Received 17 July 2024, accepted 6 August 2024, date of publication 12 August 2024, date of current version 22 August 2024.
Digital Object Identifier 10.1109/ACCESS.2024.3441762

RESEARCH ARTICLE

Efficient Personalization of Amplification in Hearing Aids via Multi-Band Bayesian Machine Learning

AOXIN NI[1], (Student Member, IEEE), EDWARD LOBARINAS[2], AND NASSER KEHTARNAVAZ[1], (Fellow, IEEE)
[1]Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA
[2]Callier Center for Communication Disorders, The University of Texas at Dallas, Richardson, TX 75080, USA

Corresponding author: Nasser Kehtarnavaz (kehtar@utdallas.edu)

This work was supported in part by the Seed Program Interdisciplinary Research funding provided by The University of Texas at Dallas as part of the National Research University Fund.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board at the University of Texas at Dallas.

ABSTRACT Personalization of the amplification function of hearing aids has been shown to be of benefit to hearing aid users in previous studies. Several machine learning-based personalization approaches have been introduced in the literature. This paper presents a machine learning personalization approach with the advantage of being efficient in its training based on paired comparisons which makes it practical and field deployable. The training efficiency of this approach is the result of treating frequency bands independent of one another and by simultaneously carrying out Bayesian machine learning in each band across all of the frequency bands. Simulation results indicate that this approach leads to an estimated hearing preference function close to the true hearing preference function in fewer number of paired comparisons relative to the previous machine learning approaches. In addition, a clinical experiment conducted on eight subjects with hearing loss indicate that this training efficient personalization approach provides personalized gain settings which are on average six times more preferred over the standard prescriptive gain settings.

INDEX TERMS Hearing aids, personalization of hearing aids amplification, machine learning, Bayesian learning.

I. INTRODUCTION

Widely used prescriptions of hearing aids, such as DSLv5 [1] and NAL-NL2 [2], involve setting gain values in a number of frequency bands based on a user's audiogram. An audiogram indicates the lowest level of sound pressure level (SPL) that a person can hear across audible frequency bands in a quiet audio environment [3]. There is a need to tailor the amplification function of hearing aids to noisy audio environments that are of particular interest to an individual user. Furthermore, with the recent introduction of more affordable Over-The-Counter (OTC) hearing aids, there is a growing need for their self-adjustment [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Carmen C. Y. Poon.

Any personalization or self-adjustment needs to be done in a simple and easy-to-use manner for it to be adopted by users. A complex personalization or self-adjustment involving too many "knobs" to adjust would be a major hindrance to its utilization. A number of simple methods, such as sliders, wheels, and pairwise comparisons, have been considered by researchers [5], [6], [7], [8]. Among these methods, the pairwise comparison method is often chosen due to its simplicity in hearing preference studies, e.g. [9], [10], [11], [12]. This method places minimal cognitive load on users as it merely involves selecting one out of two options similar to the pairwise comparisons in an eye exam.

A number of machine learning approaches have been developed in the literature to encode pairwise comparisons in a systematic way and to conduct personalization of the amplification function of hearing aids including the ones by our research team [13], [14], [15]. These approaches normally involve a trade-off between the size of the search space and the duration of the fitting process or training. In our latest work, the machine learning approach of Maximum Likelihood Inverse Reinforcement Learning (MLIRL) [14], [15] was introduced in order to achieve personalization of amplification in an on-the-fly or online manner. Although this approach was shown to produce personalized settings that were preferred over the standard settings by about 10 times, it required a training duration of at least one hour to go through a large number of paired comparisons. This relatively long training time poses a bottleneck that restricts deployment in the field.

In this paper, a new machine learning approach is developed in order to address the above shortcoming. This approach involves conducting Bayesian machine learning in each frequency band in an independent manner. As a result, the number of paired comparisons to reach a personalized setting is substantially reduced, thus making the online training process more feasible to deploy in the field.

The sections that follow are organized as follows. In section II, the independence aspect of the frequency bands is discussed. In section III, the developed Bayesian machine learning approach in each band is described allowing an efficient multi-band personalization. Then, the clinical setup for the subject testing conducted is mentioned in section IV. Finally, in section V, the results of our hearing preference and word recognition experiments are reported for eight subjects with hearing loss. The paper concludes in section VI.

II. INDEPENDENCE OF MULTI-BAND AMPLIFICATION

To personalize the amplification function in hearing aids, prescriptive gains across a number of frequency bands are taken to be the initial condition or the starting gain set. In this work, the DSLv5 hearing aid prescription is considered although any other prescriptions could be equally used to set the initial gain set. By adjusting these prescriptive gains within a lower and an upper bound, a set of personalized gains or a personalized gain curve can be reached as shown in Figure 1. The lower bound is established based on a user's audiogram. The upper bound is determined by the loudness discomfort level (LDL) to ensure that the output audio signal does not exceed this level.

Due to the nonlinearity of the human hearing perception, an audiogram consists of frequency bands in an increasing frequency range manner. The number of frequency bands in an audiogram is normally 5, 7, 9, or 11. Although the personalization approach developed in this paper is applicable to any number of frequency bands, for the clinical experiments reported in this paper, the number of frequency bands is considered to be five (0-500Hz, 500-1000Hz, 1000-2000Hz, 2000-4000Hz, and 4000-6000Hz) to make it easier for participants to express their hearing preferences. Furthermore, the number of gain adjustment levels in each band is considered

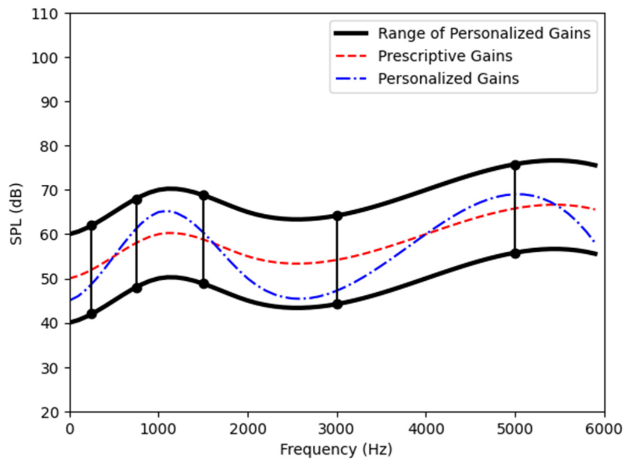

FIGURE 1. Depiction of the range of personalized gains around the prescriptive gains.

to be eight indicated by the set $\mathcal{X} = \{x^i \mid i = 1, \ldots, 8\}$ which correspond to these signal level changes $\{+12, +9, +6, +3, 0, -3, -6, -9\}$ dB. Personalization of gains is expected to provide improved hearing in an audio environment in which a user is having difficulty hearing. Note that regardless of the number of gain adjustment levels, the developed personalization approach is applicable to any number of gain adjustment levels. The size of the space that needs to be searched to find the optimum setting is given by the number of levels raised to the power of the number of frequency bands. For example, for five bands and eight adjustment levels, the size of the search space becomes 8^5.

Personalization is achieved via paired comparisons of audio signals conducted by users. A paired comparison involves a user listening to an audio signal which is passed through two different gain sets. Then, the user selects the preferred gain setting from the pair. Paired comparisons are easy to conduct and demand a simple feedback from the user. For S adjustment levels, it takes $S(S\text{-}1)/2$ paired comparisons to reach the optimum setting via the round robin tournament (RRT) method [10]. When the search space is too large, such as 8^5 noted above, it would not be practical or feasible to expect users to go through a very large number of paired comparisons. In our previous MLIRL approach [14], only three levels were used to keep the search space size to 3^5 or to keep the training time to about one hour for about 200 paired comparisons. One hour is a long time as far as users are concerned and thus a more efficient training approach is needed.

In order to reduce the number of paired comparisons, each band is assumed to be independent of one another and via the simulations described next it is shown that this is a valid assumption. Let us consider a true personalized preference function or gain set that a user desires to reach. Table 1 shows an example of the true personalized gain set [3, 4, 5, 2, 3]. A similarity between the true gain set and another gain set in a paired comparison is defined to be:

$$\text{Similarity}\,(\mathbf{A},\mathbf{B}) = \frac{\sum_{i=1}^{m} A_i B_i}{\sqrt{\sum_{i=1}^{m} A_i^2 \cdot \sum_{i=1}^{m} B_i^2}} - \frac{\sqrt{\sum_{i=1}^{m} (A_i - B_i)^2}}{\max(\sqrt{\sum_{i=1}^{m} (A_i - B_i)^2})} \tag{1}$$

where A_is and B_is denote gain values of two gain sets across m frequency bands. This similarity is defined in such a way that its value is normalized between -1 and 1. For the example illustrated in Table 1, the similarity between option 1 gain set [4, 6, 5, 2, 4] of a paired comparison and the true gain set is greater than the similarity between the option 2 gain set [4, 2, 7, 3, 4] of the paired comparison and the true gain set, thus the simulated preference for this pair is gain set 1.

Next, two tables are set up to keep track of two counts. Table 2, named preference counting, is used to keep track of a count of the frequency band locations associated with the preferred hearing as determined by the similarity of a gain set with the optimum or true gain set. Table 3, named occurrence counting, is used to keep track of a count of the number of frequency band locations that the two sets differ.

TABLE 1. Example showing similarity of two hearing preference functions or gain sets with true gain set.

True Personalized Gain Set [3, 4, 5, 2, 3]				
Gain Set 1	Gain Set 2	Similarity (True, 1)	Similarity (True, 2)	Simulated Preference
[4, 6, 5, 2, 4]	[4, 2, 7, 3, 4]	0.828	0.736	Gain Set 1

TABLE 2. Example of preference counting table.

Level Index	Band1	Band2	Band3	Band4	Band5
1	0	0	0	0 **+1**	0
2	0	0	0	0	0
3	0	0 **+1**	0	0	0
4	0	0	0	0	0
5	0	0	0 **+1**	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0

TABLE 3. Example of occurrence counting table.

Level Index	Band1	Band2	Band3	Band4	Band5
1	0	0	0	0 **+1**	0
2	0	0	0	0	0
3	0	0 **+1**	0	0**+1**	0
4	0	0	0 **+1**	0	0
5	0	0**+1**	0 **+1**	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0

TABLE 4. Example of simulated hearing preference function.

Level Index	Band1	Band2	Band3	Band4	Band5
1	0.466	0.427	0.305	0.717	0.466
2	0.534	0.521	0.429	**0.722**	0.534
3	**0.591**	0.574	0.521	0.694	**0.591**
4	0.576	**0.588**	0.574	0.628	0.576
5	0.574	0.568	**0.591**	0.521	0.574
6	0.521	0.519	0.576	0.380	0.521
7	0.429	0.445	0.534	0.230	0.429
8	0.305	0.354	0.466	0.103	0.305

The above counting is repeated for all possible 8^5 gain sets corresponding to the search space. Another table is then set up based on the ratio of the two counts to represent the overall hearing preference gain set for all possible gain sets. An example of this table or the simulation outcome is shown in Table 4. As shown in this table, it can be seen that the frequency band locations of the overall hearing preference gain set matches the frequency band locations of the true gain set. This simulation was repeated for a large number of different true hearing preference gain sets and each time the same location matches were obtained, indicating that the frequency bands can be treated independently towards reaching the optimum or personalized gain set. It is worth mentioning that these simulations were repeated by using different adjustment levels and again the same location matches were obtained.

III. PERSONALIZATION BY BAYESIAN MACHINE LEARNING

By considering the independence of frequency bands discussed above, the search space for finding the optimum gain set is drastically reduced for the paired comparisons needed just in a single band. For example, a total of only $8^*7/2 = 28$ paired comparisons would be needed for 8 levels.

Let $f(x)$ be a function indicating a user's hearing preference across gain adjustment levels in a frequency band which is to be determined by paired comparisons. Furthermore, let us consider that this function belongs to a Gaussian process $f(x) \sim \mathcal{GP}\left(0, k(x, x')\right)$ with the following kernel:

$$k\left(x, x'\right) = \exp\left(-\frac{1}{2\lambda}\left(x - x'\right)^2\right) \tag{2}$$

This kernel captures the distance between two gain settings where the parameter λ controls the preference function smoothness. Let the distribution of f given $\mathcal{X}$ and λ be normal with zero mean as follows:

$$p\left(\mathbf{f} \mid X, \lambda\right) = \mathcal{N}\left(\mathbf{f} \mid \mathbf{0}, \mathbf{K}\right) \tag{3}$$

where $\mathbf{f} = [f\left(x^1\right), \ldots, f\left(x^n\right)]$ represents a vector containing preference function values corresponding to n adjustment levels, and the covariance $[\mathbf{K}]_{i,j} = k\left(x^i, x^j\right)$. Then, the optimal or personalized adjustment level is given by

$$x^* = arg\max_{\boldsymbol{x}} f(x) \tag{4}$$

Noting that a user's feedback is provided by paired comparisons, the so-called Probit likelihood for binary observation can be used to encode the user's feedback for the two options $x^a \in \mathcal{X}$ and $x^b \in \mathcal{X}$ in a paired comparison as described in [18]. The feedback is assigned -1 and $+1$ depending on which of the two gain sets is preferred by the user during paired comparisons $\mathcal{D} = \{d_k \in (-1, 1) \mid k = 1, \ldots, K\}$, where K denotes the number of paired comparisons. Now, according to [17], the probability that x^a is preferred over x^b can be written as follows:

$$p\left(d_k \mid \mathbf{f}_k, \sigma\right) = \Phi\left(d_k \frac{f\left(x_k^a\right) - f\left(x_k^b\right)}{\sqrt{2}\sigma}\right) \tag{5}$$

where Φ denotes the normal cumulative density function and the parameter σ denotes the degree of uncertainty or variations associated with a user's feedback. Hence, given the current preference function $\mathbf{f}$, the likelihood of the feedback can be written as

$$p\left(\mathcal{D} \mid \mathbf{f}, \sigma\right) = \prod_{k=1}^{m} p\left(d_k \mid \mathbf{f}_k, \sigma\right) \tag{6}$$

Next the personalization problem is formulated within the Bayesian machine learning framework. That is the posteriori probability of interest is obtained based on the priori probability and the feedback likelihood as follows:

$$p\left(\mathbf{f} \mid \mathcal{D}, \mathcal{X}, \lambda, \sigma\right) = \frac{p\left(\mathcal{D} \mid \mathbf{f}, \sigma\right) p\left(\mathbf{f} \mid \mathcal{X}, \lambda\right)}{p\left(\mathcal{D} \mid \mathcal{X}, \lambda, \sigma\right)} \tag{7}$$

By using the Laplace approximation discussed in [19], the posteriori probability can be approximated as a normal density, that is

$$p\left(\mathbf{f} \mid \mathcal{D}, \mathcal{X}, \lambda, \sigma\right) \approx \mathcal{N}\left(\mathbf{f} \mid \hat{\mathbf{f}}, \left(\mathbf{K}^{-1} + \mathbf{W}\right)^{-1}\right) \tag{8}$$

where $\mathbf{W}$ denotes the Hessian matrix. Then, an estimate of the function $\hat{\mathbf{f}}$ can be found in an iterative manner by using the Newton method via the following iterative equation shown in [19]:

$$\mathbf{f}^{new} = \left(\mathbf{K}^{-1} + \mathbf{W}\right)^{-1} \left[\mathbf{W}\mathbf{f} + \nabla logp\left(\mathcal{D} \mid \mathbf{f}, \sigma\right)\right] \tag{9}$$

As part of the above estimation, the parameters $\{\lambda, \sigma\}$ need to be determined. These parameters can be obtained via the maximum-a-posterior estimation method named L-BFGS-B described in [20] and [21], that is

$$\{\lambda, \sigma\} \approx \arg\max_{\lambda,\sigma} p\left(\lambda, \sigma \mid \mathcal{D}, \mathcal{X}\right) \tag{10}$$

Algorithm 1 illustrates the training of the personalization algorithm indicated above. The heatmap shown in Figure 2 represents an example of the preference functions in each frequency band. The darkest block in each frequency band is the optimum personalized level for that band. The line in Figure 2 indicates the personalized gain set or curve which is obtained by connecting the optimum personalized levels in each band.

A simulation study was conducted to see whether the learning approach used generated an estimated preference function close to the true preference function. A vector of length 8 with random numbers between 0 and 1 was generated to represent the true preference function for 8 adjustment levels in a band. Initializing with a flat preference function, the outcome of the simulation is shown in Figure 3. As can be seen from

Algorithm 1 Personalization Training

Initialize with the standard prescriptive gain set
Initialize preference function value $\mathbf{f}^0$ for each band
For iterations = 1: $S(S\text{-}1)/2$
  Present two amplifications based on two gain sets
  Collect corresponding user hearing preference
  For given parameters $\{\lambda, \sigma\}$, find an estimate $\hat{\mathbf{f}}$ using equation (9)
  Obtain updated parameters $\{\lambda, \sigma\}$ using equation (10)
Output: Personalized preference function value $\mathbf{f}^*$ for each band

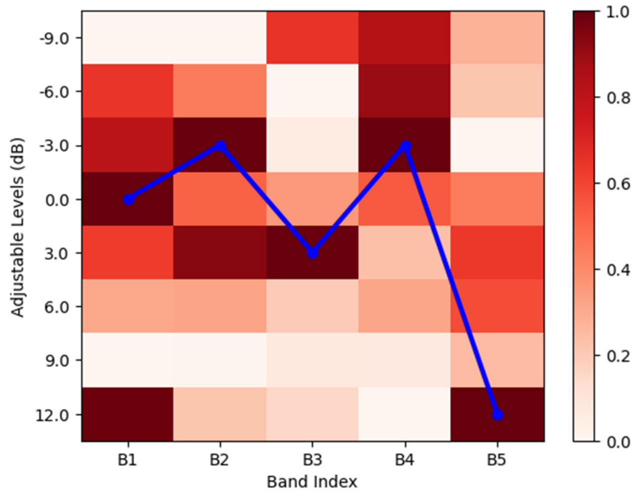

FIGURE 2. Personalized gain set or curve obtained by connecting the highest hearing preference function values (the darkest blocks) in each frequency band.

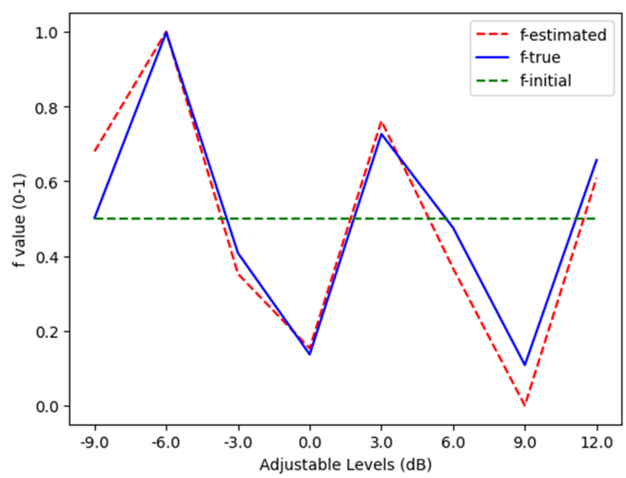

FIGURE 3. Example of estimated hearing preference function in a frequency band.

this figure, the estimated hearing preference function values $\hat{\mathbf{f}}$ came close to the true hearing preference function values.

If desired, it is possible to increase the accuracy of estimation by going through more episodes of paired comparisons. Based on a new episode of paired comparisons, the above Bayesian learning can be updated by beginning with the preference function values at the last episode. The parameters can also get computed based on the L-BFGS-B gradient for the new episode.

IV. CLINICAL EXPERIMENTS SETUP

In addition to the simulations done above indicating the usability of the developed multi-band Bayesian machine learning approach, a clinical study has been conducted in this paper to show its applicability in practice. Eight subjects with mild to moderate hearing loss were recruited for this clinical study under an approved human subject Institutional Review Board (IRB) protocol at the University of Texas at Dallas. The eligibility for the participation included: (i) symmetric mild to moderate hearing loss, (ii) being able to speak and understand English, and (iii) being an adult in the age range of 21-80 years old capable of providing informed consent.

The clinical sessions were divided into the following three sessions: audiogram measurement session, training session, and testing session. The audiograms of the participants were obtained during the first session by a hearing healthcare professional. Based on the audiogram, a DSLv5 prescriptive gain set was generated from a web-based tool [22]. This DSLv5 gain set served as the baseline or the starting point of our personalization. The range of the personalization for each band was set to $(Gain_{prescription} - 9\ dB, Gain_{prescription} + 12dB)$ for the adjustable levels around the prescriptive gain set.

During the second session, the participants underwent a training procedure aimed at determining their individualized hearing preferences using the developed machine learning personalization. The participants wore a pair of commercial hearing aids in both ears linked via Bluetooth to a dedicated laptop placed in the sound booth in which the participants were sitting. Figure 4 provides an illustration of the experimental setup. The hearing aids were configured to deliver flat amplification through their internal processors with noise reduction and sound enhancement features disabled. The environmental microphones were also turned off so all audio was from the Bluetooth connection to the laptop. This setup ensured that any variations in hearing experience were solely attributable to different gain sets derived from the personalization algorithm. For guidance and communication, the experimenter remained stationed outside the sound booth and visible to the participant through a window. The experimenter ran the personalization algorithm running on a laptop (laptop 1 in Figure 4) which was connected to another laptop in the sound booth (laptop 2 in Figure 4) via an online meeting utility. Laptop 1 with the personalization algorithm controlled the amplification of audio signals through gain sets, sending audio signals from laptop 1 to laptop 2 via the online meeting utility, and then to the hearing aids via Bluetooth. With this

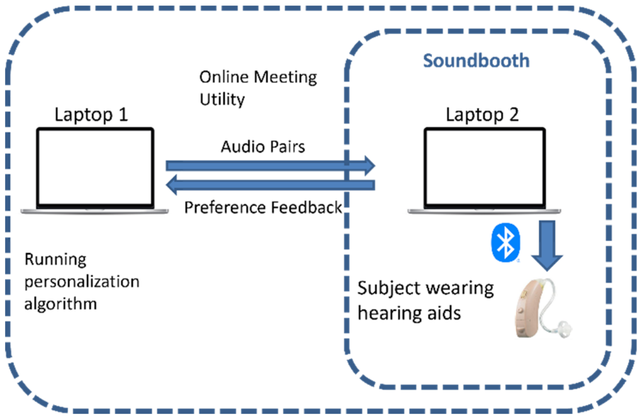

FIGURE 4. Experimental setup of clinical testing.

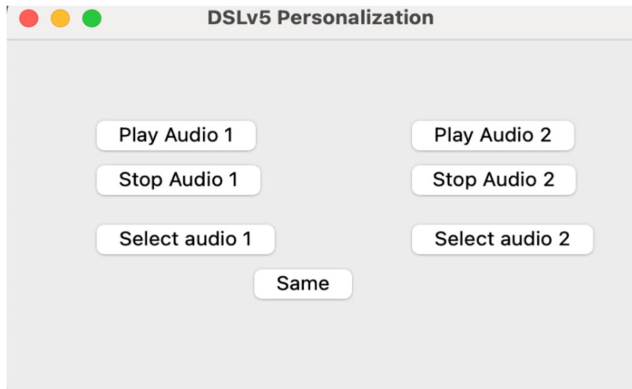

FIGURE 5. Graphical-user-interface (GUI) of paired comparisons for obtaining the hearing preference function of a hearing aid user.

setup, the experimenter was the one who operated the personalization program and the participants focused solely on stating their hearing preferences.

The graphical-user-interface (GUI) of the personalization training program is shown in Figure 5. For each paired comparison, the participants were presented with a pair of audio signals generated by applying two different gain sets on the same audio signal of a sentence lasting approximately 2.5 seconds. Then, the participants were asked to indicate their hearing preference by selecting "Audio 1", "Audio 2", or "Same" in the interface. The audio signals corresponded to the TSP speech database which is widely used in speech processing [23]. This dataset consists of recordings of 1400 utterances spoken by 24 speakers. The recordings of this database are phonetically balanced and designed to contain a diverse range of speech sounds. Babble background noise was added to the spoken utterances with a moderate Signal-to-Noise Ratio (SNR) of 5 dB. During the training session, paired comparisons were conducted for a pair of gain sets. The gain values in the five bands were passed through a cosine curve interpolator to generate a smooth frequency response curve. Subsequently, a 64-tap FIR filter, obtained

TABLE 5. Standard Dslv5 prescriptive and personalized gains.

Subject (gender, age)	Audiogram (dB)	Standard Gains (dB)	Personalized Gains (dB)
1 (M, 65)	10, 15, 25, 45, 45	4, 2, 12, 30, 28	10, 5, 21, 21, 22
2 (F, 66)	10, 25, 20, 30, 35	4, 9, 8, 22, 21	7, 18, 8, 28, 12
3 (M, 52)	0, 5, 10, 30, 30	-3, -5, 1, 22, 19	-3, 7, 1, 22, 25
4 (M, 69)	20, 20, 25, 25, 45	7, 6, 12, 21, 28	16, 18, 18, 21, 22
5 (F, 75)	5, 0, 5, 30, 40	4, -4, -2, 22, 24	10, 5, 10, 25, 33
6 (M, 64)	10, 15, 30, 40, 55	4, 2, 14, 27, 35	10, -7, 20, 21, 38
7 (F, 65)	10, 10, 20, 15, 30	16, 0, 8, 13, 19	19, 12, -1, 22, 31
8 (M, 33)	0, 30, 60, 40, 30	0, 11, 30, 27, 19	-9, 14, 21, 30, 25

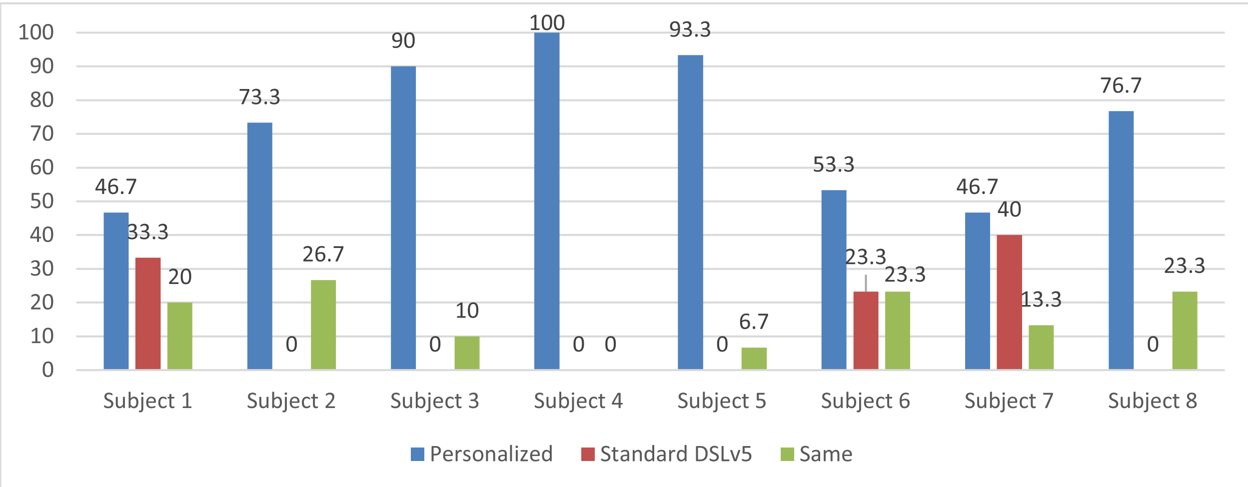

FIGURE 6. Comparison of the personalized versus standard amplification settings; the numbers indicate the percentage of times the amplification setting was preferred.

via a filter design module, was employed to generate the frequency response curve.

The training was carried out based on 28 paired comparisons. The preference function derived from the personalization algorithm was then used to establish the personalized gain setting. It is important to mention here that the entire training session was completed in less than 10 minutes, indicating the efficiency of the developed machine learning personalization algorithm.

After the personalized gain values were obtained from the training session, a test session was conducted that aimed to select between the standard DSLv5 gain setting and the personalized gain setting. This session consisted of presenting thirty audio pairs randomly chosen from the TSP dataset with added babble noise. Audio pairs were processed through the standard DSLv5 gain set and the personalized gain set, with the presentation order randomized to enable an unbiased selection. The results of the test session are presented and discussed in the next section.

V. SUBJECT TESTING RESULTS AND DISCUSSION

This section presents the outcomes of our clinical testing. Table 5 illustrates the audiograms of the participating subjects as well as the standard DSLv5 and personalized gain values for the five frequency bands and eight adjustment levels considered. As shown in Table 5, the personalized gain values differed from the standard DSLv5 prescription gain values. An important consideration is that DSLv5 targets represent the mean amplification values for a given hearing loss. In other words, there is variance around these values across frequency bands and it is not only possible but likely that an individual's optimal amplification can lie below or above these average targets. Indeed, this is supported by the literature, whereby studies have found that nearly half of hearing aid users prefer alternate settings than those prescribed from average values such as those from DSL. In clinical practice, finding these optimal gain values would be incredibly time consuming and there are no standard methods to derive them across frequency bands. The results in Table 5 indicate that our machine learning approach addresses this

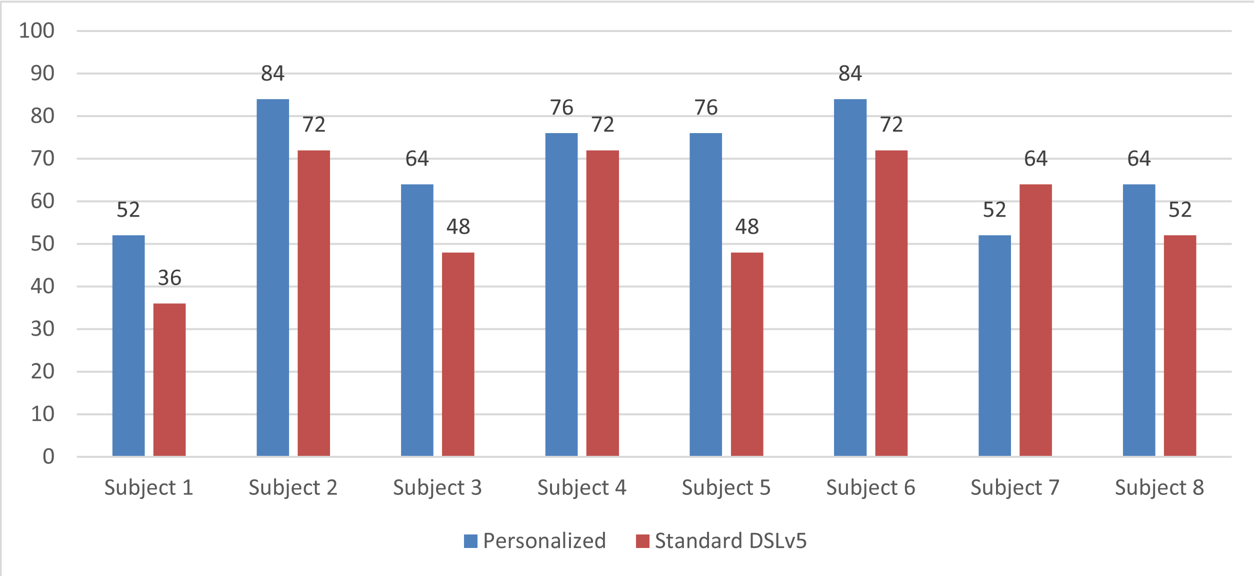

FIGURE 7. **Word recognition score percentages of the personalized and standard settings in babble background noise at 5dB SNR.**

task in a systematic way that makes clinical and field deployment possible. In this study, all of the participants exhibited greater hearing loss in high-frequency bands relative to low-frequency bands. The standard DSLv5 prescription gain values compensate for this loss by providing more gain values in the high-frequency bands. However, as seen from Table 5, the personalized gain values varied across the participants with some desiring more gains whereas others preferred less gains in the high frequencies.

The results of the hearing preference testing, represented as percentages, are shown in Figure 6. As evident from this figure, all the participants indicated that the personalized setting was preferred over the standard setting, albeit with varying degrees depending on their level of hearing loss. By dividing the sum of the percentage times that the personalized setting was preferred across the eight participants by the sum of the percentage times that the standard setting was preferred, it is seen that the personalized settings were favored on average by six times over the standard settings.

An additional experiment was conducted to assess whether the personalized settings had any adverse impact on word recognition or speech understanding. A 50-word list from the Northwestern University Auditory Test No. 6 (NU-6) dataset [24] was considered for this experiment. The list was played with babble background noise at 5dB SNR. Half of the words in the list were played using the standard DSLv5 setting, and the remaining half with the personalized setting. The order of presenting the words was randomized. The participants were instructed to repeat each word after it was presented to them once. Correct word repetition contributed to an increase in the word recognition score. The word recognition scores from this experiment are provided in Figure 7. As illustrated in this figure, the personalized setting exhibited no adverse impact on word recognition or speech understanding compared to the standard DSLv5 setting.

The work presented in this paper shows an efficient and effective approach to addressing the hearing healthcare needs of hard of hearing individuals by leveraging machine learning to personalize amplification. Our original hypothesis was that machine learning could generate an individualized prescription without degrading word recognition scores. By personalizing the hearing experience, users would likely be more satisfied with their hearing aids and would be more likely to wear them. What our work demonstrates is that not only can machine learning derive and optimize individual preferences but that these preferences can also increase word recognition scores in competing background noise. The improvement in hearing in competing background noise addresses the top complaint among hearing aid users, difficulty in background noise. The approach presented in this paper addresses a major shortcoming of the current standard of care for hearing aid programming that remains tethered to average amplification targets and provides an efficient method that could substantially improve the current practice of hearing aid fitting.

VI. CONCLUSION

A training-efficient machine learning-based personalization approach has been introduced in this paper. This personalization approach involves the use of Bayesian learning to model a hearing preference function in a frequency band independent of other frequency bands. The independence of the frequency bands has led to an efficient and thus practical training for reaching personalized gain values by conducting a small number of paired comparisons. The clinical experiments carried out on eight participants with hearing loss have shown that the personalized gain values are preferred over the standard DSLv5 prescriptive gain values on average by six times. In our future work, it is intended to turn this machine learning-based personalization approach into a smartphone app to enable its utilization in the field or in real-world audio environments.

ACKNOWLEDGMENT

The authors wish to thank Dr. Lindee Alvarez for her help in conducting the clinical experiments.

REFERENCES

[1] S. Scollie, R. Seewald, L. Cornelisse, S. Moodie, M. Bagatto, D. Laurnagaray, S. Beaulac, and J. Pumford, "The desired sensation level multistage input/output algorithm," *Trends Amplification*, vol. 9, no. 4, pp. 159–197, Jan. 2005.

[2] G. Keidser, H. Dillon, M. Flax, T. Ching, and S. Brewer, "The NAL-NL2 prescription procedure," *Audiol. Res.*, vol. 1, no. 1, p. e24, Mar. 2011.

[3] C. Portnuff and B. Bell, "Effective use of speech-in-noise testing in the clinic," *Hearing J.*, vol. 72, no. 5, p. 40, May 2019.

[4] L. Creed and M. Merkison. (Aug. 2022). *OTC Hearing Aid Final Rule is Here: What to Know, How to Prepare*. Leader Live. Accessed: May 18, 2024. [Online]. Available: https://leader.pubs.asha.org/do/10.1044/2022-0823-otc-final-rule/full/

[5] D. W. Swanepoel, I. Oosthuizen, M. A. Graham, and V. Manchaiah, "Comparing hearing aid outcomes in adults using over-the-counter and hearing care professional service delivery models," *Amer. J. Audiology*, vol. 32, no. 2, pp. 314–322, Jun. 2023.

[6] E. Convery, G. Keidser, M. Seeto, I. Yeend, and K. Freeston, "Factors affecting reliability and validity of self-directed automatic in situ audiometry: Implications for self-fitting hearing aids," *J. Amer. Acad. Audiol.*, vol. 26, no. 1, pp. 5–18, Jan. 2015.

[7] A. Boothroyd and C. Mackersie, "A 'Goldilocks' approach to hearing-aid self-fitting: User interactions," *Amer. J. Audiology*, vol. 26, no. 3, pp. 430–435, Oct. 2017.

[8] P. B. Nelson, T. T. Perry, M. Gregan, and D. VanTasell, "Self-adjusted amplification parameters produce large between-subject variability and preserve speech intelligibility," *Trends Hearing*, vol. 22, Jan. 2018, Art. no. 233121651879826.

[9] D. Vyas, R. Brummet, Y. Anwar, J. Jensen, E. Jorgensen, Y.-H. Wu, and O. Chipara, "Personalizing over-the-counter hearing aids using pairwise comparisons," *Smart Health*, vol. 23, Mar. 2022, Art. no. 100231.

[10] F. K. Kuk. *Paired Comparisons As a Fine-Tuning Tool in Hearing Aid Fittings*. Accessed: May 18, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Paired-Comparisons-as-a-Fine-Tuning-Tool-in-Hearing-Francis-K./6815f52a8d0ab5fdc5a5ccbe5a7ec23b63f09c4a

[11] A. M. Amlani and E. C. Schafer, "Application of paired-comparison methods to hearing aids," *Trends Amplification*, vol. 13, no. 4, pp. 241–259, Dec. 2009.

[12] M. Dahlquist, J. Larsson, S. Hertzman, F. Wolters, and K. Smeds, "Predicting individual hearing-aid preference in the field using laboratory paired comparisons," in *Proc. Int. Symp. Auditory Audiological Res.*, vol. 5, Dec. 2015, pp. 261–268.

[13] N. Alamdari, E. Lobarinas, and N. Kehtarnavaz, "Personalization of hearing aid compression by human-in-the-loop deep reinforcement learning," *IEEE Access*, vol. 8, pp. 203503–203515, 2020.

[14] S. Akbarzadeh, E. Lobarinas, and N. Kehtarnavaz, "Online personalization of compression in hearing aids via maximum likelihood inverse reinforcement learning," *IEEE Access*, vol. 10, pp. 58537–58546, 2022.

[15] A. Ni, E. Lobarinas, and N. Kehtarnavaz, "Personalization of hearing AID DSLV5 prescription amplification in the field via a real-time smartphone APP," in *Proc. 24th Int. Conf. Digit. Signal Process. (DSP)*, Jun. 2023, pp. 1–5.

[16] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[17] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," 2010, *arXiv:1012.2599*.

[18] L. L. Thurstone, "A law of comparative judgement," *Psychol. Rev.*, vol. 34, pp. 273–286, Jan. 1927.

[19] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[20] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, Sep. 1995.

[21] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, Dec. 1997.

[22] *Web-Based Hearing Aid Compression Fitting Using DSL-v5 By Hand*. Accessed: May 20, 2024. [Online]. Available: https://personal.utdallas.edu/ kehtar/WebBasedFitting.html

[23] P. Kabal. (2002). *TSP Speech Database*. [Online]. Available: http://www.mmsp.ece.mcgill.ca/Documents/Downloads/TSPspeech/TSPspeech.pdf

[24] T. Tillman and R. Carhart, "An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory test no. 6 SAM-TR-66-55," USAF School Aerosp. Med., Brooks Air Force Base, TX, USA, Tech. Rep. SAM-TR, pp. 1–12, Jun. 1966.

AOXIN NI (Student Member, IEEE) is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, The University of Texas at Dallas. His research interests include auditory neuroscience, signal/speech processing, and machine learning, in particular as related to hearing personalization. He has also developed real-time smartphone apps for personalization of the amplification function of hearing aids.

EDWARD LOBARINAS is currently a Professor with the Callier Center for Communication Disorders, School of Behavioral and Brain Sciences, The University of Texas at Dallas. He has been recognized by American Academy of Audiology as a Jerger Future Leader of audiology and has received grants from the National Institute of Health, American Tinnitus Association, and Tinnitus Research Initiative. His research interests include the role of hearing loss in the development of tinnitus and how inner ear damage affects higher auditory function, such as hearing in noise.

NASSER KEHTARNAVAZ (Fellow, IEEE) is currently an Erik Jonsson Distinguished Professor with the Department of Electrical and Computer Engineering and the Director of the Embedded Machine Learning Laboratory, The University of Texas at Dallas, Richardson, TX, USA, and a licensed Professional Engineer. He has authored or coauthored 11 books and more than 400 publications in these areas. His research interests include signal and image processing, machine learning, deep learning, and real-time implementation on embedded processors. He is a fellow of SPIE and AAIA and the Editor-in-Chief of *Journal of Real-Time Image Processing*.