# Analyzing the Impact of Acoustic Features on Parkinson's Disease Severity Prediction

Pien Rooijendijk
*s1054190*

**LET-REMA-LCEX** Automatic Speech Recognition *June 2024*

## Contents

**1 Introduction** **3**

**2 Related work** **3**
2.1 Parkinson's Disease . . . 3
2.2 UPDRS . . . 3
2.3 Automatic Speech Recognition . . . 4
2.3.1 ASR models . . . 4
2.3.2 Acoustic features . . . 4
2.3.3 Subsets of the features . . . 5

**3 Method** **5**
3.1 Speech task . . . 5
3.2 Participants . . . 6
3.3 Feature selection . . . 7

**4 Experiment** **7**
4.1 Feature Extraction . . . 7
4.2 Model's architecture . . . 7
4.2.1 Evaluation metrics . . . 8

**5 Results** **9**
5.1 Results based on the losses and MSE . . . 9
5.2 Results based on accuracy . . . 10

**6 Discussion** **11**
6.1 Discussion on Performance Metrics . . . 11
6.2 Discussion on Model Performance . . . 11
6.3 Discussion on the Dataset . . . 12
6.4 Discussion on most Relevant Feature . . . 12
6.5 Future Directions . . . 12

**7 Conclusion** **12**

**8 Reference list** **14**

**9 Appendices** **15**
9.1 Italian Phrases . . . . . 15
9.2 Date on the PD group . . . . . 16
9.3 All speech features . . . . . 17
9.4 Division of the features . . . . . 18
9.5 Data and Source Code . . . . . 18

# 1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder and is affecting millions of individuals worldwide. Early detection and accurate assessment of disease severity are crucial for effective management and intervention strategies. Speech impairments are common in PD patients and can provide valuable insights into disease progression. In recent years, machine learning techniques applied to speech data have shown promise in predicting PD severity based on various speech features. Features extracted from speech signals, such as Mel-Frequency Cepstral Coefficients (MFCC), Chroma, and Delta features, have been explored for their potential to capture underlying patterns related to PD severity. This research aims to investigate the influence of different speech features on predicting the severity of PD. Specifically, the focus is on MFCC, Chroma, and Delta features, hypothesizing that temporal changes captured by Delta features may be particularly informative for assessing disease severity. By examining the performance of different features, the aim is to identify the most relevant features for predicting PD severity in Italian-speaking patients. To achieve this, a Transformer-based model is built, known for its effectiveness in sequence modelling tasks, to learn the relationship between speech features and PD severity scores. Understanding the most influential speech features in predicting PD severity can provide valuable insights for developing more accurate and reliable diagnostic tools. Such tools can aid in early detection, monitoring disease progression, and optimizing treatment strategies for PD patients.

# 2 Related work

## 2.1 Parkinson's Disease

Parkinson's disease (PD) is a neurodegenerative disorder characterized by motor symptoms and non-motor symptoms, such as speech and voice impairments (Skodda, Visser, and Schlegel, 2011). The disease is characterized by neuronal loss in the brain, which causes a dopamine deficiency. The clinical diagnosis of Parkinson's disease is associated with many other non-motor symptoms that add to overall disability (Poewe et al., 2017). Dysarthria is common in PD and it affects articulation, and phonation, leading to reduced speech intelligibility. Skodda, Visser, and Schlegel, 2011 found in their research that vowel articulation was significantly reduced in PD patients compared to the control group, particularly in male patients. Furthermore, PD patients have slower verbal communication, and deficits in verb inflection which impairs spontaneous speech severely (Sonkaya et al., 2021).

## 2.2 UPDRS

The severity of Parkinson's Disease can be estimated by the Unified Parkinson's Disease Rating Scale (UPDRS) (Dimauro et al., 2017). The scales are as follows:

0: Normal: No speech problems.

1: Slight: Loss of modulation, diction or volume, while all words are easy to understand.

2: Mild: Loss of modulation, diction, or volume with a few unclear words, but the overall sentences are easy to follow.

3: Moderate: Speech is difficult to understand to the point that some, but not most, sentences are poorly understood.

4: Severe: Most speech is difficult to understand or unintelligible.

## 2.3 Automatic Speech Recognition

### 2.3.1 ASR models

Automatic Speech Recognition (ASR) is the process of automatically transcribing spoken language into text. It involves the use of computational algorithms and models to convert spoken words into written text, enabling machines to understand and interpret human speech. ASR systems typically utilize techniques from signal processing, machine learning, and natural language processing to accurately recognize and transcribe spoken language (Yu and Deng, 2016).
There has been research on language processing, which included researching which model could do this task the best (Devlin et al., 2018). As Dong et al., 2018 showed in their research, a Transformer model achieves a low word error rate (WER) and is trained relatively fast. The models are designed for sequence-to-sequence tasks, such that they are suitable for processing sequential data (Vaswani et al., 2017). Since the UPDRS prediction involves analysing sequential features from speech data, a transformer model can effectively capture the temporal dependencies and patterns in the PD data. Also, the models can utilize self-attention mechanisms to weigh the importance of different input elements when making predictions (Vaswani et al., 2017). For the UPDRS prediction, it is useful to focus on more relevant features which may be more informative than others.

### 2.3.2 Acoustic features

ASR systems can be used to analyze speech samples from individuals with PD by extracting various acoustic features. These features can include:

- Phonatory features: pitch, jitter and shimmer.
- Articulatory features: formant frequencies, speech rate, articulation rate.
- Prosodic features: intonation patterns, stress, rhythm.

Phonatory features refer in general to voice quality and loudness variations (Duffy et al., 2012). PD patients may present differences in their pitch, jitter and shimmer when speaking. Pitch variability may increase, leading eventually to a flat speech pattern known as hypophonia (Liotti et al., 2003). This speech feature is also closely related to the fundamental frequency of the PD patient's vibration during speech production. Jitter measures the variation in this fundamental frequency of the vocal vibration (Azadi et al., 2021). With PD, the jitter can be increased and this can result in irregularities in the vocal pitch. Azadi et al., 2021 describe shimmer as the variation in the amplitude of the speech wave. PD patients can have increased shimmer which may lead to fluctuations in loudness, this can be heard as a breathy or hoarse voice.
Second, the articulatory features in general are described as manner of speaking or articulation (Moro-Velazquez et al., 2021). It includes formant frequencies which are frequency peaks in the spectrum with high peaks of energy within the vocal spectrum (Malmkjaer, 2009). Another key aspect of articulatory features is the speech rate. For PD patients, speech acceleration is higher than for healthy patients. They also make longer pauses at the end of words (Skodda and Schlegel, 2008). The last feature is the articulation rate. This is the speed at which sounds are articulated. As PD patients have a form of dysarthria, this rate is influenced by their muscle control and coordination, resulting in a slower rate (Skodda, Visser, and Schlegel, 2011).
Finally, prosodic features capture changes in the fundamental frequency, amplitude of the voice and duration (Pell, 1996). It also can capture whether a sentence is a statement or a question. Furthermore, it can indicate whether the speaker is angry or sad. These features include intonation patterns. For Parkinsonian speakers, the intonation is significantly reduced (Skodda, Grönheit, and

Schlegel, 2011). Stress and rhythm are the other prosodic features, where stress contributes to the rhythm of speech.
According to (Harel et al., 2004) the most influential features are F0 variability and VOT (Voice Onset Time). These features were found to significantly change in individuals with PD following the initiation of pharmacological treatment. The study suggests that these changes in speech may manifest several years before the onset of obvious symptoms and an initial diagnosis. Additionally, the decreased dispersion around the mean F0 relates perceptually to less intonation or monotone inflection patterns, which is indicative of changes in the ability to control speech.
Based on these speech features and the literature I can make a hypothesis about which features will be the most influential when predicting the UPDRS of the PD patients. The speech features related to phonation, articulation and prosody are expected to be the most influential when predicting the UPDRS scores. Specifically, variations in pitch, jitter, and shimmer are likely to correlate strongly with UPDRS scores. Jitter and shimmer in particular would both in particular be increased in the speech analysis of the patients. Parkinsonian speech often exhibits monotonicity, reduced loudness (hypophonia), and fluctuations in pitch and voice quality, which are indicative of disease severity. Also, alterations in formant frequencies, speech rate, and articulation rate are expected to be influential predictors. The rate at which the patients speak is slower. Dysarthria in Parkinson's disease can lead to imprecise articulation, reduced speech rate, and changes in vowel production, which may reflect disease progression. Intonation patterns, stress, and rhythm are also likely to play a significant role in predicting UPDRS scores. Reduced intonation variability, abnormal stress patterns, and dysrhythmic speech are common characteristics of Parkinsonian speech and may be sensitive markers of disease severity.

#### 2.3.3 Subsets of the features

Making the hypothesis more specific, there are three subsets which capture the phonation, articulation and prosody of speech. Referring to the speech features in table 2, MFCC 1 to 13 and the Chroma Features capture acoustic properties related to both phonation and articulation and are likely to be more influential in predicting UPDRS scores compared to other features (Tracey et al., 2023). Another feature that would be the most influential are all of the Delta Features. These features represent temporal changes in the corresponding non-Delta features and may provide additional information about speech dynamics, potentially enhancing the predictive power for UPDRS scores. The other features such as Zero Crossing Rate, Energy and Spectral Centroid may capture the nuances of the speech of the patients but they are not as influential as other features or their influence is uncertain (Duffy et al., 2012). Out of these three categories, the most influential feature would be the Delta Features. These features capture the best the characteristics of the patient's speech which are the most related to PD.

# 3 Method

## 3.1 Speech task

The following tasks were used to record the speech and voice of the participants. Note that not all tasks are used for all participants, see next section which was used for which participant. The Italian text "*IL ROMARRO DELLA ZIA*" (see Appendix 9.1) was not chosen at random since the phonemically balanced text contains according to Dimauro et al., 2017 interesting features such as it is sufficiently long and requires for the patient to breathe with some effort, while it stresses for

resistance. It also contains complex phonetics to test the patient's ability to pronounce difficult sounds in a short time. Finally, it requires the patient to make changes in expression while reading.

a) 2 readings of a phonemically balanced text spaced by a pause (30 sec) (see Appendix 9.1 for the Italian text "*IL ROMARRO DELLA ZIA*").

b) execution of the syllable 'pa' (5 sec), pause (20 sec), execution of the syllable 'ta' (5 sec);

c) 2 phonation of the vocal 'a';

d) 2 phonation of the vocal 'e';

e) 2 phonation of the vocal 'i';

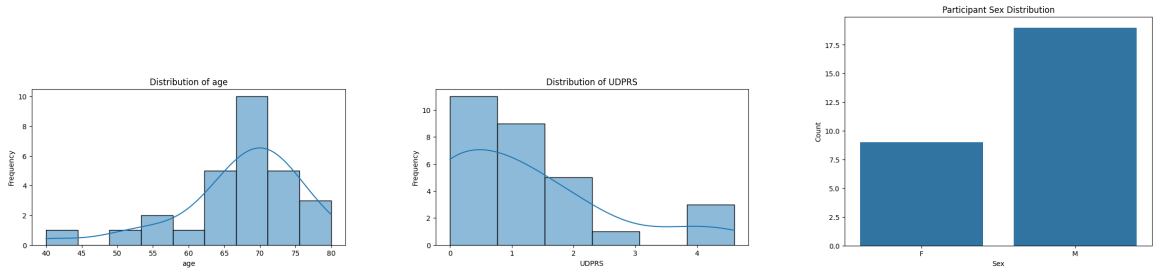
f) 2 phonation of the vocal 'o';

g) 2 phonation of the vocal 'e';

h) reading of some phonemically balanced words, pause (1 min), and reading of some phonemically balanced phrases (see Appendix 9.1 for the words and phrases).

## 3.2 Participants

The original data consisted of fifteen healthy people aged between 19 and 29 years and they were asked to perform a reading task where they precisely read a balanced Italian text and balanced words. The participants included 13 men and 2 women from the Puglia region in Italy and 13 men and 2 women from the Brindisi area.
The same reading experiment was conducted on 22 healthy elderly persons, aged between 60 and 77 years. All participants came from Bari which is in the Puglia region in Italy, of which were 10 men and 12 women. None of the healthy elderly participants reported any speech or language impairments.
The last group of the experiment is the group with Parkinsons' Disease group. This group consists of 28 patients aged between 40 and 80 years (see figure 1 for the age distribution). There were 19 men and 9 women (see figure 1 for the sex distribution) of which 27 were from the Bari area from the Puglia region in Italy and one from Venice. The patients reported that none of them had any speech or language disorders unrelated to the PD symptoms prior to the study conducted by Dimauro et al., 2017. The severity of their disease was classified by specialists on the scale of UPDRS (see the UPDRS scores in figure 6 and 1 for the distribution of the UPDRS scores). The patients performed the speech task from Appendix 9.1 in the same conditions as was described for the healthy participants.




(a) Distribution of the age of the PD patients.

(b) Distribution of the UPDRS of the PD patients.

(c) Distribution of the sex of the PD patients.

Figure 1: Distrubtion of the PD patients.

However, for some of the speech tasks some data was missing. For the reading of the phonemically balanced text (see Appendix 9.1 for the Italian text), some of the data was incomplete. The data of the PD patients with a UPDRS scale of 4 were missing. The reading of some phonemically balanced words was also missing. Lastly, for all phonation of the vocal vowels, the data was missing from one of the patients for all the second readings.

### 3.3 Feature selection

From all the extracted features, one subset was initiated with only the MFCC features, one with the Chroma features and the last one with the Delta features. The last subset was also divided into two other subsets consisting of the Delta MFCC and Delta Chroma features. See table 3 in the Appendices (section 9) for the subsets based on the most relevant features. After the features are divided, the model will be trained on these subsets and the most relevant feature will become apparent. See the next section for the model's architecture and evaluation metrics.

## 4 Experiment

### 4.1 Feature Extraction

A tool to listen to audio files was used to extract the features and understand their sounds. The Python library *pyAudioAnalysis* for audio feature extraction was used (Giannakopoulos, 2015). The feature extraction resulted in features and representations as listed in table 2. Instead of focusing on every little detail, the average of each of the features was taken. This was done for each audio file, collecting all the averages together. Then the dataset was assembled by concatenating the features for each of the different speech tasks for all of the Parkinson's patients.

### 4.2 Model's architecture

A Transformer-based regression model for predicting UPDRS scores from the extracted audio features was implemented. The input data is first passed through a stack of transformer encoder layers. These transformer layers employ self-attention mechanisms to capture relationships between different parts of the input sequence. Each transformer encoder layer is composed of multiple sub-layers, including multi-head self-attention and position-wise feedforward networks. After encoding the input sequence with the transformer layers, the output is passed through a series of feedforward neural

network layers. The hidden layers consist of linear transformation followed by activation functions (ReLU), batch normalization, and dropout regularization. Finally, the output of the feedforward neural network layers is passed through a linear transformation to produce the final prediction. The output layer maps the hidden representations to a single output value, which is the predicted UPDRS score. See the code below for the full implementation of the model.

```
class TransformerModel(nn.Module):
    def __init__(self, input_dim, hidden_dim=128, num_layers=2, dropout=0.1):
        super(TransformerModel, self).__init__()
        self.transformer_layers = nn.TransformerEncoder(
            nn.TransformerEncoderLayer(d_model=input_dim, nhead=1, dropout=dropout),
            num_layers=num_layers
        )
        self.hidden_layers = nn.Sequential(
            nn.Linear(input_dim, hidden_dim),
            nn.ReLU(),
            nn.BatchNorm1d(hidden_dim),
            nn.Dropout(dropout),
            nn.Linear(hidden_dim, hidden_dim),
            nn.ReLU(),
            nn.BatchNorm1d(hidden_dim),
            nn.Dropout(dropout)
        )
        self.output_layer = nn.Linear(hidden_dim, 1)

    def forward(self, x):
        x = self.transformer_layers(x)
        x = self.hidden_layers(x)
        x = self.output_layer(x)
        return x.squeeze()
```

Listing 1: Architecture of the Transformer model.

The model was trained using the Mean Squared Error (MSE) loss function and optimized using the Adam optimizer. A batch size of 32 was utilized and the model was trained for 50 epochs with a learning rate of 0.2.
The dataset was split into training and testing sets with a ratio of 80:20, ensuring that the same patient's data did not appear in both sets. The training set was used to train the model, while the testing set was used to evaluate its performance.

#### 4.2.1 Evaluation metrics

The model's performance was evaluated using two primary metrics:

1. **Mean Absolute Error (MAE)**: This metric measures the average absolute difference between the predicted and actual UPDRS scores. Lower values indicate better performance.

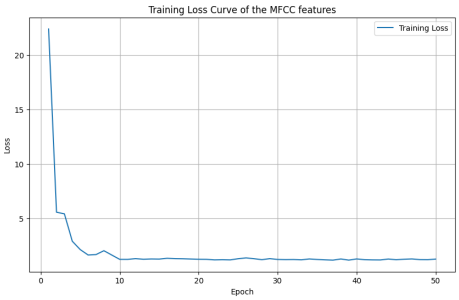
2. **Mean Squared Error (MSE)**: This metric measures the average squared difference between the predicted and actual UPDRS scores. Again, lower values indicate better performance.

The different subsets (see table 3 from Appendix 9.4) will be evaluated on the metric above. The lower the value, the better the model was able to predict the UPDRS score of that patient.

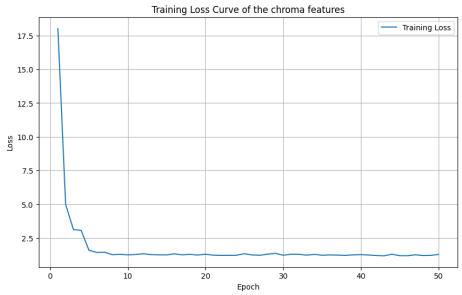
# 5 Results

## 5.1 Results based on the losses and MSE

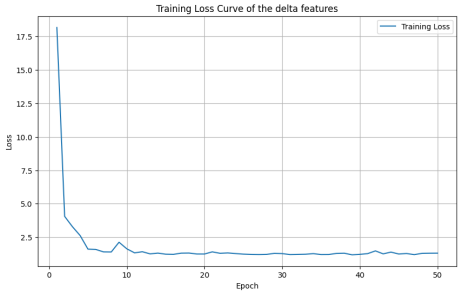
The loss curves for the three subfeatures are displayed in figure 2. The Delta features are also divided into subfeatures of which the training losses can be seen in figure 3.




(a) Training loss curve of the MFCC features.

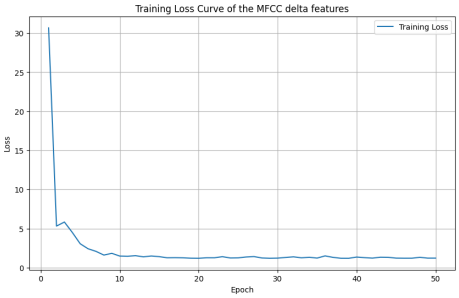



(b) Training loss curve of the Chroma features.

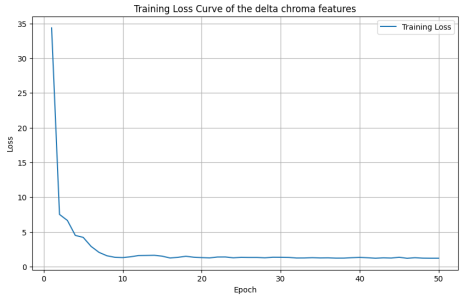



(c) Training loss curve of the Delta features.

Figure 2: Training loss curves of the features.




(a) Training loss curve of the MFCC Delta features.




(b) Training loss curve of the Chroma Delta features.

Figure 3: Training loss curves of the subfeatures of the Delta features.

As can be seen in figures 2 and 3, the training converges very quickly after 20 epochs.

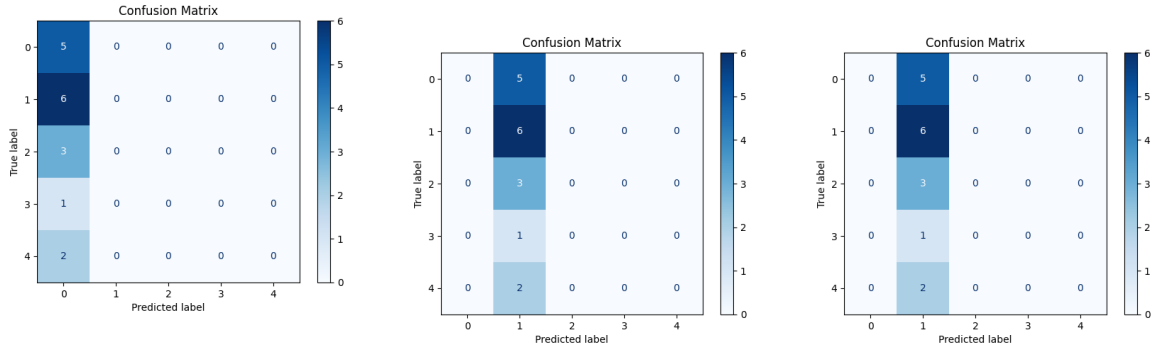
| Features | Test Loss | Training MAE | Training MSE | Testing MAE | Testing MSE |
|---|---|---|---|---|---|
| MFCC | 1.7656 | 0.8260 | 1.2417 | 0.9417 | 1.7656 |
| Chroma | 1.7504 | 0.8345 | 1.2402 | 0.9473 | 1.7504 |
| Delta | 1.7293 | 0.8488 | 1.2399 | 0.9571 | 1.7293 |
| MFCC Delta | 1.7260 | 0.8512 | 1.2401 | 0.9588 | 1.7260 |
| Chroma Delta | 1.7005 | 0.8713 | 1.2439 | 0.9727 | 1.7005 |

Table 1: The test loss, MAE and MSE of the different features.

Comparing the features, the Chroma Delta features have the lowest Test Loss (1.7005). Additionally, the Training MAE (0.8713) and Testing MAE (0.9727) for Chroma Delta are competitive, and its Testing MSE (1.7005) is also the lowest, followed closely by MFCC Delta (1.7260). Therefore, Chroma Delta is considered the best-performing feature according to the metric from table 1.

## 5.2 Results based on accuracy

After training the Transformer model for 50 epochs, a training accuracy of 32.11% and a testing accuracy of 35.29% was achieved. See figures 4 and 5 for the confusion matrices. As can be seen in the confusion matrices, the model predicts either 0 or 1 for the UPDRS for all the participants. Solely based on these results, it cannot be concluded that one of the models with the trained features is better than another feature.

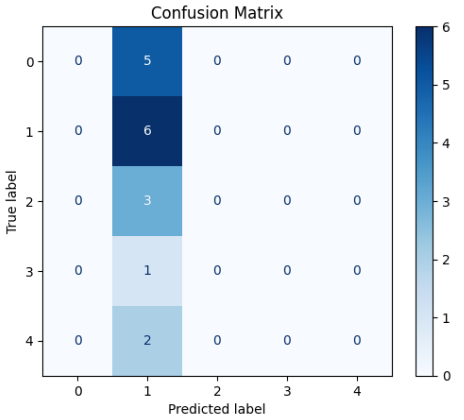



(a) Confusion matrix of the MFCC features.

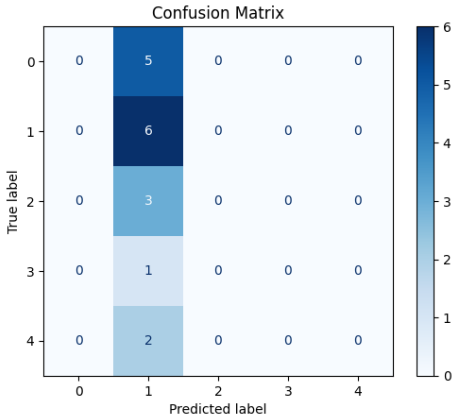
(b) Confusion matrix of the Chroma features.

(c) Confusion matrix of the Delta features.

Figure 4: Confusion matrices of the features.




(a) Confusion matrix of the MFCC Delta features.




(b) Confusion matrix of the Chroma Delta features.

Figure 5: Confusion matrices of the subfeatures of the Delta features.

# 6 Discussion

## 6.1 Discussion on Performance Metrics

The loss curves for the three subfeatures, namely MFCC, Chroma, and Delta features, are displayed in figure 2, while the training losses for the Delta features' subfeatures are shown in figure 3. Comparing these features, it's observed that the Chroma Delta features exhibit the lowest test loss (1.7005) among all features, indicating superior performance. Additionally, both the training MAE (0.8713) and testing MAE (0.9727) for Chroma Delta are competitive, and its testing MSE (1.7005) is the lowest, followed closely by MFCC Delta (1.7260). Therefore, Chroma Delta can be considered the best-performing feature based on the metrics from table 1.
Regarding the results based on accuracy, after training the Transformer model for 50 epochs, a training accuracy of 32.11% and a testing accuracy of 35.29% were achieved. The confusion matrices displayed in figures 4 and 5 illustrate that the model predicts either 0 or 1 for the UPDRS score for all participants, indicating that the model's predictions are not effectively capturing the variance in the UPDRS scores. One probable reason why the accuracy is so low is that the model is solely trained on one subset of features extracted from the speech of PD patients. To get a better prediction of the UPDRS score, the model should get all acoustic features to learn the full characteristics of PD speech deficiency.

## 6.2 Discussion on Model Performance

The obtained results suggest that although the model achieves a moderate level of accuracy, its predictions are limited in capturing the variability in UPDRS scores. This may indicate underlying issues with the model's ability to learn meaningful patterns from the features or with the representation of the UPDRS scores. The discrepancy between training and testing accuracies also suggests potential underfitting issues that need to be researched even further. Possible improvements include exploring more complex model architectures, fine-tuning hyperparameters, and incorporating additional features or data augmentation techniques to enhance the model's predictive power. Further

analysis is needed to refine the model for more accurate UPDRS score predictions. Training the model for more epochs will have no influence since the models are already converged after 20 epochs when taking into account the training loss curves from figures 2 and 3.
As for this research, it was deliberately chosen to make the model not too complex. The sole purpose was to investigate the influence of the different input features. By keeping the model relatively simple, it becomes easier to isolate the effects of the individual input features when predicting the UPDRS scores. Complex layers might obscure the impact of specific features by introducing additional layers of abstraction. Also, too complex layers have a risk of overfitting the data. Since the dataset is rather small, the model might tend to overfit this dataset.

### 6.3 Discussion on the Dataset

The dataset consists of audio recordings for different speech tasks obtained from PD patients along with their UPDRS scores. The size of dataset is a critical point when reviewing this research. Only 28 patients participated in the original study by Dimauro et al., 2017. This small dataset might have impacted the statistical power of the analysis. The distribution of the patients was not a good representation of the different disease severity levels. However, for one of the speech tasks, two patient's data with a UPDRS score of 4 was missing. This resulted in figure 4a, where the prediction of the UPDRS scores severely shifted towards only predicting a UPDRS score of 0 for all of the patients.

### 6.4 Discussion on most Relevant Feature

As stated before, Chroma Delta is the best-performing feature based on the metrics from table 1. As was hypothesised before according to the literature, the Delta Features would be the most influential. The Delta Features can capture the temporal changes in the speech features. The changes in pitch and intonation are common symptoms of PD, making Chroma Features potentially informative for disease severity assessment.

### 6.5 Future Directions

For future work, the model could be improved or another model could be used. The intricate pattern of speech features needs to be captured more effectively. For this new model fine-tuning, data augmentation or model optimization can be used. When making this new and better model, a better data set should be considered. A larger and more diverse dataset and a broader range of UPDRS scores to improve model generalization should be gathered.
Another direction would be to use the outcome of this research to focus on these features, in particular the most important feature Chroma Delta, when predicting PD in the earlier stages. These features show promise as potential indicators of early disease onset or progression. Early prediction of PD can significantly impact patient outcomes by enabling early intervention and personalized treatment strategies. Identifying individuals at risk of developing PD before the onset of motor symptoms could lead to interventions aimed at slowing disease progression and improving quality of life.

## 7 Conclusion

In this study, the influence of different speech features on predicting the severity of Parkinson's disease (PD) was investigated. Through the analysis, I found that Chroma Delta features emerged as the most relevant feature, showing a better performance in predicting UPDRS scores compared to

MFCC and Chroma features. These findings align with the hypothesis that temporal changes captured by Delta features are particularly informative for assessing disease severity. Despite achieving moderate accuracy in predicting UPDRS scores, the model succeeded in capturing the variability of UPDRS scores solely based on speech features. This suggests the need for further refinement and exploration of more complex model architectures or additional features to improve predictive accuracy. The dataset's limitations, including its small size and uneven distribution across disease severity levels, should be considered when interpreting the results. Additionally, missing data for certain patients posed challenges in model training and evaluation. Looking ahead, the insights gained from this study could be leveraged to develop more accurate predictive models for early detection of PD. In conclusion, while the research sheds light on the relevance of specific speech features for predicting PD severity, further research is needed to refine models, address dataset limitations, and translate findings into clinical practice for improved patient care and management of Parkinson's disease.

# 8 Reference list

# References


Azadi, H., Akbarzadeh-T, M.-R., Shoeibi, A., Kobravi, H. R., et al. (2021). Evaluating the effect of parkinson's disease on jitter and shimmer speech features. *Advanced Biomedical Research*, *10*(1), 54.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dimauro, G., Di Nicola, V., Bevilacqua, V., Caivano, D., & Girardi, F. (2017). Assessment of speech intelligibility in parkinson's disease using a speech-to-text system. *IEEE Access*, *5*, 22199–22208.

Dong, L., Xu, S., & Xu, B. (2018). Speech-transformer: A no-recurrence sequence-to-sequence model for speech recognition. *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 5884–5888.

Duffy, J. R., et al. (2012). *Motor speech disorders: Substrates, differential diagnosis, and management*. Elsevier Health Sciences.

Giannakopoulos, T. (2015). Pyaudioanalysis: An open-source python library for audio signal analysis. *PloS one*, *10*(12).

Harel, B. T., Cannizzaro, M. S., Cohen, H., Reilly, N., & Snyder, P. J. (2004). Acoustic characteristics of parkinsonian speech: A potential biomarker of early disease progression and treatment. *Journal of Neurolinguistics*, *17*(6), 439–453.

Liotti, M., Ramig, L., Vogel, D., New, P., Cook, C., Ingham, R., Ingham, J., & Fox, P. (2003). Hypophonia in parkinson's disease: Neural correlates of voice treatment revealed by pet. *Neurology*, *60*(3), 432–440.

Malmkjaer, K. (2009). *The routledge linguistics encyclopedia*. Routledge.

Moro-Velazquez, L., Gomez-Garcia, J. A., Arias-Londoño, J. D., Dehak, N., & Godino-Llorente, J. I. (2021). Advances in parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. *Biomedical Signal Processing and Control*, *66*, 102418.

Pell, M. D. (1996). On the receptive prosodic loss in parkinson's disease. *Cortex*, *32*(4), 693–704.

Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkmann, J., Schrag, A.-E., & Lang, A. E. (2017). Parkinson disease. *Nature reviews Disease primers*, *3*(1), 1–21.

Skodda, S., Grönheit, W., & Schlegel, U. (2011). Intonation and speech rate in parkinson's disease: General and dynamic aspects and responsiveness to levodopa admission. *Journal of Voice*, *25*(4), e199–e205.

Skodda, S., & Schlegel, U. (2008). Speech rate and rhythm in parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society*, *23*(7), 985–992.

Skodda, S., Visser, W., & Schlegel, U. (2011). Vowel articulation in parkinson's disease. *Journal of voice*, *25*(4), 467–472.

Sonkaya, Z. Z., Ceylan, M., & Sonkaya, A. R. (2021). Speech characteristics of parkinson disease. *Medical Science and Discovery*, *8*(12), 666–670.

Tracey, B., Volfson, D., Glass, J., Haulcy, R., Kostrzebski, M., Adams, J., Kangarloo, T., Brodtmann, A., Dorsey, E. R., & Vogel, A. (2023). Towards interpretable speech biomarkers: Exploring mfccs. *Scientific Reports*, *13*(1), 22787.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Yu, D., & Deng, L. (2016). *Automatic speech recognition* (Vol. 1). Springer.

# 9 Appendices

## 9.1 Italian Phrases

**The phonemically balanced text in Italian language:**
IL RAMARRO DELLA ZIA.
Il papà (o il babbo come dice il piccolo Dado) era sul letto. Sotto di lui, accanto al lago, sedeva Gigi, detto Ciccio, cocco della mamma e della nonna. Vicino ad un sasso c'era una rosa rosso vivo e lo sciocco, vedendola, la volle per la zia. La zia Lulù cercava zanzare per il suo ramarro, ma dato che era giugno (o luglio non so bene) non ne trovava. Trovò invece una rana che saltando dalla strada finì nel lago con un grande spruzzo. Sai che fifa, la zia! Lo schizzo bagnò il suo completo rosa che divenne giallo come un taxi. Passava di lì un signore cosmopolita di nome Sardanapalo Nabucodonosor che si innamorò della zia e la portò con sé in Afghanistan.

**The phonemically balanced phrases in Italian language:**

- Oggi è una bella giornata per sciare.
- Voglio una maglia di lana color ocra.
- Il motociclista attraversò una strada stretta di montagna.
- Patrizia ha pranzato a casa di Fabio.
- Questo è il tuo cappello?
- Dopo vieni a casa?
- La televisione funziona?
- Non posso aiutarti?
- Marco non è partito.
- Il medico non è impegnato.

**Words in Italian:**
pipa, buco, topo, dado, casa, gatto, filo, vaso, muro, neve, luna, rete, zero, scia, ciao, giro, sole, uomo, iuta, gnomo, glielo, pozzo, brodo, plagio, treno, classe, grigio, flotta, creta, drago, frate, spesa, stufa, scala, slitta, splende, strada, scrive, spruzzo, sgrido, sfregio, sdraio, sbrigo, prova, calendario, autobiografia, monotono, pericoloso, montagnoso, prestigioso.

## 9.2 Date on the PD group

**TABLE 5. Data on the PD group - third experimental phase.**

| | | | TEXT 1^ READING | | TEXT 2^ READING | | WORDS 1^ READING | |
|---|---|---|---|---|---|---|---|---|
| sex | UPDRS | age | time 1 | CPS1 | time 2 | CPS2 | time 3 | CPS3 |
| F | 0 | 63 | // | // | // | // | 60,64 | 4,63 |
| M | 2 | 50 | 71,73 | 7,22 | 53,82 | 9,62 | 39,75 | 7,07 |
| F | 1 | 61 | 53,40 | 9,70 | 51,4 | 10,08 | 49,25 | 5,71 |
| M | 2 | 68 | 84,05 | 6,16 | 63,32 | 8,18 | 56,66 | 4,96 |
| F | 0 | 40 | 60,92 | 7,76 | 52,4 | 9,89 | 54,58 | 5,15 |
| M | 1 | 65 | 52,40 | 9,89 | 50,23 | 10,31 | 40,2 | 6,99 |
| M | 1 | 73 | 79,35 | 6,53 | 71,22 | 7,27 | 69,62 | 4,04 |
| M | 0 | 56 | 86,81 | 5,97 | 66,64 | 7,77 | 58,7 | 4,79 |
| M | 1 | 77 | 64,75 | 8,00 | 60,9 | 8,51 | 50,41 | 5,57 |
| M | 0 | 71 | 59,84 | 8,66 | 56,76 | 9,13 | 53,52 | 5,25 |
| F | 1 | 71 | 66,22 | 7,82 | 48,38 | 10,71 | 55,56 | 5,06 |
| M | 0 | 71 | 49,95 | 10,37 | 45,35 | 11,42 | 36,86 | 7,62 |
| M | 1 | 73 | 70,98 | 7,30 | 65,07 | 7,96 | 56,87 | 4,94 |
| M | 2 | 75 | 62,20 | 8,33 | 56,58 | 9,16 | 52,98 | 5,30 |
| M | 2 | 68 | 85,37 | 6,07 | 63,35 | 8,18 | 48,67 | 5,77 |
| M | 0 | 71 | 62,33 | 8,31 | 52,56 | 9,86 | 66,38 | 4,23 |
| F | 2 | 65 | 242,50 | 2,14 | 180,09 | 2,88 | 167,83 | 1,67 |
| F | 1 | 80 | 169,29 | 3,06 | // | // | 101,19 | 2,78 |
| M | 0 | 73 | 66,90 | 7,74 | 63,5 | 8,16 | 53,04 | 5,30 |
| M | 0 | 70 | 65,46 | 7,91 | 60,4 | 8,58 | 48,25 | 5,82 |
| F | 1 | 67 | 79,30 | 6,53 | 73,8 | 7,02 | 71,67 | 3,92 |
| F | 0 | 54 | 55,00 | 9,42 | 49,8 | 10,40 | 54,7 | 5,14 |
| F | 3 | 78 | 163,60 | 3,17 | // | // | 108,3 | 2,59 |
| M | 1 | 72 | 117,60 | 4,40 | 98,8 | 5,24 | 87,51 | 3,21 |
| M | 4 | 65 | 164,10 | 3,16 | // | // | 151,3 | 1,86 |
| M | 4 | 65 | 233,00 | 1,76 | // | // | 217,3 | 0,96 |
| M | 0 | 70 | 112,00 | 4,63 | 106,6 | 4,86 | 68,31 | 4,11 |
| M | 0 | 70 | 68,30 | 7,58 | 61,47 | 8,43 | 64,5 | 4,36 |
| | | **average** | **94,35** | **6,65** | **67,50** | **8,42** | **73,02** | **4,60** |

Figure 6: The metrics from the PD participants

### 9.3 All speech features

Table 2: List of Speech Features

| **Feature** | **Description** |
|---|---|
| Zero Crossing Rate (ZCR) | Rate of sign changes in the signal |
| Energy | Total energy of the signal |
| Energy Entropy | Measure of energy distribution randomness |
| Spectral Centroid | Average frequency of the spectrum |
| Spectral Spread | Dispersion of spectral energy |
| Spectral Entropy | Measure of spectral energy distribution randomness |
| Spectral Flux | Rate of change of spectral energy |
| Spectral Rolloff | Frequency below which a certain percentage of spectral energy is concentrated |
| MFCC 1 to 13 | Mel-frequency cepstral coefficients |
| Chroma 1 to 12 | Chroma features representing energy distribution of pitch classes |
| Chroma Standard Deviation | Standard deviation of Chroma features |
| Delta ZCR | Delta feature for zero crossing rate |
| Delta Energy | Delta feature for energy |
| Delta Energy Entropy | Delta feature for energy entropy |
| Delta Spectral Centroid | Delta feature for spectral centroid |
| Delta Spectral Spread | Delta feature for spectral spread |
| Delta Spectral Entropy | Delta feature for spectral entropy |
| Delta Spectral Flux | Delta feature for spectral flux |
| Delta Spectral Rolloff | Delta feature for spectral rolloff |
| Delta MFCC 1 to 13 | Delta features for MFCCs |
| Delta Chroma 1 to 12 | Delta features for Chroma |
| Delta Chroma Standard Deviation | Delta feature for Chroma standard deviation |

## 9.4 Division of the features

| MFCC features | Delta features | Chroma features |
|---|---|---|
| mfcc_1 | delta zcr | chroma_1 |
| mfcc_2 | delta energy | chroma_2 |
| mfcc_3 | delta energy_entropy | chroma_3 |
| mfcc_4 | delta spectral_centroid | chroma_4 |
| mfcc_5 | delta spectral_spread | chroma_5 |
| mfcc_6 | delta spectral_entropy | chroma_6 |
| mfcc_7 | delta spectral_flux | chroma_7 |
| mfcc_8 | delta spectral_rolloff | chroma_8 |
| mfcc_9 | delta mfcc_1 | chroma_9 |
| mfcc_10 | delta mfcc_2 | chroma_10 |
| mfcc_11 | delta mfcc_3 | chroma_11 |
| mfcc_12 | delta mfcc_4 | chroma_12 |
| mfcc_13 | delta mfcc_5 | chroma_std |
| | delta mfcc_6 | |
| | delta mfcc_7 | |
| | delta mfcc_8 | |
| | delta mfcc_9 | |
| | delta mfcc_10 | |
| | delta mfcc_11 | |
| | delta mfcc_12 | |
| | delta mfcc_13 | |
| | delta chroma_1 | |
| | delta chroma_2 | |
| | delta chroma_3 | |
| | delta chroma_4 | |
| | delta chroma_5 | |
| | delta chroma_6 | |
| | delta chroma_7 | |
| | delta chroma_8 | |
| | delta chroma_9 | |
| | delta chroma_10 | |
| | delta chroma_11 | |
| | delta chroma_12 | |
| | delta chroma_std | |

Table 3: The features which are divided into their subsets. The red features are in the subset of delta MFCC features. The blue features are in the subset of Delta Chroma features.

## 9.5 Data and Source Code

The source data and code for reproducing the results are publicly available at: https://github.com/pcrooijendijk/asrthesis.