

Article

Not peer-reviewed version

Beyond Categorical Perception: Gradient Lexical Tone Processing Revealed by Visual Analog Scale

[Bing Cheng](#), [Xi Xiang](#), Xiangrong Dai, [Yu Zou](#), Xiaojuan Zhang^{*}, [Yang Zhang](#)^{*}

Posted Date: 23 October 2025

doi: 10.20944/preprints202510.1824.v1

Keywords: categorical perception; gradient perception; response variability



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Beyond Categorical Perception: Gradient Lexical Tone Processing Revealed by Visual Analog Scale

Bing Cheng¹, Xi Xiang¹, Xiangrong Dai¹, Yu Zou¹, Xiaojuan Zhang^{1,*} and Yang Zhang^{2,*}

¹ English Department & Shaanxi Key Laboratory of AI-Empowered Language and Culture Research, School of Foreign Studies, Xi'an Jiaotong University, 710049, China

² Department of Speech-Language-Hearing Sciences and Center for Neurobehavioral Development, University of Minnesota, Minneapolis

* Correspondence: zhangxiaojuan@xjtu.edu.cn (X.Z.); zhanglab@umn.edu (Y.Z.)

Abstract

Purpose: While the Visual Analog Scale (VAS) has revealed gradient perception in segmental speech sounds, its application to lexical tones, a critical yet understudied suprasegmental feature, has been absent. This study investigated lexical tone categorization using VAS, directly comparing it with traditional two-alternative forced-choice (2AFC). **Method:** Eighty-four native speakers categorized an 11-step F0 continuum from Mandarin Tone 1 to Tone 2 in both tasks. Four-parameter logistic functions yielded slope (categorization sharpness) and response variability. Within-category sensitivity (Δ) was quantified from VAS responses. **Results:** Paired Wilcoxon signed-rank tests showed significantly shallower slopes ($p < .001$, $r = .76$) and lower variability ($p < .001$, $r = .87$) in VAS versus 2AFC. One-sample t-tests confirmed listeners discriminated fine-grained differences within categories, with Δ reliably exceeding zero (left: $M = 0.0335$, $t(270) = 8.89$, $p < .001$; right: $M = 0.0256$, $t(316) = 8.38$, $p < .001$). Crucially, slope and response variability were weakly correlated in VAS ($\rho = .27$, $p < .05$) but strongly negatively correlated in 2AFC ($\rho = -.67$, $p < .001$). Moreover, response variability correlated significantly across tasks ($\rho = .40$, $p < .001$), while slopes did not. **Conclusion:** These findings provide the direct evidence for gradient perception at the suprasegmental level, further establishing VAS as a sensitive tool for uncovering the nature of speech categorization. The dissociation between task-dependent gradiency and stable response variability helps reconcile apparent conflicts in the categorical perception literature, suggesting that these conflicts may stem from methodological constraints rather than genuine theoretical disagreements.

Keywords: categorical perception; gradient perception; response variability

Introduction

One central issue in cognitive science is to understand how organisms map continuous sensory input onto discrete perceptual categories. In speech perception, this process has been classically exemplified by categorical perception (CP), the phenomenon in which listeners perceive a continuum of acoustic variation as belonging to distinct, discrete categories. First documented for stop consonant voicing contrasts (Liberman et al., 1957), CP has since become a cornerstone concept in speech science and psycholinguistics, inspiring numerous studies on the ontogeny of speech sounds (Kuhl, 1987), first and second language acquisition (Eimas et al., 1987; Feng & Peng, 2023; Miyawaki et al., 1975), as well as potential CP impairments in clinical populations including individuals with autism, dyslexia, and specific language impairments (Coady et al., 2005; Robertson et al., 2009; Stewart et al., 2018).

Behaviorally, CP is characterized by three key features: (1) a sharp boundary in identification function, (2) enhanced discrimination at category boundaries, and (3) reduced sensitivity to acoustic differences within categories (Liberman et al., 1957; Repp, 1984; Xu et al., 2006). This pattern has been extensively observed across a range of speech contrasts, including stop consonants, nasals, vowels,

and lexical tones (e.g., Fitch et al., 1980; Johnson & Ralston, 1994; Larkey et al., 1978; Miller & Eimas, 1977; Minagawa-Kawai et al., 2007). Neurophysiological evidence corroborates these behavioral observations: cross-category contrasts consistently elicit larger mismatch negativity (MMN) and P300 responses than within-category contrasts (Frenck-Mestre et al., 2005; Kazanina et al., 2006; Näätänen et al., 2007; Xi et al., 2010). Given these converging lines of evidence, steep identification slopes in two-alternative forced-choice (2AFC) tasks have been widely interpreted as signatures of CP. Indeed, shallower slopes have been associated with various language impairments (Joanisse et al., 2000; Manis et al., 1997; Serniclaes et al., 2001; Werker & Tees, 1987) and illiteracy (Serniclaes et al., 2005), leading researchers to view sharp categorical boundaries as reflecting normative, successful pattern of speech perception.

However, despite its enduring theoretical influence and clinical relevance, the generality of CP in speech has faced increasing empirical and methodological scrutiny (see McMurray, 2022 for a review). Empirically, CP effects vary in strength across phonetic categories (Newell & Bülthoff, 2002). While robust and replicable for stop consonants (Liberman et al., 1961; Liberman et al., 1957, 1961), categorical effects are attenuated for fricatives (Healy & Repp, 1982) and appear largely absent for certain vowel contrasts (Fry et al., 1962; Stevens et al., 1969; van Hoesen & Schouten, 1999) and lexical tones (Abramson, 1975; Francis et al., 2003). More critically, accumulating evidence suggests that within phonetic categories, listeners retain continuous sensitivity to fine-grained acoustic detail. This gradient perception emerges particularly clearly in discrimination and rating tasks (Fry et al., 1962; Gerrits & Schouten, 2004; Hary & Massaro, 1982; Massaro & Cohen, 1983; Pisoni & Lazarus, 1974; Toscano et al., 2010), challenging the notion that speech perception involves strictly discrete categories. These inconsistencies have been shown to be further modulated by cognitive and contextual factors, including individual cognitive style (Yu, 2010; Yu et al., 2013) and task demands.

Methodological challenges have further undermined confidence in CP as a universal principle. Specifically, methodological choices may shape or even create the appearance of CP. Traditional 2AFC tasks enforce binary decisions that may exaggerate boundaries and mask within-category detail (Repp, 1984; Schouten et al., 2003). Similarly, ABX discrimination tasks may reflect memory load rather than genuine perceptual discontinuities (Pisoni, 1973). When more sensitive paradigms like 4IAX or 2IFC are used, CP effects often diminish or disappear (Carney et al., 1977; Gerrits & Schouten, 2004; Pisoni & Lazarus, 1974).

In response to these mounting empirical and methodological critiques, researchers have increasingly turned to the framework of gradient perception (GP), which posits that listeners maintain flexible, graded representations rather than collapsing continuous acoustic input into rigid, discrete categories (Apfelbaum et al., 2022; Fuhrmeister et al., 2021; Kapnoula et al., 2017; Myers et al., 2024; Oden et al., 1978; Toscano et al., 2010; see McMurray, 2022 for a review). This shift is not merely terminological; it reflects a deeper reconceptualization of how speech is encoded and categorized.

Speech processing models from diverse theoretical frameworks converge on a shared principle: categorization reflects graded activation of category likelihoods rather than all-or-none classification. The connectionist TRACE model (McClelland & Elman, 1986), the auditory change-sensitivity model (Kluender et al., 2003), the Ideal Adapter Bayesian framework (Kleinschmidt & Jaeger, 2015), and exemplar/episodic models (Goldinger, 1998) all support GP, suggesting that listeners retain sensitivity to fine-grained acoustic variation. Empirical evidence for GP spans multiple methodological approaches and levels of analysis. Behavioral studies using priming and rating scales demonstrate robust within-category sensitivity (Andruski et al., 1994; Massaro & Cohen, 1983; Miller & Volaitis, 1989). Real-time measures such as eye-tracking show that listeners simultaneously entertain multiple category hypotheses (Kapnoula et al., 2021; McMurray et al., 2002, 2009). Neurophysiological data further confirm that acoustic detail is preserved in neural representations even after categorical decisions are made (Ou & Yu, 2022; Sarrett et al., 2020; Toscano & McMurray, 2015).

This theoretical shift has necessitated a reevaluation of how speech categorization is measured. Traditional metrics, particularly the slope of identification functions derived from 2AFC tasks, have proven unreliable. Slopes are contrast-specific, unstable across contexts, and often fail to generalize across different speech continua (e.g., Honda et al., 2024a; Kapnoula et al., 2021). More fundamentally, slope is conceptually ambiguous because it conflates perceptual encoding with decisional noise and response strategy (Casillas, 2020; Honda et al., 2024a; Kim et al., 2025b; Ziegler et al., 2005). A listener with a genuinely gradient representation may produce a steep slope through deterministic responding or a shallow slope through probability-matching. Conversely, a listener with a discrete representation may appear gradient due to trial-level variability (Kim et al., 2025a; 2025b). Moreover, experimental manipulations of input variability in training studies demonstrate that slope is highly sensitive to methodological context rather than reflecting stable underlying representations (Zhang et al., 2021). Thus, slope alone cannot reliably diagnose the nature of perceptual representations.

To address these limitations, researchers have introduced response variability, which measures the stability of judgments across repeated presentations of identical stimuli (Apfelbaum et al., 2022; Kapnoula et al., 2017; Kim et al., 2025a; 2025b). Unlike slope, response variability may help distinguish genuinely gradient encoding from unstable categorical responding. However, in 2AFC tasks, response variability is mathematically constrained by mean response levels, limiting its interpretability (Apfelbaum et al., 2022). This constraint has motivated adoption of the Visual Analog Scale (VAS) paradigm (Massaro & Cohen, 1983), which allows listeners to make continuous judgments along a scale anchored by phonological endpoints. VAS preserves fine-grained perceptual detail and, critically, enables independent estimation of both slope and response variability, offering a richer window into categorization processes (Apfelbaum et al., 2022; Kong & Edwards, 2016; Massaro & Cohen, 1983).

Emerging evidence suggests that response variability may be a more stable and meaningful individual difference measure than slope. It shows stronger cross-continuum correlations across different phonetic contrasts (Kim et al., 2025a) and predicts real-world outcomes including speech-in-noise comprehension, downstream lexical processing, and reading abilities (Fuhrmeister et al., 2023; Kim et al., 2025b; Myers et al., 2024). These findings suggest that traditional 2AFC slopes may primarily reflect response stability rather than underlying perceptual gradiency, with response variability capturing more trait-like individual differences in categorization ability.

Despite these advances, an important gap remains: VAS has not been systematically applied to lexical tone perception. While CP has been extensively studied in segmental speech sounds, its extension to suprasegmental features, particularly lexical tone, has yielded inconsistent results. Tonal languages, which constitute over 70% of the world's linguistic system, use pitch variations to distinguish lexical meanings even when segmental content is identical (Yip, 2002). These tone contrasts are primarily realized through dynamic F0 trajectories, varying in height, direction, and contour (Gandour, 1983), with duration and amplitude functioning as secondary cues (Liu et al., 2011; Morton et al., 2008; Whalen & Xu, 1992; Wiener & Lee, 2020).

Mandarin Chinese, the most widely spoken tonal language, offers a compelling test case. Its four contrastive tones, Tone 1 (high-level), Tone 2 (mid-rising), Tone 3 (low-dipping), and Tone 4 (high-falling), create minimal pairs such as /ma/: mā (妈, "mother"), má (麻, "hemp"), mǎ (马, "horse"), and mà (骂, "scold"). This system requires listeners to track temporally extended F0 contours across the syllable, placing unique demands on perceptual encoding.

Empirical findings on CP of lexical tones have been mixed. Some studies report categorical effects among native speakers, including abrupt identification shifts, peak discrimination at category boundaries, and enhanced neural responses to cross-category contrasts (e.g., Hallé et al., 2004; Wang et al., 2003; Xu et al., 2006; Zhang et al., 2012). However, others weaker or absent categorical patterns. For instance, Abramson (1977) reported that Thai listeners showed uniformly good discrimination across a continuum of level tones, suggesting weak categorical effects. Similarly, Francis et al. (2003)

found that Cantonese listeners exhibited categorical-like perception only for contrasts involving contour direction changes but not for level-tone distinctions based on pitch height.

This inconsistency may reflect a deeper theoretical distinction: tone perception may be inherently more gradient than segmental perception. Unlike the brief, localized cues that define consonants, tonal information unfolds continuously over time. This temporal distribution may allow or even require listeners to preserve fine-grained sensitivity to within-category F0 variation. If so, traditional CP paradigms, designed primarily for segmental contrasts, may obscure the true nature of tonal encoding.

To address these ambiguities, the present study provides the direct empirical test of gradient perception in Mandarin lexical tone using the VAS task. We compared VAS responses with the traditional 2AFC task as native Mandarin listeners categorize stimuli along a synthesized F0 continuum from Tone 1 (high-level) to Tone 2 (mid-rising). Responses in both tasks were modeled using four-parameter logistic functions to derive two key indices: slope and response variability. Specifically, slope reflects the steepness of the categorization function, traditionally interpreted as the degree of CP versus GP. Response variability measures trial-level stability of judgments, capturing perceptual consistency independent of category boundary placement. Critically, the VAS format allows these two measures to be estimated independently, overcoming the mathematical coupling inherent in binary 2AFC tasks. This independence enables a more nuanced examination of perceptual mechanisms and individual differences. Additionally, VAS permits calculation of a within-category sensitivity index (Δ), which quantifies listeners' ability to discriminate acoustic differences within a single phonological category, a direct measure of gradient perception unavailable in forced-choice paradigms.

This experimental design addresses two central questions. First, does Mandarin tone perception exhibit gradient encoding within phonological categories? Given the temporally extended and dynamically varying nature of tonal contours, we hypothesize that VAS will reveal shallower slopes and systematic within-category distinctions ($\Delta > 0$), reflecting listeners' sensitivity to fine-grained F0 variation even within nominal tone categories. Second, do slope and response variability reflect dissociable perceptual processes, with only response variability serving as a stable individual difference measure? We predict that slope and response variability will show weak or no correlation in VAS, and that only response variability will demonstrate stability across task formats (VAS vs. 2AFC). Such a pattern would support the interpretation that slope is largely task- and strategy-dependent, while response variability reflects a more fundamental, trait-like dimension of categorization ability.

By directly comparing VAS and 2AFC paradigms in the domain of lexical tone, our study aims to clarify whether tonal perception is better characterized by gradient or categorical principles, evaluate whether GP frameworks, which are well-supported for segmental speech sounds, extend meaningfully to suprasegmental features, and establish VAS as a sensitive tool for investigating tone and other temporally distributed prosodic phenomena. These findings will contribute empirical evidence toward a unified theory of speech categorization that encompasses both segmental and suprasegmental domains.

Method

Participants

Eighty-four native speakers of Mandarin Chinese (42 males and 42 females; $M_{age} \pm SD = 18.7 \pm 0.86$ years, range: 17-21) participated in the study. All participants met the following inclusion criteria: (1) native Mandarin proficiency, defined as growing up in a Mandarin-dominant environment and reporting Mandarin as their first and primary language since childhood; (2) normal hearing, with self-reported absence of hearing impairment and no reported use of hearing aids or history of auditory disorders; (3) neurological and cognitive health, with no self-reported history of speech, language, learning, or neurological disorders such as dyslexia, autism spectrum disorder, ADHD, or

traumatic brain injury; and (4) minimal formal musical training, defined as less than two years of instrumental or vocal training before age 12, to reduce potential confounding effects of musical expertise on tone perception. Written informed consent was obtained from all participants prior to testing. Participants received monetary compensation for their time. The study protocol was approved by the Institutional Ethics Committee of Xi'an Jiaotong University (Approval No. XJTU-SFS-RECA-1-004).

A priori power analysis was conducted with G-Power 3.1 (Faul et al., 2007) for correlational analysis (two-tailed, $\alpha = .05$). With this sample size $N = 84$, the study achieves 80% power to detect medium-sized correlations of $|\rho| \geq .30$ and 90% power for $|\rho| \geq .35$, but has limited sensitivity to small effects (e.g., $|\rho| \approx .07$), which we acknowledge as a constraint in interpreting null or marginal findings.

Stimuli

The stimuli consisted of an 11-step synthetic continuum spanning Mandarin Lexical Tone 1 (T1) and Tone 2 (T2), embedded in the syllable */ba/*. The endpoint tokens were originally recorded from a native Mandarin male speaker in a sound-attenuated booth at a sampling rate of 44.1 kHz. Using Praat (Boersma & Weenink, 2024), duration was uniformly normalized to 400 ms, and root-mean-square (RMS) intensity was scaled to 75 dB SPL to minimize durational and amplitude cues.

The continuum was generated using the STRAIGHT software (Kawahara et al., 1999) implemented in MATLAB. F0 contours of the T1 and T2 endpoints were linearly interpolated in semitone space to create 10 equally spaced intermediate steps, resulting in a total of 11 stimuli (Figure 1A). All non-F0 acoustic parameters, including spectral envelope, duration, and intensity, were held constant across steps to ensure that perceptual variation was driven exclusively by changes in pitch contour.

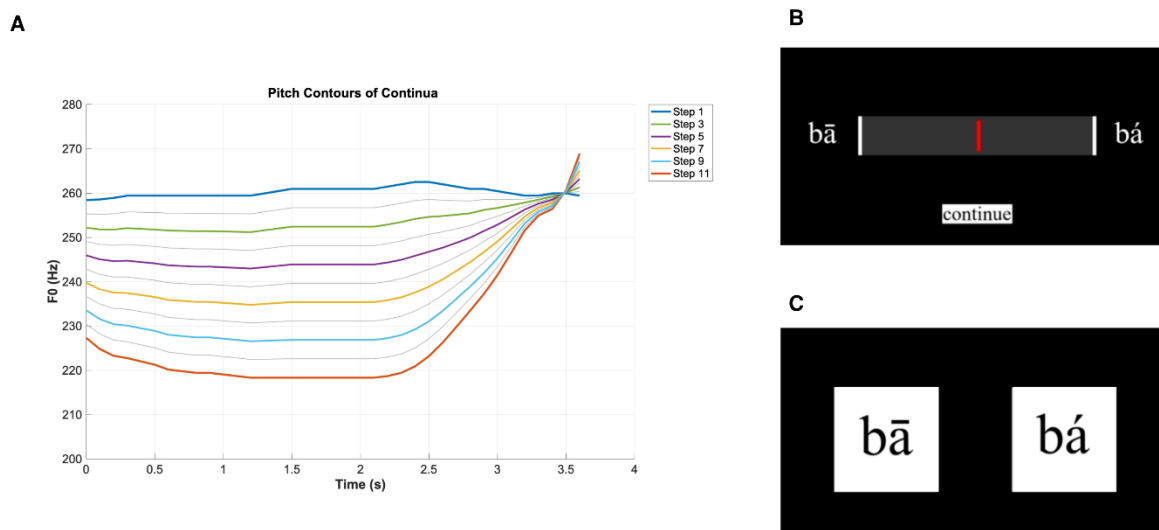


Figure 1. Overview of Experimental Stimuli and Task Interfaces. (A) Fundamental frequency (F0) contours for the 11-step lexical tone continuum, ranging from a high-level tone (T1) to a high-rising tone (T2). Six representative steps are highlighted. (B) The interface for the Visual Analog Scale (VAS) task. (C) The interface for the two-alternative forced-choice (2AFC) task.

Procedure

Testing was conducted individually in a quiet, sound-attenuated classroom equipped for phonetic experimentation. Ambient noise levels were monitored and maintained below 35 dB(A) to ensure optimal auditory presentation. Participants were seated at a computer workstation with a 24-

inch monitor and responded using a standard mouse. All auditory stimuli were delivered binaurally via high-fidelity, closed-back headphones at a calibrated level of 75 dB SPL.

Each participant completed two perceptual categorization tasks in a fixed order: the VAS task followed by the 2AFC task. This order was intentionally selected to prevent carryover effects, specifically to avoid the possibility that binary categorical decisions in the 2AFC task might bias or constrain the fine-grained, continuous judgments required in the VAS task (Kapnoula et al., 2017). Pilot testing confirmed that reversing the order led to reduced variance in VAS responses, supporting this procedural decision.

Prior to each task, participants completed a brief practice block (approximately 2 minutes) consisting of 10 trials using endpoint stimuli (Step 1 and Step 11) to ensure familiarity with the interface, response modality, and category labels. Practice trials were not included in the analysis.

In the VAS task (Figure 1B), participants rated each stimulus by positioning a slider along a 100-point horizontal continuum displayed on-screen. The left anchor was labeled “bā” (in *pinyin*, T1) and the right anchor “bá” (in *pinyin*, T2). Responses were recorded as continuous values ranging from 0 (prototypical “bā”) to 100 (prototypical “bá”), with sub-pixel resolution preserved for analysis. With each stimulus presented 10 times in a random order, yielding 110 trials total. The VAS task lasted approximately 10-12 minutes.

In the 2AFC task (Figure 1C), participants identified each stimulus as either “bā” or “bá” by clicking one of two labeled pictures. Key assignments were counterbalanced across participants to control for motor bias. As in the VAS task, each of the 11 continuum steps was presented 10 times in randomized order, yielding 110 trials total. A 500-ms inter-trial interval followed each response, and no feedback was provided. The task duration was approximately 8–10 minutes.

In both tasks, stimuli were triggered automatically upon trial onset. A fixation cross appeared 500 ms before stimulus onset and remained on-screen until response. Participants were instructed to respond only after the sound ended. Response times were logged but not used as exclusion criteria, as the tasks emphasized accuracy and perceptual judgment over speed. Total testing time, including instructions, practice, and both tasks, was approximately 25 minutes per participant.

Data Analysis

Model Fitting

We applied a four-parameter logistic function separately to each participant’s data for both VAS and 2AFC tasks using the `nlsLM()` function from the `minpack.lm` package (Elzhov et al., 2023) in R (R Core Team, 2024). The model equation is defined as:

$$f(x) = \frac{a_2 - a_1}{1 + e^{-s(x-c)}} + a_1,$$

where $f(x)$ denotes the predicted response at stimulus step x ; a_1 and a_2 are the lower and upper asymptotes, respectively; c represents the boundary parameter (i.e., the estimated category crossover point); and s is the slope parameter that determines the steepness of the categorization function.

We used the Levenberg-Marquardt algorithm over alternatives such as polynomial regression (Kong & Edwards, 2016) or Bayesian hierarchical modeling (Kim et al., 2025a; Sorensen et al., 2024) for three main reasons: (1) comparability, as both binary (2AFC) and continuous (VAS) data were modeled identically, allowing unified slope interpretation (Kapnoula et al., 2017); (2) numerical stability, as the `nlsLM()` function provides robust convergence for bounded, nonlinear functions; and (3) transparency, as individual-level fits avoid assumptions of hierarchical structure or strong priors. Model robustness was evaluated using residual plots and non-parametric bootstrapping of slope estimates.

Metrics

From each fitted curve, we extracted three indices of perceptual categorization. The slope (s) quantifies the steepness of the categorization function and is widely interpreted as an index of perceptual sharpness. Higher slope values indicate steeper, more categorical boundaries, whereas lower slopes suggest greater gradiency and sensitivity to within-category variation. The boundary

(c) corresponds to the estimated category crossover point, that is, the stimulus step at which responses are equally likely to fall into either category. This measure was used to define category regions in subsequent analyses of within-category sensitivity.

To quantify response variability, we followed prior work (Apfelbaum et al., 2022; Kapnoula et al., 2017) and calculated the dispersion of responses across repeated presentations of the same stimulus. For each participant, we computed the squared deviation of each trial-level response from the mean response to that specific stimulus step:

$$\text{Response Variability} = \sum_{i=1}^n (x_i - \bar{x}_{\text{step}(i)})^2,$$

where x_i is the response on trial i , and $\bar{x}_{\text{step}(i)}$ is the participant's mean response to the stimulus step presented on that trial. Higher values indicate greater trial-to-trial dispersion, while lower values indicate greater stability. For VAS data, values were scaled to the 0–1 range to ensure comparability with binary 2AFC responses.

Statistical Analysis

Descriptive statistics were obtained for slope, boundary, and response variability indices in each task. Normality of paired differences was assessed with Shapiro-Wilk tests. Because slope and response variability deviated from normality, Wilcoxon signed-rank tests were used for task comparisons. For boundary, the distribution of paired differences did not deviate from normality, and paired-samples t -tests were therefore conducted. Relationships between slope and response variability were assessed with Spearman rank-order correlations conducted within tasks, across tasks, and across measures.

In addition, a within-category sensitivity index (Δ) was derived from VAS ratings to directly test gradient sensitivity. The repeated presentation of each stimulus enabled reliable estimation of within-category response patterns. Δ was calculated as the average absolute difference between adjacent continuum steps within each tonal category region, excluding the step immediately adjacent to the category boundary. Larger Δ values reflected greater sensitivity to fine-grained within-category variation. One-sample t -tests against zero were used to determine whether Δ values reliably exceeded chance levels.

To capture individual variability in categorization profiles, we generated quadrant plots by median-splitting participants on slope and response variability within each task. These plots illustrated the joint distribution of gradiency and response variability, highlighting distinct response profiles across listeners.

Results

Model Performance and Overall Task Comparisons

The four-parameter logistic model successfully converged for all 84 participants in both tasks, yielding a 100% convergence rate. Figure 2 shows the fitted curves for each task. In the VAS task, slopes ranged from 0.15 to 7.02 ($M = 1.02$, $SD = 0.98$), and response variability ranged from 0.45 to 4.96 ($M = 1.90$, $SD = 0.86$). In the 2AFC task, slopes ranged from 0.57 to 10.00 ($M = 3.36$, $SD = 2.86$), and response variability ranged from 1.60 to 18.90 ($M = 5.51$, $SD = 2.35$). Category boundaries were broadly similar across tasks, with means of 5.72 ($SD = 0.79$) for VAS and 5.52 ($SD = 0.70$) for 2AFC.

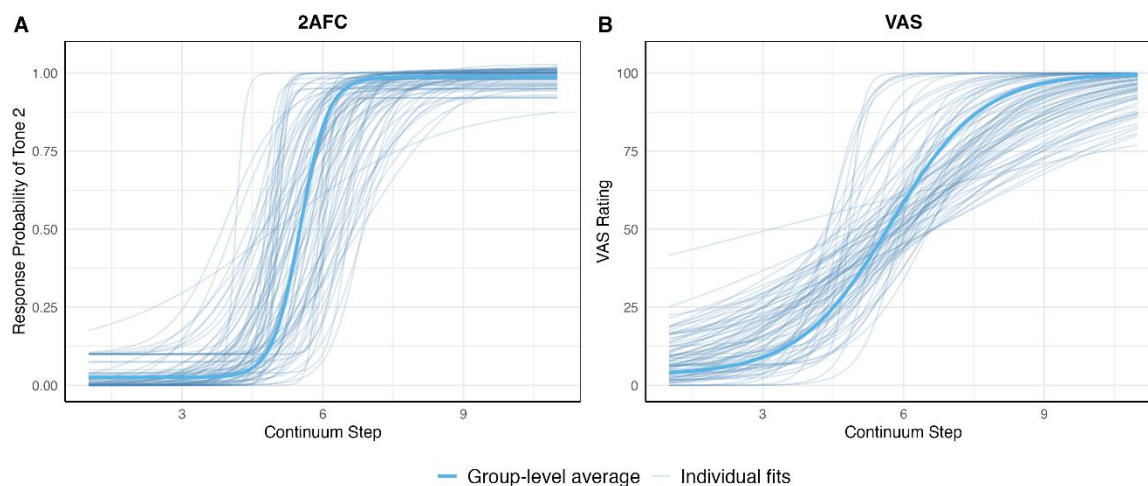


Figure 2. Psychometric Functions for Tone Identification. Four-parameter logistic functions were fitted to individual participant data (light blue lines) and averaged for the group (dark blue line) for the (A) two-alternative forced-choice (2AFC) task and (B) Visual Analog Scale (VAS) task.

Direct comparison revealed significant differences between tasks. Paired Wilcoxon signed-rank tests showed that slopes were significantly shallower in the VAS task than in the 2AFC task ($V = 233.0$, $p < .001$, $r = 0.76$; Figure 3A), and response variability was significantly lower in VAS than in 2AFC ($V = 6.0$, $p < .001$, $r = 0.87$; Figure 3B). Notably, category boundary estimates also differed significantly, with VAS boundaries shifted slightly rightward relative to 2AFC ($t(83) = 2.98$, $p = .0038$, $r = .32$; Figure 3C).

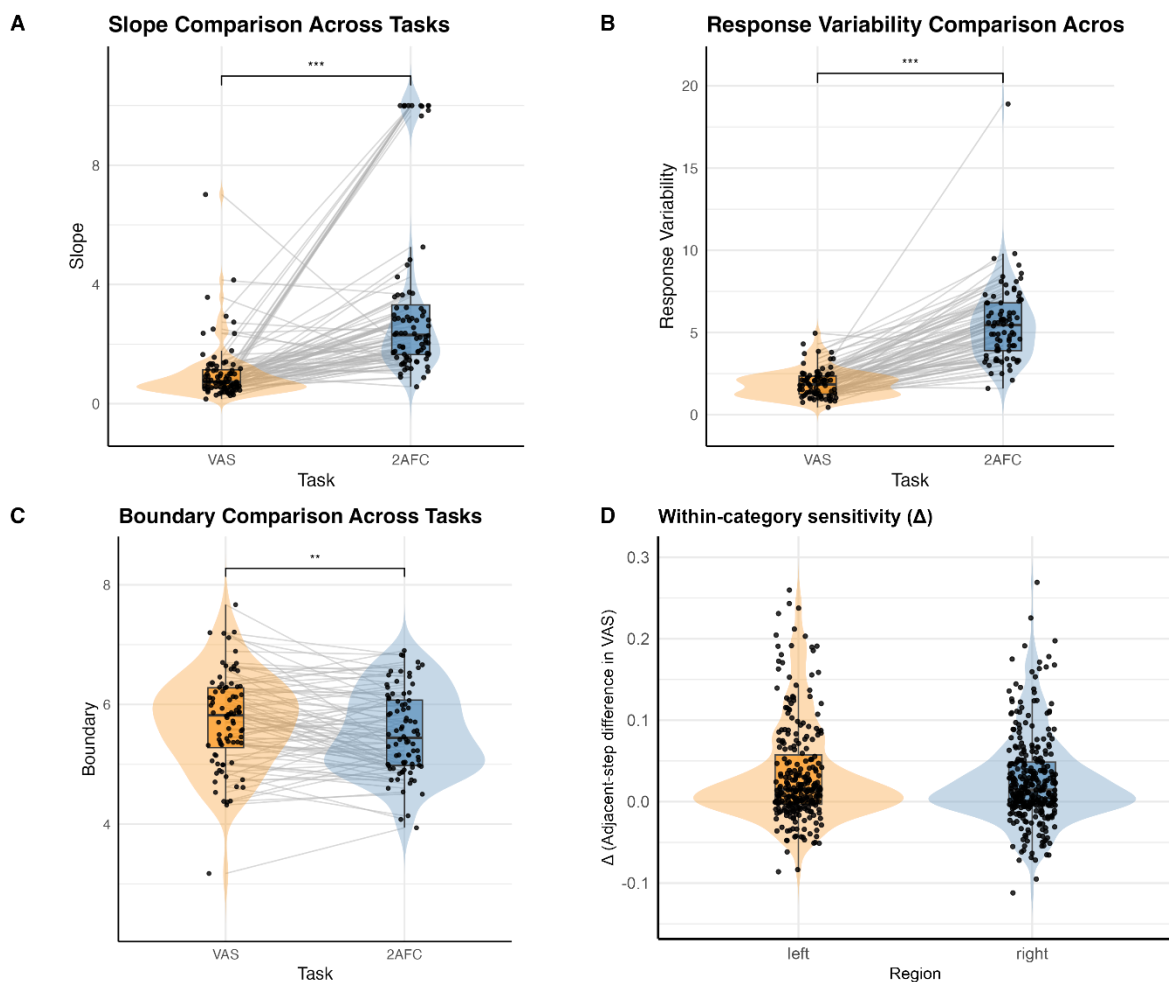


Figure 3. Comparison of Psychometric Parameters Between Tasks. Distributions of (A) slope, (B) response variability (where lower values indicate higher consistency), and (C) category boundary for the VAS and 2AFC tasks. (D) Within-category sensitivity (Δ) for the left and right category regions in the VAS task. Asterisks denote statistical significance from paired-samples tests (** $p < .001$; ** $p < .01$; * $p < .05$).

Within-Category Sensitivity

To provide a direct test of whether Mandarin tone perception exhibits within-category sensitivity, we computed the within-category sensitivity index (Δ) for each participant from VAS rating. The continuum was divided into left- and right-category regions based on each individual's boundary, excluding the step immediately adjacent to the boundary. Within each categorical region, Δ was defined as the average absolute difference in VAS ratings between adjacent steps. Larger Δ values reflect greater sensitivity to within-category acoustic variation.

One-sample *t*-tests confirmed that Δ was significantly greater than zero in both regions. In the left-category region, listeners showed a mean Δ of 0.034, $t(270) = 8.89$, $p < .001$. In the right-category region, the mean Δ was 0.026, $t(316) = 8.38$, $p < .001$. As shown in Figure 3D, the distributions of Δ values for both regions were clearly shifted above zero. These findings provide direct evidence that participants are sensitive to fine-grained F0 variation even when stimuli fall within the same tonal category, demonstrating robust within-category sensitivity in tone categorization.

Cross-Measure Correlations

We next examined the relationship between slope and response variability within and across tasks (Figure 4). In the VAS task, slope and variability showed a weak positive correlation ($\rho = .27$, $p < .05$), whereas in the 2AFC task they were strongly and negatively correlated ($\rho = -.67$, $p < .001$). Thus, steeper identification functions were associated with lower variability in the forced-choice paradigm, but this relationship was weak and reversed in the continuous VAS paradigm.

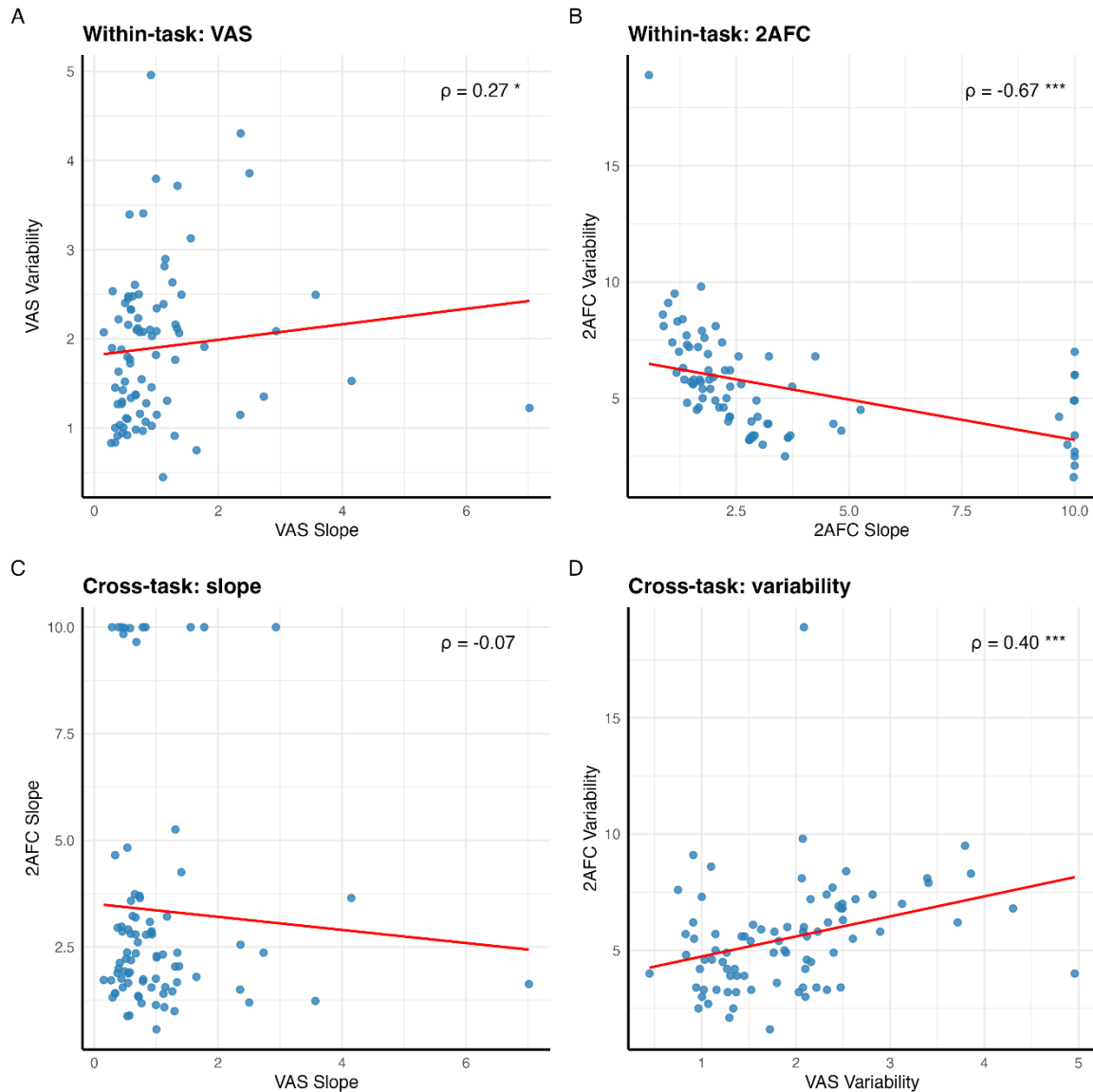


Figure 4. Correlations of Psychometric Parameters Within and Across Tasks. Scatterplots showing relationships between: (A) slope and response variability in the VAS task; (B) slope and response variability in the 2AFC task; (C) slopes across the two tasks; and (D) response variabilities across the two tasks. Lower response variability values indicate higher consistency. Asterisks denote significant Spearman correlations ($^{***}p < .001$; $^{**}p < .01$; $^*p < .05$).

Across tasks, slopes in VAS and 2AFC were not significantly correlated ($\rho = -.07$, $p = .54$), suggesting that categorization sharpness is largely task-dependent. In contrast, response variability showed a moderate positive correlation across tasks ($\rho = .40$, $p < .001$), indicating relative stability in response variability across paradigms. Moreover, cross-measure associations were negligible (Figure 5): VAS slope did not significantly correlate with 2AFC variability ($\rho = .16$, $p = .15$), and 2AFC slope was not reliably related to VAS variability ($\rho = -.10$, $p = .37$).

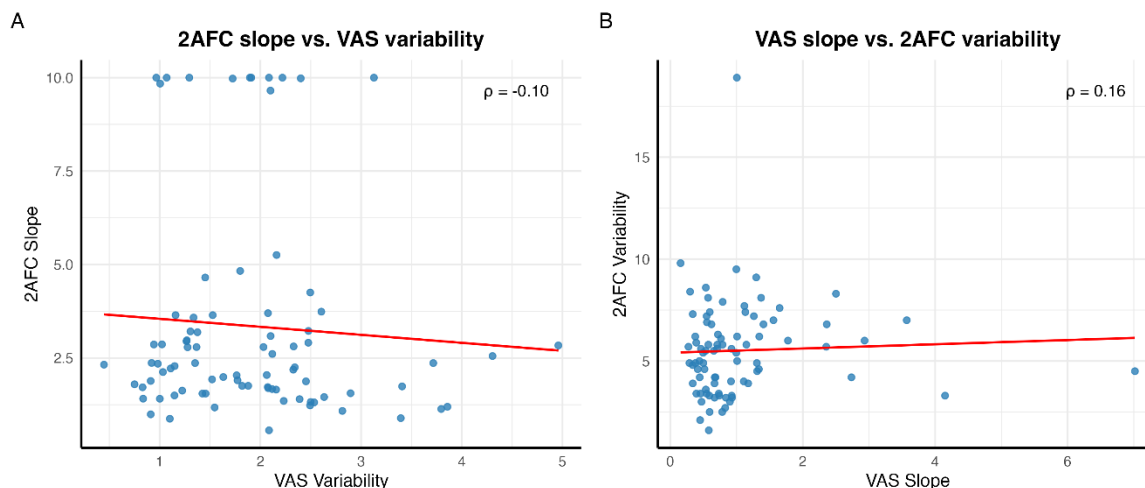


Figure 5. Cross-Task Correlations Between Slope and Response Variability. Scatterplots showing the relationship between (A) slope in the 2AFC task and response variability in the VAS task, and (B) slope in the VAS task and response variability in the 2AFC task. Lower response variability values indicate higher consistency.

Individual Differences in Slope Steepness and Response Variability

To illustrate individual patterns, we generated within-task quadrant plots by median-splitting participants on slope and response variability (Figure 6). These plots reveal both substantial heterogeneity among listeners and clear task-specific structure. Participants were distributed across all four slope–variability quadrants in both tasks, indicating that no single profile predominates. This confirms that gradient perception is present but varies considerably across individuals.

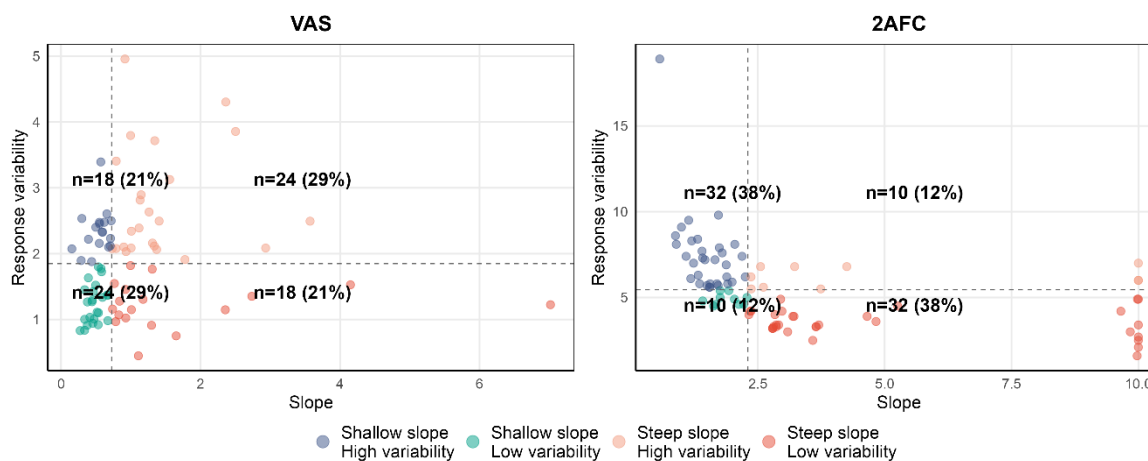


Figure 6. Classification of Listeners Based on Perceptual Parameters. Quadrant plots showing the distribution of individual listeners based on their slope and response variability values for the (A) VAS task and (B) 2AFC task. Lower response variability values indicate higher consistency.

Task-specific patterns were also evident. In the VAS task, listeners were relatively evenly distributed across quadrants (Figure 6A), consistent with the weak association between slope and response variability and suggesting that categorization sharpness and response variability operate more independently under continuous rating. In contrast, the 2AFC task showed a more polarized distribution, with participants clustering in the steep-slope / low-variability and shallow-slope / high-variability quadrants (Figure 6B). This pattern mirrors the strong negative slope–variability correlation observed in this paradigm and reflects the mathematical constraint inherent in binary response formats. These findings indicate that categorization sharpness and response variability are separable perceptual dimensions and that their interrelationship is shaped by task demands.

Discussion

The present study provides compelling evidence that Mandarin tone perception is fundamentally gradient rather than strictly categorical, and that this gradiency becomes fully visible only when assessed with paradigms sensitive to within-category distinctions. Through two complementary findings, the reliable within-category sensitivity in tone perception and the dissociation between perceptual gradiency and response variability, we reveal that tone categorization is shaped by both stable listener traits and flexible task-dependent processes. These results challenge the traditional binary view of tone perception and align with an emerging consensus that speech categories, even those as seemingly discrete as lexical tones, preserve rich acoustic detail that listeners can access under appropriate task conditions.

Gradient Encoding Within Tone Categories

Our first major finding, that Mandarin listeners exhibit reliable sensitivity to fine-grained F0 differences within tonal categories, directly challenges the long-standing assumption that tone perception for contour tones is strictly categorical. The continuous response format of the VAS paradigm allowed participants to express subtle perceptual distinctions that binary paradigms inherently suppress (Repp, 1984; Schouten et al., 2003). The resulting shallower slopes and robust within-category discrimination ($\Delta > 0$) confirm that listeners do not simply collapse continuous pitch variation into discrete categories; rather, they maintain access to gradient acoustic information even after assigning a categorical label.

This pattern mirrors findings from segmental perception, where rating and continuous response tasks have revealed within-category sensitivity for consonants and vowels (Fuhrmeister et al., 2023; Honda et al., 2024b; Kapnoula et al., 2021; Kong & Edwards, 2016; Massaro & Cohen, 1983; McMurray et al., 2009; Myers et al., 2024; Toscano et al., 2010). Critically, our results extend the empirical case for gradient encoding to the suprasegmental domain, demonstrating that gradient perception appears to be a domain-general principle operating across both segmental and tonal contrasts (McMurray, 2022; Toscano et al., 2010). They converge with theoretical models that conceptualize categorization as graded activation over competing category representations rather than discrete, all-or-none decisions (Goldinger, 1998.; Kleinschmidt & Jaeger, 2015; McClelland & Elman, 1986).

However, quadrant analyses revealed substantial individual heterogeneity in how gradient perception was expressed. Listeners were distributed across all four slope \times variability profiles, with no single pattern predominating. Some participants combined shallow slopes with low variability, consistent with genuinely gradient encoding, whereas others paired shallow slopes with high variability, reflecting unstable categorical decisions. This heterogeneity reflects the inherent ambiguity in interpreting slope and underscores that while gradient perception may be a general property of speech processing, its behavioral manifestation varies across individuals and depends critically on task demands and response strategies (Kim et al., 2025a; 2025b). This heterogeneity is not measurement noise but meaningful variation that reflects individual differences in perceptual acuity, decisional strategies, or linguistic experience. Recognizing and modeling this variation is essential for developing a complete theory of speech perception that accounts for the full spectrum of human perceptual behavior.

Gradiency and Response Variability as Independent Perceptual Dimensions

Perhaps the most theoretically significant contribution of this study lies in demonstrating that gradiency (slope steepness) and response variability are dissociable perceptual processes. In the VAS task, these two dimensions were only weakly correlated ($\rho = .27$), with listeners distributed across all four quadrants of the slope \times variability space. This dissociation reveals that shallow slopes do not necessarily reflect perceptual noise; some listeners exhibited shallow slopes and low variability, a profile indicative of genuinely gradient encoding, while others showed shallow slopes paired with high variability, suggesting unstable categorical decisions.

The stark contrast between VAS and 2AFC findings underscores a critical methodological insight: task format fundamentally shapes observed perceptual patterns. In 2AFC, slope and response variability were tightly coupled ($\rho = -.67$), producing a strong negative correlation that artificially conflates gradiency with decisional noise. This coupling effect helps explain longstanding inconsistencies in the tone perception literature regarding whether tone perception is categorical or gradient (Abramson, 1975; Francis et al., 2003; Hallé et al., 2004; Wang et al., 2003; Xu et al., 2006). Listeners who produced steep, categorical-looking functions in 2AFC often demonstrated clear within-category sensitivity in VAS, suggesting that apparent conflicts in previous research may reflect task-dependent manifestations of underlying graded representations rather than genuine theoretical disagreements.

Across tasks, slopes were not significantly correlated ($\rho = -.07$), whereas response variability showed moderate stability ($\rho = .40$), indicating that response variability captures a more stable individual difference dimension. This finding aligns with recent work emphasizing response variability as the more stable and predictive perceptual index (Fuhrmeister et al., 2023; Honda et al., 2024a; Kim et al., 2025b; Myers et al., 2024) and carries important clinical implications. Individual differences in perceptual and neural consistency have been implicated in language disorders, dyslexia, and auditory processing deficits (e.g., Centanni et al., 2018; Hornickel & Kraus, 2013; Skoe et al., 2015). In such populations, response variability may provide a more sensitive and reliable diagnostic marker than slope for characterizing perceptual stability, particularly when assessed using continuous response paradigms that avoid the artificial constraints of binary forced-choice tasks.

Cross-measure analyses further confirmed the independence of gradiency and variability: VAS slope did not significantly correlate with 2AFC response variability, and 2AFC slope was not significantly related to VAS response variability. This finding contrasts with some segmental studies demonstrating significant correlations between 2AFC slopes and VAS response variability (Honda et al., 2024a; Kapnoula et al., 2017), suggesting potential differences between segmental and suprasegmental processing. Segmental categories often rely on relatively discrete, temporally localized cues such as voice onset time, which may naturally couple categorization steepness with decisional stability. Lexical tones, by contrast, are defined by dynamic pitch contours that unfold over hundreds of milliseconds. This temporal complexity may decouple the mechanisms governing slope from those governing response variability. Future research should explore whether this decoupling generalizes to other suprasegmental phenomena, such as stress and intonation, where temporal integration is similarly critical.

Methodological Advantages of Continuous Response Paradigms

Beyond its theoretical contributions, our results underscore the methodological value of the VAS paradigm for speech perception research. Unlike forced-choice tasks that collapse continuous perceptual space into discrete bins, VAS provides a multidimensional window into perception, yielding independent estimates of gradiency and response variability (Apfelbaum et al., 2022; Massaro & Cohen, 1983). This enabled us to disentangle genuinely gradient encoding from unstable categorization, revealing patterns such as within-category sensitivity and heterogeneous individual profiles that would be conflated in 2AFC tasks.

Empirically, our results showed that variability was significantly lower in VAS than in 2AFC, indicating greater response consistency under the continuous paradigm. This contrast reflects a critical methodological artifact: in 2AFC, responses at boundary-adjacent steps fluctuate between two discrete categories, producing maximum Bernoulli variance. Even if listeners hold stable percepts, the binary format inflates trial-to-trial variability. VAS allows listeners to register partial category membership, leading responses to cluster more tightly around the perceptual mean and yielding higher apparent consistency. Thus, VAS provides a more faithful index of perceptual stability (Apfelbaum et al., 2022).

The continuous response space of VAS is particularly well-suited for phenomena defined by dynamic, temporally extended cues such as tone, stress, and intonation where binary responses may

impose artificial discreteness on inherently gradient signals. While segmental studies have established VAS as a powerful tool for capturing individual differences (Honda et al., 2024a, 2024b; Kapnoula et al., 2017, 2021; Kim et al., 2020; Kong & Edwards, 2016; McMurray et al., 2014; Myers et al., 2024), its application to suprasegmental domains remains limited. Applying VAS to lexical tone research could clarify how slope steepness and response variability interact under time-varying F0 cues, test whether slope-variability dissociations generalize beyond segmental contrasts, and provide insights into cross-linguistic, developmental, and clinical variability in tone processing. VAS should be regarded as an essential complement to traditional methods, offering the potential to resolve longstanding ambiguities in speech perception research.

Limitations and Future Directions

This study has several limitations that suggest directions for future research. First, we examined only one Mandarin tone continuum (Tone 1–Tone 2), leaving open whether the findings generalize to other tone contrasts (e.g., Tone 2–Tone 3) or to tonal languages with different phonological systems (e.g., Cantonese, Thai). Second, while VAS provides rich perceptual data, scale interpretation may vary across listeners, and slope and response variability indices do not capture all aspects of gradient perception. Complementary measures such as reaction times, eye-tracking, and neural indices (EEG, fMRI) would offer converging evidence (Kapnoula et al., 2021; Myers et al., 2024). Third, our sample consisted solely of native Mandarin speakers, so it remains unclear whether the observed patterns extend to non-native listeners, dialectal variants, or clinical populations.

These limitations point to several productive research directions. First, extending this work to clinical populations such as autism spectrum disorder (ASD) and developmental dyslexia could reveal dissociations between sensory acuity and decisional stability. Individuals with ASD often show enhanced low-level auditory discrimination but reduced categorical perception in speech (Bonnell et al., 2010; O'Connor, 2012; Serniclaes et al., 2004; Wang et al., 2017; Zhang et al., 2012; Ziegler et al., 2005). VAS paradigms could test whether this pattern extends to lexical tones, perhaps revealing heightened within-category sensitivity alongside reduced consistency due to attentional or decisional differences. Similarly, in dyslexia where segmental categorical perception is often impaired (Serniclaes et al., 2005; Ziegler et al., 2005), VAS may uncover whether suprasegmental gradient encoding is preserved, offering potential compensatory pathways for phonological learning.

Second, systematic comparisons across tone contrasts, tonal languages, and linguistic domains are needed to determine whether slope-variability dissociations are domain-general or tone-specific. Comparing gradient perception across different Mandarin tone pairs, other tonal languages, and segmental versus suprasegmental contrasts within the same participants would clarify the scope and limits of these findings (e.g., Honda et al., 2024a; Kapnoula et al., 2017; Kong & Edwards, 2016).

Third, individual differences merit deeper investigation. Factors including musical expertise, cognitive control, and working memory should be systematically examined to understand sources of variability in gradient perception. Multimodal methods such as eye-tracking, EEG, and fMRI can uncover the neurocognitive mechanisms underlying gradient encoding and response consistency.

Fourth, developmental and longitudinal studies using VAS could track how slope-variability profiles emerge in children acquiring tone languages and in second language learners, revealing whether gradient perception changes with linguistic experience. Finally, computational modeling approaches such as drift-diffusion models and neural networks could formalize the dynamics of gradient perception, informing both theoretical models of speech perception and practical applications in AI systems for tonal language processing.

Conclusions

This study provides the initial evidence that Mandarin tone perception is not exclusively categorical but preserves graded sensitivity to within-category F0 variation, a pattern revealed only through continuous response paradigms like VAS. Critically, we demonstrate that gradiency (slope steepness) and response variability reflect dissociable perceptual processes: one shaped by task

demands, the other reflecting stable individual traits. These findings challenge binary models of speech categorization and highlight VAS as a powerful tool for uncovering hidden dimensions of perceptual structure. By moving beyond forced-choice paradigms and embracing the continuous, multidimensional nature of human perception, this work lays the foundation for new research exploring gradient perception across populations, languages, and modalities. Such research holds particular promise for understanding and supporting individuals with atypical speech processing, ultimately enabling more accurate, inclusive, and dynamic models of how listeners transform acoustic signals into linguistic meaning.

Acknowledgments

The research was supported by grants from the National Social Science Fund of China (22BYY160, 24CYY096), Xi'an Jiaotong University Undergraduate Teaching Reform Research Project (2424Z), and the China Postdoctoral Science Foundation (2025T180911, 2023M742804).

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of Interest Statement: The authors declare no conflicts of interest.

References

1. Abramson, A. S. (1975). *Thai tones as a reference system* (Haskins Laboratories: Status Report on Speech Research, pp. 127–136). Haskins Laboratories.
2. Abramson, A. S. (1977). Noncategorical perception of tone categories in Thai. *The Journal of the Acoustical Society of America*, 61(S1), S66–S66. <https://doi.org/10.1121/1.2015837>
3. Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52(3), 163–187. [https://doi.org/10.1016/0010-0277\(94\)90042-6](https://doi.org/10.1016/0010-0277(94)90042-6)
4. 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), 3728–3745. <https://doi.org/10.1121/10.0015201>
5. Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* (Version Version 6.4.17) [Computer software]. University of Amsterdam. <https://www.praat.org/>
6. Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., Burack, J. A., Mottron, L., Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., Burack, J. A., & Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia*, 48(9), 2465–2475. <https://doi.org/10.1016/j.neuropsychologia.2010.04.020>
7. Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *The Journal of the Acoustical Society of America*, 62(4), 961–970. <https://doi.org/10.1121/1.381590>
8. Casillas, J. V. (2020). The Longitudinal Development of Fine-Phonetic Detail: Stop Production in a Domestic Immersion Program. *Language Learning*, 70(3), 768–806. <https://doi.org/10.1111/lang.12392>
9. 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>
10. Coady, J. A., Kluender, K. R., & Evans, J. L. (2005). Categorical Perception of Speech by Children With Specific Language Impairments. *Journal of Speech, Language, and Hearing Research*, 48(4), 944–959. [https://doi.org/10.1044/1092-4388\(2005\)065](https://doi.org/10.1044/1092-4388(2005)065)
11. Eimas, P. D., Miller, J. L., & Jusczyk, P. W. (1987). On infant speech perception and the acquisition of language. *Categorical Perception: The Groundwork of Cognition.*, 161–195.

12. Elzhov, T. V., Mullen, K. M., Spiess, A., & Bolker, B. (2023). *minpack.lm: R interface to the levenberg-marquardt nonlinear least-squares algorithm* [Computer software]. CRAN. <https://cran.r-project.org/package=minpack.lm>
13. Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
14. Feng, Y., & Peng, G. (2023). Development of categorical speech perception in mandarin-speaking children and adolescents. *Child Development*, 94(1), 28–43. <https://doi.org/10.1111/cdev.13837>
15. Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, 27(4), 343–350. <https://doi.org/10.3758/BF03206123>
16. Francis, A. L., Ciocca, V., & Chit Ng, B. K. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, 65(7), 1029–1044. <https://doi.org/10.3758/BF03194832>
17. Frencq-Mestre, C., Meunier, C., Espesser, R., Daffner, K., & Holcomb, P. (2005). Perceiving Nonnative Vowels. *Journal of Speech, Language, and Hearing Research*, 48(6), 1496–1510. [https://doi.org/10.1044/1092-4388\(2005/104\)](https://doi.org/10.1044/1092-4388(2005/104))
18. Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The Identification and Discrimination of Synthetic Vowels. *Language and Speech*, 5(4), 171–189. <https://doi.org/10.1177/002383096200500401>
19. Fuhrmeister, P., Myers, E. B., 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>
20. 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>
21. Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11(2), 149–175. [https://doi.org/10.1016/S0095-4470\(19\)30813-7](https://doi.org/10.1016/S0095-4470(19)30813-7)
22. Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3), 363–376. <https://doi.org/10.3758/BF03194885>
23. Goldinger, S. D. (n.d.). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. <https://doi.org/10.1037/0033-295X.105.2.251>
24. Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of mandarin Chinese tones by mandarin chinese vs. French listeners. *Journal of Phonetics*, 32(3), 395–421. [https://doi.org/10.1016/S0095-4470\(03\)00016-0](https://doi.org/10.1016/S0095-4470(03)00016-0)
25. Hary, J. M., & Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Perception & Psychophysics*, 32(5), 409–418. <https://doi.org/10.3758/BF03202770>
26. Healy, A. F., & Repp, B. H. (1982). Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 8(1), 68–80. PubMed. <https://doi.org/10.1037//0096-1523.8.1.68>
27. Honda, C. T., Clayards, M., & Baum, S. R. (2024a). Exploring individual differences in native phonetic perception and their link to nonnative phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 50(4), 370–394. <https://doi.org/10.1037/xhp0001191>
28. Honda, C. T., Clayards, M., & Baum, S. R. (2024b). Individual differences in the consistency of neural and behavioural responses to speech sounds. *Brain Research*, 1845, 149208. <https://doi.org/10.1016/j.brainres.2024.149208>
29. Hornickel, J., & Kraus, N. (2013). Unstable Representation of Sound: A Biological Marker of Dyslexia. *The Journal of Neuroscience*, 33(8), 3500–3504. <https://doi.org/10.1523/JNEUROSCI.4205-12.2013>
30. Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language Deficits in Dyslexic Children: Speech Perception, Phonology, and Morphology. *Journal of Experimental Child Psychology*, 77(1), 30–60. <https://doi.org/10.1006/jecp.1999.2553>
31. Johnson, K., & Ralston, J. V. (1994). Automaticity in Speech Perception: Some Speech/Nonspeech Comparisons. *Phonetica*, 51(4), 195–209. <https://doi.org/10.1159/000261975>

32. 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>
33. 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>
34. Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, 27(3–4), 187–207. [https://doi.org/10.1016/S0167-6393\(98\)00085-5](https://doi.org/10.1016/S0167-6393(98)00085-5)
35. Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, 103(30), 11381–11386. <https://doi.org/10.1073/pnas.0604821103>
36. Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics*, 81, 100984. <https://doi.org/10.1016/j.wocn.2020.100984>
37. Kim, H., Klein-Packard, J., Sorensen, E., Oleson, J., Tomblin, B., & 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. <https://doi.org/10.1016/j.cognition.2025.106194>
38. Kim, H., McMurray, B., Sorensen, E., & Oleson, J. (2025). The consistency of categorization-consistency in speech perception. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-025-02700-x>
39. Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. <https://doi.org/10.1037/a0038695>
40. Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication*, 41(1), 59–69. [https://doi.org/10.1016/S0167-6393\(02\)00093-6](https://doi.org/10.1016/S0167-6393(02)00093-6)
41. 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>
42. Kuhl, P. K. (1987). The special-mechanisms debate in speech research: Categorization tests on animals and infants. *Categorical Perception: The Groundwork of Cognition.*, 355–386.
43. Larkey, L. S., Wald, J., & Strange, W. (1978). Perception of synthetic nasal consonants in initial and final syllable position. *Perception & Psychophysics*, 23(4), 299–312. <https://doi.org/10.3758/BF03199713>
44. Liberman, A., Harris, K. S., Eimas, P., Lisker, L., & Bastian, J. (1961). An Effect of Learning on Speech Perception: The Discrimination of Durations of Silence with and without Phonemic Significance. *Language and Speech*, 4(4), 175–195. <https://doi.org/10.1177/002383096100400401>
45. 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. <https://doi.org/10.1037/h0044417>
46. Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 61(5), 379–388. <https://doi.org/10.1037/h0049038>
47. Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a Tonal Language by Attending to the Tone: An In Vivo Experiment. *Language Learning*, 61(4), 1119–1141. <https://doi.org/10.1111/j.1467-9922.2011.00673.x>
48. Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are Speech Perception Deficits Associated with Developmental Dyslexia? *Journal of Experimental Child Psychology*, 66(2), 211–235. <https://doi.org/10.1006/jecp.1997.2383>
49. Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, 2(1), 15–35. [https://doi.org/10.1016/0167-6393\(83\)90061-4](https://doi.org/10.1016/0167-6393(83)90061-4)
50. McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)

51. McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, 152(6), 3819–3842. <https://doi.org/10.1121/10.0016614>
52. McMurray, B., Munson, C., & Tomblin, J. B. (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research*, 57(4), 1344–1362. https://doi.org/10.1044/2014_JSLHR-L-13-0196
53. 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](https://doi.org/10.1016/S0010-0277(02)00157-9)
54. 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>
55. Miller, J. L., & Eimas, P. D. (1977). Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *The Journal of the Acoustical Society of America*, 61(3), 835–845. <https://doi.org/10.1121/1.381373>
56. Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46(6), 505–512. <https://doi.org/10.3758/BF03208147>
57. Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural Attunement Processes in Infants during the Acquisition of a Language-Specific Phonemic Contrast. *The Journal of Neuroscience*, 27(2), 315–321. <https://doi.org/10.1523/JNEUROSCI.1984-06.2007>
58. Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18(5), 331–340. <https://doi.org/10.3758/BF03211209>
59. Morton, K. D., Torriane, P. A., Throckmorton, C. S., & Collins, L. M. (2008). Mandarin Chinese tone identification in cochlear implants: Predictions from acoustic models. *Hearing Research*, 244(1–2), 66–76. <https://doi.org/10.1016/j.heares.2008.07.008>
60. Myers, E., Phillips, M., & Skoe, E. (2024). Individual differences in the perception of phonetic category structure predict speech-in-noise performance. *The Journal of the Acoustical Society of America*, 156(3), 1707–1719. <https://doi.org/10.1121/10.0028583>
61. Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118(12), 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>
62. Newell, F. N., & Bühlhoff, H. H. (2002). Categorical perception of familiar objects. *Cognition*, 85(2), 113–143. [https://doi.org/10.1016/S0010-0277\(02\)00104-X](https://doi.org/10.1016/S0010-0277(02)00104-X)
63. O’Connor, K., & O’Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience & Biobehavioral Reviews*, 36(2), 836–854. <https://doi.org/10.1016/j.neubiorev.2011.11.008>
64. Oden, G. C., Massaro, D. W., Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, 85(3), 172–191. <https://doi.org/10.1037/0033-295X.85.3.172>
65. Ou, J., & Yu, A. C. L. (2022). Neural correlates of individual differences in speech categorisation: Evidence from subcortical, cortical, and behavioural measures. *Language, Cognition and Neuroscience*, 37(3), 269–284. <https://doi.org/10.1080/23273798.2021.1980594>
66. Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253–260. <https://doi.org/10.3758/BF03214136>
67. Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, 55(2), 328–333. <https://doi.org/10.1121/1.1914506>
68. R Core Team. (2024). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
69. Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In *Speech and Language* (Vol. 10, pp. 243–335). Elsevier. <https://linkinghub.elsevier.com/retrieve/pii/B9780126086102500121>
70. Robertson, E. K., Joanisse, M. F., Desroches, A. S., & Ng, S. (2009). Categorical speech perception deficits distinguish language and reading impairments in children. *Developmental Science*, 12(5), 753–767. <https://doi.org/10.1111/j.1467-7687.2009.00806.x>

71. 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>
72. Schouten, B., Gerrits, E., & Van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41(1), 71–80. [https://doi.org/10.1016/S0167-6393\(02\)00094-8](https://doi.org/10.1016/S0167-6393(02)00094-8)
73. Serniclaes, W., Heghe, S. V., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*, 87(4), 336–361. <https://doi.org/10.1016/j.jecp.2004.02.001>
74. Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Demonet, J.-F. (2001). Perceptual Discrimination of Speech Sounds in Developmental Dyslexia. *Journal of Speech, Language, and Hearing Research*, 44(2), 384–399. [https://doi.org/10.1044/1092-4388\(2001\)032](https://doi.org/10.1044/1092-4388(2001)032)
75. Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition*, 98(2), B35–B44. <https://doi.org/10.1016/j.cognition.2005.03.002>
76. Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and Plasticity of Auditory Brainstem Function Across the Lifespan. *Cerebral Cortex*, 25(6), 1415–1426. <https://doi.org/10.1093/cercor/bht311>
77. Sorensen, E., Oleson, J., Kutlu, E., & McMurray, B. (2024). A bayesian hierarchical model for the analysis of visual analogue scaling tasks. *Statistical Methods in Medical Research*, 33(6), 953–965. <https://doi.org/10.1177/09622802241242319>
78. Stevens, K. N., Libermann, A. M., Studdert-Kennedy, M., & Öhman, S. E. G. (1969). Crosslanguage Study of Vowel Perception. *Language and Speech*, 12(1), 1–23. <https://doi.org/10.1177/002383096901200101>
79. Stewart, M. E., Petrou, A. M., & Ota, M. (2018). Categorical speech perception in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders*, 48(1), 72–82. <https://doi.org/10.1007/s10803-017-3284-0>
80. Toscano, J. C., & McMurray, B. (2015). The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience*, 30(5), 529–543. <https://doi.org/10.1080/23273798.2014.946427>
81. 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>
82. van Hessen, A. J., & Schouten, M. E. H. (1999). Categorical perception as a function of stimulus quality. *Phonetica*, 56(1–2), 56–72. <https://doi.org/10.1159/000028441>
83. Wang, X., Wang, S., Fan, Y., Huang, D., & Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An Event-Related Potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports*, 7(1), 43254. <https://doi.org/10.1038/srep43254>
84. Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, 113(2), 1033–1043. <https://doi.org/10.1121/1.1531176>
85. Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, 41(1), 48–61. PubMed. <https://doi.org/10.1037/h0084150>
86. Whalen, D. H., & Xu, Y. (1992). Information for mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1), 25–47. <https://doi.org/10.1159/000261901>
87. Wiener, S., & Lee, C.-Y. (2020). Multi-Talker Speech Promotes Greater Knowledge-Based Spoken Mandarin Word Recognition in First and Second Language Listeners. *Frontiers in Psychology*, 11, 214. <https://doi.org/10.3389/fpsyg.2020.00214>
88. Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, 170(1), 223–231. <https://doi.org/10.1016/j.neuroscience.2010.06.077>

89. Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, 120(2), 1063–1074. <https://doi.org/10.1121/1.2213572>
90. Yip, M. (2002). *Tone*. Cambridge University Press; Cambridge Core. <https://www.cambridge.org/core/product/B511461D134BE12E23AA9E30CD311036>
91. Yu, A. C. L. (2010). Perceptual Compensation Is Correlated with Individuals' "Autistic" Traits: Implications for Models of Sound Change. *PLoS ONE*, 5(8), e11950. <https://doi.org/10.1371/journal.pone.0011950>
92. Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLoS ONE*, 8(9), e74746. <https://doi.org/10.1371/journal.pone.0074746>
93. Zhang, L., Xi, J., Wu, H., Shu, H., & Li, P. (2012). Electrophysiological evidence of categorical perception of Chinese lexical tones in attentive condition. *NeuroReport*, 23(1), 35–39. <https://doi.org/10.1097/WNR.0b013e32834e4842>
94. Zhang, X., Cheng, B., Qin, D., & Zhang, Y. (2021). Is talker variability a critical component of effective phonetic training for nonnative speech? *Journal of Phonetics*, 87, 101071. <https://doi.org/10.1016/j.wocn.2021.101071>
95. Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y., Li, P. (2012). Universality of categorical perception deficit in developmental dyslexia: An investigation of Mandarin Chinese tones. *Journal of Child Psychology and Psychiatry*, 53(8), 874–882. <https://doi.org/10.1111/j.1469-7610.2012.02528.x>
96. Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences*, 102(39), 14110–14115. <https://doi.org/10.1073/pnas.0504444102>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.