# Behavioral Pattern Analysis between Bilingual and Monolingual Listeners' Natural Speech Perception on Foreign-Accented English Language Using Different Machine Learning Approaches

**Md Tanvir Ahad** [1,*] , **Md Manjurul Ahsan** [2,*] , **Ishrat Jahan** [3], **Redwan Nazim** [4], **Munshi Md. Shafwat Yazdan** [5] , **Pedro Huebner** [2] **and Zahed Siddique** [1]

1   School of Aerospace & Mechanical Engineering, University of Oklahoma, Norman, OK 73019, USA; zsiddique@ou.edu
2   School of Industrial & System Engineering, University of Oklahoma, Norman, OK 73019, USA; pedro@ou.edu
3   Computer Science Engineering, East West University, Dhaka 1212, Bangladesh; arnajahan00@gmail.com
4   School of Chemical, Biological and Materials Engineering, University of Oklahoma, Norman, OK 73019, USA; redwan.nazim-1@ou.edu
5   Civil and Environmental Engineering, Idaho State University, Pocatello, ID 83209, USA; yazdmuns@isu.edu
*   Correspondence: Md.Tanvir.Ahad-1@ou.edu (M.T.A.); ahsan@ou.edu (M.M.A.);
    Tel.: +1-409-678-0082 (M.M.A.)

**Abstract:** Speech perception in an adverse background/noisy environment is a complex and challenging human process, which is made even more complicated in foreign-accented language for bilingual and monolingual individuals. Listeners who have difficulties in hearing are affected most by such a situation. Despite considerable efforts, the increase in speech intelligibility in noise remains elusive. Considering this opportunity, this study investigates Bengali–English bilinguals and native American English monolinguals' behavioral patterns on foreign-accented English language considering bubble noise, gaussian or white noise, and quiet sound level. Twelve regular hearing participants (Six Bengali–English bilinguals and Six Native American English monolinguals) joined in this study. Statistical computation shows that speech with different noise has a significant effect ($p = 0.009$) on listening for both bilingual and monolingual under different sound levels (e.g., 55 dB, 65 dB, and 75 dB). Here, six different machine learning approaches (Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-nearest neighbors (KNN), Naïve Bayes (NB), Classification and regression trees (CART), and Support vector machine (SVM)) are tested and evaluated to differentiate between bilingual and monolingual individuals from their behavioral patterns in both noisy and quiet environments. Results show that most optimal performances were observed using LDA by successfully differentiating between bilingual and monolingual 60% of the time. A deep neural network-based model is proposed to improve this measure further and achieved an accuracy of nearly 100% in successfully differentiating between bilingual and monolingual individuals.

**Keywords:** bilingual; monolingual; artificial intelligence; foreign accent; natural speech; language; machine learning

## 1. Introduction

Listeners have demonstrated difficulties in understanding speech when exposed to various background noise and reverberation degradation conditions [1,2]. Speech perception is a complex process during which the auditory system perceives the sound and interprets it into linguistic information. A complex interaction between the auditory system and the cognitive skills of a listener requires alternation between target speech and competing noise for speech perception in noise [3]. Background noise, the air interface between speakers, poor room acoustics, foreign accents, and reverberation are common reasons

for an individual listener's inability to recognize speech completely [4,5]. Researchers found that older listeners showed a negative effect by the spoken speech on fast rate time compression simulation, which happens due to rapid rate of speech. [6,7]. Foreign accented speech causes temporal characteristics alternation in everyday speech conditions [1]. Changes in the rhythm and tonal patterns, as well as the signal identity of consonants and vowels, can affect and influence the timing structure of the total utterance of the accented speech [1,8,9]. In quiet conditions, researchers investigated the performance of the ability of older listeners to understand accented English [10,11].

Supervised segregation (regulated facilities for different groups' race, class, or ethnicity) has shown significant enhancement of human speech intelligibility in noisy environments, with clear indication that Deep Neural Network (DNN)-based supervised speech segregation is a promising approach to new acoustic environments [12]. Healy et al. [13,14] showed improvement in the intelligibility of noisy speech. The exploration of DNN-based speech separation in noisy environments has also been presented in the literature [15–18]. However, behavioral pattern recognition (Chains of behavior indicating particular groups' foreground nature in complex segments of behavior which foist the sameness for input data, e.g., image, speech, speech rating, text refers to behavioral pattern recognition) for bilingual individual's speech perception under noisy environments have not been evaluated and explored yet. Thus, there is a need to investigate both machine learning and DNN-based behavioral pattern recognition for Bengali–English bilingual and native English speaker monolingual individuals' speech in noise (SIN) perception for foreign-accented English language under quiet and noisy environments.

Approximately 19.7% of the U.S. population speak a language other than English at home, according to the U.S. Census Bureau [19], which projected that bilingualism would continue to rise in the United States in the near future. Previous study results showed that, in 2014, the U.S Hispanic population reached 60 million, and this number is estimated to reach 106 million by 2050 [20]. This growing population diversity will lead to language diversity as well. Understanding foreign-accented speech in noise can also be more challenging for bilingual listeners. Additionally, different racial groups may be more prone to developing hearing deficiencies. For instance, non-Hispanic white male adults report more hearing loss than other racial adult groups [21]. The language background of listeners can play a vital role in speech perception in adverse acoustic conditions [22]. Therefore, studying the mechanism of speech perception by listeners of different language backgrounds may be of interest to auditory research. Research efforts show that listeners have the ability to quickly adapt to foreign-accented speech, which also improves over time [23]. Cristia et al. (2012) [23] compared the neuronal response (in the form of EEG) of normal-hearing individuals to both foreign-accented and native-accented speech. They monitored brain activity through the EEG and suggested that the brain may respond differently to different accents [24]. Tabri et al. conducted an experiment on English speech perception in quiet and different noise levels (50, 55, 60, 65, and 70 dB) using the speech perception in noise (SPIN) test [25]. Their results showed that the bilingual and trilingual listeners performed similarly to monolingual, but the performance declined rapidly at 65- and 70-dB SPL. Lotfi et al. [26] studied 92 individuals to evaluate the differences between Kurd–Persian bilingual versus Persian monolingual speech perception in noise. Their results demonstrated that Kurd–Persian bilinguals had a poor performance in the quick speech in noise (Q-SIN) test; however, they had a better performance on consonant–vowel in the noise (CV) test than monolingual Persians. Krizman et al. [27] investigated linguistic processing demands between Spanish–English bilingual and English monolingual to identify the performance on different task demands. Skoe et al. [28] investigated the source of difficulties experienced by English proficient bilingual listeners while listening to English speech in noise and found that the performance declined with the drop of signal to noise ratio (SNR). Barbosa et al. found that, in the background noise condition, bilingual individuals make more errors than monolinguals; in addition, they found that individuals who learn English at an earlier age make fewer errors in a noisy situation [29].

Bidelman et al. [30] showed in their study that bilinguals require around 10 dB SNR more to match monolingual listeners in adverse conditions; in addition, they found that Broca's area activity does not compensate bilingual but compensate monolingual SIN perception. Other studies [31–35] investigated monolingual and bilingual listeners' speech-in-noise performance in an everyday listening environment.

Human speech intelligibility is a key research topic exploring and analyzing various subjects, such as acoustics engineering, audiometry, phonetics, and human factors. In the twenty-first century, with the increase in bilingualism, it is critical to assess the challenges faced by monolingual and bilingual individuals during communication in noisy acoustic environments to improve speech intelligibility. The development of fine-tuned automatic Artificial Intelligence (AI)-based hearing aids for Hearing Impaired (HI) individuals will be a successful contribution to increase speech intelligibility in adverse acoustics conditions. Therefore, there is a distinctive variety of elements (e.g., bilingualism, language, foreign accent, behavioral pattern) that needs to be considered. This study investigates the question regarding the effects of foreign accent on speech between Bengali–English bilingual and native American accent English listeners, specifically: (1) Does human behavior show any significant difference on foreign accent language under quiet or adverse noisy environment? and (2) How Bengali–English bilingual and native American English speakers show significance under quiet and adverse condition? The overall purpose of this study is to investigate the significant difference between Bengali–English bilingual and native American English monolinguals effects of a talker's accent in (quiet and noise) listening conditions to predict behavioral pattern recognition using Artificial Intelligence (AI).

## 2. Related Work

Intelligibility is designated by a listener's experience and accuracy in decoding the acoustic signal of a speaker. Assessing a listener's intelligibility has been practiced clinically over the years. To assess the speech intelligibility reception, a handful of detection applications have been introduced already, such as an automatic intelligibility detection system. An object is distinguished by a set of features or variables to a class denoted as a classification task [36]. The applications of the classification task in daily human activities are wide [37,38]. Classification methods have been used to classify speech intelligibility in the context of recognition or detection. This is a binary classification problem. Artificial intelligence, fuzzy logic, statistical, and the formal way of classification have been used in many recognitions or detection problems. The classification methods in speech recognition or intelligibility applications have been explored by many research groups. Fook et al. [39] carried out an experiment for the classification prolongations and repetitions among speakers using the Support Vector Machine (SVM) algorithm. Classification of speech intelligibility of Parkinsonian speakers using SVM has been explored by Khan et al. [38]. Using NKI CCRT and the TORGO database with the help of SVM, LDA, and k-NN classifier, Kim et al. [40] showed the effort in impaired speech to classify pronunciation and voice quality. Elfahal et al. [41] examined the automatic recognition system for mixed Sudanese Arabic–English Languages Speech. For the Ngiemboon language, Yemmene et al. [42] explored various characteristics of a deep learning-based automatic speech recognition system. Automatic classification of speech intelligibility for listener's using Long Short-Term Memory based system was proposed by Miguel et al. [43]. Listening effort during sentence processing has been explored by Borghini et al. [44]. In addition, several research efforts have been published showing the Deep Neural Network (DNN) for listener's speech recognition and ineligibility applications [45–48]. Based on the available literature and author's best knowledge, binary classification or recognition of bilingual and monolingual listener's speech ineligibility reception have not been reported yet.

### 3. Data Acquisition and Methods

*3.1. Data Acquisition*

Data available from the literature [49] were used in this study. Data were collected at the Applied DSP Research Laboratory of Lamar University, Beaumont, Texas, USA. Participants included eighteen college student volunteers between the ages of 20 to 27. Six native English speakers and six Bengali–English bilinguals formed two mutually exclusive experimental groups. It was verified (confirmed by the LU Speech and Hearing clinic) that all subjects had normal hearing.

Short duration (10–12 s) audio fragments spoken by adult British English speakers (male and female) were used as the speech stimuli. The recordings were obtained from the free online depository http://listentogenius.com/ (accessed on 19 January 2018). Speech fragments were delivered at three sound levels: 55 dB, 65 dB, and 75 dB. Some fragments were contaminated by either Gaussian or bubble noise at the same three sound levels to produce various signal-to-noise ratios of −10 dB, 0 dB, 10 dB, and infinity (no noise). Stimuli were delivered diotically to participants using Etymotic insert earphones at a variety of sound levels. One hundred twenty audio stimuli were presented in total in a randomized order with 2 s of silence between them.

Experimental details consisted of continuous EEG recording and behavioral data. Additionally, participants were asked to provide their subjective evaluations regarding the quality of the audio fragments that they listened to. For that purpose, the same randomized sequence of 120 audio fragments was used. The quality was evaluated on a 1 to 10 scale where 1 corresponded to "inferior" and 10 represented "excellent" quality.

The primary purpose of the survey was to understand the participant's experience on a different kind of speech with different types of frequency considering bubble noise, white noise, and quiet sound level environment.

*3.2. Methodology*

The experiment was conducted among 12 participants of native English speaker monolingual and Bengali–English speaker bilingual individuals. Since the monolingual individuals represented a higher percentage of the participant population, systematic sampling techniques were used to choose 6 participants from the total pool of participants at regular intervals. The audio fragments contain a total of 120 questions against 120 types of speech with different sound (bubble noise, white noise, and quiet) levels. However, among 120 samples, 25 speech samples were with a quiet condition, which was the lowest number of samples compared to bubble and white noise sound speech. Thus, 25 samples were chosen from each group to conduct further statistical analysis. Table 1 shows the demographic characteristics of the subject with the mean value from 120 audio fragments.

**Table 1.** Mean value of 12 subjects on a different sound level (55 dB, 65 dB, and 75 dB).

| Participants | Bubble Noise | White Noise | Quiet level |
| --- | --- | --- | --- |
| Monolingual | 5.4 | 4.2 | 9.94 |
| Monolingual | 3.875 | 3.2 | 6.4 |
| Monolingual | 4.67 | 3.68 | 9.1 |
| Monolingual | 4.45 | 3.8 | 9.58 |
| Monolingual | 4.25 | 2.6 | 6.83 |
| Monolingual | 6.17 | 7.76 | 6.54 |
| Bilingual | 7.79 | 7.64 | 7.875 |
| Bilingual | 4.29 | 3.56 | 3.16 |
| Bilingual | 5.25 | 6.24 | 5.45 |
| Bilingual | 5.41 | 5 | 4.16 |
| Bilingual | 5.79 | 6.4 | 5.5 |
| Bilingual | 6.625 | 6.28 | 6.04 |

MANOVA was used to test for group differences/variances on two or more dependent variables. This experiment considered the following independent and dependent variables for MANOVA analysis:

- Independent variables: Language (monolingual, bilingual)—2 factors
- Dependent variables: Speech sound (quiet, white noise, bubble noise)—3 factors.

## 4. Results

There was a significant difference between bilinguals and monolinguals considered jointly on the variables bubble noise, white noise, and quiet speech, Wilk's η = 0.256, F (3,8), P (Significant) = 0.009, partial eta square = 0.744. A distinct ANOVA was conducted for each dependent variable, with each ANOVA evaluated at an alpha level of 0.016. It did not show any significant difference separately on monolingual and bilingual individuals for the Multivariate test, refers to Table 2.

**Table 2.** Multivariate Test.

|  | Effect | Value | F | Hypothesis df | Error df | Sig. |
|---|---|---|---|---|---|---|
| Intercept | Pillai's Trace | 0.974 | 101.331 [b] | 3.000 | 8.000 | 0.000 |
|  | Wilks' Lambda | 0.026 | 101.331 [b] | 3.000 | 8.000 | 0.000 |
|  | Hotelling's Trace | 37.999 | 101.331 [b] | 3.000 | 8.000 | 0.000 |
|  | Roy's Largest Root | 37.999 | 101.331 [b] | 3.000 | 8.000 | 0.000 |
| Language | Pillai's Trace | 0.744 | 7.740 [b] | 3.000 | 8.000 | 0.009 |
|  | Wilks' Lambda | 0.256 | 7.740 [b] | 3.000 | 8.000 | 0.009 |
|  | Hotelling's Trace | 2.903 | 7.740 [b] | 3.000 | 8.000 | 0.009 |
|  | Roy's Largest Root | 2.903 | 7.740 [b] | 3.000 | 8.000 | 0.009 |

[b] indicates each of the four types of testable hypothesis result is not unique.

### 4.1. Correlation Analysis

In order to determine whether there is any correlation among all three sound levels based on the user experience, a co-relation analysis was conducted using "Pearson Correlation". Table 3 presents the correlation between bubble noise, white noise, and quiet speech sound. There was a significant negative relationship between bubble and white noise, r (10) = 0.883, *p* = 0.000, as shown in Table 3.

**Table 3.** Correlation between noisy and quiet sound level.

|  |  | Bubble Noise Speech | Gaussian or White Noise Speech | Quiet Speech |
|---|---|---|---|---|
| Bubble Noise Speech | Pearson Correlation | 1 | 0.883 ** | 0.051 |
|  | Sig. (2-tailed) |  | 0.000 | 0.876 |
|  | N | 12 | 12 | 12 |
| Gaussian or White noise Speech | Pearson Correlation | 0.883 ** | 1 | −0.130 |
|  | Sig. (2-tailed) | 0.000 |  | 0.688 |
|  | N | 12 | 12 | 12 |
| Quiet Speech | Pearson Correlation | 0.051 | −0.130 | 1 |
|  | Sig. (2-tailed) | 0.876 | 0.688 |  |
|  | N | 12 | 12 | 12 |

** Correlation is significant at the 0.01 level (2-tailed).

### 4.2. Machine Learning Algorithm

Another study was conducted using different machine learning techniques on survey data of 12 participants. Different machine learning algorithms were selected for the study as follows: Logistic Regression (LR), Linear Discriminant Analysis (LDA), K Nearest Neighbor

(KNN), Gaussian Naïve Bayes (NB), Classification and Regression Trees (CART), and Support Vector Machine (SVM), with default parameters. The whole experiment was carried out using Scikit learn tools with Python interpreter language. To evaluate the performance, fivefold cross-validation was utilized, and the results are presented by averaging (avg.) those five folds.

Table 4 present a summary of the performance of all the algorithms on survey data.

**Table 4.** Algorithm performance on survey data.

| Algorithm | Accuracy (Avg) | Std |
|:---:|:---:|:---:|
| LR | 0.30 | 0.24 |
| LDA | 0.50 | 0.44 |
| KNN | 0.50 | 0.44 |
| NB | 0.50 | 0.316 |
| CART | 0.30 | 0.4 |
| SVM | 0.20 | 0.244 |

LDA, KNN, and NB showed the best performance by achieving a constant accuracy of around 50%. Contrarily, SVM showed the worst performance across all measures.

Note that, in this first experiment, the performance of all the machine learning algorithms was significantly low. To improve the existing computational performance, another experiment was carried out by standardizing the dataset. After standardizing the dataset, some improvement was observed on the performance of those algorithms.

Table 5 summarizes the overall performance of algorithms after scaling the dataset. The noticeable changes were observed once the overall accuracy of LR increased from 30% (Table 4) to 60% (Table 5), still which is not up to the mark as a final result of general data analysis.

**Table 5.** Algorithm performance after Scaling the dataset.

| Algorithm (Scaled) | Accuracy (Avg) | Std |
|:---:|:---:|:---:|
| LR | 0.60 | 0.37 |
| LDA | 0.50 | 0.44 |
| KNN | 0.50 | 0.44 |
| CART | 0.20 | 0.24 |
| NB | 0.50 | 0.32 |
| SVM | 0.30 | 0.244 |

In Figure 1, a clustered bar chart was used to compare the performance of six machine learning algorithms in terms of data standardization. The LR method showed the highest accuracy improvement among all the different algorithms, from 30% to 60%. Additionally, the CART machine learning algorithm's performance significantly decreased by up to 66% (from 30% to 20%).

Since most of the machine learning algorithm's performance was significantly low on the dataset, another experiment was conducted using a deep learning approach.

*4.3. Behavioral Pattern Recognition Using a Deep Learning Approach*

To develop a neural network model, Keras Python library was used. It is a Python library that can run on top of Theaona or Tensorflow.
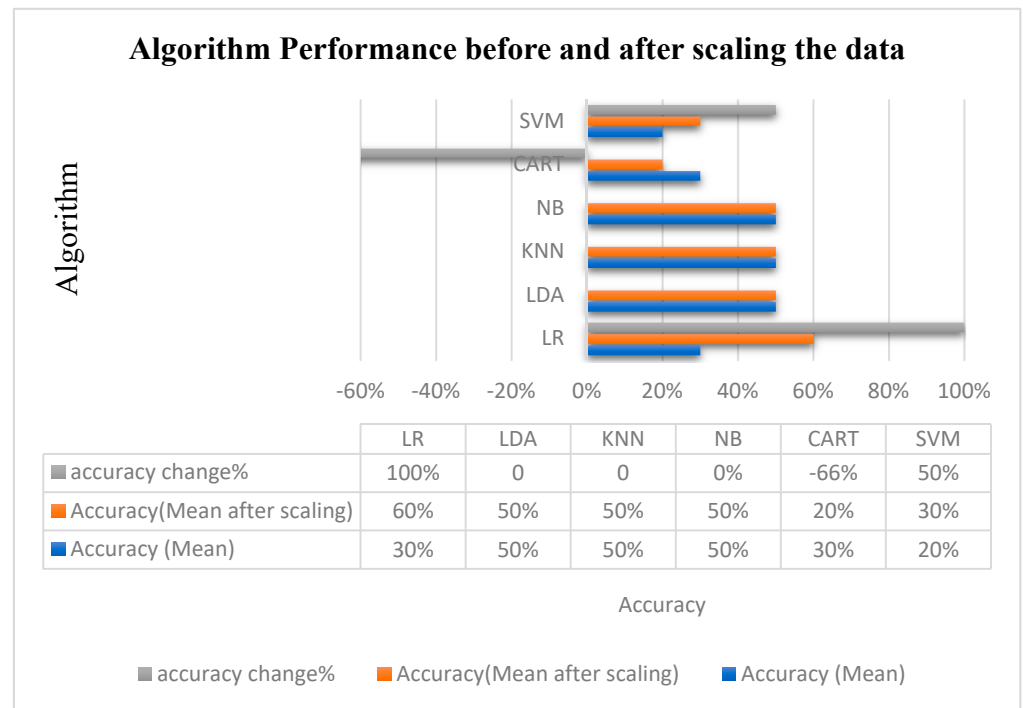
**Figure 1.** Algorithm performance before and after data standardization.

Proposed Model

   A sequential model was created, and some additional layers were also added until a significant amount of improvement was observed during the training phase. One hundred and twenty input variables were used as the data set contained 120 input parameters. The most optimal network was chosen after several trials with random input features. Note that the defined neural network was a fully connected layer using Dense Class. More details on how a deep learning architecture is developed can be found here [50]. Figure 2 shows the architecture of the network:
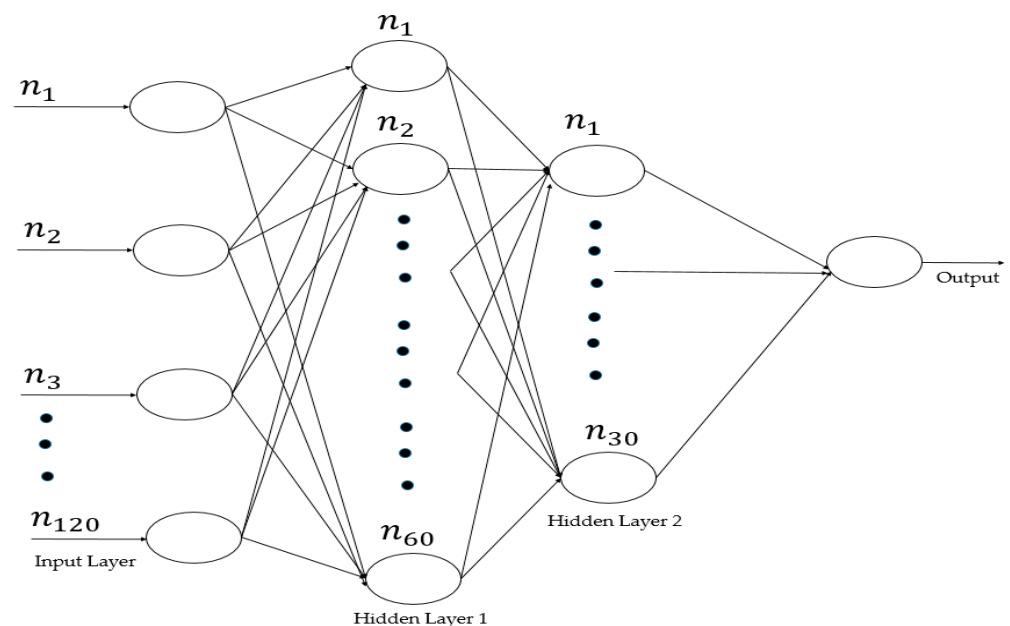


**Figure 2.** Proposed neural network architecture.

Figure 2 shows that the network was initiated by 120 inputs, and two hidden layers contained 60 and 30 neurons, respectively. To initialize the network, an activation function was necessary, and here, the network utilized the rectifier activation function on the first three layers and the sigmoid activation function as the output layer.

The sigmoid activation function was used to ensure the network output would remain between 0 and 1 since the network was designed for binary classification. Details regarding "Relu" and "Sigmoid" can be found at [51]. Note that training a network means finding the right set of weights to make a better prediction. Thus, it is necessary to specify the loss function to evaluate a set of weights. In this case, the logarithmic loss was used, which is defined in Keras as "binary_crossentropy". An adaptive learning rate optimization algorithm (Adam) was used as an optimization algorithm due to its robust performance on binary classification. More details regarding the 'Adam' optimizer can be found in [52].

The training process runs for a fixed number of iterations through the dataset called epochs which need to be specified while fitting the model. Here 150 epochs were used with a batch size of 10. Note that the batch size and the epochs were chosen experimentally by trial and error. While training the model, each iteration adjusted the loss to the next epoch. During this experiment, after 35 epochs, the accuracy reached 100%, while the loss recorded was only 76%. To understand the network performance, training loss, validation loss, training accuracy, and validation accuracy were also calculated. Figure 3 shows the training and validation loss, as well as accuracy.



(**a**) Training and validation loss　　　　　　　　(**b**) Training and validation accuracy
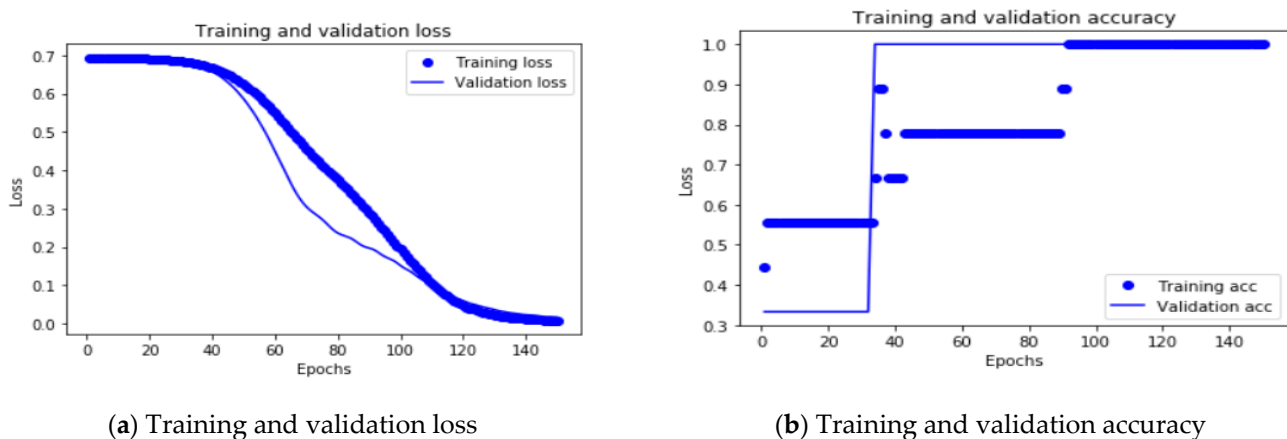
**Figure 3.** Performance measure during the training phase (**a**) training and validation loss, (**b**) training and validation accuracy.

During the training phase, both the training and validation loss curve touched at about 120 epochs, and it was decided that no further training was required after that point (Figure 3a). On the other hand, both training and validation curves in Figure 3b show some discrepancies during 35 epochs. Considering both Figure 3a,b, it is possible to assume that the proposed model performed well on "EEG_data_lamar" with an accuracy of 100%.

## 5. Discussion

As a means of understanding the effects of foreign accents on speech between Bengali–English bilingual and native American English listeners, the study observed 12 participant's behaviors under bubble noise, white noise, and quiet speech sound level environments. The results showed a significant difference ($p = 0.009$) between the two groups' (bilinguals and monolinguals) behavior under various noisy conditions. Additionally, the behavioral performance was analyzed with different machine learning approaches, such as LR, LDA, KNN, CART, NB, and SVM. The resulting analysis showed it was possible to differentiate between two groups 60% accurately using LR. Hence, a small deep neural network (DNN) was proposed, which achieved 100% accuracy in differentiating between bilinguals and monolinguals. It is relevant to emphasize that none of the reference studies considered the effects of noisy environment between two distinct groups—the bilinguals and

monolinguals—using machine learning/deep learning-based approach, which hinders the opportunity of a direct comparison with the existing literature. Therefore, this study may help researchers and practitioners in the near future to evaluate the effect of noise on multilingual individuals. Apart from the aforementioned advantages, this study also has some limitations which shall be addressed in future projects:

- During this study, only a limited number of individuals (12 participants) were considered.
- We did not consider other widely bilingual people who speak English–Arabic, Hindi–English who need to be taken into account for the proper evaluation of the effect of noise on bilingual people on a large scale.
- The performance of the proposed deep neural network may fluctuate when applied to a larger data set.

### 6. Conclusions

This study evaluated the participant's experience on a foreign-accented speech with different types of frequency considering bubble noise, white noise, and quiet speech sound level (e.g., sound levels: 55 dB, 65 dB, and 75 dB) and with signal-to-noise ratios of $-10$ dB, 0 dB, 10 dB, and infinity (no noise) environments between bilingual and monolingual individuals. The study focused on young adults. The findings suggest that foreign-accented speech with different noise has a significant effect on listening regardless of whether the person is bilingual or monolingual. A significant difference was also observed between the two groups in quiet and white noise-contaminated speech; however, no such significant difference was measured under bubble noise-contaminated speech. This indicates that the performance of listening will be mostly similar regardless of one's multi-linguistic capabilities. It seems that no additional advantages are enjoyed through the comprehension of multiple languages. Finally, we tested and evaluated six different machine learning algorithms on the 12 participant's dataset in terms of speech quality ratings in mild-to-moderately by listeners, and higher accuracy was achieved using LDA—60%, after data standardization. Speech quality ratings by monolingual and bilingual listeners observed were somewhat confounded because of the ineligibility. In addition to this, a deep neural network was developed that differentiated between bilingual and monolingual participants by achieving an accuracy of 100%. Some of the limitations associated with this work can be addressed by conducting experiments with large and imbalance datasets [53,54], comparing the performance of the proposed methods with other bilingual participants, and explaining the analytic results using explainable AI [55].

**Author Contributions:** Conceptualization, M.T.A. and M.M.A.; methodology, M.T.A. and M.M.A.; software, M.T.A. and M.M.A.; validation, M.T.A., M.M.A., I.J., R.N. and M.M.S.Y.; formal analysis, M.T.A. and M.M.A.; investigation, M.T.A., M.M.A., I.J., R.N. and M.M.S.Y.; resources, M.T.A., M.M.A., I.J., R.N. and M.M.S.Y.; data curation, M.T.A. and M.M.A.; writing—original draft preparation, M.T.A., M.M.A., I.J., R.N. and M.M.S.Y.; writing—review and editing, M.T.A., M.M.A., I.J., R.N. and M.M.S.Y.; visualization, M.T.A. and M.M.A.; supervision, P.H., Z.S.; project administration, M.T.A., M.M.A. and Z.S.; funding acquisition, Z.S. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Lamar University.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not Applicable.

**Conflicts of Interest:** Authors declare no conflict of interest.

# References

1. Gordon-Salant, S.; Yeni-Komshian, G.H.; Fitzgibbons, P.J. Recognition of Accented English in Quiet and Noise by Younger and Older Listeners. *J. Acoust. Soc. Am.* **2010**, *128*, 3152–3160. [CrossRef]
2. Nábělek, A.K.; Robinson, P.K. Monaural and Binaural Speech Perception in Reverberation for Listeners of Various Ages. *J. Acoust. Soc. Am.* **1982**, *71*, 1242–1248. [CrossRef] [PubMed]
3. Arbab, H.; Moossavi, A.; Javanbakht, M.; Arbab Sarjoo, H.; Bakhsh, E.; MahmoodiBakhtiari, B.; Lotfi, Y. Development and Psychometric Evaluation of Persian Version of the Quick Speech in Noise Test in Persian Speaking 18–25 Years Old Normal Adults. *J. Rehabil. Sci. Res.* **2016**, *3*, 51–56. [CrossRef]
4. Crandell, C.C.; Smaldino, J.J. Classroom Acoustics for Children with Normal Hearing and with Hearing Impairment. *Lang. Speech Hear. Serv. Sch.* **2000**, *31*, 362–370. [CrossRef] [PubMed]
5. Nabelek, A.K.; Mason, D. Effect of Noise and Reverberation on Binaural and Monaural Word Identification by Subjects with Various Audiograms. *J. Speech Lang. Hear. Res.* **1981**, *24*, 375–383. [CrossRef] [PubMed]
6. Gordon-Salant, S.; Fitzgibbons, P.J. Temporal Factors and Speech Recognition Performance in Young and Elderly Listeners. *J. Speech Lang. Hear. Res.* **1993**, *36*, 1276–1285. [CrossRef]
7. Ferguson, S.H.; Jongman, A.; Sereno, J.A.; Keum, K. Intelligibility of Foreign-Accented Speech for Older Adults with and without Hearing Loss. *J. Am. Acad. Audiol.* **2010**, *21*, 153–162. [CrossRef] [PubMed]
8. Fox, R.A.; Flege, J.E.; Munro, M.J. The Perception of English and Spanish Vowels by Native English and Spanish Listeners: A Multidimensional Scaling Analysis. *J. Acoust. Soc. Am.* **1995**, *97*, 2540–2551. [CrossRef] [PubMed]
9. MacKay, I.R.; Flege, J.E.; Piske, T. Persistent Errors in the Perception and Production of Word-Initial English Stop Consonants by Native Speakers of Italian. *J. Acoust. Soc. Am.* **2000**, *107*, 2802. [CrossRef]
10. Burda, A.N.; Hageman, C.F.; Scherz, J.A.; Edwards, H.T. Age and Understanding Speakers with Spanish or Taiwanese Accents. *Percept. Mot. Ski.* **2003**, *97*, 11–20. [CrossRef]
11. Gordon-Salant, S.; Yeni-Komshian, G.H.; Fitzgibbons, P.J. Recognition of Accented English in Quiet by Younger Normal-Hearing Listeners and Older Listeners with Normal-Hearing and Hearing Loss. *J. Acoust. Soc. Am.* **2010**, *128*, 444–455. [CrossRef]
12. Chen, J.; Wang, Y.; Yoho, S.E.; Wang, D.; Healy, E.W. Large-Scale Training to Increase Speech Intelligibility for Hearing-Impaired Listeners in Novel Noises. *J. Acoust. Soc. Am.* **2016**, *139*, 2604–2612. [CrossRef]
13. Healy, E.W.; Yoho, S.E.; Wang, Y.; Wang, D. An Algorithm to Improve Speech Recognition in Noise for Hearing-Impaired Listeners. *J. Acoust. Soc. Am.* **2013**, *134*, 3029–3038. [CrossRef] [PubMed]
14. Healy, E.W.; Yoho, S.E.; Chen, J.; Wang, Y.; Wang, D. An Algorithm to Increase Speech Intelligibility for Hearing-Impaired Listeners in Novel Segments of the Same Noise Type. *J. Acoust. Soc. Am.* **2015**, *138*, 1660–1669. [CrossRef] [PubMed]
15. Kim, G.; Lu, Y.; Hu, Y.; Loizou, P.C. An Algorithm That Improves Speech Intelligibility in Noise for Normal-Hearing Listeners. *J. Acoust. Soc. Am.* **2009**, *126*, 1486–1494. [CrossRef]
16. May, T.; Dau, T. Requirements for the Evaluation of Computational Speech Segregation Systems. *J. Acoust. Soc. Am.* **2014**, *136*, EL398–EL404. [CrossRef]
17. Wang, D.; Chen, J. Supervised Speech Separation Based on Deep Learning: An Overview. *IEEE ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 1702–1726. [CrossRef]
18. Chen, J.; Wang, D. Dnn based mask estimation for supervised speech separation. In *Audio Source Separation*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 207–235.
19. Shin, H.B.; Kominski, R. *Language Use in the United States, 2007*; US Department of Commerce, Economics and Statistics Administration, USA Census Bureau: Suitland, MD, USA, 2010.
20. Krogstad, J.M. With Fewer New Arrivals, Census Lowers Hispanic Population Projections. *Pew Res. Cent.* **2014**, *16*.
21. Hoffman, H.J.; Dobie, R.A.; Losonczy, K.G.; Themann, C.L.; Flamme, G.A. Declining Prevalence of Hearing Loss in US Adults Aged 20 to 69 Years. *JAMA Otolaryngol. Head Neck Surg.* **2017**, *143*, 274–285. [CrossRef] [PubMed]
22. Takata, Y.; Nábělek, A.K. English Consonant Recognition in Noise and in Reverberation by Japanese and American Listeners. *J. Acoust. Soc. Am.* **1990**, *88*, 663–666. [CrossRef] [PubMed]
23. Cristia, A.; Seidl, A.; Vaughn, C.; Schmale, R.; Bradlow, A.; Floccia, C. Linguistic Processing of Accented Speech Across the Lifespan. *Front. Psychol.* **2012**, *3*. [CrossRef] [PubMed]
24. Grey, S.; van Hell, J.G. Foreign-Accented Speaker Identity Affects Neural Correlates of Language Comprehension. *J. Neurolinguistics* **2017**, *42*, 93–108. [CrossRef]
25. Tabri, D.; Chacra, K.M.S.A.; Pring, T. Speech Perception in Noise by Monolingual, Bilingual and Trilingual Listeners. *Int. J. Lang. Commun. Disord.* **2011**, *46*, 411–422. [CrossRef] [PubMed]
26. Lotfi, Y.; Chupani, J.; Javanbakht, M.; Bakhshi, E. Evaluation of Speech Perception in Noise in Kurd-Persian Bilinguals. *Audit. Vestib. Res.* **2019**, *28*, 36–41. [CrossRef]
27. Krizman, J.; Bradlow, A.R.; Lam, S.S.-Y.; Kraus, N. How Bilinguals Listen in Noise: Linguistic and Non-Linguistic Factors. *Biling. Lang. Cogn.* **2017**, *20*, 834–843. [CrossRef]
28. Skoe, E.; Karayanidi, K. Bilingualism and Speech Understanding in Noise: Auditory and Linguistic Factors. *J. Am. Acad. Audiol.* **2019**, *30*, 115–130. [CrossRef] [PubMed]
29. Barbosa, B.A.; Coles-White, D.; Regal, D.; Kijai, J. Analysis of Language Errors in Speakers Who Are Bilingual Under Quiet and Background Noise Conditions. *Perspect. ASHA Spec. Interest Groups* **2020**, *5*, 1687–1697. [CrossRef]

30. Bidelman, G.M.; Dexter, L. Bilinguals at the "Cocktail Party": Dissociable Neural Activity in Auditory–Linguistic Brain Regions Reveals Neurobiological Basis for Nonnative Listeners' Speech-in-Noise Recognition Deficits. *Brain Lang.* **2015**, *143*, 32–41. [CrossRef]
31. Skoe, E. Turn up the Volume: Speech Perception in Noise for Bilingual Listeners. *J. Acoust. Soc. Am.* **2019**, *145*, 1820. [CrossRef]
32. Schmidtke, J. The Bilingual Disadvantage in Speech Understanding in Noise Is Likely a Frequency Effect Related to Reduced Language Exposure. *Front. Psychol.* **2016**, *7*. [CrossRef] [PubMed]
33. Kuipers, J.R.; Thierry, G. Bilingualism and Increased Attention to Speech: Evidence from Event-Related Potentials. *Brain Lang.* **2015**, *149*, 27–32. [CrossRef] [PubMed]
34. Reetzke, R.; Lam, B.P.-W.; Xie, Z.; Sheng, L.; Chandrasekaran, B. Effect of Simultaneous Bilingualism on Speech Intelligibility across Different Masker Types, Modalities, and Signal-to-Noise Ratios in School-Age Children. *PLoS ONE* **2016**, *11*, e0168048. [CrossRef]
35. Marian, V.; Hayakawa, S.; Lam, T.Q.; Schroeder, S.R. Language Experience Changes Audiovisual Perception. *Brain Sci.* **2018**, *8*, 85. [CrossRef]
36. Rosdi, F.; Salim, S.S.; Mustafa, M.B. An FPN-Based Classification Method for Speech Intelligibility Detection of Children with Speech Impairments. *Soft Comput.* **2019**, *23*, 2391–2408. [CrossRef]
37. Ahsan, M.M.; Li, Y.; Zhang, J.; Ahad, M.T.; Gupta, K.D. Evaluating the Performance of Eigenface, Fisherface, and Local Binary Pattern Histogram-Based Facial Recognition Methods under Various Weather Conditions. *Technologies* **2021**, *9*, 31. [CrossRef]
38. Ahsan, M.M.; Li, Y.; Zhang, J.; Ahad, M.T.; Yazdan, M.M.S. Face Recognition in an Unconstrained and Real-Time Environment Using Novel BMC-LBPH Methods Incorporates with DJI Vision Sensor. *J. Sens. Actuator Netw.* **2020**, *9*, 54. [CrossRef]
39. Fook, C.Y.; Muthusamy, H.; Chee, L.S.; Yaacob, S.B.; Adom, A.H.B. Comparison of Speech Parameterization Techniques for the Classification of Speech Disfluencies. *Turk. J. Elec. Eng. Comp. Sci.* **2013**, *21*, 1983–1994. [CrossRef]
40. Kim, J.; Kumar, N.; Tsiartas, A.; Li, M.; Narayanan, S.S. Automatic Intelligibility Classification of Sentence-Level Pathological Speech. *Comput. Speech Lang* **2015**, *29*, 132–144. [CrossRef]
41. Elfahal, M.O.E. Supervisor, -Mohammed Elhafiz Mustafa; Co-Supervisor, -Rashid A. Saeed Automatic Recognition and Identification for Mixed Sudanese Arabic–English Languages Speech. Ph.D Thesis, Sudan University of Science & Technology, Khartoum, Sudan, 2019.
42. Yemmene, P.; Besacier, L. Motivations, Challenges, and Perspectives for the Development of an Automatic Speech Recognition System for the under-Resourced Ngiemboon Language. In Proceedings of the First International Workshop on NLP Solutions for Under Resourced Languages (NSURL 2019) Co-Located with ICNLSP 2019-Short Papers, Trento, Italy, 11–12 September 2019; pp. 59–67.
43. Fernández-Díaz, M.; Gallardo-Antolín, A. An Attention Long Short-Term Memory Based System for Automatic Classification of Speech Intelligibility. *Eng. Appl. Artif. Intell.* **2020**, *96*, 103976. [CrossRef]
44. Borghini, G.; Hazan, V. Listening Effort During Sentence Processing Is Increased for Non-Native Listeners: A Pupillometry Study. *Front. Neurosci.* **2018**, *12*. [CrossRef]
45. Wang, Y.; Wang, D. Towards Scaling Up Classification-Based Speech Separation. *IEEE Trans. Audio Speech Lang. Process.* **2013**, *21*, 1381–1390. [CrossRef]
46. Chen, J.; Wang, Y.; Wang, D. Noise Perturbation for Supervised Speech Separation. *Speech Commun.* **2016**, *78*, 1–10. [CrossRef] [PubMed]
47. Bolner, F.; Goehring, T.; Monaghan, J.; Van Dijk, B.; Wouters, J.; Bleeck, S. Speech Enhancement Based on Neural Networks Applied to Cochlear Implant Coding Strategies. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 November 2016; pp. 6520–6524.
48. Goehring, T.; Bolner, F.; Monaghan, J.J.; Van Dijk, B.; Zarowski, A.; Bleeck, S. Speech Enhancement Based on Neural Networks Improves Speech Intelligibility in Noise for Cochlear Implant Users. *Hear. Res.* **2017**, *344*, 183–194. [CrossRef]
49. Ahad, M.T. *An EEG-Based Comparative Analysis of Natural Speech Perception by Native Speakers of American English vs. Bilingual Individuals*; Lamar University-Beaumont ProQuest: Beaumont, TX, USA, 2018.
50. Keras: The Python Deep Learning API. Available online: https://keras.io/ (accessed on 1 March 2021).
51. Brownlee, J. *Deep Learning with Python: Develop Deep Learning Models on Theano and TensorFlow Using Keras*; Machine Learning Mastery, 2016. Available online: https://books.google.com.hk/books/about/Deep_Learning_With_Python.html?id=K-ipDwAAQBAJ&printsec=frontcover&source=kp_read_button&redir_esc=y#v=onepage&q&f=false (accessed on 22 July 2021).
52. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
53. Ahsan, M.M.; Alam, T.E.; Trafalis, T.; Huebner, P. Deep MLP-CNN Model Using Mixed-Data to Distinguish between COVID-19 and Non-COVID-19 Patients. *Symmetry* **2020**, *12*, 1526. [CrossRef]
54. Ahsan, M.M.; Ahad, M.T.; Soma, F.A.; Paul, S.; Chowdhury, A.; Luna, S.A.; Yazdan, M.M.S.; Rahman, A.; Siddique, Z.; Huebner, P. Detecting SARS-CoV-2 From Chest X-Ray Using Artificial Intelligence. *IEEE Access* **2021**, *9*, 35501–35513. [CrossRef]
55. Ahsan, M.M.; Gupta, K.D.; Islam, M.M.; Sen, S.; Rahman, M.L.; Shakhawat Hossain, M. COVID-19 Symptoms Detection Based on NasNetMobile with Explainable AI Using Various Imaging Modalities. *Mach. Learn. Knowl. Extr.* **2020**, *2*, 490–504. [CrossRef]