声に対する印象を用いた合成音声ライブラリ探索システムの提案

情報　太郎　　情報　花子
情報大学情報学部

1　はじめに

人の歌声や喋り声を人工的に再現する音声合成ソフトは数多く存在しており，それらソフトのほとんどが複数種類の声を切り替えて使用できる．また，その中でもいくつかのソフトでは個人が声の元となる合成音声ライブラリを作成し，第三者による利用を前提とした配布を行える．例えば，喋り声を対象とした合成音声ソフト COEIROINK ではユーザの作成した音声合成モデルが 350 キャラクタ分以上配布されているほか [1]，歌声を対象とした合成音声ソフト UTAU では同ソフト上で使用できる UTAU 音源ライブラリが 7000 キャラクタ分以上存在する [2]．このように，今や合成音声ソフトの利用者は使える声に対し非常に多くの選択肢を持っており，その全ての把握は現実的ではない．合成音声を利用するシーンにおいて，声が持つイメージや印象は声を選ぶ上で考慮すべき要素であり，例えば喋らせるアナウンスの内容や，歌わせる曲調など用途に合った声質を持つライブラリの選定は重要なプロセスである．しかし，現状声の持つ印象を知るには実際に聴いてみることが最も有力な手段であり，数多あるライブラリの生み出す声を十分な数聴き比べ適切な声を採択するには多大な手間と時間を要する．さらにその結果として，多くのユーザがライブラリを選ぶ際，普段の生活の中で聞いた経験のある声の中から声を選択し，結果としてユーザ全体の中で使われる声に大きな偏りが生じる問題も発生する．万に近い数存在するライブラリのうち実際にユーザに用いられる声は一握りであり，ほとんどのライブラリはユーザに用いられずなく埋もれてしまう．

そこで本研究では，ライブラリごとの声に対する印象を事前に数値化し，それを用いてユーザの求める声に近いライブラリを探索するシステムを提案する．探索対象とする合成音声ソフトは特に利用できるライブラリが多く，後述する声質に関するアンケートが存在している UTAU 音源ライブラリを対象とする．本システムでは声に対する印象を複数の印象軸ごとに評価スコアとして数値化し，ユーザは理想としてイメージする声の評価スコアを入力することで，目的に合った音源を探索できる．評価スコアの軸には，例えばスコアの高低を女性らしい声・男性らしい声に対応させた”性別感”など，ユーザが声からスコアを，あるいはスコアから声をある程度想定できるような直感的な軸が望ましい．各音源に対する評価スコアは，アンケート調査によって集められたデータをもとに音源ファイルから各評価スコアを推定できる機械学習モデルを作成し，それを用いて付与する．また，本システムは開発する上で実際にユーザに利用されることを想定し，多くの人が手軽に利用できるよう Web アプリケーションとして実装する．

2　関連研究

2.1　アマチュア歌唱者に向けた歌声可視化方法の検討

本研究に関連する研究として，人間の歌声から印象やイメージされる色を推定する研究 [] がある．この研究では，ある程度の長さがある歌声と，そのうちの瞬間的な長さの歌声を用い，それぞれで印象を推定するモデルを作成している．ある程度の長さがある歌声では，迫力性，丁寧さ，明るさの 3 軸に対してそれぞれのスコアを推定するモデルを作成した．結果として，人の間でも印象の評価が大きく揺れるような歌唱などの例外を除き，十分な精度で印象を推定できた．また歌声のうち瞬間的な音声からは印象に加え，声に対してイメージされる色を推定する試みも行っている．彩度など色の要素と表現語の一つとして挙げられた活動性との関係が見られるなど，声質の色での表現に対して一定の有効性が示されたものの，他の要素との関係については今後の課題とされている．

2.2　声を探索するシステムの先例

本研究と同じく UTAU 音源を対象に声質に対し評価スコアを付与し，そのスコアを用いて音源を探索するシステムを提案する研究 [] が存在する．この研究でも本研究と同じく UTAU に評価スコアを付与することで探索システムを構築し，実際にユーザが求める声を探索できるかを確認している．スコアの推定には UTAU を用いて合成された音声データを用い，重回帰分析とカーネル回帰分析での推定制度の比較を行なっている．また，推定されたスコアを用いて音源を探索する際には，ある 2 つのスコア行列間のユークリッド距離を目標類似距離とし，その逆数として定義した目標類似度を用いて音源を探索している．評価実験では，ユーザのイメージする声に近いスコアを入力し，目標類似度の高いライブラリを提示することで，ユーザが求めるような声を持つライブラリを探索できることが示された．一方でこの研究では，探索システムの実装に留まっており，探索アルゴリズムの検討や実際にユーザが利用することを想定したシステムの提案は行われていない．

3　*声質に対する評価スコアの推定

本研究では，UTAU 音源ライブラリに対して声質に対する評価スコアを付与するための機械学習モデルを作成する．評価スコアの推論には学習データとして UTAU 音源声質アンケートのデータを，モデル作成には Python ライブラリである PyCaret を用いた．

3.1　UTAU 音源ライブラリと UTAU 音源声質アンケート

まず，今回用いる UTAU 音源ライブラリについて説明する．UTAU 音源ライブラリは，無償で公開されている歌唱用音声合成ソフトである UTAU 上で使用できる音源ファイルであり，ソフトと同じく無償で公開されているものが多い．UTAU は波形接続型合成音声と呼ばれる手法を用いており，音声データを切り貼りすることで音声を合成する．そのため，UTAU 音源ライブラリは主に合成に用いるためにできる限り一定の音程と音量になるように収録された収録者の肉声が収録されている．収録形式には複数の手法があり，単独音であれば各 wav ファイルにひらがな 1 文字に対応する音素が収録されており，連続音（VCV）であれば「あんああいあうあ」[3] といった形で複数の音素が連続して収録されている．音素と音声ファイルの対応は oto.ini ファイルによって定義されており，UTAU はこのファイルを参照して子音や母音の開始位置を把握し合成に利用する．一部のライブラリでは，掠れ声や甘い声など，声質を意図的に変化させたものや，違う音程で追加収録したものが存在し，それぞれ表情音源，多音階音源と呼ば

れる．また UTAU 音源ライブラリはその声のみをデータとして持つため，歌唱時のピッチ遷移や発音の癖などが存在せず，歌声に対する印象はその声質のみに依存すると考えられる．

次に，UTAU 音源声質アンケート [4] について説明する．UTAU 音源声質アンケートはニコニコ大百科上で提言された UTAU 音源ライブラリに対する声の特徴を評価するためのアンケート規格であり，現在までにこの規格を用いて 250 種以上の UTAU 音源ライブラリに対してアンケートが行われている．このアンケートは，声の性別，滑舌，特有性，声の年齢，透明感，声の強さ，声の明度の 7 項目について，それぞれ 1 から 7 までの 7 段階評価で 10 件以上のアンケート調査を行い，その平均を評価値としている．アンケートは各 UTAU 音源ライブラリごとに行われ，表情音源が複数存在する場合は各表情音源ごとに独立してアンケートが行われている．

このように，UTAU 音源ライブラリは数が多い点だけでなく，声質の依存先が生の音声ファイルである点や，また音素情報を関連づける ini ファイルも一般的に用いられる形式であるため非常に扱いやすい点，UTAU 音源声質アンケートが存在する点などから，ライブラリの声質を評価する上で都合が良いため，本研究の対象として選定した．

3.2 機械学習モデルの作成手法

UTAU 音源声質アンケートのデータを用いて，UTAU 音源ライブラリに対する評価スコアを推定する機械学習モデルを作成する．

(1) UTAU 音源の形式とどのように使われており，また今回の研究に対しどう都合が良いかについて説明する
(2) UTAU 音源アンケートはニコニコ大百科上で行われている音源に対するアンケートであり，そのデータを用いて評価スコアを推定する．
(3) データの前処理・理由と目的
(4) モデルの構築・理由と目的
(5) 結果

3.3 *結果の評価

(1) モデルの評価
(2) 評価スコアの自動推定

4 *ここに探索システムの名前を入力

4.1 *システムの概要

(1) システムの機能
(2) システムの画面イメージ
(3) システムの利用方法
(4) システムの実装

4.2 *システムの評価

(1) システムの有用性
(2) システムの拡張性

5 おわりに

参考文献

[1] COEIROINC，MYCOEIROINK，https://coeiroink.com/mycoeiroink/list

[2] Vocaloid Database，https://vocadb.net/Search?searchType=Artist&artistType=UTAU

[3] 巽式 連続音の録音リスト配布 - 巽のブログ，https://tatsu3.hateblo.jp/entry/ar426004

[4] ニコニコ大百科，UTAU 音源声質アンケートとは，https://dic.nicovideo.jp/a/utau%E9%9F%B3%E6%BA%90%E5%A3%B0%E8%B3%AA%E3%82%A2%E3%83%B3%E3%82%B1%E3%83%BC%E3%83%88