

動作コーチングにおける手本動作の強調と言語的注意表現の統合に関する分析とモデル化

Analysis and Modeling of Integration of Emphatic Motion and Verbal Expression Usage for Motion Coaching

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)

As a step of a robotics research toward integration of motion patterns and verbal expression, in this paper, we attempt to model how human integrates motion and verbal expression for motion coaching. Through experiments of a tennis forehand swing coaching task for beginners, we observed and analyzed learning target swing performed by human coaches, swings performed by learners and emphatic motions by human coaches. We modeled relationship between similarities among each motions and used verbal expression in a phase space and discussed how the model can be applied for realizing an efficient motion coaching system.

1. はじめに

ロボットシステムが人間に動作をコーチングする時には、動作の学習者のパフォーマンスに応じて、コーチングロボットは動的に動作提示と言語表現を適切に組み合わせる事が大事である。動作のコーチングのチャレンジの一つとして、学習者が実行した動作とコーチが示す学習ターゲットである手本動作との違いを、いかに動作学習者に認識させる事ができるかという事がある。

Fasola らは、椅子に座って運動を行う高齢者に運動を指導するロボットインストラクターを開発した [Fasola 10]。このシステムは被験者の感情を評価しているが、動作のコーチングタスクでは動作の学習度合いを定量的に評価する必要がある。岩橋らは、模倣学習を通して、ロボットが動作と言語の結び付けを獲得するシステムを実現している [Iwahashi 07]。高野らは、動作プリミティブのシークエンスと文章を統計的手法を用いて結び付けるシステムを実現している [Takano 08]。これらの手法では、異なるが似ている複数の動作を一つの言語表現に結び付けている。しかし、動作コーチングでは、似ているが異なる複数の動作を、異なる言語表現に結び付ける必要がある。なぜならば、動作の学習者が実行した動作とコーチが示す学習ターゲットである手本動作との違いを認識する為には、似ているが異なる動作の微妙な違いを明確に区別して動作学習者に認識させる必要がある。

動作学習の向上を目的に動作の微妙な違いを動作学習者の人間に認識させる為に、デフォルメ動作と言語的注意表現を用いたロボットコーチングシステムを我々は以前の研究で提案した [Okuno 11]。デフォルメ動作とは、動作学習者が実行した動作がコーチが示した手本動作と異なっている時に、同じ動作を再提示するのではなく、学習者が実行した手本動作とは微妙に異なる動作には含まれず、手本動作に含まれている要素を強調した動作である。言語的注意表現とは、「more」等、注意を促す副詞表現である。このロボットコーチングシステムでは、第2

章で説明する原始シンボル空間 [Inamura 08] 上のスカラーパラメータを用いて、デフォルメ動作のデフォルメ度合いと言語的注意表現との統合をコントロールした。ロボットコーチングシステムを用いて、人間のテニス初心者にテニスのフォアハンドの動作をコーチングする実験を通して、手法の有用性を確認した。更に、個人毎に最適化されたデフォルメ動作のデフォルメ度合いが動作学習の向上因子であることを統計的な有意差をもって確認できた。言語的注意表現が動作学習の向上因子であるかは有意差をもって確認は出来なかったが、正相関がある事は確認できた [Okuno 11]。

しかしながら、デフォルメ動作のデフォルメ度合いや言語的注意表現の選択は、予備実験から経験的に筆者が与えているという問題が残っていた。従って、本稿の目的として、人間コーチがどのようにデフォルメ度合いと言語的注意表現の選択をするかのモデル化を実験的に試み、その結果を議論する事とする。動作学習者としてロボットシステムを用い、学習者の間違い度合いを定量的にコントロールした。人間コーチが、観測したロボット学習者の動作に応じて、デフォルメ動作と言語的注意表現を用いて、ロボット学習者に動作をコーチしてもらった。

以下、第2章にて動作の認識・生成手法の説明をし、第3章にて、デフォルメ動作の合成手法とデフォルメ度合いのコントロール手法を用いたコーチングシステムを簡潔に説明する。第4章にて、本稿の手法と実験結果を示す。最後に、まとめ、議論、今後の課題を記述する。

2. 動作パターンの認識とデフォルメ動作の生成方法

原始シンボル空間法 [Inamura 08] を応用して動作の認識、内挿・外挿動作の生成、複数動作間の類似度分析、動作へのラベル付けを1つの枠組みで実現できる。

2.1 動作パターンの認識方法

動作パターンのデータベースを $D = \{M_1, \dots, M_i\}$ (i は動作パターンのインデックス) と定義する。原始シンボル空間 (PSS) を表す P は、構築プロセス $P = F_{build}(D)$ により、 D から構築され、各 M_i に対応した静止点 x_i からなる。この構築プロセス F_{build} では、まず、この動作パ

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 X_s

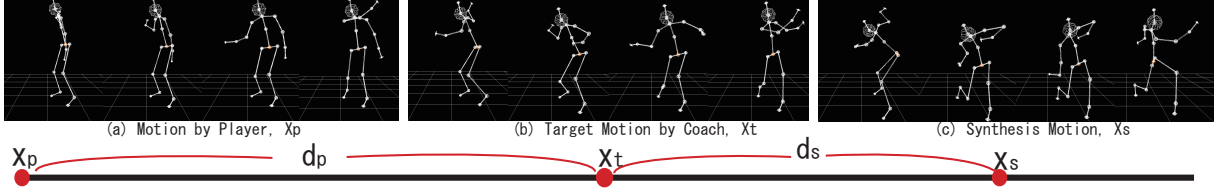


図 1: 動作例を用いた, α とデフォルメ動作, x_s と x_t 間の距離 d_s , x_p と x_t 間の距離 d_p の関係性. (a): 動作学習者の動作 x_p , (b): コーチによる手本動作 x_t , (c): デフォルメ動作 x_s (デフォルメ度合い $\alpha = 2.0$, $\alpha = (d_p + d_s)/d_p$, $d_p = |x_t - x_p|$, $d_s = |x_t - x_s|$)

ターンを表す行列 $M_i = [\theta[0]\theta[1]\dots\theta[t]]$ を, left-to-right 連続隠れマルコフモデル ($CHMM_i$) を用いて抽象化する. ($\theta[t] = [\theta_1[t]\dots\theta_j[t]\dots\theta_n[t]]^T$ は, 関節角度の時系列パターンを示すベクトルであり, j は総関節数 n における関節のインデックス) その次に, 全ての $CHMM_i$ 間の Bhattacharyya 距離 [Bhattacharyya 43] を計算し, PSS 内の静止点 x_i としての配置を決める. D に存在しない動作パターン M_x の認識も同様に, まず, Baum-Welch アルゴリズムを用いて認識対象の M_x を $CHMM_x$ に抽象化する. 次に, D 内の全動作パターン M_i に対応する PSS 内の全 $CHMM_i$ と認識対象の M_x の抽象化表現である $CHMM_x$ との間の Bhattacharyya 距離を計算し, PSS 内の静止点 x_x としての座標を決める. つまり, PSS 内の静止点 x_x の座標が動作の認識結果であり, 他の静止点 x_i との距離が動作パターン間の類似度を表す. このような動作パターンの認識プロセス F_{recog} を,

$$x = F_{recog}(M) \quad (1)$$

と表す事とする.

2.2 デフォルメ動作の生成手法

原始シンボル空間内の複数の静止点の内分・外分操作をする事で, D がない新規の内挿・外挿動作 M_s の生成が可能である. 新規に求めた内分・外分点 x_s から動作生成機能,

$$M_s = F_{gen}(x_s) \quad (2)$$

を用いて M_s を生成できる [Inamura 08].

これを応用する事で, 以下の式 (3) のように, スカラーパラメータ α を用いて, デフォルメ動作 M_s をコントロールする事が可能になる.

$$x_s = x_p + \alpha(x_t - x_p) \quad (3)$$

x_p と x_t は, それぞれプレーヤが実行した模倣動作 M_p とコーチが提示した模倣の手本動作 M_t に対応する PSS 内の静止点である. x_s から式 (2) を用いてデフォルメ動作 M_s を生成できる. 図 1 に, 動作パターンのデフォルメ度合いと α の関係を示した. M_s とは, M_t にあるが M_p に欠けている要素を M_p に加えた動作である. 例えば, M_p の膝の曲げ具合が少ない時は, M_s は膝の曲げが強調されたデフォルメ動作になる.

3. 前回提案した動作コーチングシステム

動作学習者であるプレーヤのパフォーマンスの分析結果に基づいて, デフォルメ動作のデフォルメ度合いを 1 つのスカラーパラメータ表現でコントロールする手法を提案した [Okuno 11].

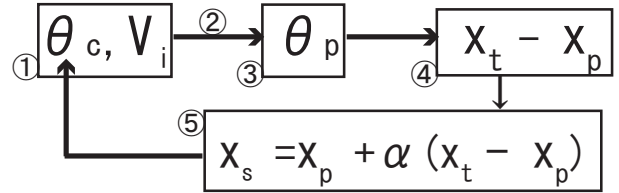


図 2: デフォルメ動作と言語的注意表現を用いた動作コーチングの流れ. θ_c は人間/ロボットコーチによる手本動作. V_i は用いられた言語的注意表現. θ_p は学習者である人間/ロボットプレーヤによる動作. x_s は人間/ロボットコーチによるデフォルメ動作. α はデフォルメ度合いを表す

デフォルメ動作のコントロールは, 第 2 章で説明した式 (3) を使用した手法を用いた. 言語的注意表現のコントロール手法としては, 「more like this」という言語的注意表現を使用するかしないかというプリミティブな手法を用いた. 単純な手法である理由は, 第 1 段階として, このような言語的注意表現がどう動作の学習に影響するかを確認する為であった. デフォルメ動作と言語的注意表現の結び付け方法は, デフォルメの有・無と言語的注意表現の有・無の全ての組み合わせる手法を試して, 異なる組み合わせが動作学習にどう影響するかを実験で調べた. 実験的に, 今後のシステムにおいて, 何をどうパラメータ表現する事が有効な結び付け手法なのかの知見を得る事を目的とした.

提案した手法を用いてロボットコーチングシステムを準備した. 初めに, 1 つの手本動作 θ_t と 1 つのプレーヤの模倣動作 θ_p からなるデータベース $D = \{\theta_t, \theta_p\}$ を準備した. 次に, D から第 2.1 章で簡略に説明した F_{build} を使用して, P で表された原始シンボル空間を構築した. ここで x_t を手本動作 θ_t に対応した PSS 内の静止点と定義する. 図 2 に図示したように, 以下の手順で動作コーチングのタスクを行う.

1. コーチ (仮想環境内のアバターを壁に投影) は, 手本動作として θ_c を動作提示した. 一回目の試行では x_t を用いて, 二回目以降はステップ (5) で合成したデフォルメ動作 x_s から, 式 (2) を用いて θ_c を生成する.
2. 人間のプレーヤは動作 θ_c を模倣する.
3. コーチはプレーヤの模倣動作 θ_p を観測し, 式 (1) を用いて P 内の静止点 x_p に変換する.
4. x_p が x_t に近くなければ, プレーヤの模倣動作は不完全とみなし, コーチはプレーヤ動作の不足要素 $x_t - x_p$ を計算する.

5. コーチは式(3)を用いて, 不足要素 $x_t - x_p$ を x_p に加えて外分点 x_s を計算する. x_s と式(2)から, 次の試行で使用する手本動作 θ_c を生成する.

プロセス(1)-(5)のループを必要に応じて繰り返す. 実験では, 1ループを1試行とした.

このロボットコーチングシステムによるデフォルメ動作の出力例を, 図1-(c)として示した. 図1-(b)は, コーチによる模倣ターゲット動作 x_t である. x_s の提示目的は, プレーヤ自身が間違いを修正して x_t を可能な限り正確に模倣する事を手助けする事である. 図1-(a)は, プレーヤが実行した模倣動作 x_p であり, それは不完全である. 例えば, 左腕の使い方は x_t とは異なる. 図1-(c)に示されたデフォルメ動作 x_s を見てみると, 左腕の動きのみならず, 右膝の曲げ具合等も不足要素として, x_s では曲げ具合が強調されている. このようなデフォルメ動作を提示する事でプレーヤに不足していた膝の曲げや腕の使い方を意識させる事ができる. これは今まで議論されていなかった, 原始シンボル空間法による内挿・外挿動作の実世界のタスクにおける有用性を定性的に示した事になる.

4. 人間コーチの手本動作と言語表現の分析とモデル化

前回の研究 [Okuno 11] では, 予備実験をとおして経験的にデフォルメ度合い α を決め, 言語的注意表現に関しては使用するかしないかの初歩的なものであったという問題が残っていた. 本稿では, 効率の良いコーチングを実現する, 最適な α の値と言語的注意表現の選択を動作学習者のパフォーマンスに応じてシステムが自動的に決定できるようになる事を目的にモデル化を試みる. テニススイングのコーチを題材として, 手本動作, 学習者が行った動作, コーチが強調した手本動作(デフォルメ動作)およびコーチが用いた言語表現を観察・分析した. 各動作間の類似性と使用された言語表現の関係性を, 位相空間上でモデル化を試みる.

実験として, 図1-(a)で例示した動作学習者の動作を, 予め用意した複数種類の動作からランダムに10種類選び, ロボットプレーヤに実行させた. これらのロボットプレーヤの動作 x_p を, テニスの素人である人間コーチに観測してもらい, x_p と手本動作 x_t を見比べて, 適切と考えるデフォルメ動作 x_s と言語的注意表現 V_i (i はインデックス)を人間コーチに実行・使用してコーチをしてもらった. 被験者はテニスの初心者である16人とした. 図3に実験の流れをしめす. 実験では, 図3のステップ2として, ロボットプレーヤは人間コーチを動作を模倣する代わりに, ランダムに手本動作とは多少異なる動作を実行した. 人間コーチの言語的注意表現の使用方法の分析とモデル化の為に, 3種類の言語的注意表現 $V_1 = \text{"little-bit-more"}$, $V_2 = \text{"more"}$, $V_3 = \text{"much-more"}$ から使用する言語表現を選んでもらい記録した.

実験により人間コーチの手本動作の強調方法を分析しモデル化を目的に $d_p = \|x_t - x_p\|$, $d_s = \|x_s - x_t\|$ として,

$$\alpha = 1 + \frac{d_s}{d_p} \quad (4)$$

と α を定義した. 図5で示される d_s と d_p の関係を示す回帰直線の傾きの値を, この定義式(4)の $\frac{d_s}{d_p}$ と示さる比の部分に代入する事で, α の値を求める, つまり, 人間コーチの手本動作の強調方法のモデル化を試みる.

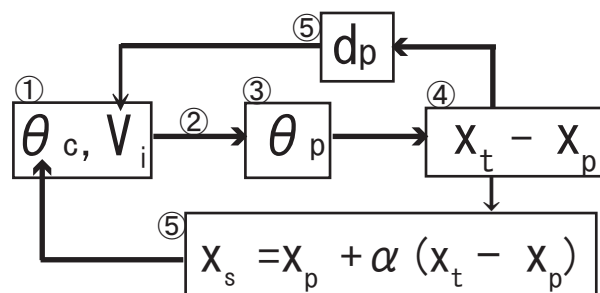


図3: モデル化した α と V_i の決定法を用いたデフォルメ動作と言語的注意表現を用いた動作コーチングの流れ. 図2との違いは, x_t と x_p 間の距離 d_p を用いて V_i を決定し, d_p を用いて式(4)から α 決定する事

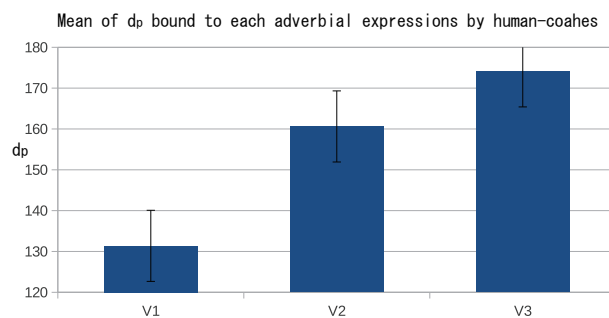


図4: 実験結果: 得られた言語的注意表現と x_p と x_t 間の距離 d_p との関係性

実験の結果, 言語的注意表現の使用と d_p の間には図4で示された関係を確認できた. V_1, V_2, V_3 それぞれに対応する d_p の値の平均値の間には統計的に有意差がある事を確認できた. 従って, 図4で示された d_p と V_i の関係から, d_p の値に応じて V_i を選択するモデル化が可能になる.

図5に, ロボットプレーヤの動作 x_p と手本動作 x_t との距離 d_p に対して, 人間コーチが使用したデフォルメ動作 x_s と x_t との距離 d_s の関係を示した4人の被験者の結果例を示した. 図5-(1)や, 図5-(2)の例のように多くのデータが回帰直線に乗る場合, 図5-(3)のように殆ど乗らない場合, 図5-(4)のように殆どならず, かつ, 回帰直線の傾きが負である, という3パターンの結果が得られた. 回帰直線の傾きの値を式(4)の d_s と d_p の比の部分に代入して α の値が計算した. α の値としては, 0.75~1.46 が得られた. この結果は, 前回の研究 [Okuno 11] で使用した経験的に選んだ1.25~2.00という値とある程度整合性がある事を確認できた.

人間コーチが使用するデフォルメ動作のモデル化という観点から結果を分析すると, 使用するデフォルメ度合い α は個人毎で異なる事が示唆された. しかしながら, 各被験者毎に観測されたデータが必ずしも回帰直線上に綺麗に乗る事を確認する事はできなかった. また, 動作学習者のパフォーマンスが良かった場合は, 修正が少なくても良いはずであり, デフォルメ動作のデフォルメ度合いは小さい値になると予想されたが, そうならない事が多かった. これらの原因として, テニスの素人を人間コーチの被験者としている事が考えられる. コーチや上級者と比較してテニスのスイングが上手には出来ないのに, デフォルメ動作を用いて修正する事は簡単なタスクではないと推測される. この問題に対応すべく, 今後はテニス上級者やテニ

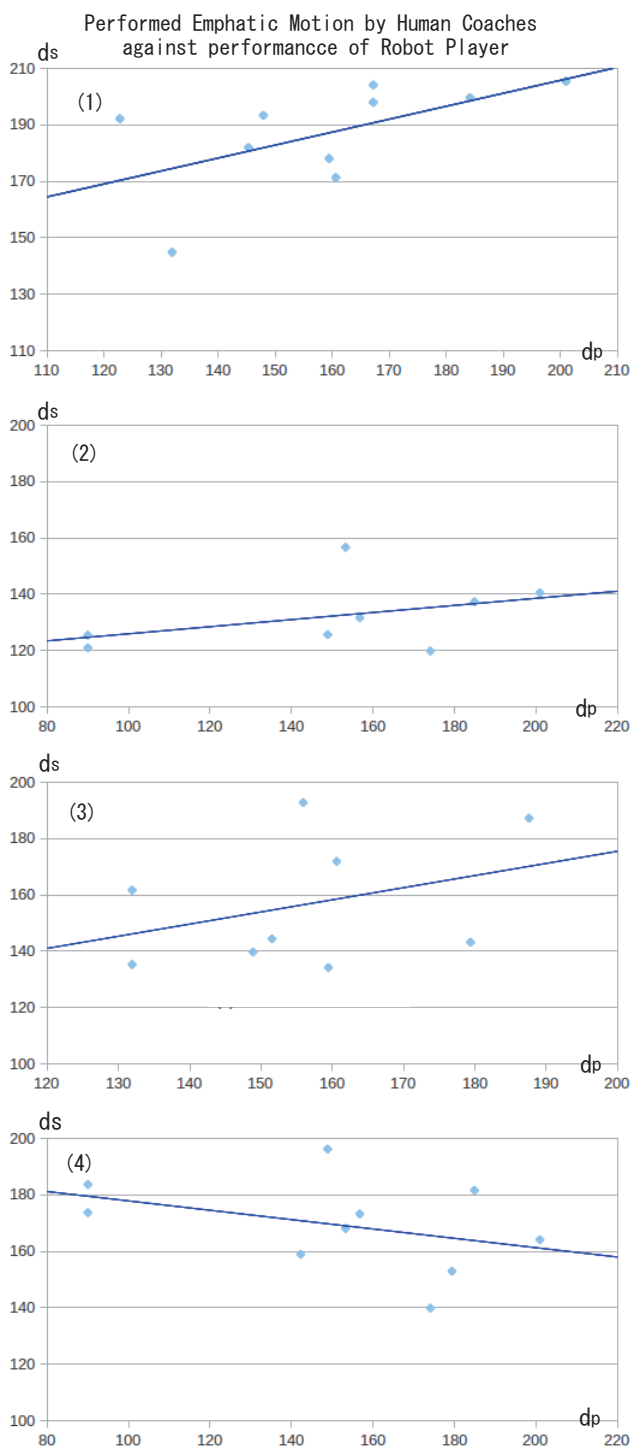


図 5: 実験結果 (抜粋): x 軸- x_t と x_p 間の距離 d_p . y 軸- x_s と x_t 間の距離 d_s . グラフ上の回帰直線の傾きは式 (4) の比の部分 $\frac{d_s}{d_p}$ に代入して α を求める為に使用

スコッチ経験者を被験者にして実験を行いモデル化を試みる .

5. 結論

本稿では, 身体運動と言語表現の統合を目標とした研究の第一段階として, コーチングに用いられる動作と言語表現の統

合法のモデル化を目指した . テニススイングのコーチを題材として, 手本動作, 学習者が行った動作, コーチが強調した手本動作およびコーチが用いた言語表現を観察・分析した . 各動作間の類似性と使用された言語表現の関係性を, 位相空間上でモデル化し, 効率の良いコーチングを実現するための応用について議論した . その結果, 運動学習者のパフォーマンスとコーチが使用する言語的注意表現の選択のモデル化は可能である事を図 4 で示された実験結果で示すことができた . しかしながら, 運動学習者のパフォーマンスとコーチが使用するデフォルト動作のモデル化ができる事は確認する事が出来なかった (図 5) . 理由として, 実験データがきれいに回帰直線上にのる結果が多くは得られていないからである . 原因として考えられる事は, 動作学習者のスイングと手本動作のスイングの微妙な違いを認識する事, そして, 2つのスイング間の違いからデフォルト動作を実行する事は, テニスの素人にとって簡単なタスクではないという事がある . コーチのモデル化のためには, テニス素人ではなくテニス上級者やテニスコーチの経験者を被験者にする事が必要であると考えられる .

本稿では運動情報として, 関節角度のみを扱っているが, 用いた手法では運動情報に対応した感覚情報も同時に原始シンボル空間に抽象化する事が可能である . 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, 感覚情報を表す言語表現を用いた動作コーチングが可能になる . 今後は, 動作と直接観測が不可能な重心の移動と言語統合した手法を取り組む予定である . 運動と感覚情報と言語表現を統一的に扱う事が原始シンボル空間法を使用する事で可能になると考える .

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 in Robot and Human Interactive Communication* (2010)
- [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)
- [Iwahashi 07] Iwahashi, N.: *Robots That Learn Language: A Developmental Approach to Situated Human-Robot Conversations*, chapter Nilanjan Sankar (ed.), Human-Robot Interaction, pp. 95-118, I-Tech Education and Publishing (2007)
- [Okuno 11] Okuno, K. and Inamura, T.: Motion Coaching with Emphatic Motions and Adverbial Expressions for Human beings by Robotic System -Method for Controlling Motions and Expressions with Sole Parameter-, in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3381-3386 (2011)
- [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)