

Title	Acoustic-articulatory DNN Model based on Transfer Learning for Pronunciation Error Detection and Diagnosis( Abstract_要旨 )
Author(s)	Duan, Richeng
Citation	Kyoto University (京都大学)
Issue Date	2018-09-25
URL	<a href="https://doi.org/10.14989/doctor.k21391">https://doi.org/10.14989/doctor.k21391</a>
Right	学位規則第9条第2項により要約公開; 許諾条件により要約は2019-08-28に公開
Type	Thesis or Dissertation
Textversion	none

( 続紙 1 )

京都大学	博士 (情報学)	氏名	Duan Richeng
論文題目	Acoustic-articulatory DNN Model based on Transfer Learning for Pronunciation Error Detection and Diagnosis (発音誤りの検出と診断のための転移学習に基づく音響・調音DNNモデル)		
(論文内容の要旨)			
<p>Computer-assisted language learning (CALL) is becoming popular because of its flexibility of allowing students to practise in a stress-free environment at their convenient time and pace. Computer-assisted pronunciation training (CAPT) as an indispensable component of CALL systems aims to improve learner's speaking skill through providing corrective feedbacks for pronunciation errors just like an experienced teacher does. Acoustic model, which plays a crucial role in CAPT, requires a large amount of training data especially for DNN models. However, it is difficult to collect a large amount of non-native speech and also to precisely annotate them.</p> <p>This thesis addresses effective and efficient learning of DNN acoustic-articulatory model without using non-native training data. Several transfer learning based methods are investigated to effectively learn the acoustic-articulatory model. The key idea of the transfer learning is making use of knowledge and resources in related tasks to improve the performance of the target task. Transferring cross-lingual knowledge between two native languages, which have a large speech database, allows for implicit learning of the acoustic representation of non-native speech.</p> <p>Chapter 1 provides an overview of the research topic of foreign language pronunciation learning. It points out the issue that the pronunciation feedback in current CAPT systems is not instructive enough to guide the learners, and presents the proposed solution of providing articulatory feedback.</p> <p>Chapter 2 reviews the relevant studies on CAPT. The three main aspects of non-native speech recognition, pronunciation error detection, and pronunciation error diagnosis are reviewed.</p> <p>Chapter 3 describes the acoustic modeling for the CAPT system. The training procedure of acoustic-phonetic and acoustic-articulatory models are described, both of which are employed in the following chapters.</p> <p>Chapter 4 presents transfer learning based acoustic-articulatory modeling for Mandarin Chinese pronunciation error detection and diagnosis. It starts with enhancing the non-native acoustic-articulatory models with two kinds of knowledge (related-task and cross-lingual). Combination of both transfer learning methods is explored as well. Experimental results demonstrate that these methods significantly improve the classification accuracy of native articulatory attributes and the detection performance of pronunciation errors produced by the non-native learners.</p> <p>Chapter 5 extends the cross-lingual knowledge transfer to English learning. It is applied to not only articulatory level but also phone level. The acoustic-phonetic model is used to recognize non-native speech and</p>			

perform pronunciation error detection while the acoustic-articulatory model conducts diagnosis of the detected pronunciation errors. The one-pass framework and the system based on two-pass pronunciation error detection and diagnosis are compared in terms of accuracy and flexibility. The effects of cross-lingual knowledge based transfer learning on acoustic-phonetic modeling are confirmed in both non-native English speech recognition and pronunciation error detection. For diagnosis of pronunciation errors, the proposed acoustic-articulatory modeling method is effective for the non-native English speech.

Chapter 6 introduces additional two improvements for acoustic-articulatory modeling. It first presents a multi-label learning scheme, which considers the interaction effects among different kinds of articulation attributes and largely reduces the model training time. The other method employs a label correction process based on the consistency of the articulatory attribute labels. Experimental results show that proposed methods significantly improve the DNN based acoustic-articulatory models.

Chapter 7 concludes the thesis and suggests directions for future research.

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し  
審査結果の要旨は日本語500～2,000字程度で作成すること。

(論文審査の結果の要旨)

外国語の発音訓練を行うシステムを実現する上で、非母語話者の発音を正確に認識する音響モデルが必要となるが、非母語話者の発音ラベル付きデータの大規模な収集は容易でない。本論文は、調音レベルのモデル化に着目し、母語話者のデータベースから転移学習に基づいて、非母語話者の音響・調音DNN（ディープニューラルネットワーク）モデルの構築に取り組んだ研究をまとめたもので、主な成果は以下の通りである。

1. 発音訓練においては調音（調音位置と調音様式）レベルのフィードバックが有用であることから、DNNの転移学習に基づく調音モデルの構築法を検討した。学習対象言語の音素識別と学習者の母語の調音分類の2つのタスクから転移学習を行う方法を提案した。日本人による中国語学習のタスクにおいて、母語（中国人）話者の調音分類と非母語（日本人）話者の調音誤り検出の両方で、各々の転移学習の有効性と両者を組み合わせる効果を確認した。
2. 上記の枠組みでは学習者は指定された単語を発声する必要があるが、自由に発声した単語から発音誤りを検出し、診断を行うために、音素レベルのモデルと調音レベルのモデルの両方を言語間の転移学習に基づいて構築した。日本人による英語学習のタスクにおいて、音声認識・誤り検出・診断を1パスで行う方法と2パスで行う方法を実装し、いずれにおいても転移学習の効果を確認した。
3. 音響・調音モデルの転移学習をさらに改善するために、複数ラベルを用いた同時学習と誤り訂正の方法を提案した。これらは複数の調音ラベルの関連性と一貫性を利用したもので、母語話者の調音分類と非母語話者の調音誤り検出の両方で有効性を確認した。

以上のように本論文は、外国語の発音のモデル化において転移学習の方法を提案し、非母語話者の発音データを一切用いずにその音響・調音モデルが構築できることを示したもので、学術上・実用上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、平成30年 8月28日に論文とそれに関連した内容に関する口頭試問を行った結果、合格と認めた。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。  
要旨公開可能日： \_\_\_\_\_ 年 \_\_\_\_\_ 月 \_\_\_\_\_ 日以降