

クラウドソーシングによる字幕情報保障に関する基礎的検討 —ウェブベース遠隔文字通訳システム captiOnline スマートフォン版の開発—

塩野目剛亮^{†1} 若月大輔^{†2} 白石優旗^{†2} 張建偉^{†2} 森嶋厚行^{†3} 平賀瑠美^{†2}

概要：聴覚障害者の情報保障を行なうために、音声を文字に変換して表示する文字通訳字幕が活用されている。近年では、PCを用いた2人1組での連係入力による字幕情報保障が多く活用されているが、その業務を担うことができるPC文字通訳者の養成やコーディネートには多くのコストを必要とする。本検討では、開発中のスマートフォン向け字幕情報保障プラットフォームを用いて、クラウドソーシングによる字幕情報保障を想定し、2人、および7人の特別な入力訓練経験を持たない人たちが暗黙に協力した場面でのスポーツ実況音声のリアルタイム字幕入力を実験的に行っている。専門家による文字起こしを行なった字幕と、クラウドソーシングによって生成された字幕とを比較し、スマートフォンを用いた文字通訳の可能性について考察する。

キーワード：クラウドソーシング、遠隔文字通訳、情報保障、聴覚障害、スマートフォン

A Basic Study on Speech-to-Text Interpretation via Crowdsourcing — Development of Web based Remote Speech-to-Text Interpretation System “captiOnline” —

TAKEAKI SHIONOME^{†1} DAISUKE WAKATSUKI^{†2} YUHKI SHIRAISHI^{†2}
ZHANG JIANWEI^{†2} ATSUYUKI MORISHIMA^{†3} RUMI HIRAGA^{†2}

1. はじめに

ISeeプロジェクト[1]では、「オープンな、誰もが誰かの助けになる情報保障」のコンセプトを掲げ、情報保障技術にクラウドソーシング技術を応用し、健常者も障害者も関係なく、各人の得意なことを活かして、支援し合う環境の構築を目指している。

著者らは、この技術の適用の場として、2020年の東京オリンピック、パラリンピックの競技会場を想定している。多様な特性を持つ人たちが集まる中で、支援を必要としている人と支援が可能なお互いが、ともに楽しめる場の創造が可能となる。すなわち、不特定多数が支援者（ワーカー）として情報保障を担当することによって、互いに助け合い、楽しめる場を創造する。

実際に、クラウドソーシングを活用して、複数の非熟練者で文字化を行う方法も提案されている[2]。たとえば、張ら[3]、および白石ら[4]の研究では、情報源（手話）をタスクの単位に分割したクラウドソーシングによる通訳システムの実現可能性を提示している。こういった技術は日本語の手話化や、手話の日本語化だけでなく、書記日本語の音声化、多国語の通訳にも利用が期待される。

先行研究[5]では、文献調査と障害者スポーツ競技団体へのアンケート調査を通して、スポーツ実況のクラウドソー

シングによる情報保障実施のための課題について検討している。

本検討では、クラウドソーシングによる字幕情報保障の可能性を検討するため、スマートフォン版 captiOnline を用いて、スポーツ実況音声の文字化を行なっている。2名、および7名を対象とした字幕入力実験の結果から、特別な訓練を積んでいない人でも、使い慣れた機器を用いて、わずかな練習で支援に携われることがわかる。

2. captiOnline 概要

captiOnline[6]は遠隔情報保障システムのうち、遠隔で話者の音声をPCで文字化するPC文字通訳を、ウェブ上で行なうことができるシステムである。

2.1 captiOnline の機能

captiOnline は、文字通訳者が使用する入力ページ（図1参照）と、入力された字幕を見るための閲覧ページとに分かれている。

また、連係文字通訳を円滑に進めるため、以下の機能を備えている。

- ① メインバー：サーバとの接続状況、接続者数の確認、およびページ全体の外観設定を行なう。
- ② 表示パネル：送出された字幕を表示する。
- ③ 入力パネル：文字を入力し、送出する。
- ④ 連係パネル：同時に接続している他者の入力文をリアルタイムで表示する。
- ⑤ 連絡パネル：通訳者間での連絡に用いる。
- ⑥ 交代タイマパネル：通訳者の交代時間、交代メッセージを自動で通知する。

^{†1} 東北文化学園大学
Tohoku Bunka Gakuen University
^{†2} 筑波技術大学
Tsukuba University of Technology
^{†3} 筑波大学
University of Tsukuba

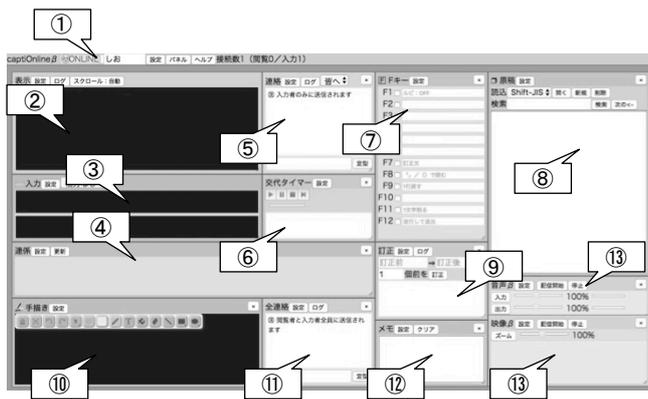


図1 captiOnlineの画面構成

- ⑦ ファンクションキーパネル: ファンクションキーで呼び出す定型文字列の設定をする。
- ⑧ 原稿パネル: テキスト原稿を読み込ませ、文を送出する。
- ⑨ 訂正パネル: 既に出した文中の単語などを訂正する。
- ⑩ 手描きパネル: 図や数式を含めた通訳を行なうために、手描きした画像を送出する。
- ⑪ 全連絡パネル: サーバに接続している全クライアントにメッセージを送る。利用者と通訳者間の連絡に使用する。
- ⑫ メモパネル: 通訳内容のメモなどに利用する。
- ⑬ 音声パネル: 他のクライアントに音声を配信、受信した音声を再生する。
- ⑭ 映像パネル: 他のクライアントに映像を配信、受信した映像を表示する。

以上のパネルのうち、閲覧ページでは主に表示パネルと設定変更のためのメンバーが表示・利用される。

2.2 スマートフォン版 (SP版) の構成

SP版では、画面の大きさや画面構成の制約、および非専門家である不特定多数の入力者を想定し、文字通訳者が特に必要とするパネルを除き、シンプルな構成としている(図2)。公開チャット(全連絡に近いが、メッセージが字幕表示部に表示される)、音声配信パネルも用意されているが、本検討での実験では使用していない。

また、通常版ではファンクションキーで実装されていた機能をSP版ではツールボタンで実装している(図3)。

3. 実験

ここでは、2名、および7名を支援者としたサッカー実況音声の文字化実験について述べる。

3.1 実験参加者

実験には、19歳から21歳の大学生9名(男性5名、女性4名、平均年齢20.2歳)が参加した。スマートフォンの使用歴は2年から5年で、平均して3.3年程度であった。うち、6名がフリック入力、3名がガラケー打ち(トグル入

a 閲覧ページでは他に関係パネル、全連絡パネル、ビデオチャットパネル、音声・映像パネルが利用可能となっている。



図2 スマートフォン版の入力ページ (SP版入力ページ) (iPod Touchでの表示例)

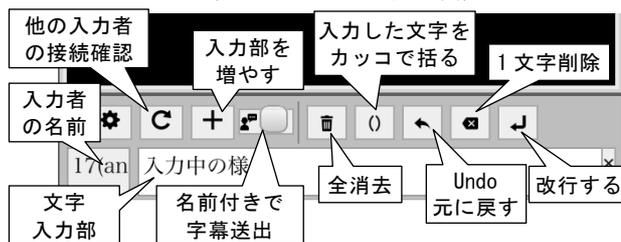


図3 SP版のツールボタン

力)を使用していた。いずれの参加者も文字通訳の経験はなかった。

3.2 実験の流れ

- ・ 実験概要、および captiOnline の説明

はじめに、クラウドソーシング、字幕情報保障に関する簡単な説明を行った。各参加者が持ち寄ったスマートフォンで captiOnline のサイトにアクセスし、接続が可能であることを確認した。

実験で使用する実況音声の一部を3分程度、事前に作成した文字起こし文を captiOnline の閲覧ページでプロジェクタで提示しながら聞いてもらい、字幕情報保障の様子を見てもらった(図4参照)。

- ・ 文字入力速度の測定

3つの日本語例文(付録A.1)について、メモ帳やメール画面等の任意のアプリを使って、紙で提示した文章を見ながら文字入力をしてもらい、入力時間を秒単位で計測した(表1)。その際、漢字の変換の違いや句読点の抜けなどは気にせず、最後まで入力するよう指示した。

このようなモバイル機器の文字入力の使いやすさについて、大塚らの検討[7]がある。スマートフォンに相当する機器として iPod Touch のフリック入力では歌を入力する時間を計測している。47文字の「いろは」、および「逆いろは」では平均119.5秒、および120.9秒かかっている。ガラケー入力(トグル入力、マルチタップ入力)では、それぞれ平均60.4秒、71.8秒となっている。

大塚らの検討が5年ほど前のもので、入力する文章も特殊なものであることから単純な比較は難しいが、本検討での測定ではフリック入力の参加者の方が、入力速度が速くなっている。

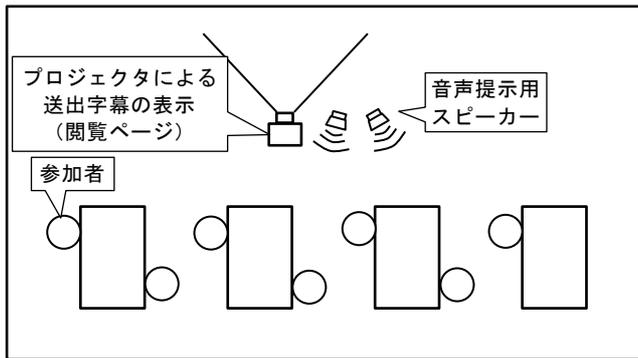


図4 実験状況 (7人入力時)

表1 例文入力時間, および平均入力速度

参加者	入力方式	例文1 (秒)	例文2 (秒)	例文3 (秒)	平均速度 (文字/秒)
S1	F	102	117	129	1.014
S2	G	196	180	198	0.613
S3	F	84	84	84	1.397
S4	F	94	83	73	1.433
S5	F	62	76	76	1.653
S6	G	125	125	147	0.887
S7	F	74	80	73	1.556
S8	F	121	102	122	1.024
S9	G	132	130	135	0.886
	平均	110.0	108.6	115.2	1.163

*入力方式はF:フリック入力, G:ガラケー打ちを示す.

・ 練習

5分程度の実況音声聞きながら, 参加者全員で字幕を入力してもらった. 参加者には, (1) 聞こえた音声をそのまま打っても, 要約した文章を打ってもかまいません, (2) 読んだ人が試合の状況がわかるような字幕を作ってください, (3) 他の入力者の様子も参考にしながら, 協力してみてください, の3つを教示した.

・ 本番

10分程度の実況音声聞きながら, 参加者全員で字幕を入力してもらった. 練習時と同様の教示を与えた. 練習時, 本番時ともに, 送出された字幕はプロジェクタで実験室の前方に提示している.

なお, 練習部分の音声の文字起こしは約2500文字, 本番部分は約5500文字となっている.

・ 感想の記入

実験に参加した感想や, 研究に関するコメントを自由記述形式で記入してもらった.

3.3 実験に使用した音声

実験には, 昨年行われたサッカー日本代表の国際試合のうち, 前半部分の音声を使用した. 実況音声は専門家による文字起こしを行なっている. 音声の長さは約49分(2940秒), 文字数は約24000文字^bである. 1秒あたりの文字数

^b 段落や句読点も1文字としてカウントしている.

は8.163文字/秒となる. 以下にその一部を示す. 文字起こしの際は, 話者の音声を忠実に入力すること, 感嘆符や疑問符, 三点リーダーは入力者の判断に委ねることとした.

【実況音声の文字起こし (□□, ○○, △△は発言者)】

□□/吉田.
 全員で押し込んでますが, シュートまで行ってほしいところ。
 香川!
 日本のコーナーキック, いや判定はゴールキックです。
 ○○/そうですね, 向こうの選手, 触らなかったですね。
 ただやっぱり, あのペナルティエリアの, 相手のペナルティエリアの, そのゴールの近いところでボールをね, 自分たち, この支配してますから。
 □□/はい, ええ。
 ○○/この時間帯はね, 点取れる, そのリズムだと思いますよ。
 □□/△△さん, 本当にこういった時間帯を大事にしたいですし, あとは日本の場合は最近押し込んでいても奪われての相手の速攻, カウンターということから1つ失点してますよね。

4. 結果と考察

4.1 字幕入力結果

ここでは, 実際に入力, 送出された字幕の一部を示す.

(1) 2人入力

【練習から一部抜粋】

ボールを出せばチャンス浅野今回がんばる
 ゲームにかける!
 シュートまで
 ゴールキック
 点を取るリズム
 この時間帯は点が取れる
 ボールをどんどん動かす
 サイドを使わず縦を
 (全160文字)

【本番】

実況/日本はサイドを使うことで
 実況/シュートだ
 状況/いいですね!
 状況/このあたりでボールが…
 状況/積極的にどんどんシュートを
 実況/長谷部から香川へ
 状況/スペースもあります
 実況/左サイドいける
 状況/上手く使って左サイド
 実況/原口がまたいく
 状況/はるかにボールが動いている
 状況/長谷部キャプテンのファウル
 実況/香川を監督が呼び止める
 (全490文字)

(2) 7人入力

【練習から一部抜粋】

日本の方が前線にプッシュャーをかけるという5分、10分にして欲しいですね。

〇〇/どんどんまえにうごいてほしいです。

△△/すでにうらに出してる。そういう所に監督は期待。

□□/本田のクロスの幻のゴール！

どんどんまえのことはわすれていってほしい

□□/いいですね。

前のゲームは忘れて、今日のゲームに全てをかける。

今のプレーは日本のプレー。大観衆のなかで日本の方がいいプレー坂井さかいひろき！吉田よしだ押し込んでる判定はゴールキック判定はゴールキックです

さかい！！相手のペナルティエリアを支配しているゴールに近いところでせめてる、全員で押し込んでますが、ゴールまでもっていきたいところ。奪われてもカウンターで日本は奪われてもカウンターに。

(全647文字、名前付き送送出を利用している参加者がいる)

【本番から一部抜粋】

グラウンドが荒れている雨の影響でグラウンドがあれてるんですけど砂がめくれているピッチはスコールの影響であれている。芝がめくらは。選手が踏み込んだだけでも。注意して欲しい。足への負担もある。日本、パスワークもありますがこういうところをつけてやってほしいですね。足への負担がいつもよりある両チームにとって大変では？あの長谷部が公式練習で、体を冷やす面ではいい。雨雨を恵みの雨へまさに！大切な予選！プレー再開プレー再開原口

日本は左サイドに雨を日本にとって恵みの雨としてほしいですね。

(中略)

素晴らしいタイミング本田とのタイミングいい本田とびこんだ！4ほんめ本田踏みこんだ原口からこの前半4本目のコーナーキックです。よく戻った本田！本田にとおる！本田、飛び込んだ！うて！！

(全2021文字)

4.2 自由記述

実験後に記入してもらった自由記述(一部抜粋)を示す。

- ・ 相談した後に練習が必要。予測変換機能をもう少し活かす。出場選手のリストがほしい。音声だけでなく、音声認識ソフトを要約したほうがやりやすい。
- ・ 実況、状況の役割を決め実験に臨んだが、果たしてその役割に沿ってできているのかわからなくなり、また、同時に速かったこともあり難しかった。
- ・ 初体験でかなり大変な実験だった。
- ・ 完璧な文じゃなくてもいいということ。周囲の人達と合わさって伝われば良いことがわかると、やりやすくて良い。ただ、実際にあの文章そのままではわかりにくい気がしたので、そこが気になる。
- ・ 音声を聞くのと同時にというのができないので、音声を聞いてから打っていると、どんどんと打たなければならないことを忘れてしまうので難しかったです。打つのが遅れてしまうと、出来事はどんどん過ぎていってしまう

ものなので、今起きたことをすぐに届けることもできないと思いました。今日は他にも実験に参加している人たちがたくさん私が字幕に出来なかった部分を補ってくれました。

- ・ この研究はとても素晴らしいものであると思った。聴覚障害者の方はサッカーの試合を観戦してもテレビから聞こえてくる実況がわからないので、今回のようなことが普及すれば良いと感じた。
- ・ 聞いたことを入力しつつ、文を覚えるということができなかった。周りの状況もわかるので、入力している部分と同じだと遠慮してしまう。
- ・ オリンピックだけじゃなく災害時の迅速な情報提供に使えるかと思います。他の人が自分が打とうとしている部分が被ってしまうことが多かった。他人がどこを打つかを決めることはできないでしょうか。変換が賢い携帯だともっと早く打てるかもしれません。
- ・ はじめて打ち込んだということもありますが、他の人が打っている文章とかぶって打ってしまったたり、文章が繋がらなくなってしまうところがあって難しいと感じました。

4.3 考察

2人入力の練習時では、入力文字数が少なかったが、本番前に参加者同士が話し合い、「実況」と「状況」とに名前を変更し、“名前付きで字幕送送”の機能を用いて役割分担を試みている。

7人入力の練習、および本番では、「グラウンドが荒れている」「グラウンドがあれてる」、「プレー再開プレー再開」「雨雨を恵みの雨…恵みの雨…」のように、複数人が同じ内容を入力していることがわかる。すなわち、多くのワーカールから入力された単語・文章の集約・統合も必要であると考えられる。同様に、ゴール前でのプレーでは、「本田」の選手名が多く送送送されている。文献[8]で指摘されているように、文字起こしされた字幕からは、多くの選手名のみでの発話、および文の送送が見られる。今回使用したSP版の画面構成では、キーボード部の表示が画面の半分程度を覆ってしまい、十分に他者の入力の様子が見えないため、重複した入力が増加してしまったと考えられる。

2人入力と7人入力とで送送送文字数を比較すると、2人練習時には160文字であったが、本番では480文字となっている。7人練習時には647文字であったが、本番では2021文字となっている。練習では約5分、本番では約10分の長さの音声であるが、本番では送送送された文字数が大幅に増加している。すなわち、実践を繰り返していくにしたがって、音声の聞き取り・入力や、他者との協力の仕方を理解し、熟達していく可能性が示唆されている。さらに、スマートフォンの予測変換・学習機能によって、過去に入力した単語が履歴として残ることによって、入力効率は向上していくことが期待できる。文献[7]の実験に比べて、入力速

度が高くなっていることは、一般的な単語の多さや予測変換機能の充実、スマートフォン操作の十分な熟達などが要因であると考えられる。

自由記述からは、「出場選手のリストがほしい」との記述が見られ、事前情報が必要であることがわかる。このことは、手話通訳者や文字通訳者に講演のスライドや原稿などの資料を提供して、事前準備をしたうえで情報保障を行なってもらうように、通常の情報保障の現場と同様なニーズがあることを示している。

音声認識を活用した復唱による入力方法も可能とすれば、肢体不自由者や視覚障害者も復唱者としてクラウドソーシングのワーカーとなりうること[5]を考慮すると、参加者のコメントにあるような音声認識文章からの要約字幕入力も現実的なものであるといえる。

さらに、複数の参加者から、「周りの状況がわかる」「周囲の人達と合わさって…」「他の参加者が補ってくれた」といった感想が出ていることから、協力しあうことに楽しさを感じてもらえていることがうかがえる。これまで、字幕情報保障に携わった経験のない人たちにとっても、使い慣れた機器を通して他者とつながりながら、必要とされる支援を提供することに意義を見出してもらえたと考える。

5. おわりに

本検討では、クラウドソーシングによる字幕情報保障を想定し、スマートフォン版 captiOnline を使った字幕入力実験を行なった。実験を通して、2人入力の場合は話し合いによる明確な協力が見られ、7人入力の場合は暗黙の協力がなされていることがわかった。また、7人入力の際は他者の入力の様子を見られず、多くの文字送出手が重複することがわかり、情報の統合が必要であることがわかった。さらに、特別な訓練経験のない人たちによる実験を通してクラウドソーシングによる情報保障の可能性を確認できた。

今後の課題としては、各人の入力・字幕送出手のログの解析によって、個人ごとの入力の癖や入力速度の変動を調べることで、ワーカーのグループ形成最適化のための指標を作ることがあげられる。また、人数を2人、4人、8人、16人…と増やし、経時的に観察する実験を行なうことで、どのように協力体制を作っていくか、各人のモチベーションがどのように変動するかを明らかにしたい。

謝辞 実験に参加いただいた東北文化学園大学の学生諸氏に感謝します。なお、本研究は科研費(15K01056)、国立情報学研究所、ならびに筑波技術大学の支援を受けています。

参考文献

[1] 平賀瑠美, 白石優旗, 張建偉, 若月大輔, 塩野目剛亮, 熊井克仁, 橋本大空, 森嶋厚行. “ISsee プロジェクト: オープンな, 誰もが誰かの助けになる情報保障”. <https://crowd4u.org/ja/projects/ISsee>, (参照 2017-02-01).

[2] W.S. Lasecki, C.D. Miller, A. Sadilek, A. Abumoussa, D. Borrello, R. Kushalinagar, and J.P. Bigham, “Real-Time Captioning by Groups of Non Experts,” In Proceedings of the 25th annual ACM symposium on User interface software and technology, 2012, p.23–34.

[3] 張建偉, 白石優旗, 櫻井恵美, 森嶋厚行. “クラウドソーシングによる聴覚障害者の情報保障手法の検討”. 情報処理学会第77回全国大会, 2015, 6B-01, p.1-467-468.

[4] 白石優旗, 張建偉, 熊井克仁, 森嶋厚行. “クラウドソーシングに基づく聴覚障害者によるリアルタイム文字情報保障手法の検討”. 第8回データ工学と情報マネジメントに関するフォーラム(DEIM2016), E-8-6, 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] 若月大輔. “ウェブを活用した聴覚障害者のための情報保障実験サイト”. <https://capti.info.a.tsukuba-tech.ac.jp/>, (参照 2017-02-01).

[7] 大塚正雄, 丁井雅美. “モバイル機器の文字入力の使いやすさ—モバイル機器の文字入力速度の比較検討—”. 広島国際大学医療経営学論叢, 2012, p.1-16.

[8] 山本浩. “ワールドカップ実況放送の現場から”. マス・コミュニケーション研究, No.62, 2003.

[9] 藤原英司訳. “シートン動物記 1, オオカミ王ロボ・ぎざ耳坊やの冒険ほか”. 集英社, 1987.

[10] アンリ・ファールブル, 舟崎克彦訳. “ファールブル昆虫記 子どものための世界文学の森 20”. 集英社, 1994.

[11] スーザン・パーレイ, 小川仁央訳. “わすれられないおくりもの”. 評論社, 1986.

付録

付録 A.1 入力速度測定用例文

例文 1 母ぎつねは、すぐに子ぎつねをくわえて走り出しました。しかし、くさりがのびきったところで、子ぎつねは、母親の口からもぎとられてしまいました。そのとき、人が窓をあける音がして、母ぎつねは、材木の山をこえてにげていきました。[9] (110文字)

例文 2 ファールブルは、フルネームをジャン・アンリー・ファールブルとって、1823年、南フランスの地方都市サン・レオンで生まれました。貧しい農家の子ではありませんでしたが、そのかわり、ファールブルのふるさとは、豊かな自然に恵まれていました。[10] (112文字)

例文 3 アナグマは賢くて、いつもみんなに頼りにされています。困っている友達は、誰でも、きっと助けてあげるのです。それに、たいへん年をとって、知らないことではないというぐらい、もの知りでした。アナグマは自分の年だと、死ぬのが、そう遠くはないことも、知っていました。[11] (130文字)