SOUND MODELING USING ARTIFICIAL NEURAL NETWORKS FOR VIRTUAL MINE DETECTION TRAINING

Hui He,
Department of Mechanical and Aerospace,
Missouri University of Science and Technology
E-mail: hhq2c@mail.mst.edu

ABSTRACT
In real field demining, soldiers can only judge the landmine location and type by the sound generated from a landmine detector. Therefore, a virtual landmine detection training system can replicate the sound in a realistic manner is imperative. In this paper, several sound datasets for various targets have been collected. Each dataset contains about 500 instances, each representing a different radius and height of the detector head away from the target. To study the characteristics of different landmine targets and to devise a mathematical model for sound generation, a multilayer perceptrons (MLP) artificial neural network (ANN) utilizing back propagation (BP) is implemented to represent the sound model. Neural networks including a particle swarm optimization (PSO) based neural network and a genetic algorithm (GA) based neural network are also applied to the datasets to obtain a good mathematical model for sound estimation and generation. The mean squared error (MSE) resulted from the different methods is compared with each other, and it is shown that PSO based neural network has the least MSE.

1. INTRODUCTION
Landmines continue to pose a serious threat to civilians and military forces worldwide. Current landmine detection technologies, including the military’s AN/PSS-14 landmine detector [1], serve as a mainstay in combating this threat. Each type of landmine gives a different auditory signal depending on such variables as metal content, buried depth, and proximity to other metal objects. The AN/PSS-14 mine detector has proven to be an invaluable asset to today’s war-fighter and will remain so for the foreseeable future. However, many soldiers cannot get the training required to build and maintain the skills needed to operate this and other landmine detectors.

Advanced military equipment (AME) corporation completed the development of a prototype landmine detection simulator in 2008. The prototype was a successful proof of concept and showed that a simulator can refine a soldier’s sweep techniques and increase his/her ability to discriminate between landmines and clutters. In that system, pre-recorded sounds were used for the auditory cues. It is easy to record and play pre-recorded sounds but there are many drawbacks [2]. The sound recorded cannot be dynamically varied in response to changes within an interactive simulation environment. Also, a large sound library is required for using pre-recorded sounds. Furthermore, it is difficult and impractical to obtain sounds for all instances (heights and distances of detector, etc.) of an application.

The project described in the present paper is aimed at developing a virtual reality simulator to generate synthetic sounds to realistically replicate the functionality and auditory output of the AN/PSS-14 detector. Since generating a synthetic sound is flexible and dynamic, it is especially good for user interaction in a virtual reality environment compared to use of a pre-recorded sound.

2. SYSTEM OVERVIEW

Fig. 1. Physical training environment and four LEDs on the detector head.

The virtual landmine detection training system uses two wiimotes, which forms a stereo vision system to detect the position and orientation of the landmine detector on the training field. When the soldier sweeps the detector on the training lane, the motion tracking sensors (wiimotes) capture the position of the four LEDs mounted on the landmine detector and sends the position and orientation information to the motion tracking software via bluetooth. Motion tracking software reads the position information sent by the sensors and computes the location of the landmine detector in 3D coordinate system. The
3D location of the landmine detector is then sent to the landmine detection training simulator which utilizes those data to generate sounds with a sound model. Figure 1 shows the physical setup of the virtual reality training environment and the four LEDs mounted on the head of the detector.

3. SOUND MODELING

The primary objective of sound modeling for virtual landmine detection training is to generate the metal detection sound in response to user movement of the physical detector. The virtual landmine detection training system development consists of sound acquisition in the real world, sound data analysis, mathematical modeling of the sound, and sound rendering in the virtual landmine training system. The block diagram for the procedure of sound data collection, analysis, modeling and rendering is given in Fig. 2.

3.1 Frequency Analysis of Sound Clips

Sounds obtained from the landmine detector are composed of many frequencies. By using the fast Fourier transform (FFT) [3] to perform frequency analysis of the recorded sounds from a real landmine detector, we find that the major frequencies generated by the detector have a base frequency and several harmonic components (Fig. 3). By examining the frequency plot and the measured sound level with respect to the detector/target position, it is possible to determine the audio frequencies and sound level as functions of the detection location and the target type.

3.2 Sound Rendering

Based on the data collected, the lowest frequency is 533 Hz, and the highest amplitude is -3 dB, which are used as baseline tone. The advantage for 533 Hz being the base frequency is ease of modeling, since 800 Hz is 1.5 times that of 533 Hz and 1600 Hz is 3 times of 533 Hz. Hence, all the frequencies can be exported using numbers ranging from 1 to 3 as multipliers, and the amplitude multiplier spans from 0 to 1. Baseline tone can be generated in Adobe Audition using a base frequency and harmonic components. FFT spectral of the generated tone reveals a similar pattern as the recorded tone, but with much less noise in the signal as shown in Fig. 4.

To replicate a sound, we first obtain the frequency and amplitude information of the sound with respect to the distance and height between the detector and the landmine target. Then we determine the functions of the sound frequencies and amplitudes with respect to the distance and height of the detector head from the target. With the frequency and amplitude functions, synthetic sounds are generated, continuously, depending on the location of the detector relative to a mine target. In the present paper we will discuss only sound frequency. And the sound level can be obtained using the same methodology.

4. SOUND DATA COLLECTION

Sound data was collected for several landmine targets. Each dataset contains around 500 instances, each representing a different radius and height of the detector head from the target. The sound level and main frequency for each target were collected by placing the AN/PSS-14 detector at various distances from the target starting from 0 cm to 20 cm, with 0.625 cm increments. Sound data was also collected at various heights from 0 cm to 10 cm, with 0.625 cm increments. The dominant frequencies and levels of the sound obtained at different radii and heights can be read directly from the graphic user interface of the spectral surplus software. We investigate representing the sound frequency using a mathematical equation with the least mean squared error (MSE) by utilizing artificial neural network (ANN). Table 1 provides sample sound frequency data obtained from a landmine target. The sound frequencies are collected in terms of Hz.
The swarm adjusts the weights, i.e., neural network parameters, of the ANN methods are compared to each other. Between the expected output and the actual output, the error found in the output of the network is the difference between the expected output and the actual output.

For a given input X, the output of a three-layer MLP neural network can be computed with the following equations:

\[
y = \sum_{i=1}^{2} v_i \sigma(\sum_{j=1}^{2} w_{ij} x_j) = [v_1, v_2][z_1, z_2]^T
\]

where \(z_1 = \sigma([w_{11} w_{12} w_{13}] \cdot X)\),
\[
z_2 = \sigma([w_{21} w_{22} w_{23}] \cdot X),
\]
\[
X = [x_1 \ x_2 \ x_3]^T,
\]

and \(\sigma(\cdot)\) is an activation function. The overall nonlinear relationship between the input and the output is realized by the activation functions typically have the form of a smooth switch function, e.g., the sigmoid function:

\[
\sigma(\alpha) = \frac{1}{1+e^{-\alpha}}.
\]

For the input layer, \(x_i\) is the horizontal distance between the detector and the virtual mine, \(x_j\) is the vertical distance (height) between the detector and virtual mine, and \(x_k\) is the bias, which is set to be 1. In order to train the values of \(w_{ij}\) and \(v_i\), three types of training methods are utilized: back propagation (BP), particle swarm optimization (PSO), and genetic algorithm (GA). These three methods are compared to examine which of them will make the ANN model-predicted output match that of the desired data set. All of the three training methods are programmed with MATLAB.

5.1 Back Propagation (BP)

BP is one of several possible learning rules to adjust the connection weights during ANN learning by example. Learning occurs when the network weights are adjusted as a function of error found in the output of the network, which is the difference between the expected output and the actual output. The weights are adjusted backwards (back-propagated) through the ANN network until the error is minimized for a set of training data.

5.2 Particle Swarm Optimization (PSO)

PSO is a stochastic global optimization method based on simulation of social behavior [5]. PSO exploits a population of potential solutions to probe the search space. It relies on the exchange of information between individuals (called particles) of the population (called swarm). Each particle adjusts its trajectory towards its own previous best position, and also towards the previous best position attained by any member of its neighborhood [4]. The previous best value of the particle position is called the \(p_{best}\). It has another value called \(g_{best}\), which is the best value of all the \(p_{best}\) positions in the swarm. The basic concept of PSO is that each particle in the swarm move toward its \(p_{best}\) and \(g_{best}\) locations at each time step [5].

In our study, PSO is used to train the MLP neural network with the same architecture and training epochs as the previous one. To train this neural network, all neuron weights are put together as one particle. Each particle is updated toward the global best position, which minimizes the difference between the neural network output and the desired value.
5.3 Genetic Algorithm (GA)

GA generally includes three fundamental genetic operations: selection, crossover and mutation. These operations are used to modify the chosen solutions and select the most appropriate offspring to pass on to succeeding generations. GA considers many points in the search space simultaneously and has been found to provide a rapid convergence to a near optimal solution in many types of problems. This MLP neural network has the same architecture and training epochs as the above two methods (BP and PSO). In the training procedure, all neuron weights are put together as parents. Through selection, crossover and mutation, the best offspring is selected for the final weight vectors to be used for representing the datasets.

6. RESULTS AND DISCUSSION

The three types of ANN training methods as described above are compared by checking the MSE between the desired output and the trained output for each method. Figure 6 shows the MSE vs. number of iterations for the three techniques. As shown in the figure, the results provided by BP and GA are not as good as that provided by PSO when the number of iterations is higher than 200.

Table 2 provides the sample sound frequency (Hz) comparison between the actual data and the estimated data from the three types of ANN methods. Each iteration number is set to be 200; the range of the radius is from 12.8 cm to 15.3 cm, and the height is between 0 cm and 1.2 cm. By comparing the three sets of estimated data, it is shown that PSO has the least MSE and the PSO estimated frequency is very close to the actual frequency. Thus, PSO based ANN gives the best result among the three methods, and is capable of providing the sound frequency estimation for virtual landmine detection training simulation.

Table 2: Sound frequency (Hz) comparison between the actual data and estimated data from PSO/GA/BP at various positions.

<table>
<thead>
<tr>
<th>Radius (cm)</th>
<th>Height (cm)</th>
<th>Actual frequency (Hz)</th>
<th>PSO estimated</th>
<th>GA estimated</th>
<th>BP estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.8</td>
<td>0</td>
<td>533</td>
<td>528</td>
<td>505</td>
<td>621</td>
</tr>
<tr>
<td>12.8</td>
<td>0.6</td>
<td>533</td>
<td>538</td>
<td>530</td>
<td>626</td>
</tr>
<tr>
<td>12.8</td>
<td>1.2</td>
<td>546</td>
<td>554</td>
<td>555</td>
<td>632</td>
</tr>
<tr>
<td>13.5</td>
<td>0</td>
<td>568</td>
<td>567</td>
<td>554</td>
<td>637</td>
</tr>
<tr>
<td>13.5</td>
<td>0.6</td>
<td>575</td>
<td>580</td>
<td>578</td>
<td>642</td>
</tr>
<tr>
<td>13.5</td>
<td>1.2</td>
<td>600</td>
<td>598</td>
<td>600</td>
<td>648</td>
</tr>
<tr>
<td>14.1</td>
<td>0</td>
<td>616</td>
<td>615</td>
<td>594</td>
<td>651</td>
</tr>
<tr>
<td>14.1</td>
<td>0.6</td>
<td>623</td>
<td>620</td>
<td>615</td>
<td>656</td>
</tr>
<tr>
<td>14.1</td>
<td>1.2</td>
<td>643</td>
<td>644</td>
<td>635</td>
<td>662</td>
</tr>
<tr>
<td>14.7</td>
<td>0</td>
<td>655</td>
<td>651</td>
<td>630</td>
<td>664</td>
</tr>
<tr>
<td>14.7</td>
<td>0.6</td>
<td>660</td>
<td>654</td>
<td>649</td>
<td>670</td>
</tr>
<tr>
<td>14.7</td>
<td>1.2</td>
<td>678</td>
<td>677</td>
<td>665</td>
<td>675</td>
</tr>
<tr>
<td>15.3</td>
<td>0</td>
<td>691</td>
<td>684</td>
<td>662</td>
<td>678</td>
</tr>
<tr>
<td>15.3</td>
<td>0.6</td>
<td>693</td>
<td>687</td>
<td>678</td>
<td>684</td>
</tr>
<tr>
<td>15.3</td>
<td>1.2</td>
<td>709</td>
<td>702</td>
<td>691</td>
<td>689</td>
</tr>
</tbody>
</table>

Mean Squared Error: 23.13, 268.14, 2594.18

Fig. 6. Comparison of different ANN methods at various numbers of iterations.

7. CONCLUSIONS

The sound modeling method for developing a virtual landmine detection training system has been investigated. In order to determine the mathematical equation for the frequency and sound level, MLP-ANN training methods including the use of BP, PSO and GA are implemented and compared. The PSO proves to be capable of providing the best sound estimation.

8. ACKNOWLEDGMENTS

This research is funded by the Leonard Wood Institute and by the Intelligent Systems Center at the Missouri University of Science and Technology.