



Physiologically-Informed Feature Analysis of Acquired Speech Disorders for Stroke Assessment

Giulia Sanguedolce^{1,2,3}, Jón Guðnason⁵, Dragos C. Gruia³, Emilie d’Olne², Fatemeh Geranmayeh^{3,4}, Patrick A. Naylor²

¹Department of Computing, Imperial College London, UK

²Department of Electrical and Electronic Engineering, Imperial College London, UK

³Department of Brain Sciences, Imperial College London, UK

⁴Imperial College Healthcare NHS Trust, London, UK

⁵Department of Engineering, Reykjavik University, Iceland

gs2022@ic.ac.uk, jgudnason@ru.is, dragos.gruia@imperial.ac.uk, e.dolne@imperial.ac.uk, f.geranmayeh@imperial.ac.uk, p.naylor@imperial.ac.uk

Abstract

Post-stroke speech disorders impair communication and rehabilitation outcomes, often requiring prolonged, intensive therapy sessions. The diversity of symptoms, coupled with the high cost and logistical burden of traditional speech therapy, underscores the need for accurate, automatic assessment to support tailored interventions. Leveraging SONIVA, our purpose-built database of stroke patients’ speech, this study introduces a feature-driven framework integrating traditional acoustic features with physiologically informed glottal parameters for classifying impaired speech after stroke. Evaluating unimodal, combined, and SHAP-derived (SHapley Additive exPlanations) feature configurations, our approach achieved a 97% F1-score in distinguishing pathological from healthy speech. These results highlight the potential of combining clinically meaningful glottal and acoustic information to support early speech deterioration detection, enhancing accessibility and personalised rehabilitation strategies for improved patient outcomes.

Index Terms: Speech disorders, Machine Learning, Stroke Assessment, Interpretability

1. Introduction

The global burden of stroke affects approximately 12 million people annually [1], with 35% of survivors developing speech and language impairments [2]. Such speech disorders vary greatly in nature, depending on the location and extent of the lesioned brain tissue. Furthermore, there is heterogeneity in impairments across individuals, as well as within the same patient over time, making standardised speech assessment highly challenging. The nature of these impairments includes aphasia, apraxia, dysarthria, and dysphonia, which can co-occur to different extents [3]. Aphasia primarily affects the cognitive aspects of speech and language, leading to difficulties in understanding and forming coherent sentences [4]. Dysarthria results from impaired muscle control of the lips, tongue, and neck, impacting articulation, pace, and rhythm, often causing slurred speech [2]. Apraxia is a motor planning deficit that can cause slow, halting, and distorted speech, while dysphonia affects specifically the larynx activity, leading to abnormal changes in pitch and loudness [5]. People with Stroke (PwS) may also experience various degrees of orofacial and body paralysis. Compensation by overusing the unaffected side, respiratory weakness, and tremors can result in abnormal speech patterns, in-

cluding reduced volume, a strained or hoarse voice, limited speech to short phrases, and gasping between utterances [6]. This heterogeneity in speech and language impairments necessitates personalised intensive speech therapy with frequent sessions. However, delivering such effective speech therapy, along with accurate diagnosis and ongoing monitoring, is both logistically and financially burdensome for healthcare providers and patients [7, 8]. Automated day-to-day analysis may offer valuable insights for the evaluation of post-stroke speech in an objective, efficient, potentially remote and non-invasive manner. Such analyses have traditionally been prioritising acoustic features like Mel-frequency cepstral coefficients and Linear Predictive parameters [9]. This focus is partly due to the difficulty of accurately detecting more laryngeal related information in patients with abnormal voice characteristics, given that the glottal opening and closing events are brief and easily masked by vocal tract effects. Yet, recent advancements such as the YAGA algorithm [10], have begun to address these challenges, providing more reliable methods for glottal activity detection. Indeed, the utility of glottal analysis with these algorithms has been demonstrated to successfully assessing numerous respiratory tract diseases involving larynx conditions [11, 12], Parkinson’s [13, 14], or dysarthria caused by other neurological conditions [15, 16]. However, this approach has not been explored yet in post-stroke, largely due to the scarcity of speech data from PwS.

Given the promising performance achieved using acoustic features alone [17, 18, 19, 20], particularly with toolkits like openSMILE [21], a combined approach integrating acoustic and glottal source analyses presents significant potential for objective and efficient speech evaluation. Acoustic features capture the overall characteristics of the speech signal, reflecting resonances and articulatory movements. In contrast, glottal features provide detailed insights into laryngeal function, independent of vocal tract resonances and articulatory movements, which are subject to voluntary control [22].

To leverage these complementary aspects of speech analysis, we used SONIVA, below described, our purpose-built database of stroke-affected speech and extracted features from YAGA and openSMILE. The contributions of this work are: (a) introducing a new large-scale PwS database; (b) demonstrating that glottal parameters are effective in classifying neurologically impaired patients vs. healthy controls, highlighting their potential for automated screening and neurological monitoring;

and (c) providing clinical interpretability of our model’s output using SHAP (SHapley Additive exPlanations) [23], which reveals the most influential parameters for the classification. To the best of our knowledge, this study pioneers the use of glottal features for stroke assessment. By automating the process of speech-disorders analysis, our approach potentially offers a rapid, objective alternative to time-intensive clinician evaluations, ultimately improving patient outcomes.

2. Methods

2.1. Database

The Imperial Comprehensive Cognitive Assessment in Cerebrovascular Disease (IC3) [24, 25] is a longitudinal research study conducted in the UK, designed to investigate post-stroke cognitive and language recovery. It is part of SONIVA (Speech recOgNition Validation in Aphasia) speech database, which currently holds data from ≈ 1000 stroke survivors and ≈ 7000 healthy controls. IC3 speech recordings were collected as part of a digital cognitive battery and include both spontaneous and structured speech tasks. In particular, picture description tasks, featuring both custom-designed stimuli and standard images from the Comprehensive Aphasia Test (CAT; [26]), were used to elicit connected speech suitable for acoustic and linguistic profiling. For this study, we analysed speech data from 388 audio recordings of 125 unique patients which undertook the same picture description tasks, including some with longitudinal data spanning their recovery trajectories. The total duration of the PwS data was 5 hours, while the control group sample, consisting of 125 age-matched healthy individuals ($\mu = 61.51$ years, $\sigma = 10.55$ years), had a duration of 3 hours. All speech recordings were sampled at 16 kHz with a 16-bit resolution.

2.2. Feature Extraction

For the extraction of glottal parameters, PEFAC algorithm [27] was first employed to retrieve voiced segments where phonation occurred. We then applied the YAGA algorithm [10], which uses an N -best dynamic programming technique to identify Glottal Closure Instant (GCI) and Glottal Opening Instant (GOI) in speech signals. YAGA was selected for its demonstrated reliability and accuracy in GCI detection for clean speech [28], and the parameter chosen are based on prior literature [11, 12, 13, 14, 15, 16]. These parameters can be categorised into time-domain and frequency-domain features. In the time domain, key parameters include the *Opening Quotient* (OQ) and *Closing Quotient* (CQ). The OQ measures the proportion of the glottal cycle during which the glottis is open, defined as the ratio between the fall time (the time between a glottal opening instance, GOI, and the subsequent glottal closing instance, GCI) and the cycle duration (the interval between two consecutive GOIs) [29]. The CQ complements the OQ and is computed as: $CQ = 1 - OQ$. *Speed Quotient* (SQ) captures the asymmetry of the glottal pulse by taking the ratio between the *Rise Time* (from a GCI to the following GOI) and the *Fall Time* [29]. The *Amplitude Quotient* (AQ) and its normalised variant (NAQ) describe the glottal flow’s closing phase. AQ is defined as:

$$AQ = \frac{f_{ac}}{d_{peak}} \quad (1)$$

where f_{ac} is the maximum amplitude of the glottal flow and d_{peak} represents the negative peak amplitude of its derivative during closure. The NAQ is obtained by normalising AQ by the fundamental period T . Variations in these measures may

indicate irregularities including impaired vocal fold closure, affecting voice quality and vocal efficiency.

In the frequency domain, the *Harmonic Richness Factor* (HRF) serves as a vital metric of harmonic content. It quantifies the ratio between the energy of higher harmonics and the fundamental frequency F_0 calculated at each identified GCI, given by:

$$HRF = \frac{\sum_{k=2}^N X(kF_0)}{X(F_0)}, \quad (2)$$

where $X(F_0)$ is the amplitude at the fundamental frequency and $X(kF_0)$ denotes the amplitude of the k -th harmonic [30]. Higher HRF values correspond to richer harmonic content, typically associated with healthy phonation. The H1H2 and H2H4 parameters represent amplitude differences between successive harmonics (H1-H2 and H2-H4, respectively) [31]. These differences reflect the balance of vocal fold vibrations, where lower ratios generally signify more stable and healthy speech production. The *Parabolic Spectrum Parameter* (PSP), as defined by [32], measures how closely the power spectrum near each GCI follows a parabolic shape. PSP is calculated as the ratio of $a_{optimal}$, the coefficient from fitting a parabolic curve to the signal’s power spectrum, to a_{max} , obtained from a DC-level approximation. This ratio provides insight into the spectral smoothness and energy distribution. Finally, the *Peak Slope* (PS), extracted using the COVAREP toolkit [33], quantifies the sharpness of glottal closure by measuring the steepness of the glottal pulse. This feature is associated with the degree of vocal effort and closure abruptness. For each of these 10 features, descriptive statistics—mean, standard deviation, minimum, maximum, kurtosis, median, skewness, and range—were computed across glottal frames of each sample, for a total of 80 parameters.

Acoustic features were extracted using the openSMILE Python library (v. 3.0.1), based on the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPSv02; [21]) and added to the glottal parameter set for the classification models. This eGeMAPSv02 set includes 88 parameters, chosen for their proven relevance in speech analysis, and widely validated for evaluating pathological speech. The acoustic feature set encompasses prosodic measures (e.g., F_0 stability, jitter, shimmer) to assess pitch and phonatory control, spectral features (e.g., MFCCs, flux) for vocal resonance and articulatory precision, and energy-related metrics (e.g., loudness, harmonics-to-noise ratio, pauses) to evaluate fluency and rhythmic disturbances.

2.3. Model Architecture

The dataset was split into training and testing sets using a 90%-10% ratio, ensuring that all samples from the same patient were confined to a single partition through a stratified group split, keeping the test portion completely unseen. This step was essential due to the presence of multiple recordings from the same patient, collected to assess intra-subject variability over time. To evaluate the contribution of different modalities, classification was conducted on acoustic features only, glottal features only, and then their combination. Features were then scaled using `MinMaxScaler`, applied separately for each modality to maintain consistent feature distributions while enabling stable neural network training.

On the 90% split, a five-fold `StratifiedGroupKFold` cross-validation was used to ensure patient-level consistency across splits while maintaining balanced class distributions. Given the class imbalance between healthy and patient samples, SMOTE [34] was employed within each fold to the mi-

Table 1: Classification results across different feature modalities, including the SHAP-selected model.

Modality	Validation (5-Fold CV)		Unseen Set	
	F1-score	Accuracy	F1-score	Accuracy
Acoustic	0.920	0.890	0.930	0.898
Glottal	0.910	0.880	0.923	0.898
Combined	0.941	0.910	0.941	0.918
SHAP-Derived Set	0.886	0.832	0.969	0.959

nority class instances by interpolating nearest neighbours, thus mitigating bias toward the majority class. The model architecture consisted of a feed-forward neural network with three fully connected layers of decreasing size (128, 64, and 32 units). Each layer was followed by batch normalisation and a 0.3 dropout rate to prevent overfitting and stabilise training, with LeakyReLU activation functions used to mitigate vanishing gradient issues. The final output layer comprised a single sigmoid unit. The model was trained using binary cross-entropy loss with Adam optimiser, configured with a learning rate of 0.001 and weight decay of 1×10^{-5} for regularisation. Performance was evaluated using accuracy and F1-score, with the latter prioritised due to its robustness under class imbalance. The final performance was reported both from the validation sets and from the held-out 10% test set, providing an unbiased measure of the model’s generalisation capability.

Finally, SHAP was applied to the neural network classifier on the combined model, using an independent masker to approximate conditional expectations. These features were then selected to train an additional model reported in Table 1. SHAP values were computed by considering all possible feature subsets and quantifying the marginal contribution of each feature, ensuring consistency and local accuracy. The explainer generated SHAP values for each feature, enabling the identification of the most influential parameters in the classification.

3. Results

The classification results across the four evaluated configurations—acoustic, glottal, combined, and the SHAP-derived subset—demonstrate the superior performance achieved by integrating acoustic and glottal features (Tab. 1). Specifically, the combined modality reached an accuracy of 0.918 and an F1-score of 0.941 on the unseen test set, outperforming the acoustic-only (accuracy: 0.898, F1-score: 0.930) and glottal-only (accuracy: 0.898, F1-score: 0.923) configurations. A SHAP-derived model was then created using only the top 15 features identified by SHAP analysis. Despite the reduced feature set, this clinically interpretable model achieved an accuracy of 0.959 and an F1-score of 0.969 on the unseen test set. These results not only surpass the performance of the combined model but also demonstrate that a streamlined subset of clinically meaningful features can yield robust classification performance while maintaining interpretability, offering potential for practical clinical deployment.

The SHAP analysis (Fig. 1) further highlights the underlying model behaviour by ranking feature contributions to classification decisions (low in blue and high in red). Positive SHAP values (towards the right) increase the predicted probability of *patient* classification, while negative values (towards the left) push predictions towards *healthy*. This distribution

indicates how each feature influences the probability rather than indicating a strict classification boundary. The most influential feature predominantly originated from the acoustic set was *mfcc4V_sma3nz_amean*, indicating that spectral and temporal dynamics play a critical role in patient classification. However, glottal features including *SQ_median*, *NAQ_std*, *OQ_min*, and *CQ_kurtosis* also exhibited substantial contributions, suggesting that glottal flow characteristics related to vocal fold closure and amplitude variability significantly affect the classification boundary. Additionally, features such as *F0st_27.5Hz_sma3nz_meanFallingSlope* and *H2H4_skewness* provided discriminative power, emphasising the relevance of frequency modulation and harmonic structure in the decision process. The polarity of SHAP values revealed that higher values in glottal irregularity features (e.g., *NAQ_std*) and reduced spectral richness are associated with patient classification, whereas consistent acoustic patterns were linked to healthy predictions. Overall, the combined modality’s improved performance can be attributed to the integration of acoustic features (6 out of 15 top features) that capture the speech signal’s spectral properties and glottal features (9 out of 15 top features) that encapsulate physiological vocal fold behaviour. The SHAP interpretation not only validates the complementary effect of these modalities but also identifies critical parameters that may be indicative of underlying pathophysiological mechanisms in disordered speech.

4. Discussion

The SHAP analysis revealed a complementary set of acoustic and glottal features that not only effectively distinguish between healthy controls and stroke patients but also provide deeper insights into how stroke-induced neurological disruptions impact speech production. The most influential feature, *mfcc4V_sma3nz_amean* (Mel-frequency cepstral coefficient), reflects the spectral envelope of speech, capturing subtle variations in vocal tract resonances and articulatory movements. Clinically, its prominence underscores how stroke impairs articulatory coordination and resonance characteristics, often resulting in reduced clarity and distorted speech. The high SHAP attribution to this feature aligns with previous findings where spectral degradation signified compromised neuromotor control in post-stroke speech [35].

Among glottal features, several parameters indicative of laryngeal biomechanics emerged as critical discriminators. The Speed Quotient (*SQ_median*), quantifying the ratio between the glottal opening and closing phases, highlights the importance of vocal fold vibratory asymmetry. Neurological impairments following stroke can disrupt this balance, leading to inefficient phonation and breathy voice qualities. Similarly, the Opening Quotient (*OQ_min*)—representing the minimal open phase duration—offers insights into laryngeal muscle tonicity. Reduced OQ values may indicate hyperadduction of vocal folds, a condition clinically associated with strained phonation or spastic dysphonia [36].

The Normalised Amplitude Quotient variability (*NAQ_std*) was another high-impact feature, reflecting the stability of glottal flow amplitude across speech cycles. From a clinical perspective, increased variability in NAQ suggests impaired neuromuscular control of the vocal folds, often manifesting as voice instability or pitch breaks. Such measures are essential for clinicians, offering quantifiable proxies for voice stability—critical in the assessment of dysphonia severity post-stroke [37].

In the frequency domain, features such as

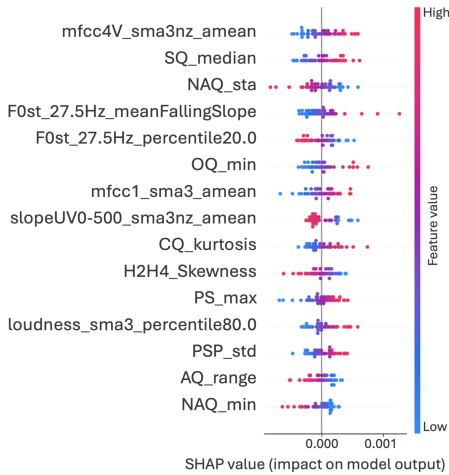


Figure 1: SHAP summary plot showing the top 15 features impacting the model's output probability.

F0st_27.5Hz_sma3nz_meanFallingSlope and *H2H4_skewness* revealed strong discriminative power. The first, the falling slope parameter, reflects pitch modulation control, critical for prosody and intonation patterns in speech. Stroke patients often exhibit monotonic speech due to impaired pitch control, a symptom captured by this feature. *H2H4_skewness*, on the other hand, quantifies harmonic distribution, providing insights into glottal source spectral tilt, which affects voice quality attributes like breathiness and harshness. Deviations in this parameter may indicate incomplete vocal fold closure or altered subglottal pressure regulation. These patterns are consistent with the roughness, breathiness, and instability observed in PwS, due to the neural damage for paralysis conditions [6].

Further glottal parameters such as Closing Quotient kurtosis (*CQ_kurtosis*) and PSP standard deviation (*PSP_std*) contribute critical insights into glottal closure regularity, parameters sensible for pathological speech discrimination also in Parkinson's Disease [14]. *CQ_kurtosis* measures the consistency of vocal fold closure patterns—higher kurtosis values suggest irregular closures, potentially linked to spastic or ataxic dysarthria. *PSP_std* reflects the abruptness of glottal closure, with greater variability indicating compromised vocal effort regulation. Lastly, the inclusion of loudness-related features (i.e., *loudness_sma3_percentile80.0*) within the top SHAP-ranked parameters emphasises the role of quantifications of vocal projection capabilities, characteristics of respiratory-laryngeal coordination. Indeed, stroke-induced respiratory weakness often results in reduced voice intensity and vocal fatigue [38].

Regarding the classification, the better performance of the combined acoustic-glottal model over single-modality approaches demonstrates the value of comprehensive voice assessment in stroke patients. The SHAP-derived subset model yielded even better performance with just 15 features, suggesting that focused assessment of specific voice parameters can provide highly accurate diagnostic information. This result is expected to be attributed to the reduction of noise and redundancy present in the larger feature set. Indeed, the full feature set, while comprehensive, may introduce complexity that potentially masks the most discriminative characteristics of impaired speech.

The classification accuracy observed across our models aligns with and surpasses previous findings in speech disorder

research. For instance, analyses employing openSMILE to extract acoustic features differentiating dysarthric from healthy speech based on single-word utterances [39] have reported accuracies of 96.18% and 93.24% on the TORGO [40] and UA-Speech[41] datasets, respectively. Similar to our approach, a study on dysarthric conditions combined openSMILE and glottal features reaching a classification accuracies of 91.88% (UA-Speech) and 82.12% (TORGO) [42]. However, these studies focus solely on phonation-centric conditions, where acoustic features are the primary discriminators. In contrast, stroke-induced speech disorders involve more complex neurological disruptions affecting motor control, laryngeal function, and language abilities. To our knowledge, this is the first study that classify speech with glottal features in post-stroke patients, making direct performance comparisons challenging. Thus, the strong performance of our model is particularly significant, as it not only addresses the complexity of these conditions but also establishes an important baseline for future research.

5. Limitations and Future Work

Several limitations should be considered when interpreting these results. First, only simple feed-forward neural network has been used. Alternative architectures could also be explored in next works, testing different fusion strategies beyond simple concatenation. Additionally, while our IC3 test set comprised unseen data, it may not fully represent the diversity of stroke-related speech impairments in real-world clinical settings, so the scores obtained should always be interpreted with caution. This limitation is particularly relevant given the heterogeneous nature of stroke-induced disorders and varying recording conditions across healthcare facilities. Therefore, further validation using both additional data from our ongoing collection efforts and external open-access pathological speech data would be beneficial to establish the generalisability of our approach. Such broader testing would help verify the robustness of our selected feature set across different patient populations, recording conditions, and clinical settings, ultimately supporting its transition into real-world clinical applications.

6. Conclusion

The findings of this study have significant implications for clinical practice, as they demonstrate that a streamlined set of acoustic features (capturing overall voice output) and glottal parameters (reflecting underlying physiological function) provides both surface-level and deeper physiological insights into voice production deficits. By combining these modalities, the classification model of the SHAP-selected features achieved an F1-score of 0.969, surpassing the unimodal acoustic (0.930) and glottal (0.923) models by 3.9 and 4.6 percentage points, respectively. This comprehensive approach may contribute to improving clinical practice by supporting more accessible, targeted, and cost-effective rehabilitation. By facilitating earlier detection and more precise monitoring through the use of IC3, it could help streamline clinical workflows and inform personalised recovery strategies, with the potential to improve patient outcomes and reduce strain on healthcare systems.

7. Acknowledgment

The authors would like to thank A. Coghlan, O. Burton, S. Brooks and N. Parkinson for their IC3 study assistance. Infrastructure provided by the NIHR Imperial Biomedical Re-

search Centre and the NIHR Imperial Clinical Research Facility. Funding: G.S. is UK Research and Innovation [UKRI Centre for Doctoral Training in AI for Healthcare:EP/S023283/1] and for F.G. Medical Research Council P79100; PSP415 – EP-SRC IAA; PSP518 – MRC IAA.

8. References

- [1] V. Feigin et al., “World stroke organization (WSO): global stroke fact sheet 2022,” *Int. J. of Stroke*, vol. 17, no. 1, pp. 18–29, 2022.
- [2] Z. Ghoreyshi, R. Nilipour, N. Bayat, S. S. Nejad, M. Mehrpour, and T. Azimi, “The incidence of aphasia, cognitive deficits, apraxia, dysarthria, and dysphagia in acute post stroke persian speaking adults,” *Indian J. of Otolaryngology and Head & Neck Surgery*, vol. 74, no. Suppl 3, pp. 5685–5695, 2022.
- [3] F. Geranmayeh, R. J. Wise, A. Mehta, and R. Leech, “Overlapping networks engaged during spoken language production and its cognitive control,” *Journal of Neuroscience*, vol. 34, no. 26, pp. 8728–8740, 2014.
- [4] G. Sanguedolce, P. A. Naylor, and F. Geranmayeh, “Uncovering the potential for a weakly supervised end-to-end model in recognising speech from patient with post-stroke aphasia,” in *Proc. of the 5th Clinical NLP Workshop*, 2023, pp. 182–190.
- [5] G. Dyukova et al., “Speech disorders in right-hemisphere stroke,” *Neuroscience and Behav. Physiol.*, vol. 40, pp. 593–602, 2010.
- [6] E. Charters et al., “Oral incompetence: Changes in speech intelligibility following facial nerve paralysis,” *J. of Plastic, Reconstructive & Aesthetic Surgery*, vol. 87, pp. 472–478, 2023.
- [7] G. Sanguedolce, S. Brook, D. C. Gruia, P. A. Naylor, and F. Geranmayeh, “When whisper listens to aphasia: Advancing robust post-stroke speech recognition,” in *Interspeech*, 2024.
- [8] G. Sanguedolce, D.-C. Gruia, S. Brook, P. Naylor, and F. Geranmayeh, “Universal speech disorder recognition: towards a foundation model for cross-pathology generalisation,” in *Advancements In Medical Foundation Models: Explainability, Robustness, Security, and Beyond*, 2024.
- [9] J. Sueur and J. Sueur, “Mel-frequency cepstral and linear predictive coefficients,” *Sound Analysis and Synthesis with R*, pp. 381–398, 2018.
- [10] M. R. Thomas, J. Gudnason, and P. A. Naylor, “Estimation of glottal closing and opening instants in voiced speech using the yaga algorithm,” *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 20, no. 1, pp. 82–91, 2011.
- [11] M. Kohler, M. M. Vellasco, E. Cataldo et al., “Analysis and classification of voice pathologies using glottal signal parameters,” *J. of Voice*, vol. 30, no. 5, pp. 549–556, 2016.
- [12] S. R. Kadiri and P. Alku, “Analysis and detection of pathological voice using glottal source features,” *IEEE J. of Selected Topics in Signal Process.*, vol. 14, no. 2, pp. 367–379, 2019.
- [13] E. A. Belalcázar-Bolanos et al., “Glottal flow patterns analyses for parkinson’s disease detection: Acoustic and nonlinear approaches,” in *Int. Conf. on Text, Speech, and Dialogue*. Springer, 2016, pp. 400–407.
- [14] P. Corcoran et al., “Glottal flow analysis in parkinsonian speech,” in *Proc. of the 12th Int. Joint Conf. on Biomedical Eng. Syst. and Technol. (BIOSTEC)*, 2019, pp. 116–123.
- [15] N. Narendra and P. Alku, “Dysarthric speech classification using glottal features computed from non-words, words and sentences,” in *Interspeech*. Int. Speech Commun. Assoc. (ISCA), 2018, pp. 3403–3407.
- [16] —, “Dysarthric speech classification from coded telephone speech using glottal features,” *Speech Commun.*, vol. 110, pp. 47–55, 2019.
- [17] M. Shahin et al., “Automatic screening of children with speech sound disorders using paralinguistic features,” in *2019 IEEE 29th International MLSP workshop*. IEEE, 2019, pp. 1–5.
- [18] P. Barche, K. Gurugubelli, and A. K. Vuppala, “Towards automatic assessment of voice disorders: A clinical approach,” in *INTERSPEECH*, 2020, pp. 2537–2541.
- [19] D. Kumar, U. Satija, and P. Kumar, “Pathological speech and electroglottography signals analysis using invariance scattering network,” *Circuits, Systems, and Signal Processing*, pp. 1–18, 2024.
- [20] G. Sanguedolce, D.-C. Gruia, P. Naylor, and F. Geranmayeh, “Latent representation encoding and multimodal biomarkers for post-stroke speech assessment,” in *ICLR 2025 Workshop on World Models: Understanding, Modelling and Scaling*.
- [21] F. Eyben et al., “The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing,” *IEEE trans. on affective computing*, vol. 7, no. 2, pp. 190–202, 2015.
- [22] L. Juvela, B. Bollepalli, V. Tsiaras, and P. Alku, “Glottnet—a raw waveform model for the glottal excitation in statistical parametric speech synthesis,” *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 27, no. 6, pp. 1019–1030, 2019.
- [23] S. Lundberg, “A unified approach to interpreting model predictions,” *arXiv preprint arXiv:1705.07874*, 2017.
- [24] D.-C. Gruia, W. Trender, P. Hellyer, S. Banerjee, J. Kwan, H. Zetterberg, A. Hampshire, and F. Geranmayeh, “Ic3 protocol: a longitudinal observational study of cognition after stroke using novel digital health technology,” *BMJ open*, vol. 13, no. 11, p. e076653, 2023.
- [25] D. C. Gruia, V. Giunchiglia, A. Coghlan, S. Brook, S. Banerjee, J. Kwan, P. J. Hellyer, A. Hampshire, and F. Geranmayeh, “Online monitoring technology for deep phenotyping of cognitive impairment after stroke,” *medRxiv*, 2024.
- [26] K. Swinburn, G. Porter, and D. Howard, “Comprehensive aphasia test,” *APA PsycTests*, 2004.
- [27] S. Gonzalez and M. Brookes, “Pefac—a pitch estimation algorithm robust to high levels of noise,” *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 22, no. 2, pp. 518–530, 2014.
- [28] T. Drugman, M. Thomas, J. Gudnason, P. Naylor, and T. Dutoit, “Detection of glottal closure instants from speech signals: A quantitative review,” *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 20, no. 3, pp. 994–1006, 2011.
- [29] R. Timcke et al., “Laryngeal vibrations: Measurements of the glottic wave: Part i. the normal vibratory cycle,” *AMA Archives of Otolaryngology*, vol. 68, no. 1, pp. 1–19, 1958.
- [30] D. G. Childers and C. K. Lee, “Vocal quality factors: Analysis, synthesis, and perception,” *The Journal of the Acoustical Society of America*, vol. 90, no. 5, pp. 2394–2410, 1991.
- [31] G. Fant, “The lf-model revisited. transformations and frequency domain analysis,” *Speech Trans. Lab. Q. Rep., Royal Inst. of Tech. Stockholm*, vol. 2, no. 3, p. 40, 1995.
- [32] P. Alku, H. Strik, and E. Vilkman, “Parabolic spectral parameter—a new method for quantification of the glottal flow,” *Speech Commun.*, vol. 22, no. 1, pp. 67–79, 1997.
- [33] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, “Covarep—a collaborative voice analysis repository for speech technologies,” in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2014, pp. 960–964.
- [34] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [35] T. S. Kay, *Spectral analysis of stop consonants in individuals with dysarthria secondary to stroke*. Louisiana State University and Agricultural & Mechanical College, 2012.
- [36] N. Young and A. Blitzer, “Management of supraglottic squeeze in adductor spasmodic dysphonia: a new technique,” *The Laryngoscope*, vol. 117, no. 11, pp. 2082–2084, 2007.
- [37] K. W. Godin and J. H. Hansen, “Physical task stress and speaker variability in voice quality,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, pp. 1–13, 2015.
- [38] R. D. Pollock et al., “Respiratory muscle strength and training in stroke and neurology: a systematic review,” *International Journal of Stroke*, vol. 8, no. 2, pp. 124–130, 2013.
- [39] A. A. Joshy et al., “Automated dysarthria severity classification using deep learning frameworks,” in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 116–120.
- [40] F. Rudzicz et al., “The TORGO database of acoustic and articulatory speech from speakers with dysarthria,” *Language resources and evaluation*, vol. 46, pp. 523–541, 2012.
- [41] H. Kim, M. H. Johnson, J. Gunderson, A. Perlman, T. Huang, K. Watkin, S. Frame, H. V. Sharma, and X. Zhou, “UAspeech,” 2023. [Online]. Available: <https://dx.doi.org/10.21227/f9tc-ab45>
- [42] N. Narendra and P. Alku, “Glottal source information for pathological voice detection,” *IEEE Access*, vol. 8, pp. 67 745–67 755, 2020.