

京都大学教育研究振興財団助成事業
成 果 報 告 書

平成21年 7月 1日

財団法人京都大学教育研究振興財団

会 長 辻 井 昭 雄 様

所属部局・研究科 大学院情報学研究科

職 名・学 年 助教

氏 名 駒 谷 和 範

事業区分	平成20年度・長期派遣助成		
研究課題名	VoIP上での音声対話システム構築ソフトウェアの開発		
受入機関	米国・ペンシルバニア州・カーネギーメロン大学		
渡航期間	平成20年5月26日 ~ 平成21年6月5日		
成果の概要	タイトルは「成果の概要/報告者名」として、A4版2000字程度・和文で作成し、添付して下さい。「成果の概要」以外に添付する資料 無 有()		
会計報告	交付を受けた助成金額	2,550,000 円	
	使用した助成金額	2,550,000 円	
	返納すべき助成金額	0 円	
	助成金の使途内訳 (使用旅費の内容)	航空賃, 空港使用料, サーチャージ等:	274,390円
		日当:	568,900円
		宿泊料:	1,706,710円
		(規程に基づく出張旅費の一部として上記金額を充当)	

成 果 の 概 要

駒谷 和範（京都大学大学院情報学研究科・助教）

2008年5月26日から2009年6月5日の間、京都大学教育研究振興財団からの支援を受け、米国ペンシルバニア州ピッツバーグにあるカーネギーメロン大学（CMU）に滞在し、Visiting Scientistとして研究を行った。受け入れ先はSchool of Computer Science, Computer Science DepartmentのDr. Alexander Rudnickyである。CMUにはRI (Robotics Institute)やLTI (Language Technology Institute)のような、目的の特化した研究所が5つあり、それぞれに数多くのFacultyが集まっている(例えばRIは69名、LTIは31名など)。このため先進的かつ多様な研究が活発に行われており、博士論文の提案審査や公聴会を含め、数多くの公開セミナーも開催されていた。これらに参加し聴講したことも滞在中に得た貴重な経験のひとつである。

滞在中に実施し、得た成果として主に下記3点が挙げられる。

1. 音声対話システムにおける Implicit label を用いた音声認識誤りの予測
2. Olympus への Julius 接続モジュールの開発
3. 他大学や研究所でのトークの実施

まず1点目として、Implicit label という、対話中におけるユーザの応答から得られる情報を生かして、書き起こし（正解ラベル）なしでシステムの能力を向上させる方法を開発した。通常、音声対話システムの性能を向上させる際には、データを大量に収集したうえで、正解ラベルを手で付与する必要がある。この際、対話中に現れる、ユーザ発話の音声認識結果が正しいことを示す対話パターンを活用することにより、自動的に正解ラベルの取得を試みる。このような手法により、人手を介さずにデータを収集しながらシステムの能力向上が実現できる。

具体的には、本研究では「はい」「いいえ」といったユーザの肯定・否定応答は比較的高精度に認識可能であると仮定したうえで、これらの確認内容に対応する音声認識結果の正否を、肯定・否定応答の内容が指す点を活用した。また本研究で利用しているデータでは各コールにユーザの電話番号が記録されているため、これに基づき各ユーザのそれまでの音声認識率を推定する。ユーザの音声認識率は、以前に我々が開発した各ユーザのバージョン率と同様に、ユーザのシステムに対する習熟度を表す尺度のひとつである。このため、ユーザの習熟度に応じてシステムのふるまいを変化させるのに用いることができる。

予備実験として、まず「ユーザの肯定・否定応答が正しく認識できる」という仮定の検証を行った。この結果、全体の音声認識率（発話単位）は72.6%であったのに対して、肯定応答は97.4%、否定応答は87.5%と高い精度で認識できていることを確認した。さらに人手による書き起こしから計算した各ユーザの音声認識率と、本研究で利用する Implicit label に基づく音声認識率の相関係数は0.806となり、高い相関を示すことも確認した。

これにより前提が確認されたため、評価実験として今回はバーズイン発話（ユーザがバーズインをしながら入力した発話）の誤りの予測を行った。バーズイン発話は、発話の途中に言い淀みが含まれやすいという傾向が知られており、実際に我々のデータにおいても発話音声認識率は45.6%と低い。この誤りを、各ユーザのバーズイン率と推定音声認識率をプロファイルとして用いることにより予測した。バーズイン率とともに当該ユーザの音声認識率を用いることで、バーズイン発話の誤り予測精度が71.8%から74.3%へと向上し、これらの尺度の有効性を示した。なお本研究の内容を2009年8月に開催されるACL-IJCNLP09にshort paperとして投稿し、採択された。採択率は25%(=96/378)であった。

2点目として、CMUで開発されているOlympusに対して、日本語の音声認識エンジンJuliusを接続するプログラムを開発した。Olympusは2008年に公開された音声対話システムのアーキテクチャであり、音声対話システムの各モジュール群をGalaxy architecture上で動作させたものである。具体的には、音声認識部としてSphinx、言語理解部としてPhonex、対話管理部としてRavenclaw、インタラクション管理部としてApollo、文生成部としてRosetta、音声合成器としてKalliopeというモジュールが用いられている。音声認識器Sphinxは英語を対象として開発されたものであるため、日本語での音声認識を可能とするため、日本で広く用いられている音声認識エンジンJuliusをOlympusに接続するプログラムを新たに開発した。この際、Sphinxの部分認識仮説の生成機能にApolloが依存しているのを解消するため、Juliusの基本的な出力で動作するようApolloを改修し、またOlympus内のシステムの不具合を修正するなど、CMUでのシステム開発にも貢献した。また、言語理解部Phonexの日本語での動作を確認するために、日本語でのタスク文法の記述も行った。この際にもPhonexが日本語を処理する際の不具合を指摘し解消した。

3点目として、滞在先であったCMUに加え、ピッツバーグ大学やニュージャージー州Florham ParkにあるAT&Tシャノン研究所で約1時間の研究のトークを行った。自分のトークに対する質疑応答に加え、訪問先の研究者らと個別にディスカッションを行い、意見交換や交流を行った。とりわけピッツバーグ大学のDiane Litman教授らや、AT&T研究所のJason Williams博士らと意見交換できたのは非常に貴重な体験であった。これらを通じて、通常の学会発表よりも深い議論を行うことができ、またその前後を通じて個人間の関係を深めることもできた。

上記の主な成果以外にも、ピッツバーグ大学の英語教師から、個人的に英会話や発表の発音指導を継続的に受ける機会があったなど、当地を訪問したからこそできた貴重な経験を数多くできた。京都大学教育研究振興財団からのご支援とともに、渡航を許可してくださった奥乃教授をはじめとした研究室・学科・専攻の先生方には深く感謝したい。