音声認識による専門講義の情報保障の基礎的検討

加藤伸子 [1)]，三好茂樹 [2)]，内藤一郎 [1)]

筑波技術大学　産業技術学部　産業情報学科 [1)]
筑波技術大学　障害者高等教育研究支援センター　障害者支援研究部 [2)]

要旨：聴覚障害者の情報保障の一つとして，音声認識を用いた字幕の利用拡大が期待されている。しかし，音声認識には誤認識等の問題があるため，講師の音声をそのまま音声認識を行う方式での利用では，誤認識が多く校正が困難という課題がある。本研究では，大学の専門の講義に対して講師の音声に対して直接音声認識を行った場合の情報保障の実現可能性とその場合の校正の方式について検討を行った。実際に講義で音声認識を実施した実験の方法と実験の結果について述べる。
キーワード：聴覚障害，情報保障，音声認識，キーワード

1．はじめに

近年，大学等の高等教育に進学する聴覚障害者の数が飛躍的に増加し，講義における情報保障の必要性が高まってきている。本学においても学外講師等の授業において情報保障を行っている。大学における専門科目への情報保障では，できるだけ講師の発言に正確で情報量の多い情報保障方法が望まれる傾向があり，遠隔情報保障システムを用いて情報保障を行っている [1]。図1は，インターネット回線等を用いて，全国にいる文字通訳者の協力を得て実施している例であるが，情報量の多い専門の講義に対して文字通訳を行うには一定以上の技量を必要とすること，同時間帯に複数の講義が重なった場合の通訳者の確保等の課題がある。

一方，聴覚障害者の情報保障の一つとして，音声認識を用いた字幕の利用拡大が期待されている。音声認識には誤認識等の問題があるため，復唱方式による音声認識が提案されており [2]，復唱者の養成が望まれている。また，放送場面においては，話者の音声を直接音声認識するダイレクト方式がすでに実用化されているが，音声認識に適したニュースでのアナウンサーの発話に利用するなど，限定的なものとなっている [3]。

本研究では，大学の専門講義において講師の音声を直接音声認識させて作成した文字列を複数の校正者が修正するダイレクト方式による字幕作成方法の検討を行った。ダイレクト方式で情報保障を行うことの実現可能性とその際の校正の方法の検討を行うものである。

図１　遠隔情報保障システムを用いた講義の様子

図２　音声認識を用いた情報保障

2．音声認識字幕

2.1　音声認識を用いた情報保障

音声認識でこれまで広く利用されているのは，復唱方式による音声認識である。これは講師の発言を復唱者が復唱し，その音声を音声認識ソフトウェアで文字化する方法

である [2]。この方式では復唱者が音声認識ソフトウェアに適した話し方をすることで認識精度を向上させることができる。また，校正者が得られた文に含まれている誤認識文字を校正している（図 2 (a) 参照）。

これに対して，講師の音声を直接音声認識ソフトウェアで文字化し，得られた文に対して校正を行うのが，ダイレクト方式である（図 2 (b) 参照）。この方式では，復唱方式に比べて，校正者はより多くの誤りを訂正する必要があり，校正可能かどうかが問題となってくる。

2.2　ダイレクト方式のシステム構成

ダイレクト方式での実験を行うためのシステムとして以下のソフトウェア等を用いた。

●音声認識ソフトウェア：
AdvancedMedia 社の AmiVoice EX

●音声認識辞書：
前年度の講義保障で作成した字幕を元に専用の辞書を作成したものを利用した

●校正ソフトウェア：
連携作業用通信ソフトウェア SR-LAN

校正ソフトウェアで用いている SR-LAN は音声認識結果の文字列を自動的に複数人に振り分けることで，複数人同時に校正を行うことを可能にしているソフトであり [4]，復唱方式の講義保障でも同じソフトウェアが用いられている。

3. ダイレクト方式での字幕作成

3.1　講義におけるダイレクト方式の試行

実際の大学の専門講義において，ダイレクト方式での校正を試行した。

一般に文字通訳では経験と技量が必要となるが，音声認識を用いることでより幅広い人材での字幕作成が可能となるかどうかを確認するため，校正担当者としては比較的経験の浅い通訳者（通訳歴3年以内）に校正を依頼した。

●校正担当者：本学が養成をし，本学の講義で要約筆記を行った経験のある文字通訳者。同時に3〜4名で校正を試す。

●講義：産業情報学科の選択科目
「エコ・環境システム」

90 分の講義で，4回にわたり同時に校正する校正者の人数や音声遅延の時間等を調整しながら，校正を行った。音声認識ソフトウェアの音響学習機能はオンにして実施したが，音響学習のための特別な期間は設けていない。

3.1.1　校正者の選定と人数

同時に校正する人数を1人，2人，3人の場合で何回か試し校正者から意見を伺ったところ，

●校正は，同時に1人または2人で行う方式がよい
●望ましい校正人数には個人差がある

ことがわかった。同時に1人で校正をする場合には，前後の文節とのつなぎを考えて個人の判断のみで自由に校正できるメリットがあるのに対して，一旦校正文章がたまってしまうと，音声と校正しようとしている文章の間で時間的なズレが大きくなり，校正できなくなる場合が見られた。

図 3　付加情報を提示しての校正

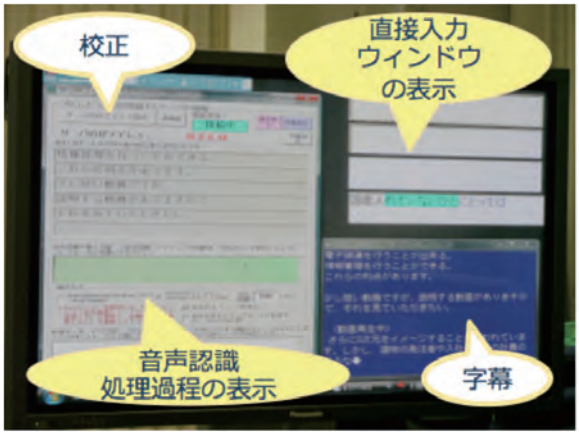

図 4　校正者への付加情報提示画面

3.1.2　音声遅延

校正者が聞く講師の音声には一定の時間の遅延をかけるものとし，校正者は遅延音声をヘッドホンで聞きながら字幕の校正を行った。講師の音声を直接聞いた場合には，振り分けられた校正予定の文章との時間差が大きく，より多くの聞き貯めが必要になるためである。校正者にとっては，音声認識が行われ校正担当者に校正する文が振り分けられた後に校正する段階で音声が聞こえてくるのが適当である。また，音声認識のシステムによっては，校正予定の文章をクリックすることでその音声を再生する機能を持つものもある。聴覚障害者にとってよりわかりやすい字幕を作成するためには，前後の文節や文章を確認し，文の接続を考える

必要があるため，校正担当文だけの音声を聞くのでは不十分であり，音声を遅延させる方法を利用した。

何回か自由に音声遅延時間を設定できる状況で試した結果，今回の実験における音声遅延の時間としては4～6秒が適当であった。

3.1.3 校正者に必要な情報

校正者にとって必要な情報として校正者にヒアリングを行ったところ，以下の2項目があげられた。

●音声認識ソフトウェアの処理状況
●直接入力表示

音声認識ソフトウェアは一定量の文章に対して認識結果を確定するため，認識途中と最終結果では文が異なる場合がみられる。前もってこの認識途中の文章を見ることが，校正の参考となる場合があるとの意見があり，音声認識ソフトウェアの処理状況を表示するものとした。

またSR-LANには，音声認識の過程で文章や単語が欠落した場合等に，直接入力で文を入力できる機能がある。この機能で校正者が入力をしている状況を他の校正者が把握することで，校正者間の連携がスムーズになるとの意見があった。このため直接入力の過程を示す画面を提示することとした。

実際の講義で試行を繰り返すことで，誤認識の少ない復唱方式とは異なるニーズが存在することがわかった。

3.2 講義におけるダイレクト方式の実験

3.2.1 実験方法

これまでの校正者の意見を受けて，付加情報を提示して校正をする実験を行った。付加情報を提示して実験を行っている様子を図3に，その時の提示画面を図4に示す。

実験に参加した校正者は各回4名で，同時に校正する人数を1人または2人とし，他の1人が校正文の訂正や直接入力を行う補助者として，交代で校正を行った。

90分の「エコ・環境システム」の2回の講義の中で，同時に校正する人数を切りかえるものとし，前半1人・後半2人で1回，前半2人・後半1人で1回の実験を行った。講義終了後に校正者に対してアンケートを行った。

アンケートの質問は，

●音声は聞き取りやすかったか
●講義の内容把握はできたか
●校正しやすかったか
●音声認識，直接入力ウィンドウ表示は役に立ったか

の4項目で，各々，7:非常にそう思う，4:どちらともいえない，1：非常にそう思わない，の7段階で評価をしてもらった。

また，これまでの実験を通して校正作業の条件として、どちらが入力をしやすいかを選択するアンケートを行った。

●同時に校正する人数（2名，それ以外）
●音声遅延（有，無）
●音声認識の表示（有，無）
●直接入力ウィンドウの表示（有，無）

3.2.2 実験結果

校正者によるアンケートを行った評価値の平均と標準偏差を図5に示す。

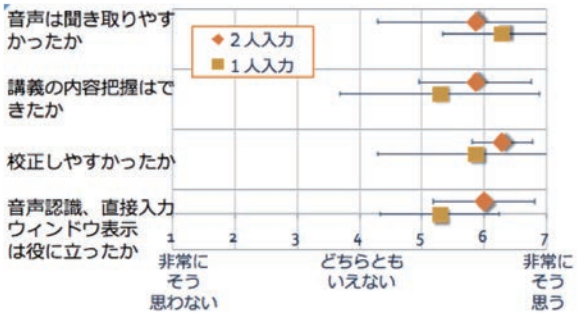

図5 校正者へのアンケート結果（1）

表1 校正作業の条件
「どちらが校正作業をしやすいか」

	2人	それ以外
同時に校正する人数	4人	0人

	有	無
音声遅延	4人	0人
音声認識部表示	4人	0人
直接入力ウィンドウ表示	4人	0人

実験の結果から，「校正しやすかったか」という質問に対する評価の平均値は6.0をこえて非常に高いものとなっている。1人入力でも同程度の値となっているが，標準偏差が非常に大きく，1人入力での校正のしやすさには個人差が大きいことがわかる。

また付加情報として提示している音声認識過程の表示や直接入力ウィンドウの表示が役に立ったかという問に対して，評価の平均値は5.0以上と高い値が得られている。

校正作業の条件の選択を行ったアンケートでは，4人全員が，同時に校正する人数は2人，音声遅延有り，音声認識部有り，直接入力ウィンドウの表示有りを選択した。

4. 異なる講義での音声認識実験

4.1 実験方法

大学のどのような講義であれば，音声認識+校正方式で情報保障が可能となるかを検討するために，これまで実験

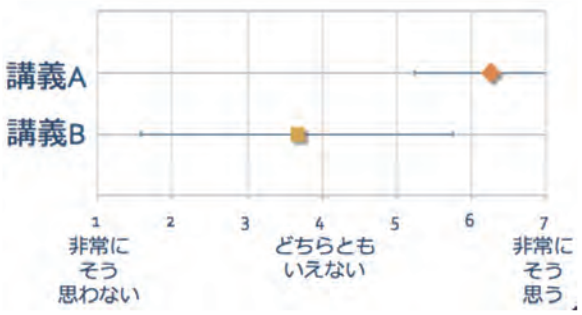

図 6　校正者へのアンケート結果 (2)
「音声認識+校正で情報保障が可能だと思うか」

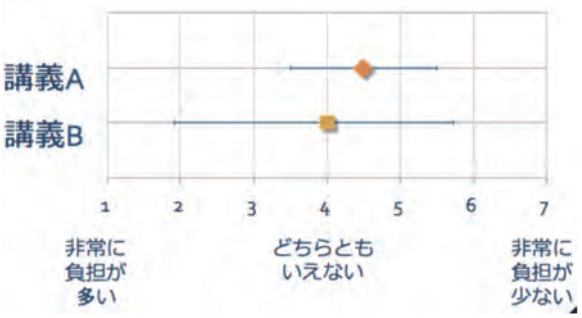

図 7　校正者へのアンケート結果 (3)
「要約筆記と比べて負担が少ないか」

を行ってきた講義（講義 A）とは異なる講義 B での実験を行い，2つの講義での校正を比較するアンケートを行った。

講義 A と講義 B の特徴を以下に述べる。

【講義 A】「エコ環境システム」
一文が短く，講師の話速が遅い。

【講義 B】「管理システム論」
一文が長く，講師の話速が速い。

音声認識ソフトウェアが認識する一文は，講師の話の間が一定時間以上であった場合に，文の切れ目と判断され，認識が行われる。間が短いと文の切れ目と判定されず，一文が長くなり，講師の発話から校正者への振り分けまでにより多くの時間を要することになる。このため，講義 A では音声の遅延時間を4秒としたが，講義 B では遅延時間を6秒に設定した。

講義 A，B を担当した後に，通訳者に以下のアンケートを行った。質問は，

●音声認識+校正で情報保障が可能だと思うか
●要約筆記と比べて負担が少ないか

の2項目で，各々，7:非常にそう思う，4:どちらともいえない，1：非常にそう思わない，の7段階で評価をしてもらった。アンケートの自由記述として，意見・感想を書いてもらった。

4.2　実験結果

講義 A，B を比較した実験結果を図 6，図 7 に示す。

「音声認識+校正で情報保障が可能だと思うか」の問に対して，講義 A では評価の平均値が 6.0 以上と非常に高かったのに対して，講義 B はどちらともいえないという結果であった（図 6 参照）。

また，通常のパソコン要約筆記との作業量の比較を問う「要約筆記と比べて負担が少ないか」という問いに対しては，講義 A，B 共に 4.0—5.0 程度であり，要約筆記と同程度かやや負担が少ない程度で，要約筆記より負担が多いものではないことがわかる。

アンケートの自由記述では，以下の記述が見られた。

●音声認識率が高ければ，経験が浅くても校正が可能ではないか。
●作業自体は要約筆記より楽。
●各自の担当のみを処理していくので，整文作業（句読点のバランス等）が困難。
●講義 B は話が早いので，要約筆記でも困難。

音声認識率の高い講義であれば，経験の浅い通訳者でも校正が可能であり,作業が容易であることが伺える。一方，担当の文章が文の一部であった場合等，整文作業が逆に困難になることも明らかになった。

5. まとめ

本稿では，講師の音声を直接音声認識ソフトウェアに入力し，校正を行うダイレクト方式での講義字幕作成の試みについて述べた。復唱者なしの方式で実験を行い校正者にアンケートを行った結果，講義によっては，校正者からはダイレクト方式（音声認識+校正）で情報保障が可能だと思う，との回答が得られた。また，今回は通訳経験が浅い文字通訳者に校正を依頼しており，経験が浅くてもある程度音声認識字幕の校正が可能であると考えられ，人材確保が進むことが期待される。また，音声認識+校正方式での校正を容易にするために，音声遅延 4 ～ 6 秒，音声認識の処理過程や直接入力ウィンドウなどの付加情報の提示が望まれることがわかった。

音声認識は辞書登録することで専門用語の認識が可能になるため，文字通訳者にとって困難な専門講義には適している面もあると考えられる。今後このような観点から，講義の選択方法等も検討を行っていく予定である。

参照文献

[1] 加藤伸子，河野純大，若月大輔，他．聴覚障害学生のための新任教員等の専門講義における情報保障の検討．筑波技術大学テクノレポート．2010; 17(2): p.7-11.

[2] 三好茂樹, 河野純大, 他. 音声認識字幕における円滑な連携作業を実現するためのソフトウェア開発と情報保障者の技能. 電子情報通信学会技術研究報告 . HIP, ヒューマン情報処理. 2009; 109(28):p.171-178.
[3] 今井亨. リアルタイム字幕放送のための音声認識. NHK 技研 R&D. 2012; 131: p.1-13.
[4] 日本聴覚障害学生高等教育支援ネットワーク. 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. PEPNet-Japan ホームページ (cited 2013-11-30), http://www.tsukuba-tech.ac.jp/ce/xoops/file/seika/onseininshiki-manual.pdf。

Basic Study of Real-time Speech Recognition Captioning for Major Subjects

KATO Nobuko[1)], MIYOSHI Shigeki[2)], NAITO Ichiro[1)]

[1)]Department of Industrial Information, Faculty of Industrial Technology,
Tsukuba University of Technology
[2)]Division of Research on Support for the Hearing and Visually Impaired,
Research and Support Center on Higher Education for the Hearing and Visually Impaired,
Tsukuba University of Technology

Abstract: Expanding use of real-time captioning using speech recognition for hearing-impaired students is expected. It is believed that university lectures contain many technical terms, and are suitable for speech recognition. However, direct recognition is often incorrect. We conducted an experiment to confirm the feasibility of direct speech recognition and the method of proofreading. In this paper, we describe the method and results of this experimental test of direct speech recognition of university lectures.

Keywords: Speech recognition, Real-time captioning, Communication support, Hearing impaired student