

Classification Of Small Arms Shock Wave Data By Statistical Clustering Of Actual Waveforms

L.J. Hamilton

Defence Science And Technology Group (DSTG), 13 Garden St, Eveleigh, Australia

ABSTRACT

Collections of acoustic shock waves generated by small arms fire for 5.56 and 7.62 mm bullet calibre have previously been classified using the two waveform features peak amplitude and duration [Ferguson, Lo, Wyber (2007). Acoustic sensing of direct and indirect weapon fire. ISSNIP 2007, 167-172]. In a very different approach, classification is investigated using unsupervised statistical clustering of the actual shock waveforms. Waveforms are essentially treated as geometrical objects, and are formed into groups with different shapes. Waveforms representative of the central tendencies of groups can then be used to classify other data. For shots fired from the same range in good weather conditions, 5.56 mm calibre waveform peak pressures and durations are less than those for 7.62 mm. However, the clustering revealed previously undetected artefacts which acted to extend some durations, causing each calibre to have two sets of waveforms instead of one, and leading to overlaps in durations for the two calibres. Clustering was able to isolate the anomalously extended waveforms and other types of anomalies in some cases, and is in principle a better data exploration and classification method for shock waveforms than feature based approaches.

1 INTRODUCTION

A bullet travelling at supersonic speed generates a near conical shock bow wave which travels with the bullet (Figure 1). The properties of the shock wave (Figure 2) can be used to estimate the calibre and closest distance to an acoustic sensor (the miss distance) of the bullet (Ferguson et al, 2007). Ferguson et al (2007) classified calibres 5.56 and 7.62 mm by partitioning a log-log graph of peak pressure amplitude and waveform duration into two areas with a smooth curve (Figure 3). Miss distances for the two calibres were monotonically spaced along either side of the separating curve. The log-log graph allowed Ferguson et al (2007) to classify calibres without first having to estimate miss distance as required by the usual conventional techniques. They found that errors due to unknown bullet speed became significant as miss distance increased, however data reduction and classification based on two features appears practical and effective.

Data of Ferguson et al (2007) are now reanalyzed using a statistical clustering technique which uses the actual waveforms, not features. Note that Ferguson et al (2007) investigated detection, classification, localization and tracking of sound sources for direct and indirect weapons fire, but the present analysis deals only with classification of small arms fire.



Figure 1. NASA shadowgraph of a supersonic bullet moving at 1.5 times the speed of sound showing the near conical bow shock wave, a turbulent wake, and weaker shock waves attached to the rifling of the bullet (https://www.nasa.gov/mission_pages/galex/2007 0815/f.html, public domain).





Figure 2. Shock waveforms. (a) 100 m range. As miss distances increase, waveform amplitudes decrease and durations increase. Waveforms rise rapidly to the zero upcrossing. (b) The initial rapid rise to the upwards zero crossing is not sustained or is not present. Note the high frequency modulation in some waveforms.



Figure 3. The feature space of Ferguson et al (2007) - a log-log graph of waveform peak amplitude vs waveform duration. Miss distances are color coded as 2.2 m (blue), 3 m (red), 5 m (green), 10 m (cyan), 20 m (black), 40 m (magenta), and 100 m (yellow). Points for 7.62 mm bullet calibre lie above the black curve, and points for 5.56 mm lie below it. Nominal range from the fire position is 475 m, wind speed is 15 knots.



2 DATA

Shots were fired at a target from six ranges (nominally 50, 100, 200, 300, 400, 500 m) in trials conducted in 2005 by Ferguson et al (2007). An array of seven high frequency quartz piezoelectric sensors perpendicular to the line of fire was placed on one side of the target to record shock waves corresponding to seven miss distances (1, 2, 5, 10, 20, 40, 80 m). An eighth sensor was placed on the line of fire 5 m behind the target to allow bullet speed to be estimated. Sample rate was 2.5 MegaSamples/second. Up to 300 bullets of 7.62 mm calibre, and 300 bullets of 5.56 mm calibre, were fired from each range in any one serial. Some ranges were repeated on the same day and on different days in different atmospheric and wind conditions.





The baseline pressure for all sensors has a large negative offset, and for some channels the baseline wanders, sometimes with a pseudo sinusoidal variation (Figure 4). Waveforms were located by their positive peak, and the baseline level was calculated for consecutive 2000-point segments prior to the peak, excluding the segment containing the peak. Records were then shifted to have approximately zero background mean pressure level.

Some waveforms were very noisy and others were affected by periodic high frequency amplitude modulations of unknown cause (Figure 2). Waveforms were used as recorded, although smoothing or filtering would likely improve classifications. It proved difficult to determine waveform starts and ends. Some waveforms have a small ramp-up in level above the background just before the actual start which can be confused with the actual start. The endpoint of the waveforms is a zero upcrossing from the negative waveform peak. Sometimes the initial rapid increase in value after the negative peak continues almost vertically upwards to the zero crossing, and the endpoint is well marked (Figure 2a). For other waveforms the initial rapid increase after the negative peak either lessens well before the zero level is approached, or is not seen (Figure 2b), and noise or signal modulation makes estimation of the endpoint ambiguous. These problems were avoided by *using only the portion of the waveform between the positive and negative peaks* for the clustering, both of which are well defined and simply calculated, except for some low energy almost flat topped signals. Lower energy signals with peak to peak ranges of 50 units were excluded.



3 METHODS

3.1 Conventional statistical clustering

Statistical clustering is a relatively simple operation of multidimensional analysis designed to find statistical objects having some measure of similarity. In the usual case the data objects are row or column vectors which hold *n* attributes describing points of a real or a mathematical space. The function of clustering is to automatically find if any of these points group together in the *n*-D space, that is, to find if there are sets of points with similar positions (i.e. similar properties). This process begins by finding which points are close together and which are far apart. The distance between points is often calculated as multi-dimensional variations of the power norm, defined by the Minkowski metric (Kaufman and Rousseeuw, 1990):

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}.$$
(1)

where $p \ge 1$, and x_i and y_i are vectors with the same number of elements (*n* in the present notation). The extended or *n*-dimensional Euclidian metric is given for p=2, and the *n*-dimensional Manhattan metric for p=1. Many other metrics exist, including the so called entropy metric. Clustering algorithms typically begin by randomly assigning objects to groups, then transferring objects to other groups if this minimises within-cluster variability and maximises between-cluster variability. Movement ceases when no more improvement in a global cost function based on these variabilities is obtained. See Kaufman and Rousseeuw (1990) or other clustering texts for further details.

3.2 Statistical clustering of single-valued curves

Rather than analysis of multi-dimensional points, statistical clustering methods can also be used to automatically sort or filter sets of single-valued curves into groups or classes with similar geometrical properties (shape, location, central tendency) without recourse to data reduction through feature analysis (Hamilton 2007, 2010, 2011, 2013). In essence each profile or curve is viewed as a discrete geometrical entity. Curves in any one cluster either have a different basic shape from curves in other clusters, or have essentially the same shape as curves in other clusters but a different average location.

The effectiveness of the clustering can be checked by examining overplots of the curves forming each cluster for uniformity of properties (shape, location, and central tendency), and by examining overplots of cluster medoids. Clustering of curves is optionally made with non-standardisation of curve values and Manhattan distance metric. The action of these selections for two single-valued curves is to calculate the difference in area between them (Hamilton, 2007), providing a simple geometrical basis for the clustering. Mathematically, this is very different in concept to clustering *n*-dimensional points by multi-dimensional distance estimates. Results can be summarised by a set of characteristic curves (medoids) representing the cluster central tendencies. The medoid waveforms can be used to make supervised classifications of other data.

3.3 The CLARA Clustering Algorithm

Hamilton (2007) showed the CLARA (Clustering LARge Applications) algorithm of Kaufman and Rousseeuw (1990) to be suitable for clustering of single-valued curves when configured to use non-standardisation of parameters and Manhattan distance metric as noted above. Use of standardisation allowed identification of outlier curves and sets of curves most different from others. Outlier detection, and removal if they are found to arise from errors or if they can be treated separately, can be made as a separate program run prior to the actual clustering. The CLARA algorithm can be very efficient at outlier detection, isolating for example a set of 4 differently shaped cumulative grain size curves from a total of 1281 when only 10 clusters were requested with data standardisation (Hamilton, 2007). It is also observed that CLARA forms clusters based on curve shape, not on the numbers of curves with similar shapes, i.e. CLARA is not overtly a density based algorithm. This and the facility for outlier detection make it suitable for shape based clustering of curves.

3.3.1 Sampling Techniques Used By The CLARA Algorithm

Many clustering algorithms require too much processing power, computer memory, or processing time to be tenable for analysis of large data sets (thousands to tens of thousands of data objects). Software program CLARA overcomes these limitations by coupling statistical sampling and clustering techniques. The algorithm first clusters several sets of randomly chosen subsamples (configured in this analysis as five subsamplings), then uses the particular subsampling returning best results to cluster the entire data set. This provides a fast



algorithm, at the possible expense of accuracy, although this is not observed in practice. Alternatively, all data can be clustered without the subsampling procedure. Run times of seconds to minutes can be obtained.

4 RESULTS

The data were initially examined in their entirety, regardless of bullet calibres or firing ranges. An initial 20clustering was made of 11,049 waveforms recorded over three days for both calibres and six ranges from 50 to 500 m. Some clusters were unique to a calibre, but this data exploration confirmed range to be a dominant factor in determining waveform characteristics. Waveforms for 500 m range were then clustered separately. Two data sets are examined, one for high wind of 15 knots, and one for low wind (0.5 knot), with three miss distances for each.

For a particular range to target each sensor should have two clusters of waveforms corresponding to the two bullet calibres. However, ten or more clusters were formed to allow identification of outliers or data errors. This may seem overkill, but it is better to have too many clusters in the initial data exploration than too few in order not to miss less obvious patterns or trends.

4.1 Miss distance 2 m, range 500 m, 15 knots wind

The data contained 42 records of 5.56 mm, and 143 records of 7.62 mm. Ten clusters were formed using five CLARA subsamplings of 50 samples each. *The two calibres each had two waveforms, rather than one each* (Figure 5). Many waveforms (defined from peak to peak) were extended towards the end by a section with a local maximum. The extended waveforms appear erroneous. The reason is unknown, but sensor ringing is a possibility. Extended 5.56 mm durations approach and exceed normal 7.62 mm durations, but a near 100% successful classification is obtained for both calibres because the four sets of waveforms have different shapes.

There are only two misclassifications, one extended 5.56 mm is classed as a normal 7.62 mm, and one normal 7.62 mm is classed as an extended 5.56 mm waveform. If these two cases are ignored, then two clusters contained only normal 5.56 mm waveforms, one contained only extended 5.56 mm waveforms, three contained only normal 7.62 mm waveforms, and four clusters contained only extended 7.62 mm waveforms. It is seen that the success rate of the 2-parameter feature classification for the extended 5.56 mm waveforms depends on the well defined separation of peak amplitudes for the two calibres at this miss distance. Ferguson et al (2007) state 99% correct classification for miss distances up to 10 m.



Figure 5. Results for 500 m range, miss distance 2 m, high wind. (a) Medoids - 5.56 mm (mauve, black) and 7.62 mm (red, green). (b) Overplots - 5.56 mm extended (black) and 7.62 mm normal (red).



4.2 Miss distance 10 m, range 500 m, 15 knots wind

The data contained 42 records of 5.56 mm, and 143 records of 7.62 mm. Ten clusters were formed using five CLARA subsamplings of 50 samples each (see Figure 6a). Note the reduction in peak amplitudes compared to 2 m miss distance. No waveforms obviously had the anomalous extensions in duration of section 4.1. *However, some waveforms for each calibre had much longer durations than others.* These also had a strong high frequency modulation and slightly higher peak amplitudes, indicating another type of artefact or changes in environmental conditions. They occurred randomly over the shoot, not at particular times (Figure 6b).

Two clusters contained a total of 31 bullets of 5.56 mm and 2 of 7.62 mm, six clusters had 132 of 7.62 mm and 2 of 5.56 mm, and two clusters had both 5.56 and 7.62 mm (total of 16). Overall 31 of 42 (=74%) of 5.56 mm bullets, and 134 of 143 (=93%) of 7.62 mm bullets could be correctly identified by the clustering. The clustering would have misclassed two 5.56 mm as 7.62 mm, and two 7.62 as 5.56, and could not separate 16 bullets of both calibres. Ferguson et al (2007) state 99% correct classification for miss distances up to 10 m.



Figure 6. Results for 400 m range, miss distance 10 m, high wind. (a) Cluster medoids. (b) Cluster assignments with time. Diamonds for 5.56 mm, squares for 7.62 mm.

4.3 Miss distance 40 m, range 500 m, 15 knots wind

The data contained 41 records of 5.56 mm, and 138 records of 7.62 mm. *Many waveforms did not have the well defined shapes or durations seen for smaller miss distances,* indicating the viable upper limit of miss distances at this wind speed (Figure 8). In an attempt to account for this, 20 clusters were formed using five CLARA sub-samplings of 50 samples each.

Eight clusters contain both calibres, two contain one 5.56 mm waveform each, ten contain 1 to 33 of the 7.62 mm waveforms (total of 66). *Overall the 20-clustering provides no discrimination for 5.56 mm calibre*. A cluster with four 5.56 mm and thirteen 7.62 mm regular waveforms indicates this is *because the two calibres often have much the same waveforms* (Figure 7a), and are no longer differentiated by amplitude or duration at this miss distance. Ferguson et al (2007) give no results for 5.56 mm calibre. They state 80% (=111/139) correct classification for 7.62 mm. For clusters containing only 7.62 mm waveforms the clustering percentage is much less at 48% (a count of 66). Adding clusters for which the 7.62 counts dominate the 5.56 counts would produce 82% correct, plus 16 of the 5.56 mm bullets incorrectly classed as 7.65 mm. Assessment of correct classification can differ depending on how it is calculated, and can be biased in favour of the higher number of 7.62 events.





Figure 7. Results for 500 m range, miss distance 40 m, high wind. (a) A cluster containing both 7.62 mm (black) and 5.56 mm shockwaves (green). (b) Ten examples of 7.62mm waveforms.

4.4 Miss distance 2 m, range 500 m, 0.5 knots wind

The data contained 134 records of 5.56 mm, and 131 records of 7.62 mm. Ten clusters were formed using five CLARA subsamplings of 50 samples each. *Two sets of waveforms are seen for each calibre* (Figure 8), but because the whole waveform is used there is only one misclassification, a 7.62 mm classified as 5.56 mm. The success rate of the 2-parameter feature classification for the extended 5.56 mm waveforms depends on peak amplitudes for the two calibres having distinctly different values. Ferguson et al (2007) state 99% correct classification for miss distances up to 10 m.



Figure 8. Cluster medoids for 400 m range, miss distance 2 m, low wind. Black curves for 5.56 mm, red for 7.62 mm.



4.5 Miss distance 10 m, range 500 m, 0.5 knots wind

The data contained 129 records of 5.56 mm, and 135 records of 7.62 mm. Ten clusters were formed using five CLARA subsamplings of 50 samples each. The ten medoids are shown in Figure 9. No waveforms obviously had the anomalous extensions in duration seen for other firings. However, *as for the high wind case for this con-figuration, some waveforms for each calibre had much longer durations than others*. These occurred randomly over the shoot, not at particular times.

Four clusters had only 7.62 mm calibres (126 bullets in all, 93% of the 135), two had only 5.56 mm (54 total), one had six of 5.56 and four of 7.62 mm, and three had dominantly 5.56 mm (total of 74 for 5.56 mm and 5 for 7.62 mm). Ferguson et al (2007) state 90% correct classification for miss distances out to 80 m.



Figure 9. Cluster medoids for 400 m range, miss distance 10 m, low wind. Black curves for 5.56 mm, green for a mixed cluster, red for 7.62 mm.

4.6 Miss distance 40 m, range 500 m, 0.5 knots wind

The data contained 108 records of 5.56 mm, and 133 records of 7.62 mm. *Many 5.56 mm waveforms did not have well defined shapes or durations, but 7.62 mm waveforms were well formed*. Ten clusters were formed using five CLARA subsamplings of 50 samples each. Figure 10a shows medoid waveforms. Four clusters have a total of 105 of 7.62 mm and one of 5.56 mm calibres. One has three of 5.56 mm. The other five contain a total of 28 of 7.62 mm and 104 of 5.56 mm. *The 28 misclassified 7.62 mm have shorter durations than expected*.

If the four clusters with only 7.62 mm bullets are used to estimate correct classifications, then the overall figures are 78.9% correct (105 of 133). In the other six clusters 107 of 135 bullets are 5.56 mm (=79.3% correct). *Waveforms in the four clusters for 7.62 mm have longer durations than all but four waveforms in the other six clusters (Figure 10b), so that a classification equivalent to the overall clustering results could be made on duration alone.* Ferguson et al (2007) give no results for 5.56 mm, and 90% probability of correct classification for 7.62 mm out to 80 m miss distance.





Figure 10. Results for 500 m range, miss distance 40 m, low wind. (a) Cluster medoids – 5.56 mm (green), 7.62 mm (red, black). (b) Overplots of waveforms, red and blue for 7.62 mm, black and green for 5.56 mm.

5 DISCUSSION

In the presence of suitable atmospheric conditions the peak amplitudes and durations recorded at any one sensor for the 5.56 calibre waveforms for smaller miss distances should be less than for 7.62 mm waveforms. The two-feature analysis method of Ferguson et al (2007) then provides an effective classification of these two calibres. However, statistical clustering of waveforms shows that even in low wind conditions the data for both calibres were occasionally affected by an apparent artefact which extended peak to peak waveform duration. This artefact had two forms. The first form appeared as a segment with an anomalous local maximum immediately before the negative peak. In the second form, durations for both calibres were extended without the obvious anomaly of the local maximum. Some extended 5.56 mm waveforms had greater durations than normal 7.62 mm waveforms, leading to potential misclassifications. Sometimes this problem was lessened because extension of 7.62 mm waveforms caused them to have durations not approached by extended 5.56 mm waveforms. Other types of artefacts also exist - in low wind conditions and larger miss distances some 7.62 mm waveforms were anomalously short, and were confused by the clustering with 5.56 mm calibre waveforms.

Clustering of waveforms with the CLARA algorithm of Kaufman and Rousseeuw (1990) as used by Hamilton (2007) is a fast effective method of data exploration for the shockwave data. It has revealed previously unknown artefacts in the shockwaves, and can sometimes account for them or lessen their effect because it essentially compares waveform shape. Feature analysis does not have this versatility. It is not known what causes the artefacts, but sensor ringing is a possibility. Once it is known that extended waveforms exist, some particular types of them can be routinely identified, and adjustments of sensor mounting or placement could be trialled to see if this removes them.

It is not possible to definitively compare errors for the feature detection and clustering methods because the clustering is used more to examine the data, rather than to make a full classification. It is also noted that the 7.62 mm calibre sometimes has three times more shots than the 5.56 mm for a particular set of conditions, and this biases any confusion matrix assessments.

ACKNOWLEDGEMENTS

This work was suggested by Principal Scientist Dr Brian Ferguson. Dr Kam Lo provided a Matlab function to read the shock wave data. Jane Cleary made initial plots of waveforms.



REFERENCES

- Ferguson, W.G., Lo, K.W. and Wyber, R.J. 2007. Acoustic sensing of direct and indirect weapon fire. *ISSNIP* 2007, 167-172.
- Hamilton, L.J. 2007. Clustering Of Cumulative Grain Size Distribution Curves For Shallow-Marine Samples With Software Program CLARA. *Australian Journal of Earth Sciences* 54, 503-519.
- Hamilton, L.J. 2010. Characterising spectral sea wave conditions with statistical clustering of actual spectra. *Applied Ocean Research* 32(3), 332-342.
- Hamilton, L.J. 2011. Acoustic Seabed Classification For Echosounders Through Direct Statistical Clustering Of Seabed Echoes. *Continental Shelf Research* 31, 2000-2011.
- Hamilton, L.J. 2013. Methods to classify or group large sets of similar underwater signals. *Acoustics Australia* special issue. Vol 40, No. 3, 167-172.
- Kaufman, L. and Rousseeuw, P.J. 1990. Finding Groups In Data: An Introduction To Cluster Analysis. John Wiley, New York, 1990.