

音声自動認識による字幕情報保障トライアル (2)

秋田 祐哉¹ 塩野目 剛亮² 白石 優旗³ 平賀 瑠美³

概要: 様々な場面で音声自動認識(音声認識)を用いた字幕情報保障への期待が高まっているが、学会などにおける研究発表では実際の運用例は少なく、利用者の視点からの字幕品質の検討も十分に行われていない。そこでアクセシビリティ研究会では、第1回研究会に引き続き、健聴者の研究発表とそれに対する質疑で音声認識の結果を編集なしに即時字幕化する、字幕情報保障トライアルを行う。この形式は、研究会や全国大会で音声認識による字幕情報保障を行う場合を模したものであり、このような字幕表示が実際の研究発表の場面で有用なのかどうかを研究会参加者と考えることが本トライアルの目的である。

Captioning by Speech Recognition for Research Presentation: Second Trial

YUYA AKITA¹ TAKEAKI SHIONOME² YUHKI SHIRAISHI³ RUMI HIRAGA³

Abstract: Automatic speech recognition (ASR) is considered promising as an efficient tool to make captions for various kinds of speeches. However, ASR-based captioning for academic presentations has been rarely carried out so far, and the quality of resulting captions has not sufficiently been examined by real users. Thus, following the previous AAC meeting, we again conduct a trial of ASR-based real-time captioning for research presentations and discussions, where ASR results are presented promptly without any editing. This format simulates a realistic way of support in actual academic meetings, and we hope every participant takes part in discussion on the feasibility of ASR-based real-time captioning in this style.

1. はじめに

2016年に障害者差別解消法^{*1}が施行され、障害による社会的障壁除去のための「必要かつ合理的な配慮」が行政機関や事業者等の努力義務(国及び地方公共団体は義務)となっている。情報処理学会アクセシビリティ研究会も、合理的配慮をできるだけ行いたいと考えている。

このうち聴覚障害に対する配慮としては、音に代わる情報伝達の手段として字幕や手話が用いられているが、これらのスキルを持つ人的資源は限られており、実際に実施できる場面は多くない。これに対して、音声の自動認識(音声認識)を用いてコミュニケーションの支援を行う取り組

みが提案されており、たとえば「こえとら」^{*2}や「UDトーク」^{*3}[1]といったアプリケーションが開発されている。

アクセシビリティ研究会でも、音声認識を用いて研究発表にリアルタイムの字幕を付与することを検討している。学会講演や大学講義などの学術的な発表に対して、講師の音声認識することによりリアルタイムに字幕を付与する試みはいくつか行われている[2]。これらは音声認識を含む要素技術の研究・開発の視点から行われていることが多く、また実際の場面での実施例も少ないことから、字幕の利用者側の視点を十分に取り込んでいるとはいえない。聴覚障害者の観点からの音声認識字幕の分析も報告されているが[3]、音声の復唱入力と音声認識出力の編集に基づいており、本稿のリアルタイム字幕とは前提が異なる。

そこでアクセシビリティ研究会では、情報処理技術をアクセシビリティの現場に生かすという立場から、音声認識

¹ 京都大学
Kyoto University

² 東北文化学園大学
Tohoku Bunka Gakuen University

³ 筑波技術大学
Tsukuba University of Technology

^{*1} <http://www8.cao.go.jp/shougai/suishin/sabekai.html>

^{*2} <http://www.koetra.jp/>

^{*3} <http://udtalk.jp/>

による研究発表のリアルタイム字幕付与を実際に研究会で実施し、その現状を研究会参加者と検証して、問題点や期待を洗い出そうとするトライアルを第1回研究会（2016年7月）に実施した[4]。第2回（今回）も、この趣旨に則り、音声認識による字幕付与トライアルを実施する。第1回では大学で開発されたシステムを利用したのに対して、第2回では企業で開発された複数のシステムを用いる。

なお、本稿における字幕とは事前に音声内容を得て字幕をつけるものではなく、リアルタイムで付与を行う字幕を指し、要約筆記とほぼ同じ意味で用いる。また、聴覚障害者の情報保障について、字幕情報保障に限ったものと考えられるものではまったくない。

2. トライアル

トライアルでは、音声認識による字幕表示システムを用いて研究発表と質疑応答の字幕付与を行う。現実的な運用を想定して、システムに対する読み上げ原稿の提供は行わず、認識結果の編集作業も行わない。今回のトライアルでは2件の講演を実施し、それぞれ別のシステムにより字幕を付与する。以下にそれぞれの講演のあらましを述べる。

2.1 その1（塩野目）

発表内容は、2016年6月7日～9日にかけて福岡県で開催された「人工知能学会（JSAI 2016）」で著者の一人が発表したものと同じである。タイトルは「クラウドソーシングによるパラリンピック競技実況の情報保障に関する一考察」[5]であり、発表内容を以下に簡単に記す。

我々の研究は、パラリンピックの試合会場において、クラウドソーシングによる試合実況の字幕情報保障を実現することで、より多くの障害者、健常者が試合を観戦しに行き、興奮と感動に満ちた時間と場の共有を実現することを最終目的とする。クラウドソーシングによる情報保障は、多数の事前登録者により行われ、被支援者と支援者の境界が非常に柔軟であるところが、従来の少数の専門家による字幕情報保障と異なる。聴覚障害者は勿論であるが、言語を異にする健常者も、視覚障害者も被支援者となり、文字入力を出来る者ならば字幕作成を行う支援者となり得る（図1）。ある場面を想定した図1について簡単に説明すると、左側に多様な特性を持つ人たちが集まっており、その中には情報を得られない人がいたり、何らかの支援が可能なたちもいる。そこで、クラウドソーシングの仕組みを使って、支援者の集まりをつくり、不特定多数が情報保障を担当する。すると、図の右のように、その場面で助け合い、楽しめる場が作られることになる。本検討では、障害者スポーツを1つの適用の現場として想定し、スポーツ実況に関する文献調査と競技団体への問い合わせを行ない、クラウドソーシングによる情報保障実施のための課題について考察している。

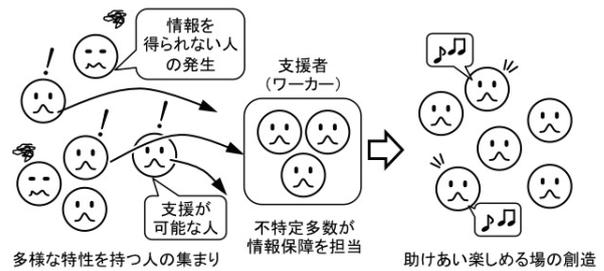


図1 クラウドソーシングによる情報保障

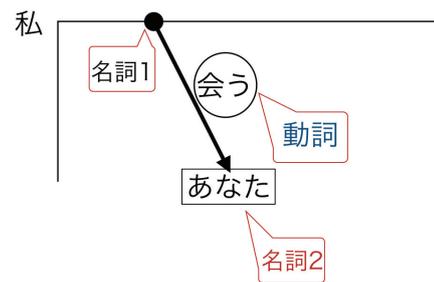


図2 提案書記法の例

2.2 その2（白石）

発表内容は、2015年3月2日～4日にかけて福島県で開催された「第7回データ工学と情報マネジメントに関するフォーラム（DEIM 2015）」で著者の一人が発表したものと同じである。タイトルは「手話言語の空間的特徴を考慮した書記法の提案」[6]であり、発表内容を以下に簡潔に記す。

手話言語は視覚言語の一種であり、日本語や英語等の音声言語とは異なる独自の語彙や文法体系を持つ言語とみなされている。しかし、手話言語は専ら対面での会話に用いられており、日常生活で手話言語をそのまま記述しコミュニケーションすることはない。そこで、本研究では、手話の「文法」、特にその空間的特徴を反映した、常用に耐えうる書記法の提案を行う（図2）。その際、手話使用者だけでなく日本語使用者も容易に使用可能となる書記法を開発することで、手話使用者と日本語使用者の円滑なコミュニケーション支援を可能とすることを目指す。更に、音声言語とは異なる発想に基づく書記法を用いることで、情報・知識の新たな表現・思考ツールにもなり得ると考える。本発表では、提案書記法の概略、並びに、アンケート調査による評価実験について紹介する。

3. おわりに

本トライアルでは、2件の講演の後に議論を行う。研究会の各参加者には、音声認識を利用したリアルタイム字幕について、利用者の立場からの忌憚のない意見や議論をお願いしたい。

謝辞 本トライアルの音声認識字幕システムは、株式会社東芝・株式会社富士通ソーシャルサイエンスラボラトリ (SSL) よりご提供をいただきます。布日光生様 (東芝)・福岡寿和様 (富士通 SSL) をはじめ、関係の皆様へ深くお礼申し上げます。また、本トライアルについては京都大学大学院情報学研究科教授 河原達也先生にご支援とご助言をいただいております。音声認識の研究の一部は科学研究費補助金 16H02847 による。

参考文献

- [1] 青木秀仁, コミュニケーションのユニバーサルデザイン ~ 音声認識技術の活用~, 日本音響学会秋季研究発表会講演論文集, 2-6-12, 2016.
- [2] 河原達也, ICT・音声認識の活用による講演・講義の字幕付与, 情報処理, Vol.56, No.6, pp.543-546, 2015.
- [3] 中野聡子, 金澤貴之, 牧原功, 黒木速人, 上田一貴, 井野秀一, 伊福部達, 音声認識技術を利用した字幕表示システムの活用に関する研究—聴覚障害者のニーズに即した呈示方法—, メディア教育研究, Vol.5, No.2, pp.63-72, 2008.
- [4] 平賀瑠美, 秋田祐哉, 音声自動認識による字幕情報保障トライアル, 情報処理学会研究報告, AAC-1-6, 2016.
- [5] 塩野目剛亮, 張建偉, 白石優旗, 平賀瑠美, クラウドソーシングによるパラリンピック競技実況の情報保障に関する一考察, 2016 年度人工知能学会全国大会, JSAI2016, 4L4-2-177, 2016.
- [6] 白石優旗, 田村誠志, 手話言語の空間的特徴を考慮した書記法の提案, 第7回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015), F4-1, 2015.