Markerless Analysis of Articulatory Movements in Patients With Parkinson's Disease

*,†Andrea Bandini, *Silvia Orlandi, ‡Fabio Giovannelli, *Andrea Felici, ‡Massimo Cincotta, §Daniela Clemente, ‡Paola Vanni, ‡Gaetano Zaccara, and *Claudia Manfredi, *,‡,§*Firenze, Italy; and* †*Bologna, Italy*

Summary: Objectives. A large percentage of patients with Parkinson's disease have hypokinetic dysarthria, exhibiting reduced peak velocities of jaw and lips during speech. This limitation implies a reduction of speech intelligibility for such patients. This work aims at testing a cost-effective markerless approach for assessing kinematic parameters of hypokinetic dysarthria.

Study Design. Kinematic parameters of the lips are calculated during a syllable repetition task from 14 Parkinsonian patients and 14 age-matched control subjects.

Methods. Combining color and depth frames provided by a depth sensor (Microsoft Kinect), we computed the three-dimensional coordinates of main facial points. The peak velocities and accelerations of the lower lip during a syllable repetition task are considered to compare the two groups.

Results. Results show that Parkinsonian patients exhibit reduced peak velocities of the lower lip, both during the opening and the closing phase of the mouth. In addition, peak values of acceleration are reduced in Parkinsonian patients, although with significant differences only in the opening phase with respect to healthy control subjects.

Conclusions. The novel contribution of this work is the implementation of an entirely markerless technique capable to detect signs of hypokinetic dysarthria for the analysis of articulatory movements during speech. Although a large number of Parkinsonian patients have hypokinetic dysarthria, only a small percentage of them undergoes speech therapy to increase their articulatory movements. The system proposed here could be easily implemented in a home environment, thus, increasing the percentage of patients who can perform speech rehabilitation at home.

Key Words: Hypokinetic dysarthria–Markerless–Kinect–Parkinson's diseaseSpeech articulation.

INTRODUCTION

Idiopathic Parkinson's disease (PD) is a neurodegenerative disorder of the central nervous system resulting from the death of neurons in the zona compacta of the *substantia nigra* of the midbrain and other pigmented nuclei.[1–3] Cardinal motor symptoms of PD are rest tremor, bradykinesia, and rigidity.[4,5] Other common motor signs involve the control of speech production, affecting motion and coordination of the articulatory organs (ie, tongue, lips, and jaw). These signs are commonly known as "dysarthria".

Nowadays, it is well known that a high percentage of PD patients (around 90%) have dysarthria.[6,7] In particular, dysarthria associated with PD can affect all the speech dimensions: respiration, articulation, phonation, and prosody, which results in reduced speech intelligibility.[8] In most of PD patients, dysarthria is usually "hypokinetic". This term refers to the reduced range of movements involved in speech production. In fact, hypokinetic dysarthria is characterized by reduced peak velocities and displacements of the articulators during speech movements.[7] This articulatory undershoot may lead to alterations in some acoustical parameters that are typical of articulation, such as the reduction of the second formant transition slope or a reduction of the triangular vowel space area.[9–11]

Several approaches and methods have been implemented to describe the kinematic characteristics of the articulators in hypokinetic dysarthria associated with PD. Most studies[12–18] pointed out a reduction in terms of velocity and range of movements of the articulators, although results of unimpaired articulatory movements in Parkinsonian patients were also presented.[19]

One of the most widespread tasks for kinematic analysis of the articulators is the production of syllables. Through this test, several authors[12,14,15] found a reduction of the displacement and of the peak velocities of lower lip and jaw during the opening and the closing of the mouth. A slowdown of jaw and lower lip movements was demonstrated also by Forrest et al[13]; on the other hand, they found an increase of the closing velocity of the lower lip in PD patients. This result might reflect an alteration of the motion control due to the severity of the dysarthria. In fact, as mentioned above, although the majority of PD patients have hypokinetic dysarthria, a small percentage experiences symptoms of hyperkinetic dysarthria, whose occurrence could be related to the prolonged administration of drug therapies that causes the onset of involuntary movements (dyskinesias).[7]

Other works on the articulatory kinematics in PD patients with hypokinetic dysarthria focused on tongue movements during speech. Yunusova et al[16] studied movements of tongue, jaw, and lower lip during the pronunciation of words, finding that tongue movements of PD patients could be more discriminative as compared with control subjects, although a reduction in lips and jaw kinematics still exists. Wong et al[17] studied tongue movements

Accepted for publication October 22, 2015.

From the *Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy; †Department of Electrical, Electronic and Information Engineering (DEI) "Guglielmo Marconi", Università di Bologna, Bologna, Italy; ‡Unit of Neurology, Florence Health Authority, Ospedale "Nuovo San Giovanni di Dio"; and the §Private Practice, Firenze, Italy.

Address correspondence and reprint requests to Andrea Bandini, Department of Information Engineering, Università degli Studi di Firenze, Via di S. Marta 3, 50139 Firenze, Italy. E-mail address: andrea.bandini@unifi.it

Journal of Voice, Vol. 30, No. 6, pp. 766.e1–766.e11
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2015.10.014

during a sentence repetition task to discriminate between dysarthric and nondysarthric PD patients. Unlike most of the previous findings, Wong et al demonstrated that dysarthric PD patients exhibited wider ranges of movements with an increase of peak velocities and accelerations.

Although it is well accepted that PD patients with hypokinetic dysarthria exhibit a reduced articulatory kinematic, some conflicting results were found. Walsh and Smith[18] tried to elucidate this point by studying jaw and lower lip movements during the opening and the closing gestures in the case of bilabial consonants. The authors demonstrated that PD patients exhibited a reduced articulatory kinematic, highlighted by reduced velocities of jaw and lower lip. These results support the hypothesis that a "downscaling" in speech production occurs in PD patients with hypokinetic dysarthria.

Another important point concerns the implemented methodologies. In the past decades, the kinematic analysis of the articulators was performed through several motion capture technologies. The most important are optoelectronic systems,[14,18,19] electromagnetic articulography (EMA),[17] X-ray techniques,[16] and magnetic resonance imaging.[20] However, all these techniques are quite expensive and their use is limited to research within highly specialized laboratories. Moreover, some of the most widely used techniques (optoelectronic systems and EMA above all) are marker based and need quite long preparation protocols to achieve good results.

Although a large number of PD patients have hypokinetic dysarthria, today, only a small percentage of PD patients undergoes speech therapy with specific protocols aimed at increasing their articulatory movements. This is due to several factors:

- during group sessions, the speech therapist has difficulty paying equal attention to each patient to evaluate the performance during exercises and provide an immediate feedback;
- several patients with hypokinetic dysarthria due to neurodegenerative diseases are elderly people who often have difficulties in moving to specialized centers;
- patients should perform the exercises also at home. However, they do not because they lack the presence of the therapist.

In the last years, several results in monitoring and rehabilitation of dysarthria in home environment were achieved with the help of acoustical analysis.[21,22] Instead, few results were obtained for the automation of exercises that involve facial muscles like those commonly performed in speech therapy protocols. The main reasons are related to the high costs of the methods used to study articulatory movements.

Recently, we proposed a fully markerless and low-cost method to study the articulatory movements in three dimensions (in particular lip movements) during speech. We tested its accuracy against an optoelectronic marker-based method during the repetition of words, sentences, and syllables.[23,24] This system is able to track lip movements during speech, combining a face tracking algorithm and a 3D depth sensor. The obtained results were promising, with mean errors (against the marker-based reference) between 1 and 4 mm for lips.[23] Moreover, good results were achieved for the estimation of velocity and acceleration of the lower lip during a syllable repetition task.[24]

This markerless technique, whose accuracy was already proven for tracking lip movements, is thus applied here to analyze movements of the lower lip during a syllable repetition task, both in PD patients and in healthy control (HC) subjects. The aim is to test the reliability of a cheap and markerless approach for assessing signs of hypokinetic dysarthria (in particular, alterations of peak velocities and accelerations) as already demonstrated by Walsh and Smith.[18]

This system could be used for the analysis of the articulatory movements during speech, leading to a significant improvement in monitoring disease progression, and in speech therapy in patients with dysarthria due to neurodegenerative diseases. In fact, this system could be easily implemented in a home environment, increasing the percentage of patients who undergoes speech therapy. Merging efficient algorithms and low-cost devices could lead to the development of applications capable of providing a real-time feedback about the right facial movements to be performed by the patient during speech therapy protocols.

MATERIALS AND METHODS

Subjects

Fourteen PD patients were recruited at the Unit of Neurology of the Florence Health Authority ("San Giovanni di Dio" Hospital, Firenze, Italy) and at the Associazione Italiana Parkinsoniani—Sezione di Firenze, Firenze, Italy. PD patients' age ranged between 62 and 80 years (mean: 71.6 years; standard deviation: 7.0 years). Nine patients were male and five were female. Before carrying out the experiment, each patient underwent a neurological examination. The Hoehn and Yahr disease stage[25] ranged from 1.5 to 2.5 (2.0 ± 0.3) and the Unified Parkinson's Disease Rating Scale (UPDRS) motor score (UPDRS part III[26]) ranged from 5 to 43 (16.0 ± 12.0), whereas the speech task (item 18 of the UPDRS part III protocol) gave results equal to 0 or 1. Through this item, the neurologist judges the Parkinsonian spontaneous speech, paying attention on those signs related to dysphonia and dysprosody, but still considering global aspects as the intelligibility. In this way, ratings do not directly concern the issue addressed in this paper (the articulatory undershoot). Thus, PD patients assessed through the perceptual evaluation of the neurologist showed no speech problems (speech item = 0) or slight problems (speech item = 1) consisting in loss of modulation, diction, or volume, without major alterations of speech intelligibility. All PD patients were under levodopa medication and were tested during their "on" state.

An age-matched control group composed of 14 HC subjects with no history of neurological disease was recruited. HC subjects' age ranged from 60 to 85 years (mean: 69.0 years; standard deviation: 7.4 years). Eight subjects were male and six were female. Table 1 summarizes the features of the two groups, both consisting of Italian native speakers. Signed informed consent was obtained from all the participants. These groups are a part of the dataset described in Reference 27, where the acoustical

TABLE 1.
Characteristics (Mean Values and Standard Deviations) of the Two Groups Analyzed in This Work

	PD Patients		HC Subjects	
	Mean	SD	Mean	SD
Age (years)	71.6	7.0	69.0	7.4
Male	9		8	
Female	5		6	
Disease duration (years)	8.4	6.1	—	
Hoehn and Yahr stage	2.0	0.3	—	
UPDRS motor score	16.0	12.0	—	
UPDRS speech	0.6	0.5	—	

analysis was performed on a sentence repetition task to detect acoustical features related to dysprosody. Indeed, both tasks (sentence and syllable repetition) are parts of an experiment performed on these groups, whose aim was to detect relevant acoustical and kinematic features of PD patients during speech. However, not all patients analyzed in the previous study[27] gave their consent to be video recorded during the test, thus, here a subgroup was considered.

Experimental settings

The experiments were carried out in a quiet room of the "San Giovanni di Dio" Hospital, Firenze, Italy. The speech task consisted in the repetition of the syllable /pa/ for at least 25 times within a single breath, in a comfortable steady pace. In Reference 28, participants were asked to avoid acceleration and slowdown of the articulatory velocity. As shown in Skodda et al,[28] a difference between PD patients and HC subjects might be present in the pace of repetition (ie, PD patients tend to accelerate the rhythm of repetition). However, the analysis of this point goes beyond the aims of the present paper and is not considered here. Subjects were seated during the experiment, avoiding abrupt head movements during the whole task.

The subjects' face was recorded using the Microsoft Kinect for Windows sensor. The aim was to detect the 3D coordinates of some facial points during the speech task. The Microsoft Kinect is a structured light sensor that provides two video streams: a color stream (in the RGB color space) and a depth stream where each pixel codes the distance of the points in the scene from the camera plane. The Kinect sensor was placed in front of the subject's face at a distance between 0.5 and 0.7 m from the mouth and at a height close to that of the subject's eyes. This distance was chosen as a trade off between the technical specifications provided by the manufacturer (in "near range" mode, the minimum distance is 0.4 m[29]) and the need of having the subject's face as close as possible to the camera, to achieve the best accuracy in tracking the 3D facial points. In fact, our previous experiments[23] showed that the error in depth estimation grows significantly for distances greater than 0.8 m.

Video processing

Video specifications

The resolution of both streams was 640 × 480 pixels at 30 frames per second. Color frames were recorded in 24-bit RGB images (8 bits per channel), whereas depth frames were recorded in 16-bit, one-channel images. These features are the best trade-off, in terms of spatial and temporal resolution provided by this kind of sensor, to achieve good accuracies for tracking fast movements like those of the lips during the syllable repetition task. Moreover, our previous study on the accuracy in tracking articulatory movements with a depth sensor[23] showed that with an image resolution of 320 × 240 pixels for the depth stream, in some cases we obtained errors around 4 mm. Therefore, we concluded that a depth resolution of 640 × 480 pixels might improve the results.

Both streams were recorded and stored in avi files through the *OpenNI* (Open Natural Interaction, ver. 2.2) and *OpenCV* (Itseez, ver. 2.4.9) libraries using a customized code written in C++ language.

Depth-color registration

Facial feature points were tracked on the color images by the face tracking algorithm that will be described in the next section. To provide the 3D location of these points, the alignment between depth and color frames is required. In fact, as the two streams come from two different and unaligned cameras (color and infrared cameras, as reported in Figure 1), a stereo calibration step is firstly required. This step involves two parts:

- Intrinsic calibration for both cameras: We retrieved the internal parameters of each camera (ie, focal length, principal point, skew coefficient, and distortions). An exhaustive description of these parameters is reported in Reference 30. These parameters allow the estimation of the 3D coordinates of facial points, as described in the next section.
- Extrinsic calibration: The rotation matrix and the translation vector that define the position of one camera with respect to the other one are estimated. In our case, we estimated the roto-translation matrix required to map the position of each depth pixel in the color reference frame.

In this work, the calibration step was performed using the Camera Calibration Toolbox for Matlab[31] by simultaneously recording with the two cameras 25 images of a checkerboard in different positions and angles. This procedure was performed just once before making the video recordings on both groups.

Once the intrinsic parameters of each camera and the extrinsic relationship between the two reference frames were estimated, it was possible to map the pixels of the depth image into the color reference frame, as shown in Figure 1.

The alignment process is of great relevance to achieve good results in the next steps, in particular in the estimation of the 3D coordinates of facial points. In fact, the face tracker used to automatically locate and track the facial landmarks relies only on the color frames, providing two-dimensional (2D) points in the image plane. Therefore, to estimate the 3D position of these points, their distance from the camera plane must be known. After

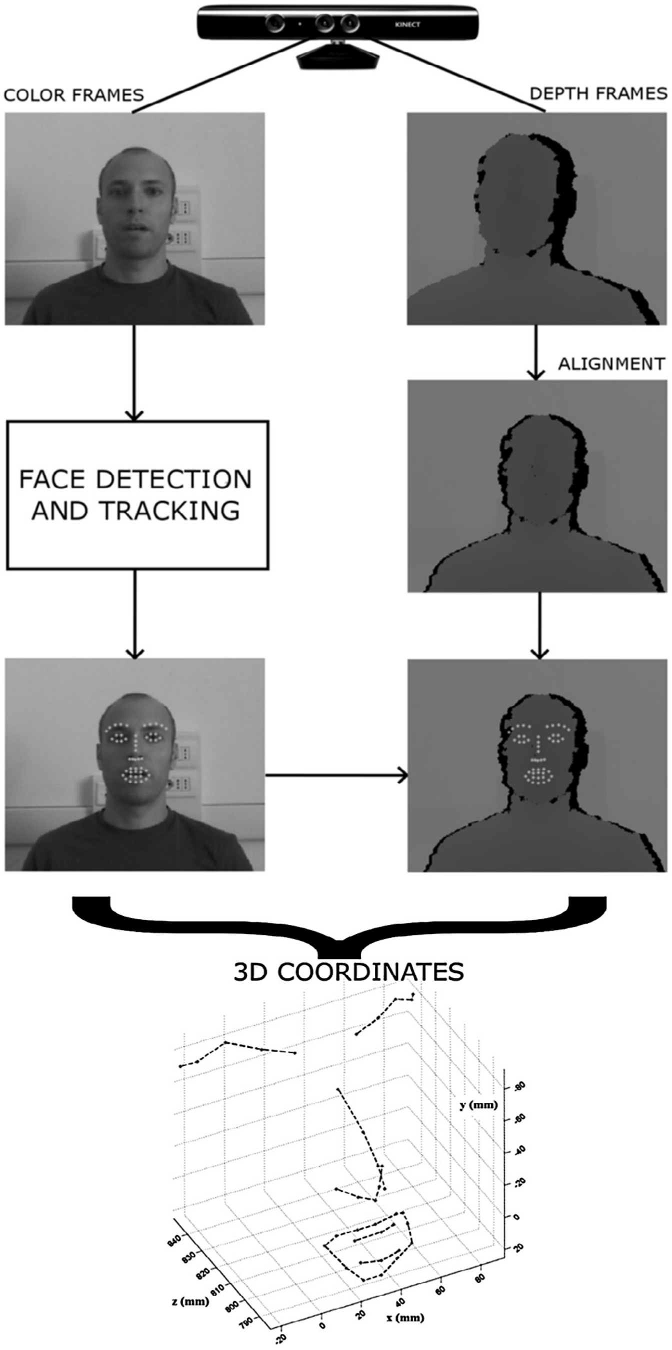

FIGURE 1. Main video processing steps. (*Left*) The face tracking algorithm detects the facial feature points in the color video frames. (*Right*) The information of the depth frames allows computing the 3D coordinates of the face points (*bottom plot*).

the alignment, these distances are given by the values of the depth image in the same pixel coordinates of the points of interest identified in the color image, sampled at the same time instant.

Although a depth to color registration was possible by means of the *OpenNI* libraries, a manual calibration of the sensor was carried out. In fact, the intrinsic parameters on which this "factory calibration" is based might not be very accurate. Thus, we preferred to perform the calibration before carrying out the experiments.

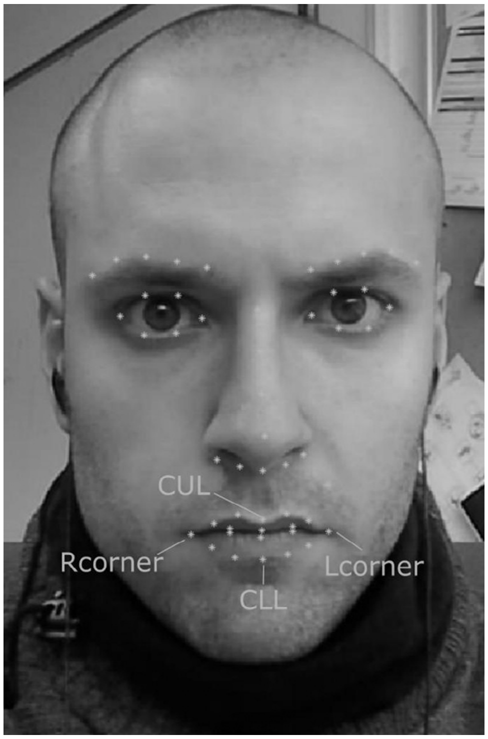

FIGURE 2. Face model used in this work. The *Intraface* tracker allows detecting and tracking 49 face points (*dots*). For the kinematic analysis, the following points are taken into account: CUL, central upper lip; CLL, central lower lip; and the two mouth corners: *right*—Rcorner, *left*—Lcorner.

Face tracking

As previously introduced, the facial feature points were located and tracked by means of a face tracking algorithm. In this work, we used the *Intraface* tracking algorithm (Human Sensing Laboratory, Robotics Institute, CMU and Affect Analysis Group, University of Pittsburgh, PA, USA) as implemented in References 23 and 24, where its performance (combined with a depth sensor) in tracking lip movements during speech was tested.[23,24] This tracking algorithm fits a face model made up by 49 feature points to the color images provided by the camera. These points are distributed as follows: 10 for the eyebrows, 12 for the eyes, nine for the nose, and 18 for the lips (12 on the outer contour, six on the inner contour) as shown in Figure 2. To solve the optimization problem that consists in minimizing the distance between the model and the image, the algorithm uses texture descriptors (scale invariant features transform). These descriptors make the tracker robust against illumination changes.[32,33] Moreover *Intraface* was chosen for its ability to generalize situations and face movements never seen in the training set, like asymmetrical movements of the mouth and eyelid movements. This could lead to a greater flexibility in view of the development of a system for speech therapy purposes, where exercises of facial muscles that involve asymmetrical movements are of great importance.

As the face tracker works on 2D color images, the facial feature points have coordinates in the image plane. To compute the 3D coordinates of these points, we used the depth value (in millimeter) retrieved by sampling the depth image in the same pixel

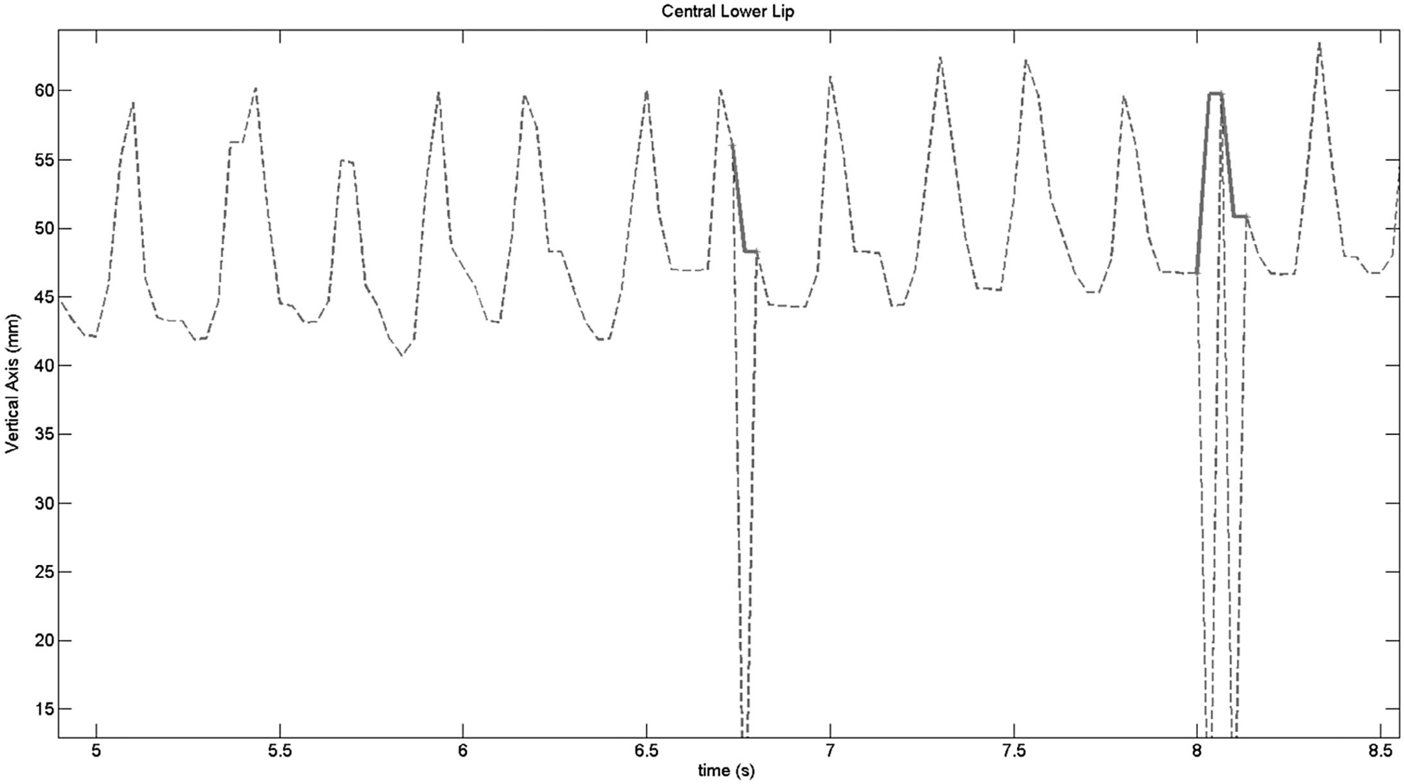

FIGURE 3. Artifacts correction for lip trajectories. The time intervals where the trajectory (*dashed line*) is equal to zero are detected and corrected through a nearest neighbor interpolation (*solid segments*). This figure refers to the vertical trajectory of the central lower lip (CLL) point during the syllable repetition task.

coordinates as the facial points identified in the color image. Through the intrinsic parameters of the color camera, estimated during the calibration step, we computed the coordinates on the lateral axis (X) and on the vertical axis (Y) starting from the depth information (frontal axis—Z):

$$X = Z\frac{(x - c_x)}{f_H} \tag{1}$$

$$Y = Z\frac{(y - c_y)}{f_V} \tag{2}$$

where x and y are the coordinates on the image plane (in pixels) of a point of coordinates $[X\ Y\ Z]^{\mathrm{T}}$ in the 3D space, f_H and f_v are the focal distances expressed in units of horizontal and vertical pixels of the color camera, and c_x and c_y are the coordinates (in pixels) of the principal point of the color camera, which is the point where the optical axis intersects the image plane.

In this work, the lens distortion parameters were not taken into account because the Kinect color camera uses lenses with low distortion.[34]

Artifacts correction

The depth images are estimated by means of a structured light coding, thus in some cases (in particular during the opening phase), it was difficult to estimate the distances from the camera plane for the area inside the mouth. Frequently, this area assumes zero values. Because lip movements are fast during the repetition of the syllable /pa/, sometimes the tracked points on the external contour of the lips could be located on "border" areas were the depth value is zero. Indeed, according to Equations 1 and 2, if Z (frontal axis) is equal to 0, X (lateral axis) and Y (vertical axis) are also equal to zero. This problem results in the presence of artifacts in the trajectories, as shown in Figure 3. The time intervals where the trajectory is equal to zero were detected and corrected through a nearest neighbor interpolation to remove these artifacts. This method was preferred to other techniques (such as linear or cubic interpolation) for its low memory requirements and the fast computation time that make it suitable for real-time applications. An example of the artifacts correction process is reported in Figure 3.

3D kinematic parameters

After the 3D coordinates of the facial feature points were computed, the velocity and acceleration on the vertical axis were calculated for the central point of the lower lip (CLL point, Figure 2). We considered just the movements on the vertical axis because they are the most important ones during the repetition of the syllable /pa/. First, the trajectory of interest was smoothed with a five-point moving average window (165 ms) to avoid large distortions of the curve especially during fast repetitions. Then the minimum values (that correspond to the closure time instants) were detected. The time interval between two consecutive minima corresponds to the period of a single syllable repetition. Thus, for each time interval, the velocity and the acceleration

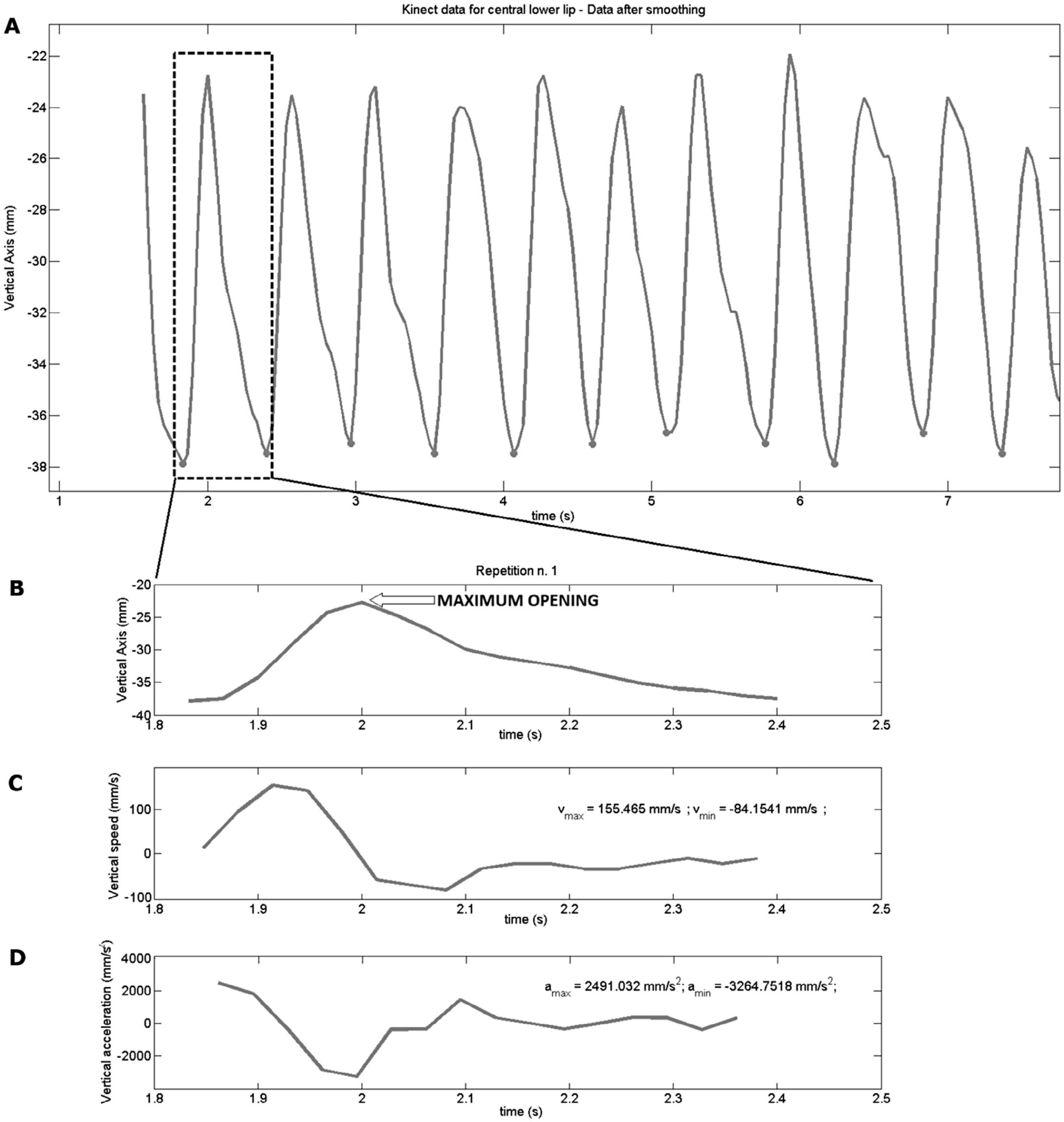

FIGURE 4. Computation of the kinematic parameters (velocity and acceleration). **A.** The minima of the vertical trajectory of point CLL (*solid line* in the upper plot) are detected (*dots*) to separate each repetition. **B.** Vertical trajectory of point CLL of a single repetition. **C.** Velocity on the vertical axis calculated as the first time derivative of the trajectory. **D.** Acceleration on the vertical axis calculated as the second time derivative of the trajectory. The maximum velocity and acceleration values are relative to the opening phase, whereas the minimum values refer to the closing phase.

on the vertical axis were computed as the first and the second time derivatives, respectively. The maximum and the minimum values were computed for both velocity and acceleration of each repetition. The maximum value is relative to the opening phase ($v_{opening}$ and $a_{opening}$), whereas the minimum value refers to the closing phase ($v_{closing}$ and $a_{closing}$) (Figure 4).

For each repetition, we calculated also the opening of the lips as the difference between the vertical coordinates of the CLL

TABLE 2.
Results of the Kinematic Analysis of the Lower Lip. Bold = statistically significative results.

		PD Patients	HC Subjects	*t* Test
3D	$v_{opening}$ (mm/s)	64.45 ± 30.94	94.94 ± 33.40	***t*(28) = 2.49, *P* = 0.02**
	$v_{closing}$ (mm/s)	61.54 ± 28.49	87.85 ± 31.28	***t*(28) = 2.32, *P* = 0.03**
	$a_{opening}$ (mm/s^2)	1277.67 ± 569.89	1768.78 ± 645.226	***t*(28) = 2.13, *P* = 0.04**
	$a_{closing}$ (mm/s^2)	2017.35 ± 1017.53	2766.29 ± 966.75	*t*(28) = 1.99, *P* = 0.06
	$\Delta Opening_{norm}$	0.46 ± 0.23	0.65 ± 0.36	*t*(28) = 1.70, *P* = 0.10
	$MaxOpening_{norm}$	0.36 ± 0.07	0.39 ± 0.10	*t*(28) = 0.91, *P* = 0.37
2D	$v_{opening}$ (s^{-1})	0.20 ± 0.10	0.26 ± 0.11	*t*(28) = 1.65, *P* = 0.11
	$v_{closing}$ (s^{-1})	0.19 ± 0.09	0.24 ± 0.09	*t*(28) = 1.44, *P* = 0.16
	$a_{opening}$ (s^{-2})	3.75 ± 1.73	4.65 ± 1.54	*t*(28) = 1.45, *P* = 0.16
	$a_{closing}$ (s^{-2})	6.00 ± 2.92	7.05 ± 2.33	*t*(28) = 1.05, *P* = 0.30
	$\Delta Opening_{norm}$	0.44 ± 0.24	0.61 ± 0.28	*t*(28) = 1.63, *P* = 0.11
	$MaxOpening_{norm}$	0.64 ± 0.13	0.68 ± 0.23	*t*(28) = 0.51, *P* = 0.61

and the central upper lip (Figure 2). This distance was normalized with respect to the head roll angle (rotations around the frontal axis). We did not take into account the rotations around the other two axes as the camera was placed in front of the subject's face at a height close to that of the eyes. Finally, for each repetition, the following parameters were computed:

- Normalized range of opening ($\Delta Opening_{norm}$): The difference between the maximum and the minimum opening values divided by its mean value;
- Normalized maximum opening value ($MaxOpening_{norm}$): The maximum opening value within a repetition, divided by the width of the lips. The width was calculated as the difference between the X-coordinates of the points Rcorner and Lcorner (Figure 2) normalized with respect to the roll angle of the head.

These normalized values allowed taking into account the anatomical variations among subjects that result in different values of opening and width of the mouth.

2D kinematic parameters

As the repetition of the syllable /pa/ involves main movements on the vertical axis, we computed the same kinematic parameters defined in the previous section (opening and closing velocity and acceleration, normalized range of opening and normalized maximum value of opening) from only 2D images. Indeed, if differences between groups could be detected also through 2D analysis, the advantages of this markerless system for monitoring and rehabilitation of speech diseases might be extended to the use of a simple webcam without requiring a depth stream. This would allow a further reduction of the costs, making the system easier and applicable to smartphones and tablets.

Starting from the 2D face points tracked by *Intraface*, the x and y coordinates of the points in the image plane were normalized with respect to their maximum values along the whole task. The maximum y coordinate is the vertical coordinate of the point CLL in the maximum opening instant, whereas the maximum x value is that of the external point of the left eyebrow (Figure 2). Thus, the coordinates assume values between 0 and 1 for all the subjects. After the normalization, the same procedure applied to the 3D kinematic parameters was used here.

Statistical analysis

A two-tailed *t* test was performed to assess the significance of differences between PD patients and HC subjects. The degrees of freedom of the distribution are equal to 28. The difference was considered significant for $P < 0.05$.

RESULTS

The parameters obtained with the kinematic analysis ($v_{opening}$, $a_{opening}$, $v_{closing}$, $a_{closing}$, $\Delta Opening_{norm}$, $MaxOpening_{norm}$) are reported in Table 2 for both 2D and 3D analyses. Specifically, $v_{opening}$ and $a_{opening}$ refer to the maximum speed and acceleration of point CLL during the opening phase; $v_{closing}$ and $a_{closing}$ refer to the maximum speed and acceleration of point CLL during the closing phase; and $\Delta Opening_{norm}$ and $MaxOpening_{norm}$ are the normalized range of opening and the normalized maximum opening value, respectively. For simplicity, values related to the closing phase ($v_{closing}$ and $a_{closing}$, Table 2) are reported in absolute value because they should be negative (Figure 4). Indeed, in this experiment, the values on the vertical axis increase downward (Figures 1 and 4). This is due to the fact that the 3D coordinates are computed starting from the 2D image coordinates where the plane origin is located in the upper left corner and the y axis is positive downward. The closing movement direction is opposite to that axis, leading to negative values of speed and acceleration.

Concerning the 3D analysis, the results show significant differences in $v_{opening}$ ($t[28] = 2.49$, $P = 0.019$), $v_{closing}$ ($t[28] = 2.32$, $P = 0.028$), and $a_{opening}$ ($t[28] = 2.13$, $P = 0.043$). All these values are lower in PD patients. In particular, the opening and closing velocities are reduced by more than 25 mm/s (94.94 ± 33.40 mm/s vs 64.45 ± 30.94 mm/s for $v_{opening}$, 87.85 ± 31.28 mm/s vs 61.54 ± 28.49 mm/s for $v_{closing}$). Lower values were found also for $a_{opening}$, although not significant. Concerning the opening parameters ($\Delta Opening_{norm}$, $MaxOpening_{norm}$), the normalized range of opening is lower in PD patients (0.65 ± 0.36 vs 0.46 ± 0.23), although not significant, whereas the normalized maximum value is comparable with values around 0.4 in both groups (ie, the

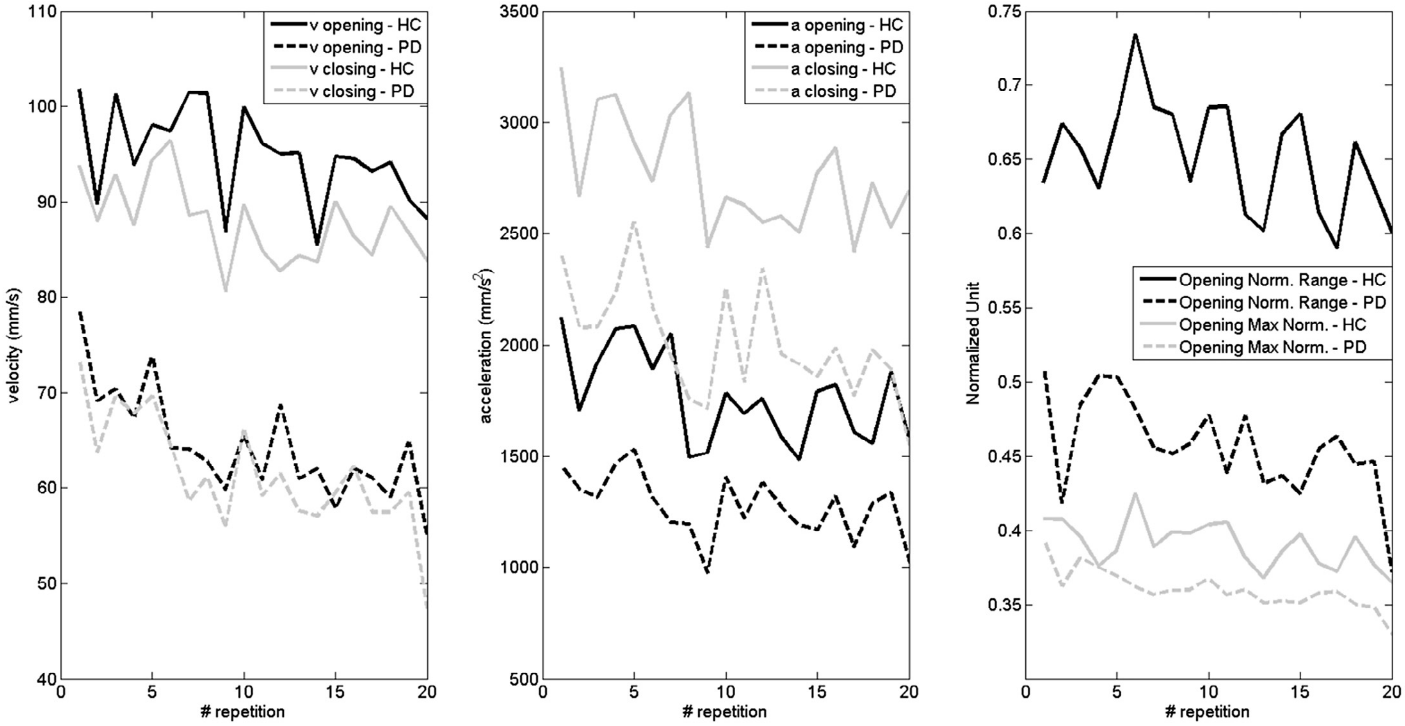

FIGURE 5. Kinematic parameters along the whole repetition task. Solid lines, HC subjects; dashed lines, PD patients; left plot, trend of peak velocities; middle plot, trend of peak acceleration values; right plot, trend of normalized opening values. All the parameters show a decrease in both groups; however, it seems to be more pronounced in PD patients especially for velocities (*left plot* and Table 3).

maximum opening is about the 40% of the mouth width). These values are normalized to take into account the anatomical variations among subjects that result in different values of opening and width of the mouth, as explained in the previous section.

As far as the 2D kinematic analysis is concerned, similar considerations can be drawn. Velocities and accelerations are reduced in PD patients, although no significant differences were found.

Figure 5 shows the trend of the six 3D parameters described above during the whole repetition task. The plots show a decrease of all parameters for both groups, more pronounced for velocity in PD patients. To provide a quantitative evaluation of this trend, a linear regression was applied. The slope of the regression line is reported in Table 3 for each parameter and both PD and HC subjects. Indeed, Table 3 shows that the decrease of the velocities in HC subjects is slower than in PD patients (–0.41 vs –0.72 for $v_{opening}$ and –0.38 vs –0.77 for $v_{closing}$), whereas for the other parameters, the trend is comparable.

TABLE 3.
Slope of the Regression Line of the Kinematic Parameters Along the Entire Repetition Task

	PD	HC
$v_{opening}$	–0.72	–0.41
$v_{closing}$	–0.77	–0.38
$a_{opening}$	–12	–19
$a_{closing}$	–26	–25
$\Delta Opening_{norm}$	–0.0034	–0.0027
$MaxOpening_{norm}$	–0.0018	–0.0015

DISCUSSION

Our results confirm the findings reported by Walsh and Smith,[18] where PD patients exhibited reduced peak velocities of lower lip, both for opening and closing phases. In addition, the peak values of acceleration are reduced in PD patients with significant differences only for the opening phase. Thus, our work supports most of the literature on the kinematic analysis of the articulators in PD patients with hypokinetic dysarthria, which states that most of these patients exhibit a downscaling of the articulatory movements.[12–16,18]

The novel contribution of this work concerns the assessment of this downscaling by means of a fully markerless and low-cost method. We described in detail the video-processing framework: camera calibration, face tracking, 3D coordinates estimation, and calculation of the kinematic parameters. All of these steps can be easily automated, providing an easy-to-use method for speech therapy and disease progression monitoring. By exploiting an existing face tracking algorithm and computer vision methods, the application of contact sensors or markers to the subject’s face is no longer required and no other manual setting is needed with obvious advantages for the patient.

The accuracy of this technique was already demonstrated in previous works,[23,24] although those experiments rely only on healthy subjects. Thus, a further development should be focused on testing this markerless method against a marker-based reference also for dysarthric patients.

Because of the limited number of PD patients, we did not make any distinction among different levels of severity of dysarthria. Thus, future studies will be devoted to increasing the dataset, recruiting PD patients at different stages of the disease, and with different levels of speech impairment, to investigate the evolution of the kinematic parameters. We expect a reduction of the kinematic parameters proportional to the severity of dysarthria, although further investigations are strictly necessary. As far as the normalized range of opening ($\Delta\text{Opening}_{\text{norm}}$) is concerned, we found lower values in PD patients with respect to HC subjects, although this difference is not significant. This result supports previous findings,[18] where the displacement of the lower lip during the utterance of bilabial sounds was lower in PD patients with respect to HC subjects. In our case, PD patients have lower values of $\Delta\text{Opening}_{\text{norm}}$. This could be due to the fact that the maximum/minimum interval of lower lip displacement is reduced in PD patients, because this value is defined as the difference between the maximum and the minimum opening values divided by its mean value. Differently from Reference 18, here, we normalized the opening values. This allows taking into account the anatomical variations of different subjects. Conversely, the results of $\text{MaxOpening}_{\text{norm}}$ are similar between the two groups. As the maximum/minimum interval is reduced in PD patients, but the maximum value of opening is similar between groups, the decrease of the normalized range of opening in PD patients might be due to an increase of the minimum value of opening. This means that during the pronunciation of the bilabial plosive /p/, the occlusion phase of the lips before the “burst” (releasing phase of the airflow) is less pronounced in PD patients. This might be due to a weakness in tightening the lips.

Concerning the 2D analysis, we found a decrease in the kinematic parameters of PD patients with respect to HC subjects as well as for the 3D analysis. However, the comparison among the 2D parameters did not show any significant difference between groups. For this reason, we recommend the use of both video streams (color and depth) to evaluate the articulatory kinematic to get highly accurate results. New experiments should be performed on larger patient groups to better assess if these differences could be detected with a single-color video stream. Thanks to the proposed methodology, the simple 2D parameters proposed here could be easily implemented in a user-friendly smartphone app for rehabilitation purposes. This would increase the number of PD patients undergoing monitoring and provide further clinical data to the clinicians. As the only significant differences between the groups were found for 3D parameters, we evaluate the trends of these measures during the task. Figure 5 shows a decrease of all the parameters along the syllable repetition, although from Table 3, a clear decrease of the regression line slope is present for v_{opening} and v_{closing} only. This could be due to speech-related fatigue that mainly occurs in PD patients. However, a recent study[35] shows that fatigue manifestations would not be so noticeable in PD patients. In our case, the performed speech task is too short to draw conclusions about speech-related fatigue, whereas in Reference 36, a speech task of 1 hour of duration was used. Nevertheless, we could apply our method to assess those findings from a kinematic point of view.

In contrast to Walsh and Smith,[18] we use a syllable repetition task, a widely used speech task already implemented in many studies on the kinematic analysis of the articulators in PD patients.[12,14,15] However, the analyzed bilabial movements are similar to those proposed there (/pa/ in our study, /paIp/ in Reference 18). Further developments will be devoted to the implementation of this technique to other speech tasks, in particular, those that involve spontaneous speech (passage reading or monologue).

Some considerations concerning the methodology should be drawn. Two main differences exist between the current work and our previous studies where we tested the accuracy of the markerless system[23,24]:

1. In References 23 and 24, we used the depth sensor Primesense Carmine 1.09 (Primesense LTD), whereas in the current study, we carried out the experiments by means of the Microsoft Kinect for Windows. However, both devices have very similar hardware, and also, the Kinect sensor used in our experiments can work in a near-range mode.[36]
2. To assess the accuracy of the proposed method, we used an image resolution of 320 × 240 pixels.[23,24] Although this is a quite low resolution for our purposes, promising results were found in terms of tracking errors of the lips during speech production. We concluded that higher resolutions should be adopted for the experiments that involve the study of the articulatory movements with 3D depth sensors. A recommendation was to use at least 640 × 480 pixels for both streams (color and depth) according to the device features.

As already demonstrated in Reference 23, the error introduced in the depth estimation by the sensor is fairly constant (around 1 mm) between 0.5 and 0.7 m far away from the camera. Considering these values and the working range of the Kinect (greater than 0.4 m), this interval is the optimal range in which to perform the experiments. Moreover, subjects were seated during the experiments; therefore, these boundaries were reasonably kept even in the case of PD patients with involuntary movements.

In Reference 24, we demonstrated that with low image resolutions, our method was able to accurately track the trajectories and the trends of velocity and acceleration of the lower lip. Despite these good results, with correlation coefficients over 0.95 for trajectories and velocities when compared with the marker-based reference, we noted that the markerless method underestimates the peak velocities of about 20 mm/s. In the present work, we cannot state if the estimation of the peak velocities and accelerations were underestimated as we have not yet compared our markerless method with a marker-based technology with higher image resolutions. Thus, further experiments will concern also this topic.

With this simple markerless system, we are able to track lip movements with good accuracy and detect significant differences in the kinematic parameters of the lower lip between PD patients and HC subjects. However, the main limitation of this system is its capability of tracking only the external articulators being based on color and infrared cameras. Thus, as speech involves a complex coordination of several articulators, complete and exhaustive results could be obtained only by combining markerless and marker-based techniques (for instance, EMA for studying tongue movements).

CONCLUSION

Our results confirm the findings that PD patients have reduced articulatory kinematic, adding further confirmations to this hypothesis. The major contribution of this work is the implementation of a fully markerless technique to analyze the articulatory movements during speech that is able to identify signs of the hypokinetic dysarthria. Several implications result from this work:

- This system could allow tracking the disease progression in patients with dysarthria also in a domestic environment. In fact, it requires just a personal computer and a cheap 3D depth sensor like the one used in this work.
- As PD patients exhibit a reduced articulatory kinematic, we support the fact that these patients should perform speech therapy exercises focused on the increasing of speech and articulatory movement efforts, as already stated by Walsh and Smith.[18] This technique is cost-effective and the use of efficient algorithms allows tracking lip movements in real time, thus, providing an immediate feedback about the right facial movements to be performed during speech therapy protocols. This could be obtained with the development of a graphical user interface displaying the actual extent of the articulatory movements as a visual feedback.

Further developments will be focused on the study of jaw movements, combining the two video streams of the Kinect sensor (color and depth) with computer vision algorithms. In fact, at present, we are able to track just lip movements. To provide a more detailed description of the kinematic parameters related to hypokinetic dysarthria, other articulatory organs should be studied. The methodology used in this work is based only on video analysis. A major drawback of this is the impossibility to study tongue movements that are very important in PD patients with hypokinetic dysarthria. To overcome this problem at least partially, we will analyze patients with hypokinetic dysarthria by combining acoustical and kinematic analyses. Moreover, we will study different speech task merging markerless and marker-based technologies for the analysis of the articulatory movements.

Acknowledgments

The authors greatly acknowledge the contribution of Ente Cassa di Risparmio di Firenze project n. 2013.0668 “A Markerless system for the automated analysis of audio and video signals as an aid to rehabilitation: application to the Parkinson’s disease”.

REFERENCES

1. Damier P, Hirsch EC, Agid Y, et al. The substantia nigra of the human brain II. Patterns of loss of dopamine-containing neurons in Parkinson’s disease. *Brain*. 1999;199:1437–1448.
2. Braak H, Del Tredici K, Rüb U, et al. Staging of brain pathology related to sporadic Parkinson’s disease. *Neurobiol Aging*. 2003;24:197–211.
3. Dickson JM, Grünewald RA. Somatic symptom progression in idiopathic Parkinson’s disease. *Parkinsonism Relat Disord*. 2004;10:487–492.
4. Dickson DW. Parkinson’s disease and parkinsonism: neuropathology. *Cold Spring Harb Perspect Med*. 2012;2:1–15.
5. Chaudhuri KR, Healy DG, Schapira AHV. Non-motor symptoms of Parkinson’s disease: diagnosis and management. *Lancet Neurol*. 2006;5:235–245.
6. Hartelius L, Svensson P. Speech and swallowing symptoms associated with Parkinson’s disease and multiple sclerosis: a survey. *Folia Phoniatr Logop*. 1994;46:9–17.
7. Tjiaden K. Speech and swallowing in Parkinson’s disease. *Top Geriatr Rehabil*. 2008;24:115–126.
8. Darley FL, Aronson AE, Brown JR. *Motor Speech Disorders*. Philadelphia, PA: Saunders; 1975.
9. Skodda S, Visser W, Schlegel U. Vowel articulation in Parkinson’s disease. *J Voice*. 2011;25:467–472.
10. Skodda S. Erratum: vowel articulation in Parkinson’s disease. *J Voice*. 2011;25:467–472.
11. Sapir S, Ramig LO, Spielman JL, et al. Formant Centralization Ratio (FCR): a proposal for a new acoustic measure of dysarthric speech. *J Speech Lang Hear Res*. 2010;53:1–20.
12. Caligiuri MP. Labial kinematics during speech in patients with Parkinsonian rigidity. *Brain*. 1987;110:1033–1044.
13. Forrest K, Weismer G, Turner GS. Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric speakers. *J Acoust Soc Am*. 1989;85:2608–2622.
14. Svensson P, Henningson C, Karlsson S. Speech motor control in Parkinson’s disease: a comparison between a clinical assessment protocol and a quantitative analysis of mandibular movements. *Folia Phoniatr (Basel)*. 1993;45:157–164.
15. Forrest K, Weismer G. Dynamic aspects of lower lip movement in Parkinsonian and neurologically normal geriatric speakers’ production of stress. *J Speech Hear Res*. 1995;38:260–272.
16. Yunusova Y, Weismer G, Westbury JR, et al. Articulatory movements during vowels in speakers with dysarthria and healthy controls. *J Speech Hear Res*. 2008;51:596–611.
17. Wong MN, Murdoch BE, Whelan BM. Lingual kinematics in dysarthric and nondysarthric speakers with Parkinson’s disease. *Parkinsons Dis*. 2011;2011:352838. doi:10.4061/2011/352838.
18. Walsh B, Smith A. Basic parameters of articulatory movements and acoustics in individuals with Parkinson’s disease. *Mov Disord*. 2012;27:843–850.
19. Ackermann H, Hertrich I, Daum I, et al. Kinematic analysis of articulatory movements in central motor disorders. *Mov Disord*. 1997;12:1019–1027.
20. Earnest MM, Max L. En route to the three-dimensional registration and analysis of speech movements: instrumental techniques for the study of articulatory kinematics. *Contemp Issues Commun Sci Disord*. 2003;30:5–25.
21. Tsanas A, Little MA, McSharry PE, et al. Accurate telemonitoring of Parkinson’s disease progression by a non-invasive speech test. *IEEE Trans Biomed Eng*. 2010;57:884–893.
22. Goetz CG, Stebbins GT, Wolff D, et al. Testing objective measures of motor impairment in early Parkinson’s disease: feasibility study of an at-home testing device. *Mov Disord*. 2008;24:551–556.
23. Bandini A, Ouni S, Cosi P, et al. Accuracy of a markerless acquisition technique for studying speech articulators. In: *Proceedings of the 16th Annual Conference of the International Speech Communication Association, INTERSPEECH, September 6–10, 2015*. Dresden, Germany: 2015 (in press).
24. Bandini A, Ouni S, Orlandi S, et al. Evaluating a markerless method for studying articulatory movements: application to a syllable repetition task. In: Manfredi C, ed. *Proceedings of the 9th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications MAVEBA, September 2–4, 2015*. Firenze: Italy Firenze University Press; 2015 (in press).
25. Goetz CG, Poewe W, Rascol O, et al. Movement disorder society task force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov Disord*. 2004;19:1020–1028.
26. Fahn S, Elton R. Recent development in Parkinson’s disease. In: Fahn S, Marsden CD, Calne DB, et al., eds. *Macmillan Health Care Information*. 2. Florham Park, NJ: 1987:153–163, 293–304.
27. Bandini A, Giovannelli F, Orlandi S, et al. Automatic identification of dysprosody in idiopathic Parkinson’s disease. *Biomed Signal Process Control*. 2015;17:47–54.

28. Skodda S, Flasskamp A, Schlegel U. Instability of syllable repetition as a marker of disease progression in Parkinson's disease: a longitudinal study. *Mov Disord*. 2011;1:59–64.
29. https://msdn.microsoft.com. Accessed August 20, 2015.
30. Heikkilä J, Silven O. A four-step camera calibration procedure with implicit image correction. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 17–19 1997; San Juan, Puerto Rico*. Los Alamitos, CA: IEEE Computer Society; 1997:1106–1112.
31. http://www.vision.caltech.edu/bouguetj/calib_doc/. Accessed August 20, 2015.
32. Xiong X, De la Torre F. Supervised descent method and its applications to face alignment. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23–28, 2013, Portland, OR, USA*. Los Alamitos, CA: IEEE Computer Society; 2013:532–539.
33. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis*. 2004;60:91–110.
34. Khoshelham K, Elberink SO. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors (Basel)*. 2012;12:1437–1454.
35. Makashay MJ, Cannard KR, Solomon NP. Speech-related fatigue and fatigability in Parkinson's disease. *Clin Linguist Phon*. 2015;29:27–45.
36. DiFilippo NM, Jouaneh MK. Characterization of different Microsoft Kinect sensor models. *Sensors (Basel)*. 2015;15:4554–4564.