
半自動音声ラベリングシステムの開発

Development of Semi-Automatic Speech Labeling System

鷹見 淳一*
Junichi TAKAMI

加藤 喜永*
Yoshinaga KATO

呂 彬*
Bin LU

佐藤 奈穂子**
Nahoko SATO

要 旨

音声認識や音声合成の研究開発に不可欠な良質の音声データベースを構築する際に最も労力を要する作業は、発話内容に適した音素ラベルの付与とその境界位置の決定である。そこで、音声認識の技術を駆使し、複雑な作業をできるだけ自動化し、使い易いGUIエディタによって結果の修正を容易にすることで、未経験者でも極短期間の訓練で本格的な音声ラベリング作業の実施を可能にする「半自動音声ラベリングシステム」を開発した。システムの概要や自動推定のための内部処理手順、エディタ部の構成、特筆すべき付加機能などを検討する。

ABSTRACT

In building a high quality speech database, which is indispensable to research and development of speech recognition and voice synthesis, the most costly work is the granting of a suitable phoneme label and decision of boundary positions between phonemes. The “Semi-Automatic Speech Labeling System” is developed, which makes it possible to automate this complicated work using speech recognition technology. With its user-friendly graphic user interface (GUI), the system offers a fast learning curve, even for inexperienced users who need to accomplish difficult tasks regularly. Outline of the system, processing procedures for automatic estimates, the constitution of editor parts and several useful additional functions are described.

* 画像システム事業本部 ソフトウェア研究所

Software Research Center, Imaging System Business Group.

** 研究開発本部 オフィスシステム研究所

Office System Research and Development Center, Research and Development Group.

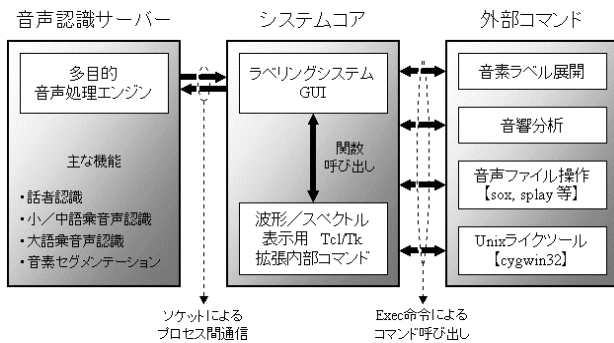


Fig.2 A block diagram of the Semi-Automatic Speech Labeling System.

[1] システムコア

システムの核となる全体制御部やシステムの操作性に大きく関与するGUI部など、本システムの主要な部分は、プログラミングの容易さと柔軟性の観点からTcl/Tk[2]のスク립トで記述されている。

Tcl/Tkは、PerlやPythonと並ぶ三大スク립ト言語として世界中で広く利用されているフリーソフトであり、その最大の特徴はマルチプラットフォームで動作可能なGUIプログラムを非常に簡単に作成できるという点にある。

更に、Tcl/Tkには、CやC++等で記述したプログラムを内部コマンドとしてインタプリタに組み込むための標準的な手段が提供されており、機能の拡張をユーザーが手軽に実現できる[3]という、もう一つの特徴がある。これによって、他のスク립ト言語と同様、Tcl/Tkでも最大の欠点となっている処理速度の遅さを克服することができる。

本システムでは、GUI部と密接に関連しながら処理の高速性が要求される波形やスペクトルの表示を行うための機能をTcl/Tkの拡張内部コマンドとして実装している。

[2] 音声認識サーバー

音声認識サーバーは、様々な音声情報処理機能を有する汎用的な知的音声処理エンジンであり、高速な処理を実現するためにC++で記述されている。

本エンジンで実現可能な主な機能は、話者認識、大語彙向き音声認識(A*探索法ベース)[4, 5]、小/中語彙向き音声認識(One-Pass DP法ベース)、音声セグメンテーションなどである。

一般に、上記のような複雑な音声処理を実現するプログ

ラムでは、起動時にメモリ空間の確保や音響モデル/文法情報などの読み込み、文法情報から有限状態オートマトンへのコンパイルなど、ある程度の計算量を要する一連の初期化処理を行う必要があるため、単純なコマンドとして実装し、処理の度に毎回起動するという使用法では効率が悪い。

そのため、本エンジンはサーバーとして動作するように設計されており、処理対象音声データの転送や機能選択、動作条件(音響モデル、文法、各種パラメータなど)の変更などは、全てTCP/IPのsocketを利用したプロセス間通信によって動的に行うことができる。

また、音響分析機能は、本エンジンとは独立に外部コマンドとして実装されている。これによって、比較的単純な音響分析コマンド側を変更するだけで、エンジン本体のプログラムには一切手を加えずに任意の音響パラメータが使用できるという高い汎用性が実現されている。本システムでは、音声処理エンジンを以下の条件で駆動している。

音響モデル	音素環境依存HMM[6], 総状態数: 400, 分布混合数: 5
音響分析	サンプリング周波数: 16kHz, 分析フレーム長: 10mS, 分析フレーム周期: 5mS(サンプリング周波数が16kHzより高いサンプルに対しては、自動的に16kHzにダウンサンプリングされる)
音響特徴量	10次元LPCメルケプストラム+10次元 Δ LPCメルケプストラム+ Δ 対数パワーの計21次元ベクトル

[3] 各種外部コマンド

これまでに述べたもの以外で、比較的単純ではあるが高速性が要求される処理や、フリーソフトとして配布されているいくつかの有用なプログラムは、全て外部コマンドとしてTcl/Tkのexec命令によって直接呼び出して使用している。

具体的には、仮名文字列から音素ラベルリストへの展開を行うためのフィルタ、音声データから音響パラメータを計算するコマンド、音声ファイルの変換や再生のためのアプリケーション、プラットフォーム間の細かい差異を吸収するためのユーティリティなどを外部コマンドとして使用している。

2-1-2 自動ラベリングのための内部処理手順

本システムでは、音素ラベル列や音素境界位置の高精度かつ高速な自動推定を、以下の処理手順で実現している。

【処理1】最適音響モデルの選択

事前に構築されている複数の音響モデルの中から、作業対象の音声データに対して最も高い推定精度が期待できるモデルを探索し、それ以降の処理で使用する標準の音響モデルとする。

この処理は、その時点で処理対象となっている音声データに対して、【処理3】で述べる音声認識処理を、音響モデルを入れ換えながら実行して音声認識のスコア(尤度)を計算し、全ての音響モデルの中で最大の尤度を示すものを探索するという方法で実現されている(Fig.3参照)。

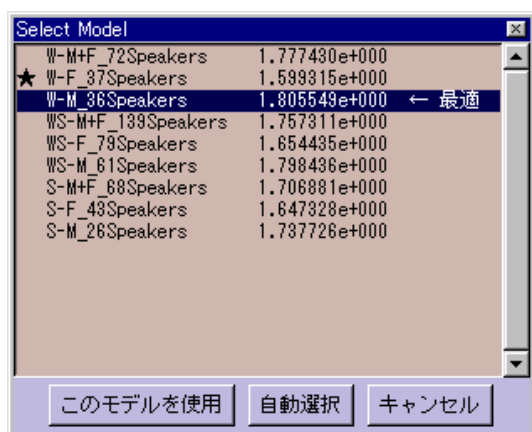


Fig.3 Automatic selection of the optimal acoustic model.

現在の音声認識技術では、統計的な音響モデル(HMM)を用いるもの[7]が主流となっており、その基本的な認識性能は、音響モデル自体の推定精度や発話者と音響モデルの相性

によってほぼ決定される。本システムで実施される音素ラベル列や音素境界位置の自動推定精度も、音声処理エンジンで使用する音響モデルに大きく依存する。

もしも真の意味での不特定話者用音響モデルが存在するのであれば、話者に依らず高い推定精度が得られるのかもしれないが、そのような汎用的な不特定話者用音響モデルの獲得を目指してデータベースの整備を進めている現段階で、真の不特定話者用音響モデルを用意することは難しい。

そこで、音響モデルとして、利用可能な全てのサンプルを用いて学習した不特定話者向きモデルの他に、男女別、発話スタイル別など、様々な観点で分類した学習サンプルのサブセットを使用して学習したモデルも合わせて用意し、それらを適切に使い分けることで推定精度の向上を図っている。

【処理2】読み仮名文字列から音素ラベルリストへの変換

音声データに付与されている読み仮名文字列に基づき、IPA (International Phonetic Alphabet : 国際音声字母)の定義に準ずる独自の音素ラベル体系に従った音素ラベル候補のリストを生成する。

本システムで使用する音素ラベルは、作業の効率化を考慮して、IPAの音素記号に対して多少の拡張を行ったものとなっている。

この拡張の主な目的は、日本語の詳細な特徴への特化と作業の効率化である。特に作業の効率化に関しては、無声子音と後続の無声化母音の境界や、半母音と母音の境界、あるいは撥音と鼻子音やバズの境界など、その前後で音響的な特徴にほとんど差異がなく、その位置の特定が極めて困難であるようなものについては、複数の音素ラベルを間にカンマを挟んで連結させた「融合ラベル」として表記することで、その位置を特定することなしに作業が進められるようになって

Table 1 Examples of conversion from Japanese KANA strings to phoneme label lists.

読み仮名	音素ラベルリスト	全展開時のラベル列総数
がっこう	{G (+ g) a Q- k {{o u} oL}	6
よぼうや けんこうかんり りはびりてーしょんの ため の せいでを じゅうじつし ていく ひつようが だろう	y,o {+ {} b {{o u} oL} y,a {- } k e x- k {{o u} oL} - k a x r i { } r i h a {+ {} b i r i- t e L S y,o n,n o {- } t a m e n o { } s {{e i} eL} {+ {} d {{o o} oL} {{+ } _ + } j {{y,u u} y,uL} {+ {} j i {- } {{ts U} ts,u0} {{S i} S,i0} - t {{e i} eL} - {{k u} k,u0} { } {{H i} H,i0} {- } {{ts U} ts,u0} {{y,o u} y,oL} {G (+ g) g} {{a _ a} aL} r {{o u} oL}	603,979,776

いる。これによって、難しい判断を作業者に強要した場合に危惧される作業精度の劣化と作業時間の増加を回避している。

読み仮名文字列から音素ラベルリストへの展開結果の一例をTable 1に示す。

【処理3】最適音素ラベル列の選択

音素ラベルリストを文法として認識処理を行い、音素ラベルリストを展開した場合に得られる全ての音素ラベル列の中から作業対象の音声データに対して最も相応しいものを決定する。

この処理は、音声認識サーバーが【処理2】で得られた音素ラベルリストを受け取り、内部でそれを有限状態オートマトンの形にコンパイルした後、音声認識を実行して最適な音素ラベル列を求めるという方法で実現される。この際に、Table 1にも示した通り、発話内容次第では認識対象の候補数が数億以上に及ぶ場合もあるが、その有限状態オートマトンは比較的単純な形になるため、ここでの認識処理は小ノ中語彙音声認識向けのOne-Pass DP法を用いて行っている。

なお、読み仮名文字列から展開された音素ラベルリスト中に一通りの音素ラベル列しか存在しない場合には、この部分の処理はスキップされる。

【処理4】音素境界位置の推定

最適な音素ラベル列に対して、Viterbiアルゴリズムによる音素対応付けを行うことで、各音素の境界位置を推定する。

Viterbiアルゴリズムによる音素対応付けは、One-Pass DP法を用いた音声認識処理過程で得られる全ての部分仮説に対して各時刻で滞留しているHMMの状態番号を保存しておき、最適経路が算出された後に、その経路に沿って対応するHMMの状態番号をトレースバックすることで実現される[7]。これは、音素対応付け処理がOne-Pass DP法による音声認識機能を完全に包含しており、【処理3】で実施している最適音素ラベル列決定のための処理を音素境界位置の推定と同時にを行うことが原理的には可能であることを意味している。

しかし、全ての部分仮説に対してHMMの状態番号を保存する必要がある音素対応付け処理は、その情報の保存を必要としない通常の音声認識処理に比べて、使用メモリ量や計算時間が大幅に増加する。そのため、候補数が数億以上に及ぶ可能性のある最適音素ラベル列の決定を音素境界位置の推定と同時に行うことは非常に効率が悪く、現実的ではない。実際に、音声データが2~3音節程度の極短い単語に対しては、同時推定を行うことで計算時間が若干短くなる場合もあるが、通常の音声サンプルに対しては本システムのように2段階で処理を行う方がトータルの計算時間は大幅に短縮されることが確認されている。

【処理5】結果の表示

自動推定処理で得られた結果を、マウスの操作によって簡単に修正可能な可動部品の形で視覚的に表示する。

2-2 エディタ部の構成要素と特徴

実際の音声ラベリング作業を進めるに当たり、心身の疲労を最大限に抑えながら単調になりがちな作業を長時間に渡って継続させるためには、自動化部分の推定精度を高めて修正が必要な箇所を減らすことももちろん大切ではあるが、最も重要なのは、作業者が直接操作するGUIの部分を如何に使い易く設計するかという点であろう。

ここでは、本システムの最重要部分であり、その操作性に最も大きな影響を与えるエディタ部の構成要素とそれぞれの特徴について述べる。

2-2-1 読み仮名表示窓および音素ラベル列表示窓

これらは共に画面の左上部に位置し、読み仮名表示窓内には発話内容を表す読み仮名が、また音素ラベル列表示窓には仮名文字列に対応する音素ラベル列が、それぞれ表示されている。

特に、音素ラベル列表示窓は、自動推定された音素ラベル列に対して部分的な変更を行うための機能を有している。この窓に表示されている内容は、数億通り以上に及ぶ可能性のある膨大なラベル列候補の中から正しいものを効率良く指定することができるよう、自動推定によって得られた最適候

補だけでなく、複数通りの可能性がある部分ラベル列(Fig.4で文字色が反転表示されている部分)をハイパーテキスト形式で表現したものとなっている。

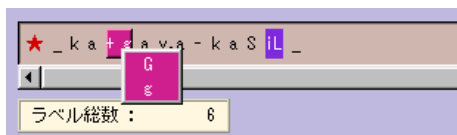


Fig.4 Expression of a phoneme label sequence by hyper-text expression.

これによって、修正したい音素ラベルをマウスの左ボタンでクリックして別候補を表示し、正しいものを選択するという簡単な操作で音素ラベルの変更を行うことができる。

また、後述の境界位置編集用窓内で音素境界カーソルを移動している間は、そのカーソルに対応する音素ラベルの部分を点滅させることで、編集中の境界が音素ラベル列のどの部分のものであるかを簡単に知ることができるようになっている。

更に、文発話のように一つのサンプルの発話内容が長くなると、音素ラベル列が音素ラベル列表示窓内に納まり切らなくなることもあるが、その場合でも窓下部のスクロールバーの操作によって任意の区間を表示させることができ、音素境界カーソル移動時にはそのカーソルに対応する音素ラベルが窓内に点滅表示されるように自動的に音素ラベル列をスクロールする機能も有している。

2-2-2 コントロールパネル

ラベリング作業に関連する様々な機能の実行やシステムの動作条件の設定、各種モードの切替えなどは、画面の右上部を中心に表示されているボタン類、および画面最上部のボタンから辿ることのできる階層化メニューによって実行する。

使用頻度が高いボタンほど選択し易い位置に配置し、利用頻度の低いもの、あるいは不適切な設定を行った場合に操作性の劣化を招く恐れがあるものについては階層化メニューに納めるという配慮を行っている。

2-2-3 データ表示窓

音声ラベリング作業は、主として音声の聴取、およびデータの視察に基づいて進められるため、データの表示に関連する諸機能がシステム全体の操作性に与える影響は極めて

大きい。

本システムでは、上方の全体波形表示用窓と下方の境界位置編集用窓という2種類のデータ表示窓を用意してそれぞれに適切な機能を持たせ、ほとんど全ての機能を簡単なマウス操作のみで実行可能にする(一部の特殊な機能はキーボードとの併用によって実現している)ことで、使い易いGUIを実現している。

[a] 全体波形表示用窓

全体波形表示用窓は常に作業対象データ全体の波形が表示されており、その上に音素境界カーソルや、現在の編集対象エリアが表示されている。ここに表示されている情報は閲覧のみ可能で、直接編集することはできない。

全体波形表示用窓が有する主要機能は、以下の通りである。

- 窓内の任意の範囲を左ボタンで選択
⇒ 編集対象エリアを設定
- 編集対象エリアの右端/左端を左ボタンでドラッグ
⇒ 編集対象エリアを微調整
- 窓内の任意の位置で左ボタンをダブルクリック
⇒ 編集対象エリアをリセット
- 編集対象エリア内で右ボタンをクリック
⇒ 編集対象エリア内の音声データを再生

[b] 境界位置編集用窓

一方、境界位置編集用窓は、実際の音声ラベリング作業を進めるためのエリアであり、編集対象エリアのデータのみが時間方向に拡大されて表示されている。この窓の音素境界カーソルはマウス操作でその位置を移動することができ、マニュアル操作が許可されている場合(マニュアル操作モードについては後述)には、音素境界の追加や削除、音素ラベルの直接入力などを行うこともできる。これらの編集内容は、全体波形表示用窓内の情報や音素ラベル列表示窓内のハイパーテキストと完全に連動しており、音素ラベル列表示窓や境界位置編集用窓で実行された変更箇所は、他の窓の表示内容にも直ちに反映される。

更に、境界位置編集用窓では、(1)波形、(2)周波数スペクトル、(3)波形と周波数スペクトルの重ね合わせ(Fig.5参照)と

いう、3種類の情報を瞬時に切替えて表示することができる。この機能を利用し、音素の性質に応じてその表示内容を適切に使い分けることで、高精度なラベリング作業が可能になる。

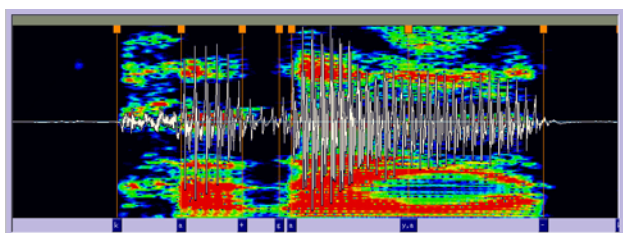


Fig.5 Simultaneous indication by overlapping a waveform on a spectrum.

境界位置編集用窓が有する主要機能は、以下の通りである。

- ・ 窓内の任意の範囲を左ボタンで選択
⇒ 編集対象エリアを設定
- ・ 窓内の任意の位置で左ボタンをダブルクリック
⇒ 現在の位置を中心に編集対象エリアを2倍に拡大
- ・ 窓内の任意の位置で右ボタンをクリック
⇒ その位置の直前、および直後にある音素境界で挟まれた音素区間の音声データを再生
- ・ 音素境界ハンドルを左ボタンでドラッグ
⇒ 音素境界を移動
- ・ 窓内の任意箇所を中ボタンでクリック
⇒ スペクトル表示時に、その上に重ねて波形を表示するか否かのトグル(波形表示時には無視、2ボタンマウス使用時にはキーボードのシフトキー+左ボタンで代用)
- ・ 音素ラベルを中ボタンでクリック
⇒ 発話誤りや著しいノイズの混入など、データとして使用できない区間に対応する音素ラベルに無効フラグを付与するか否かのトグル(無効フラグが付与された音素ラベルは赤色で表示される)

更に、マニュアル操作が許可されている場合には、以下の操作が可能になる。

- ・ 境界位置編集用窓上部の境界追加/削除エリアを左ボタンでドラッグして境界位置編集用窓内へドロップ

- ⇒ 新しい音素境界カーソルを挿入
- ・ 音素境界ハンドルを左ボタンでドラッグして境界位置編集用窓上部の境界追加/削除エリアへドロップ
⇒ 既存の音素境界カーソルを削除
- ・ 音素ラベルを左ボタンでクリック
⇒ 音素ラベルを変更(音素ラベルはキーボードより入力する)

2-3 半自動音声ラベリングシステムの付加機能

ここでは、通常の作業での利用頻度は低いが、目的に応じて適切に活用することで大きな効果が期待できるいくつかの有用な付加機能について紹介する。

2-3-1 音声ラベリング作業訓練機能

音声ラベリング作業を円滑に行うためには、本システムを用いたとしても、ある程度の作業訓練を要する。そこで、未熟練者が独力でノウハウを習得できるよう、本システムには作業訓練機能が搭載されている。この機能によって、訓練用サンプルに対する未熟練作業者のラベリング結果と、同サンプルに対して予め熟練者がラベリングした結果(これを「正解」とみなす)との差異が、その差異の程度に応じた表示色のマークで表示される。この様子をFig.6に示す。

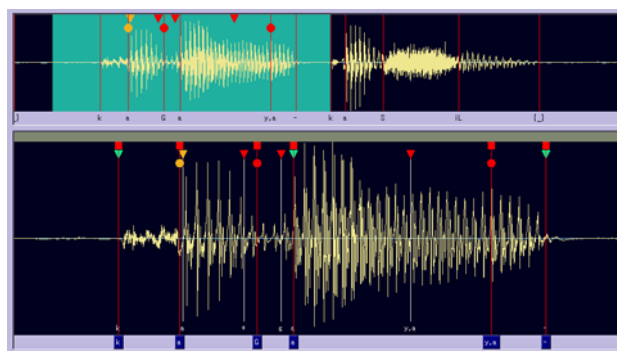


Fig.6 A screen view in job-training mode.

この画面で、波形上部に表示されている逆三角印(正解マーク)は「正解」の位置を表している。また、丸印(誤回答マーク)は、作業者のラベリング結果の中で「正解」と音素ラベル自体が異なるもの、あるいは「正解」との間の境界位置のずれ(誤差)が一定の許容範囲を越えているものを表している。作業者は、これらのマークの色や位置を参考にしながら結果の見直しを繰り返すことにより、短時間で作業のノウ

ハウを習得することができる。

2-3-2 チェック機能付きマニュアル操作モード

通常モードでは、音素ラベル列の変更や音素境界位置の修正など、基本的な作業は全てマウス操作のみで行われ、音素ラベルの定義から外れてしまう可能性のある操作はできないように工夫されている。しかし、熟練者にとっては、このようなインタフェースは必ずしも作業効率が高いとは言えない。

そこで、本システムには、キーボード入力により音素ラベルを任意の記号に変更したり、音素境界カーソルを直接挿入、あるいは削除するような操作を許可するマニュアル操作モードが用意されており、オプションメニューの中で切替えることができるようになっている。

ただし、マニュアル操作モードでの作業中であっても、システム内部では、常に作業結果と音素ラベル定義との整合性のチェックを行っており、不適切な作業結果に対しては、速やかにその旨が通知される。なお、この通知は、音素ラベル表示窓の背景色の変更によってのみ行われるため、円滑な作業の妨げになることはない。

2-3-3 完全マニュアル操作モード

本システムには、音声データベースとは全く関係のない任意のデータに対して波形やスペクトルを観測したり、なんらかのマーキングを行うような作業にも使用できるよう、自動化のための機能を全て停止し、単純なエディタとして動作させることのできる完全マニュアル操作モードも用意されている。

完全マニュアル操作モードでは、音声データだけでなく、各種のノイズや楽音など、あらゆるサウンドデータを扱うことができるため、「音」に関連する研究開発業務において、様々な目的で汎用的に利用することができる。

2-3-4 音素ラベル情報と音声データの一元管理

一連の作業の結果として得られる音素ラベル列や音素境界位置の情報の管理方法についても工夫を行っている。

一般の音声データベースでは、音声データとそのラベル情報が別々のファイルとして保存されている場合が多い[8, 9, 10]。しかし、このような管理方法では、音声データファ

イルとそれに対応するラベル情報のファイルとの間の対応関係を利用者が常に意識している必要があり、ファイルの移動やコピー、ファイル名の変更などを行った場合に、その対応関係が見失われてしまう恐れがある。

そこで、我々が構築を進めている音声データベースでは、音声データのファイルフォーマットとして広く利用されているWAVEフォーマットが有する、音声データ以外の様々な付加情報を同一ファイル内に記述するための仕組み(RIFFチャック構造)を利用し、ラベル情報を音声データ中に埋め込むことで、音声データとラベル情報の一元管理を実現している。

なお、このような情報の埋め込みは、全てWAVEファイルフォーマット定義の範囲内で実現しているため、一般のアプリケーションからは通常のWAVEファイルと全く同様に扱うことができる。また、特にWindows系のOSでは、標準機能として用意されているファイルのプロパティ表示画面で、埋め込まれた情報を手軽に閲覧することも可能である(Fig.7参照)。



Fig.7 Indication of embedded information by Windows's property window.

3. 成果

本システムの開発により、アルバイトや派遣社員等、音声に関する知識や経験の無い作業員でも、極短期間の訓練で本格的な音声ラベリングを実施することが可能になった。これによって、現在までに単語発話音声や文発話音声などのバリエーションに富んだ数百話者分の音声データベースの構築が進められている。

この成果は、音声認識や音声合成に関連する技術の研究開発に活用され、不特定話者対応音響モデルの性能向上(例えば、本データベースの利用によって、従来の不特定話者音声認識の誤認識率を半減させることに成功している)や音声合成用音声素片の獲得などを通して、オフィス機器のアプリケーション性向上を目指した音声利用技術の応用展開に大きく貢献している。

4. 今後の展開

本システムで使用している音響モデルは、ラベリング作業結果を用いて適宜更新しているため、作業の進行に伴って音響モデルの性能も向上し、その結果、自動推定精度が高まって作業速度が一段と加速するという効果的なサイクルが成立する。しかも、この音響モデルは、現在我々が開発を進めている種々の音声認識方式のためのモデルと本質的には全く同じもの(本システム用のものは、構築された音声データベースが音声合成技術の研究開発でも利用できるよう、音素ラベルの定義が音声認識用モデルに比べて、より詳細なカテゴリに分類されたものとなっている)であるため、本システムの自動ラベリング精度と音声認識性能の間には高い相関関係が見られ、本システムの性能向上が、そのまま音声認識方式の性能向上につながる事が確認されている。

今後、本システムの性能向上を行いながら、話者、発話様式、音素環境などを幅広く網羅するように音声データベースの更なる拡充を図り、その成果を日常の研究開発業務の中で大いに活用していきたい。

謝 辞

長期間に渡る音声ラベリング作業を通して、音声データベースの充実、および本システムの操作性改善や信頼性向上に多大なる貢献をして頂いた岩淵佐和子さん、木村久美枝さん、黛裕子さん、野口美香さんに感謝の意を表します。

参考文献

- [1] 鷹見 他:“半自動音声ラベリングシステムによる音声データベース構築作業の効率化”, 音響学会講演論文集, 3-P-4, (2001.3) pp. 171-172.

- [2] <http://tcl.activestate.com/software/tcltk> 等.
- [3] <http://club.pep.ne.jp/~am0250/dustproof.html>.
- [4] 鷹見:“A*探索法を用いた大語彙音声認識における音素継続時間制御と高速化の検討”, 音響学会講演論文集, 2-Q-19, (1998.3) pp. 157-158.
- [5] 鷹見:“A*探索法による大語彙音声認識”, RICOH Technical Report, No. 25, (1999) pp. 61-70.
- [6] 鷹見, 嵯峨山:“逐次状態分割法による隠れマルコフ網の自動生成”, 電子情報通信学会論文誌, J76-D-II, No. 10, (1993) pp. 2155-2164.
- [7] 中川:“確率モデルによる音声認識”, 電子情報通信学会, コロナ社(1988).
- [8] 匂坂, 浦谷:“ATR音声・言語データベース”, 音響学会論文誌, Vol. 48, No. 12 (1992).
- [9] 小林 他:“日本音響学会連続音声データベース”, 音響学会論文誌, Vol. 48, No. 12 (1992).
- [10] 安藤, 宮坂:“ニュース音声データベースの構築”, 音響学会講演論文集, 2-Q-9, pp. 157-158 (1997.9).

注1) Windowsは、米国Microsoft Corporationの米国及びその他の国における登録商標です。