

## 教授力向上のためのスピーチ訓練システム

新井 浩\*1, 杉森 公一\*2, 瀬戸 就一\*1, 川辺 弘之\*3, 下村 有子\*3

\*1 金城大学 短期大学部

\*2 金城大学 医療健康学部

\*3 金城大学 社会福祉学部

arahiro@kinjo.ac.jp

**概要**：スピーチ訓練を受けていない教員は自分自身の話がどのように受け取られているのか、よく分からない。聞き取りにくい話しぶりだとせっかくの講義内容も学生によく伝わらない。人前で話すためには訓練が必要であるが、自己学習で訓練できるシステムがあると、時間のない大学人でも空き時間に訓練ができる。我々はスピーチの問題点を指摘するシステムを構築する。このシステムでは自分で自分のスピーチを聞かずにスピーチを改善することができる。手本データとユーザーデータの比較には音声特性パラメータ（明瞭さ・音量の大小・音声の高低・間合い）を用い、レーダーチャートまたは顔型グラフによる直感的な結果表示を試みる。

### 1 はじめに

近年、大学教育の「質」に対する社会的要請が高まってきている。文部科学省は『学士力』を定義し、大学の組織的なファカルティ・ディベロップメント(FD)活動、教員の自己研鑽を求めている[1]。この背景には、大学全入時代を迎えての大学入学者の質的变化のみならず、大学教員の教授能力に疑問符がつけられていることがある。大学教員は、研究の専門性に特化した知識・知力がある一方、教育者としての訓練を受けていない。また、年齢層が高く、他者から評価されることを嫌う傾向があることも原因として挙げられる。

我々はこれまで、聴覚障害学生支援を目的に教員の講義の内容や講義の雰囲気伝えるため、講義内容を聴覚障害者に正確に伝える技術について調査・研究を行っている[2]。また、授業における非言語的な情報である臨場感や感情という教室のドラマ性が、学生の講義理解を豊かなものとするのに欠かせないものであることを主張し、聴覚障害者のための「感情フロント」の提唱と構築を行っている[3]。

上記の研究において、音声認識技術を用い音声から言語情報と非言語情報を抽出することを試みている。これら 2 つの情報は、話し方の音声特性データとみなすことができる。これらの評価は教員自身による授業評価・分析の深化を促す一方、蓄積された音声特性データは教員教育研究のための基礎データとして高い価値がある。

これまで、聴覚障害者のためのスピーチ訓練シ

ステムの研究は多く行われてきており、低学年児童用のろう学校の補助システムである「スピーチトレーナー」[4]、発話訓練システム「トーキングトレーナー」[5]、画像化音声を利用する音声分析・表示ツール「Speech-ART」[6,7]など、音響特性量を利用した訓練システムが提案されている。

本研究では、大学教員の教授力向上のために教員の話し方、特に言語的・非言語的コミュニケーション能力を成長させることを目標とした「音声認識技術を活用した自習可能な口話評価システム」を開発することである。教員の多くはスピーチ訓練が十分でないために、自分自身の話がどのように学生に受け取られているのか、十分な理解と評価手段を持っていない。人前で話すためには訓練が必要であるが、自己学習で訓練できるシステムがあると、時間のない大学人であっても空き時間に訓練することが可能である。我々はスピーチの問題点を指摘するシステム構築により、自分で自分のスピーチを聞かずにスピーチを改善することを目指す。

### 2 システムの概要

我々のプロジェクトは次の3つのサブシステムから構成される（図1）。

A) スピーチ評価システムの構築

B) GE(Good Example:手本) 音声データベース  
(声量・スピード・明瞭さなど)の構築と時系列キーワード・音声特性分析システム

C) 顔型グラフによるユーザーインターフェースの構築

マイクより入力されたユーザー音声は、文字認識と付加情報抽出が行われる。文字認識においては認識された文字と認識できた割合でスピーチの評価を行う。一方、付加情報抽出では、声の高さ、声の大きさ、声の速さ、間合いを抽出する。その結果から、話者のスピーチの欠点をレーダーチャートグラフ（図 2）や顔型グラフ（図 3）で表示し、その改善を促す。

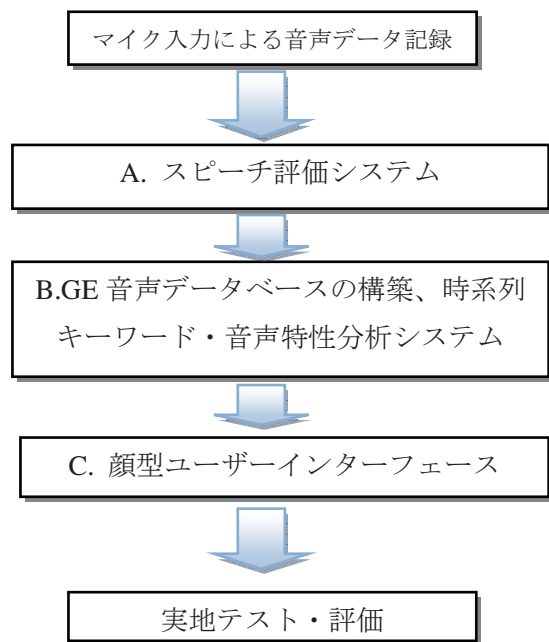


図 1 システムの構成

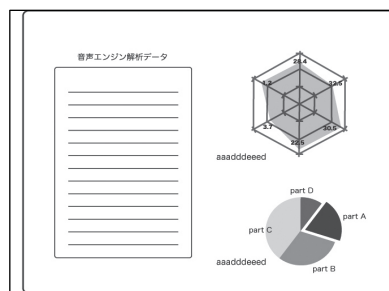


図 2 グラフ出力の例

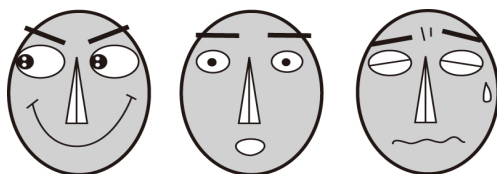


図 3 チャーノフの顔型グラフ

### 3 スピーチ評価システム

スピーチ評価のために開発中のシステム概要を図 4 に示す。UV(ユーザー音声: UV)データとアナウンサー・ナレーターなどの専門家により録音された GE 音声データを比較する。GE データを聴きながら、ユーザーは同時に音声を録音する。その差の評価は、認識結果・音声波形・スコアとして結果表示される。

システムは、以下のように処理される。

- (1) 入力された音声は同時に 2 つの異なる処理が行われる。
- (2) A 処理では、入力音声は音声認識エンジンにかけられ、音声認識文字を取り出す。音声認識された文字が正確であるか、サンプルデータと照らし合わせながら、正解文字データと認識文字データを並列で表示し、最後に認識率を表示する（図 4）。
- (3) B 処理では、入力音声は声の大きさ、話の速さ、間合いの長さなどをサンプルデータと比較しながら表示する（図 5）。
- (4) これらのデータを音声特性データベースの値と照らし合わせる。
- (5) A 処理と B 処理の結果を簡単に理解してもらうために、レーダーチャートで訓練がどれくらい上達したのかを表示する。今回はチャーノフの顔型グラフは用いずに、レーダーチャートで表示する（図 6）。

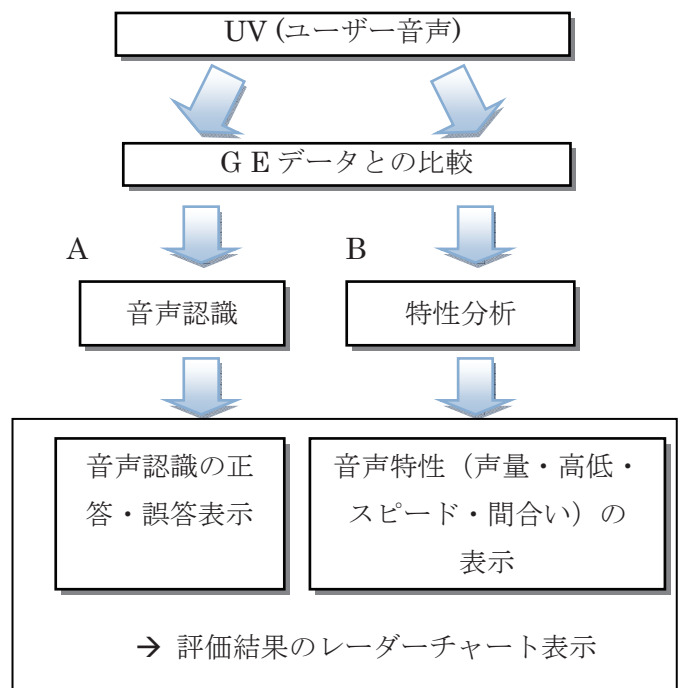


図 4 スピーチ評価システム

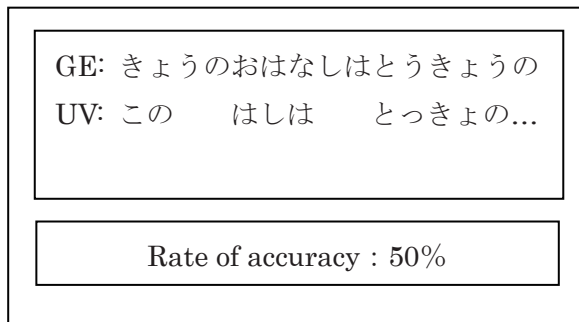


図5 処理Aの例

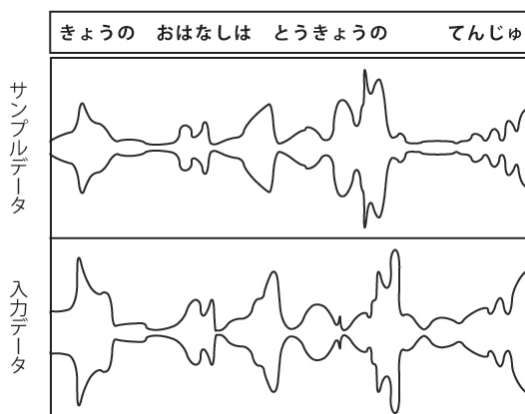


図6 処理Bの例

これらの表示インターフェースと音声特性データ処理(B処理)には、Python 2.5の標準ライブラリとmatplotlibモジュールを利用し、音声認識エンジンJulius [8]による音声認識文字の出力結果(A処理)を同時出力する。また、レーダーチャートは単位時間ごとに別画面で表示する。最終的にはリアルタイム処理を目指す。今回の試作システムでは、60秒単位で音声ファイルを切り分けて使用した。

#### 4 GE 音声データベースの構築と音声特性分析システムの構築

音声認識により各自のスピーチの認識率が評価された後に実行されるA処理とB処理について詳細を述べる。GE音声は「GE音声データベース」に蓄積され、その比較分析のために「音声特性分析システム」が実行される。

##### 4.1 GE 音声データベース

GE音声データベースはアナウンサーなどの専門家のスピーチ音声データに基づいて構成される。男女のアナウンサーが読み上げた文章から、5つの

パラメータが抽出される。

- 1) 「明瞭さ・滑舌さ」を表す音声認識率(%)
- 2) 1分毎の文字認識数から求められる「スピード」(word/min)
- 3) 「声の大小」(dB)
- 4) 「声の高低」(Hz)
- 5) 音声波形の間から抽出される間隔からの「間合い」(sec)

である。

#### 4.2 UV データベース

UV(ユーザー音声)は、1文毎に録音されデータベースに蓄積される。パラメータ抽出はGE音声データベースと同様に行われる。図7はGEデータとUVデータを同期させ表示させたもので、評価結果はレーダーチャートとして表している。

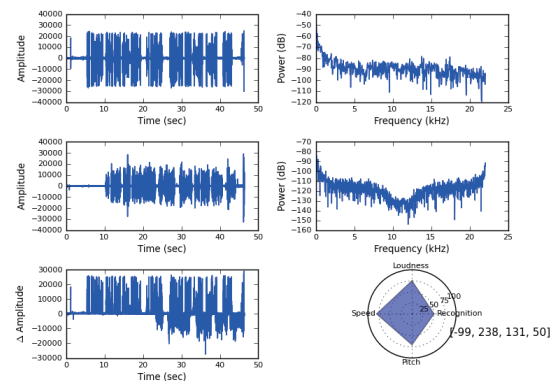


図7 スピーチ評価システムによる結果の例(一行目はGE波形データ、二行目はUV波形データにそれぞれ対応する。三行目は波形の差分とレーダーチャート結果表示である。)

#### 5 顔型ユーザーインターフェースの構築

1文ごとに評価された結果は、将来的にはチャーノフの顔型グラフで表現する予定である。評価するたびに、顔の変化がパラパラマンガのように変化する。顔型グラフの眉、目、口、顔面などのパーツは、スピーチの声の大きさ、高低、話の速さ、認識

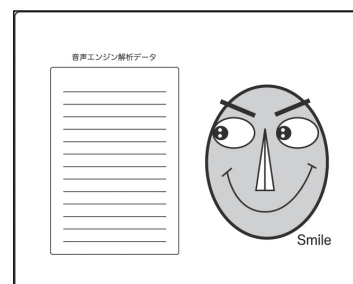


図8 顔型グラフ評価の例

率に対応して変化させる。また、全体評価は音声終了後に再び顔型グラフで表され、訓練者は直感的に結果を理解でき、改善に結び付けることが容易になる（図8）。

## 6 おわりに

教員のためのスピーチ訓練システムを提案し、現在までの開発状況について示した。今後、音声特性データベースの構築、顔型ユーザーインターフェースの構築と開発を進める予定である。このシステムは、教員の自己訓練の用途に限らず、企業内外でのプレゼンテーション・講演会用のスピーチ訓練などへの応用も期待できる。

## 謝辞

本研究の一部は、平成23年度文部科学省科学研究費補助金（挑戦的萌芽研究：23650556）の支援を受けた。

## 参考文献

- [1] 文部科学省中央教育審議会、平成19年度答申（2008年12月24日）、2008
- [2] Kawabe, H., Sugimori, K., Shimomura, Y. and Seto, S., Computer Simulation of Text Reproduction Based upon “Quantity-Rather-Than-Quality” Concept, Proceeding of the 40th International Conference on Computers and Industrial Engineering, 2010
- [3] Seto, S., Arai, H., Sugimori, S., Shimomura, Y. and Kawabe, H., Subtitle system visualizing non-verbal expressions in voice for hearing impaired - Ambient Font -, Proceeding of the 10th Asia-Pacific Industrial Engineering and Management Systems, 2010
- [4] 菅野文武、鎌田弘之、石田義久、「音声認識技術を利用した単音節・単語の発生練習とスピーチトレーナーへの応用」、信学技報、Vol.ET93-122(1994-01)、59-66、1994
- [5] 梅崎太造、倉谷和彦、藤吉弘宣、「ネットワーク環境を利用した聴覚障害児用発生訓練支援システム」、電子情報通信学会論文誌 D-II、Vol.J80-D-II、No.4、pp.925-932、1997
- [6] 上田祐市、塔本麻友美、青木敏裕、坂田聡、リアルタイム音声特徴ベクトルを用いた発生・発語学習ツールの開発、信学技報、Vol.ET2008-78(2009-1)、25-30、2009

- [7] 上田祐市、坂田聡、池田博、渡邊亮、「画像化音声を利用する音声分析・表示ツール（Speech-ART）の開発」、映像情報メディア学会誌、Vol.61、No.5、692-698、2007
- [8] Lee, A., Kawahata, T. and Shikano, K., Julius—an open source real-time large vocabulary recognition engine, Proc. European Conf. on Speech Communication and Technology (EUROSPEECH), 1691-1694, 2001