Speech Filters for Speech Signal Noise Reduction

E.S Kasthuri and A.P. James
School of Computer Science and Information Technology
Center for Excellence in Applied Machine Intelligence and Pattern Analysis,
Indian Institute of Information Technology and Management- Kerala

ABSTRACT

The implementation complexity of the conventional speech enhancement techniques increases with increased sampling rates and increased levels of noise. In order to address this issue, we propose a hardware friendly perceptive speech filter implemented using RLC filters. The proposed filters when compared with the conventional filterbanks such as based on Mel and Bark scale show significant reduction in noise levels as measured through the distance distributions.

General Terms:

Speech Enhancement, Automatic Speech Recognition

Keywords:

Filtering, RLC filters, Speech, Noise Reduction

1. INTRODUCTION

Noise in the speech signal can significantly reduce the performance of automatic speech recognition systems[1]. Preserving the speech content and reducing the noise present in a recorded speech signal is essential to improve the recognition performance of automatic speech recognition systems. In addition, the ability to implement the speech enhancement techniques in a real-time hardware[2] is important for large-scale and high speed speech processing applications. However, majority of present day approaches to noise reduction and signal enhancements are difficult to implement in hardware, due to the increase in design complexity and limitations of semiconductor process technology. In order to address the issues, we present a hardware friendly perceptive speech filters implemented as RLC filterbank in a view to present them as a front-end for a speech enhancement system.
The proposed speech enhancement system is biologically-inspired such that the main part of this front-end is a bank of filters with bandwidths in log scale, that resemble the processing of sounds by the human cochlea [12].

2. PROPOSED METHOD

Figure 1 shows the block diagram describing the speech signal enhancement system using the proposed perceptive series RLC filter bank. In the proposed system the filter bank consists of 50 discrete time series RLC filters, where the bandwidth of the successive filters are increasing logarithmically and these 50 filters together cover the entire audio spectrum. The voiced speech signal is applied to the filter set. The filters designed are discrete time domain filters. The short time fourier transforms (STFT) of the filter outputs are then calculated to generate the spectrograms[9], i.e., by using short sized sliding windows, the fast fourier transforms (FFT) of the filter outputs are calculated to transfer the time domain information into frequency domain. The FFT values corresponding to the bandwidth of each of the filters are extracted from the respective filter spectrograms. The values extracted from the spectrograms, ie from the 50 spectrograms, are appended vertically to form the final spectrogram. The FFT values extracted from the spectrograms correspond to the high gain region of each of the filters. Thereby the system ensures the quality of the speech perception.

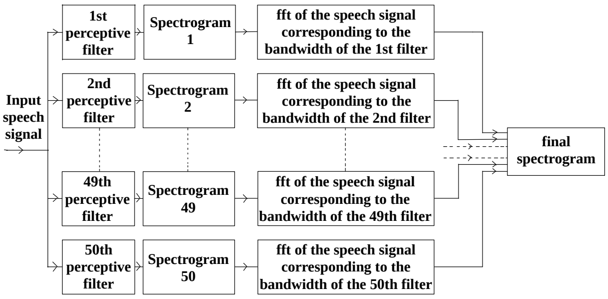
Input speech signal
1st perceptive filter
Spectrogram 1
fft of the speech signal corresponding to the bandwidth of the 1st filter
2nd perceptive filter
Spectrogram 2
fft of the speech signal corresponding to the bandwidth of the 2nd filter
49th perceptive filter
Spectrogram 49
fft of the speech signal corresponding to the bandwidth of the 49th filter
50th perceptive filter
Spectrogram 50
fft of the speech signal corresponding to the bandwidth of the 50th filter
final spectrogram

Fig. 1. Block diagram representing the speech enhancement technique.The voiced speech signal is given to 50 different perceptive RLC filters.The bandwidth of successive filters are increasing logarithamically and they cover the entire audio spectrum.A plot of log energy across time and frequency is obtained by taking the spectrograms of all the filter outputs of the speech signal.FFT values corresponding to the bandwidth of each of the filters are extracted from the spectrograms.The obtained values are vertically concatenated in accordance with the bandwidth of all the filters to form the final spectrogram.

2.1 Speech Processing

A typical speech sentence signal consists of two main parts: one that carries the speech information, and the other that includes silent or noise sections that are between the utterances, without any verbal information. The verbal (informative) part of speech can be further divided into two categories: (a) the voiced speech and (b) the unvoiced speech. Voiced speech consists mainly of vowel sounds. It is produced by forcing air through the glottis, proper adjustment of the tension of the vocal cords results in opening and closing of the cords, and a production of almost periodic pulses of air. These pulses excite the vocal tract. Psychoacoustics experiments [8] show that this part holds most of the information of the speech and thus holds the keys for characterizing a speaker. Unvoiced speech sections are generated by forcing air through a constriction formed at a point in the vocal tract (usually towards the mouth end), thus producing turbulence. A male voiced speech sentence signal in WAV file format having 19374 samples in 1 channel was used as the input signal to the enhancement system.The duration of the speech is 1.2109 seconds. Figure 2 shows the speech signal input used for the simulations.

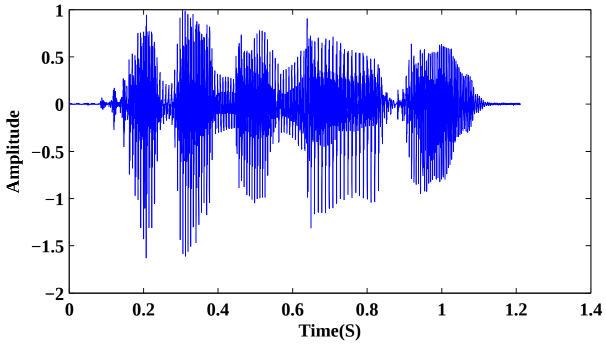

Fig. 2. Voiced speech signal recorded as .wav file.The wav file contains 19374 samples in 1 channel.The duration of the speech is 1.2109 seconds

2.2 Perceptive RLC Filters

The speech signal is applied to filterbank and these 50 filters in the bank are discrete time domain filters. The filters can be represented by difference Eq (1):

$$y(n) = -\sum_{k=1}^{N} a_k y(n-k) + \sum_{k=0}^{M} b_k x(n-k) \tag{1}$$

From this equation, note that $y(n-k)$ represents the outputs and $x(n-k)$ represents the inputs, $a_k, k = 1, 2...N$, $b_k, k = 1, 2...M$ are called the filter coefficients. The value of N represents the order of the difference equation and corresponds to the memory of the system being represented. The filter bank covers the frequency spectrum from 0Hz to 3.874KHz. Bandwidth of the first filter is 3Hz and for the following filters it is increasing logarithmically. The filter transfer function is expressed as the ratio of laplace transform of the output current to the input voltage for series RLC circuit:

$$\frac{(s/L)}{(s^2 + Rs/L + 1/LC)} \tag{2}$$

The transfer function spectral characteristics of any filter tell us how the filter influences input signals at different frequencies. From the transfer function of the RLC filter the characteristic equation of the filter can be written as:

$$(s^2 + Rs/L + 1/LC) = 0. \tag{3}$$

The bandwidth of the perceptive filter is R/L rad/s and the center frequency is $\sqrt{1/LC}$. The circuit set up of a perceptive RLC filter is shown in Figure 3. The inductance value is kept constant as $1H$, and then by varying the resistance and capacitance values in accordance with the characteristic equation the 50 perceptive RLC filters of required bandwidth and centre frequency are designed. Frequency response of all the 50 filters are shown in Figure 4. In particular, it should be noted that the overlap bands are truncated to ensure that the high amplitude responses are only passed. Time domain response of the 25th filter in the filter set to the speech signal is shown in Figure 5.

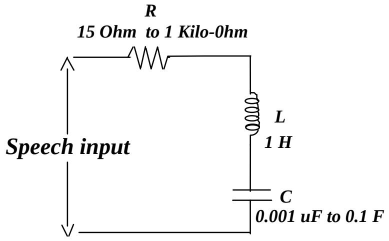

Fig. 3. Perceptive RLC filter circuit for the speech input.The inductance value is fixed as 1H. By varying the resistance values from 15 Ohm to 1 Kilo-ohm and the capacitance values from 0.001uF to 0.1F all of the 50 filters are designed. Depending on the center frequency and bandwidth, each filter gives different outputs for the voiced speech input

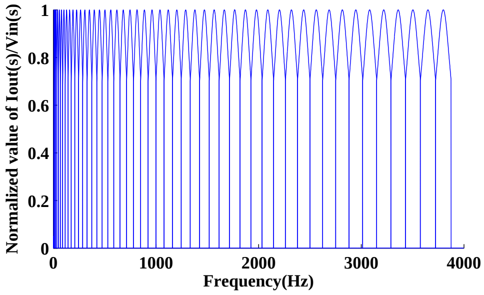

Fig. 4. Normalized gain versus frequency of 50 perceptive RLC filters. The bandwidth of the first filter is 3Hz and it is increasing logarithamically for successive filters.Frequency spectrum of the filter bank varies from 0Hz to 3.874KHz

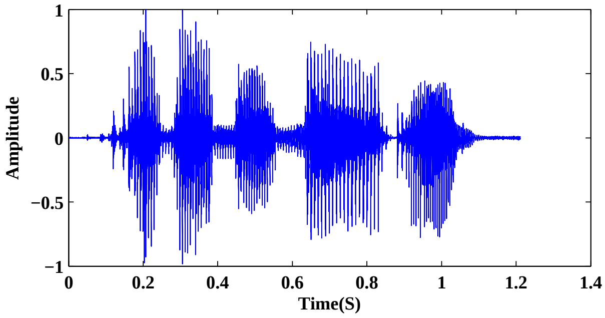

Fig. 5. Time domain response of 25th perceptive RLC filter.

2.3 Spectrogram Measurements

The filter outputs of the speech signal are in time domain. For measuring the enhancement achieved in the quality of the speech signal the time domain data need to be transformed into frequency domain. For transforming the time domain data into frequency domain we use fourier transformation. Since for non-stationary signals, whose statistic characteristics vary with time, the classic Fourier transform is not very suitable for analysis. It cannot provide information on how the frequency changes over time. Short-time Fourier transform (STFT)[7], is a method of analysis used for analyzing non-stationary signals[3]. It extracts several frames of signals with a window that moves with time. If the time window is sufficiently narrow, each extracted frame can be viewed as stationary such that Fourier transform can be applied. With the window moving along the time axis, the relation between the variance of frequency and time is identified. STFT performed on a sequence, $x[n]$, can be defined as:

$$STFT\{x[n]\} \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]\omega[n-m]e^{-j\omega n} \quad (4)$$

where $\omega[n]$ represents the sliding window that emphasizes local frequency components within it.

In the proposed system, 50 different spectrograms are calculated from the respective filtered outputs of the speech signal. The window used was Kaiser window with length 500. The overlap maintained for the signal was 5 and the sample rate selected was 8000Hz. Figure 6 shows the spectrogram calculated from the output of the 25th perceptive RLC filter.

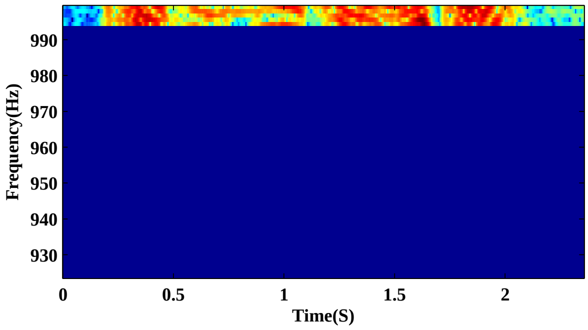

Fig. 6. Spectrogram of the speech signal taken after passing through the 25th filter in the filterbank.

The log values of STFT corresponding to the bandwidth of each of the filters in the filter bank are extracted from the respective spectrograms. These arrays (50 in number) of samples are then vertically concatenated in the order of frequency spectrum of the filters to arrive at a final spectrogram. From the 50 spectrograms the STFT values are extracted only from their high gain regions. Hence we can consider this final spectrogram as the spectrogram corresponding to the enhanced speech. Figure 7 shows the spectrogram of the speech enhanced by the perceptive RLC filtering method. Figure 8 shows the spectrogram of the input speech signal obtained by the conventional method.

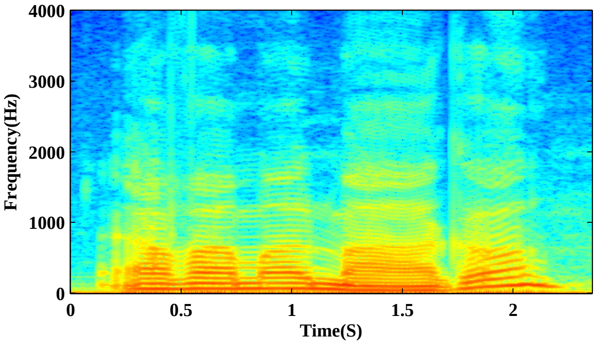

Fig. 7. Spectrogram of the enhanced speech signal.FFT values corresponding to the bandwidth of each of the 50 filters are extracted from the respective spectrograms.These values are then, according to the bandwidth, vertically concatenated to form the final spectrogram

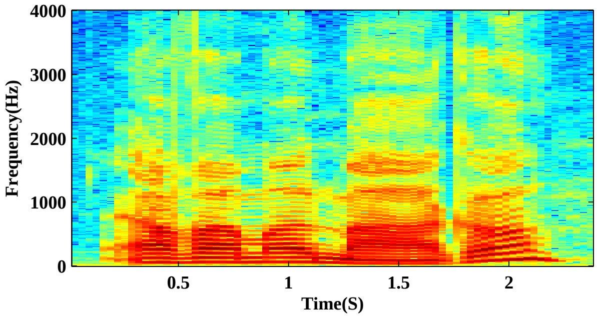

Fig. 8. Spectrogram of the input speech signal taken by the conventional method.

3. RESULTS AND DISCUSSIONS

In order to verify the speech enhancement capability of the designed filterbank, several experiments were carried out. For this purpose, to a single word speech signal, noise is added in varying quantities (in db), the speech signal is then passed through the filter bank. Spectrograms are calculated for each of the filtered speech. Spectral values corresponding to the high gain regions are extracted from the respective spectrograms. The extracted values are vertically concatenated to form the final spectrogram. For the noised words, the spectrograms obtained through the proposed method is more informative comparing to the conventional spectrograms.The spectrograms generated by the conventional method and by the discussed filter method for the noiseless single word "hello" are shown in Figure 9 and Figure 10 respectively. Followed by that, 0db noise is added to the word "hello" and spectrograms are generated by the conventional method and by the filter bank method. The spectrograms of the noise added word "hello" generated by the conventional method and by the proposed method are shown in Figure 11 and Figure 12 respectively.

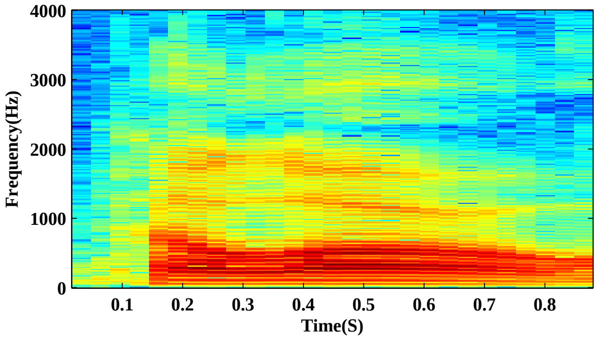

Fig. 9. Spectrogram of the single word speech signal. The word "hello" is extracted from the TIMIT database.

4. PERFORMANCE TEST FOR THE PERCEPTIVE FILTERS

4.1 Experiments for checking the enhancement of the speech signal

For analysing the performance of the proposed filters, certain tests were carried out. From the TIMIT database 200 different words are extracted and to each of the words 0db white gaussian noise is added. Spectrogram of each pair of words, i.e., both noised and noiseless, are calculated using the filter method.Then a matrix indicating the similarity between these two spectrograms is found out using the dynamic time warping (DTW)

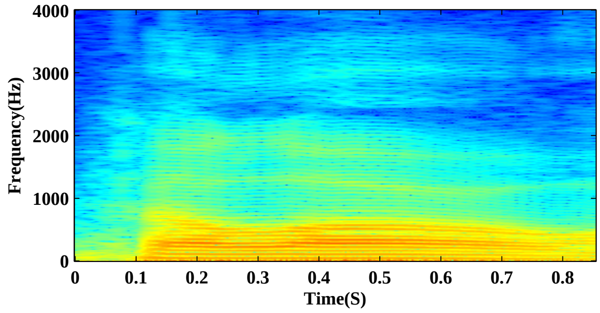

Fig. 10. Spectrogram of the single word speech signal using the proposed filterbank method. The word "hello" is extracted from the TIMIT database. It is then applied to the filterbank. The FFT values corresponding to the high gain regions of each of the filters are taken out from the corresponding spectrograms.These values are vertically concatenated to form the final spectrogram

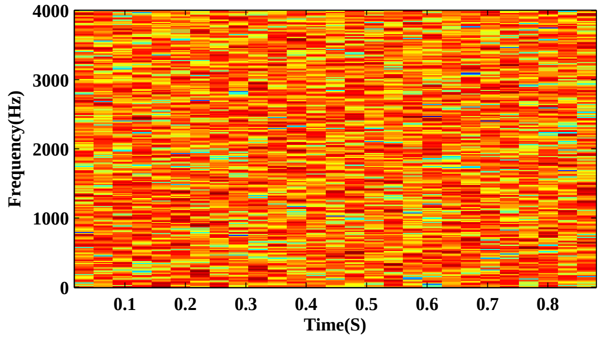

Fig. 11. Spectrogram of the noised speech signal. 0db noise is added to the "hello" word. Then using the conventional method spectrogram is generated

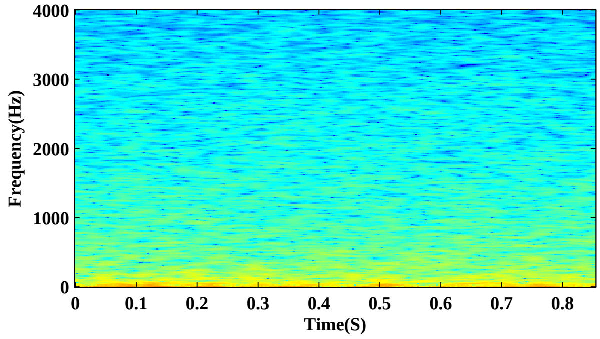

Fig. 12. Spectrogram of the 0db noise added single word speech signal using the proposed filterbank method. 0db noise is added to the word and it is then applied to the proposed filterbank.The FFT values corresponding to the high gain regions of each of the filters are taken out from the respective spectrograms.These values are vertically concatenated to form the final spectrogram

method[4][10]. From this match matrix, we could estimate the distance between the two words, which is an indication of the match between two words. If the distance values are large, the match or similarity is poor. The distance values are calculated for all the 200 word pairs (noiseless and noised version). Then, similar experiment is also done with the same set of 200 noised-noiseless pair of words using the conventional method. The histograms of these match values i.e. both for the conventional method and for the filter method is shown in Figure 13. From the histograms it is clear that by using the new filter bank we could enhance the quality of speech perception. It is because the distance or mismatch between the noiseless word and its noised version is very low in the proposed filter bank method. Another pair of histograms is shown in Figure 16, in this experiment, the match between each of the 200 unnoised word and the remaining 199 noised words are found out, after passing each pair through the filter bank, and when is passed through the conventional method.

4.2 Comparison of the logarithmic scale used in the perceptive filters with conventional scales

The MEL scale is a perceptual scale of pitches judged by listeners to be equal in distance from one another. The reference point between this scale and normal frequency measurement is defined by assigning a perceptual pitch of 1000 mels to a 1000 Hz tone. A popular formula to convert frequency in hertz into frequency in mel is

$$m = 2595log_{10}(1 + f/700) \qquad (5)$$

For analysis purpose the proposed filter bank is redesigned in melscale[5][10]. Then the match(distance) between each pair of noised and noiseless words for a set of 200 words are calculated. Same experiment is done for the new filterbank. Then the histograms of the match values are plotted for the new filter bank and for the filterbank designed in melscale as Figure 15. From the histograms it can be understood that better match of a word with its noised version is there for the proposed filterset compared with the filterset designed in melscale. Similar experiment is done to make a comparison with the bark scaled[10] filterbank as shown in Figure 16. To convert a frequency in hertz (Hz) into Bark scale use:

$$Bark = 13\arctan(0.00076f) + 3.5\arctan((f/7500)^2) \qquad (6)$$

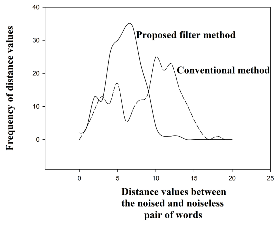

Fig. 13. Histograms showing the match between 200 noiseless words and their noised versions. For plotting the first histogram the match(distance) between the noised and noiseless pairs are found out after passing each pair of words through the proposed filter bank and the match values for the second histogram are found out by the conventional method

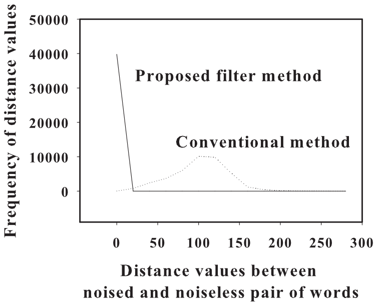

Fig. 14. Histograms showing the match between each word with the remaining set of words which are added with 0db noise. For plotting the first histogram the distance between the noised and noiseless pairs are found out after passing each pair of words through the proposed filter bank and the match values for the second histogram are found out by the conventional method. 200 different words from the TIMIT database are used for this experiment.

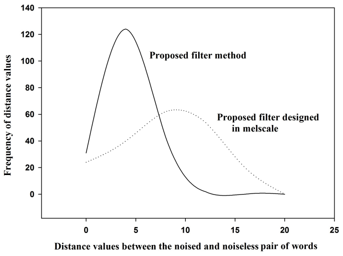

Fig. 15. Histograms showing the match between 200 noiseless words and their noised versions. For plotting the first histogram the distance between the noised and noiseless pairs are found out after passing each pair of words through the proposed filter bank and that designed using melscale

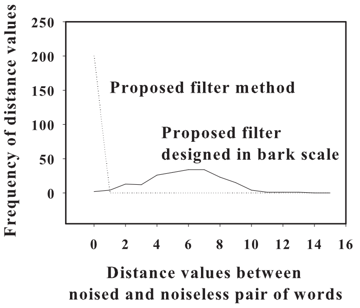

Fig. 16. Histograms showing the match between 200 noiseless words and their noised versions. For plotting the first histogram the match(distance) between the noised and noiseless pairs are found out after passing each pair of words through the proposed filter bank and that designed using barkscale

5. CONCLUSION

The paper presented and introduced the concept of perceptive RLC for reducing the presence of noise in speech signal. We demonstrated that the proposed approach shows improved similarity values for intra-class comparisons when compared with Mel filters and Bark filters. The proposed method can be fully integrated into a VLSI hardware and can offer a high speed and robust solution to automated speech processing and recognition.

6. REFERENCES

[1] Cheng, Abdulla, W. ; Salcic, Z. HardwareSoftware Codesign of Automatic Speech Recognition System for Embedded Real-Time Applications In*Industrial Electronics, IEEE Transactions* March 2011

[2] Jigar Shah and Satish Shah. VLSI Implementation of Hybrid Algorithm Architecture for Speech Enhancement In*International Journal of Computer Science Issues* Vol. 9, Issue 4, No 2, July 2012

[3] Somsak Sukittanon , Les E. Atlas , James W. Pitton , and Jack McLaughlin. Non-stationary signal classification using joint frequency analysis. citeseerx.ist.psu.edu/viewdoc/download

[4] T. Bin Amin. Speech recognition using dynamic time warping. 2008.

[5] E.H.C Choi. On compensating the mel-frequency cepstral coefficients for noisy speech recogition. In *Proceedings of the 29th Australasian Computer Science Conference*, volume 48, pages 49–54, 2006.

[6] S Umesh. Studies on inter-speaker variability in speech and its application in automatic speech recognition. In *Sadhana , Indian Academy of Sciences*, Vol. 36, Part 5, October 2011, pp. 853883.c

[7] Zbigniew Leonowicz, Tadeusz Lobos, and Krzysztof Wozniak. Analysis of non-stationary electric signals using the s-transform. *International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 28(1):204–210, 2009.

[8] G. Li and M. E. Lutman. Independent component analysis:a new frame work for speech processing in cochlear implants? http://www.spars05.irisa.fr/ACTES/PS1-9.pdf.

[9] R. R. Mergu and S. K. Dixit. Multi-resolution speech spectrogram. *International Journal of Computer Applications*, 15(4), 2011.

[10] Lindasalwa Muda, Mumtaj Begam, and I. Elamvazuthi. Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques. *Journal of Computing*, 2(3), 2010.

[11] Zohra Yermeche, Per Cornelius, Nedelko Grbic, and Ingvar Claesson. Spatial filter bank design for speech enhancement beamforming applications. In *Sensor Array and Multichannel Signal Processing Workshop Proceedings*, pages 557–560, 2004.

[12] Novlene Zoghlami and Zied Lachiri. Application of perceptual filtering models to noisy speech signals enhancement. *Journal of Electrical and Computer Engineering*, 2012. doi:10.1155/2012/282019.