Phoneme-Specific Challenges to Intelligibility in Hearing Impairment Under Noisy Condition

Denawati Junia and Candy Olivia Mawalim
Japan Advanced Institute of Science and Technology, Japan
E-mail: {junia, candylim}@jaist.ac.jp

***Abstract*—Hearing impairment significantly reduces speech intelligibility, particularly in noisy acoustic environments, due to impaired auditory sensitivity and phoneme recognition. This study investigates whether speech intelligibility can be accurately predicted by integrating phoneme-level error patterns in noisy conditions. Using the Clarity prediction challenge dataset, phoneme-level errors were computed based on the international phonetic alphabet transcriptions and quantified by word error rate (WER). Our key findings reveal that high-frequency fricatives (/f/,/ʒ/) and affricates (/dʒ/), along with voiced phonemes (/g/,/ɔ/), showed the highest average WER. This indicates their particular vulnerability to masking and the effects of high-frequency hearing loss. While higher SNR generally improves intelligibility, we observed a weak correlation ($\rho \approx 0.20$), underscoring the critical role of individual hearing loss profiles. To further our analysis, we used the five most challenging phonemes and SNR as features to predict speech intelligibility with Random Forest and XGBoost models. This approach yielded slightly better prediction performance compared to the Hearing Aid Speech Perception Index.**

I. Introduction

Hearing loss is a significant global public health challenge. According to the World Health Organization (WHO), unaddressed hearing loss annually imposes a global economic burden that exceeds USD 980 billion, primarily due to reduced productivity, healthcare costs and social exclusion [1]. Verbal communication fundamentally depends on speech intelligibility, defines as the ability to understand spoken target speech [2]. Given these significant challenges, hearing aids (HAs) are widely recognized as crucial assistive devices designed to help hearing-impaired (HI) listeners.

Although HAs are crucial, their development aimed at improving speech intelligibility encounters substantial challenges, especially concerning their evaluation. Traditional speech intelligibility prediction methods relies on expensive, time-consuming, and often subjective testing with HI listeners [3]. Therefore, objective evaluation is a preferable alternative for evaluating system performance. Various approaches have been explored, including Articulation Index (AI) [4], [5], Speech Intelligibility Index (SII) [6] and Speech Transmission Index (STI) [7]. Most recent intrusive methods, including the Short-Time Objective Intelligibility (STOI) [8] and Hearing Aid Speech Perception Index (HASPI) [9].

More recent deep learning models have shown promising results in speech intelligibility prediction. These models can leverage acoustic features derived from Automatic Speech Recognition (ASR) systems and integrate auditory models, as seen in methods incorporating the Cambridge hearing loss model [10], [11]. Some recent approaches combine hearing loss modeling with cross-domain feature extraction, while others propose binaural processing and hybrid machine learning models to improve the prediction accuracy for HAs [12].

Despite advancements in speech intelligibility modeling, significant challenges persist, particularly for HI listeners attempting to recognize phonemes in noisy environments. While numerous studies have individually examined acoustic factors like signal-to-noise ratio (SNR) or phoneme-specific errors, few have explored their combined predictive value within an integrated framework [13], [14]. Consequently, existing models often fall short in capturing listener-specific intelligibility, especially when phoneme vulnerability and acoustic masking interact in complex ways.

Certain phonemes, such as high-frequency fricatives and voiced affricates, are particularly prone to misperception under noisy conditions due to their unique acoustic and articulatory mechanisms [12]. Yet, these phonetic weaknesses are rarely modeled explicitly, thereby limiting the granularity of prediction for HI listeners [14]. Furthermore, most available datasets lack phoneme-level alignments paired with perceptual outcomes, which constrains fine-grained error analysis [2]. As a result, a detailed understanding is still needed regarding which specific phonemes consistently impair intelligibility across different SNRs, and how these patterns vary among individual hearing profiles.

To address these gaps, this study aims to evaluate speech intelligibility by integrating phoneme-level error patterns and acoustic conditions. Specifically, we quantified phonetic representations based on the International Phonetic Alphabet (IPA) and identified the top phonemes exhibiting the highest Word Error Rate (WER). We also utilized SNR as the representation of acoustic conditions. These integrated features were then used to train supervised machine learning (ML) classifiers to examine how phoneme misperception frequencies and SNR collectively influence binary correctness outcomes across listeners.

II. Related Work

A. Evolution of Speech Intelligibility Prediction Models

1) Traditional Methods:

- Articulation index (AI): AI was designed to measure accessible information in speech by considering frequency-band-specific differences between speech and noise levels

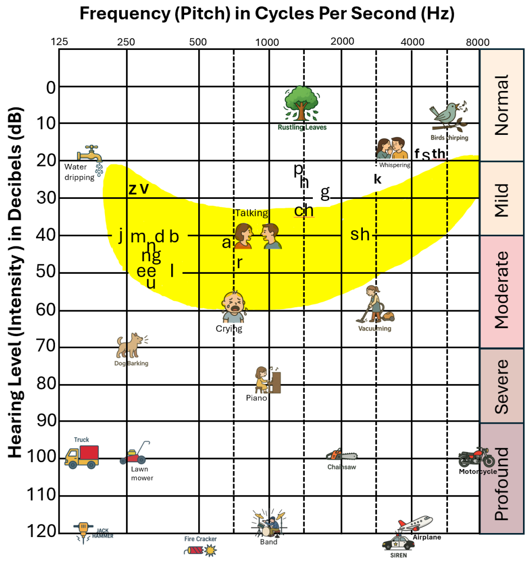

Fig. 1: Audiogram illustrating the "Speech Banana." The shaded "speech banana" region indicates the typical frequency and intensity range of human speech sounds, which are crucial for speech intelligibility.

[4], [5]. Its calculation involves summing contributions across frequency bands, with adjustments for masking effects and speech levels.

- Speech Transmission Index (STI): STI extends predictions to account for reverberant speech and nonlinear distortions. It analyzes the effect on modulation depth within frequency bands using a speech-like modulated noise signal [6], [15].
- Speech Intelligibility Index (SII): a 1997 revision of the 1969 ANSI S3.5 Standard [7], evolved from the Articulation Index (AI). It offers a more flexible framework by incorporating additional variables and correction factors. As audibility is paramount for word recognition, SII serves as a valuable tool for its estimation [16].

2) Modern Intrusive Methods:

- Short-Time Objective Intelligibility (STOI): introduced in 2011, is an intrusive measure that calculates the correlation between the degraded speech spectrum and a reference stimulus [2], [15]. A modified binaural STOI version enhances predictions by incorporating an equalization-cancellation stage, which aligns and cancels interfering noise to maximize the correlation between signals.
- Hearing Aid Speech Perception Index (HASPI): was proposed in 2014 and updated in 2021 [9], [10], is an intrusive metric specifically designed for evaluating hearing aid processing. It utilizes an auditory model to simulate impaired hearing, employing a fourth-order gammatone filterbank (GTFB) with bandwidth adjustments tailored for HI listeners.

3) Auditory Models and Hearing Loss Simulations: Recent development incorporates auditory system modeling to predict speech intelligibility and guide HA signal processing. These models simulate stages from outer to inner ear.

- The Cambridge hearing loss model has been utilized to preprocess binaural signals based on audiogram, simulating effects like loudness recruitment, reduced frequency selectivity, and outer/inner hair cell damage, resulting in attenuated or blurred high frequency consonants [17].
- The "Speech Banana" Audiogram, adapted from the American Academy of Audiology, represents a fundamental construct in audiometry and speech science, serving as a crucial visual representation within an audiogram to illustrate the relationship between and individual's hearing thresholds and their capacity to perceive sounds. The yellow banana shape in 1, outlines the approximate frequency and intensity range (20 - 70 dB) of the majority of human speech phonemes, encompassing both vowels and consonants.This speech banana highlights the critical frequencies that are essential for everyday speech comprehension.

4) Machine Learning Approaches: The integration of machine learning (ML) is significantly advanced speech intelligibility prediction, particularly in addressing the complexities of non-stationary noise and nonlinear distortions. Key aspects of these data-driven methodologies, include:

- MBI-Net and Whisper Integration: The Multi-Branched Speech Intelligibility Prediction Network (MBI-Net) [11] is a non-intrusive model designed for HAs applications. Its performance can be enhanced by incorporating pre-trained models like Whisper, which enriches the acoustic features. Whisper itself is an ASR model, extensively pre-trained on a vast multilingual dataset [10].
- Hybrid ML Models and Binaural Processing: These approaches utilize binaural processing and combine hybrid ML models like Long Short-Term Memory and Light Gradient Boosting Machine to enhance prediction accuracy for HA [12].

III. Methodology

Figure 3 summarizes our experimental workflow. This study utilized the dataset from the 2nd Clarity Prediction Challenge (CPC2), which contains speech intelligibility test results involving multiple hearing-impaired listeners across various acoustic conditions. The dataset includes percentage correct scores and SNR values. Phoneme-level features were extracted via IPA conversion. Phoneme-level features and SNR values were input to machine learning models for predicting speech intelligibility.

A. Phoneme-Level Error Analysis

A phoneme-level analysis was performed to identify the most difficult phonemes for HI listeners, achieved by calculating the WER for each IPA symbol. Table I details the top five phonemes exhibiting the highest average WER across all listening conditions, including their voicing and articulatory characteristics.

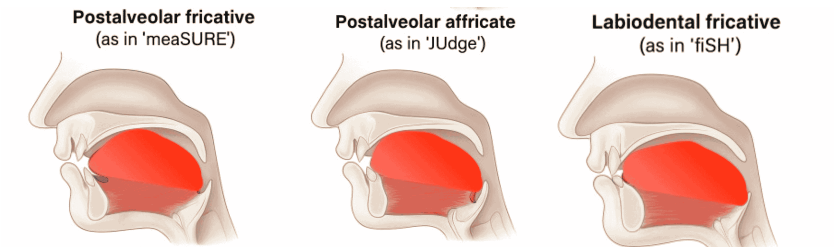

Fig. 2: Articulatory positions for selected fricative and affricate consonants: (a) postalveolar fricative /ʒ/ (as in */mɛʒər/*), (b) postalveolar affricate /dʒ/ (as in */dʒʊdʒ/*), (c) labiodental fricative /f/ (as in (*/fɪʃ/*)).

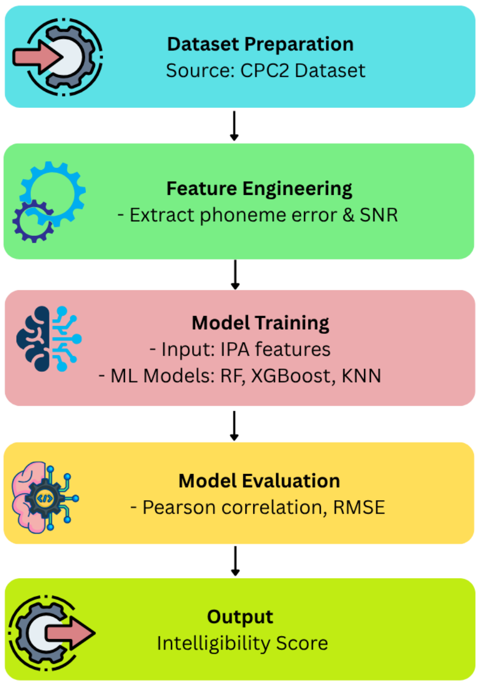

Fig. 3: Overview of the intelligibility prediction pipeline. The process includes dataset preparation, phonetic pre-processing, WER computation, model training, evaluation, and evaluation of speech intelligibility prediction.

As shown in Table I the phoneme /ʒ/ has the highest error rates at 32.1%, followed by /f/ 24.4%, /dʒ/ 24.2%, /ɔ/ at 24.1% and /g/ at 23.8%. These results suggest that both high-frequency fricatives (/ʒ/, /f/) and voiced affricates /dʒ/ are particularly susceptible to misperception. This trend is consistent with prior studies indicating that HI listeners often experience difficulty distinguishing high-frequency phonemes, especially under noisy conditions.

Among the five phonemes with the highest WER, four are voiced sounds: /ʒ/, /dʒ/, /ɔ/, and /g/.Voiced consonants typically require greater acoustic energy and involve vocal fold vibration, making them more vulnerable to masking by background noise. This finding supports the hypothesis that phonemes involving voicing are more difficult to process for impaired auditory systems under noise interference.

TABLE I: Voicing and Articulatory Features of the Top 5 Most Challenging IPA Phonemes in CPC2 Dataset. (Rank denotes the difficulty ranking, and Char. represents the IPA character.)

Rank	**Char.**	**WER (%)**	**Voicing**	**Articulation**
1	/ʒ/	32.1	Voiced	Postalveolar fricative (as in *measure*)
2	/f/	24.4	Voiceless	Labiodental fricative (as in *fish*)
3	/dʒ/	24.2	Voiced	Postalveolar affricate (as in *judge*)
4	/ɔ/	24.1	Voiced	Open-mid back rounded vowel (as in *thought*)
5	/g/	23.8	Voiced	Velar plosive (as in *go*)

Articulatory, the most error-prone phonemes include fricatives and affricates, which are known for their complex spectral structure and short duration. According to Roach [18], the phoneme /ʒ/, a postalveolar fricative, is produced by narrowing the airflow between the tongue and the rear part of the alveolar ridge. Similarly, the affricate /dʒ/ involves a complete closure followed by a turbulent release at the same articulatory location. In constrast, /f/ is a voiceless labiodental fricative, produced by constricting airflow between the lower lip and the upper teeth. These articulatory mechanisms are illustrated in Fig. 2, which depicts the tongue and lip position involved in the production of these consonants.

B. Effect of SNR on Speech Intelligibility

Beyond phoneme characteristics, acoustic environments significantly influence speech intelligibility. As illustrated in Fig.4, the mean WER shows a decreasing trend with increasing SNR, indicating that better SNR conditions generally improve intelligibility. However, the relationship between SNR and speech intelligibility is often nonlinear, with improvements more noticeable at low SNR and leveling off at higher SNR. In this study, the Pearson correlation between SNR and percentage correctness was $\rho = 0.2$, indicating a weak positive correlation. This suggests that while intelligibility generally improves as SNR increases, SNR alone explains only a small fraction of the variance in performance.

Fig.5 shows a shift toward higher correctness as SNR increases. At low or negative SNR (e.g., -6 and -4 dB), the plots are wider at the bottom, indicating a greater number of listeners with poor recognition. Despite this, some listeners still achieve relatively high scores under difficult conditions, highlighting

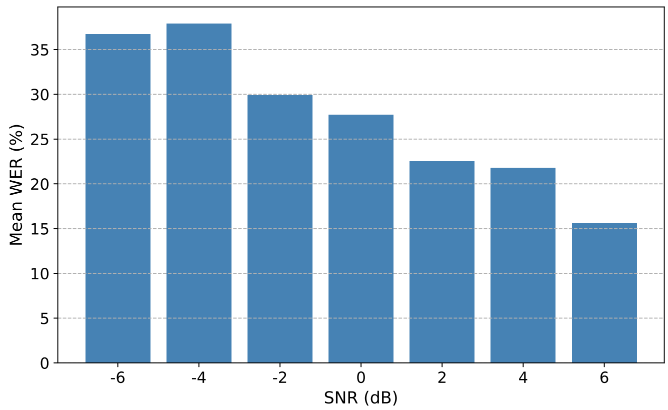

Fig. 4: Mean WER per SNR bin, with a decreasing trend in WER as SNR increases.

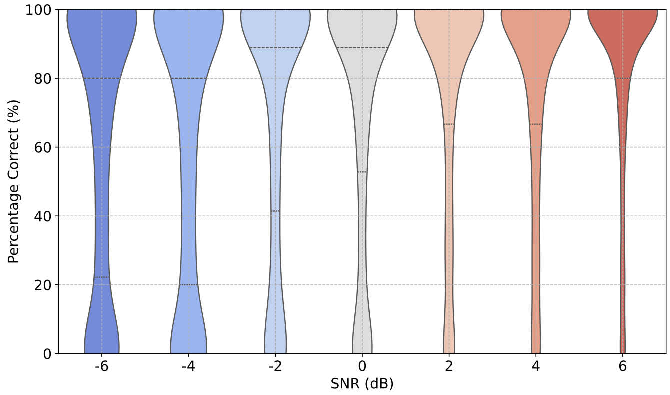

Fig. 5: Violin Plot Diagram of Percentage Correct Across Different SNR Values, with the plot shape changing from wide at low SNR to narrower and shifted to the right (higher percentage correct) at high SNR, but still showing wide "tails" or outliers at all SNR.

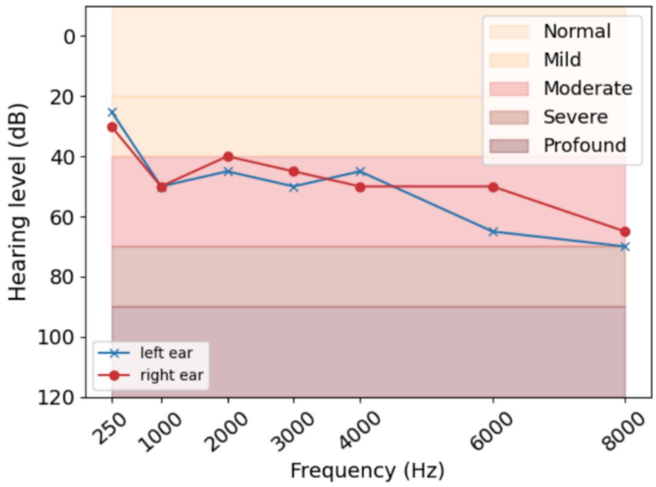

Fig. 6: Audiograms showing hearing thresholds for listener L0202, with right (red) and left (blue) ears plotted across frequency bands. The shaded areas represent different levels of hearing loss severity: normal (beige), mild (light orange), moderate (pink), severe (brown), and profound (dark brown).

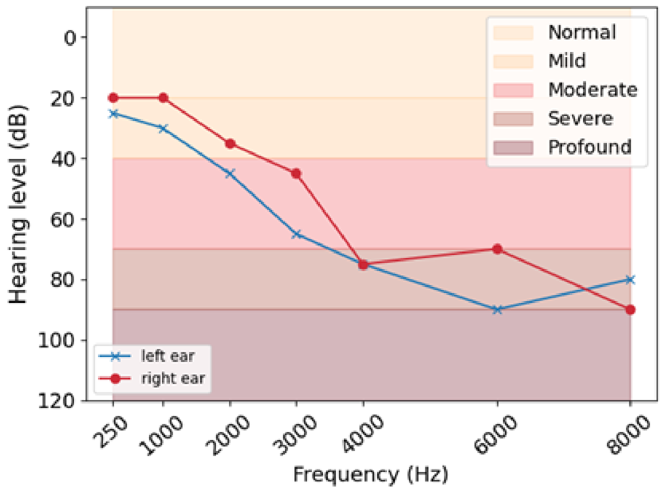

Fig. 7: Audiograms showing hearing thresholds for listener L0215. Description is equivalent to Fig. 6.

individual resilience. At mid to high SNR (+4, +6 dB), the plots are narrow and shift rightward, showing higher median performance, although some outliers with poor scores remain. These findings suggest that SNR alone does not fully determine speech intelligibility. Individual listener characteristics, such as hearing loss profiles or phoneme-specific deficits, may override the benefits of favorable SNR, underscoring the need for personalized intelligibility assessments.

C. Listener-Specific Audiogram Analysis

To further explore individual variability in speech intelligibility, audiogram-based analysis was conducted on two representative listeners: L0202 and L0215. Their audiograms, presented in Fig.6 and Fig.7 respectively, display hearing thresholds across frequencies for both ears, illustrating how high-frequency hearing loss can critically affect speech intelligibility under noisy conditions.

Listener L0202 exhibits a moderately asymmetric hearing profile, with more pronounced high-frequency hearing loss in the right ear. As seen in Fig.6, threshold remain within the normal to mild range below 2000 Hz but begin to decline sharply above 3000 Hz, particularly in the right ear, reaching the moderate loss zone. Despite relatively preserved low-frequency hearing, L0202 demonstrated poor intelligibility performance, consistent with prior studies showing that high-frequency loss disproportionately impairs the recognition of fricatives and affricates such as /ʒ/, /f/, and /dʒ/ in noisy conditions [12], [19].

Listener L0215, in contrast, presents a more symmetrical but more severe pattern of high-frequency hearing loss. As illustrated in Fig.7, both ears show thresholds falling into the moderate to severe range between 3000-8000 Hz, covering the spectral region crucial for perceiving high-frequency consonants. Notably, even when tested under favorable SNR conditions (+4 dB), L0215 exhibited poor recognition accuracy. This outcome highlights that peripheral hearing loss, particularly in high-frequency regions, can override the benefits of improved acoustic conditions.

Together, these audiogram profiles confirm that listener-specific auditory characteristics, especially high-frequency hearing loss, play a central role in shaping speech intelligibility outcomes.

IV. Speech Intelligibility Prediction

This section details the experimental setup for predicting speech intelligibility, utilizing phoneme-level error analysis results and SNR to represent acoustic conditions.

A. Dataset

Our study utilized the Clarity Prediction Challenge 2 (CPC2) dataset[1], which is specifically designed for speech intelligibility research involving hearing-impaired listeners. The CPC2 dataset comprises two subsets: CEC1 (simple acoustic scenes) and CEC2 (more complex, realistic conditions) [20]. As the label of original evaluation set was unavailable, we conducted our experiments on the provided training data, splitting it into a 70% training set and a 30% test set.

B. Features

To specifically evaluate phoneme-level intelligibility prediction, our models were exclusively trained using a highly targeted set of input features. These features comprised the five phonemes with the highest average WER, indicating their significant difficulty for listeners, alongside the SNR. The selected phonemes, identified through prior analysis includes: (/ʒ/, /dʒ/, /ɔ/, and /g/). These phonemes represent specific speech sounds: /ʒ/ is a voiced postalveolar fricative, /dʒ/ is a voiced postalveolar affricate, /ɔ/ is an open-mid back rounded vowel, and /g/ is a voiced velar plosive [18]. This targeted approach allows us to focus the prediction on the speech elements most susceptible to intelligibility issues.

C. Evaluation Metrics and Baseline

Model performance was evaluated using two metrics: Pearson correlation coefficient (ρ) and Root Mean Square Error (RMSE). Pearson correlation assesses the linear relationship between predicted and actual intelligibility scores, with higher values indicating better prediction accuracy. RMSE quantifies the average magnitude of the errors between predicted and actual values, where lower RMSE indicates greater accuracy. To provide a contextual benchmark, our prediction method's performance was compared against the Hearing Aid Speech Perception Index (HASPI), which served as a baseline from the Clarity Challenge. HASPI is an established objective intelligibility measure that simulates aspects of the auditory periphery, including the effects of hearing loss, to predict speech perception by comparing signal envelopes and temporal fine structures. This comparison highlights the efficacy of our phoneme-level and SNR-integrated approach in predicting speech intelligibility.

[1] https://claritychallenge.org/docs/cpc2/cpc2_data

TABLE II: Comparison of Speech Intelligibility Prediction Model Performance on the CPC2 Dataset, separated into CEC1 and CEC2 subsets. (Bold type indicates the best performance for each metric.)

Subset	Method	ρ (↑)	RMSE (↓)
CEC1	HASPI (Baseline)	0.6847	27.0815
	KNN	0.6665	28.4151
	XGBoost	0.6992	26.5195
	Random Forest	**0.6999**	**26.4936**
CEC2	HASPI (Baseline)	0.6464	29.9394
	KNN	0.7317	27.1044
	XGBoost	0.7243	27.4322
	Random Forest	**0.7465**	**26.8915**

D. Results

We utilized regression algorithms, specifically Random Forest (RF), k-Nearest Neighbors (kNN), and XGBoost, for their demonstrated efficacy in handling complex datasets. Table II shows that both RF and XGBoost slightly outperformed the HASPI baseline on CEC1. They achieved higher Pearson correlations (ρ = 0.6999) and lower RMSEs (approximately 26.52 for RF/XGBoost versus 27.08 for HASPI). This indicates that even in relatively simple acoustic environments, a phoneme-based approach can surpass a complex, physiologically motivated model like HASPI.

Despite the significantly more challenging conditions on CEC2, which involved multiple competing noise sources, the RF model continued to perform better than the HASPI. They achieved higher correlations (ρ = 0.7465) and lower RMSEs (RMSE = 26.89 for RF versus 29.93 for HASPI). This result is particularly important as it suggests that a simplified model utilizing a minimal set of strategically chosen phoneme features and SNR can maintain or even surpass the accuracy of a complex benchmark in demanding real-world scenarios.

This research demonstrates that utilizing interpretable, phoneme-based features provides transparent insights into speech intelligibility problems, offering a clear advantage over "black-box" models. Crucially, this approach reduces model complexity while maintaining high predictive performance, with a simple five-phoneme model competing effectively against complex, physiologically-motivated systems like HASPI. These advancements are vital for developing adaptive HA algorithms and personalized assesment tools in clinical practice. The non-intrusive nature of this method is especially critical for real-world HA development as it circumvents the need for clean reference speech signals. This capability allows for more accurate and individualized HA adjustments, marking a significant step towards optimizing outcomes for individual with hearing loss.

V. Conclusion and Future Work

This study demonstrates that speech intelligibility in hearing-impaired listeners can be effectively predicted by integrating phoneme-level errors and signal-to-noise ratios (SNR). Our findings reveal that high-frequency fricatives (/f/, /ʒ/) and affricates (/dʒ/), along with specific voiced phonemes (/g/,/ɔ/), consistently exhibited the highest error rates due to their susceptibility to masking and high-frequency hearing loss. While an increased SNR generally improves performance, its weak correlation ($\rho \approx 0.20$) and cases where severe high-frequency hearing loss resulted in 100% WER even at a high SNR highlight the dominant role of hearing profiles. Furthermore, utilizing these integrated features with Random Forest and XGBoost models achieved superior predictive performance compared to the Hearing Aid Speech Perception Index.

This study has several limitations. The use of the CPC2 dataset may not fully capture the diversity of acoustic environments and hearing loss profiles. Our reliance on WER might overlook nuanced phoneme-level errors. Additionally, the current feature set could be expanded to include richer auditory cues. Future work should therefore focus on utilizing larger, more diverse datasets and incorporating more detailed phonetic analyses. Furthermore, improving prediction accuracy through advanced auditory and listener-specific models for adaptive hearing aids are crucial next steps.

Acknowledgment

This work was partially supported by JSPS KAKENHI No. 25K21245.

References

[1] S. Chadha, K. Kamenov, and A. Cieza, *The world report on hearing*. Bulletin of the World Health Organization, 2021, vol. 99, 242–242A. DOI: https://doi.org/10.2471/BLT.21.285643.

[2] Y. Feng and F. Chen, "Nonintrusive objective measurement of speech intelligibility: A review of methodology," *Biomedical Signal Proc. and Control*, vol. 71, p. 103 204, 2022. DOI: 10.1016/j.bspc.2021.103204.

[3] L. L. Mendel, "Objective and subjective hearing aid assessment outcomes," *Am. J. of Audiology*, vol. 16, no. 2, pp. 118–129, 2007. DOI: 10.1044/1059-0889(2007/016).

[4] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.*, vol. 19, no. 1, pp. 90–119, 1947. DOI: 10.1121/1.1916407.

[5] K. D. Kryter, "Methods for the calculation and use of the articulation index," *J. Acoust. Soc. Am.*, vol. 34, no. 11, pp. 1689–1697, 1962. DOI: 10.1121/1.1909094.

[6] T. Houtgast and H. J. M. Steeneken, "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.*, vol. 67, no. 1, pp. 318–326, 1980.

[7] American National Standards Institute (ANSI), *Methods for Calculation of the Speech Intelligibility Index*, 1997.

[8] K. Payton and M. Shrestha, "Evaluation of short-time speech-based intelligibility metrics," 2008.

[9] J. M. Kates and K. H. Arehart, "The Hearing-Aid Speech Perception Index (HASPI)," *Speech Comm.*, vol. 65, pp. 75–93, 2014, ISSN: 0167-6393. DOI: 10.1016/j.specom.2014.06.002.

[10] R. Mogridge, G. Close, R. Sutherland, S. Goetze, and A. Ragni, "Pre-Trained Intermediate ASR Features and Human Memory Simulation for Non-Intrusive Speech Intelligibility Prediction in the Clarity Prediction Challenge 2." [Online]. Available: https://api.semanticscholar.org/CorpusID:276234173.

[11] R. E. Zezario, F. Chen, C.-S. Fuh, H.-M. Wang, and Y. Tsao, *Mbi-net: A non-intrusive multi-branched speech intelligibility prediction model for hearing aids*, 2022. eprint: 2204.03305. [Online]. Available: https://arxiv.org/abs/2204.03305.

[12] X. Zhou, C. O. Mawalim, and M. Unoki, "Speech intelligibility prediction using binaural processing for hearing loss," *IEEE Access*, vol. 13, pp. 25 817–25 831, 2025. DOI: 10.1109/ACCESS.2025.3538708.

[13] C. O. Mawalim, B. A. Titalim, S. Okada, and M. Unoki, "Non-intrusive speech intelligibility prediction using an auditory periphery model with hearing loss," *Applied Acoustics*, vol. 214, p. 109 663, 2023. DOI: 10.1016/j.apacoust.2023.109663.

[14] M. Huckvale and G. Hilkhuysen, "Evaluating a 3-factor listener model for prediction of speech intelligibility to hearing-impaired listeners," in *Proc. of Interspeech*, Kos, Greece, 2024.

[15] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech," *IEEE TASLP*, vol. 19, no. 7, pp. 2125–2136, 2011. DOI: 10.1109/TASL.2011.2114881.

[16] M. R. F. d. S. C. A. da Silva Andrade and M. C. M. Iorio, "Speech recognition and speech intelligibility index in intra-aural hearing aids users: A comparative study," *Audiol. Commun. Res.*, vol. 25, p. e2362, 2020. DOI: 10.1590/2317-6431-2020-2362.

[17] Y. Nejime and B. Moore, "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise," *J. Acoust. Soc. Am.*, vol. 102, pp. 603–15, Aug. 1997. DOI: 10.1121/1.419733.

[18] P. Roach, *English Phonetics and Phonology: A Practical Course*, Fourth. Cambridge, UK: Cambridge University Press, 2009, ISBN: 978-0-521-71740-3.

[19] S. Fuchs and P. Birkholz, "Phonetics of Consonants," in *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, 2019. DOI: 10.1093/acrefore/9780199384655.013.410.

[20] M. A. Akeroyd, W. Bailey, J. Barker, *et al.*, *The 2nd Clarity Enhancement Challenge for Hearing Aid Speech Intelligibility Enhancement: Overview and Outcomes*, 2023. DOI: 10.1109/icassp49357.2023.10094918.