

韻律を用いたテキスト非限定型発話アクト識別方法*

◎藤尾 茂 ニック キャンベル 樋口 宜男 (ATR 音声翻訳通信研究所)

1 はじめに

対話型応答システムの制御や音声翻訳を行なう上で、発話アクトや発話意図を理解することが重要である。文字列を認識するだけで発話アクトや発話意図を識別できることもあるが、同じ文字列で数種類の意図が考えられる場合は韻律や文脈などを利用する必要がある。例えば、「行く」という文字列が発声された場合、文末で基本周波数が上昇していれば、疑問を意図しており、下降していれば情報の提供を意図するというように発話アクトの識別には韻律が必要である。

我々は同一表記で複数の発話アクトの可能性のある発話に対して、韻律を用いてその発話アクトを識別する方法を既に報告しているが[1]、この方法は限られた文字列の発話に対してのものであり一般化されたものではなかった。対話型応答システム等のアプリケーションへの適用を考えた場合、限られた文字列だけでなく任意の文字列に対応することが必要となる。そこで、任意の文字列の発話に対して適用可能な発話アクトの識別方法について提案し、その方法を用いた発話アクトの識別実験を行なった。本稿ではその結果について報告する。

2 テキスト非限定な発話アクトの識別

我々が既に提案している発話アクトの識別方法では、同一の文字列からなる発話で観測された特徴量の分布に基づいて閾値を決定することが必要であった。このため同じ発話タイプを扱う場合であっても任意の文字列の発話に適用できるわけではなかった。このような問題点を解決するためには、文字列に制限のない発話アクトの識別方法を確立する必要がある。本稿では、任意の文字列からなる発話に対して躊躇やためらいの発話アクト(以下、temporizerとする)と、疑問の発話アクト(以下、questionとする)を検出する手法を提案する。

なお、疑問の発話アクトは「はい/いいえ」で答えられる疑問、確認の疑問等細かく複数に分けることもできるが、本稿では全てをまとめて1つの発話アクトとして扱う。

2.1 temporizer の識別方法

temporizerの「そうですか」の場合には、「そう」の/ou/と「か」の/a/の時間長が長くなることが確認されたことから、躊躇やためらいがある発話には時間長の長い音素が出現する傾向があると考えられる。また、経験的に、伸長される音素は文字列によって異なっていると思われる。temporizerには非常に多くの文字列の発話が考えられるので、伸長される音素とそ

れが伸長しているか否かの閾値をすべての文字列の発話に対して個別に求めることは不可能である。

そこで、本稿ではテキスト音声合成システムによって予測された時間長を標準として各音素毎に標準との時間長の比を求め、その最大値を入力として判別分析によりtemporizerの識別を行なう。テキスト音声合成システムとしてはATR ν -Talk 音声合成システム[2]を用いる。ただし、ATR ν -Talkの時間長の予測値は朗読調の場合のもので、全発話の各音素ごとに実際の発話の平均値とATR ν -Talkの予測値の平均値が一致するように係数を掛けて補正を行なう。また、temporizerは躊躇やためらいの発話であり、長い発話は出現しないことが予備的検討で確認されたため、発話の長さが10モーラ以下の場合についてのみ識別を行なうこととした。

2.2 question の識別方法

我々は日常会話において文字列から明らかに疑問とわかる場合を除き、文末の基本周波数の上昇などの韻律情報や文脈の流れなどを用いて疑問か否かを識別している。また、文末に「ですか」や「ですね」を持つ発話の分析の結果から、文末の基本周波数パターンの傾きがquestionの識別に有効な要因であることも、既に確認されている[1]。一般に「です、ます」調の発話ではこのような終助詞が付くことが多いが、くだけた発話では同一の文字列を用いて、韻律の変化によって異なる発話アクトを表す傾向が見られる。

文末の韻律を表すパラメータとして以下のものを用いる。

- 文末の2音素の区間内で抽出された基本周波数の回帰直線の傾き
- 文末の1音素の時間長のz-score[3]
- 文末の1音素の区間内で抽出されたパワーの回帰直線の傾き

3 実験

提案するtemporizerおよびquestionの識別方法の有効性を確認するために識別実験を行なった。

3.1 使用データ

実験には2種類のデータを用いた。1つは自由発話データベース[4](以下、旅行会話とする)であり、もう1つは、くだけた発話の収集を目的として新たに作られたものである。後者は2人の親近度の高い話者(関東地区出身)に雑談する時のような口調で構わないと指示を与えた上で、表裏にまたがる迷路の表側のみの情報を片方の話者に、裏側のみの情報をもう一方の話者に与え、お互いに情報交換しながら迷路を解いてもらい収録した(以下、迷路会話とする)。

*Identification of Communicative Acts from Unrestricted Phrases using Prosody.
S. Fujio, N. Campbell and N. Higuchi
(ATR Interpreting Telecommunications Research Labs.)

表 1. temporizer の識別結果

正解率		Temporizer以外
Temporizer		
そうですか	84.0% (21/25) [91.2%]	78.6% (246/313)
そうですね	32.4% (11/34) [38.0%]	

(注) []内の数字は人間による文脈なしでの Temporizer 識別の聴取実験における正解率

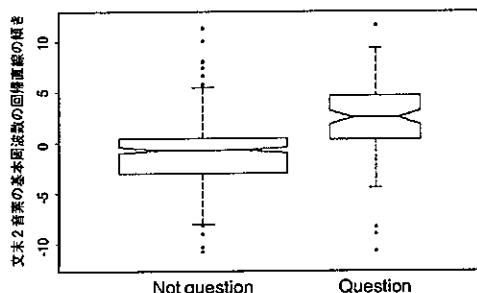


図 1. 文末の 2 音素の基本周波数の回帰直線の傾きの分布の違い

旅行会話には temporizer の発話は出現するが、話者間の親近度が低いために終助詞を伴った丁寧な表現の疑問の発話しか出現しなかった。一方、迷路会話には temporizer の発話は出現しないが、終助詞を伴った疑問だけでなく終助詞を含まない様々な疑問の発話が出現した。したがって、temporizer の識別実験には旅行会話を、question の識別実験には迷路会話をを用いた。

3.2 temporizer の識別

temporizer の識別実験の結果を表 1 に示す。「そうですね」の場合には 32.4% しか識別されていないが、人間による文脈無しでの temporizer 識別の聴取実験の正解率を考慮すると決して悪い数字とはいえない。

なお、表 1 からわかるように使用したデータベースには「そうですか」「そうですね」以外に temporizer は無かった。このため、他の文字列の temporizer への適用可能性を検証する必要がある、今後引き続きデータ収集を行なう予定である。

3.3 question の識別

まず、先に示したパラメータについて、各パラメータ単一での question の場合とそれ以外の場合の分布の違いの分析を行なった。図 1 に示すように文末の 2 音素の基本周波数の回帰直線の傾きに違いが見られたが、他の 2 つのパラメータにははっきりとした違いが見られなかった。また、文末の 2 音素の基本周波数の回帰直線の傾きを入力として判別分析で識別したところ、約 70% の question の発話を識別できた。

次に、全パラメータを用いて tree を作成し、question の識別を行なった。なお、tree の学習に際しては question のサンプル数が他に比べて少なかったため、このサンプル数の差が学習に悪影響を及ぼさないよう

表 2. question の識別結果

文末のアクセント型の区別	アクセント型	正解率	
		Question	Question以外
あり	HL	76.9% (10/13)	98.1% (154/157)
	LL or HH	85.2% (52/61)	84.8% (195/230)
	LH	82.8% (24/29)	92.3% (36/39)
	平均	83.5% (86/103)	90.4% (385/426)
なし	—	89.3% (92/103)	86.4% (368/426)

にデータに重みをつけた。この時、文末のアクセント型（文末の 1 モーラで上がる LH 型、変わらない LL 型または HH 型、下がる HL 型）の違いによる識別精度への影響を調べるために、アクセント型で分類した場合と分類しない場合の両方について識別を行なった。表 2 に示す結果のように約 80% の question の発話を識別できており、全パラメータを用いたことによる精度の向上が確認された。また、アクセント型により 3 種類に分類して識別した場合と、分類せずに識別した場合での精度の差はあまりなかった。この結果から、question の文末の基本周波数パターン等に文末のアクセント型が大きな影響を与えていないと考えられる。したがって、question の識別過程において文末のアクセント型による分類は必要ないと思われるが、これは 2 名の話者のデータでの実験結果であるので、さらに話者を増やして分析および実験する必要がある。

4 むすび

任意の文字列に適用可能な発話アクトの識別方法を検討し、temporizer および question の識別方法を提案した。また、識別実験結果から、音素の時間長の伸長度が temporizer の識別のために有効であり、これを用いることにより約 8 割の精度で識別可能であることが確認された。また、文末の基本周波数パターンの傾きが question の識別のために有効であり、これを用いることにより約 7 割の精度で識別可能で、さらに、文末の音素の時間長および文末のパワーの傾きを入力として加えることにより、識別の精度が約 8 割に向上することが確認された。しかし、temporizer の識別については「そうですか」と「そうですね」以外の文字列の発話に対する有効性の検証をする必要がある、question の識別については使用データの話者数を増やした実験を行なう必要がある。

参考文献

- [1] 藤尾, キャンベル, 樋口: “韻律を用いた発話タイプの識別”, 音響学講論, 1-1-1(1995,9)
- [2] 匂坂, 海木, 岩橋, 三村: “ATR ν -Talk 音声合成システム,” 情報処理学会「音声言語情報処理と音声入出力装置」研究グループ, 電子情報通信学会「音声認識の実用化をめざす新手法」時限研究専門委員会 研究会資料(1992-10).
- [3] W.N.Campbell and S.D.Isard: “Segment durations in a syllable frame”, Journal of Phonetics vol.19, pp.37-47, 1991
- [4] T.Morimoto et al.: “A speech and language database for speech translation research”, Proc. ICSP94, vol.IV, pp.1791-1794(1994)