

メタデータを利用した機械学習による DTM(DeskTop Music)での音色づくりの効率化

齋藤 創^{1,a)} 大場 みち子^{2,b)}

概要: 本論文では、DTM(DeskTop Music)での作曲をより容易にする手法と構築したシステムに関して述べる。DTMでの作曲プロセスの一つとして、音色づくりがあり、DTM作曲者は音色づくりをするために音色づくり情報の検索を行っている。一方、各DTM作曲者が思い浮かべる音色はそれぞれ異なるため、音色の雰囲気を用いて求める音色づくり情報を発見する事は困難である。本研究の目的は、DTMでの作曲を効率よくできるようにすることである。目的達成のため、DTM作曲者に対して音色づくり情報を適切に提示するシステムを構築する事を目標とする。本研究では、音色づくり情報のメタデータから音響的特徴を取得し、DTM作曲者ごとに機械学習をする事で、DTM作曲者ごとの検索フィルタを作成する。検索フィルタを導入した音色づくり支援システム(TASS: Timbre Adjustment Support System)をリリースし、ユーザのアクセスログを収集する。収集したアクセスログを解析することにより、機械学習アルゴリズムの妥当性を検証する。

キーワード: DTM(DeskTop Music), メタデータ, 音響特徴量, 機械学習

Efficiency of Timbre Adjustment by Machine Learning Using Metadata

Abstract: In this paper, we propose a method of creating DTM music more easily. It is difficult for DTM composer to search the desired timbre information using atmosphere data, because users' imagines are not equal to each other. Our approach is to apply machine learning based on acoustic features that will be obtained from metadata and create Searching Filter. We develop a Timbre Adjustment Support System (TASS) based on our approach and release it. Then we collect TASS each user's access log. We evaluate the proposed method by TASS.

Keywords: DTM (DeskTopMusic), Metadata, Audio Features, Machine Learning

1. はじめに

近年、PCを用いて作曲を行うDTM(DeskTop Music)が発展している。これはYouTube、ニコニコ動画等の消費者生成系メディア(CGM: Consumer Generated Media)の急速な発展が要因として考えられる[1], [2]。DTMでの作曲におけるプロセスの一つとして、音色づくりと呼ばれる

作業がある。音色づくりとは、DTM作曲者が作成する楽曲に適した音色を作成するため、ソフトウェア音源のパラメータを調整する作業のことである。ソフトウェア音源のパラメータは複雑であり、任意の音色を作成するために、DTM作曲者は多くのパラメータを操作する必要がある。DTM作曲者に対するヒアリングにより、DTMでの作曲において音色づくりに時間と労力がかかっている事が判明した[3]。以上より、音色づくりを効率化することで、DTMでの作曲効率が良くなると考えられる。

本研究の目的は、DTMでの作曲を効率よくできるようにすることである。目的を達成するため、DTMでの音色づくりをする際に利用する情報(以下、音色づくり情報)を適切に提示するシステムを構築することを目標とする。

¹ 公立はこだて未来大学 大学院
Graduate School of Future University Hakodate,
Kamedanakanocho, Hokkaido 041-8655, Japan

² 公立はこだて未来大学
Future University Hakodate, Kamedanakanocho, Hokkaido
041-8655, Japan

a) g2115015@fun.ac.jp

b) michiko@fun.ac.jp

本論文の構成は次のとおりである。第2章では、本研究に関連する先行研究に関して述べ、3章で課題と課題解決アプローチに関して述べる。第4章では、解決アプローチを適用した提案システムに関して述べる。第5章では、課題解決アプローチの手法を検証するための予備実験に関して述べる。第6章では、本論文のまとめと今後の展望に関して述べる。

2. 関連研究

本章では、本研究の関連研究と筆者らの先行研究を紹介し、先行研究における問題点を述べる。

2.1 関連研究

作曲者が任意の音色を検索し、効率的な再利用を可能にするため吉岡らが開発した電子音色辞書がある [4]。このシステムは、作曲者が作曲時に用いる環境音を記録し、この研究で独自定義している音色記号を用いた検索や、音色を3次元表示した検索などさまざまな手法で検索することが可能である。

一方、電子音色辞書では参照し利用できるデータが環境音のオーディオファイルに限られているため、発見した音源にアレンジを加えたり、音色の再現が困難であるという問題点がある。

2.2 先行研究

筆者らの先行研究として、DTM作曲者が効率的な音色づくりを可能にするため、図1のような音色づくり情報を記録・共有することができる音色づくり支援システム(TASS: Timbre Adjustment Support System)[3]を構築した。効率的な音色づくり情報の検索を可能にするため、音色づくり情報に対して以下のメタデータを付与して記録した。

- ソフトウェア音源
DTM作曲者が用いたソフトウェア音源を記録する。
- ジャンル
作成した音色に適するジャンルを記録する。
- 雰囲気
作成した音色に適する雰囲気を記録する。

先行研究では、DTM作曲者に対して過去に作成された音色を再現する実験を行った。実験は、TASSを利用した場合と利用しなかった場合の対照実験とした。実験の結果、TASSを用いることで音色再現の満足度が高くなることが判明した。

一方、一つの音色に対してDTM作曲者ごとに異なるイメージを有するため、雰囲気のメタデータを用いてDTM作曲者が求める音色づくり情報を発見することが困難であるという問題がある。



図1 TASSの音色づくり情報一覧画面

Fig. 1 Image of Timbre adjustment information list on TASS

3. 課題と課題解決アプローチ

本章では本研究における課題とその解決アプローチを示す。

3.1 課題

音色には厳密な定義が存在せず、基準となる指標が存在しない [5]。よって、DTM作曲者が音色を言葉で表現する際、同じ音色でも表現に差異がある。DTM作曲者ごとに用いる表現方法が異なるため、TASSでの雰囲気を用いた検索ではDTM作曲者ごとに適切な音色づくり情報を提示できていないという課題がある。

3.2 課題解決アプローチ

3.1節で述べた課題に対して、以下のアプローチで課題の解決を図る。

(1) DTM作曲者ごとの検索フィルタの作成

登録されたオーディオファイルの音響特徴量と、DTM作曲者が用いた雰囲気を表すデータを用いた機械学習を行うことにより、DTM作曲者ごとに適用した検索フィルタを作成する。

(2) 検索フィルタを用いた音色づくり情報の提示

DTM作曲者ごとに検索フィルタを用いることでDTM作曲者が求める音色づくり情報を適切に提示する。

図2は上記の課題解決アプローチを図示したものである。

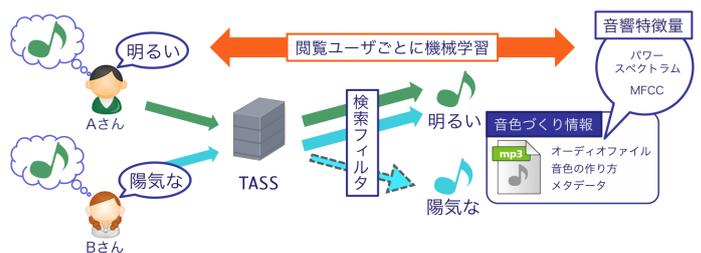


図2 課題解決アプローチのイメージ

Fig. 2 An image of problem solving approach

DTM 作曲者が音色の雰囲気表現する基準として、HEVNER の環状形容詞 [11] を用いる。音響特徴量と HEVNER の環状形容詞および機械学習に関する詳細を次項に記す。

3.2.1 音響特徴量

本研究では、音色を定量的に扱うためにオーディオファイルの音響特徴量を取得する。取得する音響特徴量はパワースペクトラムと MFCC である。各特徴量の詳細を次に示す。

- パワースペクトラム

パワースペクトラムとは、音の大きさに関する特徴量であり、波形振幅の 2 乗の時間平均の値のことであり [6]。音はアナログ的に変化する情報であり、音響特徴量の場合も取得するタイミングが重要となる。本研究では次に説明する MFCC を取得するタイミングを決定するために用いる。

- MFCC

MFCC とはメル周波数ケプストラム係数 (Mel Frequency Cepstral Coefficient) の略称であり、人の聴覚特性を考慮しながらスペクトルの外形を表現する特徴量である [6], [7], [8]。この特徴量は音声認識技術にも用いられており、本研究では 0 次から 12 次元までの係数を用いる。

3.2.2 HEVNER の環状形容詞

HEVNER の環状形容詞とは、図 3 のような KATE.HEVNER が提唱した音楽を評価する 8 つの形容詞クラスターである。この形容詞クラスターでは、C1:C5, C2:C6, C3:C7, C4:C8 がそれぞれ対として表される。図 3 は HEVNER の環状形容詞の各クラスターの 1 例を抽出したものである。

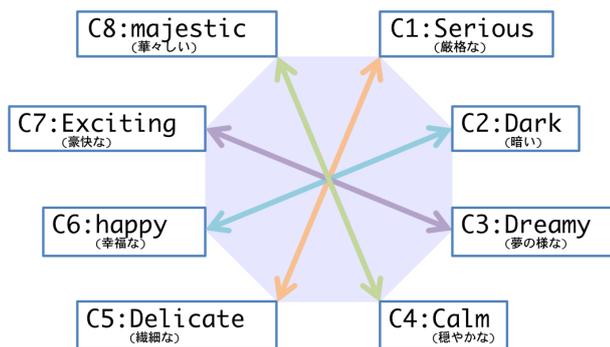


図 3 HEVNER の環状形容詞

Fig. 3 HEVNER's adjective cluster

3.2.3 機械学習

本研究では、DTM 作曲者が音色づくり情報を検索し閲覧した際、機械学習をすることで DTM 作曲者ごとの検索フィルタを作成する。次のデータを用いて、DTM 作曲者ごとの検索フィルタを作成する。

(1) 既存の検索フィルタ

* 初期値は雰囲気の親データと DTM 作曲者に登録時に入力させた値から算出する。

(2) 閲覧した音色づくり情報のオーディオファイルの音響特徴量

(3) 検索時に利用した雰囲気の親データの音響特徴量

(4) 検索時に入力した雰囲気の値

4. 提案システム

本研究では、研究の目的を達成するため先行研究で開発した TASS を改良し、3.2 節で述べた課題解決アプローチを適用した TASS V2 を開発する。本章では音響特徴量の取得手法および機械学習アルゴリズムに関して述べる。

4.1 音響特徴量の取得

3.2.1 項のパワースペクトラムの項目で記述した通り、音響特徴量は取得するタイミングによって値が大きく変わるため、動的に取得する必要がある。本研究ではパワースペクトラムを用いて、図 4 のようにオーディオファイルの音の立ち上がり音と中間音での MFCC を取得する。



図 4 音響特徴量取得タイミングイメージ

Fig. 4 Acoustic features of the acquisition timing

本研究では、これらの特徴量を Web Audio API[9] とその拡張ライブラリである Meyda[10] を用いて取得する。Web Audio API を用いることで、音響特徴量を動的に取得することが可能となる。JavaScript を用いて Web Audio API を操作するため、サーバに負荷をかけずに音響特徴量を取得することが可能となる。また、Web Audio API の拡張ライブラリである Meyda を利用することで、取得したい特徴量を指定するのみで取得し、本来複雑な計算処理をして取得する MFCC 等の音響特徴量を素早く取得することが可能になる。

4.2 機械学習アルゴリズム

本システムでの学習データは HEVNER の環状形容詞のクラスタに準拠する。HEVNER の環状形容詞を検索の基準とし、YouTube Audio Library[12] から HEVNER の各クラスタに合致するオーディオファイルの音響特徴量を機械学習の親データとする。機械学習のプロセスを次に示す。

- ユーザ登録
 - (1) DTM 作曲者： ユーザ登録画面で HEVNER の環状形容詞のクラスタごとのサンプル音を聞き、入力
 - (2) TASS： 入力データと親データを用いて、ユーザの登録時に検索フィルタの初期値を登録
- 音色づくり情報検索
 - (1) DTM 作曲者： 検索項目を入力し検索
 - (2) TASS： 検索に合致する順に結果を提示
 - (3) DTM 作曲者： 検索結果から登録されているオーディオファイルを聞き、求める音色づくり情報にアクセス
 - (4) TASS： アクセスされた音色づくり情報と DTM 作曲者の既存の検索フィルタを用いて機械学習し、検索フィルタを更新

本研究では機械学習アルゴリズムとして式 (1) を用いる。

$$x \in (1, 2, \dots, 8), y \in (0, 1, 2, \dots, 12), W \in C_x$$

d_x ： 検索フィルタ m_y ： 閲覧情報の MFCC

P_x ： ユーザの既存の検索フィルタ (初期値は $\overline{C_x}$ と登録値の差)

$\overline{C_x}$ ： 0 点から各クラスタの親データの距離

$\overline{m_x}$ ： 0 点から m_x の距離 W ： 検索語句

$$d_w = P_w + \frac{\{\overline{m}_y - P_w\}}{1 - 1/\{C_x - \overline{m}_y\}} \quad (1)$$

機械学習アルゴリズム (1) では、ユーザが学習した値と閲覧した音色づくり情報の差が大きいほど大きく値が変動し、差が小さいほど値の変動が小さくなる。本アルゴリズムを用いることで、機械学習の誤差の影響を受けづらいついという利点がある。

5. 予備実験

4.2 節で述べた機械学習アルゴリズム (1) が適用できるかどうかを検証し、実際にユーザが入力する雰囲気の結果のデータ例を取得する目的で予備実験を行なった。

5.1 手順

3 名の被験者 A, B, C に対して次の手順で予備実験を行った。

事前準備

機械学習アルゴリズム (1) を反映させた Microsoft Excel を作成した。

手順 1 20 種類の音色を提示

被験者に昨年度システムに登録されている 20 種類の

音色を提示した。提示する音色は何度聞いて良いものとした。

手順 2 雰囲気値の入力

被験者には雰囲気値を HEVNER の形容詞クラスタに合致するように、図 5 のような入力フォームを用いて入力するように指示した。図 5 の入力フォームでは、入力される値は 0 から 100 までであり、中心部ほど雰囲気値が小さい。値が 0 の場合は該当しないものとし、値が大きいほど該当のクラスタに合致するものとした。

手順 3 入力データの分析

被験者の入力したデータに関して、3 名の被験者ごとにどのような違いがあるのかを確認した。また、事前準備で用意した Microsoft Excel と手順 2 で得た被験者ごとの雰囲気値の値の遷移を確認した。

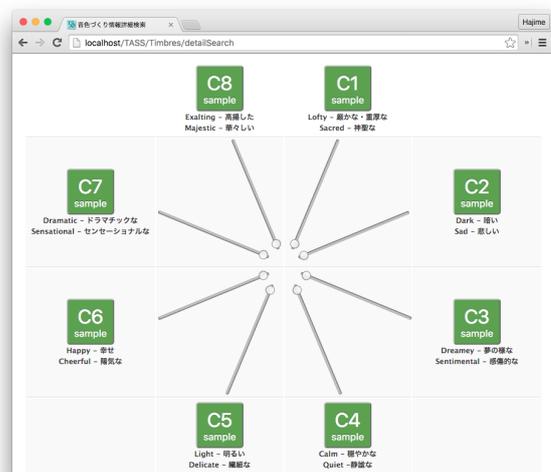


図 5 HEVNER の形容詞クラスタに対応した入力フォーム
 Fig. 5 Input form corresponding to adjective HEVNER's cluster

5.2 結果

予備実験から実際にユーザが入力する雰囲気値のデータのサンプルを取得することができた。被験者 A, B, C それぞれが音色を聞いて入力した雰囲気値の結果の一例を表 1 および図 6 に示す。表 1 は、行が各被験者を示し、列が HEVNER の環状形容詞の各クラスタを示している。表 1 の数値は 0 から 100 までの自然数であり、数値が大きいほど該当のクラスタに合致していることを示し、全く当てはまらない場合は 0 ではなく『-』として表記した。

図 6 は、中心を 0, 外周を 100 として表 1 の結果をレーダーチャートにプロットしたものである。図 6 では、クラスタに該当しない場合はプロットせず、該当するデータのみプロットした。

図 7 は、Microsoft Excel を用いた機械学習アルゴリズム

表 1 雰囲気の入力データ例

Table 1 An example of input data of atmosphere

		HEVNER の環状形容詞クラスタ							
		C1	C2	C3	C4	C5	C6	C7	C8
被験者	A	-	-	-	-	95	70	-	10
	B	-	-	-	-	85	80	35	43
	C	25	-	-	-	-	-	70	15

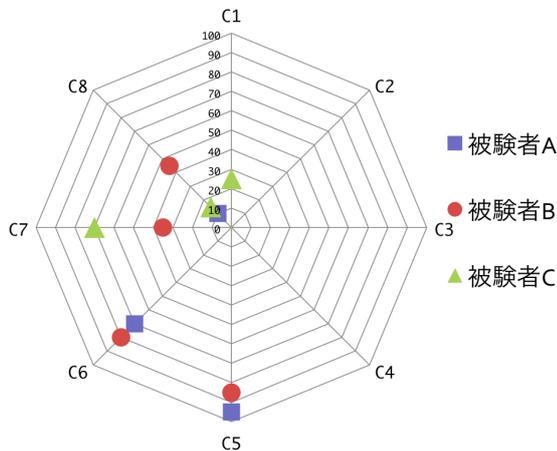


図 6 予備実験結果レーダーチャート

Fig. 6 Radar chart of preliminary experimental results

の学習値の推移を表した図である。図 7 は被験者 A, B, C のそれぞれの学習値の変移を示している。グラフの縦軸は学習値を示し、横軸は学習回数を示している。図 7 より、被験者 A では 18・19 回目の C7、被験者 B では 16・19・20 回目の C5、被験者 C では 8・9 回目の C7 の学習値が大きく変動していることを確認した。

5.3 考察

5.2 節より、次の 3 点のことが判明した。

- ある音色に対する雰囲気の評価はユーザごとに異なる
同一の音色を評価させた際、被験者ごとに該当する
と考えている形容詞クラスタに違いがあり、その値も
ユーザによって異なっていた。よって、音色に対する
雰囲気の評価はユーザごとに異なっていることを確認
することができた。
- 機械学習アルゴリズム (1) によりユーザごとの学習値
の差異を確認できる
図 7 より、ユーザごとに学習値の変移に違いがある
ことが分かる。これは機械学習アルゴリズムを用いる
ことで、ユーザごとに異なる雰囲気の評価を取得でき
ていると判断できる。
- 現状の機械学習アルゴリズム (1) をそのまま用いてし
まうと、適切な学習を行うことができない
機械学習アルゴリズム (1) はユーザごとに学習値が
収束することが求められるが、5.2 節より機械学習アル

ゴリズムが大きく変動している事を確認した。この結果から、現状の機械学習アルゴリズムをそのまま用いると、ユーザの学習値が大きく変化してしまい、ユーザごとの正確な学習値が取得できない可能性がある。

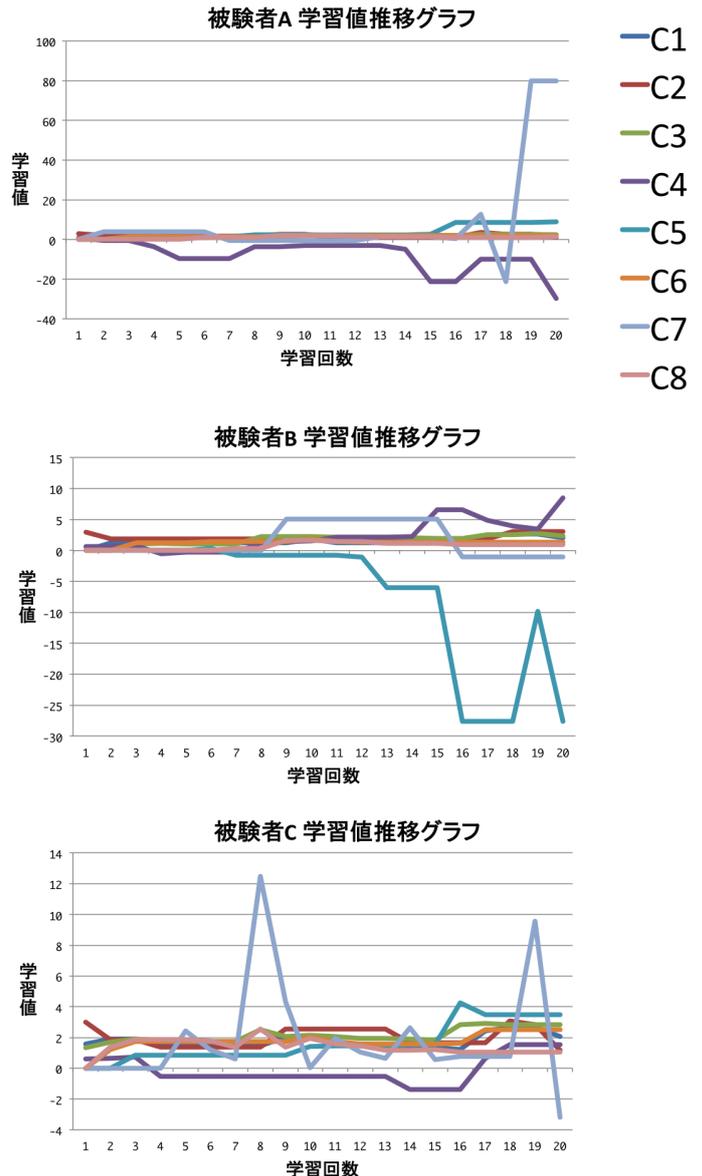


図 7 機械学習学習値推移グラフ

Fig. 7 Transition graph of the machine learning value

ゴリズムが大きく変動している事を確認した。この結果から、現状の機械学習アルゴリズムをそのまま用いると、ユーザの学習値が大きく変化してしまい、ユーザごとの正確な学習値が取得できない可能性がある。

機械学習アルゴリズムが学習値に大きく影響を与えた原因として、機械学習アルゴリズム (1) における分母の $1 - 1/\{\overline{C_x} - \overline{m_y}\}$ が 0 に近づきすぎたため、大きく変動した事が要因の一つだと考えている。

6. おわりに

本章ではまとめと、予備実験をふまえた今後の展望に関して記述する。

6.1 まとめ

DTMでの作曲を効率化するため、本研究ではDTMにおける音色づくりという作業に着目し、音色づくり情報をDTM作曲者ごとに適切に提示するシステムを構築することを目標とした。本研究の課題は、DTM作曲者がイメージする音色と表現する語句が作曲者ごとに異なるため、作曲者ごとに適切な音色づくり情報を提示することができていない事である。課題の解決アプローチとして、DTM作曲者が検索に用いた雰囲気と実際に閲覧した音色づくり情報の音響特徴量を用いた機械学習を行い、DTM作曲者ごとの検索フィルタを作成する。本研究では課題解決アプローチのアルゴリズムを検証する目的で予備実験を行なった。予備実験の結果、現状の機械学習アルゴリズムでは、適切な学習値が得られないことが判明した。

6.1.1 今後の展望

今後は、予備実験で得た雰囲気の詳細データを用いて、ユーザーごとに学習値が収束するように機械学習アルゴリズムの修正を検討し、TASS V2へ適用する。

その後、TASS V2をリリースし、Google Analytics[13]を用いて、DTM作曲者の検索時間やアクセスするページ等のアクセスログを取得・解析することでTASS V2の有効性を確認する。

参考文献

- [1] 齋藤明, “サービス・ドミナント・ロジックにおける価値共創概念と市場創造: 「初音ミク」という市場創造を中心として,” 情報文化学会誌, vol. 21, no. 1, pp. 2936, Aug. 2014.
- [2] 辰巳直也, 馬場隆, 森勢将雅, 片寄晴弘, “ロックボーカルレゾネータ Vocaloid 歌唱をロックボーカリスト風の歌い方に変換するシステム,” 情報処理学会研究報告. SLP, 音声言語情報処理, vol. 2012, no. 12, pp. 16, Jan. 2012.
- [3] 齋藤創, 大場みち子, “メタデータを活用したDTM(DeskTop Music)での音づくり支援システムの構築,” presented at the 情報処理学会第77回全国大会, 2015.
- [4] 吉岡靖博, 石塚慎也, 齋藤佳紀, 高橋奈緒子, 小坂直敏, “電子音色辞書における音色探索システムおよび3次元音色表示インタフェース(インタフェース),” 情報処理学会研究報告. 音楽情報科学, vol. 2009, no. 13, pp. 5358, Feb. 2009.
- [5] Jason Leung, “Timbre Recognition,” *MUMT*, vol. 621, no. Assignment 4, pp. 0-5, Mar. 2014.
- [6] 千葉祐弥, “Q: さまざまな音響特徴量それぞれの使い方や意味を教えてください,” 音響学会ペディア. (Online). 入手先 (http://abcpedia.acoustics.jp/acoustic_feature_2.pdf) (30-Dec-2015).
- [7] aidiary, “メル周波数ケプストラム係数(MFCC) - 人工知能に関する断創録,” <http://aidiary.hatenablog.com/>. (Online). 入手先 (<http://aidiary.hatenablog.com/entry/20120225/1330179868>) (07-Jul-2015).
- [8] 宮澤幸希, “メル周波数ケプストラム(MFCC) - Miyazawa's Pukiwiki 公開版,” Miyazawa's Pukiwiki 公開版, 29-Mar-2013. (Online). 入手先 (http://shower.human.waseda.ac.jp/m-kouki/pukiwiki_public/66.html) (07-Jul-2015).
- [9] “Web Audio API (日本語訳).” (Online). 入手先 (<http://g200kg.github.io/web-audio-api-ja/>) (30-Dec-2015).
- [10] Hugh Rawlinson, Nevo Segal, and Jakub Fiala, “Meyda: an Audio Feature Extraction Library for the Web Audio API,” in *WAC 2015*, Paris, 2015.
- [11] K. HEVNER, “experimental studies of the elements of expression in music,” *American Journal of Psychology*, vol. 48, pp. 246268, 1936.
- [12] “オーディオライブラリ - YouTube.” (Online). 入手先 (<https://www.youtube.com/audiolibrary/music>) (29-Dec-2015).
- [13] “Google アナリティクス公式サイト - ウェブ解析とレポート機能 Google アナリティクス.” (Online). 入手先 (<https://www.google.com/analytics/>). (29-Jan-2016).