Individual differences in the perception of phonetic category structure predict speech-in-noise performance

Emily Myers,[1,2] Matthew Phillips,[1] Erika Skoe[1]

[1] *Department of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, CT, 06268, USA*

[2] *Department of Psychological Sciences, University of Connecticut, Storrs, CT, 06268, USA*

Speech sounds exist in a complex acoustic-phonetic space, and listeners vary in the extent to which they are sensitive to variability within the speech sound category ("gradience") and the degree to which they show stable, consistent responses to phonetic stimuli. Here we investigate the hypothesis that individual differences in the perception of the sound categories of one's language may aid speech-in-noise performance across the adult lifespan. Declines in speech-in-noise performance are well-documented in healthy aging, and are, unsurprisingly, associated with differences in hearing ability. Nonetheless, hearing status and age are incomplete predictors of speech-in-noise performance, and long-standing research suggests that this ability draws on more complex cognitive and perceptual factors. In this study, a group of adults ranging in age from 18 to 67 years performed online assessments designed to measure phonetic category sensitivity, questionnaires querying recent noise exposure history and demographic factors, and crucially, a test of speech-in-noise perception. Results show that individual differences in the perception of two consonant contrasts significantly predict speech-in-noise performance, even after accounting for age and recent noise exposure history. This finding supports the hypothesis that individual differences in sensitivity to phonetic categories mediates speech perception in challenging listening situations.

I. INTRODUCTION

Perception of the sounds of speech is a prerequisite for mapping the auditory signal onto meaning. Listeners need to detect and analyze the fine-grained spectral and temporal qualities of speech sounds, a process that is complicated by the presence of background noise. Yet, listeners do not detect and analyze speech sounds in precisely the same way. Individual differences in perception of phonetic detail have been well-documented and linked to other aspects of language processing (Fuhrmeister et al., 2023; Kapnoula et al., 2017; Kong & Edwards, 2016). Of interest is how individual differences in phonetic sensitivity are related to speech perception-in-noise (SPIN) performance. SPIN declines are well-documented in aging, and crucially, these are not fully explained by differences in peripheral hearing (e.g. Goossens et al., 2017). This leads to the possibility that individual differences in sensitivity to the properties of speech categories might partially account for differences in SPIN, especially those that emerge as a function of aging.

In this study we aimed to answer three questions about individual differences in the perception of phonetic category structure. First, we asked whether tasks of phonetic category sensitivity measured by two-alternative forced choice (2AFC), visual analogue scale (VAS), and AX discrimination (AX) tasks tap individual differences in shared skills in perception and representation of phonetic categories, and further whether these skills are phonetic contrast-specific or reflect a general trait of the individual. Second, we evaluated age-related changes to phonetic category sensitivity. Finally, we asked to what extent individual differences in performance on these tasks predicts performance on a speech-in-noise task, after accounting for age and recent noise exposure.

A. Individual differences in the perception of phonetic category structure

Classic studies of categorical perception (A. M. Liberman et al., 1957) established that when listeners are asked to identify sounds drawn from a phonetic continuum, they will typically show a sharp boundary between categories, exhibiting a steep psychometric function. More notably, listeners also show asymmetric patterns of discrimination, with better discrimination of sound contrasts that span the category boundary than those that fall within the category, leading to the proposal that listeners are either insensitive to variability within the category, or that this information is discarded as phonemes and words are identified. These discontinuities, or warping in sensitivity according to phonetic category structure, led to the description of phonetic perception as "categorical."

Nonetheless, researchers have long noted that listeners are quite sensitive to within-category phonetic detail (McMurray et al., 2002; Myers, 2007; Pisoni & Tash, 1974; Toscano et al., 2010), and use within-category variability when accessing the lexicon (Andruski et al., 1994; McMurray et al., 2009; Sarrett et al., 2020). Of interest, when performing behavioral tasks assessing sensitivity to phonetic detail, listeners show individual differences in the gradience or categoricity of phonetic sensitivity. As discussed thoroughly elsewhere (Apfelbaum et al., 2022; McMurray, 2022), tasks vary in the extent to which they encourage or afford listeners the option of demonstrating sensitivity to phonetic gradience. 2AFC tasks (e.g., "do you hear 'da' or 'ta'?") force listeners into a binary decision, such that perception of variability might be masked. As pointed out by Apfeblaum et al., (2022), a well-defined boundary between phonetic categories (characterized by a steep slope in the categorization function) in this task does not necessarily entail that listeners cannot detect variation within the category. AX discrimination tasks may have more power to detect sensitivity to within-category detail; in these tasks, listeners are asked to decide whether two items from the same continuum are the same or different, and responses can be made without reference to any specific category label. Visual analogue scale (VAS) measures of phonetic sensitivity have been argued to provide some of the attributes of 2AFC and discrimination tasks. In this task, listeners are asked to rate tokens along a scale in terms of their fit to the category (Kong & Edwards, 2016). Even among typical listeners, substantial variability has been found in sensitivity to phonetic category structure (e.g., Fuhrmeister et al., 2023; Fuhrmeister & Myers, 2021; Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021; Kong & Kang, 2023), with some listeners showing a more graded pattern of sensitivity, and others showing a more categorical response function.

Individual differences in graded perception (as measured by the VAS) have some functional consequences for online language comprehension. Gradient listeners tend to use more secondary cues to phonetic perception (Kapnoula et al., 2017, 2021; Kong & Edwards, 2016), and gradience may aid online lexical access, particularly recovery from misidentification of words in a "lexical garden path" paradigm (Kapnoula et al., 2021). Individual differences in gradience can be seen quite early in the auditory processing stream, such that gradient listeners show correspondingly gradient patterns of neural responses to voice onset time (VOT) in the N1 EEG component (Kapnoula & McMurray, 2021). However, it remains unclear if patterns of gradience in the VAS task are characteristics of the listener, or are particular to the way that listener processes some very specific acoustic-phonetic cues but not all (e.g., Kapnoula et al., 2017, Kapnoula & McMurray, 2021, Fuhrmeister et al., 2023). Finally, the notion that gradience *per se* reflects generally better phonetic processing has not, of yet, been strongly supported. Gradience has not been shown to correlate well with speech-in-noise performance (Kapnoula et al., 2017, 2021), nor with perception of non-native contrasts (Fuhrmeister et al., 2023).

In addition to the dimension of gradience, listeners also differ in the degree to which they show trial-to-trial consistency in rating phonetic tokens (Fuhrmeister et al., 2023; Fuhrmeister & Myers, 2021; Kapnoula et al., 2017). Notably, some listeners show gradient perceptual patterns alongside highly consistent responses to each token on the continuum, whereas others show the same gradient function but much more stochastic or inconsistent responses to individual tokens. This notion of "response consistency" resonates with theories proposing that there are downstream consequences for individual differences in the stability of auditory encoding arising early in the auditory processing stream (Centanni et al., 2018; Hornickel & Kraus, 2013; Neef et al., 2017; Tecoulesco et al., 2020). Indeed, consistency of brainstem and early cortical responses to repeated auditory tokens differs in people with a history of language disorder, and may be modulated by auditory expertise (Krizman et al., 2014; Skoe & Kraus, 2013). Response consistency in the VAS task for both stop and fricative continua is linked to individual differences in the structure of the bilateral transverse temporal gyri (Fuhrmeister & Myers, 2021), a structure responsible for early cortical processing of sound. Further, individuals with higher response consistency on a VAS task were more adept at discriminating an unfamiliar non-native sound contrast (Fuhrmeister et al., 2023; Honda et al., 2024), suggesting that stability in the mapping between the auditory input and the perceptual response may allow listeners to tune into the unfamiliar acoustic details that signal non-native contrasts.

Research thus far corroborates that individual differences in phonetic judgments do reflect meaningful differences in how they process the speech signal. Nonetheless, several pertinent questions remain that we address in this study. First, while AX discrimination was classically used to establish patterns of categorical perception (A. M. Liberman et al., 1957), it has not yet been directly compared to the VAS task. If gradience in the VAS taps individual differences in fine-grained sensitivity to acoustic detail, then AX patterns should correspond to VAS patterns, such that those with more gradient VAS functions should show better ability to detect differences between tokens, especially those falling within the phonetic category. 2AFC tasks, while also a popular option for studies of phonetic category structure, have been argued to underestimate an individual's ability to detect within-category differences by forcing a binary response (e.g. Apfelbaum, et al., 2022). Prior studies comparing slope on the 2AFC task and responses on the VAS task suggest that slope of the function in 2AFC is more related to response consistency than gradience (e.g. Kapnoula et al., 2017). Finally, the jury is still out on whether or not gradience and response consistency are a property of individuals or specific phonetic contrasts. By understanding the relationships between these measures, we are able to answer how phonetic sensitivity changes during aging, and how, if at all, these measures relate to speech-in-noise performance.

B. Changes in sensitivity to phonetic category structure as a function of aging.

During healthy aging, changes in hearing are nearly inevitable (Goman & Lin, 2016), with more than 25% of adults having mild-to-moderate hearing declines by the age of 70. Even among those with relatively intact hearing as measured by the pure-tone audiogram, differences in access to the speech signal can be stark, especially for noise- masked speech (e.g., Goossens et al., 2017). Of interest, speech-in-noise performance is only moderately predicted by pure-tone hearing assessments in aging, suggesting that age-related changes extend beyond the auditory periphery to include the neural systems involved in sound-to-meaning mapping (Anderson et al., 2011; Goossens et al., 2017; Prendergast et al., 2019). Changes in sensitivity to phonetic category structure have been investigated during childhood and adolescence (McMurray et al., 2018), with evidence showing increasingly gradient sensitivity as children gain experience with their native language (see McMurray, 2022 for review). Comparing older and younger adults in 2AFC tasks, older adults have been reported to show shifted boundary locations for stop consonants, a fricative/affricate contrast, and a stop-glide contrast (Baum, 2003; Dorman et al., 1985; Gordon-Salant et al., 2006). These findings might reflect changes in sensitivity or resolution of certain types of cues, especially those that rely on temporal distinctions (Gordon-Salant et al., 2006). Notably, however, the slope of these functions is quite stable across age (Dorman et al., 1985; Gordon-Salant et al., 2006), suggesting that although older adults may rely on somewhat different cues, on balance, categorization decisions remain stable among older adults with hearing within normal limits. Mattys & Scharenborg (2014) also incorporated an AX discrimination task on a nasal contrast, showing that older adults were somewhat less sensitive across the continuum, but age-related differences were not stark. To our knowledge, no studies have investigated changes in gradient phonetic perception using a VAS task as a function of age across the adult lifespan.

We can imagine several patterns that might be associated with aging. First, if age-related declines in peripheral and central auditory function result in less neural stability in the auditory system (Skoe et al., 2015), we might observe decreased behavioral response consistency in the VAS, poorer sensitivity to subtle acoustic differences in AX discrimination, and a flattening of the categorization function in the 2AFC task. Second, age-related hearing threshold changes tend to affect higher frequencies first, which might lead to less sensitivity to specific contrasts that are distinguished by high-frequency information, for instance the fricative s-sh contrast used in this study. Notably, however, language ability is among the best-preserved functions during healthy aging (Ansado et al., 2013; Diaz et al., 2021) and increased experience with a language over one's lifespan might actually serve to fine-tune and stabilize native phonetic category representations, leading to the opposite patterns from the ones described above.

C. Consequences of individual differences in phonetic perception for speech-in-noise processing.

Comprehension of speech in noise is cognitively and perceptually demanding (Peelle, 2018). Understanding speech in a noisy environment depends not only on the audibility of the signal but also on attention, working memory, and a host of other capacities that help the listener direct attention to the most relevant portions of the acoustic signal (Pichora-Fuller et al., 2016). It is less well-understood how individual differences in sensitivity to phonetic category structure (e.g., perception of small differences within the category; consistent perceptual responses to speech) might play out in speech-in-noise processing. In theory, a listener who is sensitive to fine-grained details of speech may be better equipped to detect these properties when mixed with noise.

Similarly, a listener with greater consistency in their perceptual response to speech may be able to calibrate to noise levels more accurately in service of separating the speech signal from noise.

As described above, evidence thus far linking gradient phonetic perception to speech-in-noise performance has been weak (Kapnoula, et al., 2017) or absent (Kapnoula et al., 2021). However, perception of speech-in-noise was not the primary goal of previous studies, and the more limited age range in this prior work might limit variability in speech-in-noise performance to the extent that an association would be difficult to find. In the current study, we collected data from adult participants performing 2AFC, VAS, and AX discrimination tasks on two phonetic contrasts, a stop place of articulation contrast ('ba'--'da'), and a fricative place of articulation continuum ('sign'--'shine'). Our expanded age range (18-67) also allowed us to tap greater variability in speech-in-noise performance and we controlled for exposure to environmental noise over the previous 12 month window, given that experience in noisy environments is linked to speech-in-noise ability (M. C. Liberman, 2017; Prendergast et al., 2019; Skoe et al., 2015; but cf. Shehabi et al., 2022). We predicted that individual differences, especially in response consistency, but potentially also discrimination accuracy, would be related to differences in speech-in-noise performance after accounting for age and noise exposure.

This dataset allows us to pursue three questions. First, we ask how 2AFC, VAS, and AX performance relate to each other, specifically testing the hypothesis that discrimination as measured by AX will correlate with slope in the VAS task, and asking whether individual differences in phonetic tasks cluster by phonetic contrast. Second, we ask how behavior on phonetic tasks changes over the course of aging. Finally, we ask whether (and which) phonetic tasks best predict speech-in-noise performance.

II. METHODS

A. Participants

Participants were recruited from the online recruitment platform, Prolific, for online testing. The study was advertised to adult participants who reported being native, monolingual speakers of English and living in the US. Subjects gave informed consent according to the guidelines of the UConn Institutional Review Board and were compensated $10 per hour for their participation.

Participants were recruited in five age bands from 18-67 years, and data collection continued until each age band contained at least 19 usable participants. 143 participants completed all study procedures. Data quality checks (see below, Phonetic Decision Tasks) led to the elimination of 17 participants. Another 10 participants were excluded for failing the headphone check (see below, First Steps and Headphone Check) resulting in 116 participants whose data ultimately contributed to subsequent analyses (female=74, male=42; see Table I for complete participant demographics). Participants recruited from Prolific often have substantial experience participating in behavioral studies. Our participants were no exception: data extracted from Prolific indicates that on average, our subject pool has been approved for completing an average of 457 studies on Prolific, with high participant ratings.[1]

TABLE I: Demographic characteristics of the sample that contributed to all subsequent analyses. Speech-in-noise performance score represents the signal-to-noise ratio (SNR) at which 50% of the key words are correctly repeated, with lower scores indicating better performance. Refer to text for descriptions of the noise exposure metrics.

By age bands:	**18-27**, N = 26[1]	**28-37**, N = 19[1]	**38-47**, N = 25[1]	**48-57**, N = 24[1]	**58-67**, N = 22[1]
Age (yrs)	23 (21, 25)	30 (29, 33)	42 (39, 44)	52 (50, 55)	62 (60, 64)
Sex					
Female	17 (65%)	7 (37%)	13 (52%)	20 (83%)	17 (77%)

Male	9 (35%)	12 (63%)	12 (48%)	4 (17%)	5 (23%)
Childhood caregiver education, (yrs)	14 (12, 16)	14 (13, 16)	14 (12, 16)	12 (12, 15)	12 (12, 16)
Speech-in-noise score	1.13 (0.00, 2.19)	0.25 (-0.50, 1.25)	1.00 (0.25, 3.50)	1.00 (0.75, 2.44)	1.50 (0.56, 3.69)
Annual Noise Exposure (ANE) Estimate (dB LAeq8760h)	71.6 (69.6, 74.6)	70.6 (68.8, 77.1)	70.1 (67.5, 72.9)	70.0 (65.7, 75.8)	66.1 (64.7, 70.9)
Noise Exposure Dose (%)	18 (12, 37)	14 (9, 66)	13 (7, 25)	12 (5, 48)	5 (4, 17)

[1]Median (IQR); n (%)

B. Procedure

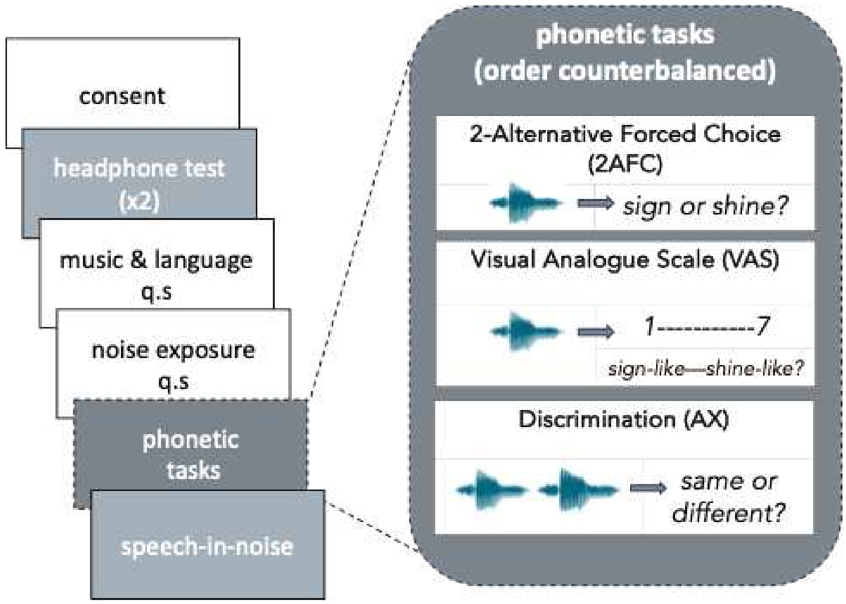

FIG 1. Task schematic. Participants performed tasks from top to bottom according to the left-hand column. Phonetic tasks were conducted for both ba-da and sign-shine continua, using a counterbalanced Latin squares design (see text for details).

1. First Steps and Headphone Check

A schematic of the study procedures can be found in Figure 1. Study participants were required to use either a laptop or desktop computer (i.e., no mobile devices), and were instructed to wear headphones. After providing informed consent, participants were directed to the online experimental software platform, Gorilla (www.gorilla.sc; Anwyl-Irvine et al., 2020). First, participants were directed to a headphone check described by Woods et al. (2017). Participants were instructed to initially set their volume to approximately 25%, listen to a burst of white noise, and then adjust their computer's volume until it was a comfortable listening level. Participants were then instructed to listen to three tones of various intensities and select which tone was the softest. This headphone check uses phase cancellation such that participants would only perceive the softest tone as being the softest if they were wearing headphones. If a participant passed the headphone check (i.e., selected the correct tone in at least four out of six trials), they continued on with the study and completed a series of questionnaires. If a participant failed the headphone check (i.e., selected the correct tone in less than four trials), they were reminded that it was important to wear headphones, and then completed the headphone check a second time. If a participant failed the headphone check a second time, they were allowed to continue with the experiment, but their data were excluded from subsequent analyses.

2. Questionnaire Data

Next, participants were directed to a series of questionnaires to collect basic demographic data, experience with musical training, experience with languages other than English, and the Noise Exposure Questionnaire (Johnson et al., 2017). Data on musical experience and language backgrounds are beyond the scope of the current investigation.[2]

The Noise Exposure Questionnaire (NEQ) is a short survey developed as a low-cost and rapid way to estimate environmental noise exposure risk. The NEQ estimates annual noise exposure based on self-reported frequency engaging in noisy activities (e.g., attending events with amplified music, riding motorized vehicles, using power tools, wearing personal listening devices, and playing a musical instrument) during the past 12 months. Annual noise exposure (ANE) is estimated using representative sound levels from the literature for each activity type. ANE is expressed in dB LAeq8760h, and represents the continuous sound level averaged over 8760 (24 hours x 365 days) hours using a 3-dB exchange rate and A-weighted sound levels. Refer to Johnson et al., (2017) for details. From the dB estimate, a noise dose is then derived, with 79 dB LAeq8760h corresponding to the National Institute for Occupational Safety and Health (NIOSH) recommended exposure limit, i.e., 100% dose. Doses above 100% place the listener at increased risk of noise-induced hearing loss. For the purposes of interpreting the NEQ data, it is important to note that the online data collection occurred between November 11, 2020 and February 4, 2021.

3. Phonetic Decision Tasks

Immediately before completing the phonetic decision tasks, participants were given the opportunity to adjust their volume. Participants were presented with an audio token at the same intensity of the phonetic stimuli and were instructed to adjust their volume until it was "comfortably loud" and they could "hear the sound easily." Participants completed three different phonetic decision tasks: a two-alternative forced choice task (2AFC), a discrete version of the visual analogue scale task (VAS), and an AX discrimination task (AX). Participants heard stimuli drawn from a voiced stop continuum ("ba" to "da") as well as a fricative place-of-articulation continuum ("sign" to "shine"). Tasks were presented in a fixed sequence, with the task schedule rotated using a Latin squares procedure such that across participants, the task that participants completed first (i.e. the one that begin the sequence) was counterbalanced. Namely, given the task order represented as ABCDEF, participants were counterbalanced across orders ABCDEF, BCDEFA, CDEFAB, etc. The fixed task order was: VAS: ba-da, VAS: sign-shine, 2AFC: ba-da, 2AFC: sign-shine, AX: ba-da, AX: sign-shine. This ordering meant that participants almost always performed the ba-da version of the task before the sign-shine version of the task. Below we describe the characteristics of the phonetic stimuli as well as the specific tasks.

Phonetic continua. A seven-point continuum from /ba/ to /da/ was synthesized at Haskins Laboratories using a Klatt synthesizer. This continuum manipulates the trajectory of the first and second formants, and the vowel information after the initial short transition is shared across all stimuli (see Supplementary Materials for details). A continuum from "sign" to "shine" was created by modifying naturally-produced tokens of "sign" and "shine." Stimuli were produced by a female, native speaker of English, and the initial fricative was excised. Blends of the excised /s/ and /sh/ tokens were created through waveform averaging using Praat (Boersma & Weenink, 2013) to create blends from 80% /s/ to 20% /s/ in 10% steps, and fricative blends were re-concatenated onto to the original “-ign” file, resulting in a seven-point perceptual continuum extending from “sign” to “shine.” Stimuli were selected such that no more than two tokens on each end of the continuum received fairly unambiguous judgements, in order to optimize sampling of the more variable responses to tokens approaching the category boundary.

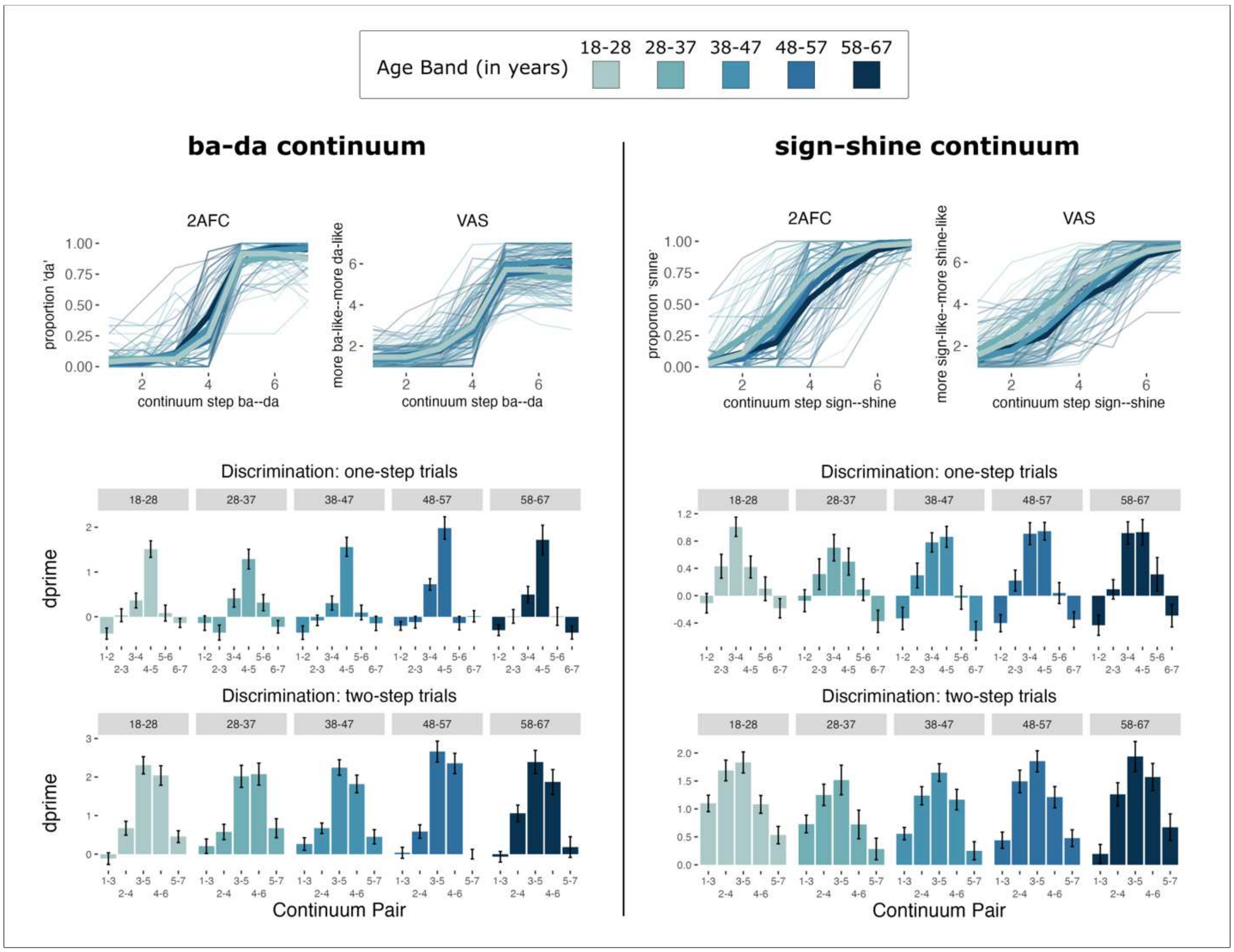

FIG 2. Behavioral data for phonetic tasks, grouped by age band. Left panel shows data for the ba-da continuum, right panel shows data for the sign-shine continuum. For 2AFC tasks, the y axis indicates proportion, for VAS tasks, this axis indicates the rating position between the two ends of the continuum, and for discrimination, the units displayed are d’ values. Error bars indicate standard error of the mean.

Two-alternative forced choice (2AFC). Participants heard 15 instances of each point along the seven-point continuum, presented in random order, for a total of 105 trials per continuum. For each token, the listener was asked to categorize the token (e.g., "ba" or "da"?) by pressing a corresponding button on the keyboard. The dependent measure was the participant response for each token. To ensure that participants perceived the endpoints of the continuum at above-chance levels, only participants who correctly categorized endpoint tokens at least 60% of the time were included in the study. This led to the exclusion of nine participants on the basis of the ba-da continuum, and one additional participant on the basis of the sign-shine continuum. Individual data and mean response curves by age band are plotted in Figure 2.

Visual analogue scale (VAS). Participants completed a "discretized" version of the visual analogue scale task (cf. Kapnoula et al., 2017, Fuhrmeister et al., 2023). In the original version of the VAS, participants are asked to rate each token from "most {ba/sign}-like" to "most {da/shine}-like" along a continuous scale by moving a slider. In our version of the task, adapted for easier online administration, participants instead rated tokens along a seven-point numeric scale. Participants heard 15 examples of each point on the phonetic continuum, presented in random order, for a total of 105 trials per continuum. Since there was no in-principle "correct" answer for this task, data quality checks ensured that participants showed some difference in rating tokens across the continuum. To pass this quality check, a participant had to demonstrate a mean difference of two points along the rating scale for any two continuum tokens for each continuum. This resulted in the exclusion of an additional seven participants on the basis of performance on the ba-da continuum (five additional participants had poor VAS data but had already been excluded on the basis of quality checks for the 2AFC task). Figure 2 displays individual response curves by continuum as well as mean response curves aggregated by age band.

AX discrimination (AX). Participants heard two tokens drawn from the seven-point continuum per trial, separated by a 1000 msec ISI.[3] Stimuli were either identical ("Same" trials, e.g. ba1-ba1, n=10 per pair), separated by one step on the continuum (e.g., ba1-ba2, "One-step" n=10 per pair) or two steps on the continuum (e.g., ba1-ba3, "Two-step", n=10 per pair). Pairs were presented in both orders (e.g., ba1-ba3 and ba3-ba1) collapsing across orders for analysis purposes. Participants completed a total of 180 discrimination trials per continuum. Data were transformed into d' scores by subtracting z-scored rates of hits for each different trial from z-scored rates of false-alarms for "same" trials. Figure 2 displays d' scores for one-step and two-step trials, aggregated by age band, for each phonetic continuum.

4. Speech-In-Noise Test

Participants were administered a modified version of the Quick Speech-in-Noise test (*QuickSIN^{TM} Speech-in-Noise Test, Etymotic Research, Inc.)*. In this test, participants listened to 24 fixed-level low-context sentences spoken in varying degrees of four-talker babble noise (i.e., signal-to-noise ratio, SNR), ranging from 25 dB (the easiest SNR level) to 0 dB (the hardest SNR level) in 5 dB intervals. When the QuickSIN is used in clinical settings, patients verbally repeat each sentence; in our modified online version of the test, participants were asked to "repeat" the sentence back verbatim by typing into a text response field, and then press the enter key once they were finished to advance to the next sentence. Modeling the clinical protocol, sentences were divided into four lists with six sentences, and each sentence within a list was presented at a different, descending SNR level. QuickSIN lists 1-4 were selected. Within each list, trials were presented in a fixed order of increasing difficulty, such that the first and last sentence within each list had a SNR level of 25 dB and 0 dB, respectively. Each sentence contained five keywords worth one point each; therefore, participants could earn a maximum of five points per sentence and a maximum of 30 points per sentence list, based on each keyword correctly repeated. The total score for each sentence list was subtracted from 25.5 to calculate a participant's SNR loss. The SNR loss represents the SNR at which 50% of keywords can be accurately repeated. Each participant's average QuickSIN score was then calculated by averaging their SNR loss across all four sentence lists, with higher scores indicating poorer performance.

Participants completed two practice QuickSIN sentences to familiarize themselves with the task (one practice sentence at 25 dB, the other at 5 dB) and adjusted their volume prior to completing the 24 main trials. Participants were instructed not to adjust their volume after the practice sentence trials. The test was scored using automatic routines, then manually checked. Speech-in-noise responses were scored automatically in R (R Core Team, 2023) to detect if each keyword was present in a participant's response, regardless of letter case. Each participants' response received a score of "0" if the keyword was not present in their response and a score of "1" if the keyword was present. After automatic scoring, speech-in-noise data were then manually checked by one of the authors to validate the automatic scoring and to rescore any unambiguous typos or homophone substitutions (e.g., typing "wait" instead of "weight" or "steal" instead of "steel") as correct. Homophones were marked as correct even when they produced a semantically or syntactically anomalous sentence, given that our primary interest was in the acoustic access to the signal. We note also that when the QuickSIN is administered under standard clinical conditions using a verbal response that the rater would be unaware of homophonic substitutions as they would by definition sound the same to the rater. After scoring keywords, the average speech-in-noise score was calculated as described above.

C. Analysis approach

All analyses were carried out using R (R core team).

1. Summary individual differences measures for each phonetic task.

To characterize individual differences in the perception of stop and fricative continua, we computed several summary measures for each participant and continuum. For the 2AFC task, we used a two-parameter logistic regression to estimate the *slope* of the categorization curve at the inflection point for each continuum and participant. Following prior work, we estimated two measures for the VAS task, the *slope* and *response consistency* for each participant and continuum (see Fuhrmeister et al., 2023). The slope was estimated by fitting a four-parameter logistic regression to estimate the minimum, maximum, inflection point (boundary), and slope of the response function for each participant and continuum. Response consistency is estimated by taking the mean of the squared residuals for each response for each subject, and can be thought of as a measure of the fit of the raw data to the estimated response function. This value is multiplied by -1 so that the lowest values reflect low consistency and higher values reflect higher consistency. For discrimination data, for each participant and continuum, we calculated a mean *sensitivity* score by averaging all d' values for both one-step and two-step trials, intended to capture general sensitivity to contrasts across the entire continuum. We also wished to capture the asymmetry in discrimination of near-boundary pairs vs. within-category pairs that is a hallmark of categorical perception. Because the precise estimation of the location of the individual phonetic category boundary (i.e., by using the psychometric function for the VAS or 2AFC task) can be unreliable if the participant has an atypical or noisy response function, we opted to calculate this measure by subtracting the d' value for the worst-discriminated one-step pair from the best-discriminated one-step pair, wherever that pair fell along the continuum. We refer to this measure as the *categoricity* measure. Notably, for the vast majority of participants, the best-discriminated pair was in the boundary region (involving a token that falls close to the boundary for that contrast), and the worst-discriminated pair tended to be distant from the boundary. Each participant therefore had five distinct phonetic scores for each of the two continua (2AFC slope, VAS slope, VAS response consistency, AX sensitivity, AX categoricity). These data were joined with measures from the demographics and questionnaire data, namely age in years, caregiver education in years, annual noise exposure (ANE), and the speech-in-noise score (expressed as SNR Loss).

2. Outlier removal and imputation.

Outliers were defined as any score that fell more than 2.5 SD from the group mean. This resulted in removal of 41 values from the dataset, or 1.7% of the total data. Missing values were replaced by imputation using the *mice* package (Buuren & Groothuis-Oudshoorn, 2011) and the predictive mean mapping (PMM) method to multiply-estimate missing values.

III. RESULTS

A. Relationships between measures of phonetic category sensitivity.

To characterize relationships between phonetic measures, Pearson correlations between all ten measures (five different measures, two continua) were calculated (Figure 3). Every measure showed a significant relationship with at least one other measure; notably all measures for the ba-da continuum were correlated at a level of at least $p<0.05$ (uncorrected for multiple comparisons, correlations between measures taken on the same phonetic contrasts are highlighted within the dashed boxes), but correlations within the sign-shine measures and between phonetic contrasts were more mixed. Discrimination metrics for ba-da correlated not only with both VAS measures but also 2AFC slope, whereas for sign-shine, discrimination categoricity and mean sensitivity were related to 2AFC slope and mean sensitivity was related to VAS consistency, but no relationships with VAS slope were detected.

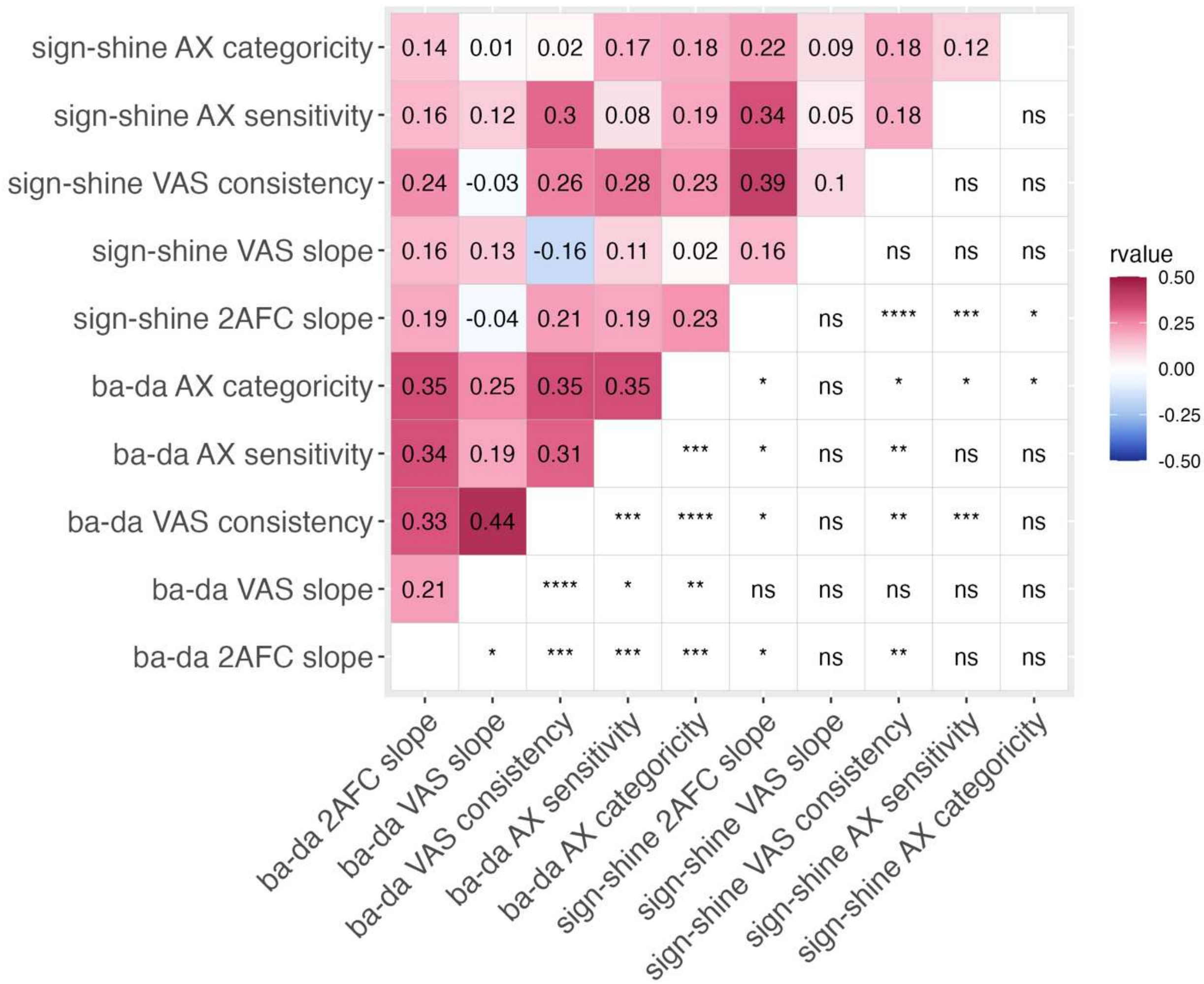

FIG 3. Correlations between all phonetic decision measures. Upper triangle displays Pearson correlations, lower triangle displays significance codes for p-values, with p<0.05=*, p<0.01=, p<0.001=***, p<0.0001=****, uncorrected for multiple comparisons.**

To address the question of the relationships between these measures in a more principled way, we performed a confirmatory factor analysis, comparing two models using the *lavaan* package in R (Rosseel, 2012). In the One-Factor model, all behavioral measures loaded on one latent variable which we term “phonetic skill.” This was compared to the Contrast-Specific model where two separate latent variables were constructed (“ba-da” and “sign-shine”), such that behavioral measures for each phonetic contrast load on separate latent variables (see Figure 4). Models that maximize the comparative fit index (CFI) and minimize Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are judged to be better-fitting. Model fit estimates suggested that the Contrast-Specific model was a better fit to the data (One-Factor Model: CFI=0.779, AIC=2724.5, BIC=2779.6; Contrast-Specific Model: CFI=0.888, AIC=2709.6, BIC=2767.4). This was confirmed by performing a chi-squared test comparing the two models; here the two-factor model was a significantly better fit to the data ($\chi 2= 16.95$, $p<0.001$). Significant loadings for the Two-Factor model are displayed in Figure 4.

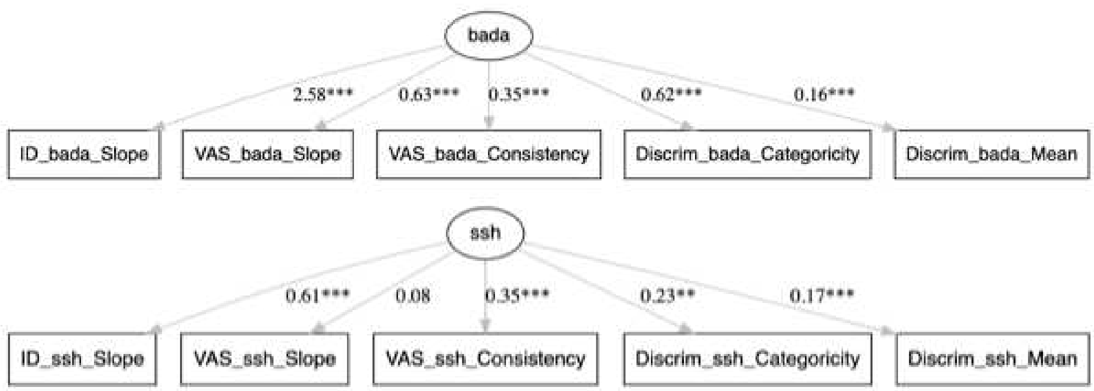

FIG 4. Results of a confirmatory factor analysis, constructed with two latent variables, one for sign-shine decisions (ssh) and the other for ba-da decisions (bada). Phonetic decision measures load on phonetic contrast-specific latent variables. Loadings displayed for all paths, (*)s indicate significance values.

B. Differences in sensitivity to phonetic category structure as a function of aging

Changes in sensitivity to phonetic category structure as a function of aging were evaluated by entering all phonetic measures into one model to predict age. Using the *lme4* package in R (Bates et al., 2014), we constructed a linear model in which all ten phonetic measures were entered to predict age (in years). Given the mild collinearity between measures (see Fig. 3), we used the *step* function in the *lmerTest* package (Kuznetsova et al., 2017) to iteratively remove predictors from the model that do not significantly contribute to model fit. The resultant model (Table II) contained three surviving predictors: 2AFC slope for the ba-da continuum, VAS consistency for the ba-da continuum, and VAS consistency for the sign-shine continuum. Of these, only VAS consistency for the sign-shine continuum was significant, with lower consistency associated with advancing age. In general, phonetic factors accounted for a small proportion of the variance in age (adjusted R^2=0.058, F(3,112=3.39, p=0.021). Results of the full model are displayed in the Supplementary Materials.

TABLE II. Best-fit linear model predicting age from all ten phonetic decision measures.

Predictor	β	95% CI	t	df	p
Intercept	37.18	[29.17, 45.18]	9.20	112	< .001
ba-da 2AFC slope	0.51	[-0.01, 1.03]	1.94	112	0.055
sign-shine VAS consistency	-4.43	[-8.65, -0.21]	-2.08	112	0.040*
ba-da VAS consistency	3.67	[-0.99, 8.32]	1.56	112	0.122

C. Predictors of speech-in-noise performance

Thus far, analyses show that there are mild associations between phonetic measures, especially between measures assessed on the same contrast, and that in general, these phonetic measures are not strongly related to age. Next we asked whether individual differences in phonetic measures predict speech-in-noise performance, together with other potentially explanatory factors (age, noise exposure, and childhood caregiver education, a proxy for socio-economic status that has been suggested to be predictive of language ability, e.g., Calvo & Bialystok, 2014). We approached this question in two ways, first by entering all measures into the same model, and second by using a PCA approach to summarize phonetic scores for use in the regression.

First, we built a model to predict scores on the speech-in-noise test based on all ten of the phonetic measures, as well as age, annual noise exposure, and childhood caregiver education in years. As above, we used a backwards-stepping approach in the *step* function in *lmerTest* to drop low-performing predictors from the model. The final model results are displayed in Table III (adjusted R^2=0.25, F(7,108)=6.52, p<0.001). Notably, five phonetic measures survive in this model, (VAS consistency and AX categoricity measures, for both continua, as well as ba-da 2AFC slope), in addition to age and noise exposure. Model results before model selection are reported in the Supplementary Materials.

TABLE III. Best-fit linear model predicting speech-in-noise score (expressed as SNR loss), from age, caregiver education, noise exposure, and all ten phonetic decision measures, best-fit model after backwards-stepping procedure.

Predictor	β	**95% CI**	t	df	p
Intercept	-7.54	[-13.18, -1.89]	-2.65	108	0.009
Age	0.04	[0.01, 0.06]	2.88	108	0.005*
Annual Noise Exposure (ANE)	0.07	[0.01, 0.14]	2.27	108	0.025*
ba-da VAS slope	0.27	[-0.01, 0.56]	1.88	108	0.062
ba-da VAS consistency	-0.78	[-1.51, -0.05]	-2.12	108	0.037*
sign-shine VAS consistency	-0.53	[-1.14, 0.09]	-1.70	108	0.092
ba-da AX categoricity	-0.69	[-1.10, -0.28]	-3.33	108	0.001*
sign-shine AX categoricity	0.64	[0.10, 1.18]	2.34	108	0.021*

Second, acknowledging the degree of overlap between our phonetic measures, we performed a principal components analysis (PCA) using singular value decomposition on all ten phonetic measures using the *prcomp* function as part of the *stats* package, provided in base R (R Core Team, 2023). Visualizing the top five dimensions (see Supplementary Materials for a table depicting all loadings, Figure 5 for a visualization of the loadings), we see that dimension 1, accounting for 28.1% of the variance, contains loadings from nearly all phonetic measures, reflecting a high degree of overlap between most measures. First, we constructed a base model to predict speech-in-noise performance using age, caregiver education in years, and noise exposure only. Model comparison using the *anova* function in the base R package (R Core Team, 2023), and showed that adding the top five phonetic dimensions extracted from the PCA significantly improved model fit ($F(5)=5.3939$, $p<0.0005$). Specifically, PCA dimensions 1, 3, 4, and 5, which have fairly heterogeneous loadings from most phonetic measures, were all significant predictors of speech-in-noise, even after accounting for demographic factors. (adjusted $R^2=0.23$, $F(8,107)=5.21$, $p<0.001$, Table IV).

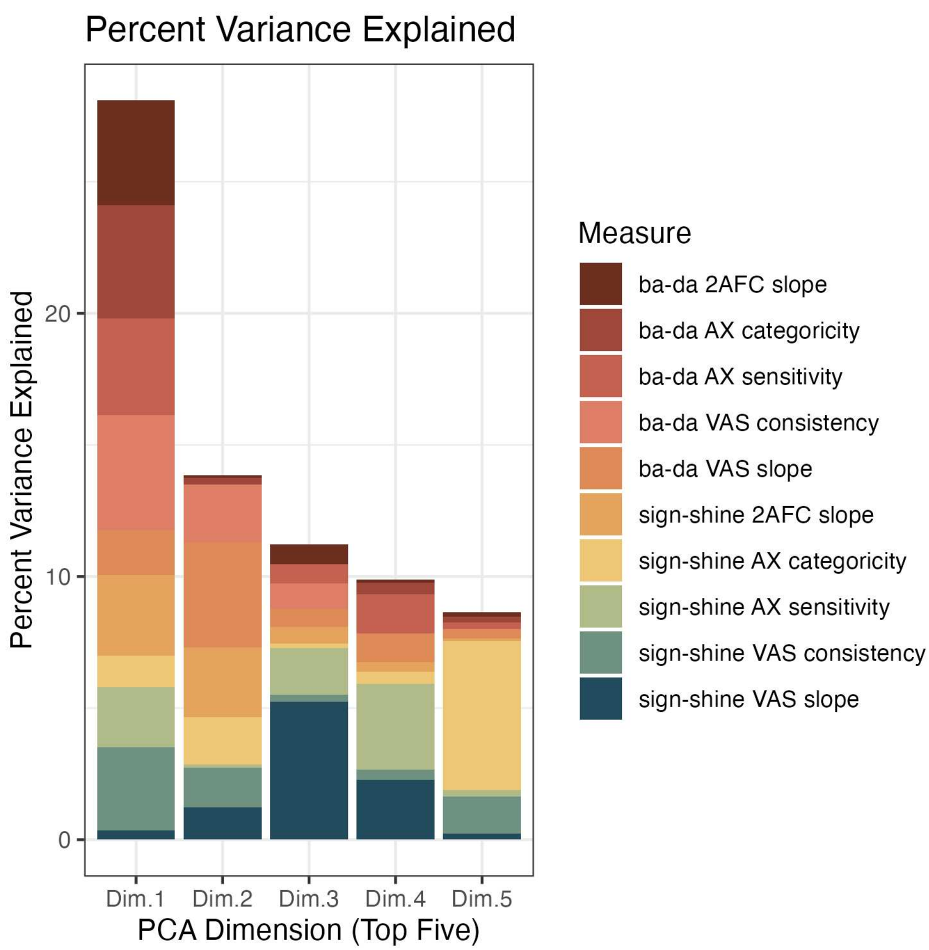

FIG 5. Loadings on each dimension in the PCA analysis of the ten phonetic decision measures. Overall height of the bar displays the percent variance explained by each dimension. Colors within the bar show the proportion of each dimension composed of each corresponding measure.

TABLE IV. Linear model predicting speech-in-noise performance score (expressed as SNR loss) from age, caregiver education, noise exposure, and the top five dimensions identified by subjecting the ten phonetic decision measures to PCA.

Predictor	β	95% CI	t	df	p
Intercept	-4.66	[-9.97, 0.64]	-1.74	107	0.084
Age (yrs)	0.03	[0.01, 0.06]	2.39	107	0.018*
Annual Noise Exposure (ANE)	0.09	[0.02, 0.16]	2.68	107	0.008*
Caregiver education (yrs)	-0.10	[-0.25, 0.05]	-1.35	107	0.180
Dimension 1	-0.35	[-0.57, -0.12]	-3.08	107	0.003*
Dimension 2	-0.16	[-0.47, 0.16]	-0.99	107	0.326
Dimension 3	-0.43	[-0.78, -0.08]	-2.44	107	0.016*
Dimension 4	0.40	[0.03, 0.78]	2.16	107	0.033*
Dimension 5	0.46	[0.06, 0.85]	2.30	107	0.023*

IV. DISCUSSION

Adult listeners are known to vary substantially in their patterns of phonetic perception, with variability in the degree of sensitivity to distinctions across acoustic-phonetic continua, as well as differences in the sharpness of the boundary between categories. Using three tasks and five measures of phonetic perception, we found that all extracted measures (with the exception of sign-shine VAS slope) were at least weakly correlated with other phonetic measures, suggesting that at least some underlying aspects of phonetic decisions rely on shared mechanisms. Coherence between tasks performed on the same stimulus set was stronger than relationships across continua, supporting the assertion that, rather than fully gradient or fully categorical, participants may have idiosyncratic patterns of perception that are fairly specific to certain continua. Counter to predictions, in this study, performance on phonetic tasks did not differ substantially as a function of age. Perhaps most importantly, individual differences in phonetic perception individually and collectively predicted performance on a speech-in-noise task, even after accounting for age and noise exposure history, lending support to the hypothesis that differences in sensitivity to phonetic detail aids in comprehending speech in challenging listening conditions. Below we discuss the interpretation and implications of these findings.

Although perception of phonetic categories has often been described as "categorical," implying an all-or-none access to the phonetic category, recent attention to this issue suggests that listeners show substantial sensitivity to acoustic variability within the category (Fuhrmeister et al., 2023; Kapnoula & McMurray, 2021), challenging the entire notion of "categorical perception" as a phenomenon (McMurray, 2022). Our data corroborate that listeners show substantial sensitivity to within-category variation. Note, for instance, the patterns of discrimination (Fig. 2), with most tokens showing above-chance discrimination, whether the pair straddles the category boundary or not (a pattern that is particularly evident for the fricative continuum). Using the VAS task, a task argued to afford listeners the opportunity to demonstrate within-category sensitivity, a wide range of psychometric functions was observed, with some listeners responding more or less categorically. Yet others rated tokens gradiently, showing remarkable correspondence to their actual position on the acoustic-phonetic continuum.

These tasks have been well-described elsewhere, but our dataset contributes to two outstanding questions regarding the underlying skills that are tapped by these tasks. First, we tested the hypothesis that gradient responses in the VAS task reflect an ability to discriminate between items along the phonetic continuum, and thus should converge with AX discrimination tasks. Here, results diverged between the two continua. We found that mean sensitivity to discrimination across the continuum ('AX sensitivity') did not relate to the steepness of the psychometric function in the VAS task ('VAS slope") for the fricative, sign-shine continuum, but did correlate with VAS slope–and indeed with all measures–within the ba-da continuum. Correlations here were relatively weak and diffuse, making it difficult to firmly argue that these tasks tap distinct aspects of phonetic perception.

Second, we asked whether individual profiles of phonetic perception are best thought of as a general trait, or whether these profiles more closely reflect an individual's response to a specific acoustic-phonetic continuum. Here, evidence was also somewhat intermediate between these two options. While the strongest correlations were between measures tested on the same acoustic-phonetic contrast (especially within the ba-da continuum), between-continuum correlations were weaker (Fig 3). In explicit comparisons of these two models using confirmatory factor analysis, a "contrast-specific" model where the tasks loaded on phonetic contrast-specific latent factors was a better fit to the data than a model where all factors loaded on one latent factor.

The direction of the relationships between phonetic decision measures are quite consistent across comparisons. Listeners who show greater sensitivity in the discrimination task are also more likely to show a strong discrimination peak at the category boundary, more likely to show steeper VAS and 2AFC response functions, and are also more likely to be consistent responders in the VAS task, particularly within-contrast. These patterns might reflect subtle differences in peripheral or central aspects of the auditory system, differences in how sound is mapped to phonetic category representations, or (less compellingly) differences in task strategy that happen to affect multiple tasks.

During development, children show increasingly gradient patterns of perception as they transition into adolescence (McMurray et al., 2018). It is unclear whether or how this trajectory evolves in the adult lifespan. We hypothesized that well-documented age-related declines in the peripheral and central auditory system would result in changes in performance on phonetic decision tasks (Slade et al., 2020). Unexpectedly, age was related to only three phonetic measures, and only one of these, response consistency on the sign-shine continuum, was reliably related to age on its own, with greater age being associated with lower consistency. Of all the measures related to age, this one perhaps makes the most sense. First, fricative continua rely more heavily on high-frequency spectral information, and accurate perception of high-frequency information tends to decline with aging (Slade et al., 2020). Second, neural consistency (i.e., the stability of the response upon repeated measurement) declines with age (Skoe et al., 2015).

Nonetheless, age-related changes in performance on phonetic tasks were not striking in our sample. This might be due to a protective effect of language experience, or to the fact that our sample extends to age 67, but does not encompass older ages where sensorineural declines are more pronounced. Despite there being no striking relationships between age and our phonetic decision measures, age was nonetheless strongly related to speech-in-noise performance This replicates a well-established pattern of decreased perceptual acuity in noise with age (Slade et al., 2020; Holder et al., 2018), suggesting that our older adult sample was not entirely atypical in their perception of speech-in-noise[4]. In further support of the typicality of our dataset, for speech-perception-in-noise performance, we found expected relationship with noise exposure, with more noise exposure relating to worse performance (Casey et al., 2017; M. C. Liberman, 2017).

If individual differences in performance on phonetic tasks had no consequences for functional outcomes for comprehension, these differences would be interesting, but entirely academic. Instead, as reviewed in the introduction, VAS measures have been linked to aspects of native and non-native processing. However, prior attempts to link the slope of the VAS function to speech-in-noise perception accuracy showed weak or absent relationships (Kapnoula et al., 2017, 2021). Here we used two approaches to investigate the relationship between phonetic decisions and speech-in-noise performance. Using a backwards-stepping linear model selection approach, we showed that five phonetic decision measures predicted speech-in-noise performance, even after accounting for age and noise exposure dose. The steepness of the psychometric phonetic decision functions (VAS and 2AFC slopes) were not strong predictors of speech-in-noise (although 2AFC ba-da slope did survive model selection)–instead, the "categoricity" measure in the AX task (both continua) and response consistency in the VAS task (ba-da) were stronger predictors of speech-in-noise performance. In prior work (Apfelbaum, 2022), 2AFC slope was argued to be more closely related to response consistency than to gradience as measured by VAS slope—in our data, 2AFC slope was weakly related to both VAS slope and VAS consistency, suggesting that these measures do not cleanly dissociate. As in prior work, we find that response consistency is a useful predictor of language tasks (c.f. Fuhrmeister, et al., 2023), lending support to the notion that stability in the perceptual response or acuity in detecting acoustic-phonetic detail may be crucial for efficient mapping of auditory input onto meaning. A new contribution was the predictive power of the AX "categoricity" measure. This measure, which assesses the advantage conferred in discrimination when tokens cross the category boundary, may reflect an exaggeration of perceptual distances near the category boundary, which may help listeners to tune to critical acoustic-phonetic details in the input. An alternative interpretation (and perhaps more likely given the one-second ISI in our design) is that this task taps a listener's ability to hold auditory detail in memory, a task that will be easier when the tokens map to distinct phonetic categories. Future research, including investigating relationships between this task and other measures of auditory memory, will be needed to disambiguate these options.

Since there was mild collinearity among our set of phonetic decision measures, we also employed a PCA approach to identify common sources of variance within phonetic measures, essentially creating several phonetic decision "summary scores" for each participant (see Supplementary Materials). Here, too, addition of these summary dimensions explained speech-in-noise perception better than a model including only age, noise exposure dose, and childhood caregiver education, with four dimensions (1, 3, 4, and 5) showing significant contributions. Dimension 1, in particular, has loadings that are fairly evenly distributed across all measures except VAS slope for sign-shine (whereas the equally well-performing, Dimension 3 primarily loads on

VAS slope for sign-shine), leading to the conclusion that the cluster of performance identified above may constitute a general profile of phonetic skill.

Indeed, we cannot rule out the possibility that performance on speech-in-noise and phonetic tasks emerge from a common underlying trait, perhaps related to differences in auditory acuity, or more elaborated/stable language ability or working memory–this question awaits further study. Other limitations of the current dataset include the lack of hearing screening and a lack of precise control of the auditory testing environment. Although we are confident that our sample does not include participants with known hearing deficits, age-related hearing deficits often go undiagnosed. However, we are dubious that hearing declines, writ large, account for our results–notably, widespread age-related declines in phonetic performance were not obvious. A lack of control of the listening environment is inevitable for online studies. Results from our labs replicating well-known phenomena in speech perception (Fuhrmeister et al., 2023; Luthra et al., 2021) using online testing give us confidence in the quality of online data for speech perception research. We note that the participants in the current study are primarily Prolific "super-users" who have participated in hundreds of online studies, and tend to be very technically adept. We also required listeners to wear headphones, instituted a strict check for the presence of headphones, and allowed listeners to adjust the volume to a comfortable listening level. Nonetheless, we cannot rule out the possibility that individual differences in access to the auditory signal (whether because of hearing status, technological limitations of headphones, or ambient noise in the test environment) might explain our results. Indeed, allowing the listener to set their listening level, might, if anything, decrease the effect of aging. Another study limitation is that the noise exposure measurement was based on the previous 12 months and may for a variety of reasons (including the pandemic conditions under which the data were collected) not be representative of lifetime noise exposure. Ongoing efforts in our labs are aimed at these questions.

V. CONCLUSION

In summary, for both a stop continuum and fricative continuum, we found individual differences across a range of measurements of phonetic perception. Interestingly, individual differences in phonetic perception, specifically measures of consistency in the VAS task and a measure of near-boundary sensitivity in AX discrimination were found to predict speech-in-noise performance, suggesting that speech communication in noise is mediated by the structure of listeners' phonetic category representations. The constellation of findings suggests, however, that individual differences in phonetic measures are not listener-level traits (that are fixed across stimuli); instead, for a given listener, perceptual patterns/strategies appear to be specific to the particular speech continuum. These continuum-specific listener strategies may then aggregate with demographic factors (age, noise exposure) to influence the perception of naturalistic speech composed of multiple speech categories (i.e., sentences in background noise).

SUPPLEMENTARY MATERIAL

See supplementary material at [URL will be inserted by AIP] for acoustic details of the stimuli and tables reporting the full regression models before model selection procedures.

ACKNOWLEDGMENTS

This work was supported by NSF grant 1947883 to (Skoe & Myers, PIs). Myers and Phillips were additionally supported by NIH 2R01 DC013064 (Myers, PI) and Skoe was additionally supported by NIH 1R01AG075271. The authors would like to thank Hannah Mechtenberg for help with figure design, Hannah Mechtenberg, Anne Marie Crinnion, Naomi Sellers, and Portia Washington for helpful comments on an earlier version of this manuscript, and Travis White-Schwoch for help with NEQ data analysis, and Bhamini Sharma for help with the acoustic analysis.

AUTHOR DECLARATIONS

Conflict of Interest

The authors declare that they have no conflicts of interest.

Ethics Approval

Informed consent was obtained from all participants according to the regulations of the Institutional Review Board of the University of Connecticut

DATA AVAILABILITY

Analysis code and de-identified data are available at: https://osf.io/j5gpb/

ENDNOTES

[1]Prolific tracks participant-level data surrounding study approval (i.e., how many studies participants have completed providing high quality data) and study rejection (i.e., participants who did not complete a study in good faith by providing nonsense responses, completing study tasks in such a short amount of time they would be considered a statistical outlier [e.g., 3 standard deviations below the mean], etc.). The average number of study rejections per participant was 1.17 Prolific studies. Overall, our sample had a high study approval rating: of the 53,158 total Prolific studies completed by our sample, 53,022 were approved by study organizers (99.7%), indicating a high degree of data quality. See Figure 1 for an overview of the tasks and procedures.

[2] Despite the fact that participants reported that they "only know English," there were some contradictory responses to other questions. Namely, a fair number of participants reported early exposure to languages other than English, and some reported high proficiency in non-English languages. To explore whether these factors affected our analysis, we chose to categorize participants who reported exposure to a language other than English before the age of 10 and also reported high proficiency in a non-English language as "bilingual." 42 participants met this criterion. For each phonetic measure (described in detail under "Phonetic Decision Tasks"), we performed a two-tailed t-test comparing "bilingual" to "monolingual" groups. Of the ten phonetic measures, two showed significant differences between groups. Participants who reported "bilingual" language experience showed a shallower slope for the VAS task in the ba-da continuum ($t(93.54) = -3.04$, $p=0.003$) and showed a smaller "categoricity" measure for the ba-da continuum ($t(82.08)=-2.53$, $p=0.013$). Further, we added "bilingualism" as a factor to the best-fit model predicting speech-in-noise performance (see Section C. below). This factor did not improve model fit.

[3]The choice of ISI in discrimination tasks is not a neutral one. Although classic studies establishing categorical perception (e.g. Liberman, et al., 1957) used a 1000 msec ISI in discrimination tasks, it has been argued that longer ISIs encourage access to phonetic category labels, whereas shorter ISIs may come closer to tapping low-level, acoustic processing of stimuli (e.g.

Van Hessen & Schouten, 1992). Since in the VAS and 2AFC tasks, participants were asked to explicitly map acoustic information to phonetic category concepts, we chose a longer ISI in order to encourage similar access to category labels. We acknowledge that this may have the effect of placing a heavier burden on working memory than had we used a shorter ISI.

[4]To ensure the comparability of NEQ data collected online to data collected in person, the NEQ scores from online participants were compared to 312 NEQ collected in person, representing a similar age range and gender distribution. NEQ dB LAeq8750h did not differ as a function of study administration medium (Online mean = 71.8, SD = 6.5; Offline mean = 72.9, SD = 5.2; t(116.5) = 1.13, p = .26).

REFERENCES

Anderson, S., Parbery-Clark, A., Yi, H.-G., & Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and Hearing*, *32*(6), 750–757. https://doi.org/10.1097/AUD.0b013e31822229d3

Andruski, J. E., Blumstein, S. E., & Burton, M. W. (1994). The effects of subphonetic differences on lexical access. *Cognition*, *52*(3), 163–187.

Ansado, J., Marsolais, Y., Methqal, I., Alary, F., & Joanette, Y. (2013). The adaptive aging brain: Evidence from the preservation of communication abilities with age. *European Journal of Neuroscience*, *37*(12), 1887–1895. https://doi.org/10.1111/ejn.12252

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. C. (2022). Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America*, *152*(6), 3728. https://doi.org/10.1121/10.0015201

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *arXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Baum, S. R. (2003). Age differences in the influence of metrical structure on phonetic identification. *Speech Communication*, *39*(3–4), 231.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* (5.3.57) [Computer software]. http://www.praat.org/

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278–288. https://doi.org/10.1016/j.cognition.2013.11.015

Casey, J. A., Morello-Frosch, R., Mennitt, D. J., Fristrup, K., Ogburn, E. L., & James, P. (2017). Race/Ethnicity, Socioeconomic Status, Residential Segregation, and Spatial Variation in Noise Exposure in the Contiguous United States. *Environmental Health Perspectives*, *125*(7), 077017. https://doi.org/10.1289/EHP898

Centanni, T. M., Pantazis, D., Truong, D. T., Gruen, J. R., Gabrieli, J. D. E., & Hogan, T. P. (2018). Increased variability of stimulus-driven cortical responses is associated with genetic variability in children with and without dyslexia. *Developmental Cognitive Neuroscience*, *34*, 7–17. https://doi.org/10.1016/j.dcn.2018.05.008

Diaz, M. T., Karimi, H., Troutman, S. B. W., Gertel, V. H., Cosgrove, A. L., & Zhang, H. (2021). Neural sensitivity to phonological characteristics is stable across the lifespan. *NeuroImage*, *225*, 117511. https://doi.org/10.1016/j.neuroimage.2020.117511

Dorman, M. F., Marton, K., Hannley, M. T., & Lindholm, J. M. (1985). Phonetic identification by elderly normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *77*(2), 664–670. https://doi.org/10.1121/1.391885

Etymotic Research | QuickSIN™ Speech-in-Noise Test—Speech-In-Noise Tests—Auditory / Research. (n.d.). Retrieved November 22, 2019, from https://www.etymotic.com/auditory-research/speech-in-noise-tests/quicksin.html

Fuhrmeister, P., & Myers, E. B. (2021). Structural neural correlates of individual differences in categorical perception. *Brain and Language*, *215*, 104919. https://doi.org/10.1016/j.bandl.2021.104919

Fuhrmeister, P., Phillips, M. C., McCoach, D. B., & Myers, E. B. (2023). Relationships between native and non-native speech perception. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *49*(7), 1161–1175. https://doi.org/10.1037/xlm0001213

Goman, A. M., & Lin, F. R. (2016). Prevalence of Hearing Loss by Severity in the United States. *American Journal of Public Health*, *106*(10), 1820–1822. https://doi.org/10.2105/AJPH.2016.303299

Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2017). Masked speech perception across the adult lifespan: Impact of age and hearing impairment. *Hearing Research*, *344*, 109–124. https://doi.org/10.1016/j.heares.2016.11.004

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *The Journal of the Acoustical Society of America*, *119*(4), 2455–2466. https://doi.org/10.1121/1.2171527

Honda, C. T., Clayards, M., & Baum, S. R. (2024). Exploring individual differences in native phonetic perception and their link to nonnative phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/xhp0001191

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: A biological marker of dyslexia. *Journal of Neuroscience*, *33*(8), 3500–3504.

Johnson, T. A., Cooper, S., Stamper, G. C., & Chertoff, M. (2017). Noise Exposure Questionnaire:

A Tool for Quantifying Annual Noise Exposure. *Journal of the American Academy of Audiology*, *28*(1), 14–35. https://doi.org/10.3766/jaaa.15070

Kapnoula, E. C., Edwards, J., & McMurray, B. (2021). Gradient activation of speech categories facilitates listeners' recovery from lexical garden paths, but not perception of speech-in-noise. *Journal of Experimental Psychology. Human Perception and Performance*, *47*(4), 578–595. https://doi.org/10.1037/xhp0000900

Kapnoula, E. C., & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language*, *223*, 105031. https://doi.org/10.1016/j.bandl.2021.105031

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology. Human Perception and Performance*, *43*(9), 1594–1611. https://doi.org/10.1037/xhp0000410

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

Kong, E. J., & Kang, S. (2023). Individual Differences in Categorical Judgment of L2 Stops: A Link to Proficiency and Acoustic Cue-Weighting. *Language and Speech*, *66*(2), 354–380. https://doi.org/10.1177/00238309221108647

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: Evidence for sensory and cognitive coupling. *Brain and Language*, *128*(1), 34–40. https://doi.org/10.1016/j.bandl.2013.11.006

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13).

https://doi.org/10.18637/jss.v082.i13

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Liberman, M. C. (2017). Noise-induced and age-related hearing loss: New perspectives and potential therapies. *F1000Research*, *6*, 927. https://doi.org/10.12688/f1000research.11310.1

Luthra, S., Mechtenberg, H., & Myers, E. B. (2021). Perceptual learning of multiple talkers requires additional exposure. *Attention, Perception & Psychophysics*, *83*(5), 2217–2228. https://doi.org/10.3758/s13414-021-02261-w

Mattys, S. L., & Scharenborg, O. (2014). Phoneme categorization and discrimination in younger and older adults: A comparative analysis of perceptual, lexical, and attentional factors. *Psychology and Aging*, *29*(1), 150–162. https://doi.org/10.1037/a0035387

McMurray, B. (2022). The myth of categorical perceptiona). *The Journal of the Acoustical Society of America*, *152*(6), 3819–3842. https://doi.org/10.1121/10.0016614

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology*, *54*(8), 1472–1491. https://doi.org/10.1037/dev0000542

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42. https://doi.org/10.1016/S0010-0277(02)00157-9

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from “lexical” garden paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, *60*(1), 65–91. https://doi.org/10.1016/j.jml.2008.07.002

Myers, E. B. (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia*, *45*(7), 1463–1473.

Neef, N. E., Müller, B., Liebig, J., Schaadt, G., Grigutsch, M., Gunter, T. C., Wilcke, A., Kirsten, H., Skeide, M. A., Kraft, I., Kraus, N., Emmrich, F., Brauer, J., Boltze, J., & Friederici, A. D. (2017). Dyslexia risk gene relates to representation of sound in the auditory brainstem. *Developmental Cognitive Neuroscience*, *24*, 63–71. https://doi.org/10.1016/j.dcn.2017.01.008

Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics*, *15*, 289–290.

Prendergast, G., Couth, S., Millman, R. E., Guest, H., Kluk, K., Munro, K. J., & Plack, C. J. (2019). Effects of Age and Noise Exposure on Proxy Measures of Cochlear Synaptopathy. *Trends in Hearing*, *23*, 2331216519877301. https://doi.org/10.1177/2331216519877301

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Sarrett, M. E., McMurray, B., & Kapnoula, E. C. (2020). Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and sentential expectations. *Brain and Language*, *211*, 104875. https://doi.org/10.1016/j.bandl.2020.104875

Shehabi, A. M., Prendergast, G., Guest, H., & Plack, C. J. (2022). The Effect of Lifetime Noise Exposure and Aging on Speech-Perception-in-Noise Ability and Self-Reported Hearing Symptoms: An Online Study. *Frontiers in Aging Neuroscience*, *14*, 890010. https://doi.org/10.3389/fnagi.2022.890010

Skoe, E., & Kraus, N. (2013). Musical training heightens auditory brainstem function during sensitive periods in development. *Frontiers in Psychology*, *4*, 622. https://doi.org/10.3389/fpsyg.2013.00622

Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex (New York, N.Y.: 1991)*, *25*(6), 1415–1426. https://doi.org/10.1093/cercor/bht311

Slade, K., Plack, C. J., & Nuttall, H. E. (2020). The Effects of Age-Related Hearing Loss on the Brain and Cognitive Function. *Trends in Neurosciences*, *43*(10), 810–821. https://doi.org/10.1016/j.tins.2020.07.005

Tecoulesco, L., Skoe, E., & Naigles, L. R. (2020). Phonetic discrimination mediates the relationship between auditory brainstem response stability and syntactic performance. *Brain and Language*, *208*, 104810. https://doi.org/10.1016/j.bandl.2020.104810

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous Perception and Graded Categorization: Electrophysiological Evidence for a Linear Relationship Between the Acoustic Signal and Perceptual Encoding of Speech. *Psychological Science*, *21*(10), 1532–1540. https://doi.org/10.1177/0956797610384142

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception & Psychophysics*, *79*(7), 2064–

2072. https://doi.org/10.3758/s13414-017-1361-2