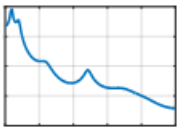


Neuro-Speech

User manual version 1.1

Juan Camilo Vasquez-Correa
Juan Rafael Orozco-Arroyave
Jesus Francisco Vargas-Bonilla
Elmar Nöth

General description

Neuro-Speech is an open source software platform designed to perform speech analysis of people with neuro-degenerative disorders. Particularly patients with Parkinson's disease. The software is designed to be used by medical examiners such as speech therapists and neurologists, but it can also be used by patients to perform the analysis, and by general population interested in the analysis of pathological speech.

The software computes several measures to evaluate the communication capabilities of the patients and includes analyses of phonation, articulation, prosody, and intelligibility. The software calculates also specific bio–markers related to the dysarthria levels of the patients and perform the prediction of the Movement disorder society–Parkinson's disease rating scale, part III (MDS-UPDRS-III), which is a general evaluation of the motor capabilities of the patients.

All of the results obtained are compared to reference patterns obtained with a group of 50 healthy control speakers.

Juan Camilo Vasquez-Correa
Faculty of Engineering, University of Antioquia UdeA, Medellín, Colombia. and Pattern Recognition Lab, Friedrich–Alexander Universität, Erlangen–Nuremberg, Germany, e-mail: jcamilo.vasquez@udea.edu.co

Juan Rafael Orozco-Arroyave
Faculty of Engineering, University of Antioquia UdeA, Medellín, Colombia. and Pattern Recognition Lab, Friedrich–Alexander Universität, Erlangen–Nuremberg, Germany, e-mail: rafael.orozco@udea.edu.co

Jesus Francisco Vargas-Bonilla
Faculty of Engineering, University of Antioquia UdeA, Medellín, Colombia. e-mail: jesus.vargas@udea.edu.co

Elmar Nöth
Pattern Recognition Lab, Friedrich–Alexander Universität, Erlangen–Nuremberg, Germany, e-mail: elmar.noeth@fau.de

Finally a medical report can be generated, which describes the different speech deficits of the patients, and how the measures are deviated respect to those computed with information from healthy controls

1 Requeriments and Running

Neuro–Speech is a software platform designed in C++, which runs python scripts in background for the speech analysis. The software uses some third party software that can be freely downloaded and installed for the correct operation of Neuro-Speech. The list of the third party software that must be installed previous to Neuro–Speech is as follows:

1. Ananconda: Python environment. It can be installed from https://www.continuum.io/downloads
2. Praat: software for speech analysis. Available at http://www.fon.hum.uva.nl/praat/
3. ffmpeg: a solution to record, convert and stream audio and video. Available at http://ffmpeg.org/download.html

To execute Neuro-Speech, please go to the folder *Release*, and then click in the icon of PDTool.exe. Then the main window shown in Figure 2 is displayed for the analysis.

2 Main Window

Figure 2 shows the main window of Neuro-Speech. It contains buttons to record and play the speech signals. The recording can be visualized also in this window. The main window contains also six different buttons to perform each one of the speech analysis: (1) phonation, (2) articulation, (3) prosody, (4) diadochokinetic (DDK), (5) intelligibility, (6) PD evaluation, and the last button to create the medical report. Table 1 details the description of each part of the main window.

2.1 First name

This field is to input the first name of the patient. A folder will be created in the directory *data*. The name of the folder is formed with the current date, the first name, and the last name of the patient. The recorded audios will be stored in this folder.

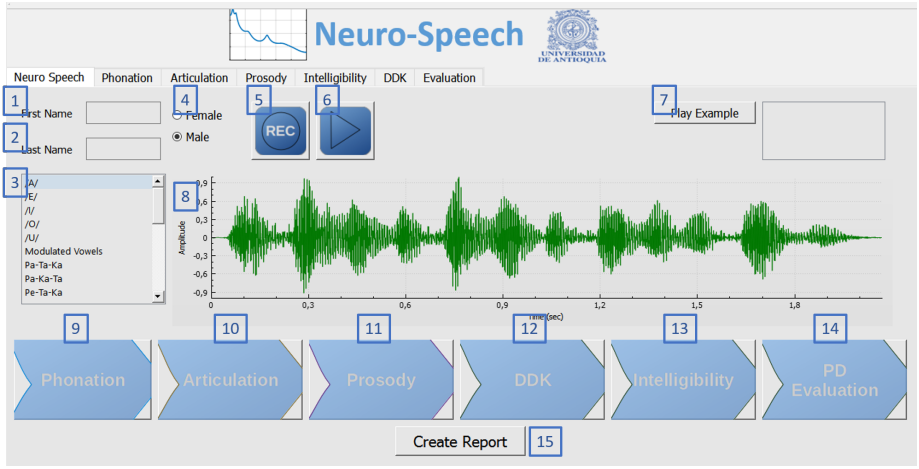


Fig. 1 Main Window of Neuro-Speech

Table 1 Description of each part of the main window

Button	Description
1	Input of the first name of the patient
2	Input of the last name of the patient
3	List of tasks for the speech recordings (one of them must be selected to record)
4	Select the gender of the speaker
5	Record button
6	Play button
7	Play an example of the task to record
8	Field to visualize the speech signal after recording
9	Perform the phonatory analysis
10	Perform the articulatory analysis
11	Perform the prosody analysis
12	Perform the DDK analysis
13	Perform the intelligibility analysis
14	Perform the dysarthria and PD evaluation
15	Generate the medical report

2.2 Last name

This field is to input the last name of the patient.

2.3 List of Tasks

This field contains a list of all the tasks that can be recorded. It contains 25 different tasks, which are described in Table 2. The selected task will be recorded.

Table 2 List of tasks for speech recording

Number	Task	Description
		Sustained Vowels
1	/A/	Sustained phonation of vowel A
2	/E/	Sustained phonation of vowel E
3	/I/	Sustained phonation of vowel I
4	/O/	Sustained phonation of vowel O
5	/U/	Sustained phonation of vowel U
6	Modulated Vowels	The five vowels pronounced with Kaiser effect
		Diadochokinetic evaluation
7	Pa-Ta-Ka	rapid repetitions of syllables pa-ta-ka
8	Pa-Ka-Ta	rapid repetitions of syllables pa-ka-ta
9	Pe-Ta-Ka	rapid repetitions of syllables pe-ta-ka
10	Pa	rapid repetitions of syllable pa
11	Ta	rapid repetitions of syllable ta
12	Ka	rapid repetitions of syllable ka
		Isolated words
13	Words	25 isolated words
		Continuous Speech
14	Sentence 1	”Mi casa tiene tres cuartos”
15	Sentence 2	”Omar, que vive cerca trajo miel”
16	Sentence 3	”Laura sube al tren que pasa”
17	Sentence 4	”Los libros nuevos no caben en la mesa de la oficina”
18	Sentence 5	”Rosita Nio, que pinta bien, donó sus cuadros ayer”
19	Sentence 6	”Luisa rey compra el colchón duro que tanto le gusta”
20	Sentence 7	”Viste las noticias, yo vi ganar la medalla de plata en pesas, ese muchacho tiene mucha fuerza”
21	Sentence 8	”Juan se rompió una pierna cuando iba en la moto”
22	Sentence 9	”Estoy muy triste, ayer vi morir a un amigo”
23	Sentence 10	”Estoy muy preocupado, cada vez me es ms dificil hablar”
24	Read Text	read text with 36 words *
25	Monologue	The patient says What he/she does in a normal day

* Read text: Ayer fu al médico. Qué le pasa? Me preguntó
Yo le dije: Ay doctor. Donde pongo el dedo me duele.
Tiene la uña rota? Sí. Pues ya sabemos que es.
Deje su cheque a la salida.

2.4 Gender

The correct radio button must be selected depending on the gender of the speaker with the aim of selecting appropiate reference for the analysis.

2.5 Record

This button starts to record the task selected from the task list. The speech utterances are recorded by default with a sampling frequency of 16 kHz, and 16–bit resolution. After finishing the recording, the audio file is saved with the name of the task in a directory created with the name of the patient and the current date in the *data* directory.

2.6 Play

This button plays the current recorded audio.

2.7 Play Example

This button plays an audio example about what the patient should pronounce in each speech task.

2.8 Signal Visualization

The space in the middle of the main window shows the recorded signal. Additional to the visualization, segments of the audio can be zoomed with the mouse scroll, clipped by selecting an initial and end time with double click.

2.9 Phonation

This button performs the phonatory analysis on sustained vowels. More details in section 3

2.10 Articulation

This button performs the articulatory analysis both on sustained vowels and continuous speech. More details in section 4

2.11 Prosody

This button performs the prosody analysis from the monologue or the read text. More details in section 5

2.12 DDK

This button allows to analyze dedicated speech tasks such as the rapid repetiton of the syllables /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, /ka/. Additional information in section 6

2.13 Intelligibility

This button performs the intelligibility analysis from the ten sentences and the read text. More details in section 7

2.14 PD Evaluation

This button performs the analysis of the dysarthric levels of the PD patients, and perform the prediction of the MDS-UPDRS-III. More details in section 8

2.15 Create Report

This button creates a medical report in *.pdf format that contains all the measures computed from the speech of the patients relative to the phonation, articulation, prosody, intelligibility and dysarthric levels of the speakers.

3 Phonation Analysis

Phonation is related with the capability of the speaker to make the vocal folds vibrate, and it has been analyzed in terms of features related to perturbation measures such as jitter, shimmer, the amplitude perturbation quotient, the pitch perturbation quotient, and non-linear dynamics measures.

The phonation analysis is based mainly in perturbation measures of the fundamental frequency of the speech, and the energy. The phonation analysis is computed on utterances from sustained vowels (mainly /A/). The analysis is performed when the user clicks in the phonation button of the main window. If a recording was performed, the analysis is performed on the utterance of the sustained vowel /A/, else the user should choose a speech recording from the hard drive (it must be a sustained vowel). Figure 3 shows the window for the phonation analysis.

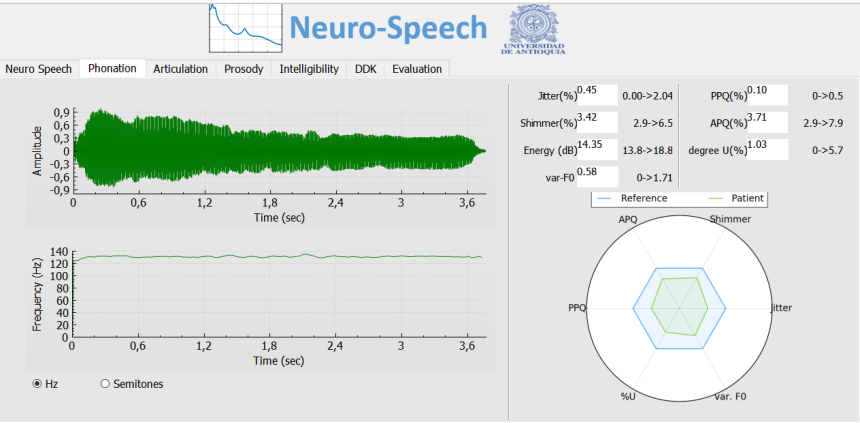


Fig. 2 Phonation analysis in Neuro-Speech

The window is divided in two fields. At the left, the user can observe the speech recording in the time-domain, and the contour of the fundamental frequency. The stability during phonation process can be analyzed through the fundamental frequency. The different measures computed from the speech recording can be observed in the right field. The computed measures include the jitter (temporal perturbation of the fundamental frequency), the shimmer (temporal perturbation of the amplitude of the signal), the APQ and PPQ (long term perturbation measures from amplitude and pitch), the degree of unvoiced (percentage of non–periodic phonation), and the variablity of the fundamental frequency in semitones. Low and High reference values of the measures are included at the right side of the value obtained for the recording. Different reference are obtained for male and female speakers.

When the value of the features are out of the normal range, the value is highlighted in red.

At the right–bottom side there is a radar plot that contains normalized values of the measures computed with the speech recording (green), and the average computed with recordings from healthy control speakers (blue). Each corner of the polygon corresponds to one perturbation measure. If one corner of the green polygon is out of bounds of the blue polygon, the measure associated with such corner is out of the normal reference.

4 Articulation Analysis

Articulation is related with the modification of the position, stress, and shape of several limbs and muscles to produce the speech, and it has been described with features such as the vowel space area, the vowel articulation index, the formant centralization ratio, and the energy in the voiced/unvoiced transition.

The articulation analysis computes measures such as the formant frequencies with the aim of model the movement of the vocal tract when the speaker pronounce the five Spanish vowels. Measures such as the vowel space area (VSA), vocal pentagon area (VPA), and formant centralization ratio (FCR) are also computed to evaluate the articulation capabilities of the speakers. The articulation capabilities of the patients in continuous speech is evaluated considering the energy content both in the transition from unvoiced to voiced segments (onset), and in the transition from unvoiced to voiced segments (offset). Figure 4 shows the Articulation analysis window.

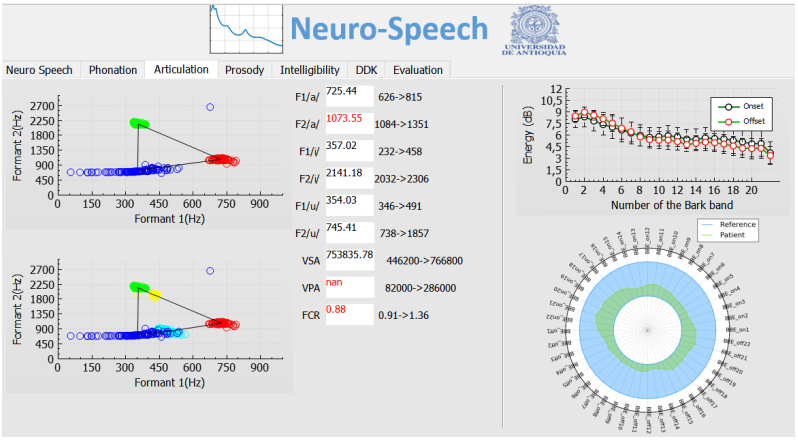


Fig. 3 Articulation analysis in Neuro-Speech

The articulation analysis is performed when the user clicks in the articulation button of the main window. If speech signals were recorded, the analysis is performed on utterances from the five Spanish vowels, and the monologue, else the user should choose speech utterances from the five vowels (A, E, I, O, U), and the monologue, in that order.

The left part of the window shows the articulation analysis from sustained vowels. It contains figures of the vocal triangle, and the vocal pentagon. The window shows also the average value of the formant frequencies for each vowel, the VSA, the VPA, and the FCR, which model the articulatory capabilities of the patients. From the figures of the vowel space can be analyzed the dispersion of the formant frequencies in the sustained vowels.

The right part of the window contains the articulation analysis in continuous speech. The speech signal is segmented into voiced and unvoiced frames, and the energy content in the border between such frames is calculated according to the Bark scale. The window includes the energy of onset (transition from unvoiced to voiced), and offset (transition from voiced to unvoiced), and the radar plot of such energies for the patient (green), and the average for healthy controls (blue). When the green figure is completely contained in the blue figure, the energy the measures are in the normal range.

5 Prosody Analysis

Prosody reflects variation of loudness, pitch, and timing to produce natural speech and it is commonly evaluated with features related to the fundamental frequency, the energy contour, and the speech rate.

The prosody analysis is performed with the read text, or the monologue. Figure 5 shows the Prosody analysis window.

The prosody analysis compute measures related to duration, speech rate, fundamental frequency and energy. The left part of the window shows the speech signal, the contour of the fundamental frequency, and the contour of the energy of the signal. In the middle there is the values of the computed measures, and include the following measures:

- The voiced rate: how many voiced segments appear per second; it is a measure related to the speech velocity.
- The average duration of voiced segments.
- The standard deviation of the duration of voiced segments.
- The silence rate: how many silence segments appear per second; it is a measure also related to the speech velocity.
- The average duration of silence segments.

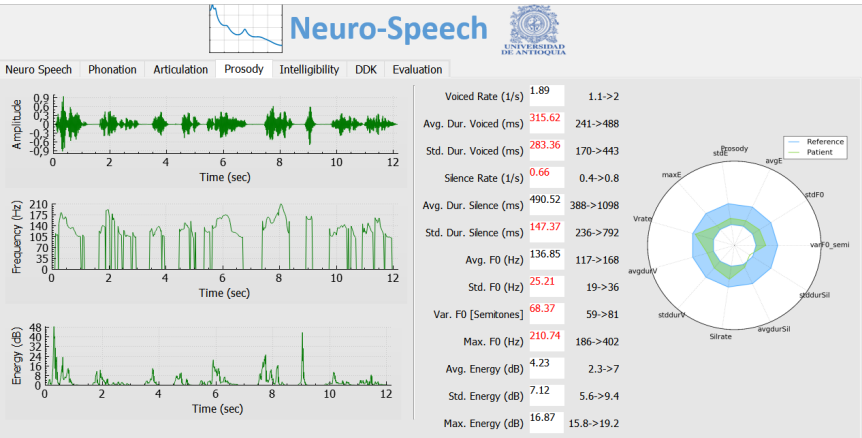


Fig. 4 Prosody analysis in Neuro-Speech

- The standard deviation of the duration of the silence segments.
- The average fundamental frequency.
- The standard deviation of the fundamental frequency.
- The maximum value of the fundamental frequency.
- The variablity of the fundamental frequency in semitones
- the average energy.
- The standard deviation of the energy.
- The maximum value of the energy.

Finally, the prosody analysis also includes the radar plot of the computed measures (green), and the comparison with the healthy controls (blue). When the green figure is out of bounds of the blue figure, there is difference between the value obtained for the speaker, and the value for healthy controls.

6 Diadochokinetic Analysis (DDK)

DDK analysis consists in the rapid repetition of syllables such as /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, and /ka/ to analyze the capability of the speaker to move articulators such as the velum, the jaw, and the tongue.

The DDK analysis is performed with the rapid repetition of the syllables described above. Figure 6 shows the DDK analysis window.

The DDK analysis compute measures related to duration, speech rate, fundamental frequency and energy on the DDK tasks. The left part of the window shows the

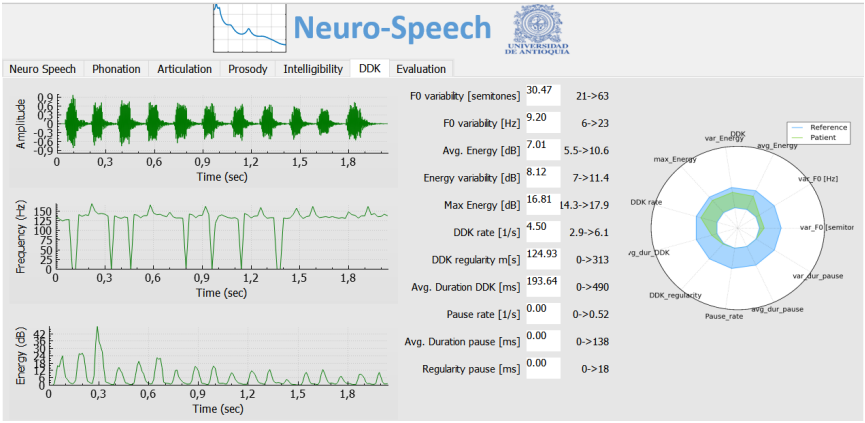


Fig. 5 DDK analysis in Neuro–Speech

speech signal, the contour of the fundamental frequency, and the contour of the energy of the signal. In the middle there is the values of the computed measures, and include the following measures:

- The variability of the fundamental frequency measured in semitones.
- The variability of the fundamental frequency measured in Hz.
- the average energy of the DDK task.
- The variablity of the energy along the utterance.
- The maximum value of the energy.
- The DDK rate: how many syllables are uttered by second.
- The DDK regularity, which is measured as the variability of the duration of each syllable.
- The average duration of the syllables.
- The pause rate: how many silence segments appear per second.
- The average duration of pauses.
- The regularity of pauses, which is measured as the variability of the duration of the pauses.

The DDK analysis also includes the radar plot of the computed measures (green), and the comparison with the healthy controls (blue). When the green figure is out of bounds of the blue figure, there is difference between the value obtained for the speaker, and the value for healthy controls.

7 Intelligibility Analysis

Intelligibility is related to the capability of a person to be understood by other person or by a system. Intelligibility is also deteriorated in patients with neurological disorders causing loss of communication abilities and social isolation specially at advanced stages of the disease.

The intelligibility analysis is performed with the ten read sentences and with the read text, using the automatic speech recognizer provided by Google Inc. (This analysis requires Internet connection). Figure 7 shows the Intelligibility analysis window.

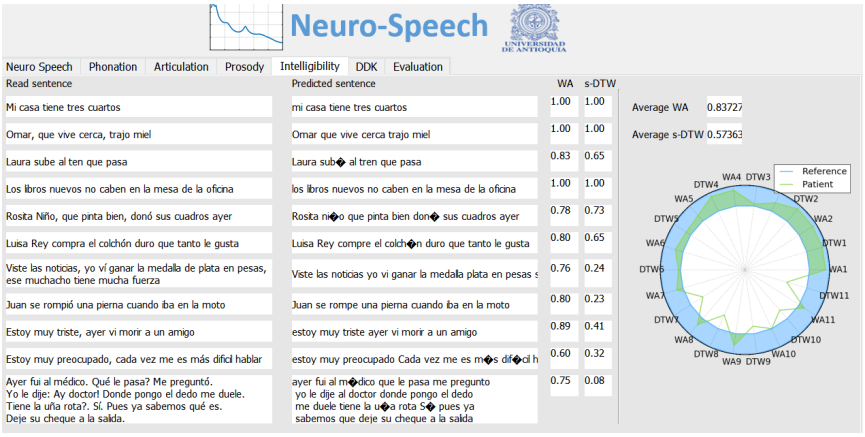


Fig. 6 Intelligibility analysis in Neuro-Speech

The window for intelligibility analysis contains at the left field the original read sentence by the patients, and the predicted sentence by the automatic speech recognizer. There are also two measures to evaluate the intelligibility capabilities of the patients: the word accuracy (WA), which measure how many words are correctly recognized, and a similarity score computed with the dynamic time warping (DTW), which is more robust than the WA due to the comparison in DTW is performed at phoneme level, rather than at word level. Finally, the radar plot is also included as in the previous analyses.

8 PD evaluation

Additionally to the UPDRS score, we introduce a modified version of the Frenchay Dysarthric Assessment score. The main aim was to evaluate only the speech impairments that the patients develop. The original version of the FDA needs the patient to be with the examiner. We introduced a modified version that considers only the speech recordings and evaluates 13 items including among others the movements of the lips, larinx, palate and tongue, the respiration, and the intelligibility. The evaluation of each item ranges from 0 to 4, for a total range from 0 to 52 (0 normal, and 52 completely dysarthric)

The dysarthric level of the patients is analyzed with the repetition of pa-ta-ka. The analysis includes the assessment of the respiration capability, the lips movement, the palate movement, the larinx movement, the tongue velocity, the intelligibility, and the total dysarthria score. The software also performs a prediction of the MDS-UPDRS-III score of the patients. Figure 8 shows the window for the analysis.

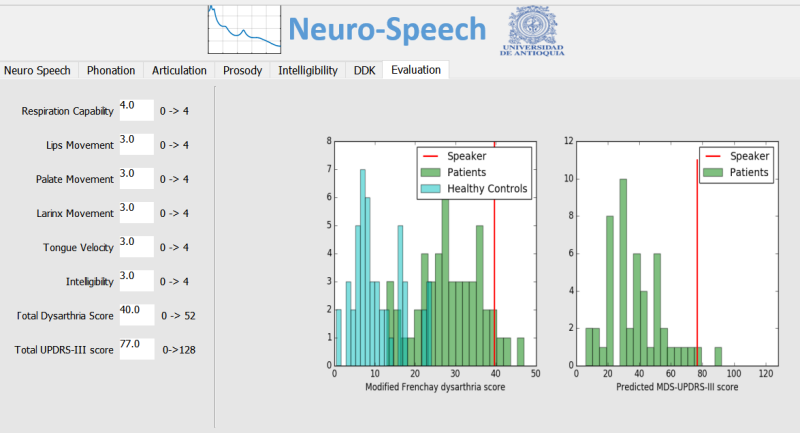


Fig. 7 Dysarthria and MDS-UPDRS-III analysis in Neuro-Speech

9 Report Generation

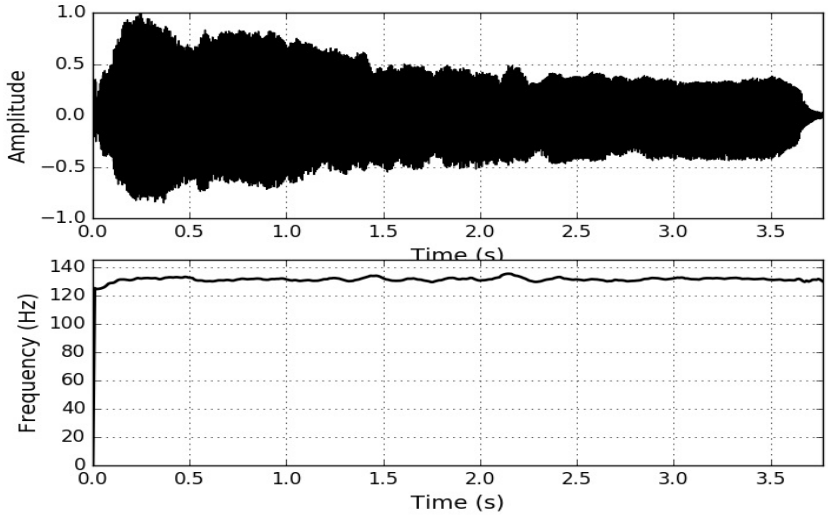
After performing all the analyses, a medical report can be generated. The report summarizes the obtained results and include the analysis of phonation, articulation, prosody, intelligibility and dysarthria. An example of the report is included at the end of this manual.

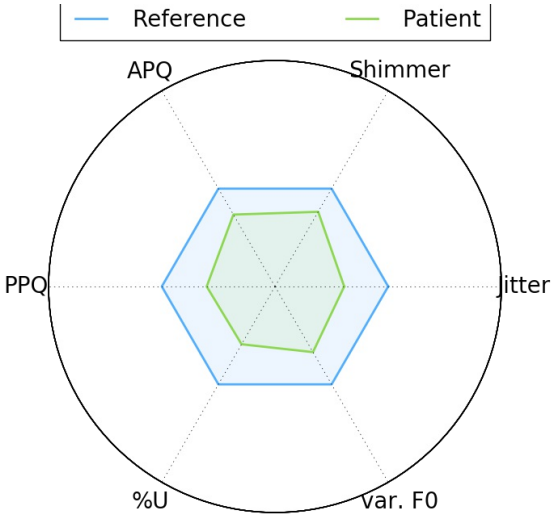
Neuro-Speech

1. Patient

- First Name:
- Last Name:

2. Phonation Analysis

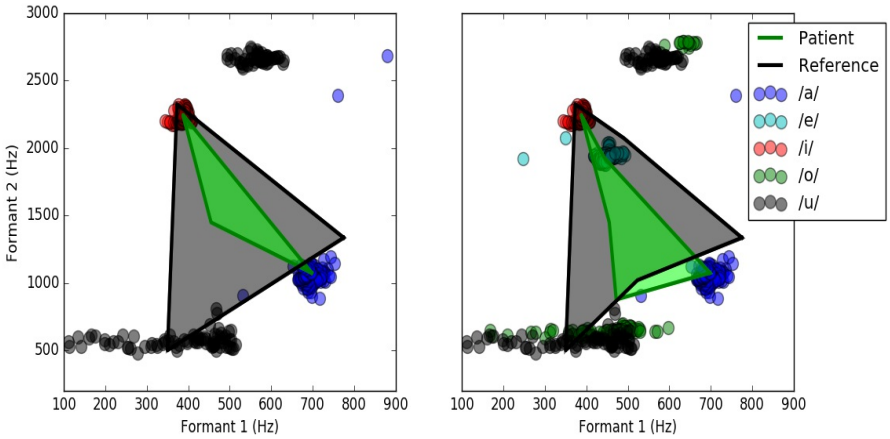


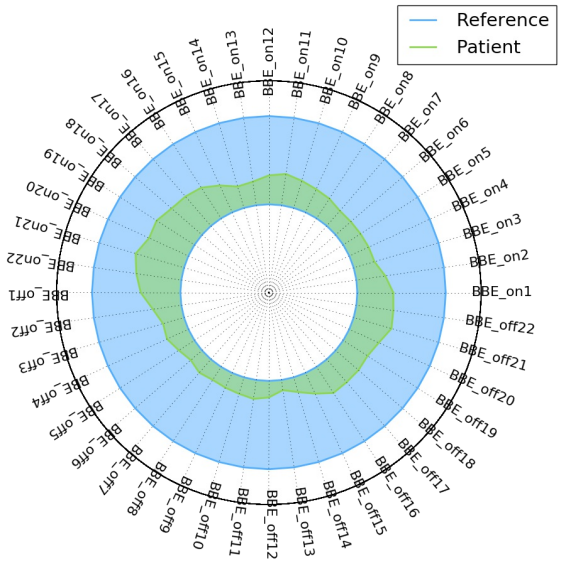


Feature	Value	Reference for Male	Reference for Female
Jitter (%)	0.45	0.00→ 2.04	0.00→ 1.32
Shimmer (%)	3.42	2.90→ 6.50	2.40→ 6.40
APQ	3.71	2.90→ 7.90	1.90→ 8.70
PPQ	0.10	0.00→ 0.50	0.00→ 0.77
degree U (%)	1.03	0.00→ 5.70	0.27→ 3.60

Feature	Value	Reference for Male	Reference for Female
variance F0 (semitones)	0.58	0.00→ 1.70	0.00→ 2.69

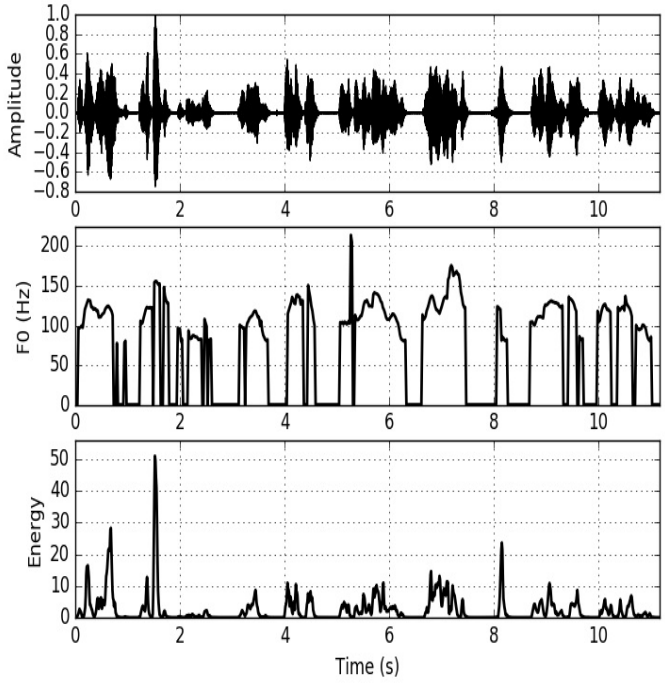
2. Articulation Analysis

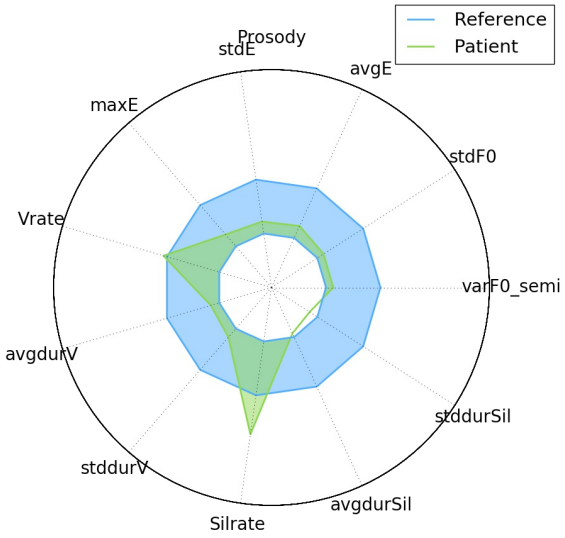




Feature	Value	Reference for Male	Reference for Female
F1/a/ (Hz)	697.57	625→ 815	688→ 974
F2/a/ (Hz)	1075.15	1084→ 1351	1340→ 1573
F1/i/ (Hz)	387.59	232→ 458	331→ 471
F2/i/ (Hz)	2244.92	2032→ 2306	2303→ 2666
F1/u/ (Hz)	455.09	368→ 494	346→ 491
F2/u/ (Hz)	1449.03	738→ 1857	545→ 1316
VSA (Hz2)	616226.15	446200→ 766800	806400→ 1206800
FCR	1.14	0.91→ 1.36	0.82→ 1.13

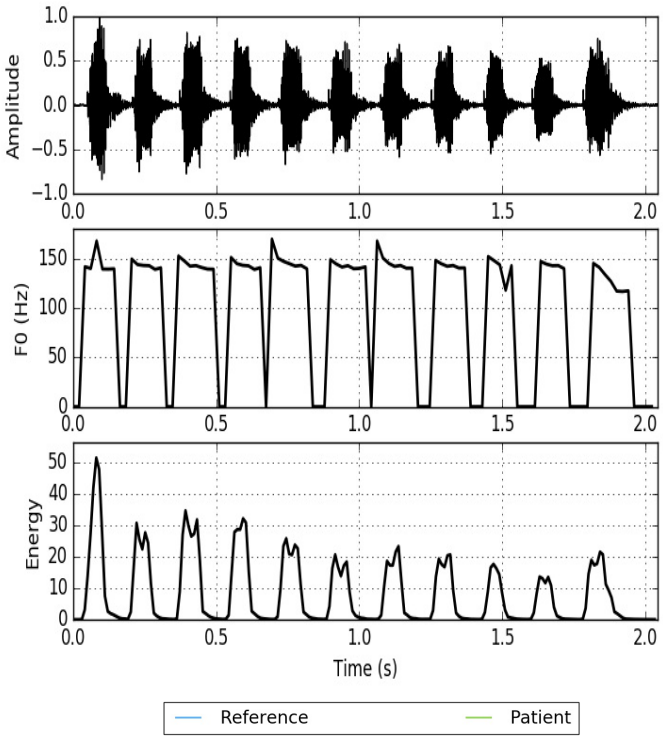
3. Prosody Analysis

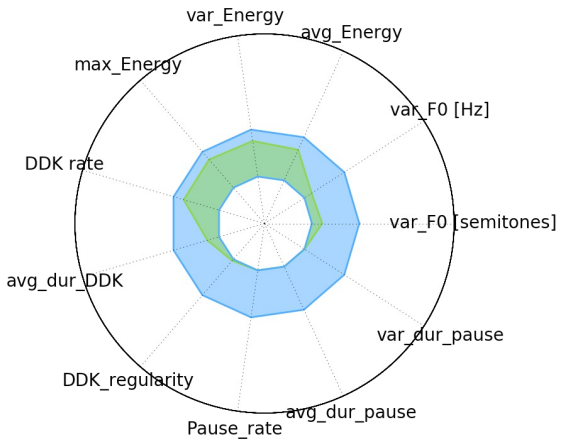




Feature	Value	Reference for Male	Reference for Female
Voiced Rate (1/s)	2.06	1.1→ 2.0	1.3→ 2.1
Avg. Dur. Voiced (ms)	280.82	241→ 488	272→ 463
Std. Dur. Voiced (ms)	226.77	170→ 443	200→ 377
Silence Rate (1/s)	1.07	0.4→ 0.8	0.4→ 0.7
Avg. Dur. Silence (ms)	334.80	388→ 1098	360→ 835
Std. Dur. Silence (ms)	142.83	236→ 792	64→ 722
Avg. Energy (dB)	3.97	2.3→ 7.0	1.9→ 6.8
Std. Energy (dB)	6.77	5.6→ 9.4	5.4→ 9.3
Avg. F0 (Hz)	116.06	117→ 168	191→ 240
Std. F0 (Hz)	21.26	19→ 36	30→ 47

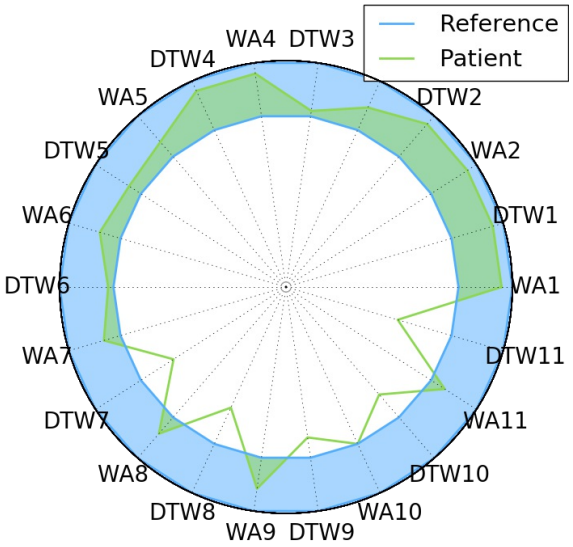
4. DDK Analysis





Feature	Value	Reference for Male	Reference for Female
DDK Rate (1/s)	5.38	3.4→ 6.2	2.9→ 6.1
Avg. Dur. DDK (ms)	125.53	20→ 490	37→ 341
DDK regularity (ms)	12.40	0→ 313	0→ 220
Pause Rate (1/s)	0.00	0→ 0.52	0→ 0.32
Avg. Dur. Silence (ms)	0.00	388→ 1098	360→ 835
Std. Dur. Silence (ms)	0.00	236→ 792	64→ 722
Avg. Energy (dB)	9.12	5.5→ 10.6	4.2→ 8.8
Std. Energy (dB)	10.34	7.0→ 11.4	6.6→ 9.9
Var. F0 (Hz)	9.10	6→ 23	8→ 29
Var. F0 (semitones)	30.09	21→ 63	31→ 73

5. Intelligibility Analysis



Feature	Value	Reference
Word Accuracy	0.837273	0.85→ 1.0
Sim DTW	0.573636	0.9→ 1.0

5. Dysarthria Level (m-FDA)

Feature	Value	Reference
Respiration capability	0.0	0→ 4
Lips movement	0.0	0→ 4
Palate movement	0.0	0→ 4
Larinx movement	0.0	0→ 4
Tongue velocity	0.0	0→ 4
Intelligibility	0.0	0→ 4
Total m-FDA	10.0	0→ 52
Total MDS-UPDRS-III	6.0	0→ 128