Design, Improvement and Investigation of Neural Frequency Compression Method in Hearing Aid for Marathi Speech

Prashant G. Patil, Arun Kumar Mittra and Vijay S. Chourasia

Department of Electronics Engineering, Manoharbhai Patel Institute of Engineering and Technology, Gondia-441614, Maharashtra, India; patilpg232@gmail.com, akmittra@gmail.com, chourasiav@gmail.com

Abstract

Objectives: To design and evaluate neural frequency compression method to improve speech intelligibility for Marathi language hearing aid users. Methods/Statistical Analysis: In Recurrent Neural Network - Frequency Compression algorithm (RNN-FC), classification and processing are two important stages. After segmentation of input speech into discrete frames. Features are extracted in terms of signal to noise ratio, Pitch, formant frequency and gain frequency spectral coefficient. Extracted features will classify into two segments for processing and improvement in SNR level. Based on classification sample data is divided in two categories; wanted and unwanted samples for processing. Findings: Extracted feature vectors, Training date rate are key performance parameter of RNN-FC method. Testing of RNN-FC was performed on Marathi spoken HA user. In regional Marathi language 14-15 consonants are located over frequency band 7-13.5 kHz. Proposed algorithm shows improvement of classifier with Min 94- Max 96% sensitivity, Specificity and Accuracy. Results reports improved recognition rate of Marathi vowel, Consonants and short words. Unwanted vowel, consonants process reduced from 5.67% to 3.56%. The inability to access the high-frequency speech contents in terms of speech and consonant recognition ability enhanced for Marathi HA users. Application/Improvements: Frequency Compression method is extensively adopted by researchers in which high frequency speech is compressed by certain compression factor which causes distortion at lower speech frequencies. Distortion occurred during processing will results in loss of information in Lower band of speech. This challenge is overcomes by using frequency compression approach with neural network classifier (RNN-FC).

Keywords: Frequency Compression, Neural Network, Marathi, Speech

1. Introduction

Speech is a method of communication between people. From last 6-7 decades speech Enhancement with Processing and recognition is an attractive region for research. Speech is at the heart of human activity, as it helps humanity to collaborate more commonly and effectively. Hearing loss strongly affects society, which causes diseases at the initial detection level that affect the middle ear, or inner ear, age or slackness. A survey shows that India suffered from hearing loss in 1.06% of the total population. In 2014, the National Sample Survey (NSS) shows that the second major disability in India is hearing impairment. High - frequency hearing loss in urban areas is about 9% and 10.50% of total disability in rural areas. The degree of hearing impairment was determined depending on the level of a person’s inability to hear. According to the NSS report in Maharashtra, up to 2.291% of total disability is the number of people with high frequency hearing impairment. In rural areas, the percentage was higher (2.310%) than in urban areas (2.236%). The survey implies; about 0.4% infants suffered from hearing loss. Hearing is the primary sense of action through which we first learn speech and mother tongue.
is important to learn speech, language and auditory processing skills to hear clearly after 6 to 8 months of birth.

2. Hearing Aid and Current Technology

Hearing aid is widely used to recover the communication capability in hearing disabled persons. Hearing disabled persons with high-frequency hearing loss is not significantly benefitted from conventional HA. Limited bandwidth of the hearing device will hamper HA Accessibility of listeners to high-frequency signals which lead to difficulty for the identification of Consonants with a spectral crest higher than 5-7 kHz.2

Commonly a frequency Compression technique is used, which is categorized in Linear Frequency Compression (LFC) and non linear frequency compression (NLFC). LFC reduces all components of high frequency by the same constant (CF)2. New frequencies can be derived by multiplying a constant factor from the original frequencies (Compression factor). In this method high frequency speech is lowered without preserving low frequency speech contents. Vowels positioned after consonants are not preserved in LFC processing. Mean in NLFC as originally proposed3. High-frequency speech the components of the original speech are compressed disproportionately more than the low-frequency components. NLFC method was progressed and Modified in last few decades, now being widely used in numerous commercial Hearing aid Devices4,5. In case of multi frequency band input speech spectrum is sequentially segmented into a number of target bands where spectral samples in each band are compressed by a constant compression factor6, which results affect increased intra speech spectral masking7. Wen-Hsuan implemented NLFC approach for Mandarin speech in which Consonants relates to high frequency information is important for recognition of speech. Mandarin language has seven consonants located near to range of 10-16 kHz. Xianbo Xiao Implemented two FC and FT frequency-lowering methods and conducted weekly hearing tests to track the benefits of such methods, finally speech intelligibility improvement was found for Chinese language in their experiment8.

Compression based auditory critical bandwidth using spectral segment mapping shows better results with improvement up to 10-20% in the recognition score. Extended-Bandwidth (EB) using NLFC approach provides better results for recognition of Mandarin words9. They investigate the effects of NLFC and EB - NLFC on the recognition of mandarin words for high frequency hearing disabled10. Spectral Subtraction is a method where speech spectrum is divided into different continuous frequency bands with uniform spacing and spectral over - subtraction in each band, which is useful to improve speech recognition12.

Tobias Goehring, Federico Bolner Proposed neural based speech Enhancement in which an estimation of frequency channels contain more perceptually information (higher signal-to-noise ratio)13. Speech enhancement using NN shows better results, While NNSE was tested for noise-specific purpose.

In proposed methodology input speech is segmented into number of frames. From each frames feature vectors are extracted using NN. Signal to noise power of each frame is calculated by using spectral subtraction method, which will gives energy related each segmented frame. Higher Signal to noise power frame will be considered for compression, while lower SNR power will neglected for processing. Corresponding frame feature vector relates information to critical band FFT filter with

![Figure 1. Block diagram of RNN-FC algorithm.](image-url)
frequency range 50-3590 Hz. Frequency above 3590 Hz
compressed by certain compression factor (0.1-0.9).
Neural network training plays vital role in estimation
of accuracy, Specificity, Sensitivity and False acceptance,
Rejection rate. The results from the above studies are
promising but limited too few extend. Existing Methods
was tested for English and mandarin speaking HA users.
Our Proposedhybrid frequency Compression with Feed
Forward Neural Network basedtechnique shows improve-
ment in sensitivity, Specificity etc.

3. Developed Methodology

3.1 Recurrent Neural Network and
Frequency Compression Algorithm (RNN-FC)

The Recurrent neural network based frequency compres-
sor (RNN-FC) was designed by using MATLAB Block set
as shown in Figure 1.

Receiving speech from microphone is fed to Pre-
emphasis, which increases dBSPL of high frequency
bands and decrease the amplitudes of lower bands.

\[
y(t) = a.x(t) + (1-a)x(t) - 1
\]  

After pre-emphases input speech is segmented into
frames with the range from 20 ms to 40 ms with 50% (+/-
10%) overlap between consecutive frames. If the frame
length is short, then resolution of narrow band compo-
nents is sacrificed which affects frequency resolution and
if it is longer, signal properties changes that affects time
resolution Therefore standard 25 m-sec frame length
selected\(^\text{15}\). Feature extraction is the most important step
for speech intelligibility. Consider the speech signal sam-
ped at 16 kHz and quantized in 16 bits which utilized for
the feature extraction. Feature extraction was performed
on each segment of the noisy signal, and the output was
fed to the RNN. It takes the high dimensional character-
istic information into the low dimensional characterized
by the method of mapping or transformation. The trans-
formation from the input signal space to the feature space
is domain specific.

\[
S(n) = -\sum_{k=1}^{p} a^k s(n-k)
\]  

Linear Predictive Analysis method is used to extract
features from speech;it is used to compress the signal
without any loss of information. The prediction of cur-
rent sample as a linear combination of past \( p \) samples
form the basis of linear prediction analysis where \( p \) is
the order of prediction. The predicted sample \( s(n) \) can be
represented in eq. (2). Hamming window function used
for silence detection & pitch detection, Window function
expressed in eq. (3).

\[
W(N) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)
\]  

The primary objective of LP analysis is to compute
the LP coefficients which minimized the prediction
error ‘e(n)’. The total prediction error can be represented.
Predication error is function of time shown is Figure 2.

\[
E(n) = \sum_{n=-\infty}^{\infty} e^{2(n)}
\]  

\[\text{Figure 2. Time dependent linear predictive error ‘e(n)’}.
\]

The normalized error \( V(n) \) can be represented as

\[
V(n) = \frac{E(n)}{R}
\]  

The LPC gain Coefficient is given by

\[
G^2 = R0 - \sum_{k=0}^{p} \alpha k.Rk = En
\]  

\( E(n) \) is the minimum mean squared error prediction.

3.2 Key Performance Parameter

Following parameters will decide performance of neural
compression technique:
• Input/Output Function- Keeps processed speech at certain dBSPL with help of certain added Gain (G).
• Input/Gain Function- Gain is variable parameter which is decided by Input dBSPL.
• Frequency Response of processed speech with keeping unchanged shape.
• Role of Frequency and Gain needs to maximize speech intelligibility.
• Loudness Limiter to avoid uncomforted situation for HA user.
• Stationary features- threshold compression frequency, Compression factor.
• Variable features – Processing Start Point (PSP) and Processing End Point (PEP).
• Linearity and Non Linearity method of compression.

The Input speech Frequency range is divided into six bands according to following octaves shown in Table1.

| I-Octave | 0-250 Hz | IV-Octave | 1000-2000 Hz |
| II-Octave | 250-500 Hz | V-Octave | 2000-4000 Hz |
| III-Octave | 500-1000 Hz | VI-Octave | 4000-8000 Hz |

4. Design of Frequency Compressor with NN Approach

The key role of the RNN is to transform the inputs into meaningful outputs. Figure 3 shows Recurrent Neural network architecture; it consists of an input layer, 5 hidden layers and 1 linear featured output corresponding to input. Back propagation was used for training the NN in full-batch mode over 500 epochs with a variable learning rate of 0.01-0.03 and weight 1.2-0.2. No of iteration 100 with different Neuron Size.

In case of back propagation network, we are for changing the weights to classify the input patterns correctly. The selected hidden layer and number of neurons are utilized in this neural network. In back propagation Neural network first, we check error value (E) lies in range of threshold or not.

If E > Threshold Value (Th) ...............update all Weights,
Else repeat learning Procedure,

\[
\Delta w^p_k = -(T_k - Y_k)*h_j
\]  

(6)

\[
\Delta w^h_{ij} = -H_j(1 - H_j)S_i \sum_{k=1}^c [(T_k - Y_k)*w^p_{jk}] 
\]  

(7)

These six band octaves incorporate with frequency compressor, feature vector obtained from trained neural network will gives solution for selecting related octave for further processing. In this methodology training of neural network plays key role. Marathi Vowel and Consonant data Set is used for different training and testing Neural Network. Training and testing conditions are given from 50% - 90 % in the incremental stage of 10% in training, decrement of 10% in testing up to maximum level of 80:20.

5. Performance Parameters of System

Figure 4(a) Shows spectrogram for Marathi short word “aaj” on time scales of 1.2 sec. It was segment into a time-frequency frame unit of 0.64 sec. with 0-5000 Hz Speech frequency. Input Word was processed by frequency compression technique, the spectrogram of frequency compressed word shown in Figure 4(b) frequency above 4000 Hz compressed where overlapping of spectrum occurs over low speech frequency. Figure 4(c) shows spectrogram of Marathi word compressed by neural compression technique. Using these method frequencies above 4000 Hz is compressed with preserving lower speech frequency contents.
RNN-FC algorithm were developed in MATLAB (The Math Works) and tested on regional Marathi spoken Hearing aid users.

Marathi consonant “cha” Spectrum spoken by Female Speaker shown in Figure 5(a) which has pitch frequency range from 2KHz to 12.5 kHz. Figure 5(b) shows segmented speech by using RNN+FC method and Figure 5(c) shows spectrum difference between original and processed spectrum using proposed method. Sensitivity and specificity are statistical measures of the performance of RNN-FC system, classification of processed and unprocessed vowel/consonant is related to sensitivity, Specificity. Sensitivity is an ability of detection which measures the proportion of positives processed alphabets, which are correctly identified for the compression purpose. Using this methodology sensitivity changes from 94.23% to 96.552% (Unwanted vowel, consonants processed from 5.67% to 3.56% of Trained Data). Mean while Specificity measures the proportion of negatives that are correctly identified, specificity changes from 5.26% to 31.25% (amount of Vowels and consonants left from processing) average accuracy for proposed methodology in the range of 68-76%. The Matthews’s Correlation Coefficient (MCC) is used learning as a measure of the quality of classifications in terms of vowel and consonants. Training NN from lowest to highest training data gives MCC from 0.21-0.37.

The objective of this methodology was to design and test the usefulness of Neural network based multi-band frequency compression in improving speech Recognition by Hearing aid listeners, HA listeners was selected with high frequency hearing loss ranging between 2 to 4 KHz. Developed Matlab based algorithm was tested on Intel core- 2 processor laptop with help of earphone. A group of six listeners were involved in this experimentation;
recognition rate of proposed method was compared with frequency compression, frequency transposition as shown in Table 2. In Marathi language 2-3 group of consonants has same pronunciation; lip movement which is difficult to understood by HA users. This consonant group recognition is tested by using RNN-FC method. Each vowel from 15 vowels (Total 45 vowels using FC, FT, NN-FC) are play backed randomly to measure recognition score. Same method was carried for consonant and other words.

|                | FC | FT | RNN-FC |
|----------------|----|----|--------|
| vowels         | 12/15 | 11/15 | 10/15  |
| consonants     | 32/45 | 31/45 | 38/45  |
| short words    | 12/20 | 11/20 | 9/20   |
| confusing words| 26/40 | 27/40 | 32/40  |

6. Conclusions and Outcomes

A neural network based frequency approach shows improvement in terms of specific characteristics of classification for processed and unprocessed alphabets. Training given to neural network is complex and time consuming task but it shows remarkable improvement in way of classification and decision making for processing. In Marathi language many consonants lies in same range of pitch frequency, so distinguish them for further processing is challenging issue. Using the proposed algorithm, the classification of these consonants was done effectively. This approach is useful to avoid unwanted processing of Marathi alphabets; this will be helpful to improve speech intelligibility while recognizing vowel, consonants, words, short sentences and confusing words for Hearing aid users.

7. Acknowledgement

This research work was carried out at the Priyadarshini Deaf Residential School, Shirpur Dist-Dhule (India). We have obtained all ethical approvals from an appropriate ethical committee of institutional review board. If any issue arises hereafter then we will be solely responsible. The Scientific responsibility is assumed by the authors. The Institute had given the researcher permission to use our data as part of their experimental study. Institute has no objection to publish the experimental data in conference or journal Paper.

8. References

1. Mao Y, Yang J, Hahn E, Xu L. Auditory perceptual efficacy of nonlinear frequency compression used in hearing aids: A review, Journal of Otology. 2017; 12(3):97–111. https://doi.org/10.1016/j.joto.2017.06.003. PMid: 29937844, PMCid: PMC5963461.
2. Liang R, Xi J, Zhou J, Zou C, Zhao L. An improved method to enhance high-frequency speech intelligibility in noise, Applied Acoustics. 2013; 74:71–78. https://doi.org/10.1016/j.apacoust.2012.06.010.
3. Sekimoto S, Saito S. Nonlinear frequency compression speech processing based on the PAR-COR analysis-synthesis technique, Annual Bulletin of the RILP. 1980; 14:65–72.
4. Desai N, Dhameliya K, Desai V. Feature extraction and classification techniques for speech recognition: A review, International Journal of Emerging Technology and Advanced Engineering. 2013; 3(12):1–5.
5. Duarte MF, Baraniuk RG. Spectral compressive sensing, Applied and Computational Harmonic Analysis. 2013; 35:111–29. https://doi.org/10.1016/j.acha.2012.08.003.
6. Kulkarni PN, Pandey PC, Jangamashetti DC. Multi-band frequency compression for improving speech perception by listeners with moderate sensorineural hearing loss, Speech Communication. 2012; 54:341–50. https://doi.org/10.1016/j.specom.2011.09.005.
7. Vihari S, Murthy S, Soni P. Comparison of Speech Enhancement Algorithms. International Multi-Conference on Information Processing-2016, Procedia Computer Science. 2016; 89:666–76. https://doi.org/10.1016/j.procs.2016.06.032.
8. Tseng WH, Hsieh DL, Shih WT, Liu TC. Extended bandwidth nonlinear frequency compression in Mandarin-speaking hearing-aid users, Journal of the Formosan Medical Association. 2017.
9. Xiao X. Evaluation of frequency-lowering algorithms for intelligibility of Chinese speech in hearing-aid users, Progress in Natural Science. 2009; 19:741–49. https://doi.org/10.1016/j.pnsc.2008.08.007.
10. Liu YT, Tsao Y, Chang RY. Non Negative matrix factorization based frequency lowering technology for mandarin Speaking Hearing aid users. 2016.
11. Liu YT, Chang RY, Tsao Y, Chang Y. A New Frequency Lowering Techniques for Mandarin Speaking hearing aid users. IEEE Global Conference on Signal and Information Processing (Global SIP); 2015. p. 722–26. https://doi.org/10.1109/GlobalSIP.2015.7418291.
12. Upadhyaya N, Karmakar A. An improved multi-band spectral subtraction algorithm for enhancing speech in various noise environments, International Conference On
Design and Manufacturing. 2013; 64:312–21. https://doi.org/10.1016/j.proeng.2013.09.103.

13. Goehring T, Bolner F. Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users, Hearing Research. 2017; 344:183–94. https://doi.org/10.1016/j.heares.2016.11.012. PMCID: PMC5256482.

14. Icer S. Classification with the neural network application of basic hearing losses determined by audiometric measuring, Journal of Networking Technology. 2010; 1(2):1–6. http://www.dline.info/jnt/fulltext/v1n2/1.pdf.

15. Bondya J, Beckerb S, Brucea I, Trainorb L, Haykina S. A novel signal-processing strategy for hearing-aid design: Neuro compensation, Signal Processing. 2004; 84:1239–53. https://doi.org/10.1016/j.sigpro.2004.04.006.