



## 10. キーワード

(1) 国際研究者交流	(2)	(3)	(4)
(5)	(6)	(7)	(8)

## 11. 現在までの達成度

(区分)(2) おおむね順調に進展している。

(理由)

本プロジェクトにより、音声処理関連の会議に出席する機会を得ることができ、少数言語の音声認識研究に関して多くの研究者と議論を交わすことができた。少数言語のデータ収集について、インドネシアに現存する4つの主要な少数言語からデータ収集を始めた(ジャワ語、スンダ語、バリ語、バタック語)。結果、下記少数言語のコーパスを収集することができた:(1) テキストコーパス: インドネシア少数言語のテキストコーパスは、地元の新聞・雑誌から収集し、ジャワ語、スンダ語、バリ語、バタック語の各少数言語につき2000~3000文を収集できた。うち、1000文を選んでネイティブの話者による校閲を行った。その後、Greedyアルゴリズムを用いて音素バランス付き文を処理し、各言語につき225の音素バランス付き文を生成した;(2) 音声コーパス: 上記の音素バランス付き225文を用い、各少数言語につき10人の話者による音声データの収集を行った。音素バランス付きコーパスの他に、インドネシア語からジャワ語、スンダ語、バリ語、バタック語に翻訳された50文の対訳音声コーパスも収集した;(3) 音声認識システムベースライン: インドネシア語の音声認識ベースラインは、既存のインドネシア語音声データ(ニュースおよび電話会話)を用いて学習した。このコーパスでは、400人の話者の発話が収録されている。話者の発話は、標準インドネシア語のアクセントと3つのネイティブ言語(バタック語、ジャワ語、スンダ語)のアクセントが付けられている。各話者は210文を発話しており、合計84,000発話、80時間の音声が取られている。上記の音声リソースおよび音声認識ベースラインに基づき、インドネシア少数言語の特性を分析することが可能となった。

## 12. 今後の研究の推進方策等

(今後の推進方策)

今後の課題として、研究計画にあげた項目の継続的作業を検討している。

1. より多くの少数言語の音声源を収録する。それで、少数言語の音声コーパスで音声認識システムのベースラインを適応/再学習する。
2. 知識統合フレームワーク: 知識統合フレームワークの設計および開発、提案したフレームワークを音声認識システムの各レベルに適用
3. 音声認識の性能評価, プロトタイプの実証実験, プロジェクトまとめ

(次年度の研究費の使用計画)

1年の計画で終了予定であったタスクの大半を達成することができたが、データ収集は、事務手続き等が原因で当初の予定よりも長くかかった。したがって、データ収集は特定の少数言語に限定され、様々な知識源を統合したフレームワークによる音声認識のベースラインを構築するには至らなかった。これが最初の年に予算が残っている主な理由です。次年度の研究費の使用計画は以下の通りです: 研究成果発表(100), 研究成果発表(430) 話者謝金, 書き起こし謝金, 主観評価(800), 計(1330)

## 13. 研究発表(平成24年度の研究成果)

〔雑誌論文〕計(1)件 うち査読付論文 計(1)件

著者名	論文標題【掲載確定】				
Auliya Sani	Towards Language Preservation: Preliminary Collection and Vowel Analysis of Indonesian Ethnic Speech Data				
雑誌名	査読の有無	巻	発行年		最初と最後の頁
O-Cocosda	有	Vol. 1	2	012	pp. 118-122
掲載論文のDOI(デジタルオブジェクト識別子)					
なし					

〔学会発表〕計(1)件 うち招待講演 計(0)件

発表者名	発表標題【発表確定】		
Auliya Sani	Towards Language Preservation: Preliminary Collection and Vowel Analysis of Indonesian Ethnic Speech Data		
学会等名	発表年月日	発表場所	
International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques	2012年12月11日	Macau, China	

〔図書〕計(0)件

著者名	出版社		
書名	発行年	総ページ数	

## 14. 研究成果による産業財産権の出願・取得状況

〔出願〕計(0)件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

(取得) 計( 0 )件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別
				出願年月日	

15.備考

--