E-015

音声に含まれる感情量の定量化のための検討
A study on quantitation of emotional intensity of speech

川村 剛 † 松澤 直之 † 政倉 裕子 ‡ 大野 澄雄 ‡
Tsuyoshi Kawamura Naoyuki Matsuzawa Yuko Masakura Sumio Ohno

1. はじめに

感情認識において、感情音声の感情量を人間の主観評価によって定量化するのが一般的である。近年では、高次の感情 (プルチックなど) のほか、音声から知覚される話者の感情状態 (以降、印象) の定量化も行われる [1][2]。ラベリングには尺度法や SAM などの手法がある [1][4]。森らは話者の感情状態に加え、話者の対人関係や態度を表す 6 軸を用いて定量化している [2]。これらは一般的によく用いられる PAD(Pleasure-Arousal-Dominance) スケールと共通点が多い。

本報告では尺度法を用いて自然発話の印象を評価する場合において、森らの 6 軸（以降、印象軸）と PAD スケール（以降、PAD）の関係について主成分分析を用いて分析した。その結果、2 個の主成分が得られ、第 1 主成分は「積極的-消極的」、第 2 主成分は「覚醒-睡眠 (快-不快)」であった。第 2 主成分までの累積寄与率は 87.7%であった。また各主成分と音響的特徴量との関係を調べるため、重回帰分析を用いて推定するモデルを作成して結果を述べる。

2. 音声資料

本実験では大学生男女 4 人でゲームをした際に交わされた自然発話を収録した。前後 100ms の無音区間を有する発話を切り出しランダムに 560 発話を抽出してラベリングを行った。うち始めと終りに配置した計 60 発話はダミーとし以降の分析には用いていない。なお、音声の形式は PCM フォーマット、サンプリング周波数 16kHz、量子化ビット 16bit である。

次に各発話に対してラベリングおよび音響的特徴量の抽出を行った。

2.1. ラベリング

得られた自然発話は森らの印象軸 (話者の感情状態:「快$^+$ - 不快$^-$(I_1)」「覚醒$^+$ - 睡眠$^-$(I_2)」、対人関係:「支配$^+$ - 服従$^-$(I_3)」「信頼$^+$ - 不信$^-$(I_4)」、態度:「関心$^+$ - 無関心$^-$(I_5)」「肯定$^+$ - 否定$^-$(I_6)」) についてラベリングを行った。対話において話者自身の感情を観測することは不可能であるため、本実験では話者の音声を聞いた聞き手が受けた印象を印象値としてラベリングした。ラベリングは 7 段階（例えば、「非常に不快 (-3)」「かなり不快 (-2)」「やや不快 (-1)」「どちらでもない (0)」「やや快 (1)」「かなり快 (2)」「非常に快 (3)」）で、大学生男女 12 名にラベリングしてもらった。

2.2.EWE による評価値の統合

ラベリングで得られた各評価者の評価値を EWE 法を用いて統合した [4]。評価者ごとに重み付けをするため、EWE 法を用いた方が評価者間の評価のばらつきによる影響を抑え、推定精度を向上させることができる。推定モデル構築をする場合、単なる平均値とした場合より高い精度が得られることは既に報告されている。

従って、以降の分析で用いる評価値は全てラベリングで得られた値を EWE 法によって統合した値とする。

2.3. 音響的特徴量抽出

音声から表 1 のような言語構造に依存しない音響的特徴量を抽出する [3]。$F0, P, C1$ はそれぞれ声の高さ、声の強さ、声質を表現する特徴量であり言語に依存しない。特徴量抽出には MATLAB を使用した。

表 1: 音声から抽出した音響的特徴量

特徴量	説明
$F0_{mean}$	個人差正規化 $F0$ の発話内平均
$F0_{max}$	個人差正規化 $F0$ の発話内最大値
$F0_{min}$	個人差正規化 $F0$ の発話内最小値
$F0_{stdv}$	$F0$ の発話内標準偏差
P_{mean}	短時間平均パワーの発話内平均
P_{max}	短時間平均パワーの発話内最大値
P_{stdv}	短時間平均パワーの発話内標準偏差
P_{magn}	短時間パワーの変動量
$C1_{mean}$	第 1 次ケプストラム係数の発話内平均
$C1_{max}$	第 1 次ケプストラム係数の発話内最大値
$C1_{min}$	第 1 次ケプストラム係数の発話内最小値
$C1_{stdv}$	第 1 次ケプストラム係数の発話内標準偏差

表 2: 印象軸の相関係数

	I_1	I_2	I_3	I_4	I_5	I_6
I_1	-					
I_2	0.68	-				
I_3	0.25	0.63	-			
I_4	**0.89**	0.55	0.18	-		
I_5	0.68	**0.79**	0.49	0.61	-	
I_6	**0.85**	0.46	0.04	**0.89**	0.55	-

3. 感情状態の評価軸に関する検討

3.1. 印象軸の主成分分析

表 2 に印象軸間の相関係数を示した。I_3 を除いて中程度から 0.8〜0.9 前後の高い相関があることがわかった。これにより各軸に影響を及ぼす共通因子があることが推測できる。

従って、印象軸と PAD との関係性を検討するため、印象軸に対し主成分分析 (相関行列法) を行った。図 1 は固有値を示しており、カイザー基準により主成分を 2 個とした。表 3 は得られた主成分負荷量、寄与率、累積寄与率を示している。主成分負荷量が 0.4 以上の項

†東京工科大学大学院, Graduate School, Tokyo University of Technology
‡東京工科大学, Tokyo University of Technology

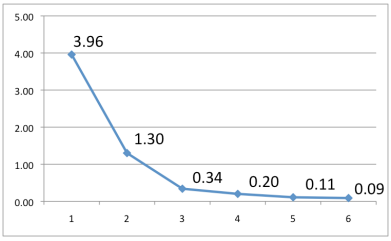

図 1: 印象軸の固有値

表 3: 主成分負荷量、寄与率 (%)、累積寄与率 (%)

	第 1 主成分	第 2 主成分
I_1	**-.930**	.233
I_2	**-.833**	**-.414**
I_3	**-.461**	**-.804**
I_4	**-.885**	.359
I_5	**-.848**	-.259
I_6	**-.827**	**.485**
寄与率 (%)	66.0	21.7
累積寄与率 (%)	66.0	**87.7**

目に着目した。

表 3 から第 1 主成分には全ての項目が含まれ、かつ I_3(-.461) を除き高い負荷量である。逆に、第 2 主成分は I_3(-.804) でもっとも高いほか比較的低い。第 2 主成分までの累積寄与率が 87.7%となっていることから十分な再現性があると言える。

3.2. 主成分解釈

図 2 に表 3 の主成分負荷量をもとに、第 1 主成分を x 軸、第 2 主成分を y 軸とし I_1〜I_6 をプロットしたものを示す。図 2 で I_1, I_2, I_3 の破線に注目すると I_1 と I_2 が主に第 1 主成分に影響されていることから「Pleasure/Arousal に相当する次元」であると考える。第 2 主成分は I_3 に対する負荷量 (-.804) から「Dominance に相当する次元」であると考える。従って、森らの印象軸を主成分分析した結果 PAD に相当する軸が観測された。

上記の結果から、音声に含まれる印象量を定量化する軸としては PAD で十分な効果が得られると言える。

3.3. 音響的特徴量との関係性

3.2 で得られた主成分と音響的特徴量間の関係を調べるために、それぞれの主成分得点を従属変数、音響的特徴量を説明変数として重回帰分析 (AIC を用いた変数選択) を行った。得られた標準化偏回帰係数を図 3 に示す。第 1 主成分の決定係数は 0.49、残差の RMS は 1.41、第 2 主成分の決定係数は 0.20、残差の RMS は 1.01 であった。

図 3 から第 1 主成分は特に $C1$ の影響を強く受けているほか、$F0_{mean}$, P_{mean}, P_{max} が影響している。モデルの決定係数は 0.49 と中程度の適合度を示しているが残差がやや大きい。一方、第 2 主成分はモデルは決定係数が 0.20 と精度が悪い。$C1$ 全体と $F0_{mean}$, P_{magn} の影響があるがどれも強い影響とは言えない。これは有効な特徴量がないためだと考えられる。

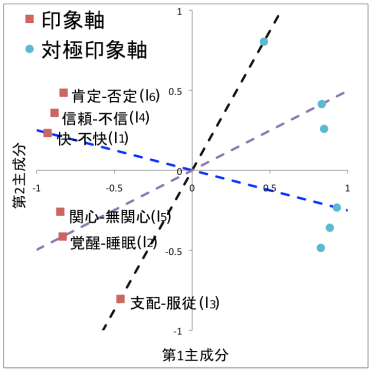

図 2: 主成分分布 (x:第 1 主成分, y:第 2 主成分)

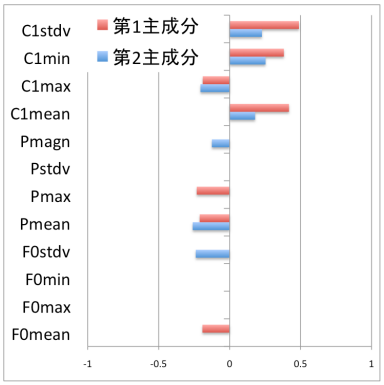

図 3: 標準化偏回帰係数

4. おわりに

本報告では森らの印象軸が PAD と多くの共通点があるため、主成分分析を用いて検証した。その結果得られた成分と PAD との関連性は高いということが分かった。また、重回帰分析を用いて音響的特徴量との関係性を調べた結果、第 1 主成分 (Pleasure/Arousal に相当する次元) は $C1$ の影響を受けやすいということが分かった。一方で、第 2 主成分 (Dominance に相当する次元) については今回用いた音響的特徴量だけでの推定は難しい結果となった。推定に有効な特徴量に関しては今後の検討課題である。

参考文献

[1] Grimm, M., Kroschel, K., Mower, E., Narayanan, S., ”Primitives-based evaluation and estimation of emotions in speech”, Speech Communication Volume 49, Pages 787-800, 2007.

[2] Mori, H., Satake, T., Nakamura, M., and Kasuya, H. ”Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics”, Speech Communication Volume 53, Issue 1, Pages 36-50, 2011.

[3] Arimoto, Y., Ohno, S., and Iida, H. ”Acoustic Features of Anger Utterances during Natural Dialog.”, in Proc. of Interspeech2007, pp.2217-2220, 2007.

[4] Grimm, M., Kroshel, K. ”Evaluation of natural emotions using self assessment manikins”, Automatic Speech Recognition and Understanding, 2005 IEEE Workshop, 381, 2005.