A Qualitative and Quantitative Characterisation of Style in Sign Language Gestures

subtitle

Alexis Heloir[1] and Sylvie Gibet[1]

Laboratoire Valoria, Université de Bretagne Sud.
Vannes, France.
firstname.surname@univ-ubs.fr

Abstract. This paper addresses the identification and representation of the variations induced by style for the synthesis of realistic and convincing expressive gesture sequences. A qualitative and quantitative comparison between styled gesture sequences is performed. This comparison leads to the identification of temporal, spatial, and stuctural processes that are described in a theoritical model of sign language. Insights raised by this study are then considered in order to enhance existing gesture specification systems.

1 Introduction

Embodying a virtual humanoid with expressive gestures requires to take into account the properties that influence the perception of convincing movements. One of these properties is style. Style, on the one hand, enhance, augment and colours the inherent meaning of a message. On the other hand, style carries information about the speaker age, gender, cultural background, and emotional state. As a consequence, style contributes to make a virtual humanoid more convincing and may increase its acceptance towards human users.

This paper presents both a qualitative and quantitative analysis of styled motion gesture data. The temporal, spatial and structural differences between styled gestures are confronted with the phenomenons described in the *Prosodic Model of Sign Language Phonology*[1].

It is already well known that gesture style modifies the temporal and the spatial aspects of gestures. In this paper, we address how style may as wel influence the structure organisation of some lexical units.
We then briefly present new insights for taking into account both the spatio-temporal and structural variations induced by style in existing gesture specification frameworks.

This paper is organised as follows : related work is reviewed in section 2, section 3 presents the gesture data upon which the study was performed and the manual segmentation scheme we applied. Section 4 presents the temporal variations that have been identified while section 5 shows up the spatial and structural variations that have been observed. Section 6 raises insights towards the enhancement of existing expressive gesture specification models. Section 7 concludes and presents pointers to future work.

2 Related Works

Studies on sign language which have been carried out since the 60's lead to dedicated description/transcription systems [2, 3]. Several gestural generation systems inspired by those works have appeared since. The eSign project [4] designed and set up a communicative gesture synthesis system inspired by phonological description of sign language. Gibet & al. [5] propose an expressive gesture synthesis system where the task is expressed as a discrete sequence of targets in the euclidian space around the virtual signer. More recently appeared projects dedicated to geometrically based modelling of sign language [6–8].

Style and emotion centric studies around human motion appeared by the end of the XIX^e century and have been continued and enhanced until today [9, 10]. The underlying theory derived from those works served as a base for procedural motion synthesis systems [11]. Other procedural systems step on psycho cognitive studies [12] in order to convey emotional content [13]. Finally, inter-subject variability of style has been addressed in [14].

Recent studies dedicated to expressive gesture rely on segmentation and annotation of gesture. Those studies aim towards characterising the spatial structure of a Sign Language phrase [8], investigating the systematic synchony between modalities [15] or transcribing and modelling gestures with the goal of further resynthesis [16, 17].

In this paper, we rely on a slightly augmented version of the transcription scheme proposed by Kita et al. [18]. This transcription scheme provide a modelling of the temporal structure of gesture by identifying basic movement phases. The goal of the segmentation is to confront styled gesture sequences with the timing description, enhancement and surfacing processes described by Brentari [1]. The subsequent section present in more details the segmentation process.

3 Motion Acquisition and Segmentation

3.1 motion data

The motion data on which we drove the study has been obtained thanks to a motion capture process described in [19]. The capture sequences have been performed by a professional deaf interpret. The sequences are mainly composed of a succession of lexical units.

Those sequences depict the same weather forecast presentation. The difference among the sequences stays in the emotional content the signer was asked to mimic during the gesture performance. Three gesture sequences have been used in this study: on the first one, the signer was asked to be neutral, for the second one, the signer was asked to mimic anger, and weariness on the third one.

3.2 Segmentation Scheme

In order to highlight the succession of gesture phases occurring during a gesture sequence, we relied on the segmentation method proposed by Kita [18]. This method was originally applied to manual annotation of gesture recorded as video streams. However, in our case it makes sence to take profit of the extra information that is provided by the three dimensional reconstruction of the motion. Kita's model has already been adapted to automatic segmentation of three dimensional motion data in [20].Unfortunately, this method did not provide us with relevant segmentation over sign language gesture sequences, we thus prefered to rely on motion data analysis only to provide hints that could help a human annotator to take decision during a manual segmentation process.

According to [18], movement phases are characterised by either an abrupt change in the direction in the hand *movement* and discontinuity in the velocity profile of the hand *movement* before and after the abrupt direction change. As it may be challenging for a human annotator to precisely detect hand configuration change, the original segmentation process was enhanced by introducing information obtained from the three dimentional gesture representation of the motion into two supplementary annotation channels. Those supplementary channels are dedicated to represent how hand configuration change along a gesture sequence. It is computed for the both hands, along the entire sequence. for each frame i in the sequence, A distance D_i is computed over two consecutive hand configuration (A_i, A_{i+1}) aligned in the Cartesian space [21]. The sum of squared distances between the corresponding joints positions gives the distance between two consecutive frames.

$$D(A_i, A_{i+1}) = \sum_j \|\overrightarrow{(P_j, P'_j)}\|$$

We assume that a hand configuration change occurs when $D(A_i, A_{i+1})$ reaches two standard derivation from the mean value over the entire sequence.
The remaining of the segmentation proccess was manually performed over the video recording of the gesture sequences and segmented thanks to the Anvil software [16]. The segmentation was performed along the three styled gesture sequences and leaded to the identification of the movement phases described in [18] and resumed in the following.

- **stroke:** a phase characterised by a lexical-internal movement;
- **preparation:** a phase with a moving limb between two stroke;
- **retraction:** a phase that arrives at the resting position;
- **partial retraction:** an interrupted retraction phase followed by a preparation phase;
- **hold:** a phase in which the hand is held "still".

A both qualitative and quantitative comparison of the gesture sequences is presented in the next section.

4 Temporal variations and gesture phases between styles

4.1 Coordination between channels

A study was driven on the coordination between the channels involed in gesture movements. In order to identify coordination characteristics, the channels dedicated to the identification of hand motion were confronted to the channel dedicated to gesture phase identification. This study is insipred by the work by the stements relative to the temporal organisation of sign units presented in Brentari[22].

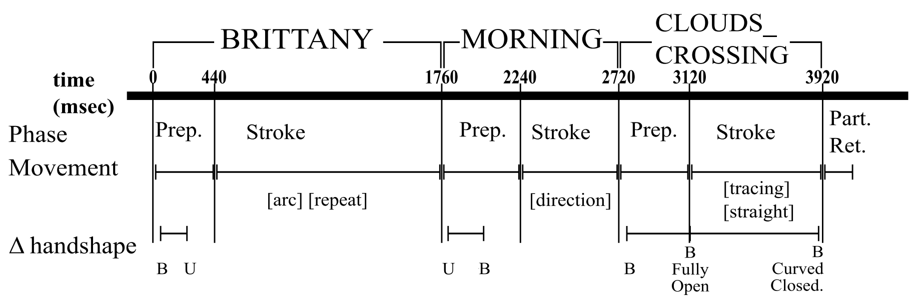

Fig. 1. Segmentation of gesture phases and temporal relations between hand and arm movement.

By making comparisons between the channels dedicated to handshape change and gesture phase, we are able to compare the *handshape change duration / movement duration* ($HS\Delta/Mov$ Ratio). This measure represents the amount of time a subject takes to execute a given handshape change simultaneously with a given movement [1]. This measure reveals relations about the coordination between arm and hand in the production of expressive gesture. We compared the $HS\Delta/Mov$ Ratio between two types of gesture phases where a handshape change was observed: preparation phases and stroke phases (see fig. 1 for details). The comparison was driven for each of the three captured gesture sequences. Results are summarised in table 1.

Table 1. $HS\Delta/Mov$ Ratio regarding gesture phase and style of gesture sequence

	preparation (mean, std. deviation)	stroke (mean, std. deviation)
neutral	38%, 19	93%, 5
angry	43%, 16	95%, 8
weary	37%, 12	86%, 8

Table 1 highlights on the one hand the similarities between handshape and motion coordination regarding style and on the other hand, the differences in the coupling of handshapes and movement between gesture phases. It appears that word-internal movements present high coupling while non-meaningful gesture phases present low coupling. Existing studies dedicated to sign language phonology state that the coupling and decoupling of handshape changes and movements constitutes evidence on the representation of word-internal movements around global timing units [1, 23]. Considering gesture specification, Those facts suggest that gesture phases should take into account different coordination schemes regarding their type: whereas word internal gesture phases should be specifed according to timing units and should reveal strong coordination between articulators involved in motion, non meaningful gestures like preparation or retraction should reveal a loose temporal coupling between articulators.

4.2 Timing Variations Between Single Path Signs

According to the litterature dedicated to sign language linguistics, timing variation may occur at several linguistic levels, such as phrase, word and syllabe [24]. in this subsection, we focus on the temporal variation occuring between simple gesture movements (syllabic level). In order to highlight the temporal variations induced by style between simple gesture units, a non linear distance evaluation was performed on each lexical unit containing a single movement path (identified as a non repetitive,non alternating stroke phase). According to Brentari [1], such movements contain two timing units[1]. The timing distance evaluation was performed thanks to a dynamic time warping algorithm, described in details in [25]. Dynamic Time Warping finds a non-linear timing adjustment between two time series that minimizes the global distance between two warped series. The warp path presented in fig.2 were computed by taking into account the preparation phase to the strokes, as it is difficult to identify precisely the separation between preparation phase and stroke phase in a lexical unit. The warp path calculated between the corresponding signs are presented in figure 1.

Figure 1 highlights the non linear relationship that links styled gestures. The light curves reprepresent warp path that were obtained by the DTW algorithm beetween a styled lexical unit and its neutral equivalent. The lexical unit contains a preparation phase plus a single movement path, segmented as a stroke phase. The bold curve represents a cubic regression of the average path over the actual computed paths. The DTW calculation was performed over 9 motion sequences. All composed by a preparation phase and a stroke phase. The comparison was done between the neutral and angry style fig 1.a, then between the weary and neutral style fig 1.b.

Those paths highlight three distinct phases occuring along the lexical unit. The first one represents the preparation phase and is subject to a linear timing deformation. The second phase represents the first timing unit of the simple

[1] Timing units are sometimes designed as segments in the litterature dedicated to sign language phonology [1, 24]

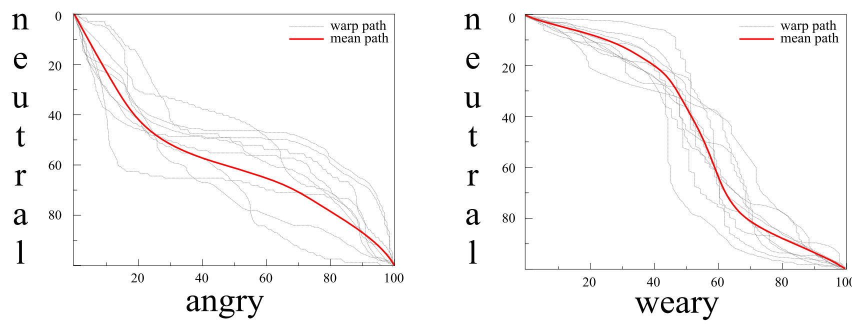

Fig. 2. in-sign temporal alignement between styles

movement in the lexical unit and the third phase stands for the second timing unit of the lexical unit. It appears on the one hand that the maximum timing deformation is observed during the first timing unit of the simple movement. On the other hand, timing deformation induced by style may be generalized, at a word level by a temporal variation profile that may be characterized according to a simple analytical formulation, like cubic Hermit curves.

4.3 Phase Repartition

Table 2 presents the number and the repartition of the annotated phases along the gesture sequences according to gesturing style. This table reveals that there is a very small difference between the number of preparation phases and the number of stroke phases. According to [18], retraction phases occur at the end of movement phrases or movement units. Thus, retraction phases highlights the end of *idea unit*, as proposed in [26]. It finally appears that hold phases occur more frequently in the *weary* sequence and less frequently in the *angry* sequence.

Table 2. Frequency of phases types

gesture sequence	preparation	stroke	retraction	partial retraction	hold	total
neutral	27 (40%)	27 (40%)	2 (3%)	6 (9%)	5 (7%)	67
angry	26 (39%)	27 (40%)	8 (12%)	2 (3%)	4 (6%)	67
weary	26 (37%)	27 (38%)	9 (13%)	2 (3%)	7 (10%)	71

The results depict in table 2 suggest that the election of gesture phases involded in a gesture phrase vary according to style. The most noticeable difference concerned retraction phases. For example,the increase of the frequency of complete retraction phases, in the angry-styled gesture tends to convey a more perceptive rhytm [27] in the gesture sequence.

5 Spatial and Structural Features

5.1 Proximalisation/distalisation

As stated in [1], a sign may be executed by a set of joints that differs form the specified set in the initial description of the sign. This phenomenon is described in the context of enhancement and visual sonority. Thus, a sign may surface in a way that the initial joints mainly involved in the execution of the movment migrate to a more proximal (closer to the body) or distal (closer to the fingers) joint. One example of proximalisation/distalisation is expressed in fig. 2.a , where an initial movement, [WIND] (fig. 2.a center) may surface as proximalised (fig. 2.a left) (the movement is mainly performed by the wrist) or may surface as distalised (fig. 2.a right) in this other example, the movement is performed by both the shoulders and the elbows. Another example is given in fig. 2.b where the initial movement ([YOU_MUST]), originally performed mainly by the elbow (fig. 2.b center) may surface as distalised (fig. 2.b left) where the movement is mainly perfomed by the wrist joints. Finally, the movement may surface as proximalised (fig. 2.b right), with the shoulder joints participating in the movement.
Those phenomenons, may convey, as highlighted in the examples, insights about the internal and emotional state of the signer, but, as stated in [13, 14, 28], they may as well give informations about the personnality, cultural and social background of the performer.

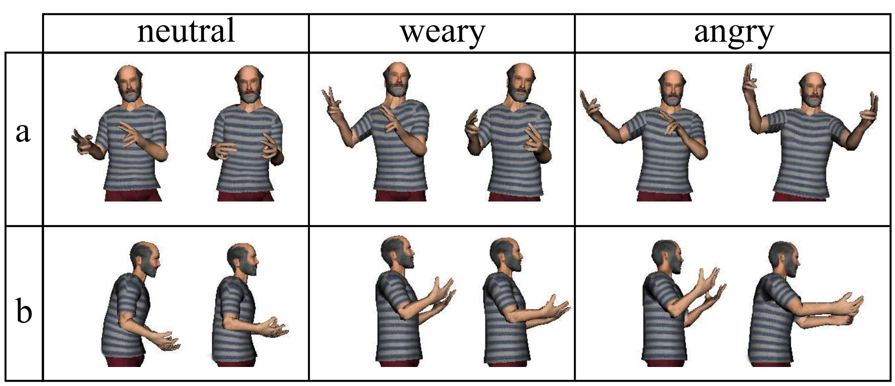

Fig. 3. Proximalisation and distalisation examples performed by a synthetic agent

5.2 Weak Drop

The weak drop phenomenon is an optional operation in which a two-handed sign is actually realised (surface) as a one handed sign. The weak drop phenomenon is observed once in the sequence that is performed in a relaxed style. This phenomenon occurs for the lexical entry [RAIN]. [RAIN] is a sign that involves two hands in a non alternating fashion and without constraint. This sign fulfils the requirements to surface as a one handed sign, according to [1]. The question is to know why this sign is elected to surface ad one handed sign (to undergo weak drop) while other valid candidates don't {[GO], [MUST], [CLOUDS], [NIGHT], [SEA], [NIGHT], [PLEASE], [BRITTANY]}.
Although the context favoritising the weak drop operation to occur is not clearly identified, we emit the assumption that the inner state and emotional state of an expressive agent may influence the occence of the weak drop operation.

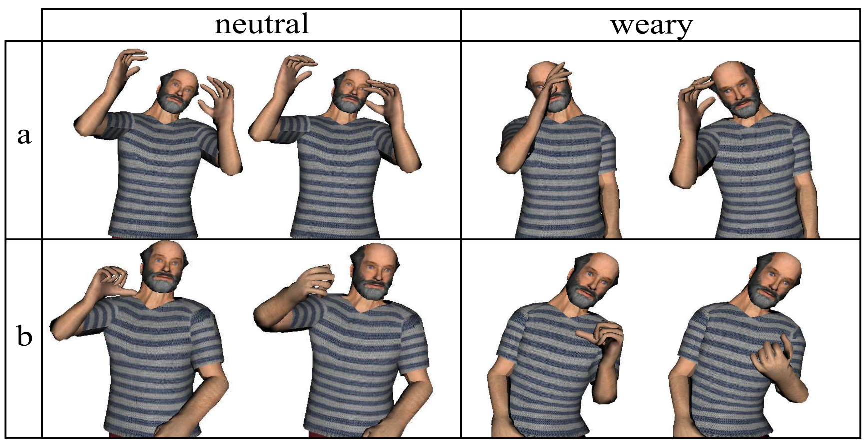

Fig. 4. example of Weak Drop and Hand Inversion performed by a synthetic agent

5.3 Strong Hand/Weak Hand Path Migration

Along the gesture sequences, we were able to identify one occurence of the hand inversion phenomenon. This phenomenon consists into performing a one handed sign thanks to the wak hand (H2) rather than the strong hand (H1). The operation is depicted in fig. 3.b, for the sign word [TOMORROW], successively performed according to neutral and weary style. This sign is a one handed sign that is (in weary style), surrounded by two other one handed signs : [RAIN], surfaced as a one handed sign (see above pararaph) and [sun], which is a one handed sign. Thus, the migration of the path movement in [rain] from H1 to H2 leads to an alternation of Right-Left-Right arm movements. This phenomenon illustrates an important change in the structure of gesture. and suggest the desciption of gesture according to hierarchical graphs.

6 Gesture Specification and Generation Framework

Our goal is to propose a framework dedicated to the specification and the generation of convincing expressive gestures. In order to take into account the variability induced by style, we choose to rely on the qualitative and quantitative study presented in sections 3 and 4. We thus plan to extend Lebourques and Gibet gesture specification system [29] by introducing new features in order to :

- take into account parts of the motion like retraction or preparation phases.
- take into account temporal features observed in gesture phases.
- take into account spatial variation between styles gesture phases.
- take into account changes in the structure of gestures.

The analysis of captured motion data should lead to the creation of gesture libraries taking gesturing style into account. Generation of expressive gesture phrases should therefore be realised thanks to a concatenation of elected gesture phases from the gesture libraries according to a specified style. The style should be dependant on the Agent's characteristics and emotional state, wich is in accordance with the style dictionary paradigm proposed in [14]. We briefly present the main insights introduced by the gesture analysis presented in this paper.

- By introducing gesture phases and specifying different classes of gestures { *preparation,retraction, partial retraction, stroke and hold* }, we are able on the one hand to represent gesture phases in a more accurate and realistic fashion.
- By specifying explicit timing relations between timing units and gesture articulators, we are able to specify faithful description of gestures. Plus, the correspondence between timing profiles highlighted in 4.1 provide time warp patterns leading to realistic timing deformations. Timing deformation should provide more freedom in the specification of phases length, or int smooth transition between styles.
- We would like to take into account phenomenons presented in section 5, such as proximalisation/distalisation, weak drop or hand inversion in the gesture generation system. Although not dedicated to sign gestures, we believe that taking into account such phenomenons should lead to more convincing utterances.

7 Conclusion and Future Works

We have presented both a qualitative and quantitative study of the influence of style on expressive gesture sequences. By considering spatial, temporal and structural aspects of gesture we were able to propose insights towards the enhancement of existing gesture specification frameworks. We plan to take into account some of the features described in the *prosodic model of sign language phonology* in order to improve existing expressive gesture specification models. Furhter works will be dedicated to the specification of a computational model of expressive gestures taking into account gesturing style and emotional state of the agent.

References

1. Brentari, D.: A prosodic model of sign language phonology. Bradford Book (1998)
2. Stokoe, W.: Sign language structure: An outline of the visual communication systems of the american deaf. In: Studies in Linguistic, Occasional Papers. Volume 8. (1960)
3. Prillwitz, S.L., Leven, R., Zienert, H., Zienert, R., T.Hanke, Henning, J.: HamNoSys. Version 2.0. International Studies on Sign Language and Communication of the Deaf (1989)
4. Zwiterslood, I., Verlinden, M., Ros, J., van der Schoot, S.: Synthetic signing for the deaf: Esign. In: Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment, Granada, Spain (July 2004)
5. Gibet, S., T.Lebourque, Marteau, P.: High level specification and animation of communicative gestures. Journal of Visual Languages and Computing **12**(6) (December 2001) 657–687
6. Filhol, M., Braffort, A.: A sequential approach to lexical sign description. In: LREC 2006 - Workshop on Sign Languages, Genova, Italy (2006)
7. Braffort, A., Lejeune, F.: Spatialised semantic relations in french sign language: Toward a computational modelling. Gesture in Human-Computer Interaction and Simulation **5** (2006) 37–48
8. Lenseigne, B., Dalle, P.: Using signing space as a representation for sign language processing. Gesture in Human-Computer Interaction and Simulation **6** (2006) 25–36
9. Laban, R.: The Mastery of Movement. Northcote House (1988)
10. Maletic, V.: Body, Space, Expression: The Development of Rudolf Laban's Movement and Dance Concepts. Mouton de Gruyte, New York, NY, USA (1987)
11. Chi, D.M., Costa, M., Zhao, L., Badler, N.I.: Emote. In Akeley, K., ed.: Siggraph 2000, Computer Graphics Proceedings, ACM Press / ACM SIGGRAPH / Addison Wesley Longman (2000) 173–182
12. Wallbot, H.: Bodily expression of emotion. European journal of social psychology **28** (1998) 879–796
13. Pelachaud, C., Hartmann, B., Mancini, M.: Implementing expressive gesture synthesis for embodied conversational agents. In: gesture in human-Computer Interaction and Simulation. (2006)
14. Ruttkay, Z., Noot, H.: Variations in gesturing and speech by gestyle. International Journal of Human-Computer Studies **62** (2005) 211–229 Special Issue on Subtle Expressivity for Characters and Robots.
15. Attina V., C.M.A..B.D.: Temporal measures of hand and speech coordination during french cued speech production. Gesture in Human-Computer Interaction and Simulation **3881** (2006) 13–24

16. Kipp, M.: Gesture Generation by Imitation - From Human Behavior to Computer Character Animation. Dissertation.com (2004)
17. Martin, J.C., Abrilian, S., Devillers, L., Lamolle, M., Mancini, M., Pelachaud, C.: Du corpus vido l'agent expressif: Utilisation des diffrents niveaux de reprsentation multimodale et motionnelle. Revue en Intelligence Artificielle RIA, Special Edition Interaction Emotionnelle **20** (2006)
18. Kita, S., van Gijn, I., van der Hulst, H.: Movement phase in signs and co-speech gestures, and their transcriptions by human coders. In: Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction, London, UK, Springer-Verlag (1998) 23–35
19. Heloir, A., Gibet, S., Multon, F., Courty, N.: Captured motion data processing for real time synthesis of sign language. In: Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop, GW 2005, Revised Selected Papers. Volume 3881 of LNAI., Berder Island, France, Springer (2005) 168–171
20. Majkowska, A., Zordan, V.B., , Faloutsos, P.: Automatic splicing for hand and body animations. In: ACM SIGGRAPH / Eurographics Symposium on Computer Animation. (2006) pp. 309–316
21. Kovar, L., Gleicher, M., Pighin, F.: Motion graphs. In: Proceedings of the 29th Annual Conference on Computer Graphics and InteractiveTechniques, ACM Press (2002) 473–482
22. Brentari, D.: Prosodic constraints in american sign language. In: Proc. of the 20th Annual Meeting of the Berkeley Linguistics Society. (1994)
23. Brentari, D., Poizner, H.: A phonological analysis of a death parkingsonian signer. Languages and cognitive processes **9** (1994) 69–99
24. Miller, C.: Regards sur la phonologie des langues signes. Recherches Linguistiques de Vincennes. **29** (2000) 101–120
25. Heloir, A., Courty, N., Gibet, S., Multon, F.: Temporal alignment of communicative gesture sequences. Computer Animation and Virtual Worlds **17** (July 2006) 347–357
26. Kendon, A.: Gesticulation and speech: Two aspects of the process of utterance. In: The Relation Between verbal and Nonverbal Communication., Mouton, The Hague (1980) 207–227
27. Wachsmuth, I.: Communicative rhythm in gesture and speech. Lecture Notes in Computer Science **1739** (1999) 277–??
28. Neill, D.M.: Hand and Mind. What Gesture Revevals About Thoughts. University of Chicago Press (1992)
29. Lebourque, T., Gibet, S.: High level specification and control of communicative gestures: the gessyca system. In: Computer Animation, 1999. Proceedings. (26-29 May 1999) 24–35