

Relating Facial Myoelectric Activity to Speech Formants

Pedro Gómez-Vilda¹(✉), D. Palacios-Alonso¹, A. Gómez-Rodellar¹,
José Manuel Ferrández-Vicente², A. Álvarez-Marquina¹,
R. Martínez-Olalla¹, and V. Nieto-Lluis¹

¹ Neuromorphic Speech Processing Lab, Center for Biomedical Technology,
Universidad Politécnica de Madrid, Campus de Montegancedo,
28223 Pozuelo de Alarcón, Madrid, Spain
pedro@fi.upm.es

² Universidad Politécnica de Cartagena, Campus Universitario Muralla del Mar,
Pza. Hospital 1, 30202 Cartagena, Spain

Abstract. Speech articulation is conditioned by the movements produced by well determined groups of muscles in the larynx, pharynx, mouth and face. The resulting speech shows acoustic features which are directly related with muscle neuromotor actions. Formants are some of the observable correlates most related to certain muscle actions, such as the ones activating jaw and tongue. 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. A relevant question is how far speech correlates and neuromotor action are related. This question is answered by the present study using electromyographic recordings on the masseter and the acoustic kinematics related with the first formant. Correlation measurements help in establishing a clear relation between the time derivative of the first formant and the masseter myoelectric activity. Monitoring disease progress by acoustic kinematics in one case of Amyotrophic Lateral Sclerosis ALS is described.

Keywords: Speech neuromotor activity · Facial myoelectric activity · Dysfluency · Dysarthria · Amyotrophic Lateral Sclerosis

1 Introduction

Speech is the result of different cognitive processes planned in the brain cognitive cortex, interpreted as specific neuromotor actions in midbrain, transformed into agonist-antagonist actions in muscles, and transferred to air as sounds. Speech production is planned and instantiated in the linguistic neuromotor cortex (Broadmann's areas 4, 6, 8, and 44–47 [1]). The neuromotor speech sequence activates the muscles of the pharynx, tongue, larynx, chest and diaphragm through sub-thalamic secondary pathways. The source-filter model of speech production, hypothesizes that an excitation source is generated either

by the joint action of chest and larynx muscles (phonation) or by turbulent airflow induced by air exhalation in different parts of the oro-naso-pharyngeal tract (ONPT). The excitation source, in its propagation through the ONPT is acoustically filtered, resulting in the enhancement or reduction of certain spectral bands. These can comprise either message codes in a specific language, personal biometric features, or reveal emotional, physiological and psychological state conditions. Among these, possible alterations of the speaker's neurological conditions are also encoded. Certain diseases affect mainly to neuromotor units in the basal ganglia, brain stem and cerebellar structures (from motor neuron connection to muscle spindle activation, or even motor fibre degeneration), and are known as neuromotor diseases. Speech alterations produced by neurological diseases may affect different levels of speech production, these being mainly the phonation and articulation levels in neuromotor diseases, or fluency levels in cognitive diseases. Failures in neurotransmission, or in improper speech planning will produce perturbations in the respiration, phonation and articulation giving place to specific dysphonias and dysarthrias, poor prosody (monotonous speech), poor VOT (especially when switching nasal to oral sounds), and deficient fluency (low syllable rate, longer inter-syllable pauses, etc.) [2]. The aim of the present work is to establish a clear connection between acoustic correlates and neuromotor muscular activity, therefore, fluency correlates of cognitive etiology will not be considered. The working hypothesis assumes that on the one hand the acoustic-phonetic correlates observed on the first two formants derived from the speech signal, and the myoelectric surface signal [3] measured on the masseter have to present common features corresponding to a close relationship based on the neuromotor actions governing the muscle contractions modulating the open/close (O/C) features of vowels and diphthongs. For such, Sect. 2 is devoted to the description of the articulation biomechanical model involved, in Sect. 3 the procedures and methods for signal collecting and processing will be shown. Results are given in Sect. 4, and a discussion follows on their significance and relevance. Section 5 is devoted to conclusions.

2 Neuromotor Articulation Model

Regarding articulation activity, the muscular structures implied are those modifying the ONPT, basically at the level of the naso-pharyngeal switch, and the jaw, tongue and lip gestures, which condition the properties of the equivalent acoustic filter. These properties will affect mainly to resonances of the tract, modifying both their static (vowel) and dynamic (consonantal) patterns. Vowels are well defined from their resonances, associated to the concept of formants, which are frequency bands especially enhanced by the ONPT resonances. The relation between the first two formant positions and the nature of the resulting vowel is well established in literature [4]. The relation between resonances and the articulation gestures (velum, tongue, jaws and lips) is less straight forward, but in general it can be said that certain gestures are clearly related to static vowel properties as defined by their first two formants. In general, 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 [4]. The specific relation between the jaw position and the first formant (F) is given in Fig. 1. The lower jaw position is mainly fixed by two muscles which act as agonist and antagonist, these being the masseter and the geniohyoid, respectively. The action of gravity is to be added as a third force (f_g) to the forces produced by these muscles (f_m , f_h). The jaw and tongue act solidly as a dynamic structure in some way, therefore it is difficult to separate their independent movements.

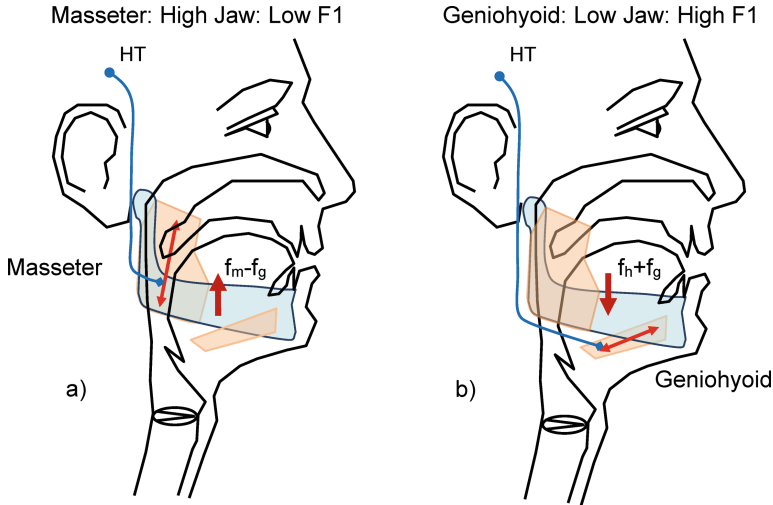


Fig. 1. Agonist-antagonist neuromotor actions regulating jaw position and the vowel feature open/close. (a) The neuromotor action promoted from hypothalamus neurons (HT) activate the corresponding trigeminal branch of the masseter. The result is a force (f_m) acting against gravity (f_g) moving the jaw upwards. (b) The neuromotor action activating the geniohyoid muscle produces a force (f_h) in the sense of gravity (f_g) pulling the jaw downwards.

For such, a certain jaw-tongue dynamic reference point (JTDRP) in the biomechanical system will be considered (equivalent center of actions), which could be related to formant positions. When its coordinates (x_A , y_A) experience a change, a corresponding change in the first two formants (F_1 , F_2) can be expected. Lowering the JTDRP will result in an elevation of F_1 , and vice-versa. Similarly, advancing the JTDRP will result in an elevation of F_2 , and vice-versa. The back-front (B/F) and open-close (O/C) features may be justified on this basis. This relationship can be established as:

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

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,

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 possible articulation positions associated to a given formant pair [4]. The first time derivative of (1) allows associating formant dynamics with the JTDRP kinematics as:

$$\begin{bmatrix} \frac{dF_1(t)}{dt} \\ \frac{dF_2(t)}{dt} \end{bmatrix} = \mathbf{A} \begin{bmatrix} \hat{v}_x(t) \\ \hat{v}_y(t) \end{bmatrix} \quad (2)$$

where it has been assumed that the system given by matrix \mathbf{A} is linear and time-invariant, and \hat{v}_x and \hat{v}_y are the B/F and O/C velocity estimates of the JTDRP. Extending the biomechanical chain one step more, the derivatives of the velocity estimates will allow evaluating the accelerations experienced at the JTDRP:

$$\begin{bmatrix} \frac{d^2F_1(t)}{dt^2} \\ \frac{d^2F_2(t)}{dt^2} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \frac{d\hat{v}_x(t)}{dt} \\ \frac{d\hat{v}_y(t)}{dt} \end{bmatrix} = \mathbf{A} \begin{bmatrix} \hat{a}_x(t) \\ \hat{a}_y(t) \end{bmatrix} \quad (3)$$

In what follows it will be assumed that the contribution to the first formant kinematics is mainly the result of vertical dynamics, as expressed by the following relation:

$$\frac{dF_1(t)}{dt} \cong a_{12}\hat{v}_y(t) \quad (4)$$

The time derivative of F_1 can be evaluated from Linear Predictive Spectral Estimation [5], whereas the jaw movement can be obtained from accelerometers [6]. In the present work, a different approach will be followed, where biomechanical dynamic variables are to be inferred from the surface myoelectric activity recorded on skin covering the facial surface over masseter [7], as illustrated in the next section.

3 Materials and Methods

The present study has a marked exploratory nature, as very few publications address the issue of using the myoelectric facial surface signals on the masseter to estimate neuromotor decay in neurological disease evaluation [7]. 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 fibre contractions under the neuromotor commands travelling the corresponding branch of the trigeminal nerve. The masseter is one of the most powerful body muscles, 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 sequence /*ayayayayay...*/ (phonetically [*ajajajajaj...*]) during four sequences, because it



Fig. 2. Recording the myoelectric surface signal produced by the contraction of the masseter.

implies an intensive masseter activity. The acoustic signal was recorded with a Sennheiser microphone at 44100 Hz and 16 bits. The electromyographic signal was recorded with the equipment Biopac MP150 EMG100 at 2000 Hz and 16 bits. The fixture to record surface myoelectric signals from the masseter is shown in Fig. 2. Typically, 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 cheekbone, and a third reference electrode is placed in the forefront over the ipsilateral eyebrow. This fixture showed high signal and low noise levels. A 10-order Butterworth low-pass filter at 20 Hz cut-off frequency was applied on the resulting myoelectric signal for de-noising and artefact removing.

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

- Myoelectric surface signals are recorded simultaneously with voice signals. Synchronization is not a strict requirement, although the subject is asked to produce a sharp click with the tongue, which is recorded both as a myoelectric and an acoustic event (see the star mark on plots (a) and (d) in Fig. 3).
- The voice recordings are undersampled to 8 KHz.
- A ten-cycle segment of voice for the diphthong [aj] signalled by the vertical arrows in Fig. 3b is selected for formant kinematic estimation.
- The ONPT transfer function of the voice segment is evaluated by an 8-pole adaptive inverse LP filter [5] 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 of the LP predictor each 2 ms with a frequency resolution of 2 Hz. The derivative of the first formant is given in Fig. 3c.
- The myoelectric signal is low-pass filtered with a 10-order Butterworth at a cut-off. Frequency of 20 Hz, producing as a result the signal given in Fig. 3d.

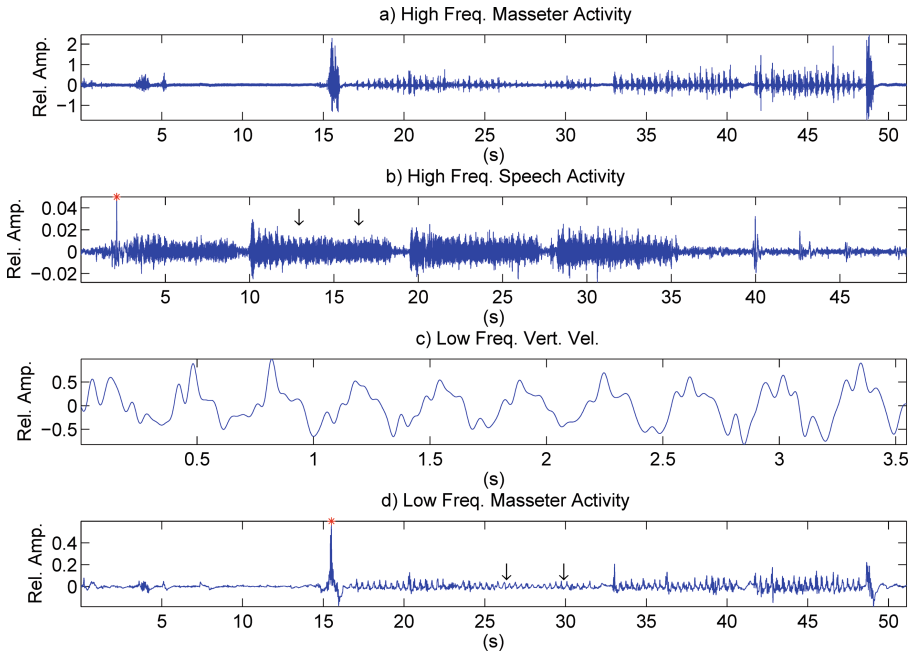


Fig. 3. Signals recorded for the study: (a) myoelectric surface signal; (b) acoustic signal showing the ten-cycle segment used in the estimation of the kinematic acoustic signal (vertical arrows); (c) ten-cycle segment of the first formant time derivative (kinematic acoustic signal); (d) low-pass filtered myoelectric surface signal, showing the segment that correlates best to the kinematic acoustic signal (vertical arrows). The stars in (b) and (d) show the alignment events.

- The derivative of the first formant as given in Fig. 3c is correlated along the low-pass filtered myoelectric signal given in Fig. 3d, producing a best match for the segment signalled by vertical arrows in (b) and (d).

The signal segments which best match under a correlation criterion are shown in Fig. 4. Each correspond to a ten [aj]-cycle segment. The upper template (a) shows the low-pass filtered time derivative of F_1 , whereas the mid-template (b) shows the low-pass filtered surface myoelectric signal. They match almost exactly as far as their pseudo-periods are concerned. It may be seen that the myoelectric signal (b) has more contents of higher harmonics than (a). There are several possible explanations for this fact, the possible influence of a further low-pass effect contributed by the jaw-tongue biomechanical system being among them.

The bottom template (c) is the result of estimating the least-square error between (a) and (b). It may be seen that this error concentrates most of the high-frequency components of (b). The details of the correlation and best match process are given in the next section.

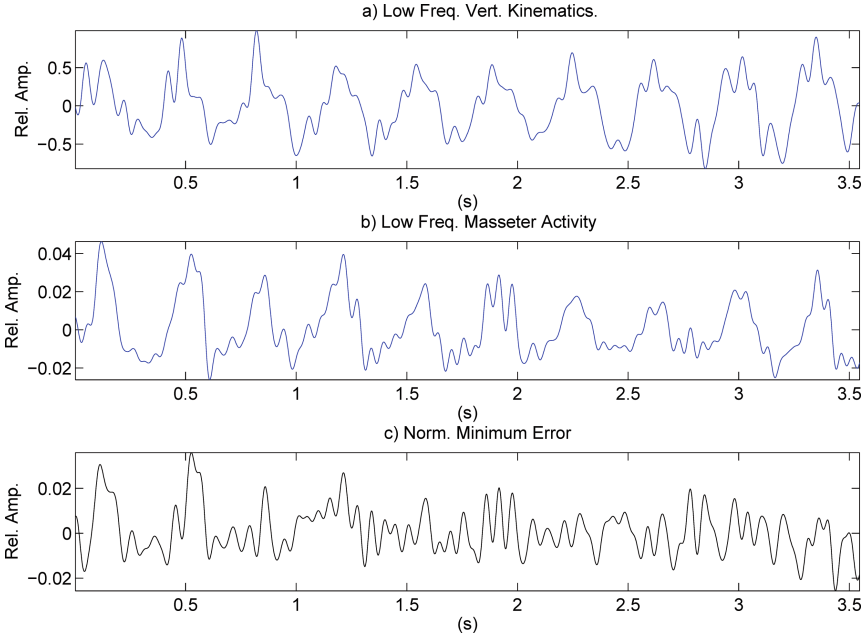


Fig. 4. Segments selected for correlation: (a) first formant time derivative (kinematic acoustic signal); (b) low-pass filtered myoelectric surface signal (best fit); (c) normalized minimum error between the acoustic kinematic and the myoelectric signals.

4 Results and Discussion

The correlation process consisted in sliding the time derivative of the first formant (kinematic acoustic signal shown in Fig. 3c), aligning it over the low-pass filtered surface myoelectric signal (Fig. 3d), and for each alignment step the correlation coefficients of Pearson and Spearman were evaluated. The alignment producing the largest Spearman's coefficient was marked as the best fit, and the corresponding segment of the myoelectric signal was selected:

$$t_{align} = \arg \max_{\zeta \in W} \left\{ \rho_S \left[s_m(t + \zeta), \frac{dF_1(t)}{dt} \right] \right\} \quad (5)$$

where W is the recording window, $s_m(t)$ is the low-pass filtered surface electromyogram on the masseter facial skin, and $\rho_S[s_1, s_2]$ is the function evaluating Spearman's correlation coefficient between two time series $s_1(t)$ and $s_2(t)$. The best-fit segments are plotted in Fig. 4a and b. The linear and nonlinear regression measurements are given in the next section (Fig. 5).

The linear regression corresponds to a Pearson's coefficient of 0.735, with an estimation confidence given by an almost null p-value (p-v: 0.000). The nonlinear regression corresponds to a Spearman's rank-coefficient of 0.776, with an estimation confidence almost null (p-v: 0.000). At this point it would be interesting

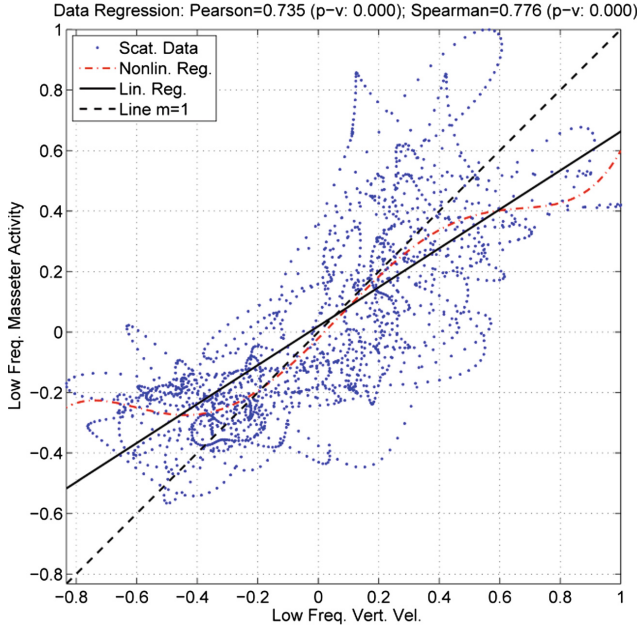


Fig. 5. Regressions between the acoustic kinematic and the myoelectric signals. The scatter plot of the dataset is given in blue dots. The full straight line gives the linear regression fit. The dash-dot red line shows a 5-th order polynomial nonlinear fit. The dash line slope is the unity. (Color figure online)

to analyze other alternatives, for instance, it should be reasonable to consider that being the myoelectric signal related with neuromotor actions, perhaps a better fit could be obtained on the second time derivative of the first formant (related to accelerations, and thus to forces). When this signal is used for the alignment with the same acoustic kinematic segment, the results are less relevant (Pearson's and Spearman's coefficients of 0.464, respectively). Therefore, it can be concluded, as far as the present experiment is concerned, that the logical association of the surface myoelectrical signal is more correlated with the acoustic kinematic signal than with its time derivative. This conclusion is of large relevance for the diagnose 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 JTDRP as given by (2), a possible correlate to neuromotor disease grade could be defined by the absolute velocity of the system, given as:

$$|v_{JTDRP}(t)| = \sqrt{\left(w_{21} \frac{dF_1(t)}{dt}\right)^2 + \left(w_{12} \frac{dF_2(t)}{dt}\right)^2} \quad (6)$$

where w_{12} and w_{21} are the coefficients of the inverse of matrix \mathbf{A} , assuming that invertibility conditions fulfill, and that vertical kinematics is expressed only in

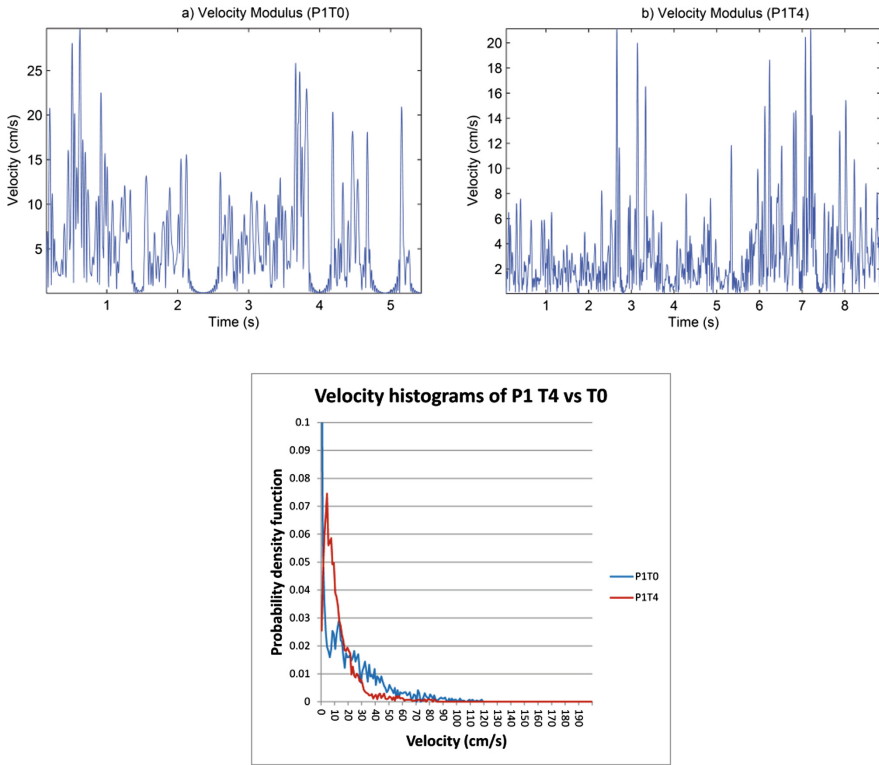


Fig. 6. Absolute velocity of the JTDRP for two recordings from an ALS patient: (a) First recording at month 0; (b) Last recording at month 12; (c) Comparison of the two probability density functions (blue-P1T0: month 0; red-P1-T4: month 12). (Color figure online)

F_1 , and horizontal kinematics only in F_2 . No matter how strong these conditions may seem, there are indications that this can be a plausible hypothesis, to be tested in a future study. The absolute velocity of the JTDRP given in (6) is a relevant correlate to articulation dysarthria and dysfluency, and can be used in comparing speech features from different utterances. In what follows the results of evaluating the degradation of the speaking performance of a 64-year old female patient suffering from Amyotrophic Lateral Sclerosis (ALS) [8] will be shown. The patient was recorded five times with a 3-month interval separation, uttering a pre-established sentence. The absolute velocity from each utterance was estimated and a probability density function was evaluated by Kolmogorov-Smirnov histogram fitting [9]. The velocity profiles for the first and last recording sessions, and the comparison of the relative probability density functions are given in Fig. 6. It may be seen that in the first recording (Fig. 6a) the speaker spent 5.3s, whereas in the last one (Fig. 6b) the same sentence was uttered in almost 9 s. Besides, pauses between phonated segments can be appreciated in the

first recording, whereas in the last one, such pauses were not found. The velocity activity is better organized in the first recording and it is distributed in a wider range of speed values (in cm/s), whereas the last recording shows smaller speed values. This can also be appreciated in the probability density functions given in Fig. 6c, where the first recording velocity distribution shows a gentle decay, whereas the last one shows a reduction in large velocities, which is transferred to low velocities (sharper decay). Based on these probability density functions, the Kullback-Leibler Divergence [10] was used to estimate the distance from these distributions with respect to a control subject. The results are given in Table 1.

Table 1. KLD for the first and last recordings relative to a female control subject.

Recording	KLD to control
P1T0	0.42657477
P1T4	0.93337742

It may be seen that the divergence between the first recording and the last recording is almost as large as the divergence from the first recording to the control subject, thus showing the degradation of the patient's condition regarding speech production.

5 Conclusions

Speech articulation is conditioned by the movements produced by well determined 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. A relevant question is how far speech correlates and neuromotor action are related. This question was to be answered by the present study using electromyographic recordings on the masseter and the acoustic kinematics related with the first formant. Although the study presented is limited to one case and one sequence, there are interesting findings to be reported, among them the following:

- A clear association between the electromyographic signal and the acoustic kinematic one may be established. Correlation measures based in second order statistics satisfy the necessary conditions for this relation to be confirmed.
- The variable to which the electromyogram fits better is the time derivative of the first formant, which may be associated to the velocity of the JTDRP.
- The main difference between both the electromyographic signal and the acoustic kinematic one seems to be due to a larger contents of harmonics present in the electromyogram. This could indicate that the jaw-tongue bio-mechanical system is acting as a low-pass electromechanical filter.

- The absolute velocity associated to the JTDRP estimated as the time derivatives of the first and second formants seems to be a good index to disfluency and disarthria of neuromotor origin.
- These conclusions are to be validated on a wider database of speakers, including gender dependence.

Acknowledgments. This work is being funded by grants TEC2012-38630-C04-01, TEC2012-38630-C04-04 and TEC2016-77791-C4-4-R from the Ministry of Economic Affairs and Competitiveness of Spain.

References

1. Demonet, J.F., Thierry, G., Cardebat, D.: Renewal of the neurophysiology of language: functional neuroimaging. *Physiol. Rev.* **85**, 49–95 (2005)
2. Yunusova, Y., Weismer, G., Westbury, J.R., Lindstrom, M.J.: Articulatory movements during vowels in speakers with dysarthria and healthy controls. *J. Speech Lang. Hear. Res.* **51**(3), 596–611 (2008)
3. Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F., Laurillau, Y.: EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert Syst. Appl.* **40**, 4832–4840 (2013)
4. Dromey, C., Jang, G.O., Hollis, K.: Assessing correlations between lingual movements and formants. *Speech Commun.* **55**(2), 315–328 (2013)
5. Deller, J.R., Proakis, J.G., Hansen, J.H.L.: *Discrete-Time Processing of Speech Signals*. Macmillan, New York (1993)
6. Green, J.R.: Mouth matters: scientific and clinical applications of speech movement analysis. *Perspect. Speech Sci. Orofacial Disord.* **25**, 6–16 (2015)
7. Wu, P., Gonzalez, I., Patsis, G., Jiang, D., Sahli, H., Kerckhofs, E., Vandekerckhove, M.: Objectifying facial expressivity assessment of Parkinson’s patients: preliminary study. *Comput. Math. Methods Med.* **2014** (2014). Article no. 427826. <http://dx.doi.org/10.1155/2014/427826>
8. Gómez-Vilda, P., Londral, A.R.M., Rodellar-Biarge, V., Ferrández-Vicente, J.M., de Carvalho, M.: Monitoring amyotrophic lateral sclerosis by biomechanical modeling of speech production. *Neurocomputing* **151**, 130–138 (2015)
9. Webb, A.R.: *Statistical Pattern Recognition*. Wiley, Chichester (2002)
10. Salicrú, M., Morales, D., Menéndez, M.L., Pardo, L.: On the applications of divergence type measures in testing statistical hypotheses. *J. Multivar. Anal.* **51**, 372–391 (1994)