摘要

以大量語料庫為基礎（Corpus-Based）的中文語音合成系統，因為單元選取的不一致，在片段接合上容易造成自然度下降，而且語料的收集亦不容易。因此，有別於前人的設計，我們發展了一套中文的語音合成系統，採用承載句（Carrier Sentence）語料庫設計為基本的合成單元，以解決語料收集困難的問題，並且建構適當的韻律參數模型，以期能在語音合成實作上，同時達到合成速度、語料庫大小、與自然度皆有不錯的水準。

本論文探討中文語音合成之韻律參數產生的幾種常見的方法：（1）使用類神經網路（Neural Network）方法為基礎的韻律產生器，（2）線性迴歸器（Linear Regression）與（3）支撐向量機（Support Vector Machine, SVM）的迴歸模型訓練，來設計中文語音的韻律模型。

為了使模型最佳化，我們對倒傳遞類神經網路（Back Propagation Network, BPN）與支撐向量機的設定參數做實驗，並且把各種韻律參數分別訓練，以提高預測力。

我們對上述各個模型做內外部測試，並以 Root Mean Square Error (RMSE)值的高低作為比較基準，最後選取最佳模型進行聽測評估。根據 RMSE 值的的實驗結果，以類神經網路與支撐向量機較低，線性迴歸法的 RMSE 誤差值較其他二者稍高。因為支撐向量機的模型穩定度較高，所以我們又針對支撐向量機的設定參數作測試，以提高支撐向量機的預測準確性。由聽測實驗結果得知，經過韻律模型產生的合成語句，其自然度比原先以承載句為基礎的語音合成系統有較佳的表現。

Abstract

A corpus-based TTS system is likely to have degradation in naturalness due to the acoustic mismatch of between selected synthesis units. Moreover, the collection of the speech corpus is also a labor-intensive task. Therefore, we have developed a carrier-sentence-based TTS system for Mandarin Chinese. Our lab is consistently trying to improve the TTS system such that a balance can be achieved considering synthesis speed, corpus size, and naturalness of the output utterances.

In this thesis, several methods that generate the prosodic parameters of a Mandarin TTS system are investigated. These methods include linear regression, the artificial neural network, and the regression model of support vector machine (SVM). We compare the RMSE of both inside and outside tests of these methods to find out the best regression model for prosody generation, and carry out a listening test. The neural network and SVM can achieve better performance in terms of RMSE. We have also performed additional optimization on the parameters of SVM.

Listening test shows that after our prosody modification, the TTS system indeed generates more natural-sounding utterances.