

調音運動に基づく発声難易度の指標化と 歌詞の歌いやすさ評価への応用の検討

宋 健智^{1,a)} 齋藤 大輔^{2,b)} 峯松 信明²

概要：本研究では歌詞に存在する「単語の連続」としての側面と「音の連続」としての側面のうち、後者に着目し、歌いやすい歌詞の自動作詞を行うための歌いやすさの評価手法の検討を行う。具体的に、本研究では従来までは経験的な知見があるに留まっていたテキストの発声難易度を、調音運動に着目してその物理的特徴を用いることで指標化し、これを用いて歌い易い歌詞かどうかを評価することを提案する。最後に主観評価実験により、提案する発声しにくさの指標が人間の感覚に合致するものかどうかを検証する。

キーワード：自動作詞, 歌い易さ, 発声難易度, 調音運動

1. はじめに

コンピュータを用いた芸術創作活動の歴史は長く、それは音楽の分野においても例外ではない。本研究ではこれらのうち、詞先^{*1}を想定した自動作詞に関する検討を行う。日本語の自動作詞の主だった関連研究として、樺島 [1], 伊藤 [2], [3], 堀ら [4], 堀・嵯峨山 [5] によるものが挙げられる。これらの研究では、いずれも確率的言語モデルに基づいた行単位の自動作詞手法の提案を行っている。

歌詞には単語の連続としての側面と音の連続としての側面があるが、関連研究は前者にのみ着目している場合が殆どである。そこで本研究では歌詞の音の連続という側面に着目して、歌い易い歌詞を生成するような自動作詞を行う事を考える。ここで歌い易さというのが主観的で曖昧な概念であることが問題となるが、本研究では歌い易さが発声難易度と強く関係すると仮定し、調音^{*2}運動という人体の動きに基づいて発声難易度を指標化することで歌い易さを捉える事を提案する。即ち、従来までは経験的な知見があるに留まっていたテキストの発声難易度を、調音運動に着目してその物理的特徴を用いることで数量的に指標化し、この指標を用いて歌い易い歌詞かどうか評価することを目指す。また、提案した指標を自動作詞結果のリランキンクに用いることの有効性を主観評価実験により検証した。

2. 関連研究

2.1 自動作詞に関する関連研究

ここでは、本研究の歌詞評価実験においてキーワードを用いた自動作詞を行う際に用いた堀・嵯峨山によるバイグラムモデルに基づく自動作詞手法の概要を説明する [5]。

2.1.1 バイグラムモデル

単語の出現確率が直前の $N-1$ 単語だけに依存すると仮定する確率的言語モデルを N グラムモデルと呼び、特に $N=2$ の場合をバイグラムモデルと呼ぶ。即ち、バイグラムモデルは単語の出現確率が直前の 1 単語のみに依存すると仮定するモデルであり、遷移確率 a_{ij} (単語 w_i の次に単語 w_j が出現する確率) を与えると定まる。このバイグラムモデルをベースとし、遷移確率に対して次節で述べる分散的意味表現に基づく修正を加えた確率的言語モデルを用いて、ユーザが入力したキーワードを含む行単位の歌詞を生成する。

バイグラムモデルを推定するために、歌詞コーパスの各行を形態素解析し、各行の解析結果の最初と最後に文頭と文末を示す特殊記号 BOS と EOS をそれぞれ付与する。次に、形態素解析の結果から形態素間の遷移回数を数え上げ、これをもとに相対頻度として遷移確率を推定する。さらに、歌詞コーパスに現れなかった単語の遷移確率がゼロになる問題 (ゼロ頻度問題) を回避するため、グッド・チューリング法による遷移確率の平滑化 (スムージング) を行う。

2.1.2 分散的意味表現

バイグラムモデルに基づき、遷移確率だけを利用して歌詞生成を行うと、歌詞コーパスの持つ情報のうち隣接しな

¹ 東京大学大学院情報理工学系研究科電子情報学専攻

² 東京大学大学院工学系研究科電気系工学専攻

a) sou@nae-lab.org

b) dsk_saito@gavo.t.u-tokyo.ac.jp

*1 歌唱曲制作において作曲より先に作詞を行う手法

*2 舌や唇等の調音器官の運動により様々な種類の音声を作り出すこと

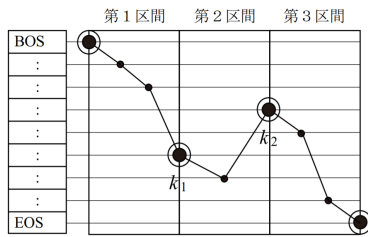


図 1: 経路探索による歌詞の生成. [5]p34 より引用.

い単語間の関連性が反映されなくなる. この問題を解決するため, 単語の分散的意味表現を利用して, 生成される歌詞に非隣接単語間の関連性を反映させる.

分散的意味表現は, ベクトル空間に単語を埋め込む手法であり, 意味的に近い単語はベクトル空間においても近いベクトルとして表現される. 単語 w_i, w_j の表現ベクトルを $\mathbf{w}_i, \mathbf{w}_j$ とすると, 両者が意味的に近い場合, コサイン類似度 $\mathbf{w}_i \cdot \mathbf{w}_j$ の値が大きくなる. ここでは, 各単語 w_i の表現ベクトル $\mathbf{w}_i$ は単位ベクトルに正規化されているものとしており, そのためコサイン類似度が内積と一致している. 分散的意味表現の学習は, Mikolov らによる word2vec を用いて行う [6]. コーパスの持つ非隣接単語間の関連性を反映させるため, ユーザが入力したキーワードと遷移先の単語のコサイン類似度が大きいほど遷移確率が大きくなるように修正を行う. 具体的には, 前小節の a_{ij} を以下のように修正する.

$$\frac{1}{C_i} a_{ij} \cdot \exp(\lambda(\bar{\mathbf{k}}) \cdot \mathbf{w}_j - 1) \quad (1)$$

ここで, $C_i = \sum_j a_{ij} \cdot \exp(\lambda(\bar{\mathbf{k}}) \cdot \mathbf{w}_j - 1)$ は遷移先に関する確率の合計を 1 とするための定数, λ は修正の重みを表す定数, $\bar{\mathbf{k}}$ はユーザが入力したキーワード $k_1, \dots, k_n$ の表現ベクトル $\mathbf{k}_1, \dots, \mathbf{k}_n$ の重心を規格化したものである. ユーザが入力したキーワードとの類似度が大きい単語の出現確率を大きくすることにより, キーワードを媒介として非隣接単語間の関連性を反映させる.

2.1.3 動的計画法による歌詞の生成

歌詞の生成は, 歌詞コーパスを形態素解析して得られた形態素リストから適切なものを選択して並べていくことにより行う. これは図 1 のような形態素リスト上の経路探索と考えられる. 図 1 の左側は歌詞コーパスを形態素解析して得られた形態素のリストであり, このリスト上で BOS から始まり EOS で経路が 1 行の歌詞を与える. 前小節で修正を加えた遷移確率に基づき, 確率が最大となる経路を探索することにより, ユーザが所定するキーワードを含み歌詞コーパスの単語の並びの特徴を保つ歌詞の生成が可能となる.

この問題のようなキーワード $k_1, \dots, k_n$ を通過する経路探索の問題は迫江らによる 2 段 DP (動的計画法) に基づく次のような手順で解くことができる [7]. 第 1 段階で

は, ユーザが入力したキーワードに便宜的に $k_0 = \text{BOS}$ と $k_{n+1} = \text{EOS}$ を付け加え, k_{i-1} と k_i をつなぐ各経路 (第 i 区間と呼ぶ) を順に探索する. この段階では経路上の形態素数にある程度の幅を持たせ, 各区間について経路の候補を複数求める. また, 計算時間短縮のためにビームサーチを行う. 第 2 段階では, 各区間について求めた候補を総当たりで組み合わせて 1 行の歌詞とし, ユーザが指定した文字数からのずれが一定値以下のものを最終的な候補とする. そして, 最終的な候補を確率が大きい順にソートし, 上位から所定の数を表示する.

2.2 発声難易度に関する先行研究

奏らは, 昭和 30 年代に NHK 放送文化研究所がアナウンサー 282 人に対して行った発声しにくい言葉に関するアンケート調査の結果に基づいて, 発声難易度を高める要因を以下のように要約している [8].

- (1) 音節主音として母音音素 /i/ あるいは /u/ を持つ音節の連続: /hihihifito/ 「ひしひしと」
- (2) 音節主音として母音音素 /a/ を持つ音素の連続: /atatakai/ 「暖かい」
- (3) それぞれ別の音節に属している /-ei/ の /e/ と /i/, /-ii/ の /i/ と /i/ の並列: /iiNkai/ 「委員会」
- (4) 子音音素の次に半母音音素 /j/ が位置している音節の連続: /sjuzjucu/ 「手術」
- (5) 半母音音素 /j/ あるいは /w/ を間に挟んでの二つの母音音素の並列: /miajamaru/ 「見誤る」
- (6) 音素 /N/ の連続: /moNdaikoNdaNkai/ 「問題懇談会」
- (7) 硬音的な無声子音素を音節副音としてもつ音節の連続: /hitotaci/ 「人たち」
- (8) 摩擦音・歯擦音系列の子音音素を音節副音としてもつ音節の連続: /shicizi/ 「7 時」
- (9) 両唇音の子音音素を音節副音としてもつ音節の連続: /mimamoru/ 「見守る」
- (10) 歯茎の系列子音音素を音節副音としてもつ音節の連続: /banananado/ 「バナナなど」

奏らの他にも発声難易度に関する議論を行っている研究は存在するが, それらは発話者が病気等により発音・発声が困難な状態であるという前提のものが殆どであり, 一般的な発声難易度について検討したものとなっていない. 従って本研究では, 奏らによる発声難易度を高める要因の考察を出発点として調音運動に基づく発声難易度の指標化の検討を進める.

3. 調音運動に基づく発声難易度の指標化

3.1 提案手法

本研究では, 調音運動に基づきテキストの発声難易度を指標化する手法を提案する. まず, テキストを調音運動に変換するため, 隠れマルコフモデル (Hidden Markov

表 1: 調音ベクトル

名称	概要
TD _x	舌後方センサの水平方向
TB _x	舌中部センサの水平方向
TT _x	舌前方センサの水平方向
LI _x	顎（下切歯）のセンサの水平方向
LL _x	下唇のセンサの水平方向
UL _x	上唇のセンサの水平方向
TD _y	舌後方センサの垂直方向
TB _y	舌中部センサの垂直方向
TT _y	舌前方センサの垂直方向
LI _y	顎（下切歯）のセンサの垂直方向
LL _y	下唇のセンサの垂直方向
UL _y	上唇のセンサの垂直方向

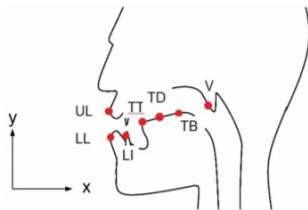


図 2: 磁気センサ取付位置。[9]p.4 図 2 より引用

表 2: HMM の学習条件

学習データ	99 文
特徴量	36 次元（調音ベクトル + $\Delta + \Delta\Delta$ ）
HMM の状態数	5
1 状態の混合数	1
フレームシフト	5ms
音素数	37

Model, HMM) に基づく音声合成法を応用する形で HMM に基づきテキストから調音位置の系列を生成するシステム (Text to Articulation) の構築を行う。続いて Text to Articulation を用いて 2.2 節で言及した発声難易度の高いテキストを調音運動に変換して考察することで発声難易度を高める要因となる調音運動の物理的特徴の分析を行い、これを数値化することによって発声難易度を指標化する。

3.2 Text to Articulation の構築

3.2.1 調音データの概要

本研究では HMM の学習用の調音データとして、日本語発話文-調音パラレルデータを用いた [10]。発話文として ATR 音素バランス 503 文 [11] から選ばれた 99 文を使用した。調音データは、磁気センサシステム (ElectroMagnetic Articulograph, EMA) によって測定された顎・唇・舌の位置のベクトル系列である。ベクトルの次元数は 6 センサー $\times$ 2 次元の計 12 次元（単位は mm）であり、サンプリング周波数は 200Hz である。各次元の詳細を表 1 に示し、また

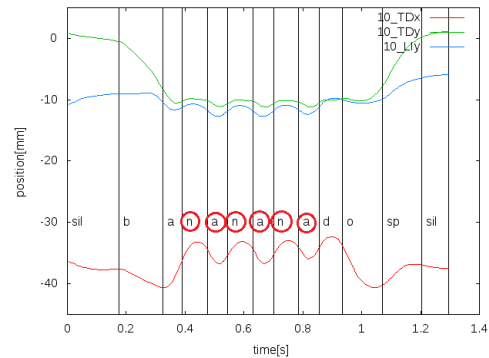


図 3: 水平方向の舌の往復

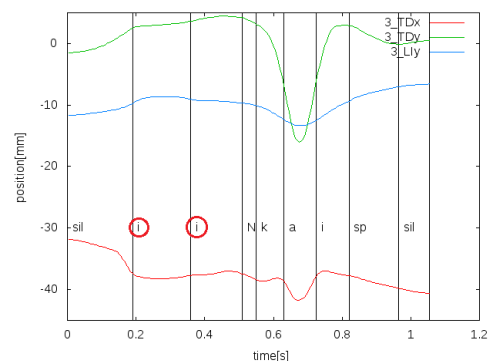


図 4: 垂直方向の唇の停滞

図 2 にセンサ取付位置を示す。

3.2.2 HMM の学習

学習の前準備として、発声データを元に無音区間のラベリングを手動で行い、モノフォン及びトライフォンのラベルを作成した。また、Julius^{*3}を用いて音素セグメンテーションを自動で行い、モノフォンラベルに対し各音素の継続長の情報を付与した。またラベリングの際には、長母音を短母音として扱っている。これらのラベル及び調音データを用いて表 2 に示す条件で HTS^{*4}を用いてトライフォン HMM の学習を行った。

3.3 発声難易度の指標化

3.3.1 Text to Articulation を用いた発声難易度の分析

構築した Text to Articulation を用いて、2.2 節で示した発声難易度を決定する要因を調音運動のレベルで分析した。分析では 2.2 節で示した 10 種類の例文を用い、調音ベクトルのうち TD_x, TD_y, LI_y の 3 成分を観察した。観察の結果、(1) 水平方向の舌の往復（TD_x の正弦波的な運動）(2) 垂直方向の唇の停滞（LI_y の直線的な運動）の 2 つの要因に着目した。

例文の調音位置系列のうち水平方向の舌の往復が見られる例を図 3 に、垂直方向の唇の停滞が見られる例を図 4 に

^{*3} Julius, <http://julius.osdn.jp/>

^{*4} HMM/DNN-based Speech Synthesis System (HTS), <http://hts.sp.nitech.ac.jp/>

示す。図3は2.2節の例文(10)「バナナなど」の発声の調音運動であり、図中の赤丸が付いた発声難易度を決定する要因となっている区間において TD_x が振動している、即ち水平方向に舌が往復運動をしていることが読み取れる。同様に、図4は2.2節の例文(3)「委員会」の発声の調音運動であり、図中の赤丸が付いた発声難易度を決定する要因となっている区間において LI_y の変動が殆どない、即ち垂直方向で唇が停滞していることが読み取れる。 TD_y で表現される舌の垂直方向の運動と発声難易度の関係については、舌の水平方向及び唇の垂直方向の場合のように分かりやすい関係性を見出すことが出来なかった。そのため10種類の例文のうち、調音運動レベルで発声難易度を決定する要因を特定出来ないものはいくつか存在する。本研究では、水平方向の舌の往復及び垂直方向の唇の停滞を調音運動レベルでの発声難易度を決定する要因と定め、これらの特徴を数値化し発声難易度を指標化した。

3.3.2 調音運動の物理的特徴に基づいた発声難易度の指標化

本研究で扱う発声難易度を決定する要因となる調音運動の特徴のうち、「水平方向の舌の往復」に関しては舌の水平方向の運動の周期性に、また「垂直方向の唇の停滞」に関しては唇の垂直方向の運動の定常性に依存している。そのため、前者については舌の運動として考えた場合の高周波数帯域に着目してパワーの大小を見ることで発声難易度の指標化を行えると考えられる。後者については周波数のパワーに着目して捉えることが難しいため調音パラメータを直線近似した際の、平均二乗誤差平方根に着目して指標化を行う。以下「水平方向の舌の往復」及び「垂直方向の唇の停滞」のそれぞれについて具体的な評価手法を述べる。

まず「水平方向の舌の往復」に関して評価手法を述べる。評価の前準備として TD_x 成分の正規化パワースペクトルを見るため、データに対して全時間の平均が0となるように補正を加え、その系列に対しハミング窓をかけて高速フーリエ変換(Fast Fourier Transform, FFT)を行った。3.2節で示した例文の中で TD_x が周期的な運動を示しているもの(2.2節の例文(2), (5), (6), (7), (10))について観察を行ったところ、周期が0.13-0.16sの範囲に収まっていることが分かった。例えば図3に関しては TD_x が約0.15s周期で動いている。そこで周期の逆数を取り6-8Hzの帯域について帯域内の最大パワーを見たところ、水平方向の舌の往復が見られる文の方がそうでない文に比べてパワーが大きくなる傾向があることが分かった。3.2節で示した例文10文における水平方向の舌の往復運動の有無による最大パワーの違いを図5に示す。サンプル数は振動あり5文(2.2節の例文(2), (5), (6), (7), (10))、振動なし5文(2.2節の例文(1), (3), (4), (8), (9))である。ここまでの議論より、「水平方向の舌の往復」に関しては発声難易度を $\log_{10} P$ と定める。ただし、 P は6-8Hzの帯域における

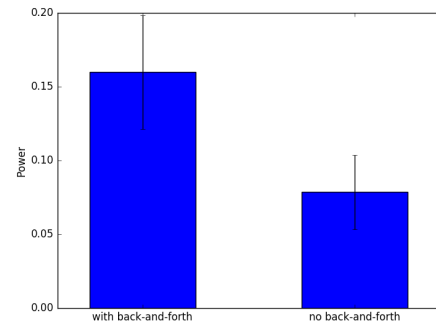


図5: 水平方向の舌の往復の有無による最大パワーの違い (エラーバーは95%信頼区間)

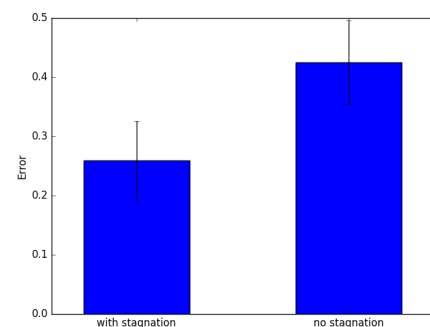


図6: 垂直方向の唇の停滞の有無による平均二乗誤差平方根の違い (エラーバーは95%信頼区間)

最大のパワーである。

続いて「垂直方向の唇の停滞」に関して評価手法を述べる。垂直方向の唇の停滞を調音位置の直線的な変化と捉えることにすれば、線形回帰を行い直線近似の平均二乗誤差平方根を求めることで発声難易度の評価を行うことが出来る。しかし、単純に全区間を対象に線形回帰を行うと、短時間における停滞運動の影響が無視されてしまう。本研究ではこの問題を解決するため、調音運動のデータを一定間隔ごとに区切り、各区間においてそれぞれ平均二乗誤差平方根を求め、さらにこれらの平均を取り発声難易度の評価に用いる。区切りの間隔は、0.25s(50フレーム)とした。3.2節で示した例文10文における垂直方向の唇の停滞の有無による平均二乗誤差平方根の違いを図6に示す。サンプル数は停滞あり4文(2.2節の例文(1), (3), (4), (8))、停滞なし6文(2.2節の例文(2), (5), (6), (7), (9), (10))である。ここまでの議論により、「垂直方向の唇の停滞」に関しては発声難易度を $-\epsilon$ と定める。ただし、 ϵ は本段落で定義した平均二乗誤差である。

ここまでの議論で定義した水平方向の舌の往復による発声難易度及び垂直方向の唇の停滞による発声難易度を統合して、最終的に調音運動に基づく発声難易度の指標を次式で定める。

$$a_{bf}r_{bf} + a_{st}r_{st} \quad (2)$$

ここで r_{bf}, r_{st} はそれぞれ水平方向の舌の往復による発声難易度及び垂直方向の唇の停滞による発声難易度である。即ち $r_{bf} = \log_{10} P, r_{st} = -\epsilon$ である。また, a_{bf}, a_{st} はそれぞれ r_{bf}, r_{st} に対応する係数である。

4. 歌詞の発声難易度に関する評価実験

4.1 実験概要

2.1 節で示した手法に基づき自動作詞を行い, 得られた歌詞を 3 章で提案した発声難易度のスコアに基づき歌い易い順にリランキングして提案手法の効果を確認する主観実験を行う。

実験において発声難易度のスコアは, 式 (2) の係数を $a_{bf} = 1, a_{st} = 1.86$ として計算した。この係数は「水平方向の舌の往復」による発声難易度の値 r_{bf} 及び「垂直方向の唇の停滞」による発声難易度の値 r_{st} が, 発声難易度の高低の区別に同程度影響するように設定した。即ち, 発声難易度の高い文とそうでない文の間の r_{bf} の差 Δr_{bf} 及び r_{st} の差 Δr_{st} を考えると, Δr_{bf} と Δr_{st} は共に発声難易度の差の度合いを表していると言えるが, この時に係数を加味した値 $a_{bf}\Delta r_{bf}$ と $a_{st}\Delta r_{st}$ が等しくなるようにする事で 2 つの要因による影響の大きさを均等にするを図った。各 Δ の計算には図 5, 図 6 に示すデータを用い, その結果の $\Delta r_{bf} = 0.308, \Delta r_{st} = 0.165$ の逆比をとって係数を上記のように定めた。

自動作詞は, 以下の条件下でキーワードを 2 語与えることにより行った。

・歌詞コーパス

実験で用いる歌詞コーパスとして日本語ポピュラー歌唱曲約 22 万曲のデータを用いた。歌詞コーパスのデータ総量は約 774 万行に上る。歌詞に含まれる英数字や記号等は取り除いている。

・形態素解析

バイグラム構築の前段階として Mecab^{*5}を用いて形態素解析を行い, 歌詞コーパス中の文に対し「分かち書き」(形態素の区切りで文を分割すること)を行った。単に分かち書きを行うのみでは語彙数が膨大になるため, 本実験では分かち書きしたのちに形態素を原形に修正することで語彙数を抑制した。

・バイグラムモデル構築

バイグラムの構築には palmkit[12] を用いた。本実験では歌詞コーパスにおいて出現回数が 20 回以上の単語を語彙として扱い, その他を未知語とした。その結果語彙数は 31506 語となった。また, 平滑化にはグッドチューリング法を用い, 2 単語間の遷移のうち遷移回数が 7 回以下のものをディスカウンティングした。

^{*5} Mecab: Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/>

表 3: word2vec の学習条件

次元数	100
窓幅	10
学習率	0.025
学習回数	5
バッチサイズ	10000

表 4: リランキング前の結果

順位	歌詞	生成確率	発声難易度
1	夢見慣れることを描き続ける	-37.715	-1.542
2	夢見慣れすぎる空を描く	-37.722	-1.760
3	夢見慣れることを描いてる	-37.735	-1.865
4	夢見慣れることを描く未来	-37.847	-1.773
5	あなたの夢を描き続けてる	-38.248	-1.923
6	あなたの夢を描き続けたい	-38.273	-1.931
7	あなたを見る夢を描き続ける	-38.436	-1.751
8	夢見慣れることを描ける	-38.454	-1.510
9	あなたを描く夢を描き続ける	-38.626	-1.689
10	いつから見る夢を描き続ける	-38.658	-1.462

表 5: リランキング後の結果

順位	歌詞	生成確率	発声難易度
1(18)	あなたは夢を描き続けること	-39.726	-2.058
2(12)	あなたの夢を描き続けること	-38.840	-2.044
3(17)	あなたの夢を描き続けるから	-39.442	-2.018
4(22)	あなたは夢を描き続けるから	-40.328	-2.012
5(28)	夢見慣れた夢を描く未来	-41.565	-1.993
6(14)	あなたは夢を描き続けてる	-39.133	-1.931
7(29)	夢見慣れたことを描いた	-41.791	-1.965
8(26)	夢見慣れた夢を描いてる	-41.452	-1.962
9(27)	夢見慣れることを描く未来へ	-41.524	-1.961
10(6)	あなたの夢を描き続けたい	-38.273	-1.931

・word2vec

word2vec の学習条件を表 3 に示す。学習データは歌詞コーパスと同一で, 学習時には skip-gram モデルを用いた。

・探索条件

ビームサーチの探索幅は 10 とした。また, 各キーワード間で探索する経路の長さは 5 単語以下とした。

自動作詞では, 生成確率上位から 30 文までの歌詞の生成を行った。続いて, これらの歌詞について調音運動に基づく発声難易度のスコアを算出し, 発声難易度の低い順にリランキングを行った。例として, キーワードとして「夢」, 「描く」を与えて自動作詞を行った場合のリランキング前後の上位 10 位までの歌詞を表 4, 表 5 にそれぞれ示す。表 5 の順位のかっこはリランキング前の順位を表す。

次に, リランキングにより歌詞が発声し易い順に並べ変えられているかどうかを評価するため, 主観評価実験を行った。実験では, 被験者に対してリランキング後の順位

が $3n+1$ ($n=0,1,\dots,9$) であるような歌詞計 10 文をランキンング前の順序に従って提示し、発声し易いと思う順に並べ替えて貰った。被験者には、4 つのキーワードセット (1) 「夢」と「描く」、(2) 「明日」と「会う」、(3) 「君」と「思う」、(4) 「今」と「始まる」から自動作詞された歌詞セット計 4 種を提示した。実験の結果得られた歌詞の順序 A と発声難易度のスコアに基づいた歌詞の順序 A' を比較し、ケンドールの順位相関係数 τ を求めることで本研究で提案した発声難易度が人間の感覚に合致するものであるかどうかを評価した。

$$\tau = \frac{c^+(A, A') - c^-(A, A')}{T(T-1)/2} \quad (3)$$

ここで、 $c^+(A, A')$ は並べ替えた歌詞セット中の任意の 2 つの歌詞について A と A' における順序の高低が一致した回数であり、 $c^-(A, A')$ は一致しなかった回数である。 T は順序の総数でこの場合は 10 である。ケンドールの順位相関係数 τ は -1 から 1 までの値を取り、 1 に近いほど 2 つの順序が正の相関を持ち、 -1 に近いほど負の相関を持つことを表す。

4.2 実験結果

9 人の被験者による主観評価実験の結果、キーワードセット毎のケンドールの相関係数 τ の平均は (1) 「夢」と「描く」: 0.052, (2) 「明日」と「会う」: -0.33 , (3) 「君」と「思う」: 0.24, (4) 「今」と「始まる」: 0.13 となった。キーワードセットによっては緩やかな正の相関が見られたものの、全体的には必ずしも本研究で提案した発声難易度指標の有効性を示す結果ではなかった。この結果からは文章の発声難易度を、生成された文章のみから評価するというタスクそのものが非常に難しいとも考えられる。また式 (2) の係数にも改善の余地がある。実際の歌詞読み上げ音声の分析等を取り入れるなど、実験設定を見直すことでより人間の感覚に即した指標設計が期待される。

5. まとめ

本研究では、歌詞の持つ音の連続としての側面に着目して、歌い易い歌詞を生成するような自動作詞の実現を目的として調音運動に基づく発声難易度の指標化を行うことを提案した。まず始めにテキストを調音運動に変換するシステム (Text to Articulation) を構築し、これを用いて発声難易度の高いテキストの調音運動を分析した。次に分析結果から「水平方向の舌の往復」及び「垂直方向の唇の停滞」という 2 つの運動を発声難易度を決定する要因とし、これらの要因と関連する物理量を用いて発声難易度の指標とした。また実際に自動作詞を行い、主観評価実験によって発声難易度の指標が人間の感覚に合致するものかどうか検証し、今後の展望を示した。

ところで、本研究で用いた日本語発話文-調音パラレル

データは通常の発話の収録データであるから、本研究で言う発声難易度はテキストを普通に読み上げた場合を想定したものとなっている。そのため、本研究で用いた発声難易度の指標を実際の歌声にも適用できるかは現段階では不明であり、検証が必要である。今後歌声データを用いた Text to Articulation を構築して分析を行ったり、歌詞を入力とする自動作曲システム Orpheus[13] を用いて歌詞にメロディ・音長を付けたものを用いて実験したりすることが期待される。

参考文献

- [1] 樺島忠夫：流行歌をつくる—国語学・国語表現法資料，計量国語学，No. 52, pp. 8–40 (1970).
- [2] 伊藤雅光：ユーミンの言語学 (17)，日本語学，Vol. 17, No. 9, pp. 92–97 (1998).
- [3] 伊藤雅光：第三次近似 (3-gram) 自動作詞実験のためのテキスト生成語彙表 (1) 80 年代前期の松任谷由実作品を素材テキストとして，計量国語学，Vol. 28, No. 8, pp. 297–334 (2013).
- [4] 堀 玄，齋藤大輔，嵯峨山茂樹：研究発表会形態素リスト上の経路探索による自動作詞 (計量国語学会第五十七回大会報告)，計量国語学，Vol. 29, No. 3, pp. 104–106 (2013).
- [5] 堀 玄，嵯峨山茂樹：分散的意味表現を利用した自動作詞，2017 年度人工知能学会全国大会 (第 31 回)，pp. 1N1–2 (2017).
- [6] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013).
- [7] 迫江博昭，千葉成美：動的計画法を利用した音声の時間正規化に基づく連続単語認識，日本音響学会誌，Vol. 27, No. 9, pp. 483–490 (1971).
- [8] 奏 吉正，木村圭子：「発音しにくい言葉」の調査とその分析，NHK 放送文化研究所年報第 9 集，pp. 191–206 (1964).
- [9] 内田秀継，齋藤大輔，峯松信明，広瀬啓吉：統計的音声-調音マッピングにおける声質変換を利用した話者正規化の検討，信学技報，Vol. 114, No. 303, pp. 59–64 (2014).
- [10] 内田秀継，橋本哲弥，齋藤大輔，峯松信明，末光厚夫：三次元磁気センサシステムを用いた日中二カ国語話者の音声-調音パラレルデータの収録，日本音響学会講演論文集，1-Q-32, pp. 301–302 (2017).
- [11] 匂坂芳典，浦谷則好：ATR 音声・言語データベース，日本音響学会誌，Vol. 48, No. 12, pp. 878–882 (1992).
- [12] 伊藤彰則，好田正紀：単語およびクラス n-gram 作成のためのツールキット，情報処理学会研究報告音声言語情報処理 (SLP)，Vol. 2000, No. 119 (2000-SLP-034)，pp. 185–190 (2000).
- [13] 深山 覚，中妻 啓，米林裕一郎，酒向慎司，西本卓也，小野順貴，嵯峨山茂樹：Orpheus: 歌詞の韻律に基づいた自動作曲システム，情報処理学会研究報告音楽情報科学 (MUS)，Vol. 2008, No. 78 (2008-MUS-076)，pp. 179–184 (2008).