ORIGINAL ARTICLE

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

Three-dimensional motion analysis – an exploratory study. Part 1: Assessment 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

Structured Abstract

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

Objectives – To objectively quantify facial movement in response to facial expression and spoken word.

Design – Experimental study.

Setting – Department of Dental Health and Biological Sciences, University Dental Hospital, Cardiff, UK.

Experimental Variable – Facial movement was assessed in response to a standardized smile expression and the utterance 'puppy'. The sequences were recorded using a non-invasive, three-dimensional motion analysis image capture system (3dMDface^{TM} Dynamic System) at 48 frames per second.

Outcome Variable – To quantify the facial movement, sequential frames of a sequence were aligned to the baseline/reference frame three-dimensionally using best fit on non-moveable points in the upper half of the face. Accuracy of the alignment process for each sequence was tested using the percentage of stable points (i.e. within ±0.5 mm) within the upper half of the face.

Results – Quantifiable changes in facial topology were seen during both the standardized smile expression and the utterance 'puppy'. The mean percentage of points (SD) that remained stable within the upper half of the face during the utterance 'puppy' was 88.8% (4.7). During the standardized smile expression, there were a much lower percentage of stable points in the upper half of the face with a mean (SD) of 60.9% (3.2).

Conclusion – The 3dMDface^{TM} Dynamic System allows objective, three-dimensional, non-invasive assessment of facial movement. The utterance 'puppy' is a more appropriate measure of facial movement when compared with the standardized smile expression.

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

Dates:
Accepted 9 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 1. Assessment of facial movement
Orthod Craniofac Res 2008;**11**:216–223

Copyright © 2008 The Authors.
Journal compilation © 2008 Blackwell Munksgaard

Introduction

Craniofacial assessment for diagnosis, treatment planning and outcome has traditionally relied on imaging techniques that provide a static image of the facial structure. Evaluation and quantification of facial movement however is becoming particularly important, for example in children undergoing facial surgery (cleft lip and palate), in the assessment of patients with motor deficits (facial nerve palsies) and in the evaluation of psychomotor function associated with depression or pain.

Until recently, the only tools available for the evaluation of facial function/movements were based on subjective scaling assessments (1–3) or two-dimensional measurements (4). Subjective assessments have the drawback that they are based on scales that are discontinuous and ambiguous (5) and although two-dimensional measurements are objective, studies have cast doubt on their validity (6–8).

This study represents experimental work that centres on a novel, non-invasive imaging system capable of three-dimensional, soft tissue image capture during facial movement. The methodology behind image capture is outlined and facial movement is assessed in response to facial expression and spoken word.

Fig. 1. Photographic representation of the 3dMDFace™ Dynamic System taken with Fujifilm F31fd digital camera.

Materials and methods

Equipment and apparatus

The scanning apparatus used in this study was the 3dMDface™ Dynamic System (3Q Technologies, Atlanta, GA, USA). It is a three-dimensional, non-invasive system capable of static image capture as well as dynamic capture with sequential framing of up to 18 s at 48 frames per second (fps) under full resolution. It captures images based on active stereophotogrammetry using a random infra-red speckle projection to capture both pattern-projected and non-pattern projected white-light images simultaneously.

The specifications of the system include six A202kc, 1 mega pixel area scan cameras with Kodak® KAI-1020 Interline Transfer Progressive Scan CCD Sensors (Basler Electric®, Highland, IL, USA) and six Pentax® (Pentax UK Limited, Slough, UK) C1641-M 16 mm 1.1:4 TV lenses.

Two custom made, infra-red projectors with 24-V halogen lamps, 50-mm lens and filters fitted enable projection of a random infra-red speckle pattern and data is processed using an Intel® Pentium® 4 Computer Processing Unit (CPU) 2.40 GHz, 0.98 RAM, 220 GB with six Seagate® SCSI 15k RPM 18 GB external hard disk drives. The imaging components are secured to a rig which enables standardization of image capture (Fig. 1).

Image acquisition and file output

The capture time for a static image is 1.5 ms. Ten output files are generated per static image. The three-dimensional geometry definition files (TSB and OBJ) are generated as one continuous point cloud produced from the two camera viewpoints. Distortion of the infra-red speckle as it projects onto a subject's face allows the CPU to generate three-dimensional co-ordinates or points of the face through complex algorithms. Several thousand points make up the cloud for a three-dimensional image of the face.

Images were handled, managed and processed using the reverse design/engineering software application, Geomagic Qualify 7© SR2 (Raindrop Geomagic Inc., Research Triangle Park, NC, USA).

Image processing

Following image acquisition, unwanted or erroneous data are first cropped or removed – this includes any hair, the ears, the neck up to the throat point and any visible clothing.

Images are then 'cleaned' whereby a sophisticated shape-correction algorithm within the Geomagic Qualify 7© SR2 software application is applied. It can remove dents, smooth cylindrical sections, or sharpen edges but is designed to change only the connection topology, and does not add points nor alter the current position of points.

For a sequence, an image must first be designated as the reference or baseline frame. Each remaining image is then designated as a test frame, and aligned on non-moveable points (the upper half of the face) within the three-dimensional point cloud. This is carried out individually for each test frame and the reference under a tolerance of 0.5 mm. This allows very little error in alignment of two or more objects. A colour deviation map of the differences (mm) between the two objects then enables quantification of the changes in facial movement over time.

Assessment of facial movement

Natural head posture was adopted for all image acquisitions. A standardized smile expression was used based on 'action unit 12' of the Facial Action Coding System (2). This involved the subject to perform a maximal smile without opening their mouth and without moving any other parts of the face, i.e. no variations other than lower face variations caused by the smile (so no eye widening or eyebrow movement). The word 'puppy', which is a bilabial speech posture involving articulation of both the upper and lower lips, was also used to assess facial movement. Image capture began with an audio cue (clapperboard) to signal the start of phonation to the subject and a concurrent audio–visual recording on a home digital video camera and microphone.

Analysis of the audio recording allowed the sequence of 3D geometry files to be related to a particular point in time. This was carried out on the digital audio editor, Audacity^{TM} 1.2.6 (Audacity Developer Team).

As the upper half of the face is used to superimpose frames, accuracy of this superimposition is important and therefore both the standardized smile expression and the utterance 'puppy' were investigated for this trait. The frames for each sequence were cropped at a coronal plane through subnasale, 90° to the mid-saggital plane and superimposed to the reference frame to detect any movement of the upper part of the face during the sequence. Stability of the upper part of the face was recorded as the percentage of points within ±0.5 mm.

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, interquartile 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

Assessment of facial movement

Figure 2 shows 3D colour deviation maps of three time points along the standardized smile expression sequence. Corresponding 3D surface images are shown adjacent. The maps show that from rest (T0), significant facial movement occurs during this expression. Strong positive changes (up to 5 mm) are seen in the region that corresponds with the zygomaticus muscles overlying the maxilla as they contract to bring the lips upwards. The regions corresponding to the upper lip and depressor anguli oris muscles flatten as they lift upwards showing negative changes (up to −4.3 mm).

Figure 3 shows three sample 3D colour deviation maps during the utterance 'puppy'. Corresponding 3D surface images are shown adjacent and the time point below. Two distinct segments are seen from the time line, each containing a collection of mouth postures or visemes; the first segment contains the viseme sequence /p/ /u/ /p/ and the second contains /p/ /y/. Fuller descriptions of the distinct mouth postures capable by a person have been published in classic texts (9). As the 'puppy' sequence progresses, negative changes (up to −5.0 mm) are seen affecting the regions corresponding to the orbicularis oris and mentalis muscles and positive changes (up to 5.0 mm) to the buccinator and floor of mouth.

During the utterance 'puppy', the mean percentage of points (SD) that remained stable (i.e. within ±0.5 mm) within the upper half of the face was recorded at 88.9% (4.7). The 95% CI was recorded at 86.8–90.8%. During the standardized smile expression, there were a much lower percentage of stable points in the upper half of the face with a mean (SD) recorded at 60.9% (3.2) and a 95% CI of 59.6–62.3% (Fig. 4).

Fig. 2. (A) Colour deviation map of time point T0 (baseline) of the standardized smile expression sequence with corresponding 3D surface image adjacent. $T = 0.02$ s. Scale in mm. 3D data obtained using reverse design/engineering software application, Geomagic Qualify 7© SR2. (B) Colour deviation map of time point T10 of the standardized smile expression sequence against the baseline (T0) with corresponding 3D surface image adjacent. $T = 0.02$ s. Scale in mm. 3D data obtained using reverse design/engineering software application, Geomagic Qualify 7© SR2. (C) Colour deviation map of time point T22 of the standardized smile expression sequence against the baseline (T0) with corresponding 3D surface image adjacent. $T = 0.02$ s. Scale in mm. 3D data obtained using reverse design/engineering software application, Geomagic Qualify 7© SR2.

The 95% CI of the mean of difference in the points that remained stable in the upper half of the face between the two groups was 25.5–30.2%. The magnitude of the difference in the 95% CI's implies that there may be a difference in the accuracy of using the upper part of the face as a stable reference between the two sequences.

Discussion

Materials and apparatus

Recent studies have begun to describe three-dimensional techniques capable of objective measurements of facial movements and therefore provide a means of evaluating soft tissue function (6–8, 10–14). Using commercially available video-based tracking systems or in-house designed systems, facial movement is assessed by tracking markers placed on various facial landmarks.

Other studies have also implemented videotaping to record facial movement but without the use of facial markers. Following digitization of the video recording, absolute pixel counts from sequential frames can be subtracted from the reference or baseline frame to give computer generated x–y plots of relative facial movement (15). Results have shown that disease-specific profiles can be constructed using computer methods that are not possible with subjective grading systems highlighting the importance of the development of computerized objective systems capable of facial movement analysis (16).

There are currently two commercially available three-dimensional motion analysis systems available; the 3dMDface™ Dynamic System (17) used in this study which is based on active stereophotogrammetry and the 4D Capture System from DI3D™ (Dimensional Imaging, Glasgow, UK) (18) based on passive stereophotogrammetry. The specification for the standard 4D system from DI3D™ is a three camera, single-pod comprising two 2 mega pixel greyscale cameras and one colour camera functioning at 30 fps. Their latest high-performance 4D system is also a single-pod with two 4 mega pixel greyscale cameras

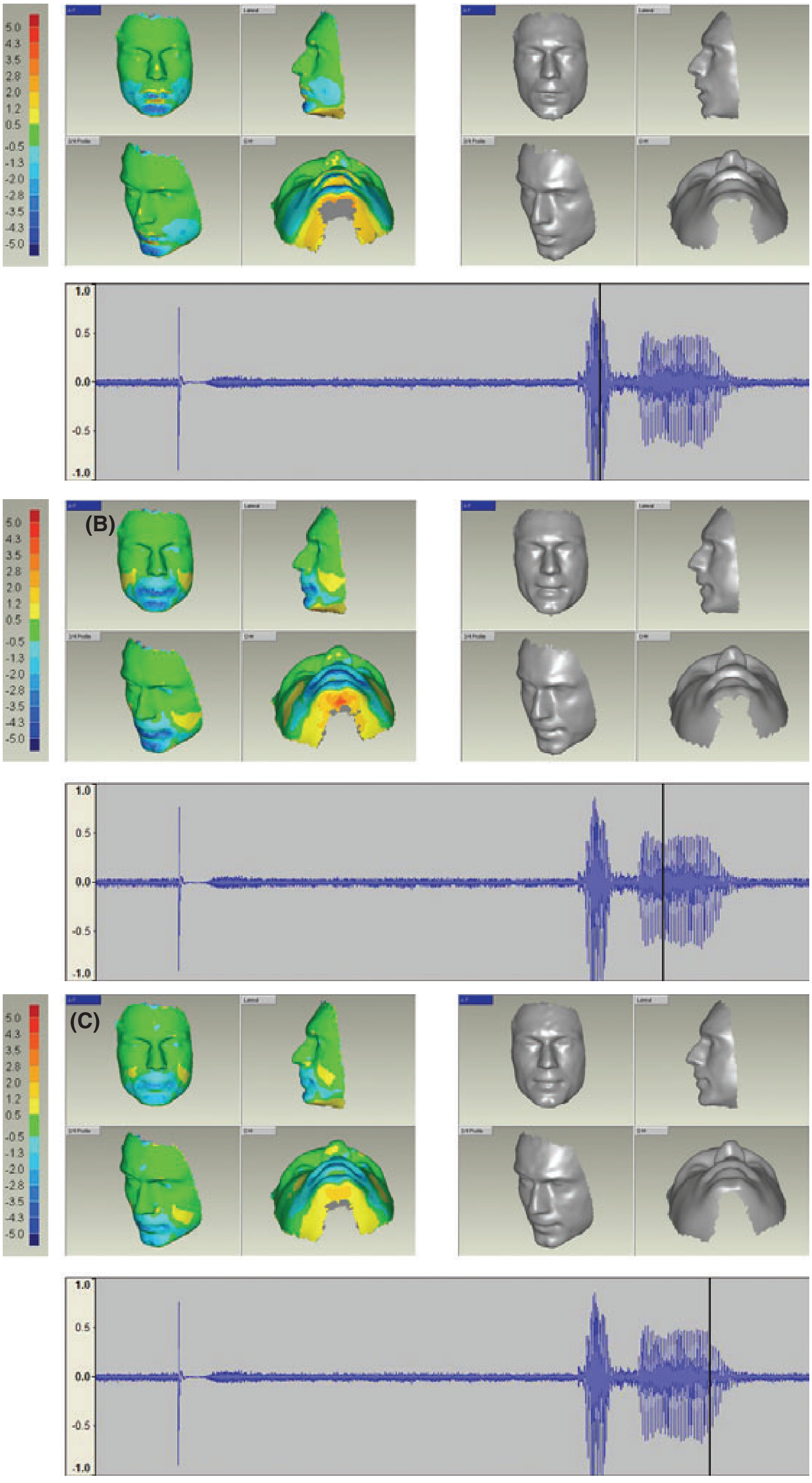

Fig. 3. (A) Colour deviation map of T5 against T0 (baseline) during the utterance of 'puppy'. Corresponding 3D surface image is shown adjacent and time line below. $T = 0.02$ s. Scale in mm. 3D data obtained using reverse design/engineering software application, Geomagic Qualify 7© SR2. Audio data obtained using digital audio editor, Audacity™ 1.2.6. (B) Colour deviation map of T20 against T0 (baseline) during the utterance of 'puppy'. Corresponding 3D surface image is shown adjacent and time line below. $T = 0.02$ s. Scale in mm. 3D data obtained using reverse design/engineering software application, Geomagic Qualify 7© SR2. Audio data obtained using digital audio editor, Audacity™ 1.2.6. (C) Colour deviation map of T33 against T0 (baseline) during the utterance of 'puppy'. Corresponding 3D surface image is shown adjacent and time line below. $T = 0.02$ s. Scale in mm. 3D data obtained using reverse design/engineering software application, Geomagic Qualify 7© SR2. Audio data obtained using digital audio editor, Audacity™ 1.2.6.

functioning up to 60 fps. Two pod versions of both DI3D™ systems are also available (19). The most recent 4D system from 3dMD™ is a six camera, two-pod system comprising of six 1.3 mega pixel cameras (four greyscale and two colour) functioning up to 60 fps and includes an integrated time and audio

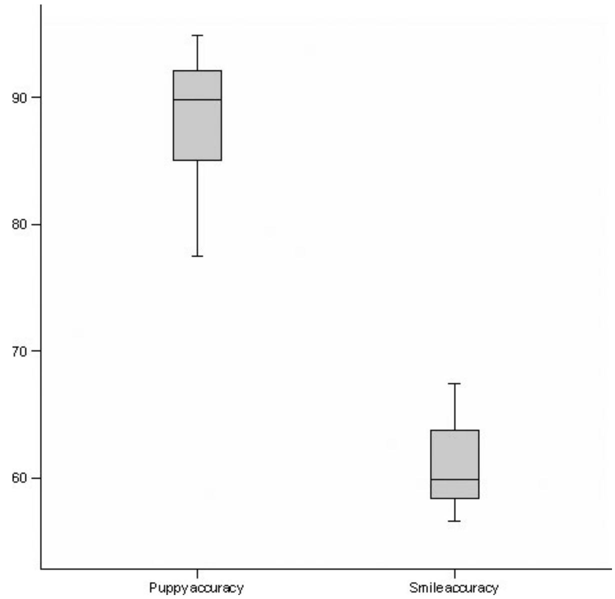

Fig. 4. Boxplots showing accuracy of superimposition of the utterance 'puppy' sequence and the standardized smile expression sequence. Data obtained using predictive analytic software application SPSS 14.0 Release 14.0.0.

recording facility. For the analysis of speech, these higher frame capture rates are preferential to adequately capture high-frequency distinct speech sounds – many of which have a duration of only 20 ms (20). A system with an integrated time and audio facility is a highly attractive feature which would save valuable manual processing hours and also make the process of image capture using the methodology described in this paper much simpler.

An important aspect of the 3dMDFace™ Dynamic System is that white light is used to capture colour texture images of the face simultaneously with the infra-red pattern produced three-dimensional data. There is inevitably a small infra-red component to the white light which may have the potential to drown some of the infra-red speckle pattern therefore compromising image capture. The authors have found the use of cold-illumination in the form of two Bowens Tri-lites© (Bowens International Limited, Essex, UK) minimizes this effect. The passive stereophotogrammetric DI3D™ system uses a stereo pair of low-noise monochrome cameras to capture three-dimensional shape and a single colour camera to capture the colour texture images using only regular white light projection which means there is no possibility of cross talk between visible and infra-red light compromising image capture. There are currently no published studies comparing the two systems for reliability or validity.

Image processing

The three-dimensional point cloud data acquired by the 3dMDFace™ Dynamic System was matched for individual frames using surface matching. Although this was carried out under a tolerance of 0.5 mm to minimize error, small errors in alignment of the forehead could lead to larger errors in the other regions of the face. A surface matching approach (21), which separately matches the eyes, the forehead and the nose regions, (i.e. the facial regions which are relatively less sensitive to movement during facial movement) using a modified ICP algorithm (22) can convey higher accuracy. Voxel-based algorithms are another documented method of matching three-dimensional data and are often used in matching data from MRI or CT scans to high precision (23). As the data set acquired in these image capture techniques is based on voxel representation, this can be considered the more appropriate matching system for these image capture techniques. The point cloud data provided by the 3dMDFace™ Dynamic system could be approximated by voxel coordinates but this would use more memory and therefore frames would also take longer to match especially if the surface to voxel conversion time is included.

Assessment of facial movement

A review of the literature showed that in previous studies, either speech or facial expressions have been used to assess facial movement (24, 25). As the majority of the facial surface moves during the standardized smile expression, the number of stable or 'still' points on the face decreases. Therefore alignment of sequential frames using surface matching may not be accurate as a small error in the region of the forehead could result in a larger error in the region of the chin. The mean (SD) number of points that remained stable in the upper part of the face during superimposition of frames was only 60.91% (3.24). Further methods of assessing facial movement were therefore investigated.

Lip function during speech has been another reported method of assessment of facial movement (14) as observation and evaluation of the visible articulators during speech can enable the clinician to determine functional normality or a potential disorder (26). The utterance 'puppy' was used in this study as a review of speech literature found this to be a bilabial speech posture involving movement of both lips (27). This would therefore give a good representation of facial movement in the lower third of the face while leaving stable or 'still' landmarks on the upper thirds of the face allowing accurate alignment of sequential frames to the reference or baseline. It was shown that almost 90% of points remain stable in the upper part of the face when superimposing frames during the utterance 'puppy'.

Conclusion

- The 3dMDFace™ Dynamic System allows objective three-dimensional, non-invasive assessment of facial movement.
- The utterance 'puppy' is a more appropriate measure of facial movement when compared with the standardized smile expression.

Clinical relevance

Objective analysis of facial movement forms an important consideration in the assessment and outcome of several medical disciplines. Of the little work published in this area, facial expressions and speech have been the main methods used to assess facial movement. This research centres on a non-invasive, three-dimensional imaging system capable of capturing facial movement at high frame capture rates. Facial movement was assessed in response to facial expression and spoken word. Surface matching of frames revealed quantifiable changes in facial topology during both sequences. Spoken word was deemed to be a more appropriate measure of facial movement using this methodology.

References

1. House J, Brackman D. Facial nerve grading system. *Otolaryngol Head Neck Surg* 1985;93:146–7.
2. Ekman P, Friesen WV. *Facial Action Coding System: A technique for the measurement of facial action*. Palo Alto, CA, USA: Consulting Psychologists Press; 1978.
3. Morrant D, Shaw W. Use of standardised video recordings to assess cleft surgery outcomes. *Cleft Palate Craniofac J* 1996;33:134–42.
4. Neely J, Jekel J, Cheung J. Variations in maximum amplitude of facial expressions between and within normal subjects. *Otolaryngol Head Neck Surg* 1994;110:60–3.
5. Sargent E, Fadhli O, Cohen R. Measurement of facial movement with computer software. *Arch Otolaryngol Head Neck Surg* 1998;124:313–8.
6. Trotman CA, Stohler CS, Johnston LE Jr. Measurement of facial soft tissue mobility in man. *Cleft Palate Craniofac J* 1998;35: 16–25.
7. Trotman CA, Faraway JJ, Silvester KT, Greenlee GM, Johnston LE Jr. Sensitivity of a method for the analysis of facial mobility. I. Vector of displacement. *Cleft Palate Craniofac J* 1998;35:132–41.
8. Trotman CA, Faraway JJ. Sensitivity of a method for the analysis of facial mobility. II. Interlandmark separation. *Cleft Palate Craniofac J* 1998;35:142–53.
9. Parke F, Waters K. *Computer Facial Animation*. Wellesley: AK Peters, Ltd.; 1996.
10. Weeden JC, Trotman CA, Faraway JJ. Three dimensional analysis of facial movement in normal adults: influence of sex and facial shape [erratum appears in Angle Orthod 2001 Dec;71(6):525]. *Angle Orthod* 2001;71:132–40.
11. Nooreyazdan M, Trotman CA, Faraway JJ. Modeling facial movement: II. A dynamic analysis of differences caused by orthognathic surgery [erratum appears in J Oral Maxillofac Surg. 2005 Jan;63(1):158]. *J Oral Maxillofac Surg* 2004;62:1380–6.
12. Trotman CA, Faraway JJ, Phillips C. Visual and statistical modeling of facial movement in patients with cleft lip and palate. *Cleft Palate Craniofac J* 2005;42:245–54.
13. Trotman CA, Gross MM, Moffatt K. Reliability of a three-dimensional method for measuring facial animation: a case report. *Angle Orthod* 1996;66:195–8.
14. Trotman CA, Phillips C, Faraway JJ, Ritter K. Association between subjective and objective measures of lip form and function: an exploratory analysis. *Cleft Palate Craniofac J* 2003;40:241–8.
15. Neely J, Cheung J, Wood M, Byers J, Rogerson A. Computerised quantitative dynamic analysis of facial motion in the paralysed synkinetic face. *Am J Otol* 1992;13:97–107.
16. Neely J, Joaquin A, Kohn L, Cheung J. Quantitative assessment of the variations within grades of facial paralysis. *Laryngoscope* 1996;106:438–42.
17. 3Q Technologies. 3dMD 3D Scanner. Available at: www.3dmd.com [cited 16th June 2008].
18. Dimensional Imaging. World Leading 3D and 4D Surface Imaging. Available at: www.di3d.com [cited 16th June 2008].
19. Urquhart CW, Green DS, Borland ED. 4D capture using passive stereo photogrammetry. In: Grau O, Hilton A, editors. *3rd European Conference on Visual Media Production, IEEE*, 2006. pp. 196.
20. Qifeng Z, Alwan A. On the use of variable frame rate analysis in speech recognition. *IEEE Int Conf Acoust Speech Signal Processing Proc.* 2000;3:1783–6.
21. Mian AS, Bennamoun AM, Owens R. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Trans Pattern Anal Mach Intell* 2007;29:1927–43.
22. Besl P, McKay N. A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Mach Intell* 1992;12:239–56.
23. Ciulla C, Deek FP. Performance assessment of an algorithm for the alignment of fMRI time series. *Brain Topogr* 2002;14:313–32.

24. Dawes KS, Kelly SW. An instrument for the non-invasive assessment of lip function during speech. *Med Eng Phys* 2005;27:523–35.
25. Johnston D, Millett DT, Ayoub AF. Are facial expressions reproducible? *Cleft Palate Craniofac J* 2003;40:291–6.
26. Morrison M, Rammage L. *The Management of Voice Disorders*. London: Chapel and Hall Medical; 1994.
27. Duffy J. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, 1st edn. St Louis, MO: Mosby Inc.; 1995.