

仮想ダンスインストラクターのための教示モデルの構築

Construct the Instruction Model of a Virtual Ballroom Dance Instructor

関 優樹

Yuki SEKI

黄 宏軒

Hung -Hsuan HUANG

上 丞正樹

Masaki UEJO

李 周浩

Joo-Ho LEE

川越恭二

Kyoji KAWAGOE

立命館大学情報理工学部

College of Information Science and Engineering, Ritsumeikan University

The use of virtual conversational agents is awaited in the tutoring of physical skills such as sports or dances. This paper describes about an ongoing project aiming to realize a virtual instructor for ballroom dance. In order to design the instruction model which is the most essential part of the virtual instructor, this work starts from a corpus collection of human-human experiment. One professional dance instructor and six dance learner subjects are recruited for the experiment. A state transition model is constructed based on the analysis of the utterances and nonverbal behaviors of both the instructor and the learner. A prototype virtual instructor system is then built based on the model.

1. はじめに

スポーツなど身体動作を伴う技能の習得は、学習者の反復的な練習の蓄積によってのみ上達が見込まれる。しかし、上達の進度や技術的な到達度には個人差があり、また、練習の効率によってもそれらは大きく左右される。一般的に初心者は、ある一定以上の技能レベルに到達するまでは、効率の良い練習を自ら考案し実行することは難しく、その種目の技術と教え方に精通した教示者に指導を受ける場合が多い。優秀な教示者の一对一の指導であれば、個々の学習者の癖を把握し丁寧かつピンポイントな指導が可能である。また、教示者の一方的な説明ではなく、学習者がよく理解できなかった部分を質問することにより重点的に苦手な身体動作の指導をしてもらうことができる。対話から教示者も学習者の理解度に合わせてフィードバックの仕方を調整しながら教えることによって技能の習得効率も上がることが期待できる。

このように技能習得において効率の良い練習方法の提示とは、個人の技能段階を的確に判断し、その個人に合った練習方法を提示し実行させることであり、特に基盤となる技術を身につける段階である初心者には教示者の指導は不可欠なものである。しかし、教示者からの指導を受ける場合に学習者には以下の負担と制約が存在する。

第一に、時間的負担である。学習者、教示者の時間的な都合により学習者自身が指導を受けたい時間に時間を確保できない可能性がある。さらに、人気のある教示者であればレッスンの予約を取る事が難しく指導してもらう事さえ困難な場合がある。

第二に、金銭的負担である。通常プロのインストラクターと呼ばれる教示者に指導を受ける場合、継続的に受講料を払わなければならない学習者の金銭的負担は大きく、さらにスポーツを教示されるような施設や設備を個人で所有することは難しい。次に第一の制約として、場所的制約がある。教示者は通常レッスンを行う為の施設に勤務しているが、教示を受けたい教示者がいるにも関わらず、家から遠い場所にあり通うのが難しいまたは不可能であるという場合がある。さらに上述した金銭的な負担に関連して、移動費などがかかる事も多く、その点も学習者にとっての大きな障害となっている。

第三に、能力的制約がある。個人により上達の差があるために、全ての学習者が平等に同じ長さのレッスンを受講した場合でも、動作の習得に時間がかかる学習者がいるために動作の習得に差が開いてしまう。この制約を解消するためには、本来苦手な部分を反復的に練習する必要があるが、教示者につきっきりで指導することは難しい。また、教示者によっても指導の仕方や流れ、身体動作の評価基準が異なり、学習者の適性もありその上達進度に違いが出てくる。

以上のような負担、制約があるが、技能動作の習得において教示者の重要性は高く、またダンスやフィギュアスケート、体操などの表現型の運動は客観的な身体動作への評価が重要となるため、十分な技能知識と教示手法を持つ教示者から習う事が多い。今回、これらの制約を解消し、誰もが好きな時、好きな場所で十分な教示を受けることができるためのシステム構築の足掛かりとして、社交ダンスを取り上げた「擬人化会話エージェントを用いた社交ダンス教示システム」を提案する。

本研究では、社交ダンスの教示者の教示内容を体系化し、それらをシステム化することにより場所、時間の制約に関係なく自分の決めた場所で教示を受けられることを目指す。学習者とエージェントのインタラクションを行うために、学習者の動作の入力にはモーションキャプチャを利用し、入力された動作データを解析し会話エージェントによる教示に生かす。また、人対人のコミュニケーションのように自然言語と表情やジェスチャーといった非言語行動を併用することで学習者とエージェントの円滑なインタラクションを図る。これにより誰も専門的な知識や訓練なしにシステムを活用できることが期待される。本研究では、実際にプロのダンサーの教示内容を分析し、社交ダンスの教示を体系化し、会話エージェントの行動モデルを作成する。またそれを元にエージェントを実装し、学習者にとって効率的な練習法を提示できる教示システムの構築を行う。

2. 先行研究

モーションキャプチャを用いた舞踊動作の解析は様々な先行研究がある。さらに舞踊を題材とした研究例として、白鳥らは後藤らが考案した音楽からリズムを抽出するビートトラック法 [Goto 01] を改良して、舞踊動作における最少の単位である動作プリミティブを音楽のリズムを条件として抽出することによって舞踊動作の保存と再構成を目指している [白鳥 05]。

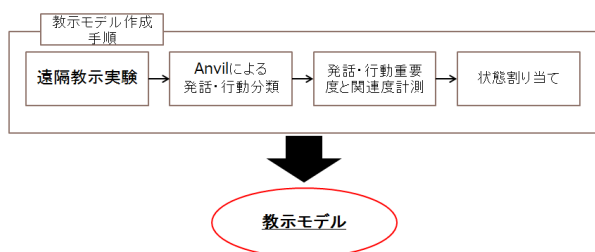


図 1: 教示モデル作成手順

会話エージェント分野では Kopp らによる MAX[Kopp 05] は実験室などの特定の環境から抜け出して実際に展示会という場で実用的に用いられている数少ない会話エージェントである。MAX は入力インターフェイスとしてキーボードを用い、ユーザの入力に対して、音声合成を利用した返答や、いくつかの動き、表情、視線、移動などで反応を示すことができる。MAX は、これらの機能が優れ実際の展示会でもその感情表現について高い評価を得ている。提案手法と似た先行研究の例として Chua ら [Chua 03] は、HMD(Head Mounted Display) とモーションキャプチャによって太極拳の教示を行うシステムを構している。学習者は HMD からバーチャル空間を見渡す。バーチャル空間には 3D CG によって表現されたキャラクターがあり、学習者がキャラクターの動作を模倣することによって動作学習し、モーションキャプチャでその動作を入力する。しかし現在、人間がシステムから動作学習するという手法は一般的ではなく、その理由としてシステムの動作解析に対して人間が分かりやすい形で学習するための教示モデル及び情報提示手法が構築されていない事が考えられる。そこで、本研究では実際に人対人の動作教示を分析し、それらから教示モデルを得る。そして情報提示手法として会話エージェントを用いることでユーザが人対人のような教示を受けることができるシステムの構築を目指す。

3. 教示モデルの作成とエージェントシステムの実装

社交ダンスを教示する会話エージェントを構築するために、実際のダンスインストラクターがどのようにしてダンスを教えるのか、その教示の流れ及び、発話・行動をモデル化しなければならない。教示モデルを作成する際に、抽出元となった教示者の教え方が正しいものであるという信憑性を得るため、本研究では抽出元としてプロのダンス講師の教示を選び、遠隔教示実験を行った。さらに、遠隔教示実験で得られたデータの中から教示モデルを構成するため、学習者・教示者毎の発話・行動を分類するためのツールとして Anvil[Kipp 08] を用いる。分類された項目の関連性、及び重要性から構築すべき教示モデルを作成する。図 1 は教示モデルの作成のために行った手順のフローである。

3.1 遠隔教示実験について

教示モデルを作成するために、人対人の社交ダンスの教示が、お互いが体を触れ合えない、スクリーン越しでどのように行われるかを観測するために遠隔教示実験を行った。(図 2) 教示者には、プロの社交ダンスの講師を招き、被験者 6 人に対

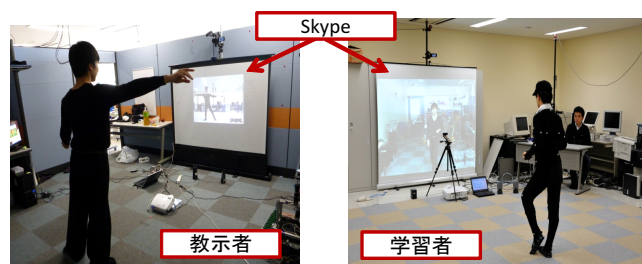


図 2: 遠隔教示実験

して以下のような条件で遠隔教示を行ってもらった。被験者は社交ダンス経験のある者を集め、それぞれのダンス技能に大きな差異が発生しない被験者を集めた。以下のような条件で実験を行った。

- 教示者と被験者は別室にて、ビデオ通話をスクリーン越しに行いダンスの教示を行う
- 教示者にはルンパの基本的な型であるベーシックを教えてもらい時間内に、型の全てについて触れながら教示してもらうように指示する
- 教示時間は一人 30 分
- 学習者は、モーションキャプチャスーツを着てもらい教示中の動作データを取得する

また、実験中は以下のデータを記録していた。

- 教示中の教示風景
- 教示中の被験者の動作データ

3.2 Anvil による分析

遠隔教示実験から教示モデルを得るために Michael Kipp が開発した Anvil というアノテーションツールを利用して遠隔教示実験で撮影したビデオに対して教示者及び被験者の発話及び行動の分析を行った。ビデオ分析に関してはダンス経験のあるコーダ二名にラベル付けを行った。その結果コーダのラベルの一致率は 80%を超えたのでよりラベル付の経験が長いほうのラベル結果を分析に用いた。分析の結果から表 1 に示すように被験者は教示中にほとんど発話しない事に対して教示者の発話時間は行動よりも長い。このことから、ダンスの教示というコミュニケーション形態は学習者が身体情報を主に用いて、教示者は発話情報を用いて行われていること分かる。これにより、教示者の発話から求めるべき教示モデルの要素が含まれていると考え、分析を進めた。ラベルとして付けられた教示者の発話の種類を表 2 に示す。

図 3 は被験者 6 人に対する教示者の発話の平均である。教示者の発話は大きく分けて、ユーザのダンスを評価し、修正する発話とユーザのダンスを促す進行発話に分類される。特に、ダンスの技術を教える前者の発話は教示の大部分を占めている。中でもユーザのダンスを体の部位毎に細かく指導する Parts of the body force ラベルに関してはどの被験者に関しても、ラベル付け回数が最も多く、細かい体の部位の修正を重点的に行う教示者個人の教示スタイルの特徴が出ていると言える。そこで今回作成する教示モデルは、この教示モデルの抽出元となった教示者の特色を生かし、体の細かな部位を修正することに重きをおいた教示モデルを作成していく。

表 1: 教示者と被験者の発話と非言語行動に対するアノテーションの結果．時間の長さ (mm:ss.00) で示す

	被験者	1	2	3	4	5	6	平均
被験者	発話	01:23.15	01:28.49	04:19.69	00:04.44	01:59.12	00:08.24	01:33.85
	非言語行動	22:21.21	19:08.92	18:19.28	23:08.86	17:30.76	18:15.00	19:47.34
教示者	発話	11:20.20	09:57.42	08:58.56	07:06.82	09:59.12	10:23.61	09:37.62
	非言語行動	14:37.86	13:05.24	12:26.00	10:56.51	14:07.66	13:05.74	13:03.17

表 2: アノテーションされた教示者の発話内容

ラベル	意味
Parts of the body force	体のどこを修正すべきかという発話
Centroid for clearly dance	重心 (ライン) を安定させるという発話
Count force	カウントの使い方に関する発話
Count	カウントのとり方に対する発話
Type to put a count	ダンスのピクチャをカウントで強調するという発話
Type	ダンスの型に関する発話
Praise	学習者のダンスに対してのほめる言葉
Stop-users-dance	学習者のダンスを止めさせる発話
Progression	教示の進行に関わる発話．例えば「ここから、ここまで踊ってください」
Reply-users-question	学習者の質問に対する返答
Reply-users-request	学習者の要望に対する返答
Reply	返事．例えば「はい」など
Make sure users hear	学習者が聞こえているか確認
Make sure users understand	学習者が教示内容を理解しているか確認
Unknown	上記以外

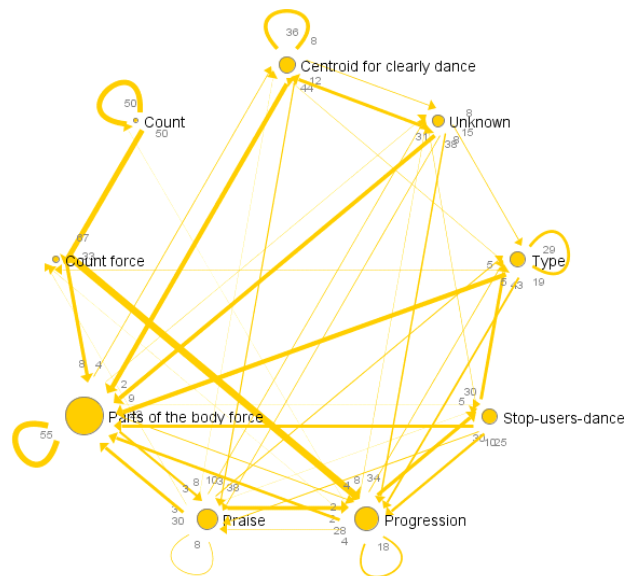


図 4: 発話遷移図

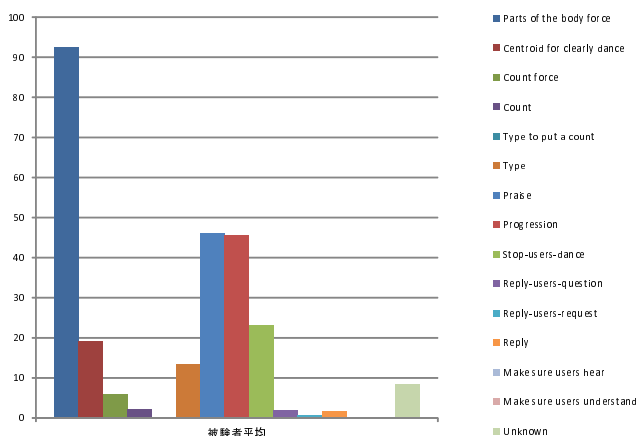


図 3: 教示者発話被験者平均ラベル数

3.3 各ラベルの状態分け

発話遷移図 (図 4) を作成した．丸は教示者の発話ラベルを表し、丸の大きさはラベル付け回数を表している．矢印の方向は各ラベル間の遷移方向を表し、矢印の太さは遷移率の高さを表している．また数字が実際の遷移率 (百分率) を表す．このダイアグラムを見て分かる通り、Parts of body ラベルはどのラベルからでも最も多く遷移されていることが分かる．これより、各体部位への教示は最も多く教示されるようにモデル化しなければならない．また、各教示発話を行った後、教示者は Progression ラベルに遷移し、学習者に対してダンスをして欲しいという進行を行っている．さらに Stop-user-dance や Praise はユーザがダンスをしている時に、教示者がそのダンスを見て評価している時に現れる発話である．さらに、各発話の遷移を五種類の状態に分類し、教示モデルとした (図 5) ．

1. カウントからカウントまでの部位毎の修正を行う教示
2. 強調カウントから強調カウントまでのアピールすべきポイントを修正する教示
3. ダンス全体を手本を見せながら教える教示
4. ユーザのダンスを見て評価する状態
5. 教示者のダンスを促す状態

3.4 教示システムの実装

教示システムために会話エージェント開発プラットフォームである黄らの GECA (Generic Embodied Conversational Agent)

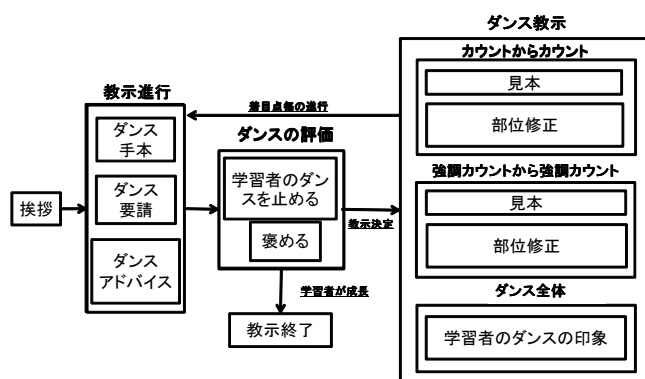


図 5: 教示モデル

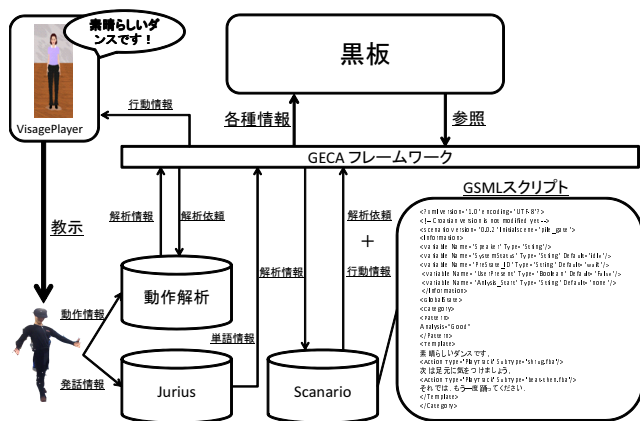


図 6: 仮想ダンスインストラクタープロトタイプシステムの構成

フレームワーク [Huang 08] にて実装を行った。コンポーネントの一つである Scenario コンポーネントにて遠隔教示実験から得られた教示モデルを基にエージェントの状態遷移及び発話と行動をエージェントのインタラクションとしてデザインする。Scenario コンポーネントは外部情報からの入力を受信と、その入力情報を条件として呼び出される、GSML(GECA Scenario Markup Language) で記述されたエージェントの発話・行動を外部に送信する機能を持っている。図 6 はシステムの構成を表したものである。動作解析部は AMSS 法 (Angular Matrix for Shape Similarity [Nakamura 06]) を用いてシステム利用者がダンスをする際の関節の移動軌跡と事前に収録したプロ教示者の物との相関値を算出し、システム利用者の動きに対する評価とする。

4. まとめ

本研究では、CG キャラクターを用いた社交ダンス教示システムの構築を目指した。システムを実現するために、モーションキャプチャによる社交ダンスの動作情報による解析を行い、解析結果を分かりやすくユーザにフィードバックするために、会話エージェントを実装した。動作解析は AMSS 法をベースとし、プロの教示者によるダンスの軌跡との相関値を評価に用いた。会話エージェントの実装に関しては、プロの社交ダンスの教示風景を撮影し、学習者とのインタラクションの中で現れた発話・行動に対する分析を行った結果、教示者主導に学習者の身体動作に対して細分化された教示が見られた。この事が

ら、教示者の各発話を分類し、それらを各教示の状態に割り当てた。そして、状態の遷移を調べることによって教示者が行う教示の流れを示す、獲得した状態遷移教示モデルに基づいて仮想ダンスインストラクターのプロトタイプを構築した。

今後の課題として動作教示をする際に行うエージェントの動作をモーションキャプチャによってキャプチャしアニメーションとして作成したが、動作の見た目がリアリティに欠けるものとなってしまった。また、エージェントが動作するアニメーションは事前に用意したものでありバリエーションに欠ける。そのため、さらなる効果的な教示をユーザに提供するため、学習者の動作に合わせて教示動作を動的にアニメーションとして生成する必要がある。また獲得した教示モデルは学習者の発話によるインタラクションをほとんど考慮していない。今回モデルの抽出元となった教示実験が学習者の発話が少なかったためこのような教示モデルとなったが、本来、会話エージェントは活発な発話によるインタラクションを想定して開発されてきたものであり、さらに学習者の発話によるインタラクションを想定した教示モデルとならなければならない。会話エージェントにあった真の教示モデルを得るためには、さらなる実教示の観測とエージェントを用いた教示の分析を行う必要がある。

参考文献

[Chua 03] Chua, P. T., Crivella, R., Daly, B., Hu, N., Schaaf, R., Ventura, D., Camill, T., Hodgins, J., and Pausch, R.: Training for Physical Tasks in Virtual Environments: Tai Chi, in *IEEE Virtual Reality (VR 2003)* (2003)

[Goto 01] Goto, M.: An Audio-based Real-time Beat Tracking System for Music With or Without Drumsounds, *Journal of New Music Research*, Vol. 30, No. 2, pp. 159–171 (2001)

[Huang 08] Huang, H.-H., Cerekovic, A., Nakano, Y., Pandzic, I. S., and Nishida, T.: The Design of a Generic Framework for Integrating ECA Components, in Padgham, L., Parkes, D., and Muller, J. P. eds., *The 7th International Conference of Autonomous Agents and Multiagent Systems (AAMAS'08)*, pp. 128–135, Estoril, Portugal (2008), Inesc-Id

[Kipp 08] Kipp, M.: Spatiotemporal Coding in ANVIL, in *Proceedings of the 6th international conference on Language Resources and Evaluation (LREC-08)* (2008)

[Kopp 05] Kopp, S., Gesellensetter, L., Kramer, N. C., and Wachsmuth, I.: A Conversational Agent as Museum Guide - Design and Evaluation of a Real-World Application, in *Proceedings of the 5th International Conference on Intelligent Virtual Agents (IVA '05)*, Kos, Greece (2005)

[Nakamura 06] Nakamura, T., Makio, K., and Uehara, K.: Discovering and Translating Skills from Motion Data, Technical Report CS24-2006-3, Department of Computer and Systems Engineering, Kobe University, Japan (2006)

[白鳥 05] 白鳥 貴亮, 中澤 篤志, 池内 克史: モーションキャプチャと音楽情報を用いた舞踊動作解析手法, *電子情報通信学会論文誌*, Vol. J88-D-II, No. 8, pp. 1662–1671 (2005)