台詞内容と声の印象に合致した最適声優リコメンド方式

青木 絢太† 岡田 龍太郎‡ 峰松 彩子† 中西 崇文‡

†武蔵野大学データサイエンス学部データサイエンス学科 〒135-8181 東京都江東区有明 3-3-3

E-mail: †s2122070@stu.musashino-u.ac.jp, ‡{ryotaro.okada,ayako.minematsu, takafumi.nakanishi}@ds.musashino-u.ac.jp

あらまし 本稿では，台詞内容と声の印象に合致した最適声優リコメンド方式について示す．本方式は，予め準備された声優の声の音声データからなる声優データベースを対象として，各音声データから，その話された内容を示すテキストデータとその声質の印象を表す印象タグを抽出したメタデータベースを構築した上で，ユーザが与える台詞テキストデータと声質の印象タグからなるクエリによって，合致した最適声優のリコメンドを実現する．本方式は，台詞内容による類似度計量方式と声質印象による類似度計量方式を統合することにより，声優の特徴を捉えたリコメンドが可能となる．本方式の実現により，台本が存在する作品において，その台本の台詞内容と声質の印象から，それに合致した声優のキャスティングが可能となる．

キーワード 台詞，印象，音声特徴

1. はじめに

古くから日本ではマンガ，アニメなどのサブカルチャーが発展してきた．近年では海外でも非常に高い人気を持っており，海外に誇る日本の一大分化となっている．そんなマンガをアニメ化する際に必要なのがキャラクターの声を演じる声優である．声優はアニメやゲームのみならず，ラジオやナレーションなどさまざまな場面にキャスティングされる．サブカルチャーが発展してきた日本では 2021 年時点で 1500 人以上のプロが存在する．声優志望者も年間で 3 万人以上いるとされ，キャスティングにも多様な選択肢がある．

一般に声優のキャスティングは台本や台詞の確認の後，声優を探索しそれぞれの声優のプロフィール，サンプル音源などを参考にオーディションを実施する．ここで，先にも述べたように近年の声優の多様化により今までよりもさらにイメージに近い声優をキャスティングできる可能性がある．それと共に，多様な声優の中から最適な声優を発掘しキャスティングを行う難易度も高くなっていると考えられる．

台詞や場面，イメージに合致した声優をキャスティングすることは，作品や世界観をよりリアルに表現し相手に伝えることができるため，作品の質や魅力の向上に繋がる．また，キャスティングされた声優自身も自分の声にあった演技を行いやすくなり，より声優自身の魅力も活かすことができる．これらによって作品全体のプロモーションなど幅広いメリットがある．

本稿では，声優が実際に読み上げた台詞に注目する．本方式は，台詞ごとに込められた感情と声優の音声特徴量を分析することで，実際に使用する台本とキャスティングしたい声の印象から最適な声優のリコメンドを可能とする．

本稿は次の通り構成される．2 節では，関連研究について紹介する．3 節では，本方式である台詞内容と声の印象に合致した最適声優リコメンド方式について述べる．4 節では，本方式を実現する実験システムを構築し，5 節で本稿をまとめる．

2. 関連研究

本章では，本方式に関連する研究について挙げる．ここでは主に台詞や台本，音声特徴量を入力とした読み上げに関する研究や，最適なキャスティングに関連する研究について挙げる．

2.1 漫画の読み上げ音声に関する研究

WANG[1]らは，デジタル化された漫画を入力とし，キャラクターの視覚的な見た目と合致する発話を合成し読み上げさせている．まず，漫画の構成などの見た目を分析し，コマ，吹き出し，テキスト，キャラクタを抽出する．抽出した要素を物語順に並べ，登場人物と関連づける．さらに，登場人物の性別や年齢，アイデンティティを認識することで漫画の内容を総合的に把握している．こうして抽出されたコマの情報をもとに音声を合成している．

2.2 ゲームキャラクタと声質の傾向分析

酒井[2]らはゲームキャラクタにおける声質印象と音響特徴量を結びつけることから最適な声優をキャスティングする研究を行っている．まず，ゲームキャラクタの台詞の音響特徴量を抽出した後，被験者に対して印象値の調査を行う．こうして得られた音響特徴量と印象値に自己組織化マップを用いて印象値を与えることで音響特徴量を推定し，新しいキャラクタに最適な声優のリストを作成している．

最後にキャラクタのテキストデータを Word2Vec[3]を用いて多次元ベクトルとして表現し，主成分分析によってそれらの距離を可視化している．

2.3 アニメ動画より声優の性別の認識に関する研究

榮田ら[4]はアニメ動画の台詞音声やその声優を推定する研究を行っている．まず，Wikipedia[5]等のキャラクタを紹介する Web テキストから登場人物の性別を推定する．この際にあらかじめ男性，女性を判別するための辞書を作成している．Web テキストに対して CaboCha[6]を用いて形態素解析によって単語抽出を行い，その出現回数に関数を作成して重みづけを行い，性別を判定している．その後に,「アニメ声」であっても男女によって基本周波数に違いが生じるという仮説のもと，アニメ動画を入力としキャラクタの音声の基本周波数を抽出し，声優データベースに登録された標本データで作成された確率密度関数を用いて性別を判定している．

2.4 本研究の位置付け

WANG[1]らは漫画から得られる視覚的情報をもとに音声が割り振られていないものに対して機械的に音声を合成することを試みている．音声情報がないものに関して音声を付与しようとする目的，その入力としてテキスト等の音声に関わる情報を使用する点は本研究との共通点だと考えられる．その上で本研究では，人間的な声を割り当てることに注目している．テキスト情報などその物語を構成する情報だけでなく，人間の声における印象を用いることでより人間の感覚に合致した音声の割り当てを実現することを目指す．

酒井ら[2]は音響特徴度とその音声の印象値を結びつけている点，特定のキャラクタを演じるのに最適な声優の推薦を目指している点で本研究と大きく類似している．本研究を実現することが社会的な需要を持っていることもわかる．そこで，本研究ではより音声特徴量と人間が感じる音声印象をより正確に結びつけるため，MFCC 等の細かく多様な情報を得ることができる音声特徴量を使用している．また，酒井らはキャラクタごとの台詞の類似度を可視化することで音響特徴量と声質印象の自己組織化マップと併せてユーザにキャスティング判断をすることを促しているが，本研究では台詞データを声優の候補を自動的に提示するシステムの中に一つの指標として直接組み込むことを試みている．これによって，よりユーザの意志を反映した自動推薦システムを作成することを目指す．

榮田ら[4]はアニメ動画等の声優を探す手間を解消するという目的のもと音声から声優を特定する研究を行なっており，その一部として性別を特定する方式をここで述べている．キャラクタの性別を判定するために台詞データ，声優の性別を判定するために基本周波数をそれぞれ用途を分けて使用している．本研究では台詞データと声優の音声特徴量を結びつけて使用することで，性別も含めた総合的な声優推薦を実現することを目指す．

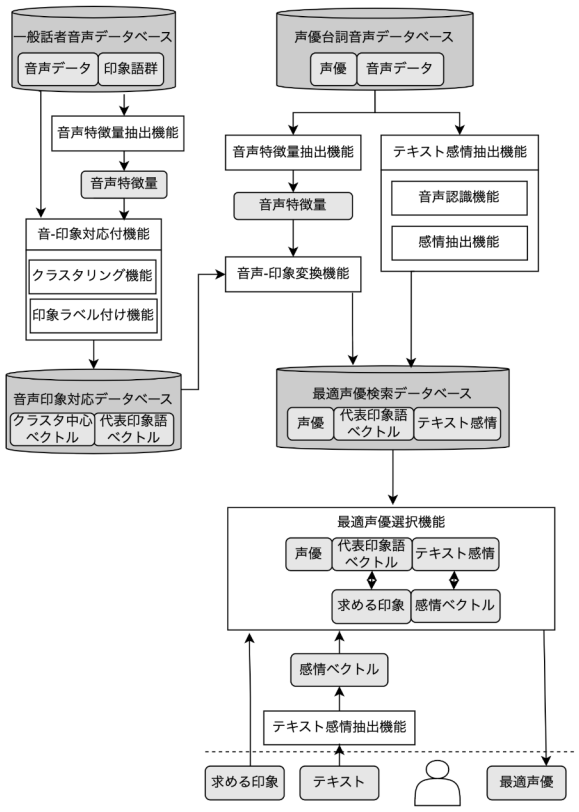

図 1 本方式の全体図

3. 台詞内容と声の印象に合致した最適声優リコメンド

本章では，提案方式である台詞内容と声の印象に合致した最適声優リコメンド方式について提示する．

3.1 全体像

本節では，本研究における提案手法の概要を述べる．本研究では声優を決定するに当たって台詞の感情による分類，印象を決定するに当たって音声特徴量を用いた推定が有効であるという仮説の元進めていく．提案システムの全体像を図 1 に示す．本システムは，声優台詞データベース ，音声印象対応データベース，音声特徴量抽出機能，テキスト感情抽出機能，音声印象変換機能，最適声優検索データベース，最適声優選択機能で構成される．音声特徴量抽出機能は，Librosa[7]を用いて，声優台詞データの音声特徴量を抽出する機

表 1 付与した印象の例

ファイル名	付与した印象
生島一二三.mp3	男性的，落ち着いた声，中年の声
箱森ゆめ.mp3	女性的，明るい声，通る声
綾瀬裕子.mp3	子供っぽい声，かわいい声，明るい声

能である．テキスト感情抽出機能は，声優台詞データセットから Whisper[8]を用いて文字起こししたテキストデータを ML-Ask[9]を用いて，台詞を 10 個の感情に分類するものである．テキスト感情分類データ群はそれによって声優台詞データセットが 10 個の感情にクラスタリングされたものである．音声-印象変換機能は，音声特徴量抽出機能によって得られた音声特徴量をあらかじめ作成された音声印象対応データベースを通してその声が持つ印象語に変換するものである．最適声優選択機能は最適声優検索データベース内のテキスト感情と音声印象に対して，それぞれユーザーが入力したものとの cos 類似度計算を行い，二つの cos 類似度を掛け合わせた値を計算する機能である．声優台詞データベースは，81 プロデュース[10]の公式サイトより得られる各所属声優のボイスサンプル 325 個によって構成される．音声印象対応データベースは，Mozilla が公開する多話者データセットである Mozilla Common Voice[11]と声優ナレーションスタジオ[12]が公開するサンプル音源から抽出した 500 の音声データに 43 語の印象語を複数付与したものである．本研究の目的は，ユーザが入力したテキストとその印象に対して合致する声優をリコメンドすることである．ユーザが読ませたい台本を入力する際に同時に求める声優の声質印象を入力し，それに合致した声優を推薦することで同じ台詞でもどのように読むのか，どんな場面で読ませたいのかという状況に応じた細かな配役に対応したリコメンドを可能とする．

3.2 音声印象対応データベース

本研究では，Mozilla の Mozilla Common Voice[11]データセットとナレーション声優スタジオ[12]が公開する声優のサンプル音源を用いる．この中から 500 の話者を抽出し，それぞれの音声に対して表 2 に示す声質を表す印象語の中から複数を付与しデータベースを作成する．実際に音声に印象語を付与した例を表 1 に示した．使用する印象語は木戸ら通常発話の声質に関連した日常表現語の抽出[12]で使用されている 25 語に独自に 18 語を追加した 43 語である．このデータベースに対し，3.3 節と同様に音声特徴量抽出を行い k-means 法で 10 個のクラスタに分類するクラスタリングを行う．クラスタの数については，印象語が 43 語あることことから，各クラスタにそれぞれの特徴的な印象語が当てはまるためにはある程度のクラスタ数が必要であるという仮説のもと，まずはクラスタ数を 10 個としクラスタリングを行った．クラスタの数の議論については 4 章の実験で検証することとする．これらの事前に印象語が付与され，音声特徴量によってクラスタリングされたデータに対して tf-idf を用いて，クラスタ内の音声データに付与された印象語のクラスタごとの出現頻度と回数から計算を行い，クラスタの傾向を表す印象語ベクトルを作成し，これを音声印象対応データベースとする．

表 2：付与した印象語の一覧

かわいい声	クールな声	優しい声	明るい声
澄んだ声	色っぽい声	機械的な声	張りのある声
通る声	鼻の詰まった声	かすれた声	がらがら声
だみ声	くもった声	潰れた声	ドスの効いた声
恐怖を感じる声	迫力を感じる声	気分が良い声	気分が悪い声
子供っぽい声	若い声	中年の声	老人の声
渋い声	男性的な声	女性的な声	中性的な声
金切り声	高い声	弱々しい声	ハキハキした声
生き生きした声	響きのある声	感情の薄い声	落ち着きのある声
落ち着きのない声	品のある声	裏のありそうな声	疑問のある声
舌足らずな声	太い声	逞しい声	

3.3 音声特徴量抽出機能

本節では，声優台詞データセットから音声特徴量を抽出する方法について述べる．Librosa を用いて各声優台詞データからゼロクロス率，MFCC(20 次元)，スペクトラルセントロイド，スペクトラルバンド幅，の計 24 次元の特徴量を抽出する．

ゼロクロス率とは，振幅の正負が何回入れ替わるかを表すものである．この値は音声の雑音感が高い時に大きくなる傾向にある．声優の声において雑音感が低い音声ほどクリアな声となり，特徴量として重要であると考えた．

MFCC とは，人間の声の音声波形スペクトラムに対

表 3 ML-Ask によって抽出した感情

ファイル名	akiyoshi_tooru02.mp3
emotion	{‘iya’:[‘面倒’]}
orientation	NEGATIVE
activation	NEUTRAL
emotion	None
intention	1
intensifier	[‘じゃ’,’んで’]
representative	‘iya’:[‘面倒’]

表 5 各クラスタのデータ数

クラスタ番号	データ数
1	105
2	129
3	67
4	143
5	100
6	42
7	83
8	118
9	55
10	50

して，フーリエ変換を複数回行ったものをヒトの周波数知覚特性を考慮した重み付けした特徴量である．ヒトの聴覚上重要な特徴量が重み付けられて表現されるため，声優の声質の印象を探る上で重要な指標であると考えられる．

スペクトラルセントロイドは，スペクトルの重心がどこにあるかを示すものである．知覚的に音の明るさの印象と強い相関があると言われている．相手に声が明るいという良い印象を与えることは声優において重要な指標の一つになると考えられるため，使用する．

スペクトラルバンド幅は音声の周波数成分の数値の幅である．音声の周波数や音圧によって変動するとされている．音圧は人間の声を聞いた時の印象において大きく関わると考えられるため，選定した．

3.4 テキスト感情抽出機能

本節では，声優台詞データセットから発話した内容の感情を抽出する方法について述べる．まず，各声優台詞データの冒頭 30 秒について Wisper を用いて文字起こしを行い，得られたテキストデータに関して ML-Ask を用いて喜，怒，昂，哀，好，怖，安，厭，驚，恥の 10 感情のベクトルを抽出し，2L 正規化を行う．表 4 には，ある音声に対しての文字起こしの結果を，表 3 には ML-Ask による分析の結果を示す．文字起こしの結果を見ると所々で誤字が見られるが，今回はその他の入力値に重点を置くため，そのままの状態で感情を抽出した．

3.5 音声印象変換機能

本節では，声優台詞データに対して，その印象語を抽出する方法について述べる．3.2 節で述べた音声印

表 4 文字起こしの結果

ファイル名	akiyoshi_tooru02.mp3
テキスト	親じがいなくなったとたんおじさんから一緒に住まないかって言われたその後もずっと面倒を見てくれたけど一緒には進めなかったなんで勝ってそれは遠慮だけが理由じゃないこの家はおれにとって宝物なんだ

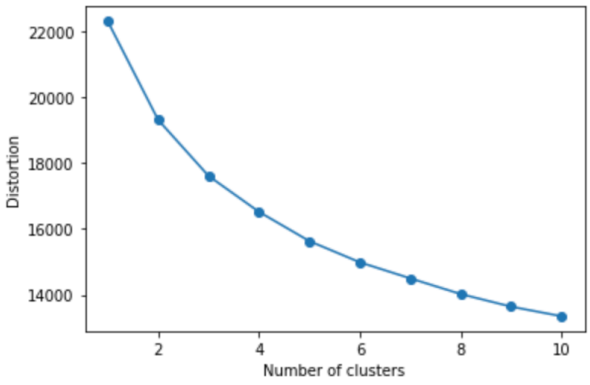

図 2 エルボー法の結果

象対応データベースに新規データとして各声優台詞データを入力し，各クラスタとの距離計算を行い一番近いクラスタの印象語ベクトルを付与し，2L 正規化を行う．

3.6 最適声優検索データベース

本節では，最適声優検索データベースの構成について述べる．各データは声優の名前を保持し，それぞれに対し 3.4 節より抽出された 10 のテキスト感情ベクトル，3.5 節より抽出された 43 の音声印象語ベクトルが格納され，全体で 325×54 のデータベースである．

3.7 最適声優選択機能

本節では，入力されたデータを元に最適声優検索データベースから最適な声優を選択する方法について述べる．ユーザーはテキストと求める印象語を任意の数入力する．入力されたテキストに対しては 3.4 節と同様にテキスト感情抽出機能を用いてテキストの感情ベクトルを抽出する．こうして得られた入力データと最適声優検索データベースのそれぞれの感情ベクトルと印象語ベクトルの間で cos 類似度計算を行う．その後，得られた感情ベクトルと印象語ベクトルの cos 類似度を掛け合わせ，これを最適声優を選択する際の値とする．この値を降順にソートしてユーザーに提示することとする．

4. 実験

本章では，本手法の実験内容と結果，考察について述べる．

4.1 節では，実験環境について述べる．4.2 節では，

表 6 クラスタ 1~5 の代表印象語

クラスタ1		クラスタ2		クラスタ3		クラスタ4		クラスタ5	
代表印象語	重み	代表印象語	重み	代表印象語	重み	代表印象語	重み	代表印象語	重み
男性的な声	0.601	男性的な声	0.422	女性的な声	0.417	情緒の薄い声	0.486	女性的な声	0.546
機械的な声	0.324	落ち着きのある声	0.308	落ち着きのある声	0.376	女性的な声	0.420	逞しい声	0.317
澄んだ声	0.285	優しい声	0.296	優しい声	0.344	クールな声	0.322	ドスの効いた声	0.292
中年の声	0.269	情緒の薄い声	0.279	情緒の薄い声	0.303	かすれた声	0.252	クールな声	0.288
女性的な声	0.245	クールな声	0.254	クールな声	0.275	ハキハキした声	0.235	恐怖を感じる声	0.234
クールな声	0.185	澄んだ声	0.211	男性的な声	0.275	中年の声	0.215	中年の声	0.199
優しい声	0.185	子供っぽい声	0.211	ハキハキした声	0.251	落ち着きのある声	0.202	子供っぽい声	0.199
落ち着きのある声	0.168	恐怖を感じる声	0.206	通る声	0.251	弱々しい声	0.184	優しい声	0.192
潰れた声	0.161	逞しい声	0.186	中年の声	0.229	子供っぽい声	0.184	澄んだ声	0.144
張りのある声	0.153	中性的な声	0.172	裏のありそうな声	0.152	機械的な声	0.184	色っぽい声	0.144

表 8 クラスタ 1~5 におけるクラスタ中心に近いデータ

クラスタ1		クラスタ2		クラスタ3		クラスタ4		クラスタ5	
ファイル	付与した印象語	ファイル	付与した印象語	ファイル	付与した印象語	ファイル	付与した印象語	ファイル	付与した印象語
075.mp3	機械的な声 ハキハキした声 落ち着きのある声 男性的な声	047.mp3	優しい声 かすれた声 落ち着きのある声 男性的な声	137.mp3	優しい声 ハキハキした声 女性的な声	151.mp3	機械的な声 情緒の薄い声 落ち着きのある声 女性的な声	272.mp3	逞しい声 ドスの効いた声 女性的な声
243.mp3	機械的な声 男性的な声	092.mp3	クールな声 優しい声	240.mp3	落ち着きのある声 優しい声 通る声	187.mp3	ハキハキした声 女性的な声	220.mp3	クールな声 優しい声 女性的な声
252.mp3	澄んだ声 恐怖を感じる声 老人の声	041.mp3	優しい声 恐怖を感じる声 子供っぽい声	278.mp3	優しい声 情緒の薄い声 裏のありそうな声	066.mp3	弱々しい声 情緒の薄い声	170.mp3	優しい声 恐怖を感じる声 子供っぽい声

音声印象対応データベースにおけるクラスタごとの印象語の値の結果とクラスタ数の妥当性，クラスタごとの音声特徴量と印象語の関係性や特徴について議論する．4.3 節では構築した方式に実際に新規データを入力し，その精度についてどのような結果が得られるかを実験する．4.4 節では，本研究による実験結果について考察を行う．

4.1 実験環境

本節では，3 章で提案したシステムを実装し実際の台詞と印象語を入力した．

本研究では，音声特徴量を抽出するにあたり，Librosa とテキストから感情を抽出するために，Wisper, ML-Ask を用いた．

入力したデータとデータベース検索の類似度を求める際には cos 類似度を使用している．

4.2 実験 1(音声印象対応データベースを対象とした音声特徴量によるクラスタリングの検証とそのクラスタごとの印象語抽出の確認)

本節では，3.2 節で述べたように音声印象対応データベースを作成する際に決定したクラスタの数の妥当性と，実際に作成されたクラスタがどのような印象語と音声特徴量を持ち，全体として各クラスタが差別化できているのか，どのような傾向を持つのかについて検証を行う．

まず，クラスタの数を決定する際によく用いられるエルボー法を用いて検証を行った．エルボー法はクラスタリングにおいて，最適なクラスタ数を求めるための手法である．一般にグラフが急激に降下した点が最適なクラスタ数と見なす．その結果を図 2 に示す．図 2 から明確にグラフが急激に降下しているといえる場所が現れず，エルボー法ではクラスタ数を決定することはできなかった．そのため，3.2 節で述べた 43 語の印象語が各クラスタに現れるためにはある程度のクラスタ数が必要であるという仮説のもと，このままクラスタは 10 個で検証を続ける．

よって，ここからは実際にクラスタごとのデータを確認し，クラスタリングが適当であるかを検証する．まずは各クラスタのデータ数を表 5 に示す．クラスタごとに差は見られるがどのクラスタもデータ数に極端な差はなかった．次にクラスタごとにクラスタにおける印象語上位 10 個とその tf-idf の値，およびクラスタからクラスタ中心に近い音声を 3 つ抽出し．どのような印象タグが付与されているのかを確認する．表 6 にクラスタ 1~5 における代表印象語，表 7 にクラスタ 6~10 における代表印象語を提示した．また，表 8 にクラスタ 1~5 における抽出データに付与されている印象

表 7 クラスタ 6~10 の代表印象語

クラスタ6		クラスタ7		クラスタ8		クラスタ9		クラスタ10	
代表印象語	重み	代表印象語	重み	代表印象語	重み	代表印象語	重み	代表印象語	重み
金切り声	0.759	男性的な声	0.604	女性的な声	0.563	生き生きした声	0.468	子供っぽい声	0.642
かすれた声	0.314	中年の声	0.365	中年の声	0.326	かわいい声	0.461	かわいい声	0.579
かわいい声	0.249	情緒の薄い声	0.302	迫力のある声	0.262	子供っぽい声	0.383	落ち着きのない声	0.262
くもった声	0.249	鼻の詰まった声	0.298	落ち着きのある声	0.238	ハキハキした声	0.35	明るい声	0.213
機械的な声	0.249	響きのある声	0.242	色っぽい声	0.218	女性的な声	0.350	女性的な声	0.146
弱々しい声	0.166	優しい声	0.220	老人の声	0.199	色っぽい声	0.230	金切り声	0.098
がらがら声	0.151	落ち着きのある声	0.200	逞しい声	0.192	落ち着きのある声	0.158	生き生きした声	0.078
子供っぽい声	0.138	クールな声	0.165	子供っぽい声	0.181	かすれた声	0.117	だみ声	0.064
通る声	0.101	通る声	0.150	だみ声	0.175	落ち着きのない声	0.104	機械的な声	0.064
恐怖を感じる声	0.101	裏のありそうな声	0.121	クールな声	0.175	鼻の詰まった声	0.104	舌足らずな声	0.064

表 9 クラスタ 6~10 におけるクラスタ中心に近いデータ

クラスタ6		クラスタ7		クラスタ8		クラスタ9		クラスタ10	
ファイル	付与した印象語	ファイル	付与した印象語	ファイル	付与した印象語	ファイル	付与した印象語	ファイル	付与した印象語
036.mp3	金切り声 かすれた声	139.mp3	高い声 情緒の薄い声 男性的な声	173.mp3	迫力のある声 ハキハキした声 女性的な声	047.mp3	生き生きした声 かわいい声 子供っぽい声	007.mp3	落ち着きのない声 明るい声 女性的な声
015.mp3	かわいい声 子供っぽい声 かすれた声	124.mp3	鼻の詰まった声 男性的な声 中年の声	222.mp3	落ち着きのある声 優しい声 女性的な声	165.mp3	ハキハキした声 女性的な声 舌足らずな声	284.mp3	子供っぽい声 かわいい声
002.mp3	機械的な声 弱々しい声 かすれた声	146.mp3	中年の声 落ち着きのある声 男性的な声	225.mp3	優しい声 色っぽい声 中年の声 女性的な声	204.mp3	ハキハキした声 子供っぽい声 かわいい声	204.mp3	生き生きした声 落ち着きのない声 子供っぽい声

語，表 9 にクラスタ 6~10 における抽出データに付与されている印象語を提示する．全体的にどのクラスタでも，クラスタ中心に近いデータに付与された印象語がクラスタの代表印象語となっていることがわかる．特に抽出したデータは複数の上位の代表印象語を保持しているため，今回はクラスタがそれぞれの特徴をもつように上手く分類されていると評価した．

4.3 実験 2(最適声優選択機能による検索結果出力例)

本節では，3 章で提案したシステムに実際に新規データを入力しその出力結果を示す．入力したデータを表 10 に，声優名，印象語と感情の類似度及び最終的な類似度の計算結果を表 11~13 に示す．

今回は 3 つのクエリを用いて実験を行った．クエリ 1 とクエリ 2 では台本を変えずに求める印象のみを変更した．同じ台本でも求める印象を変えることでイメージの異なる声優を出力できるのかを検証する．クエリ 1 とクエリ 3 では求める印象は変えずに入力する台本を変更した．同じ印象を与える声でも，台本を変えることで検索される声優にどのような違いが出るかを検証する．

その結果，表 11~13 のような結果が得られた．それぞれのクエリで全く異なる声優が推薦された．クエリ 1 とクエリ 2 を比較すると，クエリ 1 では全て女性声優，クエリ 2 では全て男性声優が推薦されている．またクエリ 3 でも，全て女性声優が推薦されている．数値的な特徴としては，印象語類似度に関しては多様な値が得られているが，感情類似度では二つの値しか得られていないことがわかる．

4.4 考察

本章では，クラスタの数とクラスタ内のデータの妥当性の検証や実装した方式に対しての新規データ付与を行い，クエリの対照実験を行った．

音声印象変換対応データベースに使用したクラスタの数についてはエルボー法で検証を行ったが最適な数は見つからなかった．よって，クラスタ内のデータを抽出して照らし併せた所印象語を上手く反映しきれていないクラスタも存在することがわかった．それらのクラスタの特徴としてはどれも分類されたデータ数が 100 より少ないということが挙げられる．ここから，音声印象変換対応データベースにおいて，クラスタの代表印象語がそのクラスタの印象語の特徴を反映するには最低でも 100 程度のデータが分類される必要があると考察できる．

次に，実際に 3 つのクエリを入力し対照実験を行った．その結果，3 つのクエリに応じて全く異なる声優を得ることができた．特にクエリ 1 では全て女性，ク

表 10 入力したデータ

	クエリ 1	クエリ 2	クエリ 3
台本	君の歌を待っている人がいるんだよ君の声を好きになったんじゃない君の、その姿にみんな惚れてついてきてるんだよ	君の歌を待っている人がいるんだよ君の声を好きになったんじゃない君の､その姿にみんな惚れてついてきてるんだよ	なんということだ。 僕は何度も考えを改めようと思った｡どうしてもその方向に推理を進めていくのが嫌だったんだ。
求める印象	生き生きした声 ハキハキした声	落ち着いた声 渋い声	生き生きした声 ハキハキした声

表 11 クエリ 1 における実験 2 の結果

声優名	印象語 類似度	感情 類似度	総合 類似度
goda aoi(女)	0.684	1.0	0.684
eda takuhiro(男)	0.578	0.707	0.409
nitta saki(女)	0.512	0.707	0.362

表 12 クエリ 2 における実験 2 の結果

声優名	印象語 類似度	感情 類似度	総合 類似度
sensaki kodai(男)	0.459	1.0	0.459
ishino ryuzo(男)	0.381	1.0	0.381
sakai keikou(男)	0.534	0.707	0.378

表 13 クエリ 3 における実験 2 の結果

声優名	印象語 類似度	感情 類似度	総合 類似度
sonoda rei(女)	0.574	1.0	0.574
takase yu(女)	0.427	1.0	0.427
fukuo yui(女)	0.405	1.0	0.405

エリ 2 では全て男性となったことから求める印象によって同じ台本でも性別から全く異なる特徴を持つ声優が適していることがわかる．しかし，感情類似度の値を見ると，値の違いが少ないという問題点も見ることができた．これは，各台詞にたいして ML-Ask で感情を付与した際に 10 個の感情から 1~2 個しか付与されておらず,他の 感情の値は 0 となっていることが原因だと考察できる．実際にクエリ 3 による検索では感情類似度が全て 1.0 であり，検索の順位はほぼ印象語類似度に依存している．これに関しては，文章の類似度検索において ML-Ask から得られる変数をより加えるか，word2vec などで文章そのものの類似度を測るという方式に変更するといった改善策が挙げられる.また，この感情類似度を大きな絞り込みとして利用するということも可能だと考察する．今回は Whisper で文字起こしを行ったものを誤字などを無視してそのまま使用したがこの部分を変更することでより精度の向上が期待できる.

5. おわりに

本稿では，台詞内容と声の印象に合致した最適声優リコメンド方式について述べた．本研究では，声優のサンプルボイスに関して台詞文章からの感情抽出とその音声特徴量から声質の印象語推定を行うことで，読ませたい台詞や台本と求める声質印象を入力した際にそれに合致する最適な声優をリコメンドする方式を実現した.

この方式を利用することで，アニメや CM 等のキャスティング時に膨大な声優の中から台本を読む上で理想的な候補を絞り込むことができる.

今後の課題として，感情の類似度を求めた際のその値の差異の出にくさ，感情の類似度の値によって全体の類似度が大きく左右されてしまう問題を解決すること，さまざまな配役をこなせる声優のため 1 人の声優にあらかじめ結びつけておく台詞の拡張などが挙げられる.

参 考 文 献

[1] YUJIA WANG , WENGUAN WANG , WEI LIANG , LAP-FAI YU “Comic-Guided Speech Synthesis” ACM Trans. Graph., Vol. 38, No. 6, Article 187. Publication date: November 2019.I. Tanaka and J. Suzuki, “Web and Database Technologies”, Proc. of ACM SIGMOD, pp. 10-22, 2010.

[2] 酒井 えりか，伊藤 彰教，伊藤 貴之， “ゲームキャラクタと声質の傾向分析”， ITE Technical Report Vol. 41， No. 12 AIT2017-99(Mar. 2017)

[3] Tomas Mikolov , word2vec , https://github.com/dav/word2vec

[4] 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”， IEICE Technical Report AI2017-16 (2017-11)

[5] Wikipedia，https://en.wikipedia.org/wiki/Main_Page

[6] 工 藤 拓 , CaboCha/ 南 瓜 , https://github.com/taku910/cabocha

[7] Librosa， https://github.com/librosa/librosa

[8] OpenAI , whisper , https://github.com/openai/whisper

[9] Ykino Ikegami , ML-Ask ,

https://github.com/ikegami-yukino/pymlask/blob/master/mlask/mlask.py
[10] 81 プロデュース, https://www.81produce.co.jp/
[11] Mozolla , Mozolla Common Voice , https://dev.commonvoice.allizom.org/ja
[12] ナレーション声優スタジオ https://wis2.jp/
[13] 木戸博, 粕谷英樹, “通常発話の声質に関連した日常表現語の抽出”, 日本音響学会誌, 55(6), pp.405-411,1999.