Subsidia: Tools and Resources for Speech Sciences

16

Detecting neuromotor disease in speech articulation

Pedro Gómez[1], Daniel Palacios[1], Andrés Gómez[1], Cristina Carmona[2], Ana R. Londral[3], Victoria Rodellar[1], Víctor Nieto[1], Miguel A. Ferrer[2], Agustín Álvarez[1]

[1] Universidad Politécnica de Madrid
[2] Universidad de Las Palmas de Gran Canaria
[3] Universidade de Lisboa
e-mail: pedro@fi.upm.es

Citation / Cómo citar este artículo: Gómez, P., *et al.* (2019). Detecting neuromotor disease in speech articulation. In J. M. Lahoz-Bengoechea & R. Pérez Ramón (Eds.), *Subsidia. Tools and resources for speech sciences* (pp. 103–108). Málaga: Universidad de Málaga.

ABSTRACT: Speech articulation may be an important resource to study neuromotor activity in relation with neurological diseases, such as Parkinson's, Alzheimer's or Amyotrophic Lateral Sclerosis. Through the present work a biomechanical model for the joint structure of the jaw and tongue is introduced. This model allows putting into relation formant trajectories with jaw-tongue kinematics in terms of dorso-ventral and rostro-caudal movements of the structure. The reconstruction of the absolute velocity of the system from inverse filtering estimates of the first two formants allows the introduction of the velocity distribution histograms as descriptors of articulation behavior. A study case using recordings separated in time from a patient suffering from Amyotrophic Lateral Sclerosis, compared to distributions from a normative control is presented. It may be seen that patient's distributions show larger contents for low ranges of velocity and smaller contents for high velocity ranges compared to the control distribution. Global comparisons of velocity distribution profiles between the control and patient's records using Information Theory measurements may be used to monitor illness conditions and progress.

Keywords: speech processing; speech articulation; speech kinematics; neuromotor dysarthria; amyotrophic lateral sclerosis.

RESUMEN: La articulación del habla puede ser un recurso importante para estudiar la actividad neuromotora en relación con enfermedades neurológicas, como el Parkinson, el Alzheimer o la Esclerosis Lateral Amiotrófica. Este trabajo presenta un modelo biomecánico de la estructura que forman conjuntamente la mandíbula y la lengua. Este modelo permite relacionar las trayectorias de los formantes con aspectos cinemáticos de la mandíbula y la lengua en términos de movimientos dorso-ventrales y rostro-caudales de dicha estructura. La reconstrucción de la velocidad absoluta del sistema a partir de un filtrado inverso de las estimaciones de los dos primeros formantes permite introducir histogramas de la distribución de la velocidad como descriptores del comportamiento articulatorio. Se presenta un estudio de caso a partir de grabaciones separadas en el tiempo de un paciente que sufre Esclerosis Lateral Amiotrófica, comparadas con las distribuciones de un sujeto control normativo. Se observa que las distribuciones del paciente muestran una mayor actividad en los rangos de velocidad bajos y una menor actividad en los rangos de velocidad altos, en comparación con la distribución control. Se propone hacer comparaciones globales de los perfiles de distribución de la velocidad entre las grabaciones control y las del paciente utilizando medidas basadas en la Teoría de la Información, y que este uso puede ayudar a monitorizar las condiciones y la evolución de la enfermedad.

Palabras clave: procesamiento del habla; articulación del habla; cinemática del habla; disartria neuromotora; esclerosis lateral amiotrófica.

1. INTRODUCTION

Speech articulation is a result of oro-naso-pharyngeal tract modifications produced by the movement of four main groups of muscles (Jürgens, 2002): jaw muscles, lingual extrinsic and intrinsic, oro-facial, and velo-pharyngeal.

As neuromotor units can produce only muscle contractions, when activated, each group of muscles is fitted to an agonist-antagonist pair (Kandel, Schwartz, & Jessell, 2000) to produce smooth movements in both directions. When the agonist group contracts and pulls in one direction, the antagonist group must relax and yield, and vice-versa. This delicate action-reaction

stimulus is regulated by an excitatory-inhibitory network of neurons.

The steady movement of muscles is a result of these agonist-antagonist actions, and has been very well described and modelled for the hand-writing function (Plamondon, Djioua & Mathieu, 2013). It has been shown that this model can also be applied to the study of speech articulation (Gómez-Vilda, Londral, Rodellar-Biarge, Ferrández-Vicente, & de Carvalho, 2015).

The intention of the present study is to show that the implications of the agonist-antagonist neuromotor function can be taken one further step ahead when dealing with speech fluency features (Singh, Bucks, & Cuerden, 2001), such as verbal rate, mean duration of pauses, standardized phonation time, and standardized pause rate, or articulation features, such as: speech rate (syllables / total duration time) or articulation rate (syllables / total locution time).

The fluency of speech production is subject to a temporal dynamic flow, which may be observed at different time scales. In the first level, the presence or absence of speech can be characterized by the presence of phonic groups, which are segments of time where a signal activity over the background noise level of the channel can be observed over a significant value. The intervals between phonic groups are considered as pauses. In the second level, phonic groups can be divided into phonated and non-phonated intervals, depending if there is vocal fold activity involved or not. In the third level, phonated intervals are divided into segments where formant activity is stable under a modulation frequency limit (in Hz/s) or not.

Having into account the ubiquity of speech recording and transmission on IP platforms, the present work is intended to search for a statistical description of the dynamic phenomena present in the three levels to describe articulation dynamics for its application in neurologic disease characterization from speech.

The paper is organized as follows. The basic model explaining articulation neuro-motor foundations is presented in section 2. Section 3 is devoted to introduce a study case used as an example for the characterization of pathologic speech (a case of amyotrophic lateral sclerosis), and presents the basic algorithmic procedures and the application supporting them. Section 4 presents the results of analysis and their statistical analysis. And section 5 summarizes the main conclusions of the study.

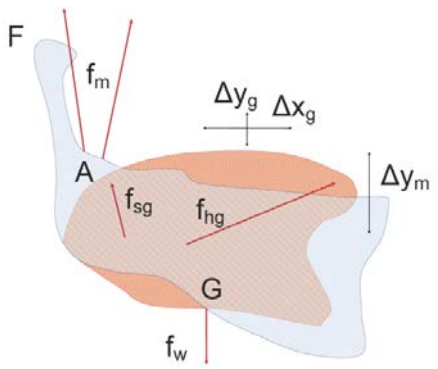
2. BACKGROUND MODEL

The process of articulation is a complex biomechanical task which implies the timely organized activation of different muscles under the control of the neuromotor cortex. The concurrent activation of respiration and phonation muscles must be accompanied by the joint action of at least the jaw, lingual (extrinsic and intrinsic) and facial, as well as the gloso-velo-pharyngeal muscles. Many neurologic pathologies result in the inadequate functioning of some of these biomechanical systems, and altered phono-acoustic correlates appear as a consequence, which may serve as indicators of the pathological condition of the speaker, and help in the characterization of the pathological process under a functional point of view.

The complexity of the whole articulation system requires a divide-and-conquer approach for its study. In the present work the study will be focused in the jaw-tongue biomechanical system (Gerard, Perrier & Payan, 2006), considered as a first-order mass-spring structure as described in Figure 1.

The jaw-tongue subsystem is presented as a third-class lever (Röhrle & Pullan, 2007), which is fixed to the cranial structure at point F (fulcrum), experiencing the force of gravity at G, and supported mainly by the action of the masseter pair at A. In this model several other acting muscles and substructures have been omitted for the sake of computational simplification. The model assumes that all forces acting on the jaw-tongue massive structure (other than gravity, which is referred to the centre of gravity) can be referred to a certain joint-mandible dynamic reference point (JMDRP), with implicit coordinates x and y in the sagittal plane, such that the axis x refers to movement in the dorsal-ventral direction (DV), and the axis y refers to movement in the caudal-rostral direction (CR). Lateral movements orthogonal to the sagittal plane are assumed small enough not to be considered (thus giving a system with two degrees of freedom). The kinematic variables relevant to the study are the displacements Δx and Δy relative to the JMDRP, which will be contributed mainly by the CR displacement of the jaw (Δy_m) and the DV and CR displacements of the tongue (Δx_g Δy_g). Important additional assumptions to this model are that the tongue system is the main surface opposite to the palate ceiling, acting as a solidary hydrostatic bulk, and that the relative displacements between these surfaces configure the main articulation cavity (Gerard *et al.*, 2006). On its turn, for the dynamic part, the main forces acting on the JMDRP (besides gravity acting on the center of gravity, as said) are the masseter contraction (f_m), the styloglossus action (f_{sg}) and the hyoglossus action (f_{hg}, both superior and inferior muscle bundles).

Figure 1: Biomechanical model of the jaw-tongue subsystem. The jaw bone is represented in light grey; the tongue structure is represented in light orange.



Under these assumptions, the resonance model involving the first two formants f_1 and f_2 can be put into relation with the positions of the jaw-tongue bulk relative to the JMDRP (Sanguineti, Laboissiere & Payan, 1997) by the dynamic system in (1):

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

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 to 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).

Under certain invertibility assumptions, which will not be given here for the sake of brevity, the system in (1) may be written in opposite terms to help expressing the algorithmic methodology implied in the process of deriving kinematic variables from acoustical ones, as in (2):

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

where w_{ij} are the weights of the inverse system. The first time derivative of this system allows associating formant derivatives in time with the JDMRP kinematics, as in (3):

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

where it has been assumed that the system is linear and time-invariant, and v_x and v_y are the DV and CR velocities of the JMDRP.

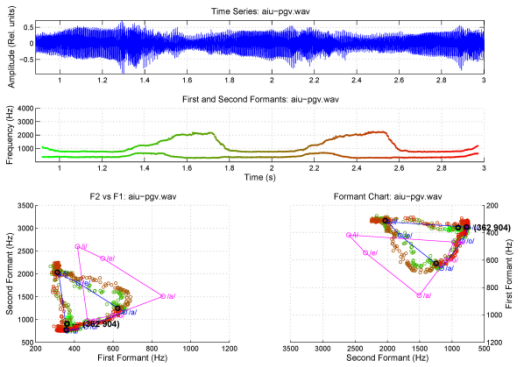
The values of the weights w_{ij} are related to the kinematics of the specific speaker, and can be considered as a biometrical mark of the person. It may be hypothesized that the DV velocity will be mostly related to changes in the second formant (back-front), and that the CR velocity will be related to the dynamics of the first formant (up-down).

An estimate of the absolute velocity of the JMDPR may be evaluated as in (4):

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

Therefore it will be hypothesized that w_{11} and w_{22} will be negligible compared to w_{12} and w_{21}. Reliable estimates for these scale factors may be obtained from diphthong articulations as for instance [aj] or [ja], involving changes in the positions of the JMDRP which are not affected by strong labialization. In a practical case, estimations of the scale coefficients may be obtained from sequences as the one shown in Figure 2, using the first two formant changes comprised between 1.46 s and 1.6 s, or 2.26 s and 2.4 s.

Figure 2: Formant structure of the monotonically repeated sequence /aiu/ uttered as [...ajijuwa...]. Upper template: time domain signal; middle template: first two formant patterns in time; lower template: two representations of formant positions over the vowel triangle.



The averaged estimations for both coefficients from the data in Figure 2 are respectively w_{12}=1.62.10^{-3} cm.s and w_{21}=1.47.10^{-3} cm.s.

3. MATERIALS AND METHODS

In what follows, results from a study case involving a patient of Amyotrophic Lateral Sclerosis (ALS) will be shown. Five recordings were taken from a 64-year old female patient from the Neurology Department of Hospital de Santa Maria, in Lisbon, Portugal, suffering from ALS, who had been diagnosed 1 year before. Recordings were taken at each neurological control, spaced 3 months (respectively: P1T0, P1T1, P1T2, P1T3 and P1T4). The recordings consisted in the utterance of a popular sentence from Fernando Pessoa: /*tudo vale a pena quando a alma não é pequena*/ (IPA: [tʊðʊɥɑlɐ pɐ̃næ kwændʊaaɫmɐ nɐ̃ɛ̃ pkɐ̃næ]) by the ALS patient during the evaluations. The recordings were taken at 44.1 kHz and 16 bits. The same recording was taken from a 36-year old healthy female control. The selection of ALS patients for this study was not coincidental. It is well known that these patients suffer from a continuous degradation of neuromotor functions which contribute to a progressive reduction of the vowel triangle (Yunusova, Weismer, Westbury, & Lindstrom, 2008). The main aim of the present study is to check the extent of the deterioration of articulation functions in the dynamic time domain as well. The basic methodological protocol consists in the following steps:

- Recordings are undersampled to 8 kHz.
- The vocal tract transfer function of the speech segment is evaluated by a 9-pole adaptive inverse LP filter (Deller, Proakis & Hansen, 1993) with a low-memory adaptive step to grasp fine time variations.
- The first two formants are estimated by evaluating the roots of the associated inverse polynomials. The formant estimation resolution used is 2 Hz, and an estimation is produced every 2 ms.

Gómez, P., *et al*. (2019). Detecting neuromotor disease in speech articulation. In J. M. Lahoz-Bengoechea & R. Pérez Ramón (Eds.), *Subsidia. Tools and resources for speech sciences* (pp. 103–108). Málaga: Universidad de Málaga.

- The derivatives of the first two formants are used to estimate the absolute velocity of the JMDRP following Equation (4).
- The values of the absolute velocity are used to build a histogram as a function of the absolute velocity distribution.
- The histograms are used to estimate probability density functions by Kolmogorov-Smirnov approximations (Webb, 2003).
- Kullback-Leibler's Divergence (Webb, 2003) is estimated between each patient's recording distribution $p_{Pi}(v)$ vs that of the control subject $p_C(v)$ as by (5), where the absolute velocity of the JMDRP has been defined in a given interval, which for the present study is in the range R_v={0, 200 cm.s^{-1}}. The above described steps are programmed into a Graphical User Interface (BioMet®Ling: www.glottex.com), which is shown in Figure 3.

$$D_{KL}(p_{Pi}, p_C) = \int_{v \in R_v} p_C(v) \log \left| \frac{p_{Pi}(v)}{P_C(v)} \right| dv \quad (5)$$

4. RESULTS AND DISCUSSION

In what follows the absolute velocity profiles for the control subject (CF) and the first (P1T0) and last (P1T4) patient's utterances are given as a reference in Figure 4. These profiles correspond to low-frequency movements under 20 Hz.

It may be seen that the upper template (Figure 4a) shows a fast and well organized action pattern, where the strongest neuromotor spikes reach values between 20 and 40 cm.s^{-1}. The total utterance duration is around 2.7 s, with two pauses. The middle template (Figure 4b), corresponding to the first patient's evaluation shows a slower utterance lasting 5.3 s, where four pauses may be seen, and the strongest spikes are between 10 and 30 cm.s^{-1}. The lower template, corresponding to the last patient's evaluation (Figure 4c) needs almost 9 s to complete the same utterance, there are no pauses, and the largest spikes are between 8 and 20 cm.s^{-1}.

On its turn, Figure 5a shows the results of estimating the probability density functions of the absolute velocity profiles (low and high frequency) for CF and the five patient recordings, the comparison between CF and the first patient's one (Figure 5b), between CF and the last patient's one (Figure 5c) and between the patient's first and last ones (Figure 5d).

Figure 3: Graphical User Interface of BioMet®Ling

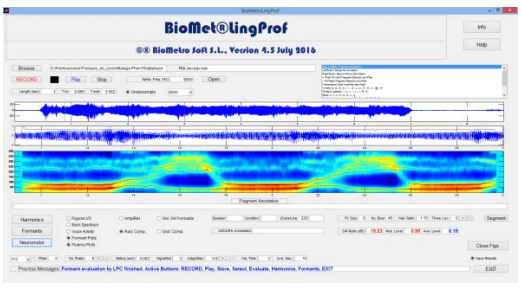


Figure 4a: Absolute velocity profile from the control subject.

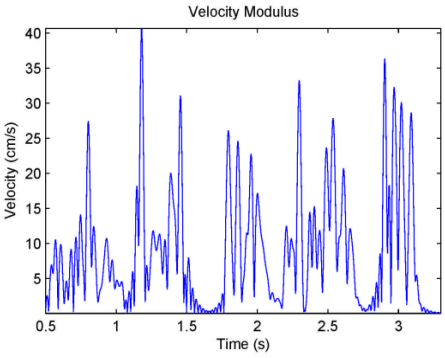


Figure 4b: Absolute velocity profile from the patient's first utterance.

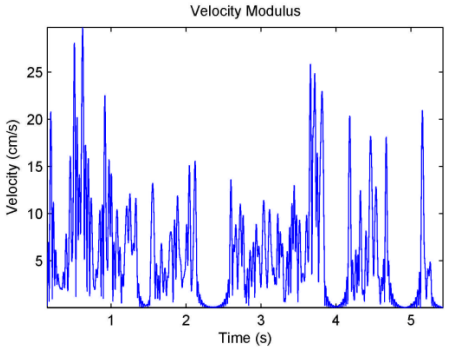
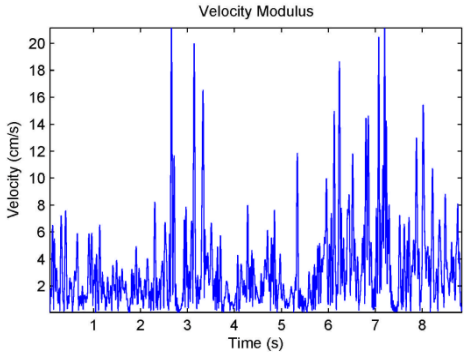


Figure 4c: Absolute velocity profile from the patient's last utterance.



It may be seen that the control subject probability density function shows velocity profiles over 80 cm.s^{-1}, which can be barely appreciated in the distributions from the ALS patient. The velocity range is larger than in the plots of Figure 4 because high and low frequency contents have been taken into account. As illness progresses, the distribution contents displace to lower velocity bins.

Figure 5b shows that the velocity distribution from the patient's first recording displays more contents for lower velocity bins than the one from the control subject, whereas for larger velocity bins the control subject shows more contents than the patient's one. This tendency is more evident when comparing the

Figure 5a: Absolute velocity probability functions for the control and patient's utterances.

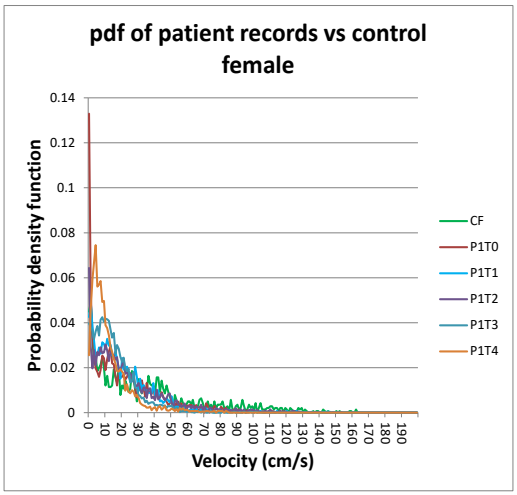
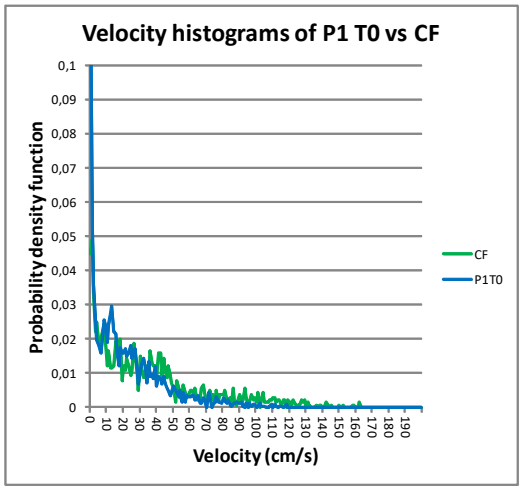


Figure 5b: Absolute velocity probability functions for the control and patient's first utterance.



patient's last distribution against the control subject (Figure 5c). Finally, in Figure 5d the patient's first and last distributions are compared, showing that the articulation dynamics of the patient has lost components in the high velocity range, which are transferred to the low velocity range. This is a clear indication of illness progression, which manifests itself as a sloth and less vivid articulation activity due to the degradation of jaw and tongue neuromotor units characteristic of ALS (known also as the motor neuron, or Lou Gherig disease).

Table 1 shows the results from estimating Kullback-Leibler's Divergence for the patient's five recordings vs the control subject.

Table 1: Kullback-Leibler's Divergence for the patient's five chronological recordings vs the control subject.

Recording	KLD
P1T0	0.42657477
P1T1	0.53332203
P1T2	0.51182122
P1T3	0.70738088
P1T4	0.93337742

Figure 5c: Absolute velocity probability functions for the control and patient's last utterance.

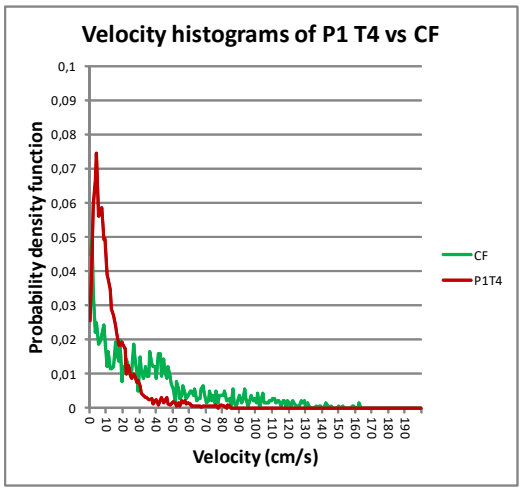
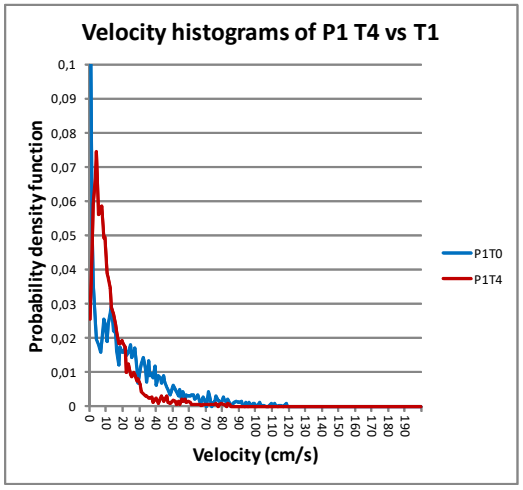


Figure 5d: Absolute velocity probability functions for the patient's first and last utterances.



It can be concluded again that the absolute divergence from the normative articulation profile shown by the control subject and that of the ALS patient increases as illness progresses except for recording P1T2, where a possible stagnation in the degradation of speech articulation may be circumstantially observed.

Generally speaking, it may be concluded that the articulation dynamic features have been captured by the absolute velocity probability distribution as derived from the velocity histogram in the range 0–200 cm.s^{-1}, which may become a comparison standard when information-theory derived statistics as Kullback-Leibler's Divergence are used. A very interesting property of the velocity histogram is that it may comprise different articulation dynamics in its structure, such as at the level of pauses and phonic groups (very low level velocity profiles), phonated and non-phonated intervals (low level velocity profiles), and formant modulation in syllabic segments (mid- to high level velocity profiles). The characterization of the velocity profiles accordingly to speech own dynamics could open its possible application to other neurodegenerative diseases which present fluency and

dynamic dysarthrias, such as Parkinson's or Alzheimer's as well.

5. CONCLUSIONS

Several conclusions can be derived from the present study, the following seem to be the most relevant ones among them:

- The articulation dynamics may be derived from the temporal evolution of the first two formant derivatives.
- A frequency resolution of 2 Hz and a time resolution of 2 ms give enough accuracy to characterize fast formant changes.
- The absolute velocity profile of the JMDRP is a rather semantic correlate, relating high and low motion with the normal behaviour of the main biomechanical system related to speech articulation.
- The application of the methodology to a case study of progressive speech deterioration produced by ALS shows the viability and applicability of the methodology.
- The use of statistical distance measurements derived from Information Theory may be a powerful means to produce objective estimates to track illness progress, opening new ways for distant patient monitoring by e-Health platforms.

These conclusions are to be confirmed on a larger number of cases, and the application of the methodology to other neurologic pathologies is foreseen in the next future.

6. REFERENCES

Deller Jr, J. R., Proakis, J. G., & Hansen, J. H. (1993). *Discrete time processing of speech signals*. Englewood Cliffs, NJ: Prentice Hall.

Dromey, C., Jang, G. O., & Hollis, K. (2013). Assessing correlations between lingual movements and formants. *Speech Communication*, *55*(2), 315–328.

Gerard, J. M., Perrier, P., & Payan, Y. (2006). 3D biomechanical tongue modeling to study speech production. In J. Harrington, & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 85–102). New York: Psychology Press.

Gómez-Vilda, P., Londral, A. R. M., Rodellar-Biarge, V., Ferrández-Vicente, J. M., & de Carvalho, M. (2015). Monitoring amyotrophic lateral sclerosis by biomechanical modeling of speech production. *Neurocomputing*, *151*, 130–138.

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, *26*(2), 235–258.

Kandel, E. R., Schwartz, J. H. & Jessell, T. M. (2000). *Principles of neural science*. New York: McGraw-Hill.

Plamondon, R., Djioua, M., & Mathieu, P. A. (2013). Time-dependence between upper arm muscles activity during rapid movements: Observation of the proportional effects predicted by the kinematic theory. *Human Movement Science*, *32*(5), 1026–1039.

Röhrle, O. & Pullan, A. J. (2007). Three-dimensional finite element modelling of muscle forces during mastication. *Journal of Biomechanics*, *40*, 3363–3372.

Sanguineti, V., Laboissiere, R., & Payan, Y. (1997). A control model of human tongue movements in speech. *Biological Cybernetics*, *77*(1), 11–22.

Singh, S., Bucks, R. S., & Cuerden, J. M. (2001). Evaluation of an objective technique for analysing temporal variables in DAT spontaneous speech. *Aphasiology*, *15*(6), 571–583.

Webb, A. R. (2003). *Statistical Pattern Recognition*. Chichester, UK: Wiley.

Yunusova, Y., Weismer, G., Westbury, J. R., & Lindstrom, M. J. (2008). Articulatory movements during vowels in speakers with dysarthria and healthy controls. *Journal of Speech, Language, and Hearing Research*, *51*(3), 596–611.