

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）

京都大学大学院 情報学研究科

河原達也 松浦孝平 三村正人

概要

- アイヌ語の自動音声認識・合成を初めて実現
- 音声認識は、90%超の音素認識率を実現し、アーカイブ構築の効率化に貢献
- 音声合成は、故人の伝承者の音声と区別するのが困難なレベル

1. 背景

平成31年4月に「アイヌの人々の誇りが尊重される社会を実現するための施策の推進に関する法律」が成立し、令和元年9月には「アイヌ施策の総合的かつ効果的な推進を図るための基本的な方針」が閣議決定され、アイヌ文化の継承・振興が推進される中、今年7月には北海道白老町に国立アイヌ民族博物館を含む民族共生象徴空間（ウポポイ）の開業に至っている。アイヌ文化の多くが口頭で伝承されてきたが、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る。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、多くが未整備となっている。そのため、本研究グループは、文化庁「アイヌ語のアーカイブ作成支援事業」を活用してアイヌ語アーカイブの構築に取り組んでいるアイヌ民族博物館や平取町立二風谷アイヌ文化博物館と協力しながら、2017年からアイヌ語の自動音声認識・合成（AINU 語 AI）の研究開発に取り組んできた。

2. 音声認識

近年、音声認識はスマートスピーカや会議録作成・字幕付与など様々な用途で実用化されているが、実用的な認識精度を得るには非常に大規模なデータが必要であり、主要な言語に限られている。世界には5千以上の言語が存在するが、音声認識が実現されているのは百程度で、少資源言語への対応が大きな課題になっている。

本研究グループでは、前記の2つの博物館から提供頂いた計10名・約40時間の民話（ウエペケレ）の音声データを用いて、音声認識の単位・構成・学習法について様々な検討を行ってきた。その結果、音節（子音-母音または子音-母音-子音）を単位として用いることで、深層学習に基づく End-to-End モデル¹⁾により、94%の音素²⁾認識率・80%の単語³⁾認識率が実現できることを示した。これにより、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る。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。

ただし、未知の新しい話者に対して認識率が大きく低下するので、日本語モデルの転移学習⁴⁾に加えて、学習データをその話者らしい声質に変換して活用する枠組みを提案し、音素認識率90%を実現した。この研究については、9月の日本音響学会で報告済みであり（9月9日）、10月の国際会議 INTERSPEECH でも報告を行う予定である（10月28日）。

今後は、長時間の録音の中から、日本語の会話とアイヌ語で語っている区間を自動的に識別し、アイヌ語区間について書き起こしにとどまらず、話者や話題を同定する方法について取り組む予定である。

3. 音声合成

前記の音声データのうち1人当たり10時間以上ある話者について、深層学習に基づく End-to-End モデル¹⁾により音声合成を構築した。

9月16日に白老町で行われたアイヌ語アーカイブ研究会で実演したところ、博物館関係者も「本当に本人が話しているように聞こえた」と感想を述べられた。また博物館職員が、音源のない民話を口演する際の参考にもされた。著作権や肖像権などの問題から慎重に扱う必要があり、本物の録音と区別ができる仕組みを検討して欲しいという意見もあった。このような問題をクリアし、教材やナレーションの作成に活用されることが期待される。

4. まとめ

アイヌ語の音声認識・合成システムの構築は初めてのものであるが、アイヌ語を専門とする研究者やアイヌ語アーカイブに従事されている方からも高い評価を得ることができた。今後、アイヌ語アーカイブ構築の効率化により、アイヌ語に誰もが触れやすい環境の整備にとどまらず、アイヌ語の伝承・学習などの幅広い場面で応用されることが期待される。

謝辞： 本研究の実施にあたり、文化庁国語課から助言を頂いた。アイヌ民族博物館及び平取町立二風谷アイヌ文化博物館から音声データの提供と様々な議論を頂いた。札幌学院大の奥田統己教授には多くの指導を頂いた。科研費挑戦的研究（萌芽）「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」の支援で行った。

注釈：

- 1) End-to-End モデル：ニューラルネットワークにより音声から文字列（または文字列から音声）に直接変換するモデル
- 2) 音素：ローマ字書き起こしの1文字に相当
- 3) 単語：書き起こしにおいて空白で分かち書きされた単位
- 4) 転移学習：既存の大規模データで学習したモデルを別のタスクに転用する手法

参考文献：

- [1] K.Matsuura, S.Ueno, M.Mimura, S.Sakai, and T.Kawahara. Speech corpus of Ainu folklore and end-to-end speech recognition for Ainu language. In Proc. Int'l Conf. Language Resources & Evaluation (LREC), pp.2622--2628, 2020.
<http://www.sap.ist.i.kyoto-u.ac.jp/lab/bib/intl/MAT-LREC20.pdf>
- [2] 松浦孝平, 上乃聖, 三村正人, 坂井信輔, 河原達也. 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. 情報処理学会研究報告, SLP-130-16, 2019.
<http://www.sap.ist.i.kyoto-u.ac.jp/lab/bib/report/MAT-slp19-12.pdf>

- [3] 松浦孝平, 三村正人, 坂井信輔, 河原達也. 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. 日本音響学会研究発表会講演論文集, 1-2-8, 秋季 2020.
<http://www.sap.ist.i.kyoto-u.ac.jp/lab/bib/oral/MAT-asj20-9.pdf>
- [4] K.Matsuura, M.Mimura, S.Sakai, and T.Kawahara. Generative adversarial training data adaptation for very low-resource automatic speech recognition. In Proc. INTERSPEECH, (発表予定), 2020.