AN ANIMATION SYSTEM TO SPECIAL PEOPLE SPEECH ORIENTATION

Jesuliana N. ULYSSES[1] and Aura CONCI[2]
[1]Federal University of de Juiz de Fora, Brazil [2]Federal University Fluminense, Brazil

Abstract. In the human evolution, the language is the most important process [2]. It constitutes a critical element not only for the acquisition of other symbolic systems, like reading, writing, or developed mathematical conventions, but also for the development of human relationship [15]. Through the language the man translates and transmits his thought and inserted himself in a society. When a person loses the capacity to communicate in one way, he must find alternatives that propitiate the development of others kind of language. In the people deprived of the sense of hearing, the alternatives should be the substitution of the audition for other channels as the vision, the touch and movement, besides the use of the existent vestiges of audition. Those people can develop an oral language or a sign language, depending on their capability and stimulus received. The specialized literature (for mute and deaf persons) relates alternative communication modes using series of symbols, allowing people's communication without production of oral language: Symbols Bliss System [3], Pictogram Ideogram Communication System[19] and Picture Communication Symbols [18] [19] are some example traditionally used by disabled peoples. More recently, these systems presents a computerized versions [9][10][11] and other systems were built for such endeavor like the ImagoAnaVox [10]. Deficient hearing people using with language of signs can also make use of computerized systems of communication, just as the system LOGOFONE [9]. These computerized systems present certain characteristics that turns them more adapted to the users' specific needs, facilitating the communication process. However, they are not related to learning deficient hearing people. Deaf people can know how to hear by lips reading. If they are not mute (but only incapable of speech because not learn it naturally by listen), they can know how to speech by emitting sounds at the same time that they mimic mouth elements position and movement. This technique involves a willingness to share efforts of professionals from psychology, medicine and audio-phonology and have been proved beneficial to help people since it is adopted by specialized instructional program like INES (Instituto Nacional de Educação de Surdos - http://www.ines.org.br/).The oral language uses body and facial language of the interlocutor. This is denominated oro-facial or oro-labial, for understanding it identifies the corporal expression, the eyes expression, the facial muscles contractions and the lips movements. Due this characteristic of the oral language, facial animation can be used in aid deaf learning systems. The use of the facial animation for special people education systems improves the perception of the speech and its understanding. The use of facial animation in such system is reinforced by the visual and audio characteristics of the speech. Through the construction of the visual image of the speech, that is named *visemes* in facial animation, disable people can identify and learn the formats of the lips on pronouncing a certain phoneme, aiding in the learning of the language. The development of those systems needs the integration of professionals of several areas. The main proposal of the research line of this work is develop affordable tool to aid this therapy. This paper describes the animation techniques used and some preliminary results of this ongoing project

Keywords: Facial animation, controlled facial expressions, mouth modeling, real time.

1. INTRODUCTION

In the human evolution, the language is the most important process [2]. It constitutes a critical element not only for the acquisition of other symbolic systems, like reading, writing, or develop mathematical conventions, but also for the development of human relationship [15]. However, one in each 200 people presents speech deficiency [9]. Understandably, it is important and possible to teach such people how to communicate or how to use information systems to interchange thoughts or feelings.

The deaf and dumb specialized literature relates alternative communication modes using series of symbols, allowing people's communication without production of oral language. Symbols Bliss System [3], Pictogram Ideogram Communication System - PIC [19] and Picture Communication Symbols - PCS [18][19] are some example traditionally used by disabled peoples. More recently, these systems present a computerized versions [9][10][11] and other systems were built for such endeavor like the ImagoAnaVox [10], that uses a varied of peripheral devices (sensitive the touch or to blow screen, specials mouse, or even guttural commands) resources for system assessment. Deficient hearing people using with language of signs can also make use of computerized systems of communication, just as the system LOGOFONE [9]. These computerized systems present certain characteristics that turns they more adapted to the users' specific needs, facilitating the communication process. However, they are not related to learning deficient hearing people. Deaf people can know how to hear by lips reading. If they are not mute (but only incapable of speech because not learn it naturally by listen), they can know how to speech by emitting sounds at the same time that they mimic mouth elements position and movement. This technique involves a willingness to share efforts of professionals from psychology, medicine and audio-phonology and have been proved beneficial to help people since it is adopted by specialized instructional program and institutes like INES (Instituto Nacional de Educação de Surdos - http://www.ines.org.br/). The therapy constitutes of *in vivo* teaching and auto-training sessions, and must rely on previous researches on what is the influence of each mouth element on the formation of the basic sound of a language. Figure 1 exemplifies this for the "u" sound [16]. In Portuguese, words are parsed in syllables, where vowels (and then diphthong and consonants+vowel sounds) are mainly important [17]. The idea of the research line, in which this work constitutes a first step, is development of affordable tool to aid this therapy. This paper describes, in next sections 2 and 3, the basis of the animation techniques used and the abstract muscle-based model. Then, in section 4, the approach used for measurements of the mouth movement and lips position are described. Implementation aspects (section 5) and its results (section 6) are commented then, as well as some preliminary conclusions (section 7) of this ongoing project on speech aid.

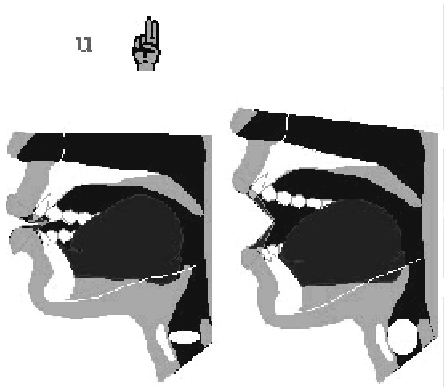


Figure 1: Mouth elements for "u"

2. FACIAL EXPRESSION ANIMATION

Nowadays, virtual characters are present in our audiovisual environment and become more and more interactive. They appear on television screens as presenters and in the cinema as main characters of stories. In computers systems much effort is concentrated into giving to the virtual characters a human appearance and behavior. These characters (avatar) use human characteristics and mimic humans in order to communicate with us. The avatar can be a human projection in a graphical virtual world or a representative of this world. A virtual characters representing a person is generally use to communicate another person. The autonomous avatar also establishes a link between the machine and the exterior world. For instance, they may help the computer game user to find his way, or they can play the role of a house seller and show the photos of future accommodation [1]. The main aspects of human communication are speech and facial expressions. The controlled facial animation is a complex task due to the huge complexity of the real structure of the face, which is composed of muscle, bone and skin. Motions on this complex structure are difficult to simulate because small changes make different expressions and humans are used to reading them intuitively. The facial expression is related with speech mainly by mouth shape. The lips and other elements of the face are controlled by the muscle movements. Muscles change the face expression combining theirs effect in one area. The possible movements are caused by contraction of some muscles and relax of others. There are two types of muscle in the human face: linear and sphincter. A linear muscle becomes thickened and shortened with contractions, and pull the face in a direction. Sphincter muscles are circular band muscles that encircle and close an orifice of the body or one of its hollow organs. On the face they surround the mouth closing the lips. Complex expressions are due to combination of these unidirectional movements.

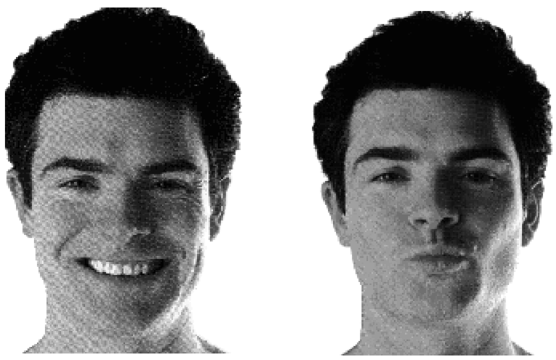


Figure 2: On the left expression, mainly two linear muscle (the Zygomatic Major Left and Right) are stressed. On the right the sphincter Orbicularis is used to close the lips.

Face animation is a difficult issue due to this muscular anatomic complexity (figure 3). Facial animation techniques uses different scheme to describe the relation of the muscle actions and its effects on the facial expressions. One of these schemes, called Facial Action Coding System (FACS), uses the idea of Action Units (*AU*) in order to associate them with the muscles responsible of these changes [12]. The *AU* are described as the smaller visible changes on the human face. For instance *AU 10*, raises the upper-lip region of the face. The muscle for this movement runs from roughly the center of the cheeks to area of the naso-labial groove. Associated with this muscular action the skin above the upper-lip is pulled upward and toward the cheeks, pulling the upper-lip up. Considering the muscular anatomy of the face (see figure 3), the association between this *AU* and the facial muscles is: the *Levator Labii Superioris* and *Caput Infraorbitalis* muscles are responsible for the AU 10, which action in a controlled face expression is: *Upper-Lip Raiser*.

3. THE USED FACIAL MODEL

Nowadays, the *abstract muscle-based model*, first reported in 1987 by Waters (http://www.crl.research.digital.com/projects/ facial/facial.html), is one of the most popular and complete models. It used in many other works [1][2][7]. It is based on the human facial anatomy due to the fact it uses abstract muscles to modify the polygon mesh that represent the face (figure 5). Two types of abstract muscle are modeled: the linear muscle to pull the mesh and the sphincter to squeeze the mesh.

This work uses the Waters' *abstract muscle-based model* to control the synthetic face movement. That is, the mouth can be adequately controlled by the two types of muscle behavior. The linear muscles can move their influential zone (figure 4) in the direction showed on figure 7. On the face model this muscles deformed a polygonal mesh which patches after render represent the skin. [7]. In order to modify the polygon mesh (figure 5) through a muscular control, the algorithm travels the vertex list to find those that are in the influence zone.

In addition to these two types of abstract muscle, a control for the jaw rotation was developed. The rotation of the jaw is done by rotating the vertices that compose the jaw in relation to the point of the face (figure 7).

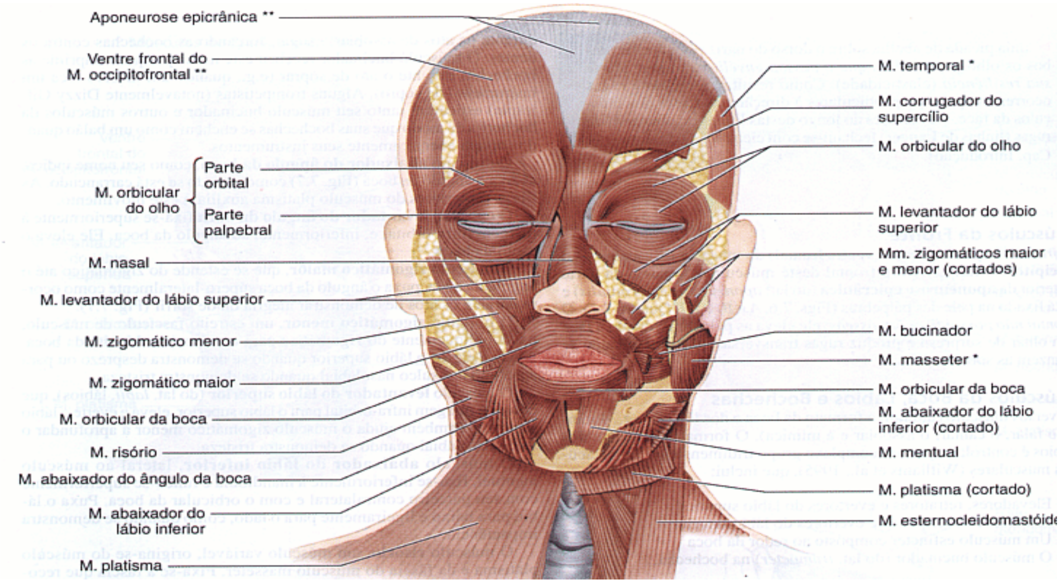


Figure 3: The facial muscular anatomy.

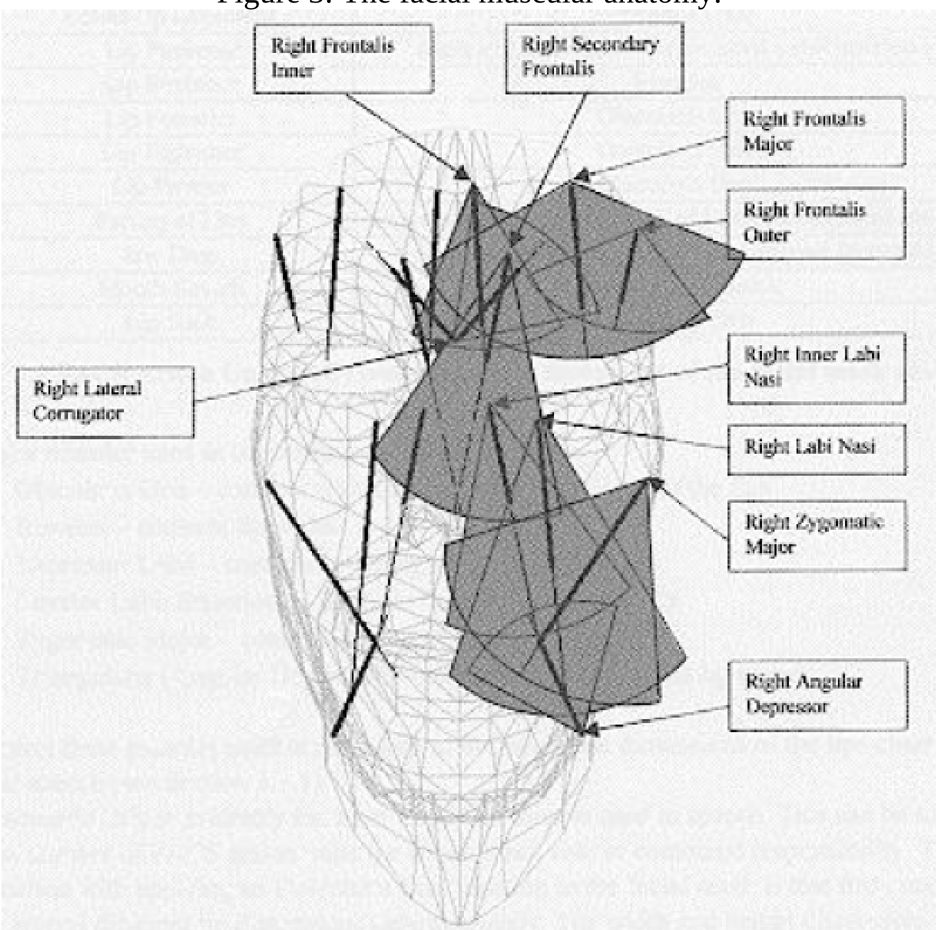


Figure 4: Overlapping muscular in fluencies upon mask facial, as uses in Water’s original model

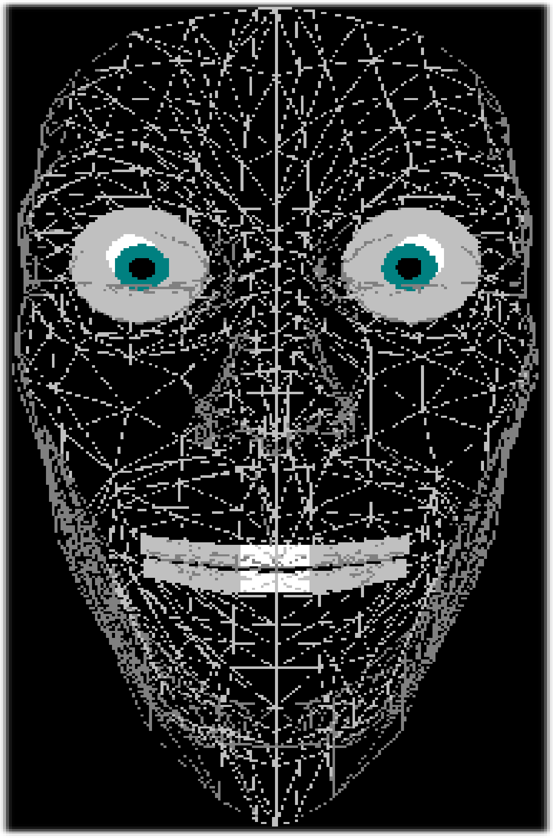


Figure 5: Polygon mesh

Through these few controls, muscles and jaw rotation, the polygon mesh can be deformed and thus produces large range labial expression. A main fault of this model is that it is just a vague approximation of the anatomy; the linear muscles do not pull the skin towards a single attached point. The implementation uses macros to create the animation of the mouth. These macros control duration of the expression using the influence zone of the muscle movements. On figure 7 these control areas can be seen. The thin gray lines determine the direction of the movements and the muscle associated with the deformation of the region in the face. The numerical value of this deformation is obtained from the experimental parameters, explained in next section 4 (Methodology). These values attributed adequately to muscle control generate the phonetic facial animation.

4. METHODOLOGY

Methodology is based on the observation that the position of the mouth over a short period can be correlated with the phonetic sound of the speech over the same time interval. The motion of the lips, tongue, mouth and jaw of a speaker can be deduced from the phonetic speech [16][15] Although, previous researches help a lot on identification of the important elements related with each sound, their objective is not animation. Therefore, we have to develop a proper approach to measure the mouth position during the speech process. This approach begins by the identification of the relevant mouth parameters. The parameters controlling the mouth elements position and shape are predicted from an analysis of Portuguese phonemes and the possible control of the facial animation (figure 6). After some analysis we conclude for the measure of four parameters. They are named JAW, FLARE, CORNERS and EDGES [4,][20]. Figure 6 shows the used parameters. JAW is the jaw position and is measured as the distance between the two (upper and lower) jaws. FLARE is the height of the maximum vertical aperture between lips. The other two parameters characterize the horizontal movement of lips. The EDGE measures the lips joint points, where the upper and lower lips connect. CORNERS represent the horizontal aperture of the lips during speech. All parameters will be measured in a video analysis and image processing software and scaled in accordance with the interlocutor distance of the speaker. The mouth measurements are in millimeters. The emphasis is on Portuguese vowels sounds, then diphthongs and finally combination of consonant with vowel. The map of consonant plus vowel sound is not finished by the time we are writing this work.

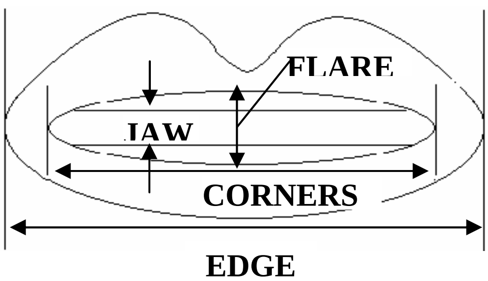
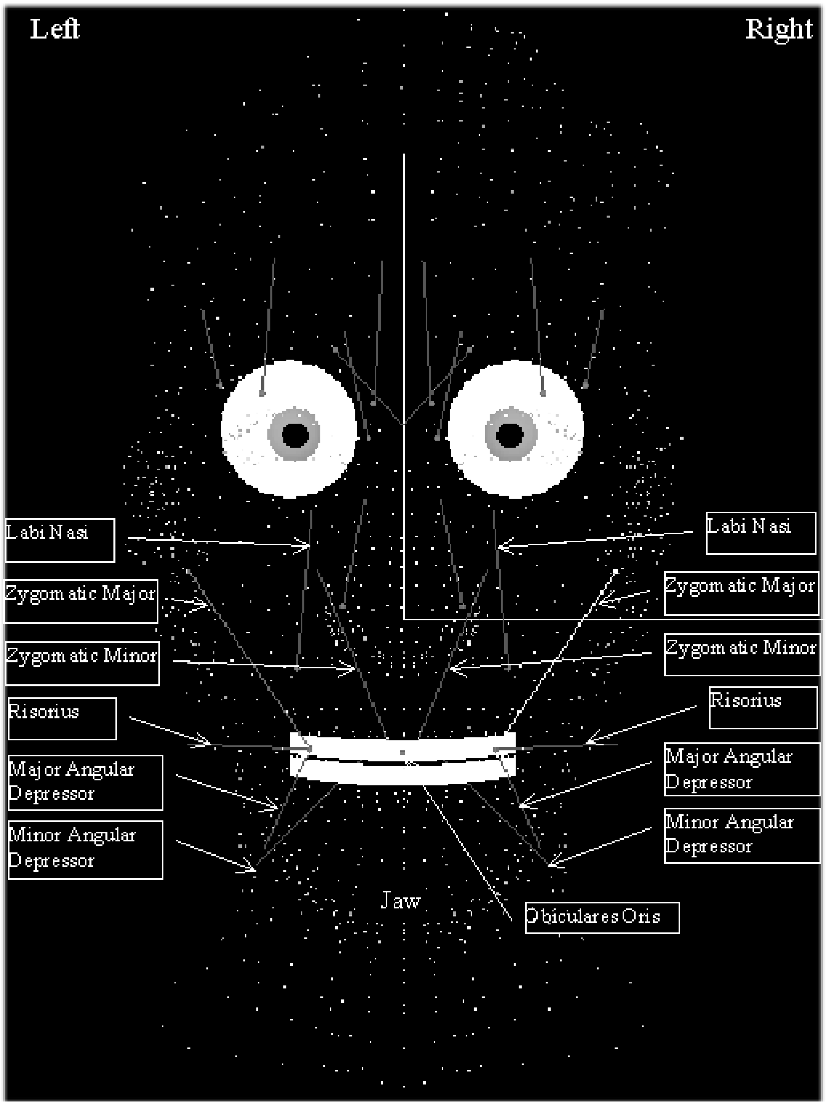


Figure 6: Mouth Parameters



Muscle simulated	Shape Motion
ObicularisOris (sphincter)	increases lip rounding, reduces lip width, increases lip protrusion
Risorius (ledft/right linear)	increases lip width
Depressor Labii (left/right linear) Labii Superioris (left/right linear)	lowers bottom lip Raises upper lip
Zygomatic Major (left/right linear)	Raises the lip corners
Traingularis (left/right linear)	lowers the lip corner
Jaw Rotation	mouth openness

Figure 7: Muscle association and control directions used for phoneme animation

5. IMPLEMENTATION

This section explains the main characteristics of the implementation of the parameter of measure in the abstract muscle-based model, the process to animate this model and the graphical interface.

The system was implemented in C++ using OpenGL, in PCs. The main objects of the face are: *Face* object, *FaceIO* object, *FaceStructure* object, *Skin* object, *Expression* object, *ListeExpressions* object and muscle hierarchy. The *Face* object is the first interface in the implementation of Water's model. The *Face* object can: *load* a face from files (Create function), *save* a face in files, *contract a muscle* given its name, *open the mouth, rotated the jaw* , etc.

FaceIO object is used to read data from the files and construct a new *FaceStructure* object with this data. *FaceStructure* object is composed of: list of muscles; two *Eyes* object; a *Teeth* object; a *Skin* object and a vector of 3D points. This vector represents the displacement of the mask from vertices of the original polygon mesh.

The *Skin* object is composed of 6 elements. (1) A vector of vertices (3D point) used as the position of the face after the loading. (2) A second vector of vertices representing the facial mask which is displaced by the muscle and drawn. (3) A vertex index vector to know the vertices composing the polygons to draw. (4) A vertex index vector representing the vertices contained in the jaw. (5) A vector of *discontinuity*. (6) And a value representing the rotation of the jaw.

To create an *Expression* the constructor needs: an expression name, a vector of name muscle, a vector of contraction and a value for the rotation of the jaw. *Expression* object is composed of the 4 following elements. (1) A vector of muscle names used to make the expression. (2) A vector of muscle contraction corresponding to the name muscle at the same index and indicating the contraction of this muscle. (3) A value corresponding to the jaw rotation used to make the expression. (4) And the names of the expression.

The *ListeExpression* object is a vector of expressions. This object is able to load or save itself from or in a file. To facial animation the application loads face from a file (.fat), loads an animation (.exp) file and asks the *face* object to show an expression, at a given time (in the animation function *ShowExpression (int time)* of the object *face*). There is only one *FaceAnimation* object by *Face*, but different animation can be loaded from a ".exp" file.

The muscles are represented by an abstract class *Muscle* from which two other class are derived: *LinearMuscle and SphincterMuscle*. This structure is used because many functions exit in these two types de muscle and have or have not the same implementation. For efficiency reasons a vertex index list of vertices in the influence zone of the muscle is set up. Consequently when a muscle is activated the algorithm does not go through whole vertex list of the *Face*. Also for same reason, a level of details (*LoD - Low level of details*) is associated with each muscle indicating from which level the muscle is active.

There are the six levels of *LoD* (table 1). Muscle are active if their level is superior or equal to the *FaceStrucuture LoD* level. The attribution of a certain level to a certain muscle is guided by the importance of the facial deformation that their action makes.

6. EXAMPLE AND CONCLUSION

To evaluate the performance of the approach the sound "a"+ "e" is presented on figure 8. These mouth movements use the file listed on figure 9. This file uses four macros. Each one of these macros is composed of the date like:

{MACRO}//begin macro
<name> *<time>*
<muscle_name> *<measurement>*
{/MACRO} //end macro

LoD Level	Muscle name
0	There is not muscle actif
1	Left Frontalis Major, Right Frotalis Major and jaw rotation
2	Obicularis Oris, Left Zygomatic Major, Right Zygomatic Major, Left Major Angular Depressor, Right Major Angular Depressor

LoD Level	Muscle name
3	Left Risorius, Right Risorius, Left Zygomatic Minor, Right Zygomatic Minor, Left Minor Angular Depressor, Right Minor Angular Depressor, Left Secondary Frontails, Right Secondary Frontails, Left Labi Nasi, Right Labi Nasi
4	Left Frontails Outer, Right Frontails Outer, Left Lateral Corigator, Right Lateral Corigator, Left Frontails Inner, Right Frontails Inner, Left Inner Labi Nasi and Right Inner Labi Nasi

Table 1 - LoD - Low level of details

In each macro, *name* is the macro's name; *time* is the animation time. After these parameters all the muscles used in the motion are listed with the corresponding deformation (figure 9).

Although, we are at the initial steps in direction of a complete system to aid special people to speech, we can see that the presented methodology and the facial model used enable the construction of such system.

Next steps in this direction will be (1) systematized of the methodology of measurements proposed in order to turn possible easy map of the four parameters for many other Portuguese phonetic sounds. (2) Composition of a dictionary of macros relating the measures with the muscle. (3) Add improvements on the interface to turn the system adequate to different levels of user (children, adult).

```
{MACRO}
Normal
Jaw 0.0
{/MACRO}
{MACRO}
a 10
Jaw 8.7
Left_Major_Angular_Depressor 1.0
Right_Major_Angular_Depressor 1.0
Obicularis_Oris 1.3
Left_Labi_Nasi 0.2
Right_Labi_Nasi 0.2
Left_Minor_Angular_Depressor 0.6
Right_Minor_Angular_Depressor 0.6
Left_Risorius 0.4
Right_Risorius 0.4
Left_Zygomatic_Major 0.3
Right_Zygomatic_Major 0.3
Left_Zygomatic_Minor 0.6
Right_Zygomatic_Minor 0.6
Left_Labi_Nasi 0.2
Right_Labi_Nasi 0.2
{/MACRO}
{MACRO}
e 10
Jaw 5.8
Left_Major_Angular_Depressor 0.6
Right_Major_Angular_Depressor 0.6
Obicularis_Oris 1.0
Left_Labi_Nasi 0.2
Right_Labi_Nasi 0.2
Left_Minor_Angular_Depressor 0.6
Right_Minor_Angular_Depressor 0.6
Left_Risorius 0.4
Right_Risorius 0.4
Left_Zygomatic_Major 0.8
Right_Zygomatic_Major 0.8
Left_Zygomatic_Minor 0.6
Right_Zygomatic_Minor 0.6
Left_Labi_Nasi 0.2
Right_Labi_Nasi 0.2
{/MACRO}
{MACRO}
Normal
Jaw 0.0
{/MACRO}
```

Figure 9: File macro

ACKNOWLEDGMENT

This work was supported by CNPq and CAPES. Our deepest gratitude goes to Emmanuel Tanguy and Fabio Policarpo for they help with the facial model and for the

great improvement given to the first steps of this works by opening to us all their C++ codes of this model.

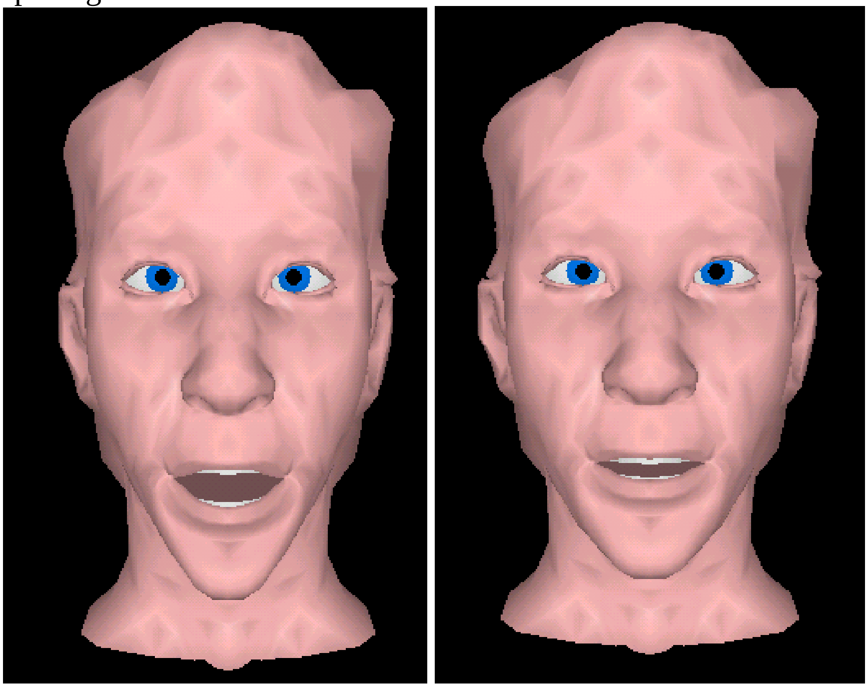


Figure 8: Vowel sounds “a+e”

REFERENCES

[1] A. Watt, (2001). *3D Games Technology, Real Time Rendering and Character Animation*, XV ECI - Escuela de Ciencias Informáticas , 23 a 28 de Julho - Departamento de Computación - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, http://www.dc.uba.ar/eci/2001/watt.html

[2] A. Watt and F. Policarpo, (2001). *3D Games: Real-time Rendering and Software Technology*, vol. 1, (1a Ed.), Addison-Wesley, UK.

[3] B. Hehner (Ed.) 1980. Blissymbols for use. (4th. printing). Blissymbolics Communication Institute, Ontario, Canada.

[4] D. McAllister, R. Rodman, and D. Bitzer. Lip Synchronization of Speech, *Proceedings of the Audio-Visual Speech Processing Conference '97 (AVSP'97)* 1997, 133...136.

[5] E. C. Macedo, F. C. Capovilla, M. J. Gonçalves, A. G. Seabra, , V. O. Thiers & M.D. Feitosa. Adaptando um sistema computadorizado pictográfico para comunicação em paralisia cerebral tetra-espástica. *Anais da II Jornada USP-SUCESU-SP de Informática e Telecomunicações*, 1994, 353 ... 361, São Paulo.

[6] E. C.Macedo, F. C. Capovilla, V. O. Thiers, A. G. Seabra & M Duduchi. Instrumento computadorizado para exploração de comunicação pictográfica em paralisia cerebral com comprometimento cognitivo leve: PCS-Comp v40s. *Resumos do I Encontro de Técnicas de Exame Psicológico: Ensino, Pesquisa e Aplicações*, 1994 pp 12, São Paulo.

[7] E. Tanguy (2001), Abstract muscle model for three dimensional facial animations, dissertation supervised by A. Watt, Computer Science Dep. University of Sheffield.

[8] F. C. Capovilla. Pesquisa e desenvolvimento de novos recursos tecnológicos para educação especial: Boas novas para pesquisadores, clínicos, professores, pais e alunos. In: E. Alencar (Org.), Tendências e desafios de Educação

Especial 1994, 196 ... 211. Brasília: Secretaria de Educação Especial.

[9] F.C. Capovilla, E. C. Macedo, M Duduchi et al. Instrumento computadorizado para exploração e habilidades lingüísticas e de comunicação simbólica em paralisia cerebral sem comprometimento cognitivo: Bliss-Comp v40s. in: *Resumos I Encontro de técnicas de exame psicológico: ensino, pesquisa e aplicações.* 1994a, São Paulo.

[10] F.C. Capovilla, E. Macedo, M. Duduchi, & A.G. Capovilla, W.D. Raphael & M. Guedes. UltrAACtive: Computerized multimedia expert AAC system. *Proceedings of the VII Biennial Conference of the International Society for Augmentative and Alternative Communication*, 1996, 467 ... 468, Vancouver, B.C., Canada.

[11] F. C. Capovilla, E. C. Macedo, M. Duduchi, V. O. Thiers, A. G. Seabra & M. Guedes. Instrumento computadorizado para exploração de habilidades linguísticas e de comunicação simbólica em paralisia cerebral sem comprometimento cognitivo. Bliss-Comp v40s. *Resumos do I Encontro de Técnicas de Exame Psicológico: Ensino, Pesquisa e Aplicações*, 1994, pp 8, São Paulo.

[12] F.I. Parke and Waters (1996) Computer Facial Animation, A.K. Peters Ltd.

[13] J. M. Martino; L. P. Magalhães. Um Conjunto de Visemas para uma Cabeça Falante do Português do Brasil. Faculdade de Engenharia Elétrica e de Computação Universidade Estadual de Campinas. Campinas, SP, 2004

[14] J. N. Ulysses. Uma proposta de sistema para ensino da linguagem oral a deficientes auditivos. Dissertação de mestrado: IC/UFF. Niterói, RJ. Abril, 2003

[15] J. R. Schumaker and J. A. Sherman, (1978). Parents as intervention agents: from birth onward. In R. L. Schiefelbusch (Ed.), Language intervention strategies, Baltimore: University Park Press, 237...315.

[16] P. M. Tujal and M. N. Souza, “Speech Aid for Deaf based on a Representation of the Vocal Tract: The Vowel Model “, *Proceedings of the 19TH International Conference on the IEEE Engineering in Medicine and Biology Society* 1997, 1757... 1759, Chicago, USA.

[17] P. M. Tujal na M. N. Souza "Realimentação Visual para Oralização de Surdos: Módulo de Ditongos", *Anais do XIV Encontro Internacional de Audiologia* 1999, pp. 107, Rio de Janeiro.

[18] R. Johnson, 1981. The Picture Communication Symbols, Solana Beach, CA: Mayer-Johnson Co.

[19] R. Johnson, 1985. The Picture Communication Symbols, Book II, Solana Beach, CA: Mayer-Johnson.

[20] R. Rodman, D. McAllister, and D. Bitzer. Lip Synchronization as an Aid to the Hearing Disabled. Proceedings of the American Voice Input/Output Society 1997, 233 ... 248.

[21] S Maharaj, (1980). Pictogram Ideogram Communication. Regina, Canada. The George Reed Foundation for the Handicapped.

[22] S. P. Baptista. Estudo Preliminar de Animação Facial Realista para o Ensino de Deficientes Auditivos. Dissertação de mestrado, IC/UFF. Niterói, RJ, Fevereiro,2005

[23] S. Warren and A. Kaiser. Incidental language teaching. A critical review. Journal of Speech and Hearing Disorders 51 (1988), 291...299.