Acoustic to kinematic projection in Parkinson's disease dysarthria

Acoustic to Kinematic Projection in Parkinson's Disease Dysarthria

A. Gómez[a,b], A. Tsanas[a], P. Gómez[b], D. Palacios[b,c], V. Rodellar[b], A. Álvarez[b]

[a]*Usher Institute, Medical School, University of Edinburgh, Old Medical School, Teviot Place Edinburgh, EH8 9AGUK*

[b]*NeuSpeLab, Center for Biomedical Technology, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28223, Pozuelo de Alarcón, Madrid, Spain*

[c]*Escuela Técnica Superior de Ingeniería Informática - Universidad Rey Juan Carlos, Campus de Móstoles, Tulipán, s/n, 28933 Móstoles, Madrid, Spain*

Corresponding author:

Andrés Gómez-Rodellar
Usher Institute
University of Edinburgh, UK
e-mail: s1859119@sms.ed.ac.uk

Abstract

Speech signal analysis is a powerful tool that facilitates the monitoring and tracking of symptom deterioration caused by neurodegenerative disorders, typically achieved using either sustained vowels, diadochokinetic exercises or running speech. This study expands our previous work on the study of the movement produced by the jaw-tongue biomechanical system. The aim is to further investigate the effects of neuromotor activity during muscular exertion that translates formant acoustics into speech articulatory movements affected by hypokinetic dysarthria in Parkinson's Disease (PD). The objective of this study is to estimate the parameters of an inverse acoustic-to-kinematic projection model that takes as an input the variations of the first and second formants and estimates as output the spatial variation of the jaw-tongue biomechanical system. The spatial variations have been extracted from 3D accelerometry (3DAcc). These serve as ground truth for comparison with the estimated activity projected from speech kinematics, as a measure of fitness of the inverse model. The estimation method is a two step process: first initial weight values are produced using multiple regression between each of the formant dynamic signals (acoustical analysis) and the estimated spatial variations (accelerometry). The second step uses a weight refinement method based on gradient-descent. Additionally, a time-realignment study has been carried out on the acoustic-to-kinematic projection model, based on the estimation of relative time displacements as to maximize the cross-correlation between signals. The study is complemented with an estimation of the model weights on a dataset from PD participants and Healthy Controls (HC). This methodology opens up new ways to investigate the underlying physiological voice production mechanism which may offer new insights into PD symptoms.

Keywords: Neuromotor diseases, speech articulation biomechanics, speech kinematics, speech neuromotor degeneration, remote monitoring, hypokinetic dysarthria.

1 Introduction

The influence of neurological and cognitive processes on speech is a well-established and recognized fact [1], [2], [3]. Many studies in the last decade have explored diverse signals such as EEG, MEG, fMRI and other non-invasive methods to provide new insights into the vocal production process [4], [5]. This is of particular interest when investigating

neurological diseases (cognitive and neuromotor) such as Alzheimer, PD, Amyotrophic Lateral Sclerosis (ALS), Huntington's Chorea, and others related [6]. Speech allows the contactless remote recording on smart terminals, as phones, tablets or laptop computers. Speech offers the added value of mapping acoustic estimates to neuromuscular activity, providing an advantage in the detection and monitoring of diseases dependent on neuromotor pathway transmission remotely [7]. A comprehensive study on the effects of PD on speech [8], [9] could provide insights into the underlying physiology, associating speech characteristics to the physical manifestations of the disease. This can be achieved through the study of phonation, articulation, prosody and fluency [10] which would offer valuable information on the activity of specific brain areas involved in speech production, such as motor planning, premotor and motor, and working memory. There is an unmet need to establish a robust and reliable methodology to map estimates extracted from the speech acoustics to motor actions in certain muscles involved in speech articulation and production. One such example is the masseter muscle, responsible of raising the lower jaw. Such a projection methodology is proposed in this research work to transform speech formant dynamics to articulatory kinematics [11], [12]. First proposals of an inverse model (relating formant dynamics and articulation) were presented; as a result several indicators were developed to encompass articulatory movements from speech alone (e.g. Absolute Kinematic Velocity, *AKV*) [13], [14]. The problem with this first attempts was the lack of a robust model parameter estimation. This led to further exploratory work, were the relationships between sEMG, accelerometry and Speech were investigated [15]. After an in-depth study of the affectations of PD on these biometric signals [16], the conclusions were applied to the characterization of PD hypokinetic dysarthria [17], [18].

The aim of the present study is to provide an insight into acoustic-to-kinematic projection, which could eventually allow to extract and transfer acoustically relevant articulation features to neuromotor actions, to be used in the characterization and monitoring neuromotor activity in specific diseases such as PD which is used as testbed in this study, this being the objective of future research already in progress. This approach can provide new insights into the physiological voice production mechanism and tentatively assign any effects of PD on specific vocal production model components. Such a model would add new semantic value over other standard approaches in the state of the art. In this regard, the previously proposed mapping model is reformulated in terms of time and space variables to allow a dynamic description of the model coefficients to be used in further mapping processes in remotely monitoring PD. This description includes also estimations from HCs.

The paper is structured as follows. In Section 2 the acoustic-mechanical model proposed is reformulated in the time-space domain to allow a more robust estimation of its coefficients. Section 3 is devoted to describe the data acquisition framework, platform and protocols, the biometrical data of participants, and the estimation methodology for the model coefficients in the space-time domain. The results derived from the PD and HC subsets are presented and described in Section 0. Section 5 is devoted to discuss the robustness of the methodology considering the results, and to analyze their impact and limitations in possible online applications. The study's key findings are summarized in Section 6.

2 The Neuromechanical Model of the Lower Jaw Articulation

The present study is based on a simplified jaw-tongue articulation model [16] which is known to be representative of PD dysarthria [18]. It allows to create a relationship between acoustic and kinematic variables relating the first two formants $\mathbf{F}=\{F_1, F_2\}$ to the horizontal and vertical coordinates $\mathbf{S}=\{x_r, y_r\}$ of the joint Jaw-Tongue Reference Point

(P_{rJT}) in the sagittal plane This point represents the center of moments of the biomechanical system integrated by the maxillary bone, tongue and facial tissues associated [17] (see Figure 1).

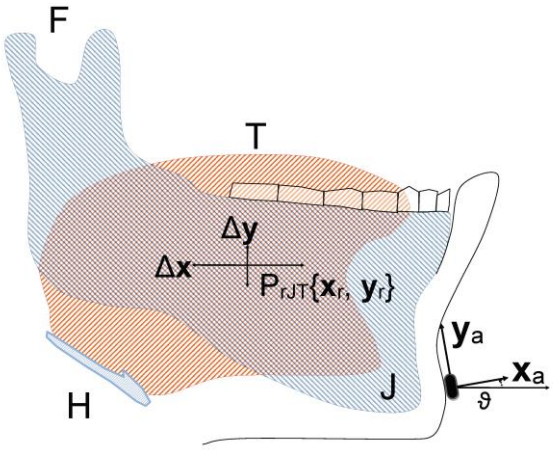

Figure 1. Jaw-tongue biomechanical model. The kinematic variables **S** = {s_1, s_2} = {x_r, y_r} are the horizontal and vertical PrJT coordinates. Their relative displacements with respect to the origin will be Δ**S** = {Δs_1, Δs_2} = {Δx, Δy}, represented as vectors in the time domain. Similarly {x_a, y_a} are the tangential and normal components of acceleration in the sagittal plane referred to the accelerometer coordinates, also vectors in the time domain. F: condylomaxillary joint, J: mandible bone, T: Tongue, H: hyoid bone.

The model assumes that a Linear Time-Invariant (LTI) relationship may be established between the P_{rJT} sagittal coordinates and the first two formant relative displacements, which may be summarized as

$$\begin{aligned} \Delta\mathbf{S} &= \mathbf{W} \times \Delta\mathbf{F}; \\ \mathbf{W} &= \{w_{ij}\}_{i=1,2}^{j=1,2} \end{aligned} \tag{1}$$

where Δ**S** = {Δs_1, Δs_2} is the vector of the horizontal and vertical displacements of PrJT in the time domain, which may be obtained from the rotation and integration of the tangential and normal acceleration components {x_a, y_a} on the participant's chin, **W** in (1) is a 2x2 matrix expressing the LTI projection model, which will be referred to as the acoustic-to-kinematic projection, and Δ**F** is the relative displacement in frequency of the first two formants with respect to their means in the time domain, as the first two formants are strongly associated with articulation kinematics [17], defined as

$$\Delta\mathbf{F} = \{\mathbf{F}_1 - mean(\mathbf{F}_1), \mathbf{F}_2 - mean(\mathbf{F}_2)\} \tag{2}$$

3 Materials and Methods

3.1 Data Acquisition framework

The study cohort comprises 8 PD participants (four male, four female, all Spanish native speakers, stage 2 in H&Y scale) who were recruited from a PD patient association in the metropolitan area of Madrid (Asociación de Pacientes de Parkinson de Alcorcón y Móstoles, APARKAM). For comparison purposes four male and four female healthy control participants have been included in the study. The biometrical data of participants are given in Table 1.

Table 1 Biometrical Description of the participants included in the study.

Label	Age	Gender	H&Y	State	Label	Age	Gender	H&Y	State
CM1	69	M	-	-	CF1	66	F	-	-
CM2	70	M	-	-	CF2	62	F	-	-
CM3	61	M	-	-	CF3	65	F	-	-
CM4	68	M	-	-	CF4	65	F	-	-
PM1	73	M	2	on	PF1	69	F	2	on
PM2	71	M	2	on	PF2	73	F	2	on
PM3	73	M	2	on	PF3	71	F	2	on
PM4	69	M	2	on	PF4	70	F	2	on

The study was approved by the Ethical Committee of UPM (MonParLoc, 18/06/2018). The voluntary participants were informed about the experiments to be conducted, the protection of their personal data and provided informed consent. The methodology was strictly aligned with the Declaration of Helsinki. Figure 2 presents an example of the process of multiple signal recording from PD patients.

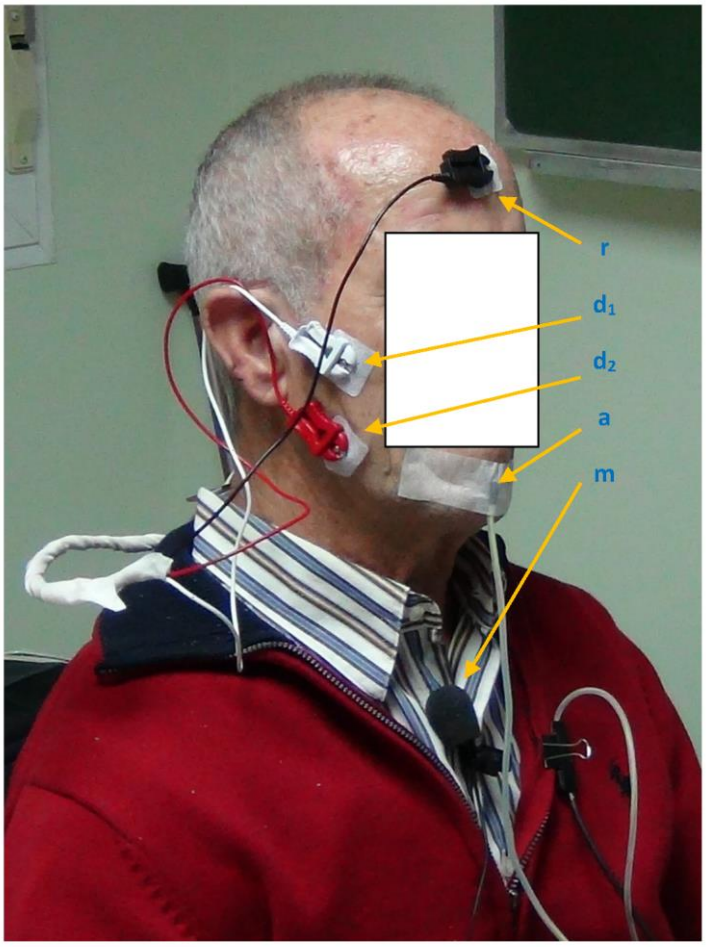

Figure 2. Signal acquisition set-up. Two sEMG electrodes are placed on the longitudinal ends of the masseter (differential pair: d_1, d_2) and one on the forehead (reference: r). The 3D accelerometer is fixed to the chin (a). A cardioid clip microphone is attached to the collar (m).

The equipment used allows the simultaneous and synchronous recording of surface electromyography (sEMG), 3DAcc, and speech, as illustrated in Figure 2. The sEMG is taken by the two attachments of the masseter complex to the jaw and skull, and the 3DAcc signals are obtained from an accelerometer attached to the chin. These signals were

digitized and collected with a Biopac MP150 EMG100 at 2 kHz and 16 bits. A Sennheiser cardioid wireless microphone (ew320 g2) on a MOTU Traveler MK1 sound card was used to record speech at 40 kHz with 32 bit resolution. Speech was latter down-sampled to 8 kHz for the analysis of the first and second formants, since the ranges for both formants are below 4 kHz [19]. The formant estimation is based on adaptive lattice filters [20] estimating a formant pair every 2 ms. Consequently, sEMG and 3D accelerometer signals were down-sampled to 500 Hz to match this time resolution. Recordings were carried out following a protocol that comprises the sustained vowels [*a*: *e*: *i*: *o*: *u*:], the fast repetition of the syllables [*pa*], [*ta*] and [*ka*], the three connected syllables [*pataka*] and [*pakata*] and the diphthong […*aja*…]. In the present study only the recordings from this last exercise were used in the estimation of **W** by multiple regression, because this diphthong produces the widest sweeps of formant dynamical patterns associated to the high-low and forward-backward displacement of P_{rJT}.

3.2 Formant estimation

The accurate estimation of the first two formants considered in **Error! Reference source not found.** is essential to the study. The procedures used in formant estimation are based on adaptive linear prediction [20], building on a previous in-depth study [16]. The details of formant estimation are briefly described as follows:

- The speech signal **x**(*n*) was bandlimited (low-pass filtered) to 4 kHz by a 4-th order Butterworth filter.
- Speech was divided in consecutive segment windows of 64 ms separated by a 2 ms stride (62 ms overlap). A Hamming window was used.
- A radiation-compensation first-order high-pass filter with a drop-off coefficient of 0.6 was used to remove radiation effects.
- The glottal formant was eliminated by a first-order inverse lattice filter [20].
- A 9-order inverse lattice filter estimated the error-predictor polynomial $H_k(z)$.
- The roots z_k of the error predictor polynomial $H_k(z= z_k)=0$ were estimated.
- The formants were obtained from the positive angles of z_k: $F_k=f_s.\varphi_k/\pi$, $\varphi_k>0$ (f_s: sampling frequency).
- The root modules were used as a quality factor for formant selection: $r_k=|z_k|>q_f$.

3.3 Data Processing Methods

The purpose of the model described in (1) is to allow an indirect estimation of the spatial oscillations $\Delta\mathbf{S}$ solely from the dynamics of the recorded signal and the acoustic formants $\Delta\mathbf{F}$ by an acoustic-to-kinematic model described by its weight matrix **W** which will be the main objective of this study

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \tag{3}$$

The individual weight values w_{ij} are to be estimated from healthy controls and PD participants, to establish possible regression models on the kinematic variables associated to the reference point P_{rJT} exclusively from acoustic estimates ($\Delta\mathbf{F}$), in other words, to establish a methodology for estimating articulatory kinematic features solely from the speech signal. The methodology proposed is based on solving for the model weights **W** using standard optimization methods to establish the relationship between the observed variables $\Delta\mathbf{S}$ and $\Delta\mathbf{F}$. The problem is stated as the minimization of a cost function *C*

$$C = \|\Delta\mathbf{S} - \mathbf{W} \times \Delta\mathbf{F}\|^2 \; subject\ to\ \mathbf{W}_{est} = argmin_{\mathbf{W}}\{C\} \tag{4}$$

where $\|\cdot\|^2$ denotes the module of a vector. Given the structural properties of C, the estimation of $\mathbf{W}$ may be decomposed in the independent minimization of each of its separate components ($C=C_1+C_2$)

$$C_{\mathrm{i}}(w_{\mathrm{i}1}, w_{\mathrm{i}2}) = \|\Delta\mathbf{s}_{\mathrm{i}} - w_{\mathrm{i}1}\Delta\mathbf{F}_1 - w_{\mathrm{i}2}\Delta\mathbf{F}_2\|^2; i = 1,2 \tag{5}$$

If expression (4) is expanded, it can be easily observed that the partial cost functions depend only on a single row of the matrix $\mathbf{W}$ (3). Therefore the error minimization problem can be split into minimizing each of the partial error functions $C_{\mathrm{i}}(w_{\mathrm{i}1},w_{\mathrm{i}2})$ for the weights $w_{\mathrm{i}1}$ and $w_{\mathrm{i}2}$.

$$[w_{\mathrm{i}1}, w_{\mathrm{i}2}] = argmin_{[w_{\mathrm{i}1},w_{\mathrm{i}2}]}\{C_{\mathrm{i}}\} \tag{6}$$

The minimization methodology is based on an iteration using gradient descent with a variable step size to estimate each individual weight as

$$w_{\mathrm{ij}}^{\mathrm{k}} = w_{\mathrm{ij}}^{\mathrm{k}-1} - \gamma_{\mathrm{i}}^{\mathrm{k}-1}\boldsymbol{\nabla}_{\mathrm{i}}C_{\mathrm{i}}^{\mathrm{k}-1};\ i,j = 1,2 \tag{7}$$

where k is the iteration step which can be estimated by the Barzilai–Borwein method [21]

$$\gamma_{\mathrm{ik}} = \frac{\langle \mathbf{w}_{\mathrm{i}}^{\mathrm{k}} - \mathbf{w}_{\mathrm{i}}^{\mathrm{k}-1}, \boldsymbol{\nabla}_{\mathrm{i}}C_{\mathrm{i}}^{\mathrm{k}} - \boldsymbol{\nabla}_{\mathrm{i}}C_{\mathrm{i}}^{\mathrm{k}-1}\rangle}{\left\|\boldsymbol{\nabla}_{\mathrm{i}}C_{\mathrm{i}}^{\mathrm{k}} - \boldsymbol{\nabla}_{\mathrm{i}}C_{\mathrm{i}}^{\mathrm{k}-1}\right\|^{\mathbf{2}}} \tag{8}$$

the weight and the gradient vectors being defined as

$$\begin{aligned} \mathbf{w}_{\mathrm{i}}^{\mathrm{k}} &= \left[w_{\mathrm{i}1}^{\mathrm{k}}, w_{\mathrm{i}2}^{\mathrm{k}}\right] \\ \boldsymbol{\nabla}_{\mathrm{i}} &= \left[\frac{\partial}{\partial w_{\mathrm{i}1}}, \frac{\partial}{\partial w_{\mathrm{i}2}}\right] \end{aligned} \tag{9}$$

Practical convergence is typically reached after a few iteration steps as shown in the next Section. The initial estimation (step k=0) for the weights $\mathbf{W}^0$ is achieved using simple linear regression [22] between the input and output signals of the inverse model

$$\mathbf{W}^0 = \begin{bmatrix}\mathbf{w}_1^0 \\ \mathbf{w}_2^0\end{bmatrix} = \begin{bmatrix} w_{11}^0 & w_{12}^0 \\ w_{21}^0 & w_{22}^0 \end{bmatrix} \tag{10}$$

$$w_{\mathrm{ij}}^0 = R_{ij}\left(\Delta\mathbf{s}_{\mathrm{i}}, \Delta\mathbf{F}_{\mathrm{j}}\right) = \frac{\Delta\mathbf{s}_{\mathrm{i}}\Delta\mathbf{F}_{\mathrm{j}}^{\mathrm{T}}}{\Delta\mathbf{F}_{\mathrm{j}}\Delta\mathbf{F}_{\mathrm{j}}^{\mathrm{T}}};\ i,j = 1,2 \tag{11}$$

This estimation process is represented in the regression plots shown in Section 0. It has been observed while estimating the initial values of the weights that there seems to be a misalignment between $\Delta\mathbf{s}_{\mathrm{i}}$ and $\Delta\mathbf{F}_{\mathrm{j}}$. To improve the estimation of the model matrix $\mathbf{W}$ it may be interesting to reduce this misalignment by maximizing the correlation function R_{ij} after introducing a shift. The relative misalignment may be a consequence of formant insertion dynamics associated to resonance in tubes with losses (this assumption needing further study), and it results in a non-optimal estimation of $\mathbf{W}$. To compensate it, each weight may be re-estimated after the realignment of signals derived from the following optimization problem

$$C_{\mathrm{i}}(z) = \|\Delta\mathbf{s}_{\mathrm{i}}(z) - w_{\mathrm{i}1}\Delta\mathbf{F}_1(z)z^{-n_{\mathrm{i}1}} - w_{\mathrm{i}2}\Delta\mathbf{F}_2(z)z^{-n_{\mathrm{i}2}}\|^2;\ i = 1,2 \tag{12}$$

where $\Delta\mathbf{s}_{\mathrm{i}}(z)$ and $\Delta\mathbf{F}_{\mathrm{i}}(z)$ are the z-transforms of $\Delta\mathbf{s}_{\mathrm{i}}$, and $\Delta\mathbf{F}_{\mathrm{i}}$ [20], and $n_{\mathrm{i}1}$, and $n_{\mathrm{i}2}$, are the relative misalignments between each of the components of $\Delta\mathbf{S}$ and $\Delta\mathbf{F}$, given in numbers of samples, assumed to be independent with each other. Similarly to the problem in (5), a solution is sought as

$$[n_{\mathrm{i}1}, n_{\mathrm{i}2}]_{min} = argmin_{[n_{\mathrm{i}1},n_{\mathrm{i}2}]}\{C_{\mathrm{i}}(z)\}; i = 1,2 \tag{13}$$

The independent minimization of $C_1(z)$ and $C_2(z)$ allows estimating the misalignments on ensuring that

$$n_{\mathrm{ij}} = argmax_{n_{\mathrm{ij}}}\{|\Delta\mathbf{s}_{\mathrm{i}}(n - n_{\mathrm{ij}})\Delta\mathbf{F}_{\mathrm{j}}(n)|\} \tag{14}$$

The alignment fitness may be evaluated by means of the root mean square error between the real displacement and the value predicted from regression for each weight w_{ij} as

$$\varepsilon_{\mathrm{ij}} = \frac{\|\mathbf{e}_{\mathrm{ij}}\|}{\|\Delta\mathbf{s}_{\mathrm{i}}\|};\ \mathbf{e}_{\mathrm{ij}} = \Delta\mathbf{s}_{\mathrm{i}} - w_{\mathrm{ij}}\Delta\mathbf{F}_{\mathrm{j}};\ i,j = 1,2 \tag{15}$$

4 Results

4.1 Data recording examples

The speech signal, the sEMG and the 3 acceleration channels from two repetitions of the […*aja*…] by a female HC participant (CF1) are shown in Figure 3 as an illustrative example. The sEMG signal has been included in the plots (channel b) with the purpose of witnessing that the acceleration and speech signals are concordant with the action of the masseter.

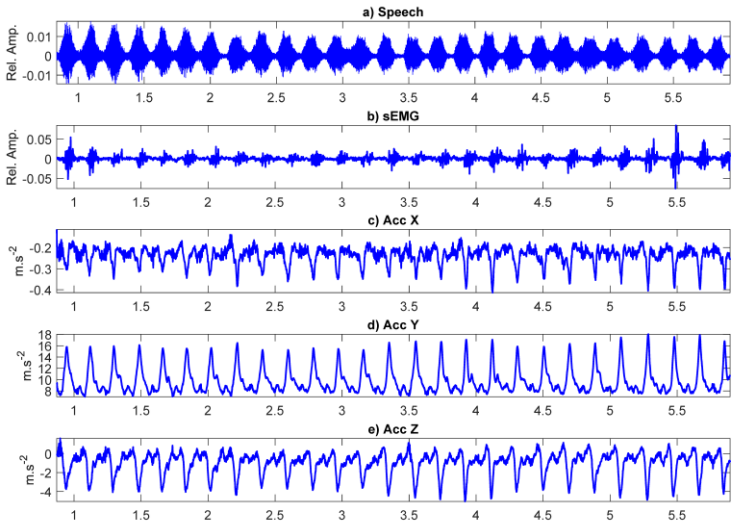

Figure 3. Signal acquisition example from the repetition of the phonetic sequence […*ajajaj*…] for a female control participant (CF1): a) speech signal; b) surface electromyographic signal on the masseter; c) channel X accelerometer signal; d) channel Y accelerometer signal; e) channel Z accelerometer signal.

Similarly, the same set of recordings from one of the PD female participants included in the study (PF1) are shown in Figure 4.

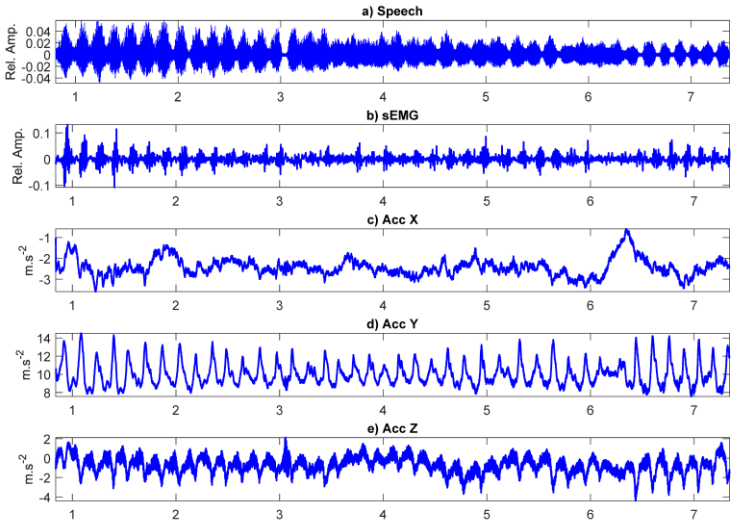

Figure 4. Signal acquisition example from the repetition of the phonetic sequence […*ajajaj*…] for a female PD participant PF1: a) speech signal; b) surface electromyographic signal; c) channel X (Acc); d) channel Y (Acc); e) channel Z (Acc).

The repetition of […*ajajaj*…] as fast as possible has been used in the present study to estimate the model parameters of matrix **W** in kinematic terms. It may be seen from Figure 3 and Figure 4 that the selection of the diadochokinetic repetition of […*ajajaj*…] presents the advantage of being mainly controlled by the action of the masseter, a powerful muscle producing good sEMG records, as it may be observed. The data used in this study are the first two formants derived from the speech recording produced by the fast repetition of the exercise. The unbiased and smoothened formants are to be compared with the jaw-tongue reference displacements obtained after rotation and integration of the acceleration signals [15]. As an example, the estimations of $\Delta\mathbf{S}$ and $\Delta\mathbf{F}$ from CF1 are given in Figure 5.

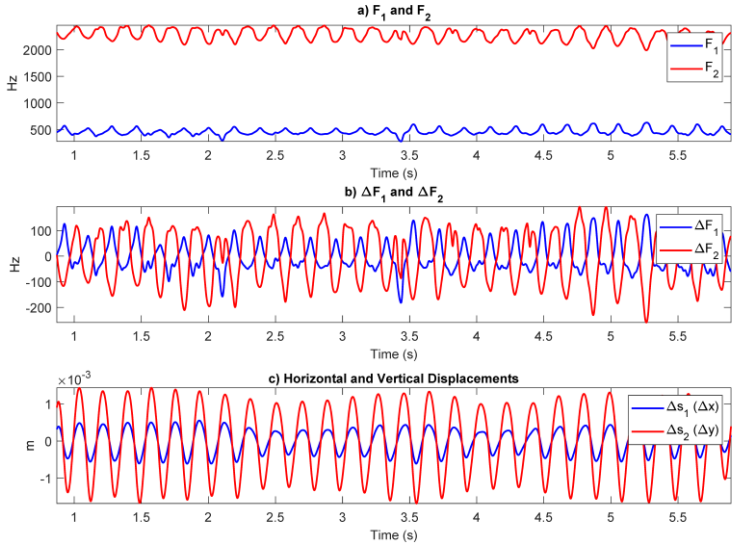

Figure 5. Formant deviations and reference point displacements obtained from CM1, corresponding to a HC participant: a) formants F1 and F2; b) formant deviations $\Delta\mathbf{F}$; c) reference point displacements $\Delta\mathbf{S}$.

The estimations of $\Delta\mathbf{S}$ and $\Delta\mathbf{F}$ corresponding to PF1 are shown in Figure 6.

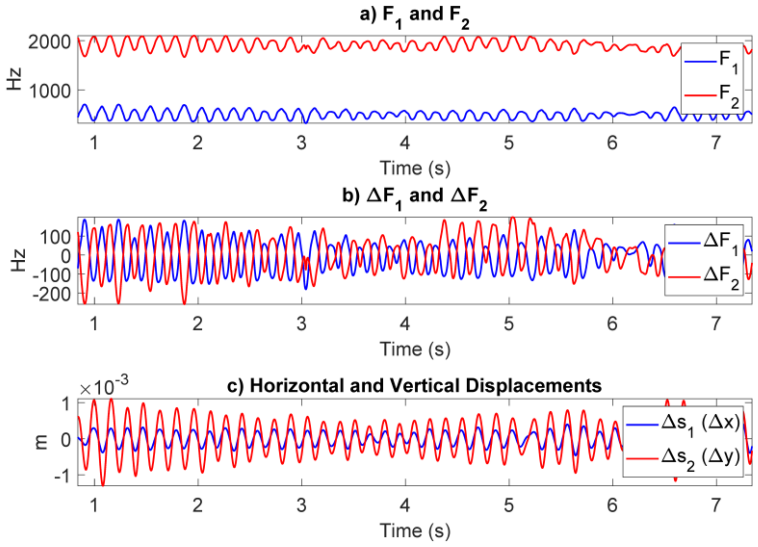

Figure 6. Formant deviations and reference point displacements obtained from PF1, corresponding to a PD participant: a) formants F1 and F2; b) formant deviations $\Delta\mathbf{F}$; c) reference point displacements $\Delta\mathbf{S}$.

4.2 Weight estimation from linear regression

The initial estimation of $\mathbf{W}^0$ is illustrated using a healthy control participant (CF1) in Figure 7. The scatter plots show the distribution patterns of each pair of $\Delta\mathbf{s}_i$, related to each pair of $\Delta\mathbf{F}_i$. A a regression line is fitted to each of these distributions with structure $\Delta\mathbf{s}_i = w_{ij}\ \Delta\mathbf{F}_i + b_{ij}$, as printed within each scatterplot. The slope of the regression line is the respective initial weight w_{ij} of matrix $\mathbf{W}^0$.

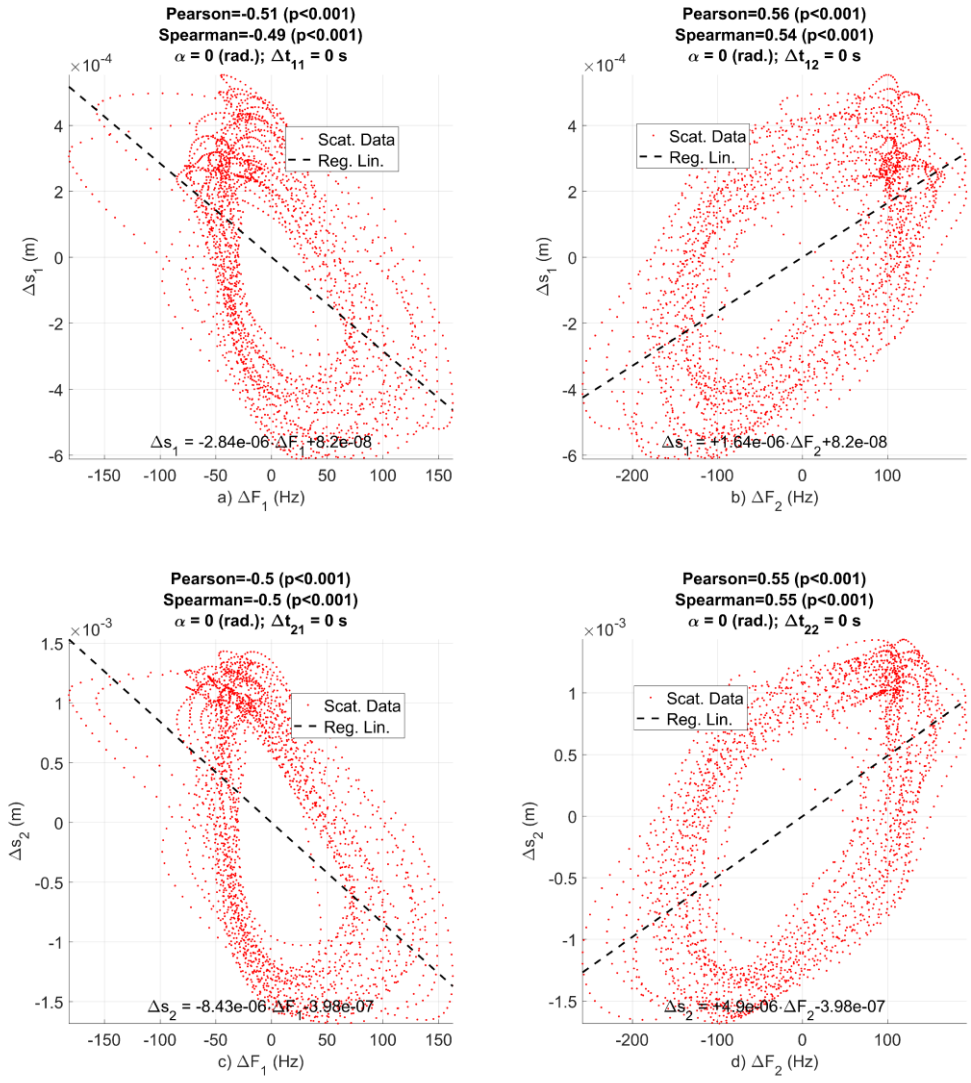

Figure 7. Scatter plots and regression results from CF1. The regression analysis is carried out for each pair of input signals (ΔF_i) and output signals (Δs_j): a) w_{11}^0=-2.84.10^{-6} m.Hz^{-1}; b) w_{12}^0=1.64.10^{-6} m.Hz^{-1}; c) w_{21}^0=-8.43.10^{-6} m.Hz^{-1}; d) w_{22}^0=4.90.10^{-6} m.Hz^{-1}.

This initial analysis shows what was expected from the hypothesized dynamic relation between formant dynamics (ΔF_i) and the kinematic outcome (Δs_j); the first formant (ΔF_1) increases with a descent and retraction of the P_{rJT}, whereas the value of the second formant is assumed to descend under the same movement conditions [16]. This is shown in the negative sign of the weights w_{11} and w_{21}, while in the case of w_{12} and w_{22} a positive sign is obtained. This method is applied to the signals from all male participants in the cohort and the healthy control (see **Error! Reference source not found.** for details), producing the initial results of the acoustic to kinematic projection given in Table 2, allowing for an initial inter-participant comparison.

Table 2 Male cases: Model weights and correlation coefficients per participant (P: Pearson; p-values <0.001); *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	$P_{\Delta x \Delta F1}$	$P_{\Delta x \Delta F2}$	$P_{\Delta y \Delta F1}$	$P_{\Delta y \Delta F2}$
CM1	-6.08	3.77	-9.90	6.31	-0.61	0.71	-0.57	0.68
CM2	-2.44	1.47	-2.71	1.56	-0.52	0.42	-0.43	0.33
CM3	-4.04	4.92	-3.45	4.34	-0.42	0.43	-0.36	0.38
CM4	-4.37	3.63	-5.65	4.71	-0.60	0.63	-0.63	0.66
PM1	-1.26	0.97	-4.56	3.45	-0.35	0.38	-0.36	0.38
PM2	-2.12	0.78	-1.05	0.45	-0.14	0.10	-0.23	0.19
PM3	-1.41	1.29	-3.17	2.88	-0.71	0.72	-0.84	0.85
PM4	-1.04	0.24	-1.13	-0.18	-0.30	0.09	-0.18	-0.04

The values of the model weights are accompanied by the correlation coefficients (Pearson) between each pair of signals, confirming the correlation relationships expected from the acoustic-to-kinematic projection properties (see further comments in Section 5). The same study has been conducted on the set of female participants (one HC and four PD participants) summarized in **Error! Reference source not found.**.

Table 3 Female cases: Model weights and correlation coefficients per participant (P: Pearson; p-values <0.001); *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	$P_{\Delta x \Delta F1}$	$P_{\Delta x \Delta F2}$	$P_{\Delta y \Delta F1}$	$P_{\Delta y \Delta F2}$
CF1	-2.84	1.64	-8.43	4.90	-0.51	0.56	-0.50	0.55
CF2	-3.78	2.15	-5.39	3.17	-0.47	0.68	-0.44	0.65
CF3	-2.15	1.19	-7.70	4.51	-0.72	0.73	-0.68	0.72
CF4	-0.28	0.30	-0.67	0.67	-0.32	0.37	-0.49	0.53
PF1	-2.09	1.56	-5.24	4.03	-0.81	0.75	-0.81	0.77
PF2	-1.46	1.15	-1.22	0.91	-0.54	0.52	-0.37	0.33
PF3	-1.88	1.75	-6.54	6.73	-0.24	0.17	-0.32	0.25
PF4	-1.52	0.95	-2.90	1.81	-0.54	0.55	-0.51	0.51

The values of the model weights are accompanied by the respective correlation coefficients (Pearson) between each pair of signals as before (more comments in Section 5).

4.3 Weight estimation from regression iteration (adaptive step = 0.5)

Following the initial weight estimation of $\mathbf{W}^0$, an iterative adjustment has been carried out. The procedure is based on the aforementioned iterative gradient-descent with variable step size (as described in (7) and (8)). This process aims to find a minimum of the error surfaces corresponding to the partial cost functions $C_i(w_{i1},w_{i2})$. The plots given in Figure 8 show this process illustrated for the control participant CF1. The trend of the descent for the pair of weights $\mathbf{w}_i^k = (w_{i1}^k, w_{i2}^k)$ can be observed, as the estimation for the k-iteration can be represented as a point on the surface C_i.

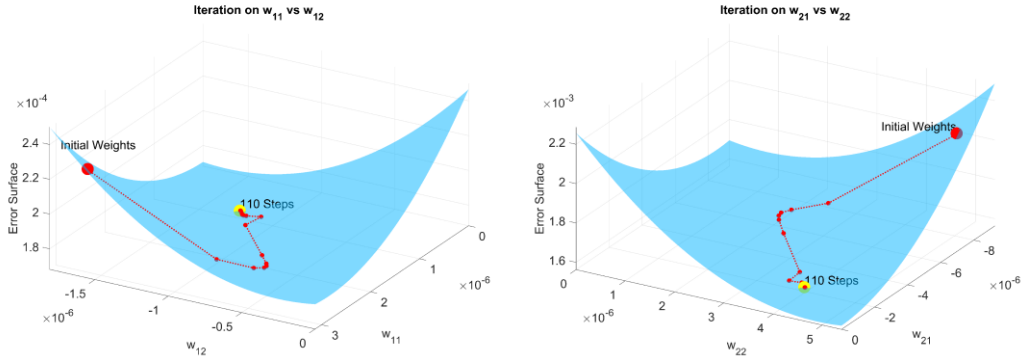

Figure 8. Error surfaces $C_1(w_{11}, w_{12})$ and $C_2(w_{21}, w_{22})$ corresponding to the iteration process on participant CF1. Left: $C_1(w_{11}, w_{12})$. Right: $C_2(w_{21}, w_{22})$. The starting position (in red) shows the values of the weights obtained from linear regression $\{w_{11}^0, w_{12}^0\}$ and $\{w_{21}^0, w_{22}^0\}$, whereas the stop position (in yellow) corresponds to the values of the weights after the iteration refinement.

It may be seen that although the error surfaces show a similar behavior to Rosenbrock's function [23], displaying a kind of *wadi*–shaped shallow valley, the variable step tracker based on the Barzilai–Borwein method is capable of reaching the minimum point of the curve in a reasonable number of iteration steps (in this case 110). The new weights after iteration refinement are given in Table 4.

Table 4 Male cases: model weights, number of iterations, and error reduction (ΔE, in percent), *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	No. Ite.	ΔE (%)
CM1	-0.71	-4.11	2.44	7.46	142	23.7
CM2	3.86	1.16	-4.85	-1.74	142	10.68
CM3	1.49	-3.32	-0.70	3.59	123	7.37
CM4	1.07	-2.83	-1.17	3.84	142	21.67
PM1	-0.25	0.82	-1.30	2.63	111	6.05
PM2	-6.24	-2.31	-2.13	-0.61	193	1.37
PM3	-0.55	0.84	-1.37	1.75	140	41.58
PM4	-2.37	-1.20	-4.09	-2.68	101	5.05

The results of the iteration refinement from the female participants are given in Table 5[1].

Table 5 Female cases: model weights, number of iterations and error reduction (ΔE, in percent), *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	No. Ite.	ΔE (%)
CF1	-0.82	1.28	-2.27	3.89	110	12.82
CF2	-0.03	2.14	0.31	3.25	108	14.45
CF3	-1.26	0.71	-3.85	3.06	88	22.75
CF4	-0.15	-0.43	0.09	0.75	153	11.55
PF1	-1.94	0.14	-4.08	1.03	132	38.47
PF2	-0.97	0.47	-1.03	0.19	111	8.07
PF3	-1.80	0.17	-5.69	1.73	77	2.34
PF4	-0.70	0.55	-1.36	1.04	163	13.00

[1]The iterative process stops once the gradient change becomes negligible (<10^{-6}).

The values of these weights would be the basis of a study towards a definition of a possible unified weight model for HC and PD participants, which is left as a future line.

4.4 Time realignment

When comparing the signals it was observed that the input $\Delta\mathbf{F}$ and output $\Delta\mathbf{S}$ showed similar patterns (number of cycles and periods), but there appeared to be a misalignment between them. The realignment method is based on maximizing correlations between $\Delta\mathbf{S}$ and $\Delta\mathbf{F}$ following (12)-(14). This process has been performed on the same female HC participant, the resulting changes from the initial estimation (**Error! Reference source not found.**) in the scatter plots and regression analysis can be observed in Figure 9.

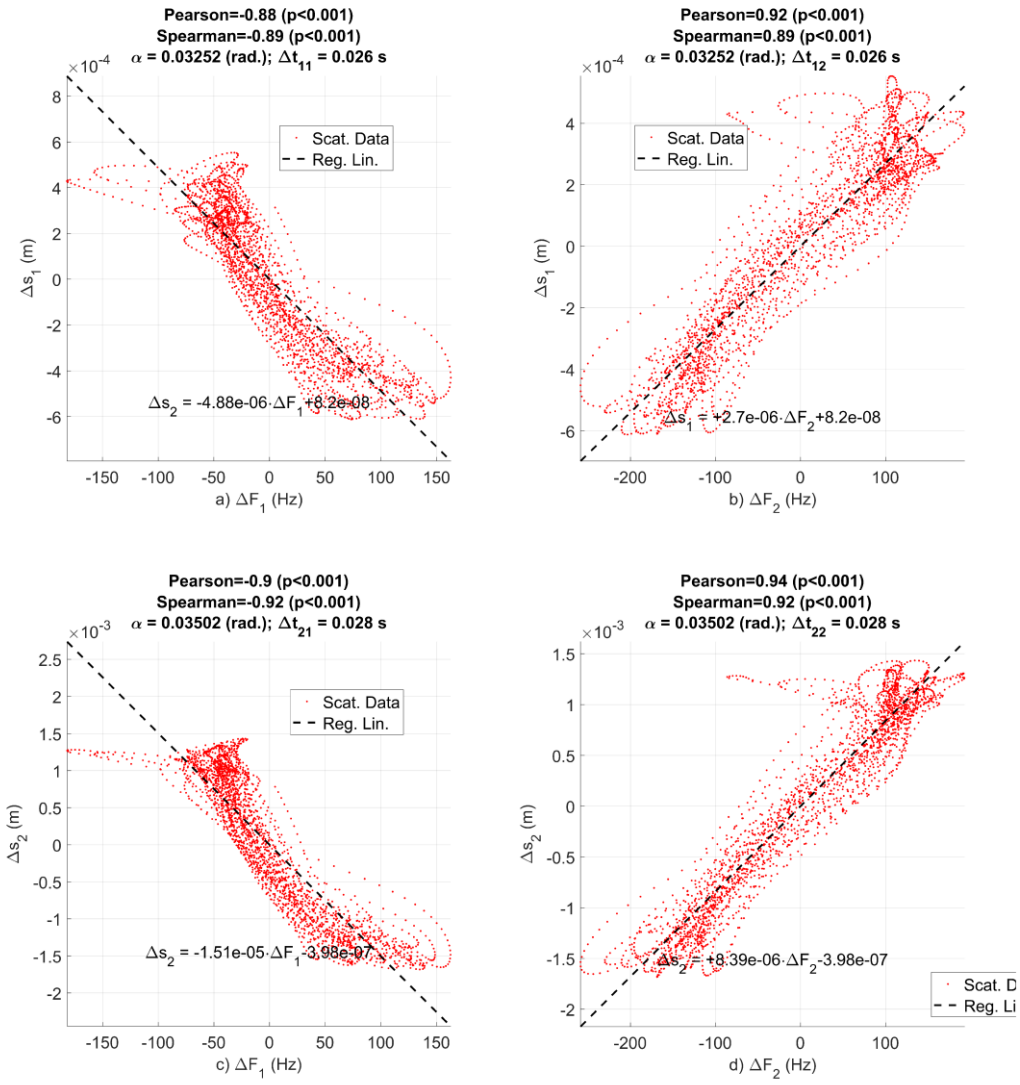

Figure 9. Scatter plots and regression results from CF1 after signal realignment denoted as $R'(\Delta F_i,\Delta s_j)$: a) $R'(\Delta F_1,\Delta s_1;)$: w_{11} =-4.88.10^{-6}; b) $R'(\Delta F_2,\Delta s_1)$: w_{12}=2.70.10^{-6}; c) $R'(\Delta F_1,\Delta s_2)$: w_{21}=-1.51.10^{-5}; d) $R'(\Delta F_2,\Delta s_2)$: w_{22}=8.39.10^{-6}. The size of the realignment time shift is given as Δt in seconds (Δt_{11}=26 ms, Δt_{12}=26 ms, Δt_{21}=28 ms, Δt_{22}=28 ms). The coefficients w_{ij} are given in cm.Hz^{-1}.

As it may be seen the realignment has reduced sensibly the dispersion of data in the new scatter plots, making the relationship between $\Delta\mathbf{F}$ and $\Delta\mathbf{S}$ more linear, as the dispersion along the perpendicular dimension to the regression line has been reduced (see the relative

quadratic errors after realignment in Table 6 and Table 7). The scatter plots and regression analysis from a male PD participant (PF1) are given in Figure 10 as a complementary example to be contrasted with Figure 9 of the HC participant.

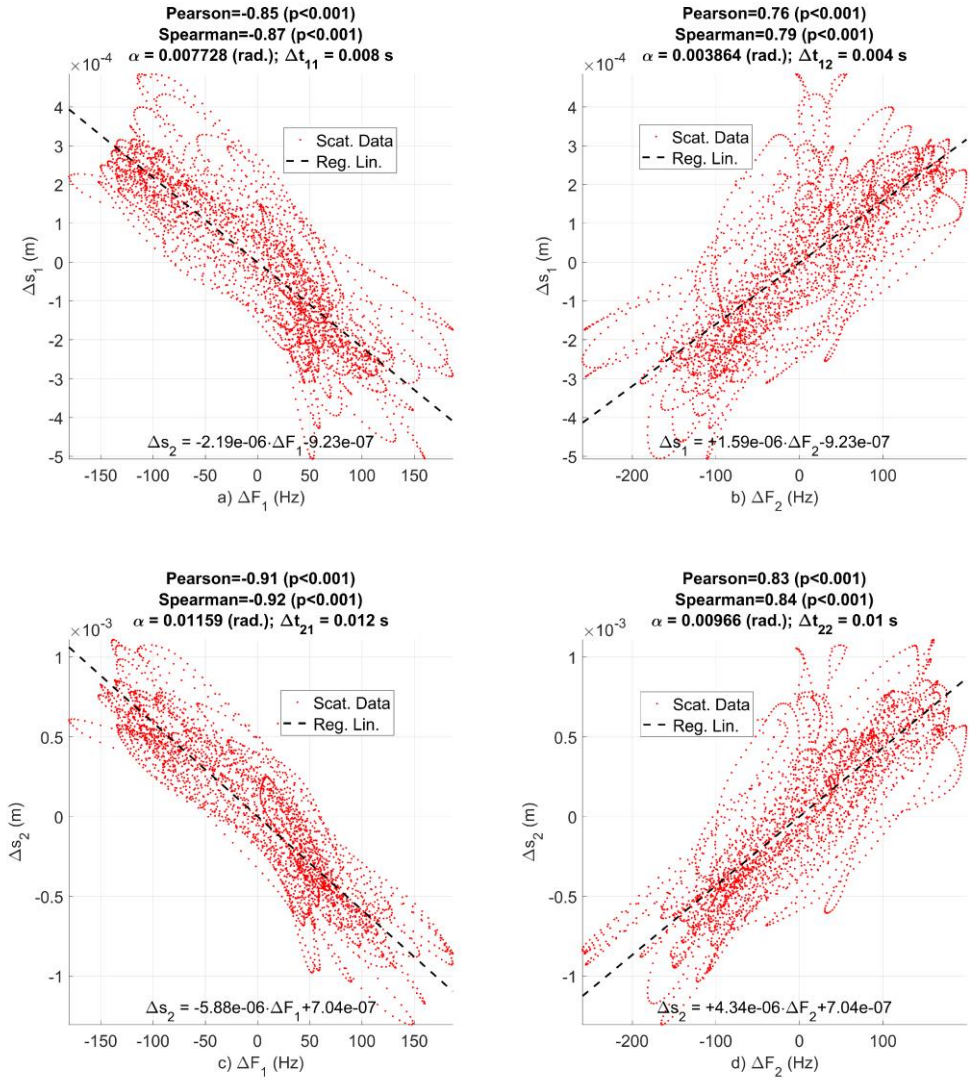

Figure 10. Scatter plots and regression results from PF1 after realignment: a) $R'(\Delta F_1,\Delta x)$; b) $R'(\Delta F_2,\Delta x)$; c) $R'(\Delta F_1,\Delta y)$; d) $R'(\Delta F_2,\Delta y)$. The size of the realignment time shift is given as Δt in seconds (Δt_{11}=8 ms, Δt_{12}=4 ms, Δt_{21}=12 ms, Δt_{22}=10 ms).

After realignment the same cross-correlation analysis (as in the one shown in Table 2 and **Error! Reference source not found.**) is then carried out for the male and female datasets. The results are shown in **Error! Reference source not found.** and **Error! Reference source not found.**.

Table 6 Male cases: Model weights, correlation coefficients and relative rms errors after realignment per participant (P: Pearson; p-values <0.001; ε_r: relative rms error in %); *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	$P_{\Delta x \Delta F1}$	$P_{\Delta x \Delta F2}$	$P_{\Delta y \Delta F1}$	$P_{\Delta y \Delta F2}$	$\varepsilon_{r\Delta x \Delta F1}$	$\varepsilon_{r\Delta x \Delta F2}$	$\varepsilon_{r\Delta y \Delta F1}$	$\varepsilon_{r\Delta y \Delta F2}$
CM1	-8.27	4.51	-14.53	7.97	-0.83	0.85	-0.84	0.86	0.56	0.53	0.54	0.51
CM2	-4.19	3.08	-5.75	4.20	-0.89	0.87	-0.90	0.89	0.46	0.49	0.43	0.47
CM3	-8.90	9.83	-9.12	10.12	-0.92	0.86	-0.94	0.88	0.38	0.51	0.34	0.47
CM4	-6.42	4.81	-8.23	6.21	-0.89	0.83	-0.92	0.87	0.46	0.56	0.39	0.50
PM1	-3.15	2.04	-11.08	7.08	-0.87	0.80	-0.87	0.78	0.49	0.61	0.50	0.63
PM2	-14.30	7.24	-4.08	2.05	-0.93	0.89	-0.90	0.85	0.38	0.46	0.45	0.52
PM3	-1.62	1.43	-3.40	3.02	-0.81	0.80	-0.90	0.89	0.59	0.60	0.44	0.45
PM4	-2.81	1.96	-5.15	-3.47	-0.81	0.76	-0.80	-0.73	0.59	0.65	0.60	0.69

Table 7 Female cases: Model weights, correlation coefficients and relative rms errors after realignment per participant (P: Pearson; p-values <0.001; ε_r: relative rms error in %); *x10^{-6} cm.Hz^{-1}.

Participant Labels	w_{11}*	w_{12}*	w_{21}*	w_{22}*	$P_{\Delta x \Delta F1}$	$P_{\Delta x \Delta F2}$	$P_{\Delta y \Delta F1}$	$P_{\Delta y \Delta F2}$	$\varepsilon_{r\Delta x \Delta F1}$	$\varepsilon_{r\Delta x \Delta F2}$	$\varepsilon_{r\Delta y \Delta F1}$	$\varepsilon_{r\Delta y \Delta F2}$
CF1	-4.88	2.70	-15.05	8.40	-0.88	0.92	-0.90	0.94	0.48	0.40	0.44	**0.33**
CF2	-6.34	2.65	-9.86	4.16	-0.79	0.84	-0.80	0.85	0.61	0.55	0.60	0.52
CF3	-2.24	1.20	-9.00	4.94	-0.75	0.73	-0.79	0.79	0.66	0.68	0.61	0.61
CF4	-0.65	0.63	-1.07	1.03	-0.75	0.78	-0.79	0.82	0.67	0.63	0.61	0.57
PF1	-2.19	1.59	-5.88	4.34	-0.85	0.76	-0.91	0.83	0.53	0.65	0.41	0.56
PF2	-2.12	1.75	-2.62	2.14	-0.79	0.79	-0.78	0.78	0.62	0.61	0.62	0.63
PF3	-5.39	7.75	-14.04	19.95	-0.68	0.74	-0.70	0.74	**0.73**	0.68	0.72	0.67
PF4	-2.31	1.44	-4.86	3.04	-0.82	0.83	-0.85	0.86	0.58	**0.73**	0.58	0.72

A further comparison between the model weights may be carried on normalizing each weight set w_{ij} to its vector norm as $\hat{w}_{ij}=w_{ij}/|w_{ij}|$. The results of the normalization are shown in **Error! Reference source not found.**.

Table 8 Weight normalization results. A scale factor of 10^{-6} cm.Hz^{-1} is to be assumed.

Male Set	$\hat{w}_{11}$	$\hat{w}_{12}$	$\hat{w}_{21}$	$\hat{w}_{22}$	Female Set	$\hat{w}_{11}$	$\hat{w}_{12}$	$\hat{w}_{21}$	$\hat{w}_{22}$
CM1	-0.43	0.24	-0.76	0.42	CF1	-0.27	0.15	-0.83	0.46
CM2	-0.48	0.35	-0.65	0.48	**CF2**	-0.50	0.21	-0.78	0.33
CM3	-0.47	0.52	-0.48	0.53	**CF3**	-0.21	0.11	-0.85	0.47
CM4	-0.49	0.37	-0.63	0.48	CF4	-0.38	0.36	-0.62	0.59
PM1	-0.23	0.15	-0.81	0.52	**PF1**	-0.28	0.20	-0.75	0.56
PM2	-0.86	0.43	-0.24	0.12	PF2	-0.49	0.40	-0.60	0.49
PM3	-0.32	0.28	-0.68	0.60	PF3	-0.21	0.30	-0.54	0.76
PM4	-0.40	0.28	-0.73	-0.49	PF4	-0.36	0.23	-0.77	0.48

Mann-Whitney tests between the normalized weights from the HC and the PD samples failed to reject the null hypothesis of equal means $\mu(\hat{w}_{ij})$ with a p-value of 0.965. A similar test between the normalized weights from the male and female participants failed to reject the null hypothesis of equal means with a p-value of 0.904. These results indicate that a general model may be built independently of gender and alteration condition, pending on a generalization relying on a larger sample database. The medians of the normalized weights may serve as a robust estimation of the model weight matrix $\{\hat{w}_{11}=-0.39, \hat{w}_{12}=0.28, \hat{w}_{21}=-0.70, \hat{w}_{22}=0.48\}x10^{-6}$ cm.Hz^{-1}.

The realignment sample shifts (n_{ij}) expressed as time shifts (in ms) are given in Table 9.

Table 9 Realignment time shifts between Δ**S** and Δ**F** per participant in ms. Delays under 24 ms are marked in bold.

Male Set	Δt_{11}	Δt_{12}	Δt_{21}	Δt_{22}	Female Set	Δt_{11}	Δt_{12}	Δt_{21}	Δt_{22}
CM1	**18**	**14**	**20**	**16**	CF1	26	26	28	28
CM2	26	28	28	32	**CF2**	**20**	**14**	**22**	**16**
CM3	28	26	30	28	**CF3**	**8**	**4**	**14**	**10**
CM4	30	26	30	24	CF4	28	26	22	20
PM1	28	26	28	26	**PF1**	**8**	**4**	**12**	**10**
PM2	34	36	32	32	PF2	26	28	36	38
PM3	**16**	**14**	**12**	**10**	PF3	28	32	26	30
PM4	32	38	36	**-40**	PF4	26	26	28	28

4.5 Formant Dynamics and Articulation Kinematics

From the regression study results it may be observed how the different magnitudes (Δ**S** and Δ**F**) relate to each other. Based on these observations a transformation function is defined by **W**, projecting formant dynamics to spatial displacements. An interesting indicator to compare between speakers is to observe the ranges that they are able to produce in these spaces. Table 10 shows the range of variation covered by Δ**S** and Δ**F**. The ranges are estimated by the 0.05 to 0.095 interquartile distance.

Table 10 Formant (Hz) and displacement (mm) ranges per participant $r(\cdot)$.

Males	$r(\Delta x)$	$r(\Delta y)$	$r(\Delta F1)$	$r(\Delta F2)$	Females	$r(\Delta x)$	$r(\Delta y)$	$r(\Delta F1)$	$r(\Delta F2)$
CM1	1.37	2.27	115	203	CF1	0.97	2.88	178	307
CM2	0.74	0.81	163	239	CF2	1.34	1.96	169	435
CM3	1.90	1.65	229	187	CF3	0.63	2.33	180	282
CM4	2.60	3.37	308	380	CF4	0.13	0.29	213	225
PM1	0.78	2.8	284	360	PF1	1.00	2.55	246	309
PM2	1.11	0.58	276	495	PF2	1.52	1.24	521	725
PM3	0.84	1.88	255	314	PF3	0.43	1.56	156	118
PM4	0.28	0.22	213	279	PF4	0.99	1.89	324	586

It may be observed that the size of the ranges shows a broad direct relationship between the formant and displacement oscillation ranges. Whether this observation could be the basis to define markers of hypokinetic dysarthria is subject to further study.

5 Discussion

In Section 3 an inverse linear model based on an acoustic to kinematic projection has been presented. This model has been validated by the results shown in Section 0, consequently the following findings may be highlighted:

- The relationship between acoustic to kinematic variables (Δ**F** to Δ**S**) has been established and may be explained using the inverse model described in expression (1).
- The initial estimation of the model weights has been carried out using least squares linear regression.
- A gradient-descent method usign a variable step size has been used in the iterative refinement of model weights to minimize the error cost function implicit in the inverse model.

- In order to linearize the relationship between the acoustic and kinematic estimates in the time domain a realignment procedure has been introduced with a considerable improvement of correlation results.

Figure 3 and Figure 4 present the speech, surface electromyography and X, Y and Z accelerations in the accelerometer system of coordinates, which is projected onto the one of the sagittal plane shown in Figure 1 (the X in the accelerometer system corresponds to the coordinate normal to the sagittal plane, the accelerometer Y corresponds with the sagittal y (s_2), and the accelerometer Z corresponds to the sagittal x (s_1)). It may be seen that each sEMG burst (channel b) is followed with an abrupt spike-like in the accelerometer axes Y and Z (channels d and e), whereas the accelerometer X (channel c) shows a much slower activity. Correspondingly the amplitude of the speech signal (channel a) shows a strong reduction immediately after each sEMG burst, as the activity of the masseter reduces the opening of the radiation end, and less energy is projected outwards, this observation being aligned with what it could be expected. The main difference found between both figures is that the interval cadence and the amplitude and pattern of the sEMG and X, Y and Z accelerometer signals are more regular in Figure 3 (corresponding to an HC female participant) that those in Figure 4 (corresponding to a PD female participant). The behavior presented in both figures cannot be generalized, but it may give a graphical view of what is being measured. We had described sEMG as part of the battery of signals that have been collected. It is true that sEMG is not used for the construction of the model, however it is useful to visualize it along with the speech signal and accelerometer data to provide an overview of this multimodal representation. For example, we present sEMG in Figures 3 and 4 as an illustration of how the raw data appear. We believe there is value in presenting this information here, even though we are not formally proceeding in full exploratory analysis of the use of sEMG further in this study; we plan to investigate this area further in future work.

Figure 5 and Figure 6 present the results from estimating the acoustic and kinematic variables in the sagittal plane ($\Delta\mathbf{F}$ to $\Delta\mathbf{S}$) from the same participants. Interestingly, more regularity may be observed in the estimates from the HC participant than in those from the PD participant. A closer observation to the relationship between acoustic to kinematic variables from the HC participant (CM1) is presented as scatter plots in Figure 7, from the regression association of the signals in Figure 5 (b and c). It may be observed that all the plots show a loop-like pattern associated to phase shifts between each pair of acoustic and kinematic variables. This is due to time misalignments resulting from formant dynamics, and explains the modest values of Pearson's correlation coefficients given in the four rightmost columns of Table 2 and Table 3.

The relationship between acoustic to kinematic variables ($\Delta\mathbf{F}$ to $\Delta\mathbf{S}$) given in both tables, expressed by the weights obtained from least squares linear regression requires a detailed analysis. The weights w_{11} and w_{12}, relate the first and second formant increments ΔF_1 and ΔF_2 with the horizontal displacement Δs_1. Weights w_{11} are negative and weights w_{12} are positive, relating a forward horizontal displacement of the jaw-tongue reference point with a descent of F_1 and with an ascent of F_2. Similar relationships may be observed on weights w_{21} and w_{22}, with respect to the first and second formant increments ΔF_1 and ΔF_2 regarding the vertical displacement Δs_2. In this case, w_{21} is always negative, and w_{22} is always positive, the upwards movement of the jaw-tongue reference point is related to a descent of F_1 and to an ascent of F_2. This behavior is aligned with the prediction of the acoustic-to-kinematic projection in the sense that increments of the first formant and decrements in the second formant are associated with the vertical pull up action of the masseter (negative values of w_{21} and positive values of w_{22}). A reflection is due at this

point in respect to the classical convention under the assumption of independent movements of jaw and tongue in static vowel positions. The underlying phenomenon is a bit more complicated when studying dynamic diphthong movements, as jaw and tongue cannot be considered moving independently. This is particularly relevant regarding the diadochokinetic exercise used in the study. As examples of non-independent movement, it must be considered that depending on the position of the tongue (back or front), the sole movement of the jaw may produce the diphthong [wa] as in /wah-wah/ when the position of the tongue is back (static), or the diphthong [jeə] as in /yeah/ when the tongue position is front (static). In the first case both F1 and F2 ascend to higher values when the jaw descends, whereas in the second case F1 ascends and F2 descends when the jaw descends. In both cases the tongue has not changed its position, but both formants move, as the jaw per se may modify completely the oral cavity, conditioning the movement of both formants. Conversely, should the jaw be kept in a stable medial position, the tongue per se could produce the diphthong [jʊ] like in /you/, where both formants descend from high to low values without the intervention of the jaw. These observations question the conventional view of independent relationships among dynamic formant movements and tongue and jaw positions, showing that the whole configuration of jaw and tongue is responsible for the production of important changes in formant positions, each system independently. In the present study no independent movement of jaw and tongue has been assumed.

It may be said about the gradient-descent iteration dynamics expressed in Figure 8 that the patterns shown by the error surfaces of E1 and E2 are quite similar, and correspond to a convex surface with a single minimum, in the shape of a *wadi* producing a large descent at the beginning followed by shorter descent steps once the bottom of the *wadi* is approached. This effect may produce some unstable predictions of the step size in (8). The shape of this narrow valley distorts the space of solutions, as their geometrical place is the set of possible values of the pairs of coefficients $\{w_{11}, w_{12}\}$ and $\{w_{21}, w_{22}\}$. Slight variations in the estimation conditions may lead to different numerical solutions, all of them sharing the property of producing a quasi-optimal approximation. The shape of this geometrical place is a kind of narrow ellipse, approaching in the limit a straight line: $w_{12}=m_1.w_{11}+b_1$; $w_{22}=m_2.w_{21}+b_2$. The results of the estimation refinements in the model weights after the iteration process, given in Table 4 and Table 5 are modest, as expressed by the relative error reduction in percent given in the rightmost columns. Reductions larger than 20% have been highlighted in bold.

The realignment process, exemplified in Figure 9 and Figure 10 from a female HC participant (CF1) and a female PD participant (PF1), produces a substantial increment in the correlation coefficients at the cost of introducing a delay, which in the case of the HC participant is around 26-28 ms, whereas in the case of the PD participant it is much shorter (between 4-12 ms). In this second case substantial increments in the correlation coefficients are also observed. This different behavior may be explained by resonance dynamics in non-rigid tubes with losses. It may be said that realignment improves correlation in all the HC and PD cases studied.

Regarding the model weights after realignment, as given in **Error! Reference source not found.** and **Error! Reference source not found.** compared to those before realignment in Table 2 and **Error! Reference source not found.**, it may be seen that realignment does not change acoustic-to-kinematic projection properties of the model, as displacements and formant oscillations maintain the relative concordance observed before the realignment. A quadratic relative error between the horizontal and vertical reference point displacements estimated from accelerometry, and the regression-predicted values as

obtained from expression (15) has been calculated for each model weight after signal realignment. These errors are reported in the rightmost columns of **Error! Reference source not found.** and **Error! Reference source not found.**. These errors are larger for the cases where the distribution of $\Delta\mathbf{S}$ is more dispersed with respect to the regression line $w_{ij}\Delta\mathbf{F}$, and therefore they serve as an indication of the goodness of fit. The best case corresponds to the prediction of ΔS_2 relative to ΔF_1 from CF1 (0.33), and the worst case (0.73) corresponds to PF3 and PF4 (ΔS_1 vs ΔF_1 and ΔS_1 vs ΔF_2, respectively). The HC subset behaves slightly better (0.52 ± 0.09) than the PD subset (0.58 ± 0.10), although given the small sample sizes this observation needs to be further verified on an external larger cohort.

An important remark comes from the observation that no relevant differences may be observed in the general pattern of the normalized model weights between HC and PD subsets given in Table 8, which allows for the definition of an overall average model, as shown in Section 4.4.

The realignment shifts (Δt_{11}, Δt_{12}, Δt_{21} and $\Delta t22$) associated to pair-wise weights $\{w_{11}, w_{12}\}$ and $\{w_{21}, w_{22}\}$ shown in Table 9 are in most cases within the range from 24-40 ms with some exceptions marked in bold (CM1, PM3, CF2, CF3 and PF1), and do not show relevant intra-speaker differences. All of them are multiples of the formant estimation time sampling rate of 2 ms. Their origin may be a consequence of algorithmic delays introduced in the insertion of formants as a result of resonance effects in non-rigid tubes with losses. The narrower the pole bandwidth associated to the formant, the shorter the time interval for the resonance format to grow in amplitude. As pole bandwidths are associated with the viscoelastic properties of the resonant cavities (oro-pharyngeal tract) more rigid and less viscid tissues would produce sharper poles and faster formant insertion, in opposition to more elastic and viscid tissues, producing duller poles and slower formant insertion, explaining the differences found in Table 9. It is known that the alterations in the viscoelastic properties of mucosal tissues are due to aging and living style (loss of elastin and collagen, irritating agents, respiratory diseases, etc.), among other factors [24]. Should this hypothesis be confirmed in further work, these delays could serve as features of tissue aging and decay [25].

The estimations of formant ascents and jaw-tongue reference point displacements ($\Delta\mathbf{S}$ and $\Delta\mathbf{F}$) are given in Table 10. There is not a clear tendency of displacements regarding gender, but it seems that PD participants produced larger displacements compared to HC participants in the average. It may also be seen that PD participants produced larger reference point displacements than HC participants on the average. Whether these results might be associated with hypokinetic dysarthria is a question which requires also further study, in the sense that a large weight magnitude means that small sweeps in formants are associated with large displacements in the reference point, otherwise, small weight magnitudes mean that small displacements in the reference point may produce large sweeps in formants. In this case, it may be hypothesized that if the effective oral cavity is reduced by hypokinetic dysarthria, small changes in its cross-section could produce a substantial change in the formants.

The down-sampling procedure, as mentioned in Section 3.1, has the added benefit of making the methodology presented in this work compatible with telephonic recordings not necessarily reliant on high quality data,. This is possible due to the characteristics of the first and second formant ranges being below 3kHz [19], the bandwidth of the telephonic channel not being an issue as its restricted to 4kHz (sampling frequency of 8kHz), allowing for enough spectral resolution. With this in mind a future line of study would be to explore the characterization of PD dysarthria on data collected remotely, such

as the database taken within the project Teca-Park [26][27]. It contains recordings taken for the eight diadochokinetic exercises mentioned in Section 3.1, including data from 45 PD participants of both genders with the collaboration of PD associations of Spain and Portugal, containing 696 valid utterances (by males) and 637 (by females) from diadochokinetic analysis. This platform is to be adapted to monitor also patients from respiratory diseases, including covid-19, as this technology allows contact-free testing.

As a general comment derived from the overall perspective of the study, it must be highlighted that time realignment is a more relevant procedure than iteration refinement to reduce the estimation error, although a combination of both techniques could improve the estimation accuracy. This is an area we aim to pursue in further work.

The limitations of the present study are the low number of participants included, which does not allow the generalization of results. The intrinsic non-linear behavior of the model needs further study, and its time variance evidenced by the correlation modelling reported needs a specific modelling effort out of the limits of the present study. The dependence of time alignment shifts on formant estimation is also an important issue. An effort in this sense is done to establish reliable relationships between formant bandwidths and delay estimations. Some difficulties arise from the data acquisition procedures, which demand direct physical contact with participants. Besides, data gathering is complicated by the difficulty found on participants perceiving and correctly implementing data acquisition protocols. This issue may become a source of variability affecting the robustness of the methodology and deserves a specific treatment in itself.

6 Conclusions

The present study has been conceived to provide further insights into the acoustic-to-kinematic model of the jaw-tongue articulation joint, based on preliminary approaches. In summary, the key findings derived from this study are:

- An acoustic-to-kinematic model to predict the jaw-tongue joint kinematics from acoustic dynamics expressed in formants has been examined in depth, with special emphasis on weight estimation procedures.
- A weight estimation refinement method based on an iterative gradient algorithm has been explored. It has been found that a reduction in the estimation error functions is always possible at a reasonable number of iteration steps, although its benefit in terms of error reduction is not uniform, depending on specific participant data.
- A complementary correlation optimization study based on signal realignment has been also proposed, and a method to estimate the relative time displacements to be included eventually in the acoustic-to-kinematic projection model has also been defined. Time delays from the male and female datasets used in the study have been estimated and discussed.
- A comparative study on the common characteristics of the estimated projection weights has also been carried on. An average gender-independent model has been estimated on the dataset available, valid for both the HC and PD datasets.

As a summarizing reflection, although many questions still remain open and will require a deeper study in future work, it is essential that progress on these methodologies allowing a remote monitoring of different diseases using convenient and cost-effective technology.

Acknowledgments

This research has been funded by grants TEC2016-77791-C4-4-R (MINECO, Spain) and CENIE_TECA-PARK_55_02 INTERREG V-A Spain-Portugal (POCTEP). The authors

thank Asociación de Parkinson de Alcorcón y Móstoles (APARKAM), and Azucena Balandín (director) and Zoraida Romero (speech therapist) for their help and advice.

References

[1] Skodda, S., Grönheit, W., Mancinelli, N., and Schlegel, U., "Progession of Voice and Speech Impairment in the Course of Parkinson's Disease: A Longitudinal Study", *Parkinson's Disease*; Article ID 389195, 2013.

[2] Sapir, S., "Multiple Factors Are Involved in the Dysarthria Associated With Parkinson's Disease: A Review With Implications for Clinical Practice and Research", *Journal of Speech, Language and Hearing Research*; 57: 1330-1343 (2014).

[3] Rusz, J., Cmelja, R., Tykalova, T., Ruzickova, H., Klempir, J. Majerova, V., Picausova, J., Roth, J. and Ruzicka, E., "Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task", *Journal of the Acoustical Society of America*; 134: 2171–2181 (2013).

[4] Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M. and Acharya, U. R.: A deep learning approach for Parkinson's disease diagnosis from EEG signals, *Neural Computing and Applications*; (2018), doi.org/10.1007/s00521-018-3689-5.

[5] Stam, J. C., Use of Magnetoencephalography (MEG) to study functional brain networks in neurodegenerative disorders, *J. Neu. Sciences*; 289: 128-134 (2010).

[6] Skaper, D., Facci, L., Zusso, M. and Giusti, P., An Inflammation-Centric View of Neurological Disease: Beyond the Neuron, *Frontiers in Cellular Neuroscience*; (2018), doi: 10.3389/fncel.2018.00072.

[7] Arora, S., Baghai-Ravary, L. and Tsanas, A., "Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality speech", *Journal of the Acoustical Society of America*; 145 (5): 2871-2884 (2019).

[8] Yunusova, Y., Weismer, G. and Lindstrom, M. J., "Classifications of Vocalic Segments From Articulatory Kinematics: Healthy Controls and Speakers with Dysarthria", *Journal of Speech, Language and Hearing Research*; 54: 1302-1311 (2011).

[9] Skodda, S., Visser W. and Schlegel, U., "Vowel Articulation in Parkinson's Disease", *Journal of Voice* 25 (4): 467-472 (2011).

[10] Mekyska, J., Janusova, E., Gómez, P., Smekal, Z., Rektorova, I., Eliasova, I., Kostalova, M., Mrackova, M., Alonso, J. B., Faúndez, M., López de Ipiña, K., "Robust and complex approach of patohogical speech signal analysis", *Neurocomputing*; 167: 94-111 (2015).

[11] Dromey, C., Jang, G. O. and Hollis, K., "Assessing correlations between lingual movements and formants", *Speech Communication*, 55: 315-328 (2013).

[12] Whitfield, J. A. and Goberman, A. M., "Articulatory-acoustic vowel space: Application to clear speech in individuals with Parkinson's disease", *Journal of Communication Disorders*; 51: 19-28 (2014).

[13] Gómez, P., Mekyska, J., Ferrández, J. M., Palacios, D., Gómez, A., Rodellar, V., Galaz, Z., Smekal, Z., Eliasova, I., Kostalova, M., Rektorova, I., "Parkinson Disease Detection form Speech Articulation Neuromechanics", *Frontiers on Neuroinformatics*; 11: 1-17 (2017).

[14] Gómez, P., Mekyska, J., Gómez, A., Palacios, D., Rodellar, V. and Álvarez, A., "Articulation Dynamics in Parkinson Dysarthria", *Proc. of MAVEBA 17*; 81-84 Firenze University Press, December 13-15 (2017).

[15] Gómez, A., De Arcas, G., Gómez, P., Álvarez, A. and López, J. M., "Estimating Facial Neuromotor Activity from sEMG and Accelerometry for Speech Articulation", *Proc. of the 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* 1-6, Rome (2018) doi: 10.1109/MeMeA.2018.8438744.

[16] Gómez, P., Mekyska, J., Gómez, A., Palacios, D., Rodellar, V. and Álvarez, A., "Characterization of Parkinson's disease dysarthria in terms of speech articulation kinematics", *Biomed. Signal Proc. and Control*; 52: 312-320 (2019).

[17] Gómez, A. Tsanas, A., Gómez, P., Palacios, D., Álvarez, A., Martínez. R., "A Neuromechanical Model of Jaw-Tongue Articulation in Parkinson's Disease Speech". *Proc. of MAVEBA 19*; 25-28, Firenze University Press, December 17-19, (2019).

[18] Gómez, P., Gómez, A., Ferrández, J. M., Mekyska, J., Palacios, D., Rodellar, V., Eliasova, I., Kostalova, M. and Rektorova, I., "Neuromechanical Modelling of Articulatory Movements from Surface Electromyography and Speech Formants", *International Journal on Neural Systems*, 29 (2), 1850039 (2019).

[19] Huang, X., Acero, A. and Hon, X. W., *Spoken Language Processing*, Prentice Hall, Upper Saddle River, NJ (2001).

[20] Deller, J. R., Proakis, J. G. and Hansen, J. H. L., *Discrete-Time Processing of Speech Signals*. Macmillan, NewYork (1993).

[21] Barzilai, J. and Borwein, J. M., Two-point Step Size Gradient Methods, *IMA Journal of Numerical Analysis*; 8: 141-148 (1988).

[22] James, G., Witten, D., Hastie, T. and Tibshirani, R., *Introduction to Statistical Learning with Applications in R*, Springer, New York (2017).

[23] Rosenbrock, H. H., "An automatic method for finding the greatest or least value of a function", *The Computer Journal*; 3 (3): 175–184 (1960), doi:10.1093/comjnl/3.3.175.

[24] Inamoto, Y., Saitoh, E., Okada, S., Kagaya, H., Shibata, S., Baba, M., Onogi, K., Hashimoto, S., Katada, K., Wattanapan, P. and Palmer, J.B., Anatomy of the larynx and pharynx: effects of age, gender and height revealed by multidetector computed tomography, *J. Oral Rehabil.*, 42: 670-677 (2015).

[25] Hidalgo, I., Gómez, P., and Garayzábal, E., Biomechanical Description of Phonation in Children Affected by Williams Syndrome, Journal of Voice 32(4) (2017) 515.e15-e28.

[26] Project MonParLoc, CENIE_TECA-PARK_55_02 INTERREG V-A Spain-Portugal (POCTEP), https://monparloc.github.io

[27] Palacios, D., Meléndez, G., López, A., Lázaro, C., Gómez, A., and Gómez, P., 2020. MonParLoc: A Speech-Based System for Parkinson's Disease Analysis and Monitoring. IEEE Access, vol. 8, pp. 188243-188255 doi: 10.1109/ACCESS.2020.3031646.