Proceedings of the 7th International and 45th National Conference on Fluid Mechanics and Fluid Power (FMFP)
December 10-12, 2018, IIT Bombay, Mumbai, India

FMFP2018 – PAPER NO. 189

Automatic Classification of Acoustic Signals from Gunshots

Tushar Singh
Research Assistant
Dept. of Aerospace Engineering,
IIT Bombay
Email: tusharsingh62@gmail.com

Sourav Bhattacharya
Master's Student
Dept. of Aerospace Engineering,
IIT Bombay
Email: srvb1995@gmail.com

Aniruddha Sinha
Assistant Professor
Dept. of Aerospace Engineering,
IIT Bombay
Email: as@aero.iitb.ac.in

Abstract
We describe several automatic classification techniques for distinguishing the acoustic signature of a gunfire's muzzle blast from that of the bullet's shock wave. This is a prerequisite for automatic acoustics-based localization of gunshots. We use the spectral content of the acoustic signal to perform the classification. The simplest method uses thresholding based on the full-width-at-half -maximum bandwidth of the signature. It demonstrates excellent results for a database of gunshots that we have collected. A more sophisticated approach based on convolutional neural networks (ConvNet) is also evaluated, with a view to future use in classification of signatures by the make of the gun and the calibre of the bullet. This method also delivered encouraging results in validation tests.

Keywords: *Convolutional neural network; Spectrum image; Gunshot detection; Pattern recognition*

I. INTRODUCTION

A gunshot is characterized by two acoustic signatures that are useful for its localization: the muzzle blast associated with the explosion of the charge, and the shock wave trailed by the bullet if it is moving supersonically. We assume here that an acoustic event of the gunshot, be it the muzzle blast or the shock wave, has been detected in the recorded microphone signal already. Then, the next step prior to the application of the localization algorithm is the classification of the event. In this paper, we propose to pursue this in two ways – a simple frequency-domain thresholding, and a more complicated application of artificial intelligence in the form of a convolutional neural network (ConvNet). The former is appropriate for the gunfire data that we have at present; the latter is necessary for a richer dataset [2].

A. Muzzle Blast

A conventional firearm uses an explosive propellant in its muzzle to discharge the bullet. The muzzle blast wave diverges with propagates spherically at the speed of sound in the ambient. The blast wave follows the inverse-square law of decay. Although the sound from this explosion travels in all directions at sonic speed, it is loudest in the direction of firing. Also, The blast wave interacts with the ground, buildings or any other objects that it encounters, which introduces the effects of reflections and absorption in the acoustic signature. A typical muzzle blast time-signature is presented in Fig. 3(a).

B. Shock wave

If the bullet fired by a gun travels faster than the speed of sound then it trails a shock wave (actually a Mach wave) in the form of a cone behind it. The leading edge of the bullet suddenly compresses the air in front of it and its trailing edge creates a corresponding expansion. This gives rise to the characteristic 'N' wave of a shock shown in Fig. 3(a). The period of this wave is related to the calibre of the bullet. The amplitude is primarily a function of the 'miss distance' – how far away from the microphone the bullet passes.

II. METHODOLOGY

A. Data Collection

An experiment was conducted at a firing range to record gunshots using an omni-directional Sennheiser MD-42 dynamic microphone. Frequency response and sensitivity of the microphone was 40-18000 Hz and 2.0mV/Pa ± 2.5dB respectively. Two firearms – a Glock pistol and an AK-47 rifle – were fired several times, and the microphone was placed at different distances with range between 5-25 meters and orientations (12°-243°) with

respect to the guns and their firing direction. Data was recorded for about 20 seconds at a time, within which expert shooters were requested to fire two shots from one of the two firearms. In total, data was recorded for 16 firings of the AK-47 rifle and 16 shots of the Glock pistol. The data were acquired at 50 kHz sampling rate using Measurement Computing data acquisition card USB-1690FS.

B. Feature Extraction

The spectral content of the time series recorded at a microphone is most suitable for event classification. In order for the method to be appropriate in a real-time scenario, we use a short-time segment of the timeseries. A study of the gunfire signals recorded in the firing range suggests that the event signature lasts no longer than 3 msec (see Fig. 3(a)). Moreover, the initial rise time of the signal from the noise floor is very sharp. The gunshot event detection algorithm was implemented using constant size sliding window technique based on thresholding to detect the time indices of the events and computed standard deviation of signal values in time window for thresholding. Thus, we select a window around the event peak starting from 1 msec before the peak and ending at 2 msec after the peak. Window is further used to center the data cloud around zero. It involves subtracting the mean from each window before applying FFT. It is therefore used to focus on the fluctuating part of the data, and retains only the relevant variations for analysis. Further, the amplitude of the signal in each segment was normalized to have an absolute maximum of unity. Fig. 3(b) shows the power spectral density of the signals within a segment calculated using short-time Fourier transform (STFT)[4]. Evidently, the broader muzzle blast signature in the time domain corresponds to a narrower peak in the frequency domain, and vice versa for the shock wave. The full width at half maximum (FWHM) is calculated in the standard manner, and they come out to be 0.6 and 2.6 kHz respectively for the two events shown. Thus, thresholding based on FWHM is a natural choice for primary classification of the two gunshot signatures.

C. Spectrum Generation

Apart from requiring a distinction between muzzle blasts and shock waves, advanced localization algorithms also use more specific information regarding the calibre of the bullet[3]. Such details require deeper classification of the shock wave event signals than is possible from a simple thresholding based on bandwidth. In this paper, we evaluate the suitability of ConvNets for this task; in particular, we use a flavour of ConvNets designed for image classification. To convert the event time series into an image suitable for application of ConvNet, we generate a spectrogram from the data. Since we are using a single segment of data (usually 3 msec long) once the event has been detected, and since the STFT is performed on this segment as before, the time axis of the spectrogram is trivial (see Figs. 1 and 2). Its frequency axis matches that of the PSD in Fig. 3(b), and the curve plotted there is converted to a contour plot now with the same information content. The ConvNet works under assumption that points close to each other in the image share some correlations, and these complex features may be learnt automatically to classify the signatures.

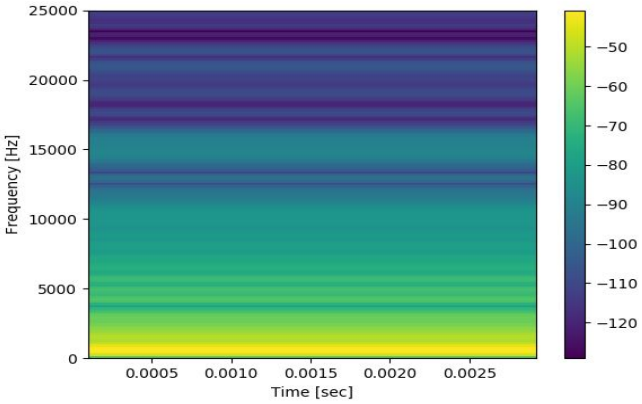

Fig 1: Single segment spectrum representation of AK-47 muzzle blast

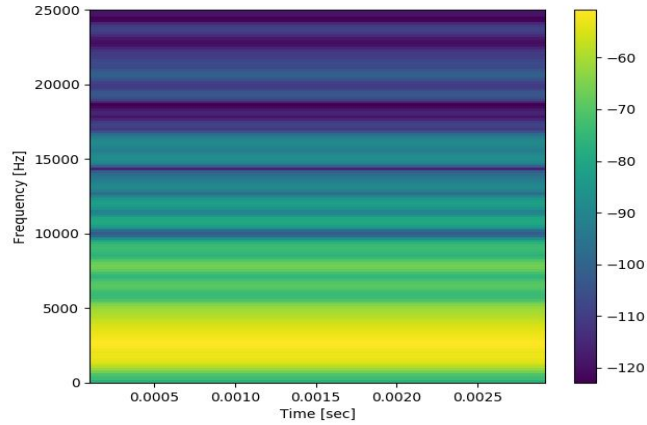

Fig 2: Single segment spectrum representation of AK-47 shock wave

D. Convolutional Neural Network Architecture

Convolutional neural networks was built by using Keras with Tensorflow as the computation back end and it take advantage of the fact that the input consists of images and they constrain the architecture in a more sensible way [1]. In particular, layers of ConvNets have neurons arranged in 3 dimensions: width, height, depth. Spectrogram images constitute an input volume of activations, and the volume has dimensions 64x64x3 (width, height, depth respectively). The final output would be a class score (probability) using sigmoid activation function with single output neuron because by the end of the ConvNet architecture we will reduce the full image into a single vector of class scores.

Layers Used to Build ConvNets

INPUT [64x64x3] will hold the raw pixel values of the image, in this case an image of width 64, height 64, and with three colour channels R,G,B.

CONV layer will compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. This may result in volume such as [64x64x32] if we decided to use 32 filters.

RELU layer will apply an element-wise activation function, such as the $max(0, x)$ thresholding at zero. This leaves the size of the volume unchanged ([64x64x32]).

POOL layer will perform a downsampling operation along the spatial dimensions (width, height), resulting in volume such as [32x32x32].

FLATTENING layer applied to create single large vector that contains all the different cells of all the different features maps. We manage to convert input image to one dimensional vector that contains some information of the spatial structure or some pixel pattern in the image.

FC (i.e. fully-connected) layer acts as a hidden layer and will compute the class score.

OUTPUT layer consist of a single neuron with sigmoid activation function for binary classification; it calculates the probability the class prediction.

Binary Cross Entropy Loss Function

We need to know the derivative of the loss function to backpropagate. Cross-entropy loss, or log loss, measures the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverge from the actual label. A perfect model would have a log loss of 0. In binary classification, cross entropy can be calculated as

$$-(y\log(p) + (1-y)\log(1-p)),$$

where, y is the binary indicator (0 or 1) if class label c is the correct classification for an observation o, and

p is the predicted probability that observation o is of class c.

An Adam Optimizer is used instead of the classical stochastic gradient descent procedure to update network weights iteratively based on training data. Stochastic gradient descent maintains a single learning rate for all weight updates and the learning rate does not change during training. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients.

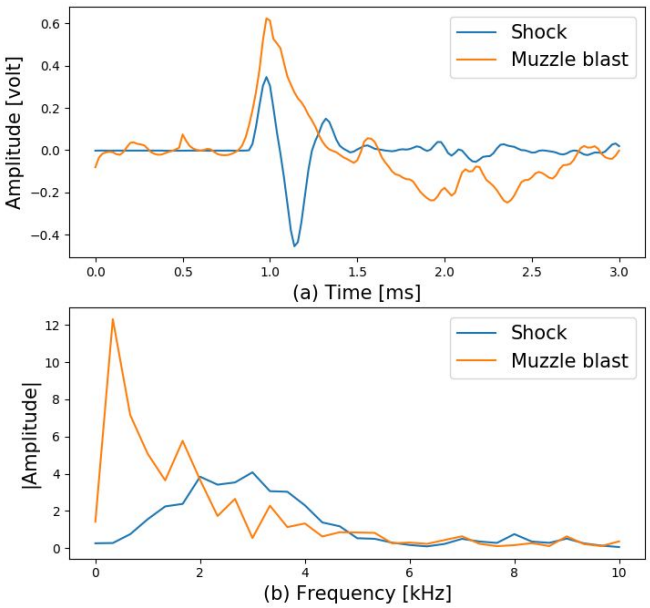

Fig 3: (a) Time-domain and (b) frequency-domain acoustic signatures of an AK-47 rifle firing

III. RESULTS AND DISCUSSION

Fig. 4 shows that 2 kHz is a suitable bandwidth-based threshold for classification of gunshot-related acoustic events as either muzzle blast or shock wave. A successful classification is obtained in all the cases available in our database. An evaluation was also performed using ConvNets on a validation set. We used the dataset consisting of 48 spectrum images for binary classification. They are divided into training and validation sets

containing 40 and 8 spectrum images respectively. At each epoch, the model was repeatedly learned under various condition, especially the structure of the network and its parameters on the training set, and at the same time evaluated on a validation set. Fig. 5 shows the 100% success rate on the validation set throughout after 42 epochs. Overfitting problem did not occur because training accuracy never reached 100%. Time estimation for training and classification is highly dependent on system specifications. Our training was performed on Intel® Core™ i7-4790 CPU @ 3.60GHz × 8 and it took 1.98 sec to perform learning on training set and recorded 0.0452 sec to made classification on single image.

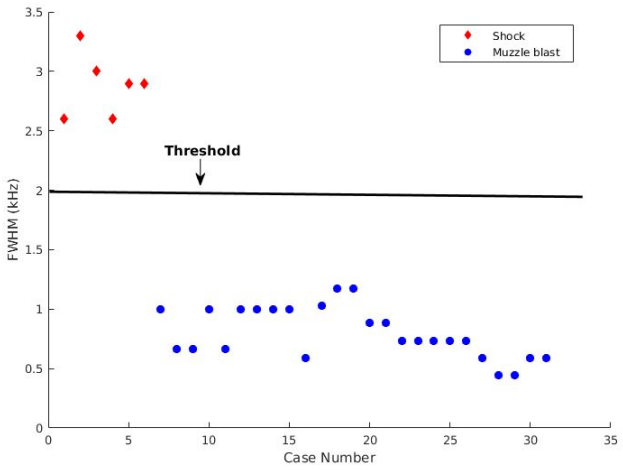

Fig 4: FWHM bandwidth for shockwave and muzzle blast

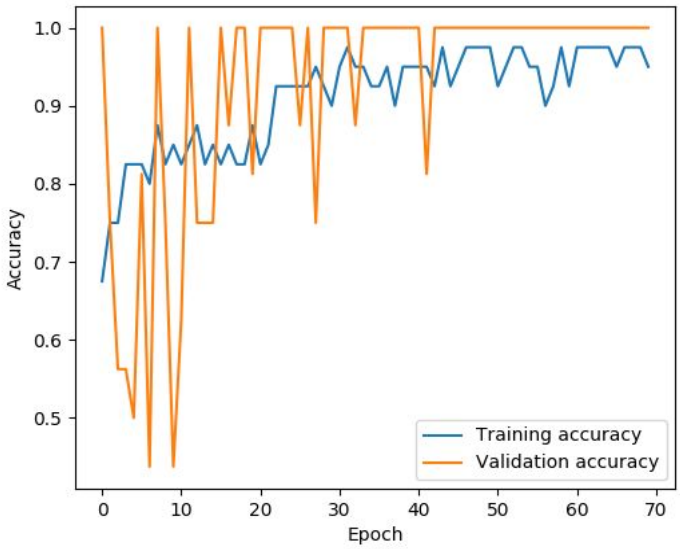

Fig 5: Classification accuracy on training and validation sets

IV. CONCLUSIONS

In this paper, frequency bandwidth features appears to provide sufficient information for accurate preliminary labelling of gunshot-related acoustic events. A more detailed classification may be obtained using ConvNets based off of spectral data treated as images. Although we do not have a rich enough database to fully evaluate this approach, preliminary results presented here indicate its promise for the future.

Acknowledgements

The authors acknowledge support from the National Center of Excellence in Technology for Internal Security (NCETIS) at Indian Institute of Technology Bombay.

References

[1] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 2006.

[2] Lorenzo Luzi, *Acoustic firearm discharge detection and classification in an enclosed environment*, Journal of Acoustical Society of America **139**, 2723 (2016)

[3] Volgyesi, Peter, et al. "Shooter localization and weapon classification with soldier-wearable networked sensors." *Proceedings of the 5th international conference on Mobile systems, applications and services*. ACM, 2007.

[4] M. J. Lighthill, *Introduction to Fourier Analysis*, Cambridge University Press, 1958