

# アニメ動画における性別判定を用いた声優認識のための 音声の高低に基づく判定クラスタの細分化

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 flowed 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。また、候補となる声優を絞り込むために、声優認識対象の音声から得られる基本周波数を利用して性別判定を試みたが、声が高い男性声優など、良好な結果を得ることが出来なかった。そこで本稿では、音声の高低に基づく判定クラスタの細分化によって性別判定の精度向上を狙う。また、性別判定を用いた声優認識の精度評価を行う。

キーワード 声優認識, 性別判定, アニメ動画, 細分化

## Pitch-based Cluster Segmentalization for Voice Actor Recognition Using Gender Determination 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. And also we tried to determine the gender of the owner of a target voice using its fundamental frequency in order to narrow down candidates for voice actors, but our system could not give enough good performance, especially for the male voice actors whose voice is high pitched. Therefore, this paper tries to improve the accuracy of gender determination by pitch-based cluster segmentalization. In addition, we conduct an evaluation on the accuracy of voice actor recognition using the improved gender determination.

**Key words** Voice Actor Recognition, Gender Determination, Anime Video, Segmentalization

### 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 flowed 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したりするであろう。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、ユーザが余計な労力を掛けずに、音声の主の声優名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ら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] では、声優認識の精度向上の為、実際にアニメ動画から流れる音声を用いて、アニメキャラクターの性別判定と音声の持ち主の性別判定を行う手法を提案した。Web テキストから抽出されたキャスト情報で声優データベースに絞り込みを掛けた後、アニメ視聴中のリアルタイムに取得される音声の持ち主の性別判定を行うことで、更に声優の候補を絞り込む為である。

また、これまでの我々の研究 [2,3] では、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、予めデータベースに声優の数だけ「特有パワースペクトル」を登録しておくことで、その音声の持ち主が誰であるか認識する手法を提案している。しかし、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、声優認識だけでは精度が良くなかったシステムに対して、従来よりも精度良く性別判定を行えることが期待出来るシステムと組み合わせることで、より良い声優認識精度を目指す。本稿で我々は、声優の音声の高低に基づく判定クラスターの細分化を新たに導入することで性別判定の精度向上を狙い、その性別判定を組み込んだ声優認識の実験も行う。

## 2. システム概要

図 1 のように、本稿で提案する声優認識システムは入力アニメ動画中の音声、出力を声優名とするが、声優認識の事前準備として、まず初めに、処理 1 で、まうまう [4] や Wikipedia [5]

といった、アニメ作品に出演した声優をまとめた Web サイトのテキストからキャスト情報を取得して、声優データベースに対して声優名で絞り込みを掛ける。

次に図 1 の処理 2 では、声優認識精度を上げる為、実際にアニメ動画から流れる音声から、その音声の持ち主の性別を判定して、更に声優データベースに対して絞り込みを掛ける。これまで我々は、この処理 2 における性別判定手法で基本周波数 (以下、F0) に基づいた音声の持ち主の性別判定 [1] を研究開発して来た。音声の持ち主の性別判定を行う為、声優の特徴である「アニメ声」であっても男女によって F0 の値に差が生じると仮説を立て、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、性別判定が可能になると考えて、その手法として、正規分布の確率密度関数に基づく性別判定を行った。

本稿では従来の性別判定手法の発展として、声質をより細かく解析することで性別判定をより精度良く行うことが出来ると考えて、図 1 の処理 2 の部分にターゲットとなる音声の声域の高低を判定して分類する新たな手法を処理 2 に組み込む。

また、図 1 の処理 3 に当たる声優認識では、過去に我々が研究して来た周波数スペクトルに基づく声優認識システム [2,3] を参考にして、周波数スペクトルを用いて声優データベースに登録されている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のパワースペクトルとの類似度計算を行う。

本稿のシステムは声優認識を行う際、声優データベースに対してキャスト情報の絞り込みを行った上で更に声優の性別で絞り込みを掛けるパターンと、性別判定を行わないでキャスト情報の絞り込みだけを行ったパターンの 2 種類に大別出来る。性別判定の手法にも、著者らが提案した従来の性別判定手法を組み込んだ場合、そして、本稿で新たに提案する声域を分割してから性別判定する手法の 2 つのパターンがある。性別判定を組み込むことで、声優認識の精度にどのような影響を与えるのか、詳細に評価実験していく。

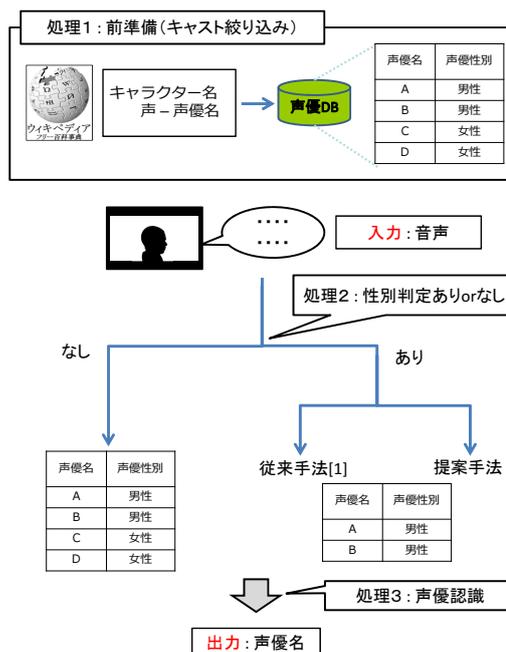


図 1 システムの流れ

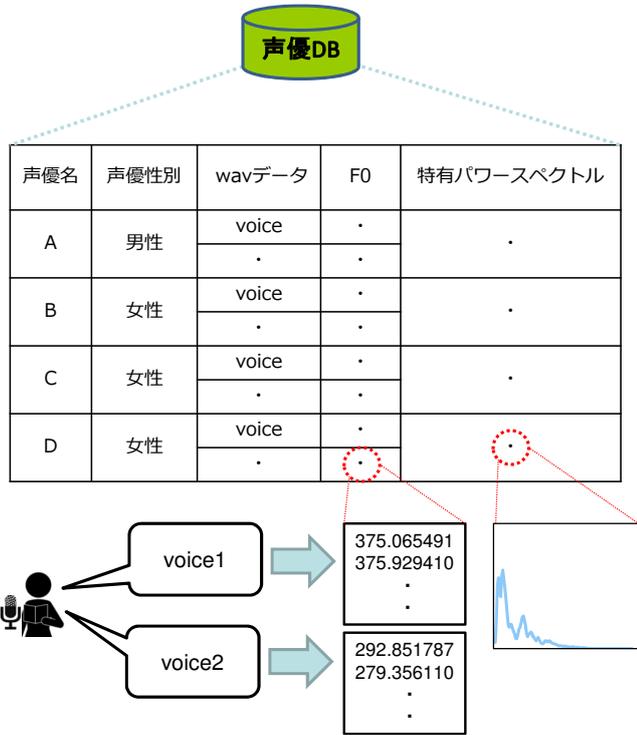


図 2 声優データベースの詳細

### 3. 声優データベース

声優認識、及び、性別判定を行う為には図 2 に示すように声優データベースを整えておく必要がある。

そこで、まず初めに性別判定の為の事前準備として、男女声優それぞれのサンプル音声データから、再生時間の長さ分だけ取得される F0 値のベクトルを声優データベースに登録しておく。一人の声優に対するサンプル音声のデータ数はばらけており、また、ノイズを除く為に BGM や効果音といった雑音が出る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 [6] 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 に設定する。この声優データベースに登録されている声優は予め性別毎に分けられており、音声の持ち主の様々なセリフの音声特徴量が、音声の持ち主と紐付けられて登録されている。

また、声優認識を行う為の音声特徴量の一つとして、特有パワースペクトルも登録しておく。この特徴量は、我々が以前の研究 [3] で定義した個々の声優の声質の特徴である波形データであり、声優毎に一つに特定されている。この特有パワースペクトルを取得する為の音声データは、F0 値を取得する際に使用した音声データと同等のものである。

### 4. 声優の音声を用いた性別判定手法の改良

#### 4.1 従来手法による性別判定

我々は前回、3 章で導入した声優データベースに含まれている F0 値を用いて、確率密度関数の一つである正規分布の計算

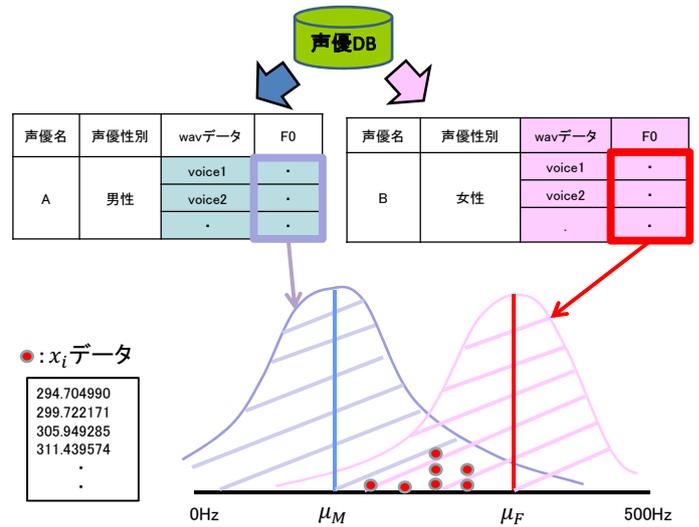


図 3 正規分布による性別判定

式を用いた声優の音声の性別判定手法を提案した [1]。

まず初めに正規分布の計算の為、図 3 に示すように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  $\mu_M$  及び  $\mu_F$ 、分散  $\sigma_M^2$  及び  $\sigma_F^2$  を求めておく。分析対象であるアニメ動画のセリフの音声から得られた F0 値のベクトルを  $x_i (i \in \{1, 2, 3, \dots, T\})$  と表し、その  $x_i$  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。

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

$$PS = \sum_{i=1}^T f(x_i, \mu, \sigma) \quad (2)$$

$x_i$  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。

$$PS_M > PS_F \rightarrow \text{男}$$

$$PS_M < PS_F \rightarrow \text{女}$$

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

#### 4.2 声優の音声を用いた声域分割による性別判定

前回の著者らの研究 [1] で前節の手法による性別判定を行った結果、精度良く性別判定を行えたが、実験対象のある男性声優に対してだけは、全体の精度よりも比較的悪かったと言える。その男性声優の声質は他の男性声優と比べて、主観ではあるが比較的高いと感じた。この考察から、声域が高い男性声優のセリフの F0 の集合が、女性である特徴を表していると従来手法による性別判定システムに判定されたと考えられる。

以上のように、これまでの研究で行った評価実験に対する考察から、男女共に声優はボイストレーニングを行っており、一

般人の声域とは違い広域であると仮説を立てた。この仮説から、性別毎に声域を高低に分割することが可能であると考えた。予め声優データベースに登録されている各声優の音声データを、性別毎に声域の高低に基づいて分類を行っておくことで、分析対象である音声は男性で声質が高い人、女性で声質が低い人であっても性別判定を精度良く行えると考えた。本稿では、声優データベースに登録されている各声優のセリフの F0 値の標本データ（以下、 $d_j$ ）が、声域の高低のどちらかに属するのかを判定する為に、F0 に基づいて分割する手法を提案する。

その手法として、高低における Pitch range (75~500Hz) のパラメータの組み合わせ ( $b_{low}$ ,  $b_{high}$ ) と、 $d_j$  との平均二乗誤差 (MSE) が最小になる組み合わせを探索する。分類の最適解を求める為、図 4 に示すように、ある時点での  $b_{low}$ ,  $b_{high}$  と  $d_j$  との MSE を求めて、 $b_{low}$ ,  $b_{high}$  のうち MSE が小さかった方の値を SUM に加算するという作業を全ての  $d_j$  に対して行う。この工程を全パターンでの組み合わせ ( $b_{low}, b_{high}$ ) に対して行い、SUM の値が最小になった組み合わせの時の分類結果を最適解とする。高低それぞれに分類された  $d_j$  を用いて男女毎における声高の平均値  $\mu_{high}$  と分散  $\sigma_{high}^2$ 、声低の平均値  $\mu_{low}$  と分散  $\sigma_{low}^2$  を求める。しかし、性別各々において分割された後の分散  $\sigma_{high}^2$ ,  $\sigma_{low}^2$  が分割する前の分散  $\sigma_M^2$  あるいは  $\sigma_F^2$  よりも値が大きかった場合、分割は不適切として行わない。

分析対象の音声の性別判定の手法として、前節と同様に、あるアニメのセリフの音声から得られた F0 値のベクトル  $x_i$  が入力された時、男女毎の高低それぞれにおいて数式 (1) と数式 (1) において算出された確率を加算していく数式 (2) を用いて計算を行う。 $x_i$  が  $T$  まで計算された後、男女毎に算出された値で大小比較を行い、最も大きい値であった性別を採択する。

## 5. 声優認識の従来手法

### 5.1 特有パワースペクトル探索アルゴリズム

人それぞれには周波数毎に個人差があると考えて [7,8], 人の声質にはバンド幅毎にそれぞれ違う強さ (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」を特定することを目的とした。

その手法として、各声優のセリフ毎に切り取った動画から取得した周波数スペクトルの時系列パターンをその声優毎にまとめたファイルから、パワースペクトル一つずつをベースとして、もう一つの比較対象である自身の全パターン各々との類似度や相関を計算する。計算した類似度や相関が予め定めた閾値を上回った場合、ベースにしているパターンのカウント数を増やす。全ての組み合わせの計算を終えた後、閾値を上回ったカウント数を一番多く獲得したものを、動画の音声データから得られた声優の特有パワースペクトルと特定する。閾値を上回ったカウント数が同数になった場合、計算から得られた類似度や相関のうち閾値を上回った時だけ加算した平均で比較して、類似度や相関の平均が最も大きかったパターンを動画の音声データから得られた声優の特有パワースペクトルと特定する。

### 5.2 声優認識アルゴリズム

声優データベースに予め用意されている各声優の特有パワー

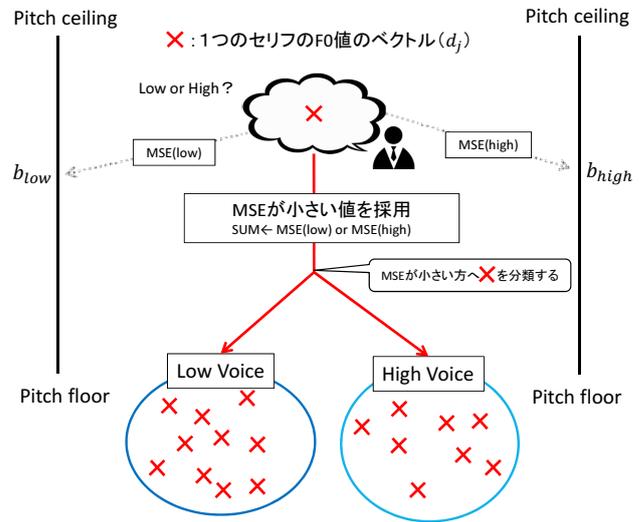


図 4 性別判定のクラスタ分割イメージ

スペクトルと、分析対象である動画から流れた音声の周波数スペクトルとの、それら 2 つの情報を用いて算出された類似度や相関を利用して、その動画内に流れた音声の持ち主の声優を判定する。声優認識する為に、互いのパワースペクトルがどのくらい類似しているかを算出する計算式として、コサイン類似度や相関係数のいずれかを用いた。類似度や相関に対する閾値を予め定めておき、その定めた閾値を上回った特有パワースペクトルを持つ声優にカウントする。最終的に、各声優において加算されたカウント数で比較して、閾値を上回ったカウント数を一番多く獲得した声優をそのアニメ動画から流れた音声の声優であると判定する。閾値を上回ったカウント数が同数の場合、特有パワースペクトルを特定する時と同様に、閾値を上回った時だけの類似度や相関の値をそれぞれ加算した平均で比較して、平均が最も大きかった声優を、そのアニメ動画から流れた音声の声優であると認識する。

## 6. 評価実験

本章では、声優認識の精度向上を狙って性別判定を導入した声優認識手法の精度に関する評価実験を行う。本評価実験では、Web からアニメ動画のキャスト情報を取得して、データベースの声優の人数を各話毎に絞り込めた状態であると仮定する。特有パワースペクトルを特定する時の閾値を  $t_1$  ( $\in \{0.50, 0.51, \dots, 1.00\}$ ) と表し、声優認識に使う閾値を  $t_2$  ( $\in \{0.50, 0.51, \dots, 1.00\}$ ) と表す。また、周波数スペクトル同士の類似度計算式は特有パワースペクトルを特定する時、声優認識を行う時と共にコサイン類似度を用いる。音声信号処理には Python のライブラリの一つである scipy を用いた。評価実験を行う時に設定した内容を以下に示す。

- 音声信号処理
  - サンプリング周波数 : 16000Hz
  - 窓関数 : ハミング窓
  - サンプル数 :  $2^N$  ( $N \in \{9, 10, 11\}$ ) (3 種類)
  - バンド幅 : 0~500Hz, 0~8000Hz (2 種類)

- 声優データベース
  - 声優のセリフの数、長さは多様なデータを収集
  - 公開しているサンプル音声やセリフ集から収集
  - BGM を出来るだけ含まないもの
  - 実験動画のアニメの各話に対応するキャスト情報の分だけを登録
    - \* 1 話：男性 1 人，女性 5 人
    - \* 2 話：男性 1 人，女性 4 人
    - \* 3 話：男性 2 人，女性 2 人
- 実験動画
  - BGM を出来るだけ含まないもの
  - あるアニメタイトルの 3 話分から音声を集
    - \* 1 話：男性 26 音声，女性 26 音声
    - \* 2 話：男性 14 音声，女性 21 音声
    - \* 3 話：男性 17 音声，女性 19 音声
    - \* 総計：男性 57 音声，女性 66 音声

### 6.1 2 種類の閾値の評価

5 章で説明した声優認識の従来手法では，特有パワースペクトルを特定する時の閾値  $t_1$  と声優認識のアルゴリズムで用いる閾値  $t_2$  を設けた．本節の閾値の評価実験では， $t_1, t_2$  の最適な組み合わせのパラメータを，サンプル数の 3 パターンとバンド幅の 2 パターンの組み合わせにおいて，それぞれ求める．実験動画の総計数である男性 57 音声，女性 66 音声に対して，著者らが 4 章で提案した従来の性別判定と声域分割による性別判定，加えて性別判定を行わなかった場合の 3 パターンによる正解数をまとめた割合で正解率を算出した．表 1 にサンプル数とバンド幅の組み合わせ毎に，最も精度が良かった  $t_1, t_2$  の組み合わせを示す．

まず初めに表 1 に注目してみると，一番精度が良かった組み合わせの閾値は小さいパラメータである印象を受ける．この結果から，周波数スペクトルの類似性は閾値を厳しく設定するよりも，大まかに設定した方が精度が良くなると思われる．

また，それぞれの組み合わせの最高精度である正解率に注目すると，どの組み合わせに依っても変わらない精度の結果である．しかし， $N$  の値が小さければわずかではあるが精度が良くなっている傾向がある．このことから，音声データを細かく FFT（高速フーリエ変換）することで精度が上がるのではないかと考えられる．しかし，サンプル数を細かく切り出していくほど処理時間が掛かる為，著者らが理想とするリアルタイムでのアニメ動画の声優認識システムには好ましくないと言える．

### 6.2 性別判定による声優認識の評価

本節では 4 章で提案した性別判定を導入することで声優認識の精度が向上すると考えて精度評価を行う．その為に，性別判定を行わなかった場合と，従来手法の性別判定と，本稿で提案した声域分割による性別判定の 3 パターンを組み込んだ声優認識システムの精度を比較する．また，特有パワースペクトル探索手法や，実験動画から取得される周波数スペクトルと声優データベースに登録されている特有パワースペクトルとの類似性の計算に影響を及ぼすと考えられるバンド幅の長さの違いや，サンプル数の違いに依る影響度にも注目して同時に比較実験を行

う．閾値  $t_1, t_2$  に関しては，前節で求めた最適値にそれぞれ固定して実験を行う．図 5 から図 10 は実験動画対象の男女声優の音声の男性 57 音声，女性 66 音声ずつを声優認識した結果の正解した数の割合をサンプル数毎に求めた結果を表している．

まず初めに，性別判定を組み込んだ声優認識の精度の比較を行う．性別判定を行わない図 5, 6 と，性別判定の声域分割を行わない図 7, 8 を比較すると，性別判定を組み込むことで精度が向上すると仮説を立てたが，仮説に反して精度が低くなった．しかし，性別判定の声域分割を行うことで男性声優に対して精度が格段に向上したが，女性に関しては精度は変わらなかった．考察として，キャスト情報によって絞り込みを掛けた声優データベースの男女比率の影響や，声優の声質といったデータから性別判定の精度に大きな影響を与えたと考えられる．本実験で用いた実験動画のキャスト情報は，男性が一人，女性多数の場合が多く，性別が女性と判定され絞り込みを掛けられても利点が小さい．しかし，それと相反して男性と判定することが出来れば声優は一人に絞られる為，男性声優の精度が上がったと考えられる．また，本実験で用いた女性声優の音声の声域は主観ではあるが低い印象を受けた．このことから，声質による影響を受けて図 7, 8 の声域分割を行わない性別判定手法の精度が好ましくない反面，図 9, 10 の本稿で提案した声域分割を行う性別判定手法では声域の分割を行ったことにより，男性キャラクターの正解率が向上したと考えられる．

次にサンプル数とバンド幅の精度評価を行う．性別判定を行う図 7, 8, 9, 10 の場合，性別判定に重みが傾き，サンプル数やバンド幅の変動による影響が見られない．しかし，性別判定を行わない図 5, 6 によると，サンプル数やバンド幅が大きな影響を表しているのが分かる．男性声優の場合，サンプル数  $2^N$  が大きいほど精度が良くなっていく傾向が見られるが，女性声優の場合は精度が下がる傾向が見られる．これは，サンプル数の変動により周波数分解能が上がることで男性声優の精度が上がり，一方で周波数分解能が下がることで女性声優の精度が上がることを表している．細かく Hz の間隔を解析することで，低周波数の値をより多く解析出来るようになる為，男性声優の精度が上がったのではないかと考えられる．また，バンド幅の違いに依る影響は小さいが，図 6 において男性，女性の認識精度が図 5 と比較して取束していない為，バンド幅は 500Hz までを声優の音声情報として用いた方が好ましいと考えられる．

表 1 閾値の組み合わせに依る正解率の評価

		バンド幅	
		0~500Hz	0~8000Hz
サン プル 数	$N = 9$	0.444 ( $t_1 = 0.67, t_2 = 0.62$ )	0.439 ( $t_1 = 0.65, t_2 = 0.58$ )
	$N = 10$	0.404 ( $t_1 = 0.54, t_2 = 0.50$ )	0.415 ( $t_1 = 0.67, t_2 = 0.50$ )
	$N = 11$	0.423 ( $t_1 = 0.57, t_2 = 0.61$ )	0.428 ( $t_1 = 0.61, t_2 = 0.51$ )

閾値  $t_1$ : 5.1 節の特有パワースペクトル探索アルゴリズムで用いる  
閾値  $t_2$ : 5.2 節の声優認識アルゴリズムで用いる

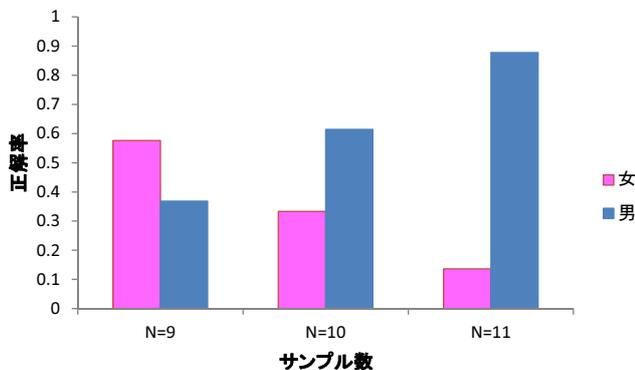


図5 性別判定なし (バンド幅 0~500Hz)

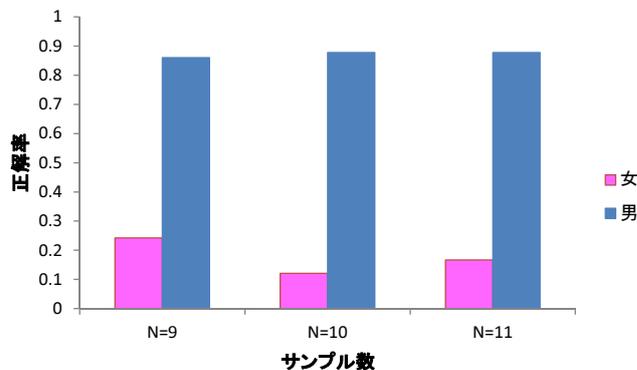


図9 性別判定あり (声域分割あり, バンド幅 0~500Hz)

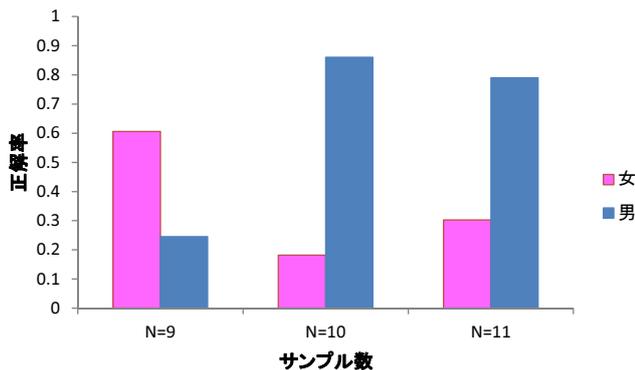


図6 性別判定なし (バンド幅 0~8000Hz)

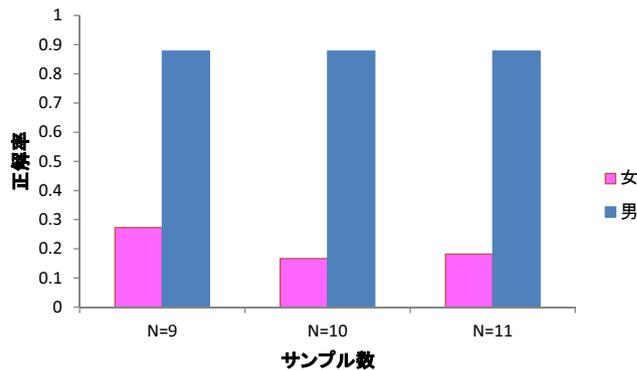


図10 性別判定あり (声域分割あり, バンド幅 0~8000Hz)

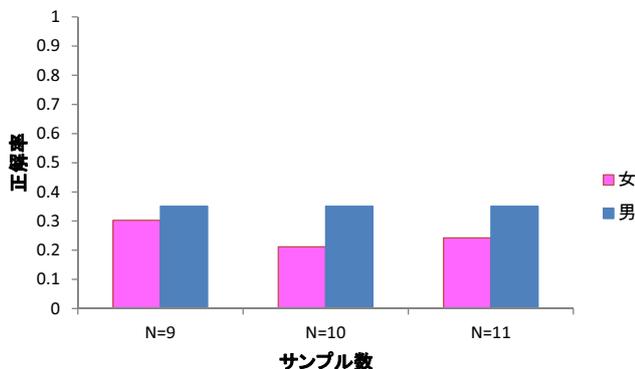


図7 性別判定あり (声域分割なし, バンド幅 0~500Hz)

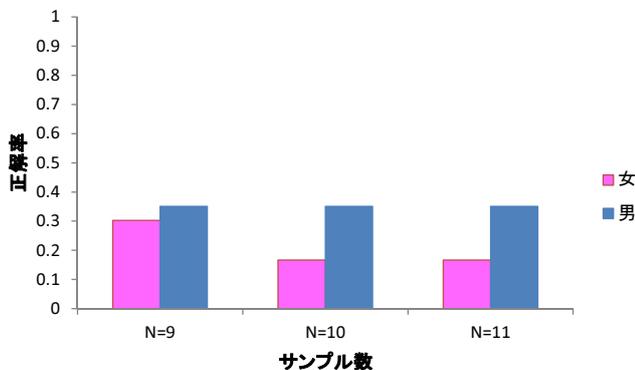


図8 性別判定あり (声域分割なし, バンド幅 0~8000Hz)

## 7. 今後の課題

今後の課題として、声優データベースと性別判定の分割手法の改善が挙げられる。これまで提案して来た手法の声優データベースの構造では、声優一人に対して予め特有パワースペクトルを一つだけ登録していたが、声優一人に対して様々な声質の特徴があると考えて、複数のセリフの特有パワースペクトルを予め登録しておく手法が考えられる。また、本稿で提案した声域分割手法が正しく行われたかを詳しく検証する為に、既存で提案されているSVMやGMMなどの分割手法と比較して評価を行う必要がある。

## 文 献

- [1] 榮田 基希, 服部 峻, “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,” 電子情報通信学会 情報ネットワーク研究会, 信学技報, Vol.117, No.326, pp.55-60 (2017).
- [2] 榮田 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).
- [3] 榮田 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).
- [4] まうまう☆, <https://www.mau2.com/> (2017).
- [5] Wikipedia, <https://www.wikipedia.org/> (2017).
- [6] Praat, <http://www.fon.hum.uva.nl/praat/> (2017).
- [7] 早川 昭二, 板倉 文忠, “音声の高域に含まれる個人性情報を用いた話者認識,” 日本音響学会誌, 51 卷 11 号, pp.861-868 (1995).
- [8] 横山 雅夫, “音声に含まれる個人性情報,” 福島大学行政社会学会, 行政社会論集, 第 4 巻第 3 号, pp.96-113 (1992).