ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095325

USING MACHINE LEARNING TO UNDERSTAND THE RELATIONSHIPS BETWEEN AUDIOMETRIC DATA, SPEECH PERCEPTION, TEMPORAL PROCESSING, AND COGNITION

Rana M. Khalil[⋆], *Alexandra Papanicolaou*[†], *Renee Ti Chou*[⋆], *Bobby E. Gibbs, II*[†], *Samira Anderson*[†], *Sandra Gordon-Salant*[†], *Michael P. Cummings*[⋆], and *Matthew J. Goupell*[†]

[⋆]Center for Bioinformatics and Computational Biology, University of Maryland, College Park
[†]Department of Hearing and Speech Sciences, University of Maryland, College Park

ABSTRACT

Aging and hearing loss cause communication difficulties, particularly for speech perception in demanding situations, which have been associated with factors including cognitive processing and extended high-frequency ($>$ 8 kHz) hearing. Quantifying such associations and finding other (possibly unintuitive) associations is well suited to machine learning. We constructed ensemble models for 443 participants who varied in age and hearing loss. Audiometric, perceptual, electrophysiological, and cognitive data were used to predict speech perception in noise, reverberation, and with time compression. Speech perception was best predicted by variables associated with audiometric thresholds (including new across-frequency composite variables) between 1–4 kHz, followed by basic temporal processing ability. Cognitive factors and extended high-frequency thresholds had little to no predictive ability of speech perception. Future associations or lack thereof will inform the field as we attempt to better understand the intertwined effects of speech perception, aging, hearing loss, and cognition.

Index Terms— Machine Learning, Audiology, Speech Perception, Temporal Processing, Cognition

1. INTRODUCTION

Over 50% of adults in the US >65 years suffer from significant hearing loss [1]. As people age, their ability to understand speech decreases and places a great burden on their ability to effectively communicate, particularly in demanding situations that include background noise, reverberation, multiple talkers, rapid talkers, or accented talkers [2]. These declines are a result of peripheral hearing loss (i.e., some speech sounds are less audible), central age-related temporal-processing deficits (i.e., neural encoding of speech sounds is distorted), and a decrease in higher-level cognitive abilities (e.g., memory, non-auditory processing speed) [3].

The complexity of understanding how age, hearing loss, and cognition intertwine as factors to predict speech-perception performance argues for modern data-driven analysis approaches on human measurements [4]. This includes machine learning, if the data set is large enough. For the present data study, we collected a large battery of audiometric, perceptual, electrophysiological, and cognitive data in 443 participants [5]. Novel to this data set is the number of subjects and suprathreshold perceptual measures of auditory temporal processing (e.g., pulse-rate discrimination).

Numerous studies have shown that among older listeners, speech perception is related to working memory [6, 7] and speed of processing [7, 8]. There is a possible relationship between hearing loss and cognitive decline as a possible precursor to dementia and Alzheimer's disease [9, 10].

The large multidimensional data set contained highly correlated variables, a challenge requiring variable reduction for improved predictions. Variable reduction for hearing thresholds has involved cluster analyses [11, 12]. Our variable reduction approach involved multiple hearing threshold composite measures from 0.25–14 kHz. Evaluating extended high-frequency thresholds between 8–14 kHz has become increasingly common because they appear to predict speech perception in noise [13, 14].

Here we determine variables predictive of speech perception to verify previously identified associations in the literature and identify possible novel associations. The goal is to produce hypotheses for future studies.

2. METHODS

2.1. Participants

We recruited 443 adult participants (293 female, 125 male, 25 no response) for a separate study [5]. They had a range of ages (18–85 y, $\bar{x}$ = 59.7, SD = 22.5) and hearing loss. All participants had a rigorous audiometric evaluation and cognitive assessment done. A subset of 130 participants performed additional speech perception and non-speech auditory temporal processing tasks, and participated in an electrophysiological assessment.

2.2. Audiometric Tests

Hearing thresholds were assessed at octave intervals between 0.125 and 8 kHz using standard audiometric techniques and a calibrated audiometer in a sound-attenuating booth. Audiometric thresholds were also collected at 3, 6, 10, 12.5, and 14 kHz, and tympanometry was performed in each ear. Many participants had audiometrically normal hearing, defined as pure-tone thresholds $\leq$ 25 dB hearing level (HL) [15] at 0.25, 0.5, 1, 2, and 4 kHz in the right ear. There were 100 younger normal-hearing listeners with a high-frequency pure-tone average (HF-PTA: 1, 2, 4 kHz) = 7.3 $\pm$ 12.3 dB HL and 187 older normal-hearing listeners with a HF-PTA = 24.4 $\pm$ 17.2 dB HL. Some older listeners (n = 120) had a mild-to-moderate hearing loss [defined as a HF-PTA $<$ 30 dB HL and thresholds at 2 kHz and 4 kHz $<$ 70 dB HL], with an average HF-PTA = 40.9 $\pm$ 8.0 dB HL.

2.3. Perception, Electrophysiological, and Cognitive Tests

Temporal processing was assessed for several psychoacoustical tasks, including pulse-rate discrimination, gap detection, gap duration discrimination, and tempo discrimination [5]. Auditory brainstem responses were recorded to 100-μs broadband clicks. Auditory steady-state responses (ASSR) were recorded to band-limited pulse trains presented at rates of 100, 200, 300, and 400 Hz [5]. Sentence recognition in quiet was measured using the IEEE corpus [16] in five conditions: a normal rate with no reverberation (i.e., clean speech), with 40% and 60% time compression, and with reverberation (0.6-s and 1.2-s RT_{60}). There were 10 sentences presented in each condition. Sentence recognition in noise was measured using the QuickSIN [17]. Word recognition scores were collected for single 25-word lists of the NU-6 test [18] presented bilaterally at 75 dB HL in quiet. The Montreal Cognitive Assessment (MoCA) [19], assessments from the National Institutes of Health Cognition Toolbox (List Sorting Working Memory Test, the Flanker Inhibitory Control and Attention Test, the Pattern Comparison Processing Speed Test, and the Dimensional Card Sort Test) [20], and a subset of the Speech, Spatial, and Qualities of Hearing Scale [21] were administered.

2.4. Machine Learning Analysis

We derived a total of 147 calculated features from the audiometric measurements. We recorded HL of each audiogram and computed new features based on the mean and difference between the two ears for each frequency. Using these features, we computed profile variables by applying a simple linear model to each participant: slope, intercept, coefficient of determination of the line of best fit, and sum of hearing thresholds across frequencies. We calculated these profile features for three different frequency ranges: standard frequencies (SF; 2.5–8 kHz), extended high frequencies (EHF; 10–14 kHz), and all frequencies.

We extracted additional features from the audiograms of both ears including total length of segments joining consecutive thresholds, highest frequency each subject can hear, inflection point, notch index [22], and a notch presence using to three definitions. One type of notch definition was a point at 3, 4, or 6 kHz where there are differences of at least 15 dB and 10 dB compared to the previous and next frequencies, respectively. The two other definitions were described by others [23, 24]. Finally, we added features that describe the symmetry, configuration, and severity of an audiogram [25].

Some subjects did not respond at the highest-intensity limits of the equipment, indicating a profound hearing loss at that frequency (typically at extended high frequencies $>$ 8 kHz). To reflect this hearing loss, we replaced the missing hearing thresholds with +5 dB above the maximum intensity level that our equipment allows at each frequency, as is typical in other publications [26]. This way, the hearing thresholds reflect the profound degree of the true thresholds but remain distinguishable from those who actually heard the signals at the equipment limits. For the other variables, missing categorical values were imputed by using the most frequent category, and the numeric values were replaced with variable medians. All features are grouped into the following seven categories: subject information, audiometric measures, behavioral-speech tests, behavioral-non-speech tests, electrophysiological measures, cognitive tests, and subjective questionnaires. We reduced the number of features by selecting those that minimized the total Akaike information criterion based on the results of random forest analyses.

We built super learner [27] stacked generalization models with an ensemble of methods using SuperLearner package [28] with ten-fold cross-validation and non-negative least square error as a loss function. For base models, we included random forests, lasso and elastic-net regularized generalized linear models, extreme gradient boosting, and feed-forward neural networks. To evaluate the performance of the super learner model, an additional layer of 5-fold cross-validation was performed and the whole process was repeated five times. The performance measure (percent variance explained) was then averaged across all folds and repeats. All analyses for this study were carried out in the Statistical Computing Programming Language R (Version: 4.1.3) [29].

Model performance was measured as the percentage of variance explained, which is defined as the fraction of the variance of the response variable that can be explained using the predictors. Permutation-based feature importance was calculated from 25 replications and reported as relative values such that the maximum score has a value of 100.

3. RESULTS

We used super learner to model the scores of three different speech-perception tests: QuickSIN speech-in-noise, sentence recognition with 60% time compression, and sentence recog-

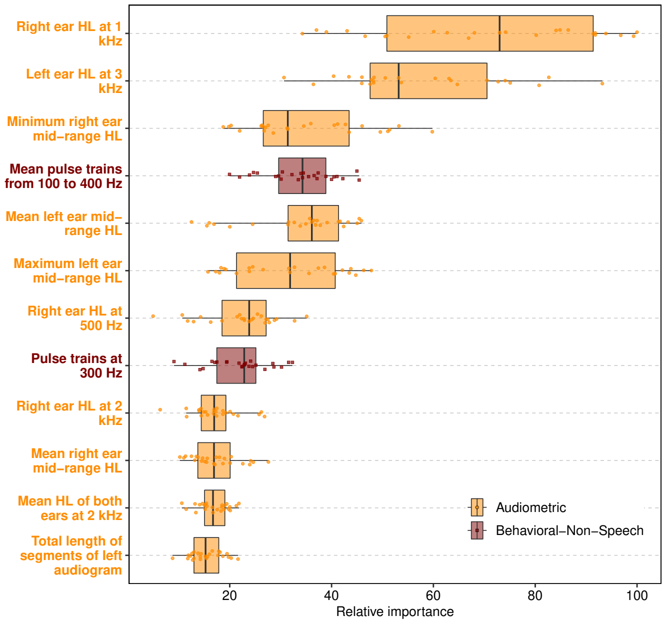

Fig. 1: Variable importance for predicting QuickSIN Speech-in-Noise. Points represent importance scores from 25 replications. Box plots indicate median (middle line), 25^{th}, 75^{th} percentile (box). Whiskers extend 1.5 $\times$ interquartile range.

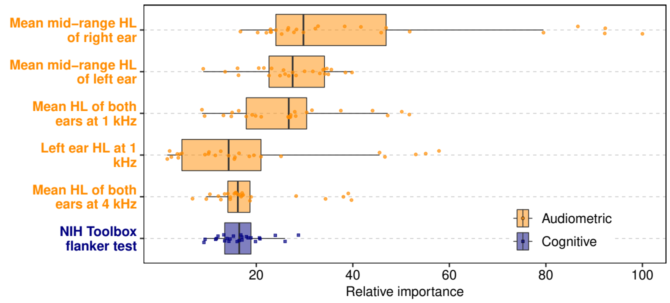

Fig. 2: Variable importance for predicting sentence recognition with 60% time compression. Points represent importance scores from 25 replications. Box plots indicate median (middle line), 25^{th}, 75^{th} percentile (box). Whiskers extend 1.5 $\times$ interquartile range.

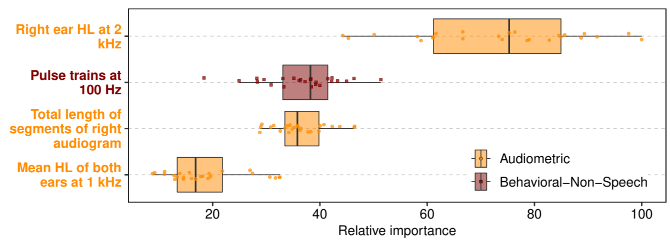

Fig. 3: Variable importance for predicting sentence recognition with 1.2-s reverberation time. Points represent importance scores from 25 replications. Box plots indicate median (middle line), 25^{th}, 75^{th} percentile (box). Whiskers extend 1.5 $\times$ interquartile range.

nition with 1.2-s reverberation time. To eliminate the dependence between different behavioral-speech features, we excluded all other features that fall under the same category before performing feature elimination and building the super learner model. For simplicity, the plots are restricted to features with average relative importance > 15 (an arbitrarily chosen value). Group variable importance analysis selected the audiometric category as the most important predictor for the scores of the three speech-perception tests.

The proportion of variance in QuickSIN test scores explained was 66.91%, and 10 of 12 important features were from the audiometric category including (ordered by importance) HL of the right ear at 1 kHz, HL of the left ear at 3 kHz, minimum threshold of the right ear at the mid-frequency range (1–3 kHz), mean and maximum of thresholds of the left ear at the mid-frequency range, HL of the right ear at 0.5 and 2 kHz, mean of thresholds of the right ear at the mid-frequency range, mean of HL of two ears at 2 kHz, and total length of segments joining consecutive left ear thresholds (Fig. 1). The other two were from the behavioral-non-speech category (i.e., basic temporal processing), which includes the average of pulse-rate discrimination values across frequencies (100–400 Hz) and at the single pulse rate of 300 Hz.

The proportion of variance in 60% time compression sentence recognition scores explained was 74.47%, and 5 of 6 important features were from the audiometric category including (ordered by importance) mean of thresholds of the right ear at the mid-frequency range, mean of thresholds of the left ear at the mid-frequency range, mean of HL of two ears at 1 kHz, HL of the left ear at 1 kHz, and mean of HL of two ears at 4 kHz (Fig. 2). The last feature was the Flanker test (inhibitory control and attention cognitive test).

The proportion of variance in 1.2-s reverberation time sentence recognition scores explained was 71.77%, and 3 of 4 important features were from the audiometric category including (ordered by importance) HL of the right ear at 2 kHz, total length of segments joining consecutive right ear thresholds, and mean of HL of two ears at 1 kHz (Fig. 3). The last feature was from the behavioral-non-speech category, pulse-rate discrimination for 100-Hz pulse trains.

Information about model parameters, variance explained, model weight, R code and models are available in the supplementary material at `https://osf.io/ta7kf`.

4. DISCUSSION

We used machine learning to determine variables predictive of demanding speech-perception conditions to verify previously identified associations and identify possible novel associations, producing hypotheses for future studies. We also developed new variables that had significant contributions to the variance explained by our models, such as the total length of segments joining consecutive thresholds (Fig. 1 and 3). Such variables capture properties of the entire hearing range, rather than single frequencies, showing the importance of using composite variables to reduce the audiometric variable space.

Although others have used clustering and ensemble-based approaches to examine communication difficulties in relation to audiometric profiles [30, 31], our results suggest that the largest amount of speech-perception variance is explained by audibility from 1–4 kHz, followed by temporal processing ability, and attention (Fig. 1, 2, and 3). Audiometric variables were the most predictive of demanding speech perception, particularly for variables that were associated with thresholds between 1–4 kHz (e.g., the mid-range composite variables; Fig. 1, 2), consistent with previous studies [32]. Conspicuous in their absence were extended high-frequency (> 8 kHz) threshold variables. It could be that the importance of extended high-frequency thresholds is diminished by hearing loss around 1–4 kHz to a negligible level. Similarly, hearing loss can obscure effects of aging, and one needs to carefully design their study to include groups that vary both by hearing loss and age [5] to see effects of the relatively weaker factor of age.

We found a limited role of cognitive processing abilities in explaining the variance in speech-perception performance. The only cognitive factor to appear in our results was the Flanker score, which is a measure of inhibition of irrelevant stimuli and attention; this factor was the second most important factor in a meta-analysis [7] and was negatively correlated with cortical envelope in older listeners in [33]. Absent was working memory and processing speed, which was unexpected given past studies [5, 7, 8]. Noteworthy is that the MoCA, which is prevalent in its use as a cognitive screener in the field of hearing and aging, had no predictive power for our outcome variables. In fact, in a separate analysis (not shown) the combined MoCA score or any individual question was not predictive of any variable in this dataset.

These findings may suggest that researchers need to be cautious in future study design when investigating extended high-frequency threshold and cognitive effects on auditory processing like speech perception. Controlling for more prominent variables like hearing thresholds appears critical.

5. ACKNOWLEDGMENTS

Funding was provided by a University of Maryland Brain and Behavior Institute Seed Grant (MPC, MJG), award T32-DC000046 from the National Institute on Deafness and Other Communication Disorders (trainee BEGII), and award P01AG055365 from the National Institute of Aging of the National Institutes of Health (SGS).

6. REFERENCES

[1] F.R. Lin, L. Ferrucci, E.J. Metter, Y. An, A.B. Zonderman, and S.M. Resnick, "Hearing loss and cognition in the Baltimore Longitudinal Study of Aging," *Neuropsychology*, vol. 25, no. 6, pp. 763, 2011.

[2] S. Gordon-Salant and P.J. Fitzgibbons, "Recognition of multiply degraded speech by young and elderly listeners," *J Speech Lang Hear R*, vol. 38, no. 5, pp. 1150–1156, 1995.

[3] Working Group on Speech Understanding and Aging, "Speech understanding and aging," *J Acoust Soc Am*, vol. 83, no. 3, pp. 859–895, 1988.

[4] L.M. Heller, B. Elizalde, B. Raj, and S. Deshmukh, "Synergy between human and machine approaches to sound/scene recognition and processing: An overview of ICASSP special session," DOI:10.48550/ARXIV.2302.09719, 2023.

[5] S. Anderson, L. DeVries, E. Smith, M.J. Goupell, and S. Gordon-Salant, "Rate discrimination training may partially restore temporal processing abilities from age-related deficits," *J Assoc Res Oto*, pp. 1–16, 2022.

[6] S. Anderson, T. White-Schwoch, A. Parbery-Clark, and N. Kraus, "A dynamic auditory-cognitive system supports speech-in-noise perception in older adults," *Hearing Res*, vol. 300, pp. 18–32, 2013.

[7] A. Dryden, H.A. Allen, H. Henshaw, and A. Heinrich, "The association between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis," *Trends Hear*, vol. 21, 2017.

[8] S. Gordon-Salant and S.S. Cole, "Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing," *Ear Hearing*, vol. 37, no. 5, pp. 593–602, 2016.

[9] G. Livingston, A. Sommerlad, V. Orgeta, S.G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, J. Cohen-Mansfield, and C. Cooper, "Dementia prevention, intervention, and care," *Lancet*, vol. 390, no. 10113, pp. 2673–2734, 2017.

[10] F.R. Lin, "Hearing loss and cognition among older adults in the United States," *J Gerontol A-Biol*, vol. 66, no. 10, pp. 1131–1136, 2011.

[11] A. Parthasarathy, S. Romero Pinto, R.M. Lewis, W. Goedicke, and D.B. Polley, "Data-driven segmentation of audiometric phenotypes across a large clinical cohort," *Sci Rep-UK*, vol. 10, no. 1, pp. 1–12, 2020.

[12] R. Sanchez-Lopez, M. Fereczkowski, T. Neher, S. Santurette, and T. Dau, "Robust data-driven auditory profiling towards precision audiology," *Trends Hear*, vol. 24, 2020.

[13] S. Jain, V.K. Narne, N.P. Nataraja, S. Madhukesh, K. Kumar, and B.C. Moore, "The effect of age and hearing sensitivity at frequencies above 8 kHz on auditory stream segregation and speech perception," *J Acoust Soc Am*, vol. 152, no. 1, pp. 716–726, 2022.

[14] I. Yeend, E.F. Beach, and M. Sharma, "Working memory and extended high-frequency hearing in adults: Diagnostic predictors of speech-in-noise perception," *Ear Hearing*, vol. 40, no. 3, pp. 458–467, 2019.

[15] Accredited Standards Committee S3, Bioacoustics, "American national standard specification for audiometers," Tech. Rep. S3.6-2018, American National Standards Institute, June 2018.

[16] E.H. Rothauser, "IEEE recommended practice for speech quality measurements," *IEEE T Audio Speech*, vol. 17, pp. 225–246, 1969.

[17] M.C. Killion, P.A. Niquette, G.I. Gudmundsen, L.J. Revit, and S. Banerjee, "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," *J Acoust Soc Am*, vol. 116, no. 4, pp. 2395–2405, 2004.

[18] T.W. Tillman and R. Carhart, "An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test no. 6," Tech. Rep., Northwestern Univ Evanston Il Auditory Research Lab, 1966.

[19] Z.S. Nasreddine, N.A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J.L. Cummings, and H. Chertkow, "The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment," *J Am Geriatr Soc*, vol. 53, no. 4, pp. 695–699, 2005.

[20] P.D. Zelazo, J.E. Anderson, J. Richler, K. Wallner-Allen, J.L. Beaumont, K.P. Conway, R. Gershon, and S. Weintraub, "NIH toolbox cognition battery (CB): Validation of executive function measures in adults," *J Int Neuropsych Soc*, vol. 20, no. 6, pp. 620–629, 2014.

[21] S. Gatehouse and W. Noble, "The speech, spatial and qualities of hearing scale (SSQ)," *Int J Audiol*, vol. 43, no. 2, pp. 85–99, 2004.

[22] P.M. Rabinowitz, D. Galusha, M.D. Slade, C. Dixon-Ernst, K.D. Sircar, and R.A. Dobie, "Audiogram notches in noise-exposed workers," *Ear Hearing*, vol. 27, no. 6, pp. 742–750, 2006.

[23] R.R.A. Coles, M.E. Lutman, and J.T Buffin, "Guidelines on the diagnosis of noise-induced hearing loss for medicolegal purposes," *Clin Otolaryngol*, vol. 25, no. 4, pp. 264–273, 2000.

[24] A.S. Niskar, S.M. Kieszak, A.E. Holmes, E. Esteban, C. Rubin, and D.J. Brody, "Estimated prevalence of noise-induced hearing threshold shifts among children 6 to 19 years of age: The Third National Health and Nutrition Examination Survey, 1988–1994, United States," *Pediatrics*, vol. 108, no. 1, pp. 40–43, 2001.

[25] F. Charih, M. Bromwich, A.E. Mark, R. Lefrançois, and J.R. Green, "Data-driven audiogram classification for mobile audiometry," *Sci Rep-UK*, vol. 10, no. 1, pp. 1–13, 2020.

[26] J.R. Dubno, M.A. Eckert, F.S. Lee, L.J. Matthews, and R.A. Schmiedt, "Classifying human audiometric phenotypes of age-related hearing loss from animal models," *J Assoc Res Oto*, vol. 14, pp. 687–701, 2013.

[27] M.J. Van der Laan, E.C. Polley, and A.E. Hubbard, "Super learner," *Stat Appl Genet Mol*, vol. 6, no. 1, 2007.

[28] E. Polley, E. LeDell, C. Kennedy, and M. Van der Laan, *SuperLearner: Super Learner Prediction*, 2021, R package version 2.0-28.

[29] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.

[30] P.D. Allen and D.A. Eddins, "Presbycusis phenotypes form a heterogeneous continuum when ordered by degree and configuration of hearing loss," *Hearing Res*, vol. 264, no. 1-2, pp. 10–20, 2010.

[31] R. Sanchez-Lopez, T. Dau, and W.M. Whitmer, "Audiometric profiles and patterns of benefit: A data-driven analysis of subjective hearing difficulties and handicaps," *Int J Audiol*, vol. 61, no. 4, pp. 301–310, 2022.

[32] J.R. Dubno, D.D. Dirks, and A.B. Schaefer, "Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. II: Articulation index predictions," *J Acoust Soc Am*, vol. 85, no. 1, pp. 355–364, 1989.

[33] A. Presacco, J.Z. Simon, and S. Anderson, "Effect of informational content of noise on speech representation in the aging midbrain and cortex," *J Neurophysiol*, vol. 116, no. 5, pp. 2356–2367, 2016.