情報処理学会論文誌 Vol. 51 No. 10 2007–2019 (Oct. 2010)

テルミンの音高・音量特性のモデルに基づくテルミン演奏ロボットの開発

水本武志[†1] 辻野広司[†2] 高橋徹[†1]
駒谷和範[†1,*1] 尾形哲也[†1] 奥乃博[†1]

本論文では，テルミンを演奏するロボットのためのテルミンの特性モデルと演奏動作生成手法について報告する．テルミンとは，演奏者の手の位置を動かして演奏する電子楽器である．楽器との物理的接触なしで連続的に音高と音量を操作できるので，ハードウェア構成が異なるロボットにも適用可能であるという点で，移植性が高い．テルミン演奏ロボットの主たる課題は，(1) 動作生成の物理的な基準点がないので，演奏法学習に要する学習サンプル数が多くなること，および (2) 演奏特性が静電的環境によって変化するので，適応的な演奏動作生成が必要であることの 2 点である．これらの課題に対して，我々は環境の影響をパラメータで表現した音高・音量特性モデルを構築し，少数の測定で音域内の任意の音を演奏できる制御手法を開発した．実験の結果，約 12 点の測定で音高が任意に制御できること，環境が変化しても所望の音高や音量で演奏できることを 3 種類のロボットで確認した．

Development of a Theremin Player Robot Based on Arm-Position-to-Pitch and -Volume Models

Takeshi Mizumoto,[†1] Hiroshi Tsujino,[†2]
Toru Takahashi,[†1] Kazunori Komatani,[†1,*1]
Tetsuya Ogata[†1] and Hiroshi G. Okuno[†1]

We present a theremin player robot towards an ensemble between humans and robots. A theremin, whose pitch and volume change continuously, can be played without any physical contacts. We thus expect that a robot system has high portability because it requires only few physical constraints. The problems for theremin playing are: (1) we have no physical reference points and (2) an environment affects sound characteristics seriously. To solve them, we develop a model-based feedforward arm control method based on our novel models of theremin's pitch and volume characteristics, which method realizes play an arbitrary sound using a few measurements. Experimental results show that our method works under four environments and three different robots.

1. はじめに

近年の制御理論，画像処理，音声認識技術の向上により，ロボットは工業用途からエンタテインメント用途にも応用が広がっている．たとえば，ソニーのペットロボット AIBO [1] や，タカラトミーのヒューマノイドロボット i-SOBOT など多くのエンタテインメントロボットが発売されている．エンタテインメントロボットの中でも音楽に関連するロボット（以降は**音楽ロボット**と呼ぶ）は，現在音楽はレジャーとして重要な位置を占めており，今後も成長が期待されるという報告[*1]にもあるとおり，今後重要になると期待される．特に楽器演奏は言語から受ける制約が少ないので，文化の異なる人々との楽しさの共有が期待できる．

音楽ロボットの中でも我々が本論文で議論する楽器演奏ロボットは，人とロボットがともに演奏（合奏）することで，インタラクティブなエンタテインメントを提供できる有望な分野である．また，合奏は，相手の演奏音という音響的な情報やアイコンタクトなどの視覚的な情報といったマルチモーダルな情報から相手の演奏に合わせた演奏動作の生成をリアルタイムで行うという困難な問題でもある．合奏を実現するうえでの課題は以下の 4 点である：

(1) 楽器演奏ロボットの開発
(2) ロボット自身の演奏音の認識
(3) 他の演奏者の演奏音と動作の認識
(4) 他の演奏者の演奏と音高やタイミングが合う演奏動作の生成

本論文では，(1)，(2) に着目し，**テルミン**[*2]を演奏するロボットを開発したので報告する．楽器演奏ロボットの観点から見たテルミンの利点は，物理的な接触なしに演奏できる点で

†1 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University
†2 株式会社ホンダ・リサーチ・インスティチュート・ジャパン
Honda Research Institute Japan, Co., Ltd.
∗1 現在，名古屋大学大学院工学研究科
Presently with Graduate School of Engineering, Nagoya University

⋆1 財団法人日本生産性本部の報告「レジャー白書 2009」によると，カラオケ，音楽鑑賞，コンサートといった音楽に関連する余暇の参加人口はすべて 20 位以内に位置している．

⋆2 テルミンとはロシアの物理学者 León Theremin によって発明された世界最古の電子楽器である．テルミンが生まれた歴史的な背景やオリジナルのテルミン（特許の内容の解説）に関しては文献 2) に詳しい．

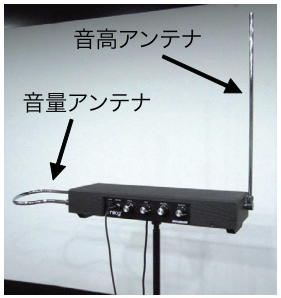


図 1　テルミンの写真
Fig. 1　Picture of a theremin.

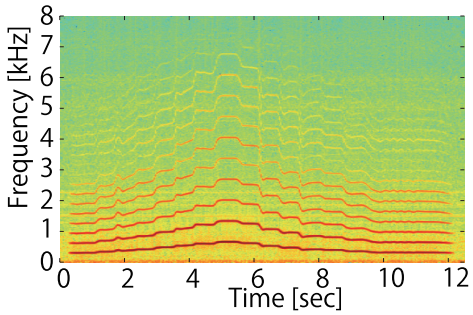


図 2　テルミン演奏の例
Fig. 2　Example of theremin playing.

ある．テルミン奏者は，演奏者の両手の位置と，テルミンに取り付けられている 2 本のアンテナ（**図 1** 参照）との距離をそれぞれ変化させることでテルミンの音高・音量を制御する．通常，演奏者の右手で音高を，左手で音量を制御する．また，テルミンは**図 2** に示す演奏例のように音高と音量が連続的に変化するので，トロンボーンやバイオリンと同様に固定された音階はない．そのため，グリッサンドなどの連続的な音高変化を用いた表現や，微分音などの音階を持つ楽器では演奏できない音高の表現が可能な楽器である．これらの利点から，人型ロボットがテルミンを演奏することには，次の 2 つの実用上の価値があると期待できる：合奏のための音楽ロボットを演奏専用の改造を施すことなく開発できる点，人の共演者と類似した人型のロボットを用いることで動作が予測しやすい合奏が実現可能である点．テルミンを演奏する際の困難な点は，(1) ピアノの鍵盤やギターのフレットような，演奏時の基準となる物理的な基準点が存在しないこと，および (2) テルミンの音高，音量と両腕の位置の関係（以降，**音高特性**，**音量特性**と呼ぶ）が気温や人の数といった周囲の環境（以降，**環境キャパシタンス**と呼ぶ）に応じて敏感に変化することである．事前に環境に関する情報が分かる場合は作り込み可能ではあるものの，様々な環境で実演するためには少数の測定で未知環境に適応できる演奏法が不可欠である．特に合奏の場合には，他の演奏者が環境キャパシタンスに影響を与えるので，適応的な演奏法が合奏の鍵になる．また逆に，手の位置によって音が変化する特徴をジェスチャ認識のための近接センサに応用した双方向音楽システムの報告[3),4)] もある．

通常，合奏では相手が近くにいるので，テルミン演奏ロボットはそのような状況でメロディを演奏することが必要である．これを実現するための要求条件は次の 2 点である．

テルミン演奏ロボットの要求条件

(1) 素早い腕の制御
(2) 異なる環境キャパシタンスへの適応

項目 (1) はメロディ演奏で要求される．テルミンは連続的に音が変化するので，素早く目標位置へ腕を移動させて移動先で静止しなければ，聴衆はロボットが演奏している音列を認識できない．項目 (2) は，上述のように他の演奏者が近くにいる合奏で要求される．これらの要求条件を満足するため，我々は**音高・音量特性を表すパラメトリックなモデルを構築**し，**モデルに基づくフィードフォワード制御手法**を開発した．本手法の利点は次の 2 点である．(1) 音を聞かずに目標位置へ腕を移動するので素早い腕の制御が実現できる．(2) 少数の腕位置の音高・音量を測定するだけで任意の音高と音量が出力できるので環境キャパシタンスが変動しても短時間でモデルパラメータを再推定できる．

従来のテルミン演奏ロボットは，演奏したい音高ごとに対応する腕の位置を探し，それらの対応関係をテーブルとして保持していた[5)]．この手法は，テーブルの初期化に時間がかかるために項目 (2) は満足できない．なぜなら，環境やロボットが変わるたびに演奏したい音高すべてに対して腕の位置を探す必要があるからである．さらに，計測した点以外は特性の情報がないため音高の正しさが保証できない．一方，本論文で述べる特性モデルは，(a) 少数の点の測定のみで任意の腕位置と音高の対応関係が求められ，(b) 異なる形状を持つロボット，異なる環境を表現可能である．特に音量に関しては，従来は定量的な制御が行われておらず，本モデルによって初めて可能になる．

本システムは，モデルのパラメータを推定するキャリブレーションフェーズと，パラメータに基づいて演奏を行う演奏フェーズの 2 フェーズに分割して動作する．このことは，キャリブレーション後に環境キャパシタンスは変化しないという仮定を意味する．ただし，変化がゆるやかであれば定期的にキャリブレーションを行うことで近似的に仮定が成立する．したがって，キャリブレーションは時間を短く抑えて何度も行う必要がある．そのため，モデル構築で時間のかかる測定点の数をできるだけ少なく抑えることが重要となる．

本論文の構成は以下のとおりである．2 章で関連する音楽ロボット研究について述べる．次に 3 章でテルミンの動作原理について述べ，4 章でテルミンの特性とモデル化について述べる．続いて本モデルに基づく制御システムを 5 章で述べ，そして 6 章で評価実験の結果を示し，7 章で本論文をまとめる．

2. 関 連 研 究

本章では，まず音楽ロボット研究における本研究の位置づけについて述べ，次に関連するテルミン演奏ロボットについてまとめる．

2.1　音楽ロボット研究の概要

音楽ロボットの表現方法は，(1) 歌などの音声，(2) ダンスなどの動作，(3) 楽器演奏に分類できる．(1) 歌うロボットには Mizumoto らのビートを数えるロボット[6] や，Murata らのビートに合わせて足踏みをしながら歌う 2 足歩行ロボット[7] がある．いずれも，ロボットに装着されたマイクから得られた音を入力音に用いている．(2) ダンスロボットには，Nakaoka らの人の全身動作を模倣して踊る 2 足歩行ロボット[8] や Kosuge らの腕と車輪移動で社交ダンスを行うロボット[9]，自分で聞いた音に合わせて足踏みするロボットの報告[7),10] がある．(3) 楽器演奏ロボットには，単独演奏と合奏の報告がある．単独演奏に関しては，キーボードを演奏する歩行機能のない全身人型ロボット WABOT-2[11] や，両腕と上半身を持つ人型バイオリン演奏ロボット[12]，人工唇を用いているが人型ではないトロンボーン演奏ロボット[13] など，多くの報告がある．WABOT-2 を除くとほとんどのロボットがメロディではなく単一の音の演奏を扱っていたが，近年では Solis らの人型フルート演奏ロボット WF-4RIV[14] や人工唇と指のみのサックス演奏ロボット WAS-1[15] が，実際に複雑なメロディの演奏にまで到達している．これらのロボットは MIDI 形式の楽譜や目標音高を入力とし，それを忠実に演奏することをタスクとしていた．人との合奏に関しては，Petersen らの交互に演奏する形式の簡単な合奏[16] や，Weinberg らの即興的な合奏ロボット[17),18] の報告がある．

テルミンが従来研究で扱われていた楽器と比較して最も異なる点は，楽器とロボットとの物理的接触なしに演奏が可能なことである．したがって，フルート[14] とサックス[15] の演奏に必要な人工唇や，キーボード演奏[11] に必要な精密な指の機構のような特殊なハードウェアを必要としない．このため，テルミンは腕が 2 本あるという要件を満たす多くロボットで演奏可能である．したがって，テルミン演奏システムは既存の多くのロボットに実装できるという点で高い移植性を持つと期待できる．

一方，従来の楽器に比べてテルミン演奏で困難な点は，適応的な制御が要求されることである．なぜなら，音高特性，音量特性は環境キャパシタンスによって変化するので，仮に事前に十分な準備ができたとしても，その環境は刻々と変化してしまうからである．しかし，従来研究では演奏動作と演奏音の関係は事前に得られると仮定していた．

2.2　テルミン演奏ロボットに関する従来研究

テルミンは単音楽器なので，制御すべき要素は音高と音量の 2 種類である．合奏の実現には，楽譜を正しく演奏するための音高の定量的な制御に加えて，音量の定量的な制御も必要である．なぜなら，合奏においては演奏相手と音量のバランスをとる必要があるので，ロボット自身の演奏音における相対的な音量制御ではなく，絶対的な “n [dB]” で演奏するといった音量制御が必要だからである．しかし，定量的な音高と音量の制御を同時に実現している報告は後述するように今までなかった．以下で各問題に対する従来研究をまとめる．

2.2.1　音高制御問題

音高制御へのアプローチには大きく分けてフィードバック制御[19] とフィードフォワード制御[5] の 2 種類がある．フィードバック制御は，演奏時にテルミンの音を聞いて腕位置を調整するので正確な音高制御が可能である．しかし，フィードバック制御は次の 2 つの理由で 1 章で述べた要求条件を満たさない：第 1 に，目標音高に到達するまでの時間が長いので，曲によっては 1 秒未満で次々と目標値が変化するメロディ演奏ができない．第 2 に，フィードバック制御を実現するには徐々に音高を変化させなければならないが，それでは演奏しているメロディを聴衆が認識できない．それに対してフィードフォワード制御は，素早い腕の制御が可能なので要求条件 (1) を満たすが，腕の目標位置の予測に用いるモデルのロバスト性（要求条件 (2)）への対応が不可欠である．なぜなら，フィードフォワード制御を行うには適切な目標位置を実際の音を聞かずに予測する必要があるが，変化する環境キャパシタンスをモデルに組み込まなければ予測がずれていくからである．

Alford らは，ルックアップテーブルに基づくフィードフォワード制御手法を提案した[5]．この手法は，事前に演奏するすべての音名（C3，D4 など）に対して適切な関節角度を出力するテーブルを作成しておき，演奏時はテーブルに基づいて腕を制御する．彼らの手法は，テーブルの作成時に演奏する音名ごとに適切な関節角度を試行錯誤しながら発見する必要があるので時間がかかる．このため，環境キャパシタンスが変動するたびに，時間のかかるテーブルの再構築が必要である．

それに対して Mizumoto らは，パラメトリックモデルに基づくフィードフォワード制御を提案した[20]．この手法では，ロボットの手の位置とテルミンの音高をパラメトリックにモデル化することで，少数の点で音高を測定するだけで任意の音高の制御を可能にしている．本論文は，文献 20) をベースとして，音量制御を拡張したものである．

2.2.2　音量制御問題

音量制御は合奏に不可欠な機能であるが，音量制御を扱った報告は少なく，前節で述べた

Hulst ら[19] も Mizumoto ら[20] も音量の制御はまったく扱っていない．Alford らが文献 5) で，ON/OFF の制御と相対的な音量の制御手法を実装しているのみである．音量制御が扱われてこなかった理由の 1 つは，音高制御への依存関係が原因である．次章で詳述するように，同じ音量腕の位置でも音高によって音量は異なる．

3. テルミンの動作原理

図 3 の概略図に示すとおり，テルミンは主に音高制御と音量制御の 2 種類の回路から構成されている．どちらの回路も，2 つの発振回路を持ち，片方の回路のうち 1 つのコンデンサ[⋆1]がアンテナとして外に出ている．テルミンの音高，音量は各発振回路の発振周波数の違いよって生じるうなりを用いて制御されているので，演奏者の手の位置を変化させることで音高，音量を制御できる．うなりとは，2 つの発振回路の出力を乗算したときに生じる波で，それぞれの発振周波数を f_1，f_2 とすると，$f_1 - f_2$ と $f_1 + f_2$ の周波数成分を持つ．音高制御回路は，Low Pass Filter（LPF）を用いてうなりの $f_1 - f_2$ の周波数成分のみを取り出し，それを出力する．ここで，取り出された周波数がテルミンの出力音の音高となる．一方，音量制御回路は，同様に LPF を用いて周波数成分を取り出し，それを積分して，音高を出力する直前の増幅器の制御入力とする．

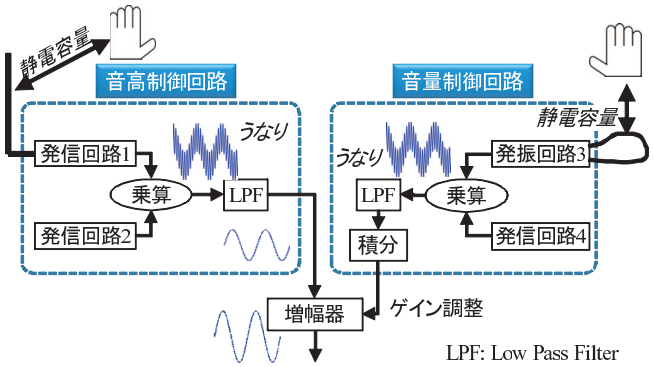


図 3　テルミンの内部構造
Fig. 3　Internal structure of a theremin.

⋆1 2 つの導体と，その間の不導体から構成される受動回路素子で，導体間に電荷を貯められる．貯められる電荷の量は，導体間の距離と不導体の特性（誘電率）で決定される．

テルミンの演奏において，(1) 音高アンテナと演奏者の右手のそれぞれを導体，両者の間の空間を不導体と見なしたときの仮想的なコンデンサと，(2) 音量アンテナと演奏者の左手のそれぞれを導体，両者の間の空間を不導体と見なしたときの仮想的なコンデンサの，それぞれの静電容量（環境キャパシタンス）が，テルミンの出力音の (1) 音高と (2) 音量を決定する．したがって，手の位置を動かしてコンデンサの導体間距離を変化させると環境キャパシタンスは変化し，結果として発振回路の発振周波数が変化する．

テルミンの音高と音量を決定する環境キャパシタンスは時間によって様々に変化していくので，事前に求めておくことは困難である．なぜなら，手とアンテナ間の空気がコンデンサを構成する不導体の部分になるので，テルミンの周囲の変化，たとえば気温や周囲の人の位置や人数が静電容量に影響を与えるからである．したがって，環境キャパシタンスがつねに一定であると仮定するのは現実的ではない．我々のアプローチは，音量特性と音高特性をパラメトリックにモデル化するので，定期的にパラメータを再推定することで本問題を解決できる．

4. テルミンの音高・音量特性のモデル化と制御手法

本章では，テルミンの音高・音量特性のパラメトリックなモデルと，モデルに基づく制御手法を提案する．4.1 節で音高・音量特性の計測条件を示し，4.2 節で計測した音高・音量特性を考察し，モデルを構築する．そして 4.3 節でモデルパラメータの推定手法を述べ，4.4 節で本モデルに基づく制御方法を示す．本制御手法のポイントは，Alford らのようにテルミンの音高を検索すべき対象ではなく，パラメータ推定のための学習データとして扱った点である．したがって，学習データに用いていない音高も，適切に補間することができる．これによって，少数のデータを学習データとしても任意の音高が出力可能になる．

まず，本論文で用いる記号を定義する．x_p をロボットの右腕（音高を操作する腕のこと．以降は『音高腕』と呼ぶ）の手先と音高アンテナとの最短距離とし，x_v を左腕（音量を操作する腕のこと．以降は『音量腕』と呼ぶ）の手先と音量アンテナとの最短距離とする．ただし，x_p と x_v は実際の距離である必要はなく，距離が増加すれば増加し，減少すれば減少する尺度でよい．なぜなら，距離の定義の違いによって生じる特性の差は高い自由度で設計したモデルのパラメータ推定で吸収できるからである．たとえば関節角度を尺度に用いても測定結果との誤差が小さいパラメータを推定できる．次に，x_p と x_v を離散化した点を $x_{p0}, \cdots, x_{pi}, \cdots, x_{pN}$，および $x_{v0}, \cdots, x_{vj}, \cdots, x_{vM}$ とする．テルミンの音高と音量は，ロボットの音高腕が x_{pi}，音量腕が x_{vj} のときに音高が p_i，音量が v_{ij} とする．また，

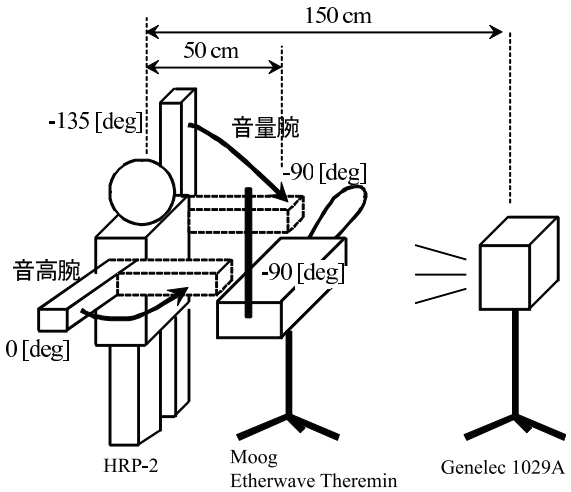


図 4　実験環境
Fig. 4　Configurations for experiments.

それぞれの値のモデルによる推定値は，ˆ を付けて表す．

4.1　音高・音量特性の計測条件

図 4 に計測環境の概略を示す．テストベッドとしては，ヒューマノイドロボット HRP-2 を用いた．テルミンには Moog Music 社の Etherwave Theremin を用い，テルミンの音はロボットの頭部に設置した 1 チャネルのマイクロフォンで収録した．人の演奏者と同様，ロボットはテルミンの演奏時に，右腕で音高操作，左腕で音量操作を行う．両腕の自由度は 1 とし，音高腕の可動域は音高最小の $-90°$ から音高最大の $0°$ までの $90°$，音量腕の可動域は音量最大の関節角度 $-135°$ から音量最小の関節角度 $-90°$ までの $45°$ とした．ロボットとテルミンとの距離は 50 [cm]，テルミンの音を再生するスピーカとの距離は 150 [cm] である．

テルミンの音高特性と音量特性は，音高腕と音量腕の可動域をそれぞれ 40 等分し，両端を含む各 41 カ所に腕の位置を固定した状態で，腕位置の全組合せ（$41 \times 41 = 1681$）について，テルミンの音を 1 [sec] ずつ測定した．

以下に示す実際に測定したテルミンの音高・音量特性は環境によって異なるが，以降で述べる傾向は変わらない．まず音高特性を**図 5** に示す．縦軸は音量腕の位置を，横軸は音高

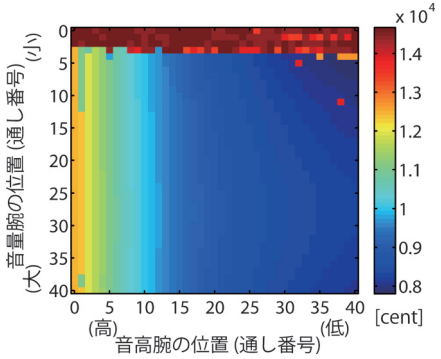


図 5　測定したテルミンの音高特性
Fig. 5　Theremin's pitch characteristics.

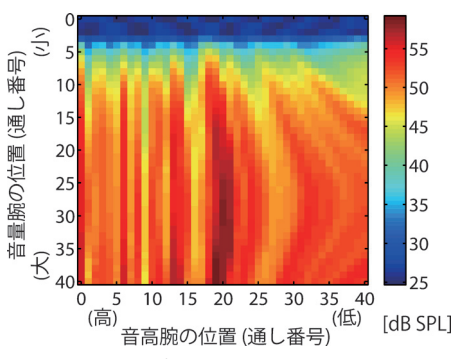


図 6　測定したテルミンの音量特性
Fig. 6　Theremin's volume characteristics.

腕の位置を表す．音高の推定には自己相関関数に基づく手法[21] を用いる．ただし，フレーム幅は 42.6 msec（2048 samples），シフト長は 10.6 msec（512 samples）とし，各フレームの音高推定値の中央値をその点での音高とした．音高は次式で定義される cent [⋆1] の単位で表す．

$$c\ [\text{cent}] = 1200 \log_2 (f\ [\text{Hz}]) \tag{1}$$

図 5 の上部の音高が不自然に高くなっているのは，音量が小さすぎるために音高推定に失敗しているからである．図 5 から，音高腕がアンテナに近い図の左側ほど音高が高く，アンテナから遠い図の右側ほど音高が低いことが分かる．次に音量特性を**図 6** に示す．縦軸は音量腕の位置を，横軸は音高腕の位置を表す．音量の単位は，収録した音の分散を次式で定義される音圧レベル [dB SPL] を用いる．

$$v\ [\text{dBSPL}] = 20 \log_{10} \frac{\sqrt{1/N \sum_{t=0}^{N} x(t)^2}}{20 \times 10^{-6}} \tag{2}$$

ただし，$x(t)$ はテルミンの出力波形を，t は時間を表す．図 6 から，音量アンテナに近い図の上側ほど音量が小さく，音量アンテナから十分離れた図の下半分は音量が比較的大きいことが分かる．しかし，詳しくは 4.2.2 項で考察するが，図 5 のように単調な増加はしていない．

⋆1 ある 2 つの音高が 100 [cent] 違うとき，それらは半音だけ異なっていることを示す．

4.2　特性の考察とモデルの構築

4.2.1　音高特性の考察と音高モデルの構築

図 5 より，音高は音高腕の位置のみによってほぼ決定され，かつ，音高は音高腕の距離に対して非線形に単調増加していることが分かる．したがって，音高特性は音高腕の位置と環境キャパシタンスの非線形関数として次式のように定式化した：

$$\hat{p} = M_p(x_p; \boldsymbol{\theta}) = \frac{\theta_2}{(\theta_0 - x_p)^{\theta_1}} + \theta_3 \tag{3}$$

ただし，$M_p(x_p; \boldsymbol{\theta})$ は音高モデル，x_p は音高腕の位置，$\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)$ はモデルパラメータ，$\hat{p}$ は音高モデルによって推定される音高（[Hz]）を表す．右辺の第 1 項がパラメータ $\boldsymbol{\theta}$ に対応する環境キャパシタンスにおける音高の増加の速さを表し，第 2 項が音高腕がアンテナから十分遠いときの音高を示す．

本モデルと関連して，Skeldon らがテルミンの環境キャパシタンスの物理モデル[22] を提案しているので，彼らのモデルと比較して議論する．Skeldon らは環境キャパシタンスを，音高腕位置を x としたときの静電容量と x が無限遠ににあるときの静電容量の和で表現している．また，静電容量が腕とアンテナの距離に対して指数関数的に増加するというモデル化を行っている．静電容量をモデル化するアプローチは汎用的ではあるが，対象としているのが特殊な簡易型のテルミンであり，一般的に流通している市販のテルミンより単純化された問題を扱っている．実際，予備実験の結果より市販のテルミンの場合は環境キャパシタンスによっては指数関数と増加の速さが異なっていた．したがって，本論文で扱う一般的な，周辺回路を含んだテルミンを演奏する問題においては，増加の仕方も可変にする必要がある．そのために，増加の速さと非線形性を制御するために指数と係数をパラメータ θ_1，θ_2 で表し，最適な増加の仕方を表現できるモデルとした．

4.2.2　音量特性の考察と音量モデルの構築

図 6 の音量特性について議論する．もし音量が音高に独立であれば，水平方向は同じ値になるはずであるが，図の上端の，無音領域以外は水平方向でも音量が変動している．つまり，音量は音量腕の位置（縦軸）だけでなく，音高腕の位置（横軸）にも影響を受けている．したがって，音量モデルは音量腕の位置と環境キャパシタンスに加えて，音高腕の位置も変数に持つ必要がある．さらに，音量特性には次の 2 つの特徴が見いだせる：

(1)　音量腕が音量アンテナに十分近い無音領域と，それ以外の有音領域が存在する．

(2)　音量特性を音高腕の位置ごと（縦方向）に見ると，それぞれの音量変化の多くは 1 つか 2 つのピークを持っている．

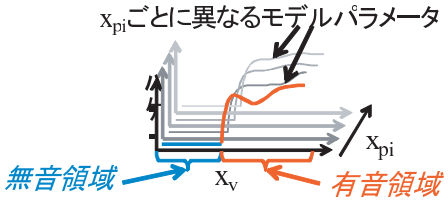


図 7　音量モデルの概略図
Fig. 7　Overview of our volume model.

我々は音量特性を音高腕の位置ごとに異なるパラメータを持つ多項式で以下のようにモデル化した：

$$\hat{v} = M_v(x_v, x_p; \mathbf{a}(x_p), b(x_p)) = \begin{cases} \displaystyle\sum_{n=0}^{d} a_n(x_p) x_v^n & (b(x_p) < x_v) \\ v_{min} & (\text{otherwise}) \end{cases} \tag{4}$$

ただし，$\hat{v}$ は音量モデルによって推定される音量を，$\mathbf{a}(x_p) = (a_n(x_p), \cdots, a_0(x_p))$ は有音領域の多項式モデルの係数を，d は多項式の次元を，$b(x_p)$ は無音領域と有音領域との境界となる x_v を，v_{min} は無音領域での音量，すなわち背景雑音を表す．

本モデルのアイデアの概要を**図 7** に示す．本モデルは音高腕の位置ごとに，異なる 3 種類のパラメータ $\mathbf{a}(x_p)$，$b(x_p)$，v_{min} を持つ．したがって，式 (4) の多項式の次元を d とし，音高腕の可動域を N 等分したとすると，合計のパラメータ数は $(1+1+d)N$ 個である．多項式の次元 d は，低すぎると特性を表現できず，高すぎると過適応や必要な学習データが増えるという問題がある．本論文では，予備実験により $d = 4$ を用いる．

4.3　音高・音量モデルのパラメータ推定法

4.3.1　音高モデルのパラメータ推定

$N+1$ カ所の音高腕位置が $\boldsymbol{x}_p = (x_{p0}, \cdots, x_{pN})$ のときのテルミンの音高がそれぞれ $\boldsymbol{p} = (p_0, \cdots, p_N)$ であるとする．所与の音高に対する関節角を求めるには，式 (3) が非線形単調増加関数なので，次式で示す評価関数の L^2 ノルムを最小化するパラメータ $\boldsymbol{\theta}$ を求める非線形最適化問題を解けばよい．

$$\mathbf{f}(\mathbf{x_p}, \mathbf{p}; \boldsymbol{\theta}) = (f_0(x_{p0}, p_0, \boldsymbol{\theta}), ..., f_{N+1}(x_{pN}, p_N, \boldsymbol{\theta}))^T$$
$$\text{where} \quad f_i(x_{pi}, p_i, \boldsymbol{\theta}) = p_i - M_p(x_{pi}, \boldsymbol{\theta}) \tag{5}$$

つまり，最適なパラメータ $\boldsymbol{\theta}^*$ は次式で表せる

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}(||\mathbf{f}(\mathbf{x_p}, \mathbf{p}; \boldsymbol{\theta})||^2) \tag{6}$$

本最適化問題を Levenberg-Marquardt 法（LM 法）[23] を用いて解く．LM 法は，次式に従ってパラメータを更新する：

$$\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^{old} - \left(\mathbf{J^T J} + \mu \mathbf{I}\right)^{-1} \mathbf{J}^T \mathbf{f}(\mathbf{x_p}, \mathbf{p}; \boldsymbol{\theta}^{\mathbf{old}}) \tag{7}$$

ただし，μ はモデルによる推定値と学習データとの誤差に基づいて各反復ごとに自動的に決定される[24] 学習パラメータ，$\boldsymbol{I}$ は単位行列，$\mathbf{J}$ は次に示す評価関数のヤコビ行列である．

$$(\mathbf{J}(\theta))_{ij} = \partial f_i / \partial \theta_j \qquad (i = 0, \cdots, N, j = 0, .., 3) \tag{8}$$

ヤコビ行列の各要素は評価関数の偏微分であり，次式で表現される．

$$\begin{cases} \partial f_i/\partial\theta_0 = -\theta_1\theta_2(\theta_0 - x_i)^{-(\theta_1-1)}, \\ \partial f_i/\partial\theta_1 = \theta_2(\theta_0 - x_i)^{-\theta_1}\log(\theta_0 - x_i), \\ \partial f_i/\partial\theta_2 = -\frac{1}{(\theta_0 x_i)^{\theta_1}}, \\ \partial f_i/\partial\theta_3 = -1. \end{cases} \tag{9}$$

ここで，$\boldsymbol{\theta}$ に適当な初期値を与えることで，適切なパラメータが推定できる．

4.3.2　音量モデルのパラメータ推定

音量モデルを同定するために求めるべきパラメータは，(1) 多項式モデルの係数 $\boldsymbol{a}(x_p)$，(2) 無音領域と有音領域の境界 $b(x_p)$，(3) 背景雑音 v_{min} である．

第 1 のパラメータである有音部分の多項式モデルの係数 $\mathbf{a}(x_{pi})$ は，音高腕の位置 x_{pi} を固定し，音量腕の位置と音量の組を (x_{vj}, v_{ij}) と表したときの次の連立方程式を解いて求めることができる．

$$\begin{cases} a_d(x_{pi})x_{v0}^d + \cdots + a_1(x_{pi})x_{v0} + a_0(x_{pi}) &= v_{i0} \\ \vdots & \\ a_d(x_{pi})x_{vM}^d + \cdots + a_1(x_{pi}x_{vM} + a_0(x_{pi}) &= v_{i(M+1)} \end{cases} \tag{10}$$

次に，第 2 のパラメータである無音領域と有音領域の境界 $b(x_p)$ は，音量腕の位置 x_{vj} が大きいほど，つまりアンテナから遠いほど音量は大きくなる傾向にあるので，音量が閾値 T_{th} を下回るという条件下で腕位置の最大値を $b(x_{pi})$ とする．つまり，$b(x_{pi})$ は次式で求める：

$$b(x_{pi}) = \max_{j \in \{j | v_{ij} < T_{th}\}} x_{vj} \tag{11}$$

第 3 のモデルパラメータ v_{min} は，音量がないときの音量とすればよい．

4.4　モデルに基づくフィードフォワード制御

4.4.1　フィードフォワード音高制御

音高制御の目標は，目標音高列を出力する音高腕の位置を求めることである．目標音高列は，与えられた楽譜の音名を平均律に基づいて音高に変換することで得る．ある音名 n $(C = 0, C\sharp = 1, \cdots, B\flat = 10, B = 11)$ と，オクターブ o（o は整数）が与えられたとき，平均律に基づいて対応する音高 p に変換する式を次に示す．

$$p = 440 \cdot 2^{(o-4)\ \sqrt[12]{2^{n-9}}} \tag{12}$$

次に，音高を音高腕の目標位置に変換する式を求める．モデルパラメータ推定の際は音高腕位置を独立変数，音高を従属変数として扱ったが，音高腕の制御にはその逆モデルが必要となる．逆モデルは，次式に示す逆関数を計算することで解析的に求めることができる．

$$\hat{x}_p = M_p^{-1}(p, \boldsymbol{\theta}) = \theta_0 - \left(\frac{\theta_2}{p - \theta_3}\right)^{1/\theta_1} \tag{13}$$

以上の方法で，所与の音名に対応する音高腕の目標位置を求められる．

4.4.2　フィードフォワード音量制御

音高腕の位置が，可動域を N 等分した i 番目の位置，すなわち x_{pi} であるとしたときの式 (4) の逆関数を以下に示す：

$$M_v^{-1}(v; \mathbf{a}(x_{pi}), \mathbf{b}(x_{pi})) = \begin{cases} v_{min} & (v < v_{min}) \\ \hat{x}_v & (\text{otherwise}) \end{cases} \tag{14}$$

楽譜は音高と音量の組で与えられるので，4.4.1 項で求めた音高腕位置 $\hat{x}_p$ と事前に与える目標音量 v を用いて音量腕の位置 x_v が求まる．このとき，目標音量が v_{min} よりも小さければ音量腕の位置は $b(x_p)$ とする．それより大きければ，多項式 $a_{n-1}x_v^{n-1} + \cdots + a_1x_v + a_0 = v$ を解いて，音量腕の位置を推定する．$\hat{x}_p$ が測定位置の間にあれば，両隣の音高腕位置から音量腕位置をそれぞれ求めて線形補完する．

5.　テルミン演奏システム

本演奏システムでは，前章で述べた**特性の逆モデルに基づくフィードフォワード制御**によって音高と音量を制御する．音高・音量の制御は，**キャリブレーションフェーズ**と**演奏フェーズ**の 2 段階に分ける．キャリブレーションフェーズでは，5 章で述べた方法で特性を

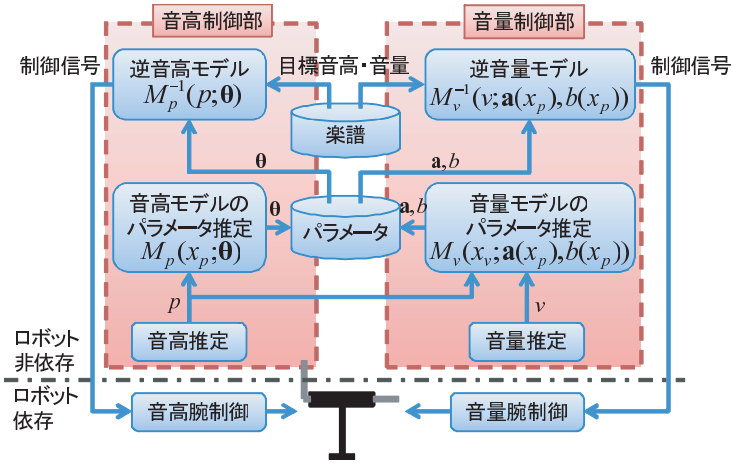


図 8　テルミン演奏システムの構成図
Fig. 8　Overview of our thereminist robot system.

測定し，モデルのパラメータを推定する．演奏フェーズでは，音高モデルと音量モデルの逆関数を用いることで，所与の楽譜に対応する両腕の位置列を求めて制御する．このとき，パラメータにはキャリブレーションフェーズで推定した値を用いる．

図 8 に我々が開発したテルミン演奏システムの概略図を示す．我々のモデルはテルミンの特性のみをモデル化しており，ロボットの物理的な制約を含んでいない．したがって，本システムはロボットのハードウェアへ依存する部分と独立な部分に分離できるので，他のロボットに実装したい場合は，そのロボットに依存する部分のみを入れ替えるだけでよい．

6. 実　　験

本章では，テルミンの音高モデルと音量モデルに関する 4 種類の評価実験について述べる．実験 1 では音高モデルのパラメータ推定に必要な学習サンプル数を明らかにする．実験 2 では音高モデルの環境変化に対するロバスト性を評価するために，異なる環境下でモデル誤差を評価する．実験 3 では，音量を一定に保ったまま音高を変えるというタスクを用いて本音量制御手法を評価する．最後に実験 4 で，本システムの移植性を 3 種のロボットに実装することで示す．各ロボットの演奏の質の尺度には，実際に演奏された音高軌跡と楽譜から求めた音高軌跡との平均 2 乗誤差を用いる．また，それらの尺度自身を議論するため，

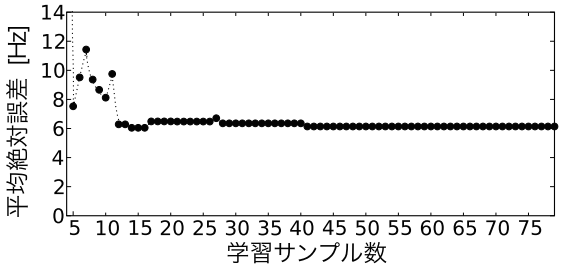


図 9　実験 1：学習サンプル数とモデル誤差
Fig. 9　Result 1: Error when the number of training data varies.

基準としてアマチュアのテルミン演奏者の音高軌跡と誤差も示し，各ロボットと比較する．なお，実験条件は 4.1 節で述べた特性の計測条件と同様に設定した．

6.1　実験 1：学習サンプル数変化に対する音高モデルのロバスト性の評価

本実験では，音高モデルパラメータの推定に必要な学習サンプル数（音高腕の分割数 N）を明らかにする．推定されたパラメータの適切さは，音高腕の各位置 x_{pi} ごとに測定した音高 p_i と，推定したパラメータ $\hat{\boldsymbol{\theta}}$ を代入した音高モデルで予測した音高 $M_p(x_{pi},\hat{\theta})$ の平均絶対誤差（MAE，Mean Absolute Error）で評価する．MAE が小さいほど，測定した音高を正確に予測するパラメータを推定できたといえる．MAE の定義式は以下のとおりである：

$$MAE = \frac{1}{N+1}\sum_{i=0}^{N}\left|p_i - M_p(x_{pi},\hat{\theta})\right| \tag{15}$$

なお，本実験では音高のみを評価するので，音量腕は使用しない．評価の手順は以下のとおりである．まず，音高腕の可動域を 80 等分し，81 点のテルミンの音高を収集する．そして，収集した音高と対応する音高腕の位置の組 (p_i, x_{pi}) を用いてパラメータを推定する．このとき，パラメータ推定に用いるデータの数（N）を 5 から 80 まで変化させ，MAE を評価する．MAE が十分小さくなる最小の N が，本モデルに適したデータ数である．

結果を**図 9** に示す．横軸は学習サンプル数を，縦軸は MAE（[Hz]）を示す．図より，サンプル数が増加すると MAE が減少するが，$N = 12$ 付近で収束する．$N > 12$ の範囲では誤差が減少しておらず，MAE は約 6 [Hz] で飽和している．この誤差の影響の大きさは演奏している音高によって異なるが，通常のメロディの音高は数百 [Hz] なので，それに比べれば十分小さい．

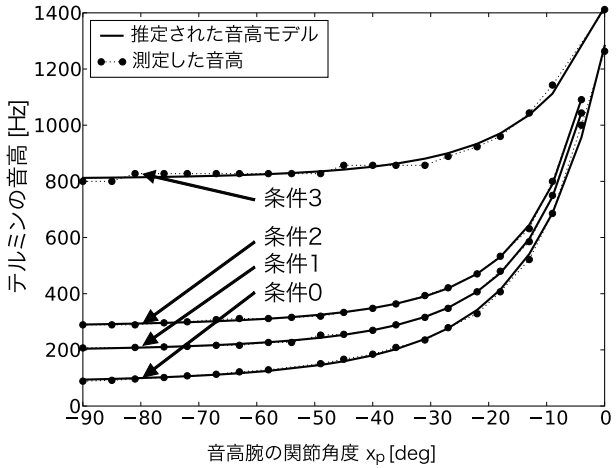


図 10　実験 2：異なる環境キャパシタンスでの音高モデル推定結果
Fig. 10　Result 2: Estimated pitch model under different environmental capacitances.

以上の結果より，音高モデルのパラメータの推定には 12 点を観測すれば十分であることが示された．文献 5) では演奏する音高の種類に依存してキャリブレーションフェーズの時間が変化するのに対して，本論文で提案する方法は依存しない．なお，本実験では各点では 1 秒ずつ録音したので，音高腕の移動時間や音高の推定時間を含めても 1–2 分で終わる．

6.2　実験 2：環境キャパシタンス変化に対する音高モデルのロバスト性の評価

本実験では，環境キャパシタンスを変化させることで，異なる音高特性における音高モデルの精度を評価する．具体的には，金属製の箱をテルミンの付近に設置し，箱の位置を変化させて環境キャパシタンスを人為的に変化させ，各位置でパラメータを推定した．実験条件は 4 種類で，条件 0 が箱がない状態，条件 1 が最もテルミンから遠く，2 が中間，3 が最も近い位置に金属製の箱を置いた状態である．それぞれの条件で，音高腕の可動域を 20 等分し，21 点の音高を計測して音高モデルのパラメータを推定した．

図 10 に各条件で推定したモデルを示す．実線が推定したパラメータで描いた音高特性の曲線，破線で接続された点が実際に測定した音高である．曲線の近傍に点が描かれているので，直感的に妥当なパラメータが推定されていることが分かる．次に，**表 1** に推定されたパラメータと MAE を示す．各条件で，右端列の MAE は 3 [Hz] から 10 [Hz] 程度であり実験 1 で飽和した誤差と同程度である．したがって，本モデルは妥当であると考えられる．

表 1　実験 2：推定したパラメータ
Table 1　Result 2: Estimated pitch model parameters and errors.

Condition	θ_0	θ_1	θ_2	θ_3	**MAE**
0	56.78	4.65	1.70×10^{11}	79.55	7.91
1	39.36	3.97	2.73×10^{9}	193.0	3.05
2	30.20	3.42	1.44×10^{8}	279.2	3.21
3	71.17	5.86	4.39×10^{13}	807.2	10.7

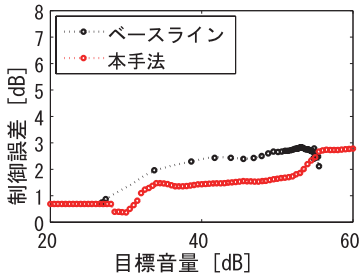


図 11　実験 3：条件 3 の制御誤差
Fig. 11　Result 3: Volume control error under condition ♯3.

6.3　実験 3：音高操作時の音量制御性能の評価

本実験では，音量を一定に保ちながら音高を変化させたときの音量の変化，つまり制御誤差を評価する．音高・音量の両特性は，両腕の可動域をそれぞれ 40 等分・5 等分（$N = 41$, $M = d + 1 = 6$，音量モデルの多項式の次元 d は 5 に設定）して測定した．条件は，テルミンの音量調節用つまみの位置を変化させることで設定した．音量調節用つまみを使えば音量特性のみを顕著に変えられるからである．特性は次の 3 条件で測定した：条件 1 が音量を小さくしたとき，条件 3 が音量を最大にしたとき，条件 2 はその中間である．

全条件で同様の傾向が見られたので，代表して条件 3 の実験結果のみを**図 11** に示す．横軸は目標音量，縦軸は制御誤差を示す．赤点は本モデルの結果，黒点はベースラインとして音量腕を固定した場合の結果を示す．本手法の制御誤差は目標値との差とすればよいが，ベースラインの場合は目標値がない．そこで，音高を変化させたときの音量の平均値を目標値とし，そこからの標準偏差を制御誤差と定義した．本実験でのタスクは一定の音量を保つことなので，音量の変動が小さいほどうまく制御できているといえる．したがって，縦軸の値は，小さい方が性能が高いことを示す．図 11 に示すとおり，ベースラインでも本手法で

も目標音量が約 25 [dB] 以下のときに誤差が小さい．これは，元々の音量が小さいからである．一方，目標音量が 30–50 [dB] のときに本手法の方が誤差が小さく，正確に制御できている．しかし，無音領域から有音領域の境界にあたる 30 [dB] の周辺では誤差が大きい．理由は，境界の推定誤差と，多項式モデルでは実際の音量の増加の速度に追いつかないことだと考えられる．特に後者は，単純に多項式モデルの次元を増加すると測定回数が増加してしまうというトレードオフがあるので，本論文の実装では測定回数を少なく抑えることを重視した．目標音量を 35 [dB] から 50 [dB] の範囲で演奏すれば安定して制御誤差は小さいので，実用上はこれで問題ない．そして，目標音量が大きくなるにつれて，再度制御誤差が大きくなっている．これは，出力可能な最大音量を超えるからである．

6.4　実験 4：テルミン演奏システムの移植性

本システムの移植性を示すため，本システムを 3 種類のヒューマノイドロボット：ホンダの ASIMO，川田工業の HRP-2 と HIRO に実装した．これらのロボットは，身体構造だけでなく制御方法も異なっている．ASIMO は 3 次元位置で手先座標を指定し[25]，HRP-2 と HIRO は関節角を直接指定する．また，それぞれ異なる部屋に設置されているので，環境キャパシタンスも異なる．使用する楽譜は**図 12** に示す童謡「かえるの歌」とした．ただし，楽譜から求める全目標音高をテルミンが出力可能な音高の範囲内に入れるために，図 12 における C は，平均律における 7 オクターブ目の C（C7）に設定した．

なお，音高・音量モデルのパラメータ推定に用いる腕の座標 x_p，x_v がスカラーなので，ASIMO においては，単純に 3 次元座標位置を用いると次元が合わない．そこで，腕の移動範囲を，座標 $\mathbf{r_1} = (x_1, y_1, z_1)$ と $\mathbf{r_2} = (x_2, y_2, z_2)$ を結ぶ直線を媒介変数 t で表した式

$$\mathbf{r} = \mathbf{r_1} + t(\mathbf{r_1} - \mathbf{r_2}) \tag{16}$$

に従って移動させる．こうすることで，両腕の移動範囲を表す媒介変数をそれぞれ x_p，x_v に用いれば本手法を適用できる．

次に，演奏の評価尺度について述べる．本実験では，単独演奏の質を定量的に議論するため，テルミンによるメロディ演奏を構成する 3 つの要素：(1) 音高，(2) 音量，(3) タイミングのうち，要素 (1) を楽譜と実際の演奏の cent を単位とした 2 乗誤差で，要素 (3) を演奏開始から終了までの全区間の平均値で評価する．まとめると，「与えられた楽譜と，実際に演奏された音高軌跡との 2 乗誤差の平均値」を評価尺度とする．演奏の質の評価尺度に (2) を含めない理由は，音量は合奏においてパートナが存在するときに重要な要素であり，本論文で議論する単独演奏の範囲ではその重要性は比較的低いからである．したがって，本論文の範囲においては音量制御性能の評価は 6.4 節で行った実験で十分である．

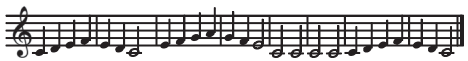


図 12　かえるの歌の楽譜
Fig. 12　Score of Frog Song.

各ロボットの音高軌跡を**図 13** に示す．横軸は時間，縦軸は C7 に対応する音高（9600 [cent]）を 0 とし，cent のスケールで表示した音高である．赤線は上から ASIMO，HRP-2，HIRO，人の音高軌跡，黒線は楽譜から求めた音高軌跡の正解を表す．人はキーボードとテルミンの演奏経験を持つ．ただし，4 段目の人の演奏は，演奏者が絶対音感を持っていなかったために正解より約 300 [cent] 低く演奏されていたので，音高軌跡を実際より 300 [cent] だけ高く表示している（長 2 度移調する操作に相当）．また，時間軸も他のロボットに合わせて調整している．なお，以上の操作によって以降の考察の結論は変わらない．

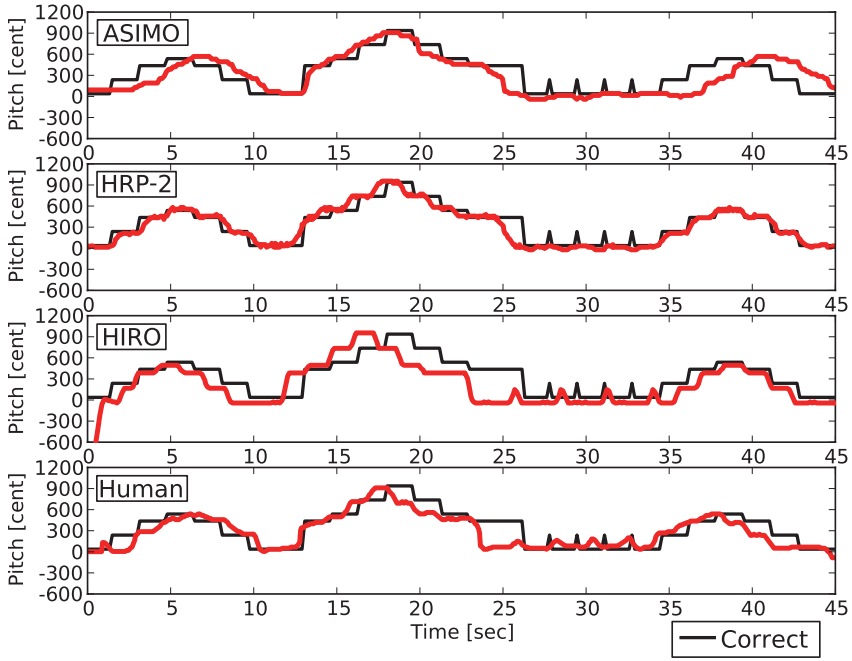


図 13　実験 4：3 種のロボットおよび人の音高軌跡
Fig. 13　Result 4: Pitch trajectries of three robots and a human.

まず平均2乗誤差による定量的な評価結果を示す．ASIMOの演奏は126.8 [cent]，HRP-2は52.9 [cent]，HIROは154.8 [cent]，人は112.4 [cent]であった．ただし，単位を合わせるために誤差の平方根を表示している．本論文で音名と音高との変換に用いている平均律では100 [cent]の差が半音の差を表すので，HRP-2を除くといずれも半音の1〜1.5倍だけずれている．誤差を人（アマチュアの演奏者）の評価結果と比較すると，HRP-2が人より正確に演奏しているものの，現状ではいずれも人とほぼ同程度の演奏の上手さであるといえる．

次に，図13の1段目のASIMOの結果を考察する．全体的に音高変化は緩やかなので，メロディの変遷はやや聞き取りにくい．さらに，直前の姿勢に依存して目標値が異なる制御手法を用いているので，同じ目標音高の12秒前後と26秒前後で異なる音高を演奏している．ただし，この全身制御のために，見た目は人の動作のように自然である．

2段目のHRP-2の結果を考察する．全体的に1段目の音高変化はASIMOに比べて急峻であり，楽譜上で同じ音符は同じ音高が出力されている．したがってメロディは比較的聴き取りやすい．また10–13秒や27–35秒の周辺に音高の振動が見られるが，これはHRP-2の質量の大きい腕の速い移動が原因である．また，ここでは関節角度で制御したために外見は機械的である．

3段目のHIROの結果を考察する．他のロボットに比べて，音高変化が最も急峻である．この理由は，単純な関節角度指定による制御で振動の生じない軽い腕を用いたからである．特に25–35秒の，細かく音高が変化する部分が楽譜どおりに演奏できている．ただし，外見はHRP-2と同様に機械的である．

最後に，4段目の人の結果を議論する．まず，音高軌跡は全体的に緩やかな曲線を描いており，ASIMOやHRP-2の軌跡に近い．しかし，25–35秒付近ではHIROと似て，楽譜どおりの急激な音高変化が見られる．次に，10–13秒や23–25秒の部分で，楽譜は一定値だが音高軌跡にはゆらぎが見られる．これは，人は腕の固定が困難だからである．似た振動は2段目の10秒付近にもあるが，これは体の振動によるものなので，性質が異なる．歌声において音高の微細振動が自然性に良い影響を与えるという知見[26]から，体の振動を抑制して，このようなゆらぎを意図的に与えれば演奏の質が向上する可能性がある．

7. ま と め

本論文では，人とロボットの合奏を目的として開発した，テルミン演奏ロボットについて報告した．テルミンは，環境キャパシタンスに応じて敏感に音高・音量特性が変化するので，テルミン演奏ロボットには適応的な制御が必要である．本問題を解決するため，我々はテルミンの音高特性と音量特性をパラメトリックにモデル化し，フィードフォワード制御によってテルミンの音高と音量を操作する方法を提案した．評価実験の結果，テルミンの音高・音量モデルは環境変化や学習データの変化にロバストであることを示した．また，音量モデルに基づく音量制御によって，音量を一定に保てることを確認した．最後に本システムを3種のロボットに実装し，本システムの移植性を確認した．

今後の課題は，プロのテルミン奏者とロボットの演奏の比較，音量特性モデルのパラメータ削減とパラメータ推定手法の洗練化である．本論文では楽譜と演奏された音高軌跡の2乗誤差の平均値を評価尺度に用いたが，プロのテルミン奏者と比較するには尺度の改善が必要である．特に，ビブラートなどの演奏表情を含めて評価するには，音高軌跡の動的な特性をとらえる尺度を設計する必要がある．また，合奏を評価するには音量を尺度に含める必要もある．フィードバック制御も重要な課題の1つである．なぜなら，速い曲なら本手法のみでも演奏できるが遅い曲や長い音を演奏する際はフィードバック制御も有効だからである．また，定期的にモデルパラメータを更新することで，動的な環境変化にも追従させられる．動的に変化する環境における音高・音量制御手法を確立すれば，人とロボットの合奏の実現に取り組めると考えている．合奏は，上記のような相手の演奏に関する音響的な情報だけでなく，ジェスチャなどの視覚的情報も含むマルチモーダルなインタラクションである．したがって，文献27)で報告されている演奏中のジェスチャ認識のような，視覚情報を用いた相手の演奏の認識にもあわせて取り組む予定である．

謝辞　本研究の一部は，科研費（S）とGlobal COEの援助をうけた．

参 考 文 献

1) Fujita, M.: AIBO: Toward the Era of Digital Creatures, *Intl. Journal of Robotics Research*, Vol.20, No.10, pp.781–794 (2001).
2) Glinsky, A.V.: The Theremin in the Emergence of Electronic Music, Ph.D. Thesis, New York University (1992).
3) Overholt, D., Thompson, J., Putanam, L., Bell, B., Kleban, J., Sturm, B. and Kuchera-Morin, J.: A Multimodal System for Gesture Recognition in Interactive Music Performance, *Computer Music J.*, Vol.33, No.4, pp.69–82 (2009).
4) Smirnov, A.: Music and Gesture: Sensor Technologies in Interactive Music and the Theremin based space control systems, *Proc. Intl. Computer Music Conf.* (*ICMC*), pp.511–514 (2000).
5) Alford, A., Northrup, S., Kawamura, K., Chan, K.W. and Barile, J.: A Music Playing Robot, *Proc. Intl. Conf. on Field and Service Robotics* (*FSR*), pp.29–31

(1999).
6) Mizumoto, T., Takeda, R., Yoshii, K., Komatani, K., Ogata, T. and Okuno, H.G.: A Robot Listens to Music and Counts Its Beats Aloud by Separating Music from Counting Voice, *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems* (*IROS*), pp.1538–1543 (2008).
7) Murata, K., Nakadai, K., Yoshii, K., Takeda, R., Torii, T. and Okuno, H.G.: A Robot Uses Its Own Microphone to Synchronize Its Steps to Musical Beats While Scatting and Singing, *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems* (*IROS*), pp.2459–2464 (2008).
8) Nakaoka, S., Nakazawa, A., Kanehiro, F., Kaneko, K., Morisawa, M. and Ikeuchi, K.: Task model of Lower Body Motion for a Biped Humanoid Robot to Imitate Human Dances, *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems* (*IROS*), pp.2769–2774 (2005).
9) Kosuge, K., Hayashi, T., Hirata, Y. and Tobimyama, R.: Dance partner root – Ms DanceR, *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems* (*IROS*), pp.3459–3464 (2003).
10) Yoshii, K., Nakadai, K., Torii, T., Hasegawa, Y., Tsujino, H., Komatani, K., Ogata, T. and Okuno, H.G.: A Biped Robot that Keeps Steps in Time with Musical Beats while Listening to Music with Its Own Ears., *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems* (*IROS*), pp.1743–1750 (2007).
11) Sugano, S. and Kato, I.: WABOT-2: Autonomous robot with dexterous finger-arm – Finger-arm coordination control in keyboard performance, *Proc. IEEE Intl. Conf. on Robotics and Automation* (*ICRA*), pp.90–97 (1987).
12) Shibuya, K., Matsuda, S. and Takahara, A.: Toward Developing a Violin Playing Robot – Bowing by Anthropomorphic Robot Arm and Sound Analysis, *Proc. IEEE Intl. Symp. on Robots and Human Interactive Communication* (*RO-MAN*), pp.763–768 (2007).
13) Kaneko, Y., Mizutani, K. and Nagai, K.: Pitch controller for automatic trombone blower, *Proc. Intl. Symp. on Musical Acoustics* (*ISMA*), pp.5–8 (2004).
14) Solis, J., Bergamasco, M., Chiba, K., Isoda, S. and Takanishi, A.: The Anthropomorphic Flutist Robot WF-4 Teaching Flute Playing to Beginner Students, *Proc. IEEE Intl. Conf. on Robotics and Automation* (*ICRA*), pp.146–151 (2004).
15) Solis, J., Petersen, K., Ninomiya, T., Takeuchi, M. and Takanishi, A.: Development of Anthropomorphic Musical Performance Robots: From Understanding the Nature of Music Performance to Its Application to Entertainment Robotics, *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems* (*IROS*), pp.2309–2314 (2009).
16) Petersen, K., Solis, J. and Takanishi, A.: Development of a Aural Real-Time Rhythmical and Harmonic Tracking to Enable the Musical Interaction with the Waseda Flutist Robot, *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems* (*IROS*), pp.2303–2308 (2009).
17) Weinberg, G. and Driscoll, S.: The Interactive Robotic Percussionist – New Developments in Form, Mechanics, Perception and Interaction Design, *Proc. ACM/IEEE Intl. Conf. on Human-Robot Interaction* (*HRI*), pp.456–461 (2007).
18) Weinberg, G., Raman, A. and Mallikarjuna, T.: Interactive Jamming with Simon: A Social Robotic Musician, *Proc. ACM/IEEE Intl. Conf. on Human-Robot Interaction* (*HRI*), pp.233–234 (2009).
19) van der Hulst, F.: Robotic Theremin Player, *Proc. National Advisory Committee on Computing Qualifications*, p.534 (2004).
20) Mizumoto, T., Tsujino, H., Takahashi, T., Ogata, T. and Okuno, H.G.: Thereminist Robot: Development of a Robot Theremin Player with Feedforward and Feedback Arm Control based on a Theremin's Pitch Model, *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems* (*IROS*), pp.2297–2302 (2009).
21) Camacho, A.: SWIPE: A sawtooth waveform inspired pitch estimator for speech and music, Ph.D. Thesis, University of Florida (2007).
22) Skeldon, K.D., Reid, L.M., Mclnally, V., Dougan, B. and Fulton, C.: Physics of the Theremin, *American Journal of Physics*, Vol.66, No.11, pp.945–955 (1998).
23) Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters, *SIAM Journal on Applied Mathematics*, Vol.11, No.2, pp.431–441 (1963).
24) Madsen, K., Nielsen, H.B. and Tingleff, O.: *Methods for Non-Linear Least Squares Problems*, *2nd ed.*, Informatics and Mathematical Modeling, Technical University of Denmark, DTU (2004).
25) Toussaint, M., Gienger, M. and Goerick, C.: Optimization of sequential attractor-based movement for compact behaviour generation, *Proc. IEEE/RAS Intl. Conf. on Humanoid Robots* (*Humanoids*) (2007).
26) Saitou, T., Unoki, M. and Akagi, M.: Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis, *Speech Comm.*, Vol.46, pp.405–417 (2005).
27) Lim, A., Mizumoto, T., Ohtsuka, T., Takahashi, T., Komatani, K., Ogata, T. and Okuno, H.G.: Robot Musical Accompaniment: Real-time Synchronization using Visual Cue Recognition, 情報処理学会全国大会 (2010).

(平成 22 年 2 月 3 日受付)
(平成 22 年 7 月 9 日採録)

水本　武志（学生会員）

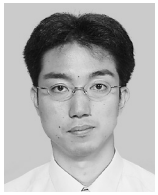
2008 年京都大学大学院情報学研究科知能情報学専攻修士課程修了．同年同専攻博士課程に進学．現在，在学中．主にテルミン演奏ロボットの開発と人とロボットの合奏の研究に従事．IROS2008 Award for Entertainment Robots and Systems Nomination Finalist，第 71・72 回情報処理学会全国大会学生奨励賞，IEA/AIE2010 最優秀論文賞を受賞．IEEE，日本ロボット学会各会員．

辻野　広司

1984 年東京工業大学理学部情報科学科卒業．1986 年同大学院情報科学専攻修士課程修了．1987 年（株）本田技術研究所入社．2003 年より（株）ホンダ・リサーチ・インスティチュート・ジャパン，チーフ・リサーチャ．脳型コンピュータ，知能システム，ヒューマンロボットインタフェース，画像認識等の研究に従事．IEEE，SFN，INNS，日本ロボット学会，人工知能学会，日本ソフトウェア科学会各会員．

高橋　　徹（正会員）

1996 年名古屋工業大学知能情報システム学科卒業．和歌山大学を経て，2008 年より京都大学大学院情報学研究科グローバル COE 助教．博士（工学）．ロボット聴覚，音声コミュニケーション，音声合成の研究に従事．IEEE，電子情報通信学会，日本ロボット学会，日本音響学会各会員．



駒谷　和範（正会員）

1998 年京都大学工学部情報工学科卒業．2000 年同大学院情報学研究科知能情報学専攻修士課程修了．2002 年同大学院博士後期課程修了．京都大学博士（情報学）．同年京都大学大学院情報学研究科助手．2007 年同助教．2010 年より名古屋大学大学院工学研究科准教授．主に音声対話システムの研究に従事．2008 年から 2009 年まで米国カーネギーメロン大学客員研究員．情報処理学会平成 16 年度山下記念研究賞，FIT2002 ヤングリサーチャー賞等を受賞．電子情報通信学会，言語処理学会，人工知能学会，ACL，ISCA 各会員．

尾形　哲也（正会員）

1993 年早稲田大学理工学部機械工学科卒業．日本学術振興会特別研究員，早稲田大学理工学部助手，理化学研究所脳科学総合研究センター研究員，京都大学大学院情報学研究科講師を経て，2005 年より同助教授（現・准教授）．博士（工学）．JST さきがけ研究「情報環境と人」領域研究員．この間，早稲田大学ヒューマノイド研究所客員准教授，同大学理工学研究所客員准教授，理化学研究所脳科学総合研究センター客員研究員等を兼務．研究分野は人工神経回路モデルおよび人間とロボットのコミュニケーション発達を考えるインタラクション創発システム情報学．2001 年日本機械学会論文賞，IEA/AIE2005，2010 最優秀論文賞等を受賞．日本ロボット学会，日本機会学会，人工知能学会，計測自動制御学会，IEEE 等各会員．

奥乃　　博（正会員）

1972 年東京大学教養学部基礎科学科卒業．日本電信電話公社，NTT，JST，東京理科大学を経て，2001 年より京都大学大学院情報学研究科知能情報学専攻教授．博士（工学）．この間，スタンフォード大学客員研究員，東京大学工学部客員助教授．人工知能，音環境理解，ロボット聴覚，音楽情報処理の研究に従事．1990 年度人工知能学会論文賞，IEA/AIE-2001，2005，2010 最優秀論文賞，IEEE/RSJ IROS-2001，2006 Best Paper Nomination Finalist，IROS-2008 Award for Entertainment Robots and Systems Nomination Finalist 2 件，第 2 回船井情報科学振興賞等受賞．本学会理事，人工知能学会，日本ロボット学会，日本ソフトウェア科学会，ACM，IEEE，AAAI，ASA 等各会員．