Virtual Agent for Deaf Signing Gestures

Sylvie Gibet, Alexis Héloir, Nicolas Courty, Jean-François Kamp, Philippe Gorce, Nasser Rezzoug, Franck Multon, Catherine Pelachaud

Virtual Agent for Deaf Signing Gestures

[1]S. Gibet, [1]A. Héloir, [1]N. Courty, [1]J.F. Kamp, [2]N. Rezzoug, [2]P. Gorce, [3]F. Multon, [4]C. Pelachaud

[1]VALORIA-Samsara, University of Bretagne Sud, rue André Lwoff, 56017 Vannes,
[2]LESP, University of Sud-Toulon-Var, avenue de l'université, 83957 La Garde,
[3]LBPEM, Unversity of Rennes 2, place du Recteur Henri Le Moal, 35043 Rennes,
[4]LINC, University of Paris 8, 2 rue de la Liberté, 93526 Saint-Denis

Abstract

We describe in this paper a system for automatically synthesizing deaf signing animations from motion data captured on real deaf subjects. Moreover, we create a virtual agent endowed with expressive gestures. Our attention is focused on the expressiveness of gesture (what type of gesture: fluidity, tension, anger) and on its semantic representations. Our approach relies on a data-driven animation scheme. From motion data captured thanks to an optical system and data gloves, we try to extract relevant features of communicative gestures, and to re-synthesize them afterwards with style variation. Within this framework, a motion database containing the whole body, hands motion and facial expressions has been built. The analysis of signals makes possible the enrichment of this database by including segmentation and annotation descriptors. Analysis and synthesis algorithms are applied to the generation of a set of French Sign Language gestures.

Key words

Communication for deaf people, sign language gestures, virtual signer agent, gesture database.

1. Introduction

Representing and modelling human motion stands for an important challenge in researches dedicated to virtual reality and communication systems. Major difficulties arise from the translation of fine dexterous human motion onto a virtual human representation. The quality of this representation relies upon the ability of the animation system to be faithful to biomechanics laws as well as higher level rules giving sense to gestures. In recent years, the huge development of new technologies made possible the capture of human motion and the emergence of new animation systems with highly realistic characteristics.

Among these animation systems, some of them try to facilitate access by deaf signing citizens to information and services expressed in sign language, using virtual humanoids or avatars endowed with communicative behaviours. We present in this paper a set of software tools resulting from a French CNRS project, and dedicated to representation, manipulation and synthesis of French Sign Language (FSL) gestures. Our approach consists into integrating within an animation framework some properties of human motions extracted from data captured on real sequences. Such a data-driven scheme fits an analysis/synthesis paradigm.

The major originality of this work consists into focusing on the expressiveness of generated gestures as well as their symbolic representations: which style of gesture and emotion? How to express smoothness, jerkiness, or tension? How these properties influence the sense of gestures? Our analysis/synthesis scheme makes possible to exploit the richness of existing gestures while enabling a parameterization of those gestures.

As a first trial, we apply this method on a subset of French Sign Language (FSL) gestures but a larger spectrum of applications concerning the design of communication virtual agents is at the crossway between various domains such as movement cognitive perception, robotics or rehabilitation. The contributions of the several teams involved in this project concerns: the definition of a semi structured gesture database including body, arm, hand movements and facial expressions, the development of tools to edit, annotate gestures and play motion onto a virtual humanoid, the analysis of sign language gesture sequences by extracting spatio-temporal and structural features, and finally the synthesis of whole body and facial movements.

This paper is organized as follows. A state of the art of the domain is proposed in section 2. After an overview of the project presented in section 3, more details are given in the following sections: construction of a structured multimodal database (section 4), motion post-processing (section 5), data-driven synthesis and high level description (section 6). Finally, conclusions are proposed in section 7.

2. Related works

An important number of studies are devoted to the linguistic analysis and representation, as well as to the synthesis of sign language gestures. Stokoe in his pioneering work (Stokoe, 1972), proposed a description of the American Sign Language (ASL) in terms of movement semantic units called *cheremes* and a system of transcription based on the combination of *cheremes* (Stokoe, 1978, Stokoe, 2005). The original notation consists of a limited number of symbols distributed in three classes, each representing a parameter for the sign formation: location (TAB), hand posture (DEZ) and movement (SIG). During the last decade, several computer applications developed a sequence of transformations including the following ordering steps: semantic representation, sign-language morphological representation, signing gesture specification language, and animation command. Losson (Losson, 1998) proposed an exhaustive grammatical description and a computer language based upon the linguistic description from Liddell and Johnson (Liddell, 1989). Lebourque designed a communication gesture specification language (GessyCA) based on a high level qualitative description of gestural command and a synthesis system based on a gradient descent strategy method (Lebourque, 1999, and Gibet, 2001). The commands are translated into parameters controlling an animation engine based on sensorimotor principles. More recently, the European ViSiCAST and eSign projects lead to the elaboration of a system for sign language communication (Kennaway, 2001, and Kennaway, 2003). Within these projects, an XML description language called SiGML was defined. SiGML is based on the HamNoSys notation (Prillwitz, 1989) and on the main concepts of the GessyCA language. A translator allows the transformation from the HamNoSys coding into SiGML expressions. In this system, two synthesis methods are proposed: one based on inverse kinematics and the second on motion capture data.

In this paper we propose a methodological approach for building a structured multimodal database, and the integration of a set of analysis and data-driven synthesis methods. To our knowledge our framework constitutes one of the first attempts to capture several modalities and try to handle them in a unified approach. Sign language phrases are registered, the same sequences being performed with different styles. The enriched database takes into account semantics as well as biomechanics features. The analysis of gesture sequences leads to a low dimension representation space used both for segmenting motion, and for smooth alignment between sequences performed with different styles. The results of the analysis process are used to generate sign language expressive sequences with various styles, thanks to the synthesis platforms developed in the research teams.

3. Project overview

In this section we present an overview of our framework. Fig.1 shows a representation of the different modules involved in this project. The motion analysis module aims at acquiring motions data from a professional signer with respect to our acquisition protocol (see next section for further information). Within this module, captured data are enriched (or augmented) through two processes: an edition/segmentation/annotation tool that requires user input, and an extraction of pertinent parameters (such as biomechanical invariants) that runs automatically by using statistical learning methods and model identification. The synthesis module controls the execution of the virtual humanoid movements. Finally, the high level description module allows the specification of the gesture to be executed from a high level representation. In the framework of this vast project, we focused our attention on the construction of a multimodal database that includes a limited set of FSL gestures executed with variations of styles and dynamics. This database constitutes the heart of the project. For modularity purpose, we designed this architecture to be as generic as possible, in order to switch easily between different animation models, and to exploit the most effective one with regards to a particular task or communicative gesture.

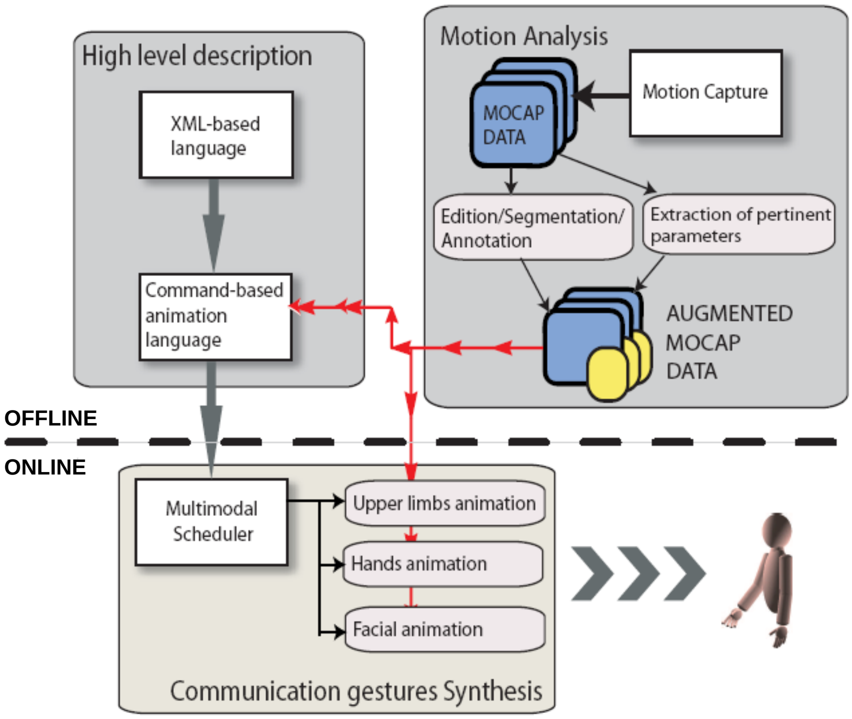


Fig.1. Functional architecture of the system

4. Acquisition protocol and multimodal database construction

One subject participated in this study. 24 reflective markers were placed on standardized anatomical landmarks, as shown in figure 2.a. A Vicon Mx system (product of Oxford Metrix) composed of 12 infrared cameras cadenced at 120Hz was used to capture the 3D displacements of those markers. This device also recorded facial expressions using small semi-spherical markers (3mm) at locations compliant with Mpeg4 specification (Fig. 2.b). Hand movements were tracked with two Cybergloves (Ascension technologies) composed of 22 sensors. Four recording sessions were organized to set up and synchronize the different devices. During the last session more than forty minutes of FSL gestures were recorded.

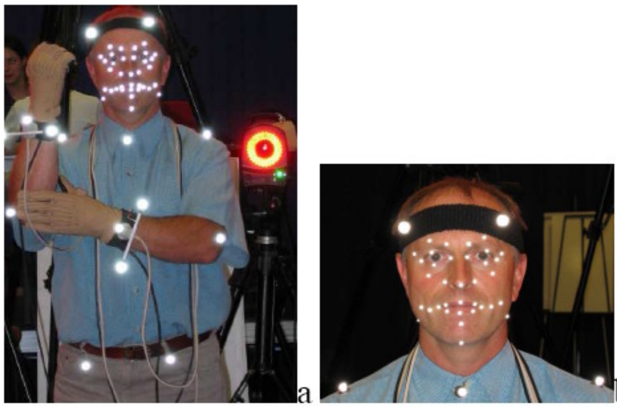


Fig.2. Alain Cahut signing: a) Whole body set up, b) Facial set up

A first corpus contains 16 sequences of FSL gestures representing several versions of a bulletin weather performed with different styles, relative to the subject's dynamics and emotional state. A second corpus gathers several incident reports relative to railway traffic. Finally, several isolated sequences were executed. One corresponds to the spelling of the FSL alphabet (B. Moody, 1998) and the other one describes the name of several French cities.

As shown in figure 3, anthropometric body and hand skeletons have been reconstructed, using specific methods from the MKM system and calibration methods for finger motions (N. Rezzoug, 2006). The fusion of the different signals (body, left and right hands) was realized after synchronization, resampling at 60 Hz, and interpolation with cubic splines. The elaboration of the motion database was achieved by gathering the different information sources into a standard format (Bio-Vision Hierarchical data), taking into account the hierarchical structure of the skeleton and the structure of the multidimensional data associated to each joint.

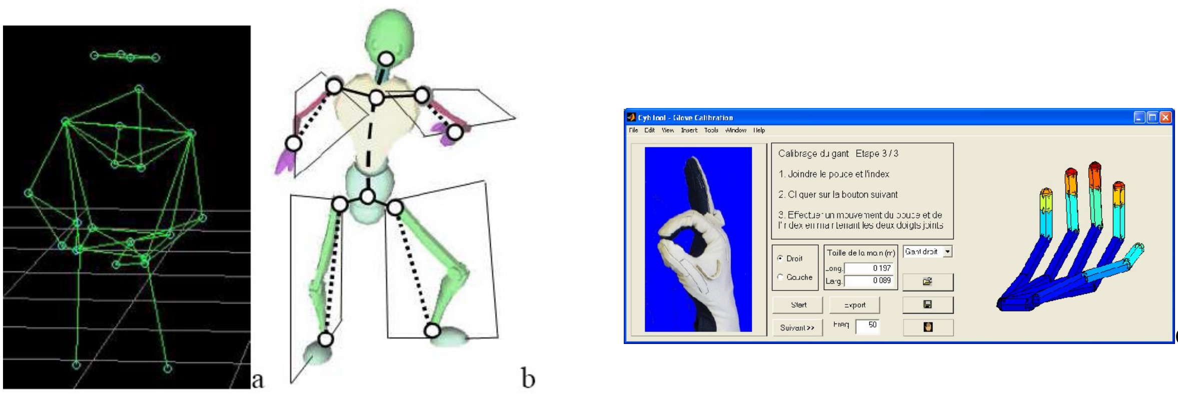


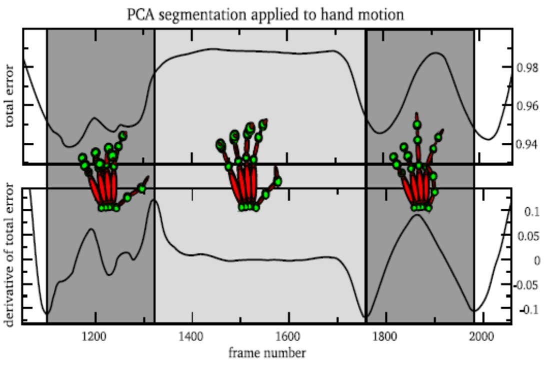
Fig.3. a) Markers location, b) Body normalized skeleton, c) Hand reconstruction

5. Motion post-processing

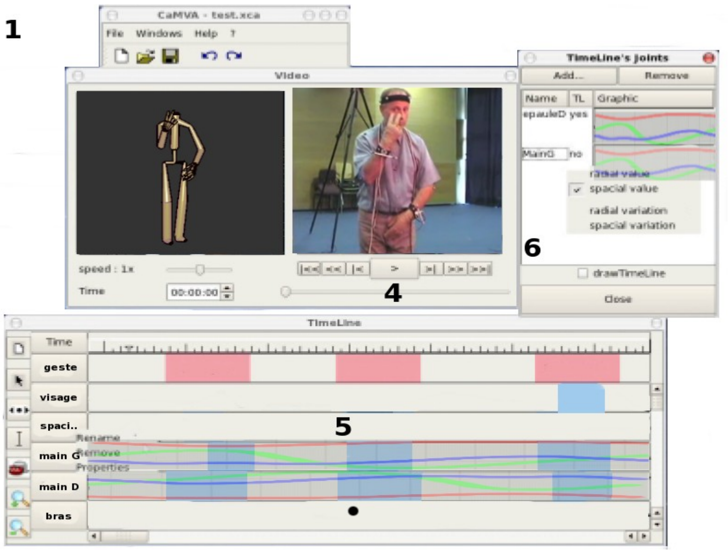
Gesture analysis and segmentation methods make possible the enrichment of movement databases by introducing structural elements that simplify movement specifications during the synthesis phase. In this section, we describe the segmentation tools we have developed and implemented. They can be classified in two categories: automatic and manual segmentation.

Automatic segmentation

Works from Barbic et al. (J. Barbik, 2004) revealed that dimension reduction tools such as Principal Component Analysis (PCA) may lead to segmentation close to the semantic segmentation made by a human subject. In (A. Héloir, 2005), this technique is applied to segment hand gestures. The assumption is that these movements can be represented by a succession of poses and configuration changes characterizing non linear transitions among motion. From an algorithmic point of view, this means that during transitions between motion segments, the reconstruction error induced by the projection of the posture on an adequate hyperplane of fixed dimension might grow quickly. These segmentation results are used to globally separate chunk of movements with the underlying assumption that the transition zones constitute articulation areas of the sign language discourse (Fig. 4 a).



a)



b)

Fig. 4. a) PCA-based hand gesture automatic segmentation;
b) Screenshot of the annotation and segmentation tool

Interactive tools for visualization, manual segmentation and annotation

A set of software tools was developed to visualize and annotate the captured gestures (Fig.4.b). The animation of the captured motion is replayed after reconstruction and synchronized with the video recordings. It is also possible to visualize, edit and annotate the temporal evolution of the skeleton joint parameters. This signal representation forwards a manual segmentation along different channels and the incorporation of temporal markers that enhance the signal description. The movement phases can be characterized by the addition of a set of descriptors and annotations tags (A. Héloir, 2005).

6. Data-driven gesture synthesis

Gesture synthesis with adaptation

In order to synthesize gestures in real time, a specific library called Manageable Kinematic Model (MKM) was developed. This library is dedicated to real time animation from captured data (S. Ménardais, 2004) and to motion retargeting, traducing the generation of recorded movement independent of the model morphology (R. Kulpa, 2005). Thanks to this representation, it is possible to replay captured motions over various synthetic actors and various environments. In order to insure that this operation does not deteriorate the initial movement, it is necessary to preserve its intrinsic related constraints. For instance, in the frame of sign language communication, if the hand has to contact the mouth for keeping the sense of the sentence, the system should be able to insure that this constraint is verified by the synthetic skeleton. This functionality is satisfied with MKM library, which makes possible the specification of different kinds of constraints: keeping a distance between two points on the skeleton, or between one point on the skeleton and one point in the environment, or verifying a segment orientation or angular velocity. These constraints are resolved in the real time environment by optimized inverse kinematics and kinetics algorithms (R. Kulpa, 2005).

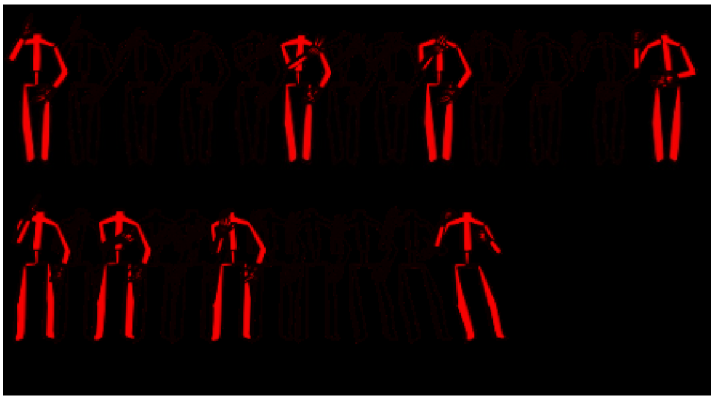


Fig. 5. Style-based FSL gesture synthesis.

Style-based motion synthesis

Recently, two articulated hands have been included into the animation framework (A. Héloir, 2005). Moreover, an original temporal alignment algorithm has been developed, which proves to be robust to the temporal and spatial variability that is induced by long gesture sequences performed with different styles (A. Héloir, 2006), as can be shown in figure 5.

High level description for facial expressions specification and generation

The specification and generation of facial expressions is based on the GRETA system which was initially developed for the animation of embodied conversational agents (C. Pelachaud, 2001, and I. Poggi, 2000). We intend to export in the near future this software into the MKM system for facial animation.

An expressiveness model of non verbal behavior was also defined by developing an intermediate representation of behavior parameterization, deduced from studies on movement perception (C. Pelachaud, 2005). This model enables, from a high level parametric specification, to produce behavior variations from one subject to another or for the same subject during a succession of emotional states. Thus, the personal style of a subject can be represented as well as variations according to the emotional state. From these high-level descriptions, we aim at specifying FSL gestures and generate motion using pre-recorded elementary units of motion. This part of the work is still in progress.

7. Conclusion

In this paper, an experimental platform dedicated to French Sign Language communication for storage, representation, and gesture animation by virtual agents was presented. Within this platform, three main objectives were achieved. Firstly, we elaborated a multimodal motion database including whole body, hand and facial synchronized gestures in a standardized format. Secondly, we developed motion post-processing tools, for the manual and automatic segmentation of motion signals, as well as visualization and edition tools for gesture annotation. These tools make possible the enrichment of the database with semantic descriptors and timing tags delimitating phases inside the gestures. Finally we developed a set of new algorithms for data-driven gesture synthesis.

In the near future, we plan to elaborate a more complete corpus of FSL gestures, and to constitute an indexed multimodal database which can be used to perform statistical studies on FSL gestures. This structured database which contains both semantic information and temporal signals should improve the design of virtual expressive and realistic agents for deaf signing gestures.

Acknowledgement

We wish to thank the subjects who participated in the motion capture sessions, in particular Alain Cahut, and their interpreters without whom nothing could have been possible. Their expertise in FSL allowed us to construct and interpret the several gesture corpuses.

References

1. W. C. Stokoe. Semiotics and Human Sign Language. Mouton, The Hague, 1972.
2. W.C. Stokoe, D. Casterline, and C. Croneberg. A dictionary of American Sign Language on Linguistic principles. Linstok Press, SilverSpring, 1978.
3. Stokoe, W.C. Sign language structure: an outline of the visual communication systems of the American deaf. 1960. J. Deaf Stud. Deaf Educ, 2005, 10, 3-37.
4. O. Losson and J.M. Vannobel. Sign language formal description and synthesis. Int. Journal of Virtual Reality, 3(4) :27–34, 1998.
5. S. K. Liddell and R.E. Jonhson. American Sign Language: the phonological base. In Sign Language Studies 64, pages 195–277, 1989.
6. T. Lebourque and S. Gibet. High level specification and control of communication gestures: the GESSYCA system. In Proc. of IEEE Computer Animation, Geneva, May 1999.
7. S. Gibet, T. Lebourque, and P.F. Marteau. High level specification and animation of communicative gestures. In Journal of Visual Languages and Computing, volume 12, pages 657–687, 2001.
8. R. Kennaway. Synthetic Animation of Deaf Signing Gestures. In 4th Int. Work. In Gesture and Sign-Language based Human-Computer Interaction, LNAI, Springer, 2001.
9. R. Kennaway. Experience with, and requirements for, a gesture description language for synthetic animation. In 5th Int. Work. In Gesture and Sign-Language based Human-Computer Interaction, LNAI, Springer, 2003.
10. S. L. Prillwitz, R. Leven, H. Zienert, R. Zienert, T.Hanke, and J. Henning. HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide. International Studies on Sign Language and Communication of the Deaf, 1989.
11. N. Rezzoug, P. Gorce, A. Heloir, S. Gibet, N. Courty, J.F. Kamp, F. Multon, and C. Pelachaud. Virtual humanoids endowed with expressive communication gestures: the HuGEx project, IEEE International Conference on Systems, Man, and Cybernetics, October, 2006.
12. B. Moody. In International Visual Theatre, editor, La langue des signes : Introduction à l'histoire et à la grammaire de la langue des signes, 1998.

13. J. Barbic, A. Safonova, J.Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Pollard. Segmenting Motion Capture Data into Distinct Behaviors. In Proceedings of Graphics Interface 2004, May, 2004, pp. 185 – 194, 2004.
14. A. Héloir, S. Gibet, F. Multon, and N. Courty. Captured motion data processing for real time synthesis of sign language. In Gesture in Human-Computer Interaction and Simulation, GW 2005 (revised short paper), Lecture Notes in Artificial Intelligence, Vol. 3881, Springer, 2005.
15. A. Héloir, S. Gibet, N. Courty, M. Raynaud. Système d'annotation et de segmentation de gestes de communication capturés. In Proc. of Traitement Automatique des Langues des Signes (Atelier TALS 2005), pp. 367-370, Dourdan, France, June 2005.
16. S. Ménardais, R. Kulpa, F. Multon, and B. Arnaldi. Synchronization for dynamic blending of motions. In Proceedings of ACM SIGGRAPH/ Eurographics Symposium on Computer Animation, pages 325–336, Grenoble, France, August 2004.
17. S. Ménardais, F. Multon, R. Kulpa, and B. Arnaldi. Motion blending for real-time animation while accounting for the environment. In Computer Graphics International, June 2004.
18. R. Kulpa and F. Multon. Fast inverse kinematics and kinetics solver for human-like figures. In Proceedings of IEEE Humanoids, pages 38–43, Tsukuba, Japan, December 2005.
19. R. Kulpa, F.Multon, and B. Arnaldi. Morphology-independent representation of motions for interactive human-like animation. Computer Graphics Forum, Eurographics 2005 special issue, 24(3) :343–352, 2005.
20. R. Kulpa and F. Multon. Fast inverse kinematics and kinetics solver for human-like figures. In Proceedings of IEEE Humanoids, pages 38–43, Tsukuba, Japan, Dec. 2005.
21. A. Héloir, N. Courty, S. Gibet, F. Multon. Temporal alignment of communicative gesture sequences. Computer Animation and Virtual Worlds (selected best papers from CASA'06), 17, July 2006.
22. C. Pelachaud. Contextually Embodied Agents, chapter Deformable Avatars, pages 250–263. Kluwer Publishers, 2001.
23. I. Poggi, and C. Pelachaud. Performative facial expressions in animated faces, chapter Embodied Conversational Agents, pages 155–188. MIT press, 2000.
24. B. Hartmann, M. Mancini, C. Pelachaud, Implementing Expressive Gesture Synthesis for Embodied Conversational Agents, Gesture Workshop, LNAI, Springer, May 2005.