

FPGA を用いたパルスニューロンモデルによる音源定位及び音源種類の識別システムの実装

岩佐 要^{†a)} 黒柳 奨[†] 岩田 彰[†]

Implementation of Sound Localization and Sound Recognition System by Pulsed Neuron Model Using FPGA

Kaname IWASA^{†a)}, Susumu KUROYANAGI[†], and Akira IWATA[†]

あらまし パルスニューロンモデルは音信号のような時系列情報を扱うのに適しており、かつ演算が単純であるためハードウェア化が容易であるという利点をもつ。本論文では、その利点をいかし実際に FPGA を用いてハードウェア上にパルスニューロンモデルを実装した。更に実装されたパルスニューロンモデルを組み合わせることで時間差検出ネットワーク及び競合学習ネットワークを電子回路として構成し、入力音から音源定位と音源種類の識別を同時に行うシステムを構築した。実験の結果、ハードウェア実装をした本システムが音源定位と音源種類の識別を計算機シミュレーションと比較して高速に行うことができ、かつ従来手法と比較して少ない回路数で実装できることが確認できた。

キーワード パルスニューロンモデル、ハードウェア実装、音源定位、音源種類の識別

1. ま え が き

人間は聴覚から得られた音に関する情報から、どのような音が（音源種類の識別）どこから（音源定位）聞こえてくるのかという周囲の状況を把握することができる [1]。この能力を工学的に応用することができるれば、聴覚障害者への支援や危険を通知する装置など様々な利用が期待できる。

このような障害者支援や危険通知のための装置を実用化する上で留意すべき点としては、装置の規模や入力信号に対する応答速度などが挙げられる。音源定位については精度良く定位を行う方法としてマイクロホンアレーを利用した研究が行われている [2] ~ [4]。しかしマイクロホンアレーを用いた場合集音器の数が多く大きな装置になるため、日常生活で使用するには不便である。また音源種類の識別に関しては、従来より数多くの研究がなされ集音器一つと計算機を利用した方法が一般的に用いられている [5], [6]。しかし音源定

位や画像処理などの他の処理を同時に行う場合には、高性能な計算機が必要となる。

一方、音源の方向や種類を“識別する”という点については、パターン認識の分野では人間の脳細胞をモデル化したニューロンモデルが広く用いられている。しかし音のような時系列情報を扱うにはパターンの時間的な変化や順序を許容することが必要であるため、各時間ごとの入力データに対してのみ考慮する従来のニューロンモデルは不向きである。時間的な変化に対応するために時間情報を空間的に展開して扱う TDNN [7] などの様々な形のニューラルネットワークが使われているが、構成の複雑さに伴って使用する素子も多くなるため計算量も膨大になり、入力信号に対する十分な応答速度が得られない可能性がある。またニューロンモデルのほかに音源定位では CSP 法 (Crosspower Spectrum Phase analysis) [4], [8]、音源識別では GMM (Gaussian Mixture Model) を用いた方法 [9], [10] など認識精度のよい手法が個別に提案されている。しかしこれらは積算を用いた計算が多く、ハードウェア化する場合には DSP などのコストの高い回路が必要となる。

これらに対して、入出力信号をパルス列に限定した

[†] 名古屋工業大学, 名古屋市

Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya-shi, 466-8555 Japan

a) E-mail: kaname@mars.elcom.nitech.ac.jp

パルスニューロン (Pulsed Neuron 以下 PN) モデルが提案されている [11]. PN モデルでは時間的に減衰する膜電位で情報を表現するため時系列情報を扱うのに適しており, 演算が単純なため簡単な電子回路によって実装が可能であるという利点をもつ [12]. また, 基本素子である PN モデルのハードウェア化が容易であれば, システム全体を一つのハードウェアに搭載することが可能となり従来手法より小規模なシステムを構築することができる. 更に各 PN モデルはハードウェア化することで並列に計算することができるため, 高性能な計算機などを用いなくても高速な処理能力が期待できる.

以上のことから, 本研究ではハードウェア上に多数の PN モデルを実装し, これらを組み合わせることで音源定位と音源種類の識別を同時に行うことのできる聴覚情報処理システムを提案する. なおハードウェアへの実装には Field Programmable Gate Array (FPGA) を用いる. 音源定位には左右の二つの耳に入力される音の時間差を用い, PN モデルで構成された時間差検出ネットワーク [13] により時間差情報を抽出し, 同じく PN モデルで構成された自己組織化マップ [14] の出力により音源の位置を判断する. 音源種類の識別には入力音の周波数情報を用い, 自己組織化マップによりベクトル量子化を行い次段の自己組織化マップの出力にて識別を行う. このように音源定位と音源種類の識別を行う回路を同じ PN モデルで実現することにより, 特殊な機構を個別に用意することなく本システムを小規模なものにすることができている. 次章より, PN モデルと提案する聴覚情報処理システム及びその FPGA への実装方法について説明した後, 音信号を入力することで本システムが音源定位及び音源種類の識別を行うことができることを示す.

2. パルスニューロンモデル

音信号のような時系列データに関して情報処理を行う場合には, 時間的な関係を扱うことができるモデルであることと, リアルタイム処理のために演算が簡単であることが望ましい. 一般的に用いられている従来のニューロンモデルでは各単位時間における静的な入力ベクトルのみを扱うため時系列情報のような動的な入力を扱うにはフィードバックを用いるなどのネットワーク構成を変更しなければならず, 各素子における演算コストも高くなる. これに対して, パルスニューロン (Pulsed Neuron 以下 PN) モデルが提案されて

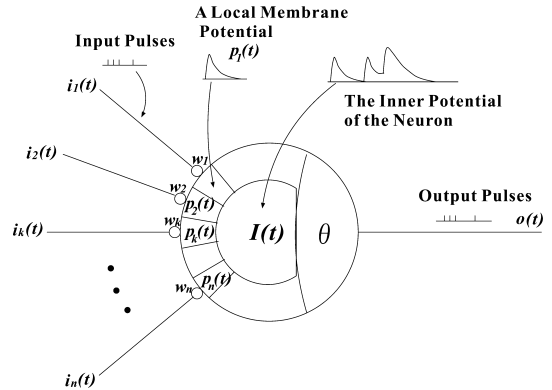


図1 パルスニューロンモデル
Fig. 1 A pulsed neuron model.

いる.

PN モデルの構造を図 1 に示す. PN モデルは他のニューロンと結合する n 本のシナプス, 各シナプスにおける局所膜電位 $p_k(t)$, 各局所膜電位の総和である内部電位 $I(t)$, 出力を他のニューロンに伝達する軸索から構成される. 時刻 t の値は連続値として扱うことも可能であるが, 本研究では計算機上及び FPGA 上で PN モデルを用いるために離散値とし, 最小時間単位を dt [sec] とする. サンプリング周波数を f_s とすると, $dt = 1/f_s$ である.

PN モデルにおいては, まず各シナプス k において入力パルス $i_k(t)$ が入力されると, それぞれのシナプスに対応した結合重み w_k の値だけ局所膜電位 $p_k(t)$ が増加し, その後時間の経過とともに時定数 τ_k で指数関数的に減衰する (式 (2)). 次に PN モデルは単位時間ごとに局所膜電位 $p_k(t)$ の総和である内部電位 $I(t)$ を計算する (式 (3)). この内部電位 $I(t)$ がしきい値 θ を超えると発火, 時刻 t において出力 $o(t)$ として 1 を出力する (式 (3)). またある PN モデルが一度発火した後, 一定時間 RP の間は内部電位がしきい値を超えた場合はこの PN モデルは発火できない. この時間を不応期と呼ぶ.

$$p_k(t) = w_k i_k(t) + p(t-1) \exp\left(-\frac{dt}{\tau}\right) \quad (1)$$

$$I(t) = \sum_{k=1}^n p_k(t) \quad (2)$$

$$o(t) = f_{usf}\left(I(t) - \theta\right) \quad (3)$$

PN モデルと通常のニューロンモデルとの違いは,

PN モデルが入出力に神経パルスを想定し 0 か 1 の値をとるパルス信号を用いている点である．このため各素子における演算が単純であり，大規模なネットワークを少ないハードウェア資源で構成することができる．また入力的情報を時定数で減衰する膜電位として保持しており，音信号のような時系列信号の扱いに適している．

3. 聴覚情報処理システム

本論文で提案する聴覚情報処理システムについて概略図を図 2 に示す．本システムは大きく分けて周波数-パルス変換部 (図 2 中 A)，時間差検出部 (B-1)，音源位置識別部 (B-2)，周波数パターン量子化部 (C-1)，音源種類識別部 (C-2) の五つからなる．時間差検出部，音源位置識別部，周波数パターン量子化部，音源種類識別部については PN モデルで構成され FPGA 上に実装されている．入力された音信号は計算機上の周波数-パルス変換部 (A) で複数の周波数成分に分割された後パルス情報へと変換される．変換されたパルス情報は FPGA へ入力され音源定位及び音源種類の識別が並列に行われる．得られた音源定位結果及び音源種類の識別結果は再び計算機へ返され，計算機上で出力結果を表示する．

音源定位手法としては，黒柳らの提案した時間差検出手法 [13] を用い，時間差検出部 (B-1) にて入力される音信号に対する左右の信号の時間差情報を検出する．次に，検出された時間差情報を音源位置識別部 (B-2) である競合学習ネットワークに入力することで音源位置の識別を行う．出力部として競合学習ネットワークを用いた理由として，雑音などの誤った情報に

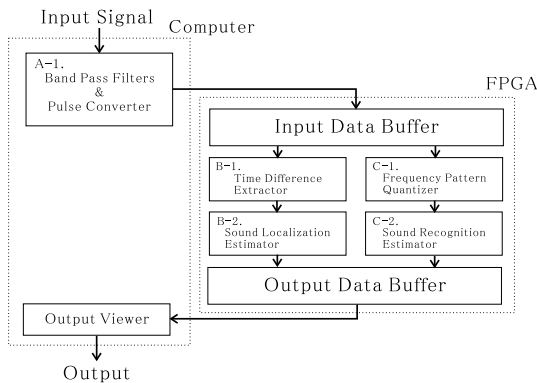


図 2 聴覚情報処理システムの概略
Fig. 2 The abstract of aural information system.

対して頑健にする他に，適用する環境で学習させることによって教師となる信号を与えることなくシステムが自動的に識別するためのパラメータである各ニューロンの重みの状態を環境に合わせ利用者が手動で設定するという不便さをなくすという目的がある．

音源種類の識別手法としては，周波数-パルス変換部で得られた周波数パターンを用いる．しかしこの周波数パターンを用いて直接教師あり学習を行うと同じ音の周波数パターンでも時間的に周波数成分が変化する場合に学習が難しくなるという問題がある．そこで本システムでは周波数パターンを一度周波数パターン量子化部 (C-1) である競合学習ネットワークによってベクトル量子化し，ベクトル量子化されたパターンを後段の音源種類識別部 (C-2) である競合学習ネットワークに入力することで入力音源の種類を識別する．

3.1 周波数-パルス変換部

まず音信号は周波数-パルス変換部に入力される．周波数-パルス変換部では，左右の信号ごとにまず周波数情報に変換された後，Band Pass Filter 群によって各中心周波数の特性をもった周波数帯域に分割される．入力信号のサンプリング周波数は 16 kHz としているため各フィルタの周波数特性は中心周波数が対数尺度において 100 Hz ~ 8 kHz までの間で 43 分割し，それぞれが等間隔になるように決定してある．また図 3 に示すように音源種類の識別には 43 すべての周波数帯の情報を使用し，音源定位には 43 のうち三つ間隔で間を抜き，合計 15 の周波数帯の情報を使用する．次に各周波数帯の情報ごとにパルス列に変換する．パルス列への変換は，周波数変換後の値が大きいほど確率的に多くのパルス列が含まれるようにする．

3.2 時間差検出部

時間差検出部ではパルス変換部で変換されたパルス列を図 4 に示す時間差検出ネットワークに入力する．本システムで用いる時間差検出ネットワークは，Jeffress のモデル [15] をもとにしており，左信号を左か

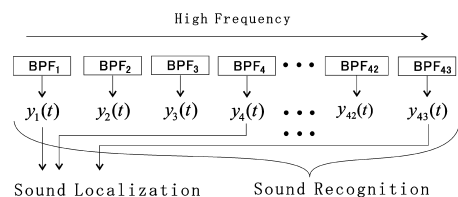


図 3 定位と識別に用いる周波数帯
Fig. 3 The channel number used each processing.

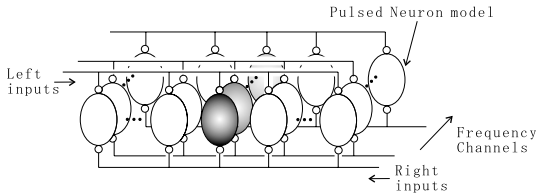


図 4 時間差検出ネットワーク

Fig. 4 The network of extracting time difference.

ら、右信号を右から入力する。入力されたパルス信号は単位時間ごとに矢印の方向に移動し、対応した PN モデルに入力される。各 PN モデルは二つの入力と同時に入力されたときに発火する。例えば音源が中央にある場合には、左右の耳に同時に入力されるためパルス情報は同じ回数だけシフトし、中央の PN モデルにおいて同時に入力され発火することになる。音源が中央より右にある場合には、音信号は右の方が左より早く入力されるため、シフトする回数が多くなり中央より左の PN モデルにおいて同時に入力され発火する。このように音源の位置によって発火する PN モデルの場所が変わることになる。

Jeffress のモデルでは入力信号を単純にシフトしていくだけであるため、入力波形中に含まれる別の周期のパルス信号同士が同時に各 PN モデル入力されることで発火し、本来求めるべき真の時間差とは異なった偽の時間差情報が発生することになる。偽の時間差は入力信号の周期に関係するため、複数の周波数成分を比較することで真の時間差を判断することができる。しかし入力信号に多くの周波数成分が含まれてない場合には比較することが困難になるので真の時間差を判断することは難しくなる。この問題に対処するため、提案システムでは PN モデルが発火した場合、発火に関与した左右のパルス信号を以降は不必要であると判断し削除するという機構を加えた [16]。これにより偽の時間差情報の影響を小さくし真の時間差情報を明確にすることができる。

3.3 音源位置識別部

音源位置識別部には PN モデルによって構成された競合学習ネットワーク [14] として自己組織化マップ (Self Organizing Map : SOM) を用いている。SOM は、任意の次元をもつ入力信号を表現空間の次元数の削減を行いながら位相的な関係を保存することが可能であるアルゴリズムである [17]。このため、SOM を用いることで入力信号の実際の空間における近さと

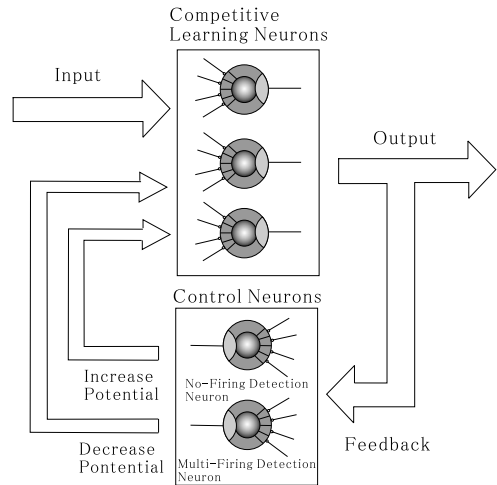


図 5 パルスニューロンモデルのための競合学習ネットワーク

Fig. 5 Competitive learning network for pulsed neuron model.

SOM の素子の近さを学習によって自動的に対応づけることができる。SOM では教師なし学習を行うため厳密な角度の校正は行われませんが、入力信号が左右均等に用意されていれば各素子が角度ごとに均等に対応づけられるため問題はない。以下本論文では、この PN モデルによって構成された競合学習ネットワークを CONP (COmpetitive learning Network using Pulsed neurons) と記す。図 5 に CONP を示す。

SOM の学習において位相的な関係を保存するには、入力に対して最も類似する重みをもつニューロン (勝者ニューロン) を決定し、勝者ニューロンのみが発火するようにしなければならない。しかし出力値が実数値ではなく 0 or 1 のパルス信号であり各素子が独立非同期に動作している PN モデルでは勝者を決定することが困難である。そこで、図 5 のように外部の素子 (制御ニューロン) が出力の状態を判断し各競合学習ニューロン (COmpetitive learning Neuron 以下 CO ニューロン) のもつ内部電位を一律に調整し、CO ニューロンが一つだけ発火するような状態にするという制御を行っている [14]。このような内部電位の制御によって CO ニューロンが一つだけ発火し、入力信号に対して発火した頻度が多くなった CO ニューロンの番号を音源の相対的な位置として決定する。なお制御ニューロンとして、CO ニューロンが一つも発火していないことを検出する無発火検出 (No Firing Detection) ニューロンと、CO ニューロンが複数個発

火していることを検出する複数発火検出 (Multi Firing Detection) ニューロンを用意する。

CONP においては、各 CO ニューロンのしきい値は外部素子を用いて増減されるため、指定値以下の入力に対して発火しないようにするという従来のニューロンモデルにあるしきい値制御はできない。このため各 CO ニューロンには学習により変化する重みのほかに、常に 1.0 とする重みとそれに関する電位をもたせる。重みが固定値であるためこれによって計算される電位は入力パルスの頻度を示すことになる。そしてしきい値 θ の他に制御しきい値 θ_{gate} を設定することで、入力パルス頻度が制御しきい値以下であれば発火しないという制御を行う。また入力パルス頻度が制御しきい値以下であれば制御ニューロンからのしきい値の調整は行わない。各 CO ニューロンにおいては内部電位から入力パルス頻度を一定の割合で引くことで、内部電位の急激な変化を抑え、CO ニューロンの内部電位の調整における追従性を向上させている。

CONP における内部電位の調整を考慮すると各 CO ニューロンにおける出力の式 (3) は、式 (4) のようになる。ここで $p_{nfd}(t)$ は無発火検出ニューロンに対する膜電位、 $p_{mfd}(t)$ は複数発火検出ニューロンに対する膜電位、 $p_{in}(t)$ は入力パルス頻度、 β は入力パルス頻度に対する割合である ($0 \leq \beta \leq 1$)。

$$o(t) = f_{usf} \left(\sum_{k=1}^n p_k(t) - \theta + p_{nfd}(t) - p_{mfd}(t) - \beta \cdot p_{in}(t) \right) \quad (4)$$

3.4 周波数パターン量子化部

周波数パターン量子化部では周波数-パルス変換によって得られたパルス列に対してベクトル量子化を行う。構成に関しては 3.3 で説明した PN モデルによる競合学習ネットワークと全く同じである。音源種類識別用の入力データとして、両側のデータを加算して使用した場合には左右データ間の時間差が認識結果に悪影響を及ぼす。このため音源が存在する側の入力データに逐次切り換えて使用するのが最適な手法であるが、予備実験から左側の入力データのみでも十分な識別結果が得られることを確認したため、本実験では簡略的に左側の入力データのみ用いている。

3.5 音源種類識別部

音源種類識別部では周波数パターン量子化部でベクトル量子化されたパターンが入力され、入力パターン

がどの音であるかを識別する。音源種類識別部は音源位置識別部や周波数パターン量子化部と同じ CONP で構成されており、必ずどれかのニューロンが発火するように制御することができ、入力音の大きさに合わせてしきい値などを設定する必要がない。逆に、必ずいずれかのニューロンが出力結果を返すことになってしまうため、後述の実験においては学習した音以外の音が入力された場合のために“その他”と識別するための素子を用意した。

学習に関しては、識別するべき音を明示的に示す必要があるため黒柳らの提案した教師あり学習法 [18] に基づき計算機上で各 PN モデルの重みを更新する。具体的には、まず学習のために出力に関する膜電位 $p_{out}(t)$ と教師パルスに関する膜電位 $p_{sup}(t)$ を用意しておく。教師パルスが与えられ、かつ $p_{sup}(t) > p_{out}(t)$ の場合に重みを増加させ、パルスを出力し、かつ $p_{out}(t) > p_{sup}(t)$ の場合重みを減少させる。以上の教師あり学習を行った後に、各 CO ニューロンごとの重みベクトルの大きさが 1 になるよう重みを正規化し、認識時には他の部位における CONP と同様に勝者ニューロンのみが発火するように制御を行う。

4. FPGA への実装

4.1 パルスニューロンモデルの実装

FPGA 上で電子回路として実現した PN モデルのブロック図を図 6 に示す [12]。従来のニューロンモデルにおいては入力信号がアナログ値であるため、“入力 × 重み” という計算を行うために積算器を必要とする。しかし PN モデルでは入力が 0 か 1 のパルス信号であるため、“入力 × 重み” の計算が積算ではなく、入力信号が入力されたときに重みの値だけ出力する AND 回路で実現することができる。一般に AND 回路は積算器に比べて回路規模が小さく、更に各回路は入力の本数だけ必要となるため PN モデルは従来のニューロンモデルに比べて非常に小さい回路規模で実装することができる。AND 回路より出力されたそれぞれの値は加算器を用いて加算され、比較器にてしきい値と比較して出力が決定される。

時定数による減衰は、ビットシフトと補数演算で実現しておりもとの内部電位の何分の一かを減算するという演算を行う。本来減衰は各膜電位ごとに用意するものであるが、本実験では各膜電位の時定数を同じ値としたため膜電位の総和である内部電位に対し減衰をもたせることで等価な回路とした。図 7 のように減

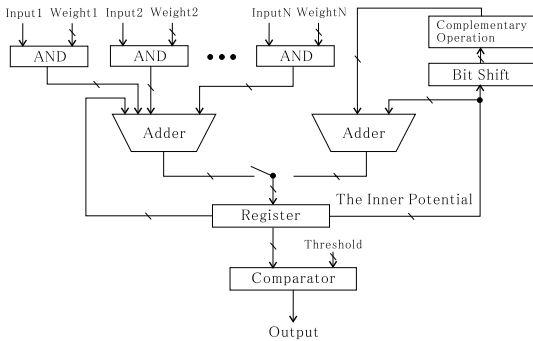


図 6 FPGA におけるパルスニューロンモデル
Fig. 6 A pulsed neuron model on FPGA.

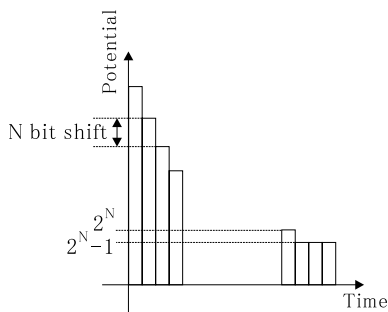


図 7 FPGA における電位の減衰の表現
Fig. 7 The expression of decaying inner potential in FPGA.

衰において N ビットシフトを行う場合、電位の値は $2^N - 1$ より小さくならず、0 にはならない。このため、各重みやしきい値を設定する際に電位の基準値を $2^N - 1$ に設定し、基準値に対して重みやしきい値が何倍であるかを算出してから与えるという方法を用いた。

なお今回の実装にあたっては、PN モデルのもつ学習機構は FPGA 上には実装されていない。このため、計算機上であらかじめ学習して生成された重みを固定小数点表現に変換して回路に与えることで代用し、認識動作のみを行うこととしている。

4.2 時間差検出ネットワークの実装

時間差検出ネットワークを実装するには、図 8 に示すようにパルス情報を移動させるためのシフト演算器と PN モデルを横一列に配置する。動作については 3.2 で説明したとおりである。また改良点である PN モデルが発火した場合にパルスを削除するという機構については、各 PN モデルの出力を対応したシフト演算器へと送り、PN モデルの出力があった場合にシフト演算器の入力をクリアするという方法を用いている。

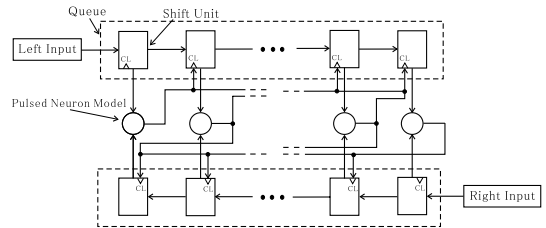


図 8 FPGA における時間差検出ネットワーク
Fig. 8 The network of extracting time difference on FPGA.

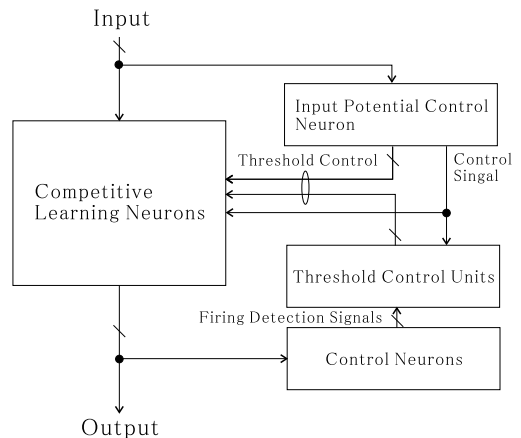


図 9 FPGA における競合学習ネットワーク
Fig. 9 The competitive learning network on FPGA.

4.3 競合学習ネットワークの実装

PN モデルによる競合学習ネットワークは図 9 のように実装した。図 6 の FPGA による PN モデルを基本とし、図 10 の競合学習ニューロン（以下 CO ニューロン）と図 11 制御ニューロンを配置する。入力パルスの頻度の計算については、すべての CO ニューロンにおいて時定数及び重みは同じであるため外部に一つ、図 12 に示す入力パルス頻度を計算するためのニューロン（Input Potential Control Neuron：以下 IP ニューロン）を用意することで等価な回路とした。

そして単一発火のための制御方法は、内部電位を調整することはしきい値を調整することと等価であるため、各 CO ニューロンの内部電位に対して直接加減算するのではなく外部からしきい値を与えて発火の判定を行うことで実現した。すなわち入力パルスの頻度から内部電位の減少させるには IP ニューロンから CO ニューロンに対してしきい値 $\theta_{in}(t)$ を与えることで等価な回路とした。同じく制御ニューロンからのしきい値の制御に関しても外部に図 13 のようなユニット

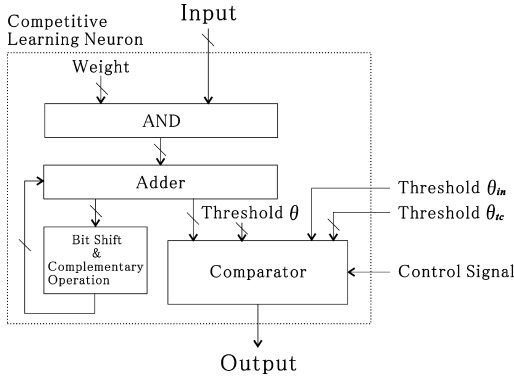


図 10 FPGA における競合学習ニューロン
Fig. 10 The competitive learning neuron on FPGA.

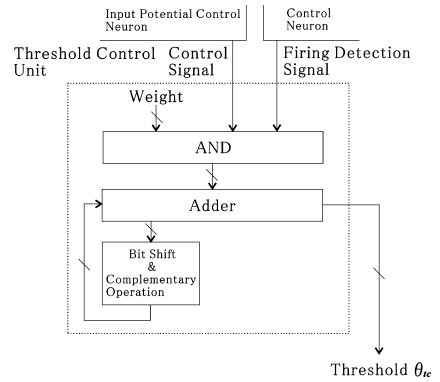


図 13 FPGA におけるしきい値制御を行うユニット
Fig. 13 The threshold control unit on FPGA.

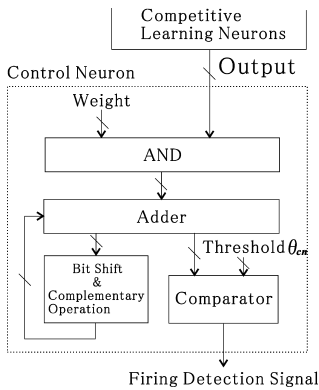


図 11 FPGA における制御ニューロン
Fig. 11 The control neuron on FPGA.

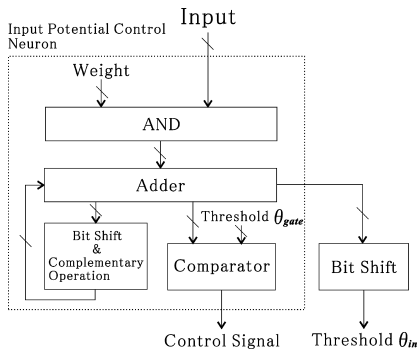


図 12 FPGA における入力パルス頻度を計算するニューロン
Fig. 12 The input potential control neuron on FPGA.

(Threshold Control Unit : 以下 TC ユニット) を無発火検出用と複数発火検出用につづつ用意してしきい値 $\theta_{tc}(t)$ ($\theta_{nfd}(t)$ と $\theta_{mfd}(t)$ の 2 種類がある) を CO ニューロンに与えることで等価な回路とした。以

上より, FPGA に実装した場合の各 CO ニューロンにおける出力は式 (5) のようになる。これは式 (4) と等価である。

$$o(t) = f_{usf} \left(\sum_{k=1}^n p_k(t) - \theta - \theta_{nfd}(t) + \theta_{mfd}(t) - \theta_{in}(t) \right) \quad (5)$$

続いて動作について説明する。まず入力パルスは各 CO ニューロンと IP ニューロンに入力され各ニューロンが動作する。CO ニューロンの出力は二つの制御ニューロンに送られ各制御ニューロンにて CO ニューロンが一つも発火していない若しくは複数個発火していることを検出して出力する (図 9 中 Firing Detection Signal)。一方, IP ニューロンの出力は CO ニューロンの比較部に送られ, IP ニューロンの出力が 0 の場合には入力パルス頻度が十分でないとして CO ニューロンは発火しないようにする (図 9 中 Control Signal)。また, 各制御ニューロンの出力は TC ユニットに送られる。TC ユニットにおいては, Control Signal が出力されているときのみ, Firing Detection Signal が送られた際に TC ユニット中のしきい値 θ_{tc} が加算されるようになっていいる。このしきい値は各 CO ニューロンに与えられ CO ニューロンのしきい値が調整される。IP ニューロンの出力には上記の Control Signal の他にしきい値があり, 入力パルス頻度である IP ニューロンの電位より一定の割合 β を与えたしきい値 θ_{in} が各 CO ニューロンに与えられる。これら二つのしきい値調整 (図 9 中 Threshold Control) により各 CO ニューロンのうち一つだけが発火できるよう制御さ

れる。

5. 実験

5.1 実験条件

前章までに説明した聴覚情報処理システムを用いて音源定位及び音源種類の識別実験を行う。

本システムの実装に用いた計算機及びFPGAの仕様を表1に示す。録音用マイクにはダミーヘッドKU100を用いた。また音源はダミーヘッドの左耳に対して垂直に約1m離れた位置を開始点とし、ダミーヘッドを中心とした円状に右耳まで移動させた。学習用の音には残響などの影響がない無響室で録音した音を用い、認識実験用の音には実環境で動作することを想定し広さ5.6m×6.0m、コンクリート壁とガラス窓のある室内で録音した音を用いた。

実験に用いた入力音源の種類としては、音源定位の学習のためにすべての周波数成分をもつ信号であるホワイトノイズと、音源種類の識別のための学習音に“警報ベル”、“インターホン”、“やかんでの水の沸騰音”、“雑踏”、“黒電話の呼び出し音”、“音声(成人男性の母音「あ」)”の合計7種類を用意した。

定位の解像度としては前方向30度ずつ合計7方向としている。ただし音源位置識別部では教師なし学習が行われているため各出力素子が正確な角度を示しているわけではない。音源位置識別部の出力素子を必要に応じて増やせば解像度を上げることも可能であるが[14]、本システムの応用先として聴覚障害者支援や室内での警告装置を考えており、細かな定位精度は必要なく大まかな方向さえ定位できればよいため、7方向という解像度でも問題なく、正確な角度は必要ないと判断した。

また音源種類の識別用に用いた学習音について、周波数-パルス変換部の出力を図14に示す。図14の横軸は時間、縦軸は周波数帯域番号であり上にあるほど高い周波数帯域である。各マス目における濃淡は出力

パルス信号の頻度を示しており、黒いほど多くのパルス信号が出力されたことを示す。なお“雑踏”に関してはその他に分類させる目的で用意してある。はじめにこれらの音を用いて表2のパラメータに従い計算機上で各CONPの学習を行った。その後、学習の結果

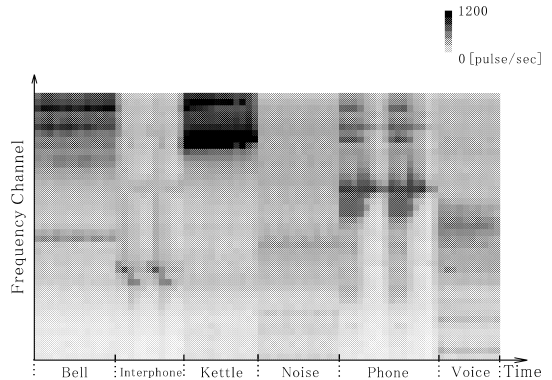


図14 音源種類識別実験のための学習音

Fig. 14 The learning signal for sound recognition.

表2 実験におけるパラメータ

Table 2 The parameters of each unit on experiment.

入力音	
サンプリング周波数	16[kHz]
量子化ビット数	8[bit]
周波数-パルス変換部	
音源定位に用いる周波数帯域数	15 [Channel]
音源種類の識別に用いる周波数帯域数	43 [Channel]
時間差検出部	
総シフト回数	41 [回]
出力ニューロン数	21 [個]
しきい値 θ_{mso}	1.0
時定数 τ_{mso}	0.35 [ms]
不応期 RP_{mso}	1.0 [ms]
競合学習ニューロン	
音源位置識別部 CO ニューロン数	7 [個]
周波数パターン量子化部 CO ニューロン数	10 [個]
音源種類識別部 CO ニューロン数	6 [個]
しきい値 θ	1.0×10^{-4}
制御しきい値 θ_{gate}	150.0
入力パルス頻度に対する割合 β	3.125×10^{-2}
膜電位減衰の時定数 τ_p	20 [ms]
不応期 RP_{CL}	1.0 [ms]
学習係数 α	2.0×10^{-7}
学習回数	1000
学習半径の収束回数	800
無発火検出ニューロン	
時定数 τ_{no}	0.5 [ms]
しきい値 θ_{no}	-1.0×10^{-3}
CO ニューロンへの結合重み	0.5
複数発火検出ニューロン	
時定数 τ_{multi}	1.0 [ms]
しきい値 θ_{multi}	2.0
CO ニューロンへの結合重み	0.5

表1 使用する機器の仕様

Table 1 The specifications of using instruments.

計算機		FPGA	
CPU	Pentium4 3.8 GHz	開発ボード	Altera 社 StratixII EPS60
メモリ	2 GB	総回路数	48,352 [ALUTs]
OS	WindowsXP	計算機との 通信規格	USB 2.0
コンパイラ	Microsoft Visual Studio C++ .NET	論理合成 ソフトウェア	Altera 社 QuartusII 5.0

表 3 FPGA 上での各回路の回路数
Table 3 The number of each circuit on FPGA.

部位	回路数 [ALUTs]
FPGA 入出力部	91
時間差検出部	15,092
音源位置識別部	8,882
周波数パターン分類部	4,353
音源種類識別部	1,591
合計	30,009

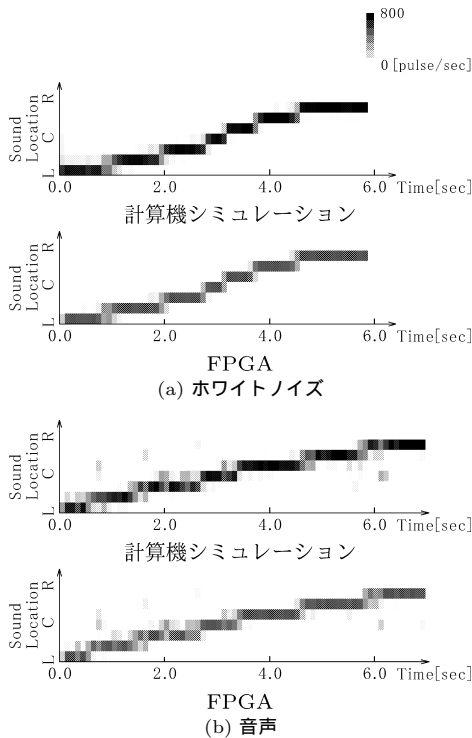


図 15 音源位置識別部における出力結果
Fig. 15 The results of sound localization.

得られた重みやパラメータを用い FPGA 上に各部を実装した．実装の際に必要なとした各部の回路数を表 3 にあげる．合計が 30,009 [ALUTs] であるため使用した開発ボードにすべての回路を搭載することが可能である．

以降の節では，学習音とは別に録音した音を使用して実験を行った際の各出力結果について検討していく．

5.2 音源定位実験

音源位置の定位を行うにはまず時間差検出部において左右の音信号の時間差を検出する．次に，その時間差情報を音源位置識別部に入力することで音源の定位を行う．図 15 にホワイトノイズと音声における音源位置識別部の出力結果を示す．図 15 における横軸は

時間を示し，縦軸は CONP の出力素子番号であり下から音源位置の左 → 中央 → 右と対応している．

結果図 15 (a) より学習音であるホワイトノイズにおいては音源の位置に対応した素子が多く発火しており，音源の定位ができているといえる．図 15 (b) の音声についてはほとんどころ音源の位置と対応していない素子が発火しているが，最も発火頻度の高い素子は音源の位置に対応できている．

今回の音源定位実験では学習データにはホワイトノイズしか用いていないが音声などの他の音に関しても定位できている．この理由は，CONP へ入力されるパターンは各周波数帯ごとの時間差情報であり，音が違っていても出力される周波数帯が違うだけで時間差が同じであれば同じ箇所に入力される．よってすべての周波数成分を含むホワイトノイズの時間差パターンで学習を行えば部分的な周波数成分をもつ音の時間差パターンを包括することになり，各方向を学習した重みが最も近くなることが保証されるからである．しかし音声のように部分的な周波数成分のみもつ音の場合，すべての周波数成分を含むホワイトノイズに比べて時間差情報として得られる情報は少なくなるため定位が困難になる．このため図 15 のように音声の場合ではほとんどころ異なった素子が出力することになる．

5.3 音源種類の識別実験

音源種類の識別を行うにはまず周波数-パルス変換部で得られたパルス出力を周波数パターン量子化部へ入力しベクトル量子化を行う．次に得られたベクトル量子化されたパターンを音源種類識別部に与えて音源種類の識別を行う．音源定位実験と同じ音を入力した場合の FPGA 上及び計算機シミュレーションでの音源種類識別部の各音に対応する素子の出力頻度を表 4 に示す．表中の数字は各入力音を与えた場合における各出力素子の単位時間当りの平均出力数 ([pulse/s]) である．また下線は最大頻度となる出力結果を示す．

表 4 よりそれぞれ各入力音に合わせて対応した素子が最も多く出力しており正しく識別ができていることが分かる．インターホン及び電話の結果に関しては雑踏を示す素子も多く発火している．これらの音は断続的な音であり音と音の間で雑踏を示しているためである．理想的には計算機シミュレーションにおける結果のように対応した素子のみが多く出力することが望ましいが，FPGA における結果では雑踏などの素子も多く発火している．これは FPGA 上では膜電位や内部電位を固定小数点で表現しているため，各段における

表 4 音源種類識別部の出力結果
(FPGA/計算機シミュレーション)

Table 4 The results of sound recognition.

出力結果	入力音		
	警報ベル	インターホン	水の沸騰音
警報ベル	470.6 / 997.0	0.1 / 0.0	70.3 / 59.0
インターホン	0.1 / 0.0	251.7 / 530.2	0.1 / 0.0
水の沸騰音	0.1 / 1.2	0.0 / 0.0	365.9 / 922.3
雑踏	0.1 / 0.0	201.6 / 239.3	0.1 / 5.7
電話	1.1 / 0.2	0.9 / 0.0	0.7 / 0.1
音声	0.1 / 0.0	5.4 / 0.0	0.1 / 0.0

出力結果	入力音		
	電話	音声	雑踏
警報ベル	5.6 / 8.3	0.0 / 0.0	0.2 / 7.0
インターホン	0.1 / 2.0	0.1 / 0.0	0.2 / 0.0
水の沸騰音	0.1 / 0.0	0.0 / 0.0	0.0 / 0.0
雑踏	182.9 / 252.0	1.1 / 6.8	487.6 / 989.9
電話	274.4 / 723.0	0.4 / 0.0	1.0 / 1.5
音声	0.1 / 0.3	486.5 / 992.4	0.0 / 0.0

表 5 各部の処理時間の比較
Table 5 The time cost of each unit.

部位	計算機	
	シミュレーション [s]	FPGA [s]
入力音長	8.76	8.76
時間差検出部	36.82	0.83
音源位置識別部	4.59	0.85
周波数パターン量子化部	1.11	0.71
音源種類識別部	0.36	0.61
全動作	42.28	1.09

CONP において計算誤差が多くなっていたことが原因である。しかし最大頻度となる出力素子は入力音と対応しており、システムの最終的な出力として最大頻度となる出力素子のみを考慮すれば実用上は問題ない。

5.4 処理時間及び従来手法との回路数の比較

PN モデルは演算が単純であり各素子が独立・非同期で動作することができるためハードウェア上に実装することで非常に高速な演算を期待できる。これを確認するため入力音として音長が 8.76 [s] のホワイトノイズを用い本システムを動作させた場合の、計算機シミュレーションと FPGA における各部の単独の処理時間及び全動作を行った場合の処理時間を表 5 に示す。なお計算機シミュレーションで利用した環境は表 1 のとおりである。また表 6 にハードウェア上に実装した場合に並列に動作させることができる各 PN モデルの個数や入力の個数などによる演算の個数を示す。

結果より、音源種類識別部以外では FPGA による処理時間は計算機シミュレーションと比べて短い処理時間となっている。この原因は、計算機上では表 5 に示した個数だけ計算を順次行わなければならないから

表 6 ハードウェア実装時に並列化が可能な演算個数

Table 6 The loop number of operation in each unit.

部位	個数
時間差検出部	14,480
音源位置識別部	2,534
周波数パターン分類部	493
音源種類識別部	82

表 7 各手法における並列化した際に用いる演算回路の個数

Table 7 The circuit number of pallareled operation in each method.

	提案システムの音源定位部分 ($F = 15, M = 21, C = 7$)
加算	$2F \times 10 + 2F(10M + 82)$ $+ (C + 1)(F \cdot M + 4) + 2(C + 8) \rightarrow 11,612$ [11,312]
積算	$2F \times 16 \rightarrow 480$ [0]
	提案システムの音源種類識別部分 ($F = 43, Q = 10, C = 6$)
加算	$F \times 10$ $+ (Q + 1)(F + 4) + 2(Q + 8)$ $+ (C + 1)(Q + 4) + 2(C + 8) \rightarrow 1,109$ [679]
積算	$F \times 16 \rightarrow 688$ [0]
	CSP 法 ($N = 256$)
加算	$2 + 2N + 4 \rightarrow 518$
積算	$2N \log_2 N + 7N + 2N \rightarrow 6,400$
	GMM 法 ($N = 256, C = 6, L = 26$)
加算	$L + 1 + C(L + 1) + C + 1 \rightarrow 196$
積算	$N \log_2 N + 4L$ $+ 2C^2 + C(L + 1) + 1 \rightarrow 2,387$

である。逆に音源種類識別部では PN モデルの数が少ないため、FPGA に実装して並列化するよりも高速な計算機で計算した方が処理が速くなる。しかし現実問題におけるパターン認識では多数の入力次元・識別数を必要とするため、表 5 中の「全動作」のように FPGA に実装することで計算機と FPGA の間に大きな処理時間の差を得ることが可能である。

続いて本研究で提案するシステムと CSP 法 [4], [8], GMM による方法 [9], [10] における実装上の回路規模について比較する。表 7 に各手法における単位時間当りに並列化が可能な加算・積算の演算回数を示す。

ただし CSP 法に関しては高速化のために DFT の代わりに FFT を用い、cos などの演算についてはあらかじめ計算されたテーブルを用いる。時間窓の長さは $N = 256$ とする。GMM 法についても Mel Frequency Cepstral Coefficients (MFCC) を計算する前処理部分に関して FFT とテーブルを用い、時間窓の長さは $N = 256$ とする。MFCC の次元数は $L = 26$ 次元とし、識別数は $C = 6$ とする。提案システムに関して

は計算機上で行われる周波数-パルス変換における演算も含み、PN モデルにおける減衰には積算ではなく 4.1 で説明したビットシフトと補数演算による実現とし 2 個分の加算と数えた。また使用する周波数帯域数を F 、時間差検出部の出力ニューロン数を M 、周波数パターン量子化部の出力ニューロン数を Q 、各識別数を C と表記する。かぎ括弧内は本研究で FPGA に実装した部分を示す。

表 7 より各手法を個別に見た場合には提案システムの方が全回路の個数が多い。しかし提案システムでは CSP や GMM と比較して積算の個数が少ないことが分かる。ここで例として 16 bit 同士の二つの数について加算と積算を行う回路を論理合成した場合の回路数について述べる。論理合成には本研究で実装に用いた開発ソフトウェアである Altera 社の QuartusII を使用した。この場合、加算を行う場合の必要な回路数が 16 [ALUTs] になるのに対して、積算の場合には 235 [ALUTs] の回路が必要となる。よって表 7 に示すように実装を行った場合、必要な回路数の概算は提案システムで 477,000 [ALUTs] 程度、CSP 法と GMM 法を用いた場合で 2,076,000 [ALUTs] 程度となり最終的な回路規模では提案システムの方が小さくなる。加えて提案システムにおいてはパルス変換までの前処理は同じ回路を用いることができるため、音源定位や音源種類の識別のために個別に手法を用いるよりも回路規模を更に小さくすることが可能である。

以上、本節の実験結果から提案システムを FPGA に実装することにより、計算機シミュレーションと比較して非常に短い時間で結果を出力することができ、入力信号に対してリアルタイムに結果を出力することも可能であり、また、従来手法と比較して少ない回路数で実装できることが確認できた。

6. む す び

本論文ではパルスニューロンモデルを FPGA 上に実装して時間差検出モデル、競合学習ネットワークを構成し、入力される音情報から音源定位と音源種類の識別を同時に行うことのできる聴覚情報処理システムを構築した。実験の結果、実装したシステムが計算機シミュレーションと同様に音源定位と音源種類の識別ができることを示し、更にパルスニューロンモデルをハードウェア化することにより処理時間が大幅に短縮され、更に従来手法と比較して少ない回路数で実装できることが明らかになった。

本システムは、基本素子であるパルスニューロンモデルの組合せのみで構成されており、組合せの方法や学習の内容によって音以外の識別を行うことも可能である。更にハードウェア化により高速な処理が可能であるため様々な機器への応用が期待できる。

聴覚障害者支援システムとして実際の環境で動作させることを想定すると複数の音源が存在する場合への対処が必要である。これについては、各音の発生する時間さえ異なれば認識部分に黒柳らの提案した方法 [19] を適用することで対応できると考えられ、今後の検討課題としたい。

謝辞 本研究の一部は財団法人堀情報科学振興財団、文部科学省科学研究費・若手研究 (B)、及び岐阜県知的クラスター創生事業「ロボティック先端医療クラスター」による支援を受けた。記して感謝致します。

文 献

- [1] J.O. Pickles, "An Introduction to the Physiology of Hearing," ACADEMIC PRESS, 1988.
- [2] P. Julian, A.G. Andreou, L. Riddle, S. Shamma, D.H. Goldberg, and G. Cauwenberghs, "A comparative study of sound localization algorithms for energy aware sensor network nodes," IEEE Trans. Circuits Syst. I, vol.51, no.4, pp.640-648, April 2004.
- [3] M.S. Brandstein, J.E. Adcock, and H.F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," IEEE Trans. Speech Audio Process., vol.5, no.1, pp.45-60, 1997.
- [4] T. Nishiura, T. Yamada, S. Nakamura, and K. Shikano, "Localization of multiple sound sources based on a CSP analysis with a microphone array," IEEE ICASSP SP-L4-6, pp.1053-1056, June 2000.
- [5] C. Couvreur, Environmental Sound Recognition : a Statistical Approach, PhD Thesis, Faculté Polytechnique de Mons, Belgium, June 1997.
- [6] K. Kashino, K. Nakadai, T. Kinoshita, and H. Tanaka, "Organization of hierarchical perceptual sounds : Music scene analysis with autonomous processing modules and a quantitative information integration mechanism," Proc. 1995 IJCAI, 1995.
- [7] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K.J. Lang, "Phoneme recognition using time-delay neural networks," IEEE Trans. Acoust. Speech, Signal Process., vol.37, no.3, pp.328-339, 1989.
- [8] 奥村 元, 趙 國, 西浦敬信, 山下洋一, "分散マイクロホンシステムを用いた音源位置同定," 信学技報, SP2006-139, 2007.
- [9] J.C. Segura, A. de la Torre, M.C. Benitez, and A.M. Peinado, "Model-based compensation of the additive noise for continuous speech recognition. Experiments using AURORA II database and tasks," Proc. Eurospeech 2001, vol.1, pp.221-224, 2001.

- [10] 李 晃伸, “大語彙連続音声認識エンジン Julius の開発の進展,” 信学技報, SP2005-111, 2005.
- [11] W. Maass and C.M. Bishop, Pulsed Neural Networks, MIT Press, 1998.
- [12] 二俣宣義, 黒柳 奨, 岩田 彰, “FPGA のためのパルスニューロンモデルの実装手法,” 信学技報, NC2001-211, 2002.
- [13] 黒柳 奨, 岩田 彰, “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,” 信学論 (D-II), vol.J79-D-II, no.2, pp.267-276, Feb. 1996.
- [14] 黒柳 奨, 岩田 彰, “聴覚情報処理システムのためのパルスニューロンモデルを用いた競合学習ネットワーク,” 信学論 (D-II), vol.J87-D-II, no.7, pp.1496-1504, July 2004.
- [15] L.A. Jeffress, “A place theory of sound localization,” J. Comp. Physiol. Psychol., vol.41, pp.35-39, 1948.
- [16] 岩佐 要, 黒柳 奨, 岩田 彰, “パルスニューロンモデルによる時間差検出ネットワークの改良,” 信学技報, NC2005-150, 2006.
- [17] T. Kohonen, Self-Organizing Maps, Springer-Verlag, 1995.
- [18] 黒柳 奨, 岩田 彰, “パルスニューロンモデルのための教師あり学習則,” 信学技報, NC97-151, 1998.
- [19] 黒柳 奨, 岩田 彰, “聴覚情報処理のための競合学習パルスニューラルネットワーク CONP-II,” 信学論 (D), vol.J90-D, no.1, pp.115-125, Jan. 2007.
- (平成 18 年 12 月 18 日受付, 19 年 4 月 23 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ゼ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 同大・電気情報・教授。平 14 同大・副学長, 平 16 同大大学院・教授, 現在に至る。ニューラルネットワーク, 情報セキュリティに関する研究に従事。工博。H5 年度本会論文賞受賞, H10 年度情報処理学会 Best Author 賞受賞。情報処理学会, 日本生体医工学会, 日本心電図学会, 日本神経回路学会, 日本医療情報学会各会員, IEEE Senior Member。



岩佐 要 (学生員)

平 15 岐阜高専専攻科・電子システム工学専攻了。平 17 名工大大学院工学研究科情報工学専攻博士前期課程了。現在, 名工大大学院工学研究科情報工学専攻博士後期課程に在学中。ニューラルネットワーク, 時系列情報処理に関する研究に従事。



黒柳 奨 (正員)

平 3 名工大・電気情報卒。平 5 同大大学院博士前期課程了。平 8 同大学院博士後期課程了。同年名工大・電気情報・助手。平 15 同大大学院・助手, 平 18 同大学院・助教授, 現在に至る。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。博士(工学)。日本音響学会, 日本神経回路学会, 日本生体医工学会各会員。