Challenges for the Animation of Expressive Virtual Characters: The Standpoint of Sign Language and Theatrical Gestures

Sylvie Gibet, Pamela Carreno-Medrano, Pierre-François Marteau

Challenges for the Animation of Expressive Virtual Characters: the Standpoint of Sign Language and Theatrical Gestures

Sylvie Gibet[1] and Pamela Carreno-Medrano[1] and Pierre-Francois Marteau[1]

1 IRISA, University of Bretagne Sud, Campus de Tohannic, rue Yves Mainguy, F-56017 Vannes
e-mail: sylvie.gibet@univ-ubs.fr

Abstract

Designing and controlling virtual characters endowed with expressive gestures requires the modeling of multiple processes, involving high-level abstract representations to low-level sensorimotor models. An expressive gesture is here defined as a meaningful bodily motion which intrinsically associates sense, style, and expressiveness. The main challenges rely both on the capability to produce a large spectrum of parametrized actions executed with some variability in various situations, and on the biological plausibility of the motion of the virtual characters. The goals of the paper are twofold. First we review the different formalisms used to describe expressive gestures, from notations to computational languages. Secondly we identify and discuss remaining challenges in the generation of expressive virtual characters. The different models and formalisms are illustrated more particularly for theatrical and sign language gestures.

1 Introduction

Performing skilled gestures and movements, eventually in interaction with users and the environment, requires a thorough understanding of the different levels of representation that underlay their production, from the construction of sense to the elaboration of motor programs that prefigure motion performances. This is even truer for meaningful and expressive gestures which involve high level semiotic and cognitive representations, and require rapidity, accuracy, and physical engagement with the environment.

During the past decades, methods and approaches describing and modeling realistic human movements have been largely investigated by the research community in many areas such as motion analysis, recognition and synthesis, and contributed to a wide range of applications using virtual characters or humanoid robots.

However, if technologies of interactive embodied conversational agents are now available to researchers, there are still many open challenges concerning the animation of virtual characters endowed with expressive behavior. As pointed out by Thiebaux et al., virtual humans must be *believable*, *interpretable*, and *responsive*.[1]

A believable character can be defined as ”one that provides the illusion of life” .[2] This can be characterized by the perceivable behavior, in terms of motion consistency and quality, as well as the avatar’s appearance. The interpretability, that we will call here comprehension, concerns both the user’s ability to understand the message conveyed by the avatar’s gestures, and the expressive information encoded in the movement. The responsiveness, which involves the property of reactivity, is related to the ability of the virtual character to respond to events from the environment, and in particular it makes possible the interaction with the user. To go further, researchers aim at designing compelling characters capable of creating a more intuitive and engaging interaction with a user.

In this paper, beyond the properties of believability, comprehension, and responsiveness, we also aim at highlighting the main difficulties in characterizing, modeling and producing expressive behaviors driven by an underlying semantics. We review the existing formalisms and concepts used to describe expressive gestures, from notations to computational languages, how this specification may influence the produced movements, and discuss the remaining challenges for animating expressive virtual characters. We do not pretend to provide an exhaustive overview of the different technologies used to create credible virtual characters, as proposed for human-computer dialog,[3] but focus more specifically on full-body movements which draw the user’s attention, and express through body language some meaningful and emotional intent. The different issues will be addressed for two categories of movements (i) theatrical gestures which are demanding in the production of believable and engaging movements, and, (ii) sign language gestures which involve highly structured movements driven by constrained linguistic rules, thus pushing the comprehension to a demanding level. Both categories of movements implicitly contain a strong semantics and involve complex cognitive, linguistic and sensorimotor mechanisms, from story telling to the production of movements whose detailed components may contain significant elements perceivable by humans.

2 Requirements for Producing Expressive Gestures

Understanding the mechanisms involved in the production of meaningful and expressive gestures implies strong directions for future research. Some essential and not yet accomplished characteristics that make the production of such gestures highly complex are exposed below, such as multi-modality, spatial content, coordination/synchronization rules, and expressiveness. These requirements are highlighted and illustrated in the context of the execution of sign language gestures and theatrical movements.

2.1 Multimodal components

Gestures are not restricted to conveying meaning solely through body movements; instead, they require the simultaneous use of body, hands' movements, facial mimics, gaze direction, and speech. In sign languages (SL), we generally separate the manual components which include hand configuration, orientation, and placement or movement, expressed in the signing space (the physical three-dimensional space in which the signs are performed), from non-manual components consisting of the posture of the upper torso, head orientation, facial expression, and gaze direction. For example, eye gaze can be used to recall a particular object in the signing space; it can also be necessary to the comprehension of a sign, as in the sign DESSINER(v) corresponding to the action of drawing, and for which the eyes follow the motion of the fingers as in drawing.

In dance, the eye gaze may be used for balance purpose, or in a magical trick for distracting the attention of the spectator. In theatrical gestures (TG), a role has been attributed to each non-manual component according to Delsarte's movement theory. For instance, the torso is considered as the main channel of emotional content while arms act as thermometers indicating how expressive a movement can be.[4]

In SL, facial mimics may serve as adjectives (e.g., inflated cheeks make an object large or cumbersome, while squinted eyes make it thin) or indicate whether the sentence is a question (raised eyebrows) or a negation (frowning). It is therefore very important to preserve this information during facial animation. Both in TG and SL, facial mimics convey the emotion and the state of mind of the signer/actor, which is useful for the comprehension of the various situations, but also for the credibility and the correctness of the movement as well as the expressiveness intent.

In theater, speech and its paralinguistic characteristics such as pitch, loudness, tempo, among others, "[..] constitute one of the most ancient objects of the actor's art."[5] Actors control the flow of information and segment discourse in order to increase the engagement and comprehension levels of the audience. They also use speech as a means of strengthening the information about the motivations and emotions behind the characters they personify.[5]

2.2 Spatial content

TG or SL use by nature spatial mechanisms, in particular for strongly iconic gestures, i.e. gestures that describe the shape of an object or an entity, or mime a situation in space. Both SL and TG execute movements involving a part or several parts of the body, for example raising a shoulder or nodding and shaking head.

In SL, spatiality is intrinsically linked to meaning and is mostly expressed through hand movement trajectories, or hand shapes (called configurations). Thus, spatial mechanisms, classically used in LSF (French Sign Language) are depicting or directional verbs which mimic spatial movements, as well as size-and-shape configurations which are static spatial descriptions. In TG, similarly to SL, the ori-

Fig. 1 LSF signs LIKE / LIKE-NOT (left), and GIVE / TAKE (right): the hand trajectories are reversed

entation of gesture in space, the starting and ending point in space of a movement, all impact the meaning of a gesture.[6]

Movement trajectories. In SL, indicating and depicting verbs require the signer to manipulate targets in the signing space by effecting pointing-like movements towards these targets. They include such signs as the LSF sign DONNER (v. give), in which the hand moves from the giver to the receiver. Depending on the intended subject and object, the initial and final placements of the hand, as well as its orientation vary greatly within the signing space; these placements may have syntactic meaning (subject, object, pronoun, etc.). Note that for such categories of signs, the target can be located on a specific part of the body. Other more accurate spatial mechanisms involve hand trajectories within signs, which are not only transitions in space between two key positions, but take the shape of a line, an arc, or a more complex form such as an ellipse, a spiral, etc. For example the sign AIMER (v., like) is represented by an upward arc-movement. An interesting inquiry is whether playing reversible indicating verbs backwards would be convincing to other signers, and whether altering the hand-shape of a stored depicting verb would be understood as a change in meaning. Thus, we can reverse the movement of AIMER to produce the meaning NE-PAS-AIMER (v., dislike, see Figure 1, right). In the same way, reversing the sign DONNER (v., give) may result in the LSF sign PRENDRE (v., take, see Figure 1, left). These mechanisms can lead to a gestural language where the trajectory is spatially coded.[7]

In TG, it has been suggested that the trajectory a movement describes in space can have an important expressive content as well as capture the overall tendency of the motion.[4] Changes in motion trajectory can also be used to emphasize and accentuate a movement in the eyes of the audience.

Shapes. Moreover, the depicting verbs can be performed with a generic hand-shape, or with a hand-shape that indicates the object has a cylindrical shape. Using hand-shapes, we also find signs in which one hand acts as the dominated hand, while the other is the dominant one. For example, in the case of the LSF sign AVION (plane), the flat dominated hand is representing the runway, the other one the plane taking off. Finally, the signing space may be decomposed into different relational spaces: the discourse space, in which the entities of the discourse are represented

and located. The topographic space, which expresses the spatial relationships between the entities, and may use embedded layouts of spatial descriptions. In TG, meanings and emotions are encoded into specific body postures that may be exaggerated in shape and duration to ensure the understanding of the interlocutor. The shape taken by the body concerns all contours of the bodies made in space. Shapes can be of three types: lines, curves, or combination of both performed in the three main planes (frontal, sagittal, or horizontal). They are stationary or moving through space and depict information about how a person interacts with her surroundings.[8]

Precision. In SL, comprehension of signs requires accuracy in their formation. Some hand-shapes differ only by the position of one finger or by whether or not it contacts another part of the hand or the body. In addition, the degree of openness of the fingers can be the sole differentiating factor between signs. This calls for notable accuracy in the motion capture and data animation processes. In TG, precision and simplicity in an actor's motion are fundamental for a clear and successful communication with the audience. TG are based on a combination of simplification (to bring focus to a particular element by eliminating all possible superfluous and ambiguous movement) and exaggeration (after simplifying a movement, emphasize its meaning by exaggerating it) principles.

2.3 Coordination/synchronization rules

In order to increase the controllability of the movements, it appears essential to be able to manipulate the movements at a finer grain than the postures themselves. This requires the precise decomposition of the body along channels that are significant to the manipulated motion type, and the definition of coordination/synchronization rules. Both in SL and TG, altering the spatio-temporal properties of the movements may deeply modify the meaning of the gestures. We give hereinafter different temporal aspects and constraints that characterize the execution of gestures.

For SL, the question of timing and dynamics of gesture is crucial. In fact, three elements are of interest for these gestures. Firstly, in SL, the kinematics characterizing the transition movements and the stroke conveying a specific meaning shows specific profiles that should be respected when synthesizing such gestures. Secondly, spatio-temporal synchronization rules between different parts of the body is a major component. In particular, phonetic studies have shown structural patterns with regular temporal invariants,[9] such as the hand configuration target which is systematically reached before the hand movement begins,[10] or the motion of the two hands which are very often synchronized, the dominant hand slightly preceding the non dominant hand. We may also observe some timing invariants between eye gaze and head movements, or eye-gaze and hand configuration. Thirdly, the dynamics of the gesture (acceleration profile along time) can be used to distinguish two meanings. An example is the difference between the LSF signs JOUER(v) (to play), and DÉTENDU (relaxed), which have the same hands configurations, the same tra-

jectories in space, but different dynamics. Let us finally note that the dynamics of contacts between the hand and the body (gently touching or striking) is particularly relevant.

In TG, temporal characteristics as tempo, duration and repetition are highlighted and can be used as a language of gesture for creating theatrical compositions.[8] Additionally, it is important to spend the correct amount of time on each movement. Otherwise, the audience will not be able to follow and interpret the inner intentions and motivations of the character personified by the actor. It is also possible to indicate a character's current state of mind by modifying the timing of its actions.

3 Previous Work

Many studies have addressed the problem of describing, categorizing, and formalizing computational models for generating movements. We summarize below some knowledge from experts in theatrical and sign language movements in terms of gesture descriptions and notations. We then derive the main trends used to animate expressive and meaningful virtual characters, and discuss the good points as well the main drawbacks in meeting the previous requirements.

3.1 Gesture Descriptions and Notations

Early work in linguistics has attempted to describe and categorize movements and gestures. For gestures conveying a specific meaning, called *semiotic* gestures, taxonomies have been proposed.They define *semantic* categories, i.e. semantic classes that can be discriminated and characterized by verbal labels. Kendon[11] is the first author to propose a typology of semiotic acts, making the hypothesis of a continuum between speech utterances and information conveyed by gestures. McNeill extends this typology with a theory gathering the two forms of expression, speech and action.[12] In these studies, both modalities are closely related, since they share a common cognitive representation. Furthermore, Kendon and McNeill[11,12] have proposed a temporal structure of gestures, above all for co-verbal gestures. This structure can be described in terms of phases, phrases, and units. It has been extended by Kita[13] who introduced the different phases (Preparation, Stroke, and Retraction) composing each significant unit, and used in the context of SL generation.[14]

In order to memorize and transcribe the gestures and their structural organization, movement notations and coding systems have also been developed. These systems generally aim at describing in an exhaustive and compact way labeled elements whose structure relies on a predefined vocabulary depending on the studied movement and context.

Laban Movement Analysis (LMA). Among these structural descriptions, the La-

ban Movement Analysis (LMA) theory initially defined for dance choreography identifies semantic components that describe the structural, geometric and dynamic properties of human motion .[15,16] This theory comprises four major components: Body, Space, Effort and Shape. Body and Space components describe how the human body moves, either within the body or in relation with the 3D space surrounding the body. Shape component describes the shape morphology of the body during the motion, whereas Effort component focuses on the qualitative aspects of the movement in terms of dynamics, energy and intent. LMA has been largely used in computer animation,[17–19] motion segmentation,[20] gesture recognition and affect analysis.[21,22]

Eshkol-Wachman notation system. Although initially developed for dance, it was also intended to notate and analyze any possible movement in space in a rather mathematical way.[23] The moving body is treated as a system of articulated axes in which each axis corresponds to a line segment of constant length connecting either two joints or a joint and a free extremity. The path described by each axis's end is parameterized using spherical-like coordinates. The notation system also describes three types of movements: rotatory, plane and conical and describes the degree of interdependence between limbs as *light* or *heavy*. All limbs are thus divided into relative classes: every limb is *heavy* relatively to any limb that it carries while moving, and *light* relatively to any limb by which it is being carried. This system has been used in a wide variety of fields like sports,[24] sign language,[25] medicine,[26] etc.

Delsarte notation system. It is a notation system based on Delsarte's (a French musician and actor) methodical observations of the human body and its interactions with others. Through his notation system, Delsarte described the relationship between meaning and motion, and how attitude and personality are conveyed by body parts and gestures.[27] Motions are classified into three categories according to the direction of movement: *eccentric*, motion away from the body center and having a relation to the exterior world; *concentric*, motions toward the body center and having relation to the interior; and *normal*, balanced motions moderating between concentric and eccentric motions. The body is divided into body zones, each zone having nine possible poses i.e., all combinations of the three types of motion. For each pose and each zone a meaning is attributed. Delsarte identified three orders of movement: *oppositions*, *parallelisms*, and *successions* as well as nine laws of motion (attitude, force, motion, sequence, direction, form, velocity, reaction and extension) that further modify the meaning of each movement. Delsarte system has been already used for the generation of virtual agents motions.[6,28]

Sign Language descriptions. The notion of decomposing signs into various components is not new to the linguistic community. In 1960, William Stokoe started his system of *Tab* (location), *Dez* (handshape), and *Sig* (movement) specifiers that were to describe any sign.[29] Since then, other linguists have expanded on Stokoe's decompositional system, keeping the location, hand-shape, and placement parameters, and introducing wrist orientation, syllabic patterning, etc.[9,30] All systems allow for

multiple configurations during the course of a single sign. In her 1981 work, Sutton presents the SignWriting model, using pictographic symbols placed spatially to represent sign components including their associated facial expressions. HamNoSys,[31] another pictographic notation system, is a Stokoe-based phonetic notation system, using a linear presentation of custom symbols with diacritics. However, this system provides no timing information in the form of sign segments, and thus makes an analysis of sign timing rather difficult. The system proposed by Liddell and Johnson introduces timing segments that divide signs into sequential parts in addition to the simultaneous parts that Stokoe had previously identified. Hand configuration, points of contact, facing, and orientation are described as static articulatory postures; movements allow for spatial and temporal progression between postures. Following Liddell and Johnson's phonetic patterns, and the grammatical decomposition proposed by Johnston and de Beuzeville,[32] Duarte et al. have developed their own annotation scheme which is used for the synthesis of signing utterances in French sign language.[33]

3.2 Animation of Virtual Characters

The modeling of human-like behavior leads to an intelligent virtual agent generally considered as deliberative, since it has the capability of decision, and reactive in the sense it can react to events. This requires the integration of both cognitive and reactive aspects, based on the will and intention of the agent, as well as on perceptuomotor processes occurring during the motor performances. Different trends in the research on cognitive architectures have recently emerged, highlighting the role of memory and learning in the design of intelligent systems that have similar capabilities to those of humans. Two surveys review various paradigms of cognition,[34] and various architectures among *symbolic*, *emergent*, and *hybrid* models.[35] However, there are very few cognitive architectures that are implemented and applied to the animation of virtual characters. Among these systems, the concept of Action / Perception / Decision has given rise to a programming environment for behavioral animation.[36]

Many levels have been defined for behavior planning and control, and for specification languages dedicated to expressive virtual characters.[3] Two major classes of approaches can be distinguished: (i) those that specify explicit "intelligent" behaviors dedicated to embodied conversational agents, and (ii) those offering data-driven animation techniques. Some hybrid frameworks combine these two approaches to respond to the requirements stated above.

Embodied Conversational Agents (ECA). Creating ECAs requires designing high-level behavior (planning, handle communicative acts, etc.), and producing coordinated and synchronized movements of multiple parts of the body, possibly associated with speech production: upper and lower body, head/neck, hands, facial expression, eye movements, speech. Regarding high-level gesture specification, his-

torical and current methods range from formalized scripts to dedicated gestural languages. The Behavior Expression Animation Toolkit (BEAT), as one of the first systems to describe the desired behaviors of virtual agents, uses textual input to combine gesture features for generation and synchronization with speech.[37] XML-based description languages have been developed to describe various multi-modal behaviors, some of which are dedicated to complex gesture specification,[38] describe style variations in gesture and speech,[39] or introduce a set of parameters to categorize expressive gestures.[40] More recently, some computational models consider the coordination and adaptation of the virtual agent with a human or with the environment in interacting situations. The models in such cases focus on rule-based approaches derived from social communicative theories.[41]

To facilitate the creation of interactive agents, recent work has proposed the SAIBA architecture, in which three main stages are identified, namely the intent planner, the behavior planner, and the surface realizer.[42,43] This software architecture is the basis for implementing various embodied characters with unified and abstract interfaces. The functional markup language (FML) is used to encode the communicative intent, whereas the behavior markup language (BML) specifies the verbal utterance and the nonverbal behaviors such as gesture or facial expression (e.g., pointing gesture, shaking hands, nodding head, etc.).

Passing from the specification of gestures to their generation has given rise to a few research work. Largely, this work aims at translating a gestural description, expressed in any of the above-mentioned formalisms, into a sequence of gestural commands that can be directly interpreted by a real-time animation engine. Most of the animation models rely on pure synthesis methods, for example by using inverse kinematics techniques (e.g., [44,45]).

More recently, novel languages and architectures, based on the SAIBA-BML behavior language have been proposed. The SmartBody[1] is an open source modular framework which hierarchically interconnects controllers to achieve continuous motion. It employs various animation algorithms such as key-frame interpolation, motion capture or procedural animation. The real-time system EMBR introduces a new animation layer of control between the behavioral level and the procedural animation level, thus providing the animator with a more flexible and accurate interface for synthesizing nonverbal behaviors.[46] This system also incorporates into the language expressive parameters (spatial extent, temporal extent, fluidity, and power).[40] Most of the proposed languages describe the behaviors in an explicit way, thus preventing the system's ability to respond reactively to external events, or to anticipate the movement of some body parts in complex tasks. Without offering an animation specification language, the PIAVCA architecture[47] proposes a functional abstraction of character behavior. It provides a range of motion editing filters that can be combined to achieve animations reactive to events.

The approaches using high-level specification languages coupled with interpolation or procedural animation methods allow for building complex behaviors, essentially by combining different controllers associated to different modalities. Moreover, based on psycho-linguistic rules or manual annotations, the generated move-

ments are consistent and precise. Finally, some scripting language may take into account expressiveness in terms of semantic labels or expressive parameters.

However, building by hand complex animations by specifying key postures or targets and synchronizing the different body parts in space and time has revealed to be a tedious task. Another drawback of such methods is the lack of believability for generating motion, except for those that use motion capture controllers. In order to ease and automatize the generation of novel movement sequences, it is necessary to take into account some movement knowledge in terms of structural spatio-temporal patterns, human motion rules (such as invariant motion laws), or statistical motion properties. To summarize, the most significant benefits of the ECA related methods are the controllability and the precision of the behavior of the virtual character, but this is achieved at the expense of ease of specification and believability.

Data-driven Synthesis. Alternatively, to achieve animation of highly believable and life-like characters, data-driven methods have replaced pure synthesis methods. In this case the movements of a real user are captured with different combinations of motion capture techniques. Motion graphs allow to generate realistic, controllable motion through a database of motion capture.[48] The authors automatically construct a graph that encapsulates connections among different motion chunks in the database and then search this graph for motions that satisfy user constraints. One limitation of the approach is that the transition thresholds must be specified by hand, which may prove to be a very tedious task.

Furthermore, machine learning techniques can be used to capture style in human movements and generate new motions with variations in style or expressiveness.[49–52] In these studies authors consider a low-level definition of style, in terms of variability observed among several realizations of the same gesture. If some relevant studies rely on qualitative or quantitative annotations of motion clips (*e.g.*,[53,54]), or propose relevant methods to create a repertoire of expressive behaviors (*e.g.*,[55]), very few approaches deal with both motion-captured data and their implicit semantic and expressive content. Within their framework, Stone et al. synchronize meaningfully gesture and speech by specifying the organization of characters' utterances and generating automatically the animation of the conversational character.[56] The authors rely on an annotation process that indicates the perceptually prominent moments of emphasis in speech and gesture. To animate gesturing characters, Jörg et al. develop a motion retrieval method to automatically add plausible finger motions to body motions, extracting the finger motions from a database, according to the similarity of the arm movements and the smoothness of finger motions.[57] To create natural-looking motions of characters that follow users scenarios, Safonova et al. provide a sketched-based method associated to a motion graph representation to approximatively specify the path of the character and adapt the existing motions through interpolation.[58]

These approaches give satisfactory results in terms of believability, since they use postures or motion chunks selected in a pre-defined database. It still remains difficult to parameterize motion and to produce controllable and flexible behaviors.

In addition, the reuse of motion data does not give the ability to generate novel movements far from existing ones. Another drawback is the lack of responsiveness of such fully data-driven approach.

4 Remaining Challenges to Animate Expressive Virtual Character

In this section only data-driven methods are considered, as they particularly meet the necessary requirement of believability of the produced animated sequences. Though these methods significantly improve the believability of the animations, there are nonetheless several remaining challenges to the reuse of motion data. The main one is the transformation and recombination of motion capture data elements in the production of behaviors that preserve the movement's sense and the emotional intent. We discuss hereafter these challenges following the requirements evoked in section 2.

4.1 Constructing resources with meaning and expressiveness

Data acquisition. Signs and theatrical gestures are by nature expressive and dexterous gestures, which simultaneously involve several modalities (arms, hands, body, gaze and facial expressions). Capturing accurately and synchronously all these channels with an appropriate frequency ($>$ 100 Mhz) actually pushes motion capture equipment to their limits. Novel technologies such as surface capture,[59] that captures simultaneously geometry and animation, are very attractive, but yet the resolution is not sufficient to capture the body and the face with an adequate precision, and only few methods exist to manipulate this complex data in order to produce new animations.

Nature of the gesture corpus. For the purpose of corpus design, several questions have to be addressed. The first one concerns the corpus definition and the compromise that exists between breadth and depth in its design. If the objective of the synthesis system is to have a lexicon that covers a broad area, including several thematic domains, then a corpus with a breadth approach would be suitable. If, instead, the goal is to have a limited vocabulary and reuse it in different sentences, then the depth approach would be best. In this case, many tokens of the same signs or actions will be preferred in the predefined vocabulary, with variations depending on the scenario context. The second question concerns the nature of the variations that should be included in the corpus for further editing and synthesis. Several levels of signs variability can be considered: we may think about incorporating multiple tokens of the same action/sign in different contexts, in order to be able to construct new sentences that take into account the spatial variations as well the co-articulation

aspects. For example, a magician in theatrical gestures might want to perform his trick in different locations in space, or describe objects of different shapes and sizes. The problem is similar in SL, but with finer motion elements (manipulating short hand movements or hand configurations). The signing context can also be taken into account by capturing the same sign in varying its predecessors and successors (e.g. influence of hand shape and placement). The inclusion of such sequencing in the corpus allows for the study of co-articulation. Therefore, if the actions/signs involving such inflection processes are contained into the corpus, then the editing operations will be less complex. For example, including many depicting verbs with spatial variation will facilitate the construction of novel utterances with verb declination, without recording new signs. Another source of variation is the style of the actor/signer, or the deliberate emotional state contained in the scenarios, which lead to kinematic variations. A closely linked question concerns the acted or spontaneous nature of the produced scenarios.

4.2 High level language for multichannel editing

The choice of the computing language allowing the description of behaviors that can be interpreted by the animation controllers is still very challenging to the computer animation community, above all for communicative and expressive behaviors involving high level semantic rules. Most of the time, these behaviors concern the combination and scheduling of elementary behaviors attached to dedicated controllers such as keyframe interpolation, motion capture, or procedural animation. This approach does not consider the coordination of finer-grain motion which is necessary when dealing with communicative gestures. Using a predefined dual database, one containing the raw motion and the other annotated data, it becomes possible to build novel phrases, by selectively composing and blending pre-existing elements along temporal segments and spatial channels.[60] In this scope, it is necessary to consider all the unsolved spatial and temporal issues raised by the editing process.

Inflecting spatial variations. When dealing with motion capture data, it is very difficult to generate new movements that are not contained in the database. However, a main challenge would be to define generic and parameterized controllers that enable the generation of similar motions varying in space, for example if we want to modify the location of a gesture (up-right or down-left), or the size of an object (showing a small or big box).

Spatial coherency. Another challenge would be to combine different spatial channels with different meanings simultaneously. This coordination differs from the classical blending approaches which mix whole skeleton motions to produce new ones.[53] An example is given in Figure 2 which illustrates the construction of the sentence: ”I don’t like orange juice” in LSF. Different channels are combined, by

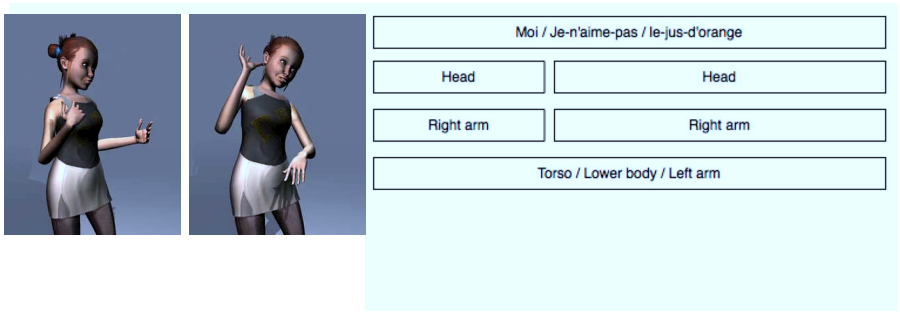

Fig. 2 Combination of three signs: moi / je-n'aime-pas / le-jus-d'orange (I don't like orange juice).

keeping the torso/lower body/left arm of one sequence, and substituting the head, facial expression and right arm movements of another sequence. In such a composition process, the spatial constraints should be preserved, in particular the sign should be executed near the corresponding body part, whatever the torso or the head orientation is. This clearly reveals that the combination process should be driven at a more abstract level, expressed by rules or constraints incorporated into the controllers.

Inflecting temporal variations. It is likely that the different motion elements have not the same duration. The subsequent problem is twofold: *i)* a common timeline has to be found, eventually as the result of a combinatorial optimization, or driven by linguistic rules. Up to our knowledge though, no existing gestural language describes such temporal rules or models the synchronization of the different channels *ii)* once a correct time plan has been devised, the temporal length of the motion chunks has to be adapted, while preserving the dynamics of the motions. To this end, time warping techniques can be used.[61] However, inter channels synchronizations may exist (for example between the hand and the arm motions[62]). Thus synchronization schema can be extracted from analysis, but the proper way to introduce this empirical knowledge in the synthesis process has not been explored yet.

4.3 Dealing with expressiveness

Virtual characters portray their emotions through movement[63], thus as stated by Byshko[64], the perception of a virtual character has everything to do with how it moves. Unfortunately, in spite of the numerous psychological studies and computer animation and machine learning applications trying to decode and exploit the most salient features to human expressiveness, there is still no common understanding about how affect, style and intent are conveyed through human movement. We know for example that in SL, the spatio-temporal variability of signs can be used to inflect

the nature of a sentence and enhance the global expressiveness and style of the virtual signer. However, small spatial or temporal variations may deeply alter the meaning of a sentence.

No field has studied character movement more intently than the performing arts, since their prime goal is to create visually affective and believable characters capable of communicating meaning and emotion to an audience. Therefore, the theatrical body movements can be of interest and employed as a source of inspiration in the understanding of expressive human movement it-self, and in turn exploited in the synthesis of avatars' expressive movements. The reasons behind this idea are threefold:

i) In the creation of a theater act it is required to develop a deep understanding of "the language of gesture"[65], since it is through movement/gesture that an actor transforms feelings, emotions, intentions, and passions into performance and meaning. By analyzing and understanding the conventions, ideas and techniques employed by theater actors while creating and embodying a character, we may be able to apply similar principles while designing virtual characters with rich emotional expressions.

ii) While in stage, every movement is deliberately chosen and executed to induce/involve the audience with emotion[4], and thus make every character in scene to be perceived as believable. By using TG as the knowledge base of a motion synthesis system, it is likely that any virtual character will also be perceived as believable and hence the user will be part of a very engaging and meaningful interactive experience.

iii) In physical theater, the body and its movement are both the center of attention and the center of the theater making process.[66] As spectators, we invest every performer's action and gesture with significance, meaning and emotional/affective content. By studying and analyzing theatrical gestures, we think it is possible to gain an additional insight on how meaning and emotions are conveyed through movement.

5 Conclusion

We have examined in this article the different challenges posed by the animation of advanced expressive virtual characters, according to different aspects that are essential to enhance the believability, comprehension, and responsiveness properties. While data-driven animation techniques clearly show the best believable results, a lot of improvements are still mandatory to fulfill the requirements of theatrical gestures and sign languages production which both require highly nuanced variations, while keeping a strong semantic. Among others, motion capture and corpus building are difficult issues which require significant studio time with experts, and are very costly in post processing. The design of a new computer specification language, including some reactivity in the specification of gestures, and controllers that respect the adaptive motor program constraints should enable the synthesis of responsive gestures. Incorporating both procedural and data driven models with machine learn-

ing capabilities is still very challenging, since it allows to combine the generation of realistic movements while giving the possibility to manipulate parameterized behaviors, thus leading to a better control of the synthesis. Finally, the usability and acceptability of virtual characters to the expert communities should also be evaluated thoroughly, notably through the help of professional actors and native signers. Though those issues have recently attracted the attention of several research groups, a lot remain to be done before comprehensive, believable and reactive avatars can be truly effective in our everyday life virtual environments.

Acknowledgements This work is part of the *Ingredible* project, supported by the French National Research Agency (ANR).

References

1. M. Thiebaux, S. Marsella, A.N. Marshall, and M. Kallmann. Smartbody: Behavior realization for embodied conversational agents. In *Proceedings of AAMAS '08 - Volume 1*, pages 151–158, Richland, SC, 2008.
2. J. Bates et al. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.
3. Y. Jung, A. Kuijper, M. Kipp, J. Miksatko, J. Gratch, and D. Thalmann. Believable Virtual Characters in Human-Computer Dialogs. In *EUROGRAPHICS*, Llandudno, United Kingdom, April 2011.
4. M. Neff. *Nonverbal Communication in Virtual Worlds: Understanding and Designing Expressive Characters*, chapter Lessons From The Arts: What The Performing Arts Literature Can Teach Us About Creating Expressive Character Movement. ETC Press, 2014.
5. K. Elam. *The Semiotics of Theatre and Drama*. Routledge, 2002.
6. S. C Marsella, S. M. Carnicke, J. Gratch, A. Okhmatovskaia, and A. Rizzo. An exploration of delsartes structural acting system. In *Intelligent Virtual Agents*, pages 80–92. Springer, 2006.
7. S. Gibet, T. Lebourque, and PF. Marteau. High level specification and animation of communicative gestures. *Journal of Visual Languages and Computing*, 12:657–687, 2001.
8. A. Bogart and T. Landau. *The Viewpoints Book: A Practical Guide to Viewpoints and Composition*. Theatre Communications Group, 2005.
9. R.E. Johnson and S.K. Liddell. Toward a phonetic representation of signs: Sequentiality and contrast. *Sign Language Studies*, 11:2:241–274, 2011.
10. K Duarte. Motion capture and avatars as portals for analyzing the linguistic structure of signed languages. In *PhD thesis, Université de Bretagne Sud*, 2012.
11. A. Kendon. Gesticulation and speech two aspects of the process of utterance. In *The Relation Between Verbal and Nonverbal Communication*, pages 207–227, 1980.
12. D. McNeill. *Hand and Mind - What Gestures Reveal about Thought*. The University of Chicago Press, Chicago, IL, 1992.
13. S. Kita, I. van Gijn, and H. van der Hulst. 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*, volume 1371 of *Lecture Notes in Computer Science*, pages 23–35. Springer-Verlag, London, 1997.
14. C. Awad, N. Courty, K. Duarte, T. Le Naour, and S. Gibet. A combined semantic and motion capture database for real-time sign language synthesis. In *IVA*, pages 432–438, 2009.
15. R. Laban. *The Mastery of Movement.* Plays, Inc., 1971.
16. V. Maletik. *Body, Space, Expression: The Development of Rudolf Laban's Movement and Dance Concepts.* Mouton de Gruyte, New York, 1987.

17. D.M. Chi, M. Costa, L. Zhao, and N.I. Badler. The EMOTE model for effort and shape. In *SIGGRAPH*, pages 173–182, 2000.
18. L. Zhao and N.I. Badler. Acquiring and validating motion qualities from live limb gestures. *Graphical Models*, 67(1):1–16, 2005.
19. L. Torresani, P. Hackney, and C. Bregler. Learning motion style synthesis from perceptual observations. In *Advances in Neural Information Processing Systems*, pages 1393–1400, 2006.
20. D. Bouchard and N. I. Badler. Semantic segmentation of motion capture using laban movement analysis. In *Intelligent Virtual Agents (IVA 2007)*, pages 37–44, 2007.
21. M. Karg, A. Samadani, R. Gorbet, K. Kühnlenz, J. Hoey, and D. Kulic. Body movements for affective expression: A survey of automatic recognition and generation. *T. Affective Computing*, 4(4):341–359, 2013.
22. A. Kleinsmith and N. Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *T. Affective Computing*, 4(1):15–33, 2013.
23. N. Eshkol and J. Harries. Ewmn: Eshkol-wachman movement notation, 2011. [Online; accessed 10-March-2015].
24. J.G. Harries. Symmetry in the movements of t'ai chi chuan. *Computers & Mathematics with Applications*, 17(46):827 – 835, 1989.
25. N. Eshkol. *The hand book : the detailed notation of hand and finger movements and forms*. [Tel Aviv] : Movement Notation Society, 1971.
26. O. Teitelbaum, T. Benton, P. K Shah, A. Prince, J. L Kelly, and P. Teitelbaum. Eshkol–wachman movement notation in diagnosis: The early detection of asperger's syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 101(32):11909–11914, 2004.
27. J. Tanenbaum, M. Seif El-Nasr, and M. Nixon. *Nonverbal Communication in Virtual Worlds: Understanding and Designing Expressive Characters*, chapter Basics of Nonverbal Communication in The Physical World. ETC Press, 2014.
28. M. Nixon, P. Pasquier, and M. S. El-Nasr. Delsartmap: Applying delsartes aesthetic system to virtual agents. In *Intelligent Virtual Agents*, pages 139–145. Springer, 2010.
29. W. C. Stokoe. *Semiotics and Human Sign Language*. Walter de Gruyter Inc., 1972.
30. D. Brentari. *A Prosodic Model of Sign Language Phonology*. MIT Press, Cambridge, MA, 1999.
31. S. Prillwitz, R. Leven, H. Zienert, T. Hanke, and J. Henning. *Hamburg Notation System for Sign Languages - An Introductory Guide*. University of Hamburg Press, 1989.
32. T. Johnston. The lexical database of AUSLAN (Australian Sign Language). In *Proceedings of the First Intersign Workshop: Lexical Databases*, Hamburg, 1998.
33. K. Duarte and S. Gibet. Heterogeneous data sources for signed language analysis and synthesis: The signcom project. In *Proceedings of LREC'10*. ELRA, 2010.
34. D. Vernon, G. Metta, and G. Sandini. A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *Trans. Evol. Comp*, 11(2):151–180, April 2007.
35. W. Duch, R. J. Oentaryo, and M. Pasquier. Cognitive architectures: Where do we go from here? In *Proceedings of the 2008 Conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference*, pages 122–136, Amsterdam, The Netherlands, The Netherlands, 2008. IOS Press.
36. F. Devillers, S. Donikian, F. Lamarche, and J.F. Taille. A programming environment for behavioural animation. *Journal of Visualization and Computer Animation*, 13(5):263–274, 2002.
37. J. Cassell, J. Sullivan, S. Prevost, and E. F. Churchill. *Embodied Conversational Agents*. The MIT Press, 2000.
38. A. Kranstedt, S. Ko, and I. Wachsmuth. MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents. In *Proceedings of the AAMAS02 Workshop on ECA*, July 2002.
39. H. Noot and Z. Ruttkay. Variations in gesturing and speech by gestyle. *Int. J. Hum.-Comput. Stud.*, 62(2):211–229, 2005.
40. B. Hartmann, M. Mancini, S. Buisine, and C. Pelachaud. Implementing expressive gesture synthesis for embodied conversational agents. In *Gesture Workshop*. Springer, 2005.

41. C. Pelachaud. *Studies on Gesture Expressivity for a Virtual Agent*, volume 63:1. Springer, 2009.
42. Hannes Högni Vilhjálmsson, N. Cantelmo, J. Cassell, N. Ech Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. van Welbergen, and R. J. van der Werf. The behavior markup language: Recent developments and challenges. In *Intelligent Virtual Agents, 7th International Conference, IVA 2007, Paris, France, September 17-19, 2007, Proceedings*, pages 99–111, 2007.
43. S. Kopp, B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. R. Thórisson, and H. H. Vilhjálmsson. Towards a common framework for multimodal generation: The behavior markup language. In *Intelligent Virtual Agents, 6th International Conference, IVA 2006, Marina Del Rey, CA, USA, August 21-23, 2006, Proceedings*, pages 205–217, 2006.
44. D. Tolani, A. Goswami, and N.I. Badler. Real-time inverse kinematics techniques for anthropomorphic limbs. *Graphical Models*, 62(5):353–388, 2000.
45. S. Kopp and I. Wachsmuth. Synthesizing multimodal utterances for conversational agents. *Journal of Visualization and Computer Animation*, 15(1):39–52, 2004.
46. A. Heloir and M. Kipp. Real-time animation of interactive agents: Specification and realization. *Applied Artificial Intelligence*, 24(6):510–529, 2010.
47. M. Gillies, X. Pan, and M. Slater. Piavca: A framework for heterogeneous interactions with virtual characters. In *Intelligent Virtual Agents, 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008. Proceedings*, pages 494–495, 2008.
48. L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. *ACM Transactions on Graphics*, 21(3):473–482, 2002.
49. M. Brand and A. Hertzmann. Style machines. In *ACM SIGGRAPH 2000*, pages 183–192, 2000.
50. A. Hertzmann. Machine learning for computer graphics: A manifesto and tutorial. In *Computer Graphics and Applications, 2003. Proceedings. 11th Pacific Conference on*, pages 22–36. IEEE, 2003.
51. K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popović. Style-based inverse kinematics. *ACM Transactions on Graphics*, 23(3):522–531, 2004.
52. E. Hsu, K. Pulli, and J. Popović. Style translation for human motion. In *ACM Transactions on Graphics (TOG)*, volume 24:3, pages 1082–1089. ACM, 2005.
53. O. Arikan, D. A. Forsyth, and J. F. O'Brien. Motion synthesis from annotations. *ACM Transactions on Graphics*, 22(3):402–08, July 2003.
54. M. Müller, Andreas B., and H.P. Seidel. Efficient and robust annotation of motion capture data. In *Proceedings of the ACM SIGGRAPHEurographics Symposium on Computer Animation*, pages 17–26, August 2009.
55. C. Rose, B. Bodenheimer, and M. F. Cohen. Verbs and adverbs: Multidimensional motion interpolation using radial basis functions. *IEEE Computer Graphics and Applications*, 18:32–40, 1998.
56. Matthew Stone, Doug DeCarlo, Insuk Oh, Christian Rodriguez, Adrian Stere, Alyssa Lees, and Chris Bregler. Speaking with hands: Creating animated conversational characters from recordings of human performance. *ACM Trans. Graph.*, 23(3):506–513, August 2004.
57. Sophie Jörg, Jessica K. Hodgins, and Alla Safonova. Data-driven finger motion synthesis for gesturing characters. *ACM Transactions on Graphics*, 31(6):189:1–189:7, 2012.
58. Alla Safonova and Jessica K. Hodgins. Construction and optimal search of interpolated motion graphs. In *ACM SIGGRAPH 2007 Papers*, SIGGRAPH '07, 2007.
59. J. Starck and A. Hilton. Surface capture for performance-based animation. *IEEE Computer Graphics and Applications*, 27(3):21–31, 2007.
60. S. Gibet, N. Courty, K. Duarte, and T. Le Naour. The signcom system for data-driven animation of interactive virtual signers : Methodology and evaluation. *Transactions on Interactive Intelligent Systems*, 1(1), 2011.
61. A. Héloir, N. Courty, S. Gibet, and F. Multon. Temporal alignment of communicative gesture sequences. *Computer Animation and Virtual Worlds*, 17:347–357, 2006.

62. A. Héloir and S. Gibet. A qualitative and quantitative characterisation of style in sign language gestures. In *Gesture in Human-Computer Interaction and Simulation, GW 2007, Lecture Notes in Artificial Intelligence, LNAI*, Lisboa, Portugal, 2009. Springer Verlag.
63. Ed Hooks. *Acting for animators.* Routledge, 2013.
64. L. Bishko. *Nonverbal Communication in Virtual Worlds: Understanding and Designing Expressive Characters*, chapter Our Emphatic Experience of Believable Characters. ETC Press, 2014.
65. J. Lecoq, J.G. Carasso, J.C. Lallias, and D. Bradby. *The Moving Body (Le Corps Poétique): Teaching Creative Theatre*. Methuen Drama Modern Plays. Bloomsbury Academic, 2009.
66. S. Murray and J. Keefe. *Physical Theatres: A Critical Introduction.* Taylor & Francis, 2007.