Check for updates

Evaluating and optimizing hearing-aid self-fitting methods using population coverage

Dhruv Vyas[1]*, Erik Jorgensen[2], Yu-Hsiang Wu[3] and
Octav Chipara[1]

[1]Department of Computer Science, University of Iowa, Iowa City, IA, United States, [2]Department of
Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, WI, United States,
[3]Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA, United States

Adults with mild-to-moderate hearing loss can use over-the-counter hearing aids to treat their hearing loss at a fraction of traditional hearing care costs. These products incorporate self-fitting methods that allow end-users to configure their hearing aids without the help of an audiologist. A self-fitting method helps users configure the gain-frequency responses that control the amplification for each frequency band of the incoming sound. This paper considers how to guide the design of self-fitting methods by evaluating certain aspects of their design using computational tools before performing user studies. Most existing fitting methods provide various user interfaces to allow users to select a configuration from a predetermined set of presets. Accordingly, it is essential for the presets to meet the hearing needs of a large fraction of users who suffer from varying degrees of hearing loss and have unique hearing preferences. To this end, we propose a novel metric for evaluating the effectiveness of preset-based approaches by computing their population coverage. The population coverage estimates the fraction of users for which a self-fitting method can find a configuration they prefer. A unique aspect of our approach is a probabilistic model that captures how a user's unique preferences differ from other users with similar hearing loss. Next, we propose methods for building preset-based and slider-based self-fitting methods that maximize the population coverage. Simulation results demonstrate that the proposed algorithms can effectively select a small number of presets that provide higher population coverage than clustering-based approaches. Moreover, we may use our algorithms to configure the number of increments of slider-based methods. We expect that the computational tools presented in this article will help reduce the cost of developing new self-fitting methods by allowing researchers to evaluate population coverage before performing user studies.

KEYWORDS

audiology, hearing aids, genetic algorithm, hearing-aid self-fitting, over-the-counter hearing aids

1. Introduction

Hearing loss is an epidemic in the United States that is too often left untreated. The primary treatment for hearing loss is hearing aids (HAs). However, of the 48 million Americans with hearing loss, only 14-34% use HAs (Lin et al., 2011; Chien and Lin, 2012; Powers and Rogin, 2019). A major reason for the low rate of HA adoption is their high cost. One study estimated that the average cost of HAs bundled with several audiologist visits is $2,500, representing a significant financial expense for 77% of Americans (Jilla et al., 2020). Economic barriers to hearing healthcare disproportionately affect minorities; most HA users are affluent,

educated, and white (Nieman et al., 2016; McKee et al., 2019; Reed et al., 2021). Thus, there is a critical need to improve access to hearing care. One solution is the advent of over-the-counter (OTC) HAs. The OTC Hearing Aid Act, signed into law in 2017, with subsequent rules delivered in 2022, has enabled HAs to be sold over the counter without the need for the user to see an audiologist or medical professional. OTC hearing aids are designated for adults who perceive that they have mild-to-moderate hearing loss, even if they have not had a formal hearing evaluation from an audiologist. In theory, OTC hearing aids should be cheaper than prescription hearing aids. However, they require the user to self-fit the device. The primary function of HAs is to divide the incoming sound into several frequency bands and amplify each band preferentially. For HAs to address a user's needs it is essential to fit the HA by configuring the gain-frequency response (gains, henceforth) of each band to compensate for their hearing loss in that frequency band. Traditionally, HA fitting is performed by an audiologist, who first measures the user's hearing loss as an *audiogram*. Audiologists construct audiograms by presenting pure tone to measure the user's hearing thresholds at frequencies important for speech perception (typically 0.25–8 kHz). A user's hearing loss is characterized by their thresholds across the frequency range relative to the average hearing thresholds of normal-hearing listeners. The amount of gain applied in each band is traditionally determined using a prescription formula, commonly NAL-NL2 (Keidser et al., 2011). NAL-NL2 is based on theoretical models of speech intelligibility and loudness comfort, as well as empirical data showing differences in gain preferences between different population subgroups (e.g., men vs. women, experienced vs. new HA users). Since NAL-NL2 uses theoretical models and population-level statistics, the prescribed NAL-NL2 configuration estimates the average configuration for a sample of users with similar audiograms (and hearing loss). However, a user's preferred configuration may deviate significantly from their prescribed NAL-NL2 configuration due to individual perceptual, lifestyle, and HA usage factors that differ from person to person (Søgaard Jensen et al., 2019). To customize a user's configuration, a series of visits to the audiologist are generally required to fine-tune the HA's configuration based on their feedback. OTC HAs reduce cost, time, and other barriers to HA access by shifting the burden of configuring these devices from the audiologist to the end user. Traditional HA fitting requires sophisticated software, a programming interface, and specialized knowledge typically acquired through years of graduate-level training. For the lay user, successful self-fitting must be achieved without any of these helpful resources, which have traditionally been bundled with purchasing a HA. To address this limitation, OTC HAs usually provide a user interface that allows users to configure their gain. Two common strategies are used to design user interfaces—collection-based and slider-based approaches. Both approaches operate on a set of predetermined and fixed configurations that we will refer to as **presets**. In the case of collection-based methods, the presets are configurations associated with representative types of hearing loss. In contrast, slider-based methods directly manipulate various aspects of configurations. For example, a loudness (i.e., overall amplitude) slider increases the overall gain. The presets are the union of all the possible configurations that the controllers may

reach. It is common for collection-based methods to include a small number of presets, whereas slider-based methods typically include more presets. We provide additional details about these two methods, including concrete examples of their use with existing HAs in Section 2.

A key challenge associated with advancing the OTC fitting methods is that their design requires extensive and costly user studies. Similar challenges also occur when comparing different self-fitting approaches. These questions will become increasingly important as more OTC HAs are commercialized. Therefore, we pose the question of whether it is possible to create metrics that we can use to assess OTC HA fitting methods without resorting to extensive user studies. We propose a new metric—**population coverage**—to evaluate, compare, and optimize fitting methods. The population coverage estimates how well a set of presets $\mathcal{P}$ meets the needs of users with mild-to-moderate hearing loss, the population that would benefit from an OTC HA the most. We will develop statistical models to estimate the population coverage of a general self-fitting method given (1) statistics regarding the typical hearing loss of users in a population of interest and (2) the set of presets used in the fitting method. Moreover, we will show that it is possible to optimize various parameters of a fitting method to maximize population coverage.

A key benefit of population coverage is that it provides a computational method to evaluate the performance of self-fitting methods. In the remainder of the paper, we describe how to use it to drive the development of two popular self-fitting approaches that use presets and sliders. We expect population coverage to be a useful design tool for the development of future self-fitting methods by allowing designers to narrow down the large space of possible gain-frequency responses to those that would benefit a population of interest. Of course, any fitting method must be evaluated through rigorous empirical studies that go beyond population coverage.

Our work builds on a line of research in audiology that aims to identify representative audiograms (or equivalently different types of hearing loss). Ciletti and Flamme (2008) used audiograms provided from two publicly available datasets and applied cluster analysis to identify audiograms representative of the US population. The audiograms within a cluster were more similar than those in different clusters. The clustering analysis results may be used to generate a set of presets by converting the centroid audiogram of each cluster to a REAR configuration using NAL-NL2. A similar approach to generating presets was proposed in Jensen et al. (2020).

We extend these approaches in two significant directions. First, the above approaches assume that NAL-NL2 accurately predicts a user's most preferred configuration. Empirical evidence shows that this assumption is incorrect and yields suboptimal configurations to be selected as presets. We develop a statistical model that characterizes how a user's preference may deviate from their NAL-NL2 prescription (see Section 3). Second, our techniques are more general in that we can use them for both collection- and slider-based approach. We present algorithms to select presets for both collection- and slider-based approaches that maximize the population coverage. Finally, our results show our approach's superiority over standard cluster-based techniques in constructing presets. We note that the paper's novelty is using probabilistic models to drive the

development of self-fitting methods, which helps avoid extensive user studies.

The remainder of the paper is organized as follows. Section 2 summarizes prior work on self-fitting methods. A formal definition of population coverage and methods for computing it for preset-based self-fitting methods is described in Section 3. Algorithms to optimize the presets for collection- and slider-based approaches are included in Section 4. Results comparing different methods for generating presets are provided in Section 5, and their importance is discussed in Section 6. Section 7 provides conclusions and discusses future work.

2. Related work

2.1. Slider-based approaches

A common approach to allow users to personalize OTC HAs is to use sliders (or wheels) that manipulate the gains HAs used to amplify each channel. For example, the Ear Machine approach presents users with two wheels, one which sliders loudness and one which controls fine-tuning (Nelson et al., 2018). The loudness slider enables users to vary the overall gain, and compression parameters of the HA, and the fine-tuning wheel enables users to vary the tilt of the gains around a 2 kHz fulcrum. In another approach, called Goldilocks, the user is presented with three parameters they can adjust using up and down arrows: fullness (low-frequency cut), crispness (high-frequency boost), and loudness (overall amplification) (Boothroyd and Mackersie, 2017; Mackersie et al., 2019). The user first adjusts loudness, then crispness, then loudness again, then fullness, then any parameter until they find a gain that is "just right." It is important to note that these interfaces enable the user to only select from a *predetermined and fixed set of presets* by manipulating the sliders.

2.2. Collection-based approaches

Slider-based approaches can lead to successful self-fitting (Brody et al., 2018; Boothroyd et al., 2022). However, because these methods require the user to manipulate various parameters on continuous scales, the self-fitting process can be cumbersome and cognitively challenging. For example, six of the 26 users in the evaluation of the Goldilocks required help with the interface, and there was a wide range in time to completion among users, indicating that some users likely found the procedure more difficult than others (Boothroyd and Mackersie, 2017). Recent MarkeTrak survey data found up to half of potential OTC users would not be comfortable tuning their own HAs (Edwards, 2020). One way to simplify the self-fitting process is to use an unordered collection of presets (Sabin et al., 2020; Urbanski et al., 2021). A gain that can accommodate most users is predetermined in a collection-based approach. The self-fitting process then involves arriving at the user's most preferred preset. Presumably, the preset approach can both shorten the self-fitting process and make the procedure easier, reducing failures during the self-fitting process. Although some studies have investigated the use of presets in OTC HAs, it remains an open problem to determine the right set of presets and

the best strategy for identifying the user's preferred preset in an easy and efficient manner. In this paper, we focus on the former problem while not directly addressing the latter.

In Urbanski et al. (2021), presets were derived using audiograms of adults with mild-to-moderate hearing loss from an extensive national database. First, they found all possible gains based on the audiogram set. Then, they found the four presets that fit the largest number of NAL-NL2 targets within ± 5 dB from 0.25 to 4 kHz. The four presets could match NAL-NL2 targets within 10 dB for 70% of the audiograms in the database. The presets were then empirically tested by first assigning different presets to participants using various fitting methods and having participants complete speech testing using the presets selected in each method. Presets were determined either based on the participant's audiologist-administered audiogram, a self-test audiogram, listening to the different presets in quiet and in noise, a questionnaire, or random assignment. The presets determined by each selection method were generally different from one another. However, none of the fitting methods besides the questionnaire and random assignment methods resulted in speech perception scores that differed significantly from those obtained when participants were individually fit to NAL-NL2 targets. Taken together, the findings from that study suggest that the process by which presets are determined for individual listeners impacts what preset is chosen–but several different presets can yield comparable efficacy for speech perception for any given listener.

3. Coverage of self-fitting methods

In this section, we consider the problem of estimating the fraction of users with mild-to-moderate hearing loss whose hearing needs can be met by a set of N presets $\mathcal{P}$ (i.e., $|\mathcal{P}| = N$). To do so, we need to answer three questions:

1. How do we define the subset of users with mild-to-moderate hearing loss?
2. When are the hearing needs of a user met by a preset?
3. How do we estimate the fraction of users that are covered by a set of presets?

We will answer each of these questions in each of the subsequent subsections.

3.1. Target population

The population of users that would benefit the most from OTC HAs includes those with mild-to-moderate hearing loss. The National Health and Nutrition Examination Surveys (NHANES) (National Center for Health Statistics, 2022) from 1996-2016 includes audiograms characterizing common hearing loss configurations. For each user u, NHANES includes a **population weight** w_u representing the prevalence of that type of hearing loss in the population. We focus on a subset of older individuals with mild to moderate sensorineural hearing loss from NHANES who meet the following criteria: (1) They are between 55 and 85 years old (inclusive); (2) Their audiometric thresholds from 0.25 to 6 kHz are less than or equal to 75 dB HL, with a threshold

at 8 kHz that does not exceed 120 dB HL; (3) A four-frequency pure-tone average (0.5, 1, 2, and 4 kHz) greater than 20 dB HL but less than 50 dB HL (World Health Organization of mild-to-moderate hearing loss, Olusanya et al., 2019); (4) Normal middle ear function defined as peak tympanometric peak pressure greater than or equal to −50 daPa and less than or equal to 50 daPa, ear canal volume equal to or greater than 0.5 ml but less than or equal to 2 ml, and compliance equal to or greater than 0.3 ml and less than or equal to 1.5 ml. If only one ear qualified, only that ear is included in the dataset. Finally, all audiometric and tympanometric data is complete with no missing values. The selected subset included 1,979 audiograms (mean age = 67.77 years, SD = 8.02 years; 1,018 female, 961 male; 813 bilateral, 1,166 unilateral).

For each user u, we compute the prescribed **NAL-NL2 configuration** $\hat{c}_u$. The NAL-NL2 configuration is computed using the NAL-NL2 software (NAL-NL2, 2022), configured similar to Sabin et al. (2020) and Urbanski et al. (2021): (1) thresholds entered as air conduction; (2) broadband input signal level of 65 dB SPL; (3) 18-channels with multichannel compression limiting; (4) default directionality (0 degrees azimuth); (5) default microphone (head surface); (6) experienced HA users; (7) compression speed to dual; (8) non-tonal language. The user's gender and age are set according to the NHANES data. NHANES dataset includes audiograms for unilateral hearing loss (hearing loss in one ear) or bilateral hearing loss (hearing loss in both ears). People with bilateral hearing loss may choose to wear one or two hearing aids (Cox et al., 2011). Thus, for bilateral audiograms, we compute four sets of NAL-NL2 configurations: unilateral left, unilateral right, bilateral left, and bilateral right. Thus, we compute a total of 4418 NAL-NL2 configurations.

3.2. Population coverage

A configuration c is a six-dimensional vector ($c_i \in \mathbb{R}^6$) representing the Real-Ear-Insertion-Gains (REIGs) at frequencies 0.5, 1, 2, 3, 4, and 6 KHz. The notation $c[f]$ ($f \in \{0.5, 1, 2, 3, 4, 6\}$) refers to the gain associated with frequency f in the vector c. We will select a subset of N of configurations as presets. We define the **preset coverage** of a preset p_i to include all the configurations c located within the ball centered at p_i and with a radius R. The distance between a configuration c and p_i is the maximum absolute difference computed across the six frequencies:

$$B_R(p_i) = \{c \in \mathbb{R}^6 \mid \max |c - p_i| \leq R\} \tag{1}$$

We assume that all the configurations within the ball are covered by preset p_i. Audiologists are typically satisfied to find a configuration within ± 5 dB, so we set $R = 5$ dB.

The coverage of all presets in $\mathcal{P}$ is $\mathbf{C}(\mathcal{P})$:

$$\mathbf{C}(\mathcal{P}) = \bigcup_{p_i \in \mathcal{P}} B_R(p_i) \tag{2}$$

Ideally, the presets included in $\mathcal{P}$ maximize the fraction of users whose configurations are included in $C(\mathcal{P})$.

Next, we consider determining whether a user's configuration is covered. A naive approach would be to assume that a user's u

preferred configuration coincides with that prescribed by NAL-NL2 (as done in prior work, Jensen et al., 2020; Urbanski et al., 2021). Consistent with this assumption, a user would be covered if its prescribed NAL-NL2 configuration $\hat{c}_u$ would be included in $C(\mathcal{P})$. Then, the fraction of the population covered would be the sum of the population weights of covered users.

Emerging empirical evidence shows, however, that a user's preferred configuration often deviates from their NAL-NL2 prescription. In the following, we will propose a statistical model that characterizes how the users' preferences may deviate from the NAL-NL2 model. We observe that fitting rationales like NAL-NL2 accurately estimate the average configuration for a group of patients with similar hearing loss but do not account for the high variability within the group due to individualized preferences. Moreover, empirical evidence shows that a user's preferred configuration often differs from the NAL-NL2 configuration in either its "overall gain" and/or "slope" (Mackersie et al., 2020; Boothroyd et al., 2022). Therefore, the goal is to model how a user's preferred configuration may deviate from the NAL-NL2 prescription in overall gain and slope.
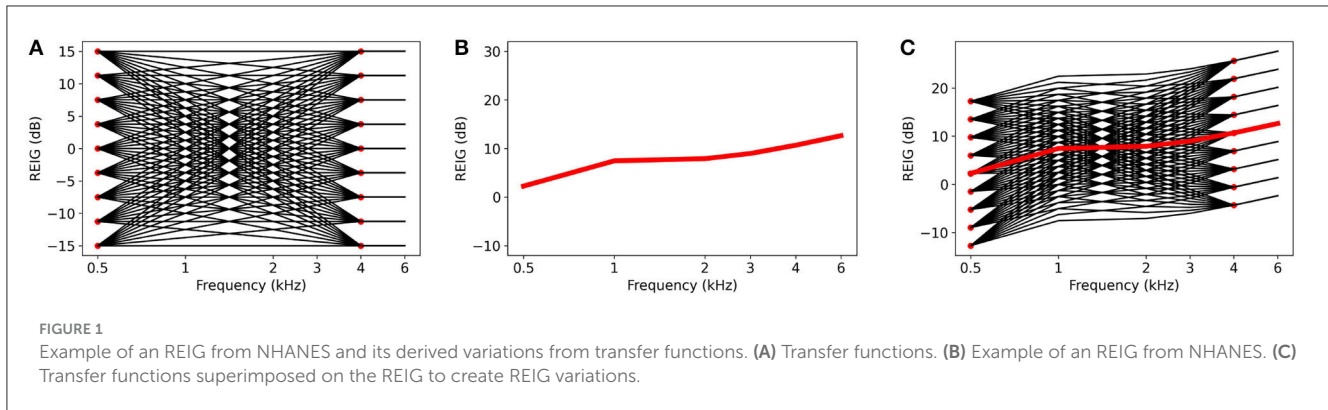
To this end, we introduce **transfer functions** (see Figure 1A). A transfer function is a log-linear function ($T_j : \mathbb{R} \to \mathbb{R}$) describing how a user's u preferred configuration c_u may deviate from their NAL-NL2 configuration $\hat{c}_u$ at f Hz.

$$c_{u,j}[f] = \hat{c}_u[f] + T_j(f) \tag{3}$$

We capture a user's u potential **preferred configurations** $c_{u,j}$ using several transfer functions T_j, each yielding changes in overall gain and slope.

We construct the transfer functions as follows. Empirical evidence suggests that the magnitude of the deviations at 0.5 and 4 KHz are within $\pm$ 15 dB (Mackersie et al., 2020; Boothroyd et al., 2022) and at remains unchanged at 6 KHz. We divide the range from −15 dB to 15 dB in increments of 3.75 dB and create an anchor point at each increment at frequencies of 0.5 and 4 KHz (see Figure 1A). We select increments of 3.75 dB as individuals cannot perceive the difference between configurations with smaller REIG differences (Caswell-Midwinter and Whitmer, 2019). Then, each transfer function is constructed by fitting a line for each pair of anchor points. By connecting each pair of anchor points from 0.5 to 4 KHz, we create 81 possible transfer functions. Accordingly, for a user u, there are 81 possible preferred configurations, 80 of which differ from the user's NAL-NL2 configuration. For example, Figure 1B plots the NAL-NL2 prescription for one of the users in NHANES. Figure 1C plots the user's preferred configurations computed by adding each transfer function shown (shown in Figure 1A) to the user's NAL-NL2 configurations (shown in Figure 1B) according to Equation (3).

The transfer functions are designed to cover the range of possible deviations from NAL-NL2. However, not all potential preferred configurations are equally likely — configurations that are closer to NAL-NL2 tend to be more likely. With each of the 81 possible configurations, we associate a likelihood that is determined by a 2D Gaussian parameterized by average deviation in the low (5 and 1 KHz) and high frequencies (2, 3, and 4 KHz). Studies Mackersie et al. (2020) and Boothroyd et al. (2022)

FIGURE 1
Example of an REIG from NHANES and its derived variations from transfer functions. **(A)** Transfer functions. **(B)** Example of an REIG from NHANES. **(C)** Transfer functions superimposed on the REIG to create REIG variations.

collected statistics about the deviations between the configuration prescribed by NAL-NL2 and those preferred by users. Using this data, we empirically determine the mean and standard deviation of the 2D Gaussian (see Figure 2). The weight $l_{u,j}$ of each potential preferred configuration is $c_{u,j}$ determined according to the fitted 2D Gaussian.

Recall that hearing loss can be either unilateral (i.e., only in one ear) or bilateral (i.e., in both ears). We define a user with unilateral hearing loss u to be covered if the sum of weights of the covered potential preferred configurations exceeds a threshold (γ). A preferred configuration $c_{u,j}$ is covered if it is included in $C(\mathcal{P})$.
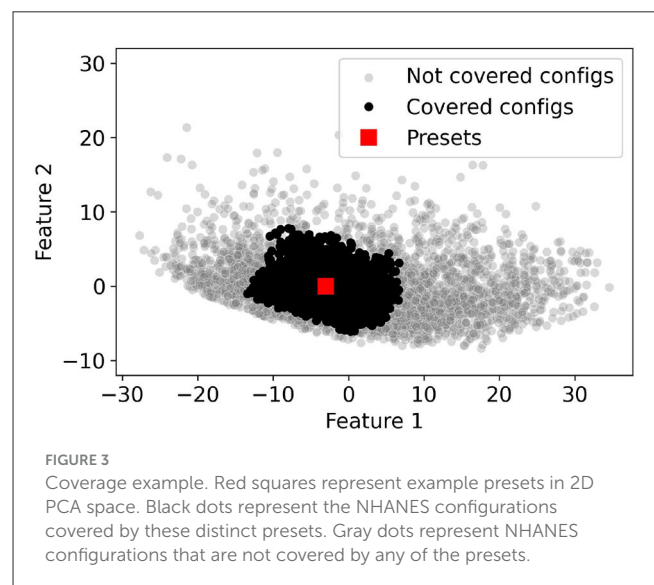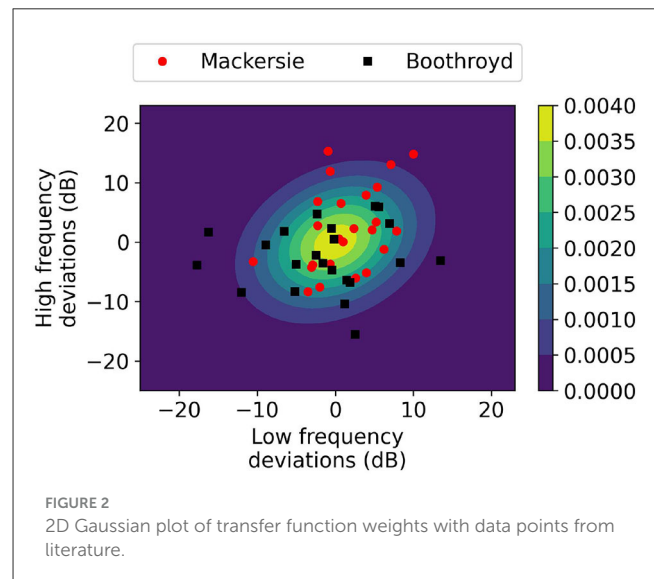
Computing the coverage for a user with bilateral hearing loss is a bit more involved since their hearing needs might be met in three different ways:

1. Use a single device worn on the left ear (uni-left).
2. Use a single device worn on the right ear (uni-right).
3. Use two devices, each device configured for the hearing loss in that ear (bi-left and bi-right).

We can evaluate whether each of the four configurations (uni-left, uni-right, bi-left, and bi-right) are covered according to the unilateral fitting criterion. We then say that a subject with bilateral hearing loss is covered if all four configurations are covered. The **population coverage** is then the sum of the population weights of the covered users. The details of the coverage computation are included in Algorithm 1.

3.3. Example

To illustrate the concept of coverage, we plot as black and gray dots the potential configurations of the considered users in Figure 3. The graph is obtained by reducing the dimensionality of the configuration to two dimensions using Principle Component Analysis (PCA). The red square in the figure indicates one of the selected presets. The black dots indicate the configurations that are covered by the preset i.e., they fall within the ball of radius $\pm$ 5 dB consistent with Equation (1).

FIGURE 2
2D Gaussian plot of transfer function weights with data points from literature.

FIGURE 3
Coverage example. Red squares represent example presets in 2D PCA space. Black dots represent the NHANES configurations covered by these distinct presets. Gray dots represent NHANES configurations that are not covered by any of the presets.
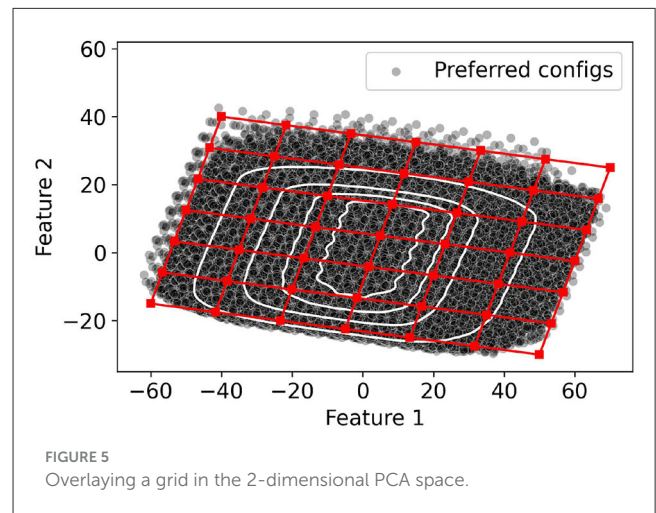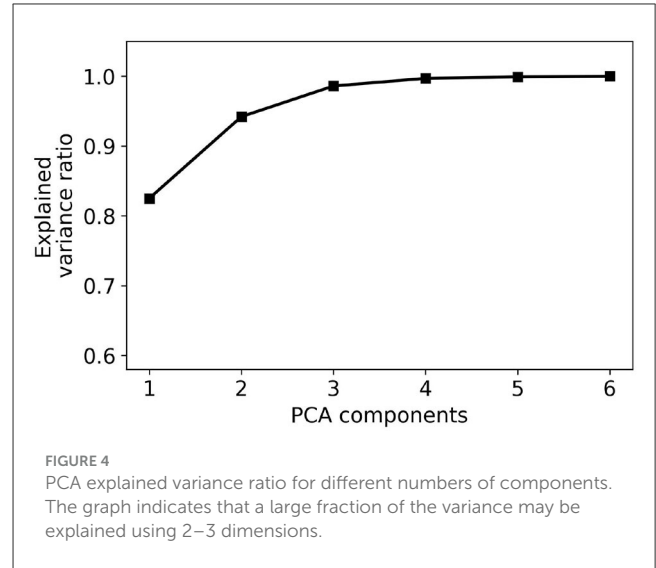
```
Input  : 𝒫- presets
         U - set of users
         losstype_u - unilateral or bilateral loss
         w_u - population weights
         ĉ_u - NAL-NL2 prescription for user u
         T_j - the j-th transfer function
         l_{u,j} - likelihood of potential config. c_{u,j}
         γ - user coverage threshold (γ = 0.8)
Output : PC(𝒫) - population coverage
```

1 **PC**($\mathcal{P}$) = 0
2 **for** $u \in U$ **do**
3   **if** *losstype*$_u$ == *unilateral* **then**
4   user_coverage = 0
5   **for** $j = \overline{1, 81}$ **do**
6   /* Compute the user's potential
            config.   */
7        $c_{u,j} = \hat{c}_u + T_j$
8   **if** $c_{u,j} \in \mathbf{C}(\mathcal{P})$ **then**
9   user_coverage = user_coverage + $l_{u,j}$
10   **end if**
11   **end for**
12   **if** *user_coverage* $\geq \gamma$ **then**
13   /* User is covered   */
14   **PC**($\mathcal{P}$) = **PC**($\mathcal{P}$) + w_u
15   **end if**
16   **else**
17   **foreach**
      fit_type $\in \{(uni; left), (uni; right), (bi; left), (bi; right)\}$ **do**
18   user_coverage$_{fit\_type}$ = 0
19   **for** $j = \overline{1, 81}$ **do**
20   /* Compute the user's potential
             config.   */
21          $c_{u(fit\_type),j} = \hat{c}_{u(fit\_type)} + T_j$
22   **if** $c_{u(fit\_type),j} \in \mathbf{C}(\mathcal{P})$ **then**
23   user_coverage$_{fit\_type}$ = user_coverage$_{fit\_type}$ + $l_{u,j}$
24   **end if**
25   **end for**
26   **end foreach**
27   **if** *user_coverage*$_{uni;left}$ $\geq \gamma$ & *user_coverage*$_{uni;right}$ $\geq \gamma$ &
      user_coverage$_{bi;left}$ $\geq \gamma$ & *user_coverage*$_{bi;right}$ $\geq \gamma$ **then**
28   /* User is covered   */
29   **PC**($\mathcal{P}$) = **PC**($\mathcal{P}$) + w_u
30   **end if**
31   **end if**
32 **end for**

Algorithm 1. Compute population coverage of presets $\mathcal{P}$.

FIGURE 4
PCA explained variance ratio for different numbers of components. The graph indicates that a large fraction of the variance may be explained using 2−3 dimensions.

FIGURE 5
Overlaying a grid in the 2-dimensional PCA space.

4. Preset optimization

In this section, we focus on the problem of fitting OTC HAs by either using preset collection- or slider-based approaches. In both cases, the challenge is to balance the population coverage of the $\mathcal{P}$ and the number of presets included. The number of presets included is a proxy for the complexity of the user interface provided to the end user and the increased time to identify the user's preferred configuration. Generally, user interfaces that expose fewer presets are likely easier to use but provide reduced population coverage. Conversely, increasing the number of presets increases population coverage at the cost of a more complex user interface. The algorithms discussed in this section can evaluate the trade-off between population coverage and the number of presets. Researchers may use this information to evaluate the effectiveness of existing self-fitting methods or to identify promising new approaches.

4.1. Collection-based self-fitting methods

Collection-based self-fitting methods maximize the population coverage of their presets. The input to the optimization problem is the sample of users with mild-to-moderate hearing loss, each with an associated population weight w_u. A feasible solution $\mathcal{P}$ must include N configurations (i.e., $|\mathcal{P}| = N$).

A key challenge to determining $\mathcal{P}$ is the high dimensionality of the configuration space. To simplify the problem, we use PCA to reduce the dimensionality of the configuration space to two dimensions. Figure 4 plots the fraction of the explained variance as the number of dimensions is increased from 1 to

6. The PCA transformation is applied for all the NAL-NL2 prescriptions of the selected users. The figure indicates that using two dimensions accounts for 95% of the total variance. Therefore, it is possible to map the high-dimensional configuration space into two dimensions with little accuracy loss. This result can be explained by a significant correlation between gains at different frequencies.

We can use the reduced two-dimensional space to simplify the original optimization problem. We observe that it is possible to construct a two-dimensional grid with G vertices that span the space of possible configurations, as shown in Figure 5. Then, a solution to the optimization problem involves selecting N out of the G configurations in the grid. Note that the granularity of the grid influences the population coverage of the selected presets ($\mathcal{P}$). For course-grained grids (i.e., when G is small), we can identify the best solution by evaluating the C_N^G possible combinations and selecting the one which provides the most population coverage. In contrast, it is computationally expensive to brute-force all solutions for finer-grained grids (i.e., when G is large), and more efficient algorithms are necessary.

We propose two approaches to solve the problem in the case of fine-grained grids. A simple strategy is to use a greedy algorithm that iteratively adds to $\mathcal{P}$ the preset that improves the population coverage the most. The algorithm maintains a set X of candidate configurations on the 2D grid that can be selected as presets. Additionally, the set $\mathcal{P}'$ contains the configurations already selected to be presets. Initially, $\mathcal{P}' = \emptyset$ and X include all the points in the grid. For each configuration c ($c \in X$), the algorithm computes population coverage of $\mathcal{P}' \cup \{c\}$. Then, the configuration that produces the largest increase in population coverage is added to $\mathcal{P}'$ and removed from X. While the greedy algorithm is computationally efficient, the greedy choice does not always identify the best set of presets.

A better alternative is to use a genetic algorithm (GA) to find the set of $\mathcal{P}$. We encode an individual's chromosome as a bit vector whose size equals the number of points on the two-dimensional grid. If a grid point is selected to be a preset, then its associated bit is set to one; otherwise, the bit is set to zero. The GA starts with an initial population containing individuals generated by setting N bits to equal 1 at random locations in the chromosome. The fitness of each individual is determined by its population coverage.

The next generation of individuals is constructed as follows. First, the individual with the highest fitness is added to the next generation. Then, half of the next generation is generated using cross-over. Cross-over involves two individuals with fitness in the top 50% of the population as parents. Then, a random location, L, is selected. A new individual is created by copying the first L bits from one parent and the remaining $N - L$ bits from the other parent. The remaining 50% of the population is obtained by mutating the chromosomes of individuals. Specifically, we select a random individual and mutate its chromosome: we flip two bits, one whose value was originally one and one whose value was originally zero. This operation enforces that the number of ones in the chromosome still sums up to N. The algorithm terminates after a predetermined number of iterations (set to 500 iterations in our experiments).

4.2. Slider–based self–fitting methods

Another common approach to configuring OTC HAs is providing a user interface with one to three sliders. An example of an interface with two sliders is shown in Figure 6A. An intuitive way of thinking about how the user selects the configuration is to imagine a ball representing current configuration selection (as show in Figure 6B). The ball moves when the sliders are manipulated. The ball's movement is constrained to a two-dimensional grid which we refer to as the controller grid. As expected, the slider "x" controls the ball's horizontal position, whereas the "y" controller controls its vertical position. The open question is how to map the ball's position in the controller space to a configuration in the space of possible configuration.
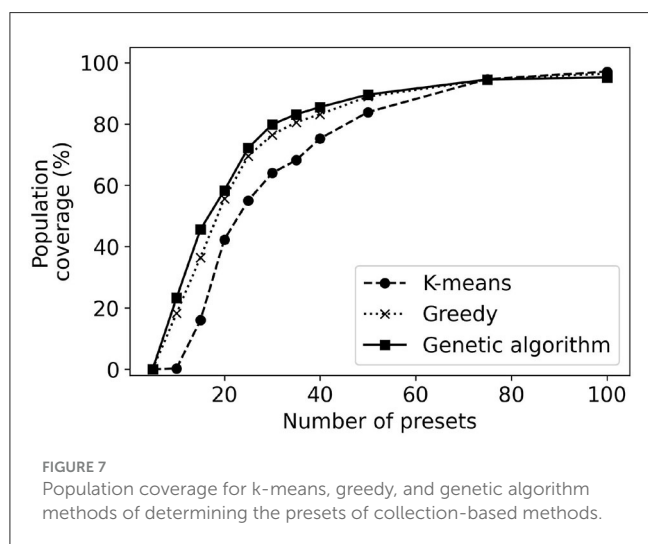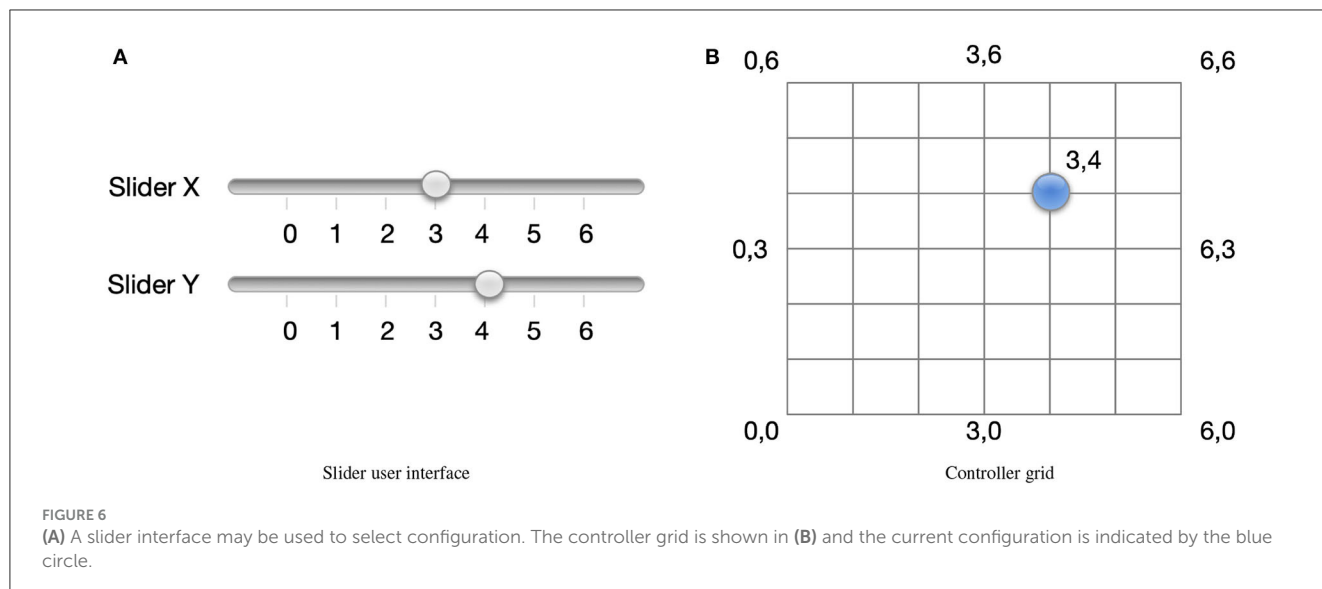
We will create a simple isomorphic mapping between the controller and configuration spaces. We start by identifying a bounding box that covers all the potential configurations in the configuration space. An example of such a bounding box is shown in Figure 5. Consider a coordinate system that originates at one of the corners of the bounding box and has its axis along the two sides of the bounding box. We will refer to one axis as bx and the other as by. Furthermore, let one of the corners of the bounding box be the origin with coordinates $(0, 0)$. We map the origin in the controller space to the origin in the configuration space. Next, we divide the bounding box to create a 2D grid such that the sides bx and by are divided in the same number of increments as sliders x and y, respectively. Each point on the configuration grid is assigned a coordinate according to the number of increments from the origin. The controller and configuration space are mapped such that a point with coordinates in the controller space (x, y) is mapped to a point with coordinates (bx, by) in the configuration space. Next, for each point on the grid, we determine its 2D coordinates in the embedded configuration space. The configurations at the grid points constitute the presets and are included in $\mathcal{P}$.

As in the case of collection-based interfaces, it is possible to evaluate and optimize the population coverage of slider-based interfaces. To evaluate the population coverage of the slider-based approach, Algorithm 1 is invoked with $\mathcal{P}$, including all the grid points. The population coverage depends only on the number of increments used for each of the sliders (and implicitly in the configuration grid). We may optimize these parameters to improve the population coverage.

5. Experiments

Our experiments answer the following questions:

- What is the trade-off between the number of presets and coverage of collection- and slider-based techniques? How is this trade-off impacted by the algorithms used to determine the presets?
- What is the impact of user demographics (e.g., sex or age) on coverage?
- How robust are our results with respect to our modeling assumptions?

FIGURE 6
(A) A slider interface may be used to select configuration. The controller grid is shown in **(B)** and the current configuration is indicated by the blue circle.

FIGURE 7
Population coverage for k-means, greedy, and genetic algorithm methods of determining the presets of collection-based methods.

5.1. Trade-off between population coverage and number of presets

Let us start by considering the performance of the greedy and genetic algorithms in building presets for collection-based self-fitting methods. We also ran k-means clustering algorithm as a baseline algorithm similar to Jensen et al. (2020). Specifically, we ran k-means on all the preferred configurations to construct N clusters. The presets were determined by computing the mean of all preferred configurations in the same cluster.

Figure 7 plots the population coverage as the number of presets is varied from 5 to 100. As expected, the general trend is that the population coverage increases with the number of presets. Initially, the population coverage increases fast with the number of presets. As the number of presets continues to grow, there are diminishing returns and further improvements in population coverage are harder to achieve. For example, in the case of the genetic algorithm, an increase from 15 to 20 presets results in an increase in coverage
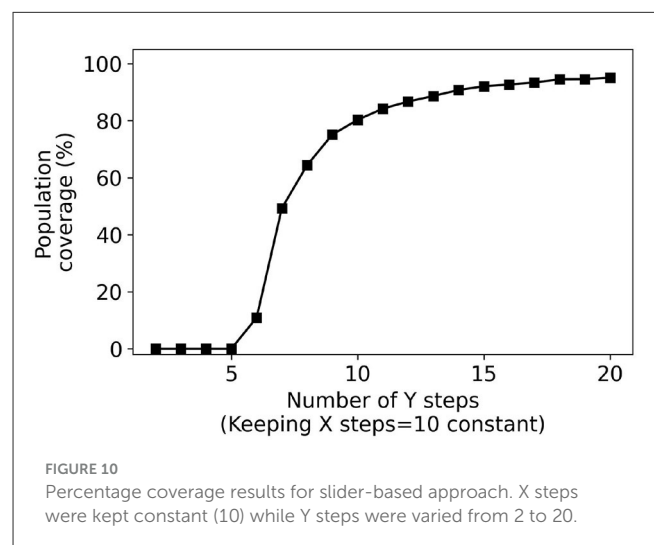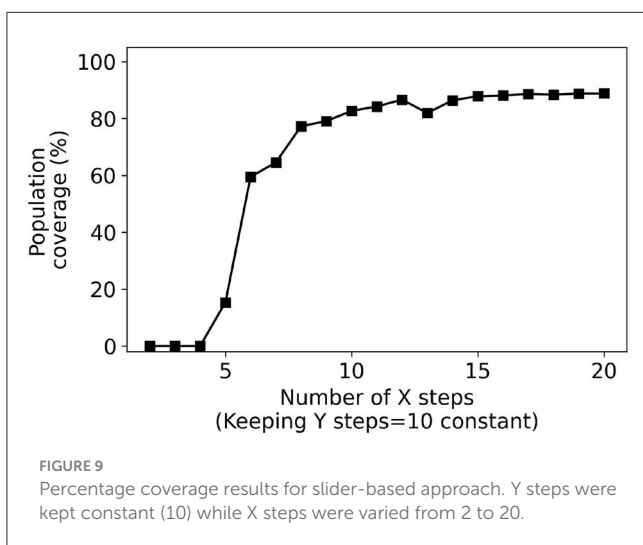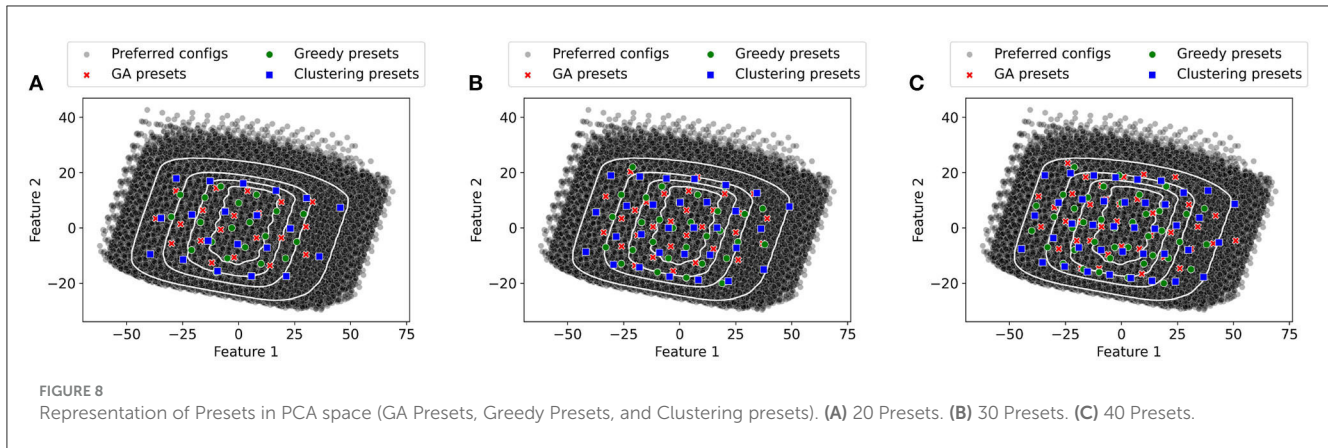
of 46–60%, an improvement of 14%. In contrast, increasing the number of presets from 35 to 40 results in an increase in population coverage of only 2.5%. The population coverage when the number of presets is 100 is approximately 95%. Improving the population coverage further would require a significant number of additional presets as the remaining 5% of the users have configurations that differ significantly from the remainder of the population.

Result: The population coverage of a collection-based method increases with the number of presets; however the improvements diminish with additional presets. Achieving a population coverage higher than 90% requires a large number of presets.

The method used to derive the presets has an impact on the obtained population coverage. The k-means clustering has consistently lower population coverage, particularly when the number of presets is below 35. The reason behind this worse performance is that k-means is not designed to maximize population coverage. The genetic algorithm provides a maximum of 9% improvement in coverage over the greedy algorithm. Note that even though a 9% improvement may seem minor, in practice this means tens of thousands of users who might find a configuration they prefer. When the number of presets is large, all considered methods of generating presets provide similar performance. Note that usually, the most important cases are when the number of presets is in the range of 4–30 when the algorithms differ the most in their performance. There are self-fitting methods that envision users performing simple auditory tests and being prescribed an OTC HA using one of four presets (Urbanski et al., 2021). On the other end, users might be provided with an interface to select one out of 30 presets for their use.

We use the greedy, genetic algorithm, and k-means to generate the 20, 30, and 40 presets. The generated presets are projected in two dimensions using PCA and shown in Figure 8. A common trend across the presets is the presets selected using k-means tend to be more spread out than those generated by greedy and genetic algorithms. As the presets picked by k-means tend to be more spread out some of them are in areas where they provide little population coverage. In contrast, the other two methods pick

FIGURE 8
Representation of Presets in PCA space (GA Presets, Greedy Presets, and Clustering presets). **(A)** 20 Presets. **(B)** 30 Presets. **(C)** 40 Presets.

FIGURE 9
Percentage coverage results for slider-based approach. Y steps were kept constant (10) while X steps were varied from 2 to 20.

FIGURE 10
Percentage coverage results for slider-based approach. X steps were kept constant (10) while Y steps were varied from 2 to 20.

are more densely packed toward the center of the figure. This result highlights the importance of picking an algorithm that is designed to maximize the population coverage (greedy or genetic algorithms) rather than as different objectives (for k-means).

Result: For the cluster-based methods that use four to twenty presets, the genetic algorithms provide better performance than the greedy and k-means approaches.

Next, we consider user interfaces that are slider-based where the number of steps on the sliders is increased from 2 to 20. Figures 9, 10 plot the population coverage for a different number of steps used for the x and y sliders respectively while keeping y and x steps constant at 10. As the number of steps is increased, the two-dimensional grid in the configuration space becomes denser and the associated population coverage also increases. A user interface with the two sliders each having ten steps provides a population coverage of 78.24%. In general, for the same number of presets sliders provide lower coverage than the same number of presets for collection-based presets. This is a consequence of the slider-based approaches constraining presets on the 2D grid whereas no such constraints are imposed for collection-based approaches.

Results: The proposed approach can be used to optimize the number of increments for slider-based approaches.

5.2. Impact of demographics

One approach that may yield improvements to the population coverage is to consider the impact of various demographics. In the following, we will consider whether dividing the population into four subgroups based on age and gender can yield presets that are tailored for those specific groups and provide higher coverage than the original presets that do not differentiate for gender or age. Historically, hearing loss in the population has varied as a function of demographics, with older adults having greater degrees of hearing loss than younger adults, and males having greater degrees of hearing loss than females (Pearson et al., 1995; Hoffman et al., 2017). Thus, dividing the population by age and gender may improve coverage estimates. The subsets into which we divide the population of users with mild-to-moderate hearing loss are:

- Male, age > 65
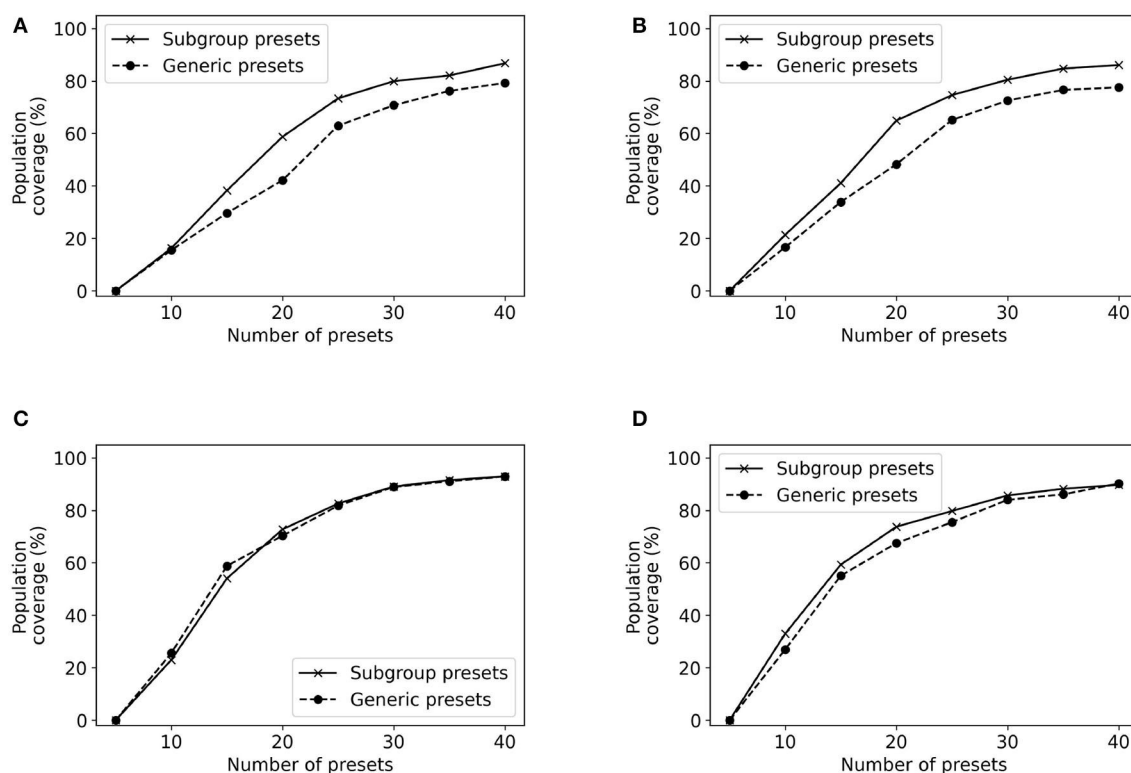- Male, age ≤ 65
- Female, age > 65
- Female, age ≤ 65

FIGURE 11
Population coverage for different subgroups. **(A)** Subgroup: male, age >65; **(B)** Subgroup: male, age <= 65; **(C)** Subgroup: female, age >65; **(D)** Subgroup: female, age <= 65.

Figure 11 plots the population coverage for each of the four groups with GA presets and presets obtained by running GA only on the subgroup population. The results indicate that in some cases higher population coverage may be achieved if the presets are built for that specific subset of the population. The most promising results are for the subgroup which includes males who are over 65. In this case, constructing presets specialized for this subgroup provides an improvement of 16% in population coverage over using the presets constructed for the general population. However, these results do not hold for all subgroups. Our approach provides no meaningful improvements for the subgroup which included women over 65 over constructing presets for the overall population.

Result: Using demographics to refine the target population of users may yield significant improvements in population coverage.

5.3. Robustness of results

An important question is to evaluate how robust are our methods to different ways of characterizing a user's deviation from NAL-NL2. We will use two different approaches—using bootstrap and scaling the variance of the Gaussian. We used bootstrap (Mackersie et al., 2020; Boothroyd et al., 2022) to obtain a distribution of possible values for the variance of the 2D Gaussian. Accordingly, we sample with replacement from the empirical data
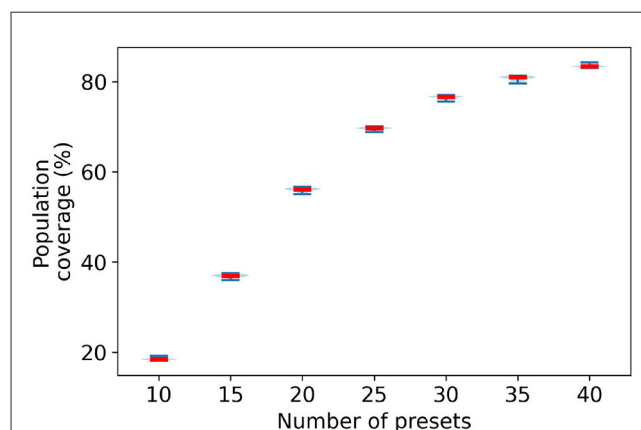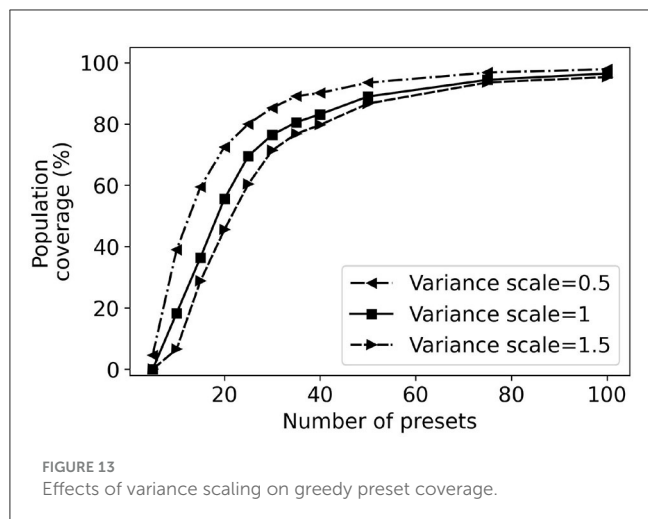

FIGURE 12
Distribution of population coverage by greedy presets using bootstrap technique.

of studies (Mackersie et al., 2020; Boothroyd et al., 2022) to generate 50 bootstrap samples. For each bootstrap sample, we run the greedy algorithm for 5 to 40 presets and compute their associated population coverage. Figure 12 plots the probability density function of the population coverage for the 20, 25, 30, 35, and 40 presets.

FIGURE 13
Effects of variance scaling on greedy preset coverage.

The figure shows that there is little variation in the population coverage for the generated bootstrap samples. Moreover, the variation tends to diminish with the number of presets. At 20 presets the population coverage range fluctuates by ± 0.75%. With an increasing number of presets, this variation in range significantly goes down. At 40 presets, the fluctuation range is ± 0.25%. Therefore, there are small variations in the population coverage within the bootstrap samples. This indicates that if the data from original studies used to build the 2D Gaussian is representative, then we should observe little variation in the population coverage.

Another approach to evaluate the robustness of our approach is to consider the impact that a higher Gaussian variance may have. We have scaled the fitted variance using the data from studies (Mackersie et al., 2020; Boothroyd et al., 2022) by scalars 0.5, 1, and 1.5. The expectation is that increasing the variance will result in lower coverage for the same number of presets. Stated differently, a larger number of presets is necessary to achieve the same coverage. We have evaluated the impact of scaling the variance in the case of using the greedy algorithm to build presets.

Figure 13 plots the effects of variance scaling on population coverage by greedy presets. When the variance is 1.5 times larger, there is a reduction of 3.5% in population coverage. Conversely, when the variance is 0.5 times smaller, there is an increase of 7% in population coverage at 40 presets. Overall, these results indicate that population coverage results are robust with respect to larger variances than those expected based on the empirical data alone.

Intuitively, if the variance is lower, there is a higher variation weight associated with the variations closer to the user's REIG. Thus, it takes less number of presets to cover the same population. It reflects in our results i.e., population coverage difference is as high as 32% for 10 greedy presets generated between a variance scale of 0.5 and 1.5. This wide difference narrows down to almost 7% for 50 greedy presets.

Result: The experiments demonstrate that the population coverage results are robust different assumptions about the underlying Gaussian distribution controlling the weights of transfer functions.

6. Discussion

The paper uses population coverage to evaluate and optimize various settings for preset-based methods. We caution the reader that population coverage is not the only metric that may be used to assess self-fitting methods. To fully evaluate self-fitting methods, it is necessary to run detailed user studies to evaluate their performance. However, population coverage can guide how user studies may be set up. For example, in the case of collection-based approaches, the region of interest is when the number of presets is in the range of 20–40. Using fewer than 20 presets yields would cover only a small fraction of users. In contrast, increasing the number of presets beyond 40 results in minor improvements in population coverage. Further user studies should particularly focus on this range. Similarly, in the case of slider-based approaches, the region of interest is when the number of increments is about ten for both sliders.

The computation of population coverage depends on a number of modeling assumptions and hyper-parameters. The hyper-parameters of our model include γ, R, and the likelihood of each possible configuration as given by the two-dimensional Gaussian. Our experiments show that the population coverage values are fairly robust to increases in the variance of the Gaussian distribution. The parameters γ and R are configured based on audiology expertise. Our experimentation with different values for γ and R shows that although population coverage changes, the same overall trends are observed.

Both studies (Mackersie et al., 2020; Boothroyd et al., 2022) that we used to create transfer functions utilize speech listening to conduct self-fitting. Therefore, the user preference could be mainly driven by speech understanding or clarity. In a different listening situation, such as music listening, other factors, such as sound quality (e.g., fullness), could drive the preference. Therefore, our study may not generalize to situations other than speech listening.

Audibility is only one factor among a constellation of complex factors that might be expected to affect hearing aid setting selection and hearing aid outcomes. This study investigates a theoretical approach to optimizing the selection of hearing aid gain-frequency configurations such that a preferable setting was included for the majority of potential users. Importantly, this study did not test the identified presets empirically among a sample of hearing aid users; such a study is a critical next step to determine the real-world viability of our approach. This study also did not consider factors in the selection process that might affect outcomes such as pre- or post-selection self-assessment or service delivery model, though such factors have not been shown to have large impacts on hearing aid satisfaction for OTC hearing aids (e.g., Humes et al., 2017; Mackersie et al., 2020).

7. Conclusion

This paper proposes a novel metric—population coverage—to evaluate, compare, and optimize preset-based self-fitting methods. The population coverage estimates how well a given number of presets meet the needs of populations of users with different characteristics. The unique aspect of our approach is that the population coverage is computed to account for

how a user's preferred configuration may deviate from the NAL-NL2 prescription. We apply the population coverage to optimize the presets used by collection- and slider-based methods. Specifically, greedy and genetic algorithms are proposed presets that maximize population coverage. Similarly, algorithms are proposed to configure the number of intervals on slider-based interfaces to maximize coverage. Our experiments indicate that the proposed algorithms can effectively identify presets for both collection- and slider-based methods. Moreover, we may use population coverage to narrow down the different configurations of preset-based methods with good population coverage and whose performance should be further characterized through user studies.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

DV conception or design of the work. DV and EJ data collection, data analysis and interpretation, and drafting the article. DV, EJ, Y-HW, and OC critical revision of the article and final approval of the version to be published.

All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Boothroyd, A., and Mackersie, C. (2017). A "Goldilocks" approach to hearing-aid self-fitting: user interactions. *Am. J. Audiol.* 26, 430–435. doi: 10.1044/2017_AJA-16-0125

Boothroyd, A., Retana, J., and Mackersie, C. L. (2022). Amplification self-adjustment: Controls and repeatability. *Ear Hear.* 43, 808–821. doi: 10.1097/AUD.0000000000001141

Brody, L., Wu, Y.-H., and Stangl, E. (2018). A Comparison of personal sound amplification products and hearing aids in ecologically relevant test environments. *Am. J. Audiol.* 27, 581–593. doi: 10.1044/2018_AJA-18-0027

Caswell-Midwinter, B., and Whitmer, W. M. (2019). Discrimination of gain increments in speech-shaped noises. *Trends Hear.* 23, 2331216518820220. doi: 10.1177/2331216518820220

Chien, W., and Lin, F. R. (2012). Prevalence of hearing aid use among older adults in the United States. *Arch. Intern. Med.* 172, 292–293. doi: 10.1001/archinternmed.2011.1408

Ciletti, L., and Flamme, G. A. (2008). Prevalence of hearing impairment by gender and audiometric configuration: results from the national health and nutrition examination survey (1999–2004) and the Keokuk county rural health study (1994–1998). *J. Am. Acad. Audiol.* 19, 672–685. doi: 10.3766/jaaa.19.9.3

Cox, R. M., Schwartz, K. S., Noe, C. M., and Alexander, G. C. (2011). Preference for one or two hearing aids among adult patients. *Ear Hear.* 32, 181. doi: 10.1097/AUD.0b013e3181f8bf6c

Edwards, B. (2020). Emerging technologies, market segments, and MarkeTrak 10 insights in hearing health technology. *Semin. Hear.* 41, 37–54. doi: 10.1055/s-0040-1701244

Hoffman, H. J., Dobie, R. A., Losonczy, K. G., Themann, C. L., and Flamme, G. A. (2017). Declining prevalence of hearing loss in US adults aged 20 to 69 years. *JAMA Otolaryngol. Head Neck Surg.* 143, 274–285. doi: 10.1001/jamaoto.2016.3527

Humes, L. E., Rogers, S. E., Quigley, T. M., Main, A. K., Kinney, D. L., and Herring, C. (2017). The effects of service-delivery model and purchase price on hearing-aid outcomes in older adults: a randomized double-blind placebo-controlled clinical trial. *Am. J. Audiol.* 26, 53–79. doi: 10.1044/2017_AJA-16-0111

Jensen, J., Vyas, D., Urbanski, D., Garudadri, H., Chipara, O., and Wu, Y.-H. (2020). Common configurations of real-ear aided response targets prescribed by NAL-NL2 for older adults with mild-to-moderate hearing loss. *Am. J. Audiol.* 29, 460–475. doi: 10.1044/2020_AJA-20-00025

Jilla, A. M., Johnson, C. E., and Huntington-Klein, N. (2020). Hearing aid affordability in the United States. *Disabil. Rehabil. Assist Technol.* 18, 246–252. doi: 10.1080/17483107.2020.1822449

Keidser, G., Dillon, H., Flax, M., Ching, T., and Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiol. Res.* 1, e24. doi: 10.4081/audiores.2011.e24

Lin, F. R., Niparko, J. K., and Ferrucci, L. (2011). Hearing loss prevalence in the United States. *Arch. Intern. Med.* 171, 1851–1852. doi: 10.1001/archinternmed.2011.506

Mackersie, C., Boothroyd, A., and Lithgow, A. (2019). A "Goldilocks" approach to hearing aid self-fitting: ear-canal output and speech intelligibility index. *Ear Hear.* 40, 107–115. doi: 10.1097/AUD.0000000000000617

Mackersie, C. L., Boothroyd, A., and Garudadri, H. (2020). Hearing aid self-adjustment: effects of formal speech-perception test and noise. *Trends Hear.* 24, 2331216520930545. doi: 10.1177/2331216520930545

McKee, M. M., Choi, H., Wilson, S., DeJonckheere, M. J., Zazove, P., and Levy, H. (2019). Determinants of hearing aid use among older Americans with hearing loss. *Gerontologist.* 59, 1171–1181. doi: 10.1093/geront/gny051

NAL-NL2 (2022). *NAL-NL2 Software.*

National Center for Health Statistics (2022). *National Health and Nutrition Examination Survey.*

Nelson, P. B., Perry, T. T., Gregan, M., and VanTasell, D. (2018). Self-adjusted amplification parameters produce large between-subject variability and preserve speech intelligibility. *Trends Hear.* 22, 2331216518798264. doi: 10.1177/2331216518798264

Nieman, C. L., Marrone, N., Szanton, S. L., Thorpe, R. J., and Lin, F. R. (2016). Racial/ethnic and socioeconomic disparities in hearing health care among older Americans. *J. Aging Health.* 28, 68–94. doi: 10.1177/0898264315585505

Olusanya, B. O., Davis, A. C., and Hoffman, H. J. (2019). Hearing loss grades and the international classification of functioning, disability and health. *Bull. World Health Organ.* 97, 725. doi: 10.2471/BLT.19.230367

Pearson, J. D., Morrell, C. H., Gordon-Salant, S., Brant, L. J., Metter, E. J., Klein, L. L., et al. (1995). Gender differences in a longitudinal study of age-associated hearing loss. *J. Acoust. Soc. Am.* 97, 1196–1205. doi: 10.1121/1.412231

Powers, T. A., and Rogin, C. M. (2019). Marketrak 10: hearing aids in an era of disruption and DTC/OTC devices. *Hear. Rev.* 26, 12–20.

Reed, N. S., Garcia-Morales, E., and Willink, A. (2021). Trends in hearing aid ownership among older adults in the United States From 2011 to 2018. *JAMA Intern Med.* 181, 383–385. doi: 10.1001/jamainternmed.2020.5682

Sabin, A. T., Van Tasell, D. J., Rabinowitz, B., and Dhar, S. (2020). Validation of a self-fitting method for over-the-counter hearing aids. *Trends Hear.* 24, 2331216519900589. doi: 10.1177/2331216519900589

Søgaard Jensen, N., Hau, O., Bagger Nielsen, J. B., Bundgaard Nielsen, T., and Vase Legarth, S. (2019). Perceptual effects of adjusting hearing-aid gain by means of a machine-learning approach based on individual user preference. *Trends Hear.* 23, 2331216519847413. doi: 10.1177/2331216519 847413

Urbanski, D., Hernandez, H., Oleson, J., and Wu, Y.-H. (2021). Toward a new evidence-based fitting paradigm for over-the-counter hearing aids. *Am. J. Audiol.* 30, 43–66. doi: 10.1044/2020_AJA-20-00085