Preliminary Acoustic Analysis of Emotional Speech Production from New Zealand English Speakers with Parkinson's Disease

Itay Ben-Dom, Catherine I. Watson, Clare M. McCann

The University of Auckland

iben350@aucklanduni.ac.nz, c.watson@auckland.ac.nz , c.mccann@auckland.ac.nz

Abstract

This study explores the effects of Parkinson's disease (PD) on emotional speech production among native New Zealand English speakers, utilising a newly developed simulated emotional speech corpus. We engaged twelve participants diagnosed with PD to generate speech samples expressing five distinct emotions, yielding 1800 sentences. Our analysis focuses on key acoustic parameters, namely fundamental frequency and intensity, to evaluate their efficacy in depicting emotional states in PD-affected speech. Results indicate these parameters effectively differentiate between emotional expressions and correlate with emotional arousal levels. This provides insights into the complex interplay between physiological alterations due to PD and emotional vocal expression.

Index Terms: Emotional Speech Corpus, New Zealand English, Parkinson's disease, fundamental frequency, intensity.

1. Introduction

The human voice is a powerful conveyor of both biological and psychological information, capable of providing nuanced cues reflective of an individual's emotional state [1]. Acoustic and physiological cues in speech act as principal carriers of emotion, influencing sound production through alterations in muscle tension and respiration. These changes are key physiological indicators for affective vocalisation [2]. Emotional states profoundly impact voice production, necessitating modifications in respiration, phonation, and articulation, thereby requiring the quantification of these changes through various acoustic parameters to distinctly identify emotions [3].

Parkinson's disease significantly compromises the speech production mechanism through hypokinetic dysarthria, affecting the respiratory, phonatory, articulatory, and resonant systems. This leads to substantial challenges in production of emotional speech, where people with PD often exhibit reduced modulation of pitch and intensity, resulting in monotonous or hypophonic speech that negatively impacts their social communication and quality of life [4, 5].

F_0 and intensity are pivotal acoustic parameters for investigating vocal expressions [6]. F_0, the perceptual quality of pitch, is linked physiologically to vocal fold vibrations during speech production and varies across different emotional states. For example, higher F_0 values compared to the speakers' mean are associated with high-arousal emotions such as anger and fear, while lower values correspond to low-arousal emotions like sadness [7, 8, 9]. Intensity, or the loudness of the voice, also plays a critical role in emotional expression [4]. In PD, vocal intensity is often diminished due to weakened respiratory support and reduced control over vocal fold adduction, leading to softer speech that can further reduce speech intelligibility [10, 11]. These changes in intensity are critical for conveying emotional nuances and are influenced by muscular adjustments and respiratory effort, which are particularly compromised in PD. This results in a reduced dynamic range in vocal intensity, which can make emotional states more difficult to discern [12, 13, 14].

Despite advancements in understanding PD's effects on speech, there is a stark lack of research focusing on emotional speech production within this demographic, especially among native New Zealand English speakers. Existing speech corpora primarily feature other languages and are often limited to extracted acoustic features rather than comprehensive speech samples [15]. This gap underscores the necessity for a specialised corpus that addresses the unique needs of PD patients in New Zealand, facilitating detailed analysis of affective speech and its variations due to PD.

This study introduces the EMOPARKNZ corpus, developed to investigate the emotional speech production of New Zealand English speakers with PD. By analysing both fundamental frequency and intensity, this research aims to delineate the alterations in speech attributable to PD and provide insights into the complex interplay between physiological changes and vocal expression in affected individuals. This corpus represents a significant step toward understanding and improving communication aids for individuals with PD [15].

2. Methodology

2.1. Speakers

The corpus includes speech recordings from twelve individuals diagnosed with PD, balanced between six males and six females, all native speakers of New Zealand English. The age range for male participants is 56 to 74 years (mean 66.2 $\pm$ 6.2 years), and for female participants, 50 to 80 years (mean 64.3 $\pm$ 11.9 years). All participants were diagnosed by neurologists and recorded within 3 hours of medication intake to ensure they

Table 1. *Age, severity, and time after the PD diagnosis of male and female patients.*

M-PD			F-PD		
Age	**Severity**	**T**	**Age**	**Severity**	**T**
56	mild-mod	6	50	mild	4
62	mild-mod	8	53	mild	7
65	mild-mod	4	55	mild	6
67	mild-mod	20	74	mild-mod	3
73	mild	4	74	mild-mod	7
74	mild	14	80	mild-mod	10

M-PD: Male-PD, **F-PD**: Female-PD, **Severity**: mild/mild-moderate (mild-mod), and **T**: time post PD diagnosis (years).

were in the ON-state during speech sample collection, i.e. no more than 3 hours after taking medication [16]. Severity was determined by the researcher on the basis of overall impact of PD on daily functioning. Details regarding the age, severity, and years since onset diagnosis of each speaker are provided in Table 1.

2.2. Speech Materials

A speech corpus was developed to simulate strictly-guided emotions. Emotions were induced using a variation of the Stanislavski method [17], with specific s ituations a nd corresponding images presented to speakers to evoke and maintain the desired emotional state. This setup helped align the speakers' mental states with the targeted emotions throughout the recording session. Emotions were recorded in sequence from high to low arousal: *excited, angry, happy, neutral, sad*. To ensure consistency, the same leading phrases and images were provided to all speakers, who were instructed to maintain consistent emotional intensity. The corpus was designed with the JL corpus [18] in mind, a related NZ English speech data set developed for strictly-guided emotions. The sentences and prompts used in this study were adapted from those in the JL corpus. Fifteen different sentences, chosen for their neutrality across all primary emotions to prevent emotional bias, were included [19]. Short sentences of 4-7 syllables were selected to minimise potential emotional deviations within longer phrases. An example sentence is "Jack views an art piece." A complete list of all sentences used is shown in Table 2.

2.3. Recording process

Recordings for this study were approved by the University of Auckland Human Participants Ethics Committee (UAHPEC23603). The setup involved an omnidirectional condenser microphone (AKG, HC 577L) positioned 7 cm at a 45-degree angle to the right of the speaker's mouth. This microphone was connected to a Microsoft Surface Laptop 3 via a pre-amplifier (M-Audio MobilePre [MK II]), and audio capture was conducted at a sample rate of 44.1kHz with a 16-bit quantisation. The recording environment was prepared in a sound-treated room within the Speech Science laboratory at the University of Auckland, New Zealand. Participants were seated in a comfortable chair facing an iPad that displayed both the sentences and corresponding emotional images. This setup helped ensure participants were well-acquainted with the recording environment. Prior to recording, they were given five minutes to familiarise themselves with the settings and procedures, and they had the opportunity to ask any questions. The recording session commenced with a warm-up task where participants read the sentence "Tom beats that farmer" at a high intensity level. This initial task allowed for the adjustment of amplification gain levels to avoid clipping; these levels were set once and maintained throughout the session. Subsequently, the main recording task involved participants reading sentences displayed on the monitor, one at a time, each accompanied by an image corresponding to the intended emotion. The researcher monitored the recording in real-time to ensure technical quality. The sentence prompt was changed at the completion of each emotion expression. Each recording session lasted up to one hour, and the entire process was repeated to ensure two variations of each sentence for each emotion, resulting in a comprehensive corpus. In total, the study produced 1800 primary emotion sentences, encapsulating responses from 12 participants across 5 primary emotions, each repeated twice with 15 different sentences, all part of the EMOPARKNZ corpus.

Table 2. *Fifteen recorded sentences.*

No.	Sentence
1.	Tom beats that farmer.
2.	John laughs like your father.
3.	The seed is buried in deep.
4.	Taylor likes stewed Asian food.
5.	The lord swims in the sea.
6.	Jack views an art piece.
7.	Carl leaps into a jeep.
8.	Linda asks for more darts.
9.	Find your boot in this chute.
10.	Water harms the newborn boy.
11.	I have not seen my tooth.
12.	Work hard or you lose.
13.	Jim saw the port.
14.	They should start to talk.
15.	Sound the horn if you need more.

2.4. Data preparation

Each participant's acoustic recording was preserved in its entirety as a .wav file. Speech tasks were extracted from these files and stored individually to facilitate detailed analysis. The complete speech recordings were examined again for clippings and none were found, ensuring the quality of the data before further processing. The audio files were imported into Python (version 3.12.0) for initial processing, where leading and trailing silences were removed using the pydub package [20]. This refined acoustic signal was subsequently imported into MATLAB [21] for more advanced processing. Here, the fundamental frequency values of the speech samples were derived using the summation of residual harmonics method, implemented in the *covarep* package [22]. Additionally, intensity measurements (in dB) were extracted using the phonetic software Praat, employing its intensity function. Mean F_0 and intensity values were calculated for each sentence and aggregated for each speaker across all five primary emotions. After processing the signals and extracting the necessary acoustic features, a comprehensive statistical analysis was conducted. This analysis utilised the R language [23] and its environment for statistical computing and graphics (version 4.3.3), alongside the integrated environment R-Studio [24]. The data analysis aimed to statistically evaluate the relationships between extracted acoustic features and study variables, providing insights into the effects of Parkinson's disease on speech characteristics.

2.5. Statistical analysis

Our study employs a linear mixed-effects model, which is well-suited for data that include both fixed effects and random effects, allowing for the analysis of data with multiple sources of variability [25]. In this analysis, emotion (categorised as *excited, angry, happy, sad, neutral*) and gender (male/female) were treated as fixed effects. Speaker identity and the specific sentences they recited were considered as random effects, accommodating individual differences and sentence variability. The model formula used is represented in Equation 1 as follows:

$$\text{acoustic feature} \approx \text{emotion} * \text{gender} + (1|\text{subject}) + (1|\text{sentence}) \quad (1)$$

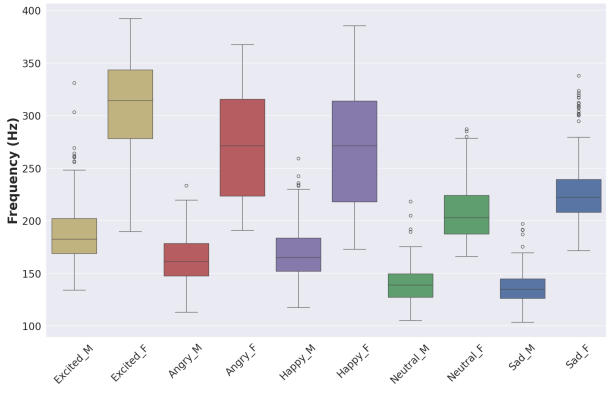

Figure 1: *Boxplot of F0 variation across five primary emotions for male (M) and female (F) speakers (n=12).*

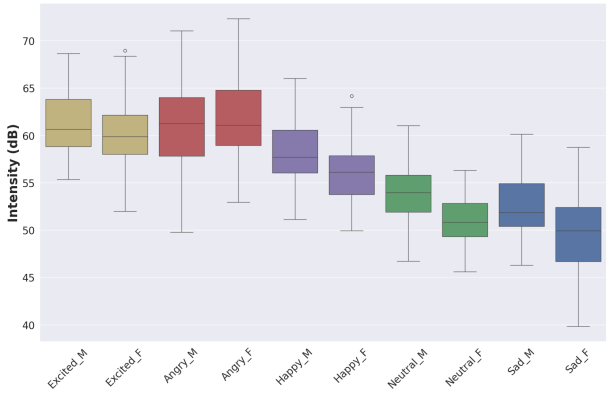

Figure 2: *Boxplot of intensity variation across five primary emotions for male (M) and female (F) speakers (n=12).*

The acoustic data was analysed using the linear mixed-effect model implemented in the R environment with the *lme4* package [26]. To refine the model and select the best fit, we used the step function from the *lmerTest* package [27], which simplifies the model selection process by comparing nested models. Subsequently, a likelihood ratio test was conducted to determine the statistical significance of the fixed effects, ensuring robustness in our findings. After determining the significant effects, post-hoc pairwise comparisons between levels of the fixed effects were performed. These comparisons were executed using the *emmeans* package [28], which adjusts *p*-values using the Tukey Honestly Significant Difference method to control for type I error across multiple tests. Effects were considered statistically significant if the *p*-values were less than 0.05, indicating robust differences in acoustic features across the groups defined by emotion and gender.

3. Results

3.1. Fundamental frequency

The analysis of F_0 across all twelve participants is depicted in Figure 1. Aggregated mean values and their standard deviations revealed the highest F_0 for the emotion *excited* (247.8 $\pm$ 65.2 Hz), followed in sequence by *happy* (219.8 $\pm$ 61.1 Hz), *angry* (216.1 $\pm$ 64.0 Hz), *sad* (182.8 $\pm$ 50.7 Hz), and *neutral* (173.7 $\pm$ 39.5 Hz). Statistical analyses utilising a linear mixed-effects model indicated significant interactions between most emotion pairs, except between *angry* and *happy*, detailed in Table 3. The linear mixed-effects model identified a statistically significant two-way interaction between emotion and gender, with a Chi-square value of $\chi^2_{(4)} = 1195$ and a *p*-value less than 0.0001. As expected, analysis revealed that females consistently exhibited higher F_0 values compared to males across all emotions studied. Specifically, for the emotion *excited*, females demonstrated an average F_0 of 306.9 Hz ($\pm$48.8 Hz), substantially higher than males who averaged 188.8 Hz ($\pm$31.4 Hz). Similarly, for *angry*, female speakers had an average F_0 of 270.6 Hz ($\pm$50.2 Hz), compared to 161.6 Hz ($\pm$24.2 Hz) in males. The significance continued with *happy* (females: 269.7$\pm$54.8 Hz, males: 169.9$\pm$26.2 Hz), *neutral* (females: 208.0$\pm$28.4 Hz, males: 139.4$\pm$19.0 Hz), and *sad* (females: 229.4$\pm$34.2 Hz, males: 136.1$\pm$16.8 Hz).

Table 3. *Statistical analysis comparison of emotion pairs.*

	Fundamental Frequency		Intensity	
Emotion Pairs	**t-ratio**	**p-value**	**t-ratio**	**p-value**
angry - excited	-15.413	**<.0001**	3.242	**0.0106**
angry - happy	-1.808	0.3692	22.139	**<.0001**
angry - neutral	20.604	**<.0001**	47.601	**<.0001**
angry - sad	16.195	**<.0001**	55.078	**<.0001**
excited - happy	13.605	**<.0001**	18.897	**<.0001**
excited - neutral	36.017	**<.0001**	44.360	**<.0001**
excited - sad	31.607	**<.0001**	51.836	**<.0001**
happy - neutral	22.412	**<.0001**	25.462	**<.0001**
happy - sad	18.002	**<.0001**	32.939	**<.0001**
neutral - sad	-4.410	**0.0001**	7.477	**<.0001**

3.2. Intensity

Corresponding analyses of acoustic intensity are illustrated in Figure 2. The mean intensity values, aggregated across all speakers, demonstrated the highest levels for the emotion *angry* (61.2 $\pm$ 3.6 dB), with descending values noted for *excited* (60.6 $\pm$ 2.7 dB), *happy* (57.1 $\pm$ 2.7 dB), *neutral* (52.4 $\pm$ 2.5 dB), and *sad* (51.0 $\pm$ 3.2 dB). The linear mixed-effects model revealed significant interactions across all pairs of emotions, detailed in Table 3. The linear mixed-effects model demonstrated a statistically significant two-way interaction between emotion and gender, with a Chi-square value of $\chi^2_{(4)} = 1195$ and a *p*-value less than 0.0001. Analyses revealed higher intensity values in males compared to females for all emotions, with the exception of *angry*. Specifically, for *excited*, intensity levels were higher in males (61.2 $\pm$ 3.4 dB) compared to females (60.0 $\pm$ 3.0 dB). Conversely, for *angry*, females exhibited a slightly higher intensity (61.6 $\pm$ 4.0 dB) than males (60.8 $\pm$ 4.4 dB). This trend persisted across other emotions, with males showing higher intensities for *happy* (58.3 $\pm$ 3.3 dB for males versus 56.0 $\pm$ 2.9 dB for females), *neutral* (53.9 $\pm$ 2.7 dB for males versus 50.9 $\pm$ 2.5 dB for females), and *sad* (52.6 $\pm$ 2.9 dB for males versus 49.4 $\pm$ 3.6 dB for females).

4. Discussion

Our acoustic analysis of the EMOPARKNZ corpus has demonstrated a systematic influence of arousal levels on fundamental frequency, confirming that higher arousal emotions such as

anger, *excitement*, and *happiness* are characterised by elevated pitch compared to *neutral* speech. The observed trend shows fundamental frequency serves as an arousal-differentiating feature, as previously reported [18]. Interestingly, *sad* speech exhibited a higher average pitch than *neutral*, which deviates from common findings and suggests variations in the definition of neutral across different studies [29]. This may indicate that the emotional state labelled as "neutral" in this context could have involved subtle emotional undertones not typically accounted for in standard classifications.

Our findings reveal that high arousal emotions consistently show elevated mean pitch, likely due to increased tension in the cricothyroid muscle and heightened sub-glottal pressure [30, 31], which are necessary for producing the vocal intensity associated with these emotions. In contrast, lower arousal emotions like sadness are associated with a reduced mean pitch, which can be attributed to stiffer vocal folds and limited vibratory capacity [29]. These physiological responses highlight the impact of PD on the vocal apparatus's ability to modulate speech according to emotional demands. The study also uncovered significant gender differences in vocal expression, with females consistently displaying higher mean pitch across all emotions. This gender disparity may be influenced by anatomical differences in the laryngeal size, which are exacerbated by PD, affecting how emotions are vocally expressed between males and females [32].

The majority of studies report elevated mean fundamental frequency in individuals with PD compared to healthy controls [13, 32]. While data from a control group is not currently available, it is possible to compare the F_0 values for *neutral* speech with the mean F_0 values reported in the literature. The mean F_0 values for male and female healthy speakers are approximately 120 Hz and 180 Hz, respectively [33]. The results for EMOPARKNZ indicate mean F_0 values for male and female speakers to be 140 Hz and 208 Hz, respectively. These findings corroborate previous reports of elevated mean fundamental frequency in individuals with PD. This increase in fundamental frequency can be attributed to laryngeal muscular impairment due to rigidity [34].

In terms of intensity, our analysis indicated that emotions typically associated with high arousal were distinguished by their higher mean intensity levels, aligning with the increased energy demands these emotions impose on speakers. The observed trend shows that like fundamental frequency, intensity serves as an arousal-differentiating feature. The significant variability in intensity for emotions like anger and sadness suggests that individual differences in the interpretation of these emotions can substantially affect their acoustic expression [35, 36]. For example, anger can vary from cold to hot anger, impacting intensity levels differently, while the unexpected variability in sadness among professional actors points to possible methodological nuances in how these emotions were elicited during recordings [37]. Furthermore, we found notable differences in mean intensity between male and female speakers, reinforcing the need for gender-specific approaches in clinical settings [10, 11]. These findings are crucial for developing more effective therapeutic strategies that accommodate the unique speech and voice characteristics of men and women with PD.

While intensity demonstrated greater discriminatory power compared to fundamental frequency, it presents potential issues due to the variable placement of the microphone relative to the speaker's mouth, especially across different datasets. In contrast, fundamental frequency can always be derived from speech samples, making it a robust and popular acoustic descriptor. This consistency in measurement makes fundamental frequency particularly advantageous for machine learning training, as it provides a reliable and stable feature across diverse datasets.

The limitations of the EMOPARKNZ corpus arise from the scarcity of participants and the inherent challenges of the data collection process. PD is a movement disorder, which makes recruiting participants difficult, especially due to the associated anxiety and stress. Anxiety, a common non-motor symptom of PD, often leads to the avoidance of everyday social situations out of fear of embarrassment caused by PD symptoms [38]. PD symptoms tend to be exacerbated by stress when required to perform new tasks, further discouraging participants from volunteering for studies. With regards to the data collection process, the EMOPARKNZ corpus includes speech samples elicited through induced emotions. This approach allowed participants to mentally prepare themselves to produce the targeted emotions and to repeat sentences if necessary. While this method facilitates data collection, it may not accurately reflect the subtleties of emotional expression in natural conversation. Capturing natural emotional speech is challenging, as such expressions are often more subtle and harder to detect. During the interviews that preceded each recording session, participants expressed frustration about their emotions not being perceived accurately in conversations, particularly with unfamiliar people. When asked to demonstrate the difficulties they face in conveying emotions, several participants mentioned that when producing *angry* speech, they need to "talk louder in their head" to make their voice sound louder. This suggests that while fundamental frequency and intensity have been shown to be strong acoustic descriptors of emotion, it will be interesting to examine their effectiveness in natural emotional speech.

Overall, the results from this study not only enhance our understanding of how PD influences emotional speech but also underscore the importance of considering both fundamental frequency and intensity as key parameters in assessing and treating speech disorders associated with neurological conditions. The insights gained here pave the way for future research to explore additional acoustic features and apply these findings in clinical practice, potentially through advanced speech therapy techniques and the development of supportive communication technologies tailored to the needs of people with PD.

5. Conclusion and Future Directions

This study introduced the EMOPARKNZ corpus, a new strictly-guided simulated emotional speech corpus in New Zealand English from speakers with PD, highlighting the critical roles of F_0 and intensity in distinguishing emotional expressions in PD-affected speech. Acoustic analysis demonstrated the variability of prosody parameters across emotions, affirming their efficacy in capturing emotional content despite vocal impairments associated with PD. Notably, participants retained some ability to modulate their voices according to emotional context, with significant differences observed between genders. Future research will build upon these findings by conducting a comprehensive acoustic analysis incorporating temporal, spectral features, and glottal-based features. The EMOPARKNZ corpus will also be used to train neural network models for speech emotion recognition tasks, contributing to the development of user-facing healthcare robots. Upon completion of the author's doctoral thesis, the EMOPARKNZ corpus will be made available, contributing further to the scientific community's efforts to support people with PD in maintaining effective communication and improving their social interactions.

6. References

[1] A. Karpf, *The human voice: The story of a remarkable talent*. London: Bloomsbury, 2006.

[2] A. Kappas, U. Hess, and K. R. Scherer, "Voice and emotion," in *Fundamentals of nonverbal behavior*, R. S. Feldman and B. Rime, Eds. Cambridge University Press, 1991, ch. 6, p. 200–238.

[3] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Communication*, vol. 40, no. 1, pp. 227–256, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167639302000845

[4] J. Möbes, G. Joppich, F. Stiebritz, R. Dengler, and C. Schröder, "Emotional speech in parkinson's disease," *Movement Disorders*, vol. 23, no. 6, pp. 824–829, 2008.

[5] M. Pell, H. Cheang, and C. Leonard, "The impact of parkinson's disease on vocal-prosodic communication from the perspective of listeners," *Brain and language*, vol. 97, pp. 123–34, 2006.

[6] I. R. Murray and J. L. Arnott, "Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech," *Computer Speech & Language*, vol. 22, no. 2, pp. 107–129, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0885230807000393

[7] R. Banse and K. R. Scherer, "Acoustic profiles in vocal emotion expression," *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 614–636, 1996.

[8] C. Breitenstein, D. V. Lancker, and I. Daum, "The contribution of speech rate and pitch variation to the perception of vocal emotions in a german and an american sample," *Cognition and Emotion*, vol. 15, no. 1, pp. 57–79, 2001.

[9] D. Guo, H. Yu, A. Hu, and Y. Ding, "Statistical analysis of acoustic characteristics of tibetan lhasa dialect speech emotion," *SHS Web of Conferences*, vol. 25, p. 01017, 2016.

[10] C. Dromey, L. O. Ramig, and A. B. Johnson, "Phonatory and articulatory changes associated with increased vocal intensity in parkinson disease: A case study," *Journal of Speech, Language, and Hearing Research*, vol. 38, no. 4, pp. 751–764, 1995.

[11] Cynthia M. Fox and Lorraine Olson Ramig, "Vocal sound pressure level and self-perception of speech and voice in men and women with idiopathic parkinson disease," *American Journal of Speech-Language Pathology*, vol. 6, no. 2, pp. 85–94, 1997.

[12] R. J. Holmes, J. M. Oates, D. J. Phyland, and A. J. Hughes, "Voice characteristics in the progression of parkinson's disease," *International Journal of Language & Communication Disorders*, vol. 35, no. 3, pp. 407–418, 2000.

[13] C. Dromey, "Spectral measures and perceptual ratings of hypokinetic dysarthria," *Journal of Medical Speech-Language Pathology*, vol. 11, pp. 85–94, 06 2003.

[14] J. E. Huber and M. Darling, "Effect of parkinson's disease on the production of structured and unstructured speaking tasks: Respiratory physiologic and linguistic considerations," *Journal of Speech, Language, and Hearing Research*, vol. 54, no. 1, pp. 33–46, 2011.

[15] C. Williams and K. Stevens, "Emotions and speech: some acoustical correlates." *The Journal of the Acoustical Society of America*, vol. 52 4, pp. 1238–50, 1972.

[16] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rátiva, and E. Nöth, "New spanish speech corpus database for the analysis of people suffering from parkinson's disease," in *International Conference on Language Resources and Evaluation*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:1735319

[17] J. Benedetti, *Stanislavski: An Introduction, Revised and Updated*, 2nd ed. Routledge, 2004. [Online]. Available: https://doi.org/10.4324/9780203998182

[18] J. James, L. Tian, and C. Watson, "An open source emotional speech corpus for human robot interaction applications," in *Interspeech 2018, 19th Annual Conference of the International Speech Communication Association*, B. Yegnanarayana, Ed. Hyderabad: ISCA, September 2018, pp. 2768–2772.

[19] R. Cowie, E. Douglas-Cowie, and C. Cox, "Beyond emotion archetypes: Databases for emotion modelling using neural networks," *Neural Networks*, vol. 18, no. 4, pp. 371–388, 2005, emotion and Brain. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608005000353

[20] J. Robert, M. Webbie *et al.*, "Pydub," http://pydub.com/, 2018.

[21] T. M. Inc., "Matlab version: 9.13.0 (r2022b)," Natick, Massachusetts, United States, 2022. [Online]. Available: https://www.mathworks.com

[22] T. Drugman, P. Alku, A. Alwan, and B. Yegnanarayana, "Glottal source processing: From analysis to applications," *Computer Speech & Language*, vol. 28, no. 5, pp. 1117–1138, 2014.

[23] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[24] RStudio Team, *RStudio: Integrated Development Environment for R*, RStudio, PBC., Boston, MA, 2020. [Online]. Available: http://www.rstudio.com/

[25] B. Winter, "Linear models and linear mixed effects models in r with linguistic applications," 2013.

[26] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, p. 1–48, 2015. [Online]. Available: https://www.jstatsoft.org/index.php/jss/article/view/v067i01

[27] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmertest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, p. 1–26, 2017. [Online]. Available: https://www.jstatsoft.org/index.php/jss/article/view/v082i13

[28] R. V. Lenth, P.-C. Buerkner, M. Herve, J. Love, F. Miguez, H. Riebl, and H. Singmann, "emmeans: Estimated marginal means, aka least-squares means," https://CRAN.R-project.org/package=emmeans, 2022.

[29] J. James, "Modeling prosodic features for empathetic speech of a healthcare robot," Ph.D. dissertation, University of Auckland, 2021.

[30] A.-M. Laukkanen, E. Vilkman, P. Alku, and H. Oksanen, "Physical variations related to stress and emotional state: a preliminary study," *Journal of Phonetics*, vol. 24, no. 3, pp. 313–335, 1996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0095447096900170

[31] Y. Li, J. Li, and M. Akagi, "Contributions of the glottal source and vocal tract cues to emotional vowel perception in the valence-arousal space," *The Journal of the Acoustical Society of America*, vol. 144, no. 2, pp. 908–916, 2018.

[32] S. Skodda, H. Rinsche, and U. Schlegel, "Progression of dysprosody in parkinson's disease over time — a longitudinal study," *Movement Disorders*, vol. 24, no. 5, pp. 716–722, 2009.

[33] J. Clark, C. Yallop, and J. Fletcher, *An introduction to phonetics and phonology*, 3rd ed. Malden, Mass: Blackwell Pub, 2007, vol. Series: Blackwell textbooks in linguistics.

[34] I. Midi, M. Dogan, M. Koseoglu, G. Can, M. A. Sehitoglu, and D. I. Gunal, "Voice abnormalities and their relation with motor dysfunction in parkinson's disease," *Acta Neurologica Scandinavica*, vol. 117, no. 1, pp. 26–34, 2008.

[35] T. Johnstone and K. Scherer, "The effects of emotions on voice quality," *Proceedings of the XIVth International Congress of Phonetic Sciences*, p. 5, 1999.

[36] K. Scherer, "Vocal affect expression: a review and a model for future research," *Psychological bulletin*, vol. 99 2, pp. 143–65, 1986.

[37] J. Wilting, E. J. Krahmer, and M. Swerts, "Real vs. acted emotional speech," in *Interspeech*, 2006. [Online]. Available: https://api.semanticscholar.org/CorpusID:14024275

[38] B. K. Gultekin, B. Ozdilek, and E. E. Bestepe, "Social phobia in parkinson's disease: Prevalence and risk factors," *Neuropsychiatric Disease and Treatment*, vol. 10, pp. 829–834, May 21 2014.