



A Neuromotor to Acoustical Jaw-Tongue Projection Model With Application in Parkinson's Disease Hypokinetic Dysarthria

Andrés Gómez^{1,2}, Pedro Gómez^{2*}, Daniel Palacios^{2,3}, Victoria Rodellar², Víctor Nieto², Agustín Álvarez² and Athanasios Tsanas¹

¹ Old Medical School, Medical School, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom, ² NeuSpeLab, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain, ³ Escuela Técnica Superior de Ingeniería Informática–Universidad Rey Juan Carlos, Móstoles, Spain

OPEN ACCESS

Edited by:

Yury Ivanenko,
Santa Lucia Foundation (IRCCS), Italy

Reviewed by:

Jan Mucha,
Brno University of Technology,
Czechia
Juan Camilo Vasquez-Correa,
University of Erlangen Nuremberg,
Germany

*Correspondence:

Pedro Gómez
pedro.gomezv@upm.es

Specialty section:

This article was submitted to
Motor Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: 29 October 2020

Accepted: 17 February 2021

Published: 15 March 2021

Citation:

Gómez A, Gómez P, Palacios D, Rodellar V, Nieto V, Álvarez A and Tsanas A (2021) A Neuromotor to Acoustical Jaw-Tongue Projection Model With Application in Parkinson's Disease Hypokinetic Dysarthria. *Front. Hum. Neurosci.* 15:622825. doi: 10.3389/fnhum.2021.622825

Aim: The present work proposes the study of the neuromotor activity of the masseter-jaw-tongue articulation during diadochokinetic exercising to establish functional statistical relationships between surface Electromyography (sEMG), 3D Accelerometry (3DAcc), and acoustic features extracted from the speech signal, with the aim of characterizing Hypokinetic Dysarthria (HD). A database of multi-trait signals of recordings from an age-matched control and PD participants are used in the experimental study.

Hypothesis: The main assumption is that information between sEMG and 3D acceleration, and acoustic features may be quantified using linear regression methods.

Methods: Recordings from a cohort of eight age-matched control participants (4 males, 4 females) and eight PD participants (4 males, 4 females) were collected during the utterance of a diadochokinetic exercise (the fast repetition of diphthong [ai]). The dynamic and acoustic absolute kinematic velocities produced during the exercises were estimated by acoustic filter inversion and numerical integration and differentiation of the speech signal. The amplitude distributions of the absolute kinematic and acoustic velocities (AKV and AFV) are estimated to allow comparisons in terms of Mutual Information.

Results: The regression results show the relationships between sEMG and dynamic and acoustic estimates. The projection methodology may help in understanding the basic neuromotor muscle activity regarding neurodegenerative speech in remote monitoring neuromotor and neurocognitive diseases using speech as the vehicular tool, and in the study of other speech-related disorders. The study also showed strong and significant cross-correlations between articulation kinematics, both for the control and the PD

cohorts. The absolute kinematic variables presents an observable difference for the PD participants compared to the control group.

Conclusion: Kinematic distributions derived from acoustic analysis may be useful biomarkers toward characterizing HD in neuromotor disorders providing new insights into PD.

Keywords: speech kinematics, surface electromyography, neuromechanics, hypokinetic dysarthria, neuromotor disorders

INTRODUCTION

Background

Speech production is a dynamic neuromechanical activity which involves cognitive and neuromotor resources of extreme complexity, and which is not yet well understood (Duffy, 2013). The natural way in which it is acquired and used shades the sophisticated processes which are placed into work during its normal expression. Speech is instantiated in the linguistic neuromotor cortex (Demonet et al., 2005), and its execution demands the concurrence of cognitive, neuromotor, neuromuscular, and musculoskeletal processes (Duffy, 2013). Through speech, thoughts and emotions are projected to the knowledge of others by cognitive-linguistic messages. These are programmed for their neuromotor expression by the activation, time-alignment and sensorimotor projection, extension, and strength of a large amount of diverse muscles and associated biomechanical structures. The neuromotor areas from the Central Nervous System (CNS), where planning, programming and control is provided are responsible of activating the respiratory, phonatory and articulatory muscular structures innervated by the Peripheral Nervous System (PNS), see Kandel et al. (2013). The resulting speech is a sequence of acoustic interactions between the glottal source signal and the vocal tract cavities, both driven by neuromotor impulses. This imprint conveys the cognitive-linguistic message. The alteration or dysfunction of any key vocal production mechanisms will result in a deficient production of speech known as a speech disorder. Among them, Motor Speech Disorders (MSD) are the result of dysfunctional neurologic structures involved in the planning, sequencing, activating, and monitoring the neuromuscular structures responsible of speech sound production, modulation and projection. One of the most active neuromuscular structures involved in speech production is the masseter-jaw-tongue complex, including part of the facial muscles and tissues attached to the mandible (Duffy, 2013). This structure is responsible for the production of open or closed, and front or back phonations perceptible in vowels and vowel-related sounds (Greenberg, 2004). Specifically, the quasi-steady positioning of this structure (for more than 30–50 ms) gives rise to vowel-like phonations, whereas its rapid movement is responsible for the acoustical representation of many consonant-like sounds. Neuromotor diseases affect the functional operation of this structure, and its central role in speech articulation suggests it could likely reflect key pathological changes reflected the neuromotor behavior. A well-known indicative neuromotor

disorder known for its prevalence and social impact is Parkinson's Disease (PD), also known as shaking palsy. It is a well-known neurodegenerative disorder since it was first described by J. Parkinson (2002). Its etiology is unclear in most of the cases, but evidence is accumulated in the sense that it may be due to different dysfunctions taking place in the fine control of muscular actions in the interplay of the brain subsystems responsible of musculoskeletal control, as the hypothalamus, the cerebellum, the primary and secondary neuromotor control areas, and the frontal lobes (Brown et al., 2009). A compelling and comprehensive overall view is given in Duffy (2013): "*The motor system is present at all of the major anatomic levels of the nervous system and is directly responsible for all motor activity involving . . . to the planning, control, and execution of voluntary movement, including speech.*" It is a well-established fact that PD causes considerable alterations in speech and phonation (Ricciardi et al., 2016; Brabenec et al., 2017). Broadly speaking, speech alterations may be classified as dysphonia (on voice production), dysarthria (on speech articulation), dysprosody (on the fundamental frequency sequence), and dysfluency (on the rhythm and speech sequence of intersyllabic and intersegment blocks). These alterations are jointly referred to as Hypokinetic Dysarthria (HD). Harel B. et al. (2004) give a summary of the symptoms associated with HD "*Hypokinetic dysarthria, a speech disorder characterized by indistinctness of articulation, weakness of voice, lack of inflection, burst of speech, and hesitations and stoppages, is an integral part of the motoric changes in PD.*" In this same sense, there is "*compelling evidence to suggest that speech can help quantify not only motor symptoms... but generalized diverse symptoms in PD*" (Tsanas, 2012). Godino et al. (2017) stress the fact that "*The low levels of dopamine that appear in patients with PD lead to dysfunctions of the basal ganglia. . . These deficits negatively affect the three main anatomic subsystems involved in the speech production: respiration, phonation and articulation.*" A good description of the neuromotor systems involved in speech production, and how they may be affected by neurodegenerative diseases is to be found in Duffy (2013). Therefore, the search of neuromotor degenerative biomarkers in speech is to be concentrated on phonation (study of the glottal signals in terms of distortion and biomechanics), on speech articulation (study of acoustic and biomechanical clues as formants and jaw-tongue kinematics), on the prosodic flow (concentrated in the time evolution of the fundamental frequency and speech energy stability), and on fluency (syllabic and intersyllabic intervals, duration, stability, and fluctuation of the speaking rate). This is well documented in the work of Mekyska et al. (2015).

Moreover, speech can be used to investigate the nature and extent of vocal impairment in individuals who are at risk of developing PD and can provide a crucial opportunity to intervene in prodromal stages. For example, we have recently demonstrated some very compelling findings when comparing speech signals between a control group, people diagnosed with Sleep Behavior Disorder (SBD), which is among the strongest known predictors of PD risk, and a PD cohort (Arora et al., 2021). Furthermore, we have demonstrated that we can accurately telemonitor PD symptom severity using speech signals collected over the standard telephone network, thus alleviating the need for frequent physical patient visits to the clinic (Tsanas et al., 2021). The main compelling facts favoring speech-based PD biomarkers are the low cost of the required equipment, smart-phones and tablets being increasingly affordable and generally accessible devices, and the contact-less factor, which is particularly useful to facilitate remote studies. Summarizing, the acoustic markers induced by HD in PD speech allows to conclude that speech analysis might become a non-invasive and cost-effective tool to characterize and monitor PD. The role of speech as a possible biomarker in PD detection is well established in the state of the art research literature, with many studies discussing speech-based PD features sensitive to HD. In the present study the focus is placed on the study of acoustic and biomechanical clues, as formants, and jaw-tongue kinematics.

Previous Work

The number of studies focusing on the field of speech and neurodegenerative disorders has been consistently growing over the last 10–15 years, and for the sake of brevity only the most relevant ones will be mentioned in what follows. Brown et al. (2009) give a good description of the relationship between the premotor and motor cortex areas with speech production. Good descriptions of the neuromotor system of the phonation and articulation including neuromotor to muscular pathways affecting the larynx, pharynx, oral, and nasal cavities may be found in Jürgens (2002). Bourchard et al. (2016) offer a description of the relationship between simultaneously recorded neural activity and the kinematics of the lips, jaw, tongue, and larynx. An interesting review on neurophysiology of language may be found in Demonet et al. (2005). Relevant information on the neuromotor pathways on vocal control can be found in Goodman and Hasson (2017). A classical and detailed biomechanical description of tongue movement control may be found in Sanguinetti et al. (1997). The influence of PD on facial muscle sEMG is given in Wu et al. (2014), and for the articulatory and acoustic changes due to Amyotrophic Lateral Sclerosis (ALS) see Mefferd and Dietrich (2020). Relating acoustic features to articulation gesture the reader may see, Gerard et al. (2006), Buchaillard et al. (2009); Dromey et al. (2013). On diphthong articulation kinematics see Tasko and Greilick (2010); for tongue positioning during vowel articulation in speakers with dysarthria see Yunusova et al. (2011); for tongue movements in speakers with ALS see Yunusova et al. (2012). For the use of speech signals from longitudinal assessment of PD we refer to Tsanas et al. (2011) and Tsanas (2012), and the specific ones by Green et al. (2013); Sapir (2014), Mekyska et al. (2015), or

Brabenec et al. (2017). On PD and multiple sclerosis see Tjaden et al. (2013), for vowel articulation positions as a marker of neurodegenerative progress see Skodda et al. (2012).

Objectives

The present study builds on previous work to characterize the relationship between acoustic, 3D accelerometry traces, and electromyographical correlates of speech, in relation with the jaw-tongue structure when carrying on diadochokinetic exercises of clinical interest in neuromotor degenerative disorders as PD (Gómez et al., 2019a,b). We have three primary objectives in the study. Firstly, it is focused to evaluate the functional relationship between neuromotor action in the masseter derived from sEMG and 3Dacc with respect to the acoustic outcomes measured by the two first formants on signals produced both by male and female controls and PD participants exercising on a specific voiced diadochokinetic utterance (repeated sequence of [aI], according to the International Phonetic Alphabet, IPA, 2005). It is assumed that cross-correlations between sEMG, 3D acceleration, and acoustic features may help in determining the optimal time-delays to estimate the weights of projection models by linear regression methods. Secondly, we aim to explore the possibility of establishing inverse relationships between acoustic features and neuromotor activity estimated from sEMG measured on the masseter. Thirdly, we aim to establish whether there are notable differences between controls and PD participants in terms of the absolute kinematic behavior derived from the diadochokinetic exercise from the repetition of [aI] to define new HD biomarkers.

MATERIALS AND METHODS

Speech Neuromechanics

Speech is a complex activity which involves the coordinated joint action of respiration, phonation and articulation muscles, controlled by different cranial nerves (V, VII, IX, X, and XII) and phrenic nerves from the spinal cord, which are responsible for producing speech (Duffy, 2013). Motor activity related to speech is expressed by the cortical areas for motor speech planning and programming of the Central Nervous System (CNS) in Brodmann Area 4 (Primary Motor Cortex: PMC) located in the dorsal portion of the frontal lobe (see **Figure 1**). The programmed activity is transferred to the Peripheral Nervous System (PMS) by connections from the PMC, working in association with other motor areas (premotor cortex, supplementary motor area, posterior parietal cortex, and other subcortical brain regions, which play a role on motor planning and execution). The Upper Motor Neurons (UMNs) are neurons in the PMC, which together with other cortical areas connect with the Lower Motor Neurons (LMNs) in the PMS by the corticobulbar and corticospinal tracts. This is known as the direct pathway. LMNs are alpha-type motor neurons whose axons directly activate the muscles. The LMN and the muscle fibers it activates is known as the Motor Unit (MU). The MUs activating a given muscle, are known as the Final Common Pathways (FCPs). In the case of the masseter (the subject of the present study), the FCPs are grouped in

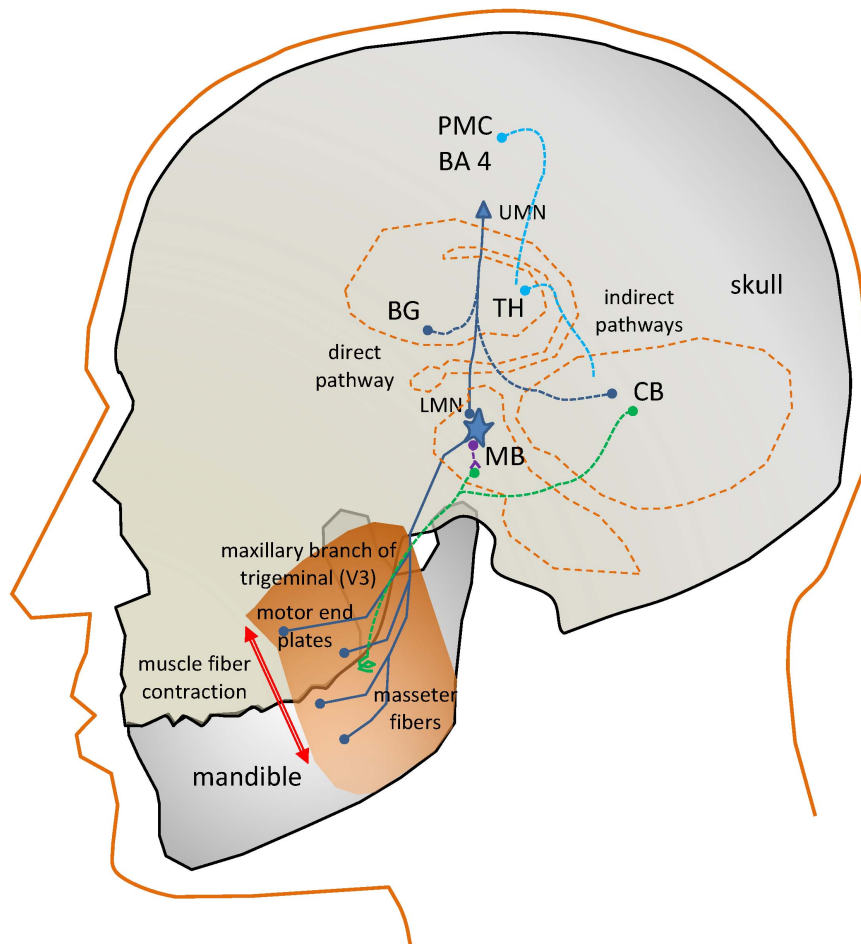


FIGURE 1 | Simplified view of the neural circuits involved in the action of the masseter neuromotor units. PM: Pre-motor Cortex (BA4, Brodmann's Area 4); UMN, Upper Motor Neuron; BG, Basal Ganglia; TH, Thalamus; CB, Cerebellum; LMN, Lower Motor Neuron; MB, Midbrain; dark blue, direct and final common pathways; light blue, indirect control pathways from CB and BG; green, sensory pathways; purple, inhibitory pathways.

the trigeminal maxillary branch, the third subdivision of the cranial nerve V (V3).

In the case of interest for the present study, the motor end plates of the FCPs innervate the masseter fibers producing the muscle contraction. The masseter activation is induced as well by indirect pathways. The fine control of a muscle movement requires some kind of feedback. This is provided by sensory pathways (in green) consisting in neurons activated by spindles (terminal sensors detecting fiber stretching) attached to the muscles, providing proprioceptive sensing to the LMNs in two ways. A direct feedback loop is provided by inhibitory interneurons (in purple). A more complex feedback loop connects sensor units with the Basal Ganglia (BG) and the Cerebellum (CB). These structures serve feedback information to the motor and frontal cortices, as well as to the LMN (blue lines). The BG control circuit assists the PMC in accurate and fine motor speech programming. The CB control circuit coordinates PMC motor planning from proprioceptive information.

Motor speech disorders are produced by specific problems affecting some of the described direct or indirect pathways of

activation, or the muscle fibers. These disorders are commonly referred to as dysarthrias. In the case of PD, the speech disorder is known as HD, related with the pathological behavior of the complex BG control circuit. It will affect any or all the mentioned systems responsible for the neuromechanical control of speech: respiration, phonation, and articulation. The term “hypokinetic” refers to weak small range and rigid movement, giving the impression of flat, soft, and expressionless speech (Duffy, 2013). The activity of the BG control circuit is inhibitory on the PMC areas to modulate cortical activity. An excessive inhibition may result in HD. Most motor problems related with the BG motor circuit have to do with neurotransmitters. Specifically, dopaminergic deficits due to the progressive death of dopaminergic neurons in the substantia nigra *pars compacta* (located at the MB) are the main reason behind PD neuromotor symptoms: “The substantia nigra is the origin of the nigrostriatal pathway, which travels to various structures within the basal ganglia. . . The dopamine deficiency in this nigrostriatal pathway and the basal ganglia account for most of the typical features of PD. Once the brain is no longer able to compensate for this

dopamine loss, there are a number of effects which can occur. Typical symptoms include muscle rigidity, akinesia, bradykinesia, and tremor. . .” (Goberman and Coelho, 2002), more specifically “The essential neuropathological changes in PD are a loss of melanine-containing dopaminergic neurons in the substantia nigra pars compacta. . . This results in a dysfunction of the basal ganglia circuitry, which is an integral part of cortico-basal ganglia-cortical loops that mediate motoric and cognitive functions” (Harel B.T. et al., 2004).

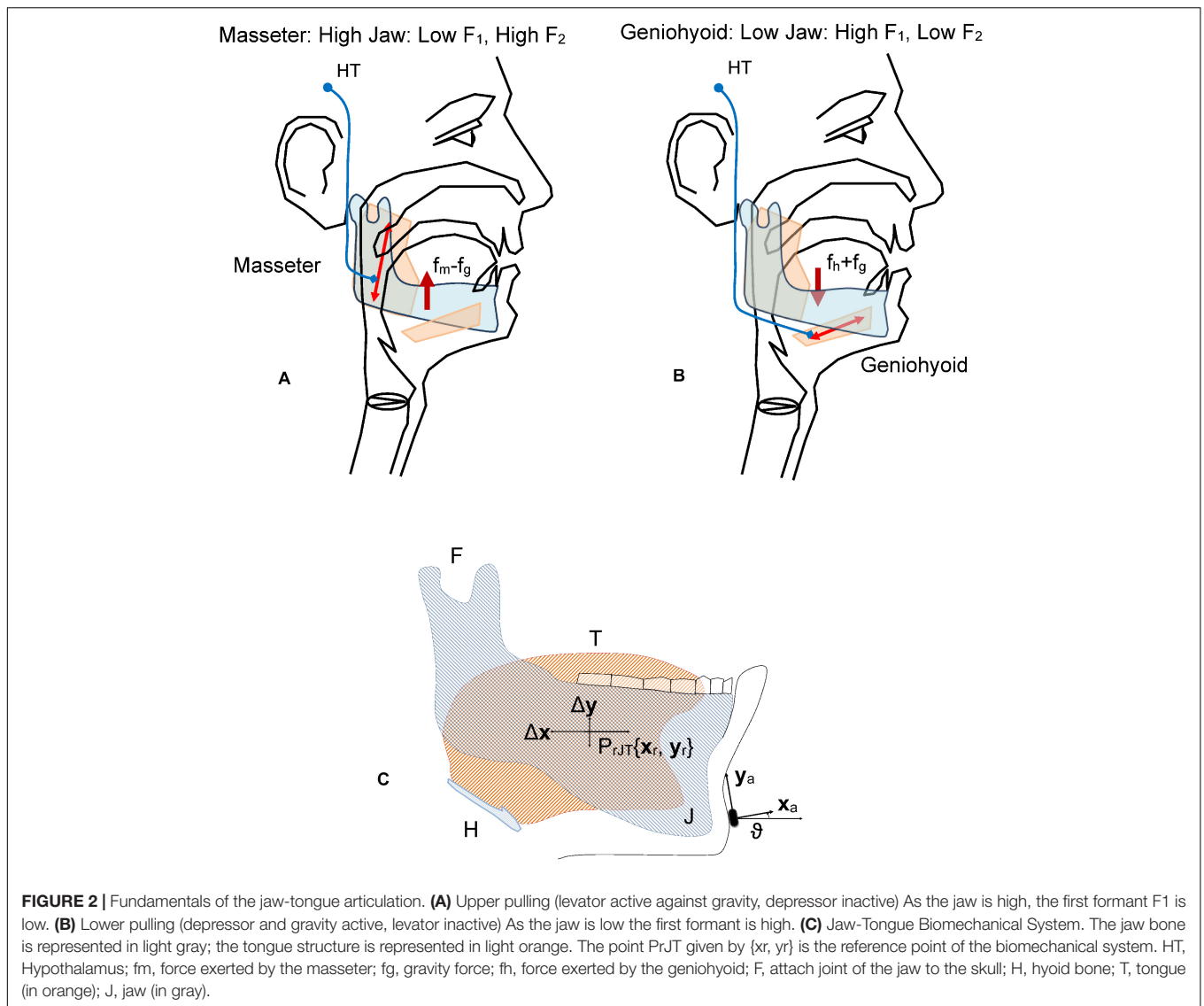
In the present study, a biomechanical system of the jaw-tongue will be used, which may be modeled to estimate the neuromotor behavior of the system, and provide specific markers of proper or improper neuromotor activity. The selection of the masseter as the target muscle obey to the following reasons: it is a powerful muscle developing a strong sEMG when contracting, it is accessible (beneath the caudal section of the cheek), it may modify strongly the oral cavity when contracting or relaxing leaving a clear acoustic signature in formants, and its biomechanical activity is well understood. The study of the masseter neuromechanics is based on the electrical, dynamic, and kinematic activity of the muscle as a functional structure.

A Model Mediating Jaw Kinematics on Acoustic Features

The present study builds on a previous masseter-jaw-tongue biomechanical model (Gómez et al., 2019b), where the active contribution of other muscles (styloglossus, geniohyoid, intrinsic glossal, etc.) has not been taken into account, and the passive contribution of other tissues, as the oro-facial substructures are implicitly included in the inertial moment. The articulation gestures based on the masseter-jaw-tongue structure as considered in the present study are described as follows: the jaw (J) is fixed to the skull bone at fulcrum (F) as in a third-class lever system. The tongue (T) is a complex muscular and vascular structure supported on the jaw and the hyoid bone. Some other extrinsic muscles fix it to the cranial structure (mainly the styloglossus and geniohyoid). These muscles and the tissues surrounding the jaw will be considered part of the lumped equivalent of masses, forces and moments at the reference point of the jaw-tongue system P_{JT} , defined at $\{x_r, y_r\}$, where forces acting on the system induce movements in the sagittal plane (x : horizontal, or rostral-caudal, y : vertical, or dorsal-ventral); these forces are f_m (exerted by the masseter), f_{sg} (by the styloglossus), f_{gi} (by the intrinsic glossus), f_h (by the geniohyoid), and f_g (gravity). Consequently, the P_{JT} may be defined as a hypothetical point in the sagittal plane (x : caudal-rostral; y : dorsal-ventral) with static coordinates $\{x_r, y_r\}$ where the sum of dynamic and inertial forces is null. The force exerted by the masseter f_m will pull up the low mandible acting as a third-order lever with fulcrum at the maxillary attachment (F). In the present study we will not consider extrinsic actions other than by geniohyoid, and by the intrinsic glossus system, therefore, the only forces to be considered contributing to the lever momentum will be f_m, f_h , and f_g . The kinematic displacements experienced by P_{JT} are given as $\{\Delta x_r, \Delta y_r\}$. Lateral movements orthogonal to the sagittal plane are assumed small enough not to be considered

(system with only two degrees of freedom). The functioning of the speech articulation neuromotor and biomechanical system is illustrated **Figure 2**. Speech articulation is defined by the configuration of the articulation organs, such as the mandible, tongue, lips, and velo-pharyngeal tissues, among others of lesser interest for the present study (Dromey et al., 2013; Cattaneo and Pavesi, 2014; Whitfield and Goberman, 2014). Therefore, vowel articulation has been traditionally described in terms of the open-close gesture (also low-high attending to lower jaw position relative to upper jaw), the front-back gesture, and round-oval configurations (Dromey et al., 2013). These gestures determine certain acoustic features known as formants, which are defined as frequency positions where the harmonic structure of the vowel is especially enhanced, by means of the resonances of the Oro-Naso-Pharyngeal Tract (ONPT, see **Figure 2**). The open-close gesture, mainly dominated by the jaw, is affecting the first formant F_1 . Pulling up the jaw by the masseter against gravity ($f_m - f_g$) is the dominant gesture in the phonation of high vowels as [i:] and [u:], whereas depressing the jaw under the force of gravity and the geniohyoid muscle action ($f_h + f_g$) is the gesture to phonate low vowel as [a:]. The front-back gesture is controlled also by the jaw position, although in high vowels like [i:] and [u:] the tongue position affects mainly the second formant F_2 (pushing the tongue forward is the gesture for [i:], pulling it back results in [u:]). Therefore, the articulation gesture of the jaw may be studied to relate articulatory gestures and acoustic features as the first two formants (F_1, F_2), as proposed in the Articulation Kinematic Model (AKM) shown in **Figure 2C**. The study is based on the dynamic tracking of the kinematic activity of the jaw-tongue reference point (P_{JT}) by means of 3D accelerometry (3Dacc).

The articulation dynamics foresees that the resonant properties of the ONPT will change in time regarding the position of the P_{JT} under the action of the forces mentioned, modifying, and producing dynamic changes in the first two formants $\{F_1, F_2\}$. From the perspective of speech production, the general problem of acoustic to articulatory inversion may be presented in general terms as the mapping $A(s) = F$ (Ouni and Laprie, 2005), where s is the articulatory vector which presents the general shape of the vocal tract at a specific time instant for instance, the k parameters of an articulatory synthesis model, and F is the acoustics vector (typically the first m formants). $A: s \rightarrow g$ is a non-linear many-to-one mapping transforming the articulatory feature space s to the acoustic feature space g . In this sense, the inverse mapping presents several important problems: on the one hand the mapping A is non-linear, on the other hand, it is a many-to-one (Qin and Carreira-Perpiñán, 2007). Besides, g is an approximation to the values of the true resonances of the vocal tract, obtained usually from Linear Prediction Inverse Filtering (LPIF), and therefore, subject to strong assumptions on the representation of the real oro-naso-pharyngeal tract (ONPT) by a concatenation of rigid-wall cylindrical-tube sections on a straightened medial axis, the number of sections, depending on the sampling rate and the order of the LPIF (Deller et al., 1993). On the other hand, the feature vector s will depend on the specific generative model used. In many studies the seven following features are used: jaw position (high-low),



tongue dorsum position (backward-forward), tongue dorsum shape (rounded-unrounded), tongue tip shape (up-down), lip height (open-close), lip protrusion (forward-backward), and larynx height (high-low). The solutions proposed are based on reproducing a reduced articulatory feature space \mathbf{s} which may generate an approximation to the acoustic feature vector \mathbf{g} , following an optimization process reducing the cost function $|\mathbf{g} - \mathbf{g}|$ to a minimum value. Several approaches are used such as codebook-based inversion, articulatory modeling, articulatory modeling, and statistical mapping (Sivaraman et al., 2019). The present study proposes a mapping model which is a simplification of the general one following several important assumptions:

- Only the vertical and horizontal components of the jaw-tongue position (high-low) are considered from the set of articulatory features \mathbf{s} . These features are estimated using a 3D accelerometer fixed to the participant's chin.

A transformation of coordinates from the 3D accelerometer to absolute ones in the sagittal plane is used (rotation and projection).

- The set of acoustic features is reduced to the first two formants F_1 and F_2 .
- A kinematic-acoustic variable estimated from formant acoustics (AFV).
- A kinematic-dynamic variable estimated from jaw kinematics (AKV).
- A non-linear model projecting acoustic features to a kinematic-acoustic variable is proposed
- A linear model mapping acoustic features to dynamic variables is proposed.

Having these assumptions in mind, the present study is focused on the evaluation of a linear model on data produced by control and PD participants in the fast and repeated utterance of the diphthong [aI]. The reasons for selecting such

a diadochokinetic exercise are that, on the one hand this pattern ensures that the jaw-tongue system moves through a range where linearity governs the link between articulation and sound features “For /a/ the strength of the acoustic-kinematic association was robust across movements that differed in duration or displacements” (Dromey et al., 2013). On the other hand, this exercise is mainly governed by the dynamic activity of the masseter, a very accessible facial muscle producing good sEMG recordings. Another important fact is that this exercise does not involve lip protrusion or rounding changes (Dromey et al., 2013). Therefore, the association between the jaw-tongue reference point movement may be associated with the first two formants in a one-to-one mapping. The complementary use of a 3D accelerometer on the jaw helps in providing a second assessment method to avail this association by an alternative acoustic-to-dynamic mapping.

The proposed model for projecting articulation kinematics to changes of the first two formants F_1 and F_2 can be described as:

$$\begin{bmatrix} \Delta F_1(t) \\ \Delta F_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \Delta x_r(t) \\ \Delta y_r(t) \end{bmatrix} \quad (1)$$

where $\{a_{ij}\}$ are the parameters relating the dynamic changes of the formant values with the oscillations of the P_{rJT} . The utility of these relations is conditioned by the possibility of estimating the set of parameters $\{a_{ij}\}$ from the signals produced by a 3D accelerometer fixed on the chin of a participant under test, as shown in **Figure 2C**. The accelerometer reference axes are the chin-normal (\mathbf{x}_a) and tangential (\mathbf{y}_a), which will be changing following jaw displacements. The accelerometry signals may be transformed to the reference coordinates $\{x_r, y_r\}$ by means of a rotation in terms of ϑ , the angle between the axes \mathbf{x}_a and \mathbf{x}_r . The set of parameters $\{a_{ij}\}$ relating acoustic features and articulation dynamics may be estimated by regression-based methods using specific repetitive diadochokinetic exercises, such as the repetition of the gliding sequence [. . . aIa . . .] as it will be shown in what follows. Assuming that (1) is time-invariant and invertible we would have:

$$\begin{bmatrix} dx_r(t)/dt \\ dy_r(t)/dt \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} \partial F_1(t)/\partial t \\ \partial F_2(t)/\partial t \end{bmatrix} \quad (2)$$

where $\{w_{ij}\}$ are the coefficients of the transference matrix inverse $\mathbf{W} = \mathbf{A}^{-1}$:

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} \quad (3)$$

With these expressions in mind it will be possible to define the Absolute Formant Velocity (AFV) of the reference point (P_{rJT}) as:

$$|v_f(t)| = \left[H_1 \left(\frac{dF_1(t)}{dt} \right)^2 + H_2 \left(\frac{dF_2(t)}{dt} \right)^2 + H_{12} \frac{dF_1(t)}{dt} \frac{dF_2(t)}{dt} \right]^{1/2} \quad (4)$$

where H_1 , H_2 and H_{12} are quadratic forms of $\{w_{ij}\}$ (see Gómez et al., 2019b). Reliable estimates for $\{w_{ij}\}$ may be obtained from articulations involving changes in the positions of the reference point showing predictable dynamic changes, as in this and other

diadochokinetic exercises. The AFV may be estimated in the following steps:

- Speech $s(t)$ is sampled at 50 kHz and 16 bits, down-sampled to 8 kHz and inverse filtered (Alku et al., 2019) to obtain a K -th order adaptive prediction vector $\{b_i\}$ representing the inverse vocal tract.
- The first two formants F_1 and F_2 are estimated by detecting the zeros of $B(z)$ in the complex plane:

$$B(z = z_i) = 1 - \sum_{i=1}^k b_i z_i^{-i} = 0; \quad z_i = r_i e^{j\varphi_i}; F_i = \varphi_i / (2\pi\tau); \varphi_i \geq 0 \quad (5)$$

τ being the sampling rate in the time domain.

Similarly, an Absolute Kinematic Velocity (AKV) of the reference point may be derived from the normal and tangential acceleration components on the sagittal plane $\{a_{Xa}(t), a_{Ya}(t)\}$ measured directly by the 3D accelerometer as:

$$|v_d(t)| = \left[\left(\int_{\zeta=0}^t a_{xd}(\zeta) d\zeta \right)^2 + \left(\int_{\zeta=0}^t a_{yd}(\zeta) d\zeta \right)^2 \right]^{1/2} \quad (6)$$

where $\{a_{xd}(t), a_{yd}(t)\}$ are the acceleration components in the sagittal plane rotated from the measurements recorded on the native accelerometer axes $\{a_{Xa}(t), a_{Ya}(t)\}$.

The estimation procedure of the AKV would involve the following steps:

- The acceleration component means $\{\hat{a}_{Xa}(t, W), \hat{a}_{Ya}(t, W)\}$ are used to estimate the 3D accelerometer angle ϑ on short time windows (W) to preserve time invariance. The acceleration components are rotated to the reference axes to produce $\{a_{xd}(t), a_{yd}(t)\}$.
- Rotated acceleration components are used to estimate the AKV following (6).

Distributions of Absolute Velocity Values

The probability distributions of both AFV and AKV from (4) and (6) are very relevant statistical descriptors of articulation kinematics. They can be directly estimated from their normalized amplitude histograms over bins between 0 and 20 ($\text{cm}\cdot\text{s}^{-1}$) as:

- The speech signal is low-pass filtered to 4 kHz (antialiasing) and downsampled to 8 kHz. The 3D acceleration signals are low-pass filtered to 250 Hz and downsampled to 500 Hz.
- The first two formants are estimated every 2 ms on short time windows of 64 ms, equivalent to 512 samples with an overlap of 97% (62/64).
- The weights $\{w_{ij}\}$ are estimated from linear regression between dynamic displacements $\{\Delta x_r, \Delta y_r\}$ and formant deviations $\{\Delta F_1, \Delta F_2\}$. The quadratic coefficients H_1 , H_2 , and H_{12} are calculated from $\{w_{ij}\}$.
- The AFV ($|v_f(t)|$) is estimated from (4).

- The accelerations $\{a_{xd}(t), a_{yd}(t)\}$ are obtained by rotating the signals recorded by the 3D accelerometer downsampled to 500 Hz and integrated numerically. Detrend filtering must be used to avoid integration drifts.
- The AKV ($|v_d(t)|$) is estimated from (6).
- An N-bin histogram of counts by amplitudes is built from each participant's AFV and AKV. The interval covered for speeds is $[0, |v_r|_{max}]$, with $|v_r|_{max} = 20 \text{ cm.s}^{-1}$, and $N = 100$, therefore each bin size is $\Delta b_k = [|v_r|_{max}/N] = 0.2 \text{ cm.s}^{-1}$ wide.
- The following histogram of counts is built for each bin $b_k = k \bullet \Delta b_k$:
if $b_{k-1} \leq |v_r(t)| < b_k$ then $c_k = c_{k+1}$
 c_k being the number of counts for bin b_k .
- Count histograms c_k ($0 \leq k \leq N$) are normalized to their total number of counts $C_T = \sum b_k$ ($0 \leq k \leq N$), therefore they could be considered estimators of probability density functions $p_k = c_k/C_T$.

Thence $p(|v_{rk}|) = p_k$ will be an estimate of the AFV and AKV probability density functions. It has been proven that this feature is relevant in separating dysarthric from normative speech (Gómez et al., 2017).

Entropy-Based Detection

The AFV and AKV distribution show a two-degrees of freedom χ^2 behavior, therefore they are an estimation of the speaker's jaw mobility in terms of similarity with a certain "articulation temperature." The probability distributions may be used to estimate the divergence between utterances from different speakers in terms of entropy based metrics (Cover and Thomas, 2006) between two probability distributions (i, j) in terms Jensen-Shannon's Divergence (JSD) as:

$$D_{Jsj} \{p_i(|v_r|), p_j(|v_r|)\} = \frac{1}{2} D_{KL} \{p_i, p_m\} + \frac{1}{2} D_{KL} \{p_j, p_m\};$$

$$p_m = \frac{p_i + p_j}{2} \quad (7)$$

where $p_i(v_r)$ and $p_j(v_r)$ are two distributions (either AFV or AKV) from two different participants, and D_{KL} is Kullback-Leibler's Divergence (Cover and Thomas, 2006):

$$D_{KLij} = D_{KL} \{p_i(|v_r|), p_j(|v_r|)\} = - \int_{\zeta=0}^{\infty} p_i(\zeta) \log \left[\frac{p_i(\zeta)}{p_j(\zeta)} \right] d\zeta \quad (8)$$

The main advantage of JSD is that it is bounded and symmetrical ($D_{Jsj} = D_{Jji}$). The AFV and AKV estimates from (4) and (6) and their normalized histograms were evaluated as described before. Two sets of distributions were produced, respectively, for the control participants (CP: p_{CP}) and the PD participants (PD: p_{PP}). The average of the p_{CP} distributions was used as the control reference:

$$p_{CP} = \frac{1}{k_C} \sum_{i=1}^{k_C} p_{CPi} \quad (9)$$

where k_C is the number of participants in the CP set separately for AFV and for AKV. The JSD between each participant in the

sets CP and PD was estimated with respect to the average p_{CP} . Therefore, the divergences used in the present study are:

$$D_{Jsf_{i \in CP}} = D_{JS} \{p_{if}, p_{CPf}\}; D_{Jsd_{i \in PD}} = D_{JS} \{p_{if}, p_{CPf}\};$$

$$D_{Jsd_{i \in CP}} = D_{JS} \{p_{id}, p_{CPd}\}; D_{Jsd_{i \in PD}} = D_{JS} \{p_{id}, p_{CPd}\} \quad (10)$$

where the suffixes f and d specify JSDs estimated from AFV or AKV, respectively.

Masseter sEMG

The selection of the masseter as the reference muscle in studying the relationship of neuromotor activity and acoustic speech dynamics is based on the following reasons:

- The masseter changes the position of the jaw-tongue structure toward high and slightly forward positions when activated. Correspondingly, the vocal tract is modified from low-mid to high-front vocal resonances (Dromey et al., 2013). The relationship between neuromuscular action and acoustics seems to be quite direct.
- The masseter is a powerful muscle, its neuromotor activity induces strong sEMG signals on the back lower part of the cheek. Signal recording is feasible and very productive.
- The neuromotor pathways from the midbrain to the masseter are short and fast, the delays due to impulse propagation on the neural pathways, and the motor unit action potentials are small. This allows producing faster oscillation movements than other larger and more distant muscles, extending the high frequency limits of voluntary and involuntary tremor.

The procedure to record sEMG activity on the masseter is represented in **Figure 3** following the works of De Luca (1997) and Castroflorio et al. (2005).

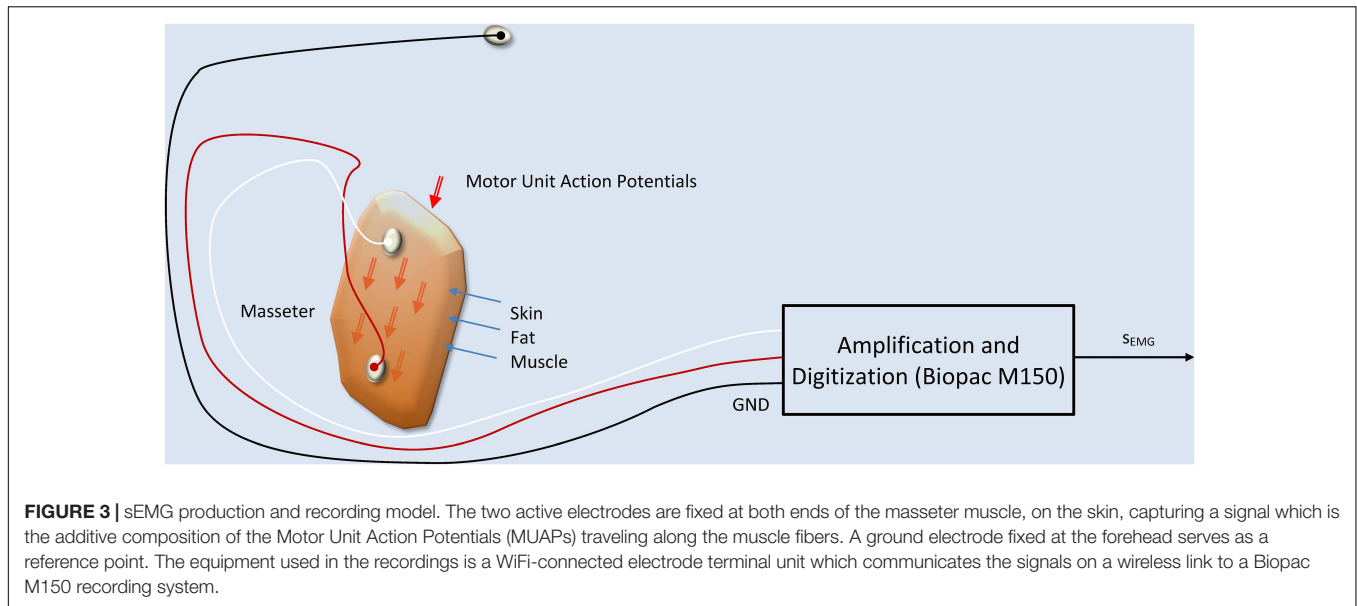
The sEMG signal recorded is related with the summation of the MUAPs traveling along the muscle fibers on the muscle cell membranes as described in Martínez-Valdés et al. (2017):

$$s_{EMG}(n) = \sum_k H(k) s_{MUAP}(n-k) e(n) \quad (11)$$

where $s_{EMG}(n)$ is the recorded sEMG signal, $H(k)$ is a transfer weight having into account skin and fat conductivity and time propagation effects, $s_{MUAP}(n)$ are the MUAPs traveling along the muscle fibers, $e(n)$ is the recording noise, and n is the time index (Farina and Merletti, 2001; Teklemariam et al., 2016). Concerning the dynamic action exerted by the muscle during activation it will be assumed that the force exerted by the muscle in the upper vertical direction $f_m(t)$ is a correlate of the sEMG recorded on the bulk of the masseter $s_m(t)$ (Roark et al., 2002; Lee, 2010) and therefore, it may be expressed as the joint action of these individual actions, therefore:

$$f_m(t) = J_m r_m(t); J_m = \frac{T_m}{l_m \cos \theta}; r_m(t) = \int_{\zeta=0}^t |s_m(\zeta)| d\zeta \quad (12)$$

where J_m is the mioelectric proportionality parameter when small oscillations are assumed, T_m is the angular neuromotor torque,



l_m is the effective jaw arm length (considering the jaw-tongue system as a lumped load), ϑ is the rotation angle (Hogan, 1984) and $r_m(t)$ is the integral of the rectified sEMG on the masseter. The sEMG is low-pass filtered at 250 Hz and downsampled to 500 Hz. Its rectified value is numerically integrated following (12). Detrending filtering must be used to avoid integration drifts. The reason behind force being related to the integral of the rectified sEMG, and not to sEMG, has to do with the way in which sEMG is recorded, using pairs of electrodes symmetrically placed at both sides of the neuromotor innervation zone on the muscle (see **Figure 3**), as suggested by the experts (De Luca, 1997; De Luca et al., 2010). The integral of the rectified sEMG will be referred in the sequel as the Electromyographic Correlate of the Masseter Force (ECMF).

Linear Regression-Based Statistical Mapping

The present study is intended to link the masseter neuromotor activity (estimated from the sEMG signal) with jaw-tongue kinematics (estimated from 3D accelerometry) and acoustic dynamics (estimated from the first two formants of voice). Cross-correlation (time lag correlation) is used to estimate the optimum time-lag alignment between two signals, determined by the maximum absolute values of Pearson's correlation coefficients. Linear regression between the aligned signals is used to estimate the weights of the corresponding projection model. Cross-correlation methods have been already used in acoustic-kinematic mapping are standard procedures (see Ouni and Laprie, 2005; Dromey et al., 2013; Mitra et al., 2017; Sivaraman et al., 2019). The aim of the present study is based on the inherent relationship between the masseter activity and jaw movement, to show that the front-to-end causality chain from Neuromotor Activity \rightarrow Masseter sEMG \rightarrow Vertical Force \rightarrow Vertical Position \rightarrow $\{\Delta F_1, \Delta F_2\}$ may be quantitatively described by a simple model, and that the model parameters might be

estimated in a first approximation by linear methods. Having in mind that the force exerted by the masseter could not be inferred directly, the present study opens the possibility of making this estimation possible relating acoustic features with neuromotor activity in an inverse relationship, using speech, sEMG and 3D accelerometry.

The methodology used is based on linear regression and cross-correlation on the mentioned signals and estimates, as represented in **Figure 4**.

The systemic approach is intended to establish the strength of the relationships in the cause-effect chain expressed by the following links: Neuromotor Activity \rightarrow Masseter sEMG \rightarrow Vertical Force \rightarrow Vertical Position \rightarrow $\{\Delta F_1, \Delta F_2\}$. In this problem s_{NA} is the ground truth (not directly observable), s_{EMG} is its observable correlate, and a_{cx} and a_{cy} are the observable dynamic correlates. The estimates of the acoustic accelerations a_{fx} and a_{fy} help in establishing a validation for the indirect estimation of kinematic variables directly from acoustics. These kinematic variables would allow estimating the neuromotor activity directly from acoustics using speech recordings from remote devices (Palacios et al., 2020).

Materials

Eight volunteering speakers (four males and four females) were recruited among PD participants from the APARKAM association of Alcorcón and Leganés, two cities in the southwest of the community of Madrid, Spain. The inclusion criteria were non having had a diagnosis of any laryngeal pathology, being in H&Y stage 2, and non-smokers. Another set of eight participants of both genders with not known laryngeal or neurological pathologies in a similar age range were selected to serve as controls. Their biometrical description is given in **Table 1**.

The study was approved by the Ethical Committee of Universidad Politécnica de Madrid, and the participants signed an informed consent. The experimentation protocols were

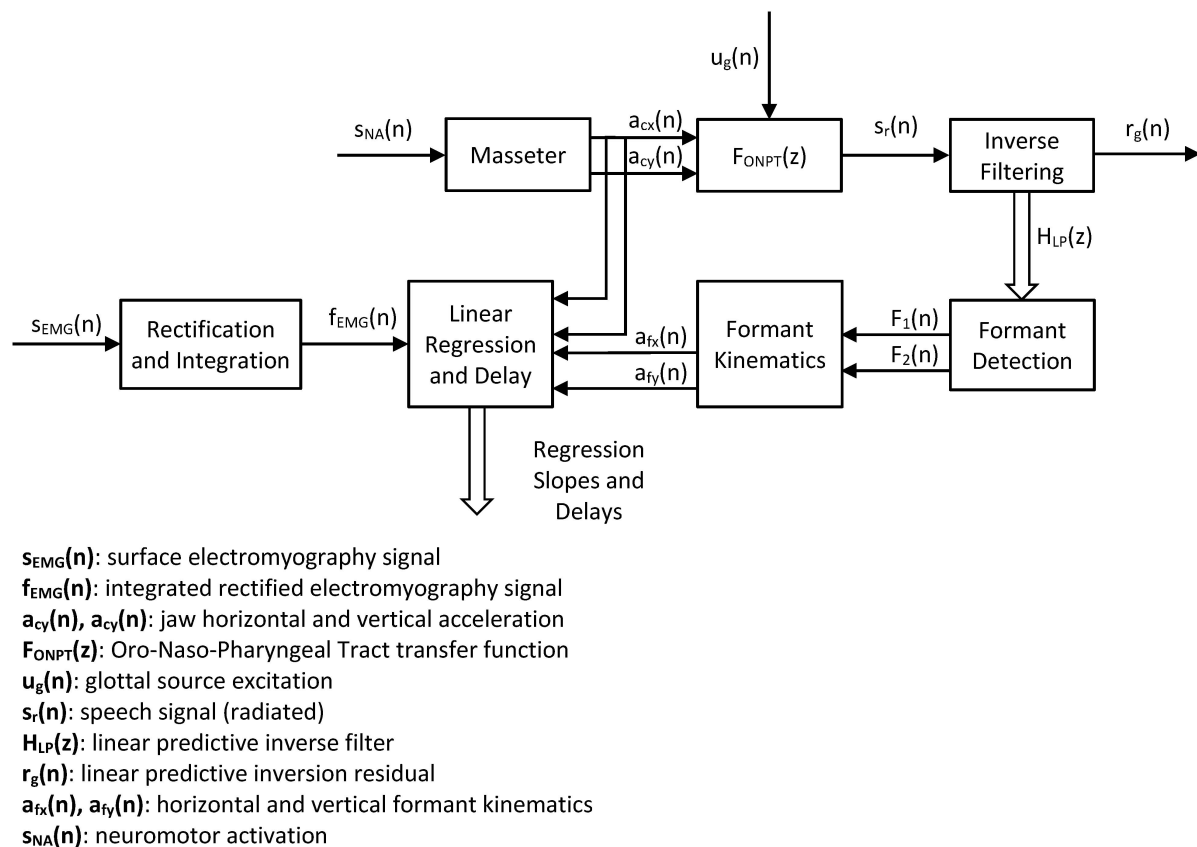


FIGURE 4 | Systemic view of the study. The neuromotor activity induced in the masseter $s_{NA}(n)$ produces accelerations in the horizontal and vertical directions of the jaw-tongue structure on the sagittal plane [$a_{cx}(n)$ and $a_{cy}(n)$] which may be directly estimated from the 3D accelerometer. The dynamic action on this structure changes the configuration of the ONPT, and its filtering function $F_{ONPT}(z)$. The radiated speech signal $s_r(n)$ is the result of filtering the glottal excitation $u_g(n)$ by the ONPT. The inverse filtering of speech reconstructs the inverse transfer function $H_{LP}(z)$, which is used to determine the first two formants $F_1(n)$ and $F_2(n)$. The dynamic behavior of the first two formants is used to indirectly estimate the correlates of the horizontal and vertical accelerations in the sagittal plane $a_{fx}(n)$ and $a_{fy}(n)$. Cross-correlation between dynamic and acoustic estimations of acceleration and force from sEMG measurements is used to optimally align formants and accelerations. Linear regression between dynamic and acoustic variables is used to estimate the model weights.

aligned with the Declaration of Helsinki. The data recording protocol consisted in the synchronous and simultaneous recording of voice, 3D accelerometry and sEMG as shown in **Figure 5**.

The speech samples were recorded at 50 kHz with 16 bits resolution on a MOTU Traveler sound card by a wireless link, using a clip cardioid microphone (Audio Technica) at the speech therapist studio. No special soundproofing room was required due to the short distance to the microphone and its high directionality. The samples from the sEMG and the 3Dacc were recorded at 2.5 kHz. The five signal channels were acquired and digitized by a Biopac M150.

RESULTS

The methodology described refers to concurrent recording of speech, sEMG and 3D accelerometry. The results presented here refer to the use of rotated vertical accelerometry only with respect to the first and second

formant oscillations. As an example, the corresponding records for the male participant CM_a when executing the diadochokinetic exercise of repeating [..aI..] are shown in **Figure 6**.

The ECMF, the rotated vertical acceleration (dynamic) and the vertical acceleration estimated from the first two formants (acoustic) are shown in detail in **Figure 7**.

The relationship between the vertical displacement Δy and the formant oscillations ΔF_1 and ΔF_2 may be studied using linear regression on the respective signals. The results are presented in **Figure 8**.

Other aspects of interest are the relationships between the ECMF and the vertical rotated accelerations measured by the accelerometer (dynamic) and estimated from formants (acoustic). As the relationship between these two variables is also of interest for the study the results are presented in **Figure 9**.

The results of similar evaluations for each participant in the study are given in **Table 2** for comparative purposes.

As HD is a manifestation of the neuromotor failure in responding to rapid movements, the concepts of absolute velocity

TABLE 1 | Participants' biometrical data.

Participant's code	Gender	Age	Condition	H&Y
CFa	F	67	C	
CFb	F	62	C	
CFc	F	66	C	
CFd	F	67	C	
CMa	M	69	C	
CMb	M	71	C	
CMc	M	68	C	
CMd	M	69	C	
PFa	F	73	P	2
PFb	F	66	P	2
PFc	F	65	P	2
PFd	F	70	P	2
PMa	M	69	P	2
PMb	M	73	P	2
PMc	M	72	P	2
PMd	M	69	P	2

defined in (4) and (6) may help in evaluating and quantifying this manifestation. The AKVs and AFVs from a control and a PD participant are represented in **Figure 10**.

The JSD of each participant with respect to the average distribution of the control participants p_{CP} defined in (9) to explain individual kinematics with respect to a common reference is given in **Table 3**.

The results from the PD cohort are compared with the ones from the control participants using three types of tests: Student's t -test, Kolmogorov-Smirnov (KS) and Mann-Whitney (MW). Previously, the results from each cohort have been tested for normality using the Lilliefors test. The normality hypothesis was not rejected for any of the cohorts (controls and PD). The results of the three tests are given at the bottom of the table. In this case, t -test and KS test rejected the null hypothesis of same means under a 0.05 significance level. MW test also rejected the null hypothesis of equal distributions from both cohort results under a 0.05 significance level.

DISCUSSION

Starting with the recordings from the example presented in **Figure 6** it may be seen that the sEMG shows spike-like bursts which correspond to muscle contractions during the articulation movements produced to utter the diadichokinetic exercise (a). The placement of the electrodes is of crucial importance, as well as cleaning the skin with deionized water or a soft de-makeup napkin prior to electrode placement. The ECMF, shown in (b) is well aligned with the rotated vertical acceleration given in (c). Both signals show similar oscillations, which presents narrow peaks, possibly due to the antagonist action of the geniohyoid muscle. The speech envelope (d) shows a larger intensity when the vocal opening is larger (lower sounds when the articulation moves to [a]) as the release of acoustic energy is larger during these intervals. It must be mentioned that formants do not arrive

to extreme positions on the vowel triangle (\rightarrow [a] and \rightarrow [I]), swinging instead between a low and a high vowel, which can be described phonetically as [æ \rightarrow I \rightarrow æ]. In (e) the first two formants (in blue and red) show a counter-phase oscillation, which is delay-aligned with the acceleration. Similar plots were also produced from each participant.

Regarding **Figure 7**, it may be seen that the rotated vertical acceleration as measured by the accelerometer (b) shows a sharper behavior during vertical pull-ups when compared to the ECMF (a), although their positive peaks are well aligned. This may be due to the action of the antagonist muscles (mainly the geniohyoid), which apparently retain the vertical movement up to a point where they suspend the retention and a sharp vertical peak is released. The acoustical vertical acceleration (c) shows a smoother behavior, and although it is also aligned with (a) and (b) a small delay may be observed in the acoustical acceleration, due to the inertial response of the positioning of the jaw, which conditions the establishment of formants within some delay. An added factor to explain this delay is the time required for the acoustic wave to produce the sustained standing waves in the ONPT which are responsible of formants. These delays can be inferred from the regression study shown in **Figure 9**.

The results of applying linear regression are given in **Figure 8**, showing that the influence of the vertical displacement affects more the second formant than the first one (an oscillation of 2 mm produces a formant swing of almost 100 Hz on F_1 , and about 200 Hz in F_2 , peak-to-peak). The effect is reversed on each formant: whereas positive displacements in the vertical axis result in negative displacements on F_1 , they induce a rise on F_2 , as expected from the description given in **Figures 2A,B**. The correlation coefficients (Pearson and Spearman) are both large and significant. The delays suggest that the formants react to the vertical displacement with a delay between 14 and 18 ms. This delay is due to the inertial moment of the jaw-tongue structure, and to the time required for the formants to emerge acoustically as standing waves within the acoustical structures, a factor which is also regulated by the quality factors of the resonances, and ultimately on the energy losses caused by viscoelastic factors on the biological tissues, an interesting problem which would deserve further study.

Regarding the regression results shown in **Figure 9** between the ECMF and vertical acceleration A_{yd} measured from the accelerometer (a: dynamical) and A_{yf} from the indirect estimation on the first two formants (b: acoustical) it may be seen that there is a direct regression in both cases, showing large and significant correlation coefficients (Pearson). Nevertheless, the dynamical acceleration shows a small delay with respect to the ECMF (2 ms, which corresponds to a single sample interval at an effective sample frequency of 500 Hz). On its turn, the delay between the acoustical acceleration A_{yf} and the ECMF is of 32 ms. When checking the correlation between the dynamical and the acoustical acceleration estimates, a delay of 28 ms is observed in the second one with respect to the first one. The disagreement when considering the delays of both signals in regard to the ECMF is of 4 ms (2 sampling intervals at 500 ms), which could be attributed to small signal misalignment errors.

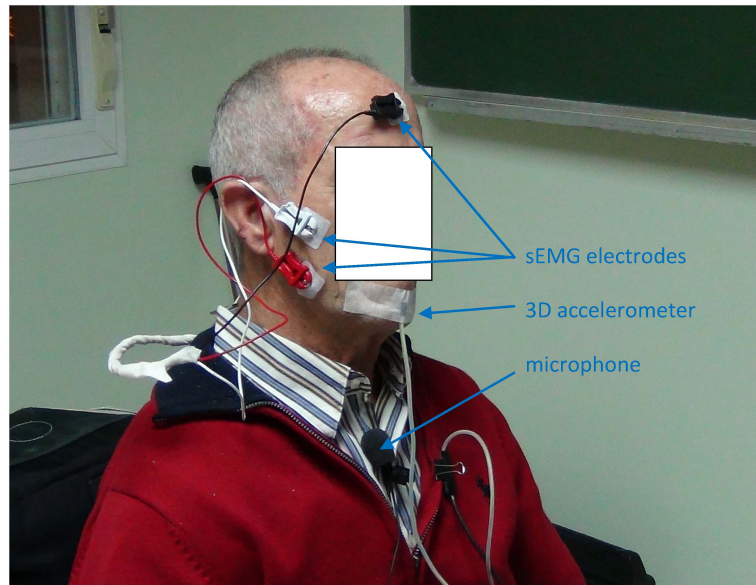


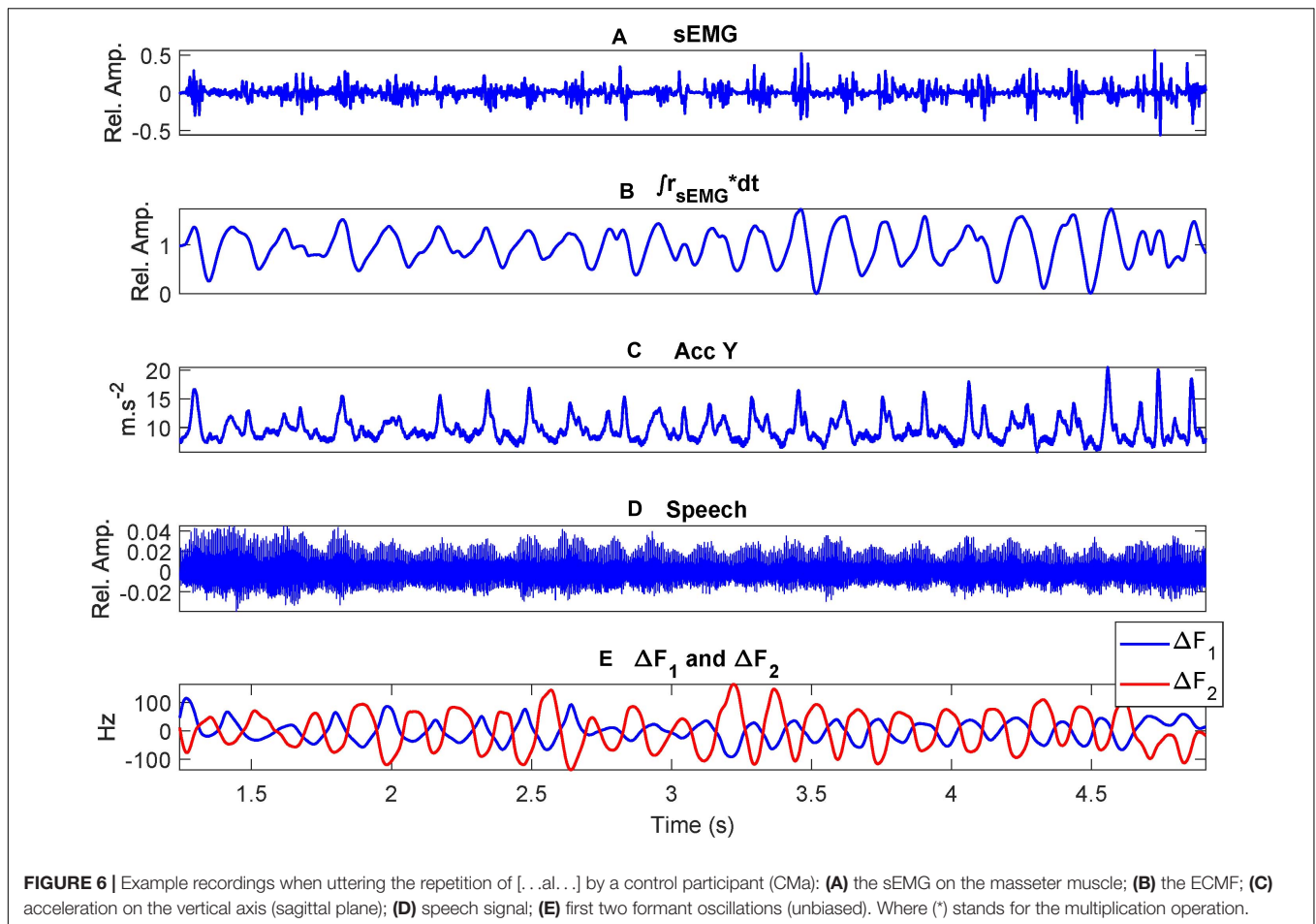
FIGURE 5 | Data recording from a normative male participant: a) the sEMG on the masseter muscle is taken by white and red contact electrodes on the masseter and a reference electrode (black) on the forehead; b) the 3D acceleration is acquired by a 3D accelerometer attached to the chin; the speech signal by a clip wireless cardioid microphone on the chest at 25 cm from the mouth. Jaw-fixed accelerometers are to be used only during signal recording to establish and validate a hands-off acoustic to kinematic projection model, to ultimate use only acoustic signals for at-home monitoring, producing indirect estimations of the neuromotor kinematic characteristics of the participant using the projection model resulting from the study. The participant consented the publication of this anonymized picture.

An exhaustive examination of the results presented in **Table 2** shows some interesting observations. For instance, the regression between the vertical displacement and the first formant oscillations (r_{21}) is the largest in absolute value for CMB, whereas Pfb shows the smallest one. No significant difference is appreciated between control and PD participants in this respect. The gain factor a_{21} is the largest for PMb, whereas for CMA it is the smallest. In this case, this factor is significantly largest for PD participants than for controls, a fact that would require further study. The regression between the vertical displacement and the second formant oscillations (r_{22}) is the largest for CFa, and the smallest for Pfb. No significant differences were found between control and PD participants. The smallest oscillation gain (a_{22}) was found for CFd, and the largest for PMb. In this case the estimations from controls and PD participants were also significant. The regression between ECMF and the dynamic vertical acceleration (r_{sd}) showed the largest value for Pfc, and the smallest for PMb. No significant differences were observed between the two datasets. The oscillation gain (w_{sd}) produced rather disperse estimations in this case, the largest value is observed for CMc and the lowest for Pfa. The regression results between ECMF and the vertical acoustic acceleration (r_{sf}) show the largest value for Pfa, and the smallest one for PMb (negative in this case, a counter-intuitive result, possibly due to an inefficient small amplitude recording due to low conductance, this fact requiring a further study). No significant differences were found between the two datasets. The oscillation gain (w_{sf}) showed again a strong dispersion, and the singularity of PMb producing a large and negative value again. This dispersion in the values of w_{sd} and w_{sf} may be related with a less efficient recording of the

sEMG in some cases, due to skin and fat conductance, as well as electrode placement. On its turn, the correlation between vertical dynamic and acoustic accelerations (r_{df}) showed large values, the difference between the largest (CMB) and the smallest (Pfb) being relatively small. No significant differences were observed between the two datasets. Finally, the gain factor between both accelerations (w_{df}) showed the largest and lowest values for PMd and Pfb. In this case, a significant difference was observed.

The results presented in **Figure 10** show that the distributions corresponding to the control participant extend to higher values of the absolute velocity ($5\text{--}6\text{ cm}\cdot\text{s}^{-1}$) than the values from the PD participant, which barely extend to $3\text{ cm}\cdot\text{s}^{-1}$ (AKV) and do not surpass $1.5\text{ cm}\cdot\text{s}^{-1}$ (AFV). This may be an indication of hypokinetic behavior, compatible with HD. This same situation was present in all the PD participants tested, as showed in the results found in **Table 3**. The statistical relevance of this different behavior has been assessed by three statistical tests: t -test, Kolmogorov-Smirnov and Mann-Whitney. The rejection of the null hypothesis of equality of distributions may be interpreted in the sense that the sets of JSD from the control and PD participants come from different distributions and are separable based on the value of their respective JSDs. Therefore, the p -values of the tests avail a statistical differentiation between controls and patients using JSDs estimated from AKV (dynamic) as well as from AFV (acoustic).

As a general summary it may be said that the correlation studies on sEMG relative to A_{yd} and A_{yf} presented relevant results both for control and PD participants, which means that the disease in itself is not a differentiation factor regarding the association of myoelectric, dynamic, and acoustic signals,



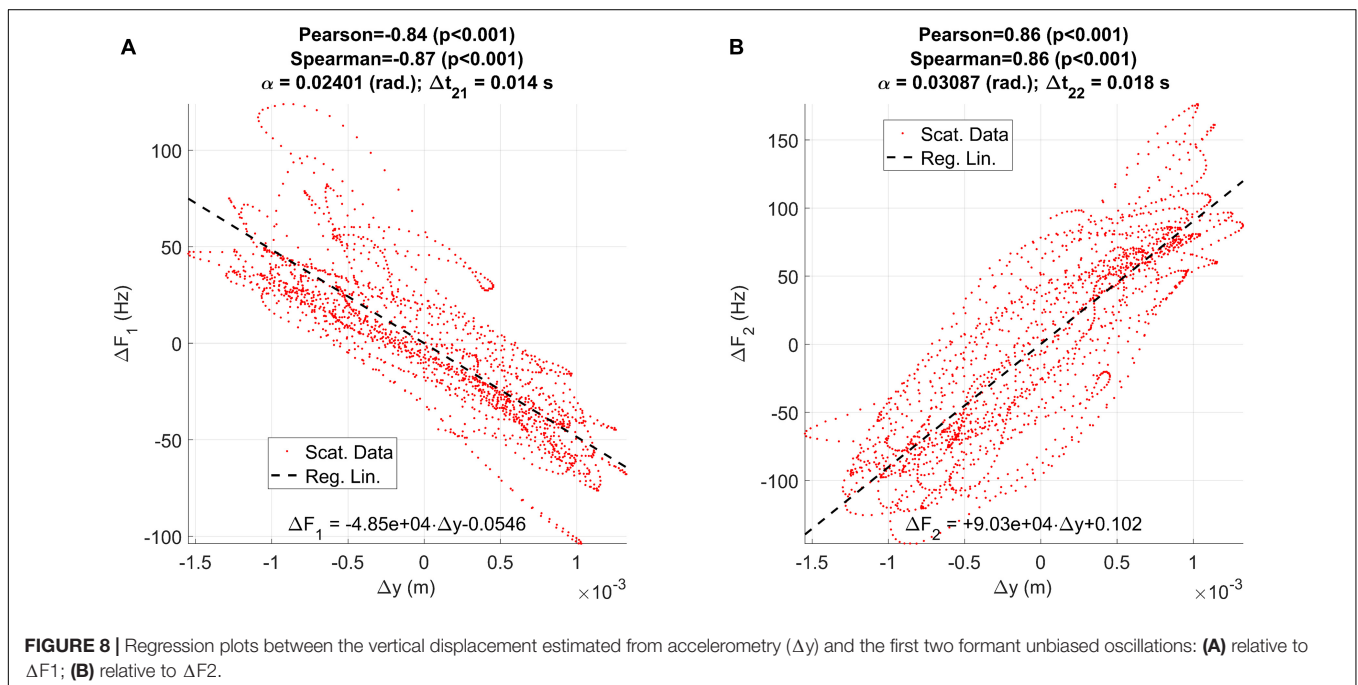
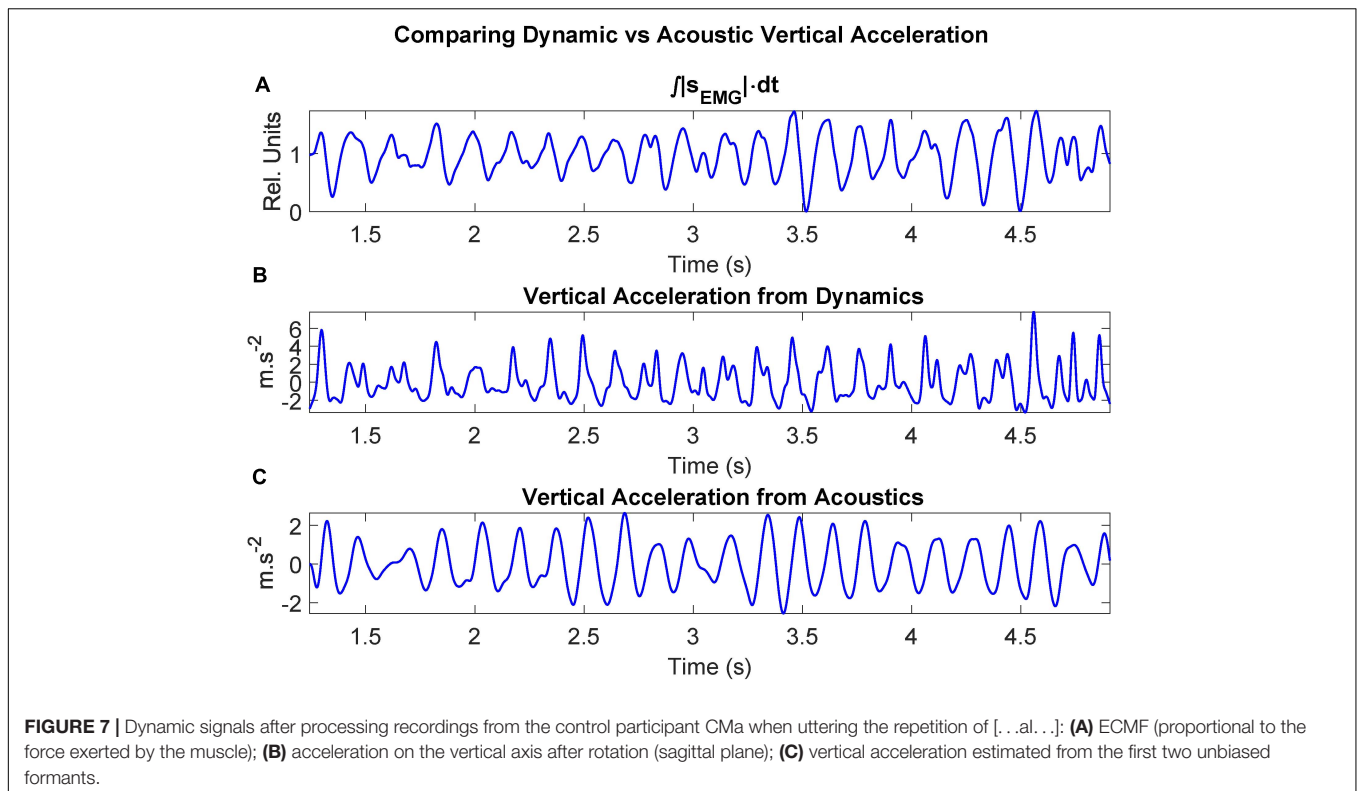
allowing to build chain models to infer the neuromotor activity in the masseter exclusively from acoustic estimates, therefore the assignment from acoustics to neuromotor (the objective of the present work) will work equivalently for both groups, allowing the design and use of an inverse model to project acoustic estimates to neuromotor ones.

Regarding the statistical tests reported on **Table 3**, a clear different behavior may be appreciated in the PD dataset with respect to the control dataset, as the JSDs of the PD dataset are larger to the average control distributions. The smallest JSDs are found in the control dataset, whereas the largest ones are found in the PD dataset. It may be seen also that the tests reject the null hypothesis of equal distributions between the control and PD participants' JSDs either estimated from dynamics (AKV) or from acoustics (AFV) under a 0.05 significance level. The differentiation between both groups is of most interest to assess HD by telemonitoring devices recording speech remotely. The validation of this possibility has been already studied using acoustic estimates only (Gómez et al., 2017, 2019a), but a wider study using both methods and ECMF is still pending.

The cross-correlation between the ECMF and the kinematic variables estimated also the delays for signal alignment. The following causes have been determined to explain the delays: the inertial dynamic behavior of the jaw-tongue structure,

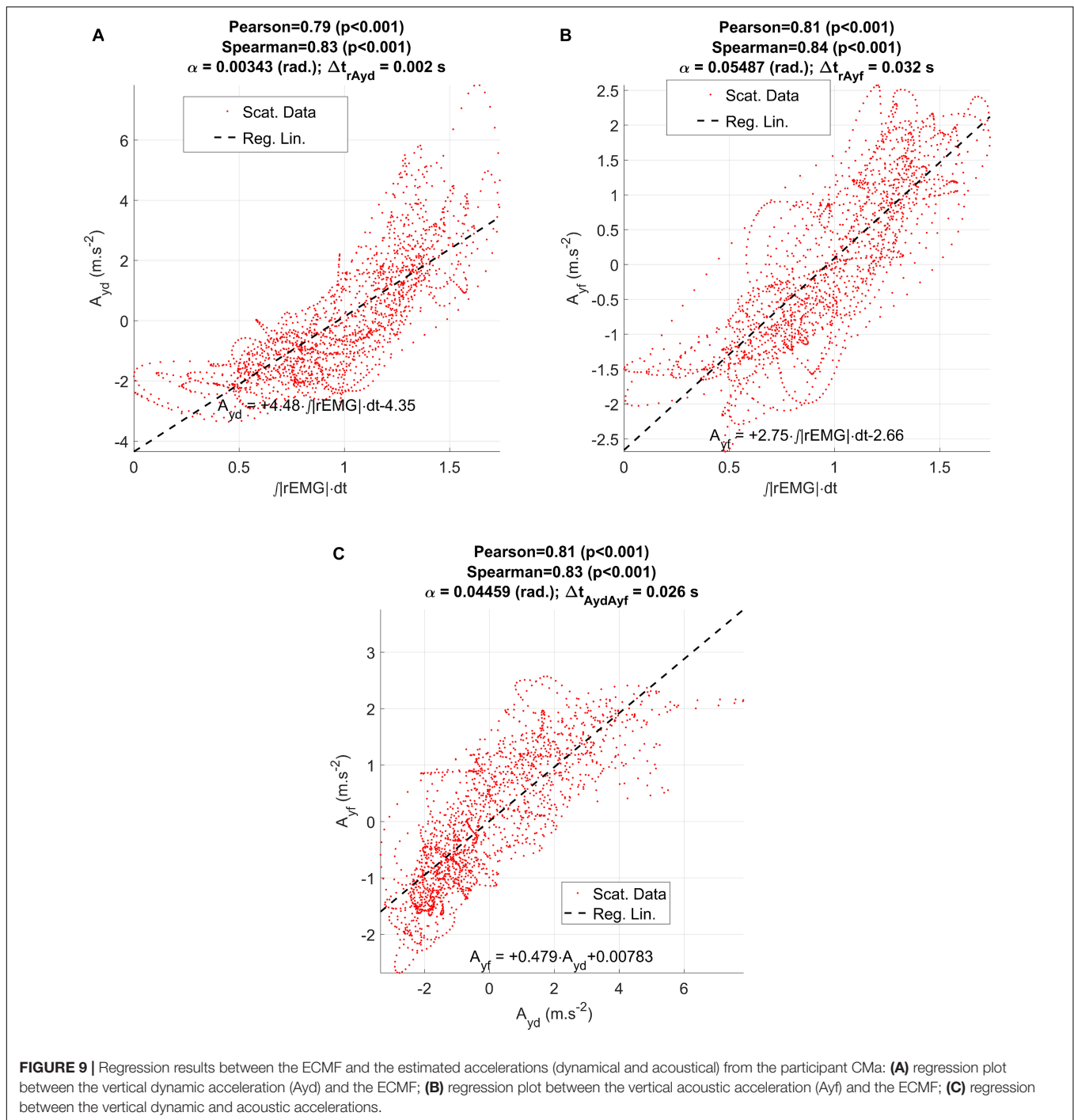
included in numerical integration (delay); the acoustic kinematic estimation, included in numerical differentiation (anticipation), and the most plausible hypothesis, yet to be tested, which would have to see with the time required for the standing waves in the resonances of the ONPT to attain enough intensity to create an emerging formant, and this process would have to see with the quality factor involved in the equivalent generating resonances, ultimately linked to the losses in the ONPT tissue walls (assumed to be rigid and non-viscous). This hypothesis may be checked tracking the pole positions after each time step (2 ms in the present study).

Apparently, the projection model for PD and control participants does not reveal strong differences as far as regression results show. On the contrary, the differentiation between both datasets is relevant in terms of absolute dynamic and acoustic kinematics. This observation must be taken with some precaution, given the important weaknesses and limitations affecting the study, among them the low number of participants studied, although a steady recruitment process is ongoing facing future studies. A non-deniable factor to be taken into account is the potential HD due to aging in the control group, as a confounding factor regarding neuromotor degeneration by disease. Another important limitation is that the low number of participants per gender did not allow a gender-separated study.



Another possible limiting fact is that all PD participants that have been recruited in the study were H&Y stage 2 and hence the generalization of our findings for different PD stages need to be evaluated in a follow-up study. This was a pragmatic constraint mandated by practical challenges in the recruitment. The study was conducted on members of PD associations in the southwest

area of Madrid. These participants had been diagnosed and followed by the public healthcare system and accepted willingly and enthusiastically participating. However, for ethical reasons we could not test them in the OFF state. Therefore, participants in an early H&Y stage 1 did not manifest motor problems associated to speech quite clearly, and most of the recordings were not valid.



On the other hand, participants in more advanced PD (above state 2) suffered from other co-morbidities, and it was challenging to recruit them for the needs of the study. Besides, we observed on the available participants in H&Y stage 3, that the recording session, although not lasting more than 25–30 min for a control participant, was for them an exhausting task. Indeed, most of the time was spent in the preparation of the face skin, fixing electrodes, accelerometer, and microphone. Recordings which would typically last 5–10 min for a control participant, almost

doubled in these cases, due to misunderstanding of instructions, repetitions, and participants' fatigue.

The effects of choosing PD participants in H&Y stage 2 only, could be assessed observing if the estimations of each feature listed in **Table 2** differed significantly between HC and PD participants. If both distributions did not differ significantly between HC and PD participants in H&Y 2, possibly it could be expected that distributions from PD participants in H&Y < 2 would not differ significantly as well. Extrapolating

TABLE 2 | Linear regression results.

Participants	CFa	CFb	CFc	CFd	CMA	CMB	CMc	CMd	PFa	PFb	PFc	PFd	PMa	PMb	PMc	PMd
r_{21}	-0.9	-0.77	-0.79	-0.88	-0.84	-0.94	-0.92	-0.91	-0.93	-0.64	-0.85	-0.9	-0.86	-0.93	-0.81	-0.91
a_{21} ($\times 10^3$)	-55.5	-58.1	-76.1	-55	-44.8	-78.4	-94.8	-67.5	-205	-74.1	-95.9	-115	-145	-289	-185	-159
r_{22}	0.94	0.86	0.76	0.85	0.86	0.88	0.87	0.86	0.84	0.53	0.93	0.84	0.81	0.89	0.65	0.83
a_{22} ($\times 10^3$)	110	152	135	43.1	86.1	62.0	99.6	71.5	258	119	173	151	173	267	153	181
r_{sd}	0.7	0.78	0.83	0.66	0.79	0.58	0.53	0.71	0.74	0.78	0.81	0.86	0.6	0.49	0.63	0.84
w_{sd}	36.9	8.81	5.01	8.84	4.48	9.24	95.8	35.0	3.94	5.52	30.9	7.27	32.2	26.4	14.2	15.5
r_{sf}	0.72	0.8	0.79	0.69	0.81	0.53	0.59	0.74	0.84	0.79	0.82	0.76	0.67	-0.41	0.68	0.83
w_{sf}	20.4	5.05	3.4	4.9	2.77	4.06	44.9	15.9	0.75	0.96	14.1	3.67	8.15	-14.4	4.23	14.4
r_{df}	0.92	0.87	0.82	0.87	0.81	0.95	0.79	0.93	0.87	0.77	0.91	0.91	0.81	0.91	0.78	0.93
w_{df}	0.50	0.49	0.58	0.46	0.49	0.46	0.40	0.38	0.15	0.13	0.41	0.52	0.18	0.59	0.22	0.87

r_{21} and r_{22} : Pearson's correlation coefficients between vertical displacement and the first two formants; a_{21} and a_{22} : projection weights defined in expression (1); r_{sd} , and r_{sf} : Pearson's correlation coefficients between vertical accelerations estimated from dynamic and acoustic signals and the force estimated from the sEMG; r_{df} : Pearson's correlation coefficient between the vertical accelerations estimated from dynamic and acoustic signals; w_{sd} , w_{sf} : regression slopes between vertical displacement and the dynamic force estimated from the sEMG; w_{df} : regression slope between the vertical accelerations estimated from dynamic and acoustic signals.

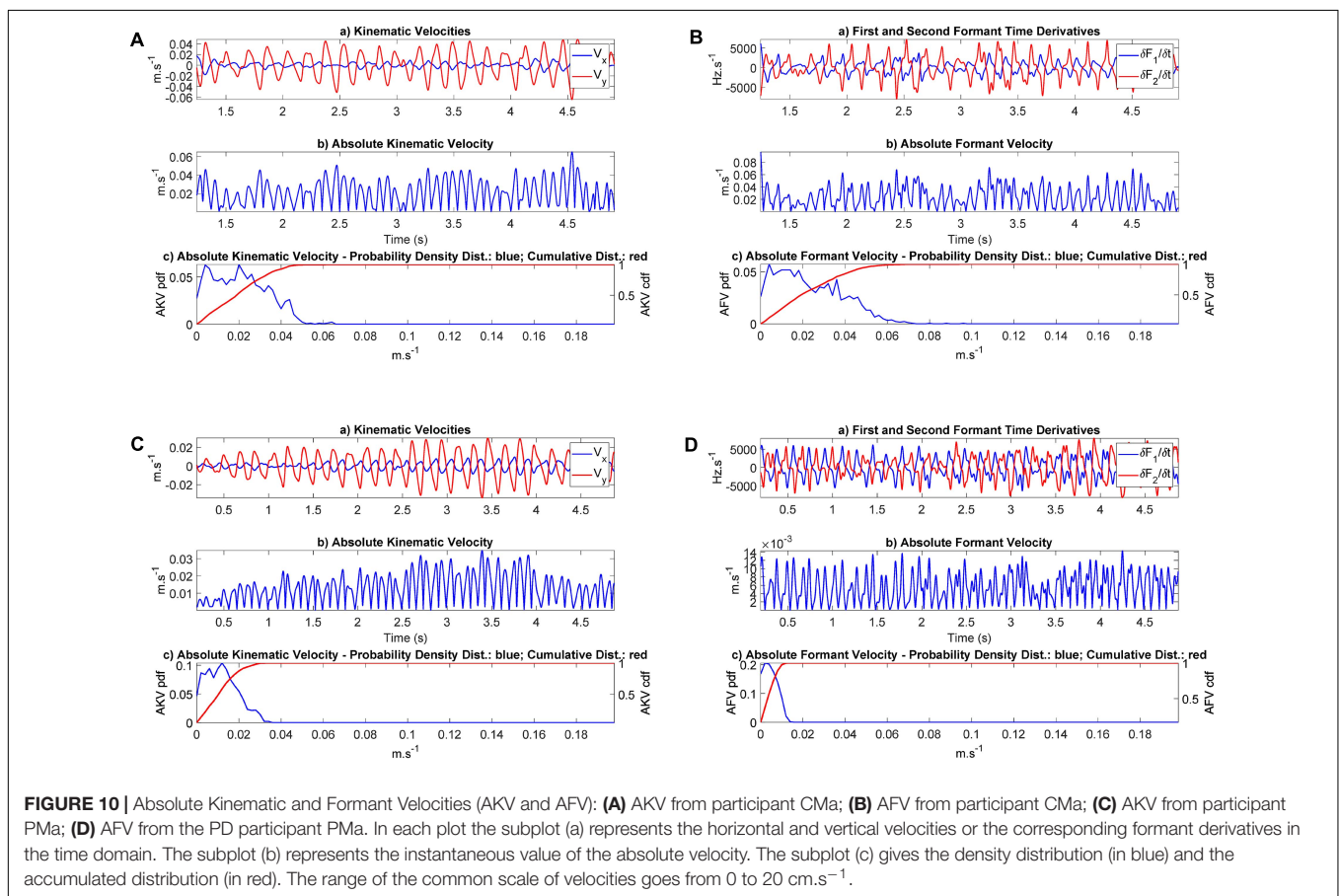


FIGURE 10 | Absolute Kinematic and Formant Velocities (AKV and AFV): **(A)** AKV from participant CMA; **(B)** AFV from participant CMA; **(C)** AKV from participant PMA; **(D)** AFV from the PD participant PMA. In each plot the subplot (a) represents the horizontal and vertical velocities or the corresponding formant derivatives in the time domain. The subplot (b) represents the instantaneous value of the absolute velocity. The subplot (c) gives the density distribution (in blue) and the accumulated distribution (in red). The range of the common scale of velocities goes from 0 to 20 cm.s^{-1} .

this observation to PD participants in $H\&Y > 2$ is not possible considering the data available right now. The comparisons have been done using three statistical tests on the null hypothesis of equal distributions: a parametric t -test, and two non-parametric ones, Kolmogorov-Smirnov (KS), and Mann-Whitney (MW) on the correlation and model projection coefficients given in **Table 2** (10 features, one per row). We adjusted the significance level of 0.05 per each feature accordingly to Holm-Bonferroni's

correction (Holm, 1979). The adjusted significance levels per feature (in parenthesis) were 0.0050 (1), 0.0056 (2), 0.0063 (3), 0.0071 (4), 0.0083 (5), 0.0100 (6), 0.0125 (7), 0.0167 (8), 0.0250 (9), 0.0500 (10). The sorted p -values per row (in parenthesis) were 0.0017 (4), 0.0064 (2), 0.1790 (8), 0.1966 (3), 0.4386 (10), 0.4881 (6), 0.5901 (7), 0.7176 (1), 0.7263 (5), 0.7832 (9). Therefore, the null hypothesis could only be rejected for feature a_{22} (4), because its p -value (0.0017) was under the lowest significance

level (0.0050). Comparing the results from the KS test in a similar way, the null hypothesis could only be rejected for feature a_{21} (p -value of 0.0014). The results from the MW test rejected the null hypothesis for both a_{21} and a_{22} (p -values of 0.0011), but not for the remnant features. The overall comparison results point to the global acceptance of the complementary hypothesis of equal means, except for the specific features a_{21} and a_{22} .

The effects of the H&Y stage choice could also be assessed observing if the JSD distances considered in **Table 3** differed significantly between HC and PD participants. As it may be seen, all the tests reject the null hypothesis of equal distributions of JSD scores respect to a significance level of 0.05, although the case of KS with JSD scores produced from the AKV distribution, is only slightly below the significance level (0.049). This means that probably this test (KS on JSD from AKV) would fail rejecting the null hypothesis for a set of PD participants in stage H&Y < 2. How the other tests would work under the same conditions is less predictable.

The unexplained contribution to the variance in linear regression results (in terms of $1-r^2$, where r is the respective correlation coefficient) by the effects of the ECMF on the vertical acceleration may be due to the action of the geniohyoid and other muscles contributing to the elevation and depression of the jaw-tongue system. The activity of the geniohyoid muscle in depressing the jaw-tongue structure has not been estimated in the present study. The inclusion of sEMG recording electrodes on geniohyoglossus muscle could complement it.

TABLE 3 | Results from JSD comparisons on the AKV and AFV distributions.

Participant's code	JSD/AKV	JSD/AFV
CFa	0.169	0.117
CFb	0.091	0.103
CFc	0.110	0.091
CFd	0.257	0.227
CMa	0.120	0.090
CMb	0.101	0.170
CMc	0.133	0.110
CMd	0.283	0.102
PFa	0.326	0.597
PFb	0.223	0.436
PFc	0.213	0.336
PFd	0.124	0.145
PMa	0.345	0.596
PMb	0.453	0.396
PMc	0.423	0.590
PMd	0.332	0.120
p -value t -Test	0.008	0.002
p -value KS	0.049	0.010
p -value MW	0.010	0.002

The mid and right columns give the JSDs of each participant with respect to the average AKV and AFV distributions of the control set, p_{PCd} and p_{PCr} , respectively. It may be seen that the JSDs of the PC participants are higher than the ones estimated on the control group. The p -values at the bottom of the table give the results of testing the JSDs from the PD set to the JSDs from the control participants. Values in bold specify the largest and smallest divergences with respect to the control averages.

There are several limitations affecting the present study, which have to be mentioned here. First of all, independent jaw-tongue movements are not analyzed, however, it must be stressed that the present study is mainly focused on proposing acoustic features which may be mapped to potential biomarkers of degenerative speech HD of neuromotor etiology, not on a general model in the sense of Ouni and Laprie (2005). Therefore, the proposed model does not cover the whole span of articulatory movements to acoustic features. Instead, reducing the movement space to vertical activity, helps in providing more accurate estimations of neuromotor activity on the masseter. A possible future extension of the study to independent tongue, jaw and lip movements would require monitoring the sEMG activity of other muscles, as the genio-hioglossus or the orbicular muscles. Besides, the model assumed linear relationship between formants and kinematics. Having in mind that the relationship with the biomarkers (acoustic and dynamic AKF and AKV) become non-linear, this is a necessary approach to allow a first conclusive study. Another important limitation is the small size of the dataset included in the study. This limitation will be removed when the pandemic conditions allow for the inclusion of more recordings. Once the projection model is validated using the multi-trait signals (speech, sEMG and 3DAcc) the detection properties of the proposed biomarkers (AKV and AFV) will be tested on larger PD speech databases (Sakar et al., 2013; Orozco-Arroyave et al., 2014), however, as they only contain speech data, we cannot use them for this validation phase, considering the novel direction that this work is proposing. Hopefully we will be able of using the projection model to produce the AKV and AFV biomarkers, to validate the statistical relevance of a speech-based home-monitoring approach on these databases and others recruited by our own using a tele-health platform (see Palacios et al., 2020). Another limiting factor is that PD participants were in a moderate stage of motor activity deterioration (H&Y stage 2). This limitation would require extending the analysis to PD participants across all stages to investigate the changes across the spectrum of the disease, from very early onset (HY stage 1) to very late PD stages.

Another important open issue is how speaker dependency and inter-speaker variability may affect absolute kinematic distributions as disease biomarkers. In this sense, it may be seen from the work of Dromey et al. (2013) that the variability of inter-speaker results for diphthong [aI] is the lowest from all diphthongs these authors have studied per class (control vs. PD participants). Moreover, it may be shown that the projection weights in **Table 2**, if normalized to the $[a_{ij}]$ vector norm as $\hat{a}_{ij} = a_{ij}/|a_{ij}|$ (see Gómez et al., 2021) do not show relevant inter-speaker variability per class. The way that inter-speaker variability may affect the absolute kinematic distributions (the proposed biomarkers), works different for controls than for PD participants, because these biomarkers are the normalized histogram distributions of the absolute kinematic velocities, and they only express if movement is more distributed toward low absolute velocities (hypokinetic), as is the case for the PD participant plotted in **Figures 10C,D** or if it spreads along low and high absolute velocities (normokinetic), as is the case for the control participant plotted in **Figures 10A,B**.

Therefore, inter-speaker variability is expressing if the speaker is hypokinetic or not, or to put it in other words, inter-speaker variability between controls and PD participants is an essential role of the proposed biomarkers, provided they are designed to differentiate hypokinetic from normokinetic speech. This separation is availed by the results shown in **Table 3**, where it may be seen that biomarker distributions (kinematic AKV, as well as acoustic AFV) from PD participants show larger distances (in terms of Jensen-Shannon Divergence) to distributions from control participants. Summarizing, the biomarkers proposed show an inter-speaker variability oriented to differentiate control participants from PD participants, in closed relationship with the second objective of the study as stated in section “Objectives.”

CONCLUSION

This study presents a multi-trait evaluation of HD, including speech, 3D accelerometry and sEMG signal acquisition on the masseter. A cross-correlation signal alignment is carried on these signals. A further regression analysis on the re-aligned acoustic formants, kinematic accelerations and dynamic-related electromyography signals allows the estimation of the projection model parameters. The projected acoustic-kinematic and dynamics-kinematics absolute movement distributions are used as HD biomarkers. An evaluation of these biomarkers in terms of Jensen-Shannon Divergence is conducted on data from control and PD participants, showing the capability of both biomarkers to detect HD in PD speech. The main findings derived from the study are the following:

- A multimodal framework for the assessment of the masseter’s neuromotor activity based on sEMG, 3DAcc and speech, has been used on diadochokinetic exercises from PD and control participants.
- Large cross-correlations between the measured and estimated signals have been observed using linear regression on a small-size data sample of control and PD participants.
- The cross-correlation study did not show relevant differences between control and PD participants.
- The articulation kinematics estimated from 3D accelerometry and from acoustics showed a relevant similarity for all the participants included in the study.
- It was possible to differentiate the speech kinematic behavior of the control and PD participant sets used in this study, using absolute velocities (dynamic and acoustic) of the joint jaw-tongue reference point. Therefore, their amplitude distributions could be used as potential biomarkers of PD speech HD.

The findings of the current study are preliminary, given the limited sample size. We are currently extending our efforts

REFERENCES

Alku, P., Murtola, T., Malinen, J., Kuortti, J., Story, B., Airaksinen, M., et al. (2019). OPENGLot-An open environment for the evaluation of glottal inverse filtering. *Speech Commun.* 107, 38–47. doi: 10.1016/j.specom.2019.01.005

toward the collection of a larger sample size and investigating how well the presented findings generalize across larger cohorts. The definition of an inverse model to infer neuromotor activity from acoustics is also a future objective. The inclusion of sEMG from other muscles active on the jaw-tongue articulation as the geniohyoid is also foreseen. A study regarding the delay on formant-estimated kinematics based on the non-linear properties of the ONPT is considered as well.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Comité de Ética de la Universidad Politécnica de Madrid. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AG: manuscript preparation, model conceptual design, and implementation, instrumentation setup, and data recording. PG: research guidelines, manuscript preparation, and model implementation. DP and VR: data recording and manuscript preparation. VN: instrumentation design and setup and data recording. AA: data processing. AT: research guidelines and manuscript preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was being funded by the grants TEC2016-77791-C4-4-R from the Government of Spain, and CENIE_TECA-PARK_55_02 INTERREG V-A Spain—Portugal (POCTEP).

ACKNOWLEDGMENTS

The authors wish to thank Asociación de Parkinson de Alcorcón y Móstoles (APARKAM), their members for their enthusiast collaboration, and to Azucena Balandín (director) and Zoraida Romero (speech therapist) for their help and advice.

Arora, S., Lo, C., Hu, M., and Tsanas, A. (2021). Smartphone speech testing for symptom assessment in rapid eye movement sleep behavior disorder and Parkinson’s disease. *IEEE Access* (in press).

Bourchard, K. E., Conant, D. F., Anumanchipalli, G. K., Dichter, B., Chaisanguanthum, K. S., Johnson, K., et al. (2016). High-Resolution,

- non-invasive imaging of upper vocal tract articulators compatible with human brain recordings. *PLoS One* 11:e0151327. doi: 10.1371/journal.pone.0151327
- Brabenec, L., Mekyska, J., Galaz, Z., and Rektorova, I. (2017). Speech Disorders in Parkinson's disease: early diagnostics and effects on medication in brain stimulation. *J. Neural Transm.* 124, 303–334. doi: 10.1007/s00702-017-1676-0
- Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Tourkeltaub, P., and Liotti, M. (2009). The somatotopy of speech: phonation and articulation in the human motor cortex. *Brain Cogn.* 70, 31–41. doi: 10.1016/j.bandc.2008.12.006
- Buchallard, S., Perrier, P., and Payan, Y. (2009). A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning. *J. Acoust. Soc. Am.* 126, 2033–2051. doi: 10.1121/1.3204306
- Castroflorio, T., Farina, D., Bottin, A., Piancino, M. G., Bracco, P., and Merletti, R. (2005). Surface EMG of jaw elevator muscles: effect of electrode location and inter-electrode distance. *J. Oral Rehabil.* 32, 411–417. doi: 10.1111/j.1365-2842.2005.01442.x
- Cattaneo, L., and Pavesi, G. (2014). The facial motor system. *Neurosci. Behav. Rev.* 38, 135–159. doi: 10.1016/j.neubiorev.2013.11.002
- Cover, T. M., and Thomas, J. A. (2006). *Elements of Information Theory*. New York, NY: Wiley.
- De Luca, C. J. (1997). The use of surface electromyography in biomechanics. *J. Appl. Biomech.* 13, 135–163. doi: 10.1123/jab.13.2.135
- De Luca, C. J., Gilmore, L. D., Kuznetsov, M., and Roy, S. H. (2010). Filtering the surface EMG signal: movement artifact and baseline noise contamination. *J. Biomech.* 43, 1573–1579. doi: 10.1016/j.jbiomech.2010.01.027
- Deller, J. R., Proakis, J. G., and Hansen, J. H. L. (1993). *Discrete-Time Processing of Speech Signals*. New York, NY: Macmillan.
- Demonet, J. F., Thierry, G., and Cardebat, D. (2005). Renewal of the neurophysiology of language: functional neuroimaging. *Physiol. Rev.* 85, 49–95. doi: 10.1152/physrev.00049.2003
- Dromey, C., Jang, G. O., and Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Commun.* 55, 315–328. doi: 10.1016/j.specom.2012.09.001
- Duffy, J. R. (2013). *Motor Speech Disorders*. St. Louis, MO: Elsevier.
- Farina, D., and Merletti, R. (2001). A novel approach for precise simulation of the EMG signal detected by surface electrodes. *IEEE Trans. Biomed. Eng.* 48, 637–646. doi: 10.1109/10.923782
- Gerard, J. M., Perrier, P., and Payan, Y. (2006). “3D biomechanical tongue modeling to study speech production,” in *Speech Production: Models, Phonetic Processes, and Techniques*, Chap. 6. eds J. Harrington and M. Tabain (New York, NY: Psychology Press), 85–102.
- Goberman, A. M., and Coelho, C. (2002). Acoustic analysis of Parkinsonian speech I: speech characteristics and L-Dopa therapy. *Neurorehabilitation* 17, 237–246. doi: 10.3233/NRE-2002-17310
- Godino, J. I., Shattuck-Hufnagel, S., Choi, J. Y., Moro, L., and Gómez, J. A. (2017). Towards the identification of Idiopathic Parkinson's Disease from the speech. New articulatory kinematic biomarkers. *PLoS One* 12:e0189583. doi: 10.1371/journal.pone.0189583
- Gómez, A., Tsanas, A., Gómez, P., Palacios, D., Rodellar, V., and Álvarez, A. (2021). Acoustic to kinematic projection in Parkinson's Disease Dysarthria. *Biomed. Signal Process. Control* 66:102422.
- Gómez, P., Gómez, A., Ferrández, J. M., Mekyska, J., Palacios, D., Rodellar, V., et al. (2019a). Neuromechanical modelling of articulatory movements from surface electromyography and speech formants. *Int. J. Neural Syst.* 29:1850039. doi: 10.1142/S0129065718500399
- Gómez, P., Mekyska, J., Ferrández, J. M., Palacios, D., Gómez, A., Rodellar, V., et al. (2017). Parkinson disease detection from speech articulation neuromechanics. *Front. Neuroinformatics* 11:56. doi: 10.3389/fninf.2017.00056
- Gómez, P., Mekyska, J., Gómez, A., Palacios, D., Rodellar, V., and Álvarez, A. (2019b). Characterization of Parkinson's disease dysarthria in terms of speech articulation kinematics. *Biomed. Signal Process. Control* 52, 312–320. doi: 10.1016/j.bspc.2019.04.029
- Goodman, S. E., and Hasson, C. J. (2017). Elucidating sensorimotor control principles with myoelectric musculoskeletal models. *Front. Hum. Neurosci.* 11:531. doi: 10.3389/fnhum.2017.00531
- Green, J. R., Yunusova, Y., Kuruvilla, M. S., Wang, J., Pattee, G. L., Synhorst, L., et al. (2013). Bulbar and speech motor assessment in ALS: challenges and future directions. *Amyotroph. Lateral Scler. Frontotemporal Degener.* 14, 494–500. doi: 10.3109/21678421.2013.817585
- Greenberg, S. (2004). *Speech Processing in the Auditory System*. New York, NY: Springer.
- Harel, B., Cannizzaro, M., and Snyder, P. J. (2004). Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. *Brain Cogn.* 56, 24–29. doi: 10.1016/j.bandc.2004.05.002
- Harel, B. T., Cannizzaro, M. S., Cohen, H., Reilly, N., and Snyder, P. J. (2004). Acoustic characteristics of Parkinsonian speech: a potential biomarker of early disease progression and treatment. *J. Neurolinguistics* 17, 439–453. doi: 10.1016/j.jneuroling.2004.06.001
- Hogan, N. (1984). Adaptive control of mechanical impedance by coactivation of antagonist muscles. *IEEE Trans. Aut. Cont.* 29, 681–690. doi: 10.1109/TAC.1984.1103644
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.
- IPA (2005). *International Phonetic Association*. Available online at: <https://www.internationalphoneticalphabet.org/ipa-charts/ipa-symbols-chart-complete/> (accessed January 08, 2021).
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neurosci. Behav. Rev.* 26, 235–258. doi: 10.1016/S0149-7634(01)00068-9
- Kandel, E., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J. (eds). (2013). *Principles of Neural Science*. New York, NY: McGraw-Hill.
- Lee, K. S. (2010). Prediction of acoustic feature parameters using myoelectric signals. *IEEE Trans. Biomed. Eng.* 57, 1587–1595. doi: 10.1109/TBME.2010.2041455
- Martínez-Valdés, E., Negro, F., Laine, C. M., Falla, D., Mayer, F., and Farina, D. (2017). Tracking motor units longitudinally across experimental sessions with high-density surface electromyography. *J. Physiol.* 595, 1479–1496. doi: 10.1113/JP273662
- Mefferd, A. S., and Dietrich, M. S. (2020). Tongue- and jaw-specific articulatory changes and their acoustic consequences in talkers with dysarthria due to amyotrophic lateral sclerosis: effects of loud, clear, and slow speech. *J. Speech Lang. Hear. Res.* 63, 2625–2636. doi: 10.1044/2020_JSLHR-19-00309
- Mekyska, J., Janousova, E., Gómez-Vilda, P., Smekal, Z., Rektorova, I., Eliasova, I., et al. (2015). Robust and complex approach of pathological speech signal analysis. *Neurocomputing* 167, 94–111. doi: 10.1016/j.neucom.2015.02.085
- Mitra, V., Sivaraman, G., Nam, H., Espy-Wilson, C., Saltzman, E., and Tiede, M. (2017). Hybrid convolutional neural networks for articulatory and acoustic information based speech recognition. *Speech Commun.* 89, 103–112. doi: 10.1016/j.specom.2017.03.003
- Orozco-Arroyave, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J. F., González-Rátiva, M. C., and Nöth, E. (2014). “New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease,” in *Proceedings of the 9th Language Resources and Evaluation Conference, (LREC)*, Reykjavik, 342–347.
- Ouni, S., and Laprie, Y. (2005). Modeling the articulatory space using a hypercube codebook for acoustic-to-articulatory inversion. *J. Acoust. Soc. Am.* 118, 444–460. doi: 10.1121/1.1921448
- 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 Early Access* 8, 188243–188255. doi: 10.1109/ACCESS.2020.3031646
- Parkinson, J. (2002). An essay on the shaking palsy. *J. Neuropsychiatry Clin. Neurosci.* 14, 223–236. (Re-edited in *Neuropsychiatry Classics from the 1817 original monograph with the same title by J. Parkinson*, edited by Sherwood, Neely and Jones).
- Qin, C., and Carreira-Perpiñán, M. A. (2007). “An empirical investigation of the non-uniqueness in the acoustic-to-articulatory mapping,” in *Proceedings of the INTERSPEECH, August 27–31, Antwerp*, 74–77.
- Ricciardi, L., Ebreo, M., Graziosi, A., Barbuto, M., Sorbera, C., Morgante, L., et al. (2016). Speech and gait in Parkinson's disease: when rhythm matters. *Parkinsonism Relat. Disord.* 32, 42–47.
- Roark, R. M., Li, J. C. L., Schaefer, S. D., Adam, A., and De Luca, C. J. (2002). Multiple motor unit recordings of laryngeal muscles: the technique of vector laryngeal electromyography. *Laryngoscope* 112, 2196–2203. doi: 10.1097/00005537-200212000-00014

- Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., et al. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE J. Biomed. Health Inform.* 17, 828–834. doi: 10.1109/JBHI.2013.2245674
- Sanguinetti, V., Laboisière, R., and Payan, Y. (1997). A control model of human tongue movements in speech. *Biol. Cybern.* 77, 11–22. doi: 10.1007/s004220050362
- Sapir, S. (2014). Multiple factors are involved in the dysarthria associated with Parkinson's disease: a review with implications for clinical practice and research. *J. Speech Lang. Hear. Res.* 57, 1330–1343. doi: 10.1044/2014_JSLHR-S-13-0039
- Sivaraman, G., Mitra, V., Nam, H., Tiede, M., and Espy-Wilson, C. (2019). Unsupervised speaker adaptation for speaker independent acoustic to articulatory speech inversion. *J. Acoust. Soc. Am.* 146, 316–329. doi: 10.1121/1.5116130
- Skodda, S., Grönheit, W., and Schlegel, U. (2012). Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. *PLoS One* 7:e32132. doi: 10.1371/journal.pone.0032132
- Tasko, S. M., and Greilick, K. (2010). Acoustic and articulatory features of diphthong production: a speech clarity study. *J. Speech Lang. Hear. Res.* 53, 84–99. doi: 10.1044/1092-4388(2009/08-0124)
- Teklemariam, A., Hodson-Tole, E. F., Reeves, N. D., Consten, N. P., and Cooper, G. (2016). A finite element model approach to determine the influence of electrode design and muscle architecture on myoelectric signal properties. *PLoS One* 11:e0148275. doi: 10.1371/journal.pone.0148275
- Tjaden, K., Lam, J., and Wilding, G. (2013). Vowel acoustics in Parkinson's disease and multiple sclerosis: comparison of clear, loud, and slow speaking conditions. *J. Speech Lang. Hear. Res.* 56, 1485–1502. doi: 10.1044/1092-4388(2013/12-0259)
- Tsanas, A. (2012). *Accurate Telemonitoring of Parkinson's Disease Symptom Severity Using Nonlinear Speech Signal Processing and Statistical Machine Learning*. Ph.D. thesis, University of Oxford, Oxford.
- Tsanas, A., Little, M. A., McSharry, P. E., and Ramig, L. O. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *J. R. Soc. Interface* 8, 842–855. doi: 10.1098/rsif.2010.0456
- Tsanas, A., Little, M. A., and Ramig, L. O. (2021). Remote assessment of Parkinson's disease symptom severity using the simulated cellular mobile telephone network. *IEEE Access* 9, 11024–11036. doi: 10.1109/ACCESS.2021.3050524
- Whitfield, J. A., and Goberman, A. M. (2014). Articulatory-acoustic vowel space: application to clear speech in individuals with Parkinson's disease. *J. Comm. Disord.* 51, 19–28. doi: 10.1016/j.jcomdis.2014.06.005
- Wu, P., González, I., Patsis, G., Jiang, D., Sahli, H., Kerckhofs, E., et al. (2014). Objectifying facial expressivity assessment of Parkinson's patients: preliminary study. *Comput. Math. Methods Med.* 2014:427826. doi: 10.1155/2014/427826
- Yunusova, Y., Green, J., Greenwood, L., Wang, J., Pattee, G. L., and Zinman, L. (2012). Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. *Folia Phoniatr. Logop.* 64, 94–102. doi: 10.1159/000336890
- Yunusova, Y., Weismer, G. G., and Lindstrom, M. J. (2011). Classifications of vocalic segments from articulatory kinematics: healthy controls and speakers with Dysarthria. *J. Speech Lang. Hear. Res.* 54, 1302–1311. doi: 10.1044/1092-4388(2011/09-0193)

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Gómez, Gómez, Palacios, Rodellar, Nieto, Álvarez and Tsanas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.