

多段接続砂時計型ニューラルネットワークを用いた 音量補償雑音除去フィルタの構成

1 P-5

小野隆広 古本友紀 並木寿枝 吉村宏紀 清水忠明 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

デジタル補聴器の開発を目的として、砂時計型ニューラルネットワーク (SNN) を用いた雑音除去フィルタを提案する。中間層ユニット数が 1 個の 3 層 SNN を多段に接続した構造を持つ多段接続 SNN 雑音除去フィルタ (CSNNRF) は、教師信号を入力信号と等しくして学習を行うことで KL 変換と等価な処理を行う [1]。この性質を音声信号の雑音除去に利用し、SNN を多段に接続した構造の雑音除去フィルタを構成した。CSNNRF は、入力信号による学習と接続段数の変化によってフィルタ特性を変化させることが可能であり、信号の濾波と雑音低減効果は接続段数の違いによって異なる。これまでの研究では、音声の大きさを一定とした条件の下でパワーが周期的に変化する白色雑音を付加した信号を対象に、雑音の大きさの関数として接続段数を変えることで適応的な雑音除去を行い、音声の聞き取りやすさを向上させる結果を得てきた [2]。

本研究では、音量が変化したり断続的な無音区間を含むような音声信号に対して、適応的な音量の補償および雑音除去を行うフィルタ処理によって、より聞き取りやすい音声を得る手法を提案する。

2. フィルタの構成

本研究では、音量が時間的に変化する音声信号を対象とし、パワーが周期的に変化する白色雑音を付加して CSNNRF への入力信号とした。多段接続 SNN の各段の中間層出力を用いて、入力信号に含まれる音声および雑音成分のパワーを逐次推定し、音声成分の音量が一定となるように音量の補償を行った。音量の補償後、雑音強度の関数として多段接続 SNN の接続段数を決定し、適応的な雑音除去を施した。

2.1 多段接続 SNN 雑音除去フィルタ

図 1 に CSNNRF の構成を示す。CSNNRF は入出力層のユニット数が 20 個、中間層のユニット数が 1 個の 3 層 SNN を 20 段直列に接続した構造を持つ。第 1 段 SNN に音声信号系列を 20 点ずつに分割して入力し、第 2 段以降の SNN には前段の出力誤差を入力した。各段の教師信号は各段の入力信号と等しい信号とし、SNN の学習には RLS アルゴリズムを用いた。学習誤差最小時には、各段の出力から入力信号を主成分に分解した信号が得られる。第 1~ P ($1 \leq P \leq 20$) 段の出力信号の総和をフィルタの出力信号とした。

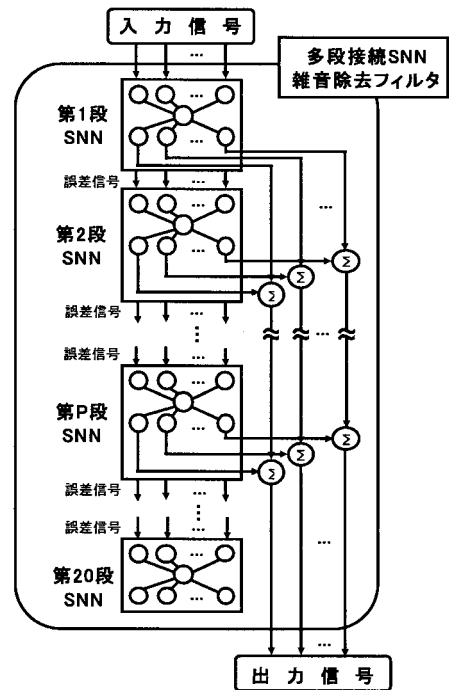


図 1. 多段接続 SNN 雑音除去フィルタの構成

2.2 信号パワー、雑音パワーの推定

通常発話される音声信号では、有声音が占める時間的割合が大きい。有声音を主成分分析すると、各成分の寄与率に大きな差が見られる。一方、白色雑音は全ての主成分のパワーが等しい。この主

Construction of noise reduction filter for speech with volume compensation by use of cascaded sandglass type neural networks

Takahiro Ono, Yuki Furumoto, Toshie Namiki, Hiroki Yoshimura, Tadaaki Shimizu, Naoki Isu and Kazuhiro Sugata

Dept. of Information and Knowledge Engineering, Tottori Univ., 4-101 Koyama-minami, Tottori 680, JAPAN

成分間の寄与率の相違を利用して音声信号および雑音のパワーを推定した。CSNNRF の第 i 段 ($i=1,2,\dots,20$) の中間層出力の 2 乗の平均値は、入力信号の第 i 主成分のパワーに等しい。まず、雑音を含まない 5 母音を CSNNRF に入力して各段の中間層出力の 2 乗平均値を求め、5 母音それぞれの第 i 主成分の寄与率 c_{ij} ($j=1,2,\dots,5$) の標準モデルを構成した。音声信号を CSNNRF に入力した時に得られる第 i 段中間層出力のパワー w_i を最小 2 乗法により $w_i = Sc_{ij} + N$ に回帰した。誤差自乗和が最小となる母音 j により得られた S , N の値を音声および雑音成分のパワーと推定した。

2.3 音量の計算と信号の増幅

音声のパワーの ($\log 2$) 乗を感覚的な音の大きさ (以下、音量と呼ぶ) と定め、推定した音声成分のパワー S を音量に変換した。音声の聞こえの大きさはアクセント部分の音量によるものと考え、0.5 秒毎に音量の最大値を求めることでアクセント部分の音量に代えた。音声信号には無音区間がしばしば含まれるため、0.5 秒毎の区間最大音量の系列の移動平均と標準偏差を求め、平均値から標準偏差の 1/2 を引いた値を閾値として、閾値を越えた区間最大音量のみを取り出した。遮断周波数 0.25 Hz の 2 次のバターワースローパスフィルタを通過させて移動平均を求め、聞こえの大きさの推定値とした。聞こえの大きさの逆数を遮断周波数 1Hz の 2 次のバターワースローパスフィルタに通し、その $1/(2\log 2)$ 乗を増幅率 α として CSNNRF の出力を増幅した。

2.4 接続段数の決定とフィルタ出力

雑音強度が変化した際に最良の聞き取りやすさを与えるように、音量の補償後の雑音成分のパワー $\alpha^2 N$ の関数として、CSNNRF の接続段数 P を次式により決定した [2]。ただし、 P は四捨五入により整数化した。

$$P = 7 + 13 / (1 + 7.668 \cdot (\alpha^2 N)^{2.322})$$

第 1~ P 段までの対応する出力ユニット毎に出力の総和を求め、フィルタの出力信号とした。

3. 音声信号のフィルタリング

日本音響学会編集の「研究用連続音声データベース」より選択した文章音声標本 (サンプリング周波数 16kHz、量子化ビット数 16 ビット、データ長 3.7~10 秒) に対して聞こえの大きさが等し

くなるように音量調整を施し、聴覚実験で等しい大きさに聞こえた音声標本を選択してそれらの平均音量を基準の音量と定めた。

音量が基準の 5.0 倍の音声と 0.4 倍の音声、約 7 秒間に音量が基準の 0.4 倍から 5.0 倍に緩やかに変化する音声を作成し、それらを約 5 秒間の無音区間を挟んで連結した (図 2. (a))。この原音声信号に 3 秒周期でパワーが変化する白色雑音を付加して入力信号とし (図 2. (b))、CSNNRF によるフィルタリング処理を行った。図 2. (c) にフィルタの出力信号を示す。

CSNNRF による音量補償によって各音声は基準の大きさになり、聞き取りやすくなった。また、雑音成分が低減された。無音区間では増幅率を低下させることで雑音が大幅に低減された。

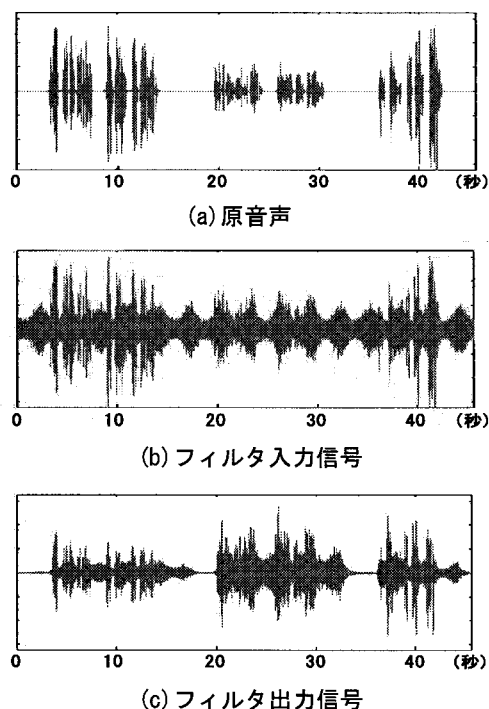


図 2. 音量補償雑音除去フィルタによる音声処理

【参考文献】

- [1] 渡辺一央, 伊東英彦, 増田一, 大堀隆文, "KL 変換用多段接続パーセプトロン", 信学論 (D-II), J75-D-II, pp. 1925-1932 (1992)
- [2] 古本友紀, 吉村宏紀, 清水忠昭, 井須尚紀, 菅田一博, "秒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", 情報処理学会第 61 回全国大会, pp2-41 (2000)