A Preliminary Study of MLSE/ACE-III Stages for Primary Progressive Aphasia Automatic Identification Using Speech Features

Amable J. Valdés Cuervo[1](✉), Elena Herrera[2], and Enrique A. de la Cal[1](✉)

[1] Computer Science Department, Faculty of Geology, University of Oviedo, Oviedo, Spain
{UO232486,delacal}@uniovi.es

[2] Psychology Department, Faculty of Psychology, University of Oviedo, Oviedo, Spain
herreraelena@uniovi.es

Abstract. Primary Progressive Aphasia (PPA) is a syndrome causing progressive deterioration of language and speech due to brain degeneration. Three variants exist: non-fluent variant (nfvPPA), semantic variant(svPPA) and logopenic variant (lvPPA). While fMRI is the most accepted diagnostic tool (and neurological exploration), it is expensive and takes even months to deliver results. Cheaper and faster tools are needed for earlier diagnosis and treatment initiation. Some studies have attempted automatic diagnosis using acoustic and linguistic features with ML and DL techniques. However, none have included Latin language patients or analyzed the effect of cognitive tests. This work proposes a methodology based on three main steps: i) a new assessment tool (PPATool) combining ACE-III and MLSE with three language tasks: verbal fluency, repetition and naming, and ii) an IDA process to obtain an ML model trained with our own two-class (PPA/Healthy) dataset, and iii) ranking the relevance of tasks in PPATool from models performance. The results obtained after deploying the IDA process on the dataset obtained from an early-stage clinical trial, show that the verbal fluency data outperforms the rest of the tasks.

Keywords: Primary Progressive Aphasia · ACE-III · MLSE · Voice silence removal · Machine Learning Classification · voice features · MFCC · Imbalanced datasets

1 Introduction

Primary progressive aphasia (PPA) is a syndrome characterized by a progressive deterioration of language and speech due to the degeneration of language-related brain systems. Three variants of PPA have been identified: non-fluent

P. García Bringas et al. (Eds.): SOCO 2023, LNNS 750, pp. 323–333, 2023.
https://doi.org/10.1007/978-3-031-42536-3_31

variant PPA (nfvPPA), semantic variant PPA (svPPA), and logopenic variant PPA (lvPPA). Diagnosis is typically made by a specialized neurologist using subjective complaints, observations during the examination, clinical criteria, and cognitive/neuroimaging tests. However, cheaper and faster tools to support clinical diagnosis are needed for earlier disease diagnosis and treatment initiation.

There are already some studies that have worked on the approach to the automatic diagnosis of PPA, Fraser et al., 2014 [6], Hoffman et al., 2017 [8], Cho et al., 2020 [4] and Themistocleous et al., 2021[15]. Most of them include extracting acoustic and/or linguistic features, different classical Machine Learning (ML) and Deep Learning (DL) techniques, datasets with up to 100 English speakers, and tackling two-class (PPA-Healty) and three-class (lvPPA-svPPA-nfvPPA) problems. But, none included Latin language patients in the study nor analysed the effect of the typology of cognitive tests in the performance of the ML technique.

This work proposes a methodology based on three main pillars: 1) design and deployment of a new assessment tool specific for PPA (PPA-Tool) combining the screening test ACE-III and the cognitive test MLSE, 2) development of an Intelligent Data Analysis (IDA) process trained with a set of prosodic and spectral transformations from the recorded tasks to obtain an ML model capable of classifying between PPA and Healthy patients, and 3) ranking the relevance of the different tasks defined in the PPA-Tool driven by the performance of the models obtained in the previous pillar. The proposed methodology aims to provide a cheaper and faster tool for supporting clinical diagnosis and treatment initiation, ultimately improving and accelerating the management of PPA patients.

This work is arranged in the following sections; the next section describes the methodology proposed, including the design of proposed examination tests and the IDA process to obtain the classification models. The experimental setup and discussion of the obtained results can be found in Sect. 3. Finally, conclusions and future work are included.

2 Proposed Methodology

The proposed methodology is composed of three main steps (see Fig. 1): i) Designing of the assessment tool (PPA-Tool), ii) Intelligent data analysis, and iii) Ranking the PPA-Tool' tasks.

2.1 Design of the Assessment Tool

Two tests were selected for assessing the participants: the ACE-III, a cognitive screening test widely used for the neuropsychological evaluation of patients and the Mini Linguistic State Examination (MLSE), a specific PPA test.

The ACE-III is a test in which all cognitive domains (orientation, attention, memory, visuospatial skills, executive functions and language) are assessed. ACE-III helps to rule out the presence of dementia and also offers a fairly comprehensive cognitive profile

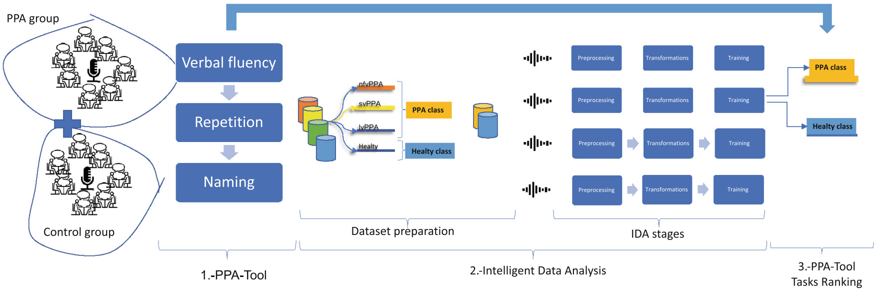

Fig. 1. The overall process of this proposal.

The MLSE is a specific test to classify the different variants of PPA recently developed [11], which evaluates the key linguistic domains affected by PPA according to diagnostic criteria[7]. It consists of eleven subtests that assess the following language skills: naming, word and sentence repetition, word and sentence comprehension, semantic association, reading, writing and a connected speech task. The MLSE is the only PPA-specific test with a Spanish version recently adapted [9].

For the purpose of this report, some of the tasks of the ACE-III and the MLSE were pooled for joint analysis, keeping the following tasks (now on PPA-Tool):

- Fluency: the verbal fluency tasks of the ACE-III.
- Repetition: the repetition of words and sentences of both tests.
- Naming: the picture naming of the MLSE.

The verbal fluency task involves three subtests: Phonological, semantic and actions. In these tasks, participants have to generate as many words of each category as they can in one minute of time (words beginning with “p”, animals and actions, respectively).

The picture naming test is a standard task used to assess language disorders normally. It is based on the presentation of 20 images, and the participant must generate the name of the picture being shown.

Finally, the repetition task involves both word and sentence repetition. In this test, the examiner states each item and the patient is required to repeat them as accurately as possible.

2.2 Intelligent Data Analisys

After carrying out the PPA-Tool tasks, one voice file per patient/participant will be obtained. To obtain an optimal classification model to identify automatically PPA recordings, the following steps are proposed: i) Data preparation, ii) Preprocessing, ii) Voice features, and iv)Classification models training.

Data Preparation: Tasks Segmentation and Labelling. The first step is to separate the audios into smaller audios that correspond to the three different tasks chosen in the PPA-Tool: fluency, naming and repetition. Although the proposed PPA-Tool just considers these 3 tasks, this splitting step could be valid for any number of tasks. In this stage of our research, this process will be deployed manually, but in the future could be run through any automatic segmentation technique.

As the number of patients in this stage of the PPA clinical trial is just six: four PPAs and two controls, a multi-class (the three variants of PPA) problem will be discarded, and just a two-class problem will be dived: PPA and Healthy classes.

Preprocessing: Silence Removing. This study aims to use prosodic and spectral features of voice waveforms to distinguish between patients with PPA impairment and healthy individuals. In long recordings, it is common to encounter sound fragments other than voice, such as unvoiced segments and silence. Since the recordings were conducted in a controlled environment, there were few or no unvoiced events present, necessitating a silence removal algorithm. The most widely accepted algorithms for silence removal rely on thresholding the Short Time Energy (STE) and Zero Crossing Rate (ZCR) features, as well as the statistical behaviour of background noise, as reported in the literature [12] (Fig. 2).

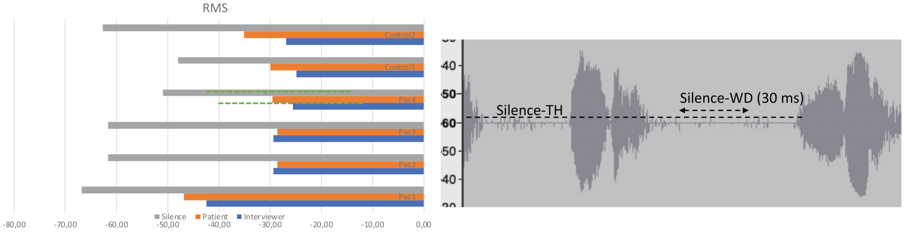

Fig. 2. left) RMS for each kind of sound event, right) Silence threshold computing.

This work proposes a simplified approach based on thresholding the Root Mean Square (RMS) feature. A representative silence reference fragment is manually selected from the beginning of each patient recording, and the mean and standard deviation of the RMS in dBFS are calculated for this segment. Consequently, the silence threshold is determined for each patient recording using Eq. 1.

$$SilenceTH_p = mean(RMS_p) - std(RMS_p) \tag{1}$$

Finally, the calculated SilenceTH is applied using an overlapped window of size Silence-WD, removing any sample windows that exceed this threshold. Figure 3 illustrates an example of silence removal, where the original recording

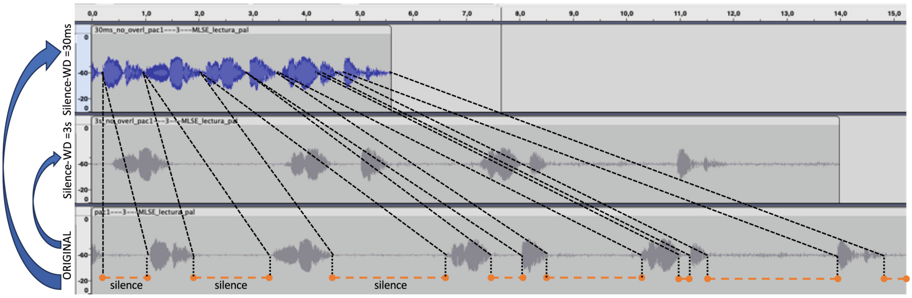

Fig. 3. Example of silence removal: Silence-WD = 30 ms (top), Silence-WD = 3 s (middle), ORIGINAL (bottom)

(ORIGINAL) is processed with the proposed algorithm with Silence-WD, 3 s and 30 ms obtaining the two subfigures Silence-WD = 3 s and Silence-WD = 30 ms. It can be observed in this case that Silence-WD = 30 ms obtains a much more clean recording than Silence-WD = 3 s.

It is important to note that PPA patients may exhibit mumbling in certain parts of the recording, which can be considered as silence. However, the applied silence-removal algorithm does not eliminate these revised mumble fragments.

Preprocessing: Framing. Window framing in voice applications has three relevant fields: Automatic Speech Recognition (ASR), Automatic Diseases Identification (ADI), and Speech Emotion Recognition (SER) [1,5,10]. ASR typically utilizes a narrow window length of approximately 25 ms, prioritizing changes over time and achieving a higher time resolution. In contrast, for SER, a wider window length of around 65 ms up to 2–3 s [2] can be used, resulting in frames with greater frequency information and higher frequency resolution. Concerning ADI applications, a crucial problem in the acoustical analysis of pathological voices is the correct Pitch Period (To) evaluation since most of the voice parameters are computed using the already determined values of To. As it's stated in [1], a window of two, even three To are suitable to segment the signal (typically 2 To correspond to 30 ms).

In this case, in order to reduce the computational cost of the training algorithms, two conservative window sizes taken from the SER field have been selected: 100 ms and 3 s, with a sliding size of 50 ms and 1 s, respectively.

Voice Features. This study focuses on using common features in Speech Emotion Recognition (SER) to characterize voice events, drawing from existing literature [3]. SER commonly utilizes prosodic and spectral features, which are combined for improved performance. Prosodic features, such as intonation and rhythm, are perceptible to humans, while spectral features capture vocal tract characteristics. Spectral features are obtained by transforming the time domain signal into the frequency domain using Fourier transform. Among the spectral features, Mel Frequency Cepstral Coefficient (MFCC) is particularly useful for SER.

Segmental transformations, rather than spectrographic ones, were considered in this research. Two groups of segmental features were chosen: ProSodic Features (PSF) to capture the rhythm and SpecTral Features (SPF) to capture frequency. The selected features include Root-mean-square (PSF), MFCC (SPF), Chroma_stft (SPF), Spectral_centroid (SPF), Spectral_bandwidth (SPF), Spectral_rolloff (SPF), and Zero_crossing_rate (SPF).

To create the dataset, each feature is computed for each frame. For MFCC, specific parameters are required:

- n_mfcc: number of MFCCs (mel coefficients) to be returned.
- n_fft: length of the FFT window in samples or milliseconds.
- hop_length: number of samples between successive frames.

Typical values for MFCC applied to Automatic Speech Recognition (ASR) are n_mfcc = 13, n_fft = 12 ms, and hop_length = 12 ms (non-overlapping frames).

Additionally, this proposal considers the mean and standard deviation of MFCC coefficients for each frame, resulting in a total of $7 + n_m fcc * 2 = 33$ features.

ML Classification Algorithms. This work only focuses on a good PPA screening test design based on an ML technique's result to rank these tasks. So, eleven representative Classification Machine Learning Techniques (CML), belonging to well-known classification typologies like linear (L), Tree-based (T), Probabilistic (P), Nearest neighBor (NB), Embeddings (EB) and Neural Networks (NN), have been deployed on the PPA/Healthy dataset.

The set of selected algorithms was: BernoulliNB (P-Ber), DecisionTree (T-DT), RandomForestClassifier (T-RF), ExtraTrees (T-XT), KNeighbors (NB-KN), RidgeClassifierCV (L-RC), SVC(L-SVC), AdaBoost (EB-AB), GradientBoosting (T-GB), Multi-Layer Perceptron (NN-MLP) and XGB.

3 Numerical Results

3.1 Material and Methods

Inclusion and Exclusion Criteria. As this is a preliminary study for a running clinical trial, just six participants have been selected for this project. Four were diagnosed with PPA in two variants, and the other two were healthy controls. The patients with PPA ranged in age from 65 to 79 years, and were diagnosed by a specialized neurologist. The two control subject were aged 67 and 73 years, with no history of neurological pathology and a normal neuropsychological profile. All participants gave informed consent before participating in the study. The ethics committee of the Principality of Asturias has approved this research.

Audio Capture Issues. A Yotto YDM-20 USB microphone connected to a MacBook Pro was used to collect the participants' recordings. All recordings were taken using a frequency 44100 Hz and mono-micro.

Validation Strategy Y Scoring Metric. All the experiments have been deployed using a repeated 5 × 2CV validation strategy, using as a scoring metric the geometric mean of the sensitivity and specificity of a two-class problem with PPA and Healthy classes (See Eq. 2).

$$GeometricMean(SE, SP) = \sqrt{Sensitivity \cdot Specificity} \tag{2}$$

3.2 Results

Tasks Splitting. The voice recording for each patient has been split manually, by an experimental psychologist, into the three tasks stated in the PPA-Tool(see Sect. 2.1). Most of the interviewer fragments were removed, but some short parts contain a mixture of both voices: interviewer and patients. It's assumed these parts will not affect the results. Finally, 2045 s (34 mins) have been obtained with a minimum of 136, 69 and 28 s for the corresponding tasks (see Table 1).

Table 1. Recorded time in secs per task (Fluency, Repetition and Naming) and patient (pac1, pac2, pac3, pac4, cont1 and con2)

Task	pac1	pac2	pac3	pac4	cont1	cont2	Subtotal
Fluency	191	182	181	**136**	181	181	1046
Repetition	238	209	98	112	79	**69**	791
Naming	176	75	217	45	31	**28**	568
Subtotal	605	466	496	293	291	278	2405

Silence Removal. After splitting the recording, the silence was removed using the method proposed in Subsect. 2.2. Four windows and sliding size configurations were deployed: i) 3 s with an overlapping size of 1 s, ii) 3 s without overlapping, iii) 30 ms with an overlapping size of 10 s, and iv) 30 ms without overlapping. For the shake of space, just the results of the best window configuration have been included in Table 2 (30 ms without overlapping). It can be stated that an average of 32% of samples were removed as silence, paying attention to the task Fluency that included a 42% of silence since it's the task where the patients have more freedom to talk, so it's more probable finding silence fragments. On the side, it can be remarked that control #1 recordings were reduced by 61%.

Table 2. Tasks time in secs per task and patient before and after silence removal deployment

Task	pac1	pac2	pac3	pac4	cont1	cont2	Subtotal
D1.Fluency before	191	182	181	136	181	181	1052
D1.Fluency after	116,4	137,1	89,9	74,3	38,3	151,4	607,5
% Removed	39	25	50	45	**79**	16	**42**
D2.Repetition before	238,0	209,0	98,0	112,0	79,0	69,0	805,0
D2.Repetition after	175,1	151,5	86,7	101,7	57,9	65,9	638,8
% Removed	26	28	12	9	27	5	21
D3.Naming before	176,0	75,0	217,0	45,0	31,0	28,0	572,0
D3.Naming after	103,8	45,9	161,9	38,4	18,7	26,3	395,0
% Removed	41	39	25	15	40	6	31
Subtotal before	605,0	466,0	496,0	293,0	291,0	278,0	2429,0
Subtotal after	395,3	334,5	338,5	214,4	114,9	243,6	1641,3
% removed	35	28	32	27	**61**	12	**32**

3.3 Features Computing, Framing and Datasets Naming

All the selected features (see Sect. 2.2) were computed for the two framing configurations proposed obtaining two datasets including the three tasks Fluency, Repetition and Naming (now on D1, D2 and D3 respectively):

- DatasetA: with a window size of 3 s with an overlapping size of 1 s.
- DatasetB: with a window size of 100 ms with an overlapping size of 50 s.

Pay attention that DatasetB deploys some oversampling, obtaining a factor by 20 of DatasetA (1 s divided by 50 ms). Synthetic new data using classical oversampling techniques like SMOTE or ADASYN has not been included in this work.

3.4 Classification Models Training, Results Discussion and Tasks Ranking

The selected CML techniques (see Sect. 2.2) have been run with the default hyperparameters provided by sklearn[1] and XGB Library[2], on each task data separately obtaining the three boxplots per CML algorithm and dataset (see Fig. 4).

On one side, the main finding is that D1.Fluency is the task that allows a good classification performance for most ML techniques in both datasets, A and B. In addition, it can be observed that the dispersion of the results of Tree-based

[1] https://scikit-learn.org/.

[2] https://xgboost.readthedocs.io.

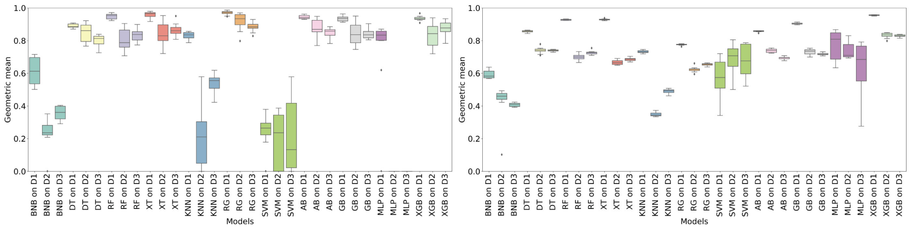

Fig. 4. left) Boxplot representing geometric mean for DatasetA, right) Boxplot representing geometric mean for DatasetB

techniques is quite reduced for DatasetB, improving the robustness of the winner models (D1-RF, D1-XT and D1-XGB).

On the other side, since DatasetA (Window/Sliding size: 3 s/1 s) is relatively smaller than DatasetB (Window/Sliding size: 100 ms/50 ms), it can be observed that Tree-Base techniques outperform the remaining ones, but the linear model ReagerC. That issue reveals the presence of multicollinearity in the data since a feature selection was not carried out. Concerning DatasetB, the Tree-Based and Embedding models outperform the rest of the models, but in this case, the RigerC obtains worse performance, but MLP performance has improved since the size of the Tasks-datasets is bigger. Thus, it can be stated that after increasing the sampling frequency (window reduced to 100 ms/50 ms), there exists more variability in each task/class dataset, so it's needed a model capable of extracting these no-linear relations (Tree-based models).

4 Conclusions and Future Work

Regarding the PPA-Tool tasks, the chosen verbal fluency task demands high cognitive effort and engages extensive brain regions [13], involving executive functions, semantic memory, and language processes. Participants rely on their own lexical-semantic system in this semi-directed task with a significant spontaneous language component. In contrast, the picture naming task limits spontaneity by providing specific pictures, while the repetition task is entirely guided, requiring patient comprehension and phonological production.

In summary, the verbal fluency task yields the best results as it assesses semi-spontaneous language, allowing participants to utilize their cognitive resources and generate fewer psycholinguistically distinct words [14]. Analyzing patients' voices offers new prospects for evaluating and diagnosing PPA, comparable to psycholinguistic analysis.

Additional language tasks from the tests used can be included to improve classification and develop more specific language assessment tests for accurate speech classification.

From a computational perspective, addressing the following is necessary: i) conducting an in-depth study on new voice features specific to PPA, ii) exploring silence removal and diarization techniques to obtain a clean signal for training, and iii) advancing CML and DL algorithms, including AutoML.

Finally, completing the clinical trial with more patients, including a balanced representation of different PPA variants, is essential to obtain a three-class dataset.

Acknowledgement. The research has been funded by the Spanish Ministry of Economics and Industry, grant PID2020-112726RB-I00, by the Spanish Research Agency (AEI, Spain) under grant agreement RED2018-102312-T (IA-Biomed), and by the Ministry of Science and Innovation under CERVERA Excellence Network project CER-20211003 (IBERUS) and Missions Science and Innovation project MIG-20211008 (INMERBOT). Also, by Principado de Asturias, grant SV-PA-21-AYUD/2021/50994. By European Union's Horizon 2020 research and innovation programme (project DIH4CPS) under Grant Agreement no 872548. And by CDTI (Centro para el Desarrollo Tecnológico Industrial) under projects CER-20211003 and CER-20211022 and by ICE (Junta de Castilla y León) under project CCTT3/20/BU/0002.

References

1. Boyanov, B., Hadjitodorov, S.: Acoustic analysis of pathological voices: a voice analysis system for the screening and laryngeal diseases. IEEE Eng. Med. Biol. Maga. **16**(4), 74–82 (1997)
2. de la Cal, E., Gallucci, A., Villar, J.R., Yoshida, K., Koeppen, M.: Simple meta-optimization of the feature MFCC for public emotional datasets classification. In: Sanjurjo González, H., Pastor López, I., García Bringas, P., Quintián, H., Corchado, E. (eds.) HAIS 2021. LNCS (LNAI), vol. 12886, pp. 659–670. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86271-8_55
3. de la Cal, E., Gallucci, A., Villar, J.R., Yoshida, K., Koeppen, M.: A first prototype of an emotional smart speaker, pp. 304–313 (2022)
4. Cho, S., Nevler, N., Shellikeri, S., Ash, S., Liberman, M.: Automatic classification of primary progressive aphasia patients using lexical and acoustic features. In: Proceedings of Language Resources and Evaluation Conference 2020 workshop on Resources and Processing of Linguistic, Para-linguistic and Extra-linguistic Data from People with Various Forms of Cognitive/Psychiatric/Developmental Impairments, June, pp. 60–65 (2020)
5. Fan, W., Xu, X., Xing, X., Chen, W., Huang, D.: LSSED: a large-scale dataset and benchmark for speech emotion recognition. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, June 2021, pp. 641–645 (2021)
6. Fraser, K.C., et al.: Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex **55**(1), 43–60 (2014)
7. Gorno-Tempini, M.L., et al.: Classification of primary progressive aphasia and its variants. Neurology **76**(11), 1006–1014 (2011)
8. Hoffman, P., Sajjadi, S.A., Patterson, K., Nestor, P.J.: Data-driven classification of patients with primary progressive aphasia. Brain Lang. **174**(July), 86–93 (2017)
9. Matias-Guiu, J.A., et al.: Spanish version of the mini-linguistic state examination for the diagnosis of primary progressive aphasia. J. Alzheimer's Dis. **83**(2), 771–778 (2021)

10. Orozco-Arroyave, J.R., Arias-Londoño, J.D., Vargas-Bonilla, J.F., González-Rátiva, M.C., Nöth, E.: New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In: Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, December 2014, pp. 342–347 (2014)
11. Patel, N., et al.: A 'Mini Linguistic State Examination' to classify primary progressive aphasia (2022)
12. Ranjan Sahoo, T., Patra, S.: Silence removal and endpoint detection of speech signal for text independent speaker identification. Image Graph. Signal Process. **6**, 27–35 (2014)
13. Riello, M., et al.: Neural correlates of letter and semantic fluency in primary progressive aphasia. Brain Sci. **12**(1), 1 (2021)
14. Rofes, A., De Aguiar, V., Ficek, B., Wendt, H., Webster, K., Tsapkini, K.: The role of word properties in performance on fluency tasks in people with primary progressive aphasia. J. Alzheimer's Dis. **68**(4), 1521–1534 (2019)
15. Themistocleous, C., Webster, K., Afthinos, A., Tsapkini, K.: Part of speech production in patients with primary progressive aphasia: an analysis based on natural language processing. Am. J. Speech-Lang. Pathol. **30**(1s), 466–480 (2021)