

音源方向定位聴覚神経系モデルによる ITD,ILD の  
脳内マッピングの実現

黒柳 奨<sup>†</sup> 岩田 彰<sup>†</sup>

Auditory Pulse Neural Network Model for Sound Localization  
— Mapping of the ITD and ILD —

Susumu KUROYANAGI<sup>†</sup> and Akira IWATA<sup>†</sup>

あらまし 人間は、両耳間の信号の時間差 (ITD:Inter-aural Time Difference) と音圧差 (ILD:Inter-aural Level Difference) を用いて、音源の方向を知ることができる。我々は神経細胞の機能を模倣したパルスニューロンモデルを用い、2 信号間から ITD と ILD を独立に抽出する聴覚神経系モデルを提案している。今回、この聴覚神経系モデルの上層に新たなパルスニューラルネットワークを追加し、入力信号間の特定の ITD,ILD にそれぞれ特異的に反応するマッピングニューロン群を構築した。この新たな聴覚神経系モデルに対して、計算機上でシミュレーションを行った結果、入力信号間の ITD,ILD の変化によってマッピングモデル内での発火ニューロンの位置が変化することが確認され、本モデルが単純なパルスニューロンモデルの組合せのみで ITD,ILD をマッピング可能であることが示された。

キーワード 音源方向定位, 聴覚神経系, ニューラルネットワーク, パルスニューロンモデル

1. ま え が き

人間は二つの耳に入ってくる音の違いを利用して、音源の位置を知ることができる [1], [2]。この音源定位機構は、音の情報から周囲の空間を認識するための重要な機能であり、その仕組みを解明し工学的に応用することは人工聴覚システムの構築にとどまらず、ある特定の方向や特定の音源からの音だけを収集する集音装置、仮想空間における音源位置の再現など、多方面にわたってその実現が期待されている。音源の位置には音源の水平面内の方向、正中面内の方向、音源までの距離の三つのパラメータがあるが、ここでは水平面内の方向について考えることとする。

人間の音源定位機構は、心理学、神経生理学それぞれのアプローチから研究が進められており [1]~[3]、音源定位が両耳に入ってくる音の時間差 (ITD:Inter-aural Time Difference) と音圧差 (ILD:Inter-aural Level Difference) を手掛かりとして用いていること

や、ITD, ILD の検出に上オリブ核が関与していること、またこれら細胞核においては音の情報はすべて神経パルス列によって伝達されており、情報処理はこの神経パルスを入出力データとする神経細胞の組合せで行っていることなどが明らかになってきた。

これら心理学、神経生理学上の知見をもとに種々の音源定位モデル [4]~[8] がこれまでにも提案されているが、これらモデルには、計算コストの大きい相関値計算を用いている、音源方向の推定に用いられている階層型ニューラルネットワークはダイナミクスをもち、音のように時系列情報を扱うのには向いていない、ITD, ILD の検出と音源の方向の推定の三つをそれぞれ別のモデルを用いて実現している、などの問題点がある。

そこで我々は、パルスニューロンモデル (以降 PN モデルと省略) を用いた音源定位のための新しいパルスニューラルネットワークモデルを提案し、これにより聴覚系の MSO (the Medial Superior Olive), LSO (the Lateral Superior Olive) の機構を実現した [9]。本モデルは、帯域フィルタ群からなる蝸牛に相当する周波数分析部と各周波数成分を音圧に比例した発火

<sup>†</sup> 名古屋工業大学電気情報工学科, 名古屋市  
Department of Electrical and Computer Engineering,  
Nagoya Institute of Technology, Nagoya-shi, 466 Japan

頻度をもつパルス列に変換するパルス変換部、そしてPNモデルから構成されているMSOモデル、LSOモデルからなる。シミュレーションの結果、ITD、ILDをもつ二つの信号を入力した場合、MSOモデルが入力信号間のITDに、LSOモデルがILDにそれぞれ特異的に反応しており、このモデルを用いれば複雑な相関値計算を行わなくてもITDの情報を抽出可能であること、またITDとILDを基本的に同じモデルの上で抽出できることが示された。

しかし、このネットワークはITD、ILDの変化に対してモデルの発火パターンが規則的に変化するものであり、モデルの出力から直接入力信号間のITD、ILDを知ることはできない。そこで、今回このモデルの上位に、MSO、LSOモデルと同じPNモデルから構成される新たなネットワークを追加し、ITD、ILDの値を、発火するPNモデルの位置にマッピングするネットワークを構築した。そしてこの新たなネットワークを加えた聴覚神経系モデルに対して、計算機上でシミュレーションを行い、2入力信号間のITD、ILDの変化に対するモデルの動作を確認した。

本論文では、まず2. で現在までに得られている心理学的、生理学的知見とこれらを用いた既存の音源定位モデルを説明し、これらモデルの問題点とそれに対するパルスニューラルネットワークの利点を明らかにする。そして3.、4. で我々の提案するパルスニューラルネットワークを説明し、5. にて計算機上でシミュレーションの結果を報告する。

## 2. 聴覚系の音源定位機構およびそのモデル

人間の音源定位機構は、心理学、神経生理学それぞれのアプローチから研究が進められている。

まず心理学のアプローチからは、人間はITDとILDを手掛かりとして音源の水平面内の方向を検出していることがわかってきた[1]。ITD、ILDをもたない全く同一の信号を両耳に与えた場合、人間が心理的に認識する音源、つまり音像は正面にあるように感じるが、信号間にITD、ILDを与えてやると、音像は相対的に音の早く到達する方、若しくは相対的に音の大きい方に片寄る[10]。また、入力音として持続する純音を用いた場合、ITDは約1kHz以下の低い周波数帯の信号の音源定位に対して有効であり、ILDは頭部および耳介の影響によって1kHz以上の高い周波数帯の信号に対して有効であることが知られている[1]。但し、AM変調された信号の場合は搬送波が高い周波数帯の信号

であってもその包絡線のITDが音源定位に有効である[11]など、ITD、ILDの有意性と信号周波数の関係は単純なものではなく、二つの特徴量がお互いに補いあって音源定位に寄与している。

神経生理学のアプローチからも、次第に聴覚系の機構が明らかになってきている[1]~[3]。耳介から入った音は外耳、中耳を通して蝸牛に達し、蝸牛において液体振動に変換され、ここで周波数分析が行われる。周波数分析が行われた信号は蝸牛内にある有毛細胞によって各周波数成分ごとに整流、検波が行われ電気信号に変換される。そして、有毛細胞に接する蝸牛神経によって神経パルス信号に変換されて脳へと送られていく。

聴覚系において、最初に音源定位に関与する細胞核は上オリブ核(SOC:the Superior Olivary Complex)で、特にこのうちの内側上オリブ核(MSO)がITDの処理に、外側上オリブ核(LSO)がILDの処理に関与していると考えられている。フクロウの聴覚系のうち人間のMSOに相当する細胞核の中では、さまざまなITDに対して左右からの信号の同時到達を検出している細胞が見つまっている[12]。LSOでのILDの検出機構についてはまだ詳しいことがわかっていないが、多くの細胞が対側の耳からの入力で抑制を受けており、その抑制の度合いが規則的に変化していることから、ILDの大きさに応じてLSO内の発火する細胞の領域が変化することによってILDの大きさを検出しているのではないかと考えられている[2]、[12]。MSO、LSO以降の音源定位に関する機構についてはフクロウを用いた実験から、MSO、LSOの上層にあたる下丘という神経核に、ある特定の音源の方向に対応した神経細胞が規則的に並んでおり、脳内に聴空間が写像されているのが見つまっている[12]。

以上のような心理学、神経生理学上の知見をもとに種々の音源定位モデルがこれまでも提案されている[4]~[8]。水平面内の音源方向を推定するモデルで実現されている機能は大きく分けると次のようになる。

- a) 蝸牛における周波数分析と神経パルスへの変換
- b) ITDの検出
- c) ILDの検出
- d) 検出されたITD、ILDからの音源方向の推定
  - a) は中心周波数の順に規則的に並んだ一連の帯域フィルタ群とそれぞれのフィルタの出力に対する整流作用、および神経パルスへの変換部分から構成される。しかし実際にパルスを発生してそれによってITD、

ILD の評価を行うモデルはなく、多くのモデルは神経パルスの発火確率の時間関数または帯域フィルタの出力で代用をしている。b) については、ほとんどのモデルは Jeffress の一致モデル [13] を基本としている。このモデルにおいて各ニューロンは左右の耳からのパルスの同時入力を検出し、左右からのパルス信号は時間遅れを伴ってそれぞれのニューロンに到達する。このため ITD を発火ニューロンの位置として検出することができる。但し、既存のモデルでは神経パルスを発火確率の時間関数で表しているため、実際には各ニューロンはそれぞれ特定の ITD に対する 2 信号の相互相関を計算することになる。しかしこの方法には各ニューロンごとに畳込み積分が必要となり、その計算コストの大きさが問題となる。Shamma のモデル [6] は、蝸牛内の進行波の伝搬の遅れを利用することで Jeffress の一致回路を用いることなく ITD の検出を行っているが、特徴量の検出にはやはり相関値計算を必要としている。c) に関しては ITD の検出手法ほど確立されたものはまだなく、ILD 検出機構を ITD 検出機構を補完するものとするもの [5] や、ITD の検出に伴って ILD の検出も可能であるとするもの [6] もある。ILD の計算方法としては各周波数成分ごとにある時間区分内のパワースペクトルの差を求める方法がある [7]。また ILD を特徴量として用いているが ILD は計算せず、次に説明するニューラルネットワークにモデル a) の両耳からの信号を入力することで、二つの入力音の周波数特性から直接音源の方向を求めるモデルもある [8]。

d) については、心理学実験の結果から音像の変移量は ITD、ILD の変化量にほぼ比例することがわかっているため [1]、ITD、ILD の検出までを行うモデルや音源の方向を ITD または ILD の簡単な関数として推定するモデルもある [4]~[6]。しかし、ITD と ILD の両方の特徴を有機的に組み合わせて、これより音源の方向を推定するには、このような簡単なモデルでは困難である。これに対して、近年ニューラルネットワークの応用が盛んに行われている。特に 3 層の階層型ネットワークに誤差逆伝搬法を学習則に用いたモデルは、入力データとそれに対する教師データを繰り返し与えるだけで入出力データ間の写像が形成できる、学習機能と呼ばれる機能があるため、音源定位モデルにおいても ITD や ILD の値を入力データとし、そのとき推定されるべき音源の方向を教師データとして与える方法が提案されている [7], [8]。

このように現在までにもいくつかの音源定位モデルが提案されているが、これらにはまだいくつかの問題点がある。一つには特徴量の検出の主要部分である ITD の検出に相関値計算のように大きな計算コストが必要な演算を用いているため、リアルタイムでの音源定位が困難な点が挙げられる。また、音源方向の推定に階層型ニューラルネットワークを用いる方法は、学習機能によって入出力データ間の高度な写像関係を実現できるため有効であるが、現在用いられている階層型ニューラルネットワークは文字認識のような時間的に不変な情報を扱うことに主眼をおいているためダイナミクスをもたない。このため音のように本質的に時系列情報であるものを扱う場合には入力信号のある時間窓で分割して特徴量を求め、それを順次ネットワークに入力する必要がある、時系列信号の処理に向きである。

これに対して生体の音源定位機構は、たくさんの神経細胞の組合せによって多くの機構を実現しており、各細胞自身は非常に単純な情報処理しか行っていない。MSO で行われている ITD の検出においても、その構成素子はこのような神経細胞である。よって神経細胞に相当するような簡単な演算素子の並列処理によっても ITD の検出は可能であると考えられる。

各細胞間の伝達情報は時系列信号である神経パルスである。このため信号が途切れることなく次々に上位の細胞群に伝えられることで、多くの情報処理がパイプライン処理的に高速に行われている。時系列信号を処理するためのニューラルネットワークを考えた場合、時系列情報であるパルス信号を伝達情報とするニューラルネットワークを用いることは有効である。また、現在の音源定位モデルは、ITD の検出、ILD の検出、そして特徴量からの音源方向の推定を別々の手法を用いて実現しているが、ILD の検出も神経細胞を単位とした機構で実現できたならば、これら三つの部分を同一のモデルで表現可能となる。

そこで我々は、上記のような特徴をもちパルスを伝達情報とする PN モデルを基本単位とし、この PN モデルを組み合わせた細胞群を層状に配置したモデルを提案する。このモデルは PN モデルの配置やパラメータの変更だけで ITD と ILD という異なる特徴量を入力信号から検出でき、かつ相関値計算のような大きな計算コストを必要としない。現在のところ PN モデルにおける結合重みの学習則が確立されていないため、音源方向の推定を行う部分は構築されていないが、こ

の部分についても同じPNモデルを用いて実現が可能であると考えられる。また各PNモデルは独立、非同期に動作が可能であり、加えて常に出力パルスを発生しているため、PNモデルから構成される各細胞層はパイプライン的に並列動作が可能である。

### 3. パルスニューロンモデルを用いた音源定位聴覚系モデル

図1に本モデルのブロック図を示す。蝸牛モデルからMSO, LSOモデルまでの詳細については文献[9]を参照されたい。なお以降の文において、MSO, LSO, ITD-Mapping, ILD-Mappingの各モデルを構成しているPNモデルのことを単にニューロンと呼ぶことにする。

入力信号はまずFIRフィルタを用いた帯域フィルタ群である蝸牛モデルにより周波数別の信号に分離され、有毛細胞モデルにおいて非線形変換が加えられる。非線形変換を受けた各チャンネルの信号は蝸牛神経に相当する連続値-パルス変換が行われ、MSO, LSOの各モデルでITD, ILDが抽出される。図2, 図3にMSO, LSOモデルの模式図を示す。この図においてだ円はそれぞれ一つのニューロンを表しており、各モデルはニューロンを2次元上に配列している。またいずれにおいても、x軸に平行な同じ行のニューロンは同じチャンネルの信号に反応し、手前から奥に向かって中心周波数は高くなっている。

MSOモデルの各行のニューロン群はJeffressの一致

回路[13]を実現しており、y軸に平行な同じ列のニューロンは同じ時間差に反応することになる。MSOモデルの各チャンネルあたりのニューロン数は121となっている。LSOモデルにおいて $L_n$ は左耳からの入力が右耳からの入力に比べてndB大きいときに発火確率が1/2となるニューロンであり、 $R_n$ は右耳からの入力が左耳からの入力に比べてndB大きいときに発火確率が1/2となるニューロンである。このため、同じ列のニューロンは同じ音圧差に反応することになる。LSOモデルの各チャンネルあたりのニューロン数は51となっている。

図4, 図5にMSO, LSOモデルの出力発火頻度の例を示す。これらの図において、x軸y軸はニューロンの2次元配列に対応しており、MSOモデルの出力の方は発火頻度を濃淡画像で、LSOモデルの方は発火頻度をz軸で表している。入力信号としては両耳に同じホワイトノイズを与えておりITD, ILDはない。MSOモデルは入力信号間にITDがあった場合、この出力パターンの形を維持したままパターンが左右にシフトする。LSOモデルは入力信号間にILDがあった場合、音の大きい側のニューロンの発火頻度が全体に増大し、音の小さい側のニューロンの発火頻度が全体に減少する[9]。またMSOモデルの最小検出可能ITDは入力信号のサンプリング周波数 $f_s$ で制限され $2/f_s(\text{sec})$ である[9]。

なお、以前提案したLSOモデルにおいては、center neuron に一番近いニューロンが $L_0, R_0$ であったが、後述のILD-Mappingモデルでエッジ検出処理を行った場合に、ILD=0dBの場合にも良好に検出できるように、 $L_0, R_0$ の更に内側に $L_{-1}, \dots, L_{-5}$ と $R_{-1}, \dots, R_{-5}$ の各ニューロンを用意した。今回新たに構築したITD-Mapping, ILD-Mappingの両モデルについては次章で詳しく述べる。

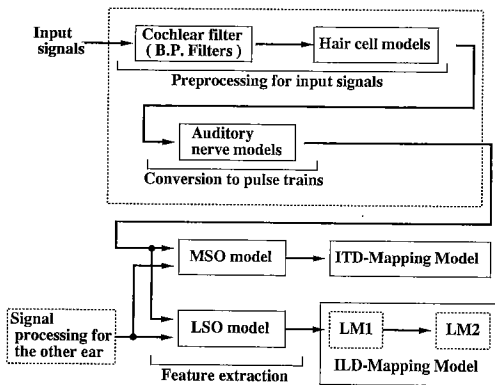


図1 聴覚神経系モデルのブロック図

Fig. 1 Block diagram of the auditory neural network model.

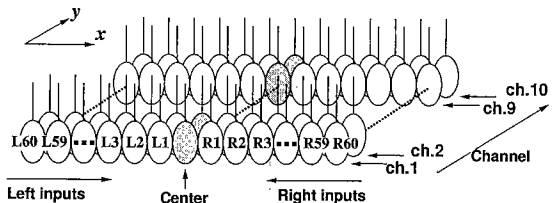


図2 MSOモデル

Fig. 2 The MSO model.

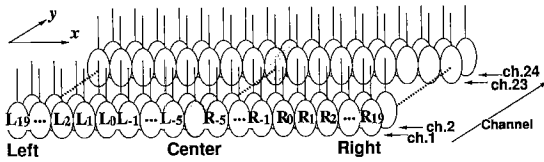


図3 LSOモデル  
Fig.3 The LSO model.

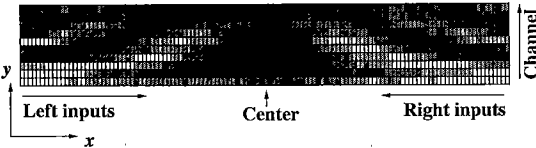


図4 濃淡画像で表したMSOモデルの出力発火頻度の例  
Fig.4 Pulse frequencies of the MSO model as a gray level image.

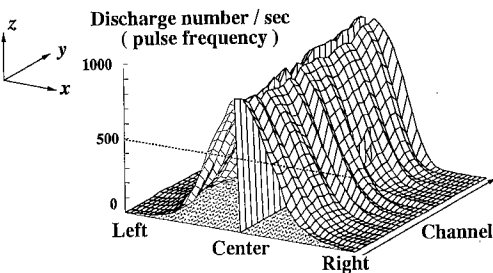


図5 LSOモデルの出力発火頻度の例  
Fig.5 Pulse frequencies of the LSO model.

#### 4. ITD-Mapping, ILD-Mappingモデル

ITD-Mapping, ILD-Mapping の両モデルについて説明する。前述のように今回のモデルは水平面上の単一音源からのITD, ILDの検出を目的としているため、最終的な出力はすべてのチャンネルの信号を統合したITD, ILDそれぞれ1次元上のPNモデルの発火で表すこととする。

図6にITD-Mapping, ILD-Mappingモデルを構成するPNモデルの模式図を示す。PNモデルは前段からのパルス入力*i*によって各局所膜電位*p*が結合重み*w*分上昇し、時定数 $\tau$ で減衰する。そしてこれら膜電位の総和*I*があるしきい値を超えたとき、発火(出力パルスの発生)する。但し、ある発火から1msecの間は細胞の不応期として、膜電位の総和がしきい値を超えた場合にも発火はしない。

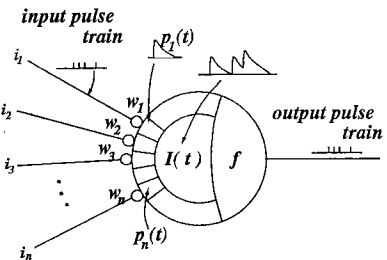


図6 パルスニューロンモデル  
Fig.6 Pulse neuron model.

PNモデルは、その時定数 $\tau$ の設定によって動作が大きく変化する。 $\tau$ を最少パルス間隔に相当する不応期の長さに対して大きくとった場合は、入力パルス列の時間的な加算が可能となり、入力パルス頻度の変化によって*I*が変化する。逆に $\tau$ を不応期の長さに対して小さくとった場合は、入力パルス列の時間的な加算が不可能となり、複数のPNモデルから入る瞬時的なパルスの量で*I*が変化をするようになる。

なおMSOモデルでは、 $\tau$ を0.35msecと不応期に対して小さく設定することで左右の耳からパルスが同時に入ってきたときを検出しており、一方LSOモデルにおいては、 $\tau$ を10msecと不応期に対して大きく設定することで音圧を表す蝸牛神経モデルの出力パルス頻度の差を検出している [9]。

#### 4.1 ITD-Mappingモデル

図7にITD-Mappingモデル(以降、TMモデルと省略)の構成とMSOモデルとの結合の様子を示す。ニューロンの個数はMSOモデルの1チャンネル当りの個数と同じ121である。TMモデルの各ニューロンはMSOモデルの1-10ch(ch:channel)のニューロン、1210個(121×10ch)からそれぞれ出力パルスを受け取る。高い周波数に対応した11ch以降の出力は位相の情報が保存されていないので、時間差の抽出には使えない [9]。本モデルにおいて両耳に与えた信号間のITDは、MSOモデルのいずれのチャンネルでも発火しているニューロンの列の位置として検出できるので、TM Layerの各ニューロンは、MSOモデルの自分と同じ列およびその両側3列のニューロンに対して興奮性の結合を、それ以外のニューロンに対して抑制性の結合を与える。例えばTMモデルのL<sub>3</sub>ニューロンは、MSOモデルの1-10chのL<sub>4</sub> ~ L<sub>6</sub> およびL<sub>2</sub> ~ CenterNeuronと興奮性の結合をもち、そのほかのニューロンと抑制性の結合をもつ。TMモデルのニューロンの膜電位減衰

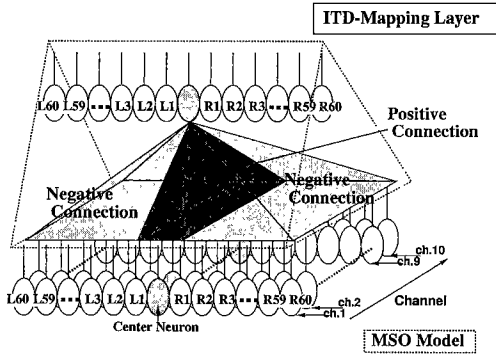


図7 ITD-Mapping モデルの構成およびMSOモデルとの結合の様子  
Fig.7 The ITD-Mapping model.

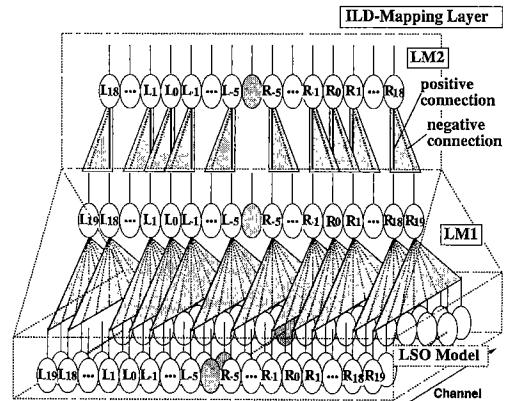


図8 ILD-Mapping モデルの構成およびLSOモデルとの結合の様子  
Fig.8 The ILD-Mapping model.

の時定数  $\tau$  は 10msec と不応期に対して大きく設定し、MSOモデルの各出力パルス頻度の変化を検出するようにした。

#### 4.2 ILD-Mapping モデル

図8に ILD-Mapping モデル (以降、LMモデルと省略) の構造を示す。LMモデルは2層構造をしており、それぞれLM1, LM2とする。LSOモデルは両耳に与えた信号間のILDを発火確率が1/2になるニューロンの列の位置として抽出できる。そこでLMモデルは、LSOモデルにおいて発火確率が1/2となるニューロンの列の位置を検出することを目的としている。なお、ここで発火確率は「1秒間当りの発火回数/1秒間当りの最大発火回数」で表される。1秒間当りの最大発火回数は不応期の存在により1000回であるから、発火確率1/2とは500回/secの発火に相当する。

本モデルを構成するニューロンは入力信号の積和演算としきい値処理を行う。このため、“そのニューロンモデルに対する入力パルス群の発火頻度があるしきい値を越えた場合に発火”や“あるしきい値を超えなかった場合に発火”の処理には適しているが、“ある特定の発火頻度に対してのみ発火”という処理は困難である。そこで、LMモデルをLM1, LM2の2層構造とし、まずLM1において“1/2以上の発火確率をもつニューロン”を検出する。LSOモデルはILDの大きさが大きいほど、各列において center ニューロンから離れたニューロンまで発火することになるので、LM1において発火しているニューロンのうち、center ニューロンから最も離れたニューロンがちょうど1/2の発火確率をもつこととなる。そこでLM2においてこのような動作を行う“発火のエッジの検出”を行うこ

ととした。

##### 4.2.1 LM1

LM1はLSOモデルの1ch当りのニューロン数と同じ41個のニューロンから構成され、それぞれはLSOモデルの対応する列のニューロン24個から入力を受ける。各ニューロンの結合重みはすべて1、発火のしきい値は12としてあるため、LSOモデルの24chのうち12chがほぼ同時に発火したとき、つまり、LSOモデルのその列が1/2以上の発火確率をもったときにLM1のニューロンは発火をする。LM1のニューロンの膜電位減衰の時定数  $\tau$  は不応期に対して小さく設定し、LSOモデルの瞬時的な出力の変化を検出するようにした。ある頻度以上の発火を検出する場合は、 $\tau$  を不応期に対して大きく設定し、LSOモデルの各ニューロンの出力パルス頻度を検出する方法も考えられる。しかし今回のモデルにおいては、単一音源からの唯一のILDを検出することを目的としており、LSOモデルの24chの出力パルスがLM1の各ニューロンに同時に入ってきているので、入力パルスの時間的加算を用いなくても空間的加算のみを用いるだけでLM1は良好な動作をすることが可能である。将来的に周波数帯ごとに異なるILDを検出する場合には、チャンネル数を増やして、隣り合う数チャンネルごとにILDを検出するようにすればよい。

##### 4.2.2 LM2

LM2は、LM1の発火しているニューロンのうち center ニューロンから最も離れたニューロンの位置を検出する。LM2の各ニューロンの結合重みは、対応するLM1のニューロンの出力に対して1、その一つ外側の

ニューロンの出力から  $-1$  とする。例えば LM2 の  $L_0$  ニューロンは LM1 の  $L_0$  ニューロンから  $1$ ,  $L_1$  ニューロンから  $-1$  の結合を受ける。発火のしきい値は  $0$  に設定した。LM1 の  $L_{19}$ ,  $R_{19}$  の二つのニューロンの出力に対しては抑制がかけられないため、LM2 は計 39 個のニューロンから構成される。

LM2 の各ニューロンの膜電位減衰の時定数  $\tau$  は  $10\text{msec}$  とし、LM1 の出力パルスに対して、時間的な加算を行うようにする。

## 5. シミュレーション結果および考察

本モデルの動作を計算機上のシミュレーションで確認した。入力波形として振幅がガウシアン分布に従うホワイトノイズを用い、サンプリング周波数を  $48\text{kHz}$  とする。このため本システムの最小時間分解能は  $\frac{1}{48}\text{msec}$  (約  $20\mu\text{sec}$ ) となる。以下の文では、この  $\frac{1}{48}\text{msec}$  を  $1\text{clock}$  とする。

シミュレーションに用いた ITD, ILD の組合せは

- (1) [same]
- (2) [ITD:+12], [ITD:+6], [ITD:-6], [ITD:-12]
- (3) [ILD:+5.0], [ILD:+2.5], [ILD:-2.5], [ILD:-5.0]
- (4) [L2], [L1], [R1], [R2]

である。上記の組合せにおいて、[same] は左右同じデータを与えた場合を意味し、[ITD: $n$ ] はオリジナルの信号に対して  $n$  clock 早い信号を、[ILD: $n$ ] はオリジナルの信号に対して  $m$  dB 大きい音を右耳から与え、左耳からはオリジナルの信号を与えることを意味する。

また、(4) は音源が正面から左右にずれた位置にある場合に対応したデータである。例えば音源が人間の左側にある場合、左耳への信号は大きくなり、かつ早く耳に到達するようになるが、右耳への信号は小さくなり、かつ遅く耳に到達するようになる。そこで、[L2] は左から  $-6\text{clock}$  早く、 $2.5\text{dB}$  大きい音を、右から  $6\text{clock}$  遅く、 $2.5\text{dB}$  小さい音を入力することで、相対的に左側が  $12\text{clock}$  早く、 $5.0\text{dB}$  大きい音を入力することとした。[L1] は相対的に左側が  $6\text{clock}$  早く、 $2.5\text{dB}$  大きい音を入力する。[R2][R1] はこの逆である。ここで  $6\text{clock}$  は  $125\mu\text{sec}$  に相当し、これは人間において正面から約  $10$  度程度左右にずれた位置にある音源からの音によって生じる時間差に相当する。同様に  $12\text{clock}$  は正面から約  $20$  度のずれに相当する [14]。ILD の値については頭部や耳介の影響のために、実際には入力信号の周波数によって音源のずれる度合は変化するが、

$3\text{kHz}$  の純音に対して  $2.5\text{dB}$  が約  $10$  度、 $5.0\text{dB}$  が約  $20$  度のずれに相当し、周波数が更に高くなった場合同じ ILD に対する音源のずれる角度は更に大きくなる [14]。

なお、信号をパルス列に変換するとき用いる乱数列は、左右それぞれに別の系列を用いているため、発生した二つのパルス列はそれぞれ独立である。また (4) の各組合せについても、例えば [L2] の左からの入力と [R2] の右からの入力は与える時間差、音圧差としては同じものであるが、パルス列変換時に用いる乱数列が異なるため、パルス列としては異なるものとなっている。

### 5.1 ITD-Mapping モデル

図 9 に ITD-Mapping(TM) モデルの出力を各ニューロンの発火頻度で示す。横軸はニューロンの位置、縦軸は各ニューロンの発火頻度である。図 9(a) は ITD の変化に対する TM モデルの出力の変化、図 9(b) は ILD の変化に対する TM モデルの出力の変化、図 9(c) は [L2], [L1], [same], [R1], [R2] に対する TM モデルの出力の変化を示す。なお比較のためいずれの図も [same] の出力を示してある。

図 9(a) より、TM モデルは ITD の大きさによって発火するニューロンの位置がシフトしており、また最大の発火頻度をもつニューロンは入力信号間の ITD に対応していることがわかる。MSO モデル、TM モデルは“右の方が  $n$  clock 遅い”や“左の方が  $n$  clock 遅い”を検出するため、音源の変位とニューロンの発火位置の変位は逆になる。それぞれの入力信号に対する発火頻度を見てみると、その大きさに多少の変動があるが、これはパルス列を生成するとき用いた乱数列の差異や、入力信号間の ILD の違いによるもの(図 9(b))である。

図 9(b) の場合は左耳からの入力を [same] に固定し、右耳への入力の音圧を変えているため、右耳への入力音圧が高くなるほど音圧の総量が大きくなり、MSO モデルの出力は全体的に発火頻度が高くなる [9]。このため右耳への音圧が大きくなるほど MSO モデルの出力パターンの S/N 比が低下し、TM モデルでの発火頻度の低下につながっている。しかし、本モデルは最大発火頻度を持つニューロンの“位置”として ITD を検出するものなので、その絶対量の変化はあまり問題とはならない。

図 9(c) は、実際に音源の位置が左右にずれているときのように左右の耳の ITD と ILD の両方を加えた場合である。この場合も図 9(a) 同様に ITD の大きさに

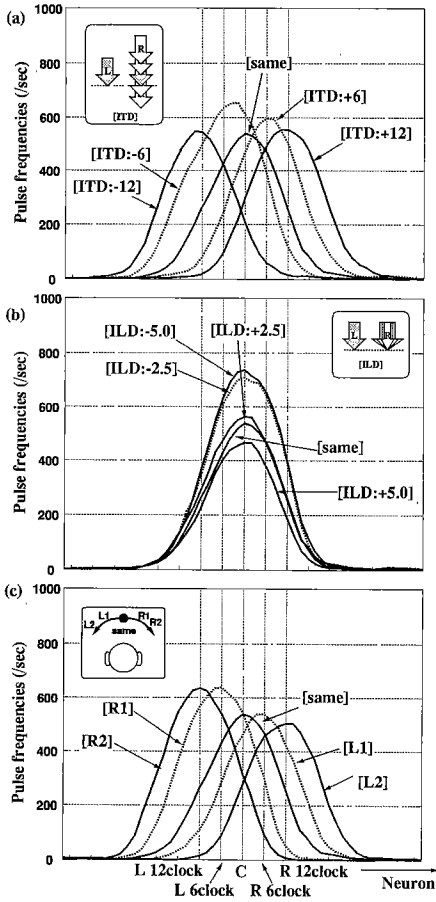


図9 ITD-Mapping モデルの出力発火頻度  
Fig. 9 Pulse frequencies of the ITD-Mapping model.

応じて、発火するニューロンの位置がシフトしている。また、この場合は図9(b)とは異なり、左右の耳への入力信号の音圧変化分を等しくしているため、全体としての音圧の総量には変化がない。このため TM モデルの出力発火頻度が図9(b)のようには大幅に変化しないことがわかる。

このように TM モデルは最大発火頻度をもつニューロンの位置として ITD を検出が可能になったが、各 ITD に対する出力には重なりが大きく、このままでは TM モデルのある発火頻度から正確な ITD の値を知ることは難しい。これは現在のニューロンに結合重みの学習機能が採り入れられていないため、結合重みをモデル化する際に人間が与えていることが原因の一つであろう。今後 TM モデルにも学習機能を付加し、TM モデルの各ニューロンの結合重みを MSO モ

デルの出力とそのとき検出されるべき ITD の組合せから学習によって獲得するようにすることで、TM モデルの出力値の精度を上げることができると考えられる。また、TM モデルの上層に側抑制をかけるためのニューロン層を設けた場合、この側抑制層の発火頻度は左右への拡がりが増えらるため、これによっても ITD 検出の精度を向上できると考えられる。

今回のモデルではサンプリング周波数が 48kHz なので、入力信号における最小時間差は約 20 $\mu$ sec であり、検出可能な最小の ITD は約 40 $\mu$ sec となる。これに対して人間やフクロウの ITD の弁別しきい値は約 10 $\mu$ sec 程度である [15], [12]。本モデルでこの 10 $\mu$ sec の ITD を検出しようとした場合、サンプリング周波数および各チャンネルあたりのニューロンの個数を 4 倍にしてやれば理論的には可能である。しかし、単純にこれらの値を 4 倍にただけでは出力パターンが x 軸方向に拡大されるだけであるので、上述のような機構を用いて発火パターンの左右への拡がりを抑え、ITD 検出の精度を向上させることが必要となる。

### 5.2 ILD-Mapping モデル

図 10 に ILD-Mapping (LM) モデルの出力を各ニューロンの発火頻度で示す。LM モデルは 2 層構造をしているので、LM モデルの出力とは LM2 の出力である。図 10(a), (b), (c) の入力データはそれぞれ TM モデルの出力 (図 9) と同じである。横軸はニューロンの位置、縦軸は各ニューロンの発火頻度を示す。

図 10(a), (b) より LM モデルは入力データ間の ITD の変化に対しては反応をせず、ILD の変化のみによって発火するニューロンの位置がシフトしており、最大の発火頻度をもつニューロンの位置として入力信号間の ILD が表現できていることがわかる。また、図 10(c) が図 10(b) と同じ出力パターンを示していることから、LM モデルは ITD, ILD をもったデータに対しても、その中の ILD 成分のみに反応していることがわかる。

人間の ILD に関する弁別しきい値は約 1dB である [15]。本モデルの ILD 検出の精度に関しては、現在のモデルにおいても隣り合うニューロン同士の検出 ILD の差は 1dB であるからほぼ十分な検出精度をもっていると言える。また、ニューロンの数を倍にして対応する ILD を 0.5dB 刻みで配置するなどの改善を加えることで更に検出精度を上げることも可能である。

今回のシミュレーションでは LSO モデルにおいて 500/sec の発火頻度をもつニューロンの位置が 2 箇所存在するため、ILD-Mapping モデルの出力に関して



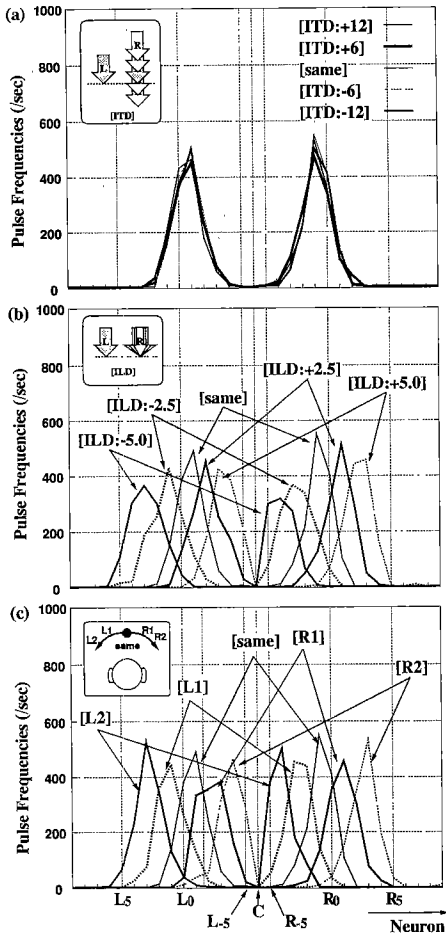


図 10 ILD-Mapping モデルの出力発火頻度  
Fig. 10 Pulse frequencies of the ILD-Mapping model.

もピークが 2 箇所存在している。これについては更に一つ上位のニューロン層を設け、この層のニューロンがこれら 2 箇所のピークの位置から信号を受けるようにすることでピークの 1 本化を図ることは可能である。しかし、今後音源の方向を推定するネットワークを構築する際に、これら二つのピークの位置からの信号を一つにまとめて送れば実用上問題は無い。そこで今回は ILD-Mapping モデルの出力としてピークが二つある状態をそのまま用いることとした。

## 6. むすび

聴覚系の水平面内の音源方向の定位機構を実現するためのパルスニューラルネットワークモデルを構築した。本モデルは大きく分けて音信号の神経パルス列への

変換部と PN モデルから構成される ITD, ILD のマッピング部からなる。シミュレーションによってこのモデルの動作を確認したところ、左右の耳に与えた信号間の ITD, ILD を今回新たに構築した ITD-Mapping, ILD-Mapping モデルの発火する PN モデルの位置としてそれぞれ検出された。

従来の音源定位機構のモデルとは異なり、本モデルは ITD, ILD の検出を同じ PN モデルを用いて行っており、特徴抽出およびマッピングの各部位をすべて PN モデルから構成される細胞層として実現している。これら細胞層の動作の違いは用いられている PN モデルのパラメータおよび PN モデル間の結合の違いだけである。

また、PN モデルの入出力データはともにパルス列であり、入力パルス列を受け付けると同時に常に出力パルス列を発生しているため、これらの細胞層はパイプライン処理的に並列動作が可能であり、今後更に多くの細胞層を設けた場合にも非常に高速の情報処理が可能である。また同層内の各 PN モデルも互いに独立、非同期に動作するので同層内の PN モデルの並列動作も可能である。既存のニューラルネットワークをハードウェア上で実現する場合、その計算コストが問題となっているが、PN モデルは簡単なアナログ回路で構築することが可能である。このためパルスニューラルネットワークモデルは音源定位システムをハードウェア上で構築する場合に有利である。

ニューラルネットワークを用いる最大の利点は、学習による画像の獲得機能にある。現段階では PN モデルのための学習則はまだないため、各結合重みは定性的にあらかじめ与えておく必要がある。今回のモデルでは ITD-Mapping モデルの出力の精度が問題となっていたが、この点についても PN モデルの学習則が確立できれば学習によるマッピングの形成が可能となり、検出精度の向上が期待される。今後は PN モデルの学習則を確立することで学習による各結合重みの変更を可能とし、マッピングモデルにおける検出精度を向上させると共に、現在のモデルの上層に位置する、検出した ITD, ILD の値から音源の方向を推定する部分についても PN モデルを用いて実現していきたい。

## 文 献

- [1] 勝木保次, 村田計一, 吉田登美男, 亀田和男, “新版 聴覚と音声,” pp 1-240, 電子情報通信学会, 1980.
- [2] J.O. Pickles, “An Introduction to the Physiology of Hearing,” Academic Press, 1988.
- [3] G.M. シェパード著, 山元 大輔訳, “ニューロバイオロジー

(第2版),” 学会出版センター, 1990

- [4] J. Blauert, and W. Cobben, “Some consideration of binaural crosscorrelation analysis,” *Acustica*, vol.39, pp.96-104, 1978.
- [5] R.M. Stern, Jr. and H.S. Colburn, “Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position,” *J.Acoust.Soc.Am.*, vol.64, no.1, pp.127-140, 1978.
- [6] S.A. Shamma, “Stereoausis: Binaural processing without neural delays,” *J.Acoust.Soc.Am.*, vol.86, no.3, pp.989-1006, 1989.
- [7] 勝 拓二, 杉山 精, 宮田裕之, 水島昌英, “ニューラルネットワークを用いた音源位置推定,” 日本音響学会講演論文集, pp.563-564, Nov. 1994.
- [8] C. Neti and E.D. Young, “Neural network models of sound localization based on directional filtering by the pinna,” *J.Acoust.Soc.Am.*, vol.92, no.6, pp.3140-3156, 1992.
- [9] S. Kuroyanagi and A. Iwata, “Auditory pulse neural network model to extract the inter-aural time and level difference for sound localization,” *IEICE Trans.*, vol.E77-D, no.4, pp.466-474, April 1994.
- [10] W.A. Yost “Lateral position of sinusoids presented with interaural intensive and temporal differences,” *J.Acoust.Soc.Am.*, vol.70, no.2, pp.397-409, 1981.
- [11] D.M. Leakey, B.M. Sayers, and E.C. Cherry, “Binaural fusion of low- and high- frequency sounds,” *J.Acoust.Soc.Am.*, vol.30, pp.322, 1958.
- [12] 小西正一, “メンフロウの両耳による聴覚情報処理,” *日経サイエンス*, vol.23, no.6, pp.90-99, 1993.
- [13] L.A. Jeffress, “A place theory of sound localization,” *J.Comp.Physiol.Psychol.*, vol.41, pp.35-39, 1948.
- [14] W.E. Fedderson, T.T. Sandel, D.C. Teas, and L.A. Jeffress, “Localization of high-frequency tones,” *J.Acoust.Soc.Am.*, vol.29, pp.988-991, 1957.
- [15] イェンス ブラウエルト, 森本政之, 後藤敏幸, “空間音響,” 鹿島出版会, 1986.

(平成7年3月23日受付, 8月21日再受付)



岩田 彰 (正員)

昭48名大・工・電気卒。昭50同大大学院修士課程了。同年名工大・情報・助手。昭57年4月より昭58年10月まで、ドイツ連邦共和国ゲーセン大学医学部医用情報研究所客員研究員。昭59名工大・情報・助教授。平5名工大・電気情報・教授、現在に至る。生体情報処理, 医用画像処理, ニューラルネットワーク, 遠隔医療情報システムに関する研究に従事。工博, 日本ME学会, 情報処理学会, IEEE各会員



黒柳 奨 (学生員)

平3名工大・電気情報卒。平5同大大学院博士前期課程了。現在, 同大大学院博士後期課程(電気情報工学専攻)在学中。ニューラルネットワーク, 聴覚情報処理に関する研究に従事。日本音響学会会員。