

# N-gram による歌声合成に向けた芸術言語構築法の検討

石川 真大<sup>1,a)</sup> 森勢 将雅<sup>1,b)</sup>

**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、映画や音楽などにおいて世界観を強調するなどの目的で用いられる。楽曲においては、聴取者の母語ではない言語の歌声であっても魅力的に感じられることがある。芸術言語による魅力的な歌声の自動生成を実現することは、母語によらず楽しむことができる歌声として、歌声合成の表現拡張に寄与すると考えられる。本研究では、日本語モーラにおける N-gram モデルを用い、歌声合成に向けた芸術言語の構築法について検討した。本稿では、複数の N-gram モデルを用いた芸術言語の構築及び、歌詞の快さを主観評価実験により検証し、提案する芸術言語構築方法の有効性について議論する。

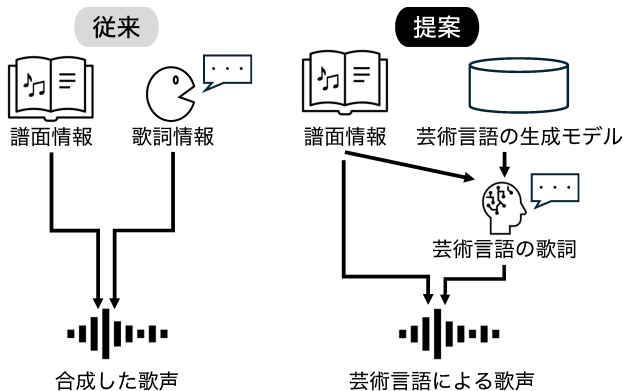


図 1 従来の歌声合成と比較した、本研究における歌声の合成。

## 1. はじめに

芸術言語とは、人間が特定の目的のために意識的・人為的に設計をする人工言語のうち、美的要素を追求したものであり、視覚的・音声的形式を持つ。中でも映画や音楽などのコンテンツにおいて、社会言語学的文脈で用いられる芸術言語を架空言語と呼ぶ。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ナヴィ語 [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。

歌唱表現においても、聴取者の母語とは異なる言語や非語彙的な音声を用いて美的効果を高める試みが見られる。既存の楽曲作品においては、平沢進の『Sign』 [3] や土屋暁の『謳う丘』 [4] の様に、芸術言語を用いることで、意味内容よりも音韻・音響の特徴による魅力を重視し、母語を共

有しない聴取者に対しても、歌声による魅力を伝達している。他にも、スキヤットのように音声や音節を楽器的に利用し、即興的に歌唱する方法も、ジャズや作曲時の仮歌として広く利用されている。

芸術言語を歌声として設計・運用する場合には、言語体系の構築段階において、それぞれの作品テーマに沿った言語的な特徴と音響的な自然さの両立を図る必要がある。Beinhoff(2015) の研究では、人工言語制作者に対して行った言語構築の意図に関する調査の結果、発音のしやすさや音象徴など、6つに分類できる音韻システムの構築が、言語制作者にとって馴染み深い要素として示された [5]。加えて、歌声合成技術を用いる際の合成可能な音素列への対応も、制作難易度の向上を招いている。芸術言語による歌詞も同様であり、言語構築そのものに対する難しさがある。歌詞と譜面に基いて歌声を生成する歌声合成による芸術言語歌詞も存在するが、基本的には既存の言語が対象である。 [6]。従って、非語彙的な言語や極端に出現頻度の低い音素列の歌声における合成は十分にサポートされていない。そのため、制作者は表現上の要請と言語設計上の制約の双方に対応せねばならず、コンテンツ制作のコストや専門性の上昇を引き起こしている。

こうした背景から、本研究では歌声合成に適した形で言語の音響的な魅力を備えた芸術言語を自動生成し、既存の歌声合成器により歌声を合成することを目指す。これにより、芸術言語の構築を容易にするだけでなく、母語によらず楽しむことができ、スキヤットやコーラスとは異なる、発展的な歌唱法として歌声合成の表現拡張に貢献が期待できる。研究では、図 1 のように、既存の歌声合成器に対し、芸術言語の生成モデルから出力した歌詞を用いることで歌

<sup>1</sup> 明治大学

<sup>a)</sup> cs242003@meiji.ac.jp

<sup>b)</sup> mmorise@meiji.ac.jp

声を作成し、歌声の評価により検討を行う。

## 2. 関連研究

### 2.1 関連研究

人工言語 (constructed languages) は、伝統的に自然発生的な言語に焦点を当ててきた言語学研究において、個人によって意識的に作り出された言語として、近年、関心が高まっている研究対象である [7][8]。人工言語の構造、習得、歴史、及び使用法に関する研究を通じて、自然言語を批判的に考察する機会を提供しており、中でも言語の普遍性、学習過程、および言語構造の本質に関する知見の解明に貢献している [9][10]。

人工言語の音響的設計と印象形成の関係を扱う研究において、Beinhoff(2015) は人工言語制作者に対して行った言語構築の意図に関する質問を通して、音韻システムを構築する際の主要な考慮事項として、発音のしやすさ、審美的評価、現実味/自然さ、理論言語学的考察、他の言語からの影響、そして音象徴の六つのテーマを特定した [5]。特に、音象徴は回答者グループにとって馴染み深い概念であり、大/小、肯定的/否定的、ジェンダーといった特定の概念を表現するために、意図的に特定の音声 (音と意味の関係) が割り当てられていることが明らかになった。さらに Beinhoff(2023) では、既存の架空言語の知覚評価を通じて、音響的特徴の選択が設計意図の伝達に寄与し得ることを示した [11]。特に、クウェンヤ語が「快い」言語として高く評価された背景には、制作者が意図的に避けた「荒々しい音」の欠如が認知的手がかりとして機能していた点が指摘されている。ナヴィ語に関しては、設計者の意図通り「異質」かつ「馴染みのない」音体系として知覚された一方、その評価は舌頭音・放出音・声門閉鎖音など、聴取者の母語に存在しない音素に対する具体的な言及を伴っており、聞き手の第一言語に存在したい特定の音の選択が、設計意図の達成に極めて効果的であることを示した。しかし、研究では全ての設計意図が成功したわけではなく、クウェンヤ語やシンダール語は自然言語の要素を取り入れた設計にもかかわらず、一部では「人工的」「不自然」と評価されたことが報告されている。同様に、ナヴィ語は「快さ」や「友好性」といった側面では意図よりも低い評価が与えられ、馴染みのない音韻要素の多さが評価判断に不確実性をもたらし、聴取者が中庸な反応を選択しやすくなる可能性が示唆されている。すなわち、人工言語における音響設計は一定の審美的効果をもたらし得るものの、その成功は音素選択や異質性の程度以外にも、それが提示される世界観やキャラクター・テキストなどの文脈的情報との整合性にも依存することが示されている。

こうした研究を踏まえ、Mooshammer(2023) は社会的背景知識の影響を最小限に抑えた音声を用いて、言語の音構造そのものが持つ影響について検証を行った [12]。具体的

には、12 種類の既存の架空言語の音声に対して、快/不快、善/悪、平和的/攻撃的の 3 つの尺度により評価実験を行った。結果、言語制作者の意図通り、クリンゴン語やドスラク語はエルフ語 (クウェンヤ語、シンダール語) よりも一貫して不快、悪、攻撃的であると評価された。さらに、この評価の違いは音構造特性によって予測され、非ドイツ語音が多いほど否定的に、有声音の割合が多いほど肯定的に評価されることが判明した。この知見は、人工言語における音響的設計が文脈的情報を考慮しない状況においても、審美的評価に一定の効果をもたらすことを示している。

言語に対する知覚要因について焦点を当てた研究では、印象形成に影響する要素として、母語への親密度と音象徴による普遍性のある印象知覚の双方が作用しうることが知られている。Trudgill(1978) らは、言語評価が社会的・文化的背景に大きく影響されることに加え、聴取者の母語における音韻的な親密度が他言語における自然性や使用する際の許容度に影響することを示した [13]。一方、W.Köhler(1947) のブーバ/キキ効果 (bouba-kiki effect) で知られるように、無意味な音声に対する印象知覚について、母語によらず普遍的な評価を示すことが確認されている [14]。また、近年の研究では発話音声における音響的・音韻的な複雑さによる聴覚システムに与える処理負担が、言語の審美的魅力に影響を与える可能性について示唆されており、脳領域が十分に活性化されつつも、処理が極度に困難にならない最適な負荷レベルの存在が示されている [15][16]。これらの知見を総合すると、芸術言語の歌声・音声における審美的評価は、話者性や社会的背景が要因として作用する可能性がある一方で、それらの要因を排除した環境下においても、音韻構造自体が印象形成に関与する可能性が示唆される。従って、音韻構造の影響を検証するためには、話者性などの条件を統制した環境での評価が必要になる。

### 2.2 本研究の位置付け

人工言語や通常と異なる音声・歌唱を生成する研究分野では、音声・歌声を意味依存から切り離し、音響表現の材料として利用する試みが行われている。通常と異なる歌唱として、音声そのものを楽器として扱うスキャット歌唱の生成における音声合成の研究では、声と個性、情緒、感動などの関係を追求し、人間らしいスキャット生成システムの構築を目指す研究や [17]、ユーザーの好みに合ったスキャット生成の研究が挙げられる [18]。他にも、スキャットとは別のアプローチによる言語音声の生成に関する研究の事例として、WaveNet を用いた感情情報を含む音声合成の研究がある [19]。この研究では、WaveNet による 2 ステップ学習を用い、小規模な感情音声データによって高品質な感情情報を含む音声合成方式を検討し、コーパス音声に近い感情表現の部分的な再現を実証した。また、CEDEC+KYUSHU 2022 においては、Text-to-Speech

(TTS) と呼ばれる、任意の文章に対応する音声波形を合成する技術 [20] を応用した架空言語音声の生成手法が紹介された [21]. 具体的には、TTS で用いられる音声特徴量変換器及び音声変換器に、入力文字列とは異なる言語の変換器を用いることで架空言語音声の生成を行う. この手法は、ある種の法則性を持ってテキストの入力長と同じ長さで音声を出し、言語的な音韻設計を必要としない利点がある. しかし、これらは主として自然言語類似の音声生成を対象としており、歌声合成の枠組みにおいて音韻構造そのものを設計し、その印象評価を検証した本研究とはコンセプトが異なる.

本研究の主眼は、歌声合成に向けた芸術言語の音韻構造が、聴取者の快さ評価に与える影響について検証することである. そこで、本研究では、日本語楽曲に基づく N-gram モデルを用いてモーラ遷移の自然性を制御した芸術言語を構築し、話者性を統一した歌声合成による音源を用いて主観評価実験を行うことで、音韻設計が歌声合成の美的印象に与える影響について検証を行う.

### 3. N-gram モデルによる芸術言語構築

#### 3.1 日本語モーラに基づく N-gram モデル

筆者らは初期検討として、本稿での mono-gram に相当する、単一モーラのモーラバランスに基づく生成手法による芸術言語の歌唱合成を行い、主観評価実験により無作為に生成されたものと比較して快さが向上することを確認した. しかし、この初期検討ではモーラ間の遷移確率を考慮しておらず、配列の流暢さに関する効果は十分に検証できていない. 言語学的な知見として、Frisch(2000) の研究では、辞書に基づく音韻パターンの出現確率が高いほど、創作した英単語の音声的自然性が高く評価されることが示されている [22]. この知見を踏まえると、歌声合成における芸術言語構築手法においても、モーラ遷移の統計的自然さが聴取印象に寄与する可能性がある. また、芸術言語の印象評価に関する既存研究では、聴取者の母語との音韻的親和性が快・不快や肯定・否定といった印象評価に影響を与えることが報告されている [11][12]. 日本語を母語とする聴取者を対象とする本実験において、日本語音韻の出現頻度を操作することは、音韻構造と聴取印象の関係を検討する上で適していると考えられる.

芸術言語の構築手法は多岐にわたっており、多様なコンセプトの一つとして他言語を基にした構築方法も考えられる. しかし、言語学において、自然言語の多くは音韻的等時性に基づいて分類することが難しく、言語のリズムは聴覚的印象や主観的評価の影響を強く受けることが知られている [23][24][25]. 一方で、日本語は、モーラ単位での等時性を持ち、多くの音節が子音と母音 (CV) から構成されるという言語的特徴を有する [26]. そのため、拍リズムとモーラ単位の対応が明確であり、他言語と比較して歌詞

表 1 モーラ出現数調査結果表

	母音					
		a	i	u	e	o
子音	#	1102	3818	1637	647	709
	b	328	193	199	111	271
	by	—	—	—	—	—
	ch	66	410	42	—	47
	d	1068	—	—	825	509
	dy	—	—	—	—	—
	f	—	—	243	—	—
	g	991	103	89	108	142
	gy	2	—	2	—	7
	h	956	362	—	80	256
	hy	3	—	—	—	6
	j	115	325	54	—	61
	k	1856	1238	1393	561	1053
	ky	17	—	42	—	90
	m	1046	796	192	433	1168
	my	—	—	5	—	2
	n	2259	1131	50	242	1600
	ny	8	—	4	—	—
	p	38	26	23	8	19
	py	—	—	—	—	—
	r	1276	685	1177	1017	265
	ry	7	—	4	—	20
	s	642	1367	661	431	500
	sh	44	—	56	—	103
	t	1825	—	—	1677	1408
	ts	—	—	724	—	—
	ty	66	—	42	—	47
	v	—	—	2	—	—
	w	650	—	—	—	—
	y	360	—	211	—	671
	z	51	325	214	134	46
	cl			1043		
	N			1583		
	:			258		

行は子音、列は母音に対応する.  
また、# は子音無し、cl は促音、N は撥音、: は長音に対応する.  
調査において出現の認められなかったモーラについては — とした.

と音符のマッピングが容易であると言える. また、既存のモーラを基礎単位として用いることで、未知の音素列を生成することなく、日本語歌声データベースを利用し、既存の歌声合成器により利用が可能となる点でも実装上の利点を持つ. 従って、日本語を基盤とすることは、音韻構造と聴取印象の関係を明らかにする上で合理的であり、本研究では、日本語モーラの遷移確率を N-gram モデルとして芸術言語構築に導入し、生成した歌詞の歌声が聴取者の快さ評価に与える影響を検証する.

#### 3.2 芸術言語の構築

本実験における芸術言語の構築には、3 種類の N-gram

モデル (mono-gram, bi-gram, tri-gram) に加え、モーラを無作為に並べた Random 手法をベースラインとして用いることとした。芸術言語の構築に用いる N-gram モデルの作成には、初期検討と同様、既存の日本語楽曲の歌詞データを用いる。また、モーラバランスを考慮した初期検討を踏まえ、作曲者の影響によるモーラバランスの偏りを減らすため、楽曲数を増やして歌詞のモーラを集計することとした。モーラの調査には Billboard JAPAN Japan Hot 100 (2024 年 5 月 8 日時点) から使用する楽曲を選定し、そのうち歌詞が日本語の楽曲 81 曲を対象とした。調査はまず、全楽曲の歌詞に対して記号及び英数字の削除と、漢字を平仮名へ変換する編集を行う。次に、編集後の全楽曲の歌詞に対し、それぞれの N-gram (mono-gram, bi-gram, tri-gram) の範囲におけるモーラ及びモーラペアの出現数を集計し出現確率・遷移確率を計算した。全楽曲において出現する単一モーラの集計結果を表 1 に示す。

歌詞の作成には、調査により作成した日本語モーラにおける各 N-gram 確率モデルを用いたプログラムを実装し、自動生成を行った。プログラムは、任意の N-gram 確率モデルを選択し、モーラ数を指定することで、任意のモーラ数の歌詞を出力するよう設計が行われている。具体的な生成として、mono-gram では、始めの 1 モーラから順に、mono-gram 確率モデルに従い、1 モーラずつ生成が行われる。bi-gram, tri-gram では、始めの 1 モーラを無作為に生成したのち、遷移確率に基づいて次のモーラが生成される。N-gram モデルの遷移確率に従った出力を行う際、次のモーラが確率の候補として存在しない可能性が考えられる。こうした状況においては、モデルに存在するモーラから無作為にモーラを選択し生成を再開することとした。Random 手法については、mono-gram 確率モデルの確率分布を一様分布に平滑化することでモデルを構築し、同様に歌詞の生成を行った。

## 4. 評価実験

本実験では、芸術言語の歌声について、快さの評価及び、音質が歌声の評価に影響を与えているか判断するため、作成した歌声音源を用いて、音質評価実験及び、主観評価実験を行った。実験で用いる歌声のメロディには、既知の楽曲に対する聴取経験の差による影響を排除するため、新たに作成した 10 曲のメロディを使用した。メロディに対して、割り当てる歌詞は 1 音につき 1 モーラの対応関係で、それぞれの芸術言語構築手法に基づき生成した歌詞を割り当てた。音源に使用する歌声の合成には、Dreamtonics 株式会社の Synthesizer V [27] 及び歌声データベースに『Synthesizer V AI 夏色花梨ライト版』[28] を使用し、作成したメロディと割り当てた歌詞を入力として与えた。歌声音源の合成後、Audacity [29] を用いて、1 音源当たり 15 秒前後の長さでサンプリングを行い、手法ごとに 21 種類

の歌声を作成。合計 84 の音源により実験を行った。

実験は 2 つ連続して行い、主観評価実験及び音質評価実験の順序は完全に無作為であり、実験の間には 5 分間の休憩を設けた。以下表 2 に実験条件を示す。

音質評価実験において、参加者は無作為に提示される各音源を聴取したのち、音質について、「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。主観評価実験においては、同様に提示される音源を聴取したのち、音源の歌声に対する快さについて、音質評価実験と同様に 5 段階で評価を行った。

表 2 実験条件

参加者	日本語母語の大学生 20 名
再生環境	騒音レベル 20–30 dB の防音室
ヘッドフォン	HD 660S / Sennheiser
オーディオ・インターフェイス	ADI-2 Pro FS R / RME
サンプリング	48 kHz / 32 bit

## 5. 結果と考察

### 5.1 結果

音質評価実験及び主観評価実験のスコアを図 2, 3 に示す。縦軸に評価スコア、横軸に歌詞に使用した芸術言語の手法を示している。また、エラーバーは 95% 信頼区間を示している。いずれの実験においても、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、ノンパラメトリック検定である Kruskal-Wallis 検定に基づき実施した。音質評価実験においては、いずれの手法間においても有意な差は認められなかった。一方、主観評価実験においては、有意差が認められたため、Dunn-Bonferroni 検定による多重比較を行った。それぞれの有意差について、グラフ中の項目ごとに示す。

主観評価実験の結果、N-gram モデルによる各手法は、いずれも Random 手法より有意に高い快さの評価を得た。ただし、N-gram モデルの各手法間における有意な差は確認されなかった。

### 5.2 考察

本研究で使用した音源の品質については、音質評価実験の結果から、実験音源における明確な音質の差は示されなかった。また、歌詞の合成段階において、使用する歌声合成システムにおいて、一部のモーラにおける歌声の合成が不完全となり、歌声の欠損がみられた。特に、Random の手法の歌詞による合成では、他手法よりも多く欠損が見られた。

歌声の快さについては、主観評価実験の結果から、無作為なモーラの並べ替えによる Random 手法と比較し、N-gram モデルによる各手法がいずれの手法においても快

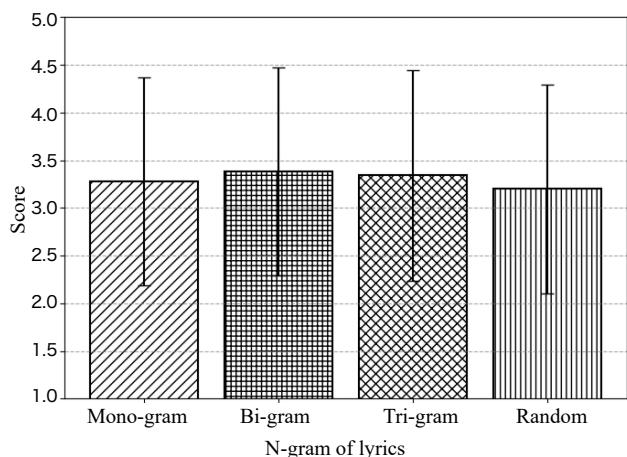


図 2 音質評価実験 Score.

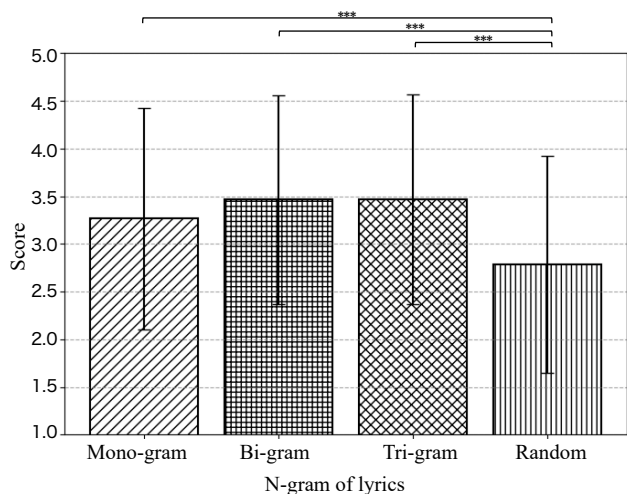


図 3 快さの主観評価実験 Score.

\*\*\*:  $p < 0.001$

く受聴されていると考えられる。この結果は、これまで行った検討と同様に、芸術言語の構築において、出現する日本語のモーラバランスを考慮することは、歌声の快さに一定の有効性があることを示唆している。一方で、Random手法が他手法より快さの評価が低いことに注目したとき、関連研究 [11][12] にて示された、聴取者の母語に親和性の低い音韻の出現増加が、否定的な評価を誘発することを支持すると考えられる。これについては、日本語のモーラバランスを用いた芸術言語構築手法において、あえて出現率が低いモーラを頻繁に用いた構築が、歌声において否定的に知覚されるのか評価を行うことが必要になると考えられる。さらに、N-gram モデルの各手法の比較において、有意な差が得られなかったことから、歌声の快さにおいて、モーラ間の遷移確率を考慮しモーラバランスを複雑化させる必要性は低いことが示唆された。

## 6. おわりに

本稿では、合成歌唱に適した芸術言語の構築に向け、日本語モーラに基づく複数の N-gram モデルを用いた芸術言

語生成を行い、合成した歌声の快さについて主観評価実験により検証を行った。実験の結果、モーラバランスを考慮した芸術言語は、無作為に生成されたものと比較して快さの向上が認められ、言語構築においてモーラバランスを考慮することの有効性が示唆された。一方で、N-gram モデルの各手法間に有意な差は得られず、N の変化によってモーラ遷移を複雑化させることによる知覚効果は限定的である可能性が示された。

歌声の合成段階においては、一部の歌詞において合成の不完全な箇所が確認されたことから、より多様なモーラに対応可能な歌声合成器の導入を検討する必要がある。また、現在は手作業で行っている歌詞入力を、楽譜情報を基に自動化することで、歌詞とメロディを同時に入力として与え、歌声作成の作業効率化を図る予定である。今後は、モーラバランスが聴取印象に与える影響をより詳細に分析するため、快さ以外の審美的要素を含む多面的な印象評価を行うことで、歌声における音韻特徴と聴取印象の関係を統計的に明らかにすることを目指す。

**謝辞** 本研究は、明治大学科学技術研究所重点研究 (B) による支援を受けて実施された。

## 参考文献

- [1] Learn Na'vi, "Kaltxi! 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] CHAOS UNION, Inc, "平沢進 サウンドトラック映像のための音楽〜平沢進サウンドトラックの世界," <https://susumuhirasawa.com/discography/item-43/> (閲覧日: 2025/10/16).
- [4] ヤマハ [mysound], "謳う丘-Harmonics EOLIA - 志方あきこ," <https://mysound.jp/song/1877096/> (閲覧日: 2025/10/16).
- [5] B. Beinhoff, "Why are Alien Languages Inherently Human?," *Foundation*, vol.44, no.122, pp. 5-19, 2015.
- [6] 剣持秀紀, "歌声合成技術の動向: 「初音ミク」を支える技術," *日本音響学会誌*, vol.67, no.1, pp. 46-50, 2010.
- [7] G. Grant, "Constructed Language," *Annual Review Of Linguistics*, vol.9, pp. 419-437, 2023.
- [8] S. Christine, and A. David, "Comparing prehistoric constructed languages: world-building and its role in understanding prehistoric languages," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol.376, no.1824, pp. 1-8, 2021.
- [9] A. Eran, "Language as an instrument of thought," *Glossa: a journal of general linguistics*, vol.1, no.46, pp. 1-23, 2016.
- [10] H. Tily, M. Frank, and F. Jaeger, "The learnability of constructed languages reflects typological patterns," *Cognitive Science Society*, vol.33, no.33, pp. 1364-1369, 2011.
- [11] B. Beinhoff, "Design intentions and actual perception of fictional languages: Quenya, Sindarin and Na' vi," *Edinburgh University Press*, pp. 76-92, 2023.
- [12] C. Mooshammer, D. Bobeck, H. Hornecker, K. Meinhardt, O. Olina, M. C. Walch, and Q. Xia, "Does Ork-



- ish Sound Evil? Perception of Fantasy Languages and Their Phonetic and Phonological Characteristics,” *Language and Speech*, pp. 1–40, 2023.
- [13] P. Trudgill, and H. Giles, “Sociolinguistics and linguistic value judgements: Correctness, adequacy and aesthetics,” *Functional studies of language and literature*, pp. 167–190, 1978.
- [14] P. S. Nathan, and C. Laurent, “The cerebral bases of the bouba-kiki effect,” *NeuroImage*, vol.186, pp. 679–689, 2019.
- [15] D. Jeroen, L. Esther, J. Clio, and V. Tom, “Do some languages sound more beautiful than others?,” *Proceedings of the National Academy of Sciences*, vol.120, no.17, pp. 1–7, 2023.
- [16] A. Anikin, N. Aseyev, and N. E. Johansson, “The sound of beauty: How complexity determines aesthetic preference,” *Acta Psychologica*, Vol.192, pp. 146–152, 2019.
- [17] 河原英紀, 片寄晴弘, “高品質音声分析変換合成システム STRAIGHT を用いたスキット生成研究の提案,” *情報処理学会誌*, vol.43, no.2, pp. 208–218, 2002.
- [18] 鶴田穰士, 岡夏樹, 田中一品, “ユーザー好みのスキ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.
- [19] 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.
- [20] 全炳河, “テキスト音声合成技術の変遷と最先端,” *日本音響学会誌 (第 32 回)*, vol.74, no.7, pp. 387–393, 2018.
- [21] CEDEC Digital 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](https://cedil.cesa.or.jp/cedil_sessions/view/2685) (閲覧日: 2024/5/8).
- [22] S. A. Frisch, N. R. Large, and D. B. Pisoni, “Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Nonwords,” *J Mem Lang*, vol.42, no.4, pp. 481–496, 2000.
- [23] R. M. Dauer, “Stress-timing and syllable-timing reanalyzed,” *Journal of Phonetics*, Vol.11, no.1, pp. 151–62, 1983.
- [24] R. Peter, “On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages,” *Linguistic Controversies*, pp. 73–79, 1982.
- [25] M. Nespor, M. Shukal, J. Mehler, “48 Stress-timed vs. Syllable-timed Language,” *The Blackwell Companion to Phonology*, pp. 1147–1159, 2011.
- [26] 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).
- [27] Dreamtonics, “Synthesizer V Studio,” <https://dreamtonics.com/ja/synthesizerv/>, (閲覧日: 2024/1/17).
- [28] 重音テト, “Synthesizer V AI,” <https://kasaneteteto.jp/sv/#sv1>, (閲覧日: 2024/1/17).
- [29] Muse Group & contributors., “Audacity,” <https://www.audacityteam.org>, (閲覧日: 2024/5/28).