

本著作物の著作権は情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。利用にあたっては、著作権法ならびに[情報処理学会倫理綱領](#)に従ってください。

The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). The material has been made available on the website by the author(s) under the agreement with the IPSJ. Please be complied with Copyright Law of Japan and [Code of Ethics of the IPSJ](#) if any users wish to reproduce, distribute, or modify the whole or any part of the material.

ポピュラー音楽のセクション識別における メロディ情報の有効性

宮澤 響^{1,a)} 平賀 譲^{1,b)}

概要: ポピュラー音楽の多くはAメロ、Bメロ、サビと呼ばれるセクションから構成されている。これらのセクションは、単独でもそれぞれの音楽的特徴によりかなりの程度識別し得ることが分かっているが、どのような要素（メロディライン、和声／コード、伴奏、歌詞など）によって識別されるのかは明らかでない。そこで、本研究では、未知のポピュラー音楽を対象として、各セクションのメロディラインのみを単独で実験参加者に提示し、どのセクションだと思うかを回答させる心理実験を実施した。その結果、いずれのセクションの正答率もチャンスレベルよりも高く、十分にセクション識別が可能であることが分かった。このことから、メロディラインはセクション識別の重要な要素の一つであると考えられる。

キーワード: セクション、メロディライン、ポピュラー音楽、サビ、キャッチネス

Significance of melodic information in section identification tasks of popular music

HIBIKI MIYAZAWA^{1,a)} YUZURU HIRAGA^{1,b)}

Abstract: Popular music typically consist of sections called verse-A, verse-B, and chorus. These sections can be identified to a certain extent based on their inherent music properties alone. However, what features of the music (melody line, harmony/chord, accompaniment, lyrics, etc.) are dominant in such judgments are still unclear. In this study, we conducted a psychological experiment which tried to identify whether melody line information alone is sufficient for section identification. The participants were presented with melody lines extracted from unknown songs, and were asked to identify whether the segment was from verse-A, verse-B or chorus. The results show that the participants performed above chance level and sections were identified sufficiently, suggesting that melodic information include significant cues for section identification.

Keywords: section, melody line, popular music, chorus, catchiness

1. はじめに

ポピュラー音楽の多くはAメロ、Bメロ、サビと呼ばれるセクションから構成されている。これらのセクションには、それぞれ独自の特徴が存在すると考えられており、中でもサビは、人々の好みに最も影響を与えるセクションであるとされる [1][2]。

村松は、21 名の実験参加者に小室哲哉作曲の楽曲 12 曲を提示し、それぞれの楽曲のどこがサビであると思うかを回答させる実験を実施した [3]。その結果、90% 以上の実験参加者が正しいサビの位置を回答した。しかし、この実験では、小室哲哉作曲の楽曲を刺激に用いたため、実験参加者の中には、サビの位置を最初から知っていた者も存在した。また、楽曲をそのまま提示したため、刺激を聴いたことがなかった実験参加者も、Aメロ→Bメロ→サビという楽曲の基本構造からの推測により、楽曲の終盤部分を答えることで正解することも可能であった。

¹ 筑波大学大学院 図書館情報メディア研究科
Graduate School of Library, Information and Media Studies,
University of Tsukuba

a) s1821634@s.tsukuba.ac.jp

b) hiraga@slis.tsukuba.ac.jp

そこで我々は、村松の実験では不明確であった、未知の楽曲であってもサビを識別できるのかということや、1曲のセクションの順序を並び替えても、或いはセクション単独でもサビを識別できるのかということを調べるために、相対評価実験と絶対評価実験という2つの実験を実施した[4]。

これらの実験には、RWC 研究用音楽データベース：ポピュラー音楽データベース [5] に収録されている日本のポピュラー音楽形式の楽曲のうち、1 コーラスの基本構造がAメロ→Bメロ→サビである楽曲30曲を、AIST Annotation for the RWC Music Database[6] に従い、Aメロ、Bメロ、サビに分割したものを使用した。

相対評価実験では、1曲の中の3つのセクションの順序をランダムに並び替えたものを刺激とした。なお、それぞれの刺激には、Aメロ、Bメロ、サビは必ず含まれている。これを18名の実験参加者に提示し、それぞれの刺激のどこがどのセクションであると思うか、つまり、刺激のセクションがどのような順序であると思うかを回答させた。その結果、サビだけでなく、いずれのセクションにおいても90%程度の正答率を得た。このことから、未知の楽曲であっても、また、セクションの順序が並び替わっていても、これらのセクションは識別可能であるということが分かった。

絶対評価実験では、各楽曲のそれぞれのセクションを単独で刺激とした。これを18名の実験参加者に提示し、それぞれの刺激がどのセクションであると思うかを回答させた。その結果、いずれのセクションにおいても80%程度の正答率を得た。このことから、それぞれのセクションは同じ楽曲の他のセクションと比較せずとも、単独で識別可能であるということが分かった。

しかし、これらの実験では、メロディライン、和声／コード、伴奏、歌詞など、様々な要素がセクション識別の手がかりとなっているため、それぞれのセクションがどのような要素によって識別されるのかは明らかでない。

このような背景から、本稿では、ポピュラー音楽のそれぞれのセクションが、メロディラインのみを聴取した場合であっても識別され得るのかを、心理実験により調査した結果を報告する。

表 1 各セクション長の最大値および最小値

Table 1 maximum and minimum lengths of sections

	Aメロ	Bメロ	サビ
最大値 (秒)	39.45	26.31	39.18
最小値 (秒)	11.29	10.00	12.90

2. 研究手法

2.1 提示刺激

実験には、RWC 研究用音楽データベース：ポピュラー音楽データベース [5] に収録されている楽曲のうち、相対評価実験および絶対評価実験で使用したのと同じ楽曲30曲のメロディラインを、AIST Annotation for the RWC Music Database[6] に従い、Aメロ、Bメロ、サビに分割したもの（30曲 × 3セクション = 90刺激）を使用した。具体的には、[6]に含まれている、楽曲を可能な範囲で再現したSMFから、メロディラインのトラックのみを切り出し、音色をMIDI音色番号No. 1のピアノ音に、音量を一定にしたものを、GarageBand内蔵音源によりwav形式の音響信号データとして書き出し、各セクションに分割したものを刺激とした。なお、調やテンポ、セクションの長さなどは刺激ごとに異なり、各セクション長の最大値および最小値は表1の通りである。

2.2 実験概要

実験では、20歳から25歳の健聴者18名（男性9名、女性9名、平均年齢22.3歳）に対して、2.1で説明した刺激を単独で提示する試行を90回繰り返した。提示順序については、全実験参加者に対して90刺激全ての順序をランダム化した。ただし、それぞれの刺激の提示は一度きりであり、聴き直しをすることはできない。また、実験参加者は絶対評価実験や相対評価実験の18名とは異なり、これらの実験に重複して参加した者はいない。実験は、実験参加者一人一人に対して個別に実施した。

実験参加者には、

- ・刺激がAメロ、Bメロ、サビのどれであると思うか
- ・回答に対する確信度（5件法）
- ・刺激のキャッチネス（5件法）

の3つを順に回答してもらった。ここで、確信度とは、回答に対してどれだけ自信があるかを表す、自信の度合いのことである。また、キャッチネスとは、刺激がどれだけキャッチーに感じられたかを表す、キャッチーさの度合いのことである。どちらも1が最低値（自信がない／キャッチーでない）であり、5が最高値（自信がある／キャッチーである）である。

刺激はPCの左右に配置された2台のスピーカーにより提示し、各刺激の提示直前には500msのホワイトノイズによるキャンセリングを行った。また、何を回答するかの指示はPCの画面に表示し、回答はPCのキーボードから入力させた。

3. 実験結果

3.1 正答率

図 1 は、セクションごとの刺激に対する正答率の平均値を表しており、エラーバーは標準誤差を表している（以下、エラーバーは全て標準誤差を表す）。正答率の主効果について、実験参加者内計画による一要因分散分析を行ったところ、 $F(2, 34) = 2.066, p > .05$ となり、有意ではなかった。また、それぞれのセクションの正答数に対してカイ二乗検定による適合度検定を行ったところ、いずれのセクションの正答率においても有意差 ($p < .001$) が見られた。

つまり、メロディラインのみを聴取した場合のセクション識別においては、各セクションの正答率の間には差があるとは言えず、いずれのセクションの正答率も、チャンスレベル ($\approx 33.3\%$) よりも高い。

3.2 確信度

図 2 は、セクションごとの刺激に対する確信度の平均値を表している。確信度的主効果について、実験参加者内計画による一要因分散分析を行ったところ、 $F(2, 34) = 20.649, p < .001$ となり、有意であった。また、Bonferroni 法による多重比較を行ったところ、A メロと B メロの間には有意差 ($p < .05$) が、A メロとサビの間には有意差 ($p < .01$) が、B メロとサビの間には有意差 ($p < .001$) が、それぞれ見られた。

つまり、B メロに比べて A メロの確信度が高く、A メロ、B メロに比べてサビの確信度が高い。

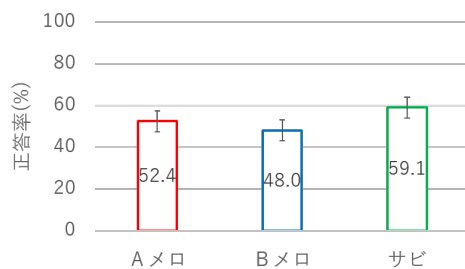


図 1 正答率

Fig. 1 percentage of correct answers

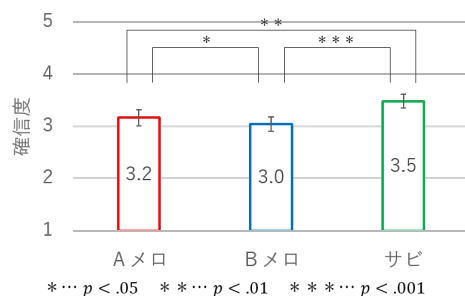


図 2 確信度

Fig. 2 degree of confidence

3.3 キャッチネス

図 3 は、セクションごとの刺激に対するキャッチネスの平均値を表している。キャッチネスの主効果について、実験参加者内計画による一要因分散分析を行ったところ、 $F(2, 34) = 57.964, p < .001$ となり、有意であった。また、Bonferroni 法による多重比較を行ったところ、A メロと B メロ、A メロとサビ、B メロとサビ、いずれの間も有意差 ($p < .001$) が見られた。

つまり、B メロに比べて A メロのキャッチネスが高く、A メロ、B メロに比べてサビのキャッチネスが高い。

3.4 相関

図 4, 図 5, 図 6 はそれぞれ、正答率、確信度、キャッチネスそれぞれの組み合わせの散布図を表しており、黒色の直線は回帰直線である。また、表 2, 表 3, 表 4 はそれぞれ、セクションごとの刺激に対するそれぞれの値の組み合わせのピアソンの積率相関係数を表している。

いずれのセクションにおいても、確信度とキャッチネスの間にそれぞれ正の相関が見られる。また、A メロに関しては正答率と確信度、正答率とキャッチネスの間にそれぞれ負の相関が見られる一方、サビに関してはそれらの間に相関係数 0.8 以上の強い正の相関が見られる。

3.5 誤答の内訳

図 7, 図 8, 図 9 はそれぞれ、セクションごとの刺激に対する誤答の内訳を表している。これらのグラフの横軸は全て、セクションごとの全 30 曲分の刺激であり、例えば、A1, B1, C1 は、同じ楽曲 ID. 1 の楽曲の A メロ、B メロ、サビ (chorus) を意味する。また、表 5 は、それぞれのセクションの誤答率をまとめたものであり、例えば、上段中央の“22.6”は、A メロの刺激に対して B メロと回答する誤答が全体の 22.6% であったことを意味する。

A メロの刺激に対しては、B メロと回答する誤答に比べてサビと回答する誤答が多く、B メロの刺激に対しては、サビと回答する誤答に比べて A メロと回答する誤答が多い。また、サビの刺激に対しては、B メロと回答する誤答に比べて A メロと回答する誤答が多い。

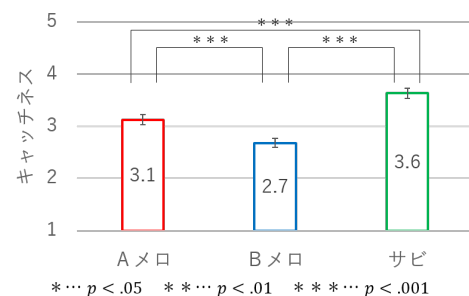


図 3 キャッチネス

Fig. 3 catchiness

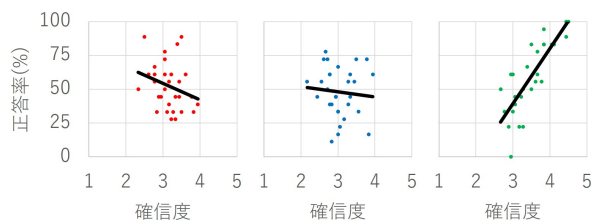


図 4 散布図 (正答率-確信度)

Fig. 4 scatter plot
(percentage of correct answers - degree of confidence)

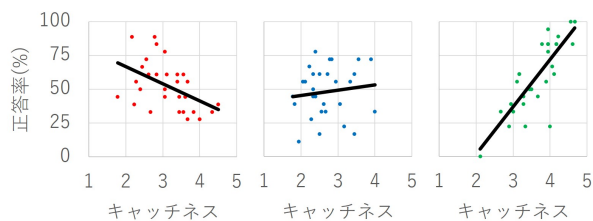


図 5 散布図 (正答率-キャッチネス)

Fig. 5 scatter plot (percentage of correct answers - catchiness)

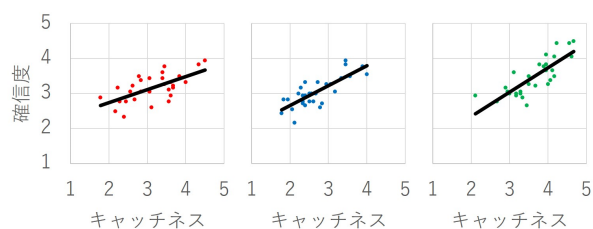


図 6 散布図 (確信度-キャッチネス)

Fig. 6 scatter plot (degree of confidence - catchiness)

表 2 相関係数 (Aメロ)

Table 2 correlation coefficients (verse-A)

Aメロ	正答率	確信度	キャッチネス
正答率	1		
確信度	-0.27	1	
キャッチネス	-0.48	0.64	1

表 3 相関係数 (Bメロ)

Table 3 correlation coefficients (verse-B)

Bメロ	正答率	確信度	キャッチネス
正答率	1		
確信度	-0.09	1	
キャッチネス	0.12	0.79	1

表 4 相関係数 (サビ)

Table 4 correlation coefficients (chorus)

サビ	正答率	確信度	キャッチネス
正答率	1		
確信度	0.82	1	
キャッチネス	0.81	0.80	1

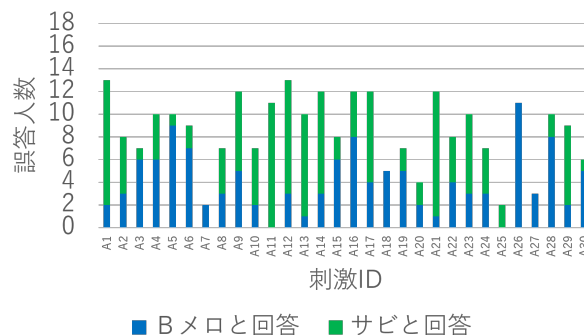


図 7 誤答の内訳 (Aメロ)

Fig. 7 incorrect answers (verse-A)

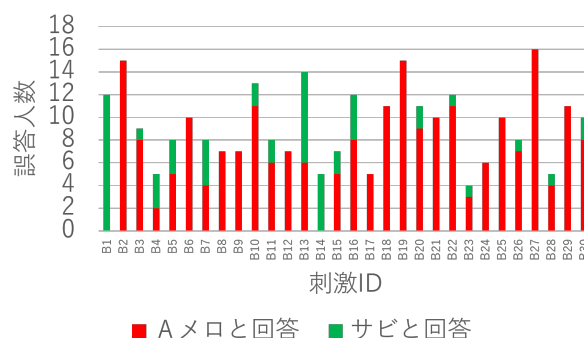


図 8 誤答の内訳 (Bメロ)

Fig. 8 incorrect answers (verse-B)

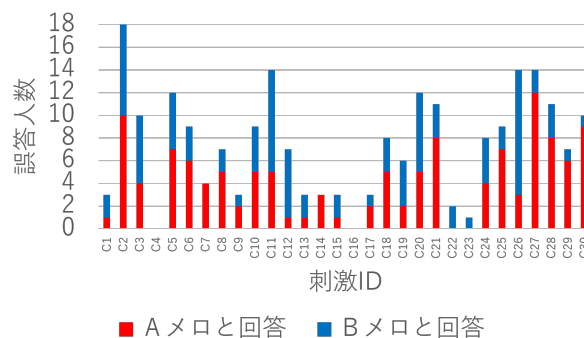


図 9 誤答の内訳 (サビ)

Fig. 9 incorrect answers (chorus)

表 5 誤答率

Table 5 percentage of incorrect answers

刺激		回答		
		Aメロ	Bメロ	サビ
刺激	Aメロ	—	22.6	25.0
	Bメロ	42.0	—	10.0
	サビ	23.3	17.6	—

3.6 絶対評価実験との比較

図 10 は、本実験での正答率と絶対評価実験での正答率とをセクションごとに併記したものであり、“melo”が本実験、“abs”が絶対評価実験のものである。2つの実験の正答率について、評価方法（本実験、絶対評価実験）とセクション（Aメロ、Bメロ、サビ）の混合計画による二要因分散分析を行ったところ、評価方法の主効果については、 $F(1, 34) = 98.240, p < .001$ となり、有意であった。一方、セクションの主効果については、 $F(2, 68) = 1.855, p > .05$ となり、有意ではなかった。また、評価方法とセクションの交互作用についても、 $F(2, 68) = 1.153, p > .05$ となり、有意ではなかった。

つまり、いずれのセクションにおいても、絶対評価実験での正答率に比べて本実験での正答率が低い。

図 11 は、絶対評価実験において正答率が 80% 以上であった刺激、すなわち、セクション識別が容易であった刺激のみの、本実験における正答率を集計したものである。

セクション識別が容易なサビの刺激に関しては、メロディラインのみを聴取した場合であっても、70% 以上と高い正答率を示している。

表 6 は、本実験での誤答率と絶対評価実験での誤答率とをまとめたものである。

本実験と絶対評価実験との共通点としては、Bメロの刺激に対してAメロと回答する誤答が多いことが挙げられる。

一方、相違点としては、Aメロの刺激に対して、本実験では、サビと回答する誤答が多いのに対し、絶対評価実験では、Bメロと回答する誤答が多いことが挙げられる、また、サビの刺激に対して、本実験では、Aメロと回答する誤答が多いのに対し、絶対評価実験では、Aメロと回答する誤答、Bメロと回答する誤答がそれぞれ同程度であることが挙げられる。

4. 考察

4.1 正答率

3.1 より、全てのセクションの正答率がチャンスレベルよりも有意に高いため、メロディラインのみを聴取した場合であっても、セクション識別は十分に可能であると言える。このことから、メロディラインはセクション識別の重要な要素の一つであると考えられる。一方、各セクションの正答率の間に差があるとは言えないため、サビのみが識別しやすいということはなく、セクション識別の難易度に差はない。

また、3.6 より、いずれのセクションにおいても、絶対評価実験での正答率に比べて本実験での正答率が有意に低いことから、メロディラインのみでも識別は可能である場合が多いものの、伴奏などのメロディライン以外の要素も、セクション識別において一定の役割を担っていると言える。

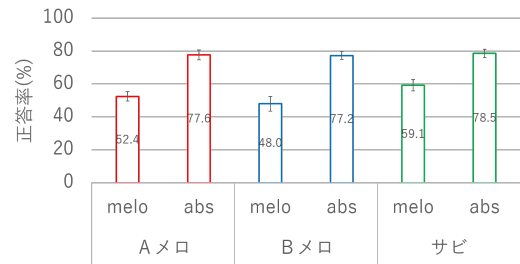


図 10 本実験と絶対評価実験との正答率の比較

Fig. 10 percentage of correct answers compared to absolute evaluation

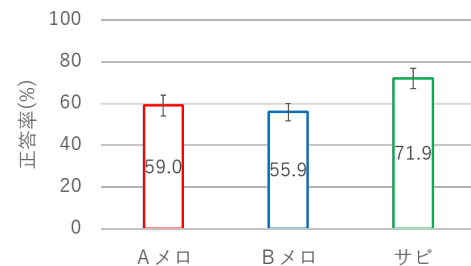


図 11 絶対評価実験において正答率が 80% 以上であった刺激の本実験における正答率

Fig. 11 percentage of correct answers of the stimulation answered more than 80% in absolute evaluation

表 6 本実験と絶対評価実験との誤答率の比較

Table 6 percentage of incorrect answers compared to absolute evaluation

		回答		
		Aメロ	Bメロ	サビ
本実験	Aメロ	—	22.6	25.0
	Bメロ	42.0	—	10.0
	サビ	23.3	17.6	—
絶対評価実験	Aメロ	—	14.6	7.8
	Bメロ	15.9	—	6.9
	サビ	10.4	11.1	—

このことに関して、絶対評価実験からの正答率低下の原因は、本実験ではメロディライン以外の情報が欠落しているということに他ならないが、その中でも特に、音の厚みやドラムパターンといった伴奏の盛り上がり具合に関する情報や、伴奏から感じ取ることができるコード進行の情報が欠落していることが、正答率低下に繋がる大きな原因であったと考えられる。

加えて、本実験では、2.1 の通り、メロディラインの音量は一定とし、メトロノームのように拍を刻む音も付与しなかった。そのため、刺激の中には拍の認識が難しいものも存在したと考えられるが、拍が分かりづらい刺激の場合、拍をとらえることに注意を払ってしまい、メロディライン自体の聴取に注意が向かなくなってしまうおそれがあるため、拍の認識のしやすさもセクション識別の難易度に影響を与えている可能性がある。

また、絶対評価実験において正答率が高かったサビの刺激は、本実験においても正答率が高かったことから、サビに関しては、セクション識別に必要な情報全体に対してメロディラインが占める割合がAメロ、Bメロに比べて高いのではないかと考えられる。

4.2 確信度

3.2 より、Aメロ、Bメロの確信度に比べてサビの確信度が有意に高いため、サビに対しては、より自信をもって回答していると言える。にもかかわらず、3.1 の通り、サビの正答率とAメロ、Bメロの正答率との間に差があるとは言えないため、Aメロ、Bメロに比べて、サビは自信のない回答が誤答となりやすいと考えられる。なお、この傾向は絶対評価実験の結果とも共通している。

4.3 キャッチネス

3.3 より、Aメロ、Bメロのキャッチネスに比べてサビのキャッチネスが有意に高いため、サビはキャッチーに感じられる。これは、一般的な理解や経験則とも一致しており、絶対評価実験の結果とも共通している。このことから、サビをキャッチーに感じる要因には、メロディラインが大きく関係していると言える。

4.4 相関

3.4 より、サビの正答率、確信度、キャッチネスの間にそれぞれ相関係数 0.8 以上の強い正の相関が見られたため、キャッチーに感じられる刺激は自信をもってサビだと回答し、そして正解していると言える。一方、Aメロの正答率とキャッチネスの間には負の相関が見られた。これは、キャッチーなAメロは誤答しやすいことを意味しており、3.5 の結果や、サビはキャッチーなもの、Aメロはキャッチーでないもの、という一般的な理解とも一致する。

4.5 誤答の内訳

3.5 より、Aメロの刺激に対してはサビと回答する誤答が多く、サビの刺激に対してもAメロと回答する誤答が多い。また、3.4 より、キャッチーなAメロは誤答しやすい。これらのことから、Aメロのメロディライン、その中でも特にキャッチーなメロディラインとサビのメロディラインとの間には、非常に高い類似性があると考えられる。

また、Bメロの刺激に対しては、Aメロと回答する誤答が多いものの、確信度がそれほど高くないため、BメロはAメロと識別されやすいというよりも、サビではないことは分かるが、Aメロ、Bメロどちらであるかの判断はつかない、という場合が多いのではないかと考えられる。

これらを踏まえると、Aメロとサビは取り違えやすいが、Bメロとサビを取り違えることはあまり多くはないと言える。

これは、ポピュラー音楽には、明るい印象のAメロから、少し印象を暗くしたBメロを経て、サビで再び明るい印象に戻る、というような構成の楽曲が多いため、Aメロとサビとでメロディラインから受ける印象が近いことに起因するのではないかと考えられる。また、Bメロはサビへの繋ぎとしての役割を担っているセクションであることから、サビとBメロのメロディラインは、違いがはっきりと分かるように作られていることも理由の一つではないかと考えられる。

5. おわりに

本稿では、未知のポピュラー音楽を対象として、各セクションのメロディラインのみを単独で実験参加者に提示し、どのセクションだと思うかを回答させる心理実験を実施した。その結果、いずれのセクションの正答率もチャンスレベルよりも高く、十分にセクション識別が可能であることが分かった。このことから、メロディラインはセクション識別の重要な要素の一つであると考えられる。

なお、現在、伴奏のみの刺激を使用した同様の実験手続きの心理実験を計画中である。そのため、相対評価実験、絶対評価実験、本実験、そしてその実験の実験結果を横断的に分析し、セクション識別に対するメロディラインや伴奏の影響を考察していくことが今後の課題である。

謝辞 本研究には RWC 研究用音楽データベースを使用いたしました。後藤真孝氏をはじめとするデータベースの構築、管理に携わる方々に感謝いたします。お忙しい中、実験協力を快く引き受けていただいた実験参加者の皆さんに感謝いたします。人と音の情報学研究室の寺澤洋子先生、先輩、同期、後輩には多くの助言とお力添えを頂きました。ここに深く感謝いたします。

参考文献

- [1] Goto, M.: SmartMusicKIOSK: Music Listening Station with Chorus-Search Function, *Proc. 16th Annual ACM Symposium on User Interface Software and Technology*, pp. 31–40 (2003).
- [2] 大野直紀, 中村聡史, 山本岳洋, 後藤真孝: 音楽動画への印象評価データセット構築とその特性の調査, 情報処理学会研究報告, Vol. 2015-MUS-108, No. 7, pp. 1–9 (2015).
- [3] 村松 純: 歌謡曲における「さび」の楽譜情報に基づく特徴抽出 —小室哲哉の場合—, 情報処理学会研究報告, Vol. 2000-MUS-35-1, No. 49, pp. 1–6 (2000).
- [4] 宮澤 響, 平賀 譲: ポピュラー音楽における人間のサビ認識に関する研究, 情報処理学会研究報告, Vol. 2018-MUS-118, No. 27, pp. 1–6 (2018).
- [5] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC 研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース, 情報処理学会研究報告, Vol. 2001-MUS-42-6, No. 103, pp. 35–42 (2001).
- [6] Goto, M.: AIST Annotation for the RWC Music Database, *Proc. 7th International Conference on Music Information Retrieval*, pp. 359–360 (2006).