情報処理学会論文誌 Vol.54 No.4 1263–1274 (Apr. 2013)

歌詞情報に基づくWeb画像検索を利用した歌詞連動スライドショー生成システム

石先 広海[1,a)] 舟澤 慎太郎[2] 帆足 啓一郎[1] 小野 智弘[1] 甲藤 二郎[2]

受付日 2012年7月2日, 採録日 2013年1月11日

概要：本研究ではユーザが入力した楽曲に対して，楽曲の歌詞に基づいて検索したWeb画像と楽曲を同期させて再生するスライドショー生成システムについて提案する．楽曲歌詞の内容に適した画像を楽曲と同期させて再生することで，楽曲の情景表現を向上させ，より印象深い音楽体験の実現を目指す．具体的には，歌詞中の単語と，歌詞から推定した全体印象語から最適な画像検索クエリを抽出し，表示候補となる画像をWeb上から取得する．取得した画像に付与されているソーシャルタグと全体印象語の適合度を用いることで，歌詞の各行と連動して表示させる画像を選定する．さらに，歌詞の各行を表示する時間の最頻値を利用してスライドショー再生時の画像切替えを自動化する．最終的に，被験者評価実験により本システムの有効性を示す．

キーワード：マルチメディアアプリケーション，楽曲スライドショー，歌詞，Web画像

Music Slideshow Generation Based on Web Image Retrieval with Queries Constructed from Lyrics

Hiromi Ishizaki[1,a)] Shintaro Funasawa[2] Keiichiro Hoahsi[1] Chihiro Ono[1] Jiro Katto[2]

Received: July 2, 2012, Accepted: January 11, 2013

Abstract: In this paper, we propose a system to generate slide shows for music which users selected, by utilizing Web images retrieved by queries constructed from song lyrics. The proposed system aims to provide new and impressive user experiences by using synchronized Web images with lyrics line. First, we propose a method to select images to compose the slideshow from the result of Web image retrieval based on queries extracted from lyrics. The system selects matched images with the whole impression of lyrics and removes images which have many social tags not related to the image content. Furthermore, we propose a method to adjust presentation periods of the images in the slides. Finally, subjective experiments are conducted to evaluate effectiveness of our system.

Keywords: multimedia application, music slideshow, lyric, Web image

1. はじめに

音楽と映像や画像を効果的に組み合わせることで，それらを単体で視聴するよりもエンタテインメント性の高いコンテンツを作成し，ユーザに新たな音楽視聴体験を提供することができる．映画やテレビ番組，プロモーションビデオなどのコンテンツでは，映像と音楽を効果的に融合することで作品の価値を高めている．たとえば，テレビドラマのシーンに悲しい音楽をBGMとして再生することで悲しみの感情を強調している．このような聴覚と視覚の相互作用により，各コンテンツが持つ効果を増幅させることが可能である[1]．表示画面がある一般的な音楽プレーヤでは，楽曲再生時に音楽とともに画像を表示させる機能（ビジュ

[1] KDDI研究所
KDDI R&D Laboratories, Fujimino, Saitama 356–8502, Japan
[2] 早稲田大学理工学術院基幹理工学部
Graduate School of Fundamental Science and Engineering, Waseda University, Shinjuku, Tokyo 169–8555, Japan
[a)] ishizaki@kddilabs.jp

アライザ）が備え付けられていることがある．このような音楽プレーヤ上で楽曲の雰囲気や状況を表現する画像を表示させることにより，視聴者により印象深い音楽体験を提供することができる．

しかし，映画やテレビ番組のような聴覚・視覚効果を付与したコンテンツをユーザが制作する場合，コンテンツを構成する素材となる映像・画像の収集や選択，さらには構成の検討など様々な作業が必要となる．このため，映像作品の制作に慣れていないユーザにとって，自身が所有する楽曲や映像・画像を用いて新たな映像作品を制作することは多くの労力を要する．たとえば，ユーザが所持する画像やインターネット上の画像を楽曲再生時のビジュアライザとして利用する際に，楽曲の情景や歌い手の心情と一致した画像を提示することが重要であるが，このような画像をユーザ自身が探し出すことは多くの時間が必要である．このように，一般ユーザが楽曲や映像・画像などを利用して新たなコンテンツを制作することは困難であり，コンテンツ制作を支援するシステムの必要性が高まっている．

そこで本稿では視覚効果として複数の画像を切り替えて表示するスライドショーに着目し，画像と楽曲を同期させて再生する楽曲スライドショーを生成するシステムを提案する．本システムは楽曲の雰囲気・状況に適した画像を，楽曲とともに再生するビジュアライザとしての利用に焦点をあて，画像共有サイトの画像（以下，Web 画像）から歌詞の雰囲気に適した画像を自動で検索・選定し，再生中の楽曲・歌詞と効果的に同期表示する．具体的には，歌詞の各行（以下，歌詞行）から画像を検索するための検索クエリ選定方法と，画像群と歌詞全体の印象との適合度に基づく画像選定，歌詞行の再生時間の最頻値を指標とした画像の切替えタイミングの自動制御方法を適用することで，高品質な楽曲スライドショーを生成する．歌詞は楽曲の内容を直接的に表現する特徴であるため，歌詞の特徴に基づいて Web 画像を検索することにより，楽曲の内容・雰囲気にあったスライドショーを作成できると考えられる．最終的に被験者による主観評価実験により，提案システムの有効性を検証する．

本稿では 2 章で本研究の関連研究と問題点を提示する．3 章では前記課題を解決するための方式・システムについて説明し，4 章で提案システムの有効性検証実験および考察を記述する．最後に 5 章でまとめと今後の課題について述べる．

2. 関連研究と問題点

本章では楽曲と画像を連動させたコンテンツを生成する手法・システムと，それに対する課題について説明する．これまでに楽曲の表現する情景や歌い手の感情を推定するための手法が考案されてきた [2], [3], [4]．たとえば，文献 [2] では音響信号に基づいて，楽曲が表現する感情情報（喜怒哀楽など）を推定する手法について提案している．文献 [3] では事前に用意した学習データにより学習させた Supervised Multi-Class Labeling 手法を利用することで，楽曲に対して複数の状況単語（パーティなど）を自動で付与させる手法について提案している．文献 [4] では，大規模データセットにより音響特徴からソーシャルタグを楽曲に付与するシステムについて提案している．これらの手法では共通して楽曲の音響的特徴を利用して状況・感情情報をタグとして付与している．

1 章に記述したような聴覚・視覚効果を付与したコンテンツを生成する手法・システムとして，楽曲と画像を連動させたコンテンツを生成する手法・システムが提案されている [5], [6], [7]．文献 [5], [6], [7] では，ユーザ自身が撮影した写真や映像を楽曲と連動させることで，スライドショーなどのコンテンツを生成している．これらのシステムではユーザ自身が撮影した画像や映像を用いることで，ユーザにとって親しみのあるコンテンツを生成できるという利点がある．一方で，これらのシステムではコンテンツを生成するためにユーザ自身が画像や映像など大量の素材を準備する必要があり，一般ユーザにとっては敷居が高い．そのほかに，楽曲の歌詞を利用して Web 画像を検索し，楽曲と Web 画像を連携させて再生するミュージックビデオを生成するシステムや楽曲特徴に合わせて画像の同期再生タイミングを制御するシステムが提案されている [8], [9], [10]．しかし，これらのシステムでは以下に記述する主な 3 項目の課題がある．

(1) 画像と楽曲情景の関連性低下による視聴者への違和感
(2) 画像検索精度および信頼性の低下
(3) 画像切替えタイミングによる視聴者に対する違和感

文献 [8], [9] では歌詞中の単語を利用して Web 画像を検索・利用する方式が提案されているが，課題 (1) および (2) が存在する．課題 (1) の主な要因として楽曲の表現する情景とスライドショーを構成する画像が一致しないことや，画像の内容が急激に変化することがあげられる．たとえば，歌詞中の単語を利用して得られた画像群から適した画像を選定する処理に重点を置いているが，検索クエリの選定では stop word の排除や，利用する品詞の限定などの最低限の処理しか適用していない．したがって，歌詞に出現する単語が一般的かつ，歌詞の情景を表現する単語ではない場合に関連性が低い画像が表示される可能性があるため，楽曲情景を考慮した画像検索方法が必要となる．

さらに，視聴者に対する違和感の原因としてスライドショー全体の統一感の欠如があげられる．文献 [8], [9] では表示する画像の連続性については考慮していない．したがって，単一の検索クエリや抽象的な検索クエリを利用した場合に検索結果の画像群のが表現する情景が多様となる場合がある．たとえば，歌詞中の「あなた」という単語を検索クエリとした場合に「人間，山，昼」の写真の次に

「人間，海，夜」となる画像が表示されるなど，風景・時間帯が異なる画像が表示されることがある．さらに，「あなた」を指す内容が「人間」ではなく猫や犬などのペットに関する画像が表示されるなど，楽曲スライドショーを構成する画像間の関連性が低くなり，スライドショー全体としての統一感が損なわれる可能性がある．そのため，スライドショーを構成する画像の連続性を考慮する必要がある．

課題 (2) の主な要因として，画像共有サイトの画像に対して内容と関係のないソーシャルタグが多く付与されていることがあげられる．たとえば画像共有サイトの 1 つである Flickr [*1] では画像の内容とは関係のないソーシャルタグが付与される場合がある．このようなソーシャルタグはメタノイズ [11] と呼ばれ，画像検索の信頼性や精度を乱す原因として知られている．文献 [8] では単純に名詞情報を利用して画像共有サイトを検索しているため，メタノイズが原因で検索クエリに適さない画像が検索結果として得られる可能性がある．さらには，検索クエリが変化しても同一の画像が検索結果の上位に存在するため，スライドショーの中で何度も同じ画像が表示されるなど，ユーザに倦怠感を与えるなどの問題もあげられる．

文献 [8], [10] では，楽曲テンポに合わせて楽曲と画像を同期させる方式が提案されているが，課題 (3) が存在する．テンポ・ビート推定技術では，推定結果が正解テンポ情報に対して 2 倍もしくは半分の値で得られるなどの精度の問題がある [12]．推定結果が正解テンポ情報に対して 2 倍もしくは半分の値が与えられた場合，本来はテンポの遅い楽曲に対して画像が頻繁に切り替わるなど楽曲の雰囲気に適さない表示となる可能性がある．改善案として楽曲歌詞の表示される行を基準として画像を切り替えることも可能であるが，画像を表示する時間が歌詞の行の長さに依存するため，画像が短時間で切り替わる，もしくは長時間表示されることがある．たとえば，バラード調のテンポの遅い楽曲に短い行の歌詞が含まれていた場合，画像が急激に切り替えられることがある．このような急激な画像の切替えは視聴者に違和感を与える可能性があるため，画像の表示時間の調整が必要となる．

3. 提案システム

そこで，本稿ではスライドショーの素材として Web 画像を利用する楽曲スライドショーシステムに着目し，2 章で記述した 3 項目の課題を解決する．課題 (1) については，歌詞から人が受ける全体的な印象（季節，時間帯，天候など）を全体印象語として定義し，システムで利用する楽曲の歌詞から全体印象語を推定する．次に，歌詞行および歌詞行が含まれる段落単語の組合せを検索クエリとして利用することで歌詞の情景に適した画像を検索し，画像間の統一感の向上を図る．課題 (2) については，検索結果の画像群に付与されたソーシャルタグと全体印象語との関連性を考慮することでノイズタグが多く付与された画像を排除し，歌詞行と同期して再生するための最適な 1 枚を選定する．課題 (3) については，スライドショーを構成する各画像の表示時間を，歌詞に付与した同期再生情報に基づいて均一化することで，画像表示の急激な切替えや，同一の画像が長時間表示されるという問題の改善を図る．

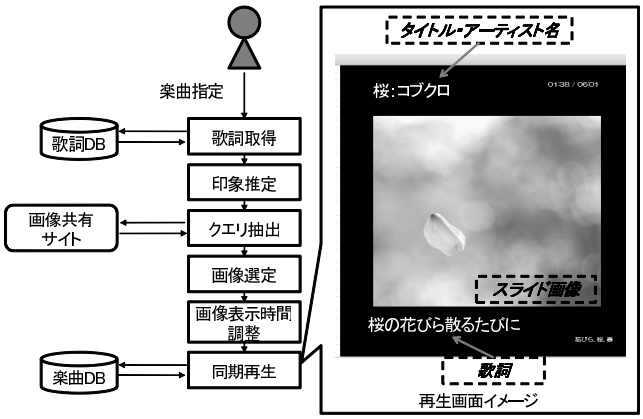


図 1 提案システムの処理の流れと画面イメージ
Fig. 1 Conceptual illustration of procedures of the proposal.

本システムの概要図を**図 1** に示す．本システムはユーザが楽曲を指定すると楽曲歌詞から全体印象語を推定し，行および行が属する段落に含まれる単語との組合せから検索クエリとなる単語群を抽出する．抽出した検索クエリにより，Web 上の画像共有サイトから各行と同期して再生する候補となる画像群を取得する．取得した候補画像群から歌詞行に適した画像を 1 枚選定し，楽曲歌詞と同期させて再生する．以下に各処理の詳細を説明する．

3.1 歌詞全体の印象推定

本節では楽曲スライドショーに統一感を与え，画像と歌詞の関連性をより強化するために楽曲歌詞から楽曲の情景を表現する全体印象語を推定する．再生画像の選定の際に全体印象語を利用することで歌詞と再生画像の関連性を強化する．さらに本処理で推定した全体印象語を考慮して画像を選定することでスライドショー全体に統一感を与えることが可能になる．

まず，歌詞から受ける印象の評価データを予備実験により収集する．印象の評価データは Support Vector Machine [13]（以下，SVM）の学習データとして利用し，入力された歌詞に対して印象の正負を判別する分類器を構築する．最終的に，楽曲データベースのすべての楽曲に対して分類器を適用し，分類された印象を全体印象語として歌詞に付与する．以下に詳細な処理について説明する．

3.1.1 全体印象語推定のための学習データ収集

予備実験により，本稿で利用する全体印象語の選定と全

[*1] http://www.flickr.com

体印象語を推定するSVMを構築するための学習データを収集する．具体的には被験者20名に歌詞を提示し，事前に設定した全体印象語候補群の中から歌詞から受ける印象に適したものを選択してもらう．各歌詞について評価者の半数以上が選択した全体印象語をSVMの学習データとして利用する．20名の被験者から240曲に対する評価データを収集するために，1楽曲あたり5名分の評価が得られるように被験者ごとに閲覧リストを作成した．最終的に，各歌詞について評価者5名のうち3名以上が選択した全体印象語を学習データとして利用する．被験者に提示した歌詞は市販されているJ-POP 240曲を用いた．なお，歌詞を提示する際に被験者には歌詞以外の情報（タイトルやアーティスト情報など）は提示していない．

被験者に提示した全体印象語候補群は，梶らの文献[14]で利用されている楽曲情景アノテーションと楽曲に対するソーシャルタグを提供しているmonstar.fm[*2]より収集したソーシャルタグ情報および，Hevnerにより分類された感情を表現する形容詞[15]を利用した．文献[14]では，予備実験によって，いつ・どこで・どのような心理状態であるかという項目を楽曲情景アノテーションとして定義している．また，文献[15]では，音楽中に利用される66種類の形容詞（ユーモア，喜びなど）を8種のカテゴリに分類し，円環状に配置したadjective circleを作成することで楽曲感情表現の関係性について提案している．

3.1.2 歌詞解析による楽曲への全体印象語付与

前項で得られた評価データを利用して，未知の楽曲歌詞に全体印象語を付与するための識別器をSVMにより作成する．テキスト解析分野ではSVMは文書作成者の性別判定などの2クラス分類に広く用いられており，識別精度は一般的に高いことが知られている[16], [17]．本処理ではSVMにより，歌詞に対する全体印象語の正否を識別することで全体印象語を付与する．

まず，全体印象語候補となる単語から利用可能な全体印象語を抽出するために，すべての印象語の正否を判別する識別器を構築する．構築した識別器を用いて，同一の概念中（季節，時間帯，天候，心理，テーマ）のすべての単語における識別精度を交差検定法による識別精度が70%以上となるものを本稿における全体印象語として選定した．**表1**に実際に選定した全体印象語を記載する．また括弧内には実際に被験者が楽曲に対して全体印象語として選択した数を示してある．本処理の流れを**図2**に示す．楽曲の歌詞はTF*IDF法によりベクトル形式で表現する（以下歌詞特徴ベクトル）．ベクトルの要素は，歌詞データベース内の3,062楽曲において10曲以上で使用されている名詞を利用し，1,070次元で表現する．各要素の値はTF*IDF値を用いる．

[*2] http://monstar.fm

表1 学習データ収集実験から抽出した全体印象語
Table 1 Concepts and category labels for describing general impression of music in collecting training data for support vector machine.

概念	全体印象語
季節	春 (88)，夏 (232)，秋 (61)，冬 (127)
時間帯	朝 (79)，昼 (125)，夕方 (111)，夜 (351)
天候	晴 (351)，曇 (36)，雨 (134)，雪 (67)，虹 (8)

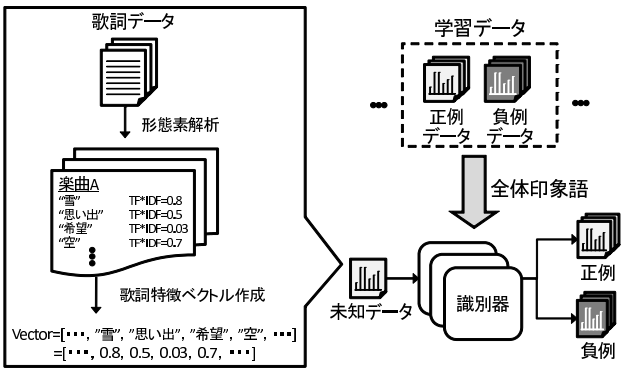


図2 Support vector machineによる全体印象語付与
Fig. 2 Impression labeling with support vector machine.

次に歌詞特徴ベクトルと，3.1.1項で収集した学習データを用いてSVMを構築する．正例データとして，学習データの全体印象語が付与された楽曲の歌詞特徴ベクトルを利用し，負例データは全体印象語が付与されていない楽曲の歌詞特徴ベクトルを利用する．全体印象語ごとに識別器を構築し，未知の楽曲歌詞を入力した際にすべての識別器を用いて，正負の判断をする．このとき，同一概念内で複数の正例が付与される場合には，SVMの識別境界平面に対する距離値が大きい全体印象語を付与する．SVMは，SVMLight[*3]を使用して構築し，線形カーネルを用いて分類する．

3.2 ソーシャルタグを利用したクエリ抽出

Web画像から楽曲全体の印象に適した画像を取得するために，楽曲歌詞の各行から抽出した検索クエリを利用して画像共有サイトの画像を検索する．本処理では「多数のユーザから，多数の画像に付与された単語はソーシャルタグとして重要な意味を持ち，画像の内容を表現する単語としての可能性が高い」という仮説をたて，画像共有サイトの検索結果画像に付与されたソーシャルタグの出現頻度に基づいて検索クエリを抽出する．これにより，各行の情景表現に有効な画像群を取得するための単語を検索クエリとして抽出する．ソーシャルタグを利用したクエリ抽出の流れを**図3**に示す．

検索クエリとして実際に利用した単語群をW，Wを利用して画像共有サイトから得た検索結果に含まれる画像

[*3] http://svmlight.joachims.org

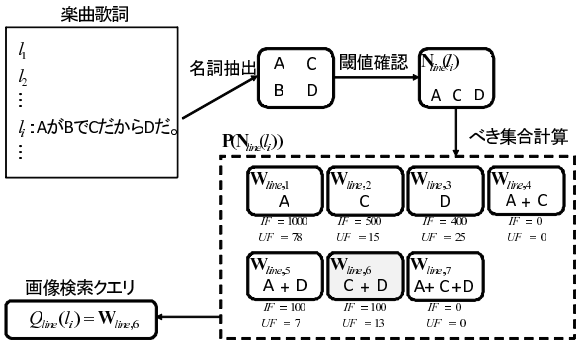


図 3 ソーシャルタグを利用したクエリ抽出の流れ
Fig. 3 Query extraction based on social tag.

数を $IF(W)$（Image Frequency），検索結果に含まれる画像を投稿したユーザ数を $UF(W)$（User Frequency）と表現する．歌詞は複数の段落によって構成されており，本処理は段落ごとに処理を行う．各段落を構成する歌詞行の i 行目で使用されている名詞群を抽出しクエリ候補とする（$N_{line}(i)$）．なお，前述した仮説に基づき，ノイズとなる検索結果を排除するために検索結果の IF および UF に対して閾値を設定した．閾値は単語の組合せがある程度一般的に使用されていると判断ができる画像数とユーザ数を考慮し，$IF = 40$，$UF = 10$ と設定し，閾値以下となる名詞を排除した．

クエリ候補 $N_{line}(i)$ から得られるすべての単語の組合せを考慮するために，べき集合 $P(N_{line}(i)) = \{W_{line,1}, W_{line,2}, \cdots, W_{line,x}\}$ を計算する．ここで x はべき集合の要素数を表し，$W_{line,x}$ はべき集合 $P(N_{line}(i))$ の部分集合（名詞群）を表す．得られたすべての部分集合を検索クエリとして画像共有サイトから画像を検索し，$IF(W) \neq 0$ かつ，W 内の単語数が最大となる部分集合を W_{max} として抽出し，実際の検索クエリ $Q_{line}(l_i)$ として利用する．なお，W_{max} が複数存在する場合には $UF(W)$ が最大となる W_{max} を検索クエリとする．

次に，検索クエリの単語数を拡張するために歌詞の段落情報に含まれている名詞情報を利用する．段落情報は意味段落と形式段落の 2 種類が存在するが，意味段落は複数の形式段落により構成され，楽曲の表現する情景や意味的な区切りとして理解できる．そのため，歌詞に含まれる段落情報から抽出する名詞は行が表現する内容を含有している可能性が高い．したがって，段落中の単語により検索クエリ数を拡張することで段落ごとに画像の統一感を高める効果が期待できる．また，段落情報の確定性についてインターネット上で歌詞を提供している 3 種の Web サイト[*4]を調査したところ，サイト上で提供されている歌詞検索機能を利用して入力したすべての J-POP 楽曲（10 曲）において段落情報が一致していた．この結果より，段落情報はある程度の確定性が保たれていると考えられる．以下に段落情報含まれている名詞情報を利用した検索クエリ拡張について具体的な手順を示す．

検索クエリ $Q_{line}(l_i)$ と，歌詞の i 行目が属している段落で利用されている名詞（群）$N_{para}(l_i)$ とのべき集合の要素から最終的な検索クエリを抽出する．具体的には，$Q_{line}(l_i)$ と $N_{para}(l_i)$ のべき集合 $P'(N_{para}(l_i)) = \{W_{para,1} \cup Q_{line}(l_i), W_{para,2} \cup Q_{line}(l_i), \cdots, W_{para,y} \cup Q_{line}(l_i)\} = \{W'_{para,1}, W'_{para,2}, \cdots, W'_{para,y}\}$ を計算し，前段落と同様の処理によって W'_{max} を計算し，最終的な検索クエリ $Q'_{line}(l_i)$ を抽出する．最終的にすべての行に対して，検索クエリ $Q'_{line}(l_i)$ を用いた AND 検索により候補画像群を取得する．なお，$Q'_{line}(l_i)$ が抽出されない場合には，楽曲 m に付与された全体印象語 $N_{all}(m)$ を検索クエリとして用いる．

3.3 全体印象語との適合度を利用した画像選定

本処理では全体印象語との適合度を利用することで，検索クエリから得られた画像群から各行ごとに同期して再生する画像を選定する．全体印象語との適合度は，1 枚の画像に付与されているすべてのソーシャルタグと全体印象語との関連度に基づいて計算する．関連度は全体印象語が属する概念を基準に計算し，同一概念において特定の全体印象ごと特に共起する単語の関連度が高くなるように設計することで，ソーシャルタグのメタノイズを考慮した画像の選定を図る．

まず，入力楽曲の全体印象語と関連が強いソーシャルタグが多く付与されている画像を選定するために，全体印象語とソーシャルタグの関連の強さを表す関連度を画像共有サイト上のソーシャルタグの条件付き確率をもとに算出する．1 つの画像に対して付与された複数のソーシャルタグは情景表現として関連性があると考えられ，多くのユーザから同時に付与されたソーシャルタグは関連性が高いと推定できる．たとえば，歌詞に付与された春という全体印象語との条件付き確率が高いソーシャルタグは春との関連性が高いと判断できる．一方で，同じ概念に属する他の全体印象語（夏，秋，冬）との条件付き確率も同様に高い場合には，ノイズタグである可能性が高い．そのため，春に対して条件付き確率が高く，同じ概念に属する他の全体印象語との条件付き確率が低いソーシャルタグを関連タグとして抽出する．

本処理では，条件付き確率 P は，UF を用いて計算し，以下のように表現する．

$$P(t, n_{all}) = \frac{UF(t \cap n_{all})}{UF(n_{all})} \quad (1)$$

$UF(n_{all})$ は，全体印象語 n_{all} を検索クエリとした場合に得られた結果の UF 値，$UF(t \cap n_{all})$ はソーシャルタグ t と n_{all} が同時に付与された画像の UF 値を表す．タグ t が同一概念中のある特定の全体印象語のみとともに使用され

[*4] UTA-NET: http://www.uta-net.com/，うたまっぷ.com: http://www.utamap.com/，歌詞 GET: http://www2.kget.jp/

る場合に関連度が高いという考えに基づいて，タグ t の全体印象語 n_{all} に対する関連度 $R(t, n_{all})$ を設計する．具体的には，式 (1) を利用してソーシャルタグ t の全体印象語 n_{all} に対する関連度 $R(t, n_{all})$ を以下のように表す．

$$R(t, n_{all}) = P(t, n_{all}) - \frac{\sum_{n \in C, n \neq n_{all}} P(t|n)}{|C| - 1} \times w \quad (2)$$

ここで，C は全体印象語 n_{all} が属する概念内のすべての全体印象語を表す．たとえば，$n_{all} =$ 春 のとき，春は季節概念に属するため $C =$ { 春, 夏, 秋, 冬 } となる．w は式 (2) 中の第 2 項による影響を調整するための係数で，事前に作成した関連単語リストと非関連単語リストを手動で生成し，w を変化させることで前記リストに記載された単語の関連度総和の差が最も高くなるように設定する（$w = 3$）．また，関連度の出力結果を目視により確認し，関連単語リストに記載されている単語が含まれるような関連度を閾値として抽出する（0.024）．また，前述した仮説に基づき $UF(t) \geq 5$ を満たすソーシャルタグを関連タグと呼ぶ．

画像と全体印象語の適合度は，画像に付与された関連タグの関連度を用いて計算する．適合度は楽曲に付与された全体印象語との関連が高いソーシャルタグが多く付与されるほど大きい値を示すように設計した．以下に適合度の計算式を示す．

$$score(i) = \sum_{n_all \in N_{all}(m)} \frac{\sum_{t \in T_i \cap T_{related}(n_{all})} R(t, n_{all})}{|T_i| - |T_i \cap T_{related}(n_{all})|} \quad (3)$$

T_i は画像 i に付与されているソーシャルタグを表し，$T_{related}(n_{all})$ は全体印象語 n_{all} に対する関連タグとする．適合度はすべての歌詞行の候補画像に対して計算し，適合度が最大となる画像を各行ごとに選定する．式 (3) により，ノイズとなるソーシャルタグが多く付与された画像を排除し，ソーシャルタグ中の関連タグの割合が高い画像を優先的に選定することができる．ここで，ある全体印象語において $|T_i| - |T_i \cap T_{related}(n_{all})| = 0$ となる場合には，その印象語におけるスコア値を 1 として扱う．

3.4 行再生時間の最頻値を利用した画像表示時間調整

本処理ではスライドショーにおける画像表示時間の単位を歌詞行に着目し，画像表示時間が視聴者に与える違和感の解消を図る．歌詞行再生時間の最頻値を利用して再構成することで，視聴者の違和感の要因と考えられる急激な画像の切替わりや，同一画像が長時間表示されるという問題の解決を図る．具体的には，表示時間の長短を閾値により判断し，閾値以下の行は前後の行と結合し，閾値以上の行は分割する．さらに，楽曲の画像表示時間の最頻値を利用して，間奏区間における画像切替えタイミングを設定する．最終的に選定した画像を歌詞行と連動させて再生する．以下に手順の詳細を示す．

(1) 楽曲の歌詞の各行における表示時間を算出し，それらの最頻値を基本表示時間 I として定義する．
(2) 段落の切り替わる箇所において，段落の間における演奏時間が d_{min} [sec] 以上ならば，その区間を間奏として抽出する．
(3) 表示時間が d_{min} 以下の行を次の行と結合する．次の行がなければ，前の行と結合する．ただし，結合は同段落に属する行どうしでのみ行う．このように，行が統合された場合，統合後の行に対して画像を 1 枚検索する．
(4) 表示時間が d_{max} [sec] 以上の行を等分割する．ただし，分割後の表示時間が基本表示時間 I に最も近くなるように分割数を調節する．このように，行が n 分割された場合，分割前の行に対して検索された候補画像から上位 n 枚を選択し表示する．また，間奏区間に対しても同様に分割し，画像検索クエリには全体印象語を用いる．

4. 主観評価実験

提案システムの有効性を検証するため，被験者による主観評価実験を実施した．本章では以下 2 種類の実験について記述する．

実験 1　画像表示時間に対する比較実験
実験 2　提示画像の品質に対する比較実験

実験 1 では，3.4 節に記載の画像表示時間調整の有効性を確認するために，被験者により画像の表示時間について評価した．実験 2 では 3.1 節，3.2 節および 3.3 節に記載の方法により取得・生成したスライドショーの品質について有効性を検証する．具体的には比較対象となる 2 種類のシステムを実装し，各システムによって作成された楽曲スライドショーの被験者による評価を収集した．以下に各実験の詳細について記述する．

4.1 画像提示時間に対する比較実験

4.1.1 実験方法

本実験では画像表示時間調整の有効性を確認するために，3.4 節に記載の画像表示時間調整方式の有無が被験者の評価値に与える影響について分析する．被験者は大学生 20 名（男性：18 名，女性：2 名，20 代：20 名）を対象に，市販されている J-POP（20 曲）を利用した．すべての楽曲に対して楽曲中の歌詞が再生される時間情報（同期情報）を手動で付与した．実験に利用するスライドショーは，3.1 節，3.2 節および 3.3 節に記載の方法により作成し，画像表示時間調整方式を適用したスライドショー（時間調整方式）と，同期情報に基づいて行単位で画像を切り替える方式を適用したスライドショー（行切替え方式）を各楽曲について作成した．なお，本実験では，事前に表示時間の許容範囲をアンケートにより調査し，暫定的に表示

表 2 画像表示時間に関する評価基準
Table 2 Table of evaluation criteria in terms of image presenting time presented to subjects.

評価値	評価基準
5	大多数の画像において適切であった
4	多くの画像において適切であった
3	半分程度の画像において適切であった
2	多くの画像において不適切であった
1	大多数の画像において不適切であった

時間の閾値を $d_{max} = 12$ [sec]，$d_{min} = 4$ [sec] と設定した．

また，比較方式として MusicStory [8] を追加した．MusicStory では歌詞中に含まれるすべての名詞を検索クエリとして抽出し，OR 検索を利用して画像共有サイトより画像を抽出する．抽出した画像群は，楽曲の BPM（Beats Per Minute）を基準として 1 小節に相当する 4 拍子に該当する時間単位を計算し，該当時間あたり画像 1 枚を切り替えて表示した．BPM は楽曲の BPM 推定ツールである *Beatroot* [18][*5] を用いて推定した．なお，MusicStory では画像の表示単位が BPM の自動推定結果に基づく 1 小節分であることや，検索クエリ抽出範囲が歌詞全体であるなど，条件を統一させることが困難であるため結果は参考値として取り扱う．

被験者は各スライドショーを閲覧したのちに，スライドショーを構成する画像を表示する時間は適切であったかどうかについて評価値を付与した．評価基準を**表 2** に示す．被験者へは画像表示時間の適切さの判断基準として，「楽曲スライドショー中に表示される画像を十分な時間閲覧することができたか，また，同じ画像が長時間表示されていないか」という教示を提示した．

4.1.2 実験結果

本節では被験者より付与された評価値を集計し，時間調整方式と行切替え方式の評価値平均を比較した．**図 4** に各スライドショーにおける表示時間の適切さに対する評価値平均を示す．なお，図 4 では分散値を用いた誤差範囲が示されている．図 4 からも明らかなように，提案手法の評価値平均は比較 2 手法に比べて大きく評価値を向上させることができた．MusicStory によって生成されたスライドショーの評価値平均は 2.54，調整方式が非適用だったスライドショーの評価値平均は 3.66，時間調整方式を適用したスライドショーの評価値平均は 4.29 であった．また，各スライドショーに対する評価値の分散値は，MusicStory では 0.14，行切替え方式を適用したスライドショーでは 0.21，時間調整方式を適用したスライドショーについては 0.76 となった．

この結果から明らかなとおり，時間調整方式を適用したスライドショーの評価値は行切替え方式を適用したスライドショーに比べ，安定して全体的に評価値が高い．t 検定により評価値平均の差は有意であることが確認できた（$p < 0.001$）．これらの結果より，3.4 節に記載の行再生時間の最頻値を利用した画像表示時間調整方式が楽曲スライドショーの画像切替えに対して有効であるといえる．なお，本実験により画像表示時間の閾値（$d_{max} = 12$ [sec]，$d_{min} = 4$ [sec]）について有効性を確認できた．

[*5] http://www.eecs.qmul.ac.uk/šimond/beatroot/

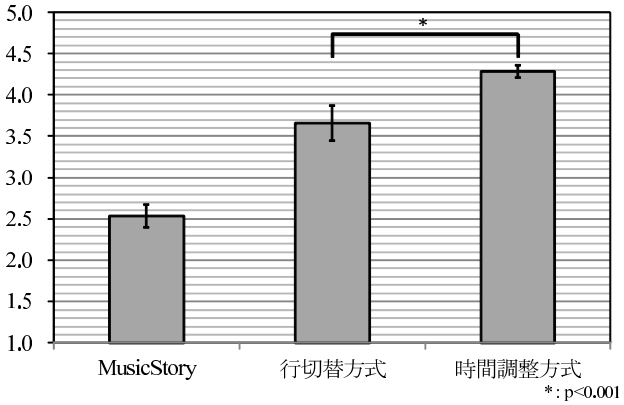


図 4 各スライドショーにおける表示時間の適切さに対する評価値平均
Fig. 4 Relationship between averages of user ratings of each slideshow and presenting time of images.

4.2 提示画像の品質に対する比較実験

4.2.1 評価システム詳細

次に，楽曲スライドショーを構成する画像およびスライドショー全体の品質を評価するために，複数のシステムによって作成したスライドショーに対する被験者評価値の比較実験を実施した．評価実験で使用した比較対象となる 2 方式の詳細を説明する．まず 1 つ目の比較方式として，4.1.1 項に記載の MusicStory を利用した．2 つ目の比較方式として，TF*IDF によるクエリ選定（TF*IDF 方式）による方式を利用した．TF*IDF は楽曲を特徴づける歌詞中の単語の重要度を表現する指標である．歌詞中から重要キーワードを抽出し，画像検索をするための比較方式が存在していないため，本稿ではテキスト情報検索の分野で一般的に利用されている重要語抽出の一般的な方式である TF*IDF を利用して画像検索を行う方式を比較対象として実装した．

TF*IDF 方式では，i 番目の歌詞行から名詞を検索クエリとして抽出し，AND 検索により画像共有サイトから画像を抽出する．検索結果が得られない場合，検索クエリから最も TF*IDF 値が小さい名詞を削除し，再び AND 検索を実行する．削除処理を画像が得られるまで順次繰り返す．この処理によって画像が取得できない場合には，前の行の検索処理によって得られた画像群を再利用する．検索結果によって得られた画像は，Flickr が提供するソート方法の中で Interesting という指標（Interesting 指標）を利用し，

表 3 被験者に提示した評価基準
Table 3 Table of evaluation criteria presented to subjects.

評価値	content	unity	quality
5	合っていた	統一性があった	完成度が高い
4	どちらかといえば合っていた	どちらかといえば統一性があった	どちらかといえば完成度が高い
3	どちらともいえない	どちらともいえない	どちらともいえない
2	どちらかといえば合っていなかった	どちらかといえば統一性がない	どちらかといえば完成度が低い
1	合っていなかった	統一性がない	完成度が低い

検索結果の Interesting 指標によるランキングが最上位となる画像を 1 枚選定し，歌詞行と表示させた．Interesting 指標[*6]は誰がいつコメントしたか，誰がお気に入り投稿したか，タグやその他の継続して変化する特徴量を利用して，サイト内での興味の度合いを利用してランキングを計算している．画像の切替えタイミングに関しては手動で付与された同期情報を利用して，行の切替えと連動させて画像を切り替えた．本稿では，TF*IDF における DF は 3,062 曲の J-POP 楽曲の歌詞を利用して計算した．

4.2.2 実験方法

本実験では提案と比較方式によって生成された楽曲スライドショーに対して被験者による評価を付与させた．被験者は一般から募集した 42 名（男性：21 名，女性：21 名，20 代：14 名，30 代：14 名，40 代：14 名）により実施した．なお，すべての被験者はスライドショーを視聴した経験を持つ．楽曲スライドショーを作成するために使用した楽曲は市販されている J-POP（10 曲）を利用し，すべての楽曲に対して楽曲中の歌詞が再生される時間情報（同期情報）を手動で付与した．このとき，3.4 節に記載の画像表示時間調整方法の有効性を検証するために，3.4 節の (3) に記載した歌詞行統合処理のみが実施される楽曲を 5 曲（統合セット），3.4 節の (4) に記載した歌詞行分割処理のみが実施される楽曲を 5 曲（分割セット）利用し，3.2，3.3 節に記載の方法により画像を選定した．1 曲に対して総評価数の 3 分の 2 となる 28 名もしくは 29 名分の評価が付与されるように，42 名を複数のグループに分割して評価を実施した．本実験では，検索対象となる画像共有サイトとして Flickr を利用した．また，名詞を抽出するための形態素解析器として MeCab[*7]を利用した．

被験者へは，1 楽曲に対して 3 種類のシステムから作成されたスライドショーを提示する．提示するシステムの順番はすべてランダムに設定されており，被験者へはどのスライドショーがどのシステムによって生成されたかどうかは通知していない．評価項目は，歌詞と画像の調和度合い（content），スライドショー全体の統一性（unity），スライドショー全体の完成度（quality）の 3 項目を設定し，被験者により 5 段階で評価を付与させた．**表 3** に評価基準を示す．なお，被験者へは評価項目の基準を与えるために以下の説明を教示してある．

- 歌詞と画像の調和
  表示される歌詞と画像の内容が一致しているか，もしくは，歌い手の心情を表現しているか．例，歌詞中に花という単語が出現した際に花の画像が表示される．失恋に関する歌詞のときに，夕暮れで人が佇んだ画像が表示される．
- スライドショー全体の統一性
  スライドショー中に表示される画像の内容が似通ったものであったか，また，全体を思い起こしてみて，違和感を覚えたか．
- スライドショー全体の完成度
  歌詞と画像の調和，画像切替えの適切さ，統一性のすべてを考慮したスライドショーの評価．

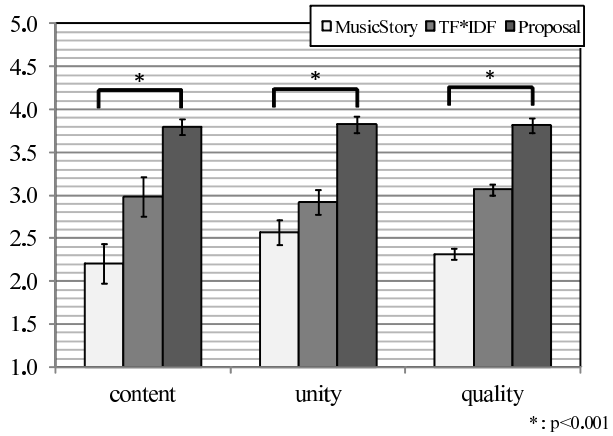


図 5 被験者による 5 段階評価値の平均（content，unity，quality）
Fig. 5 Result of user evaluation: average of user ratings (content, unity, quality).

4.2.3 実験結果

被験者から収集した，すべての方式に対する 5 段階評価値の平均値を**図 5** および**表 4** に示す．図 5 では分散値による誤差範囲を示してある．図 5 から明らかなとおり，比較方式に比べて提案方式の主観評価値はすべての項目に対して高い評価を得た．特に quality の項目に着目すると，quality は content，unity の要素が影響すると考えられるが，提案方式の評価値平均は 3.81，比較方式は MusicStory が 2.31，TF*IDF 方式は 3.07 となり，提案方式の評価値は

[*6] http://www.flickr.com/explore/interesting/
[*7] MeCab:Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/

表 5 提案方式および TF ∗ IDF 方式で利用された検索クエリの比較表
Table 5 Comparison table of image search query from Lyrics on proposal and TF*IDF based method.

段落番号	行番号	歌詞例	提案方式	TF*IDF 方式
1	1	逢いに行くわ 汽車に乗って	汽車	汽車
	2	幾つもの朝を 花の咲く頃に	朝，花	汽車
	3	泣き疲れて 笑った	春，花	汽車
	4	つないだ手と手を 離せないままで	手，花	手
2	5	季節が終わる前に あなたの空を 流れる雲を	空，雲，春	季節，雲
	6	深く眠る前に あなたの声を 忘れないように	前	あなた
3	7	窓を開けたら ホラ	春，窓	窓
	8	飛びこんでくるよ	春	窓
	9	いつか見た 春の夢	春	夢
4	10	雨上がり 胸を染めて	雨上がり	春，夢
	11	幾つもの朝を 花の咲く頃に	朝，花	春，夢
	12	鴨川 越えて 急ごう	鴨川，春	鴨川
	13	古びた景色に はしゃぐ人達も	景色，人	景色，人
5	14	桃色の宴よ 桜の花よ	桜，花，春	桃色，花

表 4 MusicStory，TF*IDF 方式および提案方式の各評価項目における評価値平均
Table 4 Table of averages of each evaluation item on MusicStory, TF*IDF and Proposal.

評価項目	MusicStory	TF*IDF	Proposal
Content	2.21	2.99	3.80
Unity	2.57	2.93	3.83
Quality	2.32	3.07	3.81

比較方式に比べて高い平均値を得ることができている．t 検定によって，すべての評価項目の平均の差は有意であることが確認できた（$p < 0.001$）．これらの結果から提案方式は比較方式に比べて歌詞の情景表現に関連した画像を表示することができたといえる．

5. 考察

5.1 画像と楽曲情景の関連性向上の効果

画像と楽曲情景の関連性向上に対する効果を分析するために，TF*IDF に基づく方式と，提案方式のそれぞれについて検索クエリとして利用された単語を比較した．**表 5** に実際に実験で利用した楽曲の歌詞と，提案方式および TF*IDF 方式で利用された検索クエリの例を示す．表 5 からも分かるように，提案方式では，全体印象語および段落に含まれる単語を利用した検索クエリが抽出できている．たとえば，段落番号 1 では段落から抽出された「花」，段落番号 3 では全体印象語の「春」が利用されるなど，歌詞の全体印象に基づいた検索クエリが抽出できている．

また，9 行目「いつか見た 春の夢」という歌詞に対して，提案方式では「春」が検索クエリとして抽出されていたのに対し，TF*IDF 方式では，「夢」が抽出されていた．TF*IDF 方式ではそのほかに，「別れ」や「勇気」など視覚的に表現することは難しい抽象的な概念を表現する単語が抽出されていた．抽象的な概念を表現する単語を検索クエリとして利用した場合に，投稿者と閲覧者の意図が異なることが多く，被験者が歌詞と適合していないと判断する可能性がある．これに対し，提案方式では抽象的な概念を表す単語を抑制することが可能である．Flickr に投稿された画像群では，風景を構成する物体などを直接的に表現するソーシャルタグが多く付与されており，「夢」や「勇気」などの抽象的な概念がソーシャルタグとして付与されている画像は少ない．したがって，3.2 節に記載した検索クエリ抽出方式により，抽象的な概念を持つ画像を抑制できたと考えられる．さらに抽象的なソーシャルタグは特定のユーザの画像に付与される傾向があり，クエリ選定時に UF 値を考慮することで，歌詞との関連性が高く，より一般的な検索クエリを抽出できる．

5.2 楽曲スライドショーの統一感向上の効果

本節では，楽曲スライドショーの統一感向上効果について分析する．TF*IDF に基づく方式では，重要な単語を抽出することはできているが，全体の統一感が得られる検索クエリが抽出できていないことが分かる．たとえば，行番号 3「泣き疲れて 笑った」および行番号 4「つないだ手と手を 離せないままで」から抽出した検索クエリを利用して得られた画像の例を**表 6** に示す．TF*IDF に基づく方式では，それぞれ「汽車」と「手」という単語が検索クエリとして抽出されており，自動車が題材となる画像に続いて，手の平にハムスターが乗っている画像がスライドショーに利用されていた．このように TF*IDF 方式では，スライドショーに利用される画像群の歌詞に対する関連性は低く，画像遷移において全体の統一感を考慮できていないため，被験者の評価値を下げる要因となったものと考えられる．

一方で提案方式では，行番号 3 では「春」と「花」，行番号 4 では「手」と「花」が検索クエリとして利用されており，行番号 3 では桜並木の画像が，行番号 4 では手の平に花弁が乗っている画像が利用されていた．このように全体印象語や段落番号を考慮して検索クエリを抽出することで歌詞に対する情景に適した画像を検索することが可能である．さらに，スライドショーの全体の統一感の観点においても効果的な検索クエリを抽出することができたといえる．

表 7 に全体印象語と抽象的な概念を表現する単語を Flickr に対する検索クエリとして利用したときの検索結果画像数（IF 値），および検索結果画像群を投稿した総ユーザ数（UF 値）を示す．「春」の IF 値は 36969，UF 値は 677 に対し，「夢」の IF 値は 1402，UF 値は 187 であった．表 7 からも明らかなように，画像に収めることが困難な「夢」や「勇気」などは UF 値が小さい傾向がある．UF 値が小さい検索クエリを利用した場合，視聴者に対して同じ印象を与えることができない可能性があり，被験者の評価値が低くなった要因として考えられる．さらに，提案手法では，行自体に検索クエリとして適する単語が存在しない場合にも，段落や全体印象語より検索クエリを利用することで，全体の雰囲気を損なわない画像を抽出することが可能である．たとえば，表 5 の行番号 3 では，検索クエリとして適した名詞が存在していないが，段落や全体印象語から「春」と「花」という検索クエリを抽出することができている．このことからも提案方式のソーシャルタグを利用した検索クエリ抽出方式の有効性が示された．

5.3 メタノイズによる画像検索精度低下への効果

本節では，ソーシャルタグのメタノイズによる画像検索精度低下に対する効果について考察する．提案方式では，比較方式と検索クエリが同一のものとなった場合にも，候補となる画像群から全体印象語と関連するソーシャルタグの割合が高い画像を選定することで，提案方式では楽曲の雰囲気により適した画像を選定することが可能である．たとえば，表 5 の 7 行目（「窓を開けたら ホラ」）に着目して，「春」と「窓」を検索クエリとして利用した場合を提案方式と，Interesting 指標を用いた画像選定で比較する．**表 8** に提案方式と Interesting 指標によって選定された画

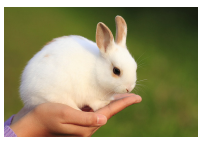
表 6 行番号 3 および 4 から抽出した検索クエリより得られた画像例
Table 6 Sample images obtained by using search queries from line 3 and 4.

行番号	TF*IDF 方式	提案方式
3 行目		
4 行目		

表 7 全体印象語と抽象的な概念を表現する単語（抽象語）の IF 値および，UF 値比較表
Table 7 Comparison table of IF and UF of impression words and conceptual words in Flickr.

属性	単語	IF 値	UF 値
全体印象語	春	36969	677
	夏	24815	589
	秋	89722	709
	冬	27309	515
抽象語	夢	1402	187
	別れ	7	7
	勇気	43	4
	孤独	155	82

表 8 検索クエリ「窓」，「春」を利用した場合の，全体印象語との適合度を利用した画像選定（提案方式）と Interesting 指標を用いた場合の画像とタグ比較表
Table 8 Image and Tag comparison table of image selection method based on proposal and interesting measure. (case of “Window”, ”spring”).

提案方式		桜，日本，花，木，道路，窓，家光，路，春，影，子供，さくら，はな，色，樹，道，町田，陰影，陰翳，machida，japon，couleur，road，light，shadow，house，color，tree，window，fleur，japan，canon，flower，cherry，spring，child，path，ombre，lumiere，maison，enfant
Interesting 指標		春，写真，5 月，手，目，うさぎ，兎，埼玉県，耳，ミニウサギ，指，鼻，キャノン，ペット，ウサギ，日本，緑，twitter，minirabbit，netherlanddwarfrabbit，kuneho，twitpic，miniusagi，updatecollection，mixrabbit，pet，white，macro，rabbit，bunny，green，eye，art，animal，japan，canon，geotagged，nose，photography，eos，spring，interesting，kitten，asia，flickr，colours，hand，image，little，wordpress，finger，small，nail，may，picture，ears，blogger，whiskers，livejournal，tiny，kit，vox，companion，usagi，gettyimages，2010，facebook，friendster，multiply，saitamaken，canoneos7d，

像と付与されたソーシャルタグの例を示す．Interesting 指標を用いた画像選定では，特に天候・季節などを考慮せずに，窓を撮影した画像が選定されていた．このような事例は TF*IDF に基づく方式で得られた多くの画像で確認できた．TF*IDF に基づく方式では，Interesting 指標による画像選定方式を利用しているため，検索クエリのみを考慮した画像が抽出される傾向があった．そのため全体の統一感が損なわれ，unity 評価値が低下したものと考えられる．一方で，提案方式では「窓を含む春」が描写されている画像を選定することができた．さらに表 5 で利用した歌詞行（合計 28 行）を対象に選定された画像のタグ付与数の平均値は，提案方式が 46.3，Interesting 指標では 62.1 であり，タグ付与数が少ない画像を検索する傾向があった．表 8 の提案方式と Interesting 指標によって選定された画像と付与されたソーシャルタグを比較すると，Interesting 指標では Web サービス名や年号，カメラのメーカ名などがノイズとして含まれているのが確認できる．提案方式では，カメラのメーカ名は含まれているものの，画像に含まれている物体がタグとして多く付与されている．このことからも，提案方式では画像の全体印象語に対する適合度を計算することで，ノイズとなるタグを含む画像が選定されるのを抑制している可能性が高い．これらの結果より，全体印象語と関連性の高いタグを持つ画像を優先的に選定することで統一感のある画像群をスライドショーとして利用することができたといえる．

6. まとめ

本研究ではユーザが入力した楽曲に対して，楽曲の歌詞情報を基に Web 画像を検索し，楽曲歌詞と同期させて再生する歌詞連動スライドショー生成システムについて提案した．提案システムでは，既存技術の主な 3 項目の課題，画像と楽曲情景の関連性低下による視聴者に対する違和感，ソーシャルタグのメタノイズによる画像検索精度の低下，画像切替えタイミングによる視聴者に対する違和感について言及し，解決するために各課題について改善案を提案した．具体的には，画像の楽曲情景の関連性を向上させるために，歌詞から推定した全体印象語と歌詞に含まれる単語を検索クエリとして抽出するクエリ抽出方式を提案した．さらに，抽出したクエリを利用して画像共有サイトから表示候補となる画像を取得し，画像に付与されているソーシャルタグと全体印象語の関連度に基づき，より関連性が高いと推定されたソーシャルタグが多く付与されている画像を選定する方式について記述した．また，画像切替えタイミングについては，歌詞行表示の最頻値に基づく制御方式により制御し，視聴者の違和感低減を図った．被験者による評価実験では，歌詞と画像の調和度，画像表示時間の適切さ，スライドショー全体の統一性，スライドショー全体の完成度の 4 項目を設定した．実験結果として，提案方式の主観評価値の平均はすべてにおいて比較方式を上回っており，提案方式の有効性が示された．また，実際に実験で利用した楽曲と画像により，提案システムのソーシャルタグを利用したクエリ抽出方式および，全体印象語との適合度に基づく画像選定方式の有効性を示した．今後は，画像に付与されたノイズタグを抑制する効果の高精度化に加え，歌詞中の形容詞や英詞の考慮，画像切替え時の効果（ズーム・パンなど）など適切な効果を自動で付与する機能について検討を進める．

謝辞 本稿を作成するにあたり，日頃ご指導いただく KDDI 研究所安田豊取締役会長，中島康之代表取締役所長，滝嶋康弘執行役員に深く感謝する．

参考文献

[1] 岩宮眞一郎：オーディオ・ヴィジュアル・メディアによる音楽聴取行動における視覚と聴覚の相互作用，日本音響学会誌，Vol.48, pp.146–153 (1992).
[2] Eerola, T., Lartillot, O. and Toiviainen, P.: Prediction of multidimensional emotional ratings in music from audio using multivariate regression models, *Proc. ISMIR 2009*, pp.621–626 (2009).
[3] Turnbull, D., Barrington, L., Torres, D. and Lanckriet, G.: Towards Musical Query-by-Semantic-Description using the CAL500 Data Set, *Proc. 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.439–446 (2007).
[4] Bertin-Mahieux, T., Eck, D., Maillet, F. and Lamere, P.: Autotagger: A model for Predicting social tags from acoustic features on large music databases, *Journal of New Music Research, special issue: From genres to tags: Music Information Retrieval in the era of folksonomies*, Vol.37, No.2 pp.151–165 (2008).
[5] Terada, T., Tsukamoto, M. and Nishino, S.: A System for Presenting Background Scenes of Karaoke Using an Active Database System, *Proc. ISCA 18th International Conference on Computers and Their Applications*, pp.160–165 (2003).
[6] Hua, X.-S., Lu, L. and Zhang, H.-J.: P-Karaoke: Personalized Karaoke System, *Proc. 12th Annual ACM International Conference on Multimedia*, pp.172–173 (2004).
[7] Xu, S., Jin, T. and Lau, F.C.M.: Automatic Generation of Music Slide Show using Personal Photos, *Proc. 10th IEEE International Symposium on Multimedia*, pp.214–219 (2008).
[8] Shamma, D.A., Pardo, B. and Hammond, K.J.: MusicStory: a Personalized Music Video Creator, *Proc. 13th Annual ACM International Conference on Multimedia*, pp.563–566 (2005).
[9] Cai, R., Zhang, L., Jing, F., Lai, W. and Ma, W.-Y: Automated Music Video Generation Using Web Image Resource, *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, pp.737–740 (2007).
[10] Shoji, N. and Miura, M.: Slideshow System That Automatically Switches Photographs Based on a Musical Acoustic Signal, *Proc. 20th International Congress on Acoustics*, p.550 (2010).
[11] Wu, H., Zubair, M. and Maly, K.: Harvesting social knowledge from folksonomies, *Proc. 17th Conference on Hypertext and Hypermedia*, pp.111–114, ACM (2006).
[12] Gouyon, F., Klapuri, A., Dixon, S., Alonso, M.,

Tzanetakis, G., Uhle, C. and Cano, P.: An Experimental Comparison of Audio Tempo Induction Algorithms, *IEEE Trans. Audio, Speech, and Language Processing*, pp.1832–1844 (Sep. 2006).
[13] Cortes, C. and Vapnik, V.: Support Vector Networks, *Machine Learning*, Vol.20, pp.273–297 (1995).
[14] 梶 克彦，長尾 確：楽曲に対する多様な解釈を扱う音楽アノテーションシステム，情報処理学会論文誌，Vol.48, No.1, pp.258–273 (2007).
[15] Hevner, K.: Experimental studies of the elements of expression in music, *American Journal of Psychology*, Vol.48, pp.246–68 (1936).
[16] Koppel, M., Argamon, S. and Shimoni A.R.: Automatically Categorizing Written Texts by Author Gender, *Library and Linguistic Computing*, Vol.17, No.4 (2003).
[17] Corney, M., De Vel, O., Anderson, A. and Mohay, G.: Gender-preferential Text Mining of E-Mail Discource, *18th Annual Computer Security Applications Conference* (2002).
[18] Dixon, S.: Evaluation of the Audio Beat Tracking System BeatRoot, *Journal of New Music Research*, Vol.36, pp.39–50 (2007).

石先 広海 （正会員）

2004 年早稲田大学理工学部電子情報通信学科卒業．2006 年同大学大学院修士課程修了．同年 KDDI 株式会社入社，2011～2012 年インディアナ大学客員研究員．現在，株式会社 KDDI 研究所知能メディアグループ研究主査．この間，マルチメディア情報検索，音楽情報処理，CMC 等の研究に従事．FIT2009 ヤングリサーチャー賞受賞．電子情報通信学会会員．

舟澤 慎太郎 （正会員）

2007 年早稲田大学理工学部コンピュータネットワーク工学科卒業．2009 年同大学大学院修士課程修了．現在，株式会社野村総合研究所勤務．在学中は音楽情報処理の研究に従事．

帆足 啓一郎 （正会員）

1995 年早稲田大学理工学部情報学科卒業．1997 年同大学大学院修士課程修了．同年国際電信電話株式会社（現，KDDI 株式会社）入社，現在，株式会社 KDDI 研究所アプリケーションプラットフォームグループグループリーダー．この間，マルチメディア情報検索等の研究に従事．2001～2005 年早稲田大学メディアネットワークセンター非常勤講師．工学博士．FIT2004 ヤングリサーチャー賞受賞．電子情報通信学会，ACM 各会員．

小野 智弘 （正会員）

1992 年慶応義塾大学理工学部電気工学科卒業．1994 年同大学大学院理工学研究科修士課程計算機科学専攻修了，同年国際電信電話株式会社入社．1999 年 9 月～2000 年 9 月スタンフォード大学電気工学科客員研究員．現在，株式会社 KDDI 研究所知能メディアグループグループリーダー．この間，利用者嗜好抽出，ソフトウェアエージェント，データベース等の研究に従事．情報処理学会全国大会学術奨励賞受賞．電子情報通信学会会員．博士（工学）．

甲藤 二郎 （正会員）

1987 年東京大学工学部電気工学科卒業．1992 年同大学大学院博士課程修了．同年日本電気株式会社入社．1996～1997 年米国プリンストン大学客員研究員．1999 年早稲田大学理工学部助教授．2004 年早稲田大学理工学部教授．2007 年早稲田大学基幹理工学部教授．主にマルチメディア通信，信号処理の研究に従事．工学博士．電子情報通信学会，IEEE，ACM 各会員．