Vol.2024-MUS-140 No.26
Vol.2024-SLP-152 No.26
2024/6/14

合成歌唱に適した芸術言語構築の基礎検討

石川 真大[1,a)]　森勢 将雅[1,b)]

概要：芸術言語とは美的要素を追求した人工言語の一種であり，芸術言語の一種である架空言語はゲームや映画，音楽などで用いられる．芸術言語を用いる既存の歌声合成手法では，コンテンツの要求に応じて手動で作成するため，構築にコストがかかる．そこで，歌声合成に特化し，自動生成を実現することで，関連コンテンツ制作の幅が広がり，省力化にも繋がると考えられる．本研究では，既存の楽曲に対するモーラ出現数を反映し，日本語のように聞こえることを目的とした芸術言語の自動構築について検討した．比較となる言語を含めた 3 種類の言語で音源を作成し，歌声の日本語らしさ及び流暢さについて主観評価実験を実施した．本稿では，芸術言語の構築方法について述べ，歌声の与える印象を定量的に検証し，芸術言語による歌声の有効性について議論する．

1. はじめに

芸術言語とは美的要素を追求した人工言語の一種であり，文字や記号，シンボルの他にも音声として創出される．エンターテインメントコンテンツにおいて，主に架空世界の創作を補助することを目的として使用され，架空世界の集団や個人のアイデンティティを形成する社会言語学的文脈で用いられる．こうした言語は，架空言語と呼ばれる．代表的な架空言語の例として，映画作品の『Avatar』シリーズでは舞台となる惑星の母語として Na'vi language [1] が登場する．他にもゲーム作品では『Monster Hunter』シリーズのモンスターハンターの言語（モンハン語）[2] などが挙げられる．

これら芸術言語の開発には，それぞれの作品テーマに沿った言語的一貫性を担保するために文法・音韻を設計する必要がある．加えて，高品質な芸術言語音声を収録するには，ディレクターや話者が言語の特性を十分に習熟しなければならない．

芸術言語を歌詞とした歌声に着目すると，歌詞はネイティブ話者であってもしばしば聞き取ることが困難であることから [3]，言語の文法的破綻はある程度無視できることが期待される．そこで本研究は，言語的一貫性を保つために，モーラバランスのみを考慮した芸術言語構築手法を提案し有効性を議論する．全く新しい芸術言語の言語的一貫性を評価することは十分に習熟していない実験参加者にとって困難である．そこで，初期検討として日本語楽曲のモーラ出現数を反映することで「日本語のように聞こえることを目的とした」芸術言語を構築した．提案手法では，文法の構築が不要であるほか，「特定の言語らしさ」を有する．提案手法により合成した芸術言語歌声のプロトタイプを用い，その「日本語らしさ」を主観評価により検証した．実験では比較対象として，日本語と，構築法の異なるベースラインを含む，3 種類の言語を用意した．歌声音声として，使用する言語の影響を調べるため，英語と日本語の 2 種類の音声データベースから歌声を生成し，違いを検証した．

2. 関連研究

特定の言語情報に依存しない歌声合成の研究例として，スキャット生成の研究がある [4]．この研究では，声と個人性，情緒，感動などの関係の追求を目的としている．他にもスキャット生成の研究として，歌声合成技術の長所を活かしユーザーの好みに合ったスキャットを生成する例が挙げられる [5]．また，スキャットとは別のアプローチによる言語情報に依存しない研究の事例として，言語情報を含まない感情音声合成の研究が挙げられる [6]．この研究では，WaveNet [7] を用いて感情情報を含む音声の合成を行っている．一方で，人工言語の音声的特徴について研究を行った例として，ファンタジー言語の知覚とその音声学的・音韻論的特徴の研究が挙げられる [8]．この研究では，オークやエルフといったファンタジーキャラクターに用いられる言語音声の受聴者による美的評価とその言語の音声学的・音韻論的特性の関連性を調査している．また，CEDEC+KYUSHU2022 では Text-to-Speech による「架

[1] 明治大学
[a)] cs242003@meiji.ac.jp
[b)] mmorise@meiji.ac.jp

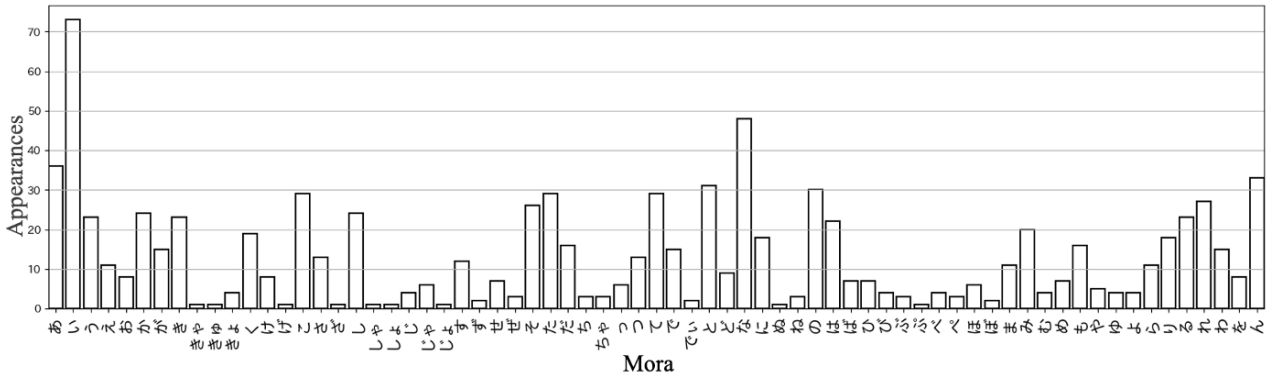

図 1 楽曲 1『Idol』モーラ出現数

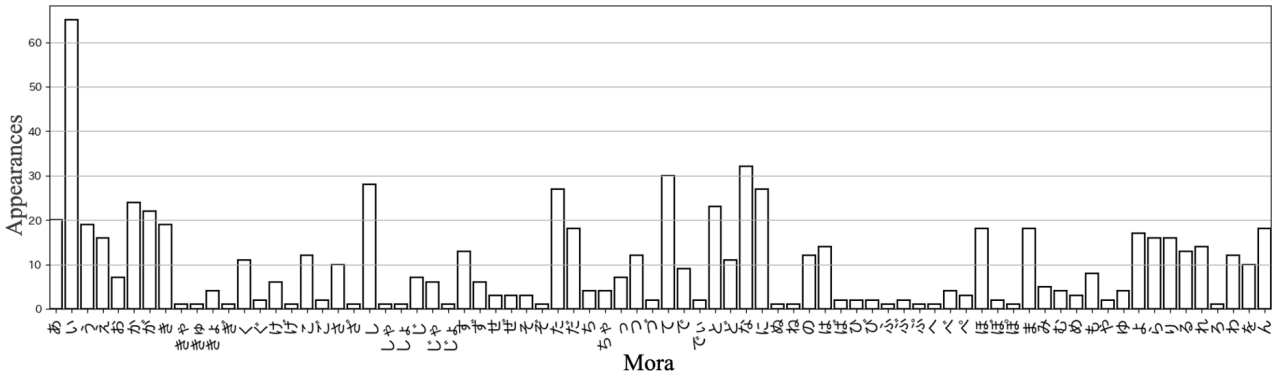

図 2 楽曲 2『Soranji』モーラ出現数

空言語」音声の生成手法が紹介された [10]．具体的には日本語のテキストを英語の音声特徴量変換器と音声変換器に読み込ませることで，英語風の架空言語音声を生成している．

本研究は，合成歌唱に適した芸術言語構築を目指しており，言語的一貫性の保持においてモーラバランスのみを考慮している点でコンセプトが異なる．具体的には，芸術言語による歌声の評価方法及び，芸術言語の構築方法をどのように決定するかが課題設定である．

3. モーラバランスに基づく芸術言語の構築

日本語の音声的特徴として，日本語特有のモーラバランスと，子音＋母音または母音単体の音素構造 [11] によるモーラがある．これら特徴を保持できるように芸術言語の構築を検討することで，日本語らしさの実現を目指した．

実験では，使用する楽曲に対して，日本語のほかに，ルールに沿って作成した芸術言語（以後 Original と呼ぶ），比較対象の芸術言語（以後 Random と呼ぶ）の 3 つを使用して歌声を作成した．

Original は日本語のモーラを活用することで作成することとした．これにより，日本語モーラの音素構造に手を加えず特徴を活かすことができる．また，日本語特有のモーラバランスを保持するために，楽曲に対して行うモーラ出現数の調査結果を反映した．具体的な Original の構築方法を以下に記す．

- 芸術言語を作成する対象となる楽曲に対し，歌詞データを取得する．
- 歌詞データからそれぞれのモーラごとに出現数をカウントする．
- カウントの結果から，出現数の高い順にペアを作成する．
- 原曲の歌詞データのモーラを，ペアに従い入れ替え，Original の歌詞を作成する．

本実験で調査を行った 2 曲に対するモーラ出現数を図 1，図 2 に示す．

これに対して，Random はベースラインとして全ての日本語モーラを無作為に入れ替えることにより，構築されている．このようにして用意された 2 種類の言語と日本語による歌詞データを使用し，既存の楽曲から音源を作成した．

4. 実験

4.1 実験音源の作成

本実験では，YOASOBI の『Idol』と Mrs.GREENAPPLE の『Soranji』の 2 曲から歌声音源を作成した．サンプルの歌声合成には，Dreamtonics 株式会社の Synthesizer V [9]

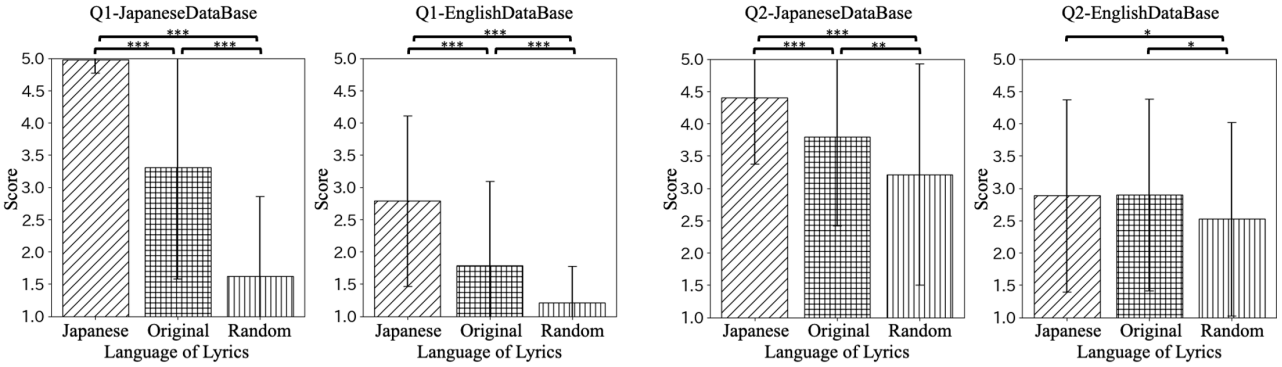

図 3 主観評価実験の Score．Q1 は日本語らしさについて，Q2 は流暢さについて．
*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

を使用した．また，歌声データベースには，日本語音声として『Synthesizer V AI 夏色花梨 ライト版』，英語音声として『Synthesizer V 弦巻マキ English AI ライト版』を使用した．

初めに，それぞれの楽曲に対して，3 種類の言語による歌詞データを用意し，合計 6 曲について日本語と英語それぞれの音声データベースを使用して歌声を作成した．この時，Synthesizer V に対する歌詞の入力をローマ字で行うことにより，両方の音声データベースで歌声合成を可能としている．実験では，作成した歌声を 4 つのサンプルに分割した，合計 48 サンプルをランダムに提示することで評価を行った．また，伴奏によるマスキングを防ぐため，伴奏は付与せず歌声のみの音源とした．

4.2 主観評価実験

作成した音源について日本語らしさを実現できているか評価を行うため，5 段階の主観評価実験を行った．実験条件を表 1 に示す．

各参加者は，音源を受聴したのち，歌声に対して 2 つの質問について，「非常に良い」，「良い」，「普通」，「悪い」，「非常に悪い」の 5 段階で評価した．質問項目として，1 つめに日本語らしさの達成を評価するために「歌声として日本語らしく聞こえるか」を設定し，2 つ目に歌声の聞きやすさを評価するために「流暢だと感じるか」という質問を設けた．

表 1 実験条件

参加者	大学生 10 名
再生環境	騒音レベル 20–30 dB の防音室
ヘッドフォン	HD 650 / Sennheiser
オーディオ・インターフェイス	ADI-2 DAC FS / RME
サンプリング	48 kHz / 32 bit

5. 結果と考察

5.1 結果

本実験 のスコアを調査項目ごとに図 3 に示す．グラフは条件に従い，4 つに分けて作成した．

- 質問 1 で日本語音声データベースを使用した条件（Q1-JapaneseDataBase）
- 質問 1 で英語音声データベースを使用した条件（Q1-EnglishDataBase）
- 質問 2 で日本語音声データベースを使用した条件（Q2-JapaneseDataBase）
- 質問 2 で英語音声データベースを使用した条件（Q2-EnglishDataBase）

縦軸に質問に対する評価，横軸に歌詞に使用した言語を示している．また，エラーバーは 95% 信頼区間を示している．Shapiro-Wilk 検定により，データは正規分布しないと判断されたため，対応のあるノンパラメトリック検定を，Wilcoxon の符号付順位検定に基づき行った．それぞれの有意差について，グラフ中の項目ごとに示す．

実験の結果，4 つの条件のうち，英語音声データベースを使用した質問 2 の条件を除く 3 つの条件において，データ間に有意差が認められた．質問 1 について，いずれの音声データベースを使用した条件においても，スコアは高い順に，日本語の歌詞，Original の歌詞，Random の歌詞となった．また，質問 2 についても，日本語の音声データベースを使用した条件において，スコアは高い順に，日本語の歌詞，Original の歌詞，Random の歌詞となった．

5.2 考察

本研究で使用する音源において，歌声合成システムに対する歌詞の入力は手作業となってしまうものの，芸術言語の構築段階においてはそのシンプルな構造により，構築コストの削減ができていると考えられる．

芸術言語の日本語らしさについては，質問 1 の結果から，いずれの音声データベースにおいても Original が Random よりも日本語らしく受聴されていると考えられる．このことから，当初定めた「日本語のように聞こえる」芸術言語の構築をある程度達成できたといえる．

流暢さについては，質問 2 の結果から，日本語音声データベースにおいて，Original の手法は Random より流暢だ

といえる．このことから，日本語音声データベースによるOriginal の歌声は Random より聞きやすいといえる．しかし，英語の音声データベースにて有意差が認められなかったことを考慮すると，流暢だと感じる要因として歌詞と音声データベースの組み合わせが考えられる．これについては，日本語以外の言語に対しても同様の手法で芸術言語の構築及び音声データベースを使用した歌声の評価を行うことが必要になるだろう．また，流暢さについて言語構築の点から考えたとき，Random は無作為なモーラの入れ替えにより構築されているため，日本語では出現率の低いモーラが多く出現している可能性が考えられる．これにより，Random による歌声では発話に違和感が生じやすくなっていると考えられる．よって，芸術言語の構築においては，基準となる言語の音素の出現頻度を加味することが重要と考えられる．

6. 終わりに

本稿では，歌声に適した芸術言語構築の基礎検討を行うため，芸術言語のプロトタイプを用いた歌声音源により主観評価実験を行った．実験により，歌声に適した言語構築の検討において，モーラ出現数を考慮することは一定の有効性があることが示された．一方で，本実験では複数楽曲に対するモーラの出現数調査を行っていないため，作曲者の影響によるモーラ出現数の偏りが考えられる．

今後の展望として，より優れた芸術言語構築の手法を確立するため，今回得られた知見を元に，再度検討を行う．具体的には，モーラ出現数の調査を行う曲数を増やし，普遍的なモーラ出現頻度に基づく芸術言語構築を検討する．

謝辞 本稿完成にあたり助言をいただいた小口純矢氏，俣野文義氏に感謝の意を表します．なお，本研究の一部は，JSPS 科研費 JP21H04900 の支援を受けました．

参考文献

[1] Learn Na'vi, "Kaltxì! Welcome to Learn Na' vi," https://learnnavi.org (閲覧日：2024/5/8).
[2] Monster Hunter Rise キャラクター声優アンケート，"オトモ雇用窓口のイオリ役花江夏樹さん," https://www.monsterhunter.com/rise/ja/topics/enquete/hanae.html (閲覧日：2024/5/8).
[3] 須田仁志, 中村友彦, 深山覚, 緒方淳．"FruitsMusic: 音楽情報処理のためのアイドルユニット楽曲コーパス," 音楽情報科学研究会（MUS）, vol.2024–MUS–139, no.13, pp. 1–10, 2024.
[4] 河原英紀, 片寄晴弘．"高品質音声分析変換合成システム STRAIGHT を用いたスキャット生成研究の提案," 情報処理学会誌, vol.43, no.2, pp. 208–218, 2002.
[5] 鶴田穣士, 岡夏樹, 田中一晶．"ユーザー好みのスキャットを強化学習する初音ミクとのジャムセッションシステムの開発," 人工知能学会全国大会（第 32 回）, vol.2018–MUS–119, no.23, pp. 1–4, 2018.
[6] 松本剣斗, 原直, 阿部匡伸．"WaveNet による言語情報を含まない感情音声合成方式の検討," 情報処理学会研究報告, vol.2019–MUS–123, no.61, pp. 1–6, 2019.
[7] Oord Aaron van den, Dieleman Sander, Zen Heiga, Simonyan Karen, Vinyals Oriol, Graves Alex, Kalchbrenner Nal, Senior Andrew, and Kavukcuoglu Koray. "WaveNet: A Generative Model for Raw Audio," arXiv preprint arXiv:1609.03499, 2016.
[8] Mooshammer, C., Bobeck, D., Hornecker, H., Meinhardt, K., Olina, O., Walch, M. C., and Xia, Q. "Does Orkish Sound Evil? Perception of Fantasy Languages and Their Phonetic and Phonological Characteristics," Language and Speech, pp. 1–40, 2023.
[9] Dreamtonics. "Synthesizer V Studio," https://dreamtonics.com/ja/synthesizerv (閲覧日：2024/5/8).
[10] CEDEC Degital Library．"意味が分からないからこそ、リアル ～「架空言語」音声合成による、没入感の高いボイス付きコンテンツの実現～," https://cedil.cesa.or.jp/cedil_sessions/view/2685 (閲覧日：2024/5/8).
[11] 高見澤孟，ハント蔭山悠子．"新・はじめての日本語教育 1 日本語教育の基礎知識," アスク出版, pp. 37–39, 2004. https://books.google.co.jp/books?id=xEbmEJobbcYC (閲覧日：2024/5/8).