Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Closing the Loop: Towards Tightly Synchronized Robot Gesture and Speech

Maha Salem, Stefan Kopp, Frank Joublin

2013

Preprint:

This is an accepted article published in International Conference on Social Robotics (ICSR). The final authenticated version is available online at: https://doi.org/[DOI not available]

Closing the Loop: Towards Tightly Synchronized Robot Gesture and Speech

Maha Salem[1], Stefan Kopp[2], and Frank Joublin[3]

[1] Technical Faculty, Bielefeld University, Germany
msalem@cor-lab.uni-bielefeld.de
[2] Center of Excellence Cognitive Interaction Technology, Bielefeld University, Germany
skopp@cit-ec.uni-bielefeld.de
[3] Honda Research Institute Europe, Offenbach, Germany
frank.joublin@honda-ri.de

Abstract. To engage in natural interactions with humans, social robots should produce speech-accompanying non-verbal behaviors such as hand and arm gestures. Given the special constraints imposed by the physical properties of a humanoid robot, successful multimodal synchronization is difficult to achieve. Introducing the first closed-loop approach to speech and gesture generation for humanoid robots, we propose a multimodal scheduler for improved synchronization based on two novel features, namely an experimentally fitted forward model and a feedback-based adaptation mechanism. Technical results obtained with the implemented scheduler demonstrate the feasibility of our approach; empirical results from an evaluation study highlight the implications of the present work.

Keywords: Multimodal Interaction and Conversational Skills, Robot Gesture, Speech-Gesture Synchrony, Social Human-Robot Interaction.

1 Introduction

Speech-accompanying hand and arm gestures are a fundamental component of human communicative behavior, often used to illustrate what is expressed in speech [12]. At the same time, human listeners are very attentive to information conveyed via such non-verbal behaviors [7]. Thus, gestures represent ideal candidates for extending the communicative capabilities of humanoid robots that are to interact socially with humans. In addition, providing multiple modalities helps to dissolve ambiguity and thus to increase robustness of communication.

In human communication, co-verbal gestures are tightly connected to speech as part of an integrated utterance, yielding semantic, pragmatic and temporal synchrony between both modalities [11]. A hand gesture typically consists of three phases ordered in a sequence over time to form a so-called gesture phrase: *preparation, stroke*, and *retraction* [7]. The *stroke* is the expressive phase of the gesture and bears meaning in relation to the corresponding part of speech, which is referred to as the *affiliate* [12]. Based on studies of human communication (e.g. [7,11]), gestures are found to be tightly synchronized with the accompanying linguistic affiliate to convey their intended communicative meaning: the gesture

G. Herrmann et al. (Eds.): ICSR 2013, LNAI 8239, pp. 381–391, 2013.
© Springer International Publishing Switzerland 2013

stroke typically precedes or ends at – but does not follow – the stressed syllable of the affiliate. To achieve this phonological synchrony, gesture may adjust to speech, e.g. by means of adequately timed preparation or additional hold phases, or speech may adapt to gesture, e.g. by means of pauses [7].

Ensuring such co-expressive synchrony between the two modalities poses a major challenge for artificial communicators such as virtual agents or social robots. Up to the present time, most technical systems dedicated to co-verbal gesture generation merely approximate crossmodal synchronization, typically by means of one modality adjusting to the other without any mutual adaptability. However, relevant literature (e.g. [2]) emphasizes the importance of accurate speech-gesture synchrony even for artificial agents, since asynchrony may result in conveying incorrect meaning.

2 Related Work

An increasing number of robotic systems incorporate both speech and gesture synthesis; however, in most cases the robots are equipped with a set of predefined gestures or even prerecorded behaviors that are not generated on-line but simply replayed during human-robot interaction (e.g. [18]). Crucially, to the best of our knowledge, no previous approach in robotics has realized exact (i.e. empirically sound), closed-loop synchronization of robot gesture and speech. For example, the system by Mead et al. [13] roughly attempts to synchronize the robot's gesture with speech at the phrase level rather than with the exact affiliate. Bremner et al. [1] define timings for speech-gesture synchronization based on estimations at the coding stage, i.e. gesture movements are preplanned in form and duration; during execution, movements of each gesture phases are then triggered by events in the speech output. Generally, in all existing robotic systems that aim to synchronize synthetic speech and gesture, running speech dictates the timing of the generated gestures, while gestures cannot affect the produced speech. While some approaches obtain exact speech timing information from the text-to-speech system, timing of gesture is typically based on estimations which lack accuracy. Another characteristic of related approaches is the open-loop control of the implementations, i.e. once speech and gesture have been planned and scheduled for the robot, both modalities are generated ballistically and cannot be further adjusted at run-time, e.g. based on sensory feedback. Such unidirectional synchronization as well as the open-loop production of multimodal behaviors are characteristic of all currently existing approaches to robot gesture generation (e.g. [14,10]). As a result, the present work stands out from other robotic systems for speech-gesture generation by incorporating reactive closed-loop feedback for a more flexible approach to multimodal robot expressiveness.

In contrast to the research area of robotics, the challenge of generating speech and co-verbal gesture has already been tackled rather extensively in various ways within the domain of embodied virtual agents (e.g. [2,3,15]). However, despite their elaborate architectures, these system are typically limited with regard to timing and synchronization of multimodal behaviors: to achieve crossmodal synchrony, gesture timing is heuristically adapted to the timing of generated speech,

i.e. start and end times of phonemes are used to parametrize gesture time. In addition, speech and gesture are executed ballistically, i.e. no feedback-based adjustments can be made once behavior execution has been initiated.

Complementing these approaches, Kopp and Wachsmuth [8] introduced the Articulated Communicator Engine (*ACE*) as the first multimodal behavior realizer for virtual agents that provides for mutual adaptation mechanisms between speech and gesture. Inspired by theories from human gesture research, ACE assumes that the production of continuous speech and gesture is organized in successive segments. Each of these segments forms a multimodal *chunk* of speech-gesture production which, in turn, comprises a pair of an intonation phrase and a co-expressive gesture phrase [8]. That is, complex utterances with several gestures consist of multiple successive chunks. Given such a multi-chunk utterance, ACE coordinates the concurrent modalities at two levels. First, for *intra-chunk* scheduling, similar to the method used in the above discussed virtual agent systems, temporal synchrony is accomplished by adapting gesture to the timing of speech. Once scheduled and ready to be executed, speech and gesture run ballistically within a chunk, unaffected by the progress of the other respective modality. Second, for *inter-chunk* scheduling between two successive chunks, both speech and gesture can anticipate the forthcoming chunk and adapt to it if required. For gesture, this adaptation may range from the insertion of an intermediate rest pose to a direct transition movement skipping the retraction phase or parts thereof. For speech, a silent pause may be inserted between two intonation phrases for the duration of the preparation phase of the following gesture. Such adaptive inter-chunk synchrony is achieved by a global scheduler which plans the following chunk in advance while monitoring the chunk currently in execution. In summary, ACE overcomes some of the issues found in previous frameworks by providing a greater level of flexibility. However, despite mutual adaptation mechanisms at the inter-chunk level, its lacking adjustability within a chunk as well as the ballistic generation of complete gesture and intonation phrases conflict with findings from psychology (e.g. [7]).

The present work thus draws inspiration from the innovative quality of the ACE system regarding its incremental on-line scheduling of multimodal behavior at the *inter-chunk level*. Crucially, we expand the model of the underlying scheduler by countering its conceptual shortcomings at the *intra-chunk level* to realize more tightly synchronized robot gesture and speech. The concept of the proposed scheduler using the Honda humanoid robot[1] is presented in the following.

3 Closed-Loop Scheduler

An improved multimodal behavior scheduler needs to address two questions. First, in the behavior *planning* phase, how can a more reliable prediction of gesture execution time be obtained to allow for better multimodal scheduling? Second, in the *execution* phase, how can planning and scheduling errors be accounted for during execution if they would cause mistimed synchrony otherwise? Addressing the first issue, the proposed scheduler incorporates an empirically verified

[1] `http://asimo.honda.com/downloads/pdf/honda-asimo-robot-fact-sheet.pdf`

forward model that predicts an estimate of the gesture preparation time required by the robot prior to actual execution Addressing the second issue, an on-line adjustment mechanism was integrated into the synchronization process for cross-modal adaptation within a chunk based on afferent sensory feedback. Generation and continuous synchronization of gesture and speech are flexibly conducted at run-time. Fig. 1 illustrates the 'planning-execution' procedure of the scheduler.

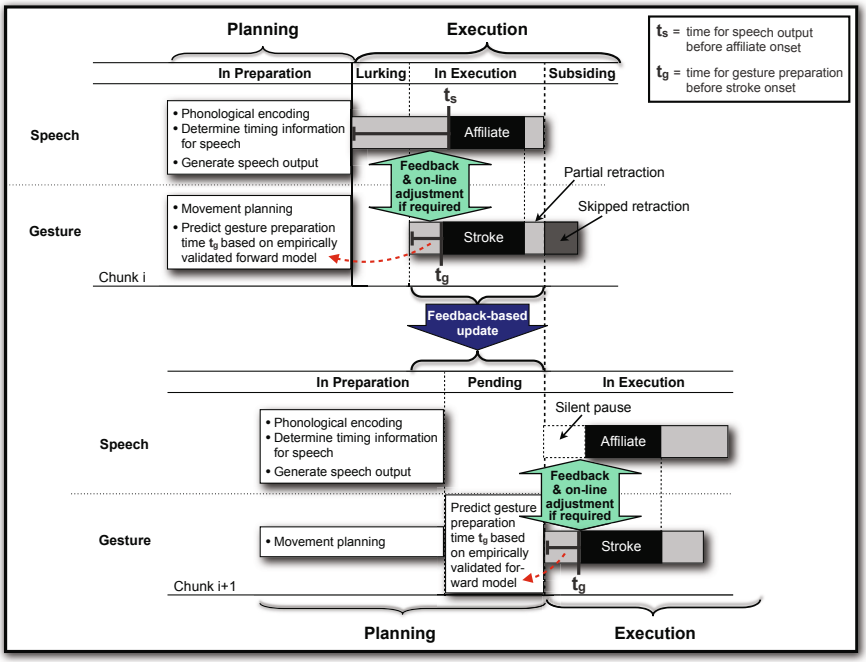

Fig. 1. Model of 'planning-execution' procedure of the proposed multimodal scheduler

3.1 Concept Overview

We specify multimodal utterances using the XML-based Multimodal Utterance Representation Markup Language (MURML [9]). It provides specific tags to label the speech affiliate based on time identifiers, while the posture of the gesture stroke can be specified in an overt form. An example MURML file is illustrated in Fig. 2. the proposed scheduler operates as follows.

Phase 1: Planning

Speech preparation. The planning phase begins with the phonological encoding of the designated speech output. During this process, the text-to-speech synthesis system MARY [17] establishes a complete list of phonemes and their respective durations, based on which the affiliate onset time is determined. Speech output is then generated in an audio file which is further processed depending on the position of the affiliate. If the affiliate is located at the beginning of the utterance, the audio file remains as it is. If the affiliate is in the middle or at the end of the utterance, the sound file is split into two parts: the first part contains speech

```
<definition>
 <utterance>
  <specification>
    <time id="t1"/> This one <time id="t2"/> is my favorite picture.
  </specification>
  <behaviorspec>
   <gesture id="gesture_1" scope="hand">
    <affiliate onset="t1" end="t2"/>
     <constraints>
      <parallel>
        <static slot="HandShape" value="BSffinger"/>
        <static slot="ExtFingerOrientation" value="DirA"/>
        <static slot="PalmOrientation" value="DirL"/>
        <static slot="HandLocation" value="LocShoulder LocRight LocNorm"/>
      </parallel>
     </constraints>
   </gesture>
  </behaviorspec>
 </utterance>
</definition>
```

Fig. 2. Example of a MURML specification for multimodal utterances

output to be uttered before the affiliate onset; the second part contains the speech affiliate and, if applicable, any subsequent remaining parts of speech.

Gesture preparation. Based on the specification of the designated location for the gesture stroke onset, a suitable trajectory is calculated, resulting in a detailed movement plan. Meanwhile, the predictive forward model further described in Section 3.2 computes the estimated execution time required by the robot for the gesture preparation phase before the stroke onset. Following a comparison of derived timing information for both modalities, start times for speech and gesture are determined so that temporal synchrony is expected to be achieved at the affiliate-stroke level.

Phase 2: Execution

Before the affiliate onset. Speech-gesture production begins as scheduled in the final step of the planning phase. If speech output is to precede the affiliate, the first sound file containing this part is replayed. While the robot performs the gesture preparation movement, variance between target and actual wrist position is constantly monitored utilizing afferent feedback from the robot controller.

Ensuring synchrony between affiliate and stroke onset. Once the robot's wrist has reached a position within a predefined range of the target location for the stroke onset, playback of the sound file containing the speech affiliate is triggered. If the predicted time was accurate, there should be no noticeable interruption in speech flow. If gesture preparation takes longer than scheduled, the feedback-based adaptation mechanism described in Section 3.3 becomes effective.

3.2 Predictive Forward Model

Neurophysiological findings suggest that human motor control relies more on sensory predictions than on sensory feedback, as feedback loops are considered too slow for efficient trajectory control [4]. Drawing inspiration from this neurobiological perspective, the concept of internal forward models is of particular interest to the realization of the proposed scheduler. Given the robot's physical properties and inability to move its arms with arbitrary speed, precise movement times are difficult to determine a priori. However, to successfully synchronize the

robot's gesture with concurrent speech, the gesture preparation phase must be completed before the onset of the speech affiliate. Thus, the execution time of this gesture phase in particular needs to be estimated as accurately as possible so that the behavior onset times of the two modalities can be adequately scheduled.

Our approach to a robot-specific forward model was realized using the whole body motion (WBM) software controlling the Honda humanoid robot [6]. The robot's WBM controller allows for accelerated internal simulations of designated trajectories prior to – and temporally decoupled from – the actual movement process [5]. The WBM-based prediction model can account for the actual path of the simulated trajectory with regard to the robot's physical properties, e.g. joint limits. To simulate the robot's future state, the internal predictor iterates the robot model, but ten times faster than real-time control and at negligible computation time. As a result, it offers a viable option for the estimation of movement time required for the gesture preparation phase. The predicted trajectory time t_g is influenced by several parameters specific to the WBM software, such as speed response and maximal velocity. Thus, the controller-based forward model was fitted to determine appropriate values for these parameters. For this, a range of different values was empirically tested for each parameter on a set of experimental training data comprising 60 different target trajectories within a typical range of gesture space, and the combination yielding the smallest overall mean error was selected (see [16] for more details). For the given training data, the mean prediction error of this combination was 0.1458 seconds with a minimum deviation of 0.0024 seconds for the best predicted trajectory time, and a maximum deviation of 0.6180 seconds for the poorest trajectory time prediction.

3.3 Feedback-Based Adaptation Mechanism

Despite the improved accuracy of timing estimation for the gesture preparation phase based on the implemented forward model, actual timing of multimodal utterances might still deviate from the prediction during execution. Thus, the second feature realized in our proposed scheduler provides a feedback-based adaptation mechanism that allows for reactive crossmodal adjustment during the execution phase (see Fig. 1). Generally, two types of real-time deviation from scheduled timing plans are possible: the forward model may either *overestimate* or *underestimate* the time required for gesture preparation, resulting in either a premature or delayed gesture stroke onset. Findings from human gesture research state that the gesture stroke may precede but not follow the speech affiliate [11], thus the first type of deviation seems less problematic than the second. In fact, for the 60 tested training trajectories, the mean error values obtained for the WBM-based predictive model yielded a maximal overestimation error of only 0.276 seconds with a rare general occurrence of errors of more than 0.1 seconds (only 7 out of 60 cases); therefore, such cases may well be tolerated.[2] But given the model's inherent tendency to underestimate the time required by the robot, our approach focuses on this more problematic type of prediction error.

[2] If, however, the model exposed a pattern of frequent overestimation of gesture preparation time, the implementation of a pre-stroke hold phase would be reasonable.

Findings from human gesture research (e.g. [7]) demonstrate that speech may pause before the affiliate onset if gesture preparation has not yet been completed, ensuring temporal synchrony of the gesture stroke with speech. To account for the human ability to align speech to gesture also within a chunk, our proposed scheduler replaces the commonly used open-loop execution mechanism with a more flexible closed-loop approach. As outlined in Section 3.1, reactive on-line adjustment of the multimodal synchronization process within a chunk is achieved by utilizing afferent feedback from the robot's WBM controller as follows. Sensory feedback received from the controller provides information about the current wrist position of the robot's arm at each time step, which enables the scheduler to constantly check for variance between the target and actual position. This information is used to trigger the speech affiliate once the robot's hand has reached a position within the range of the stroke onset position.

If gesture preparation time is overestimated or correctly predicted, feedback about successful completion of this phase is transmitted to the speech processing module before or at the very latest at the scheduled affiliate onset time. As a result, speech output is uttered according to the previously established generation plan without any audible disruption. If, however, gesture preparation time is underestimated, the described crossmodal processing dependency results in a deviation of the actual multimodal production process compared to the scheduled behavior plan. That is, if the robot's hand has not yet reached the stroke onset location by the predicted time, the reactive feedback mechanism intervenes in the speech production process: if the affiliate is located at the beginning of the utterance, speech onset is delayed; if the affiliate is in the middle or at the end, a speech pause is inserted right before the affiliate. To re-establish speech-gesture synchrony in the latter case, speech is paused until sensory feedback confirming the completion of the gesture preparation phase triggers its continuation.[3] Such anticipation and reactive adaptation of the flow of speech to gesture timing complies with psycholinguistic models from gesture literature promoting an interactive view of the relationship between speech and gesture (e.g. [12]).

4 Technical Results

The following examples were generated using the proposed multimodal scheduler incorporating both the experimentally fitted WBM controller-based predictive model and the reactive adaptation mechanism. Ideally, the movement time required for gesture preparation is estimated precisely enough during the planning phase so that the actual execution is consistent with the scheduled multimodal plan and the speech production is not notably interrupted. However, to demonstrate the scheduler's crossmodal adjustment capability, we provide examples in which sensory feedback induced the insertion of pauses into the flow of speech.

[3] Alternatively, speech rate could be modulated (e.g. 'slowed down') until gesture preparation has completed, however, this is difficult to realize with a TTS engine that preproduces the audio files.

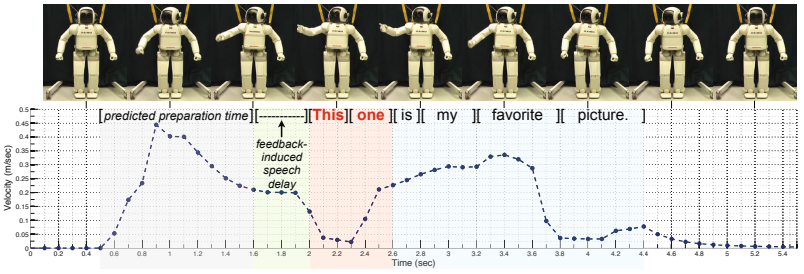

Fig. 3. Multimodal utterance generated with the proposed multimodal scheduler demonstrating a feedback-induced speech delay prior to speech onset

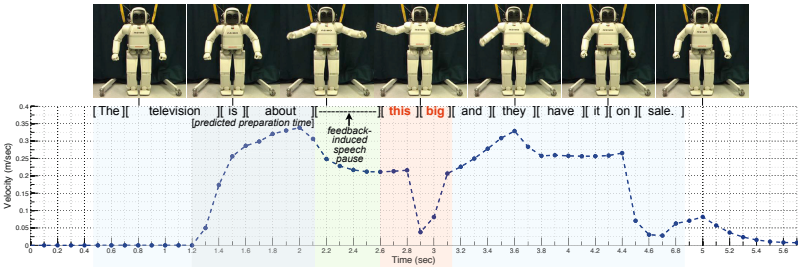

Fig. 4. Multimodal utterance generated with the proposed multimodal scheduler demonstrating a feedback-induced speech pause within a chunk

Such cases emerge when the gesture preparation time is considerably underestimated by the predictive model during planning.

Affiliate at the Beginning of a Chunk. Fig. 3 depicts an example of a multimodal utterance that illustrates the feedback-based adaptation mechanism effective at the beginning of the utterance. The plotted graph visualizes the velocity profile of the robot's right wrist during gesture execution; speech output is transcribed in temporal alignment to the generated gesture trajectory with words of the affiliate highlighted in red color. The figure illustrates the underestimation of gesture preparation time by ∼0.4 seconds resulting in a feedback-induced delay of the speech affiliate onset until the stroke onset position has been reached.

Affiliate in the Middle of a Chunk. Fig. 4 shows an example of a multimodal utterance which illustrates the feedback-based adaptation mechanism effective in the middle of the chunk. Initially, speech and gesture commence according to their assigned onset times as scheduled in the planning phase. In this case, playback of the first audio file containing speech to be uttered before the affiliate onset is initiated first and is accompanied by gesture movements after ∼0.72 seconds. Gesture preparation time is underestimated, resulting in a feedback-induced speech pause immediately preceding the affiliate onset. Once the preparation phase of the gesture has been completed, the feedback mechanism triggers

the continuation of speech so that the affiliate and remaining speech output are replayed from the second audio file.

5 Evaluation

To evaluate the optimized generation of synchronized co-verbal robot gesture with our new scheduler, we conducted an empirical study with 20 participants (10m, 10f), ranging in age from 19 to 36 years ($M = 28.5$, $SD = 4.53$). Each participant was invited to view and rate a set of video clips showing the Honda humanoid robot while delivering various multimodal utterances (speech was generated in German language for the study, as all participants were German native speakers). For each utterance, two different versions were presented in randomized order: one version was recorded using the original ACE scheduler, while the other one employed the new extended scheduler. Participants were asked to identify the version in which they found the robot's gesture to be better synchronized with the speech output. Evaluation was based on twelve video sequences covering pairs of six different multimodal utterances (including the two examples presented in Section 4); in three of these utterances the speech affiliate was located at the beginning of the chunk, while the affiliate in the other three utterances was situated in the middle of the chunk. Participants were allowed to watch the video sequences as many times as needed to come to a decision.

In summary, participants showed a significant preference for the multimodal robot behavior generated by our proposed scheduler, $\chi^2(1) = 5.63$, $p = .018$, thus attesting a more positive perception of the speech-gesture synchronization resulting from the new system. These findings not only validate our technical approach but also contribute new insights into the role of speech-gesture synchrony by providing an HRI perspective for the systematic study of human perception of (a)synchrony in robotic agents. Despite the limited scope of the present evaluation, the results suggest that humans may indeed be aware of differences in the quality of synchronized multimodal behaviors generated by a humanoid robot.

6 Conclusion and Future Work

We presented a novel multimodal scheduler that overcomes the conceptual shortcomings of existing behavior schedulers based on two features. First, the scheduler incorporates a biologically inspired forward model that predicts a more accurate estimate of the robot's gesture preparation time prior to actual execution. Future work should extend the forward model to incorporate machine learning algorithms for a more accurate and, from a neurobiological perspective, more intuitive prediction model. The second feature of the extended scheduler incorporates an on-line adjustment mechanism into the synchronization process for crossmodal adaptation not only between two successive chunks, but also within a single chunk based on sensory feedback. In future, other strategies to modulate speech (e.g. using vocal timing instead of pauses) should be explored.

The functionality of the reactive adaptation mechanism provided by our scheduler at the intra-chunk level was illustrated by means of two examples. They show how synchrony between robot gesture and speech is ensured during execution in spite of prediction errors previously made during behavior planning and scheduling. In this way, the extended closed-loop scheduler enables the robot to plan, generate and continuously synchronize gesture and speech at run-time.

The realized scheduler was evaluated in a video-based user study. Our results show significantly higher ratings regarding better synchronization for the multimodal behaviors realized with the new scheduler compared to the utterances generated by the system that does not incorporate the extended features. The findings further suggest that humans are capable of perceiving differences in the quality of speech-gesture synchrony based on utterances generated by a humanoid robot. Since human perception of speech-gesture (a)synchrony has not yet been sufficiently investigated with regard to communicative robots, this observation should be systematically studied in more exhaustive studies in the future. Our generation system will provide a useful tool for this purpose.

Acknowledgement. The work described was supported by the Honda Research Institute Europe and the Center of Excellence Cognitive Interaction Technology.

References

1. Bremner, P., Pipe, A., Melhuish, C., Fraser, M., Subramanian, S.: Conversational Gestures in Human-Robot Interaction. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, pp. 1645–1649 (2009)
2. Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., Yan, H.: Human Conversation as a System Framework: Desigining Embodied Conversational Agents. In: Embodied Conversational Agents, pp. 29–63. MIT Press, Cambridge (2000)
3. Cassell, J., Vilhjálmsson, H., Bickmore, T.: BEAT: the Behavior Expression Animation Toolkit. In: Proceedings of ACM SIGGRAPH 2001 (2001)
4. Desmurget, M., Grafton, S.: Forward modeling allows feedback control for fast reaching movements. Trends in Cognitive Sciences 4(11), 423–431 (2000)
5. Gienger, M., Bolder, B., Dunn, M., Sugiura, H., Janssen, H., Goerick, C.: Predictive Behavior Generation – A Sensor-Based Walking and Reaching Architecture for Humanoid Robots. In: Autonome Mobile Systeme 2007, pp. 275–281 (2007)
6. Gienger, M., Janßen, H., Goerick, S.: Task-Oriented Whole Body Motion for Humanoid Robots. In: Proceedings of the IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan (2005)
7. Kendon, A.: Gesture: Visible Action as Utterance. Gesture 6(1), 119–144 (2004)
8. Kopp, S., Wachsmuth, I.: Synthesizing Multimodal Utterances for Conversational Agents. Computer Animation and Virtual Worlds 15(1), 39–52 (2004)
9. Kranstedt, A., Kopp, S., Wachsmuth, I.: MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents. In: AAMAS 2002 Workshop on Embodied Conversational Agents - Let's Specify and Evaluate Them (2002)
10. Le, Q.A., Hanoune, S., Pelachaud, C.: Design and implementation of an expressive gesture model for a humanoid robot. In: Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots, Bled, Slovenia, pp. 134–140 (2011)

11. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, Chicago (1992)
12. McNeill, D.: Gesture and Thought. University of Chicago Press, Chicago (2005)
13. Mead, R., Wade, E., Johnson, P., St. Clair, A., Chen, S., Matarić, M.: An Architecture for Rehabilitation Task Practice in Socially Assistive Human-Robot Interaction. In: Proceedings of IEEE RO-MAN 2010, pp. 404–409 (2010)
14. Ng-Thow-Hing, V., Luo, P., Okita, S.: Synchronized Gesture and Speech Production for Humanoid Robots. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4617–4624 (2010)
15. Niewiadomski, R., Bevacqua, E., Mancini, M., Pelachaud, C.: Greta: An Interactive Expressive ECA System. In: Proceedings of 8th International Conference on Autonomous Agents and Multiagent Systems, pp. 1399–1400 (2009)
16. Salem, M.: Conceptual Motorics – Generation and Evaluation of Communicative Robot Gesture. Dissertation. Logos Verlag, Berlin (2012)
17. Schröder, M., Trouvain, J.: The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. International Journal of Speech Technology, 365–377 (2003)
18. Sidner, C., Lee, C., Lesh, N.: The Role of Dialog in Human Robot Interaction. In: International Workshop on Language Understanding and Agents for Real World Interaction (2003)