

デフォルメ動作と注意的言語表現を用いて人間の動作を コーチングするロボットシステムの研究 -デフォルメ度合いと言語表現の個人適応に関する考察- Motion Coaching with Emphatic Motions and Adverbial Expressions for Human beings by Robotic System -Discussion on Optimization of Emphasis and Expression for Subjects-

奥野敬丞*1

Keisuke OKUNO

稲邑哲也*1*2

Tetsunari INAMURA

*1総合研究大学院大学（総研大）

The Graduate University for Advanced Studies(SOKENDAI)

*2国立情報学研究所

National Institute of Informatics(NII)

To develop effective and intelligent human-robot interaction systems that use whole body gestures and verbal expressions, verbal expressions and gesture expressions should be strongly connected according to given tasks and current situation. In our previous work, a robotic motion coaching system has been proposed. It does not use fixed expressions, but dynamically changing emphatic motions and adverbial expressions to assist players learning motions. The emphatic motions are synthesized dynamically according to players' performance. However, we were not able to conclude how both emphatic motions and adverbial expressions effect to the HRI. Thus, in this paper, we propose a method that uses optimized degree of emphasis for each subjects. Then, we discuss how both subject-optimized emphatic motions and adverbial expressions effect to human robot interaction based on result of motion coaching of forehand tennis swing using the proposed method.

1. はじめに

全身を使用したジェスチャーと言語表現を用いた知的なヒューマンロボットインタラクションのシステムの実現には、目的とするタスクと現在の環境の認識に基づいて、言語表現とジェスチャーを適切に結びつける事が必要と考える。これに加えて、固定された表現だけではなく、インタラクションにおけるユーザの反応に応じた表現の柔軟な適応が必要と考える。例えば、ユーザの反応に応じて提示する動作や言語表現に変化を加える機能である。

また、動作や言語表現に変化を加える為にはユーザの反応の認識が必要になるので、1): 反応の認識, 2): タスクを達成する為のプランニング, 3): 反応に応じた柔軟な、提示表現の適応, 4): 1)-3) からなるインタラクションを繰り返す、という4つ要素を統合する事が重要と考える。しかしながら、従来の研究 [Fasola 10][Hoshino 04][Rose 98] [Takano 08] では、これらの必要な要素が個別に論じられていて、統合した議論がなされていないという問題があった。

そこで、これらの要素の統合を議論する為に、我々の前回の研究 [奥野 11] では、動作学習者であるプレーヤの動作学習が向上する様に、プレーヤの反応に応じて再提示する動作や言語表現に変化を加えて、人間に動作をコーチングするロボットシステムを提案した。提示動作に対する変化としては、「デフォルメ動作」を用いた。これは、プレーヤが実行した動作が目的の動作と異なっている時に、同じ目標動作を再提示するのではなく、間違っただけで学習者が実行した動作には含まれず、目的の動作に含まれている要素を強調した動作である。また、言語表現における変化としては、「注意的言語表現」を用いた。これは「more」等、注意を促す副詞表現である。そして、これら、動

作の評価、デフォルメ動作の生成、注意的言語表現の選択を一つのパラメータで統合する手法を提案した。テニスのフォアハンドの動作コーチング実験を通して、デフォルメ動作と注意的言語表現が、どのように人間とロボット間のインタラクションへ影響を及ぼすかを議論した。

以下の二点が、ロボットシステムによる動作コーチングを応用例とした理由である。一つ目の理由は、動作をコーチングする為には以下の四つの要素を満たす事が必要であり、我々が興味のある四つの要素の統合に対応すると考えるからである。1) 反応の認識: プレーヤのパフォーマンスを認識し、コーチングのターゲット動作との類似性・相違を定量的・定性的に評価する。2) プランニング: プレーヤのパフォーマンスが向上する様な動作提示と言語表現を含むフィードバックを、ロボットはプランニングする。3) 提示表現の適応: 1) の評価結果に応じて、デフォルメ動作、注意的言語表現、および、それらの結び付け方の修正をする。4) インタラクションの繰り返し: パフォーマンスを向上させる為に、1)-3) のプロセスを繰り返す事で動作コーチングが進行する。二つ目の理由は、ヒューマノイドが足を含めた人間と似た身体を有する事で、動作の共有とその言語化を可能にしえる事という点である。

我々の前回の研究では、*case1*: ターゲット動作の提示のみ、*case2*: ターゲット動作と注意的言語表現の提示、*case3*: デフォルメ動作の提示のみ、*case4*: デフォルメ動作の提示と注意的言語表現の提示、の4通りのシナリオで動作コーチングの実験を行った。しかしながら、以前の研究で提案した手法では、*case1* と *case2* 間と、*case1* と *case3* 間では、プレーヤの動作学習の向上に有意差を確認できたが、*case2* と *case4* 間と、*case3* と *case4* 間では、有意差を確認する事が出来ないという問題が残っていた。

有意差がこれらの *case* 間で見つからない問題の原因を実験とデータから考察した。その結果、プレーヤによってはデフォルメ動作に過剰反応して、逆に学習エラーが大きくなる事実がある事を確認できた。従って、この問題に取り組む為に、各

連絡先: 奥野敬丞, 総合研究大学院大学(総研大), 情報学専攻; 稲邑哲也, 国立情報学研究所/総研大, 101-8430 東京都千代田区一ツ橋 2-1-2, 03-4212-2518, {k-okuno, inamura}@nii.ac.jp

Example Synthesized Emphatic Motion

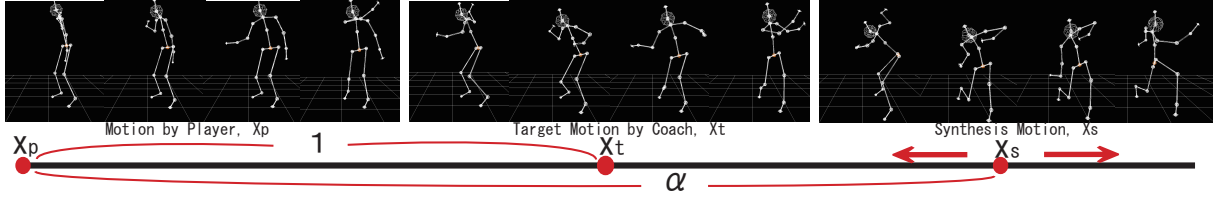


図 1: Example relationship between the weight α and a synthesis motion

プレーヤによって、デフォルメ動作に対する感度が異なるという仮説を採用する。そして、プレーヤ毎に最適化されたデフォルメ度合いを使用した動作コーチング手法を提案し、実験を通してその有用性を議論する。実験結果のプレーヤの動作学習の向上を指標として、デフォルメ動作と注意的言語表現、及び、それらの組み合わせの、HRI への影響を議論する事とする。

以下、第 2 章にて今までに提案している手法とコーチングの流れを説明する。第 3 章にて本稿での提案手法を説明し、第 4 章にて提案手法を用いた実験結果を示し、動作コーチングにおける有効因子に関して結論づける。

2. 以前の提案手法とコーチング実験

言語表現と任意の動作パターンの相互変換を実現するための手段として、原始シンボル空間法が提案されている [Inamura 03]。この手法は、動作パターンを連続隠れマルコフモデル (以下 CHMM) で抽象化した表現を原始シンボルと呼び、原始シンボル間の類似度合いを幾何学的空間での距離に対応させて、原始シンボルに対応する静止点を空間に配置することで、動作の内挿・外挿を、原始シンボルに対応する空間上の点の内分・外分操作に帰着させている [Inamura 08]。原始シンボルを幾何学的空間に配置する時に、CHMM によって抽象化された動作パターン間の類似度の尺度として、Bhattacharyya Distance [Bhattacharyya 43] を用いている。

我々の前回の研究 [奥野 11] では、この原始シンボル空間を用いた動作パターンの内挿・外挿システムを動作の強調を行う手段として応用し、HRI に重要と考える四つの要素を一つのパラメータで統合する手法を提案した。具体的には、まず、コーチングの対象となる手本動作 θ_t (関節角度の時系列データからなるベクトルで表される動作) をコーチ (本稿では仮想環境内のロボット) が提示する。ここで、手本動作 θ_t に対応する原始シンボル空間上の静止点を x_t とする。

次に、以下の 1-5 のループからなるインタラクションをとおして動作をコーチングする。

1. コーチに提示された動作 θ_c を、(初回は x_t から、以降はステップ 5 で合成される x_s から動作生成)
2. 学習者であるプレーヤが真似て動作を実行する。
3. この様子を計測し動作パターン θ_p を原始シンボル空間上の点 x_p に変換する。(詳細なアルゴリズムは文献 [Inamura 08] を参照のこと)
4. ここで、 x_p が x_t の近傍に無い場合には、修正が必要とみなし、模倣が実現できていなかった成分 $x_t - x_p$ を修正分として、

5. その修正分を x_p に加えて、 x_s を求める。

$$x_s = x_p + \alpha(x_t - x_p) \quad (1)$$

α は動作外挿の重み係数を表す。この点から生成される動作を、修正に必要なデフォルメ動作として再度提示する動作 θ_c とする。

1-5 のループ 1 回を 1 試行とする。

ここで、式 (1) の α は図 1 で示される様に、デフォルメ動作のデフォルメ度合いと注意的言語表現に用いる副詞をコントロールするパラメータである。 α を大きくすれば、デフォルメ動作の強調度合いは増し、小さくすれば強調度合いは減少する。また、負の値をとれば、プレーヤのパフォーマンスに不足している要素を補完する方向に強調する代わりに、プレーヤのパフォーマンスの間違いを強調した動作を合成する事も可能である。

前回の研究 [奥野 11] では、このシステムを用いて、まず、図 2 に結果を示した実験をプレーヤ 11 人にて行った。この実験では、「case1: 注意的言語表現なし」、 「case2: 注意的言語表現あり」の二つのケースにおいて、異なる α の値を用いたデフォルメ動作の提示に対して、1 試行での各プレーヤ毎の動作学習の度合いを調べた。そして、結果を評価する為に、被験者の学習エラー d_l

$$d_l = \|x_t - x_p\| \quad (2)$$

を導入した。ここで、 l は、実験の試行回であり、 $l = 1, 2, 3, 4$ とした。 d_l の値が小さい方が、学習エラーが少なく良い動作模倣となる。また、試行 l の時の、各被験者 i の学習エラー d_{il} の平均として、

$$\bar{d}_l = \frac{\sum_{i=1}^m d_{il}}{m} \quad (3)$$

と定義した。ここで、 i は被験者の ID 番号とした。この結果、case1 と case2 両方において、11 人の平均として最良の結果が得られる値が $\alpha = 2.0$ となる事が分かった。

次に、デフォルメ度合いとして $\alpha = 2.0$ に固定して、表 1 で示した、case1 - case4 のデフォルメ動作と注意的言語表現の四種類の組み合わせ条件で、テニスのフォアハンドスイングを初心者 15 人にコーチングする実験をおこなった。

その結果、表 2 に示された様に、4 試行後における結果からは、動作学習の向上が観測される事は確認でき、デフォルメ動作と注意的言語表現それぞれが有効である事は確認できた。しかしながら、デフォルメ動作と注意的言語表現の組み合わせが有効であるかどうかは、各 case の 4 試行目の \bar{R}_l の値の間からは、有意差が確認出来なかった。 \bar{R}_l とは、 l 回目の試行の時に、初回試行時に対する模倣エラー割合の全被験者平均であり、

$$\bar{R}_l = \frac{\sum_{i=1}^m \frac{d_{il}}{d_1}}{m} \quad (4)$$

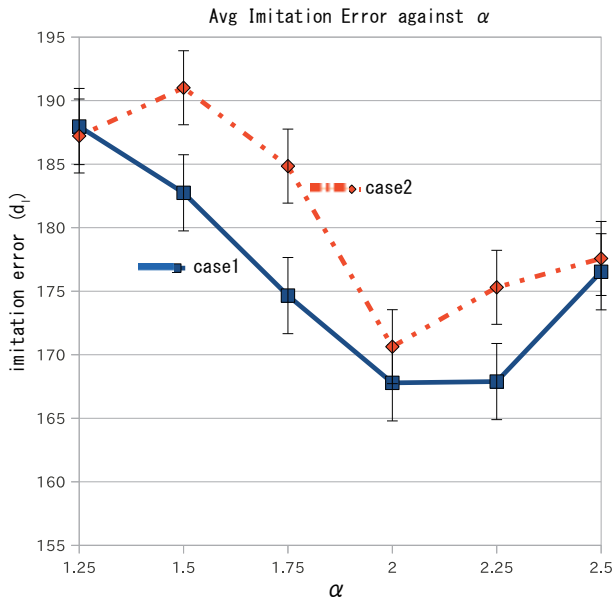


図 2: Imitation error against different values of α , as x-axis . y-axis is \bar{d}_l in Eq.(3). (\bar{d}_l adjusted, so that the maximum values of each subjects would be 200)

と定義した。同一 case において，初回試行時と比べて動作模倣のエラーが少なくなっていれば，1.0 より小さくなる。逆に，エラーが大きい時は，1.0 より大きくなる。完璧に動作を模倣できた時は，0 となる。

3. 提案手法

以前の研究では確認出来なかった，デフォルメ動作と注意的言語表現の同時提示が動作学習の向上に貢献するかを調べる為に，本稿では，デフォルメ度合いを決める手法として，平均値ではなく，各プレーヤ毎に最適化された α を使用する手法を提案する。なぜならば，各プレーヤの運動経験・運動神経や動作模倣能力の違いによって，デフォルメ動作のデフォルメ度合いに対する感度が異なるという仮説が考えられるからである。

実際に，異なるデフォルメ度合い α に対する動作学習エラー d_l (式 ()) のグラフを見た時，図 2 に示した被験者平均では， $\alpha = 2.0$ が最小の動作学習エラーをしていても，プレーヤによって最適な α が異なる事が図 3 から読み取る事ができる。例えば，図 3 がしめす様に，被験者 A は case1 の時は $\alpha = 1.5$ ，case2 の時は $\alpha = 2.0$ を使用したデフォルメ動作を用いる事が動作学習に最適である。同様に，被験者 B は case1 の時は $\alpha = 2.5$ ，case2 の時は $\alpha = 2.25$ である。

4. 実験

実験の流れは，前回の研究 [奥野 11] 同一であり，表 1 の case5，case6 で示される条件で実験を行った。実験の具体的な条件は，以下の通りである。

4.1 条件

被験者数は 8 人であった。ターゲット動作として，図 1 の Target motion で示される動作を用いた。コーチングエージェントは，壁のスクリーンに表示して動作の提示を行った。模倣ターゲット動作の表示は，常に正面からの角度に固定した。HMM で抽象化されたターゲット動作は，17 関節からなり，

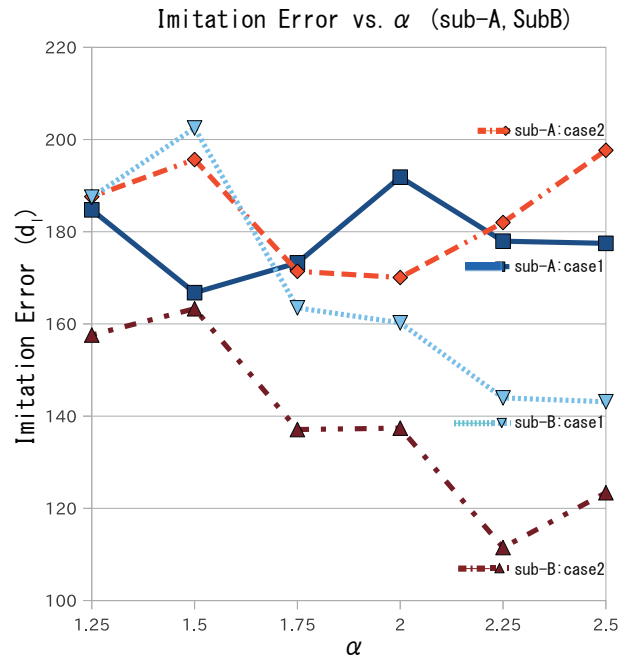


図 3: example individual imitation error against different values of α s (subject-A and subject-B): \bar{d}_l in Eq.(3)

各関節の自由度は 3 であった。テニスラケットの軌道は考慮に入れなかった。各回の試行において，HMM の学習に 5 スイングを用いて模倣動作の評価に用いた。各試行は 5 分間のインターバルで行われた。各被験者はランダムな順番で各 case の実験を行った。各 case にて各プレーヤの最適 α を用いてデフォルメ動作を合成・提示した。case5 は「注意的言語表現なし」，case6 は「注意的言語表現あり」で行った。

表 1: Cases in the Experiments

-	$\alpha = 1.0$	$\alpha = 2.0$	optimized α
without adverb	Case 1	Case 3	Case 5
with adverb	Case 2	Case 4	Case 6

4.2 結果

得られた結果を，図 4 に示した。また，各 case の第 4 試行目の \bar{R}_l の値の間で one-tailed T-test を行い，有意差があるかどうかを確認した結果が表 2 である。

表 2: one-tailed T-test with \bar{R}_l in Eq.(4) for the exp. (note: ** ($p < 0.01$), * ($p < 0.05$), + ($p < 0.10$))

-	at 4th trail (l=4)
case1 vs case2	0.00091**
case1 vs case3	0.0000079**
case2 vs case4	0.061+
case3 vs case4	0.48
case1 vs case5	0.0000062**
case2 vs case6	0.027*
case5 vs case6	0.33



図 4: Result of experiment. Average ratio of imitation error, \bar{R}_l in Eq.(4), with optimized α for each subjects

この結果から、あらたに、最適化された α を用いる事で、「デフォルメ動作と注意的言語表現の組み合わせ (case6)」と「注意的言語表現のみ (case2)」には有意差がみられた。しかしながら、「デフォルメ動作と注意的言語表現の組み合わせ (case6)」と「デフォルメ動作のみ (case5)」の間に有意差は見られなかった。ただし、前回と比べた場合、最適化された α を用いる事で T-test の値 p が、被験者数が約半数にも関わらず小さくなっているため、被験者数が増加した時に、有意差が生じるのではと予想する。

5. 結論

全身を使用したジェスチャーと言語表現を用いた知的なヒューマンロボットインタラクションのシステムの実現には、目的とするタスクと現在の環境の認識に基づいて、言語表現とジェスチャーを適切に結びつける事が必要と考える。また、固定された表現だけではなく、インタラクションにおけるユーザの反応に応じた表現の柔軟な適応が必要と考える。そこで、本稿では、プレーヤ毎に最適化されたデフォルメ度合いを実験で求める手法を提案し、以前提案したロボット動作コーチングシステムを用いて、人間に対するテニスのフォアハンドスイングの動作コーチングの実験を行った。プレーヤの動作学習の向上を指標として、デフォルメ動作と注意的言語表現の人間とロボット間のインタラクションへの影響を議論した。結果としては、「デフォルメ動作と注意的言語表現の組み合わせ (case6)」と「注意的言語表現のみ (case2)」には有意差がみられた。しかしながら、「デフォルメ動作と注意的言語表現の組み合わせ (case6)」と「デフォルメ動作のみ (case5)」の間に有意差は見られなかった。つまり、「デフォルメ動作と注意的言語表現の組み合わせ」が HRI に影響を及ぼしているかどうかを完全には結論できなかった。ただし、被験者数が少ない時点で、前回の提案手法よりも結果は向上している事から、今後被験者数が増えた時により良い結果が出ると予想している。

今後は、原始シンボル空間を構成する動作プリミティブを適切に与える事で、定性的に体のどの部位の動作を、定量的にどの程度、修正するべきかを評価した上で、デフォルメ動作の合

成を行い、それを説明するインストラクション文章の自動生成と、それらの組み合わせ方の研究を進める予定である。

また、本稿では運動情報として、関節角度のみを扱っているが、提案手法は感覚情報も考慮にいれる動作コーチングが可能になる。推測した他者の感覚情報を用いる事で、見た目の動きのみならず、感覚を表す言語表現を用いた動作コーチングが可能になる。推測した他者の感覚情報の言語化を用いたインタラクションが、より知的な HRI には重要であると考えられる。これら二つの事を実現する事で、コンセプトを示す為に現在は単純な言語表現を用いているシステムが、ユーザの癖をも含めた、よりリッチな言語表現の実現に近づくと考えている。そして、これこそがミメシスモデルを使用する事で得られる最大の利点であると考えられる。

参考文献

- [Bhattacharyya 43] Bhattacharyya, A.: On a measure of divergence between two statistical populations defined by their probability distributions, *Bulletin of the Calcutta Mathematical Society*, Vol. 35, pp. 99–109 (1943)
- [Fasola 10] Fasola, J. and Mataric, M.: Robot Exercise Instructor: A Socially Assistive Robot System to Monitor and Encourage Physical Exercise for the Elderly, in *Proc. of 19th IEEE International Symposium on Robot and Human Interactive Communication* (2010)
- [Hoshino 04] Hoshino, K.: Interpolation and extrapolation of repeated motions obtained with magnetic motion capture, *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E87-A, pp. 2401–2407 (2004)
- [Inamura 03] Inamura, T., Tanie, H., and Nakamura, Y.: From Stochastic Motion Generation and Recognition to Geometric Symbol Development and Manipulation, in *International Conference on Humanoid Robots* (2003)
- [Inamura 08] Inamura, T. and Shibata, T.: Geometric proto-symbol manipulation towards language-based motion pattern synthesis and recognition, in *Proc. of Int'l Conf. on Intelligent Robots and Systems*, pp. 334–339 (2008)
- [Rose 98] Rose, C., Cohen, M., and Bodenheimer, B.: Verbs and adverbs: multidimensional motion interpolation, *IEEE Computer Graphics and Applications*, Vol. 18, No. 5, pp. 32–40 (1998)
- [Takano 08] Takano, W. and Nakamura, Y.: Integrating Whole Body Motion Primitives and Natural Language for Humanoid Robots, in *Proc. of International Conference on Humanoid Robots*, pp. 708–713 (2008)
- [奥野 11] 奥野敬丞, 稲田哲也: デフォルメ動作と言語注意を使用したロボットシステムによる動作コーチングの研究 - ミメシスモデルによるコーチングの定量的・定性的評価の実現 -, 第 16 回ロボティクスシンポジウム予稿集, pp. 436–411 (2011)