表現豊かな発話様式の韻律的特徴の分析 *

石井カルロス寿憲 & ニック・キャンベル（JST/CREST at ATR/HIS Lab.）

1. はじめに

人間と機械との音声を介した情報伝達においては、言葉のレベルを超えて、発話様式による表現の違いも考慮する必要がある。

本研究は文字が伝達する意味以上に、話し手が発話に含まれたイントネーションや声の質等をいかに制御してさまざまなレベルの感情、意図や態度を表現しているのかを調べ、音声合成・認識に生かすことが狙いである。そのため、音声信号から抽出可能な音響・韻律的特徴と人間が感じた要素との関係を調べる必要がある。

「喜怒哀楽」を強制的に発声した感情音声を分析した研究は多く、講演などのモノローグの自発音声を分析した研究[1]も報告されている。本研究では、日常会話の自然発話の音声データを基に、さまざまな発話様式のカテゴリー化とそれらに関する音響・韻律的特徴の分析を行った。

2. データと分析単位

データとしては、CREST/ESP プロジェクト[2]上で録音されている自然発話音声データを用いた。発話単位としては、韻律句を扱うことにした。韻律句の区切りとしてはあきらかなポーズ、または明らかなピッチの立ち上がりが知覚される場合に半自動・手動で行った。日本人話者一人における自然日常会話（親との会話、友達のと会話、会社への電話、子供との会話等）を含めた 670 個の韻律句を分析対象とした。

3. 発話様式におけるカテゴリー化・ラベリング

発話様式の分類化として、次のような要素を提案した。

- 話者の気分・感情状態： N（ニュートラル）、N（暗）、N（楽）、心配、楽しい、可笑しがっている、不満、不機嫌な、怒り、怪訝な、悲しい、がっかり、等
- 相手に対する話者の態度： [丁寧さ、暖かさ、気の使い方、優しさ]の度合いをそれぞれ 5 段階のスケール
- 内容に対する話者の態度： [自信度、積極性、興味度]をそれぞれ 5 段階スケール
- 声の質：

エネルギー（声の強さ）：沈んだ・低い・抑えた・普通・活発・高い・興奮した・全開・爆発的
機嫌の度合い（声の明るさ）：楽しくない・あまり・普通・少し機嫌がいい・機嫌がいい
硬柔（声帯の緊張度合い）：5 段階スケール

ラベリングの際は、話者の本来の感情は別として、各発話から感じ取られる要素をラベルするよう指示した。感情状態においては、聞き手の印象を適切に表現できるよう自由に単語を選択することを許した。ラベリング作業は母語話者一名が行い、ラベリングに疑問をもったサンプルは研究員 3 名と共に議論して決めたものである。

* Analysis of prosodic features on Expressive Speech by Carlos Toshinori Ishi & Nick Campbell (JST/CREST at ATR/HIS Lab.)

4. 音響・韻律的特徴

日本語はピッチアクセント言語であり、フレーズを構成している単語のアクセント型によって、主に F0 の動きが大きく変化するので、今回はまず声の全体的な特徴を表す音響・韻律的特徴に注目し、次のものを提案した。

- 声の全体的な高さに関わる平均 F0 値（***f0avg***）、及び F0 の幅（全体的な動き）に関わる最大 F0 値と平均 F0 値との差分（***f0dif***）。F0 の幅としては、最大値と最低値を用いることが多いが、最低値はあまりロバストでないため、平均値を基準として扱うこととした。
- 声の全体的な強さに関係するパワーに関しても F0 と同様に ***rmsavg*** と ***rmsdif***。
- 声の柔らかさに関する AQ（*amplitude quocient*）パラメータの平均値（***aqavg***）。AQ パラメータとは、フォルマントの影響を除いた声帯音源波形の peak-to-peak 値と、その微分波形の最大 negative peak の比として定義され、声の硬さ・柔らかさに関わることが報告されている[3]。
- 声道の形（主に声道長）に関わる第 3 と第 4 フォルマントの平均値（***f3avg***, ***f4avg***）。フォルマント情報は[4]で提案されたケプストル・フォルマントマッピング方法を使用した。第 1 と第 2 フォルマントは韻律よりも母音の種類に関わる要因が強いという理由で扱わないことにした。

平均値の計算に関しては、すべてのパラメータにおいて、母音の区間の値のみを使用した。

5. 分析結果

5.1. 発話様式における要素の分析結果

図 1 は現時点で分析した韻律句における各感情状態の頻度を示している。電話会話が多かったので感情的にはニュートラル(N)なものが多かった。

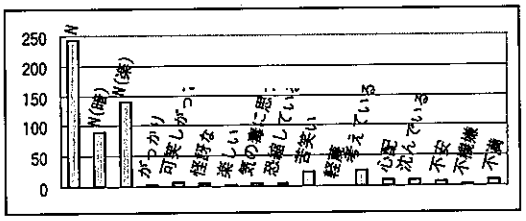

図 1. 各感情状態の頻度

図 2 では丁寧度において頻度の多かった感情状態の分布を示している。この分布から「N（暗）、沈・不満・不機嫌、苦笑い」は「だらけた・カジュアル」に、「心配・不安」は「カジュアル・少し丁寧」にデータが傾き、「丁寧・改まった」は「N・N（楽）」の要素のみが見られる結果となった。

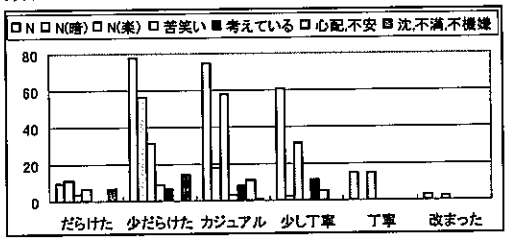

図 2．丁寧度における各感情状態の分布

第3課で紹介したラベルのうち、定量化された要素の主成分分析を行い、その結果を図1に示す。表1は要素同士の相関係数を示している。

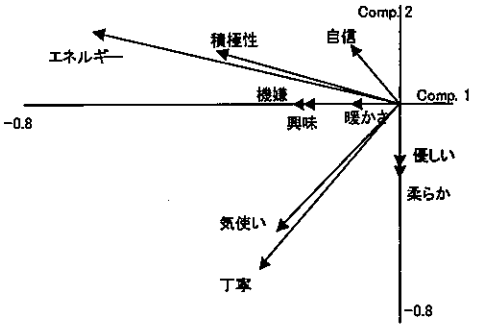

図1．発話様式における要素の主成分分析結果

表1．発話様式における要素同士の相関係数

	丁寧	暖か	優し	気使	興味	積極	自信	エネ	硬柔
機嫌	0.18	0.28	-0.02	0.18	0.29	0.39	0.26	**0.46**	0.12
丁寧		0.13	0.28	**0.63**	0.20	0.20	-0.07	0.28	-0.17
暖かさ			0.01	0.14	0.24	0.11	-0.01	0.24	-0.21
優しさ				0.24	0.01	-0.05	-0.22	-0.12	-0.28
気使い					0.33	0.21	-0.19	0.29	-0.12
興味						**0.42**	0.17	**0.50**	-0.02
積極性							0.41	**0.72**	0.31
自信								0.34	0.26
エネルギー									0.29

図1と表1から、要素同士の関係が導ける。主にエネルギー（活動性）と積極性の間、そして丁寧度と積極性の間の相関が大きいことが見られる。これらのデータから一貫したラベリングが行われたといえ、それぞれ意味を持つ要素といえる。

5.2. 音響･韻律的特徴の分析

表2．音響･韻律的特徴同士の相関係数

	f0dif	rmsavg	rmsdif	f3avg	f4avg	aqavg
f0avg	0.03	**0.56**	0.11	0.06	0.14	**0.51**
f0dif		0.26	0.32	-0.23	0.01	0.13
rmsavg			0.17	-0.24	0.06	**0.51**
rmsdif				-0.26	0.03	0.14
f3avg					0.12	-0.23
f4avg						0.04

表2の結果から、平均F0と平均パワーの相関が見られ、高い声は強く発声される傾向を示している。同様に、平均AQに関しても、F0とパワーとの相関が見られ、硬い声は高くて強い声で起こりやすいことを示している。

また、他の特徴同士の相関が小さいということから、各特徴が異なった情報をもたらしているといえる。

5.3. 発話様式における要素と音響･韻律的特徴との関連

ここでは発話様式において定量化した要素がどの程度音響･韻律的特徴と結びついているのかを調べた。図3は声の「エネルギー、機嫌、硬柔」において、音響･韻律的特徴に対するそれぞれの主成分分析の第1と第2成分を示している。

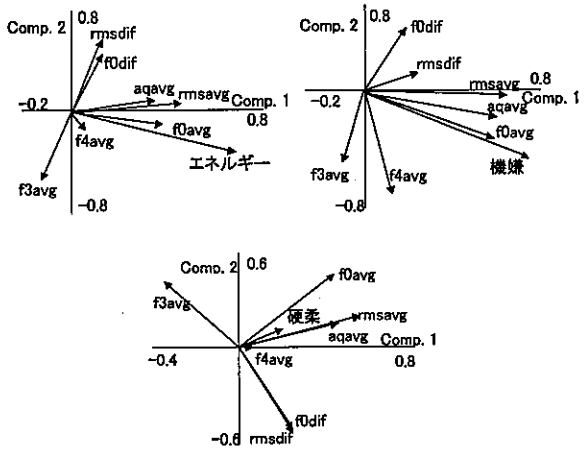

図3．「エネルギー、機嫌、硬柔」と音響･韻律的特徴の主成分分析結果

図3の結果から、「エネルギー」においては、*rmsavg* よりも、*f0avg* との相関が大きいことが見られる。「機嫌」においても、「エネルギー」と似たような傾向を示しているが、f4avgの影響がより大きいことが見られる。これは、機嫌がいい時は、笑顔になることが多く、唇の広がりによって声道長が短くなり、全体的にフォルマントが高くなり、*f4avg* との関連が強くなっていると解釈できる。*f3avg* の場合は、どの要素とも相関があまり見えず、音素タイプの影響が大きいのではないかと考えられる。「硬柔」においては、今回評価した音響的特徴とは相関があまり強くないが、*rmsavg* と *aqavg* との関連が示され、硬い声は、パワーが大きく、AQパラメータも大きいという傾向を示している。

6. おわりに

本研究では日常会話の自然発話音声を聞き手の印象によってカテゴリー化したものと全体的な要素を表す音響･韻律的特徴の関連を調べた。今後は、発話速度も含め、各カテゴリーの判断において人間が注目している韻律特徴の動きを考慮して、動的パラメータを提案し、評価する予定である。

謝辞

発話様式ラベリングに貢献している木村美名子氏に感謝する。また、AQパラメータ及びフォルマント情報を提供して頂いたパーハム･モクタリ氏に感謝する。

参考文献

[1] 篭宮、槙、菊池、前川「自発音声コーパスにおける印象評定とその要因」日本音響学会秋季2001年, Vol. I, 381-382. (2001)

[2] The JST/CREST Expressive Speech Processing project, introductory web pages at: www.isd.atr.co.jp/esp

[3] Mokhtari, P. & Campbell, N. "Perceptual validation of a voice quality parameter AQ automatically measured in acoustic islands of reliability",日本音響学会春季2002年, Vol. I, 401-402. (2002)

[4] Broad, D.J. & Clermont, F. "Formant estimation by linear transformation of the LPC cepstrum," *J.Acoust.Soc.Am.* 86(5), 2013-2017. (1989)