An effective frame-based high frequency speech transposition by using neural network

Prashant G. Patil*, Arun K. Mittra and Vijay S. Chourasia

Department of Electronics Engineering,
Manoharbhai Patel Institute of Engineering and Technology,
Gondia (MS), India
Email: patilpg232@gmail.com
Email: akmittra@gmail.com
Email: chourasiav@gmail.com
*Corresponding author

Abstract: This paper investigate design methodology and performance of neural network based frequency transposition algorithm for hearing aid users. High frequency hearing loss associated with hearing disabled person is promising issue for research. Frequency compression and frequency transposition schemes are key solution to overcome high frequency hearing loss. Neural approach to frequency transposition makes algorithm more sensitive, accurate and specific towards processing. The proposed neural network frequency transposition (NNFT) algorithm is based on framing of speech into feature vector for NN with comprehensive training and processing. The parameter to set in NNFT algorithm was calculated by evaluative study. Using these algorithm Marathi alphabets, words, confusing words are efficiently classified. Classification will improve acceptance and rejection rate for FT processing. Validation and testing result of algorithm shows improvement in sensitivity, accuracy, specificity of NNFT method compared to FT method.

Keywords: neural network; frequency transposition; speech frames; FFT; hearing loss; sensitivity.

Reference to this paper should be made as follows: Patil, P.G., Mittra, A.K. and Chourasia, V.S. (2018) 'An effective frame-based high frequency speech transposition by using neural network', *Int. J. Intelligent Systems Design and Computing*, Vol. 2, No. 1, pp.88–98.

Biographical notes: Prashant G. Patil is a Research Scholar in Department of Electronics Engineering at Manoharbhai Patel Institute of Engineering and Technology, Gondia (MS), India. His researches focus on speech processing, and biomedical signal processing.

Arun K. Mittra is a Professor and Head in Department of Electronics Engineering at Manoharbhai Patel Institute of Engineering and Technology, Gondia (MS), India. His area of research for doctoral studies was medical signal processing on fatal heart rate statistical analysis. His research interests concern with biomedical signal processing and speech processing. He has published 29 research papers in various national and international journals.

Vijay S. Chourasia is an Assistant Professor in Department of Electronics and Communication Engineering at Manoharbhai Patel Institute of Engineering and Technology, Gondia (MS). His area of research for doctoral studies was an antenatal care system using abdominal acoustic signals. His research interests concern signal and image processing and artificial intelligence. He has published 40 research papers in various national and international journals.

1 Introduction

High frequency hearing loss affects the sensitivity of speech at high frequencies. Certain level of loss is significant then speech perception becomes especially difficult, as many Marathi consonants locate near to high-frequency information. A high-frequency hearing loss ranging from an average threshold across 4,000 to 8,000 Hz greater than or equal to 75–90 dB HL. Hearing aid user with severe high-frequency hearing loss when listening without assistive devices feels tricky to discriminate certain groups of Marathi words, e.g., vkt&rkt(pky&dky. Currently, frequency compression and frequency transposition methods are adopted by many hearing aid manufacturers. High frequencies amplification will become audible results in an increase in speech understanding (Skinner, 1980). Hearing loss is related with a high-frequency dead region (DR) of hearing disabled person a region in the cochlea where inner hair cells and/or neurons are functioning so poorly that a tone producing peak basilar membrane vibration in that region is detected by off-place listening (Baer et al., 2002) then amplification of frequencies above 1.7.fe where fe can be defined in terms of the characteristic frequency of the inner hair cells or neurons immediately adjacent to the DR (Desai et al., 2013; Baer et al., 2002) will not advantageous (Moore et al., 2000; Mao et al., 2017). In such cases, frequency components above 1.7fe potentially provide more information for speech understanding than components between fe and 1.7fe. For fe values of 0.5, 0.75 and 1.0 kHz is selected (Moore et al., 2000). To make high frequency components audible to someone with a DR at high frequency, frequency transposition is an effective method where information from high to low frequencies could provide a solution to making this information audible. Many researchers focused to develop this method, review of transposition studies done by Sekimoto and Saito (1980) and Duarte and Baraniuk (2013). FT method in which they converted frequency components above 3,000 KHz to broadband noise using a no-linear structure modulator (Tseng et al., 2017; Xiao et al., 2009). Kulkarni et al. (2012) and Goehring and Bolner (2017) adopted same method but results show that low-frequency speech cues would sometimes be partially masked by the transposed signal, background noise is also transposed which will make speech complex at lower target band of transposition. Frequency re-coding device (FRED) was designed in which they subtracts 4,000 KHz from every frequency component between 4,000 and 8,000 KHz (Vihari et al., 2016; Liu et al., 2016). In this method, the transposed frequency components will cover or obstruct with the band of the unaltered low-frequency speech components. The unconditional transposition will results in background noise induced in high-frequency components. Current transposition method is different first we used FFT-based algorithm where transposition is conditional without slow playback and frequency compression then all subjects were tested with DRs and fe (Liu et al., 2015; Robinson et al., 2007).

Figure 1 Modified block diagram of neural network adopted frequency transposition (NNFT) method

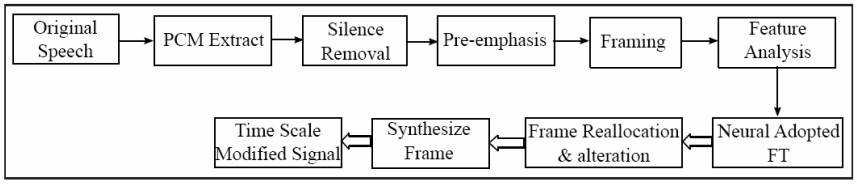
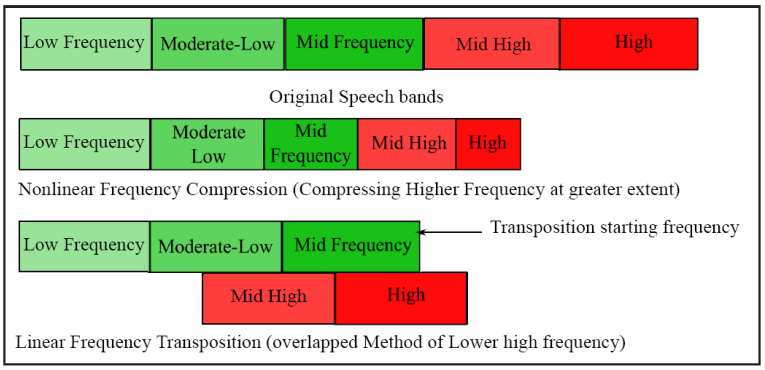
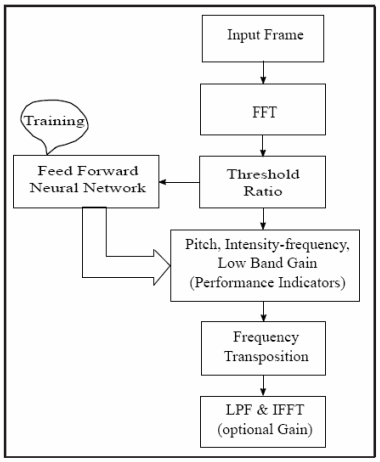


Figure 2 (a) Basic frequency transposition (b) Neural approach to frequency transposition scheme (see online version for colours)



(a)



(b)

Proposed method uses combination approach of FFT and neural network shown in Figure 2 in which features extracted from each frame is divided into two categories; under band transposed and over band transposed (Upadhyaya and Karmakar, 2013). Target and source band are calculated according to DR calculation of high frequency

region. Feature vector as format frequency, pitch and intensity-frequency level were extracted from each frame. These feature vector given as input to neural network, training and validation testing are key factors performance parameter of NN. Trained NN output decides under transposed or over transposed condition for algorithm. Set of different 'fe' are selected for every hearing aid user for trial purpose and these process is repetitive still to find optimum performance of algorithm.

2 Proposed neural frequency transposition approach

Block diagram for proposed feed forward back propagation neural network based frequency transposition in Figure 1. To represent digitally sampled analog signals pulse-code modulation (PCM) method used. Endpoint detection is used to remove the DC offset value from the signal after silence removal process. Speech is classified into voiced or silence/unvoiced sounds in terms of fundamental frequency estimation, formant extraction or syllable marking. There are several ways of classifying (labelling) events in speech. It is accepted resolution to use a three-state representation in which states are silence, unvoiced, and voiced (V). The input speech signal $s(n)$ is given to a high-pass filter.

$$s^2(n) = s(n) - \alpha \cdot s(n-1) \tag{1}$$

where $s^2(n)$ is the filtered output signal and *alpha* is usually between 0.9 and 1.0. The z-transform of the filter is given in equation (2)

$$H(z) = 1 - \alpha \cdot z^{-1} \tag{2}$$

Figure 3 Variation effect of 'α' on hamming window function (see online version for colours)

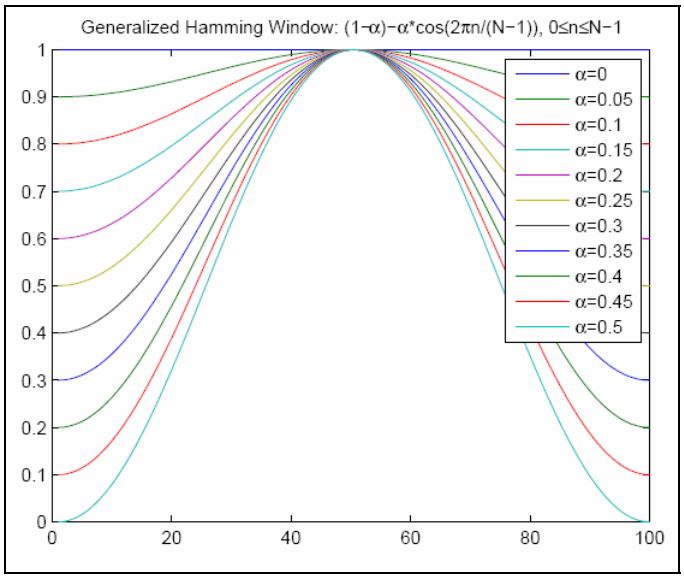
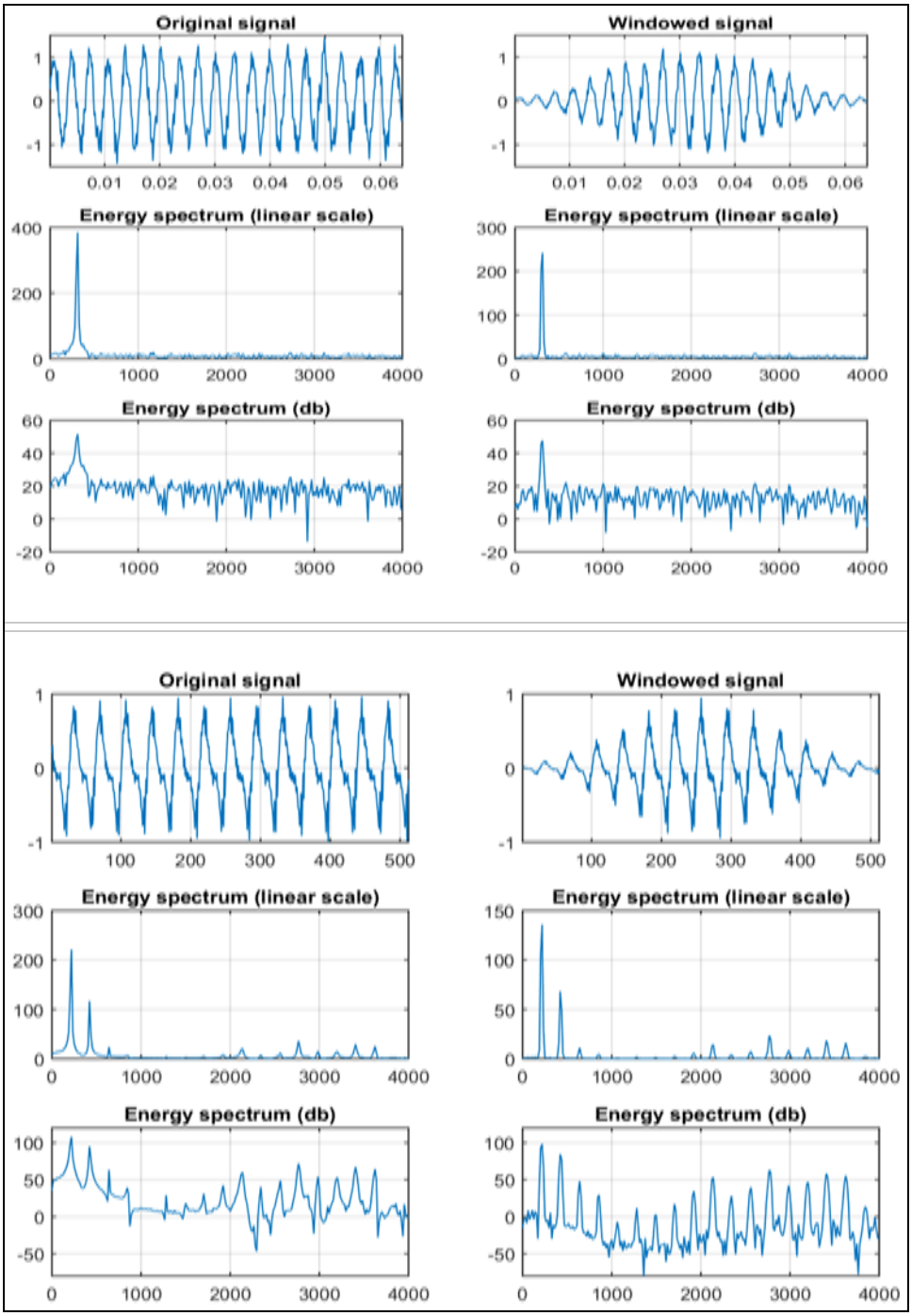


Figure 4 Hamming windowing effect on energy spectrum at first and last point on frame (see online version for colours)



First and the last points in the frame is in continued way with help of hamming window, where hamming function is multiplied with frame (Bondya et al., 2004). Signal frame is given by $s(n)$, $n = 0 \ldots N - 1$, Signal after Hamming windowing is $s(n) * w(n)$, where $w(n)$ is the Hamming window defined by equation (3), Figure 3 shows that hamming window function is key dependent of α sharper response is obtained at mid value of α.

$$W(n, \alpha) = (1 - \alpha) - \alpha \cdot \cos\left(\frac{2\pi n}{N - 1}\right) \tag{3}$$

where $0 < n < (N - 1)$.

The input speech signal is segmented into frame size of 20 ms with 33.33% to 50% overlap. Frame size is given by 2 N to enable the use of FFT. Sample rate is 16 kHz, frame size is 320 sample/points, then the frame duration is 320/16,000 = 0.02 sec = 20 msec with overlap of 50%, 160 points so effective frame rate of system is 100 frames/second. Different pitches in speech signals resembles to dissimilar energy distribution over frequencies. Magnitude frequency response of each frame is obtained by FFT. Wrapping factor, periodic nature and continuity are features of FFT operation (Liang et al., 2013). FFT will leads continuity at the frame's (start and end points) will introduce adverse effects in the frequency response. These will overcome by multiplying each frame by a Hamming window to increase its continuity at the first and last points or by taking variable frame size with integer multiple of vital period of speech. Variable frame size is ineffective as difficulty occurs in fundamental period calculation in practical approach. Hamming window function makes each frame sharper and separated at point. Figure 4 shows effect activated and non-activated functioning of the hamming window on each frame.

3 Neural network parameter

FFNN is used to transform the inputs into significant outputs. Feed forward neural network architecture has one input layer, five hidden layers and five linear featured outputs corresponding to speech frame (gives a condition in terms of under transposed and over transposed).For training of the NN in full-batch mode over 13 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. Training of data varies from 50%to 95%, other data set is used for validation of NN algorithm. Figure 5 shows regression plot for training and validation at different values (Icer, 2010). Training case of approximately 95% and 91% comparison implies that more fitted with minimum error. Target and training rate ultimately decides performance parameter like sensitivity, specific approach towards processing, accuracy, false acceptance and rejection rate, Positive and negative predicted value of NN-based FT algorithm. It is observed that more percentage training data will make NNFT system effective. False acceptance rate should decide unwanted processing of speech frame this is challenge to overcome in future. Figure 6 indicates, increase in training data will time consuming task but more improvement in sensitivity, accuracy. Significant improvement was found in all parameters with increase in 2–2.5% compared to traditional FT approach.

Figure 5 Regression plot for training, validation, testing of feed forward neural network (FFNN) (see online version for colours)

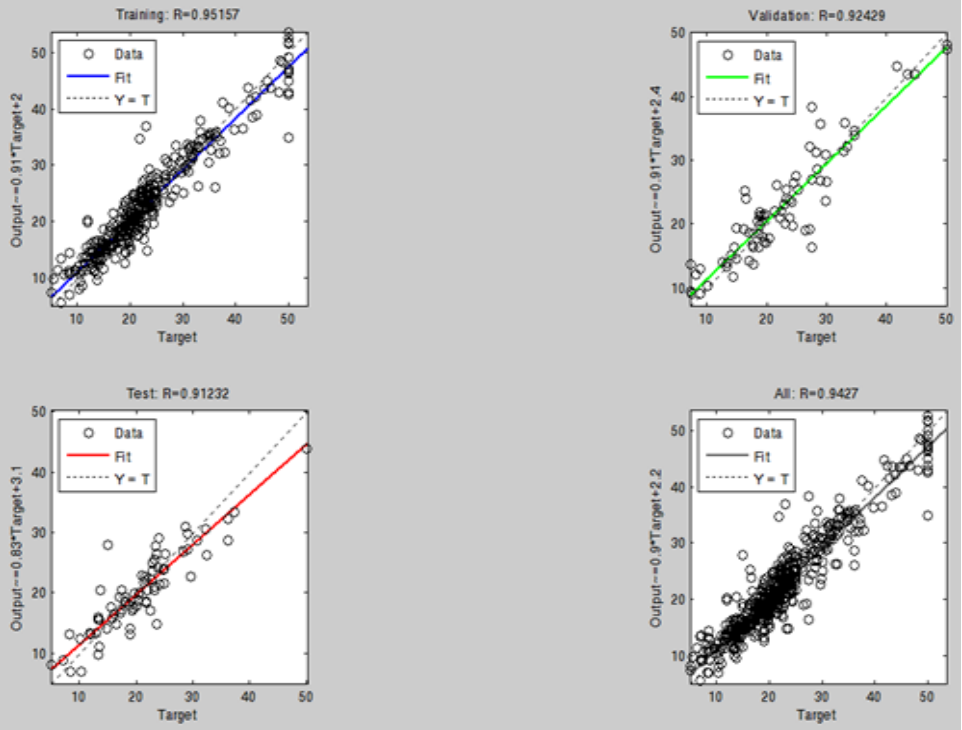


Figure 6 Key performance parameters of feed forward neural network (sensitivity, specificity, accuracy, FPR, FNR, FAR, FRR, PPV, NPV) (see online version for colours)

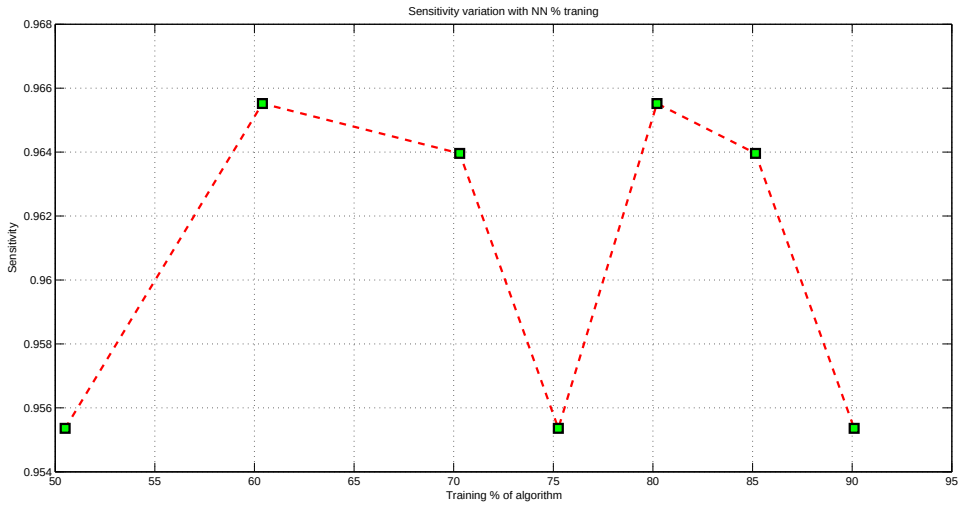
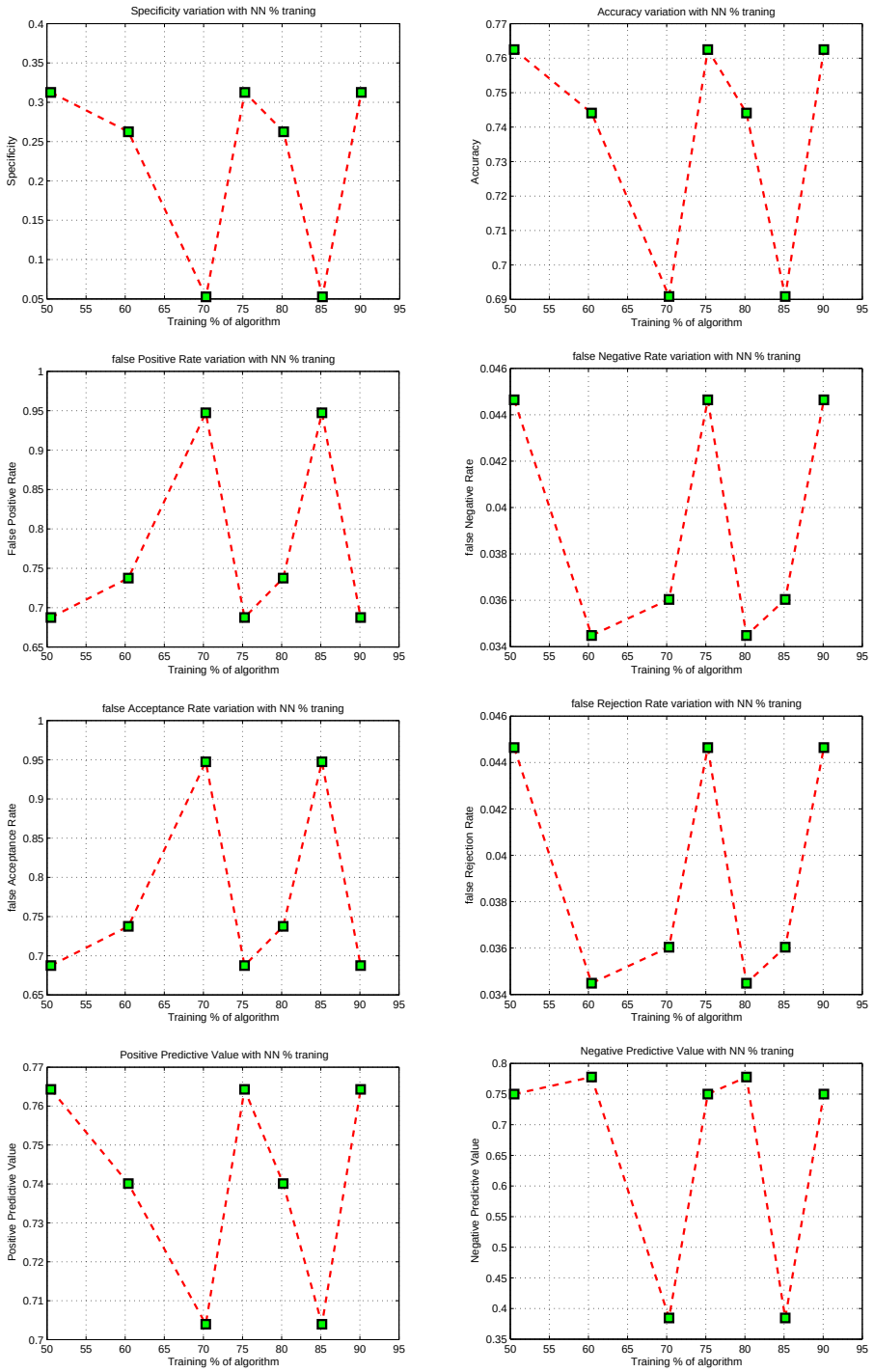


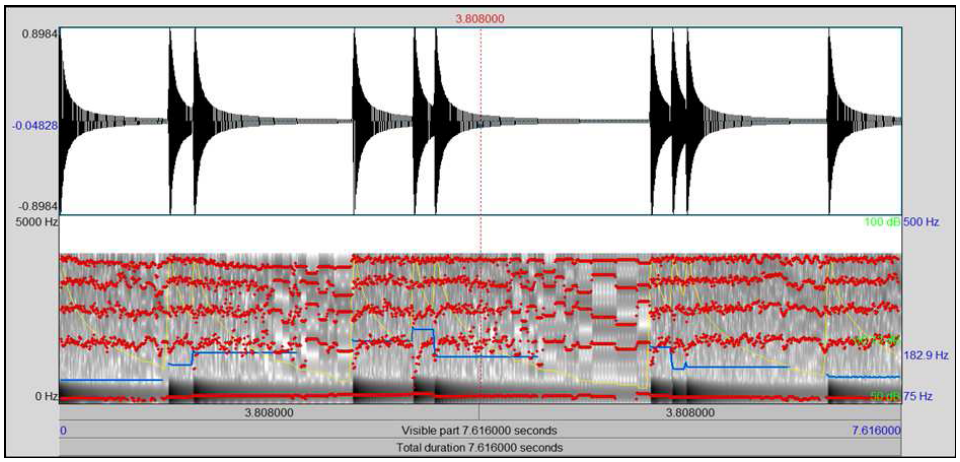
Figure 6 Key performance parameters of feed forward neural network (sensitivity, specificity, accuracy, FPR, FNR, FAR, FRR, PPV, NPV) (continued) (see online version for colours)



4 System validation and testing results

The presence and extent of their DRs using the audio-gram signal was fixed in frequency and level, and the masker level required to mask the signal was determined as a function of masker centre frequency. Frequency above 4,000 Hz is transposed over 0–4,000 Hz as shown in Figure 7. Spectrogram of transposed speech shows darker area with red dot which occurs mislaid in low frequency speech information. More amplification in range of 20–30 dB is needed for frequency range 1,500–3,500 Hz. It is observed that low speech frequency components overlapped in greater extent (0–570 Hz).

Figure 7 Transposed speech frequency spectrum for Marathi spoken word (see online version for colours)



5 Conclusions and future scopes

Objective to use NN is to avoid unwanted processing, improve better classification of processing and to find optimal set of parameters for transposition action on each frame. Performance parameter has in the range of minimum 90% maximum 94%. All validation and performance parameter was decided by way of training and percentage of data used from each frame to train neural network. Source band, target band and fe was calculated from patient requirement, with this approach process become time consuming and complex which will make the system performance with considerable delayed. Testing of system was carried for Marathi hearing aid user, in Marathi language most consonant have same frequency parameter. Classification of transposed and un-transposed data is difficult. Based on this classification false acceptance rate increased and unnecessary alphabets, words are processed. Performance results may vary for other language. Proposed neural-based frequency transposition algorithm shows improvement in sensitivity, accuracy, specificity.

References

Baer, T., Moore, B.C.J. and Kluk, K. (2002) 'Effects of low-pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies', *J Acoust Soc Am*, Vol. 112, No. 4, pp.1133–1144.

Bondya, J., Beckerb, S., Brucea, I., Trainorb, L. and Haykina, S. (2004) 'A novel signal-processing strategy for hearing-aid design: neuro compensation', *Signal Processing*, Vol. 84, No. 3, pp.1239–1253.

Desai, N., Dhameliya, K. and Desai, V. (2013) 'Feature extraction and classification techniques for speech recognition: a review', *International Journal of Emerging Technology and Advanced Engineering*, Vol. 3, No. 12, pp.367–371.

Duarte, M.F. and Baraniuk, R.G. (2013) 'Spectral compressive sensing', *Applied and Computational Harmonic Analysis*, Vol. 35, No. 2, pp.111–129.

Goehring, T. and Bolner, F. (2017) 'Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users', *Hearing Research*, Vol. 344, No. 2, pp.183–194.

Icer, S. (2010) 'Classification with the neural network application of basic hearing losses determined by audiometric measuring', *Journal of Networking Technology*, June, Vol. 1, No. 2, pp.9–16.

Kulkarni, P.N., Pandey, P.C. and Jangamashetti, D.S. (2012) 'Multiband frequency compression for improving speech perception by listeners with moderate sensorineural hearing loss', *Speech Communication*, Vol. 54, No. 3, pp.341–350.

Liang, R., Xi, J., Zhou, J., Zou, C. and Zhao, L. (2013) 'An improved method to enhance high-frequency speech intelligibility in noise', *Applied Acoustics*, Vol. 74, No. 1, pp.71–78.

Liu, Y-T., Chang, R.Y., Tsao, Y. and Chang, Y-P. (2015) 'A new frequency lowering techniques for mandarin speaking hearing aid users', *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*.

Liu, Y-T., Tsao, Y. and Chang, R.Y. (2016) 'Non-negative matrix factorization based frequency lowering technology for mandarin speaking hearing aid users', *ICASSP 2016*.

Mao, Y., Yang, J., Hahn, E. and Xu, L. (2017) 'Auditory perceptual efficacy of nonlinear frequency compression used in hearing aids: a review', *Journal of Otology*, Vol. 12, No. 3, pp.97–111 [online] http://dx.doi.org/10.1016/j.joto.2017.06.003 (accessed 26 July 2017).

Moore, B.C.J., Huss, M., Vickers, D.A., Glasberg, B.R. and Alca'ntara, J.I. (2000) 'A test for the diagnosis of dead regions in the cochlea', *Br J Audiol*, Vol. 34, No. 4, pp.205–224.

Robinson, J.D., Baer, T. and Moore, B.C.J. (2007) 'Using transposition to improve consonant Discrimination and detection for listeners with severe high-frequency hearing loss', *International Journal of Audiology*, Vol. 46, No. 6, pp.293–308.

Sekimoto, S. and Saito, S. (1980) 'Nonlinear frequency compression speech processing based on the PAR-COR analysis-synthesis technique', *Ann. Bull. RILP*, Vol. 14, p.65e72.

Skinner, M.W. (1980) 'Speech intelligibility in noise-induced hearing loss: effects of high frequency compensation', *J Acoust Soc Am*, Vol. 67, No. 1, pp.306–317.

Tseng, W-H., Hsieh, D-L., Shih, W-T. and Liu, T-C. (2017) 'Extended bandwidth nonlinear frequency compression in Mandarin-speaking hearing-aid users', *Journal of the Formosan Medical Association*, Vol. 11, No. 1, pp.109–116 [online] http://dx.doi.org/10.1016/j.jfma.2017.01.013 (accessed 10 September 2017).

Upadhyaya, N. and Karmakar, A. (2013) 'An improved multi-band spectral subtraction algorithm for enhancing speech in various noise environments', *International Conference On Design and Manufacturing, Procedia Engineering*, Vol. 64, pp.312–321.

Vihari, S., Murthy, S. and Soni, P. (2016) 'Comparison of speech enhancement algorithms', *International Multi-Conference on Information Processing, Procedia Computer Science*, Vol. 89, pp.666–676.

Xiao, X. et al. (2009) 'Evaluation of frequency-lowering algorithms for intelligibility of Chinese speech in hearing-aid users', *Progress in Natural Science*, Vol. 19, No. 6, pp.741–749.