

テルミン演奏ロボットのための Unscented Kalman Filter による適応的音高制御

水本 武志, 尾形 哲也, 奥乃 博 (京都大学大学院 情報学研究科)

1. はじめに

一緒に合奏したり、同じ音楽に合わせて手を叩くといった音楽を通じたインタラクションによって、世代、文化、話す言葉が異なった人々の間で楽しさを共有することが可能である。同様に、ロボットが人と合奏ができるようになると、人とロボットが共に楽しめる参加型エンタテインメントロボットが実現すると期待できる。我々は共演者ロボットとして複数ロボット上にテルミン演奏ロボット (図 1) を開発し [1], フルート [2] や、ギター [3] との合奏を実現してきた。共演者ロボットは近年盛んに研究されており、フルート演奏ロボットと人のサクソとの合奏 [4] や 2 体のロボット (太鼓・木琴) と 2 人の人 (太鼓・キーボード) 即興合奏 [5] なども報告されている。

本稿では、テルミン演奏ロボットのための適応的な音高制御手法について報告する。テルミンは演奏者の手とアンテナとが構成するキャパシタの容量変化で音高・音量を制御するので、共演者の位置、観客の有無などの周囲の環境に影響を受けやすい。従来は演奏を校正フェーズと演奏フェーズに分け、校正を定期的に行うことで本問題を回避していた [1]。しかし、合奏では演奏中に共演者が動くので、適応的な音高制御は不可欠である。本稿で扱うような、演奏中の楽器特性の時間変化はテルミンに限らず一般的に生じうる。例えば、弦楽器は演奏中に弦の張力が変化するために、管楽器は楽器内の気温が変化するために、音高が変化する。

適応的な音高制御を実現するためには、非線形な楽器特性 (制御入力と出力音高の関係) の頑健な実時間推定が必要である。例えば、Extended Kalman Filter (EKF)、腕位置と音高を蓄積して定期的に音高特性のモデルパラメータを再推定する方法 (区間毎再推定法) が考えられる。前者は、特性が非線形であることから推定精度が低くなり、後者は共演者が演奏中に動くなどの連続的な変動のために精度が低くなる。

我々は、非線形システムをテイラー展開の 2 次の精度で推定できる Unscented Kalman Filter (UKF)[6] を用いて、適応的な音高制御を実現する。UKF を用いることで、テルミンの非線形な音高特性モデル [1] を観測モデルとする状態推定を精度良く実現する。

2. テルミン演奏システム

2.1 音高モデル

$p \in \mathbb{R}$ をテルミンの音高, $\theta = (\theta_0, \theta_1, \theta_2, \theta_3) \in \mathbb{R}^4$ を音高モデルのパラメータ, $x_p \in [0, 1]$ を抽象化されたロボットの腕位置と定義する。ここで, $x_p = 1$ と $x_p = 0$ はそれぞれアンテナに最も近い位置, 最も遠い位置とする。音高モデル M_p は次式で定義される [1]:

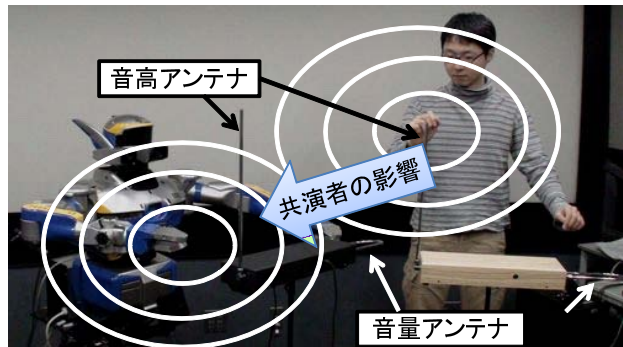


図 1 テルミンと人との合奏: テルミンの音高はロボットの左手とアンテナとの距離だけでなく、人の手の位置にも影響を受ける。

$$p = M_p(x_p; \theta) = \frac{\theta_2}{(\theta_0 - x_p)^{\theta_1}} + \theta_3, \text{ and} \quad (1)$$

$$x_p = M_p^{-1}(p; \theta) = \theta_0 - \left(\frac{\theta_2}{p - \theta_3} \right)^{1/\theta_1} \quad (2)$$

ただし、逆モデル M_p^{-1} は M_p から解析的に導出した。

抽象化した腕位置 x_p で音高モデルと制御手法を構築することで、目標腕位置の決定問題と、ロボット依存の関節角の制御問題を分離できる。この結果、個別のロボットに依存しない演奏システムの構築が可能となる。

2.2 従来の静的な音高制御手法

静的な制御手法は 2 つのフェーズ、校正フェーズと演奏フェーズから構成される。校正フェーズでは、腕の可動域のうち $L+1$ 点で音高を測定し、腕位置 $x_p^{(i)}$ とその位置でのテルミンの音高 $p^{(i)}$ とのペア ($i \in \{0, \dots, L\}$) を学習データとして、モデルパラメータ $\hat{\theta}$ を推定する。可動域を L 等分して観測すると仮定すると、

$$x_p^{(i)} = i/L \quad (3)$$

$$p^{(i)} = M_p(x_p^{(i)}; \theta_T) \quad (4)$$

ただし、 θ_T は真のモデルパラメータとする。こうして観測したペアから、次式のコスト関数を最小化するパラメータ $\hat{\theta}$ を Levenberg-Marquardt 法 [7] で求める。

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=0}^L \|p^{(i)} - M_p(x_p^{(i)}; \theta)\|^2 \quad (5)$$

なお、校正フェーズは、 $L = 14$ の場合、Lenovo ThinkPad T60p 上で 90 秒程度で終了する。

次に、演奏フェーズでは、与えられた楽譜と $\hat{\theta}$ から腕位置の軌跡を求める。楽譜は、目標音高 $s^{(i)}$ と音符長 $d^{(i)}$ のペアの系列と定義する。音符長 $d^{(i)}$ に従って

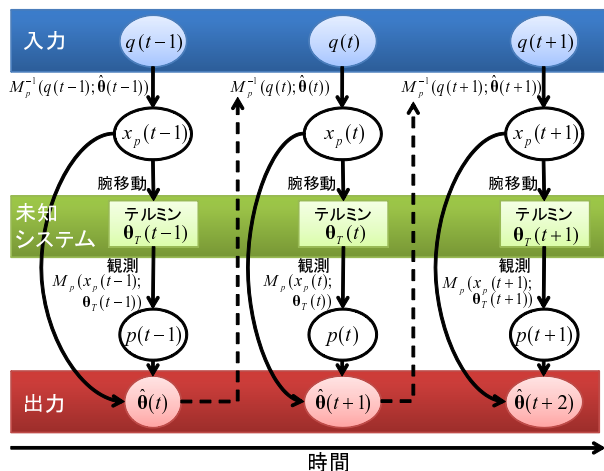


図2 データフロー: パラメータ $\hat{\theta}(t)$ は観測した音高 $p(t)$ と目標音高 $q(t)$ の2乗誤差が最小になるように逐次的に推定される

音高 $s^{(i)}$ を接続することによって、目標音高系列 $q(t)$ が求まる。 $q(t)$ と $\hat{\theta}$ を用いて、ロボットの腕位置系列が次式で求まる:

$$\hat{x}_p(t) = M_p^{-1}(q(t); \hat{\theta}) \quad (6)$$

従来のテルミン演奏手法では、 θ_T が時不変な静的環境を仮定したり [1, 8]、音高を対数軸上で表して線形近似 [9] していた。前者は楽器特性の時間変化を無視しており、後者は近似誤差のために本モデルよりも音高変化が大きい。したがって、モデルパラメータ θ_T の時間変化を許容すれば、高精度な音高制御が実現できよう。

3. UKF による適応的音高制御

3.1 問題設定

まず、本稿で解決する問題の設定について述べる。

入力: 目標音高系列 $q(t)$
 初期パラメータ $\theta(0)$
 出力: 推定パラメータ系列 $\hat{\theta}(t)$
 腕位置系列 $\hat{x}_p(t) = M_p^{-1}(q(t); \hat{\theta}(t))$
 仮定: 真パラメータ系列 $\theta_T(t)$ はランダムウォーク

各時刻 t で、目標音高 $q(t-1)$ と観測したテルミンの音高 $p(t-1)$ からモデルパラメータ $\hat{\theta}(t)$ を推定する。次に、推定パラメータを式 (2) に代入して求めた腕位置 $\hat{x}_p(t)$ に腕を移動させる。パラメータの変化は、共演者の動きや内部回路の状態などの要素が混合するので、ランダムウォークと仮定した。

3.2 UKF による適応的音高制御

UKF は Julier と Uhlman が提案した Unscented 変換に基づく Kalman Filter [6] で、パラメータ推定 [10] や Visual SLAM [11] などに応用されている。

本節では、まず Unscented 変換について述べ、その Kalman Filter への拡張について述べる。最後に UKF の適応的音高制御への応用について述べる。

3.2.1 Unscented 変換 (UT)

UT とは単峰性の確率分布に従う確率変数に任意の非線形変換を施した後の確率分布のモーメントを求める手法である。従来は次の2つの方法で求めていた:

1. 非線形変換をテイラー展開して線形近似する,
2. 変換前分布から確率的にサンプルを引き、それらに非線形変換を施した後にサンプルのモーメントを求める。

前者は非線形性が強い、すなわちテイラー展開の2次以上の項の影響が強い場合に誤差が大きくなり、後者は信頼できる推定を行うには多数のサンプルを引く必要がある。UT は (2) の方法に分類されるが、変換前分布のモーメントが既知であると仮定することで、決定的なサンプリングを行う。UT で決定的に求めるサンプルはシグマポイントと呼ばれる。従って、2つの手法の利点、(1) 決定的に推定可能、(2) 非線形変換を近似せずにモーメントが推定可能、を享受できる。

変換前分布の確率変数を x 、その次元を D 、既知の平均と分散をそれぞれ \bar{x} と Σ_x とする。これらから $2D+1$ 個のシグマポイント $\chi^{(0)}, \dots, \chi^{(2D)}$ が次のように決定的に求まる:

$$\text{for } i = 1, \dots, D \quad \chi^{(0)} = \bar{x} \quad (7)$$

$$\chi^{(i)} = \bar{x} + (\sqrt{D\Sigma_x})_i \quad (8)$$

$$\chi^{(i+D)} = \bar{x} - (\sqrt{D\Sigma_x})_i \quad (9)$$

ただし $\sqrt{\cdot}$ は行列の平方根、 $(\cdot)_i$ は行列の i 番目の列とする。行列 M の平方根 A は $M = AA^H$ と定義され、コレスキー分解で求める。次に、各シグマポイントに対応する重み $w^{(0)}, \dots, w^{(2D)}$ を次式で求める:

$$w^{(i)} = \begin{cases} \kappa/(D+\kappa) & \text{if } i=0 \\ 1/(2(D+\kappa)) & \text{otherwise} \end{cases} \quad (10)$$

ただし κ はスケールパラメータであり、変換前分布が正規分布なら $\kappa=2$ のとき近似誤差が最小になる。

上記のシグマポイントと重みから、変換後分布の平均と分散 \bar{z} と Σ_z を求める。まず、シグマポイントに非線形変換 f を施し、変換後のシグマポイント $Z^{(0)}, \dots, Z^{(2D+1)}$ を得る:

$$Z^{(i)} = f(\chi^{(i)}), \quad (11)$$

これらの重み付き和が平均と分散の推定値となる:

$$\bar{z} = \sum_{i=0}^{2D} w^{(i)} Z^{(i)} \quad (12)$$

$$\Sigma_z = \sum_{i=0}^{2D} w^{(i)} (Z^{(i)} - \bar{z})(Z^{(i)} - \bar{z})^T \quad (13)$$

3.2.2 UT の Kalman Filter への適用

Kalman Filter [12] は、状態遷移モデル f と観測モデル h を既知とし、雑音を含む入力から未知の状態系列を逐次的に推定する手法である。状態遷移方程式と観

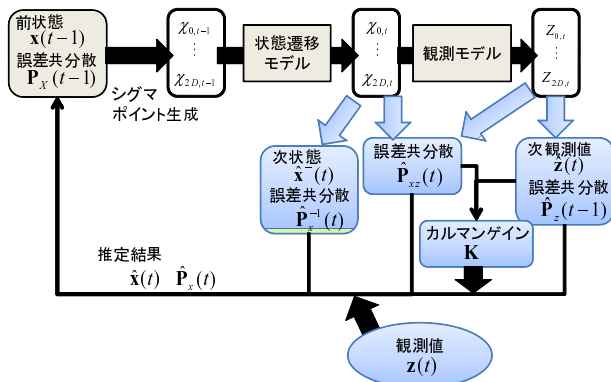


図3 UKFの概略図: UTを用いて, 現状態から次状態と次観測値を推定し, 観測値を用いて推定値を更新
測方程式は以下のとおりである:

$$\mathbf{x}(t+1) = f(\mathbf{x}(t), \mathbf{v}(t)) \quad (14)$$

$$\mathbf{z}(t) = h(\mathbf{x}(t), \mathbf{w}(t)) \quad (15)$$

ただし, t は時刻, \mathbf{v} と \mathbf{w} はそれぞれ状態遷移ノイズと観測ノイズを表す. 本来の Kalman Filter は f, h が線形関数だと仮定しており, それを非線形に拡張した Extended Kalman Filter は f, h を一次近似することで Kalman Filter に帰着させていた.

一方, UTを用いると, Kalman Filter で必要な (1) 現状態を既知とした次状態とその誤差共分散の推定と, (2) 次状態を既知とした次観測値とその誤差共分散の推定が2次の精度 [13] で行える. 従って, UTを用いることで非線形な f と h を持つシステムに対しても精度高く状態系列を推定できる. UKF のデータフローを図3に示す.

3.3 UKFの音高制御への応用

3.3.1 モデル設計

状態遷移, 観測モデルを次のように設計した:

状態遷移モデル:

$$\begin{aligned} \boldsymbol{\theta}(t) &= f(\boldsymbol{\theta}(t-1), \mathbf{v}(t-1)) \\ &= \boldsymbol{\theta}(t-1) + \mathbf{v}(t-1) \end{aligned} \quad (16)$$

$$\theta_0(t) = \max(\theta_0(t), 1 + \epsilon) \quad (17)$$

$$\theta_1(t) = \max(\theta_1(t), 0) \quad (18)$$

観測モデル:

$$\begin{aligned} p(t) &= h(\boldsymbol{\theta}(t), w(t)) \\ &= M_p(x_p(t); \boldsymbol{\theta}(t)) + w(t) \end{aligned} \quad (19)$$

状態遷移モデルは問題設定で仮定したとおりランダムウォークとした. ただし, 式 (17), (18) でパラメータが無効な値にならない範囲に写像している. UTでは微分を行わないので, このような微分不可能な式を制約に加えることが可能である. また, 観測モデルは式 (1) に観測ノイズを加えたものである.

3.3.2 ロボットの制約

物理的な制約からロボットの腕の変化速度は有限である. 従って, x_{plim} を腕の単位時間における腕位置の最大変化量と定義し, 次の条件:

$$x_{plim} > |x_p(t) - x_p(t+1)| \quad (20)$$

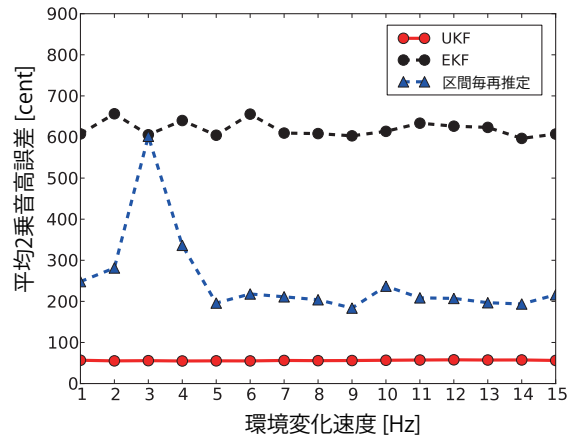


図4 音高誤差: 縦軸は音高の2乗誤差, 横軸は式 (22) で定義した環境変動の速度 ω を表す.

が真なら x_p の変化量を x_{plim} に制限することでロボットの腕の変化が x_{plim} を越えないことを保証する. x_{plim} は実験的に 0.05 に定めた. この値はヒューノイドロボット HRP-2 を用いて定めたので, ロボットが小型であればより大きく設定する.

4. 実験

4.1 シミュレーションによる実験

環境変動をシミュレートするため, 環境パラメータ $e(t) \in [0, 1]$ を定義する. $e(t)$ に基づいて $N+1$ 個の実測モデルパラメータ $\theta^{(0)}, \dots, \theta^{(N)}$ を次式のとおり線形補間してパラメータの時系列を生成し, 評価に用いる.

$$\begin{aligned} \boldsymbol{\theta}(t) &= (Ne(t) - 1)\boldsymbol{\theta}^{(i+1)} + (1 - Ne(t) + i)\boldsymbol{\theta}^{(i)} \\ &\quad (\text{if } 1/N \leq e(t) \leq (i+1)/N). \end{aligned} \quad (21)$$

具体的には, 環境パラメータは \sin 関数に従うとし, その周波数 $\omega \in \{1, \dots, 15\}$ を変動することで環境の変化速度を定める. 形式的には次のように表す:

$$e(t) = 0.5 \sin(2\pi\omega t) + 0.5 \quad (22)$$

環境が周期的に変化するという状況は, 共演者がリズムに併せて体を揺らしている場合などを想定している.

実験では次の3手法: UKF (本手法), Extended Kalman Filter (EKF), 区間毎再推定法を比較する. UKFの初期値は $\Sigma_x = \Sigma_v = \text{diag}(5, 5, 5, 5)$, $\Sigma_w = 10$, $\kappa = 2$ とした. EKFはUKFと同様の初期値とし, 式 (1) のヤコビアンで非線形な観測モデルを一次近似した. 区間毎再推定法は, 腕位置と音高を蓄積し, 5秒毎にパラメータを更新した.

評価尺度は $1200 \log(p/q)$ と定義する. ただし p, q をそれぞれ観測音高, 目標音高とする. 本尺度は cent と呼ばれ, 100 cent の差が半音の差に対応する. $\omega \in \{1, \dots, 15\}$ のそれぞれの条件で, 10回ずつ Aura Lee (楽譜は [14] 参照) を演奏した.

図4に各条件の音高誤差を示す. 青破線が区間毎再推定法, 黒破線がEKF, 赤実線が本手法の誤差である. 図からわかるとおり, 本手法では環境変動によらずに誤差が半音以内に抑えられているのに対して, EKF, 区間毎再推定法では誤差が大きい.

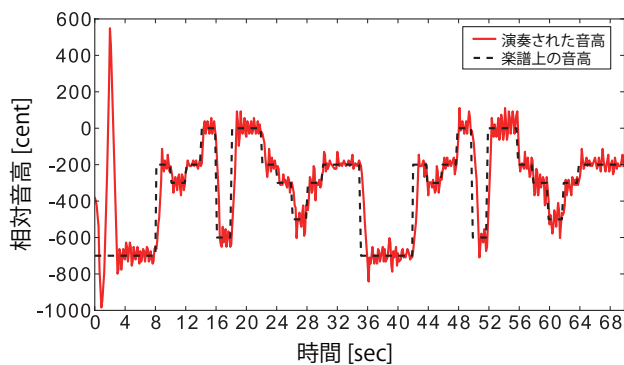


図5 テルミンの音高軌跡: 縦軸は音高 [cent], 横軸は時間を表す. 黒線は目標音高系列 $q(t)$, 赤線は実際に測定した音高軌跡を表す.

EKF で誤差が大きいのは, 状態遷移, 観測モデルがどちらも非線形であるために生じた近似誤差が原因と考えられる. 一方, 区間毎再推定法は EKF よりは低い誤差に抑えられている. これは式 (5) の非線形コスト関数を最適化しているからである. ただし, 逐次推定ではないので処理時間は長い. $\omega = 3$ の付近で誤差が急激に増加しているのは, 推定に用いる区間と環境変化の周期が逆相同期同期したために, 推定パラメータと実際のパラメータが乖離したことが原因と考えられる.

4.2 実ロボットによる実験

本手法をヒューマノイドロボット HRP-2 に実装し, 評価を行った. ロボットの腕の制御周期は 62.5 msec とし, 音高推定は従来のテルミン演奏システムと同様に自己相関に基づく方法とした.

音高軌跡を図5に示す. 縦軸は A3 (220 Hz) を 0 とする相対音高で示している. 図に示すとおり, 環境が変動しているにも関わらず正しい音高で演奏できている. 次に図6に音高制御誤差を示す. 誤差の絶対平均値が 72.9 [cent] であったことから, ほとんどの場合で半音以下の制御誤差に抑えられている. 0-4 秒付近の音高誤差は初期値を実際と変えたことに起因しており, 4 秒程度で正しい音高に収束していることから, 適応的な制御が実現できている. 8 秒や 16 秒の音高誤差が高い部分は楽譜上の音符が変化したことに起因している. 音符変化に追従した後も振動しているの原因は, (1) 腕位置系列 $x_p(t)$ と推定した音高系列 $p(t)$ との同期ずれ, (2) ロボット自体が振動と考えられる.

5. おわりに

本稿では, 環境変化に頑健なテルミンの適応的音高制御手法について報告した. テルミンの音高特性は周囲の環境変化に敏感に変化するので, 共演者の動きなどの影響は不可避である. そこで, UKF を用いて音高特性の変化を推定する手法を開発し, 楽譜どおりの音高で演奏し続けるテルミン演奏ロボットを実現した. 今後の課題は, ロボット自体の振動や共演者の動きなどによるパラメータ変化のダイナミクスを含むモデルに拡張すること, 複数ロボットへ実装してシステムの可搬性を評価することである.

謝辞 科研費 (特別研究員奨励金, S, 挑戦的萌芽, 新学術), HRI-JP, GCOE の援助を受けた.

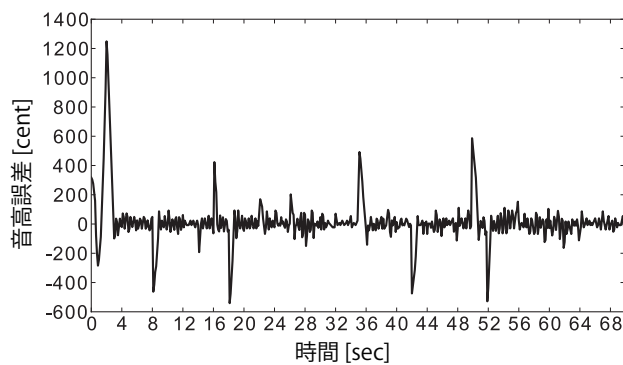


図6 音高誤差: 縦軸は音高制御誤差 [cent], 横軸は時間を表す. 急激に誤差が増加している 8, 16 秒などは音符の切り替わりに対応している.

参考文献

- [1] 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. *Proc. IROS*, pp. 2297–2302, 2009.
- [2] A. Lim *et al.* Robot musical accompaniment: Integrating audio and visual cues for real-time synchronization with a human flutist. *Proc. IROS*, pp. 1964–1969, 2010.
- [3] T. Itohara *et al.* Particle-filter based audio-visual beat-tracking for music robot ensemble with human guitarist. *Proc. IROS*, 2011 to appear.
- [4] K. Petersen *et al.* Musical-based interaction system for the Waseda Flutist Robot: implementation of the visual tracking interaction module. *Autonomous Robots J.*, 28(4):439–455, 2010.
- [5] G. Weinberg *et al.* The creation of a multi-human, multi-robot interactive jam session. *Proc. NIME*, pp. 70–73, 2009.
- [6] S. J. Julier and J. K. Uhlmann. A new extension of the kalman filter to nonlinear systems. *Proc. of SPIE*, pp. 182–193, 1997.
- [7] D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. on Appl Math.*, 11(2):431–441, 1963.
- [8] A. Alford *et al.* A music playing robot. *Proc. FSR*, pp. 29–31, 1999.
- [9] Y. Wu *et al.* Towards anthropomorphic robot thereminist. *Proc. ROBIO*, pp. 235–240, 2010.
- [10] E. A. Wan and R. van der Merwe. The unscented kalman filter for nonlinear estimation. *Proc. AS-SPCC*, pp. 153–158, 2000.
- [11] S. Holmes *et al.* A square root unscented kalman filter for visual monoslam. *Proc. ICRA*, pp. 3710–3716, 2008.
- [12] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. of the ASME-J. of Basic Engineering*, 82(Series D):35–45, 1960.
- [13] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proc. the IEEE*, 92(3):401–422, 2004.
- [14] 水本武志 *et al.* 打楽器とロボットとの合奏のための結合振動子モデルに基づく打撃時刻予測. 日本ロボット学会学術講演会, pp. 1H3–2, 2010.