Expressive Sign Language System for Deaf Kids with MPEG-4 Approach of Virtual Human Character

Itimad Raheem Ali[1,*] , Hoshang Kolivand[2]

[1]Faculty of Business Informatics, University of Information Technology and Communications.

itimadra@uoitc.edu.iq

[2]Department of Computer Science, Liverpool Jone Moores University, Liverpool, United Kingdom

h.kolivand@ljmu.ac.uk

Corresponding author: Itimad Raheem Ali, Faculty of Business Informatic, University of Information Technology and Communications, itimadra@uoitc.edu.iq

Abstract

Children with language impairments during their significant developmental periods within childhood are exposed to cognitive risk, social impairments, along with language. This is difficult with children born deaf from hearing parents who own little or no experience of communicating in sign language. This system presents the sign language in the context of British Sign Language (BSL) for producing utterances through virtual characters. In capturing, Kinect sensors use a motion capture sensor for motion actors. The connection uses sensors to read data, connect to high-quality 3D scans, and then use these high-quality scans of the animated MPEG-4 face and hand models. The main challenges of this system are the simultaneous capture of data for the whole hand and the development of the MPEG-4 approach considering the animation engines with descriptive sign language features. After synchronizing motion data from motion capture results with Kinect, the combined hand character adjusts points, frames, and time with virtual characters based on the motion of character actors. This study demonstrates the skills of this sign language system instrumental in presenting an assessment by users, highlighting the importance of the hand part in creating new accents and signs in BSL. We have validated this system by testing the reliability and functionality of the virtual characters.

Keywords: Expressive, Hand animation, Sign language system, MPEG-4 approach, Kinect.

1. Introduction

In the present era, computer animation along with the sign languages addresses the needs of linguistics communities. They collectively develop signed virtual characters ensuring real sign languages-based communication. On the other hand, signers or signing virtual avatars are considered to be critical in using sign languages to access information (Félix Bigand et al. 2019), this system mainly deals with this research category. Like all signed virtual character projects, this system focuses on truly human interactive modalities. More specifically, the system communicates in British Sign Language (BSL) and attempts to create interactions among deaf children and virtual characters by initiating real-time conversations. They can be subjected to signing by the contextual human user facing a camera that recognizes the signature. It is through providing a virtual character that culturally along with linguistically accepts responses according to the respective behavioral sign (Brian Scassellati et al. 2018, Félix Bigand et al. 2019).

This study will introduce a process of sign generation a part and parcel of an interactive system. In which its main purpose is to produce a realistic, yet highly expressive, BSL along with pronunciation data. On capture, the results of this capture are then converted to motion data. After synchronizing motion data from Motion Capture results with Kintech, the joint hand character points, images, and timing are adjusted with the virtual character based on the flow of character actors. Most of the signing characters are instrumental in projecting and adopting a synthetic animation technique in context to their virtual character systems. Going further this can have the capacity to convince audiences march toward transcend reality. In this work, we draw MPEG-4 animation techniques, such as capturing motion to record human movements. It is believed by us that all these virtual signers are data-driven. Within few days they will be experiencing a smoother operation than synthetic pieces resulting distinct signature style. Similarly, virtual signers animate expressive interactive contexts that are considered to be tedious. Most of the deaf people are critical in regards to their language misuse hence requires a detailed system's evaluation associated with output.

The study contributes toward nursery rhymes use in British Sign Language (BSL) with their specific phonetic-syllabic rhythmic virtual characters linguistic to match children's biological stimuli. This is associated with sensitivities in the course of this particular period associated with development. In original MPEG-4 the animation system is dedicated to linguistic interaction amongst human as well as virtual characters using BSL. This system in acceptable modes uses animation community along with key modes like our regression strategy for restoring speed or our streaming architecture for facial animation. This system evaluated and proved that the system acceptability depends upon technical preferences limitations.

The research is organized below. We described the historical, as well as technical background in context to virtual signature creation that discusses MPEG-4 models versus methodological models and describes our sessions in Section 2, Section 3, giving an overview of data motion capture sessions, this motion captured is stored in databases of motion for later retrieval, as well as talk about the design as well as BSL's annotation signed virtual with multichannel composition to rendering the signing avatar. In the fourth Section, we evaluate the results in detail as our animation modules system, handling simultaneous hand based on corporal animation, discussion and conclusion in Section 5.

2. Motivation

Since deaf children do not usually have access to spoken words, sign language is their native language, fully accessible through their visual modality. Moreover, by the nature living in the hearing world, deaf children acquire knowledge of the world by reading the bilingual signs hearing compatriots. Nevertheless, members of the same deaf community differ in fluency in spoken / written language and may find it difficult to rely solely on written text or captions (Wheatland N. et al. 2015).

Thus, new interactive sign language communication systems have been subjected in present years to progress the convenience of the deaf in an assortment of media. Recently linguistically-native tend to centers around assistive technologies. It refers to signed language generation using virtual character. Motion data's synchronization derived from motion capture by connecting with Kinect sensor resulting to the adjustment. At the joint real hand character with virtual hand characters according to the flow of character actors. The joints of the hand are the important hence

the components tend to comprise of hand's basic structure. For bones as well as the naming conventions that joints are derived from anatomical systems as shown in Figure 1.

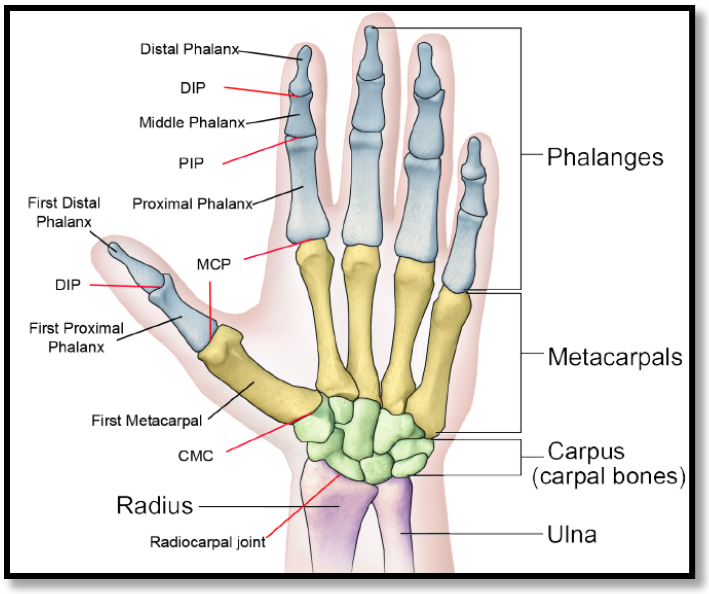

Figure 1. The forearm, wrist, as well as hand's bone's structure by Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436. - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=29987031. Acronyms: CMC – Carpometacarpal joint, MCP – Metacarpophalangealjoint, PIP – Proximal interphalangeal joint,DIP – Distal interphalangeal joint

Here we use some of the techniques used to animate interactive kids and virtual communication players, distinguishing descriptive methods from the MPEG-4 animation method which includes a full MPEG-4 method to animate the character.

2.1 Kids and Technology

Previous works have explored the educational and social activities of present technology for children screen based and physically. Anderson D. R. and Pempek, T. A. (2005) reported that children could learn less from similar real-life experiences by appearing on television based on or video program similar to the experiences of real-life. This "video deficit effect" extensively reproduces and shown to be involved in a variety of linguistic developments, together with phone phoneme dissimilarity individual word mapping (Patricia K Kuhl et al. 2003, Marina Krcmar 2011), as well as lexical category retention (Clare E Sims and Eliana Colunga 2013). However, (Judy S DeLoache et al. 2010), observed, children (over 30 years of age) learn new words when a person seems to be sitting in a social interaction among two subjects while watching a video. It addresses directly to the child's purpose by emphasizing the significant role of social interaction in children's language.

2.2 Descriptive Methods for Sign Language Virtual Character

In context to the past decade, awareness in computer applications, especially in sign language has grown significantly. These applications subject sign language dictionaries such as Auslan Dictionary (Vesel J. 2005), learning

tools like the SYNENOESSE project in Greece and WebSign (Jemni M. and O. ElGhoul 2008) that can be customized to create 3D signed content using characters. Another conversation character system on MMSSign (Jemni M. et al. 2007) has been subjected to assist Tunisian deaf to communicate on their cell phones so that they can exchange 3D signed sequences. Combined Conversation Agents (ECAs) provide animations, psychology, including artificial intelligence, linguistics and computer animation, is described at the behavioral planning stage of the building, and is built by means of animation engines. The Behavior Expression Animation Toolkit or also collectively known as BEAT. It's the 1st systems instrumental in explaining virtual character's preferred behaviors using textual input. It is done by combining gesture-based features meant for generation as well as refers to speech synchronization mentions (Cassell et al., 2004). Description languages based on XML are designed to subject multimodal behaviors variety devoted to gesture based complex specifications (Kranstedt, A.et al. 2002).

Introduces parameter's set used in the classification of the expressive gestures (Hartmann B et al. 2005). Kopp et al. (2006) the integrated BML structure that defines different abstraction levels, interpreting planned behavior associated with multimodal edited behavior. This can put together dissimilar planning or plan's controllers the real-time system. In recent times, represents EMBR. It reflects a new controlling layer amongst the different behavioral level as well as at the level of procedural animation (Kipp M. et al. 2010). Gibet et al. 2012, developed a multi-channel animated signature language system in a signature French Sign Language (LSF) and an animation engine based on the data taken into account to create virtual character pronunciation by capturing whole body data including torso, hand and facial movements. Bigand et al. 2019 addressed the problem of gestural identity in the context of animated agents in French Sign Language (FSL). Nasihati et al. 2019 evaluated an AI system that uses avatar and robots to facilitate children's skills in learning aspects of the American Sign Language (ASL). Kaneko et al. 2010 commented by stating that to focus on TVML (Television Program Writing Language) the primary focus must be on the adaptation of the sign language-based subject's animation. A text-based computer language named TVML simply writes a script in TVML as well as produces the most important computer graphics (CG). Going further the authors presented a completely new approach used to recognize as well as helps in indexing a 3D signed contents. All these 3D contents are based on the recognition as well as helps in the classification of sign language in accordance to their relevant parameters. The contextual uses avail an adaptation associated with the Longest algorithm based common subsequence combines with Minkowski and are used to measure similarity (Jaballah K. and Jemni M. 2012). Scassellati et al. 2018 was instrumental in subjecting an integrated type of multi-agent system involving robotics. Sutupo et al. 2019 proposed to synchronize the proposed animated characters. Some of the revelations were passed down from generation to generation to their generation. Basically, the authors of this method want to translate any description of the gesture as a sequence of gesture commands in the formality mentioned above which can be interpreted directly by the animation engine in real time. Most of these problems use pure synthesis methods, for example (Tolani D. et al. 2000, Kopp S. and Wachsmuth I. 2004) use retrograde kinematic techniques using systematic animation methods with a high level of specification language. The main drawback of this method is the unreality of generating generic motion, especially for complex behaviors that require coordination of different parts of the body.

2.3 Virtual Character Signers

Over the past decade, 3D Virtual Signature Virtual characters have been designed to provide increased accessibility for the deaf on a variety of computer devices. In addition, these avatars have given rise to a variety of applications, including sign production, text-to-text, and sign synthesis assessment, which can aid in the study of signed linguistics (Escudeiro P. et al. 2017). To date, signing avatars has focused on creating sign phrases with phonetic or vowel descriptions of sign sequences using the animation techniques described above. Chiu Fu-Sheng 2007 talk about a sophisticated method of translating text from Chinese in the direction of Tai-wanese sign language by creating a sign video using bilingual corpus and sign data. In project ATLAS, an interpreter that virtually translates Italian to Italian sign language or the (LIS). The system pre-captures and sets the manual animation functionality to suit the context of their speech. Although some studies have shown that the use of depth sensors and the application of their image processing techniques have yielded optimistic consequences in smaller data and noise amounts (Lombardo V. et al. 2010), isn't scalable to word's amount planned for us not it's feasible for a real time image processing translator as a heavy hardware (Verma H. V. et al. 2013). As a replacement for it, we use the Kinect sensor in detecting body skeleton with Kinect first market launch, capturing the user and various joint coordinates and the general movement of the body in front of it. The only key element missing by hand configuration processing using data gloves and body motion analysis through Kinect was facial expression processing. Kinect is also effective for this part because it has already recognized some of the first face expressions to several libraries supporting functionality (Li, B. Y. et al. 2013). This last part of the project has not been implemented yet and only some research has been subjected.

2.4 MPEG-4 Methods for Embodied Communicative Characters

MPEG-4 methods are an interesting way to animate virtual avatars. Motion Data based, methods allowing you to change the purpose of input data. Therefore, the MPEG-4 animation method introduces most of the previous work editing and composition techniques, emphasizing the reuse of capture speeds to create new motion sequences. Many data-driven approaches have also focused on creating statistical models from speed data (Grochow, K. et al. 2004, Chai J. and Hodgins J. K. 2007). It can be noted that both the movement and the data captured by its underlying semantic content work in very few ways and almost nothing concerns the gestures of communication.

Added benefit, includes the subjects of motion capture based on data analytical material that analyzes language-specific features or parse. These will be considered as a generic feature associated with signed language. Similar to the dynamics or rules of the movements associated with the spatial-temporal form of relationships amongst different production channels (Buyun Sheng et al. 2019, Yang-Kun Ou, et al. 2020). In conclusion, the main challenge of the advanced MPEG-4 animation approach is i)to be able to synthesize and manage the various forms involved in communication gestures together with data from different sources ii) and in the case of data structures, manage data and related semantics properly. Our door is taking a significant step in this direction. We know that the proposed system is present, which motivates our animation system (Yao L. et al. 2018).

3. METHODOLOGY

The purpose of the British Sign Language system is to progress in real-time interaction's quality among human as well as virtual characters with mutual communication in the British sign language. The system considers three aspects of the subject: recognition of the symbols created by the user, a conversation that subjects an adaptive

response, as well as the synthesis associated with the right by means of a virtual signer's kind of response. The, tandem with the dialogue generator is capable of system recognition that gradually state their relations entities, canceling or confirming the transmitted messages raising necessary ambiguities. Dialogue process measured in real-time restriction applied in context to vocabulary limitation, allowing synchronization of lips. As used by (Ali I. R. et al. 2018) this paper synthesis and focuses on British sign system that is our fully MPEG-4 virtual signer. We present a multi-channel animation framework provided by various organizations that represent the parts of the body that synthesize information: torso, arms, hands, head, face and vision.

The British Sign Language system's functional organization has been represented in Figure 2. The system comprises 2 large building blocks: one, working offline, a dual-index database that captures movements, and the other, automatically recognizing signals, the animation of a virtual signer, and a go-among module that produces meaningful and appropriate dialog.

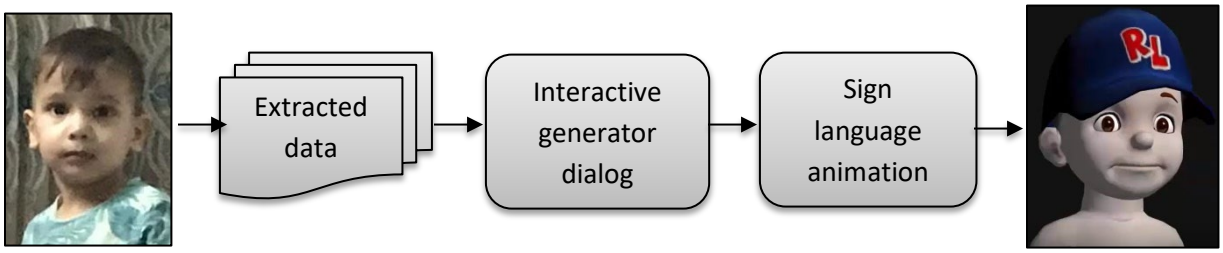

Figure 2. An overview of the proposed sign system.

In context to the aforesaid understanding of sign language in mind, we and other members of the system have created a sign specifically for registration for the lottery with signature avatars. Here we describe some of the challenges raised by BSL and the ways we choose to include such solid data for the proposed system test. These signatures form a range of BSL signaling cards for the proposed system if signatures are required to comprehend a set of grammatical and redemption processes.

Also, in order to synthesize sign language, numerous symbol repetitions are required in different contexts to the signed avatar. It allows reading new pronunciation compositions with different elements associated with varied places. In the middle of conducting multiple phonological instances based on same types of sign record. A computer animator opts for the best-fit sign and places the forcing of a single instance associated with the sign into an only one of its kind contexts. An overview of the system is given in Figure 3.

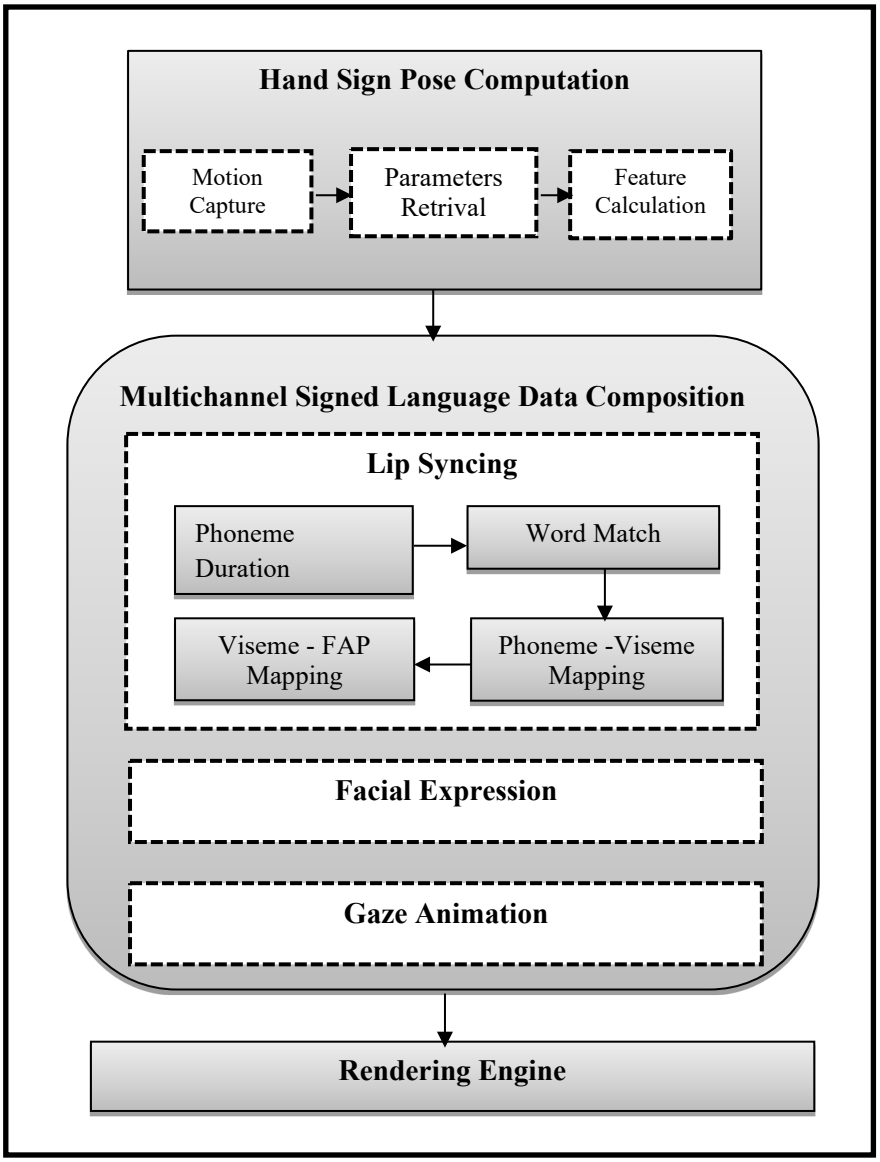

Figure 3. Overview of the sign system.

3.1 Hand Sign Pose Computation

Hand poses are associated with paired feature's distance between landmarks which are computed in real-time from the coordinate points of the hand poses according to the requirements of the MPEG-4 method.

3.1.1 Motion Capture

Understanding the signs requires the appropriateness of their structure. Especially in finger spelling, where each letter of the alphabet is named with a symbol, the degree of finger opening may be the only prominent reason among the letters. A few handshapes are separated only by the position of one finger or by the fact that it communicates with another part of the hand. It requires significant accuracy in speed capture and data animation processes. Much of our description focuses on hand setup and speed, also considering significant non-manual elements, such as shoulder

movements, head tremors, change vision, or facial imitations. For example, it can be used to remember a specific subject rather than signing in with the eye; it may also be necessary to understand a sign in this channel that we use (Ali I. R. et al. 2018).

The process of taking motion with markers must use tools on animators, unique clothing, and green background spaces so that movements performed by actors can be recorded and stored in MPEG4 file format, while without markers do not use tools, unique clothes and special rooms to make motion arrests on actors (Yang-Kun Ou et al. 2020). The author uses Kinect X-Box 360 motion capture which is a type of markerless device because the price is relatively low and does not have to use a special room for recording motion (Sutopo J. et al. 2019) as shown in Figure 4.

Figure 4. Kinect X-Box 360

In Figure 5, one of the motion capture devices from a well-known company, Microsoft, which issued a video game console named X-Box 360 or Kinect, is a Project Natal, using the Kinect sensor to read the bone structure in the human body.

Figure 5. Hands outfitted with a fairly comprehensive marker set for optical motion capture. Further joints could be added to capture the motion of additional joints

Analysis of the number of joints to avoid errors in reinforcement on animated characters because it will affect when synchronizing MEGE4 files to animated characters (Hartmann B. et al. 2005, Clare E Sims and Eliana Colunga 2013, Ali I. R. et al. 2018), to find out how many joints there are in the human body (Kranstedt A. et al. 2002, Anderson D. R. & Pempek T. A. 2005, Davani A. M. et al. 2018). Each joint will be named according to the sequence of joints that are in the anatomical structure of the human body (Blausen Medical staff 2014, Félix Bigand et al. 2019). This

paper is a development of researches where in previous studies it has not discussed the technicality of how character synchronization from motion capture data with animated characters is made so that the results of synchronization in the form of motion that is not real or smooth in accordance with the original motion of the actor. So, this study seeks to show the technical synchronization of signing avatar and motion capture results in real accordance with the motion of the actor (Brian Scassellati et al. 2018). Figure 6 shows Native signer poses with motion capture sensors on face and hands in a different pose.

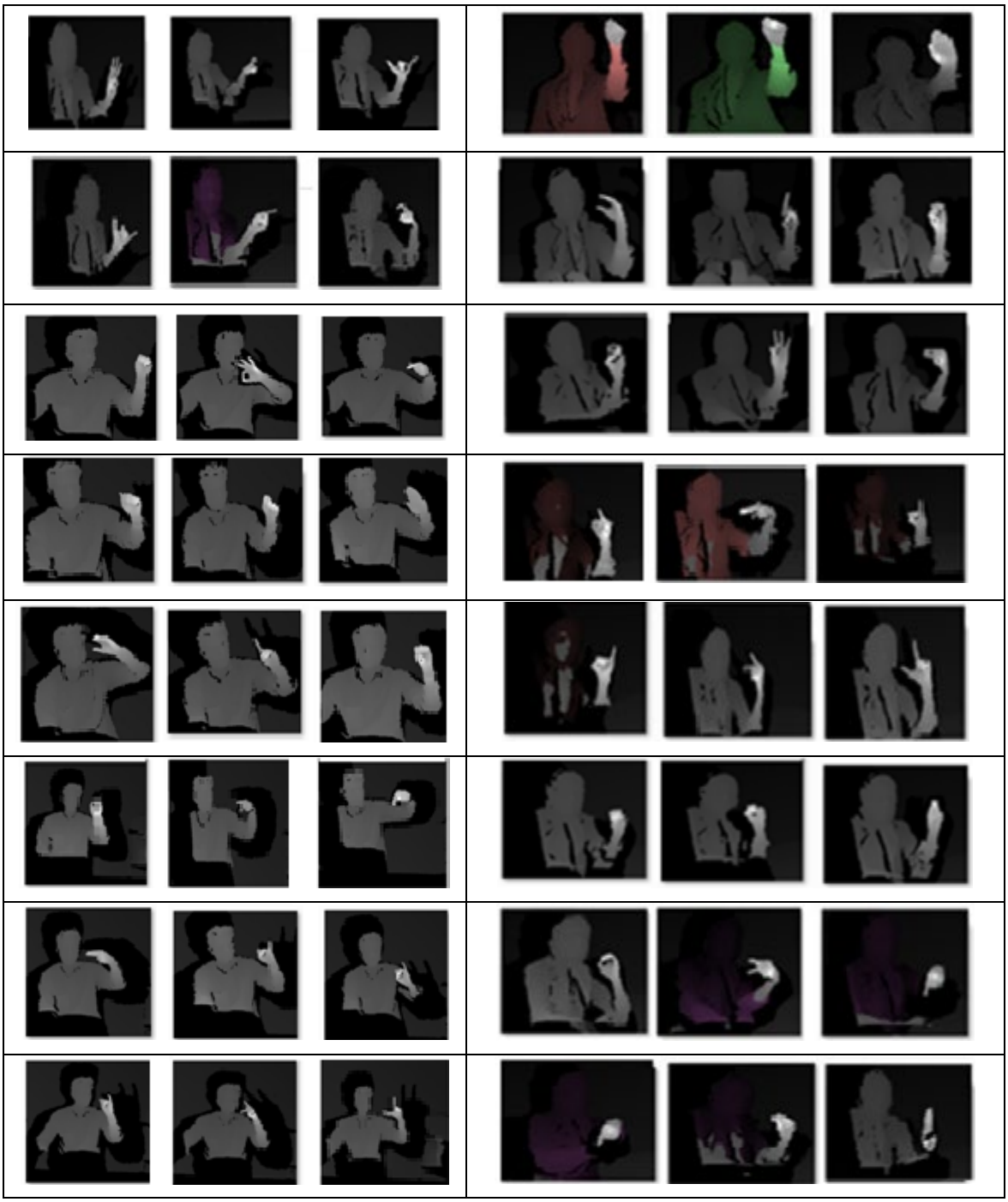

Figure 6. Native signer poses with motion capture sensors on face and hands in a different pose.

3.1.2 Parameters Retrieval

To replay a full animation and make it available for data analysis associated with motion capture, more than a few post-processing are required for the operations. Firstly, the dynamics of the finger are recreated through inverse kinematics, as the last finger's positions were recorded. Face transmission is done by capturing face motion's data through cross-mapping along with subjected blend shaped parameters (Ali I. R. et al. 2018). Thus, allowing us to complete face transmission directly from raw motion capturing of data just after mapping model learning.

To recover images from Kinect Recordings it is needed to acquire the tools provided by the Blender open source library (https://opensource.com/tags/blender). All the mentioned type of these tools subjects to getting frames images. The frames are converted to basic BMP format (Kinect frames are initially 640 x 480).

3.1.3 Feature Calculation

In Figure 6, Kinect has three cameras consisting of (A) RGB cameras with 640 x 480 resolution with 8-bit colours with 30 frames per second, Kinect RGB cameras can do face recognition and detect three colour components namely Red, Green, and Blue. (B) Camera Depth A sensor or depth sensor is an infrared projector, in the form of a moving frame that has joint points of human motion. The Dept. Sensor consists of a combination of an infrared laser projector that can take video data in 3D. (C) Is the result of recording the motion when the actor moves the hands, the camera recorded from monochrome the way the sensor works is almost the same as the Depth sensor, but on the monochrome sensor it works simultaneously to view areas or spaces in 3D without regard to light conditions (Vesel J. 2005). After that, the process retrieved from databases that contain raw data associated with motion capture. Then we build a new multichannel composition motion system expressed in sequence to the postures of the skeletal hand. Containing body and hand configurations encoding information that are associated with facial markers, as shown in Figure 7.

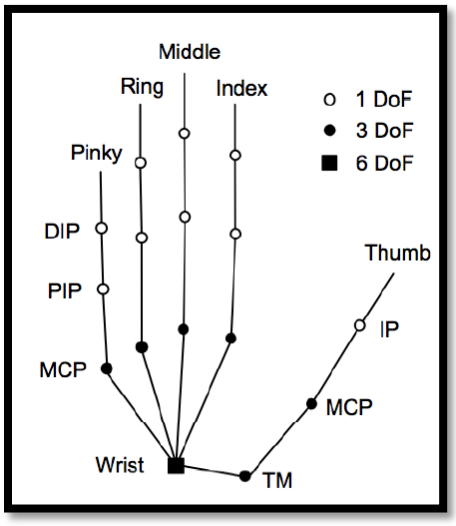

Figure 7. Hand model showing freedom of certain degrees at the individual joints (Wheatland N. et al. 2015).

The estimation of hand poses associates with paired feature's distance. Such distances among landmarks needs computation derived from the coordinates with real world. It is therefore, is based on initial conversion fora given depth frame's pixel p(x, y) associated with the coordinates of the real-world:

$$X = \frac{x-c_x}{f_x} \times (Z - f_{ir}), Y = \frac{y-c_y}{f_y} \times (Z - f_{ir}) \quad (1)$$

In which Z = p (x, y), c_x and c_y are the optical center position of the Kinect, f_x and $f_y$are the focal lengths along with X as well as the Y axis associated with the Kinect. Moreover, $f_{ir}$is infrared depth's focal length associated with Kinect. Taking into assumption that initial frame refers to a front hand image, (the coordinate's origin of location in the face's middle in context to initial posing), again by using coordinates of real-world (Equation (1)).

Equivalent to MPEG-4, the obtained values of the parameters are directed from the motion capture presented by Kinect. The matrix include all the points against time row along with Z point coordinating column. As a result, expressed as (F) parameter matrix is as presented in Equation (2).

$$p = Z + M.U.(fp - fp_0) \quad (2)$$

In which ,fp_0 refers to fp.'s rest point. M as a matrix is used to map fp(X column's) to FAPs (F's columns), as well as diagonal matrix U encloses correct scale. of fp

We should note that the feature points in our process were subjected in contradiction to manual 29 developed all 3D hand alignment and dense warping feature points. Even though the use associated with short-characterized points slightly reduces the performance of alignment as well as coarse warping, it still subjects important presentation to facilitate less computational complexity.

3.2 Multichannel Signed Language Data Composition

This compositional process is instrumental in serving as an inspirational animation system, although phonetic specifications are not required as imagined for some of our general linguistic goals. Similarly, we encode the ways multiple signer's hand or channel's parts of time parallelly described specifies channel's motion with corresponding linguistic handshape, including its location, movements, orientations, facial expression, as well as directional gazing categories demonstrated in Figure 8.

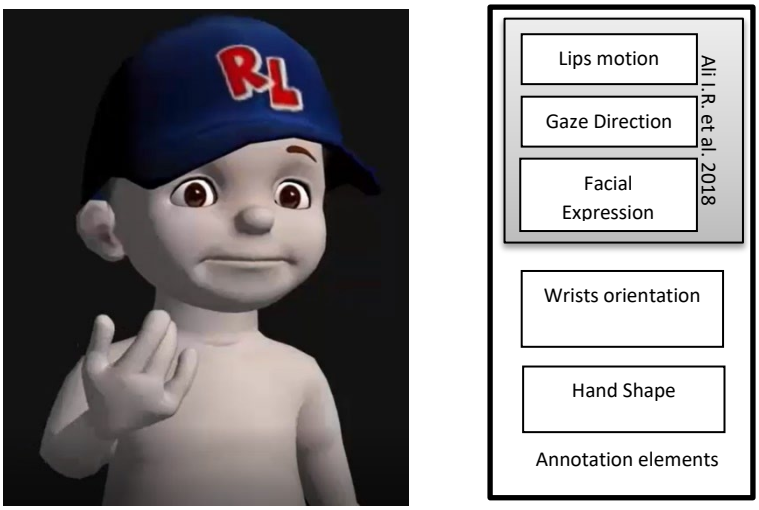

Figure 8. The manipulated channels are signed languages mechanics inspired proposed by the linguists of British Signed Language.

The goal determines static featured video sequencing points combining each image's difference from the original image is assumed to be consistent with the background of all the feature points which have an engraved difference value near zero. Thus, the median of all related attribute points gives a summary of the animation. Aftermath, we delete all the points that add the present value to this mean value. Due to the noise of the sensor orb these outliers are very bright pixels are large enough for outliers to exclude assemblies using gray quality thresholds.

3.3 Rendering

The example videos under the link "https://www.youtube.com/watch?v=deLCiBTRfdc" showed proposed method performance calculating animated model MPEG-4. This video shows the process of speech from text through the use of a model on animation output associated with this study. Another video in this article shows the manual tracking process with expression change using the MPEG-4 mathematical animated model. These videos clearly show that real, animated MPEG-4 face models can be calculated using data collected with inexpensive 3D scanners like Kinect.

Where the front part forms the front shape of the characters like hand length, head and eyes and face, while the side part forms a character with a combination of hand, chest size. The results obtained by the analysis carried out in the previous section, the author can make 3D animation by utilizing motion capture and synchronization with (Ali I.R. et al. 2018) in making animations so that the results desired by the author will describe the results obtained by the previous analysis as shown in Figure 9.

Figure 9. The results of proposed British Signed Language system.

Character creation has several types of software for image representation (blueprint) into 3D animated characters, namely, Blender, 3ds Max, AutoCAD, and others (Brian Scassellati et al. 2018, Yao L. et al. 2018). In this study the proposed system uses Blender software to create animations because the open source and specifications for running blenders are relatively light compared to other software, blending effortlessly in creating 3D characters using the prepared blueprint (Vesel J. 2005, Davani A. M. et al. 2018). Characters created are not always the same as the blueprint that exists because what is described can be different when made in a blender.

Algorithm 1. The ways a blender controls Algorithmic blend

```
for all hand feature points do
        set hand feature points weights to 1 and points transformations to Identity
end for
while feature points&& a points weight > 0 do
 if time ⊂ [startTime - fadeIn, endTime + fadeOut] then
        find weight w wrt. time and compute pointshand
        for j = 0 to hand.pointsCount do
            k = feature points index in points transformations
           if feature points weights[k] > 0 then
             if feature points weights[k]! = 1 then
               points transformations[k] = interpolate wrt. w among
              points transformations[k] and hand.point transformations[k]
              else
               points transformations [k] = hand.point transformations[k]
             end if
             feature points weights[k] = max (0, feature points weights[k] - w)
          end if
     end for
  end if
end while
```

The characters produced by using the blueprint and blender application, by utilizing the joints feature in the blender so that the characters that have been made with a blender can move according to the wishes of the animator.

4. Evaluation of the BSL signing avatar

Evaluation in context to virtual signers was completed in Iraq, using local BSL signer whose actions were captured through motion capture technique, but only hand and body measurements were recorded. In regard to our information, hand and body movements have been used simultaneously to manage movement, facial expressions, and vision, so we have created an evaluation of the proposed system that is divided into two parts: We examine the significance of expression and gaze movement, and secondly, we evaluate the effectiveness of animation systems in producing a realistic and understandable signature avatar.

Respondents see the first video shown based on monitoring general replay of motion capture footage. In the sequel, Avatar explains what his school life needs and describes the preparations he has made for their release. Respondents was asked to rate the signs of comprehension used to be comprehensive their entire story. Alongside of it, the aspect of realism associated with the avatar that too using a scale ranging from 1 – 5 known as the Likert scalene, with five being extremely practical or very comprehensible. Motion capture sequences on an average, all the participants are rated to be at 3.12 in context to realism, 3.48 in context to sign comprehension, as well as 3.88 for their story comprehension, as conferred from Table 1. Although these numbers are considered to be lower than

expected. They tend to remain at 3 just above median threshold. The 2nd as well as 3rd tests showed 2 dialogues and the respondents were asked to rate them according to those factors associated with the controlling of playback sequences. It revolves around realism, sign as well as story comprehension allowing test on our system's ability. It was instrumental in producing combined motion segments into believable linguistic utterances all across different channels. 2 dialogues comprise of school's life scenario as tokens related to the recorded corpus. We produce novel utterances in multiple numbers' all-round the similar subject. One sequence as constructed is diagramed as well as discussed to show motion retrieval and its combination using Figure 10.

.

A: What do you need for school?

B: I need pencils.

A: Anything else?

B: I need a notebook.

Respondents rated stories ratings ranged from1 to 2 to control sequential playback. Suggesting surveyed hesitations in regards to the signing of avatars.

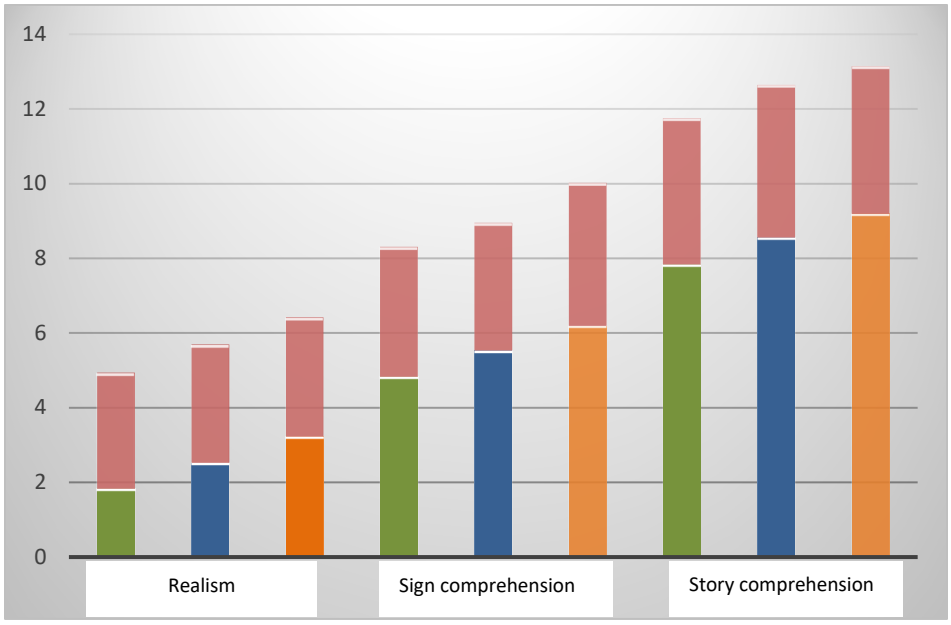

Figure 10. Ratings associated with all 3 sequences as grouped by question with visually similar responses.

The contribution is toward nursery rhymes use in British Sign Language (BSL) with their specific phonetic-syllabic rhythmic virtual characters linguistic to match children's biological stimuli in MPEG-4 approach the animation system is dedicated to linguistic interaction amongst human as well as virtual characters using BSL. The qualitative clear that respondents didn't notice much of a difference between facial animations. Kids describe help to improve the animation system. They are not instrumental in finding our concatenated sequences. Anything which is less real or can be said more understandable than those of the ones associated with simple sequences of playback. Responses are therefore quantified in context to the Table 1.

Table 1. Respondent's ratings as shared by them in the form of a simple play backing sequences with two stories that are created with or without the animation engine. The contextual response scales that are ranged from 1 aren't realistic or at least comprehensible to 5. It is further considered to be in the form of a very realistic yet highly comprehensible value. P values compare the story sequences to the control playback sequence.

Table 1: Response's rating quantified in context.

	Playback($\bar{x}, \sigma$)	Story 1 ($\bar{x}, \sigma$)	Story 2 ($\bar{x}, \sigma$)
Realism	3.5, 0.9	3.2, 0.8	3.2, 0.8
Sign Comprehension	3.35, 0.66	3.33, 0.9	3.7, 0.9
Story Comprehension	4.0, 0.9	4.1, 0.6	3.9, 0.7

Given the human signer in context to the wearing of the distractions placed in the markers during recording sessions it was evaluated that concatenated sequences in contrast to the videos showing human signer's to perform same continuous sequences. As conceived as an able form of differentiation the same phrased satiations obtained from different recording techniques can be compared with one another. They include a video, mocap playback, along with mocap concatenation for this comparative evaluation against each other to pave way for future studies. To the 3rd question mentioned above, we were instrumental in conducting a frequency-based analysis associative to different behavioral aspects of the kids when compared with avatar's behaviors. Figure 11 intend to show that the kid's behavioral rate toward each response refers to avatar's production's category. The bars present in each of the categories don't adds up to 1 necessarily. It is due to the fact that, the baby at times does responds with multiple forms of responses.

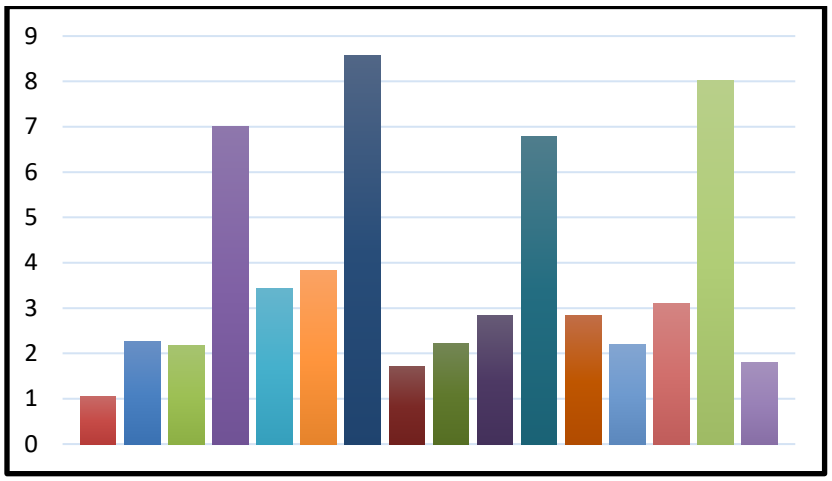

Figure 11. The kid's behavioral rate toward each response refers to category of character signer's production.

Described in Figure 11, the babies tend to respond differently during the presence of the avatar in "Linguistic Nursery Rhyme conversational mode". It was further observed in case of other modes when compared with those of the Social, Idle, along with 3-Way. The babies were instrumental in producing a huge linguistic responses percentage. It is largest amongst all the avatar's "Linguistic Nursery Rhymes" comprising of 35percent in Nursery Rhymes when

compared with 12 percent Idle, including 20 percent Social Gestures, along with to 23 percent associated with 3-Way. Not only this, these babies' responses to the avatar's "Linguistic Nursery Rhymes" when compared with avatar are other types of conversational turn. It involves largely riveted conversations turned into fixed Sustained state of them.

DISCUSSION AND CONCLUSION

In making animations there are still many who use old methods that will take a long time in the production of animation because they have to move the animated characters manually and save the movement into the frame, with synchronization animation can help in the very long animation process to be more short because without having to move the animated character manually and save it into the frame, using digital data recorded using Motion Capture Kinect X-Box360 and synchronizing with animated characters makes the animator's work faster.

The proposed system is capable of creating multi-channel presentations of sign language and real-time MPEG-4 animations, including hand, face and andeye motion. After all, it was able to do it with realism and understanding - the equivalent of realized motion capture sequences. We detailed the initial assessment of our pronunciation generation strategies at BSL. Several results are highlighted: First, our experiments confirm the importance of including facial expressions in sign language; In addition, our facial animation method demonstrates strong qualities. This result strengthens the idea of using this kind of MPEG-4 model to synthesize any facial expression from motion capture data. Surprisingly, we also came to the conclusion that the experiments conducted with our signature avatar did not contain any details on the flow of vision-oriented signs. Finally, our movement composition process allows the formation of new accents with promising results, as respondents fail to distinguish synthetic movement from the movement of our survey lessons.

Although the framework that is reported in context to this paper is convinced of providing an advanced utterances production attempts in regards to a virtual signer. All these attempts as made are considered to be believable and are accepted in a better way from deaf peoples. There is a need to address several unresolved problems while introducing or using MPEG-4 techniques. The virtual signer's appearance impacts on avatar's believability. On the other hand the mentioned type of realistic avatar could fail at the face of the problems associated with uncanny valley. Other types of avatar possess an abstract, sharper along with cartoonish representation of the tested.

In future, significant amount of improvements are expected in the animation system: with the introduction of new controllers they handle, spatial coherency, detection of collision, or constraints in dynamics movements-based constraints. It is significant in developing various partial human sensational methods helping them dealing with the coordination schemes. These are placed in among various channels that develop new methodological evaluations. That more the sign's comprehension degree is thoroughly analyzed through the inclusion of a detailed questionnaires along with addressing toward expressivity degree associated with the virtual signer.

References

Ali I R, Kolivand H, Alkawaz M H (2018) Lip syncing method for realistic expressive 3D face model. Multimedia Tools and Applications 77(5): 5323-5366.

Anderson D R, Pempek T A (2005) Television and very young children. American Behavioral Scientist 48(5): 505-522.

Bigand F, Prigent E, Braffort A (2019) Retrieving Human Traits from Gesture in Sign Language: The Example of Gestural Identity. In Proceedings of the 6th International Conference on Movement and Computing, pp 1-4.

Brian Scassellati, Jake Brawer, Katherine Tsui, Setareh Nasihati Gilani, Melissa Malzkuhn, Barbara Manini, Adam Stone, Geo Kartheiser,ArcangeloMerla, Ari Shapiro, David Traum, Laura-Ann Petitto (2018) Teaching Language to Deaf Infants witha Robot and a Virtual Human. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Canada, pp 1-13.

Buyun Sheng, Feiyu Zhao, Chenglei Zhang, Xiyan Yin, Yao Shu (2019). Parameterized representation and solution method of the lightweight 3D model virtual assembly constraint. Journal of Ambient Intelligence and Humanized Computing 10(3): 1167-1187.

Cassell J, Vilhjálmsson H, Bickmore T (2004) Beat: the behavior expression animation toolkit. In Life-Like Characters, Springer, Berlin, Heidelberg, pp 163-185.

Chai J, Hodgins J K (2007) Constraint-based motion optimization using a statistical dynamic model. In ACM SIGGRAPH 2007 papers. 8-es.

Chiu Fu-Sheng (2007) Interactive news gathering and media production control system. U.S. Patent Application 11/172, 858.

Clare E Sims, Eliana Colunga (2013) Language Development in the Age of Baby Media: What We Know and What Needs to be Done, Cascadilla Press. Retrieved on March 2: 384–396.

Davani A M, Shirehjini A N, Daraei S (2018) Towards interacting with smarter systems. Journal of Ambient Intelligence and Humanized Computing 9(1): 187-209.

Escudeiro P, Escudeiro N, Norberto M, Lopes J, Soares F (2017) Digital Assisted Communication, in Proceedings of the 13th International Conference on Web Information Systems and Technologies, pp 395–402.

Félix Bigand, Elise Prigent, Annelies Braffort (2019) Animating Virtual Signers: The Issue of Gestural Anonymization. In Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents IVA ’19, France, pp 2–5.

Gibet S, Courty N, Duarte K, Naour T L (2011) The SignCom system for data-driven animation of interactive virtual signers: Methodology and Evaluation. ACM Transactions on Interactive Intelligent Systems (TiiS) 1(1): 1-23.

Grochow K, Martin S L, Hertzmann A, Popović Z (2004) Style-based inverse kinematics. In ACM SIGGRAPH 2004 Papers, pp 522-531.

Hartmann B, Mancini M, Pelachaud C (2005) Implementing expressive gesture synthesis for embodied conversational agents. In International Gesture Workshop, Springer, Berlin, Heidelberg, pp 188-199.

Jaballah K, Jemni M (2012) Sign language parameters classification from 3D virtual characters. In 2012 International Conference on Information Technology and e-Services, IEEE, pp 1-6.

Jemni M, El Ghoul O, Yahia N B, Boulares M (2007) Sign Language MMS to Make Cell Phones Accessible to the Deaf and Hard of hearing Community in CVHI.

Jemni M, Elghoul O (2008) A system to make signs using collaborative approach. In International Conference on Computers for Handicapped Persons, Springer, Berlin, Heidelberg, pp 670-677.

Judy S DeLoache, Cynthia Chiong, Kathleen Sherman, Nadia Islam, MiekeVanderborght, Georgene L Troseth, Gabrielle A Strouse, and Katherine Oâ˘A ´ ZDoherty. (2010) Do babies learn from baby media? Psychological Science 21(11): 1570–1574.

Kaneko H, Hamaguchi N, Doke M, Inoue S (2010) Sign language animation using TVML. In Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry, pp 289-292.

Kipp M, Heloir A, Schröder M, Gebhard P (2010) Realizing multimodal behavior. In International Conference on Intelligent Virtual Agents, Springer, Berlin, Heidelberg, pp 57-63.

Kopp S, Wachsmuth I (2004) Synthesizing multimodal utterances for conversational agents. Computer Animation and Virtual Worlds 15(1): 39–52.

Kopp S, Krenn B, Marsella S, Marshall N, Pelachad C, Pirker H, Th´orisson K, Vilhjlmsson H (2006) Towards a common framework for multimodal generation: The behavior markup language. ." In International workshop on intelligent virtual agents, Springer, Berlin, Heidelberg, USA05–217, pp 205-217.

Kranstedt A, Kopp S, Wachsmuth I (2002) MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents. In Proceedings of the AAMAS02Workshop on Embodied Conversational Agents - let's specify and evaluate them. Italy.

Li B Y, Mian A S, Liu W, Krishna A (2013) Using Kinect for face recognition under varying poses, expressions, illumination and disguise. In 2013 IEEE workshop on applications of computer vision (WACV), IEEE, pp 186-192.

Lombardo V, Nunnari F, Damiano R (2010) A virtual interpreter for the Italian sign language. In International Conference on Intelligent Virtual Agents, Springer, Berlin, Heidelberg, pp 201-207.

Marina Krcmar (2011) Word Learning in Very Young Children from Infant-Directed DVDs. Journal of Communication 61(4): 780–794.

Nasihati Gilani S, Traum D, Sortino R, Gallagher G, Aaron-Lozano K, Padilla C, Petitto L A (2019) Can a Virtual Human Facilitate Language Learning in a Young Baby?. In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, pp 2135-2137.

Patricia K Kuhl, Feng-Ming Tsao, Huei-Mei Liu (2003) Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. Proceedings of the National Academy of Sciences 100 (15): 9096–9101.

Scassellati B, Brawer J, Tsui K, Nasihati Gilani S, Malzkuhn M, Manini B, Traum D (2018) Teaching language to deaf infants with a robot and a virtual human. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp 1-13.

Sutopo J, Abd Ghani M K, Burhanuddin M A, Ardiansyah H, Mohammed M A (2019) The Synchronisation Of Motion Capture Results In The Animation Character Reinforcement Process. Journal of Southwest Jiaotong University. 54(3).

Tolani D, Goswami A, Badler N I (2000) Real-time inverse kinematics techniques for anthropomorphic limbs. Graphical Models 62(5): 353–388.

Verma H V, E Aggarwal, Chandra S (2013) Gesture recognition using Kinect for sign language translation. In 2013 IEEE Second Int. Conf. on Image Inf. Process, pp 96–100.

Vesel J (2005) Signing Science! Andy And Tonya Are Just Like Me! They Wear Hearing Aids And Know My Language!?. Learning & leading with technology 32(8).

Wheatland N, Wang Y, Song H, Neff M, Zordan V, Jörg S. (2015) State of the art in hand and finger modeling and animation. Computer Graphics Forum 34 (2): 735-760.

Yang-Kun Ou, Yu-Lin Wang, Hua-Cheng Chang, Shih-Yin Yen, Yu-Hua Zheng, Bih-O. Lee (2020) Development of virtual reality rehabilitation games for children with attention-deficit hyperactivity disorder. Journal of Ambient Intelligence and Humanized Computing 11(11): 5713-5720.

Yao L, Peng X, Guo Y, Ni H, Wan Y, Yan C (2018) A Data-Driven Approach for 3D Human Body Pose Reconstruction from a Kinect Sensor. Journal of Physics: Conference Series, IOP Publishing 1098 (1): 012024.