5T-6

対話音声から受ける4種の印象と分節的特徴との関係の分析

上野 吉弘 †　政倉 祐子 ††　大野 澄雄 ††

† 東京工科大学大学院バイオ・情報メディア研究科　†† 東京工科大学コンピュータサイエンス学部

1　はじめに

近年、人と機械が接することが多くなり音声技術でのコミュニケーションの研究が盛んになってきている。

自然なコミュニケーションを行う上で重要な、印象や感情の認識を行うためには韻律的特徴に加え、分節的特徴が重要な役割を果たしていると考えられる。また、発話内容を正確に伝えるためには、言語情報以外に場面や話者の状態、感情といった感性情報 (以下発話印象) も重要であることがわかっている [1]。

そこで本研究では、発話印象の認識を目指し、発話印象を考慮した韻律的特徴、及び分節的特徴の分析を行い、4種の印象の認識モデルを作成することを目的とする。また、分節的特徴の有無により認識の精度に変化があるのかを比較し、分節的特徴の有用性を調べる。

そのためにまず録音した自然発話にSD法を用い、4つの評価軸に対して評価実験を行った。その結果とそれぞれの特徴量との関係を求め、重回帰分析を行った。

2　音声コーパス

音声はMister Oという絵本を用い、1対1の対話方式で録音を行った [2]。Mister Oは台詞のない絵が60コマ並んでいる外国の作者の絵本である。絵本の中の1つの話から間の24コマを抜き出しランダムに並べる。正しい順番を二人で話し合い、その様子を録音した。また録音はSkypeによる会話をTapurというソフトを用いてサンプリング周波数16kHz、量子化ビット数16bitで行った。

話者は男性2名、女性2名の計4名の発話数600発話の音声をそれぞれ収録した。

3　研究内容

3.1　発話印象の評価実験

上記のコーパスに対し、評価実験を行った。評価方法はSD法を用い、対立する印象に対して7段階で評価を行った。評価は1つの音声に対してそれぞれの軸について判断してもらった。また被験者数は8名である。

図1は評価実験で用いた印象軸である。本実験で用いた軸は予備実験を元に選定した図1の4つに対し評価を行っている [3]。

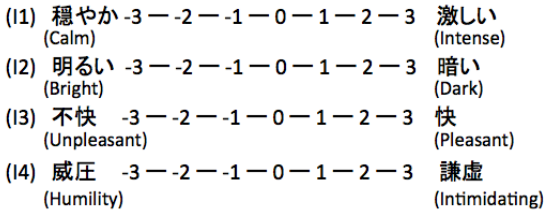

図 1: 対象とした印象軸

3.2　Evaluator Weighted Estimator(EWE)

評価実験によって得られた結果に対して、EWE[4]を用いて重み付けを行った。EWEを用いることで評価者ごとのばらつきを小さくする、これによりデータの信頼性を向上することができる。

3.3　音響特徴

本実験で対象とする音声の特徴量を表1に挙げる。特徴量はそれぞれの発話単位の音声から抽出している。F_0 と Power に関しては、一つの音声発話内の最小値として10%、最大値として90%の値を用いている。また、MFCC1、第1、第2フォルマントはそれぞれの音声に含まれる母音における最小値、最大値、平均値である。

表 1: 対象とした特徴量

韻律的特徴	F_0	10%	F_0 の 10%値
		90%	$F_0$90%値
		平均	F_0 の平均値
		標準偏差	F_0 の標準偏差
	Power	10%	Power の 10%値
		90%	Power の 90%値
		平均	Power の平均値
		標準偏差	Power の標準偏差
分節的特徴	MFCC1 (第1ケプストラム係数)	最小値	MFCC1 の最小値
		最大値	MFCC1 の最大値
		平均	MFCC1 の平均
	第1，第2フォルマント (母音)	最小値	フォルマントの最小値
		最大値	フォルマントの最大値
		平均	フォルマントの平均

Analysis of the relationship between segmental features and four impressions received from the dialogue speech

† Yoshihiro UENO(Graduate School of Bionics,Computer and Media Sciences, Tokyo University of Technolosy)

†† Yuko MASAKURA(Tokyo University of Technolosy)

†† Sumio OHNO(Tokyo University of Technolosy)

3.4 発話印象と音響特徴

発話印象と各音響特徴の関係を調べた。図2は4つの印象軸ごとに F_0 の最小値、平均値、最大値の関係を示している。また、図中の折れ線は ±0.5 の範囲にあるデータの中央値を結んだものである。

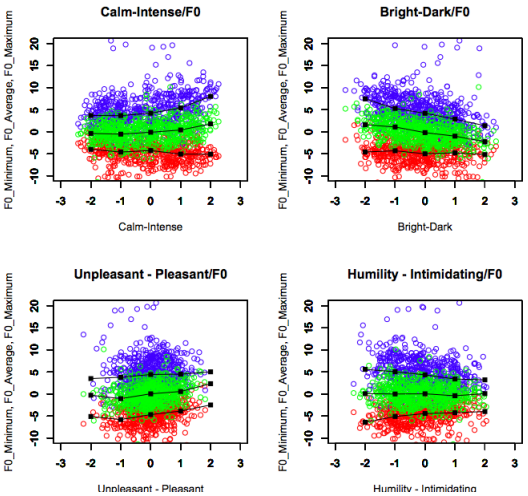

図 2: 発話印象と F_0 最小値、平均値、最大値の関係

図2を見ると、印象軸1では「激しい」に近づくにつれ、F_0 の最大値が上がり、かつ最小値が下がっていることがわかる。また印象軸3を見ると「快」が大きくなるほど最小値が下がっている。

4 重回帰分析によるモデル作成

表1の各特徴量を説明変数として重回帰分析を行い、韻律的特徴のみの場合、分節的特徴のみの場合、両方を考慮した場合について、それぞれモデルを作成した。その際、説明変数同士の独立性を保証するため多重共線性が認められる変数を除外した。

予測モデルの評価は決定係数 R^2 と 実測値 − 予測値 の残差 RMS を用いた。

図3は全特徴を用いた重回帰分析による実測値と予測値の関係を示している。「穏やか-激しい」、「明るい-暗い」の軸において高い精度があることがわかる。これより表2の、I1、I2 の R^2 値は I3、I4 に比べて高い値を示していると考える。

また表2より、4軸全てにおいて韻律的特徴のみの場合、分節的特徴のみの場合に比べ、全特徴を用いた場合の残差の RMS が小さくなっている。このことから、分節的特徴を用いることにより、印象推定の精度が向上することが言える。

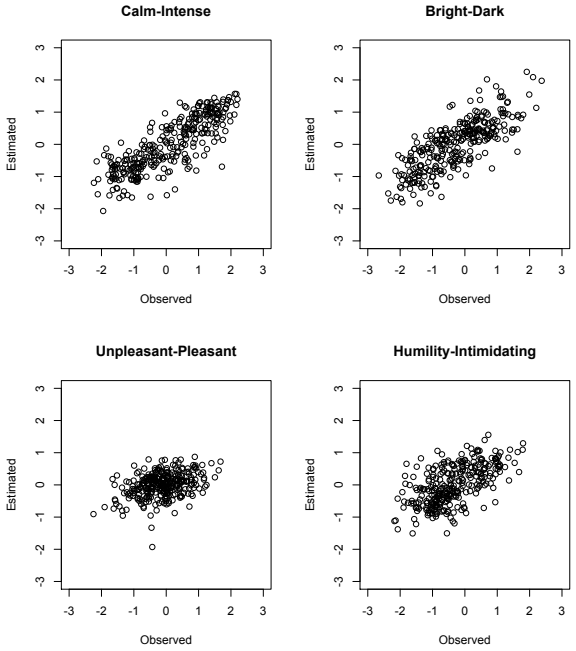

図 3: 重回帰分析による実測値と予測値

表 2: 決定係数 (R^2) と実測値と予測値の残差 RMS

韻律的特徴				
	I1	I2	I3	I4
決定係数 R^2	0.387	0.257	0.010	0.091
残差 RMS	0.697	0.706	0.622	0.684
分節的特徴				
	I1	I2	I3	I4
決定係数 R^2	0.269	0.184	0.002	0.067
残差 RMS	0.782	0.762	0.641	0.704
韻律的特徴及び分節的特徴				
	I1	I2	I3	I4
決定係数 R^2	0.444	0.351	0.016	0.121
残差 RMS	0.649	0.639	0.610	0.654

参考文献

[1] 石井 カルロス寿憲, 石黒 浩, 萩田 紀博, “韻律および声質を表現した音響特徴と対話音声におけるパラ言語情報の知覚と関連”, 情報処理学会論文誌, Vol.47, No.6, pp.1782-1792, 2006

[2] L.Trandheim, “Mister O”, 講談社, 2003

[3] 上野 吉弘, 政倉 祐子, 大野 澄雄, “種々の発話印象を表現する音声合成のための音響的特徴量の検討”, 情報処理学会, 第 74 回全国大会講演論文集, Vol.2, pp.599-600, 2012

[4] G.Michael, K.Kristian, “Evaluation of natural emotions using self Assessment manikins”, IEEE Workshop on Automatic Speech Recognition and Understanding, pp.381-385, 2005