

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

古本友紀, 吉村宏紀, 並木寿枝, 清水忠昭, 井須尚紀, 菅田一博
鳥取大学工学部知能情報工学科

1. はじめに

入・出力層のユニット数を同数とし, 中間層のユニット数をそれらよりも少なくした構造の砂時計型ニューラルネットワーク (SNN) は, 教師信号を入力信号と等しくして学習を行った場合, KL変換と等価な処理を行う¹⁾. この性質を音声信号に対する雑音除去フィルタに利用して, SNNを多段に接続した構成の多段接続SNN雑音除去フィルタを提案した. 多段接続SNN雑音除去フィルタでは, 接続段数が少ないと雑音除去率は高いが音声は歪む. 逆に, 接続段数が多いと雑音除去率は低いが音声の歪みは少ない. このため, 音声信号に含まれる雑音の強度やその変動により適切な段数が異なる. 本研究では, 雑音強度の関数として接続段数を適応的に変化させ, 音声信号に重畳した雑音が大きく変化しないように雑音を除去する手法を提案した. 本手法を用いることにより雑音を除去した音声, 雑音強度やその変動に依らず, 聞き取り易くなることを聴覚実験により明らかにした.

2. 多段接続型SNN雑音除去フィルタ

図1に提案した雑音除去フィルタの構成図を示す. 各段のSNNは全て, 入・出力層のユニット数は N 個, 中間層のユニット数は 1 個とした. 第1段SNNには, 雑音の重畳した音声信号を入力した. 第 i 段SNN ($i=2,3,\dots,P$) には, 第 $i-1$ 段SNNの入力信号から出力信号を引いた信号 (第 $i-1$ 段SNNの誤差信号) を入力した. 各段SNNの教師信号には, 各段SNNの入力信号と同じ信号を用いた. 多段接続型SNN雑音除去フィルタの出力信号は, 各段SNNの出力信号の和で与えた.

Noise Power-dependent Adaptive Filter for Speech Noise Reduction Composed of Cascaded Sandglass-type Neural Networks

Yuki Furumoto, Hiroki Yoshimura, Toshie Namiki, Tadaaki Shimizu, Naoki Isu and Kazuhiro Sugata
Dept. of Information and Knowledge Engineering,
Tottori Univ., 4-101 Koyama-minami, Tottori 680,
JAPAN

3. 雑音強度に依存した接続段数の決定

多段接続型SNN雑音除去フィルタの各段のSNNの入・出力層のユニット数を 20 個とし, 接続段数の上限値を 20 段とした. 最適な雑音除去を行うには, 音声信号のSN比および音声信号に含まれる雑音成分のパワーに応じて接続段数を可変にする必要がある. 本研究では, 雑音成分と信号成分のパワーは, 次式により推定した.

$$S(t) = \frac{1}{2\alpha - 1} (F(t) - L(t)) \quad (1)$$

$$N(t) = F(t) + L(t) - S(t) \quad (2)$$

$$\alpha = 0.97425$$

ここで, $S(t)$ は音声信号のパワー, $N(t)$ は雑音のパワー, $F(t)$ は第1段~第10段の出力のパワー, そして $L(t)$ は第11段~第20段の出力のパワーを表わしている. α は音声信号の全成分のうち第1段~第10段に現れるパワーの比の平均値である.

接続段数を固定した雑音除去フィルタについては, 入力信号のSN比の違いにより最も聞き取りやすい接続段数が調べられている²⁾. この結果を採用し, 平均SN比が $-5, 0, 5$ [dB]のときそれぞれ 7 段, 8 段, 16 段となるようにパラメータを定め, 推定した雑音成分のパワーの関数として次式により, 接続段数 $P(t)$ を決定した.

$$P(t) = 7 + \left\lfloor \frac{13}{1 + \exp(A(t))} + 0.5 \right\rfloor \quad (3)$$

$$A(t) = 10 \log_{10} \left(\frac{s}{N(t)} \right) + b \quad (4)$$

$$a = -0.5346$$

$$b = 2.037$$

$$s = 0.43842691$$

本研究では, 音声の過度の歪みを避けるため, 接続段数の下限値を 7 段とした.

4. 聴覚実験

本手法の有効性を示すため、聴覚実験を行った。聴覚実験に用いた音声は、日本音響学会発行の研究用連続音声データベースより 45 文の連続音声を選択し、音の大きさを揃えるために正規化を行った後、聴覚実験を行って等しい音量に聞こえる音声 26 文を選択した。連続音声のサンプリング周波数は 16 kHz、量子化ビット数は 16 bit とした。

上記の音声信号に、周期的にパワーが変動するようなガウス性白色雑音を付加した。音声信号に付加するガウス性白色雑音のパラメータは以下のものを用いた。

平均SN比：-10, -5, 0, 5, 10, 15 [dB]

パワー変動の振幅：0, 2.5, 5, 7.5, 10 [dB]

パワー変動の周期：0.25, 1, 4 [s]

聴覚実験は、接続段数可変の雑音強度依存型フィルタで処理した音声、接続段数を 8 段で固定した固定長フィルタで処理した音声、およびフィルタリング処理を施さなかった音声を被験者に呈示し、205 対の音声について「雑音の大きさと音声の品質を総合的に考慮して、どちらの音声がよりはっきりと聞き取ることができ、違和感や不快感が少ないか」を、どちらか 1 文を選択させることにより評価を行った（一対比較法）。被験者は 20 歳前後の男 6 名、女 4 名の計 10 名とし、各被験者に呈示する 205 対の音声の呈示回数は、呈示順序を入れ替えたものを含めて各対計 4 回とした。ただし、4 回の文章の発話内容は全て異なり、26 文の内ランダムに選択した 4 文を用いた。音声の呈示はヘッドホンを通して行った。一対比較法により得られた評価結果を基に、Thurstone の比較判断則（Case V）に従って最小ノーマット χ^2 法を適用し、聞き取りやすさを距離尺度で表した。段数可変多段接続SNNで処理した音声は、SN比および振幅の大きさによらず最良の結果が得られた（図 2, 3）。

4. おわりに

本研究では、雑音強度の関数として接続段数を換え、音声信号に重畳した雑音が大きく変化しないように雑音を除去する手法を提案した。この手法によって効果的に雑音除去が行われ、音声の聞き取りやすさを向上させるのに有効であることが示された。

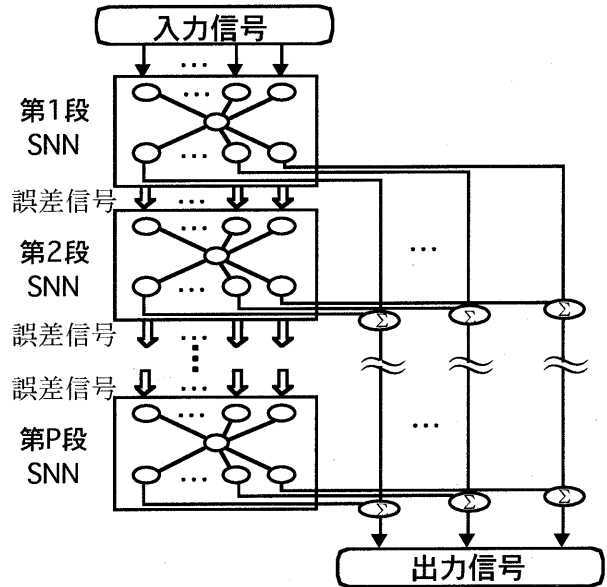


図 1. 多段接続型SNN雑音除去フィルタの構成図

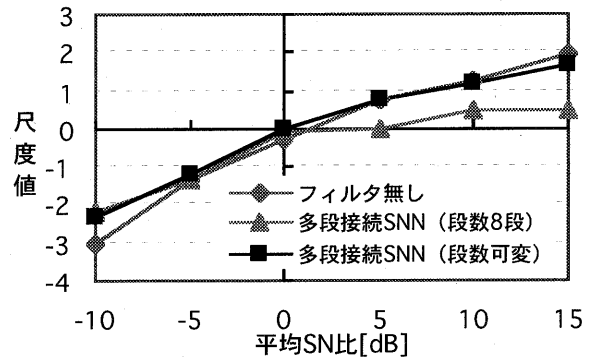


図 2. 平均SN比と聞き取り易さの比較

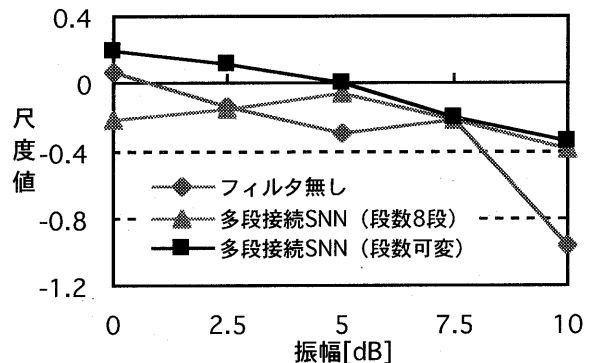


図 3. 雑音の振幅と聞き取り易さの比較

参考文献

- 1) P. Baldi and K. Hornik: "Neural networks and principal component analysis: Learning from examples without local minima," Neural Networks, 2, pp.53 - 58, February 1989.
- 2) 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", 電気学会論文誌C, Vol. 120-C, No. 4, pp.507-515, April 2000