Interspeech 2024
1-5 September 2024, Kos, Greece

Neurocomputational model of speech recognition for pathological speech detection: a case study on Parkinson's disease speech detection

Sevada Hovsepyan, Mathew Magimai.-Doss

Idiap Research Institute, Centre du Parc, Rue Marconi 19, CH-1920 Martigny, Switzerland
{sevada.hovsepyan, mathew}@idiap.ch

Abstract

This paper presents a computational model for distinguishing between healthy speech and pathological speech, specifically speech from patients with Parkinson's disease. The model is based on neurophysiologically plausible computational models of speech and syllable recognition. These models were designed to uncover the functional roles of brain activity during speech perception. The proposed model is a two-level generative model that uses predictive coding to identify whether the input syllable corresponds to the healthy or Parkinson's disease condition. During inference, the model accumulates the evidence associated with each condition. Although early results are modest (around 60% AUC), they suggest that this approach has merit and should be further investigated.

Index Terms: neurocomputational models, speech recognition, Parkinson's disease detection, predictive coding

1. Introduction

Although both automatic speech recognition (ASR) models and neurocomputational models of speech perception share a similar task of recognising words and syllables from speech waveform, their goals are largely different. While the former is more focused on achieving high accuracy scores, the latter is focused on neurophysiological plausibility, aiming to replicate and explain brain functioning during speech perception. Throughout history, there has been mutual benefit between the two fields of research. For example, artificial neural networks have been inspired by cortical networks [1], transformers have been inspired by selective attention in the brain [2], and Hopfield networks have been used as a model of associative memory [3]. Additionally, online automatic syllabification has been achieved using cortical oscillations [4, 5], and the GPT architecture has been used to explain brain activity during speech perception [6, 7]. However, while linguistic (ASR) and paralinguistic speech analysis often share similar methodologies (e.g. feature extraction), the connection between the latter and neurocomputational research has not yet been fully explored. The present paper focuses on that aspect to demonstrate the use of neurocomputational models for paralinguistic speech analysis, including the detection of pathological speech such as Parkinson's disease.

One prominent characteristic of speech perception in the brain is its hierarchical organization, spanning multiple cortical areas that communicate with each other through feedforward and feedback connections [8]. There are several mathematical models that quantify hierarchical processing in the brain, such as predictive coding [9] and Free Energy Principle (FEP) [10, 11]. The latter is of particular interest as it describes a unified theory that accounts for learning, recognition and action [11].

In a series of papers [12, 13] Yildiz and colleagues used Dynamic Expectation Maximization (DEM) [14] implementation of FEP to construct two-level generative model for birdsong generation [12] and human speech recognition [13]. The Hopfield network [3] is used in the model to represent the amplitude modulation of the spectrogram of the input sound, whether it is birdsong or speech. The input to the Hopfield network is defined with a stable heteroclinic channel (SHC) [15]. Depending on the active unit, the global attractor of the Hopfield network changes, encoding the spectro-temporal pattern of the birdsong or speech segment. The authors demonstrated that the model achieved high levels of performance in digit recognition task, comparable to state-of-the-art ASR models, despite having considerably fewer parameters [13]. In recent years, Hovsepyan and colleagues have expanded the model to include the notion of neural oscillations and applied it to an online syllable recognition task [16, 17]. By utilizing the added coupled theta-gamma oscillation, the model successfully segmented continuous speech into syllable segments and identified them with an accuracy of approximately 50-60%, compared to a chance level of around 8%.

In this paper, we draw inspiration from the aforementioned models to propose a neurophysiologically plausible model for detecting speech pathology. The proposed model also has a two-level hierarchy. The SHC is used to determine the global attractor of the Hopfield network, which encodes the amplitude modulations of the input speech spectrogram. In addition, we have included evidence accumulators (as in [16, 17]) to determine whether the input speech sample corresponds to a healthy condition (HC) or Parkinson's disease condition (PC). Our ultimate objective is to develop a neurophysiologically plausible model of speech production/perception that can manipulate (learn, generate, infer) both linguistic and paralinguistic aspects of speech independently. This study demonstrates the proof of concept by showing that existing neurocomputational models of speech perception already possess the capacity to engage with paralinguistic scenarios

The remainder of the paper is structured as follows: Section 2 details the proposed methodology, including the mathematical description of the model and classification procedure. Section 3 outlines the speech database and data preprocessing. Finally, Section 4 presents the results, and Section 5 concludes the findings and provides an outlook.

2. Proposed algorithm

In this section, we describe the proposed algorithm. First, we provide a mathematical description of the generative model using equations adapted from [13, 16]. For further details, please refer to these papers. We then outline the classification proce-

10.21437/Interspeech.2024-1041

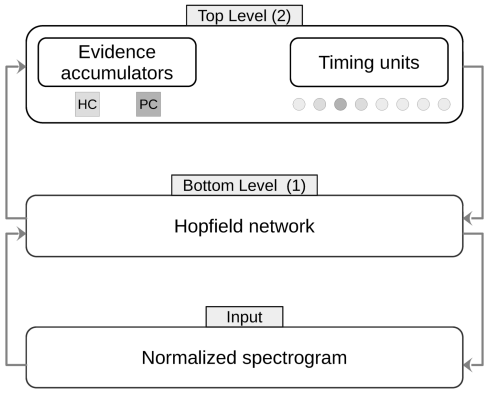


Figure 1: *This is a diagram of the proposed generative model. The top level contains evidence accumulators for PC and HC, as well as timing units. Depending on the active timing unit (dark circle), the respective spectral vector of each condition weighted with the respective amount of accumulated evidence is sent to the bottom level as input to the Hopfield network. The latter encodes the frequency amplitude fluctuations that are compared with the input spectrogram. The model operates within a predictive coding framework, with arrows indicating the flow of information between levels of hierarchy: top-down predictions and bottom-up prediction errors.*

dure.

2.1. The generative model

Similar to the models mentioned previously, we utilised a predictive coding framework to identify speech pathology. However, there is a significant conceptual difference. In Hovsepyan et al. (2020) [16], the model inferred the identity of the syllables by selecting the pattern in the model's memory that most closely matched the syllable input. In our proposed model, we only retain two patterns in memory: one for the healthy condition (HC) and one for the pathological condition (PC) (Figure 1). During inference, the model is tasked with identifying the corresponding condition of the input syllable. The syllable's identity in the input and the model's memory is always the same for all simulations. For example, if the input syllable is '-pa', the spectrotemporal patterns in the model's memory correspond to '-pa', with one for HC and the other for PC. The presented generative model has two levels of hierarchy, where the dynamics of each level are determined by the corresponding hidden states and the information from the level above (if applicable) via the causal states. The DEM [14] algorithm was used to invert the generative model for the inference process. The equations of the dynamics (time derivatives) of the hidden states for each level are described below. The information sent to the level below is described by the corresponding causal states. The observation noise for hidden and causal states for each level is described by ε and η respectively, where the superscript refers to the hierarchical level of the model and the subscript describes the functional group.

At the **Top level**, there are two functional groups: evidence accumulators and timing units. The evidence accumulators are perfect integrators (Equation 1) that accumulate evidence for either the HC or PC condition (Equation 2). The information about the accumulated evidence is then sent to the level below with the corresponding causal states (Equation 3). Here, they are scaled with the softmax, which represents the model-estimated probabilities of whether the input syllable corresponds to the HC or PC.

$$\frac{d\omega}{dt} = 0 + \varepsilon_{\omega}^{(2)} \tag{1}$$

$$\omega = [\omega_{HC}; \omega_{PC}] \tag{2}$$

$$\upsilon_{\omega}^{(1)} = \frac{e^{-\omega}}{\sum e^{-\omega}} + \eta_{\omega}^{(2)} \tag{3}$$

Timing units (Equation 4), on the other hand, are modelled as stable heteroclinic channel [15, 13], thus they have sequential activation: meaning that each moment of time only one unit is active and correspond part of the syllable is deployed in time.

$$\frac{dx}{dt} = \kappa_2[-\lambda x - \rho(1 + e^{-x})^{-1} + 1] + \varepsilon_{x}^{(2)} \tag{4}$$

here x is a vector of 8 units forming the stable heteroclinic channel, ρ is the connectivity matrix, indicating the inhibition strength from unit j to i, $\kappa_2 = 0.525$, $\lambda = 0.1$. To ensure that the heteroclinic channel is scaled between [0 1], the following transformation was used.

$$\frac{dy_i}{dt} = e^{x_i} - y_i \sum_{j=1}^{8} e^{x_j} + \varepsilon_{y_i}^{(2)} \tag{5}$$

The respective causal states (Equation 6) send the scaled SHC to the level below.

$$\upsilon_{y}^{(1)} = y + \eta_{y}^{(2)} \tag{6}$$

The **Bottom level** uses information from the top level to model amplitude fluctuations of the frequency channels (Figure 1). This is achieved through the Hopfield network [3] (Equation 7), with the input (Equation 8) changing depending on which timing unit is active. The input is constructed based on the standardized representation of syllables (Figure 2), where each spectral vector serves as a global attractor in the Hopfield network. For each condition, two sets of global attractors are calculated, as detailed in the next section (Equation 9). The weighted average of these two attractors is then used as input to the Hopfield network. During inference, the model aims to minimize prediction errors by accumulating evidence for each condition (Figure 3). The weighted average global attractor assigns weight based on the accumulated evidence for each condition.

$$\frac{dx^{(1)}}{dt} = \kappa_1[-Ax^{(1)} + Wtanh(x^{(1)}) + I] + \varepsilon^{(1)} \tag{7}$$

A is the diagonal self-connectivity matrix, while W is an asymmetric synaptic connectivity matrix with values ranging from -1 to 1. These matrices are designed to satisfy special conditions that ensure the Hopfield network has a global attractor, the location of which depends on I. For more details, refer to Yilidz et al. (2011) [12].

$$I_f = \sum_{\gamma=1}^{8} (\omega_{HC} P_{f\gamma}^{HC} + \omega_{PC} P_{f\gamma}^{PC}) \upsilon_{\gamma}^{1} \tag{8}$$

Where $P_{f\gamma}^{HC}$ and $P_{f\gamma}^{PC}$ are spectrotemporal patterns in the Hopfield space associated with HC and PC respectively. Those

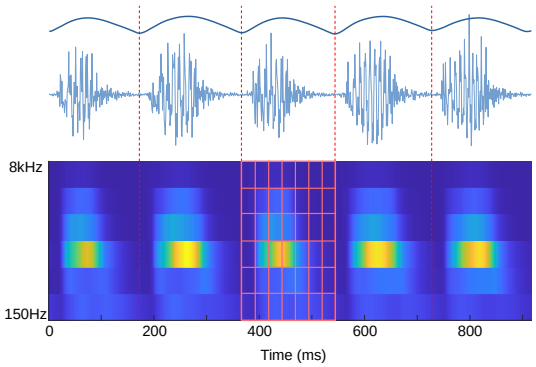


Figure 2: *The figure presents an example of a DDK utterance: the waveform is shown at the top, with dashed red lines aligned with local minima of the sonority envelope (dark blue) indicating syllable onsets. At the bottom, the collapsed 6-channel auditory spectrogram is displayed. The spectrogram of each syllable (contained between two consecutive syllable onsets) is divided into eight equal temporal bins. The average activity across six frequency channels within each bin is calculated to parameterize the syllable.*

are calculated based on the global attractor in the Hopfield space of each syllable σ with the following equation:

$$P_{f\gamma\sigma} = \sum_{i=1}^{6} A_{fi} S_{i\gamma\sigma} - \sum_{i=1}^{6} W_{fi} tanh(S_{i\gamma\sigma}) \qquad (9)$$

In the Equations 8 and 9 S_σ denotes to the standardized spectrogram of a syllable σ (Figure 2), f and γ index frequency bands and timing units respectively. Finlay, i represents the index of the row (column) during matrix multiplication.

The causal states (Equation 10) at this level determine the amplitude fluctuations of the generated frequency, which are then compared with the spectrotemporal patterns of the input syllable.

$$v^{(0)} = x^{(1)} + \eta^{(1)} \qquad (10)$$

Implementation of the algorithm was carried out by adapting the publicly available code from [16]. The latter is designed to recognize syllables with natural duration's from continuous speech, however for our current purposes, a more streamlined approach with a single, normalized (in duration) syllable as input (as in [13]) is more applicable.

2.2. Classification procedure

For the classification process the generative model was inverted with DEM algorithm [14] and an inference was performed for each syllable. Figure 3 presents a typically dynamics of evidence accumulators during the inference process. The accumulated evidence at the end of the inference and the average across the inference are stored (Figure 3), representing the model's estimated probabilities for which health condition the input syllable corresponds to.

3. Experimental setup

The proposed algorithm is, in principle, independent of the paralinguistic scenario. The objective of this study is to test, as a

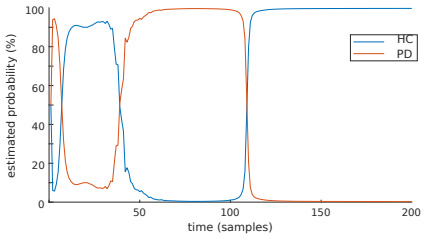


Figure 3: *An example of the dynamics of evidence accumulation during the inference process. It specifically refers to the input syllable in the HC condition.*

first attempt, whether neurocomputational models can be used for speech pathology detection. The specific question we are interested in is the ability of the model to determine whether the input speech sample corresponds to a healthy patient or a patient with Parkinson's disease - a neurodegenerative disease that is often associated with impairment of motor control and, consequently, speech articulation [18]. For a such scenario, diadochokinetic (DDK) utterances are suitable recordings [19, 20], as they assess speech articulation.

3.1. Database

In this study we utilized DDK utterances from the PC-GITA database [19], which includes recordings (in Spanish) of 50 individuals with PD (25 male, 62.2±11.2 years old, 25 female 60.1±7.8 years old) and 50 HC individuals (25 male, 61.2±11.3 and 25 female 60.7±7.7 years old). All recordings in the PC-GITA database are sampled at 44.1kHz with 16-bit resolution. The database contains recordings of 6 different DDK utterances. However, for this study, we only selected DDK utterances that repeat the same syllable (-pa, -ta, or -ka). This decision was made to test the model's ability to detect differences in the input speech sample due to pathological conditions, rather than differences in syllable identity.

3.2. Data preprocessing

The speech data from the PC-GITA database was preprocessed before being used for simulations (Figure 2). The preprocessing involved resampling recordings to 16kHz, dividing each utterance into syllables, computing the condensed spectrogram of each utterance and parameterized representation of each syllable. The preprocessing steps are explained below.

Segmentation. The DDK utterance is segmented into syllables using a neurophysiologically plausible model that identifies local minima of the sonority envelope corresponding to syllable onsets [21]. A syllable segment is defined as the segment of speech between two consecutive onsets. Although any other automatic syllabification algorithm could have been used for this task (e.g. [22, 23]), this model has the advantage of being neurophysiologically plausible and uses theta oscillations to identify syllable onsets [4, 5].

Time-frequency decomposition. The calculation of the auditory spectrogram was the second step of the preprocessing. As with segmentation, a neurophysiologically plausible model of the auditory periphery [24] was used to compute the time-frequency decomposition. The model decomposes the speech waveform into logarithmically spaced 128 frequency channels spanning from approximately 150Hz to 8kHz. Following the procedures described in previous studies [16], we then col-

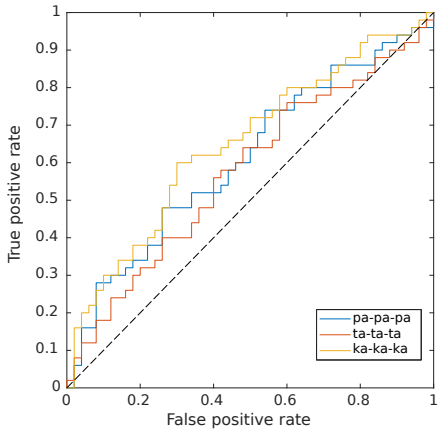


Figure 4: *Receiver operating curve for each DDK utterance*

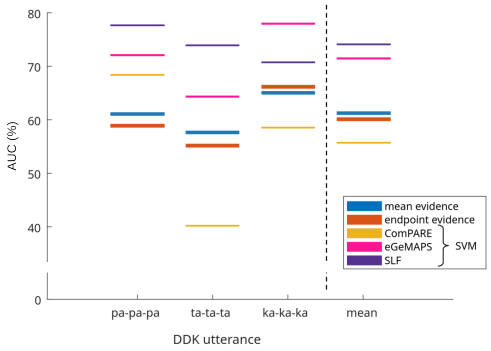


Figure 5: *The graph displays the area under ROC for each DDK utterance, as well as the mean across utterance types. Thick dashes correspond to the AUC values from the proposed algorithm, whereas narrow dashes correspond for the baseline taken from [26].*

lapsed the 128-channel auditory spectrogram into 6 channels by averaging the activity of the adjacent channels. The main reason for this step is to reduce the computational load, while still retaining enough information for intelligible speech [25]. However, it is worth noting that since our goal is not syllable recognition, the dimensionality of the collapsed spectrogram may be different and more selective (something to be explored in future iterations of the study).

Syllable parameterization. The collapsed spectrogram was segmented into syllabic segments using the syllable onsets (Figure 2). From each DDK utterance, the spectrotemporal decomposition of each syllable was extracted. To ensure consistency, we resampled each syllabic spectrogram to have a fixed number of samples (200 samples, as in [13]). Using the same procedure, the duration of each syllable was further divided into 8 equal temporal bins and the average value of each frequency band within the corresponding temporal bin was calculated, resulting in a 6x8 matrix (Figure 2). This matrix represents a syllable in a time-invariant manner, indicating a sequence of spectral vectors (each column) that must be deployed in time to generate the syllable's spectrogram.

3.3. Experimental protocol

A leave-one-subject-out protocol was used for the classification procedure. The global pattern of P^{HC} and P^{PD} for HC and PD conditions was calculated by averaged across all syllables corresponding to each condition. The syllable patterns corresponding to the left out utterance were excluded from the calculation of the global patterns for each fold. Moreover, to obtain the estimate for each utterance, the scores of the left-out utterance's syllables are averaged. The receiver operating curves (ROC) (Figure 4) as well as the corresponding AUC values (Figure 5) were calculated using the resulting scores.

4. Results

This section presents the investigation results. Figure 3 displays the time course of evidence accumulators during the inference. Initially, both conditions have equal probabilities, but as the inference progresses, the accumulated evidence changes. From around the 110th sample, the model 'selects' the correct condition. It is important to note that the selection process is not gradual, and the model can update its 'decision' and switch between selections based on the available evidence.

Figure 4 displays the resulting ROC for each DDK utterance based on the mean accumulated evidence. However, using scores based on the end-values yields very similar results, as shown by the area-under-the-curve (AUC) values in Figure 5 (thick dashes). The results show that performance is dependent on the DDK utterance. The 'ka-ka-ka' utterance generally results in better classification with an AUC value of around 65%, while the 'ta-ta-ta' utterances generally result in lower performance, reaching around 57% AUC value. Noteworthily, for all conditions, the performance is higher than the chance level (50%).

In a recent study [26], in a similar manner, syllable-level features were extracted and classified using SVM. The results showed that 'ka-ka-ka' performed the best, while 'ta-ta-ta' had the worst results. Although the classification was better overall by around 10% in terms of AUC value, the trend was similar. Additionally, the study found that traditional feature sets like ComParE2016 (6373 features) [27] and eGeMAPS (88 features) [28] resulted in 55.71% and 74.1% AUC, respectively (SVM results in Figure 5).

5. Conclusion and Outlook

This paper presented an investigation on whether (neuro-) physiologically plausible models, which are often linked with linguistic/ASR challenges, can be useful also for paralinguistic scenarios. As a case, we explored whether the neurocomputational model of speech recognition would be suitable for speech pathology (more specifically, Parkinson's disease speech) detection tasks. The results, although modest, suggest that the proposed methodology has merit and warrants further investigation. However, there are several ways to improve the model results, such as increasing the temporal/spectral resolution (number of timing units/frequency bands), changing the FEP minimisation method, interrupting the decision process based on a predefined threshold of accumulated evidence, removing the restriction of normalised syllable duration and the requirement of single syllable utterances. Overall, this paper serves as a proof-of-principle and paves the way for further research on the use of (neuro-)physiologically plausible models for paralinguistic challenges.

6. Acknowledgments

This work was funded by the Swiss National Science Foundation through the Bridge Discovery project EMIL: Emotion in the loop - a step towards a comprehensive closed-loop deep brain stimulation in Parkinson's disease (grant no. 40B2–0_194794).

7. References

[1] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," Aug. 2023.

[3] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 8, pp. 2554–2558, Apr. 1982.

[4] A. Hyafil, A.-L. Giraud, L. Fontolan, and B. Gutkin, "Neural Cross-Frequency Coupling: Connecting Architectures, Mechanisms, and Functions," *Trends in Neurosciences*, vol. 38, no. 11, pp. 725–740, Nov. 2015.

[5] A. Hyafil and M. Cernak, "Neuromorphic based oscillatory device for incremental syllable boundary detection," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2015-Janua. ISCA, 2015, pp. 1191–1195.

[6] M. Schrimpf, I. A. Blank, G. Tuckute, C. Kauf, E. A. Hosseini, N. Kanwisher, J. B. Tenenbaum, and E. Fedorenko, "The neural architecture of language: Integrative modeling converges on predictive processing," *Proceedings of the National Academy of Sciences*, vol. 118, no. 45, p. e2105646118, Nov. 2021.

[7] E. Fedorenko and S. L. Thompson-Schill, "Reworking the language network," *Trends in Cognitive Sciences*, vol. 18, no. 3, pp. 120–126, Mar. 2014.

[8] J. P. Rauschecker and S. K. Scott, "Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing," *Nature Neuroscience*, vol. 12, no. 6, pp. 718–724, Jun. 2009.

[9] R. P. N. Rao and D. H. Ballard, "Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects," *Nature Neuroscience*, vol. 2, no. 1, pp. 79–87, Jan. 1999.

[10] C. L. Buckley, C. S. Kim, S. McGregor, and A. K. Seth, "The free energy principle for action and perception: A mathematical review," *Journal of Mathematical Psychology*, vol. 81, pp. 55–79, Dec. 2017.

[11] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, Feb. 2010.

[12] I. B. Yildiz and S. J. Kiebel, "A hierarchical neuronal model for generation and online recognition of birdsongs," *PLoS Computational Biology*, vol. 7, no. 12, 2011.

[13] I. B. Yildiz, K. von Kriegstein, and S. J. Kiebel, "From Birdsong to Human Speech Recognition: Bayesian Inference on a Hierarchy of Nonlinear Dynamical Systems," *PLoS Computational Biology*, vol. 9, no. 9, pp. e1 003 219–e1 003 219, Sep. 2013.

[14] K. J. Friston, N. Trujillo-Barreto, and J. Daunizeau, "DEM: A variational treatment of dynamic systems," *NeuroImage*, vol. 41, no. 3, pp. 849–885, Jul. 2008.

[15] P. Varona, M. I. Rabinovich, A. I. Selverston, and Y. I. Arshavsky, "Winnerless competition between sensory neurons generates chaos: A possible mechanism for molluscan hunting behavior," *Chaos*, vol. 12, no. 3, pp. 672–677, Sep. 2002.

[16] S. Hovsepyan, I. Olasagasti, and A.-L. Giraud, "Combining predictive coding and neural oscillations enables online syllable recognition in natural speech," *Nature Communications*, vol. 11, no. 1, p. 3117, Jun. 2020.

[17] ——, "Rhythmic modulation of prediction errors: A top-down gating role for the beta-range in speech processing," vol. 19, no. 11, p. e1011595. [Online]. Available: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011595

[18] J. Rusz, J. Rusz, R. Cmejla, H. Ruzickova, E. Růžička, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," *Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 350–367, Feb. 2011.

[19] J. R. Orozco, J. D. Arias-Londoño, J. Vargas-Bonilla, M. González-Rátiva, and E. Noeth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," May 2014.

[20] A. Bayestehtashk, M. Asgari, I. Shafran, and J. McNames, "Fully Automated Assessment of the Severity of Parkinson's Disease from Speech," *Computer Speech & Language*, vol. 29, no. 1, pp. 172–185, Jan. 2015.

[21] O. Räsänen, G. Doyle, and M. C. Frank, "Pre-linguistic segmentation of speech into syllable-like units," *Cognition*, vol. 171, pp. 130–150, Feb. 2018.

[22] P. Mermelstein, "Automatic segmentation of speech into syllabic units," *The Journal of the Acoustical Society of America*, vol. 58, no. 4, pp. 880–883, Oct. 1975.

[23] R. Villing, T. Ward, and J. Timoney, "Performance limits for envelope based automatic syllable segmentation," in *IET Irish Signals and Systems Conference (ISSC 2006)*, vol. 2006. IEE, 2006, pp. 521–526.

[24] T. Chi, P. Ru, and S. A. Shamma, "Multiresolution spectrotemporal analysis of complex sounds," *The Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 887–906, 2005.

[25] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, no. 5234, pp. 303–304, Oct. 1995.

[26] S. Hovsepyan and M. Magimai.-Doss, "Syllable level features for Parkinson's disease detection from speech," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024.

[27] B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, and K. Evanini, "The INTERSPEECH 2016 Computational Paralinguistics Challenge: Deception, Sincerity & Native Language," in *Interspeech 2016*. ISCA, Sep. 2016, pp. 2001–2005.

[28] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. Andre, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong, "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, Apr. 2016.