

## 要約筆記者の技術を活かした音声認識結果修正インタフェースに 関する基礎的研究

三好茂樹<sup>1)</sup>, 河野純大<sup>2)</sup>, 加藤伸子<sup>2)</sup>, 若月大輔<sup>2)</sup>, 白澤麻弓<sup>1)</sup>

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

キーワード: ノートテイク, 要約筆記, 音声認識, 修正方式

### 【成果の概要】

#### 1. はじめに

音声認識技術の発展に伴い、パーソナルコミュニケーションや1対1場面での音声認識技術の利用が現実的なものとなってきている。しかしながら、会議や講演等の場面において聴覚障害者が音声認識結果を有効に活用するためには、支援者による誤認識字幕の修正や、場面や用途に応じた整文が不可欠である。しかし、音声認識字幕を的確に修正できる支援者が不足している、システム自体が有料であるなどの理由により、最新技術の恩恵に預かれる利用者は限定的である。

このような状況に対応するため、本研究では、従来から聴覚障害者への文字による情報保障に携わってきた要約筆記者（ノートテイク）を最新のテクノロジーである音声認識を活用した情報保障の支援者として活用するために必要なシステムの開発を目標として、基礎的な調査研究を学長のリーダーシップによる教育研究等高度化推進事業経費を得て開始した。

より具体的には、ソフトウェアで実現した2つの修正インタフェースを開発・調整し、それらを用いて修正作業（キーボードやマウスを用いた通常のパソコン・スキルによる操作）を実施した後、その経験に関するアンケートの実施を行った。

#### 2. 方法

図1に、我々が開発した修正インタフェース機能を含む「音声認識によるリアルタイム字幕提示システム SR-LAN (エス・アール・ラン)」を示す。SR-LAN は、情報保障者の役割に応じたクライアント・ソフトウェア（復唱クライアントおよび修正クライアント）と、それらによって処理された文字データの送受信や蓄積を管理するサーバ・ソフトウェア（管理サーバ）で構成されている。授業を例に挙げると、まず、講師の発話音声、或いは講師の発話音声を復唱した音声

音声認識ソフトウェア（使用機材：アドバンストメディア社製 AmiVoice SP2）によって誤字脱字混じりの文字データが生成される。その文字データは同じ PC 上にインストールされた SR-LAN 復唱クライアントによって、管理サーバへ送られる。管理サーバへ送られたデータは、複数の修正クライアントへ送られる。各修正インタフェースで発話音声との照合・修正された文字データは管理サーバへ戻され、学生へと順次提示されてゆく。管理サーバには、2種類の修正インタフェースが実装されており、この管理サーバにて修正インタフェースを切り替えることができる。

修正クライアントによる研究対象者の作業の流れを図2に示す。修正クライアントの画面構成は大きく分けて3つのエリアに分けられる。講師の発話に合わせて復唱クライアントから順番に送られてくる音声認識データは管理サーバから各修正クライアントへ送信される。修正クライアントで受信した修正候補の音声認識データは、エリア A の各行に順番に蓄積されてゆく。図中、修正クライアント画面に、講師音声を例として、この例と通りに発話した場合に、各エリアに修正候補どのように蓄積・配置されるかを示した。エリア A に蓄積された修正候補は、研究対象者（修正者）によって、古い順番に確認・修正作業が実施され、発話順位を保持されながら、エリア C に蓄積されて行く。修正作業に際して、確認・修正作業が未着手なもの、確認・修正作業中もの、そして確認・修正作業が完了したものの背景色をそれぞれ色分けして伝える。各エリアの役割は、確認・修正作業を行うエリア B、主に確認・修正が未着手のものが配置されるエリア A、そして主に確認・修正作業が完了したものが配置されるエリア C というように、配置される位置が下にある程新しく、上にある程古い。

本研究では、図3のように構成を変更し実験を実施する。変更点は、復唱クライアントにて録音音声データの再生による音声認識を実施すること、そして、本実験では聴覚に障

がいのある学生が介在しないため、「学生への字幕提示」を実施しないという2点である。実験では、録音音声データの再生によって音声認識された誤字脱字混じりの字幕データを、修正インタフェース（修正クライアント）を用いて修正するという作業を研究対象者に体験させ、体験終了後に修正インタフェースの使用感についてアンケート調査を実施した。

尚、本研究の中心となる修正インタフェースに関して、「強制割当」と「自由選択」の2種類を用意した。「強制割当」では音声認識ソフトウェアから順次出力される文字データを管理サーバが各研究対象者（修正者）のノートパソコンへ交互に強制的に割当・送信してゆく方式である。即ち、ある研究対象者は、音声認識から出力された文字データはすべて手元のノートパソコン上で視認できるが、自分に割り当てられた分のデータ以外に操作することはできない。一方、「自由選択」では音声認識ソフトウェアから順次出力される文字データを管理サーバが各研究対象者のノートパソコンへ交互に送信してゆき、研究対象者が修正したい文字データを自由に選んで修正できる方式である。2つの修正インタフェースの切り替えはサーバ・ソフトウェア上のボタンで実施責任者や実施分担者が行える。

尚、本研究は筑波技術大学 研究倫理委員会の承認（承認番号 H 29-48）を得て実施した。

### 3. 結果

研究対象者は連係入力による PC ノートテイク経験のある4名であった。

ノートテイク（要約筆記者）の技能のうち、役立つ技能としては、「記憶した発話内容を長時間保持し、必要に応じて思い出す能力」が挙げられた。強制割当に関しては、「こちらでタイミングを計る必要がない」という意見があった。自由選択に関しては、「相手の入力を見て文をつなげる力」や「自分の入力スピードに合わせて、修正作業が割り振られるので、経験が浅い人とペアになる時には円滑に進むと思う」という意見が挙げられた。

字幕作成の効率や、字幕完成・提示までの遅延時間に関しては、「音声認識の認識精度に大きく左右されるので、比較できない」というコメントもあった。

完成・提示される字幕の精度に関しては、「（認識精度が揺らぐので）直しきれない誤認識が表示される可能性が PC テイクより高いと思う」というコメントが挙げられた。

修正作業のストレスに関しては、同等との回答から高いと

の回答あった。2つの修正手法の比較においては、強制割当に関しては「どのくらいの量、正確さの文が割り当てられるか直前まで見えない。長文になった時、余裕をなくす。」や「間違いが多いと直すストレスがある（直す=消して打つ、より「打つ」方が早い）」などが挙げられた。一方、自由選択に関しては、「どちらかというと自由選択の方がまし」や「強制（割当）に比べると少しストレスは低いが、認識精度によっては、最初から自分で打てた方が良い」があった。

2つの修正方法を比較してどちらが効率的な修正作業ができるかという質問に関しては、同等という意見から自由選択の方ができると答えており、強制割当の方を効率的と答えたものはいなかった。自由選択に関して肯定的な意見としては「ペアの入力スピードに対応できるから」や「どちらか一人がつかずいても、修正作業が進められる」が挙げられた。

強制割当と自由選択がそれぞれどのような状況に向いているかという質問に関しては下記のとおりであった。

「強制割当」に関して：

- ・「強制割当は認識精度が良く、細切れに上がってくる場合に向いているのではないかな」
- ・「認識結果が短く、コンスタントに出る場合」
- ・「認識率が良い字幕で、言いよどみ、繰り返し、文末処理程度の対応で済む場合」
- ・「認識率が良好な話し手」

「自由選択」に関して：

- ・「少しまとまって結果が表示される場合」
- ・「修正作業を担う人の入力スキルに差がある場合」

### 4. まとめ

本報告では研究対象者は4名ではあるが、以下の可能性を示すことができた。ノートテイク（要約筆記者）の技能のうち、役立つ技能としては「記憶した発話内容を長時間保持し、必要に応じて思い出す能力」が挙げられ、修正作業のインタフェースとしては概ね自由選択に関しては肯定的なコメント等があった。音声認識の認識精度によっては、最終的な字幕の精度や遅延は PC ノートテイクより劣る可能性も挙げられた。

今後は更に実験回数を増やし、学会発表等を目指したい。それらによって今後、要約筆記者（ノートテイク）の技術を活かした最適な修正インタフェースを見出していきたい。

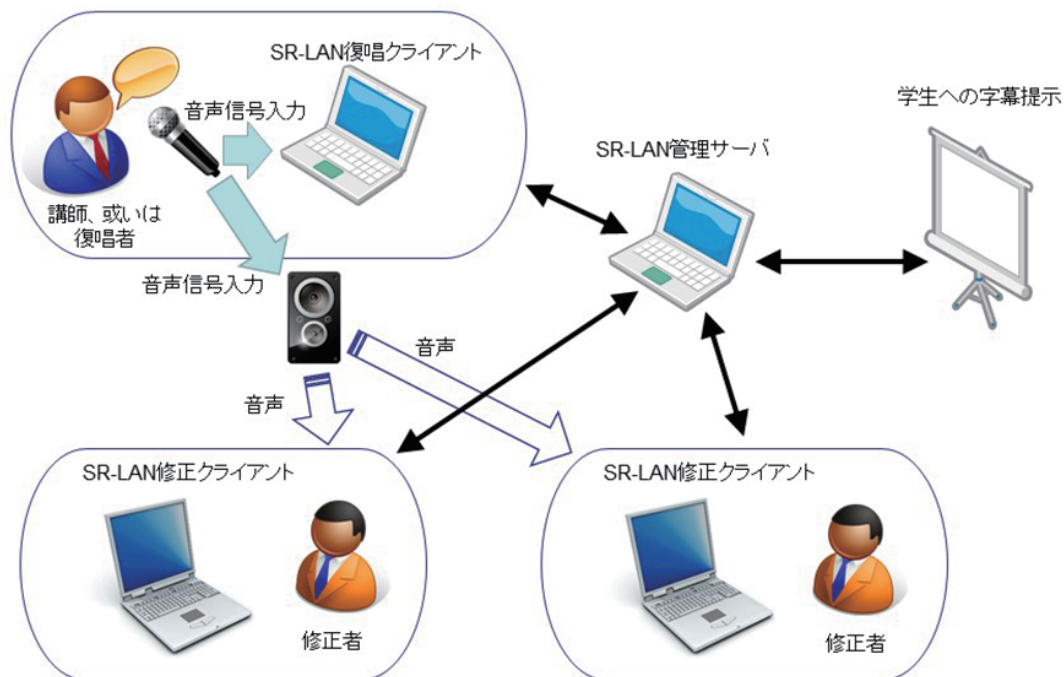


図1 音声認識によるリアルタイム字幕提示システム

講師音声(例):

こんにちは。  
今日は、天気ですね。  
先週は大雨で、  
会社からの帰り道、  
傘を差しても、  
服が濡れてしまいました。

エリアC

エリアB

エリアA



図2 修正クライアントによる研究対象者の作業の流れ

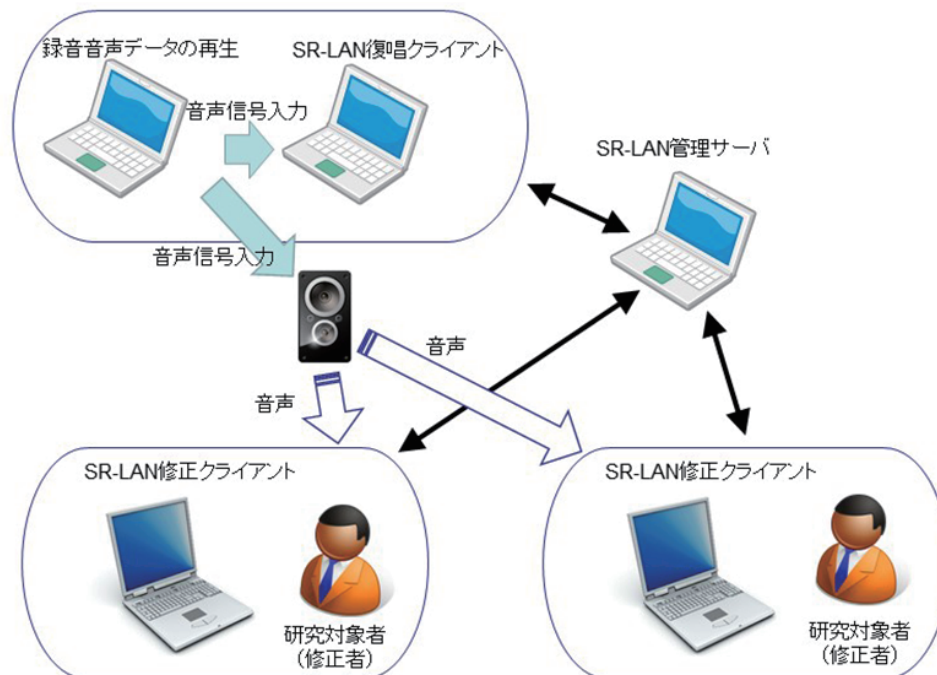


図3 実験用に構成を変更して利用するシステム