ISBN: 978-628-95207-4-3. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2023.1.1.1554

Improving Speech Classification Accuracy: A Support Vector Machine Approach

Balvin Thorpe, Ph.D.
Electrical Engineering Department
University of Technology, Jamaica
237 Old Hope Road, Kingston
Jamaica
bthorpe@utech.edu.jm

Abstract- Digital hearing aids improve hearing by reducing background noise and significantly improve sound qualities. However, digital hearing aids do not produce significant improvements for the hearing impaired, when the challenge for some hearing-impaired persons, is recognizing high frequency sounds, such as consonants. Digital hearing aids (DHAs) use parametric phoneme classifiers, such as Hidden Markov Model (HMM). These classifiers produce at best 80% phoneme classification accuracy. The main research question therefore was: - Could the digital hearing aid's phoneme classification accuracy improve, if a non-statistical/non-parametric phoneme classification algorithm was developed and used in the processor of the digital hearing aid? In this study, speech was classified using linear Support Vector Machine (SVM), a non-parametric classifier, to identify vowels differently from consonants. SVM was chosen, as literature indicate that SVM non-parametric classifier was likely to return the highest classification accuracy among all classifier algorithms. The SVM classifier was built in MATLAB by parsing the phonemes from Texas Instrument and Massachusetts Institute of Technology (TIMIT) training speech files, and generating the corresponding Mel Frequency Cepstrum Coefficients (MFCC) for each phoneme. The built SVM Classifier was tested, using files from the TIMIT speech test database. Results showed that the built SVM classifier produced phoneme classification accuracy ranging from 74% to 92.7%. These results indicate that the built SVM can be used to classify phonemes with, accuracy that is equal to or better that the existing statistical/parametric phoneme classifiers.
Keywords: Vowels, Consonants, Parametric, Non-Parametric, Phonemes, Support Vector Machine

I. INTRODUCTION

Speech is arguably the most important activity that distinguishes human from non-human species [1]. However, many problems can arise due to speech miscommunication. One method used to help persons with hearing impairment reduce speech miscommunication is the hearing aid. The problem with hearing aids however, is that they generally amplify both desired signals as well as noise. Hearing aid technology has progressed dramatically over the past 10 – 15 years. The introduction of Digital Signal Processing (DSP) into hearing aids in 1996 allowed advanced signal processing algorithms to be implemented. In 2005, 93% of the hearing aids sold in the United States contained DSP technology [2]. More than half of those hearing aids included directional microphones, providing verifiable improvements to speech understanding in noise [3]. However, with all these improvements in hearing aid technologies, hearing aids do not improve overall hearing for the hearing impaired, when the challenge for some, especially the elderly, is recognizing high frequency sounds, which typically are consonants are. Therefore, it is clear that further research into speech component processing, including hearing, is needed.

It can be said that there are still crucial, scientific and technological problems in hearing aids that have not been solved efficiently and comfortably [4]. According to [5] consonant recognition studies have shown that hearing impaired listeners make significantly more consonant recognition errors than normal hearing listener. Similarly, according to [6], normal hearing listeners typically require less contrast between spectral peaks and valley for accurate identification of vowels than do hearing impaired listeners. While the vowels create the sound volume of speech, the consonants are the bearers of information.

A fundamental distinctive unit of a spoken language is the phoneme; the phoneme is distinctive in the sense that it is a speech sound class that differentiates words of a language [7]. It is the smallest unit of language and has no inherent meaning. The English alphabet has 26 letters but approximately 44 phonemes. This means that letters combine in different ways to represent the various sounds a person can make while speaking. These sounds help to distinguish the meanings of words. There are approximately 20 vowel phonemes and 24 consonant phonemes in the American English language.

Fundamentally different acoustic cues are carried by consonants and vowels [8]. Whereas consonants are characterized by vocal tract constriction, high frequency components, and often aperiodicity, vowels are characterized by sustained voicing, lack of constriction, and dominant lower frequency structure.

The acoustic distinctiveness of vowels and consonants has been studied extensively by investigators from various fields. But to date, the distinctiveness of the different phonemes has predominantly been based on the classification of their presumed distinctive articulatory features such as lip rounding, lip opening, lip height, lip contour, and lip area [9], and tongue tip and tongue body height [10], and vocal tract shape geometry [11]. Most of these classification approaches for articulatory data, without using acoustic data, have resulted in only poor to moderate classification accuracy, only a few of these researches achieved accuracy of 80% [12]. DHAs use parametric phoneme classifiers [12]. DHAs phoneme classifiers have yielded up to approximately 80% phoneme classification accuracy [12]. Parametric classifiers make statistical assumptions about the data being processed [13]. The assumptions do not always fit practical applications [28]. Speaker-specific properties are aggregated in for the formation of statistical model – causing loss of information [14].

Hence, these classifications approaches have yielded limited success. The main research question therefore was: - Could the digital hearing aid's phoneme classification accuracy improve, if a non-statistical/non-parametric phoneme classification algorithm was developed and used in the processor of the digital hearing aid?

The goals of this study were to: -

- Develop a method to parse American English spoken words in to Phonemes, as per the International Phonetic Association Alphabet.
- Build a Non-Parametric Classifier →SVM
- Test the Spoken American English words/phrases against the built SVM Classifier.
- Determine the built classifier's accuracy.
- Compare the built non-parametric classifier accuracy to parametric classifier (used in hearing aids) accuracy.

II. LITERATURE REVIEW

Speech is produced by the movement of speech productions organs located at the top half of the human body known as articulators (see Figure 1). This consists of organs such as the lips, teeth, tongue, lungs, trachea, glottis, larynx, pharynx, oral cavity and the nasal cavity [15]. During speech production, the shape of the vocal tract varies due to movement of the articulators in the oral cavity namely, the tongue, jaws, lips and velum. If there is any abnormality in any movement of these articulators then speech impediment occurs.

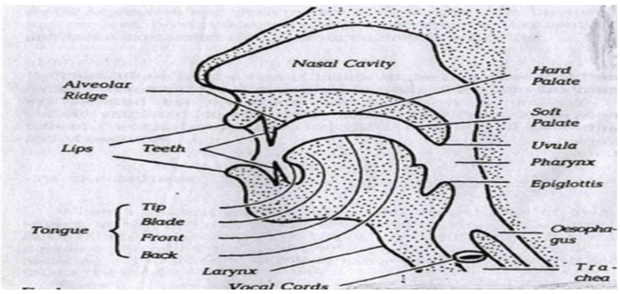

Figure 1. The Speech Production Organ [15]

Speech impediments can have severe impact on the ability to hear the intended spoken words [16], but this is made even worst for those persons with hearing impairments. Over the years, much has been put in place to help persons with hearing impediments. One such device is the hearing aid. Hearing aids have gone through five major periods, the acoustic era, carbon hearing aid era, vacuum tube era, transistor era, and the most recent, microelectronics era [17]. All of these have contributed significantly to achieve a tiny wearable device which can fit in the canal, in the ear, or behind the ear increasing the quality of speech which is delivered to a person's ears. Many hearing aid users are not satisfied with the quality signal they have hear, as these hearing aids simply amplifies all the sounds in the environment and not what the hearing aid user has a problem hearing. According to [18], advanced algorithms and more powerful signal processing have been able to produce better hearing aid units. One method of enhancing the hearing aid is to classify speech into phonemes [19], and correctly manipulate the phonemes which are not being heard clearly.

Figure 2 shows the basic layout of a speech classification/recognition system, in which a phoneme classifier can be embedded.

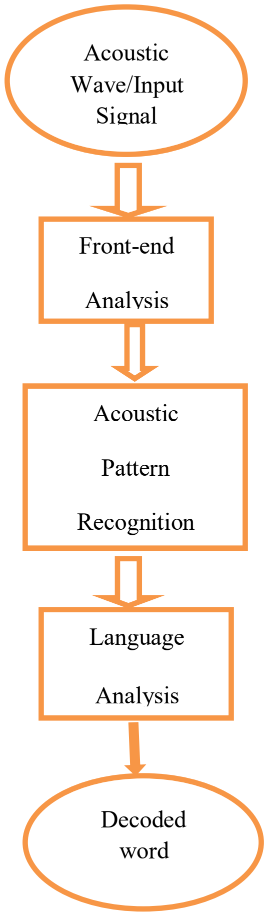

Figure 2. Speech Classification/Recognition System

Figure 2 shows the Acoustic wave block, which is the input speech from speaker; the Front-end Analysis, which extract acoustic features from input speech wave, outputs compact efficient set of parameters that represent the input speech properties, and uses one of 3 processing techniques to capture the necessary acoustic input wave properties. These techniques are: (i) Linear Predictive Coding (LPC), (ii) Mel-Frequency Cepstral Coefficients (MFCC), and (iii) Perceptual Linear Prediction (PLP). For this study MFCC was used, as it is best designed to capture positions and widths of formants (exactly the resonant frequencies of a vocal tract when pronouncing a vowel, that are acoustically perceivable, and have an easy interpretation and compact representation [28]; The Acoustic Pattern Recognition this block measure the similarity between an input speech and a reference pattern or model obtained during training, determines a reference or model, which best matches the input speech, as an output.

Acoustic model: - The incoming speech features from the front-end part are modelled by this unit. Several speech models exist. Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), and Artificial Neural Networks (ANNs), just to name a few of the most popular acoustic models [20]. HMM has been the most popular model used in speech recognition processing. However, GMMs, which includes HMM, fail to capture long-term (i.e., longer than one sentence) temporal dependency in acoustic features [20]. These weaknesses are the natural result of using statistical modes (HMM), that can generalize easily, thus speaker properties are aggregated in for the information of the statistical model and information is therefore lost [21].

On the contrary, in non-parametric-base model such as SVM, all the information from the training data is retained and not just the statistical approximations [21]. Keeping all the information from the training data results in retaining fine phonetic details, possibly resulting in more accurate speech classification/recognition [30]. SVM was used in this study.

Support Vector Machine (SVM)

SVM is a Supervised Learning algorithm, which is used for Classification as well as Regression problems. However, primarily, it is used for classification problems in Machine Learning. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification as well using a trick or parameter called as Kernel, which implicitly maps their inputs into high-dimensional feature spaces. SVM is a machine learning method based on statistic learning theory and it is classified as one of computational approach developed by Vapnik [22]. Based on the structural risk minimization (SRM) principal, SVM can get decision-making rules and achieve small error for independent tests set and hence can solve the learning problems efficiently [23]. Recently SVM is applied to solve the problems such as nonlinear, local minimum and high dimension. In many practical applications, SVM can ensure higher accuracy for a long-term prediction compared to other computational approaches. SVM is based on the concept of decision planes that define decision boundaries. SVM creates a hyperplane by using a linear model to implement nonlinear class boundaries through some nonlinear mapping input vectors into a high-dimensional feature space [24]. In SVM, there is some unknown and nonlinear dependency for example in mapping of function $\gamma = f(\chi)$ between some high-dimensional input vector x and scalar output γ (or the vector output y as in the case of multiclass SVM). No information regarding the underlying joint probability functions and one must contribute a distribution-free learning. Training data set D= {(xi, yi) ∈ X x Y}, I = 1, l where l stands for training data pairs and it is same to the size of training data set D. Frequently yi is stated as di, where d stands for desired target value. So, SVM is a part of supervised learning techniques. There are three major advantages of SVM, they are: (1) Only two parameters to be chosen, upper bound and the kernel parameter, (2) unique, optimal and global for solving a linearly constrained quadratic problem, the solution of, (3) good generalization performance due to the implementation of SRM principal. Due to these advantages, a number of studies have been conducted by researchers concerning SVM theory and application [25; 26].

Mel-Frequency Cepstral Coefficients (MFCC) Algorithm

Mel frequency Cepstral Coefficients algorithm is a technique which takes voice sample as inputs. After processing, it calculates coefficients unique to a particular sample. In this study, a simulation software called MATLAB R20163a was used to perform MFCC. The simplicity of the procedure for implementation of MFCC makes it most preferred technique for voice recognition. MFCC takes human perception sensitivity with respect to frequencies into consideration, and therefore are best for speech/speaker recognition.

Texas Instrument and Massachusetts Institute of Technology (TIMIT)

TIMIT is a corpus of phonemically and lexically transcribed speech of American English speakers of different sexes and dialects. Each transcribed element has been delineated in time. TIMIT was designed to further acoustic-phonetic knowledge and automatic speech recognition systems. It was commissioned by Defense Advance Research Project Agency (DARPA) and worked on by many sites, including Texas Instrument (TI) and Massachusetts Institute of Technology (MIT), hence the corpus' name. TIMIT corpus of read speech is designed to provide speech data for acoustic-phonetic studies and for the development and evaluation of automatic speech recognition systems. Although it was primarily designed for speech recognition, it is also widely used in speaker recognition studies, since it is one of the few databases with a relatively large number of speakers. It is a single-session database recorded in a sound booth with fixed wideband headset. TIMIT contains broadband recordings of 630 speakers of eight major dialects of American English, each reading ten phonetically rich sentences. The TIMIT corpus includes time aligned orthographic, phonetic and word transcriptions as well as a 16-bit, 16kHz speech waveform file for each utterance Corpus design was a joint effort among the Massachusetts Institute of Technology (MIT), SRI International (SRI) and Texas Instruments, Inc. (TI) [27].

III. METHODOLOGY

TIMIT, MFCC, and MATLAB were used in this study to create the SVM classification algorithm. A Computer, with the specification of 1 TB hard drive; 16 GB RAM; Intel Core i5 was used to build the SVM classifier in the MATLAB software.

Figure 3 shows the steps used in this study to build the classification algorithm, which afterwards was used to classify vowels and consonants of speech.

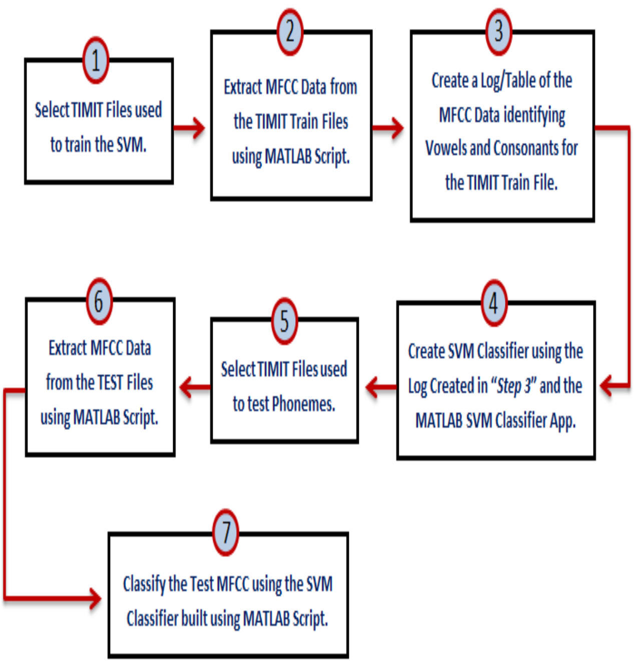

Figure 3. Steps used to Build the SVM Classifier

Step 1- Speech Selection (Input Audio Signal)

Five (5) random files were selected from the TIMIT training database to train the classifier that was built. Each of the five files was individually imported into the MATLAB software. This was done at the input command line in MATLAB.

Step 2 - Feature Extraction using MFCC

A set of codes were generated in MATLAB which created MFCCs for the individual phonemes created. All the processes involved in creating the MFCC algorithm were carried out in MATLAB. These include: Pre-emphasis, Framing, Hamming Window, Fast Fourier Transform, Mel Filter Bank processing, and Direct Cosine Transform. One thousand six hundred and forty (1640) MFCCs were created and used as the trained data from the TIMIT training folder.

A high pass filter was created in MATLAB. The filter is modeled through the mathematical equation [28]:

$$s_2[n] = s[n] - \alpha s[n-1] \qquad (1)$$

Where $s_2[n]$ is the output signal, $s[n]$ is the original input signal, and the value of "α" is usually between 0.9 and 1.0. The built classifier was tested with values of 'α' ranging from 0.9 -1.0. The result showed that the value $\alpha = 0.95$ gave the best classification results (when all other variables are held constant). Hence, 'α' was selected to be 0.95

A MATLAB code was generated to parse the phonemes of each of the selected TIMIT database files. The known range for each phoneme in the TIMIT speech files was created as individual speech signals. These segments of the speech were saved within MATLAB for feature extraction.

Step 3 – Tabulating the MFCC

The MFCCs gathered from the MATLAB processing, were placed in a table giving the 13 cepstrums, which creates a unique signature of the specified phoneme depending on the glottal disturbances which helped to form the phoneme. These were labeled as vowels and consonants.

Steps 4 & 5 - Creating the SVM Classifier Algorithm

The MFCCs obtained from step – 3 above were used to build the SVM classifier in MATLAB, by importing the data from the five TIMIT training speech files into the MATLAB Classification App. The SVM classifier algorithm was created from the compiled phoneme table. The SVM Classifier was trained and exported to the model space.

Step 6 – Extract MFCCs from TIMIT Test Speech Files

A MATLAB code was written and executed, which parsed and generated the phonemes of each of the selected TIMIT test speech files. These phonemes were saved and used later to test the built SVM Classifier.

Step 7 - Testing the New Query data

The data that was used to test the accuracy of the newly created SVM model was taken from the TIMIT speech file test folders. The phonemes obtained from the speech test files (step 6) were tested in the built SVM classifier. The resulting MATLAB generated MFCC was then classified against the built SVM model to determine if the selected phoneme was a consonant or a vowel.

IV. RESULTS

Selected TIMIT Training Speech Files

The five TIMIT training speech files used in this study accounted for 40 of the 44 IPA American English Phonetic Alphabet. This these 40 represents approximately ninety-three percent (93%) of the entire IPA phonetic alphabet. The training files used had the following path in the TIMIT database:

- DR1- FCJFO-SI648.wav - 'A sailboat may have a bone in her teeth one minute and lie becalmed the next.'
- DR3- FITM0-SX80.wav - 'It's illegal to postdate a check.'
- DR2- MRLR0- SA1.wav - 'She had your dark suit in greasy wash water all year.'
- DR4- MLBC0-SX429.wav- 'A toothpaste tube should be squeezed from the bottom.'
- DR5- FLOD0-SX117.wav - 'The mango and the papaya are in a bowl.'

Feature Extraction

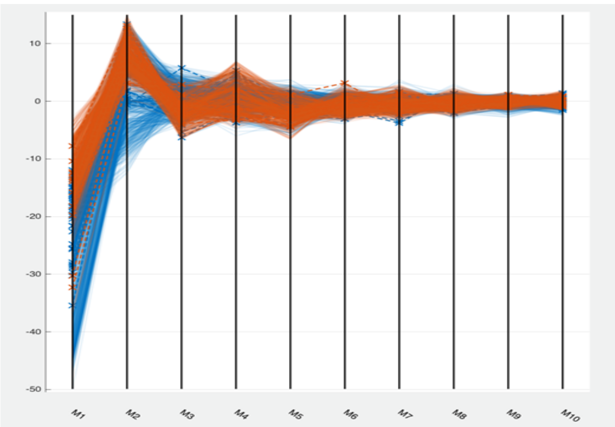

Figure 4. First 10 MFCCs for all Phonemes used to Build SVM Classifier

Figure 4 shows a plot of the first 10 extracted relevant MFCCs. The values for these 10 MFCCs for all 1640 phonemes were placed in a tabular format, and a label of consonant or vowel was assigned to each phoneme, according to IPA. Figure 4 shows that the magnitude for the eight MFCC to the tenth MFCCs number 10 and beyond were low, and hence would not have any significance contribution on the overall characteristics of the speech signal. The values for these 10 MFCCs for all 1640 phonemes were placed in a tabular format, and a label of consonant or vowel was assigned to each phoneme, in accordance to standard American English Phonetic Alphabet.

The SVM Classifier

Figure 5 shows the result from testing the built SVM classifier with the TIMIT test file "DR2-MMDB1-SA2 – "Don't ask me to carry an oily rag like that." As shown in the guide of Figure 5, 1- represented correctly classified phonemes, 0 represented incorrectly classified phoneme and 2 represent phonemes which produced an undefined error for this TIMIT test file.

```
*SA2 - Notepad
File  Edit  Format  View  Help
0 43725 Don't ask me to carry an oily rag like that.
0       2340 h#
2340    2580 d          - 2
2580    4120 ow         - 1
4120    4680 n          - 1
4680    6720 ae         - 1
6720    7690 s          - 1
7690    8117 kcl        - 1
8117    8508 m          - 1
8508    9322 ix         - 1
9322    9990 dcl        - 1
9990    10190 d         - 2
10190   10800 ix        - 1
10800   11380 kcl       - 1
11380   12970 k         - 1
12970   13827 ih        - 1
13827   15078 r         - 0
15078   16041 iy        - 1
16041   16640 ix        - 1
16640   17440 n         - 1
17440   19676 oy        - 1
19676   21054 l         - 0
21054   22475 iy        - 1
22475   24104 r         - 1
24104   26193 ae        - 1
26193   26600 gcl       - 1
26600   27187 g         - 1
27187   27738 l         - 0
27738   28920 ay        - 1
28920   29330 kcl       - 1
29330   30580 dh        - 1
30580   33817 ae        - 1
33817   36299 q         - 1
36299   43680 h#
Guide
0 - INCORRECTLY CLASSIFIED
1 - CORRECTLY CLASSIFIED
2 - UNDEFINED
```

Figure 5. Phoneme Classification Result using of the SA2 TIMIT Test File using the Built SVM

Of the thirty-one phonemes classified, three were misclassified and two produced an undefined error. Therefore, twenty-six of the phonemes which were in the speech signal were correctly classified by the built SVM classifier. This gives an accuracy of 83.4% (number of correctly classified phonemes, 26/total number of phonemes in the speech file, 31) x 100%. However, given that two of these phonemes produced an error, the actual accuracy of the built SVM model created was 89.7% for this speech file.

The classification accuracy obtained from using the built SVM classifier for the other four (4) TIMIT test speech files were:

(i) 92.7% for the TIMIT test file DR7>MPSB0>SA1128. That is, (38/41) x 100% = 92.7%
(ii) 74% for the TIMIT test file DR2>MPDF0>SX102. That is, (34/50) x 100% = 74 %
(iii) 85.1% for the TIMIT test file DR1>MDAB0>SX409. That is, (40/47) x 100% = 85.1%
(iv) 86.5% for the TIMIT test file DR3>MJES0>SI1384. That is, (32/37) x 100% = 86.5%

V. CONCLUSION

Based on the research done and the results obtained, the following conclusions were drawn:

- According to [29], existing methods of classifying phonemes accuracy range between 78% - 96% using the parametric classifiers. However, it was shown from the results of this research, that a non-parametric phoneme classifier, such as Linear SVM, can efficiently and effectively classify phonemes with similar classification accuracy as compared to parametric phoneme classifiers, but with lot less complications compared to most other speech classification algorithms
- A successful SVM model was developed using MATLAB and the TIMIT database. This model can be used to classify vowels and consonants of speech. The developed model produces accuracy between 74% to 92.7%.
- Although the built SVM phoneme classifier did not produce an accuracy of eighty percent (80%) for all the tested speech files, it did produce greater that 80% accuracy for four (4) of the five (5) tested speech files, which is equal to or better that most of the existing speech classification algorithms.

VI. RECOMMENDATIONS

The following recommendation can be put in place as it relates to speech classification using the non-parametric method of linear SVM: -

The built SVM classifier was developed using the MATLAB software, which can be uploaded to the microprocessor of most Digital Hearing Aids. Hence, this built SVM should be able to improve hearing for the community of hearing aid users, especially those who suffers from the inability to differentiate between vowel and consonant sounds.

References

[1] Moore, B.C., Glasberg, B.R. Modelling binaural loudness. J Acoust Soc Am. 2007; 121:1604–1612.

[2] Strom, K. E. "The HR 2006 Dispenser Survey." *Hearing Review* vol. 13, no. 6, 2006, pp. 16-39.

[3] Bentler, Ruth A. “Effectiveness of Directional Microphones and Noise Reduction Schemes in Hearing Aids: A Systematic Review of the Evidence.” *Journal of the American Academy of Audiology*, vol. 16, no. 07, 2005, pp. 473–84.

[4] Prasad, Bhanu, and SR Mahadeva Prasanna, eds. *Speech, audio, image and biomedical signal processing using neural networks*. Vol. 83. Springer, 2007

[5] Phatak, S. A., and Allen, J. B. (2007). “Consonant and vowel confusions in speech-weighted noise,” J. Acoust. Soc. Am. **121**, 2312–2326

[6] Billings CJ, Bennett KO, Molis MR, et al. Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. Ear Hear. 2011;32 (1):53–60.

[7] The *sounds of speech communication*. A primer of acoustic phonetics and speech perception. Austin, Texas: PR O-ED.J.M Pickett **-** 1980 **–** Inc

[8] Ladefoged, P. (2001). Vowels and Consonants: An Introduction to the Sounds of Languages (Blackwell, Oxford), pp. 1– 19

[9] Potamianos, G., Neti, C., Gravier, G., Garg, A., & Senior, A. W. (2003). Recent advances in the automatic recognition of audio-visual speech. Proceedings of the IEEE, 91(9), 1306-1326.

[10] Richardson, M., Bilmes, J., and Diorio, C. (2000b). “Hidden-articulator Markov models: Performance improvements and robustness to noise,” in Proceedings of the International Conference on Spoken Language Processing, Beijing, China.

[11] Fuchs, S., Winkler, R., & Perrier, P. (2008). Do speakers’ vocal tract geometries shape their articulatory vowel space? In R. Sock, S. Fuchs, & Y. Laprie (Eds.), Proceedings of the International Seminar on Speech Production (pp. 333–336). Strasbourg, France: INRIA

[12] Yunusova Y, Weismer G, Westbury J, Lindstrom M. Articulatory movements during vowels in speakers with dysarthria and in normal controls. Journal of Speech, Language, and Hearing Research. 2008; 51:596–611

[13] Mather, P. and Tso, B. (2009) Classification Methods for Remotely Sensed Data. CRC Press, Boca Raton.
[14] Rizwan, Muhammad, and Anderson, David V. "Comparison of Distance Metrics for Phoneme Classification based on Deep Neural Network Features and Weighted k-NN Classifier." *Workshop on Machine Learning in Speech and Language Processing*. 2016.
[15] Coleman, John. "Air and Phonation." *The Vocal Tract and Larynx*, www.phon.ox.ac.uk/jcoleman/phonation.htm. Accessed 30 Dec. 2021.
[16] Nicolosi L, Harryman E, Kresheck J. Terminology of Communication Disorders. 4. Baltimore: Williams and Wilkins; 1996
[17] R. E. Sandlin, Textbook of Hearing Aid Amplification, CA, San Diego: Singular, 2000.
[18] Levitt, H. (2007). A historical perspective on digital hearing aids: how digital technology has changed modern hearing aids. *Trends in amplification*, *11*(1), 7-24.
[19] B. Prasad, S.R.M. Prasanna (Eds.), Speech, Audio, Image and Biomedical Signal Processing using Neural Networks, Springer-Verleg, Berlin (2008), pp. 239-264
[20] L. Deng, H. Strik, et al. Structure-based and template-based automatic speech recognition: comparing parametric and nonparametric approaches. In Proc. Interspeech, pp. 898--901, 2007
[21] M. Rizwan, D. V. Anderson, "Using k-Nearest Neighbor and speaker ranking for phoneme prediction", *Machine Learning and Applications (ICMLA) 2014 13th International Conference on*, pp. 383-387, 2014
[22] Mat Deris, A. and Mohd. Zain, A. and Sallehuddin, R. (2011) *Overview of support vector machine in modeling machining performances.* Procedia Engineering, 24 pp. 308-312. ISSN 1877-7058
[23] Deng, C., Wu, J. and Shao, X. (2008). Reliability assessment of machining accuracy on support vector machine, ICIRA 2008, Part II, LNAI 5315, pp. 669–678.
[24] Samanta, B., Al-Balushi, K.R., and Al-Araimi, S.A. (2003). Artificial neural networks and support vector machine with genetic algorithm for bearing fault detection. Engineering Application of Artificial Intelligent 16, 657-665.
[25] Shin, K.S., Lee, T.S., and Kim, H.Y. (2005). An application of support vector machines in bankruptcy prediction model, Expect Systems with Application 28, 127-135.
[26] Sugumaran, V, V., Sabareesh, G.R., and Ramachandran, K-I. (2008). Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. Expert Systems with Applications 34, 3090–3098.
[27] L.R. Rabiner and R.W. Schafer, (Third Edition), Digital Processing of Speech Signals, Pearson Education, Delhi, 2009
[28] B. Thorpe and T. Dussard, "Classification of Speech Using MATLAB and K-Nearest Neighbour Model: Aid to the Hearing Impaired," SoutheastCon 2018, St. Petersburg, FL, USA, 2018, pp. 1-8, doi: 10.1109/SECON.2018.8479223.
[29] Nwe, Tin et al., "Speech Emotion Recognition Using Hidden Markov Models". *Speech Communication*, no. 41, 2003, pp. 603-623.
[30] Bzdok, Danilo et al. "Statistics versus machine learning." *Nature methods* vol. 15,4 (2018): 233-234.