Classification of Educational Backgrounds of Students Using Musical Intelligence and Perception with the Help of Artificial Neural Networks

Naciye Hardalaç¹, Nevhiz Ercan², Firat Hardalaç³, Salih Ergüt⁴

Abstract - In this study we demonstrate that machine learning can be used to classify students who had backgrounds in positive sciences (including engineering, science and math disciplines) vs. social sciences (including arts and humanities disciplines) by the help of musical hearing and perception using artificial neural networks. Our 80 test subjects had an even mixture of both aforementioned disciplines. Each participant is asked to listen to a melody played on a piano and to repeat the melody himself verbally. Both the original melody and participants repetition is recorded and frequency and amplitude response is analyzed by using Fast Fourier Transform (FFT). This information is then used to train a neural network. Our results show that by using musical perception our neural network classifies students with positive and social science backgrounds at a success rate of 90% and 85%, respectively.

Index Terms - Artificial Neural Networks (ANN), Fast Fourier Transform (FFT), education, musical hearing, pure tone audiometry.

1. INTRODUCTION

How we perceive the sounds we hear around us and how we interpret them gives a good indication of our mental capabilities. For example, ones ability to perform well in music can be a sign that he will be doing well in mathematics as well. In this paper we investigate musical perception of a student and try to build a neural network classifier to identify his educational background using this information.

Neural networks have been widely and successfully used in the classification context. In [3] it is used to distinguish between pop music and classical music. [1] uses neural networks to automatically organize music according to its sound characteristics in such a way that similar pieces of music are grouped together. In this study we focused on a different classification problem: using the musical hearing to identify educational background of a person.

Sound is a vibrational wave that propagates in the medium. Single tone is produced in a single frequency as a periodic wave. Sound that we hear in our daily life consists of multiple sinusoidal components. These are called complex sounds and they can be periodical or aperiodical in nature. A periodical complex sound repeats the complex sound in a determined manner. Musical sounds are compromised of multiple sinusoidal components.

Hearing is the collection of the sound energy by ears and transferring this energy to the brain as the action potentials. The brain senses these action potentials as a sound. Resonance regions of internal ear and wavelength change according to the sound frequencies from outer medium. There is a maximum sensitive region (resonance region) for each frequency at the base membrane in the ear. While low frequency sounds stimulate the membrane region near the apex, high frequency sound stimulate near cochlea [2,7]. Any sound that comes to the ear shifts the place of stimulated region. This shift is called pitch changing in the brain. This pitch can be soft or sharp according to the place of excited neuron. The fact that the sound is separated into frequency components when it is transferred to the brain is the motivation for using a frequency transform on the sound in our analysis.

Throughout this paper we use the term positive sciences to refer to math, science, and engineering disciplines. Social sciences include the disciplines such as humanities and arts. Students with a positive science background received an analytical focused education beginning with high school years. Those with social science backgrounds, on the other hand, received an education that stimulated their verbal abilities. In this paper we investigated the effects of education on the musical perception and hearing.

This paper is organized as follows: Section 2 gives background on neural networks and Fast Fourier Transform (FFT) as well as mathematical model used. Section 3 discusses the experimental results. Section 4 evaluates the performance of the neural network and finally Section 5 provides the conclusion.

1 Naciye Hardalaç, Faculty of Education, Department of Music, Gazi University, Turkey
2 Nevhiz Ercan, Faculty of Education, Department of Music, Gazi University, Turkey
3 Firat Hardalaç, Department of Electrical and Electronics Engineering, Firat University, Turkey
4 Salih Ergüt, Department of Electrical and Computer Engineering, University of California, San Diego, sergut@ucsd.edu

1-4244-0257-3/06/$20.00 © 2006 IEEE

October 28 – 31, 2006, San Diego, CA

36th ASEE/IEEE Frontiers in Education Conference

M3F-9
2. NEURAL NETWORK MODEL

Neural network is a non-algorithmic method, which uses parallel computing technique. They imitate functioning of the brain. Even though inter-neuron communication speed is quite slow for the brain, parallel processing allows it to analyze very complicated data in a short period of time. Neural networks learn directly from current examples rather than programming.

Feed forward neural networks with multiple hidden layers have been widely used and showed to operate successfully (see Figure 1). Multi Layer Perceptron (MLP), which is a successful learning algorithm, is used in the training of the network. MLP is a back propagation algorithm and it computes the error at the output of the network and sets weights of neurons iteratively. This operation is spread out on all layers and the error in the output is reduced.

Deviations between the real and the predicted values are computed to evaluate the learning success of the network. The performance analysis of the artificial neural network is evaluated by statistical methods. Mean square error (MSE) is used to determine the compliance between the predicted output and computed network output.

The correlation coefficient (r) may be between -1 and +1 and when r is closer to +1, the compliance between network output and aimed output is higher [5]. The stopping criterion for the supervised learning is set over the curve of MSE (e.g. when MSE is below 0.01).

During the learning process, it is possible to determine how much each input affects the output by comparing their relative weights within the neural network. After successful termination of the learning process, the classification performance is determined by applying test data to the neural network. If the performance values meet the desired criteria at the end of the test, the structure of the neural network is completed and it is ready to classify any external data.

The velocity values, which are the amplitudes of the frequencies in dB at the output of FFT analysis, are applied to train neural network. These are on dB frequency corresponding to piano notes in Table 1 and they are obtained from the students with positive and social science backgrounds through FFT analysis. In addition to them the values obtained from the piano as a control group are also applied to the neural network. The magnitude of power spectrum at 20, 100, 120, 200, 320, 380, 420, 520, 600, 680, 770, 800, 880, 900 and 980 Hz frequencies is taken as an input to the neural network. A total of 15 parameters are applied to 15 input neurons. The number of neurons at the output layer is 3; these correspond to positive science background, social science background, and piano. During the learning of the neural network, back propagation and momentum are used and tangent hyperbolic (tanh) is selected as the transfer function. Note that momentum allows a network to respond to global gradients as well as local ones. Only one hidden layer was used in this study.

2.1. Fast Fourier Transform

A finite audio signal is framed with a window that is power of 2 (such as 64, 128, 256, etc) to take the FFT. Windowing technique is used to evaluate the frequency spectrum for the corresponding frame. Using windowing prevents the non-existing frequency components to appear in the spectrum. In addition, zero padding is applied to the same signal after windowing process. This entails certain overhead on the process although it increases the readability of spectrum.

Discrete Fourier transform of a discrete time periodic signal is defined as follows:

\[ X_k = \sum_{n=0}^{N-1} x(n) \exp(-j\frac{2\pi}{N}kn) \]  

where \( X_k \) is expressed as discrete Fourier coefficient, \( N \) is the frame size, and \( x(n) \) is the input signal on time domain. To obtain the frequency spectrum of this signal, logarithmic values of the squares of absolute values of \( X_k \) are found as shown below:

\[ P_k = 10\log(X_k^2) \]
3. EXPERIMENTAL SETUP

The experimental setup is as follows: A melody is played by a piano and then it is recorded. College students from two groups of educational background (positive and social sciences) were asked individually to repeat the melody they heard. Their verbal repetition is also recorded. All these recordings are quantized in order to process and analyze by computing their spectrums with FFT. FFT was performed on 256 samples and succeeding intervals overlapped 50%.

The spectrums of the original melody as well as samples from two different groups of participants are used to train the coefficients of neural networks. Using these setup data from a new participant is fed into the neural network and his educational background is predicted as an output.

In our experiments we used a Zimmerman piano with a relative frequency of 440 Hz, a Sony recorder, and a Pentium III 600 MHz PC (see Figure 2).

4. RESULTS AND PERFORMANCE EVALUATION

Figure 3 shows the power spectrum of the recording played on the piano. Similarly, we calculated the FFT of the recordings repeated by the students who volunteered to participate in our experiments. Figure 4 and 5 displays the power spectrum of the recordings corresponding to the students with social and positive science backgrounds, respectively.

Music contains pitches, rhythm, harmony, tones, and silence and hence music theory involves mathematics and physics. Positive science disciplines involve analytical thinking and dealing with numbers extensively. We believe this helps the students with positive sciences to do better in terms of repeating the melody they hear. As a consequence, their power spectrum (see Figure 4) resembles to the piano (see Figure 3) more than the students with social science backgrounds (see Figure 5).

4.1 t-test

The t-test provides test of significance when the standard deviation of a sample set is unknown due to its small size [4]. Therefore in statistical analysis it is used to verify that the results are not due to chance.

The t-value computed this way corresponds to a probability measure, or p-value. The smaller is this value the less likely that our results are achieved by chance. If p-value is less than 5%, the result is called statistically significant, and if it is even less than 1%, then the result is called highly statistically significant [4].
The amplitudes corresponding to these frequencies and the standard deviations of these amplitudes and the results of student’s t-test are analyzed by using SPSS statistical package program (www.spss.com) and the results of this analysis is shown in Table 1. According to this Table, the amplitude values, which correspond to these frequencies, take the frequencies of piano notes. Table 1 also shows amplitudes of the subjects that have positive and social sciences background.

When p-values on 20 Hz in Table 1 are examined, we see that p-values are p<0.05 and p<0.001 for students who have positive and social science backgrounds, respectively. The results for other frequencies between 100 Hz and 980 Hz are shown in Table 1 as well. They all have small p-values which is an indication of reliability of the results.

When we examine the relative weights of inputs we can calculate the weights of participants with positive and social science backgrounds that identify the sound frequencies of the piano.

As shown in Figure 6, only at frequencies 420, 520, and 680 Hz, the relative importance of the inputs exceed 0.1 for students with positive sciences background. The corresponding frequencies for students with social sciences background are 320, 420, 520, 770, and 980 Hz. Therefore these values have great affect on the classification process.

As seen in Figure 7, the MSE curve is below the 0.001, and it is seen that the prediction success rate for students who had positive and social science based education is 85% and 90%, respectively. These results are shown in Table 3.

4.3 Performance of the Neural Network

The results seen in Table 2 are obtained by applying test data to the neural network after the learning process is completed. As seen in Table 2, 17 and 18 students among the students with positive and social science backgrounds are predicted in success, respectively. When these results are evaluated, it is seen that the prediction success rate for students who had positive and social science based education is 85% and 90%, respectively. These results are shown in Table 3.

When we examine the relative weights of inputs we can calculate the weights of participants with positive and social science backgrounds that identify the sound frequencies of the piano.

As shown in Figure 6, only at frequencies 420, 520, and 680 Hz, the relative importance of the inputs exceed 0.1 for students with positive sciences background. The corresponding frequencies for students with social sciences background are 320, 420, 520, 770, and 980 Hz. Therefore these values have great affect on the classification process.

As seen in Figure 7, the MSE curve is below the 0.001, and from Table 3 we see that the correct classification rate is 85% and 90% for two groups of students. These results indicate that the architecture of the neural network we used here is successful.

4.2 MSE Curve

The momentum coefficient is chosen as 0.9 and the step size is chosen as 0.1 for our neural network implementation. Half of the sound records obtained from the students are used for training as a control group and the remaining half is used to verify the performance as a test group. The termination criterion for the supervised learning is set over the curve of MSE. The learning process stops when the MSE is below 0.01. The MSE curve obtained in the learning is shown in Figure 7. When the MSE curve is examined, it is seen that the neural network terminates quickly with success at the 69th step.

As seen in Figure 7, the MSE curve is below the 0.001, and from Table 3 we see that the correct classification rate is 85% and 90% for two groups of students. These results indicate that the architecture of the neural network we used here is successful.

4.3 Performance of the Neural Network

The results seen in Table 2 are obtained by applying test data to the neural network after the learning process is completed. As seen in Table 2, 17 and 18 students among the students with positive and social science backgrounds are predicted in success, respectively. When these results are evaluated, it is seen that the prediction success rate for students who had positive and social science based education is 85% and 90%, respectively. These results are shown in Table 3.

When we examine the relative weights of inputs we can calculate the weights of participants with positive and social science backgrounds that identify the sound frequencies of the piano.

As shown in Figure 6, only at frequencies 420, 520, and 680 Hz, the relative importance of the inputs exceed 0.1 for students with positive sciences background. The corresponding frequencies for students with social sciences background are 320, 420, 520, 770, and 980 Hz. Therefore these values have great affect on the classification process.

As seen in Figure 7, the MSE curve is below the 0.001, and from Table 3 we see that the correct classification rate is 85% and 90% for two groups of students. These results indicate that the architecture of the neural network we used here is successful.
In order to further test the results of the artificial neural network, data obtained from 10 students with even number of positive and social sciences backgrounds are applied to the input of the neural network. The prediction values and the real output values for this experiment are shown in Table 4. A comparison of the outputs to the real values, we see that the neural network predicts successfully. For example, when the rows 1, 2, and 3 in Table 4 are closely examined, we observe that the positive-science output of the neural network is close to 1 and the social-science output is close to zero and hence the neural network has predicted these data successfully. However, when see the results corresponding to row 4, we see that outputs are 0.72 and 0.41 for positive-sciences and social-sciences, respectively. Since 0.72>0.41 the neural networks predicts the output as positive-sciences, which is 0.03. Therefore the neural network successfully classifies this case as well.

The rows 7, 8, 9 and 10 are gives a clear social-sciences output since the output of social-sciences is close to 1 while the output of positive-sciences is close to zero. Only row 6 exhibits a different behavior, where output of social-sciences is 0.31 but this is still much higher than the output of positive-sciences, which is 0.03. Therefore the neural network successfully classifies this case as well.

As it is seen in Table 4, the neural network successfully predicts the output for participants with social and positive science backgrounds.

### Table 4

<table>
<thead>
<tr>
<th>No</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input 1</td>
<td>20Hz</td>
<td>46.0</td>
<td>46.0</td>
<td>45.0</td>
<td>44.8</td>
<td>45.8</td>
<td>45.0</td>
<td>45.2</td>
<td>44.8</td>
<td>46.1</td>
</tr>
<tr>
<td>Input 2</td>
<td>100Hz</td>
<td>39.4</td>
<td>40.0</td>
<td>41.2</td>
<td>40.5</td>
<td>39.8</td>
<td>40.0</td>
<td>39.0</td>
<td>41.1</td>
<td>39.8</td>
</tr>
<tr>
<td>Input 3</td>
<td>120Hz</td>
<td>41.4</td>
<td>42.0</td>
<td>42.1</td>
<td>41.1</td>
<td>40.8</td>
<td>38.0</td>
<td>37.0</td>
<td>39.1</td>
<td>38.4</td>
</tr>
<tr>
<td>Input 4</td>
<td>200Hz</td>
<td>39.0</td>
<td>37.0</td>
<td>38.5</td>
<td>36.9</td>
<td>39.2</td>
<td>36.2</td>
<td>35.0</td>
<td>36.1</td>
<td>34.9</td>
</tr>
<tr>
<td>Input 5</td>
<td>280Hz</td>
<td>36.2</td>
<td>36.8</td>
<td>37.0</td>
<td>35.8</td>
<td>36.4</td>
<td>35.0</td>
<td>34.0</td>
<td>36.1</td>
<td>34.9</td>
</tr>
<tr>
<td>Input 6</td>
<td>320Hz</td>
<td>39.0</td>
<td>37.0</td>
<td>36.8</td>
<td>37.5</td>
<td>37.9</td>
<td>34.8</td>
<td>34.0</td>
<td>35.1</td>
<td>34.8</td>
</tr>
<tr>
<td>Input 7</td>
<td>420Hz</td>
<td>36.1</td>
<td>36.4</td>
<td>36.9</td>
<td>37.1</td>
<td>35.8</td>
<td>38.8</td>
<td>35.0</td>
<td>36.8</td>
<td>36.7</td>
</tr>
<tr>
<td>Input 8</td>
<td>520Hz</td>
<td>36.0</td>
<td>36.0</td>
<td>35.4</td>
<td>36.5</td>
<td>36.7</td>
<td>36.0</td>
<td>35.0</td>
<td>35.4</td>
<td>35.2</td>
</tr>
<tr>
<td>Input 9</td>
<td>600Hz</td>
<td>36.0</td>
<td>36.5</td>
<td>37.4</td>
<td>35.6</td>
<td>36.2</td>
<td>34.8</td>
<td>34.0</td>
<td>35.1</td>
<td>33.9</td>
</tr>
<tr>
<td>Input 10</td>
<td>680Hz</td>
<td>36.1</td>
<td>36.6</td>
<td>37.6</td>
<td>35.5</td>
<td>35.2</td>
<td>35.2</td>
<td>35.0</td>
<td>35.6</td>
<td>35.8</td>
</tr>
<tr>
<td>Input 11</td>
<td>770Hz</td>
<td>37.4</td>
<td>38.0</td>
<td>36.8</td>
<td>37.4</td>
<td>36.0</td>
<td>35.4</td>
<td>35.0</td>
<td>34.2</td>
<td>33.8</td>
</tr>
<tr>
<td>Input 12</td>
<td>800Hz</td>
<td>38.1</td>
<td>38.4</td>
<td>37.8</td>
<td>36.5</td>
<td>39.4</td>
<td>36.0</td>
<td>36.0</td>
<td>35.6</td>
<td>35.2</td>
</tr>
<tr>
<td>Input 13</td>
<td>880Hz</td>
<td>39.4</td>
<td>40.0</td>
<td>41.2</td>
<td>40.6</td>
<td>39.8</td>
<td>37.0</td>
<td>35.0</td>
<td>36.5</td>
<td>35.2</td>
</tr>
<tr>
<td>Input 14</td>
<td>900Hz</td>
<td>38.1</td>
<td>38.6</td>
<td>36.8</td>
<td>37.4</td>
<td>38.2</td>
<td>39.0</td>
<td>38.0</td>
<td>36.9</td>
<td>38.6</td>
</tr>
<tr>
<td>Input 15</td>
<td>980Hz</td>
<td>46.0</td>
<td>46.0</td>
<td>45.2</td>
<td>44.8</td>
<td>45.7</td>
<td>47.0</td>
<td>46.0</td>
<td>48.7</td>
<td>47.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output</th>
<th>Social-science</th>
<th>-0.00074</th>
<th>-0.00209</th>
<th>-0.00262</th>
<th>0.41911</th>
<th>0.00472</th>
<th>0.31001</th>
<th>1.00045</th>
<th>0.99887</th>
<th>1.00019</th>
<th>1.00026</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>Positive-science</td>
<td>1.00419</td>
<td>1.00445</td>
<td>1.00990</td>
<td>0.31001</td>
<td>0.03085</td>
<td>-0.00347</td>
<td>0.04301</td>
<td>0.01805</td>
<td>-0.00377</td>
<td></td>
</tr>
</tbody>
</table>

Real values: PS PS PS PS PS SS SS SS SS SS

In order to further test the results of the artificial neural network, data obtained from 10 students with even number of positive and social sciences backgrounds are applied to the input of the neural network. The prediction values and the real output values for this experiment are shown in Table 4. A comparison of the outputs to the real values, we see that the neural network predicts successfully. For example, when the rows 1, 2, and 3 in Table 4 are closely examined, we observe that the positive-science output of the neural network is close to 1 and the social-science output is close to zero and hence the neural network has predicted these data successfully. However, when see the results corresponding to row 4, we see that outputs are 0.72 and 0.41 for positive-sciences and social-sciences, respectively. Since 0.72>0.41 the neural networks predicts the output as positive-science, which turns out to be the correct choice.

The rows 7, 8, 9 and 10 are gives a clear social-sciences output since the output of social-sciences is close to 1 while the output of positive-sciences is close to zero. Only row 6 exhibits a different behavior, where output of social-sciences is 0.31 but this is still much higher than the output of positive-sciences, which is 0.03. Therefore the neural network successfully classifies this case as well.

### Figure 7

![Mean Square Error (MSE) Curve](image)

**Figure 7**

**Mean Square Error (MSE) Curve.**

In order to further test the results of the artificial neural network, data obtained from 10 students with even number of positive and social sciences backgrounds are applied to the input of the neural network. The prediction values and the real output values for this experiment are shown in Table 4. A comparison of the outputs to the real values, we see that the neural network predicts successfully. For example, when the rows 1, 2, and 3 in Table 4 are closely examined, we observe that the positive-science output of the neural network is close to 1 and the social-science output is close to zero and hence the neural network has predicted these data successfully. However, when see the results corresponding to row 4, we see that outputs are 0.72 and 0.41 for positive-sciences and social-sciences, respectively. Since 0.72>0.41 the neural networks predicts the output as positive-science, which turns out to be the correct choice.

The rows 7, 8, 9 and 10 are gives a clear social-sciences output since the output of social-sciences is close to 1 while the output of positive-sciences is close to zero. Only row 6 exhibits a different behavior, where output of social-sciences is 0.31 but this is still much higher than the output of positive-sciences, which is 0.03. Therefore the neural network successfully classifies this case as well.

As it is seen in Table 4, the neural network successfully predicts the output for participants with social and positive science backgrounds.

### 4.4 ROC Analysis

Receiver operating characteristic (ROC) analysis is a widely accepted method for analyzing and comparing the accuracy of tests [9]. In our study, specificity refers to the false-positive percentage and sensitivity to the true-positive percentage. Conventionally these two parameters are plotted under...
different thresholds to produce the ROC curve. The area under the ROC curve has a special meaning. It represents that the probability of a randomly chosen subject will be classified correctly [6] and thus used to evaluate the performance of the neural network. As it is seen in Figure 8 the performance of our neural network is very satisfactory.

5 Conclusion

In this study we used neural networks to classify students with different educational background with respect to their ability to repeat a melody. Students were asked to repeat a melody verbally after it was played on a piano. The neural network was trained by the sound frequencies and amplitudes of the recording of the piano and repetition of 10 students. The training of the neural network converged successfully at the 98th step with an MSE error of 0.001. This neural network classified people with educational backgrounds in positive and social sciences with a success rate of 90% and 85%, respectively.

In the light of our results, we strongly believe that musical hearing and perception in artificial intelligence settings can help to evaluate analytical and verbal performance of students.

As a future work, we would like to investigate the effects of musical education on the development of student’s analytical and verbal ability.

REFERENCES


