音声認識併用型遠隔文字支援システムの構築

森　直之

Construction of remote captioning support system with speech recognition

NAOYUKI　MORI

音声情報を文字に変える支援は、徐々にニーズが増してきている。大学においてはノートテイカーや要約筆記者の育成がなされているが、いまだに文字支援が難しい現場も数多く存在する。今回はノートテイクや要約筆記といった文字支援の場において音声認識を併用した情報保障の仕組みを構築し、実践した内容を報告する。

はじめに

本論では、講演や授業において音声情報の代わりに文字情報を提供する、いわゆる「情報保障・情報支援」に用いられるアプリケーションソフトウェアに着目し、より多くの情報を伝えるためのシステム設計及び機能について言及する。

国内では、「障害者の権利に関する条約（障害者権利条約［外務省 2013]）」が批准されたことを皮切りに、高等教育機関をはじめ教育機関や民間企業内に合理的配慮に対する対応が進められている。合理的配慮の定義では、「*障害者が他の者との平等を基礎として全ての人権及び基本的自由を享有し、又は行使することを確保するための必要かつ適当な変更及び調整であって、特定の場合において必要とされるものであり、かつ、均衡を失した又は過度の負担を課さないもの*」とされており、特にこの定義にある「過度の負担」に対する判断は主観的であり難しい。そのため、例えば文部科学省においては初等教育機関における事例検討がなされている[文部科学省, 2010]。事例には字幕作成や歩調援助システムといった技術・システムが例としてあげられており、現に支援をしている学校もある。

日本学生支援機構の調査［独立行政法人 日本学生支援機構, 2016]によると、聴覚・言語障害者に対する支援は、パソコンテイクが 39.5%、ノートテイクが 49.8%、音声認識ソフトウェアが 6.0%となっており、音声情報を活用する例が報告されてきている。昨年の調査［日本学生支援機構, 2015]では音声認識ソフトウェアの調査項目がない点から、実際に音声認識技術が活用され始めている事が推察できる。

静岡福祉大学では、情報保障について平成 17 年度にノートテイクソフトウェア「まあちゃん」を製作し、その環境づくりを研究・サポートしている。ノートテイクの現場においては、ノートテイカーの育成に時間がかかることや、卒業等による支援メンバーの入れ替わりなど、継続的にサポートを得られることが難しい現状があるが、この点に音声認識技術と遠隔技術が適用できるのではないかと筆者は考えている。特に音声認識においては、ディープラーニングなどの機械学習技術によって、近年の認識技術は飛躍的に精度を向上している。音声認識であれば、携帯電話から話しかけるだけで正しく認識し、その言葉にしたって機材が動くなど、個人の身の回りに存在し活用できるレベルまで技術が洗練され、応用できるレベルとなってきた。また、遠隔に関しても、クラウド技術などにより通信の敷居が下がってきており、SoftEther などの VPN サーバーを活用して、遠隔地であってもネットワークを組みやすい環境が比較的簡単に実現できる世の中になってきている。

1.音声認識システムとの連動方法

従来、音声認識ソフトウェアは、音声を録音しモデルに入力をするシステム、音響モデル、文章を提示するシステムなどが１つのパッケージとなっている製品を購入し、導入することで利用可能になっている。例えば、アドバンスド・メディア社の AmiVoice SP や、Nuance 社のドラゴンスピーチなどが、このシステムに該当する。専用のウィンドウの中で文章を音声入力するケースや、IME と同様にカーソル位置に文章が入力されるケースが一般的である。例えば、先行研究されている ソフトウェア SR-LAN2 [三好茂樹, 2009]で

あれば、AmiVoice ES2008 を起動してフォーカスのある入力欄に文字化された文字を挿入させることで音声入力を実現している[「音声認識によるリアルタイム字幕作成システム構築マニュアル」編集グループ, 2009]。これらの場合、音声認識を使って文章入力する設定の手数が多く、操作者がキーボードを入力操作しているケースでは併用できないというデメリットがある。そのためいくつかの会社では、API と呼ばれる直接文字のやり取りができるプログラム用通信の仕組みを用意し、システム同士が連携可能になるようにしている。まあちゃん 2016 においても、AmiVoice SP2 と連動して入力できる仕組みを用意している。

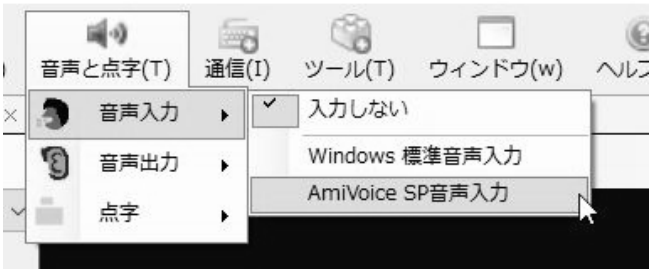

図 1　SP 連携入力システムの画面

近年、認識率が格段に向上した Apple 社の Siri などを代表するディープラーニング採用されたシステムは、主にインターネットベースの技術(クラウドや JavaScript ベースのシステム）や携帯電話 OS に特化されているため、Windows ネイティブアプリから直接扱うようにはできていないケースが多い。今回の研究では、携帯電話の音声認識アプリケーション「UD トーク　[Shamrock Records, 2016]」を用い、AmiVoice Cloud を用いた連続発話認識の文字情報を通信で連携させて利用する方法を採用し、研究することとした。

2 システム間通信方式の検討

これまで、要約筆記システムでは、いろいろな通信方式が採用されてきた。UDP を採用しているアプリケーションでは、安定した有線 LAN で使うことを前提としており、入力や表示に関する応答性を優先した構成になっている。TCP を採用しているシステムでは、応答性よりも安定性を優先した設計になっている。UDP はその性質上、パケットの入れ替わりやデータの欠落が起きうるため、UDP と TCP を併用したプロトコルを採用している「ハイブリッド式」のアプリケーションもある。当時設計された通信速度は 10Mbps の時代であるが、現在では民生品においても 1Gbps の機材が容易に入手可能な時代となっており、「文字情報を伝える」ことに関して「文字の入れ替わり」や「欠落」といったリスクよりも速度をとるメリットはなくなってきている。

表 1　プロトコル表

システム	通信方法	プロトコル
IPtalk [栗田茂明, IPtalk, 2016]	UDP	IPtalk
Mekiku [mekiku, 2016]	UDP	IPtalk Mekiku 独自
ITBC2 [森直之, ITBC2, 2014]	UDP TCP	IPtalk まあちゃん はやとくん Telnet IRC
RTD2 [神野健吾, 2014]	TCP	Telnet
まあちゃん 2 (静岡福祉大学)	UDP	まあちゃん
まあちゃん 2016 (静岡福祉大学)	TCP UDP	UD トーク IPtalk
UD トーク [Shamrock Records,Inc, 2016]	TCP	UD トーク

今回も、ノートテイクシステムに関する課題の整理と再設計 [森直之, 2015]において提示した方式を用いて UD トークと連動する方法を採用する。

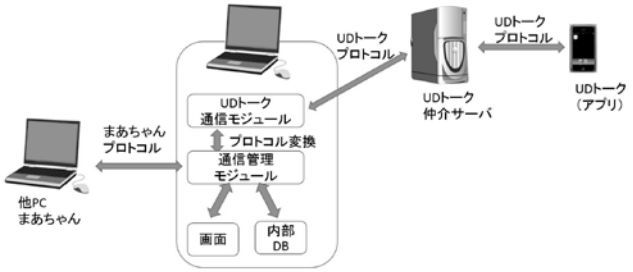

図 2 プロトコル構成図

3. 実装

音声認識の特性として、「仮確定文章の確からしさ」が認識途中で変化していく点があげられる。入力を確定することによって情報を送るノートテイクソフトと仕様が異なる部分があるため、整合を取る必要がある。今回は、音声認識とキーボード人力入力、双方のメリットを引き出せる実装を検討した。

3.1 連動の方法

UD トークの通信は TCP/IP であるため、サーバーとなる「親」がいる。あらかじめ、そのサーバーに全システムを接続したのち、UD トーク側で話す。

音声データは、音響処理されたあと、音声認識データで連続認識されて、文字情報が UD トークサーバーに接続されている全システムに通知される。その文字列を受け取ったまあちゃん 2016 は、内部 DB にデータをストアし、文章列の位置を再現する。結果、画面表示が正しく連動するようになる。

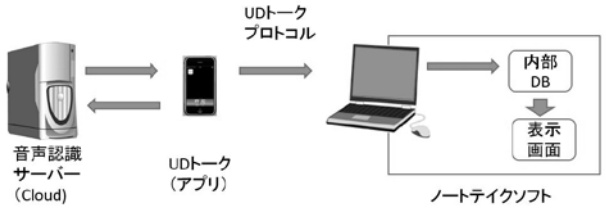

図 3 ストアの流れ

文章は確定されたあとにも修正する必要があるが、全端末の同期を取っていく必要がある。そのため、TCP 接続の親となる端末が、編集に関する制御制御機能をもち、許可をだした端末だけが編集をできるような形をとることで全端末が同期するようにしている。

3.2 入力情報の見せ方

ノートテイクや要約筆記では、連携をとるためにチームを組んでいるペアの入力が見える必要がある。UD トークで採用している連続発話音声認識では、未確定の変換過程を画面に表示することでリアルタイム性を確保する。それ故に、キーボード入力時の変換過程もが見える。できるだけ早く情報を得るためには有効な手段ではあるが、通常のノートテイクや要約筆記と同等の画面（決定文字列だけを表示）にはならない。まあちゃん 2016 の実装に関しては

① 「入力状態が見える」ウィンドウを作り、まとめて現状が把握できる画面を用意。

② 表示に関しては入力過程をそのまま見せるのか、入力確定してから見せるのかを見る側が決められるように設計。送信側は、IME の仮入力状態なのかが分かるよう、フラグデータを添付して送信。

という方法で従来と同等の表示ができる仕組みを実現している。

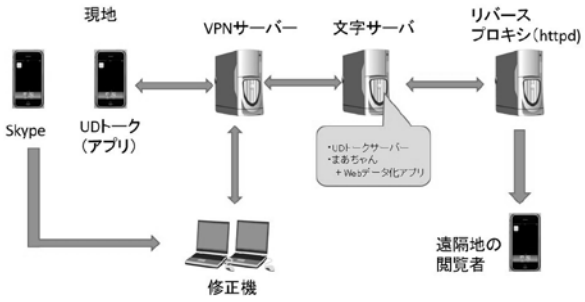

図 4 配信構成図

3.3 遠隔に関する機能

UDP 通信を採用する場合、特に情報の欠落を意識しなければならない。 VPN を使っている事例は多くあり、たとえば、非営利活動法人での実用例がある [栗田茂明，2013]。今回は実験として遠隔実験チーム EXTRA を立ち上げ、VPN を利用して字幕配信のテストを実施した。その結果 UDP をベースとした通信の場合、①ブロードキャスト通信が VPN 先のネットワークを超えられないケース②VPN 対応の仮想 IF に対してファイアウォールの設定が利かないケース③通信が不安定なときに、パケットが欠落する、などの問題に遭遇した。そのため、安定した配信を実現するためには、ある程度の機材、通信機器、サーバ管理者の知識が必要となる。その点、TCP であれば通信欠落に対しての強さがあり、ブロードキャスト固有のの問題に遭遇することもない。しかし、背反としてネットワークにつながっているグループメンバーを検知する仕組みが必要になる。今回は、UD トーク同様、マルチキャストを用いて同等グループに参加している個体を検出できる方法を採用した。

3.4 文章の訂正方法

音声入力された文章は更新速度が早いため、文章履歴画面を追ってから入力枠に視点移動すると訂正までに時間を要してしまう。例えば図 5 のような配置であれば、履歴行数と入力枠が離れているために修正位置が

直感的にわかりづらい。

図 5 入力・訂正画面

そのため、上下の位置関係を把握しつつも、入力位置が分かるようなレイアウトを採用した。

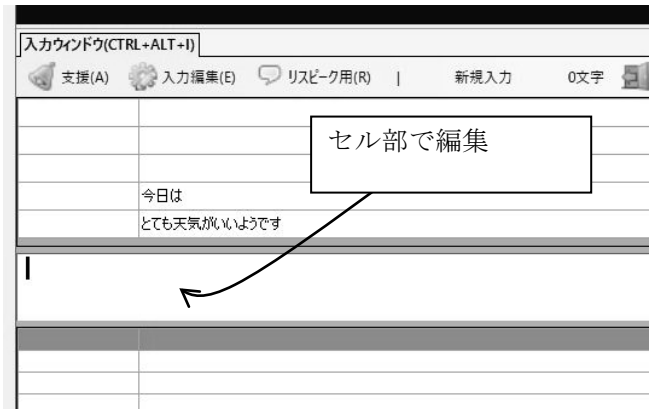

図 6 入力・訂正画面　改良版

4.実験

これらのシステムを組み込んだものを情報支援現場で実際にテスト運用してみた。

概要：

VPN を構築し、現地から文字データを送り、遠隔地で修正を実施するケース。

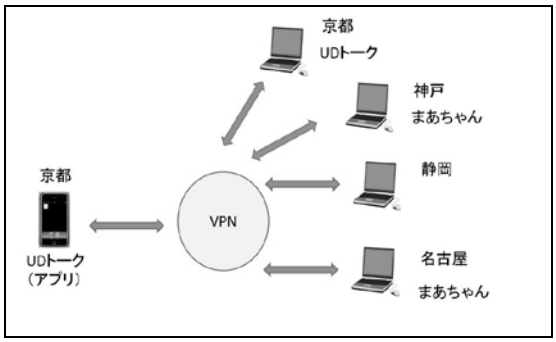

図 7　ネットワーク構成図

音声は、現場から各地方へ Skype で送り、修正者が遠隔から訂正参加する方法を採用した。現地での対応者は 1 名で、音声送信や文字情報の配信準備を担当した。字幕は UD トークアプリの QR リンク機能を用いて HTTP 通信経由で利用者の手元まで字幕が届く方式と、全体投影を併用する形で運用した。通信の負荷を分散するために、HTTP 通信で字幕を配る部分にはリバースプロキシによる配信キャッシュを持たせておき、入力者側のシステムに負担がかからないようにしている。受信者がごく少数の場合においては、このようなサーバー構成にせずとも対応できると考えているが、今回の実験では受信者数が大勢いた場合にも対応できるようにする必要があったため、この形を採用した。

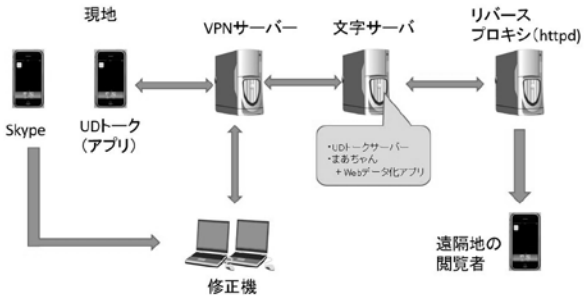

図 8　内部構成図

5.結果

VPN を用いた音声情報併用の字幕配信はシステムとしては、音声認識～文字提示まで、話終わり 2 秒程度で確定、その後修正によっておおよそ５～１０秒程度で修正完了という状況であった。訂正に関しては音声を遠隔地に送るまでの時間に数秒かかり、音声認識終了後に編集権を取得するのに 1 秒程度かかっていた。本実験を行ったイベントでは、映像の一般公開がなされており、映像配信システムの遅れが１５秒程度あったことから、現場では話し終わってからの文字確定という流れであったが、インターネット経由で字幕を受信した場合には、同じタイミングか、あるいは字幕配信のほうが速かった、という結果が得られた。そのため、インターネット経由で閲覧するときに限って言及すれば、訂正編集が終わった時点で表示を出す方法を採用しても遅れがほとんど感じさせずに提示することが可能であることが分かった。

ネットワークに関しては、モバイルインターネット回線や VPN の安定度が確保できたため、切断されることなどによる弊害や文字提供時の問題を確認すること

はできなかった。本実験への参加者からの反省としては、下記の点があげられた。

（○：良い点　　●：問題、課題）
●編集権のやり取りがうまくいかないケースあり。
●編集ロックが解除できないケースが発生すると解放されないままになるケースがある。
●UDP に比べて処理が重いと感じる。
○VPN(TCP)使用による処理のもたつきによって作業ができないことはなく、「編集権取得の操作が遅れる」などの特性を理解していれば大きな問題は感じない。
○編集は 2 名で十分対応できると感じた。
●音声認識に向く、向かない話題がある。
●流れる文字をみて酔うケースがあった。
○話した言葉が手を加えられず、そのまま表示されることにより、心理的な安心感がある。
●（拍手）などの表現が自動で入ることはない為、状況を伝えるには訂正者が対応する必要がある。
○文字化されたことにより、文字情報の必要性を理解していただけた。
●情報保障という観点でいえば、音声情報を文字に変えただけでは不十分と感じる点がある。
○音声さえ送れば訂正ができる仕組みはかなり有効的と考える。
○音声の送信に遅れがあるため、現地の編集者と遠隔の編集者が編集権を取り合うのは効率的ではなく､「修正は遠隔で行う」などルール決めは効果があることがわかった。

6.考察と今後

今回の実験では、イベントでの字幕付けで行ったが、この仕組みを用いれば、大学のキャンパス間や大学間での情報保障の仕組みを作ることも可能と考える。近年では学内 LAN がキャンパス間で接続されているケースも多いため、ブロードキャスト通信に依存しないノートテイクツールの需要も増していくことが大いに考えられる。また、副次的な効果であるが、音声認識によって情報保障を行うことによって、話し手側が内容を認識しやすく丁寧に話す傾向が見られた。内容が誤認識する場合には発話し直すなど、話者の中でフィードバックが行われることによって、後から別のものが訂正しなくても提示に耐えうる字幕を作ることも可能であることが分かった。すなわち、音声認識自体の能力向上（認識率向上）のほかに、認識しやすい話し方によって誤認識率を下げることが可能であり、これによって訂正者の負担・仕事量が減ることになる。仕事量が減ることで遠隔で実施する場合も訂正が最小化され、あるいは意味が通じる文章に限っては訂正をしない選択も可能になると考える。(例えば､「合う」「会う」などの漢字違いなどはニュアンスがわかるため訂正しない､など)。すなわち、従来のノートテイクや情報保障にあった「人間の能力に最大限依存したシステム」は、一部を最新技術に託すことで効率と内容精度を高め、要約や筆記に関係する仕事量を減らす形で提供することが可能となった。

今回は「話し言葉を最大限文字化する」技術と、遠隔地から修正する技術を用いて実験をしてきた。この技術は有用ではあるものの、一方で「話し言葉の文字量に読み取りが追いつかない」という声もある。この点は今後の検証課題としていきたい。

謝辞等
本論文における研究は科研費研究基板Ｂ（21330143）の助成を受けて構築したシステム基盤をベースに、音声認識連動部分などの応用機能を組み込み、検討を進めたものである。また、UD トークとの連携に際し、Shamrock Records 株式会社の青木秀仁氏、株式会社プラスヴォイス社に、実証実験の実施に関して遠隔実験検証チーム Project EXTRA、また、EXTRA の実験場所提供として京都大学 宇宙総合学研究ユニットにご協力頂いた。この場を借りてお礼申し上げる。

参照文献

「音声認識によるリアルタイム字幕作成システム構築マニュアル」編集グループ. (2009). 音声認識によるリアルタイム字幕作成システム構築マニュアル. 参照先: http://www.tsukuba-tech.ac.jp/ce/xoops/file/seika/onseininshiki-manual.pdf
mekiku. (2016). 参照先: http://www.mekiku.com
Shamrock Records,Inc. (2016). UD トーク. 参照先: http://udtalk.jp
外務省. (2013 年 3 月 6 日). 障害者の権利に関する条約. 参照日: 2014 年 9 月 28 日, 参照先: 外務省 : http://www.mofa.go.jp/mofaj/gaiko/treaty/shomei_32.html

栗田茂明. (2013). 運用コスト低減を目指した遠隔パソコン文字通訳システム. 参照先: http://www.nck.or.jp/shiryou/131116HIS_ReducingOpCosts.pdf

栗田茂明. (2016). IPtalk. 参照先: http://www.geocities.jp/shigeaki_kurita/

三好茂樹. (2009). SR-LAN2.

森直之. (2014 年 07 月 14 日). ITBC2. 参照日: 2014 年 9 月 28 日, 参照先: http://www.caption-sign.in.net/software/itbc2.html

森直之. (2015). ノートテイクシステムに関する課題の整理と再設計. 静岡福祉大学.

神野健吾. (2014). RTD2. 参照先: http://hp.vector.co.jp/authors/VA006163/pccap/

独立行政法人 日本学生支援機構. (2016 年 08 月 25 日). 平成 27 年度（2015 年度）障害のある学生の修学支援に関する実態調査. 参照日: 2014 年 09 月 28 日, 参照先: 独立行政法人 日本学生支援機構: http://www.jasso.go.jp/gakusei/tokubetsu_shien/chosa_kenkyu/chosa/__icsFiles/afieldfile/2016/03/22/h27houkoku.pdf

日本学生支援機構. (2015 年 3 月 27 日). 平成 26 年度（2014 年度）障害のある学生の修学支援に関する実態調査. 参照先: 独立行政法人 日本学生支援機構: http://www.jasso.go.jp/gakusei/tokubetsu_shien/chosa_kenkyu/chosa/__icsFiles/afieldfile/2015/11/09/2014houkoku.pdf

文部科学省. (2010 年 9 月 6 日). 特別支援教育の在り方に関する特別委員会（第 3 回） 配付資料 > 資料 3：合理的配慮について. 参照先: 特別支援教育の在り方に関する特別委員会: http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/044/attach/1297380.htm