博士学位論文審査要旨

２０１７年2月１７日

論 文 題 目：Phoneme Set Design for Second Language Speech Recognition
（第二言語音声認識のための音素セットの構築に関する研究）
学位申請者：王　暁芸
審 査 委 員：
主　査：　同志社大学大学院理工学研究科　教授　山本　誠一
副　査：　同志社大学大学院理工学研究科　教授　片桐　滋
副　査：　ヘルシンキ大学　現代言語学部　教授　Wilcock Graham

要　　旨：

本論文は第二言語話者発話の音声認識技術に関する研究成果である。本論文で提案されている手法は、第二言語話者の発話をネイティブ話者の発話とは音響特徴量空間で異なる頻度分布を持つ情報源とみなし、これを表現する適切な音素セットを構築する手法である。

第２章では、第二言語話者の音声認識の課題と第二言語話者の音響的特徴の違いにのみ着目した従来の適応化手法について概説している。第３章は日本人による英語発話の特徴を述べ、第４章では日本人による英語発話を認識する音声認識技術開発の基盤となる音声コーパスについて述べる。第５章では、第二言語話者の音響的および言語的な特徴を考慮した音素セット構成法について述べる。提案手法では、認識対象となる第二言語と発話者の母語との調音位置や調音様式などの類似性に加え，音素セットの決定により生じる同音異義語の発生による単語識別性能の低下を考慮した総合的な基準に基づき、最適な音素セットを決定する。更に、提案手法を日本人学生の英語発話の音声認識に適用し、種々の条件下で認識実験を実施した結果に基づき認識精度の向上を検証している。

第６章では音素セット構築の際に第二言語話者の使用単語の偏り等の言語的な特徴により生じる音響特徴量空間での頻度分布の違いを設計要素として取り入れた効果を検証する。第７章は発話者の第２言語話者としての習熟度を音素セット構築の際の要素として取り入れた手法を述べる。第８章では第７章で述べた手法を音声認識装置に組み込む具体的な手法を述べ、習熟度の異なる広範囲の第二言語話者に対しても高精度な音声認識性能を示すことを述べる。

本研究で提案した第二言語話者の音素セット構成法は、種々の第二言語話者の発話を高精度で認識する音声認識を可能とし、母語の異なる話者による第二言語発話が行き交うグローバル化した社会での音声による情報検索等への応用が期待される。

よって、本論文は、博士（工学）（同志社大学）の学位論文として十分な価値を有するものと認められる。

総合試験結果の要旨

２０１７年2月１７日

論 文 題 目：Phoneme Set Design for Second Language Speech Recognition
（第二言語音声認識のための音素セットの構築に関する研究）
学位申請者: 王　暁芸
審 査 委 員：
主　査：　同志社大学大学院理工学研究科　教授　山本　誠一
副　査：　同志社大学大学院理工学研究科　教授　片桐　滋
副　査：　ヘルシンキ大学　現代言語学部　教授　Wilcock Graham

要　　旨：

本論文提出者は、理工学研究科情報工学専攻博士後期課程に在籍している。本論文の主たる内容は、Transactions of IEICE, Vol. E98-D, No.1、Transactions of IEICE, Vol. E98-D, No.12 および Proceedings of Interspeech2016 (Transactions of IEICE 印刷中)に掲載され、十分な評価を得ている。

2016 年 12 月 22 日 10 時より約 1 時間 50 分にわたって提出論文に関する学術講演会（博士論文公聴会）が開催され、種々の質疑討論が行われたが、論文提出者の説明により十分な理解が得られた。

さらに、講演会終了後、審査委員により論文に関連した諸問題につき口頭試問を実施した結果、十分な学力を有することが確認できた。

提出者は語学試験に合格しており、また英語による論文発表および口頭発表を行っており、十分な語学能力を有すると認められる。

よって、総合試験の結果は合格であると認める。

博 士 学 位 論 文 要 旨

論 文 題 目： Phoneme Set Design for Second Language Speech Recognition
第二言語音声認識のための音素セットの構築に関する研究

氏　　名： 王 暁芸

要　　旨：

In today's environment of rapid globalization, people have increasing opportunities for speaking in foreign languages, and the ability to communicate in foreign language is more important than ever. Various applications (dialogue-based computer assisted language learning (CALL) systems, car navigation system, hotel reservation systems, mobile platforms, etc.) are incorporating spoken language interfaces for non-native speakers to provide more convenient life. The key role for these kinds of interfaces is non-native automatic speech recognition (ASR) or second language (L2) speech recognition.

Non-native speakers usually have a limited vocabulary and a less than complete knowledge of the grammatical structures of the target language. This limited vocabulary forces speakers to express themselves in basic words, making their speech sound unnatural to native speakers. In addition, non-native speech includes less fluent pronunciation and mispronunciation even in case in which it is well composed. Therefore, non-native speakers represent a significant challenge for state-of-the-art ASR.

Two major problems need to be addressed for non-native ASR: (1) For given speech with limited vocabulary and less knowledge of grammatical structures, how can a speech recognizer take advantage of these characteristics by the speakers? (2) For given speech with different pronunciation variations, how can a system recognize speakers correctly? In order to tackle above problems, I propose to derive the customized phoneme set different from the canonical one for recognition of non-native speech, particularly when the mother tongue of users is known. Then based on the efficiency of derived phoneme set for the non-native speakers with different proficiency levels, we build a proficiency-dependent phoneme set to capture their different pronunciation variations.

The dissertation focuses on several important aspects for above two problems: *the phonological knowledge between mother tongue (L1) and target language (TL), acoustic and linguistic features of speech, proficiency of non-native speakers*. The first part of the dissertation proposes the statistical method using integrated acoustic and linguistic features on the phonetic decision tree (PDT) to derive the phoneme set for L2 speech recognition. As the results of the first part of the dissertation, the effect of the derived phoneme set is different depending on the speakers' proficiency in L2. To further improve the second language ASR, the second part of the dissertation investigates the relation between proficiency of speakers and a derived phoneme set customized for them. The investigated results are then used as the basis of a novel speech recognition method using a lexicon in which the pronunciation of each lexical item is represented by multiple phoneme sets for each L2 speaker with various proficiency levels.

The dissertation verifies the efficacy of the proposed methods using second language

speech collected with a translation game type dialogue-based English CALL system. In conclusion, the dissertation shows that the feasibility of building a speech recognizer with the proposed methods is able to alleviates the problem caused by confused mispronunciation by L2 speaker. As a result, the ASR system can achieve the higher recognition accuracy with the derived phoneme set than that with the canonical phoneme set which is used in the traditional English speech recognition system.