

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 27 日現在

機関番号：12103

研究種目：基盤研究(B) (一般)

研究期間：2012～2015

課題番号：24300264

研究課題名(和文)聴覚障害学生のための専門教育に特化した授業支援システムの検討

研究課題名(英文)Study of captioning system in higher education for the hearing impaired students

研究代表者

加藤 伸子 (KATO, Nobuko)

筑波技術大学・産業技術学部・教授

研究者番号：90279555

交付決定額(研究期間全体)：(直接経費) 5,100,000円

研究成果の概要(和文)：本研究では、音声認識により作成された字幕が、専門教育の観点から聴覚障害学生にわかりやすいかどうかを判断する指標を作成すること、および誤認識を含む音声認識字幕の提示と聴覚障害者にとっての理解との関連を検討することを目的とする。

講義資料から自動的に抽出したキーワードに基づいて算出したキーワード再現率を提案し、その有効性を確認した。また、聴覚障害学生に誤りを含む音声認識字幕を提示した場合の問題点を明らかにし、授業内容を理解してもらうには、リアルタイム性の高い誤認識を含む字幕だけでは不十分で、修正した字幕を提示する必要があることがわかった。

研究成果の概要(英文)：The purpose of this study is to create an index of real-time captioning using speech recognition that indicates whether the captioning is easy to understand for the hearing impaired students. We have proposed keyword reproducibility as an index, and confirmed the effectiveness of proposed index in the experiment. We also have carried out experiments of presenting speech recognition caption including errors to the hearing impaired students. For having them understand the class contents by using speech recognition captioning, it was found to be important to present an accurate caption that was corrected simultaneously.

研究分野：コミュニケーション支援

キーワード：音声認識 聴覚障害者 直接校正 情報保障 専門用語 キーワード 障害者支援

1. 研究開始当初の背景

大学の専門の講義では、専門用語が多用され細かなニュアンスまでもくみ取るため、講師が話した内容をできるだけ全文に近い形で字幕として提示する事を希望する聴覚障害学生が増えてきている。このため提示される文字情報の量だけでも膨大なものである(90分の講義で18,000文字程度)。このような文字通訳(複数の通訳者がPCを用いて連係して文字を入力する方法)を行うことができる技量の高い通訳者の人数は国内でも限定されており、必要とされる全ての講義を担うことが困難である。

これまで我々は手話通訳やリアルタイム字幕、文字通訳などの情報保障により、遠隔より講義情報を保障する研究や聴覚障害者を支援する研究を進めてきた[1][2]。特に、音声認識・復唱技術の進歩によって、講師音声を復唱者が復唱してから音声認識を行い校正する、復唱方式(復唱+音声認識+校正)が実用レベルのものとなってきている[3]。しかし復唱者の養成は必ずしも容易ではないため、さらなる支援の整備・コストの削減のためには、復唱を行わず講師の音声に対して音声認識を行いその結果を校正する直接校正方式(音声認識+校正)での運用が期待されている。これまでの我々の検討では、

- ・ 経験の浅い担当者でも、直接校正方式での校正が可能である。
- ・ 講師の音声を直接音声認識にかけるため、認識結果には、文や単語の脱落や、単語の追加など、直接校正方式特有の現象が発生する。
- ・ 経験の浅い担当者では、100%の修正は不可能であり、校正結果の字幕には誤りを含む場合や、情報量が減る場合がある。
- ・ 文字通訳等と比較し、字幕表示までの遅延時間が増大する。

ことがわかっている。復唱方式であっても聴覚障害者に字幕を呈示することの問題点が指摘されている中[4][5]、さらに多くの誤りを含む可能性のある校正結果を聴覚障害学生に字幕として提示することは、学生に悪影響を及ぼす可能性がある。

このため、理工系の専門教育において音声認識を活用とした場合、どのような字幕であれば適用が可能なのか、表示方法を含め、検討することが急務である。

2. 研究の目的

本研究では、対象を大学での理工系の専門授業の情報保障に限定し、誤りを含むまたは情報量が減る可能性のある字幕について、専門教育の観点からのわかりやすさを判定するための指標を作成する。また、聴覚障害者にとって望ましい音声認識字幕の使用方法・表示方法を検討することを目的とする。

3. 研究の方法

本研究では、聴覚障害者のために作成された字幕が、専門教育の観点からどの程度わかりやすいのかを判定するための指標を作成すると共に、実際に誤りを含んだ可能性のある字幕が聴覚障害学生にとってどのように受け取られるか表示方法を含めて検討することを目的とする。そのために、以下の方法で研究を進める。

(1) 指標の作成

音声認識字幕の比較

復唱なしで直接音声認識を行った場合について、文字通訳で作成した字幕との比較を試みる。音声認識の指標として、認識率と正解率を以下のように定義する。本研究では形態素単位で認識率、正解率を計算するものとする。

- ・ 認識率 = (正解数) / (全形態素数)
- ・ 正解率 = $\frac{(\text{正解数}) - (\text{挿入誤り数})}{(\text{全形態素数})}$
- ・ 全形態素数 = 正解数 + 脱落誤り数 + 置換誤り数

キーワードに基づく指標の検討

大学の授業での文字通訳において、入力された文がどの程度原文を反映し話者の意図を伝えているかの評価を行った研究では、最低限注意すべき内容として文の主体やキーワードといった情報を伝えること、誤解を招くような誤変換や繰り返し表現等を極力生じさせないこと等が明らかになっている[6]。

音声認識を用いた字幕についても、同様に主体やキーワードが誤りなく最低限伝えるべき情報であると考えられ、このような観点からの指標が必要と考えられる。

そこで、講師の発言に含まれる専門用語等のキーワードを抽出し、キーワードがどの程度再現されているかを示すキーワード再現率を定義し、音声認識字幕を評価する指標となるかを検証する。

実験に用いるデータとしては、講師の音声を直接認識させるダイレクト方式を用いるものとし、講師の音声をマイクで収集し、アドバンスト・メディア社のAmiVoice EXを用いて音声認識を行った字幕を実験に用いた。また、比較対象の文字通訳で作成した字幕は、遠隔文字通訳で通訳を行ったものである。

(2) 誤りを含む字幕の表示について

字幕の表示においては、リアルタイム性と正確性の2つが求められる。音声を聞いて文字に変換するために、必ず時間遅れを含むことになるが、これを如何に短くし、リアルタイム性を保つかが、非常に重要になってくる。一方で、時間遅れを短くした結果、誤りを含んだままの字幕や、情報量が少ない字幕となり、正確性に欠けることになる場合があり、リアルタイム性と正確性はトレードオフの関係にある。このような状況において、リアルタイム性を優先し、

音声認識結果として誤りを含む字幕を提示した場合の修正実験を通して、聴覚障害学生はどの程度これを活用できるのか、問題点は何かを明らかにする。また、表示方法の工夫で理解を促進することができるのか、検討を行った。

4. 研究成果

(1) 音声認識字幕の比較とキーワードに基づく指標の検討

音声認識により作成した字幕の中で、正解率が約 60%の部分字幕 A、正解率が約 70%の部分字幕 B とし、同じ部分を文字通訳者が作成した字幕との比較を試みた。70%は音声認識率の平均と言われている精度である。A と B では、音声認識を行った場合に認識率や正解率に差があるが、文字通訳ではどちらも認識率・正解率が 90%以上であり、差がないことがわかった。認識率・正解率は、「正解文」に大きく左右され、語尾の表記をどうするか、言いよどみを表記するかどうか、などの整文の状況に応じて値が変化する。文字通訳の認識率・正解率が約 90%と抑えられているのは、文字通訳と音声認識を書き起こした「正解文」との整文の違いによるところが大きい。

本研究では、字幕の質を表す指標として、キーワードがどの程度再現されているかを表すキーワード再現率を提案する。

$$\text{キーワード再現率} = \frac{\text{正解キーワード数}}{\text{全キーワード数}}$$

ここで、キーワードは専門家である講師自身に選定してもらうことが最適であると考えられるが、毎回の講義から相当数のキーワードを選定することは講師に多大な負担をかけることになるため、ここでは、専門用語自動抽出システム TermExtract[7] を用いて、講義資料より抽出した重要度が閾値以上の単語とする。また、全キーワード数は正解文に含まれるキーワードの総数とする。

字幕 A、B についてキーワード再現率を算出し、横軸を認識率または正解率、縦軸をキーワード再現率としたグラフを図 1 に示す。

認識率とキーワード再現率、正解率とキーワード再現率は各々正の相関がある。また、キーワード再現率は正解文の整文の影響を受けにくい上、聴覚障害者にとって最低限伝えるべき内容であるキーワードが伝達できているかどうかを確認可能であり、大学での授業における情報保障としての音声認識の指標になると考えられる。

本研究においては、この指標を簡便に算出できるように Web 上で動作するシステムを作成した。試作システムでは、音声認識結果と正解文を入力することで指標が自動的に算出される。

(2) 聴覚障害学生に誤りを含む音声認識字幕を提示した場合の問題点と字幕の提示方式

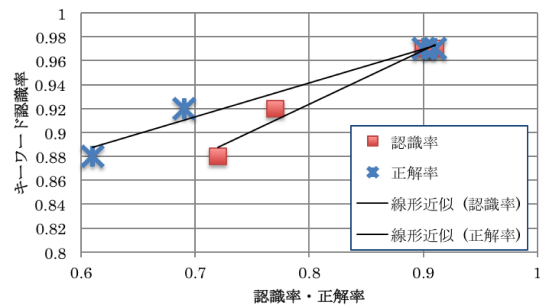


図 1 認識率・正解率とキーワード認識率の関係

聴覚障害学生に、音声認識結果の字幕を提示し、資料等を元に修正してもらった実験を行った。その結果、以下のことが判明した。

- 一目でわかる不要語の削除は容易であった(約 90%)。
- 配布資料に明記されたキーワードについては、資料を元にある程度修正が可能であったが(約 60%)、実時間以上の時間を要し、修正可能な割合の個人差が非常に大きかった。

また、実験の結果、被験者に共通して修正が困難であった誤りとして、以下の 3 つがあげられた。

- 他のキーワードへの置換誤り
元の字幕とは異なるキーワードへ置換された語句
例)【正解】3つのプロジェクト
【音声認識】3つのコレクト
- 誤ってキーワードが加わる挿入誤り
元の字幕には存在しないが、誤認識の結果キーワードとして認識された語句。
例)【正解】(えー)このプロジェクトに
【音声認識】エコのプロジェクトに
- 濁点の誤り
例)【正解】スマートグリッドは電力のインターネットとされています。
【音声認識】スマートグリッドは電力のインターネットとされています。

字幕データ中で明らかにおかしい文字や語句は容易に誤りと気付き、資料から該当する用語を探すことが可能であるが、一見正しい語句に置き換えられたもの、特に当該授業のキーワードの挿入や置換は、誤りに気付くのが困難であり、大きな誤解を与える恐れのあることがわかった。また、濁点の誤りは聴覚障害者に特徴的に見られる誤りの一つと考えられ、聴覚障害児の単語の表記の誤りを調べた研究においても、最も多くみられた書き誤りが濁点のつけ間違いと報告されている[8]。専門用語としては、濁点の有無を含めて正確に覚えることが必要であり、濁点の誤りに気付くにくい聴覚障害学生に提示する字幕では何らかの配慮が必要と考えられる。

さらに、修正には実時間以上の時間を要した上に、実験終了後に行ったアンケートでは、「内容が理解できたか」と「キーワードを修正できた割合」に相関が見られず、資料を見ながら修正を行うことができたとしても、講

義の理解にはつながらないことが推測され、講義の情報保障として課題であることがわかった。

音声認識システムを用いて作成した字幕は、誤認識を含んでいる場合が多い。誤認識を含んだままであればできるだけ早く表示できる。一方、誤認識を手で修正した字幕は、正確性は高いが修正に時間を要するために遅れて表示されることになる。そこで音声認識を行った結果の字幕のみ（誤りを含む字幕）を提示した場合と、音声認識字幕と修正字幕の双方を聴覚障害学生に提示した場合を比較した実験を行った。

音声認識した文章は、音声認識した内容をそのままにしたものは「誤り字幕」とする。また、誤りを含む字幕を修正した字幕を「修正字幕」とする。

実験終了後に行った理解度テストに関しては、誤り字幕のみを提示した場合に比べて、誤り+修正字幕を提示した場合は有意に点数が高かった。

また、同じく実験終了後に行ったアンケートでは、「字幕は講義を理解するために参考になりましたか」という問いに対し5段階に回答（1:非常にそう思わない, 5:非常にそう思う）してもらったところ、回答の平均値が誤り字幕のみでは 1.25、誤り字幕+修正字幕では、4.75 と有意に差があった。

実験の結果、誤りが多い字幕だけでは理解度が有意に低く、その場で修正を推測し、講義内容を把握するのは困難であると考えられる。また、誤り字幕と正解字幕の両方を提示した実験では、ある程度の理解度が得られたが、正解字幕のみを注視する傾向が見られた。すなわち、誤り字幕と正解字幕の両方に目を配るのは困難と見られ、修正字幕を注視すると、リアルタイム性の高い誤り字幕を確認できないと考えられる。

今回の研究の結果、キーワード再現率が音声認識字幕の指標となることが確認できた。また、研究で確認された聴覚障害学生に音声認識字幕を提示する場合に生じる問題点や表示方法の実験結果を踏まえて、今後、専門の講義においてアクティブ・ラーニングが重視されていく中、指標を元に利用方法を更に検討していく必要があることがわかった。

参考文献

- [1]加藤伸子,河野純大,三好茂樹,他:聴覚障害者の情報保障におけるパソコン要約筆記入力者に対するキーワード提示,ヒューマンインタフェース学会論文誌,Vol.9, No.2, pp.125-134, 2007.
- [2] 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,Vol.17, No.2, pp.7-11, 2010.
- [3] 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,Vol.109, No.28, pp.171-178, 2009.

[4] 中野聡子,他:聴覚障害者向け音声同時字幕システムの読みやすさに関する研究(1) - 改行効果に焦点をあてて, ヒューマンインタフェース学会論文誌, Vol.10, No.4, pp.435-444, 2008.

[5] 福島智,他,:音声認識技術を活用した字幕呈示システムの開発研究及び運用における諸課題-利用者の観点を中心に-,群馬大学教育学紀要 人文・社会科学編, Vol.55, pp.179-186, 2006.

[6] 白澤麻弓,松崎丈,吉川あゆみ,中野聡子,三好茂樹,河野純大,他:文字通訳による入力文の評価に関する研究,日本通訳学会 通訳翻訳研究, Vol.9, pp.141-157, 2009.

[7] 中川裕志,森辰則,湯本紘彰:出現頻度と接続頻度に基づく専門用語抽出,自然言語処理,Vol.10 No.1, pp.27-45, 2003.

[8] 我妻敏博:聴覚障害児における単語の表記上の誤りに関する研究,上越教育大学研究紀要, Vol.10, No.2, pp.155-164, 1991.

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)

加藤伸子,三好茂樹,内藤一郎:音声認識による専門講義の情報保障の基礎的検討,筑波技術大学テクノレポート, Vol.21, pp.1-6, 2014.

加藤伸子,佐藤知紘,三好茂樹,内藤一郎:聴覚障害学生のための音声認識字幕における誤認識の与える影響についての検討,電子情報通信学会技術研究報告(福祉情報工学), Vol.115, No.354, pp.75-78, 2015.

[産業財産権]

出願状況(計0件)

6. 研究組織

(1)研究代表者

加藤 伸子(KATO, Nobuko)
筑波技術大学・産業技術学部・教授
研究者番号:90279555

(2)研究分担者

内藤 一郎(NAITO, Ichiro)
筑波技術大学・産業技術学部・教授
研究者番号:00237182

三好 茂樹(MIYOSHI, Shigeki)
筑波技術大学・障害者高等教育支援センター・准教授
研究者番号:80310192