情報処理学会論文誌　Vol. 50　No. 2　460–468 (Feb. 2009)

書き起こしへの付与を目指した音声とテキストを対象とした発話印象の分析

西田昌史[†1]　堀内靖雄[†1]
黒岩眞吾[†1]　市川熹[†2]

近年，音声から書き起こしを自動的に作成するシステムに関する研究がさかんに行われている．これまでは，音声を正確に書き起こすことに重点をおいて研究されてきているが，見た者にとって議論の内容をより理解しやすい書き起こしの作成が重要であると考えられる．議論の内容を正確に伝えるには言語情報だけでは不十分であり，議論の場面や発話意図，感情といった情報も必要であると考えられる．そこで，本研究では会議や討論などの書き起こしに発話意図を付与することを目指し，テキストと音声の両方から発話印象について分析することを目的とした．まず，文字の太さや大きさの変化といった文字の装飾や，「！」，「？」などの記号に着目し，そのようなテキストの変化を書き起こしに付与する形で主観評価実験を行うことにより「疑問」，「驚き」などの発話印象がどの程度感じられるのかを調べた．また，音声についても同様に主観評価実験を行い，その結果と「F0」や「パワー」などの韻律パラメータを使って重回帰分析を行い，韻律パラメータと発話印象の関係を分析した．その結果，各テキスト変化，韻律パラメータとそれぞれの発話印象との関係が明らかになった．さらにそれらを総合的に分析することで，テキストと音声では発話印象の受け方が異なるものと，同じ傾向のものがあることが明らかになった．

Analysis of Utterance Impressions in Speech and Text for Indexing to Transcriptions

Masafumi Nishida,[†1] Yasuo Horiuchi,[†1]
Shingo Kuroiwa[†1] and Akira Ichikawa[†2]

In recent years, a great amount of research has been done on systems that transcribe utterances through automatic speech recognition. This research has generally been focused on transcribing utterances correctly. What is presently required, however, is a transcription method that enables the overall content of a given discourse to be more easily understood by readers. It is generally considered that linguistic information by itself is insufficient for this purpose, and that a way of showing speaker's intentions and emotions is also required. In this study, we analyzed user's impressions of utterances from both text and speech, with the aim of at indexing the impressions to the transcriptions of discourse forums such as meetings and discussions. We investigated how impressions such as "doubt" and "surprise" are felt by changing the size of written characters and indexing signs such as question marks and exclamation marks in the text. The relation between prosody parameters and utterance impressions was analyzed by using multiple linear regression. As a result, we were able to clarify the relationship between variations of text, prosody parameters, and utterance impressions.

1. はじめに

古くから，会議や討論の内容をその場にいなかった人に伝える，また会議の内容を振り返る手段として議事録などの書き起こしが作成されている．しかし，手作業で正確な書き起こしを作成することは困難であり，なおかつ莫大な労力が必要となる．そこで近年，音声認識技術を用いた書き起こしの自動作成についての研究がさかんに行われてきている[1)–3)]．しかし，そのような従来研究では，音声を正確に文字に書き起こすことを目的としているため言語情報しか扱われておらず，また人手により作成した書き起こしテキストに比べて要点がまとまっていないために理解しにくいという問題点がある．

従来のシステムでは音声から言語情報のみを抽出し，それを文字という形で出力しているが，会議や討論だけでなくあらゆる音声において，実際は同一話者であっても声の高さや大きさ，話す速さなどが頻繁に変化している．そして，それらが発話の抑揚となり相手に主張の強さや疑問などを伝えており，議論の論点をとらえるうえで大変重要となっている．このように，発話の内容を正確に把握するためには言語情報だけでなく，発話の意図や感情といったパラ言語情報が必要であると考えられる．そして，そのパラ言語情報を書き起こしに付与することができれば，より理解しやすい書き起こしが作成可能になると考えられる．そこで，本研究ではそのパラ言語情報を付与した書き起こしを作成することを目指す．

音声から知覚することのできるパラ言語情報に関する研究はさまざまなものが行われているが，特に人間の感性の尺度として，感情や発話意図，態度などを用い，それらと韻律情

†1 千葉大学
Chiba University
†2 早稲田大学
Waseda University

報やスペクトル情報といった物理パラメータとの関連性について研究を行っているものに注目した．森山らは，聞き手が音声から知覚する感情と，音声が感情を含むことによって生じる物理的変化の双方を基底空間に写像してから対応づけるモデルを提案している[4]．また，音声の聞き手の感情のステレオタイプの基底概念が，「快–不快」「緊張–弛緩」「注目–拒否」といった心理学で報告されている基本感情と一致することを示している．さらに，音声が感情を含んだときに生じる物理的な変化として，「声高さ」「抑揚」といった成分が支配的であるという結果を示している．有本らは，自然な対話の中で話者の怒りの感情を音響的な特徴でとらえる手法について検討を行っている[5]．分析には判別分析と決定木との 2 つの方法を用いており，open なデータに対し 90%を超える正解率を示している．また，異質の怒りが混在している場合にも柔軟に対応できるとして決定木の方が有効であるという結果を示している．小野寺らは，コールセンタの対話を音声資料として用い，語句の談話要素と韻律情報の関係を調べ，そこで得られた韻律特徴を利用して判別実験を試みている[6]．音声には「相槌」，「確認」，確認に対する応答である「確認返答」などの談話タグが付けられており，タグが「相槌」かつ「確認返答」と判別された発話は対話に対する理解度が高い発話としている．そのときの話速との関係を調べることで，対話に対する理解度が高い状況では話速が速くなる傾向にあるという結果を示している．藤江らは，発話態度が肯定的であるか否定的であるかをパラ言語情報を用いて認識することを試みている[7]．この研究では，パラ言語情報として韻律情報のほかに頭部ジェスチャを用いている．そして，これらの情報を統合してユーザの言語的に曖昧な応答に対し肯定的/否定的態度を判断し，対話制御を行う音声対話システムを実現している．

また，テキスト表現とパラ言語情報の関係についての研究は文献 8)，9) などがある．江尻らは，文字の大きさと色（白，黄，赤，緑，ピンクの 5 色）の変化により話者の興奮が伝わるかを調査している[8]．それにより，色によって興奮の印象の受けやすさが異なり，大きさは大きくなるにつれてより興奮した印象を受ける傾向があるという結果を示している．片山らは，テレビの字幕の色や大きさ，表示位置を変化させて感情の伝わりやすさと字幕の見やすさについて調べた．その結果，大きさを変化させることが最も感情を伝えやすく，色をつけることは見やすさにおいて評価が低いことを示している[9]．

パラ言語情報に関する研究としては上記のようなものが行われている．しかし，これまでの研究では，特定の感情に着目しそれらの印象の変化が主に分析対象とされてきており，音声とパラ言語情報，テキストとパラ言語情報それぞれについては研究されているものの，音声とテキストの関係性は分析されるにいたっていない．パラ言語情報を付与した書き起こしテキストを作成するためには，従来研究のような，音声，テキスト各々とパラ言語情報の関係についての分析に加え，それをもとに音声とテキストとの関係性を分析していく必要がある．より理解しやすい書き起こしを作成するためには，感情や発話意図といったパラ言語情報を複数推定し付与することが望ましいと考えられる．その際音声中には印象が単独で出現する場合と，複数の印象が出現する場合が考えられる．そこで，テキストと音声の関係性を分析することで，単一の印象を表現するには他の印象と区別するためのテキスト表現と韻律情報を明らかにし，複数の印象を表現する場合においてはどのようなテキスト表現で表すことが望ましいかが明らかになるのではないかと考えられる．

そこで，本研究では，音声からパラ言語情報を付与した書き起こしの作成を目指して，まずテキスト変化と音声の関係を明らかにすることを目的とした．パラ言語情報には，従来研究で示したように発話意図や感情などが含まれており比較的広い概念であると同時に，その定義もまちまちである．そこで，本研究では基本感情よりも発話意図を書き起こしに付与することを目指していること，およびテキストや音声から感じるものということで，以後本論文で扱うパラ言語情報を発話印象と呼ぶ．その際，言語情報ではなく韻律により感じられる印象でかつテキストの大きさなどの装飾や記号の付与により表現できると考えられる印象に着目した．これまでの研究でまず手始めに強調，疑問，驚き，自信，迷いの 5 つの印象を著者らで主観的に選び，1 フレーズ単位で韻律パラメータから印象を推定する手法について検討を行った[10]．しかしながら，その後の検討により強調については何度も同じ発話をしたり，発話内の一部を強調したりするといった表現を行うことが一般的であると考え，本研究では疑問，驚き，自信，迷いの 4 つを対象とした．これらの発話印象を文字の大きさや記号を付与することでテキストからどのように感じられるか，また対話音声を対象としてこれらの発話印象がどのような韻律情報により表現されるかを分析する．さらに，テキストと音声における発話印象の受け方を総合的に分析することで，テキストと音声における発話印象についての類似点ならびに相違点を明らかにする．

以下，2 章では発話印象を付与した書き起こしの例を示し，3 章ではテキストの変化と発話印象との関係の分析，4 章では音声における発話印象の分析，5 章ではテキストと音声の比較による発話印象の分析，6 章でまとめと今後の課題について述べる．

2. 書き起こしへの発話印象の付与

例として通常の言語情報のみの書き起こしと，発話印象を付与した書き起こしを**図 1** と**図 2** に示す．これらは，本研究で使用した日本語地図課題対話コーパスの対話データの中

話者A：じゃ城壁のある街の左要するに
話者B：上行く左上行く
話者A：左上の行ってでそれその絵の左側にそって
真下に行けばいいの
話者B：そうそうそうそう 真下にうん
話者A：真下に行くと今度ゴーストタウンてあるでしょう
話者B：ゴーストタウン
話者A：そっちにはないのか
話者B：ないな

図 1　通常の書き起こし
Fig. 1　Normal transcription.

話者A：じゃ城壁のある街の左要するに
話者B：上行く左上行く
話者A：左上の行ってでそれその絵の左側にそって
真下に行けばいいの？
話者B：そうそうそうそう 真下にうん
話者A：真下に行くと今度ゴーストタウンてあるでしょう？
話者B：ゴーストタウン！？
話者A：そっちにはないのか？
話者B：ないな

図 2　発話印象を付与した書き起こし
Fig. 2　Transcription-indexed utterance impressions.

から一部を抜粋して書き起こしたものである．通常の書き起こしでは，言語情報しか使われていないために話者がどのような意図で発話したのかが分かりにくく，議論の場の雰囲気が読み取りにくい．それに対して発話印象を付与した書き起こしでは，文字の大きさを変えたり記号を利用したりすることで発話印象の付与を試みている．たとえば「自信」がある発話に対しては文字を大きくし，「迷い」のある発話に対しては文字を小さくする．また，「疑問」を表現した発話には「？」を，「驚き」を表現した発話には「!?」を付与することで，話者がどのような意図で発話していたのかが想像され，議論の内容がより理解しやすくなると考えられる．

これまで筆者らは，書き起こしに付与することを目的として，対話音声を対象として韻律情報から発話印象を推定する手法について検討を行ってきた[10),11)]．しかし，実際に発話印象を付与した書き起こしを作成する場合，発話印象をテキストでどのように表現するのかも重要な問題である．そのために，本研究では発話印象として「疑問」，「驚き」，「自信」，「迷い」の 4 つに注目し，テキスト表現と音声の両方から発話印象の分析を行った．テキストにおける表現法については，文字の大きさの変化や「！」，「？」などの記号に，音声については F0（基本周波数）やパワー，平均モーラ長といった韻律パラメータに注目して分析を行った．なお，最終的には会議や討論の音声の書き起こしを作成することを目指しているが，会議や討論の音声では複数の話者の発話が重なることがよくあるために発話を明確に取り出すのが困難であると考えられる．そこで，今回は 1 対 1 の対話音声を対象とした．

3. テキストの変化と発話印象の分析

テキストの変化と発話印象との関係を分析するために，まずテキスト変化の主観評価実験を行った．テキスト変化には，文字の太さや大きさを変えるといった文字の装飾，「！」や「？」などの記号を用い，それらのテキスト変化を付与することで「疑問，驚き，自信，迷い」の発話印象がそれぞれどの程度感じられるのかを調べた．そして，その結果からテキスト変化と発話印象の関係性，ならびに各々の発話印象間の関係性を分析した．

3.1 実 験 方 法

実験を行うためのデータとして，94 年度に千葉大学で収録された「日本語地図課題対話コーパス（通称マップタスク）」というデータベースを利用した．これは 2 人一組の対話音声を収録したものである．お互いの顔がガラス越しに見え，ヘッドホンから相手の声が聞こえるだけという環境で，それぞれの話者には「湖」や「パン工場」などの目印が描かれている地図が与えられる．一方の話者（giver）に渡される地図にはスタート地点から目標地点までの正解ルートが示されている．もう一方の話者（follower）に渡される地図にはその正解ルートが示されていない．正解ルートが示されている地図を持っている話者が示されていない地図を持っている話者へ指示し，そのルートを書き込んでいくという課題を行ったときの対話を収録している．本研究では，すべてこの条件で収録された 10 対話のデータを用いた．データの長さは 1 対話 10 分程度である．この音声は対話であるが，人に伝えることを目的とした発話をしているという面では会議や討論と共通であるので，その結果は会議，討論の音声にも応用可能であると考えられる．

テキストの変化と発話印象との関係を分析するために，テキストの変化により受ける発話印象の主観評価実験を行った．発話印象には，「疑問，驚き，自信，迷い」の 4 つを用いた．これは，書き起こしに付与することで発話内容の再現性を向上させることができると考えられるものとして選んだ．実験に用いた全テキスト変化を**表 1** に示す．本研究では，疑問，

表 1　実験に用いた書体の変化
Table 1　Style variations used in experiments.

変化の種類	変化させたテキスト例
太字	**4 センチぐらい**
文字縮小	4 センチぐらい
文字拡大	4 センチぐらい
下線	4 センチぐらい
?付与	4 センチぐらい?
!付与	4 センチぐらい!
!?付与	4 センチぐらい!?
…付与	4 センチぐらい…

驚き，自信，迷いの 4 つの印象を書き起こしに付与することにしたため，これらの印象を表現できると考えられるものとして，見やすさを考慮し比較的単純なものとしてテキストの文字の大きさや記号の付与に着目しこれらの 8 つの表現を選択した．

フレーズ（言語情報）には「4 センチぐらい」，「左上」，「右上にそのまま真っ直ぐ上に」，「S の字を描きながら」の 4 種類を用いた．これらの 4 種類のフレーズは前述の日本語地図課題対話コーパスの中に含まれる発話の一部である．文字数により発話印象への影響がある可能性を予想し，さまざまな長さのものを含むようにこの 4 種類を選んだ．上記の発話印象，テキストの変化，フレーズを用いて主観評価実験を行った．

実験では，上段に変化させていない通常のテキスト，下段に変化させたテキストを提示し，上段のテキストに比べて下段のテキストから各発話印象がどの程度感じられるのかを 0（感じられない），1（やや感じられる），2（感じられる），3（とても感じられる）の 4 段階で評価してもらった．基準となる変化させていないテキストの文字は Microsoft Word において大きさ 10.5 pt の明朝体であり，A4 の紙面上に印刷したものを実験に用いた．テキスト変化を提示する順番によって発話印象に影響を与えてしまうと考えられる（たとえば，文字を縮小したものを評価した後に文字を拡大したものを評価すると，実際よりも過大評価してしまう可能性がある）ので，順番をランダムに提示するようにしてその影響を低減させた．なお，被験者は大学生，大学院生合計 10 名で，全部で 128 データ {= 4（発話印象の数）× 8（テキスト変化の種類）× 4（フレーズの種類）} について評価を行った．

3.2　実験結果と考察

各テキスト変化ごとに，被験者 10 名×フレーズ 4 種類の 40 個ずつ評定結果が得られる．今回は，40 個すべての評定結果の平均値をその発話印象の評定値とした．

まず，**表 2** に発話印象間の相関係数を示す．表 2 の結果から，自信と迷いの相関係数は

表 2　書体の変化における発話印象間の相関係数
Table 2　Correlation between utterance impressions in style variations.

	疑問	驚き	自信	迷い
疑問	1.0	0.52	−0.60	0.52
驚き	0.52	1.0	0.0	−0.08
自信	−0.60	0.0	1.0	−0.76
迷い	0.52	−0.08	−0.76	1.0

表 3　各書体の変化における発話印象の評定値
Table 3　Evaluation of utterance impressions in style variations.

	疑問	驚き	自信	迷い
太字	0.0	0.2	1.0	0.0
文字縮小	0.2	0.0	0.0	1.7
文字拡大	0.0	0.4	1.7	0.0
下線	0.0	0.1	0.7	0.0
?付与	2.4	0.1	0.1	1.2
!付与	0.0	0.7	1.6	0.0
!?付与	1.9	2.1	0.2	0.9
…付与	0.5	0.0	0.0	1.8

−0.76 となっており，自信と迷いのように直観的に反対の意味を持つと考えられる発話印象間には強い負の相関が見られていることから，妥当な評定結果が得られていると考えられる．また，「疑問」と「驚き」，「疑問」と「迷い」に正の相関が見られる．

次に，各テキスト変化における発話印象の評定値を**表 3** に示す．

また，**表 4** に各テキスト変化における発話印象の評定値の標準偏差を示す．ここで標準偏差が 0 になっているものは，すべての話者において印象がないと判断されていた．この結果から，すべての印象とテキスト変化の組合せにおいて，評定値の標準偏差がほぼ 1.0 以下であり，比較的安定した評定結果が得られていると考えられる．さらに，評定値が 1 以上になった割合を**表 5** に示す．

本研究では，評定値が 1.0（やや感じられる）以上となったものはその発話印象が現れているもの（印象あり）とし，評定値が 1.0 より小さいものはその発話印象が感じられないもの（印象なし）とした．ただし，実際に発話印象を書き起こしに付与する場合，人によって発話印象の受け方にばらつきが少ないテキスト変化を扱うことが望ましい．特に，印象ありと印象なしの間で人によって評定がばらついているものは避けるべきである．

表 5 に示した割合が高いものほど，人によって印象あり，印象なしのばらつきが少ない

表 4　各書体の変化における発話印象の評定値の標準偏差
Table 4　Standard deviation of evaluation of utterance impressions in style variations.

	疑問	驚き	自信	迷い
太字	0.0	0.5	0.9	0.0
文字縮小	0.4	0.0	0.0	0.7
文字拡大	0.0	0.7	1.1	0.0
下線	0.0	0.3	0.8	0.0
？付与	0.5	0.3	0.3	0.9
！付与	0.0	0.9	0.8	0.0
!? 付与	0.8	0.7	0.5	0.8
…付与	0.6	0.0	0.0	0.8

表 5　評定値が 1 以上になった割合（%）
Table 5　Percentage of evaluation values of one or greater.

	疑問	驚き	自信	迷い
太字	0	13	68	0
文字縮小	23	0	0	100
文字拡大	0	28	83	0
下線	0	8	50	0
？付与	100	8	3	78
！付与	0	48	85	0
!? 付与	93	100	18	63
…付与	45	0	0	95

ものであるといえる．各発話印象において，表 5 で評定者が 1.0 以上と評定した割合が最も高く，表 3 で印象ありとなったものをまとめると，「疑問」では「？付与」，「驚き」では「 !? 付与」，「自信」では「！付与」，「迷い」では「文字縮小」となった．以上のようなテキスト変化が最もその発話印象を感じやすいものであると考えられる．

「疑問」を表現するのに最も有効な「？付与」は「迷い」を，「驚き」を表現するのに最も有効な「 !? 付与」は「疑問」を同時に感じることが分かった．このことから，「疑問」と「迷い」を同時に感じる音声の場合は「？付与」，「驚き」と「疑問」を同時に感じる場合は「 !? 付与」により複数の印象を表現することが可能となる．

また，印象が単一な音声の場合は，「自信」であれば「文字拡大」あるいは「！付与」，「迷い」であれば「文字縮小」あるいは「…付与」で表現することで他の発話印象との区別がしやすくなる．

4. 音声における発話印象の分析

音声と発話印象との関係を分析するために，まずテキストの場合と同様に音声から感じられる発話印象の主観評価実験を行った．主観評価実験は，3 章で分析の対象としたテキスト変化により表現可能な「疑問，驚き，自信，迷い」の 4 つについて行った．そしてその主観評価実験の結果と，F0（基本周波数）の平均値・最大値・最小値・レンジ，パワー（声の大きさ）の平均値・最大値・最小値・レンジ，平均モーラ長の平均値の 9 個の韻律パラメータを用いて，重回帰分析により発話印象をモデル化した．

4.1 実 験 方 法

実験に用いる音声は，3 章で述べた日本語地図課題対話コーパスに含まれる音声のうち 10 対話分で，性別により発話印象の受け方が変化する可能性があると考え，話者は男性のみに統一した．話者は全部で 10 名，1 人 1 度ずつ目標地点まで誘導する側（giver）と誘導される側（follower）になり対話している．各話者十数発話ずつ，全 130 発話を音声データとして選択した．言語情報が発話印象に影響を与えることを避けるため，文脈の情報が表れにくい 2，3 秒程度までの比較的短いものを選んだ．また，書き起こしへの印象付与が有意であると考えられる，韻律情報がないと意味を理解しにくいような発話を優先した．発話例としては，「廃屋の左上」，「五時の方向にはい」，「墓地ない」，「4 センチぐらい」，「S の字を描きながら」などである．

音声を聞いて各発話印象がどの程度感じられるのかを 0（感じられない），1（やや感じられる），2（感じられる），3（とても感じられる）の 4 段階で主観評価実験を行った．話者が代わることによる影響で評定値が変動する可能性を考慮して，同一話者の音声を連続で聞かせた．また，音声の順番も入れ替え，順番による影響も現れないようにした．10 名の被験者に参加してもらい，130 音声 × 4 発話印象で計 520 回の評定を行った．1 回の音声で 4 つの発話印象を同時に評価させなかったのは，できるだけ直観的に受けた印象を評価させるためである．

4.2 韻律パラメータ

使用した音声資料は標本化周波数 20 kHz，量子化ビット 16 bit となっている．パラメータ取得のためのツールとして「WaveSurfer 1.6.3」[12] を用いた．

今回分析には，F0（基本周波数）の平均値・最大値・最小値・レンジ，パワー（声の大きさ）の平均値・最大値・最小値・レンジ，平均モーラ長の 9 個の韻律パラメータを用いた．F0 の平均値が大きいものほど全体的に声の高い発話であり，レンジの値が大きいほど抑揚

のある発話ということになる．パワーの平均値が大きいものほど全体的に声の大きな発話であり，レンジが大きい発話は声の大きさの変動が大きな発話ということになる．平均モーラ長とは話す速度を表すパラメータであり，平均モーラ長の値が大きくなるほど話す速度は遅くなる．人間が何かしらの意図をもって発話する場合，声の高さ，大きさ，話す速さをさまざまに変化させることでその意図を伝えていると考えられるためこのようなパラメータを用いている．

今回，F0 の値はフレーム間隔 10 msec で抽出したが，有声と判断されたところのみを使用し，有声部に対して F0 の発話内平均を求めた．Wavesurfer では，RAPT（A Robust Algorithm for Pitch Tracking）というアルゴリズムにより F0 を抽出している．RAPT は，相関法によって求められた複数の F0 候補を用いて，動的計画法を用いた後処理によって精度の高い F0 抽出を実現したものである[13]．また，F0 の最大値，最小値は，そのまま F0 の最高値と最低値を用いると誤差を含む可能性が高いため，百分位で発話内全サンプル中の上位 10%と下位 10%の値を求め，その値をその発話の F0 の最大値，最小値とした．F0 の発話内レンジはその最大値と最小値との差で求めた．それぞれの値は話者ごとに個人差があると考えられるので，話者ごとの F0 パラメータを話者ごとに集めて平均値と標準偏差を求めて，平均が 0，分散が 1 になるように正規化した．

パワーの値も同じく Wavesurfer を利用しフレーム間隔 10 msec ごとに対数パワーを求めた．F0 の場合と同様に百分位で発話内全サンプル中の上位 10%と下位 10%の値を求め，その値を最大値，最小値とし，その差を発話内レンジとした．発話内平均も F0 の場合と同様に有声の部分について求めた．また，F0 の場合と同様に話者ごとに正規化をした．

平均モーラ長は，発話の継続時間をモーラ数で割ることで求めた．その後 F0，パワーと同様の正規化も行っている．なお，平均モーラ長はテキストを与えて求めている．

4.3　実験結果と考察

主観評価実験の結果，各発話印象において印象ありとなったデータ数の内訳は「疑問」が 40，「驚き」が 21，「自信」が 31，「迷い」が 30 となった．なお，全 130 発話のうちどの発話印象も感じられない音声が 42 発話あり，本研究ではこれを平静音声と見なす．また，主観評価実験での印象語の評定値の標準偏差は，疑問 1.4，驚き 1.1，自信 1.2，迷い 1.2 という結果となり，印象ごとに評定の変動に偏りもなく比較的安定していると考えられる．さらに，テキスト変化より音声のほうが同じ印象語でも印象の感じ方にバラツキがあることが分かった．

音声における発話印象間の相関関係を**表 6** に示す．

表 6　音声における発話印象間の相関係数
Table 6　Correlation between impressions of speech utterances.

	疑問	驚き	自信	迷い
疑問	1.0	0.47	−0.69	0.78
驚き	0.47	1.0	−0.04	0.29
自信	−0.69	−0.04	1.0	−0.86
迷い	0.78	0.29	−0.86	1.0

表 7　変数選択により得られた韻律パラメータ
Table 7　Prosody parameters obtained by selecting variables.

疑問	F0 平均値，F0 レンジ，パワー最大値，パワーレンジ
驚き	F0 最大値，F0 最小値，F0 レンジ，パワーレンジ
自信	F0 レンジ，パワー平均値，平均モーラ長
迷い	F0 最大値，パワー最小値，平均モーラ長

表 6 において，自信と迷いのように逆の意味を持つと思われる発話印象には負の相関が見られるので，妥当な評定結果が得られていると考えられる．また，「疑問」と「迷い」に強い正の相関が得られている．

各発話印象ごとの重回帰式を式 (1) から式 (4) に示す．なお，y_a は「疑問」，y_b は「驚き」，y_c は「自信」，y_d は「迷い」の重回帰式を表している．また，式中の x_1，x_2，x_3，x_4 はそれぞれ F0 の平均値，最大値，最小値，レンジを，x_5，x_6，x_7，x_8 はそれぞれパワーの平均値，最大値，最小値，レンジを，x_9 は平均モーラ長を表している．

$$y_a = 0.58 + 0.5x_1 + 0.57x_4 - 0.56x_6 + 0.72x_8 \quad (1)$$

$$y_b = -0.11 - 0.69x_2 + 1.44x_3 + 1.83x_4 + 0.36x_8 \quad (2)$$

$$y_c = 0.5 + 0.25x_4 + 0.87x_5 - 0.23x_9 \quad (3)$$

$$y_d = 0.44 + 0.41x_2 - 0.32x_7 + 0.32x_9 \quad (4)$$

変数選択を行った結果得られた韻律パラメータを**表 7** に示す．

以上の結果から，「疑問」ではパワーレンジの偏回帰係数が最も大きく，F0 平均値・レンジ，パワー最大値が選択されており，「疑問」が感じられる音声は声の大きさと高さの変動が大きい．「驚き」では F0 レンジの偏回帰係数が最も大きく，F0 最大値・最小値，パワーレンジが選択されており，「驚き」が感じられる音声は声の高さの変動が大きく，声の高さ

の最低値が大きい．「自信」ではパワー平均値の偏回帰係数が最も大きく，F0 レンジ，平均モーラ長が選択されており，「自信」が感じられる音声は全体的に声が大きくなる．「迷い」では F0 最大値の偏回帰係数が最も大きく，パワー最小値，平均モーラ長が選択されており，「迷い」が感じられる音声は声の高さの最高値が大きく，全体的に声が小さくなる．

これらの分析結果から，各発話印象で共通に得られた韻律パラメータにより発話印象間で音響的に似ているものが明らかになり，異なる韻律パラメータが発話印象間を区別するのに有効なものであることが分かった．

各発話印象における重回帰式の決定係数は，疑問 0.42，驚き 0.69，自信 0.61，迷い 0.28 という結果が得られた．この結果から，驚きと自信に関しては今回用いた韻律パラメータで表現することができていることが分かった．一方，疑問と迷いに関しては，今回用いた韻律パラメータでは不十分であり，それ以外にも有効な韻律パラメータが存在していることが明らかになった．また，韻律パラメータによって迷いを表現することが難しいことが分かった．今後，疑問や迷いを表現するのに有効な韻律パラメータについてさらに分析する必要がある．

5. テキストと音声の比較による分析

ここでは，これまでの分析をふまえてテキストと音声の両方を比較して分析を行う．**表 8** にテキストと音声における発話印象間の相関係数を示す．

表 8 の結果から，テキストと音声で相関関係が同じ傾向を示した発話印象は，「疑問」と「驚き」でともに正の相関，「疑問」と「自信」でともに負の相関，「驚き」と「自信」で無相関，「自信」と「迷い」でともに強めの負の相関であった．

また，テキストと音声で相関関係に変化が生じている発話印象は，「疑問」と「迷い」でテキストに比べて音声のほうがより強い相関があるため，音響的に似た変動をするため互いを同時に感じやすいと考えられる．さらに，「驚き」と「迷い」でテキストでは無相関であるが音声ではやや正の相関があることから，音響的に似た変動をするため互いを同時に感じやすいと考えられる．

以上の結果から，テキストと音声とでは関係性が変化する発話印象と，同じ関係性が現れるものが存在することが分かった．また，「疑問」と「驚き」，「疑問」と「迷い」ではテキストと音声ともに正の相関が見られたことから，同時に感じられる割合が高い発話印象であると考えられる．

次に，発話印象を書き起こしに付与する際，韻律パラメータからテキスト変化を推定することが今後重要となるため，テキスト変化と韻律パラメータの関係についても分析を行った．表 3 から各テキスト表現において評定値が 1.0 以上であった印象において，表 7 で変数選択により得られた韻律パラメータのうち偏回帰係数が最大値と最小値のものを**表 9** に示す．ここで，「!? 付与」に関しては，疑問と驚きの評定値がほぼ同じだったため，互いの印象で表現される韻律パラメータを記載した．

以上の結果から，声が大きいとテキストが太く文字が大きい，声が小さいと文字が小さい，声の大きさの変動が大きいと「？」を付与，声が大きく話す速度が速いと「！」を付与，声の高さが大きく声が小さいと「…」を付与といった関係が明らかになった．

表 8　テキストと音声における発話印象間の相関係数
Table 8　Correlation between impressions of text and speech utterances.

	テキスト	音声
疑問⇔驚き	0.52	0.47
疑問⇔自信	−0.60	−0.69
疑問⇔迷い	0.52	0.78
驚き⇔自信	0.00	−0.04
驚き⇔迷い	−0.08	0.29
自信⇔迷い	−0.76	−0.86

表 9　書体の変化と韻律パラメータの関係
Table 9　Relation between style variations and prosody parameters.

	最大の偏回帰係数	最小の偏回帰係数
太字	パワー平均値	平均モーラ長
文字縮小	F0 最大値	パワー最小値
文字拡大	パワー平均値	平均モーラ長
下線	パワー平均値	平均モーラ長
？付与	パワーレンジ	パワー最大値
！付与	パワー平均値	平均モーラ長
!? 付与	パワーレンジ	パワー最大値
	F0 レンジ	F0 最大値
…付与	F0 最大値	パワー最小値

6. おわりに

本研究では，発話印象を書き起こしに付与することを目指して，テキストの変化と音声の両方から発話印象について分析を行った．まず，テキスト変化による発話印象の主観評価実験を行い，どのようなテキストの変化で各発話印象がどの程度感じられるのかを調べた．そ

の結果，テキスト変化により単独で表現できる発話印象と，単一のテキスト表現で複数の発話印象を表現できることが明らかになった．また，表 3 のテキスト変化における発話印象の評定値において，たとえば疑問では「!? 付与」で 1.9,「？付与」で 2.4 のように，同じ発話印象でもテキストの表現によって印象の感じやすさが異なることも明らかになった．次に，音声においても同様に発話印象の主観評価実験を行い，その結果と韻律パラメータを用いて重回帰分析を行うことで，韻律パラメータと発話印象との関係を分析した．その結果，各発話印象がどのような韻律パラメータと関係があるかが分かった．さらに，テキストと音声での発話印象の受け方を総合的に分析していくことで，テキストと音声では発話印象の受け方に違いがあるものと，テキストと音声で同じように感じられるものとがあることが明らかになった．

今後は，今回扱っていないその他の発話印象の種類，テキストの変化，韻律パラメータについても，さらに追加してより詳細な分析を行っていきたいと考えている．また，韻律パラメータを自動的に抽出する手法について検討を行う必要がある．さらに，発話印象を付与した書き起こしを作成した場合，見やすい書き起こしであることも重要である．そこで，実際に発話印象を付与した書き起こしを作成し，見やすさについても評価していきたいと考えている．

謝辞　在学中にテキストと音声における印象評定などの実験や分析を行った小川純平氏に感謝いたします．

参 考 文 献

1) 篠崎隆宏，古井貞熙：日本語話し言葉コーパスを用いた講演音声認識，情報処理学会論文誌，Vol.43, No.7, pp.2098–2107 (2002).
2) 堤　怜介，加藤正治，小坂哲夫，好田正紀：発音変形依存と教師なし適応による講演音声認識の性能改善，話し言葉の科学と工学ワークショップ講演予稿集，pp.93–98 (2004).
3) 秋田祐哉，河原達也：話し言葉音声認識のための汎用的な統計的発音変動モデル，電子情報通信学会論文誌，Vol.J88-D-II, No.9, pp.1780–1789 (2005).
4) 森山　剛，斉藤英雄，小沢慎治：音声における感情表現語と感情表現パラメータの対応付け，電子情報通信学会論文誌，Vol.J82-D-II, No.4, pp.703–711 (1999).
5) 有本泰子，大野澄雄，飯田　仁：「怒り」識別のための音声の特徴量の検討，人工知能学会研究会資料，SIG-SLUD-A303-03, Vol.40, pp.13–19 (2004).
6) 小野寺佐知子，落谷　亮：テキスト情報と韻律情報を利用したコールセンター対話の分析，人工知能学会研究会資料，SIG-SLUD-A303-08, Vol.40, pp.45–50 (2004).
7) 藤江真也，江尻　康，菊地英明，小林哲則：肯定的/否定的発話態度の認識とその音声対話システムへの応用，電子情報通信学会論文誌，Vol.J88-D-II, No.3, pp.489–498 (2005).
8) 江尻芳雄，金森康和：話者の興奮度合いを適用した字幕表現，電子情報通信学会技術研究報告，SP2006-106, pp.1–6 (2006).
9) 片山滋友，鈴木久仁子，谷　史織：テレビの字幕提示における感情伝達の方法とその効果，電子情報通信学会総合大会講演論文集，p.424 (2002).
10) 西田昌史，小川純平，堀内靖雄，市川　熹：対話音声を対象とした韻律情報による発話印象のモデル化，電子情報通信学会技術研究報告，SP2005-105, pp.79–84 (2005).
11) 西田昌史，小川純平，堀内靖雄，市川　熹：韻律特徴に基づく対話における発話印象の推定，日本音響学会講演論文集，1-4-7, pp.225–226 (2006).
12) http://www.speech.kth.se/wavesurfer/
13) Talkin, D.: A robust algorithm for pitch tracking (RAPT), *Speech Coding and Synthesis*, Kleijin, W. and Paliwal, K. (Eds.), pp.495–518, Elsevier (1995).

(平成 20 年 6 月 5 日受付)
(平成 20 年 11 月 5 日採録)

西田　昌史（正会員）

平成 9 年龍谷大理工学部電子情報学科卒業．平成 11 年同大学大学院理工学研究科電子情報学専攻修士課程修了．平成 14 年同大学院博士後期課程修了．博士（工学）．平成 14 年 4 月～平成 15 年 6 月科学技術振興事業団さきがけ研究 21「協調と制御」領域博士研究員．平成 15 年 7 月～平成 19 年 3 月千葉大学大学院自然科学研究科助手．平成 19 年 4 月より同大学院融合科学研究科助教．音声情報処理，話者認識，音声認識，ヒューマンインタフェース，福祉情報工学に関する研究に従事．電子情報通信学会，日本音響学会，人工知能学会各会員．

堀内　靖雄（正会員）

平成 2 年東京工業大工学部情報工学科卒業．平成 7 年同大学大学院理工学研究科情報工学専攻博士課程修了．博士（工学）．同年千葉大学工学部助手．平成 12 年千葉大学大学院自然科学研究科助手．平成 14 年同大学院助教授．平成 19 年より同大学院融合科学研究科准教授．音楽情報処理，音声言語処理，福祉情報工学に関する研究に従事．平成 4 年情報処理学会全国大会奨励賞，平成 7 年人工知能学会研究奨励賞，平成 8 年人工知能学会全国大会優秀論文賞．人工知能学会，日本音響学会，ヒューマンインタフェース学会，ソフトウェア科学会，社会言語科学会，日本手話学会各会員．

黒岩　眞吾（正会員）

昭和 61 年電気通信大学電気通信学部通信学科卒業．昭和 63 年同大学大学院修士課程修了．博士（工学）．同年国際電信電話株式会社入社．昭和 63 年〜平成 13 年同社研究所において機械翻訳システムおよび電話音声認識システムの研究・開発に従事．平成 13 年徳島大学工学部助教授．平成 19 年より千葉大学大学院融合科学研究科教授．音声認識，話者照合，音声信号処理，自然言語処理の研究に従事．平成 8 年度電子情報通信学会学術奨励賞，日本音響学会第 3 回および第 5 回技術開発賞受賞．日本音響学会，電子情報通信学会，人工知能学会各会員．

市川　　熹（正会員）

昭和 39 年慶應義塾大学工学部卒業．日立製作所中央研究所を経て平成 4 年千葉大学工学部教授，平成 10 年同大学院自然科学研究科教授．現在，早稲田大学人間科学学術院客員教授，千葉大学名誉教授．工学博士．信学会フェロー．音声対話理解，手話，指点字，福祉情報機器等に取り組み，特に対話言語のプロソディ情報に関心を持つ．最近は言語獲得や発達障害の課題にも眼を向けている．信学会理事，音声研委員長，手話工学研副委員長，福祉情報工学研委員長，音響学会評議員，人工知能学会理事，言語・音声理解と対話処理研主査，JEITA アクセシビリティ標準化対応専門委員長等を歴任．他に電子情報通信学会，IEEE，ISCA，言語処理学会，HI 学会，音声言語医学会，手話学会等の各会員．昭和 64 年信学会論文賞，平成 8 年人工知能学会研究奨励賞，平成 17 年総務大臣賞等受賞．