Monitoring Parkinson Disease from speech articulation kinematics

Pedro Gómez[1], Jiri Mekyska[2], Andrés Gómez[1], Daniel Palacios[1],
Victoria Rodellar[1], Agustín Álvarez[1]

[1] Neuromorphic Processing Laboratory (NeuVox Lab)
Center for Biomedical Technology, Universidad Politécnica de Madrid
Campus de Montegancedo, s/n, 28223, Pozuelo de Alarcón, Madrid, Spain
[2] Department of Telecommunications, Brno University of Technology
Technicka 10, 61600 Brno, Czech Republic
pedro@fi.upm.es ORCID: http://orcid.org/0000-0003-3283-378X
j.mekyska@gmail.com ORCID: http://orcid.org/0000-0002-6195-193X
andres.gomez@ctb.upm.es ORCID: http://orcid.org/0000-0001-8643-9871
daniel.palacios.alonso@alumnos.upm.es ORCID: http://orcid.org/0000-0001-6063-4898
mariavictoria.rodellar@upm.es ORCID: http://orcid.org/0000-0001-9384-3290
agustin.alvarez@upm.es ORCID: http://orcid.org/0000-0002-3387-6709

ABSTRACT: Parkinson Disease (PD) is a neuromotor illness affecting general movements of different muscles, those implied in speech production being among them. The relevance of speech in monitoring illness progression has been documented in these last two decades. Most of the studies have concentrated in dysarthria and dysphonia induced by the syndrome. The present work is devoted to explore how PD affects the dynamic behavior of the speech neuromotor biomechanics (neuromechanics) involved in deficient articulation (dysarthria), in contrast to classical measurements based on static features as extreme and central vowel triangle positions. A statistical distribution of the kinematic velocity of the lower jaw and tongue is introduced, which presents interesting properties regarding pattern recognition and classification. This function may be used to establish distances between different articulation profiles in terms of information theory. Results show that these distances are correlated with a set of tests currently used by neurologists in PD progress evaluation, and could be used in elaborating new speech testing protocols.

Keywords: neuromotor deterioration; Parkinson Disease speech; hypokinetic dysarthria; speech articulation modeling; speech neuromechanics.

RESUMEN: *Monitorización de la enfermedad de Parkinson a partir de la cinemática de la articulación del habla.* – La enfermedad de Parkinson (EP) es un trastorno del sistema neuromotor que afecta a todo tipo de movimientos regulados por el sistema muscular humano, entre ellos los que controlan la producción del habla. La importancia del habla en la monitorización del progreso de la EP se ha estudiado ampliamente durante las últimas dos décadas. La mayoría de los estudios se han concentrado en describir y modelar el comportamiento de la fonación (disfonía) y de la articulación (disartria), en relación con el síndrome de la EP. El presente trabajo se centra en explorar la forma en la que la EP afecta al comportamiento dinámico de la biomecánica neuromotora relacionada con la disartria típica mostrada por este tipo de pacientes, a diferencia de las mediciones clásicamente utilizadas hasta el momento, que se basan en las posiciones extremas y centrales del triángulo vocálico. Se propone, para ello, una nueva medida basada en la distribución estadística de la cinemática del sistema mandibular y lingual, que presenta interesantes propiedades de cara al reconocimiento de patrones utilizado en la clasificación de los rasgos disártricos del paciente. Esta medida puede emplearse para establecer las distancias entre las distintas articulaciones en términos de la teoría de información. Los resultados del estudio presentado muestran que dichas distancias se hallan sustancialmente correlacionadas con ciertas pruebas utilizadas habitualmente por los neurólogos para evaluar el progreso de la EP. La distancia cinemática propuesta puede servir para elaborar nuevos protocolos de realización de pruebas para el seguimiento de la EP.

Palabras clave: deterioro neuromotor; el habla del paciente de Parkinson; disartria hipocinética; modelado de la articulación del habla; neuromecánica del habla.

1. INTRODUCTION

The neuromotor disorder known as *Parkinson Disease* is a sickness produced by a deficit of the neurotransmitter dopamine in basal ganglia, resulting in hampered neuromotor activity during an early phase, and leading to a movement impairing disease accompanied by cognitive disability at later stages. As it affects most neuromotor paths, it also interferes with speech capability in different ways, which have been extensively well documented (Mekyska et al., 2015). Rough and asthenic phonation, monotonicity, mono-loudness, freezing, velo-pharyngeal incompetence, and low tone are some of the observed alterations of speech coined with the term *hypokinetic dysarthria* (Sapir, 2014). Historically, a pioneering study on PD phonation correlates with clinical purpose using distortion features as jitter, shimmer, and noise-harmonic ratio, on sustained vowels, was due to Gamboa et al. (1997), who monitored the effects of dopaminergic drugs using the Computer Speech Lab (CSL, 2017). Articulation, on its turn, has been studied using features as the Vowel Space Area (VSA) or Formant Centralization Ratio (FCR; Sapir et al., 2010). Illness progress is evaluated by neurologists using scales as Hoehn & Yahr (Hoehn & Yahr, 1967) or UPDRS (Goetz et al., 2007), although these scales have not been specifically designed for speech or phonation. On the one hand, to study the influence of disease progress, neurologists have also used other motor and non-motor symptoms. On the other hand, having into account that PD is an illness characterized by the failure of the peripheral neuromotor activity, it could be possible that a description of hampered speech articulation, supported by features estimated from the speech signal, could serve as a semantic descriptor of patient's conditions. A possible description of the neuromotor activity from speech can be given in terms of the dynamic changes experimented by the resonant frequencies of the vocal tract, which are known classically as *formants*. The aim of the present study is to evaluate if features derived from the dynamic behavior of formants in sustained vowels are related with some of the indices used by neurologists, and to establish to which extent dynamic measures can be used in the multimodal study of PD speech production. Initially, dynamic estimates of formant activity, such as the absolute kinematic velocity (AKV), which are highly correlated with the superficial myoelectric activity of certain facial muscles (Gómez-Vilda et al., 2017b), seem to be the adequate candidates for such study. The structure of the present paper is as follows: the biomechanical foundations explaining distortion of vowel articulation in terms of formant dynamics is explained in Section 2. Section 3 is devoted to describing the fundamentals of the experimental setup (materials and methods). The results derived from the present work are shown and discussed in Section 4. Conclusions are given in Section 5.

2. KINEMATIC EVALUATION OF PD

Speech is the result of a biomechanical sequence of actuations led by neuromotor activity of chest, neck, oral, nasal and facial muscles under strict coordination. Speech production is constructed by a set of coordinate neurologic processes taking place in the linguistic neuromotor cortex (Démonet et al., 2005). This complex activity driven by primary neurons is transformed into neuromotor actions to excite muscle fibers through the intermediation of basal ganglia, where secondary neurons connected to the muscles of the pharynx, tongue, larynx, chest and diaphragm through sub-thalamic secondary pathways produce sequences of motor actions which activate the respiratory, phonatory and articulatory systems responsible for speech production (neuromechanics). One of the most relevant structures as far as articulation is involved is the Jaw-Tongue Neuromechanical[1] System (JTNS), involving the active (muscular) and passive (bone and tissue inertia and elasticity) of the lower jaw, tongue and lip muscles. For the purposes of the present study, only the structures depicted in Figure 1 is to be considered, having in mind that the inertial part of the facial tissues associated is considered as part of the mandible.

Regarding the biomechanical model, the jaw (J) is fixed to the skull bone at fulcrum (F) as in a third-class lever system. The tongue (T) is supported by the lower jaw and hyoid bone (as well as other facial tissues associated). A reference point of the jaw-tongue system P_{rjt} is defined at $\{x_r, y_r\}$ where forces acting on the system induce movements in the sagittal plane (x: horizontal, y: vertical), such as f_m (masseter), f_{sg} (styloglossus), f_{gi}

Figure 1: Jaw-Tongue Neuromechanical System. The jaw bone is represented in light grey, the tongue structure is represented in light orange. The point P_{rjt} given by $\{x_r, y_r\}$ is the reference point of the biomechanical system.

---
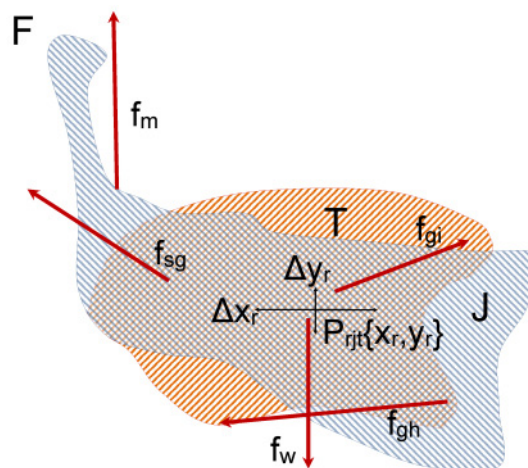
[1] Here the term *neuromechanical* is summarizing the dynamic model of the biological structure (biomechanical in considering muscles, tissues, forces and velocities) as well as the neuromotor spiking activity on the neuromotor pathways involved in the processes of speech production.

(intrinsic glossus), f_{gh} (hyoglossus) and f_w (gravity). The kinematic displacements experienced by $\mathring{P}_{rjt}$ as a result of these forces are given as $\{\Delta x_r, \Delta y_r\}$. Lateral movements orthogonal to the sagittal plane are assumed small enough to be neglected (two degrees of freedom).

The functioning of the speech articulation neuromotor and biomechanical system is explained in Figure 2. The phonation and articulation systems are governed by specific neuromotor units activated from the bulbar structures in the midbrain (1), which control the retraction of the velopharyngeal switch in nasalization (2), activate tongue movements up, down, back and forth (3: intrinsic and extrinsic lingual muscles), modify lower jaw position (4: masseter, stylo and hyoglossus) or control larynx and vocal fold configuration (5: vocalis, crico-arytenoid, transversal and oblique). Through this study only the subsystems (3) and (4) are to be considered regarding speech characterization in relation to PD neurodegeneration.

The way in which this disease alters normal speech articulation may be attributed to fine movement control, which requires proprioceptive feedback. This means that sensor neurons in the activated muscles must send information back to the midbrain in order to adjust the ongoing neuromotor flow to muscles for finely tunning the muscular forces exerted on the biomechanical structures of the jaw-tongue system. The feedback is modulated in short and long loops, depending on the role played by the cerebelum and other structures processing the flow from sensory units in the peripherical muscles. These loops are continuously over or underactivating the neuromotor flow depending on the response of the sensory neurons; therefore a slow amplitude tremor may be detected in the muscle activity in a band over 25 Hz, which is considered a correlate of healthy motor behavior. These feedback loops may not work properly when connections to neuromotor units are affected by some kind of problem, as for instance, lack of neurotransmitters. In PD the lack of dopamine due to degradation of certain cells in *substantia nigra pars compacta* may lead to improper feedback loop function, and this fact is responsible of defective or excessive excitatory flow (inducing hypo- or hypertonicity), or of a sloth overcorrecting flow, which results in wider muscle tension oscillations in the band from 5-8 Hz, and is considered a pathological tremor (Mertens et al., 2013). Therefore, hypo-, hyper- or unstable muscle tone are markers of possible PD neurodegeneration, and these can be traced from speech.

PD-induced neuromotor degeneration speech monitoring has been classicaly based on phonation (fundamental frequency, energy, jitter, shimmer, noise–harmonic ratio, biomechanical parameters (Gómez et al., 2017a; Mekyska et al., 2015; Orozco et al., 2015). Speech articulation is also a possible target to explore. It is a well established fact that muscles in the jaw-tongue structure may modify the vocal tract in a predictable and highly controllable way (speech communication is based on this principle), and acoustic phonetics provides good descriptions of vocal tract configurations, as it is manifested in the vowel polygon (International Phonetics Association, 2015) shown in Figure 3.

The association between articulation gestures (open–close and back–front positions) summarizes in a simple way what researchers have established formally (Sanguineti et al., 1997). As vowel positions are related to formant associations, a possible way to infer articulation dynamics could be formant kinematics (Carmona-Duarte et al., 2016). Let's assume that the first two formants $\{f_1, f_2\}$ of a speech segment are known. A functional relationship

Figure 2: Speech articulation neuromotor and biomechanical system.
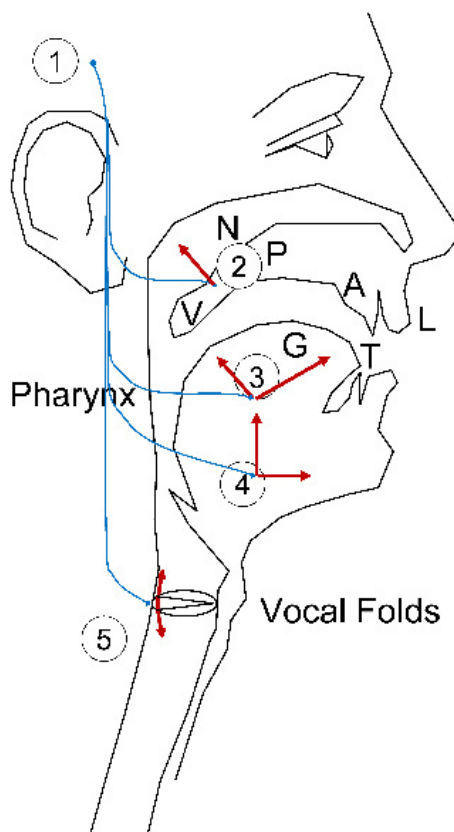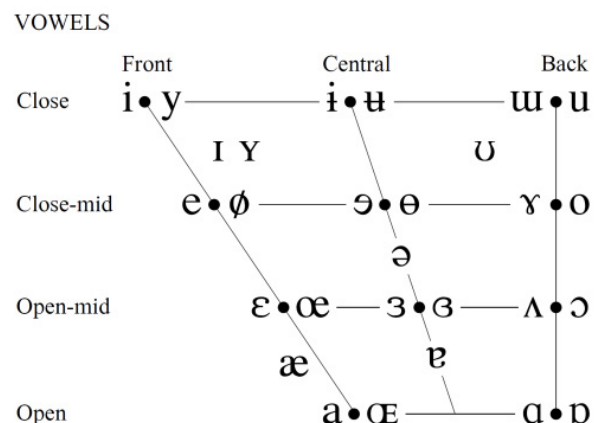

Figure 3: IPA vowel chart.

could be established between formants and the reference point as:

$$(1) \qquad \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$$

where a_{ij} are the transformation weights explaining the position-formant association, and t is the time. This relationship is known to be one-to-many, i.e., the same pair of formants $\{f_1, f_2\}$ may be associated with more than a single articulation position. This inconvenience may be handled by modelling the joint probability of all the possible articulation positions associated to a given formant pair (Dromey, Jang & Hollis, 2013).

In the sequel the following important assumptions have been taken into account:

- The tongue top surface, in opposition to the teeth, alveolar ridge, palate and velum surfaces is the primary element configuring the main articulation cavity (Gerard, Perrier & Payan, 2006).
- The transversal section normal to sound propagation of the main articulation cavity is inversely proportional to the tongue profile in the sagittal plane.
- The lower formants $\{f_1, f_2\}$ are specifically determined by the concatenate tube equivalent model of the main articulation cavity.
- The kinematic displacements $\{\Delta x_r, \Delta y_r\}$ are limited to small oscillations around the reference point P_{rjt} (small signal hypothesis).
- The tongue profile in the sagittal plane is directly related to the kinematic displacements.
- Labialized sounds will not be considered.

It will also be assumed that the system given by (1) may be considered linear, time-invariant and invertible:

$$(2) \qquad \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix}$$

where w_{ij} are the weights of the inverse system. The algorithmic methodology implied in the process of deriving kinematic variables from acoustical ones depends on the estimation and association of formant derivatives in time with the reference point kinematics as:

$$(3) \qquad \begin{bmatrix} v_x(t) \\ v_y(t) \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} \dfrac{df_1(t)}{dt} \\ \dfrac{df_2(t)}{dt} \end{bmatrix}$$

where it has been assumed that v_x and v_y are the dorsal-ventral and caudal-rostral velocities of the reference point. It may be hypothesized that the dorsal-ventral velocity will be mostly related to changes in the second formant (back–front), and that the caudal-rostral velocity will be related to the dynamics of the first formant (up–down). This is equivalent to considering that w_{11} and w_{22} will be negligible compared to w_{12} and w_{21}. Therefore, the absolute kinematic velocity (AKV) of the reference point may be stated as:

$$(4) \qquad |v_{RP}(t)| = \sqrt{\left(w_{21}\frac{df_1(t)}{dt} \right)^2 + \left(w_{12}\frac{df_2(t)}{dt} \right)^2}$$

Reliable estimates for these scale factors may be obtained from diphthong articulations involving changes in the positions of the reference point which show a fast and steady change. The average estimations for both coefficients from an utterance by a typical male speaker were respectively $w_{12} = 1.62 \cdot 10^{-3}$ cm.s and $w_{21} = 1.47 \cdot 10^{-3}$ cm.s. The algorithmic procedure consists in evaluating $\{f_1, f_2\}$ by adaptive inversion of the speech segment after removing radiation effects to produce an estimate of the vocal tract transfer function in time and frequency. Formant positions are obtained either from the local maxima of the transfer function in each time instant or from the polar positions of the prediction polynomials used to invert the speech segment. An example from an utterance of the five vowels [a:, e:, i:, o:, u:] by a 34 year-old normative female speaker is given in Figure 4.

The speech signal sampled at 8 kHz and 16 bits was inverse filtered by a 9-pole filter. Estimations of the LPC spectrum were calculated every 2 ms. The first two formants were smoothened to eliminate glitches and other hazards and artifacts on the kinematic variables by a low-order predictive filter. The kinematic variables v_x and v_y evaluated from the derivatives of f_2 and f_1 as given by (3) are depicted in Figure 5.

As formant kinematics is related to neuromotor activity driving the jaw-tongue biomechanical system, frequency contents above 20 Hz are not presumably of interest due to the inertial properties of the system; therefore, to calculate v_x and v_y formant derivatives they have been low-pass filtered at a cutoff frequency of 20 Hz. Using these data, the absolute value of the reference point velocity given in (4) is shown in Figure 6.

In Figure 6a the reference point velocity is given in horizontal and vertical components or in module (AKV) and angle. The zero degrees of the reference angle correspond to the horizontal movement of the reference point. In Figure 6b the AKV presented in the time domain shows large spikes in vowel onsets and decays, being smaller in vowel nuclei (healthy motor behavior). Figure 6c shows the AKV probability density function and cumulative distribution. It may be seen that the probability density distribution (blue) shows a peak for low velocities, and a gentle decay to 10 cm.s⁻¹ (healthy articulation). The large loops in the right part of the polar plot in Figure 6a are related to changes of the reference point due to adjust-

Figure 4: Spectrogram of an utterance with the five vowels [a:, e:, i:, o:, u:] by a female normative speaker, with the first two formants superimposed (f_1: black, f_2: blue). Top: speech trace and energy envelope (red). Bottom: spectrum and formants.
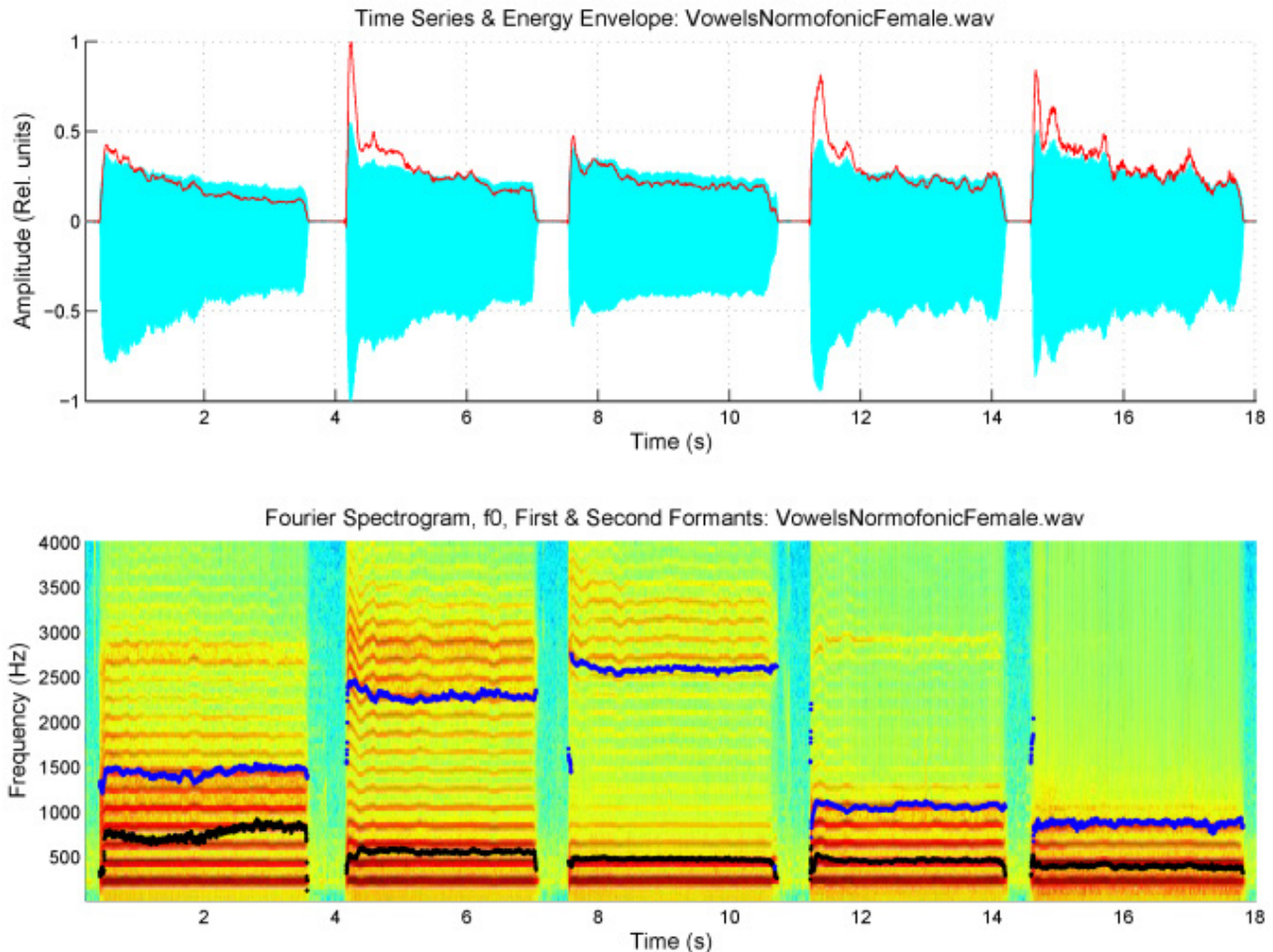
ments of the jaw-tongue system in vowel onsets, whereas vowel nuclei activity is seen as a cloud of small amplitude actions near the center. A similar analysis as the one performed for a normative female speaker is shown in Figure 7 from a 72 year-old female patient suffering from PD, grade 2 in Hoehn & Yahr scale (Hoehn & Yahr, 1967). Tremor and articulation instability can be clearly appreciated in this case, both in harmonics and in formants. It must be said that tremor in harmonics and formants need not be correlated as they depend on different neuromotor pathways, phonation being controlled by laryngeal nerves (vagus), whereas articulation depends mainly on jaw, lingual and facial nerves (branches of trigeminal), and PD may affect both systems differently.

Formant derivatives and low frequency velocities are given in Figure 8.

It may be seen that horizontal and vertical velocities behave quite differently than in the normative case. Now the movements in the vowel onsets are significantly smaller, but much stronger in vowel nuclei (at least four times wider), whereas their frequency is much smaller (around 8 Hz), which is a clear indication of pathological

tremor being present. Finally, the statistical characterization of the AKV is given in Figure 9.

In this case, it may be seen that the AKV horizontal and vertical loops move forwards and backwards as well as up and down, an indication of intense activity during vowel nuclei, as the speaker's proprioceptive system is trying to adjust vowel positions, but an improper feedback loop produces overshoot and undershoot in the articulation structures, and this instability results in tremor. The articulation quality is captured by the AKV distribution (Figure 9c), which shows activity over 20 cm.s^{-1} at a slower decay.

At this point, it will become clear that the AKV probability density function may be a good candidate marker to establish differentiation between stable and unstable articulation due to PD. Due to the nature of AKV as given in (4) it is clear that its probability density function may be modeled as a χ^2 (chi-square) distribution with two degrees of freedom (NIST, 2015), and its general shape will obey the pattern given in Figure 10. Taking into account that the AKV is a correlate of the horizontal and vertical movement speed of the jaw-tongue reference point, the

Figure 5: Formant kinematics (normative subject). Top: time derivatives of the first (blue) and second formant (red). Bottom: low frequency contents of formant derivatives.
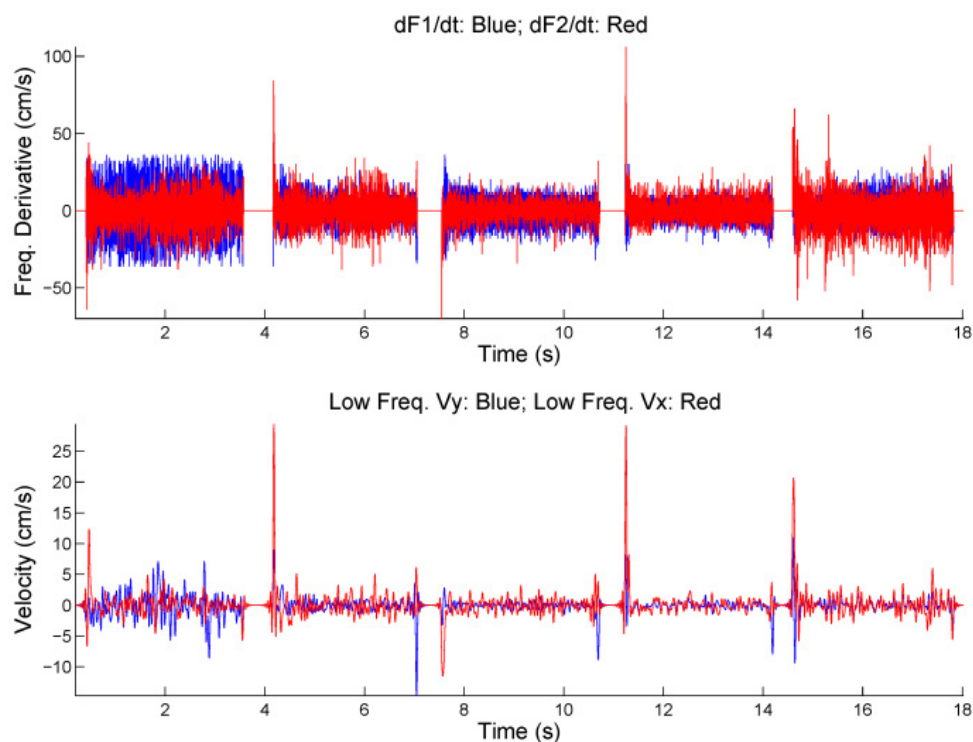

Figure 6: Reference point absolute velocity. a) Module and angle. b) Absolute velocity in time domain. c) Absolute Kinematic Velocity probability density (blue) and cumulative distribution (red).
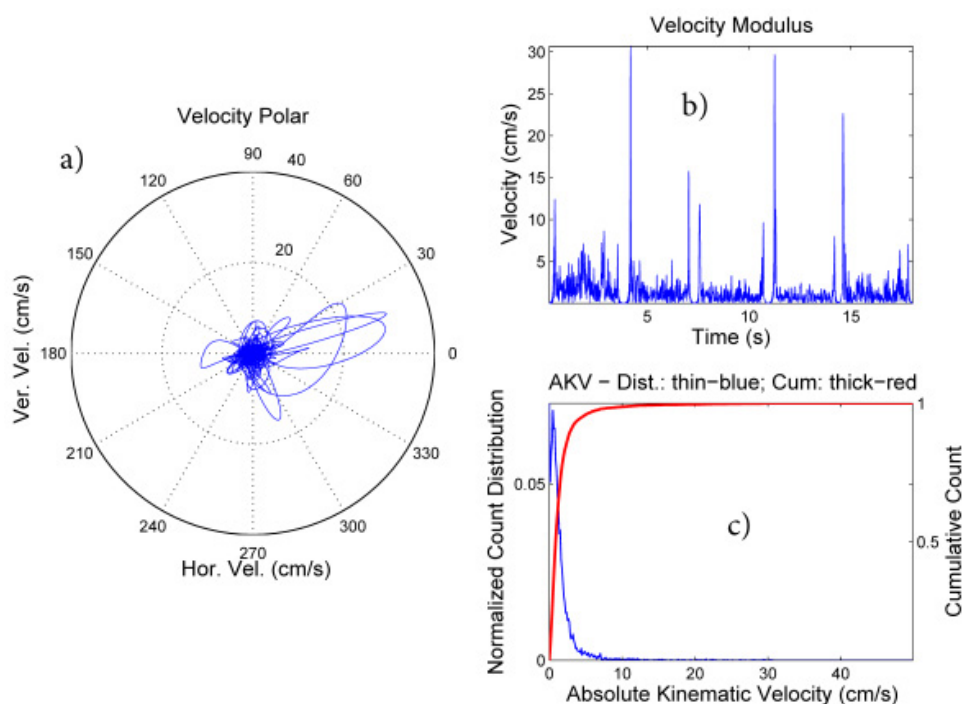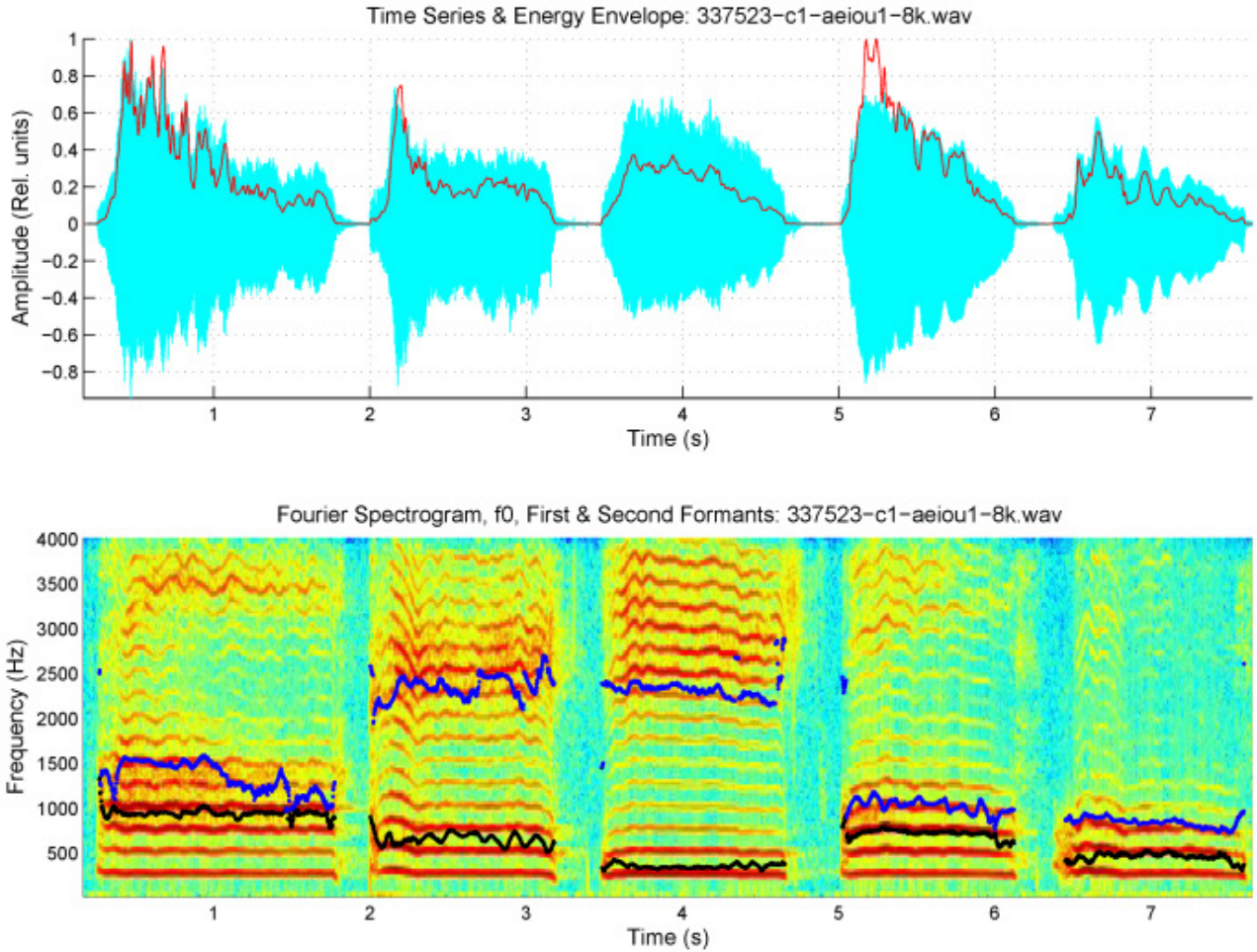
Figure 7: Spectrogram of an utterance with the five vowels [a:, e:, i:, o:, u:] by a female PD patient, with the first two formants (f_1: black, $f2$: blue). Top: speech trace and energy envelope (red). Bottom: spectrum and formants.

kinematic behavior of the utterance is coded in its probability density function, evaluated from a histogram of counts, as follows:

R1. Silent intervals and pauses are accumulated as counts in the origin of the abscise point (zero bin). The higher the value, the larger the number of pauses and the longer their duration.
R2. Stability in formants as in vowel nuclei means smaller and slower formant changes; therefore, it is expected that they will be accumulated near the origin, under 3 cm.s⁻¹.
R3. Smooth formant adjustments due to glides and approximants are to be found around 5 cm.s⁻¹.
R4. Values of the AKV probability density function for vowel onsets and decays are above 10 cm.s⁻¹, following stop consonants and other sharp phonation changes.

The way in which probability functions distribute on these four regions will depend on the kind of pathology being monitored, and on the type of utterance being analyzed. For instance, in pathologies were sentences are used to detect the residual articulation competence, as in Amyotrophic Lateral Sclerosis, it will be expected that distributions from patients will show less activity in R3 and R4 than controls. In certain cognitive pathologies where fluency is affected it will be expected a stronger number of counts in R1 relative to controls if continuous speech is monitored. In the case of PD patients uttering sustained vowels or sequences of vowels, the activity in R2, R3 and R4 is expected to be larger than in controls. To assess the different kinematic behavior between healthy controls and PD patients the mutual information of their respective AKV probability density functions is to be evaluated by Kullback-Leibler's Divergence (KLD; Cover & Thomas, 2006):

$$D_{KLij}\left\{p_{Ti}\left(v_r\right), p_{Mj}\left(v_r\right)\right\} =$$
$$= -\int_{\zeta=0}^{\infty} p_{Ti}(\zeta)\log\left[\frac{p_{Ti}(\zeta)}{p_{Mj}(\zeta)}\right]d\zeta$$

(5)

Figure 8: Formant kinematics (PD patient). Top: time derivatives of the first (blue) and second formant (red). Bottom: low frequency contents of formant derivatives.
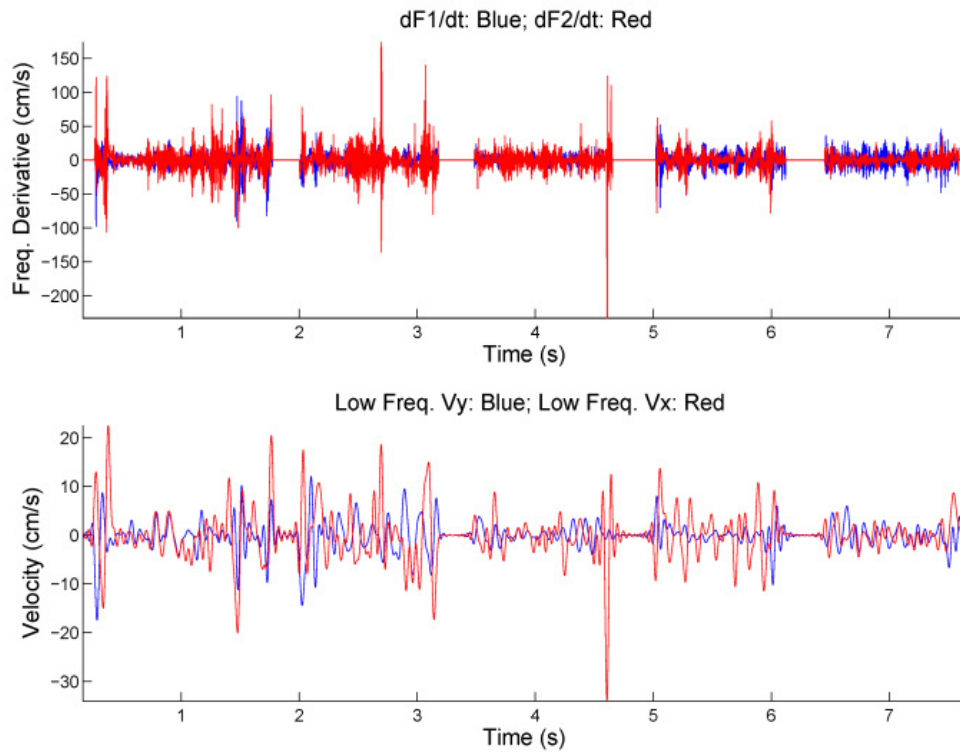

Figure 9: Reference point absolute velocity. a) Module and angle. b) Absolute velocity in time domain. c) Absolute Kinematic Velocity probability density (blue) and cumulative distribution (red).
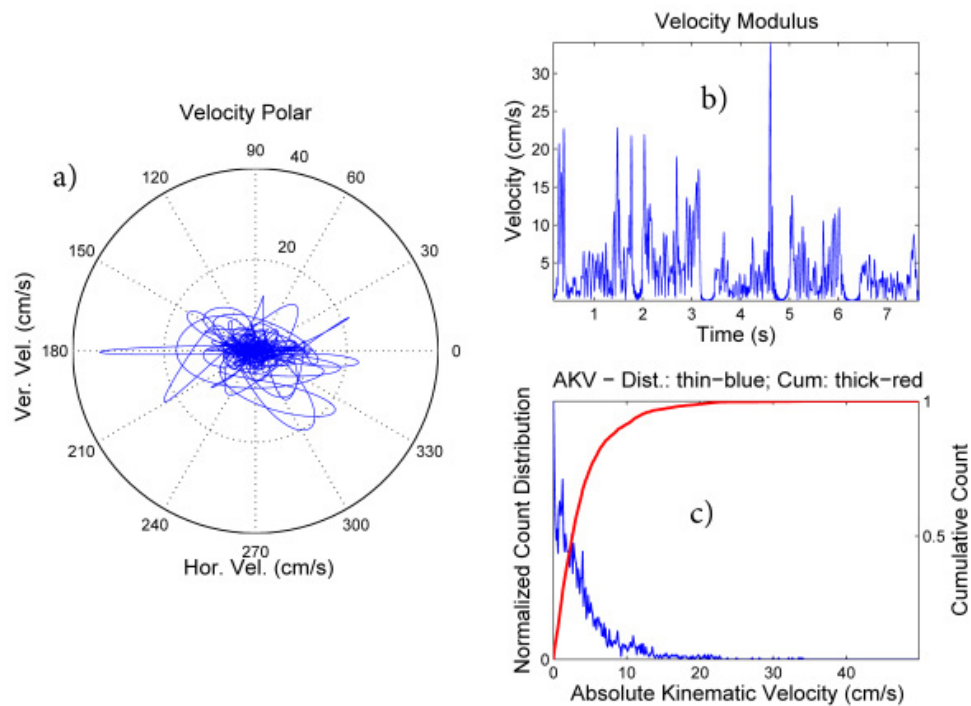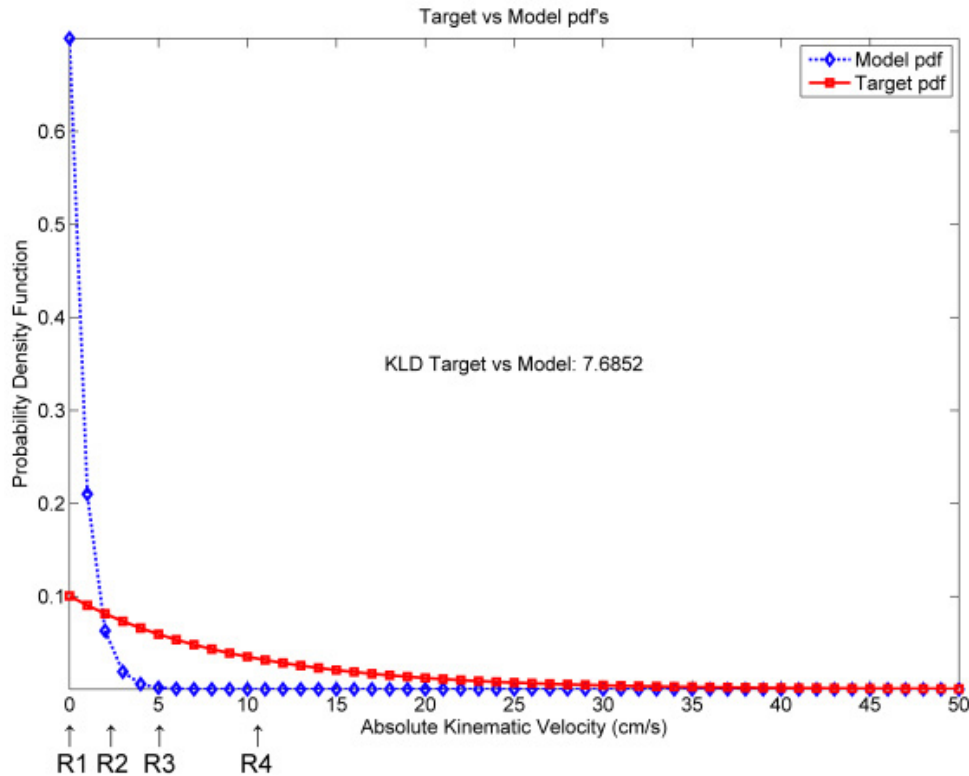
Figure 10: Simulated AKV probability density functions from a target (pathologic speaker) and a model (normative speaker).

where $p_{Mj}(v_r)$ and $p_{Ti}(v_r)$ are the probability density functions of the *j-th* model and *i-th* target subjects (control and patient) respectively, and $v_r \in \mathbb{R}_{\geq 0}$ accordingly to (4). The differential behavior of speech from PD patients with respect to healthy controls, following χ^2 distributions, has been simulated in Figure 10. The AKV probability distribution from a healthy subject (diamond marks) is expected to follow a faster decay than the one from a PD patient when uttering sustained vowels. As the area under both distributions must equal one—being probability distributions—the control subject distribution is expected to start from a larger value than the pathological one. The estimated KLD between both probability density functions in this example is 7.6852 as given in the plot. In Section 3 measurements of the distance between speech samples taken from a longitudinal study case and healthy controls are given and discussed.

3. MATERIALS AND METHODS

The described methodology is used to monitor a database of the sustained vowel [a:] produced by PD patients. In what follows utterances from PD patients is compared with similar utterances produced by healthy controls using KLD as the differentiation feature. The pathological database used is a part of the Czech Parkinsonian Speech Database (PARCZ), recorded at St. Anne's University Hospital in the Czech Republic and consisting of 4 sets of 5 Czech vowels (/a, e, i, o, u/) uttered in 4 different ways: short vowels (modal), long vowels (modal), long vowels (aloud), and long vowels (soft). Recordings were made at 16,000 Hz and 16 bits using calibrated microphones and in soundproof rooms. The subsets selected for the present experiment corresponded to utterances by 54 male and 40 female PD patients, respectively. Recordings of long aloud vowels were selected and processed to obtain the first two formants, and the respective distributions of their AKV were estimated as referred before. The control set consisted in utterances of sustained [a:] produced by 8 male and female subjects. Vowel segments 500 ms long were processed to estimate their vocal tract transfer function, their first two formants every 2 ms, and their AKV probability density estimation as described in Section 2. The AKV distributions for the female set of 40 PD patients plus 8 normative ones are depicted in Figure 11 as an example.

It may be seen that healthy controls show strong activity below 20 cm.s^{-1}, whereas PD patient distributions show activity spread over higher velocities, ranging from 0-80 cm.s^{-1} for subjects 8-25 to 0-160 cm.s^{-1} and beyond for subjects 26-48 (depending on their degree of affection).

Figure 11: Probability density functions of the AKV from female healthy controls (files 1-8, right axis) and female PD patients (files 9-48, right axis). The AKV is given in cm.s⁻1. The control and patient distributions are similar to the normative and pathologic ones given in Figure 10 for models and targets, respectively.
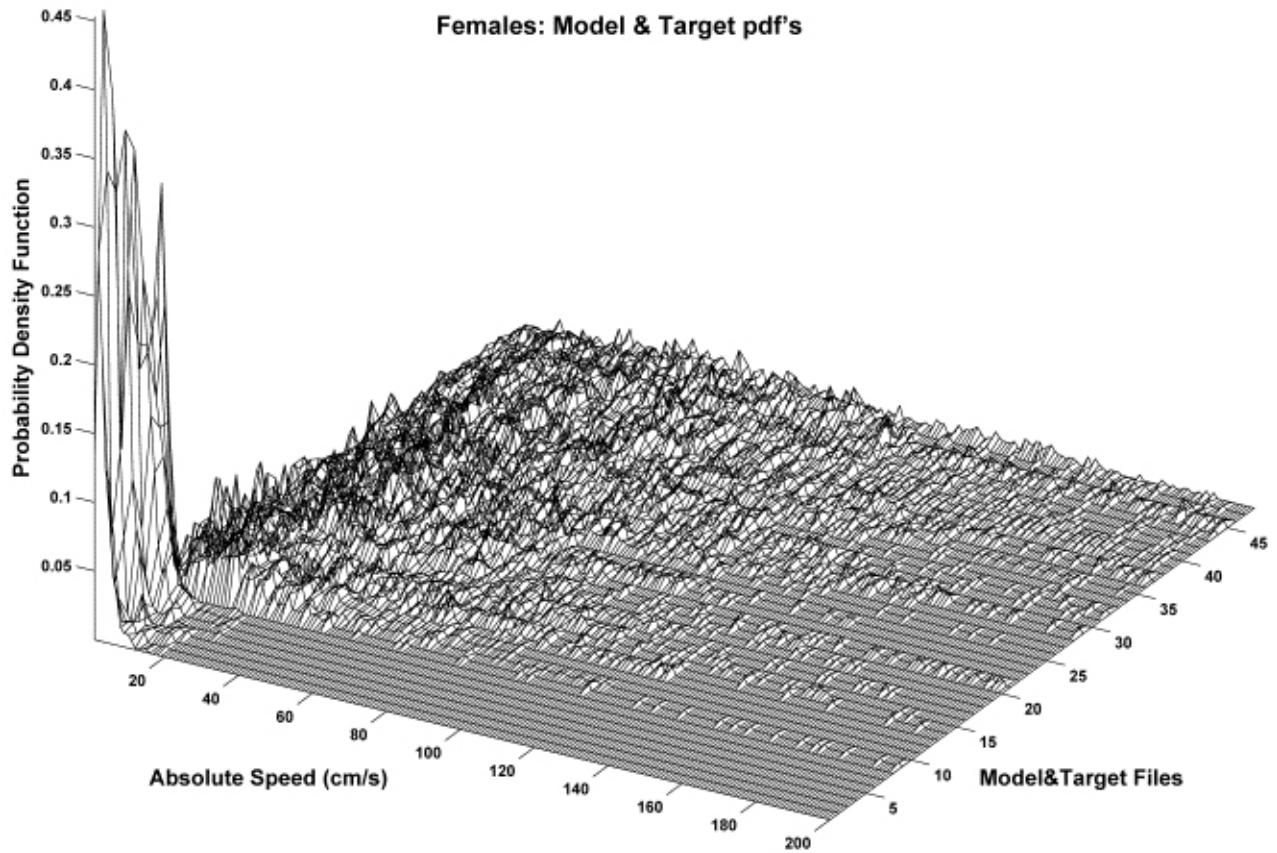
4. RESULTS AND DISCUSSION

Two normative models were defined, as $M_m = \{p_{Mmi}(v_r); 1{\leq}i{\leq}8\}$ for the male set and $M_f = \{p_{Mfi}(v_r); 1{\leq}i{\leq}8\}$ for the female set, where $p_{Mmi}(v_r)$ and $p_{Mfi}(v_r)$ are the AKV probability density functions for the normative subject i, correspondingly male or female. The target (pathological) male and female sets are similarly defined as $T_m = \{p_{Tmi}(v_r); 1{\leq}i{\leq}54\}$ and $M_f = \{p_{Tfi}(v_r); 1{\leq}i{\leq}40\}$, where $p_{Tmi}(v_r)$ and $p_{Tfi}(v_r)$ are the AKV probability density functions for the pathological subject i, correspondingly male or female. These model sets were used to estimate the accumulated D_{KL} of each PD patient of both genders against their respective model set, as

$$(6) \qquad D_{KLj}\left(T_{f,m}, M_{f,m}\right) = \sum_{i \in T_{f,m}} D_{KLij};$$
$$j \in M_{m,f}$$

It is crucial now to see if information theory distances as D_{KL} are related to the PD patient's condition, for which the correlation of D_{KL} to subjective scores used by clinicians were studied. Usually, disease progress is

clinically evaluated by means of general scales as Hoehn & Yahr (1967) or UPDRS (Goetz et al., 2007). Besides, neurologists also use other tests that evaluate freezing of gait (FOG), non-motor symptoms (NMSS), REM sleep behavior disorder (RBDSQ), faciokinesis (FK), phono-respiratory competence (PRC), or phonetic competence (PC; see Fénelon, et al., 2000; Pérez-Lloret, et al., 2014; Stiasny-Kolster, et al., 2015; Ziemssen & Reichmann, 2007). Therefore, Pearson's correlation between D_{KL} and each of the available scores in the set $\mathbf{S} = \{$Age, UPDRSIII, UPDRSIV, FOG, NMSS, RBDSQ, LED, FK, PRC, PC, OC, PRN$\}$ is evaluated, where OC is the average of FK, PRC and PC, whereas PRN is the z-scored correlate of PRC. A global composite score (CS) has been used to represent the set of objective scores in a single value as:

$$(7) \qquad CS = \sum_{i \in \Omega} \omega_{ci} S_i$$

where $\Omega = \{\omega_{ci}\}$ is the set of weights associated with the set of neurological evaluation scores in $\mathbf{S}$. The results of the comparisons for the female set are given in Table 1.

Table 1: Pearson's correlation coefficient beween KLD and clinical scores.

Gender	Age	NMSS	RBDSQ	LED	PRN	CS	*p* value
females	-0.03	-0.26	-0.25	-0.37	0.22	-0.45	0.0065
males	-0.04	-0.21	-0.34	0.15	-0.27	-0.57	0.0000

It may be seen that the most relevant clinical scores D_{KL} are the levodopa equivalent dose (LED), the non-motor symptom score (NMSS), the sleep behavior disorder screening (RBDSQ), and the z-scored phono-respiratory competence (PRN). Age and UPDRS (not shown) are irrelevant or testimonial. The correlation with the composite score is negative and relevant, with a significant *p* value to reject null correlation. The score plots of the DLK and CS for each set are given in Figure 12.

The first observation from the results presented is that the male set presents better correlation between D_{KL} and CS than the female one. This is not an uncommon situation in this kind of studies, as generally models and analysis protocols were initially designed on a male population, and, only latter on, they were adapted to a female population. This underlying fact may be behind gender skew. Besides, the larger number of male cases available in the version of the PARCZ database currently used could also have an influence in these results. Another factor possibly influencing gender skew may be the wider spread of formant frequencies in female voice, which would introduce more dispersion in the results. Another factor of dispersion to be taken into account is the variability and low reliability of subjective scoring scales. No matter how well designed they may be or well-trained raters are, a human subjective factor is implicit and difficult to be removed. This fact stresses the need of developing objective scoring methods, especially in relation to speech production. But, in general, it may be concluded that a certain degree of correlation between formant dynamics and a wide set of motor and non-motor scoring scales exist in PD, and could be conveniently exploited if fused with other articulation static features such as VSA or FCR (Sapir et al., 2010).

5. CONCLUSIONS

The conclusions to be derived from the present work can be summarized as follows:

- A neuromechanical model of the jaw-tongue seems to be efficient in providing a kinematic description of speech from acoustic correlates as formants.
- The statistical distribution estimated from the kinematic absolute velocity of the neuromechanical reference point appears to be a good, compact and normalized descriptor of the dynamic speech behavior.
- The Kullback-Leibler divergence (Cover & Thomas, 2006) is a good estimator of the distance between stable and pathological speech articulation in terms of information theory.
- This estimator appears to be related to other motor and non-motor correlates used currently in the clinical monitoring of PD patients by neurologists.

The field of speech analysis for biomedical applications is of great interest, and it opens new and exciting research lines to speech scientists. It is multidisciplinary and complex, therefore cooperation among large research

Figure 12: Scatter plots of D_{KL} vs CS. Left: female subset. Right: male subset. Linear regression lines have been drawn and formulated for comparison purposes.
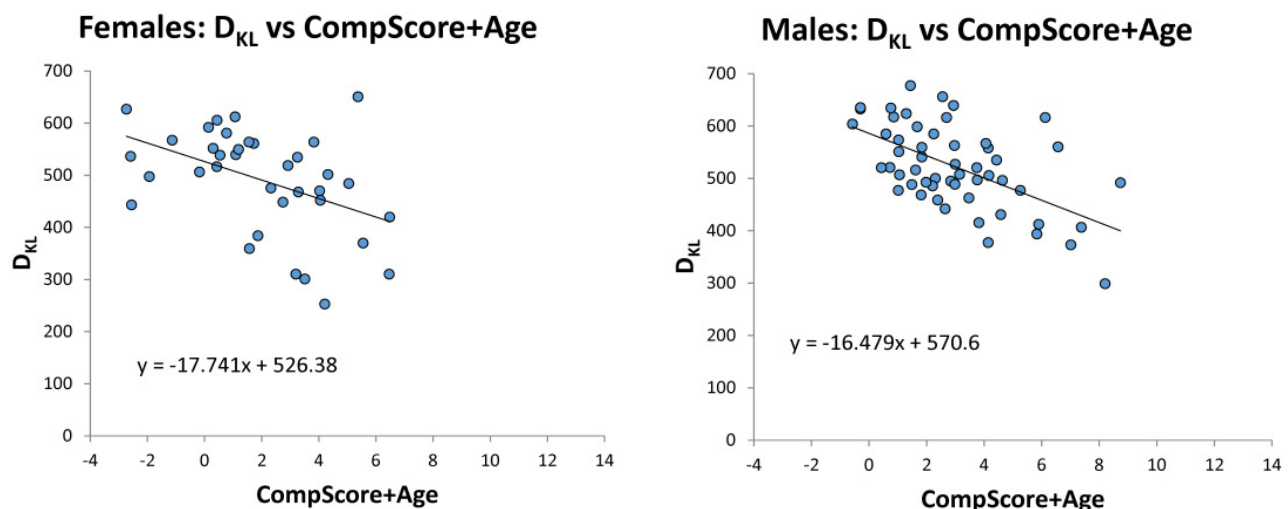
groups including linguists, speech therapists, information processing experts, neurologists, psychologists, etc., must be set up. Databases are to be collected based on new premises. Research on new feature estimation and machine learning paradigms are to be proposed. It is expected that in the next years to come neurodegenerative diseases will be monitored by speech as a daily routine (Cecchi, 2017).

ACKNOWLEDGMENTS

REFERENCES

Carmona-Duarte, C., Alonso, J. B., Díaz, M., Ferrer, M. A., Gómez-Vilda, P., & Plamondon, R. (2016). Kinematic modelling of diphthong articulation. In A. Esposito et al. (Eds.), *Recent advances in nonlinear speech processing*, Vol. 48, 53-60. Cham: Springer. doi.org/10.1007/978-3-319-28109-4_6

Cecchi, G. (2017). With AI, our words will be a window into our mental health. Retrieved from https://www.ibm.com/blogs/research/2017/1/ibm-5-in-5-our-words-will-be-the-windows-to-our-mental-health/

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory*. New York: Wiley.

CSL (2017). Computerized Speech Lab. https://www.pentaxmedical.com/pentax/en/94/1/Visi-Pitch-IV-Model-3950B-Computerized-Speech-Lab-CSL-Model-4500-and-4150B.

Démonet, J. F., Thierry, G., & Cardebat., D. (2005). Renewal of the Neurophysiology of Language: Functional Neuroimaging. *Physiological Reviews, 85,* 49-95. doi.org/10.1152/physrev.00049.2003

Dromey, C., Jang, G. O., & Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Communication,55*(2), 315-328. doi.org/10.1016/j.specom.2012.09.001

Fénelon, G., Mahieux, F., Huon, R., & Ziégler, M. (2000). Hallucinations in Parkinson's disease. Prevalence, phenomenology and risk factors. *Brain, 123*(4), 733-745. doi.org/10.1093/brain/123.4.733

Gamboa, J., Jiménez-Jiménez, F. J., Nieto, A., Montojo, J., Ortí-Pareja, M., Molina, J. A., … Cobeta, I. (1997). Acoustic voice analysis in patients with Parkinson's disease treated with dopaminergic drugs. *Journal of Voice, 11*(3), 314-320. https://doi.org/10.1016/S0892-1997(97)80010-0

Gerard, J. M., Perrier, P., & Payan, Y. (2006). 3D biomechanical tongue modeling to study speech production. In *Speech Production: Models, Phonetic Processes, and Techniques*, New York; Psychology Press, chapter 6, 85-102.

Goetz, C. G., Fahn, S., MartinezMartin, P., Poewe, W., Sampaio, C., Stebbins, G. T., … LaPelle, N. (2007). Movement Disorder Societysponsored revision of the Unified Parkinson's Disease Rating Scale (MDSUPDRS): Process, format, and clinimetric testing plan. *Movement Disorders, 22*(1), 41-47. doi.org/10.1002/mds.21198

Gómez, P., Palacios, D., Rodellar, V., Álvarez, A., Nieto, V., & Martínez, R. (2017a). Parkinson's disease monitoring by biomechanical instability of phonation. *Neurocomputing, 255,* 3-16. doi.org/10.1016/j.neucom.2016.06.092

Gómez-Vilda, P., Palacios-Alonso, D., Gómez-Rodellar, A., Ferrández-Vicente, J. M., Álvarez-Marquina, A., Martínez-Olalla, R., & Nieto-Lluis, V. (2017b). Relating facial myoelectric activity to speech formants. In J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, J. Toledo Moreo & H. Adeli (Eds.), *Proceedings of IWINAC 2017 LNCS 10338, Vol. 2* (pp. 520-530). Cham: Springer. doi.org/10.1007/978-3-319-59773-7

Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism: onset, progression, and mortality. *Neurology, 17*(5), 427-442. doi.org/10.1212/WNL.17.5.427

International Phonetic Association (2005). IPA Chart. Retrieved from http://www.internationalphoneticassociation.org/content/ipa-chart

Mekyska, J., Janousova, E., Gomez-Vilda, P., Smekal, Z., Rektorova, I., Eliasova, I., … López-de-Ipiña, K. (2015). Robust and complex approach of pathological speech signal analysis. *Neurocomputing, 167,* 94-111. doi.org/10.1016/j.neucom.2015.02.085

Mertens, C., Schoentgen, J., Grenez, F., Skodda, S., & Manfredi, C. (2013). Acoustical analysis of vocal tremor in Parkinson speakers. *Proceedings of MAVEBA, 13,* 19-22.

NIST/SEMATECH (2015). NIST/SEMATECH e-Handbook of Statistical Methods. Retrieved from http://www.itl.nist.gov/div898/handbook/

Orozco-Arroyave, J. R., Belalcazar-Bolaños, E. A., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Skodda, S., Rusz, J., … Nöth, E. (2015). Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases. *IEEE Journal of Biomedical and Health Informatics*, *19*(6), 1820-1828. doi.org/10.1109/JBHI.2015.2467375

Pérez-Lloret, S., Negre-Pages, L., Damier, P., Delval, A., Derkinderen, P., Destée, A., …Rascol, O. (2014). Prevalence, determinants, and effect on quality of life of freezing of gait in Parkinson disease. *JAMA Neurology, 71*(7), 884-89. doi.org/10.1001/jamaneurol.2014.753

Sanguineti, V., Laboissière, R., & Payan, Y. (1997). A control model of human tongue movements in speech. *Biological Cybernetics, 77*(1), 11-22. doi.org/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. *Journal of Speech, Language, and Hearing Research, 57*(4), 1330-1343. doi.org/10.1044/2014_JSLHR-S-13-0039

Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language and Hearing Research, 53,* 114-125. doi.org/10.1044/1092-4388(2009/08-0184)

Stiasny-Kolster, K., Sixel-Döring, F., Trenkwalder, C., Heinzel-Gutenbrunner, M., Seppi, K., Poewe, W., … Frauscher, B. (2015). Diagnostic value of the REM sleep behavior disorder screening questionnaire in Parkinson's disease. *Sleep Medicine, 16*(1), 186-189. doi.org/10.1016/j.sleep.2014.08.014

Ziemssen, T., & Reichmann, H. (2007). Non-motor dysfunction in Parkinson's disease. *Parkinsonism & Related Disorders, 13*(6), 323-332. doi.org/10.1016/j.parkreldis.2006.12.014