Vol.2011-AVM-73 No.9
2011/7/15

歌詞情報に基づくWEB画像検索を利用した楽曲連動スライドショー生成システム

石先　広海[†1]　帆足　啓一郎[†1]　小野　智弘[†1]

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検索したWEB画像を，楽曲と同期させて再生するスライドショーを自動生成する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，楽曲が表現する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，歌詞に含まれる単語と，歌詞情報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，行と連動して表示させる写真を選定する．さらに，歌詞行の表示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．

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

Hiromi Ishizaki,[†1] Keiichiro Hoashi [†1]
and Chihiro Ono[†1]

In this paper, we propose a system to automatically generate slide shows for music user selected, by utilizing Web images retrived by queries constructed from song lyrics. Proposed system aims to provide new and impressive user experiences with using Web images synchronized with lines of lyric. First, we propose a method to select images to compose the slideshow from the result of Web image retrieval based on query extracted from lyrics. The system selects matching images with the whole impression of lyrics and removes images which have many social tags related not to the image content. Furthermore, we propose a method to switch the images in the slides. Finally, subjective experiment is conducted to evaluate effectiveness of our system.

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，その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)．これにより，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製作する場合，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，映像作品の製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．したがって，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製作することは困難である．

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自動生成するシステムを提案する．本システムでは，画像共有サイトの画像（Web画像）を利用して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・歌詞と同期させて表示する．具体的には，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方法，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．

2. 関連研究

楽曲と画像を連動させたコンテンツを生成する研究として，ユーザ自身が撮影した写真や映像を利用してスライドショーなどのコンテンツを作成するシステム2)–4)が提案されて

†1 KDDI 研究所
KDDI R&D Laboratories

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自動で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高品質なコンテンツを生成するために多くの量の素材を準備する必要があり，ユーザ自身が画像や映像を大量に撮影・準備する必要があるなど，一般ユーザにとっては敷居が高いものであった．

その他に，楽曲の歌詞情報を利用して 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自動で生成するシステムが提案されている．Shamma らは，一般的な Web 検索エンジンや画像共有サイトから，歌詞に出現する単語を検索クエリとして用いることで画像を取得し，音楽の音量の急激な変化を検出することで楽曲テンポを推定し，テンポ情報に基づいて楽曲と画像を同期させて再生するシステム 5) を提案している．また，Cai らは歌詞に出現する単語のうち名詞，名詞句，人名，地名を抽出し，Web 検索エンジンの検索クエリとして用いる．検索結果として得られた画像群から，顔検出を適用し，人物の顔を含む風景画像などを優先的に選別することで，最終的に楽曲全体のムードに適した画像を選定するシステム 6) を提案している．

3. 問 題 点

前述のように，楽曲と画像を連動させたコンテンツを自動生成するシステムは多く研究されている．しかし，これらのシステムでは主に二項目の課題がある．一つ目の課題として，スライドショーを構成する画像の質が低い，もしくは楽曲との関連性が低いことが挙げられる．これらの研究では，歌詞中に含まれている単語を利用して素材となる画像を Web 上から取得し，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．一方で，検索の際に利用するクエリの選定という観点では，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．そのため，歌詞に出現する単語が一般的で，楽曲の特徴を表現する単語ではない場合に，楽曲の雰囲気に適さない画像が表示される可能性がある．たとえば,「今」や「誰」などの単語は歌詞中での出現頻度は高いが，これらの単語を検索クエリとして画像を検索しても，意味のある画像が取得できるとは限らず，作成したスライドショーの品質を低下させる可能性がある．

また，スライドショーの品質低下の主な要因の一つとして，画像共有サイトに投稿された画像に対して意味のないタグが多く付与されていることが挙げられる．たとえば，Filckr[⋆1]では画像に付与されたタグに基づいて画像を検索することができる．このような画像共有サイトでは，ユーザが投稿した画像にタグを付与することができ，他のユーザの画像に対してもタグを付与することが可能である．そのため，画像の中には膨大なタグが付与されているものもあり，検索インデクスを目的としたタグなど，画像に対して関連性の低いタグが付与されているものも多く存在している．このようなタグはメタノイズとよばれ 7)，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．関連研究では，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．

もう二つ目の課題として，スライドショーの画像切り替えタイミングの制御が挙げられる．既存の方式 5) では，楽曲のテンポを自動で推定し，楽曲テンポに合わせて画像を切り替えている。しかし，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という共通の問題があるなど精度の問題がある 8)．これにより，推定結果が誤っていた場合に，本来はゆったりとした楽曲に対して，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，画像の表示時間が極端に短い場合や，長い場合が発生し，楽曲が表現する情景と適合しない可能性がある．

4. 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利用する Web 画像の質を高めるためのクエリおよび画像選定方式と，各画像の表示時間を自動制御する方式を用いた楽曲スライドショー生成システムを提案する．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，楽曲の歌詞から受ける全体的な印象を推定し，各歌詞行から検索クエリとなるキーワード群を抽出する．抽出した検索クエリにより，各行について Web 上の画像共有サイトから候補となる画像群を取得する．取得した画像群から歌詞行に適した画像を一枚選定し，画像と同期させて再生する．以下に各処理の詳細を説明する．

4.1 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るために，楽曲歌詞から全体印象を推定する．楽曲が表現している状況を表現する印象カテゴリを事前に設定し，これらのカテゴリに対する楽曲分類を適用することで全体印象語を付与する．具体的には，被験者から収集した歌詞に対する印象情報を用いて Support Vector Machine 9)(以下 SVM) により，入力さ

⋆1 http://www.flickr.com

Vol.2011-AVM-73 No.9
2011/7/15

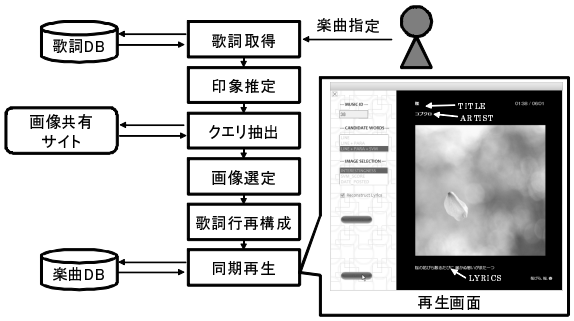

図 1 提案システムの処理の流れと画面イメージ
Fig. 1 Conceptual illustration of procedures of the proposal

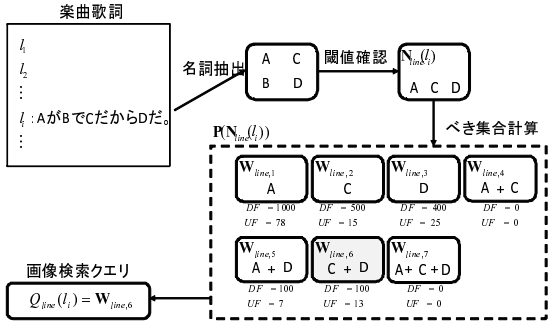

図 2 ソーシャルタグを利用したクエリ抽出の流れイメージ図
Fig. 2 Conceptual illustration of query extraction based on social tag.

れた歌詞情報の印象カテゴリを判別する分類器を構築する．SVM は多次元ベクトルで表現されたオブジェクトを分類する手法で，テキスト分類の分野でも広く利用されている．最終的に，楽曲データベースの全ての楽曲に対して分類器を適用し，分類された季節，時間帯，天候の三種類のカテゴリのラベル（**表 1**）を全体印象語として付与する．

4.2 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,「多くのユーザが付与した単語はソーシャルタグとして重要な意味を持つ単語であり，歌詞を表現する画像として適した画像を検索することができる.」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を解析し，タグの出現頻度に基づいて検索クエリを決定する．これにより，歌詞から各行の情景表現に有効な画像群を取得できる検索クエリを抽出する．ソーシャルタグを利用したクエリ抽出の流れを**図 2** に示す．

表 1 全体印象に用いるカテゴリラベル
Table 1 Concepts and category labels for describing general impression of music.

概念	印象ラベル
季節	春，夏，秋，冬
時間帯	朝，昼，夕方，夜
天候	晴れ，曇り，雨，雪，虹

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数を $DF(W)$(Document Frequency)，ユニークな投稿者数を $UF(W)$(User Frequency) と表現する．歌詞の i 行目 (l_i) で使用されている名詞（群）を抽出する．抽出した名詞に対して，事前に設定した閾値を満たさない名詞を排除し，クエリ候補とする（$N_{line}(l_i)$）．尚，閾値は経験的に $DF > 40$，$UF > 10$ と設定した．クエリ候補 $N_{line}(l_i)$ から得られる，べき集合 $P(N_{line}(l_i)) = \{W_{line,1}, W_{line,2}, \cdots, W_{line,x}\}$ を計算する．得られたべき集合の中で，$DF(W_{max}) \neq 0$ かつ，$|W_{max}|$ が最大となる集合 W_{max} を実際の検索クエリ $Q_{line}(l_i)$ として利用する．なお，W_{max} が複数存在する場合には，$UF(W_{max})$ が最大となる W_{max} を検索クエリとする．

さらに，検索クエリの単語数を拡張するために，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 $P'(N_{para}(l_i)) = \{W_{para,1} + Q_{line}(l_i), W_{para,2} + Q_{line}(l_i), \cdots, W_{para,y} + Q_{line}(l_i)\} = \{W'_{para,1}, W'_{para,2}, \cdots, W'_{para,y}\}$ 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．最終的に全ての行に対して，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．尚，$Q'_{line}(l_i)$ が存在しなかった場合には，楽曲 m に付与された印象ラベル $N_{all}(m)$ を検索クエリとして用いる．

4.3 全体印象語との適合度を利用した画像選定

本処理では，各候補画像に対して，全体印象語との適合度を利用することで，各行と同期して再生する画像を候補画像の中から選定する．全体印象語との適合度は，一枚の画像に付与されている全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．全体印象語との関連度を利用することで，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，入力楽曲の全体印象語と関連が強い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共起情報をもとに算出する．二つのタグが同一の画像に付与されたタグ同士は何らかの関係性を持つと考え，共起情報をタグ間の関連の度合いを表す指標として利用する．たとえば，歌詞に付与された春という印象ラベルとの共起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印象ラベル（夏，秋，冬）との共起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，春との共起確率が高く，同じ概念に属する他の印象ラベルとの共起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，共起確率を UF を用いて計算し，以下のように表現する．

$$Co(t, n_{all}) = \frac{UF(t \cap n_{all})}{UF(n_{all})} \tag{1}$$

これを利用して，ソーシャルタグ t と全体印象語 n_{all} の関連度 $R(t, n_{all})$ を以下のように表す．

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

ここで，C は同じ全体印象語 n_{all} に属する全ての印象ラベルを表す．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) 中の第二項による影響を調整するための係数で，経験的に 3 と設定した．最終的に，関連度が 0.024 を超え，$UF(t) \geq 5$ を満たすタグを印象ラベルに対する関連タグとして判定する．

画像に付与された関連タグの関連度を用いて，画像と全体印象の適合度を計算する．適合度は，楽曲に付与された全体印象語との関連が高いタグが多く付与されるほど大きい値を示す．以下に適合度の計算式を示す．

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

ここで，T_i は画像 i に付与されているタグを表し，$T_{related}(n_{all})$ は印象ラベル n_{all} に対する関連タグとする．適合度は全ての歌詞行の候補画像に対して計算し，各行において適合度が最大となる画像を選定する式 (3) により，ノイズとなる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．

4.4 行再生時間の最頻値を利用した歌詞行再構成

本処理では，スライドショーにおいて各画像を表示する時間を適切にするために，画像切り替えの単位となる歌詞行の再構成を行う．具体的には，表示時間の短い行は周辺の行と結合し，表示時間の長い行は分割することで画像切替えを制御する．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り替えタイミングも設定する．以下に手順の詳細を示す．なお，表示時間が短い，又は，長いと判定する閾値は，それぞれ 4[sec], 12[sec] と経験的に設定した．

(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 4[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 4[sec] 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．但し，結合は同段落に属する行同士でのみ行う．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が 1 枚検索される．
(4) 表示時間が 12[sec] 以上の行を等分割する．但し，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．

最終的に選定した画像を行と連動させた楽曲スライドショーを自動で再生する．

5. 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．本実験では，被験者 42 名を対象とし，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

評価を付与した．実験では市販されている J-POP（10 曲）を利用し，全ての楽曲に対して歌詞と同期情報を手動で付与した．このとき，4.4 節に記載の歌詞再構築方法の有効性を検証するために，歌詞行統合処理が実施される楽曲を 5 曲（統合セット），歌詞行分割処理が実施される楽曲を 5 曲（分割セット）を利用した．一曲に対して 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を分割して評価を実施した．本実験では，画像の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[⋆1]を利用した．

評価項目は，歌詞と画像の調和度合い（content），画像切替えの適切さ（transition），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 4 項目を設定し，被験者により 5 段階（５：とても良い〜１：とても悪い）で評価を付与してもらった．

5.1 評価システム詳細

本節では，評価実験で使用した比較対象となる二つの方式の詳細を説明する．まず一つ目の比較方式として，MusicStory5) を利用した．MusicStory では，歌詞中に含まれる全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）により画像を切替えて表示させた．

二つ目の比較方式として，TF*IDF によるクエリ選定 (TF*IDF 方式) を利用した方式を利用した．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．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．最終的に，得られた画像において，検索結果のランキングが最上位となる画像を一枚選定し，歌詞行と表示させた．本稿では，TF*IDF における DF は 3062 曲の J-POP 楽曲の歌詞を利用して計算した．

5.2 実 験 結 果

被験者から収集した，全ての方式に対する 5 段階評価値の平均値を**図 3** に示す．図 3 から明らかな通り，比較方式に比べて提案方式の主観評価値は全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 の項目は，content，transition，unity の全ての要素が影響する項目であると考えられるが，提案方式は比較方式に比べて高い平均値

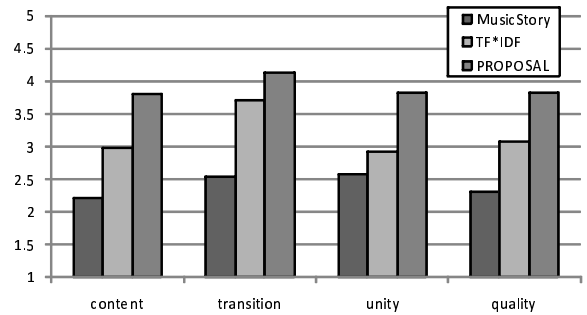

図 3 被験者による 5 段階評価値の平均（content，transition，unity，quality）
Fig. 3 Result of user evaluation: average of user ratings (content, transition, unity, quality))

を得ることができている．t 検定によって，評価値平均の差は優位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高品質な楽曲スライドショーの作成に有効であるといえる．

次に，クエリ選択の有効性を分析するために，TF*IDF に基づく方式と，提案方式それぞれで検索クエリとして利用された単語を比較した．提案方式は「街」や「笑顔」など，視覚的に表現可能な単語を抽出することができている．一方で，TF*IDF に基づく方式では，重要語を抽出することはできているが,「別れ」や「心」など視覚的に表現することは難しい単語が抽出されていた．これは，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タグが多く付与されており,「別れ」や「心」などの抽象的な概念がタグとして付与されている画像は少ないことが理由として考えられる．実際に,「別れ」という単語の DF 値は 7 となっており,「街」の DF 値は 9342 であった．また，抽象的なタグは特定のユーザの画像に付与されていることが多いため，クエリ選定時に UF 特徴を考慮することで，提案方式の検索クエリが良質となったと考えられる．たとえば,「午後」の DF 値は 31，UF 値は 15 であったのに対し,「勇気」では DF 値は 41，UF 値は 2 であった．これは勇気というタグが 2 ユーザのみからしか使用されていないことを意味しており，午後という単語の方がより一般的に利用されていることが予想できる．さらに，提案手法では，行自体に検索クエリとなる単語が存在しない場合にも，段落や全体印象ラベル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．

さらに，提案方式と比較方式で，検索クエリが同一のものとなった場合に，提案方式で

⋆1 MeCab:Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/

は候補となる画像群から，全体印象ラベル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，より楽曲の雰囲気に適した画像を選定することが可能である．たとえば，「冬」という印象ラベルが付与された楽曲において,「街」を検索クエリとして利用した場合，TF*IDF 方式では，特に天候・季節などを考慮せずに，街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 方式で得られた多くの画像で確認できた．全体の統一感を損なわせたため，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「雪が降る街」が描写されている画像を選定することができた．これは，画像の全体印象ラベルに対する適合度を計算することで，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ためである．このことから，適合度を利用した画像選定方式が有効であるといえる．

さらに，transition 項目について述べると，提案方式は，比較方式に比べて高い評価値を得ることができた．TF*IDF 方式と比べると 9 楽曲の評価値平均を上回っていた．結合セットでは 0.21 の差が，分割セットでは 0.61 の差があった．このことは，提案方式が特にゆっくりとした楽曲に対して有効であることを示している．実際に，MusicStory では BPM が正解値の 2 倍と推定された楽曲は 8 楽曲存在しており，テンポの遅い楽曲に対して，画像の切替えが頻繁に発生していた．また，TF*IDF 方式で生成した楽曲スライドショーでは，歌詞行の長さに応じて楽曲の一枚当たりの画像表示時間が長くなり，被験者を飽きさせたことが原因と考えられる．このことから提案方式の歌詞行再構築方法が楽曲スライドショーの品質を向上させることに有効であるといえる．

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検索した WEB 画像を，楽曲と同期させて再生する楽曲スライドショーを自動生成するシステムについて提案した．楽曲歌詞の内容に適した Web 画像を取得し，効果的に楽曲と同期させて再生するために，歌詞から全体印象ラベルを推定し，歌詞から WEB 画像を検索するためのクエリ選定方式を利用した．さらに，選定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いて，行と連動して表示させる画像を選定する選定方式を提案した．最終的に楽曲スライドショー再生時の画像切替えタイミングを，歌詞行表示の最頻値に基づく制御方式により制御し，自動で楽曲スライドショーを再生する．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四項目を設定した．実験結果として，提案方式の主観評価値の平均は全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．今後は，歌詞中の形容詞や，英詞の考慮に加え，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．

7. 謝　　辞

本論文を作成するにあたり，実験に協力いただいた早稲田大学，舟澤慎太郎氏，ご討論いただいた早稲田大学，甲藤二郎教授に感謝する．また，日頃ご指導いただく KDDI 研究所，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，48 巻，pp146–153 (1992)
2) Terada, T. Tsukamoto, M. and Nishino, S.: A System for Presenting Background Scenes of Karaoke Using an Active Database System, Proceedings of the ISCA 18th International Conference on Computers and Their Applications, pp.160–165 (2003)
3) Hua, X.-S. Lu, L. and Zhang, H.-J.: P-Karaoke: Personalized Karaoke System, Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 172–173 (2004)
4) Xu, S. Jin, T. and Lau, F. C. M.: Automatic Generation of Music Slide Show using Personal Photos, Proceedings of 10th IEEE International Symposium on Multimedia, pp.214–219 (2008)
5) Shamma, D. A. Pardo, B. and Hammond, K. J.: MusicStory: a Personalized Music Video Creator, Proceedings of the 13th Annual ACM International Conference on Multimedia, pp.563–566 (2005)
6) Cai, R. Zhang, L. Jing, F. Lai, W. and Ma, W.-Y: Automated Music Video Generation Using Web Image Resource, Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, pp.737–740 (2007)
7) Wu, H. Zubair,M and Maly, K.:Harvesting social knowledge from folksonomies, Proceedings of the seventeenth conference on Hypertext and hypermedia, ACM, pp.111–114 (2006)
8) Gouyon, F. Klapuri, A. Dixon, S. Alonso, M. Tzanetakis, G. Uhle, C. and Cano, P.: An Experimental Comparison of Audio Tempo Induction Algorithms, IEEE Transaction on Audio, Speech, and Language Processing, IEEE Transactions on. Sept. 2006 , pp.1832–1844 (2006)
9) Cortes, C. and Vapnik, V.: Support Vector Networks, Machine Learning, Vol. 20, pp.273–297 (1995)