

# 話し方に対するリアルタイムアドバイスシステムの提案

新井浩\*, 南保英孝\*\*, 瀬戸就一\*, 下村有子\*\*\*, 川辺弘之\*\*\*

\*金城大学短期大学部, \*\*金沢大学, \*\*\*金城大学

arahiro@kinjo.ac.jp

**概要：** 私達は音声から5つのパラメータを抽出し、お手本となる音声と比較することで、話し方を自習するシステムを研究している。音声データから声の大小・高低・速遅・間合い・明瞭さ等のパラメータを抽出し、お手本となるスピーチデータと比較している。話者はマイクに向かって話すだけで、システムが顔グラフで評価を与える。

私達はこれまで、このシステムの試作を作成し、ノートパソコン上で動作させる事ができた。このシステムを使う事で話者の話し方が改善する結果を得ることができた。今回はこの話し方自習システムを小型化し持ち歩けるように改良を行う。評価の投影装置には透過型のヘッドマウントディスプレイ (HMD) を活用する。

## 1 はじめに

日本では 18 歳人口の減少にともない、2005 年に大学進学率は 50%に達し、高等教育機関への進学率については 2014 年に 80%を超えた。[1]。その結果、以前に比べ多様な資質をもった学生が入学するようになり、学力や理解度の差が開きつつある。講義においてより丁寧な説明、分かりやすい講義が大学に求められている。これまで大学教員は研究を行う事が仕事の中心であった。研究分野については詳しいが、教え方について専門的な訓練を受けておらず、話し方があまり上手ではない人も多い。

そこで我々は、音声から得られる音量や高低といった音声特性を解析してスピーチ評価システムを開発した[2], [3]。

マイクより入力した音声は、声の大小・高低・速遅・間合い・明瞭さ等のデータパラメータとして抽出される。このシステムでは音声から抽出された5つのデータをお手本となるスピーチデータと比較し評価するため、客観的な評価を得ることができる。また自分で録画した授業データを全部みる必要がないため、使用者の心的負担も軽減される。

今回はこの音声評価システムを使用する教室などへ持ち出し、人前で話す時に使用出来るよう改良を行う。実際の講義で装着したまま話し、リアルタイムに話者の話し方を分析し、アドバイスを送るシステムをめざす。持ち歩く事ができるシステムでは、小型化する事が必要な為、

当初は音声で指示することも考えていた。しかし近年は google glass などに代表される HMD が高機能、低価格化してきた。今後の展開も見据え、今回は音声評価システムを HMD へ出力できるように研究し、システムの小型化を行う。

HMD はノート PC に比べ視界が極端に狭く、一度に理解できる情報量も減少する。瞬時に理解できるように、漫符表現の使用も検討する。

## 2 スピーチ評価システム

音声データから音声特性を解析し評価する行程を説明する。

### 2-1 スピーチ評価システムの概要

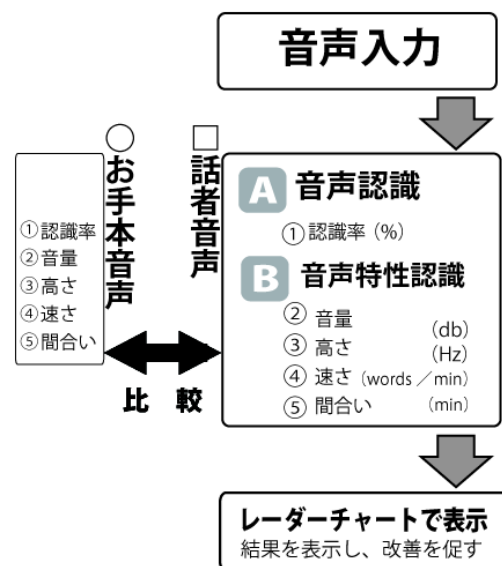


図1 スピーチ評価システムの概要

マイクより入力されたユーザー音声から、音声認識と付加情報が抽出され、音声特性分析が行われる。音声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（図1）。

音声認識にはオープンソースの音声認識エンジンである Julius を使用した。マイクから入力された音声信号は、システムと Julius に同時に入力される。Julius は音声信号を解析して認識されたテキストと、認識の際に計算された音響尤度をシステムに転送する。音響尤度は入力音声認識テキストになる確率を表しており、活舌がよいと認識テキストが正確に予測されやすくなると考える。よって、音響尤度は活舌と相関があると考え、この値を活舌の指標として扱う。一方、音声特性分析では、声の大小、声の高低、間合い、話す速さの 4 要素を抽出する。以上、5 つのデータをお手本音声データと比較し良否判定を行う。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や顔型グラフを組み合わせて作成する顔形ユーザーインターフェースで表示し、改善を促す。

## 2-2 お手本データの選出

システムでは絶対的な一つのお手本データと話者データを比較・評価するのではなく、複数あるお手本音声データの中から、話者の音声特性と近いデータを選び出し、評価に用いる。

日本語コーパスの中から多くの人が主観的に良いと評価する音声を選別し、これをお手本データとする。まず我々は良い話し方という評価基準を言語化したループブリック指標を作成した。指標には音量、音の高低、間合い、早さ、滑舌について明記し、すべて含んだ総合評価によって点数化する [4]。現在は 100 音声から男女それぞれ 5 音声のお手本データを選出したものを用いている。

## 3 漫符を用いた顔形グラフへの展開

話者の音声特性はシステムで評価され、顔型グラフを用いたユーザーインターフェースを通じて表示される。音声特性と対応した顔のキャラクターが、話者のスピーチにおける欠点を示し話し方の改善を促す。

顔型グラフは、鼻・口・眉など顔のパーツ形状に 16 変量までのデータを対応させ人間の顔として示すグラフである。顔形状が評価と連動していることから、評価の直感的な理解が可能である。評価と顔形グラフの表情が一致するように調整することで、あたかも聴講する学生が眼前にいるような感覚で表示させる事ができ、話者は直感的に評価結果を感じる事ができる [5], [6]。

音声特性と良否判定の 2 つの評価を同時に行えるよう本システム用のルール作りを考案した。音声特性は顔の輪郭形状や、目・鼻の配置間隔など顔の造形に反映される [7]。またお手本音声との類似性で比較する良否判定については、眉や黒目の形状など表情として反映する（図 2）。

| 音声特性 | 変化する部位      |
|------|-------------|
| 小・大  | 顔全体のサイズ     |
| 低・高  | 顔パーツの位置（上下） |
| 遅・速  | 輪 郭         |
| 間合い  | 目の間隔        |
| 滑 舌  | 開口部のサイズ     |

図 2 顔形グラフに反映される音声特性と変形する顔の部位

顔型グラフの視認性についてプレアンケートをとった所、面白く良否判定は分かりやすいがスピーチの改善点は分かりにくい、という意見が多くみられた。そこで漫符の追加とサブグラフを併置することで理解を促した。これを顔型ユーザーインターフェースとする（図 3）。

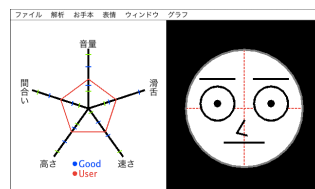


図 3 顔形ユーザーインターフェース

## 4 リアルタイムアドバイスシステムの構築

スピーチ評価システムを小型化し持ち歩けるように改良を行う。評価の投影装置には透過型のヘッドマウントディスプレイ (HMD) を活用する（図 4）。

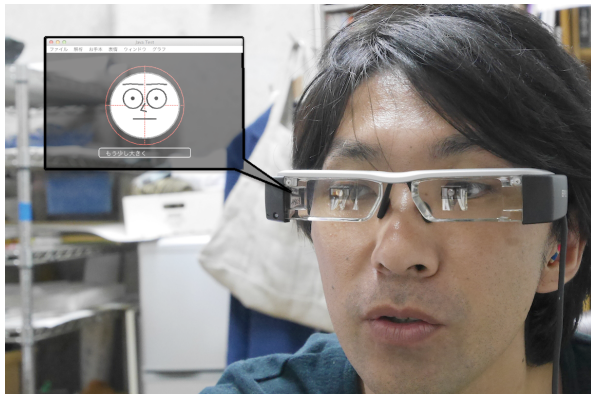


図4 リアルタイムアドバイスシステムのイメージ

#### 4-1 システムの HMD 展開と改善ポイント

音声認識を用いたこのスピーチ自習システムはノートパソコン上で動作する事ができ、一定の改善を促すことができた。次の展開として、このシステムを HMD（へ出力させシステムの小型改良化を行う。

頭部に装着する小型のディスプレイであるヘッドマウントディスプレイ(HMD)は一般的なディスプレイと異なり眼球にとっても近い位置で映像を投影できるため、視聴者の動きを妨げない。また小型であるため消費電力にもすぐれている。Bluetooth ヘッドセットで入力された音声はワイヤレスにノートパソコンへ送られる。システムで評価された結果は再び Bluetooth を経由して HMD へ投影される。

今回使用する透過型 HMD は小型であるために表示面積が小さく表示できる情報が限られる。限られた情報で理解できるよう、シンプルな画面設計と改善指示のルールを作成する。

#### 4-2 インターフェースの決定

透過型 HMD では投影面の背後にくる風景によって視認性が変わる。特に明るい場所では投影された白系の画像は見難くなるが、投影画像の背景をグレー、もしくは中間調にすることで視認性があがる。

プロトタイプで作成した顔型ユーザーインターフェースを利用し 4 種類の表示パターンを作

成した（図 4）。本システムにおいては音声特性の認識よりも、現在の話し方に対するアドバイスが中心であるため、グラフ表示は省略しても構わない。これらを HMD に投影しその見え方について検証したところ、図 4 の①または④が見えやすいという結果になった。

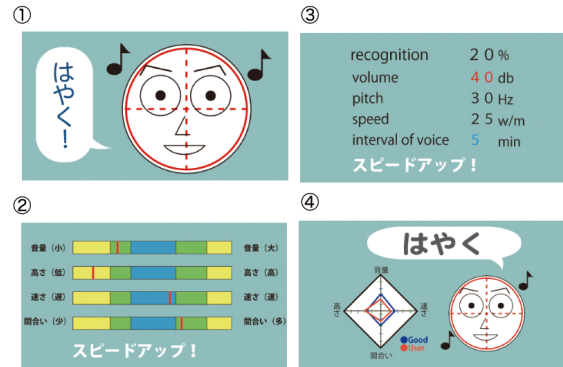


図4 アドバイスト組み合わせる顔グラフ 4 種類の表示パターン

①顔形グラフと漫符の組み合わせ②数値のみ③棒グラフ④顔グラフとレーダーチャート

#### 4-3 アドバイス表示のルール

話者の音声データはお手本データと比較され、一定誤差の範囲内にあれば加点、範囲外であれば程度に応じて減点される。点数は音量が 2 点、高さは 1 点というように、パラメータによって異なり、重要度が設定されている。話者の音声データが適正音域から最も遠い -a ゾーンに入ったパラメータについてのアドバイスが吹き出しに表示される。2 種類以上が表示される場合は、優先順に吹き出しに表示され、全く同点の場合は -a ゾーンの滞在時間が長いパラメータが表示される（図 5）。

**a = 音量 (2 点) ・ 高さ (1 点) 速さ (1.5 点) ・ 間合い (1 点)**

○お手本音声データ



□実験者音声データ



小 ちょうどよい 大

図5 HMD に投影するための採点ルール

## 5 まとめ

音声から得られる音声特性を解析してスピーチを評価する自習システムを、小型化し持ち歩けるよう改良を行った。プロトタイプシステムでは一人で学習を行うためのものであった。今回、小型化することで実際の講義で装着したまま話すことが可能な、リアルタイムに話者の話し方をアドバイスできるシステムとなった。

表示にHMDを使用するため、視認性を考慮したインターフェースを作成した。またお手本データとの比較で複数のアドバイスを送らなければならない場合には、優先順位を設け表示することとした。

今後は顔形グラフが表示する評価がより適切になるよう、スピーチ評価適正音域と非適正音域の境界の適正値をさぐっていく。

## 参考文献

- [1] 文部科学省平成 26 年度学校基本調査
- [2] Arai, H., Nambo, Seto, S., H., Shimomura, Y., and Kawabe, H. (2013) Consideration of the effectiveness of a face-graph user interface - a self-training system for professors., Proceeding of the 14th Asia-Pacific Industrial Engineering and Management Systems, 2013
- [3] 新井浩, 南保英孝, 瀬戸就一, 川辺弘之, 下村有子, 「教員のためのスピーチ評価システム」, AXIES 大学 ICT 推進協議会 2013 年度年次大会(2013)
- [4] 前川喜久雄 :『日本語話し言葉コーパス』の概要; 日本語 科学, 15, pp.111-113 (2004)
- [5] 田中豊, 脇本和昌, 「多変量統計解析法」現代数学社 (1983)
- [6] 脇本和昌, 「チャーノフの顔形グラフの変形のこころみ-体型グラフ-」, 行動計量学, 4 巻 2 号 (1977)
- [7] P・エクマン, W・V・フリーセン, 工藤力訳 編, 「表情分析入門 表情に隠された意味をさぐる」, 誠信書房 (1987)