IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 26, 2022, accepted May 17, 2022, date of publication May 27, 2022, date of current version June 7, 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3178594*

# Online Personalization of Compression in Hearing Aids via Maximum Likelihood Inverse Reinforcement Learning

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

Corresponding author: Sara Akbarzadeh (sara.akbarzadeh@utdallas.edu)

This work involved human subjects or animals 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 under IRB No. 20-13.

**ABSTRACT** A key function of modern hearing aids is compression or mapping of sound to the residual hearing range of those suffering from hearing loss. This paper presents a machine learning approach to personalize compression in hearing aids in an online manner. The online feature of this approach allows it to be deployed in the field. The significance of this personalized compression lies in enabling preferred hearing outcomes relative to the one-size-fits-all prescriptive compression rationales that are currently being used. This personalization approach utilizes maximum likelihood inverse reinforcement learning to establish a model of a hearing aid user's preference based on paired comparisons by the user. The results of the preference paired comparisons between the personalized and standard prescriptive settings from ten subjects indicated that personalized settings were preferred about 10 times more than the standard prescriptive settings. In addition, a word recognition comparison was conducted showing that the personalized settings had no adverse impact on speech understanding in either quiet or in competing noise conditions.

**INDEX TERMS** Personalization of compression in hearing aids, hearing aid fitting, maximum likelihood inverse reinforcement learning.

## I. INTRODUCTION

People with normal hearing can hear a wide range of sounds in terms of sound pressure levels (SPLs) from softest sounds that are barely audible to loudest sounds that are tolerated without pain. This range of sound pressure levels is known as the dynamic range of hearing (~120 decibels or dB) [1]. In people suffering from hearing loss, this dynamic range is reduced. To compensate for the reduced dynamic range, modern hearing aids fit a broad range of sounds into the residual dynamic range of a hearing impaired person. This process is referred to as compression.

Compression algorithms amplify soft sounds that are inaudible but provide less amplification when sound levels exceed a specified threshold called compression threshold (CT) [2]. In this manner, sounds that are below a hearing impaired person's thresholds become audible but more intense sounds receive less or no amplification to avoid loudness discomfort. In essence, compression consists of a gain-frequency response curve at soft, moderate, and loud SPLs. These curves are typically set by hearing healthcare professionals using manufacturer's software during the hearing aid fitting process. After acquiring a person's lowest hearing levels across a number of frequency bands, known as audiogram, a hearing aid is programmed or fitted to provide gains according to a prescription. The two most widely used prescriptions are DSL-v5 [3] and NAL-NL2 [4], which are derived from averages of optimum gain for speech stimuli from a group of people with similar hearing loss.

In real-world audio environments, however, amplification needs and hearing preferences of individuals with similar audiograms can vary considerably from one person to another [5], [6]. In other words, a prescription derived from the averages of a group may not be optimal for everyone. In [7],

The associate editor coordinating the review of this manuscript and approving it for publication was Yue Zhang.

it was shown that the preferred gains of more than half of the subjects who completed the study fell above or below a 6-dB window surrounding their NAL-NL1 prescriptive gains. In general, people's hearing preferences differ in real-world audio environments encountered in their daily lives [8]. The hearing aid fitting process normally takes place in a quiet office in the absence of real-world audio environments of particular interest to a hearing aid user. A solution to deal with real-world audio environments is to personalize the gains for a hearing aid user based on the user's hearing preference.

Previous studies have shown that personalization of hearing aid gains can provide preferred hearing relative to standard prescriptive gains [9]–[16]. In [9], an algorithm was developed for users to manually adjust the amplification for several input sound levels and listening situations. The manual adjustments done in [9] to find the optimal settings is a non-systematic approach which is time consuming to conduct. In addition, it is necessary for users to understand, at least to some degree, how the algorithm works. In many cases, this is not so easy for users to understand the algorithm whereas the task of picking the preferred audio in an audio pair is quite simple and straightforward for users to do. More recently, machine learning methods have been used to achieve personalization settings in a systematic way. For example, a modified simplex approach was used to find the optimal hearing aid parameters in [17] and [18]. In a study conducted by Neuman *et al.* [19], the modified simplex approach was found to be more efficient than the two methods of iterative round robin and double elimination. It is to be noted that the modified simplex method does not work effectively when the error function possesses multiple peaks. A genetic algorithm approach was considered in [20] and [21] to tune the feedback cancellation feature of hearing aids based on users' feedback. In [16] and [22], a non-parametric Gaussian model was trained by carrying out pairwise comparisons of music clips. The personalization reported in these works did not include examination of more challenging stimuli such as speech. Furthermore, a limited number of comparisons done in these works is often not adequate to model human preference in challenging listening environments such as speech in the presence of background babble noise.

In a previous study by our research team [23], a human-in-the-loop (HITL) deep reinforcement learning (DRL) framework was developed to personalize compression in hearing aids. Pairs of noisy speech stimuli compressed by different compression settings were presented to a hearing aid user to pick the preferred stimulus in a pair. A deep neural network (DNN) was trained based on the user's feedbacks to establish the optimal compression settings individualized or personalized for that user. It was shown that the subjects who participated in the study preferred the personalized settings seven times more than that of the standard prescriptive settings. In another study [24] by our research team, it was shown that this human-in-the-loop personalization did not have any negative impact on word recognition, and in fact generally produced higher word recognition scores compared to standard prescriptive settings.

Whereas the HITL-DRL personalization provided promising results, the offline training of the DNN limits its deployment in the field because its training is carried out in an offline manner and not in an on-the-fly or online manner. A user's preference varies depending on the audio environment encountered. For example, for a given hearing loss, preferred compression settings for understanding speech in the presence of babble background noise would be different from preferred compression settings when listening to music. The online training capability is essential as far as the deployment of any personalization algorithm in the field is concerned. The main contribution of this work is thus in the development of a machine learning personalization approach for hearing aid compression via Maximum Likelihood Inverse Reinforcement Learning (MLIRL) which can be trained in the field in an online manner. This approach enables optimal personalized settings to be determined in on-the-fly manner in real-world audio environments. The existing personalization approaches in the literature are conducted in an offline training manner. In other words, the novelty of our approach lies in its ability to get trained in an online manner. MLIRL was initially introduced in [25] and since then has been implemented in several applications, e.g. [26]–[28].

The rest of this paper is organized as follows. In section II, the developed personalization approach based on MLIRL is described in detail. In section III, the experimental setup is stated. The preference and word recognition results for ten subjects are then reported in section IV. The paper is finally concluded in section V.

## II. ONLINE PERSONALIZATION OF COMPRESSION

In the hearing aid fitting process, the gains across a number of frequency bands are adjusted as specified by a prescription based on a user's audiogram. For personalizing hearing aid compression, the prescriptive gains are considered to be the initial condition and are adjusted to the user's preference. Figure 1 shows a depiction of the range of personalized gains around prescriptive gains across different frequency bands. This range spans the boundaries of hearing threshold and loudness discomfort.

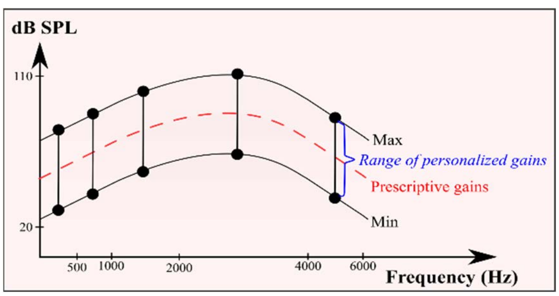


**FIGURE 1.** Depiction of the range of personalized gains across prescriptive gains as the initial condition.

The gains are related to so-called compression ratios (CRs) in each of the frequency bands. Figure 2 illustrates a typical

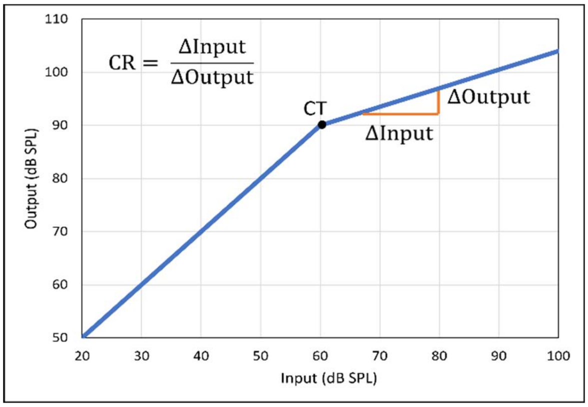


**FIGURE 2.** Typical compression curve for one frequency band.

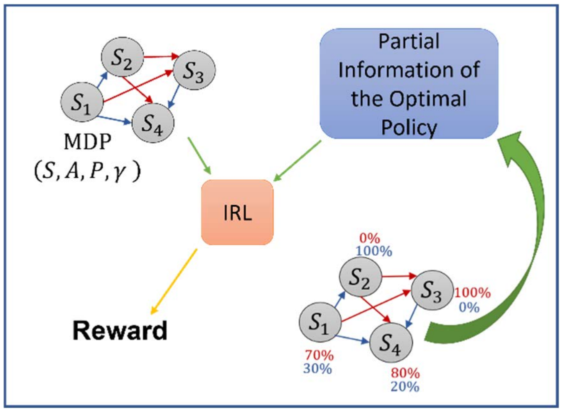


**FIGURE 3.** Inverse reinforcement learning framework.

compression curve in one frequency band. By individualizing or personalizing CRs in each band, a personalized compression can be reached.

First, the prescriptive gains are translated into prescriptive CRs and are used as the initial or starting condition. These CRs are then trained or adjusted via MLIRL. To adjust CRs to their new values from their prescriptive settings, scales are used as follows:

$$CR_{new}(i) = CR_{prescription}(i) \times scale(i) \tag{1}$$

where $CR_{prescription}(i)$ denotes prescriptive CR in the $i^{th}$ frequency band and $scale(i)$ denotes the adjustment scale in that band. The personalization of compression is aimed at finding the most preferred combination of CRs across all the frequency bands for a specific user.

Paired comparisons of audio signals are considered to provide an easy and user-friendly mechanism for users to provide their feedbacks in order to find the best combination of scaled CRs. A paired comparison means presenting an audio signal pair compressed by different sets of CRs and asking users to indicate the audio signal they prefer in the pair. Carrying out paired comparisons for all possible CRs is very time consuming and thus not practical due to users getting fatigue and tired. In this work, we have used the method of MLIRL in order to find the most preferred CRs by carrying out paired comparisons in a systematic manner within a non-fatigue time duration. In the subsections that follow, it is explained how MLIRL is utilized for personalization of compression. First, we state the general concept of Inverse Reinforcement Learning (IRL). Then, our personalization approach via MLIRL is described.

### A. PROBLEM FORMULATION

Reinforcement Learning (RL) is a machine learning method that enables an agent to interact with an environment to perform an action based on a reward received from the environment. This paradigm involves a Markov Decision Process (MDP) consisting of a 5-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \gamma, \mathcal{R})$, where

$\mathcal{S}$ is a finite set of states in the environment,

$\mathcal{A}$ is a finite set of available actions,

$\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ indicates a state transition probability;

$\mathcal{P}(s, a, s') = \mathbb{P}(s_{t+1} = s' \mid s_t = s, a_t = a)$ with $s'$ denoting next state resulted from performing action $a$ in state $s$,

$\gamma \in [0, 1]$ is a discount factor,

$\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is a reward function.

A reward is a value assigned to performing action $a$ in state $s$. A limitation of RL is that it requires the reward to be pre-defined. This restricts its applicability to the problems in which the reward function can be readily defined [29]. To avoid obtaining the reward in an offline manner as done in [23] by training a DNN model or enable field deployment, when demonstrations or partial information of a desired behavior are available, IRL can be used. IRL is thus used in this work to broaden the applicability of RL to field deployment for this hearing application. Figure 3 shows the general framework of IRL. IRL involves a MDP without the reward and the goal of the agent is to find a reward function from demonstrations reflecting a policy or the human behavior.

In the problem at hand, the human behavior is denoted by a user's preference which is unknown to us. The user's feedback on paired comparisons can be viewed as partial information of the user's preference. The reward function is thus obtained from such paired comparisons. The state space $S$ consists of sounds compressed by all possible sets of CRs. Each set of CRs corresponds to a vector $[CR_{new}(1), CR_{new}(2), \ldots, CR_{new}(N)]$ of new CRs derived from (1) across $N$ number of frequency bands.

Each action corresponds to applying a set of scales in the frequency bands and the action space consists of all permutations of the scales across $N$ frequency bands. If $\beta_i$ denotes the number of scales at the $i^{th}$ frequency band in (1), then the number of actions in the action space is given by

$$n_{\mathcal{A}} = \prod_{i=1}^{N} \beta_i \tag{2}$$

For a sound compressed by a set of CRs in state $s$, applying action $a$ will result in a sound compressed by a new set of CRs depending on action $a$. Here, the actions are stochastic but $\mathcal{P}$ is deterministic since it is known what the next state $s'$ would be by applying action $a$ in state $s$. $\mathcal{R}$, which is the reward of applying action $a$ in state $s$, needs to be found.

Similar to other works using IRL [30]–[32], the reward here is considered to be a linear function of state-action features; $\mathcal{R}_W(s,a) = W^T\varphi(s,a)$ where $\varphi : \mathcal{S}\times\mathcal{A}\rightarrow\mathbb{R}^n$ is a known $n$-dimensional state-action function and $W$ is an unknown weighting vector of the state-action function. Next, it is discussed how MLIRL is used to find the reward function or the weighting vector $W$.

A demonstration consists of a series of trajectories $D = \{\tau^i\}_{i=1}^K$, where $K$ is the number of trajectories and each trajectory is a set of state-action pairs; $\tau^i = \{(s_1,a_1),(s_2,a_2),\ldots\}$. Demonstrations are generated according to a policy. In IRL, the policy is indicated to be the probability of choosing action $a$ in state $s$ and is defined by the Boltzmann distribution as follows:

$$\pi_W(s,a) = \mathbb{P}(a_t = a\,|\,s_t = s) = \frac{e^{\alpha Q_W(s,a)}}{\sum_{a'\in\mathcal{A}} e^{\alpha Q_W(s,a')}} \quad (3)$$

where $\alpha$ controls the randomness in the policy and $Q_W(s,a)$ denotes the optimal state-action value function. The optimal state-action value function can be written as:

$$Q_W(s,a) = \mathcal{R}_W(s,a) + \gamma\sum_{s'\in S}\mathcal{P}(s,a,s') \times\left[\sum_{b\in\mathcal{A}}\pi_W(s',b)\,Q_W(s',b)\right] \quad (4)$$

where $\gamma$ is the discount factor and $\mathcal{P}(s,a,s')$ is the state transition probability stated earlier. The likelihood of a demonstration $\mathcal{D}$ having the weighting vector $W$ can be written as

$$L(\mathcal{D}\,|\,W) = \prod_{(s,a)\in\mathcal{D}}[\pi_W(s,a)] \quad (5)$$

MLIRL obtains the weighting vector $W$ which maximizes the likelihood of the demonstration $\mathcal{D}$.

In the following subsection, it is explained how the feedback of a user is collected via paired comparisons. Then, it is described how this feedback is used to find the reward via MLIRL.

## B. USER'S FEEDBACK

The agent interacts with the environment for several trajectories. In each trajectory, the agent observes the environment and collects the user's feedback on several paired comparisons and updates the reward function based on the user's feedback. The first paired comparison in the first trajectory is initialized with $CR_{prescription}$. At each paired comparison or time step $t$, a noise added speech signal is randomly selected from a dataset and presented to the user. A state $s$ corresponds to the sound compressed by a set of CRs, $CR_t^{init}$, which represents the initial CRs at time step $t$. Then, an action $a$ which corresponds to a scale set is randomly selected from the action space. Performing action $a$ takes state $s$ to state $s'$, which means having the noise added speech signal compressed by an updated CRs denoted by $CR_t^{update}$. Therefore, the state transition $s\rightarrow s'$ corresponds to $CR_t^{init}\rightarrow CR_t^{update}$. The user's feedback on performed actions is obtained via a hearing preference interface. A pair of audio signals, one compressed by $CR_t^{init}$ and the other compressed by $CR_t^{update}$, are presented to the user as audio 1 and audio 2 asking the user to select the one he or she prefers. The user is provided with three options to select from: audio 1 is preferred, audio 2 is preferred, no preference between audios 1 and 2 or same preference. If the user selects the audio compressed with $CR_t^{update}$ as the preferred audio, action $a$ is considered to be a correct action, i.e., the feedback is considered to be positive and denoted by $f^+$. If the user selects the audio compressed with $CR_t^{init}$ as the preferred audio, action $a$ is considered to be a wrong action, i.e., the feedback is considered to be negative and denoted by $f^-$. When the user expresses the same preference for both audios, the feedback on action $a$ in state s is considered to be neutral.

Getting the user's feedback on all possible actions and states would be very time consuming and not practical when human subjects are involved. To avoid human fatigue and enable a manageable training time, the action space is sub-sampled while going through each trajectory. The agent defines the next state based on the feedback received from the user. Positive feedback means that $CR_t^{update}$ is more preferred than $CR_t^{init}$. Thus, the next step gets adjusted or initialized by $CR_t^{update}$, that is $CR_{t+1}^{init} = CR_t^{update}$, and $CR_t^{init}$ is not considered in future comparisons in the same trajectory. If the feedback is negative, $CR_t^{update}$ is not considered for future comparisons in the same trajectory and $CR_{t+1}^{init} = CR_t^{init}$. If the feedback is neutral, one of $CR_t^{init}$ and $CR_t^{update}$ is randomly selected as $CR_{t+1}^{init}$ and the other one is not considered for future comparisons in that trajectory. This procedure is repeated for a predefined number of paired comparisons in a trajectory. The outcome of each trajectory defines the initialization of a next trajectory. At the end of each trajectory, the reward function is updated using the MLIRL algorithm as described in the following subsection.

## C. MLIRL OPTIMIZATION USING USER'S FEEDBACK

A feedback model $h$ is defined based on feedbacks received from a user. Based on feedback $f_{a_j}$ for performing action $a_j$ in state $s$, similar to [27], the following feedback model is defined

$$h\left(s,a_i,f_{a_j}=f^+\right) = \begin{cases} 1-\varepsilon, & \text{if } a_i = a_j \\ \dfrac{\varepsilon}{|A|-1}, & \text{if } a_i \neq a_j \end{cases} \quad \forall a_i\in\mathcal{A},$$
$$(6)$$
$$h\left(s,a_i,f_{a_j}=f^-\right) = \begin{cases} -(1-\varepsilon), & \text{if } a_i = a_j \\ \dfrac{-\varepsilon}{|\mathcal{A}|-1}, & \text{if } a_i \neq a_j \end{cases} \quad \forall a_i\in\mathcal{A},$$

This model indicates whether the performed action was in favor or against the user's preference with a feedback error of $\varepsilon$. At the end of each trajectory, the following preference function or model $\mathcal{H}(s,a_i\,|\,f_s)$ is set up

$$\mathcal{H}(s,a_i\,|\,f_s) = \sum_{j=1}^{|f_s|} h\left(s,a_i,f_{a_j}\right), \quad \forall a_i\in\mathcal{A}. \quad (7)$$

where $f_s$ is the list of all feedbacks received in state $s$. This function reflects the feedback history and incorporates the user's preference for action $a_i$ in state $s$. Then an enhanced preference function $\mathcal{H}_E$ is set up by normalizing the function $\mathcal{H}$ to [0-1] as described in [27].

In the likelihood objective function in (5), the reward $W$ is found in such a way that:

- If the feedback for action $a$ in state $s$ is positive ($f_a = f^+$), then the policy $\pi_W(s, a)$ is maximized in the likelihood objective function.
- If the feedback for action $a$ in state $s$ is negative ($f_a = f^-$), then the policy $\pi_W(s, a)$ is minimized in the likelihood objective function.

As a result, the likelihood function of a demonstration $\mathcal{D}$ based on the reward function $W$ and feedback $f$ can be stated as (8), shown at the bottom of the page.

The goal here is to maximize the probability of correct actions and minimize the probability of wrong actions. Since the user's feedback may include errors and inconsistencies, instead of using the exact correct or wrong actions, the preference model $\mathcal{H}_E(s, a)$ is used and the likelihood function is modified as follows:

$$L(\mathcal{D} \mid W, \mathcal{H}_E) = \prod_{(s,a) \in \mathcal{D}} \left[\pi_W(s,a)^{\mathcal{H}_E(s,a)}\right] \tag{9}$$

Then, the optimization problem of interest involves finding the optimum reward function $W^*$ that maximizes the likelihood of the demonstration, i.e.

$$W^* = \mathrm{arg}max_W L(\mathcal{D} \mid W, \mathcal{H}_E) \tag{10}$$

The gradient ascent approach is used to obtain the optimum $W$. The gradient is given by

$$\begin{aligned}\frac{d}{dW}\log\left[L(\mathcal{D} \mid W, \mathcal{H}_E)\right] &= \sum_{(s,a) \in \mathcal{D}} \frac{\mathcal{H}_E(s,a)}{\pi_W(s,a)} \frac{d\pi_W(s,a)}{dW} \\ &= \sum_{(s,a) \in \mathcal{D}} \frac{\mathcal{H}_E(s,a)}{\pi_\theta(s,a)} \frac{1}{B_W(s)^2} \\ &\quad \cdot \left[B_W(s)\,\alpha e^{Q_W(s,a)} \frac{dQ_W(s,a)}{dW} - e^{Q_W(s,a)} \frac{dB_W(s)}{dW}\right]\end{aligned} \tag{11}$$

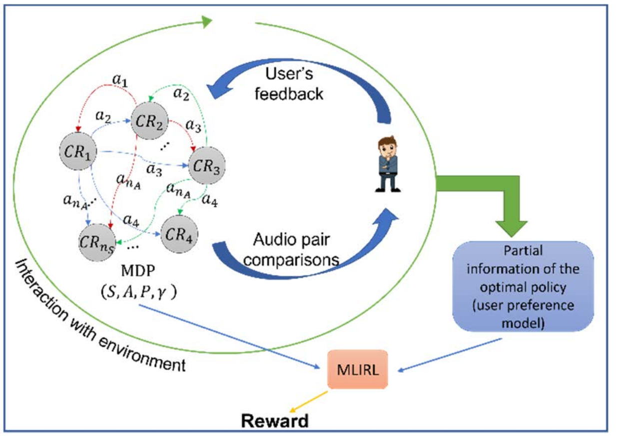


**FIGURE 4.** General framework of developed personalization approach.

where $B_W(s) = \sum'_a e^{\alpha Q_W(s,a')}$ and $Q_W(s,a)$ is defined by (4). The reward function gets updated at the end of each trajectory via gradient ascent.

Figure 4 shows the general framework of the developed online personalized compression approach defined in the context of IRL. The block diagram of this approach is presented in Figure 5 with more details and its algorithm is outlined in Table 1.

## III. EXPERIMENTAL SETUP

Ten subjects with mild to moderately severe hearing loss participated in our study under an approved human subject Institutional Review Board protocol (IRB 20-13) at the Callier Center for Communication Disorders, University of Texas at Dallas. The eligibility for participation included: (i) symmetric mild to moderately severe hearing loss, (ii) being able to speak and understand English, and (iii) being an adult in the age range of 21-80 years old who could provide informed consent.

The experiments were conducted in three sessions. In the first session, the subject's audiogram was obtained by a hearing healthcare professional. The DSLv5 prescription [33] was then used to define the subject's prescriptive gains across these five frequency bands that are commonly used [0-0.5], [0.5-1], [1-2], [2-4], and [4-6] kHz and the gains were translated to CRs. These CRs were used to initialize the MLIRL personalization algorithm described in section II.

In the second session, the subject's preferences were used to personalize the settings. Each subject sat in a sound booth wearing hearing aids in left and right ears connected via Bluetooth to a computer running the personalization algorithm. The hearing aids were programmed to have flat amplification, no compression on the hearing aid processors and all noise

$$L(\mathcal{D} \mid W, f) = \prod_{\substack{(s,a) \in \mathcal{D} \\ where\, f_a = f^+}} \pi_W(s,a) \prod_{\substack{(s,a) \in \mathcal{D} \\ where\, f_a = f^-}} (1 - \pi_W(s,a)) \tag{8}$$

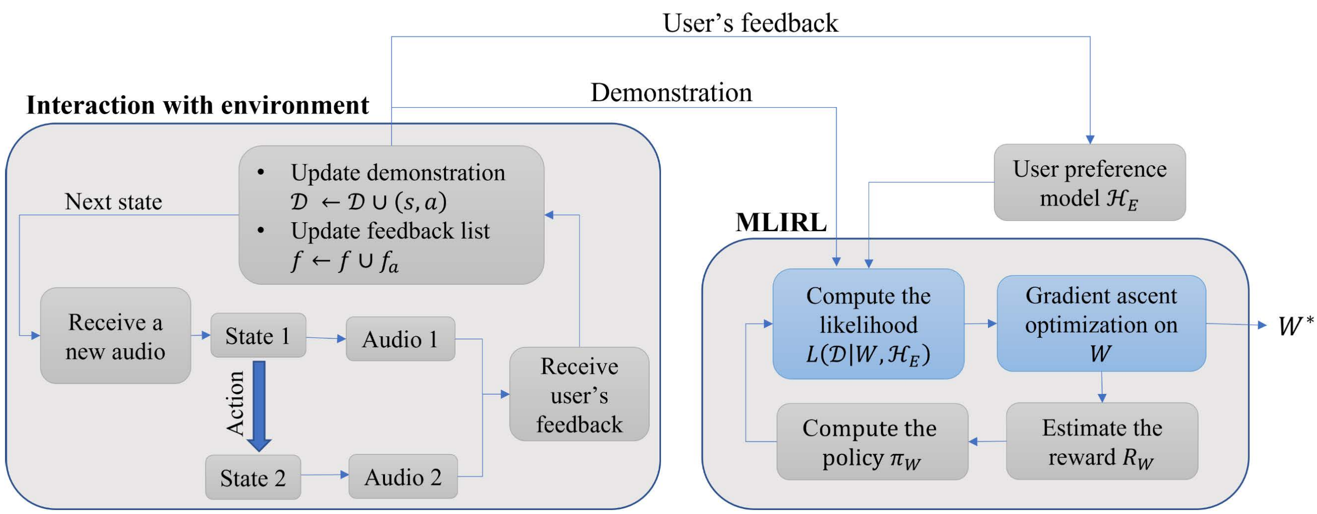


**FIGURE 5.** Block diagram of the personalized compression approach using MLIRL.

reduction and sound enhancement features were turned off. All the audio signal processing was performed on the computer and transmitted to the hearing aids via Bluetooth.

The experimenter first presented pairs of audio signals of about 2.5 second duration to the subject. Then the subject's feedback from each paired comparison was used to train the agent in the personalized compression algorithm. For each paired comparison, a stimulus was randomly selected from a dataset consisting of noisy speech sentences. The noisy speech sentences were generated by adding multi-talker babble noise to the TSP speech database [34]. The TSP speech database consists of over 1400 utterances spoken by 24 speakers (half male, half female). The SNR of the noisy sentences were set to a moderate noise level of 5 dB and the sampling rate was set to 16 kHz. To adjust the CRs, similar to our previous work [23], two scales for each frequency band were considered; $scale\,(i) \in \{1, 4\}, \forall 1 \le i \le N$. To avoid human fatigue by keeping the total duration of the sessions to less than two hours, an action space of 32 was thus considered by the above two scales and five CRs across the frequency bands of [0-0.5], [0.5-1], [1-2], [2-4], and [4-6] kHz. The preference data collection was conducted in 7 trajectories and each trajectory consisted of 31 paired comparisons. At the end of each trajectory, the reward function was updated in an online manner. Finally, the reward function of the last trajectory was used to define the policy.

The CRs with the highest probability was chosen to be the most preferred CR set defining the personalized compression

**TABLE 1.** Online personalization of compression algorithm.

1. Input: Markov Decision Process without reward, $n_{Trajectoies}$ number of trajectories, $n_{Steps}$ number of paired comparisons in each trajectory
2. $\mathcal{D} \leftarrow \emptyset$; $f \leftarrow \emptyset$; $\mathcal{D}$ denotes demonstration and $f$ denotes user's feedbacks
3. Initialize with the prescriptive compression ratios ($CR_{prescription}$)
4. For trajectory=1: $n_{Trajectoris}$
    - Interaction with environment $n_{Steps}$ times:
        a. Observe the environment $\rightarrow s$ (a randomly selected speech compressed by $CR^{init}$)
        b. Execute action $\rightarrow a$ (generate the speech compressed by $CR^{update}$)
        c. Receive feedback $f_a$
        d. Update demonstration ($\mathcal{D} \leftarrow \mathcal{D} \cup (s, a)$)
        e. Update feedback list ($f \leftarrow f \cup f_a$)
        f. Define next state based on the feedback
    - MLIRL:
        a. Compute preference model $\mathcal{H}_E$ based on the feedback list $f$ (equations 6 and 7)
        b. Gradients ascent optimization:
            i. Till convergence condition is reached:
                - Compute Q-value and policy ($Q_W$ equation 4, $\pi_W$ equation 3)
                - Compute $\nabla \log(L)$ equation 11
                - Update $W$ ($W \leftarrow W + \beta \nabla \log(L)$)
5. Output: $W^*$

**TABLE 2.** Parameters used in the MLIRL personalized compression.

| Parameter | Value | Description |
|---|---|---|
| $n_{Trajectories}$ | 7 | Number of trajectories |
| $n_{Steps}$ | 31 | Number of paired comparisons in a trajectory |
| $\alpha$ | 1 | Randomness control factor in policy definition |
| $\gamma$ | 0.9 | Discount factor |
| Attack Time | 0.01 | Time to respond to higher sound levels |
| Release Time | 0.1 | Time to respond to lower sound levels |
| CT | 60 | Compression threshold |

setting. The parameters used in the MLIRL personalized compression algorithm are shown in Table 2. The compression parameters of CT, attack time, and release time listed in this table correspond to the nominal values of the prescriptive compression.

## IV. RESULTS AND DISCUSSION

After obtaining each subject's audiogram, the DSLv5 prescriptive compression ratios were set as the initial condition to personalize the fitting. By going through the MLIRL personalization algorithm, the personalized CRs were obtained. Table 3 shows the audiograms of the ten subjects who participated in this study. Audiometric thresholds are shown for the five frequency bands used by the algorithm. This table also includes the prescriptive DSLv5 gains for the soft speech level and the corresponding CRs in the frequency bands. The last column shows the personalized CRs. The personalized CRs that are different than the prescriptive ones are bolded.

To compare preference between the personalized and the standard prescriptive settings, a preference comparison test was conducted for 30 sentences randomly selected from the TSP database. The subjects were presented with a pair of audios, one compressed by the prescriptive CRs (five CRs that are listed in the fourth column of Table 3) and the other compressed by the personalized CRs (five CRs that are listed in the fifth column of Table 3) in random order. Subjects had to define the audio that they preferred or if they had the same preference for both audios. Figure 6 shows the results of the preference testing. As can be seen from this figure, the personalized settings were preferred about 10 times more than the DSLv5 prescriptive settings. The only other personalization algorithm that has been reported in the literature for speech stimuli in the presence of babble background noise is the one in [23] named DRL. Similar to this work, DRL uses five frequency bands to personalize compression ratios associated with the DSLv5 prescriptive compression. As reported in [23], the DRL personalized settings were preferred about 7 times more than the DSLv5 prescriptive settings. Although the performance of the DRL personalization is comparable to the MLIRL personalization, the key difference is that the DRL algorithm training is done in an offline manner while the MLIRL training is done in an online manner. Note that the DRL personalization algorithm is not deployable in the field as it takes a few hours of training time on a modern computer, whereas the MLIRL personalization algorithm developed in this work is deployable in the field as it takes only a couple of seconds of training time on a modern computer.

A one-way ANOVA statistical test was performed on the preference results with compression setting (personalized or prescriptive DSLv5) as a within-subject factor to evaluate the statistical significance of the difference between the preference for the personalized versus prescriptive settings. A p-value lower than 0.05 was considered for rejecting the null hypothesis. The ANOVA test revealed that the personalized compression settings was preferred over the prescriptive settings in a statistically significant way ($F(1,18) = 122.36$, $p \ll 0.01$).

To ensure that the personalized settings did not compromise audibility or negatively impact word recognition or speech understanding, a word recognition test was conducted. One list of 50 words was selected from the Northwestern University Auditory Test No. 6 (NU-6) [30]. Half of the list was selected randomly and processed by the prescriptive settings and the other half was processed by the personalized settings. Subjects had to repeat each word back to the experimenter

**TABLE 3.** Prescriptive DSLv5 and personalized compression settings for ten subjects participated in the study.

| Subject | Audiogram for freq. bands [0.5, 1, 2, 4, 6] kHz | DSLv5 gains for soft speech | DSLv5 compression ratios | Personalized compression ratios |
|---|---|---|---|---|
| 1 | [15, 15, 15, 25, 40] | [5, 8, 10, 23, 25] | [1.2, 1.2, 1.3, 1.2, 1.4] | [1.2, **4.8**, 1.3, **4.8**, **5.6**] |
| 2 | [15, 15, 20, 20, 30] | [5, 6, 14, 15, 15] | [1.2, 1.2, 1.3, 1.2, 1.3] | [**4.8**, **4.8**, **5.2**, 1.2, 1.3] |
| 3 | [20, 25, 30, 50, 45] | [10, 12, 19, 36, 29] | [1.2, 1.3, 1.4, 1.5, 1.4] | [**4.8**, 1.3, 1.4, 1.5, 1.4] |
| 4 | [25, 30, 30, 40, 45] | [11, 14, 19, 29, 29] | [1.2, 1.3, 1.4, 1.3, 1.4] | [**4.8**, **5.2**, **5.6**, 1.3, 1.4] |
| 5 | [15, 15, 15, 30, 30] | [8, 9, 10, 25, 20] | [1.2, 1.3, 1.3, 1.3, 1.3] | [1.2, 1.3, 1.3, 1.3, **5.2**] |
| 6 | [5, 10, 25, 30, 40] | [1, 5, 17, 25, 25] | [1.2, 1.2, 1.4, 1.3, 1.4] | [**4.8**, **4.8**, 1.4, 1.3, **5.6**] |
| 7 | [15, 25, 20, 30, 35] | [11, 12, 13, 25, 22] | [1.2, 1.3, 1.3, 1.3, 1.3] | [**4.8**, 1.3, **5.2**, **5.2**, **5.2**] |
| 8 | [10, 20, 20, 50, 40] | [7, 8 13, 34, 25] | [1.2, 1.2, 1.3, 1.4, 1.4] | [**4.8**, 1.2, **5.2**, 1.4, **5.6**] |
| 9 | [20, 20, 25, 35, 25] | [9, 10, 17, 32, 17] | [1.2, 1.3, 1.4, 1.5, 1.2] | [**4.8**, **5.2**, **5.6**, 1.5, 1.2] |
| 10 | [15, 20, 25, 35, 40] | [9, 10, 17, 29, 25] | [1.2, 1.3, 1.4, 1.4, 1.4] | [**4.8**, **5.2**, 1.4, 1.4, **5.6**] |

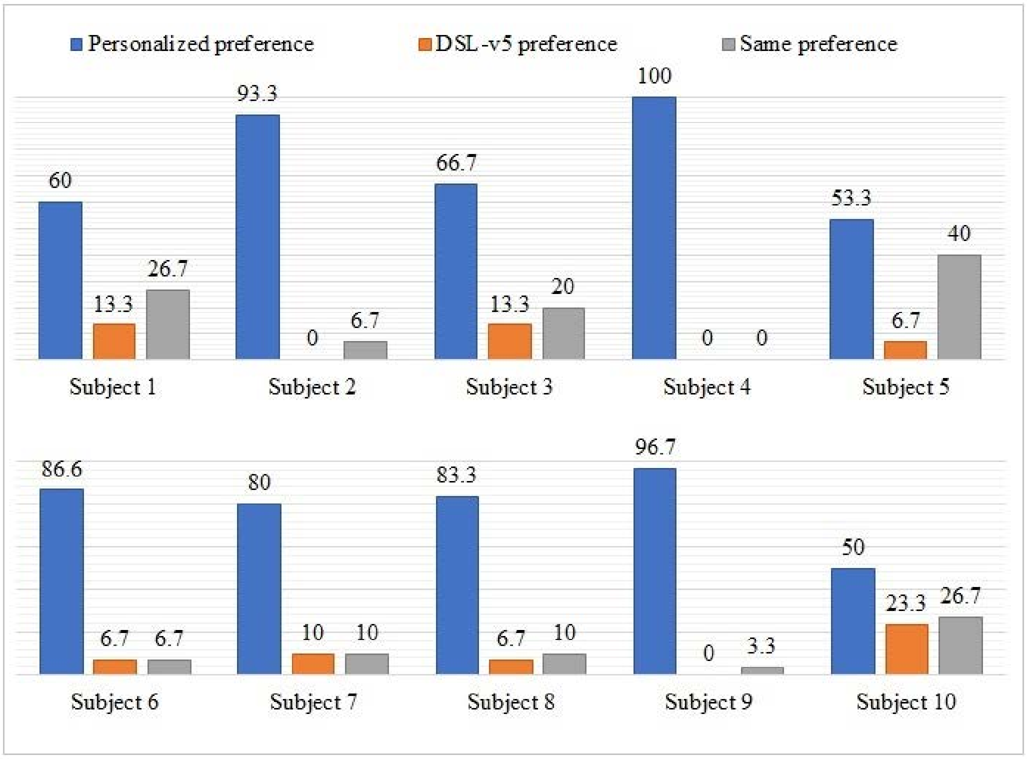


**FIGURE 6.** Preference comparison between the personalized and DSLv5 compressions in percentages.

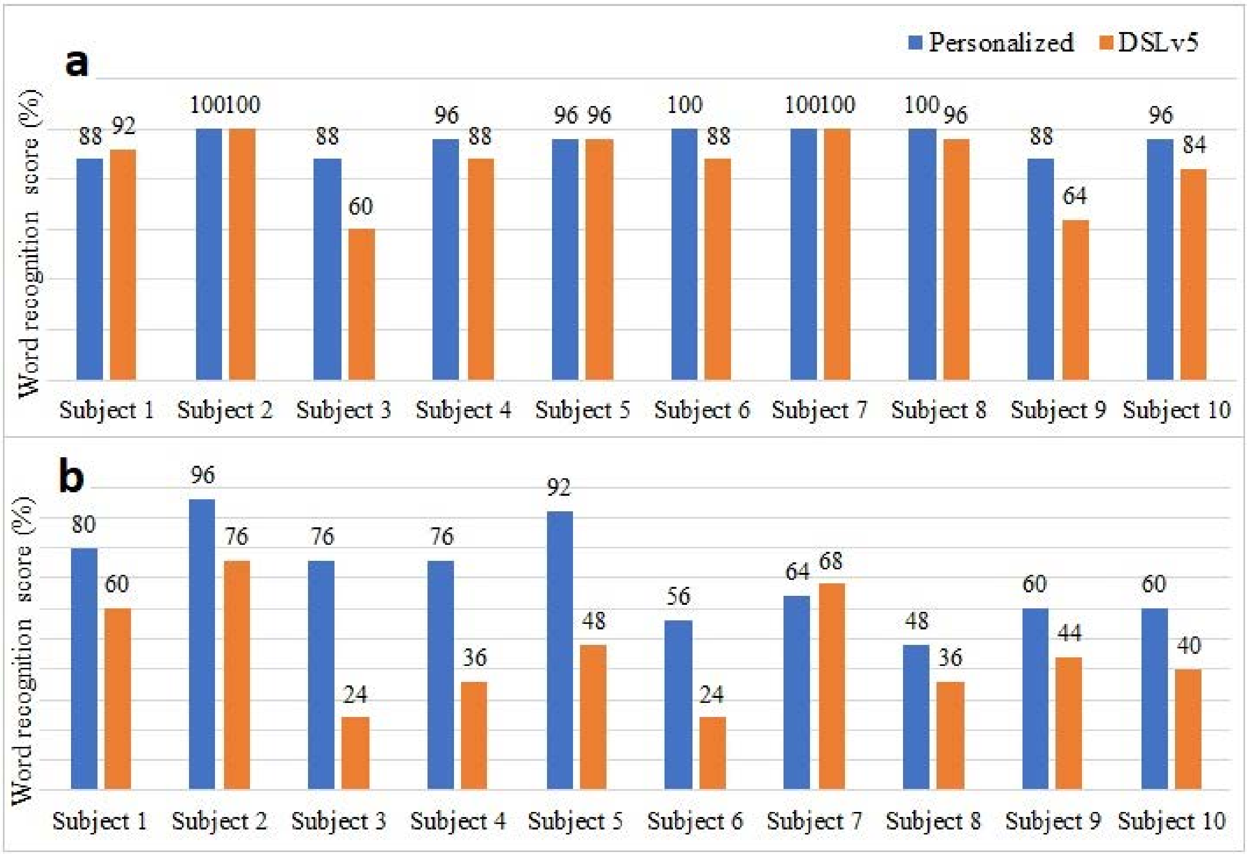


**FIGURE 7.** Word recognition score in percentages for (a) clean speech signals, and (b) noisy speech signals at 5dB SNR.

after it was presented to them. Another list of 50 words was selected from the same database for the word recognition test in the presence of noise. The noisy sentences at 5 dB SNR were created by adding the multi-talker babble noise to the clean speech signals.

The word recognition scores obtained are shown in Figure 7 in terms of the percentages of the words correctly recognized out of the total number of the presented words in quiet and noisy conditions. As can be seen from this figure, except for one subject, the same number of words or more words were recognized when using the personalized settings compared to when using the prescriptive settings in quiet condition. Adding the noise resulted in lower word recognition scores for both the personalized and prescriptive settings compared to the quiet condition. In noisy condition, for all but one of the subjects, the decrease in the word recognition score was lower for the personalized settings.

The codes developed for the personalized compression reported in this paper can be obtained via a GitHub link by contacting the authors.

## V. CONCLUSION AND FUTURE WORK

A machine learning personalization approach for performing the compression function in hearing aids has been developed in this paper. This approach utilizes maximum likelihood inverse reinforcement learning to establish personalized compression for hearing aid fitting purposes in an online manner. As a feasibility or proof-of-concept study, ten hearing impaired subjects participated in the testing of this developed personalized compression approach. The results of preference comparison between the prescriptive DSLv5 compression and the personalized compression showed that the personalized compression settings were preferred about 10 times more than the prescriptive compression settings which is higher than the previous state-of-the-art compression personalization method. Furthermore, the developed personalized compression settings did not have an adverse impact on the word recognition scores and in fact improved the word recognition scores in noisy condition compared to the prescriptive compression settings. It is worth noting here that for commercial deployment of the developed personalized compression solution, it is necessary to examine a much larger pool of hearing impaired subjects.

Noting that the developed personalized compression approach is trainable in an online manner, it can get deployed in the field or in real-world audio environments. This enables the user to find the preferred settings for different audio environments. As our future work, we plan to develop a smartphone app for this purpose so that the experimentation can be carried out live in the field.

## ACKNOWLEDGMENT

The authors wish to express their appreciation to Andie Lea Zang for her help in conducting the subjects testing.

## REFERENCES

[1] C. J. Plack and N. F. Viemeister, "Suppression and the dynamic range of hearing," *J. Acoust. Soc. Amer.*, vol. 93, no. 2, pp. 976–982, Feb. 1993, doi: 10.1121/1.405403.

[2] H. G. Hassager, A. Wiinberg, and T. Dau, "Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment," *J. Acoust. Soc. Amer.*, vol. 141, no. 4, pp. 2556–2568, Apr. 2017, doi: 10.1121/1.4979783.

[3] 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, doi: 10.1177/108471380500900403.

[4] G. Keidser, H. R. Dillon, M. Flax, T. Ching, and S. Brewer, "The NAL-NL2 prescription procedure," *Audiol. Res.*, vol. 1, no. 1S, pp. 1–3, 2011, doi: 10.4081/audiores.2011.e24.

[5] R. A. Bentler and L. J. Cooley, "An examination of several characteristics that affect the prediction of OSPL90 in hearing aids," *Ear Hearing*, vol. 22, no. 1, pp. 58–64, Feb. 2001.

[6] J. Chalupper, D. Junius, and T. Powers, "Algorithm lets users train aid to optimize compression, frequency shape, and gain," *Hear. J.*, vol. 62, no. 8, pp. 26–33, 2009, doi: 10.1097/01.HJ.0000359131.58356.8f.

[7] G. Keidser and H. Dillon, "What's new in prescriptive fittings down under?" *Hear. Care adults*, vol. 13, pp. 133–142, Nov. 2006.

[8] G. Keidser, C. Brew, S. Brewer, H. Dillon, F. Grant, and L. Storey, "The preferred response slopes and two-channel compression ratios in twenty listening conditions by hearing-impaired and normal-hearing listeners and their relationship to the acoustic input," *Int. J. Audiol.*, vol. 44, no. 11, pp. 656–670, Jan. 2005, doi: 10.1080/14992020500266803.

[9] S. H. Yoon, K. W. Nam, S. Yook, B. H. Cho, D. P. Jang, S. H. Hong, and I. Y. Kim, "A trainable hearing aid algorithm reflecting individual preferences for degree of noise-suppression, input sound level, and listening situation," *Clin. Experim. Otorhinolaryngol.*, vol. 10, no. 1, pp. 56–65, Mar. 2017.

[10] G. Keidser and K. Alamudi, "Real-life efficacy and reliability of training a hearing aid," *Ear Hearing*, vol. 34, no. 5, pp. 619–629, 2013.

[11] L. L. N. Wong, "Evidence on self-fitting hearing aids," *Trends Amplification*, vol. 15, no. 4, pp. 215–225, Dec. 2011, doi: 10.1177/1084713812444009.

[12] B. Johansen, M. K. Petersen, M. J. Korzepa, J. Larsen, N. H. Pontoppidan, and J. E. Larsen, "Personalizing the fitting of hearing aids by learning contextual preferences from Internet of Things data," *Computers*, vol. 7, no. 1, pp. 1–21, 2018, doi: 10.3390/computers7010001.

[13] G. Aldaz, S. Puria, and L. J. Leifer, "Smartphone-based system for learning and inferring hearing aid settings," *J. Amer. Acad. Audiol.*, vol. 27, no. 9, pp. 732–749, 2016, doi: 10.3766/jaaa.15099.

[14] 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 Hear.*, vol. 22, pp. 1–13, Sep. 2018, doi: 10.1177/2331216518798264.

[15] N. S. Jensen, O. Hau, J. B. B. Nielsen, T. B. Nielsen, and S. V. Legarth, "Perceptual effects of adjusting hearing-aid gain by means of a machine-learning approach based on individual user preference," *Trends Hear.*, vol. 23, pp. 1–23, May 2019, doi: 10.1177/2331216519847413.

[16] J. Brehm, B. Nielsen, J. Nielsen, J. Larsen, and S. Member, "Perception-based personalization of hearing aids using Gaussian processes and active learning," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 1, pp. 162–173, Jan. 2014.

[17] F. K. Kuk and N. M. C. Pape, "The reliability of a modified simplex procedure in hearing aid frequency-response selection," *J. Speech, Lang., Hearing Res.*, vol. 35, no. 2, pp. 418–429, Apr. 1992, doi: 10.1044/jshr.3502.418.

[18] F. K. Kuk and N. M. C. Pape, "Relative satisfaction for frequency responses selected with a simplex procedure in different listening conditions," *J. Speech, Lang., Hearing Res.*, vol. 36, no. 1, pp. 168–177, Feb. 1993, doi: 10.1044/jshr.3601.168.

[19] A. C. Neuman, H. Levitt, R. Mills, and T. Schwander, "An evaluation of three adaptive hearing aid selection strategies," *J. Acoust. Soc. Amer.*, vol. 82, no. 6, pp. 1967–1976, Dec. 1987, doi: 10.1121/1.395641.

[20] E. A. Durant, G. H. Wakefield, D. J. V. Tasell, and M. E. Rickert, "Efficient perceptual tuning of hearing aids with genetic algorithms," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 2, pp. 144–155, Mar. 2004.

[21] D. Baskent, C. L. Eiler, and B. Edwards, "Using genetic algorithms with subjective input from human subjects: Implications for fitting hearing aids and cochlear implants," *Ear Hearing*, vol. 28, no. 3, pp. 370–380, 2007.

[22] J. B. Nielsen, J. Nielsen, B. S. Jensen, and J. Larsen, "Hearing aid personalization," in *Proc. 3rd Workshop Mach. Learn. Interpretation Neuroimaging (NIPS)*, Lake Tahoe, NV, 2013, pp. 2–5.

[23] 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, doi: 10.1109/ACCESS.2020.3035728.

[24] S. Akbarzadeh, N. Alamdari, C. Campbell, E. Lobarinas, and N. Kehtarnavaz, "Word recognition clinical testing of personalized deep reinforcement learning compression," in *Proc. IEEE 14th Dallas Circuits Syst. Conf. (DCAS)*, Nov. 2020, pp. 1–2, doi: 10.1109/DCAS51144.2020.9330651.

[25] M. Babes-Vroman, V. Marivate, K. Subramanian, and M. Littman, "Apprenticeship learning about multiple intentions," in *Proc. 28th Int. Conf. Mach. Learn. (ICML)*, 2011, pp. 897–904.

[26] M. Wen, I. Papusha, and U. Topcu, "Learning from demonstrations with high-level side information," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3055–3061, doi: 10.24963/ijcai.2017/426.

[27] N. Mourad, A. Ezzeddine, B. N. Araabi, and M. N. Ahmadabadi, "Learning from demonstrations and human evaluative feedbacks: Handling sparsity and imperfection using inverse reinforcement learning approach," *J. Robot.*, vol. 2020, pp. 1–18, Jan. 2020, doi: 10.1155/2020/3849309.

[28] M. K. Ho, M. L. Littman, J. MacGlashan, F. Cushman, and J. L. Austerweil, "Showing versus doing: Teaching by demonstration," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3035–3043.

[29] S. Arora and P. Doshi, "A survey of inverse reinforcement learning: Challenges, methods and progress," *Artif. Intell.*, vol. 297, Aug. 2021, Art. no. 103500.

[30] H. B. Suay, T. Brys, M. E. Taylor, and S. Chernova, "Learning from demonstration for shaping through inverse reinforcement learning," *Proc. Int. Jt. Conf. Auton. Agents Multiagent Syst. (AAMAS)*, 2016, pp. 429–437.

[31] K. Shiarlis, J. Messias, and S. Whiteson, "Inverse reinforcement learning from failure," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, 2016, pp. 1060–1068, doi: 10.1007/s10994-018-5730-4.

[32] C. Xia and A. El Kamel, "Neural inverse reinforcement learning in autonomous navigation," *Robot. Auto. Syst.*, vol. 84, pp. 1–14, Oct. 2016, doi: 10.1016/j.robot.2016.06.003.

[33] *DSL ® v5 by Hand*. pp. 1–13. Accessed: 2014. [Online]. Available: https://www.dslio.com/wp-content/uploads/2014/06/DSL-5-by-Hand.pdf

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

**SARA AKBARZADEH** (Student Member, IEEE) is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Texas at Dallas. She has published ten papers in her research areas. Her research interests include auditory neuroscience, signal/speech processing, and machine learning, in particular as related to hearing assistive devices.

**EDWARD LOBARINAS** is a Professor with the Callier Center for Communication Disorders, School of Behavioral and Brain Sciences, University of Texas at Dallas. He is recognized by the American Academy of Audiology as a Jerger Future Leader of Audiology. He received grants from the National Institute of Health, the American Tinnitus Association, and the 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 an Erik Jonsson Distinguished Professor with the Department of Electrical and Computer Engineering and the Director of the Embedded Machine Learning Laboratory, University of Texas at Dallas, Richardson, TX, USA. He has authored or coauthored 11 books and over 400 publications in his research 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, a Licensed Professional Engineer, and the Editor-in-Chief of *Journal of Real-Time Image Processing*.