

クラウドソーシングによるライブ手話文字通訳に関する研究

若月大輔¹⁾, 田中康平²⁾, 皆川洋喜¹⁾,

筑波技術大学 産業技術学部¹⁾, 大学院技術科学研究科²⁾

キーワード: 手話文字通訳, 聴覚障害, 情報保障, クラウドソーシング, ライブ動画配信

1. 背景と目的

聴覚障害者のコミュニケーションサポートとして、手話通訳や文字通訳（ノートテイクや要約筆記など）が、聴覚情報を補助する情報保障に利用されている。例えば、発話者が話す場合は、その音声を手話や文字に変換して聴覚障害者に提供する。一方、話者が手話を使う場合は、その手話を音声や文字に変換して提供することが求められる。手話による講演を手話がわからない聴覚障害者や健聴者に伝える情報保障として、手話通訳者が読取通訳で手話を音声にし、文字通訳者が音声や文字に変換する方法がとられることが多い。しかし、現状の手話を文字へと直接通訳する方法（以下、手話文字通訳）は、手話から音声、音声から文字の2段階の変換をとるため、表出されるまでの遅延が著しく増加する。また、専門技術をもった手話通訳者と文字通訳者の双方が必要になるため、金銭的コストも高い。

そこで、我々は、より効率のよい手話文字通訳を実現するために、クラウドソーシングによる手話文字通訳について検討を進めている。不特定多数の手話がわかる者をクラウドソーシングのワーカーとして、手話のライブ映像を分割した短い時間の手話文字通訳をおこなってもらい、それらを統合して字幕化することで手話文字通訳を実現する。これは、手話を直接的に文字化するだけでなく、手話を使う聴覚障害者が自ら手話がわからない人を助けることができる機会を提供することにつながる。

これまでに、ライブの手話映像を短い区間で区切り（以下、通訳区間）、それを手話が読めるワーカーに配信して文字化をしてもらい、文字化された結果を統合してライブ手話文字通訳をおこなうシステムを試作した[1]。予備的な実験を通して、通訳区間の中で文の欠落や重複が発生すること、わからない手話表現の文字化ができないという課題が明らかになった[2]。本報告では、前者の課題を解決するために、ワーカー間で関係を促す入力インターフェースの工夫について研究を進めた結果について述べる。

2. 手話文字通訳システム

2.1 システムの概要

手話話者の映像をライブ動画としてワーカーに配信して、ウェブベースで手話文字通訳をおこなうシステムを構築した（図1）。話者を撮影したライブ手話動画は一定時間の通訳区間に分けられ、文字化を担当するワーカーに順に割り当てられる。ワーカーが担当の通訳区間を文字化した文を統合して、字幕として利用者に提供する。これらの一連のコントロールをするサーバをウェブベースで実装した。

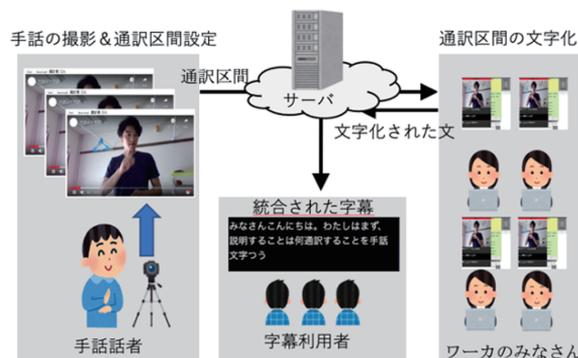


図1 システムの概要

2.2 ワーカーの入力インターフェース

ワーカーの入力インターフェースを図2に示す。ウェブブラウザでアクセスすることで開くことができる。手話話者のライブ動画が常に表示されており、割り当てられた通訳区間になったら、タスク指示のところに手話を文字化する旨のメッセージが表示される。また、プログレスバーで通訳区間の動画再生位置も示した。ワーカーは自分の通訳区間において入力領域に文を入力して手話文字通訳をする。自分の入力領域の上下には、前と次の通訳区間のワーカーの入力状況がリアルタイムに確認できるように工夫した。これによって、ワーカー間で通訳区間の間の文の欠落や重複を避ける関係が期待できる。

ワーカが担当の通訳区間を文字化する際には、動画の再生速度を下げ（標準では0.8倍速）、手話を読み取りやすくした。この再生速度を下げることや、文の入力作業によって動画の再生位置が遅れてしまうため、通訳を終了したら再生速度を上げて（1.5倍速で固定）、現在の再生位置に追いつかせる実装をした。なお、通訳区間の再生速度については、ワーカが自分で設定を変更できるようにした。

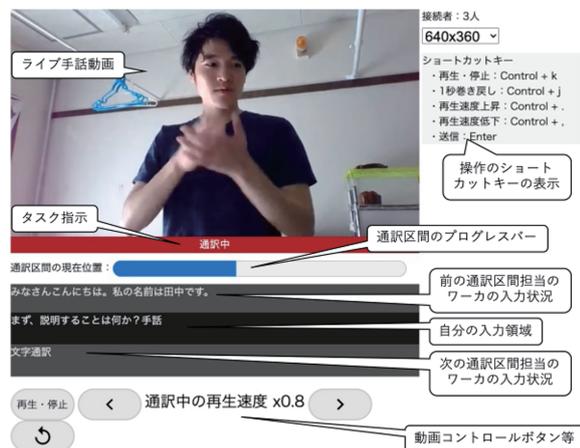


図2 ワーカの入力インタフェース

3. 実験と結果

3.1 実験の概要

本研究で構築したシステムの手話文字通訳の評価するために実験をおこなった。実験は5回実施し、各回8名程度の実験参加者（以後、ワーカ）で、それぞれ異なる手話動画に対して手話文字通訳をしてもらった。ワーカは聴覚障害がある大学生で、手話歴の平均は10.9年（SD7.1年）であった。手話動画は実験参加者以外の5名に依頼し、日本語対応手話で約7分30秒の動画を作成してもらった。実験時にワーカが入力インタフェースを操作したログを収集しておき、分析をおこなう。また、各回の実験終了時にはアンケートも実施した。

通訳区間の長さは6秒に設定した。また、手話文字通訳の欠落を防ぐため、各区間を1.5秒間重複させる設定にした。したがって、約7分30秒の動画が6-1.5=4.5秒の区間で区切られるため、1つの動画あたり約100個の通訳区間が設定された。通訳区間の標準の再生速度を0.8倍速（ワーカが変更可能）とし、動画の再生位置の遅れを取り戻す際の再生速度を1.5倍速（変更不可）にした。

3.2 結果

全5回の実験における通訳区間の間の文の欠落率の平均は、19.2%（SD 7.1%, MAX 26.3%, MIN 9.1%）であった。重複率の平均は6.1%（SD 2.3, MAX 10.1%, MIN 4.0%）であった。

アンケートについて、「通訳中に前後の入力文から欠落や重複があるか確認しましたか」では、「毎回確認した」の回答割合が約79%であった。また、「通訳は楽しかったですか」については、5段階評価（1:ネガティブ～5:ポジティブ）で回答してもらった結果の平均は、4.4（SD0.8）であった。自由記述では「助けられる側ではなく、助ける側に回れたことが心嬉しく感じる。」「通訳は好きであるため」などの回答があった。

4. 考察

通訳区間の中で文の欠落や重複が発生する課題を解決するために、前後のワーカが互いに入力文を確認して連携できる方法を導入した。その結果、予備的な実験では欠落率と重複率がともに33%であった[2]が、今回の実験では欠落率の平均が19.2%、重複率の平均が6.1%になり、特に重複については大幅な改善がみられた。アンケートの回答でも「毎回確認した」ワーカが多いことも明らかになっており、うまく関係できていたことが示唆された。欠落には入力が追いつかない他に、わからない単語に起因するものも見られたので、今後は、文の欠落と重複を含めた文の誤りを修正するタスクを設け、修正タスクもクラウドソーシングでおこなう方法を検討したい。

通訳の楽しさに関するアンケートでは、ほとんどのワーカが楽しさを感じていた。コメントから、手話文字通訳そのものへの楽しさ、通訳で他者を助ける側に回れたことへの喜びが背景にあることが示唆された。ボランティアベースでの手話文字通訳が想定された場合は、タスクの楽しさは特に重要な観点であると考えられる。今後は通訳タスクと楽しさの関連性を分析し、より楽しさや喜びを感じられるシステムの実現を目指したい。

5. まとめ

本報告では、ライブの手話映像を短い通訳区間に区切り、手話を読める人を対象としたクラウドソーシングで、手話文字通訳を実現できるシステムの改善について延べた。特に、前後の通訳区間を担当するワーカが相互に入力状況を確認できるようにし、文の欠落や重複を連係して防ぐことができるインタフェースを実装した。

実験の結果、実験参加者としてのワーカは実装したインタフェースを活用して、文の欠落率を下げることで、重複率については大幅に改善することができた。また、手話文字通訳のタスクについても、通訳作業の楽しさや、他者を支援できる喜びを感じていることがわかった。

今後は、手話文字通訳の質を高めるための、修正タスクの検討と、より楽しさや喜びが感じられる手話文字通訳システムの実現を目指した研究に取り組んでいきたい。

参考文献

- [1] 田中康平, 若月大輔, 皆川洋喜, クラウドソーシングを用いた手話文字通訳～ライブ動画配信を活用したシステムの試作～, 信学技報, vol. 119, no. 322, WIT2019-39, pp. 95-100, 2019年12月.
- [2] Kohei Tanaka, Daisuke Wakatsuki, Hiroki Minagawa, A Study Examining a Real-Time Sign Language-to-Text Interpretation System Using Crowdsourcing, Computers Helping People with Special Needs (ICCHP 2020), Part II, pp.186-194, 2020