

音響特徴量と声の印象に関する分析*

○高椋琴美, 保田千津子, 谷田泰郎 (シナジーマーケティング株式会社)

1 はじめに

音声認識によって音声を文字情報に変換することが可能になり, 音声合成によって文字情報も音声化できるようになった。一方, 音声には言葉以外にも感情や意図といった非言語情報も含まれており, コミュニケーションにおいて重要な要素である。発信側は自らの価値観を基準として言葉や表現方法を選択し, 受け手側は言葉と合わせて相手の表現を印象という形で受取り, 意図を理解しようとする。

感情については, 感情を音声合成に反映させる研究や, 音声から相手の感情を読み取る試みが盛んに行われているが, 声の印象についての研究はまだあまり行われていない。発話時の声の特徴によって, 相手への情報の伝わり方は大きく異なる。声の印象はコミュニケーションにおいて重要な要素の1つである。

音響特徴量から声の印象が推定できれば, 印象診断や印象トレーニングへの応用や, CMなどのキャスティング時の音声選択基準としても利用できる。また音声合成に声の印象を反映することが可能になると考えられる。

一方で, 自分をどう見せたいか, 伝えたい内容をどう表現するかは個人の価値観に起因する。例えば相手との信頼関係が重要と考える価値観を持つ場合, 誠実な印象を持たれる様に話をする。つまり, 価値観と声の印象には関係があり, 同じく価値観と音響特徴量にも関係があると考えられる。

音響特徴量から価値観が推定できれば, 我々が行っている個人の価値観をマーケティング活動に活用する試み[1][2][3][4][5]において, 音声で個人の価値観を推定する手段の1つとして利用可能となる。

そこで本稿では, (1) 音響特徴量と声の印象の関係 (2) 音響特徴量と価値観の関係について調査を行った結果を報告する。

2 提案モデル

1 で述べた価値観・音響特徴量・声の印象

の関係より Fig. 1 のモデルを提案する。音響特徴量と声の印象の関係については, 声の高さの様に価値観を経由せず関係するものがあると考え直接にも関係付けている。

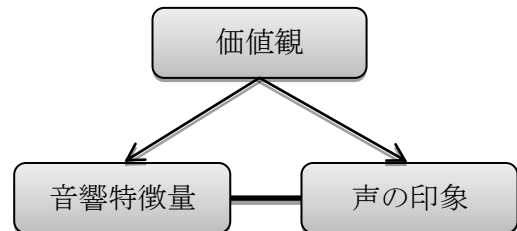


Fig. 1 価値観-音響特徴量-声の印象のモデル

3 音声録音と声の印象調査

3.1 音声録音

当社社員 20 代~50 代の計 138 人 (男性 88 人/女性 50 人) の音声データを WAV 形式 (サンプリングレート: 48kHz/ビット数: 16bit/チャンネル数: モノラル) で録音した。同条件になるように, 音声提供者は静かな会議室で, 椅子に座り, マイクを口から約 30cm 離れた机の上に置いて録音を行った。セリフを間違えた場合は再録音した。

セリフ選定のため, 予備調査として 20 人に対して 7 種類のセリフを録音し, その中から印象と音響特徴量 (基本周波数平均, 基本周波数の分散, エネルギー平均など) にある程度バラつきのある 3 種類を選択した。セリフの長さはいずれも 4~7 秒程度とした。

3.2 声の印象調査

録音した 3 種類のセリフのうち最も印象にばらつきがみられた 1 つを使って声の印象調査を行った。

設問は事前に検討した 32 項目の複数選択式である。音声提供者と面識のない評価者 11 人 (男性 6 人/女性 5 人) を会議室に集め, スピーカーから 1 音声 4~5 回繰り返して流し, アンケートへの回答がほぼ終わった段階で次の音声に移った。再生順は印象を相対評価させないよう, 男性と女性の声をなるべく交互

* Analysis of correlation between voice features, voice impressions and personal values, by TAKAMUKU, Kotomi, YASUDA, Chizuko and TANIDA, Yasuo (Synergy Marketing, Inc.)

になるよう工夫した。

なおアンケートには直感で回答するよう事前説明を行った。

3.3 使用する音響特徴量の決定

音響特徴量は VoiceSauce[6]により取得した。音響特徴量と声の印象の相関分析により、声の印象と相関のみられた次の6つの音響特徴量を採用した (Table 1)。各音響特徴量は各々値の小さい方から離散化し分析に用いた。

Table 1 音響特徴量と離散化分割数

音響特徴量	分割数
基本周波数平均 (avgsF0)	5
基本周波数の標準偏差の絶対値平均 (sdF0)	4
エネルギー平均 (avgSPL)	5
Cepstral Peak Prominence 平均 [7] (avgCPP)	2
倍音ノイズ比率平均 0- 500 Hz [8] (avgHNR05)	3
倍音ノイズ比率平均 0-3500 Hz [8] (avgHNR35)	3

3.4 声の印象の主成分分析

3.2 で得られた声の印象については回答を主成分分析し、固有値が 1 以上のものを Promax 回転して、9 個の声の印象成分 (ハキハキ/丁寧で誠実/若く可愛い/不真面目で雑/やさしく上品/かっこいい/冷たい/年老いた/個性的) を得た。また印象成分の主成分得点を、0 以下を「なし」、0 より上を「あり」として 2 値に離散化し分析に用いた。

4 音声提供者の価値観調査

4.1 価値観の定義

我々は、マーケティング活動における消費者の価値観を明らかにするために、1 万 1 千人に対し価値観調査を実施した。その結果 Table 2 の価値観成分 62 個を得た。(詳しい手法については[1][2]参照)。

Table 2 価値観成分

価値観フレーム	成分数	成分の内容 (ネーミング)
性格	11	好奇心旺盛 デリケート マイペース 協調型 勤勉 上昇志向 短気 正義感 ルーズ・不精 無気力 文系的
ポジティブ	8	自己愛 自己実現 アウトドア スポーツ 恋愛 ギャンブル ひとり時間

ネガティブ	3	否定・批判 非常識 期待外れ
家族	7	結婚願望 不仲 責任感 (主婦軸) 責任感 (扶養軸) 良好 (別居家族) 不十分 良好 (同居家族)
仕事観	6	満足 ストレス プライベート重視 キャリアアップ転職願望 堅実 社会的意義
交友関係	8	ストレス 親友中心 ネットワーク重視 社交的 大人数派 消極的 (独身) 仕事人脈中心 ノンストレス
時間	11	ゆとり 余裕がない 充実 仲間優先 家族優先 趣味優先 インドア派 アウトドア派 家事分担 退屈 自己投資
金銭観	8	ギリギリ ゆとり 貯蓄志向 家族優先 慎重派 自己投資 堅実生活 常識的

本実験では、消費者に対して最も重要と判断した 22 個の価値観成分に絞り、この 22 個の価値観成分を導き出すのに有効な小セットの質問項目を用意して定量調査を行った。使用した価値観は Table 2 に太字で示す。

4.2 価値観調査

当社社員 20 代~50 代の計 198 人に対し価値観調査を行った。価値観成分は主成分得点を、0 以下を「なし」、0 より上を「あり」として 2 値に離散化して分析に用いた。

5 分析

ここでは、3 と 4 の両方でデータを得ることができた 123 人を対象に分析を行った。分析では男女を区別していない。

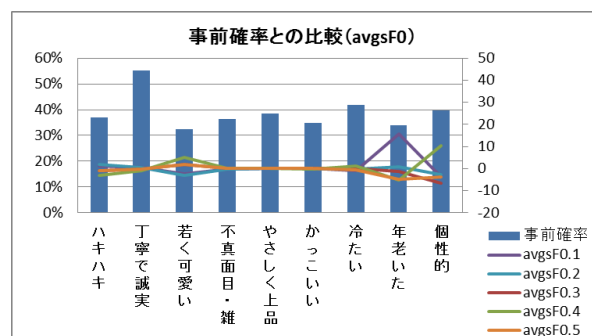
5.1 確率モデルの作成

ベイジアンネットワークによる確率モデル Fig. 1 (構造探索の制約条件は、価値観成分が親になることを許容。それ以外はフリー探索) の構築を行った。

5.2 音響特徴量と声の印象の感度分析

音響特徴量と声の印象の関係について 5.1 を用いて以下を計算し感度分析を行った。

- ① 声の印象の事前確率 (棒グラフ: 左軸)
 - ② 音響特徴量の特定ランクの値のみを与えた場合の、声の印象の事後確率
 - ③ 感度 = ② - ① (折れ線グラフ: 右軸)
- 結果を Fig. 2 に示す。



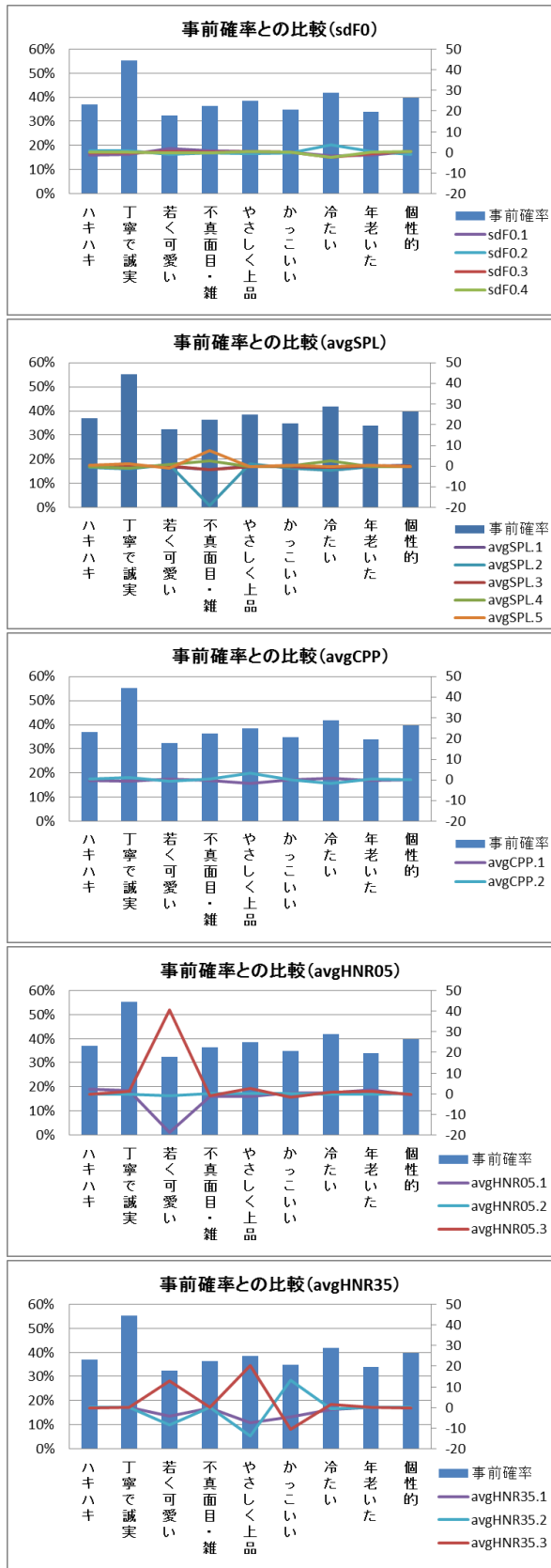


Fig. 2 音響特徴量と声の印象の感度分析

Fig. 2 より、それぞれの音響特徴量で特定の印象に反応が見られた (Table 3 に示す)。特に avgHNR05, avgHNR35 は反応が強く、中でも avgHNR35 は複数の印象に対し反応が見られた。

Table 3 音響特徴量と声の印象の分析結果

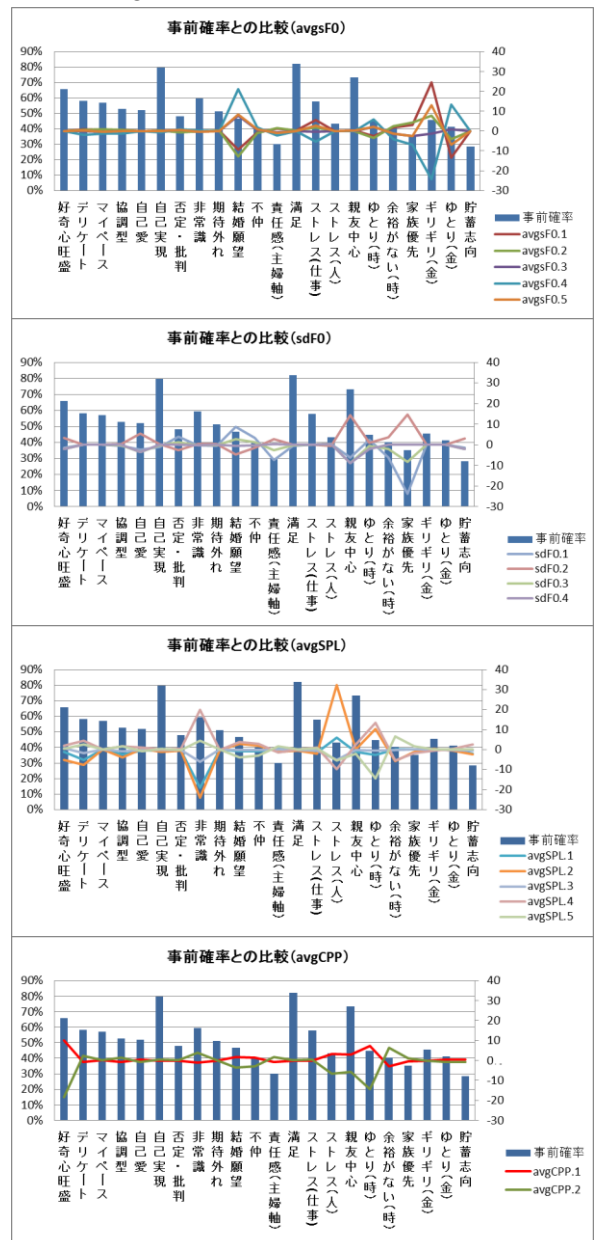
音響特徴量	関係のある印象
avgsF0	若く可愛い 年老いた
sdF0	冷たい
avgSPL	不真面目・雑
avgCPP	やさしく上品
avgHNR05	若く可愛い
avgHNR35	若く可愛い やさしく上品 かっこいい

5.3 音響特徴量と価値観の感度分析

音響特徴量と価値観の関係について 5.1 を用いて以下を計算し感度分析を行った。

- ① 価値観の事前確率 (棒グラフ: 左軸)
- ② 音響特徴量の特定ランクの値のみを与えた場合の、価値観の事後確率
- ③ 感度 = ② - ① (折れ線グラフ: 右軸)

結果を Fig. 3 に示す。



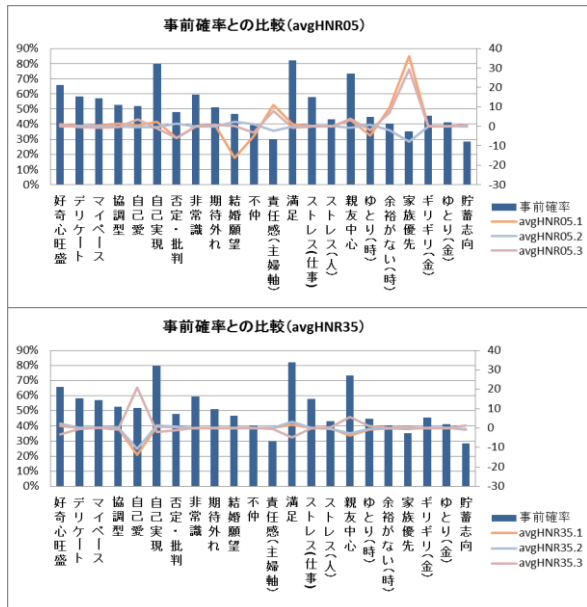


Fig. 3 音響特徴量と価値観の感度分析

Fig. 3 より、それぞれの音響特徴量で特定の価値観に反応が見られた (Table 4 に示す)。avgF0, sdF0, avgHNR05 は結婚願望や家族優先など世代に関係する価値観が反応し、似た特徴を示している。avgSPL は繊細さに関する価値観が反応していることがわかった。

Table 4 音響特徴量と価値観の分析結果

音響特徴量	関係のある価値観
avgF0	結婚願望 ストレス (仕事) ゆとり (時間) 余裕がない 家族優先 ギリギリ (お金) ゆとり (お金)
sdF0	自己愛 否定批判 結婚願望 責任観_主婦軸 親友中心 家族優先
avgSPL	デリケート 非常識 ストレス (人間関係) ゆとり (時間) 余裕がない
avgCPP	好奇心旺盛 ストレス (人間関係) 親友中心 ゆとり (時間)
avgHNR05	否定批判 結婚願望 責任観_主婦軸 余裕がない 家族優先
avgHNR35	自己愛 満足 (仕事) 親友中心

6 おわりに

今回の調査により、価値観・音響特徴量・声の印象は相互に関係があるという仮説の有効性を確認できた。これらは 5.1 で作成したモデルにより互いに推定が可能である。従って、音響特徴量から声の印象を推定することにより、印象診断、音声の選択基準としての利用、音声合成への応用等に可能性が示せた。

また音響特徴量から価値観を推定することによる、マーケティング活動への活用という新たな音声利用の可能も示せた。

本稿では述べていないが、[9]より価値観から声の印象を推定することも可能である。

課題として、データ数の不足、年代や価値観の偏り、韻律情報を扱っていないことなどがある。また今回は男女を区別せず分析したが、声の印象は男女で異なるため、区別した方がより良い結果が得られる可能性がある。

データ数や多様性の不足という課題を解決するため、音響特徴量から声の印象を推定する機能を応用したスマートフォンアプリ【声タイプ診断】によりデータ収集を行っている。

http://lab.synergy-marketing.co.jp/voice_doctor/

今後はアプリで集めたデータにより上記課題を解決しつつ、価値観・音響特徴量・声の印象についての研究をさらに進めていきたい。

参考文献

- [1] 谷田泰郎他, 価値観モデルを利用したマイクロブログ発言者の社会的類型の推定, 言語処理学会第 19 回年次大会(NLP2013)
- [2] 馬場彩子他, 社会知としての消費者価値観構造モデルと類型「Societas」の構築, 人工知能学会全国大会 (第 27 回) JSAI2013
- [3] 谷田泰郎他, マイクロブログにおける潜在的価値観の推定, 人工知能学会全国大会 (第 27 回) JSAI2013
- [4] 木虎直樹, 久保証人, Web アクセス履歴に基づくユーザの価値観の類推, 人工知能学会全国大会 (第 27 回) JSAI2013
- [5] 谷田泰郎他, 自然言語処理とマーケティング, 電子情報通信学会 思考と言語研究会 (TL), 2013
- [6] Yen-Liang Shue et al., VOICESAUCE: A PROGRAM FOR VOICE ANALYSIS, ICPhS XVII, Hong Kong, 17-21 August 2011
- [7] James Hillenbrand et al., Acoustic Correlates of Breathy Vocal Quality, Journal of Speech and Hearing Research, Volume 37, 769-778, August 1994
- [8] Guus de Krom, A cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals, Journal of Speech & Hearing Research; Apr93, Vol.36 Issue 2, p254
- [9] 高椋琴美他, 価値観から推定する発話特徴と声の印象, ことば工学研究会 (第 43 回), 2013