

年末発表

ボーカロイド楽曲の
歌唱難易度指標に関する初期検討

安藤研究室

21T317 谷知紘

ボーカロイド(VOCALOID)とは？

- ヤマハが開発した歌声合成技術の応用ソフトウェア

ボーカロイド楽曲(ボカロ楽曲)とは？

- ボーカロイドを用いて作成された楽曲の総称

主な特徴

- 機械音声がメインボーカルを担当
- 人間には表現が難しい歌声を容易に合成可能

先行研究

「歌いにくさ」が人気の要因の1つではないか



謎スペースを
埋めてくれる
重音テト

「UTAUだから
ボカロじゃない」
の申し出は受け
付けません

先行研究

人間が歌いにくい楽曲の特徴としてメロディに注目[佐々木ら, 2013]

- 「発音単位毎の継続時間」「音程」「テンポ」に基づいて分析
- 人気の曲ほど, 人間が発声しにくい音域が含まれやすいことを発見

歌詞の意味の伝わりにくさに注目[中井ら, 2022]

- ボカロ楽曲とJ-POP楽曲における品詞の使用率を分析
- ボカロ楽曲, J-POP楽曲, とともに品詞の偏りがあることを発見

課題点

歌いにくさを定量的に表現するまでには至っていない

研究目的

研究目的

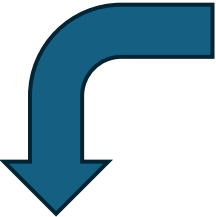
歌詞に注目した歌唱難易度推定方法の実現

検討手法

ボーカロイドには以下の制約が存在しない

- 息継ぎの制約
- 筋肉の動きの制約

→ 歌いにくさを表現する指標となる

- 
1. 単位時間あたりの音素数による算出法
 2. 歌詞に用いられる単語の難しさによる算出法
 3. 早口言葉を構成する要素の有無による算出法

1. 単位時間あたりの音素数による算出法

【仮定】

単位時間あたりの
音素数が多いほど
歌唱難易度が上がる

「音素」
意味の違いに関わる
最小の音の単位

音素数の算出

ローマ字変換済みの歌詞を使用
歌詞中に登場するアルファベット数を疑似音素数としてカウント

単位時間の基準

1分間における拍の数を表すBPM(Beats Per Minute)を採用

歌唱難易度の定義

$$\text{歌唱難易度} = \text{BPM} \times \text{疑似音素数}$$

1.実験設定

手法

有志により作成された「VOCALOID曲難易度表」を用いる
音域の広さ・リズム・低音高音の頻度のみを基準にランク付け

VOCALOID曲難易度表のうち

- 上位20曲→「歌いにくい」
- 下位20曲→「歌いやすい」

上位20曲

- 炉心融解
- ワールズエンド・ダンスホール
- 初音ミクの激唱
- ...

下位20曲

- メランコリック
- からくりピエロ
- モザイクロール
- ...

全40曲における歌唱難易度の平均値と中央値を閾値として利用 ➡ 各楽曲を2クラス分類

表1:歌唱難易度の平均値と中央値

	上位20曲	下位20曲	全40曲
平均値	240,575	131,240	185,907
中央値	223,839	117,174	160,545

- 上位20曲
- 下位20曲

全40曲の歌唱難易度を
閾値で分けたときの分類
性能を測る

1.実験結果

表2:2クラス分類のスコア

		Accuracy	Recall	Precision	F1
平均値	TP:上位20曲	0.75	0.60	0.86	0.71
	TP:下位20曲	0.75	0.90	0.69	0.78
中央値	TP:上位20曲	0.75	0.75	0.75	0.75
	TP:下位20曲	0.75	0.75	0.75	0.75

表3:2クラス分類のスコア(マクロ平均)

	Recall	Precision	F1
平均値	0.75	0.77	0.74
中央値	0.75	0.75	0.75

考察

全体的に低い歌唱難易度が広く分布した
低い歌唱難易度:多
高い歌唱難易度:少

実験データの分析結果

各性能が約0.75と, 高い有効性を確認

2-1. 歌詞に用いられる単語の難しさによる算出法

単語の難しさとは

- 使用される単語(漢字)が難しい
- 使用される単語が日常的に馴染みがない

手法

歌詞に含まれる漢字の総数を100%とし、
そのうち、漢字検定の各級(2～10級)の漢字が
何割現れるかを調査する。

各歌詞の結果を、レベルごとに平均をとる。

各レベルの平均値をグラフで可視化し、
振り分けのばらつき具合を確認する。

【仮定】

歌詞に用いられる漢字が難しいほど
歌唱難易度が上がる

級の振り分けには以下を用いる。

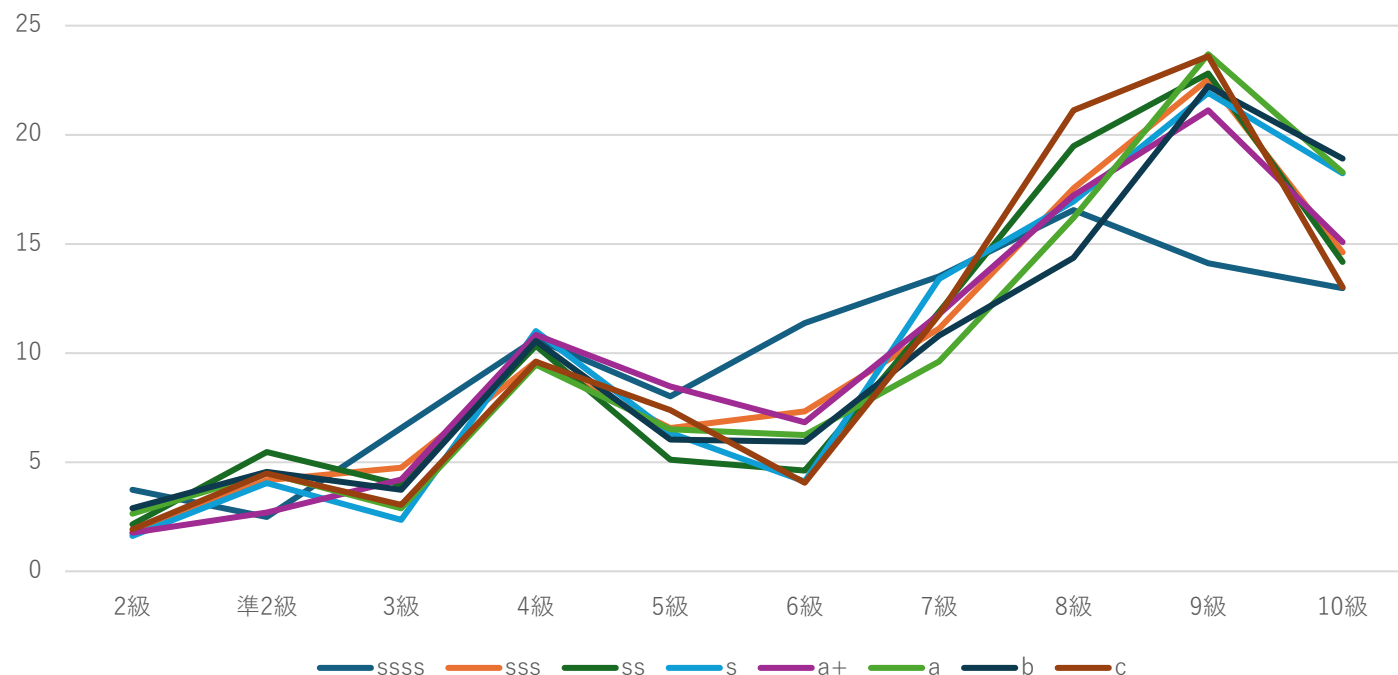
- 日本漢字能力検定 級別漢字表(2～10級)

歌詞データには以下を用いる。

- VOCALOID曲難易度表

2-1.実験結果

レベル別 級別漢字 割合



- SSSS 難
- SSS
- ...
- b
- c 易



考察

全体的に級のレベルが下がるほど出現率が高くなる傾向があった。

例外的に、4級の漢字が前後の級に比べて出現率が高くなった。

どのレベルも、似通ったグラフになることを確認できた。

実験データの分析結果

歌詞に使用される漢字より算出する歌唱難易度は、有効性が低い。

2-2. 歌詞に用いられる単語の難しさによる算出法

単語の難しさとは

- 使用される単語(漢字)が難しい
- 使用される単語が日常的に馴染みがない

手法

歌詞に含まれる各単語の親密度の平均値を測る.
単語親密度には, 以下の3種類がある.

- 文字音声単語親密度
- 音声単語親密度
- 文字単語親密度

各レベルの平均値をグラフで可視化し,
振り分けのばらつき具合を確認する.

【仮定】

歌詞に用いられる単語に馴染みがないほど歌唱難易度が上がる

単語親密度の振り分けには以下を用いる.

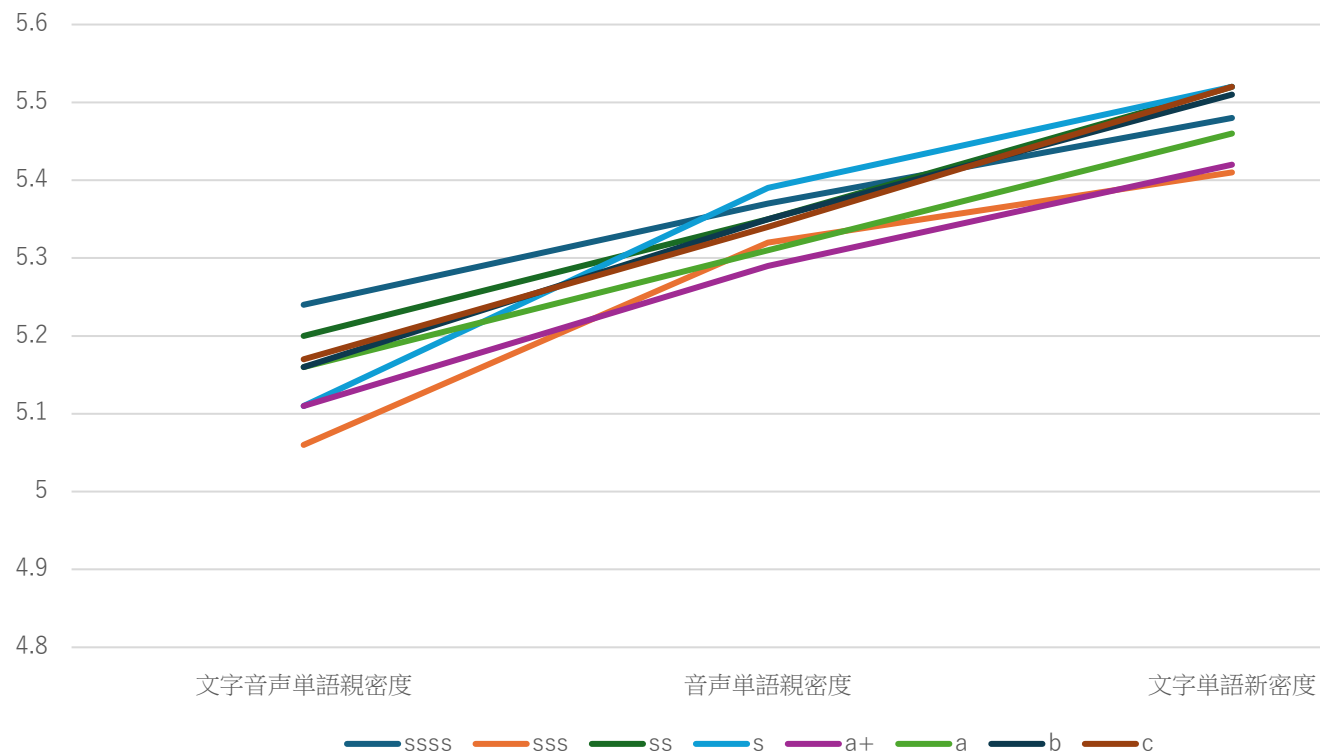
- NTT単語親密度データベース平成版

歌詞データには以下を用いる.

- VOCALOID曲難易度表

2-2.実験結果

レベル別 単語親密度 平均



考察

レベルに関わらず似た傾向になる.

どのレベルにおいても文字単語親密度が一番高い結果となった.

実験データの分析結果

歌詞に使用される単語の親密度による歌唱難易度は, 有効性が低い.

2つの実験を踏まえた全体のまとめ

歌詞に用いられる単語の難しさと歌唱難易度に関係性はない

3.早口言葉を構成する要素の有無による算出法

【仮定】

早口言葉に似た文体の
歌詞ほど歌唱難易度が
上がる

早口言葉を構成する要素(一例)

- 素早く発音しにくい音が含まれる
 - 拗音(○ゃ, ○ゅ, ○ょ)
 - マ行, パ行, バ行
- 似たような表現の繰り返し
 - 母音と母音が入れ替わる
 - 子音と子音が入れ替わる

- 老若男女
- バスガス爆発

- 炙りカルビ(あぶりかぶり)
- 神アニメ(かにあにめ)

問題点

早口言葉の定義が曖昧, 複雑
→ ルールベースに基づく判定が難しい

解決案

早口言葉の文体を機械学習
→ 作成したモデルにより早口言葉の判定を行う

3.モデル構築

早口言葉の特徴に対する着目点

- 文全体の母音と子音の組み合わせ
 - バスガス爆発
 - basugasubakuhatsu
 - 文字ベースのunigram
- 単語と単語間の音の組み合わせ
 - バスガス爆発
 - [“バス”, “ガス”, “爆発”]
 - [“su”, “ga”, “su”, “ba”]
 - 文字ベースのbigram
- その両方

【使用データ】

Google検索より, 早口言葉, 早口言葉でない文を取得
早口言葉160件

- <https://jp.quizcastle.com/dictionary/jx3kj6g75p>

ことわざ160件

- <https://proverb-encyclopedia.com/primary-school/>

手法

3つの特徴に対し,
以下3つの計算によってモデルを構築する.

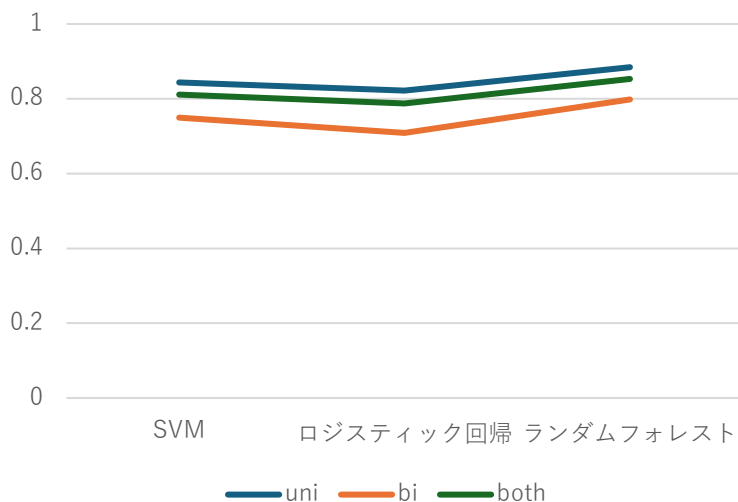
- SVM
- ロジスティック回帰
- ランダムフォレスト

ベクトル変換には, TF-IDFを用いる.
→ 文の長さの違いによる影響を受けにくい.

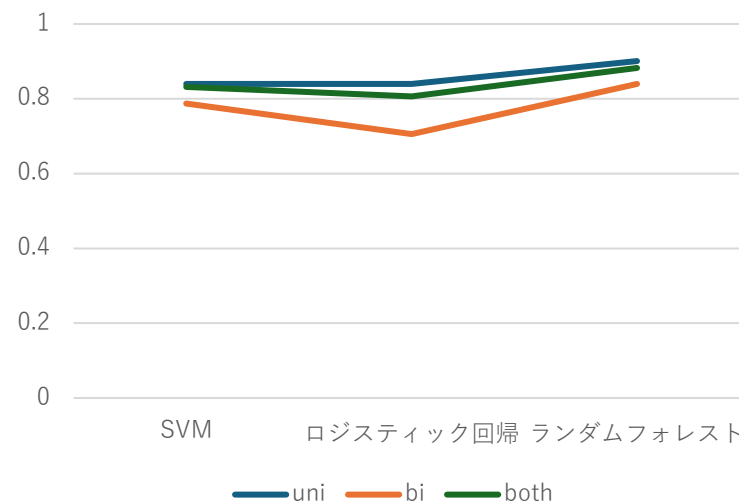
各3つの計算によって作成されたモデルを
10分割交差検証により評価する.

3.実験結果

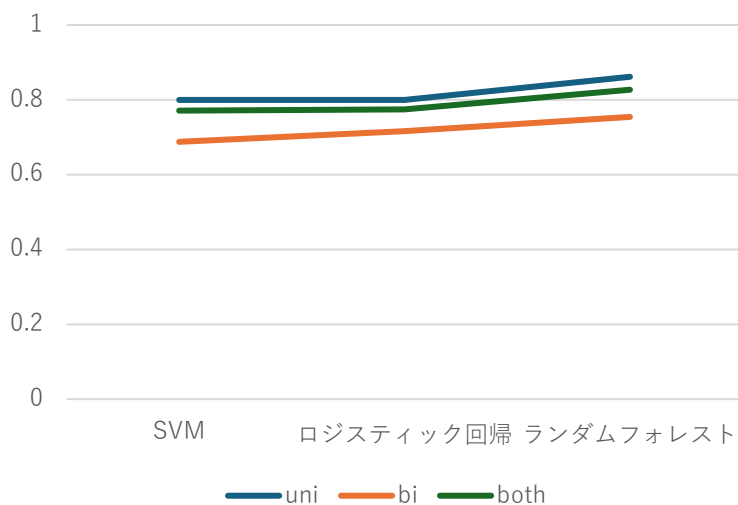
Accuracy



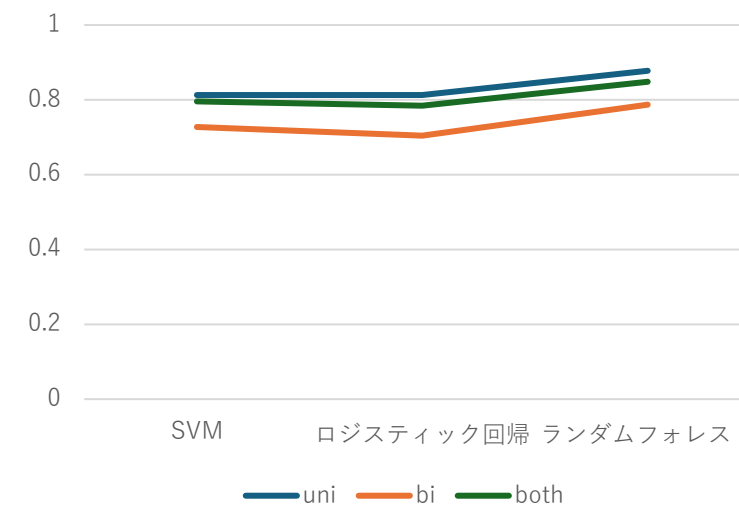
Precision



Recall



F1



考察

順位

- ローマ字変換→unigram
- 両方の手法を合体
- 文字間の音抽出→bigram

全体的にランダムフォレストが良い精度を得た.

ipadicによる分かち書きの精度がイマイチだったことに原因があると推測

実験データの分析結果

ローマ字変換 x ランダムフォレストによるモデルを採用.

今後の予定

- ローマ字変換 x ランダムフォレスト(文字ベースのunigram)の手法に対して,
 - uni-gram
 - bi-gram
 - tri-gramの3つに性能差があるかの検証.
- 最終的に選定したモデルで, 早口言葉の要素による歌唱難易度推定が有効かを判定する.
- 新しい手法の模索&検討
- 卒論執筆