HAL open science

# Validation of the Neuro-Concept Detector framework for the characterization of speech disorders: A comparative study including Dysarthria and Dysphonia

Sondes Abderrazek, Corinne Fredouille, Alain Ghio, Muriel Lalain, Christine Meunier, Virginie Woisard

## To cite this version:

Sondes Abderrazek, Corinne Fredouille, Alain Ghio, Muriel Lalain, Christine Meunier, et al.. Validation of the Neuro-Concept Detector framework for the characterization of speech disorders: A comparative study including Dysarthria and Dysphonia. Interspeech 2022, Sep 2022, Incheon, South Korea. 10.21437/interspeech.2022-10631 . hal-03837510

**HAL Id: hal-03837510**

**https://hal.science/hal-03837510v1**

Submitted on 8 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interspeech 2022
18-22 September 2022, Incheon, Korea

# Validation of the Neuro-Concept Detector framework for the characterization of speech disorders: A comparative study including Dysarthria and Dysphonia

*Sondes Abderrazek*[1], *Corinne Fredouille*[1], *Alain Ghio*[2], *Muriel Lalain*[2], *Christine Meunier*[2], *Virginie Woisard*[3]

[1]LIA, Avignon University, France
[2]Aix-Marseille Univ, LPL, CNRS, Aix-en-Provence, France
[3]UT2J, LNPL, Toulouse University & Toulouse Hospital, France
(sondes.abderrazek,corinne.fredouille)@univ-avignon.fr

## Abstract

Recently, we have proposed a general analytical framework, called Neuro-based Concept Detector (NCD), to interpret the deep representations of a DNN. Based on the activation patterns of hidden neurons, this framework highlights the ability of neurons to detect a specific concept related to the final task. Its main strength is to provide an interpretability tool for any type of DNN performing a classification task, whatever the application domain. Thanks to *NCD*, we have demonstrated the emergence of phonetic features in the classification layers of a CNN-based model for French phone classification. The emergence of this concept, of great interest in the field of clinical phonetics, has been studied considering healthy speech. Applied to Head and Neck Cancers, we have shown that this framework automatically reflects the level of impairment of the phonetic features produced by a patient, which is supported by the strong correlations with perceptual assessments performed by clinical experts. The objective of the work presented here is to validate the proposed framework by confronting it to new populations of patients, but with very different pathologies (neurodegenerative diseases/ Dysarthria and vocal dysfunction/ Dysphonia). The robustness of the approach to the phonetic content variability of read text is also studied.

**Index Terms**: Deep Learning, Interpretability, Speech disorders, Phonetic features, Intelligibility, Head and Neck Cancers, Dysarthria, Dysphonia.

## 1. Introduction

Speech and voice disorders impact millions of people and their ability to communicate. They can be caused by neurodegenerative diseases like Parkinson's disease, strokes, laryngeal or oropharyngeal cancers (also referring to Head and Neck Cancers - HNC), stuttering, or even by a dysfunction of vocal folds. Perceptual evaluation remains the most widely used method in clinical practice to assess speech and/or voice disorders [1, 2, 3]. However, this assessment remains a non-trivial task, time-consuming depending on the degree of precision sought and very subjective (important intra and inter-expert variability). Clinicians are therefore still waiting for objective evaluation tools that will allow them to better understand the evolution of these disorders in the case of a longitudinal follow-up of patients or to evaluate the benefits of a therapeutic treatment or a rehabilitation. Automatic speech processing approaches have been seen, very early, as potential solutions to provide these objective tools [4]. Studies on the application of automatic approaches are numerous in the literature for the evaluation of voice or speech disorders. The explosion of deep learning-based approaches in automatic speech processing has not escaped this field of application [5]. Related to voice disorders, we can cite [6, 7, 8, 9] dedicated to the prediction or classification tasks of diseases, for instance. Regarding speech disorders, the amount of related work is even larger, the Parkinson's disease being one of the main focuses [10, 11]. As recent studies, we can also cite [12, 13] for the assessment of speech intelligibility for patients suffering from HNC. Considering populations of dysarthric patients, [14, 15] propose deep learning-based approaches for the enhancement of speech intelligibility while [16, 17] for the assessment of speech disorder severity. Despite the wide range of work, it is very rare to see studies involving patient populations with various causes of speech disorders. From our knowledge, we can mention [18], involving adults suffering from laryngeal cancers and children with cleft lip and palate, or [19] involving dysarthric and dysphonic speakers, children with cleft lip or palate, speakers with pathological speech secondary to hearing impairment, laryngectomized and glossectomized speakers.

Recently, we have proposed an original framework, *Neuro-Concept Detector* (NCD), for the interpretability of Deep Neural Networks (DNNs) [20]. This framework highlights the ability of hidden neurons to detect a specific concept related to the final task. In our specific context, the focus has been on the phonetic features related to French phones as target concept. Indeed, the main objective of that work is to model healthy speech and to characterize potential deviation within impaired speech in order to provide relevant information in terms of phonetic feature alteration. Experiments conducted within this overall framework have shown that: (1) we are able to define a set of phonetic feature detectors on healthy speech; (2) based on this set of phonetic feature detectors, we designed a special similarity score between the reference (i.e. healthy speech) and new unseen data (i.e. impaired speech). The relevance of this score has been confirmed on HNC patients [21], since it strongly correlates with perceptual measures given by clinical experts when considered globally per speaker all features included; (3) the analysis of local similarity scores - one score per speaker and per phonetic feature - has led to very interesting observations, notably the role of the tongue mobility, pointed out as predominant on observed HNC patients. The objective of this paper is to validate these observations and the interest of the proposed methodology on different and varied populations of patients including dysarthria and dysphonia.

## 2. The NCD framework

In this section, we briefly describe the general analytic framework, *Neuro-Concept Detector* (NCD), proposed in [20]. This framework was designed for the interpretability of the deep rep-

Copyright © 2022 ISCA

10.21437/Interspeech.2022-10631

resentations of a DNN performing a classification task. Based on the activation patterns of hidden neurons, it highlights the ability of neurons to detect a specific *concept* related to the target task. The focus here is to show the emergence of phonetic feature detectors. Such a concept is of a great interest in the field of clinical phonetics since it can be directly associated with the characteristics of speech disorders.

Consequently, we firstly applied NCD to the fully connected layers of the CNN-based model proposed in [22], performing French phone classification. The model was trained on a first subset of the BREF corpus [23], composed of French read-speech records. We adopted the definition of the set of phonetic features, composing and distinguishing French phones, given in [24]. Relying on a separation between vowels and consonants, this definition as proposed by [25] is more phonetically and acoustically pertinent, especially regarding speech disorders. Here, the notion of phonetic features imposes a binary status (i.e. equals 1 if the phonetic feature is present in the phoneme, or 0 if absent). In the following, *[+phonetic_feature]* refers to the class of phonemes presenting the phonetic feature while *[-phonetic_feature]* refers to the class of phonemes presenting the opposite phonetic feature. Lastly, in order to explore whether interpretable neurons in terms of phonetic feature detection exist in the classification layers, a second subset of the corpus BREF was still implied, that we consider as our reference for healthy speech. Thus, a representative subset dedicated to test and referred to *BREF-Int*, was selected including almost 82K of frames balanced across phones.

Some basic notations are defined in the following to formulate the score used for this exploration. Let $h_{n,i}$ be the activation value of the neuron *n* given the $i^{th}$ input frame of *BREF-Int* dataset. A normalized activation $a_{n,i}$ is calculated for each neuron by dividing the initial activation values of the neuron for different input frames of the dataset by the maximum value reached over all these values; $a_{n,i} = \frac{h_{n,i}}{hmax_n}$ where $hmax_n = \max h_{n,j} \forall j$.
Let $A^{BREF}_{n,k}$ be the set of normalized activations of the neuron *n* for all the frames having the phone *k* as a true label and belonging to *BREF-Int*. We note the median activation value of the neuron $n$ for the phone *k* as $m_{A^{BREF}_{n,k}}$. The score $S_{n,T_x}$, quantifying the degree to which a unit detects the presence of a phonetic feature/the opposite phonetic feature, is therefore calculated for each neuron $n$ and phonetic feature $T_x$ as follows:

$$S_{n,T_x} = \frac{1}{|[+T_x]|} \sum_{k \in [+T_x]} m_{A^{BREF}_{n,k}} - \frac{1}{|[-T_x]|} \sum_{k \in [-T_x]} m_{A^{BREF}_{n,k}} \tag{1}$$

Since phonetic features are binary concepts characterizing vowels and consonants separately, the $x$ in the score denotes the macro-class of either vowels or consonants, $v$ and $c$ respectively ($x \in [v, c]$). Consequently, $T_v$ denotes a vowel phonetic feature (i.e. $T_v \in \{$*nasal, back, round, high, low*$\}$) and $T_c$ denotes a consonant phonetic feature (i.e. $T_c \in \{$*sonorant, continuant, nasal, voiced, compact, acute*$\}$).

$S_{n,T_x} \in [-1; 1]$, where a strong value close to 1 reflects that the neuron is a strong detector for the presence of the phonetic feature in question while a very low score close to -1 means that the neuron is a strong detector for the opposite phonetic feature. A threshold value, indicating whether or not a neuron is selected as detector, is empirically fixed to $\pm 0.25$. Indeed, we consider that the neuron $n$ is a detector of the presence of phonetic feature $T_x$, noted [+$T_x$], if $S_{n,T_x} > 0.25$. Conversely, if $S_{n,T_x} < -0.25$, then the neuron $n$ is considered as a detector of the opposite phonetic feature $T_x$, noted [-$T_x$]. Experiments conducted in [20] revealed interesting results. Indeed, it showed that interpretable neurons with phonetic feature encoding properties emerge in the fully connected layers of the CNN. These detectors cover all the phonetic features, and additionally increase numerically when going deeper in layers towards the final classification layer, ensuring potential discrimination among phone classes.
At this stage, each selected interpretable neuron, *n*, is being labelled with the specific phonetic feature *t* it detects. Let $N_t$ denotes the set of interpretable neurons selected as detectors for the phonetic feature *t* across all the examined layers (*t* is equal to either [+$T_x$] or [-$T_x$]). Based on these findings, it can be examined to which extent this set of neurons can be used to extract relevant interpretations on speech alterations in case of disorders. To this end, a similarity score, named *Artificial Neuron-based Phonological Similarity (ANPS)* is specifically defined. Its goal is to assess how well the acoustic/articulatory characteristics related to phonetic feature *t* are produced by speaker *s*, based on the corresponding set of detectors. Let $ANPS_{s,t}$ be defined as the following ratio:

$$ANPS_{s,t} = \frac{\sum_{n \in N_t} \sum_{k \in t} m_{A^s_{n,k}}}{\sum_{n \in N_t} \sum_{k \in t} m_{A^{BREF}_{n,k}}} \tag{2}$$

where, similarly to the previously $A^{BREF}_{n,k}$, we note $A^s_{n,k}$ the set of normalized activations of the neuron *n* for all the frames belonging to the phone *k* and produced by the speaker *s*. In the same way, we note $m_{A^s_{n,k}}$ the median value of this set of normalized activations. It should be noted that *ANPS* score can range from zero to an unbounded maximum value that we constrained to 1 considering that a greater value brings no more information than a perfect production of the phonetic feature by the speaker in question. Accordingly, a low score close to 0 implies that the speech production of the speaker does not exhibit typical acoustic characteristics related to the phonetic feature in question.
This scoring approach revealed interesting results on the French corpus C2SI-LEC as reported in [21]. Recorded within the C2SI project [26], this corpus includes read speech produced by patients with HNC and control speakers, coupled with their perceptual evaluations. By generating *ANPS* scores for these speakers for the different phonetic features, we were able to link the most deteriorated features, highly correlated to the perceptual measures, with the role of the tongue, which can be strongly impacted in cancers of oral cavity and oropharynx (i.e. the case of C2SI patients). On the other side, no perturbation in the vocal cord vibration was reported within the *ANPS* scores for the corresponding voiceness feature even for patients with very severe speech degradation. This is consistent with our HNC patient population and would very probably not be the case if the patients in question had laryngeal cancer.

# 3. Validation corpora

Various pathologies related to dysarthria or dysphonia are involved in this comparative study. Each pathology is associated with different populations of patients, described just below.

## 3.1. Dysarthria

Three types of dysarthrias are included in this study. They are all associated with neurodegenerative diseases implying three major neurological systems: the extrapyramidal system with Parkinson's disease, the pyramidal system with Amyotrophic Lateral Sclerosis, and the cerebellar system with Cerebellar ataxia. Patients involved were recruited in different hospitals at different

periods as described below (all patients signed a consent form when required). All of them were recorded on different speech production tasks, among which a reading task on the French text of "Le cordonnier", implied in this study. A group of 5 male and 5 female control subjects is also considered, performing the same reading task. They were all included in a past research project [27].

- Group of Parkinson's disease (PD): composed of two sub-groups of patients (min./max. age: 48/85 years; mean/standard deviation: $\mu$=66.7 years/$\sigma$=8.6). The first one, composed of 13 male and 3 female patients issued from the CCM corpus described in [27], was recorded by Dr. Claude Chevry-Muller over 30 years (between 1967 to 1997) in Paris. The second sub-group, composed of 10 male and 5 female patients, referring to the AHN corpus in [27], was recorded at the department of Neurology of Aix-en-Provence Hospital (impulsed by Prof. François Viallet). It is worth noting that this second sub-group performed a double task of reading, comprising the same text as the other groups of patients as well as the reading of the French text "La chèvre de M. Seguin".

- Group of Cerebellar Ataxia (CA): composed of two sub-groups of patients as well. The first one, still referring to the CCM corpus, includes 6 male and 5 female patients. The second sub-group was recorded at the department of Ear, Nose & Throat (ENT) of the Timone Hospital at Marseille (impulsed by Dr. Danielle Robert) and includes 7 male and 4 female patients (min./max.: 32/86 yrs; $\mu$=55.7 yrs/$\sigma$=16.3).

- Group of Amyotrophic Lateral Sclerosis (ALS): it was recorded in the Voice and Speech lab. of the European Hospital Georges Pompidou in Paris by Dr. Lise Crevier Buchman and her colleagues. It includes 14 male and 24 female patients (min./max.: 44/89 yrs; $\mu$=65.4 yrs/$\sigma$=9.4).

A perceptual evaluation of speech productions of all patients and controls was performed at the same time by 11 expert judges (10 speech pathologists and 1 neurologist). This evaluation was done according to the 9 items of a French perceptual evaluation scale of dysarthria (GEPD) [28]. Seven speech dimensions - global dysarthria severity, global speech/voice regularity, speech intelligibility, presence of nasal resonance, palilalia, articulatory accuracy and regularity of the speech rate - were rated on a 4 degree scale (0=normal to 3=severely impaired). Two last dimensions - melodic fluctuation and speech rate - were rated on a -3 to 3 scale (0 being the normal degree, negative values referring to hypo-modulation or slow rate of speech respectively). A final item regarding the presence of a regional accent was evaluated by experts as this may be very significant for some patients.

Figure 1 provides the mean scores of the different perceptual items (including the presence of regional accent) on the different dysarthric groups and control population. As reported in [29] on similar patient sub-populations, the ALS group has the most severely rated speech, with the highest scores for global dysarthria, speech intelligibility and articulation accuracy. It is followed by the CA group, and afterwards the PD group on the same scores. The ALS group also presents the largest presence of nasal resonance, which is typically due to the weak or absent closure of velo-pharynx. Although speech rate and speech melody will not be addressed later, abnormal slow speech rate is observed on CA and ALS groups, typical of ataxic or flaccid-spastic dysarthria as opposed to normal or fast observations for the PD group.

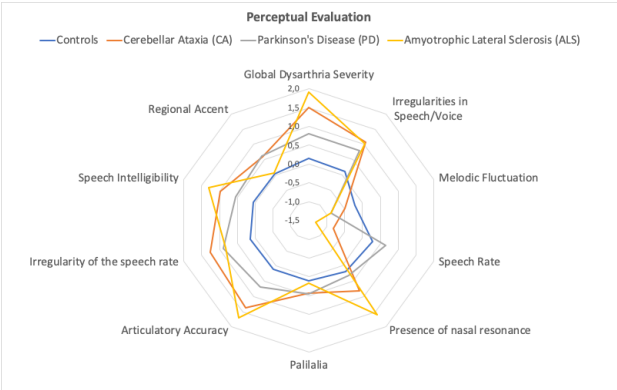


Figure 1: *Mean perceptual scores according to 9 GEPD items & the regional accent per dysarthria group and control speakers.*

### 3.2. Dysphonia

The corpus of dysphonic voices was recorded in the 2000s at the department of ENT of the Timone Hospital at Marseille. It is composed of 80 records of female voices, including 20 control subjects and 60 dysphonic patients, aged from 17 to 50 years (average 32.2 years) [30]. The set of dysphonic patients underwent a laryngoscopic examination showing dysphonia essentially of functional origin mainly due to nodules, oedemas, polyps, and cysts (gathering 53 patients among 60). All patients were recorded on a reading task of the French text "La chèvre de M. Seguin". The 80 female speakers were selected among a larger corpus in order to be equally distributed into the 4 levels of the Global item of the GRBAS scale [2]: 20 normal/control voices (i.e. grade G0), 20 voices with mild dysphonia (i.e. grade G1), 20 voices with moderate dysphonia (i.e. grade G2), and 20 voices with severe dysphonia, but still intelligible (i.e. grade G3). The GRBAS-based assessment of the larger corpus was performed by a panel of clinical experts following a consensus decision.

## 4. Experiments

This section provides a comparative analysis examining to which extent we can rely on NCD approach and the *ANPS* score to extract relevant knowledge related to the specific characteristics of each pathology: dysarthria (ALS, CA and PD) and dysphonia.

### 4.1. Comparative analysis based on ANPS scores

At this stage, the analysis is carried out for each pathology based on *ANPS* scores computed for each individual phonetic feature and the concerned patients and control speakers.
For visualization, heatmaps are used to plot *ANPS* scores, where the x-axis represents the speakers sorted, by pathology, from the least severely affected (on the right) to the most severely affected (on the left) within the same group (according to the global perceptual severity measure), while the y-axis represents the phonetic characteristics of the macro-class in question. Due to limited space, only the heatmap related to the consonant phonetic features is presented in this paper in figure 2. A sequential scale shows the progression from the most to the least opaque shades of red color, representing low to high score values. As a first global observation, we can clearly mention that cells with high opacity are concentrated in the left side of each pathology, which is consistent with the high global severity level of the corresponding patients. The heatmap visualization is also supported by a Pearson correlation analysis summarized in table 1 between the different *ANPS* scores associated with the subset of phonetic features as well as the subset of perceptual measures that we con-

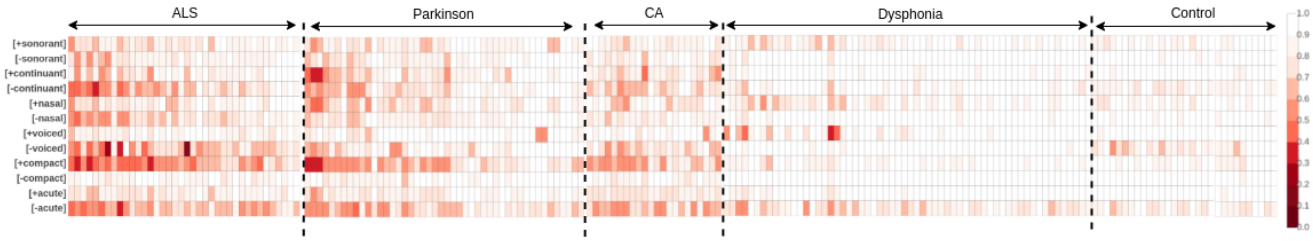


Figure 2: *Heatmap showing ANPS scores per consonant phonetic feature (y-axis) and patients grouped by pathology and sorted by Global Severity within each group (x-axis), in addition to control speakers*

Table 1: *Most relevant phonetic features regarding correlation rates between their ANPS scores and perceptual measures*

| | | Vowels | | | | Consonants | | |
|---|---|---|---|---|---|---|---|---|
| | | [-nas.] | [+bac.] | [+rou.] | [+hig.] | [-con.] | [-nas.] | [+com.] |
| **G Sev.** | ALS | 0.50 | 0.74 | 0.59 | **0.78** | 0.76 | 0.74 | **0.79** |
| | CA | 0.47 | 0.60 | 0.39 | 0.27 | 0.37 | 0.52 | 0.41 |
| | PD | 0.71 | 0.67 | **0.85** | 0.73 | **0.82** | 0.72 | 0.74 |
| **Intell.** | ALS | 0.60 | 0.74 | 0.67 | 0.67 | **0.78** | **0.77** | **0.84** |
| | CA | 0.67 | 0.65 | 0.47 | 0.43 | 0.58 | 0.61 | 0.57 |
| | PD | 0.66 | 0.64 | **0.84** | 0.73 | **0.81** | 0.70 | 0.74 |
| **Nas. R.** | ALS | **0.76** | 0.64 | 0.64 | **0.75** | **0.75** | **0.81** | 0.55 |
| | CA | 0.41 | 0.72 | 0.40 | 0.00 | 0.05 | 0.33 | 0.25 |
| | PD | 0.57 | 0.56 | 0.65 | 0.53 | 0.63 | 0.63 | 0.56 |
| **Artic.** | ALS | 0.51 | 0.73 | 0.60 | 0.68 | **0.75** | 0.70 | **0.85** |
| | CA | 0.49 | 0.63 | 0.29 | 0.20 | 0.29 | 0.44 | 0.36 |
| | PD | 0.67 | 0.69 | **0.83** | **0.76** | **0.76** | 0.73 | 0.71 |

All correlation values are in absolute value ( those $\geq 0.75$ are in bold)

sider as the most relevant according to pathologies.

Although ALS, PD and CA are associated with different types of dysarthria, it is worth mentioning that they all show difficulties in the consonant production in terms of articulatory alteration or consonant imprecision [1]. This reason is behind our choice of including the consonant feature heatmap.

**ALS:** It is worth noting that the perceptual measure of the nasal resonance is mainly correlated with the *ANPS* scores of patients suffering from ALS when compared to the rest of pathologies. More specifically, the phonetic features *[-nasal]* for the vowel and consonant macro-classes are among the top two phonetic features with which this measure correlates most strongly, with 0.76 and 0.81 respectively. These correlations, visually clear for the consonant phonetic feature *[-nasal]* in figure 2 for ALS, suggest that the ALS patients have a nasal quality voice (i.e. oral phones are badly nasalized). This finding is altogether consistent with the high nasal resonance of the patients, perceived by the expert jury as reported in section 3.1, and with the well-known hypernasality of mixed dysarthria characterizing ALS patients [1]. Furthermore, regarding the imprecision of consonants, which is one of the characteristics of ALS, it is observed in table 1 that *[+compact]* and *[-continuous]* are strongly correlated with the articulatory measure of ALS, with values equal to 0.85 and 0.75 respectively.

**PD:** the imprecision of consonants is also an important feature in Parkinson's disease and usually includes distortions due to the reduction of articulatory movements notably. In particular, this imprecision is observable in the table through the strong correlation of the feature *[-continuant]* with related perceptual measures, such as global severity, intelligibility and articulation accuracy. This can be notably explained by the parkinsonian reduced capacity of completing articulatory occlusion [31].

**CA:** Surprisingly, while the cerebellar patients show rather similar patterns to the other dysarthric groups on the heatmap, no correlation score with the perceptual assessments exceeds 0.7 value. Further patient-by-patient analysis is required here to better understand *ANPS* scores obtained and their consistency with perceptual measures and related dysarthria characteristics.

**Dysphonia:** Since dysphonia is a voice disorder, we can clearly notice in figure 2 that almost none of the phonetic features related to the place or manner of articulation are significantly impaired, as this would be the case with speech disorders, considering the most affected patients. We would have expected the phonetic feature *[+voiced]* to be affected, almost within G2 and G3 patients. However, three patients only exhibit very low scores (less than 0.5) for that feature. When checking the corresponding recordings, it turns out that two of these patients have the most severe voice disorders (compared with other patients rated G3). Indeed, they are characterized by a weak and whispered speech as well as some large difficulties to produce speech (vocal fatigue).

### 4.2. ANPS and phonetic content variability

So far, we have demonstrated the capacity of our approach to reflect some speech impairments, depending on the speech pathology observed. At this stage, our aim is to study the robustness of the approach to the variability of phonetic content. To this end, we observe the *ANPS* scores of the (AHN) PD patients described in section 3.1, who were involved in a double reading task on two different texts. Given the set of independent paired *ANPS* scores $(x_{i,j}, y_{i,j}), \forall j, i = 1, .., 15$ where $j$ refers to the phonetic feature produced by the $i^{th}$ patient on "La chèvre de M. Seg." and "Le cordonnier" texts respectively, considering vowels and consonants macro-classes independently . The Wilcoxon signed rank test is applied to identify whether a significant difference exists between the matched pairs. Results with a *p-value*=0.015 and *p-value*=4x$10^{-4}$ for vowel and consonant macro-classes resp. show a significant difference between scores, which means that the proposed scores are not independent from the text and related phonetic content (which is not so surprising). A preliminary analysis tends to show that the phonetic features are not equally affected by the text used. Further analysis is therefore required.

## 5. Conclusion

Although our model was trained on healthy speech for the task of phone classification and had never received any prior information about pathological speech, we have shown that we are able to extract relevant knowledge related to the specific characteristics of some pathologies, like ALS and the Parkinson's disease considered in this paper. Thus, despite the variability in speech production we can observe between patients, typical observations can be made in a similar way to the ones reported for the HNC cancers in our previous work. To the best of our knowledge, there is no study in the scientific literature that highlighted the capacity of an automatic tool to generate interpretations not only on different types of dysarthria, but also on cancers and dysphonia. Nonetheless, we are aware that further studies, focusing more on the interpretation of these results from a clinical point of view, must be conducted to complete this validation study.

## 6. Acknowledgements

This work was supported by the French National Research Agency under project RUGBI (ANR-18-CE45-0008-04).

# 7. References


[1] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," *Journal of Speech and Hearing Research*, vol. 12, pp. 246–269, 1969.

[2] M. Hirano, "Psycho-acoustic evaluation of voice : GRBAS scale for evaluating the hoarse voice," *Clinical Examination of voice, Springer Verlag*, 1981.

[3] P. Enderby and R. Palmer, "Fda-2: Frenchay dysarthria assessment," *(2nd ed), Pro-Ed, Tex*, 2008.

[4] L. J. Ferrier, N. Jarrell, T. Carpenter, and H. C. Shane, "A case study of a dysarthric speaker using the dragondictate voice recognition system," *Journal for Computer Users in Speech and Hearing*, vol. 8(1), pp. 33–52, 1992.

[5] N. Cummins, A. Baird, and B. W. Schuller, "Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning," *Methods*, vol. 151, pp. 41–54, 2018, health Informatics and Translational Data Analytics.

[6] M. Alhussein and G. Muhammad, "Voice pathology detection using deep learning on mobile healthcare framework," *IEEE Access*, vol. 6, pp. 41 034–41 041, 2018.

[7] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. Khanapi Abd Ghani, M. S. Maashi, B. Garcia-Zapirain, I. Oleagordia, H. Alhakami, and F. T. AL-Dhief, "Voice pathology detection and classification using convolutional neural network model," *Applied Sciences*, vol. 10, no. 11, 2020.

[8] H.-C. Hu, S.-Y. Chang, C.-H. Wang, K.-J. Li, H.-Y. Cho, Y.-T. Chen, C.-J. Lu, T.-P. Tsai, O. K.-S. Lee *et al.*, "Deep learning application for vocal fold disease prediction through voice recognition: Preliminary development study," *Journal of medical Internet research*, vol. 23, no. 6, p. e25247, 2021.

[9] I. Miliaresi, K. Poutos, and A. Pikrakis, "Combining acoustic features and medical data in deep learning networks for voice pathology classification," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1190–1194.

[10] I. Nissar, W. A. Mir, Izharuddin, and T. A. Shaikh, "Machine learning approaches for detection and diagnosis of parkinson's disease - a review," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2021.

[11] P. Klumpp, T. Arias-Vergara, J. C. Vásquez-Correa, P. A. Pérez-Toro, J. R. Orozco-Arroyave, A. Batliner, and E. Nöth, "The phonetic footprint of parkinson's disease," *Computer Speech Language*, vol. 72, p. 101321, 2022.

[12] L. Bin, M. C. Kelley, D. Aalto, and B. V. Tucker, "Automatic speech intelligibility scoring of head and neck cancer patients with deep neural networks," in *International Congress of Phonetic Sciences (ICPhS'19), Melbourne, Australia*, 2019.

[13] S. Quintas, J. Mauclair, V. Woisard, and J. Pinquier, "Automatic Prediction of Speech Intelligibility Based on X-Vectors in the Context of Head and Neck Cancer," in *Interspeech*, China, 2020.

[14] C.-Y. Chen, W.-Z. Zheng, S.-S. Wang, Y. Tsao, P.-C. Li, and Y.-H. Lai, "Enhancing intelligibility of dysarthric speech using gated convolutional-based voice conversion system." in *INTERSPEECH*, 2020, pp. 4686–4690.

[15] W.-Z. Zheng, J.-Y. Han, C.-K. Lee, Y.-Y. Lin, S.-H. Chang, and Y.-H. Lai, "Phonetic posteriorgram-based voice conversion system to improve speech intelligibility of dysarthric patients," *Computer Methods and Programs in Biomedicine*, vol. 215, p. 106602, 2022.

[16] C. Bhat and H. Strik, "Automatic assessment of sentence-level dysarthria intelligibility using blstm," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 322–330, 2020.

[17] S. Gupta, A. T. Patil, M. Purohit, M. Parmar, M. Patel, H. A. Patil, and R. C. Guido, "Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments," *Neural Networks*, vol. 139, pp. 105–117, 2021.

[18] A. Maier, T. Haderlein, F. Stella, E. North, E. Nkenke, F. Rosanowski, A. Schutzenberger, and M. Schuster, "Automatic speech recognition system for the evaluation of voice and speech disorders in head and neck cancer," *EURASIP Journal on Audio, Speech and Music Processing*, 2010.

[19] C. Middag, J.-P. Martens, G. V. Nuffelen, and M. D. Bodt, "Automated intelligibility assessment of pathological speech using phonological features," *EURASIP Journal on Applied Signal Processing*, 2009.

[20] S. Abderrazek, C. Fredouille, A. Ghio, M. Lalain, C. Meunier, and V. Woisard, "Towards interpreting deep learning models to understand loss of speech intelligibility in speech disorders step 2 : contribution of the emergence of phonetic traits," in *ICASSP, Singapore*, 2022.

[21] S. Abderrazek, C. Fredouille, A. Ghio, M. Lalain, C. Meunier, and et al., "Interpreting Deep Representations of Phonetic Features via Neuro-based Concept Detector: Application to speech disorders due to Head and Neck cancer," *Submitted in IEEE/ACM Transactions on Audio Speech and Language Processing*, February 2022.

[22] S. Abderrazek, C. Fredouille, A. Ghio, M. Lalain, V. Woisard, and C. Meunier, "Towards interpreting deep learning models to understand loss of speech intelligibility in speech disorders, step 1 : CNN model-based phone classification," in *Interspeech'20*, Shanghai, China, 2020.

[23] L. F. Lamel, J. L. Gauvain, and M. Eskénazi, "BREF, a large vocabulary spoken corpus for french," in *Eurospeech'91*, Italy, 1991.

[24] A. Ghio, M. Lalain, L. Giusti, C. Fredouille, and V. Woisard, "How to compare automatically two phonological strings: Application to intelligibility measurement in the case of atypical speech," in *12th Conference on Language Resources and Evaluation (LREC)*, France, 2020.

[25] A. Ghio, "Achile : un dispositif de décodage acoustico-phonétique et d'identification lexicale indépendant du locuteur à partir de modules mixtes," Theses, Université d'Aix Marseille, Nov. 1997.

[26] V. Woisard and et al., "C2SI corpus: a database of speech disorder productions to assess intelligibility and quality of life in head and neck cancers," *Language Resources and Evaluation*, 2021.

[27] C. Fougeron, L. Crevier-Buchman, C. Fredouille, A. Ghio, C. Meunier, C. Chevrie-Muller, J.-F. Bonastre, A. Colazo-Simon, C. Delooze, D. Duez, C. Gendrot, T. Legou, N. Lévêque, C. Pillot-Loiseau, S. Pinto, G. Pouchoulin, D. Robert, J. Vaissière, F. Viallet, and C. Vincent, "The DesPho-APaDy project: Developing an acoustic-phonetic characterization of dysarthric speech in french," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC'10*, Valletta, Malta, may 2010, pp. 2831–2838.

[28] L. Lhoussaine, *Première validation de la Grille d'Évaluation Perceptive de la Dysarthrie (G.E.P.D.) : effet du niveau d'expertise du jury et différenciation entre types de dysarthrie*. Speech therapist thesis, University of Paris VI, Pierre et Marie Curie (in French), 2012.

[29] C. Meunier, C. Fougeron, C. Fredouille, B. Biggi, L. Crevier-Buchman, E. Delais-Roussarie, L. Georgeton, A. Ghio, I. Laaridh, T. Legou, C. Pillot-Loiseau, and G. Pouchoulin, "The typaloc corpus: A collection of various dysarthric speech recordings in read and spontaneous styles," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC'16*, Portoroz, Slovenia, May 2016.

[30] G. Pouchoulin, C. Fredouille, J.-F. Bonastre, A. Ghio, and A. Giovanni, "Frequency study for the characterization of the dysphonic voices," in *Proceedings of Interspeech'07*, Antwerp, Belgium, September 2007.

[31] H. Ackermann and W. Ziegler, "Articulatory deficits in parkinsonian dysarthria: An acoustic analysis," *Journal of neurology, neurosurgery, and psychiatry*, vol. 54, pp. 1093–8, 01 1992.