情報処理学会論文誌 Vol.54 No.4 1363–1372 (Apr. 2013)

推薦論文

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

後藤 真孝[1,a)] 吉井 和佳[1] 藤原 弘将[1] Matthias Mauch[1] 中野 倫靖[1]

受付日 2012年6月20日, 採録日 2012年10月10日

概要：本論文では，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 について述べる．従来，研究開発段階の音楽インタフェースや技術を，日常生活で人々に使ってもらうのは容易でなかった．Songle では，Web 上で人々に能動的音楽鑑賞インタフェースを体験してもらうことで，音楽鑑賞がより能動的で豊かになる質的な変化を日常生活で起こすことを目指す．そして，Web 上の任意の楽曲に対して楽曲構造，階層的なビート構造，メロディライン，コードの 4 種類の音楽情景記述を自動推定して可視化することで，それを見て再生したユーザの音楽理解が深まることを可能にする．しかし，自動推定では誤りが不可避である．そこで効率的な誤り訂正インタフェースを Web 上で提供し，誤りを人手で自発的に訂正する貢献を促す．そうした不特定多数による訂正がユーザ体験の改善に結び付くことで，Songle のさらなる利用を促していく．

キーワード：能動的音楽鑑賞，音楽理解，音楽鑑賞インタフェース，集合知，クラウドソーシング

Songle: An Active Music Listening Service Based on Music-understanding Technologies for Audio Signals and Error Corrections by Users

Masataka Goto[1,a)] Kazuyoshi Yoshii[1] Hiromasa Fujihara[1] Matthias Mauch[1] Tomoyasu Nakano[1]

Received: June 20, 2012, Accepted: October 10, 2012

Abstract: This paper describes a public web service for active music listening, *Songle*, that enriches music listening experiences by using music-understanding technologies based on signal processing. Although various research-level music interfaces and technologies have been developed, it has not been easy to get people to use them in everyday life. Songle aims at bringing qualitative changes in everyday life toward more active, richer music listening by enabling people to experience active music listening interfaces on the web. Songle facilitates deeper understanding of any musical piece on the web by visualizing its music scene descriptions estimated automatically, such as music structure, hierarchical beat structure, melody line, and chords. When using music-understanding technologies, however, estimation errors are inevitable. Songle therefore features an efficient error correction interface that encourages people to contribute by correcting those errors to improve the web service. The error corrections by anonymous users lead to a better user experience, which encourages further use of Songle.

Keywords: active music listening, music understanding, music listening interface, wisdom of crowds, crowdsourcing

[1] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305–8568, Japan
[a)] m.goto@aist.go.jp

本論文の内容は 2012 年 3 月のインタラクション 2012 にて報告され，同プログラム委員長により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である．

1. はじめに

デジタル化された音楽コンテンツが持つ潜在的な可能性は，まだ十分には引き出されていない．近年，膨大な音楽コンテンツをデジタル化して保存しておくことで，いつでもどこでも視聴可能になった．また，音楽配信やオンラインストレージ等の普及により，自ら持ち運ぶ必要もなくなり，書誌情報に基づく音楽検索や，協調フィルタリング等に基づく音楽推薦も実用化された．しかし，従来は表層的なスペクトル特徴量等に基づく処理が中心で限界があり（たとえば，性能が頭打ちになる「ガラスの天井」問題 [1] が有名），音楽の音響信号の内容理解に踏み込んだ処理は，ほとんど普及していなかった．

デジタル化がもたらす価値として，こうした多量の楽曲に自在にアクセスできる量的な変化は日常生活で起きたが，本研究ではさらに，音楽の聴き方がより能動的で豊かになる質的な変化をエンドユーザの日常生活で起こすことを最終的な目的とする．その変化を起こす鍵となるのが，音楽の音響信号を自動的に理解できる技術（音楽音響信号理解技術）である．後藤らは 2002 年から，従来の受動的な鑑賞とは違う，能動的な音楽鑑賞を可能にするエンドユーザ向け音楽インタフェースの研究に取り組み [2]，その研究アプローチを**「能動的音楽鑑賞インタフェース」**と名付けて [3], [4],「音楽音響信号理解技術が，音楽の聴き方をどのように豊かにできるか」を様々な事例により明らかにしてきた．ここでの「能動的」という言葉は，音楽の創作は意味せず，音楽鑑賞を楽しむうえでのあらゆる能動的なインタラクション（音楽再生位置の変更やインタラクティブな加工，ブラウジング等）を意味する．たとえば，サビ出し機能付き能動的音楽鑑賞インタフェース「SmartMusicKIOSK」[2], [5] では，ユーザが楽曲中の興味のない区間を容易に飛ばしながら，自動検出されたサビ区間を聞いたり，楽曲中の繰返し構造を可視化した「音楽地図」を見ながら，音楽に対する理解をより深めたりすることができる．しかし，こうした研究中の音楽鑑賞インタフェースや音楽理解技術を，日常生活で誰でも自由に使える環境は実現されておらず，エンドユーザは質的な変化を実感できなかった．

そこで本研究では，誰でも Web ブラウザ上で能動的音楽鑑賞インタフェースを使用して楽しむことができる環境の実現を第 1 の目標とし，Web サービス**「Songle」（ソングル）**を実現して提供する．Songle は，音楽理解技術を用いて，Web 上で公開されている任意の楽曲（MP3 形式の音響信号ファイル）中の様々な音楽情景記述（音楽要素）[6], [7], [8] を推定する．Songle のユーザは，その結果が可視化された様子を見ながら，楽曲の再生を楽しむことができる．現在の実装では，歌声をともなうポピュラー音楽を主な対象として，

(i) 楽曲構造（サビ区間と繰返し区間），
(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を自動推定し，可視化して音楽内容に基づくブラウジングを可能にする．さらに Songle では，前述の SmartMusicKIOSK のインタフェース機能を実装し，ユーザが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．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．

さらに本研究では，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の実現を第 2 の目標とし，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を Web 上で提供する．音楽理解技術は，推定誤りが不可避だが，人間が一生かけても聞ききれない多量の楽曲を処理できるという利点を持つ．一方，人間（特に音楽家）は音楽の内容をより深く理解して記述でき，推定誤りにも気づくことができるが，何もないところからすべてを記述するのは長時間かかり限界がある．そこで両者が相補的に力を合わせることで，より的確な音楽情景記述を各楽曲に付与し，音楽鑑賞時のユーザ体験を向上させることを狙う．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．今後，十分な訂正量が蓄積された段階で，訂正結果を機械学習することによって，音楽理解技術の性能向上につなげることも可能である．

Songle により，音楽の聴き方がより能動的で豊かになる質的な変化を誰でも体験できるだけでなく，どういう音楽ジャンルや混合音に対する推定が難しいかといった音楽理解技術に対する理解が深まることが期待できる．将来的には，様々な研究者が開発した音楽理解技術に対応し，共同で社会に対して貢献していくプラットフォームに発展させていければと考えている．

2. 能動的音楽鑑賞サービス Songle の機能

Songle は，Web 上の楽曲をユーザが検索，閲覧，アノテーション可能なソーシャルアノテーション用 Web サービスである．**図 1** に Songle のタイトルページ表示例を，**図 2** に楽曲選択後の音楽再生用インタフェース表示例を示す．Songle 公開当初の現段階では，歌声をともなうポピュラー音楽を主な対象としている．ユーザは，任意の楽曲の MP3 ファイルが置かれている URL や，複数の MP3 ファイルの URL が列挙された Web ページあるいは RSS（Really Simple Syndication）の URL を指定することで，Songle に登録できる．

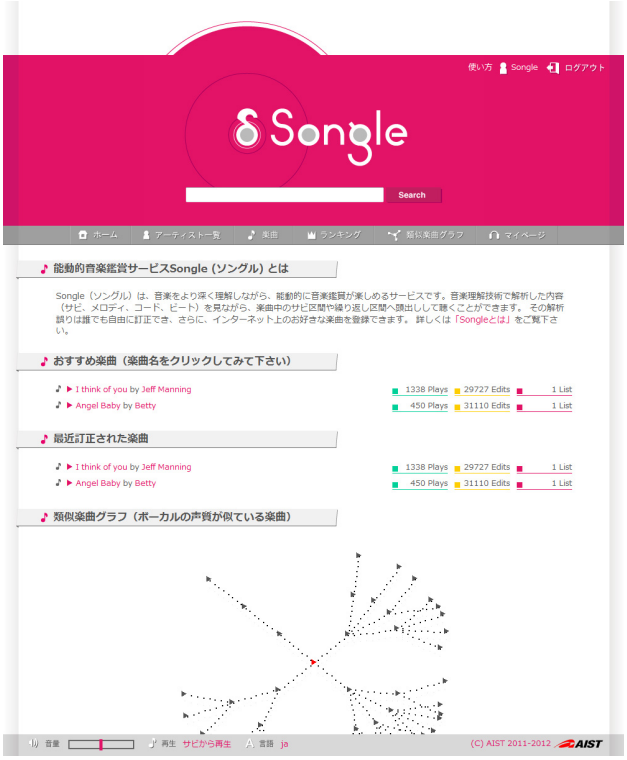


図 1　Songle のタイトルページの画面表示例．類似楽曲グラフも表示されている

Fig. 1　Songle screen snapshot of the title page with a graph of similar songs.

図 2　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．自動推定された音楽情景記述も可視化されている

Fig. 2　Songle screen snapshot of the main interface for music playback with the visualization of music scene descriptions estimated automatically.

誰でも Songle にログインせずに匿名で能動的音楽鑑賞を楽しみ，推定誤りを自発的に訂正できる．ただし楽曲を登録するには，OpenID 等を用いてログインする必要がある．ログインすると，プレイリストを作成したり，作成したプレイリストに名前をつけて他のユーザと共有したりすることができる．ログインしたユーザごとの楽曲推薦機能もあり，各曲の好き・嫌いの嗜好を入力していくと推薦精度が向上する．

Songle では，1 章で述べた 2 つの目標を達成するために，「検索」，「閲覧」，「アノテーション（誤り訂正）」の 3 つの機能を用意した．「検索」と「閲覧」機能によって能動的音楽鑑賞を可能にし，「アノテーション」機能によって，推定誤りを訂正して音楽情景記述の改善に貢献可能にする．そして，音楽情景記述が改善されると，楽曲を検索・閲覧する際のユーザ体験も向上する．

2.1 「検索」機能

曲名やアーティスト名をテキスト検索したり，曲名一覧，アーティスト名一覧から選択したりして，次の「閲覧」機能で再生する楽曲を指定できる機能である．最近推定や訂正がなされた曲名一覧や，様々なランキング（再生回数の多い楽曲やアーティスト，登録楽曲の多いアーティスト，訂正回数の多いユーザ等）から選択することもできる．

さらに，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」[9] の発想を取り入れ，声質の類似楽曲グラフを表示して（図 1），そこからの選択も可能にする．類似楽曲グラフの中心ノードには，推薦されたりユーザが指定したりした楽曲が配置され，それと声質が似ている楽曲のノードが周辺に放射状に接続されて表示される．ユーザは，ノード上の楽曲にマウスカーソルをオーバレイして試聴しながら，グラフ上を次々とたどることで，好みの声質に近い新たな楽曲を見つけることができる．

最終的に，検索結果や曲名一覧等で，興味がある楽曲かを試聴をして判断し，楽曲を選択すると次の「閲覧」機能の画面に切り替わる．プレイリストを選択した場合には，その再生順に従って，「閲覧」機能画面上で聴いている楽曲が次々と自動的に切り替わる．

2.2 楽曲内の「閲覧」機能

図 2 の上半分の再生画面のように，楽曲の内容（音楽情景記述）を可視化したユーザインタフェースにより，再生位置を自在に制御できる機能である．横軸が時間であり，再生画面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．

この機能では，自動推定した下記の 4 種類の音楽情景記述をユーザが閲覧しながら，普段見落としがちな構造や音

に気づいたりできる点が重要である．

(1) **楽曲構造（サビ区間と繰返し区間）**

大局的な表示部に，SmartMusicKIOSK [2], [5] の「音楽地図」が表示されている．音楽地図は，楽曲中の繰返し構造を可視化した楽曲構造表示で，最上段にサビ区間，その下の5段に様々な長さの繰返し区間が表示されている．各段の中で，着色されている区間どうしが似ている（繰返しである）ことを表している．楽曲を聴く前に構造が把握できるので，区間を直接クリックして再生したり，興味がある場所へ再生位置スライダを動かして再生したりすることが可能である．音楽地図よりも上には，再生位置スライダ，楽曲の先頭からの経過時間表示，再生操作ボタンが表示されている．再生操作ボタンには，通常の再生，停止ボタンだけでなく，音楽地図に対応した「前・次のサビ区間の頭出し」,「前・次の繰返し区間の頭出し」ボタンが配置されている．

(2) **階層的なビート構造（拍と小節の先頭）**

局所的な表示部の最下部では，小さい三角形が，各拍（四分音符に対応するビート）の位置を示している．三角形の上側の頂点がその時刻である．大きい三角形は，小節の先頭を示している．大局的な表示部でも，音楽地図の直下にすべての小節の先頭を赤い点列で表示し，ユーザがテンポ変化に気づきやすく工夫した．

(3) **メロディライン（メロディの歌声の基本周波数（F0））**

局所的な表示部のビート構造表示の上には，メロディラインの基本周波数（F0）がピアノロール[*1]で表示されている．半音（semitone）単位に量子化した表示もできる．大局的な表示部でも，下部に楽曲全体のメロディラインを縮小表示し，ユーザが繰返しや全体の高低変化に気づきやすく工夫した．

(4) **コード（根音とコードタイプ（構成音））**

局所的な表示部の最上部には，各区間ごとのコード名がテキストで表示されている．たとえば，コード名 Abm7 は，根音が Ab であり，その構成音を示すコードタイプが m7 であることを意味している．1 オクターブ中の 12 種類の根音をそれぞれ違う色で着色することで，ユーザがコード進行の繰返し等に気づきやすく工夫した．大局的な表示部の最下部も同様に着色した．

こうした表示は，「音楽理解力拡張インタフェース」[10] の観点からも重要であり，専門的知識のないユーザ（非音楽家）でも，各音楽的要素の存在や要素間の関係，構成上の意図に気づきやすくなる．このように，音楽の理解は再生に同期して「見る」ことで深まる．

さらに，図 2 の下半分（再生画面の下）では，ソーシャルタグや時刻同期コメント（各時刻を指定して言及することもできる任意のテキスト）を，ユーザが自由に追加・共有できる．時刻同期コメントをクリックすると，そこから再生される．音楽を聴いて気付いたことをその要素に時刻同期してコメントして共有することで，他の人とコミュニケーションしながら自分 1 人では気付けない豊かな理解を得ることができる [10]．

2.3 アノテーション（誤り訂正）機能

音楽を聴きながら推定誤りに気づいたら，それを訂正して，楽曲のアノテーションとして共有できる機能である．ここでのアノテーションは楽曲の内容に関する記述を意味し，候補選択や直接修正によって入力できる．そのために，**図 3** に示すような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．

再生画面下部の局所的な表示部のタブを切り替えることで，下記の 4 種類の音楽情景記述を編集できる．

(1) **楽曲構造** （図 3 (a)）

サビ区間，繰返し区間の各段ごとに区間長は同一となっているので，追加ボタンを押してから任意の段にマウスカーソルを移動すると，その段に応じた区間長で区間が表示され，クリックすれば追加できる．サビ区間，繰返し区間はドラッグすると移動でき，削除ボタンを押してから区間をクリックすれば削除できる．ある区間の先頭か末尾の境界線をドラッグすることで，その段のすべての区間の区間長の調整もできる．こうした訂正で音楽地図が改善されると，ユーザはより的確に再生位置の変更が可能になる．

(2) **階層的なビート構造** （図 3 (b)）

拍と小節先頭で異なるクリック音が鳴りながら楽曲とともに再生されるので，ユーザはそれらが正しいかを確認する．訂正が必要な場合には，局所的な表示部の下半分に表示された複数の訂正候補から選択し，選択時点から先を置き換えることができる．もし適切な候補がない場合には，音楽再生に同期して拍と小節先頭の位置でそれぞれの入力キーを押すか，手作業で左右に直接移動して修正する．

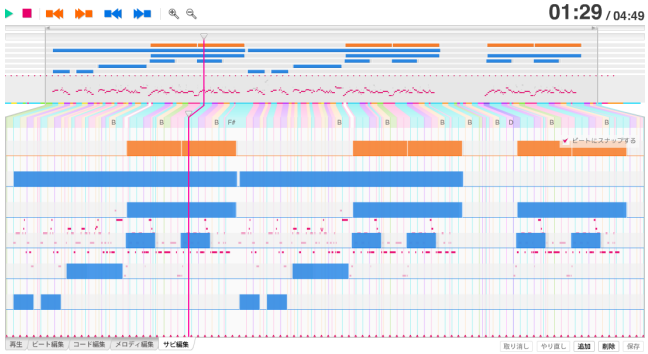
(3) **メロディライン** （図 3 (c)）

推定したメロディの高さで合成音が楽曲とともに再生されるので，ユーザは訂正が必要な箇所を確認し，必要に応じてピアノロール表示上で半音単位の線を引いて訂正ができる．ただし，メロディラインは内部的には F0 の軌跡として表現されているので，軌跡を直接描いて半音より細かい修正をしてもよい．この訂正でメロディラインが改善されると，声質の類似楽曲グラフがより的確に求まる．

(4) **コード** （図 3 (d)）

コードの構成音が楽曲とともに再生されるので，ユーザはその音とコード名表示から正しいかを確認する．訂正が必要な場合には，コード名をクリックして表示される

[*1] ピアノロールとは，横軸が時間，縦軸が音高の二次元平面上で，発音中の部分に着色する表示方法である．



(a) 楽曲構造（サビ区間と繰返し区間）の訂正
(a) Correcting music structure
(chorus sections and repeated sections)

(b) 階層的なビート構造（拍と小節の先頭）の訂正
(b) Correcting hierarchical beat structure
(musical beats and bar lines)

(c) メロディライン（メロディの歌声の基本周波数（F0））の訂正
(c) Correcting melody line (F0 of the vocal melody)

(d) コード（根音とコードタイプ（構成音））の訂正
(d) Correcting chords (root note and chord type)

図 3 アノテーション機能で音楽情景記述を訂正する際の画面表示例
Fig. 3 Songle screen snapshots of the annotation function for correcting music scene descriptions.

候補から選択するか，直接コード名をタイプ入力できる．

コードの区間（境界）も追加・修正・削除ができる．

以上の訂正機能は，いかにユーザが迅速かつ容易に訂正できるかを追求して実現した．上記で選択肢として表示される訂正候補は，いずれも音楽情景記述の推定時に事前に求めておく．また，階層的なビート構造の各ビートの位置を活用することで，他の音楽情景記述を容易に編集するためのビートスナップ機能も選択可能とした．ビートスナップ機能では，カーソルがビートの位置以外には移動できないように制約をかけることで，編集時に区間や時刻を容易に指定できる．ただしビートの位置だけでは粗いので，その間隔の 4 分の 1 ずれた各位置（四分音符に対応するビートに対して 16 分音符単位の位置）にも移動可能とした．

以上が達成する Songle の第 1 の目標は，あくまで 1 章で述べたように，誰でも Web ブラウザ上で能動的音楽鑑賞インタフェースを使用して楽しむことができる環境を提供することである．そのため，ユーザの多くは訂正をせずに，単に能動的音楽鑑賞を楽しむことを我々は想定している．実際，一部のユーザにとっては，訂正の意欲はあっても，上記の音楽情景記述（特にコード）を訂正するのは難しい．そこで，よりいっそう簡易に訂正できるインタフェースの設計にも，今後取り組んでいくことを検討している．また，訂正してもらえる場合でも，すべてが網羅的に訂正されるのではなく，ユーザが興味のある箇所が一部だけ訂正されることを期待している．

図 4 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）
Fig. 4 Original estimation results are visualized as trails with gray color after error corrections.

こうした Songle の訂正機能は，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の実現という第 2 の目標を達成する．それと同時に 1 章で述べたように，ユーザが Songle を使用して楽しむ過程で，音楽理解技術に対する理解が深まることも我々は期待している．そこで，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 4 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，すべての訂正履歴を記録して，誰でも訂

正前後の比較ができる機能も用意した．

3. Songleの実装

Songleのシステム構成図を**図 5** に示す．**Webクローラ**はユーザが登録したURLやRSSに基づいて楽曲MP3ファイルを収集し，データベースに登録する．次に，音楽情景記述のそれぞれの種類に対応した**音楽理解モジュール**が，各楽曲を処理する．たとえば，楽曲構造とビート構造は別々のモジュールが推定する．処理が終わった音楽理解モジュールから**音楽理解状態管理部**へリクエストがあると，次に処理すべき楽曲が引き渡される．音楽理解モジュールがその推定処理を終えると，推定結果は音楽理解状態管理部を経てデータベース管理部に渡される．**データベース管理部**では，その推定結果やユーザによる訂正結果を保存し，処理状態を管理する．最後に，**Webサーバ**は，Songleの**インタフェース**を提供するWebサイトとして動作する．なお楽曲MP3ファイルは，Songleにアップロードすることはできず，元のWebサイトから，Songleを経由せずに直接ユーザのブラウザ上で再生される．

一連の機能のサーバ側動作は，Webアプリケーションフレームワーク Ruby on Rails，プログラミング言語Ruby，Webサーバ Passengerおよび Apache，データベース MySQLを用いて実装した．一方，クライアント側のユーザインタフェース機能は，スクリプト言語 ActionScript 3，そのコンパイラ Adobe Flex Compiler，スクリプト言語 JavaScriptを用いて実装した．

各音楽情景記述は，以下のように推定した．

(1) **楽曲構造の推定**

サビ区間と繰返し区間は，ポピュラー音楽に対するサビ区間検出手法 RefraiD [2], [5] を用いて推定した．RefraiDは，様々な繰返し区間の相互関係を調べることで，転調の有無にかかわらず，楽曲中で繰り返されるすべてのサビ区間を網羅的に検出しようとする特長を持つ．

(2) **階層的なビート構造の推定**

拍推定は，テンポや位相の違いを表現する複数の状態を持つ隠れマルコフモデル（HMM）に基づく手法を新たに実現した．各テンポにおいて，テンポ変動や変化に追従した遷移が可能な left-to-right 型の HMMで拍をモデル化し，各状態の出力確率は発音時刻検出結果との一致度に基づいて計算した．裏拍や倍テンポ誤りが起きる可能性も考慮しながら，もっともらしい複数の拍候補を出力できるという特長を持つ．

小節先頭の推定は，コードの響きの変化に基づく手法を実現した．拍に同期した低域と高域の NNLS クロマ特徴量 [11] を求め，コード変化をしている可能性が高い拍ほど，小節の先頭となるように計算した．その際，3/4，4/4，6/8拍子のそれぞれの可能性を評価して，もっともらしい拍子と小節の先頭を決定した．

(3) **メロディラインの推定**

混合音中で最も優勢な音高を推定する手法 PreFEst [8] を，メロディの歌声に特化するように拡張した F0推定手法 [12] を用いて推定した．PreFEstにより求めた各F0候補に対し，その歌声らしさを事前に学習した歌声GMMを用いて評価することで，歌声に対する推定精度を高めた．さらに，歌声区間も文献 [13] の手法で推定した．

(4) **コードの推定**

9種類の代表的なコードタイプ（major, major 6th, major 7th, dominant 7th, minor, minor 7th, half-diminished, diminished, augmented）に加え，ベース音が異なる major コードの変種5種類（/2，/3，/5，/b7，/7）の，計14種類のコードタイプに対応し，それぞれに根音は12種類存在するので全部で168種類のコードを推定できる手法を新たに実現した．コードが存在しない区間も推定できる．本手法では，小節先頭の推定と同様のクロマ特徴量に基づくコード推定用 HMMと，3種類のスケール（major，natural minor，harmonic minor）に対応した調（キー）推定用 HMMを組み合わせた．その際，調ごとに使用されるコードが大きく異なることを文献 [14] のように考慮し，いったん，調推定用 HMMに基づいて調の事後分布を推定し，それをコード推定用 HMMによる各コードの確率の重みとして用いた．コード変化は拍の場所のみで起き，かつ，小節の先頭で起きる可能性が高いようにViterbi探索で考慮した．

声質の類似楽曲グラフは，VocalFinder [9] で用いた手法に基づいて，歌声の声質に関する2曲間の類似度を網羅的に計算して求めた．現在の実装では，楽曲の推薦機能も同一の類似度を用いている．声質の似た楽曲が推薦される点は好ましいが，多様な尺度を今後導入していく予定である．

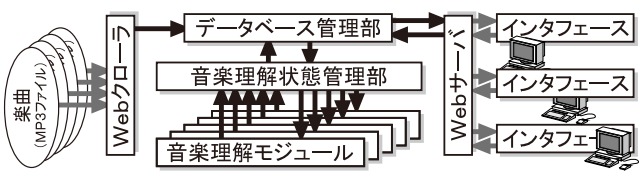


図 5 Songleのシステム構成図

Fig. 5 Implementation overview of Songle.

4. 議論

本研究のように，能動的音楽鑑賞とソーシャルアノテーションのためのWebサービスを一般公開した研究は，我々の調査した限り前例がないが，音楽のアノテーションに関しては様々な関連研究がなされてきた．以下では，関連研究を紹介するとともに，Songleの社会的，学術的貢献等を議論する．

4.1 関連研究

音楽の内容を表すアノテーションを収集する様々な手法が，従来提案されてきた．そうしたアノテーションは，機械学習手法の学習データとしても，音楽情報処理システムの評価用正解データとしても，有用である．たとえば，Lee [15] は，金銭的報酬を与えて不特定多数に仕事を依頼するクラウドソーシング型 Web サービス Amazon Mechanical Turk（MTurk）を用いて，類似度の判断結果を集めた．Mandel ら [16] は，同様に MTurk を用いて音楽を表現するタグを収集した．これらのアプローチでは，十分に高い品質のアノテーションが得られたものの，いたずら対策が不可欠で，必要とされるアノテーションの量に比例して，人間の労力と対価が増えるという問題があった．

その問題を解決する 1 つのアプローチが，ゲームとして楽しませながらアノテーションを付与させる方法である．たとえば，Turnbull ら [17], [18] は，与えられた曲に対して最も適切・不適切な表現をプレーヤに選ばせるアノテーションゲームを提案した．Mandel ら [19] も，他のプレーヤに付与されていないような有用な表現を入力すると加点されるゲームを提案した．Law ら [20] は，初期のアノテーションゲームとして有名な画像用 ESP Game [21] の考え方に基づき，ランダムに選ばれてペアになった相手が音に付与しそうな表現を推測して答えるゲームを提案した．一方，Songle のユーザは，貢献するとサービスが改善して自分を含む他のユーザの役に立てるということを明確に意識できるので，こうした従来のアプローチよりも，貢献しようというより強い動機に基づいてアノテーション（訂正）可能な点が特徴的である．

他の関連研究としては，楽曲のアノテーション用エディタとして，Sonic Visualiser [22]，Audacity extension [23]，CLAM [24]，MUCOSA [25] のようなスタンドアロンアプリケーションが開発されてきた．多様なアノテーションを利用するうえでは，Echo Nest API（http://developer.echonest.com/）も便利である．一方 Songle は，サビ，ビート，メロディ，コードという主要な音楽情景記述を，Web 上で不特定多数のユーザが協調して編集可能な世界初のシステムであり，スタンドアロンアプリケーションと違ってインストールが不要で誰でも気軽に利用できるという特長を持つ．

4.2 Songle の貢献と意義

本研究は，音楽理解技術に基づく能動的音楽鑑賞インタフェースを楽しむための世界初の Web サービスを公開して，エンドユーザの役に立つという社会的意義を持っている．さらに，音楽理解技術は，すでに普及しつつある音声認識技術や画像理解技術と違い，そもそもそういう技術があるということ自体の認知度が高くなく，Songle によって音楽理解技術の潜在的な可能性が認知され，他の応用事例開拓に波及する効果が期待できる．また，音楽理解技術で Web 上の様々な楽曲に対して推定した結果をユーザが見れば，どのような箇所で推定が難しいかが分かる．そこで誤認識が多い場合には，批判を受けるリスクはあるが，そうした現状をユーザと共有してはじめて，音楽理解技術の真の普及と発展があると我々は考える．

本研究の学術的意義は，不特定多数のエンドユーザに誤り訂正の協力をしてもらうことで，サービスの利便性とユーザによる利用率をどこまで向上させることができるかを探求することにある．こうした発想は，従来の信号処理に基づく音楽理解研究にはなかった．この新たな研究アプローチでは，(i) ユーザが音楽理解技術に基づくサービスを利用することでその性能を理解する，(ii) そのサービス改善にユーザが貢献する，(iii) その改善がより良いユーザ体験に結び付く，という 3 段階からなるポジティブスパイラルを回すことができる点が重要である．(iii) のユーザ体験の向上が，(i) のサービス利用を促進するからである．従来の GWAP（game with a purpose）や人間計算 [26]（ESP Game [21] も含む）といったゲームの楽しさをインセンティブとしたクラウドソーシングのアプローチでは，この (iii) という重要な段階が欠けていた．Songle は，多数のユーザの訂正結果を Web サービス上で共有して性能改善を図る**社会的訂正**の枠組みであり，他のユーザの利便性に貢献している実感が得られるうえに，他のユーザが訂正している活動を見ることで，訂正の意欲も高まる点が優れている．このように Songle では，集合知（wisdom of crowds）やクラウドソーシングを活用しつつ，ユーザ体験向上を実現していく．

今後取り組むべき重要課題の 1 つは，ユーザによる誤り訂正の協力で，音楽理解技術の性能をどこまで高くできるかを探求することである．訂正結果を機械学習して性能向上させる仕組みはまだ実装しておらず，今後，十分な訂正量が蓄積された段階で実現する予定である．それが実現できれば，「ユーザの貢献を増幅」する新たな音楽情報処理の枠組みとなる．Wikipedia（http://www.wikipedia.org/）等の典型的な Web 2.0 の Web サービスでは，ユーザの貢献は編集した項目に限定される．一方，こうしたユーザ貢献増幅の枠組みが実現できれば，訂正内容の学習によって音楽理解技術が向上することで，まだ訂正していない楽曲に対する性能改善も期待できる．このユーザ貢献増幅こそが，従来の Web 2.0 や人間計算 [26] にない特長であり，これまで音声認識に基づく音声情報検索サービス「PodCastle」[27] の研究でその重要性と可能性が示されてきたが，今後は Songle で音楽情報処理における可能性も示していきたい．

我々は「ユーザを信頼する」立場から，PodCastle [27] 同様に，基本的にはユーザによる訂正の質は高いものと考えている．仮にユーザが故意に不適切な訂正（いたずら）を

した場合でも，その信頼性（訂正が楽曲内容と合致するか）を音響的に検証する方法が実現できる可能性があり，新たな研究課題として興味深い．また，不適切な訂正に気づけば，誰でもその前の状態に戻すことが可能な機能も提供している．

4.3 研究プラットフォームとしての Songle の発展

今後，他の研究者が開発した音楽理解技術による推定結果も提示可能とすることで，共同で社会に対して貢献していく研究プラットフォームに Songle を発展させていきたいと考えている．そうした技術を，図 5 の音楽理解モジュールとして実装すれば，ソースコードやバイナリコードを共有せずに，世界中のどこでも（ファイアウォールの中でも）実行可能な実装にすでになっている．ある音楽情景記述について複数のモジュールの結果が得られたら，その違いを比較・可視化するのも興味深い．そうすれば，そうした異なる結果を訂正の候補としたり，まとめ上げたりする仕組みも，今後研究対象にできる可能性がある．

5. おわりに

本論文では，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 を提案した．4 種類の主要な音楽情景記述を自動推定し，Web ブラウザ上で動作するユーザインタフェースによって可視化したことで，ユーザはそれを見て音楽に対する理解を深めながら，インタラクティブに再生位置を変更して音楽を楽しむことができるようなった．自動推定結果は不完全でも，その誤りをユーザが自発的に訂正して他のユーザと共有していくことで，ユーザ体験が向上するポジティブスパイラルを回す仕組みも実現できた．

ただし，音楽情景記述の可視化方法には，多様なユーザに向けて異なる選択肢があるとよい．そこで，能動的音楽鑑賞の観点から見ても効果的なビジュアライザ（視覚エフェクトや音楽同期アニメーション）機能等の，よりいっそう分かりやすい方法を探求したい．また，音楽情景記述を訂正するインセンティブも高めていく必要がある．楽曲構造とメロディラインは，音楽地図や類似楽曲グラフが改善されるので分かりやすいが，ビートは他の音楽情景記述の位置を決めやすくする効果が主で，コードもまだ見たり聴いたりして楽しいという以上の効果がない．そこで，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 API 機能等を追加する拡張に取り組んでいる．さらに，ある音楽情景記述の訂正が他の種類の音楽情景記述の改善につながる機能の追加や，音楽理解技術の性能向上，インタフェース改善によって，訂正の労力も低減していく予定である．

謝辞 Songle の Web サービスの実装を担当していただいた川崎裕太氏，Web デザインを担当していただいた櫻井稔氏に感謝する．本研究の一部は JST CREST の支援を受けた．

参考文献

[1] Aucouturier, J.-J. and Pachet, F.: Improving Timbre Similarity: How high is the sky?, *Journal of Negative Results in Speech and Audio Sciences*, Vol.1, No.1 (2004).
[2] 後藤真孝：SmartMusicKIOSK：サビ出し機能付き音楽試聴機，情報処理学会インタラクション 2003 論文集，pp.9–16 (2003).
[3] Goto, M.: Active Music Listening Interfaces Based on Signal Processing, *Proc. ICASSP 2007* (2007).
[4] 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).
[5] Goto, M.: A Chorus-Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station, *IEEE Trans. ASLP*, Vol.14, No.5, pp.1783–1794 (2006).
[6] 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).
[7] Goto, M.: Music Scene Description Project: Toward Audio-based Real-time Music Understanding, *Proc. ISMIR 2003*, pp.231–232 (2003).
[8] 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).
[9] 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. ASLP*, Vol.18, No.3, pp.638–648 (2010).
[10] Goto, M.: Music Listening in the Future: Augmented Music-Understanding Interfaces and Crowd Music Listening, *Proc. AES 42nd International Conf. on Semantic Audio*, pp.21–30 (2011).
[11] Mauch, M. and Dixon, S.: Approximate Note Transcription for the Improved Identification of Difficult Chords, *Proc. ISMIR 2010*, pp.135–140 (2010).
[12] 藤原弘将，後藤真孝，奥乃 博：歌声の統計的モデル化とビタビ探索を用いた多重奏中のボーカルパートに対する音高推定手法，情報処理学会論文誌，Vol.49, No.10, pp.3682–3693 (2008).
[13] Fujihara, H., Goto, M., Ogata, J. and Okuno, H.G.: LyricSynchronizer: Automatic Synchronization System Between Musical Audio Signals and Lyrics, *IEEE Journal of Selected Topics in Signal Processing*, Vol.5, No.6, pp.1252–1261 (2011).
[14] Mauch, M. and Dixon, S.: Simultaneous Estimation of Chords and Musical Context from Audio, *IEEE Trans. ASLP*, Vol.18, No.6, pp.1280–1289 (2010).
[15] Lee, J.H.: Crowdsourcing Music Similarity Judgments using Mechanical Turk, *Proc. ISMIR 2010*, pp.183–188 (2010).
[16] Mandel, M.I., Eck, D. and Bengio, Y.: Learning Tags That Vary within a Song, *Proc. ISMIR 2010*, pp.399–404 (2010).
[17] Turnbull, D., Liu, R., Barrington, L. and Lanckriet, G.: A Game-Based Approach for Collecting Semantic Annotations of Music, *Proc. ISMIR 2007*, pp.535–538 (2007).

[18] Turnbull, D., Barrington, L. and Lanckriet, G.: Five Approaches to Collecting Tags for Music, *Proc. ISMIR 2008*, pp.225–230 (2008).
[19] Mandel, M.I. and Ellis, D.P.W.: A Web-Based Game for Collecting Music Metadata, *Proc. ISMIR 2007*, pp.365–366 (2007).
[20] Law, E.L.M., von Ahn, L., Dannenberg, R.B. and Crawford, M.: TagATune: A Game for Music and Sound Annotation, *Proc. ISMIR 2007*, pp.361–364 (2007).
[21] von Ahn, L. and Dabbish, L.: Labeling Images with a Computer Game, *Proc. CHI 2004*, pp.319–326 (2004).
[22] Cannam, C., Landone, C., Sandler, M. and Bello, J.P.: The Sonic Visualiser: A Visualisation Platform for Semantic Descriptors from Musical Signals, *Proc. ISMIR 2006*, pp.324–327 (2006).
[23] Li, B., Burgoyne, J.A. and Fujinaga, I.: Extending Audacity as a Grouth-Truth Annotation Tool, *Proc. ISMIR 2006*, pp.379–380 (2006).
[24] Amatriain, X., Massaguer, J., Garcia, D. and Mosquera, I.: The CLAM Annotator: A Cross-platform Audio Descriptors Editing Tool, *Proc. ISMIR 2005*, pp.426–429 (2005).
[25] Herrera, P., Celma, Ò., Massaguer, J., Cano, P., et al.: MUCOSA: A Music Content Semantic Annotator, *Proc. ISMIR 2005*, pp.77–83 (2005).
[26] von Ahn, L.: Games With A Purpose, *IEEE Computer Magazine*, Vol.39, No.6, pp.92–94 (2006).
[27] 後藤真孝，緒方　淳，江渡浩一郎：PodCastle：ユーザ貢献により性能が向上する音声情報検索システム，人工知能学会誌，Vol.25, No.1, pp.104–113 (2010).

推薦文

インタラクション 2012 では，87 名から構成されるプログラム委員会によって投稿数 43 件の中から優秀な論文 18 件を一般講演発表として採択し，インタラクティブ発表は 149 件の投稿から 19 件をファイナリストとして選出いたしました．本論文は，これらの 37 件からさらにプログラム委員会による投票によって，論文誌に推薦すべき論文であるとの評価を得たものであり，論文誌編集委員長としてもぜひ推薦したいと考えました．

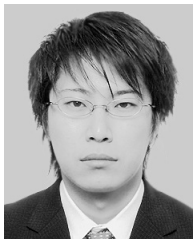
（インタラクション 2012 プログラム委員長　宮下芳明）

後藤 真孝　（正会員）

1998 年早稲田大学大学院理工学研究科博士後期課程修了．博士（工学）．同年電子技術総合研究所に入所し，2001 年に改組された産業技術総合研究所において，現在，情報技術研究部門上席研究員兼メディアインタラクション研究グループ長．統計数理研究所客員教授，筑波大学大学院准教授（連携大学院），IPA 未踏 IT 人材発掘・育成事業プロジェクトマネージャーを兼任．ドコモ・モバイル・サイエンス賞基礎科学部門優秀賞，科学技術分野の文部科学大臣表彰若手科学者賞，情報処理学会長尾真記念特別賞等，31 件受賞．

吉井 和佳　（正会員）

2008 年京都大学大学院情報学研究科博士後期課程修了．博士（情報学）．同年より産業技術総合研究所情報技術研究部門研究員．統計的機械学習技術に基づく音楽情報処理の研究に従事．山下記念研究賞，船井研究奨励賞等受賞．電子情報通信学会，IEEE 各会員．

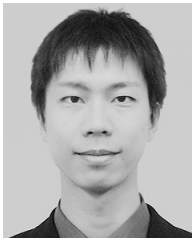


藤原 弘将　（正会員）

2005 年京都大学工学部情報学科卒業．2007 年同大学大学院情報学研究科知能情報学専攻修士課程修了．同年産業技術総合研究所に入所．博士（情報学）．2010 年京都大学大学院情報学研究科知能情報学専攻博士課程修了．音楽情報処理，音楽情報検索，音声情報処理に興味を持つ．平成 19 年度山下記念研究賞受賞．日本音響学会，電子情報通信学会各会員．

Matthias Mauch

Matthias Mauch received his Ph.D. in Electronic Engineering from Queen Mary, University of London, in 2010. He is currently a Royal Academy of Engineering Research Fellow, and Lecturer in Digital Signal Processing at Queen Mary. His research interests are centred around Music Informatics and include the automatic extraction of harmony, rhythm and temperament from audio. He is also interested in the mining of musical patterns in symbolic data, and the study of singing intonation and musical evolution.



中野 倫靖 （正会員）

2003 年図書館情報大学卒業．2008 年筑波大学図書館情報メディア研究科博士後期課程修了．博士（情報学）．現在，産業技術総合研究所研究員．日本音響学会会員．2006 年日本音楽知覚認知学会研究選奨，2007 年インタラクション 2007 インタラクティブ発表賞，2009 年情報処理学会山下記念研究賞（音楽情報科学研究会），2010 年音楽情報科学研究会（夏のシンポジウム 2010）ベストプレゼンテーション賞各受賞．