

平成 22 年 5 月 8 日現在

研究種目：基盤研究（C）

研究期間：2007～2009

課題番号：19500819

研究課題名（和文） 音声認識方式による聴覚障害学生の講義保障の研究

研究課題名（英文） Study of Lecture Security for Hearing-impaired Students  
Using Speech Recognition Technology

研究代表者

磯野 春雄（ISONO HARUO）

日本工業大学・工学部・教授

研究者番号：60337516

研究成果の概要（和文）：音声認識技術を利用して教員の音声をリアルタイムに字幕に変換し、聴覚障害学生に提示する字幕講義システムを開発した。本システムを大学の講義で運用した結果、教員の音声を約93%の高精度で字幕に変換して聴覚障害学生に提示可能となり、従来の字幕なし講義に比べて講義内容の理解度が向上することがわかった。また本システムは、従来のリスピーク方式の音声認識を用いた講義保障方式に比べ特別な訓練を必要とする復唱者が不要であるため実用性が高い。

研究成果の概要（英文）：A captioned lecture system that converts teachers' voices into captions in real-time by means of speech recognition technology and displays them to hearing-impaired students has been developed. A field test in university classes revealed that this system could convert the voice of teachers to captions with a high accuracy level of about 93%. The hearing-impaired students have stated that the lecture captioning system made the classes easier to understand than the classes that they had attended previously. Unlike the conventional lecture security system using speech recognition respeak method, there is no need for specially trained person to repeat the speech, and so the system has higher practicability.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	1,000,000	300,000	1,300,000
2008年度	900,000	270,000	1,170,000
2009年度	800,000	240,000	1,040,000
年度			
年度			
総計	2,700,000	810,000	3,510,000

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学

キーワード：授業学習支援システム

## 1. 研究開始当初の背景

近年、大学への進学率の上昇に伴い高等教育機関で学ぶ聴覚障害学生数は年々増加傾向にある。しかし、一部の高等教育機関を除いて聴覚障害学生に対する大学側の受け入れ状況や学習支援体制は十分とはいえない。従来、聴覚障害を持つ学生への講義内容の情報保障の手段には、手話通訳、要約筆記、ノートテイク（聴覚障害学生の隣席で講義内容をメモのように書き取り、それを聴覚障害学生が見る方法）などがある。このうち手話通訳、要約筆記は有償ボランティアとなっていることが多く、一般大学では聴覚障害学生が増加する中で大学側が見合う予算を捻出することは極めて困難な状況にある。このため同じ大学の学生ボランティアによるノートテイクという講義保障が行われている場合が多い。しかし、基本的に教員の講義の話す速さに追いついて書くことは困難であるほか、聴覚障害学生から質問ができない、ディスカッションができないなどの課題があり十分な講義保障とはならない。またノートテイク作業は疲労を伴うため短時間で交代しなければならず、多数の学生ボランティア要員を確保することの困難性という現実的課題に直面している。

このように国内外の大学における講義保障の手段は手話とノートテイク方式が中心であり、音声認識方式による講義保障の実施例は少ない。しかし、音声認識方式は筆記やキーボード入力に比べて長時間入力しても疲労が少なく、また約3倍の速度で文字入力可能であるため、教員の音声をほぼリアルタイムに字幕に変換して提示することが可能である。現段階では課題も多いが聴覚障害学生に対する講義保障として有望な方法である。

## 2. 研究の目的

本研究では将来の発展性を考慮し、音声認識技術を利用して教員の音声をリアルタイムに字幕に変換して聴覚障害者に提示する講義保障システムを研究対象とする。音声認識システムはすでに様々な用途に使用され始めているが、通常の音声認識使用とは異なり、大学の講義場面での使用では専門用語が多いほか、教員が音声入力を意識しない話し方をするため音声認識率が低く、そのままでは講義保障としては使用できない。

現状では音声認識方式の場合、どうしても誤認識は避けられない。このため音声認識に

よる講義保障方法としては教員の音声をインターネットを活用して復唱者に送信し、復唱者の音声をコンピュータで音声認識して字幕化する方法が実験的に試みられている。この方法はリスピーク方式と呼ばれ、特別に発声訓練した復唱者が必要であり、一般大学では復唱者の養成や訓練、費用などの面で難しい。さらに音声認識による誤字の修正や字幕表示の遅れの問題が発生するほか、聴覚障害学生と教員との双方向の質疑応答やゼミ学習等におけるグループ討議が困難など多くの課題が残されている。このように音声認識方式による講義保障システムは将来的には有望な方法と期待されるがまだ研究途上である。

本研究の目的は、一般大学でも導入可能な音声認識方式による聴覚障害学生のための講義保障システムの実現を目的としている。特に聴覚障害学生の大学における授業・学習支援という視点にたつて、聴覚障害学生の要望を取り入れながら音声認識方式によるリアルタイム講義保障システムを構築し、山積する多くの課題の解決を目指していく。今回は従来のリスピーク方式ではなく、ダイレクト方式の音声認識リアルタイム講義保障システムを試作し、大学の実際の講義で運用して聴覚障害学生を対象に授業評価や満足度に関する調査を実施して有効性を検証する。

## 3. 研究の方法

音声認識技術を利用した講義保障システムは、教員の音声をリアルタイムに字幕化し、誤認識文字を修正できる字幕の生成・修正部と字幕を無線LANを利用して聴覚障害学生に配信する字幕表示部とから構成される。本システムは聴覚障害学生に音声認識した字幕だけを提供する方法と、教員の映像に字幕付きで提供する方法を選択できる。

### (1) 音声認識方式による字幕の生成と修正

字幕生成部では音声入力する教員1名と誤認識文字を修正する学生1~2名で構成した(図1)。音声認識ソフトウェアには高性能で比較的 low コストな「Dragon Naturally Speaking 2005」を使用し、誤認識文字の修正にはPC要約筆記に用いられているIPTalkと呼ばれるソフトを使用してリアルタイムに字幕の入力・修正を行った。本方式は大学における講義や講演会など、1人の話者が長時間話す場合の情報保障に適している。

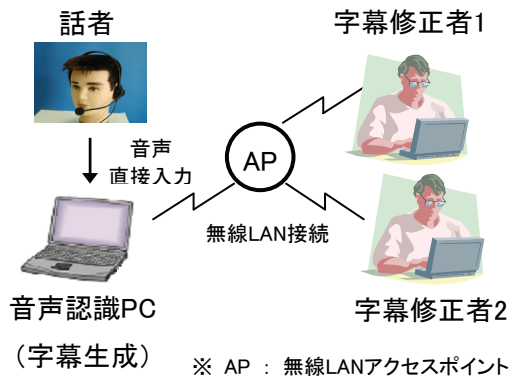


図1 音声認識による字幕の生成と修正

(2) 字幕および字幕付き映像の表示方式

字幕表示部では、無線LAN通信技術を使用し字幕表示部を無線化することで、場所を問わず字幕を見ることができるようにした。字幕だけを表示する方式として、字幕をIPTalk BroadcasterでWebブラウザに変換するWebブラウザ字幕配信方式を採用(図2)。この方式は多数の字幕表示装置に同時配信可能である。一方、字幕付き映像を配信できる方式では、実際に話者の映像を見ながら字幕を見ることができる(図3)。字幕表示装置を携帯型で無線LAN機能内蔵の小型液晶ディスプレイ端末を使用したことで字幕を見る場所の制約が大きく軽減された。

字幕 → Webブラウザ変換



図2 Webブラウザによる字幕配信方式

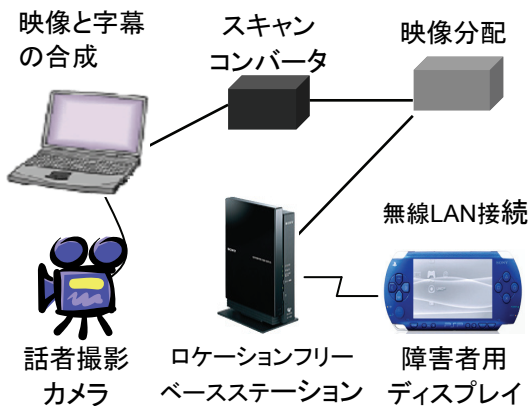


図3 字幕付き映像の配信方式

4. 研究成果

(1) リアルタイム音声認識字幕化装置の試作

音声認識技術と無線LAN技術を用いて、聴覚障害学生のために講義や講演会等における音声をリアルタイムに字幕変換し提示するシステムを試作した。試作したシステムは教員の音声をダイレクトに字幕化する方式を採用しており、従来の音声認識方式による情報保障で用いられているリスピーク方式のように特別な訓練が必要な復唱者が不要のため一般大学でも導入が容易である。本システムを大学の90分講義(合計15回実施)で使用した結果、字幕精度は平均87%、字幕表示遅れ時間3秒、誤認識字幕の修正後は字幕精度平均93%、字幕の表示遅れ時間15秒の性能が得られた。

(2) 字幕精度と講義内容の理解度の関係

音声認識方式では、人間による修正作業を行ったとしても100%完全で誤りなく字幕化することは困難である。そこで講義内容の理解に必要な字幕精度を明らかにするための主観評価実験を実施した。字幕精度を95%から80%まで4段階に変化させた字幕を被験者に提示し、5段階評価による内容理解度の実験を行った。5段階評価で評価値3.5を内容理解に差し支えない許容限界とした。その結果、字幕精度が90%で評価値が3.7となり、85%では評価値が3.2となった。このため少なくとも90%以上の字幕化精度が必要であることが明らかになった。

(3) 字幕の表示遅れ時間の許容限界

誤認識文字の修正を行うと字幕精度は向上するが、トレードオフとして字幕の表示遅れ時間が増大する。そこで字幕の表示遅れ時間を2秒から60秒まで6段階に変化させた字幕を提示して字幕の表示遅れに対する許容限界を5段階評価した。その結果、字幕の表示遅れ時間の許容限界は約10秒であることがわかった。このため教員の音声を約10秒以内に字幕化してリアルタイムに近い状態で提示することが聴覚障害学生の内容理解にとって重要であることが判明した。

(4) 誤認識文字の効率的な修正方式

音声認識方式では誤認識が避けられないため誤認識部分は人間による修正作業が必要である。修正作業が効率よく迅速に出来ない場合には字幕の表示が遅れ、聴覚障害学生の満足度を損なう。今回、効率よく字幕修正作業を行うための字幕修正方式を検討し、修正者が1人の場合や数名が連携しながら分担して修正する場合のほか字幕修正をタッチパネルで行う方法など、種々の修正方式に

ついて所要時間と作業効率との関係について実験した。その結果、タッチパネルよりもキーボードの方が修正時間が短いことが分かった。また字幕を修正する人数と修正効率の関係を検討した結果、修正者の人数が増加するにつれて修正作業効率は高くなるが、人数が3名以上では作業効率が飽和し、4人以上では逆に修正作業に競合が生じるため作業効率が低下することが分かった。このため修正者の人数は2～3名が望ましい。

#### (5) 講義に適した字幕表示方式の検討

音声認識結果を字幕表示する方式として、教員の映像と字幕を一緒に画面下部に表示する1画面表示形式(図4)と、講義資料と字幕を別々の表示端末に表示する2画面表示形式について検討した。その結果、聴覚障害学生には話者の表情が分かる1画面表示形式が好まれた。一方、聴覚障害学生と健聴学生と一緒に受講するケースにおいては、健聴学生は字幕の動きに注意が向くため講義に集中できないという意見が多かった。このため健聴学生には講義資料のみを表示し、聴覚障害学生には1画面表示形式または字幕のみの表示形式(図5)を選択視聴できる専用表示端末で対応している。

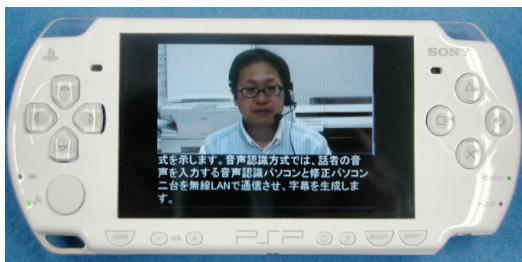


図4 字幕付き映像表示(1画面表示形式)

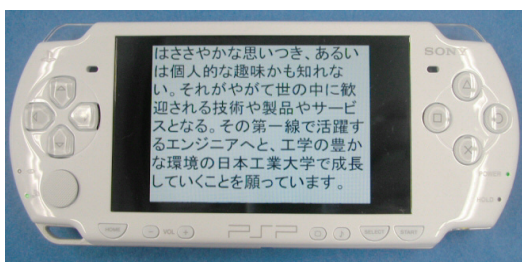


図5 字幕のみ表示する画面形式

#### (6) ゼミ形式講義やグループ討議における聴覚障害学生の支援

教員と聴覚障害学生によるゼミ形式の講義や複数の学生がグループ討議する場合のコミュニケーションシステムを構築した。グループ討議の参加者全員に無線LANで接続されたノートPCを用意し、各PCにはIPTalkソフトウェアをインストールする。

このあと参加者がキーボードで意見を入力することにより、参加者全員のPC画面上には参加者の意見が字幕で同時表示されるチャット形式で10人規模のグループ討議を実施した。その結果、この方式はキーボード使用に慣れている学生には違和感はなく、手話などに比べて導入が容易で少し慣れれば聴覚障害学生と健聴者との会話を円滑に行う手段として実用性が高いとの評価を得た。

#### (7) インターネットを利用した字幕講義の補完システム

聴覚障害学生への情報保障をより充実させるために音声認識によるリアルタイム字幕講義システムで誤認識した字幕情報をフォローアップする情報保障方法を検討した。本システムでは2人で誤認識を修正した場合、字幕精度93%が得られた。残り7%の誤認識文字については講義終了後にボランティア学生が修正し、誤りのない字幕講義データとして大学のサーバーからインターネット経由で提供する補完システムを実現した。これによってリアルタイム音声認識方式では完全には避けられない誤認識の課題を解決する情報保障として利用することができた。

#### (8) 音声認識データからの手話アニメーション映像の生成

聴覚障害者からは字幕のほかに手話映像による情報保障の要望があるため、音声認識で得られた字幕テキストデータ文字入力データをもとに手話アニメーション映像の生成・表示が可能なシステムを検討した。今回、手話アニメーションソフトには(株)日立製作所の開発した「MimehandII」を使用した。日常会話表現レベルでは手話アニメーション映像の生成ができることを確認できたが、大学の講義には専門用語や高度な表現が多く、手話アニメーション映像の生成ソフトで登録されている手話単語では不足することが分かった。また、手話表現にはいくつかの表現方法があり統一されていないことも課題となった。このため、本研究では手話アニメーション映像の生成・表示方式は、音声認識方式による字幕情報保障の補完的な利用とホームページの情報保障に応用した。音声認識の字幕データから直接手話アニメーションを生成表示するには、専門術用語に対応する手話辞書の整備や登録、手話表現方法の統一など多くの検討を要する。

#### (9) PC要約筆記方式との比較および聴覚障害学生による評価

本研究で実現した音声認識によるリアルタイム字幕講義システムの有効性を明らかにするため、代表的な情報保障方式であるパ

ソコン要約筆記方式との比較実験を実施し、聴覚障害学生から授業評価を受けた。教員1名と字幕修正者2名でリアルタイムに音声認識して字幕生成する本方式と、パソコン要約筆記の専門家2名がリアルタイムにキーボードで要約筆記しながら字幕生成する方法に関して、字幕精度や字幕の表示遅れ時間、聴覚障害学生に提供される情報量（字幕の総文字数）、運用コストについて比較した。

その結果、表1に示すように音声認識方式では字幕精度は字幕修正しない場合には87%、字幕修正した場合には93%に対し、要約筆記方式は99%であった。字幕表示時間は字幕修正しない場合には音声認識方式と要約筆記方式は共に3~5秒と同程度であった。字幕修正した場合の音声認識方式では字幕表示遅れ時間が平均15秒であった。

一方、PC要約筆記方式では教員が話す音声のすべてを字幕化しているわけではない。プロの要約者が教員音声の不要部分をカットしながら大意を判断しながら要約し字幕化している。このためキーボード入力による誤字は少なくなり字幕精度は99%と高い。比較評価実験によれば、要約筆記方式で聴覚障害者に提供された情報量（総字幕文字数）は教員の話した音声全体の74%である。これに対して音声認識方式では、93%~96%の情報量が提供されていることが分かった。

聴覚障害学生による授業評価アンケートによれば、音声認識方式はリアルタイム性や提供される情報量が多いこと、システム導入の容易さと運用コストが低いことが評価された。一方の要約筆記方式は要約筆記者の人件費等の運用コストがかかるが表示される字幕の誤字がなく見やすい点が評価された。

表1. 音声認識方式とPC要約筆記方式の比較

	音声認識 (修正無)	音声認識 (修正有)	PC要約 筆記
提供され た情報量	96%	93%	74%
字幕遅れ	平均3秒	平均15秒	平均5秒
字幕精度	87%	93%	99%
運用費用	小	中	大

#### (10)研究成果の位置づけと今後の展望

本研究では将来の発展性を考慮し、音声認識技術と無線LAN技術を利用して教員の音声をリアルタイムに字幕に変換して聴覚障害者に提示する講義保障システムを開発

した。本システムを大学の授業で教員1名と学生ボランティアによる字幕修正者2名で90分授業を15回運用した結果、教員の音声を約93%の字幕精度で聴覚障害学生に提供できることが分かった。現状では100%の認識率で字幕化することは達成できないが、聴覚障害学生による授業評価によれば90%以上の字幕精度があれば、誤認識した字幕が混入しても実際の授業においては前後の文脈から判断して内容理解が可能である。また、また本システムで採用した誤認識した字幕情報を授業後にインターネット経由で補完するシステムに対しても高い評価が得られた。

本システムはダイレクト方式の音声認識を採用しているため、特別な訓練を必要とする復唱者の人材確保が不要であり、低コストでシステム構築ができる点に大きな特徴がある。これまでに得られた研究成果から、一般大学でも導入可能な音声認識方式による聴覚障害学生のための講義保障システムを実現できる見通しが得られた。今後、コンピュータの性能や音声認識技術の進歩により、音声認識率が向上して誤認識が大幅に軽減されるほか、字幕表示のリアルタイム性は飛躍的に進歩することが期待できる。

現在、一般大学では聴覚障害学生の入学が増加する中で大学側が見合う予算を捻出することが困難な状況に置かれているため、ボランティア学生によるPCノートテイク方式が情報保障として一般的である。しかし、将来的には聴覚障害学生の情報保障方法は、音声認識方式に移行していくものと考えられる。本研究で実現したダイレクト方式による音声認識方式の講義保障方法は低コストで導入可能であることから、一般大学でも広く利用されていくものと考えられる。

今後の課題としては、本研究のような情報保障の技術的な研究開発とともに聴覚障害学生を受け入れる大学側および教職員の理解と協力、ボランティア学生の確保など、大学全体としての聴覚障害学生への支援体制の早急な整備が求められる。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計2件)

- ① 磯野春雄、音声認識方式による聴覚障害学生のための講義保障システム、日本工業大学研究報告、査読無、39巻、2009、29-30
- ② Haruo Isono, A Real-time Lecture Captioning System Using Speech Recognition Technology for Hearing-impaired Students, 日本工業大学研究報告、査読無、Vol. 38, No. 2、

[学会発表] (計 10 件)

- ① 磯野春雄、神戸 俊、IPtalk を利用したチャット形式による情報保障、第 5 回日本聴覚障害学生高等教育支援シンポジウム、第 3 部門、pp. 109, 2009 年 11 月 3 日、東京
- ② Haruo Isono, Junpei Yamaguchi, Captioned Lecture System for Hearing-impaired Students Using Speech Recognition Technology, The 10<sup>th</sup> Western Pacific Acoustics Conference (WESPAC X 2009), 2009 年 9 月 21 日、北京市
- ③ 磯野春雄、山口淳平、音声認識技術を利用した聴覚障害学生のための字幕講義システム、日本人間工学会、第 60 回記念大会、ID3-2, 2009 年 6 月 10 日、つくば市
- ④ 山口淳平、磯野春雄、聴覚障害学生のための講義保障方式の検討、電子情報通信学会技術研究報告、Vol. 108, No. 435, WIT2008-56, pp. 1-4, 2009 年 2 月 20 日、松山市
- ⑤ 山口淳平、磯野春雄、聴覚がい学生のための講義保障の取り組み (日本工業大学)、第 4 回日本聴覚障害学生高等教育支援シンポジウム、第 3 部門、2008 年 10 月 26 日、京都
- ⑥ 磯野春雄、音声認識方式による聴覚障害学生のためのリアルタイム講義保障システム、私立大学情報教育協会、平成 20 年度教育改革 IT 戦略大会、2008 年 9 月 3 日、東京
- ⑦ Junpei Yamaguchi, Haruo Isono, Real-time Captioning System for Hearing-impaired Students, Proc. Of JES/ESK Joint Symposium, JI-5, pp. 394-397, 2008 年 6 月 15 日、東京
- ⑧ 山口淳平、入倉啓輔、中野真弓、板橋佑太、磯野春雄、聴覚障害学生のためのリアルタイム字幕システム、2007 映像情報メディア学会冬季大会、10-1, 2007 年 12 月 19 日東京
- ⑨ Haruo Isono, Keisuke Irikura, Junpei Yamaguchi, Real-time Lecture Captioning System Using Speech Recognition Technology for Hearing-impaired Students, Proc. of 19<sup>th</sup> International Congress on Acoustics, Madrid, CAS-03000, 2007 年 9 月 4 日、マドリッド (スペイン)
- ⑩ 入倉啓輔、斎藤祐記、山中敬太、磯野春雄、音声認識技術を利用した聴覚障害学生のための講義保障システム、映像情報メディア学会技術報告、Vol. 31, No. 14,

[その他]

① 謝辞:

本研究は平成 19 年度から平成 21 年度の 3 年間にわたる科学研究費補助金 (基盤研究 (C)) により実施したものである。本研究の遂行に際して元筑波技術短期大学助教授の長谷川 洋博士と手話通訳士の宮下あけみ氏からは、多くの有益な資料と貴重な助言を賜り深く感謝申し上げます。また日本工業大学、情報工学専攻の大学院生、入倉啓輔氏、山口淳平氏ならびに本学の聴覚障害学生のご協力に対して厚く御礼申し上げます。

② ホームページ:

<http://leo.nit.ac.jp/~isono/>

## 6. 研究組織

(1) 研究代表者

磯野 春雄 (ISONO HARUO)

日本工業大学・工学部・教授

研究者番号: 60337516

(2) 研究分担者

なし

(3) 連携研究者

なし