Lifelike Gesture Synthesis and Timing for Conversational Agents

Ipke Wachsmuth, Stefan Kopp
Artificial Intelligence Group
Faculty of Technology
University of Bielefeld
D-33594 Bielefeld, Germany
{ipke,skopp}@techfak.uni-bielefeld.de

Contribution for GW2001 proceedings – draft, July 2000

Abstract. Synthesis of lifelike gesture is finding growing attention in human-computer interaction. In particular, synchronization of synthetic gestures with speech output is one of the goals for embodied conversational agents which have become a new paradigm for the study of gesture and for human-computer interface. In this context, this contribution presents an operational model that enables lifelike gesture animations of an articulated figure to be rendered in real time from representations of spatiotemporal gesture knowledge. Based on various findings on the production of human gesture, the model provides means for motion representation, planning, and control to drive the kinematic skeleton of a figure which comprises 43 degrees of freedom in 29 joints for the main body and 20 DOF for each hand. The model is conceived to enable cross-modal synchrony with respect to the coordination of gestures with the signal generated by a text-to-speech system.

1 Introduction, Previous Work, and Context

Besides the inclusion of gesture recognition devices as an intuitive input modality, the synthesis of lifelike gesture is finding growing attention in human-computer interface research. In particular, the generation of synthetic gesture in connection with text-to-speech systems is one of the goals for embodied conversational agents which have become a new paradigm for the study of gesture and for human-computer interface. Embodied conversational agents are computer-generated characters that demonstrate similar properties as humans in face-to-face conversation, including the ability to produce and respond to verbal and nonverbal communication. They may represent the computer in an interaction with a human or represent their human users as "avatars" in a computational environment [1].

The overall mission of the Bielefeld AI lab is interacting with virtual reality environments in a natural way. Three things have been important in our previous work toward incorporating gestures as a useful input modality in virtual reality: (1) measuring gestures as articulated hand and body movements in the context of speech; (2) interpreting them by way of classifying features and transducing them to an application command via a symbolic notation inherited from sign language; (3) timing gestures in the context of speech in order to establish correspondence between accented behaviors in both speech and gesture channels.

An important aspect in the measuring of gestures is to identify cues for the gesture stroke, i.e. the most meaningful and effortful part of the gesture. As indicators we screen the signal (from electromagnetic trackers) for pre/post-stroke holds, strong acceleration of hands, stops, rapid changes in movement direction, strong hand tension, and symmetries in two-hand gestures. To give an idea, Figure 1 shows the signal generated from hand movement in consecutive pointing gestures. We have developed a variety of methods, among them HamNoSys [12] descriptions and timed ATNs (augmented transition networks), to record significant discrete features in a body reference system and filter motion data to object transformations which are put into effect in the 3D scene. The things we have learned from investigating these issues help us to advance natural interaction with 3D stereographic scenes in a scenario of virtual construction.

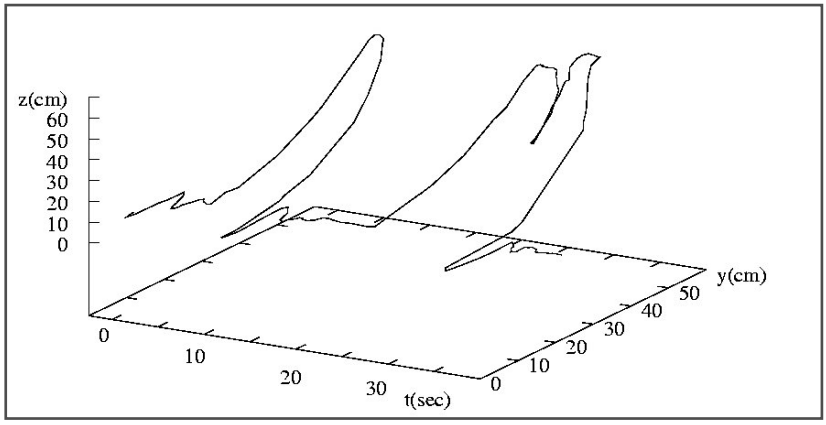


Fig. 1. Tracked hand movement in consecutive pointing gestures.

In previous years we have dealt with pointing and turning gestures accompanying speech, commonly classified as deictics and mimetics [6; 7]. In the DEIKON project ("Deixis in Construction Dialogues"), we have now started to research into more sophisticated forms of deictics that include features indicating shape or orientation, which lead into iconic gesture [14]. The DEIKON project was begun in the year of 2000 and is concerned with the systematic study of referential acts by coverbal gesture. In a scenario setting where two partners cooperate in constructing a model aeroplane, we investigate how complex signals originate from speech and gesture and how they are used in reference. One goal is to elucidate the contribution of gestural deixis for making salient or selecting objects and regions. Another goal is to make an artificial communicator able to understand and produce coverbal gestures in construction dialogues.

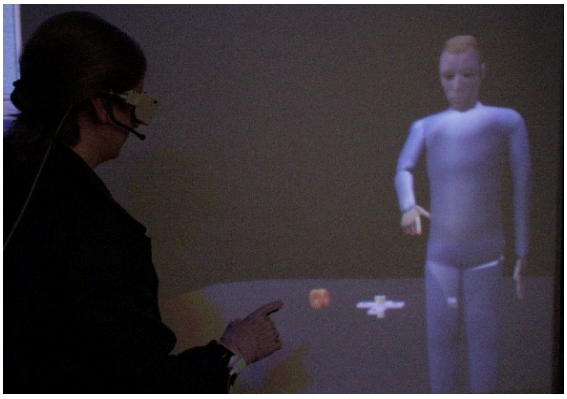


Fig. 2. Articulated communicator (target scenario).

In this context, this contribution focusses on an approach for synthesizing lifelike gestures for an articulated virtual agent, with particular emphasis on how to achieve temporal coordination with external information such as the signal generated by a text-to-speech system. The mid-range goal of this research is the conception of an "articulated communicator" (cf. Figure 2) that conducts multimodal dialogue with a human partner in cooperating on a construction task.

The paper is organized as follows. Having sketched some of our previous work and the context in which it is carried out, we next turn to the issue of lifelike gestures synthesis which is a core issue of our most recent work. In Section 3 we describe some details of the articulated communicator. The focus of Section 4 is timing, in particular, with respect to the gesture stroke. In Section 5 we give an outlook on how we have started to include speech with gesture synthesis.

2 Lifelike Gesture Synthesis

The rationales for our research on lifelike gesture synthesis are twofold. On the one hand, we seek for a better understanding of biologic and of cognitive factors of communication abilities through a generative approach ("learning to generate is learning to understand"). That is, models of explanation are to be provided in the form of "biomimetic" simulations which imitate nature to some extent. On the other hand, the synthesis of lifelike gesture is finding growing attention in human-computer interaction. In the realm of a new type of advanced application interfaces, the generation of synthetic gesture in connection with text-to-speech systems is one of the goals for embodied conversational agents [1].

If we want to equip a virtual agent with means to generate believable communicative behaviors automatically, then an important part in this is the production of natural multimodal utterings. A lot of progress has been made with respect to combining speech synthesis with facial animation to bring about lip-synchronous speech, as with so-called talking heads [10]. Another core issue is the skeletal animation of articulated synthetic figures for lifelike gesture synthesis that resembles significant features of the kinesic structure of gestural movements along with synthetic speech. Especially the achievement of precise timing for accented behaviors in the gesture stroke as a basis to synchronize them with, e.g., stressed syllables in speech remains a research challenge.

Although promising approaches exist with respect to the production of synthetic gestures, most current systems produce movements which are only parametrizeable to a certain extent or even rely on predefined motion sequences. For instance, the GeSSyCa system by Lebourque and Gibet [8] produces (French SL) sign language gestures from explicit representations, based on a limited set of motion primitives (pointing, straight line, curved, circle, wave form movements). They can be combined to more complex gestures and reproduce natural movement characteristics (cf. Figure 3, left). However, adaptation of the movement's temporal and kinematic properties as required in coverbal gesture is out of the focus of their work.

Fig. 3. Gesture production in the GeSSyCa (left), REA (middle), and MAX system (right).

The REA system by Cassell and coworkers (described in [1]) implements an embodied agent which is to produce natural verbal and nonverbal outputs regarding various relations between the used modalities (cf. Figure 3, middle). In the gesture animation process, a behavior is scheduled that, once started, causes several motor primitives to be executed. The REA gesture model employs standard animation techniques, e.g. keyframe animation and inverse kinematics. Although the issue of exact timing of spoken and gestural utterances is targetted in their work, the authors state that it has not yet been satisfactorily solved.

The goal of our own approach, demonstrable by the MAX system (cf. Figure 3, right) is to render real-time, lifelike gesture animations from representations of spatio-temporal gesture

knowledge. It incorporates means of motion representation, planning, and control to produce multiple kinds of gestures. Gestures are parametrized with respect to kinematics, i.e. velocity profile and overall duration of all phases, as well as to shape properties. In addition, emphasis is given to the issue of "peak timing", that is, to produce accented parts of the gesture stroke at precise points in time that can be synchronized with external events such as stressed syllables in synthetic speech. In more detail this is described in the following sections.

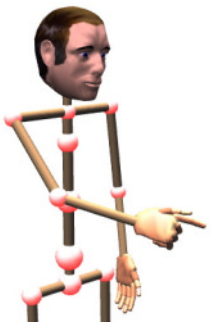
3 Articulated Communicator

In earlier work we have developed a hierarchical model for planning and generating lifelike gestures which is based on findings in various fields relevant to the production of human gesture [4;5]. Our approach grounds on knowledge-based computer animation and encapsulates low-level motion generation and control, enabling more abstract control structures on higher levels. These techniques are used to drive the kinematic skeleton of a highly articulated figure – the "articulated communicator" – which comprises 43 degrees of freedom (DOF) in 29 joints for the main body and 20 DOF for each hand (cf. Figure 4, left). While it turned out to be sufficient to have the hands animated by key-framing, the arms and the wrists are driven by model-based animation, with motion generators running concurrently and synchronized.

To achieve a high degree of lifelikeness in movement, approaches based on control algorithms in dynamic simulations or optimization criteria are often considered a first method, since they lead to physically realistic movements and provide a high level of control. But due to high computational cost they are usually not applicable at real-time. Starting from the observation that human arm movement is commonly conceived as being represented kinematically, we employ a single representation for both path and kinematics of the movement. The representation is based on B-splines which in particular make it possible to have smooth arm gestures that may comprise several subsequent, relative guiding strokes.

Our model (see Figure 5) incorporates methods for representing significant spatiotemporal gesture features and planning individual gestural animations, as well as biologically motivated techniques for the formation of arm trajectories. The fundamental idea is that, in planning a gestural movement, an image of the movement is created which is internally formed by arranging constraints representing the mandatory spatial and temporal features of the gesture. These are given either as spatiotemporal descriptions from previous stages of gesture production, e.g., location of a referent for deictic gestures, or they are retrieved from representations of common sense knowledge about gestural movements, e.g., the conventionalized hand shape during pointing. Therefore, our model comprises a gesture lexicon or *gestuary* as postulated by deRuiter [2], which contains abstract frame-based descriptions of gestural movements in the stroke phase, along with information about their usage for transferring communicative intent. The entries can hence be considered as defining a mapping from communicative function to explicit movement descriptions of the gesture stroke.

In the gestuary, gestures are described in terms of either postural features (static constraints) or significant movement phases (dynamic constraints) that occur in the gesture stroke. Features that can be defined independently are described using a symbolic gesture notation system which builds on HamNoSys [12], while others must be determined for each individual gesture. To this end, the description further accommodates entries which uniquely refer to specific values of the content the gesture is to convey, e.g., quantitative parameters for deictic or iconic gestures. The gesture's course in time is defined by arranging the constraint definitions in a tree using PARALLEL and SEQUENCE nodes which can optionally be nested. While means for defining parallel and sequential actions were provided with original HamNoSys, Version 2.0 [12], further HamNoSys-style conventions were introduced for our animation model which allow it to describe repetitions and symmetries in gestural movement.



```
(IDENT          Pointing-2)
(FUNCTION   Refer_To_Loc (RefLoc RefDir))
(CONSTRAINTS
      (PARALLEL
            (STATIC (HandShape  (BSifinger)))
            (STATIC (HandLocation  (RefLoc)))
            (STATIC (ExtFingerOrientation (RefDir)))
            (STATIC (PalmOrientation (PalmD)))
      )
)
```

Fig. 4. Articulated Communicator (left); template of a pointing gesture (right)

An example gesture template from the gestuary is shown in Figure 4; right (the concrete syntax is actually denoted in XML). The communicative function of the pointing gesture specified is to refer to a location plus indicating a pointing direction. By the use of these two parameters it is, for instance, possible to have the finger point to a location from above or from the side. The pointing gesture stroke is defined to have movement constraints that describe a target posture of the hand to be reached at the apex, namely, that – in parallel – the handshape is basic shape index finger stretched, the hand location is directed to the referenced location, the extended finger orientation is to the referenced direction, and the palm orientation is palm down. The symbols used in Figure 4 are ASCII equivalents of selected HamNoSys symbols that we use for gesture description.

In the following section we explain how gesture templates are instantiated to include timing constraints which lead to individual movement plans that are executed by the articulated communicator's motor system.

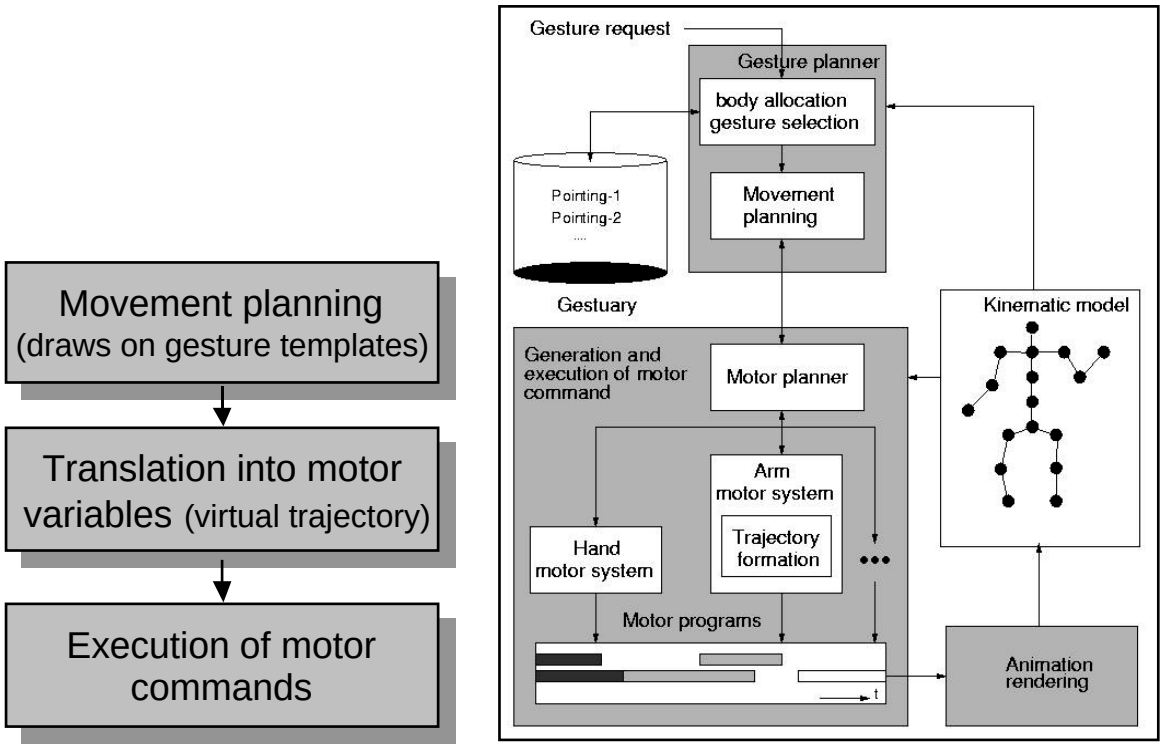


Fig. 5. Main stages and overall architecture of the gesture animation system.

4 Timing

As was said earlier, our gesture generation model is based on a variety of findings of gesture production and performance in humans which is a complex and multi-stage process. It is commonly assumed that representational gestural movements somehow derive from spatiotemporal representations of "shape" in the working memory on cognitively higher levels. These representations are then transformed into patterns of control signals which are executed by low-level motor systems. The resulting gesture exhibits characteristical shape and kinematic properties enabling humans to distinguish them from subsidiary movements and to recognize them as meaningful [2]. In particular, gestural movements can be considered as composed of distinct movement phases which form a hierarchical kinesic structure (cf. Figure 6). In coverbal gestures, the stroke (the most meaningful and effortful part of the gesture) is tightly coupled to accompanying speech, yielding semantic, pragmatic, and even temporal synchrony between the two modalities [9]. For instance, it was found that indexical gestures are likely to co-occur with the rheme, i.e. the focused part of a spoken sentence, and that the stroke onset precedes or co-occurs with the most contrastively stressed syllable in speech and covaries with it in time.

4.1 Prerequisites

In Figure 5, an outline of the main stages of the gesture animation process (left) and the overall architecture of the movement planning and execution (right) are shown. In the first gesture planning stage – movement planning – an image of the movement is created in the way indicated in the previous section, by arranging constraints representing the mandatory spatial and temporal features of the gesture. In the second planning stage, these ordered lists of constraints are separated and transferred to specialized hand, wrist, and arm motor control modules. These modules produce submovements for preparation, stroke, and retraction phases of the corresponding features that occur in a gesture phrase (cp. Figure 6).

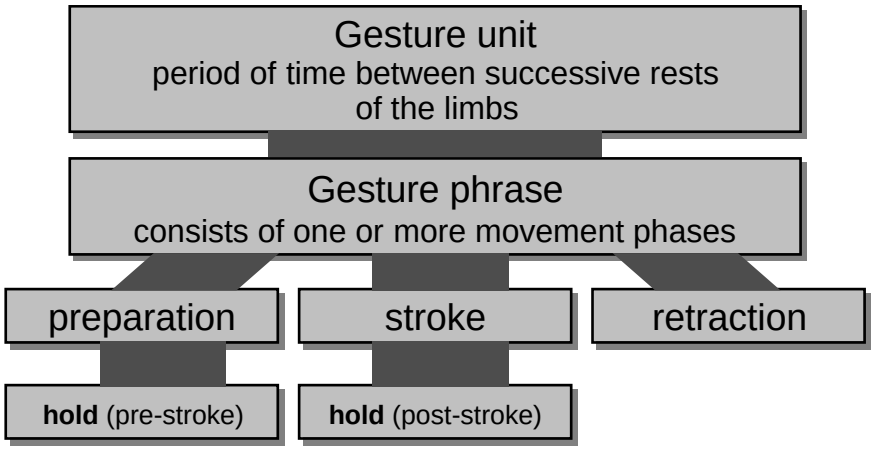


Fig. 6. Hierarchical kinesic structure of gestural movements (after [9]).

We briefly describe how motor commands are put into effect in the third (execution) stage of the animation system. The overall movement is controlled by a motor program which is able to execute an arbitrary number of local motor programs (LMPs) simultaneously (for illustration see bottom of Figure 5; right). Such LMPs employ a suited motion generation method for controlling a submovement (affecting a certain set of DOFs) over a designated period of time. LMPs are arranged in a taxonomy and share some functionality necessary for basic operations like creating, combining, and coordinating them. In detail, each LMP provides means for (1) self-activation and self-completion and (2) concatenation. In order to guarantee continuity in the

affected variables, each LMP connects itself fluently to given boundary conditions, i.e., start position and velocity.

Since different movement phases need to be created by different motion generators, LMPs are arranged in sequences during planning (concatenation). In the execution stage, an LMP is able to take over control from its predecessor or to pass it on to a successive one. The specializied motor control modules create and prepare proper LMPs from the movement constraints at disposal. In addition, concatenations are defined by assigning predecessor, resp. successor relationships between the LMPs. At run-time, the LMPs complete themselves and pass control over to one another. Note that any of the defining features may be left unspecified and does not affect the submovements within the complementary features. Moreover, the devised method accounts for co-articulation effects, e.g., fluent gesture transitions emerge from activation of the subsequent gesture (resp. its LMPs) before the preceding one has been fully retracted.

4.2 Timed movement planning – example

The most crucial part in movement planning is the issue of timing the individual phases of a gestural movement. On the one hand this is relevant to achieve a high degree of lifelikeness in the overall performance of a gesture and, in particular, the gesture stroke. On the other hand, timing is the key issue to enable cross-modal synchrony with respect to the coordination of gestures with the signal generated by a text-to-speech system. For instance, we would want the apex of a gesture stroke to be coordinated with peak prosodic emphasis in spoken output.

Our gesture planner forms a movement plan, i.e. a tree representation of a temporally ordered set of movements constraints, by (1) retrieving a feature-based gesture specification from the gestuary, (2) adapting it to the individual gesture context, and (3) qualifying temporal movement constraints in accordance with external timing constraints. The movement planning modules for both hand and arm motor systems are able to interpret a variety of HamNoSys symbols, selected with respect to defining initial, intermediate and target postures, convert them into position and orientation constraints with respect to an egocentric frame of reference, and generate a movement which lets the arm/hand follow an appropriate trajectory. It is possible to specify movement constraints with respect to static or dynamic features of the gesture stroke (see Section 3) and, further, to comfortably define timing constraints with respect to start, end, and peak times of each feature incorporated in the gesture stroke. Hence, roughly, a movement plan is generated by a specification of the following details:

HamNoSys	+ movement constraints	+ timing constraints
(selected)	{STATIC, DYNAMIC}	{Start, End, Manner }

To give an example of a movement plan, Figure 7 shows the instantiated gesture template of the stroke phase of a left-hand pull gesture which starts with an open flat hand, stretched out with the palm up, and ends with a fist-shaped hand near the left shoulder, palm still up (relative to the forearm). Dynamic movement constraints are specified for arm and hand motion, with initial and target postures given in symbolic HamNoSys descriptions. From the way the timing constraints are instantiated (Start, End, and Manner), the stroke would be performed within the period of 310 ms, with a velocity peak close to the end of that period. The peak can be placed at any time within the stroke phase. In this example, the apex of the arm motion and the apex of the hand motion are synchronous because they are carried out in parallel and their peaks are specified at equal times. Due to a STATIC movement constraint, the palm orientation remains the same over the full gesture stroke.

```
Pull-2
  (PARALLEL (Start 1.1)(End 1.41))
    (DYNAMIC (Start 1.1)(End 1.41)(HandLocation ((LocShoulder LocLeftBeside LocStretched)
                                                 (LocShoulder LocLeftBeside LocNear))) (Manner ((Peak 1.39))))
    (DYNAMIC (Start 1.1)(End 1.41)(HandShape ((BSflato)(BSfist)))(Manner ((Peak 1.39))))
    (STATIC (Start 1.1)(End 1.41)(PalmOrientation (PalmU)))
```

Fig. 7. Example of pull gesture desription (stroke phase) with instantiated start, end, and peak times.

Preparation and retraction of the pull gesture are supplied from the motor planner automatically, with smooth (C1-continuous) transitions and tentative, but not full-stop, pre- and post-stroke holds automatically generated. The resulting velocity profile for the pull gesture as specified in Figure 7 is shown in Figure 8, and snapshots from the preparation, stroke, and retraction phases are shown in Figure 9. Similarly, we have sucessfully specified a wide variety of further gestures, among them pointing gestures with and without peak beats, iconic two-handed gestures, and gestures that include several guiding strokes such as the outlining of a rectangular shape. Lab experience has shown that any specific gesture in such a realm can be specified in the amount of not much more than a minute and put into effect in real time.

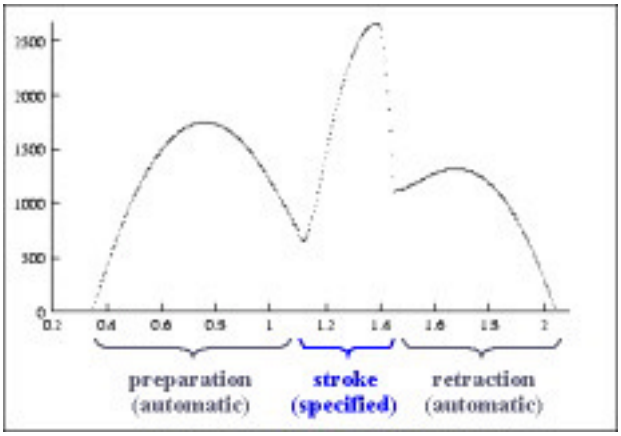


Fig. 8. Velocity profile for a pull gesture with a timed peak.

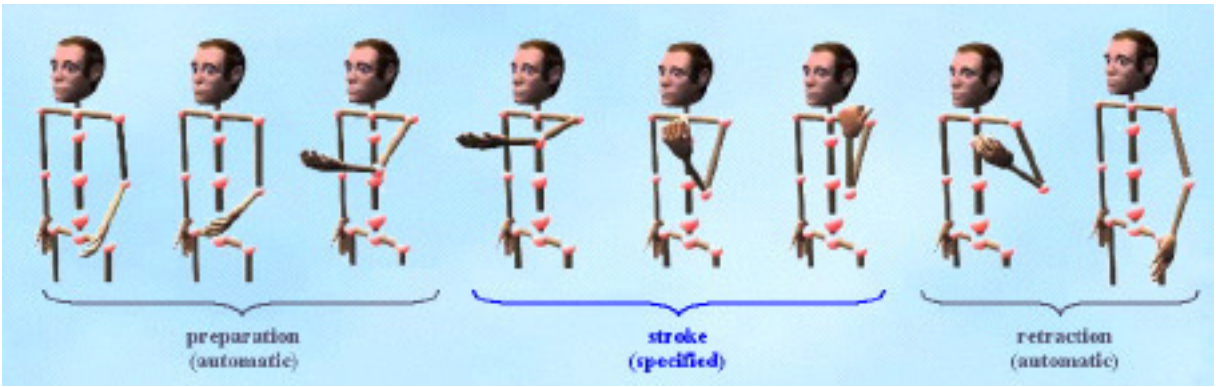


Fig. 9. Preparation, stroke and retraction phase for pull gesture.

5 Outlook: Gesture and Speech

One core issue in our work is the production of synthetic lifelike gesture from symbolic descriptions for an articulated virtual figure where natural motion and timing are central aspects. Particular emphasis lies on how to achieve temporal coordination with external information such as the signal generated by a text-to-speech system. As our model is particularly conceived to enable natural cross-modal integration by taking into account temporal synchrony constraints, further work includes the integration of speech-synthesis techniques as well as run-time extraction of temporal constraints for the coordination of gesture and speech. In this outlook, some remarks on ongoing work is given. In particular, we have managed to coordinate the gesture stroke of any formally described gesture with synthetic speech output. For instance, we can have the MAX agent say (in German) "now take *this bolt* and place it in *this hole*" and, at the times of peak prosodic emphasis, have MAX issue pointing gestures to the according locations. Thereby, the shape and specific appearance of the gesture is automatically derived from the gestuary and the motor system, while the gesture peak timing is derived from the EMPH (emphasis) parameters of the synthetized speech signal.

For text-to-speech (TTS) we currently use a combination of a module for orthographic-phonetic transcription and prosody generation, TXT2PHO, and MBROLA for speech synthesis. TXT2PHO was developed at the University of Bonn [11]. Its core part consists of a lexicon with roughly 50.000 entries and flexion tables which are used to convert German text to phonemes. Each word is marked with prominence values which support the subsequent generation of prosodic parameters (phoneme length; intonation) to produce a linguistic representation of text input. From this representation, speech output is generated by the use of MBROLA and the German diphone database provided for it [3]. MBROLA is a real-time concatenative speech synthesizer which is based on the multi-band resynthesis, pitch-synchronous overlap-add procedure (MBR-PSOLA). To achieve a variety of alterations in intonation and speech timing, pitch can be varied by a factor within the range of 0.5 to 2.0, and phone duration can be varied within the range of 0.25 to 2.0.

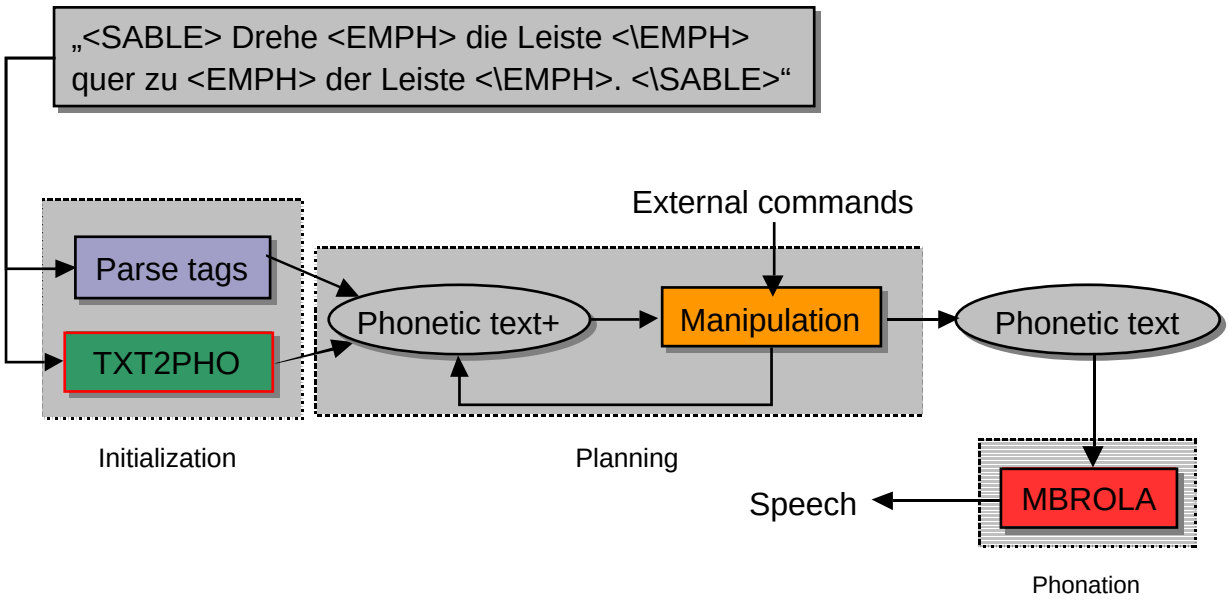


Fig. 10. "Turn *this bar* crosswise to *that bar*" – outline of our text-to-speech method allowing pitch scaling and time scaling; SABLE tags used for additional intonation commands.

Building on these features, a method was developed in our lab which allows to control a variety of prosodic functions in the TTS system by pitch scaling and time scaling. We use a markup language, SABLE [13] which is based on the extensible markup language (XML), to tag words or syllables to be emphasized in speech. Upon parsing such tags in the text input, phonetic text produced by TXT2PHO is altered accordingly and can further be manipulated to meet timing constraints from external commands generated in conjunction with the gesture planning procedure. Thus it is possible to preplan the timing of stressed syllables in the phonation for MBROLA and to synchronize accented behaviors in speech and gesture synthesis. Figure 10 roughly outlines the extended TTS method used in our lab. Not further explained in this paper, the face of the MAX agent is animated concurrently to exhibit lip-synchronous speech.

In conclusion, we have presented an approach for lifelike gesture synthesis and timing that should work well for a conversational agent and which is demonstrable by the MAX system. Our model is conceived to enable cross-modal synchrony with respect to the coordination of gestures with the signal generated by a text-to-speech system. In particular, the methods desribed can achieve precise timing for accented parts in the gesture stroke as a basis to synchronize them with stressed syllables in accompanying speech. Future work will be directed to have the system link discourse segments in a smooth way. In multimodal communication, by which we mean the concurrent formation of utterances that include gesture and speech, a rhythmic alternation of phases of tension and relaxation can be observed. The issue of rhythm in communication has been addressed widely and has been a key idea in our earlier work on synchronizing gesture and speech in HCI input devices [15]. We intend to use the idea of production pulses to mark a grid on which accented elements (e.g., stressed syllables) are likely to occur, together with a low-frequency (2-3s) chunking mechanism to achieve natural tempo in multimodal discourse output.

Acknowledgment

This research is partially supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center “Situated Artificial Communicators” (SFB 360). The authors are indebted to Dirk Stößel who provided a scalable TTS in his master thesis work, and to the members in the Bielefeld AI group who provided general support. in many ways

References

[1] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill (eds.). Embodied Conversational Agents. Cambridge (MA): The MIT Press, 2000.

[2] J.P. deRuiter. The Production of Gesture and Speech. In D. McNeill (ed.): Language and Gesture (pp. 284-311). Cambridge, UK: Cambridge University Press, 2000.

[3] Dutoit, T. (1997) An Introduction to Text-to-Speech Synthesis. Dordrecht: Kluwer Academic Publishers.

[4] S. Kopp & I. Wachsmuth. A knowledge-based approach for lifelike gesture animation. In W. Horn (ed.) ECAI 2000 Proc.14th European Conference on Artificial Intelligence, Amsterdam: IOS Press, 2000.

[5] S. Kopp and I. Wachsmuth. Planning and motion control in lifelike gesture: a refined approach. In Proceedings of Computer Animation 2000, IEEE Computer Society Press, 2000, pp. 92-97.

[6] M. Latoschik, I. Wachsmuth. Exploiting Distant Pointing Gestures for Object Selection in a Virtual Environment. In I. Wachsmuth & M. Fröhlich (eds.): *Gesture and Sign Language in Human-Computer Interaction (pp 185-196)*. Berlin: Springer (LNAI 1371), 1998.

[7] M. Latoschik, M. Fröhlich, B. Jung, I. Wachsmuth: Utilize Speech and Gestures to Realize Natural Interaction in a Virtual Environment. IECON'98 - Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society, Vol. 4, IEEE, 1998, 2028-2033.

[8] T. Lebourque & S. Gibet. A complete system for the specification and the generation of sign language gestures. In A. Braffort et al. (eds.), Proceedings International Gesture Workshop (pp. 227-238). Berlin: Springer-Verlag (LNAI 1739), 1999.

[9] D. McNeill. Hand and Mind: What Gestures Reveal about Thought. Chicago: University of Chicago Press, 1992.

[10] Pelachaud, C. & Prevost, S. (1995). Talking Heads: Physical, Linguistic and Cognitive Issues in Facial Animation. Course Notes, Computer Graphics International, June 1995.

[11] Portele, T., Höfer, F. & Hess, W. (1994) A Mixed Inventory Structure for German Concatenative Synthesis. In: Proceedings of the 2nd ESCA/IEEE Workshop on Speech Synthesis (pp. 115-118).

[12] S. Prillwitz, R. Leven, H. Zienert, T. Hamke, and J. Henning. HamNoSys Version 2.0: Hamburg Notation System for Sign Languages: An Introductory Guide, volume 5 of International Studies on Sign Language and Communication of the Deaf. Signum Press, Hamburg, Germany, 1989.

[13] Sable Consortium (2000) SABLE: A Synthesis Markup Language (version 1.0). Bell Laboratories, 25.2.2001, <http://www.bell-labs.com/project/tts/sable.html>

[14] T. Sowa & I. Wachsmuth: Coverbal Iconic Gestures for Object Descriptions in Virtual Environments: An Empirical Study. Submitted for the (Post-) Proceedings of the Conference "Gestures: Meaning and Use" (Porto, Portugal, April 2000).

[15] Wachsmuth, I. (1999). Communicative Rhythm in Gesture and Speech. In A. Braffort et al. (Eds.), *Gesture-based Communication in Human-Computer Interaction.* Berlin: Springer (LNAI 1739).