

令和 3 年 6 月 24 日現在

機関番号：14301

研究種目：挑戦的研究（萌芽）

研究期間：2018～2020

課題番号：18K19814

研究課題名（和文）アイヌ語アーカイブを対象としたEnd-to-End音声認識の研究

研究課題名（英文）End-to-End Automatic Speech Recognition for the Archive of Ainu Folklore

研究代表者

河原 達也（Kawahara, Tatsuya）

京都大学・情報学研究科・教授

研究者番号：00234104

交付決定額（研究期間全体）：（直接経費） 4,800,000円

研究成果の概要（和文）：アイヌ民話（ウエペケレ）の音声認識の研究に取り組んだ。2つの博物館から提供されたアイヌ語アーカイブのデータを元に、沙流方言を対象としたアイヌ語音声コーパスを構築した。このコーパスを用いてEnd-to-Endモデルに基づく音声認識システムを構成した。音素・音節・ワードピース・単語の4つの認識単位について検討し、音節単位が最もよいことを示した。音声認識精度が話者オープン条件において大幅に低下する問題に対して、CycleGANを用いた教師なし話者適応を提案した。さらに、日本語とアイヌ語が混合した音声に対して、音素認識と単語認識を組み合わせることで、アイヌ語の区間の検出（言語識別）を実現した。

研究成果の学術的意義や社会的意義

アイヌ文化の多くが口頭で継承されてきましたが、アイヌ語は2009年にUNESCOにより「極めて深刻な」消滅危機言語に認定される事態となっています。以前から口頭伝承を録音・記録する活動が様々に行われてきたが、その書き起こし・アーカイブ化には膨大な手間とアイヌ語の知識を必要とするため、多くが未整備のままでした。本研究により、アイヌ語のアーカイブ構築の効率化への寄与が期待されます。実際に、アイヌ民族博物館において音声と書き起こし同期のための対応付けに活用され、1時間のデータに対して、人手で1日要する作業がほぼ完全に自動化できました。

研究成果の概要（英文）：We have investigated automatic speech recognition (ASR) of Ainu folklores (Uwepeker). First, we constructed an Ainu speech corpus for the Saru dialect based on data provided by two museums which have constructed Ainu archive. Next, we built an ASR system based on the end-to-end model, and compared four recognition units of phones, syllables, word pieces, and words. With the syllable unit, we achieved phone recognition accuracy of 93.7% and word recognition accuracy of 78.3%. To address the problem of significant degradation in the speaker-open condition, an unsupervised speaker adaptation method using CycleGAN is proposed. Finally, we also implemented language identification in Japanese and Ainu mixed speech by combining phone and word recognition modules.

研究分野：知能情報学

キーワード：アイヌ語 音声認識 音声合成 消滅危機言語 End-to-Endモデル

1. 研究開始当初の背景

音声認識研究は近年飛躍的な発展を遂げて、実用的な認識性能を実現した。しかし、これは英語や日本語のように大規模な書き起こし付きデータベースがあることが前提で、主要言語に限定される。世界には5千以上の言語が存在するが、そのうち半数以上は死語になる危険性をはらんでいる。日本における先住民族としてアイヌがいる。明治以降アイヌの文化は急速に失われ、2009年にはUNESCOがアイヌ語を「極めて深刻」な消滅危機言語に認定した。一方で、**アイヌ語には大量の録音資料が作成・保存されているが、その大半が書き起こしされていない**。本研究では、この音声アーカイブを(母語話者がほとんどいない状況で)効率的に書き起こすための音声認識システムの実現を目指す。

2. 研究の目的

大規模な書き起こし付きデータベースが存在せず、しかも電子化された単語辞書や形態素解析システムもない前提での音声認識は非常に大きな挑戦的課題である。従来の音声認識システムが、単語(日本語では形態素解析結果)を単位とした言語モデルと、音素を単位とした音響モデルに基づいて構築されているという枠組みの転換を要する。これに関して近年、深層学習を発展させた End-to-End 音声認識という枠組みが検討されている。これは音素や文字を単位としたリカレントニューラルネットワークにより構成するものであるが、実用的な音声認識性能を実現するには単語辞書や単語単位の言語モデルを後処理として適用する必要がある。本研究では、純粋に End-to-End の枠組みでアイヌ語の音声認識を実現する。そのための音素や音節(カナ)などのサブワード単位を検討する。

3. 研究の方法

(1) アイヌ語音声データベースのアノテーション

アイヌの博物館でこれまでに書き起こしが作成されている音声資料を、話者や内容毎に整理した上で、発話や文章程度に区切り、書き起こしと対応づける。書き起こしは、日本語カナと音素の系列及び日本語訳となっているため、カナ(=音節)と音素の両方のモデル化が可能である。これにさらに、“カムイ/kamuy”のような重要な単語(主要な名詞を想定)をアノテーションする。

(2) 音素・音節・単語単位のEnd-to-End音声認識の構築

音素・音節・単語単位の音声認識のモデルを学習する。LSTM(Long Short-Term Memory)に基づいて、CTC(Connectionist Temporal Classification)とAttentionモデルの枠組みで学習を行う。これらはHMMを用いずに、異なる長さの系列間の写像を学習する方法で近年着目されているが、大規模な言語資源が存在しない言語において特に有効であることが期待され、この点を実証することを目指す。

(3) 書き起こしや検索における評価

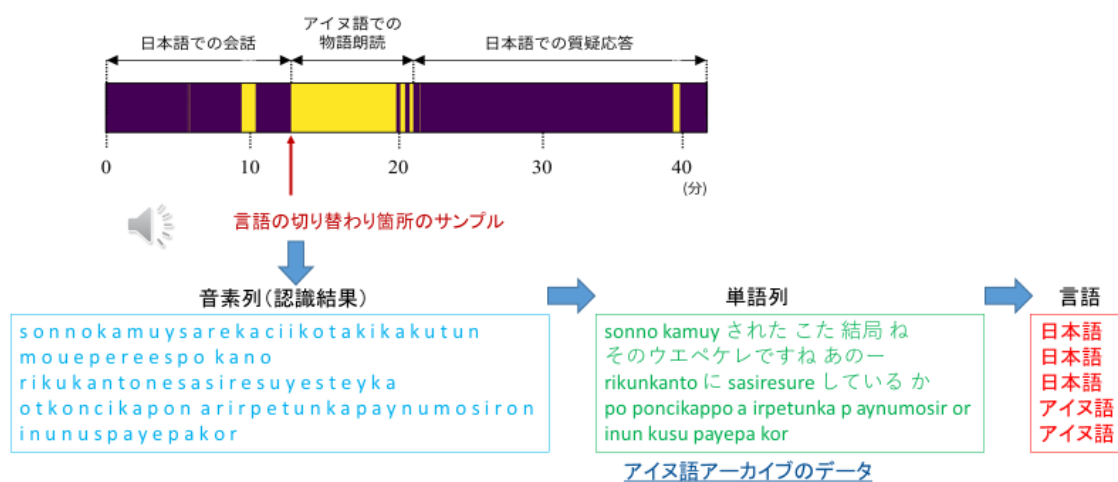
上記の音声認識システムで作成される書き起こしが、音声アーカイブの書き起こしや検索において有用であるか評価を行う。また、音声認識の話者依存性に関する検討も行う。

4. 研究成果

まず、2つの博物館から提供されたアイヌ語アーカイブのデータを元に、沙流方言を対象としたアイヌ語音声コーパスを構築した。次に、このコーパスを用いて注意機構モデルに基づく音声認識システムを構成し、音素・音節・ワードピース・単語の4つの認識単位について検討した。その結果、音節単位での音声認識精度が最も高くなることがわかり、話者クローズド条件と話者オープン条件のそれぞれについて、音素認識精度で93.7%と86.2%、単語認識精度で78.3%と61.4%を実現した。音声認識精度が話者オープン条件において大幅に低下する問題に対して、CycleGANを用いた教師なし話者適応を提案した。これは、学習データ内の話者の音声から認識対象話者の音声への写像をCycleGANに学習させ、学習データ内の音声を全て認識対象話者風の音声に変換するものである。本手法によって最大で相対60.6%の音素誤り率の改善を得た。さらに、日本語とアイヌ語が混合した音声における言語識別についても検討を行い、音素認識と単語認識を用いた構成で一定の識別性能を達成できることを示した。

実際に、アイヌ民族博物館において音声と書き起こし同期のための対応付けに活用されている。1時間のデータに対して、人手で1日要する作業がほぼ完全に自動化できた。

音声資料からのアイヌ語民話区間の音声認識



5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 松浦孝平, 三村正人, 河原達也.	4. 巻 28
2. 論文標題 アイヌ民話アーカイブに対する音声認識	5. 発行年 2021年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 (採録決定)
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計12件（うち招待講演 0件 / うち国際学会 7件）

1. 発表者名 K.Matsuura, S.Ueno, M.Mimura, S.Sakai, and T.Kawahara.
2. 発表標題 Speech corpus of Ainu folklore and end-to-end speech recognition for Ainu language
3. 学会等名 Int'l Conf. Language Resources & Evaluation (LREC) (国際学会)
4. 発表年 2020年

1. 発表者名 K.Matsuura, M.Mimura, S.Sakai, and T.Kawahara.
2. 発表標題 Generative adversarial training data adaptation for very low-resource automatic speech recognition.
3. 学会等名 INTERSPEECH (国際学会)
4. 発表年 2020年

1. 発表者名 松浦孝平, 三村正人, 坂井信輔, 河原達也.
2. 発表標題 日本語アイヌ語混合音声における言語識別.
3. 学会等名 日本音響学会研究発表会
4. 発表年 2021年

1. 発表者名 松浦孝平, 三村正人, 坂井信輔, 河原達也.
2. 発表標題 低資源言語音声認識における敵対的声質変換を用いた教師なし話者適応学習.
3. 学会等名 日本音響学会研究発表会
4. 発表年 2020年

1. 発表者名 松浦孝平, 上乃聖, 三村正人, 坂井信輔, 河原達也
2. 発表標題 民話を対象としたアイヌ語音声コーパスとend-to-end音声認識
3. 学会等名 情報処理学会研究報告 SLP-130-16
4. 発表年 2019年

1. 発表者名 松浦孝平, 上乃聖, 三村正人, 坂井信輔, 河原達也
2. 発表標題 End-to-endモデルに基づくアイヌ語音声認識におけるクロスリンガル話者拡張敵対学習
3. 学会等名 日本音響学会研究発表会
4. 発表年 2019年

1. 発表者名 松浦孝平, 三村正人, 上乃聖, 坂井信輔, 河原達也.
2. 発表標題 End-to-endモデルに基づくアイヌ語の音声認識.
3. 学会等名 日本音響学会研究発表会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

【報道発表】

人工知能によるアイヌ語の自動音声認識・合成に成功（AINU語AI）
<https://www.kyoto-u.ac.jp/ja/research-news/2020-10-15-0>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	奥田 統己 (Okuda Osami) (60224151)	札幌学院大学・人文学部・教授 (30103)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------