

ロドラムによるドラムパターン検索手法

中野 倫靖[†] 緒方 淳[‡] 後藤 真孝[‡] 平賀 譲[†]

[†]筑波大学 [‡]産業技術総合研究所

[†]{nakano,hiraga}@slis.tsukuba.ac.jp [‡]{jun.ogata,m.goto}@aist.go.jp

あらまし 本稿では、人がドラムの音を真似て口ずさんだ音声（ロドラム）を認識し、それに対応するドラムパターンを検索する手法を提案する。従来、実際のドラム音（楽器音）を対象とした認識は研究されてきたが、ロドラムは研究されていなかった。ロドラム認識では、声質とドラム音表現の両方の個人差への対処が問題となるため、従来のドラム音認識手法は適用できない。そこで本手法では、擬音語を中間形式として採用することでこの問題に対処する。擬音語の各音素をロドラム音のスペクトル構造へ対応付けるために確率モデルを用い、声質の個人差を吸収する。また、各ドラム音に対応する擬音語の辞書を用意して、表現の個人差に対処する。200 発話のロドラムデータに対して実験した結果、91.5%の認識率を得た。

A Drum Pattern Retrieval Method by Voice Percussion

Tomoyasu Nakano[†] Jun Ogata[‡] Masataka Goto[‡] Yuzuru Hiraga[†]

[†]University of Tsukuba

[‡]National Institute of Advanced Industrial Science and Technology (AIST)

Abstract This paper proposes a method of recognizing voice percussion (simulated drum sound by voice) and retrieving the corresponding drum pattern from a database. Although drum sound recognition has been the topic of existing work, there has been no previous attempt that dealt with the problem of voice percussion recognition. This problem is difficult because of individual differences inherent in voice spectrum characteristics and also in how the intended drum sounds are articulated. We solve this problem by utilizing phonemic sequences of onomatopoeia as internal representation. The sequences are estimated from the input power spectrum with a stochastic model, and are flexibly matched with dictionary entries representing typical drum patterns. This two-level scheme is intended to deal with the two types of individual differences mentioned above. In an experiment with 200 utterances of voice percussion, our method achieved a recognition rate of 91.5%.

1 はじめに

本研究では、ドラム音を真似た音声（ロドラム）から、ドラムパターンを検索する手法の実現を目指す。ロドラムは、ドラム音（パーカッション音）を音声で表現することを指し、ヴォイスパーカッション（Voice Percussion）や、ヒューマンビートボックス（Human Beat Box）などと呼ばれることもある。ロドラムの認識によってドラムパターン¹を得ることができれば、様々なアプリケーションが構築可能となる。例えば、頭に思い浮かべたドラムパターンをロドラムとして歌唱することで、実際のドラム音として聴くことができれば、作曲・編曲支援として有用である。また、ドラム以外の楽器を演奏しながら合奏を行うこともできる。

音声（歌唱）から音楽を検索する手法として、ハミング（鼻歌）や歌声を用いた音楽検索手法が提案されている [1, 2]。これらは主として、検索キーに音高・音長情報を用いる。したがって、楽器の音色に関する情報は考慮していない。しかし、ドラムパートは、音色の違いこそが重要な要素であるため、音高・音長情報のみの検索手法を、ドラムパターン検索にそのまま適用することは難しい。そこで、ドラム音を表現する手段としてのロドラムに着目し、その認識を行うことによってドラムパターンを検索する。

これまで、ドラム音を認識する研究はいくつか行われているが、ロドラム認識に関する研究は行われていない。ドラム音認識に関する研究は、ドラム単音を対象としたもの [3]、ドラムパートの演奏を対象としたもの [4]、複数の楽器音が混在した音響信号を対象としたもの [5, 6] がこれまでに提案されている。ドラム音認

¹ ドラムパートは、決まったパターンを繰り返して使用することが多く、このようなパターンをドラムパターンと呼ぶ。

識に共通する問題は、ドラム音の個体差に関するものであり、個体差への対処として、一つのテンプレートを対象曲中で適応的に変化させる手法が提案されている [5, 6]。しかし、ロドラム音をドラム音として認識するためには、音響的なマッチングでは対処しきれないため、従来のドラム音認識を、そのままロドラム認識に適用することは難しい。

ロドラム認識が困難なのは、歌唱者の声質の個人差に加え、ドラム音を表現する際の個人差も存在するためである。さらに、同一の歌唱者であっても、いつも同じ表現を用いるとは限らない。本手法は、擬音語によって表現の個人差に対処し、擬音語と音響信号との対応関係に確率モデルを用いることによって、声質の違いを吸収する。

以下、2.でロドラム認識について論ずる。3.で検索手法、4.で実験結果について述べ、考察する。そして最後に、5.でまとめと今後の展開について述べる。

2 ロドラム認識

これまで、ロドラム認識の研究は行われていないため、まずロドラム認識における問題点を論じ、続いて解決法を述べる。

2.1 問題点

ロドラム認識の問題は個人差によるところが大きい。本節では、ロドラム認識における問題点を、「スペクトル構造の差異」と「表現方法の差異」として論ずる。

(1) スペクトル構造の差異

全てのロドラム歌唱者が、同じドラム音に対して同じ表現を用いたとする。その場合でも、歌唱者が異なると声質が異なる。例えば、歌唱者 A の「ドン」という音声と、歌唱者 B の「ドン」という音のスペクトル構造は一致しない。

(2) 表現方法の差異

「ドラム音のイメージが人によって異なる」、「表現をドラムパターンにうまく適合するように変える」、「曖昧な発声をする」などの理由から、ロドラムは歌唱者によって異なる。よって、どのようなロドラムがどのドラム音を示すかの対応付けが難しい。

2.2 解決法

ロドラムの個人差による問題の解決法の 1 つは、多くのロドラムデータを収集し、パターン認識手法を用いることである。ドラム単音の認識においては、そのような手法が従来用いられている [3]。しかし現在、ロドラムのデータベースは整備されていないため、この

ような手法を用いることはできない。さらに、ロドラム認識ではロドラム音とドラム音の抽象的なつながりを求めるため、たとえ大量のデータがあったとしても、従来のドラム音認識手法をそのまま用いることは難しい。

ロドラムの歌唱者は各ドラム音の音響的性質の違いを、擬音語の違いとして表現することが考えられる。そこで、各ドラム音を擬音語で表現することを考え、その擬音語を中間形式として採用することで、個人差の問題に対処する。以下、前節で論じたそれぞれの問題の解決法を説明する。

(1) スペクトル構造の差異について

多数の話者の音素がどのようなスペクトル構造になるかを学習した確率モデルを用いることで、この問題を解決する。これにより、擬音語の各音素とロドラム音のスペクトル構造とを対応付けることが可能となり、ロドラム歌唱における声質の差異を吸収できる。

本研究では、擬音語音素の確率モデルとして、連続音声認識コンソーシアム (Continuous Speech Recognition Consortium: CSRC) で頒布されている、2002 年度版の不特定話者 HMM (Hidden Markov Model) を使用した [11]。これは、男女約 270 名、約 40,000 発話から学習された音響モデルであり、モデルのタイプは monophone HMM (音素コンテキスト独立型) である。

(2) 表現方法の差異について

各ドラム音がどのような擬音語で表現されるかをロドラム用の発声辞書として登録しておくことで、この問題を解決する。人によって、ロドラムの表現に個人差はあるが、各ドラム音の擬音語の種類はある程度限定することができ、事前に発声辞書を構築することで表現方法の差異に対処できる。

人が音をどのような擬音語へ変換するのか、という一般的な問題については、田中ら [7, 8]、比屋根ら [9]、石原ら [10] の研究がある。特に比屋根らは、単発音が擬音語へ変換される際の規則を示しており、発声辞書構築に利用できる。しかし、その規則は単発音に関するもので充分ではないため、データ収集を兼ねて、人がドラムパターンをどのような擬音語で表現するかを調査する実験も行い、その結果も発声辞書に含める。

以下、「ロドラムの表現」という言葉は、「どのような擬音語を用いるか」という意味で用いる。

3 ドラムパターン検索

本研究では、ポピュラー・ロック音楽のドラムパートを対象とする。ドラムパターンは、Bass Drum(BD)と Snare Drum(SD)のみで構成され、音は 2 つ同時に鳴らないものとし、4/4 拍子の 1 小節を対象とする。

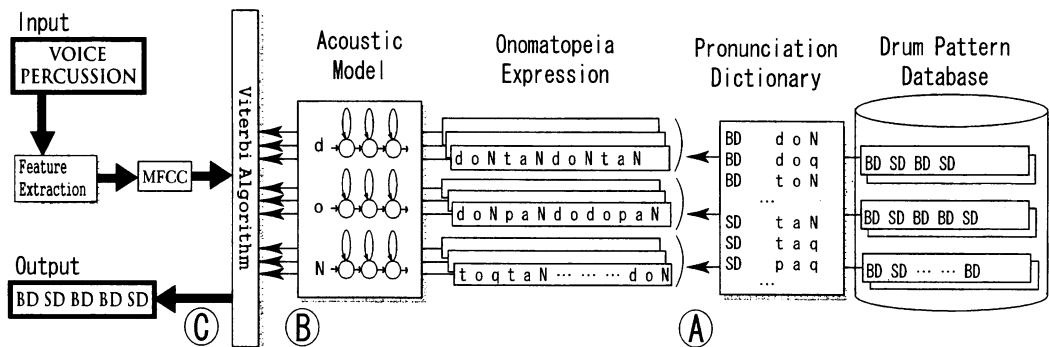


図 1: 全体の流れ

ドラムパターンデータベースは、後藤らの開発した RWC 研究用音楽データベース（ポピュラー音楽）[12] から構築する。ドラムスを含む 90 曲の SMF（Standard MIDI File）からドラムトラックの各小節を機械的に切り出し、全 536 パターンを登録した。BD と SD の発音時刻が同一であった場合は SD を優先させた。

3.1 手法

本研究では、ドラムパターンの検索と、ロドラムの認識を同時に行う。すなわち、ロドラムの個々の楽器名を決定してから検索を行うのではなく、ロドラムがどのドラムパターンに近いかを直接求める。

本手法の流れを説明する。まず、構築したドラムパターンデータベース中の各ドラムパターンを、発声辞書を用いて擬音語表現へ展開する（図 1: A）。この際、各ドラム音は複数の擬音語候補を持つため、展開される擬音語表現も複数となる。また、入力されたロドラム音に対し、音声認識で広く用いられている音響特徴量 MFCC（Mel Frequency Cepstral Coefficients）を求め、全ドラムパターンの各擬音語表現に対応する音響モデル（HMM）との尤度を Viterbi Algorithm により計算する（図 1: B）。こうして求めた全ドラムパターンの擬音語表現の尤度の中で、最も尤度の高いドラムパターンを検索結果として出力する。（図 1: C）。このように本手法では、入力のロドラム音と音響的に最も近いドラムパターンと擬音語表現を、同時に推定することができる。

3.2 発声辞書の構築

発声辞書を構築するためには、ドラム音がどのような擬音語として発声されるのかを知る必要がある。比屋根らは、短時間に減衰する単発音に対して、中心周波数や残響時間、周波数ゆらぎなどのパラメータによ

って 6 種類の擬音語の分類を報告している [9]。そこで提案された分類においては「衝突減衰音」と「多重衝突音」が BD と SD の音に当てはまる。衝突減衰音は 1 回の衝突がごく短時間に減衰する音を示し、多重衝突音は同系の音が二回以上連続して発音される音を示す。

BD の擬音語には、衝突減衰音として「ト」「トゥ」「トン」「ド」「ドゥ」「ドン」「ズ」「ズッ」「ズン」を用いた。SD の擬音語には、衝突減衰音として「タ」「タッ」「タン」「パ」「パッ」「パン」を用いた。また、多重衝突音も考慮し、BD に「コ」（トコトコ、ドコドコのため）、SD に「カ」（タカタカのため）も含めた。

このように構築した辞書を基本的な発声辞書とする。また、ロドラム歌唱の実験を行い（4.1 で後述）、そこで用いられた次の擬音語も発声辞書に追加登録した。BD の擬音語として「ドゥ」「ドゥッ」「ドゥン」「トゥ」「トゥッ」「トゥン」を追加し、SD の擬音語として「ダ」「ダッ」「ダン」「チャ」「チャッ」「ラ」を追加した。

3.3 確率モデルのロドラムへの適応

ロドラム音は、通常の会話では使わない特殊な音声であるため、音声認識で用いられる日本語音素の音響モデルをそのまま用いると、誤認識する可能性が大きい。そこで、音響モデルをロドラム音へ適応させることを考える。これには、音声認識において一般的に用いられる話者適応の手法をそのまま適用することができる。そこで本実験では“ロドラム音への適応”と“ロドラム歌唱者への適応”の 2 種類の適応を行う。ロドラム音への適応では、評価データに含まれないロドラム音を適応データとして用い、ロドラム歌唱者への適応には、適応データにその歌唱者自身のロドラム音を用いる。前者は汎用的な状況、後者は歌唱者を限定できる状況を想定している。

本研究では、音響モデルの適応に MLLR（Maximum Likelihood Linear Regression）法と最大事後確率

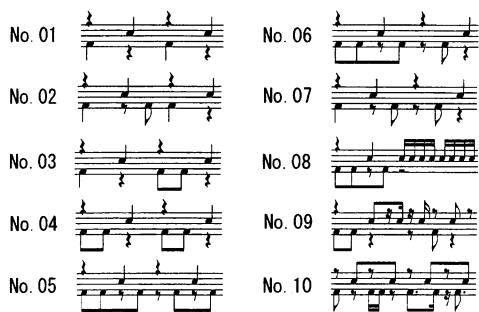


図 2: 呈示ドラムパターン

(Maximum *A Posteriori* Probability: MAP) 推定法を組み合わせた MLLR-MAP 法を用いた [13].

4 実験および考察

本研究では、ロドラム表現実験とロドラム認識実験を実施した。前者では、ドラムパターンの擬音語表現について調査し、後者では、その結果に基づいて実装した本手法の有効性を認識性能によって評価した。

4.1 ロドラム表現実験

本実験では、19 歳から 31 歳の男女 17 名の被験者（ドラム・パーカッション経験者 2 名、非経験者 15 名）によるロドラムの歌唱を収録して分析した。各被験者は、BD と SD のみで構成されるドラムパターンを聴取し、それに対応する擬音語をイメージして歌唱する。BD と SD の音源には、後藤らの開発した RWC 研究用音楽データベース（楽器音）[12] に収録されている RWC-MDB-I-2001 No.42「ロックドラムス 1」の強さ「強」のデータを用いた（421BD1N3.WAV, 421SD3N3.WAV）。

4.1.1 実験方法

各被験者は、実際のドラム音で演奏されたドラムパターンの再生音を、その楽譜も見ながら、記憶するまで何度も聴取する。その後、再生音を停止し、被験者は自分で擬音語表現を考えて歌唱する。その際、ロドラムの例は示さず、再生音も実際のドラム音のみとする。

実験に用いたドラムパターンは、ドラムの練習教本を参考に、基本的な 7 パターン（図 2: No.01-07）、SD の多重衝突音を含む 1 パターン（図 2: No.08）、複雑な構成を持つ 2 パターン（図 2: No.09-10）、の計 10 パターンを用意した。各パターンに対して、80M.M. と 120M.M. の 2 種類のテンポで演奏した再生音を用意し、被験者 1 人あたり 20 パターンのロドラムの歌唱を収録した。その収録順は、被験者ごとにランダムとした。

表 1: 擬音語の割合（CV のみ）

BD						
ド	ト	ドゥ	トゥ	ズ	レ	
9	5	8	4	3	1	
SD						
タ	ダ	バ	デ	チャ	カ	ラ
13	3	3	1	1	2	1

4.1.2 結果

被験者には、聴こえたパターンを擬音語に直して発声するように指示したが、BD と SD の違いを、同一擬音語の音高の違いのみを用いて表現する被験者も見られた。1 人が全パターンを音高の違いで表現し、3 人が部分的に音高の違いで BD と SD を区別した。呈示ドラムパターンが速い場合や複雑である場合に、このような傾向が見られた。

擬音語の違いでドラム音を表現した場合、その表現の仕方は、モーラ表記では、「CV」、「CV Q」、「CV N」、「CV R N」の 4 つの構造のいずれかの形をとった。ここで、CV は「子音+母音」、Q は促音、R は長音、N は撥音を表す。例えば、SD の CV を「タ」とすると、CV Q が「タッ」、CV N が「タン」、CV R N が「ターン」となることを示す。CV の表現の違いを表 1 に示す（数字はその擬音語を発声した人数²）。

また、以下のような傾向が見られた。

- 次の音までの間隔が広くなるごとに、「CV」、「CV N」、「CV R N」となる。
- BD の CV を「ド」と発声する被験者でも、BD の速い連続音では「ドド」ではなく「ドト」と発声する。
- 最初は音高の違いで表現していた被験者も、途中で擬音語の違いで表現し始めると、それ以後は擬音語の違いによる表現のみとなる。

さらに、ドラム・パーカッション経験者の 2 名のみ、パターン No.08 の SD の多重連続音を「タカタカ」と発声した（それ以外の被験者は「タタタタ」と発声した）。また、この 2 人は休符も発声した。例えば、No.09 のドラムパターン「BD BD SD SD SD BD SD」に対して「ズンズンタン ズ タ ズ タズタン」と、BD と同じ擬音語を用いたり「ドンドンタ ン タ ン タドンタン」のように「ン」を用いた発声を行った。

4.1.3 考察

(1) BD と SD を音高の違いで表現

本手法では、音高の違いで表現することは例外的だと考え対処しない。こうした表現を被験者がした理由

² ドラムパターンによって表現方法が変わるため、一人の歌唱者が複数の CV を使用することもある。

には、以下の二つが考えられる。

- 複雑なパターンや速いパターンを呈示された場合パターンを追うことで精一杯になり、言葉を使い分ける余裕がなかった。
- 始めだけ音高の違いで表現していた被験者各ドラム音をどのような擬音語表現で歌うとよいか、始めはうまく考えられていなかった。

(2) 多重衝突音の擬音語の使用、休符の発声

本手法では多重衝突音の擬音語には対処するが、休符の発声には対処しない。これらの現象は、ドラムやパーカッションの練習方法に起因すると考える。例えば、ドラム教則本には、リズムの取り方を理解させるために、「タカタカ」や「ンタンタ」などの記載がある場合が多い。打楽器以外の楽器の演奏者でも、休符を読んだり、多重衝突音の擬音語を発声することでリズムを取りやすくした経験を持っている可能性がある。将来は、そうした特殊な休符の発声にも対処するとよい。

4.2 ロドラム認識実験

ロドラム表現実験で収録したロドラム音を利用して、本手法が正しいドラムパターンを検索できるかを評価する。ロドラム表現実験で用いた 10 種類のドラムパターンは、3 節で述べた 536 種類のドラムパターンデータベースとは無関係に決定したため、図 2 の No.08 と No.10 がデータベースに含まれていない。そこで、この 2 パターンもデータベースに加え、全 538 種類のドラムパターンに対して検索実験をする。

本研究では、以下の 4 種類の実験を行う。

- 既存の音響モデルを用いる。
- 評価データに含まれないロドラム音で適応した音響モデルを用いる。
- 各被験者ごとに話者適応を行った音響モデルを用いる。
- 各被験者ごとに話者適応を行った音響モデルを用い、さらに辞書も各被験者のドラムパターンごとに再構築する。

本実験では、全パターンを音高の違いで表現した被験者 1 名と、ほぼ半分のパターンを音高の違いで表現した 1 名を除いた 15 名分の収録音声を用いる。10 名が発声した計 200 発話を評価データとし、それ以外の 5 名が発声した計 100 発話を (B) の適応データとする。(C)(D) の話者適応は、認識率を leave-one-out cross validation 法で評価する。具体的には、各被験者の発声 20 パターンを適応用 19 パターンと評価用 1 パターンに分け、評価用を順に変えながら 20 回実験した平均認識率で評価する。さらに (D) では、各ロドラム音声に用いられ

表 2: パターン評価の結果 (全体の認識率)

実験 (A)	実験 (B)
53.5%(107/200)	67.5%(135/200)
実験 (C)	実験 (D)
87%(174/200)	91.5%(183/200)

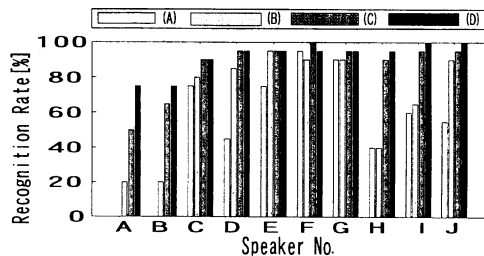


図 3: パターン評価の結果 (被験者ごと認識率)

ている擬音語だけを登録した辞書を構築して評価する。これは、辞書の適応と考えることができ、歌唱者が事前にどのような擬音語を用いるかを定める条件に相当する。

4.2.1 評価方法

ドラムパターンが正しく検索されたことを評価するには、楽器名と発音時刻がそれぞれ正しく求めたかを調べればよい。そこで、以下の二点を評価する。

(1) パターン評価

(A)(B)(C)(D) の各実験ごとに、正解のドラムパターン (楽器名の並び) が、尤度 1 位として正しく推定できた割合を認識率とする。

(2) 発音時刻評価

各ロドラムの音声に対して正解時刻のラベル付けを人手で行い、その時刻とビタビアルゴリズムによるセグメンテーション結果の時刻との「ずれ」を評価する。本評価では、実験 (D) の結果において、パターン評価で正しく推定できたパターンのみを用いる。

4.2.2 結果

(1) パターン評価

表 2 に全体の認識率、図 3,4 に、被験者ごと、パターンごとの認識率のグラフを示す。縦軸が認識率を表し、横軸は、図 3 では被験者番号、図 4 ではドラムパターン番号を示す。図 3 において、B がドラム経験者、D がパーカッション経験者である。

(2) 発音時刻評価

図 5 にずれの頻度分布を示す。横軸は、認識された発音時刻から正解発音時刻を引いた値である。ずれの

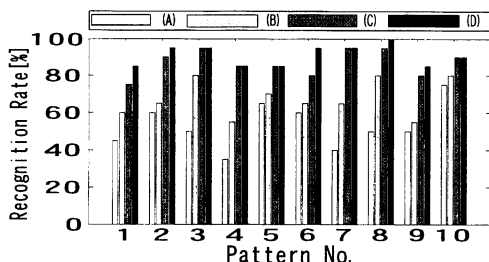


図 4: パターン評価の結果 (パターンごと認識率)

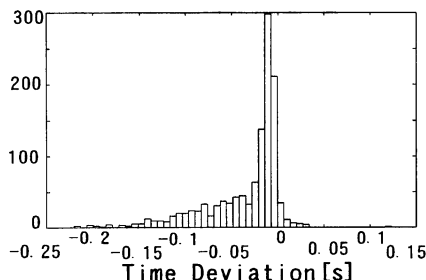


図 5: 発音時刻評価の結果 (実験 (D)) でのずれの頻度分布

平均は -0.0346 秒, 標準偏差は 0.0409 秒, 最も大きなずれ幅は -0.2200 秒であった。

4.2.3 考察

(1) パターン評価について

本実験で得られた結果は, 適応が可能な運用条件で応用システムを構築することで, 本手法が高い有効性を持つことを示している。不特定多数の歌唱者に対して運用する場合, 実験 (B) のようにロドラム音声を用いて既存の音響モデルに適応処理を行うことにより, 7 割弱の認識率を得ることができる。この場合でも, 無音声を使用しないなど, 発声スタイルを制限するとさらに良い認識率を得られる可能性がある。そして, 歌唱者を事前に登録する運用をすれば, 実験 (C) で示されたように, 19 パターン程度の少数のドラムパターンを発声するだけで, およそ 20% の認識率向上が期待できる。さらに, 歌唱者が自分のロドラムで用いる擬音語表現も事前に登録する運用をすれば, 認識率は 90% を超えて実用性はさらに高くなる。

ドラムパターン毎のパターン評価の結果では, ドラムパターンの難易度と認識率は, あまり関係していない。これは, 最も簡単なパターン No.01 の認識率よりも, 最も難しいパターン No.10 の認識率の方が高いことからわかる。さらに, 歌唱者への適応で全てのパターンの認識率が大きく向上していることから, ロドラムの認識では歌唱者の発声スタイルへの依存が大きく

いことがわかる。

(2) 発音時刻評価について

平均的なずれは -0.0346 秒であり, 120M.M. の 32 分音符のレベルで誤認識の可能性がある。検出時刻のずれは「前の音との間にある無音をうまく検出できていない場合」, 「2 つの音の間が途切れていない場合」, 「息などが原因の雑音」が理由であった。

5 おわりに

本稿では, ロドラム認識における問題点と解決法を示し, 実際にロドラムの歌唱でドラムパターンを検索できる手法を提案した。これは, 音声入力による音楽検索において, 従来の音高・音長情報以外の情報活用への新たな可能性を示すものである。

今後は, BD と SD 以外のドラムスの対応, 2 音以上が同時に鳴る場合の表現方法の調査, ロドラム固有の表現への対応等も検討していく予定である。

参考文献

- [1] 藤山哲也, 高島洋典, “ハミング歌唱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, pp.1543-1551, 1994.
- [2] 園田智也, 後藤真孝, 村岡洋一, “WWW 上での歌声による曲検索システム,” 電子情報通信学会論文誌, Vol.J82-D-II, No.4, pp.721-731, 1999.
- [3] P. Herrera, A. Yeterian, F. Gouyon, “Automatic Classification of Drum Sounds: A Comparison of Feature Selection Methods and Classification Techniques,” ICMAI, LNAI2445, pp.69-80, 2002.
- [4] 後藤真孝, 村岡洋一, “打楽器音を対象にした音源分離システム,” 電子情報通信学会論文誌, Vol.J77-D-II, No.5, pp.901-911, 1994.
- [5] A. Zils, F. Pachet, D. Olivier, F. Gouyon, “Automatic Extraction of Drum Tracks from Polyphonic Music Signals,” WEDELMUSIC, pp.179-183, 2002.
- [6] 吉井和佳, 後藤真孝, 奥乃博, “テンプレート適応を利用した実世界の音楽音響信号に対するドラムスの音源同定,” 情報処理学会研究報告 2003-MUS-53-12, Vol.2003, No.127, pp.55-60, 2003.
- [7] 田中基八郎, 松原謙一郎, 佐藤太一, “異音の表現における擬音語の検討: 衝突音等の単発音やうなり音の場合,” 日本機械学会論文集 C 編, Vol.61, No.592, pp.4730-4735, 1995.
- [8] 田中基八郎, 松原謙一郎, 佐藤太一, “機械の異常音の擬音語表現,” 日本音響学会誌, Vol.53, No.6, pp.472-482, 1997.
- [9] 比屋根一雄, 澤部直太, 飯尾淳, “単発音のスペクトル構造とその擬音語表現に関する検討,” 電子情報通信学会研究報告 SP97-125, Vol.97, No.586, pp.65-72, 1998.
- [10] 石原一志, 坪田康, 奥乃博, “日本語の音節構造に着目した環境音の擬音語への変換,” 電子情報通信学会研究報告 SP2003-38, Vol.103, No.154, pp.19-24, 2003.
- [11] 河原達也, 住吉貴志, 李晃伸, 坂野秀樹, 武田一哉, 三村正人, 伊藤克亘, 伊藤彰則, 鹿野清宏, “連続音声認識コンソーシアム 2002 年度版ソフトウェアの概要,” 情報処理学会研究報告 SLP-48-1, Vol.2003, No.48, pp.1-6, 2003.
- [12] 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,” 情報処理学会論文誌, Vol.45, No.3, pp.728-738, 2004.
- [13] E. Thelen, X. Aubert, P. Beyerlein, “Speaker adaptation in the philips system for large vocabulary continuous speech recognition,” ICASSP'97, Vol.2, pp.1035-1038, 1997.