

## キャラクター音声のステレオタイプ識別のための音響分析\*

石井 沙季<sup>†</sup>, 伊藤 克亘<sup>†</sup>

## 1 はじめに

現代の日本では、アニメやゲーム、舞台といったサブカルチャーコンテンツが人気である。アニメは年間新作が200本ほどテレビやインターネットの動画サイトで配信される。人気の作品になると様々なメディアで展開され、声優の他にも役者がキャラクターを演じるため一人のキャラクターに対していくつもの声がついたりもする。このように、多種多様なキャラクターが数多く生まれてきている。

アニメなどでは、キャラクターの姿を見なくてもしゃべっている音声を聞くだけでそのキャラクターの言動や役割を推定できる。そこでキャラクター音声のステレオタイプ性について注目する。言語学の分野では言語内容からキャラクターのステレオタイプを明らかにする研究が行われてきた。実際にお嬢様や博士などの言葉遣いの特徴が指摘されている。言葉遣いや話し方、つまり音声でどういうキャラクターかわかると物語に入りやすくなるという利点がある。本研究では、言語内容ではなく音声に着目し、キャラクター音声のステレオタイプ性を音響特徴面で明らかにすることを目標とする。

## 2 キャラクターのステレオタイプ性

ステレオタイプ [1] とは社会心理学、社会言語学の概念である。性別や年齢、容姿などの特徴で人間を分類し、そこに属するものが共通して持つとされるものがステレオタイプである。

多くの人は幼少期から物語の中でお嬢様や王子様、ヒーロー、悪役などの様々なキャラクターを目にしている。上記のキャラクターに関しては立場や言動を無意識に決めつけているところがあると感じる。このように無意識に役割等を断定しているキャラクターがステレオタイプなキャラクターであると考えられる。

また、言語学の分野で役割語というものが研究されている。ある特定の言葉遣い（語彙・語法・言い回し・イントネーション等）を聞くことと特定の人物像（年齢、性別、職業、階層、時代、容姿・風貌、性格等）を思い浮かべることができる時、あるいはある特定の人物像を提示されると、その人物がいかにも使用しそうな言葉遣いを思い浮かべることができる時、その言葉遣いを「役割語」と呼ぶ [2]。

これらを用いて言語内容からキャラクターのステレオタイプ性を明確にする研究が行われている。上記の内容を参考にし、キャラクター音声のステレオタイプ性を判断する。

## 3 パラ言語情報とキャラクターとの関係性

キャラクターのステレオタイプの明確化は先に述べたように言語学の分野で行われている。ここでは役割語に引きずられてステレオタイプ性を決定しないようにするため、音声によって分類を行っていく。例えば音声には性別、年齢、態度、性格、感情、調子などが現れる。これは言語内容とは別の情報であり、パラ言

語情報と呼ばれるものである。キャラクターには言葉遣いだけではなく、話し方にも個性がある。語尾の上昇、下降、弱弱しい声、とても力強い声など様々なものがある。以下に三つのセリフ [3] を載せる。

- 「環さまがあなたを構うのは、育ちが珍しいからよ」
  - － 若い女性、気品がある、性格きつそう、皮肉こめてる
- 「まあ、それが噂の？」
  - － 若い女性、気品がある、感嘆、疑問
- 「貧乏な方は暇がないので豆も挽けないというのは本当でしたのね」
  - － 若い女性、気品がある、驚き、若干の興奮

ここで挙げた文章はいずれもお嬢様のセリフである。セリフの下に各音声から得たパラ言語情報を記している。ここからお嬢様である音声は上のようなパラ言語情報を所持していると考えられることができる。共通して得た情報は、「若い女性」「気品がある」であった。他に共通しない情報があることから音がついていると表現の方法が広がるということが推定できる。つまり、パラ言語情報の一部がキャラクターに関係していると考えられる。これらの情報は音声から感じる言語情報以外のものであると記した。したがって、音響的に特徴があると思われる。

## 4 ステレオタイプ性を持つ音声の定義のための準備

聴取実験により3つのステレオタイプ性を持つ音声を定義する。明らかにする対象のステレオタイプは「お嬢様」「真面目」「不良」である。

## 4.1 音声データ収集

アニメ、ドラマCDから音声を集めていく。収集する際、極端な感情音声（誰が聞いても怒っていると感じるものなど）は除く。これは感情まで考慮すると細分化されすぎてしまう可能性があるからである。基本的にそのキャラクターの通常時の音声を集めていく。また、BGMがついていると、正確な判断や分析を行えない可能性が高いためBGMがついていない音声を集める。

## 4.2 音声の逆再生

収集した音声には言語情報が含まれている。本研究では音声のステレオタイプ性を明らかにすることが目的であるため、言語情報が含まれていると役割語 [2] などにより、音声を聴いた者が正確な判断を行うことが難しくなると考えられる。そこで言語情報を無くすために逆再生を施す。

逆再生をしたところで音響的な特徴には大きな違いは現れない。含まれている子音や母音は同一のものであるからである。しかし、聴こえ方の速さに違いが生じることが確認されている。逆再生すると失われる情報があることは聴取時に感じる速度が変わることから明らかである。その点も踏まえ、逆再生した音声と通常再生する音声の二種類を用意する必要があると考えられる。

\*: Acoustical analyses for stereotype identification of character voice Saki Ishii (Hosei Univ.) et al.

<sup>†</sup> 法政大学 情報科学部

### 4.3 F0 推移

逆再生の音声は言語情報を無くし、音響的な特徴を見るのが目的であった。ここでは、元の音声のF0の変化の様子を取り出し、母音の音によって音高情報のみの音声を作成する。

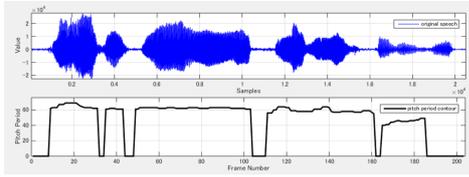


図 1. お嬢様の波形と周期変化

図1はお嬢様の元の音声波形と周期の変化の様子を示したものである。最も音声の変化を聞き取ることが出来ると感じたため、母音は「UH」を使用した。しかし音高の変化の様子はかろうじて認識できるという感じであり、どこで言語が区切れているのか聞き取ることが難しい音が多いように思った。F0の推移のみで被験者に判断を仰ぐことは厳しいかもしれない。その場合、オリジナルの音声とF0の音声の両方を聞いてもらい、何か気がつくことがあるかどうかを聞いてみることも考える。

### 5 予備実験

本格的な聴取実験の前段階として男女各1名に簡易的な実験に参加してもらった。被験者に聞かせた音声は、主に話速を見ることに利用する、通常再生音声。主に音色を見ることに利用する、逆再生音声。主に音高の変化の様子を見ることに利用する、F0 母音音声の三種類である。

今回はあらかじめこちらで明らかにステレオタイプ性を持つと感じた音声を聞いてもらい、実際どのように感じたのかを表1に載せた印象語によって1から7の7段階で答えてもらった。印象語の選定には関連研究 [4] を参考にし、全部で20対の形容詞を用意した。

表 1. 音声評価尺度

高い：低い	弱い：強い
明るい：暗い	遅い：速い
暖かい：冷たい	親しみやすい：親しみにくい
優しい：きつい	流暢な：たどたどしい
軽い：重い	力強い：弱弱しい
澄んだ：濁った	上品な：下品な
柔らかい：硬い	感じの良い：感じの悪い
細い：太い	好きな：嫌いな
丁寧な：雑な	はっきりしてる：はっきりしていない
穏やか：激しい	せわしげな：のんきな

「お嬢様」は澄んだ、丁寧な、上品な、「真面目」は流暢な、はっきりしてる、硬い、「不良」は冷たい、重い、激しいといった形容詞を選択する傾向にあった。音声だけではなく、各ステレオタイプから連想される性格等から言葉を選択されている可能性もあると考えた。

### 6 有効な特徴量の調査

聴取実験から明らかになったステレオタイプ性を持つ音声を音響特徴から有効な特徴量を調べる。現状の予定としては、音色、話速、音高の変化、基本周波数を見ていくことを考えている。

図2は仮で決定したお嬢様と真面目、不良音声のスペクトル包絡を重ねて表示したものである。第一、第

二ホルムントは言語情報を含んでいるため、音響的な特徴が出るのは第三ホルムント以降になる。図2を見ると4500Hzから5500Hzの周波数帯域で特徴が現れているように思われる。

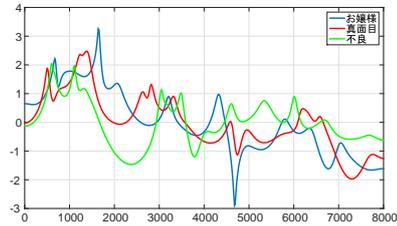


図 2. 「あ」のスペクトル包絡

予備実験の結果から、お嬢様は全体的にゆっくりめ、真面目はきびきびとした一定の速さ、不良は激しく速いといった印象があることを得た。しかし、同様のステレオタイプであってもキャラクターごとに性格があるため、話速は異なる可能性もある。また、平均的な話速では明確な違いが出ない場合も考えられる。語尾の速度など、限定的な部分でも見ていく。音韻ごと、パワー変動の二つの方法で求める。音韻ごとによく話速を求めるときに使用される方法で、発話内容に左右される [5]。一方、パワー変動は発話内容によらない話速を求めることが可能である。それぞれ一定の時間長の中にどれだけのピークが現れるかで話速を定義する。

音高の変化は、例えばお嬢様なら終助詞の「わ」が上昇調であることがわかっている [2]。これは女性的な発音の特徴である。お嬢さまには女性的な発音の特徴（語尾が上がるなど）が多く現れていると考えている。

基本周波数は、男性キャラクターと女性キャラクターでは性別が異なるため、差が出ることは性別判定の研究からも明らかである。また、年齢が異なっても高さの違いが現れることが判明している。現状、真面目と不良において基本周波数に明確な差は現れていないが、データを集めて見ていきたいと考える。

### 7 おわりに

本格的な聴取実験を行い、あらためてステレオタイプ性を持つ音声を定義する。現状は実験を行うための音声を準備した。今後は主観評価実験の具体的な方法を設計していく必要がある。そして聴取実験が済み次第、先に挙げた特徴量において分析を行い、印象語の評価なども参考にし、評価実験を行う。評価実験はシステムによる客観的な評価、人による主観的な評価の二つの方法を用いて総合的に評価する。

### 参考文献

- [1] 山岸俊男, "社会心理学", 新星出版社, 2013
- [2] 金水敏, "ヴァーチャル日本語 役割語の謎", 岩波書店, 2003
- [3] テレビアニメ, "桜蘭高校ホスト部 第一話", 2006
- [4] 山住, 他, "講演音声の音響的特徴とその印象に対する評価構造モデル", Japanese Journal of Sensory Evaluation, 2007, Vol.11, No.1, 30-36
- [5] 峯松, 他, "話者認識技術を利用した主観的高齢話者の同定とそれに基づく主観的年代の推定", 情処学論 vol.43, no.7, pp.2186-2196, 2002