INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

A Machine Learning Based Clustering Protocol for Determining Hearing Aid Initial Configurations from Pure-Tone Audiograms

Chelzy Belitz, Hussnain Ali, John H.L. Hansen

CRSS: Center for Robust Speech Systems, The University of Texas at Dallas, Richardson, TX, USA
chelzy.belitz@utdallas.edu, hussnain.ali@utdallas.edu, john.hansen@utdallas.edu

Abstract

Of the nearly 35 million people in the USA who are hearing impaired, only an estimated 25% use hearing aids (HA). A good number of HAs are prescribed but not used partially because of the time to convergence for best operation between the audiologist and user. To improve HA retention, it is suggested that a machine learning (ML) protocol could be established which improves initial HA configurations given a user's pure-tone audiogram. This study examines a ML clustering method to predict the best initial HA fitting from a corpus of over 90,000 audiogram-fitting pairs collected from hearing centers throughout the USA. We first examine the final HA comfort targets to determine a limited number of preset configurations using several multi-dimensional clustering methods (Birch, Ward, and k-means). The goal is to reduce the amount of adjustments between the centroid, selected as a fitting configuration to represent the cluster, and the final HA configurations. This may be used to reduce the adjustment cycles for HAs or as preset starting configurations for personal sound amplification products (PSAPs). Using various classification methods, audiograms are mapped to a limited number of potential preset configurations. Finally, the average adjustment between the preset fitting targets and the final fitting targets is examined.

Index Terms: Hearing Aids, audiogram, audiometry, clustering, classification

1. Introduction

A 2008 survey found that in the United States, 34.25 million people reported having hearing difficulty. The same survey found that of those who reported hearing loss, only 24.6% owned hearing aids [1]. Another study examining data from the 1999-2006 cycles of the National Health and Nutritional Examination surveys estimate that only 14.2% of Americans over fifty with hearing loss wear a hearing aid. For those above fifty, that is one in seven individuals with hearing loss that wears a hearing aid [2]. Of those who do own a hearing aid, many choose not to wear it with numbers ranging from 4.2% to 24% in various international surveys reporting individuals who own a hearing aid but do not use it [3, 4, 5].

Studies suggest that mild to severe hearing impairment is associated with adverse effects on the quality of life, but that the adverse effects may be reversible using hearing aids [6, 7, 8]. Cost, lack of insurance coverage, and the necessity of repeat visits for customized fittings are among the barriers to obtaining a hearing aid for some individuals [9]. Among those who do own hearing aids but who do not use them, some of the most prevalent reasons cited across various studies are reported low hearing aid value and issues with the fit or comfort of the hearing aid [3]. Low hearing aid value is associated with a number of factors including issues with noise, difficulty adjusting to the hearing aid, and the necessity for repeated fine-tuning of the hearing aid to achieve the best programming for an individual. By reducing the amount of necessary adjustment, the programming process can be streamlined to make amplification products such as hearing aids more readily accessible to more individuals and to improve hearing aid retention. Furthermore, in the United States, the Over-The-Counter Hearing-Aid Act of 2017 has sparked interest in investigations of hearing aid configurations.

Another motivating application in the study of amplification product configurations includes determining settings for personal sound amplification products (PSAPs). A PSAP is a device similar to a hearing aid but which is intended to be used by an individual with normal hearing (NH). Studies on clustering hearing aid configurations can provide insight into the settings of PSAPs by providing a limited number of preset configurations which best represent a group of data points.

This study focuses on examining the comfort targets associated with a hearing aid setting. The comfort target is intended to be a measure which keeps the users of amplification devices in a comfortable hearing range. It is intended to ensure that high intensity sounds are not amplified to the point of nearing the pain threshold while ensuring that low intensity sounds are amplified in such a range that they are within the audible range of the subject. The hearing tests used include hearing threshold level (HTL) pure-tone audiograms.

The data set examined includes over 800,000 clients from over 500 locations throughout the United States. Within the data set, eleven test frequencies are listed: 125, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 Hz. Only 71% of clients listed in the database have a value listed for at least one test frequency. The majority of these clients, 70.34% of total entries, have entries for the core frequencies (500, 1000, 2000, and 4000 Hz). These four core frequencies are therefore the values examined in clustering and audiogram mapping. Roughly 16% of these audiograms are associated with a comfort target fitting resulting in a data set of over 90,000 audiogram-fitting pairs.

2. Proposed Solution

This study examines combining unsupervised and supervised machine learning methods to determine a small set of hearing aid configurations which can act as preliminary hearing aid settings based on audiogram results. The proposed solution treats the problem as two separate steps. The first step involves determining a limited number of preset configurations to represent the fitting targets. To determine the similarities among the fitting targets, various clustering methods are examined. The centroids of the clusters are then selected as the representative fitting targets which act as preset starting configurations for a hearing aid and can then be adjusted to better suit each individual. The second step of the problem maps pure-tone audiograms

http://dx.doi.org/10.21437/Interspeech.2019-3091

to the selected preset hearing aid configurations by assigning a classification to each audiogram based on the results of the comfort target clustering. Each audiogram is assigned to the same cluster as its associated comfort target. Various classification methods are tested to assign each audiogram to a preset configuration based on the cluster labels.

Fig. 1 highlights the proposed machine learning fitting process. In particular, after taking a hearing test, a client may be suggested an amplification product with one of the limited number of preset starting configurations. From there, a hearing aid may be adjusted to better suit the needs of the individual.

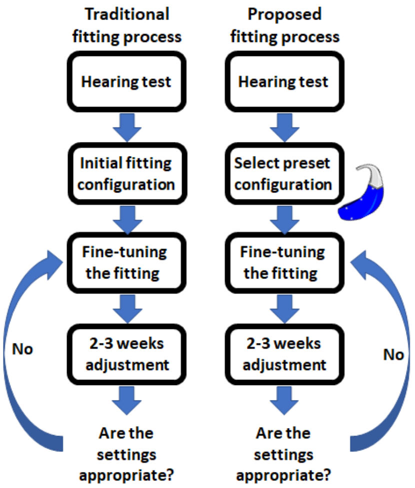

Figure 1: *An overview of the proposed solution (right) compared to the traditional process one undergoes to obtain a hearing aid (left). The goal is to optimize the hearing acoustic space for a core set of pre-selected HA configurations.*

3. Comfort Target Clustering

The goal of creating data clusters is to determine a small number of hearing aid comfort targets which, if used as a starting point for all the comfort targets within the cluster, results in the least necessary adjustment. The data point selected to represent each cluster is the mean value or the centroid value. In order to evaluate this, the average adjustment per frequency per cluster is examined as well as the average total adjustment. This value is akin to the mean absolute error with the cluster centroid or mean acting as the predicted value. Eq. 1 shows the calculation for the average adjustment per frequency per cluster. Here, y_i is the true value of the comfort target array, $\hat{y}_i$ is the estimated value (either the mean or the centroid; value selected is the one which produces the minimum MAE), and n is the number of comfort target arrays. Here, this value is considered the average total adjustment.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid y_i - \hat{y}_i \mid \quad (1)$$

Three clustering methods are explored: Ward agglomerative clustering, Birch clustering, and k-means clustering.

3.1. Clustering Investigation

In order to visualize the comfort target clusters generated by the various clustering methods, TSNE is used to project the data into a two-dimensional space. For an initial evaluation, a random sample of 10% of the data is selected and clustered.

Approaches such as DBSCAN, which is a density-based clustering algorithm, considered all data points to be noise. Spectral clustering, a computationally expensive clustering algorithm, ran on a small sample of the data set, generally considered the entire data set a single cluster.

3.1.1. Agglomerative Clustering

Agglomerative clustering is a hierarchical clustering approach which begins by considering each observation as its own cluster, then merges the mutually closest nodes into a single node in each step. The method by which clusters are merged varies per implementation, though common implementations include Ward, Complete, Average, and Single linkage [10]. Table 1 summarizes the results of experimenting with different linkage criteria. Note that Ward has the lowest average total adjustment, notably because it minimizes the variance within each cluster similar to k-means. Fig. 2 shows these clusters projected into a two-dimensional space. The clusters are well distinguished and clusters 2, 3, and 4 are of roughly equal size though cluster 1 contains nearly twice as many points.

Table 1: *A summary of the average total adjustment per linkage criteria.*

Linkage Criteria	Average Total Adjustment
$Ward$	5.15
$Complete$	5.56
$Average$	5.88
$Single$	8.63

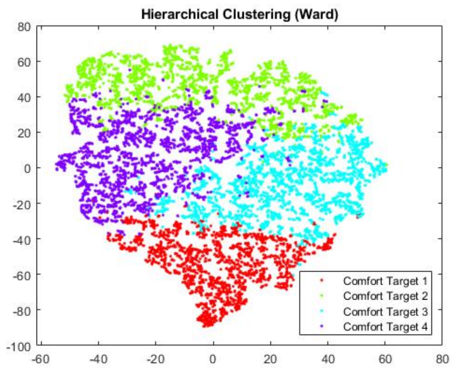

Figure 2: *The comfort target clusters created using Ward clustering.*

3.1.2. Birch Clustering

Birch is an agglomerative clustering algorithm which is used to perform hierarchical clustering over large data sets. The metric used in Birch clustering is the Euclidean distance between points [11]. Running the Birch clustering algorithm using the same data set yields the clusters seen in Fig. 3. The clusters again are well-defined and again of roughly equal size with cluster 1 being the largest however containing a different subsection of points.

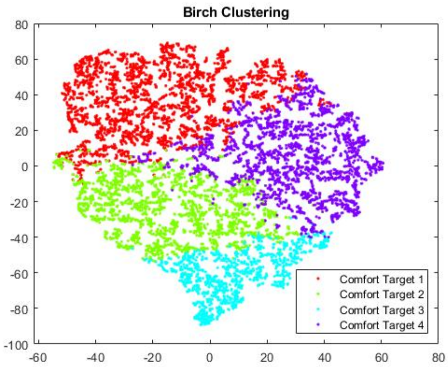

Figure 3: *The comfort target clusters created using Birch clustering.*

3.1.3. K-Means Clustering

K-means is a popular clustering algorithm which determines data clusters by iteratively selecting centroids then assigning data points to the cluster centroids until some convergence criterion is achieved. It clusters based on minimizing the inertia and creating groups with equal variance. Fig. 4 shows the clusters created using the k-means clustering algorithm. Notably, and as seen in the figure, cluster 4 contains the fewest points with the remaining clusters containing similar numbers of data points.

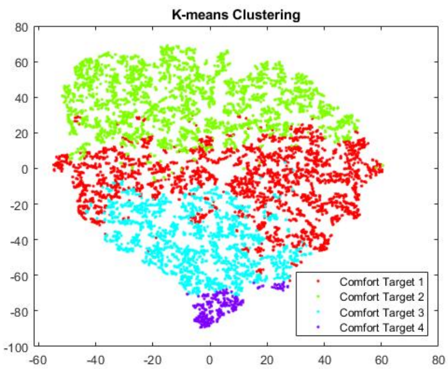

Figure 4: *The comfort target clusters created using k-means clustering.*

3.2. Clustering Comparison

Table 2 summarizes the average total adjustment for each of the clustering algorithms tested. Notably, the three tested methods performed similarly despite creating alternative cluster shapes. However, k-means performed overall the best and therefore will be the algorithm considered for subsequent audiogram mapping.

3.3. K-Means Cluster Analysis

From the results of the clustering comparison, k-means performed the best in terms of reducing the average total adjustment between the cluster centroids and the comfort targets assigned to each centroid. K-means clustering is therefore performed using the entire data set to assign each comfort target

Table 2: *A summary of the average total adjustment for each clustering algorithm tested.*

Clustering Algorithm	Average Total Adjustment
$Ward$	5.15
$Birch$	5.36
$K-Means$	4.92

to a cluster. These assignments are then applied to the audiograms associated with each comfort target. Again, using TSNE for cluster visualization, Fig. 5 shows the final comfort target clusters as well as the groupings created by assigning each audiogram the same classification as its associated comfort target. Notably, the groupings of the audiograms appear much more overlapped than the groupings of the original comfort targets.

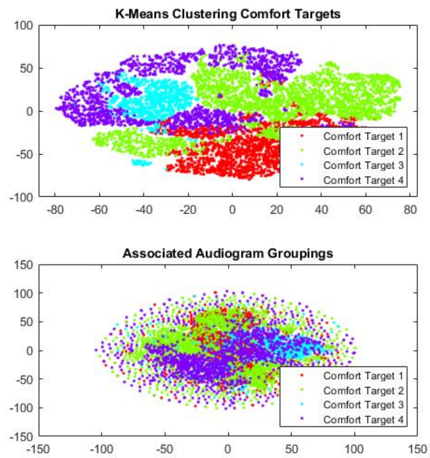

Figure 5: *Top: the comfort target clusters created using k-means clustering on all data. Bottom: the groupings creating by assigning each audiogram the same cluster labels as the associated comfort targets.*

Further analysis shows that by considering the standard deviation of each dimension independently, an average of 26.92% of comfort targets are within one-standard deviation of the centroid value of their respective clusters. This shows that a large portion of comfort targets are relatively close to the centroid. Table 3 summarizes the results for each cluster.

Table 3: *A summary of the percent of points falling within 1-3 standard deviations considered per dimension.*

Cluster	1-STD	2-STD	3-STD
$Comfort\,Target\ 1$	22.64%	82.73%	97.61%
$Comfort\,Target\ 2$	29.03%	84.90%	97.78%
$Comfort\,Target\ 3$	22.06%	83.18%	97.15%
$Comfort\,Target\ 4$	33.95%	88.39%	97.43%

Further, it may be noted that by the nature of this task, introducing more clusters will reduce the average total adjustment. The limitation, here four clusters, is selected to intentionally create a very small number of presets while achieving effective acoustic space coverage.

4. Audiogram Mapping

In order to map the audiograms to the centroid of their respective cluster, the cluster labels are treated as classification labels. Classification is then performed using 10-fold cross-validation to determine the best parameters for each classifier. Of the +90,000 audiograms with assciated comfort target fittings, 20% were set aside to determine the testing accuracies listed in Table 4. For this task, multiple classification methods were examined. Preliminary examinations looked at k-nearest neighbors (for its relative speed) and decision trees (for their easily interpretable nature). More robust classifiers were then examined including Random Forest, XGBOOST and Multi-layer perceptron classifiers resulting in slight increases in accuracy of classification as highlighted in Table 4. However, each method examined performed well, achieving similar results and mis-classifying the same subset of data. Examining Fig. 5, this result is anticipated due to the highly overlapped nature of the dataset.

Table 4: *A summary of the classification accuracy per algorithm.*

Classifier	**Test Accuracy**
$Decision\ Tree$	63.61%
$K - Nearest\ Neighbor$	63.23%
$Random\ Forest$	63.85%
SVM	64.19%
$XGBOOST$	64.29%
$Multi - layer\ Perceptron$	64.19%

To examine the effect of this misclassification, the average total adjustment is again considered to determine how much a hearing aid would need to be adjusted if the labels generated by the classifier are used to assign an audiogram to a particular preset fitting. In order to accomplish this, the comfort targets associated with each audiogram are assigned to the clusters based on the results of classification. Next, the average total adjustment is examined again. The results are summarized in Table 5. As anticipated, the correctly classified labels require an adjustment that is roughly similar to the true labels. The misclassified labels introduce more variance in the clusters, requiring a greater adjustment and increasing the average total adjustment when considering all points thereby increasing the adjustment for all classification labels produced using MLP classification.

Table 5: *The average total adjustments considering the classification labels determined using the MLP Classification Model.*

Labels	**Adjustment**
$True\ Cluster\ Labels$	4.89
$Correctly\ Classified\ Labels$	4.82
$Misclassified\ Labels$	9.48
$All\ Classification\ Labels$	6.49

In this particular application however, it is still useful to attempt to narrow down potential classifications. This limits the potential preset configurations which may be selected. By applying a softmax layer to the end of the MLP classifier and instead considering the top two classifications, it is possible to determine if the classification narrows down the potential starting configurations. For analysis, if either of the top two most likely classifications are correct, it is considered a correctly classified example when accounting for accuracy, called the top two accuracy here. This results in drastic improvements over the traditional test accuracy, indicating that it is possible, even with relatively simple classifiers, to narrow down the preset hearing aid configuration to one of two possibilities with a high degree of top two accuracy. The results are summarized in Table 6.

To further fine-tune the system, strong classifications may be specified to indicate cases where an audiogram very strongly falls into one classification and so the associated fitting may be recommended first. Defining a strong classification as one in which the likelihood associated with a particular classification is greater than the sum of the likelihoods that a data point falls within any of the other classifications, 82.68% of the data are considered strongly classified. The remaining data are considered weakly classified. Considering only strong classifications, there is a 68% accuracy selecting only the maximum argument of the softmax output (that is, considering only top individual correct classification). Applying a threshold to the definition of a strongly classified sample decreases the percentage of data considered strongly classified, however it also increases the strongly classified accuracy. For example, defining a strong classification as greater than the sum of the likelihoods of all other classes plus a threshold of 0.2 results in 59.5% of the data considered strongly classified but with a top individual classification accuracy of 73.3%.

Table 6: *A comparison of the accuracy of the system selecting the top one classification and the top two classification.*

	Accuracy	**Top 2 Accuracy**
$Strong\ Classifications$	68.00%	95.20%
$Weak\ Classifications$	43.80%	89.10%
$All\ Classifications$	63.80%	92.70%

5. Conclusions

By using unsupervised learning methods, it is possible to determine clusters of hearing aid comfort targets which share similar features within the acoustic hearing space. This, in turn, may aid in hearing aid retention by fine-tuning the programming process and reducing the time to convergence for best operation between the audiologist and user. In this case, it is advantageous to emphasize reducing the variance within clusters in order to minimize the amount a hearing aid must be adjusted from the centroid location selected to represent the cluster as a preset fitting to the final comfort targets within the cluster. Using k-means clustering and limiting the number of clusters to four, it was shown that it is possible to reduce the average total adjustment to 4.89. Assigning audiograms to the preset configurations proved to be a difficult task due to the highly overlapped data, however it was possible to select the top two configurations with a very high degree of accuracy.

Further considerations for future work could be to explore a wider range of hearing aids. It would also be beneficial to study classification methods which may reduce error in highly overlapped datasets.

6. Acknowledgements

This project was supported by a grant from Hearing Lab Technologies (HLT), who also provided access to all audiogram and hearing aid fitting data.

7. References

[1] S. Kochkin, "Marketrak viii: 25-year trends in the hearing health market," *Hearing Review*, vol. 16, 11 2008.

[2] W. Chien and F. R. Lin, "Prevalence of hearing aid use among older adults in the united states," *Archives of internal medicine*, vol. 172, no. 3, pp. 292–293, 2011.

[3] A. McCormack and H. Fortnum, "Why do people fitted with hearing aids not wear them?" *International Journal of Audiology*, vol. 52, no. 5, pp. 360–368, 2013.

[4] S. Hougaard and S. Ruf, "Eurotrak i: A consumer survey about hearing aids in germany, france, and the uk," *Hearing Review*, 2 2011.

[5] D. Hartley, E. Rochtchina, P. Newall, M. Golding, and P. Mitchell, "Use of hearing aids and assistive listening devices in an older australian population," *Journal of the American Academy of Audiology*, vol. 21, pp. 642–53, 11 2010.

[6] T. H. Chisol, C. E. Johnson, J. L. Danhaur, L. J. Portz, H. B. Abrams, S. Lesner, P. A. McCarthy, and C. W. Newman, "A systematic review of health-related quality of life and hearing aids: Final report of the american academy of audiology task force on the health-related quality of life benefits of amplification in adults," *Journal of the American Academy of Audiology*, vol. 18, no. 2, pp. 151–183, 2007.

[7] C. Mulrow, C. Aguilar, J. E. Endicott, M. R. Tuley, R. Velez, W. S. Charlip, M. C. Rhodes, J. A. Hill, and L. A. DeNino, "Quality-of-life changes and hearing impairment: A randomized trial," *Annals of internal medicine*, vol. 113, pp. 188–94, 09 1990.

[8] E. Tsakiropoulou, I. Konstantinidis, I. Vital, S. Konstantinidou, and A. Kotsani, "Hearing aids: Quality of life and socio-economic aspects," *Hippokratia*, vol. 11, no. 4, pp. 183–186, 2007.

[9] S. K. Mamo, N. S. Reed, C. L. Nieman, E. S. Oh, and F. R. L, "Personal sound amplifiers for adults with hearing loss," *The American Journal of Medicine*, vol. 129, no. 3, pp. 245–250, 03 2016.

[10] D. Mullner, "Modern hierarchical, agglomerative clustering algorithms," 09 2011.

[11] T. Zhang, R. Ramakrishnan, and M. Livny, "Birch: A new data clustering algorithm and its applications," *Data Min. Knowl. Discov.*, vol. 1, pp. 141–182, 06 1997.