International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 1014-1019 www.ijaem.net ISSN: 2395-5252

Extracting Audiogram Data Using Image Processing and Machine Learning Algorithms

Ferhat Mirac Sahin[1], Assist. Prof. Dr. Adem Ozyavas[2]
[1]*Computer Engineering Dept., Faculty of Engineering, Istanbul Aydin University Istanbul, Turkey*
[2]*Computer Engineering Dept., Faculty of Engineering, Istanbul Aydin UniversityIstanbul, Turkey*

Submitted: 25-05-2021 Revised: 01-06-2021 Accepted: 05-06-2021

ABSTRACT: Audiogram contains visual information about the hearing tests of a patient. A patient's hearing thresholds are encoded as a diagram in audiograms. Extraction such data could be useful since patients without having to visit a physician can obtain an interpretation of their tests. This study utilizes image processing and machine learning algorithms to extract important information from audiograms. First using Hue-Saturation-Brightness (HSB) technique, red (right) and blue (left) ear markings are obtained and removed from the image. The center positions of these red and blue markings are recorded for later processing. Using the OTSU method the binary image is obtained. The lines which mark the frequency and decibel values are determined using horizontal and vertical projections histograms from the binary image. Using a linear interpolation precise locations (frequency and decibel positions) are computed. The results of this study could be used to determine the hearing abilities of the patient.

KEYWORDS:image processing, machine learning, image segmentation, audiology.

I. INTRODUCTION

For people, sounds and being able to hear these sounds are as important and vital as the sense of sight. Because although the visual sense can perceive a limited area for a person standing in a fixed position, sounds from almost all directions can be perceived in hearing [1].

Sound is produced by the vibration of molecules belonging to an object that can stretch. These vibrations spread as waves at a certain speed in a medium such as air or water, and simply create the perception of sound when they reach the eardrum in humans. Sound waves change in direct proportion to the temperature of 20 °C and the increase in height. In addition, sound can have different speeds in different environments [2].

Hearing plays an important role in social communication, awareness of danger and perception of our position in space. Our stato-acoustic organ, the ear, basically consists of the balance and peripheral hearing system. The peripheral auditory system receives periodic air pressure changes and converts them into neural signals. The central auditory system, on the other hand, continues to process acoustic information and plays a role in the definition of directional hearing and sound pattern. The most important sound pattern that can be processed in human hearing is speech sounds [3].

II. THE PROPOSED METHOD

The processing of an audiogram image starts with scaling down of it. A sample audiogram image is shown in Figure 1.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 1014-1019 www.ijaem.net ISSN: 2395-5252

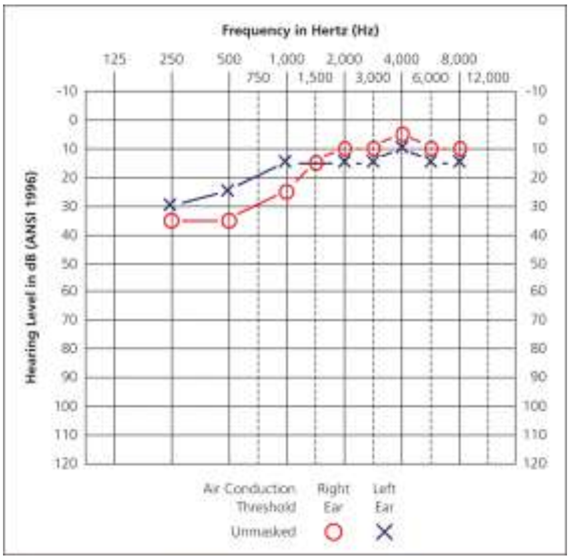

Fig. 1 Sample Audiogram Image

Scaling down the image to a certain width and its proportional height ensures that the template matching technique used for locating the blue crosses would be more accurate. It also makes it easier to determine the size of red circles in the image. Once the image is rescaled down to a certain width, the next step is to extract the red and blue pixels from the image for later processing. Since an audiogram image can have different shades of red and blue colors in it and hue-saturation-brightness (HSB) produces better results than the simple RGB color scheme in detecting different shades of colors, HSB scheme is utilized. The HSB (hue, saturation, brightness) color space defines colors as hue, saturation, and brightness, respectively [4].

•Hue determines the dominant wavelength of the color, for example yellow, blue, green, etc. It is an angular value of 0° - 360°, and in some applications 0-100 is normalized.

•Saturation determines the "vividness" of the color. High saturation will result in vibrant colors, while low probability will cause the color to approach gray tones. It varies between 0-100.

•The brightness determines the brightness of the color, that is, the ratio of white in it. It ranges from 0-100.

a. Obtaining Red Pixels and Central points of the Circles

Red and blue pixels are extracted using the HSB scheme. Figure 2 shows an example of red and blue pixels' extraction from the original image shown in Figure 1. After red and blue pixels are detected as white foreground pixels, the rest of the image is treated as the background image.

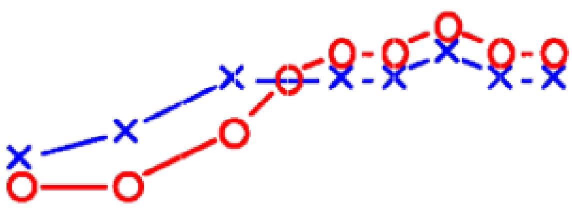

Fig. 2 Red and Blue pixels extraction using HSB scheme

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 1014-1019 www.ijaem.net ISSN: 2395-5252

The next step is to separate the red and blue pixels into their groups. The red pixels are shown in Figure 3 where the foreground pixels are shown as white and the background pixels as black. As it can be seen from the figure the circles' foreground pixels have some holes in them either because blue pixels or a line was over them in the original image.

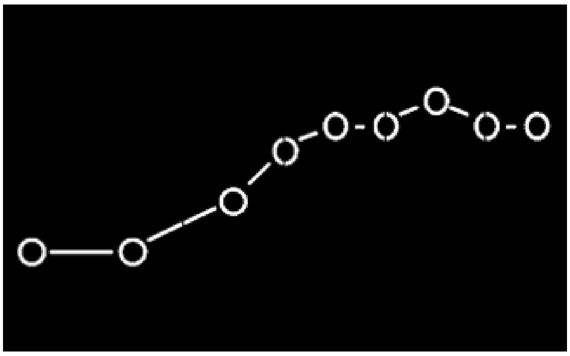

Fig. 3 Red pixels showing circles and lines in between with holes in them

The segmentation algorithm utilized is called the disjoint set forest (DSF) segmentation. In faster implementation of disjoint sets, we represent clusters with rooted trees, where each node contains a member and each tree represents a cluster. The root of each tree is its own parent and contains the delegate [5]. It is a graph-based linear time algorithm appropriate for real time applications. The algorithm first creates a graph by connecting every pixel to its neighboring 4 pixels where each pixel forms a node of the graph. Initially every node is the parent node of its own single-node graph. Then the algorithm checks to see if any two neighboring pixels can be included in the same set to form a bigger set of pixels. If two neighboring pixels are close enough which is determined by the minimum edge connecting any two nodes in these two subtrees these two pixels belong to. Eventually, once no more inclusions are possible each set of pixels is considered to be a segment. Since each pixels is dealt with one time by including or not including it to an existing set of pixels, the algorithm takes linear time, linear with the number of pixels in the image. Detection of the circles means detection of the background pixels enclosed by the circles. Since the image is converted to a binary image, determining segments is performed by connecting all background pixels in the image. A minor change is made in the DSF segmentation algorithm. Instead of using the Kruskal' s minimum edge to decide if a pixel should be included in another segment or not, equality of the representing pixels are checked. Once two pixels are of the same color, they are included in the same set of pixels. If not, they are kept in separate segments. As can be seen in Figure 3, there are ten segments in the binary image: 9 circles and the rest of the background pixels. Out of these ten segments the one with the largest number of pixels (the background black pixels) is discarded. The remaining (inner) circle pixels are the segments used to calculate the central points by averaging the x and y coordinates of the every pixel included in the segment.

b. Obtaining Blue Pixels and Central points of the Crosses

The same HSB technique is used to detect the blue pixels in audiogram images initially. All audiogram images are rescaled to a width and a corresponding height so that no distortion results. Every blue cross has almost the same size and orientation in the rescaled images. Therefore, template matching technique is used in the detection of the central points of the blue crosses. First all blue crosses are marked manually with an enclosing rectangular window. Several sample rescaled images are used for that purpose. In these enclosing windows those pixels part of the blue crosses are counted automatically as foreground pixels and those that are not as background. Majority voting of window pixels determines if a pixel should be counted as foreground or background over several images.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 1014-1019 www.ijaem.net ISSN: 2395-5252

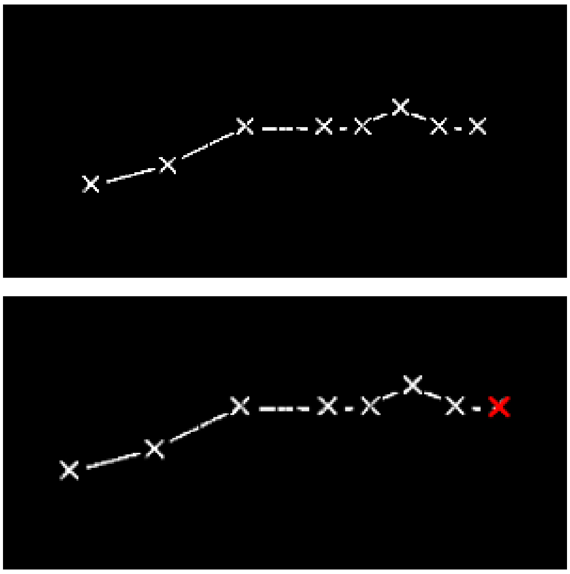

Fig. 4 Top shows blue pixels as foreground pixels. Bottom shows the rightmost blue cross' foreground pixels as red after majority voting

Once crosses are found, the central points are determined using the templates since the central points are already known in the templates.

c. Determining the Frequency and Decibel Values for the Central Points

The last step is determining the frequency and decibel values for each circle and cross central points. The proposed method uses horizontal and vertical projection histograms to compute frequency and decibel values of the vertical and horizontal lines. Vertical and horizontal projection histograms can be used for segmentation. After the red and blue pixels are removed from the original image, a binary image is obtained using the OTSU method. The lines and characters in the binary image which are the black/gray pixels in the original image are considered to be foreground pixels. Therefore, they are marked as white pixels. All remaining pixels are marked as background and hence as black pixels.

Figure 5 shows the binary image obtained from the original image shown in Figure 1. Note that the red and blue line beginnings are marked to ensure that those lines are correctly detected. Only the top portion of the image is shown below.

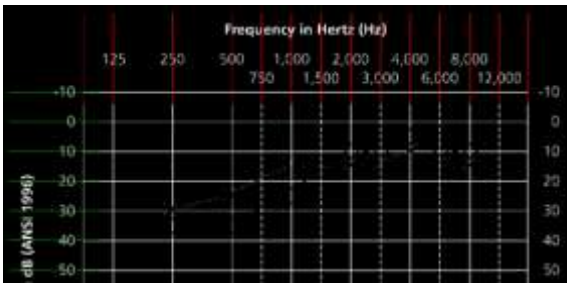

Fig. 5 The binary image obtained from the original image using OTSU's method.

The next step using the binary image is detecting the horizontal and vertical lines in the binary image. To that end, horizontal and vertical projection histograms are computed first. Horizontal and vertical projection histograms are used generally for segmentation. For these two histograms, the number of white pixels summed up horizontally and vertically. That is, for each row and column in the binary image the total number of white pixels are counted and stored. Figure 6 shows the horizontal projection histogram of the binary image shown in Figure 5.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 6 June 2021, pp: 1014-1019 www.ijaem.net ISSN: 2395-5252

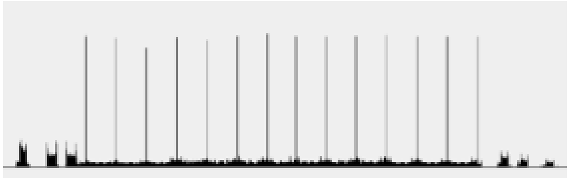

Fig. 6 Horizontal Projection Histogram

The long thin lines are the fourteen horizontal lines in the image. Vertical projection histogram is computed the same way counting the total number of white pixels for each column. Figure 7 shows the histogram of that. Note that the width of both histograms in Figure 6 and 7 is the width of the black and white image. Each horizontal and vertical projection histogram is stored in a one-dimensional array in the implementation. The length of these two arrays are the width and height of the binary image respectively. So, the array element at an index denotes the total number of white pixels at that row or column. Therefore, finding the indices of long thin lines would mean finding the row and column of where the lines are in the binary image.

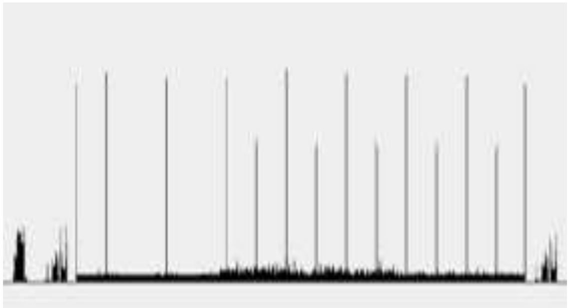

Fig. 7 Vertical Projection Histogram

Similarly, the long thin lines the vertical lines in the binary image. As you can see there are two bigger gaps on the left thin lines since the usual dashed lines that exist between solid vertical lines do not exist on the left of the binary image. Also, the longer lines represent the sum for the solid vertical lines and the shorter lines in the histogram are for the dashed lines in the binary image.

It is assumed and known that in audiogram images the first vertical line's reading starts with 125Hz. Similarly, the first horizontal line reading is marked with -10db. Every vertical and horizontal line has the set readings shown in the original image in Figure 1. Extraction of lines' positions from the binary image means to find the x and y positions of the lines in the horizontal and vertical projection histograms. A threshold value is used for that purpose. As it can be seen in Figure 6 and 7, drawing a horizontal line around the mid point would mean that any array index whose value is larger than the threshold would show the existence of a line in the binary image. In the implementation the threshold is computed by finding the standard deviation of the values in the horizontal and vertical projection histograms separately. Once a value in an array element is bigger than 30% of the standard deviation, that element's index is considered to be a line in the binary image. The red and green line beginnings in Figure 5 mark vertical and horizontal lines.
Computing lines' x and y coordinates help find the coordinates of the central points of the circles and crosses in the image. Using simple linear interpolation the central points of each circle and cross can be computed.CONCLUSION

To our knowledge, much of the study on audiograms are about interpretation of numerical audiogram data using machine learning algorithms. This study can be considered as a pre-processing step for such algorithms in that it generates the dataset for the interpretation. Also, patients with their audiogram images can use the work produced in this study to obtain the analysis of their hearing test results presented as audiograms without having to visit an audiologist. All the code in this study is implemented using the Java programming language.

REFERENCES

[1]. Mather G., Foundations of Perception, Psychology Press, 2006.
[2]. Barrett K., Barman S., Yuan J., and Brooks H., Ganong's Review of Medical

Physiology, The McGraw-Hill Companies Inc., 2003.
[3]. Probst R., Grevers G., and Iro H., Basic Otorhinolaryngology: A Step-by-Step Learning Guide 2nd edition, Germany: Thieme, 2006.
[4]. Reinhard E., Ashikhmin M., Gooch B., and Shirley P., "Color Transfer between Images", IEEE Computer Graphics and Applications 21(5):34-41, 2001
[5]. T. H. Cormen, C. E. Leiserson and R. L. Rivest, Introduction to Algorithms, The MIT Press, McGraw-Hill Book Company, 1990.