Effect of Sampling Rate Reduction and Signal Filtering for Gunfire Sound Classification with Spectral Vector using ANN

Settha Tangkawanit* and **Surachet Kanprachar*†**, Non-members

ABSTRACT

To identify the type of the shooting gun, the frequency components of the gunfire sound has been utilized as the input to an Artificial Neural Network (ANN) model [1–2] with a promising recognition accuracy. With this method, the required computational resource is large if the sampling frequency is too high. In this research, the sampling frequency for transforming the input gunfire sound into a digital signal is varied from 44.1 kHz down to 4.41 kHz. 6 different types of gunfire sound are considered. Additionally, the effect of applying different types of signal filtering is studied. It is found that only reducing the sampling frequency on the input gunfire sound signal does not deliver a good performance on gunfire sound classification. To obtain a good classification accuracy, signal filtering has to also be applied. With Chebyshev Type II filter and 4.41 kHz sampling frequency, the classification accuracy is all 100% for the practical range of SNR. This impressive classification accuracy comes with a 10-times reduction on the computational resource; that is, from the sampling frequency of 44.1 kHz to 4.41 kHz. The findings from this work can be applied to the gunfire sound classification system with limited computational resource.

Keywords: Gunfire Sound, Classification, Neural Network, Sampling Frequency, Signal Filtering, Fast-Fourier Transform

1. INTRODUCTION

Sound classification is very important in a variety of applications; for example, heart sound classification, lung sound classification, and so on. Considering the gunfire sound, it is seen that if there is a way to identify the type of the shooting gun, it would be very beneficial in many situations; for example, military service, patrolling duty along the border line, and so on. Furthermore, the location of the shooting gun can then be determined if the identification process is applied as a network. Currently, there are many commercial gunfire-sound analysis systems. Boomerang [3] is one of these systems. It was developed by DARPA and BNN Technology company in order to build a remote gunfire sound detection system using many large microphones in capturing the sound. These microphones are put on high poles located in the area of interest. The input sound from these microphones is then passed through the system so that the type and location of the gunfire sound can be determined. It is seen that this particular system is quite static; that is, the receiving devices are fixed in terms of location. It cannot be applied to other areas.

There are many techniques in analysing the gunfire sound. The muzzle blast is one key sound that has been studied [4–5]. This sound is produced by many factors while the bullet is shot from the gun; for example, the gunfire powder ignition, the movement of the bullet in the barrel, the length of the barrel, and so on. Analysing this muzzle blast, it was shown that the trajectory of the bullet and the location of the shooting gun can be estimated. In [6], many key features from the gunfire sound are considered; for example, muzzle blast, energy, zero-crossing rate, frequency spectrum determined by MFCC (Mel-Frequency Cepstral Coefficients), and so forth. These features were then put in the decision process exploiting maximum likelihood algorithm in order to identify the type of the shooting gun. It was found that the least error in classification was obtained from the case of applying LS-LDA (Least Square-Linear Discriminant Analysis) to the system. To utilize VLSI with low power consumption in gunshot detection, ROC (Receiver Operating Characteristic) was adopted as the metric in measuring the performance [7]. It was shown that to get a good performance, Wavelet Transform has to be applied. However, the complexity of such system was the drawback, which led to a high cost in production.

Apart from identifying the type of gunfire, the localization of the gunfire has also been in the focus of many research works. A technique called DOA (Degree of Arrival) has been adopted [8] in order to

Manuscript received on September 10, 2019 ; revised on September 30, 2019 ; accepted on October 10, 2019. This paper was recommended by Associate Editor Chanon Warisarn.

*The authors are with the Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University, Phitsanulok, Thailand.

†Corresponding author. E-mail: surachetka@nu.ac.th

Digital Object Identifier 10.37936/ecti-eec.2020181.215607

receive the signal from the array microphone. The input signal was estimated by using three methods; that is, ES (Exhaustive Search), SCFL (Searching Consistent Fundamental Loop), and ES-SCFL. It was shown that the best was from ES-SCFL method in terms of the localization accuracy. Additionally, in [9], asynchronous localization from many nodes over a network has been adopted in order to determine the GPS coordinates of the gunfire. The shockwave and muzzle blast characteristics were used in the process. These two parameters were also adopted extensively in [10–11] in order to estimate the loudness of the gunfire so that the direction and the shooting angle can be determined.

The techniques used in other sound analysis applications are also important. In analysing heart beats, singular spectrum analysis was used in extracting the respiratory signals from the electrocardiogram (ECG) [12]. In [13], the abnormality of the heart beats was studied using the spectrum of the signal. It is seen that these spectral-based techniques can also be applied to the gunfire analysis since both are considered as impulses. This spectral consideration has been done also in gunfire sound classification as shown in [1–2]. The frequency components of the gunfire sound in the frequency range of interest were used as feature vectors in an ANN model for performing the classification process. An impressive classification accuracy was achieved. However, in order to enable such classification system in a mobile device, the required computational resources should be reduced. One important parameter affecting the number of required resources is the sampling frequency since the higher the sampling frequency, the larger the number of frequency components. In this work, the sampling frequency is reduced from its original sampling frequency of 44.1 kHz down to 4.41 kHz. Additionally, different types of signal frequency filtering; that is, Butterworth, Chebyshev Type I, Chebyshev Type II, and Elliptic filters, are considered. The gunfire sound classification accuracy from different cases in terms of sampling frequency and filtering is studied.

The organization of this paper is done as follows. The related technology used in this work is shown in Section 2. The proposed gunfire sound classification system is described in Section 3. Then, in Section 4, the results in terms of the classification accuracy and the time required for different cases are given and discussed. Finally, the research work is summarized in Section 5.

2. RELATED TECHNOLOGY

2.1 Sampling Frequency

To transform the input gunfire sound from the receiving microphone into a digital signal, the sampling process has to be done. In this process [14–15], the input analog signal is sampled repeatedly with an equal time difference, called sampling period (T_s). The number of samples in 1 second is called the sampling frequency or f_s. Note that the sampling frequency is just the reciprocal of the sampling period; that is, $f_s = 1/T_s$. The input analog signal ($S(t)$) and its samples (S_i) are shown in Fig. 1. It is seen from the figure that successive samples are apart by equal time interval; that is, T.

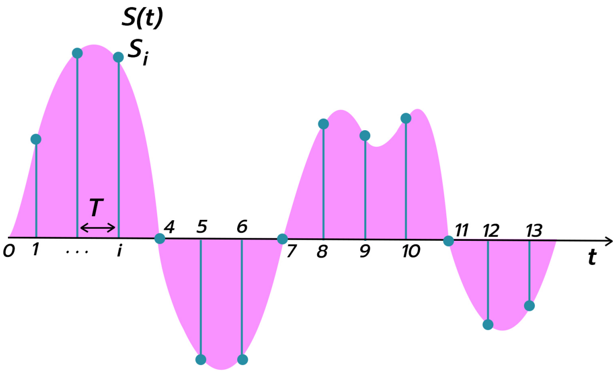

Fig.1: *Example of input analog signal and sampled signal.*

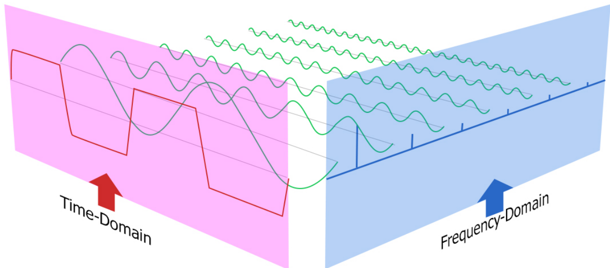

Fig.2: *Example of signal decomposition into many cosine signals.*

2.2 Fourier Transform

Generally, any physical signal for a certain time interval can be described as the combination of many orthogonal functions with different frequencies and amplitudes [16–17]. The frequencies of these orthogonal functions are in the factor of the fundamental frequency of the considered signal. The theorem behind this description is from Fourier Series. An example on this is shown in Fig. 2. It is seen that the signal on the left can be decomposed into many cosine signals each with different frequencies and amplitudes. Note also that, from the figure, it is also shown that the amplitudes of these frequency components decrease as the frequency increases.

To determine the frequency components of the given signal, Fourier transform is normally required. Dealing with the sampling data, to transform the input time-domain samples into frequency ones, Discrete Fourier Transform (DFT) is needed. This transform is given in Eq. (1); that is,

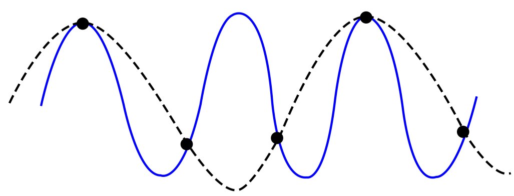

Fig.3: *Example of sinusoidal signal (solid line) and the under-sampled signal (dashed line).*

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn} \quad (1)$$

where $x(n)$ is the time-domain sampled signal and N is the total number of samples in one period.

From Eq. (1), it is seen that to determine the DFT for a particular k, many computations are needed. Such process requires both time and computational resources. To reduce these requirements, a technique called Fast Fourier Transform (FFT) proposed by Cooley and Tukey has to be applied. The Butterfly operation is adopted by FFT in order to reduce the multiplication process. This transform is given by Eq. (2); that is,

$$X(k) = \sum_{n=0}^{\frac{N}{2}-1} x(2n)W_N^{2mn} + \sum_{n=0}^{\frac{N}{2}-1} x(2n+1)W_N^{2(n+1)m} \quad (2)$$

where

$$W_N = e^{-j\frac{2\pi}{N}} \quad (3)$$

From Eq. (2), it is seen that applying this FFT to the input time-domain samples, the frequency components can be determined faster.

2.3 Aliasing Effect

In transforming the analog signal into a digital one, the sampling process is the first one to be considered. There is a theory supporting this process; that is, Nyquist sampling theorem. From this theorem, it is mentioned that in order to perfectly recover the sampled signal into its original format, the sampling frequency has to be at least twice the highest frequency of the input analog signal. However, if the sampling frequency is less than this requirement, an impairment in signal reconstruction is introduced [18–19]. This effect is called the aliasing effect.

An example of the aliasing effect is shown in Fig. 3. The input sinusoidal signal is shown by the blue solid line. The sampled signal is shown by the black dashed line. It is seen that the sampled signal is different from the original input signal since the sampling frequency used in the sampling process is lower than the Nyquist sampling frequency. If this sampled signal is used in any further applications, it will lead to a wrong interpretation.

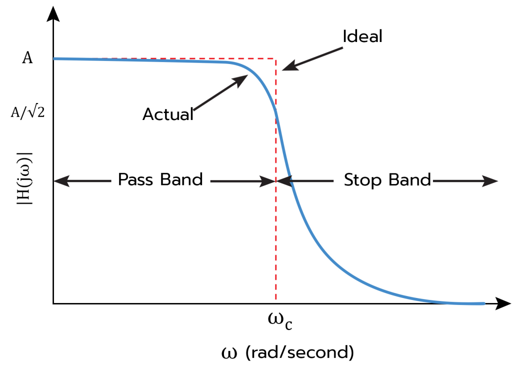

Fig.4: *Ideal and actual frequency response of low-pass filters.*

2.4 Digital Low-Pass Filter

From the previous discussion about the aliasing effect, it is seen that the key factor in preventing this effect is the sampling frequency. However, if the sampling frequency is too high, at least two drawbacks are introduced; that is, the number of frequency elements to be considered and the noise added to the system. The first drawback will lead to a requirement on the memory to be used in the computation. The second drawback can then lead to a corrupted received signal for which a wrong interpretation may be decided. To reduce these two effects and not to allow the aliasing effect to happen, a proper filtering is normally applied. Low-pass filter is used in order to allow only the signal of interest, in terms of frequency representation [20–23]. The frequency response of a low-pass filter is shown in Fig. 4. From the figure, it is seen that ideally, only frequency components lower than ω_c (called the cut-off frequency) are allowed to pass through the filter. Any components with frequency higher than ω_c will not pass through the filter. Such frequency response is shown by the red dashed line in the figure. In practice, to obtain such response is quite complex and costly. Generally, if the requirement in terms of the sharp cut-off is lessened, the frequency response of the low-pass filter can then allow some distortion at the passband and some frequency components at the frequency slightly higher than the cut-off frequency. This actual response is shown in the figure by the blue solid line.

According to the actual frequency response of the low-pass filter shown in Fig. 4, there are major 4 different designs of low-pass filter used in practice; that is, Butterworth, Chebyshev Type I, Chebyshev Type II, and Elliptic filters. Examples of frequency response of these 4 types of digital low-pass filter are shown in Fig. 5.

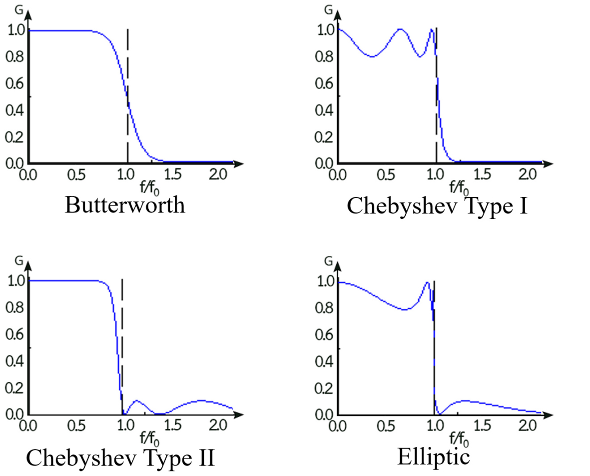

Fig.5: *Frequency responses of 4 different types of digital low-pass filter.*

From Fig. 5, it is seen that the frequency responses from different types of filter are different. For Butterworth filter, there is no ripple in both passband and stopband. However, the response rolls off slowly from passband to stopband; thus, some attenuation is introduced in the passband and some positive gain (G) is allowed in the stopband. Considering Chebyshev Type I and II filters in the figure, it is seen that the roll-off from passband to stopband is sharpened compared to that of the Butterworth filter. However, in Chebyshev Type I, an equiripple behaviour occurs in the passband; while, in Chebyshev Type II, this ripple happens in the stopband. For the elliptic filter, both passband and stopband exhibit the equiripple behaviour. Among these 4 types of low-pass filter, elliptic filter owns the sharpest roll-off in terms of changing gain from passband to stopband. Applying these filters to a signal, the output would surely be different.

2.5 Artificial Neural Network (ANN)

Currently, Artificial Intelligence (AI) and Machine Learning have been widely adopted in an automatic classification system. There are many models proposed and done in this area. One important model used in practice is the Artificial Neural Network or ANN [24–26]. ANN was developed by imitating the biological neuron networks in a human brain. The information is passed through nodes in a layer. Such information is then processed and then passed to the following layer as shown in Fig. 6. From the figure, the input data called feature vectors are fed into the input layer containing many nodes. The data are then processed and passed to the next layer with a weight multiplying to the passing data. This weight represents the importance of a particular input and it is adjusted adaptively. This process is further done in many layers (called hidden layers) depending on the design of the network. At last, the data from the hidden layers are passed to the output layer at which the final output feature is produced.

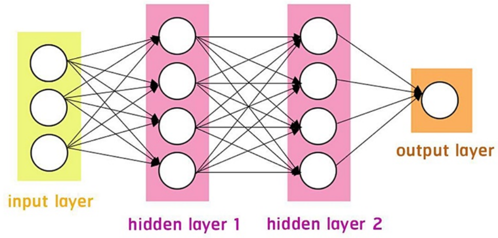

Fig.6: *General structure of an ANN.*

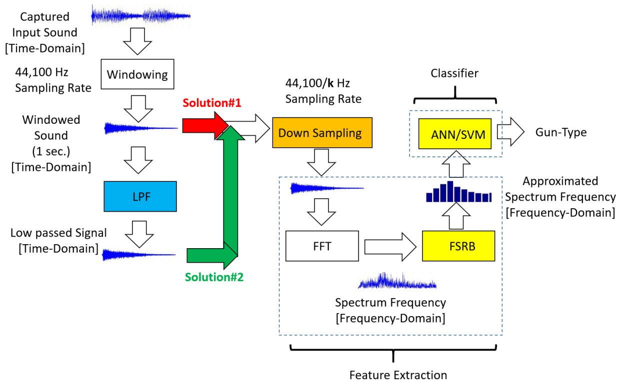

Fig.7: *Proposed gunfire sound classification system.*

ANN is a supervised-learning system; that is, the answer of the work is known and the system is trying to adjust itself so that with the given input, it can finally get the correct answer or at least get the answer with a minimum error as set by the system. Along this process, weights associated with different layers are adjusted adaptively using the back-propagation method. This process of the ANN model is called the training process. Such process has to be done repeatedly in order to reduce the error between the final output and the known answer by modifying the parameters in the model. To get a better model for a particular classification application, the number of input data and the number of runs (called epochs) in the training process have to be large enough.

3. PROPOSED CLASSIFICATION SYSTEM

In this section, the classification system proposed in this work is given. Actually, the classification model used in [1–2] is adopted. An input gunfire sound is classified into one of 6 different gun types; that is, AK-47, AK-15, Shotgun, 9-mm, .38 caliber, and .45 caliber. The proposed classification system is shown in Fig. 7. It is seen that this figure can be divided into 3 main parts; that is, the input signal preparation part (to make the captured gunfire sound into the digital format), the feature extraction part (to prepare the feature vector for ANN), and the classification part (to determine the gun type). The feature extraction and the classification process are identical to that from [1–2]. The major task in this research is to study about techniques in selecting the feature vector as the input to the feature extraction process. The aim is to reduce the computations and memory required in FFT process while not affecting the classification performance. Two solutions are proposed. The first solution is done by only performing a down sampling process to the input digital signal. The second solution is to apply both signal filtering and down sampling. These two solutions, the feature extraction, and the classification process are described next.

3.1 Solution#1: Down Sampling

From Fig. 7, the proposed gunfire sound classification system is shown. First, the input gunfire sound is captured and windowed into only a 1-second time interval. The sampling rate used is set to be 44,100 Hz; that is, there will be 44,100 samples available. To reduce the number of samples, a down sampling process is added to the system. This is the key block added to the first solution (Solution#1), as shown by the big red arrow in the figure. At this block, 44,100 input samples are sampled down by a factor of k. For example, with $k = 1$, 2, and 4; the number of samples at the output of this block will be 44,100, 22,050, and 11,025 samples, respectively. It is seen that with higher values of k, the number of samples is reduced; thus, the number of input elements to be processed in the succeeding block (that is, FFT block) is lessened. This will then reduce the number of computations and required memory. For this solution, 5 values of k are chosen; that is, 1, 2, 4, 5, and 10. Consequently, the output of the down sampling block will contain 44,100, 22,050, 11,025, 8,820, and 4,410 samples, respectively. Note that using k higher than 10 will result in an under-sampling effect; that is, part of the required signal will be strongly distorted.

3.2 Solution#2: Signal Filtering and Down Sampling

As discussed in Section 2, the aliasing effect could happen if the sampling rate is not high enough. Hence, in this solution, in order to prevent such effect that might be caused by the down-sampling block, a digital low-pass filter is placed precedingly. 4 different low-pass filters are considered; that is, Butterworth filter, Chebyshev Type I filter, Chebyshev Type II filter, and Elliptic filter. The cut-off frequency of these filters is set according to the number of output samples from the down-sampling block.

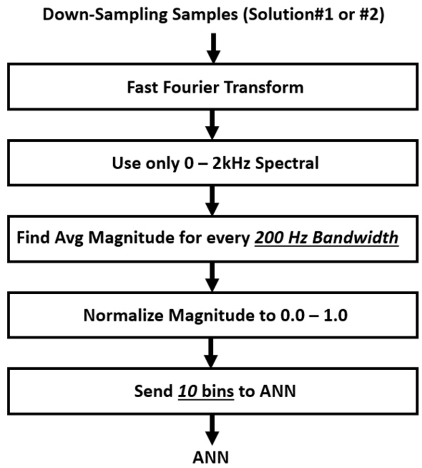

Fig.8: *Flow of the feature extraction part.*

3.3 Feature Extraction

As mentioned previously, the feature extraction in this work is the one used in [1–2]; that is, the Frequency Spectral Representative Bin (FSRB) is adopted. The diagram of this part is also shown in Fig. 7. The output samples from the down-sampling block are passed to the FFT block in order to transform the time-domain samples into the frequency-domain samples (frequency spectrum of the signal). Since the frequency of the gunfire sounds covers mainly from 0 to 2,000 Hz, only the frequency-domain samples from 0 to 2,000 Hz are selected. These samples are divided into 10 bins each covering a bandwidth of 200 Hz [2]. In each bin, the average amplitude of the samples in the bin is determined. The average amplitudes from these 10 bins are then normalized to be between 0.0–1.0. These 10 normalized amplitudes are sent to the ANN as the feature vector in the classification process. The flow of the feature extraction part is given in Fig. 8.

3.4 Classification

The ANN model used in this work is shown in Fig. 9 [1]. It consists of 4 layers; that is, Input Layer, Hidden Layer 1, Hidden Layer 2, and Output Layer, respectively. The input of Input Layer is the feature vector containing 10 normalized average magnitudes from 10 bins, as described in the previous Subsection. The feature vector is passed to the input layer which consists of 10 nodes, as shown in the figure. The output of Input Layer is passed through 2 hidden layers; that is, Hidden Layer 1 (containing 20 nodes) and Hidden Layer 2 (containing 10 nodes). The activation function used for these 2 hidden layers is Sigmoid function. The output from Hidden Layer 2 is passed to Output Layer at which only 1 node with Tansig function is deployed. The output from this ANN model is just a number. This output, ranging from 0.0 to 1.0, will be used in classifying the

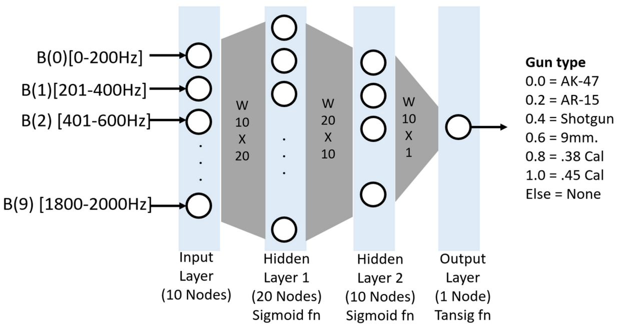

Fig.9: *Structure of the proposed ANN model in the classification part.*

gun type of the given gunfire sound. For example, to classify the given gunfire sound to be AR-15 gun, the output of ANN model has to be in the range of 0.2. Note that the maximum displacement allowed for each case is set to be 0.01; that is, for defining the input gunfire sound to be AR-15 gun, the output of the ANN model has to be between 0.19 to 0.21. Similarly, this maximum displacement is used for the other 5 types of gun. And, if the output of the ANN model is not at these ranges, the given input sound will be classified as non-gunfire sound.

4. RESULTS AND DISCUSSION

To study the performance of the proposed solutions on gunfire sound classification, the classification accuracy will be determined. As discussed, in this work, 6 different types of gun are studied; that is, AK-47, AR-15, Shotgun, 9-mm., .38-Caliber, and .45-Caliber. 10 gunfire sounds from each of these gun types are collected. There are then 60 gunfire sounds. For each sound, some noises are added so that an additional 10 sounds are produced. Totally, there are 660 gunfire sounds to be used in the training process of the gunfire sound classification system. After the training process, the parameters in the ANN model are set accordingly. This trained ANN model is then tested with the testing gunfire sounds. The testing gunfire sounds are determined from varying the signal-to-noise ratio of the 60 original gunfire sounds from 30 dB down to -10 dB by adjusting the noise power to the sound. The average gunfire sound classification accuracy is then determined.

4.1 Solution#1 Results

In this Subsection, the performance of the proposed Solution#1 (that is, the sampling down solution) is studied. The average classification accuracy from different values of sampling frequency is determined for the signal-to-noise ratios of 30, 25, 20, 15, 10, 5, 0, and –5 dB, respectively. 5 different sampling frequencies are chosen; that is, 44,100, 22,050, 11,025, 8,820, and 4,410 Hz. The obtained average classification accuracies for these cases are shown in

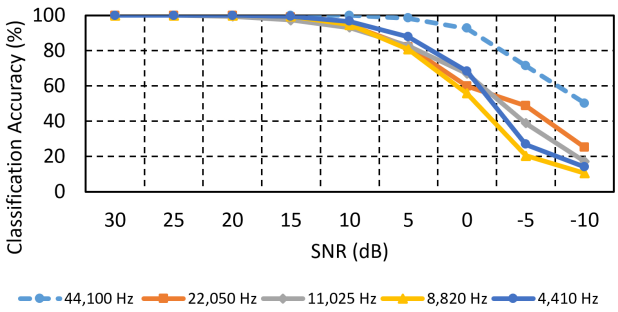

Fig.10: *Average gunfire sound classification accuracy obtained from Solution#1.*

Fig. 10.

From Fig. 10, it is seen that among 5 different curves of the obtained average gunfire sound classification accuracy, at the SNR higher than 20 dB, these 5 different sampling frequencies can deliver an impressive accuracy of 100%. However, as the SNR decreases, the obtained accuracy decreases. The decrement of the accuracy varies depending on the sampling frequency applied. It is seen that the best accuracy curve is from the case of using 44,100 Hz sampling frequency. For this case, the classification accuracy is 100% for SNR of 30 dB down to 10 dB, and decreases to 98%, 92%, 71%, and 50% for the lower SNRs. This accuracy curve is approximately 20% higher than those from the other 4 cases, at the SNR lower than 0 dB. At the SNR of –5 dB, the obtained classification accuracy can be listed in a descendent order as 72%, 49%, 39%, 26%, and 20% for sampling frequencies of 44,100 Hz, 22,050 Hz, 11,025 Hz, 8,820 Hz, and 4,410 Hz, respectively.

It is clearly seen that reducing the sampling frequency (proposed in Solution#1) degrades the performance of the gunfire sound classification. To view the reason behind this degradation, the output frequency spectrum from FFT block (shown in Fig. 7) has to be considered. The frequency spectrum of AK-47 gunfire sound is selected as an example in explaining this degradation and is shown in Fig. 11. 5 different frequency spectra are shown in the figure corresponding to 5 different sampling frequencies.

From Fig. 11, it is seen that for 5 different sampling frequencies, the frequency spectra are different. As the sampling frequency decreases, the effect of aliasing becomes stronger. For example, comparing the spectrum for sampling frequency of 44,100 Hz with that of 4,410 Hz sampling frequency, it is seen that the spectra are significantly different. Many noticeable differences in various frequency ranges of interest are given in the figure. For example, at the frequency range of 0 to 400 Hz, the obtained spectrum of 4,410 Hz case is strongly lifted up for the frequency of 0 to 100 Hz when comparing to that from the case of using 44,100 Hz sampling frequency. Note that even though there is no red circle mark on the case

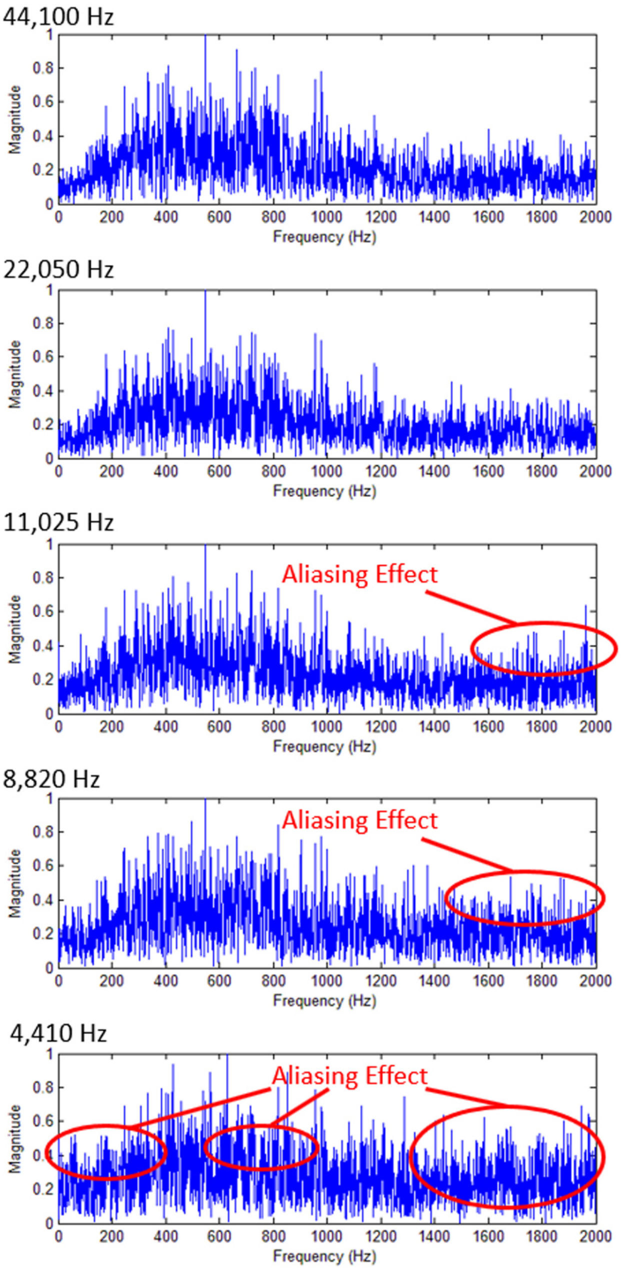

Fig.11: *Frequency spectra of AK-47 gunfire sound at SNR of 0 dB: for different sampling frequencies.*

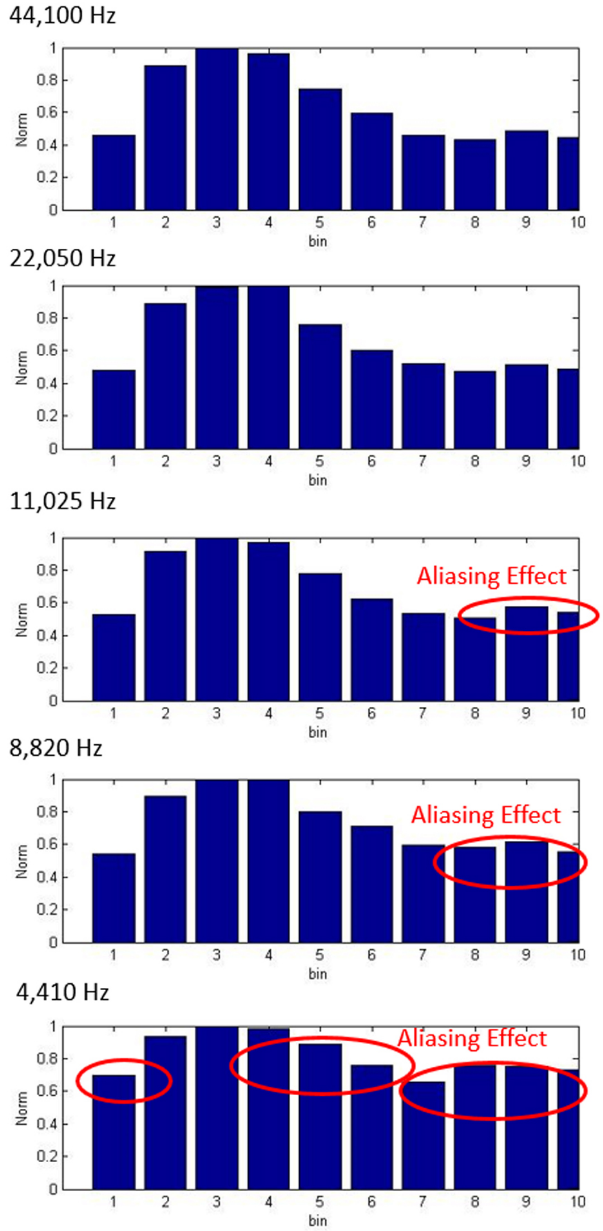

Fig.12: *FSRB feature vector of AK-47 gunfire sound at SNR of 0 dB: for different sampling frequencies.*

of 22,050 Hz sampling frequency, there are still some noticeable differences when comparing its spectrum in the 700 Hz frequency range to that from the case of 44,100 Hz sampling frequency. It is seen from the figure that, as the sampling frequency is further reduced, the obtained spectrum is significantly different when comparing to that from 44,100 Hz sampling frequency. The reason behind this effect is that once the sampling frequency is not high enough, the aliasing effect becomes significant. Additionally, in the situation where more noise is added to the input gunfire sound (that is, with the case of low SNRs), this effect can majorly corrupt the obtained frequency spectrum of the signal.

To view such effect shown in Fig. 11 on the classification process, the input feature vector to the ANN model has to be studied. The corresponding feature vectors of AK-47 gunfire sound for SNR of 0 dB with different sampling frequencies are shown in Fig. 12. As mentioned previously, in this work, the frequency spectral representative bin (FSRB) with 10 bins is adopted. The frequency range of interest is from 0 to 2,000 Hz. The frequency components in each bin covering 200 Hz bandwidth are averaged out in amplitude. The average amplitudes from these 10 bins are then normalized and included in the feature vector.

From Fig. 12, comparing between the feature vector of the case with 44,100 Hz sampling frequency to that with 22,050 Hz sampling frequency, it is seen that at bin#8 to bin#10, these two feature vectors are slightly different; that is, the amplitudes from the case of 22,050 Hz sampling frequency is slightly higher. Additionally, at bin#4, the amplitude from the case of 22,050 Hz sampling frequency is slightly higher. These variations in the input feature vector can result in the degradation to the classification accuracy achieved by the ANN model. When the sampling frequency is further reduced, it is seen from the figure that the obtained feature vector is increasingly different from the one obtained by 44,100 Hz sampling frequency. The average amplitudes in these bins are levelled up compared to those obtained from the case of 44,100 Hz sampling frequency. And, as seen from the figure, as the sampling frequency is further reduced, the number of noticeably different bins is increased. These differences are from the effect of aliasing when sampling frequency is not high enough.

From the result shown in this Subsection, it is seen that applying Solution#1 (that is, only reducing the sampling frequency) does not deliver a good performance in terms of the classification accuracy when compared to the performance obtained from the case of using 44,100 Hz sampling frequency, as seen from Fig. 10. In the next Subsection, the results from applying the second solution, Solution#2, will be given.

4.2 Solution#2 Results

In order to both keep the performance of the gunfire sound classification and reduce the amount of memory used in the classification system, in this section, the result from applying both signal filtering and down-sampling processes is given. As seen from the previous Subsection, reducing the sampling frequency alone does not provide a good classification performance. Conversely, it strongly degrades the obtained classification accuracy because of the aliasing effect. To prevent such effect, in this section, an additional process; that is, signal filtering, is done prior to the down-sampling process. 4 different types of low-pass filter are adopted; that is, Butterworth, Chebyshev Type I, Chebyshev Type II, and elliptic filters. The cut-off frequency of these low-pass filters is set according to the sampling frequency to be used in the following process. The obtained classification accuracies of applying different low-pass filters for 22,050 Hz sampling frequency is shown in Fig. 13.

From Fig. 13, with 22,050 Hz sampling frequency, the classification accuracy curves from different low-pass filters are given. It is clearly seen that the worst curve is from the case of without filtering (shown by the solid blue line with circular markers). A significant increment in classification accuracy is obtained by applying a low-pass filter to the signal before passing it to the down-sampling process, as seen from the other 4 accuracy curves in the figure. At the SNR equal or lower than 5 dB, an increase of at least 10%

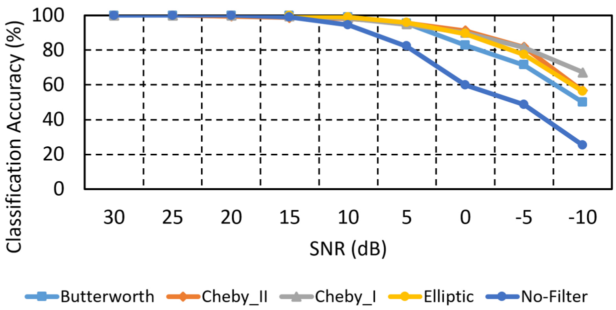

Fig.13: *Average gunfire sound classification accuracy obtained from Solution#2 with sampling frequency of 22,050 Hz.*

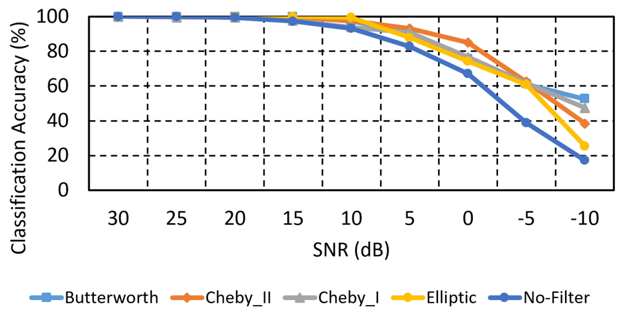

Fig.14: *Average gunfire sound classification accuracy obtained from Solution#2 with sampling frequency of 11,025 Hz.*

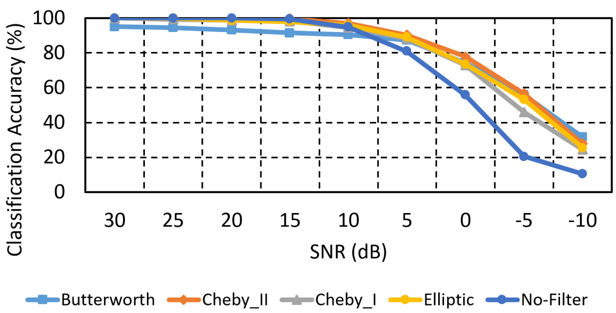

Fig.15: *Average gunfire sound classification accuracy obtained from Solution#2 with sampling frequency of 8,820 Hz.*

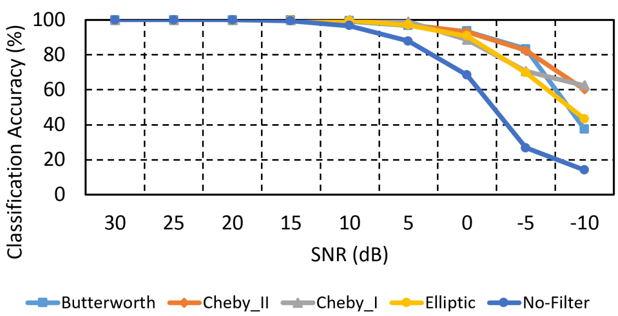

Fig.16: *Average gunfire sound classification accuracy obtained from Solution#2 with sampling frequency of 4,410 Hz.*

in classification accuracy is gained when comparing the accuracy from these 4 low-pass filtering cases to that from the case of no-filter. This improvement is achieved by applying the low-pass filtering prior to the down sampling process, preventing any possible aliasing effect to the output signal. Similar to the result shown in Fig. 13, with other sampling frequencies, the classification accuracy obtained from the case of using a low-pass filter is superior to that obtained from the case of no filtering. These are shown by Figs. 14 to 16.

It is seen from these figures that the lowest classification accuracy curve is from the case of not using filtering. Once filtering is applied, an improvement on the classification accuracy can be obtained. For example, at the SNR of –5 dB, at least 20%, 25%, and 40% increase in classification accuracy are achieved for the cases with sampling frequencies of 11,025 Hz, 8,820 Hz, and 4,410 Hz, respectively. At this point, from the results shown in Figs. 13 to 16, it can be concluded that applying low-pass filtering prior to down-sampling can help reduce the aliasing effect and, thus, increase the classification performance of the system.

Considering the effect of applying different types of low-pass filter to the system, it is seen from Figs. 13 to 16 that the obtained classification accuracy varies. These low-pass filters seem to provide similar performance in terms of the classification accuracy. However, if the figures are viewed carefully, it can be seen that the classification accuracy curve from the case of using Chebyshev Type II is better than that from the other 3 cases (that is, the cases with low-pass filtering). This can be explained by considering the responses of these filters as shown in Fig. 5. For Butterworth filters, there is a large attenuation in the passband near the cut-off frequency, this leads to a loss of the signal frequency components at that particular range of frequency. And, there is an allowable gain at the frequency higher than the cut-off frequency, this means that the unwanted signal outside the bandwidth of interest can affect the signal. For Chebyshev Type I filters, a ripple in the passband is allowed. This ripple can result in a degradation of the frequency spectrum in the band of interest. Similar effect is delivered by elliptic filters. Among these 4 types of low-pass filters, Chebyshev Type II is the only filter that allows most of the frequency spectrum of the signal to pass through the filter. Even though there are some ripples in the stopband, for the case of Chebyshev Type II filters; the gain allowed for these ripples is very low; that is, at least –10 dB compared to the highest gain of the passband region. Hence, the effect of the unwanted signal outside the passband is limited.

From the explanation given, it is seen that the frequency response of the filter affects the achieved classification accuracy. And, with this point, it can be seen that Chebyshev Type II is the most suitable to

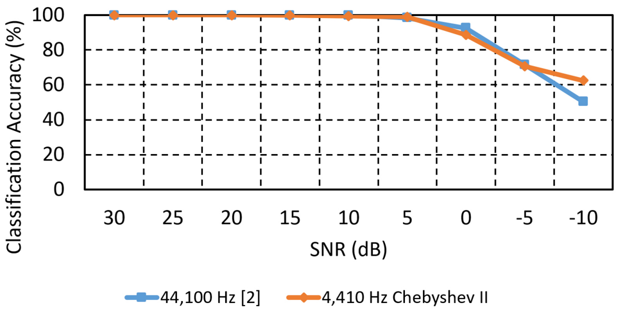

Fig.17: *Average gunfire sound classification accuracy obtained from [2] and Solution#2 with sampling frequency of 4,410 Hz and Chebyshev Type II.*

be used as the low-pass filter in the proposed classification system. Additionally, from Figs. 13 to 16, the best classification accuracy is obtained from the case of using Chebyshev Type II with a sampling frequency of 4,410 Hz. This classification accuracy curve is given and compared to the accuracy curve from [2], as shown in Fig. 17. Note that the accuracy obtained from [2] is the one with sampling frequency of 44,100 Hz as also shown in Fig. 10.

From Fig. 17, it is seen that the gunfire sound classification accuracy from both cases are almost identical. A slightly higher accuracy is obtained by the accuracy from [2] at the SNRs of 0 dB and –5 dB. While, at the SNRs of 5 dB and –10 dB, the classification accuracy from applying Chebyshev Type II low-pass filter with a sampling frequency of 4,410 Hz is higher. The differences in accuracy from both cases at the low SNRs is not significant. It can then be viewed that using a low-pass Chebyshev Type II filter with a sampling frequency of 4,410 Hz can deliver the same gunfire sound classification accuracy as offered by the classification system with 44,100 Hz sampling frequency. This is a significant finding since, with this proposed low-pass filtering and down-sampling system, the required amount of memory is lessened by 10-fold compared to what is required by [2]. Additionally, much less computational time is needed.

5. CONCLUSION

Two proposed solutions are studied in order to help reduce the computational resources in a gunfire sound classification system while keeping the classification accuracy as high as possible. The first solution is to apply only down-sampling process to the system. And, the second one is to apply both low-pass filtering and down-sampling to the system. In this work, 6 different gunfire sounds are used; that is, AK-47, AR-15, Shotgun, 9-mm., .38-Caliber, and .45-Caliber. An ANN model is adopted as the model for performing the classification process. The frequency spectral representative bins (FSRB) is used in feature vector selection with 10 bins from the input frequency spectrum ranging from 0 to 2,000 Hz.

The gunfire sound samples are captured and sampled first with 44,100 Hz sampling frequency. According to the first solution, the sampled signal is then down sampled again to the other 4 different sampling frequencies; that is, 22,050 Hz, 11,025 Hz, 8,820 Hz, and 4,410 Hz. This down sampling is done in order to reduce the amount of required memory in the FFT process at the feature extraction block. It is first found that applying only the down sampling does not offer a high classification accuracy. Actually, performing solely the down-sampling process leads to a degradation on the classification accuracy because of the aliasing effect, which introduces some unwanted signals into the frequency band of interest. The second solution is then studied. It is found that applying a low-pass filter prior to the down-sampling process can help improve the classification accuracy since most of the unwanted signal outside the band of interest is filtered out. It is found also that applying Chebyshev Type II and 4,410 Hz sampling frequency can deliver a very good classification performance while reducing the required amount of memory by 10-fold compared to that from the system with 44,100 Hz sampling frequency. The results shown in this work can certainly be applied to a classification work on gunfire sound with a small device at which the computational resource is limited.

ACKNOWLEDGEMENT

This work was supported by Faculty of Engineering, Naresuan University, Phitsanulok, Thailand.

References

[1] S. Tangkawanita, C. Pinthong, and S. Kanprachar, "Development of gunfire sound classification system with a smartphone using ANN," in *Proceedings of the 2018 International Conference on Digital Arts, Media and Technology (ICDAMT 2018)*, pp. 168–172, 2018.

[2] S. Tangkawanita, C. Pinthong, and S. Kanprachar, "Spectral Vector Design for Gunfire Sound Classification System with a Smartphone using ANN," in *Proceedings of the 2018 International Symposium on Wireless Personal Multimedia Communications (WPMC 2018)*, pp. 421–426, 2018.

[3] J. A. Mazurek, et al., "Boomerang mobile counter shooter detection system," in *Proceedings of SPIE in Sensors, and C3I Technologies for Homeland Security and Homeland Defense IV*, vol. 5778, pp. 264–282, 2005.

[4] R. C. Maher, "Acoustical Characterization of Gunshots," in *Proceedings of Workshop on Signal Processing Applications for Public Security and Forensics 2007*, pp. 1–5, 2007.

[5] R. C. Maher, "Modeling and Signal Processing of Acoustic Gunshot Recordings," in *Proceedings of 12th Digital Signal Processing Workshop & 4th IEEE Signal Processing Education Workshop 2006*, pp. 257–261, 2006.

[6] H. A. Sánchez-Hevia, D. Ayllón, R. Gil-Pita, and M. Rosa-Zurera, "Maximum Likelihood Decision Fusion for Weapon Classification in Wireless Acoustic Sensor Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1172–1182, 2017.

[7] A. Chacon-Rodriguez, P. Julian, L. Castro, P. Alvarado, and N. Hernandez, "Evaluation of Gunshot Detection Algorithms," *IEEE Transactions on Circuits and Systems I*, vol. 58, no. 2, pp. 363–373, 2011.

[8] A. Borzino, J. Apolinário, and M. Campos, "Consistent DOA estimation of heavily noisy gunshot signals using a microphone array," *IET Radar, Sonar & Navigation*, vol. 10, no. 9, pp. 1519–1527, 2016.

[9] S. Hengy, P. Duffner, S. DeMezzo, S. Heck , L. Gross, and P. Naz, "Acoustic shooter localisation using a network of asynchronous acoustic nodes," *IET Radar, Sonar & Navigation*, vol. 10, no. 9, pp. 1528–1535, 2016.

[10] J. George, and L. M. Kaplan, "Shooter localization using soldier-worn gunfire detection systems," in *Proceedings of International Conference on Information Fusion 2011*, pp. 1–8, 2011.

[11] D. Grasing and S. Desai, "Data fusion methods for small arms localization solutions," in *Proceedings of International Conference on Information Fusion 2012*, pp. 713–718 , 2012.

[12] F. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing Heart Sounds in Respiratory Signals Using Singular Spectrum Analysis," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 12, pp. 3360–3367, 2011.

[13] V. Elamaran, N. Arunkumar, A. F. Hussein, M. Solarte, and G. Ramirez-Gonzalez, "Spectral Fault Recovery Analysis Revisited With Normal and Abnormal Heart Sound Signals," *IEEE Access*, vol. 6, pp. 62874–62879, 2018.

[14] L. Tan and J. Jiang, *Digital signal processing: fundamentals and applications*, Amsterdam, Elsevier, 2013.

[15] B. P. Lathi, R. Green, *Essentials of digital signal processing*, New York, Cambridge University Press, 2014.

[16] K. Rao, D. Kim, and J. Hwang, *Fast Fourier Transform – Algorithms and Applications*, New York, Springer, 2010.

[17] A. S. Broughton, *Discrete Fourier analysis and wavelets*, New Jersey, John Wiley & Sons, 2009.

[18] M. Parker, *Digital signal processing everything you need to know to get started*, Burlington, Newnes/Elsevier, 2010.

[19] S. K. Mitra, *Digital signal processing: a computer based approach*, Boston, McGraw-Hill, 2011.

[20] S. M. Kuo, B. H. Lee and W. Tian, *Real-time digital signal processing: fundamentals, implementations and applications*, New York, Wiley, 2013.

[21] R. J. Kapadia, *Digital filters: theory, application and design of modern filters*, Weinheim, Wiley-VCH, 2012.

[22] H. G. Dimopoulos, "Optimal Use of Some Classical Approximations in Filter Design," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 9, pp. 780–784, 2007.

[23] F. Xiao, "Fast Design of IIR Digital Filters With a General Chebyshev Characteristic," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, no. 12, pp. 962–966, 2014.

[24] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 1st Edition, UK, Prentice-Hall, 2000.

[25] D. Graupe, *Principles of Artificial Neural Networks*, 2nd Edition, Singapore, World Scientific Publishing, 2007.

[26] V. Sze, et al., "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.

Settha Tangkawanit received the B.Eng. degree (second-class honors) in Computer Engineering and M.Eng. degree in Electrical Engineering from Naresuan University, in 2005 and 2008, respectively. Currently, he works at the Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University, Phitsanulok, Thailand. His research interests are in the area of embedded system, computer vision, wireless sensor network and IoT.

Surachet Kanprachar received his B.Eng. degree (first-class honors) in Electrical Engineering in 1996 from Chulalongkorn University, Bangkok, Thailand. He received both the M.Sc. and Ph.D. degrees in Electrical Engineering from the Virginia Polytechnic Institute and State University (VA Tech), Blacksburg, Virginia, USA in 1999 and 2003, respectively. Since 2003, he has been with the Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University, Phitsanulok, Thailand, where he is now an assistant professor. His research interests are in the area of optical fiber communications, coding theory, and artificial intelligence.