

機関番号：13501
 研究種目：若手研究（B）
 研究期間：2009～2010
 課題番号：21700807
 研究課題名（和文） 良い講義を行うための音声処理技術を利用した話し方トレーニング支援システム
 研究課題名（英文） Development of training system for improving the manner of speaking on lecture speech using speech processing technologies
 研究代表者
 西崎 博光（NISHIZAKI HIROMITSU）
 山梨大学・大学院医学工学総合研究部・助教
 研究者番号：40362082

研究成果の概要（和文）：

本研究の目的は、講演等での話し方が聴き手に及ぼす影響を調査し、聴きやすい話し方の特徴を明らかにすることにある。研究の主な成果として、話し言葉に含まれるフィラー（“あー”、“えーっと”等の間投詞）が、聴き手が感じる話しの上手さに一役買っていることが明らかとなった。この特徴を用いて話し方の評価・アドバイスシステムを構築し、システムの評価を行ったところ、本システムは話し方を改善する上で有効であることが分かった。

研究成果の概要（英文）：

The main purpose of this research is to investigate that how the manner of speaking on lecture speech affects the impression of listeners and to find out the characteristics of good speech. My study cleared that “filled pauses” including spontaneous speeches were the main factor for improving the manner of speaking. And, we developed an evaluation and advice system for training the manner of speaking. The system I developed worked well and gave the good advices for the speakers. The speakers followed by the advices improved their manner of speaking.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	2,500,000	750,000	3,250,000
2010年度	900,000	270,000	1,170,000
年度			
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学

キーワード：講義・講演音声、音声認識、音響特徴量、フィラー、講義評価、教師教育、話し方評価、印象評価

1. 研究開始当初の背景

大学において、講義の質改善が社会的な要望でもあり、喫緊の課題である。このような社会背景もありほとんどの大学で講義評価が実施されている。講義評価に大きく影響する項目は講師の話し方であるとの報告もされている。従って、

講師は話し方を改善するための支援システムがあれば、それを用いることで講義の質を高めることができると考えられる。

本研究者は大学に勤務していることから、音声データ収集のために講師の協力が得やすく、多くのサンプルデータを入手できるため円滑に本研究を推進できる背景にあった。さらに、2003

年に講演音声データベース「日本語話し言葉コーパス」が国立国語研究所から公開されたことから、大量の話し言葉のサンプルが利用可能となった。研究では、大学講義やこれらの講演音声进行分析し、システムを構築していくことが可能となった。

社会的・学術的背景、大量のデータが分析可能になったことから、本研究を進める環境が整った。

これまでに、学会講演や大学講義音声において講師の音声进行分析する研究に取り組んできた。

この研究では、講義音声を音響的・言語的に分析し、音声に含まれる物理量（声の高さ、大きさ、話速等）や言語的特徴であるフィラー（「あー」「えーと」等の間投詞）を抽出した。そして、それらの特徴量と学生のアンケート評価との関連性を調査した。

その結果、各種の音響的特徴やフィラーの使い方が学生の講義印象評価結果と深く関係していることを明らかにした。また、特徴量の主成分分析を行うことで講師毎の話し方の特徴を一目で表示できる方法を開発した。さらに、大学が期末毎に実施する授業評価結果との関連性を調査し、話し方評価が講義内容までも考慮した全体のアンケート評価と関連性が高いことがわかった。これは講義音声の特徴が講義評価にも影響することを示唆している。

しかし、2008年までに分析してきた10名という講師の数では、信頼性が高いとは言えない。そこで、これらの研究成果を発展させるために、さらなるサンプル音声データ収集、聴取実験、音声分析による新しい特徴量の発見が必須であった。

2. 研究の目的

本研究では、話し方トレーニング支援システムを構築することで講演・講義における講師の話し方を改善することで、聴き手の話の理解度を高めるための支援を目指す。図1に提案システムを利用した話し方改善の流れを示す。

講義音声を支援システムに入力する。システムは音声認識結果や特徴量から話し方を自動評価し、講師に評価結果とアドバイスを返す。その結果を元に講師は講義を行い、話し方の再評価を行う。

システムでは話し方の評価が基本となっているため、話し方評価を如何に正確に自動で行うかということが焦点となる。そこで、図1に示すように話し方評価を行うために以下の3つの課題に取り組む。

- (1) 講義データの収集と音声分析～「収集・分析プロセス」
- (2) 有効な音声特徴量の決定と講義評価と

の関連付け～「学習プロセス」

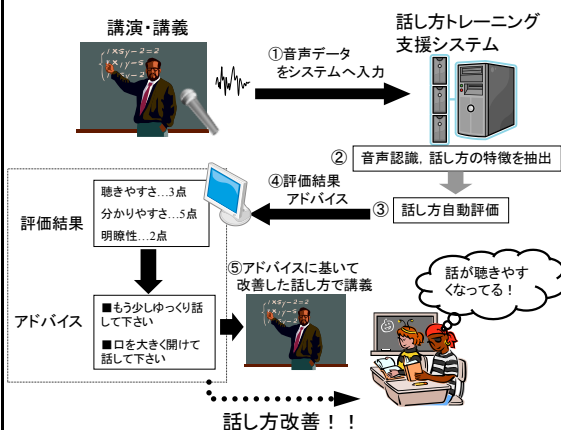


図1. 全体構想

(3) 講義音声の評価と講義改善～「評価プロセス」

これらの課題を順次解決した後、全体のシステムを構築する。

3. 研究の方法

本研究の目的を達成するために、まず、講義や講演などの話し方が聴き手に及ぼす影響を調査することが必要である。そして、聴き手にとって聴きやすく分かりやすい話し方の特徴を、音響的・言語的な側面から明らかにし、講義や講演における話し方トレーニング支援システムを構築する。目的を達成するため主に、次の3つのステップで研究を行った。

- (1) 音声データの収集と分析、
- (2) 有効な音声特徴量の決定と音声印象評価との関連付け、
- (3) 講演・講義音声の評価

(1) 音声データの収集と分析

本研究を遂行するためには、できるだけ多くの音声データ収集・分析が必要である。そこで、新たに講義・講演音声に関するデータを収集する。既存の講義・講演音声データを被験者に聴取してもらい、受ける印象や明瞭性に関するアンケート調査を行うことで、“音声”と“聴き手の印象評価”のデータセットを用意する。具体的には、音声の中のフィラーの有無等に基づく印象評価、音声の中の特定語の明瞭性判定評価等のデータを大量に収集する。

(2) 有効な音声特徴量の決定と音声印象評価との関連付け

用意した音声と人間による印象評価結果を用いて、様々な話し方の特徴量と印象評価を関連付け、聴き手の印象に与える影響が大きい特徴量を調査する。

特徴には、声の高さ・大きさ・話速・明瞭度などがあり、言語的特徴としては、フィラー（間投詞）・ポーズの入れ方等が考えられる。

これらの物理的な特徴に対して、回帰分析等の統計的な処理を施すことで、自動的に話し方を評価するための枠組みを模索していく。

また、新しい特徴量として、音声認識システムより得られる情報が自動評価に有効であるか否かを調査する。

(3) 講演・講義音声の評価

(2)のステージで得られた知見を用いて、音声の聴きやすさや理解しやすさに影響する音声特徴をモデル化することで、話し方の自動評価を行う。具体的には、重回帰分析や決定木等の統計的手法を用いることで、これを実現する。

(4) システムの評価

(3)の枠組みを用いて、話し方を自動評価するシステムを構築する。構築したシステムを用いて、講演者にアドバイスとしてフィードバックを行い、話し方が改善するかどうかを調査する。

4. 研究成果

(1) 話し方の印象評定データベースの構築

既存の講義データに加えて、新しく7名の教員の授業音声を収録し、同時に授業音声評価のアンケートを実施した。

また、既存の講義・講演音声データを被験者に聴取してもらい、受ける印象や明瞭性に関するアンケート調査を行うことで、“音声”と“聴き手の印象評価”のデータセットを用意した。

具体的には、音声中のフィラーの有無等に基づく印象評価、音声中の特定語の明瞭性判定評価等のデータを大量に収集した。最終的に、講師21名、約50講義の講義音声データを収集することができた。

また、講演・講義中の特定の単語やフレーズの明瞭性に着目し、それら特定の単語やフレーズを被験者に聴いてもらい、正しく聞き取れるかどうかの被験者実験も行った。これにより、聴きやすい単語と聴き取りにくい単語のデータを収集した。

(2) 有効な音声特徴量の決定と音声印象評価との関連付け

音声に含まれる様々な特徴と、被験者によるアンケート結果を調査した結果、次のような特徴が聴きやすさに影響していることが明らかとなった。

- ・声の抑揚、

- ・声の大きさ、
- ・話速
- ・明瞭度
- ・ポーズ
- ・フィラー（間投詞）
- ・音声認識精度

これらの特徴についての調査結果を次に述べる。

①音響的特徴量と関連性

声の大きさについては、むしろ、声が高い方が良いが、マイクを用いることで、良い講義を行うことができる。

そこで、その他の特徴である、声の抑揚、話速、明瞭性、ポーズを用いて、講師の特徴を可視化した。それを図2に示す。これは、授業音声の各特徴量を主成分分析することで2次元に圧縮し、グラフにプロットしたものである。

この図を見ると、授業科目が異なったとしても、同じ講師であれば、同じような位置にプロットされていることが分かる。図2において、グラフの右半分かつ上半分の位置にある授業ほど高い授業評価を得ており、逆に、グラフの左半分かつ下半分の位置にある授業ほど授業評価は低い。

これにより、声の特徴を用いることで、講師の音声の聞き取りやすさを自動評価できていることがわかった。

授業評価の高い音声（話し方の評価が高い音声）は、特に明瞭度の評価が高いことが明らかになった。

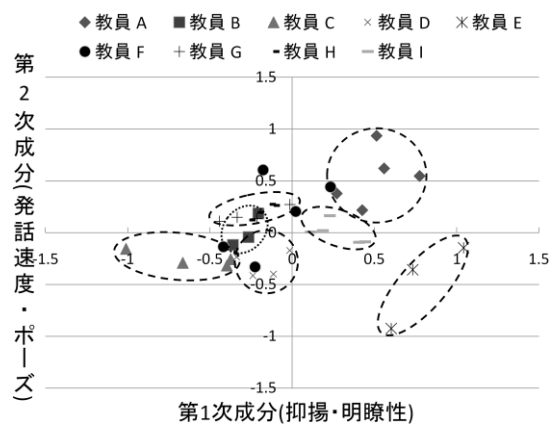


図2. 教員の特徴分析結果

②音声認識精度との関連性システムを用いた評価

講義における話し方の良さは、講義音声の明瞭性評価と関連があることが分かった。

明瞭性と高い特徴を調査したところ、講義音声の音声認識精度と講義の明瞭性評価に高い相関(0.6)があることがわかった。こ

れにより、音声認識システムを用いることで、講義の自動評価が行えるのではないかとということが示唆される。

そこで、音声認識システムを用いて、講義の評価が行えないかを検討した。音声認識システムから得られる特徴は次の通りである。

- ・母音率（音声全体に占める母音音素の割合）
- ・話速
- ・尤度（音声認識のスコア）
- ・単語信頼度

音声認識精度と明瞭性とは高い相関があるが、音声認識精度は、音声の正確な書き起こしがなければ算出することができない。

そこで、上記の特徴を用いて、まず、音声認識精度を推定し、それを用いて明瞭性の評価を行うことにした。

重相関分析により、上記の特徴を用いて明瞭性を推定したところ、アンケート結果の値との相関値が **0.74** となった（クローズドの条件）。

これにより、音声認識システムを用いることで明瞭性、すなわち、話し方の評価ができそうであることがわかった。

しかし、オープンな環境でテストをおこなったところ、推定した明瞭度とアンケート結果との相関値は **0.25** と低くなってしまった。弱い相関があるものの、このままでは利用できない。

発話ごとに評価を行う、他の特徴量と組み合わせる等、今後、改良が必要である。

③ フィラーとの関連性

次に、先行研究により明らかになった、フィラーと話し方の関係について、調査した。その結果、話の中に適切にフィラーを取り入れることによって、聴きやすい話し方になることが明らかとなった。

その知見を生かし、フィラーおよびポーズに着目した話し方評価およびアドバイスシステムを構築した。

詳細については次節で報告する。

(3) フィラーとポーズを用いた話し方の自動評価システムの構築

① 特徴量

本システムでは、音声の中でのフィラーに関する 13 種類の特徴量を用いて評価・アドバイスを行っている。使用している特徴量は以下のとおりである。なお、入力音声はあらかじめ発話単位に分割されている。文頭とは、分割された発話の先頭の単語を意味し、文中とは、分割された発話の先頭及び末尾以外の単語を意味する。

- ・フィラー率：
フィラー単語数÷全単語数
- ・フィラーの文頭率：
文頭フィラー数÷全フィラー数

- ・単独フィラー（フィラーのみの発話）率：
単独フィラー数÷全フィラー数
- ・文頭フィラー長：
文頭フィラーの継続時間の平均
- ・文中フィラー長：
文中フィラーの継続時間の平均
- ・フィラーの時間率：
全フィラー継続時間の合計÷総発話時間
- ・フィラーの文頭時間率：
文頭のフィラーの継続時間の合計÷全フィラーの継続時間の合計
- ・SP（ショートポーズ）率：
SP 数÷全単語数
- ・SP 時間率：
SP の継続時間の合計÷総発話時間
- ・ピッチ差（前）：
フィラーと直前単語の平均のピッチ差
- ・ピッチ差（後）：
フィラーと直後単語の平均のピッチ差
- ・パワー差（前）：
フィラーと直前単語の平均のパワー差
- ・パワー差（後）：
フィラーと直後単語の平均のパワー差

② 自動評価・アドバイスシステム

評価には決定木を用いている。また、評価の高くなる枝をたどるようにアドバイスを行うことで、話し方の改善点を提示できるようにしている。決定木にはADTreeという学習アルゴリズムを用いている。Wekaという機械学習ソフトウェアによりADTreeが実現できる。

学習データは日本語話し言葉コーパス

(CSJ) の72講演を用い、各講演から約1 分間の音声を切り出して用いた。それらの音声に対して、被験者に「話し方は上手いか？」という質問をおこない4 段階で回答してもらった。その結果を標準化したものを評価値とし

て学習に用いる。システム上での評価は、入力音声をGood, Normal, Bad の3 種類に分類する。被験者回答結果の標準化後の値が、+1 以上をGood, -1 以上+1 未満をNormal, -1 未満をBad として学習を行った。Good, Normal を分類する木とBad, Normal を分類する木の多数決により、評価を決定する。

ADTree はノード毎に評価が得られるので、これをアドバイスに利用した。悪い評価を得た決定木中の各ノードにおいて、逆の枝が選択されるようにそのノードで評価されているフィラーの特徴量を改善するアドバイスを提示する。

③ システムの評価

構築したシステムをクローズドテストで評価した。評価結果を表1 に示す。72 講演中、正解講演は71 であり、ほぼ正確に評価の予測ができた。

次に、アドバイス通りにフィルターの使い方を改善し、システム上での評価が改善した場合、実際に人間が聴いても、話し方が上手くなったと感じるのかを聴取実験により調査した。

使用した音声は、CSJ の講演を 1 人の人間が読み上げたものである。システムの評価が悪い音声をシステムのアドバイスに従って修正し、システムの評価が良くなった 5 音声 9 名の被験者に聞き比べてもらい、どちらの話し方がうまく感じるか回答してもらった。

結果を表 2 に示す。音声 X がシステムのアドバイスに従って話し方を改善したもの、音声 Y が元音声である。5 音声中 4 音声において、アドバイス後の音声の評価が高かった。また、システム上での評価に差がない 2 組の音声 (表 2 の下 2 組) は、被験者の評価でもあまり差が出なかった。

以上の結果より、評価は正しく行えており、アドバイスも有効であることがわかった。また、フィルターから講演の話し方のうまさに対して、評価・アドバイスを行えることもわかった。

表 1. システムの評価

		システムの評価		
		Good	Normal	bad
人間の評価	Good	10	0	0
	Normal	0	50	0
	Bad	0	1	11

表 2. システムのアドバイス効果

システム評価		人間の評価	
音声 X	音声 Y	X>Y	X<Y
Good	Normal	8	1
Good	Normal	7	2
Good	Normal	2	7
Good	Bad	8	1
Good	Bad	7	2
Normal	Normal	3	6
Normal	Normal	4	5

X>Y : X より Y の方が良いとした評価人数

(4) その他の成果

その他の研究成果として、講義音声の音声認識率改善などに取り組んだ。

講演・講義の自動評価を行うためには、音声認識率が高くないと精度良く評価できないためである。例えば、フィルターの検出精度等は、音声認識率に大きく影響する。

そこで、音声認識率を改善するためのいくつかの方法を開発した (例えば、雑誌論文③、学会発表③④等)。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 6 件)

- ① 西崎博光, 関口芳廣, 教員の話し方改善支援システムの開発に向けた講義音声の特徴分析, 日本教育工学会論文誌, 査読有, 34 巻, 2010, pp. 171-179
- ② Masatoshi Tsuchiya, Satoru Kogure, Hiromitsu Nishizaki, Kazumasa Yamamoto, and Seichi Nakagawa, "Construction and Analysis of Corpus of Japanese Classroom Construction and Analysis of Corpus of Japanese Classroom Lecture Speech Contents Lecture Speech Contents," 2009 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2009), 査読有, 2009, CD-ROM
- ③ 藤原裕幸, 西崎博光, 関口芳廣, 話題依存言語モデル構築のための LSA と単語発話情報を用いた語彙推定, 第 8 回情報科学技術フォーラム講演論文集, 第 2 分冊, RE-004, pp.35-42, 2009 年 9 月 2 日 発表, 東北工業大学 (宮城県)

[学会発表] (計 7 件)

- ① 内藤航, 西崎博光, 関口芳廣, 決定木を利用した講演音声のフィルターの使い方アドバイスシステムの試作, 日本音響学会 2011 年春季研究発表会講演論文集, 2-P-6(b), pp.427-428, 2011 年 3 月 10 日, 早稲田大学 (東京都)
- ② 石丸聡, 西崎博光, 関口芳廣, 音声認識誤りの手動訂正インタフェースにおける自動訂正の効果, 日本音響学会 2011 年春季研究発表会講演論文集, 2-P-52(b), pp. 231-232, 2011 年 3 月 10 日発表, 早稲田大学 (東京都)
- ③ 土屋厚郎, 西崎博光, 関口芳廣, 話し言葉音声の認識改善のための適切な音声分割手法の検討, 日本音響学会 2011 年春季研究発表会講演論文集, 2-P-28(b), pp.191-192, 2011 年 3 月 10 日発表, 早稲田大学 (東京都)
- ④ 張 安, 関口芳廣, 西崎博光, 講義音声認識精度改善のためのチョーク音除去, 情報処理学会, 創立 50 周年記念 (第 72 回) 全国大会講演論文集, Vol.2, pp.241-242, 3U-1, 2010 年 3 月 10 日発表, 東京大学 (東京都)

〔産業財産権〕

○出願状況（計1件）

名称：音声認識装置で利用される音声認識用
辞書、音声認識用言語モデルの学習方法

発明者：西崎博光，上平拓弥，高岡充，関口
芳廣

権利者：山梨大学

種類：特許

番号：特願 2010-014372

出願年月日：2010年1月26日

国内外の別：国内

6. 研究組織

(1) 研究代表者

西崎 博光 (NISHIZAKI HIROMITSU)

山梨大学・大学院医学工学総合研究部・助
教

研究者番号：40362082

(2) 研究分担者

なし

(3) 連携研究者

なし