

## ニューラルネットワークを用いた雑音除去フィルタの構成 — デジタル補聴器の前処理への応用 —

並木寿枝<sup>†</sup> 古本友紀<sup>†</sup> 吉村宏紀<sup>†</sup> 清水忠昭<sup>†</sup> 井須尚紀<sup>†</sup> 菅田一博<sup>†</sup>

<sup>†</sup>鳥取大学工学部 〒680-0945 鳥取市湖山町南 4-101

E-mail: <sup>†</sup>namiki@ike.tottori-u.ac.jp

**あらまし** デジタル補聴器の前処理への応用を目的とし、砂時計型ニューラルネットワーク (Sandglass-type Neural Network: SNN) が KL 変換 - KL 逆変換と等価な処理を行うことを利用し、音声・雑音の大きさの経時的な変化に対応した雑音除去と音量の補償を行う適応的アルゴリズムを提案した。

入力信号の KL 変換成分から入力信号に含まれる音声成分の期待値を求め、期待値の大きさに比例した係数を多段接続した SNN 雑音除去フィルタ (Cascaded SNN Noise Reduction Filter: CSNNRF) の各段の出力に乗算することで、出力信号に含まれる雑音成分のパワーを減少させて SN 比の改善を図った。次に、補聴器の出力音量が絶えず変動するのは、聴覚上好ましいことではないので、音量の逆数を増幅率とし CSNNRF の出力総和に乗算し、音量の補償を行った。最後に CSNNRF の段数を適応的に変化させることで雑音除去を行った。音声成分と雑音成分の大きさが変化する入力信号を用いた雑音除去実験の結果、雑音は低減され音量補償も適切に行われていることが示された。

**キーワード** 砂時計型ニューラルネットワーク、適応的雑音除去フィルタ、デジタル補聴器

## A Construction of Noise Reduction Filter by Neural Network — Application for Digital Hearing Aid Appliance as Preprocessing —

Toshie NAMIKI<sup>†</sup> Yuki FURUMOTO<sup>†</sup> Hiroki YOSHIMURA<sup>†</sup> Tadaaki SHIMIZU<sup>†</sup>

Naoki ISU<sup>†</sup> and Kazuhiro SUGATA<sup>†</sup>

<sup>†</sup> Faculty of Engineering, Tottori University 4-101minami, koyamacyou, Tottori city, 680-0945 Japan

E-mail: <sup>†</sup>namiki@ike.tottori-u.ac.jp

**Abstract** We proposed an adaptive noise reduction filter composed of Cascaded Sandglass-type Neural Network (CSNNRF) to develop a hearing aid appliance. It was known that Sandglass-type Neural Network (SNN) has property of processing signals which is equivalent to KL transformation - KL inverse transformation after learning. First, we calculated the expected value of the speech element from KL transformation element of input signal. We decided coefficients in proportional to the expected value and multiplied them to the output of each stage of CSNNRF. Next, the sound volume was adjusted by multiplying amplification rate to the total sum of output signals of CSNNRF. Here, amplification rate was determined as the inverse of the sound volume. Finally in order to improve S/N ratio, the number of stages in CSNNRF was altered adaptively. As a result of the noise reduction experiment, it was shown that the noise was reduced adaptively and the sound volume was adjusted appropriately.

**Keyword** Sandglass-type Neural network, Noise reduction adaptive filter, Digital hearing aid appliance

## 1. はじめに

補聴器では、利用者の聴覚特性や入力信号の音圧に合わせて各人に適合した周波数特性を備え、伸張増幅や圧縮増幅などの非線形な増幅を行う機能が重要である。デジタル補聴器の増幅機能を効果的に働かせるためには前処理として入力信号の雑音除去が望まれる。

しかし、実環境を想定した場合、入力される音声の大きさ、雑音の大きさが経時的に変化するために、適応的雑音除去をオンラインで行うことを難しくしている。

そこで、我々は砂時計型ニューラルネットワークを用いた雑音除去フィルタを適応的に構成するアルゴリズムを提案し、その有効性を雑音除去実験によって示した。

## 2. 砂時計型ニューラルネットワーク雑音除去フィルタ

砂時計型ニューラルネットワーク(Sandglass-type Neural Network: SNN)とは、入・出力層のユニット数( $N$ 個とする)を同数とし、中間層のユニット数( $1 \leq M \leq N$ )をそれよりも少なくした階層型ニューラルネットワークである<sup>[1]</sup>。SNN 雑音除去フィルタ(SNN Noise Reduction Filter: SNNRF)は、SNN で構成した雑音除去フィルタである。

3層 SNN は、応答関数に線形関数を用いることでニューラルネットワークの学習が収束した後は、Karhunen-Loève 変換(KL 変換) - KL 逆変換と等価な処理を行うことが知られている<sup>[2][3]</sup>。KL 変換は、入力信号の共分散行列の固有ベクトルを変換係数とする直交変換であり、高能率符号化の理論的限界とされている。この KL 変換 - KL 逆変換を用いた雑音除去では、音声成分の共分散行列のランク数が推定された時に、信号成分を欠如することなく雑音成分を最もよく除去できる<sup>[4]</sup>。

しかし、信号成分のランク数に応じて中間層ユニットの個数  $M$  を変化させなければならず、その度に SNN の結合係数の学習(更新)をやり直すことはオンライン処理には不利である。そこで、中間層ユニット1個の SNN を多段接続した雑音除去フィルタ(Cascaded SNN Noise Reduction Filter: CSNNRF)を構成した。但し、接続した SNN の個数を接続段数と呼ぶ( $1 \leq m \leq M$ )。CSNNRF は、入力信号が入力される毎に逐次的に結合係数の学習(更新)を行う。1 段目 SNN には入力信号を2 段目以降の SNN には前段の誤差信号(入力信号と出力信号の差)を入力する。CSNNRF の出力信号は、各段の SNN の対応するユニットの出力を加算することで得られる。また、CSNNRF の各段は平均パワーの大きい順に KL 変換成分を受け持つ。そのため、CSNNRF は、接続段数を適切に調節することで音声成分の KL 変換成分のみを残すことで雑音除去を行うことになる。

図 1 に提案する CSNNRF の構成図を示す。但し、図中の係数については 3~5 章で後述する。音声成分と雑音成分の大きさは経時的に変化するため、適応的に CSNNRF の接続段数を調節する必要がある。

さて、我々は、今までに音声成分の大きさを一定とし、雑音成分の大きさが変化する場合に対応した CSNNRF を提案してきた<sup>[4]~[6]</sup>。しかし、音声成分の大きさも経時的に変化すると考える方が一般的である。そこで、音声成分と雑音成分の大きさの変化にも対応した適応的アルゴリズムを提案した<sup>[7]</sup>。

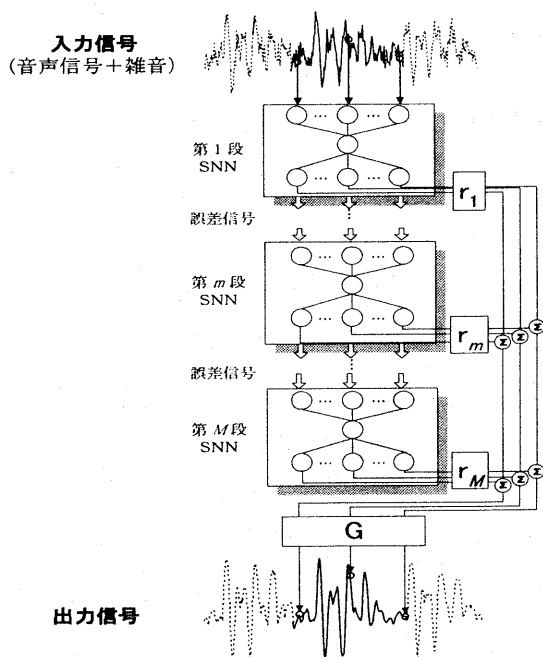


図 1 CSNNRF の構成図

## 3. 音声成分と雑音成分のパワーの推定

通常の会話音声などにおいて、有声音区間は音声の大部分を占める。音声成分の KL 変換成分パワー比(寄与率)と雑音成分の KL 変換成分パワー比の相違に着目し、それぞれのパワーの推定を行った。

まず、雑音の重量されていない文章音声試料を CSNNRF に入力し、中間層ユニットの出力パワーの区間平均を求めることで、音声成分の KL 変換成分のパワーを求めた。音声信号には、日本音響学会編集の「研究用連続音声データベース(Vol. 2)」(1991 年)より選択した男女各 12 名の文章音声試料を用いた(サンプリング周波数 16 k Hz, 量子化ビット数 16bits)。求められた KL 変換成分のパワーの分布に LBG(Linde Buzo Gray: LBG) アルゴリズムを適用し、有声音の分布の

中で代表値とみなせる値(代表ベクトル,  $M$ 次元ベクトル)を求めた. 求められた代表ベクトルを推定に用いる有声音の KL 変換成分のパワー比の標準モデルとした. 但し, モデル数は日本語母音数より多い 8 種類とした.

次に音声成分と雑音成分のパワーの推定を行った. 雑音の混入された入力信号を CSNNRF に入力し得られた中間層ユニットの出力パワーを 20Hz の IIR ローパスフィルタに通した. この値を  $\bar{y}_i$  とする. これは, 入力信号の第  $i$  KL 変換成分である. 次に最小 2 乗法を用いて 8 種類の標準モデルと  $\bar{y}_i$  をマッチングし, 誤差  $\varepsilon^{(p)} = \sum_{i=1}^M (\bar{y}_i - \hat{y}_i^{(p)})^2$  が最小となるように回帰直線に当てはめ, 音声成分のパワー  $\hat{S}^{(p)}$  と雑音成分のパワー  $\hat{N}^{(p)}$  を式(1)で推定した.

$$\hat{Y}_i^{(p)} = \hat{S}^{(p)} x_i^{(p)} + \hat{N}^{(p)} \quad (1)$$

但し,  $p$  は標準モデルの数である. ( $1 \leq p \leq 8$ )

#### 4. 音声成分の期待値

雑音成分の各段の KL 変換成分のパワーは, 白色雑音の場合同じ大きさである. しかし, 音声信号の KL 変換成分のパワー比(寄与率)は異なるため, 段数が増えるにしたがって音声成分の KL 変換成分のパワーは減少する. そのため, 入力信号の KL 変換成分のパワーが小さなおとこでは, 逆に音声成分の KL 変換成分のパワーに比べ, 雑音成分の KL 変換成分のパワーの割合が大きくなる. そこで, CSNNRF の出力パワーの小さなおとこでは, 出力信号の KL 変換成分のパワーに含まれる雑音成分の KL 変換成分のパワーを小さくすることで SN 比の改善を図った.

まず, 式(1)によって求められた雑音成分のパワーを 5Hz の IIR ローパスフィルタに通して得られた値を全段数 ( $M=20$ ) で割ったものを各段の雑音成分のパワーの推定値  $\bar{N}_i = \hat{N}^{(p)} / M$  とする. 次に中間層ユニットの出力パワーを 20Hz の IIR ローパスフィルタに通して求められた区間平均値  $\bar{Y}_i$  ( $i=1, 2, \dots, M$ ) を用いて以下の式(2)により入力信号の第  $i$  KL 変換成分中の音声信号の期待値を求めた. これを, 各段の音声成分のパワーとする.

$$\begin{aligned} \hat{S}_i &= E[s_i | y_i, \bar{S}_i, \bar{N}_i] \\ &= r_i y_i \\ r_i &= \frac{\bar{S}_i}{\bar{S}_i + \bar{N}_i} = \frac{\bar{Y}_i - \bar{N}_i}{\bar{Y}_i} \end{aligned} \quad (2)$$

但し,  $y_i$  を第  $i$  段 SNN の出力信号とし, それに含まれる音声信号を  $s_i$  とする. 各段で求めた音声信号の期待値は, 入力信号の第  $i$  KL 変換成分中における音声成分の割合が大きいほど大きくなり, 割合の小さい段ほど小さくなる. 雑音成分がほぼ均等に出力されると考えると音声成分が少なく, 雑音成分の割合が多くなる段ほど信号が小さく減衰される. このことから, 各段の出力信号の KL 変換成分のパワーに含まれる雑音成分の割合が少なくなり雑音が低減され, SN 比が改善される. 但し, 音声の過度の歪み为了避免するため  $r_i$  の下限値を 0.3, 上限値を 1 とした.

#### 5. 音量の補償

デジタル補聴器への応用を考えると, 音量が変化する場合, 小さな音は聞き取りづらく, 大きな音は耳障りな音となる. また, 断続的な無音区間に雑音が重畳されている場合, 無音区間をそのまま増幅すると雑音も増幅してしまうことになり, 聞き取りやすさの低下を招く. そこで, 聞き取りやすさを向上させるために, 適応的に音量の補償を行う方法を提案した.

音声の聴感上の聞こえの大きさ(音量)は, 大きく聞こえる部分の音圧パワーによって決まると考え, 0.5 秒毎に推定した信号成分のパワーが最大になる値を求め,  $(\log_2)$  乗した. 求められた値を 0.5Hz の IIR ローパスフィルタを通して平均化することにより音量を求め, その逆数を 1.5Hz の IIR ローパスフィルタに通して平均化することにより増幅率とした. 但し, 無音区間では推定した信号成分のパワーが小さくなることにより増幅率の跳ね上がり, 雑音のみを大幅に増幅してしまうのを避けるため, 増幅率を強制的に低減した.

無音区間は, 推定した信号成分パワーあるいは SN 比が 1 秒間持続して閾値を下回るにより検出した. また, 無音の割合が 0.9 よりも多く 2 秒間持続する場合は無音区間とした.

0.5 秒間信号成分のパワーと SN 比がどちらも閾値を超える場合は音声区間とした. また, 無音の割合が 0.5 より少なくなれば, その区間は音声区間とした. 但し, 信号成分のパワーの閾値は 0.316228(-5dB), SN 比の閾値は 0.240253(SN 比-5dB)とした.

更に, 通常の会話において, 無音区間後に入力される音量は, 無音区間以前の音量と大きく変化することはないと考え, 無音区間に入る直前の音量を 0.5 秒間維持した. 但し, 音量がある程度小さく雑音が多い場合, 音量を増幅しても元の入力信号よりも雑音が大きくなり却って耳障りな音となる. そのため, 増幅率を 1.0 とし, 入力信号の音量をそのまま維持した.

## 6. 接続段数の決定

音声の聞き取りやすさを向上させるためには、音声の歪みを小さくして、雑音をより多く除去することが必要である。CSNNRFは接続段数を少なくすることで、出力信号に含まれるKL変換成分の個数を少なくして、雑音除去率を上げるが、音声の歪みは大きくなる。一方、接続段数が多いと雑音除去率は低いが音声の歪みは小さくなる。そのため、雑音小さければ接続段数を多くし、雑音が大きければ接続段数を少なくするという適応的な段数変化を行うのがよい。

しかし、接続段数の急激な変化は、雑音除去率を急激に変化させるため、音声の聞き取りやすさを低下させる。そこで段数の変化を雑音の大きさに応じて決定するような段数決定アルゴリズムを提案したが<sup>[5]</sup>、出力信号に残った雑音成分のパワーが大きいと聴感上は聞き取りやすい音声とはならない。そこで雑音除去率を最大にするのではなく、出力信号に含まれる雑音成分のパワー変化を極力抑制させるように接続段数を変化させる方法を提案した。音量補償後の出力信号に含まれる雑音成分のパワーが一定値を越えない、つまりSN比が-5dBを超えないように式(3)によりCSNNRFの接続段数 $M$ を変化させた。但し、音声の過度の歪みを避けるため、接続段数の下限を7段、上限を20段とした。

$$N_M = G \sum_{m=1}^M N'_m \quad (3)$$

$N_M$  : 1~ $M$ 段SNNの出力信号に含まれる

雑音成分のパワーの総和

$N'_m$  : 各段SNNの出力に含まれる雑音成分のパワー

$G$  : 増幅率

## 7. 雑音除去実験

提案するアルゴリズムの有効性を調べるため、雑音の大きさのみが変化する場合、音声と雑音の大きさが変化する場合の2種類について雑音除去実験を行った。

### 1. 雑音の大きさが変化する入力信号を用いた実験

- 無雑音
- 平均SN比が5dBで定常な雑音 (小さな雑音)
- 平均SN比が-5dBで定常な雑音 (大きな雑音)
- 平均SN比が0dB、雑音のパワーの変化の振幅が5dB、周期が12秒で変化する雑音 (緩やかに大きさが変化する雑音)

- 平均SN比が0dB、雑音のパワーの変化の振幅が5dB、周期が4秒で変化する雑音

上記のような入力信号を用い、それぞれの入力信号の間は3秒の無音区間で連結した。

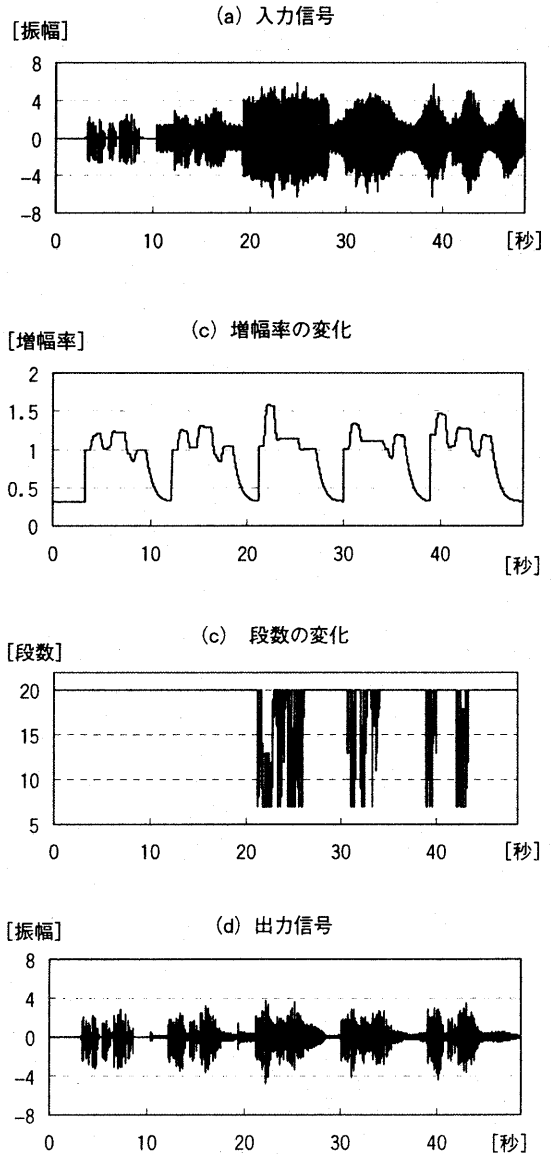


図2 雑音の大きさが変化する場合の雑音除去実験

図2(a)に入力信号を増幅率の変化を図2(b)に示す。各音声の増幅率は雑音が大きくなると若干大きくなる。これは、4章で述べたように雑音が大きくなると音声

信号の期待値が小さくなるため、出力に乗算する係数が小さくなり信号を減衰するため逆に増幅率が大きくなったためである。更に無音区間においては増幅率を下限値(0.3)まで下げることで雑音の低減がされていることがわかる。以上の結果より雑音の有無や変化にかかわらず適切な音量補償が行われていることがわかる。

次に段数の変化を図2(c)に示す。小さな雑音の時は、接続段数を上限の20段と固定し、音声の歪みを小さくしていることがわかる。また、雑音が大きいた時は、適応的に段数を変化させることで、音声は歪む多少が雑音除去率を高くしていることがわかる。

図2(d)に出力信号を示す。結果、すべての音声に対して適切な音量補償が行われ、音量がほぼ等しくなっている。また、出力信号に含まれる雑音もかなり低減されている。

## 2. 音声の大きさが変化する入力信号を用いた実験

- 基準音量  
(信号成分パワーが1倍)
- 大きな音声  
(信号成分パワーが基準音量の6倍)
- 小さな音声  
(信号成分パワーが基準音量の0.3倍)
- 徐々に大きくなっていく音声  
(信号成分パワーが0.3倍から6倍)
- 徐々に小さくなっていく音声  
(信号成分パワーが6倍から0.3倍)

上記のような入力信号を用い、それぞれの入力信号の間は1と同様に3秒の無音区間で連結した。雑音は、無雑音、定常な雑音(基準音量に対して平均SN比0dB)、大きさが変化する雑音(基準音量に対して平均SN比0dB、雑音成分パワーの変化の振幅5dB、変化の周期4秒)を付加した3種類の実験を行った。

図3(a)に雑音の大きさが変化する雑音を付加した入力信号を増幅率の変化を図3(b)に示す。結果、ほぼ適切に音量補償が適切に行われていることがわかる。

しかし、音量が大きな音声、小さな音声の場合に音声開始点において適切な音量補償が行われていない。これは、音声成分と雑音成分のパワーの推定の際に推定誤差が生じたためであると考えられる。

次に段数の変化を図3(c)に示す。接続段数は雑音の大きさの変化に応じて適応的に変化していることがわかる。

図3(d)に出力信号を示す。結果、すべての音声に対して適切な音量補償が行われ、音量がほぼ等しくなっていることがわかる。

また、無雑音時のCSNNRF処理時においても適切な

音量補償が行われ、接続段数もほぼ20段に固定されることで音声を歪ませないことが確認できた。

更に、定常な雑音付加時においても適切な音量補償と接続段数決定が行われることがわかった。

以上の結果より提案するアルゴリズムを用いたCSNNRFは音声の大きさ、雑音の大きさが経時的に変化する場合に有効なフィルタであることが示された。

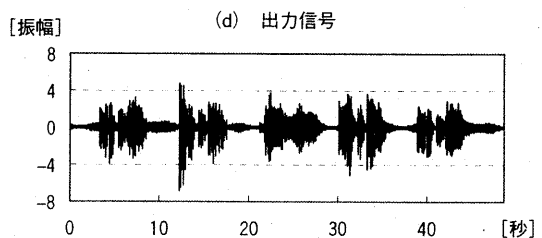
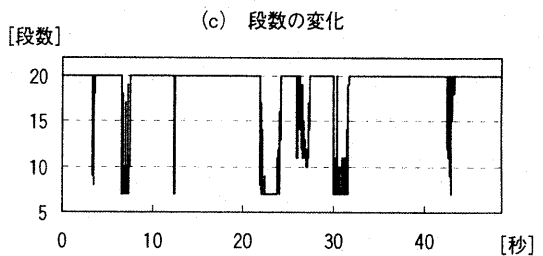
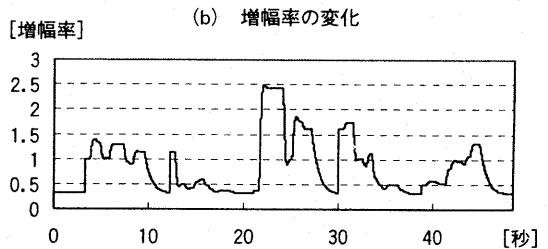
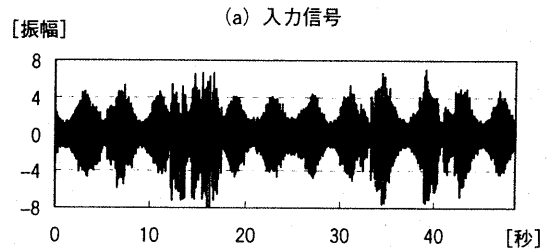


図3 音声の大きさが変化する場合の雑音除去実験

## 8. まとめ

デジタル補聴器への前処理への応用を目的として、線形な応答関数を用いた砂時計型ニューラルネットワーク (SNN) が KL 変換-KL 逆変換と等価な処理を行うことを利用し、雑音除去フィルタを構成した。

また、中間層ユニット数が 1 個の SNN を多段接続することで KL 変換-KL 逆変換では不可能なオンライン処理が可能となった。そこで多段接続 SNN を用いて適応的な雑音除去と音量補償を行うアルゴリズムを提案した。

結果、音声の大きさ、雑音の大きさが経時的に変化する場合において適応的な雑音除去と音量補償が可能なフィルタが構成できた。雑音除去実験によって適応的な雑音除去と音量補償が適切に行われていることが示された。

## 文 献

- [1] G.W.Cottrell, P.Munro and D.Zipser, "Image compression by back propagation: An example of extensional programming," *Advances in Cognitive Science*, 3, pp.208-240, Ed.Sharkey N.E., Norwood, NJ, Ablex, 1988
- [2] 渡辺一央, 大堀隆文, 下沢楯夫, "KL 変換用単位パーセプトロンの収束性に関する理論的考察," *信学論(D-II)*, vol.J75-D-II, no.11, pp.1933-1939, Nov.1992
- [3] 渡辺一央, 伊東英彦, 増田一, 大堀隆文, "KL 変換用多段パーセプトロン," *信学論(D-II)*, vol.J75-D-II, no.11, pp.1925-1932, Nov.1992
- [4] 吉村宏紀, 清水忠昭, 井須尚紀, 菅田一博, "砂時計型ニューラルネットワークを用いた雑音除去フィルタの構成," *電学論(C)*, vol.117-C, no.10, pp.1498-1505, Oct.1997
- [5] 吉村宏紀, 清水忠昭, 井須尚紀, 菅田一博, "多段接続砂時計型ニューラルネットワーク雑音除去フィルタを用いた適応的雑音除去," *電学論(C)*, vol.120-C, no.4, pp.507-515, Apr.2000
- [6] 古本友紀, 吉村宏紀, 並木寿枝, 清水忠昭, 井須尚紀, 菅田一博, "砂時計型ニューラルネットワークを用いた雑音強度依存型フィルタの研究," 2000 情報処理学会第 61 回全国大会講演論文集, 分冊 2, no.3N-1, pp.40-41, Oct.2000
- [7] 並木寿枝, 古本友紀, 吉村宏紀, 清水忠昭, 井須尚紀, 菅田一博, "カスケード接続砂時計型ニューラルネットワークによる補聴器用雑音除去フィルタの提案," 2002 信学ソサイエティ大会講演論文集, 分冊 2, no.A4-21, pp.83, Sept.2002