ORIGINAL ARTICLE

H Popat
S Richmond
R Playle
D Marshall
PL Rosin
D Cosker

Three-dimensional motion analysis – an exploratory study. Part 2: Reproducibility of facial movement

Authors' affiliations:
H. Popat, S. Richmond, R. Playle, Department of Dental Health and Biological Sciences, University Dental Hospital, Cardiff, UK
D. Marshall, P. L. Rosin, Cardiff School of Computer Science, Cardiff University, Cardiff, UK
D. Cosker, Department of Computer Science, University of Bath, Bath, UK

Correspondence to:
Hashmat Popat
Department of Dental Health and Biological Sciences
University Dental Hospital
Cardiff, CF14 4XY, UK
E-mail: popath@cardiff.ac.uk

Dates:
Accepted 10 August 2008

To cite this article:
Popat H, Richmond S, Playle R, Marshall D, Rosin PL, Cosker D:
Three-dimensional motion analysis – an exploratory study. Part 2. Reproducibility of facial movement
Orthod Craniofac Res 2008;**11**:224–228

Structured Abstract

Authors – Popat H, Richmond S, Playle R, Marshall D, Rosin PL, Cosker D

Objectives – To investigate the reproducibility of using spoken word as a measure of facial movement.

Design – Experimental study.

Setting and Sample Population – Department of Dental Health and Biological Sciences, University Dental Hospital, Cardiff. 22 normal subjects.

Experimental Variable – Subjects were asked to say the word 'puppy' in a normal, relaxed manner twice within a 10-second time interval. The sequence was recorded using a non-invasive, three-dimensional motion analysis image capture system (3DMDface™ Dynamic System) at 48 frames per second.

Outcome Variable – Corresponding frames between the two utterances were aligned three-dimensionally using best fit on non-moveable points on the upper half of the face. Reproducibility was measured as the percentage point deviation between ± 0.5 mm between two corresponding frames.

Results – Mean intrasessional reproducibility (SD) for the group was recorded at 86.2% (5.8). The reproducibility ranged from a minimum of 66.8% to a maximum of 97.5%. When the utterance was split into its two separate viseme segments (/p/ /u/ /p/ and /p/ /y/), the second part of the utterance was seen to be more reproducible than the first. The male group were more reproducible than the female group.

Conclusion – Intrasessional reproducibility of the utterance 'puppy' shows high intra- and intersubject variability in this group of normal subjects and therefore further research needs to be conducted before being able to confidently use this word as a reproducible measure of facial movement.

Key words: facial expression; facial movement; imaging; motion analysis; reproducibility; speech; three-dimensional

Introduction

The use of a non-invasive three-dimensional motion analysis system was previously outlined in the assessment of facial movement (1). When assessing facial movement, use of the word 'puppy' to assess facial

movement was preferred to that of a standardized smile expression as when the word 'puppy' is spoken; it was shown that a greater percentage of stable points remain in the upper half of the face when compared with the smile expression. This allows for greater accuracy of surface superimpositions when assessing a sequence of facial movement. The second part of this study aims to investigate the reproducibility of using this spoken word as a measure of facial movement.

Materials and method

Intrasessional reproducibility of the utterance 'puppy' was tested in a group of 22 normal individuals (mean age: 24.3 years), comprising 9 males and 13 females. Subjects were asked to say the word 'puppy' twice in a normal and relaxed manner within a 10-sec time interval upon hearing an audio cue (clapperboard). The sequence was captured using the 3DMDface™ Dynamic System (3Q Technologies, Atlanta, GA, USA) and concurrent audio-visual recordings were made on a home digital video camera and microphone. The method has been described in detail in part 1 of this study (1).

Reproducibility of the utterance 'puppy' was carried out by first analyzing the audio sequence. Two segmentations, each containing a collection of mouth visemes (2) were determined in the audio sequence (Fig. 1). The first segment contains the viseme sequence /p/ /u/ /p/ and the second segment contains /p/ /y/. From the time line, precise location of the start and end point of the utterance and both segments was possible. The corresponding three-dimensional (3D) surface frames could then be extracted from the motion capture data and mapped for reproducibility. The reliability of segmenting the audio to extract the 3D files was tested by repeating this process for all subjects after a 3-month time interval by the same observer (HP).

Statistical analyses

Statistical analysis was carried out on a visual and descriptive basis using the predictive analytic software application SPSS 14.0 Release 14.0.0, 5 September 2005 (SPSS Inc., Chicago, IL, USA). Boxplots were used to show the median, inter-quartile range, outliers, and extreme cases of individual variables. 95% Confidence intervals (CI) were used to assess the reliability of the mean and to infer differences between tested variables.

Results

The extraction of the 3D files from the segmentation of the audio was seen to be very reproducible with 100% of frames matched to within ± 2 frames or 0.4 s. More importantly, 92.4% were matched exactly (Table 1). Mean intrasessional reproducibility (SD) for the utterance 'puppy' for the subject group as a whole was recorded at 86.2% (5.8). The 95% CI was recorded at 85.7% for the lower bound and 86.7% for the upper. As this range is narrow, we can assume there is little error in the mean. Reproducibility ranged from a minimum of 66.8% to a maximum of 97.5%. When the utterance was split into its two separate viseme segments (Fig. 2), the second part of the utterance was seen to have a higher mean, less variation (smaller range), and lower standard deviation than the first. Boxplots showed that

Table 1. **Reproducibility of extracting the 3D files from audio segmentation**

Frame match (no.)	Time match (s)	Percentage match	Cumulative percentage match
Exact	Exact	92.4	92.4
±1	± 0.2	5.4	97.8
±2	± 0.4	2.2	100.0

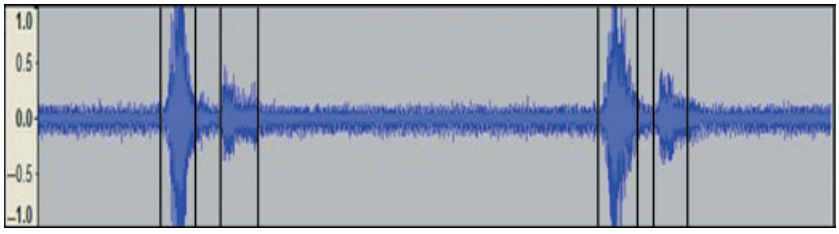

Fig. 1. Audio recording of two utterances of the word 'puppy' split into corresponding segments or visemes. Audio data obtained using digital audio editor, Audacity™ 1.2.6.

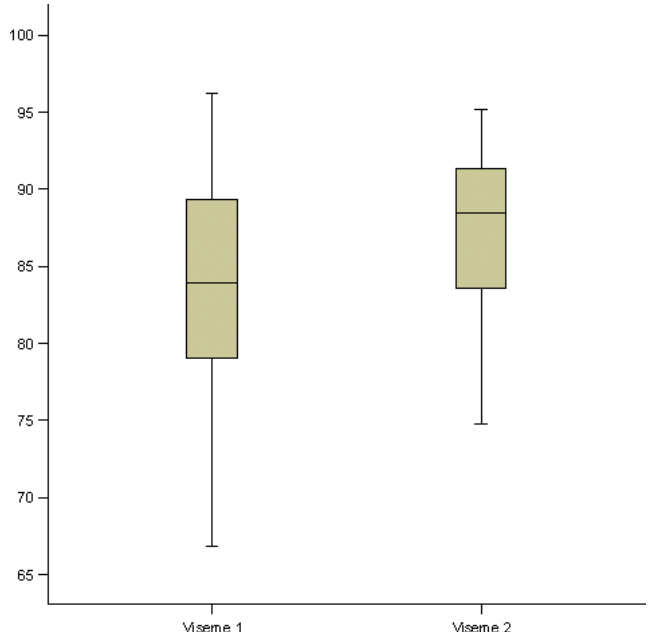

Fig. 2. Boxplots showing reproducibility of the two separate visemes of the two utterances of 'puppy'. Data obtained using predictive analytical software application SPSS 14.0 Release 14.0.0.

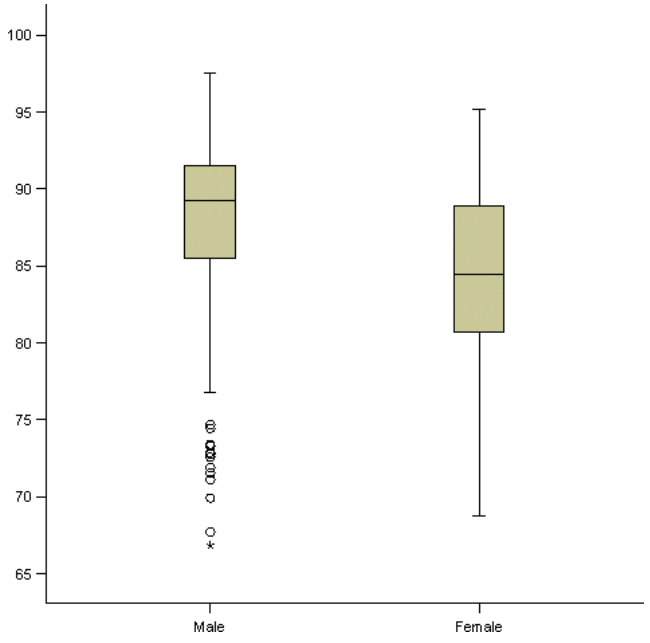

Fig. 3. Boxplots showing the reproducibility between males and females for the two utterances of 'puppy'. Data obtained using predictive analytical software application SPSS 14.0 Release 14.0.0.
*, Frames with values more than 3 box lengths from upper or lower edge of the box, the box length is the interquartile range.
○, Frames values between 1.5 and 3 box lengths from upper or lower edge of the box, the box length is the interquartile range.

males appeared to have slightly higher reproducibility than females (Fig. 3). The male group also had greater numbers of outliers or extreme cases below the lower interquartile reliability range than the female group.

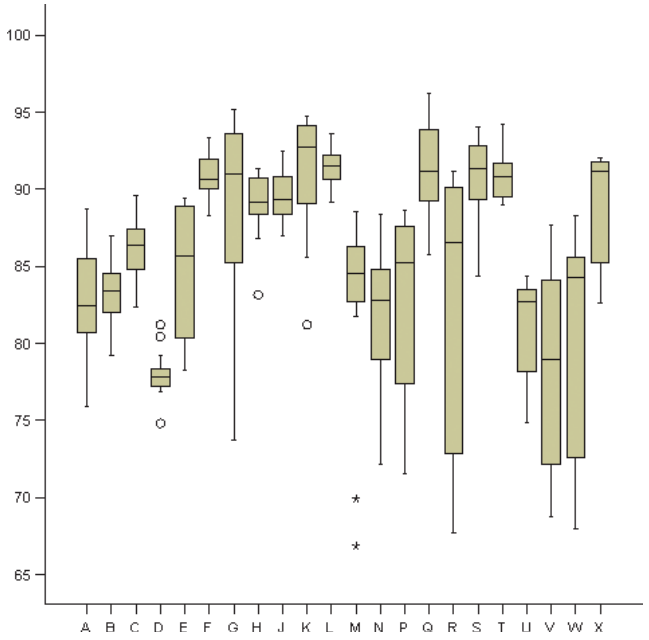

Fig. 4. Boxplots showing reproducibility ranges for the two utterances of 'puppy' by individual. Data obtained using predictive analytic software application SPSS 14.0 Release 14.0.0.
*, Frames with values more than 3 box lengths from upper or lower edge of the box, the box length is the interquartile range.
○, Frames values between 1.5 and 3 box lengths from upper or lower edge of the box, the box length is the interquartile range.

When reproducibility was assessed for each subject individually within the group, large inter- and intra-subject variability was seen (Fig. 4). Subjects with higher mean reproducibility for the two utterances of 'puppy' had lower associated standard deviations and conversely lower mean reproducibility scores were generally associated with higher standard deviations. Subject L was found to have the highest mean reproducibility of the group at 92.0% accompanied by the lowest standard deviation (1.5) showing consistently high reproducibility and very little variability between the two utterances. Subject D showed the lowest mean reproducibility of the group at 79.0% but with a relatively low standard deviation (2.3). This is in contrast to subject W who had a similar mean reproducibility of 81.7% but a much higher standard deviation (7.9) suggesting large variations between the two utterances for this individual.

Discussion

Reproducibility of head position is an important consideration in the assessment, diagnosis, and treatment planning of orthodontic and craniofacial procedures.

Natural head position has been the standard adopted for static facial pose (3, 4). With the advent of dynamic facial imaging, a reproducible measure of facial movement needs to be determined to allow accurate assessment of changes in facial movement over time and following treatment procedures. Previous work on facial expressions has shown that there is a hierarchy of expression reproducibility with the rest position being significantly more reproducible than lip purse, maximal smile, natural smile, or cheek puff (5). However, in the first part of this study (1) the standardized smile expression (6) was shown to lack sufficient stable surface areas in the upper half of the face for accurate frame superimposition and subsequent quantification of facial movement.

The word 'puppy' was the preferred method of assessing facial movement because of the upper half of the face remaining particularly stable during the sequence allowing for more accurate surface superimposition between frames from the sequence (1). Results in this part of the study showed that mean intrasessional reproducibility (SD) for the two utterances of 'puppy' was 86.2% (5.8). The male group had a higher mean reproducibility than the female group. However, as the 95% CIs of the mean reproducibility for these groups are narrow and relatively close, the difference seen could be attributed to the small sample size and at this level is unlikely to be of clinical significance. Another reason could simply be an inherent difference in phonation (Fig. 5).

The second viseme segment (/p/ /y/) appeared to be more reproducible than the first with a mean reproducibility (SD) of 87.6% (4.6) compared to 83.5% (6.8). The 95% CI's of the mean reproducibility for the two visemes are again narrow and close and a larger sample size and power calculation would be needed to confirm a significant difference. The difference could also, however, be because of the first viseme segment (/p/ /u/ /p/) being more plosive when spoken compared with the second segment (/p/ /y/). Another investigation which compares reproducibility between words with similar phonemes (for example, /cup/ to /pup/ and /bee/ to /py/ of /puppy/) could give more understanding on this area.

To determine whether facial movement during speech is reliable, a numeric value should be assigned to what is meant by reproducible. In other words, how much do repeated utterances of the same word have to vary before they are no longer considered reproducible? From a clinicians' point of view, this figure should be as low as possible. In this particular sample, although the mean intrasessional reproducibility (SD) was recorded at 86.2% (5.8) the intra- and intersubject reproducibility was extremely variable. Can we therefore apply this data confidently to use the utterance 'puppy' as a reliable measure of facial movement? On the basis of the results in this study, a reproducibility 'cut-off' of 85.7–86.7% could be used which represents the 95% CI for reproducibility of the group as a whole. If a subject was to be scanned, using dynamic imaging, uttering the word 'puppy' before and after a particular treatment and the resultant reproducibility measure falls below the lower 'cut-off' bound, it could be concluded that the treatment had an effect in changing the soft tissue surface. However, a larger sample size and power calculation would be required to test any statistical significance and a minimum magnitude of difference would need to be set to infer clinical significance.

It should also be noted that this study only tested intrasessional reproducibility of the word 'puppy'. By recording at two different time intervals and separating the capture sessions, calibration of the cameras can be repeated, natural head posture can be maintained and the memory effect of repeating the utterance can be avoided. The data collected would provide more robust

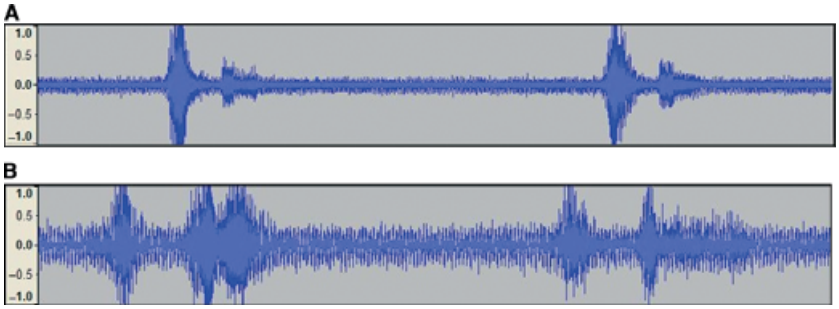

Fig. 5. (A) Audio frequency for the two utterance of 'puppy' by Subject L (Male). Audio data obtained using digital audio editor, Audacity™ 1.2.6. (B) Audio frequency for the two utterance of 'puppy' by Subject V (Female). Audio data obtained using digital audio editor, Audacity™ 1.2.6.

information of the reproducibility of facial movement. The department has recently acquired the upgraded motion capture system from 3DMD™ which boasts an increased frame capture rate and includes an integrated time and audio recording facility (1). This will significantly reduce the manual processing time and we are currently extending the work already completed to include recording at different time intervals and hope that these results will provide the basis for further research into this newly evolving field of facial imaging.

Conclusion

Intrasessional reproducibility of the utterance 'puppy' shows high intra- and intersubject variability in this group of normal subjects and therefore further research needs to be conducted before being able to confidently use this word as a reproducible measure of facial movement.

Clinical relevance

The use of the word 'puppy' to assess facial movement using three-dimensional motion analysis was described previously. With the advent of this imaging modality, a reproducible measure of facial movement is required to allow assessment of changes over time or after intervention. This study therefore aims to determine whether the word 'puppy' can be used as a reproducible measure of facial movement. Reproducibility of the word 'puppy' was highly variable for this group. The collected data provides a baseline for further research into what is a newly developing field of craniofacial imaging.

References

1. Popat H, Richmond S, Playle R, Marshall D, Cosker D, Rosin PL. Part 1: The assessment of facial movement using three-dimensional motion analysis – an exploratory study. *Orthod Craniofac Res* 2008;11:216–23.
2. Parke F, Waters K. *Computer Facial Animation*. Wellesley: AK Peters, Ltd.; 1996.
3. Kau CH, Knox J, Zhurov A, Richmond S. Validity and reliability of a portable 3D optical scanning device for field studies. In: Giuliani R, Galliani E, editors, Proceedings of the 2003 7th European Craniofacial Congress, Bologna, Italy; 2004. pp. 41–5.
4. Peng L, Cooke M. Fifteen-year reproducibility of natural head posture: a longitudinal study. *Am J Orthod* 1999;116:82–5.
5. Johnston D, Millett DT, Ayoub AF. Are facial expressions reproducible? *Cleft Palate Craniofac J* 2003;40:291–6.
6. Ekman P, Friesen WV. *Facial Action Coding System: A technique for the measurement of facial action*. Palo Alto, CA, USA: Consulting Psychologists Press; 1978.