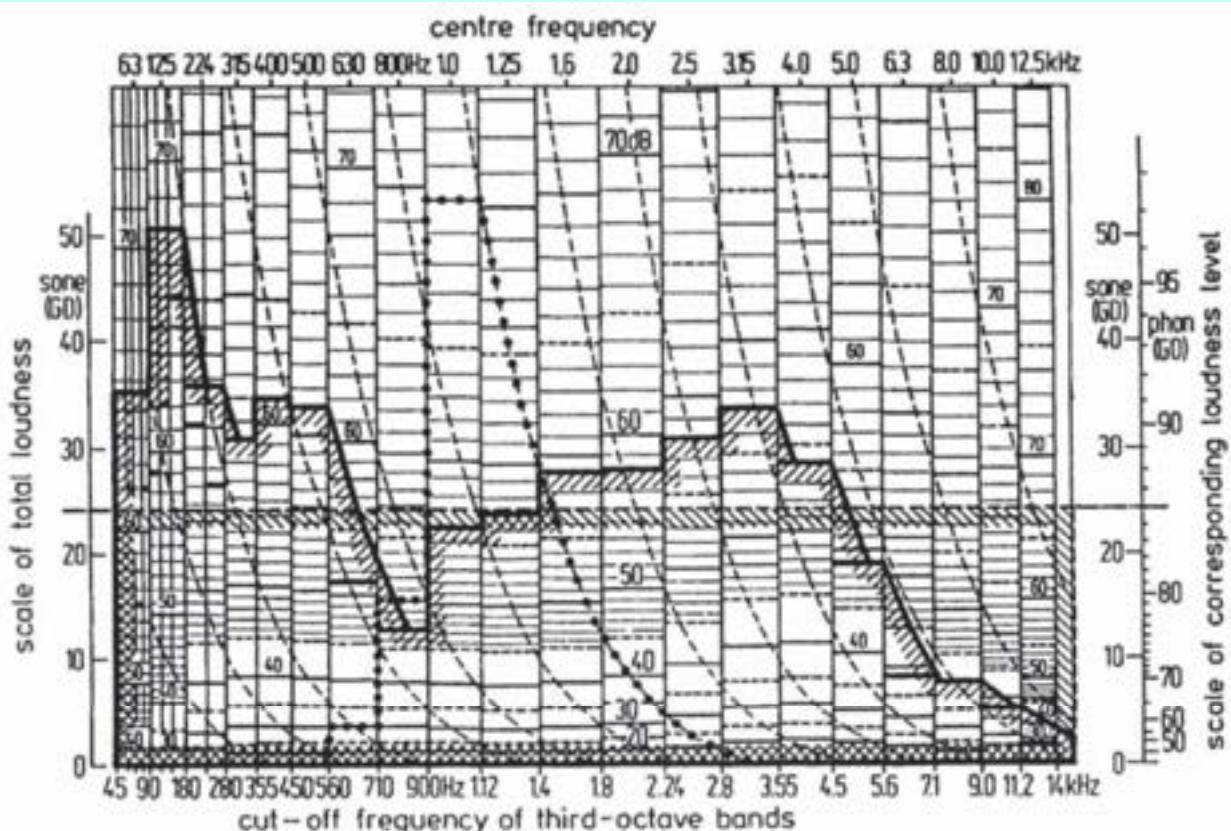
周波数スペクトル的なパーシャルラウドネス



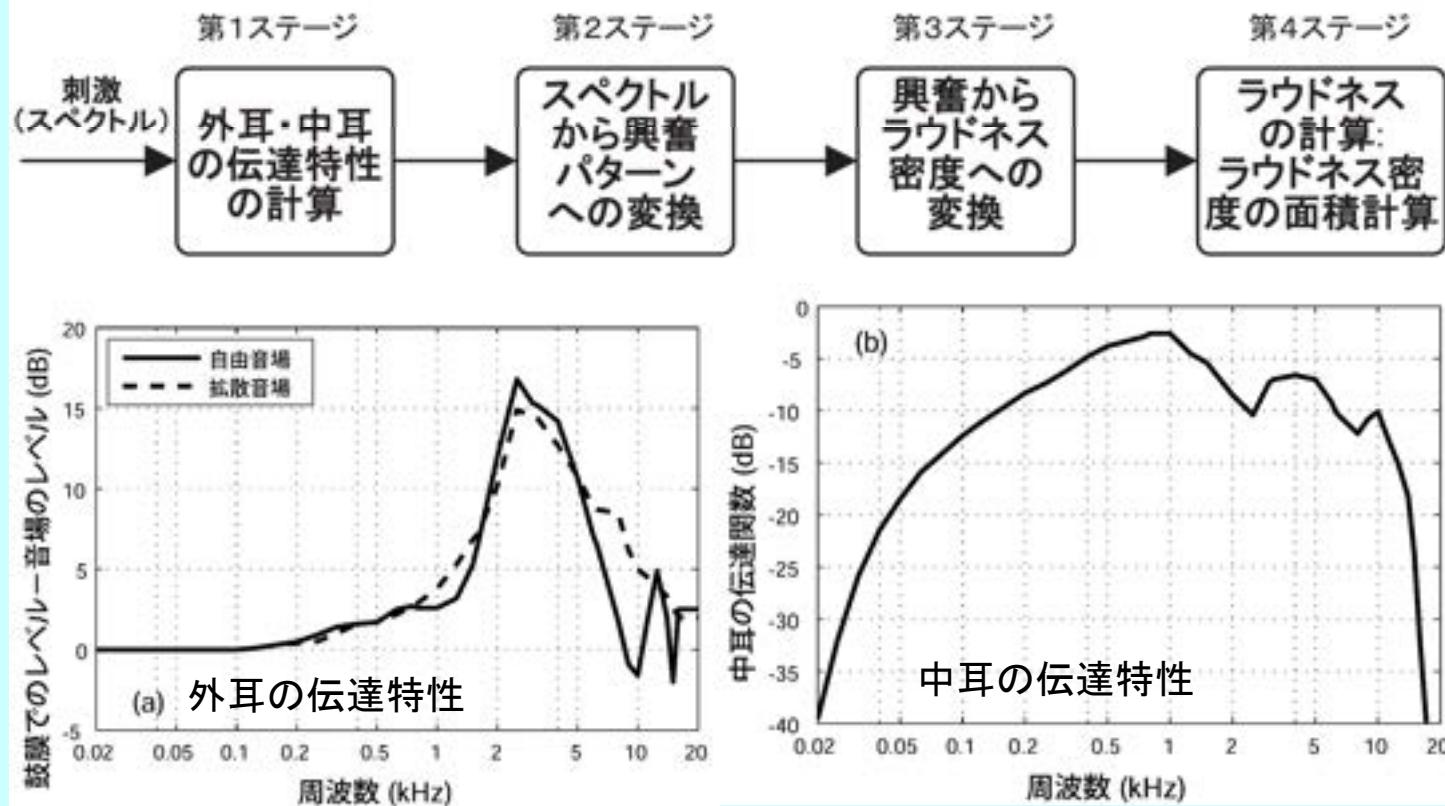
6. 定常広帯域音のラウドネスの計算



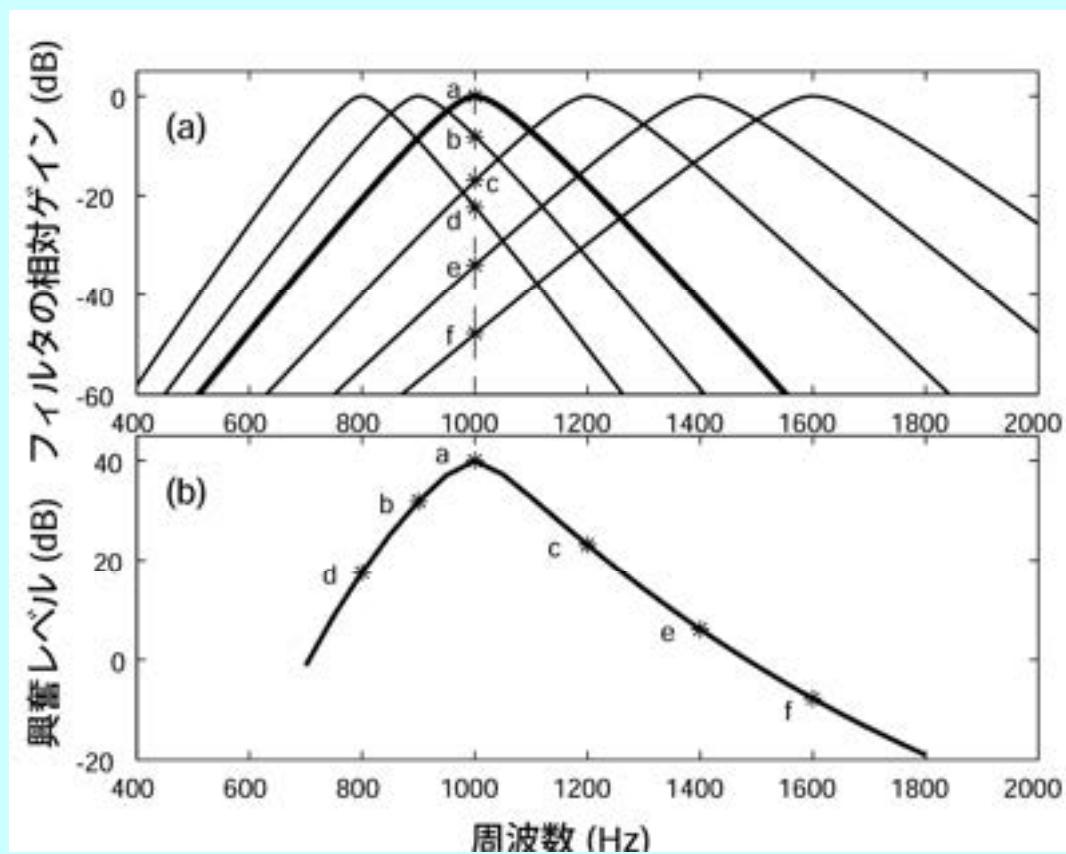
プラニメータ



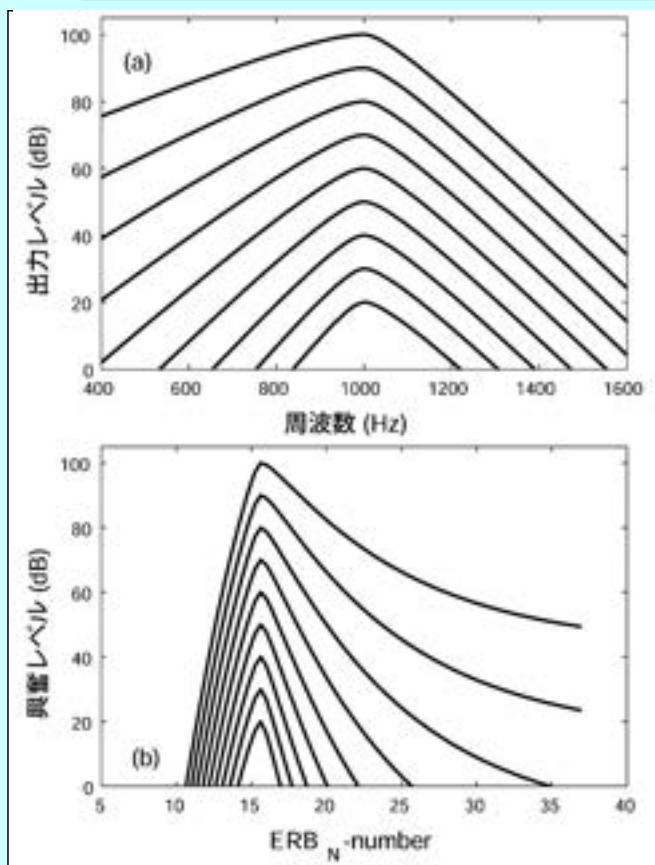
Moore & Glasberg法



興奮パターンの計算



ラウドネス密度の算出



$E < E_{THRQ}$ の場合

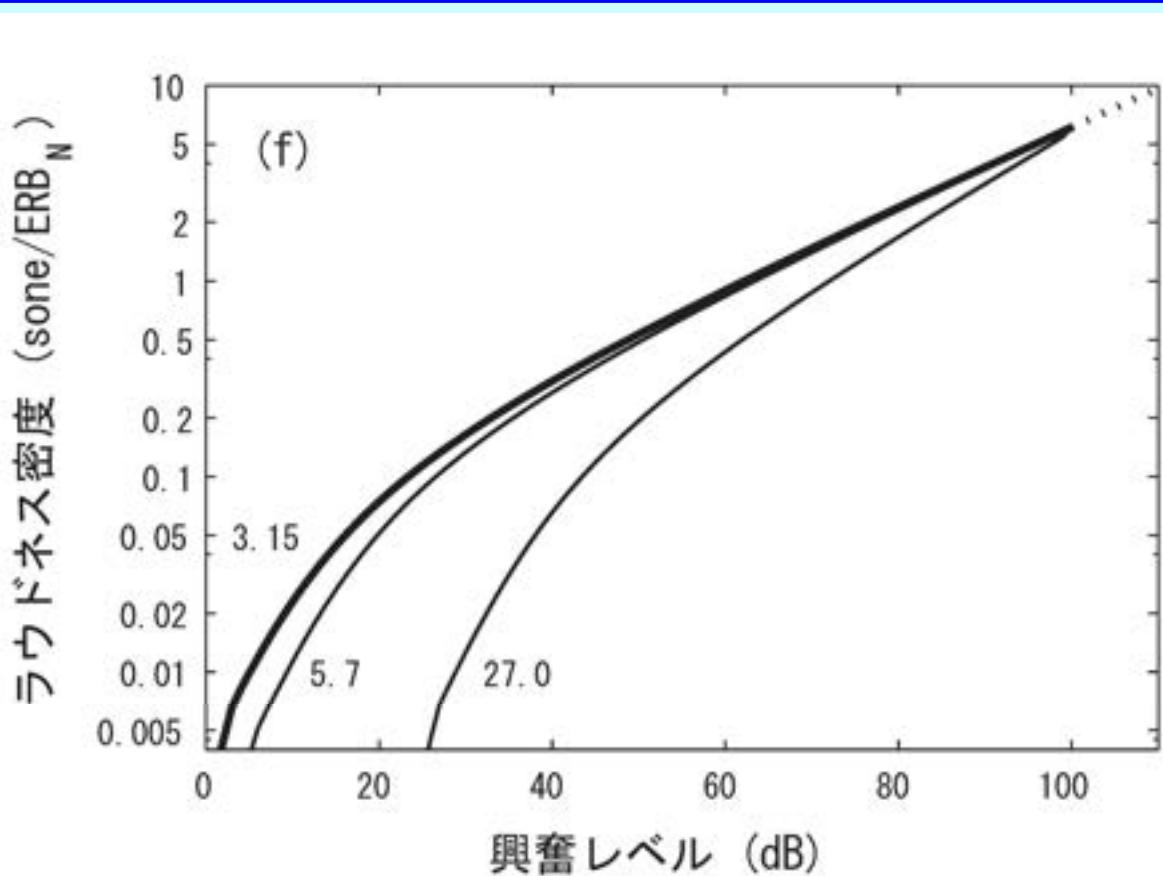
$$N' = k \left(\frac{2E}{(E + E_{THRQ})} \right)^{1.5} [(GE - A)^\alpha - A^\alpha]$$

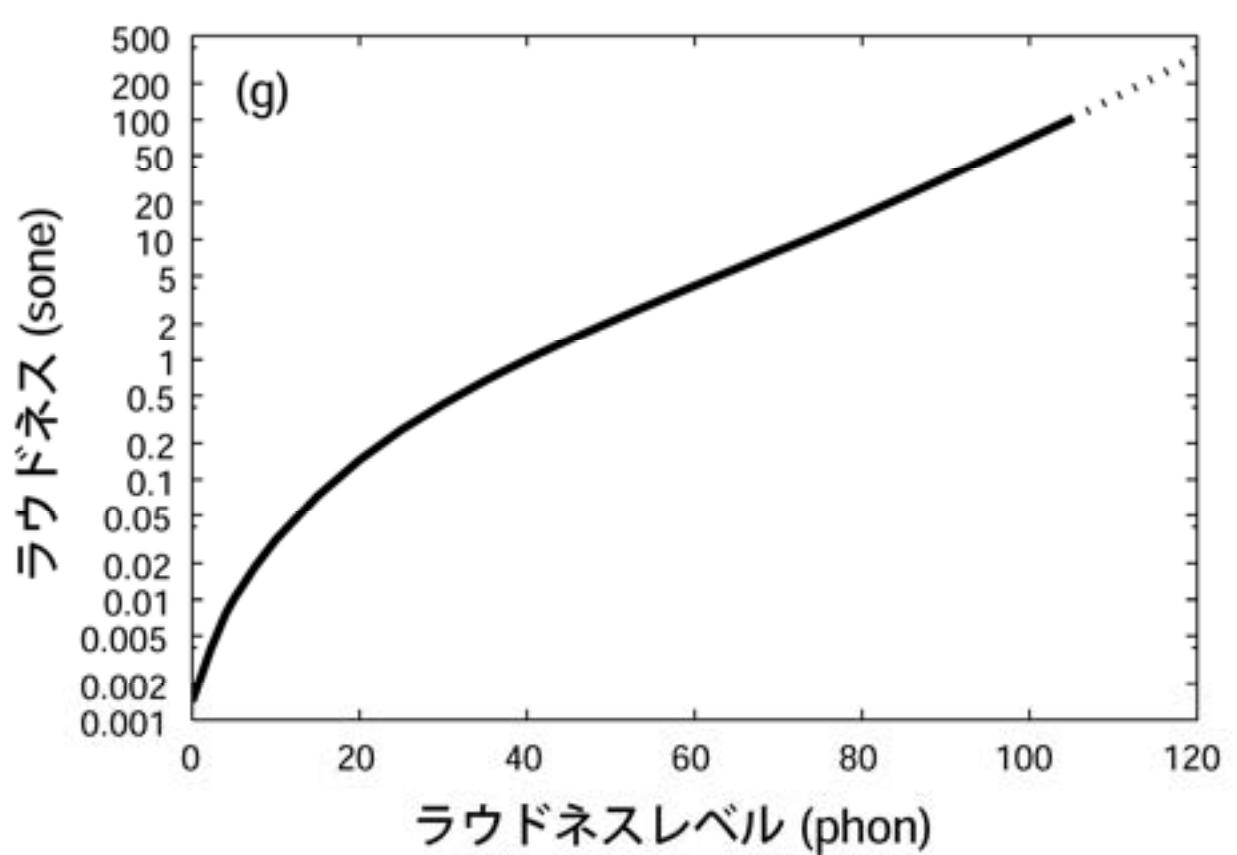
$E_{THRQ} \leq E < 10^{10}$ の場合

$$N' = k [(GE - A)^\alpha - A^\alpha]$$

$10^{10} \leq E$ の場合

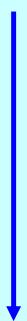
$$N' = k \left(\frac{E}{1.0707} \right)^{0.2}$$



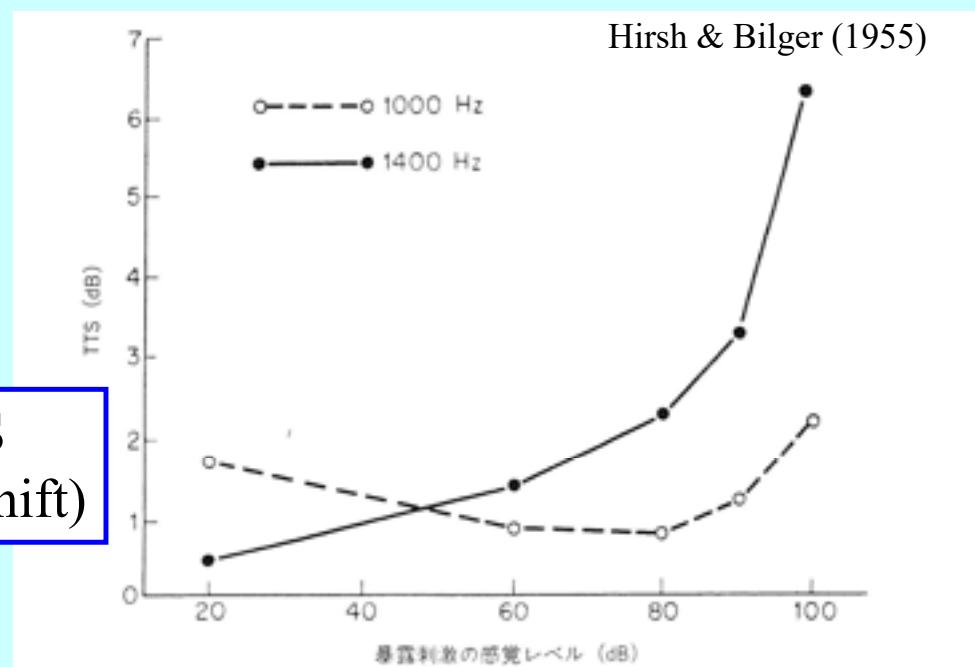


7. 音の大きさの順応、疲労

- 聴覚順応 (auditory adaptation)
- 聴覚疲労 (auditory fatigue)

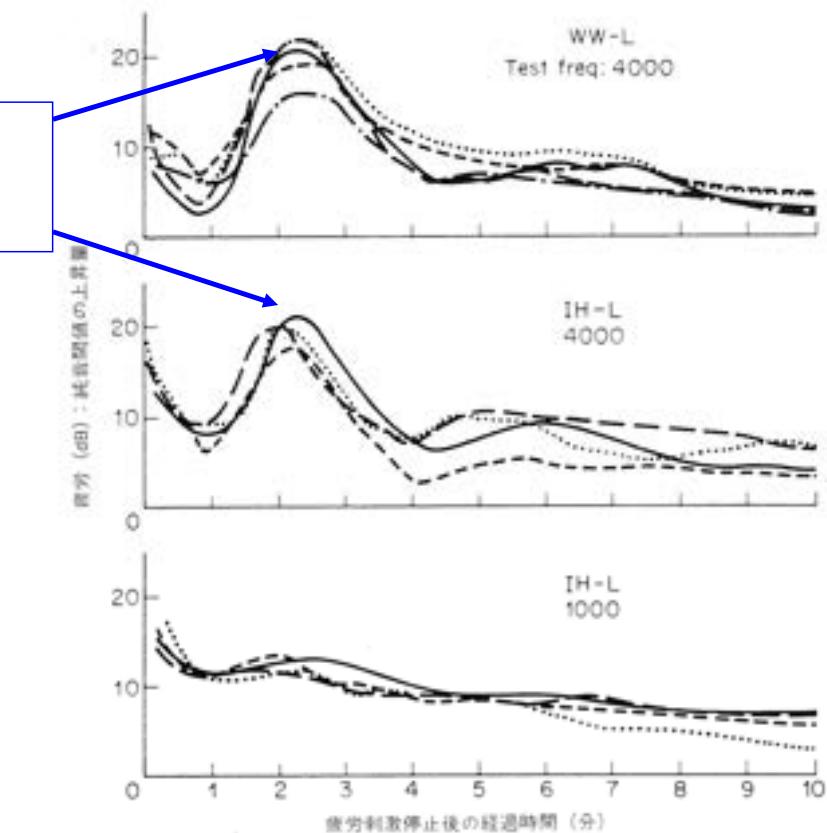


一過性閾値変動 TTS
(temporal threshold shift)



疲労刺激からの回復の度合い

跳躍
(bounce)

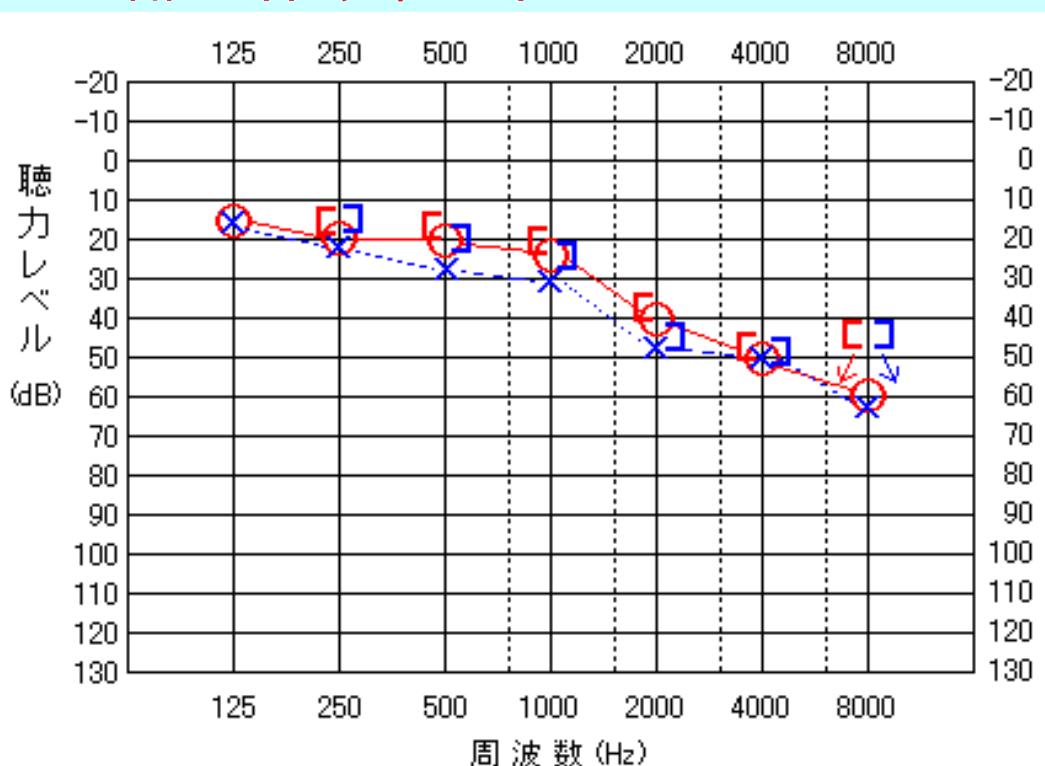


500 Hz,
120 dB
の疲労刺激
3分呈示後
の回復の度合

Hirsh & Ward (1952)

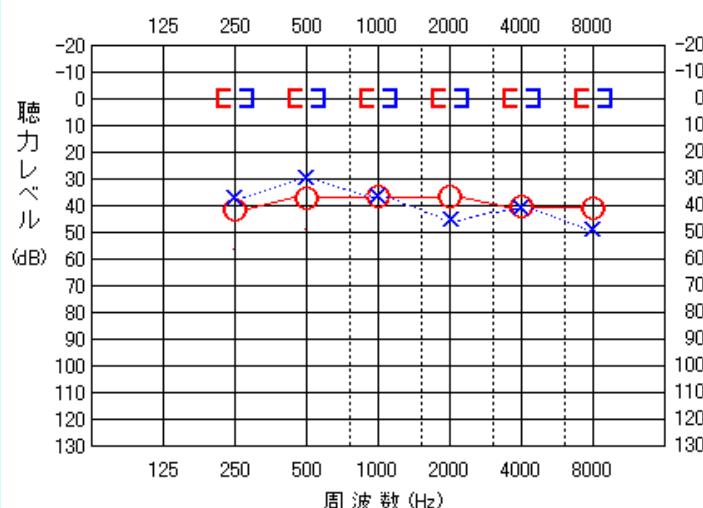
8. 聴覚障害

加齢に伴う聴力低下



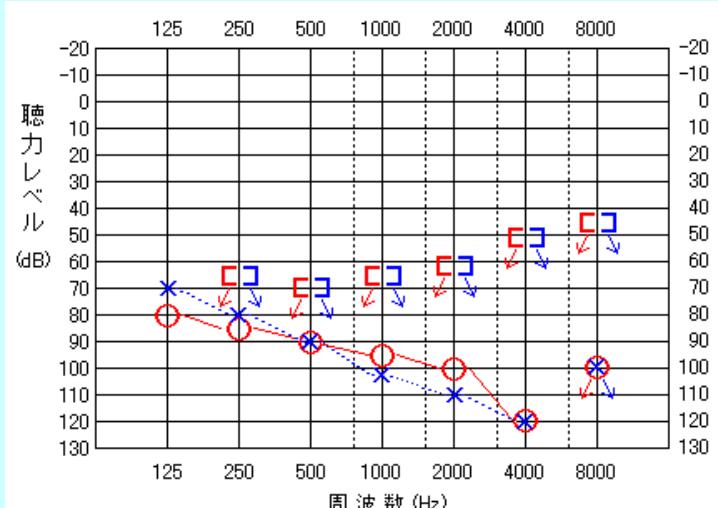
伝音性難聴

(conductive deafness)



感音性難聴

(sensorineural hearing loss)



内耳性難聴 (cochlear loss)

I485F 知覚情報処理特論

11. 音の高さの知覚とモデル

2023年5月27日(土)

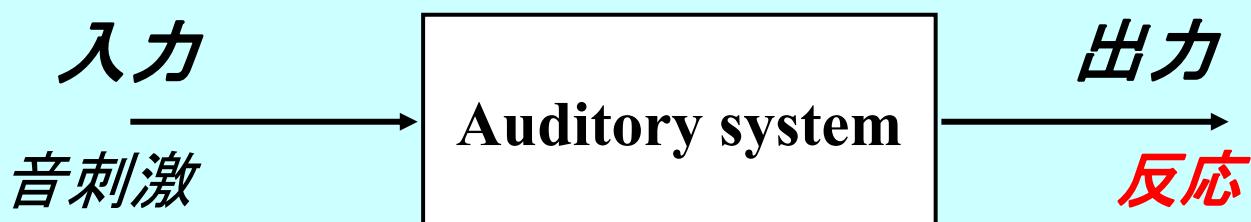
人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>
Email: unoki@jaist.ac.jp



1. 音の知覚



■ 三つの属性:

- “音の大きさ” ← “音の強さ”
- “音の高さ” ← “周期” or “調波”
- “音色” ← その他いろいろ

■ “音” の知覚 (perception) の三つの属性

- 音の大きさ ⇒ ラウドネス (loudness)

音を「小さい」から「大きい」に伸びた尺度上に順序づけられる音の強さに関する聴覚知覚の属性

- 音の高さ ⇒ ピッチ (pitch)

音を音階上に順序づけることができるような聴覚知覚の属性

- その他（音色） ⇒ 音色 (timbre)

同じラウドネスで同じピッチをもつ二つの音を異なる音であると断言できるような聴覚知覚の属性

2. 音の高さの知覚

■ 音の高さ (pitch) . . . 繰り返し速度に関係

- 純音 周波数 (frequency)

- 周期的複合音 基本周波数 (fundamental frequency)

■ 知覚に関する理論

- 場所説 (place theory) (or 周波数説)

基本周波数ないし調波に対応する場所

⇒ 興奮パターンに関係

- 時間説 (temporal theory)

神経インパルスの時間パターンに関係

3. 古典的聴覚説

■ SeebeckとOhmの論争

■ 時間説の主張

Seebeck

衝撃波（インパルス波）の発生間隔に応じた音の高さの知覚実験

■ 周波数説（場所説）の主張 Ohm

「 f Hzの高さを聞く時、刺激音の中に正弦波成分 $\sin 2\pi ft$ が必ず存在する」（オームの法則）
→ オームの純音の定義

差音説

■ ヘルムホルツによるオームの法則の拡張

「複合音は基本周波数成分とその高調波成分とに分解され、その基本周波数に応じた高さを有する」
(オームの音響法則)

■ 差音説（耳性ひずみ説）

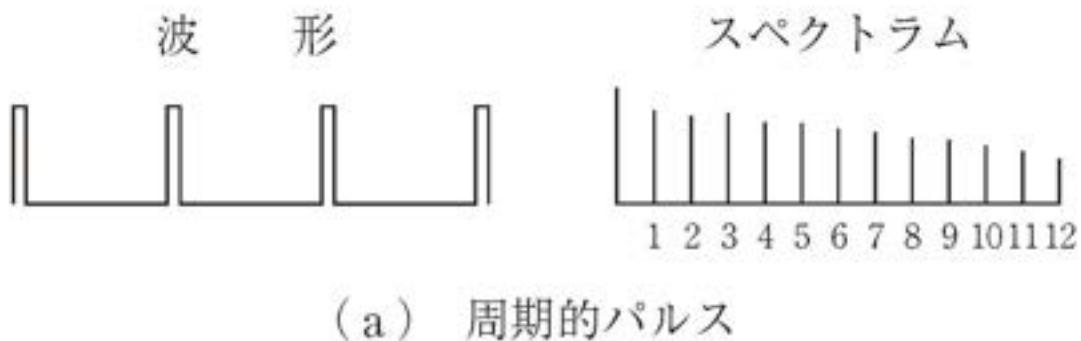
基本周波数成分を取り除いても音の高さは変わらない
⇒ 聴覚系の非線形性によるひずみ波成分（差音）の発生とその知覚で説明

⇒ 結合音と耳性内ひずみ説

4. 微細構造説

- 存在しない基本周波数成分音の知覚 (missing fundamental)
- 「低次の部分音は個別的に知覚され、それらは別々に聞くときの高さと同じ高さを有する。これに対し、高次の部分音は個別的に知覚できず、ひとつの成分音として一体的に知覚される。“residue”と名づけられたその主観的成分音は混じりあった鋭い音質を有し、その高さは一体化された波形の周期で決まる」
⇒ “residue pitch”
- ほとんどエネルギーを含まない周波数に対応した高さ
low pitch, periodicity pitch, time-separation pitch, repetition pitch

Missing fundamental



振幅変調音（AM音）

- Schoutenの研究 periodicity theoryの実験
- 振幅変調音（AM音）
 - 周波数 f の正弦波（搬送波）
 - 周波数 g の正弦波（振幅変調）
⇒ $f-g, f, f+g$ の周波数成分を有する複合音（振幅変調音）の高さ（low pitch）
(例) $f=1000$ Hz, $g=200$ Hz のとき（差音説）,
 $f-g=800, 1000, 1200$ Hz, 高さは 200 Hz
 - (例) $f=1050$ Hz, $g=200$ Hz のとき,
 $f-g=850, 1050, 1250$ Hz,
高さは 1/210 s のパルスの高さと一致

■ ピッチマッチング

調べる複合音の low pitch を高調波関係にある複合音の基本周波数で測ること

■ ピッチシフト

高調波関係にある複合音の low pitch とその部分音を一律にシフトさせて作った高調波関係にない複合音との差のこと

■ Low pitch 知覚

基本周波数成分は重要な役割を果たさない
(Small & Campbell, Patterson など)

■ de Boerの研究

■ Pseudo-period theory

高調波関係にない複合音の時間波形の山と山の時間に着目

■ Pseudo-fundamental theory

高調波関係にある複合音の基本周波数に着目

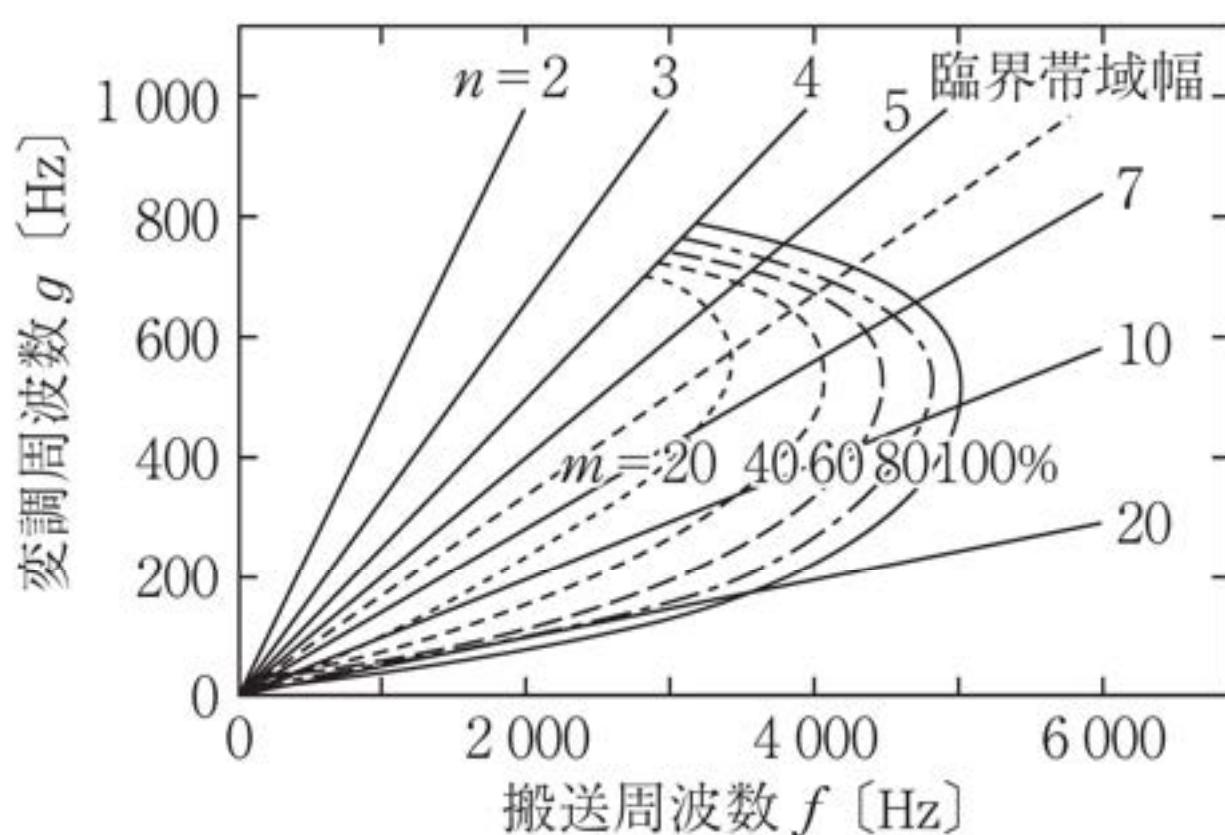
■ Low pitch知覚

■ 基本周波数は重要ではない

■ 純音のピッチのように明瞭ではない

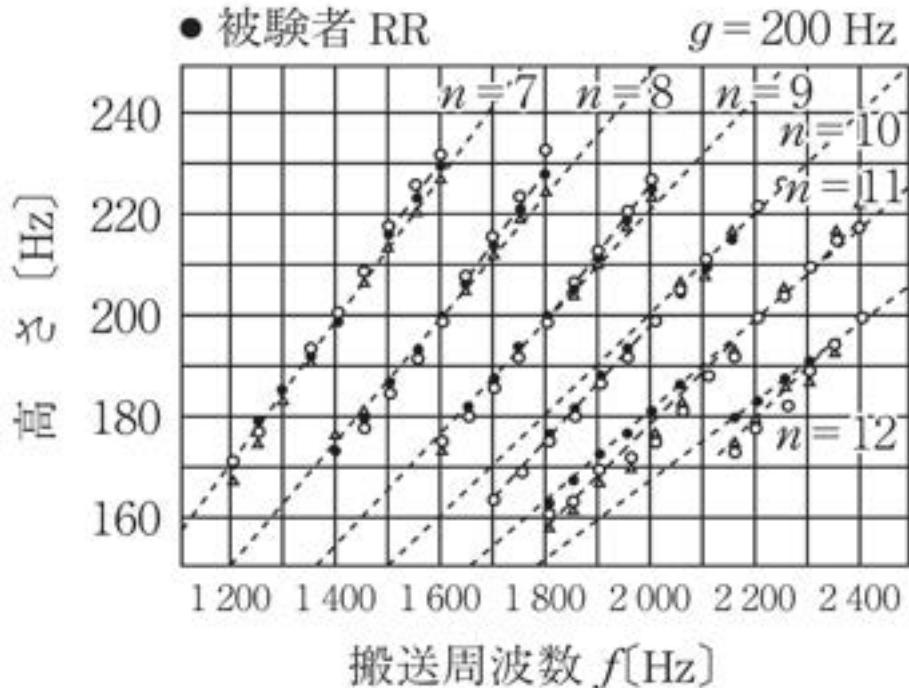
■ ピッチマッチングは難しい

Low pitch知覚



AM音の高さ

- 被験者 H
- △ 被験者 FL
- 被験者 RR



- * 搬送周波数 f
 - * 変調周波数 g
 - * f と g の関係
- $$f = ng$$
- $$f = ng + \Delta f$$

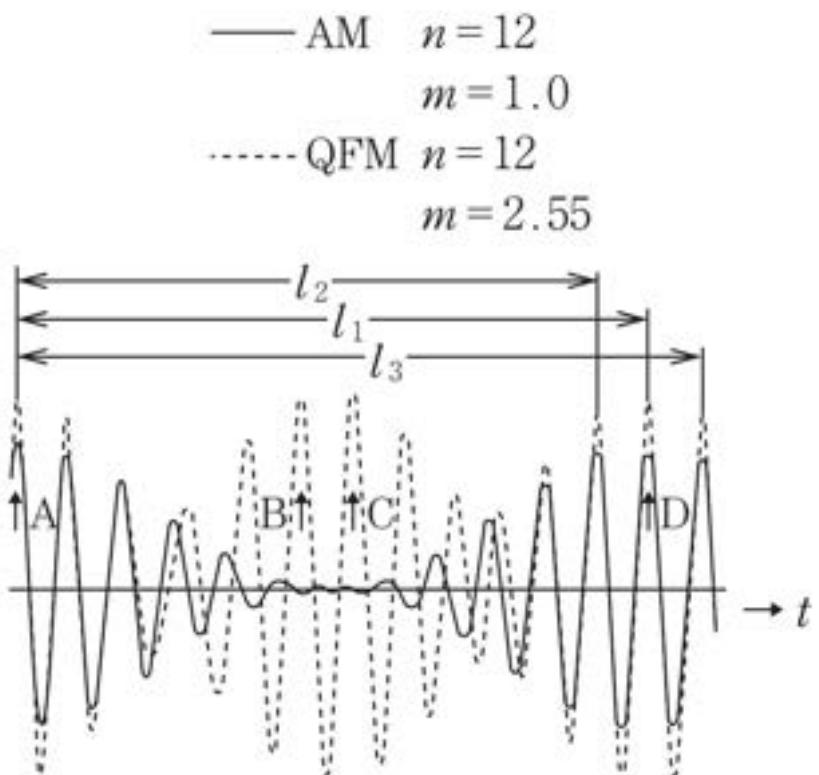
微細構造説

■ 微細構造説

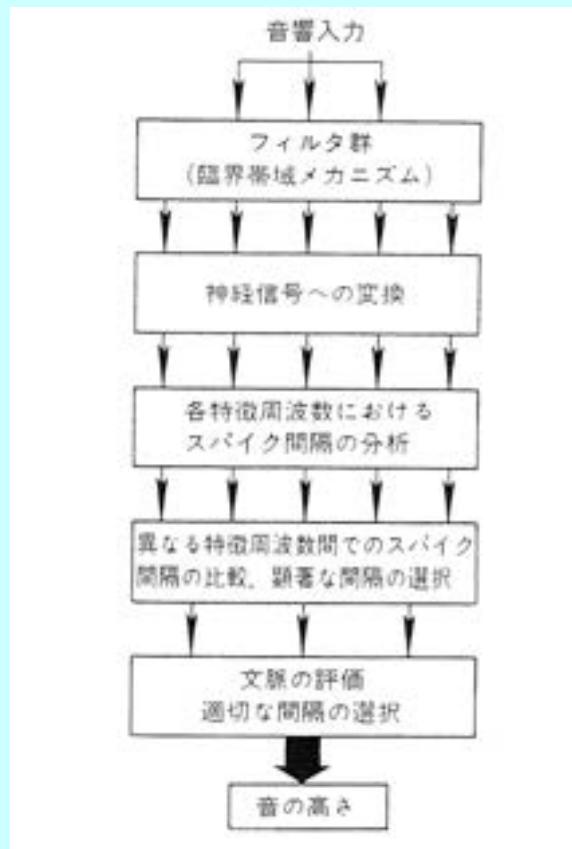
時間波形の山と山との間の時間がlow pitchを規定するという説

■ 振幅変調音 (AM)

■ 擬似周波数変調音 (QFM)

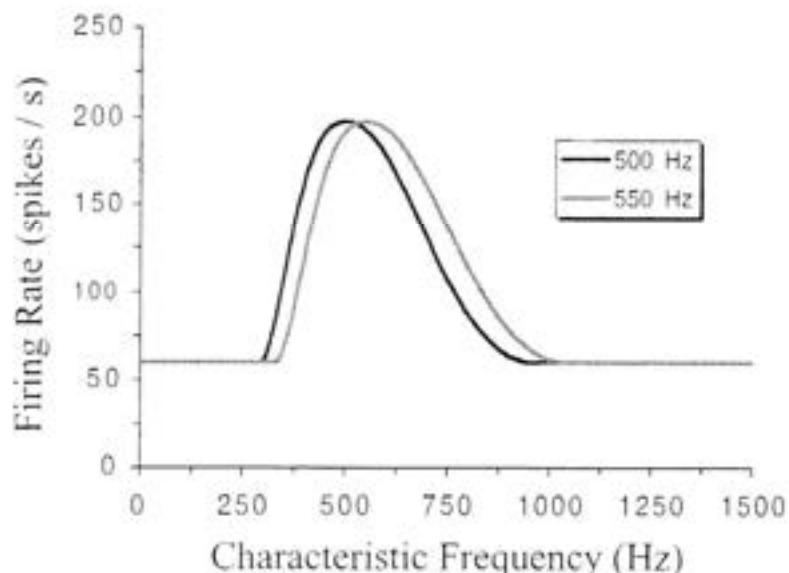


5. ピッチ知覚の図式的モデル

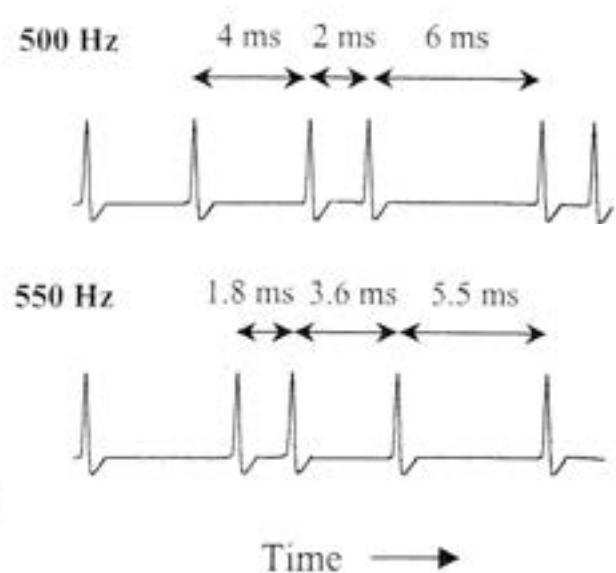


神経発火パターン

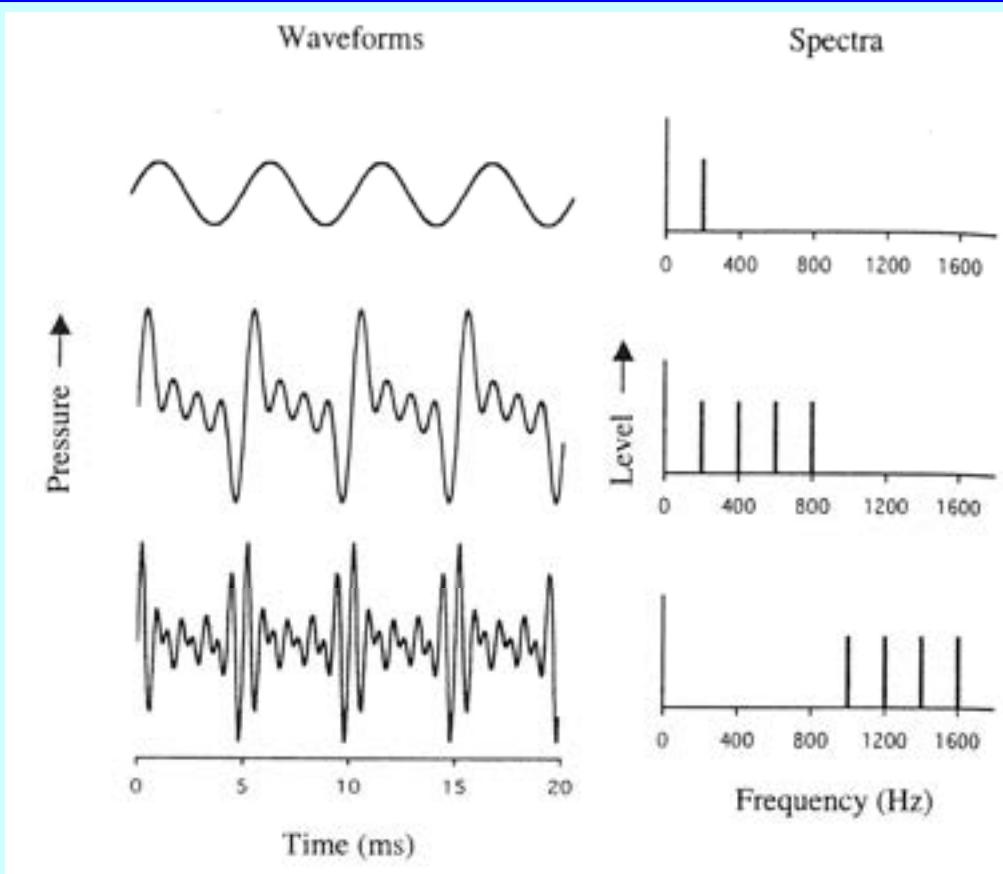
Rate-Place Coding



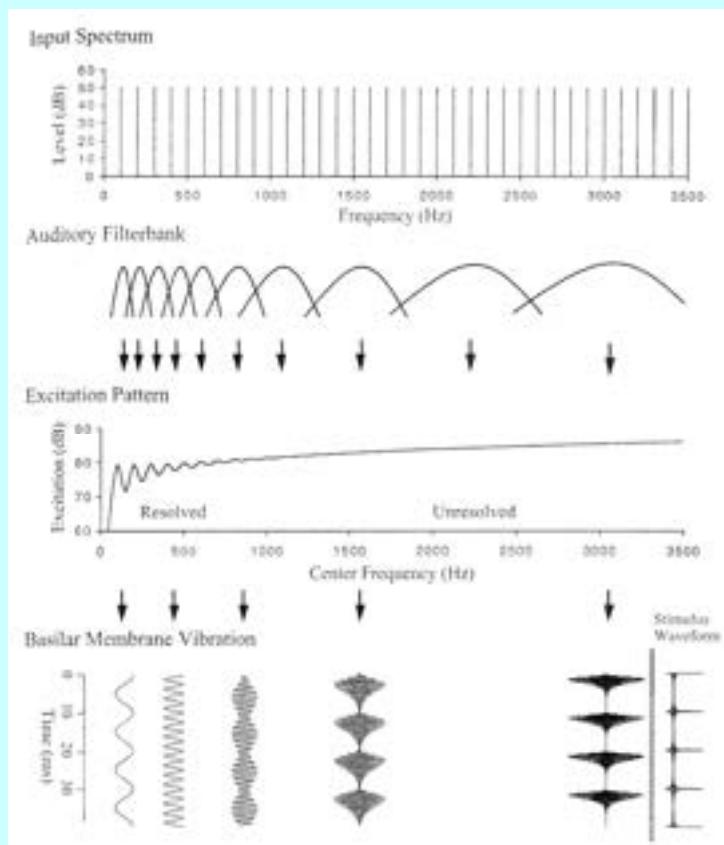
Temporal Coding



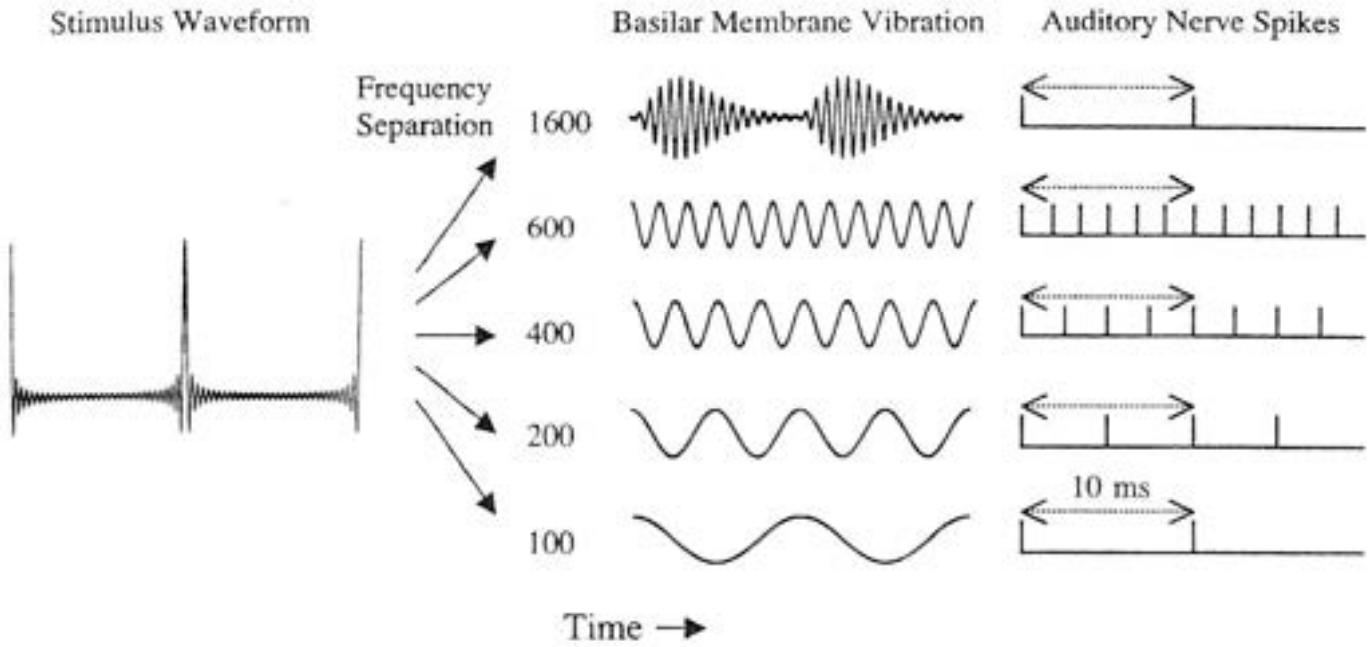
同じピッチをもつ波形



複合音に対する表現



基底膜振動と位相固定



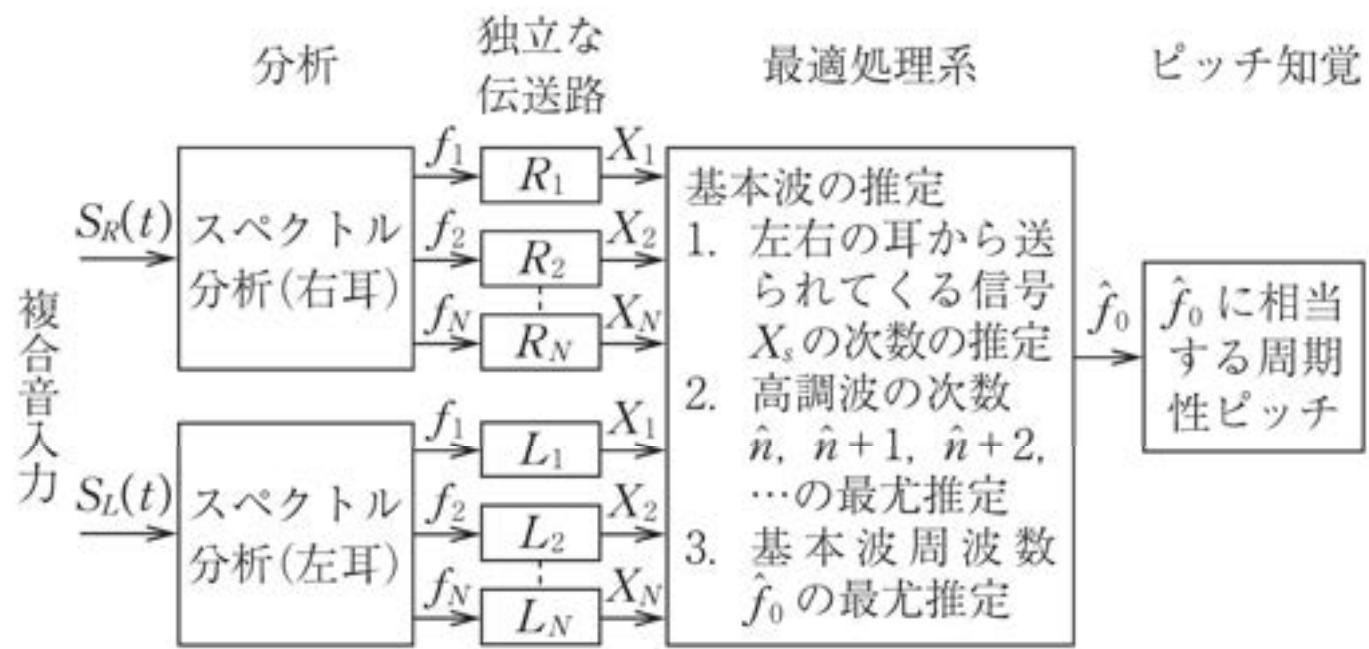
6. 代表的なモデル

■ WigghtmanのPattern transformation model

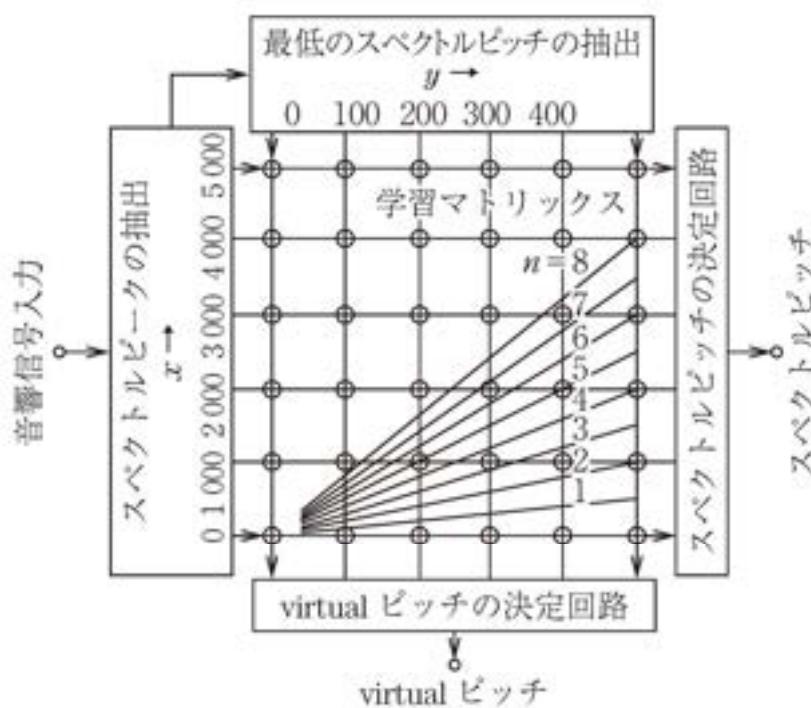


- 聴覚末梢系を模擬した周波数分解
- 神経回路で構成されたフーリエ変換器
- 出力波形のピーク位置を検出する処理

■ Goldstein の optimum processor model



■ Terhardtのvirtual pitch theory



I485F 知覚情報処理特論

12. 音色の知覚とモデル

2023年5月27日(土)

人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

Email: unoki@jaist.ac.jp



1. 音の知覚



■ 三つの属性:

- “音の大きさ” ← “音の強さ”
- “音の高さ” ← “周期” or “調波”
- “音色” ← その他いろいろ

■ “音” の知覚 (perception) の三つの属性

■ 音の大きさ ⇒ ラウドネス (loudness)

音を「小さい」から「大きい」に伸びた尺度上に順序づけられる音の強さに関する聴覚知覚の属性

■ 音の高さ ⇒ ピッチ (pitch)

音を音階上に順序づけることができるような聴覚知覚の属性

■ その他 (音色) ⇒ 音色 (timbre)

同じラウドネスで同じピッチをもつ二つの音を異なる音であると断言できるような聴覚知覚の属性

2. 音色

■ 音色 (ねいろ／おんしょく)

timbre／tone color

■ ASA (American Standards Association) での定義 timbreはもともとフランス語、英語では日常的でない

■ 日本工業規格 (Japanese Industrial Standards, JIS)

「聴覚に関する音の属性の一つで、物理的に異なる二つの音が、例え同じ音の大きさおよび高さであっても異なった感じに聞こえるとき、その相違に対応する属性」

■ ヘルムホルツの “On the sensation of tone”

「バイオリン、フルート、クラリネットおよび歌声が同じ音符を同じ高さで演奏されるとき、バイオリンの音を他のものと区別する特性を音色と呼ぶ」

■ これまでの定義に対する異論

- 音の大きさ、高さの異なる音色を説明できない
- 音の大きさ、高さの異なる場合でも楽器の違いは判断可能

⇒再定義の必要性

■ Pratt & Doakの定義

「音色とは、音の大きさ、高さ、持続時間以外の、何らかの判断基準を用いて、二つの音を違うと判断できる感覚の性質のこと」

⇒ 「音響心理学的にも比較的よくわかっている三つの性質（大きさ、高さ、持続時間）を取り除いた残りを「音色」という「ゴミ箱」に投げ込んでいる

音色の定義が難しいのは、大きさ、高さと異なり、物理量と直接対応付けができていないため

3. 音色に関する物理量

- 音色に対応する（と考えられている）物理量
⇒ 音の大きさ、高さほどわかっているわけではない
- 物理量
 - 周波数スペクトル
(パワー、振幅スペクトル、位相スペクトル)
 - 立上り
 - 減衰特性
 - 定常部の変動
 - 成分音の調波・非調波関係
 - ノイズ成分の有無

4. 音色の特徴

■ 印象的側面

音の印象を形容詞で表現すること

「明るい音」、「暗い音」、「澄んだ音」、「濁った音」，
「迫力ある音」，「もの足りない音」，「しっとりした
音」，「乾いた音」

■ 意識的側面

何の音であるかを聞き分けること

- * 人やものが変わってもその音を識別できる
- * 「～らしさ」の認識 (音の不变性)

5. 擬音語

■ 擬音語

- * 音を言葉で直接的に表現したもの
(例) 「ドーン」, 「ワンワン」など
- * オノマトペ (onomatopoeia) のひとつ

■ 擬態語や擬情語

動作の状態・感覚・心理状態を語音によって象徴的に表わしたもの

- (例) 「サラサラ」, 「スッキリ」など

6. 音色と音質

■ 音質 (sound quality)

- 音色とほとんど同義と思ってよい
- 音質の言葉が新しい

■ 音色と音質の違い

- 音質は価値判断（音のよさ, 悪さ）を含む
- 音色は価値判断を含まない（ニュートラル）

7. 音色の評価手法

■ 心理物理的測定法

恒常法, 極限法, 調整法, 適応法など

■ 心理物理的尺度構成法

音色, 音質の評価で多用される主観評価手法

■ 多次元尺度構成法

対象間の距離関係をもとに空間上の布置を求める方法

■ 多変量解析

重回帰分析, 主成分分析, 因子分析, クラスタ分析

■ SD (semantic differential) 法

形容詞尺度を利用して印象を科学的に捉える手法

心理物理的測定法

■ 恒常法

弁別限（同種の刺激の違いがわかる最小の差）を求めるときに利用される方法。呈示順序はランダム。

■ 極限法

刺激強度を一方向に変化させながら、被験者が判断していく方法。絶対閾値の測定に利用。

■ 調整法

被験者が刺激の強度を直接制御可能な方法。

■ 適応法

被験者の判断に応じて、課題の難易度を変化させて求める方法。

心理物理的尺度構成法

■ 音色、音質の評価で多用される主観評価手法

■ 尺度の4水準

- 名義尺度 クラス、符号などを当てはめるもの
- 順序尺度 数字の順位性、大小関係に意味あり
- 距離尺度 順序だけでなく相互の間隔に意味あり
- 比率尺度 絶対零点が存在し、すべての数学的操作が可能な尺度（例：長さ、重さなど）

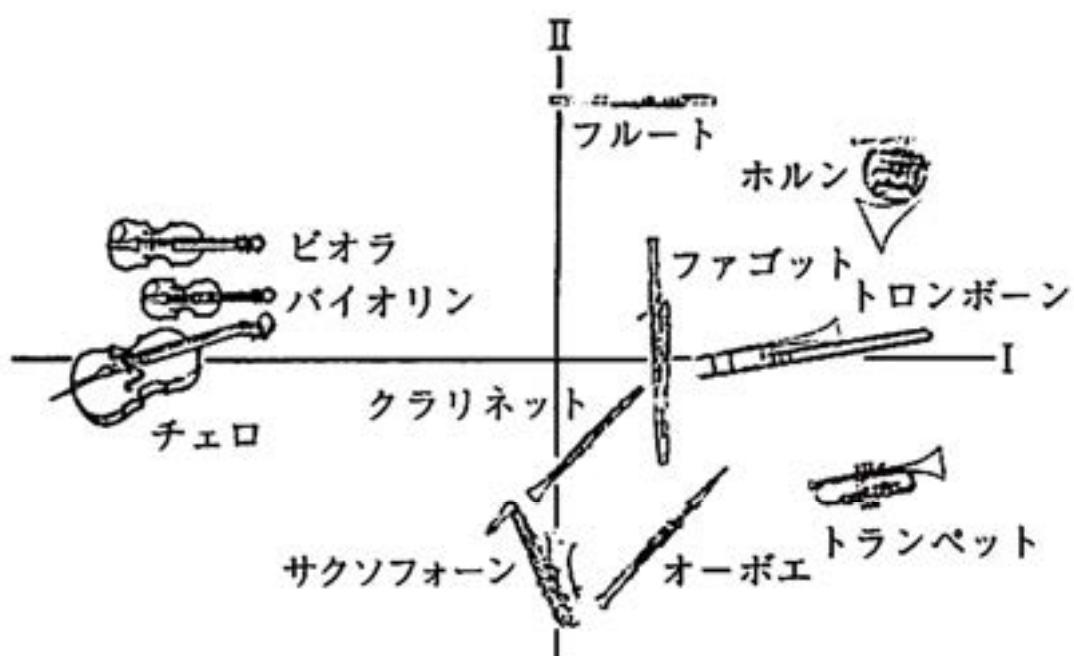
■ 評定尺度法 絶対判断

■ 一対比較法 相対判断 （順序効果に注意）

■ マグニチュード推定法

多次元尺度構成法（MDS）

■ 対象間の距離関係をもとに空間上の布置を求める方法



(b) 実際に演奏したドレミファソ音について

多変量解析

■重回帰分析

ある変量 (y) をいくつかの変量 (x_1, x_2, x_3, \dots) の線形結合で説明

$$y = a_1x_1 + a_2x_2 + a_3x_3 + \cdots + b$$

■因子分析

因子負荷

音色の場合 ⇒ 美的因子, 金属性因子, 迫力因子

■クラスタ分析

対象物を量的根拠に基づいて分類する手法

SD法

■ SD (semantic differential) 法

形容詞尺度を利用して印象を科学的に捉える手法

三つの仮定：

- 印象は形容詞による意味空間でとらえられる
- 両極性の形容詞対（反対語対）が存在する
- 両極をなす形容詞の評定尺度は連続である

■ 音色因子はSD法により得られたもの

⇒ 美的因子, 金属性因子, 迫力因子

8. 音色・音質を表現する方法

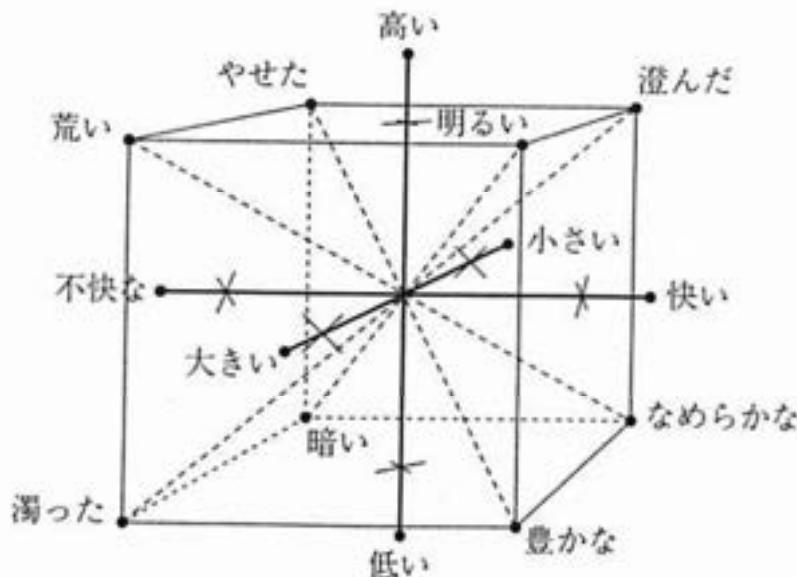
- 印象的側面
- 識別的側面
- 音色評価尺度の因子分析
 - 北村の研究
 - 美的因素 「きれいなー汚い」 「澄んだー濁った」
 - 迫力因子 「迫力のあるーもの足りない」
「豊かなー貧弱な」
 - 金属性因子 「深みのあるー金属性の」
「鋭いー鈍い」
 - 柔らかさ因子 「固いーやわらかい」

- 曾根らの研究 室内音響分野・電気音響分野
 - 美的・叙情的因素 (美的因子)
「美しい」「快い」「うるおいのある」「澄んだ」「情趣豊かな」「つやがある」「繊細な」「まとまりのある」
 - 量的・空間的因素 (迫力因子)
「ひびく」「豊かな」「音量感のある」「迫力のある」「広がりのある」「生き生きした」
 - 明るさを表わす因子 (金属性因子)
「明るい」「華やか」「軽やか」
 - 柔らかさを表わす因子
「柔らかい」「歯切れのよい」

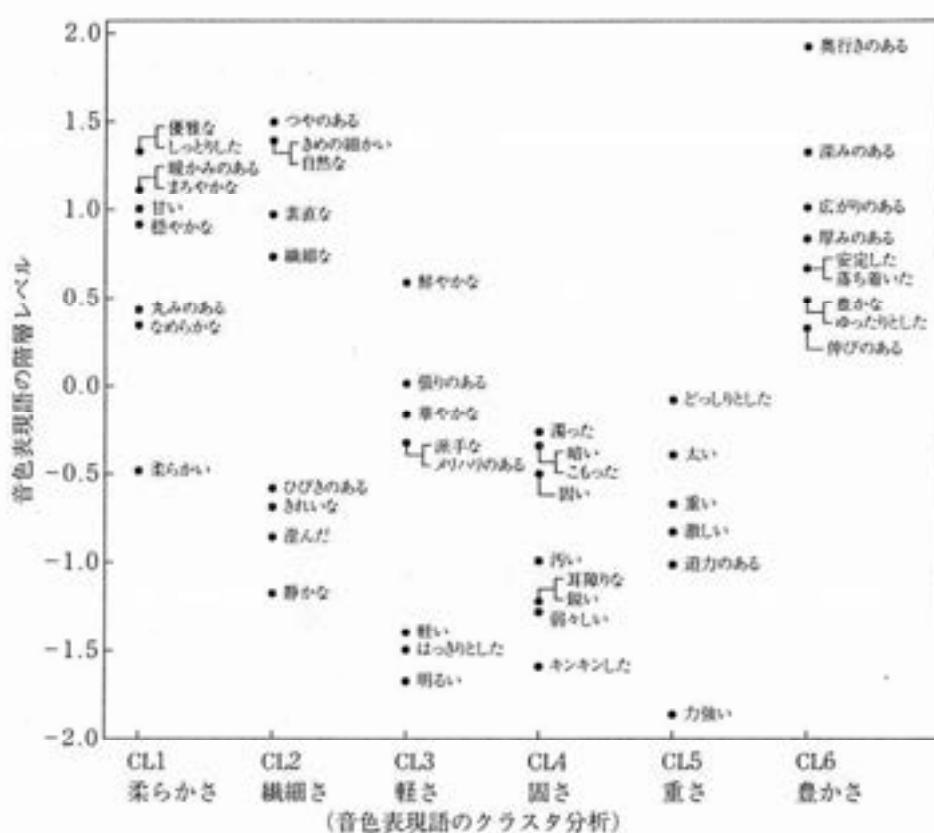
音質評価のための7属性

■厨川らの研究

- 大きさ, 高さ, 快い
- 協和性, 粗滑性, 明暗性, 豊痩性



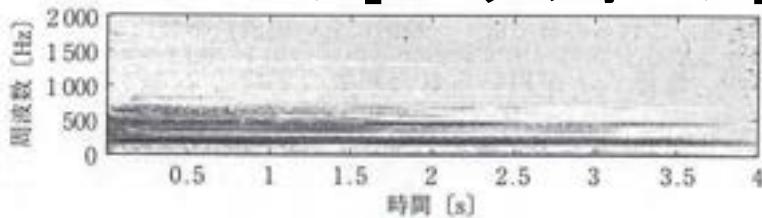
音色表現語と階層構造



環境音の音色を表わす擬音語表現

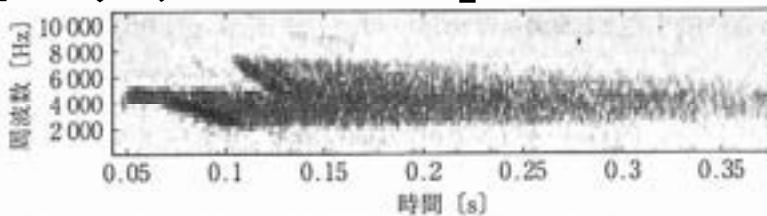
■ 鐘の音

「ゴーン」「グワーン」

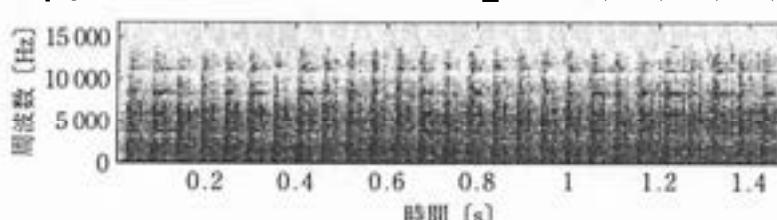


■ スズメのさえずり

「ピヨ」



■ 工事現場の音 「ガガガガガガ」 「ダダダダダダ」



9. 音色の分類

音の特徴

音色を規定する要因

■ 基礎的音色

■ 静的音色

スペクトル(S)が
が時間的に定常

振幅スペクトルの形状

■ 準静的音色

Sが時間的に定常

振幅包絡(E)の時間変化

■ 動的音色

SとEの時間変化

成分音の振幅の時間変化

■ 準動的音色

Eは時間的に定常

スペクトルの時間変化

Sは時間変化

■ 総合的音色

■ 複合的音色

4つの音色の場合

(複雑)

■ 音楽的音色

時間・空間的に集合

(複雑)

静的音色

■ 振幅スペクトルと音色の関係

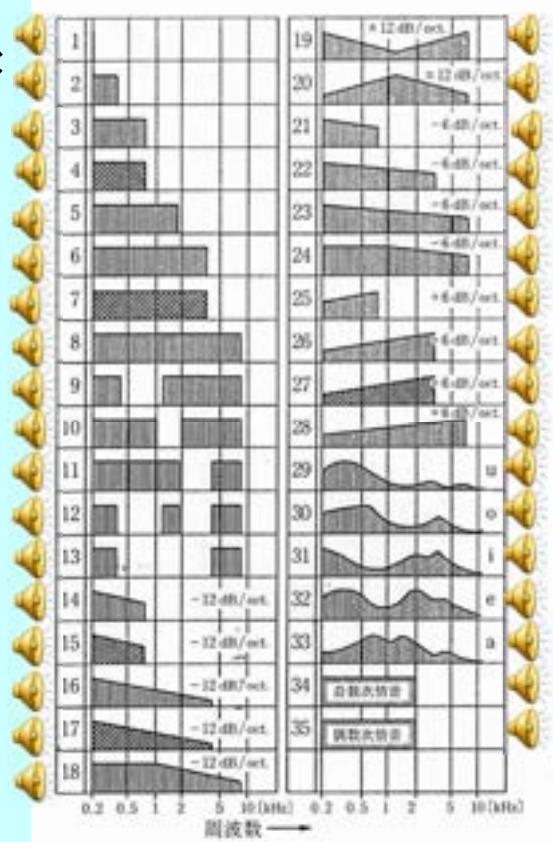
■ W.H. Lichteの実験

- 明るさ，粗さ，豊かさの調査

■ Von Bismarchの実験（右）

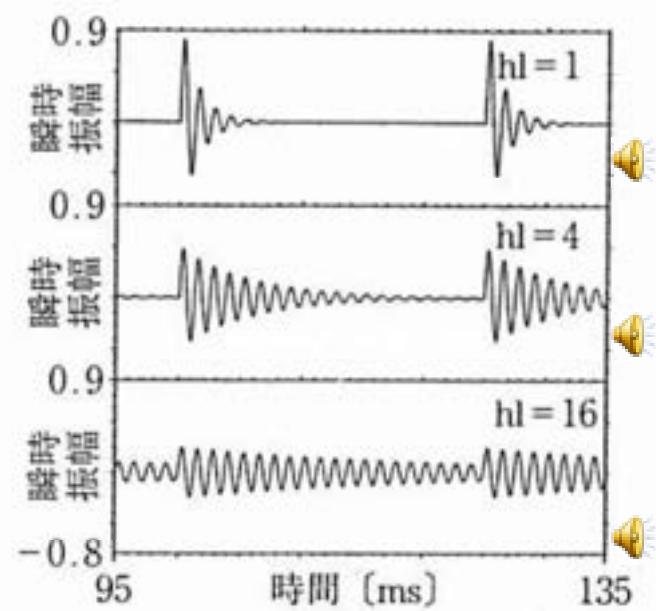
- 4因子による分析

（美的因子，金属性因子，迫力性因子，柔らかさ因子）

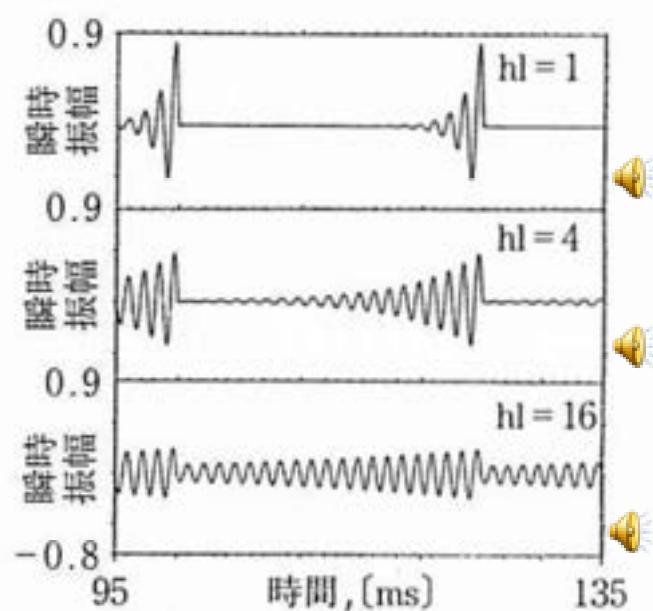


準静的音色

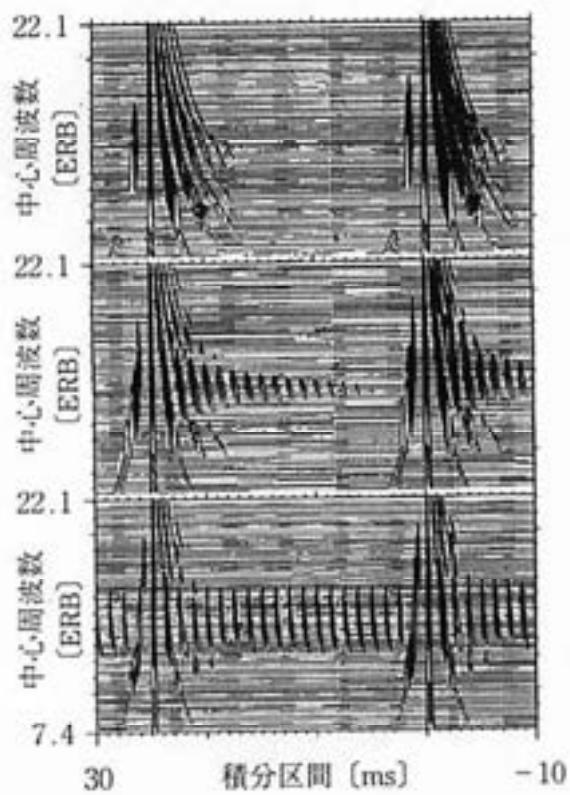
■ R. D. Pattersonの実験



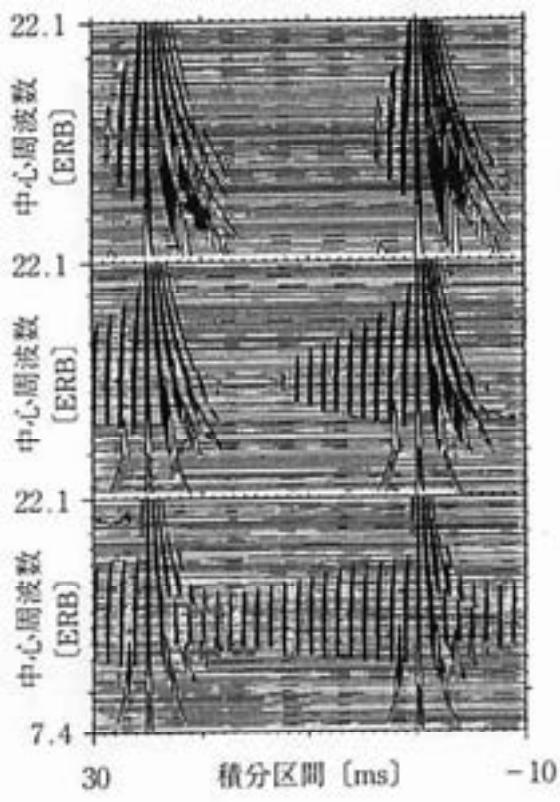
(a) damped 波



(b) ramped 波

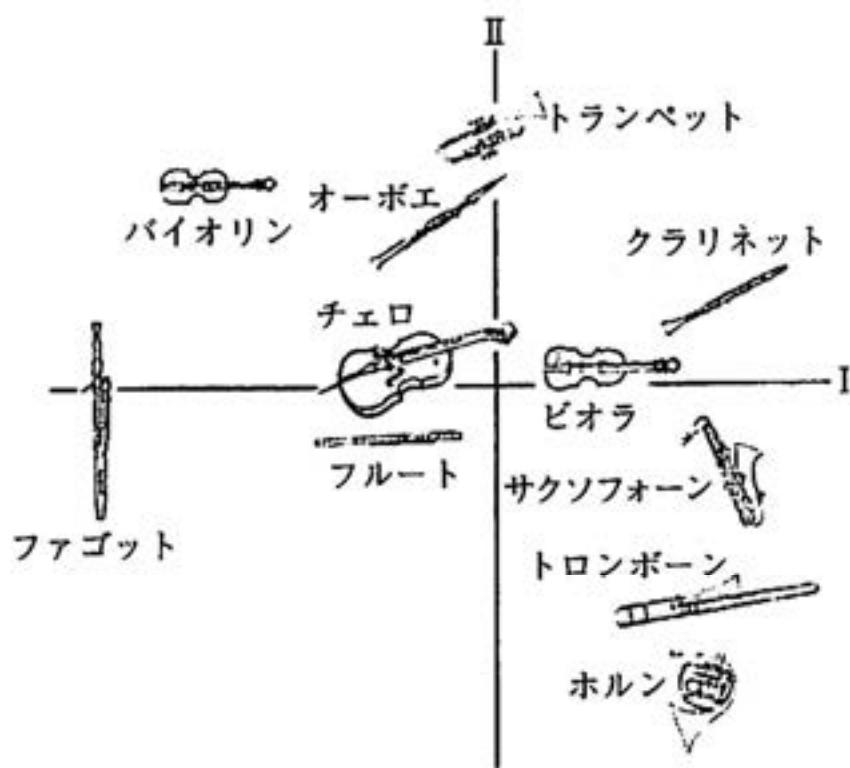


(a) damped 波

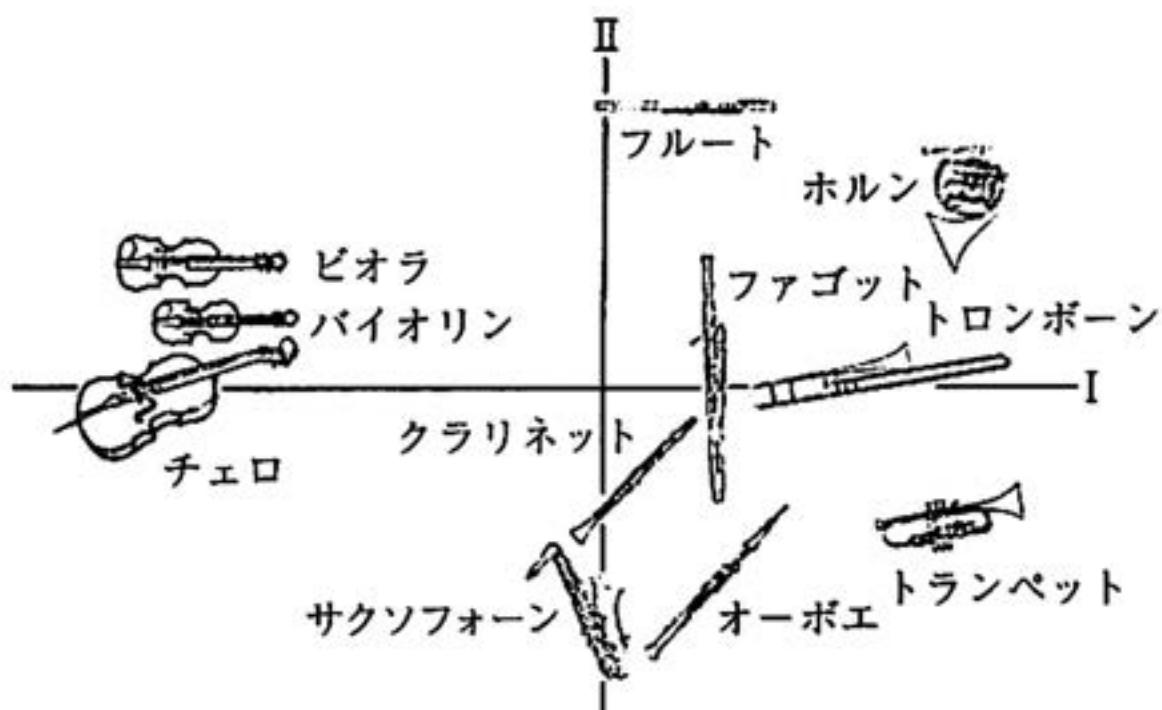


(b) ramped 波

動的音色

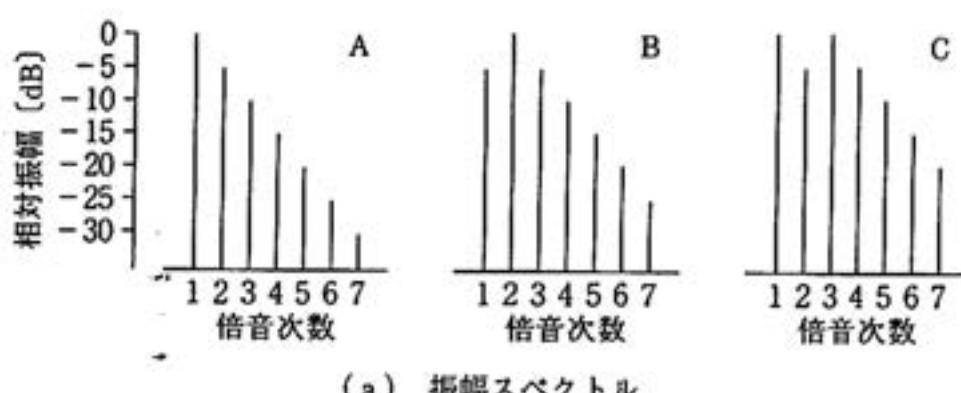


(a) 楽器音(A4 音)の波形を繰り返して接続した音について



(b) 実際に演奏したドレミファソ音について

■ Millerの実験



(a) 振幅スペクトル



(b) 振幅の時間包絡波形

準動的音色

■ 周波数変調 (FM) 音の知覚

$$f(t) = \sin(2\pi f_c t + m \sin(2\pi f_p t))$$

$$f(t) \approx \sin(2\pi f_c t) - \frac{m}{2} \sin(2\pi(f_c - f_p)t) + \frac{m}{2} \sin(2\pi(f_c + f_p)t)$$

■ ビブラート

I485F 知覚情報処理特論

13. 聽覚情景解析

2023年5月28日(日)

人間情報学研究領域

鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

Email: unoki@jaist.ac.jp



1. カクテルパーティー効果

- ざわついたカクテルパーティー会場でも有意義に会話を楽しめることに由来 Cherry (1953)



Interspeech2007
in Antwerp

カクテルパーティ効果？

音楽

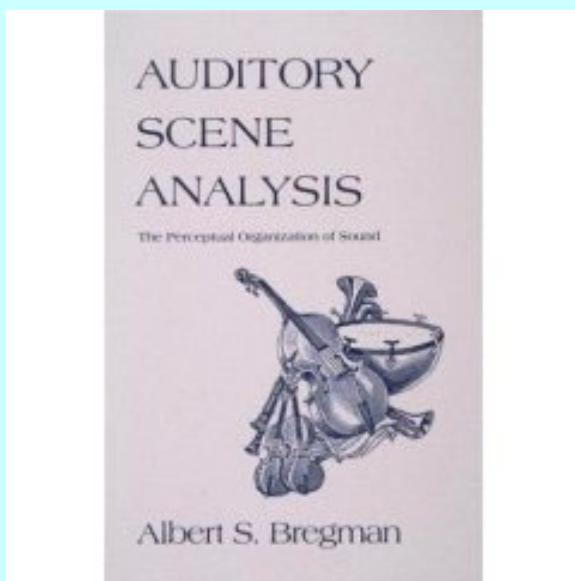


2. 聴覚の情景解析 (ASA)

Auditory Scene Analysis

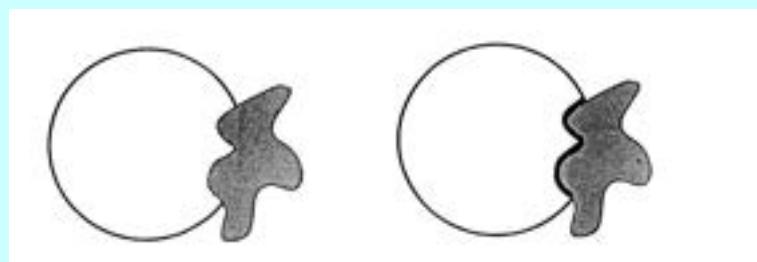
Albert Bregman (1990)

<http://www.psych.mcgill.ca/faculty/bregman.html>

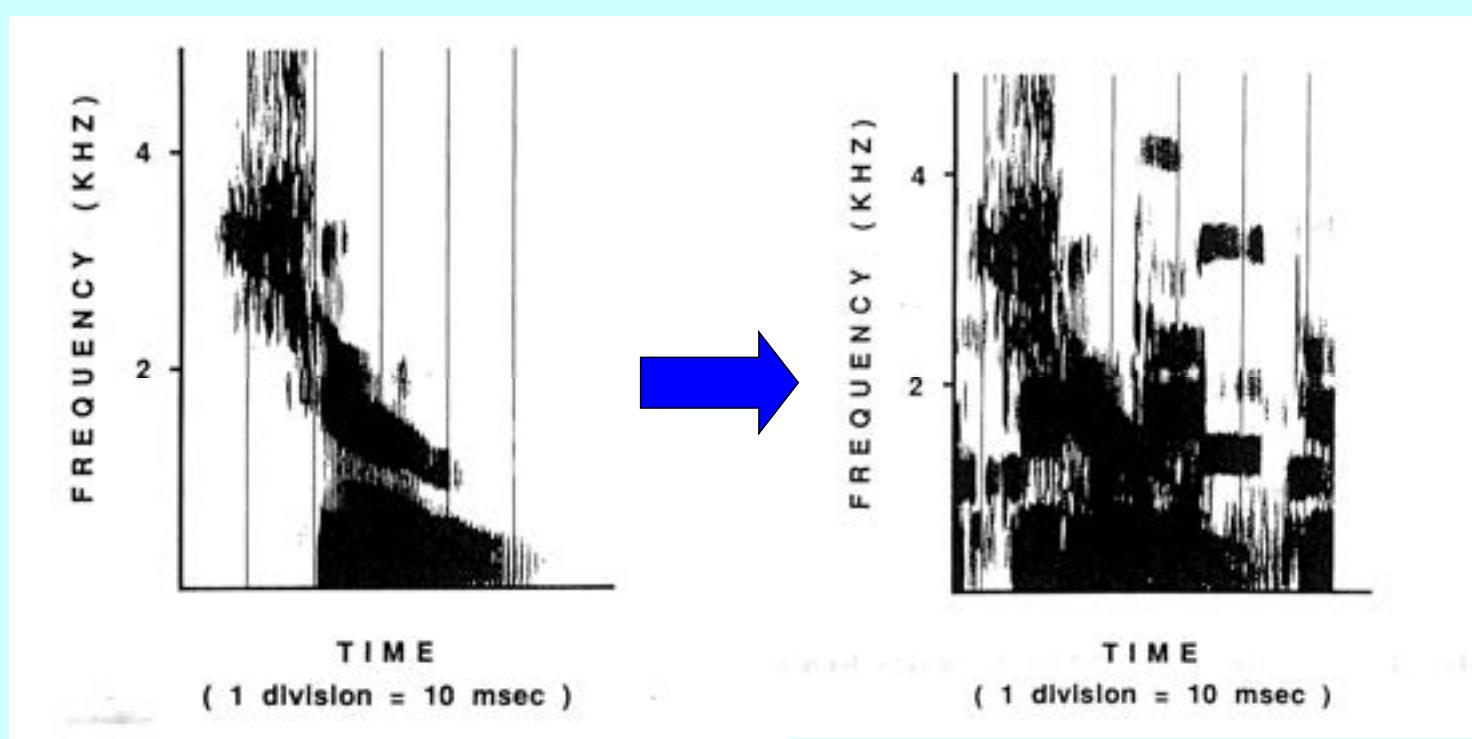


視覚に関するゲシュタルト心理学

図と地の分離



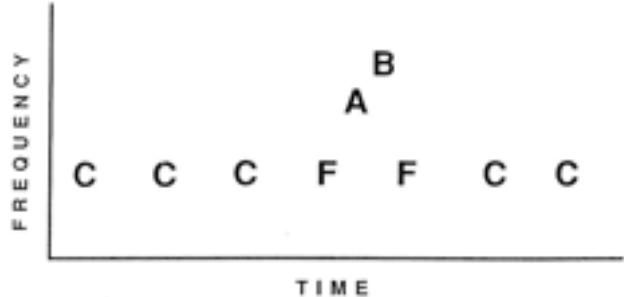
音声の場合は・・・



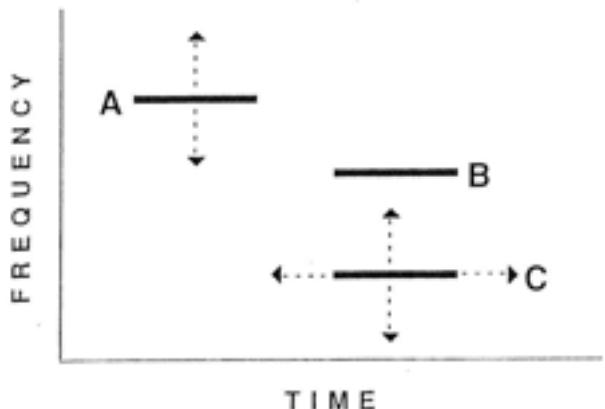
雑音／別の音が重なった場合

音の流れ

ストリーミング (streaming)



時間と周波数の近接



補完



体制化 (organization)

聴覚末梢系の働き

- 音を通じて環境を把握するためのもの
- 能動的システム
- 四つの心理物理的規則

1. 立上り・立下りの同期
2. 変化は急激には起こらない
3. 基本周波数の整数倍の変化（調波性）
4. 共通な変化

3. 同時グルーピング

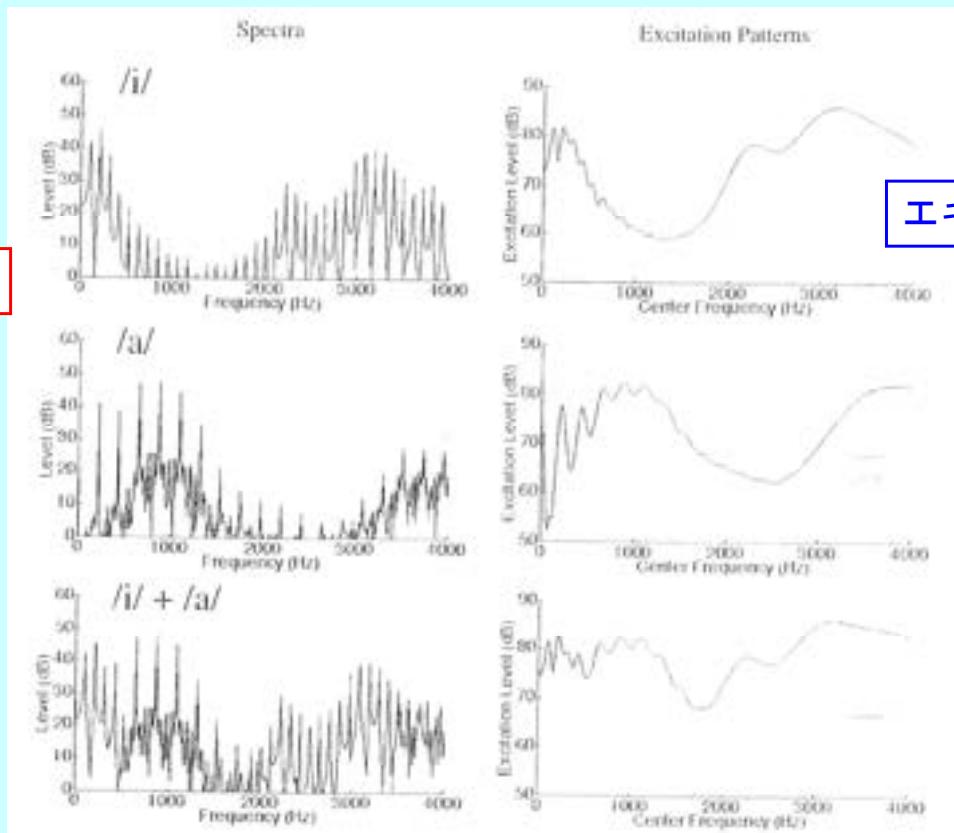
- 周波数分析に関連したグルーピング
⇒ simultaneous grouping

- 蝸牛のスペクトル分析に依存

- グルーピングの手がかり
 - 時間的な手がかり
 - 調波性の手がかり
 - 方向性の手がかり

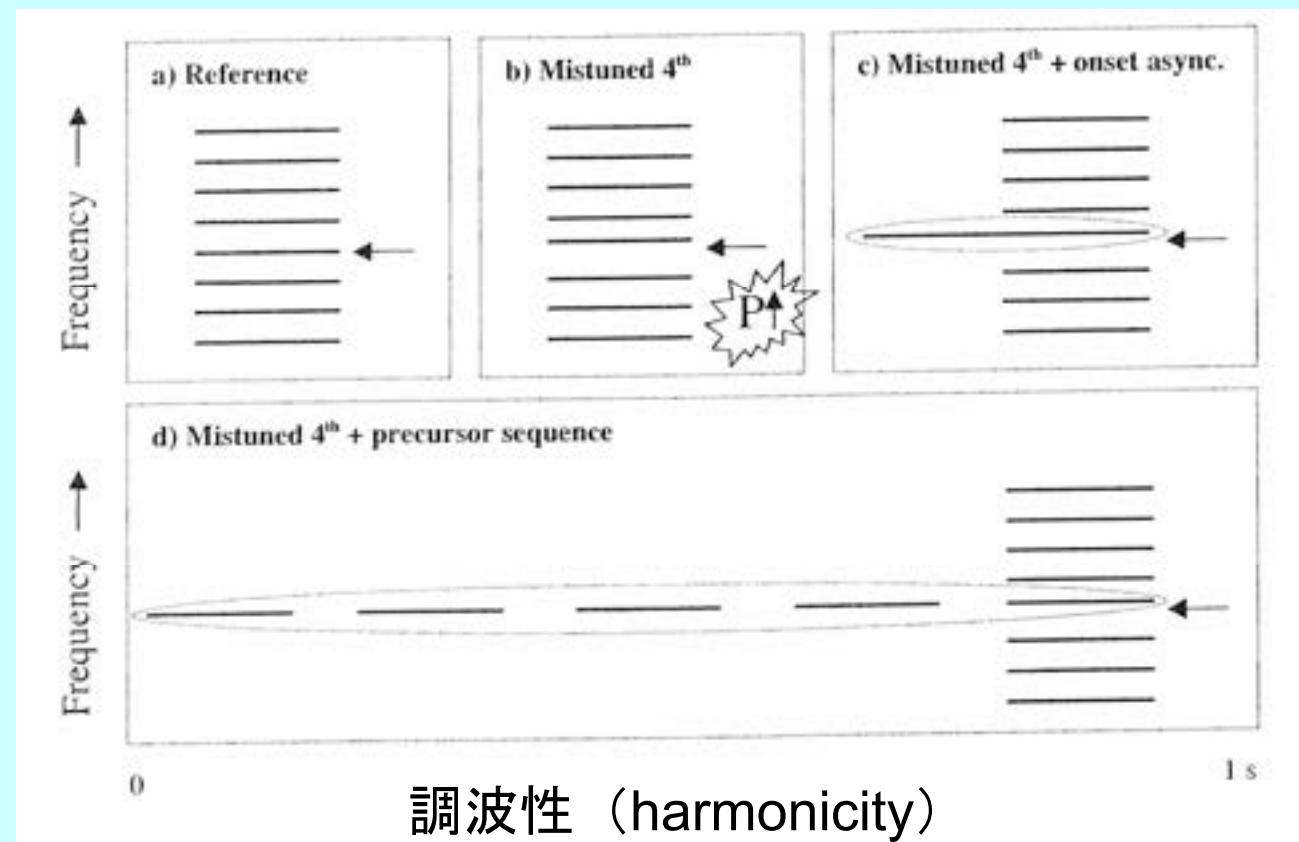
スペクトルとエキサイテーションでの違い

スペクトル

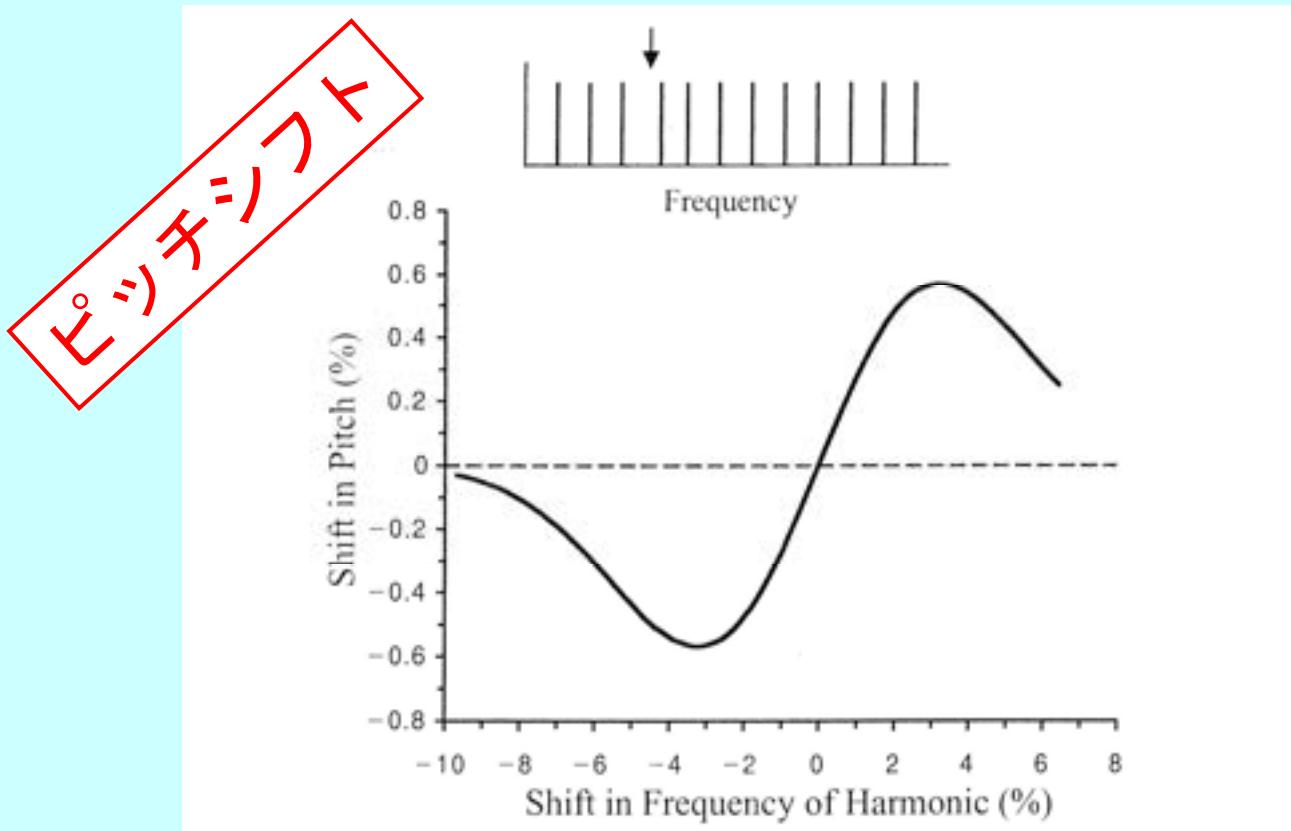


エキサイテーション

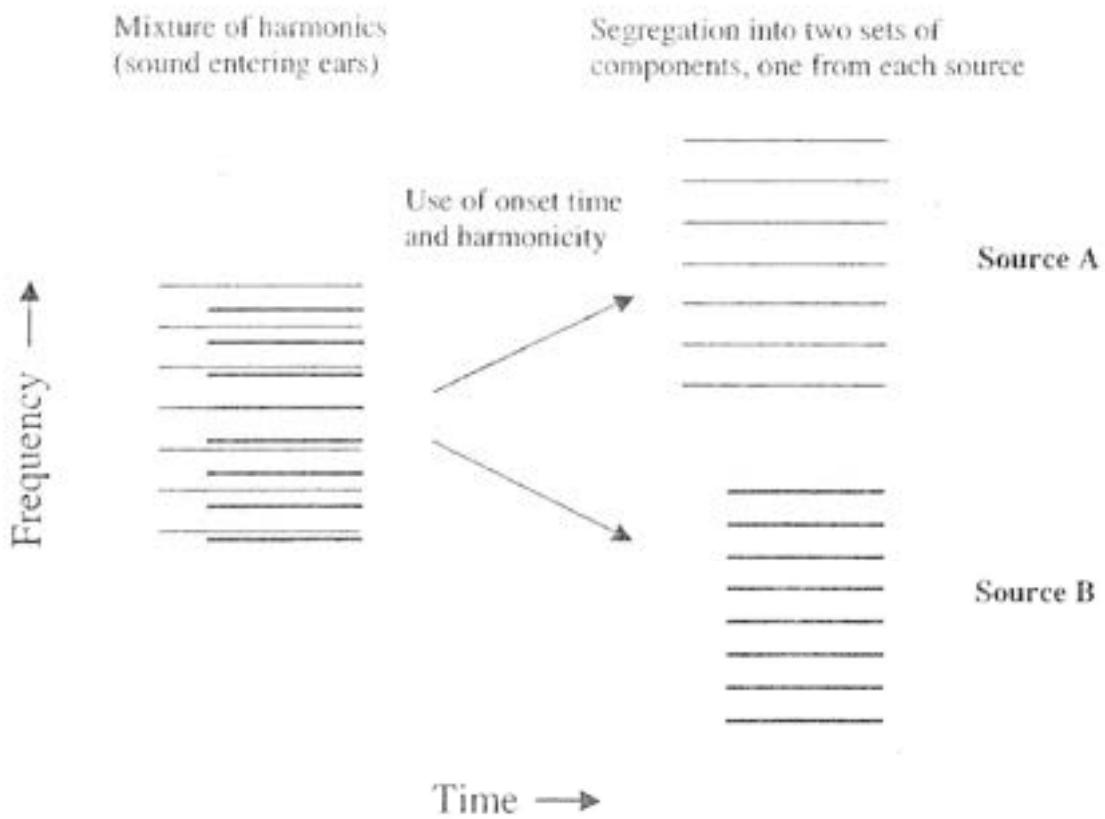
調波複合音の知覚



調波性とピッチの対応



調波複合音の分離

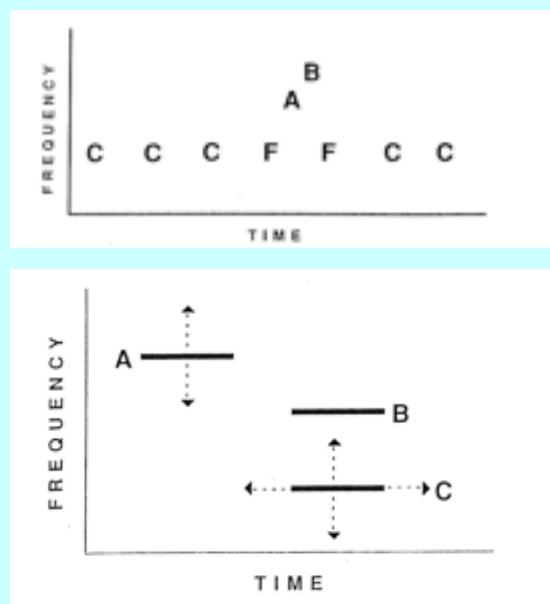


4. 逐次グルーピング

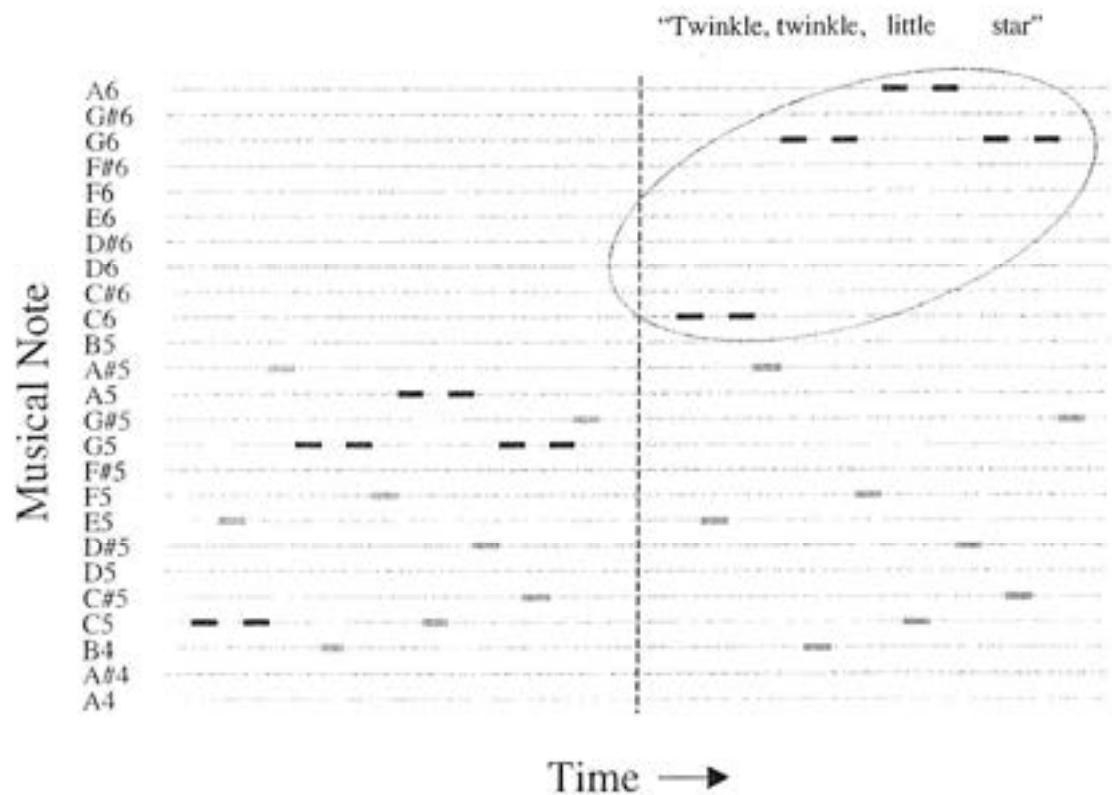
■ 音の集まりを知覚的なストリームにまとめる体制化のこと

■ グルーピングの手がかり

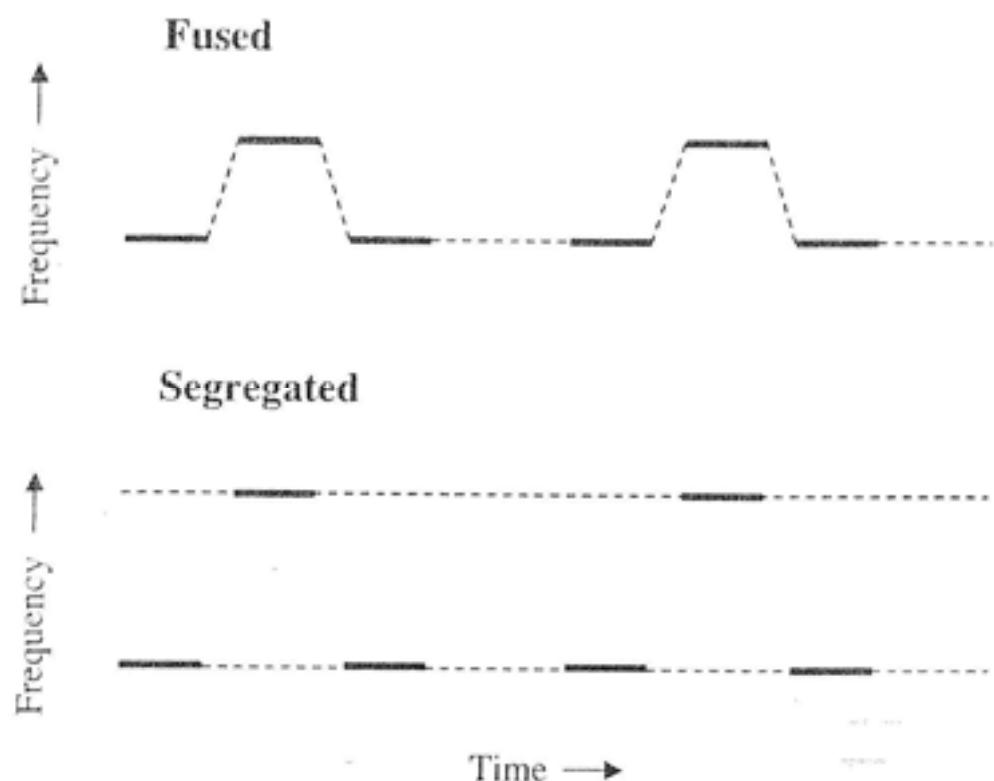
- 周期性の手がかり
- スペクトルの手がかり
- 方向性の手がかり
- 連続性の手がかり



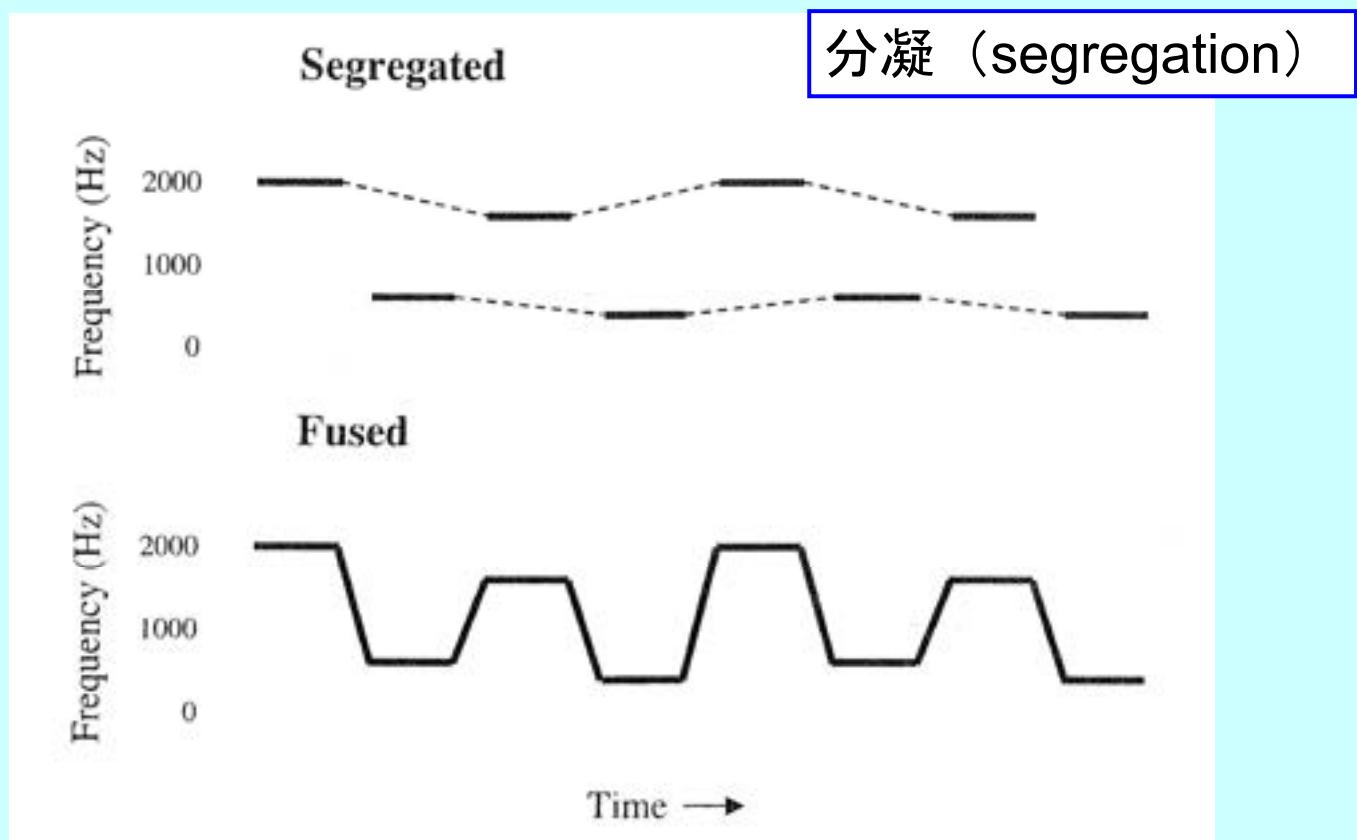
メロディー上の音の分凝



ストリーム分凝



周波数の変化に伴う分凝



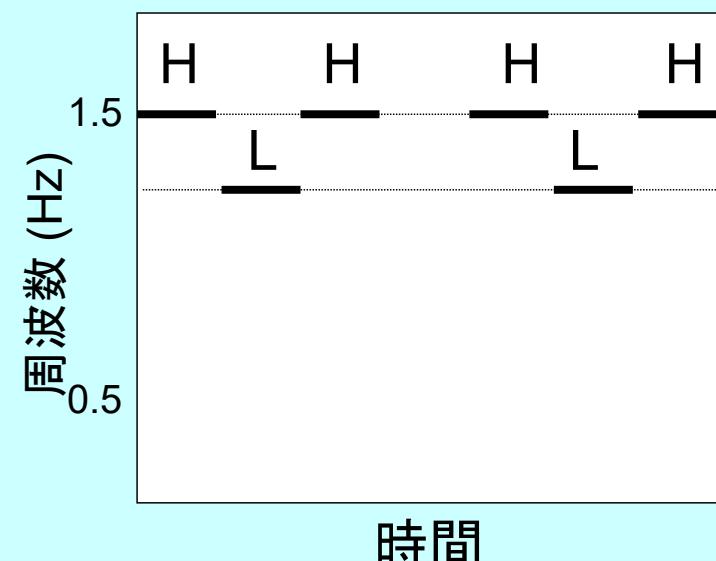
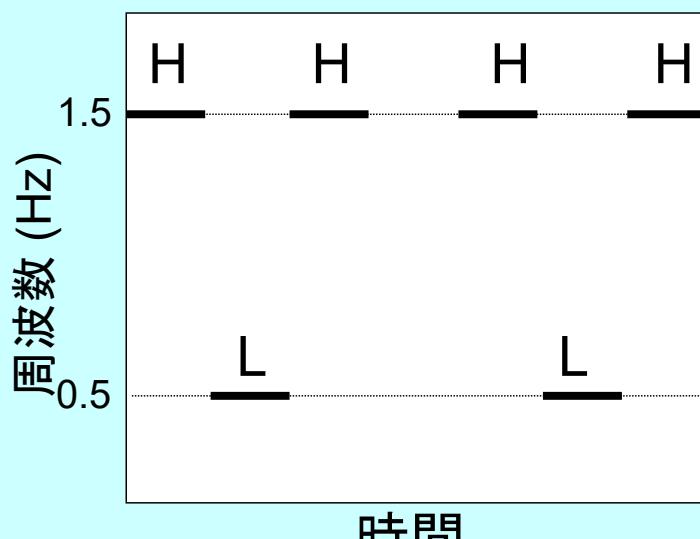
事例紹介

ストリーミング

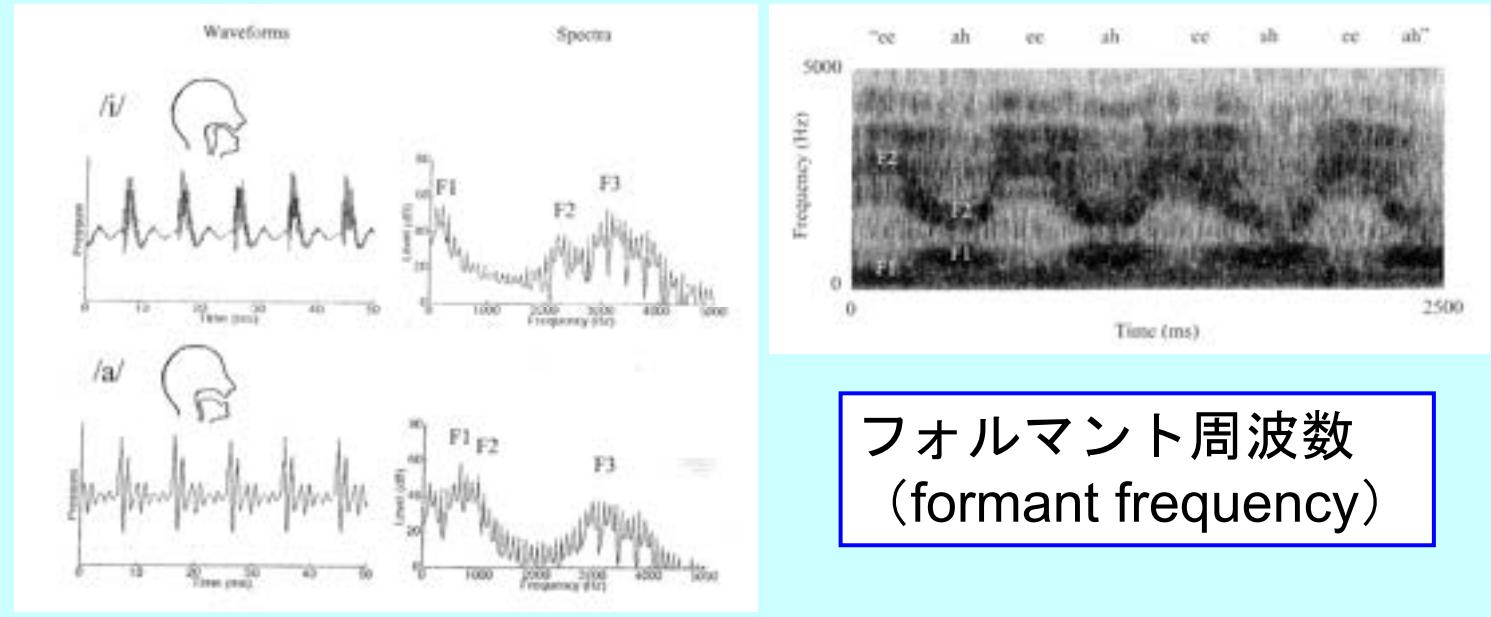
- 周波数間隔／時間間隔
- 繰り返しの速度



HLH-HLH HH-HH と -L---L-



フォルマント知覚

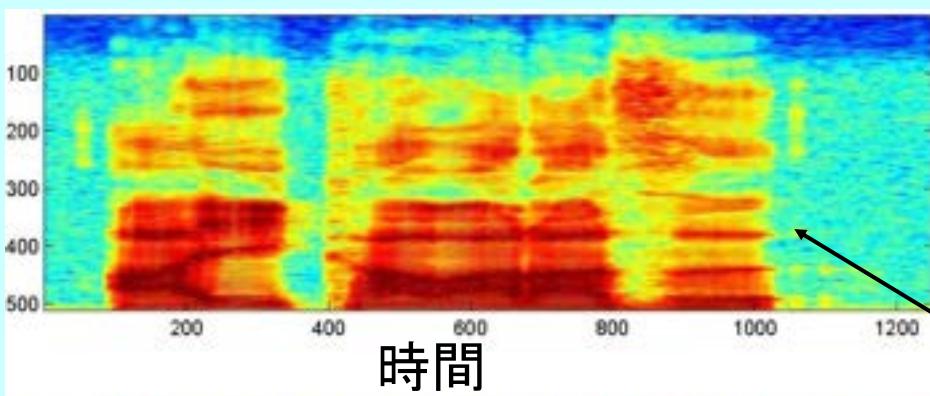


フォルマント周波数
(formant frequency)

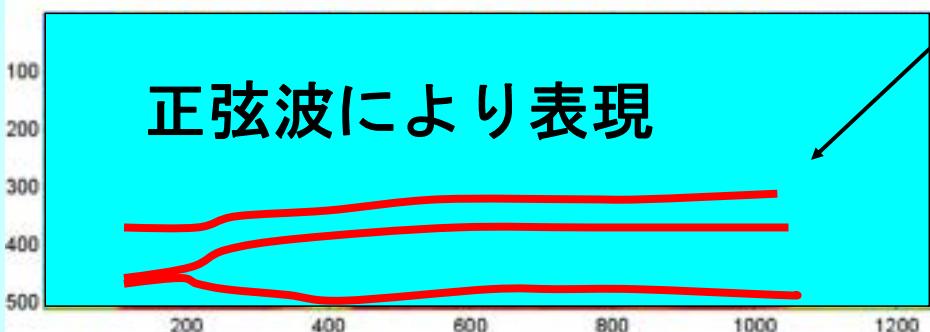
Sine-Wave音声（フォルマント知覚）

音声のスペクトルグラム

周波数



周波数

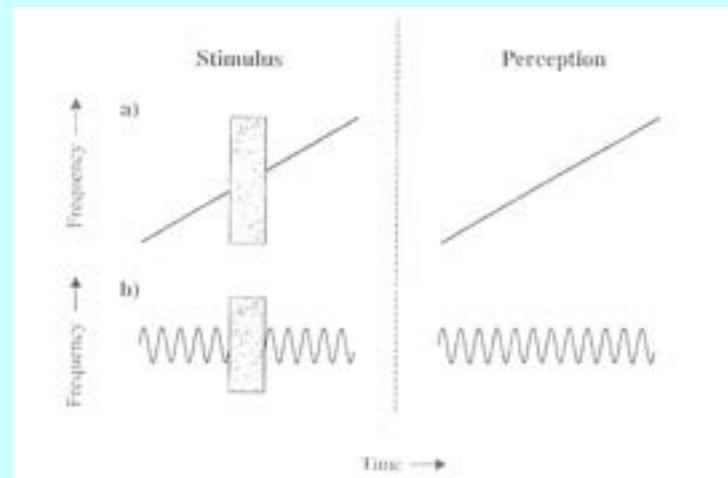
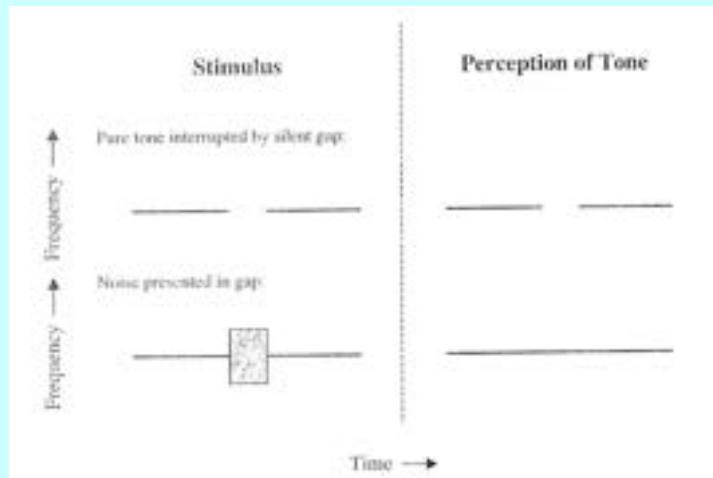


フォルマント
軌跡

正弦波により表現

5. 音韻修復現象

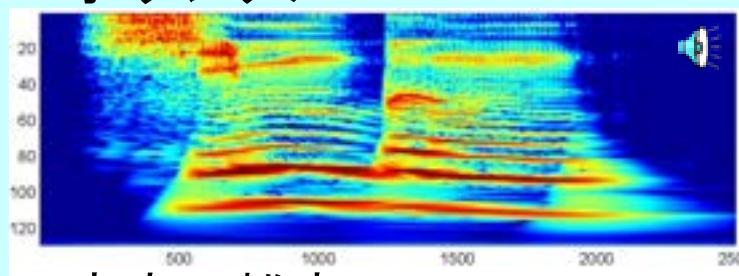
■ 音の集まりを知覚的なストリームにまとめる体制化のこと



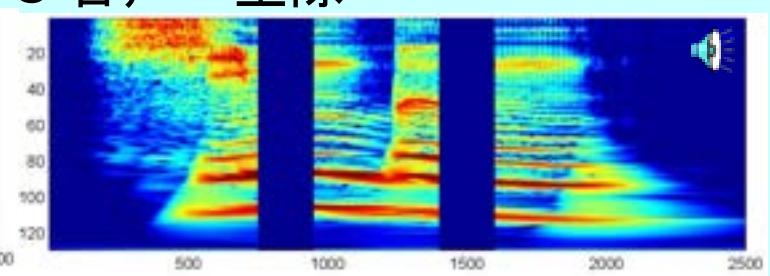
事例紹介

音韻修復現象

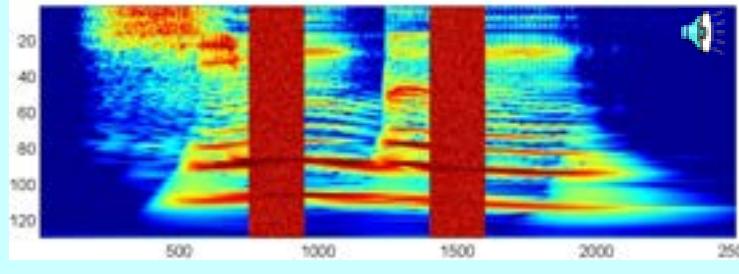
A オリジナル



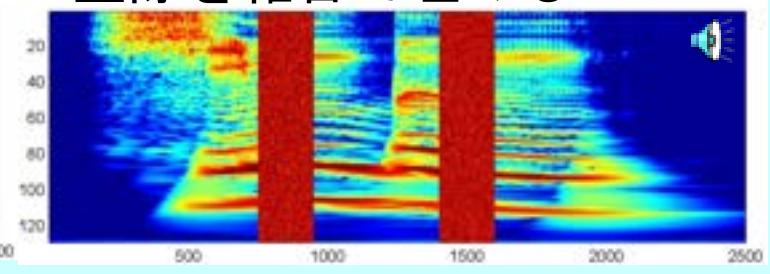
C 音声－空隙



B 音声＋雑音



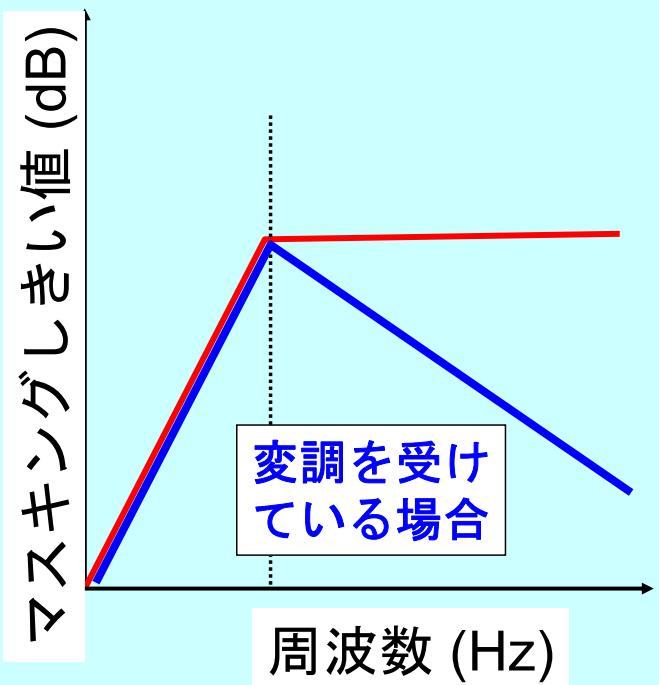
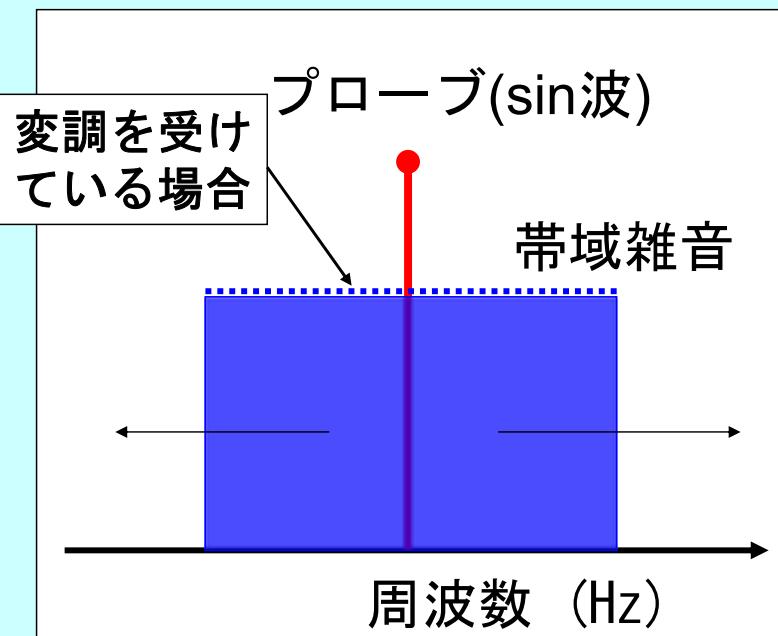
D 空隙を雑音で埋める



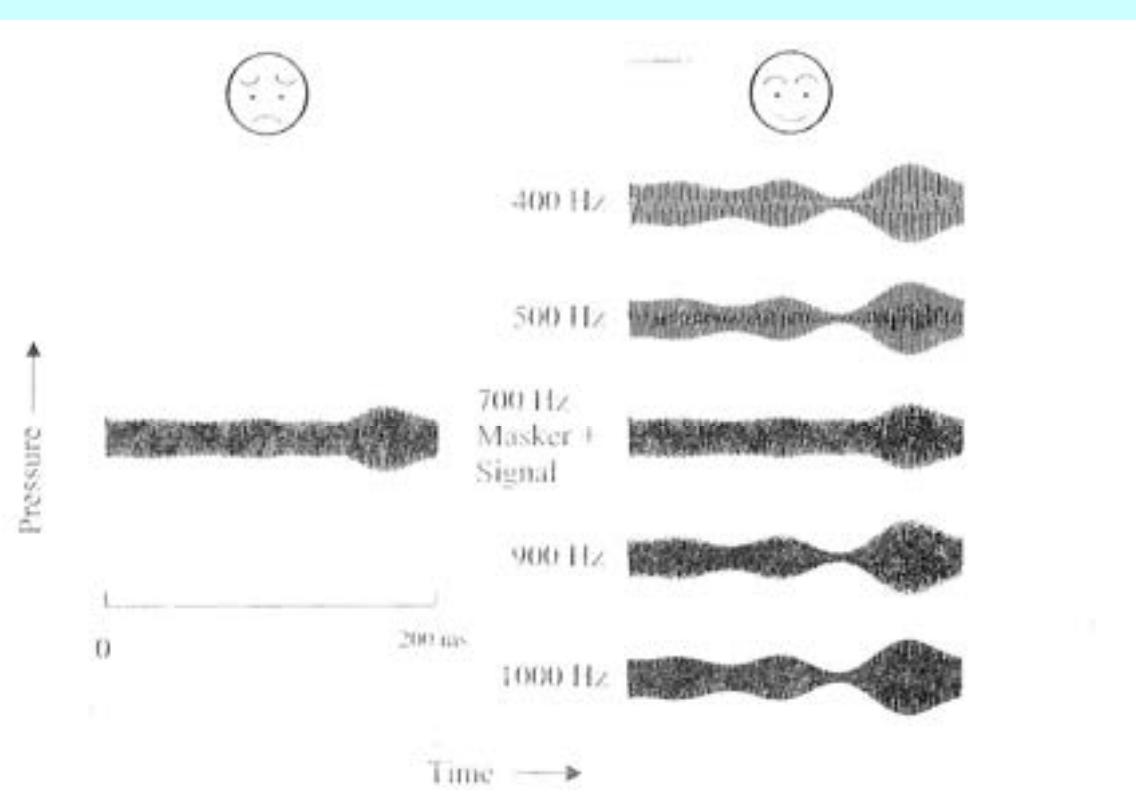
失われた情報を補って聞く能力

6. 共変調マスキング解除 (CMR)

- 雑音における共通な振幅変動により、マスキングされにくくなること



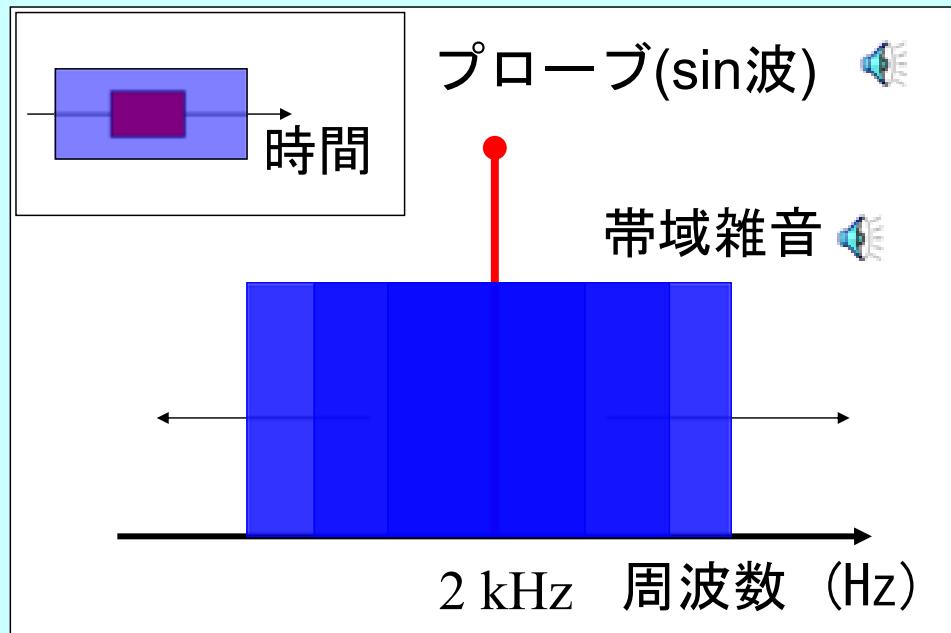
振幅変調を手がかりとした知覚現象



同時マスキング実験

■マスキング

雑音によって音が聴きにくくなる！



徐々に帯域幅を
広げると...

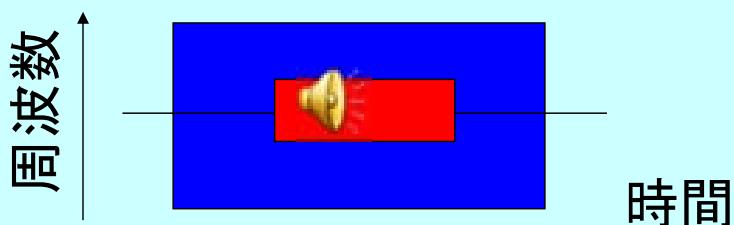


プローブの音圧レ
ベルを上げると...



共変調マスキング解除の実験

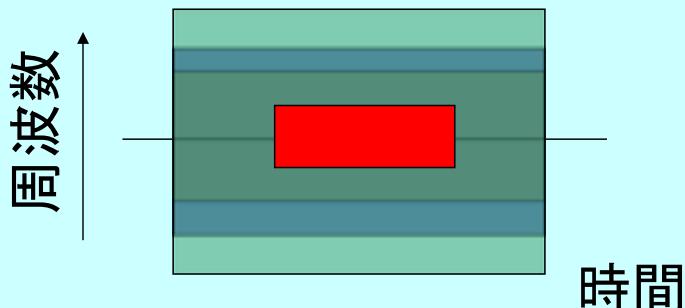
■白色雑音の場合



聴こえない...

■振幅変調雑音の場合

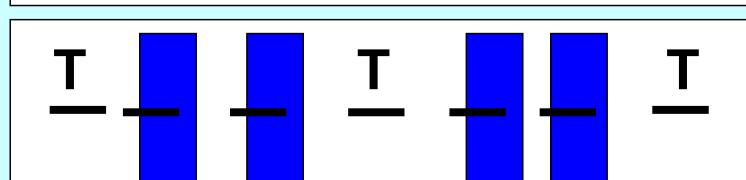
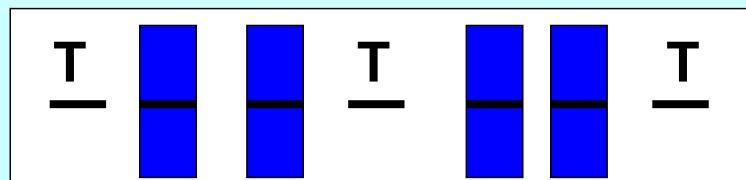
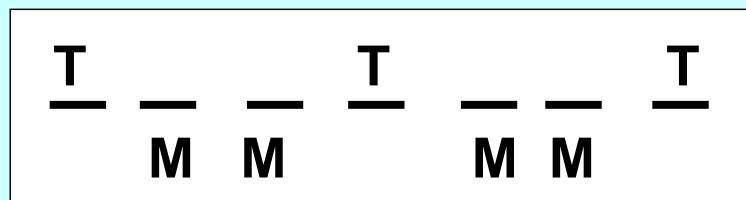
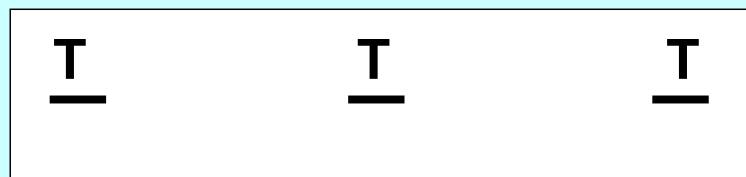
雑音が増えているのに
聴こえる？？？



事例紹介

共変調マスキング解除

■リズミックマスキング解除



Tone
Masker



同じ変動をもつ
マスカーによって
マスキングを解除

7. 計算論的聴覚の情景解析 (CASA)

■ CASA: Computational Auditory Scene Analysis ASAの現象を計算モデルで説明する研究分野

M. Cooke and D. P. Ellis (2001)

■ 国内外の研究

- 英国シェフィールド大学のチーム (M. Cooke)
- 米国MIT・コロンビア大学 (D. Ellis)
- NTT (中谷)
- 京都大学 (奥乃)
- JAIST (鶴木・赤木)

I485F 知覚情報処理特論

14. 聽覚モデルとシミュレータ

2023年5月28日(日)

人間情報学研究領域

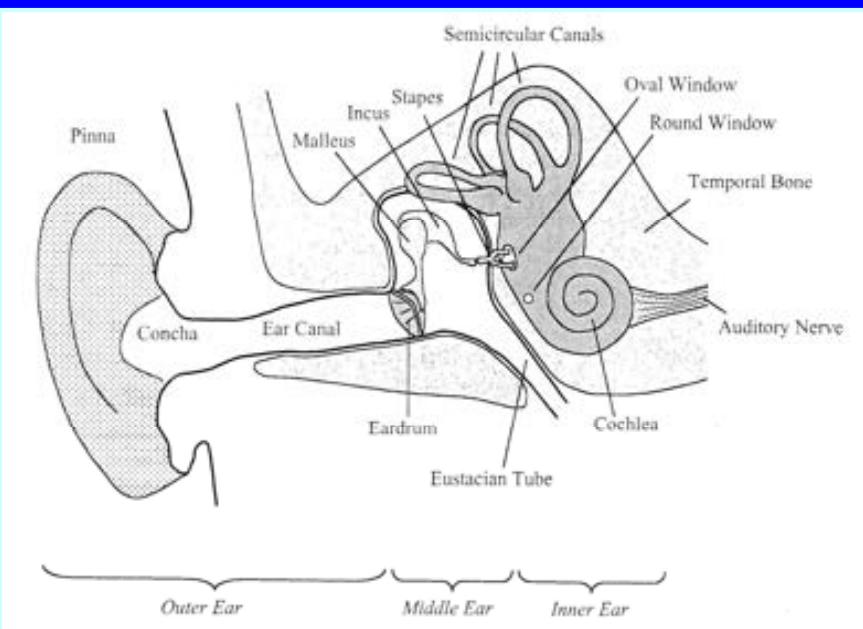
鵜木 祐史

URL: <http://www.jaist.ac.jp/~unoki/>

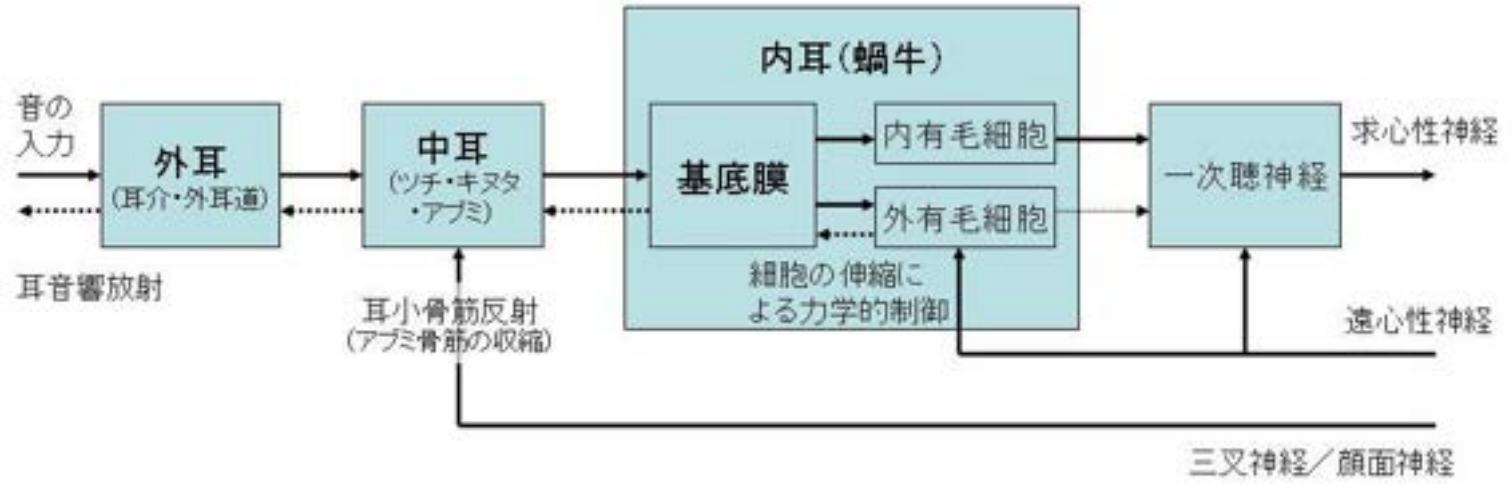
Email: unoki@jaist.ac.jp



1. 聴覚の機能の細分化



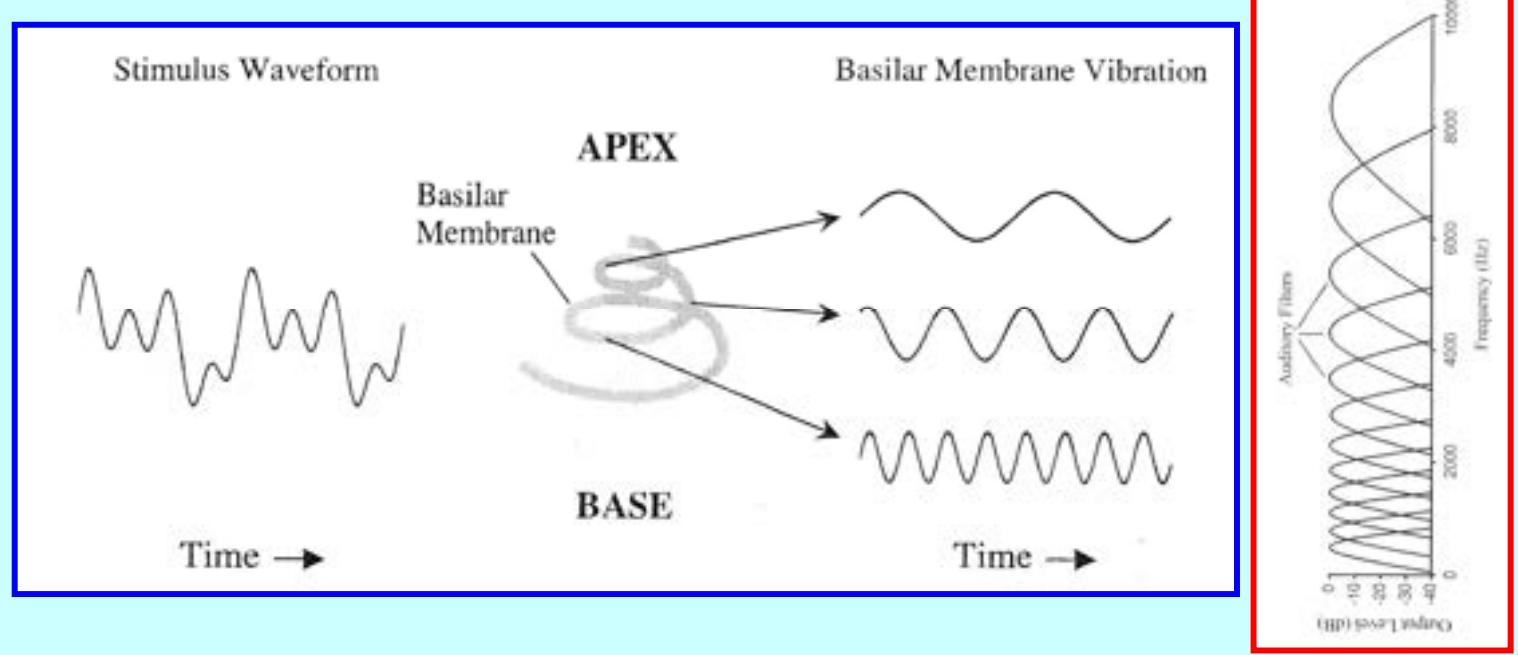
聴覚末梢系のモデル



2. 周波数選択性

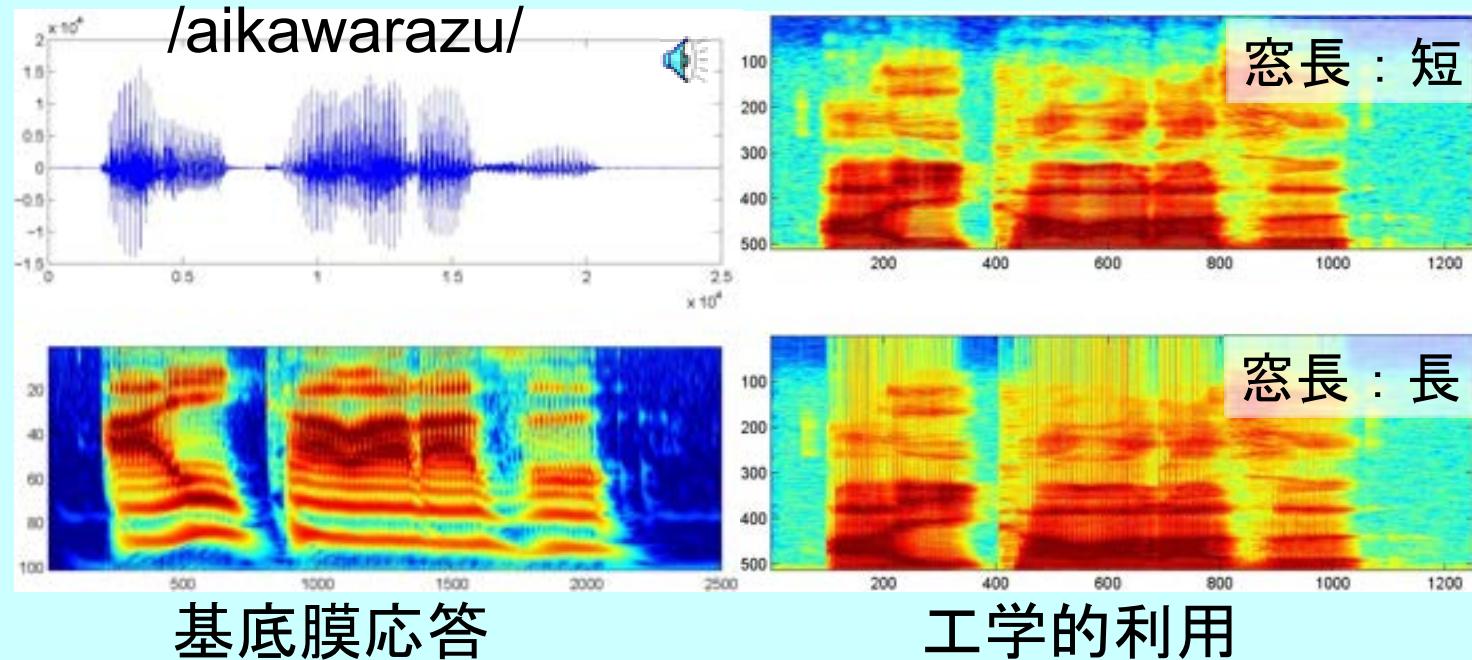
■周波数選択性 (frequency selectivity)

■複合音中の正弦波成分を分解する能力



3. 聽覚フィルタバンク

定Qフィルタバンク vs. 短時間フーリエ変換
(ウェーブレット変換) (窓関数による切取り)

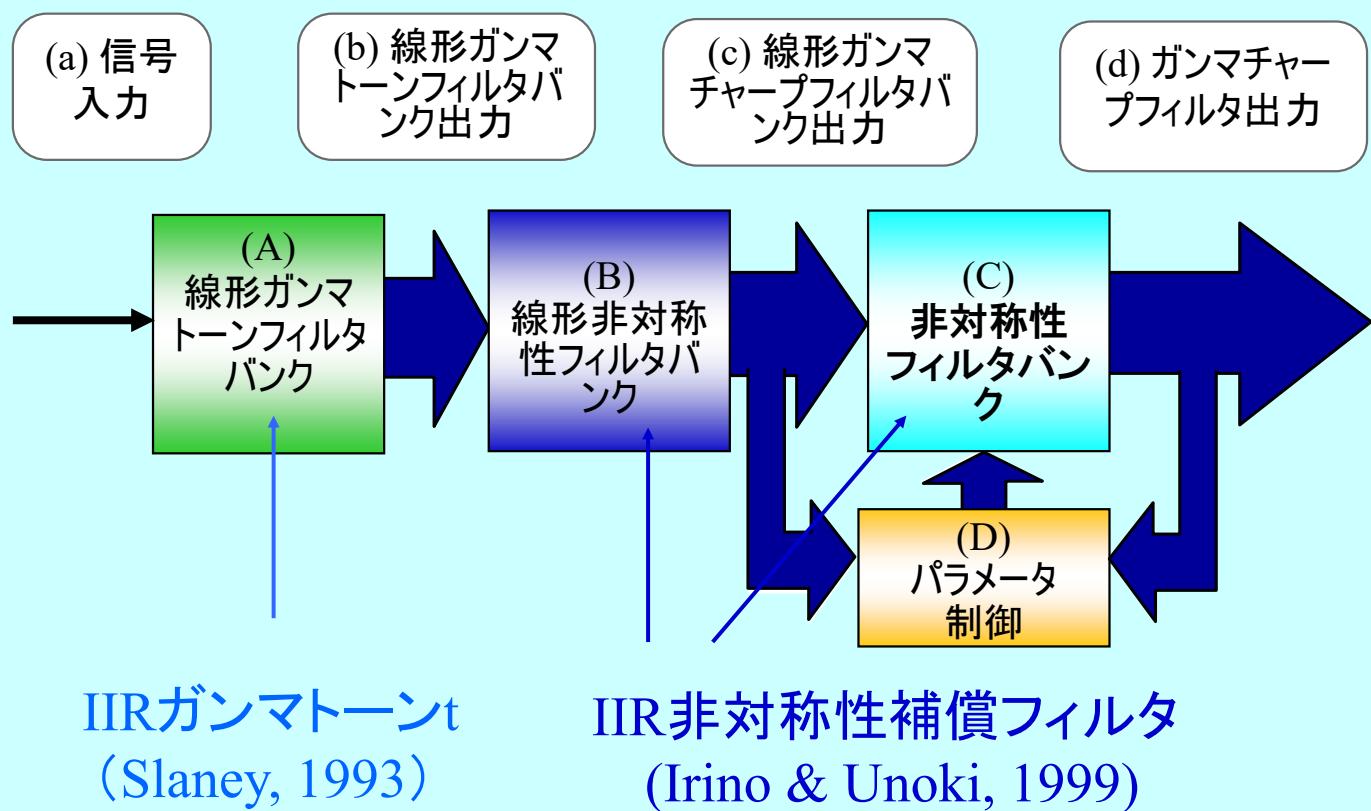


3.1 課題

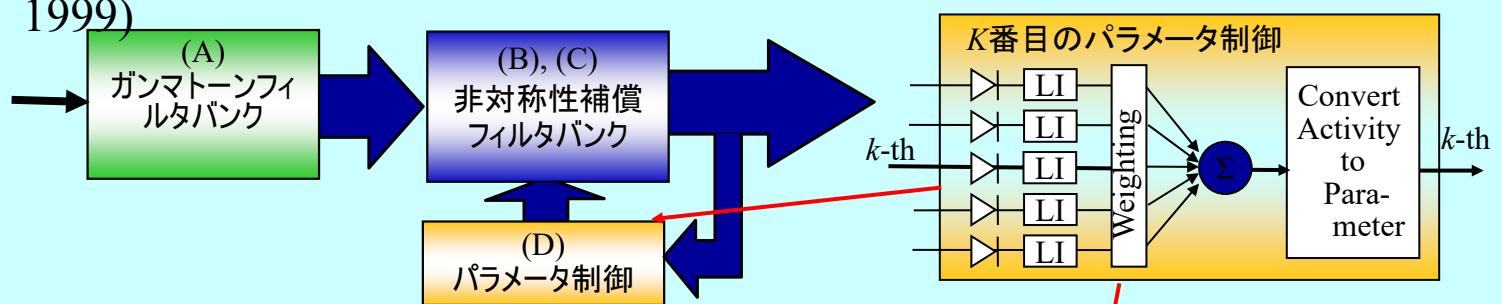
- 非線形フィルタバンク
- 音圧に依存して形状が変わること
- 圧縮効果がある（AGC機能）こと
- 様々なマスキング現象を説明し得ること

- 聴覚特性を考慮した分析技術の完成
⇒ 聴覚のメカニズムの解明
⇒ 音声信号処理などへの応答

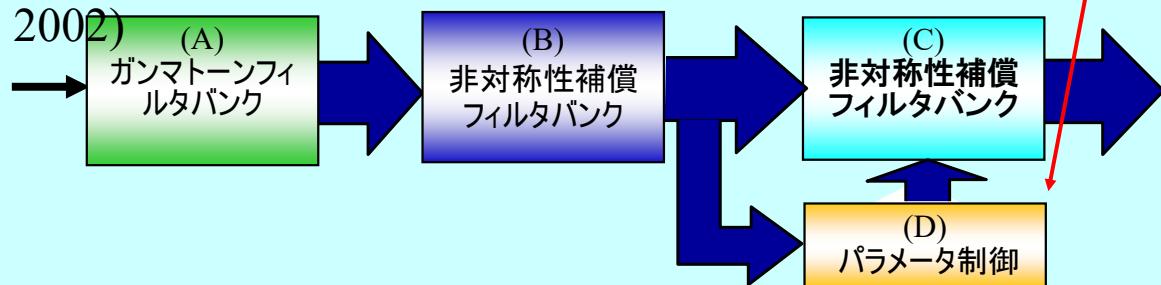
3.2 フィルタバンクの構成



- 解析的ガンマチャーブフィルタバンク (Irino & Unoki, 1999)



- 圧縮型ガンマチャーブフィルタバンク (鵜木, Patterson, 入野, 2002)



3.3 フィルタバンクの構築

- 様々な周波数のノッチ雑音マスキングデータに適合

(鵜木, Patterson, 入野, 2002)

- GM00 (Glasberg and Moore, 2000):

250, 500, 1k, 2k, 4kHz; 合計304個

- BRD98 (Baker et al. 1998):

250, 500, 1k, 2k, 3k, 4k, 6kHz; 合計973個

- 一般の音刺激に対して、フィルタの圧縮特性を決める音圧レベルをどのように求めるか？

- パラメータ推定のための音圧レベル算出法の提案

- 同じノッチ雑音マスキングデータに再適合

音圧レベル算出

(a) ノッチ雑音マスキングの状況
masker/probe

(b) 内耳入口での状況
(中耳の伝達特性を利用)

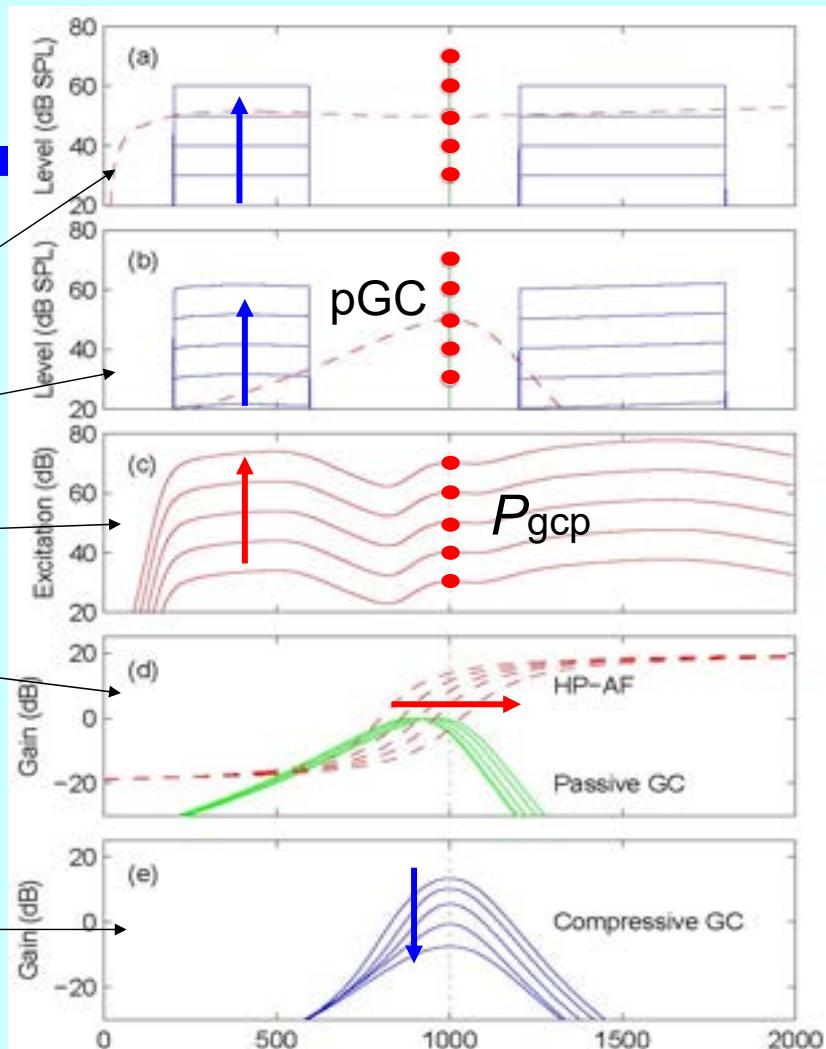
(c) エキサイテーションパターン

(d) 受動的ガンマチャープ+HP-AF

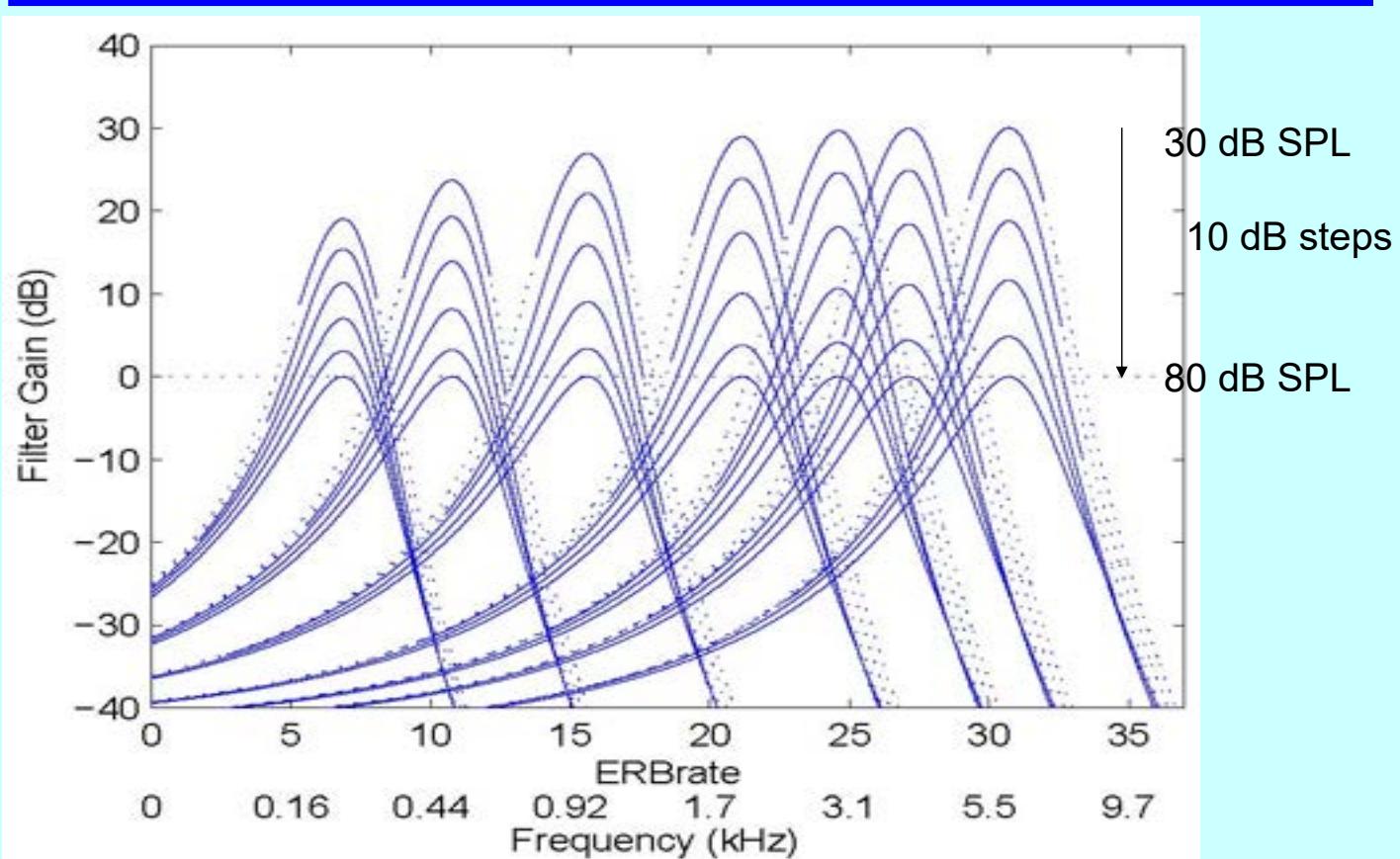
レベル依存

$$f_{r2} = (f_{rat}^{(0)} + f_{rat}^{(1)} \cdot P_{gcp}) \cdot f_{p1}$$

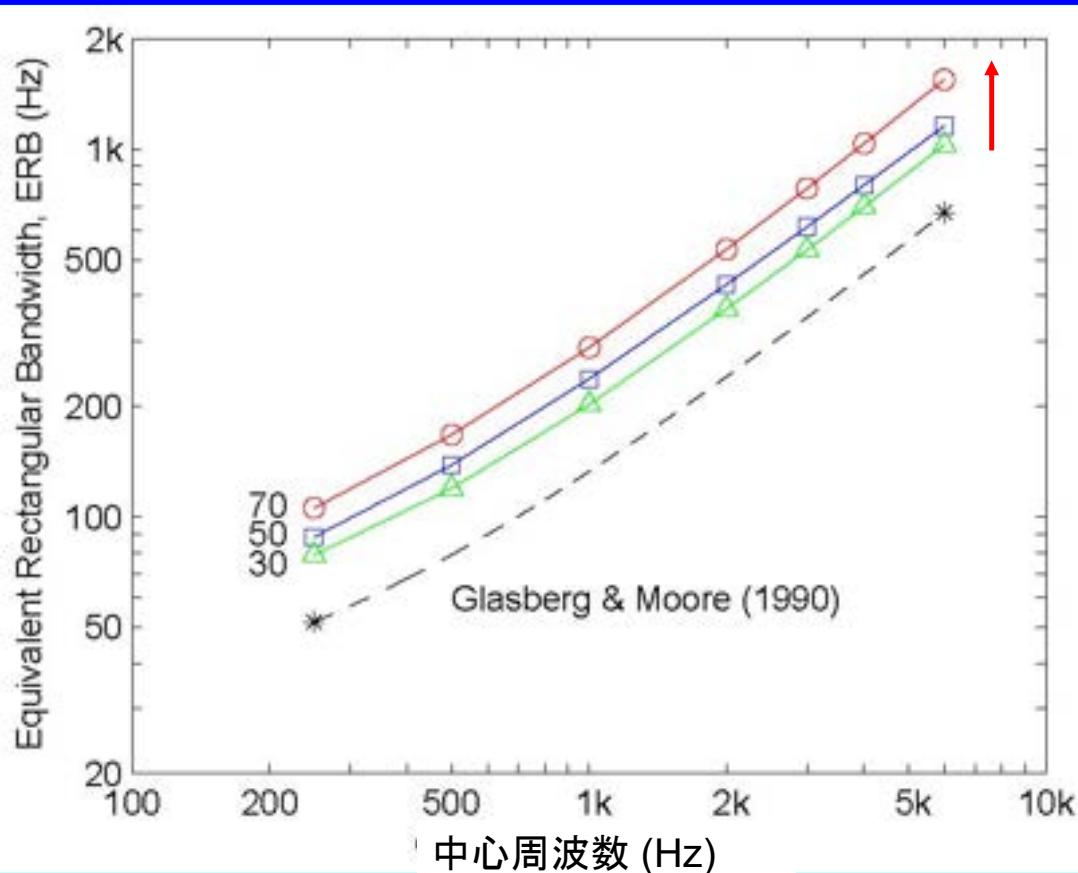
(e) 圧縮型ガンマチャープ



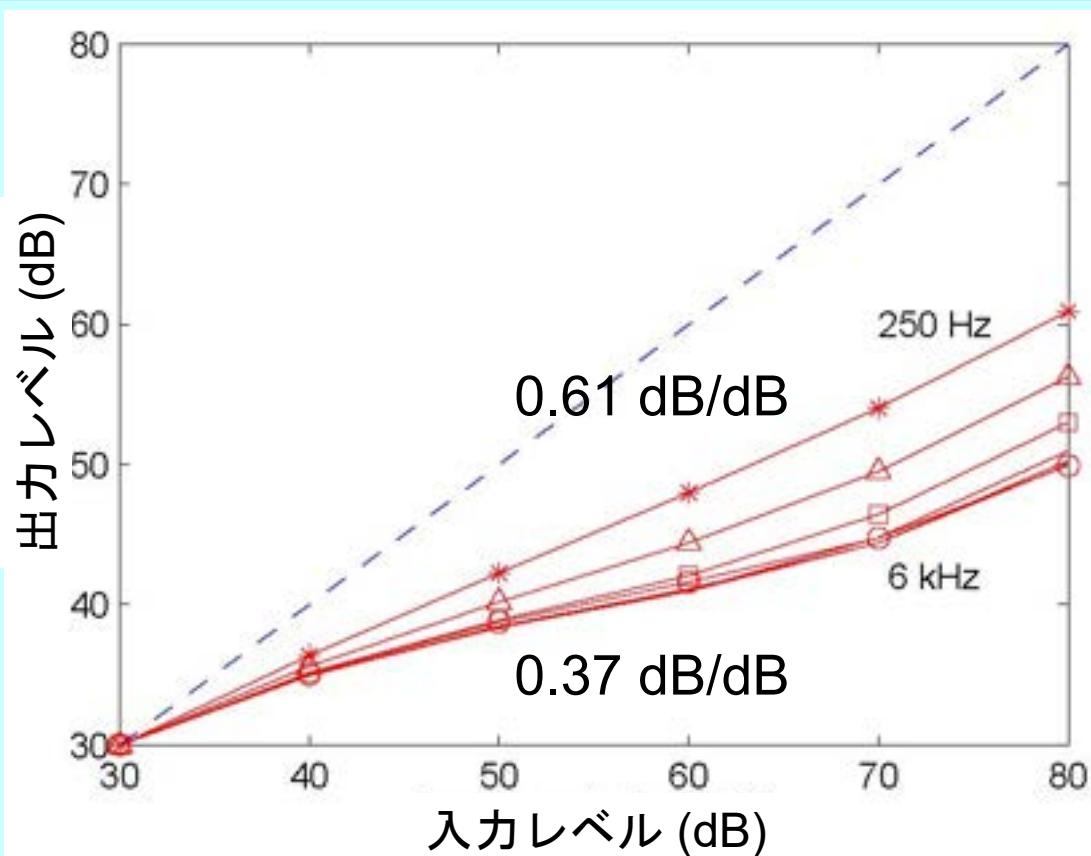
適合結果 (Patterson, Unoki, and Irion, 2003)



フィルタの等価矩形帯域幅(ERB)



フィルタのI/O特性

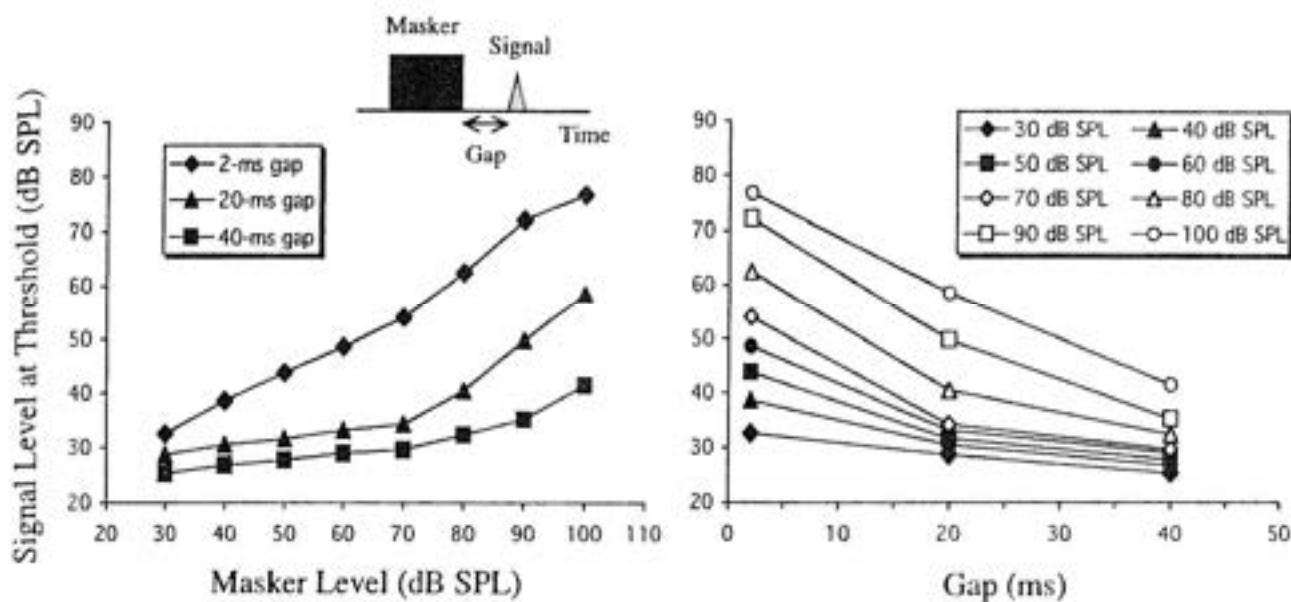


要約

- 圧縮型ガンマチャープ聴覚フィルタの適合
 - 少ないパラメータでマスキングデータを説明可能
 - フィルタ形状(帯域幅): レベル依存
 - 圧縮特性(利得の減少):
 - 周波数依存性の導入:
未計測の様々な周波数のパラメータを決定可能
- 残された課題
 - 時間特性、
 - 非同時マスキングの説明

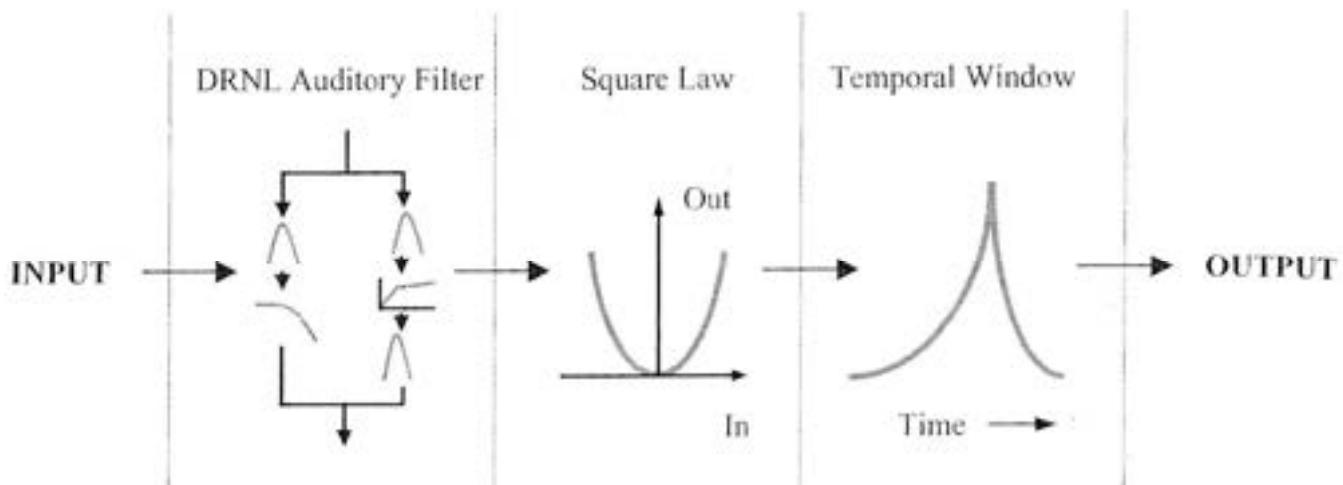
4. 順向性マスキング

- ギャップ検出
- 順向きにおける信号の検出レベル

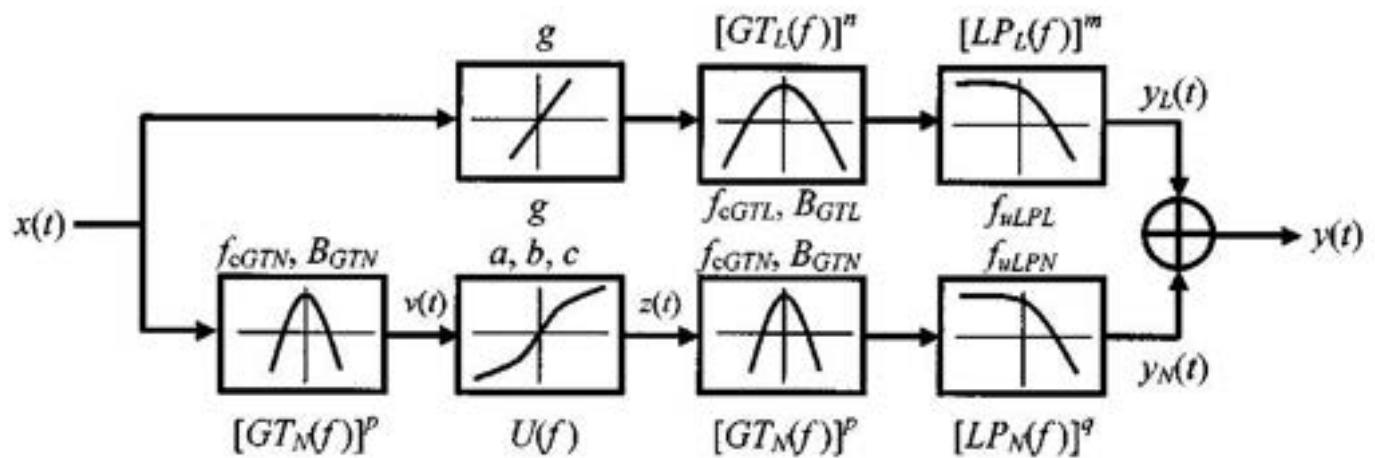


説明モデル

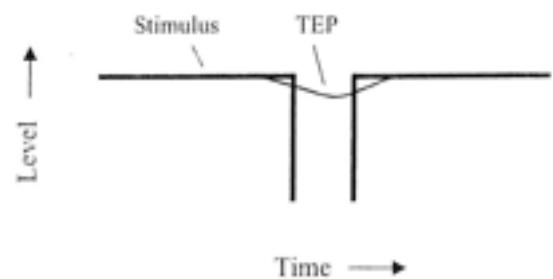
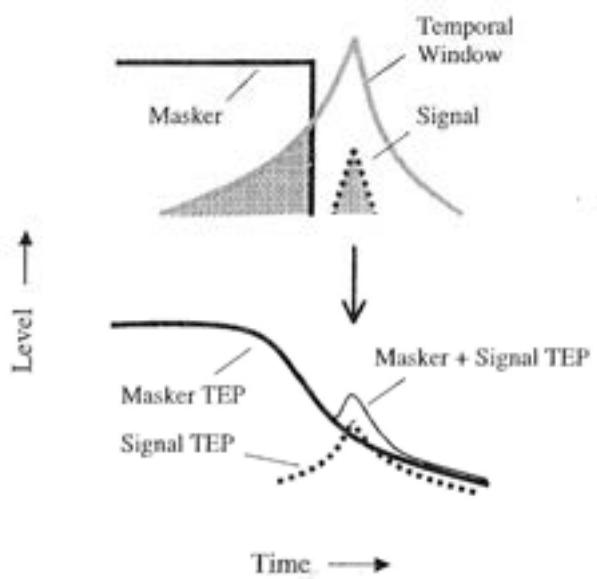
- (1) DRNL (double-resonance non-linear) auditory filter
- (2) Square law
- (3) Temporal window



DRNL聴覚フィルタモデル



時間窓の導入



5. 歴史的な概観

■ 場所 (Place) 説 と 時間 (temporal) 説

■ Ohm (場所説) vs. Seebeck (時間説)

Fourier theorem

■ Ohmのpure tone: f in Hzのピッチ $\Rightarrow \sin 2\pi ft$

■ Seebeck: パルス列のピッチ知覚 (*periodicity*)

■ Helmholtzの支持 : Ohmの法則

複合音のピッチ知覚

蝸牛フィルタリング \Rightarrow *Resonance* (場所説)

■ 生理学的なアプローチ

■ von Békésy

■ 基底膜の進行波

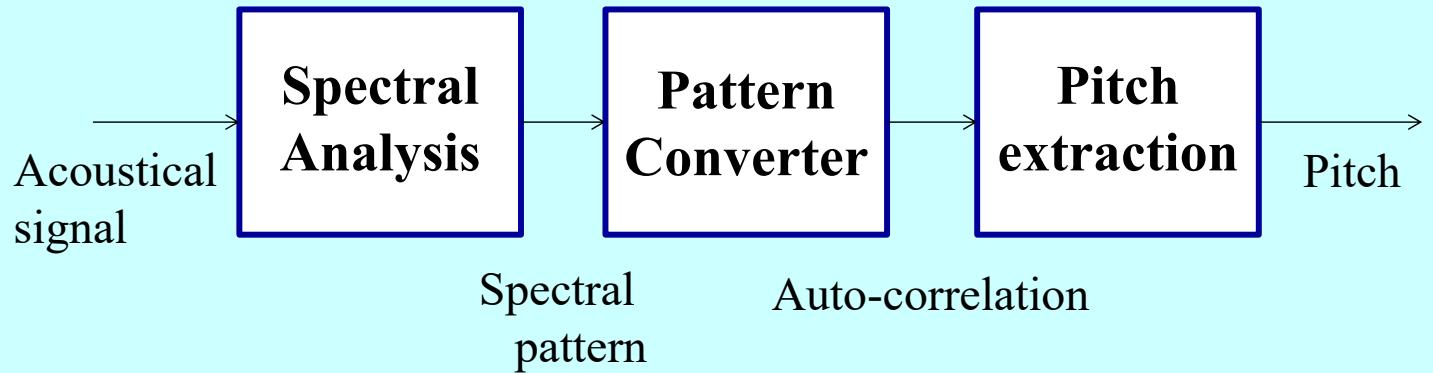
■ 蝸牛フィルタリング, 基底膜, コルチ器の調査

■ 数学的なアプローチ

■ Zwislocki, Patterson & Bogert, Fletcher

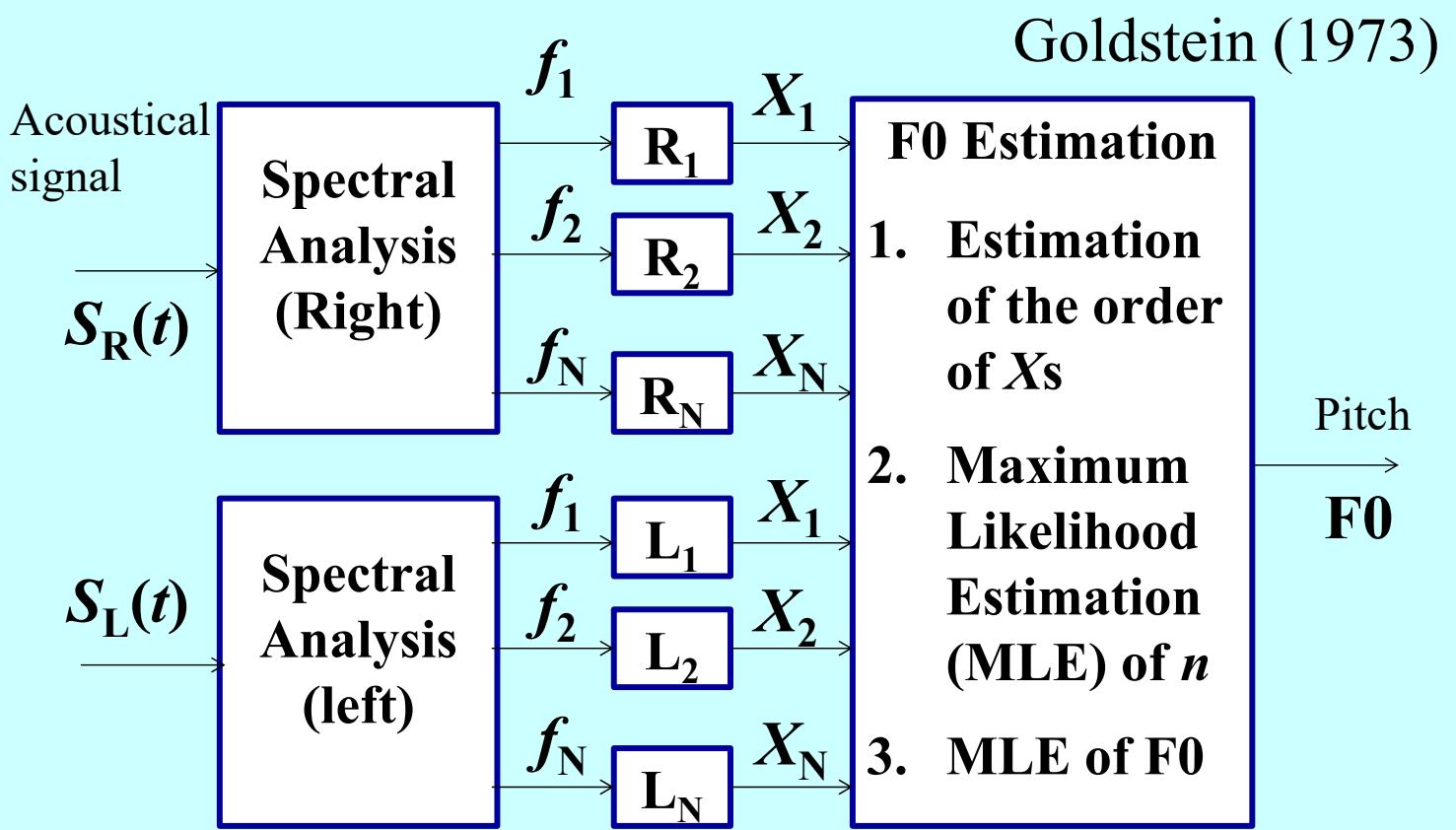
パターン変換モデル

Wightman (1973)

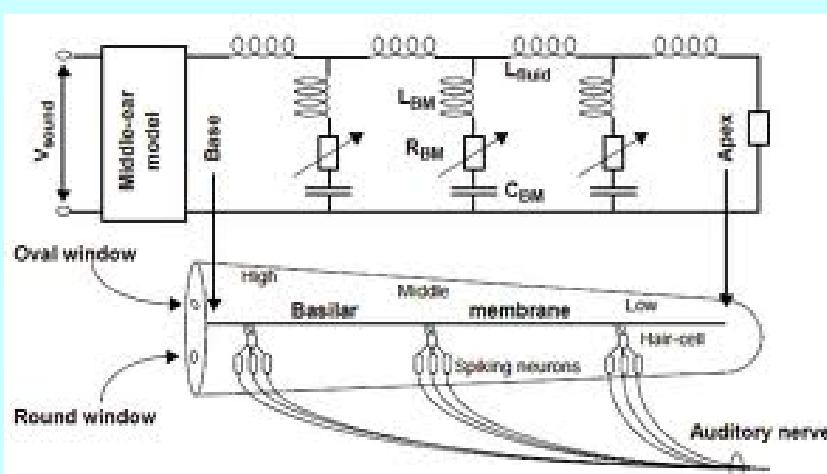
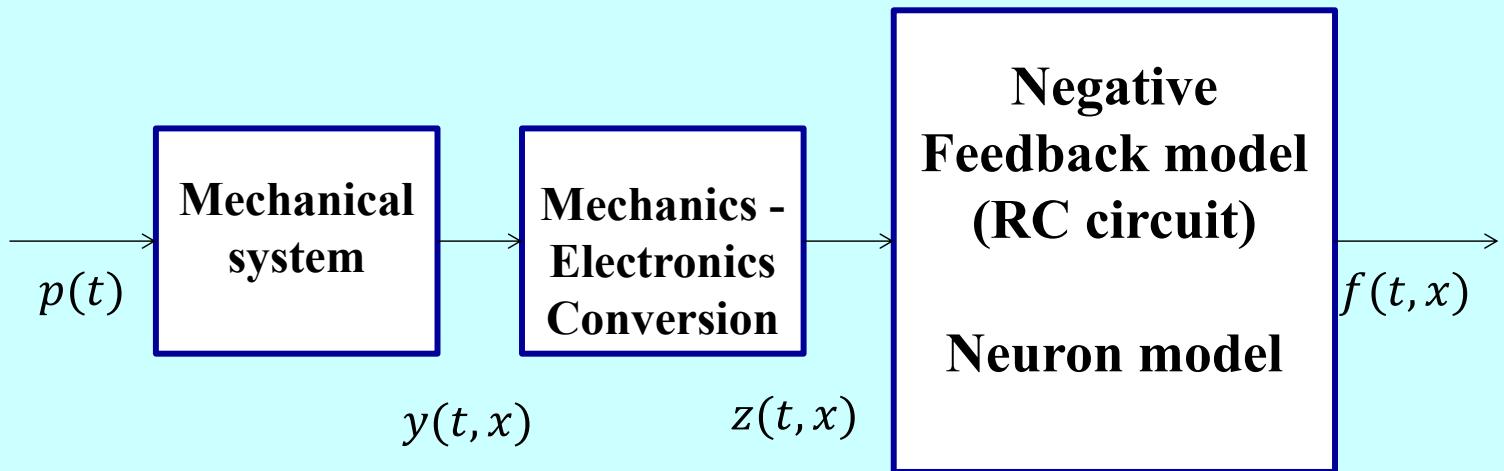


最適処理モデル

Goldstein (1973)

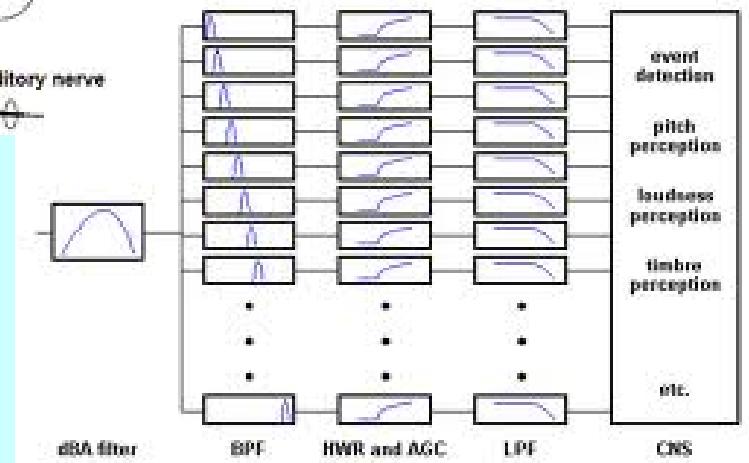


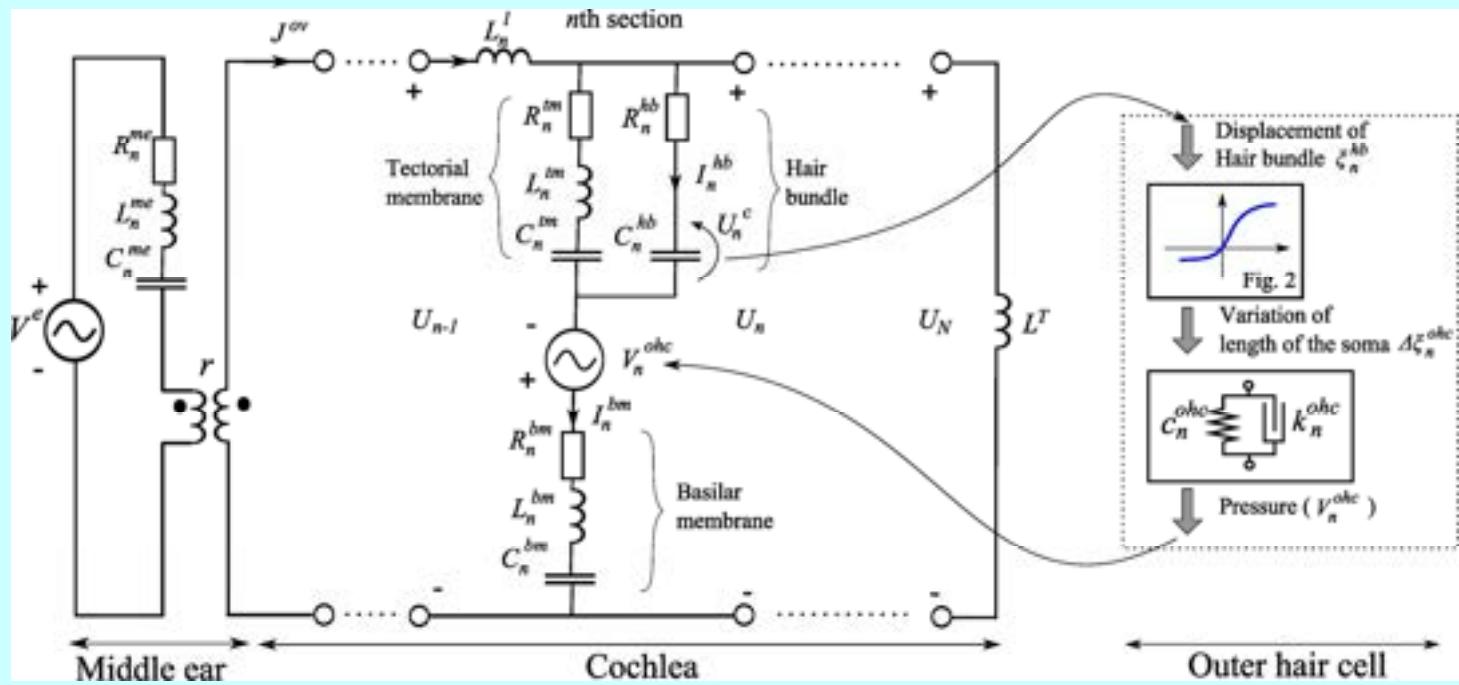
6. 生理学的モデル



フィルタバンクモデル
(Filterbank model)

伝送線路モデル
(Transmission model)

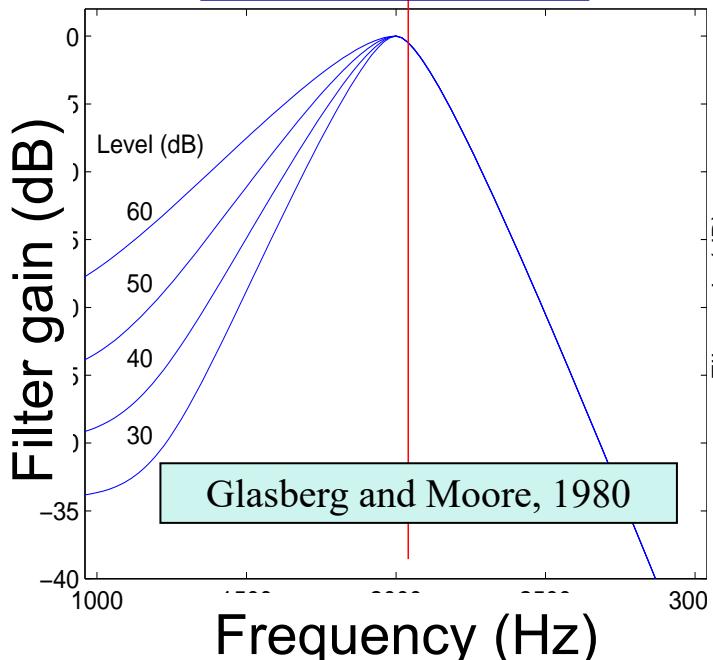




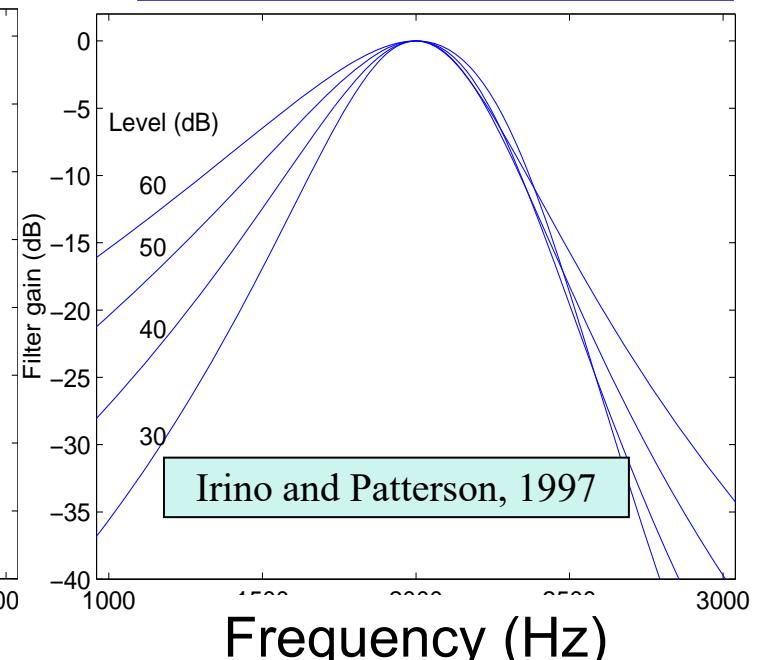
7. 心理物理的アプローチ

レベル依存／非対称性

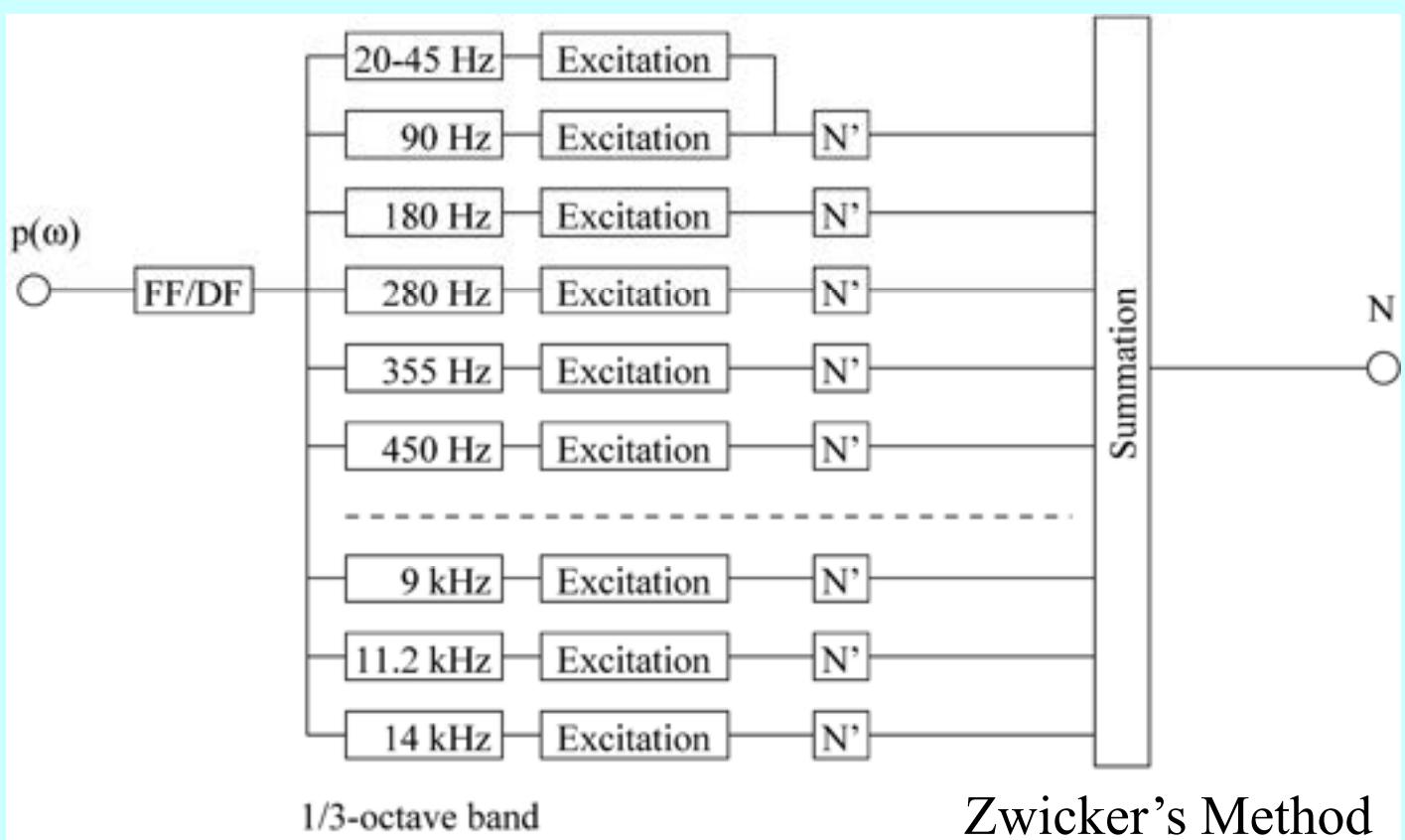
Roex filter



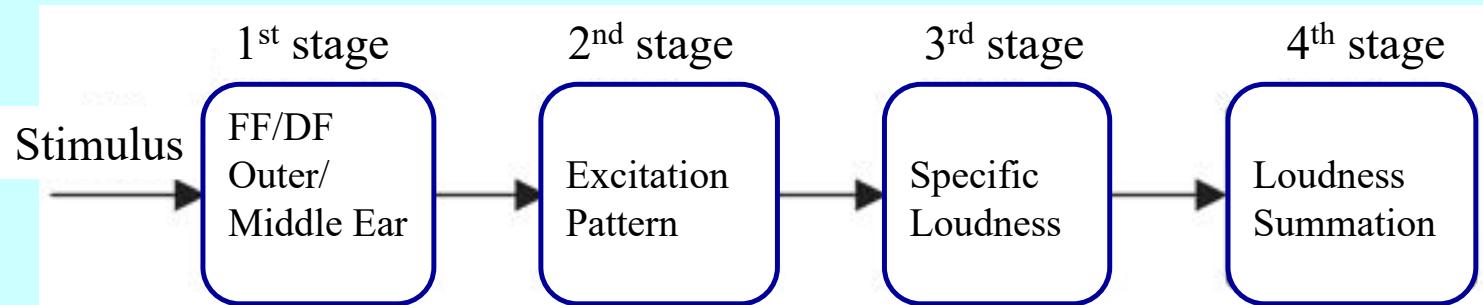
Gammachirp



Zwickerモデル



Moore & Glasbergモデル



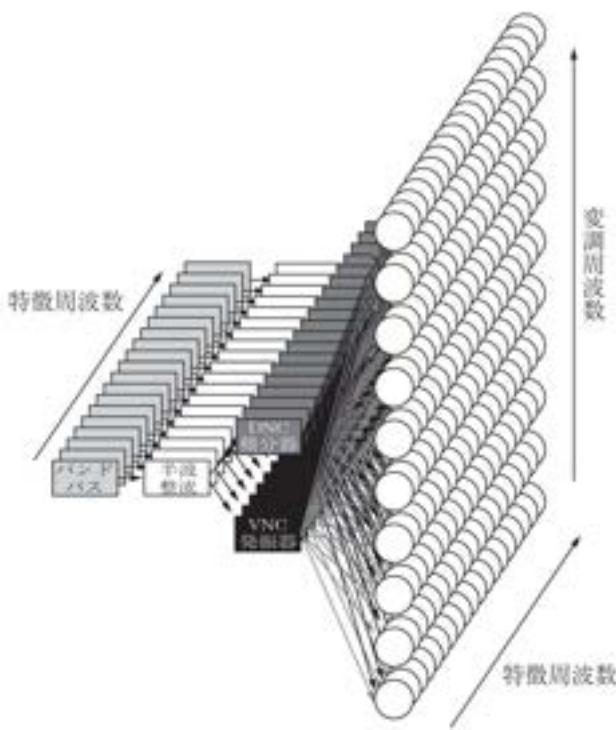
- 自由音場から鼓膜までの伝送系フィルタ
- 中耳の伝送系フィルタ
- エキサイテーションパターンへのスペクトル変換
- エキサイテーションパターン→ラウドネス密度
- ラウドネス密度における面積計算

8. モデルとシミュレータ

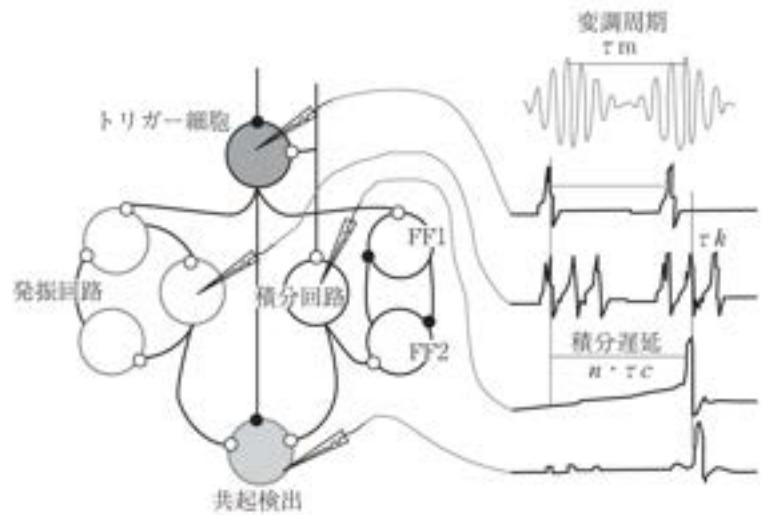
- 聴覚末梢系における周波数選択性の解明
- 聴覚の処理過程
 - 気圧変動→機械振動→基底膜振動→神経信号
 - 蝸牛における機械的な周波数分析
 - 外有毛細胞による能動的な処理
 - 内有毛細胞・聴神経における神経信号
- 「聴覚的」という意味
- 「**蜗牛における周波数分析はフーリエ分析的ではなくウェーブレット的分析である**」

9. Langnerの周期性検出モデル

- Langnerの発振・遅延による周期性検出モデル
- Missing fundamental
 - AM音のピッチ知覚：搬送波 1000 Hz, 変調周波数100 Hzで振幅変調 → 800, 1000, 1200 Hzの高調波
(講義資料 I458F-11.pdf 参照)
- 同期性検出ネットワーク (PAN, periodicity analyzing network)
 - 周波数成分 (特徴周波数)
 - 周期性 (変調周波数)



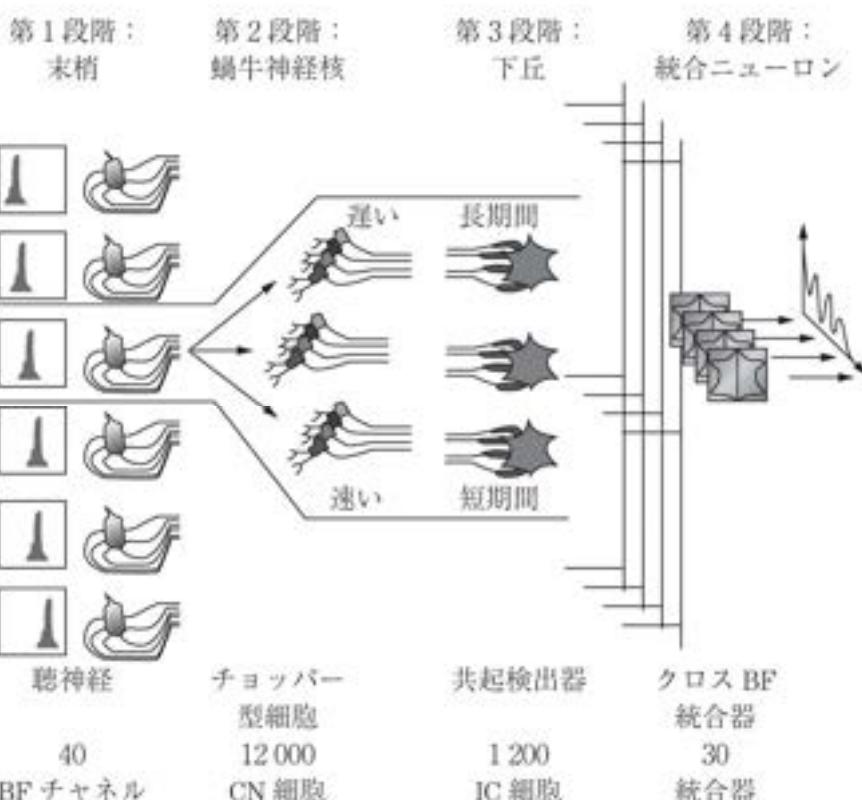
PANの模式図



神経回路的実現

10. Meddisの変調周期検出モデル

- Meddisのチョパー型細胞による変調周期検出モデル
- Lickliderの発想に基づく聴神経発火の時間パターンの自己相関によるピッチ検出機構
⇒Meddisらによる内有毛細胞の貯蔵庫モデルのシミュレータによる現実性と精度が高まった
- チョッパー型細胞の変調伝達特性と周期性検出
- MacGregor型の積分発火細胞の膜電位の変化に着目
 - ①静止電位からの逸脱として測定される膜電位,
 - ②カリウムイオンの伝導率, ③時変の閾値,
 - ④神経発火の有無を決定する制御変数



■ 第一段階

末梢の基底膜における周波数分解と蝸牛神経の活動

■ 第二段階

チョッパー型細胞による振幅変調刺激の変調周期検出機構

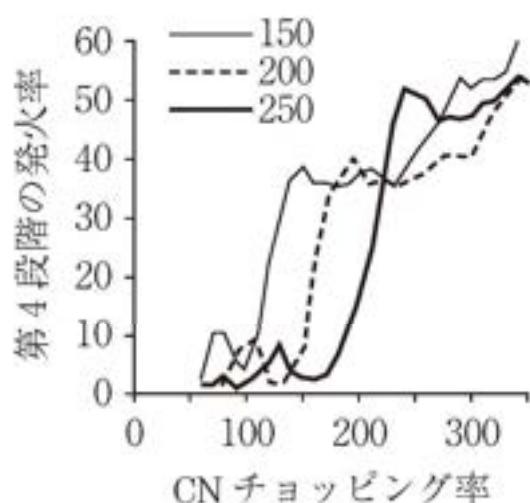
■ 第三段階

VCNの下丘の共起検出

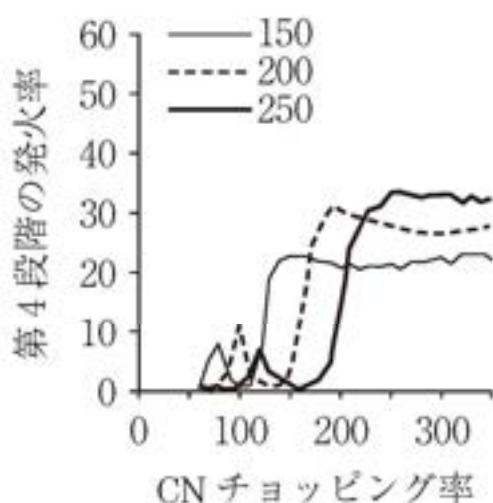
■ 第四段階

同一の最適変調周波数をもつモジュール出力を最適周波数間で統合

分解された高調波の場合



分解されない高調波の場合



講義資料 I485F-07.pdf 参照

公開パッケージ (Windows/Linux)

DSAM (development system for auditory modeling)

11. Pattersonの聴覚イメージモデル

■ 音色のモデル

時間的な非対称性の知覚

正弦搬送波に指数関数的な減衰包絡特性を持たせた繰り返し音 (damped sounds) とその時間反転 (ramped sounds) の間に存在する音色の違い

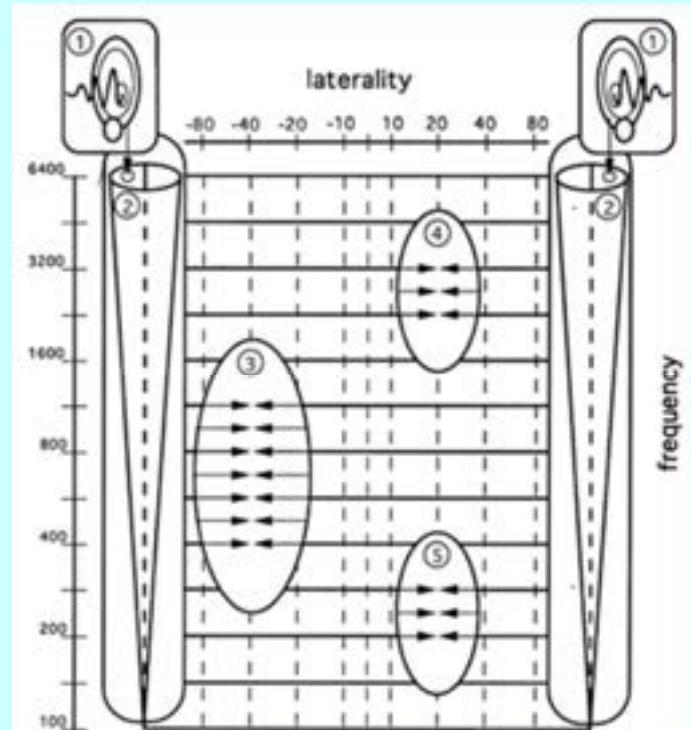
(講義資料 I485F-12.pdf 参照)

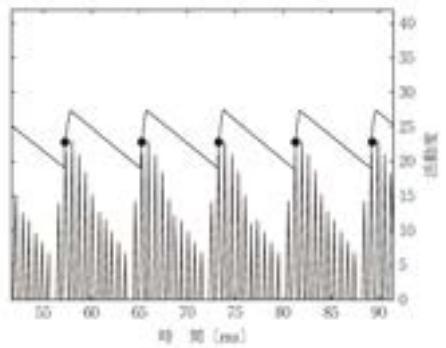
- ガンマトーン／ガンマチャーブフィルタ
- ストローブ時間積分 (strobed temporal integration, STI)
- 安定化聴覚像 (stabilized auditory image, SAI)
- 公開パッケージ aim (auditory image model)

Auditory Image Model

<http://www.pdn.cam.ac.uk/groups/cnbh/research/aim.php>
A schematic of the frequency-laterality plane of auditory perception. The first stage of image construction involves the spectral analysis performed on a sound entering the ears at a given point in time.

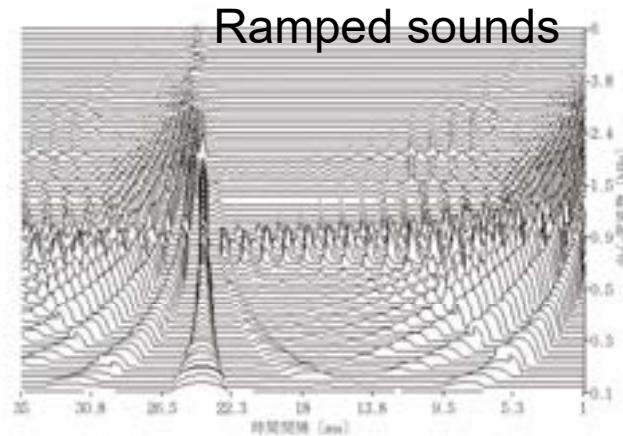
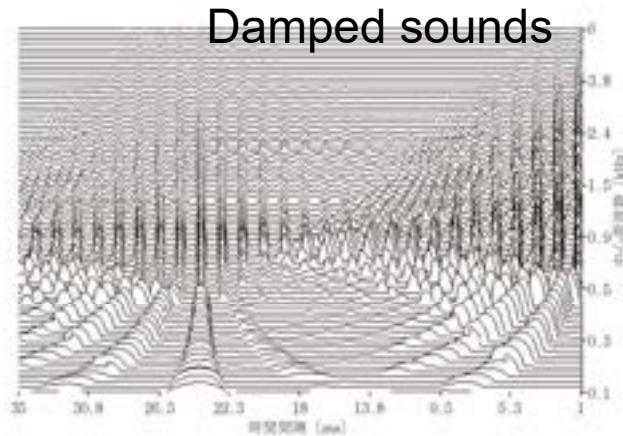
Right panel shows a schematic overview of this analysis, and the arrival-time analysis.





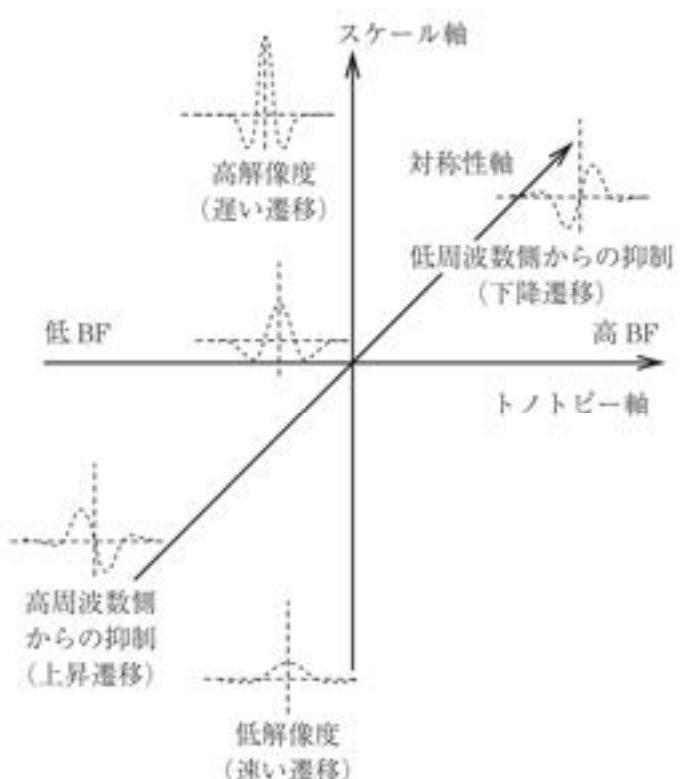
← 神経活動パターン (NAP)
におけるストローブ検出

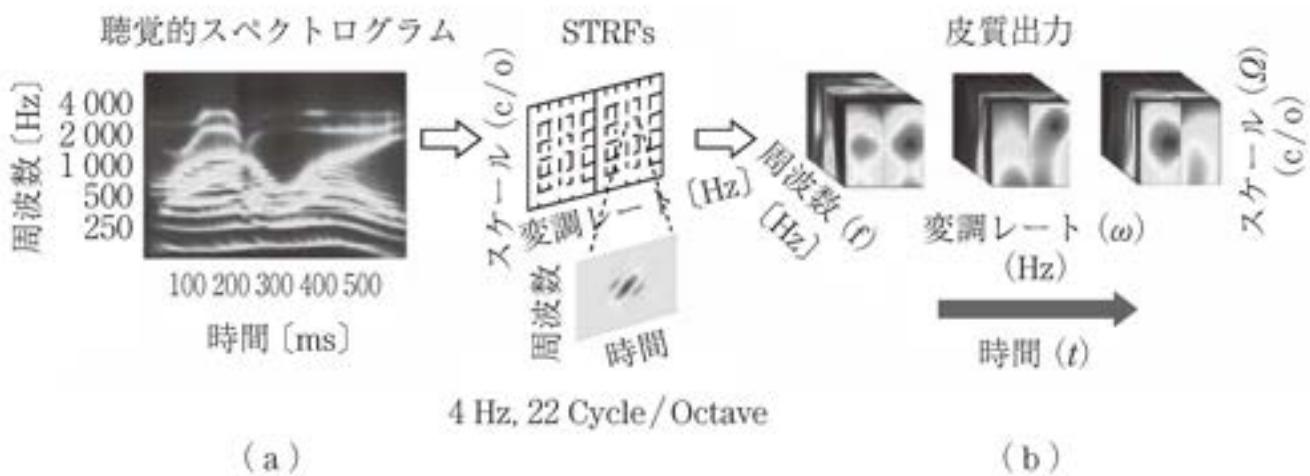
damped音/ramped音に対するSAI ↓



12. Shammaの聴覚皮質応答モデル

- 側抑制ネットワーク
側抑制結合の役割に焦点を当てたもの
- 皮質における受容野
視覚第一野と類似した多層柱状構造の発見
- 音声の表現
 - 調波成分 ⇒ 高解像度 フィルタの出力
 - 緩やかなフォルマントの変化 ⇒ 低解像度

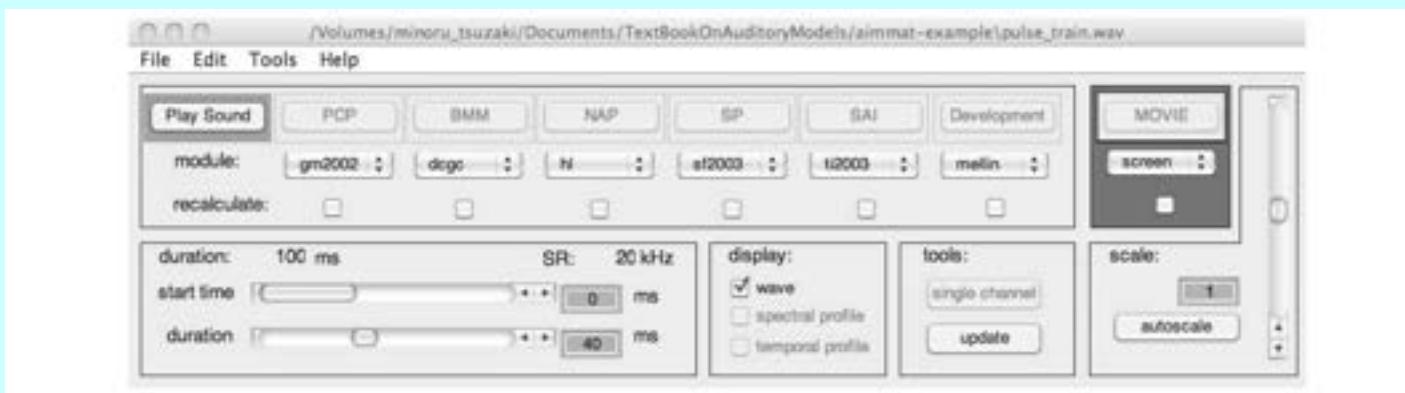




公開パッケージ : NSL Matlab toolbox
<http://www.isr.umd.edu/Labs/NSL/nsl.html>

13. AIMのデモ

- <http://www.pdn.cam.ac.uk/groups/cnbh/aim2006/>
- READMEを参照



- Development system for auditory modelling (DSAM)
- <http://dsam.org.uk/>