IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 18, 2020, accepted October 27, 2020, date of publication November 3, 2020, date of current version November 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3035728

Personalization of Hearing Aid Compression by Human-in-the-Loop Deep Reinforcement Learning

NASIM ALAMDARI[1], (Student Member, IEEE), EDWARD LOBARINAS[2], AND NASSER KEHTARNAVAZ[1], (Fellow, IEEE)

[1]Electrical and Computer Engineering Department, 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: Nasim Alamdari (alamdari@utdallas.edu)

ABSTRACT Existing prescriptive compression strategies used in hearing aid fitting are designed based on gain averages from a group of users which may not be necessarily optimal for a specific user. Nearly half of hearing aid users prefer settings that differ from the commonly prescribed settings. This paper presents a human-in-the-loop deep reinforcement learning approach that personalizes hearing aid compression to achieve improved hearing perception. The developed approach is designed to learn a specific user's hearing preferences in order to optimize compression based on the user's feedbacks. Both simulation and subject testing results are reported. These results demonstrate the proof-of-concept of achieving personalized compression via human-in-the-loop deep reinforcement learning.

INDEX TERMS Personalized audio compression, deep reinforcement learning, human-in-the-loop personalization, personalized hearing aid, hearing aid compression.

I. INTRODUCTION

In hearing impaired individuals, the relative intensity difference between barely audible and uncomfortably loud sound becomes smaller. Thus, in order to achieve optimal audibility, sound must be calibrated to occupy a smaller range of sound pressure levels (SPLs). This dynamic range adjustment is achieved through the process of compression [1]. Compression in a reduced dynamic range is the key function of modern hearing aids. This process involves squeezing or fitting sound into the residual audibility range of a hearing aid user. In hearing aid fitting, so-called compression curves are set up by adjusting gains across a number of frequency bands based on a user's audiometric profile. The two most widely used hearing aid prescriptions are NAL-NL2 [2] and DSL-v5 [3]. These prescriptions correspond to gain tables across a number of frequency bands for three sound levels, soft, moderate, and loud.

It has been reported that up to half of individuals using fitted hearing aids preferred amplification or compression settings different than the prescription provided [4]–[9]. Considering that suprathreshold hearing perception varies from person to person and that acoustic environments encountered vary from person to person, several papers in the literature have examined self-adjustment or self-tuning of hearing aid fitting relative to the one-size-fits-all prescriptive fitting [10]–[18]. In [18], it was reported that hearing aid users favored gain settings that were different from the NAL prescription settings both in quiet and in noise. Self-adjustments carried out on a custom hardware/software such as those recently reported in [19], [20] have also demonstrated improvements in hearing perception that can be gained over prescriptive fitting. In addition, the benefits of hearing aid personalization were examined in [11], [12]. In [11], the parameter space consisted of four possible combinations of microphone mode (omnidirectional and directional) and noise reduction state (active and off). First, preferences of a user were learnt in order to create a supervised trained model. Then, the model was used to derive an optimal setting among the four choices.

There have been only a few studies reporting algorithms to learn personalize audio compression gains (or compression ratios). In [21]–[23], a machine learning approach for self-adjustment or self-tuning of compression was presented.

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

In these papers, a Gaussian regression model was used to achieve personalized compression by estimating its parameters from training data. User preferences were obtained via hearing assessments by listening to music clips in [21], [22]. Although the results reported show the benefits of personalization, differences in preferences between music clips and conversation in noisy environments (e.g., in babble noise) were not addressed. Understanding speech in the presence of bothersome background noise is expressed as a major challenge by hearing aid users [24]. Furthermore, in [21]–[23], only twenty preference iterations were done for modeling the hearing preference of a user. In actual audio environments, this many iterations would be inadequate for modeling various non-linearities associated with hearing perception. Although these findings show the overall usefulness of personalization, preferred hearing cannot be achieved without a proper design of the personalization framework. In a recent study [25], an agent is trained using simulated contextual preferences within a controlled environment. There, the user model is created by hypothesizing correlation among users' preferences based on a number of their observable characteristics. However, the validity of the assumption made in [25], i.e. categorizing users' preferences, is not supported by clinical evidence. Hence, human feedback in the training loop is deemed vital in order to provide a personalization that can deliver preferred hearing to a specific user.

To address previous design limitations, a human-in-the-loop (HITL) interactive machine learning compression approach based on deep reinforcement learning (DRL) [26] is developed in this paper. In our approach, the user is placed in the learning loop. A DRL network is designed to receive preference feedback from the user. As a result, it becomes possible to deal with various non-linearities of human hearing perception. In general, placing human in the loop of training a machine learning model enables reducing the error made by a trained model. The use of the conventional reinforcement learning is not sufficient to perform personalization for hearing aid compression. For compression, it is vital to include data from human feedback to optimize and improve the model over time. That is why in this work, a user's feedback is placed in the learning loop for personalization of compression, considering that the user's feedback is sparse in practice. A combination of a convolutional neural network (CNN) [27] and a bidirectional long short-term memory recurrent neural network [28] or CNN-BiLSTM is used to model a user's preferences in those audio environments that are of interest to the user.

As discussed in [29], user feedback is affected by biases. Thus, rather than absolute feedback, pairwise or relative hearing assessments are deemed more suitable [30]. Hence, in our approach, a user is subjected to a series of compressed audios to express his/her preferences towards training the model via the reward/punishment mechanism of reinforcement learning; an approach that has been successfully applied to gaming [31] and robotics [32]. The developed DRL approach provides personalized compression that can be utilized in the field for hearing aid compression studies. It should be noted that the focus of this paper is on the development of a human-in-the-loop deep reinforcement learning approach for personalizing audio compression and to show its proof-of-concept by carrying out simulated experiments and a limited clinical subject testing. However, deployment would require carrying out extensive clinical testing.

To describe our approach in detail, the remainder of this paper is organized as follows. Section II covers the developed approach to personalize hearing aid compression or fitting via human-in-the-loop DRL as well as a protocol to perform human preference assessment. The experimental results and discussion are then presented in section III followed by the conclusion in section IV.

II. PERSONALIZED COMPRESSION APPROACH

To set the stage for the developed personalized compression, the conventional reinforcement learning (RL) is first briefly described. In a reinforcement learning framework, an agent and an environment interact over a series of steps. At each time step t, the agent receives an observation or state $s_t \epsilon S$ from the environment and sends an action $a_t \epsilon A$ back to the environment with S and A denoting the state and action sets, respectively. In a conventional RL framework, based on a given action, the environment generates the next state together with a reward $r_t \epsilon R$ with R denoting the reward set, and the goal is to maximize reward over time. Fig. 1(a) shows a block diagram of a conventional RL framework.

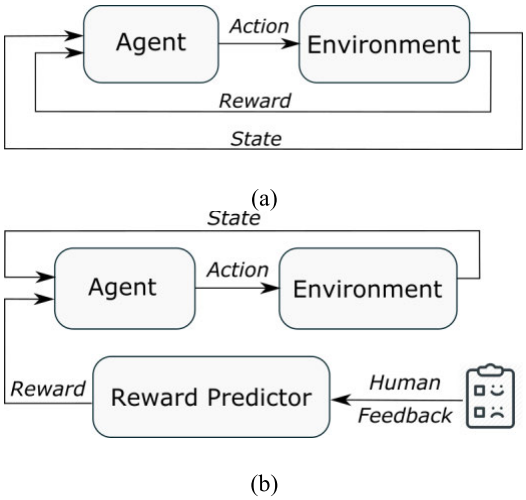


FIGURE 1. (a) Block diagram of a conventional reinforcement learning framework. (b) Deep reinforcement learning with user's feedback in which reward is obtained based on user preferences.

The success of RL heavily depends on setting up an effective reward function. Many real-world problems are complex, and it is often difficult to formulate an effective reward function. Inverse reinforcement learning (IRL) [33] can be used to design a reward function, which can then get deployed to train the agent using (deep) reinforcement learning. In order to build an effective reward function, human feedback can be

used to evaluate the behavior of the agent [34], [35]. In our case, in order to model and learn hearing preferences via deep reinforcement learning, the listener's preferences are used.

Obtaining rewards in a direct manner, based on user feedback, is labor intensive and makes the training process impractical because thousands of iterations and user feedbacks would be needed. In order to decrease the number of user feedbacks and thus enable a practical deployment of the personalized compression, first a reward function is considered to model hearing preferences of a user in an asynchronous manner. This is achieved by carrying out comparison between instances of two different compressed audios. Then, an agent is trained to maximize reward. Fig. 1(b) shows a block diagram of this approach. Unlike the conventional RL in which reward is computed by the environment, here reward is computed based on user preferences.

A. PERSONALIZED FITTING PROTOCOL

Compression in hearing aid fittings is normally performed via software tools that are provided by hearing aid manufacturers. These software tools are used to automatically set gains across a number of frequency bands using established prescriptions based on group averages or with manufacturers adding their own variations.

A user's audiogram and the prescription gains are used to set the target gains for that user. Across each frequency band, a different compression curve is used to generate multiband dynamic range compression (DRC). In this paper, the DSL-v5 by Hand prescription in [36] is considered to serve as the reference compression. In other words, the gains in the DSL-v5 tables are used to compute the reference DRC parameters consisting of compression ratio (change in gain) and compression threshold (sound level at which compression is applied). The process of personalization involves modification of the gains specified by DSL-v5 or any other generic prescription based on user preferences.

In this study, the following steps are taken to achieve personalized compression for a specific user. The first step is assessing hearing sensitivity by measuring the audiogram of a user. In the second step, the compression gains of the user are set by using fitting software. In this work, the DSL-v5 prescriptive fitting software is used. The third step is initializing the human centered-DRL framework with the compression ratios obtained in the second step as the starting point. In the fourth step, the compression ratios are adjusted by going through the training process of the human centered-DRL framework (an illustration of the gain change ranges for the agent action in the DRL framework is depicted in Fig. 2). The final or fifth step involves comparing the performance of the personalized compression with the prescriptive or reference compression.

In the personalized framework, an agent that is interacting sequentially with the environment over a number of time steps is considered similar to the one in [34]. At each time step, the agent receives a new state (observation) from

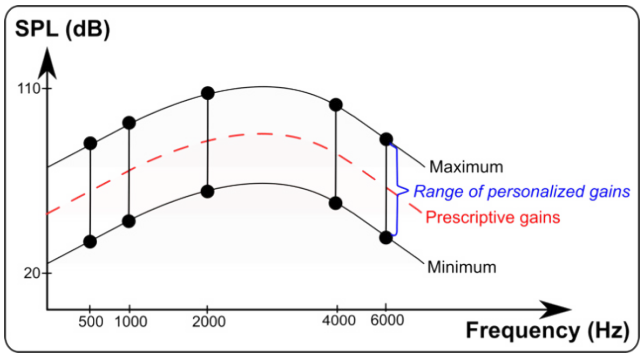


FIGURE 2. Illustration of gain ranges across different frequency bands.

the environment and performs an action. Over an episode, the agent performs interaction for a number of time steps. In typical RL settings, a reward at each time step is fed into the agent as well.

However, here rather than a predefined reward, a reward that is modeled based on a user's preferences is considered. In other words, by putting human feedback in the learning loop, it is attempted to optimize the agent learning and thus the user hearing perception. The personalization protocol and block diagram of the developed DRL-based personalized hearing aid compression framework is shown in Table 1 and Fig. 3, respectively. Each block in Fig. 3 is described in the subsections that follow.

B. ENVIRONMENT

In the personalized framework, the environment consists of three components: *audio segment creation*, *compression ratio update*, and *agent state transition function*. At each policy time step, a noisy speech audio signal is down-sampled from 48 kHz to 16 kHz to lower the computational burden. Noisy speech audio signals are generated by distorting the widely used public domain IEEE speech dataset [37] with babble restaurant noise (provided on YouTube) and SNR of about 0 dB. The IEEE dataset consists of 3600 speech audio files by 20 speakers (10 females and 10 males) in which each file is about 2 seconds long. The speakers are from two American English regions of the Pacific Northwest (PN) and the Northern Cities (NC) reading the IEEE "Harvard" sentences. A total of 3600 noisy speech audio signals are thus generated and at each time step, a randomly selected audio signal is used for preference training. Once a new action a_{t+1} is received from the agent, CR (compression ratio) in the frequency bands are updated based on the action a_{t+1}. Depending on the number of scales (β) specified for adjusting CR in each of the frequency bands (CR_{adj}), a set of actions is created by permutations and the action space A is given by

$$A = \prod_{i}^{nBands} \beta_i \tag{1}$$

where each β_i denotes the number of actions corresponding to ith frequency band. In other words, the action space A is the product space of the action space in all the frequency

TABLE 1. Personalized fitting protocol.

Step	Description
1.	Measuring audiogram of the user
2.	Defining the compression gains of the user using a prescriptive fitting software (e.g., DSL-v5 in [3]); computing compression ratios in a number of frequency bands
3.	Initializing the human-centered DRL framework with the above compression ratios
4.	Running the policy in the environment and storing a set of compression ratio adjustments that are resulted from randomly generated agent's actions;
5.	Generating pairs of compressed audio signals with compression ratios in step 4
6.	Asking the user to label each pair and add audio pairs and their labels to a buffer
7.	Training the preference (reward) predictor using the buffer
8.	Training the RL policy, based on the observation received from the environment with the reward from the trained reward predictor.
9.	**For** M iterations **do**:
10.	Training the policy in the environment for N_{steps} with the reward from the reward predictor
11.	Selecting compressed audio pairs resulted from given actions
12.	Asking the user to declare preferences and adding preferences to a buffer
13.	Fine-tuning the reward model for k batches from the above buffer
14.	**End for**
15.	Hearing assessment to compare personalized compression ratios with those specified by DSL-v5 prescriptive compression

bands. Consequently, adjusting CR in each frequency band is achieved as follows:

$$CR_{new}(f) = CR_{DSL-v5}(f).CR_{adj}(f) \tag{2}$$

where $CR_{adj}(f)$, $CR_{DSL-v5}(f)$, and $CR_{new}(f)$, respectively, stand for the compression ratio adjustment, the compression ratio computed from the DSL-v5 prescription, and the new compression ratio in the f^{th} frequency band. Permutations are defined by a dictionary in which each action is mapped to a set of compression ratio adjustments across all the frequency bands. In our experimentations, to keep the subject training time under two hours, β is set to 2 in each frequency band for an action space of 32. It should be noted that the introduced methodology is general purpose in the sense that it is applicable to higher β values. When using higher β values ($\beta > 2$), more feedbacks from a subject are needed resulting in higher subject training time.

In *agent state transition function*, a new audio signal is compressed by the updated CRs from the previous iteration using the dynamic range compression whose details are described in our previous publication [36]. Due to a better match to the human hearing perception, Mel-scale frequencies are often used instead of linear scale frequencies to represent noisy speech signals in classification tasks [38]. Similarly, compressed noisy speech signals are sampled at 16kHz and framed to 20ms by using a Hanning window with 50% overlap. This translates into a frame size of $0.020 \times 16000 = 320$ samples. The short time Fourier transform (STFT) of frames are then computed. The conversion of frequency f in a STFT-frame to Mth Mel-scale frequency is done as follows [39]:

$$M(f) = 2595 log_{10}\left(1 + \frac{f}{700}\right) \tag{3}$$

Log Mel-spectrogram features are extracted from each STFT frame using a bank of 80 Mel filters. The log Mel-spectrogram features of 240 consecutive frames from an audio segment are stacked to create a 2D feature matrix (80×240). This 2D feature matrix is reshaped to a 3D feature space creating three adjacent images ($80 \times 80 \times 3$), which is considered to be one observation for agent training. For training the reward estimator, the 2D format of the observation (80×240) is considered to be the input/observation to the network.

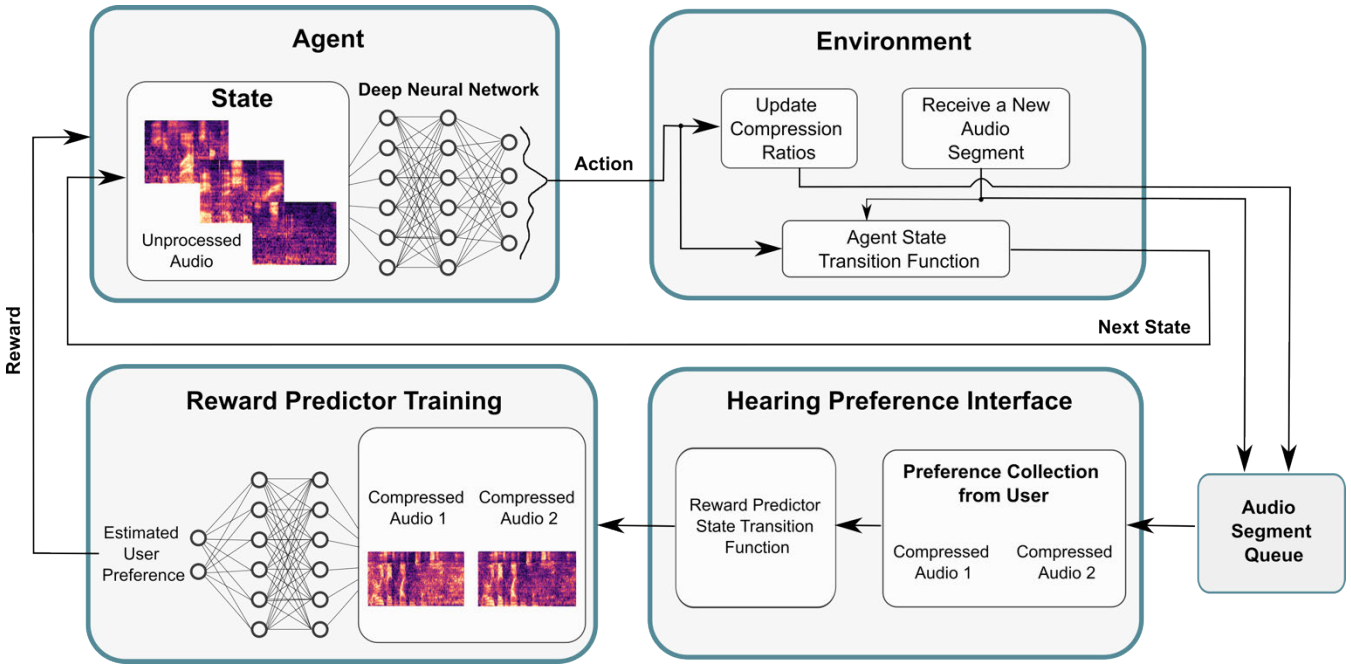
Note that in contrast to the conventional RL, here the end-of-episode sign from the environment is not shared with the agent to make the agent training one uninterrupted episode. Moreover, reward values are received from the reward predictor rather than from the environment.

Next, unprocessed audio signals and updated CRs are added to a buffer called *audio segment queue* as depicted in Fig. 3(a). The audio segment queue σ shown in this figure denotes a set or collection of audio signals U and compression ratios CR computed from a number of k actions, that is

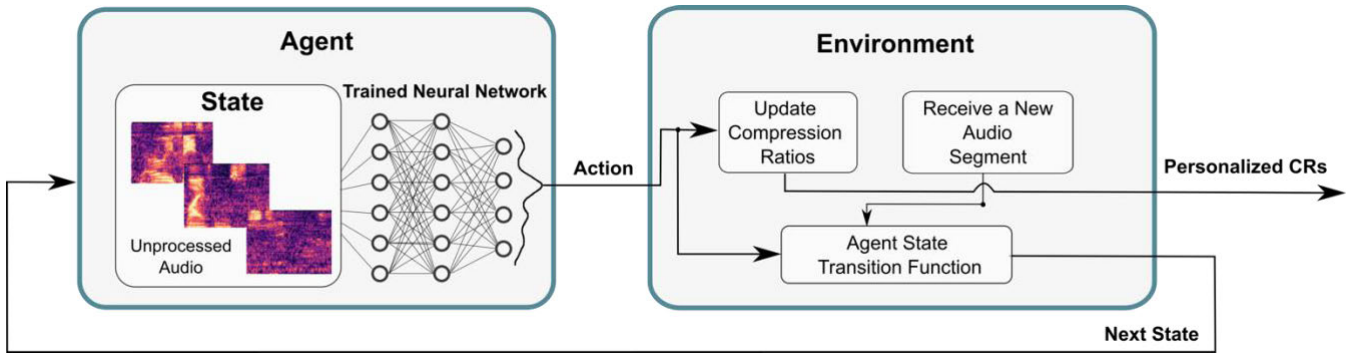
$$\sigma = ((u_0, cr_0), (u_1, cr_1), \ldots, (u_{k-1}, cr_{k-1})) \epsilon (U, CR)^k \tag{4}$$

C. HUMAN PREFERENCE INTERFACE

In order to use human input in the learning loop, a hearing preference interface is created to collect the user's hearing



(a) Training mode



(b) Operation mode

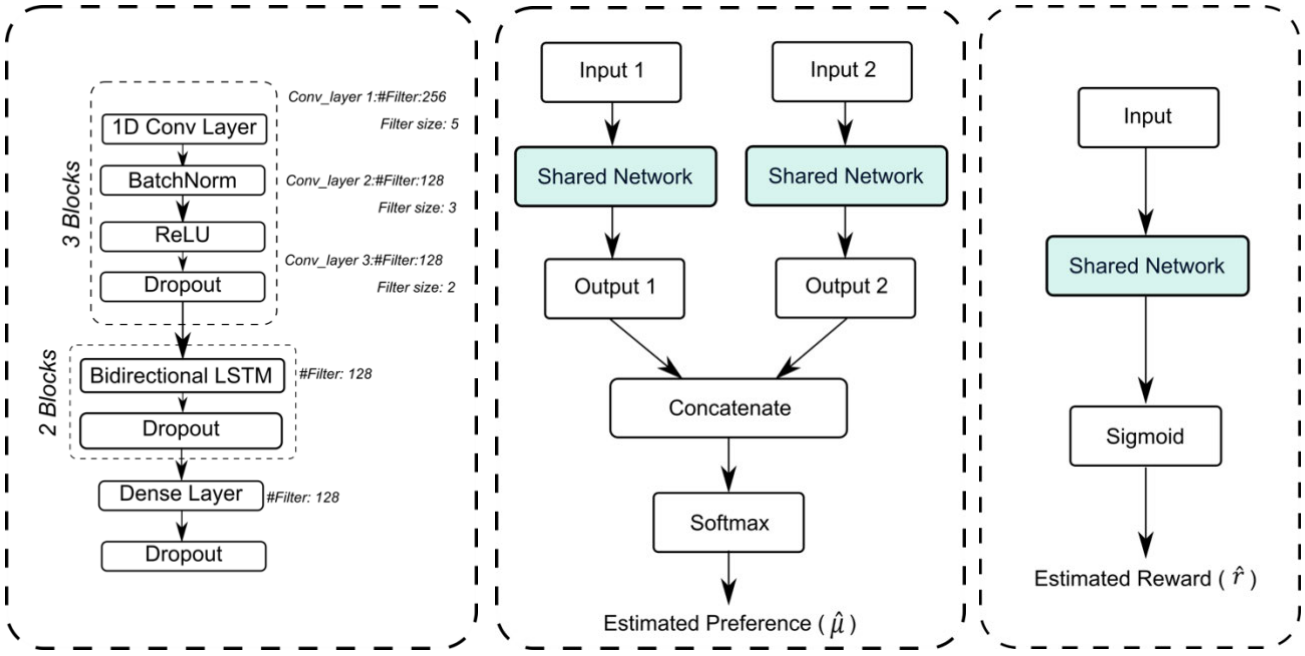
FIGURE 3. **Developed personalized compression DRL framework: (a) training mode and (b) operation mode.**

feedbacks from a group of comparisons of audio signal pairs that are compressed with two different sets of compression ratios. The goal of this user interaction is to learn the non-linearities associated with the user's preferences or reward function.

In *hearing preference interface* shown in Fig. 3, two pairs from the queue σ are selected at each time step. Then, a corresponding pair of compressed audio signals (c^1, c^2) is computed which is used for the comparison. The user is given 4 options to indicate his/her preference: (1) $\mu = [1, 0]$ if c^1 is preferable, (2) $\mu = [0, 1]$ if c^2 is preferable, (3) $\mu = [0.5, 0.5]$ if both compressed audio signals are equally preferred, and (4) neither compressed audio signals are desired. Hearing preferences are collected over a series of compressed audio signal pairs and are stored in a dataset D of triplets (c^1, c^2, μ), where μ denotes the feedback label, and c^1 and c^2 are the two compressed audio signals created by applying two different sets of CRs to the same noisy speech signal. Note that for option (4), the comparison is excluded from the dataset D.

In *reward predictor state transition function*, a batch of data from D is used to train the reward predictor model to improve the agent policy. Similar to the agent's state transition function, each compressed audio signal is framed into 20ms frames with 50% overlap. Log Mel-spectrogram features of each frame in an audio segment are computed and considered to be one observation.

Data Augmentation - The performance of the human hearing preference estimator depends on both the number of feedbacks acquired from the user and the model structure. A data augmentation is thus performed to address the limited size of training data that is available to the reward estimator



(a) Structure of shared network (b) Training mode and its network structure (c) Operation mode

FIGURE 4. Network structure of the reward predictor.

in practice. For this reason, first the data size is doubled by performing data flipping. This data augmentation consists of creating realistic samples by substituting features of audio signal 1 (c^1) with features of audio signal 2 (c^2). Their corresponding preference label is switched accordingly. The goal of the data augmentation here is to enhance the generalization capability of the preference (reward) predictor.

In addition to increasing the size of the training data, it is made sure that the training data does not suffer from unbalanced labels. Unbalanced labels can cause the model not to learn the learning preferences due to: (1) the network model not getting optimized for the unbalanced label in the original dataset, and (2) the accuracy of a validation or test set drops as it is challenging to have a complete representation with few observations. To resolve imbalanced labels, an undersampling is done by reducing the number of samples in the class with more labels to match the number of samples in the class with fewer labels.

D. PREFERENCE/REWARD PREDICTOR

In the reward predictor block shown in the Fig. 3, the parameters reflecting the reward are obtained via a combination of a convolutional neural network and a bidirectional long short-term memory (CNN-BiLSTM) in a supervised manner. The convolutional neural network model has proven effective in many applications. LSTM (long short-term memory) is a recurrent neural network architecture that has been adopted for time series forecasting. Convolutional layers on top of LSTM layers are added to capture local temporal changes. Bidirectional LSTM (BLSTM) processes inputs in two ways, once from past to future and once from future to past. Hence, it preserves information from both past and future. That is why a CNN-BLSTM model is used here to learn hearing preferences of a specific user.

Log Mel-spectrogram features of compressed audio pairs constitute the two inputs of the network and user feedbacks constitute the output of the network. The reward or hearing preference predictor provides a reward prediction $\hat{r}$ and produces the probability associated with preferring a compressed audio signal c^1 over another compressed audio signal c^2. For the prediction $\hat{r}$, the following cross-entropy loss function between the predicted reward and the actual user feedback is minimized:

$$loss\left(\hat{r}\right) = -\sum\nolimits_{(c^1,c^2,\mu)\epsilon D}\left(\mu\left(1\right)\log\hat{P}\left[c^1 > c^2\right] + \mu\left(2\right)\log\hat{P}\left[c^1 < c^2\right]\right) \quad (5)$$

Learning preferences and predicting reward from comparison pairs poses an implementation difficulty as a comparison pair does not provide a numeric feedback. To estimate the agent reward, one needs to estimate it from an intermediate model or network. The network structure of the developed hearing preference predictor is depicted in Fig. 4. Fig. 4(b) shows the overall structure of the network used for training the reward predictor based on the dataset of pair comparisons. Batch normalization [40] is applied to the convolutional layers using a decay rate of 0.90 together with a dropout with alpha = 0.5. The main purpose of the dropout is to prevent the network from overfitting. The dropout decorrelates the

weights of the hidden layers by randomly setting some hidden units to zeros at each training update step.

During the training phase, the model is trained on a batch size of 64, and optimized using the Adam algorithm [41]. Furthermore, during the training phase, early stopping and adaptive learning rate are applied to further avoid overfitting. Once the reward predictor is trained, the intermediate model or shared network as depicted in Fig. 4 (c) is used for agent training. Due to the fact that DRL is sensitive to the reward scale, a sigmoid layer is added at the end of the shared reward predictor model to bring the predicted reward between 0 and 1, see Fig. 4 (c).

E. RL AGENT

The training for a RL policy π is carried out based on the Bellman equation [42] in which at each time step t, the RL policy provides an action a_t for a given state s_t as expressed below

$$\pi : S \rightarrow A$$
$$a_t = \pi\left(s_t\right) \tag{6}$$

Action a_t influences the future state of the agent. The success of RL in learning the policy is reflected in the reward and the goal of RL is to maximize the overall reward. The parameters of the policy can get updated based on deep Q-learning [42], which is shown to be an effective RL training for personalization purposes [43], to maximize the overall estimated reward $\hat{r}$. Here, Q-value in Q-learning is optimized by a convolutional neural network. Q-value at a time step j is computed as follows:

$$y_j \longleftarrow r\left(s_j, a_j\right) + \gamma max_{a'} Q_\phi\left(s'_j, a'_j\right) \tag{7}$$

where $r\left(s, a\right)$ denotes the reward of a state and an action, and γ is a discount factor. A Q-value is basically a prediction of the future reward which allows selecting a next action for a given state. To convert the output values of the CNN into action probabilities, the so-called one-hot representation is utilized. As noted below, the action with the highest probability is then selected

$$a_j = \arg max_a Q_\phi\left(s_j, a_j\right) \tag{8}$$

The loss function in this CNN-based Q-learning is the mean-square-error (MSE) and the optimization is done by using the Adam algorithm, resulting in the CNN weights to get updated as follows:

$$\phi \longleftarrow \arg min_\phi \left\| Q_\phi\left(s_j, a_j\right) - y_j \right\|^2 \tag{9}$$

It is important to note that in contrast to supervised learning in which targets are fixed before training, here targets of the CNN-based agent depend on the network's weights that get updated gradually.

Before starting to train the RL agent and in order to reduce the chance of training a bad policy based on an untrained reward predictor, the reward predictor is trained with the dataset D of user preferences (mentioned earlier in subsection D). This means that the training of the reward predictor is performed asynchronously with respect to the DRL agent. For example, 200 comparison pairs can be conducted by the user at the beginning of the DRL training. Then, querying of the user feedback can be done every M time steps (see Table 1).

For the CNN $Q_\phi\left(s_j, a_j\right)$ model, a similar configuration used in the Atari experiment in [35] is utilized here. In this work, $80 \times 80 \times 3$ stacked log Mel-spectrogram images of an unprocessed audio segment are used as the input to the policy. The policy model consists of 3 convolutional layers having 32, 64, and 128 filters, respectively, with rectified linear unit (ReLU) activation and alpha $= 0.01$. Then, the flattened output of the last convolution layer is concatenated with the compression ratio adjustments (CR_{adj}) of the previous time step.

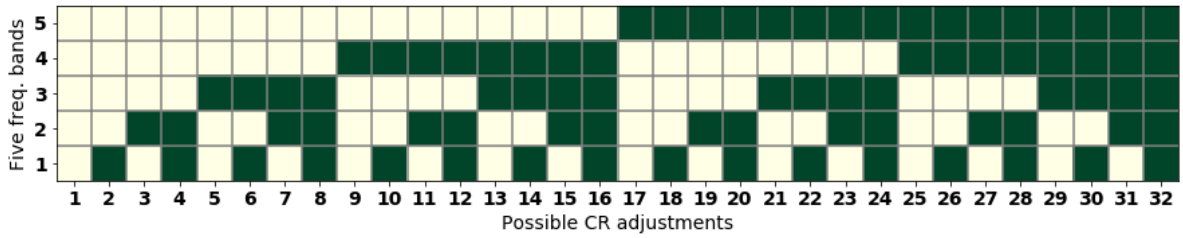
This is followed by two fully-connected layers of size 256 with ReLU activation, and a fully-connected layer with a size equal to the action space size. A fraction of the dataset is used as validation data to avoid overfitting. The agent is trained for 300 episodes, each containing 20 agent time steps. The trained reward model is fixed during the agent training. The value and description of parameters associated with the agent training are summarized in Table 2.

TABLE 2. Parameters associated with agent training.

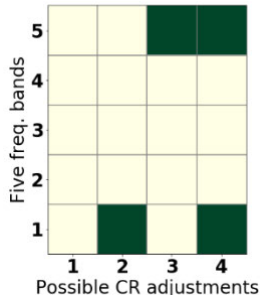
Parameter	Value	Description
$N_{Episode}$	300	Number of episodes for training
N_{Steps}	20	Number of steps in an episode
Training frequency	20	Number of steps to train agent
Batch size	50	Number of training observations used in one iteration
γ	0.99	Discount factor in updating Q-learning
No-op	30	Number of time steps before starting to train the agent

In this approach, non-numerical feedback rather than absolute feedback is obtained from a user. The goal is to learn a policy that is most consistent with the user's preferences. As a result, personalization emulates the intention of the user and finds a policy that is ideally consistent with it.

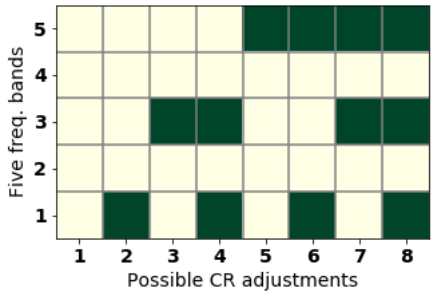
The above learning can be viewed as an active learning approach for achieving personalized compression. This approach has the advantage of being able to get trained in an online manner, thus allowing its utilization in the field or in real-world audio environments. The described approach constitutes the first attempt at personalizing compression via DRL by placing a user in the process of learning hearing preferences as reward. The key attribute of the developed personalization compression is that it is capable of modeling the non-linearities of a user's hearing preferences and dealing with noisy preference feedbacks, and thus improving over time by fine-tuning the model based on more preferences and new encountered audio environments.



(a) Adjustments in all five frequency bands that is mapped to 32 actions



(b) Adjustments only in the first and the fifth frequency bands



(c) Adjustments only in the first, the third and the fifth frequency bands

FIGURE 5. Mosaic representation of action space corresponding to simulated (a) users 1, 2, and 3, (b) simulated user 4, and (c) simulated user 5, with $\beta = 2$ ($CR_{adj} = 1$ *or* 4) and compression in five frequency bands. Light color indicates $CR_{adj} = 1$ and dark color indicates $CR_{adj} = 4$.

III. EXPERIMENTAL RESULTS

Two sets of experiments were conducted to examine the performance of the developed personalized DRL compression. The first set included simulations of the HITL deep reinforcement learning. In the second set of experiments, five adult human participants with bilateral, mild to moderate hearing loss were tested. All human subject testing was performed under an approved IRB (Institutional Review Board) protocol at the University of Texas at Dallas. In the two subsections that follow, these two sets of experiments are described. The purpose of the simulation experiments was to show the capability of the developed personalized approach to learn hearing preferences towards generating compression ratios that best matched a specific setting or user. In addition, the results of the personalized compression for five participants with hearing loss are reported.

A. SIMULATION EXPERIMENTS

In the first set of experiments, hearing preference scenarios were simulated, and the outcomes were examined to see the learning capability of the developed personalized DRL compression. The simulated experiments refer to simulating users with different hearing preferences to evaluate the effect of β value, user's feedback, and error in the training of the reward predictor and agent. In order to analyze the training performance of the DRL compression, simulations allow more control over preferences. Hence before testing the framework on subjects, five different hearing preference scenarios were simulated by using sets of if-then conditions. As described earlier, when the size of the action space is grown by increasing the number of frequency bands or the number of scalesβ, the personalization consequently demands more iterations, which poses difficulties for its real-world deployment. For simulated hearing aid users 1, 2, and 3, adjustments in all five frequency bands were considered and for simulated hearing aid users 4 and 5 only in two and three frequency bands (first, third, and fifth bands), were considered, respectively. To visualize the permutation and possible compression ratio (CR) adjustments for simulated users, a mosaic representation also known as Marimekko diagram is displayed in Fig.5. Each column represents one possible compression ratio adjustment in the action space A.

For all simulated hearing aid users, the same audiogram or the same prescription gains from [36] were used. Then, the DSL-v5 prescriptive gains expressed in nine frequency bands were mapped to five bands to reduce the computational complexity. The developed DRL methodology is general purpose in the sense that it can be applied to any number of bands. The number of bands that are commonly used are five, seven, and nine. Naturally, more training time with a subject in the loop would be needed as the number of bands is increased since the action space becomes larger.

Five frequency bands used were: [0-500] Hz, [500-1000] Hz, [1.0-2.0] kHz, [2.0-4.0] kHz, and [4.0-6.0] kHz. The compression ratios (gain changes) were computed from the gains. The attack time (time it takes to respond to higher sound levels) and the release time (time it takes to respond to lower sound levels) were set to the typical values of 0.01s and 1.0s, respectively. Basically, the attack and release time regulate the reaction pace of compression. Moderate and loud compression thresholds were also set to

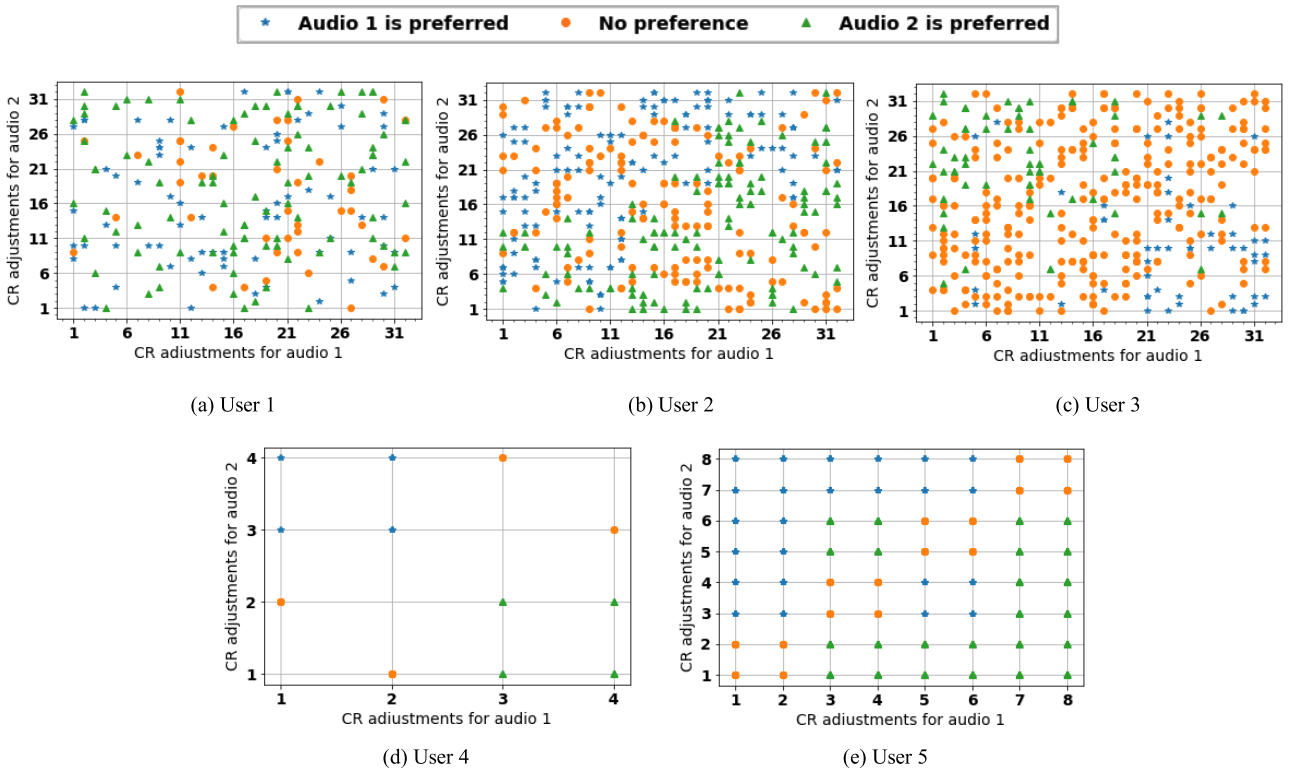


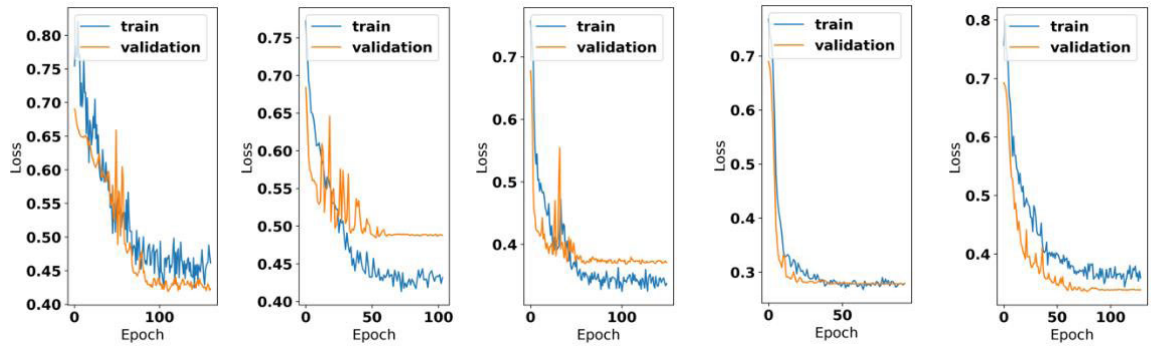
FIGURE 6. **Preference space collected from simulated (a) user 1, (b) user 2, (c) user 3, (d) user 4, and (e) from simulated user 5.**

60dB and 80dB to be consistent with the prescriptive hearing aid compression fittings.

At each training time step, based on the agent's input or state, the agent outputs one of the CR adjustments as an action in A to the environment. Each action corresponds to compression setting adjustments in the five frequency bands. Action space A of simulated users is shown in Fig. 5 via a mosaic representation for two possible CR adjustments, $\beta = 2$ ($CR_{adj} = 1$ *or* 4). The column in this figure shows all possible combinations of CR adjustments across the five frequency bands. As a result, the action space of the first three simulated users became as depicted in Fig. 5(a), exhibiting 32 possible compression ratio adjustments across all five frequency bands ($A = \prod_i^5 2 = 32$). Fig. 5(b) depicts the action space of a simpler case by changing the compression ratios in only the first and the fifth frequency bands, exhibiting 4 possible CR adjustments. Likewise, the action space for the case of adjustments in only three frequency bands (first, third, and fifth) is depicted in Fig. 5(c). In Fig. 5, light color indicates $CR_{adj} = 1$ (no adjustment) and dark color indicates $CR_{adj} = 4$ (users often prefer higher compression ratio in a noisy environment). As an example, in Fig. 5(a), the second column in Fig. 5(a) mosaic plot exhibits the CR adjustment settings as [4,1,1,1,1], and for the 20th column, the CR adjustment settings as [4,4,1,1,4]. In other words, based on the agent's input, the agent can give one of the 32 actions in A, each one corresponding to a compression setting adjustments in five frequency bands. As can be seen from Fig. 5, the action space of the agent is influenced by two factors: number of frequency bands and number of adjustments in each frequency band.

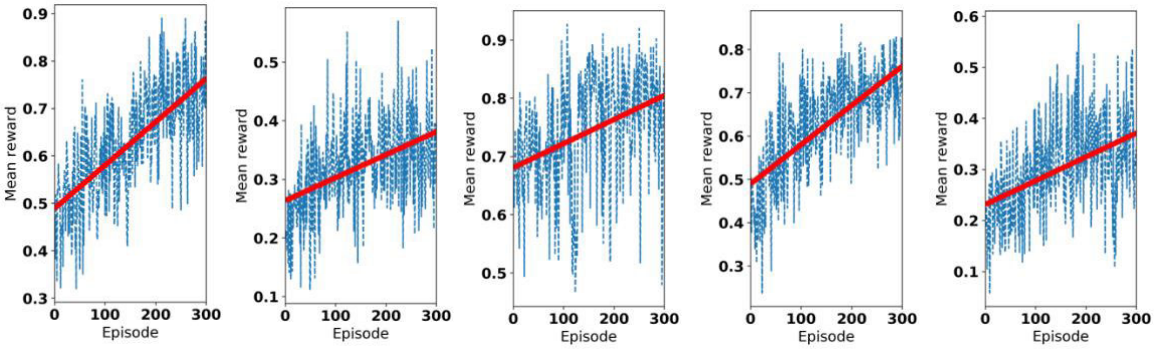
Overall, 200 hearing preferences over audio pairs were considered for each simulated user. As illustrated in Fig. 6, when the action space became larger, more user feedbacks were required to model user's hearing preferences. Preference space is displayed in Fig. 6 not only show the complexity of hearing feedbacks in larger action spaces, but also differences in preferences between the users 1, 2, and 3. To make the simulation close to reality, the following conditions for the simulated users were considered (users 1 to 3 are shown in Figs. 6a to 6c, respectively): (1) receiving noisy feedback from user, (2) receiving inconsistent and highly noisy feedback from user, (3) receiving neutral preferences across various compression settings from user. The simulated user 2 is an example of the situation when an actual user does not give proper feedback preferences when listening to audio pairs of two sets of compression ratios. This led to failure in learning preferences during the reward predictor training which consequently led to failure in the agent learning or learning the best settings for that user. The simulated user 3 is an example of the situation when an actual user is very strict about some settings and has neutral preferences over the other settings. This led to having more neutral preferences and therefore the training dataset became highly unbalanced.

The learning loss value of hearing preferences with respect to training epochs in different scenarios is displayed in Fig 7. By comparing Fig. 7(a) with Fig. 7(d), it can be seen that the



(a)User 1 (b) User 2 (c) User 3 (d) User 4 (e) User 5

FIGURE 7. Cross-entropy loss value in training reward predictor for simulated (a) user 1, (b) user 2, (c) user 3, (d) user 4, and for (e) user 5.



(a)User 1 (b) User 2 (c) User 3 (d) User 4 (e) User 5

FIGURE 8. Mean of normalized reward per episode and its trend (in solid red line) in the agent training for simulated (a) user 1, (b) user 2, (c) user 3, (d) user 4, and for (e) simulated user 5.

learning process in the reward predictor training became more challenging as the action space became larger. As can be seen from Fig. 7, simulated users 2 and 3 have worse validation loss in training the reward predictor. Failure in preference learning for simulated user 2 is due to inconsistent feedbacks from the user. For simulated user 3, due to having highly imbalanced data (more neutral preferences), failure occurs in the training of the reward predictor.

The mean of the normalized reward and the mean of the Q-values (target outputs) across the agent training episodes for each simulated user are displayed in Figs. 8 and 9, respectively. From these figures, it can be seen that both the mean reward and the mean Q-value exhibited an increasing trend, indicating that the personalized compression was gradually learning the policy that was ideally consistent with the users' hearing preferences. As mentioned earlier, the success of RL is heavily dependent on the performance of the reward predictor. That is why, although an increasing trend for user 2 is exhibited in Fig. 8(b) and Fig. 9(b), the personalization was not effective due to the poor training of the reward predictor.

B. SUBJECT TESTING EXPERIMENTS

In addition to the above simulations, actual human subject testing was performed according to the IRB protocol described earlier with one modification. Due to the Covid-19 pandemic, the original IRB was modified to allow participant testing to be conducted online instead of in a soundbooth. For the subject testing experiments, ''virtual'' visits were conducted using a video conference utility. Secure links were emailed to the participants to access online experimental sessions.

For subject testing, a crowdsourcing approach similar to the P.808 standard [44] was considered. It was ensured that the subjects listened to different audio pairs many times in a random order for a trustworthy comparison between the DSL-v5 settings and the personalized DRL settings. Eligibility conditions of the participants in our approved IRB included: (i) range of hearing loss of participants being mild to moderate, (ii) participants being native English speakers or presenting a native-level fluency of English, (iii) participants having symmetric hearing loss, and (iv) age range of participants being in the range 18-80 years old.

Initially, the participants obtained their audiograms using the web-based hearing test at https://hearingtest.online/. As per Table 1, for training the developed deep reinforcement learning personalized compression, 210 (7 sessions of 30, with breaks in between) pairs of sound files consisting of the spoken sentences, discussed earlier, in noisy (babble) background were played at SNR of 0 dB.

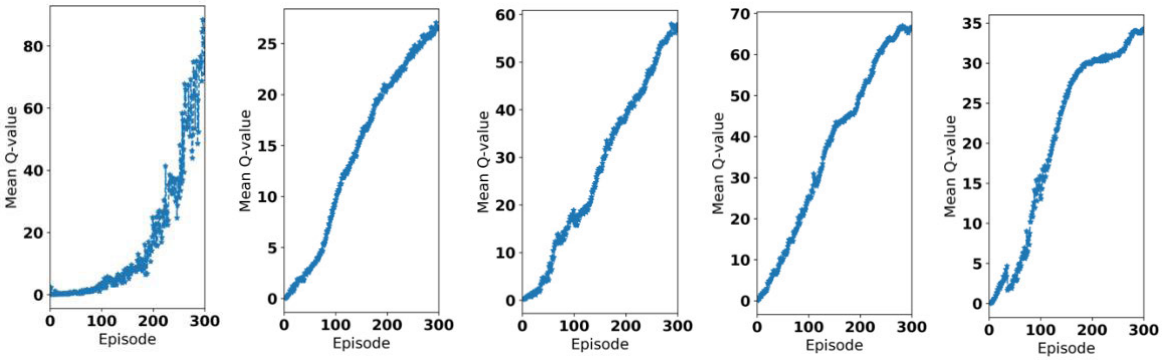


FIGURE 9. Mean Q-value per episode in the agent training for simulated (a) user 1, (b) user 2, (c) user 3, (d) user 4, and for (e) simulated user 5.

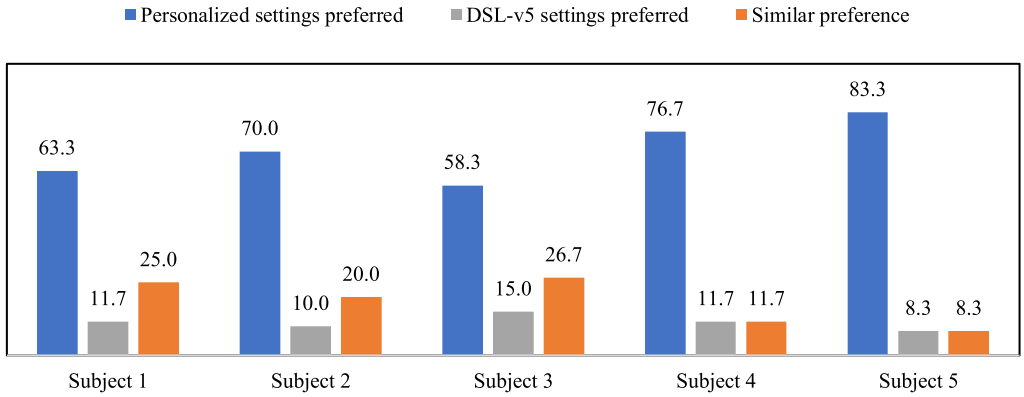


FIGURE 10. Outcome of subject testing experiments in percentages: comparison of hearing preference between personalized compression and DSL-v5 compression.

The participants were asked to indicate which sound file or clip they preferred or whether both sound clips sounded the same to them. It is worth mentioning here that increasing the number of sound files naturally improves the training and 210 audio sound files may not be adequate to cover all possible combinations of compression settings. For example, for $\beta = 2$, the number of possible actions is $2^5 = 32$, demanding $\binom{32}{2} = 496$ pairs of sound files. However, to avoid human fatigue, the test sessions were limited to 2 hours. This time frame constrained the number of audio sound files to 210 over 7 sessions. The data augmentation mentioned earlier was then applied to the collected dataset D of triplets (c^1, c^2, μ). The reward predictor was trained based on the augmented data to learn a participant's hearing preferences.

After training the policy, a comparison test was conducted between the personalized compression and the DSL-v5 reference prescriptive compression by playing 60 randomly selected sentences across different talkers in a noisy (babble) background at the same SNR level of 0dB. To remove any bias associated with the timeline of the training and testing phases, a gap of at least one-week was placed between these phases. Note that the training is carried out offline and once the offline training is completed, the actual operation/testing of the trained DRL compression only takes 71ms for a 2.5 seconds noisy speech sentence using a 2.9 GHz dual-core i5 processor computer. Thus, for all practical purposes, the developed personalized DRL compression runs in real-time during operation or testing. The "preference metric" used here is an integrated metric to enable personalization as it incorporates speech quality, speech intelligibility or word error rate, and audio comfort at the same time in a collective manner. In other words, when users judge audio pairs, they consider all these metrics together. For example, one user may prefer or prioritize speech intelligibility over speech quality and one user may prioritize speech quality over speech intelligibility.

Table 3 provides the compression ratios of DSL-v5 versus the compression ratios of the developed personalized approach for five participants with mild to moderate bilateral hearing loss who took part in this study. The outcomes of the participant testing experiments in terms of preference percentages are shown as a bar chart in Fig. 10. In this figure, the "personalized settings preferred" implies that in the audio pair assessment indicated in step 15 of Table 1, the subject preferred the audio that was compressed by personalized DRL compression settings. Likewise, the "DSL-v5 settings

TABLE 3. Subject testing experiments: DSL-v5 vs. personalized compression ratios.

Subject	Level of hearing loss	Audiogram in freq. bands [0.5, 1.0, 2.0, 4.0, 6.0] kHz	DSL-v5 gains for soft speech	DSL-v5 compression ratios	Personalized compression ratios
1	Mild	[15, 20, 20, 30, 30]	[7, 8, 14, 17, 15]	[1.1, 1.2, 1.3, 1.2, 1.3]	[1.1, 1.2, 1.3, **4.8**, **5.2**]
2	Mild	[15, 15, 20, 20, 30]	[5, 6, 14, 15, 15]	[1.1, 1.2, 1.3, 1.2, 1.2]	[**4.4**, 1.2, **5.2**, 1.2, **4.8**]
3	Moderate	[20, 20, 40, 50, 60]	[11, 12, 24, 29, 34]	[1.1, 1.2, 1.3, 1.2, 1.4]	[**4.4**, 1.2, 1.3, **4.8**, **5.6**]
4	Moderate	[25, 20, 20, 40, 30]	[13, 11, 14, 22, 15]	[1.1, 1.3, 1.3, 1.3, 1.3]	[**4.4**, 1.3, 1.3, **5.2**, 1.3]
5	Moderate	[20, 20 ,30 ,40 ,40]	[6, 11, 20, 23, 20]	[1.1, 1.2, 1.3, 1.2, 1.4]	[1.1, 1.2, 1.3, **4.8**, **5.6**]

preferred'' refers to when the subject preferred the audio compressed by the baseline or reference prescriptive DSL-v5. Similar preference denotes that the subject had equal preference over the audio compressed with the personalized DRL compression settings, and the audio compressed by the reference prescriptive DSL-v5 compression settings.

As can be seen from this figure, on average, personalized settings were clearly preferred by the participants over the DSL-v5 settings across different talkers and sentences heard. In other words, the number of times the personalized settings were preferred by the participants were nearly 7 times greater than the number of times the DSL-v5 settings were preferred. These results indicate that the developed personalized or individualized compression indeed is more effective than a one-size-fits-all DSL-v5 prescriptive compression approach. Audio samples of the subject testing experiments can be heard at this link: www.utdallas.edu/~kehtar/DRLcompression.html.

IV. CONCLUSION AND FUTURE WORK

In this paper, an active human-in-the-loop DRL-based personalized hearing aid fitting approach is developed to improve the currently practiced one-size-fits-all hearing aid fitting. The current fitting practice involves setting compression gains based on gain averages of a group of users which are not necessarily optimal for a specific user. The developed approach personalizes compression settings via a deep reinforcement learning framework. This is the first time human-in-the-loop DRL has been used to achieve improved hearing aid compression. Both simulation and experimental results show the effectiveness of the developed personalization approach in achieving preferred hearing outcomes.

The overall goal of this paper was to leverage simulation with limited clinical subject testing, to show the proof-of-concept of our novel personalized compression using HITL DRL approach. In future studies, the deployment and efficacy of our approach can be further assessed by carrying out extensive clinical testing. This would require examining a large number of subjects in controlled audio environments and in the field. It would also be useful to examine several noise types, SNRs, numbers of frequency bands, and alternative metrics other than the preference metric used in this work. One key advantage of the introduced personalization approach is that it is general purpose in the sense that all of these parameters are already built into its training/testing and additional parameters can be added to explore an even larger variable space.

ACKNOWLEDGMENT

The authors wish to express our appreciation to Ms. T. Campbell and Dr. C. Escabi for their help in putting together the IRB materials and their assistance with the subject testing. They also thank Mr. A. Salman for his help with the codes written for this project.

REFERENCES

[1] D. Giannoulis, M. Massberg, and J. D. Reiss, "Digital dynamic range compressor design—A tutorial and analysis," *J. Audio Eng. Soc.*, vol. 60, no. 6, pp. 399–408, 2012.

[2] *NAL-NL2, National Acoustic Laboratories*. Accessed: Jul. 28, 2014. [Online]. Available: http://www.nal.gov.au/nal-software_tab_nal-nl-2.shtml

[3] *Western University, DSL-v5 by Hand*. Accessed: 2014. [Online]. Available: https://www.dslio.com/wpcontent/uploads/2014/06/DSL-5-by-Hand.pdf

[4] S. Kochkin, D. L. Beck, L. A. Christensen, C. Compton-Conley, B. J. Fligor, P. B. Kricos, J. B. Mcspaden, H. G. Mueller, M. J. Nilsson, J. L. Northern, T. A. Powers, R. W. Sweetow, B. Taylor, and R. G. T. MarkeTrak, VIII, "The impact of the hearing healthcare professional on the hearing aid user success," *The Hearing Rev.*, vol. 17, no. 4, pp. 12–34, 2010.

[5] G. Keidser and H. Dillon, "What's new in prescriptive fittings down under," in *Hearing Care for Adults*, C. V. Palmer and R. Seewald, Eds. Stäfa, Switzerland: Phonak AG, Nov. 2006, pp. 133–142.

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

[7] K. Smeds, "Is normal or less than normal overall loudness preferred by first-time hearing aid users?" *Ear Hearing*, vol. 25, no. 2, pp. 159–172, Apr. 2004.

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

[9] B. Johansen, M. Petersen, M. Korzepa, J. Larsen, N. Pontoppidan, and J. Larsen, "Personalizing the fitting of hearing aids by learning contextual preferences from Internet of Things data," *Computers*, vol. 7, no. 1, p. 1, Dec. 2017.

[10] H. Dillon, J. A. Zakis, H. McDermott, G. Keidser, W. Dreschler, and E. Convery, "The trainable hearing aid: What will it do for clients and clinicians?" *Hearing J.*, vol. 59, no. 4, p. 30, Apr. 2006.

[11] 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, Oct. 2016.

[12] A. Pasta, M. Petersen, K. Jensen, and J. Larsen, "Rethinking hearing aids as recommender systems," *Proc. HealthRecSys*, 2019, pp. 11–17.

[13] D. Cuda, A. Murri, A. Mainardi, and J. Chalupper, "Effectiveness and efficiency of a dedicated bimodal fitting formula," *Audiol. Res.*, vol. 9, no. 1, pp. 6–9, May 2019.

[14] L. Brody, Y.-H. Wu, and E. Stangl, "A comparison of personal sound amplification products and hearing aids in ecologically relevant test environments," *Amer. J. Audiol.*, vol. 27, no. 4, pp. 581–593, Dec. 2018.

[15] 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.

[16] A. Boothroyd and C. Mackersie, "A 'Goldilocks' approach to hearing-aid self-fitting: User interactions," *Amer. J. Audiol.*, vol. 26, no. 3S, pp. 430–435, 2017.
[17] G. Keidser and E. Convery, "Outcomes with a self-fitting hearing aid," *Trends Hearing*, vol. 22, Jan. 2018, Art. no. 233121651876895.
[18] E. Convery, G. Keidser, L. Hickson, and C. Meyer, "Factors associated with successful setup of a self-fitting hearing aid and the need for personalized support," *Ear Hearing*, vol. 40, no. 4, pp. 794–804, 2019.
[19] C. L. Mackersie, A. Boothroyd, and H. Garudadri, "Hearing aid self-adjustment: Effects of formal speech-perception test and noise," *Trends Hearing*, vol. 24, pp. 1–16, Jun. 2020.
[20] A. T. Sabin, D. Van Tasell, B. Rabinowitz, and S. Dhar, "Validation of a self-fitting method for over-the-counter hearing aids," *Trends Hearing*, vol. 24, pp. 1–19, Jan. 2020.
[21] 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, USA, Dec. 2013, pp. 5–10.
[22] J. Brehm Bagger Nielsen, J. Nielsen, and J. Larsen, "Perception-based personalization of hearing aids using Gaussian processes and active learning," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 1, pp. 162–173, Jan. 2015.
[23] 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 Hearing*, vol. 23, May 2019, Art. no. 2331216519847413.
[24] J. R. Dubno, D. D. Dirks, and D. E. Morgan, "Effects of age and mild hearing loss on speech recognition in noise," *J. Acoust. Soc. Amer.*, vol. 76, no. 1, pp. 87–96, Jul. 1984.
[25] M. Korzepa, M. K. Petersen, J. E. Larsen, and M. Mørup, "Simulation environment for guiding the design of contextual personalization systems in the context of hearing aids," in *Proc. Adjunct Publication 28th ACM Conf. User Model., Adaptation Personalization*, Jul. 2020, pp. 293–298.
[26] C. Szepesväri, "Algorithms for Reinforcement Learning," *Synth. Lectures Artif. Intell. Mach. Learn.*, vol. 4, no. 1, pp. 1–103, 2010.
[27] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*, Aug. 2017, pp. 1–6.
[28] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
[29] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application*. Hoboken, NJ, USA: Wiley, 2007.
[30] G. R. Lockhead, "Absolute judgments are relative: A reinterpretation of some psychophysical ideas," *Rev. Gen. Psychol.*, vol. 8, no. 4, pp. 265–272, Dec. 2004.
[31] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," 2013, *arXiv:1312.5602*. [Online]. Available: http://arxiv.org/abs/1312.5602
[32] S. Gu, E. Holly, T. Lillicrap, and S. Levine, "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 3389–3396.
[33] A. Y. Ng and S. J. Russell, "Algorithms for inverse reinforcement learning," in *Proc. ICML*, vol. 1, 2000, pp. 663–670.
[34] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4299–4307.
[35] B. Ibarz, J. Leike, T. Pohlen, G. Irving, S. Legg, and D. Amodei, "Reward learning from human preferences and demonstrations in Atari," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 8011–8023.
[36] N. Alamdari, E. Lobarinas, and N. Kehtarnavaz, "An educational tool for hearing aid compression fitting via a Web-based adjusted smartphone app," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7650–7654.
[37] D. McCloy, P. Souza, R. Wright, J. Haywood, N. Gehani, and S. Rudolph. *The PN/NC Corpus. Version 1.0*. Accessed: 2013. [Online]. Available: https://depts.washington.edu/phonlab/resources/pnnc/pnnc1/
[38] J. Abeßer, "A review of deep learning based methods for acoustic scene classification," *Appl. Sci.*, vol. 10, no. 6, p. 2020, Mar. 2020.
[39] S. S. Stevens, J. Volkmann, and E. B. Newman, "A scale for the measurement of the psychological magnitude pitch," *J. Acoust. Soc. Amer.*, vol. 8, no. 3, pp. 185–190, Jan. 1937.
[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn*, 2015, pp. 448–456.
[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[42] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[43] S. Liu, K. C. See, K. Y. Ngiam, L. A. Celi, X. Sun, and M. Feng, "Reinforcement learning for clinical decision support in critical care: Comprehensive review," *J. Med. Internet Res.*, vol. 22, no. 7, Jul. 2020, Art. no. e18477.
[44] *Subjective Evaluation of Speech Quality With a Crowdsourcing Approach*, document ITU-T Recommendation P.808, International Telecommunication Union, Geneva, Switzerland, 2018.

NASIM ALAMDARI (Student Member, IEEE) received the M.S. degree in electrical engineering from the University of North Dakota, ND, in 2016. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, The University of Texas at Dallas. Her research interests include real-time audio processing, speech enhancement, preference learning, and machine learning.

EDWARD LOBARINAS received the B.S. degree from Rutgers University and the M.A. and Ph.D. degrees from the State University of New York at Buffalo. His research areas 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. He has been recognized by the American Academy of Audiology as a Jerger Future Leader of Audiology and has received grants from the National Institute of Health, the American Tinnitus Association, and the Tinnitus Research Initiative.

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 at The University of Texas at Dallas, Richardson, TX. His research interests include signal and image processing, machine learning, deep learning, and real-time implementation on embedded processors. He has authored or coauthored ten books and over 400 journal articles, conference papers, patents, manuals, and editorials in these areas. He is a Fellow of SPIE, a licensed Professional Engineer, and Editor-in-Chief of *Journal of Real-Time Image Processing*.