音声認識による
リアルタイム
字幕作成システム

構築マニュアル

PEPNet-Japan
日本聴覚障害学生高等教育支援ネットワーク

音声認識による
リアルタイム
字幕作成システム

構築マニュアル

PEPNet-Japan
日本聴覚障害学生高等教育支援ネットワーク

目　次

１．はじめに

音声認識とは、人が話す音声に様々な工学的処理を施し、文字情報に変換する手法のことです。現在、様々な工夫が施された音声認識ソフトウェアを手頃な価格で入手することが可能です。それらのソフトウェアでは、予め多数の単語を辞書登録しておき、発話された音声の文脈を考慮して、高い認識精度を維持しようとする工夫が施されているのが一般的です。また、更に利用者個人の音声特徴を把握し、音声を矯正しなくてもある程度対応できるような工夫もあります。近年、音声認識技術による情報保障では、講師等の発話速度に追従し、高度な技能に頼らずに、ほぼ全文の文字化を実現することができる手法として期待されています。

しかしながら、大まかに言って、通常の会話や講義での発話スタイルのまま音声認識ソフトウェアを利用した場合ではその認識精度は 60～70%台、また未経験者が明瞭に発話するように意識した場合では 80%台、ある程度経験を積んで初めて 95%前後という高い認識率を実現できるようになるというのが実情です。音声認識ソフトウェアからの出力結果には、必ず誤字脱字が含まれるという大きな問題があり、利用にあたっては注意を必要とします。

以下に、同じ文章を音声認識ソフトウェアで認識した精度の異なる結果を提示します。

> 一件、市販されている音声認識ソフトとパソコンの組み合わせで気軽に住んでいる。試してみると、場合によってはそこそこの字幕制度で提示できていると庁舎は感じる。松田の、話沖縄、字幕を見ると、あまりわかりにくさは関係ないため、しかし聴こえる者が考える事、音声認識による沈黙は若やしない。
>
> 音声を沖縄金額を見ることと、地獄の運用を見ることでは、全く字幕の分かりやすさに対する以上異なる。特に選定の聴覚障害学生にとっては、後世に頼ったと石が困難であるため、５認識の最も意味の類推が聴こえる学生本当容易ではない。５意識が仮にほとんどなかったとしても、そもそも話し言葉をそのまま文字化した字幕は、分かりやすいものではない。話し言葉には書き言葉にはあまり見られないボイラーが多く含まれている。

これで認識精度８０%程度です。これでは正しい情報を推測することは不可能でしょう。次の文章はどうでしょう。

> 一件､市販されている音声認識ソフトとパソコンの組み合わせで気軽にできそうに見える。試してみると、場合によってはそこそこの字幕制度で提示できていると庁舎は感じる。松田の、話を聞きながら、字幕を見ると、あまりわかりにくさは関係ないため、しかし聴こえる者が考えるほど、音声認識による沈黙は分かりやすくない。
>
> 音声を聞きながら金額を見ることと、地獄の運用を見ることでは、全く字幕の分かりやすさに対する印象が異なる。特に選定の聴覚障害学生にとっては、音声に頼った類推が困難であるため、５認識の最も意味の類推が聴こえる学生本当容易ではない。５認識が仮にほとんどなかったとしても、そもそも話し言葉をそのまま文字化した字幕は、分かりやすいものではない。話し言葉には書き言葉にはあまり見られないボイラーが多く含まれている。

認識精度は９０％程度です。かなり良いように思いますが、元の文章を理解するためには、誤変換を正しく提示しなければなりません。

一見、市販されている音声認識ソフトとパソコンの組み合わせで気軽にできそうに見える。試してみると、場合によってはそこそこの字幕精度で提示できていると聴者は感じる。なぜなら、話を聞きながら、字幕を見ると、あまりわかりにくさは感じられないため、しかし聴こえる者が考えるほど、音声認識による字幕は分かりやすくない。

音声を聞きながら字幕を見ることと、字幕のみを見ることでは、全く字幕の分かりやすさに対する印象が異なる。特に先天の聴覚障害学生にとっては、音声に頼った類推が困難であるため、誤認識の元の意味の類推が聴こえる学生ほど容易ではない。誤認識が仮にほとんどなかったとしても、そもそも話し言葉をそのまま文字化した字幕は、分かりやすいものではない。話し言葉には書き言葉にはあまり見られない文法エラーが多く含まれている。

こちらが読み上げていた原文です。誤変換されていた単語には線を付けています。

では実際に音声認識を用いた情報保障を実施する場合には、どのような体制で実施されているのでしょう。群馬大学及び筑波技術大学で実施している様子を一部紹介します。

写真１　教室内で修正作業を２名体制で行っている様子（群馬大学）

写真２　別室で復唱・修正作業を４名体制で行っている様子（群馬大学）

写真３　別室で復唱・修正作業を３名体制で行っている様子（筑波技術大学）

写真４　無音ブース内で復唱・修正作業を３名体制で行っている様子。復唱にはマスクマイクを使用している。（筑波技術大学）

多くの人員で字幕作成を行っていることがお分かりになりましたでしょうか？

現在の音声認識技術を用いて満足のいく字幕品質を保つためには、情報保障者の役割として講師等の音声を聞き取りその音声を発話し直して音声認識ソフトウェアに入力する復唱者が必要となります。また、誤字脱字を校正する校正者も必要となります。これら２つの役割を果たすためには、様々な機器やソフトウェアも必要となります。

このような情報保障に必要な機器等の準備が揃ったとしても、復唱者が音声認識ソフトウェアになれていない場合には、著しく精度の落ちた字幕が作られてしまいます。手書きのノートテイクやＰＣ要約筆記と比較しても最初の時点で日本語が書けない、または入力できないという状況は発生し得ませんが、音声認識による字幕ではシステム構築や技能レベルに関するハードルが高く、結果的にそれに近いような事態となる可能性があります。このように、現時点では情報保障実施までに多くの労力を他の手法よりも要するという特徴もあります。

その反面、音声入力による字幕の自動生成技術は将来的にも多くの可能性を秘めていると言えます。このマニュアルで紹介する手法以外にも、今後の音声認識技術の発展に合わせて様々な実施形態も考えられるでしょう。

本マニュアルの目的は、ＰＣ要約筆記の一人入力や連係入力で、すでに情報保障を実施している方々が、先駆的な試みとして音声認識による字幕提示を実施し、その経験を通して現状と可能性について正しく知って頂くこと、そしてそのために必要な最低限のシステム構築ができるようにすることにあります。

音声認識による情報保障のためのシステム構築や練習は簡単ではありません。しかし練習を繰り返すことで、音声認識ソフトウェアの認識精度を高める発話方法や復唱方法、校正の注意点、そして連携作業のコツが掴めてくるはずです。

このマニュアルを通して日本国内で特定の団体や研究グループだけが実現している試みを、より身近に捉えて頂ければ幸いです。

本マニュアルでは、復唱方式を用いた音声認識によるリアルタイム字幕作成システムを構築する際に必要な機器構成やソフトウェア、そして情報保障者に必要なスキルについて触れています。工学的なシステム構成については、いくつかの研究グループによって様々な構成で実施されています。本マニュアルでは、システムを構築する一手法について説明し、最低限、システムを実現できる手順を紹介します。

２．音声認識同時字幕システムの概要

前項で述べたとおり、現在のところ、音声認識ソフトウェアを利用したリアルタイム字幕作成では、講義中の講師の音声を直接音声認識ソフトウェアに入れただけでは、正しい字幕が出力されることは稀です。このために、通常、講師の音声を直接利用するのではなく、情報保障者が音声認識ソフトウェアに適した音声で発話し直し（復唱）、その音声を音声認識ソフトウェアに入力することで、第一段階の字幕精度を上げる方法が用いられています。これを復唱方式と言います。

図１に復唱方式による音声認識同時字幕システムの概略図を示します。この復唱方式では、まず文字化すべき講師の音声を復唱担当者が聴取します。聴取した音声を元音声の内容のまま、または音声認識ソフトウェアに適合した文語的な言い回しで発話し直します。復唱担当者の音声はマイクロホン等のハードウェアを介して復唱担当者用のＰＣにインストールされた音声認識ソフトウェアに入力されます。音声認識ソフトウェアは音声-文字変換を実施し、文字データを出力します。この文字データには誤字脱字が確率的に必ず含まれてしまいます。次に、この出力された文字データは復唱担当者用のＰＣにインストールされた通信ソフトウェアに入力され、校正担当者用のＰＣに送信されます。この際、直接校正担当者用のＰＣに送られる場合と、サーバと呼ばれる全体を管理する役割を担ったハードおよびソフトウェアを介して送られる場合があります（本マニュアルで紹介するソフトウェアの場合にはサーバ・ソフトウェアを介します）。校正担当者用ＰＣに送られた誤字脱字混じりの文字データは、校正担当者によって校正作業が実施されます。復唱担当者の認識精度によっては、校正担当者は複数人で同時に校正しなければならない場合もあります。校正後の文字データは、校正担当者用ＰＣから表示用ＰＣに送られ、聴覚障害学生に提示されます。このような流れで字幕をリアルタイムに作成するのが、復唱方式による音声認識同時字幕システムです。

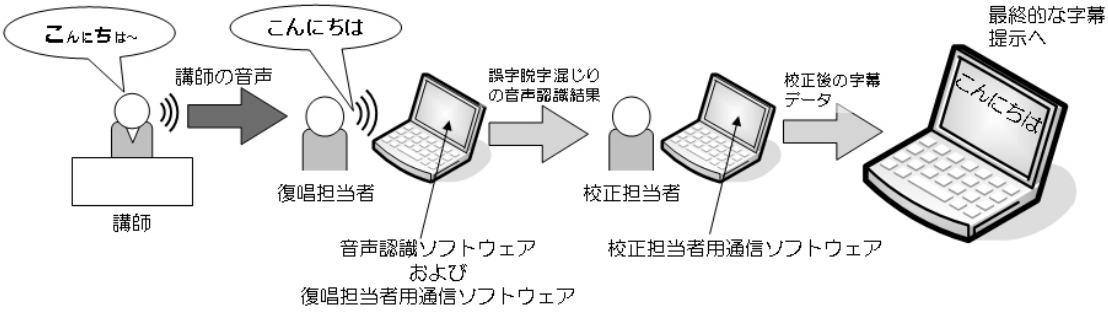


図１　復唱方式による音声認識同時字幕システムの概略図

システム構成は、復唱を別室で行うか、教室内で行うか、校正担当者はどこで作業するかなどによって様々に変わります。また、別室で行う場合には教室音声の送受信方法を検討する必要があります。これらの方法は実際に情報保障を行う環境にあわせてご検討いただければと思いますが、一般的には以下のような形態が考えられます。

２－１．別室から学内ＬＡＮやインターネットを介して情報保障を実施する場合

復唱方式による音声認識同時字幕の場合、通常のマイクロホン利用時には復唱担当者の音声自体が講義室内で広がり、講義の阻害要因となります。そのために作業は別室で行うのが一般的です。

ネットワークを利用して、別室で作業を実施する場合には、通信用の機材を別途用意する必要があります。各機器類の接続図の一例を図２に示します。

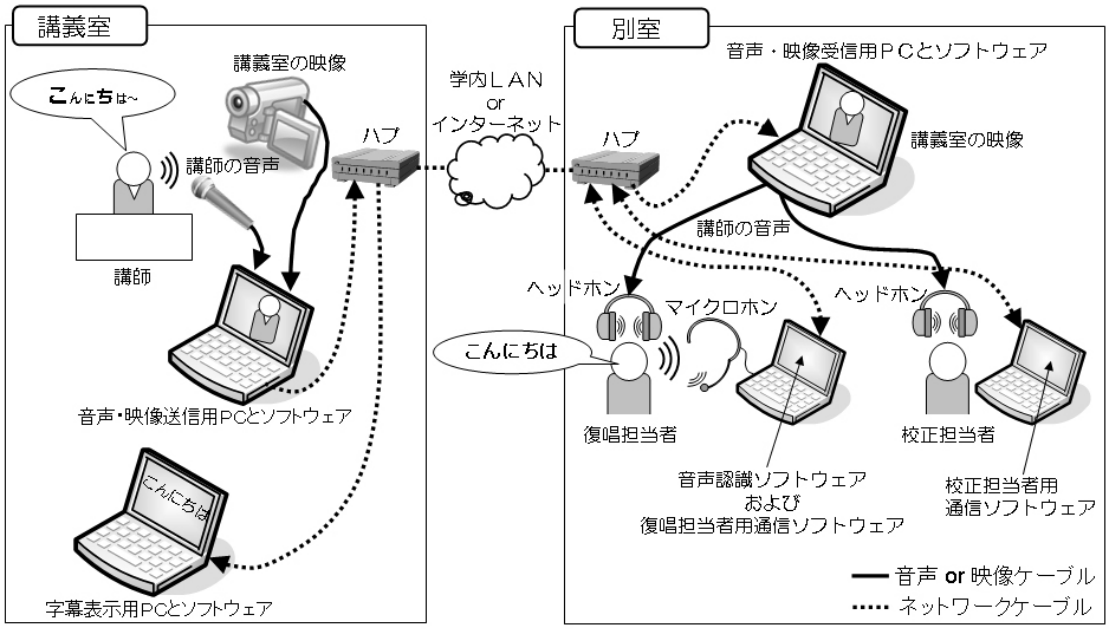


図２　別室から学内ＬＡＮやインターネットを介して情報保障を実施する場合の接続図

講義室および字幕作成作業を行う別室で、それぞれ通信用のＰＣ等を学内ＬＡＮやインターネットに接続します。この場合、各ネットワーク環境下で通信ができるように、各大学やプロバイダのルールに従って割り振られたＩＰアドレスを各ＰＣに設定する必要があります。同じルータ下に各ＰＣが配置できない場合には（字幕作成を実施する場所とネットワークの管理範囲の異なる他の学部などでの字幕提示など）、ソフトウェアの設定等を変更する必要があります。また、PacketiX 等に代表されるＶＰＮ（Virtual Private Network）機能を有するソフトウェアの利用が必要になる場合もあります。

講義室からの音声・映像の配信は、フリーウェアである Skype、MS-NetMeeting や Yahoo! メッセンジャーのようなビデオチャット用のソフトウェアの利用でも良いでしょう。また、専用のビデオ会議システムや Web カメラシステムの利用も可能です。図２の場合では、Skype のようなＰＣベースの通信システムの場合を示しています。

２－２．ＬＡＮケーブル１本で講義室と別室間を接続する場合

講義室および字幕作成作業を行う別室にそれぞれＨＵＢを設置して、それらのＨＵＢ同士を直接ＬＡＮケーブルで接続します（図３）。

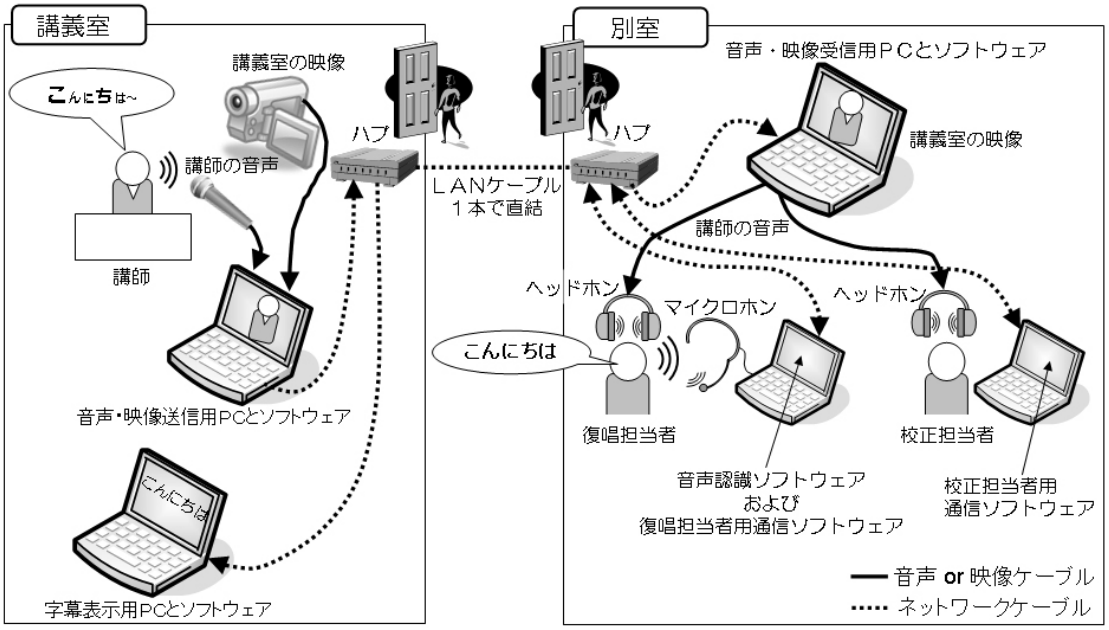


図３　ＬＡＮケーブル１本で講義室と別室間を接続する場合の接続図

この場合、１本のＬＡＮケーブルを一時的にでも設置する必要があります。ネットワーク構造はシンプルですが、同じフロア内に留まらない配線（階が異なる場所や棟の異なる場所へ配線）の場合には、学内ＬＡＮ等への接続が適しているでしょう。

ネットワークが外部と遮断されていますので、最低限、教室側の音声・映像を取得するためのシステムは、ローカル環境で動作するものを選定する必要があります。ビデオ会議システムや Web カメラシステムを使用すると、単独で講義室の音声や映像の配信が可能です。音声・映像配信用の機器や字幕作成に利用するＰＣには、プライベートＩＰアドレスを割り振って利用します。

２－３．別室から音声・映像ケーブルを使用して情報保障を実施する場合

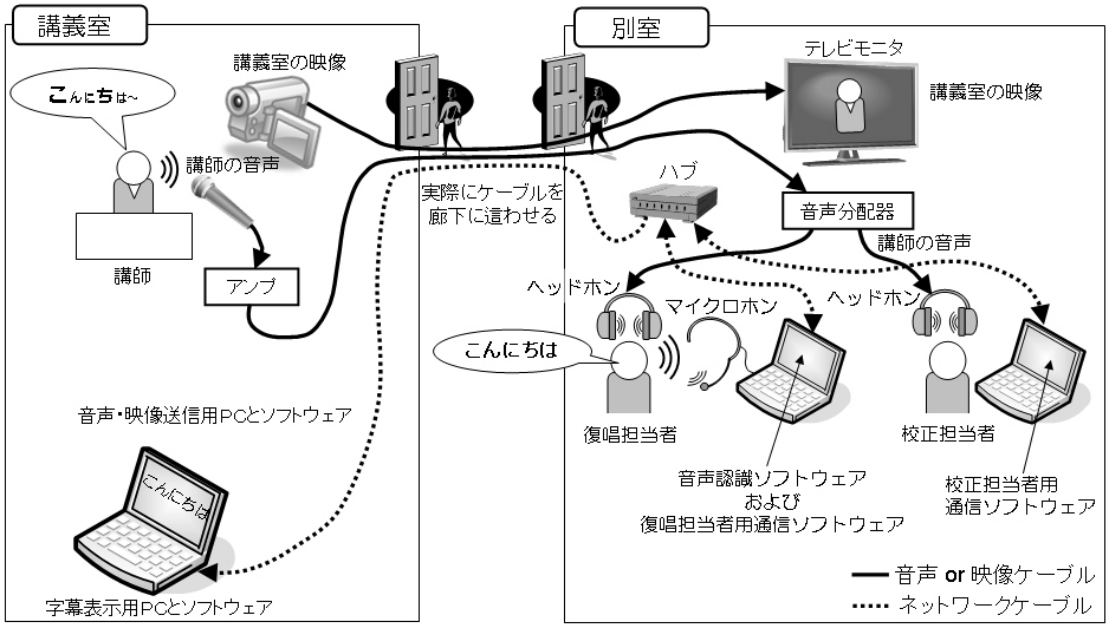


図４　別室から音声・映像ケーブルを使用して情報保障を実施する場合の接続図

別室から音声・映像ケーブルを利用して情報保障を行う際、講義室側にマイクロホンとビデオカメラを設置して別室の情報保障者に音声と映像を提示します（図４）。これらの情報をもとに、字幕作成作業を実施します。具体的には、マイクロホンが取得した講師や教室の音声信号を、アンプを介して増幅し、その信号を音声ケーブルで別室の情報保障者へ送ります。同様に、ビデオカメラで撮影した教室側の映像信号を映像ケーブルで別室に送ります。別室では音声信号を音声の分配器によって分配し、各情報保障者のヘッドホンへ送ります。映像信号はモニタ等に映し、情報保障者へ提示します。音声・映像ケーブル以外に、字幕の通信用にＬＡＮケーブルも設置する必要があります。

特に、講義室と作業を行う部屋が近い場合には、ケアレスミスを防ぐ上でも、映像・音声ケーブルで接続する方が、利点が多いでしょう。

２－４．講義室内で特殊なマイクロホンを使用して情報保障を実施する場合

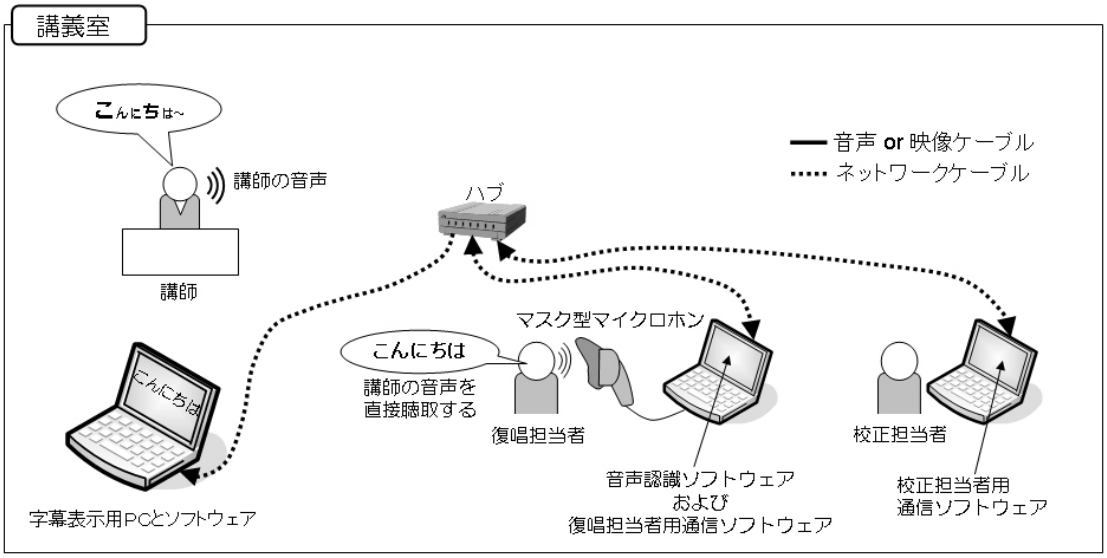


図５　講義室内で特殊なマイクロホンを使用して情報保障を実施する場合の接続図

復唱方式による音声認識同時字幕の場合、復唱作業は別室で行うのが一般的ですが、特殊な形状のマイクロホン（マスク型マイクロホン）を利用することで、講義室内で実施できる場合もあります（図５）。利用には注意を要するマイクロホンではありますが（P16 参照）講義室内での復唱作業の実施が可能となるために、多くの通信機材の準備が不要となるという大きなメリットがあります。

復唱担当者は講師の音声を直接聴取して、復唱を実施します。字幕の通信のためのネットワークもローカル接続で良く、非常にシンプルな構成となります。

3．復唱方式による音声認識同時字幕システムで使用するソフトウェア

では、実際に復唱方式を用いて字幕を作成するときの方法について説明していきましょう。字幕作成には、ネットワーク関連の機器をのぞいて、以下のような機器が必要です。

復唱用	PC	音声認識ソフトウェア	１万円～２万円
		SR-LAN2’	フリーソフト
	USB サウンドデバイス		７千円程度
	高指向性マイクロホン※1		７千円程度
校正用	PC	SR-LAN2’	フリーソフト
		音声遅延ソフト	フリーソフト
	高遮音性ヘッドホン※1		６万円弱
表示用	PC	SR-LAN2’ or IPtalk	フリーソフト

※1　遮音性の高いヘッドホンおよび指向性の高いマイクロホンを使用する代わりに、２つの機能を持ったヘッドセット（7 千円程度）を使う方法もあります。この際、復唱作業に慣れていない場合には、作業が実施しづらくなることもあります。

3ー1 音声認識ソフトウェア

現在、国内で利用可能な主なソフトウェアは、ViaVoice 10.5、Dragon Naturally Speaking 2005 Professional/Select そして AmiVoice ES 2008 が挙げられます。

これらのソフトウェアには、単語を登録する機能が用意されているものや、個人の音声特徴を登録するための機能を有するものもあります。このように個人や利用する分野に合わせることで、認識率を高めることが可能となります。ただし、人名等の固有名詞に関しては誤認識が多発しますので注意が必要です。

情報保障における音声認識ソフトウェアの選定時には、認識速度や認識精度が十分か、また単語登録機能が備わっているかなどのポイントを重視する必要があります。加えて、復唱担当者を複数体制で実施する場合、復唱担当者の数だけ音声認識ソフトウェアを用意する必要があるので、コスト的な問題も無視できないことでしょう。

	ViaVoice Academic V10.5	Dragon Naturaly Speaking 2005 Professional 日本語版	AmiVoice ES 2008
製造会社	IBM	ニュアンスコミュニケーション㈱	㈱アドバンスト・メディア
価格	15000 円程度	90000 円程度	20000 円程度

本書ではこうした点を鑑み、AmiVoice ES 2008 を利用した字幕作成について紹介します。AmiVoice ES 2008 の場合、音声特徴の登録のためのコンテンツは用意されておりませんが、自由発話時や文章の読み上げ時に音声特徴を蓄えておき、その特徴を登録す

ることができます。そのような登録作業によって各個人の特徴に適合したユーザーデータを作成することができ、認識精度を高めることができるようになります。

参考：

・エムシーツー社（開発元：アドバンスト・メディア）
AmiVoice ES 2008(マイク別売)　型番：AESSD-0802
http://www.mc2-ltd.jp/amivoicees2008.html

３－２．連携作業用通信ソフトウェア

音声認識ソフトウェア以外にも、誤字脱字を修正したり、字幕を表示される役割を持ったソフトウェアが必要です。ここでは「SR-LAN2 ダッシュ」（以下、SR-LAN）（開発：筑波技術大学　三好茂樹氏）を利用します。

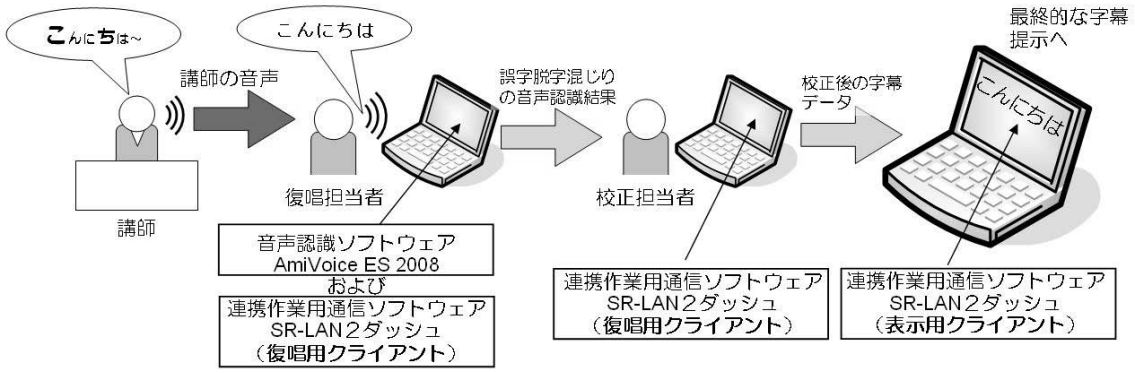


図６　復唱方式による音声認識同時字幕システムでの作業概略

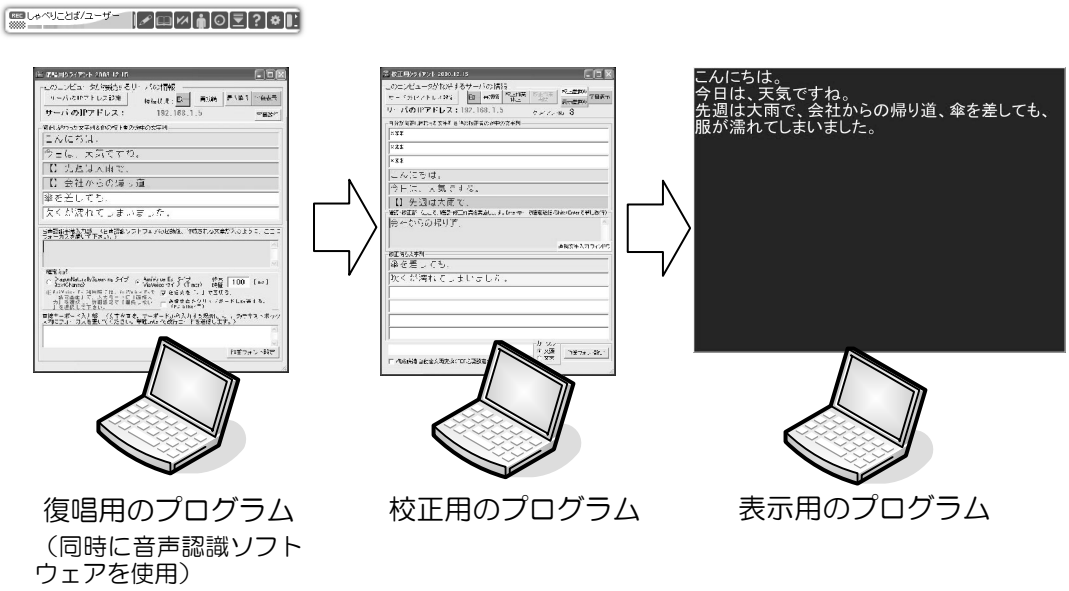


図７　各担当と SR-LAN2 ダッシュ

連携作業用通信ソフトウェア SR-LAN は、復唱方式による音声認識同時字幕システムを実現するためのプログラム群です（図６）。音声認識によるリアルタイム字幕提示実験実施のために作成されたシステムで、復唱者用、校正者用、表示用および管理サーバの４つのソフトウェアから構成されています（図 7）。

まず、「復唱用のプログラム」である「復唱用クライアント」では、音声認識ソフトウェアから得られる誤字脱字混じりの字幕データを、「校正用のプログラム」にネットワークを介して自動的に送ることができます。

次に「校正用のプログラム」である「校正用クライアント」では、復唱用クライアントから送られてきた文字を校正者に割り振り、修正後の字幕データを表示用のプログラムに送信します。修正は複数名で行えるため、誤字脱字の量に応じた人数で対応することができます。

さらに「表示用のプログラム」である「表示用クライアント」では、校正用クライアントから送られてきた字幕データを聴覚障害者に対して表示します。ここではフォントの種類や大きさを選ぶことができ、聴覚障害学生のニーズにあわせて表示方法を変えることができます。

「管理サーバ」は、復唱用、校正用および表示用のプログラム間の字幕データの相互のやり取りを管理する役割があります。この管理サーバを起動しておかないと、他のプログラム同士の通信が出来なくなり、システムとして機能しません。復唱用クライアントや校正用クライアント等、他のプログラムと一緒に起動することができるので、新たに PC を用意する必要はありません。起動後、画面を最小化しておき、必要に応じて操作すれば良いでしょう。

SR-LAN は、PEPNet-Japan より無償で配布されています。詳しくは PEPNet-Japan 事務局までお問い合わせください。

PEPNet-Japan ホームページ　http://www.pepnet-j.org

３－３．その他機材等

復唱方式による音声同時字幕を作成するためには、ここで示したソフトウェアの他に、復唱者用のマイクロホンや音声をパソコンに入力するためのサウンドデバイスなどが必要です。これらの機材の選定方法については、２９ページ以降に記載していますので、参考にしてください。

４．音声認識ソフトウェアと連携作業用システム構成の手順と字幕作成までの流れ

次に、前項で紹介したソフトウェア等を利用し、字幕作成を行うための設定について説明します。音声認識ソフトウェアを利用した字幕作成では、IPtalk を使った通信による連携作業時と同様のネットワーク機器を用います。以下で触れる校正担当者用 PC も含め各 PC を、同一の HUB にストレート・ケーブルで接続して利用します。

４－１．音声認識ソフトウェアについて

音声認識ソフトウェアのインストール後、情報保障用途で利用する場合の設定を行います。まず、各種の付加的な機能を停止させる必要があります。特にボイスコマンドは、利用中に支障をきたしますので停止させておきましょう。また、本マニュアルで説明しているシステムの運用時には、誤字脱字が発生しても、その時に修正による学習はさせられません。そのため、もしも繰り返し同じ単語の誤りが発生した場合、その単語出現率が上昇し、誤った単語を学習してしまう可能性があります。よって、そのような学習効果をユーザーデータに上書き保存しないよう、自動保存機能を OFF にしておくことをお勧めします。

この他、事前の講義資料がある場合には、単語登録を実施しておきましょう。ソフトウェアによっては、テキストデータや Word 文章をそのまま読み込ませ、そこから単語を抽出し、登録してくれる機能もあるので、有効に活用しましょう。

４－１－１．音声認識ソフトウェア「AmiVoice ES 2008」の設定

本マニュアルでは、AmiVoice ES 2008 を例にとって説明します。

AmiVoice ES 2008 のインストール後、ソフトウェアを起動すると、以下のようなツールバーが表示されます（図１０）。使用するユーザを登録した後、数分間適当な文章を読み上げて個人の音声特徴を登録するだけでもある程度の認識精度が期待できます。この事前の登録作業が少ない点もこのソフトウェアの特徴の一つと言えるでしょう。もしも認識率が低いと感じた場合には、更に登録作業を続けてみることをお勧めします。

図１０　AmiVoice ツールバー概観

次に説明する「連携作業用通信ソフトウェア」と共に AmiVoice を利用するために、AmiVoice 側を以下の設定にしておく必要があります。

- 個人の音声特徴登録後、「設定画面」で、「音響学習」の「これ以上蓄積しない」をチェック。
- 「オプション」の「転送方法」で、「標準の転送方法」を選択。
- 「入力モード」で、「直接入力」を選択。そして「詳細設定」で、「編集しない」を選択。
- 「表現」の箇所は、表示させたいスタイルに合わせて選択。

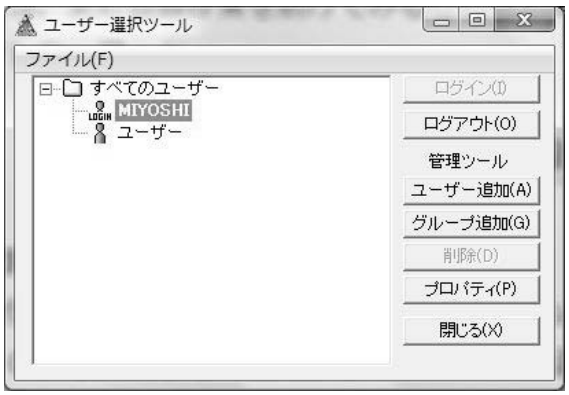


図１１　ユーザーデータの追加

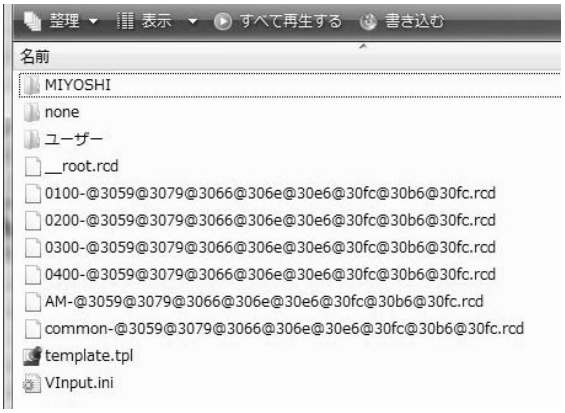


図１２　ユーザーデータのコピー

４－１－２．音声認識ソフトウェア「AmiVoice ES 2008」のユーザーデータについて

複数台の復唱用 PC にインストールした AmiVoice 間で、復唱を担当する方々のユーザーデータを統一するための手順を以下に示します。

１．１台目の復唱用 PC（復唱用 PC１）で、ユーザーを追加する。例えば、ユーザー名：MIYOSHI（図１１）

２．個人の音声登録を実施し、保存する。

３．「c:¥Program Files¥AmiVoiceEs ¥users¥すべてのユーザー」フォルダ内にある「MIYOSHI」フォルダ（図１２）をコピーし、USB メモリ等に保存する。

４．復唱用 PC２の AmiVoice で、同一の名称（MIYOSHI）のユーザーを追加します。

５．復唱用 PC2 の「…¥すべてのユーザー」フォルダ内にある「MIYOSHI」フォルダに、USB メモリにコピーした「MIYOSHI」フォルダを上書きコピーします。

６．復唱用 PC２でユーザーを「MIYOSHI」に切り替えれば、復唱用 PC１で作成した音響情報が反映されています。

このような手順で、ユーザーデータを各復唱用 PC 間で同一のものに整えることができます。

４－２．連携作業用通信ソフトウェア SR-LAN の運用手順

連携作業用の通信ソフトウェア SR-LAN は、４ つのプログラム群で成り立っています。図１４に示すインストーラを起動すると、これらのプログラム（図１４）が一度に PC にインストールされます。また同時にデスクトップや「スタート」に、ショートカットが作成されます。

図１３　インストーラ概観

※ ソフトウェアは予告無しに
更新されることもあります。

図１４　各プログラムのアイコンの概観

４つのプログラムは以下の役割があり、利用する情報保障者が異なります。

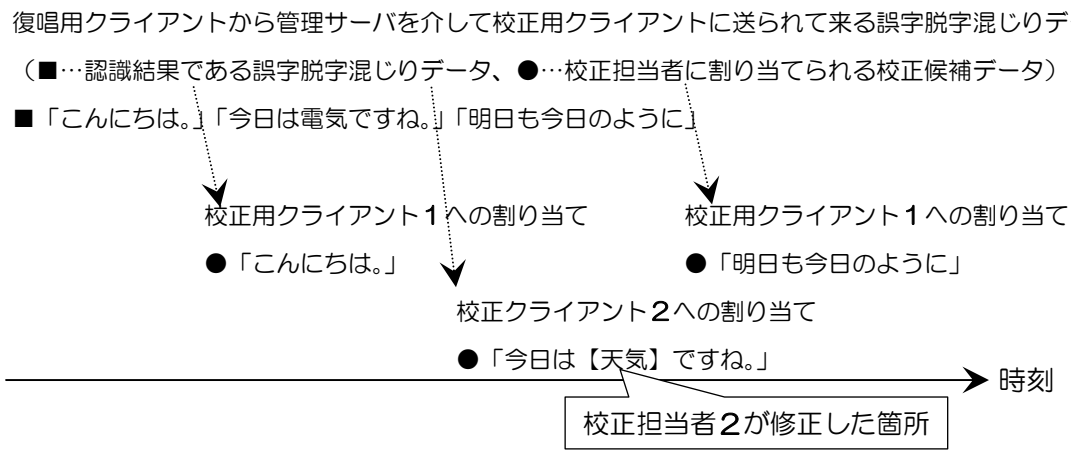
①復唱用クライアント.exe（以下、復唱用クライアント）

復唱者用の PC で起動し、復唱担当者が利用するソフトウェアです。このソフトウェアは音声認識ソフトウェアと同時に利用します。復唱者が複数人で、交代で実施する場合には、各 PC で起動しておく必要があります。

②校正用クライアント.exe（以下、校正用クライアント）

校正者用の PC で起動し、校正担当者が利用するソフトウェアです。校正者が複数人で、同時に実施する場合には、各 PC で起動しておく必要があります。復唱用クライアントから管理サーバを介して送られてくる誤字脱字混じりの字幕データを、校正する役割を担います。校正クライアントが複数個、管理サーバに接続している場合には、復唱用クライアントから送られてくる誤字脱字混じりの複数個の連続した字幕データは、管理サーバによって各校正用 PC に対して、次の例のように交互に割り振られます。

例.



③表示用クライアント exe（以下、表示用クライアント）

これは、表示用 PC で起動しておくソフトウェアです。

④管理サーバ.exe（以下、管理サーバ）

クライアント PC のどれか１台で起動して、字幕データの通信処理等を管理するソフトウェアです。このソフトウェアが稼動していないと、システムとして機能しません。

ソフトウェアである管理サーバは、各クライアントを使用する時だけ起動しておいて下さい。起動するＰＣは基本的にどれでもかまいません。ここでは、校正用ＰＣ上で校正クライアントと同時に起動しておいてください（バックグラウンドサービスではありませんので、起動している間だけ機能しています）。このソフトウェアは、復唱用クライアントからの文字列を校正用クライアントに割り振る役割を主に担っています。また校正者が複数人で同時に作業を行う場合に生じる校正済み文字データの表示の順序も管理しています。例えば、ある文章の校正作業中に、校正作業が早く終わってしまった後続の文字データを即座に表示するというような文章の前後が入れ替わることなく、時間的な順序を整えて提示する機能があります。整えられた字幕データが最終的に表示字幕として表示用ＰＣに送られます。

４－３．各クライアント PC とネットワーク接続について

このマニュアルでは、「２－２．ＬＡＮケーブル１本で講義室と別室間を接続する場合」で紹介した接続方法を例にとり、復唱者２名、校正者２名で実施する場合を想定して説明します。

必要な機材：

- ノートパソコン×**５台**
  - 内訳：
    - **音声認識ソフトウェアと復唱用クライアントを起動する復唱用ＰＣ２台**
      （今後、各 PC を「復唱用 PC#1」、「復唱用 PC#2」と呼びます）
    - **校正用クライアントを起動する校正用ＰＣ２台**
      （今後、各 PC を「校正用 PC#1」、「校正用 PC#2」と呼びます）
      （内、校正用 PC#1 で管理サーバを起動、
      校正用 PC#２目で遅延再生プログラムを起動）
    - **表示用クライアントを入れる表示用ＰＣを１台**
      （今後、この PC を「表示用 PC」と呼びます）
- ＨＵＢ１個
- ＬＡＮケーブルを各ＰＣ用にそれぞれ１本（合計５本）
- 表示用プロジェクタおよび VGA ケーブル１本
  （字幕をプロジェクタに表示する場合）

各ＰＣをＬＡＮケーブルでＨＵＢに接続します。

各ＰＣにはＩＰアドレスが割り振られていることが前提です。学内ＬＡＮにも接続しておらず、完全にローカル環境である場合には独自にＩＰアドレスを割り振って下さい。重要なポイントは管理サーバのＩＰにあります。このＩＰアドレスを各クライアントに入力して、クライアントがどこに接続するのかを指定する必要があります。すべてのクライアントが管理サーバに接続されると字幕の作成作業が出来るようになります。

４−４．各クライアントの起動と接続設定

まず、校正用クライアント PC で、管理サーバを１つだけ起動しましょう。

右のようなプログラムのウィンドウが開きます。

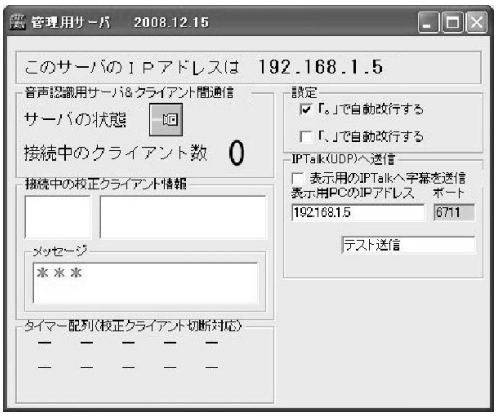


図１５　管理サーバの概観

この管理サーバのみ起動しており、まだクライアントを接続していませんので、「接続中のクライアント数　０」という表記が出ています。

次に、１台目の復唱用 PC で復唱用クライアントを起動しましょう。以下のような画面が表示されます（図１６）。左上の「サーバの IP アドレス設定」ボタンを押すと図１７のようなサーバの IP アドレスの入力を促すウィンドウが表示されます。

ここで、図１５に示した管理サーバのウィンドウに表示されているサーバの IP アドレスを入力します。OK ボタンを押すと、サーバに接続されます（図１８）。

図１６　復唱用クライアント概観

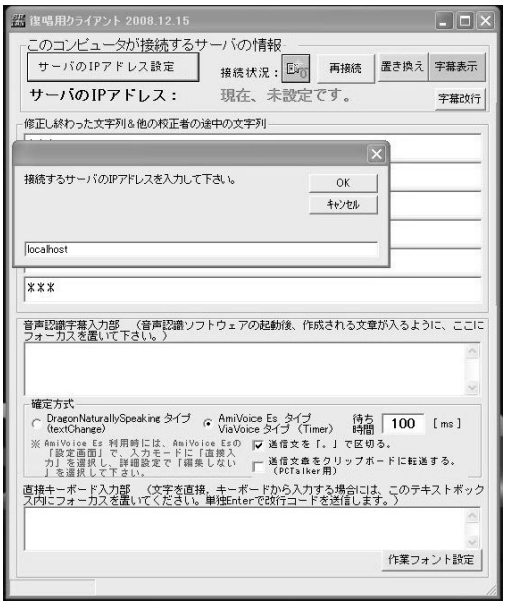


図１７　復唱用クライアント（サーバ IP 設定）

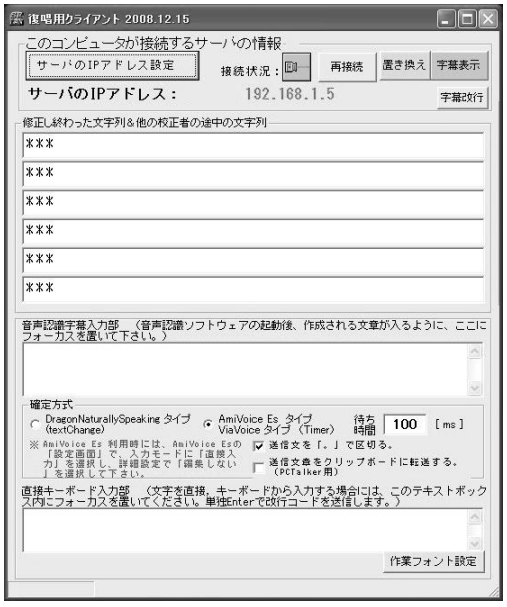


図１８　復唱用クライアント
（サーバ接続後）

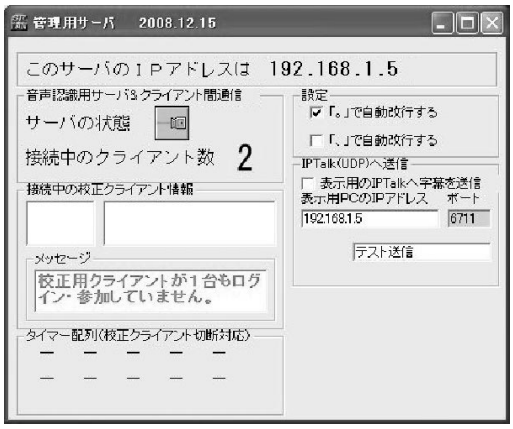


図１９　管理サーバ
（2 つの復唱用クライアント接続後）

同様に２台目の復唱用 PC で復唱用クライアントを起動し、サーバに接続します。接続された管理サーバでは、図１９のように、「接続中のクライアント数」が0から１、そして２へ増加します。

次に、１台目の校正用クライアントを接続しましょう。校正用 PC で校正用クライアントを起動すると図２０のようなウィンドウが表示されます。復唱用クライアントと同様に、管理サーバの IP アドレスを入力すると（図２０）、管理サーバ側の表示（図１９）は、クライアント数が２から３へ変化し、接続中の校正クライアントに、接続した１台目の構成クライアントの IP アドレス情報が表示されます。

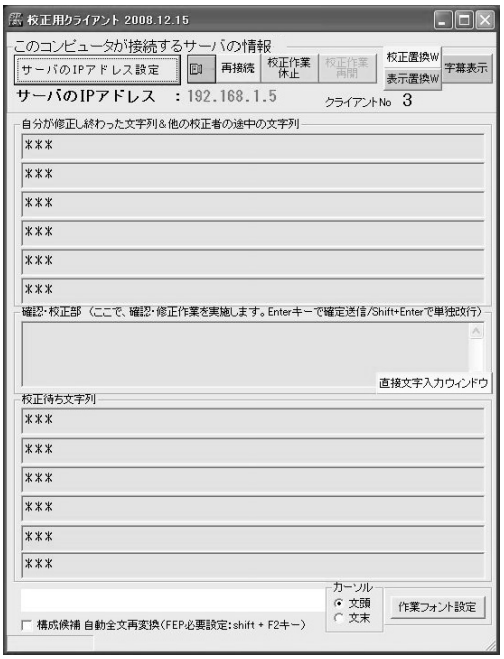


図２０　校正用クライアント概観

同様に２台目の校正用クライアントを接続して下さい。管理サーバ側の表示は３から４に変化します。

続いて、表示用クライアントを表示用 PC で起動して、管理サーバの IP アドレスを入力して下さい。すると図２２のような表示のウィンドウになります。

※ 校正用クライアントの「校正候補 自動全文再変換」機能は、愛媛大学立入哉氏の情報提供で実現しました。
機能の利用の際には、ATOK または MS-IME の「再変換」キーの割り当てを「shift キー+F2 キー」に設定する必要があります。

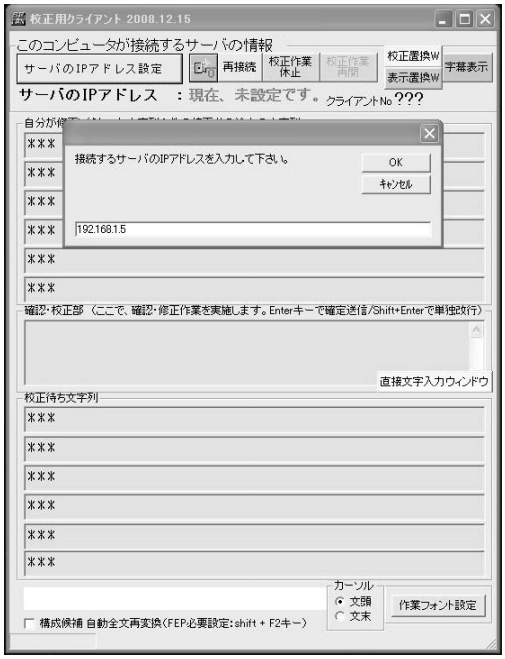


図２１　校正用クライアント
（サーバ IP 設定）

図２２　表示用クライアント
（管理サーバへの接続後）

図２３　表示用クライアントの
字幕ウィンドウ

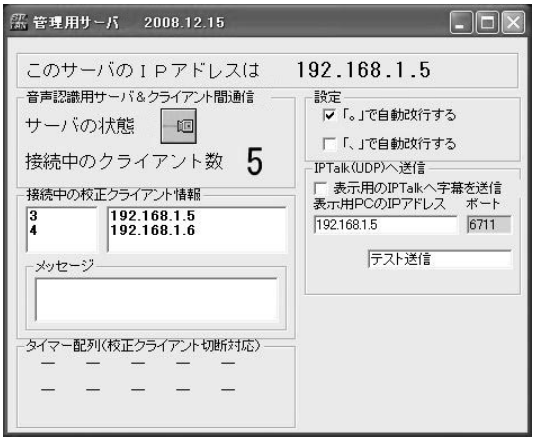


図２４　管理サーバ（2 つの復唱用クライアント、２つの校正用クライアントおよび１つの表示用クライアントが接続）

表示用クライアントの「字幕ウィンドウ表示」ボタンを押すと、図２３に示すようなウィンドウが開きます。このウィンドウ上に、聴覚障害学生へ提示するための最終的な字幕が表示されます。このウィンドウは、最大化した後に「Ｆ１キー」を押すことでツールバーを消去し、PC 画面全体に字幕を表示することが可能です。また、文字の色、フォント、背景色等の設定も可能です。

ここまでで、全種類のクライアントが管理サーバに接続したことになります。全クライアント接続後のサーバの状態を図２４に示します。復唱用クライアント×２、校正用クライアント×２および表示用クライアント×１の合計５クライアントの接続が終了しました。

４－５．各クライアントの操作と連携作業の流れ

始めに、復唱用クライアントを起動している復唱用 PC で、音声認識ソフトウェアを起動しましょう（図２５）。

ウィンドウ内の緑色の背景のテキストボックス内にフォーカスを置き、音声認識ソフトウェアを用いて音声認識結果を入力します。

図の例では、復唱者は「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。」と発話したところ、音声認識結果は「こんにちは。」「今日は、電気ですね。」･･･「吹くが濡れてしまいました。」となってしまっています（下線部が誤字を示しています）。

音声認識ソフトウェア AmiVoice ES 2008 のツールバー

図２５　復唱用クライアントと音声認識ソフトウェア

図中の音声入力部の上部に管理サーバから送られてくる認識結果のヒストリーが、白い背景のテキストボックスに表示されています。

図２６と図２７に示したのが、字幕データ受信後の２つの校正クライアントの様子です。復唱用クライアントから管理サーバを介して送られて来た字幕データが校正候補として各ウィンドウ下部に順に蓄積しています（図２６）。管理サーバは各校正用クライアントに交互に校正候補である字幕データを割り振ります。自分が校正する予定の字幕データは薄い緑色で表示され、ウィンドウ中央部の「確認・修正部」で作業を行います。青色の箇所は、他方の校正用クライアント（校正用クライアント＃２）の校正中の字幕データを表しています。校正箇所（カーソルの位置）は【】で示され、漢字変換中の様子もモニタすることができます。また、薄い青色の箇所は、他方の校正用クライアントが担当する校正候補を表しています。

一方、校正用クライアント＃２（図２７）では、誤った字幕データ「電気」を「天気」に修正する作業を実施し始めており、その様子は校正用クライアント＃１（図２６）や復唱用クライアント（図２５）にも表示されています。このようにして、校正の進捗状況を互いにモニタすることができます。これは、校正作業で困っている場合に、その箇所に関して他の担当者は口頭でアドバイスすることや、その日の情報保障として十分な人員構成かを各自が判断するための補助にもなります。

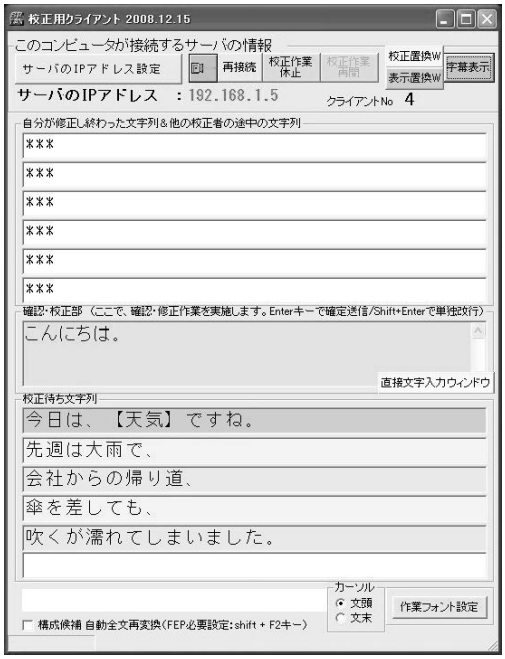


図２６　校正用クライアント＃１

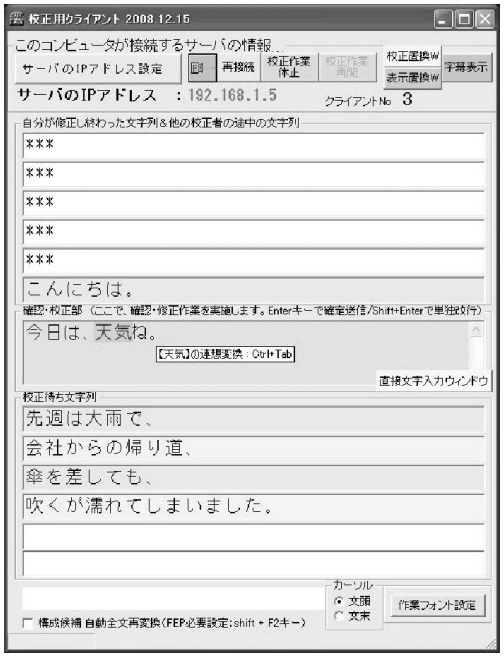


図２７　校正用クライアント＃２

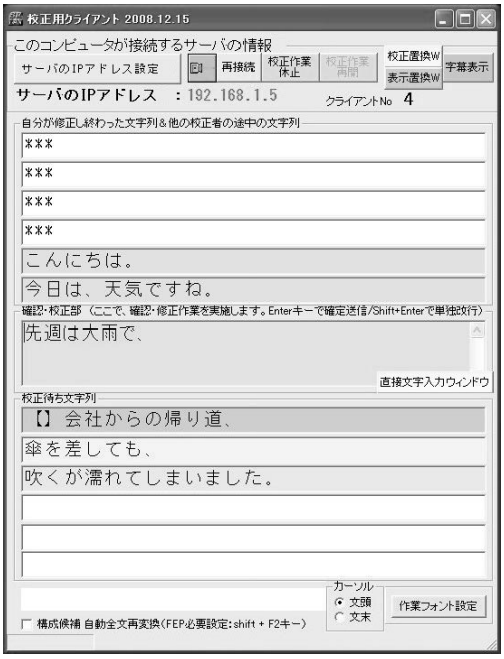


図２８　校正用クライアント＃１
（作業中）

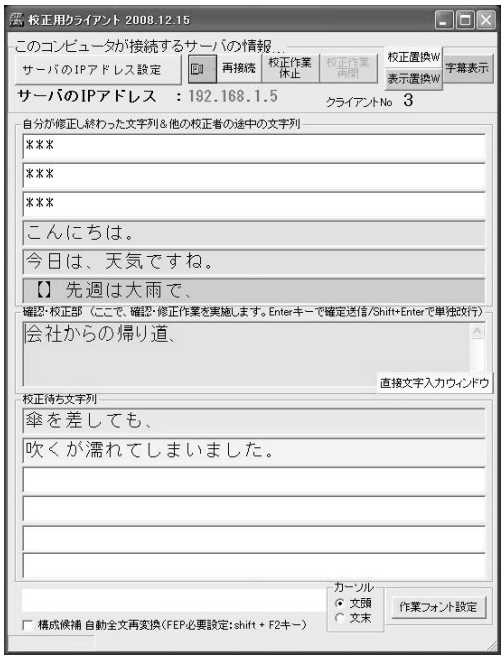


図２９　校正用クライアント＃２
（作業中）

図２７で「電気」を「天気」に修正した後、確定送信をすると、他のクライアントでも該当する字幕データのテキストボックスが灰色の背景に変化します（図２８および図 29）。この灰色の背景は「校正作業が終了し、字幕として提示されている」という意味を表します。確定送信後、担当する字幕データが自動的に「確認・修正部」に入り、修正候補も上に移動します。図２８では、「こんにちは。」を確定送信し（Enter キーを１度押して確定送信する）、次に「会社の帰り道、」の確認を始めています。この様子も、他のクライアントでもモニタできます。図２９では、「会社からの帰り道、」の確認を始めています。この様子も、同様に他のクライアントでもモニタできます。

ここまでの校正作業の様子は、復唱用クライアント＃１では図３０のようになります。

音声認識ソフトウェアによって字幕化したデータがウィンドウの上部に蓄積され、２名の校正担当者が各校正用クライアントで校正している様子がわかります。灰色の箇所は校正作業終了および字幕提示済みのデータ、そして青で示された箇所が現在校正作業を実施しているデータ、そして白い箇所はまだ手付かずの校正候補データです。

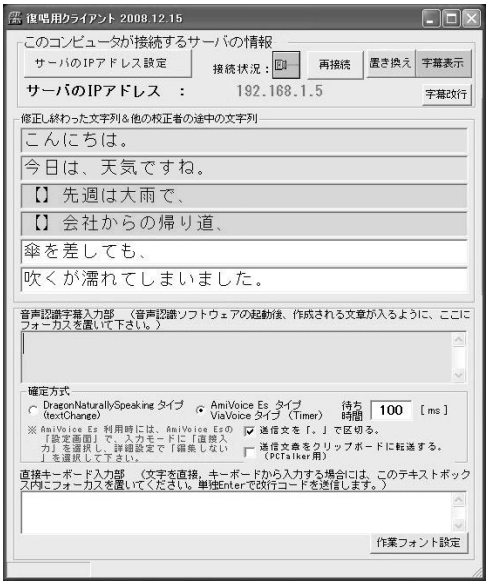


図３０　復唱用クライアント＃１
（作業中）

図３１　復唱用クライアント＃１
（全データ確定後）

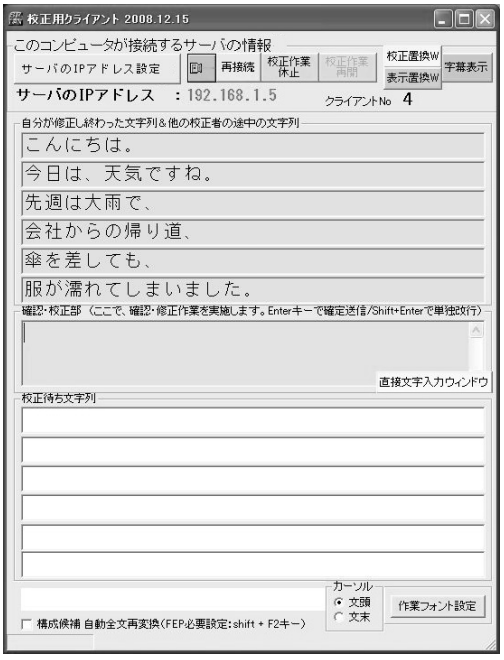


図３２　校正用クライアント＃１
（全データ確定後）

大まかに言うと、復唱担当者にとって復唱後、速やかに校正済みの灰色の“範囲”が降りてくることが、システム全体として望ましいということになります（図３１）。

もしも、なかなか校正済み範囲が降りてこない場合には、自分の発話方法と認識率を確認・改善したり、同時に作業をする校正担当者の人数増加等を考えるべきでしょう。

一方、校正担当者にとっては、「確認・校正部」下部の校正候補の字幕データが速やかに無くなるように作業を実施することが望まれます（図３２）。また、「【】」による他方の校正担当者のカーソル位置や校正作業の様子は、他方の校正担当者との力量の違いを認識することにも役立ちます。他方が遅れているようであれば、一時記憶した講師音声の内容が不明確になり、校正作業が実施しづらくなります。このような場合には、口頭でサポートする方が得策です。

復唱用および校正用クライアントには、それぞれ最終的な字幕表示画面（図３３）が用意されており、聴覚障害者に提示する字幕の表示タイミング等も確認できます（下記の表示用クライアントでの字幕と同様の内容が表示されます）。また、提示済みの字幕の変更や、直接、文字データを送信する機能など様々な補助的機能が含まれています。

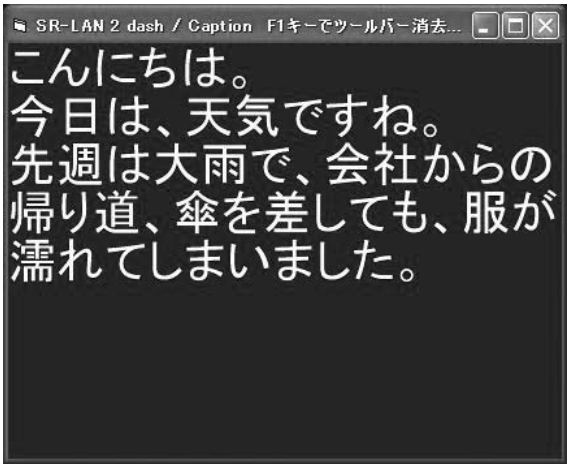


図３３　復唱用および校正用クライアントの最終字幕のモニタ画面

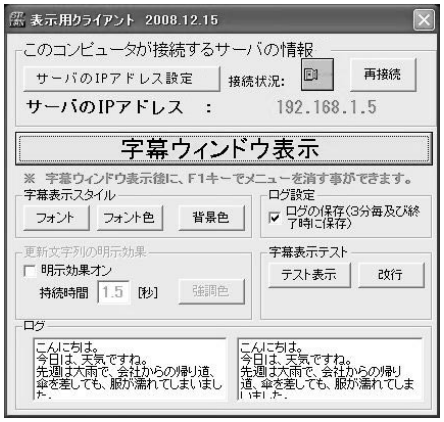


図３４　表示用クライアント

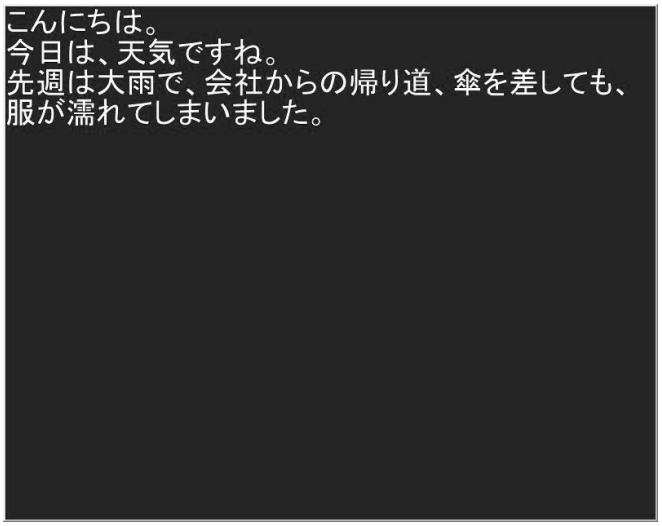


図３５　表示用クライアントの字幕表示画面（全画面表示）

聴覚障害学生に提示する字幕は、表示用クライアント（図３４）を用いて提示します。「字幕ウィンドウ表示」ボタンを押して字幕提示用ウィンドウを表示した後、ウィンドウを最

大化してF1 キーを押すと、全画面表示を行うことが出来ます。このウィンドウを図３５に示します。

表示用クライアントではログの自動保存も可能ですが、他のクライアントの字幕ウィンドウをコピー＆ペーストして保存しても、ログの保存は可能です。

また、表示用に IPtalk を利用することも可能です。管理サーバで「IPtalk(UDP)へ送信」の箇所で、IPtalk 用 PC の IP アドレスを正確に入力した後、「表示用の IPtalk へ字幕を送信」にチェックを入れると、該当する IPtalk へ字幕データを送信・表示することが可能です。IPtalk への送信では、校正クライアントによる「表示済みの字幕中に含まれる誤字の置換機能」も機能します。

ところで、校正作業を複数人で実施する場合、情報保障の途中から校正作業に新規に加わる校正クライアントの場合、単純にサーバに接続するのみで問題はありません。しかし、校正のメンバーから一時的に離脱する場合、「校正作業休止」ボタンを押した後、自分の担当分の字幕データの校正作業を終えてから、連携作業から離脱して下さい。すでに割り当てられている字幕データの校正作業を終えずに、作業を停止した場合、またはネットワークトラブルによって通信が切断された場合、校正作業予定の字幕は、数秒後に“そのまま”表示されることになります（管理サーバの「タイマー配列」の箇所に表示）。再度、作業に復帰する場合には、「校正作業再開」ボタンを押すと、押した時点から字幕データの割り振りを受けられるようになります。

なお、校正用クライアントが１台も接続していない場合には、復唱クライアントからの字幕データは一切字幕表示画面には表示されません。また、校正用クライアントが接続した時点からの字幕データが校正・表示され、接続前までの字幕データは無視されます。

５．復唱・校正作業負荷を軽減させる機材

前項では、音声認識ソフトウェアと SR-LAN を用いた字幕作成の手順について述べました。音声認識による字幕の精度を上げていくためには、復唱者や校正者が十分なトレーニングを行うことが重要ですが、これ以外にも機材やその使用方法を変更することで、復唱・校正者の負担を軽減させることも可能です。ここでは、復唱方式によるリアルタイム字幕作成作業をより効果的に実施するために用いることができる機材について紹介します。

５－１．復唱に関する作業負荷を軽減させる機材について

復唱作業を軽減させるための機材には、高性能のマイクロホンや音のノイズを軽減するサウンドデバイスなどがあります。ここで紹介している機材のうち、教室から離れた別室で作業を行う場合には 5-1-1、教室内で復唱作業を行う場合には、5-1-2 を用いることができます。

5-1-1．指向性の高いマイクロホン利用

音声認識による字幕の精度には、復唱者の音声のみでなく、周囲のノイズも大きく影響します。復唱を行う環境の外部ノイズ（ドアの開閉音やエアコン、人の話し声など）から入力音声の音質を守るためには、周辺ノイズをカットする機能を有する指向性の高いダイナミック・マイクロホンを利用することが望ましいでしょう。この場合、復唱音声が周辺に聞こえてしまうため、授業を行っている教室ではない“別室”で、聴取・復唱する必要があります。

参考：

・オーディオテクニカ社製　ハンズフリーマイクロホン　型番：AT810F
http://www.audio-technica.co.jp/products/mic/at810f.html

写真５　指向性の高いマイクロホン

５－１－２．口や鼻を覆うタイプのマイクロホン（マスク型マイクロホン）の利用

復唱方式による音声認識同時字幕の場合、復唱作業は別室で行うのが一般的ですが、鼻部と口部を覆い、音声が外部に漏れない工夫を施されたマスク型マイクロホンを利用することで、教室内で作業を行うことが可能になります。ただし、マスク型マイクロホンの利用によって、復唱担当者の音声が比較的マスク外へ漏れなくなりますが、完全に防げるという訳ではありません。比較的広く、漏れた音声が気にならないような教室環境での利用が望まれます。

また、マスク型マイクロホンは鼻口部を覆うために、発話時に口周辺の筋肉の動きに制限を与えます。これが明瞭な発話を妨げることが多く、利用には慣れが必要です。一方、マスク型マイクロホンは音響的にもマスク内部での残響（ノイズの一種）が発生しやすく、これが認識精度を落とす一因となっています。このように通常のマイクロホンを利用する場合と比較して、多くのノウハウの習得が必要となります。使用する各個人によって、音声認識精度に大きく差が出るために、万人向けとは言えません。また、息継ぎのタイミングにもコツがあり、使いこなせないうちは、疲労するのが早くなると思います。

しかしながら、別室から情報保障を行う場合のように、多くの通信機材を準備する必要が無く、メリットも大きいと言えるでしょう。

注意としては、教員の発話を要約して発話した場合、他の校正担当者にその音声が伝わらないために、認識結果の校正が実施しづらくなることが挙げられます。これに関しては、マスク型マイクロホンの出力を分岐して、校正担当者に聞かせる等の工夫が挙げられます。マスク型マイクロホンには出力端子を２つ持つタイプのものもあり、そのタイプの採用によって、簡単に実現することができるでしょう。

最近、出力信号レベルを調整する機能を付加したマイクロホンが販売されており、従来のものと比較して、ある程度容易に使用することができるようになりました。(2009年1月現在)

参考：

・TALK TECHNOLOGIES INC.社製
Sylencer® SmartMic™ Silent Speech Recognition Microphone 型番： SM300
http://www.talktech.com/pages/products.html

写真６　鼻口部を覆うマスク型のマイクロホン

５－１－３．各種マイクロホンからの音声信号を PC へ入力するための機器

各種マイクロホンからの復唱音声の信号を、復唱用パソコンに入力する際には、適正な音量と音質を維持する必要があります。ノート PC の場合、内蔵のサウンドデバイスを利用すると、入力された音声信号が電磁ノイズの影響を受け、字幕の精度が著しく低下することがあります。これを回避するためには、USB 接続タイプのサウンドデバイスを利用することをお奨めします。

また、先述の「指向性の高いダイナミック・マイクロホン」の場合、出力音声信号のレベルが低い場合があるため、信号を増幅して PC へ伝えられるよう、ボリュームコントロールのついたサウンドデバイスを選択するとよいでしょう。

参考：

・クリエイティブ社製　型番：Sound Blaster Digital Music PX

http://jp.creative.com/products/product.asp?category=1&subcategory=207&product=10319&listby=

写真７　ノイズに強い音声入力デバイス

５－１－４．遮音性の高いヘッドホンの利用

音声を聴取しながら発話（復唱）する場合、自分自身の音声が、聴取したい教員の音声と重なってしまい、聴取しづらくなります。通訳や英語学習のトレーニング手法の一つに“シャドーイング”というものがあります。聞き取った音声を囁（ささや）くように復唱する方法で

すが、音声認識ソフトウェアを使用する際には、喉頭の振動が伴う明確な有声音である必要があります。明瞭な音声であればさらに復唱しづらくなってしまいますが、これをある程度防ぐ方法があります。それは、遮音効果の高いヘッドホンで教員の音声を聴取する方法です。これにより、自分自身の音量を低減し、発話しながら“教員の音声を聴取する”ことに集中しやすくなります。

自分自身の音声は空気中だけではなく、体の中も伝って耳に届くので、完全に聞こえなくなることはありません。しかし、音量が小さくなるので、始めのうちは自分が本当に明瞭に発話できているか、また、その大きさは適正か、混乱することがあるかもしれません。実際の復唱作業を担当する前に、ある程度音声認識ソフトウェアに適した発話方法のトレーニングを積んで、“明瞭な発話と大きさ”を維持できるようにしておくことが大切です。音声認識ソフトウェアには発話中の音声の大きさをレベルメータ表示する機能が付いていますので、それを目安として適度な大きさを意識しながら発話すると良いでしょう。

また、復唱作業に慣れてくると、より正確な字幕作成を実施するために、校正担当者との積極的なコミュニケーションを取りたいと思うようになるかも知れません。完全な分業からより積極的な連携作業に移行する際、遮音性の高いヘッドホンを装着していると、なかなか円滑なコミュニケーションが取れなくなってしまいます。“教員の音声を聴取する”ことから注意がそれなくなった頃（復唱作業に余裕が出てきた場合）には、遮音性の低い通常のヘッドホンに変更することをお勧めします。そのようにして、字幕作成中に発生した誤認識結果を校正担当者に伝えたり、聞き取れなかった箇所を補い合う連携体制も字幕精度に大きく影響を与えることでしょう。

なお、手話通訳経験等のある方の中には、自分の音声が復唱の妨げにならない・なりづらく、むしろ自分の音声も明確に聞きながら復唱作業をしたいという方もいます。よって、復唱が上手く出来ない方の初期段階のサポート手段として、遮音性の高いヘッドホンの利用は考えるべきでしょう。

参考：

・ゼンハイザー社製 ヘッドホン　型番：HDA200

http://www.sennheiserusa.com/newsite/productdetail.asp?transid=002994

http://www.sennheiser.com/sennheiser/home_en.nsf/root/professional_audiology-hda-200?Open&path=professional_audiology

※　業者への発注の際には、コネクタ形状を「ステレオミニジャックに加工」する事を依頼してください。

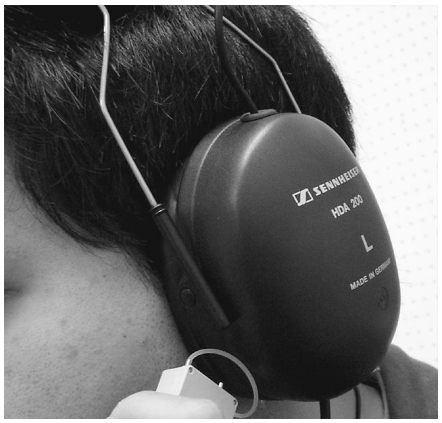


写真８　遮音性の高いヘッドホン

また、遮音性の高いヘッドホンの入手が困難な場合は、通常のヘッドホンでも音量を大きめに設定する、ヘッドホンを手で押さえるなどである程度回避できる場合もあります。

遮音性の高いヘッドホンの代替の機種について：

ある程度の遮音性を持つヘッドホンも各社から販売されています。

例えば、KOSS 社製 ヘッドホン QZ99 などがあります。

機種の選定時に注意する点としては、遮音性を謳っているイヤホン・タイプのものが挙げられます。遮音性は確かに高いのですが、自分の発話音声が強調されてしまいます。

通常のヘッドホンを利用する場合には、マイクロホンと一体となった「ヘッドセット」を用いることが出来ます。この場合、遮音性はあまり期待できませんが、機器のセッティングが手軽になるという利点もあります。また、マイクロホンのオン・オフ機能が付加されており、復唱担当者が校正担当者に指示を出したりする場合に有効です。

参考：

・Plantronics 社製　のマルチメディアステレオ PC ヘッドセット .Audio 500 USB
http://www.plantronics.com/japan/jpn/products/computer/multi-use-headsets/audio500-usb

５－２．校正に関する作業負荷を軽減させる機材について

校正者は、復唱者の復唱作業を経て生成された字幕を見ながら、誤字脱字を修正します。しかし、通常復唱者によって生成された字幕が校正者に届く頃には、元々の教員の発話からかなり時間がたってしまっているため、字幕を見ても内容の正確さが判断できないことがあります。こうした状況を改善し、教員の音声と字幕を照らし合わせながら修正作業を行うために、音声を遅延させて再生するソフトウェアを使用することができます。

５－２－１．音声遅延再生用ソフトウェア「SR-DELAY」

この音声遅延ソフトウェアは校正担当者の作業をサポートするための Windows 対応のソフトウェアです。通常、聴取した音声と、校正担当者用ＰＣへの文字データ到着または校正時までの時間差が数秒存在します。このため、「SR-DELAY」を利用して意図的に講師等の音声を遅延させ、文字データ到着との時間差を少なくさせます（図３６参照）。校正担当者は、講師または復唱担当者の発話内容を、少なくとも校正作業終了までは記憶しておかなければなりません。通常、復唱担当者から送られてくる字幕データは、講師または復唱担当者の発話開始から数秒経過した後に音声認識ソフトウェアで生成され、校正担当者 PC に送られてきます。この時間は短縮できない遅延です。この遅延にさらに、校正待ちによる時間遅延が加わります。復唱担当者からの字幕データに多数の誤字脱字が含まれていた場合、一文を校正するのに時間を要し、校正待ちの字幕データが蓄積してしまい、さらなる遅延に繋がります。

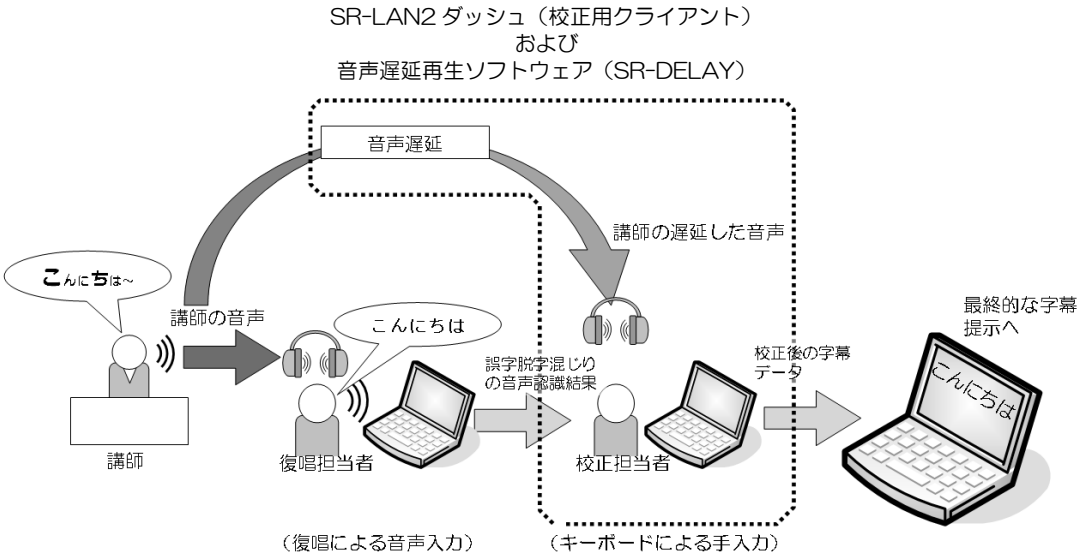


図３６　IPtalk と 遅延再生ソフトウェア

字幕データと音声との比較作業の負荷を軽減する目的で、「遅延再生プログラム」を利用します。このプログラムは、校正担当者用の PC で稼働させることができます。

図３７に、SR-DELAY の外観を示します。

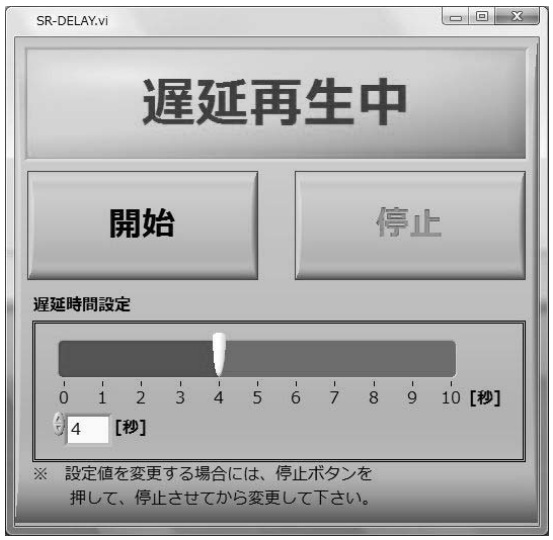


図３７　音声遅延再生用ソフトウェア「SR-DELAY」

SR-DELAY のインストール方法ならびに使用方法は以下の通りです。

①「SR-DELAY200812 インストーラ」フォルダ内の「setup.exe」を起動し、指示に従ってプログラム本体をインストールします（図３８）。

②「遅延時間設定」箇所で、遅延させたい時間を設定します。白い調整ツマミでスライドさせて設定することや、直接、数値を入力することで設定できます。

図３８　SR-DELAY インストーラ
（インストール中のウィンドウ）

「開始」ボタンを押すと、指定された「既定のサウンドデバイス」の入力端子に入力された音声が、出力端子から遅延されて出力されます。

デフォルトでは、遅延時間は４秒に設定してありますが、遅延時間設定用のスライドバーを操作、または数値の直接入力によって変更することができます（図３９および図４０）。

（※東京大学、および㈱BUG のシステムを参考にしました。）

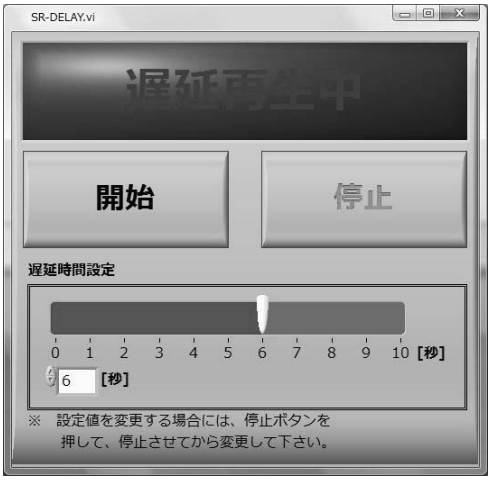


図３９　遅延時間の変更１

図４０　遅延時間の変更２

また、予め、入力端子には音声の信号源を接続しておく必要があります。音響機器のスピーカー出力端子から信号を得る必要がある場合、「抵抗入り」の変換ケーブルを利用し、音声信号が振り切れないようにして下さい。ノート PC での利用時には、サウンドブラスターPX のライン入力および出力端子を利用すると、音響機器との接続が比較的容易となります。

もしも入力音声が上手く出力されない場合には、サウンドデバイスの入力音源選択や出力の選択を確認して下さい。また、「サウンドとオーディオデバイスのプロパティ」から「オーディオ」にて、利用するデバイスを選択して、「既定デバイスのみを使用する」にチェックを入れてみる、また、入力端子がマイクになっているかラインになっているか等、ケーブルを接続している端子と設定が合っているかを確認して下さい。

（※ SR-DELAY は、㈱YAMAHA 社製 SPX2000 に代表されるような音響メーカー各社から販売されている音響信号の遅延装置や、㈱BUG 社製 VideoBOX のような音声遅延再生機能を有しています。なお、VideoBOX では映像信号も遅延表示できる機能が備わっています。）

６．復唱方式における情報保障者のタスクについて

システム構築以上に重要なのが、復唱タスクおよび校正タスクを速やかに実施する情報保障者側のスキルです。このスキルがある一定上を上回っていない場合、不十分な字幕精度や字幕表示までの大幅な遅延時間発生の原因になります。特に復唱スキルが低い場合では、校正担当者に人数を増加させざるを得なくなるなど、人員構成の変更を余儀なくされるので注意が必要です。

現在、各大学・研究グループで、復唱方式による音声認識同時字幕関連の実験的な試みがなされています。この章では、各タスクに関する一般的な説明のみに留め、以下に示す各タスク等についての事例の詳細を、「７．各大学・研究グループでの実践例」で触れます。

６－１．復唱担当者のタスク概要

復唱担当者は、講師の音声を聴取しながら、同様の内容を明瞭に「復唱」するタスクを果たす必要があります。この際、「聴く」ことから注意が離れないようにすることが重要です。聞き取れない場合、内容すら取得できなくなり、要約文すら発話出来なくなってしまいます。練習時には、「音声認識ソフトウェアに適した発話手法の習得」や「復唱能力の習得」が必要になります。

一般に、復唱担当者は同時に１名で実施し、一定時間で交代します。

６－２．校正担当者のタスク概要

校正担当者は、復唱者から送られてくる誤字脱字を含むデータを、遅延を加えられた講師の音声と比較を行い、正しい場合には確定し、誤っている場合には該当箇所を直す役割を担っています。特に、音声認識ソフトウェア特有の同音異義語の誤りや、助詞や名詞の一部を巻き込んだ誤りに対して、「音韻的な類似性」に注意しながら校正作業を実施する必要があります。校正候補の蓄積によって、講師音声の記憶が薄れ、校正作業に支障をきたさないように、手際の良い作業や人員配置が必要です。

一般に、校正担当者は同時に１～２名で実施します。また、一定時間で他のメンバーと交代することもあります。

７．事例１：筑波技術大学における情報保障実験等から得たノウハウ

７－１．復唱担当者のタスクについて

現在、筑波技術大学で実施している音声認識による情報保障実験では、復唱担当として手話通訳経験者が多くを担っています。手話通訳経験者の中には講師音声を聴取しながらそれを聴き貯めて発話し直すという「復唱」作業を問題なく実施できる方から、講師の音声が早くなると時々発話が不明瞭になる方もいます。

一方、情報保障実験や復唱能力に関する実験結果からは、通訳や復唱も未経験である方の場合には、⑴講師音声の聴取と記憶が難しい　⑵自分が発話に集中すると聴取が難しくなる　⑶発話音声が講師音声と混合して聞き分けることができず、聴取が難しくなる　というような感想もあります。講師音声が綺麗に文単位で、流暢に、且つ十分な間も伴って発話するような場合には、それらの困難さは軽減されますが、講師の発話速度が増すにつれて、上述の経験者との差が顕著になります。

復唱方式の音声認識同時字幕では、復唱担当者が作り出す音声認識結果（字幕データ）の精度をある程度見込んで人員構成を考える必要がありますので、９０％台の認識率が望ましく、練習が必要となります。復唱担当者の能力を高めるためのトレーニングとしては、北海道大学の研究グループ（現在、東京大学伊福部研究室）の研究報告が挙げられます。これらの研究報告では、トレーニング内容として数分間分の録音音声を聴取して出来るだけ正しく復唱するというタスクを、短いトレーニングとして時間を置いて数回繰り返すだけでも多くの方で復唱の精度がかなり上昇することを報告しています（中には復唱作業自体に不向きな方もいるようです。その場合には校正作業で力をふるって貰うという選択肢もあります）。発話速度がゆっくりなビデオコンテンツ等で復唱トレーニングを開始し、その後、慣れきたら速度の速いものでトレーニングを実施するのが良いでしょう。

筑波技術大学では、復唱担当者のトレーニング項目として２つが挙げられると考えています。一つ目は「音声認識ソフトウェアに向いた発話方法の習得」、そして２つ目は「講師音声を聴取しながら、少し遅れて発話し続けるという復唱能力の習得」です。

７－１－１．音声認識ソフトウェアに向いた発話方法の習得

高い認識精度を維持するためには、利用する音声認識ソフトウェアの特徴に合わせて発話する必要があります。以下、各社共通に言えると考えられる注意項目を示します。

・単語単位のような断続的な発話は避け、できるだけ１つのセンテンス単位に区切って発話すること。
・通常の発話時には気付きづらいのですが、発話時の「単語等の構成音（音素）」の省略などをできるだけ避け、明瞭に発話すること。
・１文の発話中に、発話速度を大きく変えないようにすること。
・朗読ボランティアのように発話に感情を込めず、普通のイントネーションの範囲で発話

すること。

７－１－２．トレーニング案

パソコン上で音声認識ソフトウェアを起動し、簡単な文書を用意し、それを読み上げて文字化してみる。

実施中、語認識の出た箇所を何度か注意しながら発話し直し、自分の発話音声に問題がないかチェックしましょう。また、発話音声を音声認識ソフトウェアに登録も行い、音声認識ソフトウェア上の個人用データも鍛えましょう。

ここで、音声認識向きの発話方法（自分のスキル）と発話者にぴったりとあった個人データ（音声認識ソフトウェア側の性能）が得られ、総合的な認識率向上に繋がります。

７－１－３．技術的なコツ

口とマイクとの距離を一定に保ち、ＰＣへの入力音量のブレを少なくなるように注意しましょう。正しい発話であっても音量が小さすぎたり、大きすぎたりするとそれだけで認識精度が低下してしまいます。

復唱作業中に誤変換を発見し、校正担当者がその修正に手間取っている場合には、積極的に誤変換箇所の正しい読みを音声で知らせましょう。その際、マイクロホンを手で握ったり、スイッチをオフにする必要があります（マイクロホンによってはスイッチの無いものあります。本マニュアルで紹介したマスクタイプのマイクロホンや指向性の高いマイクロホンには付加されていません）。音声認識ソフトウェア AmiVoice ES 2008 にはオン／オフを切り替えるショートカットキーを登録することができ、この機能が便利です。使用する PC のキーボード配置に合わせて押しやすいキー（例えば、キーボードの左手前に配置されている Ctrl キーなど）を選定し、登録しておけば、このキーがマイクのオン／オフスイッチになります。この機能を利用して、積極的に誤変換箇所を校正担当者に教え、字幕品質が低下しないように配慮しましょう。

７－１－４．講師音声を聴取しながら、少し遅れて発話し続けるという復唱能力の習得

講師音声を聴取・記憶し、その後少し遅れて同じ内容を発話することを「シャドーイング」と言います。通訳や語学学習で良く登場する用語ですが、このシャドーイングに「音声認識ソフトウェアに向いた明瞭な発話」を同時に実施する必要があるのが、音声認識ソフトウェアを用いた復唱方式の特徴です。

先にふれたように、発話速度がゆっくりなビデオコンテンツ等で復唱トレーニングを開始し、その後、慣れてきたら速度の速いものでトレーニングを実施するのが良いでしょう。音声認識ソフトウェアを併用して、文字データを取得してその精度を確認しながら実施しても良いですが、同時に様々なことに注意を払いながら復唱することが困難な方が多いと思います。そのような場合には、復唱のトレーニングのみ実施し、その後、音声認識ソフ

トウェアの併用で全体のトレーニングへと移行してゆく方法をお勧めします。

復唱に慣れないうちは、講師音声の聴き貯め（記憶）が難しく、それを補うために早めについて行こうと心がけてしまいがちになる方もいます。そうした場合、講師の発話の躓きや言い直しに対応できず、最終的な認識率も落ちてしまうことがあります。このような場合には特にトレーニングによって、講師音声の「聴取・記憶・理解」の過程を強化する必要があろうと考えています。聴き貯めが可能になると、講師の一つ一つの発音の間違いや言い直しに翻弄されることもなくなり、また不要語の省略等にも対応しやすくなります。

ところで、AmiVoice は精度的な利点は高い一方、連続発話時においては文字データ確定までの時間が他のソフトウェアより長いので、意識的に発話文同士に１秒程度の間を入れて発話する必要があります。そのような「間」の挿入によって、校正担当者への字幕データ送出を早める必要があります。そうしないと、字幕データが校正担当者に到着までの時間に更に遅延が発生し、それに伴って１度に送られる文字データ量も蓄積（増加）してしまいます。このような遅延の発生により、校正担当者が記憶しておくべき講師音声内容も増加し、校正作業そのものを実施しづらくさせ、実質的には文法チェック程度の処理しか出来なくなってしまいます。音声認識ソフトウェアは今後も各種のものが登場してくるかも知れませんので、各ソフトウェアの動き・特徴を実際の利用から把握し、上手く利用できるように体制を整えることが重要となるでしょう。

このような復唱作業の困難さを技術的に補う方法として、遮音性の高いヘッドホンの利用によってある程度軽減できるという手法もあります。工学的な機器で補えるタスクは機器にまかせ、どうしても人がしなければならないタスクのみ集中的にトレーニングし、スキルを身に付けられるように情報保障者の負荷の軽減が急務と言えます。

７－１－５．明瞭な発話のための準備

発声器官である口腔部の動きは、筋肉の動きで作られています。ですから、正しく発話するためには、事前の適度な「ストレッチ」が必要になります。運動の場合１０分程度のストレッチが適しているようですので、早口言葉などを発話してみて、自分が発話しづらい箇所を重点的に練習しましょう。

７－１－６．復唱作業のコツ

音声認識ソフトウェア利用中に、“言い間違え”や“躓き”が出てしまうことが必ずあります。発話した文章に、一部分だけ発話誤りを含んでいる場合には、一文丸ごと言い直しを行う方が、校正担当者の負担としては、削除作業だけですので、全体として得策な場合もあります。そのような手順について、校正担当者との申し合わせをしっかりと行っておくことをお勧めします。

また、実験的な情報保障を経験した復唱担当者側の工夫・注意点を以下にまとめました。

・ 発話しづらい言葉、または認識されづらい言葉を復唱する場合には、意識してゆっく

りと発話するように心がける。

- 認識しにくそうな単語はハッキリ言うようにする（特に、マ行やナ行など）。
- 認識されやすい復唱のために、ハッキリと発話する。
- 何を言っているのか予測・理解できない講師音声の場合には、多めに聴き貯めをして復唱する。
- ほとんど正確に認識されないことが判っている単語の場合には、直接キーボードで入力して送信する。その後、復唱で全文を完成させる。
- 講師の音声に合わせて発話速度を変化させて復唱する。
- 「指示語→具体的な名詞」への置き換えを行う。
- 文の区切りをつけ、校正担当者への１回の送信文が複数個の文章に及ばないようにする。（細かな文や文節単位で、校正担当者へ割り当てられる様に注意する）
- 誤りが一定しない場合には、自動置換の機能は意味がない。
- 専門用語は予め、音声認識ソフトウェアに登録しておく。

７－１－７．交代時間

講師の発話スピードが遅かったり、演習等を挟むような無音時間が比較的多くある講義の場合には、交代無しで９０分程度実施し、担当者の疲労も特に無いという感想を得た例もあります。通常は、２０分程度で交代しています。しかし、復唱担当者側の感想としては、講師の発話速度にも依りますが、２０～３０分程度、または３０～６０分程度でも大丈夫という人もいます。

（※ 東京大学の研究グループの研究結果では、不慣れな人程、復唱による認識制度が長続きしないというデータもあり、個人差はあるでしょうが、数十分で交代すべきでしょう。）

７－１－８．字幕による講義内容理解のための配慮

復唱方式による音声認識同時字幕は、字幕提示までの時間遅延の問題もあり、講師音声中の指示語の置き換えによる「内容理解」に対する配慮を考える場合もあります。この場合、復唱担当者のレベルで指示語を置き換えると、次の校正作業が楽になります。指示語を置き換える場合には、講師側の映像、特に講師の挙動と板書やパワーポイントやキーノートなどのプレゼンテーション資料の映像が重要になります。

７－２．校正担当者のタスクについて

校正担当者は、復唱担当者用ＰＣから受け取った音声認識結果に含まれている誤字脱字をチェックします。講師の音声または復唱担当者の音声を聞きながら、その音声と音声認識結果である文字列を照合します。合致していた場合には確定（表示用ＰＣに送信）し、間違いがある場合には、その該当箇所を修正し、その後、確定します。作業のポイントとしてはキーになる用語や数字を暗唱するような工夫をし、できるだけ頭の中で保持するよ

う心がける必要があります。特に、講師音声が速い場合など、校正箇所が増加した場合に有効と思われます。そのような状況時には、休憩中の他の担当者が、判断に困っている校正担当者に対して音声で指示を与えてあげると良いでしょう。

字幕データ到着後に該当音声（遅延音声）が聞こえてくれば、校正作業が容易なのですが、常にそういう状況が発生するとは限りません（個人差があるようです）。それを回避するために、更に講師音声に遅延を加えると校正作業自体は容易になりますが、字幕表示までの時間が増加し、リアルタイム性が失われ、聴覚障害学生の満足度も減少することでしょう。このように遅延音声との照合作業においても、講師音声の聴取と記憶・保持が重要なスキルとなります。特に、復唱担当者が聴き貯めに多くの時間を使う局面では、元音声の聴取ではなく、遅延音声の聴取が好ましい場合があります。

ところで、音声認識ソフトウェアの利用時に発生する誤字脱字には、次のような例もあります。

講師の音声：「この魚はいくらですか」
音声認識結果：「この坂俳句らですか」

もしも源音声を聴取できない聴覚障害学生が上記の変換結果を見た場合、正しい情報を推測することは不可能でしょう。聞こえる校正担当者であっても、このような「てにをは（助詞）」と単語を巻き込んだ誤変換を校正することはなかなか困難が伴います。校正担当者は、このような「音韻的には類似はしているが、全くことなる漢字仮名交じり文」の照合を行わなければならないことが多くあるということを、予備知識として知っておくべきでしょう。

７－２－１．校正作業のコツ

実験的な情報保障を経験した校正担当者側の工夫・注意点を以下にまとめました。

- 校正作業時、他の校正担当者の校正状況を確認して、各文章の接続部分に注意する。
- 校正に遅延が生じた場合には、句読点の確認よりも、数値や語句に間違いが無いかの確認を優先する。
- 特に、講師の発話内容に含まれる数値は記憶しておくようにする。
- 講師の発話内容に含まれる数値や専門用語を暗唱し続け、作業に備える。
- 校正すべき字幕データが貯まると、記憶が薄れ、正しく校正できているかどうか自信が無くなる。そうなると校正作業も遅れだし、字幕表示までの時間遅延も増えるという悪循環に陥るので、忘れた場合には他の校正担当者や手の空いている復唱担当者に積極的に助けを求めるようにすると良い。

７－２－２．交代時間と同時作業人数

２名で同時に交代無しで９０分程度実施する場合や、復唱していない復唱担当者が校正担当者のどちらかと交代したり（この場合、２０分程度）、予備のメンバーが交代したりし

ます。交代は不要という人もいれば、２０～３０分程度で交代したいという人もいます。

校正担当者１名のみでの校正作業実施の場合は、非常にゆっくりとした講師の発話速度のときのみ可能であると考えた方が良いでしょう。通常の講義時に１名で対応した場合には、字幕表示までの大幅な遅延発生と誤字脱字が多発・残存することでしょう。また、音声と発話内容の比較ができず、文法的な誤りの確認のみしか実施できなくなり、更に著しい誤字脱字文が来た場合には、全文削除または「？？？」等の内容の推測を促すメッセージしか出せないような「校正作業不能」な状態に陥ってしまいます。

７－２－３．字幕による講義内容理解のための配慮

「７－１－８．字幕による講義内容理解のための配慮」でも触れたように、復唱方式による音声認識同時字幕は、字幕提示までの時間遅延の問題もあり、講師音声中の指示語の置き換えによる「内容理解」に対する配慮を考える場合もあります。この場合、復唱担当者のレベルで指示語を置き換えると、次の校正作業が楽になります。しかし、その作業から漏れてしまった場合には、極力校正担当者が実施しなくてはならなくなります。しかしながら、校正担当者は遅延音声を聴取しているために、聴取する音声と時間的なズレがある「リアルタイム映像」では見逃してしまう可能性が多分にあります。映像も遅延させられる機材を入手または構築出来れば良いですが、かなりの工夫を要します。

できるだけ復唱担当者のレベルで実施しておきたいタスクと言えるでしょう。

７－２－４．校正担当者が聴取する音声の選択について

復唱担当者が語尾を整文したり、「えー」などの冗長語を省略する場合には、講師音声を聴取するのではなく、復唱音声を聴取したくなる場合もあります。この場合、校正担当者は復唱担当者の発話誤りに気付くことができないこともありますので注意が必要です。また、遮音性の低いオープンエア・ヘッドホンで同室の復唱担当者の音声を同時に聴くことも可能ですが、慣れていない場合には、聴取すべき２つの音声が混合し、どちらも聞き取れなくなってしまう場合もあります。

７－３．実施体制について

実験の当初は、校正担当者１名体制での実施も試みました。しかし講師の発話が瞬間的に速くなる時など、校正作業が追い着かず、個人差はありますが作業に支障をきたすケースも多々ありました。よって、講師音声が最初から非常にゆっくりであることが判っている場合を除いて、校正担当者１名体制は避けた方が良いでしょう。

現在は、復唱担当者２名（２０分交代）、校正担当者２名（基本的に連続ですが、場合によっては復唱担当者と交代する、または、補助人員がいる場合にはその人と交代することもあります）で試行しています。

７－４．別室（遠隔）からの情報保障時の講師映像の必要性

講師側の映像は、指示語を置き換えるためだけではなく、情報保障者のストレス軽減に大きく寄与するようです。映像情報が無い場合では、別室にいる情報保障者にとって教室側の状況を知るための手段は音声のみとなってしまい、もしも無音の場合では「講義開始時間なのに、講師がまだ来ていないのか？」、「講義中に、学生に何かタスクを与えて待っているのか？」、それとも「何か通信トラブルが発生したのか？」判別ができません。

また、講義の様子を確認する事によって、その各状況に合わせて「緊張」と「リラックス」のメリハリを付ける事ができ、結果的に長時間の作業実施が可能になります。特に、復唱担当者の場合には、映像によるメリットが高いようです。

８．事例２：

群馬大学における音声認識技術を活用した字幕呈示システムの運用の取り組み
−聴覚障害学生の発言権を保障していくための工夫−

金澤貴之（群馬大学教育学部　准教授）
味澤俊介（群馬大学学務部学生支援課障害学生支援室　専門支援者）

８−１．群馬大学における運用方法

音声認識技術による字幕呈示の場合、話し手の音声情報をほとんど変えることなく、情報保障ができるという点に大きなメリットがあるといえます。しかしその反面、聴覚障害学生が情報の一方的な受け手になってしまい、自ら主体的に発言をしてくことが難しくなってしまうという点に、どのように配慮していくかが課題であるともいえます（ただし、情報の遅延が生じること自体は音声認識だけの問題ではなく、パソコン要約筆記（ＰＣテイク）にしてもほとんど類似した問題を抱えることになります）。

群馬大学では、手話母語話者の聴覚障害学生（日本語の音声発話によるやりとりは困難）が参加する授業の中で、音声認識システムを使用した場合、どのような工夫が必要かについて検討を行ってきました。

群馬大学では場合、防音機能を備えた復唱室で復唱作業、修正作業を行います。復唱者２名、修正者２名が大きなテーブルを囲んで座り、字幕作成作業にあたります。教室から復唱室への音声の送信は、基本的には Skype を使用します（環境が整っていれば内線電話を優先）。修正者は、音声遅延装置により４秒遅延された音声をヘッドフォンで聞き、Skype で送られてきた教室の映像を見ながら、修正作業を行います。したがって、字幕作成はすべて復唱室で行い、作成された字幕を教室へ配信する形をとっています。これが、いわば群馬大学での音声認識システムの基本形といえます。

字幕の表示方法は、日ごろのＰＣテイクで学生自身が希望している方法となるべく近い環境で表示するようにしています。

通常のＰＣテイクでは文字色や文字の大きさは、個々の利用者にあわせた設定にしていますが、特に希望がない場合は、句読点での改行をおこない背景色を「黒」文字色を「白」として行っています。

また、話し手特有の言い回しや用語を聴覚障害学生も味わい、他の学生と授業の臨場感を共有するために、ＰＣテイクの際には、できるだけ要約はせずに、講師の話した言い回しをなるべくそのまま示すように取り組んでいます。その点は、話し言葉をほぼそのまま字幕化できる音声認識システムと親和性が高いといえるかもしれません（情報量は音声認識システムの方が多いといえますが）。

表示媒体については、聴覚障害学生が表示端末を持ち、自分の好きな場所に座り、無線ＬＡＮを使って字幕を送信する方法もありますが、特にゼミ形式の少人数制の授業で、ディスカッションが想定される場合には、プロジェクタでスクリーンに字幕を表示する方法

を採用することで、教員も含めたその場にいる参加者全員が字幕情報を共有し、字幕の進行状況を確認しながら授業を進めることができます。教員が学生に質問をする時、あるいは他の学生が意見を言う時など、話者交代のタイミングにあわせて字幕を確認するだけで、聴覚障害学生の参加のしやすさは格段に向上します。その際には、「情報保障は、聴覚障害学生のためのものだけではなく、そこにいる双方が情報を共有するために存在する」ということを参加者全員に理解してもらうことが重要です。

8−2. 手話使用者への対応の工夫について

音声で発話することができる聴覚障害学生の場合、前述した「皆が字幕を共有する」という方法で、状況の改善は図られると言えますが、聴覚障害学生の中には、音声で話をすることが難しい方もいます。聴覚に障害があることは、発音の習得に困難さが生じます（聴力や教育環境などさまざまな要因が絡むため、いちがいには言えませんが）。そのため発音自体が明瞭ではない学生もいますし、人前で音声を発することへの苦手意識やその人なりの障害観など、心理的な理由から人前で音声で話したくないという学生もいます。では、そうした学生の発言方法をどのように確保したらよいでしょうか。

手話を日常的に使用している学生の場合、学生が発言する際に備えて、手話の読み取りができる者を教室に配置するという方法が考えられます。その上で、本人が発言をするときには手話による発言を手話通訳者が音声化し、その音声について、復唱、修正作業を経て教室で表示することが考えられます。その際、表示方法としては、手話通訳者や教員、学生が確認できるよう、スクリーン等で教室全体に表示すると同時に、手話で発話した聴覚障害学生本人も字幕が確認できるように配慮する必要があります。特に聴覚障害学生本人が前に立って話をする場合には、前方と後方の両方に表示画面を用意するなどの配慮が必要となります。群馬大学では、前方にあるスクリーンと、後方にある 50 インチ液晶モニタを活用しています。

ただし、発言する手段を確保するだけでは十分ではありません。表示される字幕と元の音声との間で生じるタイムラグがあるため、聴覚障害学生が発言しにくい状況が生じるという点は、前述した通りですが、その上に手話通訳を介在させるわけですから、参加者がその状況を十分に理解していなければ、相当な遅延を実感することになります。そのためにも、「今、ここで何が起きているのか？」を参加者が皆、確認できる環境を用意しておくことが重要です。

手話の読み取りができる者がいない場合には、聴覚障害学生へのモニタ用のＰＣにチャット機能を設け、文字媒体で発言の手段を保障する方法も考えられます。その場合は本人によるタイピングで発言することになりますから、さらに遅延が発生します。その結果としての「じれったさ」によるストレスは、「聴覚障害学生がいるから」ではなく、双方にとって情報保障が必要であり、それが十分に整っていないための「歩み寄り」のために必要な時間なのだということを、十分に参加者全員が理解していることが重要になるでしょう。

8−3. 復唱者・修正者からの要望とその改善策

復唱者また修正者は、群馬大学では、日常的に連係入力によるＰＣテイクを行っており、十分に習熟している学生テイカーに依頼しています。それは、ＰＣテイクでも音声認識による字幕システムでも、復唱や修正に必要な一時保持記憶能力や理解能力など、同じ技術が必要となると考えているからです。また、修正者は迅速に文字の誤りを修正する必要があるため、一定程度のタイピングの技術も必要であり、この点でもＰＣテイクの技術は活かされるといえます。

復唱者、修正者の意見として、最も字幕作成が困難となる状況としてあげられるものが、ゼミなどの、話者交代が頻繁に起こる場面を多く含む授業です。話者が頻繁に変わる場合、第一に、教室内でなんの配慮もしなければ、話者交代の際に音声の重なりが生じます。場合によっては、瞬間的に音声が重なるだけでなく、まとまった長さで複数の人が同時に話をしてしまう場合もあります。聞こえる人たちは、その場にいて、複数の音声による相互作用の中で決定される優位な音声を選択的に聞き分けて会話に参入することもできます。しかし字幕作成者は同時に複数の音声を作成することはできず、音声が発せられる瞬間により優位な音声を選択的に決定することは極めて困難です。

また、復唱者は別室で作業をしているため、話し手が誰なのかを判断することが困難です。修正者は同じ教室にいたとしても、復唱作業の効率化のために遅延音声を聞いて作業をしているため、やはり今話している人が誰なのかを特定することは困難です。そのため、話者交代を伴う授業で音声認識システムを運用する場合には、教室内の利用者全体で字幕を意識して話をする必要があります。

具体的な工夫の方法としては、第一に、音声が重ならないようにすること。より具体的には、常に挙手をして、司会者（ゼミの場合は教員など）が指名してから発言するよう、ルールを徹底することが重要です。加えて、聴覚障害学生が発言する権利を保障するためには、字幕が十分に追いついているかどうかを確認し、聴覚障害学生が発言する意思の有無を確認してから司会者が指名することも必要になります。

第二に、話者交代が起きる際には、話し手が最初に「〇〇です。」と自分の名前を言ってから発言を始めることも重要です。通常、ＰＣテイクの場合には、テイカーの側で「〇〇／」と表示をすることができますが、音声認識による字幕システムの場合、復唱者・修正者側で話者を特定することが困難であるため、発言者が自ら名前を明示することで補う、といった工夫が求められるということです。

これらのルールを設けることは、確かに、利用者にとっては日常の会話パターンと大きく異なる方法での会話を強いられることになり、ストレスがかかります。しかし、逆にこうしたルールを設けないことは、復唱者・修正者に過度の負担がかかり、あわせて聴覚障害学生が発言できない状況を作り出してしまうことになります。情報保障は聴覚障害学生のためだけにあるのではなく、その場にいる利用者全員のためにあるということ、字幕作成者に任せれば解決できるわけではないということについて十分な共通理解をもち、利用

者側でできる工夫を行っていくことも必要であると言えるでしょう。

ルールの徹底を行いやすくするための具体的な工夫の１つとして、あえて１本のマイクを使用し、発言をする際にはそのマイクを相互に利用するという方法も有効です。利用者の話しやすさを優先するならば、複数のマイクを使い分ける方法もありますし、さらには会議用の音声集音装置を中央に置く方法もあります。しかし実際に使用した結果、一応はルールを守って話そうとしていても、日常的な会話の方略が染みついているために、挙手すると同時に発言していたり、うっかり名前を言わずに話してしまったりすることが頻発してしまいました。学生・教員からは「非常に快適であった」との感想が聞かれる一方で、復唱者・修正者からは、「誰の発言かもわからず、誰の会話を拾ったら良いのかもわからない。二度とこの方法では行いたくない」という厳しい意見が聞かれました。マイクを１本に統一することは、教室内の利用者に制約をかけることで、むしろ復唱・修正作業者の負担の軽減に役立つという面がみられました。

また、「教室の状況を把握しにくい」と言う問題を解消するために、教室に Web カメラを設置し、音声と映像を同時に送るようにしています。映像がないと、例えば、教員が来たのかどうか、音が届かないがこれはなぜなのかといったことについて把握することができないため極めて不安な心理状態におかれることになります。その不安感の解消のためにも、映像情報の確保は重要です。

復唱用の認識率を上げるためには、復唱者の技術向上も必要ですが、それだけでなく事前の単語登録も重要です。これにより人の名前や特殊な専門用語をより確実に表示することが可能になります。群馬大学での失敗例として、不要な単語の登録を消去する作業を行わなかったために、かえって誤認識が増えてしまうということがありました。そのつど登録と削除を行い、その場にあった単語登録を行っていくことも必要だと考えられます。

また、復唱者が安定して復唱を行うために、十分に良質の音声を届けることは非常に重要です。音声伝送には Skype を通常利用していますが、回線の問題またはソフトウェアの問題で、音が途切れてしまったり、聞き取りにくくなったりすることがしばしばありました。そうした場合、復唱者は前後に使われている単語や文脈から、その途切れている箇所を補いながら復唱を行っていましたが、そうした作業は非常に負担になっていました。教室の通信環境による制約が生まれることはありますが、安定した音声を届けるためには、可能であれば、現時点では、電話回線を利用したり、隣接する教室であればＡＶケーブルを用いて音声を送信するなどの方法の方が望ましいでしょう。マイクを複数用いる場合に双方の音量をそろえておくといった配慮なども、復唱者の負担の軽減につながります。

復唱者と修正者が同じ部屋で作業をすることで、復唱の音声が修正者に聞こえてしまうということについては、一見デメリットのようにも思えますが、必ずしもそうともいえないようです。復唱者の音声が耳に入ることによって、次にどのような文字が送られてくるのかを予測できる、といった意見も聞かれました。

8−4. 人員配置について

誤認識がほとんどなく、シンポジウムや式典などの公的な場での表示に十分に耐えうる字幕を作成するには、十分に作業に慣れた復唱者２人と修正者２人による体制で、事前に単語登録を十分に行うことが必要になります。

しかしその一方で、毎週恒常的に繰り返される授業で、毎回４人体制で臨むことは、予算的にも人材的にも実現が難しいかもしれません。

群馬大学では、復唱者・修正者が慣れるまでのしばらくの間は復唱者２人（１０分交代）、修正２人の４人体制で臨み、その後、復唱者２人、修正者１人とした上で、１０分ごとに交代する復唱者のうち、休んでいる側の者が、修正の補助に回るという方法を採用しました。このことにより、３人体制による運用を実現することができました。２人体制で運用ができればコスト的にもより利用する機会は増やせるのかもしれませんが、現実問題として、２人体制にした結果、作業者への負担が大きくなりすぎるか、あるいは（かつ）字幕の誤認識の修正が十分に行えず、「情報保障」として耐えうるだけの字幕の質を保てなくなるおそれがあります。

人員を減らした運用を実施していくためには、復唱者・修正者が十分に作業に習熟しているかどうか、過負荷になっていないかを十分に配慮して進めなければならないでしょう。その上で、字幕を運用する場面にあわせて、４人体制で臨むべきなのか、３人体制で臨むべきなのかといった判断をし、効率的な運用を考えていく必要があるでしょう。

９．参考

９－１．パソコン要約筆記のためのソフトウェア IPtalk

IPtalk の使用方法は、IPtalk の付属の説明書や Web 公開されている資料をご参考にして下さい。この説明書では、IPtalk9t6 を用いております。IPtalk は栗田氏が作成したパソコン要約筆記用ソフトウェアです。IPtalk に関しましては、下記の URL をご参照下さい。

パソコン要約筆記用ソフトウェア　IPtalk　　URL　http://iptalk.hp.infoseek.co.jp/

９－２．音声・映像遅延再生のための機器

・映像と音声を同時に遅延再生させることができる機器

製品名：VideoBOX　　（３０万円程度）

㈱BUG：音声認識同時字幕を有料サービスとして提供している国内唯一の企業

URL　http://www.bug.co.jp/products/vbox.html

・音声を遅延再生させることができる機器

製品名：SPX2000　　（１０万円程度）

㈱YAMAHA：音響機器メーカー

URL　http://proaudio.yamaha.co.jp/products/processors/spx2000/index.html

１０．謝辞

・マニュアル作成にあたり、主に復唱担当者の技能面に関して、東京大学・伊福部達教授の研究成果を参考にさせて頂きました。ここに厚く感謝申し上げます。
・ソフトウェア SR-LAN2 ダッシュ開発にあたり、主に表示手法等に関して、NPO 法人日本遠隔コミュニケーション支援協会（NCK）会長・栗田茂明氏開発のパソコン要約筆記用ソフトウェア IPtalk を参考にさせて頂きました。ここに厚く感謝申し上げます。

「音声認識によるリアルタイム字幕作成システム構築マニュアル」編集グループ

代表者　三好茂樹（筑波技術大学障害者高等教育研究支援センター　准教授）
磯田恭子（筑波技術大学　障害者高等教育研究支援センター　特任研究員）
金澤貴之（群馬大学　教育学部障害児教育講座　准教授）
味澤俊介（群馬大学　学生支援課障害学生支援室　専門支援者）
立入　哉（愛媛大学　教育学部　准教授）
苅田知則（愛媛大学　教育学部　准教授）
大倉孝昭（大阪大谷大学　教育福祉学部　教授）
白澤麻弓（筑波技術大学　障害者高等教育研究支援センター　准教授）
河野純大（筑波技術大学　産業技術学部　准教授）
黒木速人（筑波技術大学　産業技術学部　准教授）

音声認識によるリアルタイム字幕作成システム構築マニュアル

発行日：2009 年 9 月 25 日
編　集：「音声認識によるリアルタイム字幕作成システム構築マニュアル」
編集グループ
協　力：日本聴覚障害学生高等教育支援ネットワーク（PEPNet-Japan）
URL　http://www.pepnet-j.org
発　行：筑波技術大学
〒305-8520　茨城県つくば市天久保 4-3-15
筑波技術大学　障害者高等教育研究支援センター
ＩＳＢＮ：978-4-9904374-4-2

国立大学法人
筑波技術大学

※本事業は、文部科学省特別教育研究経費による
拠点形成プロジェクト（筑波技術大学）の活動の一部です。

ISBN978-4-9904374-4-2

音声認識によるリアルタイム字幕作成システム

構築マニュアル

PEPNet-Japan 日本聴覚障害学生高等教育支援ネットワーク