

テルミン演奏ロボットのための音高依存性を考慮した音量モデル

○水本武志 †, 辻野広司 ‡, 高橋徹 †, 駒谷和範 †, 尾形哲也 †, 奥乃博 †
† 京都大学大学院情報学研究科 ‡ (株) ホンダリサーチ・インスティテュート・ジャパン

An Arm-Position-to-Volume Model for a Thereminist Robot Robust against Pitch-dependency of the Volume

*Takeshi Mizumoto†, Hiroshi Tsujino‡, Toru Takahashi†,
Kazunori Komatani†, Tetsuya Ogata†, Hiroshi G. Okuno†

† Graduate School of Informatics, Kyoto University ‡ Honda Research Institute Japan Co., Ltd.

Abstract— In this paper, we present an arm-position-to-volume model considering a theremin's pitch dependency of the volume for a thereminist robot. The pitch dependency of the volume raises the problem that even if a robot's volume-control-arm position is fixed, the theremin's volume changes by the pitch change. To solve the problem, we constructed an arm-position-to-volume model which has different model parameters for each pitch-control-arm positions. The learning data for the model calibration is few (about 30 in our case) and parameters are analytically calculated. We experimentally evaluate the model with equal-volume-pitch-control task whose goal is to change the theremin's pitch by keeping the volume stable. The result shows that our method outperforms the fixed-volume-control-arm method.

Key Words: Theremin, Feedforward control, Arm-position-to-volume model, Pitch-dependency

1. はじめに

近年のロボット技術の発展によって、工業用途だけでなくエンターテインメント用途のロボットも開発されてきている。その中でも音楽は一般的なエンターテインメントのひとつであり、音楽と関連したエンターテインメントロボット（以後、音楽ロボットと呼ぶ）が盛んに研究されている [1-3]。特に楽器演奏ロボットは人との合奏によってインタラクティブなパフォーマンスが可能で、有望な音楽ロボットの1つである。楽器演奏ロボットが扱う楽器にはキーボード、フルート、ドラムなどがあるが、本研究では電子楽器のテルミンを採用する。なぜなら、テルミンは次の3つの利点があるからである：(1) 音を出すこと自体は容易、(2) 音高がなめらかに変化するので多彩な表現（微分音など）が可能、(3) 演奏者の手の位置で音を制御するのでロボットに向く。

テルミンの写真を図1に、それを演奏するロボット（HRP-2W改）の写真を図3に示す。テルミンは、2本のアンテナと演奏者の両手のそれぞれの距離を変えて演奏する。本稿では、図1の右側を音高アンテナ、左側を音量アンテナと呼ぶ。また、音高を制御する腕（通常は右腕）を音高腕、音量を制御する腕（通常は左腕）を音量腕と呼ぶ。また、テルミンが同時に発音できるのは1音のみなので、音高と音量の2つが制御対象である。

テルミンの音量・音高制御での問題は次の3点である。

1. 環境キャパシタンスへの鋭敏性：テルミンの音の特性が周囲の環境（気温、観衆の位置など）に強く影響を受ける、
2. 音高・音量腕位置の相互依存性：音高腕の位置が音量に、音量腕の位置が音高に影響を与える、
3. 音量の音高依存性：音量が出力している音の音高によって変化する。

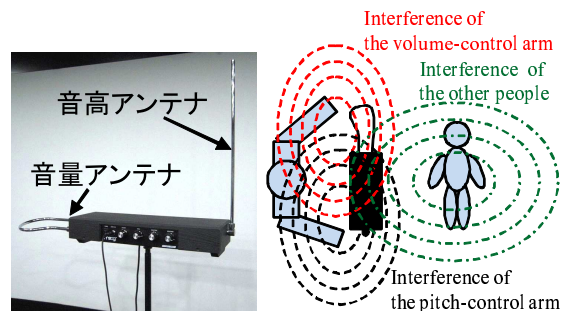


Fig.1 テルミンの写真 Fig.2 相互依存性と環境キャパシタンス干渉

テルミンの音高特性は、問題1が原因で、気温や観衆、演奏者の位置などで変化してしまう。（図 ref:interference）そのため、演奏の際に物理的な基準点が存在せず、テルミンの音を聞きながら演奏する必要がある。音高制御に関しては Frank らによるフィードバック制御 [4] や Alford らによる Lookup-table に基づくフィードフォワード制御 [5]、我々が提案した音高モデルに基づくフィードフォワード制御 [6] などによって解決されてきた。

しかし、音量の制御は上記3つの問題すべてが顕在化するので実現は困難であり、Alford らによる音の ON/OFF と音量腕単独の制御の報告があるのみであった。この方法では、相対的な音の強弱は制御できるが、定量的な音量制御はできない。その結果、休符の表現は可能であるが、クレッシェンドやフォルテといった音高と音量を同時に制御する表現は不十分であった。

そこで本稿では、テルミン演奏ロボットのための音高依存性を含む音量モデルを提案する。本モデルを用いることで、音高と独立な音量の制御が可能になる。



Fig.3 テルミンを演奏する HRP-2W 改

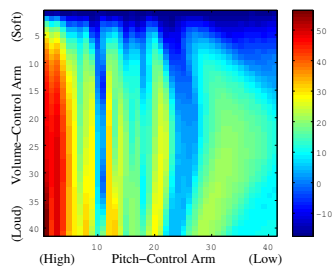


Fig.4 測定した音量マップ

2. 音量制御の課題とアプローチ

本章では、テルミンの音量制御で解くべき課題と本稿で採用したアプローチについて述べる。ただし、本稿では音量を次式で定義する。

$$volume = 20 \log(\text{Var}[x(t)]) \quad (1)$$

ただし、 $x(t)$ は $[-1, 1]$ の範囲に正規化されたテルミンの音響信号、 $\text{Var}[\]$ は分散である。

2.1 音量制御の課題

テルミンの音量を制御する上での課題は、音量腕の位置が変わらなくても、音高が異なれば音量が変化するという音量の音高依存性である。図4に音高腕と音量腕の可動域をそれぞれ40等分し、各位置で収録したテルミンの音量のマップを示す。横軸は音高腕、縦軸は音量腕の位置であり、色はテルミンの音量を表す。もしテルミンの音量に音高依存性が無ければ図の横軸は常に同じ色でなければならないが、音量腕の位置が同じでも、音高腕の位置ごとに音量が大きく異なっている。これは、音量の音高依存性の存在を示している。

2.2 音量制御のアプローチ

従来の音量制御の限界は音高依存性を扱っていないことに起因するので、音高ごとに異なるパラメータを持つモデルを構築することで音高依存性問題を解決する。その際、パラメータの推定に長い時間がかかると実環境で動作させるのは現実的ではない。なぜなら、環境キャパシタンスは頻繁に変化するので、定期的な再推定は必須だからである。よって、本稿で提案するモデルは、必要なデータを少なく、かつパラメータ推定の計算量を小さく抑える必要がある。

そこで音量は対数的に単調減少すると仮定することで、必要なデータを音高モデルの2倍に抑え、かつパラメータ推定の解析解を求めた。

2.3 音量と音高の相互依存性

2.1節で述べた音量の音高依存性とは対照的に、音高の音量依存性は見られなかった。すなわち、音量腕の位置によらず、同じ音高腕の位置ではほぼ同じ音高が出力される。この予備実験の結果から、音量モデルには音高腕と音量腕の位置が必要であるのに対して、音高モデルは音高腕の位置だけで十分である。次章で述べるテルミン演奏システムの概要図(図5)中の、音高モデルへの入力部分に音量推定結果が無いのは、予備実験の結果を反映している。

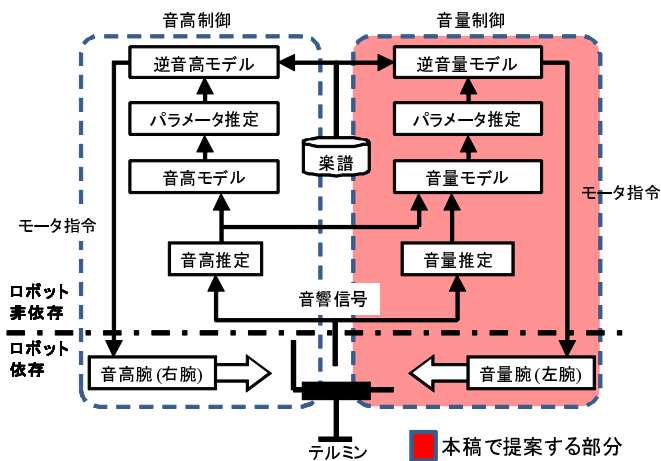


Fig.5 テルミン演奏ロボットシステムの概要

3. テルミン演奏ロボットシステム

本章では、我々が開発したテルミン演奏ロボットシステムについて述べる。まず、システムの概要を述べ、つぎに重要なモジュールである音高制御部、音量制御部について述べる。このうち、本稿で新しく提案するのは音量制御部の音量モデルと逆音量モデルである。

3.1 テルミン演奏ロボットシステム概要

図5にテルミン演奏ロボットシステムの概要を示す。本システムは、音量操作と音高操作の2モジュールから成る。また、本システムは2つのフェーズ: (1) 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、(2) 推定したパラメータを用いて演奏する演奏フェーズで動作する。

キャリブレーションフェーズでは、ロボットは音高腕を、その可動域を $N-1$ 等分した領域の境界で静止させ、それぞれの位置で約1秒間テルミンの音を収録する。この N 回の録音を音量が最小のときと最大のときの2回行う。音高モデルは音量が最大の場合の N 個のテルミンの音高をモデルパラメータ推定のデータとして用いる。一方、音量モデルは、音量が最大と最小の場合の2回、合計 $2N$ 個のテルミンの音量をモデルパラメータ推定のデータとして用いる。演奏フェーズでは、推定されたパラメータを音高・音量の逆モデルに代入し、演奏する楽譜の音高・音量を与えることで、目標とする腕位置 \hat{x}_p, \hat{x}_v を求める。

次に、本システムの特徴について議論する。本システムでは、キャリブレーション後は環境キャパシタンスは変化しないと仮定している。キャリブレーションの時間は現在の実装では $N=15$ で1~2分程度なので、環境の変化が遅ければ、再度キャリブレーションを行うことで対応できる。ただし、動的に環境が大きく変化する場合は対応できない。また、音量腕と音高腕の位置 x_v, x_p は厳密な距離でなくてもよく、距離と比例していればよい。なぜなら、その違いはパラメータで吸収できるからである。実際、前回の報告 [6] では関節角度と手先位置の座標を、今回の報告では等分割した領域の通し番号を用い、正しく動作することを確認した。

3.2 音高モデル

本節では簡単に音高モデルについて述べる (詳細は文献 [6]). 音高腕の可動域を $N-1$ 等分した境界の位置を x_{pi} , その位置での音高を p_i とする ($i = 0, \dots, N-1$). N 個の学習データセット (x_{pi}, p_i) を用いて次の音高モデルのパラメータを推定する.

$$p = M_p(x_p; \theta) = \frac{\theta_2}{(\theta_0 - x_p)^{\theta_1}} + \theta_3, \quad (2)$$

ただし, $M_p(x_p; \theta)$ はテルミンの音高と音高腕の位置の関係を表し, $\theta = (\theta_0, \theta_1, \theta_2, \theta_3)$ はある環境におけるテルミンの音高特性を表すパラメータである.

音高モデルと学習データを用いて Levenberg-Marquardt 法によってパラメータ $\hat{\theta}$ を推定し, 次の逆音高モデルに代入する.

$$\hat{x}_p = M_p^{-1}(p; \hat{\theta}) = \hat{\theta}_0 - \left(\frac{\hat{\theta}_2}{p - \hat{\theta}_3} \right)^{1/\hat{\theta}_1}. \quad (3)$$

演奏する楽譜の音高を逆音高モデルに与えると, キャリブレーションを行った環境における, 音高に対応する腕位置が得られる.

3.3 音量モデル

本節では, 本稿のキーポイントである音量モデルについて述べる. 本モデルは, 図 4 に示した音量マップの観察に基づく次の仮定を用いて構築した.

仮定: テルミンの音量は, x_v が減少するにつれて対数的に単調減少する.

本仮定に基づく音量モデルを次式に示す.

$$v = M_v(x_p, x_v; \mathbf{a}, \mathbf{b}) = a(x_p) \log(b(x_p)x_v) \quad (4)$$

ただし, パラメータ \mathbf{a}, \mathbf{b} は離散的な音高腕の位置 x_{pi} ごとに異なる値をもつので, 音高腕位置が x_p のときのパラメータを $a(x_p), b(x_p)$ と表す. また, v はテルミンの音量である.

本モデルのパラメータ群は解析的に求まる. 仮定より, 本モデルは音高腕の位置が x_p のときの最大音量 $v_{\max}(x_p)$ と最小音量 $v_{\min}(x_p)$ を対数的に結ぶ曲線の集合である. よって, 次の 2 変数の連立方程式を解けばよい.

$$v_{\min}(x_p) = a(x_p) \log(b(x_p) \min(x_v)) \quad (5)$$

$$v_{\max}(x_p) = a(x_p) \log(b(x_p) \max(x_v)) \quad (6)$$

ただし, パラメータは音高腕位置ごとに異なるので, 連立方程式は N 組存在する.

最終的に, $2N$ 個のパラメータ群は次式で求められる.

$$a(x_p) = \frac{v_{\min}(x_p) - v_{\max}(x_p)}{\log(\min(x_v) / \max(x_v))} \quad (7)$$

$$b(x_p) = \frac{\exp\left(\frac{v_{\max}(x_p)}{v_{\min}(x_p) - v_{\max}(x_p)} \log\left(\frac{\min(x_v)}{\max(x_v)}\right)\right)}{\max(x_v)} \quad (8)$$

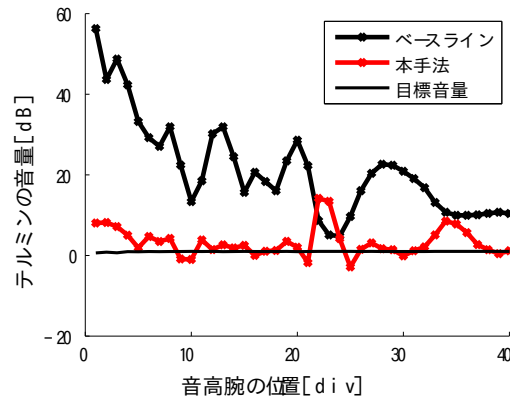


Fig.6 音高腕が音量に与える影響

次に, 逆音量モデルについて述べる. まず, 逆音量モデルを次式に示す. 本逆モデルは, 式 (4) に示した音量モデルの逆関数である.

$$\hat{x}_v = M_v^{-1}(v, x_p; \mathbf{a}, \mathbf{b}) = \frac{1}{b(x_p)} \exp\left(\frac{v}{a(x_p)}\right) \quad (9)$$

本音量モデルを用いた音量制御は, 音高腕の位置 x_p とモデルパラメータ \mathbf{a}, \mathbf{b} , そして目標とする音量 v を与えて音量腕位置 \hat{x}_v を逆算することで実現する. なお, \mathbf{a}, \mathbf{b} はキャリブレーションフェーズで用いた x_p の離散的な位置ごとに存在する. x_p が離散点の間にある場合は, 最近傍の 2 点を用いて音量腕の位置を推定し, それらを線形に補完することで \hat{x}_v を求める.

4. 音量モデルの評価実験

4.1 実験タスク

音量の定量的な制御の効果を評価するため, 一定の音量を保ったまま音高を変化させるタスク (等音量演奏タスク) を設定した. 等しい音量で演奏できれば, 強調したい場面では音量を上げるなど, 音高に独立な音量の操作が可能になる. タスクの例として実験結果の一部を図 6 に示す. ベースライン手法として音量腕を固定した場合との性能を比較した. この場合, 本手法の方が音高腕の位置によらず音量の変化を小さく抑えられているので, 本手法の方が性能が高い.

本タスクの評価は, 音高腕の位置を変化させながら各位置での音量を測定し, その平均と分散をプロットすることで行った. 音量の平均を高く保ったまま分散を低く抑えられれば, 音量制御によって音量を一定に保てたと言える.

4.2 実験条件

図 7 に実験の配置を示す. ロボットは HRP-2W 改を用い, テルミンは Moog Music 社の Etherwave Theremin を用いた. ロボットとテルミン, ロボットとスピーカの間隔はそれぞれ 50 [cm], 150 [cm] である. マイクはロボットの前頭部に 1 つ装着した. ロボットは左腕で音量を, 右腕で音高をそれぞれ 1 自由度を使って制御する. 使用した関節角は, 左腕は肩ピッチ軸, 右腕は肩ロール軸で, 可動域はそれぞれ $-90 \sim -135$ [deg], -90 [deg] から 0 [deg] である.

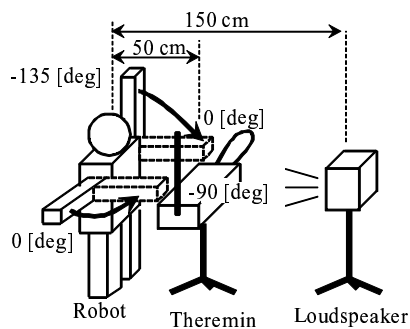


Fig.7 実験配置

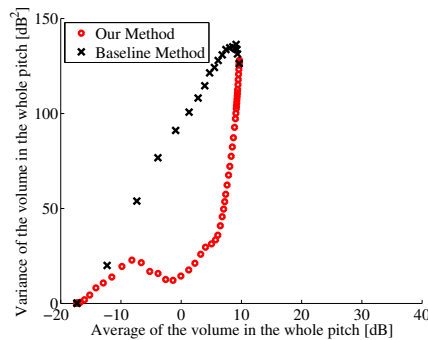


Fig.8 音量最小の条件下での音量変動

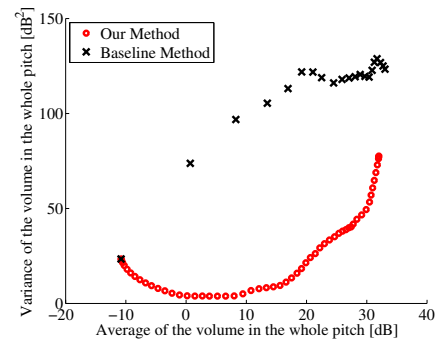


Fig.9 音量最大条件下での音量変動

両腕の可動域はそれぞれ 20 等分し ($N = 21$), 各点でテルミンの音をサンプリングレート 48kHz で 1 秒ずつ録音した。テルミンには音量を設定するつまみがあるので、このつまみを変化させることで条件を変えた。つまみは、上方向を 0 [deg] として 30 [deg] ずつ時計周りに回転させた。条件は 0, 30, 90, 120, 150 [deg] の 5 条件とした。なお、0 [deg] が音量最小、150 [deg] が音量最大の条件である。

4.3 実験手順

テルミンの音量は両腕位置の全組み合わせ 441 通り ($N^2 = 441$) で録音し、事前に計算しておく。そして、目標音量と音高腕から逆音量モデルで逆算した音量腕位置 \hat{x}_v に対応する測定音量を、本手法で制御した音量とした。 \hat{x}_v が測定した点の間であれば、最近傍の 2 点を使った線形補完で音量を求めた。また、目標音量 v は -20 [dB] から -40 [dB] の間で 1 [dB] ずつ変化させた。

4.4 実験結果

音量最小の場合と最大の場合の結果を図 8 と図 9 に示す。横軸は音量の平均、縦軸は音量の分散を表す。音量の平均によらず分散を低く抑えられれば音量を一定に保てているので、図の下側に点があれば性能が高いといえる。図に示すとおり、すべての条件で本手法が分散を低く抑えられた。

本手法、ベースライン手法ともに平均音量が低いときは分散が低い。これは平均音量が十分低いときはテルミンの音が消えているからである。逆に平均音量が高いときは両手法ともに分散が高い。これは、目標音量が $v_{\max}(x_p)$ を超えているために音量制御が不能になったのが原因である。さらに、本手法の分散は平均音量が -10 [dB] の付近で分散が大きくなっている。これは他の条件でも見られたが、音量モデルと実際の音量マップとの誤差が大きい領域だからである。

次に、本モデルの限界を議論する。平均音量をあげるにつれて、本モデルによる音量制御を使っても分散音量が増加している。この原因は、音量は対数的に単調減少する仮定が成り立たなくなるからである。実際に測定した音量とモデルから推定した音量との比較を図 10 に示す、図より、特に低音域でピークがカーブしているため、本仮定が成立していない。このカーブは、音量が小さい場合顕著に現れるため、図 8 では分散が大きい。

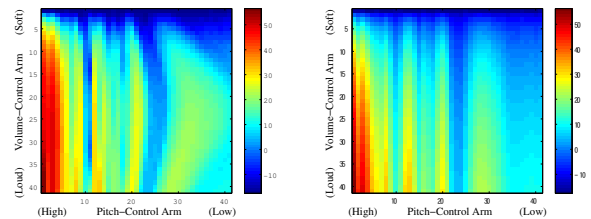


Fig.10 音量マップの測定値とモデルの比較: 左が測定値, 右がモデルに基づく推定値. 左側は Fig. 4 と同様.

5. まとめ

本研究では、テルミン演奏ロボットのための音高依存性を考慮した音量モデルについて述べた。テルミンの音量は音高に依存する性質があるが、従来研究では音量は ON/OFF や単純な位置制御などの、依存性を考慮しない制御しかされていなかった。本研究では、音高ごとに異なるパラメータで音量をモデル化することで音量の定量的な制御を実現した。実験の結果、本モデルによる音量制御によって音量の変化を押さえながら音高を変化させられることを示した。

ただし、本モデルは実際の音量特性との差が特に低音域で顕著である。今後は、モデルの精緻化による正確な音量の推定と、モデルパラメータのオンライン推定を行う予定である。

謝辞 本研究の一部は、科研費 (S) と GCOE の援助を受けた。

参考文献

- [1] J. Solis *et al.* Development of Waseda flutist robot WF-4RIV: Implementation of auditory feedback system. In *ICRA08*, pp. 3654–3659.
- [2] K. Yoshii *et al.* A biped robot that keeps steps in time with musical beats while listening to music with its own ears. In *IROS07*, pp. 1743–1750.
- [3] G. Weinberg *et al.* The interactive robotic percussionist - new developments in form, mechanics, perception and interaction design. In *HRI07*, pp. 456–461.
- [4] F. van der Hulst. Robotic theremin player. In *17th NACCQ*, p. 534, 2004.
- [5] A. Alford *et al.* A music playing robot. In *FSR99*, pp. 29–31.
- [6] T. Mizumoto *et al.* Thereminist robot: Development of a robot theremin player with feedforward and feedback arm control based on a theremin's pitch model. In *IROS09*, to appear.