

音声認識を利用したリアルタイム字幕提示システム

小林 正幸 内野 権次 西川 俊* 石原 保志 三好 茂樹 森山 利治
 聴覚部教育方法開発センター *聴覚部客員研究員

要旨：音声通訳者（入力担当オペレータ）は、話者の音声聞きながら話者の音声内容を1台目のパソコンである文字入力用パソコンへ発話する。文字入力用パソコンは、音声通訳者が発話した音声を特殊な機器を必要としない安価な音声認識ソフト（ViaVoice98）のダイレクトディクテーションモードにより、テキストウィンドウに、設定した文字サイズ、フォント、文字色で文字化して提示するとともに、RS-232Cを介して2台目のパソコンである文字修正用パソコンへ文字コードとして送出する。文字修正用パソコンは文字入力用パソコンから送出された文字コードを読みやすい文字サイズ、フォント、文字色で字幕として表示する。修正担当オペレータは表示された字幕の誤字、脱字を確認し、字幕の修正を行う。このような手法により、リアルタイムで、音声認識を活用した字幕の入力作業と、字幕の誤字、脱字の修正作業が同時に行うことができるリアルタイム字幕提示システムを開発したので、このシステムの機能、特徴等について報告する。

キーワード：聴覚障害、音声認識、字幕、リアルタイム

1. はじめに

一般教育講義室には、高速で文字が入力できるDOS/VパソコンのWindows NTに対応した日本語高速入力システム（Windows対応ステノワードPCシステム）、校正器と字幕挿入装置（VIP-4100R）を連動させ、1セット目のステノワードPCシステムで発話の内容を先ずひらがなで入力、表示し、その後かな・漢字変換を行い、かな漢字混じり文を提示し、2セット目のステノワードPCシステムで、1セット目で入力した確定文章の誤字、脱字の修正を行い、より正確な字幕を話者の映像と共にリアルタイムで提示するWindows対応連弾入力方式RSVシステム¹⁾が設置されている。しかし、このシステムで使用されている高速で文字の入力が可能なステノワードPCキーボードは、数個のキーを同時に押して文字列を入力する特殊なキーボードであるため、専門に養成されたオペレータが入力する必要がある。

そこで、我々は、専門に養成されたオペレータが入力するのではなく、特殊な機器を必要としないパソコンで動作する安価な音声認識ソフト（ViaVoice98）を活用した、次のようなシステムを開発した。

音声通訳者（入力担当オペレータ）は、話者の音声を聞きながら話者の音声内容を1台目のパソコンである文字入力用パソコンへ発話する。文字入力用パソコンは、音声通訳者が発話した音声を音声認識ソフトのダイレク

トディクテーションモードにより、テキストウィンドウに、設定した文字サイズ、フォント、文字色で文字化して提示するとともに、RS-232Cを介して2台目のパソコンである文字修正用パソコンへ文字コードとして送出する。文字修正用パソコンは文字入力用パソコンから送出された文字コードを読みやすい文字サイズ、フォント、文字色で字幕として表示する。修正担当オペレータは表示された字幕の誤字、脱字を確認し、字幕の修正を行う。このような手法により、リアルタイムで、音声認識を活用した字幕の入力作業と、字幕の誤字、脱字の修正作業が同時に行うことができる特定話者の音声認識によるリアルタイム字幕提示システムを開発したので、このシステムの機能、特徴等について報告する。

2. ハードの構成

本システムの構成を図1に示す。また、機器の型名、メーカー、及び性能は次の通りである。

- (1) 文字入力用パソコン
 - ・SOLO9100XL
（日本ゲートウェイ2000）
 - ・CPU：PentiumII300MHz
 - ・RAM：96MB
- (2) マイク
 - ・MD425
（SENNHEISER）

・ダイナミックマイク

(3) 文字修正用パソコン

・DIMENSION XPS D333

(デルコンピュータ)

・CPU: Pentium II 333MHz

・RAM: 127MB

(4) 字幕合成器

・スーパー・てろっぱー

(コンパル)

・コンポジットビデオ入出力

・Sビデオ入出力

・カラーキー設定

3. ソフトの構成

本システムのソフトウェア構成は次の通りである。

(1) 文字入力用パソコン

①OS

・Microsoft Windows 98

(マイクロソフト)

②開発言語

・Microsoft Visual Basic
Ver. 5.0

(マイクロソフト)

③音声認識ソフト

・ViaVoice 98

(日本IBM)

④日本語変換システム

・ATOK12

(ジャストシステム)

⑤開発ソフト

・Visual Basicで開発した文字表示・送
出ソフト

(2) 文字修正用パソコン

①OS

・Microsoft Windows 95

(マイクロソフト)

②開発言語

・Microsoft Visual Basic
Ver. 5.0

(マイクロソフト)

③日本語変換システム

・ATOK12

(ジャストシステム)

④開発ソフト

・Visual Basicで開発した字幕提示ソフト

4. ソフトの機能、特徴

開発した文字入力用パソコンで動作する文字表示・送
出ソフトと、文字修正用パソコンで動作する字幕提示ソ
フトの機能、特徴は次の通りである。

(1) 文字表示・送出ソフト

①文字サイズ、フォント、文字色、背景色の設定が可能

②RS-232Cポートのパラメータ設定が可能

③入力した文章をテキストファイルとして保存可能

④入力された音声は音声認識ソフト (ViaVoice
98) のダイレクトディクテーションモードによりテ
キストウィンドウへ文字として提示

⑤提示された④の文字はリアルタイムに文字コードとし
てRS-232Cポートから送出

(2) 字幕提示ソフト

①文字サイズ、フォント、文字色、背景色の設定が可能

②RS-232Cポートのパラメータ設定が可能

③入力した文章をテキストファイルとして保存可能

④CRT画面に表示されている「修正ボタン」をマウス
でクリックすることで、誤字、脱字の修正が可能

⑤修正作業中に音声通訳者が音声で入力した文字コード
をRS-232Cの受信バッファメモリへ蓄積可能

⑥修正作業後、RS-232Cの受信バッファメモリへ
蓄積されていた文字コードを取り出し字幕として提示
可能

5. システムの動作

本システムの動作は次の通りである。

(1) 音声通訳者 (入力担当オペレータ) は話者の音声
を聞きながら、話者の音声内容を発話する。

(2) 音声通訳者が発話した音声はマイクを介して文字
入力パソコンのマイク入力端子へ入力される。

(3) 入力された音声はダイレクトディクテーションモ
ードで動作している音声認識ソフト (ViaVoice
98) で認識される。

(4) 認識された音声は、開発した文字表示・送出ソフ
トにより設定した文字サイズ、フォント、文字色、背景
色でテキストウィンドウに文字として提示されるととも
にRS-232Cを介して文字修正用パソコンへ文字コ
ードとして送出される。

(5) 文字修正用パソコンは開発した字幕提示ソフトに
より送出された文字コードを字幕として読みやすい文字
サイズ、フォント、文字色、背景色でCRT画面へ提示
する。

(6) 文字修正担当オペレータは、(5) のCRT画面に
提示された文章の誤字、脱字を判断した後、修正作業を
行うかどうかを決める。

修正する場合は（誤字、脱字がある場合）、次のような手順で作業を行う。

- ①修正担当オペレータは、文字修正用パソコンに表示されている「修正ボタン」をマウスでクリックし、誤字、脱字の修正を行う。
- ②文字修正用パソコンは開発した字幕提示ソフトにより、文字修正担当オペレータの修正作業中に、音声通訳者が文字入力用パソコンへ音声で入力した文字コードを、RS-232Cの受信バッファメモリに蓄積する。
- ③文字修正担当オペレータは、修正作業終了後、文字修

正用パソコンに表示されている「修正ボタン」をマウスでクリックする。

- ④文字修正用パソコンは②で蓄積したRS-232Cの受信バッファメモリの文字コード（修正担当オペレータの修正作業中に、音声通訳者が音声で入力した文字コード）を設定した文字サイズ、文字色、文字フォント、背景色で文章として文字修正用パソコンのCRT画面へ提示する。
- (7) CRT画面に提示された文章は、ビデオ合成器によりビデオカメラで撮影された話者の映像とともに字幕としてモニタへ提示される。

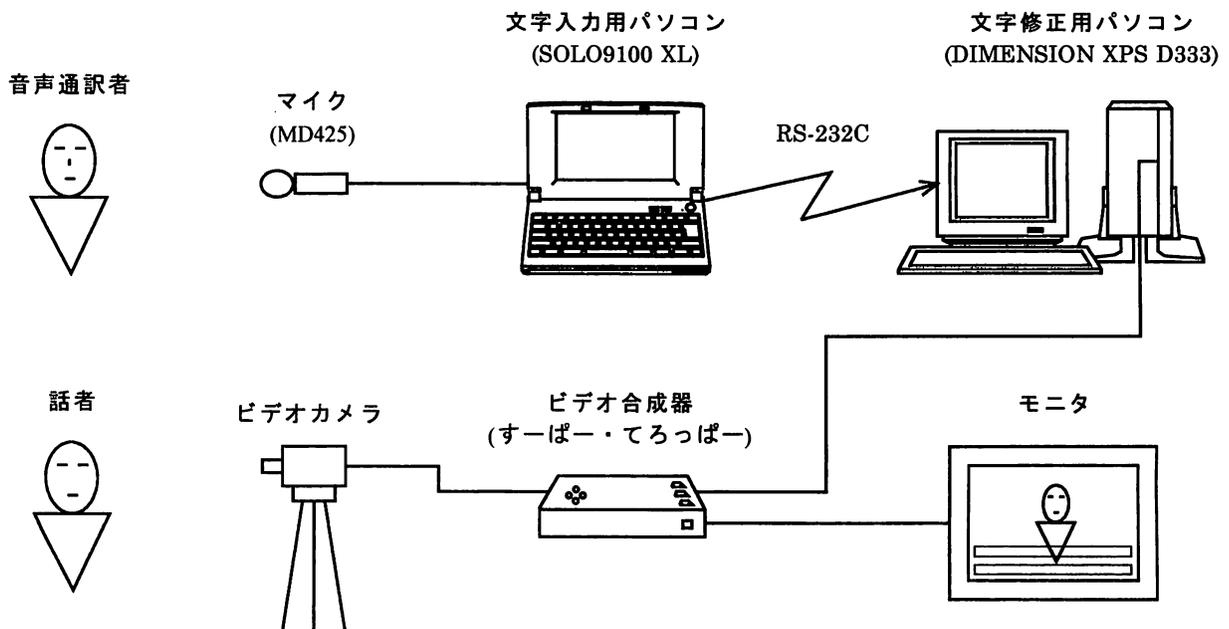


図1 システムの構成

6. 字幕提示画面

図2に文字修正用パソコンの字幕提示画面を示す。この画面は4領域で構成されている。

(1) 設定領域

字幕の文字色、文字サイズ、フォント、背景色等を設定する領域である。

(2) 映像表示領域

話者の映像等を表示する領域である。CRT画面上では青の背景色が表示されているが、モニタ（テレビ画面）では話者等の映像が提示される。

(3) 字幕提示領域

前述(1)で設定した文字色、文字サイズ、フォント、

背景色で、文字入力用パソコンで入力された字幕と文字修正用パソコンで誤字、脱字を修正した字幕を提示する領域である。背景色を青に設定すると、話者等の映像に字幕がスーパーインポーズされる。青以外の背景色に設定すると、設定した背景色に文字が上書きされる。

(4) 字幕修正ボタン領域

マウスで「修正ボタン」をクリックすると、文字入力用パソコンで入力された字幕の誤字、脱字の修正が可能となる。字幕の修正作業終了後、もう一度、「修正ボタン」をクリックすると、修正作業中に音声通訳者が入力した字幕（文字修正用パソコンのRS-232Cの受信バッファに蓄積されていた文字）が提示される。

7. 字幕提示例

図2に本システムで入力した字幕の提示例を示す。音声通訳者は音声認識の精度を向上させるため、2時間程度、音声認識ソフト（ViaVoice98）のエンロールを実行した。

音声通訳者が発話した内容は「これは音声認識を利用したリアルタイム字幕提示システムですまる」であり、提示された字幕は「...これは音声認識を利用したリアルタイム内幕を提示システムです。」であった。（但し、発話した文末の「まる」は音声認識ソフトにより自動的に「。」に変換される。）

アンダーライン「 」を付加した文字は音声認識ソフトの誤認識である。また、アンダーライン「 」を付加した文字は、開発したソフトの立ち上げ時のみの不具合であるため、特に問題ないと思われるが、ソフトの改善で削除可能である。

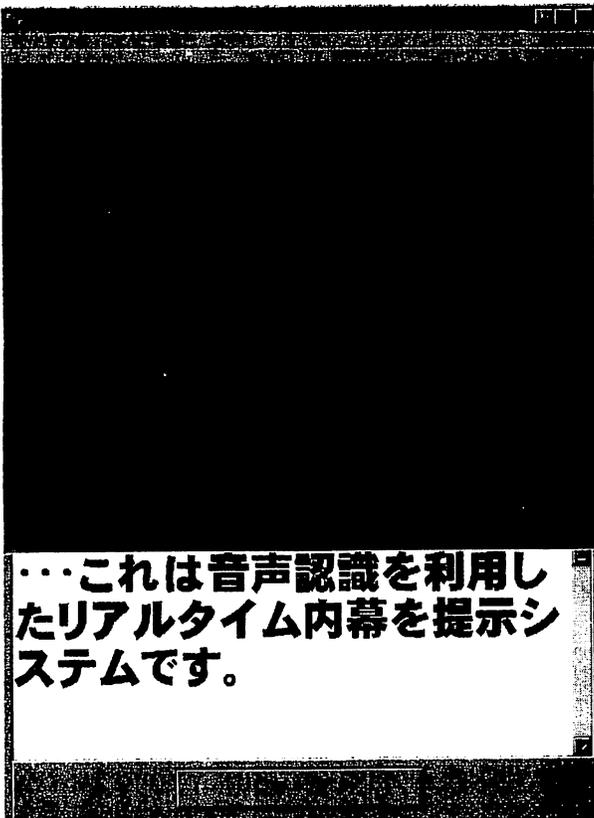


図2 字幕提示画面

8. おわりに

本研究では、一般に販売されている機器やソフトにより特定話者の音声認識によるリアルタイム字幕提示システムの開発について報告した。

今後の課題は次の通りである。

- (1) 本システムを様々な場面で使用し、システムの問題点、改良点を明確にする。
- (2) 字幕の字幕の行送りを行毎ではなく、徐々に上がるようにソフトを改良する。
- (3) 音声辞書登録を充実させ、音声認識の精度を向上させる。
- (4) 音声入力方式以外にも、開発済みの高速で文字の入力ができるステノワードPCキーボード¹⁾を文字入力パソコンへ接続し、システムの拡張性を検証する。

尚、本研究は本学平成10年度教育研究特別経費（学長裁量経費）により行われた。

参考文献

- 1) 小林正幸, 石原保志, 西川俊
聴覚障害者のための遠隔地でのキーボードの連弾入力によるリアルタイム字幕提示システム
ろう教育科学, Vol.40, No.3, pp.121-130
(Oct.,1998)
- 2) 小林正幸, 西川俊, 石原保志, 高橋秀知
聴覚障害者のためのキーボードの連弾入力方式によるリアルタイム字幕提示システム
映像情報メディア学会誌, Vol.51, No.6,
pp.886-895 (Jun.,1997)
- 3) 西川俊, 高橋秀知, 小林正幸, 石原保志,
柴田邦博
電子情報通信学会論文誌D-II, Vol.J78-D-II,
No.11, pp.1589-1597(Nov.,1995)
- 4) ViaVoice98マニュアル