Capture, analysis and measurement of images of speech and smile dynamics

Vera Lúcia Cosendey[1], Stephanie Drummond[2], Jonas Capelli Junior[3]

Introduction: Dynamic analysis of smile and speech makes it easier to identify the features that define facial esthetics while allowing researchers to study different variables and observe the effects of aging. **Objective:** The aim of this study is to present a method for capturing, analyzing and measuring video images to support the study of speech and smile dynamics. **Methods:** Natural head positioning was standardized with the aid of a head holder (cephalostat). Image acquisition is performed with a video camera attached to a tripod, positioned at a fixed distance of 0.90 m. The subjects are trained to say out loud: "Tia Ema torcia pelo antigo time da Tchecoslováquia" and then to smile. The resulting images are fragmented and yielded four pictures that best represent a resting position, the least exposure of maxillary incisors, the greatest exposure of upper and lower incisors, and a posed smile. A freeware computer program called VIDEOMED was used to carry out measurements. **Conclusion:** The method presented in this study is an effective resource to record images captured during rest, speech and smile, thereby enabling a better understanding of changes in perioral soft tissues.

Keywords: Aging. Video recording. Facial expression.

Introdução: a análise dinâmica do sorriso e da fala facilita a identificação de características da estética facial, possibilita o estudo de diferentes variáveis e viabiliza a observação dos efeitos do envelhecimento. **Objetivo:** o objetivo desse trabalho é apresentar um método de captura, análise e medição de imagens por meio de vídeos, para o estudo da dinâmica da fala e do sorriso. **Métodos:** foi padronizado o posicionamento da cabeça em posição natural com o auxílio de um cefalostato. A captura de imagens foi realizada por uma câmera de vídeo, acoplada a um tripé, posicionada a uma distância fixa de 0,90m. Os indivíduos foram treinados a pronunciar a frase "Tia Ema torcia pelo antigo time da Tchecoslováquia" e depois sorrir. As imagens geradas foram fragmentadas de forma a gerar quatro quadros que melhor representassem o repouso, a menor exposição de incisivo superior, a maior exposição de incisivos superior e inferior, e o sorriso posado. Para realização das medidas, utilizou-se um programa específico, chamado VideoMed. **Conclusão:** o método apresentado torna possível um registro eficaz para captura de imagens durante o repouso, a fala e o sorriso, possibilitando um maior entendimento sobre as alterações dos tecidos moles peribucais.

Palavras-chave: Envelhecimento. Gravação em vídeo. Expressão facial.

[1] MSc in Orthodontics and Clinic Instructor, Department of Orthodontics, Rio de Janeiro State University.

[2] MSc Student, Department of Orthodontics, Rio de Janeiro State University.

[3] PhD in Orthodontics and Full Professor, Department of Orthodontics, Rio de Janeiro State University.

How to cite this article: Cosendey VL, Drummond S, Capelli Junior J. Capture, analysis and measurement of images of speech and smile dynamics. Dental Press J Orthod. 2012 Sept-Oct;17(5):151-6.

Submitted: May 02, 2011 - **Revised and accepted:** August 17, 2012

» The authors report no commercial, proprietary or financial interest in the products or companies described in this article.

» Patients displayed in this article previously approved the use of their facial and intraoral photographs.

Contact address: Stephanie Drummond
Universidade do Estado do Rio de Janeiro – Faculdade de Odontologia
Boulevard 28 de Setembro, 157/ sala 230 – Ortodontia
Zip code: 20.551-030 – Vila Isabel, Rio de Janeiro/RJ – Brazil

© 2012 Dental Press Journal of Orthodontics

INTRODUCTION

In orthodontics three methods are commonly suggested when studying the smile, namely: Qualitative, semi-quantitative and quantitative methods. The qualitative method is strictly visual. Typically, an orthodontist will look at a patient's smile and assess, for example, the smile line height. In the semi-quantitative method, analysis of the smile is performed by means of photographs, and in the quantitative method smile line height is determined with the aid of instruments. Measurements range from the simplest to the most sophisticated approaches.[1]

Capturing the smile image through photography presents certain difficulties: Photograph standardization is difficult due to differences in camera positioning, control of the distance to the focal point of the patient, head angle and the clinician's inability to capture the social smile twice in different photographic sessions.[2]

Image capture with a camera and video edition for subsequent analysis with computer programs seems to be a highly efficient method in the dynamic analysis of speech and smile. Ackerman et al[2,3,4] spearheaded the development of this technique. According to Maulik and Nanda,[5] videos allow researchers to select frames, increasing accuracy when choosing the images that more faithfully depict posed smiles and incisors exposure during speech while concurrently enabling observation of the patient in conversation. This also facilitates the identification of strengths and weaknesses in facial esthetics, allowing observation of the effects of aging on perioral soft tissues.[4]

Exposure of anterior teeth is not the same during speech vs. smile. This is evidenced in the rest, speech and posed smile video clips. A reasonable camcorder can record thirty frames per second, producing a five-second video clip in a total of about one hundred and fifty frames.[2]

This study aims to present a method to enable the capture, analysis and measurement of images through videos clips as a foundation for studying the dynamics of speech and smile.

MATERIAL AND METHODS

Image capture

Natural head position should be chosen, and for this reason a head holder (cephalostat) was used as the gold standard to stabilize the head of each individual. In this case, ear positioners restricted excessive lateral movements and the nasion positioner limited vertical movements (Fig 1).

Image capture can be performed by any video camera that uses digital tape MINI-DV format (for example: Sony model DCR-HC15 VTSC). The camera was attached to a tripod and positioned at a fixed distance in a straight line of 0.90 m between the patient's face and the camera lens. A second tripod should be positioned with an acrylic plate having a millimeter marker, of known dimension (30 mm x 50 mm) positioned flush to the patient's lips and orthogonally to the camera, for later image calibration in the computer program. This marker must fully cover the lips of the individual, and the upper portion of the plate should be parallel to the ground. Both tripods have

Figure 1 - Image capture: **A**) Positioning participant in head holder (cephalostat) and capturing image with acrylic calibration plate, **B**) capturing posed smile, **C**) acrylic plate used for calibration of image with bubble level.

to have a bubble level indicator to ensure parallelism of the camera and the acrylic plate with the floor. An additional bubble level was then positioned on the acrylic plate, in order to ensure greater accuracy in its positioning in the horizontal plane (Fig 1).

After leveling, the bubble level can be removed, once the tripod has been properly adjusted. The camera must be raised to the level of the lower face, with the lens perpendicular to the ground. The captured image will be that related to the lower face, so that the mouth of the subject is in the center of the camera's LCD display. Due to variations in facial heights between individuals, the resulting images may suffer differences in the framework.

To standardize the analysis of exposure of the teeth and soft tissues at different time intervals, the pronunciation of the same phrase can be employed. The sentence in Portuguese: *"Tia Ema torcia pelo antigo time da Tchecoslováquia"* followed by a smile was created with the guidance of a speech therapist who translated it phonetically from its original in English: "*Chelsea eats cheesecake on the cheasapeake*" created by Ackerman and Ackerman[2] in 2002 for capturing the greatest exposure of the incisor teeth during speech. According to Morley and Eubank,[6] enunciation of the phoneme "M" is used to obtain the exposure of the incisor teeth at rest. This phoneme was therefore added to record the lowest exposure of the incisors.

Shooting then begun, with a light source and diffuser being used as indirect lighting, with the calibrator positioned in front of the mouth for further calibration. This calibrator was then removed to visualize the pronunciation of the sentence previously trained, beginning at rest, and ending the session with a smile.

Video editing

The videos obtained were transferred to a computer, using Adobe Premiere Pro 2.0 software (Adobe Systems Incorporated, USA), in order to generate files in AVI format. The resulting videos have on average 12 seconds duration with an average 47 MB file size, in a total of about 360 picture frames per video. These videos were analyzed and split up at rest, during speech and smile, in order to produce the four static frames (corresponding to a photograph) that best represented a resting position, the least exposure of maxillary incisors, the greatest exposure of maxillary and mandibular incisors, and a posed smile.

The first frame selected was the image of a resting position, where the length of the lip and the height of the commissures were measured. Following the video sequence, the next frame showed the end of the utterance of the syllable "ma" in the word Ema; where the amount of maxillary incisor exposure, considered as the lowest exposure in speech was measured. On enunciation of the syllable "tche" in Tchecoslováquia, maxillary incisor exposure considered as the greatest exposure during speech was measured, as well as the amount of mandibular incisor exposure, if it ever came into view. As regards to the posed smile, measurements were made of the maximum maxillary incisor exposure and gingival exposure (Fig 2).

To ensure an accurate choice of picture frames it was necessary to carefully observe all the video fragments of each individual, so that only those that best represented each particular frame could be selected. Figure 3 shows nine different frames extracted from the same video, where frame number 6 best represents the ending utterance of the phoneme "tche."

Measurements in the selected frames were performed with the aid of a specific freeware program for measuring distance and area on video, called VIDEOMED 1-16.9.2002 ALPHA (version PAEDD) produced in the Multimedia Laboratory of the Electronic Computer Center of the Federal University of Rio de Janeiro. This program allowed researchers to view the images and hear the speech of the subjects, which facilitated the selection of frames corresponding to each phoneme. To use VIDEOMED, it was executed initially the calibration of the image (obtained with the calibration plate positioned in front of the area to be measured) with linear correction factor for X and Y, thus enabling measurements with the lowest error coefficient possible (Fig 4). Using a cursor and a lens that magnified the view of the marked points, it was possible to carry out specific measurements in each selected frame.

Image analysis

Direct measurements were made of each individual's cervical incisal height of the maxillary right central incisor and mandibular right central incisor with a digital caliper. This measure was obtained by

Figure 2 - Selection of four picture frames: **A**) At rest; **B**) uttering syllable "ma" – considered the least exposure of maxillary incisors during speech; **C**) uttering of syllable "tche" – considered as the moment of greatest exposure of maxillary and mandibular incisors; **D**) posed smile.

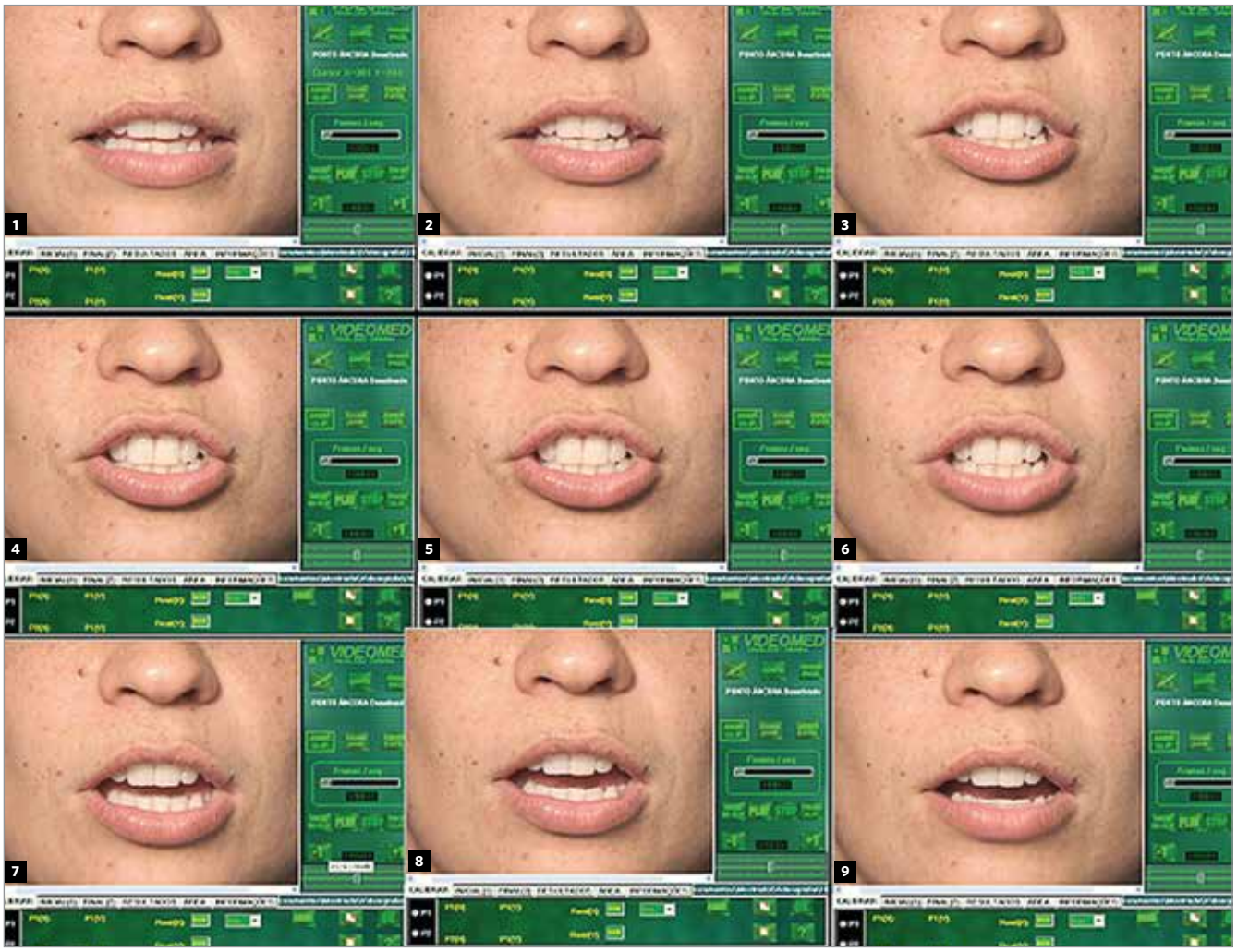

Figure 3 - Different pictures frames showing utterance of syllable "tche". Frame no. 6 best represents utterance of this phoneme.

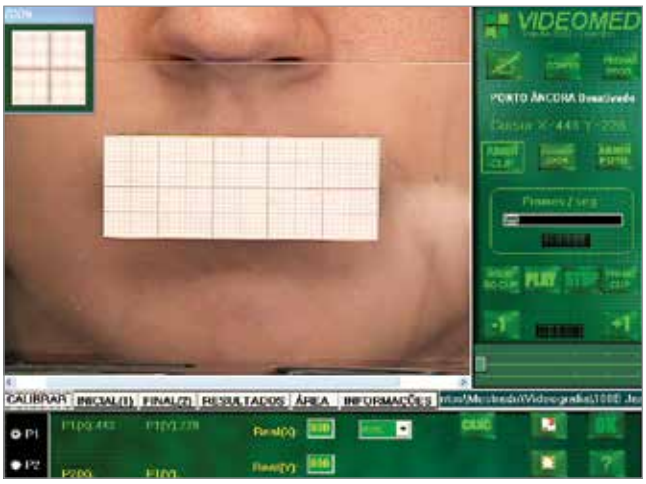

Figure 4 - Image frame captured for calibration on computer and subsequent measurement using VIDEOMED software.

the distance between the incisal edge and the neck of the tooth in question, in the direction of its long axis. In each frame analyzed, specific measures on the teeth and soft tissue could be assessed. Noteworthy among these were:

A) Lip length (LL): Measured in millimeters from the base of the nose, on the midline, as far as the lowest portion of the upper lip vermillion, on the philtral column, at rest. This imaginary line must pass through the arch of the cupid.

B) Lip commissure height (LC) and (RC): Measured from a vertical line tangent to the external commissures and perpendicular to a horizontal line passing through the lower portion of the bases of the nose wings.

C) Least exposure of maxillary incisors during speech (EMA): Measurement of the cervical incisal length of the maxillary right central incisor exposed at the end point of utterance of the syllable "ma," in the word Ema.

D) Maxillary incisor exposure during speech (TCHE): Measurement of the cervical incisal length of the maxillary right central incisor exposed during the utterance of the syllable "tche," in the word Tchecoslováquia.

E) Mandibular incisor exposure during speech (TCHE lower): Measurement of the cervical incisal length of the mandibular right central incisor exposed during the utterance of the syllable "tche," in the word Tchecoslováquia.

F) Maximum exposure of maxillary incisor during posed smile (SMILE): Measurement of the cervical incisal length of the maxillary right central incisor exposed during posed smile.

G) Gingival exposure during smile (GE): Measurement of the amount of gum exposed during posed smile: Distance between the lower edge of the upper lip and the gingival margin of the maxillary right central incisor. Exposure of a strip of gum above the maxillary right central incisor was regarded as a positive value. The value was considered zero when the lower edge of the upper lip was leveled with the gingival margin of the incisor. It was considered negative, when one cannot view the total cervical incisal length of the maxillary right central incisor, which was calculated by subtracting the measurement of incisor exposure during smile from its total length.

To determine the reliability and validity of the method presented, 124 subjects were randomly selected at the School of Dentistry of Rio de Janeiro State University to voluntary participate in the shooting in question.

The intra-class correlation test regarding the reproducibility of the operator showed a high correlation (1.0) considering the level of significance of 5%. The reliability of the method was evaluated measuring two times each variable in the subjects investigated, with an interval of 1 week. The intra-class correlation test was used for this assessment, obtaining also for all variables a high correlation coefficient considering the level of significance of 5% (Table 1).

Table 1 - Intra-class correlation coefficient between the first and second measurement of the variables in the subjects investigated.

Variables	Intra-class correlation
EMA 1 X EMA 2	0.991
TCHE 1 X TCHE 2	0.998
SMILE 1 X SMILE 2	0.989
TCHE low 1 X TCHE low 2	0.974
LL 1 X LL 2	0.975
LC 1 X LC 2	0.981
RC 1 X RC 2	0.962
GE 1 X GE 2	0.996

© 2012 Dental Press Journal of Orthodontics

DISCUSSION

Incisor and gingival exposure, as well as the lip shape are significantly different during speech vs. smile. These differences can be observed when evaluating the smile images and the pronunciation of certain phonemes.[7] The few studies that have investigated these factors report that these methods are reliable.[1]

Studies of this nature pose difficulties owing to the labiodental characteristics, facial mobility and the complexity involved in acquiring images that represent such characteristics in each patient evaluated, with faithfulness and reproducibility, and which can be repeated at different time intervals.[2]

In this method, image capture was accomplished by filming with a digital camcorder, to ensure the recording of more accurate observations of facial dynamics during a conversation. It was used a head holder to standardize head positioning in natural and orthogonal position to the camera, in order to reduce the variation in this position, which could alter the angle of observation and thus the analysis of the arch of the smile, the gingival margin, incisors length, axial tilt, in agreement with other studies.[2,8,9]

The development of a method capable of capturing, analyzing and measuring the recorded images in a reliable and relatively simple manner, was made possible thanks to the availability of a program developed at the Multimedia Electronic Computer Center from the Federal University of Rio de Janeiro. This method made it possible to observe the images of dynamics facial in the form of a video clip, split them frame by frame and select the ones that best represent the variable being examined.

Due to the great difficulty in capturing a reproducible smile to analyze labiodental characteristics, this study used what is called a posed or social smile, which is considered more static and therefore reproducible.[10,11,3]

A standardized videography provides the clinical orthodontist a greater number of images for selection of labiodental relationship parameters. Due to likely variations in the smile of adolescents, over time, photographs are rendered inadequate for evaluating treatment effects or changes caused by aging.[7]

CONCLUSION

This method makes possible an effective recording that supports image capture during rest, speech and smile, while allowing the analysis and measurement of different variables. Information gleaned from these video clips can afford a deeper understanding of the changes in perioral soft tissues, contributing to the implementation of such knowledge in the search for more effective orthodontic treatment results.

REFERENCES

1. Van der Geld PAAM, Oosterveld P, Van Waas MAJ, Kuijpers-Jagtman AM. Digital videographic measurement of tooth display and lip position in smiling and speech: Reliability and clinical application. Am J Orthod Dentofacial Orthop. 2007 Mar;131(3):301.e1-301.e8.
2. Ackerman MB, Ackerman JL. Smile analysis and design in the digital era. J Clin Orthod. 2002 Apr;36(4):221-36.
3. Ackerman JL, Ackerman MB, Bresinger CM, Landis Jr. A morphometric analysis of the posed smile. Clin Orthod Res. 1998 Aug;1(1):2-11.
4. Sarver DM, Ackerman MB. Dynamic smile visualization and quantification: Part 1. Evolution of the concept and dynamic records for smile capture. Am J Orthod Dentofacial Orthop. 2003 Jul; 124(1):4-12.
5. Maulik C, Nanda R. Dynamic smile analysis in young adults. Am J Orthod Dentofacial Orthop. 2007 Sep;132(3):307-315.
6. Morley J, Eubank J. Macroesthetic elements of smile design. J Am Dent Assoc. 2001 Jan; 132(1):39-45.
7. Ackerman MB, Brensinger C, Landis JR. An evaluation of dynamic lip-tooth characteristics during speech and smile in adolescents. Angle Orthod. 2004 Feb;74(1):43-50.
8. Zachrisson BU. Esthetic factors involved in anterior tooth display and the smile: vertical dimension. J Clin Orthod. 1998 Sep;32(9):432-45.
9. Wong NKC, Kassim AA, Foong KWC. Analysis of esthetic smiles by using computer vision techniques. Am J Orthod Dentofacial Orthop. 2005 Sep;128(3):404-11.
10. Hulsey CM. An esthetic evaluation of lip-teeth relationship in the smile. Am J Orthod. 1970 Feb;57(2):132-144.
11. Rigsbee OH, Sperry TP, Begole EA. The influence of facial animation on smile characteristics. Int J Orthod Orthognath Surg. 1988;3(4):233-239.

© 2012 Dental Press Journal of Orthodontics