Characterization of Speech in Amyotrophic Lateral Sclerosis by Neuromorphic Processing

Pedro Gómez-Vilda¹, Ana Rita M. Londral², Mamede de Carvalho², José Manuel Ferrández-Vicente³, Victoria Rodellar-Biarge¹

¹NeuVox Laboratory, Center for Biomedical Technology, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28223 Pozuelo de Alarcón, Madrid, Spain
²Instituto de Medicina Molecular, Faculty of Medicine, University of Lisbon, Lisbon, Portugal
³Universidad Politécnica de Cartagena, Campus Universitario Muralla del Mar, Pza. Hospital 1, 30202 Cartagena, Spain

e-mail: pedro@pino.datsi.fi.upm.es

Abstract. Amyotrophic Lateral Sclerosis is a severe disease, which dramatically reduces the speech communication skills of patients as disease progresses. The present study is devoted to define accurate and objective estimates to characterize the loss of communication skills, to help clinicians and therapists in monitoring disease progression and in deciding on rehabilitation interventions. The methodology proposed is based on the perceptual (neuromorphic) definition of speech dynamics, concentrated in vowel sounds in character and duration. We present the results from a longitudinal study carried out in an ALS patient during one year. Discussion addresses future actions.

Keywords: Neuromorphic speech processing, amyotrophic lateral sclerosis, speech production, speech perception.

1 Introduction

The detection and characterization of vowel spaces is of most importance in many applications, as in pathological characterization of speech, accordingly the present work will investigate inter-vowel space detection and characterisation in amyotrophic lateral sclerosis (ALS) by neuromorphic methods. ALS is a very severe and rapidly progressive neuromuscular disease of unclear origin [1] and with treatment able to halt progression. This disorder is characterized by degeneration of upper and lower motor neurons, causing generalized muscle weakness and atrophy, with death occurring, in general, between 2-5 years after disease onset, due to complications related to respiratory muscles involvement. One of the most dramatic symptoms is the progressive limitation of speech production, resulting from deterioration of the complex neuromuscular system involved in respiration, phonation, swallowing and lingual and oro-facial muscle function. This process degenerates in a specific kind of dysarthria characterized by hypernasality, reduced speech rhythm, vowel intelligibility degradation, loss of consonantal dynamics, reduced and prolonged number of inter-phonation intervals and pitch dysprosody. The patient experiences a loss of oral communication capability, usually characterized by a reduction of intelligibility and speech rate [2], which may lead to social isolation and depression. The present paper is intended to explore some of the most perceivable features among the ones described above, with the aim of helping the early detection of speech limitation, as well as to provide the speech therapist with objective tools to evaluate patients over disease progression in order to optimize exercising techniques to manage patients. Among the different numbers of speech changes mentioned above, the paper concentrates in the description of vowel colour and count as possible markers of ALS dysarthria, as well on the number and duration of inter-phonation intervals. In doing so, an important definition has to be established, which is the nature and characteristics of vowel sounds. This is not a trivial task as vowels may be defined under the acoustic-phonetic or phonologic point of view [3]. In such task, perceptual concepts of vowel production and perception may be of great help. In this study, vowels will be characterized by the following descriptors: phonation must be present (i.e. a glottal source excitation of the vocal tract must be detected), strong and narrow formant descriptors must be evident, stability in the formants has to be maintained under a certain criterion, and a mapping in the vowel triangle of the patient may be attributed to a certain phonologic attractor or vowel
representation space. The paper is organized as follows: A neuromorphic description of the phonation and articulation processes is given in section 2, to easy the understanding of the underlying neuromotor mechanisms involved; a brief description of vowel nature, based in formant characteristics and dynamics, is given in section 3; in section 4, the metrics used to characterize the articulation in dysarthria is presented, and the case study is described; section 5 presents the results, which are illustrated graphically and briefly discussed; finally, conclusions are presented in section 6.

2 Neuro-Physiological Speech and Phonation Model

Speech production is planned and instantiated in the linguistic neuromotor cortex (see Fig. 1). The neuromotor activation sequence involved in speech production is transmitted to the pharynx (2), tongue (3), larynx (4), chest and diaphragm (5) through the brainstem motor neurons. Fine muscular control is provided by a sophisticated feedback control system (6). A large number of cranial nerves are involved in motor controls of relevant muscles, including the facial, hypoglossus, glossopharyngeal and vagus nerves, innervating the facial muscles, the tongue and, in particular, the following muscles: levator veli palatini, palatoglossus and palatopharyngeous (2), which are critical for the naso-pharyngeal switch. These latter structures play a most relevant role in nasalization (hyper-, hypo- and modal). The superior, middle and inferior pharyngeal constrictors, and stylopharyngeous (3) muscles found in the mid-pharynx, are responsible for the swallowing function as well as of changes in the vocal tract during speech articulation. The cricothyroid, transverse and oblique arytenoid, as well as the posterior cricoarytenoid (4) muscles in the larynx are responsible for vocal fold stretching, adduction and abduction by acting on the cricoarytenoid joint as well as in raising and lowering the cricothyroid cartilage. The vagus nerve (5) is responsible for filling and depleting the lung cavity with air by contraction and relaxation of the crural diaphragm. Most of the muscles in the tongue, responsible of articulation gestures are innervated by the hypoglossal nerve (CN XII). Other muscle-nerve systems of interest in speech are the buccal and mental nerves, derived from the facial nerves. Any alteration in the functionality of these structures will produce perturbations in respiration, phonation and articulation, giving place to specific dysarthrias [4, 5] which may be characterized by the F2 vs. F1 positions in time [3].

Fig. 1  Simplified view of main neural pathways involved in the production of phonation and speech articulation. N: nasal cavity, V: velum, P: palate, A: alveoli, L: lips, T: teeth, G: tongue.
3 Neuromorphic characterization of speech

Vowels may be formally defined as applications between the space of acoustic representations at the cortical level to the set of perceptual symbols defined as vowels at the phonologic or linguistic level [3]. The acoustic-phonetic nature of these patterns is based on the association of the two first resonances of the Vocal Tract, which are referred to as ‘formants’, and described as F₁ and F₂. F₁ in the range of 200-800 Hz is the lowest, F₂ sweeps a wider range, from 500 to 3000 Hz. Under this point of view the nature of vowels may be described by formant stability during a time interval larger than 30 ms, and relative position in the F₂ vs F₁ space, in which is called the ‘Vowel Triangle’ (see Fig. 2). The characterization of vowels by neuromorphic speech processing requires the identification of formants as the basic instantiations to develop further knowledge. Formants are specific resonances of the articulation organs (vocal and nasal tracts, and pharyngeal cavities). Neuromorphic processing refers to processing methods directly inspired in neuronal activity (Hebbian structures) [6, 7]. The main processes mimicking the neuronal activity are lateral inhibition formant profiling, tonotopic frequency band tracking, vowel representation space grouping by space-frequency neuromorphic density functions, vowel assignment by mutual exclusion, and vowel temporal clipping, among others.

![Reference Vowel Triangle](image)

**Fig. 2** Reference Vowel Triangle used in the present study. The vowel set {i, e, a, o, u} in full circles is referred as the cardinal set. The vowel set in dash circles may have different phonological assignments in different languages.

A full description of these processes can be found in [3]. The specific procedures implemented in this study are the following:

1. The speech trace is processed by an adaptive lattice gradient filter to obtain the inverse vocal tract transfer function.
2. A spectrogram is evaluated from the coefficients of the inverse vocal tract transfer function.
3. Formants are estimated from the maxima of the spectrogram by lateral inhibition.
4. Speech activity and phonation activity are estimated by the umbralization of speech and glottal residual energy.
5. The first two formants are tracked using space-frequency density functions.
6. Vowel representation spaces are used in detecting vowel presence by formant pair associations.
7. Vowel assignments by mutual exclusion determine the most probable vowel uttered.
8. Vowel limits in time are delimited by characteristic-frequency overlapping.

Speech may be described as a time-running acoustic succession of events (or phonetic sequence, see Fig. 3.a) [8]. Each event is associated with an oversimplified phonation paradigm composed of vowels, and non-vowels. Non-vowel sounds are characterized by unstable formants (dynamic), by not having a representation inside the vowel triangle, or by lacking a neat F₂ vs. F₁ pattern. The International Phonetic
Alphabet (IPA) [9] has been used, with symbols between square brackets [a] and bars /a/ are phonemes (acoustic representations) and phonologic representations, respectively. Formants are characterized in this spectrogram (middle template) by darker energy envelope peaks. What can be observed in the figure is that the vowels and vowel-like sounds correspond to stable positions of the formants.

Fig. 3  a) time series of the utterance in Portuguese “tudo vale a pena quando a alma não é pequena” (“tudo vala peh nãm e pehmen”) uttered by a female control speaker. b) Adaptive Linear Prediction Spectrogram (grey background) and first two formants (superimposed in color). The color dots mark the positions of each pair {F1,F2} from green (the oldest) to red (the most recent). The approximate phonetic labeling is given as a reference. c) Formant plot of F2 vs F1. d) Same plot as a Formant Chart commonly used in Linguistics. The blue triangle and circles give the limit positions of the five cardinal vowels /i/, /e/, /a/, /o/, /u/ (for a typical male speaker in blue, female in melba). These plots show the formant trajectories of the utterance. There is color correspondence between the bottom and middle templates to track formant trajectories on the time axis.

4 Materials and Methods

The present study has a marked exploratory nature. Early work on formant descriptions in ALS related dysarthria has a long history [10]. Nevertheless, objective characterization of these articulation abnormalities by using objective representations on the vowel triangle is not frequently found. The study will concentrate in producing sequences of positions on the vowel triangle F2 vs. F1 corresponding to pairs {F1(n), F2(n)}, where n is the discrete time index, as given in Fig. 3 (c and d). The more stable a vowel will be the more points will be found in a given area of the vowel triangle in time. A measurement of the vowel triangle extension covered by a given sentence or utterance may be produced in terms of the distribution of pairs {F1, F2} in that specific area. Therefore the following landmarks of the vowel triangle will be defined

\[
\begin{align*}
V_{UL} &= \{ q_1^1, q_2^1 \}, \\
V_{LL} &= \{ q_1^2, q_2^2 \}, \\
V_{MR} &= \{ q_1^3, q_2^3 \}, \\
C_{MM} &= \{ q_1^4, q_2^4 \}
\end{align*}
\]  

(1)

Disclaimer: This is a draft copy (restricted to non-commercial personal use) of the paper published in the Proceedings of the JVHC 2013, I Jornadas Multidisciplinares de Usuarios de la Voz, el Habla y el Canto, Las Palmas de Gran Canaria - 27-28 de junio 2013, pp. 61-70, ISBN: 84-695-8101-5 © GIAPSU-UPM
where $V_{UL}$, $V_{LL}$, $V_{MR}$ and $V_{MM}$ are respectively the upper left, lower left and mid right vertices of the vowel triangle, and $C_{MM}$ is the median centre of the triangle, defined in terms of the generic quantiles

$$q^\theta_i = \arg \left\{ \int_{-\infty}^{\infty} \nu \gamma_i(\nu)d\nu \right\}$$

where $\gamma_i(\nu)$ is the probability distribution of the formant $i$ in frequency $\nu$, and $\theta$ is the specific quantile threshold (for instance $\theta=0.03$ would correspond to a 3% quantile). In the present study the following definitions apply: $\theta_1=0.03$, $\theta_2=0.5$ and $\theta_3=0.97$. Using these definitions the virtual centroid of the vowel triangle would be defined as

$$C_v = \left\{ \frac{q^\theta_1 + q^\theta_2}{2}, \frac{q^\theta_1 + q^\theta_3}{2} \right\}$$

whereas the median centroid of the vowel triangle would be

$$C_m = \{ q^\theta_1, q^\theta_2 \}$$

The asymmetry coefficient would be the difference between the median and the virtual centroids, which may be expressed in module and argument as

$$M_A = \left| \frac{2q^\theta_2 - 1}{q^\theta_1 + q^\theta_2} \right|^{1/2}$$

$$\psi_A = \arctan \left( \frac{2q^\theta_2 - q^\theta_1 + q^\theta_2}{2q^\theta_1 - q^\theta_1 + q^\theta_1} \right)$$

Given the exploratory character of the present study, a case study of a woman affected with ALS has been used in contrast to a control healthy woman. The case study consisted in five recordings from the patient taken at specific 3-month intervals, these being referred to as HA_T0 (November 2011), HA_T1 (January 2012), HA_T2 (March 2012), HA_T3 (July 2012) and HA_T4 (October 2012). HA_T2 had to be rejected after voice quality analysis later and was not used in the present acoustic study. In all cases the recordings contained utterances of the sentence /tudo vale a pena quando a alma não é pequena/ in Portuguese. Baseline data (HA_T0) was recorded in initial stage of disease progression, when clinical evaluation (ALSFRS [10]) indicated a high score in bulbar related functions. The results of the study conducted on these recordings are given in the next section.

5 Results and Discussion

The descriptions of the vowel triangle for each utterance produced were obtained and compared. For the sake of brevity only the first and last plots compared against the control subject are given in Fig. 4. It may be seen that the distribution of the control subject stresses the main patterns and trajectories of the target sentence. The first utterance of the ALS patient (HA_T0) stresses the differences in the vocalic space between [ʊ] and [a], but fails in weighting the respective distribution of each vowel group. The last utterance (HA_T4) is clearly unbalanced towards [a], with most of the vowels improperly articulated as [æ]. The plots in Fig. 5 help in establishing a better comparison among the different vowel triangles and to derive resolving conclusions.
Fig. 4  Positions of f1 vs f2 on the vowel triangle. The 3% quartiles are given in red circles. The median centre is given as a red diamond. a) Control female. The main vowel positions are for [ʊ] (in melba) and [a] (in green). The formant trajectories are well organized and separate. The frequency span is large both for f1 and f2. b) HA_T0. The vowel space is much more confuse, there are not clear vowel trajectories, but the [ʊ] and [a] are still differentiated. The frequency span is still wide, but it shrinks in f2. The median centre is slightly tilted to the left. c) HA_T4.

Fig. 5  The approximate vowel triangles derived from the utterances from the control subject (ContFemale), and the ALS patient in four different sessions chronologically ordered from less severity to most severity (HA_T0, HA_T1, HA_T3 and HA_T4) are compared among themselves. It may be seen that the vowel triangle for HA_T0 is slightly narrower in f2 but larger in f1 than the control one. HA_T1 shrinks clearly with respect to HA_T0. The same happens with HA_T3 and HA_T4 with respect to HA_T1, although there is not a clear change between themselves.

The progressive degradation of the vowel triangle can be clearly perceived, with strong differences between the results for HA_T0 (still comparable with the control subject) and the three last utterances. This indicates that a strong decay in articulatory ability of the patient took place from November 2011 to January 2012. These results are summarized in Table 1.
Table 1. Asymmetry measurements for the vowel triangles studied

<table>
<thead>
<tr>
<th></th>
<th>C1 (Hz)</th>
<th>C2 (Hz)</th>
<th>C1m (Hz)</th>
<th>C2m (Hz)</th>
<th>M</th>
<th>A</th>
<th>φ (deg.)</th>
<th>VSA</th>
<th>FCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaleRef</td>
<td>445</td>
<td>1578</td>
<td>406</td>
<td>1507</td>
<td>0.099</td>
<td>-152.9</td>
<td>-131.1</td>
<td>240230</td>
<td>1.07</td>
</tr>
<tr>
<td>FemaleRef</td>
<td>570</td>
<td>1828</td>
<td>523</td>
<td>1656</td>
<td>0.125</td>
<td>-355.0</td>
<td>-142.3</td>
<td>361240</td>
<td>0.93</td>
</tr>
<tr>
<td>ContFemale</td>
<td>601</td>
<td>2000</td>
<td>695</td>
<td>1555</td>
<td>0.272</td>
<td>-55.0</td>
<td>-113.4</td>
<td>147190</td>
<td>1.29</td>
</tr>
<tr>
<td>HA T0</td>
<td>640</td>
<td>1578</td>
<td>539</td>
<td>1523</td>
<td>0.200</td>
<td>-113.4</td>
<td>-63.4</td>
<td>90637</td>
<td>1.42</td>
</tr>
<tr>
<td>HA T1</td>
<td>586</td>
<td>1578</td>
<td>578</td>
<td>1531</td>
<td>0.033</td>
<td>-113.4</td>
<td>-63.4</td>
<td>90637</td>
<td>1.42</td>
</tr>
<tr>
<td>HA T3</td>
<td>640</td>
<td>1578</td>
<td>648</td>
<td>1539</td>
<td>0.027</td>
<td>-63.4</td>
<td>-63.4</td>
<td>90637</td>
<td>1.42</td>
</tr>
<tr>
<td>HA T4</td>
<td>578</td>
<td>1656</td>
<td>664</td>
<td>1726</td>
<td>0.154</td>
<td>15.7</td>
<td>105740</td>
<td>1.43</td>
<td></td>
</tr>
</tbody>
</table>

The most interesting fact to be stressed is that the angle of the normalized asymmetry coefficient swings from the third to the first quadrant in a progressive succession. This means that the orientation of the vowel median centroids is evolving from a more balanced situation to a tendency marked by vowel [æ]. The two last columns in the table give the Vowel Space Area (VSA) and the Formant Centralization Ratio (FCR) evaluated following [12, 13] as a further reference. It may be seen that the FCR gives also an indication of the anomalous articulation function. The respective values of the asymmetry modulus and phase (Relative Centre Displacement) are given also in Fig. 6. It may be seen from the plot that the cases show a progression from the third quadrant to the first one, with a transit through the fourth quadrant. This is an objective measurement expressing the migration of the articulation center of gravity towards the position of an open mid vowel as [æ].

Fig. 6 The asymmetry coefficient is the normalized distance between the triangle base and height midpoint intersection and the median center. This relative displacement is plotted in module and angle for four progressive ALS stages from the same patient (HA_T0, HA_T1, HA_T2 and HA_T3).

The last part of the results presented is the study of vowel and stop interval durations. To produce such results vowels are detected using a coincidence function between the first and second formant CF neuron firings [3], as these units are activated when a formant is relatively unchanged for a certain interval. Coinciding quasi-stable formants are an indication of vowel presence, independently of its nature. The results for the target sentence from the control and the first and last ALS patient utterances are given in Fig. 7.
Fig. 7  Vowel-Consonant Dynamics. The green line delimits vowel intervals corresponding to the high level, whereas the low level corresponds to silence or unstable formants (consonant patterns). The long intervals are associated to vowel patterns, the short intervals correspond to brief and unstable vowels near consonantal groups. a) Control female. Stable vowel groups are 200-300 ms long. Short unstable vowels around 50 ms may be observed near plosive groups [p-] and [kw-]. Vowel formants take different configurations. b). HA_T0. The number of long vowel groups is smaller, the number of short vowel groups is larger. Formant patterns deteriorate. c) HA_T4. Larger vowel groups may be appreciated again for the first part of the sentence, but formant patterns are much deteriorated pointing to the positions for [æ]. The number of silences is shorter and larger, the length of the sentence is more than twice longer than in (a).

The presence of vowel groups are divided in three categories: larger than 150 ms, between 50 and 150 ms, and smaller than 50 ms. Classically regular vowels would be included in the second category. The first category may be associated with an impaired articulation. The presence of the third category is associated with dynamic transitions between stable vowel positions, and its reduction can be also associated with impaired articulation. Silences can be also classified according to the same principles. Intervals longer than 150 ms can be associated with phrase splits, less than 150 ms with plosive consonants. The account of the different vowel and silence intervals for the three cases presented in Fig. 7 are listed in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Vowel group duration and stop and silence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
</tr>
<tr>
<td>ContFemale</td>
</tr>
<tr>
<td>HA_T0</td>
</tr>
<tr>
<td>HA_T4</td>
</tr>
</tbody>
</table>

In general it may be observed that the number of longer vowel groups is larger in pathological speech than in normal speech. It may be seen also that the duration of the utterance is much larger as pathology expresses its severity.

Disclaimer: This is a draft copy (restricted to non-commercial personal use) of the paper published in the Proceedings of the JVHC 2013, I Jornadas Multidisciplinares de Usuarios de la Voz, el Habla y el Canto, Las Palmas de Gran Canaria - 27-28 de junio 2013, pp. 61-70, ISBN: 84-695-8101-5 © GIAPSI-UPM
6 Conclusions

The results of the study avail some of the preliminary goals formulated in section 1, consisting in producing objective measurements of speech degradation phenomena which may be perceived by the expert listener or the speech therapist: increased hypernasality, reduced speech rhythm, reduction of vowel intelligibility, loss of consonantal dynamics, reduced and prolonged number of inter-phonation intervals and pitch dysprosody, among others. Due to the limitations of the present study only rhythm, duration and intelligibility of vowels have been studied using neuromorphic detection of formant dynamics to establish the presence and nature of produced vowels. The most important findings established in this sense as illness progresses are the following:

- The utterance of the same sentence is produced in a longer interval.
- The duration of vowels in syllabic nuclei are also extended.
- The number of inter-phrasal intervals is reduced, but its duration is extended.
- The number of pre-stop silence intervals is reduced. As a consequence stop consonants are lost.
- The vowel triangle shrinks, especially in F2.
- The vowel triangle centroids evolve towards [æ].

This last finding needs a further explanation. It is well known from literature [14] that F1 is very much related to the degree of opening of the vocal tract ([i] and [u] corresponding to the more closed extremes, whereas [a] gives the more open extreme), whereas F2 is more related to the articulation position (where [u] is considered a back vowel whereas [i] is a frontal, and [a] would be a middle vowel). Thus forcing frontal or rear vowels would imply the operation of the hypoglossal and facial (mental) neuromotor systems which need not be active in the neutral mid position open vowel given by [æ]. Under severely impaired neuromuscular activity this would be the only articulatory position and the relative colouring of the different vowels would be fused towards this final position. Other vowels nearby the lower vertex of Fig. 2 could also be plausible solutions to an impaired articulatory situation. There are other aspects of ALS dysarthric speech which have not been checked in the present study, as estimating the degree of hypernasality due to the failure of the levator veli palatini, palatoglosus and palatopharyngeous neuromuscular structures acting on the naso-pharyngeal switch, as these would require a spectral detector to model the zeroes in the vocal and nasal tract anti-resonances. Consonantal dynamics could also be traced using neuromorphic speech processing [7]. Dysprosody could also be characterized using well-known pitch tracking methods. These tasks are left for future research. Another important task to be accomplished is the estimation of the biomechanical parameters of phonation in ALS patients, in a similar way used in other neurological pathologies [15], which has not been conducted in this case due to the limitations of the study. Another important task to fulfil in the near future is the collection of a large database containing longitudinal studies as the one described to extend the statistical significance of the findings produced in this study.

Acknowledgments. This work is being funded by grants TEC2009-14123-C04-03 and TEC2012-38630-C04-04 from Plan Nacional de I+D+i, Ministry of Science and Technology of Spain.

References