Cognition 276 (2026) 106671

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Full Length Article

Towards an attentional theory of second-language speech perception: Evidence from cue ecology

Annie Tremblay [a,*], Taehong Cho [b], Mirjam Broersma [c], Hyoju Kim [d], Quentin Zhen Qin [e], Jiayu Liang [e]

[a] The University of Texas at El Paso, United States of America
[b] Hanyang University, South Korea
[c] Radboud University Nijmegen, the Netherlands
[d] The University of Iowa, United States of America
[e] The Hong Kong University of Science and Technology, Hong Kong, China

ARTICLE INFO

Keywords:
Cue weighting
Lexical stress
Second-language speech perception
Attentional learning
Cross-linguistic transfer
Prosody
Bilingualism

ABSTRACT

Speech perception relies on multiple acoustic cues whose relative weighting varies across languages. The present study examines how long-term language experience shapes cue weighting in second-language (L2) speech perception, refining an attentional-learning account of cross-linguistic transfer. Native speakers of English, Dutch, Spanish, Korean, and Mandarin completed a cue-weighting task targeting English lexical stress, in which vowel quality, pitch, and duration were orthogonally manipulated. Results revealed robust, dimension-specific differences across first-language (L1) groups that could not be explained solely by the presence or absence of lexical stress or lexical tone in the L1. Instead, cue weighting reflected how acoustic dimensions function within the L1 cue ecology, including their relative contribution to lexical distinctions and the stability and interpretability of these mappings across contexts. Cue redundancy constrained relative cue strength without eliminating attentional sensitivity to secondary dimensions. Machine-learning classification further showed that L1-linked attentional profiles were sufficiently structured to support prediction, even among L2 listeners with substantial English proficiency, demonstrating the persistence of L1-shaped attentional tuning. These findings support a view of cue weighting as reflecting durable, multidimensional attentional priors shaped by long-term experience and highlight the importance of L1 cue ecologies in understanding cross-linguistic transfer in speech perception.

1. Introduction

Speech perception poses a fundamental cognitive problem: Listeners must map a continuous, multidimensional acoustic signal onto discrete categories under conditions of variability and uncertainty. The *acoustic dimensions* that make up this signal are physical continua along which speech varies. Because multiple acoustic dimensions are available at any given moment, successful perception requires selective attention (i.e., the prioritization of some dimensions and the suppression of others) so that variation along those dimensions can be used as meaningful *acoustic cues* for categorization; from this perspective, cue weighting reflects selective attentional tuning to acoustic dimensions based on their relevance for recovering category distinctions from the input rather than differences in acoustic sensitivity alone (e.g., Clayards et al., 2008; Francis et al., 2000; Francis & Nusbaum, 2002; Holt & Lotto, 2006; Ingvalson et al., 2011; Iverson et al., 2003; Schertz & Clare, 2020).

Crucially, the allocation of attention to acoustic dimensions is shaped by learning (Holt, 2025; Kuhl, 2000). Through exposure to the statistical structure of the native language (L1), listeners develop dimension-specific attentional priors that reflect (i) the relative contribution of different dimensions to distinguishing lexical categories, (ii) the stability and interpretability of these mappings across contexts, and (iii) the extent to which they support robust category learning despite covariation in the signal, as proposed in attention-to-dimension models of categorization (e.g., Clayards et al., 2008; Francis et al., 2000; Francis & Nusbaum, 2002; Holt, 2025; Holt & Lotto, 2006; Kuhl, 2000; Nosofsky, 1986; Roark & Holt, 2022). These language-specific properties determine the *learnability* of an acoustic cue—that is, the extent to which listeners can discover and maintain reliable mappings between a specific acoustic dimension and linguistically meaningful distinctions.

* Corresponding author at: Linguistics, The University of Texas at El Paso, 500 W. University Ave., El Paso, TX 79968, United States of America.
E-mail address: actremblay@utep.edu (A. Tremblay).

https://doi.org/10.1016/j.cognition.2026.106671
Received 2 February 2026; Received in revised form 19 June 2026; Accepted 20 July 2026

Dimensions that reliably support lexical interpretation under these conditions tend to be prioritized, whereas dimensions whose interpretation depends on additional contextual normalization or that provide largely redundant information may be downweighted over time. Once established, these attentional biases tend to be stable and resistant to drastic changes, persisting when listeners encounter a second language (L2) in which the functional importance of the same dimensions differs (e.g., Ingvalson et al., 2011; Ingvalson et al., 2012; Iverson et al., 2003; Iverson et al., 2005). Within this framework, cross-linguistic transfer can be understood as *attentional inertia*: Listeners continue to allocate attention according to L1-trained priorities even when doing so is suboptimal for L2 category learning (Chang, 2018; W. Choi, 2022; Ingvalson et al., 2011; Iverson et al., 2003; Kim & Tremblay, 2021, 2022; Qin et al., 2017; Schertz et al., 2015; Tremblay et al., 2021; Tremblay & Kim, 2025).

The current study seeks to explain how such long-term, dimension-specific attentional tuning shapes listeners' use of acoustic cues when perceiving categories in an L2. To address this question, we focus on lexical stress as a particularly informative test case. Lexical stress contrasts are realized through multiple acoustic cues that are simultaneously present in the signal, yet languages differ markedly in the relative importance, contextual stability, and learnability of these cues for interpreting word-level distinctions. For example, pitch carries a high functional load for signaling lexical contrasts in tone languages such as Chinese (e.g., Chao, 1968; Duanmu, 2007) but not in non-tonal languages such as English. From an attentional learning perspective, such cross-linguistic differences create language-specific expectations about the relative informativeness of acoustic dimensions for extracting meaning, leading listeners to prioritize those dimensions in the L2. Accordingly, Chinese listeners rely heavily on pitch in English even if pitch is generally not informative for identifying English words (W. Choi et al., 2019; W. Choi, 2022; Ortega-Llebaria & Wu, 2021; Qin et al., 2017; Wang et al., 2024; Zhang & Francis, 2010). In this view, cross-linguistic transfer reflects the persistence of dimension-specific attentional tuning. Lexical stress thus offers a window into how attentional tuning to acoustic dimensions supports category learning and transfer in speech perception.

A number of studies have examined cross-linguistic and cross-dialectal differences in the perception of lexical stress (W. Choi et al., 2019; Chrabaszcz et al., 2014; Dupoux et al., 2008, Dupoux et al., 2010; Guo & Chen, 2022; Kim & Tremblay, 2021, 2022; Qin et al., 2017; Tremblay et al., 2021; Zhang & Francis, 2010). These studies have yielded important insights into how prior linguistic experience shapes stress perception. However, this literature has made use of a wide range of experimental paradigms (e.g., discrimination, identification, sequence recall) and auditory stimuli (e.g., real words vs. nonce words, naturally produced vs. resynthesized, with vs. without orthogonal cue manipulation), and it has not been organized around an overarching framework aimed at explaining how L1-based perceptual differences arise from general principles of attentional learning and cue learnability. As a result, it remains difficult to integrate results across studies and to determine whether observed differences reflect broader, recurring patterns of dimension-specific attentional tuning across languages. Moreover, it remains unclear whether such patterns are sufficiently systematic and distinctive to support reliable identification of listeners' linguistic backgrounds.

The present study addresses this gap by advancing a unified attentional framework grounded in the learnability of acoustic cues and testing it across multiple L1 backgrounds within a common experimental paradigm. This framework builds on transfer-based approaches but differs from them in its level of explanation. Traditional accounts of L1 transfer characterize cross-linguistic influence in terms of the presence or absence of phonological categories or typological properties (e.g., whether or not the L1 has lexical stress, and if so, whether lexical stress is predicable from phonological rules) (e.g., Peperkamp et al., 2010; Peperkamp & Dupoux, 2002). In contrast, the current framework shifts the focus to the distributional and functional properties of individual acoustic dimensions within the L1 and to how these properties shape learning. From this perspective, what is transferred is not category membership but *dimension-specific attentional priors acquired through exposure to the statistical structure of the L1*. Importantly, this view conceptualizes transfer as the emergent consequence of domain-general attentional learning processes that give rise to stable, language-specific weighting of acoustic cues (see also Holt & Lotto, 2006).

To test this framework, we examine how listeners from five phonologically distinct L1s—English, Dutch, Mandarin, Spanish, and Korean—weight vowel quality, pitch, and duration cues in the perception of English lexical stress.[1] These languages, as will be demonstrated below, differ not only in whether they have lexical stress, but also in the extent to which each of these acoustic dimensions contributes to word-level contrasts in the L1, providing a principled basis for testing an attentional-learning account. Using a cue-weighting speech perception experiment with orthogonal manipulations of vowel quality, pitch, and duration, we directly estimate how listeners allocate attention to these acoustic dimensions when perceiving English lexical stress. From these estimates, we derive a set of dimension-specific attentional profiles that capture systematic patterns in how attention to acoustic cues varies as a function of listeners' prior linguistic experience and the functional role of these cues in the L1. To evaluate the robustness and explanatory value of these profiles, we then apply machine-learning classification models to listeners' cue weights to assess whether they reliably predict listeners' L1 background (e.g., Yarkoni & Westfall, 2017). This data-driven validation approach tests whether the observed profiles are sufficiently structured and distinctive to support cross-linguistic prediction. These analyses allow us to assess how long-term attentional tuning to acoustic dimensions constrains cue weighting in L2 speech perception and gives rise to recurring, L1-linked attentional profiles.

To generate a priori predictions for each listener group, we review evidence from the L1 literature bearing on the role of vowel quality, pitch, and duration in each language. This review seeks to identify the properties of the L1 cue ecology that are most relevant for shaping long-term attentional tuning.

2. L1 cue ecologies and attentional tuning

In the present framework, a language's cue ecology refers to the functional role that acoustic dimensions play in supporting linguistically meaningful distinctions and category learning. Cue ecologies are therefore not restricted to lexical stress. Rather, they encompass any domain in which a dimension becomes a cue that provides systematic and recoverable information for distinguishing linguistic categories, including lexical tone, lexical pitch accents, and segmental contrasts. The review below focuses specifically on those aspects of each language's cue ecology that are most relevant for understanding how listeners may develop long-term attentional tuning to the three dimensions that signal lexical stress in English—vowel quality, pitch, and duration.

[1] Although lexical stress in English has been associated with additional acoustic correlates, including intensity, spectral tilt, and hyper-articulation (e.g., Beckman, 1986; de Jong, 1995; Sluijter & van Heuven, 1996b), the present study focuses on vowel quality, pitch, and duration because these dimensions have been most consistently examined in previous cross-linguistic research on lexical stress perception and have been shown to play a central role in listeners' perceptual weighting of stress contrasts (e.g., Chrabaszcz et al., 2014; Tremblay et al., 2021; Zhang & Francis, 2010). Moreover, intensity is highly correlated with pitch, and its contribution to stress perception is generally weaker and less consistent than that of the dimensions examined here (e.g., Zhang & Francis, 2010). Restricting the analysis to these three dimensions therefore allows for more direct comparison with previous studies and provides a tractable framework for investigating cross-linguistic differences in attentional tuning.

2.1. English

English has contrastive lexical stress: Shifting stress across syllables can change a word's meaning (e.g., *DEsert* vs. *deSSERT*) or its grammatical category (e.g., *REcord* vs. *reCORD*). Lexical stress distinctions in English are supported by multiple acoustic dimensions, including vowel quality, pitch, duration, and intensity (Lehiste, 1970), with vowel quality constituting a central dimension for lexical interpretation: Vowels in stressed syllables are realized as full vowels, whereas vowels in unstressed syllables are often reduced to a schwa (i.e., [ə]) (e.g., Lindblom, 1963). This pattern creates a robust association between vowel quality and lexical stress, such that attending to vowel quality is critical for distinguishing stress contrasts (e.g., Bond, 1981; Bond & Small, 1983; Fear et al., 1995; Field, 2005; Ghosh & Levis, 2021; Small et al., 1988). Although the degree of vowel reduction may vary, the contrast between full and reduced vowels provides a consistent and interpretable cue to lexical stress across intonational positions, making vowel quality a highly learnable dimension for signaling lexical distinctions during acquisition (e.g., Quam & Swingley, 2014). Consistent with these properties, perception studies show that native English listeners assign the greatest weight to vowel quality when identifying lexical stress, relying on it more strongly than on suprasegmental cues such as duration or pitch (Chrabaszcz et al., 2014; Tremblay et al., 2021; Zhang & Francis, 2010).

Duration is also a correlate of lexical stress in English, as stressed syllables are typically longer than unstressed ones (e.g., Beckman & Edwards, 1994; Lieberman, 1960). However, durational differences are strongly modulated by sentence-level prosodic structure (e.g., phrase-final lengthening) as well as global temporal factors such as speech rate (Cambier-Langeveld, 1997; Nakatani et al., 1981; Paschen et al., 2022; Turk & Shattuck-Hufnagel, 2007), which together can complicate the interpretation of duration cues at the lexical level. Moreover, duration covaries with vowel quality (i.e., reduced vowels are shorter) (e.g., Lindblom, 1963), such that much of its contribution to stress is redundant with segmental information. Correspondingly, although duration is a reliable correlate of stress in English, listeners tend to assign it relatively lower perceptual weight (Chrabaszcz et al., 2014; Tremblay et al., 2021; Zhang & Francis, 2010), likely due to its contextual variability and redundancy with vowel quality.

Pitch in English is primarily recruited to mark intonational structure, including the distribution of pitch accents and prosodic boundaries, as well as information-structure–driven focus effects (e.g., Beckman, 1986; Beckman & Pierrehumbert, 1986; Cole, 2026; Fletcher, 2010; Ladd, 2012; Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990). As a result, the pitch contour associated with a given word varies substantially depending on the prosodic structure in which it occurs and its discourse function. Additionally, not all content words receive a pitch accent in English (e.g., Beckman & Elam, 1997; Ladd, 2012), and pitch must be interpreted relative to the broader intonational context (e.g., Beier & Ferreira, 2018; Brown et al., 2016). Crucially, because vowel quality provides robust, context-stable information about lexical stress in English, pitch does not need to be strongly weighted for listeners to recover stress distinctions. Accordingly, pitch is used as a secondary cue to lexical stress (W. Choi et al., 2019; Chrabaszcz et al., 2014; Tremblay et al., 2021; Zhang & Francis, 2010).

2.2. Dutch

Dutch also has contrastive lexical stress, such that shifting stress across syllables can distinguish lexical items (e.g., *CAnon* 'canon' vs. *kaNON* 'cannon') (Gussenhoven, 2008). In Dutch, vowel quality is a weaker cue to lexical stress than in English: Although unstressed syllables tend to contain more centralized vowels than stressed syllables, the degree of vowel reduction is substantially smaller in Dutch, with centralized vowels often not being reduced to a schwa (e.g., Kager, 1989; Sluijter & van Heuven, 1996a; van Bergem, 1993). Additionally, compared to English, fewer Dutch words have reduced vowels in unstressed syllables (Campbell & Beckman, 1997). As a result, vowel quality provides less categorical information about lexical stress (van Heuven & de Jonge, 2011), limiting its contribution to distinguishing lexical items.

Duration plays a prominent role in the realization of lexical stress in Dutch: Stressed syllables are typically longer than unstressed syllables, and durational differences contribute substantially to stress perception (e.g., Sluijter & van Heuven, 1996b; van Heuven & de Jonge, 2011). Given that vowel-quality cues to lexical stress are weaker in Dutch, listeners allocate greater perceptual weight to duration cues than to vowel quality cues (van Heuven & de Jonge, 2011). At the same time, durational patterns are modulated by prosodic context, including phrase-final lengthening and speech rate (e.g., Cambier-Langeveld, 1997), which introduces variability in their surface realization. From an attentional perspective, duration thus provides probabilistic but informative evidence relevant to lexical stress.

As with English, pitch in Dutch is primarily recruited to signal intonational structure: The pitch contour associated with a word varies as a function of its position in the utterance and its discourse function (e.g., Gussenhoven, 2004). Like duration, pitch does not encode lexical stress directly but provides probabilistic information insofar as pitch accents are preferentially aligned with metrically strong syllables. From an attentional perspective, pitch and duration thus appear to occupy similar functional roles in the Dutch cue ecology: Both provide indirect, context-sensitive cues to prominence that can support stress perception when segmental evidence is weaker. Consistent with this view, Dutch listeners have been found to make effective use of suprasegmental cues to lexical stress in Dutch (Cutler & Van Donselaar, 2001; Jesse et al., 2017; Reinisch et al., 2010; van Donselaar et al., 2005). However, to our knowledge, previous research has not directly compared Dutch listeners' relative weighting of pitch and duration cues to lexical stress in Dutch.

2.3. Spanish

Spanish also has contrastive lexical stress, such that shifting stress across syllables can signal a change in meaning (e.g., *PApa* 'pope' vs. *paPÁ* 'daddy') or encode morphosyntactic information such as verb tense and person (e.g., *LLEgo* '[I] arrive' vs. *lleGÓ* '[s/he] arrived') (Harris, 1983; Hualde, 2005). Unlike English and Dutch, however, Spanish does not exhibit vowel reduction in unstressed syllables: Vowels maintain their full articulatory quality regardless of stress placement (Harris, 1983; Hualde, 2005). As a result, vowel quality does not encode lexical stress in Spanish and provides no systematic basis for distinguishing words that differ in lexical stress. Spanish also lacks a centralized reduced vowel comparable to the English schwa, and therefore such a vowel does not distinguish words at the segmental level.

In contrast, duration plays a central role in the realization of lexical stress in Spanish. Stressed syllables are consistently longer than unstressed syllables across intonational contexts, indicating that this durational contrast is stable and easily recoverable (e.g., Ortega-Llebaria & Prieto, 2007, 2011). Although durational differences are relative and can be modulated by speech rate and prosodic context, the absence of vowel-quality cues to stress in Spanish increases the functional importance of duration within the stress cue ecology. From an attentional perspective, this means that Spanish listeners allocate a moderately high perceptual weight to duration as a source of stress information (Ortega-Llebaria et al., 2013; Ortega-Llebaria & Prieto, 2009), even though duration is still variable and context-dependent.

As with English and Dutch, pitch in Spanish is primarily associated with accenting and intonational structure rather than with lexical stress: Intonational pitch accents are not obligatorily realized on all stressed syllables, and the presence, shape, and alignment of pitch movements depend on discourse and prosodic context (e.g., Face & Prieto, 2007; Hualde & Prieto, 2015; Ortega-Llebaria & Prieto, 2007; Prieto &

Torreira, 2007). As a result, pitch provides probabilistic information about syllabic prominence. From an attentional perspective, pitch thus serves as a secondary, context-dependent source of information for stress perception (e.g., Ortega-Llebaria et al., 2013).

2.4. Seoul Korean

Seoul Korean (henceforth, Korean) does not have lexical stress or lexical tones. Instead, pitch patterns form part of the language's intonational system (e.g., Cho, 2011; Jun, 1998, 2000, 2005). Importantly, these pitch patterns contribute to lexically contrastive segmental distinctions, differentiating the three-way laryngeal contrast between lenis, fortis, and aspirated stops (e.g., Cho, 2016; Kim et al., 2024; H. Lee & Jongman, 2019; Lisker & Abramson, 1964). In contemporary Korean, Voice Onset Time (VOT) differences between lenis and aspirated stops have largely merged, shifting the primary burden of contrast maintenance onto the pitch of the following vowel (e.g., K.-H. Kang & Guion, 2008; Y. Kang, 2014b; Silva, 2006). Perception studies confirm this functional role: Korean listeners rely on pitch as a primary cue for distinguishing lenis-aspirated stops and on both pitch and VOT for fortis-lenis contrasts (Kim et al., 2024; H. Lee et al., 2013; Schertz et al., 2015). Importantly, this use of pitch is largely restricted to phrase-initial position, where it is conditioned by intonational phonology—specifically, the assignment of a high tone following aspirated or fortis stops and a low tone following lenis stops, rendering the pitch cue position-specific rather than uniformly contrastive (J. Choi et al., 2020). These properties support some attentional tuning to pitch as a dimension that encodes linguistically meaningful distinctions in the L1, although the expected allocation of attention may be conditioned by the prosodic positions in which such distinctions are relevant.

Unlike pitch, duration plays a limited role in Korean, marking the final boundary of words at the end of the intonational phrase but not at lower intonational levels (e.g., Cho & Keating, 2001). Although Korean historically had a phonemic vowel length contrast, this contrast has largely neutralized in contemporary Korean and no longer distinguishes lexical items for most speakers (e.g., Y. Kang et al., 2015; Shin et al., 2012; Sohn, 1999). As a result, duration does not serve as an informative dimension for lexical interpretation in modern Korean, suggesting weak weighting of this dimension.

While Korean has traditionally been described as having a central vowel /ʌ/, which is spectrally closer to the English schwa relative to other Korean vowels and serves to distinguish lexical contrasts (e.g., Yang, 1996), more recent evidence suggests that in contemporary Korean /ʌ/ is realized further back and patterns with the back vowel class (Y. Kang, 2014a), while still remaining spectrally and perceptually closer to English schwa than other Korean vowels.

2.5. Mandarin

Lexical tones serve a primary contrastive function in Mandarin words: Lexical identity is overwhelmingly determined by pitch patterns, such that differences in pitch height and contour distinguish words (e.g., *mā* 'mother' vs. *má* 'hemp' vs. *mǎ* 'horse' vs. *mà* 'scold') (Chao, 1968; Duanmu, 2007). As a result, pitch must be attended to closely for successful Mandarin word recognition (e.g., Howie, 1976; Liu & Samuel, 2004). Lexical tones are realized on individual syllables and, despite systematic contextual variation (e.g., coarticulation, tone sandhi; Y. Xu, 1997), pitch patterns map in a stable way onto lexical distinctions across prosodic contexts (e.g., Y.-C. Lee et al., 2016; Ouyang & Kaiser, 2015; Y. Xu, 1999). Duration also plays a role in the perception of lexical tones (e.g., Howie, 1976; Liu & Samuel, 2004; Nordenhake & Svantesson, 1983; Shen, 1993). However, durational differences are far less consistent than pitch, limiting their reliability for distinguishing tonal categories (e.g., Liu & Samuel, 2004; Shen, 1993). Accordingly, Mandarin listeners develop early learning and strong, long-term attentional tuning to pitch as the primary carrier of lexical contrasts (e.g., Feng & Peng, 2023; Ma et al., 2021).

Some dialects of Mandarin (e.g., Beijing Mandarin) have also been proposed to have lexically contrastive stress, where stressed-stressed disyllabic words contrast with stressed-unstressed ones (e.g., respectively, ['tʊ¯ŋ'ɕī] 'west and east' vs. ['tʊ¯ŋɕi] 'stuff,' both written as 东西) (Chao, 1968; Duanmu, 2007). In these varieties, duration is the primary cue to lexical stress, while pitch plays a secondary role due to the presence of the neutral tone in unstressed syllables (e.g., Y. Chen & Xu, 2006; W.-S. Lee & Zee, 2010; T. Lin, 1985). The vowel in unstressed syllables may also be centralized (e.g., Duanmu, 2007; C. Xu, 2024). However, because vowel reduction in Mandarin is closely tied to the occurrence of the neutral tone and tonal contrasts already provide robust, lexically contrastive information, any vowel-quality variation associated with neutral-tone syllables is largely redundant and does not contribute independent information for distinguishing stress. Importantly, few syllable combinations of Mandarin syllables form disyllabic minimal pairs differing only in lexical stress, and the perception of lexical stress appears driven by the presence of the neutral tone rather than by duration cues alone (e.g., Yan, 2021). Thus, while duration provides the primary cue to lexical stress in Mandarin, this role is restricted to a limited set of contrasts and systematically covaries with pitch, underscoring the central role of pitch-related cues in lexical interpretation.

3. Cross-linguistic predictions for attentional tuning

The predictions presented below are derived from converging evidence from the quantitative, acoustic and perceptual studies reviewed above on the role of vowel quality, pitch, and duration in each language. Under an attentional-learning account, cross-linguistic differences in L2 lexical stress perception are expected to emerge as systematic differences in attention to individual acoustic dimensions, shaped by (i) their relative contribution to distinguishing lexical categories, (ii) their contextual stability and interpretability, and (iii) their learnability in the L1 despite covariation in the signal. The present framework generates predictions about how each cue is expected to be weighted across listener groups, based on their prior experience with that dimension as a carrier of lexical information. Accordingly, the predictions below are framed cue by cue, comparing how the same acoustic dimension is weighted across L1 groups, rather than comparing the relative strength of different cues within a single language. Comparisons across cues are not the focus of the present predictions because absolute cue weights are inherently task- and signal-dependent, reflecting properties of the experimental design and auditory stimuli (Holt & Lotto, 2006) as much as of listeners' long-term representations. In contrast, attentional-learning accounts make principled predictions about how the same acoustic dimension is differentially weighted across L1 groups as a function of prior linguistic experience.

These predictions are summarized in Table 1 and justified below. Existing L2 findings are cited as converging evidence.

3.1. Vowel quality

We predict substantial cross-linguistic variation in attention to vowel quality as a cue to English lexical stress, reflecting differences in the lexical function of vowel quality in the L1. English listeners are expected to show the strongest weighting of vowel quality of all groups, as vowel

Table 1
Predicted L1 rankings in acoustic cue reliance for the perception of English lexical stress.

Cue	Predicted L1 Ranking
Vowel Quality	English > Dutch > Korean > Spanish, Mandarin
Pitch	Mandarin > Korean > Dutch, Spanish > English
Duration	Spanish > English ≥ Dutch > Mandarin > Korean

reduction provides a robust, context-stable cue to stress that directly supports lexical distinctions and facilitates early and reliable category learning.

Dutch listeners are predicted to show weaker weighting of vowel quality than English listeners given the less categorical and frequent nature of vowel reduction in Dutch, but stronger weighting than the other L1 groups, reflecting the lexical nature of this cue in Dutch. In prior analyses of the same participant sample, Dutch L2 learners of English showed reduced sensitivity to vowel quality cues to lexical stress relative to native English listeners (Tremblay et al., 2021). However, those analyses did not situate Dutch listeners' use of vowel quality within a broader cross-linguistic framework, nor did they compare Dutch listeners directly with learners from multiple other L1 backgrounds. The present study seeks to fill that gap.

Korean listeners are predicted to show more limited sensitivity to vowel quality cues in English lexical stress perception. While vowel quality does not encode lexical prominence in Korean, the central-back vowel category in Korean may assimilate to the English schwa (e.g., Flege & Bohn, 2021; Tyler et al., 2014; Yu, 2025). This cue ecology supports greater attention to vowel quality in Korean than in Spanish or Mandarin but weaker attention than in English or Dutch. Previous spoken word recognition research has demonstrated that Korean L2 learners of English are better able to use vowel quality cues to English lexical stress than Mandarin L2 learners of English (Connell et al., 2018). However, other studies have shown no advantage for Korean over Mandarin listeners in sensitivity to vowel-quality changes resulting from incorrect stress placement in English words (C. Y. Lin et al., 2014), suggesting that Korean listeners' greater use of vowel quality information may be task-dependent.

Spanish and Mandarin listeners are expected to show the least reliance on vowel quality but for different reasons. In Spanish, vowel quality does not vary with lexical stress and therefore provides no basis for developing attentional tuning to this dimension as a prominence cue. Additionally, Spanish does not have a centralized vowel to which the English schwa could assimilate. Hence, Spanish listeners should have difficulty using vowel reduction to differentiate English words. Consistent with this prediction, previous research has shown that Spanish L2 learners of English display reduced sensitivity to vowel quality cues in the perception of English lexical stress (Tremblay & Kim, 2025).

In Mandarin, vowel quality variation associated with unstressed syllables is largely redundant with tonal information (i.e., the neutral tone), limiting its independent contribution to lexical interpretation. Additionally, relative to Spanish, English, and Dutch, Mandarin provides relatively sparse lexical evidence that lexical stress distinguishes word meanings, limiting the usefulness of vowel quality as a cue to stress across the lexicon. The redundancy of vowel reduction and pitch, together with the sparser evidence for lexical stress, should result in lower long-term attentional weighting of vowel quality as a cue to lexical stress relative to the other L1 groups. Previous research has shown that Mandarin L2 learners of English can use vowel quality to identify nonce words and English words differing in lexical stress (Chrabaszcz et al., 2014; Zhang & Francis, 2010). However, Mandarin listeners' lexical activation and competition do not appear sensitive to vowel quality cues to stress (Connell et al., 2018; C. Y. Lin et al., 2014), suggesting that Mandarin listeners' use of such cues in English may not be robust.

3.2. Pitch

Pitch is predicted to show important cross-linguistic differences in attentional weighting, reflecting differences in its functional load, interpretability, and learnability across languages. Mandarin listeners are expected to show the strongest reliance on pitch, as pitch patterns encode lexical identity and exhibit stable, interpretable mappings across contexts, supporting robust and early category learning. This strong attentional prioritization is predicted to persist in L2 speech perception even when pitch does not encode lexical stress. Existing L2 research provides converging evidence for this prediction. Multiple studies have shown that Mandarin L2 learners of English rely heavily on pitch cues when perceiving English lexical stress (Chrabaszcz et al., 2014; Qin et al., 2017; Zhang & Francis, 2010).

Korean listeners are predicted to show moderate reliance on pitch. Although pitch does not encode lexical tone or stress in Korean, it plays a crucial role in signaling lexically contrastive segmental distinctions in the laryngeal system. These properties support attentional tuning to pitch as a linguistically meaningful dimension—more so than in Dutch, Spanish, and English but less so than in Mandarin. Existing L2 research supports this prediction. Kim and Tremblay (2022) found that Seoul Korean L2 learners of English outperformed proficiency-matched French L2 learners of English in the perception of English lexical stress cued by pitch. Korean and French have broadly similar intonational systems but differ in that pitch encodes a lexically contrastive segmental distinction in Korean but not French. Korean listeners' superior performance was attributed to the lexical relevance of pitch in Korean but not in French. At the same time, Korean listeners' use of pitch in the perception of lexical stress is not predicted to be uniformly robust. Kim and Tremblay (2021) found that Seoul Korean listeners had greater difficulty than Gyeongsang Korean listeners in perceiving pitch-cued lexical stress, a difference attributed to the presence of lexically contrastive pitch accents in Gyeongsang Korean but not Seoul Korean. Taken together with evidence that pitch-based contrasts in Seoul Korean are restricted to specific prosodic positions (i.e., in phrase-initial contexts; J. Choi et al., 2020) and to a limited set of segmental contrasts (e.g., lenis–aspirated stops), these properties support limited, position-conditioned attentional tuning to pitch as a linguistically meaningful dimension in the L1. Previous research has also shown that Korean listeners make limited use of suprasegmental cues to English lexical stress in lexical tasks (Connell et al., 2018; C. Y. Lin et al., 2014). These findings are consistent with a predicted level of attentional weighting for pitch in Korean that is greater than in languages where pitch lacks a lexical function, but weaker than in languages where pitch carries a high functional weight.

Dutch and Spanish listeners are expected to show weaker but non-negligible reliance on pitch. In both languages, pitch does not encode lexical stress directly but provides probabilistic information about prominence through its proximal alignment with stressed syllables when pitch accents are realized. From an attentional perspective, pitch should therefore serve as a secondary, context-dependent cue with some degree of informativeness, supporting moderate attentional weighting. This moderate weighting is further reinforced by the reduced usefulness of vowel quality cues to stress in Dutch and the absence of such cues in Spanish. Existing findings are consistent with these predictions. In our previous work, we found that the current Dutch L2 learners of English showed greater sensitivity to pitch-cued lexical stress than native English listeners (Tremblay et al., 2021), consistent with the predicted weighting of pitch but leaving open the question of how Dutch listeners would compare to other L1 groups. Tremblay and Kim (2025) similarly reported that Spanish L2 learners of English relied more on pitch than native English listeners, providing converging evidence for moderate attentional weighting of pitch in Spanish but not situating Spanish listeners within a broader cross-linguistic framework.

English listeners are predicted to show the weakest reliance on pitch in part due to their strong reliance on vowel quality, which is far more recoverable as a cue to English lexical stress. As in Dutch and Spanish, pitch in English provides probabilistic information about prominence through its alignment with stressed syllables in words that receive a pitch accent. However, because robust and context-stable segmental cues to stress are readily available in English, pitch is expected to receive lower attentional weight than in Dutch or Spanish, where vowel quality cues are weaker or absent.

3.3. Duration

Cross-linguistic differences in attention to duration are predicted to reflect both the interpretability of durational variation in the L1 and its status relative to other available cues within the stress system. Compared to the other L1 groups, Spanish listeners are expected to show the strongest reliance on duration, as duration consistently distinguishes stressed from unstressed syllables and maintains its contrastive role across intonational contexts. In the absence of vowel-quality cues and given the probabilistic nature of pitch cues, durational differences provide the most consistently recoverable source of lexical stress information in Spanish. This degree of recoverability is expected to support attentional tuning to duration over the course of acquisition. Existing L2 research is consistent with this prediction: Spanish L2 learners of English have been found to rely more strongly on duration cues to English lexical stress than native English listeners (Tremblay & Kim, 2025).

English listeners are predicted to show more limited reliance on duration as a cue to lexical stress. In English, durational differences between stressed and unstressed syllables are highly redundant with vowel quality, which provides a more categorical and context-stable cue to stress. From an attentional perspective, the availability of a dominant segmental cue is expected to constrain long-term attentional weighting of duration, resulting in more limited reliance on this dimension compared to Spanish listeners. At the same time, lexical stress contrasts are pervasive in English and encountered across a large number of words, such that durational differences correlate systematically with stress and are available in the input. Thus, duration should remain recoverable at a broader distributional level across the lexicon even if it contributes relatively little unique information on a given token, yielding limited but non-negligible sensitivity to duration cues in English lexical stress perception.

Dutch listeners present a more complex case. On the basis of the L1 literature, Dutch listeners would be expected to rely substantially on duration, as durational differences contribute robustly to stress perception and classification in Dutch, particularly given the comparatively weaker role of vowel quality. We already know from our previous work, however, that the Dutch L2 learners of English in the current study relied less on duration than native English listeners (Tremblay et al., 2021). At the same time, the literature on Dutch lexical stress perception has not directly compared the relative contribution of duration to that of pitch, raising the possibility that duration may not function as a strong cue in Dutch stress perception. Indeed, pitch may play an equally important—if not more important—role, which would be consistent with the strong reliance on pitch observed for the same Dutch listeners in English (Tremblay et al., 2021). One possibility, therefore, is that duration functions as one of several partially diagnostic cues to lexical stress in Dutch, rather than as a dominant cue. Even if duration contributes more strongly than vowel quality to the perception of lexical stress in Dutch, it may not outweigh pitch and thus not dominate the hierarchy of stress cues. Under this view, Dutch listeners are predicted to rely less on duration than Spanish listeners. This account does not necessarily predict a difference between English and Dutch listeners, however. We suggest that the previously observed difference reflects the interaction between the multi-cue stress system in Dutch and the availability of a highly categorical vowel-quality cue in English, which may further reduce the contribution of duration in L2 stress perception.

In contrast, Mandarin listeners are predicted to show more limited reliance on duration. Although duration plays a role in the realization of neutral-tone syllables and has been proposed as a cue to lexical stress in some varieties of Mandarin, its contribution is relatively restricted and closely coupled with pitch. As a result, durational variation provides little independent evidence for stress-related distinctions in the L1. Moreover, relative to Spanish, English, and Dutch, Mandarin provides relatively sparse lexical evidence that stress alone distinguishes word meanings, further limiting the recoverability of duration as a stress cue across the lexicon. Consequently, Mandarin listeners are expected to rely less on duration than Spanish, English, and Dutch listeners. Existing research is consistent with this prediction in comparisons with English, where Mandarin L2 learners of English show weaker use of duration cues to English lexical stress than native English listeners (Qin et al., 2017; cf. Zhang & Francis, 2010).

Korean listeners are expected to show the weakest reliance on duration as a cue to lexical stress. In contemporary Korean, duration does not encode lexical contrasts and primarily marks higher-level prosodic boundaries rather than word-level distinctions. As a result, durational variation provides little stable or interpretable information for lexical interpretation and offers minimal opportunity for listeners to develop long-term attentional tuning to duration as a meaning-bearing dimension. From an attentional perspective, this cue ecology is not expected to support sustained weighting of duration in L2 lexical stress perception.

In summary, the attentional-learning framework predicts systematic cross-linguistic differences in cue weighting: English listeners should rely most strongly on vowel quality, Mandarin listeners should rely most strongly on pitch, Spanish listeners should rely most strongly on duration, and Korean and Dutch listeners are expected to occupy intermediate positions reflecting the functional role of these dimensions in their respective L1 cue ecologies (see Table 1). To test these cross-linguistic predictions, we compared listeners' relative weighting of vowel quality, pitch, and duration using the cue-weighting stress-perception from Tremblay et al. (2021). The task employs systematically controlled stimuli in which these three acoustic dimensions were orthogonally manipulated, allowing for direct estimation of listeners' attentional weighting of each cue during English lexical stress perception. As previously mentioned, the Dutch listener group in the present study is identical to the Dutch group reported in Tremblay et al. (2021). All other listener groups were recruited separately, and their data have not been published elsewhere.

4. Method

4.1. Participants

A total of 274 participants took part in this study: 42 native speakers of English (30 females, mean age: 20.8, SD: 4.26), 40 native speakers of Dutch (32 females, mean age: 21.5, SD: 2.94), 49 native speakers of Spanish (37 females, mean age: 21.3, SD: 3.49), 83 native speakers of Seoul Korean (43 females, mean age: 24.8, SD: 3.87), and 60 native speakers of Mandarin (35 females, mean age: 24.3, SD: 3.31).[2]

The English speakers were tested at a public university in the Midwestern United States. The Dutch speakers were tested in the Netherlands, and the data for this group are the same as those reported in Tremblay et al. (2021). The Spanish speakers were tested in the Southwestern United States, the Korean speakers in Seoul, South Korea, and the Mandarin speakers in Hong Kong.[3] None of the participants reported a history of hearing or speech disorders. The Dutch, Spanish, Korean, and Mandarin speakers had learned English as an L2.

All participants completed a language background information questionnaire. L2 learners reported their age of first exposure to English, the number of years of formal English instruction, and their self-rated proficiency in English listening, speaking, reading, and writing. Summary statistics for age of first exposure, years of instruction, and mean

[2] The groups were not evenly sized because the participants were recruited and tested for different studies, and their data were pooled at a later stage. This is not a problem for the Bayesian models reported below, because Bayesian modeling naturally accounts for such imbalance through partial pooling, with uncertainty appropriately reflected in posterior estimates.

[3] The Mandarin speakers were from the northern region of Mainland China and only spoke Mandarin dialects of Chinese. They had spent less than one year in Hong Kong and reported limited exposure to spoken Cantonese.

self-rated proficiency across the four skills are provided in Table 2. As shown in Table 2, the four L2 groups were not matched for experience with or perceived proficiency in English. Despite these differences, the Dutch and Spanish groups reported comparable and relatively higher proficiency, whereas the Korean and Mandarin groups reported comparable but lower proficiency levels.

All participants also completed the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012), which served as a measure of English lexical proficiency. Fig. 1 shows the distribution of LexTALE accuracy scores by language group. As expected, native English speakers achieved the highest accuracy. Among the L2 groups, Dutch and Spanish speakers performed at an advanced level and similarly to one another, whereas Korean and Mandarin speakers showed intermediate proficiency and were likewise comparable. This pattern closely mirrors the self-rated proficiency data, with the two measures correlating moderately ($\rho = .54$, $p < .001$, $n = 232$). To account for cross-group differences in English proficiency, LexTALE accuracy was included as a co-variate in all the statistical analyses.

4.2. Materials

Participants completed a cue-weighting stress-perception task adapted from Tremblay et al. (2021; see that study for full methodological details). The task was designed to assess listeners' reliance on vowel quality, pitch, and duration in the perception of English lexical stress.

The auditory stimuli were created from naturally produced recordings of the word pair *DEsert-deSSERT*, spoken by a female native speaker of American English in the carrier sentence *Click on* __, where the target word was produced with a nuclear, high-tone pitch accent (H*). One token of each word was selected as the basis for acoustic manipulation. These tokens differed systematically in vowel quality, pitch (i.e., fundamental frequency [F0]), and duration across the two syllables, reflecting canonical realizations of initial versus final stress.

Stimulus manipulation was carried out in Praat (version 6.0.46; Boersma & Weenink, 2019). Three acoustic dimensions associated with lexical stress were manipulated: vowel quality, pitch, and duration. Vowel quality was varied by interpolating the formant structure (F1-F3 and bandwidths) between the two endpoint vowels across seven evenly spaced steps, creating a continuum from the vowel realization in *DEsert* to that in *deSSERT*. Pitch and duration were likewise manipulated in seven steps between the two endpoint values. For each stimulus, two dimensions were varied orthogonally while the third was held constant at a midpoint, yielding three stimulus matrices: vowel quality × pitch, vowel quality × duration, and pitch × duration. Overall intensity was equalized across syllables and normalized across items. A visual representation of the acoustic manipulations can be found in the Supplementary Materials of Tremblay et al. (2021).

The full stimulus set comprised 147 unique items (three matrices × 49 stimuli), each presented three times in separate blocks, for a total of 441 experimental trials. Twelve practice trials were included prior to the test phase; these consisted of manipulated but canonical stress patterns with all cues aligned to either initial or final stress. This design enabled direct estimation of listeners' relative weighting of vowel quality, pitch, and duration in the perception of English lexical stress. All auditory stimuli can be found on the project's Open Science Framework page

Table 2
Participants' experience with English and self-rated English proficiency.

	Age of First Exposure (Years)	Years of English Instruction	Self-Rated Proficiency
Dutch ($n = 40$)	11.00 (1.07)	7.49 (1.69)	3.15 (0.51)
Spanish ($n = 49$)	6.86 (3.56)	9.41 (6.51)	3.15 (0.64)
Seoul Korean ($n = 83$)	8.93 (2.22)	11.70 (4.28)	2.46 (0.77)
Mandarin ($n = 60$)	7.67 (2.43)	13.70 (4.72)	2.45 (0.49)

Note. Values are means with standard deviations in parentheses.

4.3. Procedures

Participants were tested individually at a computer workstation in a quiet laboratory environment. For the English, Spanish, Korean, and Mandarin groups, the experiment was implemented and administered using Gorilla (www.gorilla.sc). The Dutch group completed the same task using E-Prime 3.0 Software (Psychology Software Tools, 2016), following the original implementation reported in Tremblay et al. (2021).[4] Task structure, stimuli, and trial sequencing were otherwise identical across groups.

On each trial, participants heard an auditory stimulus over headphones and were prompted to identify the word they heard by pressing a labeled response key corresponding to *DEsert* or *deSSERT*. The next trial began 1,000 milliseconds after a response was registered. Trials were self-paced, with participants controlling the onset of each stimulus.

The experiment began with a practice phase consisting of 12 trials in which vowel quality, pitch, and duration cues to stress were fully aligned. Participants received accuracy feedback during the practice phase and were required to achieve at least 75% correct responses to proceed. All participants met this criterion. The test phase followed immediately and consisted of three blocks in which the 147 experimental stimuli were presented in randomized order. The full stress-perception task lasted approximately 20-25 minutes.

In what follows, we report two sets of analyses: The first set models listeners' relative weighting of vowel quality, pitch, and duration, and the second set examines whether cue-weighting patterns can be used to classify listeners by L1 background. Each set of results is introduced with a description of the corresponding analytical approach.

5. Results

5.1. Cue weight Bayesian modeling

The cue-weighting analyses were conducted in R (R Core Team, 2024) using Bayesian multilevel logistic regression (e.g., Gelman et al., 2013; Kruschke, 2010) as implemented in `brms` (Bürkner, 2017) with the `cmdstanr` backend (Gabry et al., 2025). A Bayesian framework was adopted because frequentist mixed-effects logistic regression models with multiple continuous predictors and interactions are prone to convergence and boundary issues in complex crossed designs, whereas Bayesian multilevel models provide more stable estimation under these conditions. The dependent measure was a binary response (1 = *DEsert*, 0 = *deSSERT*), modeled with a Bernoulli likelihood and logit link. To examine how listeners from different L1 backgrounds weighted acoustic dimensions when categorizing lexical stress, we fit separate models for each two-cue matrix: Vowel Quality × Pitch, Vowel Quality × Duration, and Pitch × Duration. In each model, the two acoustic predictors were entered as centered continuous variables, and L1 was included as a categorical predictor with English as the reference level. To test whether cue use differed across L1 groups, each model included cue-by-L1 interactions for both cues. Models also included a centered proficiency covariate to account for individual and between-group differences in English lexical proficiency.[5] The random-effects

[4] We switched the experimental platform used to collect the cue-weighting data from E-Prime to Gorilla because the latter made it easier to collect data across different testing sites.

[5] Interactions between proficiency and the fixed effects were not included because modeling proficiency-dependent changes in cue weighting was not a goal of the present study. Proficiency was entered as a covariate solely to control for between-group differences and to isolate effects attributable to listeners' L1 experience.

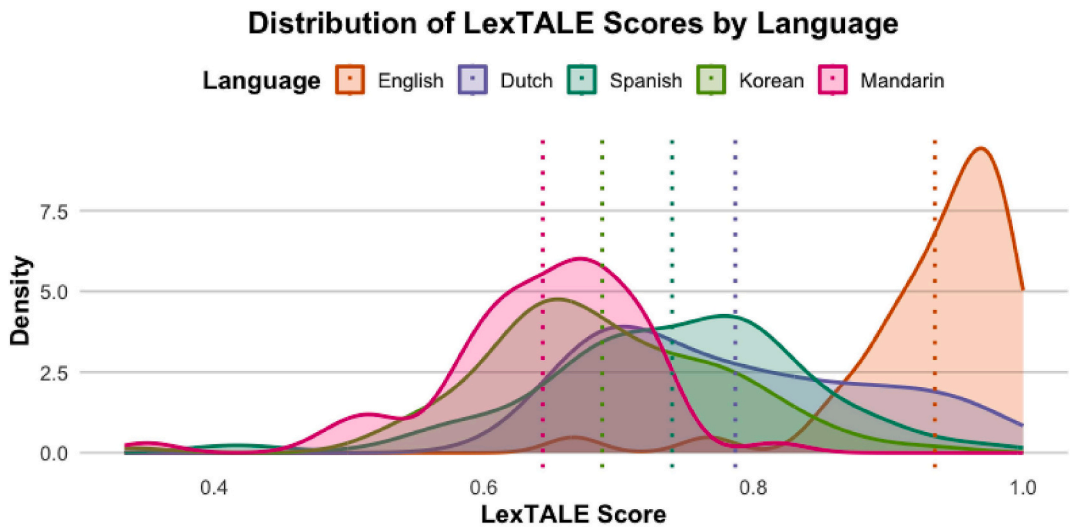

Fig. 1. Density distributions of LexTALE accuracy scores by language group (Lemhöfer & Broersma, 2012). Curves represent the distribution of individual participants' scores on the LexTALE; dottle lines represent the group means.

structure was specified to estimate population-level differences in cue weighting across L1 groups. Accordingly, models included by-participant and by-item random intercepts to account for repeated observations and baseline response variability. Because the primary aim of the analysis was to characterize systematic group-level differences in cue weighting, cue effects were modeled at the population level. Cue-by-participant random slopes were therefore not included, as they would primarily capture individual variation in cue use rather than inform the group-level contrasts of interest. Moreover, including such slopes substantially increases model complexity and can compromise parameter identifiability and interpretability in highly crossed designs of this kind. The chosen random-effects structure thus reflects a principled trade-off between adequately accounting for dependency in the data and maintaining stable estimation of the population-level effects that constitute the focus of the study.

All models were fit in `brms` using the default weakly informative priors, which are designed for standardized predictors and binary response models, and provide mild regularization without strongly constraining effect sizes (Bürkner, 2017). Given the large number of observations contributing to each model, posterior estimates were expected to be driven primarily by the observed data rather than by the prior specification. Models were estimated with four chains and four cores, using 4,000 iterations per chain (1,000 warmup), and sampling was stabilized using `adapt_delta` = 0.95 and `max_treedepth` = 12. Inference was based on posterior summaries of population-level effects and on derived cue-weight measures. For each model, posterior draws were used to compute changes in predicted response probability as each cue varied from −1 SD to +1 SD while holding other predictors constant. Because stronger cue sensitivity corresponded to *more negative changes* in predicted probability under the present coding (with Step 1 corresponding to *DEsert* [1] and Step 7 to *deSSERT* [0]), cue-weight estimates are reported with inverted sign so that larger values indicate stronger reliance on a given cue. These cue-weight estimates were summarized using posterior means, 95% credible intervals, and posterior probabilities of directional differences between language groups. Pairwise comparisons are reported as differences between cue-weight estimates on the probability scale and are used to characterize relative differences among L2 groups. Because each cue appears in two different two-cue matrices, primary inferences about cross-linguistic differences in cue weighting are based on cue-weight estimates aggregated across matrices, with matrix-specific estimates reported descriptively to illustrate context-dependent patterns. The data analysis and visualization script for the descriptive and Bayesian analyses can be found on the project's Open Science Framework page.

Fig. 2 displays participants' responses to stimuli that varied along the vowel quality and pitch continua for each L1 group. All stimuli shown in this figure had a fixed, mid-range value (Step 4) on the duration continuum.

The results of the Bayesian multilevel logistic regression examining the effects of vowel quality, pitch, and their interaction with L1 can be found in Table S1 of the Supplementary Materials. For English listeners (reference level), vowel quality was a strong predictor of *DEsert* responses ($\beta = -0.637$, 95% CI [−0.686, −0.586], $P(\beta < 0) > .999$), with proportions of *DEsert* selection decreasing as vowel quality step increases, indicating robust reliance on this cue. Pitch also exerted a reliable effect for English listeners (β = –0.170, 95% CI [–0.217, –0.123], $P(\beta < 0) > .999$), but its magnitude was smaller than that of vowel quality, consistent with a secondary role for pitch in this group. The model also revealed credible cue-by-L1 interactions: Relative to English listeners, all L1 groups showed significantly reduced weighting of vowel quality, reflected in positive vowel-quality-by-L1 interactions (βs = 0.265 to 0.348, $P(\beta > 0) > .999$). In contrast, compared to English listeners, all L1 groups except the Spanish group made greater use of pitch, as indicated by negative pitch-by-L1 interactions (βs = –0.303 to –0.208, $P(\beta < 0) > .999$; see Table S1 for more details). The positive interaction terms indicate that the effect of vowel quality observed with English listeners is attenuated (i.e., less negative) in other L1 groups, and the negative interaction terms indicate that the effect of pitch showed by English listeners is accentuated (i.e., more negative) in other L1 groups.

Fig. 3 represents model-based predicted response probabilities as a function of cue and L1 in the Vowel Quality × Pitch matrix. Because English listeners served as the reference group in the regression models, matrix-level comparisons focus on relative differences among L2 groups. In this context, Dutch listeners showed comparatively greater sensitivity to vowel quality, whereas Mandarin listeners exhibited the steepest pitch-related changes in predicted response probability. Korean listeners also displayed substantial pitch sensitivity, while Spanish listeners showed more moderate cue effects overall. Table S2 in the Supplementary Materials reports posterior summaries of pairwise differences in cue-weight estimates across L2 groups for this matrix.

Fig. 4 shows participants' responses to stimuli that varied along the vowel quality and duration continua for each L1 group. All stimuli shown in this figure had a fixed, mid-range value (Step 4) on the pitch continuum.

The results of the Bayesian multilevel logistic regression examining the effects of vowel quality, duration, and their interaction with L1 can

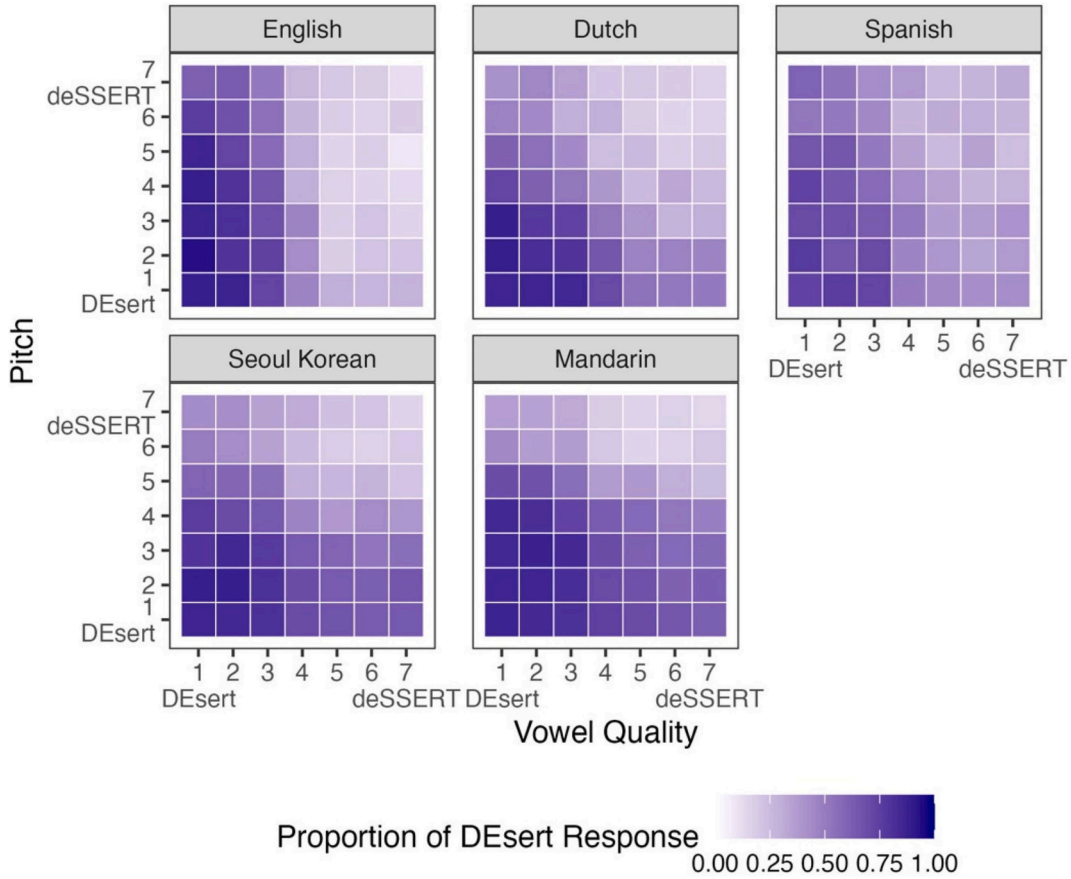

Fig. 2. Proportion of *DEsert* responses as a function of vowel quality and pitch, with duration held constant at its midpoint (Step 4). Separate panels show mean responses for each L1 group. Vowel quality and pitch vary along seven-step continua from acoustic properties associated with word-initial stress (Step 1) to those associated with word-final stress (Step 7). Color shading indicates the proportion of *DEsert* responses, with darker purple representing more frequent *DEsert* selections and lighter purple representing more frequent *deSSERT* selections.

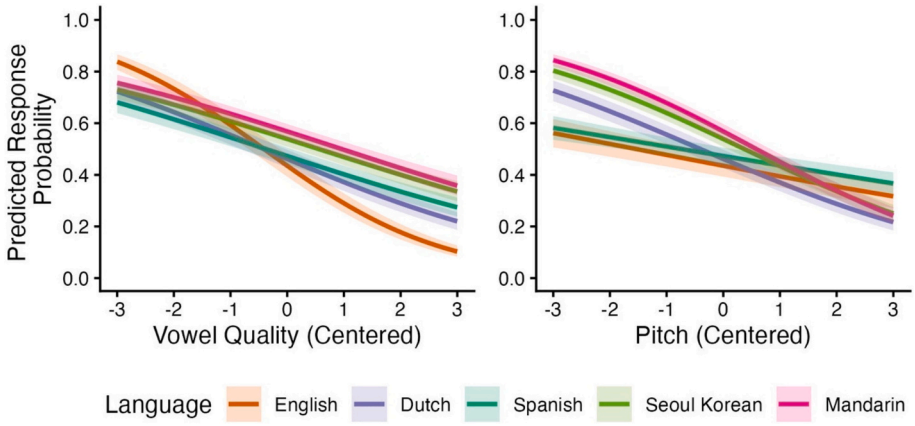

Fig. 3. Model-implied predicted probabilities of *DEsert* responses as a function of vowel quality (left panel) and pitch (right panel) in the Vowel Quality × Pitch matrix. Curves show population-level predictions from the Bayesian multilevel logistic regression, plotted separately for each L1 group. Predictions were generated by varying the target cue across its observed range while holding the other cue and proficiency at their centered means. These plots illustrate the shape and relative strength of cue effects by L1 within this two-cue context.

be found in Table S3 of the Supplementary Materials. Vowel quality was again a strong negative predictor of *DEsert* responses for English listeners ($\beta = -0.671$, 95% CI [−0.715, −0.628], $P(\beta < 0) > .999$), indicating robust reliance on vowel quality as a cue to lexical stress. Duration also exerted a reliable effect ($\beta = -0.115$, 95% CI [−0.155, −0.074], $P(\beta < 0) > .999$), but its magnitude was substantially smaller than that of vowel quality, consistent with a secondary role for duration in this group. Cue-by-L1 interaction terms showed that English listeners made greater use of vowel quality than all L1 groups (βs = 0.277 to 0.383, $P(\beta > 0) > .999$). Importantly, contrary to predictions, English listeners relied more on duration than all L1 groups (βs = 0.055 to 0.072, $P(\beta > 0) > .99$) except Spanish listeners (see Table S3 for additional details).

Fig. 5 provides a visualization of model-based predicted response probabilities as a function of cue and L1 in the Vowel Quality × Duration

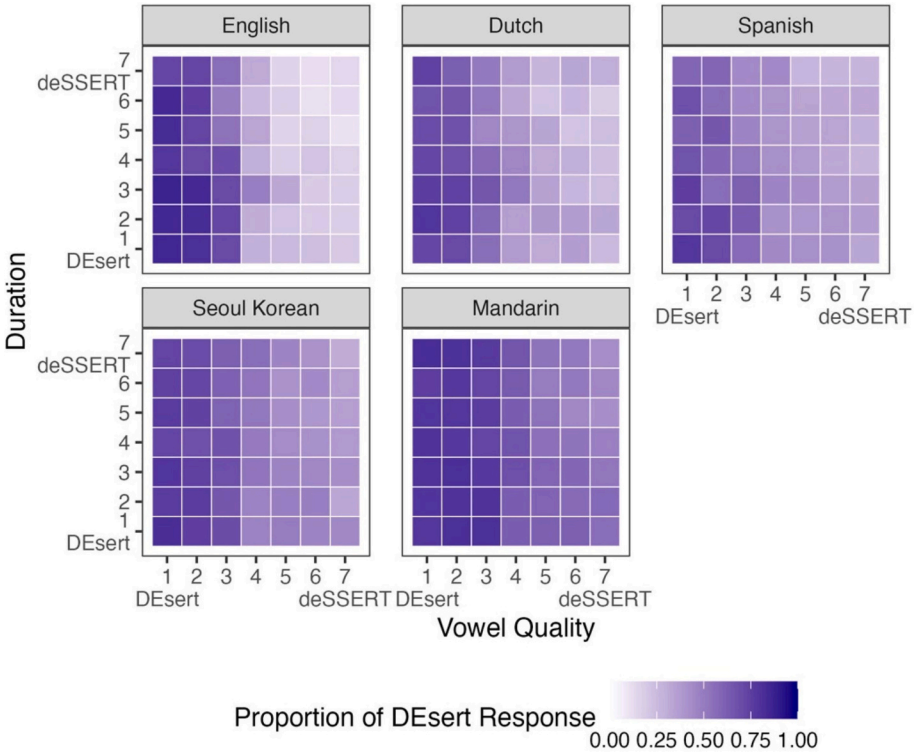

Fig. 4. Proportion of *DEsert* responses as a function of vowel quality and duration, with pitch held constant at its midpoint (Step 4). Separate panels show mean responses for each L1 group. Vowel quality and duration vary along seven-step continua from acoustic properties associated with word-initial stress (Step 1) to those associated with word-final stress (Step 7). Color shading indicates the proportion of *DEsert* responses, with darker purple representing more frequent *DEsert* selections and lighter purple representing more frequent *deSSERT* selections.

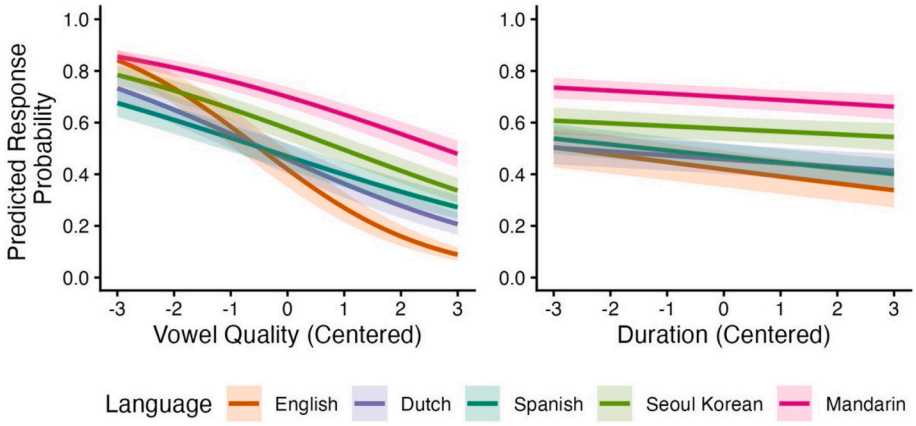

Fig. 5. Model-implied predicted probabilities of *DEsert* responses as a function of vowel quality (left panel) and duration (right panel) in the Vowel Quality × Duration matrix. Curves show population-level predictions from the Bayesian multilevel logistic regression, plotted separately for each L1 group. Predictions were generated by varying the target cue across its observed range while holding the other cue and proficiency at their centered means. These plots illustrate the shape and relative strength of cue effects by L1 within this two-cue context.

matrix. With these stimuli, Dutch listeners again showed comparatively strong sensitivity to vowel quality, while Korean listeners also displayed steeper vowel-quality-related changes in predicted response probability than the Mandarin and Spanish groups. Duration-related effects were most pronounced for Spanish listeners, whereas Mandarin and Korean listeners showed more modest sensitivity to duration in this two-cue context. Table S4 in the Supplementary Materials reports posterior summaries of pairwise differences in cue-weight estimates across L2 groups for this matrix.

Fig. 6 displays participants' responses to stimuli that varied along the pitch and duration continua for each L1 group. All stimuli shown in this figure had a fixed, mid-range value (Step 4) on the vowel quality continuum.

For English listeners, pitch was a reliable negative predictor of *DEsert* responses ($\beta = -0.176$, 95% CI [−0.214, −0.140], $P(\beta < 0) > .999$), indicating that lower pitch steps were associated with greater likelihood of *DEsert* categorization. Duration also yielded a reliable effect ($\beta = -0.048$, 95% CI [−0.085, −0.012], $P(\beta < 0) = .996$), although its magnitude was smaller than that of pitch. Cue-by-L1 interaction terms revealed systematic cross-linguistic differences in the weighting of pitch. Compared to English listeners, Dutch, Korean, and Mandarin listeners all showed greater pitch weighting (βs = −0.315 to −0.181, Ps > .999),

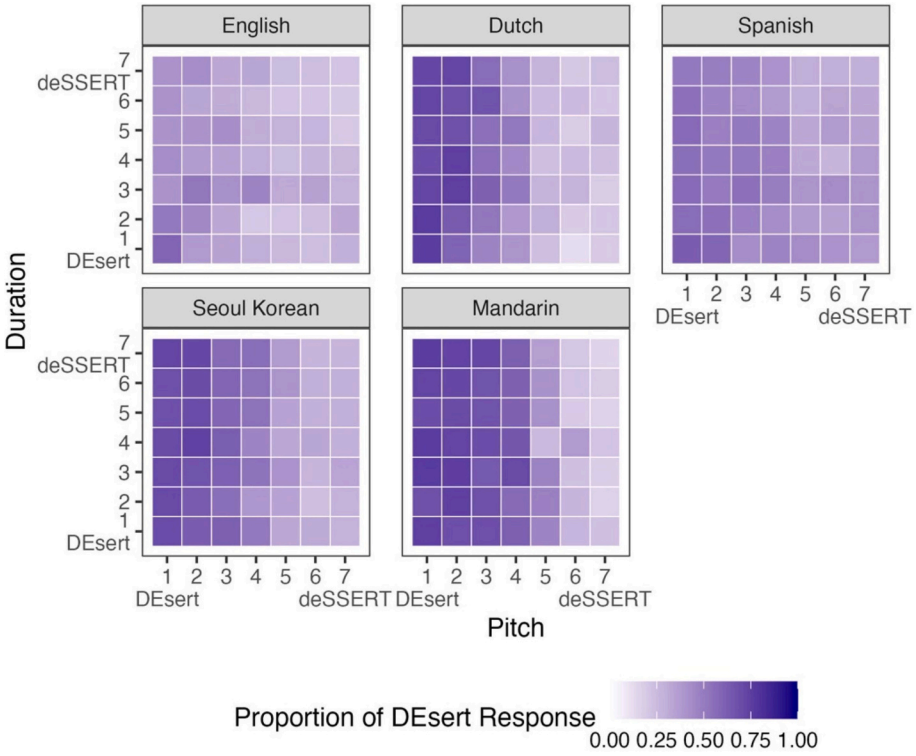

Fig. 6. Proportion of *DEsert* responses as a function of pitch and duration, with vowel quality held constant at its midpoint (Step 4). Separate panels show mean responses for each L1 group. Pitch and duration vary along seven-step continua from acoustic properties associated with word-initial stress (Step 1) to those associated with word-final stress (Step 7). Color shading indicates the proportion of *DEsert* responses, with darker purple representing more frequent *DEsert* selections and lighter purple representing more frequent *deSSERT* selections.

whereas the pitch-by-L1 interaction for Spanish listeners was small and not credibly different from zero, indicating comparable pitch weighting to English listeners. For duration, only Dutch and Korean listeners showed reliably reduced duration weighting relative to English listeners, (Dutch: $\beta = 0.080$; Korean: $\beta = 0.068$; $Ps > .999$); no reliable duration-by-L1 interaction were found for Spanish or Mandarin listeners (for more details, see Table S5).

Fig. 7 visualizes model-based predicted response probabilities as a function of cue and L1 in the Pitch × Duration matrix. In this context, Mandarin listeners showed pronounced pitch-related changes in predicted response probability, while Korean and Dutch listeners also displayed substantial sensitivity to pitch; Spanish listeners exhibited comparatively weaker pitch-related effects overall. Duration-related effects were most pronounced for Spanish listeners, whereas the remaining L1 groups showed more modest sensitivity to duration in this two-cue context. Table S6 in the Supplementary Materials reports posterior summaries of pairwise differences in cue-weight estimates across L2 groups for this matrix.

The preceding analyses showed that cue-weighting patterns varied across the two-cue matrices in which a given cue appeared, reflecting context-dependent differences in how cues were used when paired with different acoustic dimensions. To provide a concise and comparable summary of cue use across these contexts, we conducted a combined analysis that yielded a single cue-weight estimate for each cue and L1

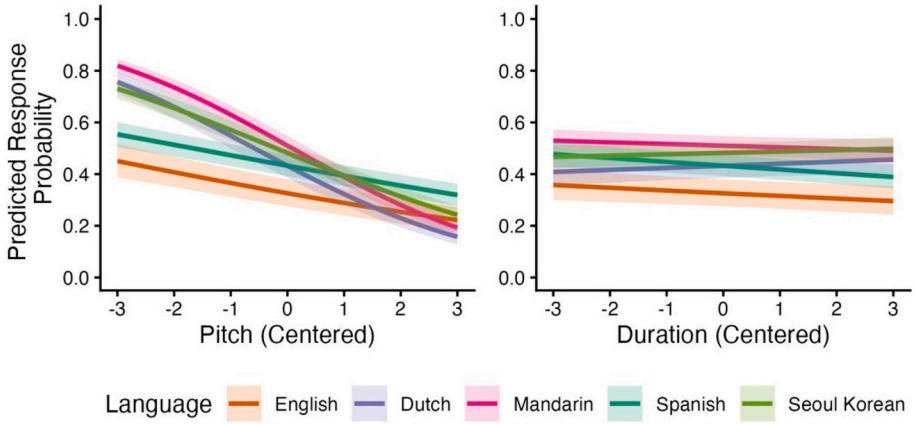

Fig. 7. Model-implied predicted probabilities of *DEsert* responses as a function of pitch (left panel) and duration (right panel) in the Pitch × Duration matrix. Curves show population-level predictions from the Bayesian multilevel logistic regression, plotted separately for each L1 group. Predictions were generated by varying the target cue across its observed range while holding the other cue and proficiency at their centered means. These plots illustrate the shape and relative strength of cue effects by L1 within this two-cue context.

group. To ensure comparability across models and matrices, cue-weight estimates were placed on a common scale using a globally defined standard deviation. For each cue, the previously computed changes in predicted response probability (ΔP) from the two matrices in which that cue appeared were averaged within each posterior draw, yielding a single, context-robust summary of cue weighting for each L1 group.

Fig. 8 provides a visualization of the cue weights aggregated across matrices, with the L1 groups ordered by cue strength, and Table 3 reports posterior summaries of pairwise differences in aggregated cue-weight estimates across all L1 groups.

For vowel quality, the aggregated comparisons revealed a clear separation between English listeners and all other groups, with English showing substantially stronger vowel-quality weighting than Dutch, Mandarin, Korean, and Spanish listeners. Dutch listeners also exhibited stronger vowel-quality weighting than the remaining L2 groups. Differences among Mandarin, Korean, and Spanish listeners were comparatively small and largely overlapping, although Korean listeners showed slightly stronger reliance on vowel quality than Mandarin listeners. Taken together, these results yield a ranking in which English listeners show the strongest reliance on vowel quality, Dutch listeners occupy an intermediate position, and the remaining L2 groups cluster together with weaker and less differentiated weighting of vowel quality.

For pitch, the aggregated comparisons showed a markedly different pattern. Mandarin listeners exhibited stronger pitch weighting than all other groups. Dutch listeners also showed stronger pitch weighting than Korean, English, and Spanish listeners. Korean listeners showed stronger pitch weighting than English and Spanish listeners, while English and Spanish listeners did not differ reliably from one another. These comparisons support a ranking in which Mandarin listeners show the strongest pitch reliance, followed by Dutch listeners, then Korean listeners, with English and Spanish listeners occupying the weakest and largely overlapping positions.

A distinct pattern emerged for duration. Aggregated cue-weight comparisons indicated that English and Spanish listeners relied more strongly on duration than Dutch, Mandarin, and Korean listeners. Mandarin listeners showed stronger duration weighting than Dutch and Korean listeners. Dutch and Korean listeners showed the weakest duration weighting overall. Hence, duration weighting followed a ranking in which English and Spanish listeners showed the strongest reliance, followed by Mandarin listeners, with Dutch and Korean listeners showing comparatively weaker and overlapping weighting of duration.

Having established aggregated cue-weight estimates and corresponding rankings of L1 groups for each acoustic dimension, we next provide an overview of whether these empirically derived L1 rankings align with our cross-linguistic predictions. Table 4 directly contrasts the predicted ordering of L1 groups for each cue with the rankings observed in the present data.

Overall, the observed L1 rankings showed partial convergence with our predictions. For vowel quality, the predicted ordering of English and Dutch listeners as the strongest groups was confirmed, although the remaining L2 groups exhibited greater overlap than anticipated. For pitch, Mandarin listeners ranked highest as predicted, but the relative ordering of Dutch and Korean listeners differed from expectations. For duration, the observed L1 rankings diverged more clearly from predictions, with English listeners ranking higher than expected and Dutch listeners ranking lower than predicted. We return to these mismatches in the Discussion section.

In the next section, we take a complementary, data-driven approach by asking whether listeners' overall cue-weighting profiles can be used to classify them into their L1 group.

5.2. *Machine-learning-based L1 classification*

To assess whether listeners' L1 can be predicted from their cue-weighting behavior, we trained machine-learning classifiers on participant-level cue-weight estimates (e.g., Yarkoni & Westfall, 2017). Because participant-specific cue weights were not identifiable from the previous Bayesian models, participant-level models were required for this analysis, which targets individual-difference structure rather than population-level effects. Hence, for each participant, we fit a logistic regression predicting categorization responses from centered formant, pitch, and duration cues, yielding individual slopes indexing reliance on each acoustic dimension. To assess the stability of these individual slopes, we conducted a split-half reliability analysis. For each participant, trials were randomly divided into two halves, cue weights were re-estimated separately for each half. Because any single random split can be influenced by the particular assignment of trials to halves, this procedure was repeated 100 times and the resulting correlations were averaged. Spearman-Brown corrected reliabilities were very high for vowel quality (.96) and pitch (.96) and moderate for duration (.66), indicating that the participant-level cue-weight estimates reflected stable individual differences rather than measurement noise. The lower reliability observed for duration is consistent with its generally weaker

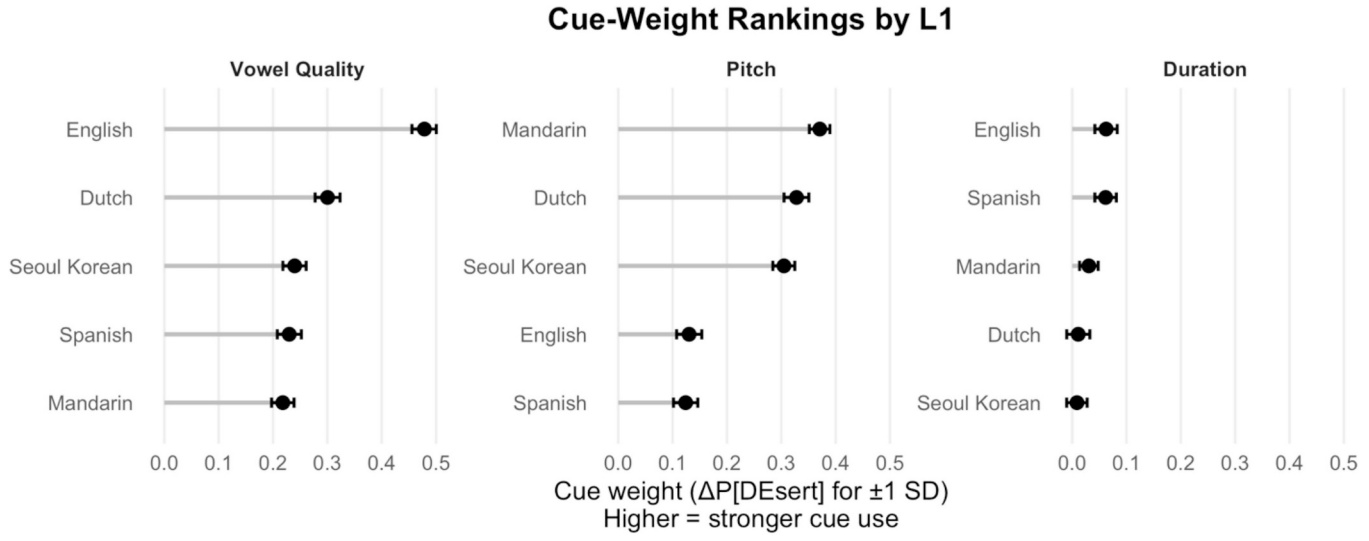

Fig. 8. Cue-weight rankings by L1 for vowel quality, pitch, and duration. Points show posterior mean cue-weight estimates aggregated across the two matrices in which each cue appeared, defined as the change in predicted probability of a *DEsert* response associated with a ±1 SD change in the cue. Error bars indicate 95% credible intervals. Within each panel, L1 groups are ordered from strongest to weakest cue use (top to bottom), with larger values indicating stronger cue weighting.

Table 3
Pairwise comparisons of aggregated cue weights across L1 groups.

Cue	Contrast	Posterior Mean	95% CI (Lower)	95% CI (Higher)	p ($\beta <$ 0)
Vowel Quality	Dutch - English	–0.178	–0.201	–0.154	<.001
	Dutch - Mandarin	0.082	0.062	0.103	>.999
	Dutch - Seoul Korean	0.060	0.040	0.081	>.999
	Dutch - Spanish	0.070	0.049	0.091	>.999
	English - Mandarin	0.261	0.238	0.282	>.999
	English - Seoul Korean	0.239	0.216	0.261	>.999
	English - Spanish	0.249	0.226	0.271	>.999
	Mandarin - Seoul Korean	–0.022	–0.040	–0.004	.009
	Mandarin - Spanish	–0.012	–0.031	0.007	.115
	Seoul Korean - Spanish	0.010	–0.009	0.029	.851
Pitch	Dutch - English	0.198	0.175	0.220	>.999
	Dutch - Mandarin	–0.043	–0.062	–0.023	<.001
	Dutch - Seoul Korean	0.023	0.003	0.043	.987
	Dutch - Spanish	0.204	0.183	0.225	>.999
	English - Mandarin	–0.241	–0.260	–0.221	<.001
	English - Seoul Korean	–0.175	–0.195	–0.154	<.001
	English - Spanish	0.006	–0.015	0.028	.722
	Mandarin - Seoul Korean	0.066	0.050	0.083	>.999
	Mandarin - Spanish	0.247	0.229	0.264	>.999
	Seoul Korean - Spanish	0.181	0.162	0.199	>.999
Duration	Dutch - English	–0.051	–0.074	–0.029	<.001
	Dutch - Mandarin	–0.020	–0.039	0.000	.025
	Dutch - Seoul Korean	0.002	–0.018	0.023	.589
	Dutch - Spanish	–0.050	–0.072	–0.029	<.001
	English - Mandarin	0.032	0.013	0.051	>.999
	English - Seoul Korean	0.054	0.033	0.074	>.999
	English - Spanish	0.001	–0.020	0.022	.531
	Mandarin - Seoul Korean	0.022	0.005	0.039	.995
	Mandarin - Spanish	–0.031	–0.049	–0.013	<.001
	Seoul Korean - Spanish	–0.053	–0.071	–0.034	<.001

contribution to lexical stress perception across listener groups, which reduces between-participant variability and consequently limits the reliability of individual-difference estimates.

The cue-weight estimates, together with each participant's centered English proficiency score, were used as predictors in supervised classification models with L1 as the outcome variable. The primary classifier was a Random Forest model, chosen for its robustness to correlated predictors and nonlinear interactions (e.g., Breiman, 2001). Classification performance was evaluated using Monte Carlo cross-validation with 100 random 80/20 train-test splits at the participant level (e.g., Hastie et al., 2009; Q.-S. Xu & Liang, 2001). For each iteration, we computed overall accuracy as well as accuracy by L1 group, and summarized performance as the mean and standard deviation of participant-level accuracy across iterations. The contribution of each predictor to L1 classification accuracy was assessed by measuring the decrease in accuracy resulting from permuting that predictor in a Random Forest model trained on the full dataset (e.g., Strobl et al., 2007). To evaluate whether more flexible nonlinear models further improved classification performance, we additionally trained two higher-capacity classifiers using the same predictors and cross-validation procedure: a gradient-boosted decision tree model (XGBoost) (T. Chen & Guestrin, 2016) and a deep feed-forward neural network (e.g., Goodfellow et al., 2016). Comparing performance across these models allowed us to assess whether L1 classification was driven primarily by stable, low-dimensional cue-weight structure or benefited from more complex nonlinear decision boundaries. The data analysis and visualization script for the classifier modeling can be found on the project's Open Science Framework page.

Table 4
Predicted and observed L1 rankings in acoustic cue reliance for the perception of english lexical stress.

Cue	Predicted L1 Ranking	Actual L1 Ranking
Vowel Quality	English > Dutch > Korean > Spanish, Mandarin	English > Dutch > Korean, Spanish, Mandarin
Pitch	Mandarin > Korean > Dutch, Spanish > English	Mandarin > Dutch > Korean > English, Spanish
Duration	Spanish > English ≥ Dutch > Mandarin > Korean	English, Spanish > Mandarin > Dutch, Korean

Using the Random Forest classifier, listeners' L1 could be predicted from participant-level cue-weight profiles with accuracy (M = .536, SD = 0.059) and well above chance (where chance = .20). To provide an empirical benchmark for classifier performance, we conducted a permutation analysis in which L1 labels were randomly shuffled and the Random Forest classifier was re-estimated using the same predictors and Monte Carlo cross-validation procedure. Under label permutation, mean classification accuracy was .225 (SD = .050), closely approximating chance performance. Thus, the observed classification accuracy substantially exceeded what would be expected from class structure, sampling variability, or characteristics of the modeling procedure alone. Returning to the observed (non-permuted) data, classification accuracy varied across L1 groups: English listeners were classified with high accuracy (M = .815), whereas Mandarin and Spanish listeners showed moderate classification accuracy (Ms = .619 and .553, respectively). In contrast, Korean and Dutch listeners were classified less reliably (Ms = .378 and .254, respectively).

Table 5 presents the confusion matrix from the Random Forest classifier. Inspection of the confusion matrix indicates that misclassifications were structured rather than random. English listeners were rarely confused with other L1 groups; when misclassifications occurred, they were most often with Dutch or Spanish listeners. In contrast, Dutch listeners were frequently misclassified as other L1s, most commonly as English but also as Spanish, Korean, and Mandarin. This diffuse pattern of misclassification suggests that Dutch listeners

Table 5
Confusion matrix from the random forest classifier with cue weight slopes and proficiency as predictors, reported as proportions within each true L1 (rows sum to ~1).

True L1	English	Dutch	Spanish	Korean	Mandarin
English	**.815**	.099	.073	.000	.014
Dutch	.222	**.254**	.191	.170	.162
Spanish	.104	.076	**.553**	.128	.139
Korean	.000	.091	.093	**.378**	.439
Mandarin	.019	.048	.049	.265	**.619**

occupied a relatively intermediate position in the cue-weighting space, consistent with a cue ecology that shares properties with multiple groups and therefore gives rise to less distinctive attentional profiles. Korean listeners were often confused with Mandarin listeners, and although Mandarin listeners showed relatively high separability from other L1 groups, their misclassifications occurred primarily with Korean listeners. This reciprocal pattern suggests substantial overlap between the Korean and Mandarin groups in their cue-weighting profiles, particularly in their reliance on suprasegmental dimensions. Spanish listeners exhibited intermediate levels of confusability across L1 groups.

Permutation-based importance analyses quantified each predictor's contribution to L1 classification accuracy as the relative reduction in accuracy following permutation of that predictor. English proficiency made the largest contribution to classification performance (mean decrease in accuracy = 45.82). Among cue-weighting measures, pitch contributed most strongly (26.67), followed by duration (17.47) and vowel quality (11.22). The intercept, reflecting baseline response bias, also showed a moderate contribution (29.36). Overall, these results indicate that L1 classification relied primarily on proficiency-related variation, suprasegmental cue-weighting differences, and baseline response bias, with more limited contributions from segmental cues.

Because English proficiency differed across L1 groups, it is important to determine whether classification performance reflected L1-linked cue-weighting structure or merely proficiency-related differences. The Random Forest classifier was therefore rerun with English proficiency excluded as a predictor. Overall classification accuracy decreased moderately (M = .450, SD = .062) but remained stable and well above chance. The largest reduction in accuracy was observed for English listeners (M = .498), consistent with the fact that this group differed most strongly from the remaining groups in proficiency. In contrast, classification accuracy decreased only slightly for the other L1 groups (Mandarin = .573, Spanish = .519, Korean = .372, Dutch = .212), suggesting that proficiency primarily contributed to L1 classification by distinguishing English listeners from the other groups.

Permutation-based importance analyses of the proficiency-excluded model confirmed the previously observed pattern of cue contributions, with pitch cue weight showing the strongest contribution to classification accuracy (mean decrease in accuracy = 32.42), followed by duration (21.80) and vowel quality (14.99). Baseline response bias showed a larger contribution (37.55), indicating that it captured variance that was previously accounted for by proficiency.

Finally, to assess whether classification performance depended on the choice of classifier, we compared the Random Forest model including proficiency to XGBoost and neural network classifiers trained on the same predictors. Overall classification accuracy differed modestly across models, with XGBoost showing slightly higher accuracy (M = .543, SD = .066) and the neural network showing lower accuracy (M = .493, SD = .062) than the Random Forest classifier (M = .526, SD = .059). Importantly, these differences did not alter the pattern of results across L1 groups (see Tables S7 and S8 of the Supplementary Materials for the corresponding L1 confusion matrices). Together, these results indicate that the observed L1 classification patterns are not specific to the Random Forest approach and do not benefit substantially from the increased capacity of XGBoost or neural network models to capture complex nonlinear decision boundaries.

6. Discussion

The present study examined how long-term, language-specific attentional tuning to acoustic dimensions shapes listeners' perception of English lexical stress. Using a cue-weighting paradigm with orthogonal manipulations of vowel quality, pitch, and duration, we compared listeners from five phonologically distinct L1 backgrounds and derived aggregated cue-weight rankings for each dimension.

Across analyses, clear and systematic cross-linguistic differences emerged. English listeners showed the strongest reliance on vowel quality, followed by Dutch listeners, with the remaining L2 groups exhibiting substantially weaker and overlapping weighting of this cue. For pitch, Mandarin listeners relied most strongly on the cue, followed by Dutch and Korean listeners, while English and Spanish listeners showed the weakest and largely indistinguishable pitch weighting. Duration revealed a different pattern: English and Spanish listeners relied most strongly on duration, Mandarin listeners occupied an intermediate position, and Dutch and Korean listeners showed the weakest reliance.

The observed cue-weighting profiles were sufficiently structured to support above-chance classification of listeners' L1 backgrounds using machine-learning models, even when English proficiency was excluded as a predictor. Importantly, classification performance was not specific to one model and did not improve substantially with higher-capacity nonlinear models, indicating that L1 differences were driven by stable, low-dimensional structure in cue-weighting behavior rather than by complex idiosyncratic interactions.

These results provide converging evidence that cross-linguistic differences in English lexical stress perception reflect persistent, dimension-specific attentional biases shaped by the functional role, interpretability, and learnability of acoustic cues in the L1. At the same time, the observed patterns reveal important asymmetries between the cue-weight rankings we predicted and those we found that refine and constrain an attentional-learning account. We discuss these points of convergence and divergence separately for each cue and then discuss the theoretical and broader implications of the present findings.

6.1. Vowel quality

The results for vowel quality largely confirmed our predictions. English listeners' strong reliance on vowel quality is consistent with the central lexical role of vowel reduction in English (e.g., Bond, 1981; Bond & Small, 1983; Fear et al., 1995; Field, 2005; Ghosh & Levis, 2021; Small et al., 1988). Dutch listeners' intermediate weighting of vowel quality aligns with the weaker and less categorical nature of vowel reduction in Dutch (e.g., Campbell & Beckman, 1997; Kager, 1989; Sluijter & van Heuven, 1996a; van Bergem, 1993), which limits its diagnostic value for lexical stress while still supporting some degree of attentional tuning. Spanish and Korean listeners' weaker reliance on vowel quality is consistent with the absence of a systematic mapping between vowel quality and lexical stress or prominence in these languages (e.g., Harris, 1983; Hualde, 2005; Yang, 1996).

Our prediction that Korean listeners might assimilate the central-back Korean vowel /ʌ/ to the English schwa was only weakly supported, with Korean listeners relying slightly more on this cue compared to Mandarin listeners, in line with previous L2 research (Connell et al., 2018). However, this assimilation was not sufficiently robust to yield stronger use of this cue in Korean listeners compared to Spanish listeners.

The results for Mandarin listeners are likewise consistent with the covariance between vowel reduction and tonal information in Mandarin (e.g., Y. Chen & Xu, 2006; W.-S. Lee & Zee, 2010; T. Lin, 1985) and with the relatively sparse distribution of lexical stress contrasts in Mandarin (Yan, 2021). As a result, Mandarin listeners have little incentive to allocate sustained attention to vowel quality in the L1. From an attentional perspective, the combination of high predictability and limited lexical payoff is expected to constrain the development of strong attentional weighting for vowel quality, in line with previous research showing weak reliance on vowel quality among Mandarin L2 learners of English (Connell et al., 2018; C. Y. Lin et al., 2014).

6.2. Pitch

The pitch results were in line with most of our predictions. Mandarin listeners' strongest reliance on pitch reflects the exceptionally high functional load of pitch for lexical identity in a tone language (e.g.,

Howie, 1976; Liu & Samuel, 2004) and the stability of pitch-meaning mappings across contexts (e.g., Y.-C. Lee et al., 2016; Ouyang & Kaiser, 2015; Y. Xu, 1999). This strong attentional prioritization persists even when pitch is not lexically contrastive in the L2, consistent with the view that cue weighting reflects long-term, dimension-specific attentional priorities shaped by the L1 and in line with existing L2 research (Chrabaszcz et al., 2014; Qin et al., 2017; Zhang & Francis, 2010).

Dutch listeners showed stronger reliance on pitch than predicted. Such results would be expected if pitch played a more central role in Dutch stress perception than what was hypothesized on the basis of the existing L1 literature (cf. Sluijter & van Heuven, 1996a). One possibility is that pitch accents may be more systematically aligned with stressed syllables in Dutch than in English or Spanish, providing recurrent and recoverable evidence for lexical stress across words and context, particularly in the presence of relatively weak vowel quality cues (e.g., Campbell & Beckman, 1997; Kager, 1989; Sluijter & van Heuven, 1996a; van Bergem, 1993). This cue ecology may therefore promote heightened attentional sensitivity to pitch, even in the absence of categorical pitch-lexical mappings.

Korean listeners' moderate reliance on pitch was as predicted and can be attributed to the lexical function that pitch plays in the language. Although pitch does not encode lexical tone or stress in Seoul Korean, it plays a crucial role in signaling certain lexically contrastive segmental distinctions in the laryngeal system (e.g., K.-H. Kang & Guion, 2008; Y. Kang, 2014b; Silva, 2006), a role that is largely restricted to phrase-initial contexts (J. Choi et al., 2020). This functional relevance, though limited in scope, appears sufficient to support greater attentional weighting of pitch than in languages where pitch provides little recoverable information for lexical stress.

English and Spanish listeners' weak reliance on pitch is consistent with pitch providing indirect and probabilistic information about lexical stress through the proximal alignment of pitch accents and stressed syllables (e.g., Beckman, 1986; Beckman & Pierrehumbert, 1986; Face & Prieto, 2007; Hualde & Prieto, 2015; Ladd, 2012; Ortega-Llebaria & Prieto, 2007; Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990; Prieto & Torreira, 2007). English listeners' stronger use of pitch than anticipated suggests that the presence of a more informative cue (vowel quality) in the language constraints but does not eliminate the relative contribution of a less informative one (pitch), particularly when pitch is enhanced by utterance-level prominence through a pitch accent. In English, a head-prominence language (Jun, 2005), utterance-level prominence is realized through a phrase-level pitch accent aligned with the lexically stressed syllable (the prosodic head). Because the stimuli were produced as independent utterances with a pitch accent aligned with the lexically stressed syllable, pitch may have become especially informative in this context by reinforcing the stress contrast via intonational prominence. Under such conditions, pitch is integrated into the prosodic realization of the contrast at the utterance level, potentially enhancing listeners' attentional sensitivity to pitch-based cues.

6.3. Duration

The results for duration diverged more clearly from our predictions. English and Spanish listeners' stronger reliance on duration relative to the other L1 groups is consistent with stress-based durational differences being encountered frequently and systematically across the lexicon in both languages. In English and Spanish, duration provides a highly consistent correlate of lexical stress (Beckman & Edwards, 1994; Lieberman, 1960; Ortega-Llebaria & Prieto, 2007, 2011). Contrary to our predictions, English listeners did not rely less on duration than Spanish listeners. Together with the pitch results, this pattern suggests that redundancy with a more informative cue (vowel quality in English) constrains the relative dominance of a cue without eliminating attentional tuning altogether: When durational patterns consistently co-occur with lexical stress across a large number of words, duration remains recoverable as a stress-related dimension and can accrue sustained sensitivity, even in the presence of more diagnostic cues.

Mandarin listeners' weaker reliance on duration aligns closely with our predictions. In Mandarin, duration has been proposed as a cue to lexical stress, but its contribution is closely coupled with pitch, with unstressed syllables being realized with a neutral tone (e.g., Y. Chen & Xu, 2006; W.-S. Lee & Zee, 2010; T. Lin, 1985). Crucially, Mandarin provides relatively sparse lexical evidence that stress alone distinguishes word meanings, with few minimal pairs differing solely in stress (e.g., Yan, 2021). As a result, durational variation offers limited recoverable evidence for stress-related distinctions across the lexicon. From an attentional perspective, this input structure is expected to weaken tuning to duration.

Dutch and Korean listeners showed the weakest reliance on duration. This pattern was not expected for Dutch listeners, given the evidence that duration contributes to stress perception in Dutch and has been shown to be more informative than vowel quality in the L1 (e.g., Sluijter & van Heuven, 1996b; van Heuven & de Jonge, 2011). The present results suggest that although duration may be exploited in Dutch, it does not accrue strong attentional prioritization when listeners allocate attention across competing acoustic dimensions (e.g., pitch). As with English listeners, the prosodic realization of the stimuli may also be relevant here. Because the target words were produced in a sentential context that elicited a nuclear pitch accent, stress detection may have been closely aligned with pitch-accent placement, encouraging reliance on pitch rather than duration. Given that Dutch, like English, is a head-prominence language in which pitch accent serves as a central cue to prominence (Jun, 2005), this alignment may help explain Dutch listeners' relatively weak reliance on duration in the present task. In Korean, where duration does not encode lexical contrasts (e.g., Y. Kang et al., 2015; Shin et al., 2012; Sohn, 1999) and is primarily associated with higher-level prosodic structure (e.g., Cho & Keating, 2001), reliance on duration was likewise weak, consistent with limited opportunities for attentional tuning to this dimension.

An important point to notice is that duration was consistently weighted less strongly than vowel quality and pitch across all listener groups, including those for whom duration plays a more important role in the L1 stress system (i.e., Spanish listeners). This convergence across distinct L1s suggests that limitations on the use of duration may reflect general constraints on its learnability and interpretability, rather than properties of any single language. Unlike vowel quality or pitch, durational information is inherently relative and requires substantial contextual normalization, which reduces its utility as a stable cue for lexical interpretation. This interpretation is further reinforced by the prosodic realization of the stimuli. Because the target words were produced in utterance-final position, durational differences across stress patterns are likely to have reflected not only lexical stress but also phrase-final lengthening on the final syllable (Cambier-Langeveld, 1997; Paschen et al., 2022; Turk & Shattuck-Hufnagel, 2007). The convergence of these multiple sources of durational variation may make it difficult for listeners to isolate duration as a stable, independent cue to lexical stress, thereby limiting the extent to which duration accrues strong attentional weighting in the present task. Hence, from an attentional perspective, duration cues to lexical stress may function as supporting cues whose influence is shaped by their alignment with more interpretable dimensions, rather than as a dimension that readily accrues strong long-term attentional weighting.

6.4. Theoretical contributions

The present findings make several theoretical contributions to models of speech perception and category learning by refining an attentional-learning account of cue weighting and by specifying the conditions under which long-term, language-specific attentional biases interact. The current results support a view in which cross-linguistic

transfer is conceptualized as durable, dimension-specific attentional priors shaped by the structure of the L1 input (Francis et al., 2000; Francis & Nusbaum, 2002; Holt & Lotto, 2006; Nosofsky, 1986; Roark & Holt, 2022) and constrained by redundancy among cues (Holt & Lotto, 2006; Schertz & Clare, 2020).

A major contribution of the current study lies in clarifying what kinds of L1 experience matter for attentional transfer in L2 speech perception. The results demonstrate that attentional tuning is shaped not merely by whether a language has lexical stress, but by how individual acoustic dimensions function within the L1 cue ecology (e.g., Chang, 2018; W. Choi, 2022; Ingvalson et al., 2011; Iverson et al., 2003). Pitch, for example, was most strongly weighted by Mandarin listeners, as predicted, but Dutch listeners showed greater reliance on pitch than Korean listeners, contrary to initial expectations. This pattern suggests that attentional tuning is sensitive to the recoverability of a dimension across contexts and across lexical items, not simply to whether that dimension encodes a categorical contrast. Pitch accents in Dutch, a head-prominence language typologically similar to English (Jun, 2005), may provide sufficiently consistent evidence for prominence to promote attentional sensitivity, even in the absence of categorical pitch–lexical mappings.

Another important contribution of the study is to clarify how cue redundancy shapes long-term attentional tuning. Previous work on cue trading has shown that listeners can flexibly shift reliance across cues under experimental manipulation or when cues are placed in conflict (e.g., Pisoni & Luce, 1987; Repp, 1983; Schertz & Clare, 2020). The present results extend this literature by showing that redundancy does not eliminate attentional sensitivity to a cue. English listeners, for example, showed sensitivity to duration and pitch despite the presence of a highly informative and context-stable vowel-quality cue. Similarly, Mandarin listeners showed sensitivity to duration despite their strong attentional focus on pitch. These findings indicate that redundancy constrains the relative strength of cues over developmental timescales without erasing attentional tuning.

This refinement has important consequences for how attentional-learning theories conceptualize long-term cue representations. Rather than assuming that attentional weights converge on a single optimal solution or that less informative cues are effectively pruned through learning, the present findings suggest that listeners maintain multidimensional attentional profiles shaped by the cumulative distributional history of each cue in the L1 (e.g., Chang, 2018; W. Choi, 2022; Roark & Holt, 2022). These profiles are not simple hierarchies but structured patterns of sensitivity that preserve information about multiple dimensions, even when those dimensions differ in interpretability, contextual stability, or diagnosticity. Such profiles support robust perception under variable conditions and may help explain why listeners can flexibly recruit secondary cues when primary cues are degraded or unreliable.

A third contribution of the study is methodological but theoretically consequential. By combining cue-weighting analyses with machine-learning classification, the study demonstrates that L1-linked attentional profiles are not only interpretable post hoc but also sufficiently structured to support prediction (Yarkoni & Westfall, 2017; see also Toscano & McMurray, 2010). Importantly, with the exception of native English listeners, participants were tested in the L2 and had intermediate to advanced levels of English proficiency, indicating that performance on the cue-weighting task cannot be attributed solely to L1-based processing. Consequently, classification accuracy would not be expected to approach ceiling levels if learners have acquired any sensitivity to English lexical stress. The observation of above-chance classification nonetheless indicates that long-term, L1-shaped attentional structure continues to constrain perception even after substantial L2 experience.

Crucially, classification accuracy differed systematically across L1 groups. Some groups (e.g., English and Mandarin listeners) were classified with relatively high accuracy, whereas others (e.g., Dutch listeners) showed substantially lower classification performance. This pattern is theoretically informative, as it suggests that some L1s give rise to more distinctive and internally coherent attentional profiles, whereas others yield profiles that overlap more extensively with those of other language groups. Such variation is expected if attentional tuning reflects the structure of the L1 cue ecology: Languages in which cue-function mappings are more categorical or consistently recoverable (e.g., English and Mandarin) should give rise to more separable profiles, whereas languages with more distributed cue systems (e.g., Dutch) may produce attentional profiles that are less sharply differentiated.

The current findings also have implications beyond lexical stress perception. Although the present study focused on stress, the attentional-learning framework advanced here is intended to apply more broadly to linguistic domains in which multiple acoustic cues jointly support categorization. The framework therefore predicts that similar dimension-specific profiles should emerge in other domains, including segmental contrasts (e.g., Escudero et al., 2009) and tonal systems (e.g., Chandrasekaran et al., 2010). Crucially, the framework shifts the emphasis from categorical inventories (e.g., Best & Tyler, 2007; Flege & Bohn, 2021) to L1 cue ecologies and dimension-specific attentional tuning (Tremblay et al., 2021), offering a unifying perspective for understanding both convergence and divergence in speech perception across languages and learners.

6.5. *Limitations and future research*

This study is not without limitations, which point to directions for future research. First, the experimental design focused on a single English word pair in order to tightly control the acoustic realization of the cues of interest (e.g., Chandrasekaran et al., 2010; Iverson et al., 2005). This design choice was essential for orthogonally manipulating vowel quality, pitch, and duration within a fixed lexical contrast, thereby isolating dimension-level attentional weighting independent of variation in lexical frequency, neighborhood structure, or phonotactic context. As such, the paradigm provided a targeted probe of listeners' attentional allocation to acoustic dimensions, rather than a measure of word-specific representations. Nevertheless, restricting the stimulus set to a single lexical item necessarily limits the extent to which the observed attentional profiles can be assumed to generalize across words and phonological contexts. Future work using larger and more lexically diverse stimulus sets will therefore be important for assessing the stability of these attentional profiles across the lexicon.

Second, listeners were tested at a single point in time, after substantial prior exposure to English. This design choice was intentional, as it allowed us to examine the persistence of L1-shaped attentional tuning under conditions in which learners have had extensive opportunity to adapt to the L2 input. By focusing on relatively stable experienced listeners, the study directly targeted whether early-acquired attentional biases continue to shape cue weighting even after prolonged L2 exposure, rather than conflating persistence effects with ongoing learning dynamics. At the same time, the present findings cannot speak to the developmental trajectory of cue weighting over the course of L2 acquisition. Longitudinal or training-based studies that track changes in cue weighting over time will therefore be critical for disentangling how attentional priors are modified through L2 experience and for identifying the conditions under which such priors are most resistant or most amenable to change (e.g., Clayards et al., 2008; Idemaru et al., 2012; Kim et al., 2024; Schertz et al., 2015).

Finally, the present study emphasized group-level patterns in order to advance theoretical accounts of how language experience shapes attentional weighting at the population level. This focus was necessary for identifying systematic L1-related regularities that would be difficult to discern from individual-level analyses, particularly in the presence of substantial within-group variability. At the same time, attentional learning is ultimately instantiated at the level of individual listeners. Future research that integrates cue-weighting paradigms with computational modeling approaches will be important for clarifying how

dimension-specific attentional tuning is represented and updated over time within individuals.

7. Conclusion

The present study demonstrated that cross-linguistic differences in the perception of English lexical stress reflect systematic, dimension-specific patterns of attentional tuning shaped by prior linguistic experience. Across five L1 backgrounds, listeners differed predictably in their weighting of vowel quality, pitch, and duration, and these cue-weighting profiles were sufficiently structured to support above-chance classification of listeners' L1 backgrounds. These findings suggest that cross-linguistic transfer is best understood in terms of how individual acoustic dimensions function within the L1 cue ecology and how those experiences shape long-term attentional priorities. More broadly, the results indicate that language experience gives rise to stable and distinctive attentional profiles that continue to influence perception even after substantial L2 exposure. These findings support an attentional-learning account of speech perception and highlight the importance of cue ecology and dimension-specific learning in explaining both commonalities and differences across language learners.

Institutional review board and informed consent statement

The study was conducted according to the guidelines of the Declaration of Helsinki and received umbrella approvals from the Institutional Review Boards at the University of Kansas, with an approval date of June 18, 2020, and at the University of Texas at El Paso, with an approval date of January 21, 2025. Informed consent was obtained from all participants involved in the study.

Data collection and availability statement

The auditory stimuli and data analysis and visualization scripts can be found on the project's Open Science Framework page available at https://osf.io/jz9kx/overview?view_only=d942cc1257fb4d758185923f2a4399a2.

Acknowledgment of AI assistance

During the preparation of this manuscript, the authors used ChatGPT (OpenAI; GPT-5.2) for the purpose of improving the data analysis and visualization scripts, and to organize and polish the writing of this manuscript. The tool was not involved in the design of the study, nor did it directly handle the analysis or interpretation of the results. All outputs were critically reviewed and edited by the authors, who take full responsibility for the content of this publication.

CRediT authorship contribution statement

Annie Tremblay: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Taehong Cho:** Writing – review & editing, Supervision, Resources, Investigation, Conceptualization. **Mirjam Broersma:** Writing – review & editing, Investigation, Conceptualization. **Hyoju Kim:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Quentin Zhen Qin:** Writing – review & editing, Visualization, Supervision, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Jiayu Liang:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis.

Funding

This material is based upon work supported by the National Science Foundation under Award No. [BCS-2016750].

Declaration of competing interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2026.106671.

Data availability

The auditory stimuli and data analysis and visualization scripts can be found on the project's Open Science Framework page available at https://osf.io/jz9kx/overview?view_only=d942cc1257fb4d758185923f2a4399a2.

References

Beckman, M. E. (1986). *Stress and non-stress accents*. Foris.
Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in laboratory phonology III* (pp. 7–33). Cambridge University Press.
Beckman, M. E., & Elam, G. A. (1997). *Guidelines for ToBI labeling [Unpublished manuscript]*.
Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook, 3*, 255–310.
Beier, E. J., & Ferreira, F. (2018). The temporal prediction of stress in speech and its relation to musical beat perception. *Frontiers in Psychology, 9*, 431.
van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication, 12*(1), 1–23. https://doi.org/10.1016/0167-6393(93)90015-D
Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In M. J. Munro, & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). John Benjamins.
Boersma, P., & Weenink, D. (2019). *Doing phonetics by computer [Computer software]*.
Bond, Z. S. (1981). Listening to elliptic speech: Pay attention to stressed vowels*. *Journal of Phonetics, 9*(1), 89–96. https://doi.org/10.1016/S0095-4470(19)30929-5
Bond, Z. S., & Small, L. H. (1983). Voicing, vowel and stress mispronunciations in continuous speech. *Perception & Psychophysics, 34*, 470–474.
Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
Brown, M. E., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2016). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology, 41*, 306–323.
Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1–28.
Cambier-Langeveld, T. (1997). The domain of final lengthening in the production of Dutch. *Linguistics in the Netherlands*, 13–24.
Campbell, N., & Beckman, M. (1997). Stress, prominence, and spectral tilt. In A. Botinis, G. Kouroupetroglou, & G. Carayiannis (Eds.), *Intonation: Theory, Models, and Applications: I* (pp. 67–70). GESCA.
Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America, 128*(1), 456–465. https://doi.org/10.1121/1.3445785
Chang, C. B. (2018). Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics, 68*, 85–102. https://doi.org/10.1016/j.wocn.2018.03.003
Chao, Y.-R. (1968). *A grammar of spoken Chinese*. University of California Press.
Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16* (pp. 785–794). https://doi.org/10.1145/2939672.2939785
Chen, Y., & Xu, Y. (2006). Production of weak elements in speech — Evidence from F(0) patterns of neutral tone in Standard Chinese. *Phonetica, 63*(1), 47–75. https://doi.org/10.1159/000091406
Cho, T. (2011). The timing of phrase-initial tones in Seoul Korean: A weighted constraint model. *Phonology, 28*, 293–330.
Cho, T. (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Lang & Ling Compass, 10*, 120–141.
Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155–190. doi:10.006/jpho.2001.0131.
Choi, J., Kim, S., & Cho, T. (2020). An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PLoS One, 15*(10), Article e0240682. https://doi.org/10.1371/journal.pone.0240682
Choi, W. (2022). Theorizing positive transfer in cross-linguistic speech perception: The Acoustic-Attentional-Contextual hypothesis. *Journal of Phonetics, 91*. https://doi.org/10.1016/j.wocn.2022.101135

Choi, W., Tong, X., & Samuel, A. G. (2019). Better than native: Tone language experience enhances English lexical stress discrimination in Cantonese-English bilingual listeners. *Cognition, 189*, 188–192. https://doi.org/10.1016/j.cognition.2019.04.004
Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research, 57*, 1468–1479. https://doi.org/10.1044/2014_JSLHR-L-13-0279
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
Cole, J. (2026). Prosody and information structure. *Annual Review of Linguistics, 12*, 387–406. https://doi.org/10.1146/annurev-linguistics-011724-121524
Connell, K., Hüls, S., Martínez-García, M. T., Qin, Z., Shin, S., Yan, H., & Tremblay, A. (2018). English learners' use of segmental and suprasegmental cues to stress in lexical access: An eye-tracking study. *Language Learning, 68*(3), 635–668. https://doi.org/10.1111/lang.12288
Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44*(2), 171–195. https://doi.org/10.1177/00238309010440020301
van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology, 58*, 251–273. https://doi.org/10.1080/02724980343000927
Duanmu, S. (2007). *The Phonology of Standard Chinese*. Oxford University Press. https://doi.org/10.1093/oso/9780199215782.001.0001
Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress "deafness" in simultaneous French-Spanish bilinguals. *Cognition, 114*, 266–275. https://doi.org/10.1016/j.cognition.2009.10.001
Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition, 106*, 682–706. https://doi.org/10.1016/j.cognition.2007.04.001
Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics, 37*(4), 452–465. https://doi.org/10.1016/j.wocn.2009.07.006
Face, T. L., & Prieto, P. (2007). Rising accents in Castilian Spanish: A revision of Sp_ToBI. *Journal of Portuguese Linguistics, 5*(6), 117–146.
Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America, 97*, 1893–1904. https://doi.org/10.1121/1.412063
Feng, Y., & Peng, G. (2023). Development of categorical speech perception in Mandarin-speaking children and adolescents. *Child Development, 94*(1), 28–43. https://doi.org/10.1111/cdev.13837
Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399–423. https://doi.org/10.2307/3588487
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. 1–83). Cambridge University Press.
Fletcher, J. (2010). The prosody of speech. *Timing and Rhythm*, 521–602. https://doi.org/10.1002/9781444317251.ch15
Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62*, 1668–1680.
Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 349–366.
Gabry, J., Češnovar, R., Johnson, A., & Bronder, S. (2025). cmdstanr: R Interface to "CmdStan" (Version 0.9.0) [Computer software]. https://discourse.mc-stan.org.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC https://sites.stat.columbia.edu/gelman/book/.
Ghosh, M., & Levis, J. M. (2021). Vowel quality and direction of stress shift in a predictive model explaining the varying impact of misplaced word stress: Evidence from English. *Frontiers in Communication, 6*. https://doi.org/10.3389/fcomm.2021.628780
Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://mitpress.mit.edu/9780262035613/deep-learning/.
Guo, X., & Chen, X. (2022). Perception of English stress of synthesized words by three Chinese dialect groups. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.803008
Gussenhoven, C. (2004). Transcription of Dutch intonation. In S. A. Jun (Ed.), *The phonology of intonation and phrasing* (pp. 118–145). Oxford University Press.
Gussenhoven, C. (2008). Vowel duration, syllable quantity and stress in Dutch. In K. Hanson, & S. Inkelas (Eds.), *The nature of the word: Studies in honor of Paul Kiparsky* (pp. 181–198).
Harris, J. W. (1983). *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge University Press.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer. https://doi.org/10.1007/978-0-387-84858-7
van Heuven, V. J., & de Jonge, M. (2011). Spectral and temporal reduction as stress cues in Dutch. *Phonetica, 68*(3), 120–132. https://doi.org/10.1159/000329900
Holt, L. L. (2025). Speech perception is speech learning. *Current Directions in Psychological Science, 34*(4), 218–224. https://doi.org/10.1177/09637214251318726
Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America, 119*(5 Pt 1), 3059–3071. https://doi.org/10.1121/1.2188377
Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge University Press.
Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge University Press.
Hualde, J. I., & Prieto, P. (2015). *Intonational variation in Spanish: European and American varieties*. S. Frota & P. Prieto, Eds. (pp. 350–391) Oxford University Press
Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *Journal of the Acoustical Society of America, 132*(6), 3950–3964.
Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r-l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition, 15*, 255–274.
Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics, 39*(4), 571–584. https://doi.org/10.1016/j.wocn.2011.03.003
Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America, 118*(5), 3267–3278. https://doi.org/10.1121/1.2062307
Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, B47–B57.
Jesse, A., Poellmann, K., & Kong, Y. Y. (2017). English listeners use suprasegmental cues to lexical stress early during spoken-word recognition. *Journal of Speech, Language, and Hearing Research, 60*(1), 190–198. https://doi.org/10.1044/2016_JSLHR-H-15-0340
de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America, 97*(1), 491–504. https://doi.org/10.1121/1.412275
Jun, S.-A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology, 15*, 189–226. https://doi.org/10.1017/s0952675798003571
Jun, S.-A. (2000). K-ToBI (Korean ToBI) labeling conventions. *UCLA Working Papers in Phonetics, 99*, 149–173.
Jun, S.-A. (2005). Korean intonational phonology and prosodic transcription. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 201–229). Oxford University Press.
Kager, R. (1989). *A metrical theory of stress and destressing in English and Dutch*. Foris Publications.
Kang, K.-H., & Guion, S. G. (2008). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *The Journal of the Acoustical Society of America, 119*(3), 1672. https://doi.org/10.1121/1.2166607
Kang, Y. (2014a). A corpus-based study of positional variation in Seoul Korean vowels. *Japanese/Korean Linguistics, 23*, 1–20.
Kang, Y. (2014b). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics, 45*, 76–90.
Kang, Y., Yoon, T.-J., & Han, S. (2015). Frequency effects on the vowel length contrast merger in Seoul Korean. *Laboratory Phonology, 6*(3–4), 469–503. https://doi.org/10.1515/lp-2015-0014
Kim, H., & Tremblay, A. (2021). Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics, 87*, 1–15. https://doi.org/10.1016/j.wocn.2021.101059
Kim, H., & Tremblay, A. (2022). Intonational cues to segmental contrasts in the native language facilitate the processing of intonational cues to lexical stress in the second language. *Frontiers in Communication, 7*. https://doi.org/10.3389/fcomm.2022.845430
Kim, H., Tremblay, A., & Cho, T. (2024). Perceptual cue weighting matters in real-time integration of acoustic information during spoken word recognition. *Cognitive Science, 48*(12), Article e70026. https://doi.org/10.1111/cogs.70026
Kruschke, J. K. (2010). Bayesian data analysis. *WIREs Cognitive Science, 1*, 658–676. https://doi.org/10.1002/wcs.72
Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 11850–11857. https://doi.org/10.1073/pnas.97.22.11850
Ladd, D. R. (2012). *Intonational phonology*. Cambridge University Press.
Lee, H., & Jongman, A. (2019). Effects of sound change on the weighting of acoustic cues to the three-way laryngeal stop contrast in Korean: Diachronic and dialectal comparisons. *Language and Speech, 62*, 509–530.
Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics, 41*(2), 117–132. https://doi.org/10.1016/j.wocn.2012.12.002
Lee, W.-S., & Zee, E. (2010). Articulatory characteristics of the coronal stop, affricate, and fricative in Cantonese / 广州话舌前塞音、塞擦音及擦音的颚位和舌位研究. *Journal of Chinese Linguistics, 38*(2), 336–372.
Lee, Y.-C., Wang, T., & Liberman, M. (2016). Production and perception of Tone 3 focus in Mandarin Chinese. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01058
Lehiste, I. (1970). *Suprasegmentals*. MIT Press.
Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0
Lieberman, P. (1960). Acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America, 32*, 451–454. https://doi.org/10.1121/1.1908095
Lin, C. Y., Wang, M. I. N., Idsardi, W. J., & Xu, Y. I. (2014). Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition, 17*, 316–346. https://doi.org/10.1017/s1366728913000333
Lin, T. (1985). Tantao Beijinghua qingyin xingzhi de chubu shiyan [On neutral tone in Beijing Mandarin]. In S. Hu (Ed.), *Beijing yuyin shiyanlu [Working papers in*

experimental phonetics] (pp. 1–26). Beijing: Daxue chubanshe [Peking University Press].
Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*, 1173–1181. https://doi.org/10.1121/1.2142410
Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing initial stops: Acoustical measurements. *Word, 20*, 384–422.
Liu, S., & Samuel, A. G. (2004). Perception of Mandarin Lexical Tones when F0 Information is Neutralized. *Language and Speech, 47*(2), 109–138. https://doi.org/10.1177/00238309040470020101
Ma, J., Zhu, J., Yang, Y., & Chen, F. (2021). The development of categorical perception of segments and suprasegments in Mandarin-Speaking Preschoolers. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.693366
Nakatani, L. H., O'Connor, K. D., & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica, 38*, 84–106.
Nordenhake, M., & Svantesson, J.-O. (1983). Duration of Standard Chinese word tones in different sentence environments. *Lund University Working Papers in Linguistics, 25*, 105–111.
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology. General, 115*(1), 39–61. https://doi.org/10.1037//0096-3445.115.1.39
Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics, 41*, 186–197. https://doi.org/10.1016/j.wocn.2013.01.006
Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 155–176).
Ortega-Llebaria, M., & Prieto, P. (2009). Perception of word stress in Castilian Spanish: The effects of sentence intonation and vowel type. In *Phonetics and Phonology* (pp. 35–50). John Benjamins. https://www.jbe-platform.com/content/books/9789027289001-cilt.306.02ort.
Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech, 54*(1), 73–97. https://doi.org/10.1177/0023830910388014
Ortega-Llebaria, M., & Wu, Z. (2021). Chinese-English speakers' perception of pitch in their non-tonal language: Reinterpreting English as a tonal-like language. *Language and Speech, 64*(2), 467–487. https://doi.org/10.1177/0023830919894606
Ouyang, I. C., & Kaiser, E. (2015). Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience, 30*(1–2), 57–72. https://doi.org/10.1080/01690965.2013.805795
Paschen, L., Fuchs, S., & Seifart, F. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics, 94*, Article 101179. https://doi.org/10.1016/j.wocn.2022.101179
Peperkamp, S., & Dupoux, E. (2002). A typological study of stress deafness. In C. Gussenhoven (Ed.), *7. Proceedings of Laboratory Phonology* (pp. 203–240). Mouton de Gruyter.
Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics, 38*, 422–430. https://doi.org/10.1016/j.wocn.2010.04.001
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished dissertation. MIT Press.
Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. P271–P311). MIT Press.
Pisoni, D. B., & Luce, P. A. (1987). Trading relations, acoustic cue integration, and context effects in speech perception. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 155–172). Netherlands: Springer. https://doi.org/10.1007/978-94-009-3629-4_11.
Prieto, P., & Torreira, F. (2007). The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics, 35*, 473–500.
Psychology Software Tools. (2016). *E-Prime 3.0 [Computer software]*.
Qin, Z., Chien, Y.-F., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied PsychoLinguistics, 38*, 541–570. https://doi.org/10.1017/s0142716416000321
Quam, C., & Swingley, D. (2014). Processing of lexical-stress cues by young children. *Journal of Experimental Child Psychology, 123*, 73–89. https://doi.org/10.1016/j.jecp.2014.01.010
R Core Team. (2024). *R: A language and environment for statistical computing [Computer software]*. R Foundation for Statistical Computing.
Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology, 63*, 772–783. https://doi.org/10.1080/17470210903104412
Repp, B. H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication, 2*(4), 341–361. https://doi.org/10.1016/0167-6393(83)90050-X
Roark, C. L., & Holt, L. L. (2022). Long-term priors constrain category learning in the context of short-term statistical regularities. *Psychonomic Bulletin & Review, 29*(5), 1925–1937. https://doi.org/10.3758/s13423-022-02114-z
Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics, 52*, 183–204. https://doi.org/10.1016/j.wocn.2015.07.003
Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. Wiley interdisciplinary reviews. *Cognitive Science, 11*(2), Article e1521. https://doi.org/10.1002/wcs.1521
Shen, X. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech, 36*(4), 415–433. https://doi.org/10.1177/002383099303600404
Shin, J., Kiaer, J., & Cha, J. (2012). *The sounds of Korean*. Cambridge University Press. https://doi.org/10.1017/CBO9781139342858
Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology, 23*, 287–308.
Sluijter, A. M. C., & van Heuven, V. J. (1996a). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the International Congress of Spoken Language Processing*. University of Delaware.
Sluijter, A. M. C., & van Heuven, V. J. (1996b). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*, 2471–2485.
Small, L. H., Simon, S. D., & Goldberg, J. S. (1988). Lexical stress and lexical access: Homographs vs. Nonhomographs. *Perception & Psychophysics, 44*, 272–280.
Sohn, H. M. (1999). *The Korean Language*. Cambridge University Press.
Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*(1), 25. https://doi.org/10.1186/1471-2105-8-25
Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science, 34*(3), 434–464. https://doi.org/10.1111/j.1551-6709.2009.01077.x
Tremblay, A., Broersma, M., Zeng, Y., Kim, H., Lee, J., & Shin, S. (2021). Dutch listeners' perception of English lexical stress: A cue-weighting approach. *Journal of the Acoustical Society of America, 149*, 3703–3714. https://doi.org/10.1121/10.0005086
Tremblay, A., & Kim, H. (2025). Perceptual plasticity in bilinguals: Language dominance reshapes acoustic cue weightings. *Brain Sciences, 15*(10), 1053. https://doi.org/10.3390/brainsci15101053
Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics, 35*(4), 445–472. https://doi.org/10.1016/j.wocn.2006.12.001
Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*(1), 4–21. https://doi.org/10.1159/000356237
Wang, F., Deng, D., Tang, K., & Wayland, R. (2024). The effect of pitch accent on the perception of English lexical stress: Evidence from English and Mandarin Chinese listeners. *Languages, 9*(3). https://doi.org/10.3390/languages9030087
Xu, C. (2024). Cross-dialectal perspectives on Mandarin neutral tone. *Journal of Phonetics, 106*, Article 101341. https://doi.org/10.1016/j.wocn.2024.101341
Xu, Q.-S., & Liang, Y.-Z. (2001). Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems, 56*(1), 1–11. https://doi.org/10.1016/S0169-7439(00)00122-2
Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*(1), 61–83. https://doi.org/10.1006/jpho.1996.0034
Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0contours. *Journal of Phonetics, 27*(1), 55–105. https://doi.org/10.1006/jpho.1999.0086
Yan, F. (2021). *A study on word stress of Mandarin disyllabic words*. Unpublished PhD Dissertation. University of Wisconsin Madison.
Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics, 24*(2), 245–261. https://doi.org/10.1006/jpho.1996.0013
Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393
Yu, H. J. (2025). Acoustic properties of English Schwa [ə] produced by Korean learners. *The Linguistic Association of Korea Journal, 33*(4), 163–181. https://doi.org/10.24303/lakdoi.2025.33.4.163
Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics, 38*, 260–271. https://doi.org/10.1016/j.wocn.2009.11.002