Hybrid Filter for Enhancing Input Microphone-Based Discriminative Model

Hani S. Hassan[1]*, **Jammila Harbi S.[2]**, **Maisa'a Abid Ali Kodher[1]**

[1]Computer Science Department, University of Technology, Baghdad, Iraq
[2]Computer Science Department, Mustansiriyah University, Baghdad, Iraq

Abstract

Voice denoising is the process of removing undesirable voices from the voice signal. Within the environmental noise and after the application of speech recognition system, the discriminative model finds it difficult to recognize the waveform of the voice signal. This is due to the fact that the environmental noise needs to use a suitable filter that does not affect the shaped waveform of the input microphone. This paper plans to build up a procedure for a discriminative model, using infinite impulse response filter (Butterworth filter) and local polynomial approximation (Savitzky-Golay) smoothing filter that is a polynomial regression on the signal values. Signal to noise ratio (SNR) was calculated after filtering to compare the results after and before adding the Savitzky-Golay smoothing filter. This procedure showed better results for the filtering of ambient noise and protecting a waveform from distortion, which makes the discriminative model more accurate when recognizing voice. Our procedure for preprocessing was developed and successfully implemented on a discriminative model by using MATLAB.

Keywords: hybrid filter, voice enhancing, ambient noise, preprocessing

<div dir="rtl">

تحسين مدخل المايكروفون القائم على النموذج التميزي بواسطة فلتر هجين

هاني سعيد حسن[1]*، **جميلة حربي سعود[2]** ، **ميساء عبد علي خضر[1]**

[1]قسم علوم الحاسبات، الجامعة التكنلوجية، بغداد، العراق
[2]قسم علوم الحاسبات، الجامهة المستنصرية، بغداد، العراق

الخلاصة

الهدف من النظام المقترح معالجة الأصوات غير مرغوب بها عند استخدامة ضمن نموذج تميزي. النظام المقترح يساعد النموذج التميزي للتعرف على الكلام ضمن محيط ضوضائي يصعب تميز الكلام داخلة، لذلك نحتاج الى فلترة الأصوات الغير مرغوبة بطريقة تحافظ على شكل موجة الكلام من التشوية لكي يتم الكشف عنها بدقة بواسطة النموذج التميزي. مبدأ عمل النظام المقترح, النظام يستحوذ على الموجة الصوتية من المايكروفون ثم يمررها الى خطوة المعالجة التي تحتوي على فلترين متتالين **هما** Butterworth و Savitzky–Golay حيث يعمل الأول على الاحتفاظ بالموجة ما بين (250−7000) هيرتز اي بمعنى, قطع الموجة الاكبر من 7000 هيرتز والأقل من 250 هيرتز. النظام يمرر الموجة المحفوظة الى الفلتر الثاني للحصول على تنعيم لشكل الموجة بدون تشوهها. نظامنا المقترح تم وضعة وتنفيذة بنجاح على نموذج تميزي بواسطة الماتلاب.

</div>

_____

*Email: Hani.co81@gmail.com

Introduction

   Voice enhancement is the process of improving the quality of the voice signal by lessening the foundation loud noise and other undesirable sounds. Voice signal quality is frequently evaluated by its clarity, consistency, and understanding. Voice improvement is an essential system in the speech recognition, including speech synthesis, speech coding, speech recognition, and speech analysis. A voice signal that is recorded in a constant situation may contain undesirable sound, for example, playing uproarious declaration by individuals, sound of fan, air conditioner and so on. These are considered under the class of loud noise. To audience members, these obstructions are profoundly unsavory and ought to be decreased in order to improve the quality and consistent of the discourse signal. Additionally, the preparing calculations of the discourse signal are based on the suspicion that the voice signal is free from foundation commotion. Nearness of foundation commotion in voice signal will bring down the presentation of the discourse handling framework essentially [1]. Hayashida *et al*. [2] proposed a creative versatile beamformer innovation for voice enhancement under loud situations and assessed speech recognition performance by using four microphones. Accordingly, the process used a single microphone for the enhancement of the input voice and the achievement of an appropriate discriminative model. Naylor *et al*. [3] proposed an approach to accomplish such objectives of speech improvement by using a state-of-the-art automatic speech recognition (ASR) system for a wide range of resonation and commotion conditions. The work employed the ACE challenge database that included the measurements of multichannel acoustics from seven unique rooms with resonation times ranging from 0.33 to 1.34 s. The approach was inappropriate for real-time speech recognition when using a single microphone because of its dependence on the ACE challenge database which makes it more time-consuming when treating noise. Hussein, *et al*. [4] proposed a wavelet analysis for noise cancellation in speech signals. A test was made among different families of wavelets. However, the approach was slow as compared with the proposed one when applied on speech recognition, since it is based on wavelet. Lee *et al*. [5] proposed a procedure based on a system that recognizes the content of speech that has ambient noise. The work consisted of two stages, the first is the automatic enhancement of the quality of speech, depending on the signal-to-noise ratio (SNR), and the second is noise reduction by using the subspace speech enhancement. Also, the work was based on two steps for reducing ambient noise. In the first step, the input microphone is passed through a passband filter for the cut-off between high and low frequencies. The second stage is the smoothing of the waveform after the cut-off and the higher reduction of the noise ratio, compared with the first step, based on SNR.

Relational Transfer Function and Local Polynomial Approximation

   The infinite impulse response filter (IIR) is called the recursive filter. The most ordinarily utilized IIR filter technique uses a reference simple model channel. It is the best technique to utilize when planning standard filters, for example, high-pass, low-pass, bandpass and band-stop filters. The IIR filters are very similar to the simple filters. There are four distinct types of IIR filter structures, namely Chebyshev I and II as well as the Butterworth and elliptic structures. The differences between their reactions are whether the reaction is monotone or it swells in the passband and stopband [6, 7, 8]. The Butterworth filter is a signal processing filter that has a flat frequency response in the passband and called as a maximally flat magnitude filter. The Butterworth is a monotone filter that is diminishing both stopband and passband [9]. Savitzky-Golay smoothing filter is a low pass filter that achieves a polynomial regression on the values of the signal that it uses, in order to render the relative heights and widths of the spectral lines noticeable in the boisterous spectrometric information. It is considered as a conceivable and least difficult averaging system. Savitzky-Golay smoothing filter is based on moving the window averaging, as referred to in equation 1, and applying the filter to continuous data values, as shown in equation 2 [10].

$$gi = \sum_{n=-nl}^{nR} Cnfi + n \qquad\qquad\qquad\qquad\qquad ………..(1)$$

$$f_i \equiv f(\text{ti}),\ t_i \equiv t0 + i\Delta \qquad\qquad\qquad\qquad ………………..(2)$$

where:

fi: *data value*

Δ: *constant sample spacing*

$i = … -2, -1, 0, 1, 2, …$

(nl,nR): the number of points used (left and right)

$$Cn = {1}/{(nL + nR + 1)}$$

The bandpass filter will enable a few signals to go through while blocking others. A bandpass filter permits a signal of a specific recurrence range to go through the filter as it stands. This scale of acknowledged frequencies is known as the bandpass. The size or scale of the bandpass is known as the bandwidth. By this way, any signal higher or lower than the selected recurrence range will be blocked. This is helpful for evacuating undesirable commotion by blocking any signal that is not needed [11,12].

The Proposed Procedure

The proposed procedure is applied in order to make the discriminative model better in recognizing the input speech from the microphone during the running of the model. It involves a preprocessing step through the suppression of noise environments such as air conditioner, fan, loudness, etc. Our procedure, as shown in Fig. (1), is a hybrid between the infinite impulse response filter (Butterworth) and finite impulse response filter (Savitzky-Golay). The procedure consists of two steps. Firstly, the Butterworth bandpass filter is used to cut-off the input microphone of the range of 250-7000 Hz, which leads to the reduction of the loudness and environmental noise. Secondly, Savitzky-Golay smoothing filter is used as a low pass filter for filtering, smoothing and keeping the form of the wave. Signal to noise ratio and intensity (db) are calculated by using equation 3 between the original "1s wav. file" and the same file after filtering by passing it through the Butterworth filter. Then the SNR is calculated between the wav. file that was filtered by the Butterworth filter and that file after passing it through Savitzky-Golay filter. The results showed that the SNR was increased from -3.4745 to -0.0497, making the discriminative model more accurate.
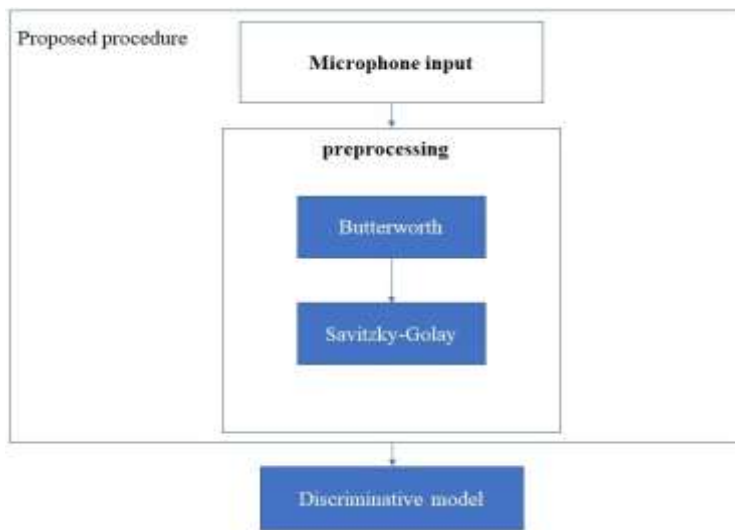

Figure 1-Diagram of the proposed procedure with the discriminative model

$$SNR = 20\, log10 \left\{ \frac{1}{N} \sum_{i=0}^{N-1} \frac{s}{n} \right\} \qquad \text{.........(3)}$$

Results

After developing the entire system, a testing was made for the proposed procedure based on the preprocessing, which is the important step in the discriminative model (using deep learning). The discriminative model is based on the trained dataset in the English language. The number of files used in the training is 1750 files of the format "wav.". The model employed the convolution neural network (CNN) with specific properties, such as the number of layers, filters, and the activation function. Thus, we obtained a trained dataset with labels, which is an important factor in the proposed procedure. The outcomes of this procedure Are shown in Figure-2 which shows the original waveform ('go') as the input to the microphone that has the properties of the format "wav." . The Bit rate was 256 kbps and its length was 1 second. It was recorded by using MATLAB via a microphone with a sampling frequency of16000, Nbits of 16 and a number of channels of 1. Figure-3 shows the filtered waveform

by using the Butterworth filter with a cut-off value of 250-7000 Hz. The ambient noise was reduced by reducing the high amplitude of the original waveform to 0.8 rather than 1, whereas the low amplitude of the original waveform became -0.7 rather than -1. The resulted Butterworth filter waveform was used as the input to Savitzky-Golay filter in order to smooth the waveform. As a result, the high amplitude was reduced from 0.8 to 0.77 whereas the low amplitude was increased from -1 to 6.5, as shown in Figure-4. Thus, a more accurate recognition was achieved by using the discriminative model. Figure-5 shows the discriminative model that recognizes the voice ('go').
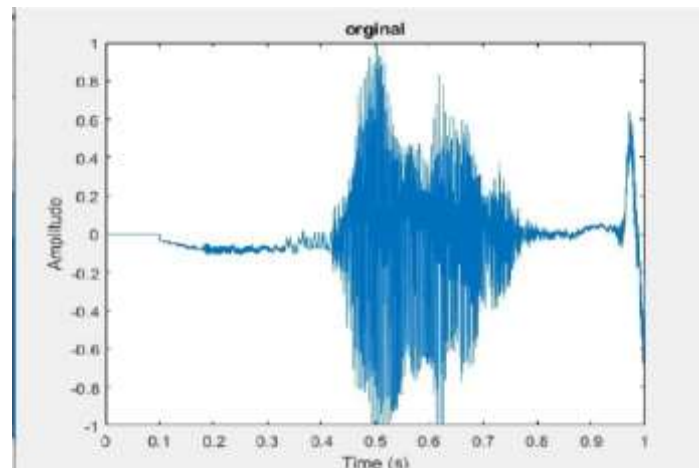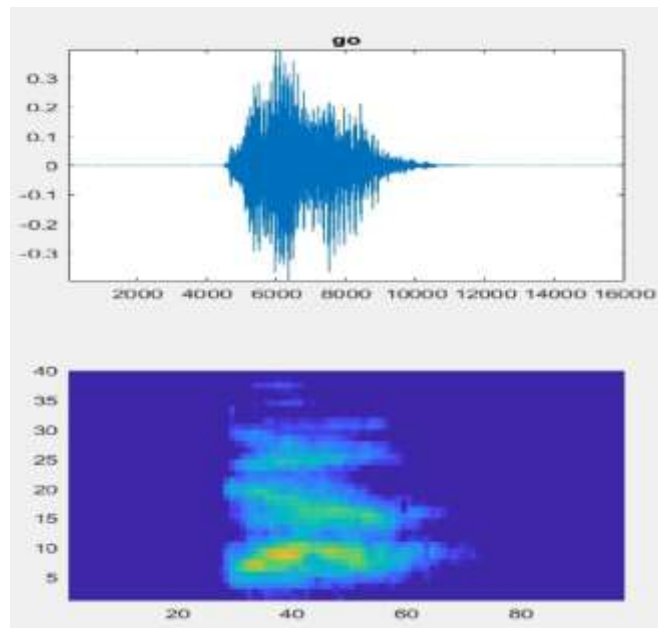

Figure 2-The original waveform 'go' of the input microphone before it processing by hybrid filters
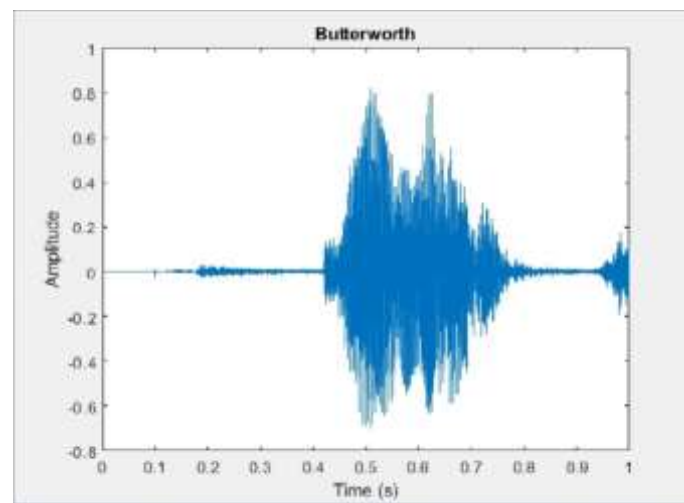
Figure 3-Butterworth filter
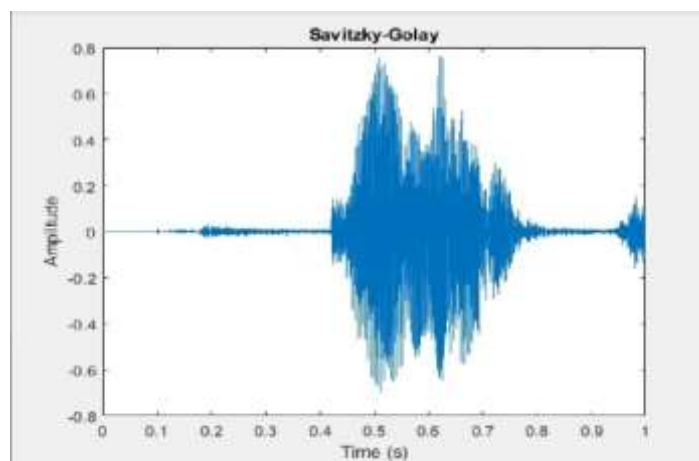
Figure 4-Savitzky-Golay smoothing filter

Figure 5-Detecting the voice 'go'

Conclusions

The main aim of our proposed procedure was the preprocessing of the signal during the recognition of the speech using the discriminative model to filter the environmental noise and protect the waveform from distortion. The speech waveform was acquired from the input microphone and passed through the preprocessing step that consisted of the Butterworth bandpass filter with a cut-off of 250-7000 Hz and the Savitzky-Golay low pass filter for smoothing the waveform and protecting it from distortion. Our procedure for preprocessing was successfully implemented on a discriminative model by using MATLAB.

References
1. Shajeesh, K. U., Sachin, S., Pravena, D. and Soman, K. P. **2012**. International Journal of Computer Applications, *Speech Enhancement based on Savitzky–Golay Smoothing Filter*, **21**: 112-6.
2. Naylor, P.A.**2017**. ELSEVIER, *Speech enhancement for robust automatic speech recognition: Evaluation using a baseline system and instrumental measures*, 46(5): 574-584.
3. Kohhei, H., Tsuyoki, N., Kazuya, N., Takeo K. and Takanori, A. **2017**. Noise reduction and evaluation of speech recognition performance by the adaptive beamformer with four microphone for wearable speech translation device. IEEE 6th Global Conference on Consumer Electronics (GCCE) Nagoya, Japan:24-27.
4. Kohhei, H. , Tsuyoki, N. , Kazuya, N. , Takeo, K. and Takanori, A. **2018**. Journal of University of Babylon, Engineering Sciences, *Reduce The Noise in Speech Signals Using Wavelet Filtering*, **5**: 117-8.
5. Sheng Lee, Jhing, F. and Miao, H. **2018**. Threshold-Based Noise Detection and Reduction for Automatic Speech Recognition System in Human-Robot Interactions, *National Center for Biotechnology Information*, **18**(7):2068. DOI: 10.3390/s18072068
6. John, Wi, Sons. **2008**. *Digital Signal Processing using MATLAB* 2nd ed.New Jersey. 285 p.
7. Thad, B., Cameron, HG. and Michael, G. **2017**. *Real-Time Digital SignalProcessing*. 3th ed. Boka.313 P.
8. AL-Abidy, M., S. and AL-Tamimi, A., A. **2010**. AL Iraqi Academic Scientific Journals. *Noise Remiving from Speech Signal Using Linear –Phase Type II Notch Fir Filter*, **23**(4): 3-6.

9. Eswara, Ra., Muralidhar, PV. and Raghuramakrishna, S. **2015**. *Audio Equalizer with Fractional Order Butterworth Filter, International Journal of Engineering and Management Research*. **5**(5): 266-272.

10. Monisha, Ch., Shreya, Da. **2012**. ELSEVIER. *Determination of Signal to Noise Ratio of Electrocardiograms Filtered by Band Pass and Savitzky-Golay Filters*, **8**(10): 831-835.

11. Thyagarajan, KS. **2019**. *Introduction to Digital Signal Processing Using MATLAB with Application to Digital Communications*. 1nd ed. San Diego, Springer.189 p.

12. Mahmood, A., F. and Hayder, M., H. **2014**.. An Optimized Adaptive Filtering for Speech Noise Cancellation, *AL Iraqi Academic Scientific Journals*, **23**(5): 44-45.