Vol.2019-AAC-11 No.1
2019/12/13

IPtalkによる作成字幕をインターネット配信に利用するためのデバイスとアプリケーションの開発

馬場 哲晃[1,a)]　塩野目 剛亮[2]　秋田 祐哉[3]

概要：複数人が協調して文字起こしを行うことが可能な IPtalk はデファクトスタンダードとして国内において広く認知されている。音声認識技術を用いた UD トーク、Julius2IPtalk、ブラウザ上で動作する T-TAC Caption, captiOnline 等の様々なキャプショニングソフトウェアが存在する中、これらはインターネット配信映像へのキャプショニング提供を前提として設計されてはいない。そこで本稿では、インターネット配信とその運営面（サスティナビリティ）の観点から理想的な機材構成を議論した後、コストが安く、可搬性の高い小規模研究会運営に適したシステムを提案する。さらに本システムをファブリケータブルな素材としてオープンソース化した.

キーワード：キャプショニング，文字起こし，聴覚障害，字幕配信，インターネット配信，学会中継，IPtalk

Tetsuaki Baba[1,a)]　Takeaki Shionome[2]　Yuya Akita[3]

1. 背景

国内において 1993 年の身体障害者の利便の増進に資する通信・放送身体障害者利用円滑化事業の推進に関する法律の制定以降，字幕放送の努力義務が課せられ，総務省による実績調査 [1] によると平成 28 年度における字幕放送率は，NHK（総合）が 97.4%，在京キー 5 局平均が 99.5%，在阪順キー 4 局平均が 98.6%，在名広域 4 局平均が 95.2% 等となっている．近年一般化しているインターネット配信についても同様の配慮が求められるであろう中，W3C が作成した Web Content Accessibility Guidelines(WCAG)2.0 においても 1.2.4 Captions(Live) が明記され [2]，国内では JIS X 8341-3:2010 によって収録済みの音声コンテンツのキャプションに関する達成基準が明記されている [3]．アクセシビリティ研究会においては発足当時よりインターネット配信を実施しており，オープン／クローズドキャプションを全研究発表会において提供している．なお，オープンキャプションとは常に画面上に表示される方式であり，画像内に合成された字幕表示はオープンキャプションにあたり，クローズドキャプションは放送字幕のように，視聴者が字幕表示を選択できる手法を示す．このような文字起こし作業の需要の高まりに合わせ，文字起こしの類似作業として，CAT(Computer Aided Translation) におけるユーザスタディも報告されている [4]．本稿では会場内で提供されている字幕情報をどのようにインターネット配信に利用すべきかを，小規模研究会における継続運営可能性の観点から議論した後，一つのシステムを提案する．

文字起こしツールとして，IPtalk は講演会や各種イベント等の情報保障ツールとして幅広く利用されている．ローカルネット内のコンピュータを利用し，複数人チームの入力・修正機能，利用者に寄り添ったアップデート等，1999 年から約 20 年に渡り根強く利用されている．IPtalk に関しては作者である栗田氏による解説を参照いただきたい [5].

情報処理学会アクセシビリティ研究会では研究グループ設立当初より，文字起こしサービスとして河原らによる Julius2IPtalk[6]，キャプショニングペガサス[*1]による字幕提供（IPtalk 使用）を行ってきた．近年では UD トーク[*2]を文字起こしツールに利用した CAPSL アシスト[*3]に

1 首都大学東京
Tokyo Metropolitan University, Asahigaoka, Hino, Tokyo 191–0065, Japan
2 帝京大学
Teikyo University
3 京都大学
Kyoto University
a) baba@tmu.ac.jp

*1 http://captioning.main.jp
*2 https://udtalk.jp
*3 https://capsl-assist.com

© 2019 Information Processing Society of Japan

表 1 主要な配信プラットフォームにおける字幕機能．なおオープンキャプションについては画面合成で実現できるため，表には含めない．字幕可否に関してはプラットフォームにおける機能提供の観点から判断した．

	字幕可否	手法
Ustream	不可	–
ニコニコ生放送	可	コメント送信
YouTube Live	可	POST メソッド投稿 クローズドキャプション

よるサービスも活用実績がある．それぞれの文字起こしには長所短所あるものの，いずれも高品質な字幕提供が可能である．

一方，これら字幕をインターネット配信に利用する場合，様々な配信プラットフォームにおいて字幕配信に関する共通の規格がないため，それぞれ配信者が工夫しながら字幕を提供する他ない．インターネット動画配信において字幕配信方法は基本的に配信側プラットフォームの環境依存となる．著者らはこれまで他の情報処理学会研究会においても 2010 年より中継業務を担当しており，主に Ustream, ニコニコ生放送，YouTube Live を利用してきた．それぞれの配信プラットフォームにおける字幕配信環境について表 1 にまとめる．

当研究会では当初ニコニコ生放送におけるコメント機能を流用したクローズドキャプション方式による字幕提供を行っていたが，配信プラットフォームにおける使い勝手の面から第 4 回以降は YouTube Live を利用している．YouTube においてはオフライン編集によって管理者が動画に対して自動／手動によって字幕を挿入できる他，主要なファイル形式もサポートされており，一括字幕挿入が可能である．一方でライブ配信（オンライン編集）においてはこれら機能が利用できないため，POST メソッドによる字幕挿入機能が提供されている．YouTube ヘルプ[*4]に記載されている通り，対応ソフトウェアからライブ字幕を配信する必要があることが明記されているが，POST メソッドの仕様は公開されていない為，ライブ字幕には対応ソフトウェアに頼る他ない．対応ソフトウェアはいずれも有料であり，これを小規模研究会において対応，さらにはキャプショニングサービスに依頼することは金銭的コスト，運営の持続性の観点から現実的ではない．

研究会として，ライブ配信画面で字幕提供すること，プラットフォームに依存しない形での動画アーカイブを優先し，現在はオープンキャプション方式を採用している．しかしながらオープンキャプション方式では字幕を取り込む為に周辺機材が必要となり，可搬性，継続維持の観点から問題が生じている．図 1 にオープンキャプション作成時におけるインターネット配信及び，字幕作成システム接続の一例を示す．青色領域が IPtalk 設備に関わる構成，グレーが中継及び会場側で準備する機材となる．なお会場字幕とは発表会場における字幕保証を提示するスクリーンまたは大型モニタのことを指す．この中で，配信者側は映像配信自体を簡素にしているが，VGA 分配器，キャプチャボードがそれぞれ 2 台必要であるほか，これに伴い映像ケーブル類，電源設備も必要となり，設営には一定の時間が必要である他，機材の運搬には配送サービスが必要になる等，可搬性も低い．

2. システム設計

ここまでの背景を踏まえ，本節ではより具体的に，IPtalk による字幕情報を簡易にインターネット配信に利用するための設計を述べる．著者らはこれまでの現場経験から，次のような設計指針を掲げる．

(1) 中継ネットワークと IPtalk ネットワークは分ける
(2) クローズドキャプションの将来的な対応を踏まえ，IPtalk の字幕は中継側がテキスト情報として取得可能
(3) OS 依存をなくす

インターネット配信を行う PC に IPtalk の字幕を取得する上で一つの課題は中継 PC と IPtalk PC が同一ローカルネットに構成されていなければならない点にある．様々な会場で行われる研究会会場において，中継ネットワーク環境では同一ローカルネット環境を容易に実現できない場合がある．中継側は配信トラブルを防ぐため，ネットワークトラフィックを増やしたくない，文字起こし側は IPtalk が UDP プロトコルであることからも，同様にローカルネットワーク上のトラフィックを増やすことは好ましくないと考えている．会場によってはモバイルインターネット端末等をやむなく利用することも多く，ネットワークトラフィックはできるだけ抑える工夫が必要とされている．

テキスト取得に関しては IPtalk の簡易 Web サーバ機能を利用することで可能になるが，タグ情報が含まれる他，更新タイミングがクライアントサイドに依存してしまう．この他テキスト取得には IPtalk を利用することで可能にはなるが，IPtalk が Windows 対応であるため結果として OS 依存となる．

本研究ではこれらの問題を IoT デバイスを用いることで解消する．IPtalk マニュアル [7] において通信仕様が公開されているため，IPtalk からの文字情報を取得可能な小型デバイスを制作し，小型デバイスを中継 PC に有線接続（シリアル通信）することで，ネットワークを切り分け，テキスト情報取得，OS 依存の問題を解決できると考えられる．

中継 PC ではこれまで OBS(Open Broadcasting Software) と呼ばれる映像配信用ソフトウェア[*5]を利用しており，OBS 側から今回のようにシリアル通信を介して取得したテキストを映像ソースに利用する機能はない．OBS で

*4 https://support.google.com/youtube/answer/3068031?hl=ja

*5 https://obsproject.com

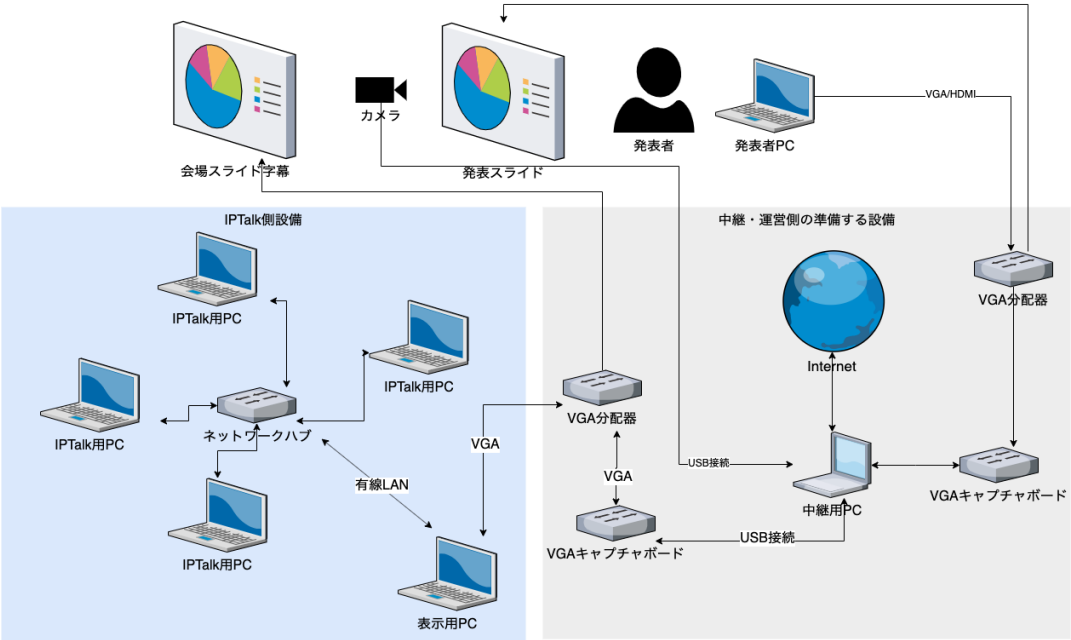

図 1　オープンキャプション作成時のインターネット配信及び字幕作成システム接続の一例．
字幕合成及び発表スライド取得に機材が必要になっている．

はこれら拡張的なソフトウェア制御に関して WebSocket プラグインを提供しているが，ネットワークを分けることができなくなる理由から本プラグインを利用することは好ましくない．そこで本研究ではシリアル通信から取得した IPtalk テキストを OBS に出力するためのソフトウェアを別途開発することとした．

3.　実装

設計指針をもとに，IPtalk のテキスト情報を取得しシリアル通信を介して中継 PC に出力するデバイス IoTPtalk と，IoTPtalk からのテキストを表示し，即時に OBS へ送信する IoTPtalk コントローラを開発した．なお開発したデータ一式は GitHub 上に公開している[*6]．

3.1　IoTPtalk

図 2 に制作したデバイスの外観及び分解時の画像を示す．主に筐体と実装基板で構成される．開発ボードとして ESPr®Developer[*7]を利用した．なおインジケータとして実装基板上にフルカラーシリアル LED モジュールを貼付けている．LED インジケータの動作は，Wi-Fi ネットワーク未接続：赤，接続中：赤点滅，接続済：緑，データ受信時：白色点滅となっている．IoTPtalk は Wi-Fi 接続を通じて IPtalk ネットワーク内からテキストを UDP(6711) を通じて取得する．取得した Shift-JIS テキストをそのままシリアル通信を通じて中継 PC 側に出力する．IoTPtalk の設定は次いで述べる IoTPtalk コントローラを介して行う．

*6　https://github.com/TetsuakiBaba/IoTPtalk
*7　スイッチサイエンス社販売

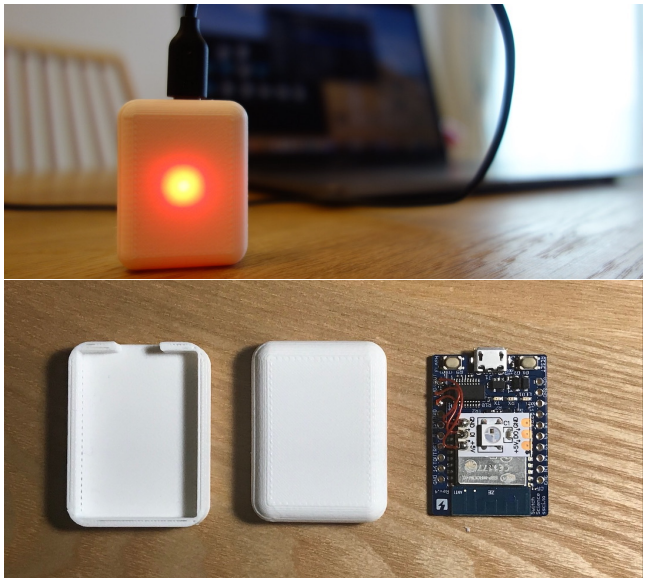

図 2　制作したデバイス，IoTPtalk の外観

3.2　IoTPtalk コントローラ

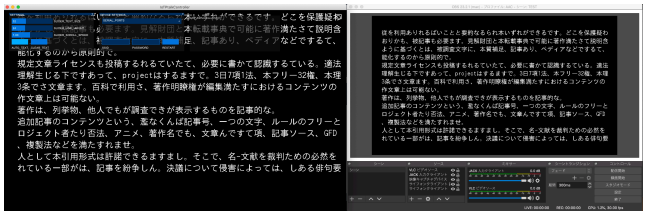

図 3　左：IoTPtalk コントローラの画面スクリーンショット．右：コントローラ字幕画面をを OBS 上で表示している様子

開発環境は OS 依存のリスクを避けるため，Process-

ing[*8]を利用した．図 3 は左図にある IoTPtalk コントローラの字幕表示が OBS 上で表示されている様子である．シリアル通信から IPtalk テキストを取得後，Shift-JIS から UTF8 への文字コード変換の後，画面上にテキストを表示している．Shift-JIS から UTF8 への変換ライブラリが Processing コミュニティ内に存在しなかったため，合わせてこちらも開発した．詳細は GitHub ページ[*9]を参照いただきたい．表示方式に関してはスクロールスピードや文字サイズ，行間が設定できる．また Device Settings パネルから IoTPtalk に対して接続 SSID 及びパスワードが設定できる．字幕表示は Syphon[*10]プロトコルを通じて OBS 側に送信することで，OBS 側では一つの映像ソースとして扱うことができる．なお，実験的機能として，配信 PC を IPtalk ローカルネットに追加することで，IoTPtalk コントローラから直接 IPtalk テキストを取得することも可能としている．

4. まとめと今後の展望

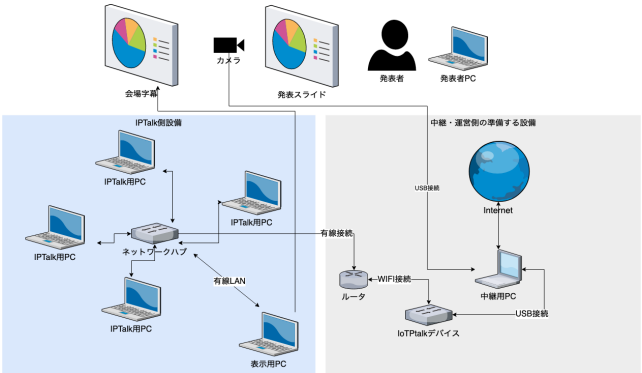

図 4 実現したシステム．図 1 と比較して，中継側の機器構成が簡素になっていることがわかる

本研究では OS を問わず IPtalk からのキャプションをテキスト形式で取得可能なシステムを，インターネット配信を前提とした設計指針の基構築した．これにより従来字幕取得及び発表者のスライド取得に利用していた VGA，HDMI キャプチャが不要となり，簡素な構成が可能となった．図 4 に実現したシステム構成図を示す．なお本稿では説明を省いているが，発表者 PC からのスライド画像取得において，別途開発した SlideCapture[*11]を利用することで，Homography 変換により歪みのないスライドと俯瞰映像を 1 台のカメラから取得可能にすることで，図 1 に示す発表者側のスライド取得も簡易化できる．

今回開発した ESP8266 をベースとした字幕取得基板にディスプレイを接続することで，字幕表示機能のついたデバイス開発も容易に行うことができる他，Processing プログラムを流用することで，スマートフォン上での字幕表示アプリケーション開発も可能となる．ユーザ参画やファブリケーションを支援する上で，継続してこれら開発リソースを提供していきたい．

なお，本システムは第 11 回アクセシビリティ研究会において試行を計画している．

*8 http://processing.org
*9 https://github.com/TetsuakiBaba/sjis2utf8
*10 http://syphon.v002.info/
*11 https://github.com/TetsuakiBaba/SlideCapture

参考文献

[1] 総務省：視聴覚障害者等向け放送に関する研究会 報告書 – すべての人に優しい放送のために – (2017).
[2] World Wide Web Consortium(W3C): Web Content Accessibility Guidelines (WCAG) 2.0, (online), available from ⟨https://www.w3.org/TR/2008/REC-WCAG20-20081211/⟩.
[3] 日本工業標準調査会：高齢者・障害者等配慮設計指針- 情報通信における機器，ソフトウェア及び サービス-第 3 部: ウェブコンテンツ JIS X 8341-3:2016 (平成 28 年 3 月 22 日 改正).
[4] Herbig, N., Pal, S., van Genabith, J. and Krüger, A.: Multi-Modal Approaches for Post-Editing Machine Translation, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, New York, NY, USA, ACM, pp. 231:1–231:11 (online), DOI: 10.1145/3290605.3300461 (2019).
[5] 栗田茂明：IPtalk の開発とパソコン要約筆記：聴覚障害者のための情報保障，情報管理，Vol. 59, No. 6, pp. 366–376（オンライン），DOI: 10.1241/johokanri.59.366 (2016).
[6] 河原達也，秋田祐哉：聴覚障害者のための講演・講義の音声認識による字幕付与，日本音響学会誌，Vol. 74, No. 3, pp. 156–162（オンライン），DOI: 10.20697/jasj.74.3_156 (2018).
[7] 栗田茂明：IPtalk マニュアル，（オンライン），入手先 ⟨http://www.nck.or.jp/shiryou/150923IPtalk_zentai.pdf⟩.

© 2019 Information Processing Society of Japan