平成 21 年度　修士論文

歌詞情報を利用した
Web 画像・楽曲連動スライドショー
自動生成システム

早稲田大学大学院基幹理工学研究科情報理工学専攻

5108B111-2　　舟澤 慎太郎

指導　甲藤 二郎　教授

2010 年 2 月 5 日

指導教授印	受付印

目次

第1章　序論

本章では，本研究を行うにあたっての背景，目的，そして，本論文の構成について述べる．

1.1. 研究背景

映画やテレビ番組，ミュージックビデオなどの作品では，映像と音楽を効果的に組み合わせることで，その作品の価値を高めている．例えば，テレビドラマの別れのシーンに BGM として悲しい音楽を流したり，ミュージックビデオにおいて歌詞のストーリーに沿った映像を提示したりすることで，それらを単独で鑑賞する以上のインパクトを生み出すことができる．このような視覚と聴覚を刺激する効果は，心理学の分野でも視覚と聴覚の共鳴現象として報告されている[1]．この効果を，普段の音楽鑑賞に適用することで，より印象深い音楽体験が実現できると考えられる．しかし，音楽に合った映像などの視覚的コンテンツを制作するとなると，構成する素材の収集や選択，構成法の考慮など，様々な作業が必要となる．したがって，このような作品の制作に慣れていないユーザが，個人で所有する楽曲を対象として，それに合う視覚的コンテンツを制作するには，多大な労力を必要とする．このような背景から，視覚と聴覚の共鳴現象を利用した音楽の楽しみ方を，誰でも手軽に実現するために，自動で楽曲に合った視覚的コンテンツを生成するシステムが望まれる．

1.2. 研究目的

本研究では，視覚と聴覚の共鳴現象を利用した音楽の鑑賞法を，誰でも手軽に楽しめるようにするために，ユーザが指定した楽曲に対して，その楽曲に合うスライドショーを自動で生成するシステムを提案する．

スライドショーを構成する素材は，楽曲の歌詞情報を基に検索した Web 画像を用いる．Web 画像は非常に豊富で多様性のある素材であり，かつ，手軽に入手可能であるため，Web 画像を用いることで，スライドショーを生成するための素材を，ユーザ自らが収集する手間を省くことができる．また，歌詞情報は，楽曲の内容を直接的に表現する特徴であるため，それを基に画像の検索を行うことで，より楽曲の内容に即したスライドショーを構成することができると考えられる．このように，楽曲に合ったスライドショーを自動で生成し，楽曲と同期再生することで，手軽に印象深い音楽体験を実現することができる．なお

本論文では，本システムにより生成した楽曲の付与されたスライドショーを，“楽曲スライドショー”と呼ぶ.

1.3. 本論文の概要

本論文ではまず，第 2 章にて関連研究として，スライドショー作成支援ツールや，ミュージックビデオ自動生成 Web サービス，および，従来において提案されているミュージックビデオ自動生成システムについて述べる．第 3 章では，我々の提案する楽曲スライドショー自動生成システムの概要について述べ，その中の，スライドショーを構成する要素の候補となる画像を取得する処理における，Web 画像検索に与えるクエリの選定法を 2 つ提案する．そして第 4 章において，それら 2 つの手法の有効性を検証するための被験者評価実験について述べる．第 5 章では，第 4 章の被験者評価実験により明らかになったシステムの問題点に対して，画像検索結果から最終的に用いる画像を選定する手法と，スライドショーにおける画像を切り替えるタイミングの改善手法を提案し, 第 6 章にて前者の手法, 第 7 章にて後者の手法の有効性を評価実験により検証する．最後にまとめとして，第 8 章で結論を述べ，本論文を締めくくる．

第2章　関連研究

本章では，関連研究として，スライドショーの生成を支援するツールとミュージックビデオ生成 Web サービス，および，過去の研究において提案されたミュージックビデオ自動生成システムについて説明し，それらの問題点について言及する．

2.1. スライドショー作成支援ツール

音楽付きのスライドショーの作成を支援するツールが提供されている[2][3].

[2]は，Microsoft 社により無償で提供されている“Photo Story”と呼ばれるツールである（図 2.1.1.）．このツールでは，スライドショーを構成する要素として，ローカルに保存してある画像を直接指定することで，WMV 形式のスライドショーを生成することができる．また，各画像の表示時間，画像切り替えの際の効果，BGM として付与する音楽を指定することも可能である．

図 2.1.1.　Photo Story [2]

一方[3]は，“Photo Flash Maker”と呼ばれるツールで，Anvsoft 社により提供されており，機能限定版であれば無料で利用することができる（図 2.1.2.）．このツールも“Photo Story”と同様，スライドショーを構成する画像と BGM として付与する音楽ファイルを指定することで，SWF 形式の楽曲スライドショーFlash 動画を生成できる．さらに，画像の周りにフレームとして付与する数多くのテーマがあらかじめ用意されており，効果的にスライドショーの飾り付けを行うことができる．

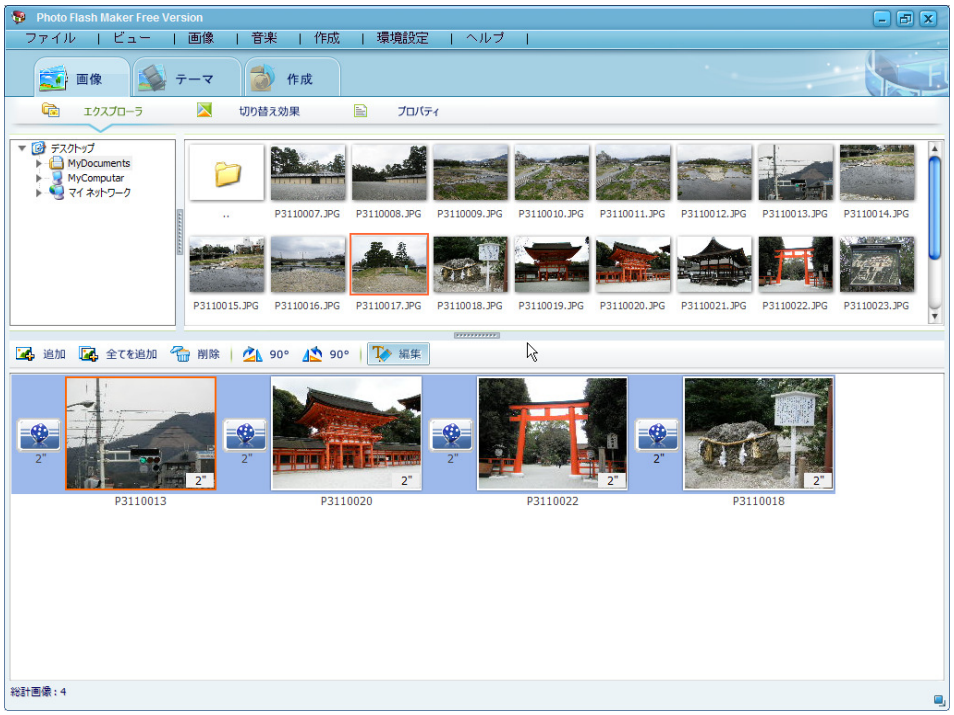

図 2.1.2.　Photo Flash Maker [3]

上記のようなスライドショー作成支援ツールは，あくまでスライドショーが主体であり，音楽は BGM として補助的な役割を果たすため，音楽を主体として扱う我々の研究とは目的が異なっている．また，これらのツールでは，使用する画像素材の選択から，画像を切り替えるタイミングの設定まで，全てユーザが手動で行うため，思い通りのスライドショーが作成できるメリットがある一方で，それらを細かく設定するために多くの労力が必要となる．

2.2. ミュージックビデオ自動生成 Web サービス

構成要素として用いる画像や映像と，同時に再生する音楽を指定するだけで，ミュージックビデオを生成する Web サービスがある[4][5].

[4]は, “shwup”と呼ばれる Web サービスであり, 構成素材となる画像, BGM となる音楽, 画面デザインが入力として指定されると，それを基に生成したミュージックビデオを出力する. 使用する画像は, ローカルファイルをアップロードする他に, 写真共有サイト Flickr[6]や写真管理ソフトウェア Picasa[7]のウェブアルバムなどから指定することができる. このサービスでは，単に画像を表示して切り替えるだけでなく，画像切り替わり時においてアニメーションを加えたり，画像表示時にズームやパンなどの効果を加えたりすることで，スライドショーの質を高めている.

一方[5]は, “animoto”と呼ばれるサービスであり, “shwup”と同様の方法で画像（映像も可）と音楽を指定することで，ミュージックビデオを生成する．“shwup”との大きな相違点は以下の 2 点である．まず 1 点目が，指定された音楽のリズムを解析し，それに合わせて画像や映像の切り替えを行う点である．こうすることで，より音楽の曲調に合ったミュージックビデオを生成することができる．そしてもう 1 点は，画像や映像の提示法が多彩な点である. “shwup”では, 画像を 1 枚ずつ表示していくことでスライドショーを構成しているが, これに対し“animoto”では，一度に表示する枚数や画像のサイズを固定しておらず，複数の画面に分割して複数枚の画像を表示したり，画像のサイズを拡大や縮小したりすることで, より芸術性の高いコンテンツを生成している．例えば，図 2.2.1.では，背景に明度を落としたサイズの大きい画像が 1 枚と，手前にサイズの小さい画像が 2 枚表示されている．さらに，これらの画像に回転の効果を加えることで，より立体的なミュージックビデオを生成している.

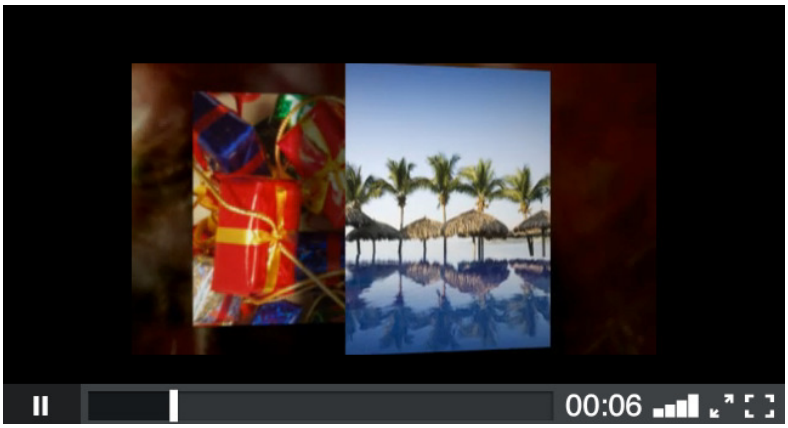

図 2.2.1.　animoto にて生成したミュージックビデオ [5]

2.3. ミュージックビデオ自動生成システム

2.3.1. 個人で所有する画像・映像を構成素材として利用する研究

個人で所有する画像や映像を素材として利用することでミュージックビデオを自動生成するシステムが提案されている[8][9][10].

[8]では，個人所有画像・映像を用いたカラオケ背景映像生成システム P-Karaoke を提案している．このシステムでは，画像や映像のコンテンツ自体を解析し，映像における重要場面の抽出[11]や，画質の低い画像を排除することで，用いる素材の選択を行っている．しかし，用いる素材の選定において，画像や映像に映っているオブジェクトは考慮していない．

同様に，カラオケ背景映像を自動生成する研究として[9]がある．[9]では，あらかじめ個人所有画像にその内容を表現するキーワードが付与されていることを前提として，それらキーワードと歌詞に出現する単語を対応付けることで画像を選択し，カラオケの背景映像を生成している．さらに，バラードの楽曲であれば画像をセピア色に変換する，などのルールをあらかじめ定義しておくことで，様々なエフェクトを用いた映像を生成している．

これに対し，個人所有画像にキーワードが付与されている，という前提を解消した研究が[10]にて提案されている．この研究では，歌詞に出現する単語をクエリとして取得した Web 画像検索結果との類似度の高い個人所有画像を用いることでミュージックビデオを生成している．このミュージックビデオでは，楽曲のリズムを解析して抽出した keyframe 単位で画像を切り替え表示している．

2.3.2. Web 画像を構成素材として利用する研究

Web 画像を素材として利用することでミュージックビデオを自動生成するシステムが提案されている[12][13].

[12]にて提案されている“MusicStory”では，楽曲の歌詞に出現する単語を画像検索クエリとして用い，一般的な Web 画像検索エンジンや写真共有サイトから取得した Web 画像を素材としてミュージックビデオを生成している. このシステムでは, 入力楽曲の BPM を基に，“Slow”，“Medium”，“Fast”を判定し，それぞれに設定されている画像 1 枚あたりの表示時間と画像切り替え時のフェード時間に従ってミュージックビデオを構成する．

[13]では，歌詞に出現する単語により Web から画像を検索し，それらに対して顔検出[14]や indoor/outdoor 判定[15]を行うことにより，人の顔の写っている割合が大きい写真，かつ，外で撮影された写真を優先してミュージックビデオを構成する候補として選択する．また，ミュージックビデオ全体の統一感を表現するために，画像の色特徴と楽曲のムード[16]を対応付けることで，楽曲全体のムードに適した画像を最終的に選出している．そして，選出した画像群に対して，Photo2Video[17]と呼ばれる技術により，パンやズームなどを効果的に

加えることでミュージックビデオを構築している．なお，この研究では，歌詞に出現する単語の中でも，名詞，名詞句，人名，地名が重要であるとし，これらを検索クエリとして用いている．

2.3.3. 課題

2.3.1.節にて述べたような，個人で所有している画像や映像をミュージックビデオの構成要素として利用する研究では，ユーザの身近な画像や映像を用いることができるというメリットがある一方で，十分な数の素材を所持していない場合，限られた中から素材を選択する必要があるため，楽曲に合った作品を生成することは難しいと考えられる．本研究では，視聴者の視覚と聴覚に働きかけ，より強い印象を表現することを目的としているため，楽曲に合ったスライドショーを生成することは必須条件である．そのため，スライドショーを構成する素材として，利用できる数に制限のある個人所有画像・映像ではなく，利用できる数が豊富で多様性に富んだ素材である Web 画像を用いる．

また，2.3.2.節にて述べたように，歌詞情報を基に検索した Web 画像を用いてミュージックビデオを生成する研究が既に提案されている．これらの研究では主に，Web から検索した画像群から最適なものを選定する処理に着目しており，検索に与えるクエリの選定には，stop word を排除する，品詞を特定のものに限定する，という処理しか行っていない．しかし，歌詞に出現する単語にも，画像検索に有用なものとそうでないものが存在する．例えば，“今”や“誰”という単語は歌詞において比較的出現頻度の高い単語であるが，これらをクエリとして画像検索を行った場合，その単語を的確に表現する画像を取得することは難しい．本論文ではまずこの点に着目し，画像検索に与えるクエリの選定法を提案する．

第3章　提案手法

本章では，提案システムの概要について述べ，さらにその中の，スライドショーに用いる候補となる画像を検索する処理について説明する．

3.1. システム概要

提案システムのインターフェースを図 3.1.1.に，処理概要図を図 3.1.2.に示す．本システムでは，入力としてユーザが指定した楽曲に対し，その歌詞の行ごとに画像を 1 枚 Web から検索することでスライドショーを構成する．そして，楽曲と共にスライドショーを再生しユーザに出力する．なお，前提条件として，データベース内には，楽曲の音源，楽曲の歌詞情報，楽曲と歌詞の同期情報を有しているとする．楽曲と歌詞の同期情報とは，歌詞の各行のフレーズの楽曲中における開始時刻と終了時刻を記した情報であり，図 3.1.3.のような形でテキストデータとして所持している．

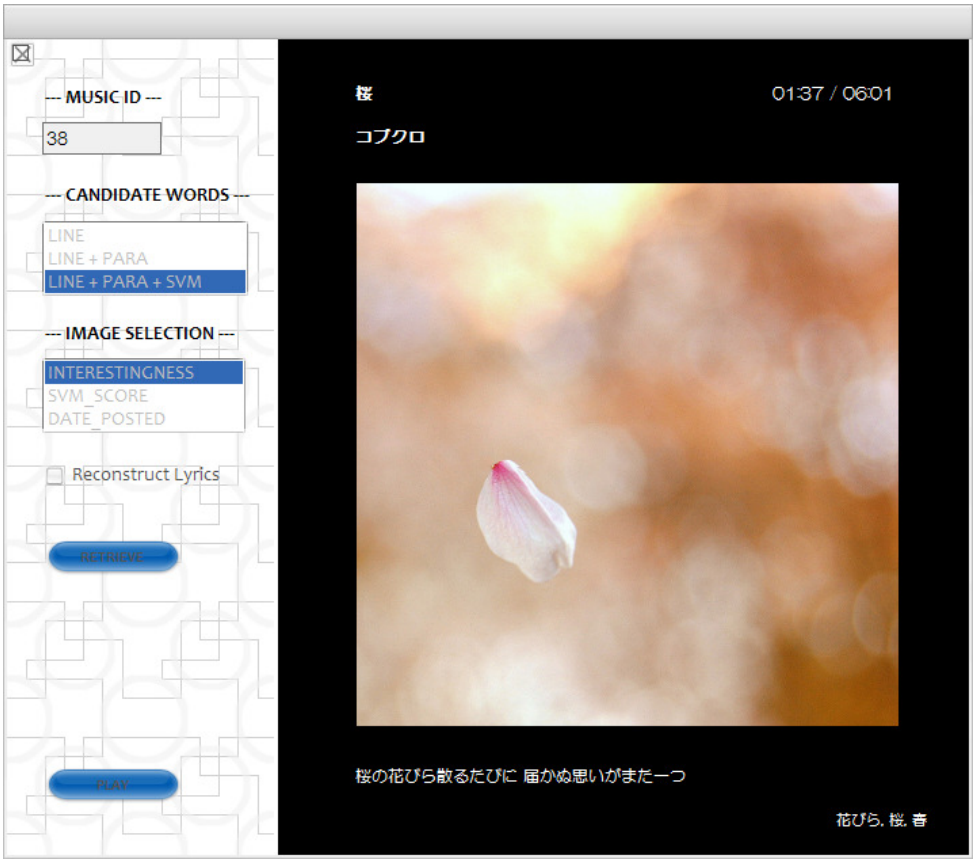

図 3.1.1.　システムインターフェース

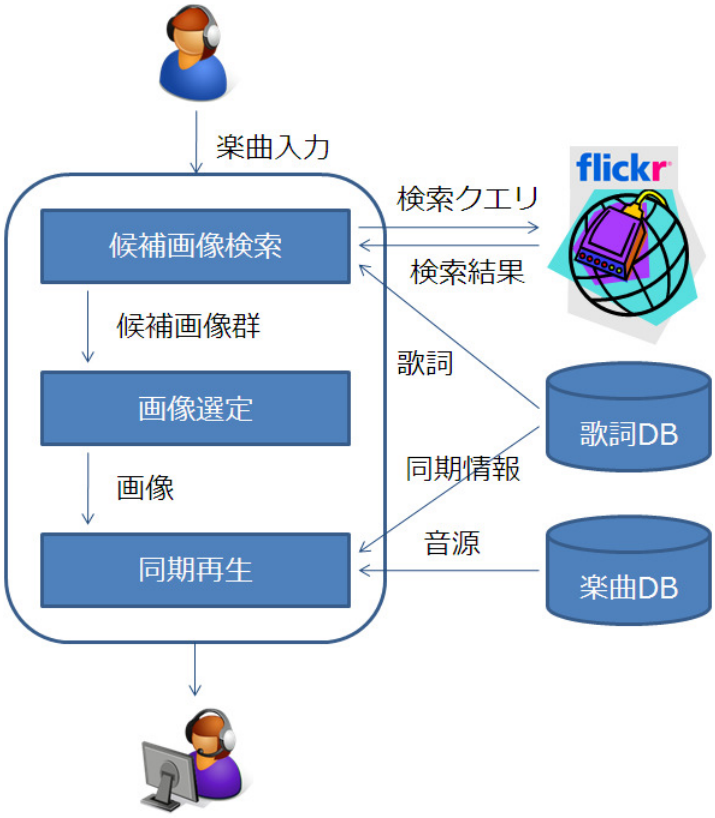

図 3.1.2. 処理概要図

本システムは，図 3.1.2.に示すように，以下の 3 つの処理から構成される．なお，本システムでは，検索対象となる画像データベースとして，写真共有サイト Flickr[6]を用い，画像に付与されているソーシャルタグを基に検索を行う．Flickr を用いることで，一般的な Web 画像検索エンジンを用いる場合と比較して，質の高い画像を得ることができる．さらに，楽曲の前奏区間で表示するアーティスト画像は，音楽のソーシャルネットワーキングサービスである Last.fm[18]により取得する．

(1) 候補画像検索

入力された楽曲の歌詞情報を基に検索クエリを構築し，スライドショーを構成する要素の候補となる画像（以下，候補画像）を Flickr から取得する．候補画像は，歌詞の各行に対して取得される．

(2) 画像選定

候補画像の中から，最終的にスライドショーに用いる画像を 1 枚選択する．ただし，同楽曲内で同じ画像は二度使用しない．このようにして，歌詞の各行に対し，画像が 1 枚対応付けられる．

⑶ **同期再生**

楽曲と歌詞の同期情報を基に，検索した画像群と楽曲を同期再生することにより，ユーザに出力する．画像の切り替えは歌詞の行単位で行い，画像の切り替わる区間にはフェード処理を施し，滑らかにつなげる．

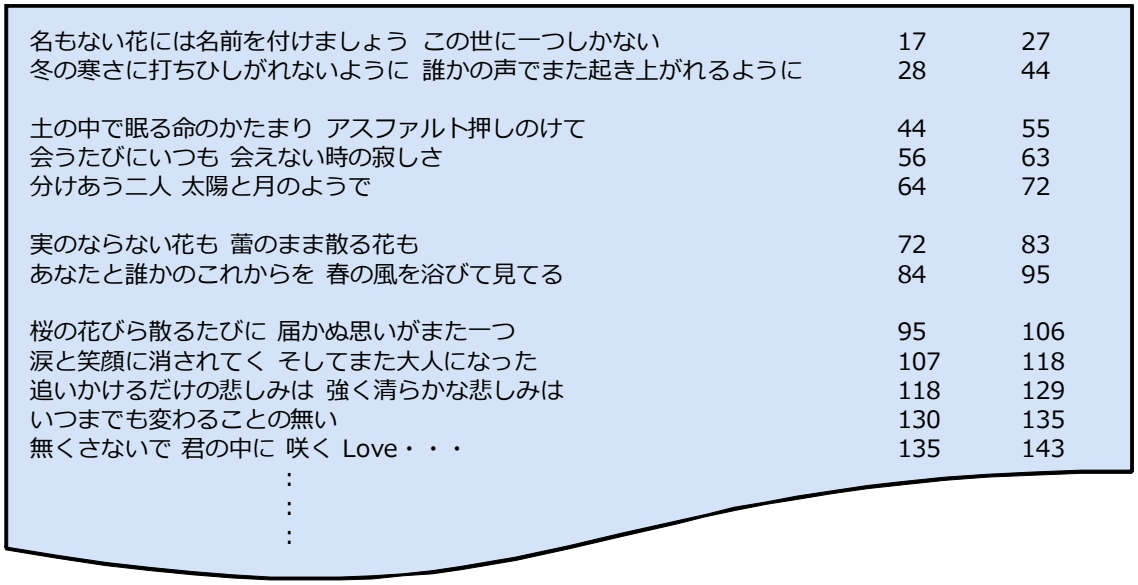

図 3.1.3.　楽曲と歌詞の同期情報

3.2.　候補画像検索処理

3.2.1.　概要

候補画像検索処理では，歌詞の各行に対し，スライドショーを構成する候補となる画像群を Web から取得する．

本処理の概要を図 3.2.1.に示す．本処理ではまず，歌詞情報を基にして，Web 画像検索に与えるクエリを構成する候補となる単語（以下，クエリ候補単語）を抽出する．この段階で，画像検索に有用でない単語は排除する．次に，クエリ候補単語から検索クエリとして最適な単語の組み合わせを選定する．最適な組み合わせを選定することで，効果的な画像絞込みが行える．そして選定したクエリ用いて Flickr により画像を検索し，その検索結果を候補画像として取得する．

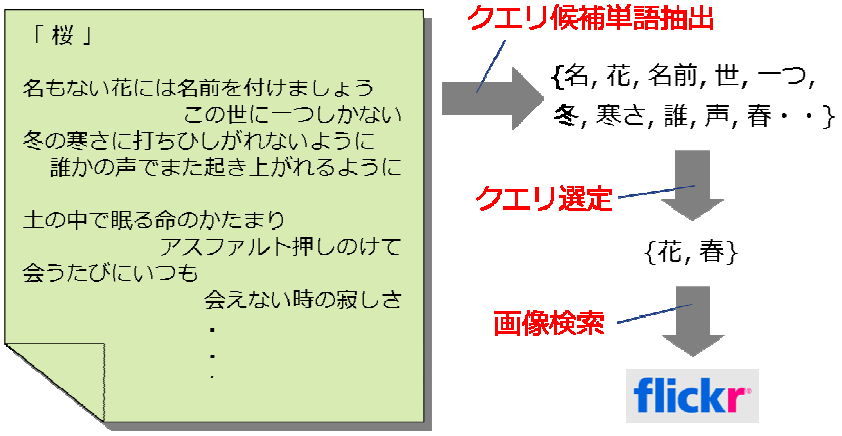

図 3.2.1.　候補画像検索処理

3.2.2.　クエリ候補単語抽出

楽曲の歌詞情報を基に，Flickr に与える検索クエリを構成する候補となる単語を抽出する．

歌詞のある行におけるクエリ候補単語は，行に出現する名詞，行の含まれる段落に出現する名詞，および，歌詞全体の印象を表現する単語（以下，全体印象語）により構成する．全体印象語に関しては，3.2.3.節にて説明する．なお，歌詞情報からの名詞抽出は，Yahoo! JAPAN により公開されている日本語形態素解析 Web API[19]を利用して行い，解析対象は日本語の歌詞のみとする．また，歌詞における段落は，歌詞データにおける空行を検出することにより判断する．

ここで，本論文では，楽曲 m における歌詞の l_i 番目の行に出現する名詞集合を $\mathbf{N}_{\mathbf{line}}(l_i)$，$l_i$ 番目の行を含む段落に出現する名詞集合を $\mathbf{N}_{\mathbf{para}}(l_i)$，楽曲 m の全体印象語集合を $\mathbf{N}_{\mathbf{all}}(m)$ と定義する．さらに，単語集合 **N** の要素全てをクエリとして用いて Flickr で AND 検索した際，検索された画像数を DF(**N**)(Document Frequency)，検索結果におけるユニークな画像投稿者数を UF(**N**)(User Frequency)と定義する．UF という値を定義する理由は，Flickr では一定のユーザが同一のタグを付与して大量の画像をアップロードする傾向があり，これにより DF の値が不当に高くなることがあるため，ユーザ 1 人の重みを等しく考慮するためである．

3.2.3.　全体印象語: 歌詞の印象に基づく楽曲自動分類

3.2.3.1. 概要

歌詞の全体の印象を表現する単語は，被験者実験により収集した学習データに基づく楽曲自動分類の分類結果を利用する[20]．具体的には，あらかじめ歌詞の印象を表現するカテゴリを設定し，各カテゴリに適していると判断される歌詞をもつ楽曲を被験者評価実験により収集し，それらを学習データとして分類器を構築することで，各カテゴリへの分類を

行う．そして，楽曲の分類されたカテゴリにおけるラベルを，全体印象語として適用する．設定したカテゴリを表 3.2.1.に示す．

表 3.2.1.　分類するカテゴリ

概念	カテゴリラベル
季節	春　夏　秋　冬
時間帯	朝　昼　夕方　夜
天候	晴れ　曇り　雨　雪　虹

3.2.3.2. 歌詞印象評価実験

各カテゴリの分類器の構築に必要な学習データを取得するため，被験者による評価実験を行った．被験者は提示される歌詞を参照し，そこから受ける印象を，表 3.2.1.に示すカテゴリラベルから該当するものを選択する．ただし，同概念に属するカテゴリは複数選択できず，該当するカテゴリが 1 つもなければ，選択しないこともできる．なお，本実験では，歌詞の印象に基づいて楽曲を分類するため，被験者に対して，音源や楽曲名・アーティスト名などの情報提示は行っていない. 本実験により最終的に，J-POP 楽曲 240 曲に対して，1 曲につき 5 名分の評価情報を収集した．

3.2.3.3. 歌詞情報ベクトル生成

楽曲の歌詞情報を数値化するために，ベクトル空間モデルを用い，TF*IDF アルゴリズムによる重み付けを行う．ベクトル空間モデルとは，文書中にどの単語がどの程度出現しているかをベクトルの形で表現する手法であり，ベクトルの各次元が単語に対応する．各次元の重みとして，単純に単語の出現回数を利用する他に，TF*IDF アルゴリズムがしばしば利用される．TF*IDF とは，文書中における重要と見なされる単語を抽出するアルゴリズムであり，文書検索や文書要約の分野で利用されている．本論文でも，この TF*IDF により文書ベクトルの重み付けを行う．

楽曲 m の歌詞における単語 t の TF*IDF(m, t)は(3.2.1)式により定義される．

$$TF * IDF(m,t) = TF(m,t) * \log_{10}\left(\frac{N}{DF(t)}\right) \tag{3.2.1}$$

ただし，TF(m, t)(Term Frequency)は楽曲 m における単語 t の出現回数，DF(t)は算出対象楽曲の中で単語 t が 1 回以上出現する楽曲数，N は算出対象楽曲数をそれぞれ示す．この定義からも示されるように，TF*IDF アルゴリズムでは，楽曲 m 中に多く出現し，かつ，他の楽曲であまり使用されていない単語ほど, 楽曲 m を特徴付ける上で重要であると見なされる．

なお，本研究ではベクトルの属性として，我々の所持しているデータベース内の 3062 曲の楽曲において，10 曲以上で使用されている名詞，および，一部の感動詞を選択した．また，TF*IDF の算出も，このデータベースを対象として行う．すなわち，(3.2.1)式における N の値は 3062 となる．

このようにして，楽曲の歌詞情報は 1070 次元のベクトルで表現される．

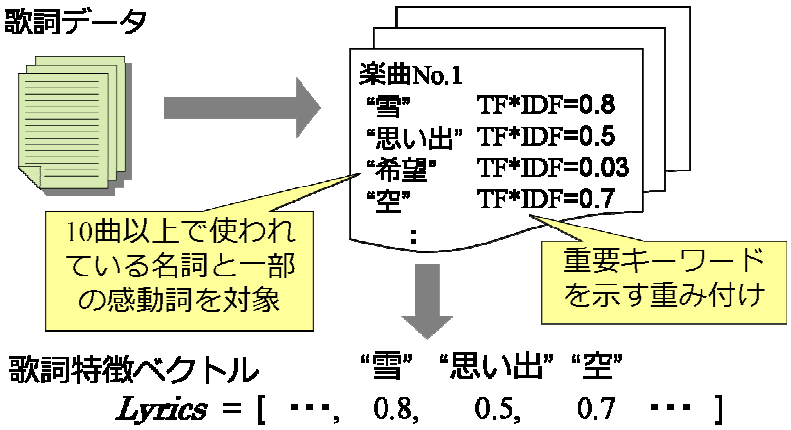

図 3.2.2.　歌詞情報ベクトル化

3.2.3.4. Support Vector Machine による分類

歌詞印象評価実験により得られた評価情報を基にして，各カテゴリの学習データを判定し，Support Vector Machine[21]（以下，SVM）による楽曲の分類を行う．SVM は多次元のベクトルで表現されたオブジェクトを二値分類する手法であり，文書分類の分野でも広く用いられている．本論文では，カテゴリごとに SVM を構築する．具体的には，各カテゴリに対して，評価者 5 名中 3 名以上が，そのカテゴリに適していると評価した楽曲を正例，それ以外の楽曲を負例と見なし, SVM の学習を行う. なお, SVM ツールとして SVMLight[22]を用い，学習には線形カーネルを適用する．このようにして，構築した SVM を用いて楽曲を分類し，分類されたカテゴリにおけるラベルをその楽曲の全体印象語として付与する．

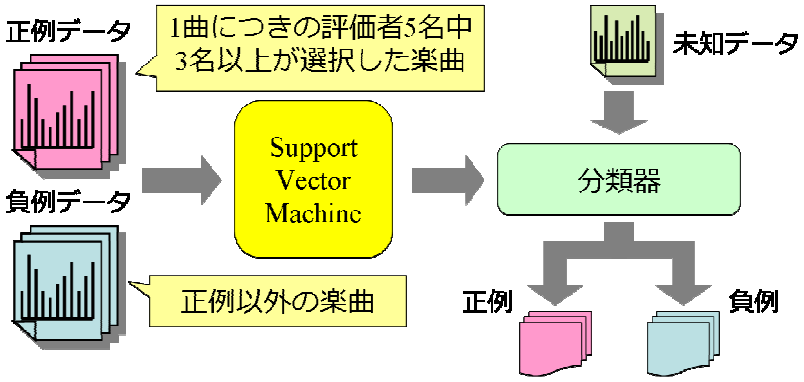

図 3.2.3.　Support Vector Machine による楽曲分類

3.2.4. クエリ選定

3.2.4.1. 概要

クエリ候補単語を基にして，最適な組み合わせを探索し，最終的に用いるクエリを選定する．

本論文では，画像検索クエリを選定するために，ソーシャルタグの傾向に基づく手法（以下，Social tag-based 手法）と歌詞における TF*IDF に基づく手法（以下，Lyrics-based 手法）の 2 つを提案する[23]．前者は，単語のソーシャルタグとしての使用頻度を基にしたクエリ選定法であり，後者は，単語の歌詞における重要度を基にしたクエリ選定法である．

3.2.4.2. Social tag-based 手法

Social tag-based 手法では，歌詞のある行に対するクエリ候補単語として，その行に出現する単語だけでなく，その周辺に出現する単語や全体印象語を用いることで，楽曲の流れや全体のテーマ性を考慮した画像を検索することができる．さらにその中から，ソーシャルタグとしての使用頻度を考慮し検索に有用な単語を選出することで，不要な画像の取得を避けることができる．そして，それらクエリ候補単語から，以下の 3 つの考えに基づき，最適な検索クエリの組み合わせを選定する．

i)　行に出現する単語を最優先して用いる．

行に出現する単語は，その行の内容を最も的確に表現していると考えられるため，最優先してクエリの要素とする．

ii)　より多くの単語を用いてクエリを構成する．

多くの単語を用いてクエリを構成することで，より詳細な絞込みができると考えられる．

iii)　タグとして付与されやすい単語を優先して用いる．

画像のタグとして付与されやすい単語を用いることで，画像を表現する上で重要な単語を用いた検索が期待できる．

Social tag-based 手法によるクエリ選定の手順を，図に例を示しながら説明する．なお，例の中では，歌詞のある行におけるクエリ候補単語として，

{“土”, “命”, “かたまり”, “アスファルト”, “時”, “二人”, “太陽”, “月”, “春”, “雨”}

という単語集合が抽出されているとする．

1. 入力楽曲 m における歌詞の l_i 番目の行のクエリ候補単語は，$\mathbf{N}_{line}(l_i)$，$\mathbf{N}_{para}(l_i)$，$\mathbf{N}_{all}(m)$ により構成される．さらに，これらの単語集合において，DF 又は UF が閾値以下の要素を排除する．なお，これらの閾値は，DF を 40，UF を 10 と経験的に設定した．このように，DF，UF によるフィルタリングを行うことで，画像として表現するのに適さない単語を排除することができる．

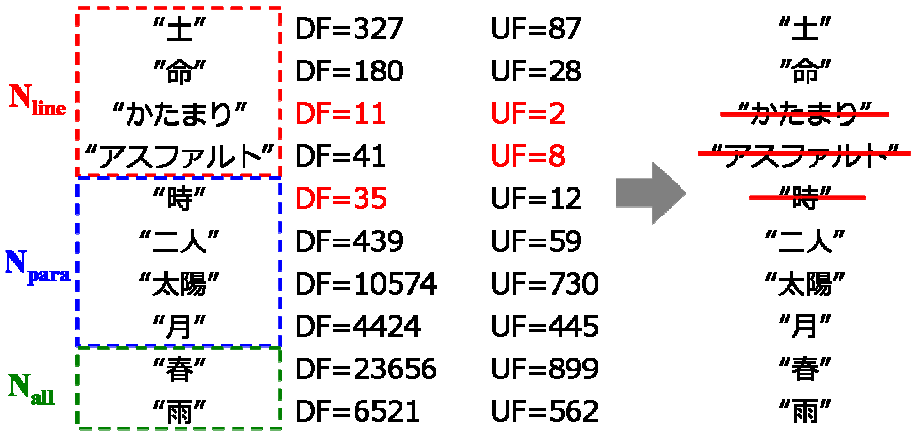

図 3.2.4.　Social tag-based 手法(1): DF・UF によるクエリ候補単語フィルタリング

2. $\mathbf{N}_{line}(l_i)$の冪集合 $\mathbf{P}(\mathbf{N}_{line}(l_i))=\{\mathbf{W}_{line,1}, \mathbf{W}_{line,2}, \ldots \mathbf{W}_{line,x}\}$において，$DF(\mathbf{W}_{line,i}) > 1$ を満たし，かつ，$|\mathbf{W}_{line,i}|$ が最大となるような $\mathbf{W}_{max}$ を選出する．ここで $|\mathbf{W}|$ とは，単語集合 $\mathbf{W}$ を構成する要素数（名詞数）を示す．$\mathbf{W}_{max}$ が複数ある場合，$UF(\mathbf{W}_{max})$が最大のものを選出する．こうして選出した $\mathbf{W}_{max}$ を行クエリ集合 $\mathbf{Q}_{line}(l_i)$とする．

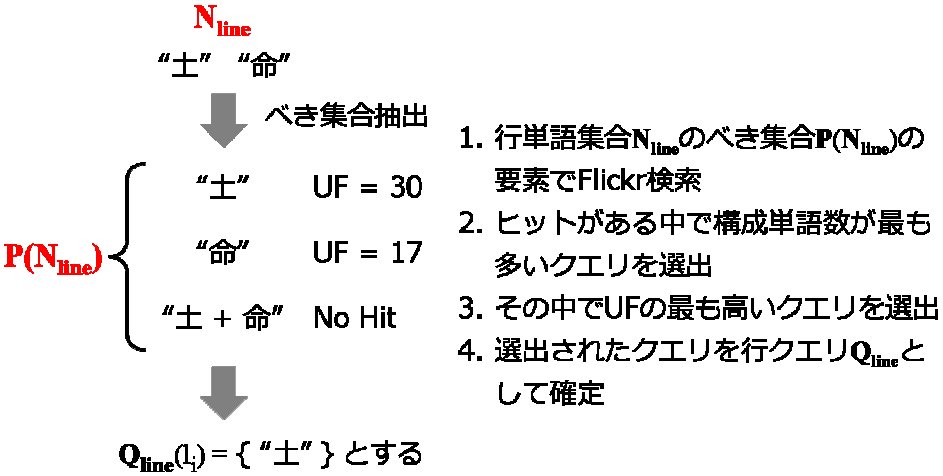

図 3.2.5.　Social tag-based 手法(2): 行クエリの選定

3. $\mathbf{N}_{para}(l_i)$と$\mathbf{N}_{all}(m)$の和集合における冪集合$\mathbf{P}(\mathbf{N}_{para}(l_i)+\mathbf{N}_{all}(m))=\{\mathbf{W}_{para,1}, \mathbf{W}_{para,2}, \cdots \mathbf{W}_{para,y}\}$の各要素に$\mathbf{Q}_{line}(l_i)$を加えた集合$\mathbf{P'}(\mathbf{N}_{para}(l_i)+\mathbf{N}_{all}(m))=\{\mathbf{W}_{para,1}+\mathbf{Q}_{line}(l_i), \mathbf{W}_{para,2}+\mathbf{Q}_{line}(l_i), \cdots \mathbf{W}_{para,y}+\mathbf{Q}_{line}(l_i)\}=\{\mathbf{W'}_{para,1}, \mathbf{W'}_{para,2}, \cdots \mathbf{W'}_{para,y}\}$において，先程と同様の処理により，$\mathbf{W'}_{max}$を選出する．こうして選出した$\mathbf{W'}_{max}$を最終クエリ集合$\mathbf{Q}(l_i)$とする．

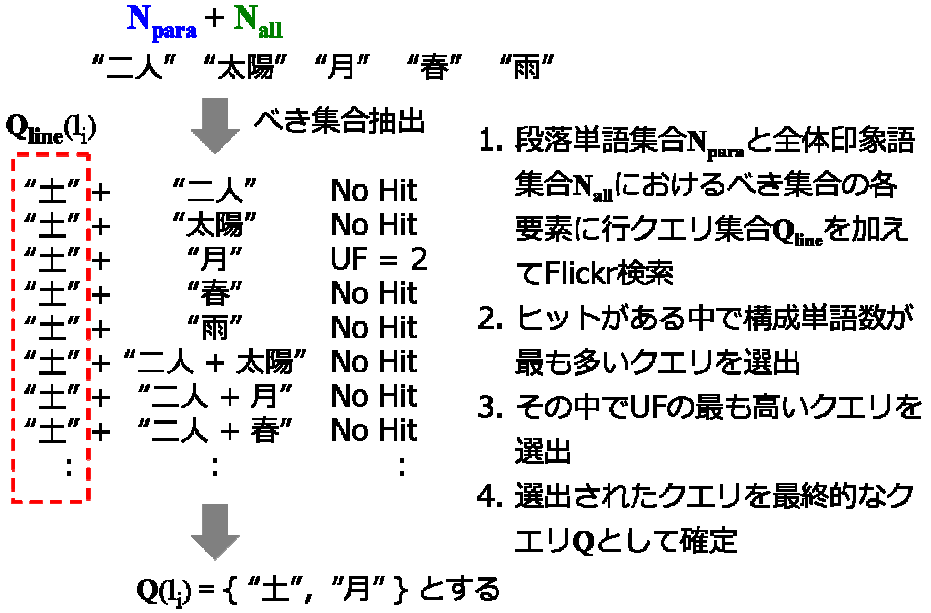

図 3.2.6.　Social tag-based 手法(3): 最終クエリの選定

4. $\mathbf{Q}(l_i)$の要素全てをクエリとして Flickr で AND 検索を行い，その検索結果をl_i番目の行における候補画像して取得する．

以上のようにして，候補画像を取得する．

3.2.4.3. Lyrics-based 手法

Lyrics-based 手法では，画像を検索するクエリとして，その行の出現単語の中で TF*IDF 値の高いものを中心に用いる．このようにして，歌詞を特徴付ける上での重要度を考慮したクエリ選定を行うことができる．

Lyrics-based 手法によるクエリ選定の手順を，図に例を示しながら説明する．なお，例の中では，歌詞のある行におけるクエリ候補単語として，

{“土”, “命”, “かたまり”, “アスファルト”}

という単語集合が抽出されているとする．

1. 入力楽曲 *m* における歌詞の l_i 番目の行のクエリ候補単語は，$\mathbf{N}_{line}(l_i)$のみにより構成される．ここで，$\mathbf{N}_{line}(l_i)$に含まれる各単語についての TF*IDF 値をあらかじめ算出しておく．TF*IDF 値は我々の所持している楽曲データベース 3062 曲を基に算出した．

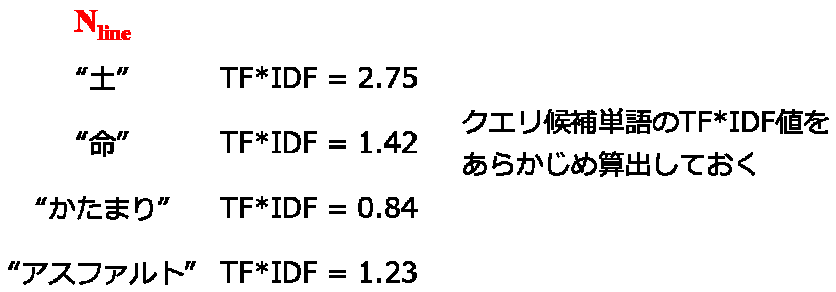

図 3.2.7. Lyrics-based 手法(1): クエリ候補単語の TF*IDF 算出

2. $\mathbf{N}_{line}(l_i)$に含まれる要素全てをクエリとして Flickr で AND 検索を行う．その結果ヒットがあれば，検索結果を候補画像として取得し，ヒットがなければ，$\mathbf{N}_{line}(l_i)$の中で最も TF*IDF 値の低い要素を取り除く．

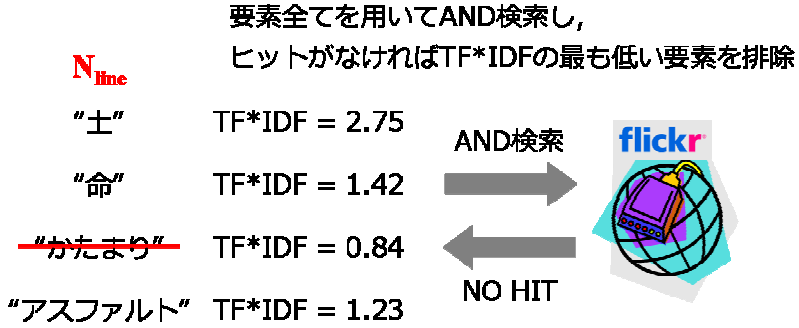

図 3.2.8. Lyrics-based 手法(2): TF*IDF の低い要素の排除

3. 残った要素を全て用いて再度 Flickr にて検索を行う．この処理を，候補画像が取得できるまで繰り返し行う．クエリの要素がなくなるまで取得できない場合，1 つ前の行において用いたクエリにより，候補画像を取得する．

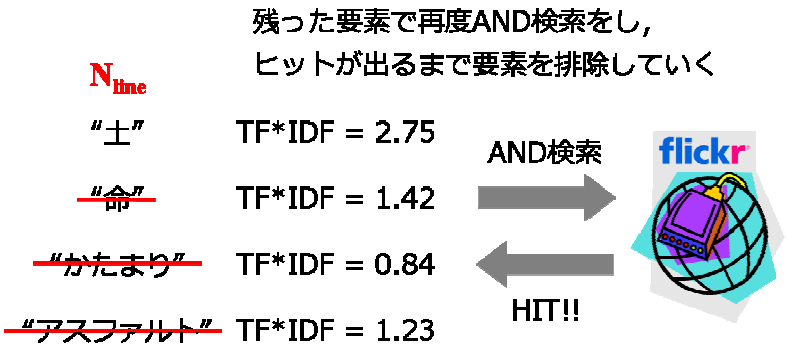

図 3.2.9. Lyrics-based 手法(3): 候補画像取得

以上のようにして，候補画像を取得する．

第4章　評価実験 I

本章では，3 章にて提案した 2 つのクエリ選定手法の有効性を検証するために実施した被験者による評価実験について述べる．

4.1. 実験内容

画像検索に与えるクエリを選定することの有効性の検証と，3 章にて提案した 2 つの画像検索クエリの選定手法における性能の評価を目的として，被験者による評価実験を実施した．被験者は，Social tag-based 手法，および，Lyrics-based 手法にてクエリを選定することで生成した楽曲スライドショーを視聴し，それぞれについて以下に示す全体評価と個別評価を行う．

- **全体評価**

楽曲と生成したスライドショー全体との適合性を，以下の質問にて 5 段階で評価する．

> Q. この手法で提示した画像群は楽曲に合っていましたか？
>
> 5: とても楽曲に合っていた
>
> 4: どちらかといえば楽曲に合っていた
>
> 3: どちらともいえない
>
> 2: どちらかといえば楽曲に合っていなかった
>
> 1: あまり楽曲に合っていなかった

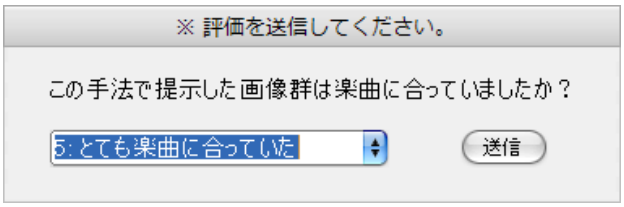

図 4.1.1. 全体評価

- **個別評価**

楽曲の歌詞における行単位での画像との適合性を評価する．具体的には，図 4.1.2.に示す評価ページにおいて，歌詞の各行に対して両手法により検索した画像を 2 枚提示し，その行と表示するのに適している画像を選択することで評価を行う．選択はチェックボックスにより行うため，両画像を選択することや，逆に，いずれも選択しないことも可能である．

図 4.1.2. 個別評価

さらに，全ての評価を行った後に，被験者に対してアンケートを実施した．アンケート内容を以下に示す．

Q1. 楽曲に適している画像を判定する際に，楽曲のどの特徴を考慮しましたか？
[いずれか 1 つ選択]
○ 音響特徴と歌詞特徴
○ 主に音響特徴
○ 主に歌詞特徴
○ その他特徴
Q2. 本システムの優れていた点があればご記入お願いします. [自由記述形式]
Q3. 本システムの改善すべき点があればご記入お願いします. [自由記述形式]

なお，本実験では，候補画像の中から最終的にスライドショーに用いる画像を選定する処理（画像選定処理）は, Flickr のランキングを基に行う. 具体的には, Flickr の“interestingness”指標における画像検索結果ランキングにおいて，最も上位の画像を用いてスライドショーを構成する．Flickr にて利用できる検索結果のソート方法は他にも存在するが，実際に全てを利用してみたところ, “interestingness”指標によるソート方法がスライドショーを構築するにあたって最も適していると判断したため，今回はこの指標を使用した．

4.2. 実験フロー

本実験の具体的な流れを以下に示す．

1. 1 つ目の手法にて生成した楽曲スライドショーを視聴し，このスライドショーに対する全体評価を行う．
2. 同じ楽曲ついて，もう一方の手法にて生成した楽曲スライドショーを視聴し，同様に全体評価を行う．
3. 評価ページにおいて，視聴した楽曲について両手法の個別評価を行う．

以上の作業を提示された 5 曲全てについて行い，最後に，アンケートに回答する．なお，両手法の提示する順序効果は考慮してある．

4.3. 実験データ

本実験の被験者は大学生 20 名で，対象楽曲データは表 4.1.1.に示す J-POP 楽曲 10 曲を用い，1 曲につき 10 名分の評価情報を収集した．なお，表 4.1.1.では用いた楽曲における ID，タイトル，アーティスト名の他に，その楽曲に付与されている全体印象語を示している．

表 4.3.1. 実験対象楽曲

Music ID	タイトル	アーティスト	全体印象語
14	散歩道	JUDY AND MARY	晴れ，夕方
38	桜	コブクロ	春，雨
88	パスポート	GOING UNDER GROUND	春
129	君の名を呼ぶ	浜田省吾	夏，夜
153	プラネタリウム	大塚愛	夜，夏，晴れ
163	One more time, One more chance	山崎まさよし	冬
452	瞳をとじて	平井堅	晴れ
473	you	倖田來未	冬，雪，夜
523	LIFE	YUI	晴れ，昼
1775	ワダツミの木	元ちとせ	晴れ，夏

4.4. 実験結果

図 4.4.1.に全体評価の結果として，各楽曲における 10 名の評価者の 5 段階評価平均値を示す．また，図 4.4.2.に個別評価の結果として，各楽曲における 10 名の評価者の平均画像適合率を示す．ここで，画像適合率とは，個別評価において楽曲に適していると判断された画像の割合を示す．なお，両図において，青いグラフが Social tag-based 手法による評価，赤いグラフが Lyrics-based 手法による評価を示す．

まず，図 4.4.1.の全体評価結果を参照すると，実験に使用した楽曲 10 曲中 8 曲において Social tag-based 手法が高い評価値を示しており，1 曲において Lyrics-based 手法が高い評価値を示している．また，Social tag-based 手法に関しては，全ての楽曲において，中間値である 3 以上の評価を得ている．これより，歌詞特徴のみを用いて適切にクエリを選定するだけで，比較的良いスライドショーを生成できることが示せた．

一方，図 4.4.2.の個別評価結果を参照すると，9 曲において Social tag-based 手法の方が高い評価となり，残り 1 曲で Lyrics-based 手法の方が高い評価となった．これより，Social tag-based 手法によってクエリ選定を選定し画像検索を行うことで，より楽曲に合った画像

を取得することができるといえる．しかし，Social tag-based 手法における 10 曲の平均画像適合率は約 50%にとどまっていることから，この点に関してはまだ改善の余地があると考えられる．

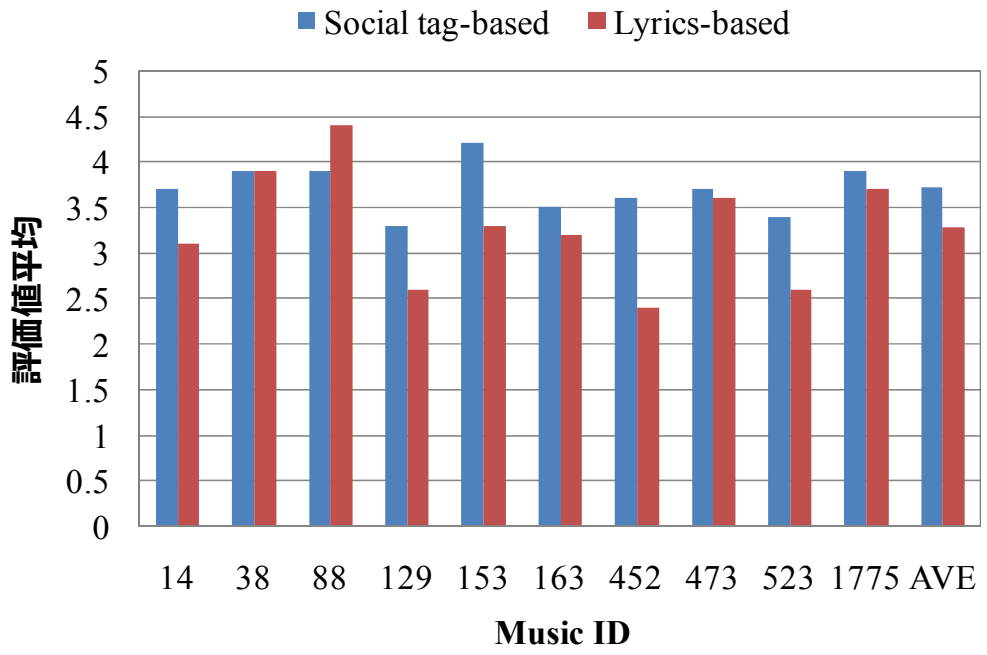

図 4.4.1. 全体評価結果(1): 各楽曲における評価値平均

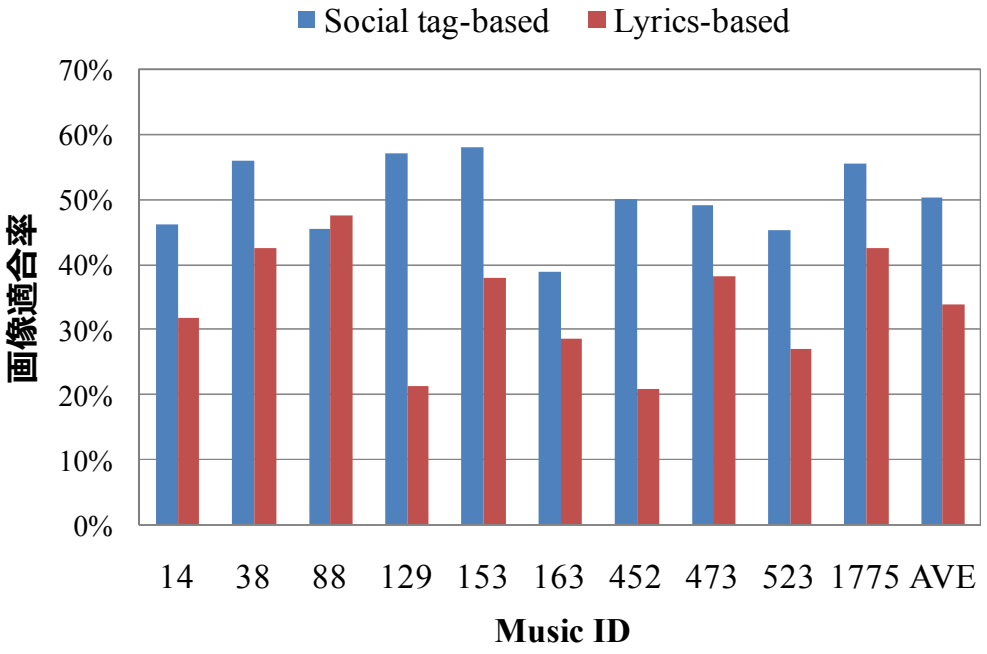

図 4.4.2.個別評価結果(1): 各楽曲における画像適合率

さらに，評価の視点を変え，各被験者の重みを同等に扱うための評価を行った．具体的には，被験者ごとにある楽曲の 2 手法に対する評価を比較し，どちらの手法により高い評価を与えたか，を判定することで，各楽曲においてどちらの手法を支持した被験者が多いか，を評価した．このように評価をすることで，各評価における絶対的な差（5 段階評価値や画像適合率の差）の情報は欠落するが，各被験者の重みを同等に扱った評価となる．この方法での評価結果を，図 4.4.3.に全体評価について，図 4.4.4.に個別評価についてそれぞれ示す．両図中では，横軸に示される各楽曲において，評価者 10 名が支持した（高い評価を与えた）手法の分布を示している．青いグラフが Social tag-based 手法を支持する被験者，赤いグラフが Lyrics-based 手法を支持する被験者，緑のグラフが両手法に同じ評価を与えた被験者をそれぞれ示す．

まず，図 4.4.3.に示された全体評価結果において，1 曲につきの評価者の半数以上（5 名以上）が，Social tag-based 手法を支持している楽曲は 10 曲中 4 曲，Lyrics-based 手法を支持している楽曲は 10 曲中 1 曲，同等の評価を与えている楽曲は 10 曲中 3 曲であり，残りの 2 曲に関しては半数を超える分布はなかった．

一方，図 4.4.4.に示された個別評価結果においては，10 曲全ての楽曲において，1 曲につきの評価者の半数以上（5 名以上）が，Social tag-based 手法に高い評価を与えている．ただし，ID:88 の楽曲のみ，Lyrics-based 手法にも半数の支持がある．

このような観点から評価した結果においても，Lyrics-based 手法と比較し，Social tag-based 手法の方が，楽曲に適したスライドショーを構成していることがわかる．

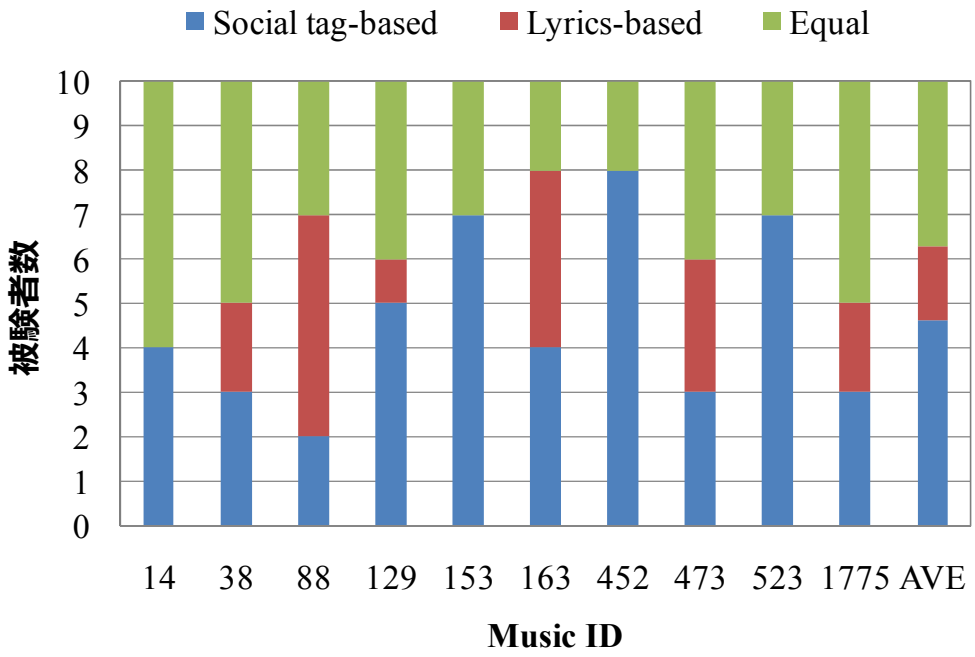

図 4.4.3. 全体評価結果(2): 各楽曲における被験者の評価分布

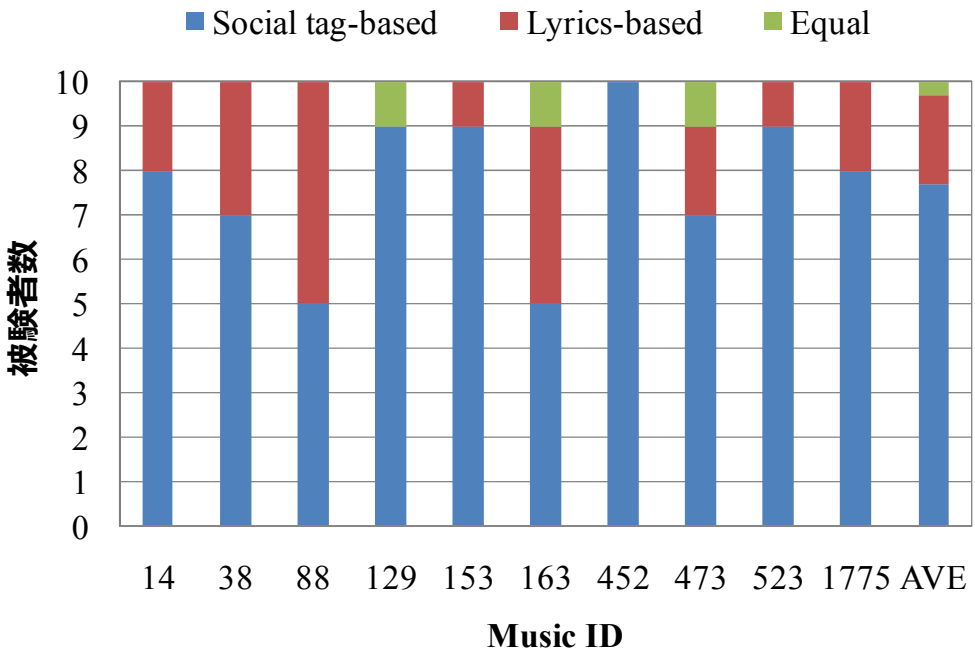

図 4.4.4. 個別評価結果(2): 各楽曲における被験者の評価分布

4.5. 考察

4.4.節に示した結果についての考察を行う.

まず, 実際に両手法にて用いられたクエリを確認してみたところ, Lyrics-based 手法では, 約 88%のクエリが単語 1 語のみで構成されていたが，Social tag-based 手法では，このようなクエリは約 16%であった．これは，より多くの単語を用いてクエリを構成することで，詳細に画像検索の絞り込みを行えていることを示す.

また，ID:153，452，523 の楽曲では，提案手法による改善の度合い，すなわち，Social tag-based 手法と Lyrics-based 手法の評価の差が比較的大きくなっている. これらの楽曲では, Lyrics-based 手法を用いた場合，“気持ち”，“痛み”, “ダメ”のような，画像検索クエリとしては相応しくない単語を用いて検索が行われていたが，Social tag-based 手法を用いた場合，これらの単語は取り除かれていた．つまり，Social tag-based 手法におけるクエリ候補単語の DF・UF フィルタリングが，効果的に機能しているといえる.

一方，実験対象楽曲の中で唯一，全体評価，個別評価ともに，Lyrics-based 手法の方が優れた評価となっていた ID:88 の楽曲では，以下のような 3 つの傾向が見られた．まず 1 つ目は，先述とは逆で，DF・UF フィルタリングによって，本来検索に用いるべき単語が取り除かれていた．2 つ目は，段落単語を用いることによって，不適切なクエリを構築していた．具体的には，同じ段落に含まれる単語だからといって，未だ歌われていない箇所の単語を参照し，それをクエリとして用いることで，（その時点では）あまり関係のない画像を取得していた．そして 3 つ目は，ノイズタグが多く付与されている画像（図 4.5.1.）が取得されていた．ここでいうノイズタグとは，画像を検索されやすくするために付与する，その画像の内容とは無関係のタグのことである．この傾向は ID:88 の楽曲に限らず，全体的に見受けられた．このように，検索クエリを適切に構築しても，そもそも画像に付与されているタグがノイズである場合には，結果的に不適切な画像を取得するおそれがあるため，何らかの対策が必要である.

tag = { “春”, “南”, “夜”, “雲”, “設計”, “心”,….}

図 4.5.1.　ノイズタグの付与された画像例

参考結果

参考として，図 4.5.2.に Social tag-based 手法における，クエリの構成要素別，また，単語数別の画像適合率を示す．ここで，クエリの構成要素とは，行に出現する単語（$\mathbf{N_{line}}$），行を含む段落に出現する単語（$\mathbf{N_{para}}$），全体印象語（$\mathbf{N_{all}}$）のうち，どの単語の組み合わせによりクエリを構成しているかを表す．また，単語数とは，クエリを構成する単語の数を表す．なお，本結果を参考としたのは，各ケースのサンプル数が十分に確保できなかったためである．例えば，$\mathbf{N_{line}}$，$\mathbf{N_{para}}$，$\mathbf{N_{all}}$，を用いて 4 単語のクエリを構成した場合，図より画像適合率は 90%となっているが, 実際はこのようなクエリを構成したケースは 2 回のみである．このように十分なサンプル数が得られていないため，参考として本結果を記す．

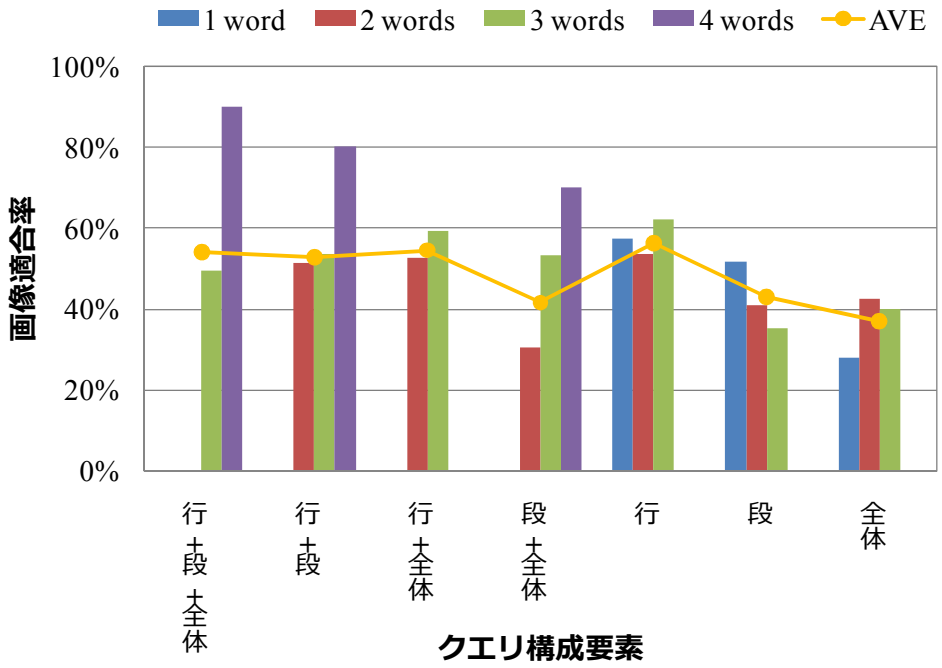

図 4.5.2.　クエリ構成要素別・単語数別画像適合率

図 4.5.2.において，クエリ構成要素に着目すると，全体的に $\mathbf{N_{line}}$ をクエリとして用いた場合に画像適合率が高くなっていることがわかる．これは，Social tag-based 手法における“行に出現する単語を優先して用いる”という考えが，間違っていないことを示唆している．一方，クエリ単語数に着目すると，クエリ構成要素が $\mathbf{N_{para}}$ のみの場合を除き，クエリを構成する単語数が多い方が，比較的画像適合率も高くなる傾向がある．これも同様に，Social tag-based 手法における“より多くの単語を用いてクエリを構成する”という考えが, 正しいことを示している．なお，$\mathbf{N_{para}}$ のみ用いた場合が例外だったのは，先述したような，未だ歌われていない箇所の単語の参照による影響があると考えられる．

4.6. 本実験の結果によるシステム修正

本実験結果を受けて，システムに以下のような修正を施した．

- 段落単語は既に歌われた箇所のみから参照する

4.5.節にて述べたように，段落単語として，未だ歌われていない箇所の単語を参照し，それを基に画像検索を行うことで，視聴者が理解できない画像を取得するケースがあった．そのため，段落単語として参照する範囲を，既に歌われた箇所のみに限定することで，上記のような問題を回避することができる．

- 各クエリ候補の組み合わせに対して UF によるフィルタリングを行う

4.5.節にて述べたように，ノイズタグによって不適切な画像を取得するケースが，実験結果において多々見られた．その対策として，UF によるフィルタリングを個々のクエリ候補単語のみでなく，それらの組み合わせに対しても行う．現状では，個々のクエリ候補単語に対する DF・UF フィルタリングによって，画像を表現する上で不要な単語は排除できているが，残った単語の組み合わせの妥当性に関しては考慮していない．図 4.5.1.に示したノイズタグの付与された画像を例にすると，“春”，“夜”，“雲”のような単語は，個々で見ると画像を示すのに重要な単語であるが，これら全てが実際にタグとして付与されている画像は少数である．つまり，これらの単語を組み合わせて利用するユーザが少ないという傾向は，その組み合わせが画像検索において広く用いられていないということを示しており，それにより取得した画像はノイズタグが付与されている可能が高いと考えられる．よって，このように UF フィルタリングをクエリ候補単語の組み合わせに対しても行うことで，ノイズによる画像の取得を軽減することが期待できる．

なお，UF の閾値は，$\frac{10}{\text{クエリを構成する単語数}}$ とする．一般的に，クエリを構成する単語数が増えるほど，それにより算出される UF は減少していくため，より多くの単語によりクエリを構成するほど，UF の閾値を寛容に設定している．

4.7. アンケート結果

被験者アンケートの結果を以下に示す.

- Q1 について

Q1 に対する回答結果を表 4.7.1.に示す.

表 4.7.1. Q1 に対する回答結果

選択肢	回答人数
音響特徴と歌詞特徴	7 名
主に音響特徴	1 名
主に歌詞特徴	12 名
その他特徴	0 名

本結果において，多くの被験者が楽曲と画像との適切性を判断するために歌詞特徴を主に意識していることから，歌詞特徴に基づいて用いる画像を選定することは有効であるといえる.

- Q2 について

Q2 に対する回答結果から主なものを以下に示す.

a) 楽曲を聞きながら歌詞に合った画像を眺めることで，受ける印象が深くなった.
b) 楽曲に合った画像が提示されると，感情移入しやすくなる.
c) 意外な画像が提示されることで，新たな解釈を発見できた.
d) 楽曲の内容をよりイメージしやすくなる.
e) ある程度文脈や前後の流れに沿った画像が検索されていた.
f) 自らの手で画像を用意しなくても，自動でスライドショーを生成してくれる.

全体的に，楽曲に合った画像を見ることで，受ける印象が強くなった，との意見が多かった．また，予期しない画像が提示されるが，それも新たな解釈として楽しめるとの意見もあった．一方で，自動でスライドショーを生成してくれる（自分で素材を用意する必要がない）という手軽さに対しても，本システムのメリットとして指摘する意見が多く見られた.

- Q3 について

  Q3 に対する回答結果から主なものを以下に示す.

a)	間奏区間でも画像を切り替えるべきである.
b)	複数の楽曲に対するスライドショーにおいて特定の画像が使用される.
c)	画像を切り替えるタイミングが適切でない.
d)	スライドショー全体の統一感に欠ける.
e)	曲調と異なる画像が選択されることがある.
f)	画質により印象が変わる.
g)	楽曲構造を考慮すべきである.

回答結果は，スライドショーを構成する画像に関する課題（曲調と合っていない，統一感に欠ける etc），画像の切り替え方に関する課題（間奏区間でも画像を切り替えるべき，切り替えるタイミングが不適切 etc）が中心であった.

まず b)に関して，この問題の 1 つの原因として，Web 画像検索結果から最終的に用いる画像を選定する処理（画像選定処理）において，Flickr におけるランキングをそのまま利用していることが挙げられる．Flickr におけるランキングは，検索クエリが同一であれば結果も同じとなるため，異なる楽曲において同一のクエリが検索に用いられた場合，最終的に用いられる画像も同一のものが選択される．このようにして，特定の画像が複数の楽曲スライドショーにまたがって使用される.

一方，c)に関しては，歌詞の行単位で画像を切り替えることにより生じる問題である．つまり，行の切り替わりと同期して画像を切り替えるため，画像を表示する時間が歌詞の行の長さに依存することで，表示時間が極端に短い，または，長い画像が生じている.

また，d)に関しては，画像検索クエリを構成する要素として，全体印象語を適用していることからも，スライドショー全体の統一感を考慮していると思われるが，全体印象語はあくまでクエリを構成する一候補に過ぎず，全ての画像検索において用いられるとは限らない．実際に，本実験で使用した楽曲に対して生成したスライドショーにおいて，クエリの構成要素として全体印象語が選択されたケースは，全体の約 60%であった．すなわち，約 40%の画像は，全体の印象を考慮していないことになる.

これらの課題を解決するための手法を，以降の章において述べる．具体的には，b)および d)の課題を解決するため，5.1.節にて，歌詞全体の印象との適合度に基づく画像選定手法を提案し，c)の課題を解決するため，5.2.節にて，画像切り替えタイミングの再構成法を提案する．また，画像切り替えタイミングの再構成に伴い，間奏区間における画像の切り替えも実現し，a)の課題の解決も図る.

第5章　改善手法

本章では，4 章に示した評価実験におけるアンケートにて指摘された課題を解決するため，5.1.節にて，歌詞全体の印象との適合度に基づく画像選定手法を提案し，5.2.節にて，画像を切り替えるタイミングを再構成する手法を提案する．

5.1. 画像選定処理

5.1.1. 概要

画像選定処理では，候補画像検索処理において取得した候補画像の中から，最終的に用いる画像を選定する．本節では，“複数の楽曲に対するスライドショーにおいて特定の画像が使用される”，“スライドショー全体の統一感に欠ける”という課題を解決するための画像選定手法を提案する．

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表現する score を算出し，その値が最も高い画像を選定する．score は，画像に付与されている全てのタグと，入力楽曲の全体印象語との間における関連の強さを基に算出する．この関連の強さは，全体印象語との関連タグ情報を基に定義する．このように，画像に付与されているタグの傾向を参照し，全体印象語との関連度を数値として表現することで，全体印象語が直接タグとして付与されていない画像でも，“全体印象語らしさ”を表現することができる．そして，この数値を基に画像を選定することで，スライドショー全体の統一感を生み出すことができる．さらに，画像に付与されている全体印象語の組み合わせによって score 算出の指標が変化するため，異なる楽曲にて同じ画像が使用されるケースが少なくなり，より多様性のあるスライドショーを構築することができる．

なお，本処理において全体印象語を用いるため，候補画像検索処理における画像検索のためのクエリ候補となる単語は，歌詞に出現する名詞のみ（$\mathbf{N_{line}}$，$\mathbf{N_{para}}$）により構成する．また，候補画像は各行において最大 1500 枚（Flickr における“interestingness”指標を用いたランクの上位 1500 枚）取得し，それら全てに対して score を算出する．

5.1.2. ソーシャルタグの共起確率に基づく関連タグ抽出

score を算出するために，各全体印象語における関連タグと関連度を抽出する．関連タグは，Flickr におけるタグの共起確率を基に算出した関連度を用いて抽出する．タグ *w* に対するタグ *t* の共起確率とは，*w* が付与されている画像において，*t* も付与されている確率により定義されるため，2 つのタグ間の関連の強さを示す指標となる．

全体印象語 *n* に対するタグ *t* の関連度 R(*t*|*n*)を(5.1.1)式により定義する．

$$R(t \mid n) = P_{DF}(t \mid n) = \frac{DF(t \cap n)}{DF(n)} \tag{5.1.1}$$

この定義により関連度を算出したところ，適切に関連タグが抽出できなかった．これは，Flickr では，特定のユーザが同一のタグを付与して，大量の画像をアップロードすることによって，DF の値が不当に高くなってしまう傾向があることに起因する．そのため，関連度 R(*t*|*n*)を(5.1.2)式に示すように，UF により定義する．

$$R(t \mid n) = P_{UF}(t \mid n) = \frac{UF(t \cap n)}{UF(n)} \tag{5.1.2}$$

このようにして，関連度を算出することで，DF を用いた場合よりは改善することができた．一方で，多くの画像に満遍なく付与されているタグにおいて不当に関連度が高くなるという傾向が見られた. そこで, このようなタグの関連度を抑えるべく, (5.1.3)式のように, *n* と *t* の共起確率だけでなく，*n* と同じ概念に含まれる印象語に対する共起確率も考慮することで，*n* に対してのみ共起確率の高いタグ，すなわち，*n* に特化して関連の強いタグを抽出することができる．

$$R(t \mid n) = P_{UF}(t \mid n) - \frac{\sum_{x \in N, x \neq n} P_{UF}(t \mid x)}{|N| - 1} \times weight \tag{5.1.3}$$

ここで，**N** は *n* の属する概念に含まれる印象語集合である．例えば，*n* = “春”の場合，“春”は“季節”という概念に含まれる印象語であるため, **N**={ “春”, “夏”, “秋”, “冬”}となる. また, |**N**|は印象語集合に含まれる要素数を示し，*weight* は同概念の他の印象語に対する共起確率を考慮する程度を設定する重みであり，本研究では，*weight*=3 としている．

全体印象語“春”に対し，上記 3 つの定義により算出した関連度を表 5.1.1.に示す．表中では，適切でないと思われる結果を赤く示してある．表より，DF ではなく UF を用いることで不適切な結果が軽減され，また，“春”に対する共起確率に加え，“夏”，“秋”，“冬”に対する共起確率も考慮することで，より精度の高い関連度算出ができていることが確認できる．

表 5.1.1.　全体印象語“春”に対する関連度算出結果

(5.1.1)式		(5.1.2)式		(5.1.3)式	
タグ	関連度	タグ	関連度	タグ	関連度
春	1.0000	春	1.0000	春	0.9187
桜	0.242	桜	0.670	桜	0.636
野球	0.173	花	0.454	花見	0.150
高校	0.172	日本	0.373	さくら	0.136
練習試合	0.167	東京	0.190	梅	0.134
高校野球	0.167	梅	0.173	花	0.094
沖縄	0.167	花見	0.153	菜の花	0.091
日本	0.154	さくら	0.149	サクラ	0.072

このように(5.1.3)式により定義した関連度を，それぞれの印象語に対して，UF(t)≧5 を満たし，かつ，全角文字のみで構成されたタグ t を対象に算出し，経験的に関連度が 0.024 以上のタグを関連タグとして判定した．以上の処理によって抽出した，一部の全体印象語に対する関連タグとその関連度を，表 5.1.2.に示す．

表 5.1.2.　全体印象語“夏”，“夕方”，“雨”に対する関連タグと関連度

夏		夕方		雨	
関連タグ	関連度	関連タグ	関連度	関連タグ	関連度
夏	0.939	夕方	0.981	雨	0.953
ｔｏｋｙｏ	0.223	ｓｋｙ	0.285	ｒａｉｎ	0.764
花火	0.180	夕暮れ	0.125	傘	0.069
ひまわり	0.092	夕焼け	0.113	台灣	0.069
海	0.080	夕日	0.112	台北	0.054
祭り	0.077	日没	0.056	水滴	0.034
浴衣	0.056	夕景	0.049	梅雨	0.033
蝉	0.049	黄昏	0.038	雫	0.030

以上の処理によって，各全体印象語に対して，関連タグとその関連度を抽出する．

5.1.3. score 算出法

全体印象語における関連タグ情報を基に，画像の score を定義する．score は，入力楽曲の全体印象語における関連タグが多く付与されている画像ほど高くなる．逆に，全体印象語との関連の低いタグが多く付与されている画像ほど，score は低くなる．このように，関連の高いタグだけでなく，関連の低いタグまで考慮するのは，Flickr におけるノイズタグへの対策である．このようにすることで，全体印象語と関連が高く，かつ，ノイズタグの少ない画像を選定することができる．

楽曲 *m* が入力された際の画像 *i* における score の定義を(5.1.4)式に示す．

$$score(i) = \frac{\sum_{n \in N_{all}(m)} \sum_{t \in T_i \cap RT_n} R(t \mid n)}{\left| T_i \cap \overline{RT_{N_{all}(m)}} \right|} \tag{5.1.4}$$

ただし，$\mathbf{N_{all}}(m)$は楽曲 *m* における全体印象語集合，$\mathbf{RT}_n$ は全体印象語 *n* における関連タグ集合，$\mathbf{RT}_{\mathbf{N_{all}}(m)}$は全体印象語集合 $\mathbf{N_{all}}(m)$の全ての要素における関連タグ集合の和集合，$\mathbf{T}_i$ は画像 *i* に付与されているタグ集合，|**T**|はタグ集合 **T** に含まれる要素数をそれぞれ示す．

以上のように定義した score を候補画像全てに対して算出し，その値の最も高い画像を選定する．

5.2. 画像切り替えタイミング再構成

5.2.1. 概要

スライドショーを構成する各画像の表示時間を適切にするために，画像切り替えタイミングの再構成を行う．画像の切り替えを歌詞の行単位で行うと，その歌詞の行の長さに画像の表示時間が依存し，適切な表示時間とならないことがある．例えば，歌詞の行が短い場合，それに伴い画像の表示時間も短くなり，画像の内容を把握するのを困難にする．逆に，歌詞の行が長い場合，画像の表示時間も長くなり，その結果，スライドショーに退屈感が生じる．このような問題を解決するために，表示時間の短い行は，周辺の行と結合し，1 枚の画像を表示し続け，表示時間の長い行は，分割を行い，複数の画像を表示する．

また，間奏区間における画像切り替えタイミングも設定する．画像の切り替えは，小節の区切りなど楽曲に対して自然なタイミングで行うことが望まれる．また，歌詞の各行のフレーズは，小節の始まり，あるいは，終わりに対応付けられていることが多い，つまり，小節の途中で歌詞の行が切り替わる箇所は少ないという傾向がある．したがって，対象楽

曲の歌詞における画像表示時間の最頻値によって，楽曲に対して適切な画像切り替え間隔が大まかに推定できると考えられる．そこで，本手法では，対象となる歌詞の画像表示時間の最頻値を基にして，間奏区間の画像切り替えタイミングを設定する．

5.2.2. 再構成の手順

以下に画像切り替えタイミング再構成の処理手順について，例を交えて説明する．なお，本手法では，画像の表示時間が短いと判断する閾値を 4[sec]，長いと判断する閾値を 12[sec] と，それぞれ経験的に設定した．

1. 対象となる歌詞の各行における画像表示時間を算出し，それらの最頻値を基本表示時間 *I* として定義する．

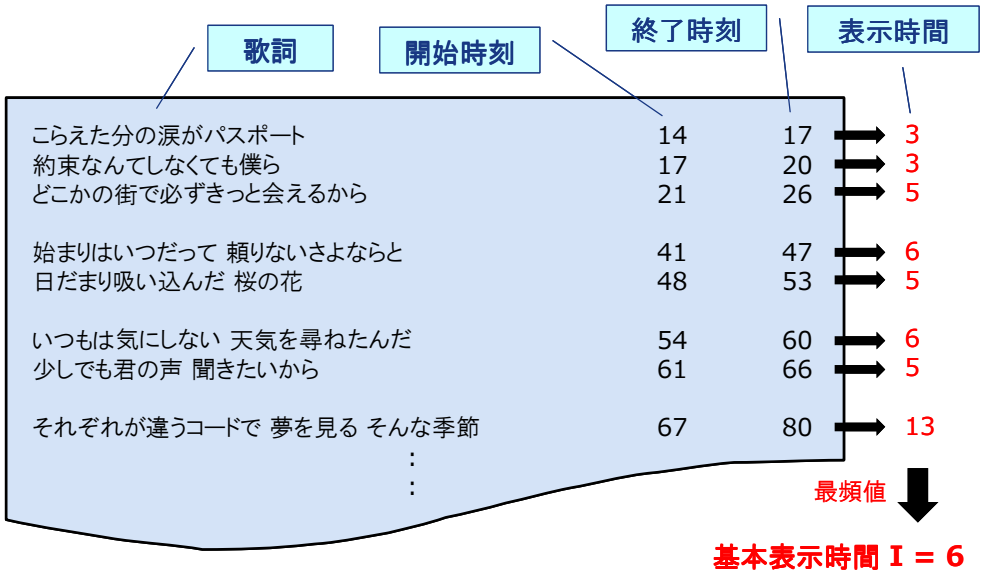

図 5.2.1. 基本表示時間算出

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．

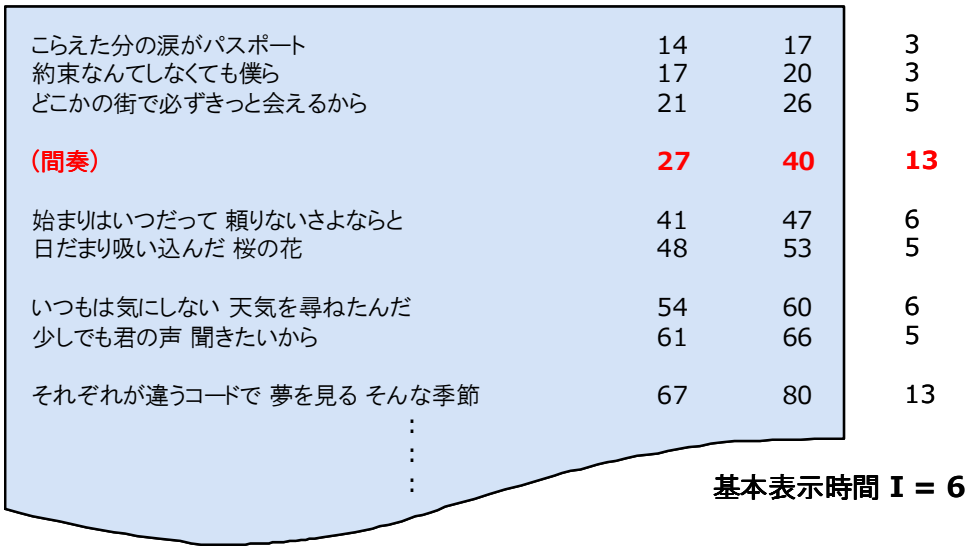

図 5.2.2.　間奏区間抽出

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．ただし，結合は同段落に属する行同士でのみ行う．このように，行が結合された場合，結合後の行に対して画像が 1 枚対応付けられる．

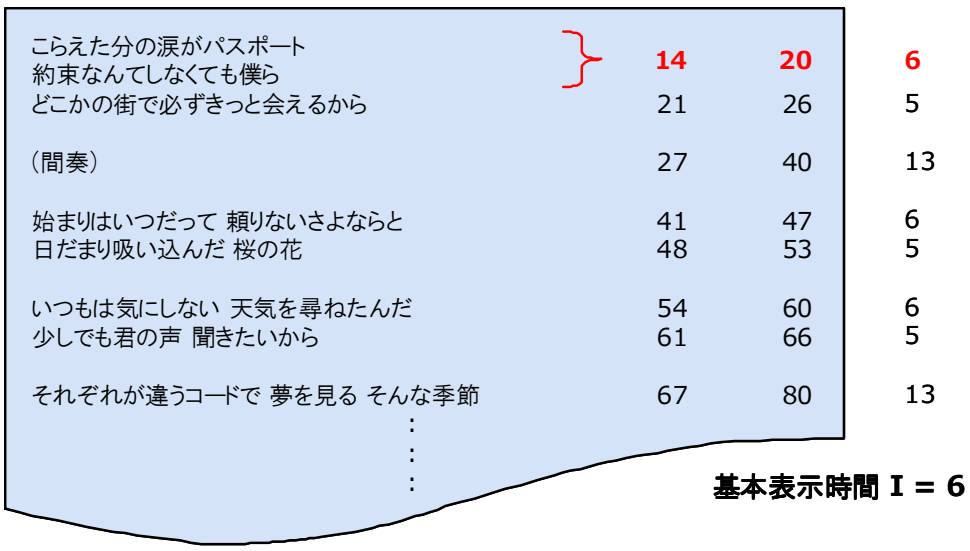

図 5.2.3.　結合処理

4. 表示時間が 12[sec]以上の行を等分割する．ただし，分割後の行における表示時間が，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 枚を選定し表示する．

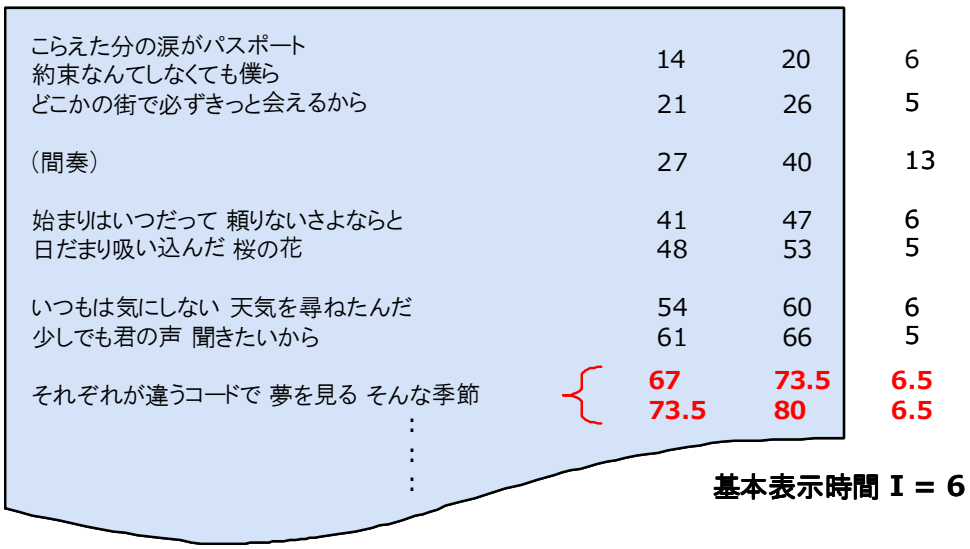

図 5.2.4.　分割処理

5. 間奏区間に対しても，12[sec]以上であれば，分割を行う．なお，間奏区間に表示する画像を取得するための画像検索クエリには，全体印象語を用いる．

以上のようにして，画像切り替えタイミングの再構成を行う．

第6章　評価実験Ⅱ

本章では，5.1.節で述べた，画像選定手法の有効性を検証するための評価実験，および，スライドショーにおいて重要視される特徴を把握するための被験者アンケートについて述べる．

6.1. 実験内容

画像選定手法の違いによる，スライドショーに用いられる画像素材の多様性への影響を評価するための実験を実施した．ここで挙げた多様性とは，複数の楽曲に対してスライドショーを生成した際に，特定の画像が多用されず，様々な画像素材を用いてスライドショーを構成する特徴である．複数楽曲のスライドショーを鑑賞する際，一部の画像のみ頻出することは，スライドショーの新鮮さを損なう原因となるため，用いられる画像が多様であることは重要な特徴だと考えられる．

本実験では，12 曲の楽曲に対して 2 つの画像選定手法によりスライドショーを生成し，それらの画像重複率を算出することで, 画像素材の多様性を評価した. 画像重複率は, (6.1.1)式にて定義される．

$$\text{画像重複率} = \frac{\text{重複して使用された回数}}{\text{使用された全画像数}} \tag{6.1.1}$$

なお，比較対象は，全体印象語との適合度に基づく画像選定手法と，Flickr のランキングに基づく画像選定手法とする．後者の手法では，Flickr における“interestingness”指標を用いて画像検索結果のソートを行い，そのランクの最も高い画像を選定する．

さらに，画像素材の多様性がスライドショーにおいて重要な特徴であることを改めて確認するため，被験者アンケートを実施した．具体的には，以下の質問について，各評価項目に対し，3: 特に重要である，2: 重要である，1: あまり気にしない，の 3 段階で評価する．

Q. 楽曲スライドショーにおいて重要だと思われる特徴は何ですか？
以下の項目について評価して下さい．

1. スライドショーと楽曲の内容が合っている．（内容）
2. 同じ画像ばかりでなく，様々な画像が使用される．（多様性）
3. 各画像が表示される時間が適切である．（表示時間）
4. 楽曲に対して自然なタイミングで画像が切り替わる．（タイミング）
5. スライドショーに何らかの一貫したテーマ性や統一感がある．（統一感）
6. 人物の写っている画像が使用される．（人物）
7. 画質の高い画像が使用される．（画質）
8. 単に画像を切り替えるだけでなく，何かしらの演出効果がある．（演出）

これらの評価項目は，4.7.節にて述べた被験者アンケートの結果と，従来研究において重要視されている特徴を基にして設定した．なお，本アンケートの対象被験者は，大学生 20 名である．

6.2. 実験結果

表 6.2.1.に，歌詞全体の印象との適合度に基づく手法（Impression-based）と Flickr のランキングに基づく手法（Flickr-based）に対する画像重複率を，それぞれ示す．

表 6.2.1. 各手法における画像重複率

手法	Impression-based	Flickr-based
重複率	10.8%	27.1%

表に示す実験結果は，12 曲の楽曲に対してスライドショーを生成した場合，Flickr のランキングを用いて画像選定を行うと，約 3 分の 1 が 1 度使用された画像が提示されるのに対し，提案手法により画像選定を行うと，この値が約 10 分の 1 まで改善できることを示している．ゆえに，提案手法により画像選定を行うことで，特定の画像が頻繁に使用される問題を緩和し，より多様なスライドショーを生成することができるといえる．

6.3. アンケート結果

図 6.3.1.に，被験者アンケートの各評価項目における評価結果を示す．図中では，横軸に示す評価項目に対して，被験者 20 名が与えた評価の分布を示している．

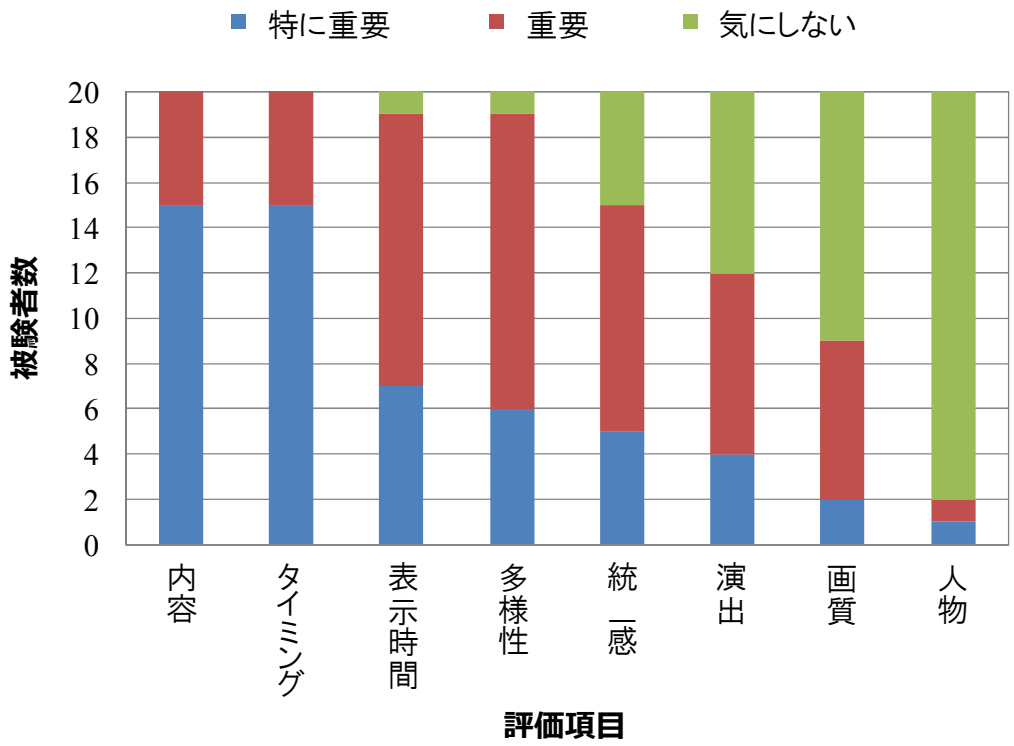

図 6.3.1.　被験者アンケート結果

本アンケートの結果より，まず，“多様性”に対して，“特に重要である”，または，“重要である”，と回答した被験者が 20 名中 19 名であることから，“多様性”という特徴は楽曲スライドショーにおいて重要視される特徴であるといえる．

同様に，“内容”，“タイミング”，“表示時間”に関して，被験者ほぼ全員が，“特に重要である”，または，“重要である”，と回答していることからも，これらの特徴の向上が，良いスライドショーを生成するための要因であるといえる．

一方で，文献[13]の手法では，スライドショーを構成する素材として，人物が写っている画像を優先して用いているが，本アンケート結果より，“人物”の特徴は，20 名中 18 名において，“あまり気にしない”，と評価されていることから，さほど重要視すべき特徴ではないといえる．

第7章　評価実験Ⅲ

本章では，5.2.節で述べた手法による画像切り替えタイミング再構成処理の有効性の評価と，それによるスライドショー全体への影響を評価する目的で実施した被験者評価実験について述べる．

7.1. 実験内容

画像を切り替えるタイミングを再構成することの有効性を評価するための被験者評価実験を実施した．被験者は，画像切り替えタイミングを再構成する前と後のスライドショーを，楽曲の 1 コーラス分鑑賞し，それぞれに対して以下の 2 項目について 5 段階評価を行う．

Q1. スライドショーの完成度を評価して下さい.
- 5: 完成度が高い
- 4: どちらかといえば完成度が高い
- 3: どちらともいえない
- 2: どちらかといえば完成度が低い
- 1: 完成度が低い

Q2. 各画像を表示する長さ（時間）は適切でしたか？
- 5: 大抵の画像において適切だった
- 4: どちらかといえば適切な画像が多かった
- 3: どちらともいえない
- 2: どちらかといえば不適切な画像が多かった
- 1: 大抵の画像において不適切だった

評価はスライドショーを鑑賞した直後に行う．なお，スライドショーを提示する順序効果は考慮してある．

さらに，これらの評価を行った後に，被験者アンケートを実施した．アンケート内容を以下に示す．

Q1. 本システムを利用してみて，今後実際に利用したいと思いましたか？
[いずれか 1 つ選択]
○ 是非使いたい
○ 機会があれば使いたい
○ どちらともいえない
○ あまり使いたいとは思わない
○ 使いたくない
Q2. 本システムの優れていた点があればご記入お願いします. [自由記述形式]
Q3. 本システムの改善すべき点があればご記入お願いします. [自由記述形式]

7.2. 実験データ

本実験の被験者は大学生 20 名で，対象楽曲データは表 7.2.1.，表 7.2.2.に示す J-POP 楽曲 20 曲を用い，1 曲につき 10 名分の評価情報を収集した．なお，実験対象楽曲は，画像切り替えタイミングの再構成において，主に結合処理を施した楽曲 10 曲（結合楽曲群，表 7.2.1.）と，主に分割処理を施した楽曲 10 曲（分割楽曲群，表 7.2.2.）により構成される．

表 7.2.1. 実験対象楽曲（結合楽曲群）

Music ID	タイトル	アーティスト	全体印象語
9	KYOTO	JUDY AND MARY	春
51	太陽の子供	関ジャニ∞	晴れ
181	Escape	ZEEBRA	晴れ，夏
196	HEART OF SWORD ～夜明け前～	T.M.Revolution	夜
476	WIND	倖田來未	晴れ
487	CRAZY GONNA CRAZY	AAA	夜
509	come again	m-flo	夜
1471	別れても好きな人	つじあやの	雨
1695	Another Days	w-inds.	夕方
1708	サボテン	ポルノグラフィティ	雨

表 7.2.2. 実験対象楽曲（分割楽曲群）

Music ID	タイトル	アーティスト	全体印象語
39	Saturday	コブクロ	夜，雨
79	同じ月を見てた	GOING UNDER GROUND	夜，冬
129	君の名を呼ぶ	浜田省吾	夏，夜
274	Everlasting	BoA	冬
339	あられ	aiko	夕方
444	Miracles	平井堅	夜
546	東京ひとり	真心ブラザーズ	夏
1514	Everything	Misia	夜
1676	Pika★★Nchi Double	嵐	夕方
1743	光と影	TUBE	雨

7.3. 実験結果

図 7.3.1.～図 7.3.4.に，各楽曲の評価項目ごとの被験者 10 名における評価値平均を示す．結合楽曲群に対する Q1 の評価結果を図 7.3.1.に，Q2 の評価結果を図 7.3.2.に示し，同様に，分割楽曲群に対する Q1 の評価結果を図 7.3.3.に，Q2 の評価結果を図 7.3.4.に，それぞれ示す．

まず，Q2 に対する評価結果（図 7.3.2.，図 7.3.4.）に着目する．対象楽曲 20 曲における Q2 に対する評価値平均は，再構成前では 3.66，再構成後では 4.29 となった．また，楽曲ごとに再構成前と後の評価値平均を比較したところ，20 曲中 15 曲において再構成後の評価値の方が高くなっていた．さらに，楽曲ごとに再構成前と後の評価値平均を，t 検定により検証した結果，ID:181，476，1695，1708，129，339，546，1514 の楽曲において，再構成後の方が有意に高いと認められた一方，再構成前の方が有意に高いと認められた楽曲はなかった（$p<0.10$）．このような結果から，本手法によって画像の切り替えるタイミングを再構成することで，画像の表示時間を適切に設定できていることが示せた．

次に，Q1 に対する評価結果（図 7.3.1.，図 7.3.3.）に着目する．対象楽曲 20 曲における Q1 に対する評価値平均は，再構成前では 3.69，再構成後では 3.85 となった．また，楽曲ごとに再構成前と後の評価値平均を比較したところ，20 曲中 13 曲において再構成後の評価値の方が高くなっており，1 曲において同等の評価値であった．さらに，再構成後の評価の方が高かった 13 曲に関しては，Q2 についても同様に，再構成による評価の向上が見られた．このような結果から，画像の表示時間を適切に設定することで，スライドショー全体の評価を向上することができるといえる．なお，Q1 における再構成による改善幅（再構成前と後の評価値の差）が Q2 のそれと比較して小さくなっているのは，スライドショーの完成度

に影響を与える要因には様々なものがあり，画像の表示時間の適切性はその中の一要因に過ぎないため，画像の表示時間の改善幅がそのままスライドショーの完成度の評価に直結していないのだと考えらえる．

なお，結合楽曲群と分割楽曲群の間においては，評価値（Q1，Q2 とも）に大きな差は見られなかった．

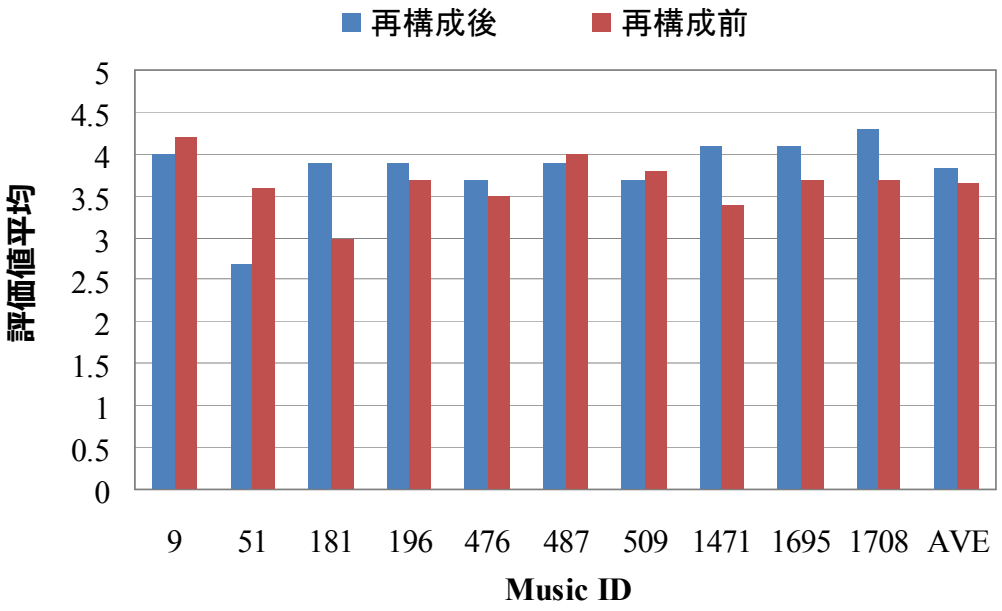

図 7.3.1.　Q1 に対する評価値平均（結合楽曲群）

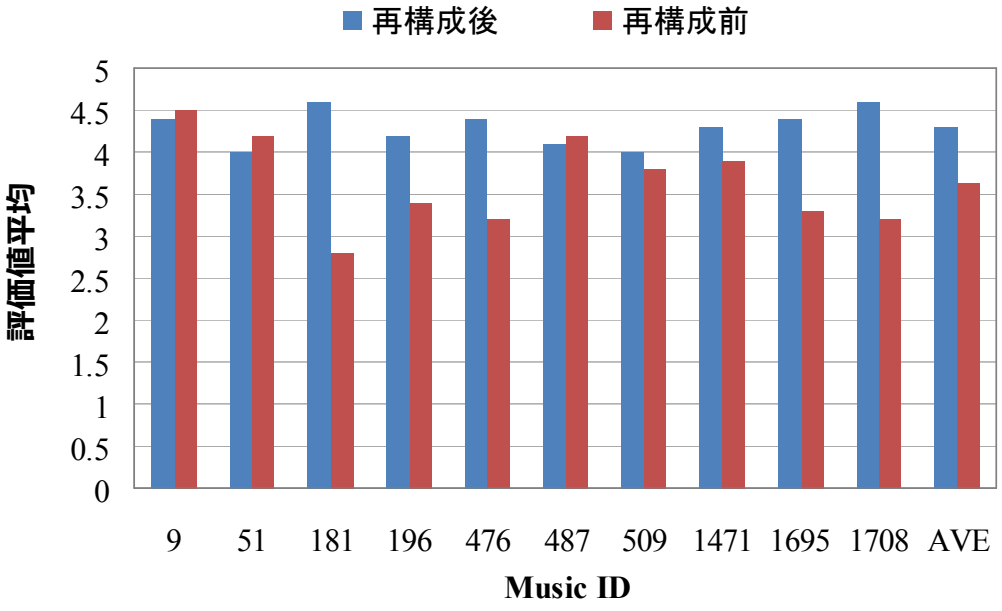

図 7.3.2.　Q2 に対する評価値平均（結合楽曲群）

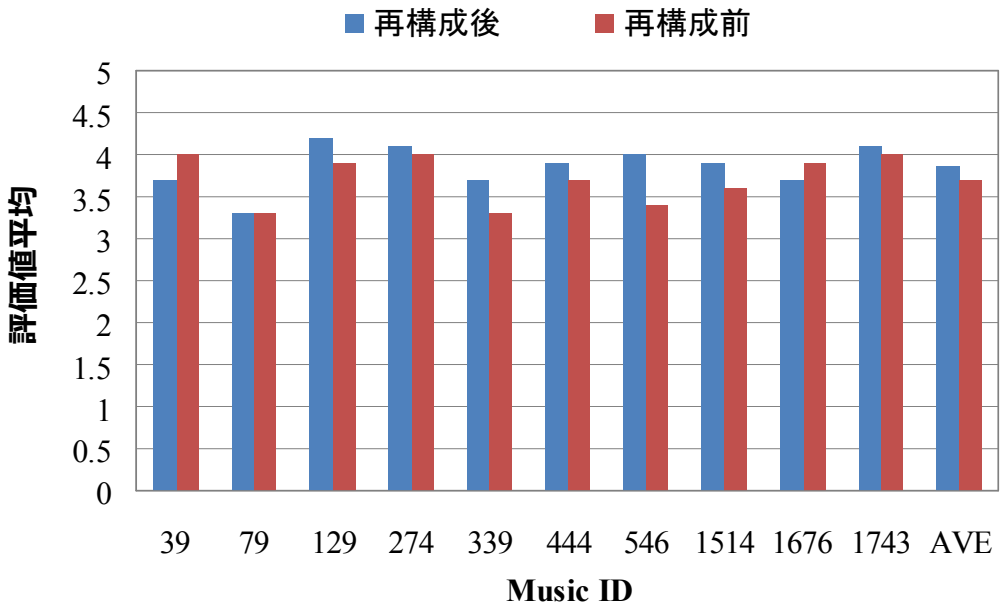

図 7.3.3.　Q1 に対する評価値平均（分割楽曲群）

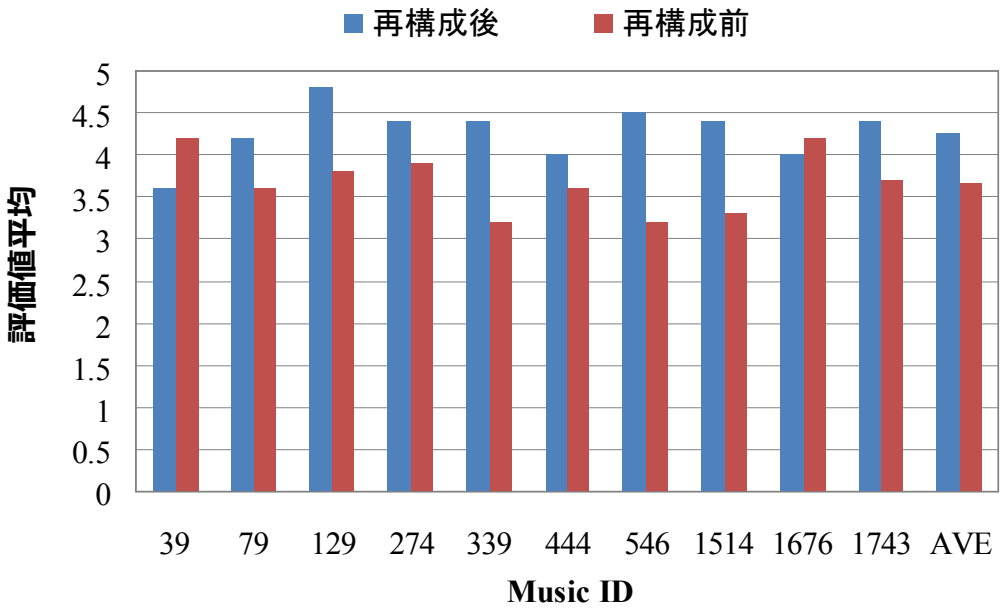

図 7.3.4.　Q2 に対する評価値平均（分割楽曲群）

7.4. 考察

Q2 に対する評価結果において，個々の楽曲に着目する．

まず，ID:181 の楽曲では，対象楽曲の中で再構成により最も大幅に評価の向上が見られた．ID:181 は，ラップが中心の楽曲であり，表 7.4.1.に示すように，再構成前における 1 コーラスの平均画像表示時間は，対象楽曲中最小である 2.52[sec]であった．そのため，短時間で頻繁に画像が切り替わるスライドショーを構成していた．同様に，ID:129，339 の楽曲においても，再構成によって大幅に評価が向上しており，これらの再構成前における平均画像表示時間がそれぞれ 17.00[sec]，16.43[sec]となっていた．これらは，対象楽曲中で値が大きい上位 2 つである．以上より，平均画像表示時間が極端に短い，または，長い楽曲においては，再構成によって大きな改善効果が得られるといえる．

逆に, ID:39 の楽曲では, 再構成前の評価値が再構成後と比較して, 大きく上回っている．この楽曲は, 表 7.4.1.に示すように, 再構成前における平均画像表示時間は 14.17[sec]であり, 再構成後が 7.0[sec]となっている．本手法では，画像の表示時間が長いと判断する閾値を 12[sec]と設定しているが，本評価結果より，ID:39 の楽曲に対しては，この閾値が適切でないといえる．ここで，本手法における画像の表示時間が長い，または，短いと判断する閾値の適切性を確認する．図 7.4.1.に，各スライドショーに対する評価値平均とその平均画像表示時間（1 コーラス分）との関係を示す．図中の赤線が，表示時間が長いと判断する閾値（12[sec]）と短いと判断する閾値（4[sec]）を示しており, 緑点線にて示したプロットが ID:39 を示している．

表 7.4.1.　一部楽曲における平均画像表示時間と Q2 評価値平均

Music ID	再構成	平均画像表示時間（1 コーラス）[sec]	Q2 評価値平均
181	前	2.52	2.8
	後	6.93	4.6
509	前	2.74	3.8
	後	6.65	4.0
39	前	14.17	4.2
	後	7.00	3.6
129	前	17.00	3.8
	後	8.29	4.8
339	前	16.43	3.2
	後	6.32	4.4

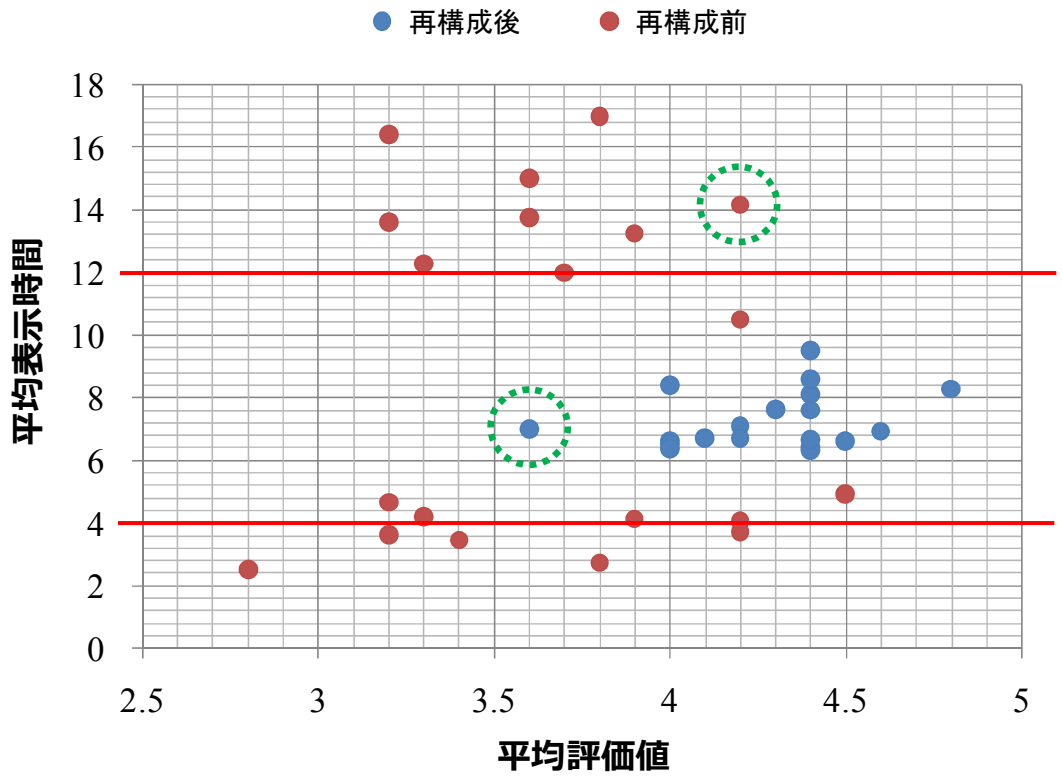

図 7.4.1.　評価値と平均表示時間の関係

図より，赤線の間に位置している多くのプロットにおいて，平均評価値が 4 以上となっており，それ以外の範囲に位置しているものの多くは評価値が低くなっている．一方，図に示す ID:39 の楽曲のみにおいて，この傾向とは逆の結果となっている．このように，多くの楽曲において，平均画像表示時間が 4[sec]～12[sec]の場合に評価値が高くなっていることから，今回設定した閾値は，妥当なものであると判断できる．ID:39 の楽曲のみが上記のような傾向を示す原因として，楽曲の曲調や雰囲気による影響が考えられる．つまり，ID:39 の楽曲のように，スローテンポなバラード曲では，表示時間が長くても楽曲の雰囲気と合っていると判断されることがあると考えられる．

さらに，ID:509 の楽曲では，再構成前の画像平均表示時間が 2.74[sec]と，対象楽曲の中で 2 番目に短いにも関わらず，再構成による改善効果がさほど大きくない．これは，楽曲内における画像の表示時間の変化が，その楽曲の場面変化を捉えていると解釈されたことが，原因の 1 つとして挙げられる．ID:509 の楽曲には，ラップの部分とそうでない通常の歌唱の部分があり，再構成を行わないスライドショーでは，ラップの箇所では画像が頻繁に切り替わり，通常の箇所では比較的長い時間画像が表示されるという特徴がある．このような特徴を，“楽曲の雰囲気の変化を的確に反映している”と良い傾向に捉えることにより，再構成前のスライドショーが若干高く評価されることで，再構成による改善効果が小さく

なっていると考えられる．実際に，被験者アンケート（Q2）においても，上記のような特徴を重要視する意見があった（20 名中 2 名）．しかし一方で，同アンケート（Q3）において，画像が頻繁に切り替わる特徴を問題視する意見も見られた（20 名中 4 名）．このように，“画像の表示時間”か“楽曲の場面変化”のどちらの特徴を重要視するかは，被験者によってまちまちであると考えられる．そのため，画像切り替えタイミングを再構成することにより，“画像の表示時間”を適切に設定した後，ズームやパンなどのスライドショーの表示における効果によって“楽曲の場面変化”を表現するなど，両特徴を同時に考慮できる対策が望まれる．

7.5. アンケート結果

被験者アンケートの結果を以下に示す．

- Q1 について

Q1 に対する回答結果を表 7.5.1.に示す．

表 7.5.1. Q1 に対する回答結果

選択肢	回答人数
是非使いたい	5 名
機会があれば使いたい	15 名
どちらともいえない	0 名
あまり使いたいとは思わない	0 名
使いたくない	0 名

本結果において，全ての被験者が，是非使いたい，または，機会があれば使いたい，と回答していることから，本システムが，ユーザの利用意欲をわきたてる新しい音楽の楽しみ方を提供できているといえる．

- Q2 について

Q2 については，4.7.節における回答結果とほぼ同様であったため，割愛する．

- Q3 について

Q3 に対する回答結果から主なものを以下に示す.

a)	明らかに楽曲に合わない画像は排除した方が良い.
b)	楽曲に対して不自然なタイミングで画像が切り替わる箇所がある.
c)	表示時間が長い画像があると，だらけてしまう.
d)	歌詞の中に人物を示す単語がある場合は，人物画像が表示された方が良い.
e)	画像が短時間で切り替わると見ていて疲れる.
f)	歌詞のストーリー性が表現できていない.

b)に関しては，画像切り替えタイミング再構成の際に，楽曲ごとの基本表示時間を考慮することで，楽曲に対して適切なタイミングを推定しているが，やはり，歌詞情報のみでは完全な推定は困難であると考えられる．したがって，基本表示時間のみではなく，音響特徴解析によって得られるビート位置などの情報と組み合わせることにより，更なる改善が期待できる．

また，d)に関して，6.3.節における被験者アンケートでは，“人物の写っている画像が多く用いられる”特徴は，さほど重要でないという結果であったが，歌詞中に人物が登場していることを示唆する単語がある場合においては，人物の写っている画像を表示することで，より良いスライドショーが生成できると考えられる．これより，歌詞を解析することで，その場面に人物を示す単語（私，あなた，笑顔など）の有無を判定し，その結果に応じて，人物の写っている画像を提示するか否かを判定する，などの対策が考えられる．

さらに，f)に関して，歌詞全体の印象語との適合度に基づく画像選定手法では，スライドショー全体の統一感は考慮しているが，歌詞におけるストーリー性は十分に考慮できていない．このような，歌詞における話の展開を考慮するためには，歌詞の内容を詳細に把握する必要がある．しかし，テキスト処理の分野でも報告されているように，ニュースや新聞などの，情報を的確に伝達することを目的としたテキストにおいては，内容把握が比較的容易である一方で，歌詞のような芸術的要素を含んだテキストでは，その内容を詳細に把握するのは難しい．そのため，完全なストーリー性の表現は難しいと考えられるが，その実現に向けた第一歩として，歌詞における名詞だけでなく，形容詞や動詞も解析の対象とすることが挙げられる．

なお，c)，e)に関しては，再構成前のスライドショーに対する意見だと考えられる．このように指摘されていることからも，表示時間を適切にすることが，重要であることが改めて示唆される．

第8章　結論

本章では，本研究の内容について簡潔にまとめ，また，更なる発展のために，今後の展望について言及する．

8.1. まとめ

本論文では，視覚と聴覚を刺激する新しい音楽の楽しみ方を，誰でも手軽に体験できるようにすることを目的として，Web 画像を用いた楽曲スライドショーの自動生成システムを提案した．その中でまず，Web 画像検索に用いるクエリの選定法を提案し，被験者評価実験にて，ソーシャルタグの傾向に基づいてクエリを選定することの有効性を示した．さらに，被験者アンケートにて指摘された，“複数の楽曲に対するスライドショーにおいて特定の画像が使用される”，“スライドショー全体の統一感に欠ける”，“間奏区間でも画像を切り替えるべきである”，“画像を切り替えるタイミングが適切でない”という問題点を解決するため，前者 2 項目に関しては，歌詞全体の印象との適合度に基づく画像選定法を提案し，後者 2 項目に関しては，画像を切り替えるタイミングの再構成法を提案した．そして，提案した画像選定法を用いることで，特定の画像が頻繁に使用される問題が緩和できることを評価実験にて証明し，また，提案手法によって画像を切り替えるタイミングを再構成することで，画像の表示時間を適切に設定でき，スライドショー全体の評価の向上につながることを，被験者評価実験により示した．

8.2. 今後の展望

更なるシステムの発展のために，以下のような改善策が挙げられる．

- スライドショーの表示法の工夫

本システムでは，スライドショーの表示に関する工夫として，画像切り替え時にフェード処理を施している．これだけでなく，ズームやパンなどのより多様な効果を加えることで，動きの豊富なスライドショーが生成できると考えられる．さらに，これらの効果を 7.4. 節にて示したように，楽曲の場面変化や盛り上がりに対応付けることで，より楽曲に合ったスライドショーの生成が期待できる．

- 歌詞情報における形容詞や英詞の考慮

本システムでは，歌詞情報における名詞のみを解析の対象としているが，名詞以外にも視覚的に重要な意味をもつ単語が存在すると考えられる．例えば，現状では，“黄色い花”というフレーズが歌詞に出現した場合，“花”の画像は提示することができるが，“黄色い”という情報までは考慮することができない．このように，形容詞を考慮することで，歌詞の内容をより詳細に表現することができる．また，7.5.節に述べたように，詳細な歌詞の内容を把握するためには，動詞の考慮も必要であると考えられる．さらに，現状は考慮していない英詞にも対応することで，スライドショーを生成できる楽曲の幅を広げることができる．

- 音響特徴の考慮

歌詞特徴のみでは抽出できない特徴を，音響特徴により補完することで，楽曲スライドショーの更なる向上が期待できる．例えば，7.5.節に記したように，音響特徴を解析し，ビート位置などの情報を適用することで，より詳細に画像を切り替える自然なタイミングが推定できる．また，歌詞情報と画像のタグによる意味ベースの対応付けに加えて，楽曲の曲調や雰囲気などによる印象ベースの対応付けも考慮し，用いる画像を選択することで，さらに楽曲に適したスライドショーを実現できると考えられる．

謝辞

本研究を行うにあたり，懇切丁寧にご指導賜りました甲藤二郎教授に厚く御礼申し上げます．また，共同研究させて頂いた KDDI 研究所において，多くの御指導と貴重な助言を賜りました，帆足啓一郎氏，石先広海氏，および知能メディアグループの方々に深く御礼申し上げます．そして，本研究における評価実験や，研究以外の面でも大変お世話になりました甲藤研究室の皆様に心から感謝致します．

2009 年 2 月 5 日

舟澤　慎太郎

参考資料

[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.43, No.3, pp.146-153 (1992).

[2] Microsoft Photo Story 3 for Windows: http://www.microsoft.com/windowsxp/using/digitalphotography/PhotoStory/default.mspx

[3] Photo Flash Maker: http://www.anvsoft.com/flash-slideshow-maker.html

[4] shwup: http://www.shwup.com/

[5] animoto: http://animoto.com/

[6] Flickr: http://www.flickr.com/

[7] Picasa: http://picasa.google.com/

[8] X. –S. Hua, L. Lu, and H. –J. Zhang: “P-Karaoke: Personalized Karaoke System”, Proceedings of the 12th Annual ACM International Conference on Multimedia, pp.172-173 (2004).

[9] 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, pp.235-244 (2002).

[10] S. Xu, T. Jin, and F. C. M. Lau: “Automatic Generation of Music Slide Show using Personal Photos”, Proceedings of 10th IEEE International Symposium on Multimedia, pp.214-219 (2008).

[11] Y. –F. Ma, L. Lu, H. –J. Zhang, and M. Li: “A User Attention Model for Video Summarization”, Proceedings of the 10th Annual ACM International Conference on Multimedia, pp.533-542 (2002).

[12] D. A. Shamma, B. Pardo, and K. J. Hammond: “MusicStory: a Personalized Music Video Creator”, Proceedings of the 13th Annual ACM International Conference on Multimedia, pp.563-566 (2005).

[13] R. Cai, L. Zhang, F. Jing, W. Lai, and W. –Y. Ma: “Automated Music Video Generation Using Web Image Resource”, Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing, 2007, Vol.2, pp.737-740 (2007).

[14] R. Xiao, M. –J. Li, and H. –J. Zhang: “Robust Multipose Face Detection in Images”, IEEE Transactions on Circuits and Systems for Video Technology, Vol.14, No.1, pp.31-41 (2004).

[15] L. Zhang, M. –J. Li, and H. –J. Zhang: “Boosting Image Orientation Detection with Indoor vs. Outdoor Classification”, Proceedings of 6th IEEE Workshop on Applications of Computer Vision, pp.95-99 (2002).

[16] L. Lu, D. Liu, and H. –J. Zhang: “Automatic Mood Detection and Tracking of Music Audio Signals”, IEEE Transactions on Audio, Speech, and Language Processing, Vol.14, No.1, pp.5-18(2006).

[17] X. –S. Hua, L. Lu, H. –J. Zhang: “Automatically Converting Photographic Series into Video”, Proceedings of the 12th Annual ACM International Conference on Multimedia, pp.708-715 (2004)

[18] Last.fm: http://www.lastfm.jp/

[19] Yahoo!デベロッパーネットワーク - テキスト解析 - 日本語形態素解析: http://developer.yahoo.co.jp/webapi/jlp/ma/v1/parse.html

[20] 舟澤慎太郎, 石先広海, 帆足啓一郎, 滝嶋康弘, 甲藤二郎: “歌詞の印象に基づく楽曲検索のための楽曲自動分類に関する検討”, 第 71 回情報処理学会全国大会, 5R-2 (2009).

[21] C. Cortes and V. Vapnik: “Support-Vector Networks”, Machine Learning, Vol.20, No.3, pp.273-297 (1995).

[22] SVMLight: http://svmlight.joachims.org/

[23] 舟澤慎太郎, 石先広海, 帆足啓一郎, 滝嶋康弘, 甲藤二郎: “歌詞特徴を考慮した Web 画像と楽曲同期再生システムの提案”, 第 8 回情報科学技術フォーラム, E-034 (2009).

発表文献

[1] 舟澤慎太郎, 北市健太郎, 甲藤二郎: “楽曲推薦システムのための楽曲波形と歌詞情報を考慮した類似楽曲検索に関する一検討”, 情報処理学会研究報告, 2008-AVM-060, No.1 (2008).

[2] S. Hamawaki, S. Funasawa, J. Katto, H. Ishizaki, K. Hoashi, and Y. Takishima: “Feature Analysis and Normalization Approach for Robust Content-based Music Retrieval to Encoded Audio with Different Bit Rates”, Advances in Multimedia Modeling: Proceedings of the 15th international multimedia modeling conference, pp.298-309 (2009).

[3] 舟澤慎太郎, 石先広海, 帆足啓一郎, 滝嶋康弘, 甲藤二郎: “歌詞の印象に基づく楽曲検索のための楽曲自動分類に関する検討”, 第 71 回情報処理学会全国大会, 5R-2 (2009).

[4] 舟澤慎太郎, 石先広海, 帆足啓一郎, 滝嶋康弘, 甲藤二郎: “歌詞特徴を考慮した Web 画像と楽曲同期再生システムの提案”, 第 8 回情報科学技術フォーラム, E-034 (2009).

[5] 舟澤慎太郎, 石先広海, 帆足啓一郎, 滝嶋康弘, 甲藤二郎: “歌詞情報を利用した Web 画像･楽曲連動スライドショー自動生成システム”, 情報処理学会研究報告, 音楽情報科学研究会 (2010).

[6] 舟澤慎太郎, 石先広海, 帆足啓一郎, 滝嶋康弘, 甲藤二郎: “歌詞情報を利用した Web 画像・楽曲連動スライドショー自動生成システムにおける画像切り替え間隔の改善”, 電子情報通信学会 2010 年総合大会 (2010).

[7] S. Funasawa, H. Ishizaki, K. Hoashi, Y. Takishima, and J.Katto: “Automated Music Slideshow Generation using Web Images Based on Lyrics”, The 11th International Society for Music Information Retrieval Conference (2010). [投稿予定]