

氏名	きもとまさや 木本雅也
学位の種類	博士(工学)
学位記番号	甲第168号
学位授与年月日	平成17年 3月18日
学位授与の要件	学位規則第4条第1項該当
学位論文題目	ニューラルネットワークを用いた小規模応用向け高品質 規則音声合成システムの研究
学位論文審査委員	(主査) 田中美栄子 (副査) 中川聖一 伊藤良生 池原悟 菅原一孔

学位論文の内容の要旨

本研究は、小容量・低コストで済む小規模な音声合成システムの構築を目指した研究開発の中で、特に以下の2点について検討したものである。一つ目は、VCV素片接続型の規則音声合成方式に於ける音韻環境の最低長の同定、2つ目は、フォルマント周波数と線スペクトル対(LSP)パラメータの利点を併せ持つ新たな合成パラメータの提案である。

第1のテーマでは、VCV素片接続型の規則音声合成法について、音韻環境を指標とした素片選択において考慮すべき音韻環境の長さを実験的に調べた。

規則音声合成において、音声合成を困難にする原因のひとつは、同じ音韻として表記される音素であっても、その前後に発声された音楽の影響を受けて性質が変化してしまう事であり、この現象を調音結合と呼んでいる。調音結合の影響を考慮するために、VCV素片の発声時に、その後でどのような音韻が発声されていたかを表す音韻環境情報を用いるが、VCV素片の前後いくつかの音韻までを考慮すれば十分であるかは判っていなかった。特に小規模な音声合成システムでは、音韻環境情報を最低限に抑える必要があった。本研究では、VCV素片選択実験を行い、音韻環境として先行2音韻、後続1音韻を考慮すれば良好な選択が行われることを示した。

第2のテーマでは、砂時計型ニューラルネットワーク(SNN)の情報圧縮能力を用いてフォルマント周波数とLSPパラメータの利点を併せ持つパラメータを抽出する手法を提案した。LSPパラメータの利点は、計算機で容易に計算できることであるが、実用的な合成音声品質を得る為には10次以上の高次のパラメータを必要とする為、制御規則の抽出には適さないという難点がある。一方、フォルマント周波数の利点は、音韻との対応関係が良いことであるが、話者個人毎の変動が複雑で、計算機での自動推定が困難であるという欠点がある。

提案手法では、日本語5母音のLSPパラメータをSNNに学習させ、SNNの中間層から圧縮したLSPパラメータ(主成分)を抽出する。この圧縮パラメータがフォルマントの様な性質を持ち、

且つ十分な精度で復元できれば音声合成に応用できる。また、圧縮したパラメータであるので、小規模な音声合成システムに適している。近年、計算機の飛躍的な処理速度向上と大容量化に伴って、開発される音声合成システムは大規模化の一途を辿ってきているが、現在でもカーナビゲーションシステムや PDA 等への組み込みに小容量・低コストで済む小規模な音声合成システムが必要とされている。

本研究の目的を達成する為に、1) 本手法での適切な SNN の構成、2) SNN を用いて次数圧縮した LSP パラメータの振舞い、3) 次数圧縮された LSP パラメータを復元して合成した音声の品質、の3点を実験的に調べた。その結果、中間層ユニットを1個とした SNN を2段にカスケード接続した SNN を用いて次数圧縮した LSP パラメータは、フォルマントに類似した分布を示すことがわかった。また、次数圧縮した LSP パラメータを復元して合成した音声に対して聴覚試験を行い、正しく聞き取れることを示した。以上より、本方式でフォルマント周波数と同様に音韻との対応関係が良く、LSP パラメータの様に音声合成に適したパラメータが得られることを示した。

本研究の成果として、VCV 素片接続型の規則音声合成において考慮すべき音韻環境長が、先行2音韻、後続1音韻と短いことを示し、2次元で音韻と対応付けの良いパラメータを得る手法を提案して、小規模向けの音声合成に有効な知見を得ることができた。

論文審査の結果の要旨

本研究は、小規模な音声合成システムの構築を目指して、VCV(母音-子音-母音)素片接続型の規則音声合成方式に於ける音韻環境の最低長の同定と、フォルマント周波数と線スペクトル対(LSP)パラメータの利点を併せ持つ新たな合成パラメータの提案を行ったものである。

前者では必要な音韻環境が先行2音韻、後続1音韻で済むことを明らかにした。VCV素片選択において、比較的短い音韻環境を考慮するだけで良いという結果を得たことは、特に小規模な音声合成システムに対して有用な成果である。

後者では、多くの素片を必要とする素片結合に依らない合成法を模索するために、日本語5母音を分離生成できるパラメータを得る手法を提案した。その手法では、日本語5母音のLSPパラメータを砂時計型ニューラルネットワーク(SNN)に学習させ、SNNの中間層から圧縮したLSPパラメータ(主成分)を抽出する。中間層ユニットを1個としたSNNを2段にカスケード接続したSNNを用いて次数圧縮したLSPパラメータが、フォルマントに類似した分布を示すことを明らかにした。更に、次数圧縮したLSPパラメータを復元して合成した音声に対して聴覚試験を行い、正しく聞き取れることを示した。これにより、素片を用いずルールのみによって音声合成系を作るための基礎となるパラメータを得ることができた。

これらは小規模システムによる音声合成に対して有用な知見を提供するものである。よって本論文は、博士(工学)の学位を授与するに十分な価値を有すると認められる。