Evaluation of Hearing Aid Personalization Algorithms

Justin Jensen[1], Dhruv Vyas[2], Harinath Garudadri[3], Octav Chipara[2], Yu-Hsiang Wu[1]

[1]Department of Communication Sciences & Disorders, The University of Iowa; [2]Department of Computer Science, The University of Iowa, [3]Qualcomm Institute, University of California San Diego

INTRODUCTION

BACKGROUND

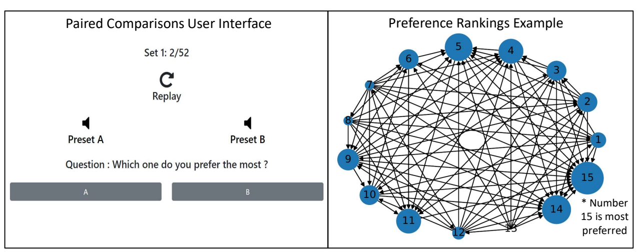
- Users of self-fitting hearing aids often need to select hearing aid configurations without assistance from professionals. While limiting the number of available configurations makes this selection process easier, it also limits the potential for personalizing the settings to a given user's needs. Conversely, offering large numbers of potential configurations would allow for more personalization but would make the process of selecting one considerably more challenging.

PURPOSE

- The question of how an individual might choose the most optimal configuration without examining or even knowing about all possible options was framed as a multi-armed bandit problem, where a solution requires balancing exploration of available options with exploitation of known preferences or probabilities. The goal is to efficiently converge on an optimal configuration without the need for exhaustively examining every option.
- The purpose of this study is to evaluate the performance of five selection algorithms that could be used to simplify the process of selecting personalized settings from among a pool of possible configurations for self-fitting hearing aids.

METHODS

- Audiometric data from a national health database were used to develop 15 hearing aid gain-frequency responses representing at least one clinically appropriate hearing aid fitting for 95% of older adults in the U.S. with mild-to-moderate sensorineural hearing loss.
- Speech was recorded from the output of a hearing aid programmed with each of the 15 gain-frequency responses and presented to 28 older adults with mild-to-moderate sensorineural hearing loss.
- The participants used a paired comparison paradigm to determine the order and strength of the preference of the gain-frequency responses for all possible combinations (105 pairwise comparisons x 4 repetitions).



- The preference data were then used as the ground truth to simulate each user's behavior through five selection algorithms. The first two algorithms do not use machine learning. The remaining three are classic machine learning algorithms designed to learn from user input and converge on an optimal solution over time.
- The performance of each algorithm was evaluated by how quickly (convergence rate), accurately (accuracy), and consistently (robustness) the algorithm converges on the gain-frequency response most preferred by the study participant.

RESULTS

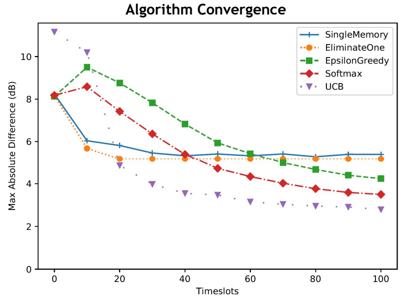


Figure 1. Average convergence of each algorithm towards users' most preferred configuration plotted as max absolute difference* as a function timeslots~ (i.e. required number of inputs from the user).

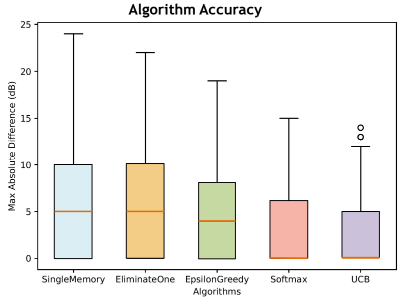


Figure 2. Accuracy plotted as the distribution of max absolute difference* across 1,000 simulations of 100 timeslots~.

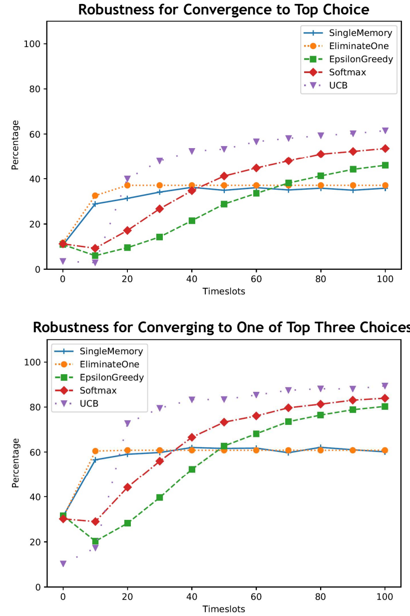


Figure 3. Robustness plotted as the percentage out of 1,000 simulations that the algorithm converges to the user's top choice (top) or to one of the user's top three choices (bottom) as a function of required number of inputs from the user (timeslots~).

- The UCB algorithm has the best overall performance across subjects in terms of convergence rate, accuracy, and robustness. The Single-memory and Eliminate-one algorithms performed significantly worse on all measures than the three machine learning algorithms.
- The UCB algorithm converges to a configuration with a max absolute difference* of 5 dB from the user's most preferred configuration after 20 pairwise comparisons. Softmax and epsilon-greedy require significantly more pairwise comparisons to achieve comparable max absolute difference.
- The UCB and Soft-max algorithms are significantly more accurate on average than the other algorithms after 100 required pairwise comparison inputs from the user.
- Robustness of the algorithms was examined for convergence to either the user's most preferred configuration or to one of the user's top three preferred configurations. The UCB algorithm is 70% consistent across 1,000 simulations in converging to one of the user's top three preferred configurations after 20 pairwise comparisons.

* Max absolute difference is the maximum difference between the algorithm's current best option and the user's most preferred option at any of eight frequencies (0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz).

~ At each "timeslot," the algorithm presents a pairwise comparison to the simulated user. The simulated user's choice is based on the empirical user preference data collected from human subjects.

SELECTION ALGORITHMS

1. Single-memory algorithm:
   - Choose the best configuration following each pairwise comparison and keep it in memory until replaced after a new comparison.
2. Eliminate-one algorithm:
   - Eliminate the rejected configuration after each pairwise comparison.
3. Epsilon-greedy algorithm:
   - During exploration, all available options are considered equally. New pairwise comparisons are chosen randomly from total option pool.
4. Softmax algorithm:
   - Stores and updates probabilities that a given option will be winner. These probabilities influence the likelihood that a given option will be presented in future pairwise comparisons.
5. Upper Confidence Bound (UCB) algorithm:
   - Tracks probability of win plus number of times a given option is presented to the user. If difference in score of top two is more than the confidence bound, the top choice is declared the overall winner.

DISCUSSION

- The performance varies across algorithms, and there is trade-off between required user inputs and the accuracy and robustness of an algorithm.
- Accuracy and robustness are poor across all algorithms for between 5 and 15 required inputs from the user. The UCB algorithm's performance improves dramatically by 20 pairwise comparisons, while the other machine learning algorithms take longer to achieve comparable results.
- Converging to one of the user's top three most preferred configurations that is within a max absolute difference of 5 dB from the most preferred option may be close enough to be an acceptable choice for the user.
- The data from this study can be used to develop smarter and more efficient algorithms for self-fitting hearing aids in the future.

ACKNOWLEDGMENTS/REFERENCES

National Science Foundation, Division of Information & Intelligent Systems (IIS-1838830)
National Institute on Disability, Independent Living, and Rehabilitation Research (90REGE0013)

Urbanski, et al. AAS 2019. Jensen et al. AAS 2019. National Health and Nutrition Examination Survey 2009-2012. *Retrieved from: https://www.cdc.gov/nchs/nhanes/index.htm.* Arabie, et al. 1996. *Clustering and Classification*, pp. 65-121. Lai and Robbins. 1985. *Science Direct, 6(1), 4-22.*

CONTACT

Contact Justin-Jensen@uiowa.edu for more information