

Songle: Web上の楽曲の中身を音楽理解技術で推定する 能動的音楽鑑賞サービス

後藤 真孝^{1,a)} 吉井 和佳^{1,b)} 中野 倫靖^{1,c)}

概要: 本稿では、音楽音響信号理解技術によって音楽の聴き方をより豊かにするための能動的音楽鑑賞サービス Songle について、2012年8月以降の主な進展を報告する。Web上の音楽コンテンツとして、従来のMP3形式だけでなく動画共有サービス上の動画形式の楽曲にも対応し、その音楽情景記述(サビ、ビート、メロディ、コード)を音楽理解技術で推定して、従来の「音楽地図」に「ビジュアルライザ」も加えた二種類の表示方法で可視化可能になった。自動推定の誤りを人手で訂正するインターフェースを提供してきたが、同一の楽曲が複数登録されていても、それらを自動的に紐付けることで、その一つをユーザが訂正するだけでよくなった。さらに、コード進行の検索機能や、音楽情景記述を他のWebサービスから活用して連携できる外部埋め込みプレーヤ機能、照明等の実世界デバイスを音楽に連動して制御する機能も実現した。今後もユーザの役に立ちつつ音楽理解技術の普及と発展を目指していく。

1. はじめに

音楽理解技術は、人々の音楽の聴き方を豊かにすることができる。それを実証し、日常生活で人々に体験してもらうために、本研究では能動的音楽鑑賞サービス「Songle(ソングル)」(<http://songle.jp>)を開発し[1-5]、2012年8月29日に誰でも利用できるWeb上のサービスとして一般公開した。これは、音楽理解技術(音楽の音響信号中の様々な要素を自動的に推定できる技術)を用いて計算機が膨大なコンテンツを「聴く」ことで、人々の鑑賞やブラウジングを支援するサービスである。

当初の文献[1-3]の時点のSongleでは、MP3形式の音響信号ファイルとして公開されているWeb上の任意の楽曲を対象としていた(図1)。4種類の音楽情景記述(楽曲構造、階層的なビート構造、メロディライン、コード)を自動推定し、それらを「音楽地図」として眺めることができる可視化機能や、興味のある箇所を容易に見つけられるサビ出し機能を実現した。しかも、ユーザが音楽地図を改善できる誤り訂正機能も実現した。これは、自動推定では誤りが不可避であるため、効率的な誤り訂正インターフェースをWeb上で提供することで、誤りを人手で自発的に訂

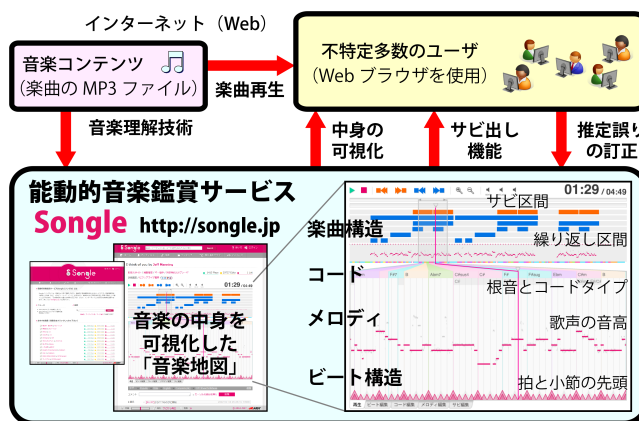


図1 音楽理解技術を活用した能動的音楽鑑賞サービス「Songle」

正してもらおう試みである。

その後、2012年8月29日にプレス発表[6]をした「2012年8月版Songle」では、再生した楽曲の進行に連動した様々なアニメーションを楽しめるビジュアルライザ画面を新たに表示できるよう拡張した。さらに、同一のコード進行をもつ複数の楽曲を聴き比べることができるコード進行検索機能や、ユーザが自分のホームページやブログなどの外部のWebページ内にSongleの小型プレーヤを埋め込んで、Songle上の楽曲を紹介できる外部埋め込みプレーヤ機能を実現した。コンテンツ投稿サイト「ピアプロ」(<http://piapro.jp>)と正式に連携したことでピアプロ上の全楽曲からSongleへのリンクが設置され、「SoundCloud」(<http://soundcloud.com>)上の楽曲の登録にも対応した。

¹ 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

a) m.goto [at] aist.go.jp

b) k.yoshii [at] aist.go.jp

c) t.nakano [at] aist.go.jp



図 2 Songle のタイトルページの画面表示例。下側に類似楽曲グラフも表示されている。

しかし、MP3 形式の楽曲に対象が限定されており、動画形式の楽曲には対応していなかった。

そこで今回、2013 年 8 月末に公開する「2013 年 8 月版 Songle」では、新たに、代表的な動画共有サービスの「ニコニコ動画」(<http://www.nicovideo.jp/>)と「YouTube」(<http://www.youtube.com/>)に対応する。動画共有サービスには同一の楽曲が違う映像で複数登録されていることがあるが、それらの関係を自動検出して紐付け、そのうちの一つの自動推定誤りをユーザが自発的に訂正すれば、それら複数の同一楽曲動画に対する音楽地図がすべて改善される仕組みも導入する。さらに、ビジュアライザ機能で画面上で音楽に連動した図形が動くだけでなく、照明等の実世界デバイスも音楽に連動して制御できる機能を実現する。

本稿では、2 章で文献 [1-3] の時点の Songle の基本機能について紹介した後に、3 章と 4 章で、これまで文献 [1-3] では説明していなかった「2012 年 8 月版 Songle」と「2013 年 8 月版 Songle」の新機能について述べる。そして、5 章で大規模な楽曲数に対応するための実装上の改善について述べる。最後に、6 章で Songle の対外的な連携と科学館等での展示について紹介する。

2. 能動的音楽鑑賞サービス Songle の基本機能

Web ブラウザ上で能動的音楽鑑賞インタフェース [7,8] を使用できる Web サービス「Songle」(図 2)では、歌声を伴うポピュラー音楽の楽曲を主な対象として、

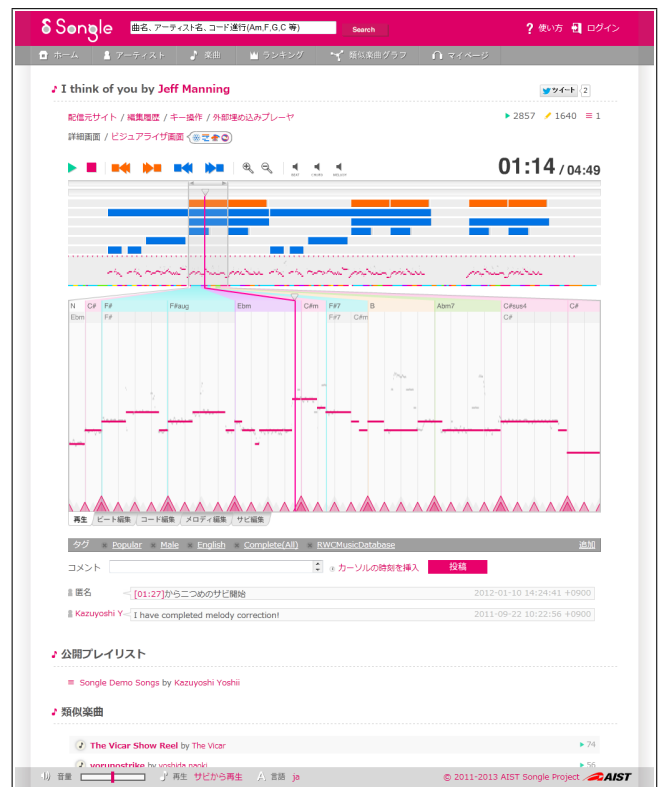


図 3 Songle の中心的な機能である音楽再生用インタフェースの画面表示例。楽曲の中身(音楽情景記述)を「音楽地図」として可視化したユーザインタフェース上で、再生位置を自在に制御できる。横軸が時間であり、上部の大局的な表示部には、楽曲全体の構造が表示され、下部の局所的な表示部には、そこで選択した区間が拡大表示されている。

- (i) 楽曲構造 (サビ区間と繰り返し区間)
- (ii) 階層的なビート構造 (拍と小節の先頭)
- (iii) メロディライン (メロディの歌声の基本周波数 (F0))
- (iv) コード (根音とコードタイプ (構成音))

の 4 つの代表的な音楽情景記述 (音楽的要素) [9-11] を音楽理解技術を用いて自動推定する。Songle のユーザは、それらの音楽的要素を把握しやすい「音楽地図」として可視化した詳細画面 (図 3) を見ながら、楽曲の再生を楽しむことができる。こうした可視化は、「音楽理解力拡張インタフェース」[12] の観点からも重要であり、専門的知識のないユーザでも、各音楽的要素の存在や要素間の関係、楽曲構成上の意図に気づきやすくなる。例えば、サビの繰り返しやイントロとエンディングの繰り返しなどの楽曲全体の構造を把握したり (サビが例外的に多く繰り返す曲や、サビから始まる曲に容易に気づくことができる)、同じハーモニー (コード進行) なのにメロディが変化する様子に気づいたり、繰り返すときの歌詞や曲調の変化を聴き比べたりすることもできる。このように、再生に同期して推定結果を「見る」ことで音楽の理解を深めることができる。

音楽内容に基づくブラウジングも可能であり、Songle では、サビ出し機能付き能動的音楽鑑賞インタフェース「SmartMusicKIOSK」[13] の機能を実装した。ユーザがサ



図 4 サビ出し機能の使用例. 楽曲中の興味のある箇所を容易に見つけて聴くことができるサビ出し機能では, 楽曲構造に対応した 4 種類のボタン「次のサビ区間の頭出し」, 「前のサビ区間の頭出し」, 「次の繰り返し区間の頭出し」, 「前の繰り返し区間の頭出し」で再生位置のジャンプができる. 各区間を直接押してもよい.

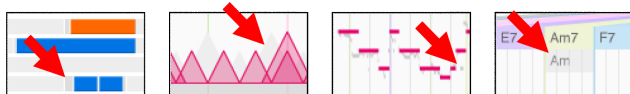


図 5 誤り訂正後に残っている元の自動推定結果の跡 (グレーで着色)

サビ出しボタンを押すことで, サビに飛んで聞くことができる (図 4). このように Songle では, 楽曲中の興味のある箇所を容易に見つけて聞くことができる.

さらに, 音楽情景記述の推定誤りを誰でも容易に訂正して貢献可能なインタフェース (図 12) も Songle 上で提供した. Songle のユーザは推定誤りを見つけたら, 自動生成された候補から選んだり, 直接編集したりして自発的に訂正する. その結果は他のユーザと共有されて, 即座にユーザ体験の向上に資することができる. これは, 音楽理解技術が不十分であっても, ユーザの貢献によってユーザ自身が利便性を感じられる仕組みである. ただし, 音楽理解技術の性能が過大評価されないように, ユーザが誤り訂正すると, 元の自動推定結果は違う色 (図 5 のグレー) で着色されて跡が残るように工夫した. これにより, ユーザは訂正された箇所を容易に区別できる. さらに, すべての訂正履歴は記録されており, 誰でも訂正前後の比較をしたり, 過去の任意の時点に戻したりできる機能も用意した.

楽曲は, 曲名やアーティスト名をテキスト検索したり, ランキング等の各種一覧表示から選択したりするだけでなく, 歌声の声質の類似楽曲グラフを表示して (図 2), そこからの選択も可能にした. 声質の類似度に基づく楽曲検索が可能な能動的音楽鑑賞インタフェース「VocalFinder」[14]の発想を取り入れた機能である.

誰でも Songle にログインせずに匿名で利用できるが, OpenID や OAuth 等を用いてログインすると, 任意の楽曲の MP3 ファイルが置かれている URL を指定して, Songle に登録できる. ユーザが Songle を使用しているときには, その元の Web サイト上にある楽曲が, Songle を経由せずに直接ユーザのブラウザ上でストリーミング再生される.



図 6 2012 年 8 月版 Songle のビジュアルライザ画面の表示例

3. 2012 年 8 月版 Songle の新機能

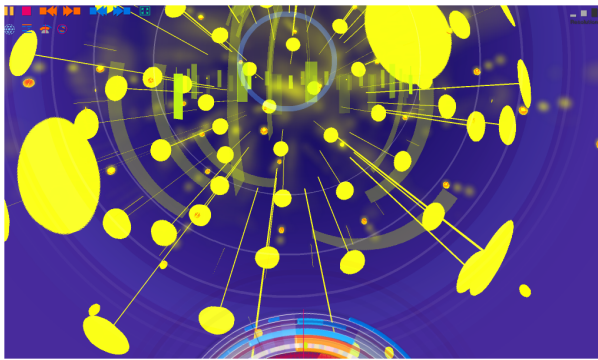
上記の Songle から拡張し, 「2012 年 8 月版 Songle」では, ビジュアルライザ画面表示機能, コード進行検索機能, 外部埋め込みプレーヤ機能の三つを追加した.

さらに, 音楽に関するコンテンツ投稿サイト「ピアプロ」(<http://piapro.jp>) と「SoundCloud」(<http://soundcloud.com>) 上の楽曲も登録可能なように拡張した.

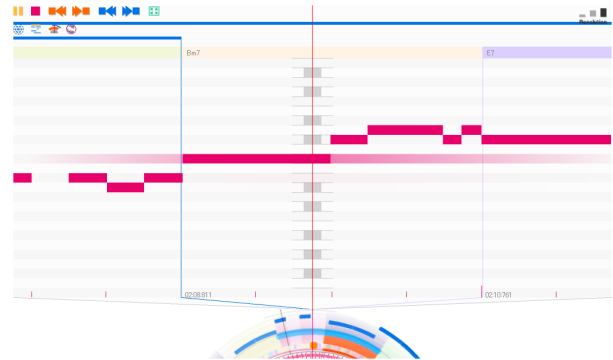
3.1 ビジュアルライザ画面表示機能

音楽の中身の可視化では, これまでの「音楽地図」を表示する詳細画面 (図 3) に加え, アニメーションによる大きな動きを伴って表示内容が動的に変わるビジュアル画面 (図 6, 図 7) の表示も可能にした. いずれも, 同じ 4 つの音楽情景記述 (音楽的要素) に密接に連動して表示内容が変わる.

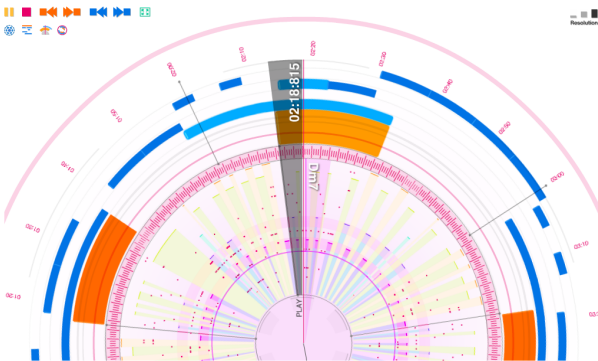
ビジュアルライザ画面では, 図 7 の 4 種類の表示形式を切り替えることができる. 図 7(a) は幾何学模様が連動して大きく動く表示で, サビは背景の色, ビートは幾何学模様が伸縮する動き, メロディは中央の 1 オクターブの鍵盤状の長方形の動き, コードは幾何学模様の色に対応している. 図 7(b) はメロディラインのピアノロール状の表示で, 右から左にスクロールして中央の鍵盤をメロディに対応した長方形が横切り, サビになるとスクロールスピードが速くなる. 図 7(c) は半円状に描かれた音楽的要素の表示で, 上記の (a) と (b) では下に隠れている円盤がせり上がった状態である. 円盤には, 音楽地図に対応した内容が表示されており, 音楽の再生に応じて反時計回りに回転する. サビのオレンジ色の区間をクリックするとそこへジャンプができ, ドラッグすることで円盤を手動で回転してシークすることができる. 図 7(d) は全体を俯瞰した円盤状の表示



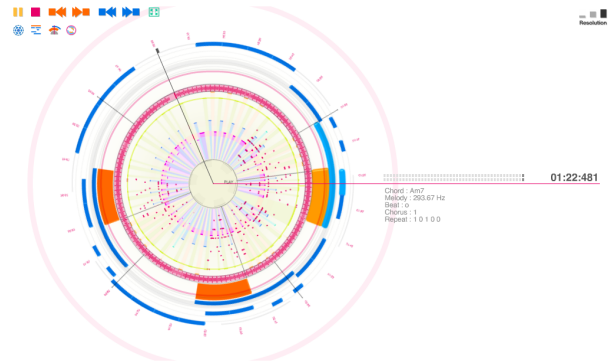
(a) 幾何学模様が連動して大きく動く表示



(b) メロディラインのピアノロール状の表示



(c) 半円状に描かれた音楽的要素の表示



(d) 全体を俯瞰した円盤状の表示

図 7 楽曲の中身をアニメーション表示するビジュアライザ画面での 4 種類の表示形式

で、(c)の半円状の円盤の全体を見ることができる。同様に円盤の操作で再生位置のジャンプやシークができ、1周が楽曲の長さに対応している。

詳細画面が音楽的要素を把握するためのインターフェースであるのに対し、ビジュアライザ画面は音楽的要素に基づいて動的に生成される表示を楽しんでもらうためのインターフェースとなっている。既存のメディアプレーヤ等に搭載されているビジュアライザでは、波形や周波数成分を単純に反映して表示していることが多いが、Songleのビジュアライザ画面では、楽曲の中身により密接に連動して表示が変化する点が異なる。

3.2 コード進行検索機能

曲名やアーティスト名といった書誌情報に基づく従来の音楽情報検索に加えて、新たに、コード名の系列を与えるとそれをコード進行として含む楽曲群を検索・列挙する音楽情報検索を可能にした(図8)。このコード進行検索機能により、同一のコード進行をもつ複数の楽曲を聴き比べることができる。

コード進行の多くは人類共有の知と言えるぐらい、様々な楽曲で同じコード進行が繰り返し使用されている。この機能により、専門的知識のないユーザでも、そうした「ありがち度」[15]が高い事象であることを確認できる。



図 8 2012年8月版 Songle のコード進行検索機能の使用例

3.3 外部埋め込みプレーヤ機能

外部の Web ページ内に Songle の小型プレーヤを埋め込んで、Songle 上の楽曲を紹介できる外部埋め込みプレーヤ機能を実現した(図9)。このプレーヤは楽曲構造の可視化機能とサビ出し機能を備えており、その Web ページを閲覧した人が手軽に試聴しながら Songle を知ることができる。曲名をクリックすれば、Songle 上のその楽曲のページに直接アクセスできる。

さらに、ユーザのホームページやブログなどの埋め込み先の Web ページの背景に、楽曲の再生に連動して変化する視覚的効果(図10)を付与するオプション機能も、埋め込み時に選択できる。この機能は、閲覧者のブラウザ上において、外部埋め込みプレーヤの埋め込み先の Web ページ上で動作する JavaScript のプログラム(視覚的効果の表示用)と、外部埋め込みプレーヤを配信している <http://songle.jp> の Web ページ上で動作する JavaScript のプログラムが、

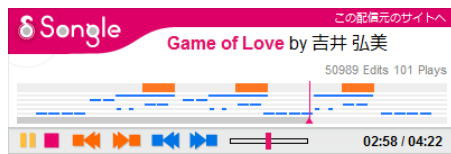


図 9 2012 年 8 月版 Songle の外部埋め込みプレーヤ

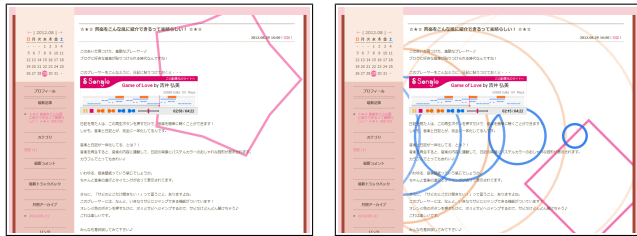


図 10 Songle 外部埋め込みプレーヤを別の Web ページに埋め込んだ例

HTML5 Web Messaging の仕組みによって異なるドメイン間で通信することで可能にしている。埋め込み先での視覚的効果の表示用プログラムを書き換えることで、様々な表示に変更できる。

つまり、Songle 外部埋め込みプレーヤは、Songle 上の音楽情景記述を Songle 以外の Web サービスから活用して連携できる枠組みとなっている。これにより、従来、音楽理解技術がなければ困難だった音楽に連動した表示を、Songle 以外の Web サービス上でも可能にした点が優れている。また、音楽情景記述に推定誤りがあっても、Songle 上で訂正すれば即座にその結果が外部埋め込みプレーヤに反映されるので、訂正するインセンティブも生じる。外部埋め込みプレーヤをアニメーション表示等に活用する際に、第三者の訂正によって、音楽情景記述が後から変更されると困る場合には、過去の特定の日時の音楽情景記述を明示的に指定することもできる。

4. 2013 年 8 月版 Songle の新機能

上記の「2012 年 8 月版 Songle」から拡張し、「2013 年 8 月版 Songle」では、動画共有サービスへの対応、同一楽曲の紐付け機能、実世界デバイスの音楽連動制御機能の三つを追加する。

4.1 動画共有サービスへの対応

代表的な動画共有サービス（ニコニコ動画、YouTube）上で公開されている動画形式の楽曲に対応する。これは、動画共有サービスが公式に提供する埋め込み動画プレーヤ機能を用いて実現した。具体的には、ニコニコ動画と YouTube に投稿されて公開されている動画を外部のサイト上に再生できる「ニコニコ動画外部プレーヤ」と「YouTube 埋め込みプレーヤ」を、Songle の楽曲ページに埋め込んだ。それらの埋め込み動画プレーヤ上の再生と同期して、「音楽地図」中の再生位置スライダが自動的に移動することで、

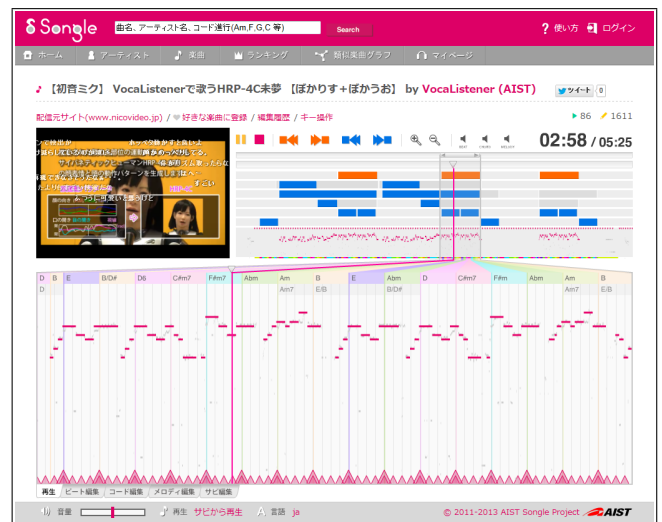


図 11 2013 年 8 月版 Songle の音楽再生用インタフェース（ニコニコ動画上の動画を再生する例）

動画内の音楽に連動した表示ができる（図 11）。音楽情景記述の推定誤りを訂正するインタフェースは図 12 のように、Songle 外部埋め込みプレーヤも図 13 のように使用できる。以上のようにユーザが Songle を使用しているときには、元の動画共有サービス上にある動画が、Songle を経由せずに直接ユーザのブラウザ上でストリーミング再生される。

この拡張により、MP3 形式で公開されている楽曲だけでなく、動画共有サービス上の動画形式の楽曲でも能動的音楽鑑賞インタフェースをユーザは体験できるようになり、Songle 上の登録楽曲数を大幅に増やすことができた。ただし現在の実装では、動画共有サービスにはビジュアル化画面表示機能は対応していない。また、動画共有サービスが提供する埋め込み動画プレーヤでは、動画圧縮符号化時のキーフレーム（I フレーム）が存在する時刻にしかシークができず、キーフレームの間隔が粗いときには、ユーザが望んだ位置にシークできないことがある。特に、ニコニコ動画上でユーザが投稿した動画が再エンコードされずに、もしキーフレームが粗い状態になっていると、「ニコニコ動画外部プレーヤ」でのシークが困難な場合があった。

4.2 同一楽曲の紐付け機能

動画共有サービスには、同一の楽曲が違う映像と組み合わせられた派生動画として複数投稿されていることがあるのでそれらを紐付けて適切に扱うための機能を実現した。Songle では音楽コンテンツ（MP3 形式あるいは動画形式での楽曲）を配信元の URL で区別するため、それらは独立に登録されることになる。そこで、新たに音楽コンテンツが Songle に登録されたときには、それと音響信号が一致する登録済みの音楽コンテンツを網羅的に自動検出し、相互に紐付けた。紐付けられた一連の音楽コンテンツはグループを形成し、自動推定結果を共有する。これにより、



図 12 誤り訂正機能で音楽情景記述を訂正する際のインタフェース (2013 年 8 月版 Songle でニコニコ動画上の動画に対して訂正する例). 候補選択や直接修正によって入力できる効率的な誤り訂正インタフェース (エディタ) を Web 上に実装した.



図 13 2013 年 8 月版 Songle の外部埋め込みプレーヤ (ニコニコ動画上の動画を埋め込む例)

ある楽曲で自動推定の誤りをユーザが訂正すれば、それと同一グループの他の音楽コンテンツを閲覧したときに、訂正後の「音楽地図」を利用できる。

同一グループの任意の音楽コンテンツで訂正ができるので、訂正したいユーザにとっては、訂正の操作がしやすい MP3 形式の楽曲や、動画圧縮符号化時のキーフレームが多い動画を容易に把握できることが重要となる。訂正ではシークを多用するが、キーフレームが少ないと前述したよ

うにシークが困難になるからである。そこで、Songle の各楽曲のページでは、同一グループに紐付けられた他の音楽コンテンツの一覧表示を見ることができるようにした。その一覧では、MP3 形式の楽曲や、動画圧縮符号化時のキーフレームが多い動画が上位に表示され、訂正が容易な音楽コンテンツへ移動しやすくなっている。

音楽コンテンツ冒頭の無音区間等の長さによって、楽曲部分の音響信号としては同一でも、動画の先頭時刻に対して時間的なずれ (オフセット) が生じることがあるが、それも自動的に補正しながら「音楽地図」を表示できるようにした。同一楽曲の自動検出の実装には、Robust Landmark-Based Audio Fingerprinting (AUDFPRINT) [16] を用いた。

以上の機能では、同一の楽曲としては紐付けられないが、楽曲の内容が同一ではなく似ている派生動画も複数投稿されることがある。そこで、楽曲間での「音楽地図」のコピー・ペーストのインタフェースも新たに開発した。4 種類の音楽情景記述 (楽曲構造、階層的なビート構造、メロディライン、コード) のそれぞれに関して、任意の区間をコピーして、同一楽曲内や別楽曲上でペーストできる。例えば、同じ伴奏を使用しながら様々な歌唱者が歌い直した派生動画群では、メロディラインは歌い直しによって違うことがあっても、楽曲構造、ビート構造、コード進行は、訂正済みの他の楽曲からコピーできる場合が多い。同一楽

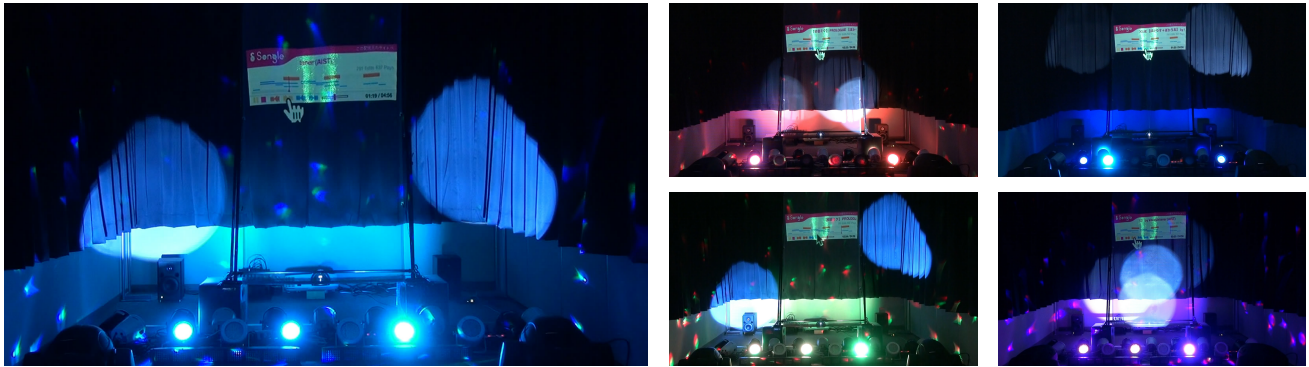


図 14 Songle 外部埋め込みプレーヤ機能を活用した舞台照明機器の音楽連動制御例

曲を同一テンポでアレンジを変えて演奏し直した派生動画群でも、楽曲構造やビート構造はコピーできることがある。

4.3 実世界デバイスの音楽連動制御機能

Songle の外部埋め込みプレーヤ機能により、音楽の中身に連動して変化するアニメーション表示をこれまで実現してきたが、実世界デバイスを制御できるように拡張した。現在の実装では、通信プロトコル「DMX512」に対応した照明等の機器の制御に対応しており、Songle 外部埋め込みプレーヤで任意の楽曲を再生すると、その音楽情景記述に連動して機器の制御内容を変えられる。DMX512 は、主に舞台照明や演出機器の制御に使われており、対応機器も多く市販されている。

Web ブラウザ上で、外部埋め込みプレーヤから音楽情景記述情報を受け取って動作するプログラム (JavaScript で記述) によって、イーサネット経由で DMX512 を制御するために、マイコンボード「mbed NXP LPC1768」に対応した DMX512 プラットフォーム「mbed DMX Platform」のキットを使用して実装した。そして、ムービングヘッド型照明機器の「ELATION / PLATINUM SPOT 5R」や、LED 照明機器の「ALKALITE / OCTOPOD75BK」, 「ALKALITE / OCTOPOD75WH」, 「AMERICAN DJ / DEKKER LED」等を制御することで、ステージに見立てた空間に様々な光を音楽に連動して投影した (図 14)。光の色や方向、投影パターンを、楽曲構造、階層的なビート構造、コードによって変化させた。

ただし、本機能はハードウェアに依存しているため、Songle の他の機能のように Web 上で誰でも自由に利用できる状態にはなっていない。より容易に様々な場面で利用可能にしていくことが、今後の課題である。

5. 大規模な楽曲数に対応するための実装改善

各楽曲に対して歌声の声質の類似楽曲グラフを生成するには、全楽曲 (将来的に 100 万曲以上を想定) との声質に基づく距離を網羅的に計算する必要があるため、従来 Songle で用いていた距離計算法よりも約 250 倍高速に楽曲間の距

離を計算できる手法を実現した。

従来用いていた距離計算法は、歌声の声質に基づく楽曲検索システム VocalFinder [14] で提案した手法である。まず、各楽曲の音響信号に対して PreFEst [11] を適用し、各時刻ごとに最も優勢な基本周波数 (F0) を推定する。次に、推定された F0 に基づいて調波構造を抽出し、正弦波重畳モデルを用いて歌声を含む音響信号を再合成する。さらに、12 次元線形予測メルケプストラム係数 (LPMCC) および $\Delta F0$ を抽出し、信頼できる歌声区間のみを選択することで、13 次元の声質特徴量群を得る。最後に、各楽曲ごとに混合ガウスモデル (GMM) を学習する。声質に基づく楽曲間の距離は、GMM 間の対称 Kullback-Leibler (KL) 距離として定義していた。

GMM 間の対称 KL 距離は解析的に計算できないため、文献 [14] では、楽曲間の距離を計算する際に、2 つの GMM と 2 つの声質特徴量群との間の 4 種類の異なる組み合わせで対数尤度の計算を必要とする近似算法を用いていた。しかし、GMM の混合数は 64 としており、声質特徴量の個数が増えると計算コストが大きい問題があった。

そこで我々は今回、ガウス分布間の対称 KL 距離に基づく Earth Mover's Distance (EMD) [17] を用いて GMM 間の対称 KL 距離を近似する手法を採用した。さらに各楽曲の GMM を学習する際には、ディリクレ過程に基づく無限混合ガウスモデル (iGMM) [18] を用いて、各楽曲ごとに適切な混合数が自動的に求まるようにした。EMD の計算量は混合数の二乗に比例するが、ポピュラー音楽では 15~20 混合程度になる場合が多かった。RWC 研究用音楽データベース [19] (ポピュラー音楽 RWC-MDB-P-2001) の全 100 曲を対象に距離行列を求めたところ、従来用いていた距離計算法 [14] で得られた距離行列との相関係数は 0.85 であり、コンパクトな GMM を用いて高精度かつ高速な距離計算が可能となった。

上記に加えて、ユーザのメロディ訂正結果を活用して、声質特徴量を抽出するうえで重要となる歌声区間の検出精度を向上させることができる仕組みも開発した。現在の実装では、歌声区間から学習した GMM と非歌声区間か

ら学習した GMM を用いて、与えられた声質特徴量の尤度を比較することで歌声の有無を判定している。従来は、RWC-MDB-P-2001 に収録されている 100 曲から学習した GMM を用いていたが、今回新たに、ユーザの訂正で歌声あるいは非歌声と判定された区間から抽出した音響特徴量を学習データに追加して、歌声および非歌声の GMM を再学習するようにした。この仕組みによって、Songle に登録されている楽曲により適した歌声区間検出が可能になる。

6. Songle による連携と展示

能動的音楽鑑賞サービス Songle の対外的な連携として、コンテンツ投稿サイト「ピアプロ」(<http://piapro.jp>) と 2012 年 8 月 29 日から正式に連携し [20]、ピアプロ上のすべての楽曲のページに Songle へのリンクをサイト運営者のクリプトン・フューチャー・メディア株式会社が設置することで、Songle での登録や閲覧を容易にした。さらに、Songle が自動推定したサビ区間を Songle からピアプロに提供し、ピアプロ上の音楽プレイヤーでもサビ出し機能を利用できるようにして、ユーザの利便性を高めた。

さらに同日、歌声合成ソフトウェア「初音ミク」発売 5 周年記念を祝う公式ページ (<http://mikubook.com/event/miku5>) に「Songle 外部埋め込みプレーヤ」が埋め込まれて応用され、同ページの背景に音楽に連動した幾何学模様が表示されるだけでなく、初音ミクのキャラクタの 2 次元の絵が音楽に連動して切り替わり、踊って見えるアニメーションが表示された。

2012 年 12 月 27 日には、歌声合成ソフトウェア「鏡音リン・レン」発売 5 周年記念を祝う公式ページ (<http://mikubook.com/event/rinlen5>) にも「Songle 外部埋め込みプレーヤ」が埋め込まれて応用され、鏡音リン・レンのキャラクタの 3 次元モデルが音楽に連動して踊る様子が、Chrome ブラウザ上での 3 次元リアルタイムレンダリングによって表示された。ドラッグ操作でカメラアングルを変えることもできた。さらに画面上にはコード名が書かれたボタンがあり、Songle のコード推定結果に応じてそれらが光るだけでなく、音楽に合わせてユーザが任意のコード名のボタンを押すと、初音ミクの歌声で押したコードをコーラスとして歌う機能も実現された。

その後、つくば市の科学館「つくばエキスポセンター」から、特別展「エジソンからはじまる音と電気のふしぎな関係」への協力依頼を受け、Songle のオフライン版 (インターネット接続のない環境で動作するデモ専用バージョン) が、2013 年 3 月 23 日から 6 月 9 日まで展示された。また、産業技術総合研究所の展示施設「サイエンス・スクエア つくば」の常設展示として、同様に Songle のオフライン版が 2013 年 4 月から展示されている。

7. おわりに

本稿では、ユーザ貢献によって徐々に改善されていく能動的音楽鑑賞サービス Songle の新機能を中心に紹介した。Songle では、計算機による音楽理解の誤りを、ユーザが Web 上で訂正できるインタフェースを備えているところが、クラウドソーシング利用の事例としても位置づけられる [4,5]。そして、不特定多数のユーザによる自発的な訂正をユーザ体験の向上に結びつけていくことで、さらなる利用を促す仕組みを持っている点が大きな特長である。

2012 年 8 月版と 2013 年 8 月版の Songle によって、当初の文献 [3] で今後の課題に挙げていた主要な点に取り組むことができた。具体的には、音楽情景記述の可視化方法の多様化や、他の Web サービスから活用して連携できる機能の提供、訂正するインセンティブの向上を挙げていたが、ビジュアライザや、外部埋め込みプレーヤ機能とそれに基づく連携によって、それぞれに進展が得られたといえる。

音楽コンテンツのデジタル化は進んだが、その潜在的な可能性はまだ充分には引き出されておらず、従来は多量の楽曲にアクセスできる量的な変化が中心であった。そこ本研究では、音楽コンテンツの聴き方や活用のされ方がより能動的で豊かで便利になる質的な変化を、エンドユーザの日常生活で起こす貢献をすることを目指している。その変化を起こす鍵となるのが、音楽理解技術である。これまでも音楽理解技術に関連して様々な研究がなされてきたが [21-28]、今後、より一層の発展が必要である。本研究でもさらに高度な音楽理解技術を研究開発していくと共に、エンドユーザにとっての利便性をより向上するための研究開発や、様々な連携にも取り組んでいく予定である。

謝辞 Songle の Web サービスのサーバとクライアントの実装を担当して頂いた川崎 裕太氏、Web デザインとビジュアライザを担当して頂いた櫻井 稔氏、外部埋め込みプレーヤと実世界デバイスの音楽連動制御を担当して頂いた井上 隆広氏、音楽理解技術等を担当して頂いた藤原 弘将氏と Matthias Mauch 氏に感謝する。本研究の一部は JST CREST 「OngaCREST プロジェクト」の支援を受けた。

参考文献

- [1] Goto, M., Yoshii, K., Fujihara, H., Mauch, M. and Nakano, T.: Songle: A Web Service for Active Music Listening Improved by User Contributions, *Proc. of IS-MIR 2011*, pp. 311-316 (2011).
- [2] 後藤真孝, 吉井和佳, 藤原弘将, Mauch, M., 中野倫靖: Songle: ユーザが誤り訂正により貢献可能な能動的音楽鑑賞サービス, 情報処理学会インタラクション 2012 論文集, pp. 1-8 (2012).
- [3] 後藤真孝, 吉井和佳, 藤原弘将, Mauch, M., 中野倫靖: Songle: 音楽音響信号理解技術とユーザによる誤り訂正に基づく能動的音楽鑑賞サービス, 情処学論, Vol. 54, No. 4, pp. 1363-1372 (2013).

- [4] Goto, M., Ogata, J., Yoshii, K., Fujihara, H., Mauch, M. and Nakano, T.: **[Keynote talk]** PodCastle and Songle: Crowdsourcing-Based Web Services for Spoken Content Retrieval and Active Music Listening, *Proc. of the 2012 ACM Workshop on Crowdsourcing for Multimedia (CrowdMM 2012)*, pp. 1–2 (2012).
- [5] Goto, M., Ogata, J., Yoshii, K., Fujihara, H., Mauch, M. and Nakano, T.: PodCastle and Songle: Crowdsourcing-Based Web Services for Retrieval and Browsing of Speech and Music Content, *Proc. of the First International Workshop on Crowdsourcing Web Search (CrowdSearch 2012)*, pp. 36–41 (2012).
- [6] 産業技術総合研究所: プレスリリース「インターネット上の楽曲の中身を自動解析する音楽鑑賞システムー誰でも利用できる能動的音楽鑑賞サービス「Songle」を一般公開ー」, http://www.aist.go.jp/aist_j/press_release/pr2012/pr20120829/pr20120829.html (2012).
- [7] Goto, M.: Active Music Listening Interfaces Based on Signal Processing, *Proc. of ICASSP 2007* (2007).
- [8] 後藤真孝: 音楽音響信号理解に基づく能動的音楽鑑賞インタフェース, 情処研報音楽情報科学 2007-MUS-70-9, pp. 59–66 (2007).
- [9] 後藤真孝: リアルタイム音楽情景記述システム: 全体構想と音高推定手法の拡張, 情処研報 音楽情報科学 2000-MUS-37-2, pp. 9–16 (2000).
- [10] Goto, M.: Music Scene Description Project: Toward Audio-based Real-time Music Understanding, *Proc. of ISMIR 2003*, pp. 231–232 (2003).
- [11] Goto, M.: A Real-time Music Scene Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Communication*, Vol. 43, No. 4, pp. 311–329 (2004).
- [12] Goto, M.: Music Listening in the Future: Augmented Music-Understanding Interfaces and Crowd Music Listening, *Proc. of the AES 42nd International Conf. on Semantic Audio*, pp. 21–30 (2011).
- [13] Goto, M.: A Chorus-Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station, *IEEE Trans. on ASLP*, Vol. 14, No. 5, pp. 1783–1794 (2006).
- [14] Fujihara, H., Goto, M., Kitahara, T. and Okuno, H. G.: A Modeling of Singing Voice Robust to Accompaniment Sounds and Its Application to Singer Identification and Vocal-Timbre-Similarity-Based Music Information Retrieval, *IEEE Trans. on ASLP*, Vol. 18, No. 3, pp. 638–648 (2010).
- [15] 後藤真孝: 招待講演「未来を切り拓く音楽情報処理」, 情処研報音楽情報科学, Vol. 2013-MUS-99, No. 33, pp. 1–9 (2013).
- [16] Ellis, D.: Robust Landmark-Based Audio Fingerprinting, <http://labrosa.ee.columbia.edu/matlab/fingerprint/> (2009).
- [17] Rubner, Y., Tomasi, C. and Guibas, L. J.: A Metric for Distributions with Applications to Image Databases, *Proc. of ICCV 1998*, pp. 59–66 (1998).
- [18] Rasmussen, C. E.: The Infinite Gaussian Mixture Model, *In Advances in Neural Information Processing Systems 12*, MIT Press, pp. 554–560 (2000).
- [19] 後藤真孝, 橋口博樹, 西村拓一, 岡 隆一: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情処学論, Vol. 45, No. 3, pp. 728–738 (2004).
- [20] クリプトン・フューチャー・メディア株式会社: プレスリリース「『ピアプロ』が『Songle』に正式対応したことについてのお知らせ」, <http://www.value-press.com/pressrelease/100718> (2012).
- [21] Klapuri, A. and Davy, M.(eds.): *Signal Processing Methods for Music Transcription*, Springer (2006).
- [22] 嵯峨山茂樹, 亀岡弘和: 解説“自動採譜技術の展望”, 日本音響学会誌, Vol. 64, No. 12, pp. 715–720 (2008).
- [23] Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C. and Slaney, M.: Content-Based Music Information Retrieval: Current Directions and Future Challenges, *Proceedings of the IEEE*, Vol. 96, No. 4, pp. 668–696 (2008).
- [24] 片寄晴弘, 白木善尚 (編): 特集「音楽情報処理技術の最前線」, 情報処理 (情報処理学会誌), Vol. 50, No. 8, pp. 709–772 (2009).
- [25] 後藤真孝, 緒方 淳: 音楽・音声の音響信号の認識・理解研究の動向, コンピュータソフトウェア (日本ソフトウェア科学会論文誌), Vol. 26, No. 1, pp. 4–24 (2009).
- [26] Muller, M., Ellis, D., Klapuri, A. and Richard, G.: Signal Processing for Music Analysis, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 5, No. 6, pp. 1088–1110 (2011).
- [27] Yoshii, K. and Goto, M.: Unsupervised Music Understanding Based on Nonparametric Bayesian Models, *Proc. of ICASSP 2012*, pp. 5353–5356 (2012).
- [28] 亀岡弘和: 非負値行列因子分解の音響信号処理への応用, 日本音響学会誌, Vol. 68, No. 11, pp. 559–565 (2012).