

リズム入力インタフェース「タタタタップ」による 大規模音楽検索

池谷 直紀 服部 正典 梅木 秀雄 大須賀 昭彦

株式会社東芝 研究開発センター

我々は1つのボタンをリズムカルに押す入力方式「タタタタップ」を提案しており、これを大規模音楽検索に適用するための新たな検索アルゴリズムを開発した。本稿ではその方式と、2500曲の実際の楽曲データを用いた実験による検索精度を示す。結果として、2500曲の用意した楽曲の中の一曲を想起してその印象的なフレーズのリズムを打鍵し検索する実験において、意図した楽曲が検索結果の上位五位以内に提示される確率が75%に達することを確認した。

Large Scale Music Search Using Rhythm Input User Interface “Tatata-Tap”

Naoki IKETANI Masanori HATTORI Hideo UMEKI Akihiko OHSUGA
Corporate Research & Development Center, Toshiba

We are researching an input user interface called “Tatata-Tap” that use only one button and rhythmical input pattern. We developed new algorithm for applying Tatata-Tap to large scale music searching systems. This paper shows the method and experimental evaluation that use 2500 of music. As a result of the experiment, we confirm the ratio that the system shows the correct answer in the best 5 candidates is 75 percent.

1. はじめに

筆者らは1つのボタンをリズムカルに打鍵する入力ユーザインタフェース「タタタタップ」を提案しており[1]、パターンマッチング技術を用いてこれを実現している。また、タタタタップを用いた音楽検索方式についても提案しており、100曲を対象とした場合の有効性を確認している[2]。しかしその後の研究により、先に提案した手法では実用を想定した大規模な楽曲検索実験においては意図した候補が提示される率が低く実用には遠いことが確認されていた。本稿では、1000曲以上の大規模な楽曲検索に適したタタタタップの新たなアルゴリズムを示す。また2500曲の実際の楽曲を用いた実験による評価結果を示し、実用可能性を考察した上で今後の課題を明らかにする。

2. 背景

2.1 タタタタップの概要

タタタタップは、1つのボタンをリズムカルに打鍵する入力方式であり、打鍵のパターンに基づいて機器の操作や楽曲の検索を行うことができる方式である。1つのボタンの押し方による入力方式、すなわちワンキーオペレーションは古来モールス信号が有名であり、最近ではUbiButton[3]が提案されている。また広義には各種機器におけるボタンの長押しもその一種である。一般にワンキーオペレーションは以下の特徴を備えている。

- ・広範な入力デバイスへの適用が可能
- ・画面を見る必要がなく簡易に入力が可能

タタタタップを既存のワンキーオペレーション方式と比較すると、楽曲の主旋律のような複雑なリズムの入力を用いることができるため、さらに以下の特徴が加わっている。

- ・入力パターンを楽曲と結びつけることに

より利用者が記憶しやすい

- ・短時間での入力パターンが多彩である

タタタタップはその性質上、文字列の入力には向いていないが、あらかじめリズムと動作を関連付けておくことで、リズムをコマンドとして入力することにより機器を操作することができる。そしてタタタタップのもう一つの主な応用が楽曲検索である。

タタタタップの動作イメージを図1に示す。仕組みの詳細については3章で述べる。

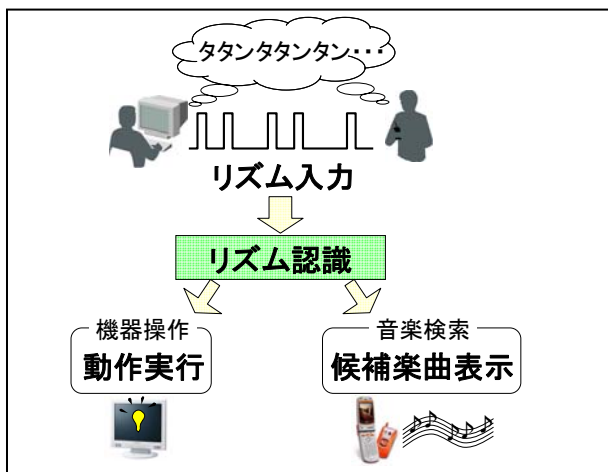


図1：タタタタップの動作イメージ

2. 2 音楽検索技術の動向

現在、携帯電話の着信メロディの市場は着うたなどによりさらに拡大している。一方では音楽配信ビジネスも普及しつつあり、楽曲の購入は店頭で買うスタイルだけでなくデジタルデータとしてダウンロードする形態へと展開している。こういった市場動向に伴い、電子的な楽曲検索のニーズは拡大している。

現行の楽曲検索の方式としては、あらかじめジャンル別などに分類したリストを用いて利用者が探す方式や、曲名や作曲者や演奏者名など曲に付随する情報の文字列をキーとするキーワード検索方式が一般的である。その他に、曲の旋律を口ずさみ音声入力で検索するハミング検索[4]や、その曲の音響信号そのものを入力として楽曲を識別する方式[5]などがあり、これらの方式には以下の共通の特徴がある。

- ・曲名などを知らない楽曲を検索可能
- ・キーワード検索よりも素早く候補曲を絞り込める場合が多い

タタタタップもこれらの特徴を備えた楽曲検索方式として用いることができ、主旋律のリズ

ムを打鍵することによりそのリズムを含む楽曲の検索をすることができる。

3. タタタタップの構成と動作

タタタタップを音楽検索に用いる場合のシステム構成を図2に示す。あらかじめ、検索対象である音楽のリズムを楽曲IDと関連付けてリズムパターンDBに登録しておく。一つの楽曲IDについても複数のリズムパターンが登録可能となっている。これは、一曲の中でも異なる部分の打鍵や、人によって異なる打鍵入力に対応するためである。

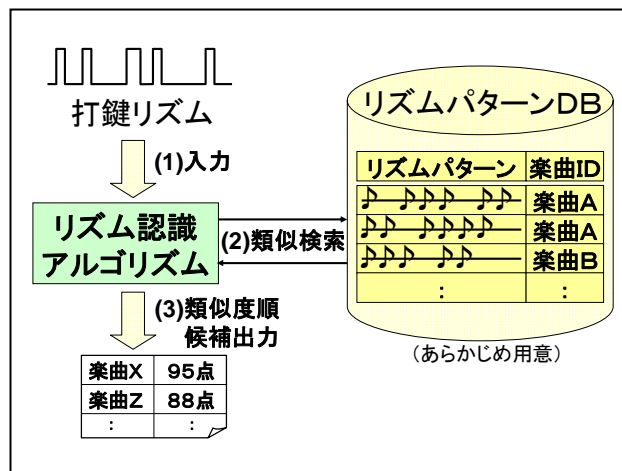


図2：タタタタップの構成と動作

そして利用者がボタンなどの入力装置を用いて打鍵した打鍵リズムを、ON/OFFの時系列信号として入力する。リズム認識アルゴリズムはリズムパターンDBを参照して、入力された打鍵リズムに類似したリズムパターンを1つあるいは複数選出し、その楽曲IDと類似度のリストを出力する。

以上がタタタタップの中心部の挙動であり、実用シーンにおいては事前のリズムパターンのセットや、楽曲IDに基づいて楽曲名を表示するなどの処理を行う。

タタタタップは簡易な入力方式であることを目指しているため、基本設計として以下の3項目を前提としている。

- (1) 打鍵のタイミングのみを扱う
ボタンを押している長さの情報は活用しない。このため、リズムパターンに音長、休符の情報は不要となる。
- (2) 打鍵のテンポ情報は用いない
登録されたリズムパターンのテンポと異なっても、パターンが類似していれば類似し

たりズムとみなす。

(3) 前方一致によるリズムの類似判定を行う

すなわち、打鍵パターンと冒頭と、登録されたリズムパターンの冒頭を一致しているものとみなす。そのためこれを前提としたリズムパターンを用意せねばならない。これは大きな制約であるが、先行研究[2]により部分一致方式では精度が大幅に落ちてしまうことが確認されているためである。ただし将来的には部分一致に近い方式に工夫する余地はある。

以上の前提から、打鍵リズムは打鍵時刻の列、リズムパターンは音符の立ち上がり時刻の列として表現され、データ形式としては同一となる。そしてリズム認識アルゴリズムは打鍵リズムと個々のリズムパターンの類似度を判定し、結果として類似度の高いリズムパターンを選出する。

一見、タタタタップによる楽曲検索はハミング検索から音高情報を省いたサブセットのように捉えられるが、入力が打鍵されたリズムのみであるため、利用シーンにおいてはマイクおよび発声が不要、技術的には少ない情報量における類似判定が必要という点で本質的に異なる。

4. リズム認識アルゴリズム

タタタタップの大きな技術的課題は、入力された打鍵リズムと用意されたリズムパターンとの類似判定である。入力者に打鍵されたリズムはテンポが不明であり誤差もあるため、入力者の意図するリズムを認識することは容易ではない。これは楽器の演奏情報から譜面を生成する「採譜」技術におけるリズムの特定に通ずるものであるが、タタタタップにおいては用意された複数のリズムパターンと入力された打鍵リズムの類似判定であるという点で異なる。

本稿では採譜の研究[6]で提案され、タタタタップの先行研究[2]で活用した「リズムベクトル方式」と、その欠点を考察して新規に開発した「曖昧検索アルゴリズム」を用いて評価を行った。

4.1 方式A：リズムベクトル方式

先行研究[2]と同様に、連続する3つの音長の比をリズムベクトルとし、打鍵リズム \mathbf{r}_i と一つのリズムパターン \mathbf{r}_p に含まれるリズムベクトルの列を比較して類似度を判定している。リズムベクトル \mathbf{r}_{in} と \mathbf{r}_{pn} の距離 $d(\mathbf{r}_{in}, \mathbf{r}_{pn})$ をユークリッド距離を用いて算出し、その総和

$D = \sum_n d(\mathbf{r}_{in}, \mathbf{r}_{pn})$ をリズムパターン間の距離として類似判定を行った。

4.2 方式B：曖昧検索アルゴリズム

本アルゴリズムが本稿で提案する新たな手法であり、タタタタップによる楽曲検索のために開発した独自の方式である。詳細は複雑化しており容易に数式では表現できないため、ここではその処理の概要を述べる。

この方式の考え方は、ベースとなるのはリズムベクトル方式と同様にリズムパターンと打鍵タイミングの細かいずれを積算するものであるが、打鍵の欠落や増加および大きなずれがある程度は許容し、曖昧性を考慮した類似判定を行うものである。これは、検索者は必ずしも用意したリズムパターンどおりには打鍵せず往々にして打鍵の欠落や増加が発生する、という事前実験により得られた知見に基づいて開発している。曖昧性の判定についてはDPマッチングの考え方を導入して実現している。

(1) 二つのリズムの全体を重ね合わせる

打鍵数が同一でない場合も考えられるため、時間間隔の長い部位を基準として二つのリズムの時間間隔の対応付けを決定する。

(2) リズムの大きな差異について積算

個々の時間間隔について、大きな差異を検出し、総合的にそれらがリズムの類似性に及ぼす影響度を算出して積算する。上述の打鍵の欠落や増加による類似度の判定である。

(3) 細かい差異の積算

時間間隔の差異が比較的小さい(2)で考慮されなかった部分について、細かい差異を積算する。このフェーズはこのアルゴリズムでは単純な差分を積算しているが、リズムベクトル方式の考え方に相当する部分である。

(4) (1)～(3)について最尤となる重ね合わせ方についての類似度を算出する

(1)～(3)の手順により積算した値が二つのリズムの距離に相当し、これを用いて類似度を算出することが可能であるが、入力の曖昧性を考慮することにより手順(1)における重ね合わせは唯一ではない。そこでDPマッチングの考え方を導入し、最も距離の小さい重ね合わせ方における類似度を算出する。類似度は単に距離を定数から引いたものとして算出した。

5. 実験による評価

上述の2つのアルゴリズムについて、実際の楽曲データを用いた実験による評価を行った。

5. 1 実験方法

実験に用いたのは日本のポピュラー音楽を中心とする比較的メジャーな楽曲 2500 曲であり、MIDI (Musical Instrument Digital Interface) 形式のデータを用いた。

実験は、2500 曲を用いて 6 人の被験者により実施した。実験の手順は、まず打鍵フェーズとして、曲名リストから知っている曲を探しその楽曲の印象的なフレーズを想起して打鍵する。その後リズムパターン作成フェーズとして、楽曲中の印象的なフレーズの箇所を指定し、リズムパターンの生成を行う。そしてリズムパターン DB を構築した上で、打鍵フェーズで収集した打鍵リズムを入力としてタタタタップによる楽曲検索を実行し、被験者の意図した楽曲が検索結果の上位何番目に提示されるかを結果として集計した。

打鍵フェーズは実用に近い状況を想定し、曲を聴く前に打鍵を実施した。同一曲を複数の被験者が別々に打鍵する場合もある。打鍵が可能な程度に被験者がフレーズを想起できる曲という条件で入力を行ったところ、結果としては 441 曲、延べ 648 種の打鍵リズムが入力された。

リズムパターン作成フェーズにおいては、全楽曲を被験者が分担して作業を実施した。楽曲中に往々にして複数存在する印象的フレーズについて、フレーズの開始時刻と、主旋律の MIDI チャンネル番号とを主観的に判断し手動で指定した上で、該当部分を機械的にリズムデータに変換し、これをリズムパターンとして用いた。知らない楽曲である場合には数回聴いて印象的なフレーズを指定している。

この際の曲数に対するリズムパターンの数を表 1 に示す。

表 1：曲数に対するリズムパターン数

曲数	500	1000	1500	2000	2500
リズム数	1900	3800	5700	7600	9500

印象的フレーズは一曲当たり 3.8 箇所指定されたことがわかる。これは主観的な作業の結果であり、例えば、曲のイントロ、Aメロ、Bメロ、サビの 4 つのようなフレーズを印象的とみなしたということである。

5. 2 実験結果

実験 1：方式 A 「リズムベクトル方式」

全ての打鍵リズムについて、方式 A のリズムベクトル方式によるリズム類似判定を用いて楽

曲検索を行った。母集団の曲数を 500 曲から 2500 曲まで変動させた場合の結果を図 3 に示す。なお理由は後述するが、この曲数別結果においては打鍵数 24 までにおける最も高い正解率を結果として採用した。

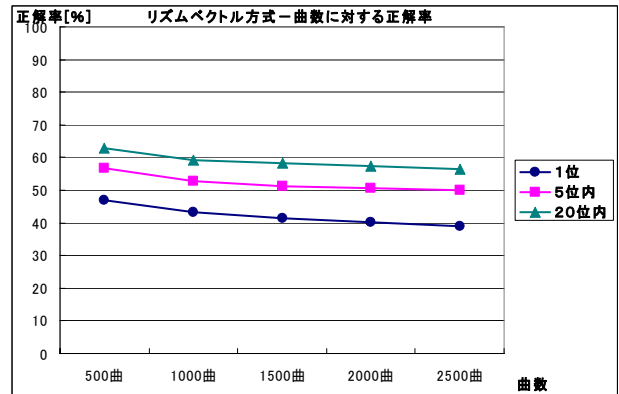


図 3：実験 1 の曲数別結果

この結果、検索の結果第一候補として提示された楽曲が意図したものである「第一位正解率」は 500 曲で 47%、2500 曲では 39% となっている。

次に、この 2500 曲の場合における打鍵数ごとの正解率を図 4 に示す。

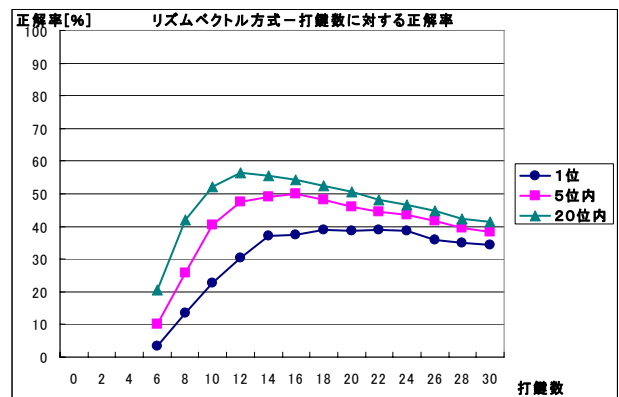


図 4：実験 1 の打鍵数別結果

図 4 から、打鍵数が一定数を超えると正解率が下落していることがわかる。例えば正解が上位五位に含まれる「第五位内正解率」においては打鍵数 16 にピーク値があり（これが図 3 のプロット値となっている）、それを超えると打鍵数を増加するにつれて正解率が下落している。これは、打鍵数が増加するにつれて、類似パターンが減り識別しやすくなる正解率上昇の要因よりも、被験者の打鍵リズムと登録されたリズムパターンの乖離が次第に広がる正解率下落の要因が強まることに起因している。これに関してさらに詳細な分析を行ったところ、打鍵リズ

ムとリズムパターンの乖離は手ぶれなどの誤差による要因でも次第に広がるが、それ以上に入力者がどのようにリズムを把握しているかという認識に起因する場合が大半であることがわかった。すなわち、入力者は譜面どおりに演奏するのではなく自分の想起したその曲のフレーズを打鍵するため、打鍵の増加や欠落や変動など、用意されたパターンとの食い違いがしばしば発生するのである。

実験 2：方式 B「曖昧検索アルゴリズム」

実験 1 を踏まえて新たに開発した曖昧検索アルゴリズムを用いた場合の結果を以下に示す。

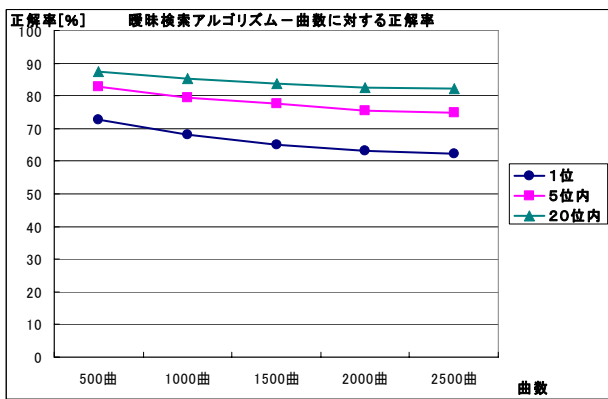


図 5：実験 2 の曲数別結果

方式 A と比較して大幅に検索精度は向上し、第一位正解率が 500 曲では 73%、2500 曲では 62% となった。また第五位内正解率は 500 曲で 83%、2500 曲で 75% となっている。これが本稿で示す主要成果である。改善された理由は、方式 A では曖昧性を許容しない方式であったため類似していないと判定された打鍵の増減や変動が含まれる入力について、曖昧検索アルゴリズムでは類似していると判定できる場合があったことである。これにより検索精度が約 20%~25% 向上していることが読み取れる。

次に実験 1 と同様に、2500 曲における打鍵数ごとの正解率を図 6 に示す。

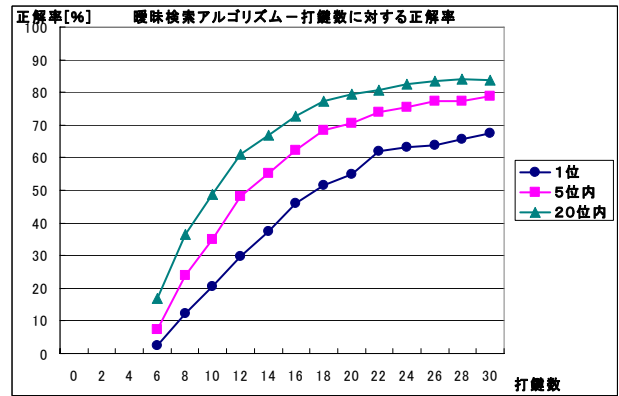


図 6：実験 2 の打鍵数別結果

実験 1 とは異なり、打鍵数が増加するにつれて正解率が増加する右肩上がりのグラフとなっている。これは曖昧性を考慮した類似判定の効果であり、打鍵数の増加による打鍵リズムとリズムパターンの乖離の広がりによる検索精度低下要因の影響を大幅に軽減できているためである。結果として、打鍵数の増加に伴うリズムパターンの多様化という検索精度上昇の要因が効くこととなり右肩上がりとなるのである。

一方で、打鍵数 1 2 以下においては方式 A の図 4 よりも値が低い。これは曖昧性を考慮したことの副作用であり、打鍵数が少ない状況では、曖昧性を考慮した類似判定により、本来は一致していないリズムパターンを類似していると誤判定してしまう場合が精度を下げている。

ここで、打鍵数に対する時間の関係を図 7 に示す。これは全試行について平均したものである。これによれば、平均すると 25 打鍵目で 10 秒を超えている。そこで、打鍵すればするほど正解率が上昇する実験 2 の場合にも、図 5 の曲数別比較においては実際の利用に近い状況として 10 秒間で入力できる 24 打鍵以内での最大値を正解率とした。

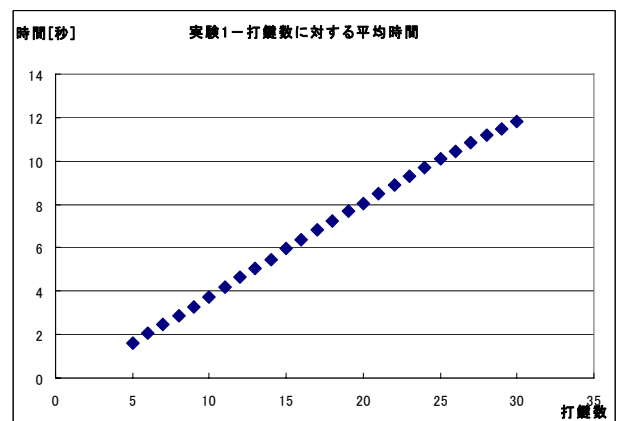


図 7：打鍵数と時間の関係

6. 考察

まず、本稿で用いた二つのアルゴリズムについて、打鍵数が12以下の範囲においてはリズムベクトル方式の方が優れていた。このことから、単純に打鍵数が少ない場合はリズムベクトル方式を用い、打鍵数が多い場合には曖昧検索アルゴリズムを用いるハイブリッド方式が考えられる。これは両方式の良いところを組み合わせた優れたものとなることは明らかである。さらにこの考えを推し進めると、打鍵数の増加に伴い、曖昧性の許容度を次第に大きくするアルゴリズムがより優れていると考えられる。この作成と検証は今後の課題の一つである。

次に、打鍵リズムに対して類似したリズムを検索する際に、正解率を低下させる要因を整理する。

- (1) テンポの相違
 - (2) リズムパターンDB中のテンポ以外が同一なリズムの存在
 - (3) 手ぶれによる誤差
 - (4) 被験者が想起したリズム範囲は正解のリズムパターンとして登録されているが、やや異なる
 - (5) 被験者が想起したリズムがリズムパターンに登録されていない
- これらの要因について影響の度合いを考察し、その対策をまとめる。

(1) テンポの相違

テンポの相違は現在のタタタタップにおいてはアルゴリズムで完全に吸収するよう設計された項目であり、今回の設計において影響は無い。

(2) テンポ以外が同一のリズムの存在

今回用いた2500曲、9500種のリズムパターン中にも、例えば開始からしばらく八分音符の連続だけで構成されるような単調なリズムパターンは存在するが、完全に同一のリズムは多くはない。20打鍵以上のリズムパターンについて完全に一致する他のリズムパターンの存在確率は数パーセントである。

タタタタップはリズムの情報のみで検索する方式であるため、この要因を完全に解決することは不可能だが、実用の際にはリズムパターンDB構築時に同一リズムが登録されないように制御することが対策として挙げられる。

(3) 手ぶれによる誤差

手ぶれによる入力誤差はタタタタップによる楽曲検索における主要な解決課題の一つであり、

今回の二つの方式でもある程度は吸収している。手ぶれが大きくなるにつれて、意図した結果ではない類似したリズムパターンが上位に提示されてしまう可能性が高くなり、意図した楽曲の順位は次第に下がることとなる。図4において、第一位正解率に対して第五位正解率、第20位正解率が高い理由は(2)と(3)の影響が大部分である。

これも完全な解決は不可能である。被験者の熟練度によっても誤差の大きさはさまざまである。対策としてはぶれの発生しやすいパターンなどのヒューリスティックを考慮したアルゴリズムの構築や、打鍵されたリズムを部分的に打鍵し直すなど入力を支援する方法が考えられる。

(4) 想起したリズムが登録リズムとやや異なる

手ぶれ要因に劣らず大きい影響力であることが今回明確となったのが、想起し入力するリズムは用意したリズムパターンどおりではない、という入力者のリズムの把握に起因する要因である。曖昧検索アルゴリズムを用いた結果の改善はこの要因をある程度解決したことによるものであり、図4と図6の比較によれば24打鍵においては解決されただけでも20~30%の検索精度向上となっている。

しかしこれも完全な解決は不可能であり、過度な曖昧性の許容は検索精度の低下につながってしまう。どのようなパターンにおいて入力者の打鍵が異なることが多いかというヒューリスティックを考慮するなどの曖昧検索アルゴリズムの改良が望まれる。

(5) リズムパターンが登録されていない

打鍵リズムが楽曲中には存在するがリズムパターンとして用意されていないという場合は本稿の実験でも存在しているが、対策を採っていない。多様な利用者に対して妥当なリズムパターンDBを如何に構築するかという方法論が必要となる。また、今回は一曲当たり平均3.8箇所の印象的なフレーズをリズムパターンとした上で前方一致検索を前提としたが、部分一致方式に近い方式とするための新たな方式の研究も考えられる。

以上がタタタタップを用いた楽曲検索における検索精度低下の要因と、その影響力と対策および今後の改良方針である。

最後に、音楽検索方式としてのタタタタップの性質を他の音楽検索方式と定性的に比較する。

表 2 : 音楽検索方式の定性比較

	検索 精度	曲名 知識	入力 簡易性	ノイズ 耐性
キーワード検索	◎	要	○	○
ハミング検索	○	不要	△	×
音響指紋方式	○	不要	×	△
タタタタップ	△	不要	◎	○

タタタタップは他の音楽検索方式と比較して検索精度では比較的劣っているものの、曲名を知らずともダイレクトに検索ができる。また他の曲名を知らずとも検索できる技術と比較した場合には、ノイズの影響を受けないことと、ボタン一つがあればよいため簡易なデバイスでどこでも入力が可能である入力の簡易性が長所である。

7. まとめと今後の課題

本稿で得られた知見を元に、タタタタップを大規模楽曲の検索方式として実用する形態を検討する。例えば 2500 曲の対象楽曲から着信メロディをタタタタップにより検索し、候補の曲リストを一画面に 5 曲提示するようなサービスを想定すれば、最初の候補画面に正解が含まれる可能性（第五位内正解率）が 75%であり、4 ページ以内なら 82%となる。ただし、多数の被験者に対して本稿の実験と同様に機能するかどうか、携帯電話のボタンで入力する場合にはどうかなどについては別途評価が必要であるが、応用形態次第では実用可能な検索精度であると、筆者は考える。

リズムパターンDBの構築はタタタタップの実際のサービスへの適用を行う際には非常に重要かつ困難な課題となる。MIDI データでさえ、今回の実験で説明した以下の 2 項目の手作業が必要となる。

- ・主旋律のチャンネル（パート）の特定
- ・印象的なフレーズの開始箇所の特定

後者については楽曲のサビ検出アルゴリズム[7]の応用による自動化が考えられる。さらに今後の課題としては、MIDI ではない音響波形の音楽データを用いるために、音響波形データから主旋律の音を抽出し譜面化するような技術により、自動か半自動でのリズムパターン生成方式の確立が望まれる。

タタタタップの今後の展開としては、検索精度の向上、より大規模な楽曲検索への適用、そして多数の被験者による検証を実施して技術の確立と応用への適用が期待される。

参考文献

- [1]池谷直紀, 服部正典, 大須賀昭彦: “リズム入力インタフェース「タタタタップ」”, 情報処理学会第 66 回全国大会 4A-4
- [2]池谷直紀, 服部正典, 大須賀昭彦: “リズム入力による音楽検索方式「タタタタップ」”, 第三回情報科学技術フォーラム FIT2004, Sep, 2004.
- [3]福本雅明, 外村佳伸, “指釦:手首装着型コマンド入力機構”, 情報処理学会論文誌, Vol. 40 No. 2, Feb 1999
- [4]蔭山哲也, 高島洋典: “ハミング検索を手掛りとするメロディ検索”, 電子情報通信学会論文誌 Vol. J77-D-II, No. 8, 1999
- [5]“GraceNote Mobile MusicID”, http://www.gracenote.com/gn_japan/mobileMusic.html
- [6]武田晴登, 篠田浩一, 嵯峨山茂樹等: “リズムベクトルを用いたリズム認識”, 情報処理学会研究報告「音楽情報科学」No. 46, 2002
- [7]後藤 真孝: “リアルタイム音楽情景記述システム: サビ区間検出手法”, 情報処理学会 音楽情報科学研究会 研究報告 2002-MUS-47-6, No.100, Oct 2002.