

(様式第5-3) 論文博士

平成 22 年 8 月 10 日

琉球大学大学院
理工学研究科長 殿

論文審査員

主査 高良富夫
副査 宮城隼夫
副査 M. R. Asharif



学位（博士）論文審査及び学力確認終了報告書

学位（博士）の申請に対し、学位論文の審査及び学力確認を終了したので、下記のとおり報告します。

記

申請者	氏名 DO TU TRONG	
現住所		
成績評価	学位論文 <u>合格</u> 不合格	学力確認 <u>合格</u> 不合格
論文題目	Vietnamese Text-To-Speech System and Perceptual Features of Glottal Tones ベトナム語のテキスト音声合成システムと声門声調の知覚特性	
審査要旨（2000字以内）		

(次頁へ続く)

審査要旨

本研究は、ベトナム語の音声合成に関するものであり、特にベトナム語に特有の声調である Broken を対象としている。まず、Broken の高精度合成のためには音声パワー・パターンの活用が重要であることを実験的に明らかにしている。次に、Broken 声調と Drop 声調との音声学的關係について、合成による分析の手法により新たな知見をえている。

ベトナム語は、Level, Falling, Broken, Curve, Rising, Drop の 6 つの声調を持っている。Broken と Drop は声門閉鎖を伴った声調である。Broken は音節の中央部に声門閉鎖の特性を有したものであり、ベトナム語だけにあり、他の声調言語とは大いに異なっている。そこで本研究では、これらの声門閉鎖のある声調の聴感上の特性を検討している。

本研究で開発したベトナム語のテキスト音声合成(VieTTS)システムは、Level 声調の半音節を基本ユニットとした規則合成システムである。VieTTS は、LMA フィルタを声道フィルタとするソース・フィルタモデルによるシステムである。声調の合成のため、本研究では、基本周波数とともにパワー・パターンを用いる方法を提案した。まず、パワー制御を声調に適用し、これが Broken 声調と Drop 声調の正聴率向上に有効であることを示した。

声門閉鎖のある声調に関して、Broken の前半部と Drop との類似性を仮定し検討した。実験結果から、これらは交換可能であることが分かった。つぎに、Broken および Drop の後半部を評価した。その結果、Drop の最後部の基本周波数パターンには聴感上 80ms の閾値があること、および Broken の後半部の基本周波数パターンは、声調の判別において聴感上さほど重要でないことを明らかにした。

本論文は以下のような学術上の新しい結果を示している。(1) 高精度のベトナム語のテキスト音声合成システムを開発した。(2) パワー制御を Broken 声調と Drop 声調の合成に応用しその有効性を示した。(3) Broken 声調の前半部と Drop 声調は交換可能であるという音声学上の知見を示した。(4) 聴感上、Drop は基本周波数パターンに関する閾値を持ち、Broken の後半部はさほど重要でないという音声学上の知見を示した。以上のことから本研究は、情報工学的および音声情報学的に重要な知見を与えているといえる。また学力の確認の結果、上記の学生は専門的分野および関連分野の十分な知識を有することが判明した。以上のことより、学位論文審査員一同は、DO TU TRONG 君を博士(工学)の学位を授与するにふさわしい者とし、学位論文および学力確認をそれぞれ合格と認める。

Form 3

論 文 要 旨

Abstract

論 文 題 目

Title : “Vietnamese Text-to-Speech system and perceptual features of glottal tones”

Japanese title : “ベトナム語のテキスト音声合成システムと声門声調の知覚特性”

Speech synthesis has a long historical development, and many Text-to-Speech (TTS) systems have been successfully developed for major languages such as English and Japanese. However, there has been little research on Vietnamese speech synthesis. It leads us to study on Vietnamese speech synthesis and to build a TTS system for Vietnamese. A visual interactive speech analysis-synthesis system is also implemented as a basic tool for the research.

Vietnamese is a tonal language with six lexical tones: LEVEL (Ngang), FALLING (Huyền), BROKEN (Ngã), CURVE (Hỏi), RISING (Sắc), and DROP (Nặng). Among the six tones, BROKEN and DROP tones have glottal features. BROKEN tone, with irregular changes in the middle of the syllable, is unique for Vietnamese comparing with other tonal languages. Hence, we study perceptual features of these glottal tones.

Vietnamese Text-to-Speech (VieTTS) system is a parametric and rule-based synthesis system, in which the fundamental speech units are demisyllables with LEVEL tone. VieTTS uses a source-filter model for the speech production and a Log Magnitude Approximation (LMA) filter as a vocal tract filter. To synthesize tones, we propose and apply both fundamental frequency (F_0) and power patterns. For the first time, the power control is applied for the tone and is effective for BROKEN and DROP tones.

For perceptual features of glottal tones, we hypothesize and check the similarity of DROP tone and the first half part of BROKEN tone. The experimental results show that they are exchangeable. Consequently, the final part of DROP and that of BROKEN are evaluated. As results, DROP tone with the glottal stop at the end of the syllable has a perceptual threshold around 80 ms slope; the final part of BROKEN had a minor effect on the perception.

This dissertation shows some new results as: (1) the TTS system for Vietnamese is developed; (2) the power control is applied for synthesizing BROKEN and DROP tones; (3) the first part of BROKEN tone and DROP tone are comparable; and (4) for the perception, DROP tone has the perceptual threshold, and the final part of BROKEN tone has the minor effect.

Name: DO, TU TRONG