Development and Comparative Analysis of a Novel Neural Model for Marathi Alphabet Recognition for Hearing Aid Users

[1]Prashant G. Patil, [2]Arun. K. Mittra, [3]Vijay.S.Chourasia
Research Scholar[1], Professor& Head[2], Assistant Professor[3]
[1, 2, 3] Department of Electronics Engineering.
Manoharbhai Patel Institute of Engineering &Technology, Gondia, India

---

***Abstract*— In this Paper we associates differences between high frequency hearing loss reduction methods which is useful in improvement in speech intelligibility of HA users. Frequency Compression (M1) and Frequency transposition (M2) was usually implemented by many researchers & accepted by numerous HA users. In certain circumstances both methods was not acceptable by HA users. We proposed two feed forward back propagation neural network based Frequency Compression (M3) & Frequency Transposition Methods (M4) methods. Four algorithm was designed using MATLAB & tested on Laptop with wired ear set. The performance parameter of all four methods was found by using audiogram of all 6 HA users. These Parameters are fitted to corresponding MATLAB based algorithm. Regional Marathi language spoken HA users are selected as participants for recognition & Performance evaluation test of all four algorithm. It is observed that average recognition rate by group of 6 HA user increased for vowels & consonants from minimum 2 % to 14 %.All these tests are considered insensitive for objectively and accurately measuring aspects of listeners' speech perception abilities as a reflection of their performance in realistic listening situations. As a result of this Experimentation, the participant will be able to compare the average performance of all four methods and determine if his or her performance as expected on various processing methods for improvement in high frequency region.**

***Index Terms*— Frequency Compression, Frequency Transposition, Neural Network, Hearing Aid, Hearing Loss, Hearing Aid users.**

I. INTRODUCTION

There are presently several people in the India suffering from High frequency hearing loss. From many years they had to rely on conventional hearing aids. Conventional hearing aids provide little benefit for moderately deaf patients. Some individuals with HA can now communicate without lip-reading, and some can communicate over the telephone. Part of the success of Hearing aid can be attributed to the signal processing techniques developed over the years for extracting electrical stimuli from speech. Frequency Compression & Frequency transposition are two major techniques adopted by many HA users. Both methods have certain limitations which will not meet user’s requirement. From Many Years researchers have discussed the Hearing aid signal processing, Validation & recognition results for improving intelligibility. Speech intelligibility varies according to background condition of speaker & Listener. Background noise makes operative impact in reducing percentage recognition of alphabets, words, short words & rhyming words (Similar Pronunciation words).In this study, we investigated the effects of existing two frequency-lowering algorithms for high-frequency hearing disabled patients with developed method. Much attention has been paid to associating the differences in the effects with the changes in lowering strategies and parameter settings by conducting a comparative study. Frequency lowering has been proposed as a promising approach to combating high-frequency hearing loss. This includes various signal processing techniques such as channel vocoder, frequency compression, and frequency transposition (FT), all presenting high-frequency information in a lower frequency region accessible for people with high-frequency hearing loss. Channel vocoder divides the speech signal into frequency bands by band pass filters and extracts the envelopes of high-frequency signals to modulate a noise source, which will be added to the unmodified low-frequency signals. Frequency compression reduces the bandwidth of a speech signal in a linear or nonlinear fashion. Frequency transposition (FT) shifts the high-frequency components to a lower-frequency band and adds them to the unprocessed lower-frequency signals. The FT method became the first frequency-lowering technique implemented in a commercial hearing aid. In developed methodology (M3&M4) a deep neural network was trained on features extracted from speech samples of different environments which are relevant for hearing aid users. Another set of sound samples was used to test the classifier. The features used in this study reflect both spectral and temporal aspects of the input Speech signal. This mimics important aspects of the auditory system, where not only frequency information is represented in a topographical way, but also a gradient of T-F, SNR and Intensity representation can be found. The results were compared to classification of the same set of sounds with hearing aids from different manufactures. This paper is organized as follows. Sec. 2 describes the Performance Details of Existing & Modified proposed methods. Sec. 3 presents Experimentation procedure for testing of all algorithm. Sec. 4 Describes Recognition results & observations. Sec 5 concludes the paper and outlines the future work.

II. Technical Comparison of Existing & Proposed Algorithm

In this section we proposed two modified algorithm to compare with traditional Frequency compression (FC) & Frequency transposition technique (FT).Table 1& 2 illustrate Processing Method, essential parameters, Limitations of both algorithms.

Table 1 Details Designing parameters of FC & FT methods

	Frequency Compression (FC) M1	**Frequency Transposition (FT) M2**
Processing Methodology	Low-frequency Speech contents are compressed as well to make room for the high-frequency contents	Higher frequency (f H) is shifted to lower frequency (f L) by a fixed frequency value. It does not reduce bandwidth.
Processing Parameters	Amount of amplification applied to Signal Gain = Output – Input (I/O Gain control)	Target band: Range of high frequency band which is transposed over low frequency.
	Compression Threshold/Knee point	Source band: Low frequency band on which high frequency is transposed.
	Compression Ratio CR= Δ Input/ Δ Output	Cutoff frequency: Extreme edge of target band.
	Attack Time and Release Time	equal-loudness (Intensity Control)
	No of channels- Single / Multiple	'N' Point selection for FFT
Performance based application for HA User		
To Avoid distortion	Gain Controlled- O/P Threshold Knee Point – High Compression Ratio- High Attack & Release Time- Fast Channels- Single /Multiple	Target Band: Low Source Band: Low Cutoff (Edge) Frequency: Low Loudness Limiter: Active 'N' Point FFT: Less Value
Loudness Perception	Gain Controlled- I/P Threshold Knee Point – Low Compression Ratio- Low Attack & Release Time- Slow Channels- Multiple	Target Band: Low Source Band: Mid Cutoff (Edge) Frequency: Low Loudness Limiter: Less-Active 'N' Point FFT: Average
Listening Comfort	Gain Controlled- I/P Threshold Knee Point – Mid Level Compression Ratio- Mid Level Attack & Release Time- Slow Channels- Single /Multiple	Target Band: Mid Source Band: Mid Cutoff (Edge) Frequency: Mid-High Loudness Limiter: Active/Less Active 'N' Point FFT: Middle
Maximize Speech Intelligibility	Gain Controlled- I/P Threshold Knee Point – Low Compression Ratio- Low Attack & Release Time- Average Channels- Single /Multiple`	Target Band: Mid-High Source Band: High Cutoff (Edge) Frequency: Mid-High Loudness Limiter: High Active 'N' Point FFT: High Value
Limitations & Drawbacks	Processing parameters can be altered according background speech environment.	Strong distortion occurs when shifting frequency is greater than signal frequency
	Less effective for Female & Child speaker.	If the target band is high, it will not preserve high-frequency information at low frequencies.

The Frequency Compression (M1) &Transposition (M2) algorithm was implemented using MATLAB. The original speech material was digitally recorded at 16000 samples per second. A frame consisted of 128 samples, lasting 8 ms, and the entire test material was processed frame-by-frame with an overlap of 64 samples. The current frame was multiplied by a half-sine window and then converted from the time to the frequency domain using the MATLAB fast Fourier transform (FFT) function. A 128-point FFT resulted in a short-term frequency spectrum with steps of 125 KHz (sampling frequency/FFT-size).All above FC, FT, and FFBPNN-FC & FFBPNN-FT algorithms were designed using MATLAB. Pre-recorded Marathi vowel, Consonant & words data base is used for testing & recognition purpose by group of 6 HA users. Performance parameter of each Method was set in coding with relevance of requirement of HA user. Recognition Test was conducted in a quiet meeting room for 15 vowels & 35 Consonants. Data base is recorded in silent room, Noisy Room & with music background conditions. To maximize speech intelligibility we need to set parameters according to Table 1 for M1 & M2 methods.

Table 2 Details Designing parameters of FFBPNN based FC & FT methods

	Feed forward back propagation neural Network +Frequency Compression (FFBPNN+FC)	**Feed forward back propagation neural Network+ Frequency Transposition (FFBPNN+FT)**
Processing Methodology	Feature Extracted from Each frame + classification of frame for Compressing.	Feature Extracted from Each band + classification of band for Transposition.
Features Extracted	1. SNR of each frame 2. Fundamental Formant Frequency 3. Frame Intensity	
Neural Network Parameters	• It consists of an input layer • 5 hidden layers • 1 linear featured output corresponding to input. • Back propagation was used for training the NN in full-batch mode over 500 epochs • Variable learning rate of 0.01-0.03 • Random Assigned weight 1.2-0.2. • No of iteration 100 with different Neuron Size.	
FFBPNN Outcomes	• Classifies Alphabets for processing (Needed/Unwanted) • Point to Point analysis of speech frames/Bands. • Training rate given to Neural Network varies Classifier performance.	
Post Processing approach	Divide Into Six bands 0-250;250-500;500-1000; 1000-2000;2000-4000;4000 -onwards	Divide Into 2 bands (Target & Source) 0-4000;4000-Onwards
Speech Intelligibility Factors	Training of Neural Network Minimized error rate N-FFT (N Should be moderate) Input Processing Time	
NN Performance Parameters	Sensitivity, specificity, Accuracy, False Positive rate(FPR),False Negative Rate (FNR), False Acceptance Rate (FAR), False Rejection Rate(FRR), Positive Predicted Value(PPV), Negative Predictive Value(NPV)	
Limitations & Drawbacks	1. Average processing time increases for all Marathi vowels & consonants 2. More Data required for neural network training. 3. Incorrect classification will creates unwanted processing. 4. Unwanted processing will degrade speech quality	

III. EXPERIMENTATION & TESTING OF ALGORITHMS

In Phase-I, the Marathi vowels, consonants data collection was done in a small room with Silent, electric fan and air conditioner switched on. The data was collected in parallel with a headset microphone connected to a Tablet PC, the built-in microphone of another Tablet PC. 9 Female and 3 Male participants are involved for data collection with age varying from 16-23 years old. The motivation behind choosing these devices for data recording was to get a good sample of the set & female speakers count chosen large as compared to male. High frequency hearing loss patient was difficult to hear female speech which is comparatively higher frequency than male speaker. This section of the article presents collection of auditory behavioral data of each listener before experimentation which is useful for finding out optimum set of performance parameters. Figure 1 shows audiogram of Six Hearing aid users for corresponding ear. Audiogram is useful for parameter fitting which reduces no of iterations in experimental for improving speech intelligibility.

Subjects and methods

No of 6 students who have high frequency hearing loss were selected (Age 7-23) randomly. Regional Marathi language is selected for testing & recognition purpose.

Procedure

The proposed system and algorithms were first evaluated in normal hearing people. No obvious rejection was reported and some listeners showed improvements in speech perception. But according to the previous studies, effective frequency-lowering algorithms could provide quite varied effects in normal hearing people. So the results of an experiment on qualified patients will be most reliable. All the fitting and testing processes were accomplished in a silent room in a School, with the earphone plugged to laptop placed 1 m in front of the listeners. The testing sounds were broadcasted at different level of speech intensity from 35-60 dB regardless of the level of background noise.

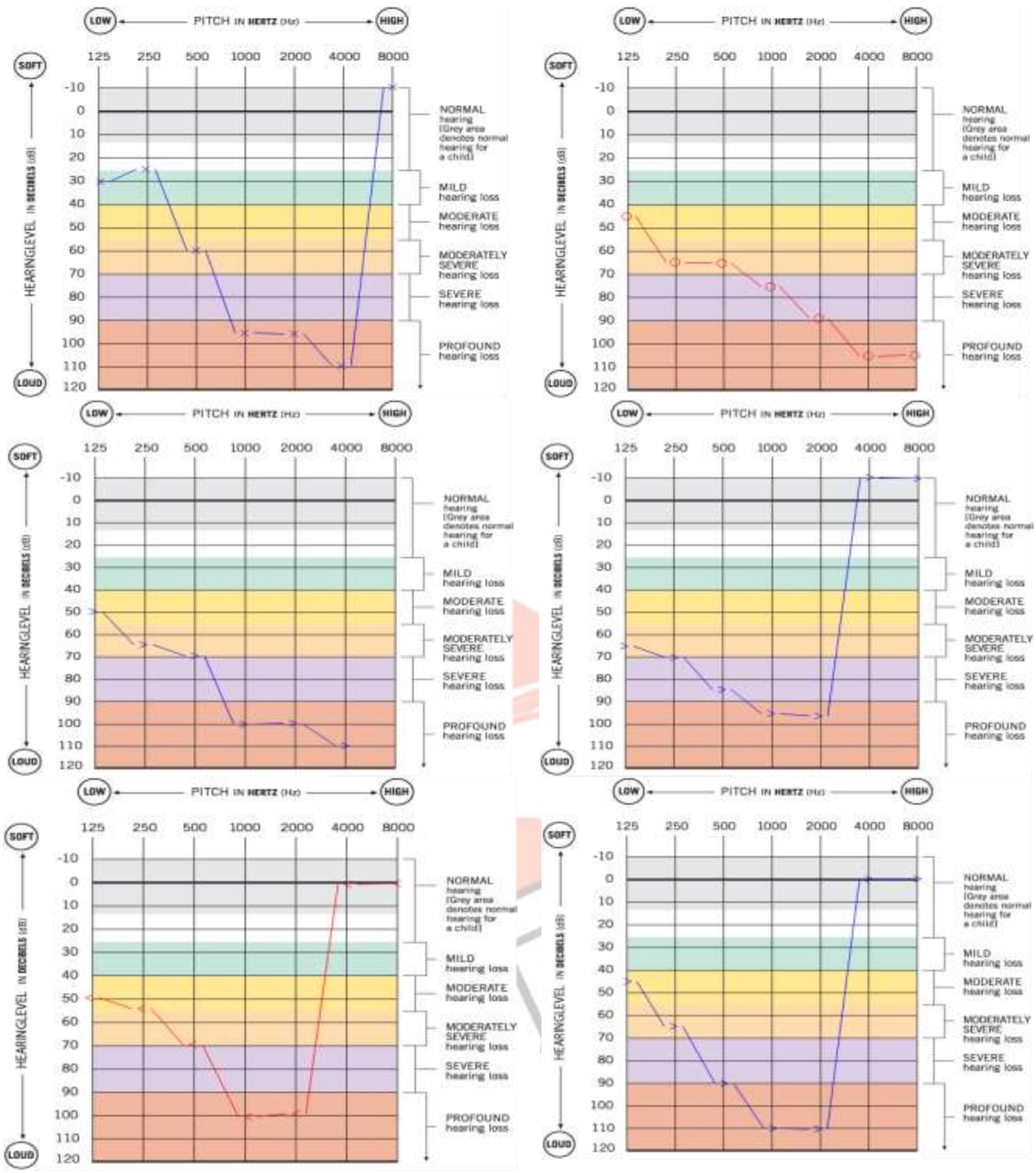

Figure 1: Mean right-ear and left-ear air-conduction pure-tone thresholds for each of the HA user. Each HA user was fitted with a different hearing-aid technology

The above-mentioned parameters in Table 1 & 2 should be fitted in the order of 'voiced speech related', 'unvoiced speech related', 'mode selector', and 'envelope shaping switcher'. This is consistent with their importance in terms of the algorithms' effect. The whole fitting procedure was carried out under the direction of audiologists. They listened to a patient's description and helped to adjust the parameters according to the subject's reaction.

Pre -recorded Marathi 15 vowels, 35 consonants, pair of 50 Marathi rhyming words (Similar Spoken words) which is not understood by HA user clearly are taken for recognition test. All these data is recorded in different background condition (Silent Room, Music background, Running Fan sound at background).No of 12 speakers (9 Female & 3 male) are participated for this data collection. We have 540 sample of individual alphabet. These will randomly processed by any one method & listened to individual HA user. The intelligibility of each HA user was rated on a six-point scale by the speech and language therapist with the assistance of written alphabet by HA user. It was possible to reach the subjective satisfaction of the HA user within two or three Week fitting sessions (including the Fine-tuning). As the test subjects had mild to moderate high-frequency hearing loss with a large dynamic range and amplification occurred with the ear canal open, it was possible to set the compression ratio to low values, resulting in both clear and natural sound output.

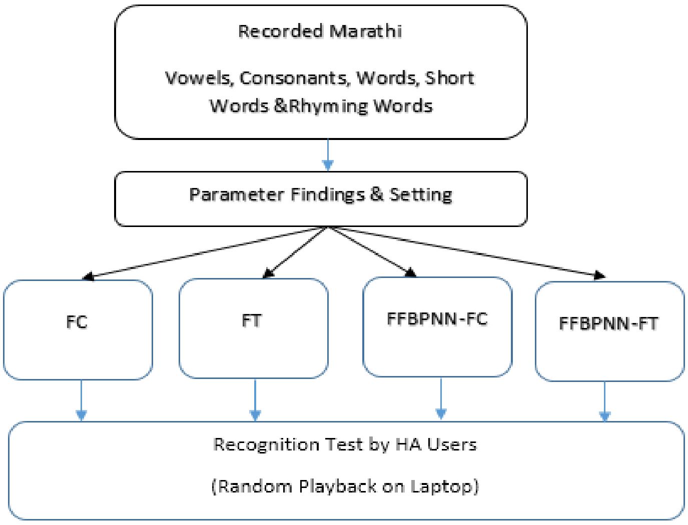

Figure 2: Software system used in this experiment for four different method Evaluation

IV. Recognition Results & Observations

Each Marathi Vowel is randomly play backed & processed 10 times using different methods for each HA users. For Example - Each vowel is play backed for 10 times for 4 different methods (1*10*4=40 Times) The Processing parameters are set by taking average of all 6 HA users .Average Recognition rate by all four Methods of individual HA user for each method & for all 15 Vowels are calculated as shown in Table 3.Where M1 – Frequency Compression, M2-Frequency transposition, M3- FFBPNN-FC, M4-FFBPNN-FT methods.

Table 3 Individual Average Vowel recognition rate by using M1, M2, M3 & M4 methods & Group Average recognition comparison for four methods.

HA User Description (Age)	M1	M2	M3	M4
L1: JSB HA002- BTE Hearing Aid :(18 years)	5.06	6.13	6.53	6.53
L2: Siemens Vita 118:(22 Years)	5	5.06	5.73	5.6
L3: ALPS PP BTE RIE: (12 Years)	4.13	3.53	6.4	6.8
L4: ALPS high gain BTE: (17 Years)	4.93	3.66	6.86	6.2
L5: RESOUND 793 UHG BTE: (16 Years)	4.06	5.93	6	7.66
L6: Rexton Arena 1S: (21 Years)	5	4.86	5.33	5.86
Average Recognition Score from Group of 6 HA users (From 10 times Playback)	4.7	4.86	6.14	6.44
Average % Recognition Score by Group	47.33 %	48.60%	61.40 %	64.40 %

$$Average\ \%\ Recognition\ Score\ by\ Group = \frac{(No\ of\ times\ Correctly\ recognised)}{(Total\ No\ of\ Playbacks\ using\ Each\ Method)} * 100 \qquad (1)$$

Table 4 Individual Average Consonant recognition rate by using M1, M2, M3 & M4 methods & Group Average recognition comparison for four methods.

HA User Description ,Age	M1	M2	M3	M4
L1: JSB HA002- BTE Hearing Aid :(18 years)	18.22	21.80	21.57	20
L2: Siemens Vita 118:(22 Years)	19.68	20.74	20.4	19.88
L3: ALPS PP BTE RIE: (12 Years)	20.88	22.05	21.91	19.6
L4: ALPS high gain BTE: (17 Years)	21.51	19.6	21.25	21.57
L5: RESOUND 793 UHG BTE: (16 Years)	22.77	21.54	22.37	20.97
L6: Rexton Arena 1S: (21 Years)	21.37	20.91	22.97	19.71
Average Recognition Score from Group of 6 HA users (From 35 times Playback)	20.74	21.10	21.74	24.10

Average % Recognition Score by Group	59.85%	60.27%	62.13%	61.11%

Average Vowel recognition score by each group decides effectiveness of individual method on Individual participant & group. It is observed that both proposed methods shows Improvement (13 % to 17 %) in Vowel Recognition as compared to existing Methods. Average Consonants recognition score by each group decides effectiveness of individual method on Individual participant & group. It is observed that both proposed methods shows Improvement (2 % to 3 %) in Consonant Recognition as compared to existing Methods. The confusion matrix in Fig. 3 shows the correct recognition count with Spoken consonant. The diagonals correspond to correct identification of each Consonants. Group of Marathi Confusing Consonant is responsible for reduction in recognition rate. Each of 6 HA user not able to identify correctly spoken Consonant.

	Spoken Alphabet (Column) Vs. Recognized Alphabets by HA user (Row)											
	क	ख	ग	घ	ज	झ	द	ध	ब	भ	श	स
क	12	7	4	2	0	1	0	2	2	3	2	0
ख	3	5	3	7	4	2	1	0	2	3	4	1
ग	4	4	6	2	3	4	5	0	1	2	3	1
घ	0	2	5	8	2	1	1	0	4	7	4	1
ज	3	1	7	1	9	6	2	0	2	3	1	0
झ	4	5	0	2	0	6	0	4	8	0	1	5
द	5	4	2	5	0	6	4	0	3	1	2	3
ध	0	0	3	7	0	5	7	11	0	1	1	0
ब	2	2	1	0	5	1	8	6	8	0	0	2
भ	0	2	0	0	6	0	3	6	0	9	4	5
श	2	0	4	0	4	0	2	2	2	3	8	8
स	0	3	0	1	2	3	2	4	3	3	5	9

Figure 3: Confusion Matrix for Marathi Spoken Consonant & recognized Consonant, Each Consonant spoken 35 times & Combined Score of recognition by FC (M1) & FT (M2) method

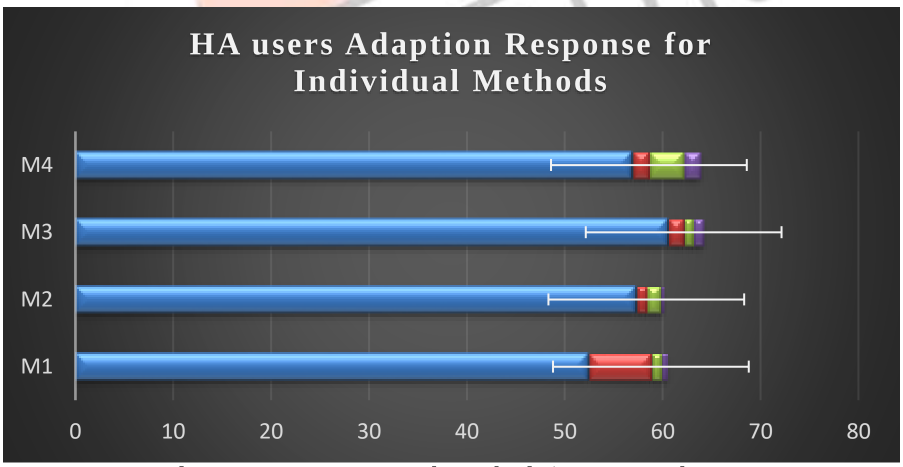

Figure 4: Group HA users Adaption Response to Each Method. (M3 & M4 shows >60 % response than other)

V. Conclusion & Future Work

Our study explores the impact of parameter setting & fitting of each algorithm on improving Marathi vowels & consonants recognition. Proposed FFBPNN Frequency Compression (M3 method) & FFBPNN frequency transposition (M4 method) shows improvement in Vowel % recognition rate from 47 % to 64 %.Figure 4 shows improvement in consonant % recognition from 59 % to 62 %. This improvement is larger in terms of vowels than consonants. In Regional Marathi language many articulated consonants are in same with spoken style, lip movement, Tongue movement which are hard to distinguish by HA user. The group of confusing consonants is challenging task in recognition process, which overall reduces % recognition & % group recognition overall rate. Hearing aid user age is also key factor in recognition process, it is observed that recognition ability for alphabets is more for older HA user in group. Confusing matrix obtained from consonants recognition indicates need of some improvement in M3 & M4 Systems. The performance of M3 & M4 system is decided by % training data given to neural network, which will classify processing & unwanted processing alphabets. Unwanted processing of alphabets degrades quality of alphabets, reduced capability of recognition. In Marathi language most words starts from consonant so more development is needed in correct classification of consonants. Re-modification & correct set of parameter finding for M3 & M4 Algorithm will progress proposed system. For NN networks error minimization is complex &challenging issue. For other language algorithm behavior will be different or may be improved.

REFERENCES

[1] Joanna D. Robinson, Thomas Baer, Brian C.J. Moore, "Using transposition to improve consonant discrimination and detection for listeners with severe high-frequency hearing loss", International Journal of Audiology 2007; 46:293-308.

[2] Cascade recurring deep networks for audible range prediction, Yonghyun Nam, Oak-Sung Choo, Yu-Ri Lee, BMC Medical Informatics and Decision Making 2017.

[3] Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users, Tobias Goehring , Federico Bolner Hearing Research 344 (2017) 183-194

[4] Haris B C, G Pradhan, A Misra, S Shukla, R Sinha and S R M Prasanna, "Multi-Variability Speech Database for Robust Speaker Recognition", 978-1-61284-091-8/11©2011 IEEE Signal Processing.

[5] Xianbo Xiao, Hui Zhang, Guangshu Hua, Chun hong Liu, Jia Liu, "Evaluation of frequency-lowering algorithms for intelligibility of Chinese speech in hearing-aid users" Progress in Natural Science 19 (2009) 741–749.

[6] Joanna D. Robinson, Thomas H. Stains, Thomas Baer, Brian C.J. Moore, "Evaluation of a frequency transposition algorithm using wearable hearing aids", International Journal of Audiology 2009; 48:384-393.

[7] Yen-Teh Liu, Yu Tsao, and Ronald Y. Chang, "Nonnegative matrix factorization-based frequency lowering technology for mandarin-speaking hearing aid users", ICASSP 2016.978-1-4799-9988-0/16.

[8] Sean Brac, "Critical Review: What is the effect of frequency transposition on speech perception in children", Journal American Academy of Audiology, 10 (1) 2017, 14–25.

[9] Francis Kuk, Denise Keenan, Petri Korhonen, Chi-chuen Lau, "Efficacy of Linear Frequency Transposition on Consonant Identification in Quiet and in Noise", Academy of Audiology 20:465–479 (2009).

[10] B.M. MacArdle, C. West, J. Bradley, S. Worth, J. Mackenzie & S.C. Bellman, "A study of the application of a frequency transposition hearing system in children", British Journal of Audiology2016,ISSN: 0300-5364 (Print) (Online) Journal homepage: http://www.tandfonline.com/loi/ijap20.

[11] Chung-Hwa Chang, Gary T. Anderson, Philipos C. Loizou, "A Neural Network Model for Optimizing Vowel Recognition by Cochlear Implant Listeners", IEEE transactions on neural systems and rehabilitation engineering, vol. 9, no. 1, march 2001.

[12] MinooLenarz, ThomasLenarz, MartinStieve ,Mark Winter, "A comparative study of the audio logical outcomes with Retro-X(semi-implantable hearing aid system) and conventional open fitting hearing aids",Eur Arch Otorhinolaryngology (2008) 265:1189–1194.

[13] Joshua M. Alexander, "Nonlinear Frequency Compression: Balancing Start Frequency and Compression Ratio", American Auditory Society Annual Meeting, March 8-10, 2012.

[14] James M. Kate's, Senior Member, IEEE, and Kathryn H. Arehart, "The Hearing-Aid Audio Quality Index (HAAQI)" IEEE/ACM transactions on audio, speech, and language processing, VOL. 24, NO. 2, FEB 2016

[15] Extended bandwidth nonlinear frequency compression in Mandarin-speaking hearing-aid users, Wen-Hsuan Tseng, Dun-Lieh Hsieh, Wan-Ting Shih, Tien-Chen Liu Journal of the Formosan Medical Association (2017) xx, 1e8. www.jfma-online.com

[16] FranciscoJ Fraga,Letica Pimenta C,S prates, Alan M Marotta, "frequency lowering Algorithms for Hearing Impaired, "A Text book of Speech technologies", Book Chapter 18 , pp. 361 -388.

[17] Danielle Glista, Susan Scollie, Marlene Bagatto, Richard Seewald, Vijay Parsa, Andrew Johnson, "Evaluation of nonlinear frequency compression: Clinical outcomes". International Journal of Audiology 2009; 48: pp 632644

[18] Harry Levitt, "A Historical Perspective on Digital Hearing Aids: How Digital Technology Has Changed Modern Hearing Aids", Trends in Amplification Volume 11 Number 1,March 2007.pp 7-24

[19] Joshua M. Alexander, "Individual Variability in Recognition of Frequency-Lowered Speech", Seminars in Hearing, VOLUME 34, NUMBER 2 , Dec 2013

[20] Ying-Yee Kong, Ala Mullangi, "On the development of a frequency-lowering system that enhances place-of-articulation perception", Speech Communication 54 (2012) 147–160, Available online at www.sciencedirect.com.

[21] Pandurangarao N. Kulkarni , Prem C. Pandey , Dakshayani S. Jangamashetti "Multi-band frequency compression for improving speech perception by listeners with moderate sensorineural hearing loss" Speech Communication 54 (2012) 341–350 Elsevier 29 September 2011

[22] Manpreet Kaur, Amanpreet Kaur, “A Review: Different methods of segmenting a continuous speech signal into basic units” International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 11 November, 2013 Page No. 3184-3186

[23] Ruiyu Liang , Ji Xi , Jian Zhou , Cairong Zou , Li Zhao , An improved method to enhance high-frequency speech intelligibility in noise, Applied Acoustics 74 (2013) 71–78, www.elsevier.com/locate/apacoust

[24] Shilpi Banerjee, Starkey Hearing Research & Technology. An overview of the characteristics and applications of compression amplification. Process handbook Published in 2014

[25] J. Fraga Francisco, C. S. Prates Leticia Pimenta, M. Marotta Alan, Martinelli Iorio Maria Cecilia , Frequency Lowering Algorithms for the Hearing Impaired , ISBN 978-953-307-996-7.