

# ポピュラー音楽のボーカル音源分離結果に対する 音響特徴量と情動の関連の可視化

吉田真尋 (指導教員：伊藤貴之)

## 1 はじめに

機械学習を用いた楽曲レコメンド機能が発展する一方で、人がなぜその楽曲を好きになるのか、楽曲に何を感じるのかはいまだに明らかになっていない。音楽と感情に関する研究は主にボーカルのない楽曲を対象にしたものが中心であるが、本研究では同じ楽曲でも歌手が違くと想起される感情が異なることに注目し、ポピュラー音楽 59 曲をボーカルと伴奏に音源分離した上でそれぞれについて Valence(感情価) と Arousal(覚醒度)[1] を評価した結果を可視化し、感情に影響を与える音響特徴量を探る。

## 2 関連研究

現在の音楽と感情に関する研究は歌詞の影響を排除するためボーカルのない楽曲を対象としたものが多く、テンポと調性が Valence に影響を与えることが明らかになっている [2]。また、音の大きさと Arousal に関しては正の相関があること [3] も示されているが、音色に関しては一貫した結果が得られていない。その一方で、声に含まれる倍音の種類が聞き手に与える印象に影響を及ぼすという研究 [4] も報告されている。このことから、本研究では音色に関連する特徴量を中心とし、感情との関連を調査する。

## 3 データと可視化手法

### 3.1 使用したデータ

本研究ではデータセットとして 1960 年～2019 年の Billboard Year-End Hot 100 singles にランクインした英詞の楽曲を 59 曲選び、各曲の最初から最後までボーカルを含む部分を 60 秒抜粋したのち demucs[5] で音源分離し、ボーカルと伴奏に分けた。原曲とその伴奏とボーカルの 157 曲を対象として、20 代の非 20 名に、GEMS[6] と Valence-Arousal (以下 V-A と称する) にもとづいて、想起される感情について 5 段階のリッカード尺度で主観評価を依頼した。V-A に関する評価の項目は表 1 の通りである。

表 1: 評価項目

The image shows a digital evaluation interface with two sections. The first section is titled '1960\_Valence\*' and contains a horizontal row of five empty circles representing a Likert scale. Below the circles are two descriptive phrases: '明るい気持ちになる、楽しい' (feeling bright and happy) on the left and '暗い気持ちになる、悲しい' (feeling dark and sad) on the right. The second section is titled '1960\_Arousal\*' and also contains a horizontal row of five empty circles. Below the circles are two descriptive phrases: '激しい、目が覚めるような' (intense, like waking up) on the left and '穏やか、眠くなるような' (calm, like falling asleep) on the right.

### 3.2 音響特徴量

本研究では、楽曲の音響特徴量として下記の特徴量の平均、分散、歪度、尖度に着目し、Librosa[7] を用いて算出した。入力データの形式は wav、サンプリングレートは 48kHz である。

#### **Boldface SpectralFlatness(SF)**

パワースペクトルの平坦度を表す。

#### **Boldface SpectralCentroid(SC)**

パワースペクトルの重心を表す。音響信号の明るさと関係している。

#### **Boldface SpectralBandwidth(SB)**

パワースペクトルの広がりを表す。

#### **Boldface RMS(RMS)**

音響信号の物理的な強度を表す。

#### **Boldface ZeroCrossingRate(ZCR)**

振幅値が 0 と交差する回数を表し、ノイズ量を示す。

### 3.3 可視化手法

本研究では、伴奏やボーカルが楽曲の印象に与える影響を確認するため、それぞれの Valence 及び Arousal の平均の相関関係をスピアマンの順位相関係数を用いて調べた。また、得られたデータを Valence と Arousal の平均および分散に着目し、原曲、ボーカル、伴奏のそれぞれについて以下の 2 種類の 방법으로可視化した。

#### Hidden を用いた平行座標プロットでの可視化

高次元データ可視化手法 Hidden[8] を用い、Valence、Arousal と関係のある特徴量をおおまかに観察する。

#### 主成分分析での可視化

Hidden で Valence、Arousal と関連のあった特徴量に対して PCA による次元削減を適用し、寄与率の高い特徴量を特定する。

## 4 実行結果・考察

本章では実行結果から得られる知見について述べる。

### 4.1 分析結果 1

表 2: 両側検定を行なった際の p 値

	p 値 (Valence)	p 値 (Arousal)
原曲 x ボーカル	9.6e-17	8.0e-24
原曲 x 伴奏	2.7e-24	7.3e-25
伴奏 x ボーカル	1.1e-10	2.8e-15

3.3 節で述べた手順に沿って、Python の Matplotlib と seaborn で可視化した結果を述べる。図 1 から、原曲、伴奏、ボーカルの全てについて Valence と Arousal の平均値には正の相関関係があることが読み取れる。ま

[H]

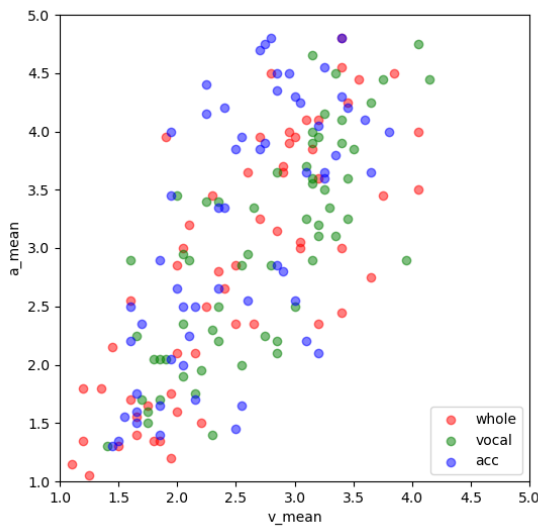


図 1: Valence と Arousal の平均を表す散布図

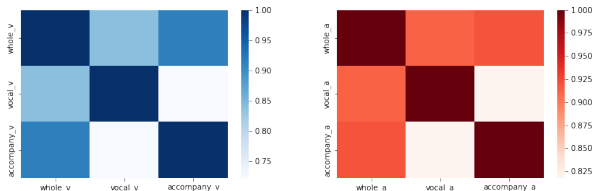


図 2: Valence の平均値から計算したスピアマン相関係数

図 3: Arousal の平均値から計算したスピアマン相関係数

た図 2, 図 3 から, Valence と Arousal どちらの場合においても原曲の印象と伴奏及びボーカルの印象は強く相関しているが, 伴奏と原曲の相関の方がボーカルと原曲の相関よりも強いことがわかる。

## 4.2 分析結果 2

### 4.2.1 原曲の可視化

Hidden による可視化の結果, Arousal の平均は SpectralFlatness の平均と負の相関を, SpectralBandwidth の尖度, SpectralBandwidth の歪度, SC の尖度と正の相関を持つとわかった。また PCA による可視化の結果から, SpectralFlatness の平均が高い楽曲は Arousal の平均が低いということがわかった。

### 4.2.2 ボーカルの可視化

Hidden による可視化の結果, Arousal の平均は SpectralFlatness の平均と負の相関を持つとわかった。また ZeroCrossingRate の平均が高く, 分散が低い楽曲ほど Arousal の平均が大きくなるとわかった。このことから, ノイズ量が一貫して大きく, 非整数次倍音を持つ歌声の方が Arousal の平均が高くなると考えられる。

### 4.2.3 伴奏の可視化

Hidden による可視化の結果, Arousal の平均は SpectralFlatness の平均と負の相関を, SpectralBandwidth

の尖度, SpectralCentroid の尖度と正の相関を持つとわかった。また PCA による可視化の結果, SpectralFlatness の平均と SpectralBandwidth の尖度が大きく, SpectralCentroid の歪度が小さい楽曲ほど Arousal の平均が小さくなるとわかった。

## 5 まとめと今後の展望

本研究では, ポピュラー音楽をボーカルと伴奏に音源分離し, それぞれについて想起された感情について評価実験を実施した。その結果, ボーカルと伴奏が楽曲の印象に強い影響を持つことや, 楽曲の形式に関わらず Arousal の平均と SpectralFlatness の平均が負の相関を持つことがわかった。今後の課題として, Arousal だけでなく Valence に関連する特徴量を発見することや, 同じ楽曲を別の歌手が歌った作品に対して印象調査を実施することがあげられる。また, 将来的にはボーカルのみを対象にして, 音楽が人間の感情に与える影響についてさらに詳しく調査したい。

## 謝辞

評価実験に協力をいただいた皆様に感謝いたします。

## 参考文献

- [1] Russell, J. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*. 39. 1161-1178. 10.1037/h0077714.
- [2] Hunter, P. G., Schellenberg, E. G., Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music : Similarities, differences, and mixed emotions. *Psychology of Aesthetics Creativity and the Arts*, 4, 47-56.
- [3] Gabrielsson, A., Lindström, E. (2010). The role of structure in the musical expression. In P. N. Juslin and J. A. Sloboda (Eds.), *Handbook of music and emotion : Theory, research, applications* (pp. 367- 400). Oxford, UK : Oxford University Press.
- [4] 菅原衣織, 伊藤貴之, ”倍音分析によるいい声作りの支援アプリ開発に向けて”, エンタテインメントコンピューティングシンポジウム 2013 論文集, 51-55, 2013
- [5] A. Défossez, N. Usunier, L. Bottou and F. Bach, ”Demucs: Deep extractor for music sources with extra unlabeled data remixed”.
- [6] Zentner, M., Grandjean, D., Scherer, K. (2008). Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement. *Emotion* (Washington, D.C.). 8. 494-521. 10.1037/1528-3542.8.4.494.
- [7] McFee, Brian, Colin Raffel, Dawen Liang, Daniel P. W. Ellis, Matt McVicar, Eric Battenberg and Oriol Nieto. ”librosa: Audio and Music Signal Analysis in Python.” *SciPy* (2015).
- [8] 高次元データ可視化手法 Hidden 伊藤貴之, 2017 年 09 月, 日本ソフトウェア科学会第 34 回大会,