

## 審査の結果の要旨

氏名 汪淼森

本論文は「Advanced Control of Prosody in HMM-based Mandarin Speech Synthesis (HMM中国語音声合成における韻律制御の高度化)」と題し、標準中国語HMM (Hidden Markov Model) 音声合成において、基本周波数パターンモデルによる無声区間の補間、声調核モデルによる声調の安定した表現などにより、韻律的特徴の学習に改善を施すことにより、従来と比較した音質の向上を達成したもので、全8章からなる。

第1章は「Introduction」であって、テキスト音声合成について、まず、歴史的経緯を含めて概説した後、特に中国語で重要な課題となっている韻律の取り扱いに触れ、次に、本論文が中国語HMM音声合成の韻律的特徴の改良を目指したものであるとしている。最後に、本論文の章構成が示されている。

第2章は「Communication Process」と題し、まず、人間における、テキスト入力-音声出力過程を概観した後、テキスト音声合成で必要となるテキスト解析から始まる種々の処理過程について簡単に言及している。次に、中国語音声の特徴について声調の観点から述べ、中国語を中心とした韻律的特徴に関する研究のいくつか紹介した上で、音声合成において着目すべき点を整理している。

第3章は「Hidden Markov Model based Speech Synthesis」と題し、本論文の基本となるHMM音声合成を概説している。まず、音声認識で一般的であるHMMの構成と、その学習、適応と認識への適用について説明し、音声合成に利用した時の音素モデルの学習の特徴について、特に決定木による状態クラスタリングについて詳細に述べている。

第4章は「Duration Modeling and Improvement」と題し、まず、HMM音声合成における、音素長、休止長の制御について述べ、決定木に基づく状態クラスタリングで注目する項目について整理している。次に、統語情報の重要性を指摘し、音素長、休止長の制御で用いる統語情報の具体的項目を提案している。その上で、連続量の取り扱いに適した決定木モデルを採用することで、音素長、休止長の推定精度が向上したとしている。

第5章は「F0 Modeling and Generation in HMM-based TTS」と題し、HMM音声合成での基本周波数発話の制御で、一般的に用いられているMulti-Space probability Distribution (MSD) HMMを紹介し、その問題点が有声・無声の切り替えにあるとしている。MSD-HMMでは、基本周波数を、有声区間では数値、無声区間では基本周期なしとして、切り替えて表現するが、学習音声における有声・無声の誤りが多く発生し、結果として、音質の低下が問題となっている。この様なMSD-HMMの問題点を整理し、次章の提案方式の導入としている。

第6章は、「Continuous F0 Modeling and Improvement」と題し、前章で提示した有声・無声の誤りに起因する問題を解消するための2つの手法を提案している。1つ目は、音素の情報を利用するもので、音素が有声音であるか無声音であるかに従い、有声・無声の切り替えに制約を加え

るものである。2つ目は、基本周波数パターン生成過程のモデルに基づく補間により、無声区間にも基本周波数を与えるもので、連続的な基本周波数パターンでモデル化する。合成に際し、有声・無声の切り替えが必要となるが、それは有声音・無声音の区別に従う。基本周波数の推定精度向上のみならず、音素長の推定精度向上も達成している。合成音声の聴取により、MSD-HMMを用いた音声合成よりも合成音声の品質向上が得られたとしている。

第7章は、「Tone Nucleus Model and its Application to HMM-based TTS」と題し、基本周波数パターンを声調核のみに着目して生成し、声調核以外の部分については、遷移区間と考え、補間する手法を提案している。声調核モデルは、音節の基本周波数パターンを、遷移部分+声調核+遷移部分として捉えるもので、声調核に着目することで、声調型のラベリングの精度を上げ、より安定した基本周波数の表現が可能となる。声調核は有声区間であるため、有声・無声の判定誤りに起因する問題も解消できる。合成音声の聴取により、音節全体の基本周波数パターンを対象とする従来手法よりも合成音声の品質向上が得られたとしている。

第8章は、「Conclusion and Future Work」と題し、本論文での成果の概要を述べ、今後の発展について議論している。

以上を要するに、本論文は、HMM音声合成において、有声・無声の誤りに起因する音質劣化を解消する手法として、音素の有声・無声の情報、基本周波数パターンの生成過程モデルによる補間、声調核モデルによる基本周波数パターンの安定した表現、に基づくものを開発したものであって、それによって標準中国語HMM音声合成の韻律的特徴の改善を達成している。音声合成の発展に大きく寄与したものであり、情報通信に貢献するところが少なくない。

よって本論文は博士（工学）の学位請求論文として合格と認められる。