

メロディの特徴と日本語歌詞のモーラ出現頻度の比較調査

宮内 隆利^{1,a)} 森勢 将雅^{1,b)}

概要：本研究では日本語楽曲のメロディの特徴と歌詞中のモーラ出現傾向との関係を分析する。歌詞とメロディの対応を扱った研究では、音節の強勢やフレーズ境界といった言語的・構造的特徴が主な分析対象とされており、母音・子音単位の音韻要素とメロディ特徴との対応関係は十分に検討されていない。本研究では、日本語のポピュラー音楽 50 曲を対象に、音高・音程の方向・音程の絶対値・持続時間の 4 つの特徴量に基づいて音符を複数の群に分類し、各群におけるモーラ出現頻度を比較した。その結果、音高や音程に関する特徴ではモーラ出現頻度に顕著な差異は見られなかった一方、持続時間については短音条件において撥音・促音の出現頻度が高くなる傾向が確認され、モーラ出現傾向がリズムの要因と関連している可能性が示唆された。

1. はじめに

歌唱において、歌詞とメロディは独立した要素ではなく、相互に影響を与えながら統合的に知覚されることが知られている [1]。作詞・作曲の実践においても、メロディを先に作成し、後から歌詞を付与する「曲先」と呼ばれる手法が用いられている。このような制作過程においては、作詞家はメロディの特徴を踏まえながら、適合する言葉を選択することが望ましいとされている [2]。

言語は、単語の意味や文の構造に関わる言語的側面と、音の響きに関わる音響的側面という 2 つの観点から捉えられる。歌詞においても、これら 2 つの側面を区別して考えることが可能である。従来の歌詞を対象とした研究では、前者の観点から、韻律の特徴に着目した分析 [3] や、歌詞とメロディの構造的対応に関する分析 [4] などが主に行われている。一方、歌詞を構成するモーラ自身がもつ音の響きとメロディの特徴との対応を定量的に分析した研究は限られている。

歌詞を構成する母音や子音といった音の響きは、単語の意味とは独立に、聴き手の印象に影響を与える可能性がある。例えば、音象徴の研究では、母音や子音の違いが知覚的印象に影響を及ぼすことが報告されており [5]、プーバ/キキ効果 [6] はその代表的な事例として知られている。作詞の現場においても、母音や子音が異なる印象を持つことが知られており [2]、音の響きと印象との間に一定の対応関係があることが経験的に示されている。一方で、こうした

音響的な印象とメロディの特徴の関係が、実際の楽曲においてどのように現れるかは明らかでない。

本研究では、音響的特徴の観点から、日本語楽曲におけるメロディの特徴と歌詞中のモーラの出現傾向の関係を分析した。メロディの特徴として音高、音程の方向、音程の絶対値、および持続時間の 4 つを用いて音符を複数の群に分類し、各群における歌詞中のモーラ出現頻度を比較した。そして、メロディの特徴の違いに応じて、歌詞を構成するの母音や子音がどのように異なるかを調査した。

2. 関連研究と本研究の位置付け

2.1 歌詞の分析に関する研究

歌詞とメロディの対応関係については、楽曲の構造を考慮した分析が行われている。Nichols らは英語のポピュラー音楽を対象に、音節の強勢とメロディの強調位置の間に対応関係があることを示した [3]。Watanabe らは日本語楽曲を対象に、休符やフレーズ境界と単語や文の境界との間に対応関係があることを示し、楽曲の構造全体においてメロディと歌詞が対応して構成されることを明らかにした [7]。これらの先行研究は、言語的側面から歌詞とメロディの対応関係に着目したものである。

日本語の音韻やアクセントに着目した研究においても、メロディとの対応関係が報告されている。堤らは、日本語の童謡や唱歌を対象として、単語のアクセントとメロディの上下に一定の対応関係が見られることを示した [8]。また、靱山は促音および撥音を対象に、特殊モーラへの音符の割り当て方の違いが歌唱音響に反映されること、および両者の持続時間特性が異なることを明らかにした [9]。こ

¹ 明治大学

^{a)} cs252038@meiji.ac.jp

^{b)} mmorise@meiji.ac.jp

これらの先行研究では、日本語においても音韻・アクセントなどの言語的側面がメロディ構造と関わることを示唆している。

2.2 自動歌詞生成に関する研究

自動歌詞生成の研究では、楽譜の構造情報と歌詞の対応関係に着目した手法が提案されている。Oliveira らは、メロディのリズムに基づき、強拍と音節における強勢の対応を考慮した歌詞生成手法を提案した [10]。その後、意味的制約や周囲の語の関連を考慮し、リズムと意味の両立を図る手法へと拡張された [11]。これらの研究は、事前に定義した制約やルールに基づいて歌詞を生成している。

深層学習を用いた自動歌詞生成の研究では、Potash らが、LSTM を用いて特定アーティストのスタイルを模倣したラップ歌詞を生成する手法を提案した [12]。言語的側面を考慮した研究としては、Lu らによって音節構造と文脈情報を同時にモデル化することで、リズムと意味の両立を図った手法が提案された [13]。また、Watanabe らは、楽曲構造とメロディの対応関係の分析に基づき、その知見をモデルに組み込んだ歌詞生成手法を提案した [7]。これらは、大量のデータから歌詞の特徴を自動的に学習して歌詞を生成する点に特徴がある。

音響的特徴に着目した歌詞生成の研究として、石川らは、N-gram を用いて日本語を基に構築した芸術言語の歌詞構築手法を提案した [14]。芸術言語は、美的要素を重視して設計された人工言語であり、言語的側面よりも音響的側面を重視して設計される。この手法は、音の響きそのものを歌詞生成の基準とすることで、言語的な意味に依存しない歌詞構築を実現している。

2.3 本研究の位置づけ

歌詞とメロディの対応関係を扱う先行研究 [3], [7] では、音節の強勢やフレーズ境界といった言語的側面が主な分析対象とされており、母音・子音単位の音響的側面とメロディ特徴との対応関係の知見は限られている。日本語固有の特性に着目した研究 [8] においても、アクセントなどの言語的特徴に基づく対応関係が中心であり、母音・子音単位の対応関係については十分に検討されていない。音響的特徴に着目した歌詞生成の研究 [14] においても、音符数などの表層的な対応は扱われているものの、メロディがもつ構造的特徴の違いに基づく対応関係は十分に検討されていない。

音象徴の研究が示すように、音の響きは聴き手の印象に影響を与えることが知られており、歌詞を構成する音韻要素もメロディの特徴と何らかの対応関係を持つ可能性が考えられる。本研究では、実際の日本語楽曲を対象に、メロディの音楽的特徴と歌詞を構成する母音・子音単位の音韻要素との対応関係を明らかにすることを目的とする。

3. モーラ出現頻度の分析手法

3.1 データセットと前処理

本研究では、日本語楽曲の歌詞を解析するため、日本語楽曲データセットである東北きりたん歌唱データベース [15] を用いた。本データセットは日本語のポピュラー音楽全 50 曲から構成されており、各楽曲について歌詞を含む MusicXML 形式の楽譜データが収録されている。本データセットを用いて、各楽曲から歌詞とそれに対応する音符を抽出し、分析を行った。

データセットには、日本語の歌詞中に一部英語の単語が含まれている楽曲が存在する。日本語のリズムはモーラを基本単位とするのに対し、英語は強勢を基本単位とするため音節構造が大きく異なり、音符とモーラの対応関係が日本語と同様には成立しない [16]。このため、本研究では日本語の歌詞のみを分析対象として扱った。

日本語の歌詞と音符の対応付けを行うため、モーラ単位に基づく規則を用いた。日本語においては、歌詞と音符の対応にモーラ単位の制約があることが知られている [2]。原則として 1 音符につき 1 モーラが対応するが、例外として、2 モーラ目が以下のいずれかに該当する場合には、1 音符に 2 モーラが含まれることがある。なお、以下の例において左側がモーラ単位の分割、右側が音符単位での対応を示す。

- 促音 (例: 「き・っ・て」 → 「きっ・て」)
- 長音 (例: 「ボ・ー・ル」 → 「ボー・ル」)
- 撥音 (例: 「さ・ん・ぽ」 → 「さん・ぽ」)
- 母音 (例: 「あ・い・た・い」 → 「あい・たい」)

上記の原則および 4 つの例外のいずれにも該当しない音符は、対応規則が不明確であるため分析対象から除外した。

3.2 メロディの特徴に基づく分類

本研究では、メロディの構造的特徴として、相対音高、音程の方向、音程の絶対値および持続時間の 4 つの特徴を用いた。これらは音の高さ・動き・長さをそれぞれ示す特徴であり、メロディを多角的に捉えるために選定した。各特徴量の値に基づいて音符を群に分類し、群ごとのモーラ出現傾向を比較することで、メロディの特徴とモーラ出現傾向との関係を分析した。

3.2.1 相対音高に基づいた分類

音高とは、各音符の高さを表す特徴量であり、本研究では MIDI ノート番号 (0~127 の整数値) として数値化して扱った。楽曲はそれぞれ固有のキーや音域を持つため、絶対音高を直接比較すると、キーや音域の違いが分析結果に影響を与える可能性がある。この影響を低減するため、本研究では各楽曲ごとに音高の中央値 $\tilde{x}$ を基準とし、各音符の相対音高 m_i を、音高 x_i との差分 $m_i = x_i - \tilde{x}$ として定義した。

図 1 に相対音高別のモーラ数の分布を示す。横軸は相対

音高を半音単位で表し、縦軸はモーラ数を表す。分布は相対音高 0 付近で最大となり、正負両方向に離れるにつれて減少する山型を示した。分布の対称性と各群のデータ数の偏りを考慮し、相対音高 ± 1 を閾値として、低音域 (Low)、中音域 (Mid)、高音域 (High) の 3 群に分類した。

3.2.2 音程に基づいた分類

音程とは、隣接する音符間の音高差として定義される特徴量であり、旋律の動きを表す指標として用いた。音高の絶対値ではなく変化量に着目することで、旋律の上下や跳躍といった特徴を捉えることができる。曲頭や休符の直後は音程が定義できないため、分析対象から除外した。また、1 つの音符に対して 2 モーラ含まれる場合は、最初のモーラのみを評価対象として扱った。

図 2 に音程の方向別のモーラ分布を示す。横軸は音程を半音単位で表し、縦軸はモーラ数を表す。音楽理論では、旋律の動きの方向は上行・同音・下行の 3 種に大別され、上行は緊張感が強くなり感情が高揚する性質を持ち、下行は緩やかに落ち着く性質を持つとされている [17]。この区分に基づき、音高が正方向に動いた場合を上行 (Ascending)、同じ音高の場合を同音 (Unison)、負方向に動いた場合を下行 (Descending) として分類した。

図 3 に音程の絶対値ごとのモーラ分布を示す。横軸は音

程絶対値を半音単位で表し、縦軸はモーラ数を表す。絶対値 2 半音付近でモーラ数が最大となり、絶対値の増加に伴って減少する傾向が確認された。音楽理論においても、メロディの進行は 2 度以内の順次進行と 3 度以上の跳躍進行に大別される。順次進行は滑らかな性質を持つのに対し、跳躍進行は比較的角ばった性質を持つとされている [17]。音楽理論におけるこの区分とデータの分布を参考に、音程の絶対値が 0 の場合を同音 (Unison)、1~2 半音の場合を順次 (Stepwise)、3 半音以上の場合を跳躍 (Skipwise) と定義し、3 群に分類した。

3.2.3 持続時間に基づいた分類

持続時間は、各音符の発音長を表す特徴量であり、歌唱における時間的特性と関係する。楽譜に含まれる単純な音価では楽曲ごとのテンポの違いを考慮できないため、本研究では音価とテンポに基づき、各音符の長さを実時間 (ms) に変換することで持続時間を定義した。

図 4 に持続時間別のモーラ数の分布を示す。横軸は持続時間をミリ秒単位で表し、縦軸はモーラ数を表す。分布は 150 ms から 200 ms 付近と 300 ms から 350 ms 付近に 2 つのピークを持ち、ロングトーンを含む音符が存在するため高持続時間域まで分布が伸びている。2 つのピークの境界付近である 300 ms を閾値として、持続時間が 300 ms 未満の場合を短音 (Short)、300 ms 以上の場合を長音 (Long) と定め、2 群に分類した。

3.2.4 分類結果

表 1 に、各特徴量における群ごとのモーラ数および合計を示す。相対音高および持続時間は計 10,928 モーラを対象に分類した。音程の方向および絶対値については、曲頭や休符の直後の音符を除外したため、計 8,625 モーラを対象に分類した。

表より、各群に含まれるモーラ数には偏りがあることがわかる。そのため、本研究では各群におけるモーラ出現頻度を、各群内の総モーラ数で正規化した出現確率として扱い、メロディの特徴別に比較した。

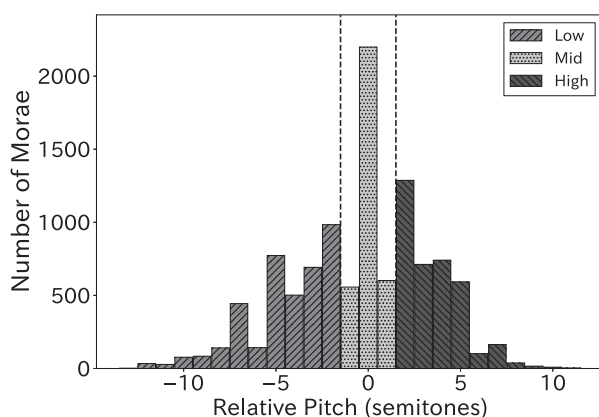


図 1 相対音高とモーラ数の分布

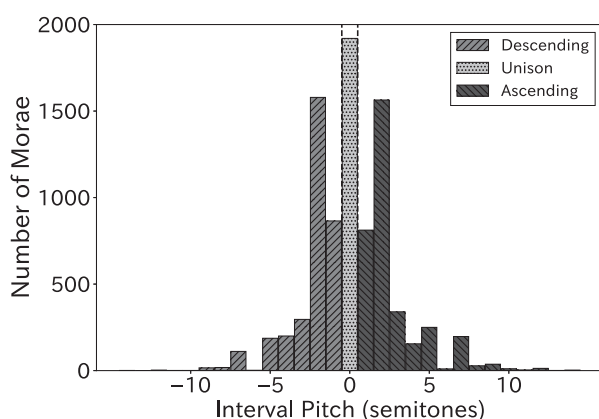


図 2 音程の方向とモーラ数の分布

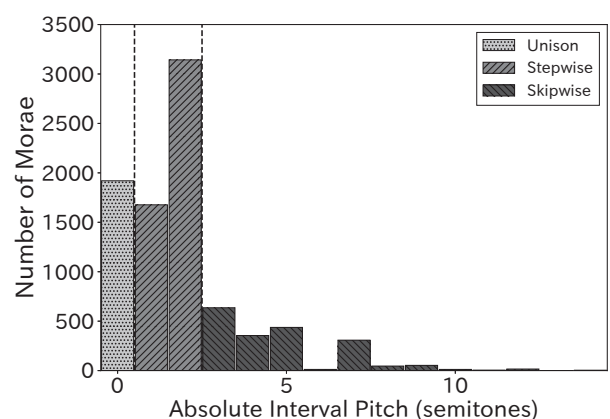


図 3 音程の絶対値とモーラ数の分布

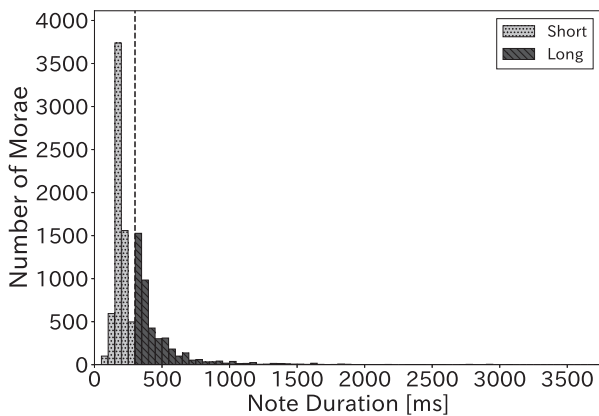


図 4 音の持続時間とモーラ数の分布

表 1 各メロディ特徴におけるモーラ数の分布

特徴量	群	モーラ数	合計
相対音高	低音域 (Low)	3,927	10,928
	中音域 (Mid)	3,314	
	高音域 (High)	3,687	
音程方向	上行 (Ascending)	3,425	8,625
	同音 (Unison)	1,920	
	下行 (Descending)	3,280	
音程絶対値	同音 (Unison)	1,920	8,625
	順次 (Stepwise)	4,820	
	跳躍 (Skipwise)	1,885	
持続時間	短音 (Short)	6,489	10,928
	長音 (Long)	4,439	

4. 結果

4.1 相対音高とモーラ出現頻度

相対音高とモーラ出現頻度のヒートマップを図 5 に示す。縦軸は母音、横軸は子音を表し、# は母音なし、/cl/ は促音、/N/ は撥音を示す。以降のヒートマップも同様の形式で示す。全体として、相対音高の違いによるモーラ出現頻度の分布に顕著な差異は見られず、低音域・中音域・高音域の各条件のヒートマップは類似した傾向を示した。

母音では /a/, /i/, /o/ の出現頻度が高い傾向が全条件で一致していた。子音では /n/, /r/, /m/, /k/, /t/ などの出現頻度が高く、音高条件間で分布の大きな変化は確認されなかった。撥音 /N/ および促音 /cl/ についても、音高による違いは認められなかった。

4.2 音程とモーラ出現頻度

音程方向とモーラ出現頻度のヒートマップを図 6 に、音程絶対値とモーラ出現頻度のヒートマップを図 7 に示す。音程方向および音程絶対値のいずれにおいても、条件間でモーラ出現頻度の分布に顕著な差異は見られず、各条件のヒートマップは全体として類似した傾向を示した。

母音では /a/, /i/, /o/ の出現頻度が高い傾向が両特徴

量の全条件で共通していた。子音では /n/, /r/, /m/, /k/, /t/ などが比較的多く出現しており、音程方向・音程絶対値ともに条件間で分布の大きな変化は確認されなかった。撥音 /N/ および促音 /cl/ についても、すべての条件間で類似していた。

4.3 持続時間とモーラ出現頻度

持続時間とモーラ出現頻度のヒートマップを図 8 に示す。相対音高および音程と異なり、持続時間の条件間では一部のモーラにおいて出現頻度が異なる傾向が示された。

母音および子音については条件間で大きな差は確認されず、両条件で類似した分布を示した。一方、撥音 /N/ および促音 /cl/ については、長音条件と比較して短音条件での出現頻度が高い傾向が確認された。

4.4 考察

本研究の結果より、相対音高、音程方向、音程の絶対値などの音の高さに関わる要因については、モーラ出現頻度に顕著な差異は見られなかった。単語アクセントに着目した先行研究 [8] では、童謡・唱歌においてアクセントとメロディの音程に高い一致が見られることが報告されている。しかし、これは単語・文節単位での分析であり、本研究のように母音・子音といったモーラ内部の構成要素単位で分析した場合には、音高との明確な対応関係は確認されなかった。このことから、音の高さや動きに関しては歌詞の言語的側面との対応が強く、モーラ単位では現れにくい可能性がある。

持続時間については、短音条件において撥音 /N/ および促音 /cl/ の出現頻度が高くなる傾向が確認された。撥音や促音は他のモーラと同じ音符に割り当てられることがあり、その場合は各モーラの持続時間が短くなる。そのため、短音条件においてこれらの特殊モーラが出現しやすくなったと考えられる。この傾向は、撥音および促音が持続時間に関わる特性を持つとする先行研究 [9] とも整合的である。この結果から、歌詞中のモーラ出現傾向は音高に関わる要因よりも、持続時間に関わるリズムの要因と関連している可能性が示唆される。

4.5 本結果の課題

本研究には、データの偏りと統計的検証の不足という 2 つの課題がある。分析対象が日本語のポピュラー音楽 50 曲に限られており、楽曲や歌手、作詞家の偏りの影響を十分に排除できていない。そのため、他のジャンルや異なるデータセットを用いた場合には、結果が異なる可能性がある。

統計的検証については、各群に含まれるモーラ数に偏りがあるため、比較結果に影響を与えている可能性がある。さらに、出現確率の差異に対する統計的検証を行っていない

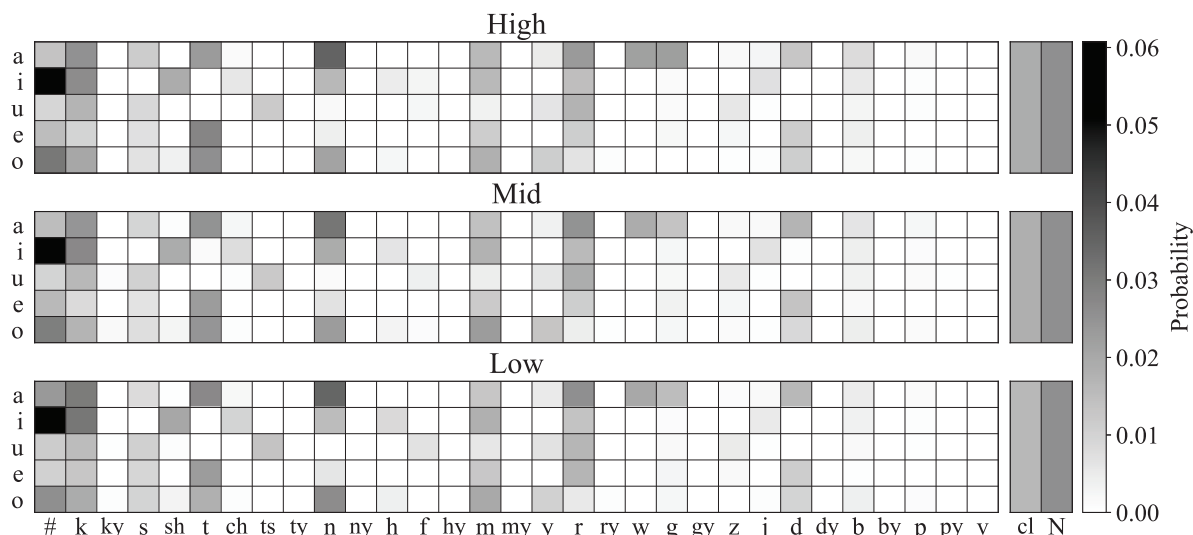


図 5 相対音高とモーラ出現頻度のヒートマップ

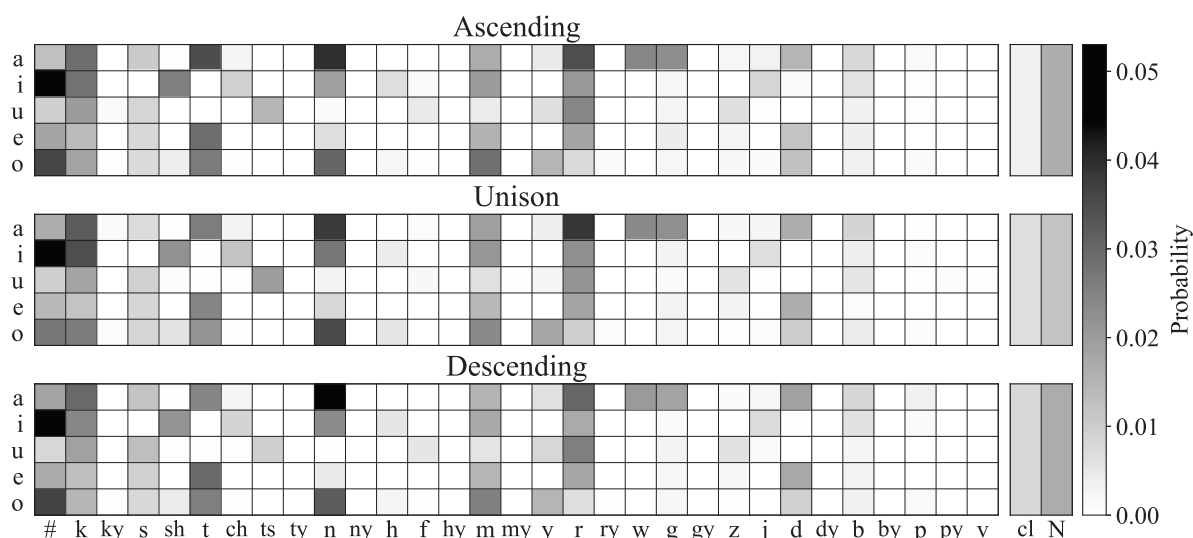


図 6 音程方向とモーラ出現頻度のヒートマップ

いため、観察された傾向の有意性については確認されていない。

5. おわりに

本研究では、日本語のポピュラー音楽 50 曲を対象に、音符と歌詞を抽出し、音高・音程の方向・音程の絶対値・持続時間の 4 つの特徴量に基づいて音符を群に分類した。各群のモーラ出現頻度を比較した結果、音高や音程に関する特徴量では条件間に顕著な差異は見られなかった。一方、持続時間については、短音条件において撥音 /N/ および促音 /cl/ の出現頻度が高くなる傾向が確認され、モーラ出現傾向がリズム的要因と関連していることが示唆された。

今後は、多様なジャンルや作詞家を含む大規模データへの拡張および統計的検定の導入により、本研究で観察された傾向をより客観的に検証していく。また、隣接音符間の比較にとどまらないフレーズ単位での分析を行うことで、

メロディの大局的な特徴とモーラ出現傾向との関係をより詳細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] Serafine, M. L., Crowder, R. G. and Repp, B. H.: "Integration of melody and text in memory for songs", *Cognition*, Vol. 16, No. 3, pp. 285–303 (1984).
- [2] 野口義修: "15 秒作詞入門", ヤマハミュージックエンタテイメントホールディングス (2019).
- [3] Nichols, E., Morris, D., Basu, S. and Raphael, C.: "Relationships between lyrics and melody in popular music", *Proc. ISMIR*, pp. 471–476 (2009).
- [4] Watanabe, K. and Goto, M.: "Lyrics information processing: Analysis, generation, and applications", *Proc. of the 1st Workshop on NLP for Music and Audio (NLP4MusA)*, Association for Computational Linguistics, pp. 6–12 (2020).

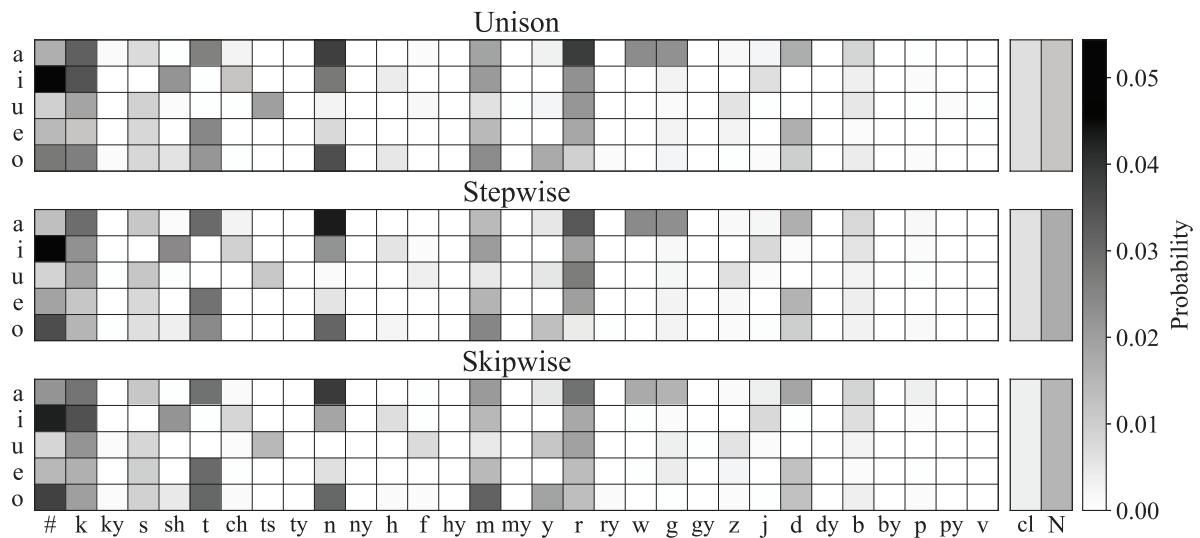


図 7 音程絶対値とモーラ出現頻度のヒートマップ

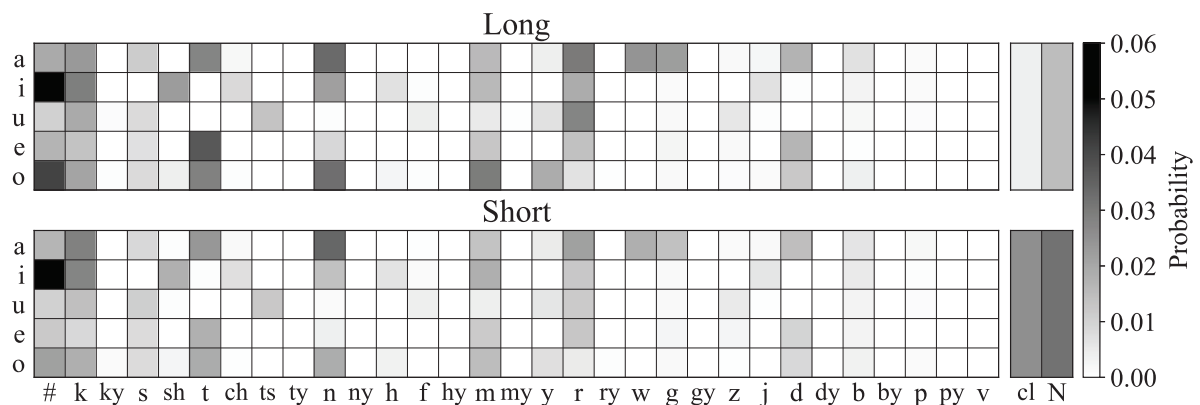


図 8 持続時間とモーラ出現頻度のヒートマップ

- [5] Ahlner, F. and Zlatev, J.: “Cross-modal iconicity: A cognitive semiotic approach to sound symbolism”, *Sign Systems Studies*, Vol. 38, No. 1/4, pp. 298–348 (2010).
- [6] Ramachandran, V. S. and Hubbard, E. M.: “Synaesthesia — A window into perception, thought and language”, *Journal of Consciousness Studies*, Vol. 8, No. 12, pp. 3–34 (2001).
- [7] Watanabe, K., Matsubayashi, Y., Fukayama, S., Goto, M., Inui, K. and Nakano, T.: “A melody-conditioned lyrics language model”, *Proc. of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, pp. 163–172 (2018).
- [8] 堤 彩香, 平賀 譲: “日本語の音韻と旋律の関係について～童謡・唱歌を中心に～”, 情報処理学会研究報告, Vol. 2014-MUS-102, No. 5, pp. 1–6 (2014).
- [9] 靱山陽子: “歌唱音響に見られる促音と撥音の歌詞付けの相違”, 情報処理学会第 79 回全国大会講演論文集, pp. 25–26 (2017).
- [10] Gonalo Oliveira, H., Cardoso, A. and Pereira, F.: “Tra-la-Lyrics: An approach to generate text based on rhythm”, pp. 47–55 (2007).
- [11] Gonalo Oliveira, H.: “Tra-la-Lyrics 2.0: Automatic generation of song lyrics on a semantic domain”, *Journal of Artificial General Intelligence*, Vol. 6, No. 1, pp. 87–110 (2015).
- [12] Potash, P., Romanov, A. and Rumshisky, A.: “Ghost-Writer: Using an LSTM for automatic rap lyric generation”, *Proc. of the 2015 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 1919–1924 (2015).
- [13] Lu, X., Wang, J., Zhuang, B., Wang, S. and Xiao, J.: “A syllable-structured, contextually-based conditionally generation of Chinese lyrics”, *Proc. of the Pacific Rim International Conference on Artificial Intelligence (PRICAI)*, Springer-Verlag, pp. 257–265 (2019).
- [14] 石川真大, 森勢将雅: “N-gram による歌声合成に向けた芸術言語構築法の検討”, 情報処理学会研究報告, Vol. 2025-SLP-158, No. 24, pp. 1–6 (2025).
- [15] Ogawa, I. and Morise, M.: “Tohoku kiritan singing database: A singing database for statistical parametric singing synthesis using japanese pop songs”, *Acoustical Science and Technology*, Vol. 42, No. 3, pp. 140–145 (2021).
- [16] Otake, T., Hatano, G., Cutler, A. and Mehler, J.: “Mora or syllable? Speech segmentation in Japanese”, *Journal of Memory and Language*, Vol. 32, No. 2, pp. 258–278 (1993).
- [17] 高山 博: “ポピュラー音楽作曲のための旋律法 一聴く人の心に響くメロディラインの作り方ー増補版”, リットーミュージック (2021).