



## OPEN ACCESS

### EDITED BY

Matthew W. Crocker,  
Saarland University, Germany

### REVIEWED BY

Filiz Tezcan,  
Maastricht University, Netherlands  
Svetlana Vetchinnikova,  
University of Helsinki, Finland

### \*CORRESPONDENCE

Andrea Hofmann  
✉ andhofma@uni-potsdam.de

<sup>†</sup>These authors have contributed equally to this work and share senior authorship

RECEIVED 08 December 2025

REVISED 17 January 2026

ACCEPTED 23 January 2026

PUBLISHED 02 March 2026

### CITATION

Hofmann A, Tuomainen O, Hanne S, Verissimo J and Wartenburger I (2026) Auditory-perceptual acuity impacts prosodic boundary prediction in a gating task. *Front. Lang. Sci.* 5:1763160. doi: 10.3389/flang.2026.1763160

### COPYRIGHT

© 2026 Hofmann, Tuomainen, Hanne, Verissimo and Wartenburger. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Auditory-perceptual acuity impacts prosodic boundary prediction in a gating task

Andrea Hofmann<sup>1\*</sup>, Outi Tuomainen<sup>1</sup>, Sandra Hanne<sup>1</sup>, João Verissimo<sup>2†</sup> and Isabell Wartenburger<sup>1†</sup>

<sup>1</sup>Department of Linguistics, Cognitive Sciences, University of Potsdam, Potsdam, Germany, <sup>2</sup>Center of Linguistics, School of Arts and Humanities, University of Lisbon, Lisbon, Portugal

Processing of prosodic phrasing requires listeners to integrate acoustic cues that unfold incrementally during speech comprehension, yet substantial individual differences exist in how listeners use unfolding prosodic information. This study investigated whether individual differences in auditory-perceptual discrimination abilities for prosodic boundary cues are related to processing of prosodic phrasing, and, more specifically, the ability to use the incremental bottom-up prosodic information for making top-down predictions about the syntactic structure of an unfolding utterance. Sixty German-speaking adults completed adaptive staircase procedures measuring Just-Noticeable-Difference thresholds for auditory-perceptual acuity in pitch, pause, and final lengthening discrimination. In addition, they performed a gating task that provided snippets of coordinate three-name sequences with or without an internal prosodic boundary in a randomized order. Performance in the gating task was analyzed using Bayesian multilevel Signal Detection Theory models to separate discriminability from response bias. Participants with higher auditory-perceptual acuity demonstrated better prediction of the upcoming structure across all gates. When all three auditory-perceptual acuity measures were modeled simultaneously, each individual effect attenuated substantially, indicating shared, rather than independent, predictive variance. These findings suggest that top-down prediction during speech comprehension is related to overall auditory-perceptual acuity rather than independent boundary-cue-specific sensitivities.

### KEYWORDS

auditory-perceptual acuity, final lengthening, gating paradigm, just-noticeable difference, pause, pitch, prosodic boundary, prosodic boundary cue

## Introduction

Prosodic phrasing refers to the organization of pitch, timing, rhythm, and intensity that accompanies spoken language. It guides listeners through an utterance by marking syntactic structure, signaling discourse relations, highlighting information structure, and conveying pragmatic and/or emotional meaning (e.g., Cole, 2015; Frazier et al., 2006; Wagner and Watson, 2010). A central function of prosodic phrasing is the segmentation of continuous speech into hierarchically organized units, enabling listeners to map the acoustic stream onto syntactic and semantic structure during comprehension (e.g., Clifton et al., 2002; Frazier et al., 2006).

Consider the sentences: “*What’s that ahead in the road?*” and “*What’s that, a HEAD in the ROAD?*” (Kjelgaard and Speer, 1999, p. 153), which contain identical segmental content. What distinguishes their meanings is the prosodic structure, specifically, where prosodic phrase boundaries occur and which words receive prominence. In the first version, continuous phrasing groups *ahead in the road* into one constituent, yielding the interpretation of something visible on the road ahead. In the second version, a boundary after *that* with nuclear accents on *head* and *road* signals a different syntactic organization, prompting listeners to interpret *a head in the road* as a noun phrase. This example illustrates how prosodic boundaries resolve structural ambiguities that cannot be disambiguated from segmental information alone (e.g., Frazier et al., 2006).

Listeners construct prosodic phrasing through real-time integration of incremental acoustic information with higher-level linguistic knowledge. This process involves both bottom-up processing—the extraction of acoustic cues from the speech signal—and top-down prediction—the use of syntactic, semantic, and pragmatic expectations to anticipate upcoming prosodic structure (e.g., Cole et al., 2010; Ferreira and Karimi, 2015; Ji et al., 2024; Wagner and Watson, 2010). Understanding how listeners balance these two sources of information, and why individuals differ in this ability, remains a central question in prosodic processing research.

## Acoustic cues to prosodic boundaries

Listeners rely primarily on three acoustic cues to identify prosodic phrase boundaries: Pitch movements, silent pauses, and pre-boundary lengthening (e.g., Cole, 2015; Kentner and Féry, 2013; Petrone et al., 2017). Pitch movements (e.g., final rises, falls, or pitch reset at phrase onset) provide the main spectral cues, while silent pauses and lengthening of vowels or syllables preceding a boundary (final lengthening) serve as major temporal cues (e.g., Schubö et al., 2023; Tyler and Cutler, 2009). These three prosodic boundary cues constitute the most robust and extensively studied boundary markers in German and are known to support reliable boundary perception and structural disambiguation (e.g., Hansen et al., 2023; Kentner and Féry, 2013; Petrone et al., 2017; Schubö et al., 2023). Beyond these three prosodic boundary cues, prosodic boundaries can be marked by other acoustic and articulatory cues. These include intensity changes (e.g., Kochanski et al., 2005), segmental lengthening patterns beyond phrase-final position (e.g., Byrd and Saltzman, 2003), voice quality modulations (e.g., González et al., 2022), and domain-initial strengthening of post-boundary segments (e.g., Cho and Keating, 2009).

Across languages, pitch consistently signals prosodic phrasing, but generally carries less perceptual weight than temporal cues (e.g., Collier et al., 1993; Ganga et al., 2024; Lin and Fon, 2010). This asymmetry reflects the inherently greater salience of abrupt temporal events—such as silence or marked segmental slowing—compared to more gradually unfolding pitch movements. Although pitch is perceptually less dominant for marking local phrase boundaries, it serves many broader functions; pitch rises, among other things, guide listeners’ attention and facilitate prosodic

chunking or memory processes (Lialiou et al., 2024). In contrast, silent pauses serve as highly reliable markers of local boundaries and often trigger categorical boundary judgments (e.g., Männel and Friederici, 2016; Petrone et al., 2017; Yang et al., 2014). Final lengthening on its own usually lacks sufficient perceptual salience to trigger boundary perception, but functions most effectively when accompanied by pitch and/or pause cues (e.g., Ganga et al., 2024; Schubö et al., 2023).

Rather than treating prosodic boundary cues in isolation, listeners integrate them incrementally in a non-additive way: When a pause is present, additional cues add little perceptual benefit, but when a pause is absent, pitch and pre-boundary lengthening become more informative (e.g., Yang et al., 2014). In German, pitch change and pre-boundary lengthening must co-occur to elicit robust boundary detection (e.g., Holzgrefe-Lang et al., 2016). Prosodic boundary cues also interact perceptually: Rising pitch increases the perceived duration of a syllable compared to level pitch, even when the objective duration of the syllable is constant. Additionally, the pitch contour of syllables flanking a silent interval modulates the perceived pause duration (also known as the auditory kappa effect Cohen et al., 1953). When the distance between pre-pause and post-pause pitch is greater (e.g., pre-boundary rise followed by low post-boundary pitch), the intervening silence is perceived as longer (Brugos and Barnes, 2014). Such interactions demonstrate that spectral and temporal cues jointly shape prosodic boundary perception through dynamic spectrotemporal integration.

## Boundary perception in coordinate structures

In the present study, we investigate how listeners use prosodic boundary cues in coordinate name sequences, where prosodic information determines syntactic grouping. Consider the question “Who is arriving at the station?” The answer could be:

- (1) [*Moni und Lilli*] # *und Manu* (*grouped* condition with internal grouping; und = and)  
or
- (2) *Moni und Lilli und Manu* (*ungrouped* condition without internal grouping; und = and).

In (1), a prosodic boundary after the second name (marked here with #, see also Hansen et al., 2023) creates an internal grouping, indicating that *Moni* and *Lilli* arrive together while *Manu* arrives separately. In (2), the absence of such a boundary implies that all three arrive together. Crucially, only the presence or absence of a prosodic boundary disambiguates the intended grouping while the segmental content is identical in both cases.

Coordinate name sequences offer an ideal testing ground for studying prosodic phrasing because they exhibit systematic acoustic variation tied to syntactic grouping but contain identical segmental content, that is, the syntactic structure is solely conveyed by means of prosody. Kentner and Féry’s (2013) Proximity/Similarity model formalizes how prosodic boundaries are distributed in such structures. The model predicts that in *grouped* sequences

like [*Name1 und Name2*] # *und Name3*, the Proximity principle weakens internal boundaries within the syntactic group (at or after *Name1*), while the Anti-Proximity principle strengthens boundaries at group edges (after *Name2*). Consequently, the *grouped* structure should exhibit a stronger prosodic boundary after *Name2* compared to the *ungrouped* structure (2) *Name1 und Name2 und Name3*, which lacks internal grouping. Empirical studies confirm that speakers mark grouping through systematic modulation of pitch, pause, and final lengthening, and that listeners reliably recover grouping from these prosodic boundary cues (e.g., Ganga et al., 2024; Huttenlauch et al., 2021; Schubö et al., 2023).

The properties mentioned above make coordinate structures valuable for examining the interplay between bottom-up and top-down processing in prosodic boundary perception. The question is not merely which boundary cues listeners use, but how much bottom-up prosodic information listeners need before they can generate reliable top-down predictions about the upcoming syntactic structure, specifically, whether the utterance belongs to a *grouped* (1) or *ungrouped* structure (2).

Hansen et al. (2023) addressed this question and examined how much prosodic information is necessary to predict the upcoming syntactic structure using a gating paradigm (see Grosjean, 1980). In this paradigm, listeners hear segments of an utterance which get progressively longer (“gates”), and each successive “gate” reveals more acoustic information. Hansen et al. (2023) presented syllable snippets of German coordinate name sequences, starting with *Mo* (Gate 1), progressing to *Moni* (Gate 2), then *Moni und* (Gate 3), *Moni und Li* (Gate 4)... through to the complete sequence *Moni und Lilli und Manu* (Gate 7). After each gate, listeners decided whether they predicted a *grouped* structure “[*Moni und Lilli*] *und Manu*” (example (1)) or an *ungrouped* structure “*Moni und Lilli und Manu*” (example (2)). The central question was whether listeners could exploit early, subtle bottom-up prosodic boundary cues to make top-down predictions about grouping before encountering the critical boundary after *Name2* itself, as reflected in the accuracy of their grouping decisions at each gate, and whether listeners differed in their predictive abilities.

Across participants, overall accuracy reached near ceiling when participants listened to *Moni und Lilli* (*Name1 and Name2*, Gate 5). However, based on the timing and stability of listeners’ responses across gates, Hansen et al. (2023) identified two distinct listener subgroups: Approximately 60% of participants seemed to update their grouping predictions as more and more prosodic information became available already from Gate 2 (“identification group”). In contrast, the remaining participants showed a “waiting” pattern, maintaining consistent responses until later gates (e.g., Gate 5) when clear boundary evidence had accumulated (“waiting group”). Both groups ultimately achieved similarly high accuracy, indicating that the “waiting group” either had reduced perceptual abilities and/or they employed a differential processing strategy. Because Hansen et al. (2023) always presented the gates in fixed ascending order (always starting from Gate 1 and progressing forward), their design could not determine whether this “waiting” pattern reflected genuine auditory-perceptual limitations or a deliberate strategic choice. Listeners who appeared to “wait” might have been unable to detect subtle early prosodic boundary cues, but

they could just as well have been applying a conservative decision strategy, waiting for stronger evidence before committing to a prediction.

The present study therefore sought to disentangle these possibilities by examining whether individual differences in bottom-up auditory-perceptual acuity—the ability to detect subtle prosodic differences—explain why some listeners can generate reliable top-down predictions earlier than others.

## The present study: aims and hypotheses

Since prosodic processing relies on the incremental detection of subtle prosodic boundary cues, such as pitch, pause, and final lengthening, differences in auditory-perceptual discrimination abilities (auditory-perceptual acuity) for these cues may be the critical factor driving the individual differences in the timing of correct boundary predictions observed by Hansen et al. (2023). Listeners with finer perceptual resolution for the respective acoustic dimensions should be better equipped to exploit early, subtle prosodic boundary cues for predicting upcoming syntactic structure. The present study investigates this hypothesis by directly linking bottom-up perceptual abilities to top-down boundary prediction performance in a gated paradigm.

We measured perceptual thresholds using Just-Noticeable-Difference (JND) tasks for pitch, pause, and final lengthening discrimination, providing an index of listeners’ perceptual resolution, more specifically, how small a change in each prosodic boundary cue dimension they can reliably detect. To test boundary prediction, we adapted Hansen et al. (2023)’s gating paradigm to investigate inter-individual differences in using bottom-up prosodic information for top-down predictive processes. Unlike Hansen et al. (2023)’s ascending presentation of successive gates, gates were presented in randomized order with the goal of reducing strategic effects such as deliberate withholding of responses. Instead, our experimental paradigm required participants to base each judgment solely on the acoustic evidence available at a certain gate.

We asked the following research question: How does auditory-perceptual acuity for prosodic boundary cues relate to the ability to predict syntactic structure (*grouped* vs. *ungrouped*) from partial prosodic information in gated speech?

We hypothesized that participants with higher auditory-perceptual acuity (as indicated by lower JND thresholds) for pitch, final lengthening, and pause would demonstrate an enhanced, that is, earlier prediction of the upcoming syntactic structure compared to participants with lower auditory-perceptual acuity (reflected in higher JND thresholds). This auditory-perceptual acuity advantage should be evident across gates, but particularly pronounced at early gates (Gates 2-4) where prosodic information is more subtle as compared to later Gates 5 and 7. Specifically, effects should be detectable already at Gate 2 (*Name1* only), because listeners who are able to discriminate smaller acoustic differences should be able to use even subtle prosodic information for their predictions.

To address limitations of accuracy measures specified in our preregistration (<https://osf.io/dgu7v>), we adopted a Signal

Detection Theory framework to separate two distinct processes: Discriminability—how well participants detect acoustic differences between *grouped* and *ungrouped* structures, and response bias—their general tendency to predict one structure over another, regardless of acoustic evidence (e.g., Hautus et al., 2021; Zloteanu and Vuorre, 2024). This distinction is essential for our research question. If better auditory-perceptual acuity enables early exploitation of prosodic boundary cues, it should manifest primarily as enhanced discriminability. In other words, listeners with better auditory-perceptual acuity should be able to detect subtle prosodic differences that signal upcoming structure. However, analyzing performance using only accuracy conflates discriminability with response bias, obscuring whether better performance reflects improved acoustic sensitivity or simply response preferences. By modeling these components separately within a unified regression framework, we can test whether auditory-perceptual acuity specifically enhances sensitivity to prosodic boundary cues at early gates.

We examined the relation of bottom-up auditory-perceptual acuity and top-down prediction of the upcoming syntactic structure through two complementary Signal Detection Theory-based model setups: First, we tested each auditory-perceptual acuity measure separately to establish whether pitch acuity, pause acuity, and final lengthening acuity each determine boundary prediction performance. Second, we modeled all three auditory-perceptual acuities simultaneously. This combined modeling approach allowed us to determine whether any single prosodic boundary cue stands out as particularly important for boundary prediction, thus testing whether one auditory-perceptual acuity provides predictive power beyond what the others explain. If an auditory-perceptual acuity effect still remains after the other auditory-perceptual acuities have been statistically controlled for, this would indicate that the specific perceptual ability in question plays a particularly important role in prosodic phrasing. Conversely, if effects attenuate in the combined model, this would indicate that all three auditory-perceptual acuity measures share predictive variance.

## Materials and methods

This study was preregistered (<https://osf.io/dgu7v>).

### General procedure

Participants were tested individually in a sound-attenuated booth. The written name sequences and response choice images were presented on a 1080×1920 pixel monitor, with keyboard input used to record responses. Auditory stimuli were presented via a Beyerdynamic DT-297 headset (80 Ohm headphones), connected to a Focusrite Scarlett 18i8 audio interface. All experimental procedures were controlled by custom Python 3.8 scripts (PyCharm, Windows 10).

## Gating task

### Stimuli

The stimulus set (adapted from Hansen et al., 2023) comprised six coordinate three-name sequences, each containing three disyllabic German names connected by “und” (“and”). Within each sequence, the first two names consistently ended with an /i/ sound (e.g., Moni, Lilli, Leni, Nelli, Mimmi, Manni), while the third name ended in either /u/ or /a/ (e.g., Manu, Nina, Lola). Each sequence was produced under two prosodic stimulus conditions: (1) A *grouped* stimulus condition featuring a prosodic boundary after the second name (*[Name1 and Name2] and Name3*), and (2) an *ungrouped* stimulus condition without such a boundary (*Name1 and Name2 and Name3*).

The name sequences were derived from audio recordings produced in a prior study by Huttenlauch et al. (2021), by four female speakers (mean age = 24 years, SD = 4.24, range: 21–30 years). These recordings were selected based on high perceptual congruence between intended and perceived stimulus conditions ( $\geq 98\%$ ), as determined through a perception check where naïve listeners categorized each recording from the complete production corpus as *grouped* or *ungrouped*.

Each name sequence recording was segmented into seven temporal gates of increasing duration, revealing progressively more prosodic information:

- Gate 1: *Name1* – 1st syllable only (e.g., “Mo”)
- Gate 2: *Name1* complete (e.g., “Moni”)
- Gate 3: *Name1* + conjunction1 (e.g., “Moni und”)
- Gate 4: *Name1* + conjunction1 + *Name2* — 1st syllable (e.g., “Moni und Li”)
- Gate 5: *Name1* + conjunction1 + *Name2* complete (e.g., “Moni und Lilli”)
- Gate 6: *Name1* + conjunction1 + *Name2* + conjunction2 (e.g., “Moni und Lilli und”)
- Gate 7: Complete sequence (e.g., “Moni und Lilli und Manu”)

Our approach deviates from Hansen et al. (2023), who employed all seven gates: We excluded Gate 1 because Hansen et al. (2023) found that performance at this gate was unstable. Some participants who scored above chance at Gate 1 dropped back to chance level at Gate 2, indicating that the single first syllable carries insufficient and unreliable prosodic information. We also excluded Gate 6 because it provided equivalent prosodic boundary cue information to Gate 7 (the full prosodic context), making it redundant for our analyses. Note that prosodic boundary cues in the *grouped* stimulus condition are maximal at/after the second name (see Huttenlauch et al., 2021).

Our experimental design yielded a total of 240 stimuli (4 speakers × 2 stimulus conditions × 5 gates × 6 name sequences).

### Prosodic boundary cue acoustics

Cue measurements and extraction: Following Hansen et al. (2023), we extracted prosodic boundary cue measurements from the Huttenlauch et al. (2021) recordings at two temporal locations: *Name1* (where early boundary-related cues emerge) and

at or immediately following *Name2* (where late boundary cues accumulate). Measurements were obtained using Praat (Boersma and Weenink, 1992–2020) and included: (i) Pitch range, defined as the difference between  $f_0$  minimum and maximum across the first and second syllable (in semitones), measured separately for *Name1* and *Name2*; (ii) pause duration, defined as the duration of any silent interval following *Name1* or *Name2*, relative to total utterance duration (in percent); and (iii) final lengthening, defined as the duration of the final vowel relative to total name duration (in percent), also measured separately for *Name1* and *Name2*.

In *grouped* productions, prosodic boundary cues on *Name1* are expected to be attenuated due to proximity within the first prosodic group, whereas prosodic boundary cues on *Name2* are expected to be enhanced, marking the major prosodic boundary. Because prosodic boundary cues unfold over time, their alignment with specific gates is necessarily approximate: Pitch range and final lengthening are expressed most clearly at Gates 2 and 5, whereas pause perception requires the onset of subsequent speech material (Gates 3 and 7).

Figure 1 presents the distribution of prosodic boundary cue strength for *Name1* and *Name2* separately for *grouped* and *ungrouped* stimulus conditions. Pitch range on *Name1* shows acoustic differentiation with modest distributional overlap between stimulus conditions, with *grouped* stimuli showing reduced pitch range relative to *ungrouped* stimuli ( $\Delta = -3.32$  st,  $\ln\text{BF}_{10} = 8.30$ , strong evidence). Final lengthening on *Name1* shows no consistent stimulus condition differences ( $\Delta = -2.37$  %,  $\ln\text{BF}_{10} = 0.63$ , inconclusive evidence, numerically favoring H1), and pauses after *Name1* were virtually absent in both stimulus conditions ( $\Delta = -0.39$  %,  $\ln\text{BF}_{10} = -0.87$ , inconclusive evidence, numerically favoring H0). In contrast, prosodic boundary cues were more distinct on *Name2*. Pitch range showed a clear separation between stimulus conditions ( $\Delta = 5.37$  st,  $\ln\text{BF}_{10} = 26.0$ , strong evidence), and pauses displayed categorical-like distributions: Minimal pausing in the *ungrouped* stimulus condition vs. substantial pauses in the *grouped* stimulus condition ( $\Delta = 16.1$  %,  $\ln\text{BF}_{10} = 22.7$ , strong evidence). Final lengthening showed moderate stimulus condition differences ( $\Delta = 9.12$  %,  $\ln\text{BF}_{10} = 8.51$ , strong evidence), though less consistently than pitch and pause cues. Overall, the pattern is consistent with stronger boundary marking at/after *Name2* in line with Kentner and Féry (2013). The stimulus condition differences at the different name positions were evaluated using Bayesian paired *t*-tests (Rouder et al., 2009), with *grouped* and *ungrouped* tokens paired by speaker and name sequence ( $n = 24$  pairs per cue).  $\ln\text{BF}_{10}$  denotes the natural-logged Bayes Factor quantifying the strength of evidence for stimulus condition differences (values  $> 1$  supporting a difference and values  $< 1$  supporting the null; see the Statistical modeling section for interpretation thresholds), whereas the corresponding  $\Delta$  values index the magnitude of those differences.

## Procedure

Participants completed a binary decision gating task (originally introduced by Grosjean, 1980) judging whether each gated stimulus snippet corresponded to the *grouped* or *ungrouped* stimulus condition (see Hansen et al., 2023). Participants received written

instructions on screen explaining that they would hear audio snippets of name sequences (e.g., only “Mimmi” or “Mimmi und Mo”). After each audio snippet, two pictograms appeared on screen, each symbolizing one of the two possible grouping structures. Participants’ task was to decide which pictogram best matched the grouping structure conveyed by the snippet they just heard, and to indicate their choice by pressing the corresponding arrow key (left or right) on the keyboard.

**Trial structure:** Each trial began with a fixation cross (1 s) at screen center, followed by the presentation of a single gated audio snippet over headphones. After a 500 ms delay, two response pictograms appeared on screen, symbolizing the two possible grouping structures. One pictogram depicted two stick figures positioned close together with a third spatially separated, representing the *grouped* stimulus condition (two arriving together, one arriving separately). The other pictogram showed three equidistant stick figures without spatial separation, representing the *ungrouped* stimulus condition (all three arrive together). Arrows below each pictogram (pointing left/right) indicated the corresponding keyboard response. Participants responded using the designated keyboard keys after the audio presentation. The condition-key-mapping (*grouped*-left / *ungrouped*-right vs. *grouped*-right / *ungrouped*-left) was counterbalanced between participants but fixed within each participant. The visual response pictograms remained on screen until participants responded.

**Experimental procedure:** The task began with a practice block of 10 randomized trials with visually presented accuracy feedback. Practice stimuli were gated snippets from another female speaker not included in the main task.

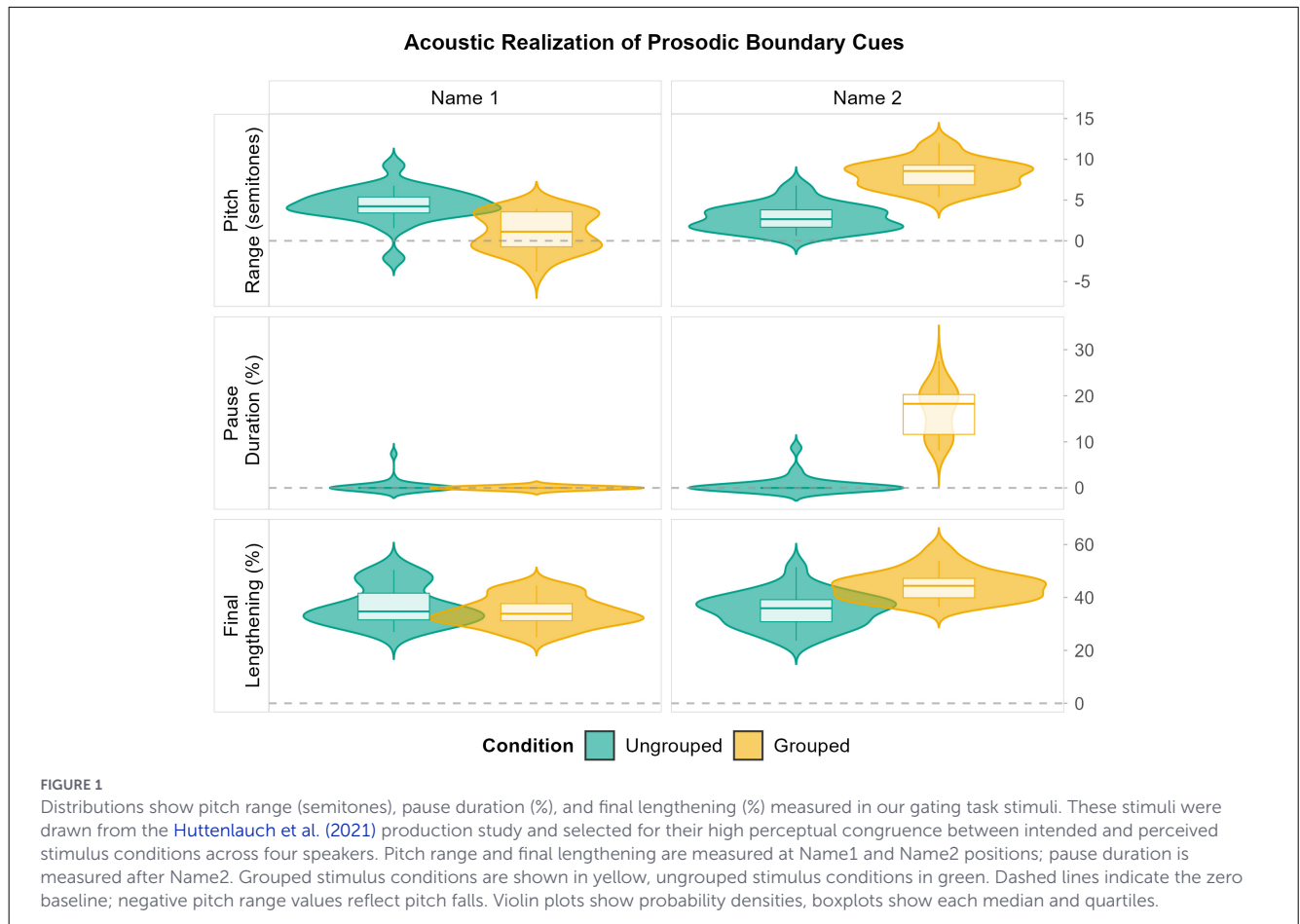
The main task consisted of four speaker blocks (60 trials per block), each containing recordings from a single speaker. The order of these blocks was randomized across participants. Within each block, all Gate 2–5 stimuli for the respective speaker ( $N = 48$ ) were presented in a randomized order first, followed by all Gate 7 stimuli ( $N = 12$ ) also in random order. This design deviates from Hansen et al. (2023), who used a fixed ascending gate order from Gate 1 to Gate 7 for each trial. In the present paper, we randomized the Gates 2–5 presentation to eliminate a strategic waiting behavior while maintaining decision difficulty, with Gate 7 serving as a fully informative baseline. Participants could take a short break between blocks. Each block lasted approximately 4 min.

## Just-Noticeable-Difference (JND) task

The JND tasks reported here follow the same procedure described in Hofmann et al. (Submitted).

## Stimuli

We developed three separate stimulus continua to measure participants’ auditory-perceptual acuity thresholds for the three prosodic boundary cues typically used in German for prosodic boundary marking: Pitch rise, pause duration, and final lengthening. Importantly, although these prosodic boundary cues originate from prosodic boundary contexts, this perception



task did not assess boundary perception but rather individual auditory-perceptual acuity to the underlying acoustic cue strength.

The base stimuli for the continua were derived from original recordings used in a perception experiment by de Beer et al. (2022), in which a phonetically trained female speaker produced coordinate three-name sequences with varied prosodic boundary realizations. From these recordings, we selected the tokens exhibiting maximal prosodic boundary cue expression, meaning those tokens where the acoustic realization on *Name2* (for pitch and final lengthening) or immediately following *Name2* (for pause) showed the strongest manifestation of the relevant prosodic boundary cue. Specifically, we selected the instance where (a) the pitch rise between the stressed and unstressed syllable of *Name2* showed the largest excursion, (b) a clear and extended silent interval followed *Name2*, or (c) the final segment of *Name2* was maximally lengthened.

These maximally realized prosodic boundary cue segments were extracted and used as base stimuli for constructing three JND continua using custom Praat scripts (Boersma and Weenink, 1992–2020). Each continuum spanned from the maximally expressed prosodic boundary cue to a minimally expressed prosodic boundary cue. Thus, the starting point of each continuum represented the original, clearly perceivable prosodic boundary cue (base stimulus), while the end point of each continuum (reference stimulus) represented an acoustically neutral version with minimal or no prosodic boundary cue expression. The intermediate steps formed comparison stimuli with systematically decreasing prosodic boundary cue strength.

**Pitch rise continuum:** The base stimulus was *Name2* — *Nelli*, produced with a pitch rise of 13 semitones. The pitch contour was progressively flattened from 13 to 0 semitones in 0.005-semitone increments, yielding a flat, non-rising contour as the reference stimulus.

**Pause duration continuum:** The base stimulus was the coordinate phrase *Name2 und Name3* — “*Moni [PAUSE] und Lilli*”, containing a 550 ms silent interval after *Moni*. The silence duration was shortened from 550 ms to 0 ms in 1 ms increments, ending with a reference stimulus without a pause.

**Final lengthening continuum:** The base stimulus was *Name2* — *Mimmi*, produced with a final vowel duration of 225 ms (approximately half of the total word duration). The final segment was progressively shortened in 0.3 ms increments until reaching 61 ms, resulting in roughly equal syllable durations (no perceivable final lengthening) as the reference stimulus.

## Procedure

Auditory-perceptual acuity thresholds were measured using an AXB oddball discrimination task with an adaptive staircase procedure (based on Smith et al., 2020). Each participant completed three separate JND tasks, one per prosodic boundary cue (pitch rise, pause duration, final lengthening), administered in randomized order. Each task began with a short practice block with visual accuracy feedback until participants achieved four consecutive

correct responses. No feedback was provided during experimental trials.

**Trial structure:** In each trial, three auditory stimuli were presented in an AXB sequence (either AAB or ABB) with 500ms inter-stimulus intervals. On every trial, participants heard two tokens: A reference stimulus with minimal (or no) prosodic boundary cue expression (0 semitones pitch rise, 0 ms pause, 61 ms final lengthening) and a comparison stimulus with a detectable prosodic boundary cue strength (from the respective prosodic boundary cue continua). The acoustic difference between these two stimuli—which we term the *cue difference*—determined how easy or difficult discrimination was on that trial. Large *cue differences* made discrimination straightforward, while small *cue differences* approached the limits of perceptual discriminability.

The order of reference and comparison tokens was randomized, resulting in trials presenting either two identical reference stimuli and one comparison, or two identical comparison stimuli and one reference. Participants identified the odd-one-out by pressing the left arrow key (for ABB patterns) or right arrow key (for AAB patterns). Visual response prompts remained on screen until participants responded, showing corresponding arrow icons and schematic stimulus patterns (with “A” boxes in green, “B” boxes in black).

**Adaptive procedure:** The adaptive staircase dynamically adjusted the *cue difference* presented on each trial to converge on each participant’s discrimination threshold - the smallest acoustic difference they could reliably detect. Each staircase began with the maximum *cue difference* to ensure initial success: 13 semitones for pitch rise, 550ms for pause duration, and 164ms for final lengthening. After each trial, the *cue difference* for the subsequent trial was adjusted based on performance: Correct responses decreased the *cue difference* (making the next trial harder by bringing the comparison stimulus closer to the reference stimulus), while incorrect responses increased it (making the next trial easier by moving the comparison farther from the reference).

The magnitude of these adjustments, i.e., the step size, also changed dynamically throughout each staircase. Initial step sizes were large to allow rapid descent from the maximum starting difference toward the participant’s approximate threshold region, then progressively decreased to enable precise threshold estimation: From 0.75 to 0.005 semitones for pitch rise, from 30 to 1 ms for pause duration, and from 7.5 to 0.3 ms for final lengthening. This combination of large initial steps (for efficiency) and small later steps (for precision) produced accurate threshold estimates within a reasonable number of trials.

The staircase initially followed a 1-down-1-up adjustment rule, where each correct response decreased the *cue difference* and each incorrect response (would have) increased it, allowing rapid initial convergence. After the first error, the procedure switched to a 2-down-1-up rule requiring 2 consecutive correct responses to decrease the *cue difference* but only 1 incorrect response to increase it. This asymmetric rule converges on a performance level of approximately 71% accuracy (Levitt, 1971), which provides a stable estimate of the smallest reliably detectable *cue difference* while avoiding ceiling or floor effects. A reversal occurred when the adjustment direction changed from decreasing to increasing *cue difference* or vice versa, indicating that the staircase had crossed the

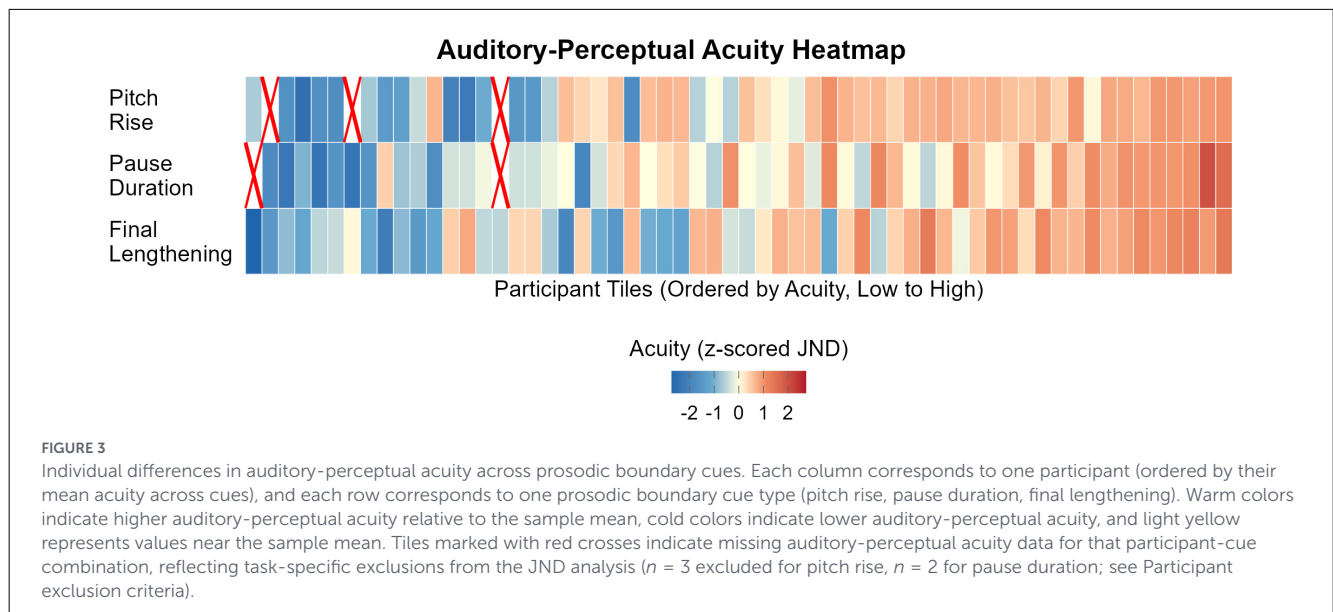
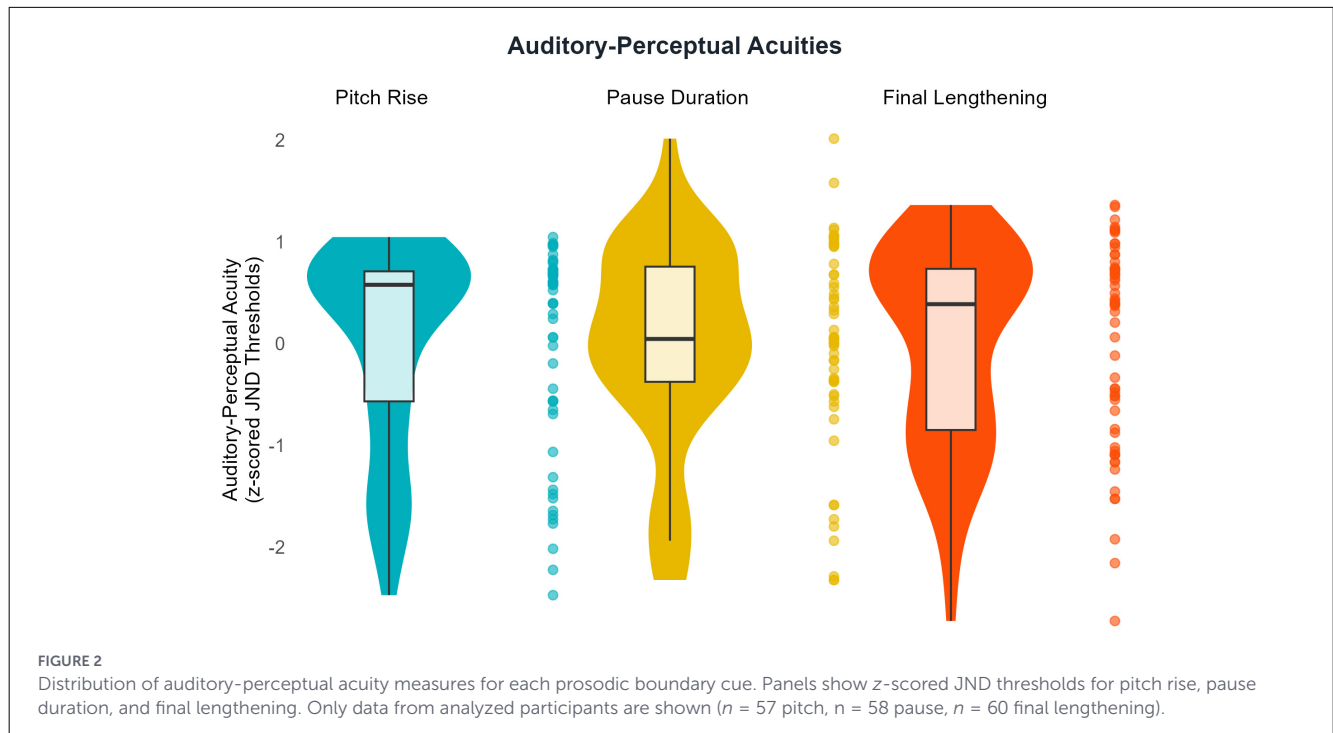
participant’s threshold. The task stopped after either 120 trials or 18 reversals, whichever came first. JND threshold calculation was based on the *cue differences* at reversal points.

## Data pre-processing

The JND threshold for each prosodic boundary cue was calculated as the mean of the six most stable consecutive reversal points, defined as the set of six consecutive reversals exhibiting the lowest standard deviation (Brunner et al., 2011; Oschkinat et al., 2022). JND thresholds represent the smallest acoustic difference a listener can discriminate; thus, lower thresholds indicate better perceptual ability. However, to facilitate interpretation throughout the analyses and to align with the intuitive meaning of “acuity” (where higher values indicate better ability), JND values were *z*-scored and direction-reversed such that higher auditory-perceptual acuity scores correspond to better perceptual sensitivity. These transformed thresholds are referred to throughout as pitch acuity, pause acuity, and final lengthening acuity, representing each participant’s *z*-scored auditory-perceptual acuity for the respective prosodic boundary cue. Figure 2 displays the distributions of the *z*-scored JND thresholds for each prosodic boundary cue.

Figure 3 visualizes individual auditory-perceptual acuities for the tested prosodic boundary cue continua. Each column represents one participant (ordered left-to-right by overall mean acuity), and each row represents a prosodic boundary cue (pitch rise, pause duration, final lengthening). The different color gradients reflect standardized auditory-perceptual acuity scores relative to the sample mean (warm = higher than mean; cool = lower than mean). The figure reveals substantial variability in auditory-perceptual profiles, both across participants and across prosodic boundary cues. Some participants show consistently elevated (or reduced) sensitivity across all auditory-perceptual acuity measures (uniform color patterns within a column), whereas others exhibit selective strengths for specific prosodic boundary cues, that is, relatively higher auditory-perceptual acuity for some prosodic boundary cues but not others (mixed colors within a column).

As is visible in Figure 3, the heatmap reveals both inter- and intra-individual patterns in auditory-perceptual acuity. Many participants display relatively consistent color patterns across prosodic boundary cues, suggesting shared sensitivity across prosodic boundary cues, while others show notable variation, indicating cue-specific strengths and weaknesses. This pattern is reflected in the moderate correlations between auditory-perceptual acuity measures: Pitch and pause acuity correlated at  $r = 0.52$  [95% CI [0.3, 0.69],  $n = 56$ ], pitch and final lengthening acuity at  $r = 0.30$  [95% CI [0.04, 0.52],  $n = 57$ ], and pause and final lengthening acuity at  $r = 0.36$  [95% CI [0.12, 0.59],  $n = 58$ ]. These correlations indicate overlapping but not identical perceptual sensitivities. Correlations were computed using Pearson’s product-moment correlation between participants’ *z*-scored JND thresholds (auditory-perceptual acuity) for each cue, based on pairwise complete observations ( $n = 56$ – $58$  per cue pair). Participants with higher auditory-perceptual acuity in one prosodic boundary cue tend to show higher auditory-perceptual acuity for the others and vice versa, but substantial individual variation remains, with some



participants showing markedly different auditory-perceptual acuity profiles across the three acoustic dimensions.

## Participants

Sixty native German speakers (48 females) participated in the study, with a mean age of 24.78 years ( $SD = 6.07$ , range: 18–49). Participants had no reported history of speech or language disorders, hearing impairments, or neurological or psychological conditions. They received either monetary reimbursement or course credit for completing two experimental

sessions (approximately 2 h each), scheduled on the same or separate days based on availability. Of the two tasks reported here, the JND task was conducted in the first session, while the second session began with the gating task. The study was conducted in accordance with the Declaration of Helsinki and approved by the University of Potsdam Ethics Committee (approval code: 99/2020). Informed consent was obtained from all participants.

## Participant exclusion criteria

**Gating task:** No participants were excluded. Following Hansen et al. (2023), we evaluated participant performance using three

criteria: (a) Above chance accuracy at Gate 7 ( $> 50\%$ ), (b) accuracy above the group mean minus 2SD at Gate 7, and (c) no systematic response patterns. All participants met criteria (a) and (c). However, we did not apply criterion (b). Hansen et al. (2023) used the group mean minus 2SD threshold to identify participants with unusually low performance, potentially indicating poor task compliance. This criterion was not suitable for our randomized-gate design, which was inherently more difficult than their ascending design. Applying it would have removed four participants with high accuracy (85–88%) who clearly demonstrated task compliance.

**JND task:** No participants were excluded based on the pre-registered criterion for adaptive staircase performance (Oschkinat et al., 2022), which required JND thresholds to decrease below 70% of the initial cue difference. However, following data inspection, some participants exhibited extreme JND values for specific prosodic boundary cues (Hofmann et al., Submitted). We thus applied *post-hoc* exclusion criteria using the interquartile range (IQR) rule, removing participants whose JND scores fell below  $Q1 - 2 \times IQR$  or above  $Q3 + 2 \times IQR$  for each prosodic boundary cue separately. This led to the exclusion of three participants from the JND pitch rise task, leaving  $N = 57$ , and two from the JND pause duration task, leaving  $N = 58$ . No exclusions were made for the JND final lengthening task ( $N = 60$ ).

## Statistical modeling

We analyzed binary responses (*grouped* vs. *ungrouped*) using Bayesian generalized mixed-effects regression within a Signal Detection Theory framework (Zloteanu and Vuorre, 2024), fitted with the *brms* package and the Stan programming language (Buerkner, 2018; Stan Development Team, 2020). Responses were modeled using a Bernoulli family with a probit link function. The outcome variable (response) was factor-coded with *ungrouped* as the reference level, such that the model estimated the probability of a *grouped* response. The probit link transforms predicted probabilities into  $z$ -scores on the standard normal distribution. On this scale, a coefficient of 1 represents 1 standard deviation (SD) shift in the underlying decision variable, corresponding to a higher or lower probability of responding *grouped*. This transformation allows model coefficients to be interpreted as the extent to which each predictor shifts a participant's internal decision tendency toward or away from responding *grouped*.

To address our primary hypothesis, we fitted four Bayesian Signal Detection Theory probit regression models: Three separate models, each including one auditory-perceptual acuity (pitch, pause, or final lengthening acuity) as a predictor, and one combined model including all three auditory-perceptual acuities simultaneously. The fixed-effects structure for each model comprised gate (levels 2, 3, 4, 5, 7), stimulus condition (*grouped* vs. *ungrouped*), the relevant auditory-perceptual acuity measure(s), and all two-way and three-way interactions among these predictors. The combined model included the same structure but with terms for pitch acuity, pause acuity, and final lengthening acuity (as well as their interactions with gate and stimulus condition), allowing

us to estimate the unique contribution of each auditory-perceptual acuity while accounting for shared variance among them (see Verissimo, 2023; Wurm and Fiscaro, 2014).

Stimulus condition was sum-coded (*grouped* = +0.5, *ungrouped* = -0.5), such that the stimulus condition coefficient directly reflected discriminability (how well participants distinguished between the two stimulus conditions). Interactions involving stimulus condition (e.g., gate  $\times$  stimulus condition and auditory-perceptual acuity  $\times$  stimulus condition) represent changes in discrimination as a function of accumulated prosodic information or individual auditory-perceptual acuity. In contrast, model terms that do not involve stimulus condition capture response bias, or more concretely, participants' tendency to respond *grouped* or *ungrouped* independent of the actual stimulus condition. Specifically, the intercept reflects overall response bias, while gate and auditory-perceptual acuity main effects represent how this bias changes across gates or with auditory-perceptual acuity level. The factor gate (levels 2, 3, 4, 5, 7) was coded using (centered) sliding-difference contrasts (Gate 3–2, Gate 4–3, Gate 5–4, Gate 7–5), quantifying how discrimination and bias change as additional acoustic information becomes available. Since auditory-perceptual acuity measures were  $z$ -scored (mean = 0, SD = 1), coefficients represent effects associated with a 1SD difference in auditory-perceptual acuity.

Following Barr et al. (2013), we aimed for maximal random-effects structures but constrained complexity to ensure model identifiability given the available data (Bates et al., 2015). The final models included random intercepts for subjects, items, and speakers, random slopes for gate and stimulus condition by subject, and random slopes for gate, stimulus condition, and the relevant auditory-perceptual acuity by item, with correlated random effects. Speaker was included only as a random intercept (not as a random-slope term), since variance estimates become unreliable with fewer than five grouping levels and we only had four speakers (Bolker, 2015).

Model results reported in the main text are based on weakly-informative priors, with normal distributions centered at zero and standard deviations varying by parameter type (see Supplementary material for specific values). These priors were chosen to rule out implausible extremes while allowing large effects in either direction, following current recommendations (see Gelman et al., 2008; Ghosh et al., 2018; McElreath, 2020; Vasishth et al., 2018). Prior predictive checks confirmed that these choices produced reasonable data-level predictions. The detailed prior specifications for all models are provided in Supplementary material.

Hypothesis testing was performed using Bayes Factors, which quantify evidence strength for an effect by comparing the model (i.e., the alternative hypothesis, H1) including the effect to a null model (i.e., the null hypothesis, H0), where the effect is excluded. Since Bayes Factors are sensitive to prior specifications, we conducted a sensitivity analysis with five different prior configurations, ranging from narrower (moderate and strong informative) to wider (moderate wide and wide) settings compared to our default weakly-informative priors. We calculated natural-logged Bayes Factors ( $\ln BF_{10}$ ) using the Savage-Dickey method (Dickey and Lientz, 1970; Wagenmakers et al., 2010), where values

> 1 indicate evidence for H1 (the respective effect), values < -1 support H0 (absence of the effect), values between -1 and 1 are inconclusive, and values > 3 represent strong evidence for H1 (Jeffreys, 1991; Kass and Raftery, 1995; Verissimo, 2025). Effects are considered reliable when they meet these thresholds ( $\ln\text{BF}_{10} > 1$  or > 3) in the base model and show consistent direction and magnitude across prior-sensitivity analyses.

Model convergence was judged by  $\hat{R} \leq 1.01$ , adequate effective sample sizes, and stable trace plots. Finally, model quality was assessed with posterior predictive checks by comparing model predictions to the observed data distribution.

## Results

Below, we report estimates from the three separate auditory-perceptual acuity models (pitch acuity, pause acuity, final lengthening acuity), unless otherwise noted. The complete results for both, the separate models and the combined model, can be found in [Supplementary material](#). Since the auditory-perceptual acuity measures are uncorrelated with our experimental factors (stimulus condition, gate), estimates for effects not involving auditory-perceptual acuity are nearly identical across specifications (see fixed effects tables in [Supplementary material](#)). For effects involving auditory-perceptual acuity, we report both, separate and combined model estimates to reveal how auditory-perceptual acuity effects change when modeled independently vs. simultaneously (i.e., when pitch acuity, pause acuity, and final lengthening acuity are simultaneously included as predictors in the same statistical model).

### Overall discriminability

Participants demonstrated strong discriminability between *grouped* and *ungrouped* stimulus conditions. Across all three separate auditory-perceptual acuity models, the main effect of stimulus condition was large and consistent (pitch acuity model:  $b = 1.93$  probit units, 95% CI [1.54, 2.28],  $\ln\text{BF}_{10} = 37.07$ ; pause acuity model:  $b = 1.92$ , 95% CI [1.54, 2.25],  $\ln\text{BF}_{10} = 40.45$ ; final lengthening acuity model:  $b = 1.87$ , 95% CI [1.51, 2.20],  $\ln\text{BF}_{10} = 42.10$ ). These estimates represent performance averaged across gates and evaluated at mean auditory-perceptual acuity, reflecting the mean-centering of the auditory-perceptual acuity predictors. These results provide robust evidence that participants could reliably distinguish the two prosodic boundary stimulus conditions, confirming that the bottom-up acoustic differences between stimulus conditions were effective in eliciting differential perceptual responses and validating the experimental paradigm.

### Discriminability changes from gate to gate

Figure 4 presents the discriminability trajectory across gates, illustrating the gate  $\times$  stimulus condition interaction. These effects capture how discriminability evolves with increasing gates, reflecting the general temporal pattern of discriminability improvement from one gate to the next, averaged across all

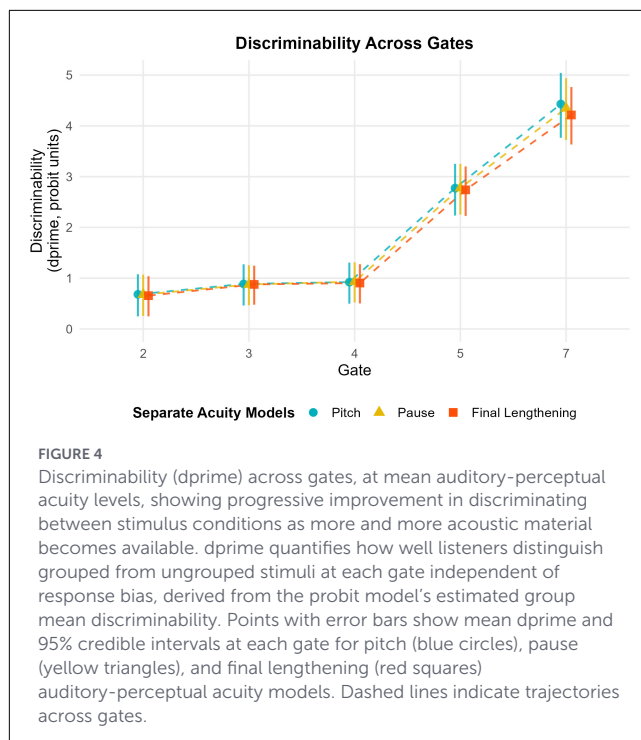


FIGURE 4

Discriminability (dprime) across gates, at mean auditory-perceptual acuity levels, showing progressive improvement in discriminating between stimulus conditions as more and more acoustic material becomes available. dprime quantifies how well listeners distinguish groups from ungrouped stimuli at each gate independent of response bias, derived from the probit model's estimated group mean discriminability. Points with error bars show mean dprime and 95% credible intervals at each gate for pitch (blue circles), pause (yellow triangles), and final lengthening (red squares) auditory-perceptual acuity models. Dashed lines indicate trajectories across gates.

participants regardless of individual differences in auditory-perceptual acuity. Since higher gates contain more acoustic information than lower gates, the interaction reveals the points in time at which critical prosodic information becomes available. Our analysis included Gates 2–5 and Gate 7, excluding Gate 1 due to unstable performance patterns (Hansen et al., 2023) and Gate 6 because it provided equivalent prosodic boundary cue information to Gate 7, making it redundant given that prosodic boundary cues in the *grouped* stimulus condition are maximal at/after the second name (Huttenlauch et al., 2021).

As can be seen in Figure 4, starting from the initial name at Gate 2 (e.g., “Moni”), discriminability (dprime) remained relatively stable throughout the addition of the first conjunction at Gate 3 (e.g., “Moni und”) and the first syllable of the second name at Gate 4 (e.g., “Moni und Li”). A sharp increase in discriminability occurred when the complete second name became available at Gate 5 (e.g., “Moni und Lilli”), providing full access to pitch range and final lengthening cues after *Name2*. Discriminability improved further at Gate 7 (e.g., “Moni und Lilli und Manu”), when the full three-name coordinate structure was revealed and the pause after *Name2* became perceivable. This temporal trajectory emerged consistently across all three separate auditory-perceptual acuity models.

The gate-by-gate contrasts confirmed this pattern: Performance at Gate 3 vs. 2 showed small improvements in discriminability with inconclusive evidence for the effect: Pitch acuity model:  $b = 0.21$ , 95% CI [0.02, 0.39],  $\ln\text{BF}_{10} = 0.18$ ; pause acuity model:  $b = 0.20$ , 95% CI [0.01, 0.39],  $\ln\text{BF}_{10} = -0.04$ ; final lengthening acuity model:  $b = 0.22$ , 95% CI [0.03, 0.41],  $\ln\text{BF}_{10} = 0.58$ . The discriminability improvement from Gate 4 vs. 3 was even smaller and non evident: Pitch acuity model:  $b = 0.03$ , 95% CI [-0.17, 0.24],  $\ln\text{BF}_{10} = -2.30$ ; pause acuity model:  $b = 0.05$ , 95% CI [-0.14, 0.25],  $\ln\text{BF}_{10} = -2.22$ ; final lengthening acuity model:  $b = 0.03$ , 95% CI [-0.17, 0.22],  $\ln\text{BF}_{10} = -2.37$ .

A critical shift, however, occurred when performance at Gate 5 was compared against performance at Gate 4: Pitch acuity model:  $b = 1.85$ , 95% CI [1.34, 2.30],  $\ln\text{BF}_{10} = 36.53$ ; pause acuity model:  $b = 1.85$ , 95% CI [1.33, 2.30],  $\ln\text{BF}_{10} = 37.84$ ; final lengthening acuity model:  $b = 1.83$ , 95% CI [1.31, 2.28],  $\ln\text{BF}_{10} = 36.80$ . This was followed by a continued improvement at Gate 7: Pitch acuity model:  $b = 1.66$ , 95% CI [1.14, 2.16],  $\ln\text{BF}_{10} = 9.32$ , pause acuity model:  $b = 1.57$ , 95% CI [1.09, 2.05],  $\ln\text{BF}_{10} = 8.41$ , final lengthening acuity model:  $b = 1.48$ , 95% CI [1.08, 1.88],  $\ln\text{BF}_{10} = 28.68$ .

## Discriminability modulation by auditory-perceptual acuities

Figure 5 displays the relationship between auditory-perceptual acuities and discriminability across the separate and combined models, illustrating the stimulus condition  $\times$  auditory-perceptual acuity interactions. These effects captured whether individual differences in auditory-perceptual acuities explain variability in overall discriminability (i.e., across all gates). Specifically, they examined whether participants with higher auditory-perceptual acuity showed an enhanced ability to distinguish *grouped* from *ungrouped* stimuli based on the partial prosodic information available in the gated stimuli compared to participants with lower auditory-perceptual acuity. We examined these relationships using two complementary analytical strategies: Fitting separate models for each auditory-perceptual acuity measure (pitch, pause, final lengthening acuity), and fitting a combined model including all three auditory-perceptual acuities together. The separate models establish whether each auditory-perceptual acuity predicts boundary discrimination when examined in isolation. The combined model additionally reveals whether any auditory-perceptual acuity provides unique predictive power beyond the others, or whether effects attenuate due to shared variance among the auditory-perceptual acuity measures (see Veríssimo, 2023; Wurm and Fiscaro, 2014).

The colored regression lines in Figure 5 show the separate models; the gray regression lines show the combined model. When examined separately, all three auditory-perceptual acuities demonstrated clear positive relationships with discriminability, evidenced by the steeper colored slopes. The substantially flatter gray slopes reveal that when all auditory-perceptual acuities are accounted for together in the combined model, the individual effects attenuated. This pattern suggests that all three auditory-perceptual acuity measures share considerable predictive variance.

**Separate acuity models:** When tested separately, both, pause acuity [ $b = 0.29$ , 95% CI [0.11, 0.47],  $\ln\text{BF}_{10} = 2.29$ ] and final lengthening acuity [ $b = 0.25$ , 95% CI [0.07, 0.44],  $\ln\text{BF}_{10} = 1.21$ ] showed evidence for effects on boundary discriminability, indicating that participants with better sensitivity to these temporal prosodic boundary cues demonstrated enhanced ability to distinguish *grouped* from *ungrouped* stimulus conditions. Pitch acuity showed a slightly more modest contribution [ $b = 0.22$ , 95% CI [0.03, 0.42],  $\ln\text{BF}_{10} = 0.22$ ], with inconclusive evidence, suggesting that pitch sensitivity may be less central to this particular boundary detection task.

**Combined acuities model:** When all three auditory-perceptual acuities were modeled simultaneously, each individual effect became substantially attenuated: Pause acuity dropped from  $b = 0.29$  to  $b = 0.17$  [95% CI [-0.07, 0.40],  $\ln\text{BF}_{10} = -1.12$ ], final lengthening acuity dropped from  $b = 0.25$  to  $b = 0.11$  [95% CI [-0.11, 0.33],  $\ln\text{BF}_{10} = -1.69$ ], and pitch acuity similarly dropped from  $b = 0.22$  to  $b = 0.11$  [95% CI [-0.11, 0.33],  $\ln\text{BF}_{10} = -1.74$ ]. Evidence actually favored the null hypothesis for all three auditory-perceptual acuity types, suggesting that none of the three auditory-perceptual acuities provides a unique discriminative advantage when the others are accounted for.

## Gate-specific discriminability modulation by auditory-perceptual acuities

Figure 6 visualizes the relationship between auditory-perceptual acuity and discriminability across gates, illustrating the stimulus condition  $\times$  auditory-perceptual acuity  $\times$  gate interaction. These effects captured whether the relationship between auditory-perceptual acuity and boundary discriminability differed across gates as acoustic information accumulated. In particular, we examined whether auditory-perceptual acuity advantages were particularly pronounced at early gates where prosodic boundary cues are more subtle. As with the two-way interactions, we examined these relationships using separate models (testing each auditory-perceptual acuity individually) and a combined model (testing all three simultaneously) to determine whether any auditory-perceptual acuity shows gate-specific effects beyond the others.

The trajectories in Figure 6 show that discriminability improves from early to later gates (consistent with the gate  $\times$  stimulus condition interactions), but critically, the separation between higher and lower auditory-perceptual acuity levels remains consistent across all gates. Participants with higher auditory-perceptual acuity demonstrated better discrimination at every gate, rather than showing particularly strong advantages at early gates where prosodic boundary cues are subtle.

**Separate acuity models:** All three-way interactions showed evidence against modulation across gates, with effect sizes ranging from -0.40 to 0.21 probit units and  $\ln\text{BF}_{10}$  values from -2.40 to -1.03. An exception occurred with final lengthening acuity at Gate 7 vs. 5, where evidence was inconclusive [ $b = -0.24$ , 95% CI [-0.60, 0.13],  $\ln\text{BF}_{10} = -0.82$ ], representing the largest three-way interaction observed in the separate models.

**Combined acuities model:** The pattern was similar to those in the separate auditory-perceptual acuity models, with all three-way interactions showing evidence against gate-specific modulation ( $\ln\text{BF}_{10}$  from -2.31 to -1.11). Again, the final lengthening acuity  $\times$  Gate 7 vs. 5  $\times$  stimulus condition interaction yielded inconclusive evidence [ $b = -0.45$ , 95% CI [-0.92, -0.01],  $\ln\text{BF}_{10} = 0.50$ ], consistent with the separate model finding.

Simple effects analysis revealed that auditory-perceptual acuity effects remained present across all gates. This demonstrates that the lack of evidence for three-way interactions reflects uniform auditory-perceptual acuity facilitation rather than absent effects.

Thus, better auditory-perceptual acuity enhances discriminability consistently across gates rather than providing differential advantages at specific gates.

### Response bias

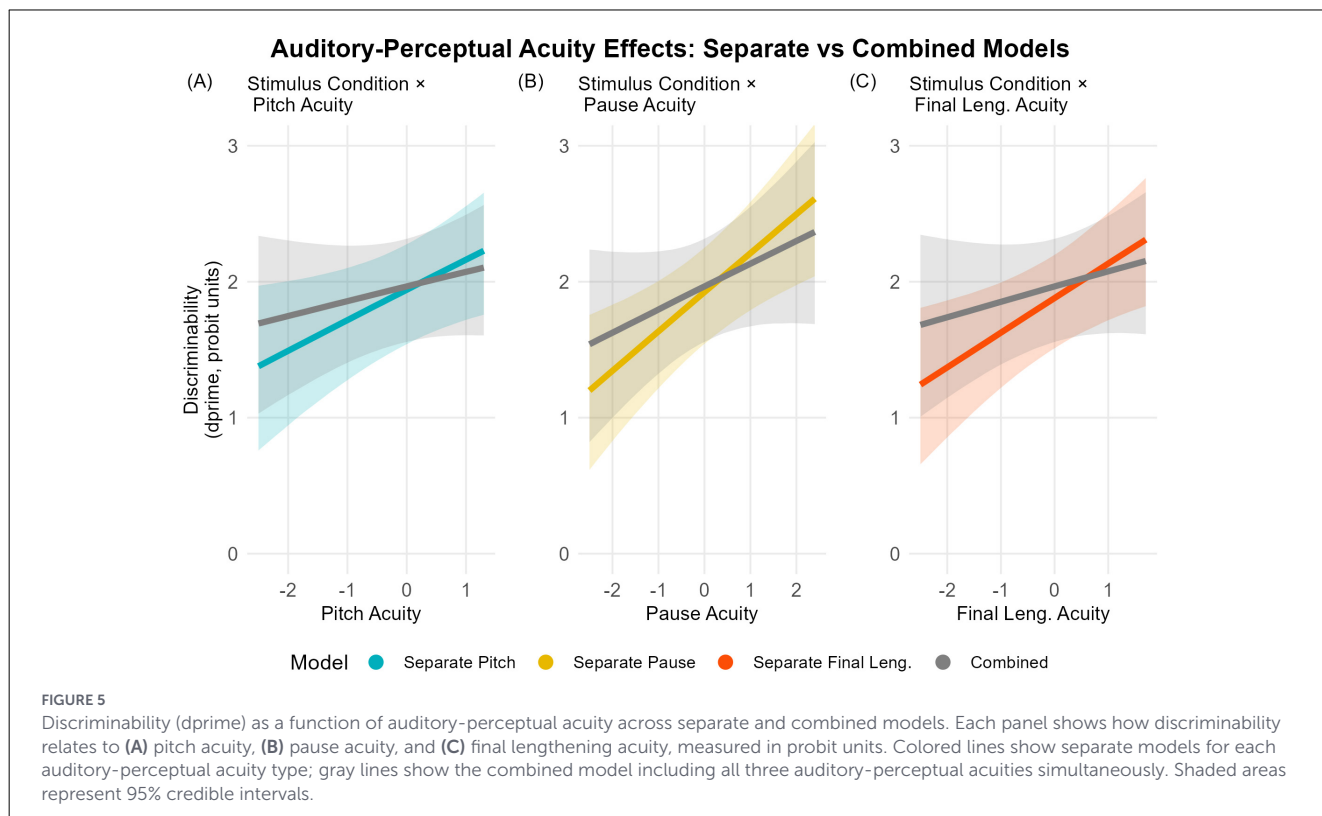
Response bias reflects participants' tendency to favor one response option over another, independent of their actual discrimination performance. These analyses thus examine whether participants showed systematic preferences for *grouped* or *ungrouped* responses and whether such biases varied with auditory-perceptual acuity or gate progression.

Participants showed no systematic response preferences overall (intercepts near zero, all  $\ln BF_{10} < -1$ ). Auditory-perceptual acuity did not influence baseline bias (all  $\ln BF_{10} < -1.5$ ). A small shift in bias emerged at Gate 3 vs. 2, with participants becoming slightly more likely to respond *grouped* ( $b = 0.22-0.24$ ,  $\ln BF_{10} > 2.8$ ), while other gate transitions showed inconclusive effects (all  $\ln BF_{10}$  between  $-0.73$  and  $0.36$ ). Furthermore, auditory-perceptual acuity did not systematically modulate bias changes across gates: Most gate  $\times$  auditory-perceptual acuity interactions showed evidence against effects ( $b = -0.09$  to  $0.01$ ,  $\ln BF_{10}$  from  $-2.33$  to  $-1.14$ ), except for the Gate 7 vs. 5  $\times$  auditory-perceptual acuity interaction, which showed inconclusive evidence ( $b = 0.14 - 0.16$ ,  $\ln BF_{10} =$  between  $0.04$  and  $0.44$ ).

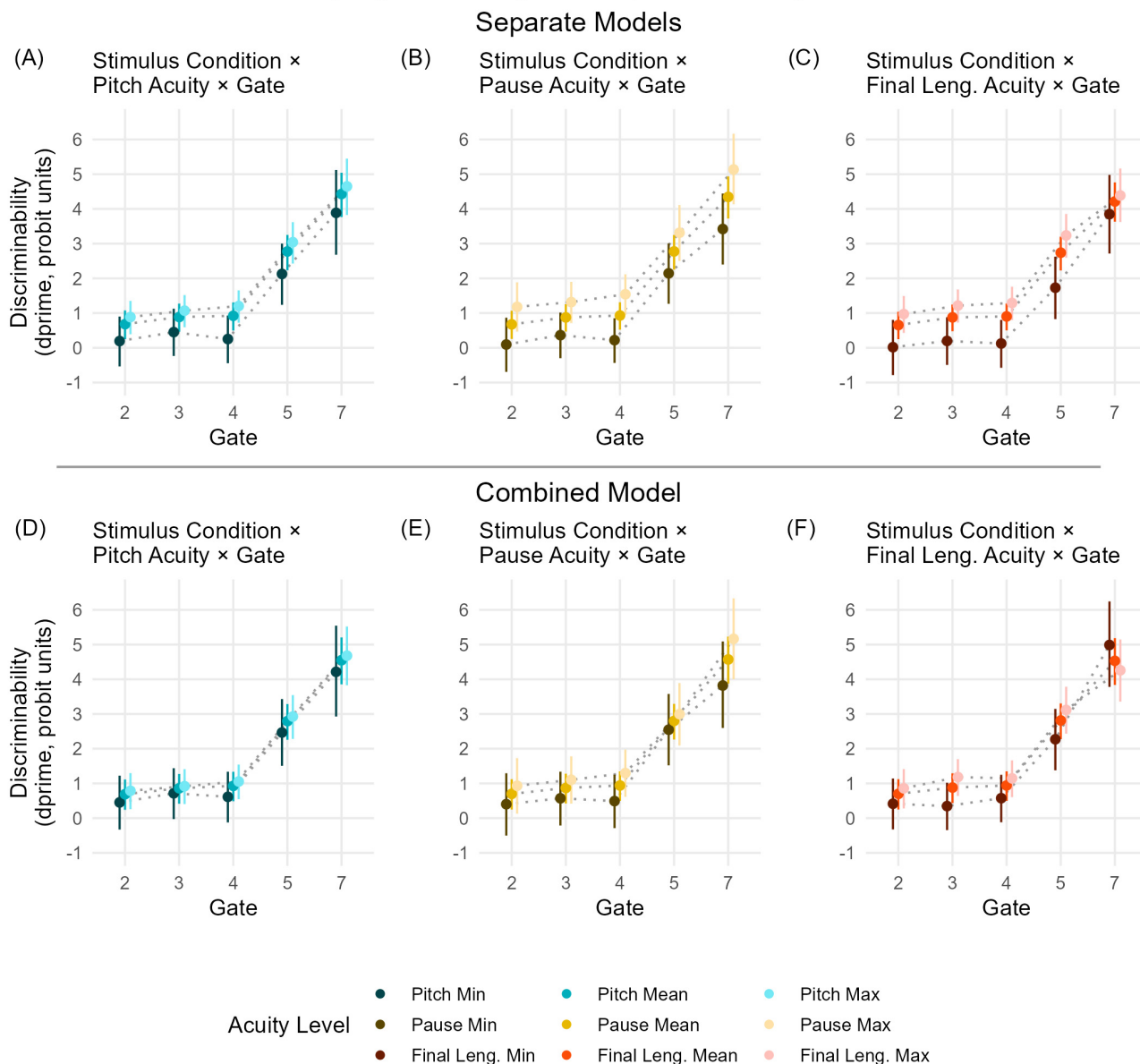
## Discussion

This study investigated how auditory-perceptual acuity for prosodic boundary cues relates to the ability to predict grouping structure (i.e., syntactic grouping) from partial prosodic information in gated speech. We modified the design from Hansen et al. (2023) who identified substantial individual differences in boundary prediction using a similar gating paradigm: In their study, approximately 60% of participants updated their predictions incrementally from early gates, while the remaining 40% maintained consistent responses until later gates when clear boundary evidence had accumulated. However, the ascending gate presentation used by Hansen et al. (2023) made it impossible to distinguish whether “waiting” listeners genuinely lacked perceptual abilities or whether they strategically withheld responses. We addressed this issue by measuring individual differences in auditory-perceptual acuity for pitch, pause, and final lengthening discrimination in a randomized-gate paradigm, testing whether facilitation was uniform across gates or gate-specific as acoustic information unfolds.

Overall, participants successfully predicted *grouped* vs. *ungrouped* structures from prosodic information. The large main effect of stimulus condition confirmed that the acoustic differences between stimulus conditions enabled reliable prediction and that the experimental paradigm worked as intended. Prediction performance improved systematically as more acoustic information became available across gates, with the critical shift occurring at Gate 5 when the complete second name (*Name2*) became available, demonstrating that listeners required sufficient pitch and final



### Discriminability by Auditory-Perceptual Acuity Level and Gate



**FIGURE 6** Discriminability ( $d'$ ) by auditory-perceptual acuity level and gate position. (A–C) show separate models (each auditory-perceptual acuity modeled independently); (D–F) show the combined model (all auditory-perceptual acuities modeled simultaneously). Points represent median  $d'$  with 95% credible intervals at three auditory-perceptual acuity levels, with darker colors indicating lower auditory-perceptual acuity (minimum, ca.  $-2SD$ ) and lighter colors indicating higher auditory-perceptual acuity (maximum, ca.  $+1SD$ ). Dotted lines connect estimates across gates. Blue = pitch acuity, yellow = pause acuity, red = final lengthening acuity.

lengthening cue information to reliably predict the grouping structure. As visualized in Figure 1, pitch range distributions showed a moderate distinction between *grouped* and *ungrouped* stimulus conditions across both name positions, while final lengthening displayed clear distributional differences only at *Name2*, making this the earliest point at which both prosodic boundary cue types provided reliable boundary information.

With respect to our research question on how auditory-perceptual acuity for prosodic boundary cues relates to predicting syntactic structure from partial prosodic information,

we found that participants with higher auditory-perceptual acuity demonstrated better structural prediction. When tested separately, both pause acuity and final lengthening acuity facilitated structural prediction, whereas pitch acuity did not show clear evidence for a facilitatory effect. However, when all three auditory-perceptual acuity measures were modeled simultaneously, the effects attenuated, that is, each individual auditory-perceptual acuity effect was reduced by approximately fifty percent and none of them retained statistical support. This attenuation indicates that the three auditory-perceptual acuity

measures share considerable predictive variance rather than contributing independently, consistent with their moderate intercorrelations.

Crucially, this facilitation pattern was observed across all gates rather than being particularly pronounced at the early gates where the acoustic information is weakest. This contradicts our hypothesis that auditory-perceptual acuity would provide its strongest advantages when subtle prosodic boundary cue distinctions require fine-grained perceptual resolution, with these advantages diminishing as acoustic evidence accumulates. Instead, better bottom-up auditory-perceptual acuity provides general processing advantages throughout the accumulation of evidence rather than selectively enhancing prediction when prosodic boundary cues are ambiguous.

Response bias analyses confirmed that participants' baseline response preferences were not influenced by auditory-perceptual acuity. Participants with higher auditory-perceptual acuity did not adopt different response strategies nor did they favor one grouping structure over the other. Instead, they predicted the upcoming grouping structure more effectively at all gates based on the available prosodic information. This supports the idea that auditory-perceptual acuities enhance prosodic boundary prediction through perceptual sensitivity rather than decision-level strategies.

Our findings extend prior research on prosodic processing by showing that individual differences in bottom-up auditory-perceptual acuity relate to top-down prediction of syntactic structure during incremental sentence comprehension (e.g., Cole et al., 2010; Ferreira and Karimi, 2015; Wagner and Watson, 2010). While previous studies have emphasized the role of acoustic boundary cues such as pitch movements, silent pauses, and final lengthening in boundary perception (e.g., Cangemi et al., 2015; Ganga et al., 2024; Holzgrefe-Lang et al., 2016; Petrone et al., 2017; Schubö et al., 2023), they have largely focused on group-level effects or prosodic boundary cue interactions without accounting for perceptual variability across listeners. Our findings are novel in that they link isolated auditory-perceptual discrimination thresholds for prosodic boundary cues to predictive processing in a gated paradigm. The shared variance among auditory-perceptual acuities indicates integrated rather than isolated cue-specific mechanisms, which contribute to enhanced top-down boundary prediction. This is consistent with current views on prosody processing emphasizing a more integrated, spectrotemporal framework for prosodic boundary perception (e.g., Brugos and Barnes, 2014; Cohen et al., 1953), where auditory-perceptual acuity may act as a domain-general facilitator of prosodic phrasing. Theoretically, these results are consistent with predictive coding models of speech perception (e.g., Park et al., 2018; Preisig and Meyer, 2025; Sohoglu et al., 2012), which posit that listeners generate top-down expectations about upcoming structure based on bottom-up sensory input. Our results suggest that finer auditory-perceptual resolution may equip listeners to better exploit unfolding prosodic information, thereby supporting the incremental syntactic analysis of the utterance (e.g., Clifton et al., 2002; Frazier et al., 2006), particularly in ambiguous contexts like coordinate structures (e.g., Kentner and Féry, 2013).

Our finding that auditory-perceptual acuity effects are not confined exclusively to the earliest gates, but persist across the

unfolding signal, has implications for the timing of prosodic-syntactic integration during grouping prediction. Classical immediate-integration accounts (e.g., Tanenhaus et al., 1995) predict that prosodic information should influence syntactic parsing as soon as it becomes available, whereas delayed-integration accounts (see Cutler et al., 1997) assume that prosodic cues are initially buffered and only affect syntactic commitments once sufficient evidence has accumulated. Neurophysiological evidence supports rapid sensitivity to prosodic boundary cues: ERP studies show early neural responses to prosodic boundaries, indexed by the Closure Positive Shift (CPS), which can emerge as soon as boundary-related prosodic information becomes available and does not depend on the presence of pauses or later syntactic input (e.g., Holzgrefe-Lang et al., 2016; Steinhauer et al., 1999; see also Holzgrefe et al., 2013). CPS effects have been observed even when boundary perception relies on subtle cue combinations such as pitch change and final lengthening, indicating early neural integration of prosodic structure. Our behavioral findings complement this evidence by showing that auditory-perceptual acuity provides a sustained processing advantage across information accumulation, rather than a sharp early-gate effect. This pattern is most consistent with continuous integration models (e.g., Kuperberg and Jaeger, 2016), in which prosodic and syntactic information are incrementally integrated and individual differences in perceptual precision modulate processing continuously rather than at discrete decision points, in line with predictive-processing accounts emphasizing continuous precision weighting (e.g., Friston, 2010).

Some limitations warrant consideration. First, our perceptual JND AXB threshold tasks capture isolated prosodic boundary cue discrimination but may not reflect how listeners integrate multiple acoustic cues in natural speech. Second, the shared variance among auditory-perceptual acuity measures leaves open whether this reflects domain-general auditory sensitivity or correlated but distinct perceptual abilities; larger samples or orthogonal prosodic boundary cue manipulations are needed to clarify this. Future work should involve larger cohorts and combine behavioral measures with neurophysiologic methods to identify the neural mechanisms underlying prosodic boundary cue processing. Examining the interaction of bottom-up perceptual abilities and top-down prediction in more naturalistic, multi-cue contexts would additionally help to determine whether the shared behavioral variance reflects common neural resources or coordinated but partly independent processes.

## Conclusion

In conclusion, individual differences in auditory-perceptual acuity affect how well listeners exploit prosodic information for structural prediction, providing consistent advantages regardless of the available acoustic information strength. Importantly, listeners do not process prosodic boundary cues independently but rather integrate them incrementally during processing. Our findings also suggest that the individual differences in exploiting bottom-up prosodic information for top-down syntactic prediction observed

by Hansen et al. (2023) likely reflect underlying differences in auditory-perceptual abilities rather than differences in task strategies or cue-specific processing mechanisms.

## Data availability statement

All materials, data, and reproducible analysis code are available through the Open Science Framework: experimental code for the gating task (<https://osf.io/ehu8g/>), experimental code for the JND task (<https://osf.io/mqy2p/>) and analysis scripts with data (<https://osf.io/4yb7j/>).

## Ethics statement

The studies involving humans were approved by University of Potsdam Ethics Committee (approval code: 99/2020). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

AH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. IW: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing, Conceptualization. OT: Funding acquisition, Project administration, Writing – review & editing, Conceptualization. SH: Funding acquisition, Project administration, Writing – review & editing, Conceptualization. JV: Formal analysis, Methodology, Supervision, Validation, Writing – review & editing.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This original research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-Project-ID 317633480-SFB 1287. João Veríssimo has been funded by the Fundação para a Ciência e a Tecnologia (FCT, Foundation for Science and Technology), grant UID/214/2025 to the Center of Linguistics of the University of Lisbon.

## References

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

## Acknowledgments

We thank our research assistants for their help with the data collection, and all of our participants for their time and effort and willingness to be part of this research project.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author IW declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. Assistance from ChatGPT (OpenAI) was used for language editing and formatting consistency. Assistance from Claude (Anthropic) was used for experimental and analysis code formatting and commenting. All content and data analyses were verified by the author(s).

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/flang.2026.1763160/full#supplementary-material>

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). *Parsimonious Mixed Models*. Available online at: <http://arxiv.org/pdf/1506.04967> (Accessed January 16, 2026).

- Boersma, P., and Weenink, D. (1992–2020). *Praat: Doing Phonetics by Computer [Computer Program]*. Available online at: <http://www.praat.org/> (Accessed January 16, 2026).
- Bolker, B. M. (2015). “Linear and generalized linear mixed models,” in *Ecological Statistics*, eds. G. A. Fox, S. Negrete-Yankelevich, and V. J. Sosa (Oxford: Oxford University Press/Oxford Academic), 309–333. doi: 10.1093/acprof:oso/9780199672547.003.0014
- Brugos, A., and Barnes, J. (2014). Effects of dynamic pitch and relative scaling on the perception of duration and prosodic grouping in American English. *Speech Prosody* 2014, 388–392. doi: 10.21437/SpeechProsody.2014-65
- Brunner, J., Ghosh, S. S., Hoole, P., Matthies, M., Tiede, M., and Perkell, J. S. (2011). The influence of auditory acuity on acoustic variability and the use of motor equivalence during adaptation to a perturbation. *J. Speech Lang. Hear. Res.* 54, 727–739. doi: 10.1044/1092-4388(2010/09-0256)
- Buerkner, P.-C. (2018). Advanced bayesian multilevel modeling with the r package brms. *R J.* 10, 395–411. doi: 10.32614/RJ-2018-017
- Byrd, D., and Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J. Phonet.* 31, 149–180. doi: 10.1016/S0095-4470(02)00085-2
- Cangemi, F., Krüger, M., and Grice, M. (2015). “Listener-specific perception of speaker-specific productions in intonation,” in *Individual Differences in Speech Production and Perception* (Frankfurt: Peter Lang), 123–145.
- Cho, T., and Keating, P. (2009). Effects of initial position vs. prominence in English. *J. Phonet.* 37, 466–485. doi: 10.1016/j.wocn.2009.08.001
- Clifton, C., Jr., Carlson, K., and Frazier, L. (2002). Informative prosodic boundaries. *Lang. Speech* 45, 87–114. doi: 10.1177/00238309020450020101
- Cohen, J., Hansel, C. E., and Sylvester, J. D. (1953). A new phenomenon in time judgment. *Nature* 172:901. doi: 10.1038/172901a0
- Cole, J. (2015). Prosody in context: a review. *Lang. Cogn. Neurosci.* 30, 1–31. doi: 10.1080/23273798.2014.963130
- Cole, J., Mo, Y., and Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Lang. Cogn. Process.* 25, 1141–1177. doi: 10.1080/016909609032552507
- Collier, R. P. G., de Pijper, J. R., and Sanderman, A. A. (1993). “Perceived prosodic boundaries and their phonetic correlates,” in *Human Language Technology*, ed. Morgan Kaufmann Publishers, Inc. (San Francisco, CA: Morgan Kaufmann Publishers, Inc.), 341–345.
- Cutler, A., Dahan, D., and van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* 40(Pt 2), 141–201. doi: 10.1177/002383099704000203
- de Beer, C., Hofmann, A., Regenberg, F., Huttenlauch, C., Wartenburger, I., Obrig, H., et al. (2022). Production and comprehension of prosodic boundary marking in persons with unilateral brain lesions. *J. Speech Lang. Hear. Res.* 65, 4774–4796. doi: 10.1044/2022\_JSLHR-22-00258
- Dickey, J. M., and Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a markov chain. *Ann. Math. Stat.* 41, 214–226. doi: 10.1214/aoms/1177697203
- Ferreira, F., and Karimi, H. (2015). Prosody, performance, and cognitive skill: evidence from individual differences. *Explicit Implicit Prosody Sent. Process.* 46, 119–132. doi: 10.1007/978-3-319-12961-7\_7
- Frazier, L., Carlson, K., and Clifton, C. JR. (2006). Prosodic phrasing is central to language comprehension. *Trends Cogn. Sci.* 10, 244–249. doi: 10.1016/j.tics.2006.04.002
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Ganga, R., Geutjes, J., van Niekerk, E., Reshetnikova, V., and Chen, A. (2024). Processing prosodic boundaries in Dutch coordinated constructions. *Speech Prosody* 2024, 985–989. doi: 10.21437/SpeechProsody.2024-199
- Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* 2, 1360–1383. doi: 10.1214/08-AOAS191
- Ghosh, J., Li, Y., and Mitra, R. (2018). On the use of cauchy prior distributions for Bayesian Logistic Regression. *Bayesian Anal.* 13, 359–383. doi: 10.1214/17-BA1051
- González, C., Weissglass, C., and Bates, D. (2022). Creaky voice and prosodic boundaries in Spanish: an acoustic study. *Stud. Hisp. Lusophone Ling.* 15, 33–65. doi: 10.1515/shll-2022-2055
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Percept. Psychophys.* 28, 267–283. doi: 10.3758/BF03204386
- Hansen, M., Huttenlauch, C., de Beer, C., Wartenburger, I., and Hanne, S. (2023). Individual differences in early disambiguation of prosodic grouping. *Lang. Speech* 66, 706–733. doi: 10.1177/00238309221127374
- Hautus, M. J., Macmillan, N. A., and Creelman, C. D. (2021). *Detection Theory*. New York, NY: Routledge. doi: 10.4324/9781003203636
- Hofmann, A., Tuomainen, O., Hanne, S., Verissimo, J., and Wartenburger, I. (submitted). *The Prosodic Perception-Production Link: Impact of Auditory-Perceptual Acuity on Prosodic Cueproduction Under Cognitive Load: Manuscript Under Review*. Available online at: <https://osf.io/brw2t> (Accessed January 16, 2026).
- Holzgrefe, J., Wellmann, C., Petrone, C., Truckenbrodt, H., Höhle, B., and Wartenburger, I. (2013). Brain response to prosodic boundary cues depends on boundary position. *Front. Psychol.* 4:421. doi: 10.3389/fpsyg.2013.00421
- Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Råling, R., Truckenbrodt, H., Höhle, B., et al. (2016). How pitch change and final lengthening cue boundary perception in german: converging evidence from ERPs and prosodic judgements. *Lang. Cogn. Neurosci.* 31, 904–920. doi: 10.1080/23273798.2016.1157195
- Huttenlauch, C., de Beer, C., Hanne, S., and Wartenburger, I. (2021). Production of prosodic cues in coordinate name sequences addressing varying interlocutors. *Lab. Phonol.: J. Assoc. Lab. Phonol.* 12:1. doi: 10.5334/labphon.221
- Jeffreys, H. (1991). *Theory of Probability (2nd Edn., repr)*. Oxford: Oxford University Press.
- Ji, J., Zhao, X., Li, Y., and Yang, X. (2024). Age effects on prosodic boundary perception. *Psychol. Aging* 39, 262–274. doi: 10.1037/pag0000811
- Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795. doi: 10.1080/01621459.1995.10476572
- Kentner, G., and Féry, C. (2013). A new approach to prosodic grouping. *Ling. Rev.* 30, 277–311. doi: 10.1515/trl-2013-0009
- Kjelgaard, M. M., and Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *J. Mem. Lang.* 40, 153–194. doi: 10.1006/jmla.1998.2620
- Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *J. Acoust. Soc. Am.* 118, 1038–1054. doi: 10.1121/1.1923349
- Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375
- Lialiou, M., Grice, M., Röhr, C. T., and Schumacher, P. B. (2024). Auditory processing of intonational rises and falls in German: rises are special in attention orienting. *J. Cogn. Neurosci.* 36, 1099–1122. doi: 10.1162/jocn\_a\_02129
- Lin, H.-Y., and Fon, J. (2010). Perception on pitch reset at discourse boundaries. *Interspeech* 2010, 1225–1228. doi: 10.21437/Interspeech.2010-388
- Männel, C., and Friederici, A. D. (2016). Neural correlates of prosodic boundary perception in German preschoolers: if pause is present, pitch can go. *Brain Res.* 1632, 27–33. doi: 10.1016/j.brainres.2015.12.009
- McElreath, R. (2020). *Statistical Rethinking*. New York, NY: Chapman and Hall/CRC. doi: 10.1201/9780429029608
- Oschkinat, M., Hoole, P., Falk, S., and Dalla Bella, S. (2022). Temporal malleability to auditory feedback perturbation is modulated by rhythmic abilities and auditory acuity. *Front. Hum. Neurosci.* 16:885074. doi: 10.3389/fnhum.2022.885074
- Park, H., Thut, G., and Gross, J. (2018). Predictive entrainment of natural speech through two fronto-motor top-down channels. *Lang. Cogn. Neurosci.* 35, 739–751. doi: 10.1080/23273798.2018.1506589
- Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., and Höhle, B. (2017). Prosodic boundary cues in German: evidence from the production and perception of bracketed lists. *J. Phonet.* 61, 71–92. doi: 10.1016/j.wocn.2017.01.002
- Preisig, B. C., and Meyer, M. (2025). Predictive coding and dimension-selective attention enhance the lateralization of spoken language processing. *Neurosci. Biobehav. Rev.* 172:106111. doi: 10.1016/j.neubiorev.2025.106111
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bull. Rev.* 16, 225–237. doi: 10.3758/PBR.16.2.225
- Schubö, F., Zerbian, S., Hanne, S., and Wartenburger, I. (2023). *Prosodic Boundary Phenomena*. Berlin: Language Science Press. doi: 10.5281/zenodo.7777469
- Smith, D. J., Stepp, C., Guenther, F. H., and Kearney, E. (2020). Contributions of auditory and somatosensory feedback to vocal motor control. *J. Speech Lang. Hear. Res.* 63, 2039–2053. doi: 10.1044/2020.JSLHR-19-00296
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* 32, 8443–8453. doi: 10.1523/JNEUROSCI.5069-11.2012
- Stan Development Team (2020). *RStan: The R Interface to Stan*. Available online at: <https://mc-stan.org/> (Accessed January 16, 2026).
- Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2, 191–196. doi: 10.1038/5757

- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863
- Tyler, M. D., and Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *J. Acoust. Soc. Am.* 126, 367–376. doi: 10.1121/1.3129127
- Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., and Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: a tutorial introduction. *J. Phonet.* 71, 147–161. doi: 10.1016/j.wocn.2018.07.008
- Veríssimo, J. (2023). When fixed and random effects mismatch: another case of inflation of evidence in non-maximal models. *Comput. Brain Behav.* 6, 84–101. doi: 10.1007/s42113-022-00152-3
- Veríssimo, J. (2025). A gentle introduction to Bayesian statistics, with applications to bilingualism research. *Linguistic, Approach, Biling.* 15, 453–486. doi: 10.1075/lab.24027.ver
- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hypothesis testing for psychologists: a tutorial on the savage-dickey method. *Cogn. Psychol.* 60, 158–189. doi: 10.1016/j.cogpsych.2009.12.001
- Wagner, M., and Watson, D. G. (2010). Experimental and theoretical advances in prosody: a review. *Lang. Cogn. Process.* 25, 905–945. doi: 10.1080/01690961003589492
- Wurm, L. H., and Fisičaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *J. Mem. Lang.* 72, 37–48. doi: 10.1016/j.jml.2013.12.003
- Yang, X., Shen, X., Li, W., and Yang, Y. (2014). How listeners weight acoustic cues to intonational phrase boundaries. *PLoS ONE* 9:e102166. doi: 10.1371/journal.pone.0102166
- Zloteanu, M., and Vuorre, M. (2024). A tutorial for deception detection analysis or: how I learned to stop aggregating veracity judgments and embraced signal detection theory mixed models. *J. Nonverbal Behav.* 48, 161–185. doi: 10.1007/s10919-024-00456-x