DEIM Forum 2017 P3-6

ユーザの嗜好や個人的コンテンツを利用したPersonalized Visual Jockey

成田　敦英†　　荻野　晃大†　　中島　伸介†

† 京都産業大学 コンピュータ理工学部　〒 603-8555 京都府京都市北区上賀茂本山
E-mail: †{g1445548,ogino,nakajima}@cc.kyoto-su.ac.jp

あらまし　イベント会場やクラブ等においてミュージシャンや DJ の出力する音楽に合わせて，その音楽にあった映像を出力するシステムとして，Visual Jockey（以下，VJ）というシステムがある．基本的にその場を盛り上げるという目的で実施されるものであり，音楽に対する即興的な映像演出の役割を担っている．ただし，従来の VJ は，基本的に家庭で個人的に楽しめるものではなく，リズムに合わせた波動や，それほど関連性が深いとは限らない映像が使用されることが多く，音楽とのマッチングはそれほど高いとはいえない．そこで我々は，ユーザの嗜好や個人的コンテンツを利用した Personalized Visual Jockey を提案する．

キーワード　Personalized Visual Jockey，メタデータ付与，映像情報推薦，コンテキスト依存

1. はじめに

Visual Jockey(以下，VJ と記載) とは，イベント会場やクラブ等においてミュージシャンや DJ の出力する音楽に合わせて，その音楽にあった映像を出力する人やシステムのことを指す．基本的にその場を盛り上げるという目的で実施されるものであり，音楽に対する即興的な映像演出の役割を担っている（図 1 参照）．しかしながら，基本的にイベント会場やクラブ等において，それなりの技術や機材を有するパフォーマーにより披露されるものであり，家庭で個人的に楽しめるものではない．また，リズムに合わせた波動や，それほど関連性が深いとは限らない映像が使用されることが多く，音楽とのマッチングはそれほど高いとはいえない．

そこで我々は，ユーザの嗜好や個人的コンテンツを利用した Personalized Visual Jockey（以下，Personalized VJ と呼ぶ）を提案する．本手法の特徴は，あらかじめ作られた映像を流すのではなく，音楽から推定される場面や印象に合致する，個人が有する思い出の写真やムービーの利用を可能とするものである．また，一緒に居るメンバー等のコンテキスト情報も考慮することが可能であり，同じ曲でもホームパーティ時，デート時，1 人でくつろいでいる時のそれぞれで異なる映像を提示することができる．写真を利用する場合には，フェードアウトやズームインなどのエフェクトを利用することで映像化することを考えているが，この際，音楽の印象やユーザのコンテキストに合致するエフェクトを選定することを可能にする．

提案手法である Personalized VJ を実現するためには，音楽データおよび写真・ビデオデータへのメタデータ付与が必要となる．音楽データへ付与されるメタデータとしては,（テンポ等）曲調の分析や歌詞の分析により，曲全体の場面や印象とフレーズ毎の場面や印象を表現可能なものを考えている．写真・ムービーへのメタデータ付与に関しては，少なくとも本稿では検討しない．各種方法によりメタデータが付与されるものと考える．

本稿の校正は以下の通りである．２章にて関連研究について説明する．３章にて提案手法に関する詳細を述べ，４章にてまとめと今後の課題を述べる．

図 1　従来の Visual Jockey のイメージ

2. 関連研究

浅田ら [1] は，VJ を 1 つのパフォーマンスとして捉え音楽に合う画像，動画を出力する AR を用いてパフォーマンスとプロジェクションマッピングを組み合わせよりインタラクティブな関係を持たせるシステム，投影スクリーンを平面ではなく立体スクリーンを使用しダンサーがパフォーマンスしている際，Kinect という赤外線センサーを用いて動きのスピードを検知し，まるで踊り手から放たれるような一体感の演出という研究を行っている．

寺田ら [2] は，リアルタイムコンテンツである VJ を最大限利用するためにメルチメディアデータからルールに基づいて自律的に動作するシステム，この場合の Active Karaoke の作成歌詞の内容や曲調，マイクの入出力に合わせて状況に合った素材をリアルタイムに選択に観点を置いている．検索結果の妥当性に関しては画像情報に付与するメタデータが大切であり逆に意外性のある画像を表示したほうが当てはまるというパターンも有るという考えを持っている．ストーリー性に関しても，撮

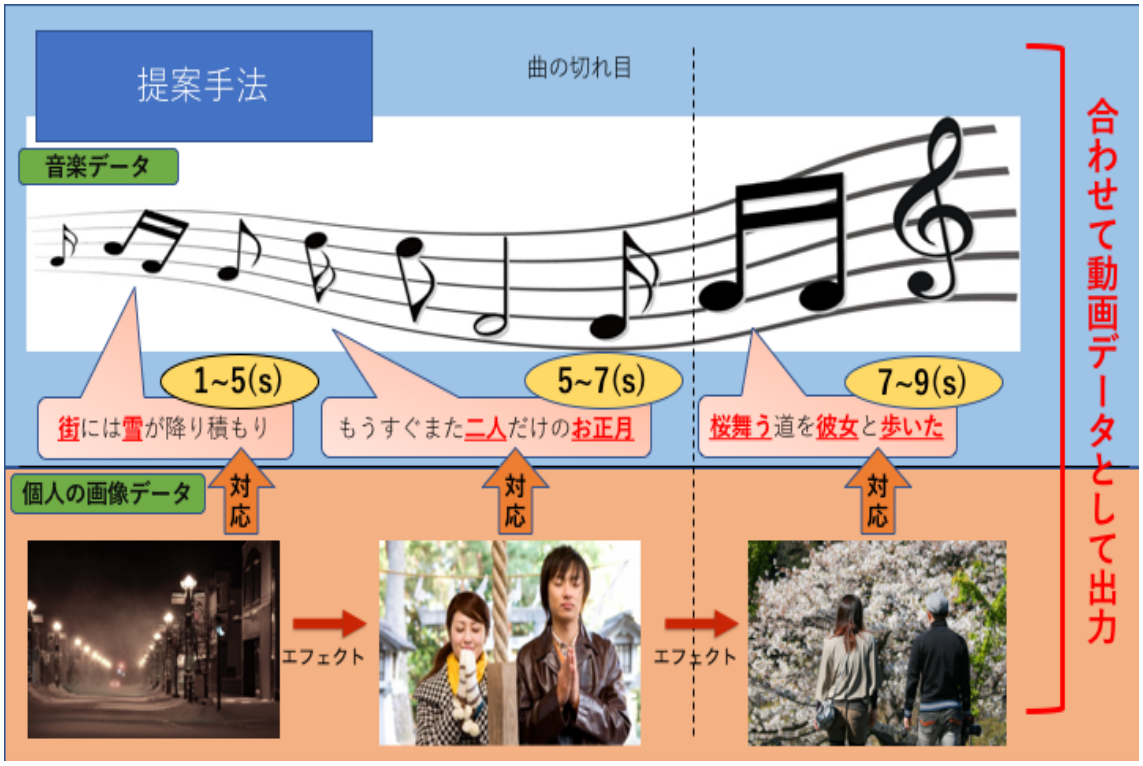

図 2　音楽データに対する個人の画像データの対応付け方式

影する場所時間情報があれば良いので容易である．

中野ら [3] は，既存の音楽動画の映像部分を切り取り新たに別の音楽に合うよう自動生成かつインタラクティブに編集できるシステムの実現，VJ を利用することで元のコンテンツを再利用し新たなコンテンツを生む事に観点を置いている．代表的な例で言うと，ニコニコ動画の MAD もこれに該当，これをマッシュアップという．今回論文では踊ってみた動画について述べていたが，それ以外の MAD 動画でも同じことが言えると考えている．

草間ら [4] は，曲名やアーティスト名などのメタデータに依存せず，旋律や調性などの楽曲特徴や印象に基づいて楽曲を検索するツールを提案している．楽曲の印象を瞬時に直感的に認識するために，各楽曲データから特徴を検出し楽曲をユーザのニーズに合わせて階層的にクラスタリングその後，抽出した特徴データを元に印象画像を自動作成し，印象画像を用いて楽曲を一覧表示するという手法．

木村ら [5] は，歌詞カードのような映像操作インターフェースを持つ新奇な VJ システムを提案している．クラブミュージックのように単純な構造の音楽ではなく，JPOP のような複雑な構造を持つ音楽に対応することを前提とし，従来の VJ システムのような各瞬間にどのような映像を表示するか，という機能ではなく楽曲構造を意識し，使用するユーザの音楽情報知識の有無にかかわらず使用できユーザの負担を軽減することが可能にしている．

以上のように，大衆に向けての映像を利用した VJ と音楽とのリンクシステムは既に存在する．だが，それらはユーザ個人の趣向にそれぞれ対応したものではない．その為，今回我々の研究では，VJ システムの個人化という点に着目し，音楽に合わせ最適な画像をユーザの特性に合わせて表示する，というアプローチを行うものとする．

3. 提 案 手 法

図 2 では，音階を曲データ，吹き出しを歌詞として表しており，赤字の部分をキーワードとし，それに該当する画像の表示している．歌詞を小節ごとに区切った後，画像とのマッチングを行うが，その際，曲のテンポ，音の高低などの音に関する情報のメタデータも考慮する必要があるので，小節の区切り以外にもテンポや音の高低が変化した場合も，再度音楽と画像のメタデータのマッチングを行う．

図 3 では，具体的な分割の手順を記載している．小節に分割された音楽データには，最終的にキーワード情報のメタデータが 6 個，その小節がどの区間にあるのかという時間情報，その区間の音の高低やテンポなどの音に関する情報がメタデータとして添付される．

3.1 概　　要

本研究では従来の VJ のように，ただ単に音楽情報をリズムの塊として捉え，そのリズムにマッチした関連性のない動画，画像をリズミカルに表示するのではなく，音楽，例えば歌の歌詞情報を単語に分解し，各単語をキーワード情報のメタデータとして捉え，画像側にも事前に付与されている場面などのキーワード情報のメタデータとのマッチングを行う．

それにより，ただ単にリズムに乗せて無関連な映像を表示するよりも，よりその音楽にあった画像を音楽を視聴しているユーザが閲覧することが出来る．図 2 を用いて実際に説明する．

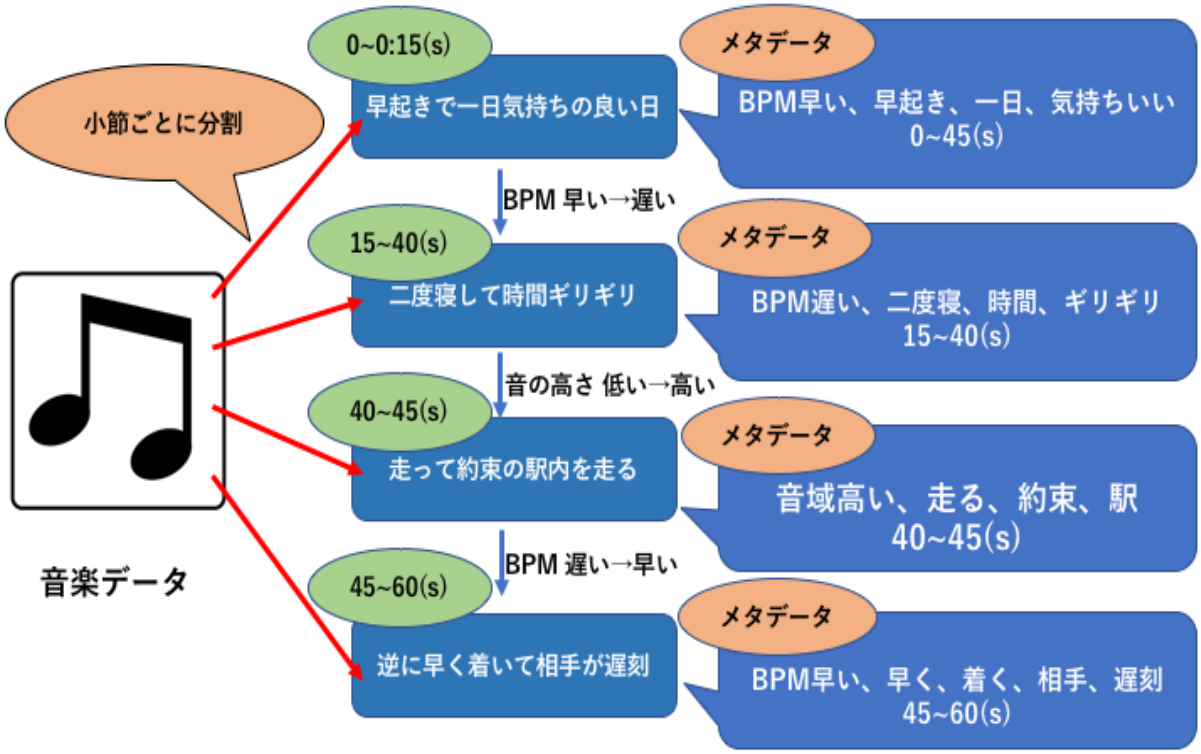

図 3　音楽データへのメタデータ付与方式

以上の図では，音楽データを音階，吹き出しを歌詞データとして表している．

以下の章で補足して説明するが，まず第一段階として歌詞データとそれが何秒から何秒の間に出てくるワードなのかというメタデータが記載されている音楽データが必要である．図 2 の「街には雪が降り積もり，もうすぐ二人だけのお正月，桜舞う道を彼女と歩いた」の部分が添付されている歌詞データであり，歌詞データの上部に記載されている「1〜5(s)」「5〜7(s)」「7〜9(s)」がそのワードが出現する秒数のデータである．

次に，添付した歌詞データを「街には雪が降り積もり」「もうすぐ二人だけのお正月」「桜舞う道を彼女と歩いた」のように小節に分割する．実際のシステムでは，分割する際のキーワードを 6 個に設定しているが，図 2 では説明のため，少なく設定している．そして，その小節内の名詞，動詞などのキーワードを抽出しメタデータとして記載する．今回のキーワードは赤字で記載している「街，雪」「二人，お正月」「桜舞う，彼女，歩いた」である．

画像側では，音楽に添付されたキーワードのメタデータに対し，画像に添付されているメタデータがより多く該当する画像をそのキーワードが割り当てられているその時間内に，エフェクトを掛けて表示する．無加工で画像を表示する事も可能だが，無加工では明暗などからマッチング率が低下してしまう可能性があるため，表示する段階で例えば「失恋」というキーワードの場合はモノクロ加工,「子供の頃」というキーワードの場合はオレンジ調を強調するなど加工を行いたいため，エフェクトにも画像データの格納はデータベース内に行われているが，それだけでは全ての音楽情報に対応しきれないため不十分である．

その為，不足分を補う為にキーワードに該当する画像を探す際，該当する画像がないときは Google 画像検索を用いて外部から画像データを取得をする．データベース内に初期段階から収納される画像に関しては，事前にシステム設計段階で設計者がキーワード情報のメタデータを添付しておく必要があるが，外部から取得した画像データに関しては，検索時に使用したキーワードをそのままメタデータとして添付する．

また，元々システム内に用意されている画像，外部から取得された画像以外にもユーザが直接データベースに自身の所持する画像を登録することも可能になっており，その場合は画像のタイトルをその画像のキーワード情報として認識し，それをメタデータとする．

ユーザが自身で登録した画像と元々システム内に存在する画像との間でメタデータが同じものが存在する場合，ユーザが自身で登録しているということで，元々の画像ではなくユーザが登録したものを優先的に表示するという仕様になっている．

3.2　音楽データへのメタデータ付与

前提として，その音楽情報に歌詞情報が付与されているとして話をすすめるが，まず記載されている歌詞情報のテキストデータを単語単位に分解する．

次にその分解した単語を小節ごとに区切り，何秒から何秒の間にその小節が存在するかというデータを音楽データに付与し，小節毎にその時間内でキーワード情報を切り替える．

図 3 で，説明しているが，曲ののキーワード情報だけでは，その曲に関しての印象を決めきれないので BPM の速い曲遅い曲，音が高音低音などの観点にも着目してメタデータ付与を行う．ただし，本研究では，曲中で印象が不規則にするのではな

く，規則的に変化するものを扱うものとする．

今回は「早起きで一日気持ちの良い日，二度寝して時間ギリギリ，走って約束の駅内を走る，逆に早く着いて相手が遅刻」というフレーズを例にして説明するが，概要で説明したが，まず最初に小節に分割し，その小節が何秒に存在するのかをデータとして割り当てる．「早起きで一日気持ちの良い日」が 0〜15(s),「二度寝して時間ギリギリ」が 15〜40(s),「走って約束の駅内を走る」が 40〜45(s),「逆に早く着いて相手が遅刻」が 45〜60(s) が今回割り当てるデータである．

次に分割した小節間で音の高さ，BPM の変化に着目し「BPM 早い」や「音域高い」などのように音の変化をメタデータとして添付する．分割した小節内の名詞，動詞に対しても「早起き，一日，気持ちのいい」のように単語のみを抽出し，同様にメタデータとして添付する．そして，音楽に添付されたメタデータと，画像の持つメタデータとのマッチングを行い，画像の表示をしながらバックグラウンドで音楽を再生する．各小節ごとに分割する際，区切る基準は秒数ではなく，その小節が持つ歌詞のキーワードの数であり画像データへのメタデータ付与の項目で補足するが，今回分割は，キーワードが 6 個の部分で行う．

理由は，画像データとのマッチング時，キーワード数が少なすぎると候補が膨大になり最適な画像が選出できないが，逆に多すぎても同じことが言えるので，検証の結果，本研究ではキーワード数を 6 個に設定している．

3.3 画像データへのメタデータ付与

画像データへは，主にその画像の色合い，シチュエーション，写っている人物などの情報からメタデータ付与を行う．概要でも述べたが，システム内の既存の画像に対しては設計者が手動でメタデータ付与を行わなければならない．

この時，付与するメタデータは出来る限り数が多い方が適している．なぜなら，添付されているメタデータが少ない場合，その画像と歌にあまり関連性がない場合でも他にそのキーワードに該当するものがなければ，その画像が選出されてしまうからである．最終的に表示する画像は歌詞のメタデータとマッチするメタデータをより多く持つものが選出される．例として，以下の図 4 のような画像の場合は,「海」「浜辺」「親友」「夕日」「綺麗」「青春」などのキーワードに該当するメタデータと,「高音」「BPM80〜90」という音に関するメタデータを付与するものとする．

次に，Google 検索から取得する外部からの画像の場合だが，これに対しては歌詞の区切り毎に，単語単位に分解された歌詞情報を検索バーに入力し，そこから最低でも画素数 680 × 480 以上のものを選出する．その際，検索バーに入力するキーワードは多ければ多いほど，システムが要求する画像により近いものを取得することが可能であるが，あまりに多い場合，該当するものがなくなってしまうので本研究では，入力するキーワードは，6 個までとしている．検索ワードの取得は google history から行い，取得した画像にそれを添付する．ユーザが自身で，追加した画像に関しては，自身でメタデータを入力してもらうと，とても負担になるので，画像のタイトルをそのままその画像のメタデータとして添付する．

これに対し、DEIM のポスター発表にて・画像のほうが音楽よりも心理的に及ぼす作用が大きいため、画像で思い出などを連想させる事は難しい・エフェクトを付ける際に必要な情報のビート、BPM などなどの情報を読み取るために、Songle というサービスが有る (有料) ・画像の特徴情報を読み取るために、Amazon Rekognition というサービスが有り、これにより画像に添付するメタデータを自動で検出し、貼付可能であるというご意見をいただきました。これを元に、画像表示の際のエフェクトを決定する際、考慮する情報を Songle から取得できる音程情報や C6、Dm7 のようなギターコードなどの情報に変更することでより精密に曲に適した画像エフェクトを選出できると考えています。

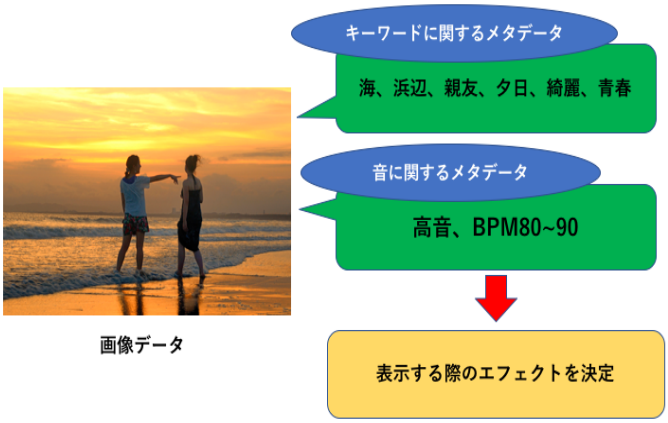

図 4　画像データへのメタデータ付与の図

図 4 は，Google 画像検索から画像取得する際の例を示している．今回検索に使用したワードは「海」「浜辺」「親友」「夕日」「綺麗」「青春」である．BPM，音の高低，キーワード情報より画像のエフェクトを決定する．

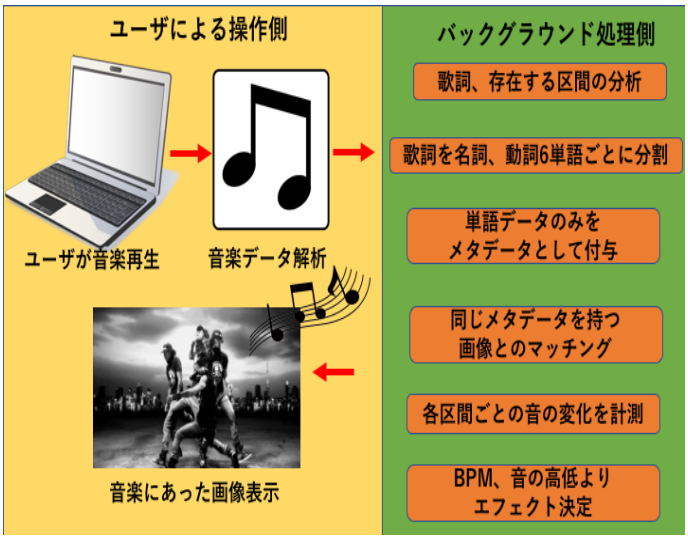

図 5　Personalized Visual Jockey の概要

3.4 Personalized VJ の構築

事前に歌詞情報より，単語単位のメタデータが付与された音楽データを用意し，それらの曲に関しては，BPM や音の高さがわかっているものとする．それを再生した時,「りんご」「ぶどう」などのキーワードのメタデータと「高音」「低音」「BPM 100〜120」などの音に関するメタデータに，より多く該当するメタデータを持つ画像を選出し，表示する．

画像側が持つメタデータの最大個数はキーワード情報のメタデータ 6 個，音に関するメタデータ 2 個の合計 8 個である．以上のことを踏まえた上で，曲を再生した際の話をすすめると，まず再生された段階で，バックグラウンドで 1 秒から 10 秒までの間で「A」というキーワードだから「A」というキーワードをメタデータとして持つものはないか，というように画像と音楽とのマッチング処理が行われる．

次にマッチングした画像に対して，曲の BPM，高低よりエフェクトをかける処理が行われる．楽しいのに曲調が暗い曲や，悲しいのにとても愉快な疾走感のある曲は今回例として適さないので，省くものとする．エフェクトに関しては,「高音× BPM 早い」,「高音× BPM 遅い」,「低音× BPM 早い」,「低音× BPM 遅い」の 4 つの組み合わせで判定するものとし，エフェクト処理に関しては，曲の小節ごとに判定するのではなく，BPM，音の高低に随時対応し，変化させるものとする．これにより音楽に合わせ，最適な画像を最適なエフェクトで表示することが出来る．

4. ま と め

本稿では，ユーザの嗜好や個人的コンテンツを利用した Personalized Visual Jockey を提案した．本手法の特徴は，あらかじめ作られた映像を流すのではなく，音楽から推定される場面や印象に合致する，個人が有する思い出の写真やムービーの利用を可能とするものである．また，一緒に居るメンバー等のコンテキスト情報も考慮することが可能であり，同じ曲でもホームパーティ時，デート時，１人でくつろいでいる時のそれぞれで異なる映像を提示することが可能になると考えている．

今後は，音楽データベースおよび，画像データベースの構築を行い，各データベースへのメタデータ付与およびプロトタイプシステムの開発を行う．

文　　献

[1] 浅田真理，高橋光輝，香田夏雄, VJ 表現を使った環境演出の展開事項と考察, 情報処理学会研究報告,Vol.2014-GN-90 No.9, 2014.

[2] 寺田努，塚本昌彦，西尾章治郎, アクティブデータベースを用いたカラオケの背景作成システム, 情報処理学会論文誌,Vol.44 No.2, 2003.

[3] 中野倫靖，室伏空，後藤真考，森島繁生, Dancer Producer : Web 上の音楽動画を再利用して新たな音楽動画を自動生成する N 次創作支援システム, 第 27 回 NICOGRAPH 論文コンテスト 論文集, I-2, pp.1-8, September 2011.

[4] 草間かおり，伊藤貴之, MusCat : 楽曲の印象表現に基づいた一覧表示の一手法，研究報告音楽情報科学（MUS），2009-MUS-81，19 号，pp.1-6，2009.

[5] 木村翔，久留島寛也，西本一志, 楽曲構造を反映した VJ 表現の全体デザインを支援する歌詞インターフェース, 情報処理学会研究報告,Vol.2014-HCI-157 No.16, 2014.