Article:

Orozco-Arroyave, J.R., Vásquez-Correa, J.C., Vargas-Bonilla, J.F. et al. (13 more authors) (2018) NeuroSpeech. SoftwareX, 8. pp. 69-70.

https://doi.org/10.1016/j.softx.2017.08.004

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

NeuroSpeech: An open-source software for Parkinson's speech analysis

Juan Rafael Orozco-Arroyave [a,b,*], Juan Camilo Vásquez-Correa [a,b], Jesús Francisco Vargas-Bonilla [a], R. Arora [c], N. Dehak [c], P.S. Nidadavolu [c], H. Christensen [d], F. Rudzicz [e], M. Yancheva [e], H. Chinaei [e], A. Vann [f], N. Vogler [g], T. Bocklet [h], M. Cernak [i], J. Hannink [b], Elmar Nöth [b]

[a] *Faculty of Engineering, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Colombia*
[b] *Pattern Recognition Lab, Friedrich-Alexander-Universtät Erlangen-Nürnberg, 91058 Erlangen, Germany*
[c] *Johns Hopkins University, USA*
[d] *University of Sheffield, UK*
[e] *University of Toronto, Canada*
[f] *Stanford University, USA*
[g] *University of California-Irvine, USA*
[h] *Intel, Germany*
[i] *Idiap Research Institute, Switzerland*

A R T I C L E I N F O

Article history:
Available online xxxx

Keywords:
Parkinson's disease
Dysarthria
Speech processing
Python
Software

A B S T R A C T

A new software for modeling pathological speech signals is presented in this paper. The software is called *NeuroSpeech*. This software enables the analysis of pathological speech signals considering different speech dimensions: phonation, articulation, prosody, and intelligibility. All the methods considered in the software have been validated in previous experiments and publications. The current version of *NeuroSpeech* was developed to model dysarthric speech signals from people with Parkinson's disease; however, the structure of the software allows other computer scientists or developers to include other pathologies and/or other measures in order to complement the existing options. Three different tasks can be performed with the current version of the software: (1) the modeling of the speech recordings considering the aforementioned speech dimensions, (2) the automatic discrimination of Parkinson's vs. non-Parkinson's speech signals (if the user has access to recordings of other pathologies, he/she can re-train the system to perform the detection of other diseases), and (3) the prediction of the neurological state of the patient according to the Unified Parkinson's Disease Rating Scale (UPDRS) score. The prediction of the dysarthria level according to the Frenchay Dysarthria Assessment scale is also provided (the user can also train the system to perform the prediction of other kind of scales or degrees of severity).

To the best of our knowledge, this is the first software with the characteristics described above, and we consider that it will help other researchers to contribute to the state-of-the-art in pathological speech assessment from different perspectives, e.g., from the clinical point of view for interpretation, and from the computer science point of view enabling the test of different measures and pattern recognition techniques.

1. Introduction and motivation

Parkinson's disease (PD) is a neurological disorder caused by the progressive loss of dopaminergic neurons in the mid-brain, leading to clinical symptoms like bradykinesia, rigidity, tremor, postural instability, and others. Non-motor symptoms like sleep disorders and problems in cognition and emotions have also been observed [1]. Most of Parkinson's patients develop hypokinetic dysarthria, which is a multidimensional impairment that affects different aspects or dimensions of speech including phonation, articulation, prosody, and intelligibility [2]. The neurological state of PD patients is evaluated subjectively by clinicians who are, in most of the cases, mainly focused on the evaluation of motor deficits rather than the speech impairments; however, the patients claim that the reduced ability to communicate is one of the most diffi-

* Corresponding author at: Faculty of Engineering, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Colombia.
E-mail address: fael.orozco@udea.edu.co (J.R. Orozco-Arroyave).

http://dx.doi.org/10.1016/j.dsp.2017.07.004
1051-2004/

cult aspects of the disease. The Royal College of Physicians in the NICE guidelines [3] recommends that every PD patient should have access to speech and language therapy among other kinds of non-pharmacological treatments [4].

The research community has shown interest in developing computational tools to help patients and clinicians to assess Parkinson's speech. There have been several initiatives towards the development of computer-aided tools to support or assist the speech therapy and the diagnosis of different disorders. In [5] the authors present VOCALIZA, a software application for computer-aided therapy in Spanish language with three levels: phonological, semantic, and syntactic. The system is mainly based on a speech recognizer and it is used to train the language skills of speakers with different pathologies. Several characteristics of using automatic speech recognition (ASR) systems for the assessment of voice, speech, and language disorders are presented in [6]. Additionally, the use of such systems to support the speech and language therapy of patients who had their larynx removed due to cancer and for children with cleft lip and palate was evaluated in [7]. In that work the authors introduced the system PEAKS, which is mainly based on prosodic features (at word and turn level) and showed to be suitable to support several applications (diagnosis and monitoring) in the clinic. Another work based on ASR systems to assist speech and language therapy is presented in [8]. In that work the authors introduce the system to evaluate the speech of children with neuromuscular disorders, e.g., dysarthria. Further to the use of an ASR system, the authors use a pronunciation verification (PV) approach to evaluate the improvement in the communication skills of the user at both phoneme and word levels. Further to the aforementioned studies, several works have been published summarizing the technology that has been developed to support the speech and language therapy. In [9] the authors present a brief description of several systems that were developed to assist the speech therapy of Romanian patients with different disorders, e.g., stammering, logoneurosis, dyslexia-dysgraphia, and others. Recently, in [10] the authors present a systematic review of the literature about computer-aided systems developed to support speech and language therapy mainly focused on articulation and phonological impairments. According to their review, "all the studies introduced their own developed tools and used them for intervention except one study, in which a previously developed software was used". Additionally, the authors mention that the intervention of each study varies based on the technical abilities of the system, the disorder targeted, and the intervention framework. The types of interventions covered on the review include phonation, articulation, phonological awareness, and general intervention.

According to the reviewed literature and the evidence collected in [10], the research community, patients, medical doctors, and therapists lack of computational tools with different characteristics including: user-friendly, easy to use, open source, able to perform several interventions/analyses, and able to be personalized or adapted according to the necessities of the users. In this paper we introduce *NeuroSpeech*, a new system for the semi-automatic analysis of speech signals. It includes measurements to model four speech dimensions: phonation, articulation, prosody, and intelligibility. *NeuroSpeech* has been designed and tested upon Parkinson's speech signals; however, its methods and techniques can be easily adapted to perform analyses of other speech disorders. Further to the computation of several speech dimensions, the system is able to perform a regression analysis based on a support vector regressor (SVR) to estimate the neurological state of the patient (according to the Unified Parkinson's Disease Rating Scale – UPDRS [11]) and the dysarthria level according to a modified version of the Frenchay Dysarthria Assessment (FDA-2) score [12]. This software and its associated documentation, i.e., source code, user and technical manuals, can be downloaded for free from the link https://github.com/jcvasquezc/NeuroSpeech. Further details of the characteristics of *NeuroSpeech* and several case study examples will be provided in the next sections.

2. Parkinson's speech analysis using *NeuroSpeech*

2.1. General characteristics, the user interface, and the speech tasks

NeuroSpeech is a graphic user interface designed in C++ which runs Python scripts. The software uses other programs (also open source) which need to be installed for the correct operation of *NeuroSpeech*. The following is the list of third-party programs:

- **Anaconda**: it is required to have a Python environment.[1]
- **Praat**: this program is required to extract the pitch values from the voice recordings and to compute the vocal formants.[2]
- **ffmpeg**: this tool allows the recording, conversion, and streaming of audio and video files.[3]

2.1.1. User interface – main window

The main window of the user interface is thought to help the user to follow the analysis process. It displays the basic information of the patient, i.e., name and last name. It has only two buttons, one to start the recording and another one to play the recorded file. On the upper right hand side the user can select sound examples of the speech task that is going to be recorded. The software enables the user to select the gender of the speaker (female or male). The raw signal is also displayed on this window. On the lower side of the window there are six arrows named phonation, articulation, prosody, DDK, intelligibility, and PD evaluation. By clicking on those arrows the user can perform the analysis of each speech dimension. The DDK arrow corresponds to diadochokinetic evaluation which consists of the repetition of syllables like /pa/, /ta/, /ka/ or combinations of them like /pa-ta-ka/ or /pa-ka-ta/. This option is added due to its importance and relevance in the assessment of dysarthric speech signals [13]. The right hand side arrow enables two evaluations, the neurological state of the patient according to the UPDRS-III score and the dysarthria level of his/her speech according to the adapted FDA-2 score (details about this evaluation will be provided in the section 2.1.7). Besides the computation of features, *NeuroSpeech* allows the user to perform comparisons with respect to measures obtained from recordings of the healthy control group in the PC-GITA database [14]. All of the comparisons are with respect to the corresponding group according to the sex of the speaker. The components of this window are displayed in Fig. 1. Each number inside the blue boxes highlights different fields on the interface. Each field is described in Table 1.

As it was already introduced above, Parkinson's disease affects several aspects or dimensions of speech including phonation, articulation, prosody, and intelligibility. Each aspect is directly related with the ability to produce an specific sound, movement, rhythm, or effect in the receiver of the oral message. Phonation can be defined as the capability to make the vocal folds vibrate to produce sound, articulation comprises changes in position, stress, and shape of the organs, tissues, and limbs involved in speech production. Prosody is the variation of loudness, pitch, and timing to produce natural speech, and intelligibility is related to the match between the message produced by the speaker and the information effectively received by the listener [15].

In order to provide more understandable or interpretable results, the computer-aided systems that support speech therapy (for

[1] https://www.continuum.io/downloads.
[2] http://www.fon.hum.uva.nl/praat.
[3] http://ffmpeg.org/download.html. Last retrieved 06.12.2016.

Please cite this article in press as: J.R. Orozco-Arroyave et al., NeuroSpeech: An open-source software for Parkinson's speech analysis, Digit. Signal Process. (2017), http://dx.doi.org/10.1016/j.dsp.2017.07.004

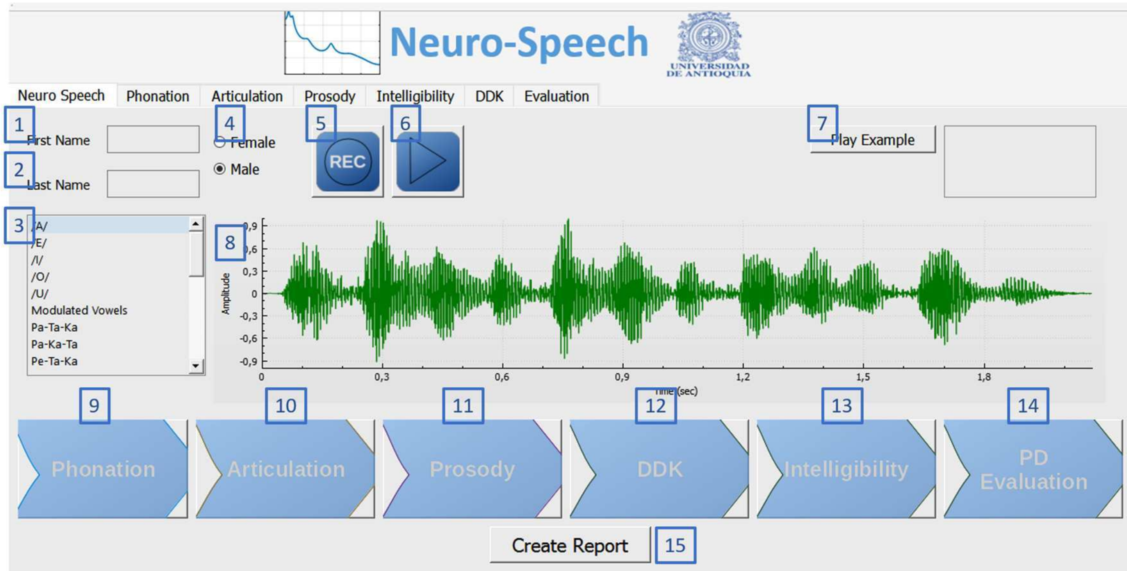

Fig. 1. Main window of *NeuroSpeech*. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Table 1
Description of the fields in the main window.

Field	Description
1	First name of the patient
2	Last name of the patient
3	List of speech tasks to be recorded
4	Allows the user to select the gender of the speaker
5	Starts recording
6	Plays the sound file
7	Plays an example of the speech task
8	Field to visualize the speech signal after recording
9	Performs the phonatory analysis
10	Performs the articulatory analysis
11	Performs the prosody analysis
12	Performs the DDK analysis
13	Performs the intelligibility analysis
14	Performs the dysarthria and PD evaluation
15	Generates the medical report

diagnosis or monitoring) may consider these speech aspects/dimensions. *NeuroSpeech* is designed to perform phonation, articulation, prosody, and intelligibility analyses separately, thus its results can be interpreted separately by the medical expert. As such methods have been presented in previous publications, they will be briefly introduced here, for further details see [16–19].

2.1.2. Phonation analysis

The phonatory capability of a speaker has been analyzed typically in terms of features related to perturbation measures such as jitter (temporal perturbation of the fundamental frequency), shimmer (temporal perturbation of the amplitude of the signal), amplitude perturbation quotient (APQ), and pitch perturbation quotient (PPQ). APQ and PPQ are long term perturbation measures of the amplitude and pitch of the signal, respectively. Further to the perturbation measures, the degree of unvoiced is also included. A brief description of the methods is presented below, further details can be found in [16] and [20]).

Jitter and shimmer

Temporal perturbations in the frequency and amplitude of speech are defined as jitter and shimmer, respectively. Jitter is computed according to Equation (1), where N is the number of frames of the speech utterance, M_f is the maximum of the fundamental frequency, and $F_0(k)$ corresponds to the fundamental frequency computed on the k-th frame.

$$\text{Jitter}(\%) = \frac{100}{N \cdot M_f} \sum_{k=1}^{N} \left| F_0(k) - M_f \right| \quad (1)$$

Shimmer is computed using Equation (2), where M_a is the maximum amplitude of the signal, and $A(k)$ corresponds to the amplitude on the k-th frame.

$$\text{Shimmer}(\%) = \frac{100}{N \cdot M_a} \sum_{k=1}^{N} |A(k) - M_a| \quad (2)$$

Amplitude and Pitch Perturbation Quotients (APQ and PPQ)

APQ measures the long–term variability of the peak-to-peak amplitude of the speech signal. Its computation includes a smoothing factor of eleven voice periods and it is calculated as the absolute average difference between the amplitude of a frame and the amplitudes averaged over its neighbors, divided by the average amplitude. Similarly, PPQ measures the long–term variability of the fundamental frequency, with a smoothing factor of five periods. It is computed as the absolute average difference between the frequency of each frame and the average of its neighbors, divided by the average frequency. Both perturbation quotients are computed using Equation (3), where $L = M - (k - 1)$, $D(i)$ is the pitch period sequence (PPS) when computing the PPQ or the pitch amplitude sequence (PAS) when computing the APQ. M is the length of PPS or PAS, k is the length of the moving average (11 for PAQ or 5 for PPQ), and $m = (k - 1)/2$.

$$\text{PQ} = \frac{1}{L} \sum_{i=1}^{L} \frac{\left| \frac{1}{k} \sum_{j=1}^{k} D(i+j-1) - D(i+m) \right|}{\left| \frac{1}{M} \sum_{n=1}^{M} MD(i) \right|} \quad (3)$$

Degree of unvoiced

This measure is calculated as the ratio between the duration of unvoiced frames and the total duration of the utterance. It is calculated upon sustained phonations, thus it gives information about the amount of aperiodicity in the phonation.

Implementation for developers

The phonation analysis in NeuroSpeech is performed with the python script named `phonVowels.py`, which is stored in the folder `/phonVowels/`. The syntax to perform the analysis is as follows:

Please cite this article in press as: J.R. Orozco-Arroyave et al., NeuroSpeech: An open-source software for Parkinson's speech analysis, Digit. Signal Process. (2017), http://dx.doi.org/10.1016/j.dsp.2017.07.004

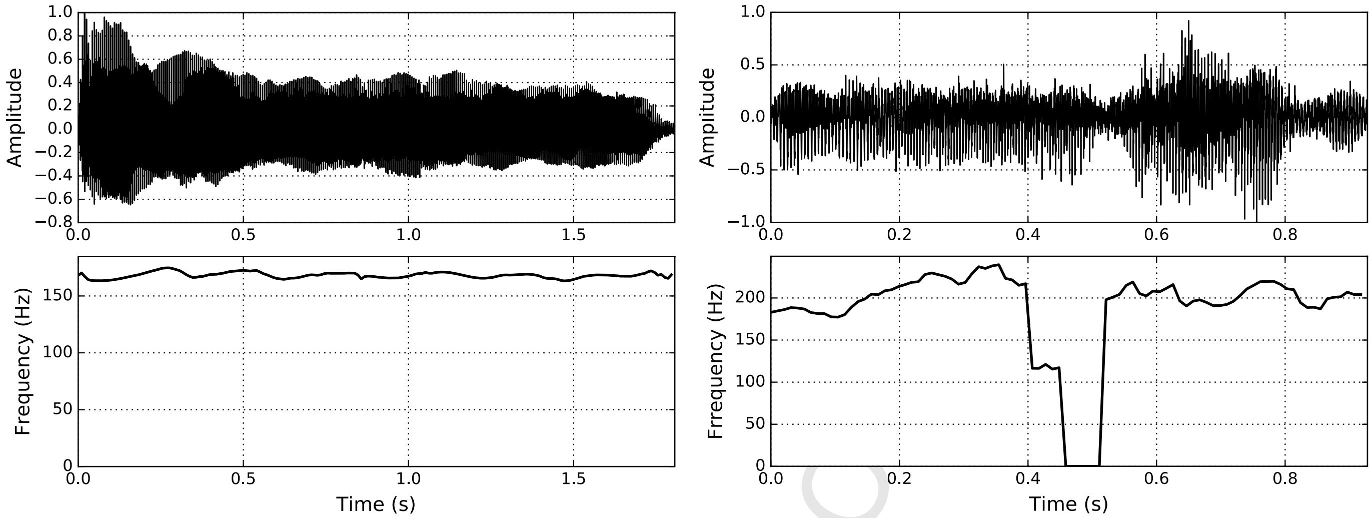

Fig. 2. Speech signal and fundamental frequency of a sustained phonation of vowel A for a healthy speaker (left), and for a PD patient (right).

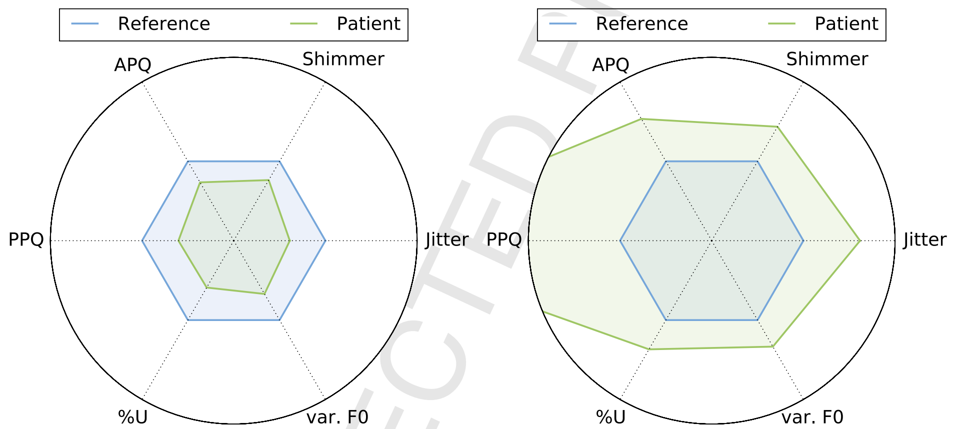

Fig. 3. Radar-type figures for a healthy speaker (left) and for a PD patient (right). (For interpretation of the colors in this figure, the reader is referred to the web version of this article.)

```
python phonVowels.py <file_audio>
<filef0.txt>
<file_features.txt> <path_ base>
```

where <file_audio> corresponds to the name of the audio file that will be analyzed, <filef0.txt> is a file that will contain the contour of the fundamental frequency (computed using Praat), <file_features.txt> is a file that will contain the values of the measures, i.e., jitter, shimmer, APQ, and PPQ. Finally, <path_base> corresponds to the path where the script is contained.

In addition to the measures computed upon the signal that is under evaluation, *NeuroSpeech* has a pre-computed reference based on speech recordings of 50 healthy speakers of the PC-GITA database [14]. All of the features are computed and averaged over those healthy speakers in order to provide a numerical and graphical reference to the user, thus besides the numerical results and the references, *NeuroSpeech* creates a plot with the contour of the fundamental frequency, and a radar-type figure. All of the comparisons are performed with respect to the corresponding healthy group according to the sex of the speaker. Such a figure allows visual comparisons of the result w.r.t. the values obtained from PC-GITA. Fig. 2 displays the contour of the fundamental frequency computed over a sustained vowel /a/ uttered by a healthy speaker (left) and a PD patient (right). Note that the contour of the healthy speaker is more stable than the contour obtained from the patient.

Fig. 3 shows the radar-type figure with the perturbation measures extracted from the vowel uttered by a healthy speaker (left) and a PD patient (right). The polygon in green corresponds to values of the features obtained from the patient, and the polygon in blue corresponds to the average values of the features extracted from the healthy speakers of PC-GITA (female or male, depending on the sex of the test speaker). Typically, when the green pentagon coincides or is inside the blue one it means that the features of the patient are in the same range of the reference. If any of the corners in the green pentagon does not match with the reference pentagon, it suggests that the vibration of the vocal folds is abnormal. For instance, the utterance of the patient in Fig. 3 shows abnormal values of shimmer and APQ, which means that the amplitude (volume) of his/her voice is not stable while producing a sustained vowel, which is a typical sign of dysarthria due to Parkinson's disease.

Implementation for end-users

The window of *NeuroSpeech* that enables phonation analysis is displayed in Fig. 4. This window is divided into two fields. On the left hand side the user can observe the speech recording in the time domain and the F_0 contour. The values of the measures extracted from the sustained phonation are displayed on the right hand side in both formats, numerical and using the radar-type figure.

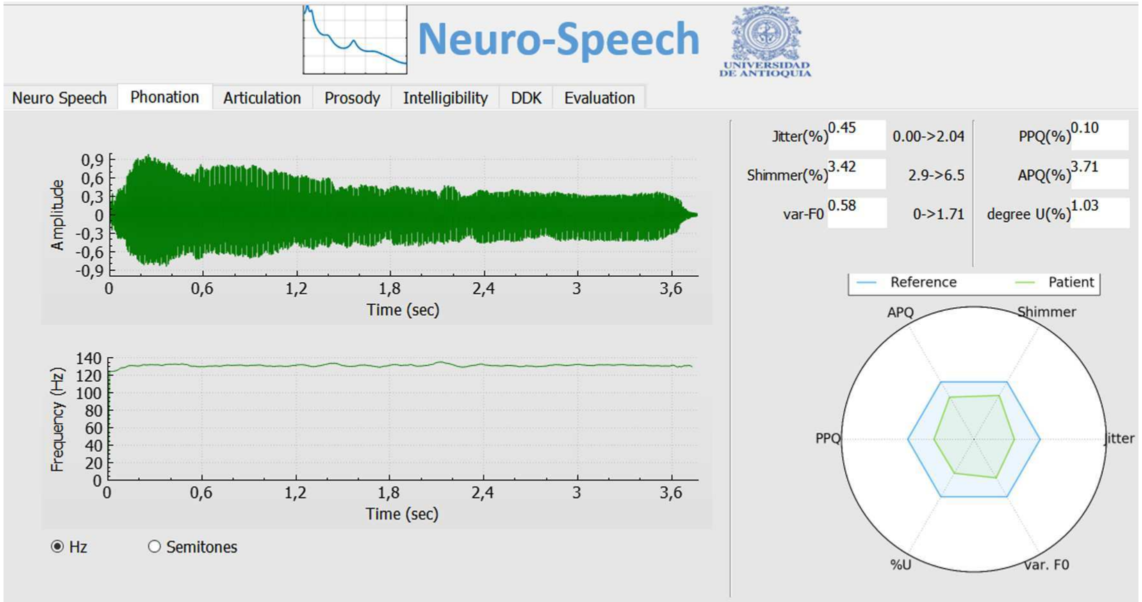

Fig. 4. Window for phonation analysis.

2.1.3. Articulation analysis

Articulation is related to the modification of position, stress, and shape of several limbs and muscles involved in the speech production process. This kind of analysis can be performed with sustained vowels or with continuous speech signals. When sustained utterances are considered, several measures to assess the position of the tongue are typically used [21,22]. For the case of continuous speech signals, a new approach to evaluate the capability of PD patients to start and to stop the vocal fold movement was introduced recently in [17] motivated on the results obtained in [18]. In *NeuroSpeech* we have implemented both approaches, using sustained vowels and continuous speech signals.

Articulation in sustained vowels

This analysis is mainly based on the computation of the first two vocal formants F_1 and F_2. Measures such as the vowel space area (VSA), vocal pentagon area (VPA), and formant centralization ratio (FCR) are calculated to evaluate the articulation capabilities of the speakers.

VSA is calculated to quantify possible reduction in the articulatory capability of the speaker. Such a reduction is observed as a compression of the area of the vocal triangle which is found by extracting the first two vocal formants from the vowels /a/, /i/, and /u/. Then the VSA is estimated using Equation (4).

$$\mathrm{VSA} = \frac{|F_{1\mathrm{i}}(F_{2\mathrm{a}} - F_{2\mathrm{u}}) + F_{1\mathrm{a}}(F_{2\mathrm{u}} - F_{2\mathrm{i}}) + F_{1\mathrm{u}}(F_{2\mathrm{i}} - F_{2\mathrm{a}})|}{2} \quad (4)$$

When the first two formants of the five Spanish vowels are considered as the vertexes of a polygon, the vocal pentagon is formed, and its area is called VPA. This measure quantifies the articulatory capabilities of the speakers when they pronounce the five Spanish vowels. VPA was introduced in [23] to evaluate articulatory deficits in PD patients. This measure is calculated using Equation (5), where $\mathrm{ps}_1 = F_{1\mathrm{a}}F_{2\mathrm{o}} - F_{1\mathrm{o}}F_{2\mathrm{a}}$, $\mathrm{ps}_2 = F_{1\mathrm{o}}F_{2\mathrm{u}} - F_{1\mathrm{u}}F_{2\mathrm{o}}$, $\mathrm{ps}_3 = F_{1\mathrm{u}}F_{2\mathrm{i}} - F_{1\mathrm{i}}F_{2\mathrm{u}}$, $\mathrm{ps}_4 = F_{1\mathrm{i}}F_{2\mathrm{e}} - F_{1\mathrm{e}}F_{2\mathrm{i}}$, and $\mathrm{ps}_5 = F_{1\mathrm{e}}F_{2\mathrm{a}} - F_{1\mathrm{a}}F_{2\mathrm{e}}$.

$$\mathrm{VPA} = \frac{|\mathrm{ps}_1 + \mathrm{ps}_2 + \mathrm{ps}_3 + \mathrm{ps}_4 + \mathrm{ps}_5|}{2} \quad (5)$$

Finally, FCR is a measure introduced in [21] to evaluate changes in the articulatory capability of people. According to the authors, this measure presents a reduced inter–speaker variability, thus it is suitable for diagnosis and monitoring of voice disorders such as dysarthria. FCR is computed with Equation (6).

$$\mathrm{FCR} = \frac{F_{2\mathrm{u}} + F_{2\mathrm{a}} + F_{1\mathrm{i}} + F_{1\mathrm{u}}}{F_{2\mathrm{i}} + F_{1\mathrm{a}}} \quad (6)$$

Articulation in continuous speech

The articulation capabilities of the patients in continuous speech is evaluated considering the energy content in the transition from unvoiced to voiced segments (onset), and in the transition from unvoiced to voiced segments (offset). These features were introduced in [16] and are described with details in [18]. The main hypothesis that motivates the modeling of such transitions in PD speech is [24]:

> *PD patients produce abnormal unvoiced sounds and have difficulty to begin and/or to stop the vocal fold vibration. It can be observed on speech signals by modeling the frequency content of the unvoiced frames and the transitions between voiced and unvoiced sounds.*

To compute those measures, the fundamental frequency of the speech signal is estimated using Praat [25] to detect the voiced and unvoiced segments. Afterwards, onsets and offsets are detected and 40 ms of the signal are taken to the left and to the right of each border. Finally, the spectrum of the transitions is distributed into 22 critical bands following the Bark scale, and the Bark-band energies (BBE) are calculated according to [26]. For frequencies below 500 Hz the bandwidths of the critical bands are constant at 100 Hz while for medium and high frequencies the increment is proportional to the logarithm of frequency. The Equation (7) reproduces the frequency distribution proposed by Zwicker et al. with an accuracy of ±0.2 Bark.

$$\mathrm{Bark}(f) = 13 \arctan\,(0.00076f) + 3.5 \arctan\left(\frac{f}{7500}\right)^2, \quad (7)$$

arctan is measured in [radians] and f is in [Hz].

Note that the minimum duration of the voiced frames to be processed is 40 ms and the maximum duration of the unvoiced frames is 270 ms (longer unvoiced frames are considered pauses). These lower and upper limits have been tuned manually by direct observation during several experiments.

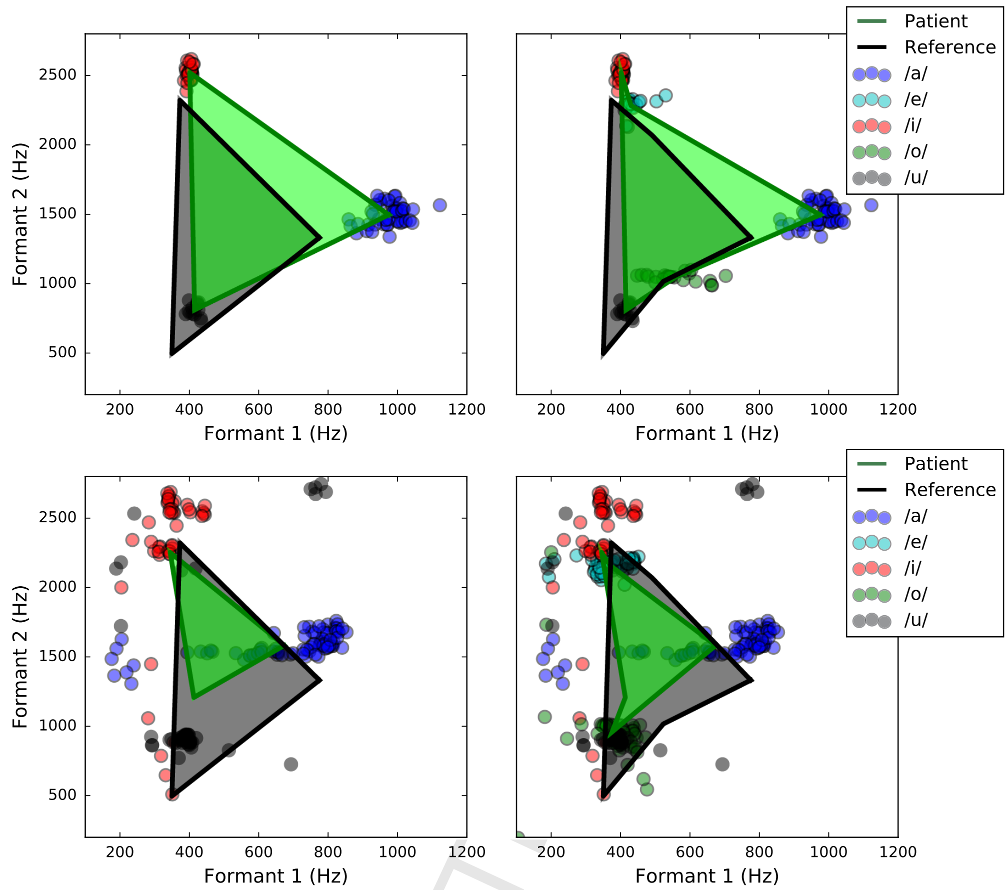

Fig. 5. Vocal triangle and pentagon for a healthy speaker (up) and for a PD patient (bottom).

Implementation for developers – case 1: sustained vowels

This analysis is performed with the python script called artVowels.py, which is contained in the folder artVowels. The syntax to perform the analysis is as follows:

```
python artVowels.py <file_audioA>
<file_audioE> <file_audioI>
<file_audioO> <file_audioU>
<file_resultsA> <file_resultsE>
<file_resultsI> <file_resultsO>
<file_resultsU> <file_features>
<path_base>
```

where <file_audioX> corresponds to the audio file of vowel X and X ∈ $\{A, E, I, O, U\}$, <file_resultsX.txt> is the file with the values of the formant frequencies of vowel X (computed using Praat), <file_features.txt> is the file that will contain the features previously described (the average of the vocal formants, VSA, VPA, and FCR), and <path_base> corresponds to the path where the script is stored.

The script generates also figures of the vocal triangle and the vocal pentagon. The polygons obtained with the speech recordings of the healthy speakers of PC-GITA are also displayed for comparison purposes. Fig. 5 shows the vocal triangle and the vocal pentagon obtained from a healthy speaker (up side), and for a PD patient (bottom side). Note the reduction in the area of the triangle and the pentagon for the case of the PD patient relative to the areas obtained for healthy speakers. This fact indicates a reduction of the articulatory capabilities of the patient. Note also that the formant frequencies for each vowel are more spread for the PD patient than for the healthy speaker, indicating a loss of control of the tongue and velum while producing the sustained phonation.

Implementation for developers – Case 2: continuous speech

This analysis is performed with a python script called artCont.py, which is stored in the folder /artCont/. The syntax to perform the analysis is as follows:

```
python artCont.py <file_audio>
<file_features> <path_base>
```

where <file_audio> corresponds to the audio file that will be analyzed, <file_features.txt> is a file that will contain the results, and <path_base> corresponds to the path where the script is stored.

The script creates also a radar figure that allows comparisons of the computed features w.r.t. those computed from the reference. Fig. 6 shows the radar-type plot of the articulation measures in continuous speech for a healthy speaker (left) and a PD patient (right). BBE_onXX corresponds to the Bark band energy number XX computed upon the onsets of the utterance, and BBE_offXX corresponds to the Bark band energy number XX computed upon the offsets of the utterance. The area in green corresponds to the values of the features obtained from the speaker under evaluation (typically a patient), and the area in blue corresponds to the values of the reference. When the green area is inside the blue one, it means the features of the current speaker are in the same range of the healthy speakers. In Fig. 6 note that in the left hand side, the values exhibited by the healthy speaker are inside the blue area, conversely in the right hand side, the patient exhibits lower energy values in BBE_on01 to BBE_on07, and in BBE_off04 to BBE_off07.

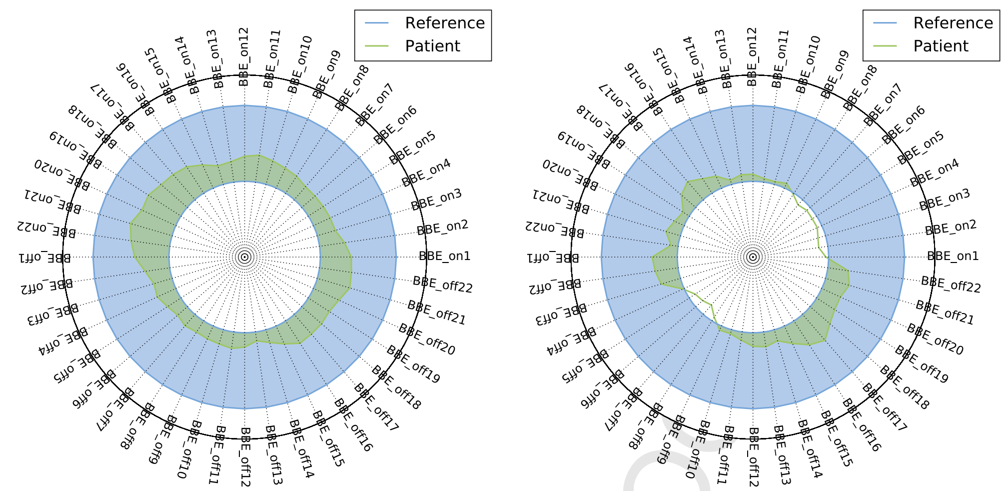

Fig. 6. Radar-type figures for a healthy speaker (left) and for a PD patient (right). (For interpretation of the colors in this figure, the reader is referred to the web version of this article.)

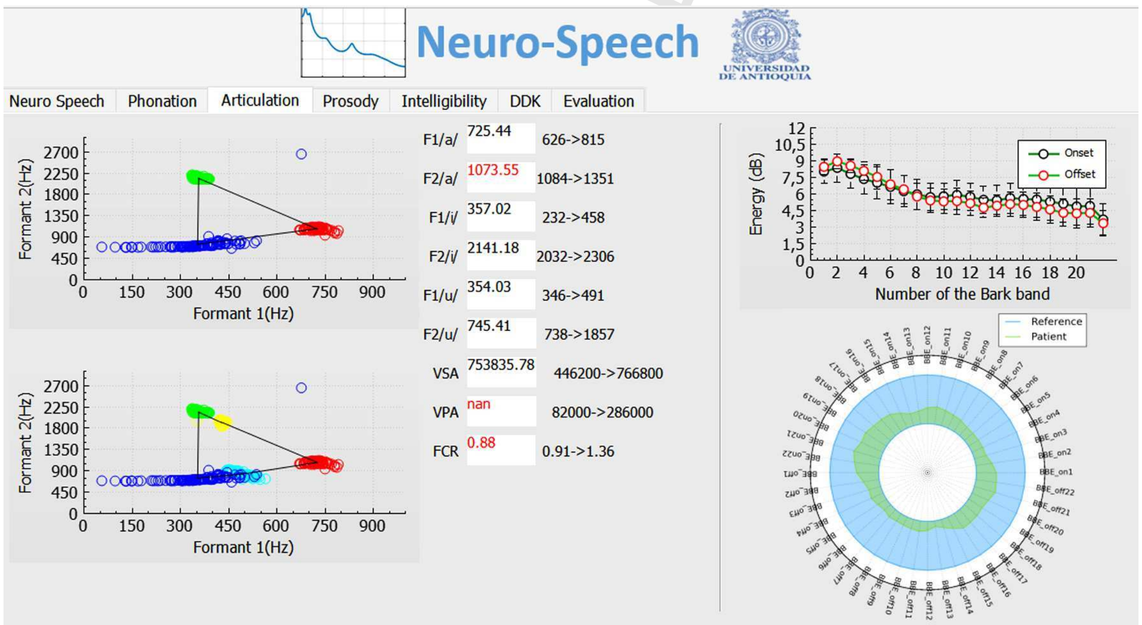

Fig. 7. Window for articulation analysis. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Implementation for end users

The articulation analysis using the graphic user interface is performed by clicking on the articulation button of the main window 1. The result is displayed in Fig. 7. This window consists of two fields. On the left hand side the articulation analysis based on sustained utterances is displayed. It includes the vocal triangle, the vocal pentagon (for cases where the five Spanish vowels are available), and the values of the measures introduced in Section 2.1.3. Those values that exceed the reference ones are displayed in red. On the right hand side the window displays the results of the analysis upon continuous speech signals. In the upper part of this field there is a plot with the log-energy values of the Bark bands extracted from the onsets and offsets. Below this plot, there is a radar-type figure displaying the result of the Bark band energies compared with respect to the reference set (healthy controls of PC-GITA).

2.1.4. Prosody analysis

Prosody studies variables like timing, intonation, and loudness during the production of speech. These features are necessary to produce intelligible and natural speech and they have been studied in the research community since several years ago [27–29], and [30]. Prosody is commonly evaluated with measures derived from the fundamental frequency, the energy contour, and duration. In *NeuroSpeech* the prosody analysis can be performed with the read text and/or the monologue.

Features related to the fundamental frequency

With the aim of modeling intonation patterns of the PD patients, the contour of the fundamental frequency is computed, and statistical functionals are computed: average, standard deviation, and maximum values are calculated to evaluate the monotonicity of the patient, and the maximum frequency that the speaker can reach.

Features related to energy

Similar to the fundamental frequency, the energy contour of the utterance is computed and statistical functionals are calculated from such a contour. The average and standard deviation of the energy besides its maximum value are calculated.

Please cite this article in press as: J.R. Orozco-Arroyave et al., NeuroSpeech: An open-source software for Parkinson's speech analysis, Digit. Signal Process. (2017), http://dx.doi.org/10.1016/j.dsp.2017.07.004

Table 2
Prosody features.

Measure	Description
Vrate	Number of voiced segments per second (voiced rate)
avgdurV	Average duration of voiced segments
stddurV	Standard deviation of the duration of voiced segments
silrate	Number of silence segments per second (silence rate)
avgdurSil	Average duration of silence segments
stddurSil	Standard deviation of the duration of the silence segments
avgF_0	Average fundamental frequency
stdF_0	Standard deviation of the fundamental frequency
maxF_0	Maximum value of the fundamental frequency
avgE	Average energy
stdE	Standard deviation of the energy
maxE	Maximum value of the energy

Features related to duration

Several duration measures are computed from the speech signal. The complete list of the prosody features extracted in *NeuroSpeech* is displayed in Table 2.

Implementation for developers

The prosody analysis is performed with a python script called `prosody.py`, which is stored in the folder `/prosody/`. The syntax to perform the analysis is as follows:

```
python prosody.py <file_audio>
<filef0.txt> <fileEn.txt> <file_fea-
tures.txt> <path_base>
```

where <file_audio> is the audio file to be analyzed, <filef0.txt> is a file with the result of the F_0 contour, <fileEn.txt> contains the energy contour, <file_features.txt> is the file that will contain the features described above, and <path_base> is the path where the script is stored. The script creates also a figure with the contours of F_0 and energy. A radar-type figure with the prosody features is also created for comparison purposes w.r.t. the reference speakers.

Fig. 8 shows the contours of the fundamental frequency and energy for a healthy speaker (left) and for a PD patient (right). These figures corresponds to results obtained from recordings of the read text. Note that both contours are more stable for the healthy speaker than for the PD patient.

Fig. 9 contains the radar-type plots obtained for a healthy speaker (left) and for a PD patient (PD). Note that most of the extracted features of the healthy speaker (green area) are inside the reference (blue area), only the standard deviation and the average value of the energy are slightly below the reference. Conversely, the PD patient (right) shows higher values of the standard deviation of the duration of silences, and the standard deviation of the duration of voiced segments. Additionally, the standard deviation of the fundamental frequency is also lower than the reference, which confirms the monotonicity of dysarthric speech. These two features are commonly mentioned in the literature as typical indexes of dysarthric speech.

Implementation for end users

When the user clicks on the prosody button of the main window of *NeuroSpeech* the software enables the prosody analysis. Fig. 10 displays an example. The window is divided into two fields. On the left hand side there are three figures, the speech signal in the time domain, the F_0 contour, and the energy contour. On the right hand side there is a list with the twelve prosodic features that were listed above. The obtained values and the references are displayed. Those values that exceed the reference ones are in red. Finally, a radar-type figure is also displayed in order to allow the user to make quick analyses and comparisons.

2.1.5. Diadochokinetic analysis

The DDK evaluation is commonly used to assess dysarthric speech. The speech tasks associated with DDK include the repetition of syllables like /pa/, /ta/, /ka/, or combinations of them like /pa-ta-ka/ or /pe-ta-ka/. The main purpose of these evaluations is to assess the capability of the speaker to move several articulators of the vocal tract including tongue, lips, and velum [18]. The DDK analysis performed in *NeuroSpeech* considers several speech tasks including the repetition of /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/.

Features extracted to perform the DDK analysis

The measures calculated upon DDK tasks are mainly related with changes in energy, pause rate, DDK regularity, duration, and others. The complete list of measures extracted in the DDK analysis is presented in Table 3.

Implementation for developers

The DDK analysis is performed with a python script called `DDK.py`, which is stored in the folder `/DDK/`. The syntax to perform the analysis is as follows:

```
python DDK.py <file_audio> <filef0.txt>
<fileEn.txt> <file_features.txt>
 <path_base>
```

where <file_audio> is the audio file to be analyzed, <filef0.txt> is a file with the result of the F_0 contour, <fileEn.txt> contains the energy contour, <file_features.txt> is the file that will contain the features described above, and <path_base> is the path where the script is stored. The script creates also a figure with the contours of F_0 and energy. A radar-type figure with the prosody features is also created for comparison purposes w.r.t. the reference speakers.

Fig. 11 shows the contours of the fundamental frequency and energy for a healthy speaker (left) and for a PD patient (right). These figures corresponds to results obtained from recordings of the rapid repetition of the syllables /pa-ta-ka/. Note that both contours are more stable for the healthy speaker than for the PD patient.

Fig. 12 contains the radar-type plots obtained for a healthy speaker (left) and for a PD patient (PD). Note that the extracted features of the healthy speaker (green area) are inside the reference (blue area). Conversely, the PD patient (right) shows higher values in six measures, confirming observations reported by other scientists about the deviation of DDK tasks exhibited by PD patients compared to healthy speakers.

Implementation for end users

When the user clicks on the DDK button of the main window of *NeuroSpeech* the software enables the DDK analysis. Fig. 13 displays an example. The window is divided into two fields. On the left hand side there are three figures, the speech signal in the time domain, the F_0 contour, and the energy contour. On the right hand side there is a list with the eleven features that were listed above in Table 3. The obtained values and the references are displayed along with the reference values which are obtained from the reference group in PC-GITA according to the sex of the test speaker. Finally, a radar-type figure is also displayed in order to allow the user to make quick analyses and comparisons.

2.1.6. Intelligibility analysis

Intelligibility is related to the capability of a person to be understood by other person or by a system. Intelligibility is deteriorated in patients with neurological disorders, and it causes loss of their communication abilities and produces social isolation specially at advanced stages of the disease, thus it is an important

Please cite this article in press as: J.R. Orozco-Arroyave et al., NeuroSpeech: An open-source software for Parkinson's speech analysis, Digit. Signal Process. (2017), http://dx.doi.org/10.1016/j.dsp.2017.07.004

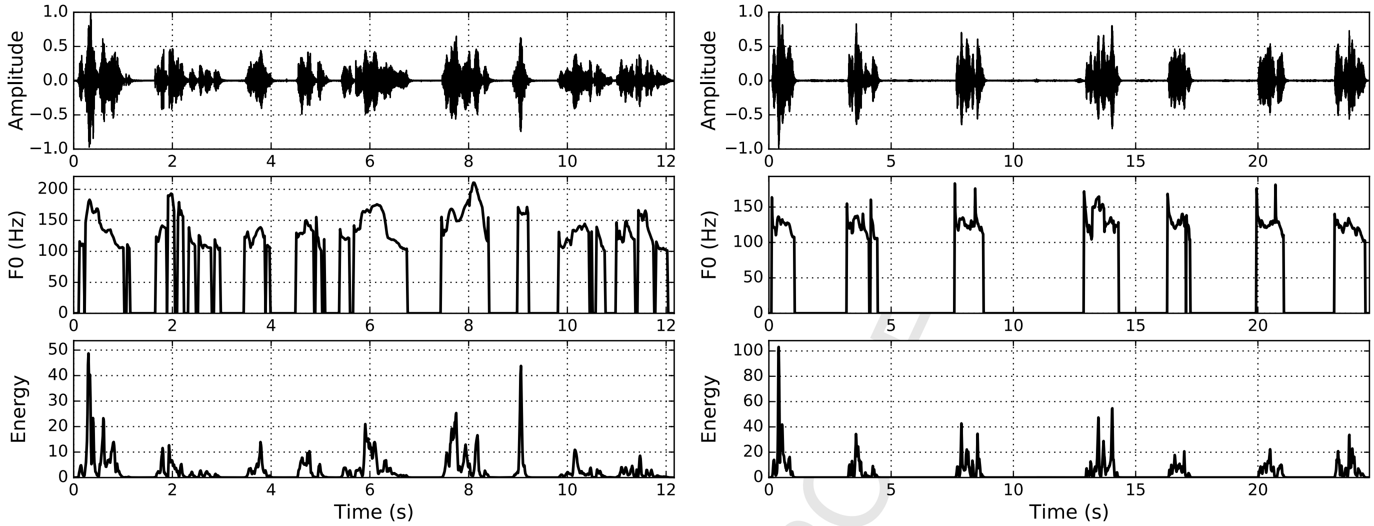

Fig. 8. Contours of the fundamental frequency and energy for a healthy speaker (left), and for a PD patient (right). Speech task: read text.

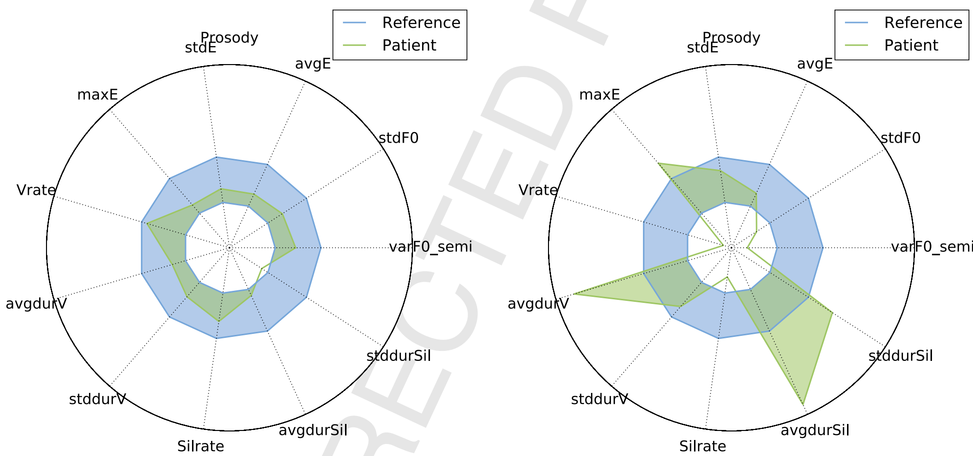

Fig. 9. Radar-type figures for a healthy speaker (left) and for a PD patient (right). (For interpretation of the colors in this figure, the reader is referred to the web version of this article.)

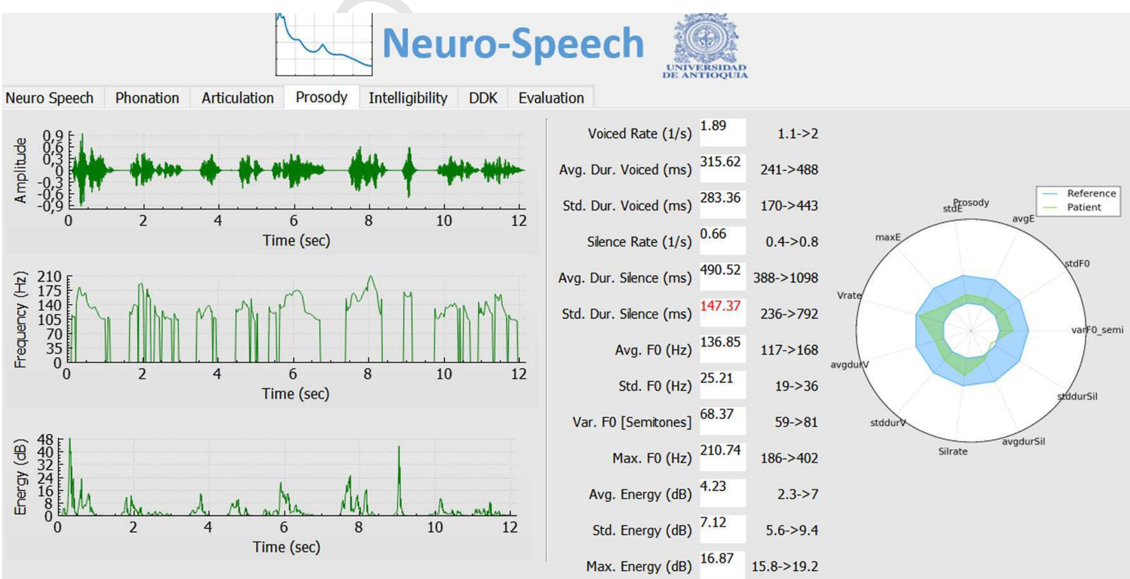

Fig. 10. Window for prosody analysis. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Table 3
DDK features.

Measure	Description
F_0 variability [Semitones]	Variance of the fundamental frequency in semitones
F_0 variability [Hz]	Variance of the fundamental frequency in Hz
Avg. energy [dB]	Average energy
Energy variability [dB]	Variance of energy
Max. energy [dB]	Maximum value of energy
DDK rate	Number of syllables per second
DDK regularity	Variance of the syllables duration
Avg. duration DDK	Average of the syllables duration
Pause rate [1/s]	Number of pauses per second
Avg. duration pause	Average duration of the pauses
Regularity pause	Variance of the pauses duration

marker that deserves attention from medical experts, care givers, patients [31]. In order to help these persons to analyze and monitor this speech dimension, *NeuroSpeech* includes a module to perform intelligibility analyses based on several speech tasks including a set with ten sentences that are read by the patient and a text that contains all of the Spanish sounds that are spoken in Colombia. The analysis is based on the automatic speech recognizer (ASR) provided by Google Inc®. It can be accessed by Internet through https://www.google.com/intl/es/chrome/demos/speech.html, thus this part of the analysis in *NeuroSpeech* needs to have Internet access (unless the developers decides to implement their own ASR, which is also possible).

Two measures are calculated for the intelligibility analysis: the word accuracy (WA), and a similitude measure based on dynamic time warping (sDTW). The measures were introduced in [19] and [32] to model the intelligibility deficits of PD patients.

Word accuracy (WA)

The WA has been established as a marker to analyze the performance of ASR systems and the intelligibility of persons. It has been successfully used to assess intelligibility of people with other kind of speech disorders [33], and the authors indicate that WA can be a good descriptor of speech intelligibility in people with pathological voice. WA is defined as the number of words correctly recognized by the ASR system relative to the total of words in the original string. It is computed with Equation (8).

$$\text{WA} = \frac{\text{\# words correctly recognized}}{\text{\# of total words}} \quad (8)$$

Similarity based on dynamic time warping

DTW is a technique to analyze similarities between two time–series when both sequences may have differences in time and number of samples. It is performed by a time-alignment between the sequences. The DTW distance is computed between the predicted string, i.e., the complete sentence recognized using the ASR system and the original sentence read by the speaker. The distance is computed over the text, at the grapheme level, then the distance is transformed into a similarity score using Equation (9). If the sequences are the same, the $DTW\_distance$ is zero, and the similarity will be 1, conversely if the strings are very different, the $DTW\_distance$ will be high, and the similarity will be close to zero.

$$\text{sDTW} = \frac{1}{1 + \text{DTW\_distance}} \quad (9)$$

Implementation for developers

This analysis is performed with a script called `intelligibility.py`, which is stored in the folder `/intelligibility/`. The syntax to perform the analysis is as follows.

```
python intelligibility.py <file_audio>
<file_txt.txt> <pred_txt.txt>
<file_features.txt>
```

where <file_audio> is the audio file to be analyzed, <file_txt.txt> is the transcription of the audio file, <pred_txt.txt> will contain the string predicted by the ASR, and <file_features.txt> is the file that will contain the intelligibility measures.

A radar figure is also created with the intelligibility features computed from different utterances. Fig. 14 contains the radar figures obtained for the intelligibility analysis of a healthy speaker (left), and for a PD patient (PD). Note the high reduction in the intelligibility of the PD patient, compared to the figure obtained for the healthy speaker.

Implementation for end users

The intelligibility analysis of *NeuroSpeech* is performed upon a set with ten sentences and a text of 36 words that are read by the speakers (see [14] for details of the proposed recording protocol for the analysis of PD speech).

Fig. 15 displays the window of *NeuroSpeech* that is activated after clicking on the intelligibility button. This window shows the

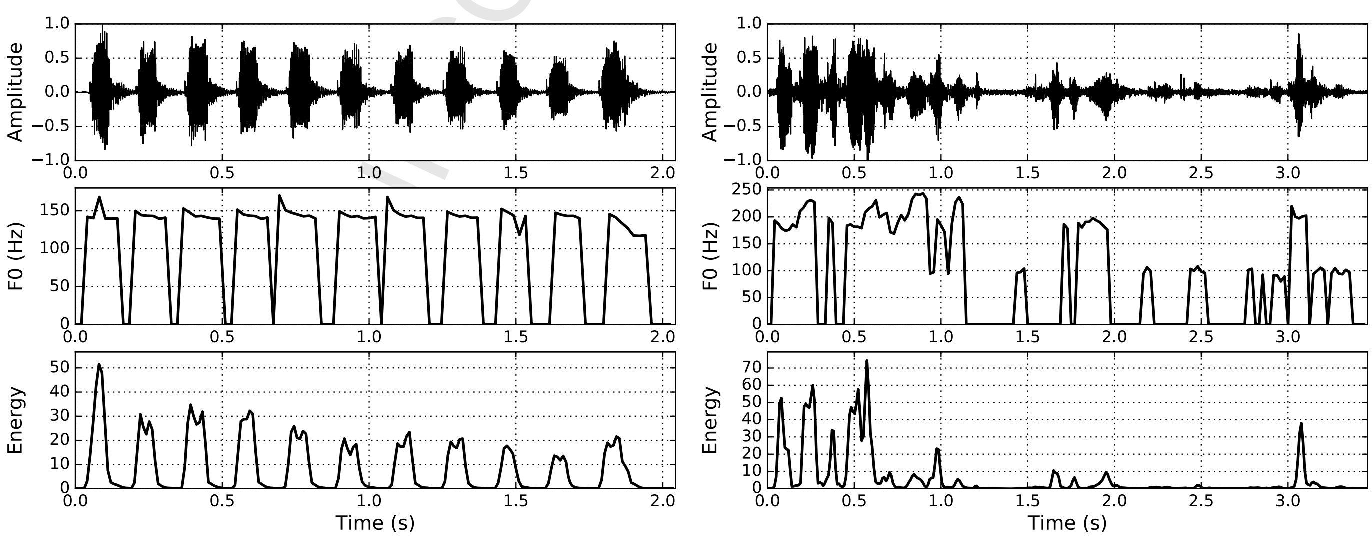

Fig. 11. Contours of the fundamental frequency and energy for a healthy speaker (left), and for a PD patient (right). Speech task: read text.

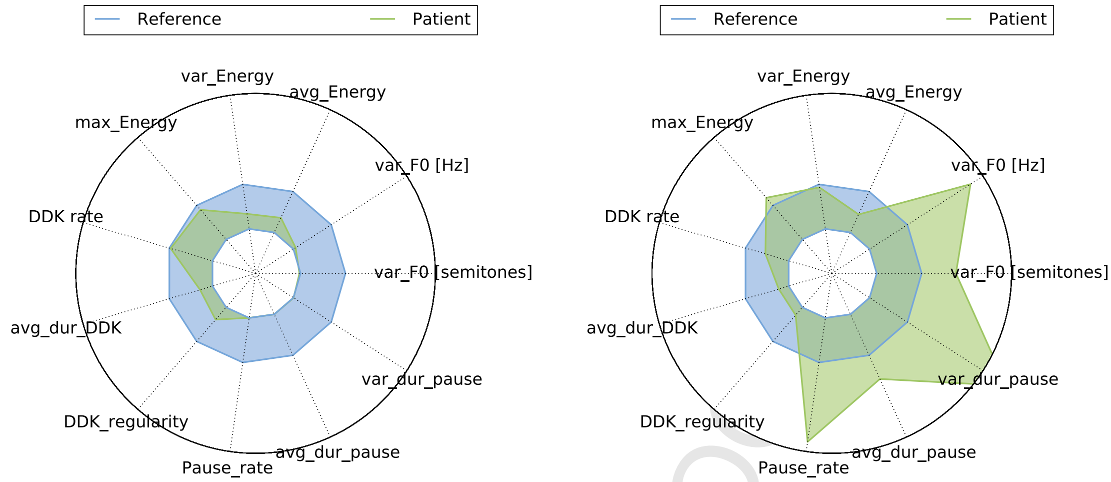

Fig. 12. Radar-type figures for a healthy speaker (left) and for a PD patient (right). (For interpretation of the colors in this figure, the reader is referred to the web version of this article.)

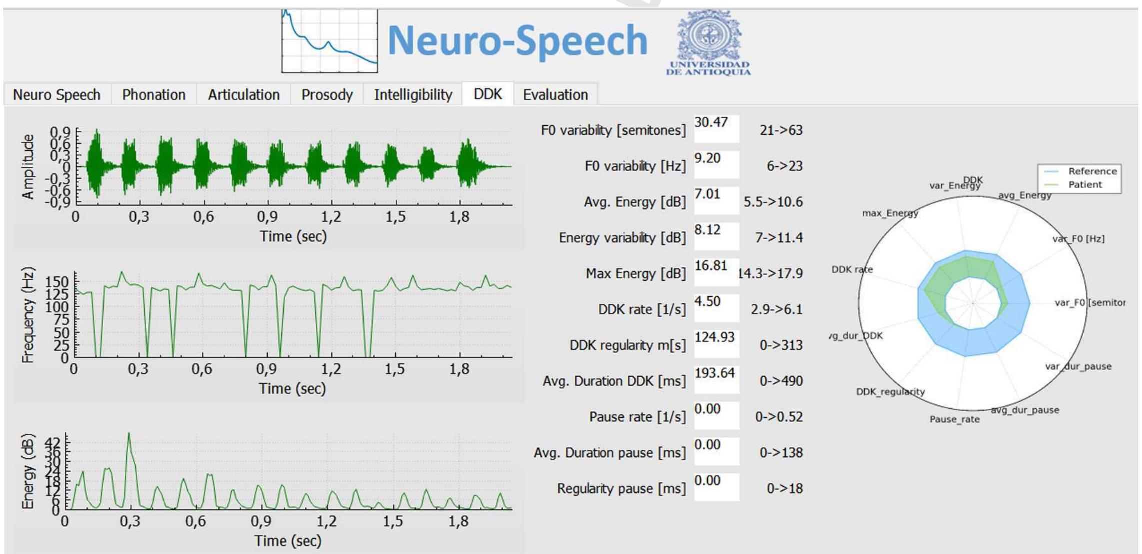

Fig. 13. Window for prosody analysis.

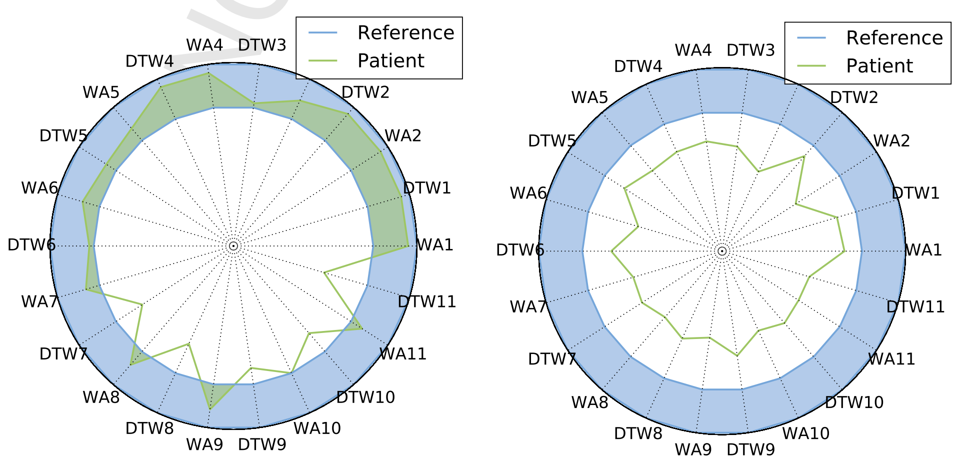

Fig. 14. Radar-type figures for a healthy speaker (left) and for a PD patient (right).

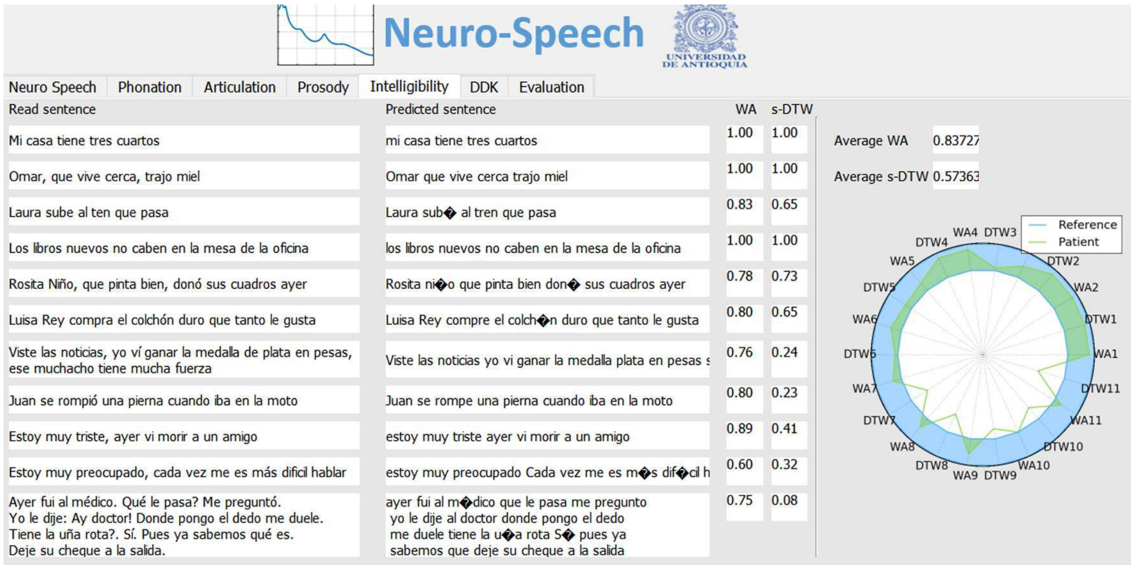

Fig. 15. Window for intelligibility analysis.

Table 4
List of items included in the mFDA-2 evaluation.

Aspect	Item and speech task
Respiration	Sustained vowel /a/ to assess the respiratory capability DDK evaluations to assess the respiratory capability
Lips	DDK evaluations to assess the strength of closing the lips Read text and monologue to assess general capability to control the lips
Palate	Read text and monologue to assess nasal escape DDK evaluations to assess the velar movement
Laryngeal	Sustained vowel /a/ to assess the phonatory capability Read text and monologue to assess the phonatory capability in continuous speech Read text and monologue to assess monotonicity Read text and monologue to assess the effort to produce speech
Tongue	DDK evaluations to assess the velocity to move the tongue Repetition of the syllable /ta/ to assess the velocity to move the tongue
Intelligibility	Read text and monologue to assess intelligibility

original text of the speech tasks (sentences and text) and the texts predicted by the ASR. The results of the WA and DTW are also displayed. On the right hand side of the window, there is a radar-type figure that shows how the intelligibility aspect of the speaker under evaluation is compared w.r.t. the reference. The average values of the DTW similarity and the WA are also displayed.

2.1.7. Neurological state and the dysarthria level assessment

The neurological state of PD patients is typically evaluated according to the Unified Parkinson's Disease Rating Scale (UPDRS) [11]. This scale evaluates motor and non-motor aspects of PD. It is divided into four parts, part one evaluates non-motor experiences of daily living and it has a total of 13 items; part two evaluates motor activities of daily living and it is composed by 13 items; part three includes the motor examination and comprises 33 items; and part four is to assess the motor complications with a total of 6 items. The ratings of each item range from 0 (normal) to 4 (severe), and the total score of each part corresponds to the sum of its items. During a typical neurological evaluation of a PD patient, the medical expert performs the evaluation of the third part of the UPDRS scale (UPDRS-III), thus each patient is typically labeled with a score ranging between 0 and 132 (33×4).

The motor evaluation of the UPDRS has shown to be suitable to assess Parkinson's patients; however, such an evaluation only considers speech in one of its items, thus the deterioration of the communication abilities suffered from PD patients is not properly evaluated. In order to help clinicians, speech and language therapists, patients, and care givers to assess and to monitor the communication abilities of PD patients, we started recently the development of an adapted version of the Frenchay Disathria Assessment (FDA-2) [12]. The original version of the FDA-2 considers several factors that are affected in people suffering from dysarthria, such as reflexes, respiration, lips movement, palate movement, laryngeal capability, tongue posture/movement, intelligibility, and others. Although this tool covers a wide range of aspects, it requires the patient to be in front of the examiner. In our case it is not possible to have a new appointment with the patients, thus we only have access to the recordings captured several months or years ago. In order to perceptually evaluate such recordings following a tool similar to the FDA-2, we introduced the modified FDA (mFDA-2). This scale includes the following aspects of speech: respiration, lips movement, palate/velum movement, larynx, tongue, and intelligibility. The m-FDA is administered considering different speech tasks including sustained vowels, rapid repetition of the syllables /pa-ta-ka/, read texts, and monologues. It has a total of 13 items and each of them ranges from 0 (normal or completely healthy) to 4 (very impaired), thus the total score of the mFDA-2 ranges from

Please cite this article in press as: J.R. Orozco-Arroyave et al., NeuroSpeech: An open-source software for Parkinson's speech analysis, Digit. Signal Process. (2017), http://dx.doi.org/10.1016/j.dsp.2017.07.004

0 to 42. Table 4 shows details of the items included in the mFDA-2 evaluation.

NOTE: this is a preliminary version and it is still under review, evaluation, and validation. We decided to include it in this paper because we consider that it could be interesting for the research community to see how they can include other reference labels or scales into *NeuroSpeech*, thus the community can train different models for the evaluation of other diseases and use the framework of our software to perform specialized assessments.

According to the rating scales introduced above, there are two different labels to be predicted per patient, i.e., UPDRS-III and mFDA-2. The prediction of the UPDRS-III labels is performed using a support vector regressor (SVR) like in [19] and [24]. For the case of mFDA-2, as each aspect of speech is evaluated with particular speech tasks and the resulting label is formed by a set of six sub-scores, we decided to perform the evaluation of each sub-score using a multi-class support vector machine (SVM).

The process of labeling the speech recordings of the PD patients in PC-GITA (reference set) was performed by three phoniatricians. They were asked in the beginning to agree in the first ten evaluations, then each expert performed the evaluation of the remaining recordings independently. The inter-rater reliability among the labelers is 0,75. Regarding the labeling of the neurological state of the patients, they were evaluated by one Neurologist certified by the Movement Disorders Society to perform such kind of evaluations. All of the predictions performed in *NeuroSpeech* are based on the criteria of the medical experts that supported this research.

Implementation for developers

The prediction of the mFDA-2 and the UPDRS-III scores is performed with a python script called `predictPD.py`, which is stored in the folder `/evaluation/`. The syntax to perform the prediction is as follows:

```
Python predictPD.py <path_base>
<file_audio>
```

where `<file_audio>` is the audio file to be analyzed and `<path_base>` is the folder where the script is stored. The script will generate a file called `<pred.txt>` in the folder `<path_base>`. The file contains a total of eight fields: the total score of mFDA-2, the six items of the scale, and the predicted score of the UPDRS-III scale.

If the user wants to re-train the models to predict the UPDRS-III or the mFDA-2 scores using another features or speech tasks, the following script has to be executed as follows:

```
Python TrainSVRNeuroSpeech.py
<file_matrix.txt> <file_labels.txt>
<file_ scaler.obj> <fileSVR.obj>
```

where `<file_matrix.txt>` is a txt file with the feature matrix, the `txt` file is named as `<file_ labels.txt>` and it contains the labels/scores originally assigned by the experts (in the folder there are files for the UPDRS-III `<labelsUPDRS.txt>` and for the mFDA-2 scores `<labelsmFDA.txt>`). `<file_scaler.obj>` is an output file that contains an object with the mean value and standard deviation which can be used for the standardization of the feature matrix, and `<fileSVR.obj>` will contain the re-trained SVR. The `<file_scaler.obj>` has to be named as `<scalerUPDRS.obj>` or `<scalermFDA.obj>` for the prediction of the UPDRS-III or the mFDA-2 scores, respectively. The file with the trained SVR has to be named as `<SVRtrainedUPDRS.obj>` for the prediction of the UPDRS-III, and for the prediction of the mFDA-2 scores the required name is `<SVRtrainedmFDA.obj>`.

To re-train the multi-class SVMs for the prediction of the mFDA-2 sub-scales the following script needs to be followed:

```
Python TrainSVMNeuroSpeech.py <file_ma-
trix.txt> <sub-scale> <file_labels.txt>
<file_scaler.obj> <fileSVM.obj>
```

Note that the script `<TrainSVMNeuroSpeech.py>` may be executed similarly to `<TrainSVRNeuroSpeech.py>`, adding the parameter `<sub-scale>`, which refers to the sub-scale that will be trained, i.e., `<r>` for respiration, `<l>` for lips, `<p>` for palate, `<x>` for larynx, `<t>` for tongue, and `<i>` for intelligibility.

Implementation for end users

This part of the analysis can be performed by the user by clicking on the evaluation button of the main window of *NeuroSpeech*. Fig. 16 shows an example of the information that is displayed after performing the analysis. The left hand side of the evaluation window displays the result of each speech aspect evaluated in the mFDA-2 scale. Besides, the total score of the predicted mFDA-2 and the predicted value of the UPDRS-III score are included. There are two figures on the right hand side of the window, one displays a histogram with the values of the mFDA-2 assigned by the experts to the people in PC-GITA, and the other one displays a histogram of the values of the UPDRS-III scores assigned by the neurologists to the same population. The predicted values of the speaker under evaluation are displayed as a red line painted on the histogram bars.

3. Generation of the report

As it was mentioned before, in order to help the patients, clinicians, and care givers, *NeuroSpeech* is able to generate a report with the results of the evaluation of phonation, articulation, prosody, and intelligibility. An example of this report is in the same folder of the software documentation.[4]

4. Contributions of *NeuroSpeech* to the state of the art

A new software for the analysis of pathological speech signals is presented in this paper. The software is based on state-of-the-art methods that have been validated in previous publications. In this paper we have presented a configuration of the software to analyze voice recordings of Parkinson's patients; however, the same platform can be easily adapted to perform the analysis of other voice pathologies like dysphonia or hypernasality. Four different aspects or dimensions of speech can be modeled using *NeuroSpeech*: phonation, articulation, prosody, and intelligibility. The analysis of pathological speech signals based on such a "division" by dimensions allows the user to make specific conclusions regarding different aspects of the speech production process. This characteristic makes *NeuroSpeech* a very good option for clinicians, patients, and care givers to assess pathological voices.

NeuroSpeech will contribute the state of the art in several aspects including, (1) its methods are based on state of the art techniques which enables the research community to address up-to-dated tests in a very easy way, (2) it can be used by patients, clinicians, and computer scientists. Patients can give feedback about its usability, clinicians can contribute with the interpretation of results, and computer scientists can contribute with new methods,

[4] https://github.com/jcvasquezc/NeuroSpeech.

Please cite this article in press as: J.R. Orozco-Arroyave et al., NeuroSpeech: An open-source software for Parkinson's speech analysis, Digit. Signal Process. (2017), http://dx.doi.org/10.1016/j.dsp.2017.07.004

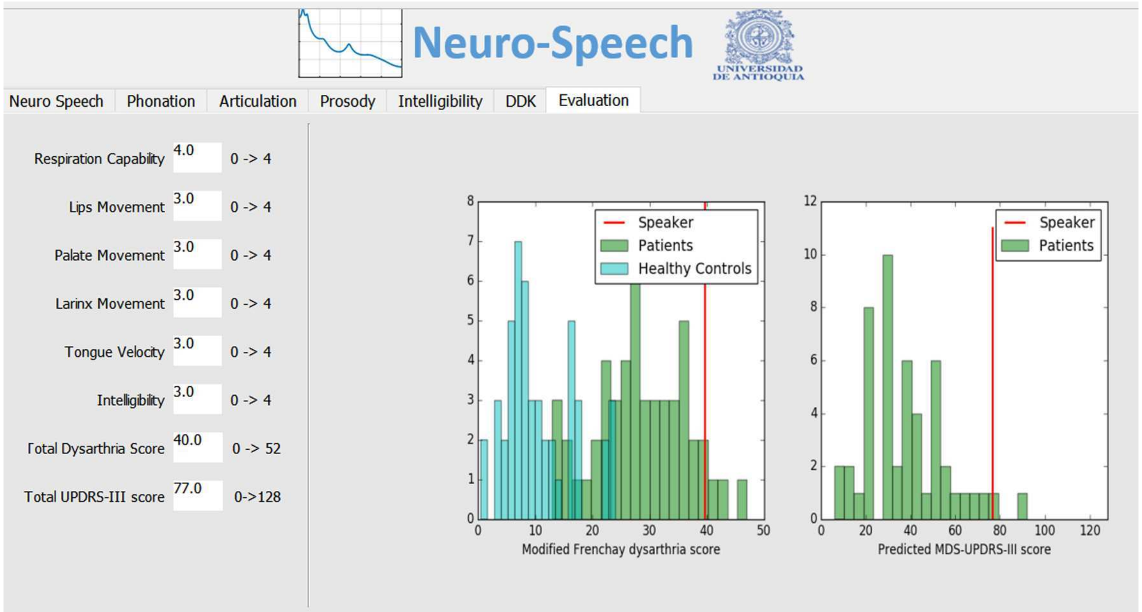

Fig. 16. Window for the evaluation. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

(3) this software is freely accessible and open source, and (4) to the best of our knowledge, this is the first attempt to launch an easy to use software, freely accessible and open source to help people of different nature (clinicians, patients, or scientists) to perform their own analyses and to make their own conclusions regarding different aspects of speech pathologies.

We are currently working on two different fields, the implementation of more measures and pattern recognition techniques, and the construction and characterization of databases with other neurological diseases, such that in the near future we expect to include models of more pathologies in *NeuroSpeech*.

Acknowledgments

This project was funded by COLCIENCIAS (project No. 111556933858), and also by CODI at Universidad de Antioquia (grant No PRG-2015-7683). Additionally, this work was partially developed during the 2016 Jelinek Memorial Summer Workshop on Speech and Language Technologies, which was supported by Johns Hopkins University via DARPA LORELEI, Microsoft, Amazon, Google, and Facebook.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.dsp.2017.07.004.

References

[1] J. Logemann, H. Fisher, B. Boshes, E. Blonsky, Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients, J. Speech Hear. Disord. 43 (1978) 47–57.
[2] F.L. Darley, A.E. Aronson, J.R. Brown, Differential diagnostic patterns of dysarthria, J. Speech Hear. Res. 12 (1969) 246–269.
[3] National Collaborating Centre for Chronic Conditions, Parkinson's Disease: National Clinical Guideline for a Diagnosis and Management in Primary and Secondary Care, Royal College of Physicians, London, 2006.
[4] P.F. Worth, How to treat Parkinson's disease in 2013, Clin. Med. 12 (1) (2013) 93–96.
[5] C. Vaquero, O. Saz, E. Lleida, J. Marcos, C. Canalís, VOCALIZA: an application for computer-aided speech therapy in Spanish language, in: Proceedings of IV Tecnologías del habla, 2006, pp. 321–326.
[6] P. Kitzing, A. Maier, V. Ahlander, Automatic speech recognition (ASR) and its use as a tool for assessment or therapy of voice, speech, and language disorders, Logop. Phoniatr. Vocology 34 (2009) 91–96.
[7] A. Maier, T. Haderlein, U. Eysholdt, F. Rosanowski, A. Batliner, M. Schuster, E. Nöth, PEAKS – a system for the automatic evaluation of voice and speech disorders, Speech Commun. 51 (5) (2009) 425–437.
[8] O. Saz, S.-C. Yin, E. Lleida, R. Rose, C. Vaquero, W. Rodríguez, Tools and technologies for computer-aided speech and language therapy, Speech Commun. 51 (10) (2009) 948–967.
[9] S. Pentiuc, O. Schipor, M. Danubianu, M. Schipor, I. Tobolcea, Speech therapy programs for a computer aided therapy system, Electron. Electr. Eng. 103 (7) (2010) 87–90.
[10] Y.-P. Chen, C. Johnson, P. Lalbakhsh, T. Caelli, G. Deng, D. Tay, S. Erickson, P. Broadbridge, A. Refaie, W. Doube, M. Morris, Systematic review of virtual speech therapists of speech disorders, Comput. Speech Lang. 37 (2016) 98–128.
[11] C.G. Goetz, et al., Movement Disorder Society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results, Mov. Disord. 23 (15) (2008) 2129–2170.
[12] P. Enderby, R. Palmer, Frenchay Dysarthria Assessment, 2nd edition, Pro-ed, 2008.
[13] M. Novotný, Automated Assessment of Diadochokinesis and Resonance in Dysarthrias Associated with Basal Ganglia Dysfunction, Ph.D. thesis, Faculty of Electrical Engineering, Czech Technical University in Prague, 2016.
[14] J.R. Orozco-Arroyave, J.D. Arias-Londoño, J.F. Vargas-Bonilla, M.C. Gonzalez-Rátiva, E. Nöth, New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease, in: Proceedings of the 9th LREC, ERLA, 2014, pp. 342–347.
[15] A. Goberman, Correlation between acoustic speech characteristics and non-speech motor performance in Parkinson disease, Med. Sci. Monit. 11 (3) (2005) CR109–16.
[16] J.R. Orozco-Arroyave, E. Belalcázar-Bolaños, J.D. Arias-Londoño, J.F. Vargas-Bonilla, S. Skodda, J. Rusz, K. Daqrouq, F. Hönig, E. Nöth, Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases, IEEE J. Biomed. Health Inform. 19 (6) (2015) 1820–1828.
[17] J.R. Orozco-Arroyave, F. Hönig, J.D. Arias-Londoño, J.F. Vargas-Bonilla, S. Skodda, J. Rusz, E. Nöth, Voiced/unvoiced transitions in speech as a potential biomarker to detect Parkinson's disease, in: Proceedings of the 16th INTERSPEECH, ISCA, 2015, pp. 95–99.
[18] J.R. Orozco-Arroyave, F. Hönig, J.D. Arias-Londoño, J.F. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, E. Nöth, Automatic detection of Parkinson's disease in running speech spoken in three different languages, J. Acoust. Soc. Am. 139 (1) (2016) 481–500.
[19] J.R. Orozco-Arroyave, J.C. Vásquez-Correa, F. Hönig, J.D. Arias-Londoño, J.F. Vargas-Bonilla, S. Skodda, J. Rusz, E. Nöth, Towards an automatic monitoring of the neurological state of Parkinson's patients from speech, in: Proceedings of the 41st ICASSP, IEEE, 2016, pp. 6490–6494.
[20] S. Hadjitodorov, P. Mitev, A computer system for acoustic analysis of pathological voices and laryngeal diseases screening, Med. Eng. Phys. 24 (6) (2002) 419–429.
[21] S. Sapir, L.O. Raming, J.L. Spielman, C. Fox, Formant centralization ratio (FCR): a proposal for a new acoustic measure of dysarthric speech, J. Speech Lang. Hear. Res. 53 (1) (2010) 1–20.
[22] S. Skodda, W. Visser, U. Schlegel, Vowel articulation in Parkinson's disease, J. Voice 25 (4) (2011) 467–472, Erratum in J. Voice. 25 (2) (2012 Mar) 267–268.

[23] J.R. Orozco-Arroyave, E.A. Belalcázar-Bolaños, J.D. Arias-Londoño, J.F. Vargas-Bonilla, T. Haderlein, E. Nöth, Phonation and articulation analysis of Spanish vowels for automatic detection of Parkinson's disease, in: Lecture Notes in Artificial Intelligence, vol. 8655, Springer-Verlag, 2014, pp. 389–396.
[24] J.R. Orozco-Arroyave, Analysis of Speech of People with Parkinson's Disease, Logos Verlag, Berlin, 2016.
[25] P. Boersma, Praat, a system for doing phonetics by computer, Glot Int. 5 (9/10) (2002) 341–345.
[26] E. Zwicker, E. Terhardt, Analytical expressions for critical-band rate and critical bandwidth as a function of frequency, J. Acoust. Soc. Am. 68 (5) (1980) 1523–1525.
[27] E. Nöth, A. Batliner, V. Warnke, J. Haas, M. Boros, J. Buckow, R. Huber, F. Gallwitz, M. Nutt, H. Niemann, On the use of prosody in automatic dialogue understanding, Speech Commun. 36 (1–2) (2002) 45–62.
[28] V. Zeißler, J. Adelhardt, A. Batliner, C. Frank, E. Nöth, R. Shi, H. Niemann, The prosody module, in: SmartKom: Foundations of Multimodal Dialogue Systems, Cognitive Technologies, Springer, 2006, pp. 139–152.
[29] B. Schuller, A. Batliner, S. Steidl, D. Seppi, Recognising realistic emotions and affect in speech: state of the art and lessons learnt from the first challenge, Speech Commun. 53 (9–10) (2011) 1062–1087.
[30] B. Schuller, A. Batliner, Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing, John Wiley & Sons, United Kingdom, 2015.
[31] National Parkinson Foundation, Guidelines for Speech-Language Therapy in Parkinson's Disease, Nederlandse Vereniging voor Lodopedie en Foniatrie, Nijmegen, 2010.
[32] J.C. Vasquez-Correa, J.R. Orozco-Arroyave, E. Nöth, Word accuracy and dynamic time warping to assess intelligibility deficits in patients with Parkinson's disease, in: XXI Symposium on Image, Signal Processing and Artificial Vision, STSIVA, 2016.
[33] M. Schuster, A. Maier, T. Haderlein, E. Nkenke, U. Wohlleben, F. Rosanowski, U. Eysholdt, E. Nöth, Evaluation of speech intelligibility for children with cleft lip and palate by means of automatic speech recognition, Int. J. Pediatr. Otorhinolaryngol. 70 (2006) 1741–1747.

Juan Rafael Orozco-Arroyave received his Ph.D. in Computer Science from the University of Erlangen (Germany) and currently he is Professor at University of Antioquia (Colombia). He is also adjunct researcher at the University of Erlangen (Germany). His main research interests include speech processing, pattern recognition, and machine learning.

Juan Camilo Vásquez-Correa received his Ms.C. from the University of Antioquia (Colombia) and currently he is Ph.D. student from the same University and from the University of Erlangen (Germany). His main research interests include speech processing, pattern recognition and machine learning.

Jesús Francisco Vargas-Bonilla received his Ph.D. Telecommunications and Cybernetics from the University of Las Palmas in Gran Canaria (Spain). Currently he is the Dean of the School of Engineering at the University of Antioquia (Colombia). His main research interests include biometry, pattern recognition, and machine learning.

Rama Arora received his Ph.D. from the University of Wisconsin-Madison. He is member of the center for language and speech processing (CLSP) and the institute for data intensive engineering and science (IDIES) at Johns Hopkins University (USA). His main research interests include big-data processing, machine learning, and pattern recognition.

Najim Dehak received his Ph.D. from School of Advanced Technology, Montreal (Canada). He is well known as a leading developer of the I-vector representation for speaker recognition. Currently he is Assisstant Professor at the center for language and speech processing (CLSP) in Johns Hopkins University (USA). His main research interests include speech processing, machine learning, and pattern recognition.

Phani S. Nidadavolu is Ph.D. student from the center for language and speech processing (CLSP) in Johns Hopkins University (USA). His main research interests include speech processing, machine learning, and pattern recognition.

Heidi Christensen receive her Ph.D. from Aalborg University Denmark. Currently she is a lecturer in Computer Science and a member of the Centre for Assistive Technology and Connected Healthcare (CATCH) at the University of Sheffield. She is part of the Speech and Hearing Research Group (SPandH) and a member of the Sheffield Medical Humanities (MHS). Her main research interests include pathological speech processing, pattern recognition, and machine learning.

Frank Rudzicz received his Ph.D. in computer science from the University of Toronto (Canada), he is Scientist at the Toronto Rehabilitation Institute and an Assistant Professor (status only) in the Department of Computer Science at the University of Toronto. His main research interests include pathological speech processing, machine learning, and pattern recognition.

Maria Yancheva is a M.Sc. student in the Computational Linguistics group in Artificial Intelligence, Department of Computer Science at the University of Toronto. Her main research interests include speech processing and natural language processing.

Hamidreza Chinaei received his Ph.D. in machine learning from Laval University (Canada). Currently he is Postdoctoral Fellow in Computer Science at the University of Toronto. His main research interests include natural language processing, dialog systems, and machine learning.

Alyssa Vann is M.Sc. student in computer science from Standford University (USA). Her main research interests include natural language processing and machine learning.

Nicola Vogler is M.Sc. student in computer science from the University of California-Irvine (USA). His main research interests include natural language processing and machine learning.

Tobias Bocklet received his Ph.D. in humanities from the Medical Faculty at the University of Erlangen (Germany). Currently he is product architect and researcher at Intel Corporation in Munich (Germany). His main research interests include speech processing, pattern recognition and machine learning.

Milos Cernak received his Ph.D. in Computer science from Slovak University of Technology in Bratislava (Slovakia). Currently he is Research Associate at Idiap Research Institute, in Martigny (Switzerland). His main research interests include speech recognition and synthesis, low bit rate speech coding, prosody parametrization, and pathological speech processing.

Julius Hannik received his B.Sc. in Physics from the University of Gótingen (Germany) and currently he is Ph.D. student from the University of Erlangen (Germany). His main research interests include movement analysis, pattern recognition, and machine learning.

Elma Nöth received his Ph.D. from the University of Erlangen (Germany). He is Professor at the computer science department from the same university. He is the head of the speech processing and understanding group. His main research interests include speech processing, pattern recognition, and machine learning.

XMLVIEW: extended

Appendix A. Supplementary material

The following is the Supplementary material related to this article.

begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent begin ecomponent

`Label:` MMC 1
`caption:` SoftwareX impact statement: general description of the NeuroSpeech software and details of the current code and executable software versions.
`link:` **APPLICATION : mmc1**

end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent end ecomponent

Sponsor names

Do not correct this page. Please mark corrections to sponsor names and grant numbers in the main text.

COLCIENCIAS, *country*=Colombia, *grants*=111556933858

Johns Hopkins University, *country*=United States, *grants*=