論文

マルチモーダル指揮シミュレータ[†]

宇佐 聡史^{*1 *2} 持田 康典^{*1}

近年、音楽へのソフトコンピューティング応用例が増えている。柔軟な処理が音楽情報処理に適していると考えられる。本稿ではマルチモーダル指揮シミュレータ試作への応用を報告する。人の指揮を認識し実時間にオーケストラのように振る舞う。指揮動作の特徴や時間構造は前後の脈絡や人により揺らぐ。それを吸収して指揮の意図を抽出する為に HMM(隠れマルコフモデル)とファジィによる柔軟な処理を導入し、拍認識処理を人の認識に倣って構築した。右手指揮動作を加速度センサで、視点を視線検出器で、呼吸を呼吸センサで検出する。狭義の動作認識に HMM を適用。楽員(オーケストラ奏者)や指揮者へのインタビューで得た知見に基づき、曲の知識や演奏情報の活用をファジィ推論で実現した。指揮法に則った普遍性ある重要な要素、即ち、曲の演奏開始(出、予備動作)や停止、目によるアインザッツ(Einsatz, 奏者への cue=出の指示)、拍タイミング(アゴーギク=テンポ変動を含む)、何番目の拍か、フェルマータ(任意の長さに音符や休符を延ばす)、デュナーミク(音の強弱)、一部のアーティキュレーション(奏法)、呼吸法を認識する。これに基づき実時間に音楽演奏を制御する。正しい呼吸法や指揮者の重要な仕事の1つであるアインザッツに対するオーケストラの反応、発音時間遅れや自律性もシミュレートした。自律度最大では指揮入力を無視して曲のテンポデータに基づいて演奏し、自律度最低では指揮に密着して動作停止と同時に演奏停止し、次の動作入力まで待ち続ける。

キーワード: ヒューマン・インターフェイス, マルチモーダル, 動作認識, HMM, ファジィ, 音楽情報処理, 指揮, MIDI

1. はじめに

音楽には作曲、編曲、解釈、演奏、鑑賞などの局面があり、いずれも人ならではの知的処理である。しかも人によって異なる柔軟な処理であり、その違いが個性として重要視される。音楽への計算機応用は1956年のイリアック組曲の自動作曲以降、これら局面毎に多くの研究がなされてきた[1,2]。近年ではソフトコンピューティングの応用例も増えている。ニューラルネットでは合奏音からの音高同定[3]、リズム認知[4]、楽音合成[5]、演奏者の特徴抽出[6]、ファジィで自動伴奏生成[7]、エージェントで音楽信号からのビート認識[8]、人と競演するシステム試作[9]、GAでテンポ解析[10]、カオスの応用[11]などの例がある。

著者らは指揮認識への応用を研究してきた[12,13]。同じ意図の指揮動作でも、音楽の脈絡や人により動作の特徴パラメータや時間構造が揺らぐ。音声や手書き文字等の認識と同様、このことが機械認識を困難にしている。この揺らぎを吸収して指揮の意図を認識するのにソフトコンピューティング技法を適用した。指揮法[15]では左手が無くてもオーケストラを完全に誘導

可能とされるので、右手指揮動作を加速度センサで採取。重要な指揮要素をパーソナル・コンピュータで認識。それに基づいて実時間に音楽演奏を制御した。特徴パラメータの揺らぎと時間構造の揺らぎを両方一度に吸収できる HMM を狭義の動作認識に適用。更に音楽家インタビューで得た知見に基づき、曲の知識や演奏情報の活用をファジィ推論で実現した。HMM は音声認識分野で高い認識性能のために標準技術となりニューラルネットの研究を下火にした技術である[16]。その特長が音声認識と同様に指揮認識にも有効なことが確認された。また、実際のオーケストラに特有の、指揮者の拍打動作から発音までの時間遅延もシミュレートした。この遅延は指揮学習の初期に問題となるため、初心者の方の指揮練習に有効と考えられる[13]。

本稿ではマルチモーダル認識シミュレータへの展開も報告する。視線検出器と呼吸センサの併用により、アインザッツや呼吸法も検出、認識する。アインザッツや呼吸タイミングが予め指定したマークと異なると、演奏が中断したり該当パートが一定時間欠落する。これによりユーザは指揮者のアインザッツのミスや正しい呼吸法の体験や練習ができる。曲データ内に自律性の程度を任意に設定することでオーケストラの自律性もシミュレートした。

指揮は動作の1種である。動作認識へのファジィ[17-19]、ニューラルネット[20]、HMM[21-23]の応用例があるが、これらは認識対象区間の切出しが自動でなく実時間音楽演奏には適用できない。また、画像処理

[†] A Multi-Modal Conducting Simulator
Satoshi USA and Yasunori MOCHIDA

^{*1} 工学院大学 大学院情報学専攻
Graduate Course in Informatics, Kogakuin University

^{*2} ヤマハ(株) 楽器研究所
YAMAHA Corp. Musical Instrument Research Laboratories

なので動作領域や背景，着衣に制限がある場合があった。指揮に限った認識も1988年頃から多種多様なセンサで研究されてきた[12-14]，しかし画像処理では撮像フレーム周波数に伴う時間分解能の限界のため拍点認識に0.1秒もの遅れがある[24]，フラッシュで誤動作するなど照明の影響を受け易いといった問題がある。また，データグローブはテンポ制御に使うには応答が遅い上に高価である。そこで近年では加速度センサの応用例が多い。発音タイミングと音量の制御だけでなく，ニューラルネットでジェスチャー10種を弁別して制御コマンドや音色に対応させたり[24]，加速度の他に指揮棒の向きや握る圧力を検出して様々なパラメータを制御した例[25]がある。また，新規な3次元位置検出装置を指揮システムに適用して大きな違和感は無かったとの報告[26]もある。しかし本研究のように指揮法に則った多くの要素を右手動作から認識したり，アインザッツや呼吸法，オーケストラの発音時間遅延，自律性までシミュレートした例は無い。

体の動きで電子音源を制御する様々な研究もされてきたが[1,2]，指揮の技法は長い歴史の推敲を経て指揮法として系統立てられ，且つ世界共通である。従って本稿では指揮法[15]の基礎に則る。

2. 指揮について

指揮法に規定されている，或いは音大指揮科で教える基本的で普遍性のある右手指揮要素の一部を表1に

示す。人のコミュニケーションで交わされるメッセージの9割以上がノンバーバルとの報告も有るが[27]，指揮者は本番では一言も発せずに，両手，目，呼吸，体，顔など異なる部位の動作を複数独立に，必要により併用して意図を大勢の楽員に伝達する。典型的なマルチモーダル・コミュニケーションと言える[12]。

指揮システム構築の問題要素を表1順に列挙する。

明瞭で揃った曲の演奏開始(出)の指示と同時に，冒頭のテンポやデューナーミクを指示するため，指揮者は演奏開始前に1拍余分に振る。これを予備運動，この拍を予備拍(extra beat)と呼ぶ。管楽器のプレスに相当するが，指揮者も軽くプレスすると良いとされる。

アインザッツは指揮者の重要な仕事の1つで，演奏開始する奏者に向けて1拍前，速いテンポでは2拍前に，目，左手，体の向き，指揮棒などで合図する。規模の大きなオペラなどでは距離の離れた歌手にも見えるよう手で指示するが多いが，オーケストラでは頻繁に手や棒で指示されると楽員がうるさく感じ，目で見ずに手だけで指示しても印象悪い上に伝わりにくい。従って目で見るのが最良の方法とされる。

右手指揮動作の軌跡図形例を図1[15]に示す。3つ振りのエスプレシボ(表情豊かに)とスタカート(音を明確に分離して)の例である。

指揮の拍打，即ち指揮棒の加速度ローカルピーク時点から，実際のオーケストラの発音までには時間遅延がある。この遅延は指揮者と楽員の間の暗黙且つ共有

表1 指揮の要素と動作 (主に右手による普遍性の高い重要な要素)

指揮要素	動作	本手法
タイミング 出(曲の演奏開始)	予備位置から1拍前(予備拍)をin tempo(テンポ一定)で振り始めテンポと強さを指示。管楽器のプレスに相当。指揮者も軽くプレスをするが良い。	○
アインザッツ	入り(奏者の演奏開始)の前に，その奏者に向けて主に目で指示。(Einsatz, cue)	○
曲の終わりや切り	手首で叩くか，はっきり手を停止。揃って切れ難い場合のみ必要。	○
拍子とテンポ	指揮動作軌跡の図形(1〜7つ振り)。小節頭は下方に叩き省略不可。	○
拍のタイミング	指揮法の軌跡図形上に規定された箇所を通過した時点。高さの極小点通過時に発音するとは限らない。タイミングは指揮者や楽団で微妙に違う。	○
アゴーギク	テンポ変動。テンポ変化開始の拍の前に指示。ルバート。	○
リテヌート	特定の拍で棒を止めずに待つ。拍打点直後を極度に遅くする。	×
フェルマータ	終わりは「切り」で指示。ピアノや休符の場合は棒を止めるのみ。フォルテでは棒を徐々に持ち上げたり細かく震わす。	○
強さ(大きさ) デューナーミク アクセント	図形の大きさ。forte, mf, mp, piano, cresc., dim., sfz,等の強弱。 1拍前に予備的運動を行い，アクセントの拍を強く振る。 予備で息を吸い，アクセントの拍で吐く。	○ ○
奏法(Articulation)		
スタカート	各拍ではっきりと停止。素早く直線的。軌跡図形は小。	○
エスプレシボ	曲線的で滑らかな図形。(レガート)	○
ノンエスプレシボ	上記2者の間。エスプレシボとスタカートの程度や切替は棒だけで示し得る。	○
マルカート	各拍で停止。拍間がスタカートより緩やか。重厚で力強い。	×
テヌート	各拍で停止。マルカートより軽い。	×
フレージング	フレーズの終わりで調子が弱まる。	×

指揮者の要件として他に人格，人柄，カリスマ性，気，相性，心などもあるが本稿では扱わない。

図1 指揮動作軌跡の例

の習慣で指揮者やオーケストラによって異なる。一般にドイツ系のオーケストラは反応が遅く、アメリカのオーケストラは早いとされる。この遅延に最も密接に関係するのはテンポで、テンポが速ければ時間遅延量は小さく遅ければ大きい。本システム試作過程で、ある日本人指揮者に試奏を依頼し、指揮テンポと時間遅延量の最も自然に感じる組合せを調べた結果、メトロノーム(1分間の打数)50で100ミリ秒の時間遅延挿入が丁度良く、メトロノーム110では時間遅延の挿入は不要だった。指揮を学び始めた頃は誰でもオーケストラの出(発音)を遅く感じ、更にその発音を聴いて指揮動作をそれに合わせたりするため、どんどん演奏が遅くなる現象が頻発する[13]。

アーティキュレーションや音のニュアンスは打点間の動作軌跡形状で指示する。

楽員は常に指揮を凝視しているわけではない。必ず指揮を見るのは、曲の演奏開始や終わり、accel.(徐々に速く)やrit.(徐々に遅く)などテンポの変化点、大事なソロのインザッツなどで、ここでは楽員は完全に指揮に従う。一方、テンポも曲想も一定のマーチなどの例では指揮不要の部分も多い。このようにオーケストラの自律性は1曲中でも変動する。常に指揮を凝視

せずとも演奏が成立するのは、全楽員と指揮者が曲の情報を共有し、指示すべき或いは見るべき箇所を知っているからである。例えば演奏中の指揮動作の停止を見て、指揮の省略かフェルマータかを弁別するためには、このような情報共有が不可欠である。

管楽器や歌など息を使う演奏に限らず音楽演奏には呼吸法が重要である。大学音楽学部指揮科でも指導する他、カラヤンでさえベルリンフィルハーモニーと共に意識的に呼吸法を実践していた[28]。例えば演奏開始の予備拍やアタックの前の予備などで息を吸い、振り下ろしに合わせて息を吐く。

3. システム構成と拍の認識処理

3.1 システム構成

システム構成を図2に示す。ユーザの右手指揮動作を加速度センサ、目によるインザッツを視線検出器、呼吸を呼吸センサで各々検出する。指揮軌跡は図1のように2次元の動作軌跡として表されるので、縦と横方向検出用の加速度センサ2個(静電型加速度センサ:東プレTPR 70 G-100)を指揮棒に装着。センサ出力は指揮動作認識に不要な周波数成分を除去する帯域通過フィルタに通す。図1の軌跡図形に合わせて丁寧に指揮した際のフィルタ通過後の加速度波形例を図3に示す。図3左の加速度絶対値時間域波形のローカルピークが拍打に相当する。

視線検出器は白目と黒目の赤外線反射率の差を検出する強膜反射式のゴーグル型両眼球運動測定システム(竹井機器TKK 2901)で、16 msec毎にRS 232 C経由で視点情報をパソコンに送る。ユーザは1m離れた29インチモニタ画面に提示されたオーケストラ画像を見ながら指揮する。パソコンではモニタ画面を8領域に

図2 指揮シミュレータのシステム構成

分割し、ユーザが見ている領域を検出する。今の視線検出器の分解能では、オーケストラ画像の中の隣り合う奏者、例えばフルートとピッコロのどちらを見ているかは弁別できないが、8分割した領域のどこを見ているかは弁別可能で、フルートとバイオリンのどちらを見ているかは十分に識別できる。

呼吸の検出法には鼻孔にサーミスタを装着する方法や、MIDI プレスコントローラを口にくわえる方法があるが、前者は口での呼吸を検出できないし、後者は息の「吐く」を検出するが「吸う」を検出できない。本システムではこれらの問題を避けるため、胸に細い伸縮センサを巻いて胸囲変化から呼吸を検出する呼吸ピックアップ(日本光電 TR-751 T)を使った。この出力を A/D 変換してパソコンに取り込み、呼吸曲線の傾斜から息を吸っているか吐いているかを弁別した。

センサ出力を取込む A/D 変換、帯域通過フィルタ、認識処理計算、音楽再生の一連の処理を 10 ミリ秒毎に繰返す。拍認識処理に要する時間は 20 ミリ秒でパソコンで余裕もって実時間処理する。パソコン上に MIDI (Musical Instruments Digital Interface) 規格の曲データ(以下、曲データ)を記憶し、指揮認識結果に基づいて拍タイミングや音の強さを制御しながら再生。これ

図3 帯域通過後の加速度信号例
(絶対値加速度時間波形と2次元軌跡1周期分)

を MIDI トーンジェネレータ(ヤマハ MU 80)で音響信号に変換した。曲データはオーケストラパートに相当する複数の MIDI チャンネルデータで構成される。

3.2 拍認識処理の構成

右手指揮動作からの拍認識の構築に、楽員や指揮者へのインタビューで得た知見を活用した。即ち(ア)演奏する曲を知っている方が指揮に追従し易い、(イ)更には慣れている曲の方が指揮に追従し易い、(ウ)奏者は普通予め曲の情報が頭に入っている、(エ)指揮を常に凝視しなくてもある程度演奏が成立する、(オ)それは曲の知識やリハーサルの情報、今演奏中の状態などを指揮者と全楽員が共有しているから、これらのことからオーケストラ奏者は、(i)単に指揮動作軌跡形状を目で見る狭義の動作認識だけでなく、(ii)演奏中のテンポや(iii)演奏中の曲の知識に基づいて「次に何拍目がいつ頃来そうか」を予測しながら演奏していると考えられる。これに倣って拍の認識処理部を構成した。人の指揮認識と本システムの処理ブロック(以下、agent と呼ぶ)の対応を表2に示す。agent-1 は HMM 動作認識、agent-2 は演奏中のテンポからの拍打生起予測、agent-3 は曲データ上の拍生起可能性分布計算、agent-4 は agent-1, 2, 3 の各出力を統合する総合判定部である。

3.3 HMM 認識(agent-1, 狭義の動作認識)[13]

HMM とはラベル系列を出力する確率遷移過程(マルコフモデル)である[29]。HMM 認識プロセスを図4に示す。まず加速度センサ入力をその特徴量に基づいて 32 種のラベルに量子化する。特徴量は、(a)前拍時との加速度縦成分の強度差、(b)拍打時つまり絶対値加速度時間波形ローカルピーク時の加速度横成分の強度、(c)2次元加速度波形の xy 平面内での回転方向である。10 ミリ秒毎に 32 種のラベルのいずれかが出力

表2 楽員の指揮(拍)認識と本システムの対応

楽員の認識手段	認識の対象	本システムでの実現方法
指揮する右手の動作軌跡形状	・拍子 ・拍打生起時点 ・何拍目か ・アーティキュレーション ・デュナーミク	HMM (agent-1, 狭義の動作認識) 認識対象に対応する複数のHMMを設定、学習済。 HMM No.0 = 1st of 1-beat, No.1 = 1st of 2-beat, No.2 = 2nd of 2-beat, ... 加速度絶対値の大きさ
演奏中のテンポに基づく拍打生起予測	・拍打生起時点	ファジィ・ルール (agent-2) If(前拍打からの演奏テンポに基づく経過時間が長い) then(拍打生起の可能性が高い)。
曲の知識の活用(打数は予め伝達)	・拍子 ・何拍目か, 他	ファジィ・ルール (agent-3) If(演奏中の箇所が曲データの拍や音符に近い) then(拍打入力にそれが相当する可能性が高い)。
総合判断(上記複数条件から)	上記各要素	ファジィ・ルール (agent-4) If(agent-1で3拍目のHMMの尤度が高い) and(agent-2で、拍打生起可能性が低い) and(agent-3で、入力が3拍目に相当する可能性が高い) then(3拍目の拍打入力生起の可能性が高い), 他。

され、そのラベル時系列が HMM 認識処理へ入力される。認識対象動作ごと、即ち 1~5 つ振りの各拍に対応して各々 1 つの HMM を設定、予め学習してある。全 HMM について入力ラベル時系列を出力する確率(これを尤度と呼ぶ)を計算する。加速度絶対値の時間域波形ローカルピーク生起時に最大の尤度を示した HMM 番号が認識結果である。例えば HMM 5 番なら「3 つ振りの 2 つ目」などである。HMM 形式は飛越し無しの left-to-right 型、離散型で、遷移時にラベルを出力。HMM は特徴パラメータの揺らぎを出力確率で、時間構造の揺らぎを遷移確率で吸収する。本システムの HMM の定義を式(1)と図 5 に示す。式(1)の HMM のモデル M がラベル系列 $L = l_1, l_2, \dots, l_T$ を出力する尤度 $P_r(L|M)$ は(2)式で得られる。ラベル系列長 T は指揮テンポにより変動する。HMM は Baum-Welch アルゴリズム[29]で学習した。学習前の初期 HMM は全ての HMM に共通で、全てのラベル出力確率を均一に設定。HMM の学習に使った教師サンプルは HMM 1 つ当り 100 前後[12]。

- ・状態数 $N=5$. (状態: $S_1, \dots, S_5$)
- ・ラベル数 $K=32$. (ラベル: $l_1, l_2, \dots, l_{32}$)
- ・遷移確率 p_{ij} : 状態 S_i から S_j に遷移する確率。
- ・出力確率 $q_{ij}(l_k)$: 状態 S_i から S_j へ遷移の際にラベル l_k を出す確率。
- ・初期状態確率 $r_i=0$. ($i=2, \dots, 5$)
 $r_1=1$: 初期状態が S_1 である確率
- ・HMM 数 $C=16$: 認識対象の数
- ・最終状態の集合 $I_f=\{5\}$: S_5 が最終状態。

(1)

$$P_r(L|M)$$

$$= \sum_{i=0} \dots \sum_{i \in I_f} r_{i0} p_{i0i1} q_{i0i1}(l_1) p_{i1i2} q_{i1i2}(l_2) \dots p_{i(T-1)T} q_{i(T-1)T}(l_T) \quad (2)$$

図4 HMM指揮動作認識プロセス

図5 本手法のHMM

3.4 曲の知識や演奏情報の活用 (agent-2, 3, 4)

前述インタビューから得た知見に基づき、演奏中のテンポからの拍生起可能性予測 (agent-2)、曲データ上の拍生起可能性計算 (agent-3)、各 agent 出力の統合 (agent-4) を、if-then 形式ファジィプロダクションルールで構築した。ファジィ推論演算には Mamdani の min-max-重心法を用いた。agent-4 では、agent-1 で計算される各 HMM の尤度、agent-2 や agent-3 から出力される各拍の生起可能性を統合して最終判定する。ファジィルール例を表 2 に、その計算式を(3)式に示す。

$$\begin{aligned} O_1(p) &= L_h(n_3) \wedge P_h(p), \\ O_2(p) &= T(t) \wedge P_h(p), \\ O_3(p) &= C_3(d) \wedge P_h(p), \\ O_4(p) &= O_1(p) \wedge O_2(p) \wedge O_3(p), \\ P_{r3} &= [\int O_4(p) p \, dp] \div [\int O_4(p) dp]. \end{aligned} \quad (3)$$

但し、 n_3 は 3 拍目の HMM 尤度をその時点の全 HMM 尤度の最大値で正規化した値、 $L_h(n_3)$ は「HMM 尤度が高い」というメンバシップ関数の n_3 に於けるグレード、 p は 0 から 1 の間の値をとる拍生起可能性、 $P_h(p)$ は「拍の生起可能性が高い」というメンバシップ関数、 $O_1(p)$ は「3 拍目の HMM 尤度が高い」というファジィ推論出力、 t はその時のテンポに基づいて計算される前拍打から次拍までの時間経過割合、 $T(t)$ は「時間経過が長い」というメンバシップ関数の t に於けるグレード、 $O_2(p)$ 、 $O_3(p)$ 、 $O_4(p)$ は各々 agent-2, 3, 4 のファジィ推論出力、 d はその時点で演奏中の箇所から 3 拍目までの距離、 $C_3(d)$ は「3 拍目に近い」というメンバシップ関数の d に於けるグレード、 P_{r3} は 3 拍目の生起可能性のスカラ値である。メンバシップ関数例として agent-2 用 $T(t)$ と $P_h(p)$ を図

6に示す. 3拍目の例を述べたが他の拍についても同様の処理で求め, 絶対加速度ローカルピーク時点で生起可能性最大のものを認識結果とした.

曲データに拍子情報は含まれているが打数(1小節を幾つに振るか)の情報は無い. 曲データ読み込み時に「拍子=打数」に設定し, それに基づいてファジィ・メンバシップ関数形式で拍の生起可能性分布を図7(a)のように自動設定した. この3拍目が $C_3(d)$ に相当する. 拍以外の位置でタイミング指示したい場合は図7(b)のようにマニュアルで設定する. なお, 拍位置推定に重み関数を用いた先行研究[30,31]が有る.

図6 ファジィ・メンバシップ関数の例

図7 曲データ上の拍打生起可能性分布の例 (agent-3用)

曲の拍子と打数は必ずしも同じではなく[12], その関係を明記した法則も無いので指揮者は必ず奏者に打数を予め伝えねばならない[32]. 本システムでは演奏中でも打数を設定変更可能とした.

試作の過程で動作軌跡だけ, 即ち HMM(狭義の動作認識: agent-1)だけで2つ振りと3つ振りの認識を試したところ, 学習用教師データ提供者の丁寧な指揮動作を 96.13%, それ以外の人の動作を 92.49%の率で認識した[12]. 他の動作認識研究[23]でも不特定者の動作認識に HMM が有効であることが確認されている. しかし短い曲にも数百の拍があるので, agent-1 だけの認識率は音楽演奏には不十分である. 従って HMM だけでなく曲や演奏の情報活用が不可欠で, agent-2,3,4 の併用により, 教師データ提供者の丁寧な指揮を 99.74%, それ以外の人の動作を 98.95%と, 誤認識無しに1曲通せる程度まで認識率が向上した.

ファジィ推論は, (イ)人のノウハウを If-then 形式で搭載し易い, (ロ)多くのパラメータが複雑に絡んでも大丈夫, (ハ)厳密な解析モデルが不要という特長を持つが, 本システムの実現に際しこれら長所を享受したと感じている.

4. アインザッツ, 呼吸, 他

4.1 視線アインザッツ

図8に視点の検出例を示す. 新世界交響曲第2楽章でイングリッシュホルンによる「家路」のメロディが

画面を下図のように8分割.

0	1	2
3	4	5
6	7	

白い太線が視点. 29インチモニタに提示されたオーケストラ画像のオーボエ付近を見ている.

図8 視点検出例

始まる1拍前に、その奏者が位置する領域(4番)を見ている例である。

曲データと一対の制御マークファイルには、演奏中のどの時点で何番の領域(番号0~7)を見るべきかと、そのアインザッツ対象パートに対応する曲データのMIDIチャンネル番号を予め記してある。これをアインザッツ・マークと呼ぶ。演奏中は29インチモニタに提示されたオーケストラ画像の中のアインザッツを出すべきパートを妥当なタイミングで見る。演奏がアインザッツ・マークに至った時点で、視線検出器から得られたユーザの視点がアインザッツマークで指定された見るべき領域に入っていた場合はそのまま演奏が継続されるが、外れていた場合は「アインザッツ出し忘れ」と判断してマーク指定のMIDIチャンネルのみを演奏中断し、3拍後に自動復帰する。こうしてアインザッツ忘れによるパート落ちをシミュレートした。視線検出器の誤差によるもの以外にはアインザッツの誤認識は発生しなかった。

4.2 呼吸法

図9に田園交響曲冒頭の譜例と、そこを演奏中の呼吸に相当する胸囲波形の例を示す。アインザッツと同様、制御マークファイルにブレスマークを予め記入してある。これには息を吸う或いは吐くタイミングが記されている。例えばフェルマータ後の最初の音の前に息を吸い、その音の発音に合わせて息を吐く、などである。演奏がブレスマークに至った時点で、センサから得られたユーザの呼吸と、マークの呼吸指定が合っていればそのまま演奏が継続されるが、タイミングがずれたり吸/吐が逆の場合には呼吸法のミスと判断して、それ以降の演奏を中断する。今の呼吸センサは胴体の動きもノイズとして拾い易いが、安静に指揮をする範囲では呼吸の検出誤りは発生しなかった。

図9 呼吸検出例

4.3 自律性

拍入力に追従して曲データを再生することでアゴーギクに対応した、拍打時点でテンポ変化係数、即ち

$$(\text{曲データの拍間時間}) \div (\text{指揮の拍打時間間隔})$$

をその時のテンポ値に乗算して演奏テンポを更新する。しかしこのままでは指揮入力への反応が過敏なため、テンポ変化に上下制限を付けたり、数拍前までのテンポ値を使って移動平均を施す。

前述のようにオーケストラの自律性の程度は曲の中で常に変動する。これをシミュレートするために自律度を導入し、自律度マークを曲データと一対の制御マークファイルに追記可能とした。自律度は、(1)指揮入力停止時に何拍先まで自動演奏を継続するか、(2)テンポ変化幅上下限、(3)演奏テンポ計算の移動平均度数で構成した。自律度最低(0%)では曲データ内のテンポ情報を無視して指揮入力に密着して追従し、指揮動作を止めると同時に音楽再生も停止し次の動作入力までそのまま待ち続ける。発音状態で動作を止めれば音が鳴り続けるので、フェルマータも任意の長さに延ばし続けることができる。休符のフェルマータも同様。自律度最大(100%)では指揮入力を無視して曲データのテンポ情報にのみ従う。自律度マーク記入は指揮者と楽員の間の情報共有に相当する。なお、独奏と伴奏の相互作用を検討したり[33]、伴奏の自主性を導入して演奏感を向上した先行研究[34]がある。

4.4 その他の処理[12,13]

本システムは指揮法に規定された予備動作に則って演奏を開始する。予備拍タイミングと強さを音で確認することも可能。

図3左の加速度時間波形のローカルピーク生起時点から、それに対応する拍の発音までには、前述のようにテンポに伴って変化する時間遅延が必要である。そこで(4)式に示す時間遅延を挿入。delayは時間遅延量[ミリ秒]、coefとconstは定数、tempoは演奏中のテンポ[拍/分]である。定数は上記の日本人指揮者による試奏に基づいてcoef=-5/3、const=550/3に設定した。

$$\text{delay} = \text{coef} \times \text{tempo} + \text{const} \quad (4)$$

デュナーミクは加速度絶対値に比例。正規化後に曲データ内の強弱情報(MIDIのVelocity)に乘算した。アクセントも局所的なデュナーミクとして対応。

アーティキュレーションの内、スタカートからエスプレッシボの間の程度を認識した。指揮法によると表1と図1のようにスタカートの動作は鋭く、棒が停止する瞬間があるが、エスプレッシボは滑らかに常に動いて

いる。この認識に、拍間の指揮動作形状の滑らかさ(絶対値加速度波形の高周波数成分強度)や動作の一時停止の程度(絶対値加速度波形の谷の深さと時間)を使った。

曲データは市販 MIDI 曲データを加工して使った。一般の曲データには、音符、音の強さ、テンポ、拍子、小節線、楽器の種類と配置(パンニング)など、スコアに相当する情報が入っている。本システムでは本番前の準備要素は予め曲データに書き込むものとした。市販 MIDI シーケンサ・プログラムを使えば曲データの変更や調整も可能。この作業はリハーサルの一部に相当する。

指揮認識結果に基づいて曲データを再生する際、アーティキュレーションだけは MIDI にパラメータの規定が無いので、指揮認識出力を使わなかった。

5. まとめ

指揮法基礎に則った丁寧な右手指揮動作から多くの要素を認識し、アインザッツや呼吸も検出した。これに基づいて実時間に音楽演奏を制御するマルチモデル指揮シミュレータを試作した。オーケストラの自律性や時間遅延もシミュレートした。指揮を仮想体験する音楽演奏システム構築の技術的な礎となる。

今の視線検出器には、視野の一部がセンサで遮られる、体を動かすと検出精度が落ちる、センサ装着の度に綿密な調整作業が必要といった問題がある。遮られるのは視野中央下部で、指揮者が楽譜を置く所に位置する。ユーザに何も装着せずにカメラと画像処理だけで視点を検出する手法[35]も提案されて始めたが、まだ至近距離に限定されている。また呼吸センサも体の動きを敏感に拾ってしまう。このためユーザは動作が制限され、多大な負担を強いられる。

これらの問題点を解決し、指揮教授者の演奏中の指揮動作、視点、呼吸を記録再生したり、指揮学習者が自分のデータと比較できれば、学習の参考になると考えられる。また、今はアインザッツや呼吸のマークを手作業で入力しているが、教授者のデータを記録できれば、その情報からマークを自動生成することも可能となるだろう。

現在クリスプな処理をしている HMM 用ラベル化へのファジィ推論の適用、拍番号連鎖の特徴を活用するファジィルールの追加等により、更に認識率向上が可能だろうと推察している。また、より詳細な指揮要素の認識と、その音への反映も今後の課題である。

謝 辞

東京芸術大学音楽学部指揮科佐藤功太郎先生には、実地の演奏活動や教育活動に基づいた、指揮法に記されていない様々な指揮について知見やアドバイスを頂戴した。深く感謝致します。本研究の機会を与えて下さったヤマハ株式会社石村社長、奥村取締役、和智取締役、鈴木楽器研究所長、永井主査に感謝致します。

参 考 文 献

- [1] 鈴木孝：音楽芸術と計算機，情報処理，Vol.35, No.9, pp.830-835(1994)
- [2] Roads, C.: the computer music tutorial, The MIT Press(1995)
- [3] 秋田真彦, 増山繁, 船橋賢一：ニューラルネットワークと主成分分析を用いた自動採譜の検討，電子情報通信学会技術報告，1992-12, NC92-81, pp.9-14, (1992)
- [4] 大矢健一：ニューラルネットのダイナミクスによるリズム認知モデル，情報処理学会音楽情報科学研究会報告 MUS7-1, 1994.8.6~7(1994)
- [5] 大矢健一：リカレントニューラルネットワークによる楽音合成，情報処理学会音楽情報科学研究会報告 MUS13-2, 1995.12.1~2(1995)
- [6] 小田安彦, 村上豊, 波多正臣, 梶川嘉延, 野村康雄：演奏者情報を加味したピアノの自動演奏システムの構築～ニューラルネットワークによる個人的揺らぎの生成～，日本音響学会講演論文集，平成8年3月，pp.669-670, (1996)
- [7] 徳丸正孝, 文室晋一, 村中徳明, 今西茂：ファジィ推論による曲調を考慮した伴奏コードの自動生成，電学論 C, 117巻4号, pp.424-431, (1997)
- [8] 後藤真孝, 村岡洋一：音響信号に対するリアルタイム・ビートトラッキングシステム，情報処理学会音楽情報科学研究会報告，1996.7.27, 16-3, pp.13-20(1996)
- [9] 川端純一, 才脇直樹, 辻本浩章, 西田正吾：エージェントを用いた感性協調型合奏システム，ヒューマンインタフェース，67-9, 1996.7.11, pp.59-64(1996)
- [10] 山内敏彰, 野池賢二, 野瀬隆, 乾伸雄, 小谷善行, 西村恕彦：GA を用いたピアノ演奏の局所テンポ解析，情報処理学会音楽情報科学研究会報告 MUS14-1(1996)
- [11] 平野砂峰旅：大域結合カオスのコンピューターミュージックへの応用，情報処理学会音楽情報科学研究会報告 MUS14-7, 1996.2.24(1996)
- [12] 宇佐聡史, 持田康典：指揮認識へのソフトコンピューティング応用，第30回ファジィ制御研究会，(1997)
- [13] Satoshi Usa, Yasunori Mochida: A Conducting Recognition System on the Model of Musicians' Process, Journal of the Acoust. Soc. of Japan(E) (1998.7掲載予定)

- [14] 渡辺孝弘, 李七雨, 谷内田正彦: インタラクティブシステム構築のための動画像からの実時間ジェスチャー認識手法-仮想指揮システムへの応用-, 電子情報通信学会論文誌, D2, Vol.J80-D2, No.6, pp.1571-1580, (1997)
- [15] ルードルフ, 大塚明訳: 指揮法, 音楽之友社(1968)
- [16] 古井貞おき: 音声認識, 電子情報通信学会誌, 11/95, Vol.78, No.11, pp.1114-1118(1995)
- [17] 吉原道宏, 山口亨: 動作認識による知的ヒューマンインタフェース, 日本ファジィ学会誌 Vol.7, No.4, pp.871-882(1995)
- [18] 北田豊浩, 萩原将文: ファジー推論 DP マッチングを用いたジェスチャー認識, 電気学会論文誌 C, Vol.115-C, No.3, pp.496-502(1995)
- [19] 牛田博英, 山口亨: ファジィ連想記憶システムによる動作認識, 画像ラボ'95.2, pp.21-24(1995)
- [20] 今川太郎, 森家みち代, 近藤堅司, 丸野進: 量子化ニューロチップによる手振り入力インターフェース, テレビジョン学会技術報告, Vol.19, No.57, pp.19-24, IPU'95-51(1995)
- [21] 大和淳司, 大谷淳, 石井健一郎: 隠れマルコフモデルを用いた動画像からの人物の行動認識, 電子情報通信学会論文誌 D2, Vol.J76-D2, No.12, pp.2556-63(1993)
- [22] 大和淳司, 倉掛正治, 伴野明, 石井健一郎: カテゴリー別 VQ を用いた HMM による動作認識法, 電子情報通信学会論文誌 D2, Vol.J77-D2, No.7, pp.1311-18(1994)
- [23] 武士展照, 春山智, 小林哲則: HMM を用いた手振り認識, 電子情報通信学会信学技報 PRMU96-8, Vol.96, No.40, pp.53-59(1996)
- [24] 澤田秀之, 橋本周司: 加速度センサーを用いたジェスチャー認識と音楽制御への応用, 電子情報通信学会論文誌 A, Vol.J79-A, No.2, pp.452-459(1996)
- [25] Teresa Anne Marrin: Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton, <http://japan.park.org/Events/BrainOpera/Archive/Thesis.html>, (Media Arts and Sci. M.I.T(1996)
- [25] T.Marrin, J.Paradiso: The Digital Baton: a Versatile Performance Instrument, Proc.ICMC97, pp.313-6(1997)
- [25] J.A. Paradiso: Electronic music: new ways to play, IEEE SPECTRUM Dec.97, pp.18-30(1997)
- [26] 片寄晴弘, 金森務, 井口征士: 簡易モーションキャプチャセンサ DigitEye3D-原理と基本特性-, 電子情報通信学会論文誌, Vol.J80-D2, No.10, pp.2889-92(1997)
- [27] 黒川隆夫: ノンバーバルインターフェース, オーム社 pp.31(1994)
- [28] チェスターマン著, 中尾正史訳: マエストロ達との対話, p.44, 洋泉社(1995)
- [29] 大河内正明: Hidden Markov Model に基づいた音声認識, 日本音響学会誌42巻12号 pp.936-941(1986)
- [30] R.B.Dannenberg, B.Mont-raynaud: Following an improvisation in real time, Proc.ICMC, pp.241-248(1987)
- [31] 片寄晴弘, 井口征士: 知的採譜システム, 人工知能学会誌, Vol.5, No.1, pp.59-66(1990)
- [32] 大塚明, 他: 新訂 標準音楽辞典アーテ, pp.749-752, 音楽之友社, (1966-96)
- [33] 澤田秀之, 磯貝昌幸, 橋本周司, 大照完: 音楽演奏に於ける人間と機械の協調動作について, 情報処理学会第44回全国大会講演論文集, pp.1,389-390(1992)
- [34] 堀内靖雄, スパワリークリッサダー, 田中穂積: 伴奏者の自主性を考慮した自動伴奏システム, 情報処理学会第44回全国大会講演論文集, pp.1,405-406(1992)
- [35] 知野哲朗, 福井和広, 山口修, 鈴木薫, 田中克己: GazeToTalk: メタコミュニケーション能力を持つ非言語メッセージ利用インタフェース, インタラクシオン98, pp.169-176(1998)

(1997年8月26日 受付)

(1998年4月23日 再受付)

[問い合わせ先]

〒430-8650 静岡県浜松市中沢町10-1

ヤマハ(株) 楽器研究所

宇佐 聡史

TEL : 053-460-2292

FAX : 053-460-2242

E-mail : usa@milab.yamaha.co.jp

— 著 者 紹 介 —

宇佐 聡史 (うさ さとし)

工学院大学 大学院情報学専攻, ヤマハ(株) 楽器研究所

1984年 慶応義塾大学工学部電気工学科卒業, 1986年 慶応義塾大学大学院工学研究科電気工学専攻修了, 同年 ヤマハ(株) 入社, 電子楽器の研究開発に従事, 現在 同社楽器研究所勤務, 音楽情報処理, ヒューマン・インターフェイス, 楽器の研究開発に従事, 日本ファジィ学会, 日本音響学会, 情報処理学会, 電子情報通信学会 各会員.

持田 康典 (もちだ やすのり)

工学院大学 大学院工学研究科情報学専攻

1948年 金沢工業専門学校電気通信科(現金沢大学工学部)卒, 1983年 ヤマハ(株) 専務取締役, 1988年 (株)リコー 専務取締役, 1992年より工学院大学工学部電子工学科情報工学コース教授, 1993年 同大学工学研究科情報学専攻教授, マルチメディア統合および電子計算機構成に関する研究に従事, 1975年 科学技術庁長官賞受賞, 1984年 藍綬褒章授章, 工学博士(東北大学), 電子情報通信学会, IEEE各会員.

A Multi-Modal Conducting Simulator

by

Satoshi USA and Yasunori MOCHIDA

Abstract :

Soft-computing applications for music are increasing in recent years. They are regarded as suitable solutions for musical information processing. In this paper, an application for a multi-modal conducting simulator is mentioned. It recognizes human conducting and behaves like an orchestra in real-time. Even for the same conducting intention, both the gesture characteristics and the time-domain structure vary depending on the conductor, and on the connection between previous or latter parts. Soft-computing is applied for recognition to cope with both these variance and to detect conductor's intention. The beat recognition process is built on the model of actual musicians' recognition processes. The right-hand conducting gesture is captured by accelerometers and recognized with HMM (Hidden Markov Models). Based on interviews with conductors and orchestra players, musical score information and performance information are also applied to the recognition on the fuzzy model. The watching point and breath of a user is detected with a eye tracker and a breath sensor, respectively. To shape the musical performance in real-time, the system recognizes important and universal conducting elements conforming to the grammar of conducting : i.e., beginning and ending of a piece, Einsatz (cue for a player or a part, an important direction by a conductor) with eye, beat timing including Agogik (rubato, changes in tempo), the beat number in a measure, fermata (to lengthen notes or rests arbitrarily, to pause), Dynamik (dynamics), some aspects of articulation (i.e. espressivo - staccato), and breathing. The system simulates orchestra response to correct breathing and Einsatz by a user. It also simulates sounding delay and autonomy of actual orchestra. When the autonomous degree is 100%, the system ignores the conducting input and plays with tempo on music data. When the autonomous degree is 0%, the system is glued to the conducting. In this case, when the conducting gesture stops, the system stops replay simultaneously and waits until the next beat input.

Keywords : Human Interface, Multi-modal, Gesture Recognition, HMM, Fuzzy, Musical Information Processing, Conducting, MIDI

Contact Address : **Satoshi USA**

Musical Instrument Research Laboratories, Yamaha Corporation.

10-1, Nakazawa-cho, Hamamatsu-shi, Shizuoka, 430-8650 Japan

TEL : 053-460-2292

FAX : 053-460-2242

E-mail : usa@milab.yamaha.co.jp