

Neuromechanical Modelling of Articulatory Movements from Surface Electromyography and Speech Formants

Pedro Gómez-Vilda^{*,††}, Andrés Gómez-Rodellar^{*}, José M. Ferrández Vicente[†],
Jiri Mekyska[‡], Daniel Palacios-Alonso^{§,*}, Victoria Rodellar-Biarge^{*},
Agustín Álvarez-Marquina^{*}, Ilona Eliasova^{¶,||},
Milena Kostalova^{**||} and Irena Rektorova^{¶,||}

**Neuromorphic Speech Processing Lab*

*Center for Biomedical Technology, Universidad Politécnica de
Madrid Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain*

†Universidad Politécnica de Cartagena

Campus Universitario Muralla del Mar, Pza. Hospital 1, 30202 Cartagena, Spain

*‡Department of Telecommunications, Brno University of Technology
Technická 10, 61600 Brno, Czech Republic*

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

*¶First Department of Neurology, Faculty of Medicine and St. Anne's University Hospital
Masaryk University, Pekarska 53, 656 91 Brno, Czech Republic*

*||Applied Neuroscience Research Group, Central European Institute of Technology
CEITEC, Masaryk University, Kamenice 753/5, 625 00 Brno, Czech Republic*

***Department of Neurology, Faculty Hospital and Masaryk University
Jihlavská 20, 63900 Brno, Czech Republic*

††pedro@fi.upm.es

Speech articulation is produced by the movements of muscles in the larynx, pharynx, mouth and face. Therefore speech shows acoustic features as formants which are directly related with neuromotor actions of these muscles. The first two formants are strongly related with jaw and tongue muscular activity. Speech can be used as a simple and ubiquitous signal, easy to record and process, either locally or on e-Health platforms. This fact may open a wide set of applications in the study of functional grading and monitoring neurodegenerative diseases. A relevant question, in this sense, is how far speech correlates and neuromotor actions are related. This preliminary study is intended to find answers to this question by using surface electromyographic recordings on the masseter and the acoustic kinematics related with the first formant. It is shown in the study that relevant correlations can be found among the surface electromyographic activity (dynamic muscle behavior) and the positions and first derivatives of the first formant (kinematic variables related to vertical velocity and acceleration of the joint jaw and tongue biomechanical system). As an application example, it is shown that the probability density function associated to these kinematic variables is more sensitive than classical features as Vowel Space Area (VSA) or Formant Centralization Ratio (FCR) in characterizing neuromotor degeneration in Parkinson's Disease.

††Corresponding author.

1. Introduction

Speech is produced by different cognitive processes, mapped through basal ganglia to agonist-antagonist neuromotor actions on articulation muscles,^{1,2} which modulate phonation by the resonances of the oronasopharyngeal tract (ONPT), and is radiated into air as sound. The speech neuromotor action sequences activate the muscles of the pharynx, tongue, jaw, lips, larynx, chest and diaphragm. The source-filter model of speech production suggests that an excitation source is generated either by the joint action of chest and larynx muscles (phonation) or by turbulent airflow resulting from air exhalation in different parts of the ONPT. Therefore, possible alterations of the speaker's neuromotor conditions may be reflected in speech,^{3,4} as it happens with certain diseases, because they affect mainly neuromotor units in the basal ganglia, brain stem and cerebellar structures.⁵⁻⁷ Neurological diseases may affect different levels of speech production, as phonation and articulation or tone⁸⁻¹² as well as improper speech planning and emotional impairment.¹³ The present work is intended to explore the connections between neuromotor actions on certain articulation muscles and meaningful acoustical correlates, as speech formants in Parkinson's Disease (PD). The working hypothesis assumes that the kinematic properties of acoustic-phonetic variables derived from the first two speech formants (F_1 , F_2) are the direct consequence of the activity of articulation muscles (mainly the masseter), and that surface electromyographic (sEMG) signals^{14,15} measured on this muscle might be highly correlated with the neuromotor actions governing muscle contractions. The force exerted by articulation muscles will modulate the transversal section of the vocal tract, opening or closing it, in what is known as the high/low (H/L) phonetic feature of vowels¹⁶ corresponding to the articulation gesture of raising or lowering the tongue with respect to the palate ceiling. As it will be later discussed, it has been assumed that the force exerted by articulation muscles may be related to the integral

of the rectified sEMG. If this is so, this force should find a correlate on the vertical position of a reference point in the center of masses of the complex jaw-tongue and associated tissues. This position would be a determinant geometrical gesture (articulation) associated with the high/low quality of the vowel produced by phonation (glottal source) being filtered by the vocal tract (filter). This theory is solidly established since the works of G. Fant in the 60s of the past century.¹⁷ This relationship would explain the behavior of the two first formants, which are the main acoustic correlates of the vocal tract resonances. Although correlation does not imply causality, when a clear physical model linking the correlate variables is known and made explicit, correlation is a valuable tool to quantify causality effects. In the case under examination, consider the following causal links: Neuromotor Activity \rightarrow Masseter sEMG \rightarrow Vertical Force \rightarrow Vertical Position \rightarrow $\{\Delta F_1, \Delta F_2\}$. An explicit relationship may be established for each of these causal links expressed by " \rightarrow " as shown in Sec. 2. Therefore, correlation is to be seen in this case as the quantitative expression of causality. Having in mind that the force exerted by the masseter could not be inferred directly, as it is done in isometric experiments, because the masseter is not working isometrically in speech, the present study opens the possibility of making this estimation possible relating force and sEMG by means of accelerometry. The use of sEMG (to infer neuromotor action and forces), accelerometry (to estimate positions) and formant kinematics (to measure acoustic properties) is fully justified if a strong correlation can be measured between neuromotor activity (sEMG) and formant kinematics. In such a case, this last feature can be used to infer neuromotor activity in further research. The main advantage in doing so is that formant kinematics may be recorded by a "contactless" microphone, making speech a simple and useful vehicular tool to further neuromotor studies.

The main objective of the paper is to show that the relationship between neuromotor activity on the

masseter estimated from its sEMG may be quantitatively associated with the kinematic activity of the jaw estimated from accelerometric measurements on the chin, and to quantify the relationship between jaw kinematics and acoustic estimates derived from speech (first two formants). Once these relationships are established through a neuromechanical model, it would be possible, on the one hand, to infer neuromotor activity from acoustic signals, which are easily obtained from microphone recordings, and on the other hand, to use correlates derived from acoustic signals to estimate kinematic activity which could be used to characterize the dysarthric speech associated to neurodegenerative causes, as in PD hypokinetic dysarthria (HD). Therefore, the paper has to be seen as divided in two main parts. The first part, including Secs. 2 and 3 is devoted to describe a biomechanical model associating quantitatively the acoustic signals to formant kinematics, and to estimate the relationship between sEMG and kinematics derived from accelerometry using regression methods. Based on the association between acoustic correlates and neuromotor activity empirically validated in the first part, the second part of the paper, including Secs. 4 and 5 is devoted to show that the kinematic activity described from acoustic correlates is a better descriptor than other classical features as the Vowel

Space Area (VSA) and the Formant Centralization Ratio (FCR)¹⁸ to characterize dysarthric speech in PD using statistical validation. Section 6 is devoted to summarize the main conclusions derived from the study.

2. Neuromechanic Articulation Model

The muscular structures implied in articulation modifying the ONPT are the naso-pharyngeal switch, and the jaw, tongue and lips, which change the properties of the equivalent acoustic filter.¹⁶ It may be said that the relation between the first two formant positions and the nature of a vowel is well established in literature.^{19–23} Specifically, it will be observed that a relation exists between the vowel space and the two main articulation gestures, which are the jaw and tongue positions.^{24–26} The jaw position determines the first two formant positions (F_1), in the sense that low (open) vowels articulated with a low jaw tend to show a high F_1 and a low F_2 , and vice versa (see Fig. 1). The lower jaw position is mainly fixed by two muscles which act as agonist (masseter) and antagonist (geniohyoid).^{27,28} The action of gravity is also to be considered (f_w) together with the forces produced by agonist and antagonist muscles (f_m, f_g).²⁹

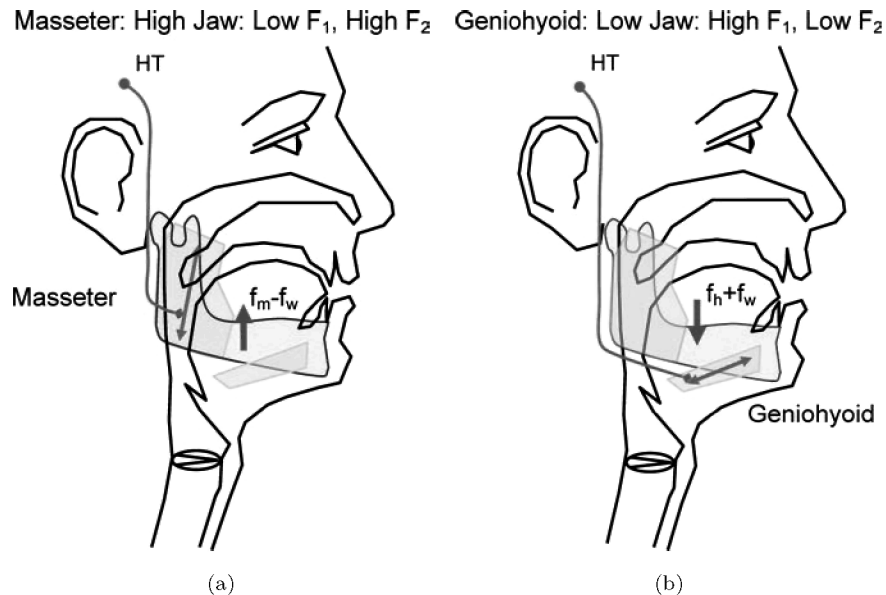


Fig. 1. Agonist-antagonist neuromotor actions relating jaw position and the vowel feature high/low. (a) The neuromotor actions produced by hypothalamus (HT) neurons activate the masseter. The result is a force (f_m) acting against gravity (f_w) moving the jaw upwards. (b) The neuromotor action activating the geniohyoid muscle produces a force (f_h) in the sense of gravity (f_w) pulling the jaw downwards.

In many articulation gestures, the jaw, tongue and surrounding facial tissues behave as a joint dynamic structure, which may be represented by a lumped moment of inertia following classical muscular biomechanics.³⁰ The biomechanics of this system may be simplified as shown in Fig. 2, considering a descriptive model of the structures, forces and displacements of a jaw-tongue reference point (P_{rJT}), equivalent to a center of actions, which could be related to formant positions.

When its coordinates experience a modification ($\Delta x_r, \Delta y_r$), a corresponding change in the first two formants $\{\Delta F_1, \Delta F_2\}$ is to be expected. Lowering the P_{rJT} will result in an elevation of F_1 , and vice versa. Similarly, advancing the P_{rJT} will result in an elevation of F_2 , and vice versa. The back-front (B/F) and high-low (H/L) phonetic features may be justified on these articulation principles.¹⁶ This relationship can be established as

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

where a_{ij} are the transformation weights explaining the position-to-formant associations, and t is the time. This relationship is known to be one-to-many in very specific articulation gestures, i.e. the same pair of formants $\{F_1, F_2\}$ could be associated to more than a single articulation position.²⁵ Nevertheless,

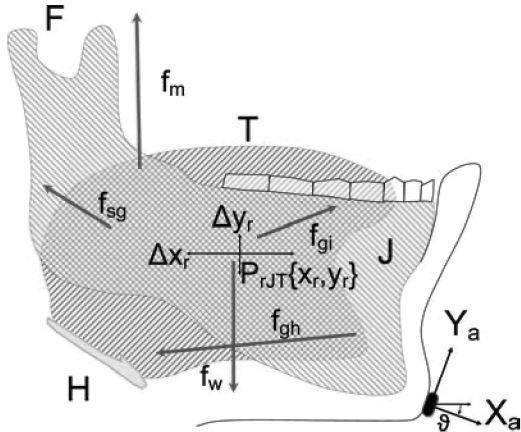


Fig. 2. Jaw-tongue biomechanical system. The jaw (J) is joined to the skull bone at fulcrum as in a third-class lever system. The tongue (T) is supported by jaw and the hyoid bone. F: fulcrum, T: tongue, J: jaw, H: hyoid; f_m : masseter force, f_{sg} : styloglossus force, f_{gh} : genio-hyoid force, f_{gi} : glosso-intrinsic forces, f_w : gravity; P_{rJT} : jaw-tongue reference point, $\{x_r, y_r\}$: sagittal plane reference; $\{X_a, Y_a\}$: accelerometer reference.

for the diadochokinetic utterances proposed in the experiments described in Sec. 3.1, a one-to-one relationship between the ONPT gesture positions and formants may be assumed (H1: Uniqueness).¹⁸ It will also be assumed that the matrix \mathbf{A} is invertible, thus, $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}$ (\mathbf{I} : identity matrix), therefore it may be said that $\mathbf{B} = \mathbf{A}^{-1}$. This assumption is substantiated by the estimation of the coefficients a_{ij} , which is based on regression methods, and on the following inequality: $a_{11}a_{22} - a_{12}a_{21} \neq 0$ (H2: Invertibility). The justification of H2 is given in Sec. 3.2 at the sight of the estimations for a_{ij} in the case under study. H3: Time invariance. This hypothesis is frequently found in speech processing problems, for instance, if spectra based on Fourier Transforms are to be estimated. Theoretically a Fourier Transform can only be defined on infinitely lasting signals (windows of infinite size). This requirement has no sense in practical cases. Therefore, a window W wide enough is defined on a signal to consider it quasi-stationary, in the sense that the moment generating function of the signal might be considered almost constant within W .^{31,32} Similarly, it will be assumed that both \mathbf{A} and \mathbf{B} will have almost time-invariant distributions during the estimation time window W . In the present study, the pseudo-stationary hypothesis is supported by the fact that the neuromechanical model of the ONPT will not change fast enough, during the diadochokinetic exercises barely producing diphthongs at a rate larger than 3–4 per second. Under assumptions H1 and H2, the system in (1) may be inverted as

$$\begin{bmatrix} \Delta x_r(t) \\ \Delta y_r(t) \end{bmatrix} = \mathbf{B} \begin{bmatrix} \Delta F_1(t) \\ \Delta F_2(t) \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}. \quad (2)$$

On its turn, under assumption H3,

$$\begin{bmatrix} a_{xr}(t) \\ a_{yr}(t) \end{bmatrix} = \mathbf{B} \begin{bmatrix} \frac{\partial^2 F_1(t)}{\partial t^2} \\ \frac{\partial^2 F_2(t)}{\partial t^2} \end{bmatrix}, \quad (3)$$

where a_{xr} and a_{yr} are the horizontal and vertical accelerations in the sagittal plane. Therefore

$$a_{yr}(t) = b_{21} \frac{\partial^2 F_1(t)}{\partial t^2} + b_{22} \frac{\partial^2 F_2(t)}{\partial t^2}. \quad (4)$$

Concerning the dynamic part of the model,³³ it will be assumed that the sEMG activity recorded on the

bulk of the masseter $s_m(t)$ is a correlate of the summation of different individual motor fibers activating the muscle,^{34,35} and therefore, the dynamic force exerted by the muscle in the upper vertical direction $f_m(t)$ may be expressed as the joint action of these individual actions on the muscle cell membranes, therefore

$$\begin{aligned} f_m(t) &= J_m r_m(t), \\ r_m(t) &= \int_{t \in W} |s_m(t)| dt, \\ J_m &= \frac{T_m}{l_m \cos \vartheta} \end{aligned} \quad (5)$$

where J_m is the myoelectric 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,³⁰ and $r_m(t)$ is the integral of the rectified sEMG on the masseter. The nature of the relation in (5) between force and the integral of the rectified sEMG is an important object of discussion. There are several important considerations in its current formulation (H4: relationship between motor unit actions and force).³⁶⁻⁴² The question of why force is related to the rectified sEMG, and not to sEMG, apparently has to see 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, as suggested by the experts.^{37,39} The force exerted by a muscle is a consequence of the neuromotor stimulation of the muscular fibers by impulse-like action potentials,⁴³ which accumulate on the muscular cell membranes as a *de facto* integration of the positive peaks of individual actions, therefore the muscular force appearing responds to the rectified integral of sEMG recordings ($r_m(t)$). On the other hand, if it is hypothesized that $r_m(t)$ as measured on the masseter is a strong correlate of the force exerted by the muscle, the vertical acceleration of the reference point P_{rJT} would be strongly related to $r_m(t)$. The biomechanical model would act as a low-pass filter, as velocities and displacements had to be estimated by integration, further smoothing the spike behavior of sEMG, to the point where high frequency components are not any more relevant, and only average oscillations are relevant to the study, provided that baseline noise and other artifacts are conveniently removed.³⁹

A further consideration has to see with attributing force only to masseter activity, having in mind that other elevator and depressor muscles are also active, as temporalis and medial pterygoid (elevators) or the geniohyoid, mylohyoid or digastric (depressors).⁴⁴ The influence of these muscles is only to be taken into account in certain jaw displacements during mastication,⁴⁵⁻⁵⁰ but not in the vertical small displacements generated by torques aligned to the normal of the sagittal plane as those appearing in speech articulation (H5: masseter dominance, meaning that the vertical displacement of the jaw and tongue resulting from the masseter activity is the main cause for low/high and high/low transitions in the first two formants, therefore horizontal displacements will be less relevant regarding formant kinematics). Another consideration to be discussed is the linear relationship assumed in (5). The nature of the relationship between integrated rectified sEMG and force in a muscle has received much attention in biomechanical studies. Most of them are devoted to main body muscles, as biceps, triceps, deltoid, biceps femoris, tibialis, etc.^{40,41} Much less attention has been devoted to facial muscles, as the masseter. Most of the works on this muscle are devoted to bite and mastication,⁴² and not to speech articulation. What can be concluded is that these studies point to the existence of a clear relationship between the integrated rectified sEMG and force in a given muscle, and that the following properties have been proven:

- The relationship between $r_m(t)$ and force is aligned and in the same sense (larger $r_m(t)$ activity is associated to larger muscle force) for all muscles studied.
- This relationship is monotonic (positive increments in $r_m(t)$ are associated with positive increments in muscle force) for all muscles studied.
- There is at least a region where the relationship may be considered linear or quasi-linear: between 0% and 30% of maximum muscle force. For larger forces, the relationship shows a nonlinear bending. This behavior seems to be associated to the type of fiber contraction speech articulation would correspond to the quasi-linear region.

Having in mind these considerations, the linear relationship was adopted in the sense that speech articulation may be considered a small force amplitude activity, at which the masseter will behave as a

linear actuator, an assumption which has to be supported by experimentation. It will be seen that the results produced in the present work support strongly this hypothesis.

The dynamic model would be a third-order lever, the pulling force of the masseter resulting in a quasi-harmonic movement of the P_{rJT} . Nevertheless, for small movements of the jaw, as those produced during normal speech, it may be assumed that the force exerted by the masseter will be related to the vertical acceleration of the point P_{rJT} as

$$a_{yr}(t) = \frac{\sum f_m}{M_m}, \quad (6)$$

where f_m will be the resultant of all force components acting vertically (excluding viscosity effects),^{30,48} and M_m may be seen as the inertial equivalent of jaw-tongue system.^{49,50} In what follows, it will be assumed that the only force acting vertically taken into account is f_m , the one exerted by the masseter. Considering only small amplitude oscillations, the force exerted by gravity (weight: f_w) may be removed from the model by unbiasing the acceleration components $\{X_a, Y_a\}$. This assumption is supported by the fact that any bias found in the linear vertical acceleration must be due to the contribution of gravity (assuming that the subject is not moving vertically). Any other constant bias in the masseter force would produce a displacement of jaw to the two extreme positions: yawn (downwards) or bite (upwards). This hypothesis will be plausible when the activity of the jaw is a harmonic oscillation, as in speech articulation, and specifically in diadochokinetic exercises, as the ones used in the experiment described in the sequel. Therefore, from (4) and (6)

$$f_m(t) = c_{21} \frac{\partial^2 F_1(t)}{\partial t^2} + c_{22} \frac{\partial^2 F_2(t)}{\partial t^2}, \quad (7)$$

$$c_{21} = M_m b_{21}; \quad c_{22} = M_m b_{22}.$$

The estimation of the corresponding scale weights c_{21} and c_{22} has to be carried on a time window W of adequate size to fulfill the conditions for pseudo-stationarity (time-invariance) requested by H3, as explained above.³² One possibility is to rely on second order moments on W as

$$R\{f_m, f_m\} = R\{f_m, F_1''\} + R\{f_m, F_2''\} \quad (8)$$

where F'' is the second derivative of F , and $\{\cdot, \cdot\}$ is the inner product in Hilbert's space

$$R\{f_m, f_m\} = \int_{t \in W} f_m^2(t) dt,$$

$$R\{f_m, F_1''\} = \int_{t \in W} f_m(t) \left[\frac{\partial^2 F_1(t)}{\partial t^2} \right] dt, \quad (9)$$

$$R\{f_m, F_2''\} = \int_{t \in W} f_m(t) \left[\frac{\partial^2 F_2(t)}{\partial t^2} \right] dt.$$

It is well known that the opening of the ONPT in its mid-front section is directly influencing the position of the first formant, and this situation is dominant over the front-back positions of the tongue in most sounds.^{25,51,52} Therefore, it will be assumed that for front vowels most of the neuromotor masseter force will produce vertical acceleration observed as a correlate on the first and second formant kinematics. The problem to be faced now is that the vertical force produced by the masseter is not accessible, and has to be estimated indirectly, having (5) into account. Therefore, assuming that the correlation between the integrated rectified sEMG given as $r_m(t)$ and the vertical acceleration $a_{yr}(t)$ is significant, the joint estimation of the scale weights c_{21} and c_{22} may be carried on by means of the following iteration:

$$\check{c}_{21}^k = \frac{R\{r_m, r_m\} - \gamma \check{c}_{22}^{k-1} R\{r_m, F_2''\}}{R\{r_m, F_1''\}}, \quad (10)$$

$$\check{c}_{22}^k = \frac{R\{r_m, f_m\} - \gamma \check{c}_{21}^{k-1} R\{r_m, F_1''\}}{R\{r_m, F_2''\}}$$

where $0 < \gamma < 1$ is a convergence parameter and k is the iteration index, with the following initial conditions:

$$\check{c}_{21}^0 = \frac{R\{r_m, r_m\}}{R\{r_m, F_1''\}}, \quad (11)$$

$$\check{c}_{22}^0 = \frac{R\{r_m, r_m\}}{R\{r_m, F_2''\}}.$$

The final estimation of c_{21} and c_{22} may be obtained by successive joint iterations of (10), having in mind that from (5)

$$c_{21} = J_m \check{c}_{21}; \quad c_{22} = J_m \check{c}_{22}. \quad (12)$$

Actual values of c_{21} and c_{22} will be estimated for the diadochokinetic exercise described in the next section.

3. Materials and Methods

3.1. Signal estimation and processing

The present study has a marked exploratory nature, as very few publications have addressed the use of sEMG signals measured on the masseter to estimate neuromotor decay in neurological disease evaluation.^{53,54} The main assumption is that the myoelectric signal recorded at the surface of the skin over the facial position of the masseter will represent the joint action of many individual muscle fiber contractions^{15,53,54} under the neuromotor commands conducted by the anterior mandibular division of the trigeminal nerve. The masseter is one of the most powerful muscles in the human body, therefore it is a good candidate for this kind of measurements. An experiment was designed to validate the working hypothesis on the relationship between the first formant dynamics and the electromyographic signal recorded on the masseter. The speaker was asked to produce the diadochokinetic sequence /*ayayayayay...*/ (phonetically [*ajajajajaj...*]), because it evokes an intensive masseter activity, and explores the low/high vowel domain.⁵⁵ The speech signal was recorded with a Sennheiser cardioid microphone at 44100Hz and 16 bits. The sEMG signal was recorded with the equipment Biopac MP150 EMG100 at 2kHz and 16 bits. The fixture to record surface myoelectric signals from the masseter and acceleration on the chin is shown in Fig. 3. Two surface contact electrodes are fixed on the skin at the masseter attachment to the mandible, and at the mid-superior part of the muscular bundle attachment to the zygomatic arch, and a third reference electrode is placed in the forefront over the ipsilateral eyebrow. This fixture produced an acceptable signal to noise ratio.³⁹ A 50-Hz IIR notch comb filter was applied on the recorded sEMG signal for electric power artifact removing.

The recording protocol and signal processing methodology is given in the following steps:

- Myoelectric surface signals (sEMG, see Fig. 4(a)) and voice signals (see Fig. 4(d)) are recorded in synchrony. The dynamic correlate of the sEMG signal is estimated by rectification and integration (see Fig. 4(b)). The accelerometer signals in axes $\{X_a, Y_a\}$ are referred to coordinates $\{x_r, y_r\}$ and integrated to estimate the vertical displacement



Fig. 3. Recording the myoelectric surface signal produced by the contraction of the masseter and the 2D acceleration in the sagittal plane.

Δy_r , which is shown in Fig. 4(c). The sEMG, initially sampled at 2kHz, is downsampled at 500 Hz, to match a formant estimation rate of 2 ms.

- An 18-cycle segment of the diphthong [*aj*] is selected. The speech recordings are undersampled to 8 kHz for formant kinematic estimation.⁵⁶
- The ONPT transfer function of the voice segment is evaluated by an eight-pole adaptive inverse LP filter⁵⁶ with a low-memory adaptive step to grasp fine time variations.
- Figure 4(d) shows the speech signal (light blue) and its energy envelope (red).
- The first two formants (F_1, F_2) are estimated by evaluating the roots of the associated inverse polynomials of the LP predictor each 2 ms (equivalent to a sampling frequency of 500 Hz) with a frequency resolution of 2 Hz. The first two formants are given in Fig. 4(e).

The upper template of Fig. 4(a) shows the activity of the sEMG as recorded on the masseter. The integral of its rectified value produces a correlate of the force exerted by the masseter,⁴² which is characterized by a sequence of pull-up actions followed by relaxing intervals, at a rate of approximately two per second. These actions provoke the raising P_{rJT} (Fig. 4(c)), moving the vowel sound from [*a:*] to [*i:*]. When the P_{rJT} is lowered (relaxed masseter), the amount of energy radiated through the lips is larger, because sound is radiated through a larger section (louder speech amplitude), whereas when the contracted masseter pulls up the jaw the radiated energy is smaller (smaller radiation section). It may be seen

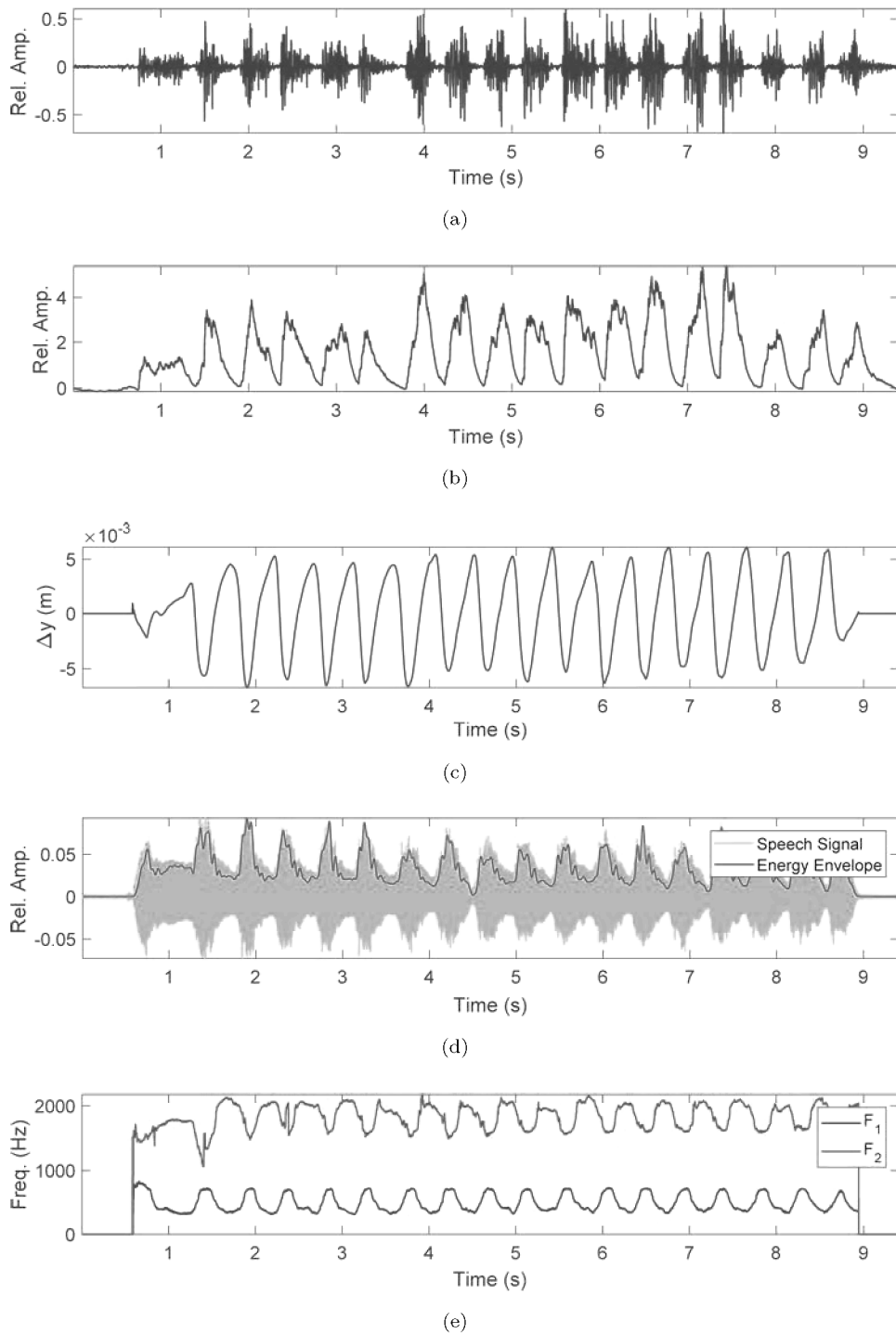


Fig. 4. (Color online) sEMG and audio signals recorded for the study, corresponding to the diadochokinetic exercise consisting in the repetition of $[aja]$ during 18 cycles: (a) Myoelectric surface signal (sEMG); (b) integrated rectified sEMG (correlate to the masseter force acting vertically); (c) vertical displacements of the reference point P_{rJT} ; (d) diadochokinetic speech signal, repetition of $[aj]$ (light blue), and its energy envelope (red thick line); and (e) first two formants (F_1 and F_2). The apparent smooth aspect of (c) is due to the effects of a double integration on accelerometry signals and not due to low-pass filtering.

that the energy envelope (red line) in Fig. 4(d)) is following these changes in the radiated loudness. In Fig. 4(e), the effect of a larger or smaller vocal tract end section (corresponding to the low or high vowel) on the first two formants is also seen, as the first formant climbs up for the low vowel [a:] and dives down for the high vowel [i:]. The second formant behaves inversely. When the masseter is relaxed the sound is closer to [a:] (higher F_1 positions, lower F_2 positions in (d)). When the masseter is contracted, the sound is closer to [i:] (lower F_1 positions, higher F_2 positions). It may be seen that F_1 in Fig. 4(e) is following Fig. 4(b) in counterphase, whereas F_2 is in phase. Therefore, a nonnegligible correlation might exist between the masseter dynamics (force) and the first two formant positions. Any correlation present among the three types of signals (speech, displacement and sEMG) may help in establishing quantitative relationships, as cause-effects are supported by dynamic and acoustic background. Quantitative correlations will be estimated in the next section.

3.2. Regression study

The physical assumptions supporting the existence of strong correlations among the three kinds of signals involved in the diadochokinetic exercise are the following:

- Articulation plays an amplitude modulation role in phonation, as shown in Fig. 4(d).
- The main cause is the jaw raising/lowering, due to masseter activations, which may be monitored from sEMG ($r_m(t)$), producing a vertical driving force $f_m(t)$.
- The vertical displacement around the reference point P_{rJT} will be given by

$$\begin{aligned}\Delta y_r(t) &= \int_{t \in W} v_{yr}(t) dt, \\ v_{yr}(t) &= \int_{t \in W} a_{yr}(t) dt,\end{aligned}\tag{13}$$

where $a_{yr}(t)$ is the vertical acceleration referred to P_{rJT} , obtained from the accelerometer estimates after proper rotation, and W is the time estimation window. The vertical displacement reduces the space between the lower jaw and the palate, modifying the ONPT resonances.

- The first two formant estimates $\{F_1(t), F_2(t)\}$ due to the ONPT resonances will follow the vertical displacement $\Delta y_r(t)$ in counterphase and in phase (respectively).

Therefore, the correlation study will first check the relationship between $f_m(t)$ (hypothetical force acting vertically) with the vertical displacement $\Delta y_r(t)$ referred to P_{rJT} and the evolution of F_1 and F_2 . As $f_m(t)$ is not accessible, $r_m(t)$ will be used instead. These correlations measured by Pearson's and Spearman's coefficients are given in Fig. 5. It may be seen in Fig. 5(a) that the linear regression between the force correlate oscillations ($\Delta r_m(t)$) and the vertical displacement Δy_r is positive and highly relevant (p -value < 0.001), with a Pearson's coefficient of 0.81 (p -value < 0.001), and a Spearman's rank-coefficient of 0.83 (p -value < 0.001). In Fig. 5(b), the regression between the vertical displacement Δy_r and the oscillations of the first formant ΔF_1 is plotted, showing that there is a strong correlation between both variables (-0.89), indicating that when the reference point moves upwards the first formant goes down. A similar relationship is found for the second formant, but the correlation is positive in this case. These correlations indicate that the displacements follow the sEMG correlate of the force exerted by the masseter, and the formants follow the displacements, but with different orientation. These regression results may be used to estimate a_{ij} . The results are $a_{11} = 561.35$, $a_{12} = -345.96$, $a_{21} = -753.84$, $a_{22} = 426.58$, all estimates given in cm s^{-1} . The reader may check that H2 is fulfilled in this case. Using the estimates provided by a similar regression study between the vertical acceleration $a_{yr}(t)$ and the masseter force, by iterating expressions (5), (6), (10)–(12) after 100 runs, the following estimates for the biomechanical parameters of the model may be derived: $J_m = -213.79 \text{ g cm s}^{-3}$, $c_{21} = -0.0273 \text{ g cm s}$ and $c_{22} = 0.0202 \text{ g cm s}$. The value for M_m has been taken as 440 g .⁵⁷

The possibility of producing these estimations is of high relevance for the diagnosis and monitoring of neurodegenerative diseases from the analysis of voice and speech. As the time derivatives of the formants can be associated with the kinematics of the P_{rJT} as given by (2), a possible correlate to rate neuromotor disease grade could be defined by the absolute kinematic velocity (AKV) of the jaw-tongue system,

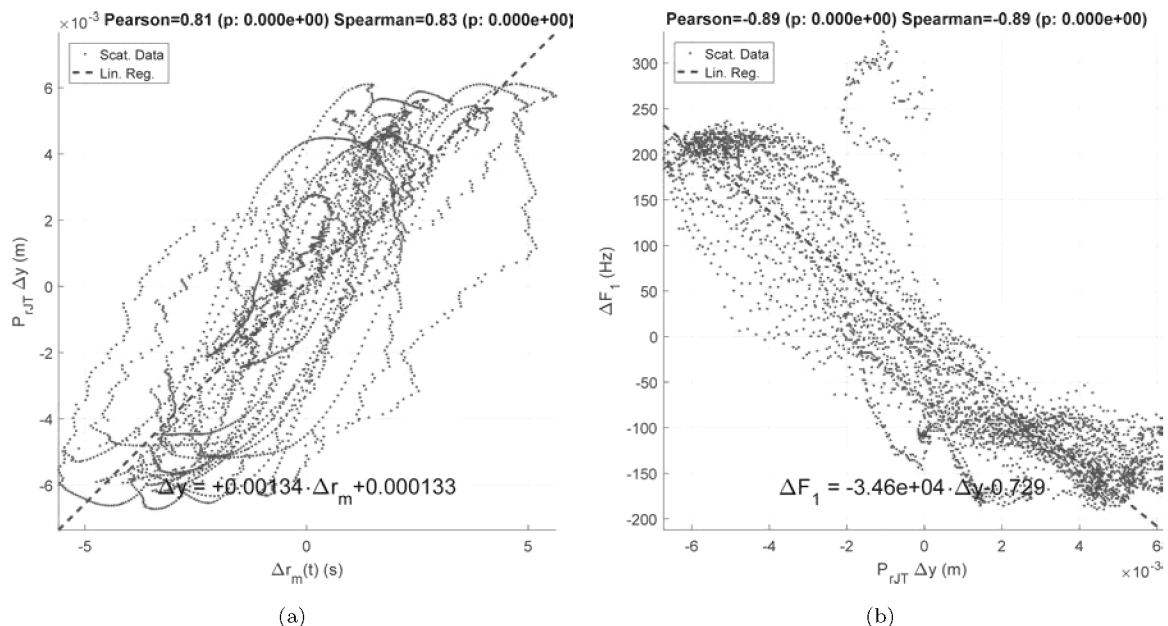


Fig. 5. (Color online) Regressions between the integral of the rectified sEMG, the vertical position of the reference point J_{rPT} , and the first formant oscillations. The scatter plot of the dataset is given in red dots. The dash lines give the linear regression fit. The regression line equation is printed within each template. (a) Regression between the integral of the rectified sEMG (correlate of the force exerted by the masseter) and the unbiased vertical displacement of the reference point. It may be seen that the correlation coefficient is reasonably high and relevant (p -value < 0.001). (b) Regression between the unbiased vertical displacement of the reference point and the unbiased first formant activity. It may be seen that the correlation coefficient is high (in counterphase) and relevant (p -value < 0.001).

given as

$$|v_r(t)| = \sqrt{B_1 \left[\frac{\partial F_1(t)}{\partial t} \right]^2 + B_2 \left[\frac{\partial F_2(t)}{\partial t} \right]^2 + B_{12} \frac{\partial F_1(t)}{\partial t} \frac{\partial F_2(t)}{\partial t}}, \quad (14)$$

where B_1, B_2 and B_{12} are quadratic forms of the coefficients of the inverse matrix $\mathbf{B} = \mathbf{A}^{-1}$, assuming that invertibility conditions are fulfilled. The estimations of a_{ij} indicate that this is a plausible hypothesis. The absolute velocity of the P_{rPT} given in (14) is a relevant correlate to articulation dysarthria and dysfluency, and can be used in comparing speech features from different utterances.^{58,59} In what follows the results of evaluating the degradation of the speaking performance in a set of 16 PD⁷ patients will be shown.

4. Application to PD Dysarthria

The database of normative and pathological speech used in PD monitoring is a part of the Parkinsonian Speech Database (PARCZ)⁶ recorded at St. Anne's

University Hospital in Brno, the Czech Republic. None of the patients showed a disease affecting the central nervous system other than PD. All patients were examined on their regular dopaminergic medication (ON state without dyskinesias). Four sets of five Czech vowels ([a:, e:, i:, o:, u:]) were recorded, uttered in four different ways: short and long vowels in a natural way, long vowels at maximum loudness, and long vowels at minimum loudness.⁶⁰ The recordings selected for this study corresponded to four subsets of speakers composed of eight normative females (NF; average age: 62.25 y; std age: 3.81 y), eight normative males (NM; av.: 63.63 y; std: 7.15 y), eight PD females (PF; av.: 69.25 y; std.: 7.11 y) and eight PD males (PM; av.: 64.88 y; std.: 8.51 y), see Table 1. Recordings of vowels [a: i: u:] at maximum loudness, sampled at 16 kHz and 16 bits were selected from the database to estimate the logarithm of the VSA (lnVSA) and the FCR.¹⁸ These are static features based on the estimation of the extreme positions and the center of gravity of the vowel triangle, which have been used classically as features of articulation disfunction.^{18,61} The VSA gives an estimation

Table 1. Subject set description, static and kinematic estimates. *Nxxxx*: normative subjects; *Pxxxx*: pathologic subjects. UPDRS refers to section III of that rating scale.⁶⁶

| Subject | Gender | Age | UPDRS | lnVSA | FCR | KLD |
|---------|--------|-----|-------|-------|------|--------|
| N1003 | F | 63 | — | 13.09 | 0.92 | 47.10 |
| N1004 | F | 65 | — | 12.72 | 0.99 | 29.89 |
| N1006 | F | 64 | — | 13.30 | 0.90 | 18.34 |
| N1007 | F | 59 | — | 13.40 | 0.81 | 38.99 |
| N1012 | F | 67 | — | 12.85 | 0.95 | 64.54 |
| N1017 | F | 61 | — | 13.30 | 0.89 | 35.53 |
| N1018 | F | 55 | — | 13.21 | 0.91 | 22.29 |
| N1019 | F | 64 | — | 13.17 | 0.84 | 25.30 |
| P1006 | F | 59 | 24 | 12.84 | 0.90 | 38.69 |
| P1007 | F | 76 | 55 | 12.85 | 0.90 | 76.31 |
| P1008 | F | 78 | 23 | 13.01 | 0.85 | 37.87 |
| P1020 | F | 64 | 8 | 12.82 | 1.03 | 100.14 |
| P1021 | F | 65 | 5 | 13.33 | 0.87 | 63.67 |
| P1022 | F | 72 | 6 | 12.96 | 0.99 | 67.75 |
| P1025 | F | 64 | 8 | 13.09 | 0.85 | 57.25 |
| P1026 | F | 76 | 12 | 13.00 | 0.93 | 44.42 |
| N2001 | M | 59 | — | 12.49 | 0.83 | 24.98 |
| N2002 | M | 68 | — | 12.60 | 0.95 | 52.77 |
| N2008 | M | 70 | — | 12.74 | 0.95 | 19.50 |
| N2009 | M | 68 | — | 12.48 | 0.93 | 40.06 |
| N2010 | M | 73 | — | 12.14 | 1.02 | 23.37 |
| N2011 | M | 55 | — | 12.62 | 0.89 | 34.80 |
| N2013 | M | 54 | — | 12.61 | 0.97 | 41.22 |
| N2014 | M | 62 | — | 12.04 | 1.04 | 15.88 |
| P2005 | M | 46 | 25 | 12.45 | 1.29 | 40.32 |
| P2009 | M | 66 | 14 | 12.46 | 0.92 | 53.92 |
| P2010 | M | 66 | 39 | 12.22 | 1.00 | 121.42 |
| P2012 | M | 71 | 35 | 12.14 | 1.03 | 62.15 |
| P2017 | M | 71 | 35 | 12.88 | 1.43 | 34.83 |
| P2018 | M | 63 | 19 | 12.08 | 1.03 | 55.47 |
| P2019 | M | 63 | 32 | 12.24 | 0.91 | 45.06 |
| P2023 | M | 73 | 12 | 12.14 | 1.00 | 77.03 |

of the area of the vowel triangle, and is larger in females than in males. Normative values are 12.7091 (females) and 12.1843 (males).

The FCR shows the tendency of formants to shrink towards the center of the vowel triangle with a slight skew to vowel [æ]. The normative value for the FCR is around 0.97 for both genders. In general, it may be said that the smaller the lnVSA and the larger the FCR, the more pathological the voice under analysis will be. The values of VSA and FCR for the normative female subject shown as an example in Fig. 6 are given in the lower templates. See that the female subject lnVSA is 12.85 (slightly over the female normative) and FCR is 0.949

(slightly under the normative). FCR and lnVSA are estimated from the vowel triangle extreme positions, representing articulation points of vowels [a: i: u:] marked with circles in the lower templates of Fig. 6, in which an example from a normative female speaker is shown based on the evaluation of the first two formants extracted from Linear-Prediction Spectral Estimation.⁵⁶ An example of the normalized histogram and cumulative distribution of the AKV for the sequence in Fig. 6 is given in Fig. 7.

It may be seen that the most active events (larger AKV) are aligned with vowel insertions (start of phonation around 1.1s and 2.3s) or during weak phonation (between 0.7 and 1.0s). The AKV distribution shows a χ^2 behavior (two degrees of freedom). Its similarity to a Maxwell-Boltzmann distribution allows establishing a parallelism with thermodynamic concepts, giving sense to the term “emotional temperature” used by some researchers in the field of neurological deterioration, as in Alzheimer’s Disease speech studies.⁶² The normalized histograms may be interpreted as probability distributions, and these can be applied to estimate the difference in terms of Information Theory^{63–65} between two probability distributions in terms of Kullback-Leibler’s Divergence (KLD) as

$$D_{KLij} \{p_i(|v_{P,rJT}|), p_j(|v_{P,rJT}|)\} = - \int_{\zeta=0}^{\infty} p_i(\zeta) \log \left[\frac{p_i(\zeta)}{p_j(\zeta)} \right] d\zeta. \quad (15)$$

KLD is an estimator of Mutual Information between two stochastic processes with probability distributions given by $p_i(|v_r|)$ and $p_j(|v_r|)$, obtained from two subjects i and j , if their articulation profiles (similar to the one in Fig. 7(a)) are given as histograms in amplitude bins (velocities) as the one in Fig. 7(b) (in blue) which is normalized to the total number of counts. Expression (15) consists in evaluating the logarithm of the ratio between both distributions at each bin, in our case, limited to 50 cm s^{-1} , and have it weighted by one of the probabilities, in this case p_i . If both probability functions are equal, the log of their ratio will be zero, and the divergence D_{KLij} will be null, i.e. there will not be divergence. In any other case, as they are distributions following χ^2 , they will present a lump, more or less shifted to the right. Speakers with a highly unstable articulation will have the lump more to the right than speakers

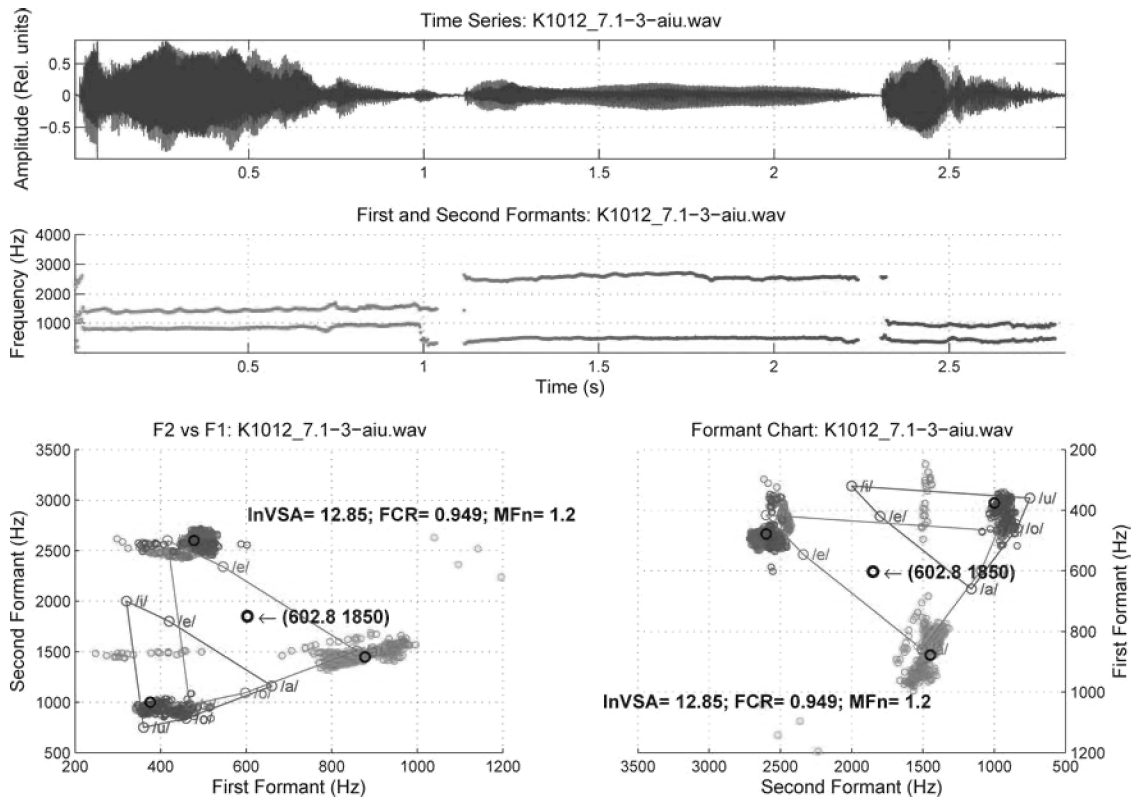


Fig. 6. Example of the first two formants extraction from a sequence [a:, i:, u:] uttered by a normative female speaker. Top: speech signal. Middle: first two formants from LP spectral estimation. Bottom left and right: formant projection on the vowel triangle. Black circles give the vowel centroids and the vowel triangle centre of gravity to evaluate the InVSA and the FCR, which are shown superimposed on the vowel triangle.

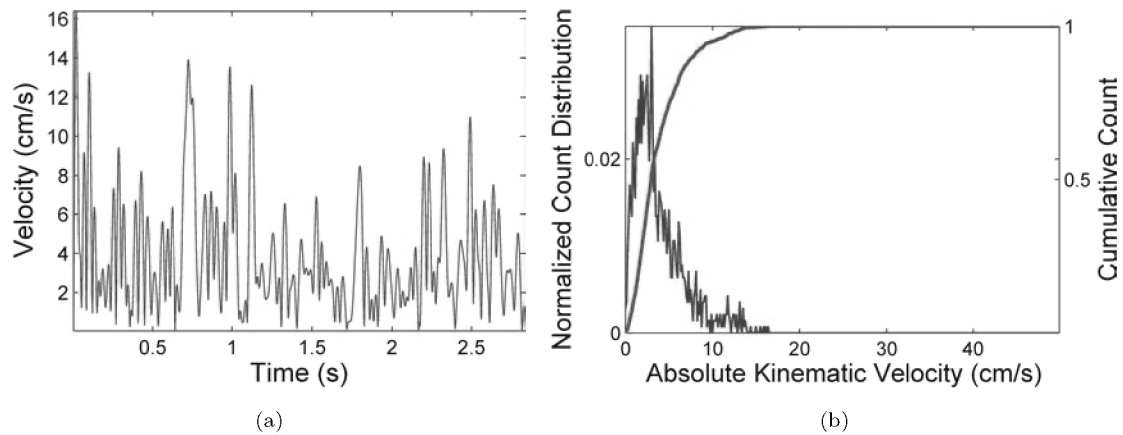


Fig. 7. AKV for the sequence shown in Fig. 4. Top: time series. Bottom: normalized histogram from 0 to 50 cm s⁻¹ (thin line), and its respective cumulative distribution (thick line).

showing a more stable articulation. If p_i represents a normative speaker, and p_j a pathologic one, it will be expected that $\log(p_i/p_j)$ will show positive values for low absolute velocities, and negative ones for high absolute velocities. As the weighting function is

also p_i , it will give more relevance to low absolute velocities, therefore the resultant divergence will be positive. As it is known, normative subjects can produce more stable vowels, with F_1 and F_2 showing little oscillations, therefore p_i will show a left-hand

histogram. PD patients on the contrary show unstable F_1 and F_2 , and their histogram will tend to show a lump to the right. In the present study, the AKV estimates and their normalized histograms by velocity bins between 0 and 50 cm s^{-1} were evaluated from (14). Four sets of normalized histograms were produced respectively for the normative male set NM: $\{p_{NM}\}$; normative female set NF $\{p_{NF}\}$; pathological male set PM: $\{p_{PM}\}$; and pathological female set PF: $\{p_{PF}\}$. The D_{KL} between each subject in the pathological sets PM and PN was estimated with respect to the averages of their respective normative sets, NM and NF.

5. Results and Discussion

The purpose of the present experiment is to show that D_{KL} as derived from (15) is a more relevant feature than lnVSA or FCR to distinguish unstable vowel articulation by PD patients from stable articulation by normative speakers. The results of evaluating lnVSA, FCR (static features) and KLD (dynamic features) from 16 normative subjects and 16 PD patients randomly selected from the PARCZ database are given in Table 1. Male (M) and female (F) subjects have been included in equal numbers (balanced sets).

This experiment has to be seen as a comparison between the discrimination capability of the KLD divergence of AKV pdf as defined in (14) and (15) and lnVSA and FCR defined by Sapir, in distinguishing normative from dysarthric subjects. The comparisons are single-feature detections, where the two-class classifier is a two-tail t -test, which decides if the distribution of the subsets being compared cannot be differentiated based on their means, assuming different variances, under a p -value lower than 0.05 (null hypothesis H_0). Test results are shown in Table 2.

Table 2. T-tests on the results between normative and pathologic sets.

| Feature/subset | p -value | Null hypothesis (H_0) |
|----------------|--------------|---------------------------|
| lnVSA/females | 0.252 | No reject |
| lnVSA/males | 0.451 | No reject |
| FCR/females | 0.885 | No reject |
| FCR/males | 0.495 | No reject |
| KLD/females | 0.016 | Reject |
| KLD/males | 0.020 | Reject |

Table 2 shows that t -tests using lnVSA as the distinguishing feature fail in rejecting the null hypothesis (H_0 : that the pathological and normative sets have the same means) for a p -value under 0.05, both in the case of males and in the case of females. FCR features fail also in rejecting H_0 for $p < 0.05$. On the contrary, when using D_{KL} as the distinguishing feature, the tests reject H_0 , pointing to the possibility that pathological and normative sets have different means with statistical relevance under $p < 0.05$. Therefore there are significant differences if D_{KL} is used as a separating feature, with respect to lnVSA or FCR, showing the superior performance of dynamic (D_{KL}) versus static features (lnVSA or FCR), as it was intended.

To clarify the relevance of this finding, a comparison to a relevant study by Skodda *et al.*⁶⁷ is commented. This study analyzed the capability of triangular VSA and the Vowel Articulation Index (reciprocal of FCR) to differentiate normative and HD speech as single features. On a database of 68 PD patients (gender balanced) uttering [a: i: u:] under highly controlled conditions compared to 32 controls (gender balanced, matched in age) significant differentiation between PD and control speech was observed using variance analysis on several features as mean and standard deviation of fundamental frequency, net speech rate, triangular VSA, and VAI. The results of the tests on VSA and VAI of 34 PD versus 16 controls produced p -values of 0.023 and 0.0003 respectively for a male set, and above 0.05 (nonsignificant) and 0.002 for the female set. Apparently VSA and VAI succeeded in differentiating PD from controls in males, and only VAI succeeded in doing so in females. In our study, neither VSA nor FDR succeeded in differentiating PD from controls, but the KLD of kinematic velocity succeeded in differentiating pathological speech from normative for both males and females. This disagreement may be based on the highly controlled vowel production and larger PD data sizes in Skodda's work, and on the lower number of patients involved in the present study (16 patients versus 16 controls). In the PARCZ database, vowel utterances were supposed to be as spontaneous as possible in both PD and controls, in phonation style and in duration, and were only grouped by intensity level. These preliminary results would have to be confirmed by an extended database under different recording conditions.

Besides the potential of AKV as a biomarker of PD, this acoustic measure could be used to monitor the effects of treatments (e.g. repetitive transcranial magnetic stimulation — rTMS, deep brain stimulation — DBS) or to adjust its settings.⁶⁸ It has been reported that especially DBS surgery can have a negative effect on speech. This is usually caused by the fact that DBS parameters, the exact targets for intracranial electrodes and used contacts are tailored based on effects of DBS on limbs and not on axial motor symptoms such as Hypokinetic Dysarthria (HD).^{7,68} Therefore, AKV could be used as a very quick and objective feedback quantifying a change in articulation and in intelligibility. It can be easily estimated during electrode navigation or during parameter adjustment so that DBS will not have such negative effects in specific HD behavior. Moreover, articulation kinematics may be used combined with clinical data to enhance computer-based monitoring systems in PD by neural-network-based classifiers.^{59,69}

6. Conclusions

Speech articulation is conditioned by the movement produced by specific groups of muscles in the larynx, pharynx, mouth and face. As the recording of speech is simple and ubiquitous, the use of speech as a vehicular tool for neuromotor action monitoring would open a wide set of applications in the study of functional grading of neurodegenerative diseases. This study has been conceived to establish the relationship between neuromotor activity (represented by the sEMG) and speech correlates (represented by formant kinematics as a functional model. This purpose has been fulfilled using electromyographic recordings on the masseter and jaw accelerometry on the chin to estimate muscular dynamics, and the acoustic kinematic features related with the first two formants. Although the study presented is limited to one case and one sequence, there are interesting findings to be reported, among them are the following:

- A strong correlation between electromyographic dynamics and acoustic kinematics was observed. Correlation measures based on second-order statistics satisfy the necessary conditions for this relation to be confirmed.
- This correlation could be formalized as explicit relationships among the integral of the rectified

sEMG, the vertical position of the reference point J_{IPT} , and the first two formant oscillations. The parameters sustaining these relationships were estimated using regression analysis. Apparently linear regression worked well because the movements and forces considered in the study were limited to speech articulation dynamics, which does not involve such strong forces as in biting or mastication. Under these conditions the relation between the integrated rectified sEMG and force may be considered linear.

- A significant correlation is observed between the integrated rectified sEMG signal and the first two formant displacements. Accepting that formants are acoustic marks related with the position of the jaw-tongue biomechanical system, this fact implies that the sEMG activity would be a correlate of the masseter neuromotor fiber firing activity, and the integrated rectified sEMG activity would be a correlate of the force produced by the masseter.
- These correlations support the use of the AKV probability density as a biomarker in monitoring different neurodegenerative processes, such as PD. This biomarker seems to be a good index to disfluency and dysarthria of neuromotor origin, complementing static features as VSA and FCR with a powerful description of articulation dynamics.
- The influence of the horizontal dynamics of the jaw-tongue system^{70,71} in conforming the kinematics of the first two formants, neglected in (7) might play a nonnegligible role which should be taken into account in future studies.
- The formulation of speech dynamics in terms of probability density functions of the kinematic variables allow the use of Information Theory principles to differentiate between dysarthric and normative speech. One of the inconveniences of KLD is its asymmetry. For this reason, other similar metrics are sought with a more balanced behavior.
- These conclusions are to be validated on a wider database of speakers, including gender dependence.

Acknowledgments

This work is being funded by Grants TEC2016-77791-C4-4-R from the Ministry of Economic Affairs and Competitiveness of Spain, Teka-Park 55_02 CENIE-0348_CIE_6_E POCTEP (InterReg

Programme) and 16-30805A, SIX Research Center (CZ.1.05/2.1.00/03.0072), and LOI401 from the Czech Republic Government.

References

1. U. Jürgens, Neural pathways underlying vocal control, *Neurosci. Biobehav. Rev.* **26** (2002) 235–258.
2. J. F. Demonet, G. Thierry and D. Cardebat, Renewal of the neurophysiology of language: Functional neuroimaging, *Physiol. Rev.* **85** (2005) 49–95.
3. Y. Yunusova, G. Weismer, J. R. Westbury and M. J. Lindstrom, Articulatory movements during vowels in speakers with dysarthria and healthy controls, *J. Speech Lang. Hear. Res.* **51** (2008) 596–611.
4. J. R. Green, Mouth matters: Scientific and clinical applications of speech movement analysis, *Perspect. Speech Sci. Orofac. Disord.* **25** (2015) 6–16.
5. O. Darbin, X. Jin, C. von Wrangel, K. Schwabe, A. Nambu, D. K. Naritoku, J. K. Krauss and M. Alam, Relationship between entropy of spike timing and firing rate in entopeduncular nucleus neurons in anesthetized rats: Function of the nigro-striatal pathway, *Int. J. Neural Syst.* **26**(2) (2016) 1550038.
6. J. Mekyska, E. Janusova, P. Gómez, Z. Smekal, I. Rektorova, I. Eliasova, M. Kostalova, M. Mrackova, J. B. Alonso, M. Faúndez and K. L. de Ipiña, Robust and complex approach of pathological speech signal analysis, *Neurocomputing* **167** (2015) 94–111.
7. L. Brabenec, J. Mekyska, Z. Galaz and I. Rektorova, Speech disorders in Parkinson's disease: Early diagnostics and effects of medication and brain stimulation, *J. Neural Transm.* **124**(3) (2017) 303–334.
8. S. Skodda, W. Grönheit and U. Schlegel, Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease, *PLoS ONE* **7**(2) (2012) e32132.
9. Y. Yunusova, J. R. Green and M. J. Lindstrom, Speech in ALS: Longitudinal changes in lips and jaw movements and vowel acoustics, *J. Med. Speech Lang. Pathol.* **21**(1) (2013) 1–13.
10. Y. Yunusova, J. R. Green, L. Greenwood, J. Wang, G. L. Pattee and L. Zinman, Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis, *Folia Phon. Logop.* **64**(2) (2012) 94–102.
11. Y. Yunusova, J. R. Green, M. J. Lindstrom, L. J. Ball and G. L. Pattee, Kinematics of disease progression in bulbar ALS, *J. Commun. Disord.* **43**(1) (2010) 6–20.
12. Y. Yunusova, G. G. Weismer and M. J. Lindstrom, Classifications of vocalic segments from articulatory kinematics: Healthy controls and speakers with dysarthria, *J. Speech Lang. Hear. Res.* **54** (2011) 1302–1311.
13. R. Yuvaraj, M. Murugappan, U. Rajendra, H. Adeli, N. M. Ibrahim and E. Mesquita, Brain functional connectivity patterns for emotional state classification in Parkinson's disease patients without dementia, *Behav. Brain Res.* **298** (2016) 248–260.
14. J. R. Cram, The history of surface electromyography, *Appl. Psychophys. Biofeedback* **28**(2) (2003) 81–91.
15. A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard and Y. Laurillau, EMG feature evaluation for improving myoelectric pattern recognition robustness, *Expert Syst. Appl.* **40**(12) (2013) 4832–4840.
16. X. Huang, A. Acero and H. W. Hon, *Spoken Language Processing* (Prentice-Hall, Englewood Cliffs, NJ, 2001), pp. 36–68.
17. G. Fant, *Acoustic Theory of Speech Production* (Mouton, Hague, 1960).
18. S. Sapir, L. O. Ramig, J. L. Spielman and C. Fox, Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech, *J. Speech Lang. Hear. Res.* **53**(1) (2010) 114–125.
19. C. Dromey, G. O. Jang and K. Hollis, Assessing correlations between lingual movements and formants, *Speech Commun.* **55**(2) (2013) 315–328.
20. R. Carré, From an acoustic tube to speech production, *Speech Commun.* **42**(2) (2004) 227–240.
21. A. Mainka, I. Platzek, W. Mattheus, M. Fleischer, A. S. Müller and D. Mürbe, Three-dimensional vocal tract morphology based on multiple magnetic resonance images is highly reproducible during sustained phonation, *J. Voice* **31**(4) (2017) 504.e11–504.e20.
22. J. Vaissière, Area functions and articulatory modeling as a tool for investigating the articulatory, acoustic, and perceptual properties of sounds across languages, in *Experimental Approaches to Phonology*, eds. M. J. Solé, P. S. Beddor and M. Ohala (Oxford University Press, Oxford, 2007), pp. 54–71.
23. K. Richmond, Estimating articulatory parameters from the acoustic speech signal, Ph. D. thesis, University of Edinburgh (2001).
24. S. Fuchs and P. Perrier, On the complex nature of speech kinematics, Universitätsbibliothek Johann Christian Senckenberg, *ZAS Papers Linguist.* **42** (2005) 137–165.
25. B. J. Kröger and P. Birkholz, A gesture-based concept for speech movement control in articulatory speech synthesis, in *Verbal and Nonverbal Communication Behaviours*, Lecture Notes on Artificial Intelligence, Vol. 4775 (Springer, Berlin, 2007), pp. 174–189.
26. J. R. Green, C. A. Moore and J. Kevin, The sequential development of jaw and lip control for speech, *J. Speech Lang. Hear. Res.* **45**(1) (2002) 66–79.
27. C. C. Peck and A. G. Hannam, Human jaw and muscle modelling, *Arch. Oral Biol.* **52**(4) (2007) 300–304.
28. S. Takahasi, G. Kuribayashi, T. Ono, Y. Ishiwata and T. Kuroda, Modulation of masticatory muscle activity by tongue position, *Angle Orthod.* **75**(1) (2005) 35–39.
29. S. Bauchillard, P. Perrier and Y. Payan, A biomechanical model of cardinal vowel production: Muscle

- activations and the impact of gravity on tongue positioning, *J. Acoust. Soc. Am.* **126**(4) (2009) 2033–2051.
30. N. Hogan, Adaptive control of mechanical impedance by coactivation of antagonist muscles, *IEEE Trans. Autom. Control* **29**(8) (1984) 681–690.
 31. S. Tasko and M. D. McClean, Variations in articulatory movement with changes in speech task, *J. Speech Lang. Hear. Res.* **47**(1) (2004) 85–100.
 32. J. Blanchet, P. Glynn and S. Zheng, Empirical analysis of a stochastic approach for computing quasi-stationary distributions, in *EVOLVE — A Bridge between Probability Set Oriented Numerics, and Evolutionary Computation II* (Springer, Berlin, 2013), pp. 19–37.
 33. A. G. Hannam, I. Stavness, J. E. Lloyd and S. Fels, A dynamic model of jaw and hyoid biomechanics during chewing, *J. Biomech.* **41** (2008) 1069–1076.
 34. K. S. Lee, Prediction of acoustic feature parameters using myoelectric signals, *IEEE Trans. Biomed. Eng.* **57**(7) (2010) 1587–1595.
 35. R. M. Roark, J. C. L. Li, S. D. Schaefer, A. Adam and C. J. De Luca, Multiple motor unit recordings of laryngeal muscles: The technique of vector laryngeal electromyography, *Laryngoscope* **112**(2) (2002) 2196–2203.
 36. L. J. Myers, M. Lowery, M. O’Malley, C. L. Vaughan, C. Heneghan, A. S. C. Gibson, Y. X. R. Harley and R. Sreenivasan, Rectification and non-linear pre-processing of EMG signals for cortico-muscular analysis, *J. Neurosci. Methods* **124** (2003) 157–165.
 37. C. J. De Luca, The use of surface electromyography in biomechanics, *J. Appl. Biomech.* **13**(2) (1997) 135–163.
 38. Y. Yang, T. Solís-Escalante, M. van de Ruit, F. C. T. van der Helm and A. Schouten, Nonlinear coupling between cortical oscillations and muscle activity during isotonic wrist flexion, *Front. Comput. Neurosci.* **10**(126) (2016), doi:10.3389/fncom.2016.00126.
 39. C. J. De Luca, L. D. Gilmore, M. Kuznetsov and S. H. Roy, Filtering the surface EMG signal: Movement artifact and baseline noise contamination, *J. Biomech.* **43** (2010) 1573–1579.
 40. J. H. Lawrence and C. J. De Luca, Myoelectric signal versus force relationship in different human muscles, *J. Appl. Physiol.* **54**(6) (1983) 1653–1659.
 41. C. Disselhorst-Klug, T. Schmitz-Rode and G. Rau, Surface electromyography and muscle force: Limits in sEMG-force relationship and new approaches for applications, *Clin. Biomech.* **24** (2009) 225–235.
 42. J. J. Mao, P. W. Major and J. W. Osborn, Coupling electrical and mechanical outputs of human jaw muscles undertaking multi-directional bite-force tasks, *Arch. Oral Biol.* **41**(12) (1996) 1141–1147.
 43. R. M. Enoka, and K. G. Pearson, The motor unit and muscle action, in *Principles of Neural Science*, eds. E. R. Kandel *et al.* (McGraw-Hill, 2013), pp. 768–789.
 44. J. H. Koolstra, Dynamics of the human masticatory system, *Crit. Rev. Oral Biol. Med.* **13**(4) (2002) 366–376.
 45. D. J. Ostry and K. G. Munhall, Control of jaw orientation and position in mastication and speech, *J. Neurophysiol.* **71**(4) (1994) 1528–1545.
 46. D. J. Ostry, E. Vatikiotis-Bateson and P. L. Gribble, An examination of the degrees of freedom of human jaw motion in speech and mastication, *J. Speech Lang. Hear. Res.* **40**(6) (1997) 1341–1351.
 47. E. Vatikiotis-Bateson and D. J. Ostry, An analysis of the dimensionality of jaw motion in speech, *J. Phon.* **23** (1995) 101–117.
 48. G. E. J. Langenbach and A. G. Hannam, The role of passive muscle tensions in a three-dimensional dynamic model of the human jaw, *Arch. Oral Biol.* **44**(7) (1999) 557–573.
 49. O. Röhrle and A. J. Pullan, Three-dimensional finite element modelling of muscle forces during mastication, *J. Biomech.* **40**(15) (2007) 3363–3372.
 50. F. Zhang, C. C. Peck and A. G. Hannam, Mass properties of the human mandible, *J. Biomech.* **35**(7) (2002) 975–978.
 51. M. Hasegawa-Johnson, S. Pizza, A. Alwan, J. S. Cha and K. Haker, Vowel category dependence of the relationship between palate height, tongue height, and oral area, *J. Speech Lang. Hear. Res.* **46**(3) (2003) 738–753.
 52. K. D. Neely, K. Bunton and B. H. Story, A modeling study of the effects of vocal tract movement duration and magnitude on the F2 trajectory in CV words, *J. Speech Lang. Hear. Res.* **59**(6) (2016) 1327–1334.
 53. A. Phinyomark, P. Phukpattaranont and C. Limsakul, Feature reduction and selection for EMG signal classification, *Expert Syst. Appl.* **39**(8) (2012) 7420–7431.
 54. S. P. Arjunan and D. K. Kumar, Recognition of facial movements and hand gestures using surface Electromyogram (sEMG) for HCI based applications, *9th Biennial Conf. of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications* (Glenelg, Australia, 2007), pp. 1–6.
 55. S. M. Tasko and K. Greilick, Acoustic and articulatory features for dyphthong production: A speech clarity study, *J. Speech Lang. Hear. Res.* **53**(1) (2010) 84–99.
 56. J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals* (Macmillan, New York, 1993).
 57. J. H. Koolstra and T. M. G. J. van Eijden, Prediction of volumetric strain in the human temporomandibular joint cartilage during jaw movement, *J. Anat.* **209** (2006) 369–380.
 58. P. Gómez, D. Palacios, A. Gómez, J. M. Ferrández, A. Álvarez, R. Martínez and V. Nieto, *Relating Facial Myoelectric Activity to Speech Formants*, Lecture Notes on Computer Science, Vol. 10338 (Springer, Berlin, 2017), pp. 520–530.

59. P. Gómez, J. Mekyska, J. M. Ferrández, D. Palacios, A. Gómez, V. Rodellar, Z. Galaz, Z. Smekal, I. Eliasova, M. Kostalova and I. Rektorova, Parkinson disease detection from speech articulation neuromechanics, *Front. Neuroinform.* **11** (2017) 1–17.
60. K. Tjaden, J. Lam and G. Wilding, Vowel acoustics in Parkinson’s disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions, *J. Speech Lang. Hear. Res.* **56**(5) (2013) 1485–1502.
61. J. A. Whitfield and A. M. Goberman, Articulatory-acoustic vowel space: Application to clear speech in individuals with Parkinson’s disease, *J. Commun. Disord.* **51** (2014) 19–28.
62. K. Lopez-de-Ipiña, J. B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C. M. Travieso, M. Ecay-Torres, P. Martínez-Lage and H. Eguiraun, On automatic diagnosis of Alzheimer’s disease based on spontaneous speech analysis and emotional temperature, *Cogn. Comput.* **7**(1) (2015) 44–55.
63. A. R. Webb, *Statistical Pattern Recognition* (Wiley, Chichester, England, 2002).
64. M. Salicrú, D. Morales, M. L. Menéndez and L. Pardo, On the applications of divergence type measures in testing statistical hypotheses, *J. Multivar. Anal.* **51**(2) (1994) 372–391.
65. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 2006).
66. C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, N. Giladi, R. G. Holloway, C. G. Moore, G. K. Wenning, M. D. Yahr and L. Seidl, Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations The Movement Disorder Society Task Force on rating scales for Parkinson’s disease, *Movement Disorders* **19**(9) (2004) 1020–1028.
67. S. Skodda, W. Visser and U. Schlegel, Vowel articulation in Parkinson’s disease, *J. Voice* **25** (2011) 467–472.
68. S. Skodda, Effect of deep brain stimulation on speech performance in Parkinson’s disease, *Parkinson’s Dis.* **2012** Article ID 850596, 10 pp, doi:10.1155/2012/850596.
69. T. J. Hirschauer, H. Adeli and J. A. Buford, Computer-aided diagnosis of Parkinson’s disease using enhanced probabilistic neural network, *J. Med. Syst.* **39** (2015) 179.
70. J. M. Gérard, R. Wilhelms-Tricarico, P. Perrier and Y. Payan, A 3D dynamical biomechanical tongue model to study speech motor control, *Res. Dev. Biomech.* **1** (2003) 49–64.
71. V. Sanguineti, R. Laboisiere and Y. Payan, A control model of human tongue movements in speech, *Biol. Cybern.* **77** (1997) 11–22.