

# ゲームキャラクタと声質の傾向分析

理学専攻 情報科学コース 酒井えりか (指導教員: 伊藤貴之)

## 1 はじめに

日本では毎年たくさんのアニメやゲームが制作され、数多くのキャラクタが生み出される。その声を担当する声優のキャスティングは、キャラクタの印象を決定づける重要な要因となる。本研究では、キャラクタと声質の関係について分析する諸手法を提案する。具体的には、ポータブルゲームの音声収録したデータベースを構築し、類似するキャラクタを担当する声優に共通する音響特徴量、また声がユーザに与える印象を調査する。以上の特徴量と印象値を自己組織化マップ SOM (Self-Organizing Map) [?] に学習させ、それによって得られた音響特徴量と印象値の関係から、印象値を与えることで音響特徴量を推定する。さらに、ゲームソフトに関するウェブ上の文書からキャラクタ間の共起度を算出し、キャラクタ間の距離関係を可視化する。これらを用いることで最終的には、音響特徴量とキャラクタ間共起度から声優をキャスティングするシステムの開発を目指す。そのシステムを利用することで、声優をゲームキャラクタに割り当てる際に議論することができると考えている。

## 2 提案手法

本手法は大きく分けて図??に表すように大きく分けて2つの処理手順から構成される。

- キャラクタの声の印象: 音響特徴量と印象値を調査し、印象値から音響特徴量を推定することで声優候補リストを生成する
- キャラクタの共起度: ゲーム作品に関するテキストデータからキャラクタ間の共起度を算出し、可視化する

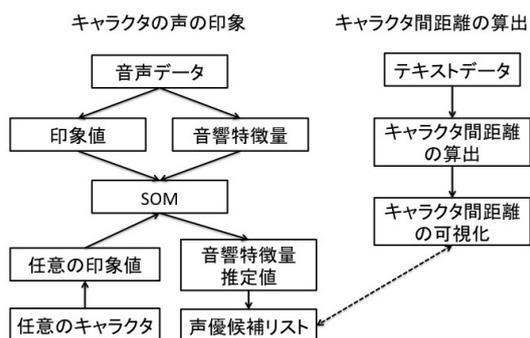


図 1: 研究概要

### 2.1 研究題材

本手法ではゲームキャラクタの声質を題材とする。理由の一つとして、ゲーム以外のメディア (例えばアニメ) では BGM と音声重なっているシーンが多く、我々の環境ではその分離が難しいから、という点があげられる。本研究では前準備として、2013 年から現在 (2015 年 8 月) までに発売された 156 作品を対象とし

て、キャラクタと担当声優のデータベースを構築した。本研究ではその中から 20 以上のキャラクタを担当している声優を選出し、その声を収集した。具体的には、我々自身で所有するゲームの中から対象となるものを PSP (PlayStation Portable) でプレイし、各キャラクタのセリフをソニーウォークマンのシンクロ録音を利用して収集した。

## 3 キャラクタの声の印象

### 3.1 音響特徴量の計算

2.1 節にて収集した音声データに対し VoiceSauce[?] を用いて音響特徴量を計算した。本手法で用いている音響特徴量は声の高さと関係がある基本周波数など計 16 種類である。この 16 の音響特徴量を 16 次元ベクトルとして扱った。続いて、各ゲームキャラクタの声の音響特徴量の範囲を計算する。以下、 $i$  番目のゲームキャラクタの音響特徴量の範囲を  $R_i = \{[f_1^{min}, \dots, f_1^{max}], \dots, [f_{n_f}^{min}, \dots, f_{n_f}^{max}]\}$  と表記する。なお  $n_f$  は現在の実装では 16 となる。この値は後述する通り、任意のゲームキャラクタについて適切と思われる声優を推薦するために用いられる。

### 3.2 ユーザの印象値調査

本研究では大学生 15 人に対して録音したセリフの印象調査を実施した。この調査では 8 人の声優の 17 キャラクタを対象とし、各々について 10 セリフを選んで音響特徴量を算出し、その平均値との距離を算出した。この距離が最小となるセリフを代表セリフとし、ユーザに与える印象を調査するために使用した。ユーザに与える印象調査法は清水らの「動画特徴量からの印象推定に基づく動画 BGM の自動生成」[?] にて提案されている評価システムを採用した。アンケート項目は木戸らの「通常発話の声質に関連した日常表現語の抽出」[?] を参考とし「高い声」「低い声」などを含めた 8 つの表現語対を使用した。以上で算出した音響特徴量と印象値を SOM に学習させた。また SOM を用いることにより任意の印象値を与えることで音響特徴量の推定を可能とした。

### 3.3 SOM の学習結果と音響特徴量推定の結果

図??は音響特徴量と印象値を SOM に学習させた結果である。図??は任意の印象値から音響特徴量を推定した結果である。A~Q までのアルファベットは各キャラクタを示している。

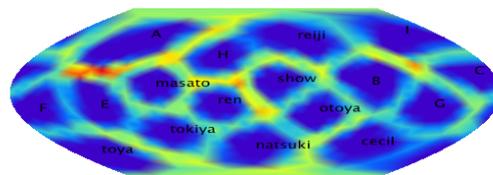


図 2: 音響特徴量と印象値の学習結果

ここで (masato, B, E, G) は全て同じ声優が担当し

ているが、(masato, E) と (B, G) の 2 か所にわかれた。(masato, E) は寡黙でクールな性格を持っており、(B, G) は明るい性格である。この配置結果から音響特徴量と印象値を学習させた結果は妥当であると解釈した。この学習結果を用いて、任意の印象値を入力し推定音響特徴量を算出した。

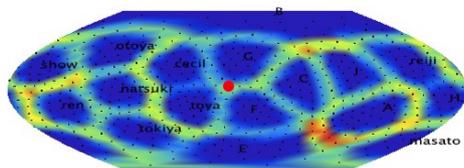


図 3: 任意の印象値から音響特徴量を推定した結果

図??中の赤い点は、任意の印象値を与えて推測した音響特徴量を表す。この音響特徴量は SOM 上の周辺のゲームキャラクターの音響特徴量を補間して求められる。

### 3.4 声優候補リストの生成

新しいキャラクターの適切な音響特徴量を  $f = \{f_1, \dots, f_{n_f}\}$  として推定し、音響特徴量の範囲  $R_i$  と比較する。 $i$  番目のゲームキャラクターのすべての音響特徴量が  $f_j^{min} < f_j < f_j^{max}$  を満たす場合、 $i$  番目のキャラクターを担当する声優が新しいキャラクターに対する声優候補の 1 人であると判断する。すべてのゲームキャラクター  $R_i$  を適用し、新しいキャラクターに対する声優候補リストを生成する。

## 4 キャラクターの共起度の可視化

本研究では研究対象としているゲーム作品に関するウェブ上の解説を収集し、Word2Vec[?] に学習させることで、キャラクターを多次元ベクトルとして数値表現した。この多次元ベクトルの集合に主成分分析 (PCA) を用いて、それらの距離を可視化した (図??)。

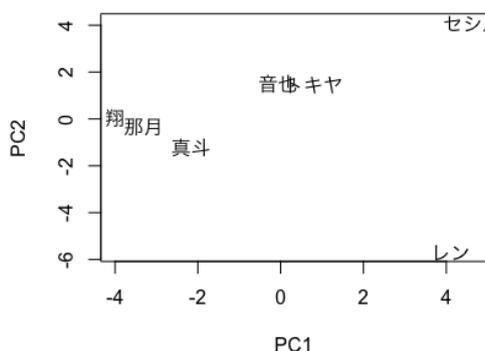


図 4: キャラクターの共起度の可視化

我々は図??,??から、この距離はウェブ上のキャラクター間の共起度を表していると判断し、共起度が高いキャラクターには異なるタイプの声優を割り当てるべきと判断した。例えば図??より、(音也, トキヤ) および (翔那月, 真斗) は共起度が高いと示唆される一方で、図??からそれぞれ異なるタイプの声であることがわかる。

声優候補リストによって得られた声優をこの可視化画面に表示することで、声優の割り当てについての議論に使用できる。また、多次元ベクトルから算出されるキャラクター間距離を用いることでキャラクター集合をグラフで表現し、そのグラフに彩色問題を適用することで、声優を自動的に割り当てることも可能である。また、主たるキャラクターの声優は決まっているが他の声優は決まっていないといった際にも、既に決定している声優を固定し、他の声優を自動的に割り当てることも可能となる。

## 5 まとめと今後の課題

本論文では、ゲームキャラクターのための声優推薦ツール開発に向けた検討を行った。このツールはまず自己組織化マップ (SOM) を適用することによって、ゲームキャラクターの印象値と音響特徴量との関係を学習する。これにより、学習結果を用いて任意の印象値から適切な音響特徴量を推定することができる。さらにこの推定結果から、新しいキャラクターの声優候補のリストを生成できる。同時に、ゲームに関するテキストデータを用いて、キャラクター間の距離を求め主成分分析を用いて可視化する。声優候補のリストによるこの可視化結果は、声優をゲームキャラクターに割り当てる方法について議論する際に有用であると考えられる。

今後の課題としては、まずキャラクターの声の印象の部分に関しては、

- 各ゲームキャラクターの声の音響特徴量の範囲を実際に計算すること
- 実際に声優候補リストを生成すること

があげられる。キャラクターの共起度の部分では、現時点では 1 作品のみでのキャラクターの共起度の可視化画面になっているので複数作品のキャラクターをプロットすることが課題としてあげられる。これらを実装した上で、共起度の可視化画面に声優候補リストにて選出された声優をプロットすることが課題である。

## 参考文献

- [1] A. Mehrabian: Silent messages, Wadsworth, Belmont, California, 1971.
- [2] 高塚, Ying: 球体 SOM のデータ構造と量子化誤差の考察およびインタラクティブ性の向上, 日本知能情報ファジィ学会誌, Vol. 19, No. 6, pp. 611-617, 2007.
- [3] Shue, Y.: VoiceSauce: A program for voice analysis, ICPHS XV, pp. 1846-1849.
- [4] 清水, 菅野, 伊藤, 嵯峨山, 高塚: 動画特徴量からの印象推定に基づく動画 BGM の自動生成, 情報処理学会第 78 回全国大会, 2Q-01, 2016.
- [5] 木戸, 粕谷: 通常発話の声質に関連した日常表現語の抽出, 音響学会誌, 55, 405-411, 1999.
- [6] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean: Distributed Representations of Words and Phrases and Their Compositionality, Neural Information Processing Systems (NIPS), 2013.