Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

BanglaVowelDataset: True AC and Synthetic BC Bangla Vowel Datasets in Speech Information System

Ohidujjaman -

ohidujjaman@cse.uiu.ac.bd

United International University

Bejoy Munshi
United International University

Mahmudul Hasan
Comilla University

Mohammad Nurul Huda
United International University

Suman Ahmmed
United International University

Hasan Sarwar
United International University

Tetsuya Shimamura
Saitama University

data-descriptor

Keywords:

Posted Date: September 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9816127/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

BanglaVowelDataset: True AC and Synthetic BC Bangla Vowel Datasets in Speech Information System

Authors & Affiliations

- Ohidujjaman[1,*], Bejoy Munshi[2], Mahmudul Hasan[3], Mohammad Nurul Huda[4], Suman Ahmmed[5], Hasan Sarwar[6], Tetsuya Shimamura[7]

[1,2,4,5,6] Computer Science and Engineering, United International University, Dhaka 1212, Bangladesh
[3] Computer Science and Engineering, Comilla University, Comilla 3506, Bangladesh
[7] Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan
*Corresponding author(s). E-mail(s): ohidujjaman@cse.uiu.ac.bd;
Co-authors Email(s): bmunshi202154@bscse.uiu.ac.bd ; mh@cou.ac.bd; mnh@cse.uiu.ac.bd; suman@cse.uiu.ac.bd; hsarwar@cse.uiu.ac.bd; shima@mail.saitama-u.ac.jp

Abstract

The BanglaVowelDataset resolves the unavailability of Bangla AC and synthetic BC vowels in the speech information system. We recorded Bangla air-conducted (AC) vowels through five male and five female speakers. The length of each recorded vowel ranges from 600 to 780 ms. We obtained a total of 120 true AC vowels from five male and five female speakers, as the Bangla alphabet has twelve vowels. In this paper, we focus on two datasets: (i) 120 raw AC vowels and (ii) 120 synthetic BC vowels. In the literature, the Bangla vowel speech dataset is not available yet. Consequently, this dataset has a significant opportunity for further research. Bangla vowel dataset is used for noisy and clean environs comparatively, for speech recognition and speaker identification. Commonly, AC speech is severely affected by ambient noise, whereas BC speech suffers less. Thus, the appropriate method is suggested for noise robustness. This dataset has potential for pitch (F0) detection, spectrum estimation, and condition number determination in speech signal processing and analysis. Performance of distinct approaches, including machine learning, deep learning, and statistical methods, is evaluated using raw AC and synthetic BC vowels.

Background & Summary

In speech signal research, data from both English and Japanese dialects are available[1,2,3,4]. There are twelve vowels in the Bangla alphabet due to diphthongs and variations. We recorded true Bangla air-conducted (AC) vowels, with five male and five female speakers participating in the recording system, set up in a soundproof room. Experts trained male and female speakers to learn the proper pronunciation of Bangla vowels. We converted the recorded true AC vowels into synthetic BC vowels using an all-pole source filter model to obtain a synthetic Bangla BC dataset. The length of each recorded raw Bangla vowel ranges from 600 to 780 ms. The sampling frequency, audio resolution correspond to 44100 Hz and 32-bit, respectively. Raw speech data are recorded in .wav format. We set the stereo in Audacity audio recording software while recording vowels to have a more natural and immersive experience. We obtained a total of 120 true AC vowels from five male and five female speakers, as the Bangla alphabet has twelve vowels. The recoded each vowel is cropped to avoid the unvoiced portions. Thus, the cropped vowel length ranges from 500 to 650 ms.

In addition, we created a synthetic dataset of the BC Bangla vowels containing 120 vowels sampled at 8000 Hz, having discarded the unvoiced portions. In this paper, we highlight two datasets: raw data consisting of 120 recorded Bangla AC vowels and synthetic data containing 120 synthetic Bangla BC vowels. Table 1 shows the overview of the dataset.

Currently, there is no available dataset for Bangla vowel speech in the literature. Consequently, this Bangla vowel dataset provides a significant scope for further research. This dataset is used for noise-robust, clean environments in speech recognition and speaker identification. However, AC speech is severely affected by additive noise, whereas BC speech is less affected [5]. Therefore, the precise method is proposed in noise-robust cases. This dataset is potentially to pitch detection, spectrum estimation, condition number calculation, formant estimation, pitch period detection, formant estimation error, and ill-conditioned measurement in a speech information system. We can evaluate the performance of machine learning, deep learning, and statistical approaches through true AC and synthetic BC vowels. Moreover, BC microphones are extremely rare for collecting original BC speech data in practice. In this paper, we employed an all-pole source filter model to create a synthetic BC Bangla vowel dataset for advanced research[6]. The BanglaVowelDataset, specifically the AC and BC vowel data, significantly encourages researchers in the field of speech signal processing and analysis. The potential reuse values of data are as follows:

- Bangla vowel data is mysterious since it contains twelve letters. At present, there is no dataset for Bangla vowel speech available in the literature. Accordingly, this dataset provides a significant scope for further research.
- This dataset is used for noisy and clean environments in a comparative fashion for speech recognition and speaker identification. Commonly, AC speech is severely affected by ambient noise, whereas BC speech suffers less [7,8]. Thus, the appropriate method is suggested in noise-robust cases.
- This dataset has potential for fundamental frequency (F0) detection, spectrum estimation, and condition number calculation in a speech information system.
- Performance of machine learning, deep learning, and statistical approaches is evaluated through true AC and synthetic BC vowels.

Methods

Experimental setup and data acquisition

We recorded raw Bangla AC vowels, with five male and five female speakers participating in the recording system, set up in a soundproof room. The speech data were acquired from Multimedia Lab (706), UIU. Five male and five female speakers, who are undergraduate students, have joined the recording process. Institutional Research Ethics Board of United International University (UIU) approved the data collection and this study, IREB Reference No.: IREB-2025-040. Positions of the equipment are regulated at each specific distance, illustrated in Figure 1. The recording microphone is 30 cm away from the speaker. The microphone was positioned 100 cm above the ground. We employed an AC microphone, model: HAVIT GAMENOTE GK61, for capturing true AC Bangla vowels. The length of each recorded Bangla vowel ranges from 600 to 780 ms. The equipment utilized is listed in Table 1. The sampling frequency and audio resolution correspond to 44100 Hz and 32-bit, respectively. Raw speech data are recorded in `.wav` format. We set the stereo in Audacity audio recording software while recording vowels to have a more natural and immersive experience. We obtained a total of 120 true AC vowels from five male and five female speakers, as the Bangla alphabet has twelve vowels. Table 2 shows the features of the recorded AC Bangla vowel dataset.

Data preparation

Each recorded vowel is cropped to avoid the unvoiced portions. Thus, the cropped vowel length ranges from 500 to 650 ms. Figure 2 shows the waveform of the raw and cropped AC vowel for the vowel /অ/ case. In this paper, we focus on two datasets: raw data consisting of 120 recorded Bangla AC vowels and synthetic data containing 120 synthetic Bangla BC vowels. The cropped AC vowels are then resampled from 44.1

Equipment	Specification
Soundproof room	Multimedia lab, UIU
AC microphone	HAVIT GAMENOTE GK61
Recording SW	Audacity
Laptop	LENOVO X1 Carbon, 7^{th} Gen.
System type	64 bits OS
Processor	Core i5
Clock speed	2.5 GHz
RAM	8 GB

Table 1: Equipment specification

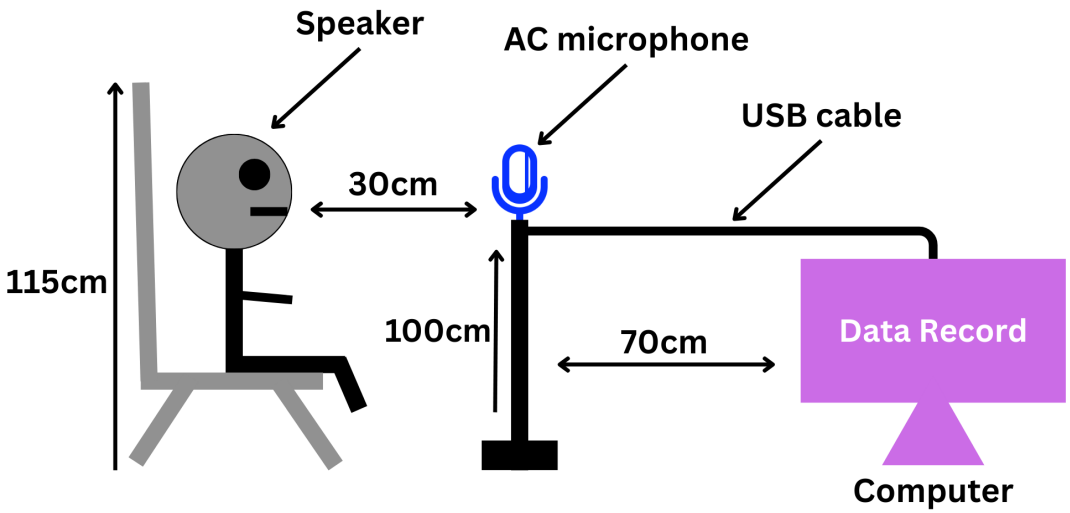


Figure 1: Data recording infrastructure

Features	Details
Dialect	Bangla vowels
Number of Bangla vowel	12
Speakers	5 males & 5 females
Total AC vowel data	12*10=120
Speech format	.wav
Speech signal length	600–780 ms
Sampling frequency	44100 Hz
Resolution in audio	32-bit
Audio mode	Stereo

Table 2: Features of recorded AC Bangla vowel

kHz to 8 kHz for voice communication purposes. The resampled AC vowels are quantized to control error, improve noise immunity, and simplify signal processing [9]. In addition, we created a synthetic dataset of the BC Bangla vowels containing 120 vowels sampled at 8 kHz, having discarded the unvoiced portions. The synthetic AC vowel is obtained from the true AC vowel using an all-pole source filter model [10]. Finally, the synthetic BC Bangla vowels are generated from synthetic AC vowels through an infinite impulse response (IIR) filter [11]. Thus, the total number of data is specified in the following Table 3.

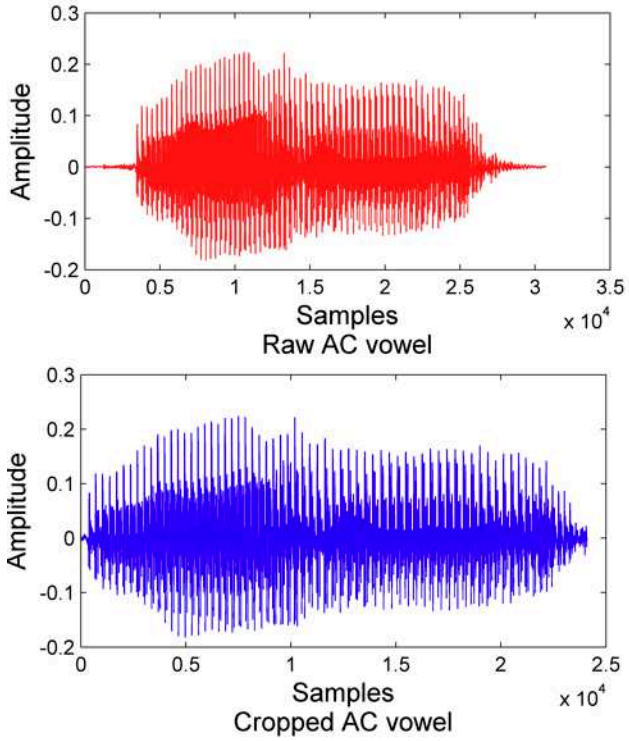


Figure 2: Waveform of raw and cropped AC vowels

Bangla vowel data	Number of vowels
Recorded raw AC vowel	120
Cropped AC vowel	120
Resampled AC vowel	120
Quantized AC vowel	120
Synthetic AC vowel	120
Synthetic BC vowel	120
Total	**720**

Table 3: Total number of Bangla vowel data

Data annotation

We annotated the recorded vowels by including the speaker identification number, gender mark, speech type, and vowel identification number. In this instance, **M** stands for male, **F** stands for female, **AC** stands for air-conduction, **BC** stands for bone-conduction, **1** stands for the first speaker, and **P** represents the vowel, with `P01` indicating the first vowel. Therefore, the first recorded vowel of the male speaker is annotated as `M1_AC_P01.wav`. Phonetic annotation for Bangla vowels is labelled as follows: /অ/ = P01, /আ/ = P02, /ই/=P03, /ঈ/ = P04, /উ/ = P05, /ঊ/ = P06, /ঋ/ = P07, /এ/ = P08, /ঐ/ = P09, /ও/ = P10, /ঔ/ = P11, and /অ্যা/ = P12. Table 4 shows instances of recorded vowels annotation.

Datasets generation framework

We generated two datasets, including raw AC Bangla vowel and synthetic BC Bangla vowel. To generate synthetic BC vowels, we face several stages. The overall data collection and preparation framework is shown in Figure 3.

Recording vowel naming	Gender	Speech type	Vowel representation	Audio format
M1_AC_P01.wav	male	AC	P01	.wav
F4_AC_P10.wav	female	AC	P10	.wav
M5_AC_P04.wav	male	AC	P04	.wav
F2_AC_P08.wav	female	AC	P08	.wav
F3_AC_P12.wav	female	AC	P12	.wav

Table 4: Instances of recorded vowels annotation

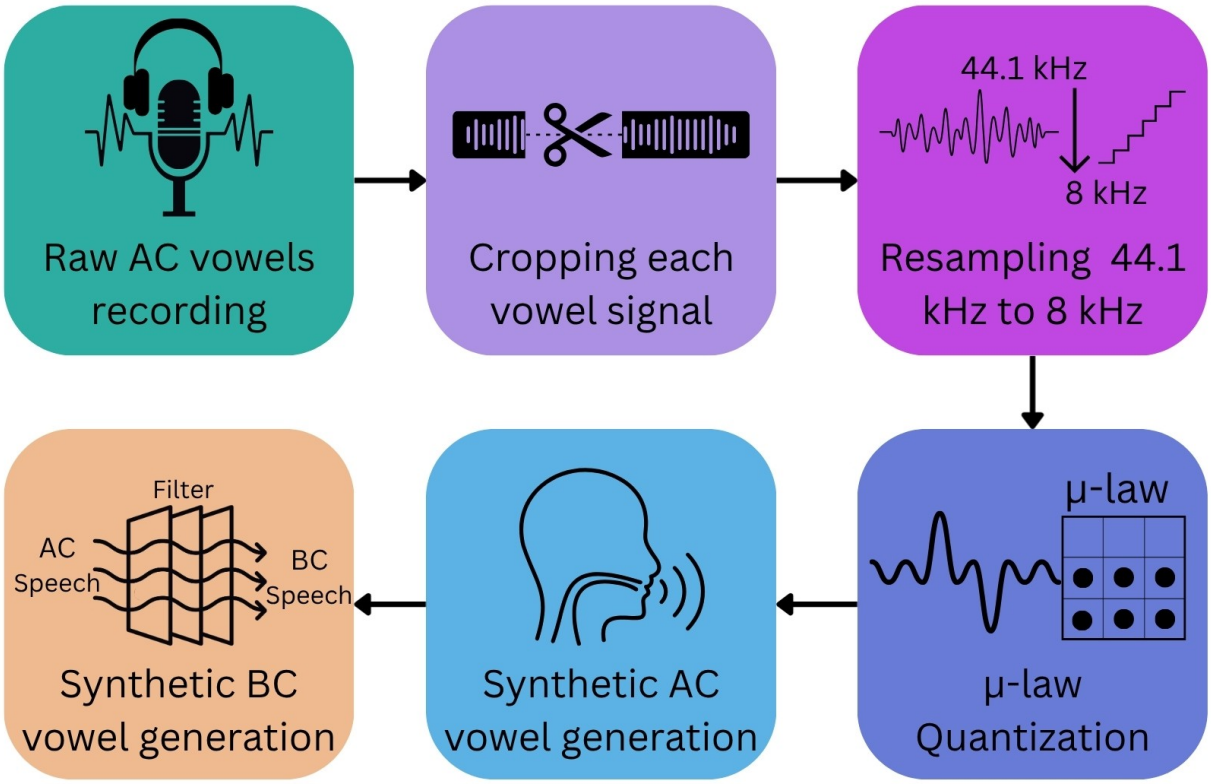


Figure 3: Framework of Bangla vowel datasets generation

Data cropping

Recorded vowel data contain silent portions on both the left and right sides of the recorded speech signal. In speech signal processing, these unvoiced segments negatively impact the performance of deployed methods for pitch detection, spectrum estimation, speaker recognition, formant estimation, spectral dynamic range estimation, and so on. Therefore, we eliminated the unvoiced portion from he recorded vowel signals. Figure 2 shows the waveform of the raw and cropped AC vowel for the vowel /অ/ case.

Data resampling

Speech signal resampling is the process of changing the sampling rate. Resampling speech signals achieves various goals, including improving compatibility between distinct systems, allowing for pitch variation, reducing computational time, and enhancing audio quality. In a lower sampling rate, the speech signal reduces computational complexity, especially in real-time processing. We resampled the original recorded vowel signal from 44.1 kHz to 8 kHz for voice communication purposes. An 8 kHz sampling frequency is suitable for voice over Internet Protocol (VoIP)[12].

Vowels	/অ/	/আ/	/ই/	/ঈ/	/উ/	/ঊ/	/ঋ/	/এ/	/ঐ/	/ও/	/ঔ/	/অ্যা/
G_f	0.176	0.201	0.142	0.988	1.478	0.581	1.005	1.362	1.845	1.932	1.436	0.154

Table 5: Gain factors for synthetic AC vowels

$\alpha(k)$	/অ/	/আ/	/ই/	/ঈ/	/উ/	/ঊ/	/ঋ/	/এ/	/ঐ/	/ও/	/ঔ/	/অ্যা/
α_1	-0.731	-0.746	-0.783	-0.658	-0.747	-0.740	-1.551	-0.771	-0.800	-0.740	-0.780	0.718
α_2	0.387	0.405	0.406	0.213	0.402	0.384	0.781	0.413	0.459	0.316	0.404	0.310
α_3	-0.705	-0.664	-0.654	-0.597	-0.710	-0.635	-0.759	-0.749	-0.739	-0.605	-0.775	-0.629
α_4	0.278	0.290	0.272	0.239	0.240	0.172	0.507	0.375	0.335	0.260	0.323	0.228
α_5	-0.065	-0.085	-0.099	0.035	-0.071	-0.035	0.277	-0.179	-0.137	-0.062	-0.083	-0.063
α_6	-0.028	0.018	-0.026	-0.098	-0.023	-0.036	-0.090	0.027	0.026	-0.017	0.041	-0.020
α_7	-0.193	-0.211	-0.219	-0.186	-0.183	-0.219	-0.137	-0.216	-0.285	-0.255	-0.157	-0.224
α_8	0.251	0.299	0.358	0.230	0.268	0.323	0.116	0.308	0.333	0.360	0.259	0.352
α_9	-0.183	-0.250	-0.262	-0.089	-0.212	-0.263	-0.521	-0.158	-0.273	-0.237	-0.208	-0.216
α_{10}	0.010	-0.020	0.029	0.070	0.050	0.102	0.388	-0.025	0.099	0.002	-0.014	-0.010

Table 6: Set of coefficients used to generate synthetic Bangla AC vowels

Data quantization

Quantization lessens the amount of data required to represent speech. Thus, Speech appears to be more efficient for both transmission and storage. Quantization allows for digital representation, manipulation, and lower bandwidth for transmission. We applied μ-law quantization to the recorded data after resampling. μ-law quantization enhances digital audio transmission quality for speech signals by addressing the varying amplitudes of speech. μ-law compresses high-amplitude signals while expanding low-amplitude signals, improving the signal-to-noise ratio (SNR), especially for quieter sounds [13].

Synthetic AC vowels generation

We applied a filter-source model for the synthetic AC vowels generation from quantized AC vowels. Synthetic AC vowels are generated by exciting an all-pole filter with a periodic impulse train [14]. The all-pole source filter mode is defined as,

$$V(z) = \frac{G_f}{1 + \sum_{k=1}^{N} \alpha(k) z^{-k}} \tag{1}$$

In (1), G_f specifies gain factors, and $\alpha(k)$ denotes filter coefficients. We consider an LP order (N) of 10. The energy of the error series is used to quantify the gain factors [10]. A set of gain factors is listed in Table 5 to produce synthetic AC vowels. Table 6 lists a set of filter coefficients acquired through the quantized AC input vowels to produce synthetic AC vowels.

Synthetic BC vowels generation

Synthetic BC vowels were derived from synthetic AC vowels. We generate synthetic BC vowels using a low-pass filter to synthetic AC. A low-pass filter is deployed using an infinite impulse response (IIR) filter. We use an IIR filter representation based on the approach described in reference [13]. An IIR filter is used on the synthetic AC signal to produce the synthetic BC signal, stated as follows:

$$\hat{b}(n) = \mu\, \hat{b}(n-1) + \phi a(n) \tag{2}$$

where $a(n)$ and $\hat{b}(n)$ correspond to the synthetic AC signal and synthetic BC signal, respectively [15,16]. We confidently designated the filter coefficients μ and ϕ as 0.80 and 0.30, respectively. These coefficients decisively demonstrate the spectral chink between AC and BC signals. Figure 4 illustrates the amplitude response of an IIR filter, indicating a calculated attenuation of 18 dB in the high-frequency region. The power spectra of synthetic AC and BC vowels (vowel /অ/ case) are shown in Figure 5. In this figure, it is evident that the spectral dynamic range is greater in the synthetic BC vowel compared to the synthetic.

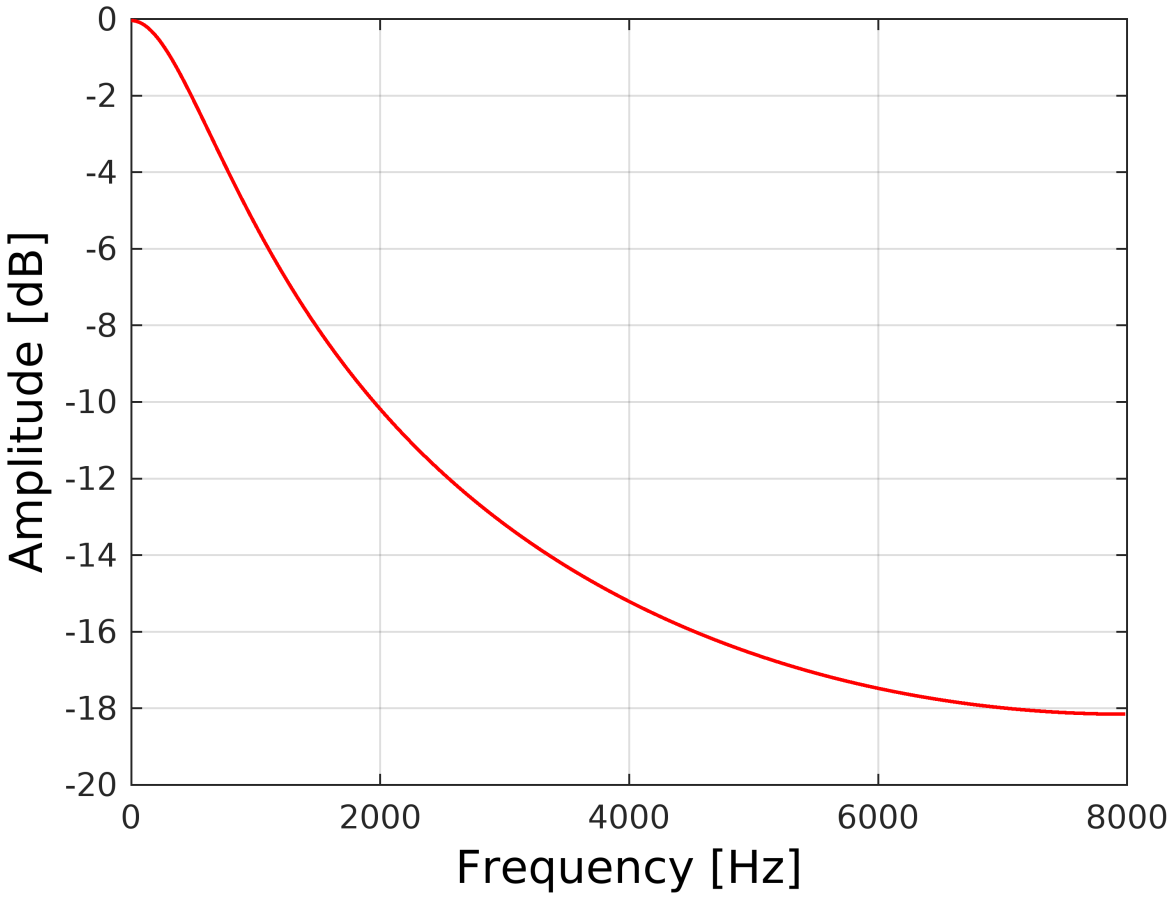


Figure 4: Amplitude response of the IIR filter

Data Record

The BanglaVowelDataset includes true AC, cropped AC, resampled AC, quantized AC, synthetic AC, and synthetic BC vowels, respectively. The dataset is publicly available on Figshare, with the DOI number: 10.6084/m9.figshare.30395665 [17]. It can be accessed through this link: https://doi.org/10.6084/m9.figshare.30395665 [17]. Additionally, the dataset contains two CSV files: BanglaVowelData.csv and BanglaVowelDataAnnotation.csv. The speech data is in *.wav* format.

Folder structure

The dataset folder is organized into six subfolders: Recorded True AC Vowel, Cropped AC Vowel, Resampled Quantized AC Vowel, Synthetic AC Vowel, and Synthetic BC Vowel. Each of these main subfolders is further divided into two subfolders containing male and female vowel data. Within each male and female subfolder, there are additional subfolders labelled Male Speaker 1 to 5 and Female Speaker 1 to 5. Each of these speaker subfolders contains 12 uniquely named vowels to avoid naming conflicts and facilitate easy traceability. The Quantized AC vowel subfolder includes a Validation folder containing all speech .wav files that are used to validate methods involving synthetic AC and BC vowels. The synthetic AC and BC subfolders are divided into train and test subfolders, where fold-out cross-validation is applied to create training and testing datasets. We allocate 75% of the data for training and 25% for testing. The folder structure is illustrated in Figure 6.

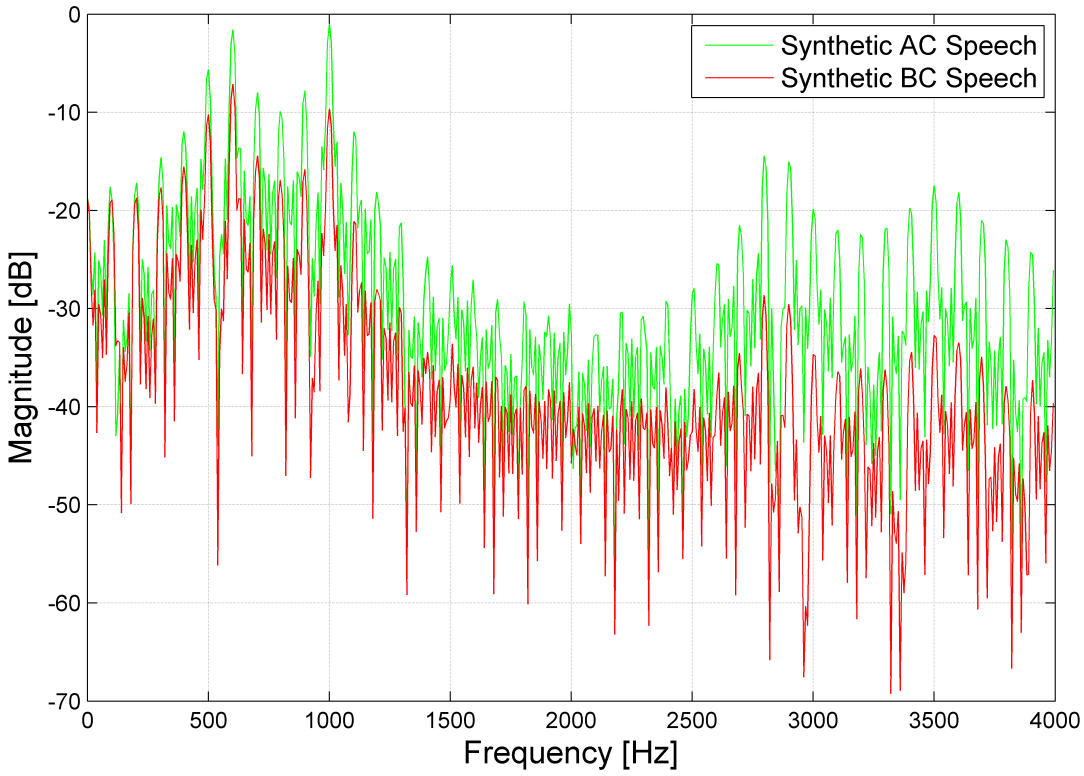


Figure 5: Power spectra of the synthetic AC and BC vowels (vowel / অ/ case)

Metadata

Metadata "BanglaVowelData.csv" contains SPEECH_SIGNAL_NAME: name of speech file with file extension, SPEECH_CATERGORY: category of speech (true AC, cropped AC, resampled AC, Quantized AC, synthetic AC, synthetic BC), GENDER_TYPE: type of gender (male or female), SAMPLING_RATE: sampling frequency (44.10 kHz or 8 kHz), SPEECH_TYPE: type of speech (training, testing, and validation). A summary of the fields is presented in Table 7.

Technical Validation

We prepared synthetic BC Bangla vowels from recorded AC vowels. In this section, we validate the synthetic BC data in various aspects. We estimate the condition number and spectral envelope of the synthetic BC vowel over the true AC vowel. Table 8 shows the condition numbers of true AC and synthetic BC vowels. Synthetic BC yields large condition numbers in comparison to the true AC vowel. In Table 8, it is noted that synthetic BC provides a wider spectral dynamic range, resulting in large condition numbers [18,19]. In summary, synthetic BC vowels are accurately prepared because they have a large condition number in comparison to original AC vowels, as shown in Table 8. Figure 7 shows the spectral envelopes of synthetic AC and synthetic BC vowels. In Figure 7, we observed that the spectral envelope curves of the synthetic BC vowel has a lower intensity due to its wider spectral dynamic range [20,21]. This fact validates that the synthetic BC vowel is prepared accurately from the synthetic AC through the IIR filter.

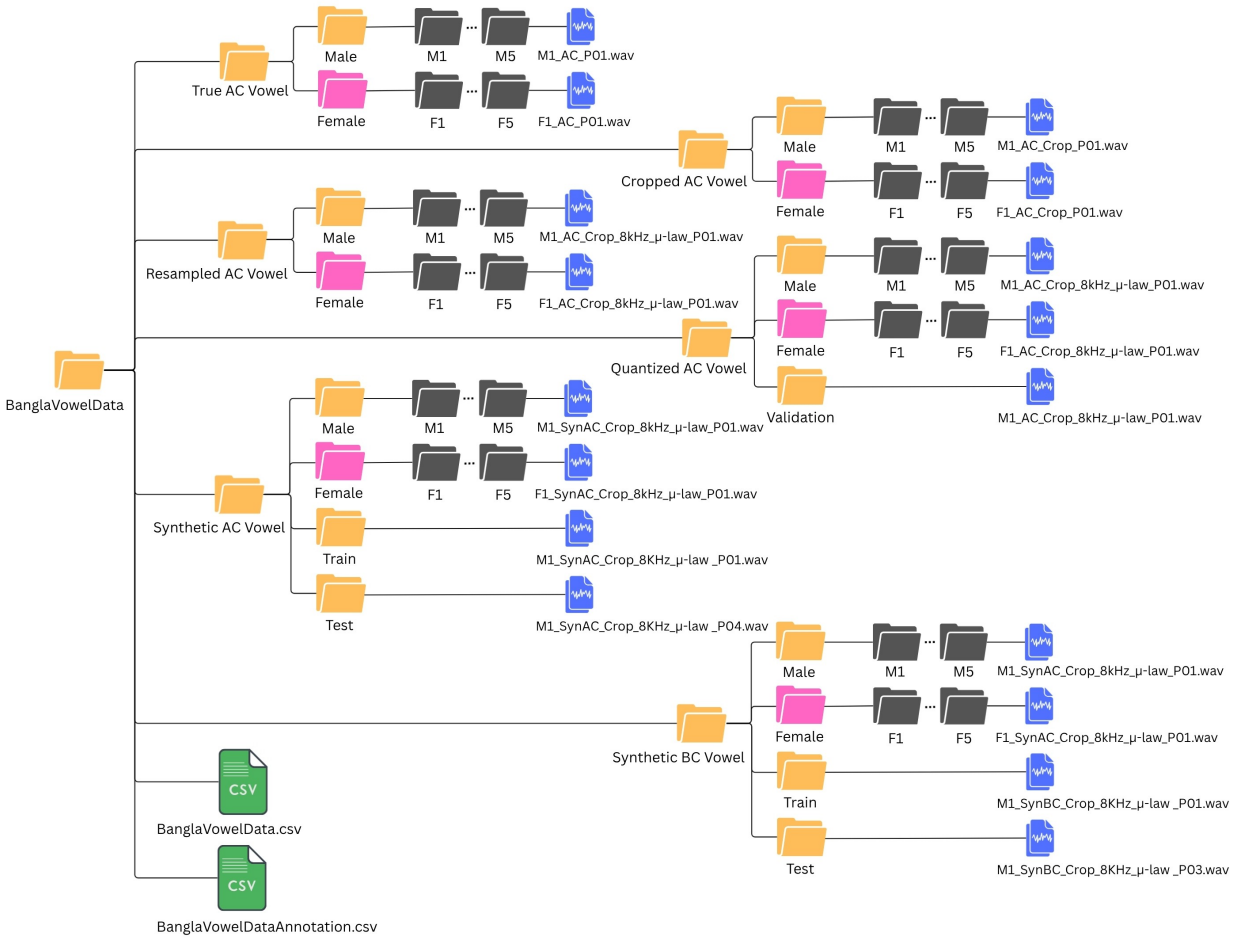


Figure 6: Folder structure of the dataset

Usage Notes

In this BanglaVowelDataset, we provide speech data: true AC, cropped true AC, resampled true AC, quantized true AC, synthetic AC, and synthetic BC vowels. We suggest reusing the BanglaVowelDataset for the researchers as follows:
(i) Quantized true AC speech vowels, totaling 120 data points, are suggested for validation.
(ii) For training purposes, 75% of the data is taken from synthetic AC and 75% from synthetic BC.
(iii) For testing, 25% of the data is drawn from synthetic AC and 25% from synthetic BC.
(iv) In both training and testing, approximately 50% of the data is sourced from male speakers and 50% from female speakers.
(v) Additionally, researchers may reuse the BanglaVowelDataset at their discretion.

Data Availability

The comprehensive dataset includes true AC, cropped AC, resampled AC, quantized AC, synthetic AC, and synthetic BC vowels, respectively. The dataset is publicly available on Figshare, with the DOI number: 10.6084/m9.figshare.30395665. It can be accessed through this link: https://doi.org/10.6084/m9.figshare.30395665. Additionally, the dataset contains two CSV files: BanglaVowelData.csv and BanglaVowelDataAnnotation.csv. The speech data is in *.wav* format. The BanglaVowelDataset contains train, testing, and validation data for method evaluation.

Field Name	Description
SPEECH_SIGNAL_NAME	name of speech file with file extension
SPEECH_CATERGORY	category of speech (true AC, cropped AC, resampled AC, Quantized AC, synthetic AC, synthetic BC)
GENDER_TYPE	type of gender (male or female)
SAMPLING_RATE	sampling frequency (44.10 kHz or 8 kHz)
SPEECH_TYPE	type of speech (training, testing, and validation)

Table 7: Summary of fields in the Metadata "BanglaVowelData.csv file"

Bangla vowels	True AC	Synthetic BC
/অ/	58.59	73.06
/আ/	44.76	56.50
/ই/	54.33	73.74
/ঈ/	50.49	70.64
/উ/	69.59	86.75
/ঊ/	78.73	92.21
/ঋ/	44.19	63.07
/এ/	37.91	51.20
/ঐ/	46.86	67.47
/ও/	70.19	88.81
/ঔ/	62.39	88.81
/অ্যা/	38.29	61.07
Average	54.69	72.19

Table 8: Condition number of true AC and synthetic BC vowels

Code Availability

We used MATLAB codes for resampling and μ-law quantization. The MATLAB codes are used to generate synthetic AC and synthetic BC vowels. For technical validation, the codes are used to estimate the condition number and spectral envelopes. All codes are provided in GitHub https://github.com/tuhiniu31/BanglaVowel.git for transparency and verification.

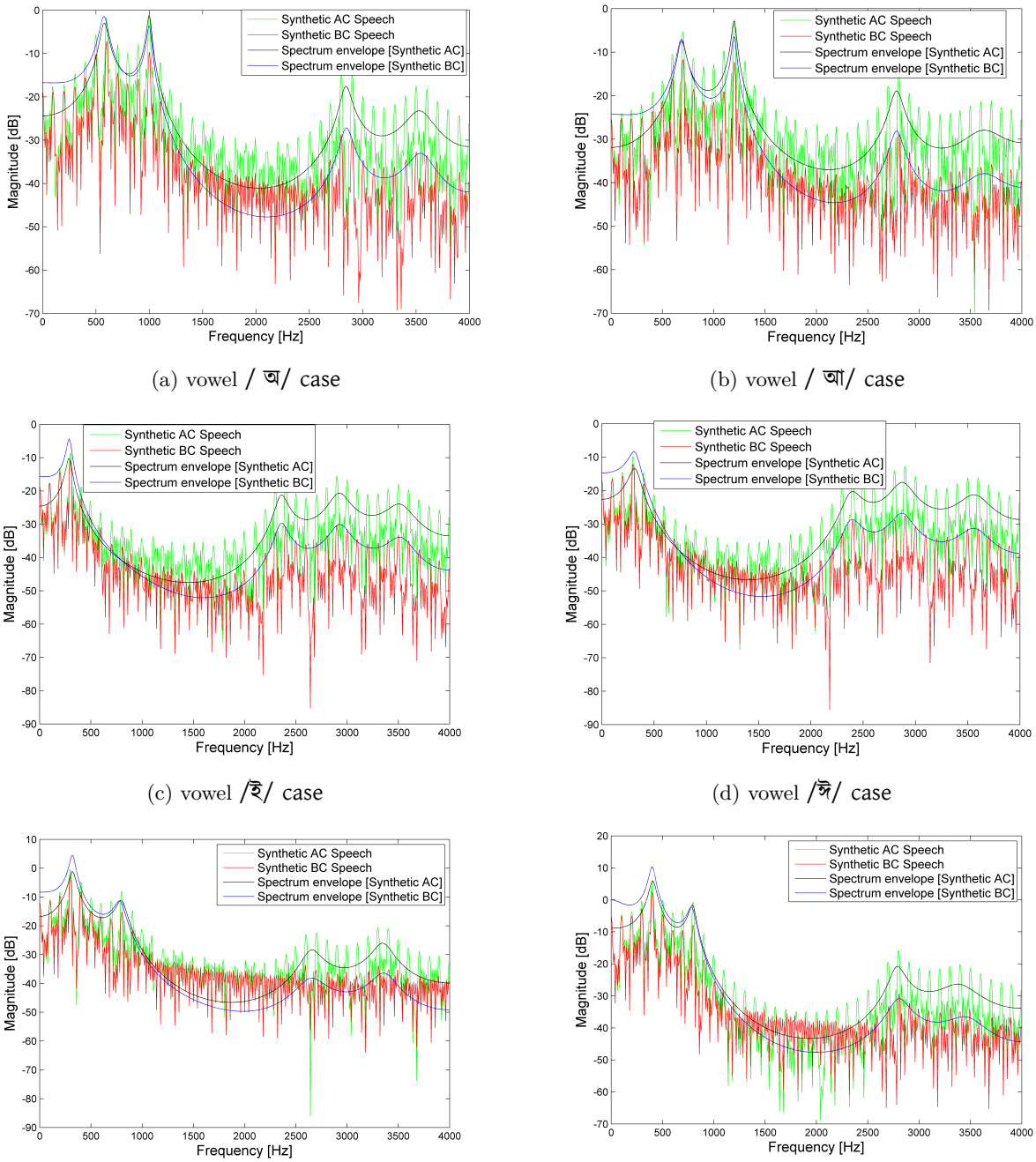
References

1. M. S. Hosain, Y. Sugiura, M. S. Rahman and T. Shimamura, EmoBone: A Multinational Audio Dataset of Emotional Bone Conducted Speech, IEEJ Transactions on Electrical and Electronic Engineering, vol. 19, no. 9, pp.1492-1506, https://doi.org/10.1002/tee.24110 (2024).

2. Ohidujjaman, M. Hasan, S. Ahmmed, M. N. Huda, M. S. Uddin, Restoration of Speech Packet Estimating Residue of Burg Linear Prediction, Results in Engineering, Elsevier, vol. 28, https://doi.org/10.1016/j.rineng.2025.107368 (2025).

3. M. S. Hosain, Y. Sugiura, N. Yasui, T. Shimamura, Deep-Learning-Based Speech Emotion Recognition Using Synthetic Bone-Conducted Speech, Journal of Signal Processing 27 (6), 151-163, https://doi.org/10.2299/jsp.27.151 (2023).

4. K. Amino, T. Osanai, T. Kamada, H. Makinae, T. Arai, "Effects of the Phonological Contents and Transmission Channels on Forensic Speaker Recognition," Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism (A.Neustein, H. Patil eds.), Springer-Verlag Berlin Heidelberg, pp. 275-308, https://doi.org/10.1007/978-1-4614-0263-3_11 (2011).



(a) vowel / অ/ case

(b) vowel / আ/ case

(c) vowel /ই/ case

(d) vowel /ঈ/ case

(e) vowel /এ/ case

(f) vowel /ঐ/ case

Figure 7: Power spectra envelopes of the synthetic AC and BC vowels for (a) vowel /অ/ case, (b) vowel /আ/ case, (c) vowel /ই/ case, (d) vowel /ঈ/ case, (e) vowel /এ/ case, (f) vowel /ঐ/ case.

5. M. S. Rahman, T. Shimamura, Pitch Determination from Bone Conducted Speech, IEICE TRANS. INF. & SYST., Vol. E99–D, No.1, https://doi.org/10.1587/transinf.2015EDL8134 (2016).

6. T. T. Vu, K. Kimura, M. Unoki, M. Akagi, A Study on Restoration of Bone-Conducted Speech with MTF-Based and LP-Based Models, Journal of Signal Processing, Vol. 10, No. 6, pp. 407-417, http://hdl.handle.net/10119/4015 (2006).

7. Ohidujjaman, Y. Nozomiko, S. Yosuke, S. Tetsuya, M. Hisanori, Packet Loss Concealment Estimating Residual Errors of Forward-Backward Linear Prediction for Bone-Conducted Speech, International Journal of Advanced Computer Science & Applications, vol. 15, no.4, pp. 1263-1268, https://dx.doi.org/10.14569/IJACSA.2024.01504126 (2024).

8. M. Hasan, Ohidujjaman, M.N. Huda, Enhancement of speech signal by originating computational iteration using SAF, 2014 IEEE International Symposium on Signal Processing and Information Technology, https://doi.org/10.1109/ISSPIT.2014.7300561 (2014).

9. ITU-T Recommendation G.711, Pulse code modulation (PCM) of voice frequencies,https://www.itu.int/rec/t-rec-g.711-198811-i/en (1988).

10. Ohidujjaman, M. Hasan, S. Zhang, M. N. Huda, M. S. Uddin, Ill-condition enhancement for BC speech using RMC method, International Journal of Speech Technology, vol.27, no. 4, pp.1085-1092, https://doi.org/10.1007/s10772-024-10159-9 (2024).

11. S. Zhang, Y. Sugiura and T. Shimamura, Bone-conducted speech synthesis based on least squares method, IEEJ Trans. on Electr. and Electron. Eng., Vol. 17, No. 3, pp. 425 –435, https://doi.org/10.1002/tee.23531 (2022).

12. J. Lecomte, T. Backstrom, Signals and Communication Technology, Chapter 12: Packet loss and concealment, Springer, pp. 173-197, https://doi.org/10.1007/978-3-319-50204-5_12 (2017).

13. Ohidujjaman, N. Yasui, Y. Sugiura, T. Shimamura, and H. Makinae, Packet loss compensation for VoIP through bone-conducted speech using modified linear prediction, IEEJ Transactions on Electrical and Electronic Engineering, vol. 18, no. 11, https://doi.org/10.1002/tee.23907 (2023).

14. K.K. Paliwal, P. V. S. Rao, A modified autocorrelation method of linear prediction for pitch-synchronous analysis of voiced speech, Signal Processing, vol. 3, no. 2, pp. 181-185,https://doi.org/10.1016/0165-1684(81)90082-7 (1981).

15. S. Zhang, Y. Sugiura, N. Yasui, T. Shimamura, Air-Conducted and Bone-Conducted Speeches Combination for Noise-Robust Pitch Extraction, IEEJ Trans. on Electr. and Electron. Eng., vol. 17 no. 7, 1061–1071, https://doi.org/10.1002/tee.23596 (2022).

16. M. A. Rahman, T. Shimamura, Linear-Prediction-Based Accurate Spectrum Estimation with Pitch Extension for Bone-Conducted Speech, Journal of Signal Processing, Vol. 22, No.6, pp.277-286, https://doi.org/10.2299/jsp.22.277 (2018).

17. Ohidujjaman, B. Munshi, M. Hasan, S. Ahmmed, and T. Shimamura, BengalVowelData: True AC and Synthetic BC Bangla Vowel Datasets in Speech Information System, Mendeley Data V4,https://doi:10.17632/ywfx54ktvn.4 (2025).

18. J. Makhoul, Linear prediction: A tutorial review, Proc. IEEE, Vol. 63, No. 4, pp. 561-580, https://doi.org/10.1109/PROC.1975.9792 (1975).

19. P. Kabal, Ill-Conditioning and Bandwidth Expansion in Linear Prediction of Speech. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong, China, 6–10, pp. 824–827, https://doi.org/10.1109/ICASSP.2003.1198908 (2003).

20. M. A. Rahman, Y. Sugiura, and T. Shimamura, Spectrum compensation method for speech signals based on prediction error filtering, WSEAS Trans. Syst. and Control, Vol. 12, pp. 213 –220, https://wseas.com/journals/articles.php?id=2808 (2017).

21. M. A. Rahman, Y. Sugiura, T. Shimamura, Accurate power spectrum estimation of speech with spectrum compensation based on prediction error filtering. WSEAS Transactions on Signal Processing, vol. 13, 21–25, https://wseas.com/journals/articles.php?id=3161 (2017).

Acknowledgements

We sincerely thank United International University (UIU) for using Multimedia Lab (706). We also appreciate the undergraduate students who delivered speeches for this work.

Author Contributions

Conceptualization and methodology, O. and B. M.; Writing-original draft, O.; Writing-review and editing, O., B. M. and M. H.; Formal analysis, data curation and validation, O., M. H., M. N. H. and S. A.; Investigation, visualization, and supervision, S. A., H. S. and T. S. All authors have read and approved the final version of the manuscript.

Ethics approval

Institutional Research Ethics Board of United International University (UIU) approved the data collection and this study, IREB Reference No.: IREB-2025-040.

Funding

The Institute for Advanced Research (IAR) Publication Grant of United International University has funded this research, Ref. No.: IAR-2025-Pub-000.

Competing Interests

The authors declare no conflicts of interest.