

MARCH 16 2026

Measuring age-related differences in phonetic gradiency with the visual analogue scale and eye tracking

Holly A. Zaharchuk ; Hannah E. Olson; Hannah Mechtenberg ; Matthew C. Phillips ; Emily B. Myers 



J. Acoust. Soc. Am. 159, 2460–2476 (2026)

<https://doi.org/10.1121/10.0042998>



Articles You May Be Interested In

Consistency in phonetic categorization predicts successful speech-in-noise perception

JASA Express Lett. (December 2025)

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

J. Acoust. Soc. Am. (September 2024)

A data-driven assessment of harmony in Quebec French [e] and [ɛ]

JASA Express Lett. (April 2024)



[LEARN MORE](#)

Advance your science and career as a member of the
Acoustical Society of America

Measuring age-related differences in phonetic gradiency with the visual analogue scale and eye tracking

Holly A. Zaharchuk,^{1,a)}  Hannah E. Olson,¹ Hannah Mechtenberg,²  Matthew C. Phillips,¹ 
and Emily B. Myers^{1,2} 

¹Department of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs, Connecticut 06269-1085, USA

²Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut 06269-1020, USA

ABSTRACT:

Listeners vary in their sensitivity to acoustic-phonetic variation. In this study, we explored different ways of measuring such sensitivity. Study 1 used a visual analogue scale (VAS) task that elicited ratings for a set of four phonetic continua. Study 2 used a two-alternative forced choice task with eye tracking, which provided both binary response data and fine-grained fixation data for the same phonetic continua as study 1. In both studies, we asked how patterns of phonetic perception changed over the course of the adult lifespan by comparing a younger adult sample (ages 18–25) to an older adult sample (age 50+). In addition, we conducted several assessments to examine how language skill, hearing acuity, and speech-in-noise perception were associated with individual differences in phonetic categorization. We found that older adults displayed more categorical (as opposed to more gradient) patterns of perception than younger adults, particularly in the VAS task, where age-related differences persisted even after controlling for hearing loss and speech-in-noise perception. There was also evidence of more categorical patterns of perception among older adults in the eye tracking measures. Overall, these data suggest that VAS tasks may be particularly sensitive to both age-related factors and differences in language skill. © 2026 Acoustical Society of America. <https://doi.org/10.1121/10.0042998>

(Received 17 June 2025; revised 17 February 2026; accepted 19 February 2026; published online 16 March 2026)

[Editor: Benjamin V Tucker]

Pages: 2460–2476

I. INTRODUCTION

Speech production is inherently variable. Socio-indexical and language-specific factors interact to produce a range of acoustic cues for any given phonetic category. Despite the continuous nature of variation in acoustic cue values, the process of mapping between these lower-level cues and higher-level phonetic categories tends to be discrete. Indeed, long-standing evidence suggests that listeners are worse at distinguishing two tokens that fall within a category (e.g., two examples of the “sh” in “shack”) compared to two tokens that fall between categories (e.g., the “s” in “sack” vs the “sh” in “shack”), even when the linear distance between the tokens is the same. This phenomenon, first described by Liberman *et al.* (1957), is known as “categorical perception” and taken as a hallmark of efficient speech processing. However, since this seminal finding, there is ample evidence supporting the existence of *gradient* rather than purely *categorical* speech perception, conditioned on the phonetic contrast and on the individual listener. In the present study, we sought to distinguish the conditions under which more gradient or more categorical patterns of speech perception emerge.

A. Task and stimulus factors in categorical perception

Previous research has shown that task design plays a large role in determining listeners’ response patterns (see Apfelbaum *et al.*, 2022; McMurray, 2022). On the one hand, binary response tasks (e.g., two-alternative forced choice [2AFC]) tend to yield categorical response patterns for tokens that vary along a phonetic continuum. A categorical response pattern is evidenced by a sudden shift in behavior at the perceptual boundary between two speech sound categories, where phonetic decisions *qualitatively* change from one category to the other despite a *quantitative* change in acoustics. On the other hand, tasks designed to elicit reaction time measures or other less discrete responses tend to yield graded response patterns. For example, listeners exhibit a steady increase in reaction times for categorizing phonetic tokens as the tokens approach a category boundary (Pisoni and Tash, 1974). Listeners are also able to make fine-grained “goodness of fit” judgments for how well a token corresponds to a phonetic category (Miller, 1994). In addition, the visual analogue scale (VAS) task, which asks participants to rate tokens along a continuous scale, was developed as a more gradient corollary to the typical 2AFC task (e.g., Kapnola *et al.*, 2017; Kong and Edwards, 2016; Massaro and Cohen, 1983). When given such a range of response options, listeners tend to show graded response patterns, where responses shift from one category to the other at a relatively slow rate compared to a 2AFC task. The

^{a)}Email: holly.zaharchuk@uconn.edu

VAS joins more implicit measures such as eye tracking (e.g., [McMurray et al., 2008](#)) and neuroimaging ([Myers, 2007](#); [Toscano et al., 2010](#)) in a suite of methods that allow researchers to sample fine-grained sensitivity to the phonetic details of speech. Indeed, these more precise methods provide listeners a better opportunity to demonstrate sensitivity to within-category contrasts than binary tasks provide ([Apfelbaum et al., 2022](#)).

Even within a given task, not all phonetic contrasts are perceived equally categorically. Listeners are better able to discriminate within-category tokens for vowel and fricative contrasts compared to stop contrasts and tend to show shallower categorization functions for vowels and fricatives compared to stops ([Fry et al., 1962](#); [Healy and Repp, 1982](#); [Massaro and Cohen, 1983](#)). Differences in the perception of stops compared to vowels and fricatives may arise from several sources. One proposal is that longer-duration sounds, like vowels and fricatives, allow listeners more time to encode acoustic differences between tokens, leading to greater within-category sensitivity ([Crowder, 1973](#); [Pisoni, 1972](#)). Another proposal is that differences in the perceptibility of particular contrasts stem from variability in production ([Newman et al., 2001](#)). If a particular contrast exhibits substantial between-talker variability (e.g., fricatives), listeners may have more gradient sensitivity to this contrast, given that they must attend to somewhat different features of the acoustic space depending on the talker ([McMurray and Jongman, 2011](#)). Overall, this work suggests that categorical perception is not so categorical; different tasks and stimuli can elicit more subtle patterns. In Sec. [1B](#), we explore the individual differences that underlie these group-level responses.

B. Individual differences in phonetic sensitivity

Listeners differ in their perceptual acuity, with some demonstrating little sensitivity to distinctions within a phonetic category and others demonstrating quite fine-grained sensitivity. This has led to two related questions in the field. First, does a particular response pattern reflect a stable trait of the listener, or a transient state? Second, do individual differences in phonetic perception—in particular gradiency (the steepness of the slope of the response function) and consistency (the degree to which listeners consistently provide the same response to a given token)—reflect phonetic processing abilities that characterize a broader language phenotype ([Fuhrmeister et al., 2023](#); [Fuhrmeister and Myers, 2021](#); [Honda et al., 2024](#); [Kapnoula et al., 2017](#); [Kapnoula et al., 2021](#); [Kim et al., 2025](#); [Kutlu et al., 2023](#); [Myers et al., 2024](#))?

Gradiency reflects the precision with which a listener provides responses for tokens ranging along a continuum. It is typically quantified as the slope of the logistic function that describes the VAS rating curve, with steeper slopes reflecting less gradient patterns of perception (e.g., [Fuhrmeister and Myers, 2021](#)). More gradient VAS rating curves have been associated with increased recovery from lexical garden paths ([Kapnoula et al., 2021](#)) and to increased sensitivity to secondary phonetic cues ([Kong and Edwards, 2016](#)). More gradient response patterns have also been

associated with increased exposure to linguistic diversity, which is thought to expand a listener's phonetic inventory ([Kutlu et al., 2023](#)).

The consistency with which listeners rate tokens in VAS tasks has been even more robustly associated with other measures. Consistent responders to native-language continua tend to be better perceivers of non-native contrasts ([Fuhrmeister et al., 2023](#); [Honda et al., 2024](#)) and better at speech-in-noise tasks ([Myers et al., 2024](#)). Grade school-age children with more consistent responses tend to have better language and reading abilities than those with inconsistent responses, a pattern not seen for gradiency ([Kim et al., 2025](#)). In addition, listeners who are more consistent in their responses also seem to have stronger weighting of acoustic-phonetic information in sentences with a word-final ambiguity ([Crinnion et al., 2026](#)). Overall, these results suggest that VAS consistency is a strong predictor of low-level acoustic-phonetic sensitivity.

It is not yet clear how patterns of gradiency and consistency evolve over the adult lifespan and whether individual differences in phonetic perception are useful predictors of other aspects of adult language. In a recent study, participants aged 18–68 years completed VAS tasks on separate stop and fricative continua ([Myers et al., 2024](#)). Consistency was found to be a robust predictor of speech-in-noise perception ability, even after factoring out age and noise exposure, suggesting that access to fine-grained phonetic detail is important for access to the speech signal under challenging listening conditions. According to episodic theories of speech perception like the ideal adapter framework ([Kleinschmidt and Jaeger, 2015](#)), an older adult listener would have accumulated more language experience over their lifetime than a younger adult listener, leading to more finely developed (and perhaps more tightly tuned) phonetic categories. Thus, aging might be associated with less gradient and more consistent responses. However, aging is unfortunately also strongly associated with hearing loss, including changes in hearing thresholds ([Goman and Lin, 2016](#)) and degraded access to other aspects of the speech signal ([Goossens et al., 2017](#)). If older adults experience less reliable and more degraded input, they would likely show less consistent responses. In turn, less consistent responses would add noise to the response function, rendering it more gradient in shape without reflecting true fine-grained sensitivity to gradience across the phonetic continuum. Although prior research has found that gradiency increases over the course of development ([McMurray et al., 2018](#)) and into middle adulthood ([Kapnoula and Samuel, 2024](#)), it remains unclear whether this trajectory continues in the aging older adult population. It is also unclear whether increased gradiency is attributable to increasingly refined phonetic categories or a potential decrease in auditory-perceptual acuity.

C. Present study

In this study we ask three primary questions. First, do long-attested differences in the perception of stops and fricatives also appear when using methods that allow for

more gradient responses (namely, VAS and eye tracking)? Second, following from evidence for developmental changes in the perception of phonetic category structure over the lifespan (Kutlu *et al.*, 2026; McMurray *et al.*, 2018), do younger and older adults differ in the gradiency and consistency of their responses? If so, do these differences emerge in both explicit (VAS) and implicit (eye tracking) tasks? Finally, we ask whether individual differences in VAS and eye tracking measures are associated with differences in language and hearing scores in a typical population.

II. STUDY 1: MEASURING PHONETIC SENSITIVITY WITH THE VISUAL ANALOGUE SCALE

A. Methods

1. Participants

Participants were recruited in two age groups: younger adults and older adults. We recruited younger adults aged 18–25 via the University of Connecticut Psychology Department participant pool and older adults aged 50 and over from the University of Connecticut and surrounding communities. All participants met the pure-tone average criteria (see Sec. II A 3 a). One participant was excluded because their language composite score did not meet the cutoff (see Sec. II A 3 b). Additionally, one participant was excluded because their CLQT+ composite score did not meet the cutoff (see Sec. II A 3 d). The 56 remaining participants included 27 younger adults [Age: 18–20, Mean (M) = 18.78, Standard deviation (SD) = 0.80; Sex: 17 female, 9 male, 1 unknown] and 29 older adults (Age: 52–80, M = 64.62, SD = 6.86; Sex: 23 female, 6 male). Three participants identified as Black, two as Asian, and 51 as White. Additionally, two participants identified as Hispanic or Latino, and the remaining 54 participants identified as not Hispanic or Latino. Participants reported that they were right-handed (handedness was confirmed with the Edinburgh Handedness Inventory, Oldfield, 1971) and that they spoke only North American English prior to the age of 13. All participants reported normal or correct-to-normal vision and either no hearing deficits or only mild hearing loss. All participants reported that they had no known history of neurological impairment or language disorders. Participants gave informed consent according to the guidelines of the University of Connecticut Institutional Review Board and were compensated with either course credit or cash payment.

2. Experimental materials

a. Stimuli. The experimental stimuli comprised four monosyllabic onset minimal pairs that crossed manner of articulation (stop, fricative) with another articulatory contrast (voicing, place of articulation): *bee-pea* (stops differing by voicing), *buck-duck* (stops differing by place of articulation), *face-vase* (fricatives differing by voicing), and *sack-shack* (fricatives differing by place of articulation). We initially piloted eight word pairs (two per type) that balanced

word frequency (to avoid rare words that may be unfamiliar to participants) with concreteness and imageability because these stimuli would later be used with eye tracking in study 2. The final set of word pairs were chosen because they elicited the most consistently accurate continuum end point ratings in a pilot study.

Stimuli were recorded in Audacity by a male native monolingual speaker of North American English with a RØDE microphone in a sound-attenuated booth (RØDE Microphones LLC (USA), Long Beach, CA). Stimuli were trimmed to the onset/offset of the word and scaled to 70 dB. The word pair recordings were manipulated to create 7-point continua from one item in the pair to the other (e.g., from *bee* to *pea*). The *bee* and *pea* recordings were manipulated in PRATT (Boersma and Van Heuven, 2001) using the “progressive cutback and replacement” technique that produced tokens with evenly spaced voice onset times (VOTs; Winn, 2020). This method produces natural-sounding tokens, enabling us to include the originally recorded end points in the final 7-point *bee-pea* continuum. The other three word pairs were morphed into a 9-point continuum using the STRAIGHT speech synthesis software (Kawahara, 2008). To blend two tokens, STRAIGHT takes measurements of multiple acoustic cues—including fundamental frequency (F0), spectral power, and aperiodicity—before iteratively blending these cues to create stepwise mixes of each end point (e.g., 20% of end point 1 and 80% of end point 2; other studies using this technique are Hubbard and Assmann, 2013; Luthra *et al.*, 2021; Rogers and Davis, 2017). The central seven tokens (i.e., steps 2–8) were used in the final continua to avoid a salient difference between the original recordings of the end points that were not manipulated and the synthesized tokens.

b. VAS task. Participants completed a VAS task, in which tokens from each 7-point continuum were presented and rated on a scale from 1 to 7 (Fig. 1). A rating of 1 always corresponded to the voiced or more frontal end point of the word pair (*bee*, *buck*, *face*, or *sack*) and a rating of 7 corresponded to the other end point (*pea*, *duck*, *vase*, or *shack*). Images depicting these anchoring end points (e.g., a “bee” or “pea”; see supplementary material, Appendix A) were presented with a visual scale with numbers 1 through 7 presented at even intervals [see Fig. 1(B)]. Participants were encouraged to use the full scale, including the intermediate points, both visually in the task instructions and verbally by the experimenter. Stimuli were presented via Gorilla and played over Sony MDR-7506 or HD 280 Pro headphones (Sony Corporation of America, New York, NY). Participants responded by pressing one of the number keys (1–7) on the keyboard or the corresponding button on the button box. The task consisted of four blocks, one per continuum. Block order was pseudo-randomized using a Latin square that ensured stops and fricatives always alternated [see Fig. 1(C)]. Within each block, there were six repeatable practice trials and 70 experimental trials that were presented

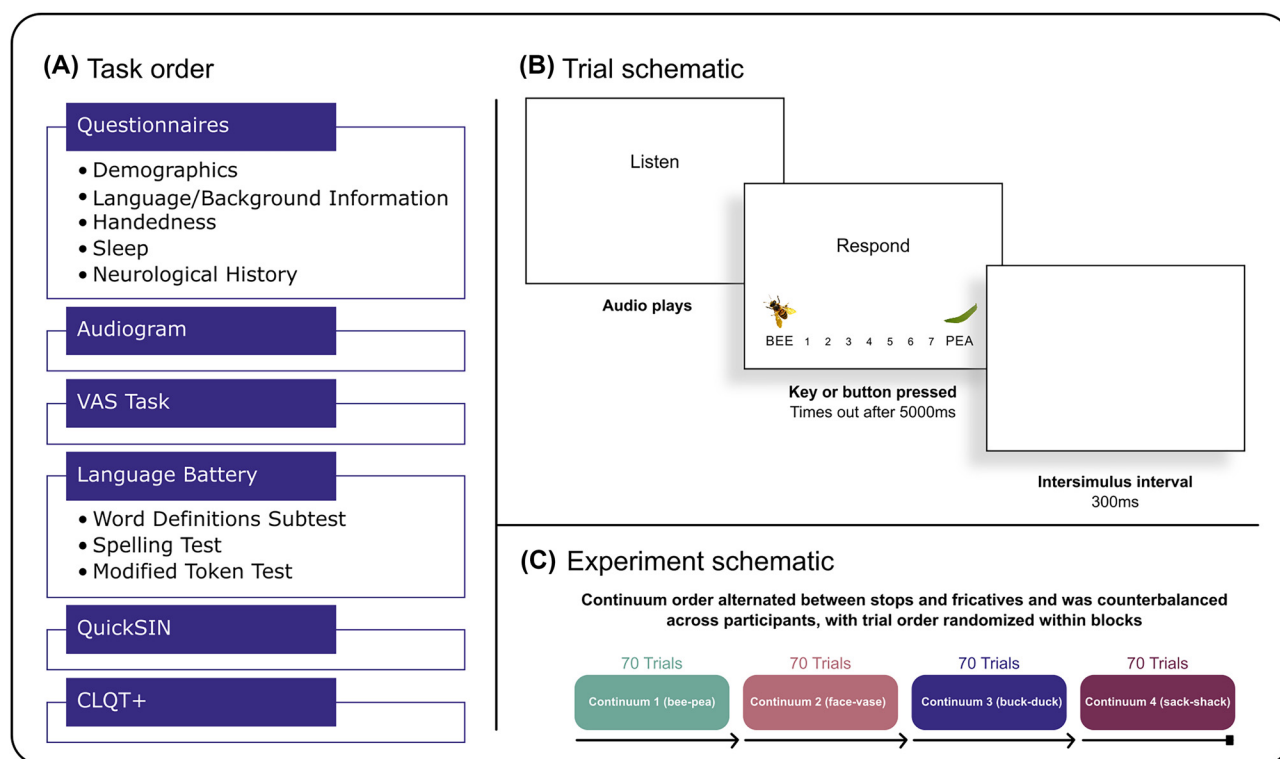


FIG. 1. Methods schematic for study 1. (A) List of tasks, in order, that participants completed as part of study 1; (B) trial schematic for the VAS task; (C) experiment schematic highlighting the blocking structure across continua for the VAS task.

in random order (10 instances of each of the 7 tokens). The task order for study 1 can be seen in Fig. 1(A).

c. Questionnaires. Participants reported their age, race, ethnicity, and sex. They completed four additional questionnaires regarding their language history and general background information, handedness, sleep habits, and neurological history. These questionnaires can be found in the Open Science Framework (OSF) repository associated with this project.

3. Assessments

Test score distributions for each age group can be found in the [supplementary material](#), Table VIII.

a. Hearing acuity. We conducted a pure-tone audiogram at 1, 2, and 4 kHz. Partway through data collection, we also added 0.5 kHz to the audiogram. Following World Health Organization guidelines for hearing screening ([Feltner et al., 2021](#)), we excluded participants with pure-tone average thresholds above 35 dB in both ears, a threshold that will include people with mild-to-moderate hearing loss. We used a fairly inclusive threshold to include a representative sample of older adults, who are likely to experience some hearing loss. Participants with a difference in pure-tone average thresholds greater than 15 dB between ears were also excluded. The better of each participant's two thresholds was used for analysis [pure-tone average (best ear)].

b. Language skill. Participants completed three language assessments designed to identify adults with developmental language impairment ([Fidler et al., 2011, 2013](#)). The Spelling Test was included as a measure of phonological skill and vocabulary knowledge, and the Modified Token Test was included as a measure of syntactic skill ([Morice and McNicol, 1985](#); see [Fidler et al., 2011](#), for task details). The Word Definitions subtest from the *Clinical Evaluation of Language Fundamentals – Fourth Edition* (CELF-4; [Semel et al., 2003](#)) was also included as a measure of semantic skill but will not be discussed further. The History-SLS formula from [Fidler et al. \(2011, 2013\)](#) was used to combine the Spelling Test and Modified Token Test scores into a single measure. This *language composite score* identifies adults who may have needed speech and language services, with a negative score indicating typical language abilities and a positive score indicating potential membership in the clinical group. Participants whose scores were greater than positive 1 were excluded.

c. Speech-in-noise perception. Participants also completed the Quick Speech in Noise (QuickSIN) task to assess their perception of speech in increasingly noisy listening conditions (see [Killion et al., 2004](#), for details). In this task, participants heard sentences embedded in four-talker babble at different signal-to-noise ratios and repeated as many words from each sentence as possible. The overall QuickSIN score was calculated by averaging the total number of correctly repeated keywords across four lists of sentences.

d. Cognitive skill. The Cognitive Linguistic Quick Test-Plus (CLQT+; Helm-Estabrooks, 2001) was administered to assess cognitive skills across five domains (attention, executive functioning, language, memory, and visuospatial skills) and screen for cognitive deficits. The first subtest, Personal Facts, was skipped for all participants to avoid collecting identifying information unnecessarily; all participants were given full points for this subtest to calculate the total scores. Any participant with a *CLQT+ composite score* below 3.5, suggesting cognitive impairment, was excluded.

B. Results

All pre-processed data (de-identified), analyses, and results can be found in the OSF repository associated with this project.

1. Data pre-processing and analysis approach

Data pre-processing and analysis were conducted with R version 4.5.0 (R Core Team, 2025). First, the VAS responses were checked to ensure that participants made full use of the rating scale for each continuum, defined as having at least two steps with mean ratings that differed by at least two points. All continua for all participants met this criterion. Next, we used the NLME package to construct a four-parameter logistic regression model for each continuum that modeled the boundary values, maximum and minimum asymptotes, and slope values for each participant (Pinheiro *et al.*, 2025). Slope estimates, reflecting the steepness of the excursion in the psychometric function, were extracted from each model. Consistency measures were estimated by computing the mean SD in VAS ratings across all continuum points by participant and continuum. This value was inverted by multiplying by -1 , such that higher values indicate more consistent (less variable) responses. Figure 2 displays the individual and group VAS rating data for all four continua, as well as summaries of the estimated slope and consistency values for each continuum. Pearson's correlation between response slope and consistency showed a modest relationship between slope and consistency values across the sample ($r(222) = 0.17$, $p = 0.012$), such that more consistent responders tended to show steeper slopes. Because a participant with a perfectly categorical function (steep slope) would also by definition have a maximum consistency value, subsequent analyses examining age and phonetic effects on slope and consistency were conducted using the opposite measure as a covariate. Subsequent analyses were conducted via linear mixed-effects models constructed with the LME4 package (Bates *et al.*, 2015). The EMMEANS package was used to obtain estimated marginal means from the models and compute contrasts (Lenth, 2025).

2. Slope and consistency measures by phonetic contrast and age

a. Slope measure. We constructed a linear mixed-effects model with individual slope estimates as the dependent variable. We included Manner (stop, fricative), Voice/

Place (voicing, place of articulation), Age Band (younger, older), and their interactions as fixed effects, and Participant as a random effect. The fixed effects were sum-contrast coded (Manner: stop = -0.5 , fricative = 0.5 ; Voice/Place: voicing = -0.5 , place = 0.5 ; Age Band: younger = -0.5 , older = 0.5), and consistency was included as a covariate. The results can be seen in Table I.

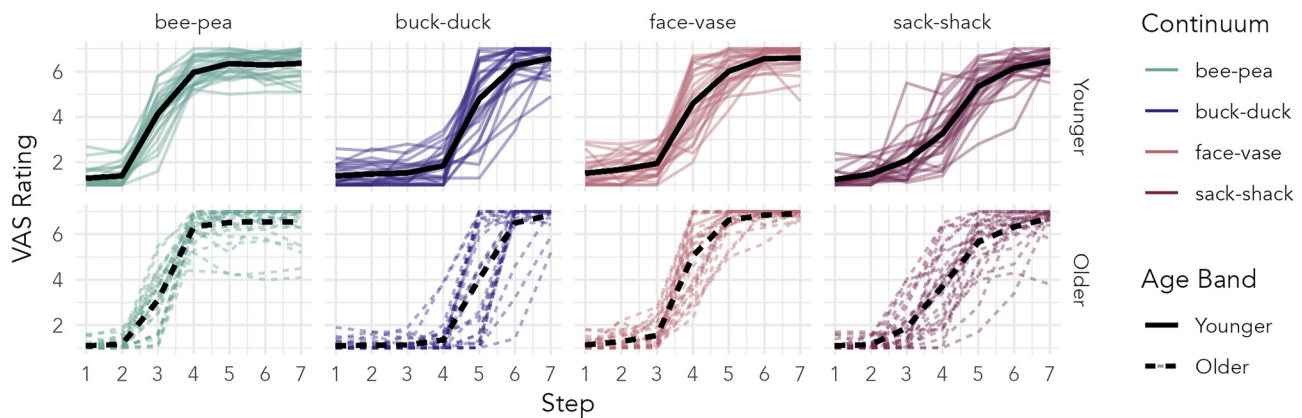
Beginning with the effects of phonetic contrast, we observed a significant main effect of Manner, such that stops had steeper slopes than fricatives. This effect was qualified by a significant Manner by Voice/Place interaction. To investigate this interaction, we compared the estimated marginal means averaged across Age Band. Overall, the steepest slopes were elicited by the stop-place continuum (*buck-duck*: $M = 0.44$, 95% CI [0.21 , 0.66]), followed by the stop-voicing continuum (*bee-pea*: $M = 0.11$, 95% CI [-0.12 , 0.34]), then the fricative-voicing continuum (*face-vase*: $M = -0.03$, 95% CI [-0.26 , 0.20]), and finally the fricative-place continuum (*sack-shack*: $M = -0.57$, 95% CI [-0.80 , -0.34]). Pairwise comparisons within Voice/Place revealed a significant difference between fricatives and stops for place of articulation (*sack-shack* vs *buck-duck*: $t(166) = -7.00$, $p < 0.001$) but not for voicing (*face-vase* vs *bee-pea*: $t(162) = -0.85$, $p = 0.399$). Within Manner, pairwise comparisons revealed a significant difference between place of articulation and voicing for both fricatives (*sack-shack* vs *face-vase*: $t(161) = -4.41$, $p < 0.001$) and stops (*buck-duck* vs *bee-pea*: $t(163) = 2.08$, $p = 0.039$).

Regarding the effects of age, there was a significant main effect of Age Band, such that older adults ($M = 0.36$, 95% CI [0.12 , 0.60]) had steeper slopes than younger adults ($M = -0.39$, 95% CI [-0.64 , -0.14]; $t(54) = -4.33$, $p < 0.001$). A three-way interaction also emerged among Manner, Voice/Place, and Age Band. Older adults had significantly steeper slopes than younger adults on all continua except *sack-shack*, where the two groups did not differ. Estimated marginal means contrasts for Age Band by Continuum are displayed in the [supplementary material](#), Table IX, for the reader's interest.

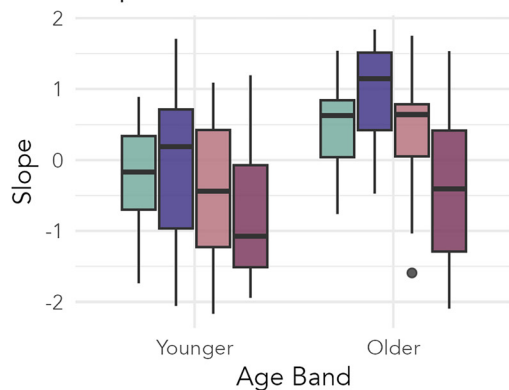
b. Consistency measure. We used the same model structure as we did for the slope measure to analyze response consistency. The results can be seen in Table II.

There was a significant main effect of Manner, with stops showing higher consistency values than fricatives. This effect was qualified by a significant Manner by Voice/Place interaction. To investigate this interaction, we compared the estimated marginal means averaged across Age Band. The most consistent continuum was the stop-place continuum (*buck-duck*: $M = 0.53$, 95% CI [0.30 , 0.76]), followed by the stop-voicing continuum (*bee-pea*: $M = -0.04$, 95% CI [-0.27 , 0.20]), then the fricative-voicing continuum (*face-vase*: $M = -0.13$, 95% CI [-0.36 , 0.10]), and finally the fricative-place continuum (*sack-shack*: $M = -0.41$, 95% CI [-0.64 , -0.18]). Pairwise comparisons within Voice/Place revealed a significant difference between fricatives and stops for place of articulation (*sack-shack* vs *buck-duck*: $t(185) = -3.89$, $p < 0.001$) but not

(A) Mean (black) and individual (color) VAS curves



(B) Slope



(C) Consistency

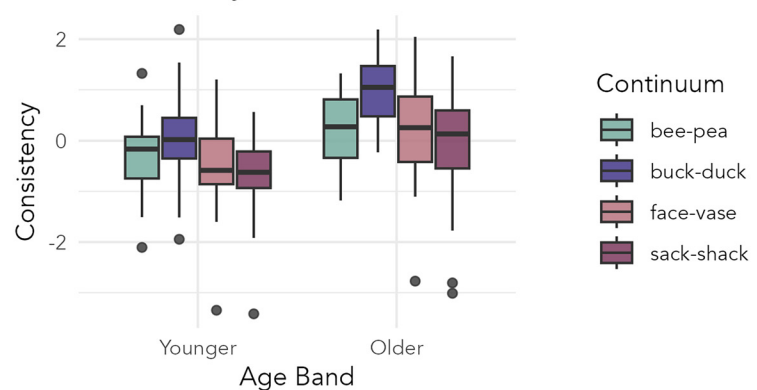


FIG. 2. Rating data from the VAS task. (A) Colored lines represent the mean ratings by participant at each continuum step. Black lines represent the overall mean across participants. Continuum step is displayed along the x axis, and the VAS rating is displayed along the y axis. (B) Box plots show the slope, (C) consistency estimates extracted from the individual response functions.

for voicing (*face-vase* vs *bee-pea*: $t(162) = -0.35$, $p = 0.724$). Within Manner, pairwise comparisons revealed a significant difference between place of articulation and voicing for stops (*buck-duck* vs *bee-pea*: $t(149) = 3.20$, $p = 0.002$) but not for fricatives (*sack-shack* vs *face-vase*: $t(169) = -0.74$, $p = 0.459$).

TABLE I. Linear mixed-effects model output predicting individual VAS slope values. Manner (stop, fricative), Voice/Place (voicing, place of articulation), Age Band (younger, older), and their interactions were included as fixed effects, and Participant was included as a random effect. df, degrees of freedom.

Term	$\hat{\beta}$	95% CI	t	df	p
Intercept	-0.01	[-0.19, 0.17]	-0.14	52	0.888
Manner	0.49	[0.31, 0.66]	5.56	164	<0.001
Voice/Place	0.14	[-0.03, 0.30]	1.61	159	0.108
Age Band	0.72	[0.36, 1.08]	3.96	52	<0.001
Consistency	0.19	[0.08, 0.30]	3.45	206	<0.001
Manner $\times$ Voice/Place	-0.77	[-1.10, -0.44]	-4.56	160	<0.001
Manner $\times$ Age Band	0.16	[-0.17, 0.48]	0.94	159	0.348
Voice/Place $\times$ Age Band	0.10	[-0.22, 0.43]	0.62	159	0.535
Manner $\times$ Voice/Place $\times$ Age Band	-0.72	[-1.37, -0.06]	-2.15	159	0.033

There was also a main effect of Age Band, with older adults ($M = -0.33$, 95% CI [0.13, 0.54]) showing more consistent responses than younger adults ($M = -0.36$, 95% CI [-0.57, -0.15]; $t(54) = 4.70$, $p < 0.001$).

TABLE II. Linear mixed-effects model output predicting individual VAS consistency values. Manner (stop, fricative), Voice/Place (voicing, place of articulation), Age Band (younger, older), and their interactions were included as fixed effects, slope was added as a covariate, and Participant was included as a random effect.

Term	$\hat{\beta}$	95% CI	t	df	p
Intercept	-0.01	[-0.15, 0.13]	-0.11	53	0.912
Manner	0.34	[0.12, 0.55]	3.09	178	0.002
Voice/Place	-0.18	[-0.38, 0.02]	-1.74	161	0.083
Age Band	0.45	[0.15, 0.76]	2.91	61	0.005
Slope	0.32	[0.18, 0.46]	4.43	196	<0.001
Manner $\times$ Voice/Place	-0.57	[-0.99, -0.16]	-2.71	171	0.007
Manner $\times$ Age Band	0.11	[-0.29, 0.51]	0.55	161	0.584
Voice/Place $\times$ Age Band	-0.18	[-0.58, 0.21]	-0.91	160	0.367
Manner $\times$ Voice/Place $\times$ Age Band	-0.41	[-1.21, 0.39]	-1.00	163	0.320

3. Language and hearing predictors of slope and consistency

a. Slope measure. Having analyzed the individual slope and consistency measures with respect to phonetic contrast and age, we next asked whether individual differences in language and hearing abilities might predict these VAS measures. We looked at three language and hearing measures (see Sec. II A 3 for details): language composite scores, where higher scores reflect poorer language outcomes; pure-tone average thresholds (best ear), where higher values reflect more impaired hearing; and QuickSIN scores, where higher values reflect better speech-in-noise perception. These values were mean-centered and scaled for analysis. We constructed a linear mixed-effects model with Age Band and the three language and hearing scores as fixed effects and Participant and Continuum (*bee-pea*, *buck-duck*, *face-vase*, *buck-duck*) as random effects. The results can be seen in Table III.

We observed a significant effect of Age Band, such that older adults had steeper slopes than younger adults. Notably, this relation between age and slope emerged even though we included other age-related factors (especially hearing) in the model. Language composite scores also significantly predicted slopes, with poorer language outcomes associated with steeper slopes.

b. Consistency measure. We used the same model structure as we did for the slope measure to analyze response consistency. The results can be seen in Table IV.

We observed a significant effect of Age Band, with older adults having higher response consistency than younger adults. As in the slope analysis, this relation between age and consistency remained after controlling for other individual differences that often decline with age.

C. Interim discussion

In this study, we replicated a long-standing finding that stops are perceived more categorically than fricatives. This was evidenced by steeper slopes in the logistic response functions estimated for each participant. We also found that stop continua exhibit greater consistency in responses than fricative continua. However, the relatively gradient and inconsistent responses to the fricative-place continuum

(*sack-shack*) largely drove these differences between stops and fricatives. Indeed, the fricative-voicing continuum (*face-vase*) did not differ strongly from the stop-voicing continuum (*bee-pea*). In general, the precise acoustics of the continuum are likely to matter as much as the phoneme class itself (a topic we return to in the general discussion).

We also observed strong age-related differences in the VAS results, with older adults responding more categorically and consistently than younger adults. These differences persisted even after controlling for hearing acuity and speech-in-noise perception, both of which typically decline with age. We will further discuss the differences between younger and older adults in the general discussion.

III. STUDY 2: MEASURING PHONETIC SENSITIVITY WITH EYE TRACKING

A. Methods

1. Participants

As in study 1, participants were recruited in two age groups: younger adults and older adults. We recruited younger adults aged 18–25 via the University of Connecticut Psychology Department participant pool and older adults aged 50 and over from the University of Connecticut and surrounding communities. One participant was excluded because they did not meet the pure-tone average criteria (see Sec. II A 3 a), six participants were excluded because their language composite scores did not meet the cutoff (see Sec. II A 3 b), and one participant was excluded because their CLQT+ composite score did not meet the cutoff (see Sec. II A 3 d). Test score distributions for each age group can be found in the [supplementary material](#), Table X. One participant was excluded because they reported receiving speech and language services, and another participant was excluded for being left-handed. Three additional participants were excluded because of technical difficulties with the eye tracking camera. The remaining 58 participants included 31 younger adults (Age: 18–20, $M=18.35$, $SD=0.61$; Sex: 25 female, 6 male) and 27 older adults (Age: 53–79, $M=66.41$, $SD=7.47$; Sex: 23 female, 4 male). Fifty-three participants identified as White, one as Asian, and four as more than one race. Additionally, two participants identified as Hispanic or Latino, one participant did not report their ethnicity, and the remaining 55

TABLE III. Linear mixed-effects model output predicting individual VAS slope values. Age Band (younger, older), language composite score, pure-tone average (best ear), and QuickSIN score were included as fixed effects and Participant and Continuum (*bee-pea*, *buck-duck*, *face-vase*, *buck-duck*) included as random effects.

Term	$\hat{\beta}$	95% CI	t	df	p
Intercept	−0.01	[−0.45, 0.42]	−0.06	3.61	0.952
Age Band	0.79	[0.39, 1.20]	3.85	51	<0.001
Language composite score	0.19	[0.01, 0.38]	2.02	51	0.049
Pure-tone average (best ear)	0.05	[−0.16, 0.26]	0.50	51	0.622
QuickSIN score	−0.15	[−0.34, 0.04]	−1.57	51	0.122

TABLE IV. Linear mixed-effects model output predicting individual VAS consistency values. Age Band (younger, older), language composite score, pure-tone average (best ear), and QuickSIN score were included as fixed effects and Participant and Continuum (*bee-pea*, *buck-duck*, *face-vase*, *buck-duck*) included as random effects.

Term	$\hat{\beta}$	95% CI	t	df	p
Intercept	−0.01	[−0.41, 0.39]	−0.05	3.42	0.962
Age Band	0.58	[0.20, 0.96]	3.02	51	0.004
Language composite score	−0.07	[−0.24, 0.11]	−0.75	51	0.456
Pure-tone average (best ear)	0.05	[−0.15, 0.24]	0.47	51	0.637
QuickSIN score	0.01	[−0.16, 0.19]	0.15	51	0.881

participants identified as not Hispanic or Latino. Participants reported that they were right-handed (handedness was confirmed with the Edinburgh Handedness Inventory, Oldfield, 1971) and that they spoke only North American English prior to the age of 13. All participants reported normal or corrected-to-normal vision and either no hearing deficits or only mild hearing loss. All participants reported that they had no known history of neurological impairment or language disorders. Participants gave informed consent according to the guidelines of the University of Connecticut Institutional Review Board and were compensated with either course credit or cash payment.

2. Experimental materials

The same auditory stimuli used in study 1 were used in study 2 (see Sec. II A 2 a). Individual auditory tokens were padded with 100 ms of silence prior to word onset, following procedures from similar eye tracking studies (McMurray *et al.*, 2002; McMurray *et al.*, 2008). End point images (see supplementary material, Appendix A for details) were centered on a white square to ensure that they were visually distinguishable from the gray background of the task (red-green-blue: 163, 191, 167). The image-to-background ratio was roughly equalized across all images such that the end point images all occupied approximately equal amounts of visual space relative to the area of the white square tile. The size of the images on the screen was 283×283 pixels. The size of each image interest area (i.e., upper and lower image)

for data analysis was set at a slightly expanded 288×288 pixels.

3. Eye tracking task and procedure

The full task order, including questionnaires and assessments, is illustrated in Fig. 3(A). For the eye tracking task, participants were seated in a soundproof booth in front of a BenQ XL2420-B 24-in monitor (BenQ America Corp., Costa Mesa, CA). A desk-mounted EyeLink 1000 Plus (SR Research, Ottawa, Ontario, Canada) collected gaze-position data from participants' right or left eye at a sampling rate of 500 Hz. The task was programmed and presented using SR Research Experiment Builder (Version 2.2.299; SR Research, Ottawa, Ontario, Canada). Auditory stimuli were delivered through Sony MDR-7506 headphones (Sony Corporation of America, New York, NY).

The eye tracking task used a 2AFC paradigm, in which participants clicked on the image that most closely matched the word they heard [see Fig. 3(B)]. At the beginning of the task, participants were familiarized with the end point words (e.g., bee and pea) and corresponding images for all four continua and walked through an example trial. Each trial began with a blank screen for 200 ms, followed by the presentation of the two end point images, aligned on the vertical axis, for 500 ms. The upper image always corresponded to the voiced or more frontal end point of the word pair (*bee*, *buck*, *face*, or *sack*) and the lower image always corresponded to the other end point (*pea*, *duck*, *vase*, or *shack*). A red circle then appeared in between the two images.

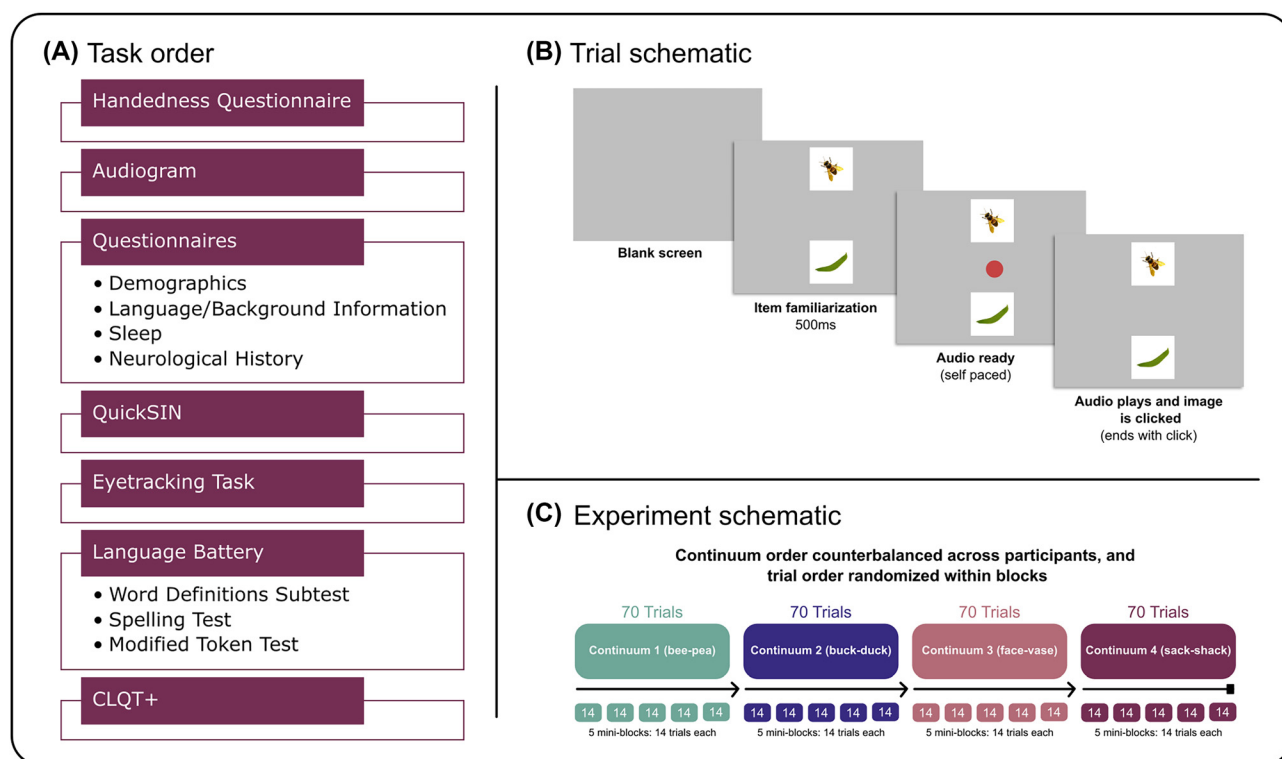


FIG. 3. Methods schematic for study 2. (A) List of tasks, in order, that participants completed as part of study 2; (B) trial schematic for the 2AFC eye tracking task; (C) experiment schematic highlighting the blocking structure within and across continua for the 2AFC eye tracking task.

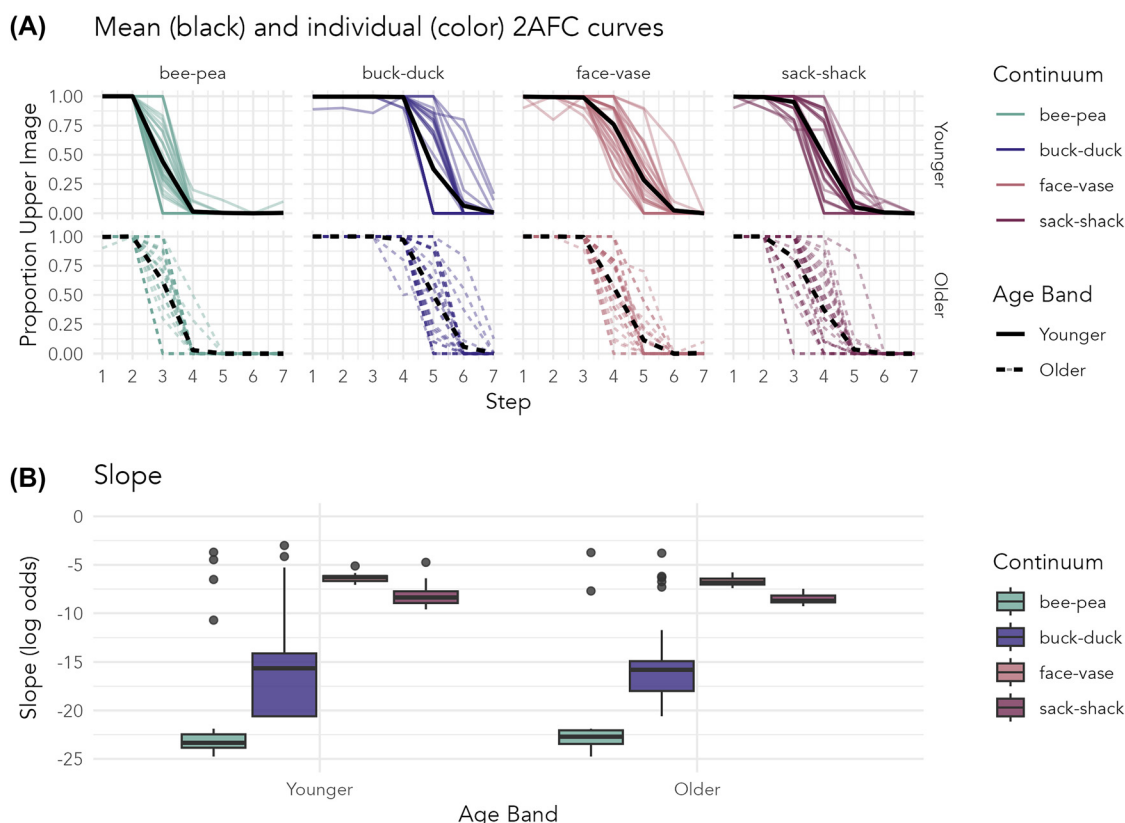


FIG. 4. Response data from the 2AFC eye tracking task. (A) Colored lines represent the mean proportion of upper image selections compared to all selections by participant at each continuum step. Black lines represent the overall mean across participants. (B) Box plots show the distributions of the individual coefficients for the effect of continuum step on upper image selections.

Participants clicked on the red circle to initiate the presentation of the auditory token. Importantly, the red circle forced participants' mouse position and directed their eye gaze to the center of the screen to ensure that all participants started in the same location. After clicking on the circle, participants heard the auditory token and were asked to click on the image that best matched what they heard. There were 280 total trials in the task (70 per continuum).

Participants completed the task in four blocks, one for each continuum [see Fig. 3(C)]. Four different continuum orders were counterbalanced across participants. The continua were grouped by their place of articulation for counterbalancing, such that the two stop continua (bee-pea, buck-duck) and the two fricative continua (face-vase, sack-shack) were always grouped, but the presentation order was switched [e.g., order 1: (buck-duck, bee-pea, sack-shack, face-vase) vs Order 2: (bee-pea, buck-duck, face-vase, sack-shack)]. Counterbalancing orders 3 and 4 began with the two fricative continua. At the beginning of each block, participants completed four practice trials (two of each end point token). Practice trials were repeated if participants scored less than three of four correct. Within each block, there were five 14-trial mini-blocks. Each mini-block began with a five-point calibration and validation procedure followed by drift correction. This mini-block procedure was adopted to reduce fatigue among our older adult sample (Dial *et al.*, 2019).

B. Results

All pre-processed data (de-identified), analyses, and results can be found in the OSF repository associated with this project.

1. Data pre-processing and analysis approach

Data pre-processing and analysis were conducted with R version 4.4.1 (R Core Team, 2025). Eye-gaze and response data were pre-processed using the VWPPE package (Porretta *et al.*, 2018). Trials in which 25% or more of the data from the onset of the visual display to the selection of the upper/lower image contained off-screen looks or blinks were removed (Younger: $N = 500$, 4.69%; Older: $N = 601$, 6.76%). Response times and fixations were time-locked to the onset of the auditory stimulus. Trials with response times under 200 ms or over 3000 ms were removed (Younger: $N = 66$, 0.80%; Older: $N = 56$, 0.82%). On average, 262 of 280 trials per participant remained [Younger: $M = 263$, $SD = 22$, Minimum (Min) = 191, Maximum (Max) = 280; Older: $M = 261$, $SD = 23$, Min = 189, Max = 280]. One older adult participant did not have data for the two fricative continua because of an equipment failure but was still included in the analyses.

For the response data, image selection was coded with respect to the upper image (upper image = 1, lower image = 0). The upper image was chosen as the reference rather than the lower image (which would have aligned the curves

Mean proportion upper image fixation

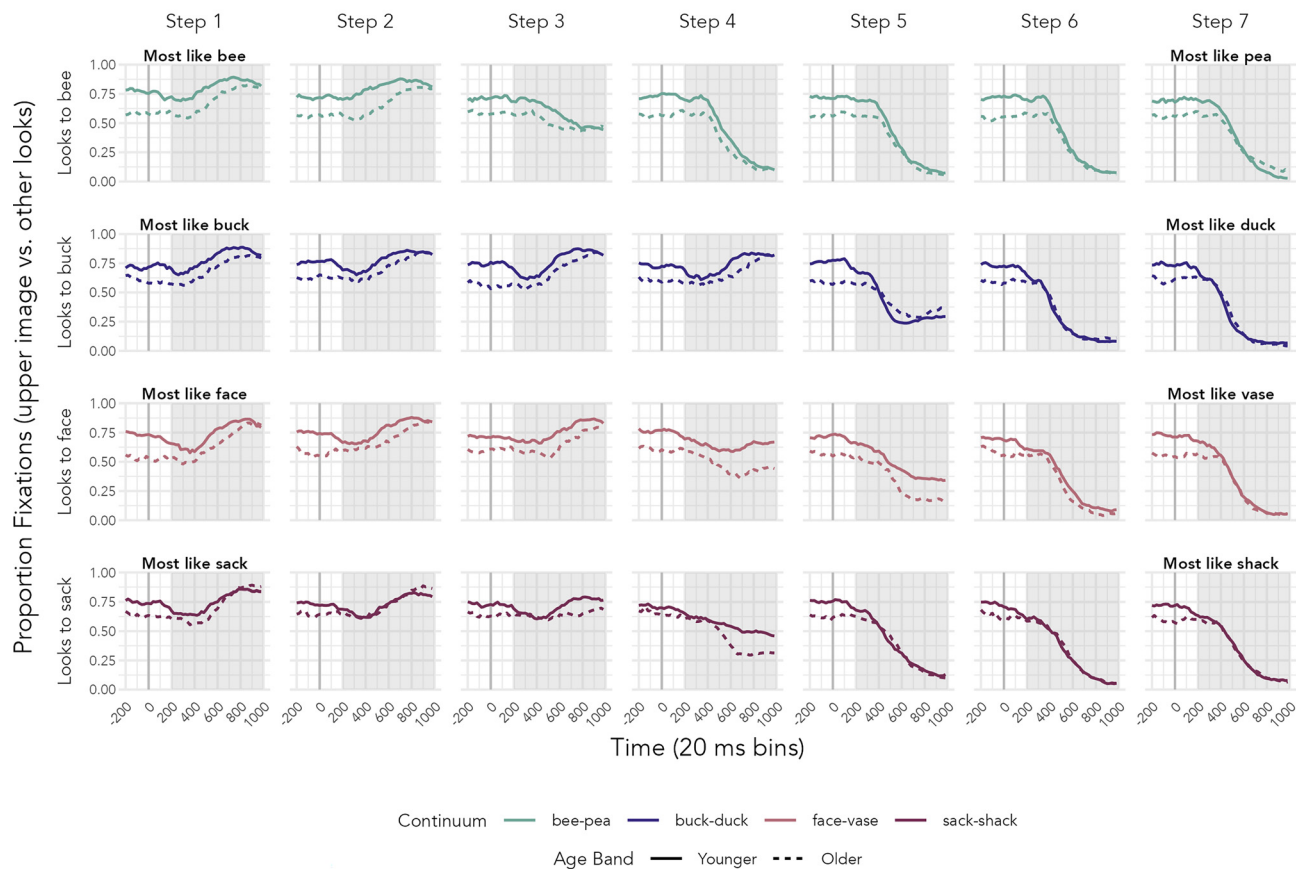


FIG. 5. Eye-gaze data while participants engaged in a 2AFC task. Colored lines represent the mean proportion of eye fixations on the upper image compared to all other fixations. The upper image always depicted the left side of the continuum (bee, buck, face, and sack). Solid gray lines indicate the onset of the auditory token. Light gray rectangles indicate the analysis window.

with the VAS rating data) because of an upper-image bias in the eye tracking data (see below). Mean individual- and group-level responses are displayed in Fig. 4(A) for each continuum and age.

For the eye-gaze data, fixations were considered in an 800 ms time window between 200 ms (accounting for the oculomotor delay) and 1000 ms post-stimulus onset. Following Kapnola and McMurray (2021), fixations were truncated or extended as appropriate to correspond to this exact time window. For trials that ended before 1000 ms, we extended the fixations using the value from the last recorded fixation location. Trials with no image fixations in the analysis window were removed (Younger: $N=67$; 0.82%; Older: $N=87$; 1.28%). Fixations were coded with respect to the upper image (upper image = 1, other = 0). The upper image was chosen as the reference after visual inspection of the data, which revealed a strong upper-image bias in the location of the eyes at the start of the analysis window (see Fig. 5). That is, for tokens that corresponded more closely to the upper image (e.g., more bee-like tokens for the *bee-pea* continuum), fixations tended to stay on the upper image over the course of the trial; however, for tokens that corresponded more closely to the lower image (e.g., more pea-like tokens for the *bee-pea*

continuum), fixations started on the upper image but decreased over the course of the trial. To analyze these data, the proportion of upper image fixations was calculated in 20 ms time bins (40 total bins per trial). Mean group-level patterns are displayed in Fig. 5 for each continuum.

We conducted two sets of analyses on these data. The first set focused on the response data and mirrored the analyses of the VAS task from study 1. In that task, participants responded on a seven-point scale to the tokens from each continuum. Four-parameter logistic regressions were used to measure individual differences in the slope and consistency of those responses. In the 2AFC task, participants made binary responses, which lent themselves better to a linear modeling approach. To estimate individual response functions, generalized linear mixed-effects models with binomial family functions were fit separately to the selection data for each of the four continua with the LME4 package (Bates *et al.*, 2015). The models included the fixed effect of continuum Step (continuous: 1–7, mean-centered and scaled) and the random effect of Participant on Step. Individual coefficients were extracted for each participant from each model and are shown in Fig. 4(B). These individual slope values were then used as the

dependent variable in subsequent models. Note that in this analysis, participants with steeper slopes in the response function have more negative slope estimates. Subsequent analyses were conducted via linear mixed-effects models constructed with the LME4 package unless otherwise noted (Bates *et al.*, 2015). The EMMEANS package was used to obtain estimated marginal means from the models and compute contrasts (Lenth, 2025).

The second set of analyses focused on the eye-gaze data. We took a growth curve analysis approach (Mirman, 2014), which involved creating third-order orthogonal polynomials with the GAZER package (Geller *et al.*, 2020). These orthogonal polynomials allowed us to model the steepness of the slope (linear), the sharpness of the central change in slope (quadratic), and the sharpness of two changes in slope (cubic) independently over time. Similarly to the response data, we estimated individual growth curves for each continuum. Linear mixed-effects models were fit separately to the proportion of upper image fixations for each of the four continua. The models included the fixed effects of trial number (continuous: 1 – 70, mean-centered and scaled), continuum Step (continuous: 1 – 7, mean-centered and scaled), the three time terms (linear, quadratic, and cubic), and the pairwise interactions between each time term and Step. The models also included the random effect of Participant on all time terms and Step and the random effect of Participant on the interactions between the linear time term and Step and the quadratic time term and Step. The random slope estimates for these two interactions were extracted for each participant from each model, and the fixed effects of Step and the relevant time term were added. These individual slope values were then used as dependent variables in the models described below in Secs. III B 4 a and III B 4 b.

To get a better sense of how to interpret the linear and quadratic polynomials, consider the *bee-pea* continuum in Fig. 5 (top row). At step 1, both age groups look at the upper image of a bee at a similar rate throughout the 200 – 1000 ms time window because this token is the most *bee*-like in the continuum. At step 4, close to the boundary between *bee* and *pea*, both age groups tend to look at the upper image at the start of the 200 ms but by 1000 ms are consistently looking elsewhere. In other words, as the continuum step increases, looks to the upper image decrease throughout the time window of interest. This change in looking behavior over time as a function of continuum step describes the linear time term by continuum step interaction. We interpret larger (more negative) estimates as evidence of more categorical perception, because a steeper linear slope reflects a more dramatic change in looking behavior from the beginning to the end of the trial across the continuum. For the quadratic effects, consider again the looking behavior in steps 1 and 4 of the *bee-pea* continuum in Fig. 5. At step 1, looks to the upper image of a bee are relatively stable across the analysis window. At step 4, we see consistent looks to the upper image for the first 200 ms of the time window of interest before they quickly drop off; that is, as continuum step increases, there is not only a decrease in looks to the upper image across the entire trial, but also a faster change in

looking behavior at a certain point in the trial compared to step 1. This shift in both the direction and rate of looking at the upper image describes the quadratic time term by continuum step interaction. We also interpret larger (more negative) estimates as evidence of more categorical perception, as a narrower quadratic curve reflects a more dramatic change in looking behavior at the inflection point of the trial across the continuum. Figure 6 in the [supplementary material](#) visualizes these linear and quadratic changes in looking behavior by comparing a categorical older adult participant (top row) to a gradient younger adult participant (bottom row).

2. 2AFC slope measure by phonetic contrast and age.

We first investigated the effects of Manner, Voice/Place, and Age Band on individual sensitivity to phonetic gradiency. Manner, Voice/Place, and Age Band were sum contrast-coded as described in study 1. We initially constructed a linear mixed-effects model with fixed effects of Manner, Voice/Place, Age Band, and their interactions and a random effect of Participant on the intercept; however, participants did not vary enough to support a random effects structure. Instead, we constructed a linear regression model that included Manner, Voice/Place, Age Band, and their interactions. The results are shown in Table V.

We observed significant main effects of Manner and Voice/Place that were qualified by a significant two-way interaction. None of the effects including Age Band were significant. The main effect of Manner reflected a steeper decrease in the likelihood of selecting the upper image with an increase in continuum step (i.e., a steeper slope) for stops compared to fricatives. The main effect of Voice/Place reflected a steeper decrease in the likelihood of selecting the upper image with an increase in continuum step for voicing compared to place of articulation contrasts. Both of these main effects were driven by the stop-voicing continuum (*bee-pea*; see below).

To investigate the Manner by Voice/Place interaction, we compared the estimated marginal means averaged across Age Band. Overall, we observed the steepest slopes for the stop-voicing continuum (*bee-pea*: $M = -21.49$, 95% CI $[-22.48, -20.49]$), followed by the stop-place continuum (*buck-duck*: $M = -15.21$, 95% CI $[-16.21, -14.22]$),

TABLE V. Linear regression model output predicting individual 2AFC slope values. Manner (stop, fricative), Voice/Place (voicing, place of articulation), Age Band (younger, older), and their interactions were included.

Term	$\hat{\beta}$	95% CI	t	p
Intercept	-12.90	[-13.40, -12.40]	-50.87	<0.001
Manner	-10.89	[-11.89, -9.89]	-21.48	<0.001
Voice/Place	-2.25	[-3.25, -1.25]	-4.43	<0.001
Age Band	0.19	[-0.81, 1.19]	0.37	0.710
Manner $\times$ Voice/Place	-8.05	[-10.05, -6.05]	-7.93	<0.001
Manner $\times$ Age Band	-0.41	[-2.41, 1.59]	-0.40	0.689
Voice/Place $\times$ Age Band	0.55	[-1.45, 2.55]	0.54	0.588
Manner $\times$ Voice/Place $\times$ Age Band	1.12	[-2.88, 5.12]	0.55	0.582

TABLE VI. Linear mixed-effects model output predicting individual coefficients for the interaction between the linear time term and continuum step. Manner (stop, fricative), Voice/Place (voicing, place of articulation), Age Band (younger, older), and their interactions were included as fixed effects, and Participant was included as a random effect.

Term	$\hat{\beta}$	95% CI	t	df	p
Intercept	− 0.55 [− 0.60, − 0.51]	− 23.27	58.28	<0.001	
Manner	− 0.10 [− 0.14, − 0.06]	− 4.68	173.26	<0.001	
Voice/Place	− 0.19 [− 0.24, − 0.15]	− 8.96	172.28	<0.001	
Age Band	0.01 [− 0.09, 0.10]	0.17	58.28	0.862	
Manner × Voice/Place	− 0.70 [− 0.79, − 0.62]	− 16.29	172.28	<0.001	
Manner × Age Band	− 0.01 [− 0.10, 0.07]	− 0.35	173.26	0.730	
Voice/Place × Age Band	− 0.04 [− 0.12, 0.05]	− 0.85	172.28	0.396	
Manner × Voice/Place × Age Band	− 0.04 [− 0.21, 0.13]	− 0.51	172.28	0.614	

then the fricative-place continuum (*sack-shack*: $M = -8.34$, 95% CI [-9.35, -7.34]), and finally the fricative-voicing continuum (*face-vase*: $M = -6.57$, 95% CI [-7.57, -5.56]). Pairwise comparisons within Voice/Place revealed a significant difference between fricatives and stops for both place of articulation (*sack-shack* vs *buck-duck*: $t(222) = -9.58$, $p < 0.001$) and voicing (*face-vase* vs *bee-pea*: $t(222) = 20.80$, $p < 0.001$). Within Manner, pairwise comparisons revealed a significant difference between place of articulation and voicing for both fricatives (*sack-shack* vs *face-vase*: $t(222) = -2.46$, $p = 0.015$) and stops (*buck-duck* vs *bee-pea*: $t(222) = 8.79$, $p < 0.001$).

3. Language and hearing predictors of 2AFC response slope

We next investigated the effects of language/hearing scores on individual sensitivity to phonetic gradiency. As in study 1, we analyzed participants' language composite scores, pure-tone average thresholds, and QuickSIN scores. Because the specific continuum was not a factor of interest in this analysis, we constructed a linear regression model including Age Band and each of the three test scores. The language/hearing scores were mean-centered and scaled, and Age Band was sum contrast-coded as before. The results

TABLE VII. Linear mixed-effects model output predicting individual coefficients for the interaction between the quadratic time term and continuum step. Manner (stop, fricative), Voice/Place (voicing, place of articulation), Age Band (younger, older), and their interactions were included as fixed effects, and Participant was included as a random effect.

Term	$\hat{\beta}$	95% CI	t	df	p
Intercept	-0.08 [-0.12, -0.04]	-4.05	58.26	<0.001	
Manner	-0.01 [-0.04, 0.02]	-0.63	173.10	0.529	
Voice/Place	-0.03 [-0.06, 0.00]	-1.85	172.25	0.067	
Age Band	0.09 [0.01, 0.16]	2.24	58.26	0.029	
Manner $\times$ Voice/Place	-0.09 [-0.15, -0.03]	-2.82	172.25	0.005	
Manner $\times$ Age Band	0.02 [-0.04, 0.08]	0.65	173.10	0.518	
Voice/Place $\times$ Age Band	0.01 [-0.05, 0.07]	0.37	172.25	0.710	
Manner $\times$ Voice/Place $\times$ Age Band	-0.02 [-0.14, 0.11]	-0.28	172.25	0.781	

are shown in the [supplementary material](#), Table XI. None of the scores significantly predicted response slope nor did Age Band.

4. Eye tracking results by phonetic contrast and age

To address our primary goal in study 2, we investigated the effects of Manner, Voice/Place, and Age Band on an implicit measure of sensitivity to phonetic gradiency: the time-course of eye fixations. We constructed two linear mixed-effects models with fixed effects of Manner, Voice/Place, Age Band, and their interactions and a random effect of Participant on the intercept. The fixed effects were sum contrast-coded as described in study 1. One model predicted the individual coefficients for the linear time term by continuum step interaction; the results are shown in Table VI. The other model predicted the individual coefficients for the quadratic time term by continuum step interaction; the results are shown in Table VII.

a. Effects on the linear term. This analysis modeled the relation between continuum step and the linear change in upper image fixations over time, with larger (more negative) values reflecting more categorical perception. We observed main effects of Manner and Voice/Place that were qualified by a two-way interaction. None of the effects including Age Band were significant. The main effect of Manner reflected a steeper decrease in fixations on the upper image over time as continuum step increased for stops compared to fricatives. The main effect of Voice/Place reflected a steeper decrease in fixations on the upper image over time as continuum step increased for voicing compared to place of articulation contrasts. Both of these main effects were driven by the stop-voicing continuum (*bee-pea*; see below).

To investigate the Manner by Voice/Place interaction, we compared the estimated marginal means averaged across Age Band. We observed the steepest (most negative) slopes for the stop-voicing continuum (*bee-pea*: $M = -0.88$, 95% CI [-0.94, -0.82]), followed by the fricative-place continuum (*sack-shack*: $M = -0.58$, 95% CI [-0.64, -0.52]), then the fricative-voicing continuum (*face-vase*: $M = -0.42$, 95% CI [-0.49, -0.36]), and finally the stop-place continuum (*buck-duck*: $M = -0.33$, 95% CI [-0.39, -0.27]). Pairwise comparisons within Voice/Place revealed significant differences between fricatives and stops for both place of articulation (*sack-shack* vs *buck-duck*: $t(179) = -8.04$, $p < 0.001$) and voicing (*face-vase* vs *bee-pea*: $t(179) = 14.56$, $p < 0.001$). Within Manner, pairwise comparisons revealed significant differences between place of articulation and voicing for both fricatives (*sack-shack* vs *face-vase*: $t(178) = -5.07$, $p < 0.001$) and stops (*buck-duck* vs *bee-pea*: $t(178) = 17.64$, $p < 0.001$).

b. Effects on the quadratic term. This analysis modeled the relation between continuum step and the quadratic change in upper image fixations over time, with more negative-going values reflecting more categorical

perception. We observed a significant two-way interaction between Manner and Voice/Place, as well as a main effect of Age Band.

To investigate the Manner by Voice/Place interaction, we compared the estimated marginal means averaged across Age Band. We observed the sharpest narrowing of the quadratic curves as continuum step increased for the stop-voicing continuum (*bee-pea*: $M = -0.12$, 95% CI $[-0.17, -0.07]$), followed by the fricative-place continuum (*sack-shack*: $M = -0.08$, 95% CI $[-0.13, -0.03]$), then the fricative-voicing continuum (*face-vase*: $M = -0.06$, 95% CI $[-0.11, -0.02]$), and finally the stop-place continuum (*buck-duck*: $M = -0.05$, 95% CI $[-0.09, 0.00]$). Pairwise comparisons within Voice/Place revealed significant differences between fricatives and stops for voicing (*face-vase* vs *bee-pea*: $t(179) = 2.40$, $p = 0.018$) but not for place of articulation (*sack-shack* vs *buck-duck*: $t(179) = -1.52$, $p = 0.131$). Within Manner, pairwise comparisons revealed significant differences between place of articulation and voicing for stops (*buck-duck* vs *bee-pea*: $t(178) = 3.26$, $p = 0.001$) but not for fricatives (*sack-shack* vs *face-vase*: $t(178) = -0.68$, $p = 0.500$). To investigate the main effect of Age Band, we compared the estimated marginal means averaged across Manner and Voice/Place. Older adults ($M = -0.12$, 95% CI $[-0.18, -0.06]$) displayed more extreme changes in looking behavior than younger adults, such that looks to the upper image began to decrease more rapidly during the trial as continuum step increased ($M = -0.03$, 95% CI $[-0.09, 0.02]$; $t(60.1) = -2.20$, $p = 0.032$). This age effect is best illustrated in Fig. 5 by comparing steps 3 and 4 of the *sack-shack* continuum.

5. Language and hearing predictors of the eye tracking data

As we did for the response data, we investigated the effects of language/hearing scores on individual sensitivity to phonetic gradiency. We constructed two linear mixed-effects models, one for each time term by continuum step measure (linear and quadratic). Both models included fixed effects of Age Band (sum contrast-coded) plus language composite scores, pure-tone average thresholds, and QuickSIN scores (mean-centered and scaled). Random intercepts for Participants and Continua were also included as in study 1. The results for the linear time term are shown in the [supplementary material](#), Table XII, and those for the quadratic time term are shown in the [supplementary material](#), Table XIII. In neither model did the language/hearing scores significantly predict the relation between time and continuum step, nor did Age Band.

C. Interim discussion

Beginning with the 2AFC response data, we observed overall steeper slopes for stop continua than for fricative continua. This finding broadly aligns with the VAS rating data from study 1 and previous research showing that listeners typically perceive stops more categorically than they do

fricatives. The largest difference in slope was observed between the two voicing contrasts, with *bee-pea* being the steepest of the four continua and *face-vase* being the shallowest. Unlike in the VAS task, we did not observe any age-related differences in sensitivity, nor any language- or hearing-related differences in the 2AFC task.

Regarding the eye tracking data, we observed an overall tendency to look at the upper image at the start of the trial. This upper-image bias was a product of our unconventional choice to use a vertical upper vs lower orientation rather than a horizontal left vs right orientation for the images. This decision was motivated by our desire to compare the healthy participant samples in this study to a brain-injured participant sample that includes individuals with hemifield neglect. Despite this unexpected pattern of initial looking behavior, we observed different patterns of sensitivity to manner of articulation and voicing/place of articulation on the linear and quadratic growth curve polynomials.

When analyzing the individual coefficients for the linear time term by continuum step interaction across continua, we found that the decrease in upper image fixations was generally steeper for stops than for fricatives. However, compared to the 2AFC response data, the distinction between stops and fricatives was not as clear-cut, with the stop-voicing continuum (*bee-pea*) having the steepest linear effects and the stop-place continuum (*buck-duck*) having the shallowest.

When we analyzed the individual coefficients for the quadratic time term by continuum step interaction across continua, only the stop-voicing continuum (*bee-pea*) differed from its counterparts, with sharper decreases in upper image looks compared to the other voicing continuum (*face-vase*) and to the other stop continuum (*buck-duck*). We also found that older adults had more extreme decreases in upper image looks over time and step than younger adults regardless of continuum; however, when controlling for language and hearing abilities, we did not observe this difference between age groups. The quadratic time term may capture uncertainty over the time course of looking behavior that the linear term cannot. This difference in the sensitivity of the time terms is best illustrated by Fig. 6 in the [supplementary material](#); notice how the tails of the dark blue lines depicting the quadratic time term deviate from the dark green lines depicting the linear time term, particularly at continuum steps 3–5, where the stimuli are the most ambiguous. It is possible that older adults experience greater uncertainty during online processing, despite being more categorical and consistent in their offline responses (at least in the VAS task in study 1).

IV. GENERAL DISCUSSION

In the present study, we investigated how sensitivity to within-category phonetic variation differs as a function of task demands, phonetic contrasts, and age. We conducted two experiments that each employed a unique measure of acoustic-phonetic sensitivity. Study 1 used a VAS rating task, which allowed listeners to make fine-grained phonetic ratings. Study 2 used a 2AFC eye tracking task that

combined binary categorization responses with time-resolved fixation data. We used these two different tasks to ask three research questions: (1) Can more gradient tasks elicit more gradient responses? (2) Do younger and older adults differ in phonetic gradiency and consistency, and if so, do these differences arise under explicit or implicit task demands? (3) Do individual differences in language and hearing abilities predict the gradiency of task responses?

A. Task differences in assessing gradiency

Researchers have many tasks to choose from when assessing sensitivity to phonetic category structure (see Apfelbaum *et al.*, 2022). There are two sets of task demands that can impact how gradient or categorical a listener appears to be. The first set of task demands relates to the response options; for example, if we compare 2AFC to VAS, as we did in the present study, we can see that expanding the set of options from a binary choice (i.e., “Is this token in category A or category B?”) to a range of choices (i.e., “Where does this token fall on a scale from category A to category B?”) increases the precision of the responses the participant can provide. The second set of task demands relates to the stage of processing. Both the VAS and 2AFC tasks provide an offline, explicit measure that reflects not only low-level speech perception but also high-level decision criteria. By contrast, eye tracking provides a more online, implicit measure that is less susceptible to top-down or strategic influences. In the present study, we compared VAS and eye tracking to understand how two gradient measures that differ in modality can capture sensitivity to phonetic detail.

To enhance our measures of gradiency and consistency, we manipulated the phonetic contrasts included in the VAS and 2AFC eye tracking tasks. Specifically, we included four phonetic contrasts that crossed two sets of articulatory features. Manner of articulation was the primary feature of interest based on previous work showing that stops elicit more categorical responses than fricatives (Fry *et al.*, 1962; Healy and Repp, 1982; Massaro and Cohen, 1983). As a secondary feature, voicing was compared to place of articulation. Of particular interest was the degree to which different measures track differences in gradiency between phonetic contrasts.

Both the VAS and 2AFC tasks provided evidence for gradiency in speech perception. In the VAS results, fricatives were overall more gradient than stops, but the type of phonetic contrast (voicing vs place of articulation) had a strong effect on both the slope and consistency measures. In particular, performance on the fricative-place continuum (*sack-shack*) was characterized by relatively inconsistent responses, which resulted in relatively shallow slopes. In the 2AFC results, we also observed overall steeper slopes for stops rather than fricatives. Although the manner of articulation pattern was consistent across tasks, the difference in gradiency can be seen by comparing the two voicing continua (*bee-pea* vs *face-vase*). In the VAS response slopes, both continua exhibited similar slope and consistency

measures despite differing in manner of place. In the 2AFC data, by contrast, the stop-voicing continuum (*bee-pea*) elicited the overall steepest response slopes and the fricative-voicing continuum (*face-vase*) the shallowest. Taken together, these results suggest that listeners are particularly sensitive to within-category phonetic variation for fricatives and that both VAS and 2AFC tasks reveal fairly similar distinctions between continua.

We also observed clear evidence of within-category sensitivity in the eye tracking results. The linear term reflected the expected pattern of steeper (more categorical) slopes for stops than for fricatives; however, the stop-place continuum (*buck-duck*) provided a particularly interesting comparison to the VAS and 2AFC response data. Although *buck-duck* elicited the steepest, most consistent VAS responses and steeper 2AFC response slopes than either fricative continuum, it elicited the shallowest linear slope across continuum step in the eye tracking data. The effect of *buck-duck* on the quadratic term was also less steep than the other stop pair (*bee-pea*). Although the latter difference was driven mostly by the strong effect of the *bee-pea* continuum, the *buck-duck* continuum in general appeared more categorical on our explicit measures and more gradient on our implicit measures. Similarly, the *sack-shack* continuum looked more categorical in the linear term results than it did in the VAS data. This is potentially because the explicit VAS responses were relatively inconsistent for *sack-shack*, in turn making the responses appear more gradient, compared to the implicit eye-gaze behavior.

What might account for these discrepancies between the explicit and implicit measures? That is, do they reflect differences in granularity, differences in the underlying neurocognitive mechanisms, or simply differences in strategy? Eye tracking measures have well-advertised strengths in that they are relatively implicit and likely less susceptible to participant strategies (Alloppenna *et al.*, 1998). Further, by capturing a listener’s decision process in real time, eye tracking measures afford more opportunities for participants to demonstrate perceptual gradiency than response measures. Recall that one hypothesis for the difference in gradiency between fricatives and stops is that fricative (and vowel) segments are longer in duration and persist longer in auditory memory (Pisoni, 1972). The online nature of eye tracking might allow listeners to demonstrate gradient patterns of perception for stops and fricatives alike, whereas this information is harder to maintain for short-duration segments in the explicit, *post hoc* decisions made in both the VAS and 2AFC tasks. It is possible that the *buck-duck* continuum showed the largest discrepancy between measures because of its implausibility in real speech; that is, unlike the other continua, it would be impossible to produce the intermediate places of articulation between /b/ and /d/. We speculate that this makes the *buck-duck* continuum especially susceptible to strategic influences during explicit tasks. Determining whether task-based differences in gradiency are meaningful requires future experimentation by comparing various measures and populations. In particular, we look forward to exploring the question of explicit vs implicit measures of

consistency with our brain-injured population, which may reveal further dissociation in online vs decision-level measures of phonetic perception.

B. Age-related differences in gradiency

Although most of the attention on the evolution of phonetic sensitivity over the lifespan has concentrated on early periods of language acquisition (e.g., Kim *et al.*, 2025; Kuhl, 1991), one's sensitivity to phonetic category may continue to evolve over the adult lifespan. Differences between younger and older adults were most evident in the VAS data. Older adults exhibited both steeper slopes and more consistent responses than younger adults, which suggests that increases in age are associated with increases in categorical perception. This was true even when controlling for individual differences in language and hearing ability suggesting that differences in VAS measures are not easily explained by other factors (such as central or peripheral hearing ability) that also change as a function of age. This pattern is consistent with the view that repeated exposure to phonetic tokens over the course of one's lifetime might refine and focus the precision of phonetic categories (e.g., Kleinschmidt and Jaeger, 2015). Intriguingly, this finding is a departure from other studies in which listeners showed increasingly gradient responses (i.e., shallower slopes) over development between 7 and 18 years (McMurray *et al.*, 2018) and between 18 and 40 years (Kapnoula and Samuel, 2024). However, we should interpret our data with some caution, given that age-related effects were not strong in the 2AFC eye tracking task; the effect of age was only evident on the quadratic term, not on the linear term, and did not emerge in the 2AFC responses.

C. Language and hearing abilities and individual differences in phonetic perception

The literature has yet to resolve whether there is an "optimal" psychometric pattern in speech perception (McMurray, 2022). One might argue that a steeper categorization slope is more optimal because attentional resources are not being used to track sub-phonemic detail. In other words, having a more precisely tuned phonetic category boundary theoretically frees up attentional resources to pursue other perceptual problems. Conversely, a shallower categorization slope could also be argued to be more optimal as fine-grained acoustic-phonetic details can be used to resolve ambiguities arising from coarticulation (Magnuson *et al.*, 2003) or accent (Xie, *et al.*, 2017). That is, sensitivity to sub-phonemic detail can be useful for dealing with variation in the speech signal.

We took an individual differences approach to understand the effects of this potential trade-off between efficiency (more categorical perception) and precision (more gradient perception). Specifically, we used hearing acuity, general language ability, and speech-in-noise perception to predict gradiency and consistency on our tasks. However, the relations between these measures were fairly minimal.

Only VAS response slopes were predicted by our language/hearing measures, with more categorical responses (steeper slopes) associated with poorer language skills. By contrast, in a recent study with school-aged children, Kim *et al.* (2025) found that poorer language and reading skills were associated with lower VAS response consistency but not with VAS response slopes. Importantly, their study included participants with language and reading disorders, whereas we specifically excluded participants with such backgrounds. We await the replication of their findings in future work. Future work will also be needed to dissociate task-based effects from true differences in phonetic processing and to determine how individual differences in phonetic processing interact with (or are isolated from) other language and cognitive abilities.

D. Conclusion

We assessed the gradiency of speech perception via three distinct tasks. The VAS task in study 1 provided an explicit measure of phonetic sensitivity that allowed for gradient responses along a seven-point scale. The 2AFC eye tracking task in study 2 provided both an explicit measure that limited participants to binary responses and an implicit measure using fixation behavior. The explicit VAS and 2AFC responses replicated well-attested differences in performance on stop contrasts compared to fricative contrasts. The implicit eye tracking measures also exhibited this pattern but did not align fully with the response data across continua, possibly reflecting differences in online processing vs *post hoc* selection. In general, the VAS response data were more sensitive to age and individual differences in language composite scores than either the 2AFC or eye tracking data. For now, it is unclear whether the degree to which VAS and eye tracking data can be associated with other measures is because of differences in the fidelity of measurement of these two tasks or differences in online vs offline phonetic decision processes; an intriguing avenue for future study.

SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for stimulus information (Appendix A) and additional tables/figures (Appendix B).

ACKNOWLEDGMENTS

The research reported in this publication was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health, Grant No. R01DC013064.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The authors obtained ethics approval for the research reported in this article from the University of Connecticut

Institutional Review Board under protocol number H21-0046.

DATA AVAILABILITY

The data that support the findings of this study are openly available on the Open Science Framework.

- Alloppenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models," *J. Mem. Lang.* **38**(4), 419–439.
- Apfelbaum, K. S., Kutlu, E., McMurray, B., and Kapnoula, E. C. (2022). "Don't force it! Gradient speech categorization calls for continuous categorization tasks," *J. Acoust. Soc. Am.* **152**(6), 3728–3745.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Software* **67**(1), 1–48.
- Boersma, P., and Van Heuven, V. (2001). "Speak and unSpeak with PRATT," *Glott Int.* **5**(9/10), 341–347.
- Crinnion, A. M., Magnuson, J. S., and Myers, E. B. (2026). "Phonetic stability predicts individual differences in acoustic-phonetic cue weighting," *Psychon. Bull. Rev.* (in press).
- Crowder, R. G. (1973). "Representation of speech sounds in precategorical acoustic storage," *J. Exp. Psychol.* **98**, 14–24.
- Dial, H. R., McMurray, B., and Martin, R. C. (2019). "Lexical processing depends on sublexical processing: Evidence from the visual world paradigm and aphasia," *Atten. Percept. Psychophys.* **81**(4), 1047–1064.
- Feltner, C., Wallace, I. F., Kistler, C. E., Coker-Schwimmer, M., and Jonas, D. E. (2021). "Screening for hearing loss in older adults: Updated evidence report and systematic review for the US Preventive Services Task Force," *JAMA* **325**(12), 1202–1215.
- Fidler, L. J., Plante, E., and Vance, R. (2011). "Identification of adults with developmental language impairments," *Am. J. Speech. Lang. Pathol.* **20**(1), 2–13.
- Fidler, L. J., Plante, E., and Vance, R. (2013). "Erratum," *Am. J. Speech. Lang. Pathol.* **22**(3), 577.
- Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). "The identification and discrimination of synthetic vowels," *Lang. Speech* **5**(4), 171–189.
- Fuhrmeister, P., and Myers, E. B. (2021). "Structural neural correlates of individual differences in categorical perception," *Brain Lang.* **215**, 104919.
- Fuhrmeister, P., Phillips, M. C., McCoach, D. B., and Myers, E. B. (2023). "Relationships between native and non-native speech perception," *J. Exp. Psychol. Learn. Mem. Cogn.* **49**(7), 1161–1175.
- Geller, J., Winn, M. B., Mahr, T., and Mirman, D. (2020). "GazeR: A package for processing gaze position and pupil size data," *Behav. Res. Methods* **52**, 2232–2255.
- Goman, A. M., and Lin, F. R. (2016). "Prevalence of hearing loss by severity in the United States," *Am. J. Public Health* **106**(10), 1820–1822.
- Goossens, T., Vercammen, C., Wouters, J., and Van Wieringen, A. (2017). "Masked speech perception across the adult lifespan: Impact of age and hearing impairment," *Hear. Res.* **344**, 109–124.
- Healy, A. F., and Repp, B. H. (1982). "Context independence and phonetic mediation in categorical perception," *J. Exp. Psychol. Hum. Percept. Perform.* **8**(1), 68–80.
- Helm-Estabrooks, N. (2001). *Cognitive Linguistic Quick Test* (Psychological Corp., San Antonio, TX).
- Honda, C. T., Clayards, M., and Baum, S. R. (2024). "Exploring individual differences in native phonetic perception and their link to nonnative phonetic perception," *J. Exp. Psychol. Hum. Percept. Perform.* **50**(4), 370–394.
- Hubbard, D. J., and Assmann, P. F. (2013). "Perceptual adaptation to gender and expressive properties in speech: The role of fundamental frequency," *J. Acoust. Soc. Am.* **133**(4), 2367–2376.
- Kapnoula, E. C., Edwards, J., and McMurray, B. (2021). "Gradient activation of speech categories facilitates listeners' recovery from lexical garden paths, but not perception of speech-in-noise," *J. Exp. Psychol. Hum. Percept. Perform.* **47**(4), 578–595.
- Kapnoula, E. C., and 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 Lang.* **223**, 105031.
- Kapnoula, E. C., and Samuel, A. G. (2024). "Sensitivity to subphonemic differences in first language predicts vocabulary size in a foreign language," *Lang. Learn.* **74**(4), 950–984.
- Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., and McMurray, B. (2017). "Evaluating the sources and functions of gradience in phoneme categorization: An individual differences approach," *J. Exp. Psychol. Hum. Percept. Perform.* **43**(9), 1594–1611.
- Kawahara, H. (2008). "TANDEM-STRAIGHT, a research tool for L2 study enabling flexible manipulations of prosodic information," *Proc. Speech Prosody* **2008**, 619–628.
- Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (2004). "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.* **116**(4), 2395–2405.
- Kim, H., Klein-Packard, J., Sorensen, E., Oleson, J., Tomblin, B., and McMurray, B. (2025). "Speech categorization consistency is associated with language and reading abilities in school-age children: Implications for language and reading disorders," *Cognition* **263**, 106194.
- Kleinschmidt, D. F., and Jaeger, T. F. (2015). "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," *Psychol. Rev.* **122**(2), 148–203.
- Kong, E. J., and Edwards, J. (2016). "Individual differences in categorical perception of speech: Cue weighting and executive function," *J. Phonetics* **59**, 40–57.
- Kuhl, P. K. (1991). "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," *Percept. Psychophys.* **50**(2), 93–107.
- Kutlu, E., Chiu, S., Sorensen, E., Oleson, J., and McMurray, B. (2023). "Network diversity leads to more flexibility in speech categorization," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, Sydney, Australia (July 26–29, 2023) (eScholarship), Vol. 45, available at <https://escholarship.org/uc/item/00r6d6kc>.
- Kutlu, E., Kim, H., and McMurray, B. (2026). "Longitudinal changes in the structure of speech categorization across school age years: Evidence against perceptual narrowing in later developmental periods," *Dev. Psy.* **29**(1), e70085.
- Lenth, R. V. (2025). "EMMEANS: Estimated marginal means, aka least squares means," <https://cran.r-project.org/package=emmeans> (Last viewed March 4, 2026).
- Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). "The discrimination of speech sounds within and across phoneme boundaries," *J. Exp. Psychol.* **54**(5), 358–368.
- Luthra, S., Mechtenberg, H., and Myers, E. B. (2021). "Perceptual learning of multiple talkers requires additional exposure," *Atten. Percept. Psychophys.* **83**(5), 2217–2228.
- Magnuson, J. S., McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2003). "Lexical effects on compensation for coarticulation: A tale of two systems?" *Cogn. Sci.* **27**(5), 801–805.
- Massaro, D. W., and Cohen, M. M. (1983). "Categorical or continuous speech perception: A new test," *Speech Commun.* **2**(1), 15–35.
- McMurray, B. (2022). "The myth of categorical perception," *J. Acoust. Soc. Am.* **152**(6), 3819–3842.
- McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., and Subik, D. (2008). "Gradient sensitivity to within-category variation in words and syllables," *J. Exp. Psychol. Hum. Percept. Perform.* **34**(6), 1609–1631.
- McMurray, B., Danelz, A., Rigler, H., and Seedorff, M. (2018). "Speech categorization develops slowly through adolescence," *Dev. Psychol.* **54**(8), 1472–1491.
- McMurray, B., and Jongman, A. (2011). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," *Psycholog. Rev.* **118**(2), 219–246.
- McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2002). "Gradient effects of within-category phonetic variation on lexical access," *Cognition* **86**(2), B33–B42.
- Miller, J. L. (1994). "On the internal structure of phonetic categories: A progress report," *Cognition* **50**(1–3), 271–285.
- Mirman, D. (2014). "Growth curve analysis: A hands-on tutorial on using multilevel regression to analyze time course data," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, Quebec City, Canada

- (July 23–26, 2014) (eScholarship), Vol. 36, available at <https://escholarship.org/uc/item/1dp5q4k2>.
- Morice, R., and McNicol, D. (1985). “The comprehension and production of complex syntax in schizophrenia,” *Cortex* **21**(4), 567–580.
- Myers, E., Phillips, M., and Skoe, E. (2024). “Individual differences in the perception of phonetic category structure predict speech-in-noise performance,” *J. Acoust. Soc. Am.* **156**(3), 1707–1719.
- 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.
- Newman, R. S., Clouse, S. A., and Burnham, J. L. (2001). “The perceptual consequences of within-talker variability in fricative production,” *J. Acoust. Soc. Am.* **109**(3), 1181–1196.
- Oldfield, R. C. (1971). “The assessment and analysis of handedness: The Edinburgh inventory,” *Neuropsychologia* **9**(1), 97–113.
- Pinheiro J, Bates D, R Core Team (2025). “NLME: Linear and nonlinear mixed effects models,” <https://cran.r-project.org/web/packages/nlme/nlme.pdf> (Last viewed March 4, 2026).
- Pisoni, D. B. (1972). “On the nature of categorical perception of speech sounds,” Diss. Abstr. Int. **32**(11-B), 6693–6694.
- Pisoni, D. B., and Tash, J. (1974). “Reaction times to comparisons within and across phonetic categories,” *Percept. Psychophys.* **15**, 285–290.
- Porretta, V., Kyröläinen, A. J., Van Rij, J., and Järvikivi, J. (2018). “Visual world paradigm data: From preprocessing to nonlinear time-course analysis,” in *Intelligent Decision Technologies 2017*, edited by I. Czarnowski, R. Howlett, and L. Jain (Springer, Cham, Switzerland).
- R Core Team (2025). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
- Rogers, J. C., and Davis, M. H. (2017). “Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds,” *J. Cogn. Neurosci.* **29**(5), 919–936.
- Semel, E., Wiig, E. H., and Secord, W. (2003). *Clinical Evaluation of Language Fundamentals*, 4th ed. (Psychological Corp., San Antonio, TX).
- Toscano, J. C., McMurray, B., Dennhardt, J., and Luck, S. J. (2010). “Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech,” *Psychol. Sci.* **21**(10), 1532–1540.
- Winn, M. B. (2020). “Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script,” *J. Acoust. Soc. Am.* **147**(2), 852–866.
- Xie, X., Theodore, R. M., and Myers, E. B. (2017). “More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories,” *J. Exp. Psychol. Hum. Percept. Perform.* **43**(1), 206–217.