

音声認識とチャットを併用した聴覚障害学生支援システム

石原 守[†] 澤口 朋丈[†] 市村 哲[†]

東京工科大学 コンピュータサイエンス学部

聴覚障害学生は全国の大学や他の教育機関に点在している。聴覚障害学生の支援方法としてボランティアが講師の話を聞き、パソコンに入力して聴覚障害学生に見せる方法が一般的である。しかし、この方法では十分な情報を伝える事が出来ない。そこで、本稿では別室で講義ストリーミング映像を見ているボランティアが講師の音声を復唱し、音声認識技術を用いる事でテキスト化して、聴覚障害学生の WEB ブラウザ上にそのテキストを表示するシステムを提案する。また、「あれ」や「これ」といった指示代名詞の代わりに映像からキャプチャーした画像を挿入する機能を実装した。実験の結果、画像挿入により講義を理解しやすくなった反面、音声認識率が悪い場合には、文章による講義理解が困難であるという意見が得られた。音声認識の際、文章を読み上げる様に音声入力することで高い認識率が得られる事がわかった。

A notetaking system with sound recognition function for hearing loss students

MAMORU ISHIHARA[†] TOMOTAKE SAWAGUCHI[†] SATOSHI ICHIMURA[†]

School of Computer Science, Tokyo University of Technology

There are many hearing loss students at universities and educational institutions in Japan. It is popular to have volunteers enter text into PCs for them. However, it is difficult to transmit accurate information. In this paper, we propose a notetaking system with speech recognition function where volunteers repeat teacher's voice and the system recognition it. We also implemented a function to insert an image substituting demonstrative pronouns. The experiment showed that hearing loss students came to understand a lecture by inserting the image, but there was a problem that speech recognition rate was bad. We found that speech recognition rate could be higher if the volunteers spoke sentences as if they read them.

1. はじめに

聴覚障害学生は全国の大学や他の教育機関に点在している。聴覚障害学生への支援方法として、ノートテイク（ボランティア学生が講義出席する聴覚障害学生のために、聴覚障害学生の隣席で講師の発話内容をノートに書き写す）や IPtalk¹⁾を利用し、講師の発話内容をパソコンにキーボード

入力して表示させる手法が一般的である。IPtalkは、パソコンを使い、リアルタイムに文字を入力したり、事前に準備した文章を表示したりすることで、聴覚に障害のある人のコミュニケーションを助けるボランティア活動用のソフトである。

しかし、この方法では、十分な情報を聴覚障害学生に伝える事は出来ず、またボランティアの数を増やした場合、誰がどこからどこまで入力するかわからないため、連携するための訓練が必要になる。また、手話通訳などの支援方法もあるが、

[†] 東京工科大学 コンピュータサイエンス学部
School of Computer Science, Tokyo University of Technology

有償ボランティアとなっていることが多く、大学が有償ボランティアを雇用する予算を確保するには困難な状況にある。さらに、講師によっては、「あれ」や「これ」といった指示代名詞を頻繁に用いて図を指す講師もいるので、非常に理解しづらくキーボード入力によるノートテイクをより一層難しいものにしている。

そこで本稿では、音声認識技術を用いて講師の音声（実際には、復唱者が復唱した音声）を文章化するシステムを提案する。音声入力はキーボード入力に比べ文章入力の速度が速く、ワープロ検定1級（約75文字/分）の人が入力するのと同程度の速度で文章を作成することが可能である。

また、既存のシステムの場合、「あれ」や「これ」と言いながら講師が指し示した箇所を文章だけで伝える事が困難な為、文章中にキャプチャー画像を挿入する機能を実装した。別室にいる復唱者への授業動画や音声の配信、音声認識ソフトを利用した復唱による文章入力、キャプチャー画像表示などのチャット表示機能を持つ。

評価実験においては、画像挿入機能により、講義を理解しやすくなったという評価が得られた反面、音声認識率が悪い場合には、文章による講義理解が困難であるという意見も得られた。音声認識の際、文章を読み上げる様に音声入力することで高い認識結果が得られることがわかった。

2. 背景と問題点

全国の大学における聴覚障害学生への支援状況については、近年の調査²⁾によると、聴覚障害学生が在籍したことのある大学は約40%で、その内の約80%の大学で1~3名と少数である。

聴覚障害学生への支援方法として、ノートテイクやIPTalkを利用し、講師の発話内容をパソコンにキーボード入力して表示させる手法が一般的である。また、手話通訳などの支援方法もあるが、有償ボランティアとなっていることが多く、大学が有償ボランティアを雇用する予算を確保するには困難な状況にある。

IPTalkを利用したノートテイクの場合、ボランティア学生1人では3割、2人でも5割程度しか講師の話した内容を入力できないと言われている。実際にノートテイクを行っている様子を見に行き、ボランティア学生の人達にインタビューした結果、

この程度の情報量では講義理解には難しいことがわかった。また聴覚障害学生からは先生の口調、雑談、たとえ話など授業を受けている雰囲気が高みたいが、現状では困難であるという意見を得た。ボランティア学生の数を増やした場合、誰がどこからどこまで入力するかわからない為、連携するための訓練が必要になる。しかし、大学がノートテイクの講習を用意することは難しく、また、タイピング能力に優れた有償ボランティアを雇用する予算を確保するには困難な状況である。

講師によっては「あれ」や「これ」といった指示代名詞を頻繁に用いて図を示す講師もいるので、非常に理解しづらくIPTalkでのノートテイクをより一層難しいものにしている。物理などの講義では、専門用語の変換に時間がかかり、その分入力が遅れているので情報量が低下してしまう。微分や積分記号、分数、ベクトルなどの図といったものを文字だけで伝えることは困難であり、入力が不可能である。

また、IPTalkはWindows上でしか動作しない為、OSに依存してしまう。Linuxの講義など、Windows以外のOSを使用する講義ではIPTalkが使用できないという問題がある。

3. 音声認識と予備実験

当初、音声認識ソフト（VaiVoice³⁾）を利用し、講師の音声を認識し文字を直接聴覚障害学生に見せるという方法を考えた。音声認識を用いればタイピングより早い速度で入力可能で、技術力も問わないと考えたからである。現在、音声認識はキーボードとマウスからの入力に代わるパソコンへの新たな文字入力として注目されている。音声認識とは、マイク等を通じて入力される人間の音声をパソコンがデータとして認識し、音声による文章入力やコマンド操作を可能にする技術である。

音声入力はキーボード入力に比べ文章入力の速度が速く、ワープロ検定1級の人が入力するのと同程度の速度で文章を作成することが可能である。

音声認識システムは既に様々な場で用いられているようになっている。音声認識を利用した情報手段としても研究⁴⁾が進んでおり、音声認識に対する期待が高まっている。NHKでは言い換えを利用したリスピーク方式を用いて字幕放送を実現させた⁵⁾。

実際に講義の音声（1000 文字）を復唱し、音声の学習を行っていない場合と、行った場合で認識率の違いを調べた。結果、音声の学習を行っていない場合の認識率は 75%、音声の学習を行った場合の認識率は 92%であった。

実験結果から、音声と専門用語の登録は必須であることがわかった。しかし、音声の登録を行うには1時間程度かかってしまう為、多くの講師の音声の登録を行ってもらうのは難しい。また、指示代名詞や同じ事を繰り返して説明する講師が多く、そのような場合情報量が多くなりすぎてしまう為、聴覚障害学生は全ての文章を読みきれなくなってしまう。よって、講師の音声を音声認識し、直接聴覚障害学生に見せる方法は実施が難しいことがわかった。

講師の声をボランティアが復唱し、さらに講義前に講義資料の専門用語を辞書に登録することで認識率を高めるノートテイキングシステムを構築した。また、訂正者を用意すれば、より正確な情報を伝えることが可能ようにした。

教室

別室

講師

陳香学生

文章送信 (修正済)

訂正者

動画・音声送信

復唱して音声 をテキスト化

復唱者 (朗字含)

文章送信 (修正済)

紙字修正

5. システム概要

また、講師が指示代名詞を使用した場合、指し示す部分を復唱者が動画からキャプチャーし、指示代名詞の代わりに画像として送る。画像の配信には、WEB サーバーを利用した。

The diagram illustrates a system architecture for exchanging text and images. It features three user icons: 陳善學生 (Chen Shan Student) on the left, 陳唱者 (Chen Chang) on the right, and 紅正者 (Hong Zheng) at the bottom. A central server icon is labeled 'WEB' and 'FCS'. Arrows show the flow of data: 陳善學生 sends '文章' (text) to the WEB and receives '圖片' (images) from the WEB. 陳唱者 sends '文章' to the WEB and receives '圖片' from the WEB. 紅正者 sends '文章' to the FCS and receives '圖片' from the FCS. The WEB and FCS are connected by a bidirectional arrow labeled '文章' and '圖片'.

6. 実装

— 75 —



図3 システムの実行画面

① システムウィンドウ

システムウィンドウでは、WEB サーバーの URL と WEB サーバー上にある各機能を持つ Flash ファイル名を入力する。また、FTP 通信を行うユーザー名とパスワードと、画像の送信を行う作業フォルダを入力する。

② 動画配信ウィンドウ

動画配信ウィンドウでは、WEB サーバー上にある動画配信用 Flash ファイルを読み込み、マイクとカメラから入力された音声と動画を、FCS を通じて配信する。マイクとカメラが接続されていれば自動的にストリーミング配信が行われる。

③ 動画受信ウィンドウ

動画受信ウィンドウでは、WEB サーバー上にある動画受信用 Flash ファイルを読み込み、動画送信ウィンドウから配信された音声と動画を、FCS を通じて受信する。

④ キャプチャーウィンドウ

キャプチャーウィンドウをマウスでクリックすると動画受信ウィンドウの画像をキャプチャーし表示する。キャプチャーした動画から、図や数式などをマウスでドラッグして範囲指定する。離すと範囲内の画像を FTP 通信で WEB サーバーに送信する。送信が成功したら、チャットウィンドウに画像を送信した事を伝え、チャットウィンドウは、WEB サーバーから画像を読み込み表示する。

⑤ チャットウィンドウ

チャットウィンドウは、復唱者は音声認識で文章を入力し、訂正者は入力された文章を修正するウィンドウである。チャットウィンドウを図4に示す。聴覚障害学生にはこのチャット用の Flash ファイルを WEB ブラウザで閲覧させる。

復唱者は、音声認識した文に誤字を見つけた場合、要修正箇所をドラッグして指定し修正要求ボタンを押すと、『』で括られ、訂正者に要修正箇所を伝える事ができる。訂正者は、修正確定ボタン右側の入力部分に正しい文章を入力し、要修正箇所をドラッグし選択した後に、修正確定ボタンを押すと、正しい文章に置換され修正する事ができる。

キャプチャーウィンドウから送られてきた画像は、右の画像ランチャーに表示する。画像をクリックするとチャットウィンドウ中央に大きく表示され閲覧する事ができる。大きく表示された画像はマウスで触れると元の小さな画像に戻る。画像ランチャーの上下にある▲▼ボタンを押すと他の画像に切り替え閲覧できる。また、キャプチャーウィンドウから画像が送信されると自動的にチャットウィンドウに画像【画像と対応した番号】が挿入される。これにより文章中に画像が挿入され文章では表現できない指示代名詞などを画像で表現できる。

チャットの設定画面では、画像フォルダを指定することで文章を FCS 上に保存することができ、ユーザー毎に文字サイズなどの文章表示に関わる設定を行うことができる。

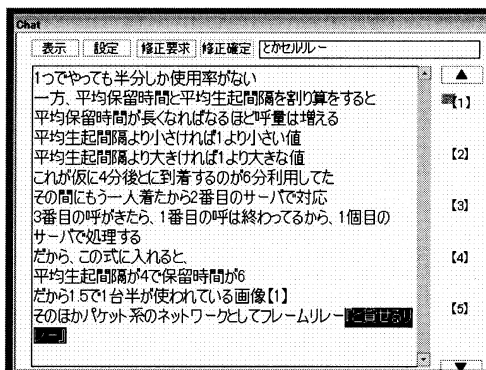


図4 チャットウィンドウ

7. 評価実験

実際に聴覚障害学生が受講している講義に参加し、本システムを使用したノートテイクを行った。本システムと IPTalk を使用させ、聴覚障害学生（1人）とボランティア学生（1人）に4段階評価（1：非常に悪い，2：悪い，3：良い，4：非常に良い）を求めた。評価内容は以下の11項目である。

- ① 表示文は見やすいか
- ② 誤認識，誤変換，誤字脱字など気になるか
- ③ 講義で利用するには十分な情報量か
- ④ 文章による講義理解度はどうか
- ⑤ 画像があることによって講義理解が向上したか
- ⑥ 講師の口調などが伝わるか
- ⑦ 講義を受けている雰囲気があるか
- ⑧ 講師の話した内容を素早く伝えられるか
- ⑨ システムは使いやすいか
- ⑩ 本システムは，今後の大学の講義で有効だと思うか
- ⑪ 事前準備は大変か

7.1 結果と考察

表1 評価結果

質問番号	聴覚障害学生	ボランティア学生
①	4	4
②	1	3
③	2	4
④	1	3
⑤	3	4
⑥	3	3
⑦	3	4
⑧	1	3
⑨	2	3
⑩	4	4
⑪	1	2

表1からわかるように，ボランティア学生からは全体的に高い評価が得られたが，聴覚障害学生の評価には最低の評価が4項目あった。①，⑤，⑥，⑦，⑩の項目に関する評価は両者とも高く，画像を取り込むことで，講義を理解しやすくなったという感想が得られた。講師の口調なども伝えることができたため，講義を受けている雰囲気を

掴むことができたと考えられる。表示文もユーザ一人一人に見やすく設定でき，本システムは今後の大学で有効であるとの評価を得た。聴覚障害学生からは WEB ブラウザさえあれば利用できる点が魅力的だという感想も得られた。復唱者は大学で復唱を行う必要はなく，自宅からノートテイクを行うことも可能である。よって，ボランティア学生も増加することが期待される。

反対に②，③，④に関する評価は低く，聴覚障害学生にとって音声認識した文章は誤字が多く理解しづらかったと考えられる。実際に講師の音声聞き，復唱した際に，おちついて音声入力を行うことができなかった為，認識率が悪くなってしまったと考えられる。短くわかりやすく復唱できるように練習を行う必要があることがわかった。なお，実際にノートテイクされた文章（800字）を音声認識した結果，87.3%と高い認識率が得られた。このことから高い認識結果を得るため，文章を読み上げるように音声入力することが大切だとわかった。また，認識率が悪かった為，誤認識した文章を正しい文章に修正する時間がかかり⑧の評価が低かった事が考えられる。

8. 遅延調査実験

講師の話した内容を素早く伝えられるかについて良い評価が得られなかった。そこで，システムと ViaVoice の遅れを調べ，どの程度遅れが生じるか実験を行った。また大学生5人を対象にシステムの操作にかかる時間を調べた。

ViaVoice が誤認識した文字を訂正者が修正するまでの時間はどの程度か調べた結果，平均 8.51秒かかることがわかった。

8.1 実験項目

① 音声の遅延

録音した音声と，FCS を通じて配信された動画を比べどの程度遅延しているか調べた。マイクのサンプリングレートは自動的に決まり最高品質で 44kHz，最低品質で 8kHz である。

② 文字の遅延

入力した文字が，FCS を通じて聴覚障害学生側のパソコンに表示されるまでどの程度遅延しているか調べた。

③ ViaVoice の音声認識の遅延

復唱者が発声した後，ViaVoice がその音声を

認識し文字として出力されるまでどの程度遅延しているか調べた。

④ 講師の音声聞き、発声するまでの時間

講師の音声聞き、復唱者が発声するまでの時間はどの程度か調べた。

⑤ 動画の遅延

撮影した動画と、FCS を通じて配信された動画を比べどの程度遅延しているか調べた。動画サイズは 640×480、帯域幅 20kHz、品質 100%である。

⑥ 画像の表示の遅延

画像を送信した後、チャットウィンドウにその画像が読み込まれるまでどの程度遅延しているか調べた。

⑦ 講義資料から単語登録するまでの作業時間

ViaVoice に講義で使用する資料をテキスト化し、ViaVoice の辞書に単語登録する時間を調べた。

8.2 結果と考察

実験の結果を表 2 に示す。

表 2 システムの遅延時間

実験番号	時間(s)
①	0.1 未満
②	0.10
③	2.33
④	1.38
⑤	0.60
⑥	0.35
⑦	108.22

実験の結果から講師の音声聞き、文字が表示されるまでの時間は約 3.9 秒 (①+②+③+④) と、講義理解に支障が出るほど特別遅いわけではない。なお、③は 2.33 秒であったが、平均 1.59 秒ずつ 7.08 文字表示された。

講師の音声聞き、誤字を修正するまでに平均 8.51 秒かかる。今回の評価実験では ViaVoice が多くの誤認識をしてしまった為、それを修正する為の時間で講師の話した内容を素早く伝えられるかという質問に対して満足できないという結果になってしまったと考える。また、音声登録に 1 時間程度かかってしまうが、一度登録作業を行えば、今後は講義に使用する資料から専門用語を単語登録する時間 108.22 秒で本システムを使用できる。

正しく音声認識されれば、平均 2.53 秒の遅延時間で使用できることがわかった。

9. おわりに

音声認識による文章入力と画像による指示代名詞の表現を提案した。これにより、キーボード入力より速い速度で入力が可能になり、画像挿入機能によって、よりわかりやすい情報を聴覚障害学生に伝達できるようになった。また FCS を活かし、文章や音声を少ない遅延時間で聴覚障害学生に送信する事が可能になった。評価実験において、画像挿入機能により、講義を理解しやすくなったという評価が得られた。しかし、音声認識率が低く、誤認識が多発してしまった為、文章による講義理解が困難であり、文字が表示されるまで遅いという評価を得た。しかし、音声認識入力はキーボード入力よりも早く、システムに極端な遅延が見られなかった為、多くの誤認識を修正するまでの時間で文章が表示されるまで遅いという結果が得られたと考える。今後は、本システムを使いやすくする為に、ViaVoice を本システムと統合させ、音声認識の精度を向上させ、より修正しやすいシステムを構築し、多くの情報を正確に速く聴覚障害学生に伝えたい。

参考文献

- 1) IPtalk.
<http://iptalk.hp.infoseek.co.jp/>.
- 2) 放送教育における音声を利用した障害者支援 電子情報通信学会誌 Vol.91. No.12 2008 pp.1024-1029(日本放送協会放送技術研究所 今井 亨 日本アイ・ピー・エム株式会社東京基礎研究所 宮本 見太郎).
- 3) ViaVoice.
<http://japan.nuance.com/viavoice/>.
- 4) 音声認識を利用した聴覚障害学生学習保証システムについて 信学技報 TL2003, NLC2003-8, WIT2003-8 (2003-6) pp.43-48 (愛媛大学教育学部 立入哉, 愛媛県立宇和豊学校 井上かおり, 神戸総合医療介護福祉専門学校 宮武由佳).
- 5) 言い換えを利用したリスピーク方式によるスポーツ中継のリアルタイム字幕制作 電子情報通信学会論文誌 D-II Vol.J87-D-II No.2 pp.427-435 2004 年 2 月(松井 淳 他).
- 6) 続 FLASH ActionScript バイブル MX のツボ オーム社 上野了亨著.