様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ６ 月 １３ 日現在

機関番号：３３９０７
研究種目：基盤研究(C)（一般）
研究期間：2016～2018
課題番号：１６Ｋ０１５７４

研究課題名（和文）パソコン要約筆記文の作成支援システムの開発と聴覚障害者支援に関する研究

研究課題名（英文）Research of development of PC captioning support system and support for the hearing impaired people

研究代表者
竹内　義則（Takeuchi, Yoshinori）

大同大学・情報学部・教授

研究者番号：６０３２４４６４

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：パソコン要約筆記文の作成支援システムの開発に関する研究を行った．要約筆記者が聞きなれていない語を自動音声認識によって講師の発話から検出して，要約筆記者に提示を行う．要約筆記者は，提示された語をファンクションキーにより要約筆記文に挿入できるシステムの研究を行った．
大学で行われている実際の講義映像の撮影を2年間行った．そのうちのいくつかの講義映像に対して，専門用語などの検出実験を行い，80%以上の検出性能を得た．

研究成果の学術的意義や社会的意義
パソコン要約筆記において要約筆記者が要約に必要な情報を簡単な操作で入力することができ，高品質な要約文を作成することが可能となる．また，これまでは要約筆記者が事前に講義資料を読み，その中の聞きなれない語を検出していたが，本研究では自動的に抽出しているため，その作業負荷が軽減される．高品質な要約文により，受講者の理解度が向上することが予想される．このことにより，聴覚障害者が教育を受ける環境が改善されると考えられる．

研究成果の概要（英文）：We have conducted research on the development of PC captioning support system. Unfamiliar words for the captionist are detected from the lecturer's speech by automatic speech recognition and presented to the captionist. The captionist can insert presented words into caption by pressing function keys.
We filmed actual lectures taking place at a university for two years. For some of the lecture videos, we conducted detection experiments and obtained more than 80% detection performance.

研究分野：福祉情報工学

キーワード：情報保障　聴覚障害　パソコン要約筆記　音声認識

１．研究開始当初の背景

本研究では，聴覚障害者が大学での高等教育を受ける際の情報保障について取り扱う．専門知識を持った外部講師に講義を依頼する場合，その講師は，手話ができないことが多い．この場合，パソコン要約筆記によって情報保障が行われてきた．これは，健聴者をその講義室に配置し，その場で講師の声をパソコンでタイプし，聴覚障害者に見せる手法である．

大学の講義において専門的な内容を扱うことが多く，専門用語などが頻出する．その為，要約筆記者はその講義に対しての知識を有する人が行うことが望ましい．また，そのような知識に加え，要約筆記におけるスキルを有する人材が求められる．このような人材を見つけるのは困難であり，講義に対する知識が無い人が要約筆記を行うことになってしまっているのが現状である．そうした場合，要約筆記者は聞き取りが慣れていない用語が発話されてしまうと，正確に聞き取ることが出来ず，字幕入力の誤りや，入力の遅れの原因となる．

２．研究の目的

本研究では，講義中に発話された内容を要約筆記者においてキーボード入力や聞き取ることが困難であると考えられる用語のことを「難入力語」と定義する．これを自動音声認識によって講師の発話から検出して，要約筆記者に提示を行うシステムを開発する．要約筆記者が用語を正確に聞き取ることができなかった場合でも，発話内容を視覚的に確認することが可能となる．また，検出された難入力語をショートカットキーで入力中の文章に挿入することができる機能を開発し，入力する際に難入力語がある場合に手間を省くことが可能となる．このシステムを用いて，専門性の高い講義についての知識が無い要約筆記者が要約筆記を行う際の支援をすることを目指す．

３．　研究の方法

３．１　概要

講義資料として講義で用いるスライドのテキストデータから難入力語を自動抽出し，Julius での認識による難入力語検出性能を向上させるために，講義それぞれにあう言語モデルを作成する．音声認識によって検出された難入力語は文字通訳用エディタに送信され，文字通訳者に提示される．図１に，提案するシステムの概要を示す．

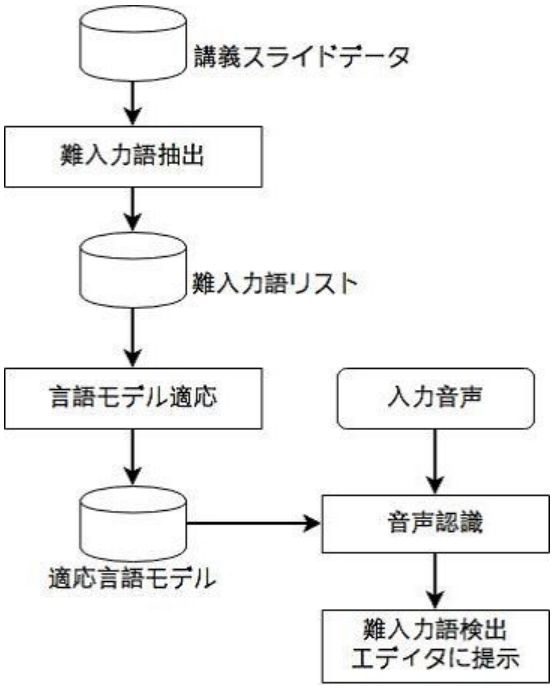

図 1　提案システム概要

３．２　難入力語自動抽出

まず，名古屋大学の聴覚障害学生支援室に所属する文字通訳者に，３．４で説明する，難入力語の提示・挿入機能を持つ文字通訳用のエディタを使ってもらった．その上で，エディタに提示されて欲しいと感じる難入力語を「経済学」,「音声言語学」,「細胞学」,「情報学」の４つの講義のスライドから手動で抽出してもらった.その結果や,文字通訳者からの意見をもとに,難入力語を以下の２つに分類した.

1.　専門用語・固有名詞
　聞き慣れていない人にとって，正確に聞き取ることや入力が困難
2.　数字，アルファベットを含む語
　聞き溜めや入力が困難

「聞き溜め」とは，文字通訳者が，講師の話した内容を入力しながら覚えておくことである．通常，キーボード入力よりも人の発話の方が速いため，入力しながら講師の話した内容を覚えておき，思い出しながら入力をする必要がある．

３．３　言語モデル適応

３．２で得られた難入力語のリストをクエリとして Web 検索を行い，その結果得られたテキストを学習データとして言語モデル学習を行い，Web 言語モデルを構築する．その後，ベースライン言語モデルを Web 言語モデルと重み付き線形補間し，適応言語モデルを生成する．この適応言語モデルを用いて認識を行うことで, 難入力語の検出性能が向上することが期待される．

３．４　文字通訳用エディタ

音声認識により検出された難入力語を提示し，またその難入力語をショートカットキーにより文章中に挿入する機能を持った文字通訳用のエディタを開発した．図 2 に，文字通訳用エディタを示す．

入力語が検出されると，エディタ下部の左側 F1-F3 のラベルの右にその難入力語が提示される．提示される順番は F1，F2，F3 の順番で，F3 まで提示された後は，再び F1 から順番に提示される．また，最も新しく提示された語のテキストボックスの背景は赤色で表示される．これらの難入力語は，対応する F1 から F3 のファンクションキーを押すことで，入力中の文章の末尾に挿入することができる．入力を行っている際，基本的に文字通訳者の視線は自分や相手の入力中の文章に集中している．そのため，入力中の視線移動を少なくするために，文章入力部分と難入力語の提示部分を近くに配置している．

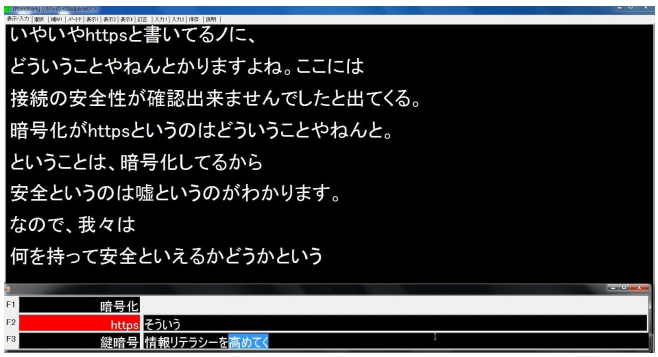

図 2　文字通訳用エディタ

３．５．実験

３．５．１　難入力語自動抽出

実験方法

4 つの講義について，講義スライドからの難入力語の自動抽出を行った．形態素解析器には MeCab を使用した．tf-idf 値の計算について，tf として講義スライドにおける単語頻度を，idf として日本語話し言葉コーパス(以下, CSJ)の 1 講演を 1 文書と見なした文書頻度に基づく値を使用した．本実験では，CSJ 内の学会講演，模擬講演の合計 2702 講演を用いた．また，抽出の際の閾値処理として，以下の 2 点を行った．

- スライド中に 1 回のみ出現し，CSJ 内には出現しない語の tf-idf 値を閾値とし，閾値未満の語を除外
- CSJ 内の 10 文書以上に出現する語を除外

2 つ目の閾値処理は，一般的な語であるが，スライド中に何度も出現する語を抽出しないために行った．本実験では，講義スライドから文字通訳者に手動で抽出してもらった語を正解の難入力語としている．

実験結果

難入力語抽出の再現率は 0.72-0.94，適合率は，0.27-0.45 であった．

考察

本実験において，適合率よりも再現率が重要だと考えられる．その理由として，誤抽出された語も，発話されなければ検出されることはなく，実際の文字通訳の場ではあまり問題にならない．一方で，未抽出語は，発話されても提示されない問題が生じる．

抽出できなかった語について，2 つの種類があった．1 つ目は，「老化のテロメア仮説」や「アレンヤングの経済思想」といったような，名詞や複合名詞の形になっていない語である．しかしながら,「テロメア仮説」や「アレンヤング」,「経済思想」といった語は抽出されているため，文字通訳を行う上で問題にならない．2 つ目は，閾値を下回った語である．idf 値の計算に CSJ

の文書を用いたが,「音声言語学」に近い内容の講演が CSJ 内にいくつか含まれており，例えば「デコーダ」や「ニューラルネットワーク」,「対数スペクトル包絡」のような語は CSJ 内に多数存在していた．その結果 tf-idf が低くなり除外された．これらの未抽出は，CSJ のような学術的なコーパスではなく，日常会話や新聞記事などの言語資源コーパスを用いることで軽減できる可能性がある．

また，誤抽出された語については，一般的な名詞で構成された複合名詞の過抽出が多く見られた．具体的には,「特殊例」や「重要概念」などである．これらは，構成する各名詞について閾値処理を行うなどすれば，軽減出来る可能性がある．しかしながら，一般的な名詞から成る複合名詞についても，入力に手間がかかるなどの理由で抽出されて欲しい，という文字通訳者の意見もあるため，今後難入力語の定義について，より精査する必要がある．

３．５．２　音声認識による難入力語の検出

実験方法

音声認識による，難入力語の検出実験を行った．音声認識エンジンには，Julius 4.3.1 を用いた．また，実際の講義でシステムを使う際には，難入力語が発話されたらすぐに文字通訳者に提示されるべきだということを考慮し，認識結果には第１パスのものを用いた．使用した音響モデルは，CSJ の性別非依存音韻モデルである．

まず，ベースライン言語モデルの構築を行った．学習テキストは３．５．１で述べた CSJ の学会講演，模擬講演の転記テキストであり，単語数は 690 万個であった．

次に,関連 Web テキスト収集による適応を行った.Web 検索のエンジンは Google を使用した．Web 検索のクエリには４．１で抽出した難入力語を用い，各クエリに対して検索結果の上位 50 件を収集した．その後，ベースライン言語モデルを Web 言語モデルと重み付き線形補間し，適応言語モデルを生成した，線形補間は以下の式で行った．

$$p_A(w|h) = \lambda p_B(w|h) + (1-\lambda)\, p_W(w|h) \qquad (1)$$

ここで，λは補間重み，pA，pB，pW はそれぞれ適応言語モデル，ベースライン言語モデル，Web 言語モデルの条件付き確率である．重みを 0.1 から 0.9 まで変化させた適応言語モデルを用いて難入力語の検出を行った．

実験結果

実験は以下の計算で評価する．

再現率=正しく検出された数/発話された数　(2)
適合率=正しく検出された数/検出された数　(3)
F 値=2(再現率)(適合率)/((再現率)+(適合率))　(4)

12 の講義音声について音声認識の言語モデルを適応させる際の補間重みλの値の変化したときの難入力語の平均検出率は，F 値で 0.794-0.830 であった．

考察

今回，収集された Web テキストには，講義トピックに関連していないものが多数存在した．その理由として，難入力語を単独でクエリとして用いていたことが考えられる．よりトピックに関連したテキストに絞って収集するためには，例えばあるスライドで抽出された難入力語の AND 検索を実行するなどが考えられる．

また，本実験では講義スライド全体を使って１つの言語モデルを構築しているが，スライド毎や，もしくは前後のスライドと合わせて 3 枚毎に言語モデルを構築しておき，講義中に動的に言語モデルを切り替えることで，より難入力語の検出性能が向上できるのではないかと考えられる．

ある講義では数式の「cos」,「sin」などが発話されていた．特に「cos」は講義内に 23 回発話されており,「sin」は 3 回発話されていた．しかし Julius で認識させてみたところ補間重みλの値が 0.1 では「cos」が 7 回，0.9 では 6 回,「sin」は 0.1 および 0.9 どちらも１回しか正しく認識されなかった．また「cos」は読みが「コサイン」であることから後ろの「(コ) サイン」が認識されて「sin」と誤認識されることがあった．ある講義では,「log」という難入力語の発話した回数が 2 回に対して検出が 7 回されていた.「6」という数字を「log」と誤認識していることが多く，そのため適合率が低下し F 値も低下していると考えられる．

３．５．３　文字通訳実験

実験方法

提案システムを用いた文字通訳実験を行った．実験には，情報セキュリティに関する「暗号と認証」という講義を用いた．事前に講義スライドから自動で難入力語の抽出を行った上で，

適応言語モデルを用いて音声認識を行い，難入力語が認識したタイミングを記録しておいた．抽出された難入力語の数は 200 であった．

実験の被験者は名古屋大学の障害学生支援室に所属するサポーター4 名であり，全員がプレゼンテーション形式の講演での文字通訳の経験がある．被験者はカウンターバランスを考慮して表 4 に示す 2 つのグループ A, B に分け, 2 人一組での連係入力を行った. 各グループに 2 回, 別の評価用映像（A，B）を見せ，文字通訳を行った．提案システムを用いる場合では，事前に行った音声認識で難入力語が検出されたタイミングで，図 4 のエディタ上に検出された難入力語が提示される．システムを用いない場合では，図 4 のエディタは用いず，IPtalk で普段用いている設定で文字通訳を行った．適応言語モデルを用いた音声認識による難入力語の検出性能は, 評価用映像 A では再現率 0.78, 適合率 0.87 であり, 評価用映像 B については再現率 0.78, 適合率 0.86 であった．また，それぞれの評価用映像を用いて文字通訳を行う前に，練習用映像を用いて文字通訳を行った．

実験終了後には，被験者に対して，行った文字通訳についてアンケートを採った．アンケート結果は紙面の都合上省略する．

また，文字通訳実験の評価観点として，以下の 2 つを用いた．

1.　　生成された字幕中に正しく入力された難入力語の割合

2.　　生成された字幕の各文が発話の意味内容を正しく伝えているか

2 つ目の評価観点に関して，発話の書き起こしと字幕の各文を比較し,「発話の内容が正しく伝わっているか」について,「まったくそう思わない」(1 点)から「とてもそう思う」(5 点)までの 5 段階評定を，健聴大学生 8 名に判断してもらった．

実験結果

2 つの評価観点の結果を表 1 に示す．

表 1　2 つの評価観点の結果

グループ	映像	システムの利用	割合	評定平均
A	A	**有**	**0.74(51/69)**	**3.8**
A	B	**無**	0.59(53/90)	3.1
B	A	**無**	0.46(32/69)	3.4
B	B	**有**	**0.67(60/90)**	**3.5**

考察

表 1 から，A，B どちらのグループにおいても，提案システムによる難入力語の検出・挿入機能を用いた場合の方が，高い割合で難入力語が入力されており，字幕が講義内容を正しく伝えているという結果となった．また，映像 A，B どちらにおいても，同様のことが言える．その要因として，正確に聞き取ることができなかった語が視覚的に確認できたことや，難入力語を聞き逃してしまった場合や入力中に忘れてしまった場合に，提示されていた難入力語を見ることで把握することができたことなどがあげられる．また，提案システムを用いない場合では，アルファベットを含む語（例えば,「IC カード」や「RC4」など）の入力に時間がかかっていたのに対し，提案システムを用いた場合では対応するファンクションキーを押すだけで入力することができ，入力時間が削減できていた．

文字通訳者からの意見として，数字は聞き溜めが難しいため，数字を聞き取って欲しいというものがあった．今回，スライド中に記載されていた数字については難入力語として抽出するようにしていたが，講師はスライド中にない数字を何度か発話していた．これに対処するために，スライド中から数字を難入力語として検出するのではなく，講師の発話を認識した際に，数字であると判断されたらそれを提示するように処理を行うことなどが考えられる．また，文字通訳者が必要ないと感じる語が多く検出され，難入力語提示用の枠がすぐに上書きされてしまって困った，という意見もあった．例えば今回の実験では,「暗号化」という語を難入力語として抽出したが，講師が「暗号化」を何度も発話し，3 つの提示用の枠が全て埋め尽くされてしまう，ということもあった．難入力語の提示用の枠を 5 つ程度に増やして欲しい，という意見もあったため，エディタのインターフェースについても今後検討する必要がある．また，文字通訳者がどのような語が提示されると助かるのかということについて，システムを用いた文字通訳を重ね，より明確にしていく必要がある．

４．研究成果

本研究では，パソコン文字通訳者の支援のために，難入力語を講師の発話から音声認識によって検出し，文字通訳者に提示し，文章中に挿入することができるシステムを開発した．パソコン要約筆記において要約筆記者が要約に必要な情報を簡単な操作で入力することができ，高品質な要約文を作成することが可能となった．また，これまでは要約筆記者が事前に講義資料を読み，その中の聞きなれない語を検出していたが，本研究では講義スライドからの難入力語

の自動抽出，および Web 検索による言語モデル適応をおこなっているため，その作業負荷が軽減された．そして，検出した難入力語を提示し，対応するキーを押すことで難入力語を文章中に挿入することができる文字通訳用のエディタを開発した．高品質な要約文により，受講者の理解度が向上することが予想される．このことにより，聴覚障害者が教育を受ける環境が改善されると考えられる．

スライドから文字通訳者が手動で抽出した難入力語を正解群とし，12 の講義に対して難入力語の自動抽出実験を行なった結果，0.7～0.9 程度の再現率を得た．提案システムを用いた文字通訳実験を行った結果，提案システムを用いた場合の方が，用いない場合に比べてより講義の内容を正しく伝える字幕が作成された．

今後，提案システムを用いた文字通訳実験を何度か実施する必要がある．文字通訳者が提案システムの使い方を十分理解した上で，どのような語が検出されると文字通訳を行う上で助けになるのかという意見を聞き，難入力語の定義を今後検討したい．また，難入力語が提示される枠の数や位置など，エディタのインターフェースについても実験を通じて改善していく必要がある．

５．主な発表論文等

〔雑誌論文〕（計　2　件）

1. Takeuchi Yoshinori, Kojima Daiki, Sano Shoya, Kanamori Shinji, “Detection of Input-Difficult Words by Automatic Speech Recognition for PC Captioning,” Computers Helping People with Special Needs, vol.10896, pp.195-202, 2018
2. Watanabe Daiki, Takeuchi Yoshinori, Matsumoto Tetsuya, Kudo Hiroaki, Ohnishi Noboru, “Communication Support System of Smart Glasses for the Hearing Impaired,” Computers Helping People with Special Needs, vol.10896, pp.225-232, 2018

〔学会発表〕（計　4　件）

1. 小島大輝，佐野翔哉，金森信治，竹内義則，“パソコン要約筆記のための音声認識による難入力語の検出と評価”，電子情報通信学会総合大会，2018
2. 渡辺大樹，松本哲也，竹内義則，工藤博章，大西昇，“聴覚障害者のための AR メガネを用いた音声理解支援システム”，電子情報通信学会スマートインフォメディアシステム研究会，2017
3. 渡辺大樹，松本哲也，竹内義則，工藤博章，大西昇，“聴覚障害者のための AR メガネを用いた音声理解”，情報処理学会アクセシビリティ研究会，2017
4. 池田直史，竹内義則，松本哲也，工藤博章，大西昇，“音声認識による難入力語の検出を用いた講義の文字通訳支援システム”，電子情報通信学会福祉情報工学研究会，2017