A Qualitative and Quantitative Characterisation of Style in Sign Language Gestures

Alexis Héloir, Sylvie Gibet

A Qualitative and Quantitative Characterisation of Style in Sign Language Gestures

Alexis Heloir and Sylvie Gibet

Laboratoire Valoria, Université de Bretagne Sud.
Vannes, France.
alexis.heloir@dfki.de
sylvie.gibet@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 sign language gesture sequences. A qualitative and quantitative comparison of styled gesture sequences is made. This comparison leads to the identification of temporal, spatial, and stuctural processes that are described in a theoritical model of sign language phonology. Insights raised by this study are then considered in the more general framework of gesture synthesis in order to enhance existing gesture specification systems.

1 Introduction

Embodying a virtual humanoid with expressive gestures requires the taking into account of the properties that influence the perception of convincing movements. One of these properties is style. Style, on the one hand, enhances, augments and colours the inherent meaning of a message. On the other hand, style carries information about the speaker's age, gender, cultural background, and emotional state. As a consequence, style contributes to making a virtual humanoid more convincing and may increase its acceptance by 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 literature dedicated to 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 also 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 deals with 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 1960's have 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. Projects dedicated to geometrically based modelling of sign language [6–8] have appeared more recently.

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

Recent studies dedicated to expressive gesture rely on segmentation and annotation of gesture. Those studies and are aimed at characterising the spatial structure of a Sign Language phrase [8], investigating the systematic synchrony between modalities [14] or transcribing and modelling gestures with the goal of further resynthesis [15].

In this paper, we rely on a slightly modified version of the transcription scheme proposed by Kita et al. [16]. This transcription scheme provides a description 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 presents the segmentation process more thoroughly.

3 Motion Acquisition and Segmentation

This section presents the motion data on which we relied for this study. We then present our gesture segmentation scheme that is inspired by previous work dedicated to the segmentation of expressive gestures.

3.1 Motion Data

The motion data on which we conducted the study has been obtained thanks to an original motion capture process [17]. This process relies on two complementary techniques. The first technique aims at capturing facial and body motions

thanks to a set of reflective markers placed on standardized anatomical landmarks and a network of 12 *Vicon-MX*[1] infrared cameras located all around the subject. The second technique aims at capturing finger motion thanks to a pair of *Cybergloves*[2] measuring finger abduction and flexion. This technique is well suited for finger motion, as it is robust to visual occlusions of fingers that may appear too often during acquisition. The two devices are synchronised together and the data acquired is made available in a single file after postprocessing. By the more, a video stream is used for manual annotation of the sequences while three dimentional data serves as a base for numerical computation and comparison on the motion.

The capture sequences have been performed by a professional deaf instructor. The sequences are mainly composed of a succession of lexical units. These sequences depict the same weather forecast presentation. The difference among the sequences is in the emotional content the signer was asked to mimic during the gesture performance. No perceptive validation has been performed over the signer's performance. Three gesture sequences have been used in this study: in the first one, the signer was asked to be neutral, for the second one, the signer was asked to mimic anger, and for the third one, weariness. The gesture sequence can be depicted as the following gloss sequences :

{ HELLO, PLEASE, LOOK_AT_ME, TODAY, SIX, JULY, WEATHER, BRITTANY, MORNING, CLOUDS_CROSSING, AFTERNOON, RAIN, TOMORROW, SUN, SWIM, GO, MUST, FRIDAY, CLOUDS, EVENING, CLOUDS, MORNING, FOG, SEE_YOU_SOON, GOOD_BYE }

The following table depicts the main characteristics of the gesture sequences :

Table 1. gesture sequences details

Sequence	style	Duration (s)	number of glosses
1	neutral	45	25
2	angry	33	25
3	weary	52	25

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 [16]. This method was originally applied to manual annotation of gesture recorded as video streams. However, in our case it makes sence to take advantage of the extra information that is provided by the three dimensional reconstruction of the motion. Although Kita's model has already been adapted to automatic segmentation of three dimensional motion data in [18]. This method did not provide

---

[1] http://www.vicon.com/products/viconmx.html
[2] http://www.immersion.com/3d/products/cyberglove.php

us with relevant segmentation over sign language gesture sequences, we thus prefered to rely on automated analysis only to provide hints that could help a human annotator to take a decision during a manual segmentation process.

According to [16], 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 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 [19]. The sum of squared distances between the corresponding joint positions ($\vec{P_j}$ and $\vec{P_j'}$) gives the distance between two consecutive frames: $D(A_i, A_{i+1}) = \sum_j \|(\vec{P_j} - \vec{P_j'})\|$ We assume that a hand configuration change occurs when $D(A_i, A_{i+1})$ exceeds two standard deviation from the mean distance value computed 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 [15]. The segmentation was performed througouth the three styled gesture sequences and lead to the identification of the movement phases described in [16] and summarized in the following.

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

Both a qualitative and quantitative comparison of the gesture sequences is presented in the next section.

4 Temporal Variations and Gesture Phases Differences Between Styles

This section presents the temporal characterization of style among the gesture sequences that have been segmented in the previous section. The results focus on the coordination between articulators and the temporal variations between simple gestures performed according to different styles.

4.1 Coordination Between Channels

A study was conducted 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.
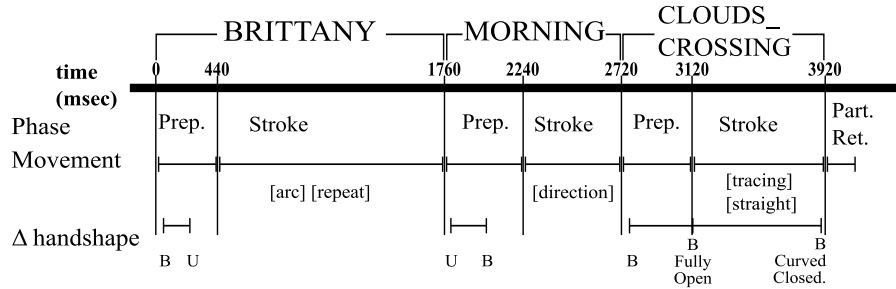

Fig. 1. Segmentation of gesture phases and temporal relations between hand and arm movement, from [1].

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 2.

Table 2. $HS\Delta/Mov$ 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 2 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]. Considering gesture specification, these 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 literature dedicated to sign language linguistics, timing variation may occur at several levels. In this subsection, we focus on the temporal variation which occurs between simple gesture movements. 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[3]. The timing distance evaluation was performed thanks to a dynamic time warping algorithm, described in detail in [20]. Dynamic Time Warping finds a non-linear timing adjustment between two time series that minimizes their global distance. The warp paths 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.
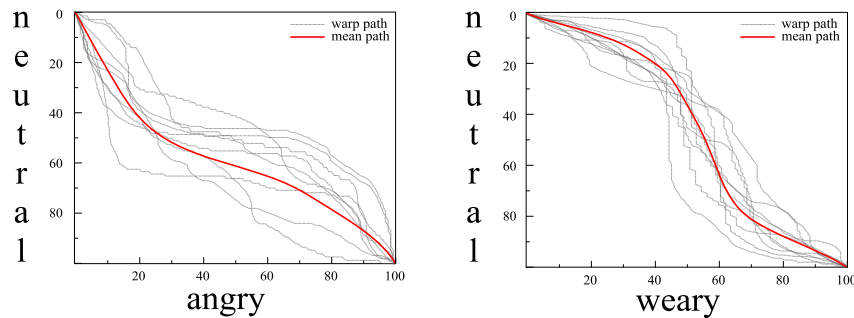

Fig. 2. in-sign temporal alignement between styles

Fig. 2 highlights the non-linear relationship that links styled gestures. The light curves represent warp paths that were obtained by the DTW algorithm

---

[3] Timing units can be referred as segments in the literature dedicated to sign language phonology [1]

between 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 of a preparation phase and a stroke phase. The comparison was done between the neutral and angry style Fig. 2.a, then between the neutral and weary style Fig. 2.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 movement in the lexical unit and the third phase stands for the second timing unit of the lexical unit. On the one hand it seems 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 Hermite splines.

4.3   Phase Repartition

Table 3 presents the number and the distribution 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 [16], retraction phases occur at the end of movement phrases or movement units. Thus, retraction phases highlights the end of an *idea unit*. Finally hold phases seems to occur more frequently in the *weary* sequence and less frequently in the *angry* sequence.

Table 3. Frequency of phase 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 depicted in Table 3 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 rythm [21] in the gesture sequence.

5 Spatial and Structural Features

This section highlights the Spatial and structural variations that have been observed between styled elementary gestures. The described phenomenons have been depicted in the sign language phonology litterature [1].

5.1 Proximalisation/Distalisation

A sign gesture may undergo variations in its spatial extend. In other words, the trajectory described by the signers wrist may be dilated or contracted in the signers signing space. This phenomenon is described in the context of enhancement as proximalisation/distalisation [1]. Thus, a sign may surface in a way that the initial joints mainly involved in the execution of the movement 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. 3.a , where an initial movement, WIND (Fig. 3.a center) may surface as distalised (fig. 3.a left) (the movement is mainly performed by the wrist) or may surface as proximalised (Fig. 3.a right): In this other example, the movement is performed by both the shoulders and the elbows. Another example is given in Fig. 3.b where the initial movement (YOU_MUST), originally performed mainly by the elbow (Fig. 3.b center) may surface as distalised (Fig. 3.b left) where the movement is mainly performed by the wrists. Finally, the movement may surface as proximalised (Fig. 3.b right), with the shoulder joints participating in the movement.
These phenomenons may convey, as highlighted in the examples, insights about the internal and emotional state of the signer, but, as stated in [12, 13], they may also provide information about the personnality, cultural and social background of the performer.
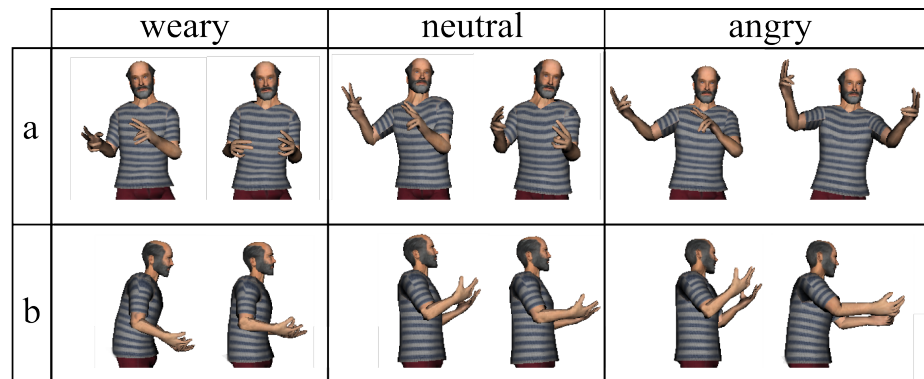

	weary	neutral	angry
a			
b			

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 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 (Fig. 4.a). 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 to undergo weak drop while other valid candidates do not {GO, MUST, CLOUDS, NIGHT, SEA, EVENING, PLEASE, BRITTANY}.

Although the context favoring the occurance of 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 occurence of the weak drop operation.
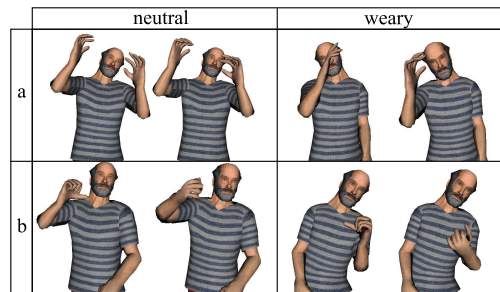

Fig. 4. example of Weak Drop and Hand Inversion performed by a synthetic agent

5.3 Strong Hand/Weak Hand Inversion

Along the gesture sequences, we were able to identify one occurence of the hand inversion phenomenon. This phenomenon consists of performing a one-handed sign thanks to the weak hand (H2) rather than the strong hand (H1). The operation is depicted in Fig. 4.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.

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 previous sections. We thus plan to extend the Lebourque et al. gesture specification system [5], by introducing new features in order to take into account:

- parts of the motion like retraction or preparation phases;
- features observed in gesture phases;
- spatial variation between styles gesture phases;
- 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. The 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, which is in accordance with the style dictionary paradigm proposed in [13]. We briefly present the main insights introduced by the gesture analysis presented in this paper and how they fit in our gesture specification and generation framework. Fig. 5 depicts the overall organisation of the framework.

1. At the phrase level, by introducing gesture phases and specifying different classes of gestures { *preparation, retraction, partial retraction, stroke and hold* }, we are able to represent gesture phases in a more accurate and realistic fashion;
2. At a phonological level, phenomenons presented in section 5, such as weak drop or hand inversion could be taken into account by considering the surfacing process as a constraint satisfaction problem (CSP);
3. Proximalisation/distalisation is handled at a gesture level by expanding or shrinking the trajectories of end effectors involved in a gesture. This operation rely on inverse kinematics (IK) algorithms;
4. By specifying explicit timing relations between timing units and gesture articulators, we are able to specify faithful descriptions of gestures. In addition, the correspondence between timing profiles highlighted in 4.1 leads to time warp patterns that provide realistic timing deformations between styles.

The variation parameters are the result of the combination of several variational layers. These layers are ordered according to their temporal extend (fig. 5.a) : for instance, agent inners characteristics (sex, culture) are permanent, while discourse context stands for the duration of a discourse and prosody has even shorter temporal resolution. The combination of these attributes leads to a set of variational attributes that parameterize the transformation steps described above.
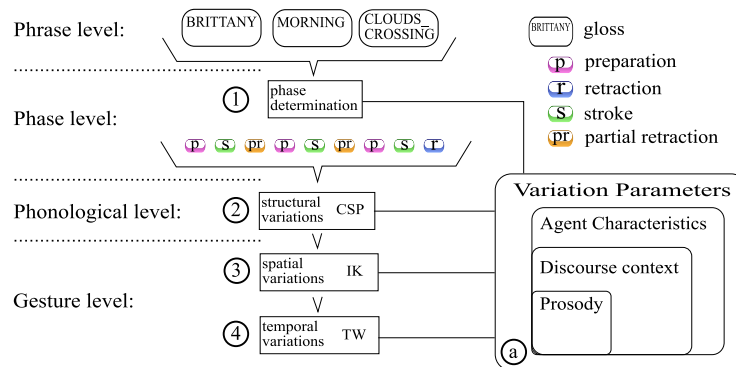
Fig. 5. Overview of the styled gesture gesneration process "clouds are crossing brittany in the morning"

7 Conclusion and Future Works

We have presented both a qualitative and quantitative study of the influence of style over expressive gesture sequences. We have highlighted how gesture may be modified by style at a structural, temporal and spatial level. This study leads to the design of an original framework dedicated to the specification and generation of styled sign language utterances. The way gestures are modified is taken into account by a set of variational parameters. These parameters are influenced by agent inner characteristics, discourse context and prosody. We are currently working on validating each block of the generative framework by implementing relevant demonstrators. a future step will consist into arranging every block in a consistent general demonstrator.

8 Acknowlegdements

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.: Ham-NoSys. Version 2.0. International Studies on Sign Language and Communication of the Deaf (1989)
4. Zwitserslood, 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.: Spatialized semantic relations in french sign language: Towards a computational modelling. In: Proc of Int. Gesture Workshop. (2006)
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. 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
11. Wallbot, H.: Bodily expression of emotion. European journal of social psychology **28** (1998) 879–796
12. Hartmann, B., Mancini, M., Pelachaud, C.: Implementing expressive gesture synthesis for embodied conversational agents. In: gesture in human-Computer Interaction and Simulation. (2006)
13. 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'.
14. 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
15. Boca Raton, F., ed.: Gesture Generation by Imitation - From Human Behavior to Computer Character Animation. Dissertation.com (2004)
16. Kita, S., van Gijn, I., van der Hulst, H.: Movement phase in signs and co-speech gestures, and their transcriptions by human coders. In: Proc. of the International Gesture Workshop, GW1997, London, UK, Springer-Verlag (1998) 23–35
17. Heloir, A., Gibet, S., Multon, F., Courty, N.: Captured motion data processing for real time synthesis of sign language. In: Proc. of International Gesture Workshop, GW 2005. Volume 3881 of LNAI., Berder Island, France, Springer (2005) 168–171
18. 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
19. 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
20. Heloir, A., Courty, N., Gibet, S., Multon, F.: Temporal alignment of communicative gesture sequences. Computer Animation and Virtual Worlds **17** (July 2006) 347–357
21. Wachsmuth, I.: Communicative rhythm in gesture and speech. Lecture Notes in Computer Science **1739** (1999) 277–289