5G－7

規則音声合成のためのニューラルネットワークによるわたり部の学習に関する研究

清水忠昭，白石雄司，菅田一博，井須尚紀
鳥取大学工学部

1.はじめに

音声の規則合成方式では、単音節や音韻連鎖といった音声の基本単位を記憶しておき、それを接続して任意の音声を合成する。単音節を規則合成の基本単位とした場合、合成に必要なデータ数は少なくて済むが、音節どうしを接続する際に調音結合が考慮されず、合成音声の品質は低くなる。音韻連鎖を規則合成の基本単位とした場合、調音結合を考慮した合成音声を合成できるが、音韻の組み合わせにより、合成に必要なデータ数が増え、データの記憶容量や音声合成時の音韻連鎖の検索の点で処理が複雑になる。

我々の研究室では、単音節を基本単位とする音声の規則合成において、調音結合を考慮にいれて音節どうしの接続を行うことにより、合成音声の品質を向上させることを目的とした研究を進めている。本稿では、単音節として日本語の5母音のみを用い、音節どうしの接続にニューラルネットワークを適用した結果について報告する。

2.ニューラルネットワークによるわたり部の学習

音節どうしの接続は、調音結合の影響を受けて非常に複雑であり、単純な補間では再現できない。音節どうしの接続のしかたをデータベースとして保持する方法は、音韻連鎖を規則合成の基本単位とする方法と同じく、記憶すべきデータ量が非常に多なる。音節どうしの接続に一定の規則を設ける方法は、音節どうしの接続規則を簡単に導く方法がないため困難である。本研究では、音節どうしの接続のしかたを音声パラメータが描くパターンとみなし、ニューラルネットワークのパターン学習能力を用いて学習させる手法を考案した。

ニューラルネットワークのパターン学習能力を生かすために、音声の調音特性の時間変化が明確なパターンとして現われる音声パラメータを用いることが望ましい。また、音声の規則合成に用いるため、音声合成を目的としたパラメータを用いる必要がある。これらを考慮して、本研究では音声パラメータとしてLSPパラメータ[1]を用いることとした。

実際の音声における調音結合の影響は、連続する音節の間で複雑に作用している。しかし、本研究では、まず隣接する2音節間の接続のみについて実験を行った。

連続して発声された2母音のデータを用いて階層型ニューラルネットワークの学習を行う。わたり部の前後の母音のLSPパラメータをニューラルネットワークの入力とし、教師データはわたり部のLSPパラメータの時系列パタンとする。実験には日本語5母音のみを用いたため、その組み合わせにより、ニューラルネットワークに学習させるパタン数は、20パタンとなる。学習が完了したニューラルネットワークは、入力として接続したい2つの母音のLSPパラメータを与えると、出力に2母音を接続するわたり部のLSPパラメータの時系列を出力する。本研究で構成したニューラルネットワークを図1に示す。

本研究ではLSP分析を12次で行ったため、ニューラルネットワークの入力層は12×2=24個のユニットを持つ。また、資料ごとに長さの異なるわたり部をLSP分析の20フレーム(40m秒)に正規化したため、

A Study of Phoneme Conjunction for Speech Synthesis from Text Based on Neural Network Approach
Tadaaki Shimizu, Yuuji Shiraishi, Kazuhiro Sugata, Naoki Isu
Department of Information and Knowledge, Faculty of Engineering, Tottori University
4-101 Koyama-cho minami, Tottori,Tottori 680,Japan

ニューラルネットワークの出力層は12×20=240個のユニットを持つ。中間層の層数やユニット数、学習に用いる重み更新パラメータ等は、経験的に定める以外に方法がないため、以下の範囲でニューラルネットワークの仕様を変更しながら学習を繰り返した。[2]

重み更新パラメータ η ：$1.0\times10^{-4}\leqq\eta\leqq0.6$

重み更新パラメータ α ：$1.5\times10^{-4}\leqq\alpha\leqq0.8$

中間層の層数　　：1層〜3層

中間層ユニット数　　：12個〜240個

3.実験結果

本研究で得た最良のニューラルネットワークの仕様を以下に示す。

中間層：1層

ユニット数：入力24個，中間39個，出力240個

重み更新パラメータ η ：1.0×10^{-2}

重み更新パラメータ α ：1.5×10^{-2}

学習回数　　：2.0×10^{5}回

出力自乗誤差　　：1.55×10^{-4}

このニューラルネットワークによって得られるわたり部のLSPパラメータの時間推移パタンは、音声合成に十分適用できる。現在、研究は途上であり、以後、聴覚実験によって合成音声の品質評価を行わなければならない。

4.今後の課題

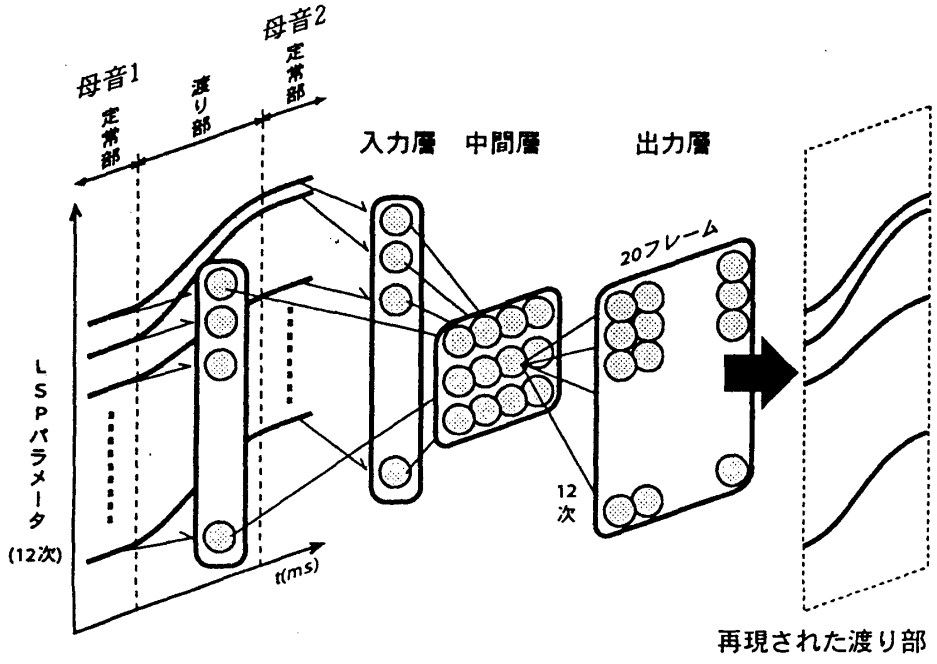
本研究では、2母音の接続について、そのわたり部を学習させたニューラルネットワークにより、母音どうしの接続を行う手法を提案した。しかし、この方法は、調音結合における音韻の調音特性の変動については考慮していない。また、母音のみを用いて実験をおこなっているため、そのままでは規則合成に適用できない。

以上の問題点のうち、前者については、すでに本手法を改良した手法が考案されており、実験準備を進行中である。また、後者については、改良手法の完成を待って子音を含んだ実験を行う予定である.

参考文献

[1]菅村昇，板倉文忠：線スペクトル対音声分析合成方式による音声情報圧縮，pp599〜606，電子通信学会論文誌，1981

[2]中野馨：ニューロコンピュータの基礎，コロナ社，1990



わたり部の学習に用いたニューラルネットワークの構成