

聴覚末梢系の準同形处理的解釈とホルマント抽出モデル

高木, 英行
松下電器産業株式会社中央研究所

小原, 和昭
松下電器産業株式会社中央研究所

<https://hdl.handle.net/2324/4481567>

出版情報 : Transaction on Technical Committee of Psychological and Physiological Acoustics. 90 (55), pp.1-7, 1990-11. 日本音響学会聴覚研究会
バージョン :
権利関係 :



聴覚末梢神経系の準同形処理的解釈とホルマント抽出モデル

Auditory Model Based on Homomorphic Signal Analysis
and Formant Extraction Model.

高木 英行 小原 和昭

Hideyuki TAKAGI Kazuaki OBARA

松下電器産業（株）中央研究所

Central Research Laboratories

Matsushita Electric Industrial Co., Ltd.

あらまし 中枢でのホルマント検出と末梢系での生理的事実を信号処理的に解釈し生理的なリフタ機能の存在を仮定することで、聴覚末梢系から中枢での信号処理を準同形処理系と解釈するモデル提案をする。モデルは聴覚末梢の生理的事実を信号処理の観点より解釈したstageIと、基底膜出力間の加重加算的な相互作用を仮定した仮説機能のstageIIとからなる。このモデルによりピッチとその高調波の影響を抑えたスペクトログラムを得た。またこの聴覚モデルとニューラルネット（NN）によるホルマント抽出の結果を述べ、さらにモデルの他の音声特徴量抽出への適応について考察した。

1. はじめに

音声認識ではHMMやニューラルネット(NN)などの認識方式の検討が盛んであるが、いずれの方法でも特徴量によって認識性能は大きく影響を受ける。特徴量の研究では、LPCパラメータに代表されるパラメトリックな方法と、聴覚モデルに代表される方法とがある。前者は数理的な把握がしやすいという特徴がある反面、一度モデル限界に達するとそのブレークスルーは大変困難になる。一方、人間は音声を聞き取るに足る特徴抽出を行なっているわけで、この人間の聴覚機能を順次解明していくという意味で、後者には究極的に人間並の特徴抽出を実現できる可能性がある。しかし、研究方向の信頼性はあるが、生理的・心理的解明に歩調を合わせる必要があり、現在までのところ急速な進展は難しい状況にある。

本研究は後者の立場で音声処理のブレークスルーを図ろうとするものである。

聴覚モデルの研究では、既に明らかになっている生理的・心理的知見を忠実に工学モデル化

していくボトムアップ的手法を採る限り、神経生理や聴覚心理の進展を越えた展開ができ難い。

本研究では、生理的に未解決な点に対して仮説を導入して大きなフレームワークを与え、この大枠を基にモデル化研究の進展を図ろうとするトップダウン的方法に基づくモデルの提案を行っている。ちょうど理論物理と実験物理のような関係で聴覚モデル研究が進んで行けば、モデルサイドから生理学サイドへの貢献も期待できると考える。

2. 信号処理的観点に基づく聴覚モデル

ここでは聴覚末梢系の機能を信号処理理論の観点から考察し、末梢系から中枢に至る処理を準同形処理[1]として捉えた聴覚モデルを示す。本提案聴覚モデルは、生理的事実に基づいたStage Iと、仮説機能のStage IIとから構成される(第1図)。

2. 1 聴覚末梢系の生理学的知見と信号処理
聴覚末梢系に関する生理学的知見として本研

究では基底膜の周波数分解機能、有毛細胞の半波整流特性と飽和特性 [2]、シナプスや軸索での低域通過フィルタ (LPF) 特性 [3] [4] を考慮してモデル化を行った。

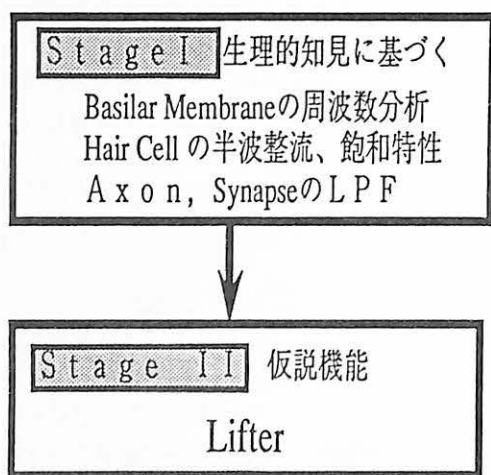
基底膜では外耳、中耳を通して入力される音声の周波数分析が行われる。

基底膜は場所によって異なる特徴周波数を持っている。また周波数選択特性は低周波数の方が高周波数よりもなだらかな減衰特性を持っている。このような基底膜の特性をここでは、減衰特性が低域側と高域側で非対象な聴覚特性を反映した 190 ch の臨界帯域フィルタバンク (CBF) [5] で表現する。この CBF の特性の一例を第 2 図に示した。

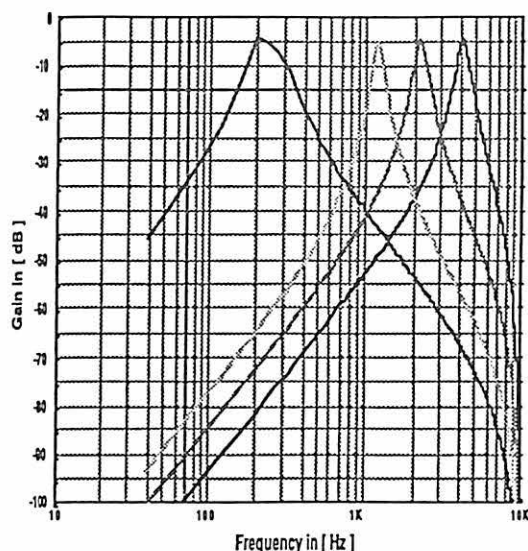
基底膜で周波数分析された信号は有毛細胞に伝えられる。

有毛細胞は基底膜の変位に対して第 3 図に示したような特性をもつ [2]。有毛細胞は基底膜の変位に対して半波整流特性を持ち、また基底膜の変位が大きくなると有毛細胞の出力 (受容器電位 receptor potential) は飽和する。

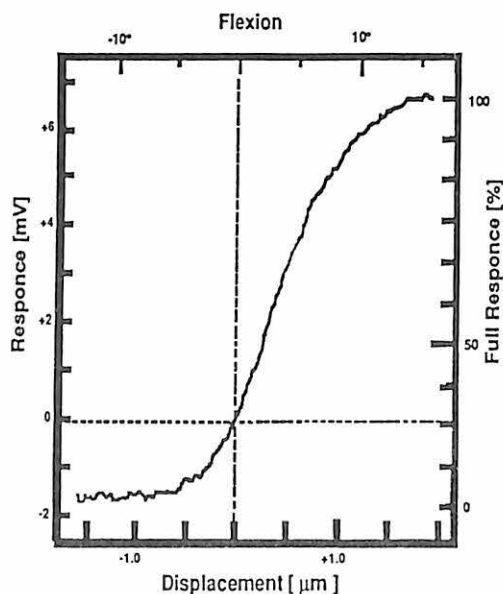
この有毛細胞の出力は求心性神経繊維にパルス列を発生させ信号の伝達を行う。このパルス列は軸索やシナプスの低域通過フィルタ (LPF) 特性を受けながら中枢へと送られる。この



第 1 図 モデルの概要
Fig.1 Outline of the Proposed Model



第 2 図 CBF の特性の一例
Fig.2 Sample characteristics of Critical Bandpass Filter.



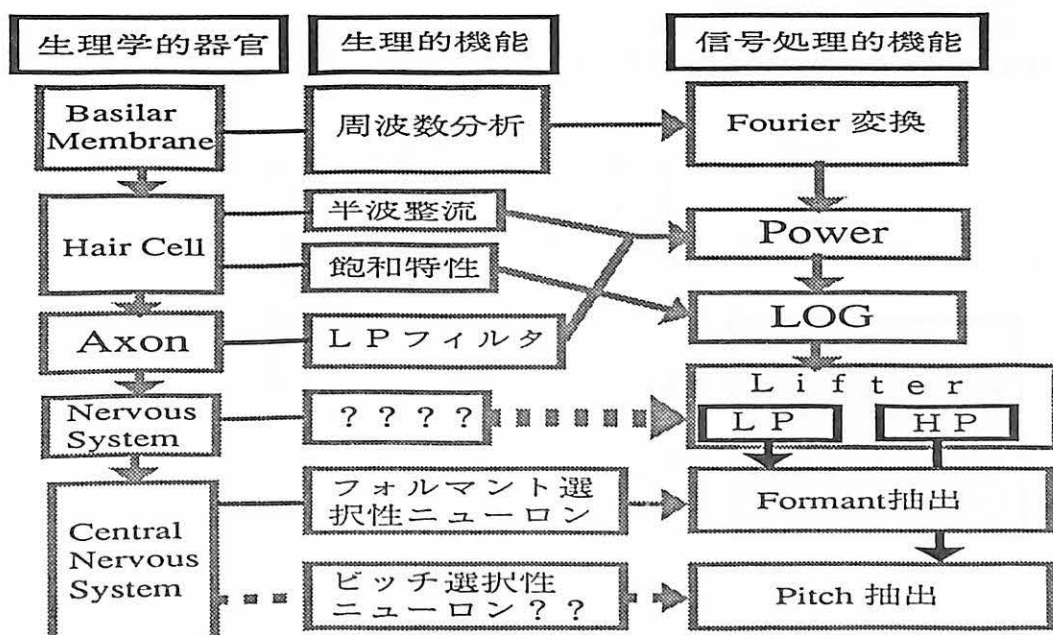
第 3 図 有毛細胞の特性 (After Hudspeth & Corey 1977)
Fig.3 Hair Cell Function (After Hudspeth & Corey 1977)

L P F特性は、ニューロン自身の不応期のために高周波成分には波形に同期したインパルスを出さないことを表わしたものであり、生理的観測でこのL P F特性がよく表れている [6] [7]。この同期発火性の現象は、Hodgkin-Huxleyの式がL P F特性を示している [4] ことから軸索で生じていると考えられる。その他にも、例えば、シナプスのイオン拡散によって生じる [6] とも言われている。また基底膜変位に対する有毛細胞の出力周波数特性は一次の低域通過型フィルタで近似できることが示されている [3]。

2. 2 信号处理的解釈に基づく聴覚モデル
一般の聴覚モデルの研究では以上に述べたような [半波整流特性→飽和特性→L P F特性] のような個々の機能を順次モデル化を行なうものが多いが、本研究ではこのような聴覚末梢系の信号処理機能を次のように工学的な信号处理的観点に立ち解釈する。すなわち、有毛細胞の半波整流の後で軸索やシナプスのL P F特性を

通すことは包絡線検波であり、パワー計算に相当すると解釈できる。そこで [半波整流+L P F特性] の代わりに2乗パワーを計算し、S.Seneffのモデル [6] に代表される (C B F + 半波整流 + 飽和特性 + L P F) のような個々の末梢機能をモデル化する代わりに (C B F + 2 乗パワー + 飽和特性) を聴覚モデルのStage Iとする。

次に、提案聴覚モデルの仮説機能に基づくStage IIについて述べる。ネコの聴覚野ではホルマントに選択的に反応するニューロンが見つまっている [8] ことから、聴覚末梢系から中枢系にかけてホルマントを抽出しやすくする機能が存在すると考えられる。ここでStage Iの機能を信号处理的観点から見てみると第4図のように捉えることができる。すなわち「末梢系機能とは [フーリエ変換 (C B F) + パワー (半波整流 + L P F特性) + 対数 (飽和特性)] である」。このような聴覚末梢での信号処理は工学的な信号処理における準同型処理の一部であると解釈することができる。従って次の仮説を立てれば



第4図 聴覚モデルと聴覚末梢機能の対応

Fig.4 Correspondence of Proposed Model and Auditory Peripheral Functions

末梢系でのホルマントを抽出しやすくする処理が説明できる：

ホルマントを抽出しやすくする機能とは周波数領域のLPF、つまりLPリフタであり、末梢系から中枢系にかけて準同形分析が行なわれている。

このStage IIのLPリフタ機能は、例えば基底膜に沿った聴神経間に加重加算的な相互作用の存在を仮定すると説明できる。加重加算的な処理は周波数特性としては低域通過フィルタと考えることができるので、基底膜に沿った出力、すなわち異なる特徴周波数をもつ聴神経の信号が加重加算的な相互作用をしながら中枢へ送られると考えると周波数方向のLPFつまりLPリフタ機能を説明できる。

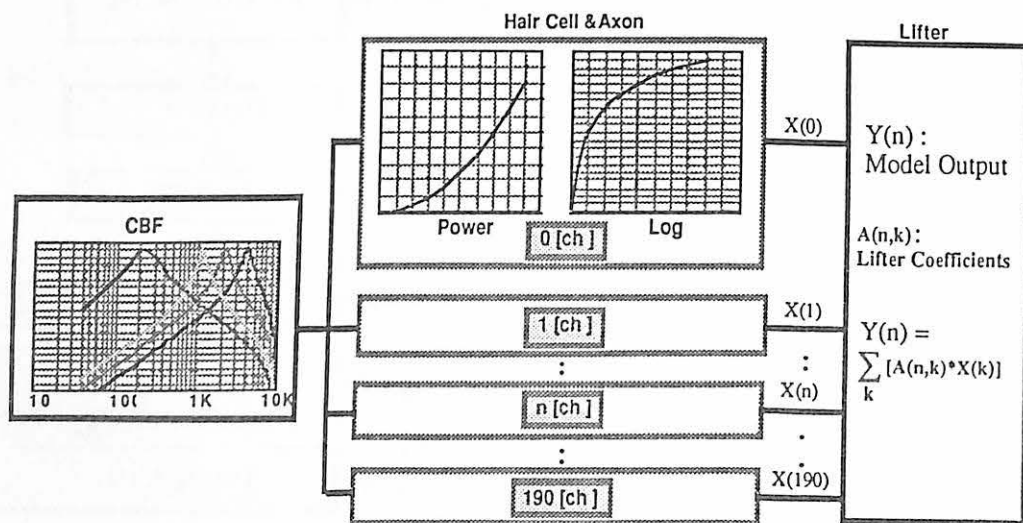
以上より、聴覚系でのLPリフタ機能の存在を仮定し、この仮説機能を組み込んだ第5図のモデルを本研究のホルマント抽出のための聴覚モデルとする。

この聴覚モデルの各レベルでの男声、女声に対するスペクトログラムを示したのが第6図、

第7図である。各々の図において(1)はDFT spectrogram、(2)はCBF出力、(3)はstage I 出力、(4)はStage II出力、(5)は後述するNNの出力である。stage Iだけの出力(3)に比べ、仮説機能を加えたstageIIのモデル出力(4)は、ピッチおよびその高調波成分が抑制され、滑らかなスペクトログラムになっている。

2.3 提案モデルの拡張性

本提案聴覚モデルは、Stage IIの仮説機能部を変更することで、ホルマント抽出に限定せず他の機能の実現するモデルに拡張が可能である。例えば、ピッチ抽出モデルであれば、第5図のLifterにはHP lifterが考えられる。HPリフタを考えれば、ピッチ高調波が抽出される。人間は基本ピッチ f_0 がなくてもピッチ高調波だけから周期性ピッチ f_0 が知覚できる[10]ので、このStage IIはピッチ抽出をしやすくする処理をしていると言えよう。この後は、Lickliderのモデル[11]のように自己相関によってピッチ知覚をしているかもしれないし、複数の線スペクトルの組合せに反応する[8] 前述のホルマント選択に似た機能を持つニューロンが周期



第5図 本研究の聴覚モデル

Fig.5 Proposed Auditory Model including Hypothetic Lifter Function

性ピッチを知覚をしているのかもしれない。

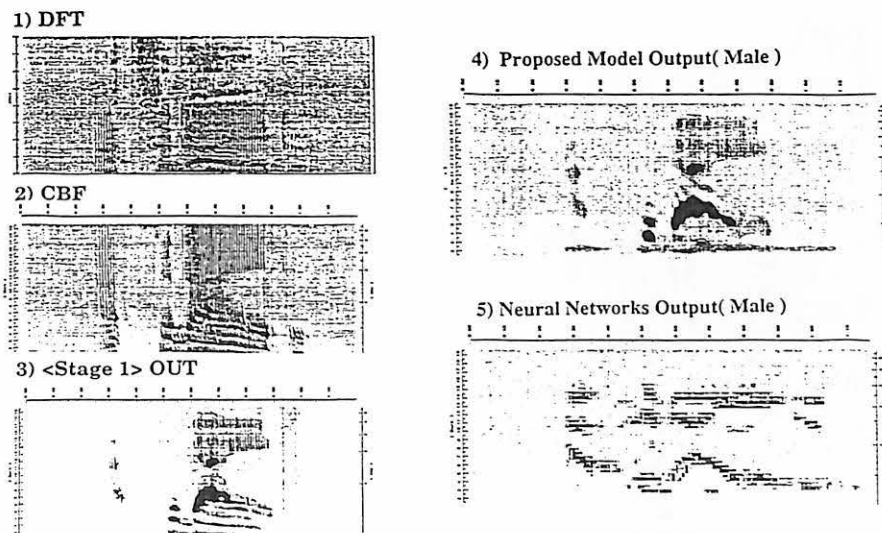
さらに、Stage IIのリフタリング特性をLPやHPにとらわれずに考えれば、その特性に応じた特徴量が得られるが、この特徴量が音声知覚に関わる特徴量であることも考えられる。したがって第5図は一般的な音声特徴量を抽出するモデルとしても拡張解釈が可能である。

3. ニューラルネットによるホルマント抽出

聴覚系では側抑制によりホルマントの強調処理を行っていると考えられる。ここではホルマント抽出を行う側抑制のモデルとして上に述べた聴覚モデル出力を用いたNNによる側抑制ホルマント抽出モデルについて述べる。筆者らは以前、3種類のNNモデルでホルマント抽出実験を行なった[9]。このNNでのホルマント抽出実験結果から、時間情報の必要性、相対位置情報の有用性、NNのprewiringの効果を示した。これらの結果と、不特定話者のホルマント

抽出への対応の2点を考慮し実験を行った。本実験のモデルは相対位置情報のみを用いるNNとし、前述の実験よりモデル入力の時空間サイズを拡大した。相対位置情報を用いたNNモデルは、入力層の各ユニットに特定の周波数成分に割り付けず、聴覚モデル出力の時空間をsweepし、入力された時空間情報のみから中心がホルマントかどうかを判断するよう学習したNNモデルである。このNN処理は時空間パターンの学習によりホルマントピークを選択的に強調するので、側抑制処理と言える。

実験に用いたモデルは3.5 bark $\times$ 50 msの時空間入力層(7 $\times$ 5 units)を持つ。NNはfeed-forward typeで、入力時空間パターンの中心にホルマントが存在する場合に発火するよう学習させた。本実験では男性話者1名が発声する80単語から得られた173,130個の学習データを先に示した聴覚モデルに入力し、その出力をこのNNに入力してホルマントを抽出した。



第6図 モデルとNNの出力(男性話者)

- (1) DFT spectrogram (2) CBF 出力 (3) Stage I 出力
(4) Stage II 出力 (5) 未学習の男性話者NN出力

Fig.6 Sample of NN output for open Male voice (1) Conventional DFT

(2) CBF Output (3) Stage I Output (4) Stage II Output (5) Output of formant extraction NN

学習に用いた男性話者のNN未学習単語の出力を第6図(5)示す。聴覚モデルのチャンネル間隔は0.1Barkであるが、NNは0.5barkを最小分解能になるよう設計している。従って、抽出結果の表示精度は5倍ほど粗くなっている。この点を考慮すると、学習データはもちろんのこと、未学習データに対しても高い抽出性能を示したといえる。さらに、ホルマント抽出NNを、女声でかつ未学習単語に適用した結果を第7図(5)に示す。未学習の女声に対してもホルマント抽出はほぼ良好に行なわれている。これは相対位置情報のみからホルマントを決定するようなNNモデルとピッチおよびその高調波を軽減するモデルを用いた学習の効果と考えられる。

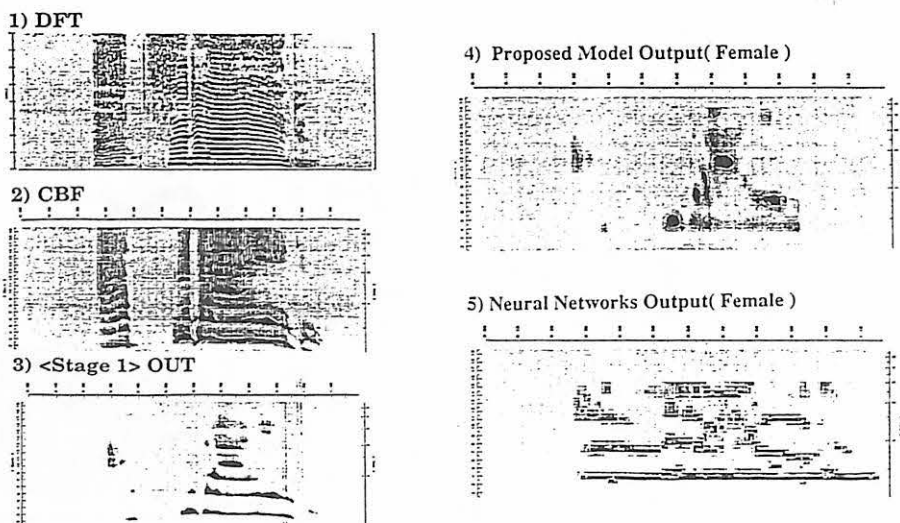
4. 考察

本提案聴覚モデルの特徴は2つある。第1の特徴は、聴覚末梢系の生理的知見を工学的に実現していく従来の聴覚モデル構築法ではなく、機能的に解釈して信号処理モデルに積極的に置

き換えていこうとする考え方である。本研究で提案したモデルでは有毛細胞の半波整流特性と軸索やシナプスなどで生じるLPF特性をまとめて、パワー抽出機能と解釈し、また有毛細胞の飽和特性を対数近似と考えることで聴覚系末梢での信号処理を準同型処理の一部とみなすことができることを示した。

第2の特徴は中枢系でのホルマントに反応するニューロンの存在と末梢系の知見の信号処理的解釈から、その中間にホルマントを抽出し易くする機能の存在を予想して準同型処理の観点からLPリフタ機能の存在を仮定してモデルを構築した点である。実際にこの仮説機能が実在するか否かは生理的知見の収集の進展を待たなければならないが、聴覚生理の分野にフィードバックできる可能性がある。

以上のようにリフタ機能の存在を仮定することで、信号処理的解釈に基づいたホルマント等の音声特徴量抽出のためのフレームワークが得られた。



第7図 モデルとNNの出力(女性話者)

- (1) DFT spectrogram (2) CBF出力 (3) Stage I 出力
(4) Stage II 出力 (5) 未学習の女性話者NN出力

Fig.7 Sample of NN output for open Female voice (1) Conventional DFT (2) CBF Output (3) Stage I Output (4) Stage II Output (5) Output of formant extraction NN

また本研究で用いたNNは聴覚モデル出力を入力としホルマントを強調して出力することから時空間の情報の学習によって形成される時空間側抑制回路と考えることができる。この側抑制回路の特徴は、従来よく用いられる周波数方向成分の演算によるスペクトルピーク情報の強調ではなく、音声のホルマント情報の時空間的な構造とパワーの情報を反映したスペクトルピーク情報の強調を行うことである。

5. まとめ

聴覚末梢系と中枢系の間にリフタリング機能の存在を仮定することで、聴覚末梢系の機能を準同形処理と解釈する聴覚モデルを提案した。ホルマント抽出を例にこのモデルの効果をシミュレーションで示した。このモデル出力をNNに入力しホルマント抽出を行った。このNNは時空間のホルマント情報を学習し、ホルマントを強調する側抑制回路として働いた。さらに提案モデルはホルマント抽出に限定されず汎用的な音声特徴抽出モデルとしての用いることのできる可能性を考察した。

本研究で述べた準同形処理は、非線形機能の入れ替えや対数近似などを行っており、工学で言う純粹の準同形処理とは言えないが、聴覚末梢系全体の機能の解釈基盤となる概念としては十分利用できる。さらに、本研究の提案モデルの中心となっている聴覚的なリフタリング機能の存在仮説は、工学サイドから神経生理サイドへの貢献につながり得る可能性を持っている。

参考文献

- [1] A.V.Oppenheim & R.W.Schafer: "デジタル信号処理(下)", (伊達訳) コロナ社 (昭53)
- [2] A.J.Hudspeth & D.P.Corey: "Sensitivity, polarity and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli", Proc. Natio. Acad. of Sci., U.S.A., 74, pp.2407-2411 (1977)
- [3] S.A.Shamma, et al: "A biophysical model of cochlear processing: Intensity dependence of pure tone responses" J. Acoust. Soc. Am. 80(1), pp133-145 (1986)
- [4] 堀川: "神経軸索上の分散関係によるフィルタ特性", 信学論誌D-II, J72-D-II, 4, pp.621-629 (平1)
- [5] 駒木根, 平原: "蝸牛の周波数分解機能を模倣するフィルタバンクの一構成法", 信学技報 SP87-45 (昭61)
- [6] S.Seneff: "A joint synchrony/mean-rate model of auditory speech processing", J. Phonetics, 16, 1, pp.55-76 (1988)
- [7] D.H.Johnson: "The response of single auditory-nerve fibers in the cat to single tones: synchrony and average discharge rate", Ph.D.Thesis of MIT (1974)
- [8] 丸山: "猫聴覚野(皮質)におけるホルマントの受容", 神経進歩, 25, 4, pp.926-933 (1981)
- [9] 高木, 坪香: "ニューラルネットを用いた疑似ホルマント抽出フィルタリング", 音響学会講演論文集 3-p-11, pp249-250 (1988)
- [10] 杉本: "音声のピッチ感覚および基本周波数抽出について", 研実報, 11, 2, p.327 (昭37)
- [11] J.C.R.Licklider: in "Foundations of Modern Auditory Theory", I, ed. J.V.Tabias, pp.47-48, Academic Press (1972)