

発話意図判定における文型と韻律の関係

益永祐吾[†] 川端豪[†]

韻律情報に基づく話者の意図判定において、文型の及ぼす効果について調べる。感情認識によく用いられる基本周波数に基づく各種の特徴量に加え、話者の緊張を表すスペクトル特徴量を用いて検討する。判別分析を用いて、特徴量の有効性や韻律と意図の関係、文型の効果などを調べていく。例えば、話し手の「期待」を聞き手に伝えるためには、「質問」など韻律を反映させやすい文型が必要である。励ますという意図の伝達のためには、貶す表現を使うと韻律が反映しにくいなどの知見が得られた。

Effects of Sentence Styles on Utterance Function Detection based on Prosodic Features

Yugo Masunaga[†] and Takeshi Kawabata[†]

This paper describes the effects of sentence styles on the prosodic-based utterance function detection system. Prosodic features based on fundamental frequencies are popular to emotion recognition. Additionally, spectrum parameters indicating the tension of speakers are investigated for the utterance function detection. By utilizing the discriminant analysis, we search for the essential factor of these parameters. The discrimination experiments of the utterance function shows that the interrogative sentence style is necessary for determining the speaker's optimistic mental state. It is difficult to notify the utterance function "encourage" using the "ill expressed" sentence style.

1. はじめに

人間とコンピュータが、音声によって情報交換を行うに際し、お互いの内部状態を把握することが重要になってくる。感情の判別についての研究は、これまでも感情(怒

り、悲しみ、喜び)に加え、平静時の音声連続音声認識システムに与える影響を示した研究[1]や喜び、怒り、悲しみ、嫌々と平静の5感情を込めて発声した単文に含まれる感情の判別の可能性を3種類の方法で示した研究がある[2]。また、パラ言語情報として、韻律情報と頭部ジェスチャに現れる発話者の発話態度を取り上げ、それが肯定的/否定的発話態度の識別を試みることで、それぞれの認識手法を提案するとともにそれらを活用した対話ロボットを扱い音声対話システムを実現した研究がある[3]。

本報告における「意図」とは、言語学的に議論される深層的な概念ではなく、発話によって相手に何らかの影響を与えようとする「機能」を指す。例えば、励ますという意図は、話し手が聞き手を元気づけるという機能、咎めるという意図は、話し手が聞き手を落ち込ませるといった機能に対応付けられる。本研究は、発話の意図を韻律情報に基づいて判別する枠組みにおいて、その発話を構成する文型に注目し、それが、判別にどのような影響を及ぼすか、いくつかの事例について検討し、意図(機能)、韻律、文型の3者の関係のモデル化を試みる。

同じ意図の発話で、文型が異なることで韻律が変わってくる事例を3種類検討し、従来から良く用いられている基本周波数及び発声速度に基づく特徴群と、N/S比として提案されているスペクトルに基づく特徴群、及び両者の併用の3種類の条件で、判別分析を行い、いろいろな組み合わせで発話意図の判別精度が、どのように変化するかを観察する。

例えば、発話意図「期待」「嫌々」については基本周波数に基づく特徴量が重要であり、そのためには「質問」など、基本周波数の変化を許容する文型が必要である。また、発話意図「励ます」「咎める」については話者の緊張を表すスペクトル歪みが重要であり、貶す表現で励ますという不自然な発話が、話者の緊張を増加させるという現象が見られる。3節で詳細に分析する。

2. 発話意図判定における文型と韻律

2.1 発話意図の判定

人間とコンピュータが、音声によってコミュニケーションを行うためには、音声認識や音声合成の技術は必須であるが、円滑に会話を行うためには人間同士の会話で行われるようなパラ言語情報の理解が必要であると考えられる。

パラ言語情報の一種である感情について「怒り」、「喜び」、「悲しみ」に平静時を表す「標準」を加え、計4感情の韻律情報を用いて、感情分析を行った研究[1]や、平静からのずれを用いて、感情分析を行っている研究[2]がある。

コンピュータが、人間の内部状態である喜怒哀楽などの感情を理解することで、人間に対して適切な反応を返す音声対話システムを構築することができると考えられるが、さらに一歩進んで人間が「励ます」や「咎める」などの意図を持って発話した音

[†] 関西学院大学 理工学研究科
School of Science and Technology, Kwansai Gakuin University

声から、その発話の持つ機能を判定できるようになれば、さらにリアルな反応を返すコンピュータを構築できると考えられる。

ここで、本報告における「発話の意図」について述べておく。ここで言う意図とは、言語学的に議論される深層的な概念ではなく、発話によって相手に何らかの影響を与えようとする「機能」を指すものとする。例えば、励ますという意図は、話し手が聞き手を元気づけるという機能、咎めるという意図は、話し手が聞き手を落ち込ませるという機能に対応付けられる。

私たちの日常生活において、このような機能は、韻律によって実現されていることが多い。例えば、武田らは皮肉や感謝など韻律情報から判定しようと試みている[4][5]。しかし、一方で、発話される文章の字面すなわち文型もまた、この意図の伝達に関係があると思われる。

本研究は、発話の意図を韻律情報に基づいて判別する枠組みにおいて、その発話を構成する文型に注目し、それが、判別にどのような影響を及ぼすか、いくつかの事例について検討した。これによって、意図(機能)、韻律、文型の3者の関係をモデル化することを試みる。

2.2 意図伝達に影響を及ぼす文型

ある発話に込められた意図は、その口調すなわち韻律的な特徴によって、相手に伝達されることが多いと考えられる。このとき、その発話の文章のスタイルすなわち文型が、何がしかの影響を及ぼさないだろうか。人間とコンピュータの音声コミュニケーションシステムにおいて、意図の伝わりやすい文型があるのならば、ぜひ、それを利用していくべきではなからうか。

さて、同じ意図の発話であっても、用いる表現が違うことにより、韻律が変わってくるような事例をいろいろ考えてみた。表1に同じ「依頼」の内容の文章に対し、2種類の文型を用いた具体的表現例を示す。このような意図に対する文型例を各々15種類ずつ準備した。文章と意図の対応付けを以下に述べる。

【事例】

話し手が聞き手にある情報を求めるに際し、その情報に対し期待感を持って聞いているのか、嫌々聞いているのかという意図判定を設定し、何らかの韻律情報によって、その判別を行う。このとき、発話の文型を制限し、質問文型を用いる場合と断定文型を用いる場合で、その判別の精度にどのような影響が表れるのかを調べる。

【事例】

話し手が聞き手に評価の表現で話しかけるに際し、聞き手を励ますつもりなのか咎めるつもりなのかという意図判定を設定し、何らかの韻律情報によって、その判別を行う。このとき、発話の文型を制限し、褒める表現を用いる場合と貶す表現を用いる場合で、その判別の精度にどのような影響が表れるのかを調べる。この設定は、実は、人間にとって不自然なものとなっており、そのような場合の観察結果が得られる。

【事例】

話し手が聞き手に何らかの警告を発するに際し、聞き手に対し、強く警告するつもりなのか軽く注意するつもりなのかという意図判定を設定し、何らかの韻律情報によって、その判別を行う。このとき、発話の文型を制限し、命令文型を用いる場合と勧誘文型を用いる場合で、その判別の精度にどのような影響が表れるのかを調べる。

以上の3種類を表2にまとめておく。

表1 異なる文型による意図の伝達

文型	表現例
質問	今日の実験内容を聞かせてくれない?
断定	今日の実験内容を聞かせて

表2 発話する文章の組とその文章に対して発話する意図

文章の組	発話する意図
質問 断定	期待感を持って相手に聞く 嫌々な気持ちで相手に聞く
褒める 貶す	励ますつもりで相手に話す 咎めるつもりで相手に話す
命令 勧誘	強く警告するつもりで相手に話す 軽く注意するつもりで相手に話す

2.3 特徴量の選択

これまでに研究されてきた感情分析において、文中の最大基本周波数、最大振幅数、基本周波数の変動範囲などの特徴量が有効であると述べられている研究[1]や韻律情報と音韻情報とを調べ、喜び、怒り、悲しみ、嫌々と平静の5感情を込めて発声した単文に含まれる感情の判別の可能性を3種類の方法で示した研究がある[2]。本研究では、これらの研究を参考に特徴量を選定し、さらに声質変換型感情音声合成の研究[6]で用いられたN/S比などを用いて、発話の意図の判別を試みた。使用した特徴量を以下に示す。

- ・基本周波数の平均値
- ・基本周波数の分散値

- ・基本周波数の最大値
- ・基本周波数の最小値
- ・発話速度
- ・N/S 比の平均値
- ・N/S 比の分散値
- ・N/S 比の最大値
- ・N/S 比の最小値

N/S 比は、残差スペクトルに含まれる雑音成分の量を定量化する指標であり、人間の声質の中でも雑音成分のレベルが感情表現に影響を与えているということが研究 [6] で報告されている。N/S 比の算出方法について簡単に説明しておく。

音声は 16kHz で標準化し、フレーム (512 点) ごとに分析窓長 (1024 点) で LPC 分析を行い、残差スペクトルを計算する。楕型フィルタによって、調波成分を抜き出し、そのパワーを S 、全体のパワーから調波成分を減算したパワーを N としたとき、N/S 比は、次式のように計算される。

$$N/S \text{ ratio} = 10 \log_{10}(N/S) \quad (dB) \quad (1)$$

これらの特徴量をパラメータとして、カテゴリー全体に対する級間分散と級内分散の比 (フィッシャー比) を最大化する基準で判別分析を行い、カテゴリーの分布の分離度を調べた。判別分析に用いるパラメータとしては、どの系統の特徴量が発話の意図の判別に影響を及ぼしやすいか観察するために、次の 3 種類の組み合わせについて検討した。

【判別分析】

基本周波数の平均値、分散値、最大値、最小値と発話速度の組み合わせ

【判別分析】

N/S 比の平均値、分散値、最大値、最小値の組み合わせ

【判別分析】

全特徴量 (基本周波数の平均値、分散値、最大値、最小値、発話速度、N/S 比の平均値、分散値、最大値、最小値)

3. 実験結果と考察

3.1 実験条件

発話の意図、韻律、文型の相互作用に対するモデルを樹立するためには、多数の話者による一般性を反映した分析を行うことが必要であるが、今回の報告では、その初期検討として、被験者 1 名による分析を行い、このようなアプローチが有望かどうかを確認することにした。

音声データの収録には、暗騒音 36dB の防音室で行った。被験者には、椅子に座ってもらい、マイクと顔の間には、三本の指が入る程度の間隔をあけて発話してもらうようにした。マイクは単一指向性のダイナミックマイクロフォンを使用し、A/D 変換機、ミキサーを経て、録音用の PC へ出力される。PC で録音される音声は、サンプリング周波数 16kHz、16 ビット、モノラルの音声ファイルになる。被験者は、大学の学生 (男性、兵庫県出身) である。表 3 (a) 表 3 (b) を 1 組とし、同様に表 4 (a) 表 4 (b)、表 5 (a) 表 5 (b) の 3 組の文章を発話した。また、音声資料を読み上げる際に話し手 (被験者) と聞き手が 1 対 1 の状況であることをイメージして発話し、録音前には十分練習した。一つの音声資料につき、15 文の発話文章を用意し、その 15 文に対して、ある意図を込めて発話する。これを 10 回繰り返した。1 つの音声資料につき、2 つの意図を込めて発話するので、収録する音声データの数は、1800 個になる。

2.2 節に述べた 3 つの事例について、2.3 節に述べた 3 つのパラメータの組み合わせに基づく判別分析を行い、その判別性能を観察する。

表 3 依頼の内容の文章の組
(a) 質問文型

No	発話内容
1	今日の実験内容を聞かせてくれない？
2	今週のスケジュールを聞かせてくれない？
3	部屋の掃除をやっておいてくれない？
4	そのレポート，代わりに提出しといてくれない？
5	このプロジェクトは，きみにまかせていいかい？
6	植木に水やりしといてくれない？
7	ガスコンロの火を消しといてくれない？
8	テーブルの上のコップをとってきてくれない？
9	今日の食事は，きみが作ってくれないか？
10	ゴミは，分別しておいてくれないか？
11	ディベートの司会役をよろしく頼める？
12	私の分の切符もついでに買っておいてくれないか？
13	君は，飲み物を持ってきてくれないか？
14	その書類を30枚コピーしておいてくれないか？
15	目的地に先に行っというてくれないか？

(b) 断定文型

No	発話内容
1	今日の実験内容を聞かせてく
2	今週のスケジュールを聞かせてくれ
3	部屋の掃除をやっておいて
4	そのレポート，代わりに提出しといて
5	このプロジェクトは，きみにまかせる
6	植木に水やりしといて
7	ガスコンロの火を消して
8	テーブルの上のコップをとってきて
9	今日の食事は，きみが作ってくれ
10	ゴミは，分別しておいて
11	ディベートの司会役をよろしく頼む
12	私の分の切符もついでに買って
13	君は，飲み物を持ってきてくれ
14	その書類を30枚コピーしておいて
15	目的地に先に行っというて

表 4 褒める内容と貶す内容の文章の組
(a) 褒める文型

No	発話内容
1	その洋服，かわいい
2	その帽子，かっこいい
3	今日は，きれい
4	君は，えらい
5	君は，器用
6	君は野球が上手
7	君は，気が利く
8	君は，素晴らしい
9	君は，優しい
10	君は，足が速い
11	君は，すごい
12	君は，強い
13	その服，似合ってる
14	君は，よくがんばる
15	君が作ったクッキーは，おいしい

(b) 貶す文型

No	発話内容
1	その洋服，かわいくない
2	その帽子，かっこよくない
3	今日は，きれいじゃない
4	君は，えらくない
5	君は，器用じゃない
6	君は野球が，上手じゃない
7	君は，気が利かない
8	君は，素晴らしくない
9	君は，優しくない
10	君は，足が速くない
11	君は，すごくない
12	君は，弱い
13	その服，似合っていない
14	君は，よくがんばらない
15	君が作ったクッキーは，おいしくない

表 5 命令する内容と勧誘する内容の文章の組

(a) 命令する文型

No	発話内容
1	そろそろ時間だから、早く起きろ
2	明日は、絶対遅刻するな
3	成績が悪いのだから、勉強しろ
4	間に合わないから、早く走れ
5	そろそろ寝なさい
6	先生の話はちゃんと聞いておけ
7	レポートの提出期限を守れ
8	時間通りに動け
9	その荷物は、壊すな
10	後片付けをちゃんとしろ
11	テレビの電源を消せ
12	先に宿題をしろ
13	忘れ物は、するな
14	水の節約をしろ
15	ごみは分別して捨てる

(b) 勧誘する文型

No	発話内容
1	そろそろ時間だから、早く起きよう
2	明日は、絶対遅刻しないようにしましょう
3	成績が悪いのだから、勉強しよう
4	間に合わないから、早く走ろう
5	そろそろ寝よう
6	先生の話はちゃんと聞いておこう
7	レポートの提出期限を守ろう
8	時間通りに動こう
9	その荷物は、壊さないようにしましょう
10	後片付けをちゃんとしよう
11	テレビの電源を消そう
12	先に宿題しよう
13	忘れ物は、しないようにしましょう
14	水の節約をしよう
15	ごみは分別して捨てよう

3.2 実験結果

表 6, 表 7, 表 8 に 3 つの事例に対して, 3 つの条件の判別分析を行った, 各組み合わせに対し, 得られた判別性能を示す. 各数値は, 発話意図に対する判別率(=正解数/資料数)である.

まず, 事例 について観察していこう. この事例は, 話し手が聞き手に情報を求めるに際し, 期待感を持つか, 嫌々聞くかという意図を韻律情報によって判定する枠組みになっている. 質問文型を用いる場合と断定文型を用いる場合で, その判別の精度にどのような影響が表れるのかを調べる.

判別分析 すなわち基本周波数と発話速度に関わる韻律情報のみから, この意図の判定を行おうとすると質問文型では, 86%の精度が得られるのに対し, 同じ意図を断定文型で, 伝達させようとすると 10%も精度が低下することがわかる. これは, 韻律によって意図を伝達するためには, 韻律を表現できる文型, ここでは, 質問文型が必要であることを意味している.

次に判別分析, N/S 比に関わる韻律情報から, この意図の判定を行うと文型によらず 95%という高い判定精度が得られた. N/S 比は, 調波構造が崩れることによって値が大きくなる特徴量であり, 緊張に基づくスペクトル歪みが反映されると考えれば, 文型が質問であるか, 断定であるかには関わらず, この韻律情報を伝達することができるので, 文型による差が表れなかったと考えられる.

全ての特徴量を利用する判別分析 では, 基本周波数と発話速度および N/S 比, 各々のもつ韻律情報が, 全て利用でき, さらに性能が向上することが観察された.

次に事例 について観察していこう. 話し手が聞き手に評価の表現で話しかけ, 聞き手を励ますかあるいは咎める. 発話の文型は, 褒めるか貶すである.

判別分析 を見ると事例 の場合とは違って, 表現による判定精度の差があまり大きくないことが分かる. どちらの文型も基本周波数の韻律には, 影響を及ぼさないためであろう.

一方, スペクトル歪みに関わる特徴量を用いた判別分析 では, 褒める, 貶す表現の違いが, 意図判定に大きな影響を及ぼしていることが分かる. これは, 励ます 咎めるという対比が, 褒める 貶すという対比と概念的に非常に近いため, 貶す表現で励ますという発話が, 極めて不自然になり, 話者の緊張を増加させたのではなからうか.

判別分析 では, 全体的に性能が向上している.

事例 について観察する. 聞き手に対し, 強く警告するつもりか, 軽く注意するつもりかという意図を判定する. 文型は, 命令文型と勧誘文型を用いる.

判別分析 では, 勧誘表現の方が命令表現よりも警告の強度を聞き手に伝達するた

めに有効であることが観察された。これは、命令表現の持つ強い意志伝達の意図が、軽く注意という意図を乗せにくかったためと思われ、判定結果が強い警告側に大きく偏っていた。

判別分析では、多少、命令文型の方が、勧誘文型より精度は良いが、どちらの文型もあまり良い結果ではない。全体として難しい事例になったと思う。

判別分析では、全特徴量を総合的に利用することによって、このような難しい事例においても比較的精度が向上していることが分かる。

表 6 発話意図「期待」「嫌々」の判定結果

発話文章の組	判別分析	判別分析	判別分析
質問	86%	95%	98%
断定	76%	95%	97%

表 7 発話意図「励ます」「咎める」の判定結果

発話文章の組	判別分析	判別分析	判別分析
褒める	75%	79%	87%
貶す	72%	69%	81%

表 8 発話意図「強い警告」「軽い注意」の判定結果

発話文章の組	判別分析	判別分析	判別分析
命令	57%	71%	71%
勧誘	71%	66%	74%

4. おわりに

本報告では、発話の意図の判別に対して、韻律情報を利用して行い、かつ、同じ意図の発話でも文章の表現・文型の違いによる判別精度の違いを被験者一人の音声データを扱い調査した。発話の意図の判別に韻律情報を利用することができることが分かった。例えば、発話意図「期待」「嫌々」については基本周波数に基づく特徴量が重要であり、そのためには「質問」など、基本周波数の変化を許容する文型が必要。また、発話意図「励ます」「咎める」については、貶す表現で励ますという不自然な発話が、話者の緊張を増加させるという現象が見られ、韻律による意図判定に対して文型が様々な形で影響を及ぼすことが確認できた。今回の検討では、初期段階ということもあり、被験者一人のデータでしか行っていないので、今後は、被験者を増やし、話者単独データでの判別精度や話者混合データでの判別精度及び発話の意図の判別における有効なパラメータについて検討し、一般性を広げていきたい。

参考文献

- 1) 門谷信愛希, 阿曾弘具, 鈴木基之, 牧野正三: 音声に含まれる感情の判別に関する検討, 信学技報, NLC2000-34, SP2000-82, pp.43-48, (2000.12)
- 2) 重永實: 感情の判別分析からみた感情音声の特性, 信学論(A), Vol.J83-A, No.6, pp.726-735 (2000.6)
- 3) 藤江真也, 江尻康, 菊池英明, 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, 信学論, D-II, Vol.J88-D-II, No.3, pp.489-498 (2005)
- 4) 浜崎敏幸, 光本浩士, 大和多寛, 田村進一, 柳田益造: 皮肉発声に対する知覚的印象の韻律からの推定, 情処研報, 99-SLP-27, Vol.99, No.64, pp.1-8 (1999.7)
- 5) 武田昌一: 声優が発声した日本語「感謝」表現音声の感情の程度に応じた韻律的特徴の比較, 音響論集, 2-Q-34, pp.441-442 (2003.3)
- 6) Takeda, S. et al : Analysis of Voice-Quality Features of Speech that Expresses "Anger", "Joy", and "Sadness" Uttered by Radio Actors and Actresses, Proc. Interspeech 2008, pp.2114-2117 (2008.9)