Biomarkers of Neurodegenerative Progression from Spontaneous Speech Recorded in Mobile Devices: An Approach based on Articulation Speed Estimation

A Study of Patients Suffering from Amyotrophic Lateral Sclerosis

Ana Londral[1,2], Pedro Gómez-Vilda[3] and Andrés Gómez-Rodellar[3]

1Translational Clinical Physiology Lab, Instituto de Medicina Molecular, University of Lisbon, Lisbon, Portugal

2Escola Superior de Tecnologia de Setúbal, Instituto Politécnico de Setúbal, Portugal

3Neuromorphic Speech Processing Lab, Center for Biomedical Technology, Universidad Politécnica de Madrid, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain

Keywords: Mobile Devices, Clinical Remote Monitoring, Speech Signal Processing, Amyotrophic Lateral Sclerosis.

Abstract: A majority of patients with Amyotrophic Lateral Sclerosis (ALS) experiment a rapid evolution of symptoms related to a progressive decline in movement function that affects different systems. Clinical assessment is based on measures of progression for identifying the need and the pace of medical decisions, and to measure also the effects of novel therapies. But assessment is limited to the periodicity of clinical appointments that are increasingly difficult for patients due to progressive mobility impairments. In this paper, we present a novel method to assess neurodegeneration process through speech analysis. An articulation kinematic model is proposed to identify markers of neuromotor functional progression in speech. We analysed speech samples that were collected with a mobile device, in 3-month intervals, from a group of six subjects with ALS. Results suggest that the method proposed is sensitive to the symptoms of the disease, as rated by observational clinical scales, and it may contribute to assist clinicians and researchers with better and continuous measures of disease progression.

1 INTRODUCTION

Motor Neuron Disease (MND) is a progressive neurological condition that affects the motor neurons, present both in the central and spinal neural systems. The most common MND is Amyotrophic Lateral Sclerosis (ALS), which is a disease with rapid progression and unknown cure, affecting both upper and lower motor neurons. The deterioration of the neuromotor system involved in respiration, phonation, swallowing, and lingual and oro-facial muscle function degenerates in a rapidly progressing speech dysarthria (Tomik and Guiloff, 2010).

The clinical support of this disease is based on the management of the symptoms, as they manifest (Andersen et al., 2012). Clinical scales for monitoring progression are either invasive, based on EMG (de Carvalho et al. 2005) or on observational evaluation and Likert-type functional scales (Cedarbaum & Stambler, 1997). Assessments are performed in 3-6 month spaced clinical appointments; frequently, when the progression of symptoms is severe, the periodicity of clinical appointments decreases due to difficulties in transportation of the patient from their residence to clinical facilities.

As speech intelligibility decreases, patients often use mobile devices (often a tablet or a smartphone) with text-to-speech to support communication (Londral et al. 2015). The aim of this work is to explore the potential of those mobile devices to continuously and remotely monitor ALS progression through speech collection, by evaluating quantitative speech parameters, using a methodology as depicted in Figure 1.

In this study, we explore speech as a biomarker of disease progression. Voice is a signal that is easily collected with non-invasive and low-cost techniques. In fact, modern mobile devices allow to collect speech with integrated *app*s that run from patient's home, for remote monitoring using e-

Londral, A., Vilda, P. and Gómez-Rodellar, A.
Biomarkers of Neurodegenerative Progression from Spontaneous Speech Recorded in Mobile Devices: An Approach based on Articulation Speed Estimation - A Study of Patients Suffering from Amyotrophic Lateral Sclerosis.
DOI: 10.5220/0006731302690275
In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Volume 4: BIOSIGNALS*, pages 269-275
ISBN: 978-989-758-279-0

Health platforms, as are example the works from Abad et al. (2013) and Vacher et al. (2006).

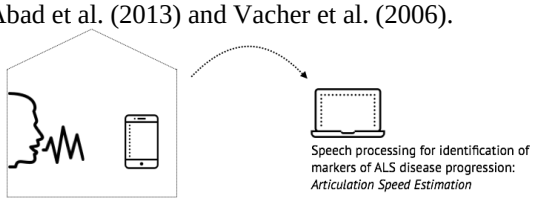

Figure 1: A scheme of the system that we aim to implement, based on speech collection from home, using regular mobile devices, and remote speech processing.

This paper firstly describes the Articulation Kinematics Model (AKM) that is proposed, then in section 3, the methodology of this study is presented, and the dataset of ALS patients is described. Results are presented in section 4; the final section includes the conclusion with the discussion of results and future work.

2 SYMPTOMS OF NEURO--MOTOR DEGENERATION IN SPEECH

Progressive dysarthria is a symptom in ALS. Dysarthria is a disorder that results from neurological impairment of the motor component of the motor-speech system. The neurological origin of dysarthria may vary in ALS but ultimately it affects the intelligibility of patient's speech causing great difficulties in communication.

It has been studied that, as ALS progresses, speech movements become smaller in extent and slower in speed (Green et al., 2013). Classical articulation measures define the vowel space area (VSA) and the Formant Centralization Ratio (FCR) as valid parameters to estimate the vowel span range and positioning produced by a given speaker (Sapir et al. 2011). Absolute span of formants F1 and F2 of a given utterance has been additionally proposed as a sensible feature to dysarthria in a longitudinal study with five persons with ALS (Gómez-Vilda et al., 2015).

While those parameters are known to be sensitive to the assessment of dysarthria, its semantic meaning is unclear. Besides, these features only express the average values of the most frequent formant positions, mainly associated to vowels. As dysarthrias with neuromotor origin express changes in dynamic activity of the articulation organs (imprinted in rapid formant changes), we used a method based on the measure of the kinematic behavior of formant dynamics, described in the following section.

3 ARTICULATION KINEMATIC MODEL (AKM)

Estimated parameters in common speech are associated to specific neuromuscular complexes involved in articulation, more specifically the masseter, the stylo-glosus and the genio-hyo-glosus muscles – as described in (Gómez-vilda et al., 2013).

The model presented in this paper allows the estimation of the articulation positions based on the indirect inference of vocal tract configuration using the simplified model depicted in Fig. 2.

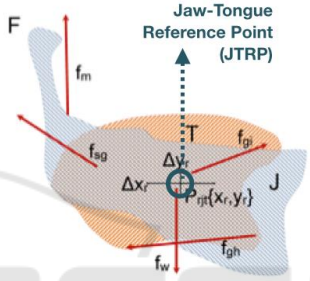

Figure 2: Articulation kinematic model used in this study. A dynamic system of muscle force vectors representing the motor articulation system can be simplified in one reference-point (JTRP), which is situated in the action centre of the oral cavity and moves in the sagittal plane, that expresses the dynamics of articulation.

The AKM includes the jaw, the tongue and the facial tissues attached to them in a dynamic system that can be approximated to a third-order lever fixed at the skull. Considering only movements in the sagittal plane, we define the *Jaw-Tongue Reference Point* (JTRP) Prjt $\{x_r,y_r\}$, where different forces act during speech, related to the neuromuscular system involved in the masseter, lips and tongue movements, as described in Gomez et al. (2017). As a result of multiple muscle forces, the reference point JTRP will move in the sagittal plane (Δx_r, Δy_r); these movements are related to formant changes as represented in Equation 1, where a_{ij} are nonlinear time-variant and multi-valued functions associating Prjt to formants, and *t* is the time.

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

3.1 Absolute Kinematic Velocity

Under the assumption that formant F_1 is related to vertical kinematics and formant F_2 is related to horizontal kinematics, the articulation kinematic velocity (AKV) may be inferred by Equation 2.

$$|v_r(t)| = \sqrt{\left(w_{21}\frac{\partial F_1(t)}{\partial t}\right)^2} + \sqrt{\left(w_{12}\frac{\partial F_2(t)}{\partial t}\right)^2} \quad (2)$$

The expected AKV profile would be that of a decaying curve, with the properties of a χ^2 distribution, illustrating the dynamics of articulation behavior.

3.2 Kullback-Leibler's Divergence (KLD)

As the disease progresses, the dynamic behavior of patient's speech is expected to be hampered by the difficulty in moving the articulation organs with enough speed. Dysarthria will limit the movements and the decaying curve of AKV probability density function (pdf) will be more marked, with higher probabilities for lower velocities and lower probabilities for higher velocities. As described by P.Gomez et al. (2017), we will use the Kullback-Leibler's Divergence (KLD) to model the dynamic behavior differences between healthy controls and ALS patients, according to Equation 3:

$$D_{KLij}\{p_{Ti}(v_r), p_{Mj}(v_r)\} = \int_{z=0}^{\infty} p_{Mi}(z)\,\mathrm{abs}\left\{\log\left[\frac{p_{Ti}(z)}{p_{Mj}(z)}\right]\right\} dz \quad (3)$$

where v_r is the absolute value of the articulation kinematic velocity, p_{Ti} is the probability density distribution of the target utterance T_i, and p_{Mj} is the probability density distribution of the model utterance T_j. Figure 3 depicts the comparison between the probability density functions (pdf) of a 'unhealthy' and a 'healthy' speech sample.

Figure 3. The model based on the AKV pdf to model the dynamic behavior differences between healthy controls (diamonds) and ALS patients (squares).

4 METHODOLOGY

In this work we applied the AKM to a dataset of sound samples from eight subjects with ALS that recorded speech in 2 to 5 assessments performed in periods of 2 to 5 month-intervals during 2 to 19 months (as described in Table A1).

4.1 Dataset

We used data from 8 women with ALS, saved in a dataset of voice samples collected for a longitudinal study that was approved by the Ethical Commission in the Hospital of Santa Maria, Lisbon, Portugal. All participants signed an informed consent to be included in the study. Speech was recorded using the microphone of a laptop or a smartphone (2-channel wav files with sample rate of 44100Hz and 16 bits). All files have the same sentence recorded by the patients in consecutive assessments.

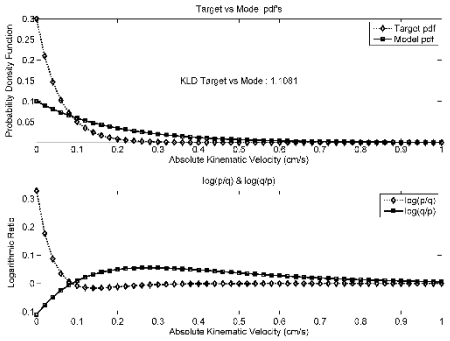

Figure 3: The model based on the AKV pdf to model the dynamic behavior differences between healthy controls (diamonds) and ALS patients (squares).

The mean age of the 8 subjects is 66.7 (±13.0) years old, the youngest with 38 and the oldest with 80 years old.

Patients were asked to repeat once, a popular sentence from the Portuguese writer Fernando Pessoa, which happens to be very well known by most people in Portugal: /tudo vale a pena quando a alma não é pequena/. This sentence was recorded from the patient in 2 to 5 assessments that were taken in successive clinical appointments, as described in Table A1. The same recording was taken from two female healthy controls of 36 and 63-years old (CF36 and CF63).

4.2 Procedure

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 8-pole adaptive inverse Linear Prediction (LP) filter (Deller et al., 2000) with a low-memory adaptive step to grasp fine time variations.
- The first two formants are estimated by evaluating the maxima and slenderness of the

LP spectrogram. The formant estimation resolution used is 2 Hz every 2 ms.
- The derivatives of the first two formants are used to estimate the absolute velocity of the JTRP following (2).
- The values of the AKV in the recording interval are used to build a histogram.
- The histograms are used to estimate probability density functions by Kolmogorov-Smirnov approximations (Webb, 2003).
- Kullback-Leibler's Divergence between each patient's histogram-derived distribution vs that of the control subject is estimated as by (3).

5 RESULTS

The AKV was calculated for all the sound files of each subject. Figure 4 represents the velocity that was dynamically calculated for a 3.5 seconds sample from the younger control and the respective histogram of the AKV. From this figure, it is possible to observe that articulation velocity is zero in pauses between words and has a maximum that is approx. 45 cm/s.

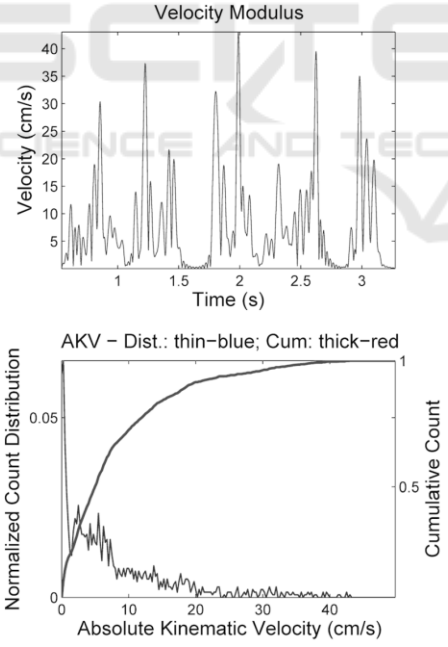

Figure 4: (up) the velocity of JTRP calculated dynamically for a sound sample with 3.5 seconds. (down) the histogram of the AKV with cumulative count (darker line).

The comparison between the average distributions of the ALS samples and the model probability from older control subject are depicted in Figure 5, with the example of subject HA. In this example, the Liljefors tests discard gaussianity ($p<0.05$); Kolmogorov-Smirnov (KS) and Wilcoxon (WX) reject the null hypothesis (H0) of similarity ($p_{wx}<0.05$) between Targets and Models. The average Kullback-Leibler distance is 0.483, the 96% of the cases reject H0 with respect to the model (CF63) under KS, and 88% reject H0 under Mann-Whitney test.

When comparing the same subject with the younger control, the similarity in the average is not rejected according to WX test, and only 76% and 52% of the files reject similarity to the model under KS and MW, respectively.

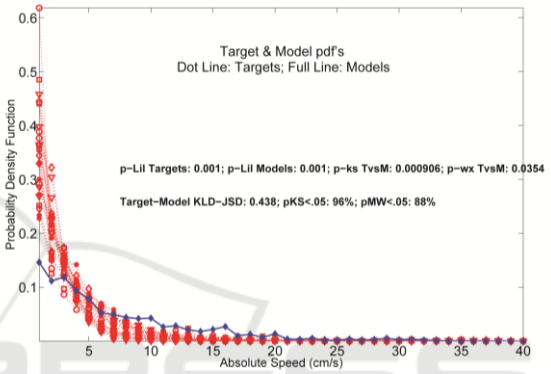

Figure 5: Comparison between the average distributions of the subject HA samples (dotted lines) and the model probability from the older control subject (full line). At the middle of the plot, the two text lines indicate results from tests of gaussianity of the distributions following Liljefors test (first text line, Targets and Models), and the Kolmogorov-Smirnov and Wilcoxon tests of Targets vs Models for the null hypothesis of similarity (H0); the Kullback-Leibler distance of the average Targets and Models, and the percentage of Targets rejecting H0 between Targets and Models according to the Kolmogorov-Smirnov and Mann-Whitney tests (second text line).

The KLD and the LLR distance were calculated between the two control subjects and all the subjects with ALS are described in Table 1 (Appendix). From Figure 6, it is clear that we can observe the disease progression within assessments, for all the subjects. It is also clear from Figure 5 and Table 1, that in some assessments, there seems to be a regression in the symptoms of neurodegeneration, as the KLD and LLR values decrease. When observing this behaviour and comparing with respective ALSFRS-B values, we can observe that these decreases are related to observed stabilization of symptoms by the doctor (the value of ALSFRS-B remains the same of the previous assessment). Subject AR, the youngest patient with ALS (38

years old), is the exception (values of KLD decrease, despite the observed symptoms were rated with lower values of speech functionality).

The LLR distance to the younger control subject had the best results for demonstrating unceasing

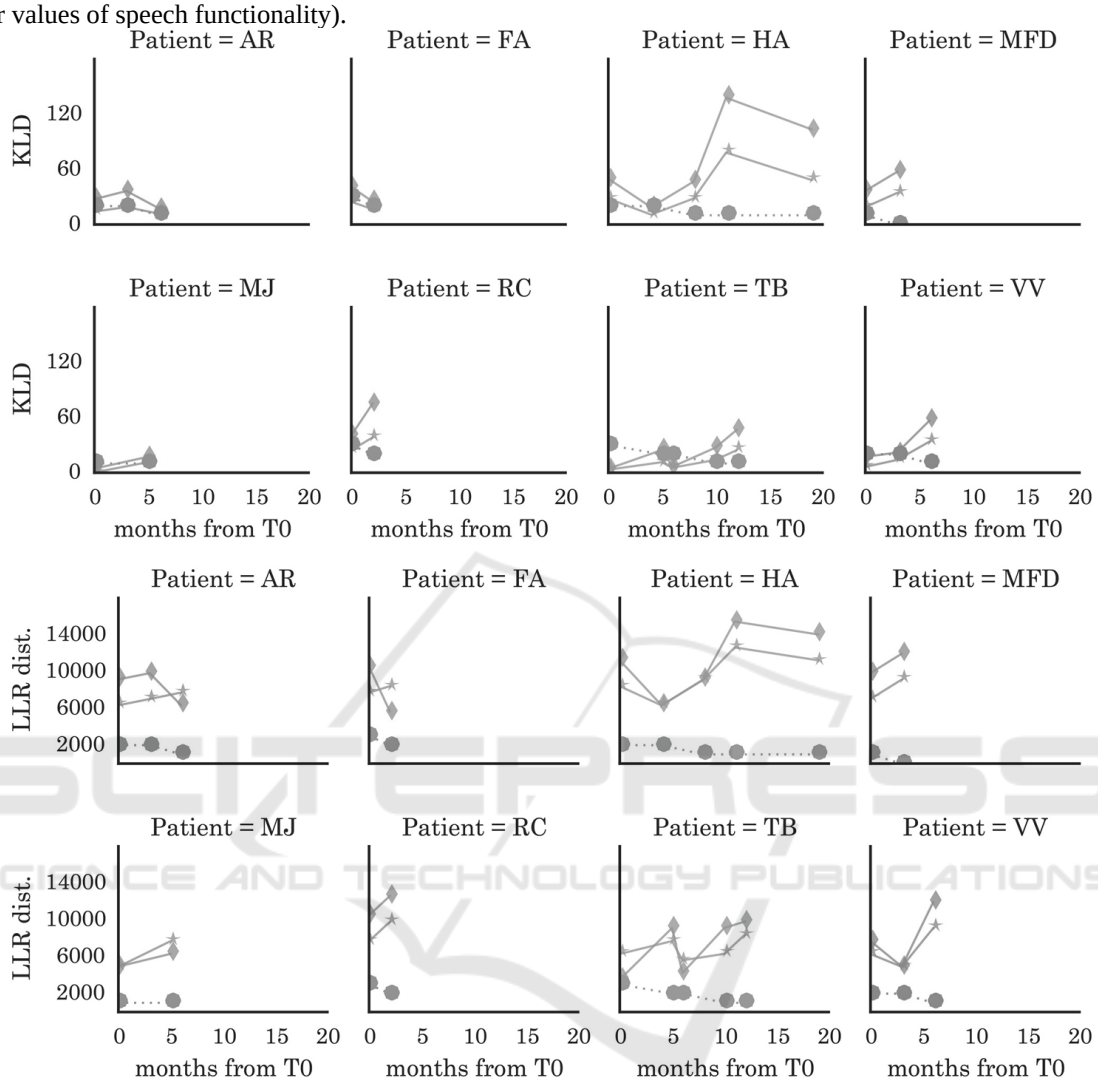

Figure 6: These plots represent the KLD and LLR distance of the 8 subjects with ALS from the younger control subject (marker *) and older control (marker ◆). Bullet points indicate the ALSFRS-B evaluation. The x-axis represents the months from the first assessment (T0).

neurodegeneration in ALS, since LLR increases along time for all the subjects with ALS (exceptions are coincident to stability of observed symptoms).

6 DISCUSSION

In this paper, we propose studying the AKV as a marker of neurodegenerative progression in ALS. We are interested in evaluating current speech that can be recorded via a current mobile device to be used in remote monitoring, outside the clinical facilities. In particular for ALS, for which standard clinical assessments are spaced of 3 to 6 months, this would facilitate:

1. A continuous evaluation of progression that could be sent to the clinician.
2. Novel markers to study the impact of new therapies.
3. The remote assessment, in particular when patient has mobility impairments and the transportation for the clinical facilities becomes difficult.

We used a database containing sound samples that were recorded from a mobile phone or a laptop computer. These samples contain the same common 20-words sentence in Portuguese. For the severity of ALS disease, a short sentence is an important requisite of our methodology, to make it valid for these patients' context. In fact, as speech becomes difficult to produce, the more complex is the sample collection, and the more dropouts will take place.

The results described in previous sections demonstrate that the KLD from a healthy control is sensitive to neurodegeneration progression in ALS. For all ALS subjects, except one, the AKV model expressed progression of neurodegenerative symptoms in speech, by increase of KLD, for both the younger and the older models. The exception was observed for the youngest subject with ALS (38 years old). In fact, we can hypothesize that the samples from this patient will fit better to the younger control model, due to the age proximities. The LLR distance for the younger model expresses a continuous increase and correlation to the ALSFRS-B values attributed to this subject.

In general, the results described in this work confirm the hypothesis that we can model the speech dysarthria in ALS as a "freezing" of the articulation process: the probability of AKV close to 0 increases as diseases progresses.

Our objective and quantitative measures are according to the qualitatively assessed clinical rating for bulbar involvement that is based on clinical experienced observation. The apparent regression in neurodegeneration from our results can be confirmed by the experienced clinical observation of stabilization of symptoms within assessments. But, our quantitative measure may have implicit information that is not observable and needs further insight on its meaning. By hypothesis, our measures may be sensible to different therapies that cause variations observed within assessments. A new dataset of samples that are collected in shorter time intervals, from home mobile devices, is needed to obtain a continuous observation of articulation measures. A continuous observation will support a novel insight on progression behaviour.

Our study has some limitations. One is the heterogeneity of our sample, since some subjects have 2 and others have 5 assessments. For this reason, it is not possible to have a solid demonstration of the progression along time. Another limitation is that we are using samples containing the same phonetic material. In all samples, subjects use the same sentence.

For future work, a larger database containing spontaneous speech from subjects with ALS will be used to further test our model and study progression from symptoms measured in speech signal.

ACKNOWLEDGEMENTS

This work was supported by Calouste Gulbenkian Foundation and the Portuguese Association of ALS (APELA), as well as by grant TEC2016-77791-C4-4-R (Plan Nacional de I+D+i, Ministry of Economic Affairs and Competitiveness of Spain).

REFERENCES

Abad, A. et al., 2013. Automatic word naming recognition for an on-line aphasia treatment system. *Computer Speech and Language*, 27(6), pp.1235–1248.

Andersen, P.M. et al., 2012. EFNS guidelines on the Clinical Management of Amyotrophic Lateral Sclerosis (MALS) - revised report of an EFNS task force. *European Journal of Neurology*, 19(3), pp.360–375.

de Carvalho, M., Costa, J. & Swash, M., 2005. Clinical trials in ALS: a review of the role of clinical and neurophysiological measurements. *Amyotrophic Lateral Sclerosis & Other Motor Neuron Disorders*, 6(4), pp.202–212. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=a2h&AN=19020138&loginpage=Login.asp&site=ehost-live.

Cedarbaum, J.M. & Stambler, N., 1997. Performance of the amyotrophic lateral sclerosis functional rating scale (ALSFRS) in multicenter clinical trials. In *Journal of the Neurological Sciences*.

Deller, J.R., Proakis, J.G. & Hansen, J.H.L., 2000. *Discrete-Time Processing of Speech Signals*, Available at: http://www.library.wisc.edu/selectedtocs/bd429.pdf

Gómez-vilda, P. et al., 2013. Characterization of Speech from Amyotrophic Lateral Sclerosis by Neuromorphic Processing. In *IWINAC 2013*. pp. 212–224.

Gómez-Vilda, P. et al., 2015. Monitoring amyotrophic lateral sclerosis by biomechanical modeling of speech production. *Neurocomputing*, 151(1), pp.130–138.

Londral, A. et al., 2015. Quality of life in ALS patients and caregivers: impact of assistive communication from early stages. *Muscle and Nerve*.

P.Gomez, P. et al., 2017. Articulation acoustic kinematics in ALS speech. In *2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI)*. IEEE, pp. 1–6. Available at: http://ieeexplore.ieee.org/document/7985522/.

Sapir, S. et al., 2011. Acoustic metrics of vowel articulation in Parkinson's Disease: Vowel Space Area (VSA) vs. vowel articulation index (VAI). In C. Manfredi, ed. *Proceedings of the MAVEBA 2011*. Florence: Florence University Press, pp. 173–175.

Tomik, B. & Guiloff, R.J., 2010. Dysarthria in amyotrophic lateral sclerosis: A review. *Amyotrophic Lateral Sclerosis*, 11(1–2), pp.4–15.

Vacher, M. et al., 2006. *Speech and sound use in a remote monitoring system for health care*,
Webb, A.R., 2003. *Statistical pattern recognition*,

APPENDIX

Table A1: Description of the results of speech features extracted from all the assessments of the 8 subjects with ALS, considered in this study.

Patient	**FA**		**MFD**		**RC**		**MJ**	
Age	73		68		64		56	
#assessment	0	1	0	1	0	1	0	1
ALSFRS-B	3	2	1	0	3	2	1	1
Months from T0	0	2	0	3	0	2	0	5
Divergence_CF63	40.88	24.76	36.51	57.44	40.65	75.43	4.45	17.30
LLRDistance_CF63	10524.84	5641.64	9837.10	11949.03	10518.08	12625.50	4945.46	6325.12
Divergence_CF36	34.45	33.86	18.89	18.85	13.98	13.90	16.62	14.13
LLRDistance_CF36	9138.28	9118.03	7020.08	7026.55	4916.18	6324.33	8415.68	6337.03
Patient	**HA**					**AR**		
Age	68					38		
#assessment	0	1	2	3	4	0	1	2
ALSFRS-B	2	2	1	1	1	2	2	1
Months from T0	0	4	8	11	19	0	3	6
Divergence_CF63	49.77	18.74	46.92	137.96	103.39	27.50	36.44	17.78
LLRDistance_CF63	11224.19	6334.85	9137.77	15437.21	14033.18	9147.58	9830.65	6340.91
Divergence_CF36	28.44	28.21	24.65	24.57	24.50	78.30	48.36	38.49
LLRDistance_CF36	9120.44	8413.34	7714.18	8428.58	7708.48	12626.64	11222.63	9814.93
Patient	**TB**					**VV**		
Age	77					80		
#assessment	0	1	2	3	4	0	1	2
ALSFRS-B	3	2	2	1	1	2	2	1
Months from T0	0	5	6	10	12	0	3	6
Divergence_CF63	4.21	24.70	6.52	27.18	46.60	17.22	22.02	56.69
LLRDistance_CF63	3526.56	9126.58	4239.57	9134.90	9841.94	7720.34	4933.06	11928.92
Divergence_CF36	11.85	11.43	11.43	10.72	6.32	5.46	3.27	0.37
LLRDistance_CF36	7719.38	7712.00	7709.37	6319.77	6308.00	5621.47	6312.70	4925.75