

Form 3

論 文 要 旨

Abstract

論 文 題 目

Title: THAI SPEECH SYNTHESIS USING A GENERATIVE MODEL OF FUNDAMENTAL FREQUENCY CONTOUR.

Thai speech synthesis by rule is developed using cepstral parameters. The system has been developed using demisyllable approach and using the log magnitude approximation (LMA) digital filter to synthesize speech sound.

Thai belongs to tonal languages. Tone, which is indicated by contrasting variations in contour of fundamental frequency ( $F_0$ ) at the syllabic level, is an important part of spoken language because the meaning of words with the same sequence of phonemes may be different if they have different tones. In Thai, there are five tonal variations traditionally named: mid (M), low (L), falling (F), high (H), and rising (R).

The purpose of the speech synthesis systems is to produce speech sound with high intelligibility and naturalness. Appropriate  $F_0$  contour is one of the most important factors to improve the intelligibility and naturalness of synthesized speech. The  $F_0$  contour plays a complex role in encoding information for the listener because it conveys not only information about syntactic structure and stress pattern, but it also helps to indicate the speaker's gender and psychological attitude toward what it is being spoken.

In order to synthesize precisely the pitch contour of Thai tone, we applied an extension of generative model of fundamental frequency contour for tonal language. Mid tone is unique for Thai comparing to Chinese. To apply the model to Thai tones, we assumed that the mid tone is neutral and we adopted the mid tone's phrase component for all tones. According to our study on pitch contour of five Thai tones using this model, the result shows that the command pattern for the local  $F_0$  components needs both positive and negative commands.

Along with this method, the pitch contours of Thai monosyllabic, disyllabic and trisyllabic words were analyzed. Based on the analysis of Thai words using this model, rules are derived and applied to synthesize Thai words and sentences. We performed listening tests to evaluate intelligibility of the model for Thai tone generation. The average intelligibility scores were 99.3%, 98.8% and 96.6% for monosyllabic, disyllabic, and trisyllabic words, respectively. The generative model of  $F_0$  contours for Thai words were shown to be effective.

Furthermore, we derived rules to synthesize suprasegmental  $F_0$  contours using the trisyllabic words' parameters. The parameters of the first, third, and second syllables were assigned to the initial, ending, and remaining syllables in a sentence, respectively. We performed listening tests to evaluate the intelligibility score and naturalness of synthesized speech. As a result, all phrases/sentences were completely identified. The MOSs (Mean Opinion Score) was 3.50 while the original and analysis/synthesis samples were 4.82 and 3.59, respectively. These results showed that the rules for tonal assimilation using the generative model of  $F_0$  contours in connected speech were shown to be effective.

Name Pusadee Seresangtakul

(様式5-3)

平成17年8月5日

琉球大学大学院

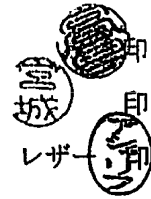
理工学研究科長 殿

論文審査員

主査 氏名 高良富夫

副査 氏名 宮城隼夫

副査 氏名 アシャリフ モハマッド レザー



### 学位（博士）論文審査及び学力確認の終了報告書

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、下記の通り報告します。

記

申請者	氏名 Pusadee Seresangtakul	
現住所		
成績評価	学位論文 ○合格 不合格	学力確認 ○合格 不合格
論文題目	Thai speech synthesis using a generative model of fundamental frequency contour (基本周波数パターンの生成モデルを用いるタイ語音声の合成)	
審査要旨 (2000字以内)		

#### 審査要旨

本論文では、ケプストラムパラメータを用いたタイ語音声の規則合成に関する研究成果が示されている。このシステムでは半音節法を採用し、音声合成用デジタルフィルタとしては、対数振幅特性近似 (LMA) フィルタが使用されている。

タイ語は声調言語に属する。声調は、音節における基本周波数( $F_0$ )パターンの違いによって特徴付けられ、音声言語の重要な構成要素である。その理由は、同じ音素の並びの単語でも、異なる声調をもつと、意味が異なるからである。タイ語には、伝統的に mid(M), low(L), falling(F), high(H), rising(R) と呼ばれる5つの型の声調がある。

音声合成システムの目標は、高い正聴率と自然性を持つ音声を作り出すことである。適切な $F_0$ パターンを用いることは、合成音声の正聴率と自然性を高めるために最も重要なことのひとつである。聞き手への情報をコード化するため、 $F_0$ パターンは複雑な役割を果たす。それは、 $F_0$ パターンが、文法的またストレスパターンの情報を担うだけでなく、話者の性および話される内容に対する心理的態度を表すためにも役立つからである。

タイ語の正確なピッチパターンを合成するため、基本周波数生成モデルを拡張して声調言語に応用した。声調 mid は、中国語と比べてタイ語に特有のものである。このモデルをタイ語に応用するため、mid を中性的と仮定し、声調 mid の句成分をすべての声調の句成分として採用する。このモデルを用いた5つのタイ語声調のピッチパターンに関する申請者らの研究によれば、局所的 $F_0$ 成分の命令パターンは正と負の両方の命令が必要であるという結果が示されている。

この方法に従って、タイ語の単音節、2音節、及び3音節単語の $F_0$ パターンを分析した。このモデルを用いたタイ単語の分析に基づいて、規則を導き、タイ語の単語及び文を合成するために応用した。タイ語の声調生成モデルの正聴率を評価するため、聴取実験を行った。平均正聴率は、単音節、2音節、及び3音節単語のそれぞれについて、99.3%、98.8%、及び96.6%であった。これにより、タイ語単語の $F_0$ パターンの生成モデルは、有効であることが示された。

さらに、3音節単語のパラメータを用いて、超分節的 $F_0$ パターンを合成するための規則を導いた。1番目、3番目、及び2番目の音節のパラメータがそれぞれ、文の最初の音節、最後の音節、およびその他の音節に割り当てられた。合成音声の正聴率および自然性を評価するため聴取実験を行った。その結果、句や文は100%正しく聴取された。MOS (平均評価値) は、3.50であった。一方、原音声および分析合成音声のMOSはそれぞれ、4.82と3.59であり、これらに近い値であるので、自然性における有効性も示された。

本論文は、タイ語音声の新しい声調合成法を提案したものであり、学術的な意義が大きいといえる。また学力確認の結果、上記の者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、Pusadee Seresangtakul さんを学術博士の学位を授与するにふさわしいものとし、学位論文および学力確認をそれぞれ合格と認める。