Language-Specific or Universal? The Nature and Roles of Consistency and Gradiency in Speech Perception

Brian W. L. WONG[a,b], Arthur G. SAMUEL[a,c,d], and Efthymia C. KAPNOULA[a,d]

[a] BCBL, Basque Center on Brain, Language and Cognition, Donostia-San Sebastian, Spain

[b] University of the Basque Country (UPV-EHU), Spain

[c] Department of Psychology, Stony Brook University, New York, U.S.A.

[d] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

Word count: 12,461

Author Note

Brian W. L. WONG https://orcid.org/0000-0002-0519-6116

Arthur G. SAMUEL https://orcid.org/0000-0001-8552-2710

Efthymia C. KAPNOULA https://orcid.org/0000-0001-6640-1948

Correspondence concerning this article should be addressed to Brian W. L. Wong, BCBL, Mikeletegi Pasealekua, 69, San Sebastián, Spain. Email: bwong@bcbl.eu

Author Note

Acknowledgements

The present study is part of the first author's PhD dissertation at BCBL. We would like to thank Brendan Costello and James Magnuson for their constructive feedback on the manuscript. We would also like to acknowledge Amets Esnal and Itziar Basterra for organizing the data collection, Ainhoa Eguiguren for the translations of the task instructions, and Daphne Weiss and Elena Alguirrebengoa for recording the stimuli for the present study. Additionally, we thank Candice Frances and Tanja Roembke for developing the Corsi block-tapping task and spatial Stroop task used in the present study.

Declaration of conflicting interests

The Authors declare that there is no conflict of interest.

Funding

This research is supported by the Basque Government through the BERC 2022-2025 program and funded by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010/AEI/10.13039/501100011033. The research is also supported by grant ref. PRE2021-097223; project CEX2020-001010-S-21-2 funded by MCIN/AEI/10.13039/501100011033 and FSE+" awarded to Brian W. L. Wong. Support for this project was provided by the Spanish Ministry of Science and Innovation through Grant # PID2020-113348GB-I00, awarded to Arthur Samuel and Efthymia Kapnoula.

Previous Dissemination of Data

Parts of the data were presented at the 64th Annual Meeting of the Psychonomic Society in San Francisco, California, U.S.A., from November 16 to 19, 2023.

Preprint: https://osf.io/preprints/psyarxiv/t5zbc

CRediT taxonomy

Conceptualization: Brian W. L. Wong, Arthur G. Samuel, Efthymia C. Kapnoula; Methodology: B. W. L. W., A. G. S., E. C. K.; Investigation: B. W. L. W.; Software: B. W. L. W., E. C. K; Formal analysis: B. W. L. W., E. C. K; Writing - original draft preparation: B. W. L. W.; Writing - review and editing: B. W. L. W., A. G. S., E. C. K; Supervision: E. C. K, A. G. S.

Abstract

Speech perception gradiency allows listeners to detect and maintain subphonemic information, thereby enhancing speech perception flexibility. However, its nature is not fully understood; it is unclear whether gradiency is a generic trait, or if it depends on language status (L1 vs. L2) and/or language-specific properties (e.g., voice onset time). To address these questions, we investigated the functions of gradiency in spoken word recognition by Spanish (L1)-English (L2) bilinguals. In addition, we examined the role of perceptual consistency, a measure reflecting the stability of auditory cue encoding. Gradiency and perceptual consistency were assessed using the Visual Analogue Scale (VAS). Spoken word recognition was evaluated through initial lexical activation and speech perception flexibility, quantified by the likelihood and speed of recovery from misleading information, using an eye-tracking Visual World Paradigm (VWP) task. Seventy Spanish-English bilinguals performed these tasks in both Spanish and English. Results revealed that listeners with higher gradiency recovered faster from misleading information in L1 Spanish and those with higher perceptual consistency relied more on early information to activate lexical candidates in both languages. This study highlights the language-specific benefit of gradiency in speech perception flexibility and provides general evidence that speech processing stability facilitates initial lexical activation.

Keywords: speech perception, gradiency, perceptual consistency, categorical perception, individual differences

Public significance statements: Accurate speech perception is challenging, but some people perform it better than others. This study tested Spanish-English bilinguals on some native and non-native speech perception tasks and found that (1) individuals who excel at noticing subtle differences in sounds can quickly correct their misunderstandings after being misled in their native language, (2) those who interpret sounds consistently tend to rely more on early information for word activation in both their native and non-native languages, and (3) noticing subtle differences in sounds is a skill that is specific to each language, while interpreting sounds consistently is a more general skill. These findings are important for enhancing our understanding of spoken word processing and could potentially improve language learning strategies.

Introduction

When processing spoken language, listeners have to map acoustic cues onto speech categories. An example of an acoustic cue is Voice Onset Time (VOT), which is the delay between the release of a stop consonant and the beginning of vocal cord vibration for the subsequent vowel. VOT serves as the primary cue enabling listeners to differentiate between voiced and voiceless stop consonants in English (e.g., /b/ and /p/; Lisker, 1986). In most cases, VOTs close to 0 ms map to a /b/ sound and those close to 30 ms or above would reflect a /p/ sound. However, because cue values depend on talker identity (Allen & Miller, 2004), speaking rate (Miller et al., 1986), etc., the same cue value does not always map onto the same category – this is known as the *lack of invariance* problem.

Early views of speech perception suggested that listeners are able to overcome this problem by largely ignoring within-category information. This idea is based on the empirical phenomenon known as "categorical perception", i.e., the finding that two speech sounds from the same phoneme category are often less distinguishable compared to an equidistant pair of sounds that map onto two different phonemes (Liberman et al., 1961; Repp, 1984; M. E. H. Schouten & van Hessen, 1992). This well-replicated finding lent support to the idea that during speech perception within-category differences are largely ignored, and speech sounds are perceived categorically. For example, an English /b/ sound with a VOT of 0 ms should be essentially indistinguishable from one with a VOT of 20 ms, whereas a /b/ with VOT of 10 ms and a /p/ with VOT of 30 ms should be more easily distinguished.

Despite the early popularity of the categorical perception account, advances in psycholinguistics have since challenged this perspective by providing evidence that listeners are sensitive to within-category information, thus suggesting that speech categorization is fundamentally *gradient* (see McMurray, 2022, for a review). According to the gradient view of speech perception, listeners are capable of detecting subtle, continuous variations in phonetic cues, rather than perceiving speech sounds in terms of distinct phonemic categories. Supporting evidence for gradiency emerges from a variety of methodologies, including various behavioral tasks (e.g., Carney et al., 1977; Kapnoula et al., 2017; Massaro & Cohen, 1983; Miller, 1994; Pisoni & Lazarus, 1974; Pisoni & Tash, 1974; Samuel, 1977, 1982), eye-tracking (e.g., McMurray et al., 2002, 2008), and EEG (e.g., Kapnoula & McMurray, 2021; Ou & Yu, 2022; Toscano et al., 2010).

Recent work on this topic has focused on individual differences as a means to examine the underlying mechanisms and functional roles of gradiency in spoken language comprehension. Building on that work, the present study further examines the mechanistic nature of speech perception gradiency by asking whether its function depends on language status (L1 vs. L2) and/or language-specific properties (i.e., VOT).

Measuring Individual Differences in Gradiency: The Case for the VAS Task

While perception of speech sounds is generally gradient, some individuals show higher gradiency than others (Kapnoula et al., 2017; Kapnoula & McMurray, 2021). These individual differences are thought to be due to variations in the degree of category-driven perceptual warping around the boundary (Kapnoula et al., 2021). These individual differences can be captured using the Visual Analogue Scale (VAS) task, initially introduced by Massaro and Cohen (1983). In this task, participants are presented with a continuum labelled, for example, with “ba” at one end and “pa” at the opposite end. Upon hearing a sound, participants are asked to indicate its position along this continuum. Gradiency measures extracted from VAS responses have illuminated individual differences in speech perception: some individuals are highly attuned to differences within phoneme categories, yielding more gradient responses characterized by a shallow slope (hereafter referred to as *more gradient listeners*). Conversely, others exhibit diminished sensitivity to such fine-grained distinctions, as evidenced by responses with a steeper slope (hereafter *less gradient* *listeners*; Kapnoula et al., 2021; Kapnoula & McMurray, 2021; Kapnoula et al., 2017; Kong & Edwards, 2016).

Importantly, Kapnoula and McMurray (2021) aimed at identifying the exact aspect of speech perception that is reflected in VAS responses. To do so, they looked at the N1 (an event-related potential approximately 100 ms after stimulus onset that reflects encoding of continuous acoustic information). The results revealed a linear relationship between VOT and N1 amplitude; however, this relationship was disrupted close to the category boundary, but only in the case of listeners who exhibited a more categorical (or less gradient) pattern of VAS responses. This finding suggests that gradiency measured by VAS likely reflects the initial phases of speech cue encoding (Kapnoula & McMurray, 2021).

Thus, relative to other tasks, like the two-alternative forced choice (2AFC) task, the VAS task likely offers a more nuanced perspective on auditory perception, capturing the subtleties of how individuals discern sound variations. Previous research has critically examined the use of 2AFC tasks (Apfelbaum et al., 2022; Hary & Massaro, 1982; McMurray et al., 2008; Munson et al., 2017; Pisoni & Tash, 1974). Such tasks require a binary response (e.g., “ba” or “pa”), potentially oversimplifying the listeners’ perceptual experience and overlooking their capacity to perceive within-category nuances. In fact, a recent study has shown that while VAS slopes are indicative of gradiency, 2AFC slopes primarily capture perceptual consistency, or the stability of acoustic cue encoding (Honda et al., 2024; details of perceptual consistency will be described below). This distinction underscores the advantage of VAS in accurately assessing gradiency[1].

Recent work has provided evidence supporting the reliability and validity of the VAS task. Gradiency, as measured by the VAS task, exhibits stability across repeated sessions

[1] Apart from 2AFC, the discrimination task is another method used to demonstrate categorical perception (Studdert-Kennedy et al., 1970). Nonetheless, it also has issues, such as the influence of working memory on results and the possibility that the task tests category rather than encoding level (see Apfelbaum et al., 2022; Gerrits & Schouten, 2004; McMurray, 2022; B. Schouten et al., 2003, for relevant discussion).

when the same stimuli are used (Kong & Edwards, 2016), which speaks to its test-retest reliability (see also Honda et al., 2024 for a relevant discussion). Furthermore, it has been shown that VAS responses are not indicative of general scale usage biases. This is partly evidenced by the minimal correlation in responses between auditory and visual VAS tasks (Kapnoula et al., 2021; Kapnoula & McMurray, 2021; Kapnoula & Samuel, 2024). VAS gradiency measures corresponding to different speech contrasts are correlated with each other (Bidelman et al., 2024; Fuhrmeister & Myers, 2021; Fuhrmeister et al., 2023; Kong & Kang, 2023), but crucially the correlation strength between slopes seems to depend on the acoustic similarity between the speech continua (e.g., higher correlation between labial and alveolar stops compared to labial stops and fricatives; Kapnoula et al., 2017, 2021; Kapnoula and McMurray, 2021). Taken together, these results are in line with the idea that VAS responses tap something fundamental about speech categorization. Finally, and perhaps most importantly, VAS gradiency aligns with corresponding measures extracted from the same participants using different methodologies. Specifically, it has been found that VAS slope is related to lexical gradiency as measured via an eye-tracking Visual World Paradigm (VWP) (discussed below) and to speech category gradiency as measured by the P3 ERP component, a pattern that further underscores the validity of VAS as a measure of speech perception (Kapnoula & McMurray, 2021). Collectively, these findings speak to the reliability and validity of the VAS task as a way of assessing listeners' speech perception gradiency in specific contrasts.

The Functional Role(s) of Gradiency in L1 and L2 Speech Perception

Higher gradiency is linked to a superior ability to process and integrate a range of acoustic cues (Kapnoula et al., 2017; Kapnoula & McMurray, 2021; D. Kim et al., 2020; Kong & Edwards, 2016; Kong & Kang, 2023; Ou et al., 2021), indicating that individuals with higher gradiency are more attuned to the fine-grained details within the acoustic signal.

Furthermore, there is some preliminary evidence that L1 gradiency is linked to L2 proficiency, as measured by a vocabulary task (Kapnoula & Samuel, 2024). This finding points to a potentially significant role of gradiency in L2 acquisition, likely by aiding the establishment of new categories or preserving the ability to discern within-category differences for non-native contrasts.

At its core, gradiency is associated with the ability to discern small cue differences, fostering enhanced flexibility amid uncertain auditory scenarios (Brown-Schmidt & Toscano, 2017; Clayards et al., 2008). One method to assess this kind of flexibility is by examining how participants with different degrees of gradiency cope when they encounter misleading or ambiguous auditory cues. For example, McMurray et al. (2009) presented participants with *lexical garden-paths* (items that sound like one word initially that ultimately turn out to be another word) and used the Visual World Paradigm (VWP; Tanenhaus et al., 1995) to track the activation of the two competing words. This allowed them to investigate how individuals dealt with the induced misunderstandings in real time. In their study, participants were exposed to auditory stimuli such as "þeachball", with the initial phoneme transitioning between /b/ and /p/ along a continuum (/þ/ represents a sound between /b/ and /p/). Notably, with a VOT of 40 ms (e.g., "peachball"), listeners were prone to first look at the picture of "peachpit" (competitor). The rate (i.e., likelihood) of making this garden-path reflects erroneous *initial lexical activation*. In this situation, since the subsequent input is consistent with "beachball" rather than "peachpit", listeners may *recover* from the garden-path and recognize "beachball" (target). The rate (i.e., likelihood) and latency (i.e., speed) of this recovery reflect *speech perception flexibility*. Crucially, both the garden-path rate and the recovery speed were linearly related to the acoustic distance between target and stimulus (i.e., more competitor-like initial sounds induced more garden-paths and slower recoveries). Thus, the findings by McMurray et al. (2009) support the notion that listeners generally exhibit gradiency, which helps them to recover from initial errors generated by ambiguous/misleading linguistic inputs.

Combining this paradigm with VAS, Kapnoula et al. (2021) investigated the relationship between gradiency and flexibility in dealing with garden-path situations. The results showed that, when faced with an auditory stimulus like “þeachball”, more gradient listeners were more likely to recover from erroneous initial lexical activations (“peachpit”), and reach the correct interpretation (“beachball”), especially when the acoustic distance from the target was high. This finding is in line with the idea that more gradient listeners are adept at considering multiple interpretations simultaneously, thereby avoiding early commitment to an incorrect lexical choice, which in turn facilitates recovery from misunderstandings. In contrast, less gradient listeners may show warping in the acoustic cue space around the category boundary, which prevents lexical-level processes from fully recovering. Essentially, the sensitivity of more gradient listeners to the nuances of speech sounds equips them with the flexibility to adjust their interpretations in the face of ambiguous or misleading information. Such flexibility is advantageous for managing the variability inherent in different linguistic environments, since successful comprehension requires continuous adjustment to variations in speech caused by differences in characteristics such as coarticulation and accent.

While Kapnoula et al. (2021) provided evidence for a positive relationship between gradiency and speech perception flexibility in monolingual English speakers, it remains unknown whether this relationship represents a universal pattern or one tied to language-specific acoustic characteristics (VOT; discussed below) or language status (i.e., a person’s L1 vs. their L2). The present study will directly address this question of whether gradiency is a generic trait affecting how listeners process speech.

Perceptual Consistency and its Functional Role(s) in Speech Perception

In addition to gradiency, another metric that can be extracted from the VAS task is the consistency in listeners' responses to identical stimuli, which is taken to reflect the stability of phonetic encoding abilities (Fuhrmeister et al., 2023). While this metric has been referred to by different names in previous studies (e.g., categorization consistency: H. Kim et al., 2024; consistency: Honda et al., 2024; noise: Bidelman et al., 2024; Kapnoula et al., 2017; Kutlu et al., 2024; Rizzi & Bidelman, 2024; Sorensen et al., 2024; response consistency: Fuhrmeister & Myers, 2021; Fuhrmeister et al., 2023; Kapnoula & Samuel, 2024; response variability: Apfelbaum et al., 2022; H. Kim et al., 2024; Kutlu et al., 2022), in the present study, we will use the term *perceptual consistency*[2].

Like gradiency, substantial individual differences have also been observed in perceptual consistency (Fuhrmeister & Myers, 2021; Fuhrmeister et al., 2023; Honda et al., 2024; Kapnoula et al., 2017). In addition, a recent study has found a positive correlation in perceptual consistency between the 2AFC and VAS tasks, suggesting that consistency is a general ability that transcends specific task contexts (Honda et al., 2024). This is further supported by neuroscientific evidence indicating that the level of gyrification in the bilateral transverse temporal gyri is inversely correlated with listeners' perceptual consistency, hinting at relatively stable individual differences (Fuhrmeister & Myers, 2021).

Regarding the relationship between VAS gradiency and perceptual consistency, mixed results have been observed: some studies indicate a negative correlation (e.g., Fuhrmeister et al., 2023), others suggest a positive link (e.g., Honda et al., 2024), and a few report no significant relationship (e.g., Kapnoula et al., 2017). In fact, as pointed out by Apfelbaum et al. (2022), the two constructs are theoretically orthogonal to each other, as

[2] The rationale behind this decision is that we wish to avoid using a term like "response consistency" that refers to the extracted measure and rather focus on the underlying construct of interest. The specific term "perceptual consistency" is based on the working hypothesis that VAS responses reflect listeners' perception, as supported by previous results (e.g., Kapnoula & McMurray, 2021).

individuals may exhibit high or low perceptual consistency regardless of their gradiency level, highlighting the importance of distinguishing between these two constructs. For example, a shallow slope on the VAS could reflect continuous and consistent responses across stimuli with different VOT steps (e.g., from /b/ to /p/), but it might also result from inconsistent responses in each step. In the latter case, the shallow slope would not accurately represent true gradiency, but rather noisy encoding (Honda et al., 2024; Kutlu et al., 2022; Sorensen et al., 2024; for a discussion on this, see Apfelbaum et al., 2022). A method to differentiate the two constructs will be discussed in the Method section.

Unlike VAS gradiency, perceptual consistency appears to tap something similar to more traditional measures of speech perception such the slope extracted from the 2AFC tasks (Honda et al., 2024). In the 2AFC tasks, steeper categorization slope is usually taken to indicate better/sharper categorization, whereas shallow slopes indicate atypical/impaired speech perception (Godfrey et al., 1981; Joanisse et al., 2000; López-Zamora et al., 2012; Serniclaes et al., 2001, 2005; Werker & Tees, 1987). Therefore, it is not surprising that higher perceptual consistency has been linked to better language and reading abilities in children (H. Kim et al., 2024) and improved learning of non-native contrasts in adults (Fuhrmeister et al., 2023; Honda et al., 2024). These findings suggest that cue encoding stability may play an important role in individual differences in both native and non-native speech perception and language learning in general. Given the emerging nature of research in the area of perceptual consistency, its impact on different aspects of speech perception, particularly the ability to recover from misleading and ambiguous auditory information, remains largely unexplored. This gap in knowledge is addressed in the present study.

The Present Study

Our *first research objective* is to examine whether the functional roles of gradiency in speech perception flexibility are modulated by language characteristics and/or language status. The main manipulation in the present experiments involves initial voicing contrasts in Spanish and English, leveraging on the VOT difference between /b/ and /p/ in these two languages. In Spanish, the typical VOTs for these two consonants are /b/ ≈ -80 ms and /p/ ≈ 16 ms (Souganidis et al., 2022), in sharp contrast to their English counterparts (/b/ ≈ 0 ms; /p/ ≈ 60 ms; Lisker & Abramson, 1964). This discrepancy indicates that, in Spanish, the /b/ sound is prevoiced, resulting in a negative VOT, while the Spanish /p/ aligns more closely with the English /b/. The different characteristics between Spanish and English provide an opportunity for us to examine whether the functional roles of gradiency are tied to language-specific acoustic properties of a speech contrast. In addition to language characteristics, it is important to understand whether the function of gradiency in spoken perception flexibility depends on language status. Hence, the present study examines how gradiency influences speech perception flexibility not only in L1 (Spanish) but also in L2 (English). L2 gradiency has not been extensively examined, with only a few studies making inroads into this domain (Kong & Kang, 2023; Kutlu et al., 2022). L2 speech perception, especially in noisy settings, necessitates the correction of errors induced by misleading or ambiguous cues, suggesting that gradient listening would be particularly valuable in these situations. More importantly, understanding the possible roles of gradiency can shed light on whether the function of gradiency is generic or if it depends on language status (i.e., L1 vs. L2). In sum, our first research question is whether the relationship between gradiency and speech perception flexibility is modulated by language characteristics and/or language status. To test this, we will examine the relationship between gradiency and speech perception flexibility in L1 Spanish and L2 English.

Regarding the first research question, we hypothesized that gradiency would facilitate recovery from lexical garden-paths in L1 Spanish, as seen in L1 English (Kapnoula et al., 2021). This hypothesis rests on the premise that greater within-category sensitivity would allow listeners to achieve more effective lexical recovery, despite the different VOTs between Spanish and English. However, it is possible that the relationship may be less obvious in Spanish because native Spanish speakers (our current participants) may show more categorical responses in the Spanish VAS. In Spanish, the contrast between /b/ and /p/ is based on the presence or absence of pre-voicing (i.e., it is more qualitative), whereas in English, it is more quantitative. Therefore, Spanish speakers may focus less on within-category differences (cf. Kapnoula & Samuel, 2024). Regarding language status, we hypothesized that, assuming participants have accurate categorical representations of English stop consonants, we will find a positive relationship between gradiency and speech perception flexibility.

Our *second research objective* is to examine the relationships between perceptual consistency and spoken word recognition, including initial lexical activation and speech perception flexibility, and whether these relationships are modulated by language. We predicted that listeners who exhibit higher stability in encoding speech sounds would have higher speech perception flexibility. This prediction is based on previous findings regarding the advantages of perceptual consistency in speech perception (Fuhrmeister et al., 2023; Honda et al., 2024). We consider this research question to be exploratory due to current gaps in our understanding of the role of perceptual consistency in speech perception. It is important to examine perceptual consistency to understand what this underexamined construct reflects and its advantages in speech perception.

Finally, our *third research objective* is to test whether gradiency and perceptual consistency are language-specific or language-general by directly comparing the measures collected in the two languages. In the present study, we extracted both measures from the same contrast (/b/-/p/), but in two different languages. This allows us to directly ask the question of whether gradiency and perceptual consistency are language-specific or not. To our knowledge, this is the first attempt to compare both gradiency and consistency with the same phonemes across languages. This comparison could shed light on the mechanisms underlying these two constructs.

The correlation between gradiency measures extracted from different contrasts has been found to depend on the acoustic similarity between these contrasts (Kapnoula et al., 2017, 2021; Kapnoula & McMurray, 2021). This suggests that gradiency is, to some degree, contrast-specific. However, previous work, albeit limited, has consistently found positive correlations between different measures of perceptual consistency extracted from the 2AFC and VAS (Honda et al., 2024), as well as between different phonetic contrasts (Fuhrmeister et al., 2023). Based on these findings, we predicted that gradiency is tied to language-specific acoustic characteristics, while perceptual consistency is a general speech perception trait.

To summarize, this study adopts an individual differences approach to examine the nature and roles of speech perception gradiency and perceptual consistency. Our three research questions include: (1) Is the relationship between gradiency and spoken word recognition modulated by language (language characteristics and/or language status)? (2) Is the relationship between perceptual consistency and spoken word recognition modulated by language? (3) Are gradiency and perceptual consistency stable across languages?

To address these three research questions, we tested native Spanish speakers in both their native language (Spanish) and their L2 (English[3]). Additionally, we included English proficiency, English exposure, working memory, inhibitory control, and musical training as covariates due to their potential links to language processing or gradiency found in previous studies (e.g., Kapnoula et al., 2017; Kapnoula & McMurray, 2021; Kapnoula & Samuel, 2024; D. Kim et al., 2020; McMurray et al., 2018; Smayda et al., 2015).

[3] In this study, English is referred to as L2, while acknowledging that for many participants, English may serve as their second, third, or even subsequent language.

Method

Participants

We recruited 70 native Spanish speakers (53 females)[4] residing in San Sebastian, Spain for this study. This sample size was informed by a previous study conducted by Kapnoula et al. (2021), in which they tested 67 participants. A power analysis (Supplementary Material I) indicates that our sample size is sufficiently large to detect the anticipated effect, if it exists.

Participants were between 18 and 40 years old (*M* = 27.8) and had normal or corrected-to-normal vision and no known hearing or neurological impairments. Apart from English, most participants were also familiar with Basque, which was taken into account in the preparation of the materials and data analyses. Detailed demographic information is presented in Table S1. Given the diverse characteristics of our participants (see Table S1), we believe that our results are relatively generalizable. The experiment was approved by the BCBL Ethics Review Board, adhering to the guidelines of the Helsinki Declaration. All participants provided written informed consent and were paid for their participation.

Overview of Design

In each language, VWP stimuli consisted word pairs with similar onsets, differing only in the voicing of the first consonant, and different offsets (e.g., "beachball" [bitʃbɔl], "peachpit" [pitʃpɪt]). Words were manipulated to create /b/-to-/p/ voicing continua (e.g., "beachball"-to-"peachball") with the goal of progressively increasing the probability that participants would first activate the competitor word (reflecting initial lexical activation) and evaluate if and how quickly they would recover after hearing the disambiguating offset

[4] All participants except two had a Spanish Age of Acquisition (AoA) of 0. Given that these two participants reported Spanish as their dominant language and had higher Spanish proficiency than Basque or English proficiency, they were included in the analyses.

(reflecting speech perception flexibility). These measures were examined using a VWP task with eye-tracking (as in Kapnoula et al., 2021 and McMurray et al., 2009). Apart from the VOT step of the stimulus onset, the two independent variables (IVs) of interest were the participant's (1) VAS slope (reflecting speech perception gradiency) and their (2) VAS response consistency (reflecting perceptual consistency). The VAS and VWP tasks were administered in English (L2) with the same stimuli and design as in Kapnoula et al. (2021); both tasks were also conducted in Spanish (L1) with Spanish stimuli in a separate session (order of sessions was counterbalanced, see "Order of Tasks" for details).

A set of additional measures were included as covariates. Specifically, participants completed an English picture naming task (assessing English proficiency), a questionnaire on language exposure and musical training, a spatial Stroop task (assessing inhibitory control), and a Corsi block-tapping task (assessing working memory). Our results showed that including these variables did not improve the model fit related to our research questions (see "Primary Analyses I: Effects of Gradiency on Initial Lexical Activation and Speech Perception Flexibility in L1 Spanish and L2 English" section for details). Therefore, the details of these tasks are not further discussed in the manuscript but are provided in Supplementary Material III. In addition, a training session was scheduled one to seven days before the English VWP to ensure participants were familiar with the English words (see Supplementary Material III for details). The tasks are summarized in Table 1.

Table 1

Summary of Tasks and Corresponding Independent and Dependent Variables

Task	Measuring	Duration (min)
Independent variables		
VAS	Speech perception gradiency and perceptual consistency	5-8 for each language
Spatial Stroop task	Inhibitory control	5
Corsi Block-tapping task	Working memory	5
English picture naming task	English proficiency	5
Questionnaire	Language exposure and musical training	3
Dependent variable		
VWP	Initial lexical activation and speech perception flexibility	52 for each language

Note. VAS = Visual Analogue Scale; VWP = Visual World Paradigm.

Measuring Gradiency and Perceptual Consistency: VAS Tasks

Stimuli

The English stimuli were the same as those used in Kapnoula et al. (2017). Specifically, we used a “buh”-“puh” continuum consisting of natural speech items that varied factorially along (a) seven VOT steps, ranging from 1 to 45 ms, and (b) five fundamental frequency (F0) steps, ranging from 90 to 125 Hz. Similarly, the Spanish stimuli formed a natural speech “ba”-“pa” continuum that had seven VOT steps from -35 to +10 ms and five

F0 steps from 179 to 193 Hz. In both cases, continua were generated employing the progressive cutback and replacement approach. This involved progressively deleting the onset of a word with the /b/ sound and replacing it with a roughly equivalent amount of the /p/ onset of its counterpart (Andruski et al., 1994). For the Spanish items, we used the Praat script developed by Winn (2020; Version 33), which was modified to account for the presence of negative VOT in Spanish voiced sounds.

Procedure

Experiment Builder (version 2022.2.5; SR Research Ltd., 2022) was used to program and run the experiment. Participants completed the VAS task in two languages (English and Spanish) and were informed about the language of the task before it started. Each trial presented participants with a line labelled at both ends according to the two speech sound categories. For English, “buh” was consistently positioned on the left, and “puh” on the right. For Spanish, “ba” and “pa” were used in the same respective positions. Participants listened to each stimulus and clicked on the line to indicate where they perceived it to fall on the continuum. After their first click, a rectangular bar appeared at the point where they had clicked, and participants had the option to modify their response or press the space bar to confirm it. Within each language, stimuli were presented in random order. Each stimulus was presented three times, resulting in 105 trials for each language. The VAS task lasted approximately five to eight minutes.

Measuring Gradiency and Perceptual Consistency

The analysis procedure for the VAS task closely adhered to previous methods (e.g., Kapnoula et al., 2017, 2021). Click locations on the x-axis were converted from pixels to a VAS rating ranging from 0 to 100. The degree of gradiency was quantified by fitting a rotated logistic function (see Supplementary Material IV for more details), the slope of which indicated the gradiency of responses, with shallower slopes reflecting more gradient responses. This approach, unlike standard logistic regression, provides orthogonal measures of gradiency and secondary cue use (i.e., use of F0).

The equation was fitted to each participant's VAS responses implemented in MATLAB (version R2022a, the MathWorks Inc., R2022a) that minimized the least squared error (free software available at McMurray, 2017). For the Spanish VAS, two participants were excluded due to problematic fits. For the English VAS, one participant was excluded for exhibiting a flat response curve (i.e., identical responses, regardless of stimulus variation from /b/ to /p/), one because of a reversed response curve (i.e., a decrease in the number of /p/ responses as stimuli changed from /b/ to /p/), and one due to problematic fit. The remaining fittings were good, with an average R^2 of .941 and .839 for the Spanish and English tasks, respectively.

In addition, we utilized the VAS ratings to extract a measure of perceptual consistency, employing the same procedure outlined in Kapnoula et al. (2017). First, we computed the residuals, or the difference between the VAS rating on each trial and the predicted value for that stimulus based on that participant's fitted curve. Subsequently, we calculated the standard deviation of these residuals for each participant. A larger standard deviation indicates lower consistency in participants' responses and, thus, lower perceptual consistency. To simplify interpretation, we reversed the standard deviation by multiplying by -1, so that higher values represent higher perceptual consistency.

In this study, our goal was to investigate gradiency and perceptual consistency as separate factors. As explained in the introduction, a shallow VAS slope may reflect either true gradiency or noisy encoding. Using Pearson correlations, we found nonsignificant correlations between perceptual consistency and slope (log) for both the Spanish VAS, *r*(66) = -.12, *p* = .329, and for the English VAS, *r*(65) = -.08, *p* = .536. However, we cannot exclude the possibility that some participants might respond inconsistently despite exhibiting a gradient pattern. Therefore, we adjusted for the effect of perceptual consistency on the slope to derive a pure gradiency measure. To accomplish this, we conducted a linear regression with VAS slope (log) as the dependent variable and perceptual consistency as the predictor and extracted the standardized residual. This residual was used as the gradiency measure (hereafter referred to as *slope*) in the subsequent analyses. A higher slope corresponds to lower gradiency.

Measuring Initial Lexical Activation and Speech Perception Flexibility: VWP Tasks

Design and Materials

To assess participants' degree of initial lexical activation and speech perception flexibility, we employed a VWP task (Kapnoula et al., 2021; McMurray et al., 2009) designed to induce lexical garden-paths. This task presented participants with auditory stimuli based on word pairs such as "beachball"-"peachpit". Each pair was characterized by differing initial voicing (/b/ vs. /p/), but shared two to five middle phonemes. Instances of initial VOT ambiguity were disambiguated by phonetic information at the word offset (e.g., the "–ball" or "–pit"). The English words and pictures utilized, including five pairs of critical stimuli (starting with /b/ and /p/) and five pairs of fillers (starting with /l/ and /r/), were the same as those developed by Kapnoula et al. (2021) (see Table 2 for details on all critical English stimuli and Table S2 for all filler stimuli). Each pair underwent VOT manipulation along a seven-step continuum, resulting in a word-to-nonword spectrum (e.g., "beachball" [bi̱tʃbɔl] to "peachball" [pi̱tʃbɔl]). The following descriptions will focus on the Spanish VWP (the design and materials of the English task can be found in Kapnoula et al. (2021)).

Table 2

Critical Stimuli used in the English VWP with International Phonetic Alphabet (IPA), as in Kapnoula et al. (2021)

Set	Voiced word		Voiceless word		Overlapping phonemes
	Spelling	IPA	Spelling	IPA	
1	bumpercar	bʌmpər**kɑr**	pumpernickel	pʌmpər**nɪkəl**	5
2	barricade	bærək**eɪd**	parakeet	pærək**it**	4
3	blanket	blæŋk**ɪt**	plankton	plæŋk**tən**	4
4	beachball	bitʃ**bɔl**	peach-pit	pitʃ**pɪt**	2
5	billboard	bɪlb**ɔrd**	pillbox	pɪlb**ɒks**	3

Note. Underlined portions mark phonemic overlap between the words in a pair; bolded portions mark offsets.

The Spanish stimuli were generated in accordance with the procedures outlined by Kapnoula et al. (2021). We developed a set of twenty Spanish words, including critical stimuli consisting of five /b/-onset and five /p/-onset words. Additionally, filler words were incorporated, featuring five /l/-onset and five /r/-onset words. The characteristics of these stimuli, along with corresponding pictures, mirrored those of the English stimuli (see Table 3 for details on all critical Spanish stimuli and Table S3 for all filler stimuli). Each word encompassed five to eight phonemes and two or three syllables, with three phonemes overlapping for each /b/-/p/ word pair.

Table 3

Critical Stimuli used in the Spanish VWP with IPA and Meanings

Set	Voiced word			Voiceless word			Overlapping phonemes
	Spelling	IPA	Meaning	Spelling	IPA	Meaning	
1	balanza	bala**nθa**	scale	palacio	pala**θjo**	palace	3
2	bañar	baɲa**ɾ**	to bathe	pañal	paɲa**l**	diaper	3
3	baraja	baɾa**xa**	deck of cards	paraguas	paɾa**ɣwas**	umbrella	3
4	vaquero	bake**ɾo**	cowboy	paquete	pake**te**	parcel	3
5	vestido	best**iðo**	dress	pestañas	pest**aɲas**	eyelash	3

Note. Underlined portions mark phonemic overlap between the words in a pair; bolded portions mark offsets.

The methods for processing the Spanish stimuli closely followed the procedures outlined by Kapnoula et al. (2021). The stimuli were created through splicing natural recordings. Initially, a native Spanish speaker recorded complete exemplars of both items in each pair, including both voiced and voiceless onsets (e.g., “balanza”, “palanza”, “balacio”, and “palacio”), within a sound-attenuated room. The recordings underwent background noise reduction using the default settings in Audacity. Then, each recording was divided into two segments: the onset (e.g., “bala-” from “balanza”) and the offset (e.g., “-nza”). The stimuli were cut at the zero-crossing point nearest to the point of disambiguation (POD), with an average POD occurring at approximately 433 ms. The recordings were adjusted to a 70 dB intensity level using Praat software (version 6.3.03; Boersma & Weenink, 2023). A 100-ms silence was added to the beginning and end of each word.

Given that the onset portions might include co-articulatory cues predicting the offset (e.g., the “bala” from “balanza” may predict “–nza” more than “–cio”), each of the two voiced onsets in a pair (e.g., “bala_{nza}” and “bala_{cio}”) was spliced onto each of the two offsets to counterbalance the co-articulatory cues in the onsets (e.g., “${}_{\text{bala}}\text{nza}$” and “${}_{\text{bala}}\text{cio}$”). Consequently, half of the stimuli were generated by combining parts from the same item (e.g., “bala_{nza}” and “${}_{\text{bala}}\text{nza}$”; *matching splice*), while the other half were formed from different items (e.g., “bala_{nza}” and “${}_{\text{pala}}\text{nza}$”; *mismatching splice*). More details can be found in Figure S4 in the Supplemental Materials of Kapnoula et al. (2021).

Finally, the VOT continua were constructed by pairing items that differed solely in the voicing of the onset consonant (e.g., “balanza” and “palanza”). These pairs were then utilized as the endpoints for generating seven-step VOT continua (Spanish: -35 – +15 ms; English: 0 – +48 ms). The construction of these continua followed the aforementioned method employed in creating the VAS stimuli. This process yielded 140 auditory items (5 pairs × 2 splice conditions × 2 offsets × 7 VOT steps). The F0 was standardized at 181 Hz for the onsets of all Spanish stimuli. Each item was presented three times, resulting in 420 experimental trials.

For filler stimuli, only the correct (e.g., “lavabo” [sink] and “regalo” [gift]) and misarticulated (e.g., “ravabo” and “legalo”) versions were recorded and used. An equivalent number of filler trials (i.e., 420 trials) were interspersed to introduce variety and obscure the task purpose, yielding a total of 840 trials. These trials were presented in a randomized order.

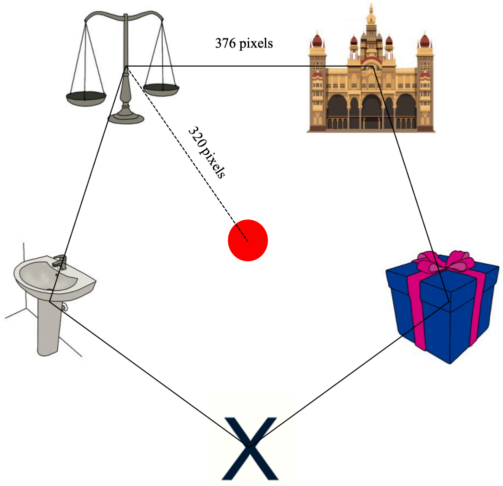
Each experimental pair was paired with a filler pair to create a four-item set (e.g., “balanza”, “palacio”, “lavabo”, and “regalo” formed one set in the Spanish VWP), with the constraint that all items within a VWP set were semantically unrelated and shared the same stress pattern. Items in a VWP set always appeared together. Five such VWP sets were created in each language.

A picture was selected for each of the four words within each of the five VWP sets, following the procedure reported by McMurray et al. (2010). Several pictures for each word were initially sourced from a clipart database, mainly utilizing the MultiPic database (Duñabeitia et al., 2018) for the Spanish task. A single picture was chosen as the most characteristic exemplar for each word. The selected images underwent refinement, including the removal of extraneous elements and optimization for clarity to ensure a faithful depiction of the intended word. The final images were approved by a lab member with extensive experience with VWP methodologies.

Experiment Builder (version 2022.2.5; SR Research Ltd., 2022) was used to program and run the experiment with the same settings across the two languages. Visual stimuli were presented on a 24” monitor with a resolution of 1024 × 768 pixels. Each display consisted of five visual stimuli, including four pictures and an “X”. The display was arranged in a pentagonal configuration, as illustrated in Figure 1. The pentagonal configuration was designed to maintain an equal distance between the center of each picture and screen center (320 pixels), as well as from one picture to another (376 pixels apart). Each picture was 240 × 240 pixels, while the “X” was 66 × 80 pixels. The position of the four pictures was randomized across trials, except for the “X”, which was always at the bottom.

Figure 1

VWP Display



Note. Lines and pixel values are for illustration and were not shown during the experiment. Images shown here were from the Spanish VWP; the same configuration was used in the English VWP.

Procedure

Participants were first fitted with the eye tracker and were given instructions. Then, a familiarization phase was administered, in which each picture and its corresponding word were shown one by one.

At the beginning of each trial of the main experiment, a red circle appeared at the center of the screen along with four pictures and an "X" presented in a pentagonal configuration. After 500 ms, the circle turned blue and participants clicked on it to hear a word. This duration between the appearance of red circle and the mouse click allowed participants a momentary visual preview of the pictures before the onset of the auditory stimulus to minimize potential eye movements driven by visual search (Apfelbaum et al., 2021). Participants then clicked on the corresponding picture or the "X" if they thought none of the four pictures matched what they had heard. While there was no explicit time constraint, participants typically responded in less than two seconds (Spanish: *M* = 1452 ms, *SD* = 138 ms; English: *M* = 1645 ms, *SD* = 195 ms). The trial ended once participants clicked on a picture (or the “X”), and the next trial began. The VWP task for each language took about 52 minutes.

Eye-tracking Recording and Analysis

Eye movements were recorded at a sampling rate of 1,000 Hz utilizing the SR Research EyeLink 1000 Plus system, configured in a head-stabilized mode with a chin-rest and a 35-mm lens, within a sound-attenuated booth. The distance between the participant's eyes and the screen was maintained at 70 cm. The overhead lighting and audio volume were set at a consistent level across participants. Auditory stimuli were presented through BeyerDynamic DT-770 Pro 250 Ohm headphones. Calibration and validation of the eye-tracker (conducted between the familiarization phase and the main experiment) used the standard nine-point procedure. A drift check was performed every 20 trials, which was also when participants could take a break. If a drift check failed, the eye-tracker was recalibrated. The eye-tracking data, recorded from the onset of the trial (the appearance of the blue circle) to the participant’s response (mouse click), was automatically parsed into saccades and fixations using default psychophysical parameters with EyelinkAnalysis (version 3.5.1). Adjacent saccades and fixations were amalgamated into a single look, commencing at the onset of the saccade and concluding at the fixation offset, consistent with established methodologies (McMurray et al., 2002, 2010).

For analyses, the eye-tracking data were downsampled to 250 Hz, and a fixed trial duration of 2,000 ms relative to stimulus onset was established. In instances where trials terminated before this point, the last eye movement was extended. Conversely, trials surpassing the 2,000-ms threshold were truncated. This approach has been used in many previous studies (e.g., Allopenna et al., 1998; Kapnoula et al., 2021; McMurray et al., 2002;

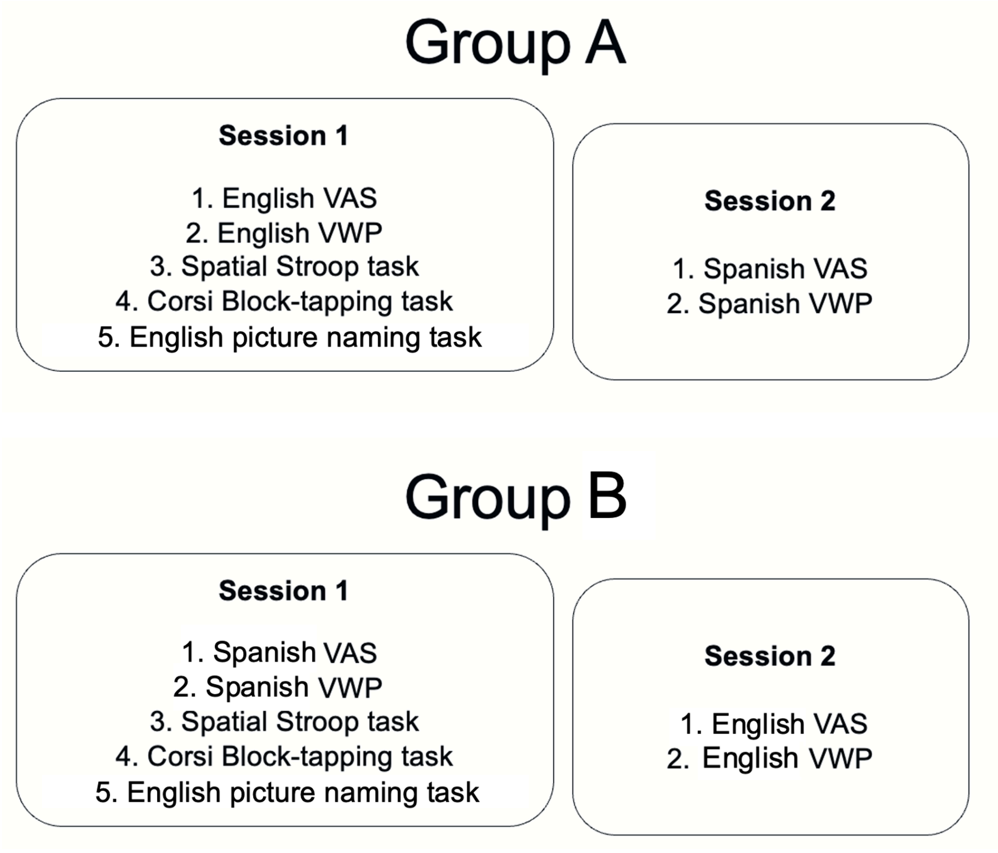
Ou et al., 2021). Picture boundaries were extended by 58 pixels to accommodate potential noise or head drift in the eye-track record. This extension did not result in any overlap, with a 118-pixel vertical and 136-pixel horizontal space maintained between pictures.

Order of Tasks

Participants were randomly assigned to either Group A (*N* = 35) or Group B (*N* = 35). The order of tasks for each group is shown in Figure 2. All participants attended two lab sessions between seven and fourteen days apart. The first session lasted approximately 1.5 hours, while the second session lasted about one hour. In addition, all participants completed an online language exposure questionnaire and a training task, designed to familiarize them with the English stimuli, one to seven days prior to their English session.

Figure 2

Order of Tasks for Each Group



Transparency and Openness

Data collection took place during the summer of 2023. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and stimuli for both the VAS and VWP tasks are available on the OSF page [https://osf.io/9g63y/]. This study's design and its analysis were not pre-registered.

Results

All 70 participants finished all the tasks. However, data from five participants were excluded from the Spanish VWP, and data from four participants were excluded from the English VWP due to problematic eye-tracking data. As mentioned above, the Spanish VAS data of two participants and the English VAS data of three participants were excluded from the analyses. To maximize statistical power, all remaining data from the participants mentioned above were included in the analyses. Therefore, the number of participants included in subsequent data analyses varied from 63 to 68, with specific numbers provided in each corresponding table and figure.

Preliminary VAS Analyses: Individual Differences in Gradiency and Perceptual Consistency

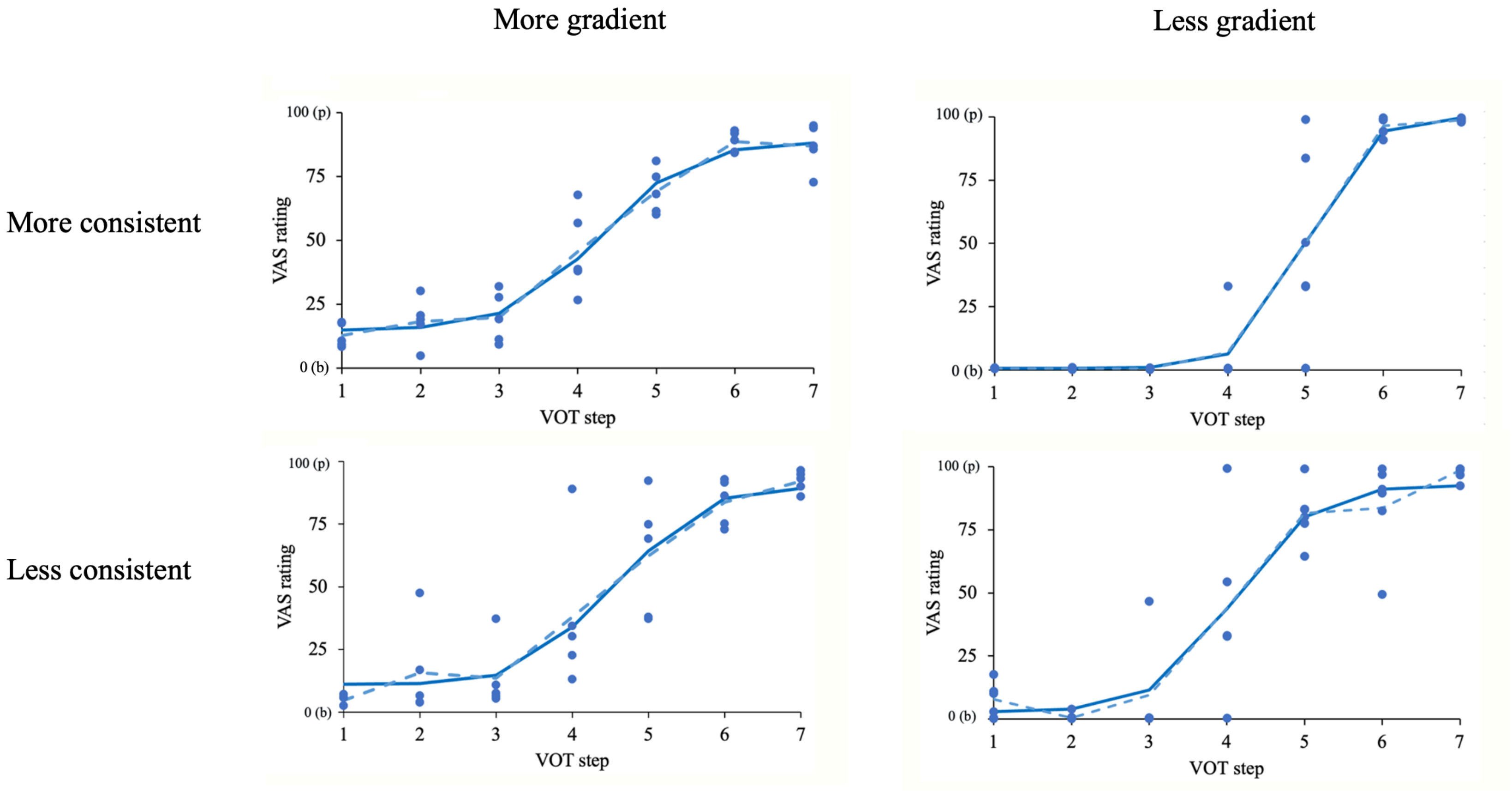
The expected patterns were observed in the VAS tasks, with substantial individual differences in both gradiency and perceptual consistency. That is, some individuals clicked more frequently on intermediate points on the line, while others displayed a strong preference for the endpoints. Similarly, participants differed in how consistently they rated the same stimuli. To illustrate the observed patterns, Figure 3 provides some examples of individual-subject Spanish VAS data with different gradiency and perceptual consistency patterns. Figure 4 shows the average VAS data for Spanish and English. The relatively high /p/

response when the VOT is low in the English VAS indicates that the recognition of the English /b/ sound is challenging for our participants; recall that the VOT range for /b/ in English is similar to that for /p/ in Spanish. In contrast, there is no bilabial consonant in Spanish with a VOT similar to that of the English /p/.

Correlation analyses were conducted between gradiency, perceptual consistency, our covariates (i.e., working memory, inhibitory control, L2 proficiency, L2 exposure, and musical training), and demographic variables to examine the data structure. Detailed results can be found in Supplementary Material VI.

Figure 3

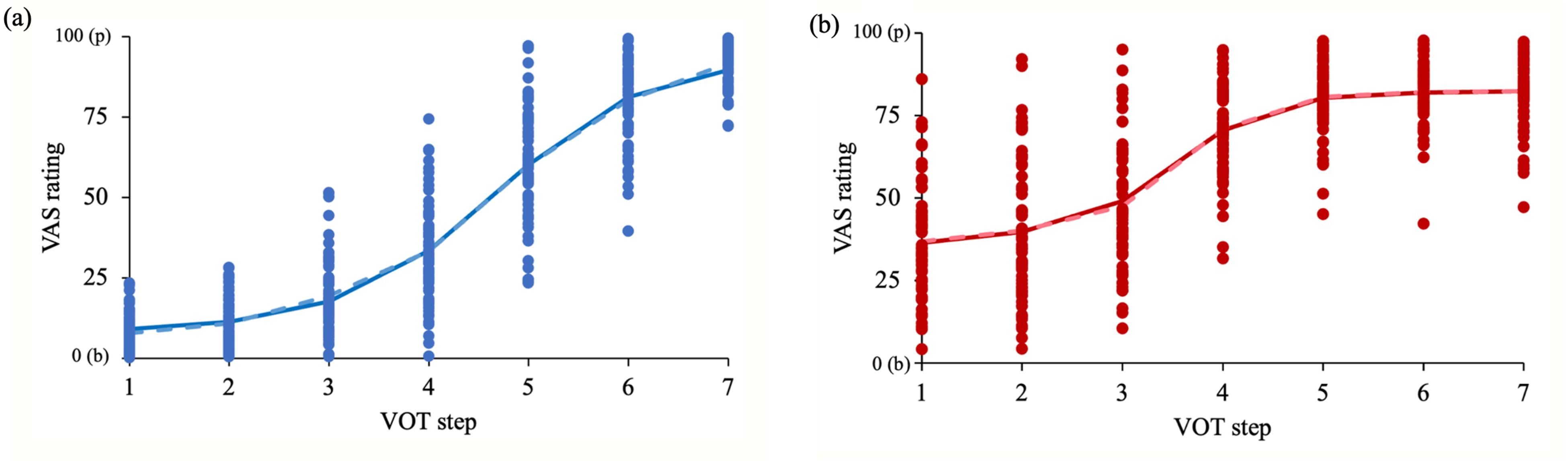
Examples of Spanish VAS Data Showing Different Gradiency and Perceptual Consistency of Four Participants



Note. Each dot represents a participant's response on each trial. The solid line is the fitting curve based on the rotated logistic function, while the dashed line connects the average response at each VOT step for that participant.

Figure 4

Average VAS Data for (a) Spanish (N = 68) and (b) English (N = 67)



Note. Each dot represents the average of a participant's response at each VOT step. The solid line is the fitting curve based on the rotated logistic function, while the dashed line connects the grand average of all participants' responses.

Preliminary VWP Analyses

Before the analyses, the VOT step of the initial phoneme of each auditory stimulus was recoded to reflect its acoustic distance from the target (henceforth: *tDist*), ranging from 0 to 6. To illustrate, consider a stimulus with VOT Step 1 (VOT = -35 ms in Spanish and 0 ms in English). In this case, tDist would be coded as 0 for the extreme /b/ targets (e.g., “balanza” and “beachball”) and 6 for the extreme /p/ targets (e.g., “balacio” and “beachpit”). Likewise, for a stimulus with VOT Step 7 (VOT = +15 ms in Spanish and +48 ms in English), tDist would be 0 for the extreme /p/ targets (e.g., “palacio” and “peachpit”) and 6 for the extreme /b/ targets (e.g., “palanza” and “peachball”). This recoding allowed us to assess the effect of acoustic distance from the target across both voiced and voiceless stimuli.

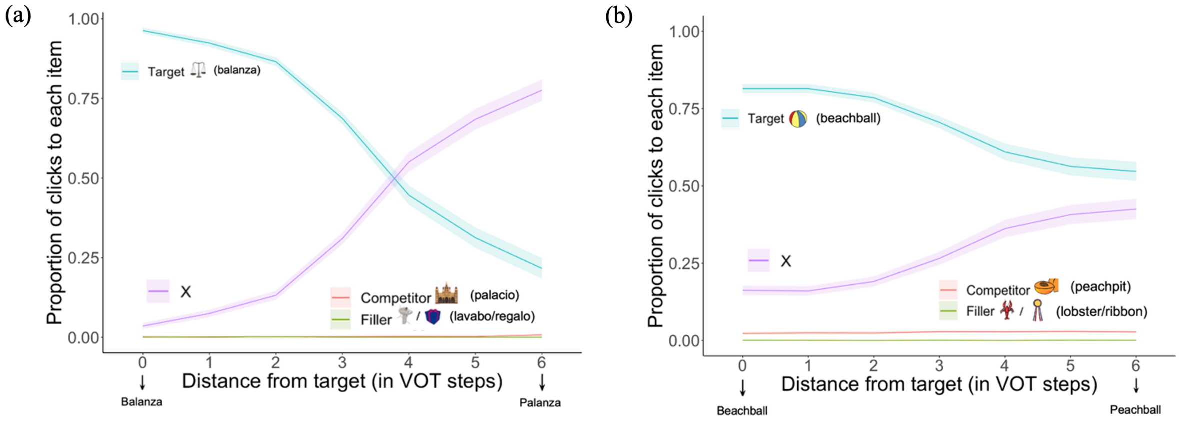
Analyses of Click Responses

For the Spanish VWP, in the case of completely unambiguous target stimuli (i.e., tDist = 0), average accuracy was 96.3% (*SD* = 7.1%). Participants clicked on the competitor on 0.1% (*SD* = 0.4%) of the trials, on the filler item 0.1% (*SD* = 0.3%), and on the “X” 3.5% (*SD* = 7.1%). RTs averaged 1,452 ms (*SD* = 138 ms). For the English VWP, average accuracy for completely unambiguous target stimuli was good but noticeably lower (*M* = 81.4%, *SD* = 11.4%). In these instances, participants clicked on the competitor on 2.3% (*SD* = 3.3%) of the trials, on the filler item on 0.1% (*SD* = 0.3%), and on the “X” on 16.2% (*SD* = 12.0%). Participants exhibited slightly slower RTs in English (*M* = 1,645 ms, *SD* = 195 ms).

The clicking response patterns for L1 vs. L2 differed more when the tDist was larger. For both languages, an increase in tDist reasonably led participants to be less inclined to click on the target and more inclined to click on the “X” to indicate that none of the pictures matched the heard word; however, this pattern was less pronounced for English. For Spanish, when the VOT was highly misleading (tDist = 6), participants selected the target on 21.6% of the trials and selected the "X" on 77.6% (Figure 5a). In contrast, for highly misleading items in English, participants selected the target on 54.7% of the trials and selected the "X" only on 42.5% (Figure 5b). Hence, participants' clicking responses in English seem to be more affected by the disambiguating (lexical) information.

Figure 5

Proportion of Responses to Target, Competitor, Filler, and "X" as a Function of Distance from the Target (tDist) in (a) Spanish VWP (N = 65) and (b) English VWP (N = 66)



Note. Shaded ribbons indicate standard errors of the mean (SEMs).

While the above results showed that participants performed the VWP reasonably, preliminary VWP analyses on accuracy, RT, and fixations were conducted to further ensure that our manipulation was successful. Importantly, as expected, when the acoustic distance from the target (tDist) was higher in both English and Spanish, participants were less likely to click on the target, clicked on the target more slowly, and fixated the target less often. For more details, interested readers can refer to Supplementary Material VII.

Primary Analyses I: Effects of Gradiency on Initial Lexical Activation and Speech Perception Flexibility in L1 Spanish and L2 English

Our first research question is whether gradiency in speech categorization influences initial lexical activation and/or speech perception flexibility in L1 Spanish and L2 English. To address this question, we first look at the initial lexical activation as indexed by the garden-path rate – the proportion of trials when a participant fixated on the competitor picture before the POD was reached (corrected for a 200 ms oculomotor delay; Hallett, 1986; Salverda et al., 2014) –as a measure of initial lexical activation. Then, we use two measures to capture speech perception flexibility: (1) recovery rate, the proportion of trials in which a participant ultimately recovered by looking back at the target picture after making a garden-path (after the POD); (2) recovery latency, the time it took a participant to recover.

Building upon the methodology established by Kapnoula et al. (2021), linear mixed-effects models (LMEMs) were employed to examine the relationship between *slope* (the inverse of gradiency; extracted from the VAS) and how participants coped with garden-path situations (i.e., garden-path rate, recovery rate, and recovery latency). Apart from slope, each analysis involved two additional factors: the *tDist* and *target voicing*. As described above, tDist represents the acoustic distance between the target and the auditory stimulus in terms of the VOT step of the initial phoneme. Target voicing depends on whether the target started with a /b/ (e.g., “balanza” in Spanish or “beachball” in English) or /p/ (e.g., “palacio” in Spanish or “peachpit” in English).

For all the following analyses, LMEMs were conducted using the lmerTest package (Version 3.1-3; Kuznetsova et al., 2017), an extension of the lme4 package (Version 1.1-35.1; Bates et al., 2015), in R (Version 4.3.2; R Core Team, 2023) via RStudio (Version 2023.12.0+369; RStudio Team, 2023). Graphs were created with ggplot2 package (Wickham, 2016). For each analysis, we compared models with increasing complexity to determine the random effect structures justified by our data using likelihood ratio tests (LRTs; Matuschek et al., 2017). We also decided whether the covariates should be included in the model as fixed effects using the same method. Based on the results, we only kept tDist, slope (or consistency), and target voicing in our main analyses. The model used in each analysis and the complete statistics are specified in Supplementary Material VIII.

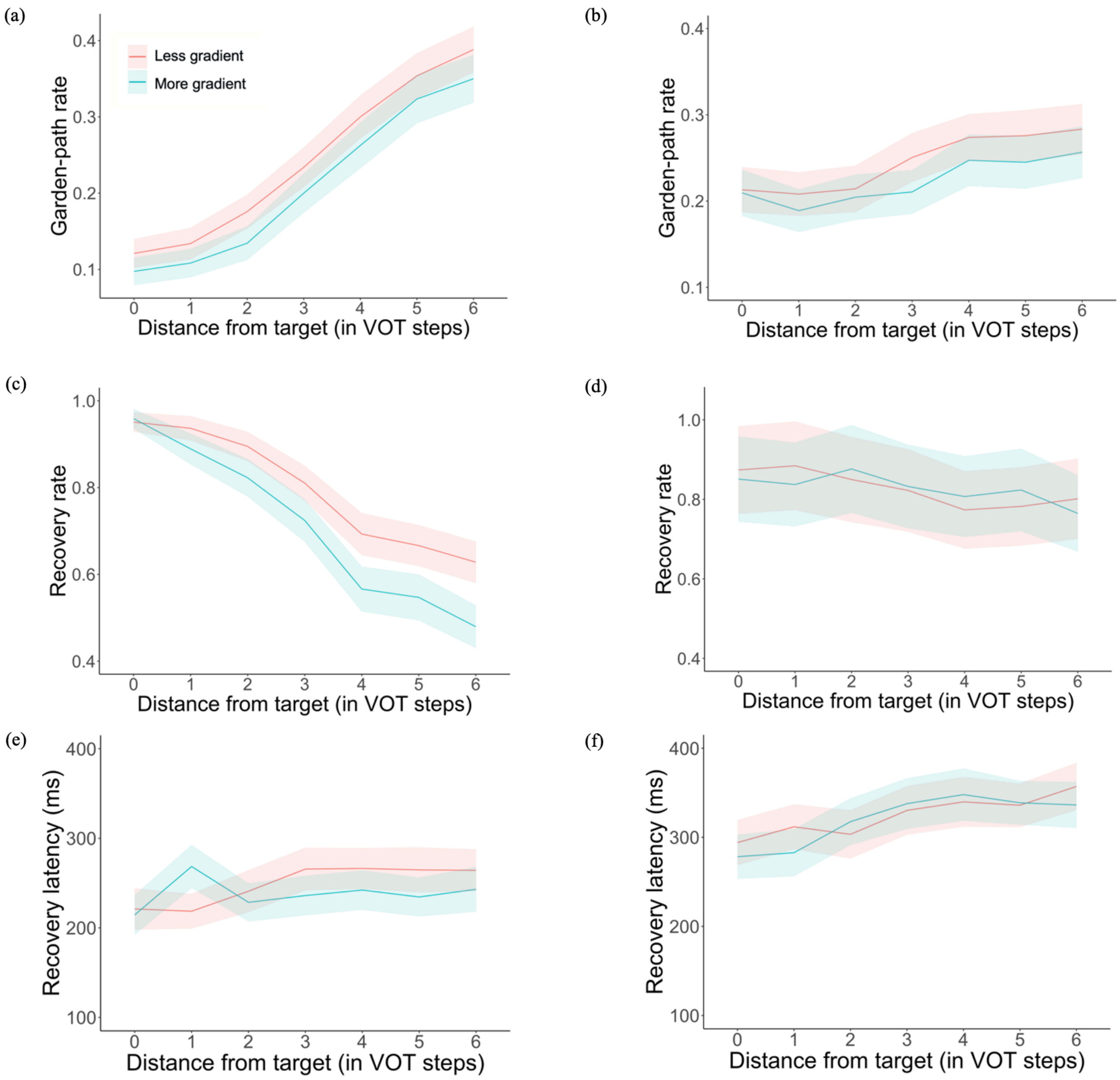
Garden-path Rate and Gradiency

The first analysis examined the likelihood of participants looking at the competitor image (e.g., the "palacio" when hearing "balanza") before reaching the POD. Each trial was given a value of 1 if the participant looked at the competitor at any time before the POD of the stimulus on that trial, and a 0 otherwise. This measure was averaged within-cell, empirical-logit-transformed, and analyzed as a function of tDist (centered), target voicing (effect-coded; /b/ target = 1; /p/ target = -1), and slope (centered). The maximal random effects structure justified by our data included random intercepts and random slopes of tDist for subjects and items (see Supplementary Material VIII for the complete model).

Greater tDist predicted a higher proportion of garden-path trials both in L1 Spanish, $B = 0.24$, $t(11) = 8.45$, $p < .001$ (Figure 6a) and in L2 English, $B = 0.06$, $t(11) = 3.32$, $p = .007$ (Figure 6b). This pattern is in line with the L1 English findings from McMurray et al. (2009) and Kapnoula et al. (2021) in showing that the strength of early lexical activation depends on fine-grained differences in VOT. Therefore, these results extend the previous finding to L1 Spanish and L2 English spoken word recognition. Crucially, VAS slope (gradiency) did not predict garden-path rate (Spanish: $B = 0.05$, $t(61) = 1.03$, $p = .307$; English: $B = 0.04$, $t(61) = 0.58$, $p = .563$), and its interaction between slope and tDist was also nonsignificant (Spanish: $B = 0.00$, $t(61) = 0.31$, $p = .754$; English: $B = 0.00$, $t(61) = 0.35$, $p = .726$), corroborating the findings from Kapnoula et al. (2021). Full results for Spanish and English are reported in Tables S9a and S9b, respectively.

Figure 6

(a, b) Garden-path Rate, (c, d) Recovery Rate, and (e, f) Recovery Latency as a Function of Distance from the Target (tDist) for Each Gradiency Group (Based on a Median Split of VAS Slope) for Spanish (left) and English (right)



Note. Slope was split (midpoint excluded) in the graph for illustration but it was analyzed as a continuous variable in the analyses. Across panels, the left column displays the data for Spanish, while the right column shows the data for English. Shaded ribbons indicate SEMs. *N* = 63 for both languages.

Recovery Rate and Gradiency

Next, we analyzed the likelihood of recovery, measured as the proportion of recovered trials out of all garden-path trials. Recovered trials were defined as instances where participants initially fixated on the competitor picture before the POD was reached, and subsequently directed their gaze to the target after the POD. This analysis includes trials where participants clicked on any picture (including "X") at the end. The proportion of recovered trials was empirical-logit-transformed and served as the DV in the LMEM. Due to singularity or convergence issues even after changing the optimizer to *bobyqa* and increasing the maximum iterations (Brauer & Curtin, 2018; Matuschek et al., 2017), by-item random slopes were excluded for the Spanish model, and by-subject and by-item random slopes were excluded for the English model (see Supplementary Material VIII for the complete models and statistics).

Again, in line with Kapnoula et al. (2021), greater tDist predicted lower recovery rates, both for L1 Spanish, B = -0.19, t(60) = -10.77, p < .001 (Figure 6c) and for English, B = -0.03, t(2994) = -3.72, p < .001 (Figure 6d). The slope (gradiency) effect was marginally significant, B = 0.09, t(60) = 1.83, p = .073 for Spanish, and nonsignificant for English, B = -0.00, t(56) = -0.03, p = .974. Unlike previous results, the interaction between tDist and slope (gradiency) was not significant (Spanish: B = 0.03, t(59) = 1.56, p = .123; English: B = 0.00, t(2996) = 0.36, p = .719). Full results are reported in Table S10a and S10b.

Recovery Latency and Gradiency

Next, we looked at the relationship between gradiency and the time it took participants to recover from garden-paths. This was calculated as the log-transformed time from the POD until the first fixation to the target. Only recovered trials were included. A mixed-effects model was employed, incorporating the same fixed effects as in the preceding analyses. By-subject and by-item random slopes for tDist were incorporated in the English model, while the by-item random slope was omitted in the Spanish model due to issues related to singularity.

For L1 Spanish (Figure 6e), tDist significantly predicted recovery latency, $B = 0.01$, $t(58) = 3.40$, $p = .001$, with slower recovery at greater tDist. The main effect of slope (gradiency) was nonsignificant, $B = 0.02$, $t(59) = 1.63$, $p = .108$. Crucially, the interaction between tDist and slope (gradiency) was significant, $B = 0.01$, $t(54) = 2.33$, $p = .024$. To examine this interaction, we divided the dataset into high and low tDist (median point excluded) and ran the model with slope (gradiency) as the predictor in each dataset. In the high tDist model, slope (gradiency) was significant, $B = 0.04$, $t(58) = 2.86$, $p = .006$. In contrast, in the low tDist model, it was not significant, $B = 0.00$, $t(58) = 0.01$, $p = .995$. These results suggest that more gradient listeners recovered more quickly than less gradient ones, particularly when the tDist was high. For English (Figure 6f), tDist once again significantly predicted recovery latency, $B = 0.02$, $t(11) = 4.87$, $p = .001$, with slower recovery at greater tDist. Neither the main effect of slope (gradiency), $B = 0.00$, $t(56) = 0.39$, $p = .698$, nor its interaction with tDist were significant, $B = 0.00$, $t(53) = -0.06$, $p = .949$. Full results are reported in Tables S11a and S11b.

Interim Summary

These analyses revealed that gradiency has an impact on recovery latency in L1 Spanish: More gradient listeners recovered more quickly when the tDist was high. This finding suggests that gradiency is helpful in recovering from garden-paths, despite the VOT differences between Spanish and English (cf. Kapnoula et al., 2021). However, this effect was not observed in L2 English in the present study, suggesting that the functional role of gradiency is qualified by language status.

Primary Analyses II: Effects of Perceptual Consistency on Initial Lexical Activation and Speech Perception Flexibility in L1 Spanish and L2 English

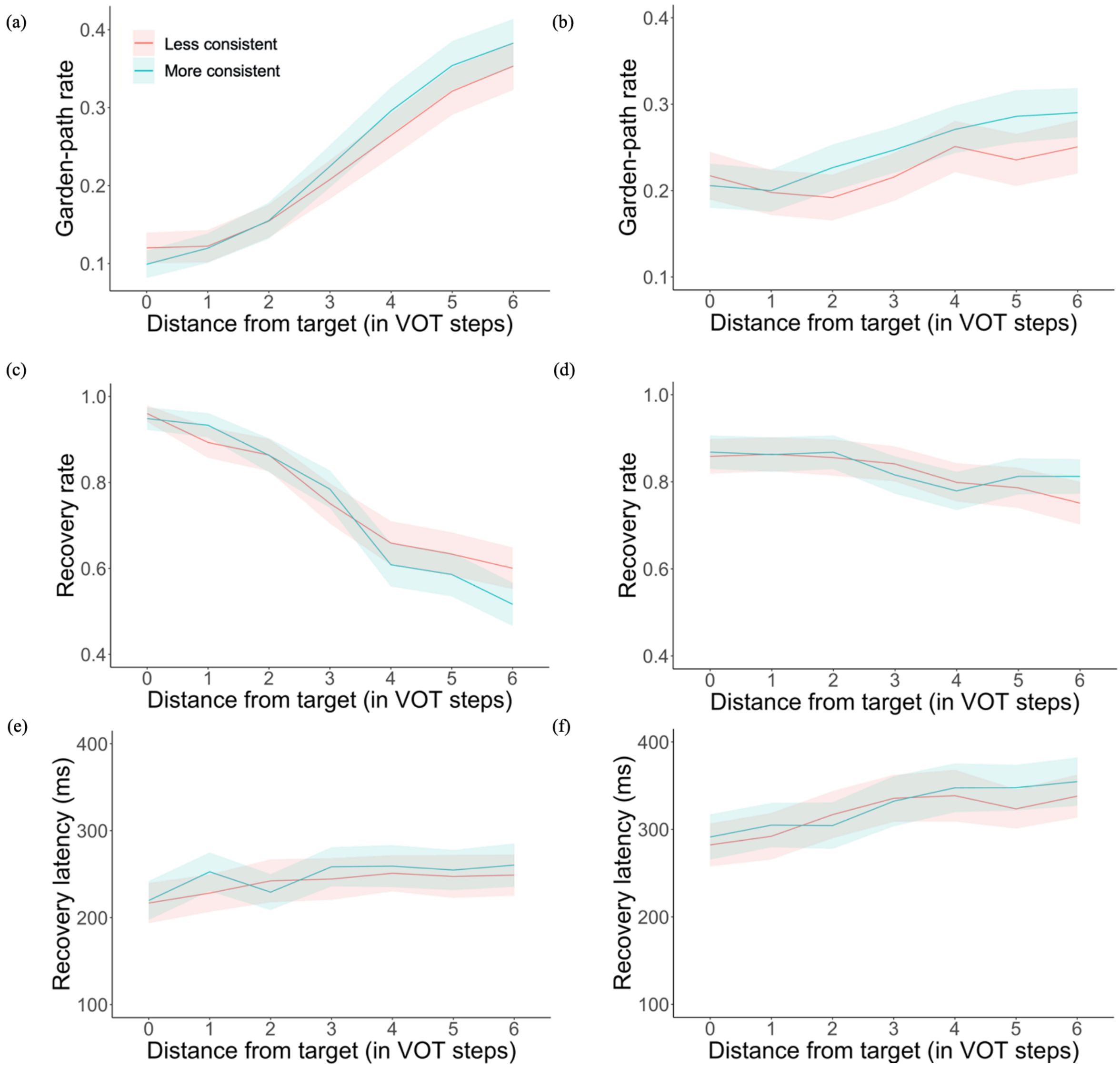
Next, we examined our second research question: whether perceptual consistency – as opposed to gradiency – affects how listeners deal with lexical garden-path situations. All the predictors were the same as in the gradiency analyses above, but instead of slope, we used our perceptual consistency measure (centered). For brevity, we focus here on the main effect and interactions of perceptual consistency because the main effects of tDist were essentially unchanged.

Garden-path Rate and Perceptual Consistency

Besides the significant effect of tDist that we have already reported, we found a marginal interaction between tDist and Spanish perceptual consistency, B = 0.01, t(61) = 1.91, p = .061 (Figure 7a). Therefore, listeners who demonstrated greater perceptual consistency in the Spanish VAS were numerically but not significantly more inclined to make more garden-paths when the tDist was higher (Table S12a). Similarly, for English (Figure 7b), more consistent listeners made slightly more garden-paths when the tDist was higher, B = 0.002, t(61) = 1.71, p = .092 (Table S12b). While the interactions between perceptual consistency and tDist were marginally significant, the same pattern emerged in both languages. This suggests that it is possible that we failed to detect a weak effect due to insufficient power. Therefore, we combined both the Spanish and English datasets and reran the same analyses, but with language (1: English; -1: Spanish) as an additional predictor. In this case, the interaction between tDist and perceptual consistency was significant, B = 0.004, t(269) = 3.03, p = .003. The three-way interaction of tDist, perceptual consistency, and language was not significant, B = -0.001, t(1243) = -1.20, p = .231, indicating that the interaction between tDist and perceptual consistency was not specific to one language. Therefore, in general, listeners with higher perceptual consistency made more garden-paths when the tDist was high, compared with listeners with lower perceptual consistency (Table S12c).

Figure 7

(a, b) Garden-path Rate, (c, d) Recovery Rate, and (e, f) Recovery Latency as a Function of Distance from the Target (tDist) for Each Consistency Group (Based on a Median Split of Perceptual Consistency) for Spanish (left) and English (right)



Note. Perceptual consistency was split (midpoint excluded) in the graph for illustration but it was analyzed as a continuous variable in the analyses. Across panels, the left column displays the data for Spanish, while the right column shows the data for English. Shaded ribbons indicate SEMs. $N = 63$ for both languages.

Recovery Rate and Perceptual Consistency

No significant main effects of perceptual consistency (Spanish: *B* = -0.00, *t*(61) = -0.06, *p* = .954; English: *B* = 0.00, *t*(55) = -0.06, *p* = .954) or interactions between perceptual consistency and tDist (Spanish: *B* = -0.00, *t*(59) = -0.54, *p* = .592; English: *B* = 0.001, *t*(2989) = 0.45, *p* = .651) were observed in both languages (Figures 7c and d; Tables S13a and b).

Recovery Latency and Perceptual Consistency

No significant main effects of perceptual consistency (Spanish: *B* = 0.00, *t*(64) = 0.10, *p* = .923; English: *B* = -0.00, *t*(55) = -0.36, *p* = .717) or interactions between perceptual consistency and tDist (Spanish: *B* = 0.00, *t*(60) = -0.20, *p* = .841; English: *B* = 0.00, *t*(54) = -0.64, *p* = .523) were observed (Figures 7e and f; Tables S14a and b).

Interim Summary

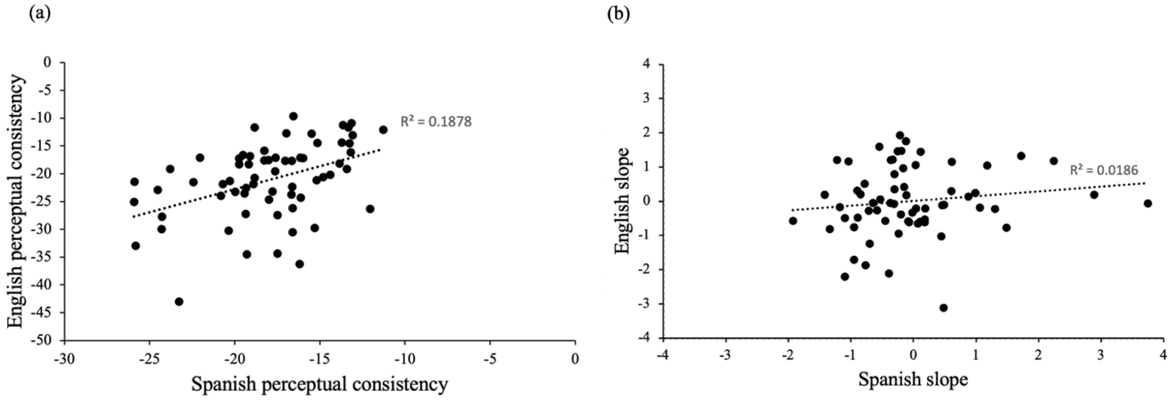
In summary, participants who responded more consistently in the VAS tended to make more garden-paths when the tDist was higher, compared with listeners with lower perceptual consistency, in both L1 Spanish and L2 English. This suggests that listeners with higher perceptual consistency tend to make higher use of acoustic information to activate lexical candidates during early stages of spoken word recognition. In addition, the functional role of perceptual consistency appears to be language-general.

Primary Analyses III: Are Gradiency and Perceptual Consistency Stable Across Languages?

To test our third research question, we looked at the relationship between L1 and L2 gradiency by conducting Spearman's correlations between the Spanish and English slopes. Similar analyses were conducted for perceptual consistency. The positive correlation between Spanish and English perceptual consistency was significant, *r*(63) = .44, *p* < .001, indicating that perceptual consistency is not language-specific (Figure 8a). In contrast, the correlation between Spanish and English slopes was nonsignificant, *r*(63) = .13, *p* = .314, indicating that gradiency is language-specific (Figure 8b). Hence, perceptual consistency seems to reflect an individual trait, while gradiency seems to be language-specific. This is in line with previous work showing that gradiency and perceptual consistency are independent (e.g., Kapnoula et al., 2017). In addition, the stability of speech sound perception is likely a more general mechanism that spans across languages.

Figure 8

Correlations on (a) Perceptual Consistency and (b) Gradiency between Spanish and English (N = 65)



Bayesian Analyses

Given the observation of some null effects in our findings, we conducted Bayesian analyses to evaluate the evidence for these effects. Following the recommendations of Dienes (2024), we re-ran all analyses (not only those corresponding to null effects) using a Bayesian approach to ensure consistency. The results were comparable to our main findings. Importantly, there was decisive evidence supporting the finding that more gradient listeners recovered more quickly than less gradient ones when the tDist was high for Spanish. Weak evidence against the interaction was found for English, which means that we cannot strongly conclude in favor of the absence of the interaction. Additionally, there was very strong evidence in Spanish that listeners with higher perceptual consistency made more garden-paths when the tDist was high, compared to listeners with lower perceptual consistency. Strong evidence for this pattern was also found for English. Furthermore, the evidence for the positive correlation between Spanish and English perceptual consistency was decisive. The absence of correlation between Spanish and English gradiency was weakly supported, which means that we cannot strongly conclude that the correlation was absent or present. Interested readers can refer to Supplementary Material IX for detailed information.

Discussion

In the present study, we investigated the impacts of speech perception gradiency and perceptual consistency on initial lexical activation and speech perception flexibility. Testing both L1 and L2 spoken word recognition allowed us to better understand the nature of these constructs and the degree to which they are language-specific vs. generic traits. Specifically, our research extends prior work in this domain in three substantial ways: (1) We examined whether the functional role of gradiency in speech perception flexibility is modulated by language-specific properties and language status, (2) we asked whether the functional role of perceptual consistency in spoken word recognition is modulated by language, and (3) we directly compared gradiency and perceptual consistency across languages. Similarly to Kapnoula et al. (2021), we found that in L1 Spanish, more gradient listeners showed enhanced speech perception flexibility in dealing with misleading auditory inputs. In contrast, the same effects of gradiency were minimal in listeners' non-native language processing. Therefore, the functional role of gradiency appears to be qualified by language status but not by acoustic (in this case, VOT) information. Furthermore, listeners with greater perceptual consistency were more likely to activate lexical candidates at an early stage in both languages, suggesting that, in contrast to gradiency, consistency is a generic speech perception trait. This conclusion was further supported by the finding that perceptual consistency is more stable across languages, while gradiency is more language-specific. These findings offer significant contributions to our understanding of speech perception, as discussed below.

Gradient Listeners Recovered More Quickly in L1

A critical finding of our study is that in their L1, listeners with higher gradiency recovered more quickly than less gradient ones, after they were initially misled by garden-path stimuli. In other words, they corrected their interpretation more rapidly after being presented with disambiguating information later in the word. This effect was supported by decisive evidence according to the Bayesian analysis. Notably, this rapid adjustment was especially evident when the acoustic deviation from the target was high. This observation is in line with the idea that more gradient listeners are better able at considering multiple hypotheses concurrently. This ability helps them avoid early commitment to a single lexical interpretation, enabling more efficient resolution of misunderstandings (Kapnoula et al., 2021). In contrast, less gradient listeners experience more category-driven warping, which may lead them to strongly commit to one word and suppress all others, slowing down their recovery from misleading information. Our results demonstrate that the ability of more gradient listeners to recover from misinterpretations is not language-specific, as we successfully extended Kapnoula et al.'s (2021) findings in English to Spanish, despite the different VOTs of stop consonants between the two languages. This suggests that a fundamental benefit of gradiency is its enhancement of speech perception flexibility.

Although the same conclusion was reached in the present study and Kapnoula et al. (2021), it is important to note that this conclusion is based on different measures. Kapnoula et al. (2021) observed that as the acoustic distance from the target increased, monolingual English speakers with higher gradiency showed a higher *recovery rate* compared with those with lower gradiency. On the other hand, our findings indicated that more gradient L1 Spanish bilinguals had shorter *recovery latency* than less gradient bilinguals. This difference may be attributed to differences in VOT distributions between Spanish and English. In Spanish, the /b/-/p/ contrast is defined by the presence (/b/) vs. absence (/p/) of pre-voicing –

a qualitative difference; the distinction in English is more quantitative, with shorter positive VOTs for /b/ than for /p/. As a result, native Spanish speakers might exhibit more categorical responses on the Spanish VAS, making them potentially less sensitive to within-category differences (for preliminary evidence for this, see Kapnoula & Samuel, 2024). Thus, the relationship between gradiency and recovery rate might be obscured by a narrower range of gradiency among Spanish speakers. Consequently, the effect was only captured by recovery latency, a measure more sensitive than the recovery rate.

Minimal Gradiency Effects on Recovery in L2

Another main question in the current study is whether the gradiency effects found in L1 operate in the same way in L2. This question clarifies whether the functional roles of gradiency are affected by language status. Contrary to our expectations, the present study revealed minimal gradiency effect on the recovery latency in L2 English, even though these same listeners showed such effect in their L1 Spanish, suggesting that the functional role of gradiency is affected by language status. The Bayesian analyses revealed weak evidence against the interaction between gradiency and recovery latency. Although we cannot strongly conclude in favor of the absence of the interaction, the weak evidence can potentially be explained by at least two factors. Firstly, Spanish learners of English may predominantly utilize the secondary cue of F0 when determining voicing distinctions within the positive VOT range (Llanos et al., 2013; see Figure S3 for present findings supporting this account). Consequently, our VOT manipulations might exert less influence on their early encoding and categorization processes compared to native English speakers.

Secondly, in general, our participants confined their ratings to a narrow range in the English VAS, for example, only responding from 40 to 60 on a scale from 0 to 100 to signify sounds from a clear /b/ to a clear /p/, suggesting that they might not have a clear representation of the category (see Figure 4b). A paired-sample t-test showed that the rating range (calculated by the difference between the maximum and minimum values based on the rotated logistic function) of the English VAS rating was significantly narrower than that of Spanish VAS, $t(64) = -15.9$, $p < .001$. Importantly, the correlation between the rating range and English proficiency score was not significant, $r(65) = .12$, $p = .341$, indicating that this difference was not driven by differences in proficiency. Rather, this aligns with the view that non-native learners often face challenges in forming categorical representations, despite extensive training (see Baese-Berk et al., 2022, for a review). Indeed, our findings highlighted a particular challenge with the English /b/ sound, as evidenced by lower accuracy rates for /b/ compared to /p/, even when the acoustic distance from the target was minimal in the English VWP (see Figure S1b). As we have noted, this difficulty can be attributed to the similarity in VOT between the English /b/ and Spanish /p/. According to the revised speech learning model (SLM-r) by Flege and Bohn (2021), L2 speech sounds that fall close to a learner's L1 categories are particularly challenging to acquire. Related to this, note that only those participants who demonstrated an increase in their responses from a complete /b/ to a /p/ sound were included in the analysis. It would be informative to test balanced bilinguals in languages with similar VOTs when examining the relationship between gradiency and speech perception flexibility, in order to better understand the relationship between language status and the functional roles of gradiency.

Perceptual Consistency Facilitates Initial Lexical Activation

Variability is a fundamental aspect of speech perception, encompassing not just external environmental factors, such as the varying VOTs articulated by different speakers, but also variability inherent to speech processing, which is what our perceptual consistency measure aimed at capturing. Previous work suggests that higher perceptual consistency is linked to more effective learning of non-native phonetic contrasts (Fuhrmeister et al., 2023; Honda et al., 2024) and better reading and language abilities (Kim et al., 2024). Notably, our study is the first to highlight another possible advantage of perceptual consistency: its role in facilitating initial lexical activation. Specifically, in our study, more consistent listeners looked at competitors more before the POD was reached when the word onset was highly misleading – an effect observed in both L1 Spanish with very strong evidence and L2 English with strong evidence in the Bayesian analyses. In this context, the absence of a garden-path can be taken to index delayed initial lexical activation due to the short period of time before the POD (around 400 ms). Therefore, this finding directly speaks to the role of perceptual consistency in early spoken word recognition. Our interpretation is the following: The speech perception system of listeners with lower perceptual consistency is likely characterized by higher noise and/or lower stability of cue-to-category mapping, leading to greater uncertainty. This uncertainty may lead listeners to use a more wait-and-see strategy in early stages of spoken word recognition. In other words, the initial lexical activation is slowed down to avoid committing to a lexical candidate until more information arrives.

This explanation aligns with the ideal observer model, which posits that high uncertainty prompts listeners to adjust their mapping of cue values onto phoneme categories, allowing them to keep multiple options open until further clarifying information is received (Clayards et al., 2008; Nixon et al., 2016). Additionally, this idea is consistent with previous findings showing that listeners delay lexical activation when the input is noisy (e.g., cochlear implant users and normal-hearing listeners presented with noise-vocoded speech; Farris-Trimble et al., 2014; McMurray et al., 2017; Smith & McMurray, 2022). This wait-and-see strategy may be a natural consequence of lower stability in perceptual encoding or a coping mechanism adopted by listeners in high uncertainty situations, allowing them to keep alternatives available and enhancing flexibility in recovering from a misperception (see McMurray et al., 2022, for relevant discussion). Regardless of the specifics of the underlying mechanism, this finding, along with prior research, collectively suggests that phonetic encoding stability confers multiple benefits in speech perception. Furthermore, the present study demonstrated that the functional roles of perceptual consistency in spoken word recognition is not language-specific, as we found evidence for this in both languages.

Gradiency and Perceptual Consistency between L1 and L2

We observed some intriguing findings regarding gradiency and perceptual consistency in L1 vs. L2. L1 and L2 gradiency were not correlated, supported by weak evidence against the alternative hypothesis in the Bayesian analyses. Although we cannot strongly conclude that the correlation was absent or present, the weak evidence may suggest that gradiency is not a generic trait of a listener but rather depends on the specific properties of different contrasts and cues. This interpretation should be approached with caution due to the narrow range of ratings by some participants on the English VAS in this study. However, previous studies also support this interpretation, showing that the correlation between gradiency measures is stronger when contrasts rely on similar acoustic cues (e.g., English /b/-/p/ and English /d/-/t/; $r(64) = .41, p = .001$; Kapnoula & McMurray, 2021) compared to different cues (e.g., English /b/-/p/ and English /s/-/sh/; $r(57) = .19, p = .16$; Kapnoula et al., 2021). This pattern suggests that gradiency depends on how listeners process specific acoustic cues, aligning with the idea that differences in gradiency stem from differences in early encoding of speech cues (Kapnoula & McMurray, 2021). Other studies have found higher correlations between gradiency measures extracted from different languages (English /d/-/t/ and Korean /t/-/th/; $R^2 = .309$; Kong & Kang, 2023) and different contrasts (English /d/-/t/ and English /s/-/sh/; $r(55) = .32, p = .02$; Fuhrmeister & Myers, 2021). However, direct comparisons are challenging due to substantial methodological differences.

On the other hand, for perceptual consistency, we found a significant positive correlation between L1 and L2 supported by decisive evidence in the Bayesian analyses. This finding suggests that perceptual consistency is a generic trait, aligning with the study by

Fuhrmeister et al. (2023), which found a positive correlation in perceptual consistency across different phonetic categories, specifically between stops (“ba”-“pa”) and fricatives (“s”-“sh”). Additionally, Honda et al. (2024) identified a positive correlation in perceptual consistency between 2AFC and VAS tasks, suggesting that cue encoding stability is a stable and task-independent trait. The current study extends these findings by showing the stability of perceptual consistency when processing the same acoustic cue but in different languages. This suggests that cue encoding stability is likely a more general mechanism compared to within-category sensitivity.

More broadly, the different correlation patterns for gradiency vs. perceptual consistency suggest that speech perception cue encoding stability and within-category sensitivity are distinct mechanisms, corroborating previous findings (Fuhrmeister & Myers, 2021; Honda et al., 2024). Indeed, prior work has highlighted the important theoretical distinction between gradiency and perceptual consistency (e.g., Apfelbaum et al., 2022; Fuhrmeister & Myers, 2021; Honda et al., 2024). Our study echoes prior work and significantly expands it by examining the link between gradiency and perceptual consistency in the context of L1 vs. L2 processing. Additionally, our study extends previous work by examining the role of perceptual consistency in speech perception and, more specifically, its relationship with early lexical activation and speech perception flexibility.

The underlying causes and functions of gradiency and perceptual consistency require more in-depth examination. Recent research has started to address these issues. For example, Kapnoula and Samuel (2024) found that gradiency is predicted by (a) temporal auditory acuity (i.e., the domain-general ability to discern subtle acoustic differences across various acoustic dimensions such as pitch, duration, and intensity; Saito, 2023) and (b) cumulative exposure to spoken language (as indicated by age). Additionally, social network diversity has been shown to promote gradient perception (Kutlu et al., 2024). Despite the increased interest in recent years, many open questions remain regarding the origins of gradiency and perceptual consistency, as well as their functional roles in spoken language processing. Addressing these questions will have high theoretical value, leading to a more comprehensive understanding of basic speech perception mechanisms, and could also be instrumental in identifying strategies to enhance L2 acquisition.

Conclusion

The current study provides evidence that both gradiency and perceptual consistency play important roles in spoken word recognition and, most importantly, it sheds new light onto the different nature of these two properties. On the one hand, perceptual consistency is crucial at the early stages of spoken word recognition and appears to be a generic trait of how listeners process speech; listeners with lower consistency are less likely to start activating words early on, possibly to avoid committing to the wrong word, in both L1 and L2. On the other hand, gradiency enhances speech perception flexibility by speeding up recovery from misleading auditory stimuli. Importantly, this effect was (once again) observed in L1, but only weak evidence was found in L2, suggesting that the functional role of gradiency in spoken word recognition is language-specific.

Apart from their theoretical value, these findings have substantial practical implications. In everyday life, individuals often encounter ambiguous and misleading auditory inputs due to factors such as background noise, speaker accents, or coarticulation. Our findings indicate that individuals with higher perceptual consistency and gradiency are better equipped to identify and deal with such complexity.

There is a clear need for further research to examine not only the functions of gradiency and perceptual consistency but also their developmental trajectories in both monolingual and multilingual settings. Addressing these issues will deepen our understanding of speech perception and may also lead to practical applications in language learning.

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *115*(6), 3171–3183. https://doi.org/10.1121/1.1701898

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

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

Apfelbaum, K. S., Klein-Packard, J., & McMurray, B. (2021). The pictures who shall not be named: Empirical support for benefits of preview in the Visual World Paradigm. *Journal of Memory and Language, 121,* Article 104279. https://doi.org/10.1016/j.jml.2021.104279

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

Baese-Berk, M. M., Chandrasekaran, B., & Roark, C. L. (2022). The nature of non-native speech sound representations. *The Journal of the Acoustical Society of America, 152*(5), 3025–3034. https://doi.org/10.1121/10.0015230

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bidelman, G. M., Bernard, F., & Skubic, K. (2024). *Hearing in categories aids speech streaming at the "cocktail party."* bioRxiv. https://doi.org/10.1101/2024.04.03.587795

Boersma, P., & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.03, retrieved 13 January 2023 from http://www.praat.org/

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods, 23*(3), 389–411. https://doi.org/10.1037/met0000159

Brown-Schmidt, S., & Toscano, J. C. (2017). Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience*, *32*(10), 1211–1228. https://doi.org/10.1080/23273798.2017.1325508

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

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809. https://doi.org/10.1016/j.cognition.2008.04.004

Dienes, Z. (2024). Use one system for all results to avoid contradiction: Advice for using significance tests, equivalence tests, and Bayes factors. *Journal of Experimental Psychology: Human Perception and Performance, 50*(5), 531–534. https://doi.org/10.1037/xhp0001202

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Farris-Trimble, A., McMurray, B., Cigrand, N., & Tomblin, J. B. (2014). The process of spoken word recognition in the face of signal degradation. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 308–327. https://doi.org/10.1037/a0034353

Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge University Press.

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

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

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

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology, 32*(3), 401–424. http://dx.doi.org/10.1016/0022-0965(81)90105-3

Hallett, P. E. (1986). Eye movements. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.11–10.112). Wiley.

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

Honda, C. T., Clayards, M., & Baum, S. R. (2024). Exploring individual differences in native phonetic perception and their link to nonnative phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0001191

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. http://dx.doi.org/10.1006/jecp.1999.2553

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

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

Kapnoula, E. C., & Samuel, A. G. (2024). Sensitivity to subphonemic differences in first language predicts vocabulary size in a foreign language. *Language Learning.* Advance online publication. https://doi.org/10.1111/lang.12650

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

Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics, 81,* Article 100984. https://doi.org/10.1016/j.wocn.2020.100984

Kim, H., Klein-Packard, J., Sorensen, E., Oleson, J., Tomblin, B., & McMurray, B. (2024). *Inconsistent speech categorization in school-age children with language and reading disabilities.* PsyArXiv. https://doi.org/10.31234/osf.io/un6bx

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

Kong, E. J., & Kang, S. (2023). Individual differences in categorical judgment of L2 stops: A link to proficiency and acoustic cue-weighting. *Language and Speech, 66*(2), 354–380. https://doi.org/10.1177/00238309221108647

Kutlu, E., Apfelbaum, K. S., Sorensen, E., Oleson, J., & McMurray, B. (2024). *Social network diversity leads to more flexible speech perception in school-aged children.* PsyArXiv. https://doi.org/10.31234/osf.io/c9u4y

Kutlu, E., Chiu, S., & McMurray, B. (2022). Moving away from deficiency models: Gradiency in bilingual speech categorization. *Frontiers in Psychology, 13,* 1033825. https://doi.org/10.3389/fpsyg.2022.1033825

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

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: General, 61(5), 379–388. https://doi.org/10.1037/h0049038

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/in trochees. *Language and Speech*, *29*(1), 34–11. https://doi.org/10.1177/002383098602900102

Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America, 134*(3), 2213–2224. https://doi.org/10.1121/1.4817845

López-Zamora, M., Luque, J. L., Álvarez, C. J., & Cobos, P. L. (2012). Individual differences in categorical perception are related to sublexical/phonological processing in reading. *Scientific Studies of Reading, 16*(5), 443–456. http://dx.doi.org/10.1080/10888438.2011.588763

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

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94,* 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McMurray, B. (2017). *Nonlinear Curvefitting for Psycholinguistic (and other) Data.*

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

McMurray, B., Apfelbaum, K. S., & Tomblin, J. B. (2022). The slow development of real-time processing: Spoken-word recognition as a crucible for new thinking about language acquisition and language disorders. *Current Directions in Psychological Science*, *31*(4), 305–315. https://doi.org/10.1177/09637214221078325

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. https://doi.org/10.1037/a0011747

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

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition*, *169*, 147–164. https://doi.org/10.1016/j.cognition.2017.08.013

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology, 60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42. https://doi.org/10.1016/s0010-0277(02)00157-9

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition.

Journal of Memory and Language, 60(1), 65–91. https://doi.org/10.1016/j.jml.2008.07.002

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition, 50*(1–3), 271–285. https://doi.org/10.1016/00100277(94)90031-0

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, *43*(1–3), 106–115. https://doi.org/10.1159/000261764

Munson, B., Schellinger, S. K., & Edwards, J. (2017). Bias in the perception of phonetic detail in children's speech: A comparison of categorical and continuous rating scales. *Clinical Linguistics & Phonetics*, *31*(1), 56–79. https://doi.org/10.1080/02699206.2016.1233292

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, *90*, 103–125. https://doi.org/10.1016/j.jml.2016.03.005

Ou, J., Yu, A. C., & Xiang, M. (2021). Individual differences in categorization gradience as predicted by online processing of phonetic cues during spoken word recognition: Evidence from eye movements. *Cognitive Science, 45*(3), Article e12948. https://doi.org/10.1111/cogs.12948

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

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

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285–290. https://doi.org/10.3758/bf03213946

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

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language*, *10*, 243–335. https://doi.org/10.1016/b978-0-12-608610-2.50012-1

Rizzi, R., & Bidelman, G. M. (2024). *Functional benefits of continuous vs. categorical listening strategies on the neural encoding and perception of noise-degraded speech.* bioRxiv. https://doi.org/10.1101/2024.05.15.594387

RStudio Team. (2023). *RStudio: Integrated Development Environment for R*. Boston, MA. Retrieved from http://www.rstudio.com/

Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing auditory acuity hypothesis-L2. *Language Teaching*, 1–17. https://doi.org/10.1017/s0261444822000453

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Samuel, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics, 22*(4), 321–330. https://doi.org/10.3758/BF03199697

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*(4), 307–314. https://doi.org/10.3758/bf03202653

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

Schouten, M. E. H., & van Hessen, A. J. (1992). Modeling phoneme perception. I: Categorical perception. *The Journal of the Acoustical Society of America*, *92*(4), 1841–1855. https://doi.org/10.1121/1.403841

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. http://dx.doi.org/10.1044/1092-4388(2001/032)

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition, 98*(2), B35–B44. http://dx.doi.org/10.1016/j.cognition.2005.03.002

Smayda, K. E., Chandrasekaran, B., & Maddox, W. T. (2015). Enhanced cognitive and perceptual processing: a computational basis for the musician advantage in speech learning. *Frontiers in Psychology, 6,* Article 682. https://doi.org/10.3389/fpsyg.2015.00682

Smith, F. X., & McMurray, B. (2022). Lexical access changes based on listener needs: Real-time word recognition in continuous speech in cochlear implant users. *Ear and Hearing*, *43*(5), 1487–1501. https://doi.org/10.1097/aud.0000000000001203

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*. https://doi.org/10.1177/09622802241242319

Souganidis, C., Molinaro, N., & Stoehr, A. (2022). Bilinguals produce language-specific voice onset time in two true-voicing languages: The case of Basque-Spanish early bilinguals. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.21081.sou

SR Research Ltd. (2022). Experiment Builder (Version 2022.2.5) [Computer software]. https://www.sr-research.com/experiment-builder/

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review, 77*(3), 234–249. https://doi.org/10.1037/h0029078

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/ 10.1126/science.7777863

Theodore, R. M., & Monto, N. R. (2019). Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review*, *26*(3), 985–992. https://doi.org/10.3758/s13423-018-1551-5

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization. *Psychological Science, 21*(10), 1532–1540. https://doi.org/10.1177/0956797610384142

Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41*(1), 48–61. http://dx.doi.org/10.1037/h0084150

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, http://ggplot2.org.

Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America, 147*(2), 852–866. https://doi.org/10.1121/10.0000692