

Animating signing avatar using descriptive sign language

Nour Ben Yahia

High School of Sciences and Techniques of Tunis
Research Laboratory: Information Technology and
Communication & Electrical Engineering (Latice)
Tunis, Tunisia
benyahia.nour@gmail.com

Mohamed Jemni

High School of Sciences and Techniques of Tunis
Research Laboratory: Information Technology and
Communication & Electrical Engineering (Latice)
Tunis, Tunisia
mohamed.jemni@fst.rnu.tn

Abstract—this paper proposes an interactive manipulation of signing avatar (complex human's figures) by means of a descriptive sign language. The objective is to improve the human-computer communication, to generate precise signs for Deaf community, to control and animate the virtual character. This approach has been developed to manipulate robots through several analytic and numerical resolution methods. These methods have been applied to the virtual character for better monitoring and to ensure a realistic animation in real time.

Keywords—component; Sign language; Avatar; Gesture; Animation

I. INTRODUCTION

Promote people's access to public places, and cultural information is also taken into account persons with Disabilities. Some people with disabilities face integrating difficulties into daily life due to environmental accessibility, the shortcomings of activities adopted to their specific needs. With the development of technology, it's a whole range of assistive technologies that now adapt to different disabilities, offering them new perspectives of integration [1], [2]. Thus, the disabled find benefit in different fields such as education, rehabilitation and appropriate vocational training. As well, accessibility appears to be an essential component of any access to information and allows those who are marginalized by their disability, to access to new technologies. Also, it is desirable that the websites, software respect accessibility criteria and incorporate specific devices such as sign language (SL) sequences, detailed images and animations simple to use. It should be noted that the SL would be that the deaf Braille for the blind that is a way for the disability community to ensure regular communication. This language [3] would have been designed recently, as a substitute for the word bypassing the ear canal in favor of a visual channel. It avoids the isolation in which the deaf can be dipped if they have no means of communication. To facilitate access to information for everyone, new methods to improve the dialogue between man and machine are required. The development of virtual characters able to generate postures sign language is a

response to this request. The information technology and communication is investing our daily space following a rapid growth. The emergence of new services resulting from these technologies facilitates consultation and the production of information for their users. However, such technological advances cannot be seen as a real step forward for a company if they continued to exclude part of the population. An individual may feel lack of access to information for various reasons, including economic, cultural, linguistic, or physical. Research on avatars and automatic translation systems in sign language have created various prototypes of systems to translate and provide statements varying according to the sign language through the channel of visual-gestural communication preferred of deaf. Recent research in the field of animation has allowed the use of virtual characters. The signers are virtual characters in three-dimensional and can translate messages to deaf by using sign language. However, the generation of animations for these characters depends on the description of signs, the categorization of gestures. The signs described by these systems are typically perfect but geometric realizations leading to unnatural and robotic movements from the signer.

In this paper, we present a new approach for animation avatar using an animation engine. This system permits to animate automatically different body parts of an avatar in order to improve realistic animation of virtual character gesture and allow deaf person to visualize realistic gestures. Our approach is based on a thorough study of sign language and gestures classification [4]. We have identified the parameters that need to be used for automatic generation postures sign language i.e. symmetry relations, that facilitate the specification of movement of the non-dominant hand knowledge of those dominant articulator and various types of repetition in sign. In addition, sign language requires contacts between different joints. Detection and avoidance of collisions between body parts of the avatar must be mastered to allow the contacts and to prevent unacceptable movement. To further facilitate editing, a graphical interface has been developed. It allows the selection of the sign components with

a direct view of the choice made on the avatar postures. The challenge of this project is to find the tradeoff between computation time and realistic representation. Indeed it must be closer to the maximum generation of real-time signs.

II. GESTURE AND SIGN LANGUAGE

A. Categorization of gestures

The notion of gesture is not precisely defined. It varies depending on the area of study, depending on whether one takes a sociological point of view, cognitive, biological ... Due to the richness of gesture modality and the variety of forms covered, it is not easy to establish a taxonomy, and there is no consensus on the matter. Gesture is defined by the field that studies it and refers to movements that accompany speech. The most common body parts used are the hands, fingers, arms, head, face, eyes, eyebrows, and trunk. Gesture does not include movement that doesn't accompany speech and doesn't include pantomime, or emblematic gestures such as the "OK" sign. Definitions for all these types of movements were provided by Kendon, who prefers the term gesticulation to the term gesture [5].

Gesticulation → Emblems → Pantomime → Sign Language

Emblems (symbolic gestures) are completing words or gestures used instead. They are used if they have a fixed meaning and commonly accepted. Frowning and is widely used to express doubt, disagreement or difficulty in understanding. In contrast, the representation of numbers with the fingers differs across cultures. The pantomime is to create an image of an object gestures, an event, and a state of mind. Its purpose is to make clear to the viewer signified by its form, it is highly iconic. Sign languages are the only ones to be considered as true languages. They have indeed some properties of decomposition, formation rules to generate new lexical items. Many different sign languages exist in the world, and even within each country. Sign languages differ from each other. Certainly, there are many similarities between certain because of their common roots. Deaf communicate much more easily than hearing by complementing their own systems of signs with facial expressions. In addition, there are in a country many differences between sign languages, according to multiple criteria: regional variations, social, ethnic, depending on the age or sex, and this differences are at all levels of the language (phonological, lexical or grammatical). McNeill [4] claims the legacy of Kendon and takes the idea of a continuum. He adds a classificatory co-verbal gestures, taking greater account of the signs of LS. He wishes not establish borders between co-verbal gestures and sign language. Later, his colleagues and himself refine the idea of continuum Kendon introducing a break between the first two elements of the continuum [5] (between gestures and gesticulations), which do not affect the integrity of the continuum. McNeill says that the actions of hearing co-verbal follow general principles but can be changed. They are closely related to the facial expression of the speaker. He insists that the actions are not fixed, they reflect imaging of the thought. Gestures are an important part of language, they can extend the usual definition of language. It uses a picture to

illustrate his point. Gesturing discover that language is not only a linear progression of segments, sounds and words but it is also instantly, nonlinear and imaged. McNeill then gives a brief chronology of research on gesture, following the work of Kendon. The modern theory designs the gesture and speech as part of a coherent whole proceeding from thinking. Kendon even defines what can be translated in French as a unit resulting from a combination of two components, verbal and the sign. Therefore, the gesture is not only economic but also communicative act, and its deletion or non-perception leading to degradation of the message understanding. The use of gesture depends greatly on the type of information to be transmitted, the degree of content iconicity (Iconicity is expressed in the grammatical structure of sign languages called classifiers. These are used to give descriptive information about a subject or verb. In American Sign Language ASL a grammatical marker denoting "intensity" is characterized by a movement pattern with two parts: an initial pause, followed by a quick completion. When this pattern is added to the adjective "good" the resulting meaning is "very-good". The ASL marker for intensity is iconic in that the intended meaning building of pressure, a sudden release is matched by the articulatory form).

B. Sign Language

Sign Language is the means of expression used by the communities of deaf and hearing person to communicate. Sign Language is a real language with a lexicon, syntax ...and is the most advanced form of gestural communication. The message in these languages is communicated through the gesture channel rather than oral sign. The expression of SL sentences cannot be reduced to only gestures produced by the two hands as the whole body can be involved to express a sentence. The signer uses the space around it to make signs and to place different elements of speech and refer to it. To better describe signs, it is essential to integrate this concept in automated systems that analyze these languages. Currently, most signs capture systems have been studied in the context of machine translation and make use of specific features of motion capture. For our part, our objective is to access this information through the design of a system that automatically generates gesture for sign language.

1) *Description of sign:* The first representation of a language is writing. If it is a relatively recent phenomenon across languages, it plays a starring role in the transmission of information for deaf person [6]. It allows individuals to communicate and is a factor that promotes the learning of the language. It is recognized by linguists that each language signs are signs consisting of several formation parameters. In 1960 in the context of structural research of LS, Stokoe [7] tries to put forward the character articulated signs [8]. It offers classes of "cheremes" required for the formation of signs. This analysis was joined by researchers as Klima and Bellugi [3] refusing to recognize any motivation iconic parameters while acknowledging that a global character could be included iconicity [9], [10]. The problem of the composition of signs is currently covered in a broader perspective syntactic and semantic where it is to describe every parameter according to

specific functionality [11]. Are retained as parameters description and training of signs:

- Manual configuration,
- The orientation of the hand,
- The location where the sign is made,
- The movement made by hand,
- The facial expression accompanying the realization of the sign.

2) *Sign transcription systems*: Three main approaches exist in the literature: the first is based on writing or drawing symbols, the second approach is based on video [12], [13], [14] and the third is based on 3D sequences and the animation of a virtual person according to a standard. Drawings were the first transcriptions of SL and the means generally used to replace writing long time ago. Later on, several transcription systems appeared such as HamNoSys (Hamburg Notation System) and SignWriting in spite of the difficulty to encode SL in a linear way. The video based systems consist in the insertion of a video sequence of human SL interpreter in the original video tape. Much more sophisticated tools exist nowadays on the market outclassing the predecessors in term of debit and quality. Some of current projects are based on video. In this new technological context, the modeling of a virtual character can be achieved either according to a segmented model, gotten by a hierarchical graph of the anatomical 3D segments, or according to a model seamless. In both cases, the surfaces are represented either with a polygonal stitch, or by a mathematical analysis, otherwise according to implicit functions of the skeleton. The creation of a virtual character is achieved either by a modeler of geometric primitives, or with the help of a 3D scanner. The approach of segmented virtual character exists in the H/Anim specifications, as well as in MPEG-4 FBA (Face & Body Animation), whereas the representation by virtual character is processed in MPEG-4 BBA (Bone-Based Animation).

3) *Descriptive languages*: Avatar animation is no longer a mere research topic. Many efforts in this field, combined with standardization such as MPEG-4, have lead to a wide use of faces animations and bodies' gestures for interactive applications. In order to describe the gestures and facial expressions made by virtual characters we use a descriptive language. This language is based on 4 criteria: facial animation, corporal animation, text-to-speech production and emotional representation. Therefore the real-time animation of the avatar can be achieved with a descriptive languages. These languages must be complete or constructed in a way that is easy to expand them and should aim to be as simple as possible and exclude any ambiguous features. That would keep the language fairly small and comprehensive. Nevertheless, this should not affect the previous criterion. In order to fulfil this criterion, elements that have the same

functionality should be merged. In addition the language must be consistent in order to make it easier for the user to learn, i.e. the syntax should follow a certain pattern. For example, the element names should be in the same form and have the same kind of attributes. Moreover, the markup language should aim to be intuitive, thus, the user will not always need to consult the specification to be able to use the language which uses a high abstraction level. That will make it easier to understand and thus to use. Many descriptive languages had been developed by researchers to describe the gestures and facial expressions made by virtual character. In 2005, Puy and al [15] have made a classification of thirteen different languages to determine the language that meets more the criteria "TABLE I". Two more languages have been added to this classification: SIGML and SWML which are dedicated to the representation of sign languages. The following table presents a comparison based on the main features required by animation engine: facial animation, corporal animation, text-to-speech production and emotional representation.

TABLE I. ANALYSIS OF THE MARK-UP LANGUAGES

Languages	Emotions	Facials	Corporal	Dialogue	Project
HumanML	Yes	Yes	Yes	Yes	-
CML	Yes	Yes	Yes	-	SAFIRA
AML	Yes	Yes	Yes	-	SoNG
APML	Yes	Yes	No	-	Greta
MPML	Yes	Yes	No	-	Mobile
MURML	-	Yes	Yes	No	Max
RRL	Yes	Yes	Yes	Yes	NECA
AIML	-	-	-	Yes	ALICE
VHML	Yes	Yes	Yes	Yes	InterFac MetaFac
XSTEP	No	No	Yes	-	-
SigML	No	Yes	Yes	No	eSign
SWML	No	Yes	Yes	No	vSign
SML	Yes	Yes	Yes	No	WebSign

III. WEBSIGN

WebSign [16], [17], [18], [19] is a Web application based on the technology of avatar (animation in virtual world) created in our laboratory. The input of the system is a text. The output is a real-time and on-line interpretation of the output into a sign language. This interpretation is constructed thanks to a dictionary of words and signs. The creation of this dictionary can be made in an incremental way by users who propose signs corresponding to words using the system described in this paper. A word and its corresponding sign are added effectively to the dictionary only after its assessment by an expert who supervises the system. For the linguistic treatment we have used a service which ensures the segmentation of a sentence and returns the set of segmented words with their grammatical category. This information is important to reduce the error of

translation of sentences. To represent and store signs, Websign uses SML (Sign Modeling Language). SML is a descriptive language defined to support a collaborative approach to write signs. In Websign, we have used SML to store description of signs which can be generated by a virtual human, or an avatar. We haven't use an existing system of transcription of sign language but we have opted to the creation of new language in which we can animate any avatar for any use. This animation can be used by different systems like chat, animation of robot or virtual characters or non-signer avatars, not necessary by systems designed for sign language based on forward kinematics. The animation on SML is a set of movement or rotation of groups of joints grouped depending to their animation on the time. Each movement or rotation has a fixed time interval, during which the rotation of every join of the group is done. Let's know that the armature description respect the H|Anim specification, in which each join have a specific name and specific initial orientation. The root element of SML is the <sentence> which represent a sentence in sign language, each <sentence> element contain one or many elements <word>. Every <word> element contains one <duration> element and one or many <join> element. The <join> element contains one <rotation> element.

```

<Sentence>
  <word>
    <mouvement>
      <duration>0.5</duration>
      <join name="l_elbow">
        <rotation type="euler">
          <heading>80</heading>
          <attitude>50</attitude>
          <bank>50</bank>
        </rotation>
      </join>
    </mouvement>
  </word>
</Sentence>

```

Figure 1. Rotation of l_elbow join using SML

IV. PROBLEMATIC

The lack of realism in the automatic generation comes with the use of a geometric approach to synthesis [20]. Add to this, this synthesis is not driven by a recording at the origin (such as the motion capture). So it's very possible, that virtual character achieve movements that a human could not do. In addition, due to the use of an automatic system, the interpolation is subject to artifacts, the most common of which is the collision of parts of the body of the virtual signer. With the absence of anatomical model, the virtual character, is not subject to the laws of physics. It therefore has the freedom to cross its members, hands or any part of the body during movement. Obviously, this kind of behavior is to be avoided if we consider a generation natural and realistic [21].

V. OUR CONTRIBUTION

Our objective is to develop an animation engine for animated virtual characters [22], and in particular for the generation of precise postures sign language. For the automatic creation of signs, WebSign use simple and intuitive interface to generate

sign. This application uses an interface that allows users to create their own descriptions of signs and participate in the development of their community's dictionary. The animation engine used is based on the technique of key frames in order to select the important moments assumed (the key moment) for training movement. But this technique raises difficulties especially in terms of time devoted to the creation of postures and lack of precision in the generation of gestures "(1)".

$$x = f(\phi) \quad (1)$$

Where x represents the final position of each joint and ϕ represents the rotation value of different joints. One of the main problem of the forward kinematics of the humanoid is the adjustment to achieve the desired position. Therefore, the first step is to rotate the elbow of the virtual character then the shoulder and wrist. This leads a multitude of solutions to generate the required posture. So the user must intervene to choose the best solution that meets their needs. This operation is costly in terms of time and presents several challenges to create a fluid motion and understandable by the deaf [23].

Now if we want to synthesize the motion of a humanoid, we are immediately faced with several problems. It is difficult to model the human in its entirety. It is therefore necessary to find a more simplified method to manipulate complex models. However, the animation engine must be enhanced to allow users with or without disabilities to capture gestures in an easy way through the technique of inverse kinematics "(2)". The best example of the use of inverse kinematics is the arm. To animate an arm by inverse kinematics, we just move the wrist. The animation engine will load automatically rotate of each joint of avatar.

$$\phi = f^{-1}(x) \quad (2)$$

Where x represents the target position of each joint and ϕ represents the rotation value of different joints calculated to achieve the desired position.

Within the set of admissible postures, the user is free to manipulate the avatar rather than specifying the value of each individual degree of freedom [24]. The Inverse Kinematics method automatically computes these values in order to satisfy a given task usually expressed in Cartesian space [25]. These techniques require the resolution of complex nonlinear equations and usually expressed as a constraint-satisfaction problem [26]. However, this is a laborious task because of the high number of degrees of freedom present in the model i.e. fifty for a human model without considering the fingers. The control of a complex figure like human model by means inverse kinematics requires that multiple tasks be simultaneously applied [27]. For example, a task may control the position of left hand and another control right hand. The balance of human model will be controlled by another task which provides some information concerning forces and mass distribution. However, this kind of problem is redundant because the number of tasks is lower than the number of degrees of freedom [28]. Besides, an infinite number of solutions may exist. Then, additional constraints must be specified in order to select the best solution that satisfies the

needs of users [29]. We are interested in the general problem of positioning a part of the human body in Cartesian space seeking a posture in the joint space [30]. We take into account, the limited travel of the joints referring to a biomechanical study in order to identify the degrees of freedom of each joint. The equations establishing the position and orientation of a segment of the body depending on the setting chosen for joints are highly non-linear [31]. Moreover, unlike the methods dedicated to real-time control, we seek to resolve this problem by involving the largest possible number of joints because we believe it can reveal more globally optimal solutions within the meaning of criteria.

VI. CONCLUSION AND FUTURE WORKS

Generation of Signs is conditioned by hands, head and face gesture, the orientation of the sign, its location and its movement. To take into account these conditions, our system deals with figures which have many degrees of freedom. We use the technique of inverse kinematics based on resolution of a nonlinear system to ensure precise movements, control and maintain the balance of characters.

Our future work aims to extend the animation engine to avoid collision between the various articulations of virtual character and add more constraints for more reality to the generation of signs. In view of the complexity of these constraints, we have to resolve these complex systems in real-time since our system will be integrated in the project Websign for automatic translation of text to sign language by the means of avatars.

REFERENCES

- [1] M. Boulares and M. Jemni, "Mobile sign language translation system for deaf community," in Proceedings of the International Cross-Disciplinary Conference on Web Accessibility. ACM, 2012, p. 37.
- [2] M. Jemni, O. Ghoul, N. Ben Yahia, and M. Boulares, "Sign language mms to make cell phones accessible to the deaf and hard-of-hearing community," in Conference and Workshop on Assistive Technology for People with Vision and Hearing Impairments (CVHI 2007), Granada, Spain (August 2007), 2007.
- [3] Klima, E., and Bellugi, U. "The signs of language". Harvard Univ Pr, 1979.
- [4] McNeill, D. "Hand and mind: What gestures reveal about thought". University of Chicago Press, 1996.
- [5] A. Kendon. "How gestures can become like words". Cross-cultural perspectives in nonverbal communication, pages 131-141, 1988.
- [6] S.K. Liddell. "Structures for representing handshape and local movement at the phonemic level". Theoretical issues in sign language research, vol. 1, pages 37-65, 1990.
- [7] Stokoe, W., Casterline, D., and Croneberg, C. "A dictionary of American Sign Language on linguistic principles". Linstok Press Silver Spring, MD, 1976.
- [8] S.K. Liddell et R.E. Johnson. "American sign language : The phonological base. Linguistics of American Sign Language", pages 267-306, 2000.
- [9] C. Cuxac. "Iconicité et mouvement des signes en langue des signes française". Le mouvement : des boucles sensorimotrices aux représentations langagières, pages 205- 218, 1997.
- [10] H. Poizner. "Visual and phonetic coding of movement : Evidence from American Sign Language". Science, vol. 212, no. 4495, pages 691-693, 1981.
- [11] Liddell, S. "Structures for representing hand shape and local movement at the phonemic level". Theoretical issues in sign language research, 37-65, 1990.
- [12] D.M. Gavrilu et L.S. Davis. "3-D model-based tracking of humans in action : a multi-view approach". In Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on, pages 73-80. IEEE, 1996.
- [13] K. Jaballah and M. Jemni, "Toward automatic sign language recognition from web3d based scenes," Computers Helping People with Special Needs, pp. 205-212, 2010.
- [14] M. Shantz et H. Poizner. "A computer program to synthesize American Sign Language". Behavior Research Methods, vol. 14, no. 5, pages 467-474, 1982.
- [15] M. PuyCarretero, D. Oyarzun, A. Ortiz, I. Aizpurua, J. Posada. "Virtual characters facial and body animation through the edition and interpretation of mark-up languages", Computers & Graphics pp.189-194, 2005.
- [16] M. Jemni and O. Elghoul, "An avatar based approach for automatic interpretation of text to sign language," in 9th European Conference for the Advancement of the Assistive Technologies in Europe, 2007, pp.3-5.
- [17] M. Jemni et O. Elghoul. "A system to make signs using collaborative approach". Computers Helping People with Special Needs, pages 670-677, 2008.
- [18] M. Jemni, O. Elghoul, "A web-based tool to create online courses for deaf pupils," International Conference on Interactive Mobile and Computer Aided Learning, Amman, 2007.
- [19] A. Othman and M. Jemni, "Statistical sign language machine translation: from english written text to american sign language gloss," arXiv preprint arXiv:1112.0168, 2011.
- [20] Lebourque, T., and Gibet, S. "Synthesis of hand-arm gestures". In Proceedings of Gesture Workshop on Progress in Gestural Interaction, Springer-Verlag, pp. 217-225, 1996.
- [21] N. Klopčar, M. Tomšič, J. Lenarčič. "A kinematic model of the shoulder complex to evaluate the arm-reachable workspace". Journal of biomechanics, vol. 40, no. 1, pages 86-91, 2007.
- [22] Ben Yahia N., Jemni M. "Toward a descriptive language for signs", Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments (CVHI), Wroclaw, Poland, 2009.
- [23] L. Unzueta, M. Peinado, R. Boulic et Á. Suescun. "Full-body performance animation with Sequential Inverse Kinematics". Graphical Models, vol. 70, no. 5, pages 87-104, 2008.
- [24] L.C.T. Wang et C.C. Chen. "A combined optimization method for solving the inverse kinematics problems of mechanical manipulators. Robotics and Automation", IEEE Transactions on, vol. 7, no. 4, pages 489-499, 1991.
- [25] CG Broyden. "Quasi-Newton methods and their application to function minimization". Math. Comput, vol. 21, no. 99, pages 368-381, 1967.
- [26] S.R.Buss. "Introduction to inverse kinematics with jacobian transpose, pseudoinverse and damped least squares methods". University of California, San Diego, Typeset manuscript, available from <http://math.ucsd.edu/sbuss/ResearchWeb>, 2004.
- [27] M. Meredith et S. Maddock. "Real-time inverse kinematics : The return of the Jacobian". Technical Report No. CS-04-06, Department of Computer Science, University of Sheffield, 2004.
- [28] C. Welman. "Inverse kinematics and geometric constraints for articulated figure manipulation". PhD thesis, Simon Fraser University, 1993.
- [29] W.A. Wolovich et H. Elliott. "A computational technique for inverse kinematics". In Decision and Control, 1984. The 23rd IEEE Conference on, volume 23, pages 1359-1363. IEEE, 1984.
- [30] L. Y. Chang et N.S. Pollard. "Method for determining kinematic parameters of the in vivo thumb carpometacarpal joint". Biomedical Engineering, IEEE Transactions on, vol. 55, pages 1897-1906, 2008.
- [31] D.G. Luenberger et Y. Ye. "Linear and nonlinear programming", volume 116. Springer Verlag, 2008.