

# アニメ動画の声優認識のためのコンテキストを意識した性別判定

榮田 基希<sup>†</sup> 服部 峻<sup>††</sup>

<sup>†,††</sup>室蘭工業大学 ウェブ知能時空間研究室 〒050-8585 北海道室蘭市水元町 27-1  
E-mail: <sup>†</sup>16043009@mmm.muroran-it.ac.jp, <sup>††</sup>hattori@csse.muroran-it.ac.jp

あらまし アニメ動画から音声の流れるとき、それが誰であるか調べようとするならば、エンドロールを探すといった手間を掛ける必要がある。音声から声優認識することが出来るようになれば、手間が掛からずに声優名が分かるだけでなく、その声優の他作品の出演情報やブログ、関連動画や関連商品、これからのイベント情報など幅広く情報を取得することが出来る。これまでの研究では、Web から取得したアニメ動画のキャスト情報や、音声の振幅、周波数パワースペクトルを活用することによって声優認識を試みたが、声優認識精度として良好な結果を得ることが出来なかった。そこで本稿では、まず音声の主の性別を判定して候補となる声優を絞り込むことにより、声優認識の精度向上を狙う。声優の性別判定を行うために、声優認識対象の音声から得られる基本周波数や、その音声の前後のセリフ内容のコンテキスト情報を活用する。

キーワード 声優認識, 性別判定, アニメ動画, コンテキスト・アウェアネス

## Context Aware Gender Determination for Voice Actor Recognition in Anime Video

Motoki EIDA<sup>†</sup> and Shun HATTORI<sup>††</sup>

<sup>†,††</sup> Web Intelligence Time-Space (WITS) Laboratory, Muroran Institute of Technology  
27-1 Mizumoto-cho, Muroran, Hokkaido 050-8585, Japan  
E-mail: <sup>†</sup>16043009@mmm.muroran-it.ac.jp, <sup>††</sup>hattori@csse.muroran-it.ac.jp

**Abstract** When we hear someone's voice from an anime video, we need to carry extra burdens of searching the end roll of the anime video in order to know about whose voice it is. If a system can recognize a voice actor from his/her voice on behalf of us, not only we can know about the voice actor's name without carrying extra burdens, but also we can acquire widely information about him/her such as his/her appearance information, blogs, related videos, related goods, and event information in the future. Our previous papers tackled a system of voice actor recognition with filtering by cast information extracted from the Web and similarity calculation based on voice amplitude and frequency power spectrum, but the system could not give enough good performance as voice actor recognition accuracy. Therefore, this paper tries to improve the accuracy of voice actor recognition by narrowing down candidates for voice actors by determining the gender of the owner of a target voice. The system utilizes the fundamental frequency acquired from a target voice and the context information of the content of spoken lines surrounding the voice, in order to determine the gender of a voice actor.

**Key words** Voice Actor Recognition, Gender Determination, Anime Video, Context Awareness

### 1. ま え が き

近年、日本には様々な娯楽メディアがあり、普段の生活の中で目や耳にする機会が多くなっている。情報通信機器の普及により多くの人にとって、パソコンやモバイル端末などの機器で番組や動画の視聴、ゲームなどが今では手軽に行うことが出来る。このような娯楽メディアに触れる機会が多くなって来る

と、どこかで聞いたことがある誰かの音声の流れて来ることがある。ミュージックビデオ中の歌手の歌声、テレビドラマや実写映画中の俳優が演じる役柄のセリフ音声、アニメ動画中の声優が演じるキャラクターのセリフ音声などが挙げられる。

音声の発生源がアニメ動画の場合、誰の音声であるかを知る為には、エンディングのスタッフロールまで飛ばしたり、Web で作品のタイトル名やキャラクター名で検索したりするなどの

余計な労力を掛ける必要が出て来る。例えば、あるユーザが適当なアニメを視聴していた際、そのアニメの中に出て来たキャラクター A の音声ユーザの聞いたことのある音声であったとする。そこで、そのユーザがキャラクター A の声優について調べようとするならば、エンディングまで飛ばしたり、アニメタイトルやキャラクター名で Web 検索して、そのアニメの公式サイトやウィキペディアなどを探そうとするであろう。しかし、知りたいキャラクター A が作中の目立たない配役だった場合、Web で検索を掛けても中々出て来ないことも考えられる。また、主要なキャラクターではない場合、キャラクター名を記憶していない可能性もあり、エンディングのスタッフロールが流れても分からないであろう。その上、脇役であった場合、スタッフロールには男の子 B、男の子 C というようにキャラクター名を不明瞭に表記していることもあり、どの場面に出て来たキャラクターか分からないことも考えられる。

ここで、ユーザが労力を掛けずに、音声の主の声優を知ることが出来るようにする為には、アニメ視聴中に音声の流れたりリアルタイムに声優名を認識して自動的に画面に表示するシステムが必要になる。

前回、我々の研究 [1, 2] では人の声質にはバンド幅毎にそれぞれ異なる強さ (dB) を持つと仮説を立て、声優が個々に持つ特有の周波数パワースペクトルのパターンである「特有パワースペクトル」を特定出来れば声優認識が可能になると考えた。そこで、各声優の音声情報が登録されているデータベースから取得した各声優の特有パワースペクトルと、アニメ動画から流れる音声データのパワースペクトルとの間の類似度計算を、データベースに登録されている声優の数だけ行うことに依って、その音声の持ち主が誰であるか認識する手法を提案した。しかしながら、予めデータベースに登録しておくサンプル音声は「アニメ声」か「素の声」かによって精度が大きく変わってしまった。また、実験する動画と同じアニメであるキャラクターのサンプル音声をデータベースに登録しておいても、音声を用いた声優認識の精度として良好な結果を得ることが出来なかった。

そこで本稿で我々は、アニメ動画に流れる音声から声優の性別を判定する手法を提案する。声優認識を行う前にまず、音声の主の性別を判定することにより、候補となる声優の絞り込みを掛けることで声優認識の精度向上を図る。性別判定の際には、これまでの研究と同様に、視聴中のアニメ動画のタイトルが特定されていることで、そのタイトルに基づいて Web 検索されたキャスト情報によって声優をキャスト陣のみに絞り込んでいる状態を想定し、音声の特徴量の 1 つである基本周波数を用いる。声優データベースには男性声優、女性声優の基本周波数をそれぞれ予め登録しておく。

## 2. システム概要

これまで我々は、音声の振幅や周波数スペクトルに基づく類似度計算を用いた声優認識システム [1, 2] を研究開発して来たが、図 1 のように声優の性別判定を新たに導入することで、声優データベースに登録されているキャスト情報から認識される声優の候補を絞りこみ、声優認識の精度向上を試みる。

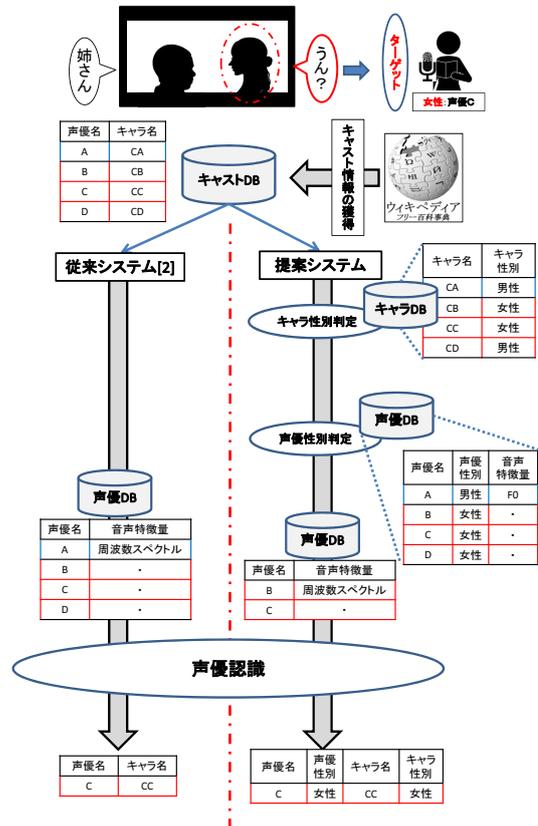


図 1 提案システム概要

歌舞伎の女形や宝塚歌劇の男役を除き、ほとんどの場合、俳優の性別と、その俳優が演じる役柄の性別は同一であるが、一方、声優の性別と、その声優が演じる（アニメや吹き替えの）キャラクターの性別とは異なる場合も少なくない。しかし、アニメ界の現状において、前述したようにアニメキャラクターの男性役を担当する声優は、男性、女性と共にキャストされている傾向が見られるが、アニメキャラクターの女性役を担当する声優においては、女性が多い傾向があることが分かった。そこで、図 1 のキャラ性別判定において、アニメキャラクターの性別を判定することが出来れば、そのキャラクターの音声の主の性別判定を精度良く行うことに繋がると考えた。例えば、あるアニメのある話数に登場するアニメキャラクターが男性一人に女性多数といった場合、アニメ動画に流れる音声からキャラクターが男性の性別であると判定出来れば自動的に声優を絞り込むことが出来たり、先に述べた傾向分析が正しければ、声優認識の対象であるアニメキャラクターの性別が女性であると判定出来た場合、その音声の主が男性声優である確率が下がるという手法を施すことが出来る。

実際に流れる動画から、キャラクターの性別判定を行う時に、声優認識のターゲットとなる音声の前後のセリフ内容の単語をコンテキスト情報として活用する。例えば、「兄さん」「姉さん」といった単語が出現した時、その呼称に当たるキャラクターがその場面に登場する確率が高いと考えた。また、性別を表す単語だけでなく、「先生」「隊長」といったキャラクターの属性情報を獲得出来れば、キャラクターの名称を判定して芋づる式に声優名が判定出来ると考えた。



### 3.2 キャラクター性別判定アルゴリズム

本節では、アニメキャラクターの情報が載っている Web テキストから、対象キャラクターの性別を判定する手法について述べる。前節で用意した辞書を用いて Web テキストを解析して、各キャラクターの性別を判定する。キャラクターの紹介が記載されている Web テキスト内には、キャラクターの性別に応じた関連語や連想語などが頻出していると仮説を立てた。例えば、男性キャラクターの紹介欄の場合、Web テキストには「お父さん」「兄」などの男性を連想する語が多く含まれ、女性キャラクターの紹介欄の場合、Web テキストには「お母さん」「姉」などの女性を連想する語が多く含まれていると予測した。さらに、キャラクター紹介の文章の大部分は一人称視点や二人称視点で記載されているのではなく、三人称視点で記載されており、男性の場合「彼」、女性の場合「彼女」といった単語で紹介されている場合が多い。また、紹介されている対象キャラクターについての詳細が、主に前半部分に集中して記載されている傾向があると考えて、男女それぞれの性別判定辞書に含まれる単語の出現順に重みを変化させて付与する必要があると考えた。

前述したキャラクター紹介の文章に関する考察から、それぞれのキャラクターに対して、その Web テキストから性別に関連する単語の出現回数を数えることでアニメキャラクターの性別判定が可能であると考えた。そこでまず、日本語係り受け解析器である CaboCha [10] を使用して、Web テキストの文章を対象に形態素解析を行い、単語を抽出する。Web テキストを対象に、初めの文章から順番に解析して抽出した単語が、男女の性別判定辞書に含まれている単語であった場合、マッチした単語順に  $k$  番目の単語を  $w_k$  ( $k \in \{1, 2, 3, \dots, N\}$ ) と表し、男女それぞれの辞書において出現回数をカウントする。単語  $w_k$  の男性の辞書  $D_M$  における出現回数を関数  $DC_M(w_k)$  で、女性の辞書  $D_F$  における出現回数を関数  $DC_F(w_k)$  で定義する。さらに、男女の辞書に含まれる単語の出現順毎に重みを付ける為、

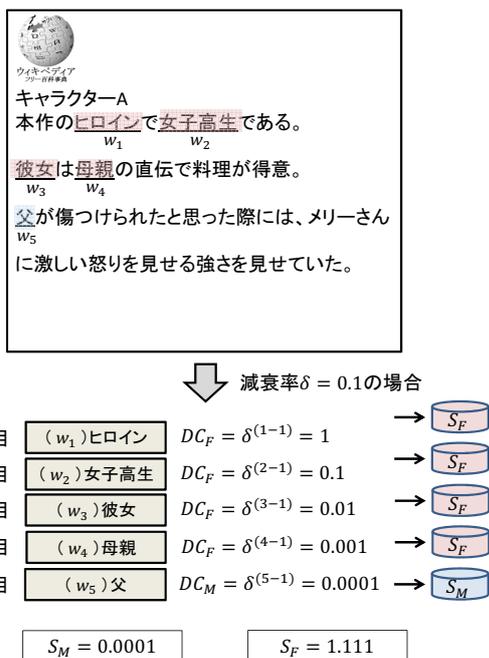


図3 キャラクター性別判定手法

減衰率  $\delta$  ( $\delta \in \{0.00, 0.01, 0.02, \dots, 1.00\}$ ) を設けて、単語  $w_k$  に対してカウントする値を  $\delta^{k-1}$  とした関数を以下に示す (前述の  $DC_{M/F}(w_k)$  では、加算されるのは 1.00 である)。

Web テキストの解析が終了するまでに行われる、場合分けによる男性の辞書における最終的な重み付き出現回数  $S_M$ 、女性の辞書における最終的な重み付き出現回数  $S_F$  のカウント方法を以下の数式と図3に示す。

(1) 男性辞書における出現回数 (減衰率  $\delta$  付き)

$$DC_M^*(w_k) = \begin{cases} \delta^{k-1} & (w_k \in D_M \text{ のとき}) \\ 0 & (\text{otherwise}) \end{cases}$$

$$S_M = \sum_{k=1}^N DC_M^*(w_k)$$

(2) 女性辞書における出現回数 (減衰率  $\delta$  付き)

$$DC_F^*(w_k) = \begin{cases} \delta^{k-1} & (w_k \in D_F \text{ のとき}) \\ 0 & (\text{otherwise}) \end{cases}$$

$$S_F = \sum_{k=1}^N DC_F^*(w_k)$$

Web テキストの文章を全て解析した後、男女に関する辞書の単語の出現回数 (減衰率付き) が大きかった性別を採択する。

$$\begin{aligned} S_M > S_F &\rightarrow \text{男} \\ S_M < S_F &\rightarrow \text{女} \\ \text{else} &\rightarrow \text{判定出来ず} \end{aligned}$$

### 4. 基本周波数に基づく声優性別判定

実際にアニメ動画から流れる音声から、その音声の持ち主である声優の性別を判定する為、声優に限定しない性同一性障害者の話声位に関する研究 [11] を参考にして、声優の特徴である「アニメ声」であっても男女によって基本周波数 (以下、F0) の値に差が生じると仮説を立てた。そこで、声優における男女それぞれの F0 に対して特徴を突き止めて、分類することが出来れば、性別判定が可能になると考えた。この仮説から、男女声優それぞれのサンプル音声を収集する。それらの音声データから、再生時間の長さだけ取得される F0 値を声優データベースに登録しておく。一人の声優におけるのサンプル音声の数はばらけており、ノイズを省く為に BGM や効果音といった雑音が出来ただけ無いものを選択している。F0 の取得には、アムステルダム大学の Paul Boersma 氏と David Weenink 氏によって開発されたオープンソース・ソフトウェア Praat [12] を用いた。標準化周波数を 16000 Hz に、Pitch の範囲については Pitch floor を 75 Hz、Pitch ceiling を 500 Hz に設定した。本稿では、アニメ動画の音声から声優の性別を判定する為に、正規分布の確率密度関数の計算式を用いた手法を提案する。

正規分布の計算の為、声優データベースに登録されている F0 の標本データを用いて、男女毎に平均  $\mu$ 、分散  $\sigma^2$  を求めておく。分析対象である声優のセリフの音声から得られた F0 の羅

列のテストデータを  $x_i$  ( $i \in \{1, 2, 3, \dots, T\}$ ) と表す。  $x_i$  を対象に、男女各々の正規分布において算出された確率を別々に加算していく (以下,  $PS$ )。男性の正規分布に基づいて加算された  $PS$  を  $PS_M$ , 女性の方を  $PS_F$  と表す。性別判定が行われるまでの計算式を以下に示す。  $f$  は正規分布の確率密度関数である。

$$f(x_i, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

$$PS_M = \sum_{i=1}^T f(x_i, \mu_M, \sigma_M)$$

$$PS_F = \sum_{i=1}^T f(x_i, \mu_F, \sigma_F)$$

テストデータ  $x_i$  が  $T$  まで計算された後、最終的な  $PS_M$  と  $PS_F$  の大小比較を行い、大きい値であった性別を採択する。

$$PS_M > PS_F \rightarrow \text{男}$$

$$PS_M < PS_F \rightarrow \text{女}$$

$$\text{else} \rightarrow \text{判定出来ず}$$

## 5. 評価実験

本章では、アニメ動画の声優認識の精度向上を狙って新たに導入した 2 種類の性別判定について、性別判定の精度に関する評価実験を行う。

### 5.1 キャラクター性別判定の評価

3 章で提案したキャラクター性別判定のアルゴリズムの精度評価を行う為に、Wikipedia, ピクシブ百科事典, ニコニコ大百科の 3 つの Web サイトを対象に実験を行った。20 タイトルのアニメを 1 タイトルにつき男女 5 人ずつキャラクターを選出して、男性 100 人, 女性 100 人の説明が載っている Web テキストをそれぞれ収集した。これらの Web テキストを対象に、3.1 節で用意した 5 種類の辞書を用いて、3.2 節で提案したキャラクターの性別判定アルゴリズムの精度評価を行った。図 4 から図 9 は、20 タイトルの男女キャラクター 100 人ずつを性別判定した結果、正解した数の割合を減衰率  $\delta$  の変動毎に求めた結果を表している。

まず初めに、5 種類の辞書における精度を比較すると、著者が主観で決めた辞書 2 の精度が良い印象を受ける。しかし、全ての Web サイトの男性キャラクターの正解率に注目してみると、日本語 WordNet を利用して作成した辞書 4 の精度が辞書 2 を上回っている。日本語シソーラス連想類語辞典で作成された辞書 3 に関しては、女性キャラクターに対して精度は良いが、男性キャラクターに対しては精度は悪い。辞書 3 が辞書 4 よりも全体的に精度が悪いのは、まず、ノイズが含まれている量が多いことが考えられる。一部例を挙げると、男性の辞書に「ルックス」「愛人」といった男女のどちらでも意味が取れる単語が見つかった。また、「お嬢様」「皇女」など、明らかに女性を意味する単語が含まれていることも分かった。日本語シソーラス連想類語辞典は、連想語の単語も取得出来ることから、ノイズが含まれるといった弊害が起きたと考えられる。また、辞

書間の単語数の違いも原因の一つとして考える。図 5, 6 を見ると、女性キャラクターを対象にした解析で、男性辞書よりも単語数が少ない女性辞書の場合、正解率が高い。図 7, 8, 9 でも、男性キャラクターを対象にした解析で、女性辞書よりも単語数が少ない男性辞書の場合、正解率が高い。これらの考察から、辞書の単語数を男女共に出来るだけ少なく均等にして、また、重要な単語を選別出来れば、精度向上を期待出来る。

最後に、減衰率  $\delta$  が及ぼしている影響を考察すると、どの図も減衰率が 0.5 の辺りで変動し始めていることが分かる。その後の変動は上昇下降とそれぞれ動きが異なるが、一番精度が良かった辞書だけを対象に、安定した精度を確保するならば 0.9 がベストであると考察する。

### 5.2 正規分布に基づく音声性別判定の評価

4 章で提案した音声の F0 を用いた正規分布に基づく性別判定の評価を行う。実験対象である音声は 3 種類のアニメタイトルを用いて、そのアニメ動画のある話数に限定を掛けて、そこから取得出来たキャラクターボイス、総計男性 85 個, 女性 87 個をそれぞれ用意した。それらは、BGM や効果音といったノイズが出来るだけ含まないものを実験対象の音声として選定した。尚、用意したセリフ 1 つずつの長さは多様である。また、声優データベースには、実験対象音声の話数に出演するキャスト陣の基本周波数が含まれている。タイトル A のキャスト情報は 8 人, タイトル B のキャスト情報は 2 人, タイトル C のキャスト情報は 5 人であった。声優データベースも実験対象音声と同様に、ノイズを出来るだけ含まないものを選択した (あるいは著者らでその部分を除去した)。声優データベースの音声は、声優が所属している事務所で公開しているサンプル音声や、ある声優に特化したセリフ集などから収集したものである。

表 2 性別判定の評価 (正解数/実験音声の数)

声優	男性 1	男性 2	女性 1	女性 2	女性 3
タイトル A	16/24	9/9	28/28	—	—
タイトル B	6/6	—	4/5	—	—
タイトル C	43/46	—	33/34	10/10	10/10

総計 : 正解率 (男性) 0.870, 正解率 (女性) 0.970

表 2 の実験結果を考察すると、全体的に男性、女性と共に高い精度であると言える。しかし、タイトル A の男性 1 の精度が比較的悪い。その要因として考えられるのは、タイトル A の男性 1 は、声質が他の男性の声優と比べると、主観ではあるが比較的高い部類である。そこで、男性の正規分布の式によって得られた確率値が、女性の正規分布の式で得られた確率値よりも小さくなり不正解が多くなってしまったと考えられる。明確に男女の声質が高低に分かれているならば、声優データベースによって作られる正規分布も男女で重ならないことが予測されて、性別判定が上手くいき易いのではないかと考えられる。

## 6. 今後の研究課題

アニメキャラクターの Web テキストから正しく性別判定する為には、男性、女性を明確に表す単語の選定が必要である。

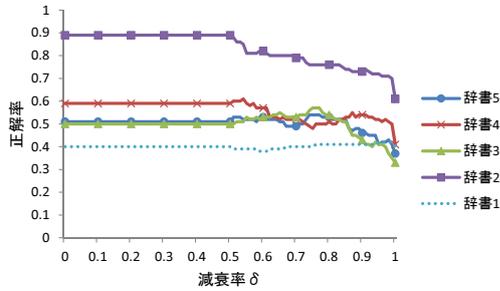


図4 Wikipedia (女性キャラ)

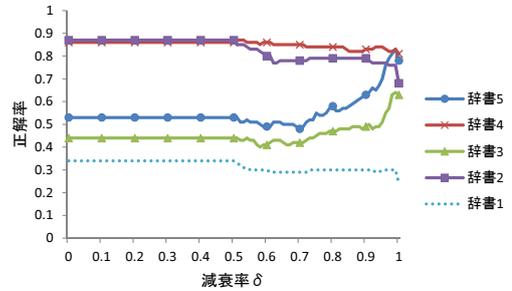


図7 Wikipedia (男性キャラ)

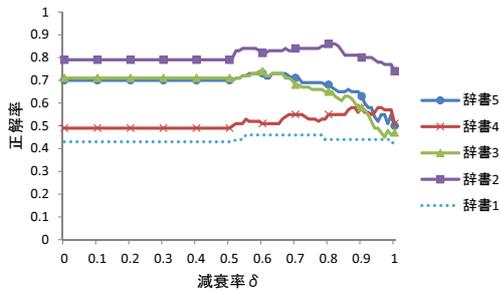


図5 ピクシブ百科事典 (女性キャラ)

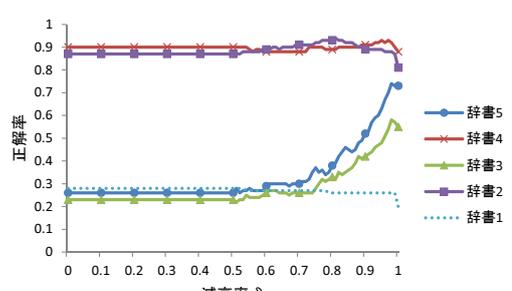


図8 ピクシブ百科事典 (男性キャラ)

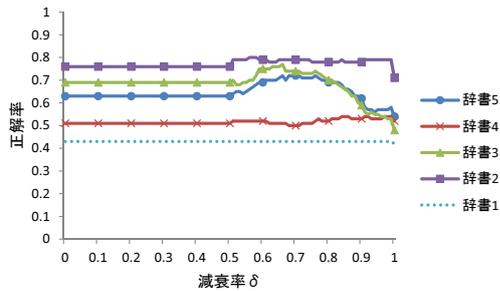


図6 ニコニコ大百科 (女性キャラ)

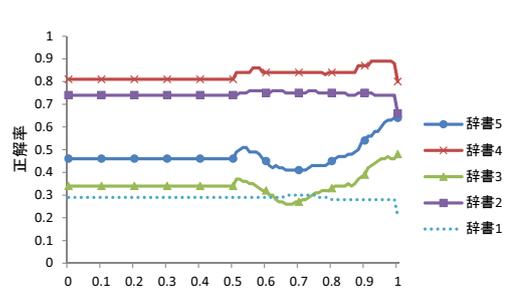


図9 ニコニコ大百科 (男性キャラ)

また、本稿で提案した、Web テキストを上から解析して、辞書に含まれている単語の順番を考慮するアルゴリズムだけでなく、Web テキスト上での、文字数または行数といった単語の出現位置を考慮した解析方法も考えられる。

また、紹介されているキャラクターとの関わり合いが深いキャラクター（親類や家族構成）やそのキャラクター特有の話題といったノイズが Web テキストに混在されて記載されていることがある。例を挙げると、男性キャラクターの紹介欄に「母は〇〇である」「彼女から初めて自身への想いを告げられた」といった事柄が記載されていた場合や、アニメタイトルに性別を表す関連語や連想語が含まれる場合がある。「母」「彼女」といった代名詞が、解析しているキャラクター本人のことを指した単語であるかどうかを判定する係り受け解析や、タイトル名に含まれているかどうかを判定する必要がある。

最後に、実際に流れる動画の音声からキャラクターや声優の性別判定の改善の為に今後行う必要がある研究としては、コンテキスト情報を活用した性別判定の評価実験や、基本周波数に基づく性別判定におけるグループの分割が挙げられる。

## 文 献

- [1] 築田 基希, 服部 峻, “アニメ動画の音声とキャスト情報を用いた声優認識,” 電子情報通信学会 情報ネットワーク研究会, 信学技報, Vol.115, No.405, pp.7–12 (2016).
- [2] 築田 基希, 服部 峻, “アニメ動画における音声の周波数スペクトルを用いた声優認識,” 電子情報通信学会 情報ネットワーク研究会, 信学技報, Vol.116, No.304, pp.25–30 (2016).
- [3] 一般社団法人 日本音響学会 Q and A (050), <http://www.asj.gr.jp/index.html> (2017).
- [4] まうまう☆, <https://www.mau2.com/> (2017).
- [5] Wikipedia, <https://www.wikipedia.org/> (2017).
- [6] ピクシブ百科事典, <https://dic.pixiv.net/> (2017).
- [7] ニコニコ大百科, <http://dic.nicovideo.jp/> (2017).
- [8] 日本語シソーラス連想類語辞典, <http://renso-ruigo.com/>.
- [9] Francis Bond, Timothy Baldwin, Richard Fothergill, Kiyotaka Uchimoto, “Japanese SemCor: A Sense-tagged Corpus of Japanese,” Proceedings of the 6th International Conference of the Global WordNet Association (GWC’12), pp.56–63 (2012).
- [10] CaboCha/南瓜: Yet Another Japanese Dependency Structure Analyzer, <http://taku910.github.io/cabochoa/>.
- [11] 櫻庭 京子, 今泉 敏, 峯松 信明, “女性と判定される声の特徴—性同一性障害者の話声位—,” 音声言語医学, Vol.50, No.1, pp.14–20 (2009).
- [12] Praat, <http://www.fon.hum.uva.nl/praat/> (2017).