Face to face: an interactive facial exercise system for stroke patients with facial weakness

P Breedon[1], P Logan[2], D Pearce[3], J Edmans[2], B Childs[4], R O'Brien[2]

[1]Design for Health & Wellbeing Research Group, Nottingham Trent University, Burton Street, Nottingham, UK

[2]Faculty of Medicine & Health Sciences, Queens Medical Centre, University of Nottingham, UK

[3]Barker Brettell LLP, 100 Hagley Rd, Birmingham, UK

[4]Maddison Product Design Ltd, Walnut Tree Yard, Lower Street, Fittleworth, UK

philip.breedon@ntu.ac.uk, pip.logan@nottingham.ac.uk, david.pearce@barkerbrettell.co.uk, judi.edmans@nottingham.ac.uk, ben.childs@maddison.co.uk, rebecca.obrien@nottingham.ac.uk,

[1]*www.ntu.ac.uk*, [2]*www.nottingham.ac.uk,* [3]*www.barkerbrettell.co.uk*, [4]*www.maddison.co.uk.*

ABSTRACT

Each year 152,000 people in the UK have a stroke. Almost all have an initial facial weakness. Many resolve in the first few days but it is estimated that 26,000 people experience some kind of long-term paralysis in their face. This may impact on their eating, drinking, speaking, facial expression, saliva management, self-image and confidence. A survey of 107 UK based clinicians found that routine treatment of facial weakness was provision of exercises with a written instruction sheet. The UK National Stroke Clinical Guidelines recommend that patients undertake 45 minutes of therapy per day, but anecdotal evidence suggests that patients have poor adherence to the exercises because they find them boring and there is no feedback to help them see a difference. A multidisciplinary team, which includes patients, researchers and therapists have produced a working prototype system to improve facial weakness. It is called Face to Face and includes a Kinect sensor, a small form PC and a monitor. Patients follow exercises given by a therapist on the screen; the system records and simultaneously gives feedback, with a facial recognition algorithm providing tracking data for each captured frame of the user's face. Results from our small clinical trial indicate that the system is more successful at getting patients to complete their exercises than using a mirror, patients liked it, and they said it had helped improve their facial symmetry. Therapists said Face to Face encouraged patients to exercise daily, they liked the fact that it could be individually programmed and could record how much the patient had exercised. Based on the initial project work and positive outcomes Face to Face aims to help patients practice their facial muscle exercises to speed their recovery, providing direct benefits in terms of costs and time, and offering patients significant improvements.

1. INTRODUCTION

The overall aim of the project was to test an effective, engaging and affordable interactive facial exercise system for use with people with facial weakness following stroke. This is to meet patient and clinical needs and lead to measurable improvements in clinical outcomes (Breedon, 2015, Breedon et al, 2014 & O'Brien, 2014).

Face to Face is a computer based system that enhances and automates the therapy provided by Speech and Language Therapists (SLTs). The system is initially aimed at stroke survivors suffering from facial paralysis, particularly with stroke patients suffering from swallowing and speech problems. The project development team also propose that the system may become a 'platform' system and be adapted and used more widely for a number of conditions and associated facial weakness.

A partnership between Nottingham Trent University, Nottingham CityCare Partnership, University of Nottingham, an industrial partner and service users within the UK have produced a system for use in the home, which uses the Microsoft Kinect games interface and a PC monitor to develop an interactive system to assist in the rehabilitation of patients with facial weakness. The aim is that patients will use this system to complete exercise programmes by copying on-screen exercises, and as they do this the system will monitor and *measure* the patient's face. Members of the University of Nottingham Stroke Research Partnership Group confirm the need for facial paralysis rehabilitation; currently there is a lack of appropriate and affordable alternatives in the market.

The system recognises facial expressions by tracking movement across the face and applying the recognised motion onto an onscreen representation of the user. This tool recognises differences between an undesirable one sided (unilateral) movement and a desired symmetrical movement across both sides of the face. This information is then conveyed via the monitor using the form of a graphical representation of the patients face. Using the representation as both a visual and auditory communicator, the system takes the patient through a series of exercises and indicates the degree of success with the exercise programme. An international patent application for the system was filed in 2015 (PCT/GB2015/053623).

1.1 Patient benefit

A key aspect of the system is the ability to visualise the improvements that the patient is making from repeated exercises. The system is simple to set up, with no need to attach any systems or markers to the face. The user interface permits small bursts of exercises to be performed throughout the day, rather than undertaking one long and tiring session. The overall aim of the initial project study was to test an effective, engaging and affordable interactive facial exercise system for use with people with facial weakness following stroke. This is to meet patient and clinical needs and lead to measurable improvements in clinical outcomes.

The system has the potential to be of benefit to patients by:

1. Improving fidelity in rehabilitative exercises prescribed for facial weakness.
2. Providing immediate feedback to patients on their progress, and in the future could provide a graded exercise programme.
3. Improving patient's facial weakness.
4. Including carers in the rehabilitation process (a Stroke Association recommendation).
5. Assisting patients in meeting the recommended target of 45 minutes daily therapy.
6. Reducing the need for patients to visit hospital for rehabilitation.

1.2 Clinical need

Approximately 16% of people suffering stroke have long lasting facial weakness (Svensson et al, 1992). A combination of physiotherapy, occupational therapy and speech therapy has been proven to be clinically effective in the treatment of those who have suffered facial weakness following a Stroke. However, whereas the National Clinical Guideline for Stroke (Intercollegiate Stroke Working Party, 2012) indicates that patients should receive 45 minutes of therapeutic treatment daily, the majority of patients receive far less than this. The Royal College of Physicians National Sentinel Stroke Clinical Audit (NSSA, 2010) states that only 33% of patients receive 45 minutes or more physiotherapy daily and only 18% of patients receive 45 minutes or more speech and language therapy daily.

It is clear from previous studies, and from involvement with partners at Nottingham University Hospital and the Stroke Association, that there are currently insufficient resources within the NHS to provide the necessary level of therapeutic care post Stroke. The most common rehabilitative exercises prescribed to facial weakness sufferers are either practising motions in front of a mirror or using an instructional DVD. These methods cannot be considered as engaging, due to their lack of feedback and quantifiable progression. Less commonly available is an electrical muscle stimulator designed for facial uses. This machine is made by the NHS electronics departments, and is in short supply, with recipients tending to get a maximum of six weeks to use the system before it is given to the next patient. The machine electrically stimulates the affected muscles to the point that they get exercise through contraction. It is reported to work well in the short term, but owing to short supply and the high price of working machines, along with a lot of training required on the part of the user, it is not currently considered a sustainable solution.

1.3 Initial system concepts

Initial system design concepts were based on the utilisation of the patient's own TV with a 'plug and play' 3D system attached. Over a development period of 18 months the preferred options were discussed and agreed with the patient and public involvement group (PPI), resulting in a PPI informed user interface and physical system design based on the Kinect sensor as shown in Figure 1.

Proc. 11th Intl Conf. Disability, Virtual Reality & Associated Technologies
Los Angeles, California, USA, 20–22 Sept. 2016
©2016 ICDVRAT; ISBN 978-0-7049-1547-3

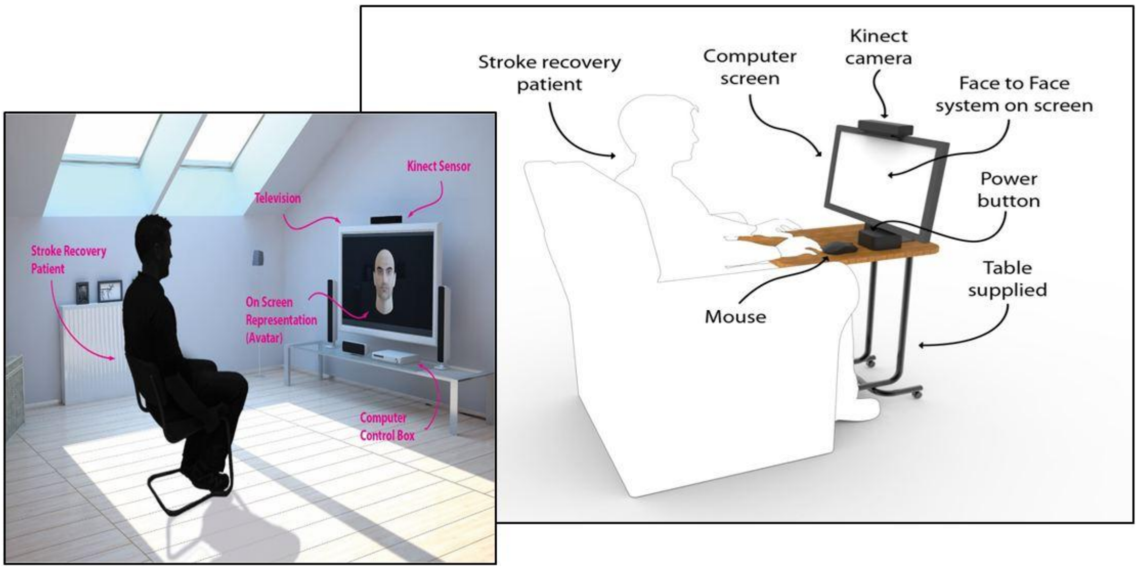

Figure 1. *Initial concept (inset) and PPI informed system design.*

The system that has been developed as a result of this project recognises facial expressions by tracking movement across the face and applying the recognised motion onto an onscreen representation of the user. How the user performs a series of facial exercises is assessed by the system and scored according to how well the user can do each of a defined set of expressions. The therapist is provided with a record of exercises as they were performed. New movements can be introduced automatically, with incrementally achievable goals throughout the week. The system will enable shorter consultations through the presentation of quantified improvement and empower the carer and the stroke survivors to take on some of the role of supervising the performance of the exercises.

It is proposed that the ‘home rehabilitation’ solution would help offset some of the problems associated with transfer from hospital to home care. The “Moving On” report (Chartered Society of Physiotherapy and Stroke Association, 2010) states that 48% of Stroke survivors believed that community and home-based therapy helped them become less long term reliant on carers.

2. SYSTEM OVERVIEW

Game based technology platforms are now widely used for physical rehabilitation of patients, recognised as a potential motivational tool for rehabilitation. New opportunities in rehabilitation have clearly arisen with emerging technologies of computer games and novel input sensors including 3D cameras. Hossein and Khademi (2014) provided a review on the technical and clinical impact of the Kinect on physical therapy and rehabilitation. Omelina et al (2012) examined serious games for physical rehabilitation and Webster and Celik (2014) provided a detailed review of elderly care and stroke rehabilitation systems utilising the Kinect. As 3D facial image processing for this research project provides clear additional pose mapping functionality, the benefits of processing depth information when compared to 2D vision systems was explored by González-Ortega et al (2014) in relation to Kinect based systems for cognitive rehabilitation exercise monitoring.

Currently the system has been developed utilising pre-existing facial recognition software known as Faceshift (available from www.faceshift.com and now owned by Apple). Image data is acquired from the Kinect and, when a face is detected, a model is generated of the user’s face. A profile can be stored for each user by capturing a pre-set series of expression, with an increased level of accuracy obtained by placing markers on a scanned face. This setup routine may be performed typically by a speech and language therapist (SLT), but should only need to be carried out once, following which the user can operate the system independently.

2.1 Pose matching

A short review on image processing was conducted by Samsudin and Sundaraj (2012) for facial paralysis and facial rehabilitation, identifying image processing as being widely utilised but with very limited applications in relation to facial rehabilitation. Further Zhang and Gao (2009) provided a critical review of face recognition techniques and the clear challenges particularly in relation to pose invariant face recognition, whilst Wang et al,

(2015) examined real time movement linked to an evaluation of skeletal pose tracking using the Microsoft Kinect first and second generation hardware.

Clearly various complications in computer vision require an assessment of pose objects in real-time with numerous reliable solutions available for pose estimation utilising feature-based 3D tracking. However difficulties can arise based on the need for fast and robust detection of known objects in view, for example the Kinect sensors used for this project were notably affected by ambient lighting conditions.

For the Face to Face system, once the facial recognition software is set up to track a user's face the facial recognition algorithm provides tracking data for each captured frame of a user's face. The tracking data relates to a number of facial gestures, together with a value associated with each gesture. In Faceshift, 48 channels of raw tracking data are provided per frame, each of which is associated with a number ranging between 0 and 1. The numbers, or scores, represent the strength of deformation on the 48 predefined gestural channels. For example, the gestural channel 'BrowsD_L' measures lowering of the left eyebrow, and 'LipsStretch_L' measures sideways movement of the left corner of the mouth. These numbers do not represent physical units, but are relative scores defined by upper and lower limits of each facial gesture. These scores are then used to analyse how well the user is performing a particular facial expression.

The speech and language therapist (SLT) defines a number of facial expressions, the facial expressions being specific to a user's rehabilitation programme. These facial expressions can each be defined as a combination of facial gestures that can be captured by the facial tracking algorithm, with each facial gesture having an associated target score. The overall aim of an exercise programme setup for a user is to encourage the user to perform the facial expressions and have these analysed and scored by the computer, using the facial recognition algorithm. To do this, a measure of how well the user is performing a particular facial expression needs to be defined. The mathematical space of tracking variables obtained from the facial recognition algorithm is mapped into a target space of facial gestures (or 'poses') for a particular facial expression. For each pose a measure termed a 'pose match strength' may be defined, for example as another number ranging between 0 and 1.

Pose match strength can be computed from the source tracking channels by defining a subset of the source channels as 'pose channels' that influence the pose match strength value for the facial expression being assessed. Source channels that are not included in a particular pose definition can either be ignored or treated as 'wildcards' for the match, for which any value will match. These represent channels that are irrelevant to the facial expression being exercised. As an example, mouth poses typically would not depend on eyebrow positions, resulting in the values for any eyebrow-related channels being discounted in an assessment of a mouth-related pose. Included pose channels may each define a target value or score for the pose, typically over the input range of between 0 and 1). This score is used to calculate an individual channel match value (again between 0 and 1). The range of input source channel values, i.e. from a minimum to a target value, is linearly mapped to an output pose channel match of 0 to 1. Beyond the target value the output pose channel match is capped at 1.

Figure 2 shows a pose tracking value over time, as defined by 14 successive samples, with a pose match peak of 0.8 and a peak value defined at 0.7 based on the development of the pose tracking value. During each 'out of pose' phase (while the system is looking for an 'in pose' event), whenever the instantaneous tracking value reaches a new highest value this 'pose match peak' value is recorded, and the time is recorded. A moving average is then maintained over the 'time since last peak', which will start at the peak and gradually decline as the instantaneous match strength diminishes.

Once a predefined time threshold is reached (which may, for example be around 0.5 seconds, and typically less than one second) without a new peak value being observed, an enter pose event is triggered and the current moving average is stored as a match value for the current 'active' pose. Conversely if a new instantaneous peak value is reached before the time threshold is reached, a new peak value is recorded, the timer is reset, and a new moving average measurement begins. This can happen several times before reaching a stable peak, particularly if the signal is noisy and/or the peak value is relatively low. The partially unsuccessful short-duration peaks before the final successful peak are analogous to 'false horizons' that are due to noise in the signal, whereas the moving average represents an adaptive threshold which continually adjusts to varying user ability and tracking response.

Each pose channel included in the pose definition may be marked as either a goal or a constraint. The output pose match strength is a relative score that can be calculated as the maximum of the set of all computed goal pose channel match values, limited by the minimum of the set of all constraint pose channel match values. This is similar to goal channels acting as logical 'OR' and constraint channels acting as logical 'AND'. The score for each facial gesture may be defined as a ratio of a score for the facial gesture provided by the facial recognition algorithm to the target score for the facial gesture. The score may range between 0 and 1, and be limited to 1 if the ratio exceeds 1.

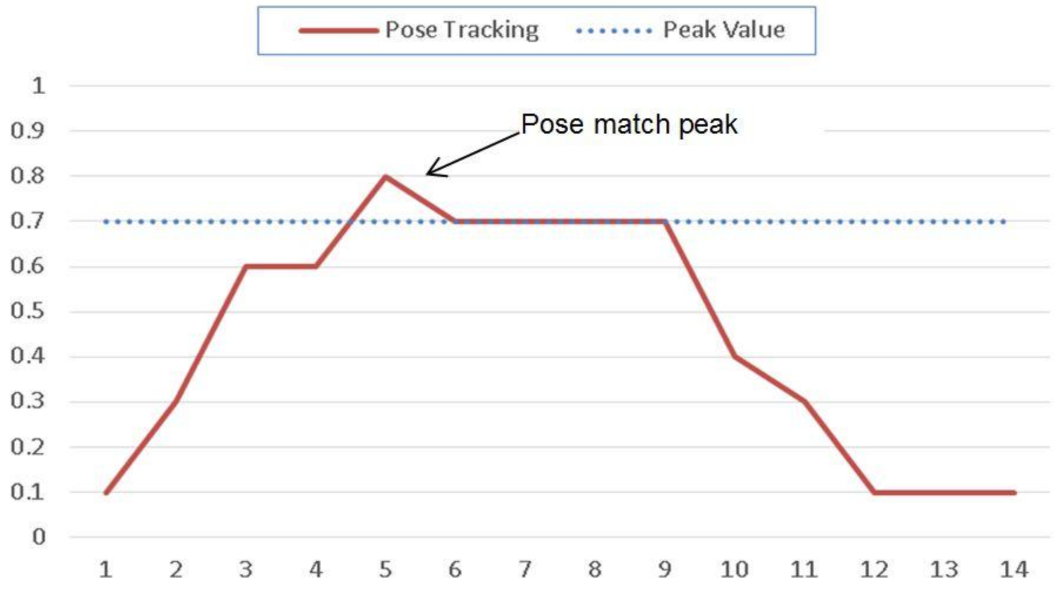

Figure 2. *Pose tracking over time (samples 1 to 14).*

2.2 *Step process*

Each of the exercise types relies on a predefined series of steps, these steps are illustrated in Figure 3. The process begins with an on-screen instruction phase with an acknowledgement when the number of repetitions has been completed. In between these two phases, the exercise involves looking for a particular target pose at any given time, which may even apply in the case of speed exercises which involve alternating between two poses. When an exercise target pose begins, the state is initially 'out of pose' and the system waits for the user to enter the pose. Once in pose, the system waits for the pose match strength to drop below a predetermined threshold value.

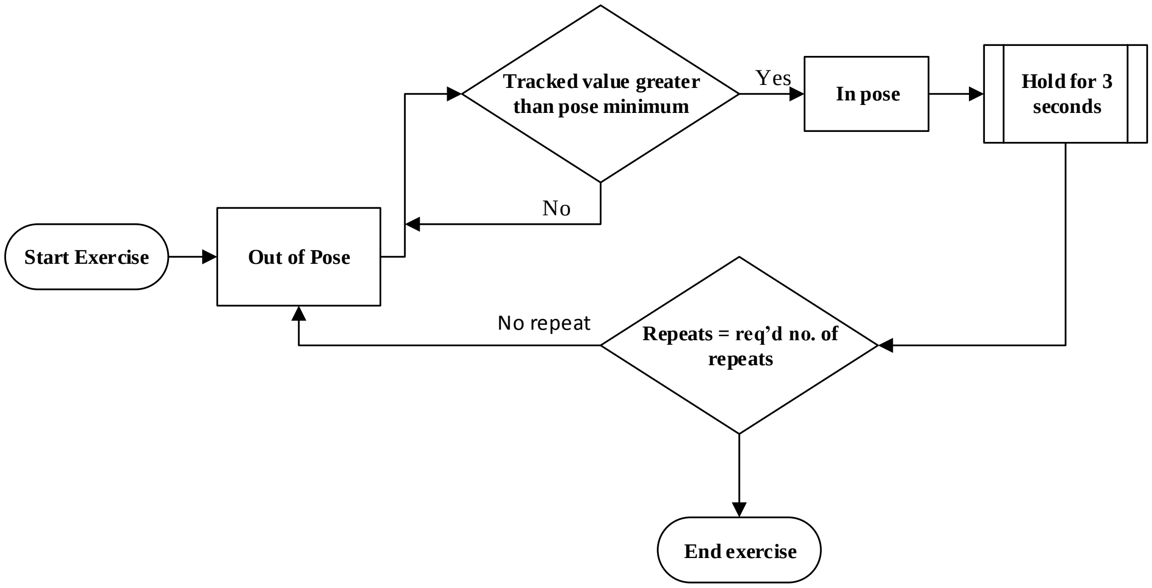

Figure 3. *Exercise steps.*

The events of entering and exiting a current target pose can trigger different responses depending on the exercise type. For example, speed exercises may switch to an alternate pose on exiting a first pose, whereas range exercises tend to set an 'anti-pose' during a relax phase which the user must enter (for example open mouth, followed by closed mouth) before returning to the original target pose for the next repetition (open mouth). Once the user is in pose the user is instructed to hold for a set period, and once the set period elapses the process can then repeat if required or complete. The series of steps involved in performing an exercise transforms a continuous range of pose match strengths into a series of discrete states and state change events. The user interface indicates the number of repetitions, which then triggers the next instructional video ("open your mouth..." / "...and relax"). A particular problem, however, arises when considering how to determine when a particular pose has been achieved, i.e. when the state should change from 'out of pose' to 'in pose'. This determination can be made on the basis of a continuous assessment of the score for a facial expression that is the subject of the exercise.

2.3 Pose example

The 'ee' pose may be defined by MouthSmile_L (left) and MouthSmile_R (right) as goal channels, each with a target value of 0.8, defining JawOpen as an inverted constraint channel with a target value of 0.5. The following two conditions then result:

Condition 1:

MouthSmile_L = 0.2 - goal match 0.2/0.8 = 0.25

MouthSmile_R = 0.6 - goal match 0.6/0.8 = 0.75

JawOpen = 0.3 - inverted constraint match = 1 (less than target 0.5, inverted, capped to 1)

Pose match strength = min (max (goal), min (constraint)) = min (0.75, 1) = 0.75

Condition 2:

MouthSmile_L = 0.2 goal match 0.2/0.8 = 0.25

MouthSmile_R = 0.6 - goal match 0.6/0.8 = 0.75

JawOpen = 0.8 inverted constraint match = (1 - 0.8) / (1 - 0.5) = 0.2/0.5 = 0.40

Pose match strength = min (max (goal), min(constraint)) = min (0.75, 0.40) = 0.40

This example could be interpreted as: an 'ee' pose requiring MouthSmile Left OR Right, AND jaw mostly closed. In the second scenario, the pose match strength is reduced to 0.4 because the jaw constraint is infringed, limiting the overall output value. In more general terms, a relative score for each identified facial gesture in a facial expression is calculated as a ratio of the score for each identified facial gesture to the corresponding target score for the facial gesture. The score for the facial expression can be calculated based on a maximum relative score for the facial gestures defining the facial expression. Other combinations of goals and constraints can also be determined, depending on the particular facial expression to be exercised and whether any particular constraints need to be considered. For example, a facial exercise regime may be made up from a set of nine different poses. These may be combined using 9 range exercises with 9 strength exercises and with 3 poses used in pairs in 'fast' and 'slow' versions to make up 6 speed exercises (AB, AC, BC x fast/slow), making a total of 24 exercises to be performed in the exercise regime. A therapist can then make a patient assessment, and simply enable or disable each exercise from this palette of 24 exercises to quickly construct a tailored exercise regime for that user. A screenshot pose is shown in Figure 4 and the patient progress screen in Figure 5.

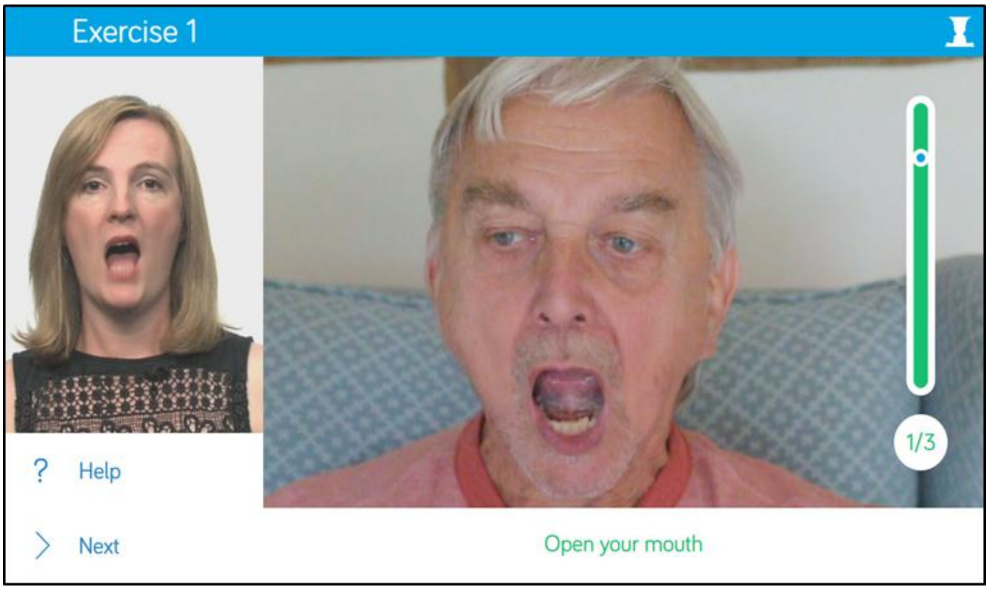

Figure 4. *Exercise pose screenshot with a graphical indication of the user's score (right) and exercise number (bottom right) overlaid on a picture of the user next to an image showing the pose to be performed. The pose indicated by a speech and language therapist (left).*

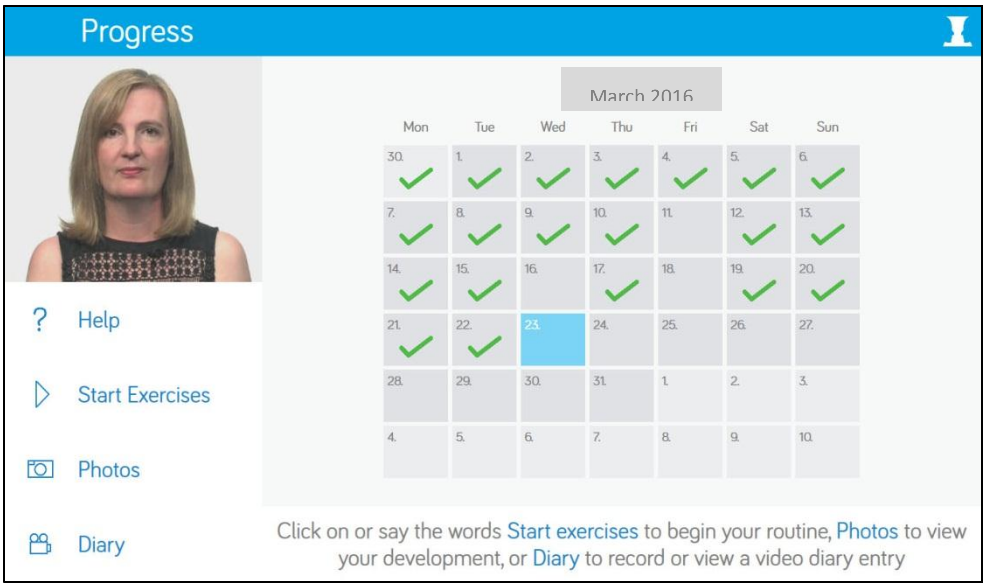

Figure 5. *System diary and progress screen.*

Looking forward, there is a growing number of technologies coming to market that will benefit and enhance this product. Of particular interest is the *new* 3D camera technology launched by Intel. These 3D cameras use the same principles as the Kinect camera but designed to be integrated into phones, tablets and laptops. Another important feature of the Intel cameras is that they have been specially designed to focus on facial recognition and tracking, this will provide an accurate and reliable facial tracking system for us to use. In contrast the Kinect system is designed for full body tracking, although the system can be used to track facial expressions it is not its primary function and therefore limits the effectiveness of the product and requires us to pay for 3rd party software licences.

In current practice, SLTs priorities are in assessing and planning the patient's therapy, and non-professional assistant practitioners (APs) are used on a day-to-day basis with the patient to monitor adherence to the program and assist the patient. We propose to extend the stored usage database to include a record of planned usage regimes for each patient, and to provide a visual comparison of actual usage against the planned regime in the web interface. This approach would allow an AP to interpret the live usage data in the web interface, identifying overuse or underuse of exercises without needing to involve an SLT in the first instance. The web-based interface would still be accessible to SLTs who could perform any further analysis they felt necessary, but they would now also be able to remotely adjust the exercise regime while the device was still in the patient's home. Further development of the product is required and to address other market opportunities, including:

- Improving the user interface to address shortcomings identified in trials and research;
- Improving the compactness and portability of the physical embodiment;
- Strengthening the robustness and functionality of the system (for example in low light conditions);
- Reducing unit cost;
- Allowing for a wider range of exercises, implying a more complex interface requirement to allow SLTs to set up systems for patients.

It is anticipated that this additional functionality would reduce the on going costs of deploying the final system within everyday healthcare practice, by reducing the required time from the SLT.

Acknowledgements: The authors also gratefully acknowledge the invaluable support and contributions to the project from: Ossie Newell MBE, The stroke service (PPI) group, Adam Russell and Paul Watts (both previously based at Nottingham Trent University, UK), Patrick Hall at Maddison and Claire Wragg at NTU. The Face to Face project was funded by the UK National Institute for Health Research Invention for Innovation (NIHR i4i) programme.

4. REFERENCES

Breedon P, Russell A, Logan P, Newell, O, O'Brien B., Edmans J, Baines D and Hall P, 2014, First for Stroke: using the Microsoft' Kinect as a facial paralysis stroke rehabilitation tool. *International Journal of Integrated Care (IJIC)*, *14*.

Breedon P, 2015, Article in conference, Face to Face: An interactive facial exercise system for stroke patients with facial, Talking Technology. Current Trends in Speech and Language Therapy, The Association of Speech and Language Therapists in Independent Practice (ASLTIP), London, UK.

Chartered Society of Physiotherapy and Stroke Association, 2010, Moving On: a vision for community based after stroke in England: The Stroke Association.

Edmans J, Breedon P, Childs B, Hall P, Newell O, O'Brien B, Logan P, 2016, Face to Face: An Exploratory Trial of a Facial Rehabilitation Tool for Facial Weakness after Stroke. College of Occupational Therapists Specialist Section – Neurological Practice newsletter, June (49):pp.40-47.

González-Ortega D, et al, 2014, A Kinect-based system for cognitive rehabilitation exercises monitoring. *Computer methods and programs in biomedicine* 113.2, pp.620-631.

Hossein HM, and Khademi M, 2014, A review on technical and clinical impact of Microsoft Kinect on physical therapy and rehabilitation. *Journal of Medical Engineering*.

Hu WL, Ross B and Nedzelski J, 2001, Reliability of the Sunnybrook Facial Grading System by novice users. *J Otolaryngol*; 30: pp.208–211.

Intercollegiate Stroke Working Party, 2012, National Clinical Guidelines for Stroke: Fourth edition, Royal College of Physicians, London. https://www.rcplondon.ac.uk/guidelines-policy/stroke-guidelines

NSSA, 2010, National Sentinel Stroke Audit, (1998 – 2010) https://www.rcplondon.ac.uk/projects/national-sentinel-stroke-audit-nssa

O'Brien R, 2014, Facial Weakness after Stroke, 2014, UK Stroke Forum, Harrogate, UK.

Omelina L, et al, 2012, Serious games for physical rehabilitation: designing highly configurable and adaptable games. *Proceedings of the 9th International Conference on Disability, Virtual Reality & Associated Technologies*.

PCT/GB2015/053623 International patent application (published May 2016).

Ross BG, Fradet G and Nedzelski JM, 1996, Development of a sensitive clinical facial grading system. *Otolaryngol Head Neck Surg*; 114(3): pp.380-386.

Samsudin WSW, and Sundaraj K, 2012, Image processing on facial paralysis for facial rehabilitation system: A review. *Control System, Computing and Engineering (ICCSCE), IEEE.*

Svensson BH, Christiansen LS and Jepsen E, 1992, Treatment of central facial nerve paralysis with electromyography biofeedback and taping of cheek. A controlled clinical trial. *Ugeskrift for Laeger* 154(50), pp.3593-3596.

Wang et al, 2015, Evaluation of pose tracking accuracy in the first and second generations of Microsoft Kinect. *Healthcare Informatics (ICHI), 2015 International Conference*. IEEE.

Webster D, and Ozkan C, 2014, Systematic review of Kinect applications in elderly care and stroke rehabilitation. *Journal of neuroengineering and rehabilitation* 11.1: 1.

Zhang Xiaozheng, and Yongsheng Gao, 2009, Face recognition across pose: A review *Pattern Recognition* 42.11: pp. 2876-2896.