3B1-OS22c-5

人とロボットの対話インタラクションにおける頭部動作効果の考察

Effects of Head Motion during Human-Robot Conversation Interaction

劉 超然[*1*2*3]
Chaoran Liu

石井 カルロス寿憲[*1*3]
Carlos T. Ishi

石黒 浩[*1*2*3]
Hiroshi Ishiguro

[*1] ATR 知能ロボティクス研究所
ATR/IRC Labs

[*2] ATR 石黒浩研究室
ATR/HIL

[*3] 大阪大学大学院基礎工学研究科 [*4] 独立行政法人科学技術振興機構, CREST
Graduate School of Engineering Science, Osaka University JST, CREST

this paper proposes a model for generating head tilting and nodding based on rules inferred from analyzing the relationship between head motion and dialogue acts, and evaluates the model using two types of humanoid robot (one very human-like android, “Geminoid F”, and one typical humanoid robot, “Robovie R2”). Subjective scores show that the proposed model including head tilting and nodding can generate head motion with increased naturalness compared to nodding only or directly mapping people’s original motions. We also find that an upwards motion of a robot’s face can be used by robots which do not have a mouth in order to provide the appearance that utterance is taking place.

1. はじめに

人とロボットの間で音声対話を介して円滑なコミュニケーションが成立するためには，言語情報の理解以上に，発話意図や話者の態度・感情などのパラ言語情報の理解も重要となる．

人間同士の対話では，発話をする際，自然に頭部動作が伴う．頭部動作は何らかの意図を示すため意識的に行う場合がある．特に話し相手の発言に対するリアクションとして頻繁に用いられる，例えば頷くことにより，同意や肯定を示し，首を横に振ることにより，否定を示す．一方，多くの場合は，発話に伴い，無意識に頭部動作が生じる．無意識に生じる頭部動作の場合も，発話となんらかの関連が存在すると考えられる．これらの頭部動作は態度，感情などのパラ言語情報を伝達する．

人間の頭部動作に含まれるパラ言語情報を完全に解明するのは難しいと思われる．しかし，自然な頭部を真似することによって，ロボットの存在感が高まり，より高度なヒューマンロボットコミュニケーションが期待出来る．

我々の研究のモチベーションとして，人型ロボット（アンドロイドなど）の遠隔操作において，音声に伴う頭部動作を音声信号から自動的に生成することを目的としている．

[1]では，数名の日本語ネーティブスピーカーの会話を分析し、発話機能（発話権の転換など）と頭部動作との関連を調べた．頷きは発話区間に最もよく見られる頭部動作であって，相槌や同意・肯定の意思を表すだけでなく，強い句境界の最後の音節でもよく生じる．話者が考えている場合や次の発話を準備している弱い句境界においては，首傾げがよく見られる．これらの結果に基づいて，ルールベースの頷きモデルが提案され，その効果を検証された．

この研究では，[1]の継続として，付加的な首傾げモデルが提案した．女性型アンドロイド“Geminoid F”とヒューマノイド“Robovie R2”を用いて，提案モデルの有効性を確かめた．

また，発話のサインとして，発話区間に顔を軽く上げる動作を追加した．口の持たない（動かない）ロボットが発話時の不自然さを低減できることが確認された．

劉超然，大阪大学大学院基礎工学研究科，大阪/ATR 知能ロボティクス研究所，京都/JST CREST，京都，chaoran.liu@irl.sys.es.osaka-u.ac.jp
石井カルロス寿憲，ATR 知能ロボティクス研究所，京都/JST CREST，京都，carlos@atr.jp
石黒浩，大阪大学大学院基礎工学研究科，大阪/ATR 石黒浩研究所，京都/JST CREST，京都，ishiguro@sys.es.osaka-u.ac.jp

2. 関連研究

頭部動作に関する研究は主に二つの問題に重心を置かれた．一つはコミュニケーションに置いての頭部動作とその役割の識別（e.g. [2-3]），もう一つは音声発話に同期した自然な頭部動作の生成．この研究は二つ目の問題に注目した．

複数の言語に置いて，頭部動作と韻律情報の相関関係の分析された[4-6]．例えば，英語に置いて頷きは単語の強調とよく伴う．声調の上昇とともに顔を上げるのはよく見られる[4]．スウェーデン語の読み上げ文に対し，強調された単語において，顔パラメータの分散が大きかったと報告されている[5]．しかし，音声の基本周波数(F0)と頭部動作の相関は英語において 0.39 から 0.52 に対し，日本語の場合は 0.22 から 0.30 であることが[6]で報告されている．故に F0周波数から自然な頭部動作を生成することは難しいである．

また，岩野ら[7]は，視覚情報を利用して対話理解を向上させることを目的とし，日本語の対話音声における頭部動作の役割を分析している．発話権や発話意図が考慮され，肯定・同意・応答・相槌では縦方向の動作，相手に応答を求める場合は顔を上げる動作が頻繁に生じることを報告している．

[8]では，頭部動作と発話機能の間，及び発話機能と韻律情報の間の強い相関関係が報告されている．この研究では頭部動作と発話機能の相関関係に注目する．

[9]で提案された発話機能タグに基づいて，スピーカーの発話に対し，フレーズに分けられ，それぞれフレーズに相応のタグが付与された．使った発話機能タグのリストを下に表す．

k (keep)：発話権の保持（強い句境界）
k2 (keep)：弱い句境界（発話権の保持）
k3 (keep)：発話末を伸ばし、発話の途中であることを表現（発話権の保持）
f (filler)：「えっとー」「あのー」など、考え中であることを表現する感動詞
f2 (conjunctions)：「じゃ」などの接続詞（短いフィラーとして捉えられる）
g (give)：対話相手への発話権の譲渡
q (question)：発話権の譲渡（対話相手に応答を求める場合）
bc (backchannels)：「うん」「はい」などの相槌を表現する感動詞
su (surprise/admiration)：「えー！」「へー」など、驚きや感心などの感情を表現する感動詞
dn (denial)：「いいえ」「ううん」などの否定を表現する感動詞

前述のように，音声発話の韻律情報と頭部動作の強い相関関係が見られない．が，発話機能タグと頭部動作の関係が報告された．[1]によると，頷きは最もよく見られる頭部動作であって，よく相槌(**bc**)や強い句境界(**k,g,q**)に現れる．そして，話者が考えている・次の発話を準備しているなど語尾を延ばした弱い句境界(**f,k3**)には首傾げ動作がよく生じる．この関連関係をベースとした頷き・首傾げ生成モデルを提案し，このあとのセッションに詳しく紹介する．

3. 首かしげの生成モデル

3.1 首傾げの生成

頭部動作と発話機能の分析結果[1][8]が示したように，人間の自然な動作の中に，弱い句境界において，頭部が動かない傾向が強い．しかし，話者が発話文の文末音素を伸ばしている，つまり考えていると思われるところ(発話機能タグの k3,f)に首傾げは高い確率で見られる．このセッションでは，新たな首傾げ生成モデルを紹介する．

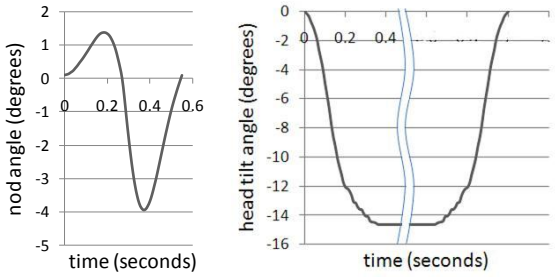

Figure 1 Representative nod and head tilt shapes used in the head motion generation model.

首傾げモデルは，既存の頷きモデルに対し，特定の弱い句境界(k3,f)に，首傾げ動作を加えたモデルである．ベースは[1]で提案された頷きモデルとなり，まず強い句境界(k,g,q)において頷きを生成させる．それに加え，データベースから首傾げのモーションサンプルを利用し，弱い句境界の k3,f に加えた．この動作の特徴は，始まりに 0.4 秒を掛けて，頭部の roll 軸を元の位置から 15 度まで回転し，終わりは同様のスピードで元の位置まで戻す．生成タイミングの制御は句境界に首傾げを開始させ，その後傾げる角度を保ち，次のフレーズの終了時刻までに頭を元の位置に戻す．生成モデルに使用した頷きと首傾げ動作の角度をFigure 1に示す．

3.2 実験設定

ヒューマノイドの Robovie R2 とアンドロイドの Geminoid F を用いて，首傾げ生成モデルによって生成した頭部動作の自然さを検証する実験を行う．

データベースからランダムに 11 の弱い句境界 k3・f を含む 10 秒から 20 秒発話サンプルを抜き出し，頷き生成モデル(NOD ONLY)と首傾げ生成モデル(NOD&TILT)を用いて，発話毎に 2 種類の動作を生成した．比較対象として，モーションキャプチャーシステムによって記録した話者のオリジナルモーション(ORIGINAL)を加えて，3 種類の動作パタンをヒューマノイド Robovie R2 と女性アンドロイド Geminoid F に再生させた．使用したロボットの写真は下に示す．

Figure 2 Robovie R2 and Geminoid F

話者のオリジナルモーションをロボットにマッピングする際，Robovie R2 は首に 3 つの回転自由度を持っていて，線形的にマッピングすることができる．アンドロイド Geminoid F へのマッピングは，[10]で提案された手法を利用する．Figure 3に使用した二つのロボットの頭頚部アクチュエータ配置を示す．ロボットに動作指令を送る時間間隔は，ロボットのハードウェアの制限に従って，Robovie R2 の場合は 100ms で，Geminoid F の場合は 20ms．Robovie R2 では口が動かないため，頭部の回転のみを再生するには 100ms の時間間隔でも支障は生じないと考えられる．Geminoid F 口の動きはモーションキャプチャーが撮った話者の鼻と顎に貼り付けたマーカーの距離の変化から求められる．

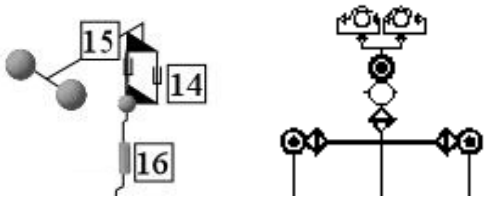

Figure 3 Head actuators for Geminoid F and Robovie R2.

41 人の被験者（男性 17 人，女性 24 人）を招いで，11 サンプル・各 3 パタンの動作，計 33 の動作を 2 種類のロボットに再生させ，発話のビデオを被験者に見せ，アンケートを取った．

比較させたペアは以下に示す

- 頷きモデル vs. 首傾げモデル
- 首傾げモデル vs. オリジナル

- 頷きモデル vs. オリジナル

ペア内の二つのビデオの順番は一定ではなく, ランダムになっている.

アンケートは 7 段階評価の様式にした.

- ロボットの動作が自然か否か

大変自然(7) | 自然(6) | やや自然(5) | どちらとも言えない(4) | やや不自然(3) | 不自然(2) | かなり不自然(1)

- 二つの動作を比較して, どちらが自然?

一つ目の方がずっと自然 | 一つ目の方が自然 | 一つ目の方がやや自然 | どちらとも言えない | 二つ目の方がやや自然 | 二つ目の方が自然 | 二つ目の方がずっと自然

3.3 実験結果

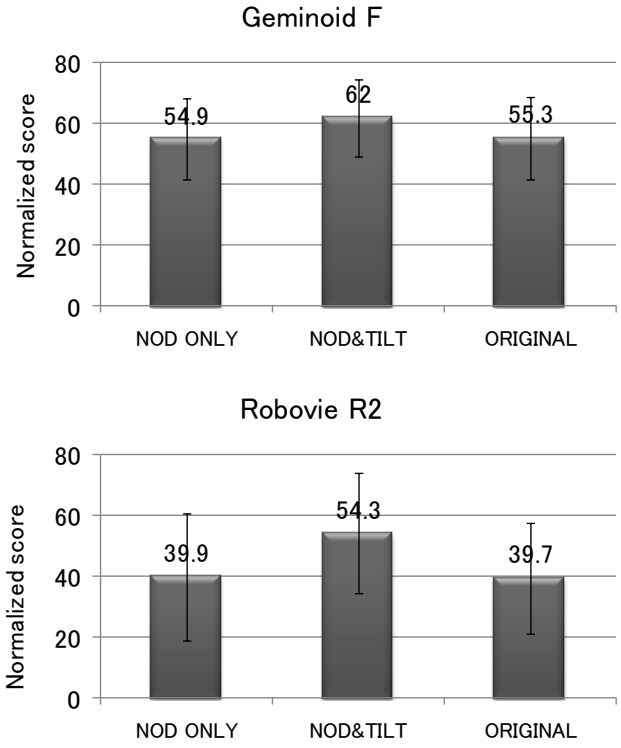

Figure 4 Distributions of the preference scores for each motion type, for Geminoid F (top) and Robovie R2 (bottom)

Figure 4は自然さに関するアンケート結果を括弧の中の数字で数値化し, その値を 0-100 の範囲内に標準化した結果を示す. 二つのロボットともに自然さに関するスコアが最も高いのは提案の首傾げモデルである. この結果は首傾げを弱い句境界の k3,f で生成することで, ロボットの動きが自然に見えることを示した. 首傾げのタイミング制御の正確性の裏づけになると思われる. そして, [1]で報告された頷きモデルの検証実験の結果と同じく, オリジナルモーション(ORIGINAL)が単純な頷きモデル(NOD ONLY)の生成モーションとほぼ同様のスコアになっている.

Figure 5は各ペア内二つの動作のどちらの方が自然という問題に対して, 各選択肢を選んだ人数のパーセンテージを表す. 真ん中の黒い部分は“比較できない”と答えた人数のパーセンテージを表す. 例えば, 左のグラフは約 25%の被験者が“比較できない”と答えた, それに 50%以上の被験者に首傾げモデル(NOD&TILT)のほうが自然・やや自然だと判断された.

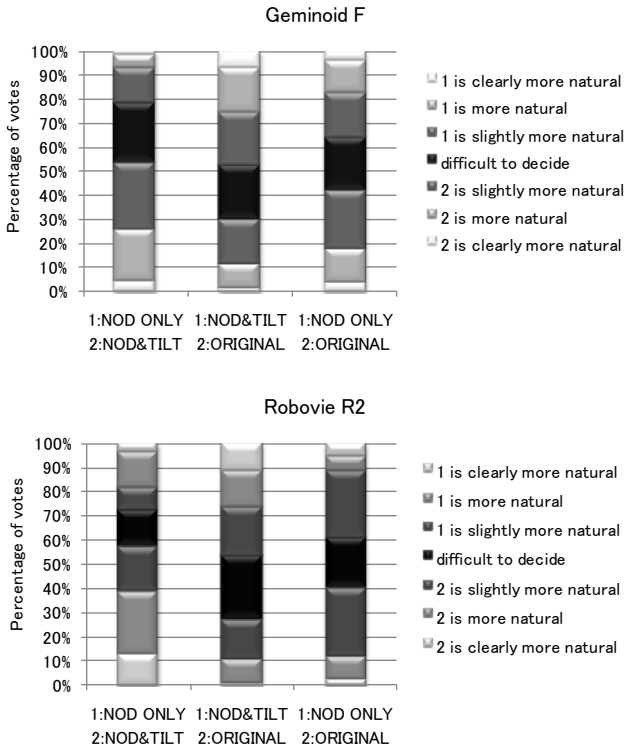

Figure 5 Distributions of the preference scores for each pair of motion types, for Geminoid F (top) and Robovie R2 (bottom)

Geminoid F における結果は, 頷きモデル(NOD ONLY)と首傾げモデル(NOD&TILT)の比較で多くの被験者が首傾げモデルの方が自然またはやや自然という答えを選んだ. 首傾げモデル(NOD&TILT)とオリジナルモーション(ORIGINAL)の比較で, オリジナルモーション(ORIGINAL)の方が自然と答えた被験者もいるが, 首傾げモデル(NOD&TILT)の方が自然と答えた被験者ほど多く無い. 首傾げモデル(NOD&TILT)の方が(やや)自然だという結果になった. 頷きモデル(NOD ONLY)とオリジナルモーション(ORIGINAL)の比較で, 各答えを選んだ被験者数の分布がほぼ対称的であった. 自然な動作は必ずしも人間そっくりの動きではないことがこの結果からわかる. これらの結果は各動作パタンの自然さの評価スコアとも一致した. 複雑な共振除去制御を施さない条件では, シンプルで且つタイミング正しい動作生成モデルのほうが効果的だという結論に繋がる.

Robovie R2 における結果は, 各答えを選んだ被験者人数が違うが, 全体として Geminoid F においての評価結果の分布と似た結果になる.

3.4 議論

二つのロボットに同様の動作を再生したものの, 全体的に自然さに関するスコアが明らかにヒューマノイド Robovie R2 の方が低かった. ロボットの外見による差が出ていると思われるが, もう一つ大きな理由として, Robovie R2 の口が動かないことだと考えられる. Geminoid F の方が口の動きが人間話者の口の動きを再現したため, ロボットが発話する際, 被験者が視覚情報・聴覚情報両方から認識することができる. 一方, Robovie R2 の場合は, 被験者はロボットが発話をしているというはっきりした視覚情報を得るのが難しい. 聴覚情報から発話を検知しても, 最初に述べたように, 頭部動作と音声の韻律情報の関連性が低いため, 視覚的に頭部の回転だけからロボットが発話をしているという再確認が困難である. 聴覚情報だけでロボットの存在感が十分に

伝えなく，それが故に Robovie R2 での自然さに関する評価が低くなっていると考えられる．

4. 口の持たないロボットのための発話サイン

前セッションで議論した口の持たないロボット（例えば，Robovie R2）の発話の不自然さの問題を解決するべく，私たちが発話サインとして，発話区間に軽く顔を上げる動作（3°）を提案した．この顔を上げる動作は自然な人同士の対話にもよく見られ，ロボットの発話の自然さの向上させることが期待される．

4.1 実験設定

セッション 3 に説明した NOD ONLY モデルと NOD&TILT モデルに，発話区間に顔を 3° 上げるという動作付加し，NOD ONLY+と NOD&TILT+という生成モデルを作成した．

セッション 3.2 に紹介した実験と同じ方法で，8 の発話サンプルを無作為に選出し，それぞれの二つの動作パタン(NOD ONLY+, NOD&TILT+)を作り出し，ロボットに再生した．

計 16 の動作サンプルをビデオに記録し，20 代から 30 代の 1 被験者 10 名に同じ発話サンプルの違う動作を見せて，評価アンケートを取った．

4.2 実験結果

Figure 6に示した数値は前セッションにも示したロボットの動きが自然かどうかに対する被験者の 1 から 7 までの 7 段階の評価を 0-100 区間にノーマライズした結果．

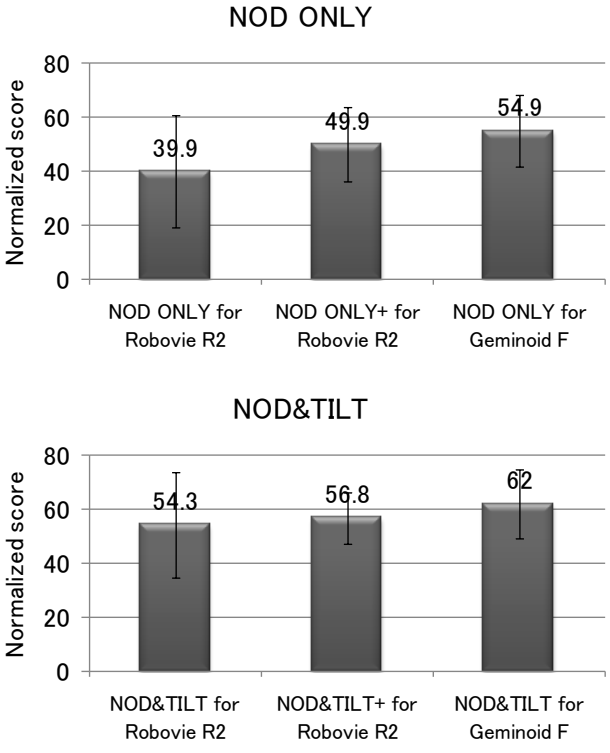

Figure 6 Subjective naturalness for NOD ONLY+ and NOD&TILT+, compared with NOD ONLY model and NOD&TILT model.

比較として二つの動作のベースとなる NOD ONLY と NOD&TILT の評価結果も用いた．真ん中のグラフは顔を上げる動作を加えた生成モデルの評価結果，これに対して，左のグラフと右のグラフはそれぞれベースモデルが Robovie R2 と Geminoid F においての評価結果に対応する．

発話音声に同期した口の動きの代わりとして，発話区間の顔を上げる動作は口の動きと同じ働きが期待される．

Figure 6の結果から，NOD ONLY と NOD&TILT 二つの動作生成モデルにおいて，発話区間の顔上げを加えることによって，自然さの評価が増していることがわかる．

しかし，発話サインを加えた動作の評価結果が Geminoid F においての評価結果に追いつくことができなかった．その理由として，Geminoid F は非常に人間そっくりのロボットであって，Robovie R2 のようなヒューマノイドとの外見の差がより自然だと評価された原因の一つと思われる．

5. おわりに

人とロボットの自然な対話インタラクションを実現するには，ロボットも発話に伴って自然な頭部動作を行うことが重要である．

本研究では，人間の対面対話における頭部動作の分析結果に基づいて，ルールベースの首傾げ生成モデルを提案した．これらの生成モデルを二種類の人間型ロボットに応用し，評価実験によってその自然性を確認した．

また，発話区間の顔を上げる動作が提案され，発話のサインとして見なすことが可能であって，口の持たないロボット（例えば，Robovie R2）の発話がより自然になることが確認された．

今後の予定は，韻律・言語情報から発話中の句境界の種類の推定を検討し，本研究の生成モデルを利用し，音声情報のみから自然な頭部動作の生成を試みる．

謝辞 本研究は，JST CREST の委託により実施したものである。

参考文献

1) C.T. Ishi, C. Liu, H. Ishiguro, N. Hagita, “Head motion during dialogue speech and nod timing control in humanoid robots,” Proc. of IEEE/RSJ Human Robot Interaction (HRI 2010), 293-300, 2010.
2) C. Sidner, C. Lee, L.-P. Morency, C. Forlines, “the effect of head-nod recognition in human-robot conversation,” Proc. HRI 2006, pp. 290-296, 2006.
3) L.-P. Morency, C. Sidner, C. Lee, and T. Darrell, “Head gestures for perceptual interfaces: The role of context in improving recognition,” Artificial Intelligence, 171(8-9): 568-585, June 2007.
4) H.P. Graf., E. Cosatto, V. Strom, F.J. Huang, “Visual prosody: Facial movements accompanying speech,” Proc. IEEE *Int. Conf. on Automatic Face and Gesture Recognition* (*FGR'02*), 2002.
5) J. Beskow, B. Granstrom, D. House, “Visual correlates to prominence in several expressive modes,” Proc. *Interspeech 2006 – ICSLP*, pp. 1272-1275, 2006.
6) H.C. Yehia, T. Kuratate, E. Vatikiotis-Bateson, “Linking facial animation, head motion and speech acoustics,” *J. of Phonetics*, Vol. 30, pp. 555-568, 2002.
7) Y. Iwano, S. Kageyama, E. Morikawa, S. Nakazato, K. Shirai, “Analysis of head movements and its role in spoken dialogue,” Proc. *ICSLP'96*, pp. 2167-2170, 1996.
8) C.T Ishi, J. Haas, F.P. Wilbers, H. Ishiguro, and N. Hagita, “Analysis of head motions and speech, and head motion control in an android,” Proc. of *IROS 2007*, 548-553, 2007.
9) C.T. Ishi, H. Ishiguro, N. Hagita, “Analysis of prosodic and linguistic cues of phrase finals for turn-taking and dialog acts,” Proc. *Interspeech'2006 - ICSLP*, pp. 2006-2009, 2006.
10) F.P. Wilbers, C.T. Ishi, H. Ishiguro, “A blendshape model for mapping facial motions to an android,” Proc. IROS 2007, 2007.