Interspeech 2018
2-6 September 2018, Hyderabad, India

Automatic detection of orofacial impairment in stroke

Andrea Bandini[1], *Jordan R. Green*[2], *Brian D. Richburg*[2], *Yana Yunusova*[1,3,4],

[1]University Health Network: Toronto Rehabilitation Institute, Toronto, Canada
[2]MGH Institute of Health Professions, Boston, USA
[3]Brain Sciences, Sunnybrook Research Institute, Toronto, Canada
[4]Department of Speech-Language Pathology, University of Toronto, Canada
andrea.bandini@uhn.ca,{jgreen2, brichburg}@mghihp.edu,
yana.yunusova@utoronto.ca

Abstract

Stroke is a devastating condition that affects the ability of people to communicate through speech, leading to social isolation and poor quality of life. The quantitative evaluation of speech and orofacial movements is essential for assessing the impairment and identifying treatment targets. However, to our knowledge, a tool for the automatic orofacial assessment, which considers multiple aspects of orofacial impairment (e.g., range of motion in addition to asymmetry), has not been developed for this clinical population. In this work, we tested a video-based approach for the automatic orofacial assessment in stroke survivors, combining low-cost depth sensor and face alignment algorithms for extracting facial features. Twelve patients post-stroke and 11 control subjects were evaluated during speech and non-speech tasks. By using a small feature-set representing range of motion and asymmetry of face movements, it was possible to discriminate patients post-stroke from control subjects with high accuracy (87%). Further insights on the choice of the task and face alignment algorithm are provided, demonstrating that a non-parametric approach such as *SDM* can provide better results. Through this work we demonstrated the feasibility of an objective tool to support clinicians in the assessment of speech and orofacial impairment post-stroke.

Index Terms: Face tracking, stroke, face kinematics, facial palsy, asymmetry

1. Introduction

Stroke is among the most common causes of death and the leading cause of permanent disability in developed countries, with nearly 800,000 cases per year in the US and an economical annual burden estimated around $34 billion [1]. One of the most debilitating aspects of the disease is in relation to speaking abilities – more than 40% of patients experience motor speech disorders after stroke [2]. Stroke is also a common cause of facial palsy, affecting emotional expression and social aspects of communication [3]. Patients post-stroke (PS) are often socially isolated, unable to express their physical and emotional needs, resulting in diminished quality of life [4].

Objective assessment tools targeting orofacial function post-stroke are currently limited in number. This impedes the diagnosis of functional impairments and progress monitoring during recovery and rehabilitation. Although many clinical scales have been developed to grade the facial nerve function post-stroke [5]–[8], they may not be commonly accepted by the clinical community due to their relatively low reliability and cumbersome administration procedures [9]. The implementation of these scales is often incompatible with the fast pace of clinical practice, where objective and interpretable measures must be readily available to clinicians. Therefore, patients PS are rarely evaluated objectively, making it harder to monitor the process of recovery and rehabilitation as well.

To overcome these issues, many authors have proposed image- and video-based approaches to assess orofacial impairments automatically. He *et al.*, 2009 [10] used texture features (local binary pattern) to automatically predict the score of a clinical scale for grading the facial nerve function (House-Brackmann – HB scale [5]), obtaining 93.1% accuracy. In [11], the authors used an active appearance model to extract the distance between mouth corners across different facial expressions. Using this measure, they obtained 87% accuracy in predicting the HB grade. Another study [12] proposed an image subtraction method to identify peak movements during various tasks (e.g., eyebrow raising, eye closure, snarl, and wide smile), predicting the HB grade with 94% accuracy. Despite showing good classification results, the above studies did not report tests on patients with orofacial impairments due to stroke and performed the analysis on 2D static images that may result in reduced robustness against head rotation.

Schimmel *et al.*, 2011 [13] used an infrared (IR) 3D video system to quantify the upper and lower facial motor function in patients with hemispheric stroke. Although the system was able to detect changes in facial muscle function, the assessment could not be performed online. Another 3D video-based approach was proposed by Quan *et al.*, 2012 [14]. The authors used 3D scans of the face and measured facial asymmetry with the Euclidean distances between corresponding facial points on the original scan and its mirrored version. The analysis was also performed on 8 patients PS, but automatic detection of orofacial impairments was not performed. It appears, therefore, that an automated and real-time tool for orofacial assessment, which considers different aspects of orofacial impairments (e.g., range of motion in addition to asymmetry) has not been developed or used for this clinical population.

The increasing availability of accurate, low-cost depth sensors, along with the development of efficient face alignment algorithms provide a great opportunity to introduce novel and intelligent systems in clinical environments [15], [16]. Previous studies demonstrated the feasibility of 3D video-based approaches for the assessment of motor speech disorders in patients with Parkinson's disease and amyotrophic

10.21437/Interspeech.2018-2475

lateral sclerosis (ALS) [17]–[19]. In this work, we aim to expand those findings to the automatic assessment of orofacial impairment in patients PS, evaluating facial movements during speech and non-speech tasks. Specifically, the aims of this work are to 1) determine if a 3D video-based system is able to automatically discriminate patients with orofacial impairment due to stroke from healthy control (HC) subjects; 2) identify the best tasks for the classification, among those commonly used by clinicians during the oral motor examination; and 3) determine the influence of the face alignment step on the overall classification performance, to identify the best approach for the automatic extraction of facial features for clinical purposes.

2. Data Collection

2.1. Subjects and clinical judgment

Twelve patients PS (7 male, 5 female; age: 62.0 ± 14.5 years) and 11 HC subjects (7 male, 4 female; age: 55.8 ± 15.7 years) were recruited for the study. 10 patients had an ischemic stroke whereas, for 2 patients the stroke was hemorrhagic. All participants were native speakers of English and showed no evidence of cognitive impairment as assessed by the Montreal Cognitive Assessment (score ≥ 26) [20]. Two trained speech-language pathologists performed a clinical oral motor exam [21] on each participant, rating facial range of motion (*ROM*) and asymmetry, among other measures (e.g., movement speed, variability and fatigue) during the speech and non-speech tasks recorded for the experiments (Sec. 2.2). Each measure was assessed with a score between 1 (normal function) and 5 (severe dysfunction). The average of the scores provided by the two raters were used to calculate the score for each participant. Patients PS had higher *ROM* and asymmetry scores as judged by the clinicians (Table 1), confirming the presence of orofacial impairment. The agreement between the two raters was found to be moderate according to the weighted Cohen's kappa statistic (κ). The study was approved by the Research Ethics Boards at UHN: Toronto Rehabilitation Institute and Sunnybrook Research Institute. Written informed consent was obtained by all the participants according to the Declaration of Helsinki.

Table 1: *Mean value and standard deviation of ROM and asymmetry scores across the participant groups considered in the study*

	HC subjects	Patients PS	% Agreement	κ
ROM	1.06±0.13	1.80±0.45	68.9%	0.50
Asymm	1.31±0.22	2.33±0.70	58.7%	0.55

2.2. Experimental setup

Each participant was asked to perform the following tasks: rest position for 20s with teeth in normal bite and neutral facial expression (*REST*); maximum jaw opening repeated 5 times (*OPEN*); lip puckering (pretending to kiss a baby) 5 times (*KISS*); pretending to blow a candle 5 times (*BLOW*); lip spreading (pretending to smile with closed lips) 5 times (*SPREAD*); repetition of the syllable /pa/ in a single breath, as fast as possible (*PA*); repetition of /pataka/ as fast as possible (*PATAKA*); 10 repetitions of the sentence "Buy Bobby a puppy" at habitual loudness and speaking rate (*BBP*).

During each task, the face of the participants was video-recorded using the Intel® RealSense™ SR300 camera, placed between 0.4-0.5 m from the face [22]. Participants were comfortably seated in front of the camera during the experiments, and their faces were illuminated by a uniform light source placed behind the SR300. Each video-recording consisted of a pair of registered and synchronized videos (color and depth) recorded at approximately 50 frames per second and 640x480 pixels of image resolution. A total of 184 pairs of color and depth videos was obtained and analyzed in this study, including 1119 speech and non-speech repetitions distributed as follows: 221 *BBP* (110 from HC subjects and 111 from patients PS); 194 *PATAKA* (107 from HC subjects and 87 from patients PS); 225 *PA* (110 from HC subjects and 115 from patients PS); 121 *BLOW* (59 from HC subjects and 62 from patients PS); 121 *KISS* (57 from HC subjects and 64 from patients PS); 119 *OPEN* (55 from HC subjects and 64 from patients PS); and 118 *SPREAD* (53 from HC subjects and 65 from patients PS). The acquisitions were performed with a customized C++ code and the Intel® RealSense™ SDK R3 2016.

3. Methods

3.1. Pre-processing

Each video recording was manually labeled by a trained annotator to identify the frames at the beginning and end of each speech and non-speech repetition. Before extracting facial geometric features for each task, the intrinsic calibration parameters, such as the focal length and principal point, were estimated for both color and infrared cameras of the SR300. This calibration was performed using the Camera Calibration Toolbox for Matlab, by recording 25 images of a black and white checkerboard pattern at different distances and angles with the SR300 camera [23].

3.2. Face alignment

The face alignment step was used for the automatic location of the facial landmarks (i.e., points of the lips, nose, eyes, and eyebrows) on the color video frames of each task. Two algorithms were considered in this study: the supervised descent method (*SDM*) for face alignment and the Intel® RealSense™ Face Tracking (*IntelFT*) algorithm [16], [24].

SDM is a non-parametric face alignment approach that solves the optimization problem (i.e., minimization of the difference between image features extracted in a generic landmark location and the same features sampled in the true landmark location) in a supervised manner. *SDM* requires a training phase to learn the descent directions for estimating the facial landmark positions given the SIFT features extracted from the color image [24], [25]. In this work, we used the pre-trained Matlab implementation proposed in [24] that allowed tracking 49 facial points (Figure 1a).

The second approach was the face tracking algorithm provided with the Intel® RealSense™ SDK R3 2016. Technical details about this approach have not been disclosed by the manufacturer. However, it was included in our experiments because of its availability with the SDK, which allowed a simple implementation with video recordings captured with the SR300 camera. *IntelFT* allowed tracking 78 facial landmarks (Figure 1b).

For each pair of color and depth frames, the frontal distance (in mm) of the facial landmarks from the camera

plane was estimated by sampling the depth image in the same location as the 2D facial points obtained on the color frame with the two algorithms. Then, using a pinhole camera model with the intrinsic parameters estimated during the calibration step, the 3D coordinates (in mm) of the facial points were obtained. The origin of the 3D coordinate system was the color camera center, and *X*, *Y*, and *Z* were the lateral, vertical, and frontal axes, respectively.

3.3. Feature extraction

For each repetition of each task, 10 geometric features (4 representing *ROM* and 6 representing asymmetry) were extracted using 3D coordinates of the following facial landmarks: right and left eyebrows (*RE* and *LE*), inner canthus of the eyes (*RIC* and *LIC*), right and left mouth corners (*RC* and *LC*), nose tip (*NT*), and central points of upper and lower lips (*UL* and *LL*) (Figure 1). A summary of the features used for this work is reported in Table 2. These features were extracted with both face alignment algorithms to determine the influence of these two approaches on the overall classification performance.

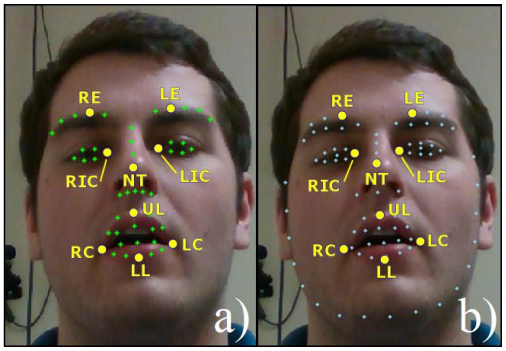

Figure 1: *Facial landmarks obtained with SDM (a) and IntelFT (b). The points of interest for feature extraction are highlighted in yellow.*

3.3.1. ROM features

Lip opening was calculated as the 3D Euclidean distance between *UL* and *LL*, whereas mouth width was the 3D Euclidean distance between *LC* and *RC*. These two distances were calculated for each frame of a repetition, and then normalized with respect to the average lip opening and width calculated during the *REST* task by using the following equation:

$$X^i_{norm} = \frac{X^i_{TASK} - X_{REST}}{X_{REST}} \cdot 100 \quad (1)$$

where X^i_{TASK} is a generic distance (lip opening or mouth width) calculated for the i^{th} frame of a speech or non-speech task, and X_{REST} is the same distance calculated during the *REST* recording. Thus, for each participant, lip opening and mouth width were expressed as percentage with respect to the baseline measures obtained at rest. After this normalization, the maximum and minimum values of lip opening (O_{MAX}, O_{MIN}) and width (W_{MAX}, W_{MIN}) within each repetition were used as features for the classification.

3.3.2. Asymmetry Features

Left and right mouth areas were calculated as the areas of two triangles with vertices *LC*, *UL*, *LL* (A_L) and *RC*, *UL*, *LL* (A_R), respectively. The absolute difference between these two areas was calculated for each frame of a repetition, and its mean value was considered as a feature for the classification (A_{diff}).

The Pearson's correlation coefficient between the *LC* and *RC* trajectories (re-expressed with respect to the nose tip to remove the effect of head rotation) was calculated within each repetition (r_{LCRC}). This feature was considered as an index of movement coordination between left and right sides of the mouth. Highly coordinated movements give rise to values of r_{LCRC} close to 1.

Following the approach proposed by Schimmel *et al.*, 2011 [13], the following distances were extracted from both sides of the face: distance between the eyebrow points and inner canthus of the eyes (d_{0L} and d_{0R}); distance between inner canthus of the eyes and mouth corners (d_{1L} and d_{1R}); distance between inner canthus of the eyes and *UL* (d_{2L} and d_{2R}); distance between the mouth corners and *UL* (d_{3L} and d_{3R}). For each distance, the absolute difference between the measure calculated on the right side and its counterpart on the left side was considered for the analysis (Table 2).

Table 2: *Description of the geometric features.*

	Feature	Description
ROM	O_{MAX}	Maximum and minimum values of lip opening with respect to *REST*
	O_{MIN}	
	W_{MAX}	Maximum and minimum values of mouth width with respect to *REST*
	W_{MIN}	
Asymmetry	A_{diff}	Absolute difference between A_L and A_R
	r_{LCRC}	Correlation between *LC* and *RC* trajectories
	d_{0diff}	Absolute difference between $d_{0\text{-}L}$ and $d_{0\text{-}R}$
	d_{1diff}	Absolute difference between $d_{1\text{-}L}$ and $d_{1\text{-}R}$
	d_{3diff}	Absolute difference between $d_{2\text{-}L}$ and $d_{2\text{-}R}$
	d_{3diff}	Absolute difference between $d_{3\text{-}L}$ and $d_{3\text{-}R}$

3.4. Classification and statistical analysis

Each classification instance consisted of a 10-dimensional feature vector (each feature was standardized as z-scores) corresponding to a single speech or non-speech repetition. The above *ROM* and asymmetry features were used to train a support vector machine (SVM) classifier with radial basis function kernel. A separate classifier was trained for each task, to compare the classification performance across tasks, and for each face alignment algorithm, to quantify the influence of the face alignment step on the overall classification performance. A binary classification was performed: a class label equal to 0 was associated to the instance if the features came from HC subjects, whereas a class label equal to 1 was assigned if the repetitions came from patients PS.

Classification performance was evaluated using a leave-one-subject-out cross-validation (*LOSO-CV*). For each fold of the *LOSO-CV*, the instances belonging to a single participant were used as a test-set, and the classifier was trained with the remaining instances/ subjects. Then, the class of a test-subject was predicted by using the majority vote among the classified test instances. If the majority of the repetitions were classified as PS, then the test subject was classified as PS, and vice-versa. If the instances classified as PS were equal to the instances classified as HC, the subject was classified as HC to produce conservative predictions.

For each test, the classification performance was evaluated using the following measures: accuracy (percentage of correctly classified instances with respect to the total number of instances), sensitivity (percentage of instances from patients

PS correctly classified as PS), and specificity (percentage of instances from HC subjects correctly classified as HC). The Wilcoxon rank-sum test was used to detect statistically significant differences in the features between the two classes. Classification and statistical analysis were performed using the Statistic and Machine Learning Toolbox (Matlab v. 2016b).

4. Results

For each test, accuracy, sensitivity and specificity are reported in Table 3. For most of the tasks (except for *PA*), features extracted with *SDM* provided higher performance than those extracted with *IntelFT*. The highest accuracies were obtained with *BBP* (87%), *BLOW*, and *SPREAD* (82.6%). Features that showed statistically significant differences between-group during these 3 tasks are reported in Table 4.

Table 3: *Classification performance (Acc: accuracy; Sens: sensitivity; Spec: specificity; values are in %). For each task, the highest value between the two face alignment algorithms is highlighted in bold.*

Task	Algorithm	*Acc*	*Sens*	*Spec*
BBP	*SDM*	**87.0**	**75.0**	**100.0**
	IntelFT	60.9	58.3	63.6
PATAKA	*SDM*	**73.9**	**58.3**	**90.9**
	IntelFT	65.2	50.0	81.8
PA	*SDM*	65.2	50.0	**81.8**
	IntelFT	**73.9**	**83.3**	63.6
BLOW	*SDM*	**82.6**	**83.3**	**81.8**
	IntelFT	69.6	58.3	**81.8**
KISS	*SDM*	**60.9**	**66.7**	54.5
	IntelFT	**60.9**	41.7	**81.8**
OPEN	*SDM*	**78.3**	**75.0**	**81.8**
	IntelFT	65.2	66.7	63.6
SPREAD	*SDM*	**82.6**	**91.7**	**72.7**
	IntelFT	56.5	83.3	27.3

Table 4: *Statistical significant differences during the BBP, BLOW, and SPREAD tasks (*p<.05; **p<.001)*

Task	Feature	HC subjects	Patients PS
BBP	A_{diff} *(mm²)*	16.56±9.60	35.20±20.95*
	d_{1diff} *(mm)*	1.12±0.65	2.28±1.27*
BLOW	W_{MAX} *(%)*	2.14±8.36	-6.34±7.82*
	A_{diff} *(mm²)*	17.74±10.33	42.62±28.12*
SPREAD	A_{diff} *(mm²)*	8.08±4.23	23.78±24.67**
	r_{LCRC}	0.76±0.22	0.40±0.47*
	d_{2diff} *(mm)*	1.25±0.98	2.70±1.76*

5. Discussion and Conclusion

This paper demonstrates the feasibility of a video-based marker-less approach for the automatic detection of speech and orofacial impairment in patients PS. By using a small set of interpretable facial features, we were able to discriminate patients PS from HC subjects with high accuracy (up to 87%). This work expands previous findings [17], [19], adding further evidence to the suitability of video-based technology for detecting orofacial impairments due to neurological diseases.

Our results demonstrated that *BBP*, *BLOW*, and *SPREAD* were the tasks that better differentiated the two groups, with accuracies >80%. These findings are in line with [19], where *BBP* and *SPREAD* allowed discriminating patients with ALS from HC subjects with 89% and 83% accuracy, respectively. The relatively good results obtained with *BLOW* may be surprising especially if we consider the lower performance in a similar task like *KISS* (Table 3). However, Denlinger *et al.*, 2008 [26] demonstrated that patients with unilateral movement disorder may exhibit larger movements during *BLOW* than during lip puckering. This increased amplitude of movements may enhance differences between the two groups leading to higher classification performance.

Table 4 shows that in the 3 best tasks, patients PS showed higher asymmetry of the lower facial muscles, as indicated by higher values of d_{1diff}, d_{2diff}, r_{LCRC}, and A_{diff}. Among the *ROM* features, W_{MAX} appeared to be decreased in patients PS during *BLOW*. Considering that during this task the mouth width decreases from the rest position, the lower values of W_{MAX} may indicate that patients PS had difficulty to come back to the rest position after pretending to blow a candle.

Features extracted with *SDM* provided higher accuracy, confirming that a non-parametric method such as *SDM* is suitable when asymmetric orofacial movements are present [24]. The only task in which *IntelFT* outperformed *SDM* was *PA*. This was probably due to the limited lateral movements during the opening-closing gestures in this relatively simple task. To the best of our knowledge, previous works have not tested how the choice of the face alignment step influences the prediction performance in patients with orofacial impairments.

In this study, we considered a binary classification between patients PS and HC subjects, when all patients PS had orofacial impairment as judged by the clinicians (Table 1). Future developments will aim to correlate the clinicians' judgments, estimating the exact scores assigned during the oral motor exam. Also, an automatic segmentation of the repetitions of interest will be implemented to fully automate the process and translate the technology into clinical environments for usability testing.

This paper shows promising results for the automatic detection of orofacial impairment in patients PS during speech and non-speech tasks. An objective and continuous assessment of the facial function is important not only to evaluate the impact of stroke on the orofacial musculature, but also to track the recovery during the rehabilitation. In fact, previous works demonstrated that targeted orofacial and speech therapy might help improve facial movements after stroke, with positive impact on orofacial functions, speech, and communication abilities [27], [28]. Further clinical validation (i.e., correlation between objective features and clinical assessment) is the next step required to test the diagnostic efficacy of an automatic tool for assessing speech and orofacial impairments post-stroke.

6. Acknowledgments

The funding for this work was provided by the Canadian Partnership for Stroke Recovery (CPSR) Collaborative Catalyst Grant and Trainee Award. The authors would also like to acknowledge the funding support from, NIH (grants R01DC009890 and R01DC013547), AGE-WELL NCE Inc., a member of the Networks of Centres of Excellence program, and UHN-TRI. We also thank Jordan Scholl and Madhura Kulkarni for their assistance with this project.

7. References

[1] E. J. Benjamin *et al.*, "Heart Disease and Stroke Statistics'2017 Update: A Report from the American Heart Association," *Circulation*, vol. 135, no. 10. pp. e146–e603, 2017.

[2] H. L. Flowers, F. L. Silver, J. Fang, E. Rochon, and R. Martino, "The incidence, co-occurrence, and predictors of dysphagia, dysarthria, and aphasia after first-ever acute ischemic stroke," *J. Commun. Disord.*, vol. 46, no. 3, pp. 238–248, 2013.

[3] D. H. Gilden, "Bell's Palsy," *N. Engl. J. Med.*, vol. 351, no. 13, pp. 1323–1331, 2004.

[4] W. H. Chang *et al.*, "Impact of central facial palsy and dysarthria on quality of life in patients with stroke: The KOSCO study," *NeuroRehabilitation*, vol. 39, no. 2, pp. 253–259, 2016.

[5] J. W. House and D. E. Brackmann, "Facial nerve grading system," *Otolaryngol. - Head Neck Surg.*, vol. 93, no. 2, pp. 146–147, 1985.

[6] J. T. Vrabec *et al.*, "Facial Nerve Grading System 2.0," *Otolaryngol. - Head Neck Surg.*, vol. 140, no. 4, pp. 445–450, 2009.

[7] B. G. Ross, G. Fradet, and J. M. Nedzelski, "Development of a sensitive clinical facial grading system," *Otolaryngol. - Head Neck Surg.*, vol. 114, no. 3, pp. 380–386, 1996.

[8] G. E. Murty, G. M. O'donoghue, P. J. Bradley, J. P. Diver, and P. J. Kelly, "The Nottingham System: Objective assessment of facial nerve function in the clinic," *Otolaryngol. Neck Surg.*, vol. 110, no. 2, pp. 156–161, 1994.

[9] M. J. Brenner and J. G. Neely, "Approaches to grading facial nerve function.," *Semin. Plast. Surg.*, vol. 18, no. 1, pp. 13–22, 2004.

[10] S. He, J. J. Soraghan, B. F. Oreilly, and D. Xing, "Quantitative analysis of facial paralysis using local binary patterns in biomedical videos," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1864–1870, 2009.

[11] J. R. Delannoy and T. E. Ward, "A preliminary investigation into the use of machine vision techniques for automating facial paralysis rehabilitation therapy," in *IET Irish Signals and Systems Conference (ISSC 2010)*, 2010, pp. 228–232.

[12] B. F. O'Reilly, J. J. Soraghan, S. McGrenary, and S. He, "Objective method of assessing and presenting the house-brackmann and regional grades of facial palsy by production of a Facogram," *Otol. Neurotol.*, vol. 31, no. 3, pp. 486–491, 2010.

[13] M. Schimmel, B. Leemann, P. Christou, S. Kiliaridis, F. R. Herrmann, and F. Müller, "Quantitative assessment of facial muscle impairment in patients with hemispheric stroke," *J. Oral Rehabil.*, vol. 38, no. 11, pp. 800–809, 2011.

[14] W. Quan, B. J. Matuszewski, and L.-K. Shark, "Facial asymmetry analysis based on 3-D dynamic scans," in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, 2012.

[15] A. Bandini, S. Ouni, P. Cosi, S. Orlandi, and C. Manfredi, "Accuracy of a markerless acquisition technique for studying speech articulators," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2015, pp. 2162–2166.

[16] F. L. Siena, B. Byrom, P. Watts, and P. Breendon, "Utilising the Intel RealSense Camera for Measuring Health Outcomes in Clinical Research," *J. Med. Syst.*, vol. 42, no. 3, pp. 42–53, 2018.

[17] A. Bandini *et al.*, "Markerless Analysis of Articulatory Movements in Patients With Parkinson's Disease," *J. Voice*, vol. 30, no. 6, pp 766.e1-766.e11, 2016.

[18] A. Bandini *et al.*, "Analysis of facial expressions in parkinson's disease through video-based automatic methods," *J. Neurosci. Methods*, vol. 281, pp. 7-20, 2017.

[19] A. Bandini, J. R. Green, B. Taati, S. Orlandi, L. Zinman, and Y. Yunusova, "Automatic detection of amyotrophic lateral sclerosis (ALS) from video-based analysis of facial movements: speech and non-speech tasks," in *13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, 2018, pp. 150-157.

[20] Z. S. Nasreddine *et al.*, "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment," *J Am Geriatr Soc*, vol. 53, no. 4, pp. 695–699, 2005.

[21] J. R. Duffy, "Motor Speech Disorders: Clues to Neurologic Diagnosis," in *Parkinson's Disease and Movement Disorders: Diagnosis and Treatment Guidelines for the Practicing Physician*, 2000, pp. 35–53.

[22] A. Bandini, A. Namasivayam, and Y. Yunusova, "Video-based tracking of jaw movements during speech: Preliminary results and future directions," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2017, pp. 689–693.

[23] J. Y. Bouguet, "Camera calibration toolbox for MATLAB," *High Speed Vis. Syst. Object Track.*, no. Figure 1, pp. 5–8, 2010.

[24] X. Xiong and F. De La Torre, "Supervised descent method and its applications to face alignment," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 532–539, 2013.

[25] G. Lowe, "SIFT - The Scale Invariant Feature Transform," *Int. J. Comput. Vis.*, vol. 2, pp. 91–110, 2004.

[26] R. L. Denlinger, J. M. VanSwearingen, J. F. Cohn, and K. L. Schmidt, "Puckering and Blowing Facial Expressions in People With Facial Movement Disorders," *Phys. Ther.*, 2008.

[27] J. A. Kang, M. H. Chun, S. J. Choi, M. C. Chang, and Y. G. Yi, "Effects of mirror therapy using a tablet PC on central facial paresis in stroke patients," *Ann. Rehabil. Med.*, vol. 41, no. 3, pp. 347–353, 2017.

[28] P. Konecny *et al.*, "Central facial paresis and its impact on mimicry, psyche and quality of life in patients after stroke," *Biomed. Pap.*, vol. 158, no. 1, pp. 133–137, 2014.