

5 層非線形砂時計型ニューラルネットワークを用いた 日本語連続母音の分析

4Q-3

國光英史 木本雅也 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

高品質な音声合成を行うために、連続音声の渡りの部分にどのような特徴があり、情報が含まれているかを調べ、自然な音韻間の推移を再現できる手法を確立することが重要である。日本語母音の音響的特徴を表現するために LSP パラメータがしばしば用いられる。本研究では、5 層非線形砂時計型ニューラルネットワーク（以下 5 層 SNN）に LSP パラメータを学習させることで音声情報の圧縮を行い、圧縮したパラメータ空間上で表現される連続母音間の渡り部分の特徴を調べた。

2. 5 層砂時計型ニューラルネットワーク

階層型ニューラルネットワークにおいて、入力層および出力層のユニット数を同数とし、中間層のユニット数を入出力層より少なくしたものものを砂時計型ニューラルネットワークと呼ぶ。5 層 SNN では、入力層から第 3 層までを圧縮層、第 3 層から出力層までを復元層と呼ぶ。

本研究では、単位 SNN と呼ばれる 5 層 SNN を 2 段カスケード接続した図 2 に示す構成のニューラルネットワークを用いた。単位 SNN の構造は、14 次の LSP パラメータを入出力信号として用いるため、第 1 層および第 5 層のユニット数を 14 個、第 2 層および第 4 層のユニット数を 8 個、第 3 層のユニット数を 1 個とする。また、5 層 SNN が非線形変換を行えるように第 2, 4 層の応答関数にはシグモイド関数を用いた。入力信号と教師信号を等しくして学習を行った場合、学習誤差最小時に各段の中間層出力から入力信号を非線形主成分に分解した信号が得られる。2 段目の入力信号には、1 段目の入力信号から出力信号を引いた残差信号を用いる。

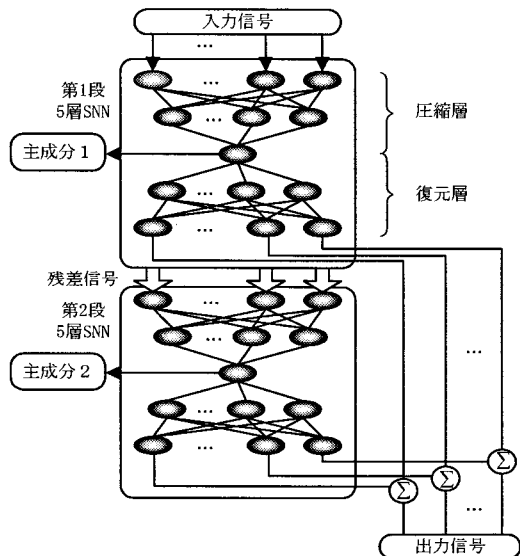


図 1. 2 段接続 5 層 SNN

3. 実験方法

本研究では、単母音 LSP パラメータの学習、単母音学習の終了した 5 層 SNN に、入力信号として 3 連続母音を与えたときに得られる 5 層 SNN の中間層出力を用いた情報の圧縮を行った。2 段 SNN から得られる 2 つの主成分の表す 2 次元平面での連続母音の主成分の軌跡を調べた。連続母音では、母音間の渡り部の調音結合による特徴は、各母音の特徴分布間に現れるので、その軌跡から渡りの部分の主成分を抽出し、その特徴を調べた。

音声試料として男性話者 1 名の単音で発声した 5 母音と、3 連続発声の母音を採取した。その際、単音 5 母音を 5 回、隣り合う母音が同一母音とならない 3 連続母音の組合せ 80 種類で各連続母音を 5 回ずつ採取した。採取した母音の原音声をサンプリング周波数 11.025 kHz、量子化ビット 16bit で A/D 変換し、14 次 LSP 分析を行った。そうしたものを 5 層非線形 SNN の入力信号として用いた。学習に用いたのは、単独発声の 5 母音である。

Analysis of the Glide in Japanese Vowel Series by use of Five-layer Non-linear Sandglass-type Neural Network

Eiji Kunimitsu, Masaya Kimoto, Tadaaki Simizu,
Naoki Isu, Kazuhiro Sugata
Department of Information and Knowledge Engineering
Tottori University

4. 実験結果

本研究では 14 次の LSP パラメータを用いているので、第 3 層が 1 ユニットである 5 層 SNN を 14 段に接続させたときに得られる 14 個の主成分によって元の LSP パラメータをほぼ完全に再現することができる。図 2 に単母音学習後の 5 層 SNN に、学習に用いた単母音のデータを入力した場合の入出力間の平均二乗誤差を LSP パラメータの分散で正規化した値を示す。その結果、2 段目までの正規化二乗誤差の値が 1.0 近くになっており、二乗誤差と分散の値がほぼ等しいことが分かる。このことから 2 段接続 SNN の出力で元の LSP パラメータの特徴を表現するのに十分であるといえる。

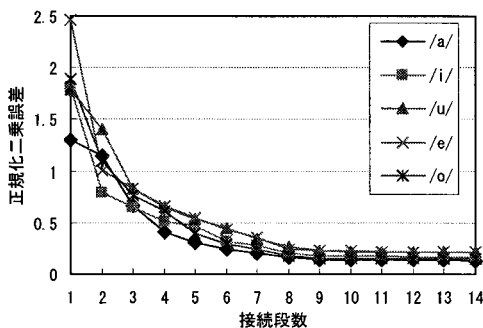


図 2. 接続段数ごとの正規化二乗誤差

単母音学習後の 5 層 SNN に入力信号として単母音データを与えたときに得られる 2 つの中間層出力を特徴平面上に表したものを図 3 に示す。

入力信号として 3 連続母音データを与えた場合の中間層出力から、連続母音の主成分の軌跡を特徴平面上に表したものを図 4 に示す。同図から連続母音間の渡り部における主成分の軌跡部分だけを取り出し、連続する 2 つの母音の種類ごとに渡り部の回帰を行ったものを図 5 に示す。2 つの音韻の特徴の分布の間に他の音韻の分布がある場合には、間にある音韻の近くを通過して渡っている。「あお」「おあ」については音韻が接近していたので今回は渡りの部分を抽出できなかったが、他の場合については渡り部分を抽出し、母音間の渡り部を推定することができた。

5. おわりに

本研究では、孤立発声 5 母音の LSP パラメータを学習させた 5 層 SNN に連続母音の LSP パラメータを入力し、連続母音の主成分の軌跡によ

り渡り部を分析する手法を提案した。連続母音間の渡り部の主成分の軌跡を抽出し、回帰分析により音韻の推移を分析することができた。本手法により、渡り部の再現が容易にできるようになり、音声合成への利用可能性を示すことができた。

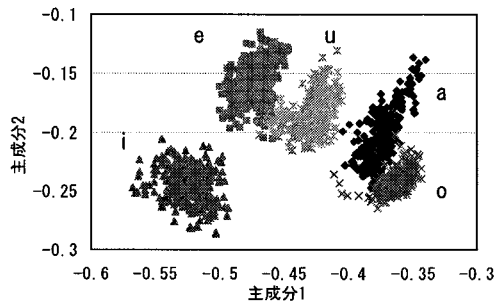


図 3. 単音 5 母音の特徴平面図

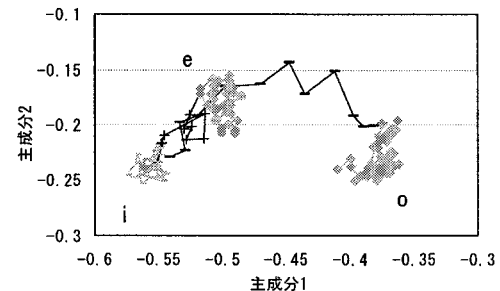


図 4. 連続母音“えいお”の特徴平面図

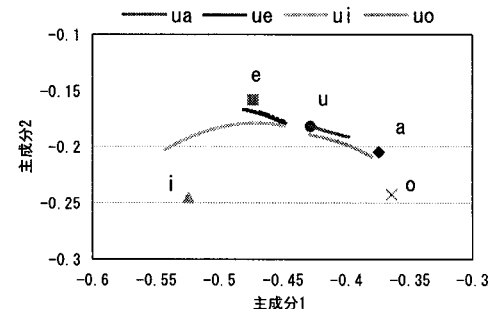


図 5. “う”から他の音韻への渡り部の回帰曲線

参考文献

- [1] 佛崎建, 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で処理した音声の聞き取り易さ, 電子情報通信学会基礎・境界ソサイエティ大会講演論文集, pp. 89, (1998)
- [2] 吉村宏紀, 清水忠昭, 井須尚紀, 菅田一博; 多段接続砂時計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, Vol. 120-C, 507-515, (2000)