身の回りのものに任意の音色を割り当てて演奏可能な電子楽器インターフェース 〜 Possessing Drums 〜

山　本　和　彦†

本研究では, 身の回りにあるもの, や声を発する, といった行為に対して任意の音色を割り当てて, あたかも割り当てられた音色を発する物体がそこにあるかのように音を鳴らして楽器のように演奏することを可能にするインターフェースを提案する. 本システムは特別なデバイスを必要とすることなくマイク一本のみで実現可能である. 身の回りのものに割り当てられた音色は, マイク音声から検出されたアタックによって PCM 波形がトリガー再生されたものではなく, マイクの入力音声そのものを加工してその音色が発せられる仕組み自体がモデリングされて生成されるため, どこまでも小さい音や速い連打, 奏法による変化など生楽器と同等の繊細な表現をすることを可能としている. また, 複数の割り当て対象物に同時に異なる音色をそれぞれ割り当てて複数音同時演奏をすることも可能である.

An Interface of Musical Instrument that Assigns Arbitrary Timbres to Personal Belongings 〜 Possessing Drums 〜

Kazuhiko Yamamoto†

In this work, I present an interface of musical instrument that assgin arbitrary timbres to arbitrary materials including personal belongings, or actions such as vocalization, to be enables us to play music as if there are the actual objects which genarate its timbres in front of us. This system requires only a microphone. It is not a special device. The assigned timbres are produced by not triggered pcm waveform in response to the detected attacks in mic input source but the modeling process of the system that generate the timbres with modifying the mic input source itself. So it enables us to play with very sensitive expression, such as very small sound, fast passages, and the effects of play style, etc... Additionally, in this system, we can assign multiple timbres to multiple objects respectively and play polyphonic music.

1. はじめに

楽器を演奏するという行為は多くの人々にとって非常に魅力的である. しかし, 実際の楽器は設置や運搬が容易ではなかったり, メンテナンスが面倒であったりといった理由で我々が演奏に至るまでにはハードルが高い. 一方, 楽器の演奏能力の有無に関わらず我々はしばしば楽器の代わりに身の回りの物体を手で叩いて音を出してリズムを取ったり, 楽器音の真似をして口ずさむといった行為を行うことがある. そういった意味では身の回りのものは我々にとって最も身近な楽器であるとも言える. 実際, 我々が鳴らしているもの音はその鳴らし方に応じてダイナミックに音が変化しており, 我々の全ての動作が意味のある入力となって反映されているという点で, 生楽器と同等の表現力を持ち得る. しかしその音は実際の楽器と比較して音色としては決して魅力的とは言い難い.

こうした身の回りのものや行為を楽器として見立てる行為をコンピュータで拡張し, 任意の音色を実際に本当の楽器と同じように演奏することが可能になれば, 手軽に身の回りのものを楽器として音楽に利用することができるようになる他, 楽器に触れるまでの敷居が下がり, 音楽教育への利用においても価値がある. また, コンピュータ上で設計した楽器等の音を身の回りのものを入力装置として利用してユーザが鳴らすことで直感的なフィードバックが可能になるなどの応用も期待できる.

そこで, 本研究では, ある音色が発せられるプロセスによって生じる特性そのもの (割り当て音色) を我々の身の回りにある任意の他の物体や音を鳴らすという行為 (割り当て対象) へ割り当て, 実際に鳴らしている

† ヤマハ株式会社 (個人)
YAMAHA(Individual Works)

割り当て対象の音自体のダイナミックな変化を活かしたままあたかもそこに割り当て音色を発する物体や楽器があるかのように発音させることを可能にするインターフェース”Possessing Drums”を提案する．これにより，例えば，ピアノの音色 (割り当て音色) をテーブルを叩いて音を鳴らすという行為 (割り当て対象) に割り当てると、あたかもテーブルがピアノになったかのように叩いて鳴らすことができる．さらに，本システムでは特別なデバイスを使うことなくマイク一本のみで容易に実現でき，かつ同時に複数の対象への割り当てを可能する．

2. 関 連 研 究

従来の電子楽器をリアルタイムに演奏するためのインターフェースをセンシング手法の観点から分類すると，2 種類のアプローチに大別される．まず一つ目として，何らかのセンサーの出力値の変化量をトリガーとして発音を開始するものがある．電子キーボードを始めとして，発音するためのプロコトルに MIDI を採用するような従来の電子楽器や，過去の楽器のための研究で提案されたインターフェースの大部分がこれに属する．ある変化量をトリガーと定義するためには閾値の定義が必要となる．閾値があるということは当然閾値を超えない入力には反応することが不可能になり，必然的に連続してセンシング可能なトリガーの時間間隔の小ささにも限界ができる．つまり，微妙な細かい演奏者の動作をセンシングすることには限界があるということである．このことは演奏している楽器に直接触れている，という感覚を著しく損ねるものであり，また，本研究の目的としているように実際に鳴らしている身の回りのものの音自体のダイナミックな変化をこの手法では活かすことができない．二つ目のアプローチとしてはマイクからの入力音やセンサーからの生のデータをそのまま加工する，というものがある。これは世に出回っている製品では KORG WaveDrum[4] に代表されるアプローチであり，これによってどんなに細かく小さい音も意味のある入力として捉えることができるようになり，生楽器と同様の表現をすることが可能となっている．広い意味では Theremin[6] も静電容量という連続量をそのまま音に変えているという点で後者に分類することができる．これらのように生のデータをそのまま利用する電子楽器の特徴は，演奏者の表現を漏らすことなく全て出音に反映することであり，これによって実際に楽器を演奏している，という感覚を強めることができる．従って，本研究では後者のアプローチを採用する．

図 1 システム構成
Fig. 1 The system.

一方，我々の身の回りにある物を楽器に変える研究としては過去にもいくつかのアプローチが存在する．The Sound of Touch[8] ではブラシ状のピエゾピックアップを内蔵したデバイスを使ってユーザがそのデバイスで任意の物体を擦ったり叩いたりしたときの振動をセンシングし，その振動に登録してあった任意の音色のインパルス応答を畳み込むことによって，身の回りのものを利用して任意の音色を鳴らすことができる．この研究ではデバイスそのものの振動の変化を活かしたまま音色を変えて演奏することを可能にしているが，身の回りのものに音色を割り当てたわけではなくブラシ状のピックアップそのものが既に単体の楽器であるため，ブラシ状のデバイスで不可能な任意の奏法は実現できず，身の回りのものそのものを特別なデバイスなしに楽器にするという本研究の目的とは異なる．また，KORG WaveDrum Mini[5] にはクリップ型のピエゾマイクが付いており，前述の WaveDrum の入力としてクリップで挟んだ身の回りの物の振動を利用することができる．この方法では割り当て対象物の振動を確実に入力することができるという利点があるが，クリップで挟むことが出来ないもの，例えば音声等には適用が不可能である他，クリップで対象物を挟む事によることによる振動への影響が少なからずあるという問題がある．TableDrum[7] ではあらかじめ登録した机を叩く音など身の回りの音の音量をトリガーとして PCM 波形を発音させる．しかし，センシング方式が前述のトリガー方式であるため，電気的なデバイスではなく実際のものを鳴らすことを入力方法とする，ということの利点を活かしきれてはいない．また，同時に認識可能なトリガー音は一種類であるため、複数音の同時発音はできないという問題もある．

3. アプローチ

図 1 が提案手法の概略である．マイクからの音声を

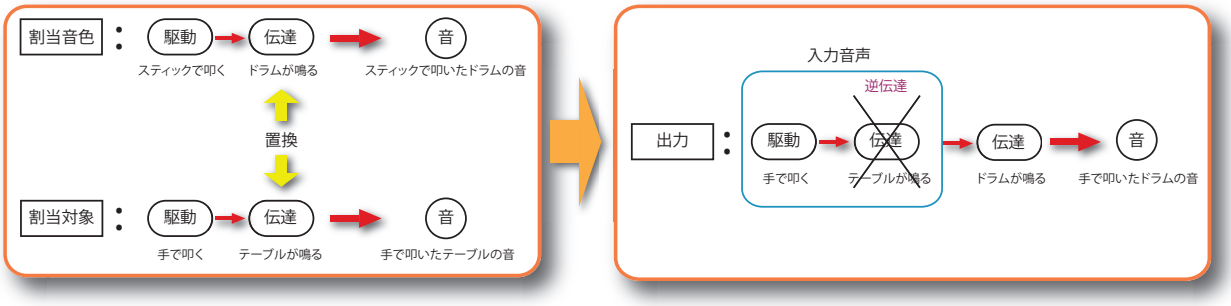

図 2　本システムでのアプローチ
Fig. 2　The approaches of my system

入力とし, それに対応した割り当て音色の音をスピーカから出力する. 本研究では, 割り当て音色の音を合成する手法として, その音色が生成される仕組みそのものを割り当て対象の音の仕組みに付加する, というアプローチをとる. どんな音であってもそれが発せられるプロセスは駆動, と伝達という 2 つの部分に分けることができる (図 2 左). 例えばテーブルを叩く音ならば, 叩くというのが駆動, テーブル自体がそれによって振動し, 音となって空気を伝わり我々の耳に届く, という部分が伝達であり, 音声ならば, 声帯の振動が駆動, それが声道を通って口から放射され, 空気中を伝わって我々の耳に届く, という部分が伝達である. Possessing Drums ではこのうち伝達の部分のみを他の音のものと入れ替えることで音生成システムの他の音への割り当てを行う (図 2 右). つまりスティックでドラムを叩いて音を出す (割り当て音色) を手でテーブルを叩いて音を出すという行為 (割り当て対象) に割り当てると, 手で叩くという駆動部分とドラムの鳴りという伝達部分が組み合わされることになり, あたかも手でドラムを叩いているかのような体験が得られる. 置き換わったのは伝達の部分だけであるので, テーブルを擦ればドラムを擦るのと同等となり. またテーブルのかわりに息を割り当て対象とすればドラムを息で鳴らす, といった現実にはありえない駆動の仕方をすることも可能である. 当然ながら, 駆動の部分はユーザが入力したものそのものであるので, どんなに繊細な演奏の変化も意味のある入力として捉えることができる.

また, 本研究では, 単独の楽器だけではなく, 複数の楽器を異なる音対象に同時に割り当てることを実現するため, 入力音の音源分離を行い, 複数の割り当て対象をそれぞれ同時に認識可能とする.

4. ハードウェア

本システムで必要なハードウェアは計算処理を行う PC の他は, 一本のマイクとスピーカのみである. ほぼ全ての仕組みがソフトウェア上で実現される. 特別なデバイスを用いることなくセンサーとしてはマイク一本のみが必要なだけであるため, 容易に実現可能である.

5. ソフトウェア

本システムには割り当て音色と割り当て対象を登録するための登録モードと実際に音を割り当てて演奏するための演奏モードが存在する. さらに演奏モードは, ユーザの入力した音声を割り当て対象の音 (複数同時も有り得る) と余計な雑音を個別に分離して次段へ入力する認識部, 分離された割り当て対象の音を擬似逆フィルターに通して駆動部分のみの信号を取り出す擬似逆フィルター部, 取り出された駆動音を入力として割り当て音色の伝達部分を付加する合成部の 3 つのブロックから成り立つ. 以下それぞれについて説明する.

5.1 登録モード

本システムではユーザはまず, 割り当て音色の音と割り当て対象の音を別個に録音し, その情報を登録する. 音を発する物体の変形が線形であることを仮定すると, その振動 $y(t)$ は指数関数的に減衰する正弦波 (モード) の線形和で表現できる.

$$y(t) = \sum_{n}^{N} A_n e^{-d_n t} sin(2\pi f_n t) \qquad (1)$$

ここで, A_n, d_n, f_n はそれぞれ n 番目のモードの振幅, 減衰係数, 周波数である. 本システムではユーザの録音した音声からこれらのモードのパラメータを推定して登録する. まず, 録音された音は解析部に回され, オンセット検出をすることによって単音の音素片に分解された後, その特徴量 (モーダルパラメータ) となるモード周波数, カップリング音量, ダンピング係数が

計算される. 理想的なインパルス応答ではない減衰音からモーダルパラメータを推定する方法にはいくつかの手法があるが, ここでは *Sirdey* らによる ESPRIT 法とガボール変換を組み合わせた手法[1] を用いた. また, 音が鳴っている間の平均パワースペクトルも保存される. この平均パワースペクトルは, 後述の認識部において入力音の音源分離を行うための教師データとして利用される. 波形データそのものに関しては保持する必要はない.

5.2 演奏モード

演奏モードでは, 実際にマイクに入力されたユーザが演奏している音を認識し, それに応じた出力音を合成する.

5.2.1 認 識 部

マイクから入ってくる音は登録された割り当て対象の音と, 全く関係が無いノイズ, そしてスピーカーから出て回りこんでくる本システムによって発せられる音のフィードバック音が混ざり合った混合音である. さらに, ユーザによって登録された割り当て対象の音は一種類とは限らず, 複数の割り当て対象の音が同時に演奏されることもある. これらの音の分離を行った後にそれぞれ後段へ受け渡す必要がある. また, マイクからの入力音をそのまま加工して出力としているためスピーカからのフィードバックが入力として認識されるとハウリングを起こしてしまうため, フィードバック音の分離は必須である. 入力された混合音のうち割り当て対象のの音とシステム自体の出力音に関してはその平均パワースペクトルは既知である. これを利用して、本研究ではβ-Divergence 規準 NMF(Non-Negative Matrix Factorization)[3] による音源分離を一部教師ありでリアルタイムで適用する. NMF ではスペクトログラム行列 $\mathbf{V}$ にスパース性を過程して,

$$\mathbf{V} \approx \mathbf{WH} \tag{2}$$

という辞書行列 $\mathbf{W}$ と励起行列 $\mathbf{H}$ の積になるように分解を行う. 辞書行列は $\mathbf{V}$ の中に含まれる音色それぞれの平均パワースペクトルを表現し, 励起行列はそれぞれの音色がある時間にどれだけ鳴っているかという情報を含む行列である. β-Divergence 規準 NMF では $\mathbf{V}$ と $\mathbf{WH}$ とのβ-divergence $D_\beta(\mathbf{V}|\mathbf{WH})$ を最小化するような $\mathbf{W}$ と $\mathbf{H}$ を求めることでこの分解を行い, 実際には,

$$\mathbf{H} \leftarrow \mathbf{H} \cdot \frac{\mathbf{W^T}((\mathbf{WH})^{\beta-2} \cdot \mathbf{V})}{\mathbf{W^T}(\mathbf{WH})^{\beta-1}} \tag{3}$$

$$\mathbf{W} \leftarrow \mathbf{W} \cdot \frac{((\mathbf{WH})^{\beta-2} \cdot \mathbf{V})\mathbf{H^T}}{(\mathbf{WH})^{\beta-1}\mathbf{H^T}} \tag{4}$$

を反復することによって辞書行列と励起行列を更新する. ここで, $\mathbf{W}$ のうち, 教師データのある音色の列については既知であるため更新を行う必要はない. ただしノイズ成分のみは教師なしとし, 対応する $\mathbf{W}$ の列の更新を行った. また, 通常 NMF による音源分離では窓長に比べ時間的にはかなり長いスペクトログラム行列に対して行われるためリアルタイムでの処理は不可能である. 本研究では, この分解をリアルタイムに行うため, 最新のフレーム (ここでは窓長 256 サンプルの片側パワースペクトル) のみにおいて分解を行う. つまり $\mathbf{H}$ はあらかじめ与えられた音色数の大きさの列ベクトルとなる. また, 入力音は時間的に連続的であるので, 前回フレームの結果を現在のフレームの初期値とすることで収束までの反復回数を減らすことができる.

5.2.2 疑似逆フィルター部

マイクへ入力された割り当て対象の音は既に駆動部分と伝達部分が含まれているため, ここからできる限り駆動部分の成分だけを取り出したい. このために本システムでは入力音にあらかじめ登録しておいたモーダルパラメータを利用して擬似逆フィルタ処理を行う. 計算済みのモーダルデータから, その逆フィルタはそのモードの数だけの BiQuad ノッチフィルターの直列接続で近似することができる. Eq.5 が周波数 f_z にディップを持つ 2 次の BiQuad フィルタの伝達関数である.

$$\frac{Y}{X} = \frac{G(1 - 2r_z cos(2\pi f_z T)Z^{-1} + r_z^2 Z^{-2})}{1 - 2r_p cos(2\pi f_p T)Z^{-1} + r_p^2 Z^{-2}} \tag{5}$$

ここで, G はゲイン, r_z はゼロ点での減衰係数, $r_p f_p$ は極を与える周波数と減衰係数である.

5.2.3 合 成 部

擬似逆フィルター処理をされた入力信号は割り当て音色を再現するようなモードの数だけ並列接続されたレ

ゾネータに入力される．ここでは，*Kees van den Doel* らによって提案されたモーダルレゾネータ[2] を採用する．n 番目のモード角周波数 ω_n を持つレゾネータの出力を $y_n(m) = u(m) + iv(m)$ と置くと，

$$
\begin{aligned}
u(m) &= c_r u(m-1) - c_i v(m-1) + a_n F(m), \\
v(m) &= c_i u(m-1) + c_r v(m-1) \\
c_r &= e^{-d_n/S_R} cos(\omega_n/S_R), \\
c_i &= e^{-d_n/S_R} sin(\omega_n/S_R),
\end{aligned} \tag{6}
$$

という漸化式で表すことができる．ここで，a_n, d_n, S_R, $F(m)$ はそれぞれ n 番目のモードの振幅，減衰係数，サンプリング周波数，駆動力である．$y_n(m)$ は複素振幅であるので，実際の音として出力するのは $Im(y_n(m))$ となる．

6. 検　　証

提案手法の妥当性を確かめるためのいくつかの検証を行った．まず，入力音が登録された割り当て対象の音ごとに正確に分離されて認識されていることを確かめる必要がある．ここでの検証はビブラフォンマレット (A)，ドラムブラシ (B)，指 (C) でそれぞれテーブルを叩いた音を割り当て対象として，それぞれピアノの C2, A4, E3 の音色を割り当てて行った．実験は 3 種類の割り当て対象音を単音，または複数同時にリアルタイムに入力し，そのときのそれぞれの割り当て音色の出力を調べた．入力した順番は，A, B, C, B+C, A+B+C, A+C, B, C, A, B, C, A, B, A+B+C である．このときの入力音と出力音の関係を図 3 に示す．図 3(縦軸:振幅，横軸:時間) では最上段にマイクからの入力音の波形を，下三段に出力された C2, A4, E3 のピアノの波形を上からそれぞれ示しており，破線で繋がれてている楕円が入出力で対応しているアタックである．A4 の音に多少の音量のうねりがみられるのは割り当てた音色による原因である．結果として，入力が単音または複数の対象音を同時に入力したときも想定通りの音色にアタックが対応しており，正確に分離して発音されていることが確認された．これにより，入力音を認識する仕組みは妥当に動作していると言える．本研究で採用した手法では，入力音は登録された音のうち，もっとも近い音としてクラスタリングされるため，奏法などによる対象音の違いを複数登録しておけば，奏法によって割り当て音色を全く別のものに切り替えることが可能であり，複数登録していなければ単独の割り当て音色の異なる奏法として認識されるようになる．つまり，テーブルを叩く音が一つしか登録されていなければ，何で叩くかということは奏法の違いとして認識され，複数種類の叩くものの素材による変化が登録されていた場合は異なる割り当て対象として認識されることが分かった．このことは割り当て対象として奏法の違いを含めるか別の対象にするかということをユーザが選択して表現できる幅を自由にコントロールできるようになるという利点がある．

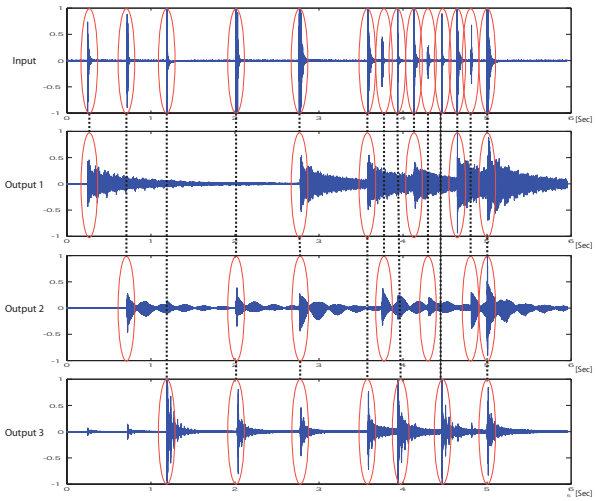

図 3　入力波形 (最上段) と以下三段それぞれの割り当て音色の出力波形．破線で繋がれたアタックが対応している．
Fig. 3　The input waveform(top) and the output waveforms of the assigned tone(below).

次に，本システムの出力波形を示す．まず，実際のピアノの A4 単音のスペクトログラムを図 4 上段に示す．このピアノ音はピアノの弦，共板をフェルトハンマーで駆動したものと考えることができる．このピアノ音を割り当て音色としてテーブルをピアノのフェルトハンマーで叩いた音を割り当て対象としたときの出力のスペクトログラムが図 4 下段である．これを実際のピアノ音との比較すると，システムの出力音では再現できるモード数に限界があるため，高域の倍音成分が少なめになっており，また比較的各モードがはっきり現れてはいるが，全体的にスペクトログラム上ではかなり近い結果が得られているのが分かる．これにより，本手法に妥当性があることが分かる．

最後に，複数の被験者に実際に本システムで演奏してもらい，その様子を観察した．被験者にはテーブルや空き缶，マイクに息を吹き掛ける，など身の回りのものや行為に対して音色を割り当てて演奏してもらった (図 5)．生楽器と同様にテーブルの叩き方や何で叩くか，といった奏法による変化やどこまでも小さい音，非常に速いフレーズまでも出力に反映されることで，非常に微妙な表現の差異を演奏に取り込むことができ，楽器としての可能性が十分あることが検証でき

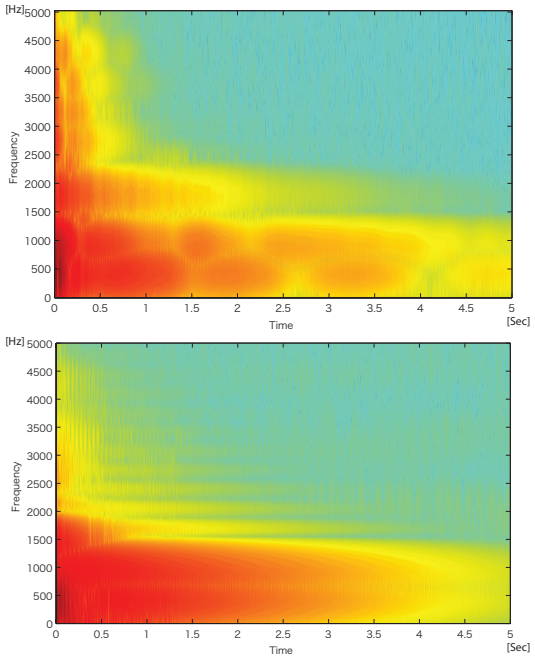

図 4 上図:実際のピアノの A4 のスペクトログラム
下図:ピアノのハンマーでテーブルを叩いた音にピアノの A4 の音を割り当てたときの出力のスペクトログラム.
Fig. 4 Above:The spectrogram of actual piano A4
Below:The spectrogram of the output.

図 5 演奏風景. 机を叩いて楽器音を鳴らしている.
Fig. 5 The play scene.

た. ここで, オーディオのバッファサイズは 256 サンプルで動作させ, このとき発音までのレイテンシーは約 6.0[ms] 程であり, リアルタイム演奏には全く支障が無かった. しかしながら, 今回の実験では認識部の精度不足が原因でマイク収録環境において発生したノイズを誤認識してしまうことがあるという問題が観測された. また, 割り当て対象の音として音声などを登録した場合, その割り当て対象の音そのものを正確に再現することが困難であることがあり, 意図通りの結果が得られにくくなるという問題があることも分かった. これらは今後の課題とする.

7. ま と め

本研究では, 身の回りのものに任意の異なる音色を割り当ててリアルタイムに演奏可能な楽器に変えるインターフェースの提案を行った. 提案されたシステムでは, マイク一本のみを使って複数の対象にそれぞれ異なる音色を割り当てて同時に演奏でき, また, 割り当てる音色は波形を単に再生するわけではなくマイクからの入力音をそのまま加工するため, 奏法などによる表現の差異も直接出力に反映される. これにより, 身の回りのものに対して生楽器と同等の表現力を与えて自由に演奏することが可能なインターフェースを構築することができ, 実験でもその有用性が示された.

今後の課題としては, 本研究の手法ではアタックの特徴的なパーカッション音やスネアドラムのスナッピーの擦れる音などは再現が難しいという問題がある. こうした音色にはトランジェント部分とトーナル部分においてそれぞれ異なるアルゴリズムを採用する必要がある. 今後はこうしたより多くの種類の音色を扱うことを可能にするシステムの構築を目指す. また, 今回の手法では入力音から正確に駆動音を抽出しているとは言えない. リアルタイムに駆動音をより正確に抽出するためのアルゴリズムの構築も今後行っていく必要がある.

参 考 文 献

1) A.Sirdey, O.Derrien, R.Kronland-Martinet, M.Aramaki: Modal Analysis of Impact Sounds with ESPRIT in Gabor Transforms, *DAFx-11*, pp.387–392 (2011).
2) Kees van den Doel, Dinesh K. Pai: Modal Synthesis for Vibrating Objects, *Audio Anecdotes III*, K. Ed. A. K. Peters, Natick, Massachusetts, Vol.10, No.2, pp.1–8 (2003).
3) Cedric Fevotte, Jerome Idier: Algorithms for nonnegative matrix factorization with the beta-divergence, *Neural Computation, Article*, pp.1–24 (2011).
4) KORG WAVEDRUM : http://www.korg.co.jp/Product/Drum/WAVEDRUM/.
5) KORG WAVEDRUM Mini : http://www.korg.co.jp/Product/Drum/WAVEDRUMmini/.
6) Theremin : http://ja.wikipedia.org/wiki/テルミン
7) TableDrum : http://www.tabledrum.com/
8) David Merrill, Hayes Raffle, Roberto Aimi: The sound of touch: physical manipulation of digital sound, *CHI ' 08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pp.739–472 (2008).