

JOURNAL ARTICLE

Iambic-Trochaic Law Effects among Native Speakers of Spanish and English

Megan Crowhurst

The University of Texas at Austin, US

mcrowhurst@austin.utexas.edu

The Iambic-Trochaic Law (Bolton, 1894; Hayes, 1995; Woodrow, 1909) asserts that listeners associate greater intensity with group beginnings (a loud-first preference) and greater duration with group endings (a long-last preference). Hayes (1987; 1995) posits a natural connection between the prominences referred to in the ITL and the locations of stressed syllables in feet. However, not all lengthening in final positions originates with stressed syllables, and greater duration may also be associated with stress in nonfinal (trochaic) positions. The research described here challenged the notion that presumptive long-last effects necessarily reflect stress-related duration patterns, and investigated the general hypothesis that the robustness of long-last effects should vary depending on the strength of the association between final positions and increased duration, whatever its source. Two ITL studies were conducted in which native speakers of Spanish and of English grouped streams of rhythmically alternating syllables in which vowel intensity and/or duration levels were varied. These languages were chosen because while they are prosodically similar, increased duration on constituent-final syllables is both more common and more salient in English than Spanish. Outcomes revealed robust loud-first effects in both language groups. Long-last effects were significantly weaker in the Spanish group when vowel duration was varied singly. However, long-last effects were present and comparable in both language groups when intensity and duration were covaried. Intensity was a more robust predictor of responses than duration. A primary conclusion was that whether or not humans' rhythmic grouping preferences have an innate component, duration-based grouping preferences, at least, and the magnitude of intensity-based effects are shaped by listeners' backgrounds.

Keywords: phonology; metrical phonology; stress; rhythm; Iambic-Trochaic Law

1 General overview

Metrical phonologists have argued that the patterned alternation of stressed and unstressed syllables in languages motivates the inclusion of the FOOT in a universal hierarchy of prosodic constituents (Halle & Vergnaud, 1987; Hayes, 1985; 1995; Selkirk, 1980; 1984). In arguing for a particular typology of stress feet, Hayes (1987; 1995) points to research by Bolton (1894) and Woodrow (1909), who discovered that the perception of recurrent groupings in patterned sound sequences can be shaped by fluctuations in acoustic duration and intensity. These early studies produced the generalizations in (1), now known as the Iambic-Trochaic Law, or ITL (Bolton, 1894, p. 232; Hayes, 1995, p. 80; Woodrow, 1909, pp. 69, 77). The statement in (1) is nearest to Woodrow (1909, p. 77; parenthesized material is the author's).

(1) *Iambic–Trochaic Law:* Intensity has a group-beginning effect (the loud-first principle), duration, a group-ending effect (the long-last principle).

Hayes (1987; 1995) posits a natural association between the prominences referred to in the ITL (greater intensity and duration) and locations of stressed syllables in feet. In particular, he interprets the long-last principle as support for his position that feet that build

in a weight asymmetry (for Hayes, the natural extension of a duration asymmetry) are iambs: the heavy syllable is both stressed and foot-final. In Hayes's typology, balanced stress feet (those with no weight asymmetry) are trochaic by default.¹

The claim of a natural association between the ITL and locations of stressed syllables in feet should be cautiously examined. On one hand, speakers of diverse languages are known to associate greater vowel intensity and duration with stress (for English, see Beckman, 1986; Turk & Sawusch, 1996; for Dutch, Sluijter & van Heuven, 1996; for Spanish, Ortega-Llebaria & Prieto, 2007; for Catalan, Ortega-Llebaria et al., 2010; and for a general review, Fletcher, 2010). However, on closer inspection, elements of Hayes's proposal invite skepticism. It is unclear to what extent ITL effects are broadly general; theorists should therefore be wary of assuming connections with the foot typology. Furthermore, listeners' use of intensity and duration patterns to locate stressed syllables does not necessarily mean that the prosodic units they live in correspond to those implied by the ITL, on Hayes's interpretation. In fact, if there is a natural association between listeners' perceptions of natural syllable groupings and stress-related intensity/duration patterning, then ITL effects should not be universal, or at least equally robust, across languages in which stress is differently distributed.

The ITL may reflect basic perceptual predispositions to some extent; that all reliable reports for intensity confirm the Loud-First Principle suggests this much (see Section 2.1). On the other hand, findings for duration have been less consistent (see Section 2.1), which suggests that rhythmic grouping preferences can be shaped by experience. It is therefore interesting to ask whether listeners' syllable grouping preferences reflect language-specific duration patterning. Phonetic length has multiple sources; two were of special interest for the current research. Increased duration may be due to preboundary lengthening, the lengthening of final syllables in prosodic domains from the prosodic word (Beckman & Edwards, 1994) to the highest levels of the prosodic hierarchy (Beckman & Edwards, 1994; Cambier-Langeveld, 1997; Turk & Shattuck-Hufnagel, 2000; 2007; Wightman et al., 1992). Length is also a physical correlate of stress in diverse languages. Importantly, this is true not only of languages with iambic stress patterns (see Buckley, 1998, for an overview of languages with so-called iambic lengthening), but also of languages with trochaic word stress, such as English (Beckman & Edwards, 1994), Dutch (Sluijter & van Heuven, 1996; Rietveld et al., 2004), Spanish (Ortega-Llebaria & Prieto, 2007; 2011), and Catalan (Ortega-Llebaria & Prieto, 2010). Different language-specific patterns may lead to different subjective estimates of the likelihood of increased length in specific prosodic positions, and these different expectations may affect listeners' preferences when they are tasked with grouping speech-like sequences that resemble alternating stress patterns.

Using the ITL as a point of departure, the primary research question was this: *Will long-lasting effects be more robust among speakers of languages in which prosodic length is more strongly associated with constituent-final syllables?* This question was tested by conducting two rhythmic grouping experiments with native speakers of Mexican Spanish and American English. Spanish and English are prosodically similar in that both have a word-level trochaic stress pattern, stress-related lengthening, and preboundary lengthening. However, they differ crucially in that prosodic length is more strongly associated with constituent-final syllables in English than in Spanish. Length on final syllables in English

¹ This is consistent with the Loud-First Principle, but does not preclude the possibility that phonetic features other than intensity are associated with the perception of trochaic units. Some research suggests that higher pitch (Bion et al., 2011) and vowel creakiness (Crowhurst et al., 2016) may also be associated with perception of trochaic groupings.

can be related to both stress and preboundary lengthening, whose effects are additive, which enhances the duration contrast between final and nonfinal syllables (Beckman & Edwards, 1994). In Spanish, final syllables are less likely to be longer both because word-final stress is less common and because stress- and boundary-related lengthening patterns are less pronounced than in English (see Section 2.3). These facts suggest that increased length on final syllables might be more natural for English than Spanish speakers, and it was therefore expected that long-last grouping effects should be stronger among the English than the Spanish speakers in a rhythmic grouping study. In fact, if listeners' grouping preferences are influenced by stress-related lengthening patterns, the Long-Last Principle might well not be found with Spanish speakers, as word-final stress in Spanish is a marked pattern. English and Spanish make a good comparison pair for the current research precisely because their similarities make it possible to attach specific predictions to the crucial differences, and test for fine-grained differences in behaviour. The demonstration of subtle differences in relation to predictions is an important component of laboratory phonology research, and this paper contributes in that area.

The study had two additional objectives beyond the theoretical motive described above. The first was to make a novel contribution by testing the rhythmic grouping preferences of Mexican Spanish speakers, a previously unstudied group, and to compare them with the better-studied preferences of English speakers (see Section 2.1). In order for any claims of language-specific differences in speech processing to be convincing, the two groups were tested using the same stimuli, streams of syllables in which intensity and/or duration levels were manipulated. Experiment 1 was a basic test of the ITL in which intensity and duration were varied singly, as in Hay and Diehl (2007), Bhatara et al. (2013), and Crowhurst and Teodocio (2014). However, as natural speech is multidimensional, the second objective was to learn more about how intensity and duration work together to influence listeners' grouping behaviour. To this end, Experiment 2 used a more complex design that included conditions in which vowel intensity and duration were varied orthogonally in the same sequences (Crowhurst & Teodocio, 2014).

Going forward, the next section provides background in three areas: the most relevant ITL literature is reviewed in Section 2.1, followed by a description of the methodological antecedent for this research (Crowhurst & Teodocio, 2014) in Section 2.2. Essential background for Spanish and English is provided in Section 2.3, followed by specific hypotheses in Section 2.4. Experiments 1 and 2 and their outcomes are described in Section 3 and Section 4, followed by a more detailed examination of the results in Section 5. The paper concludes with a general discussion in Section 6.

2 Background

2.1 The rhythmic grouping literature

The most widely cited historical antecedents for modern ITL studies are Bolton (1894) and Woodrow (1909), who tested the preferences of adult American English speakers who were asked to group patterned sequences of tones or clicks in which intensity and duration were systematically varied. Their early findings provided the foundation for the generalizations in (1). In addition to studying intensity and duration varied singly, Bolton (1894) investigated the effect of varying them orthogonally. He found that when listeners segmented sequences that alternated a long soft tone with a short loud one, listeners preferred the grouping in which the loud tone came first and the long came last. This grouping is consistent with both the Loud-First and Long-Last principles of the ITL. However, for sequences in which a short soft sound alternated with a louder sound that was twice as long, listeners preferred iambic groupings. In a contest, therefore, the dominant

influence was duration. Unfortunately, Bolton does not support his generalization quantitatively with a detailed statistical analysis, as is standard today; nor does his study answer the question of whether incrementally increasing a sound's intensity while keeping its duration constant would modulate the effect of the more dominant feature.

In more contemporary ITL studies that serve as updated methodological antecedents for the current research, adult participants are presented with streams of sounds in which different values for acoustic intensity or duration are alternated at fixed ratios, typically in a binary . . . ABABA . . . pattern. Participants are tasked with indicating whether they think these sequences break most naturally into AB or BA pairings. Studies using this or a similar method have consistently associated a loud-first grouping bias with varied intensity, as predicted by the ITL. These include studies in which non-speech sounds (tones or beeps) have been used to test English speakers (Rice, 1992); English and Japanese speakers (Iversen et al., 2008; Kusumoto & Moreton, 1997); and English and French speakers (Hay & Diehl, 2007). Bell (1977) reports a loud-first preference for small groups of subjects representing five languages (Bengali, English, French, Persian, and Polish) who were asked to group tones of different intensities that alternated in a ternary pattern. Three key studies for speech in which listeners segmented alternating sequences of syllables also report loud-first preferences for English and French speakers (Hay & Diehl, 2007); French and German speakers (Bhatara et al., 2013); and Zapotec and English speakers (Crowhurst & Teodocio, 2014).

Findings for duration in both the speech and non-speech studies cited above have been more mixed. A long-last grouping bias has been consistently reported for English speakers (Crowhurst & Teodocio, 2014; Hay & Diehl, 2007; Iversen et al., 2008; Kusumoto & Moreton, 1997), French speakers (Bhatara et al., 2013; Hay & Diehl, 2007), and German speakers (Bhatara et al., 2013). Vos (1977) reports a long-last grouping preference among Dutch speakers when the interval between non-speech tones was held constant. However, two nonspeech studies have failed to find a consistent grouping preference with Japanese speakers (Iversen et al., 2008; Kusumoto & Moreton, 1997). Interestingly, Iversen et al. (2008) report not merely inconsistent, but conflicting findings among groups of Japanese speakers they tested. One group (26% of their Japanese participants) showed a clear long-last grouping bias. A second group (nearly 45%) had a clear preference for long-short groupings. Iversen et al.'s remaining Japanese participants showed no consistent preference for either grouping.² A more nuanced set of outcomes for duration among speakers of Zapotec and English is reported by Crowhurst and Teodocio (2014), the antecedent for the current research, described in Section 2.2.

2.2 *The methodological antecedent*

Crowhurst and Teodocio (2014) studied rhythmic grouping preferences with native speakers of Zapotec (Otomanguean) and a comparison group of American English speakers. Theirs was the first ITL study to test speakers of a non-Indo-European language using speech-based stimuli, and the first modern ITL study to vary intensity and duration orthogonally as well as singly. In Crowhurst and Teodocio's first study, Zapotec speakers were presented with 10–11 second sequences in which the syllables *de* and *ge* were alternated

² Bell (1977) reports a dispreference for short-short-long groupings when tones of two lengths are alternated in a ternary fashion. However, in his length-varied sequence, the longer tone was always followed by a shorter inter-tone interval. This made Bell's results difficult to interpret, as it introduced a conflict with a second Gestalt principle: listeners tend to perceive sounds as a coherent group to the extent that the interval separating them is shorter than other intervals. Other studies reporting this finding include Bolton (1894), Woodrow (1909), and Vos (1977).

at a rate of 4–5 syllables per second. In test conditions, either vowel intensity or duration was manipulated singly. In varying intensity, a fixed disparity of 4, 8, or 12 dB was introduced between alternating syllables. For duration, fixed disparities of 40, 80, or 120 ms were used. Subjects' responses to varied intensity/duration were tested indirectly by having them indicate whether they thought sequences broke most naturally into recurrent *dege* or *gede* syllable pairs. In this study, Crowhurst and Teodocio found no evidence for consistent grouping preferences.

Outcomes were very different in a more extended follow-up study that added conditions in which vowel intensity and duration were co-varied. In a Competing condition, one alternating syllable was both louder and longer than the other (e.g., *DEE* vs. *ge*). In this condition, no syllable pairing (e.g., *ge-DEE*, *DEE-ge*) was fully consistent with the ITL and the investigators expected that intensity and duration cues should work against one another. In a Co-operating condition, a long/soft syllable was alternated with a short/loud one. Here, one grouping (e.g., *DE-gee*) was doubly favoured by the ITL and intensity/duration cues were therefore expected to work together. Crowhurst and Teodocio report a robust loud-first preference for both Zapotec and English speakers in all conditions in which intensity was manipulated. The function from intensity disparity to responses representing loud-first groupings was curvilinear: the increase from 0 to 4 dB of disparity had a greater effect on listeners' perceptions than the increase from 4 to 8 dB of disparity (and similarly for the increase from 8 to 12 dB of disparity), and this was generally true in both language groups. While this outcome has not been reported in other recent ITL studies, it is consistent with an observation made by Woodrow (1909) over a century ago: "With an increase in the ratio of the intensity of the louder sound to that of the weaker, there is an increase, first rapid and then slow, in the tendency of the more intense sound to begin the group" (Woodrow, 1909, p. 64).

Crowhurst and Teodocio report that varied duration was associated with different outcomes depending on condition, both within and between groups. When duration was manipulated singly, increasing the duration disparity increased long-last pairings in the English group (as it has done in other studies; see Section 2.1). In contrast, the Zapotec group favoured long-short groupings. Outcomes were more similar across language groups when intensity and duration were co-varied. The dominant influence in both co-varied conditions was associated with intensity, reflected in an overall loud-first grouping bias. But in the Competing condition, increasing the duration disparity reduced the effect of intensity. That is, when an intensity disparity was held constant, increasing the duration disparity increased long-last pairings (e.g., *ge-DEE*) in both language groups, although not to the point that they exceeded loud-first pairings (e.g., *DEE-ge*). The outcome for duration in the Co-operating condition was different: in both groups, increasing the duration disparity increased long-short groupings. Crowhurst and Teodocio suggested an auditory (as opposed to linguistic) explanation relating to the auditory system's integration of intensity and duration over time: when sounds of the same intensities have different lengths, the longer sound is perceived to be louder (see Gordon, 2005, and references cited there). Hence, increasing the length disparity in the Co-operating condition may have partially closed the gap between longer, soft syllables and louder, short syllables.

Crowhurst and Teodocio drew several conclusions from their findings. First, listeners may process acoustic cues differently when multiple cues fluctuate together than when they are varied singly. Second, the loud-first grouping bias in the co-varied conditions (regardless of how duration and intensity were varied) among both Zapotec and English speakers suggests that intensity may be a stronger predictor than duration of syllable grouping preferences, regardless of language exposure. Finally, effects associated with

varied duration when subjects perform a grouping task (in this and prior published studies) seem to be less stable than those for varied intensity, as they depended on whether and how duration was co-varied with intensity. These conflicting findings for duration provided the inspiration for the main focus on duration in the studies reported here. Accordingly, the current research adopted Crowhurst and Teodocio's design and procedures to study the theoretical questions of central interest in the current research and to see if the patterns Crowhurst and Teodocio describe could be replicated in a new language.

2.3 Relevant prosodic characteristics of Spanish and English

Phonetic research has shown that measures of intensity and duration are closely associated with the stress contrast. Stressed syllables are reliably longer than unstressed syllables in many languages, and listeners use this information to identify them in speech.³ Measures of intensity are less reliably associated with stress across languages, at least for unaccented stressed syllables (Ortega-Llebaria & Prieto, 2011), although intensity may be important in combination with duration (e.g., Sluijter et al., 1997; Turk & Sawusch, 1996). These generalizations hold true for the languages under study here: stressed syllables are lengthened in both Spanish (Ortega-Llebaria & Prieto, 2007; 2011) and English (e.g., Beckman & Edwards, 1994; Campbell & Beckman, 1997), although the contrast is more pronounced in English due to vowel reduction in unstressed syllables (e.g., Beckman & Edwards, 1994; Lindblom, 1963; Ortega-Llebaria et al., 2007; Ortega-Llebaria & Prieto, 2011). Stressed syllables in higher-level prosodic constituents (pitch accents) tend to be longer than lower-level stressed syllables in English (Beckman & Edwards, 1994). There is less evidence that this is so in Spanish (Ortega-Llebaria & Prieto, 2007; 2011). Evidence associating greater intensity with stress in Spanish and English has been more conflicting. While Ortega-Llebaria and Prieto (2007) concluded that spectral tilt was a reliable correlate of stress in Spanish (though less robust than duration, and in deaccented syllables), Ortega-Llebaria and Prieto (2011) did not find this to be the case. For American English, Sluijter and van Heuven (1996) report that spectral balance differences but not overall intensity differences mark unaccented stressed syllables. Campbell and Beckman (1997) did not find this difference; they associate spectral balance differences with stress-related vowel reduction (a point Ortega-Llebaria and Prieto, 2011, have made for Catalan). However, Kochanski et al. (2005) report that overall vowel intensity differences reliably mark word-level stress contrasts in some UK dialects of English. In perception, Ortega-Llebaria et al. (2007) report that their Spanish-speaking subjects relied most on duration to distinguish unstressed from (unaccented) stressed syllables, but that discrimination was better when differences in overall intensity were also present. Numerous studies have found length to be the more reliable perceptual cue to stress in English as well (e.g., Beckman & Edwards, 1994; Campbell & Beckman, 1997; Kochanski et al., 2005; Turk & Sawusch, 1996). Turk and Sawusch (1996) found that intensity differences can help listeners identify stressed syllables, but not in the absence of duration cues.

In addition to its close association with stress, increased duration is associated with constituent-final syllables in both English and Spanish. In English, preboundary lengthening has been observed across accentual contexts and at multiple levels of the prosodic hierarchy (Beckman & Edwards, 1994; Byrd, 2000). Beckman and Edwards (1994) and Turk and Shattuck-Huffnagel (2000) report a degree of preboundary lengthening as low

³ Production studies: Catalan (Ortega-Llebaria & Prieto, 2011); Dutch (den Os, 1988; Sluijter & van Heuven, 1996); German (Dogil & Williams, 1999); Italian (den Os, 1988); and Swedish (Heldner & Strangert, 2001). Perception studies: Catalan (Ortega-Llebaria et al., 2010), Dutch (Sluijter et al., 1997), and Italian (Alfano et al., 2007). For a general review of research in both areas, see Fletcher (2010).

in the hierarchy as the prosodic word, although effects are more pronounced at higher levels. For Spanish, Rao (2010) reports preboundary lengthening effects at the phonological and intonational phrase levels in the productions of speakers from Cuba, Ecuador, and Spain. In a recent tri-language study, Prieto et al. (2010) found evidence for preboundary lengthening at the ends of intermediate (i.e., phonological) and intonational phrases in Spanish, Catalan, and English. These authors report that both preboundary lengthening effects and stress-related duration differences are greater in English than in Spanish and Catalan.

Most theoretical accounts of stress posit a word-final, trochaic main stress foot in both Spanish and English (see Harris, 1983, for Spanish; Hayes, 1982; 1985, for English). However, while precise quantitative data is not available, a case can be made that they differ in the frequency with which stress occurs on word-final syllables. Domahs et al. (2014) report that primary stress fell on the penult in 61.1%; on the antepenult in 27.4%; and on the final syllable in 11.5% of polysyllabic words in a 1,160-item sample of the English lexicon. However, the prevalence of stress on final syllables in English is higher than 11.5%, given that heavy final syllables have secondary stress in words with nonfinal main stress (e.g., *sálivàte*, *níghtingàle*), a class that notably includes all compounds in which the second element is a monosyllable (e.g., *sidewàlk*, *bubblegùm*). Nor did the sample analyzed by Domahs et al. (2014) include stressed monosyllables, which have final as well as initial stress. These accounted for 11.4% of items in a 33,060-word lexical sample extracted from the Shorter Oxford English Dictionary and analyzed by Cutler and Carter (1987). There is also reason to believe that the percentage of stressed, word-final syllables in connected speech is higher than in the lexicon. The speech that English speakers hear contains a disproportionately large number of stressed monosyllables: information in Cutler and Carter (1987, p. 138, calculated from Table VII) indicates that in a 187,699-word sample of spontaneous British English conversation, 59.4% of the 76,963 tokens of lexical words were stressed monosyllables whereas only 28.2% were polysyllables with initial primary stress. The prevalence of stressed monosyllables in speech is important, because they combine with preceding unstressed function words, prepositions, or pronouns to form higher-level prosodic constituents that may be as short as a foot (e.g., *D'she GO? They're HERE!*) Therefore, while word-initial stress is very common in English and plays a role in speech segmentation (Cutler & Butterfield, 1992), stress on domain-final syllables in speech is also quite common. As stress-related and preboundary lengthening effects are additive (Beckman & Edwards, 1994), it seems likely that longer syllables in domain-final positions are natural for English speakers.

Stress on domain-final syllables is arguably less common in Spanish than in English, both in the lexicon and especially in connected speech. Estimates place primary stress on penultimate syllables in over 80% of lexical words (Morales-Front, 1999; Quilis, 1993) or 74%, weighted for frequency (Sebastian & Costa, 1997). Nonfinal stress also occurs in a minority of lexical words with antepenultimate stress (e.g., *Sábado* 'Saturday', *próximo* 'next'). While again, precise estimates for actual speech were not available, the morphological characteristics of Spanish suggest that the frequency of stress on final syllables is less common in speech than in the lexicon, and is less common than in English overall. First, due to Spanish's robust inflectional system, relatively few major class words are monosyllabic, in proportion to polysyllables, even in isolation. Most nouns and agreeing adjectives end in unstressed class-marking vowels (cf. English *cát*, Spanish *gáto*; all stressed vowels are given diacritic accents for clarity). Second, while uninflected consonant-final words tend to have final stress (Lipsky, 1997; e.g., *ciudad* 'city'), as do a small minority of vowel-final words (e.g., *menú* 'menu'), these take the plural allomorph *-es*, which adds an unstressed

syllable (e.g., *ciudadés* ‘cities’, *menúes* ‘menus’). Third, while infinitive verbs have final stress, regularly conjugated finite verbs generally have nonfinal stress (with exceptions in the future and preterite tenses). Fourth, some encliticised verbs with antepenultimate (or even preantepenultimate) stress occur frequently in discourse (e.g., *ffíjate* ‘look; think about it!’, *pásale* ‘come in’, *recuérdate* ‘remember!’). Finally, because stressable words may be deaccented in sentential contexts, some word-final syllables that might be stressed in isolation are unstressed in speech (Hualde, 2007).

Summarizing, increased length on constituent-final syllables can have more than one source (stress and preboundary lengthening), and duration contrasts are saliently marked in English. Final syllables in Spanish are less likely to be longer because final stress is less prevalent and duration contrasts are less salient than in English. For these reasons, it is reasonable to anticipate that Spanish speakers might have a lower expectation than English speakers of stress-related duration cues in final positions, and I expected this difference to be reflected in the outcomes of rhythmic grouping studies in which speech-based stimuli were used.

2.4 Hypotheses

Expectations for the current studies were to some extent determined by the findings of prior studies. Comparable ITL studies that have tested intensity-based rhythmic grouping preferences have confirmed the Loud-First Principle (see Section 2.1), and I expected to replicate these findings in the current research in both language groups. This prediction is operationalised as the *Incremental Loud-First hypothesis* in (2a). Two outcomes reported in Crowhurst and Teodocio (2014) were also important in setting expectations. The first was their finding that when intensity disparities were incrementally increased, intensity effects were greatest at lower levels, and were attenuated at the highest levels of the scale. This expectation is stated as the *Intensity Attenuation hypothesis* in (2b). Second, their findings indicated that intensity effects were consistently more pronounced than effects of duration, leading to the *Intensity Wins hypothesis* in (2c).

(2) Hypotheses for intensity

- a. *Incremental Loud-First hypothesis*: Increasing the intensity disparity will increase loud-first groupings in both the Spanish and the English participant groups.
- b. *Intensity Attenuation hypothesis*: The effect per unit of increase in intensity will be inversely correlated with the magnitude of the intensity disparity.
- c. *Intensity Wins hypothesis*: The magnitude of effects associated with intensity will be greater than effects associated with duration.

The hypotheses of greater importance for the theoretical questions of interest were associated with duration. The basic question driving the research was whether language-specific differences in the distribution of longer syllables would be reflected in the decisions of Spanish and English speakers when segmenting duration-varied sequences. Multiple studies have confirmed a long-last grouping preference for English speakers, and I expected the same outcome here. The operationalization of the Long-Last Principle adopted here is stated as the *Incremental Long-Last hypothesis* in (3b). It was argued in Section 2.3 that Spanish speakers should have a weaker association between increased length and constituent final syllables than English speakers, and I expected that this difference should be reflected in listeners’ duration-based rhythmic grouping preferences. Accordingly, one of the study’s primary hypotheses was that effects related to varied duration should be weaker in the Spanish than in the English group. This expectation is stated in hypothesis (3a). Expectations as to the nature of any duration-based grouping preferences for the

Spanish group were less clear; because distributional patterns associated with duration are less convergent in Spanish than in English (both languages have preboundary lengthening, but stress-related lengthening is associated more consistently with nonfinal syllables in Spanish than English), there were competing predictions for the Spanish group. If the Spanish speakers' rhythmic grouping perceptions were more strongly influenced by boundary-related duration cues, then the outcome should confirm the Incremental Long-Last hypothesis in (3b). Alternatively, if their preferences were more strongly shaped by stress-related duration patterning, they should prefer long-short groupings, as hypothesized in (3c). Finally, the possibility of a third outcome was considered: as duration cues related to stress and prosodic boundaries conflict more in Spanish than English, the Spanish speakers might not show a clear grouping preference when duration was varied. This would also be consistent with the hypothesis in (3a).

(3) Hypotheses for duration

- a. The magnitude of duration-related effects will be smaller in the Spanish than the English group.
- b. *Incremental Long-Last hypothesis*: Increasing the duration disparity will increase long-last groupings.
- c. *Incremental Long-First hypothesis*: Increasing the duration disparity will increase long-short groupings.

3 Experiment 1

Participants were exposed to streams of CV syllables in which vowel intensity and duration were varied singly. This study provided a baseline against which to compare results obtained in Experiment 2, which had a more complex design. A baseline study was important because Crowhurst and Teodocio (2014) established that singly-varied sequences may be perceived differently depending on the context in which they are presented (whether or not they are intermingled with covaried sequences).

3.1 Method

3.1.1 Stimulus production

The stimulus set for the study included sequences of 10–11 seconds in which two coarticulated syllables, *ba* and *ga*, were alternated in a binary ABAB... pattern at a rate of 4–6 syllables per second. Only one feature was varied, so that every other syllable was either louder or longer than its neighbours. Stimulus preparation began with natural speech. An adult male native English speaker recorded fluid, co-articulated streams of alternating *ba* and *ga* syllables at a rate of approximately 4 syllables per second. Recording was accomplished using a Marantz PMD660 solid-state digital recorder and Shure SM10A-CN dynamic head-worn microphone in a sound-treated room designed for this purpose. Because the speaker tended naturally to emphasize every other syllable, half of his recorded sequences began with *ga* and half with *ba*. A clear, modally voiced token of *ba* and one of *ga* were selected for further manipulation in Praat (version 5.3.13, Boersma and Weenink, 2011). These tokens were cut at zero crossing lines to include the sound wave from the beginning of the stop burst to the end of the vowel, visually identified as the point at which F1 and F2 decreased sharply in energy. The tokens were selected from naturally emphasized positions to avoid vowel reduction effects. They were chosen to be a close natural match for pitch, vowel duration, and intensity characteristics before further editing. Average F_0 was 98 Hz and ranged between 93.7 and 100.9 Hz for both syllables. Original vowel duration was 206 ms in *ba* and 201 ms in *ga*. Average intensity was 70 mean-level dB (henceforth dB) for *ba* and 69.7 dB for *ga*, and their intensity contours

(evenly distributed over the vowel) were similar. The consonants [g] and [b] measured 20 ms from the beginning of the burst to the onset of voicing.

The selected *ba* and *ga* tokens were copied and edited using standard Praat functions to produce 10 versions each for use in constructing stimuli. Five syllables whose vowels measured 58, 61, 64, 67, and 70 dB were produced for an intensity-varied set by changing gain over the entire syllable. A duration-varied series used five syllables whose vowels measured 110, 140, 170, 200, and 230 ms. Changes to vowel duration were made by copying and inserting or removing full voicing cycles within the steady-state portion of the original vowel. Sites for editing were subjectively chosen, taking care to avoid abrupt local changes in energy and to preserve the characteristic shape of the sound wave. Vowel duration was held constant at 230 ms in the intensity set, and intensity was fixed at 64 dB in the duration set.

To produce a set of test stimuli, four intensity-varied sequences were prepared in which the loudest (70 dB) syllable was alternated with a softer syllable to produce a fixed disparity of 3, 6, 9, or 12 dB. The maximum disparity of 12 dB was determined with reference to Hay and Diehl (2007). Four duration-varied sequences were created by alternating the longest syllable with a shorter syllable to produce sequences with a fixed disparity of 30, 60, 90, or 120 ms. A control sequence was manufactured by alternating the 64 dB *ba* and *ga* syllables, whose vowels measured 230 ms. Sequences ranged from 10 to 11 seconds in length and had between 28 and 34 syllables (depending on the syllables' durations). All sequences consisted of a whole number of syllable pairs, either *ba-ga* or *ga-ba*. In sequencing, a stop closure was simulated by inserting 100 ms of low amplitude (15 dB) white noise between syllables. This was done because the researcher and others consulted judged sequences with the low-intensity inter-syllable intervals to sound more natural than sequences with silent intervals. As in Hay and Diehl (2007), sequence onsets were masked by blending the first 5 seconds of the sequence with a segment of white noise, created in Praat. Over this 5-second period, the noise was ramped down from its maximum of 67 dB to 0 dB, and the alternating syllables were ramped up from 20 dB to their maxima.⁴ A 500-ms segment of 67 dB white noise was added to sequence endings to backwards-mask the string-final syllable, should participants listen for that long. The design according to which intensity and duration levels were manipulated in the stimulus set is shown in **Table 1**.

The complete stimulus set was larger than the 9 combinations indicated in **Table 1** because it was counterbalanced in two ways. First, the stimulus set was counterbalanced for sequence-initial syllable (*ba* vs. *ga*). Second, the prominent syllable was counterbalanced: sequences in which *ba* was the louder/longer syllable were matched by sequences in which *ga* was prominent. Counterbalancing the prominent syllable was done to reduce tedium and to discourage the expectation that a particular syllable would always be the longer/louder one, as subjects might then develop unhelpful response strategies.⁵ It was expected that listeners' preferences would be similar regardless of whether *ba* or *ga* was the more prominent syllable. These methods resulted in a total of 34 distinct sequences for Experiment 1 (32 test and 2 control sequences).

⁴ The white noise was created in Praat at a sampling frequency of 48,000 and was length-matched to sequences of approximately 5 seconds. The noise segment was scaled down using an amplitude tier. Another amplitude tier was used to increase the volume of the length-matched syllable sequence from 20 dB to the full volume of its last syllable. The noise and syllable sequences were assembled into a stereo sound file in Praat, then blended by converting the stereo to a mono sound file. The resulting sequence was concatenated with an unmodified copy of the same syllable sequence to the right.

⁵ A similar design was used in a perception experiment reported in Sluijter et al. (1997).

Set, sequence	Duration (ms)	Intensity (dB)	F_0 (Hz)	Absolute difference
Duration level 4	230 & 110	64	100	120 ms
Duration level 3	230 & 140	64	100	90 ms
Duration level 2	230 & 170	64	100	60 ms
Duration level 1	230 & 200	64	100	30 ms
Control	230 & 230	64	100	0 ms
Intensity level 1	230	70 & 67	100	3 dB
Intensity level 2	230	70 & 64	100	6 dB
Intensity level 3	230	70 & 61	100	9 dB
Intensity level 4	230	70 & 58	100	12 dB

Table 1: Parameters for alternating test sequences in Experiment 1.

Overall, the stimulus set conformed to a scale with 9 different manipulation levels, shown in **Figure 1**. This scale represented the increasing intensity/duration of *ba* relative to *ga* in increments of 3 dB or 30 ms. At the negative end of the scale, *ba* was shortest/softest, and at the positive end longest/loudest. Level 0 on the scale represented the control sequences.⁶

The statistical analysis measured the likelihood of a *baga* grouping decision (the arbitrarily chosen, positively coded response category). Returning to the ITL and specific hypothesis in Section 2.4, the Loud-First Principle in (1) would be confirmed if the proportion of *baga* responses was low at the negative end of the magnitude of intensity (MOI) scale where *ga* was louder than *ba*, and high at the positive end of the MOI scale where *ba* was louder. A preference for long-last groupings would be indicated if the proportion of *baga* responses was high at the negative end of the magnitude of duration (or MOD) scale where *ga* was longer than *ba*, and low at the positive end of the scale where *ba* was longer. The incremental versions of the Loud-First and Long-Last Principles would be confirmed if changes in *baga* responses preserved the order of incremental increases on the scale in the directions indicated in hypotheses (2a) and (3b).

3.1.2 Participants

The participants were 28 native English-speaking college-aged adults (8 males and 20 females) and 29 native Spanish-speaking college-aged adults (10 males and 19 females). The English speakers were undergraduate students at the University of Texas at Austin. The Spanish speakers were undergraduates at La Salle University in Obregon City, Sonora, and were recruited by a La Salle instructor. Participants reported having no hearing impairments and having not lived in a setting where a non-native language had been spoken as a primary or secondary language. Although most participants had studied another language in high school or college, none reported fluency in that language. The compensation was \$10 for the English-speaking and 50 Mexican pesos for the Spanish-speaking participants.

3.1.3. Experimental procedures

Experimental and informed consent procedures followed a protocol approved by the Institutional Review Board at the University of Texas at Austin. English sessions were conducted by the author or a trained research assistant in a modern campus phonetics

⁶ The scale could as well be described as representing the decreasing prominence of *ga* relative to *ba*.

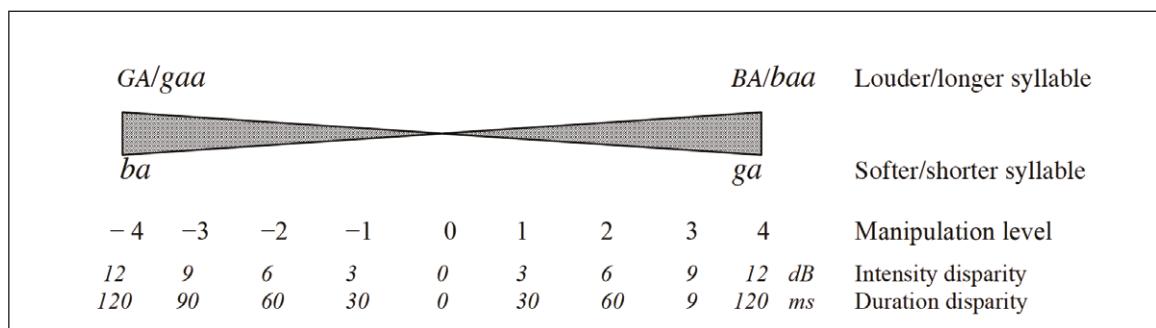


Figure 1: Manipulation scale for intensity and duration used in Experiment 1.

laboratory. The Spanish sessions were conducted in a classroom at La Salle University in Obregon City by the La Salle instructor who had recruited the participants, whose native language is Spanish. Participants heard sequences in free field, a method for presenting stimuli that has been used successfully by Sluijter et al. (1997) and Iversen et al. (2008). Sequences were played over a high-quality Bose speaker connected to an Apple MacBook computer running SuperLab 4.0. Testing was preceded by two intensity- and two duration-varying practice trials. Test trials were presented in 3 blocks of 36 (the 32 test sequences presented once and the 2 controls presented twice). Each block was presented in balanced sets of 18 followed by a brief rest. Consecutive sequences within each block of 18 were played one immediately after the other with no pause, which provided subjects with enough time to respond, but not to linger. Sequence order was automatically randomized by SuperLab every time a block was run. Participants recorded their decisions in 8-page response booklets consisting of (i) a cover sheet soliciting answers to screening questions and participants' informed consent; (ii) a page printed with instructions and four lines for the practice trials; and (iii) 6 labeled response sheets. On response sheets, 18 numbered lines were printed with arbitrary sequences of alternating *ba* and *ga* syllables. Lines with different syllable arrangements were ordered differently on successive response sheets. To the right on each line was a space in which subjects could indicate whether they were confident of their response ("yes" or "si" to indicate certainty; "no" if they were uncertain). Except for language, all subjects received the same booklets. A segment from a response sheet is shown in Appendix A.

Participants were instructed to indicate whether they thought sequences consisted of repeated *baga* or *gaba* units by bracketing any pair of adjacent syllables on the line for the current trial (e.g., ...*ba* [*ga ba*] *ga ba*...), and to provide a confidence rating by circling *yes/si* or *no*. They were instructed to wait until after the initial noise had ended and the syllables had come up to their full volumes before responding, and to do so while sequences were still playing. Participants were shown how to strike out the line for a trial if they were unable to respond in time. Sessions lasted approximately 45 minutes.

3.2 Results

The raw data obtained in the experiment are presented in Table A in Appendix B. This table indicates the absolute number and proportion of *baga* and *gaba* responses by manipulation level for each language group. The design provided for a maximum of 6,156 observations. There were 83 missing responses (1.3% of the total possible), so that a total of 6,073 data points were included in the analysis. Trends for the Spanish-speaking (solid lines, squares) and English-speaking (dashed lines, circles) groups are charted in Figure 2.

3.2.1 Basic trends

Figure 2a and the raw data in Table A, show that for duration overall, increasing MOD between -4 and 4 was associated with decreases in *baga* responses in both language groups,

consistent with a long-last bias, but there were substantial differences. Changes associated with varied duration were more robust in the English than the Spanish group, as predicted by hypothesis (3a). The proportion of *baga* responses was generally higher when *ga* was longer than *ba* (negative MODs), compared to the control condition, and the proportion of *baga* responses was lower when *ba* was longer than *ga* at MODs 3 and 4, consistent with the Long-Last Principle in (1). At a more granular level, the outcome for English was not fully consistent with the Incremental Long-Last hypothesis in (3b) because the function from MOD to *baga* responses was not fully linear: changes in *baga* responses were not proportional across units of increase on the MOD scale and were not always in the same direction. The negative correlation between MOD and *baga* responses was linear only between MODs -2 and 0, and between MODs 2 and 4. The proportion of *baga* responses plateaued between MODs -4 and -2, and unexpectedly increased at MOD 2. The plot for Spanish hints at a marginal negative trend supporting the Incremental Long-Last hypothesis, especially at the positive end of the scale where differences between levels were larger. The proportion of *baga* responses decreased between MODs -4 and 0, and between MODs 0 and 4. The negative association was for the most part more linear than in the English group in that changes across MOD levels were more proportional and ran in the same direction, apart from increases at MODs 0 and 4. However, changes were generally very small, making it difficult to confidently interpret the Spanish outcome as a meaningful trend.

For intensity, **Figure 2b** and the data in Appendix B, Table A, indicate that introducing an intensity disparity between alternating syllables was strongly associated with a preference for loud-first groupings (fewer *baga* responses at the negative end and more *baga* responses at the positive end of the MOI scale) in both language groups overall; this generally confirmed the Loud-First Principle in (1). However, the function from MOI to *baga* responses was not fully linear in either language group, as predicted by the Incremental Loud-First hypothesis in (2a). Some of the observed alinearity was unpredicted: in the Spanish group there was a decrease in *baga* responses at MOI 4 and in the English group there were small decreases at MOIs -3 and 3. However, the remainder of the alinearity observed was predicted by the Intensity Attenuation hypothesis in (2b), which predicted that the effect of a unit of increase in disparity (3 dB) would be greatest when intensity disparities were smaller. In other words, larger increases in loud-first groupings were expected between MOI 0 and MOIs -1 and 1 (from 0 to 3 dB of disparity), and gains were expected to be increasingly smaller as the disparity was increased to 6, 9, and 12 dB. The S-shaped functions associated with increasing MOI show that this was generally the case in both language groups (with the irregularities noted), although there were again differences. A larger-than-expected increase in the English group was observed at MOI 4, and the increase from MOI -2 to MOI -1 in the Spanish group was larger than might have been expected. Outcomes in the intensity condition therefore largely confirmed hypothesis (2a), within the limits predicted by (2b), which was also broadly confirmed.

The effect of varied intensity was observably stronger than the effect of varied duration in the Spanish group, as predicted by the Intensity Wins hypothesis, (2c). It is not obvious from an inspection of **Figure 2a** that this was true for the English group, where the influence of varied duration was stronger. Finally, the effect of varied intensity was generally weaker in the English group: above MOI -2, the proportion of *baga* responses was lower in the English group up to MOI 3, and was marginally lower at MOI 4 (see Table A, Appendix B).

3.2.2 Statistical analysis

To test for the significance of the observed trends, mixed effects logistic regression models were fit to the response data using the *glmer* function in the *lme4* package (Bates et al., 2015) of the statistical software program R (R Development Core Team, 2013).

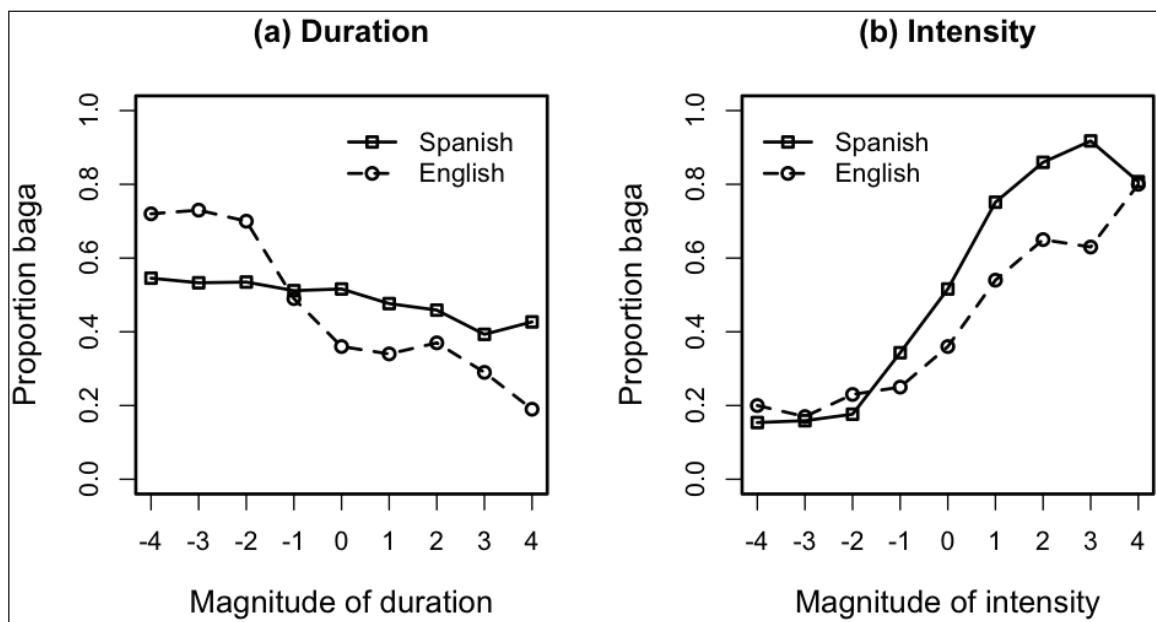


Figure 2: Proportion of *baga* responses in Experiment 1 in the (a) duration condition (as a function of increasing magnitude of duration, or MOD) and (b) intensity condition (as a function of increasing magnitude of intensity, or MOI).

This method estimated the maximum likelihood of a *baga* decision (the arbitrarily chosen, positively coded Response category). Three variables, Subject, Block (3 levels), and Order (2 levels, whether the sequence-initial syllable was *ba* or *ga*), were treated as random effects. To test for the significance of the general trends evident in **Figure 2** (a generally positive correlation between MOI and *baga* responses, and a negative correlation between MOD and *baga* responses, overall), the predictor variables Intensity and Duration were coded as 9-point scales (from -4 to 4). Language was coded as a factor with two levels, and Confidence as a factor with three levels (sure, not sure, and no response).

The output of the model that best fit the response data for Experiment 1 is shown in **Table 2**. This model, which included terms for the intercept, Intensity, Duration, and Language, provided a significantly better fit for the data than the full model that included Confidence ($\chi^2 = 1.016$; $df = 2$; $p = .6018$).⁷

	Coefficient	OR	SE	z	p	
Intercept	-0.255	.77	0.075	-3.424	0.0006	***
Duration	-0.578	.56	0.042	-13.763	<0.00001	***
Intensity	0.694	2.00	0.044	15.656	<0.00001	***
Language (Span)	0.278	1.32	0.057	4.883	<0.00001	***
Language (Span)*Duration	0.444	1.56	0.056	7.975	<0.00001	***
Language (Span)*Intensity	0.361	1.43	0.070	5.198	<0.00001	***

Table 2: Results of best-fitting logistic model predicting *baga* responses in the full data set in Experiment 1 ($n = 6,073$).

Table 2 reveals that the fixed effects of Duration, Intensity, and Language as well as the interactions Language(Span)*Duration and Language(Span)*Intensity were highly

⁷ Best-fitting models were determined using the method of backwards elimination.

significant. Consistent with the functions charted in **Figure 2**, the signs on the estimated coefficients for Duration and Intensity indicate that the proportion of *baga* responses was negatively correlated with MOD, and positively correlated with MOI. The positive sign on the estimated coefficient for Language reveals that the proportion of *baga* responses was higher in the Spanish group overall, and the positive signs on the estimated coefficients for the interactions indicates that this was true in both test conditions.

Given that the interactions were highly significant, additional models were constructed so that the language-specific trends associated with Intensity and Duration could be better understood. The outputs of the best-fitting models for Spanish and English are shown in **Table 3**.

(a) Spanish (Observations: $n = 3,055$)

	Coefficient	OR	SE	z	p
Intercept	0.021	1.02	0.074	0.279	0.7801
Duration	-0.135	.87	0.037	-3.636	0.0003 ***
Intensity	1.07	2.92	0.054	19.721	<0.00001 ***

(b) English (Observations: $n = 3,018$)

	Coefficient	OR	SE	z	p
Intercept	-0.257	.77	0.083	-3.115	0.0018 **
Duration	-0.580	.56	0.042	-13.768	<0.00001 ***
Intensity	0.696	2.01	0.045	15.656	<0.00001 ***

Table 3: Results of logistic models predicting *baga* responses for the Spanish and English groups in Experiment 1. Duration and Intensity were treated as scales.

3.2.2.1 Duration

The values associated with the Duration terms in **Table 3** indicate that the general tendency for *baga* responses to decrease as a function of increasing MOD was highly significant in both language groups (Spanish, $p = 0.0003$; English, $p < 0.00001$). The significant finding for the Spanish group means that the small size of changes between levels was offset by the fact that deviations from the negative correlation were not substantial. As noted, however, this does not represent a robust trend. The odds ratio (OR; the exponent of the estimated coefficient) provides an estimate of probability and as such, may be interpreted as a measure of effect size (Hosmer & Lemeshow, 2004, 47). Used in this way, the ORs associated with Duration in **Table 3** confirm the observation that the negative correlation associated with Duration was stronger in the English than in the Spanish group: on average, unit increases in DURATION reduced the odds of a *baga* response by 13% in the Spanish group (OR = .87) and by 44% in the English group (OR = .56). As noted, the values associated with the interaction term Language (Span)*Duration in **Table 2** revealed this difference to be highly significant ($p < 0.00001$).

To reveal which differences between individual MOD levels were significant, pairwise comparisons were produced using the *lsmeans* function in R (Tukey method with a .95 confidence interval). These comparisons provided more specific information about how listeners' grouping decisions were influenced by incremental increases in the duration disparity between syllables. The meaningful comparisons are between MOD levels at the same end of the scale (MODs 0 to -4 and MODs 0 to 4), because differences in *baga* responses at each end relate to differences in the proportion of responses representing long-last groupings when a different syllable is longer (*ba* at the positive and *ga* at the

negative end). A chart of outcomes for these meaningful comparisons is provided in **Table 4**. The second column indicates the number of steps on the MOD scale that separate the members of each comparison pair. An increase of one step represents an increase of 30 ms in the duration disparity.

In the Spanish group, only the comparison MOD 0/MOD 3 was significant, a three-step difference (**Table 4.a.vi**), comparing the no-difference control sequence with the sequence in which *ba* was 90 ms longer than *ga*. In the English group, increases of 4 steps (120 ms) significantly increased long-last groupings, as did three-step increases in all cases but one (see comparisons in **Table 4b.iii**, **4.b.v**, and **4.b.vi**). A two-step increase produced significantly more long-last groupings in comparisons **4.b.ix** and **4.b.xi**. The comparison in **Table 4.b.viii** (a two-step increase) was also significant, but due to the change in direction at MOD 2, it is not clear that this difference is consistent with the overall trend. A one-step increase made a significant difference only in comparison **4.b.xvii**.

3.2.2.2 Intensity

The fixed effect of Intensity was highly significant in both language groups, confirming the positive trend in *baga* responses as a function of increasing MOI. That is, increasing the intensity disparity increased responses representing loud-first groupings overall (fewer *baga* responses at the negative and more *baga* responses at the positive end of the MOI scale). The magnitude of the intensity effect was greater in the Spanish group: the ORs associated with Intensity in **Table 3** indicate that on average, a *baga* response was twice as likely per unit increase in MOI in the English group (OR = 2.01), and almost three times as likely per unit increase in the Spanish group (OR = 2.92). The values associated with the term Language (Span)*Duration in **Table 2** indicate that this difference was highly significant ($p < 0.00001$).

Posthoc pairwise comparisons were again made to determine which differences between individual MOI levels were significant. The results of these comparisons are shown in **Table 5**. In the Spanish group, all comparisons with MOI 0 were significant, indicating that any difference in intensity significantly increased loud-first groupings relative to the control condition. A three-step increase in the intensity disparity (9 dB) produced significantly more loud-first groupings in all cases but one: the comparison in **Table 5.a.iv** was not significant, likely due to the trend reversal at MOI 4 (see **Figure 2b**). Differences of one and two steps (3 or 6 dB) significantly increased loud-first groupings at the center of the scale. All comparisons between levels separated by two steps were significant except for the comparisons in **Table 5.a.vii** and **a.viii**. Apart from the comparisons with MOI 0, a one-step increase in disparity made a significant difference only in comparison **5.a.xvii**. In the English group, all comparisons with MOI 0 were significant except for the comparison in **Table 5.b.xx** (a one-step difference). Otherwise, a one-step increase significantly increased loud-first groupings only in comparison **b.xiv**. Given the trend reversal at MOI 3, this difference may not be consistent with the overall trend. A two-unit difference significantly increased loud-first groupings at the center of the scale (comparisons **b.ix–xii**), but not at the edges (see comparisons in **b.vii** and **b.viii**). All comparisons between levels separated by three steps were significant except for **b.iv**.

The theoretical question of central interest was whether language-specific differences in the association between prosodic duration and constituent-final positions would predict differences in listeners' segmentations of duration-varied sequences. In this regard, the interesting finding of Experiment 1 was that the influence of varied duration on grouping behaviour was in the expected direction (a long-last outcome) and was substantially greater in the English group than in the Spanish group. A more extended discussion

(a) Spanish

Comparison	Diff. (in MOD units)	Coef.	SE	z	p
i. MOD -4 – MOD 0	4	0.100	0.165	0.608	0.9996
ii. MOD 0 – MOD 4	4	0.374	0.163	2.291	0.3476
iii. MOD -4 – MOD -1	3	0.131	0.220	0.591	0.9996
iv. MOD 1 – MOD 4	3	0.203	0.220	0.923	0.9917
v. MOD -3 – MOD 0	3	0.053	0.164	0.322	1.0
vi. MOD 0 – MOD 3	3	0.520	0.164	3.162	0.0417 *
vii. MOD -4 – MOD -2	2	0.037	0.220	0.167	1.0
viii. MOD 2 – MOD 4	2	0.129	0.220	0.587	0.9997
ix. MOD -3 – MOD -1	2	0.083	0.219	0.381	1.0
x. MOD 1 – MOD 3	2	0.348	0.220	1.580	0.8160
xi. MOD -2 – MOD 0	2	0.064	0.162	0.393	1.0
xii. MOD 0 – MOD 2	2	0.245	0.163	1.507	0.8529
xiii. MOD -4 – MOD -3	1	0.048	0.222	0.214	1.0
xiv. MOD 3 – MOD 4	1	-0.145	0.221	-0.658	0.9992
xv. MOD -3 – MOD -2	1	-0.011	0.219	-0.049	1.0
xvi. MOD 2 – MOD 3	1	0.274	0.221	1.244	0.9466
xvii. MOD -2 – MOD -1	1	0.094	0.218	0.433	1.0
xviii. MOD 1 – MOD 2	1	0.074	0.219	0.336	1.0
xix. MOD -1 – MOD 0	1	-0.031	0.161	-0.190	1.0
xx. MOD 0 – MOD 1	1	0.171	0.162	1.056	0.9802

(b) English

Comparison	Diff. (in MOD units)	Coef.	SE	z	p
i. MOD -4 – MOD 0	4	1.271	0.179	7.080	<.0001 ***
ii. MOD 0 – MOD 4	4	1.130	0.203	5.560	<.0001 ***
iii. MOD -4 – MOD -1	3	1.000	0.232	4.314	0.0005 ***
iv. MOD 1 – MOD 4	3	0.751	0.256	2.927	0.0823 .
v. MOD -3 – MOD 0	3	1.331	0.182	7.324	<.0001 ***
vi. MOD 0 – MOD 3	3	0.575	0.177	3.240	0.0327 *
vii. MOD -4 – MOD -2	2	0.088	0.242	0.362	1.0
viii. MOD 2 – MOD 4	2	0.909	0.254	3.576	0.0105 *
ix. MOD -3 – MOD -1	2	1.060	0.234	4.540	0.0002 ***
x. MOD 1 – MOD 3	2	0.195	0.237	0.826	0.9961
xi. MOD -2 – MOD 0	2	1.183	0.176	6.705	<.0001 ***
xii. MOD 0 – MOD 2	2	0.221	0.168	1.318	0.9262
xiii. MOD -4 – MOD -3	1	-0.060	0.245	-0.245	1.0
xiv. MOD 3 – MOD 4	1	0.555	0.260	2.133	0.4506
xv. MOD -3 – MOD -2	1	0.148	0.243	0.607	0.9996
xvi. MOD 2 – MOD 3	1	0.353	0.234	1.510	0.8513
xvii. MOD -2 – MOD -1	1	0.912	0.229	3.976	0.0023 **
xviii. MOD 1 – MOD 2	1	-0.158	0.230	-0.688	0.9989
xix. MOD -1 – MOD 0	1	0.271	0.163	1.664	0.7685
xxi. MOD 0 – MOD 1	1	0.379	0.172	2.211	0.3982

Table 4: Pairwise comparisons between MOD levels by language group in Experiment 1.

(a) Spanish

Comparison	Diff. (in MOI units)	Coef.	SE	z	p	
i. MoI -4 – MoI 0	4	-1.693	0.220	-7.688	<.0001	***
ii. MoI 0 – MoI 4	4	-1.502	0.205	-7.336	<.0001	***
iii. MoI -4 – MoI -1	3	-1.063	0.270	-3.938	0.0027	**
iv. MoI 1 – MoI 4	3	-0.337	0.268	-1.257	0.9436	
v. MoI -3 – MoI 0	3	-1.664	0.217	-7.670	<.0001	***
vi. MoI 0 – MoI 3	3	-2.492	0.285	-8.749	<.0001	***
vii. MoI -4 – MoI -2	2	-0.152	0.295	-0.514	0.9999	
viii. MoI 2 – MoI 4	2	0.383	0.298	1.286	0.9356	
ix. MoI -3 – MoI -1	2	-1.034	0.267	-3.869	0.0035	**
x. MoI 1 – MoI 3	2	-1.328	0.333	-3.985	0.0022	**
xi. MoI -2 – MoI 0	2	-1.541	0.209	-7.387	<.0001	***
xii. MoI 0 – MoI 2	2	-1.884	0.227	-8.288	<.0001	***
xiii. MoI -4 – MoI -3	1	-0.029	0.301	-0.096	1.0	
xiv. MoI 3 – MoI 4	1	0.991	0.343	2.885	0.0924	
xv. MoI -3 – MoI -2	1	-0.123	0.293	-0.420	1.0	
xvi. MoI 2 – MoI 3	1	-0.608	0.357	-1.701	0.7461	
xvii. MoI -2 – MoI -1	1	-1.911	0.261	-3.498	0.0139	*
xviii. MoI 1 – MoI 2	1	-0.719	0.286	-2.520	0.2223	
xix. MoI -1 – MoI 0	1	-0.630	0.171	-3.682	0.0071	**
xx. MoI 0 – MoI 1	1	-1.165	0.187	-6.245	<.0001	***

(b) English

Comparison	Diff. (in MOI units)	Coef.	SE	z	p	
i. MoI -4 – MoI 0	4	-1.195	0.199	-6.015	<.0001	***
ii. MoI 0 – MoI 4	4	-1.596	0.201	-7.947	<.0001	***
iii. MoI -4 – MoI -1	3	-0.242	0.264	-0.919	0.9920	
iv. MoI 1 – MoI 4	3	-1.250	0.249	-5.016	<.0001	***
v. MoI -3 – MoI 0	3	-1.391	0.211	-6.606	<.0001	***
vi. MoI 0 – MoI 3	3	-0.736	0.169	-4.365	0.0004	***
vii. MoI -4 – MoI -2	2	-0.135	0.267	-0.506	0.9999	
viii. MoI 2 – MoI 4	2	-0.773	0.254	-3.047	0.0587	*
ix. MoI -3 – MoI -1	2	-0.438	0.273	-1.606	0.8018	
x. MoI 1 – MoI 3	2	-0.390	0.224	-1.739	0.7220	
xi. MoI -2 – MoI 0	2	-1.060	0.191	-5.541	<.0001	***
xii. MoI 0 – MoI 2	2	-0.823	0.170	-4.837	<.0001	***
xiii. MoI -4 – MoI -3	1	0.196	0.281	0.697	0.9988	
xiv. MoI 3 – MoI 4	1	-0.860	0.253	-3.403	0.0192	*
xv. MoI -3 – MoI -2	1	-0.331	0.276	-1.199	0.9569	
xvi. MoI 2 – MoI 3	1	0.087	0.229	0.379	1.0	
xvii. MoI -2 – MoI -1	1	-0.107	-0.258	-0.416	1.0	
xviii. MoI 1 – MoI 2	1	-0.477	0.225	-2.117	0.4617	
xix. MoI -1 – MoI 0	1	-0.953	0.187	-5.098	<.0001	***
xx. MoI 0 – MoI 1	1	-0.347	0.163	-2.125	0.4557	

Table 5: Pairwise comparisons between MOI levels by language group in Experiment 1.

of this and other findings of Experiment 1 is postponed until Section 5, where the outcomes of the two experiments are compared and contrasted. Experiment 1 was conducted as a baseline test of the ITL in preparation for Experiment 2. Testing listeners' responses to singly-varied sequences alone was important because prior studies have found listeners to be sensitive to the context in which sequences were presented. Listeners tested by Hay and Diehl (2007) responded at the level of chance when they were tasked with grouping a no-difference control sequence presented in the context of intensity-varied sequences. However, when the same control sequence was presented together with duration-varied sequences, listeners made more iambic responses. In their studies with Zapotec speakers, Crowhurst and Teodocio (2014) report that singly-varied sequences were associated with clear grouping preferences when the stimulus set included both singly and co-varied sequences, but not when only singly-varied sequences were used.

4 Experiment 2

Experiment 2 explored the preferences of new groups of Spanish- and English-speakers, who segmented sequences in which intensity and duration were varied orthogonally, as well as singly-varied sequences. Based on Experiment 1 outcomes, Experiment 2 was expected to confirm the Incremental Loud-Soft and Attenuated Intensity hypotheses in both language groups, and the Incremental Long-Last hypothesis in the English group. The Intensity Wins hypothesis was expected to hold for Spanish, but given Experiment 1 outcomes, it was not clear whether this would be the case for English. It was anticipated that the influence of varied intensity would again be more robust in the Spanish than the English group. As the influence of varied duration in the Spanish group was so small in Experiment 1, there was no special expectation as to whether outcomes in Experiment 2 would support the Incremental Long-Last or the Incremental Long-Short hypothesis.

4.1 Method

4.1.1 Stimuli

The design for Experiment 2 included five conditions: a no-difference control condition and two singly-varied conditions, as in Experiment 1, and two new co-varied conditions, in which sequences were characterized by variations in both vowel intensity and duration. Because Experiment 2 included more conditions, a 5-point scale was used on which magnitude levels -4, -2, 0, 2, and 4 from Experiment 1 were represented. That is, only two manipulation levels were used in test conditions: fixed intensity disparities of 6 dB and 12 dB, and duration disparities of 60 and 120 ms. Accordingly, the control sequences and singly-varied sequences prepared for Experiment 1 that represented the manipulation levels for Experiment 2 were used again.

Following Crowhurst and Teodocio (2014), new sequences were prepared for two co-varied conditions using the materials and mechanical procedures described in Section 3.1.1 for Experiment 1. In a Co-operating condition, a long, soft syllable (*baa*, *gaa*) was alternated with a short, loud complement (*GA*, *BA*). In this condition, one type of grouping (*BA-gaa* or *GA-baa*) was doubly consistent with the ITL. It was expected that intensity and duration cues would work together in the Co-operating condition and that the outcome would be the ITL-predicted one. In a Competing condition, a loud, long syllable (*BAA* or *GAA*) was combined with a soft, short complement (*ba* or *ga*). In this condition, any pairing (e.g., *GAA-ba* or *ba-GAA*) would satisfy either the Loud-First or the Long-Last Principle of the ITL, but not the other. Intensity and duration were manipulated at the same levels in the co-varied as in the singly-varied conditions. The design of the stimulus

set is shown in **Table 6**. The fully counterbalanced stimulus set for Experiment 2 included 50 distinct sequences.

↓ MOD	Co-operating		Duration	Competing		
2	<i>Ba</i> 230 / 58	<i>Ba</i> 230 / 64	<i>Ba</i> 230 / 70	<i>Ba</i> 230 / 70	<i>Ba</i> 230 / 70	
	<i>Ga</i> 110 / 70	<i>Ga</i> 110 / 70	<i>Ga</i> 110 / 70	<i>Ga</i> 110 / 64	<i>Ga</i> 110 / 58	
1	<i>Ba</i> 230 / 58	<i>Ba</i> 230 / 64	<i>Ba</i> 230 / 70	<i>Ba</i> 230 / 70	<i>Ba</i> 230 / 70	
	<i>Ga</i> 170 / 70	<i>Ga</i> 170 / 70	<i>Ga</i> 170 / 70	<i>Ga</i> 170 / 64	<i>Ga</i> 170 / 58	
0	<i>Ba</i> 230 / 58	<i>Ba</i> 230 / 64	<i>Ba</i> 230 / 70	<i>Ba</i> 230 / 70	<i>Ba</i> 230 / 70	Intensity
	<i>Ga</i> 230 / 70	<i>Ga</i> 230 / 70	<i>Ga</i> 230 / 70	<i>Ga</i> 230 / 64	<i>Ga</i> 230 / 58	
-1	<i>Ba</i> 170 / 58	<i>Ba</i> 170 / 64	<i>Ba</i> 170 / 70	<i>Ba</i> 170 / 70	<i>Ba</i> 170 / 70	
	<i>Ga</i> 230 / 70	<i>Ga</i> 230 / 70	<i>Ga</i> 230 / 70	<i>Ga</i> 230 / 64	<i>Ga</i> 230 / 58	
-2	<i>Ba</i> 110 / 58	<i>Ba</i> 110 / 64	<i>Ba</i> 110 / 70	<i>Ba</i> 110 / 70	<i>Ba</i> 110 / 70	
	<i>Ga</i> 230 / 70	<i>Ga</i> 230 / 70	<i>Ga</i> 230 / 70	<i>Ga</i> 230 / 64	<i>Ga</i> 230 / 58	
MOI →	-2	-1	0	1	2	
	Competing			Co-operating		

Table 6: Design of the stimulus set for Experiment 2. Values for vowel duration and intensity (ms/dB) for each syllable pair, by manipulation levels.

4.1.2 Participants

The participants were 20 new native English speakers and 28 new native Spanish speakers from the populations sampled in Experiment 1. One additional English-speaking participant was removed from the data set because it was clear from the response booklet that he had not followed instructions. Screening and payment procedures were as described for Experiment 1.

4.1.3 Experimental procedures

All procedures for Experiment 2 were as described for Experiment 1, with the following exceptions. Given that the stimulus set was larger, each of three test blocks consisted of 52 sequences (the 48 test sequences presented once and the two controls presented twice) with a pause at the midpoint of each block. The response sheets provided to study participants were adapted accordingly. In Experiment 2, there were two confidence categories for English-speaking participants (“a” for “very sure”, “b” for “reasonably sure”, and “c” for “guessing”). On the advice of the Spanish-speaking consultant, only two confidence categories (“sure” and “not sure”) were provided on Spanish response sheets.⁸ Sessions lasted less than an hour, including time for informed consent procedures and payment after the test.

4.2 Results

The raw data obtained in Experiment 2 are presented in Appendix B, Table B. The design of the stimulus set provided for 7,332 potential observations (2,964 for the English group and 4,368 for the Spanish group). Of this maximum, there were 126 missing responses

⁸ The revision to the scale for the English speakers was intended to give them a finer-grained range of responses. This was not done for the Mexican participants, as the Obregon consultant’s opinion was that these students were less accustomed to providing a confidence rating. The statistical analysis indicated that the variable Confidence did not contribute significantly to the model for either group.

(1.75% of the total possible), so that a total of 7,206 data points were included in the analysis (2,937 and 4,269 for English and Spanish, respectively).

4.2.1 Basic trends

Trends observed in the singly-varied conditions are plotted in **Figure 3**. **Figure 3a** shows that in the English group (dashed lines, circles), MOD and *baga* responses were again negatively correlated, consistent with a long-last preference. The correlation was more consistently linear than in Experiment 1, confirming the Incremental Long-Last hypothesis, (3b). However, the trend was also weaker in Experiment 2. In the Spanish group (solid lines, squares), increasing MOD had no clear effect on listeners' grouping choices, apart from an increase at MOD 2. That the influence of varied duration was stronger in the English than the Spanish group was again consistent with hypothesis (3a).

Plots for the intensity-varied condition appear in **Figure 3b**. Apart from a decrease at MOI 2 in the Spanish group, MOI was again positively correlated with *baga* responses in both languages, consistent with a preference for loud-first groupings. The functions were S-shaped, indicating that the greatest effects of varied intensity were again clustered at the centre of the scale in both language groups, as predicted by the Intensity Attenuation hypothesis, (2b). The functions for MOI in both language groups were more regular than in Experiment 1 and with the exception noted, outcomes were consistent with the Incremental Loud-First hypothesis (2a), within the limits predicted by Intensity Wins, (2c). The effect associated with Intensity was again stronger in the Spanish group, as predicted based on the outcome of Experiment 1: the proportion of *baga* groupings was lower at MOIs -2 and -1, and higher at MOIs 1 and 2, than in the English group. In sum, outcomes for the singly-varied conditions in Experiment 2 on the whole replicated the main tendencies observed in Experiment 1; however, effects associated with varied duration were smaller in Experiment 2 compared with the first study (absent in the Spanish and weaker in the English group).

Trends in the co-varied conditions are plotted in **Figures 4** and **5**. In **Figure 4**, the response data for the co-varied conditions is arranged to show the effect of manipulating

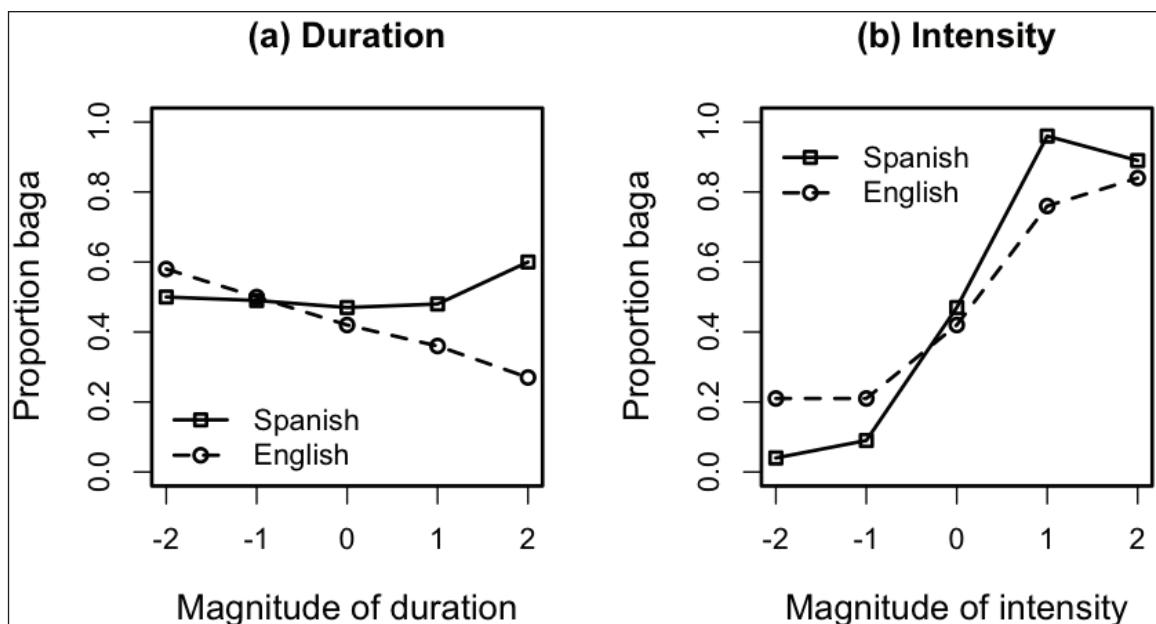


Figure 3: Proportion of *baga* responses as a function of (a) MOD and (b) MOI in conditions where intensity and duration were varied singly in Experiment 2.

Duration on outcomes for the more robust predictor, Intensity. Black dotted lines represent the baseline trends in the singly-varied intensity condition. MOI is represented on the x-axis, and MOD in the vertical dimension by the stacked circles at MOI 0. Blue lines plot trends associated with varied intensity when *ga* was longer than *ba* (positive MODs), and red lines plot outcomes when *ba* was longer than *ga* (negative MODs). Dashed lines represent a 60 ms, and solid lines a 120 ms disparity. **Figures 4a and 4b** show that

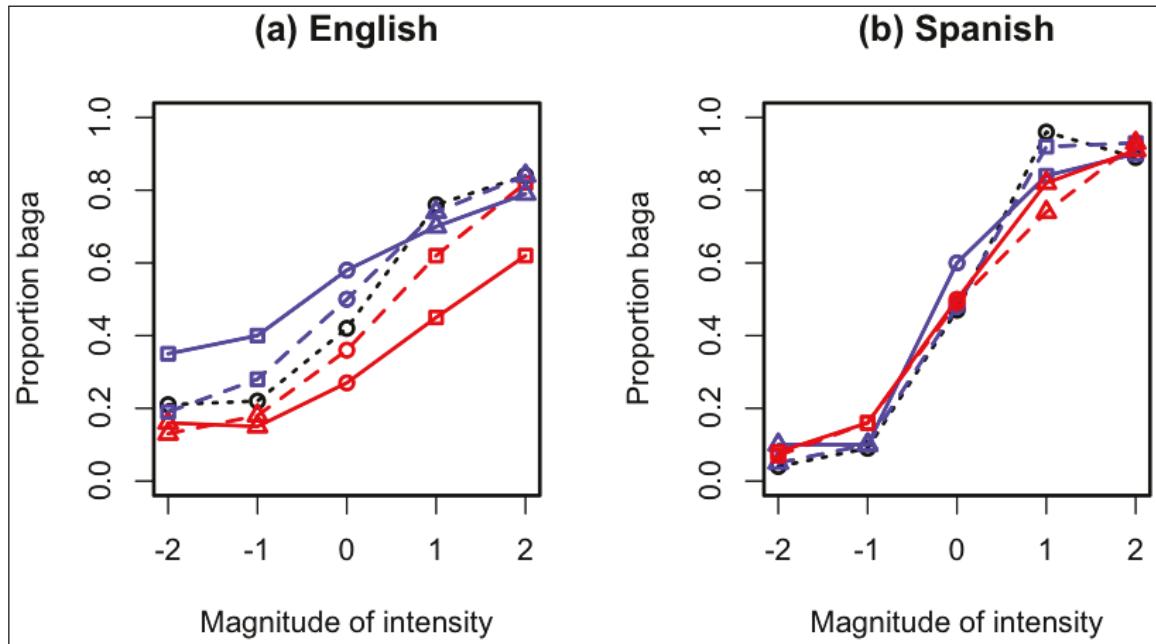


Figure 4: Proportion of *baga* responses as a function of increasing MOI (x axis) by MOD (vertical) in Experiment 2. Lines: dotted = baseline intensity (MOI 0); dashed = 6 dB disparity; solid = 12 dB disparity; red = positive MODs (*ba* is longer); blue = negative MODs (*ga* is longer).

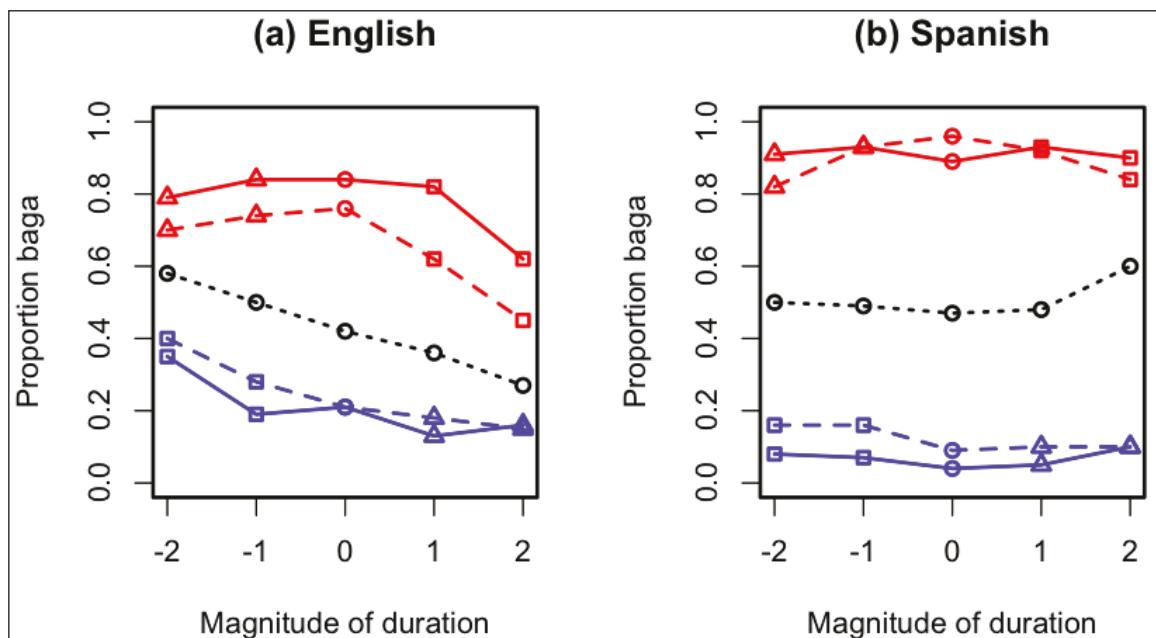


Figure 5: Proportion of *baga* responses as a function of increasing MOD (x axis) by MOI (vertical) in Experiment 2. Lines: dotted = baseline intensity (MOD 0); dashed = 60 ms disparity; solid = 120 ms disparity; red = positive MOIs (*ba* is louder); blue = negative MOIs (*ga* is louder).

the dominant effect in both co-varied conditions was one of intensity: increases in MOI increased *baga* responses regardless of whether a duration disparity was also present. The interaction between intensity and duration is seen most clearly in **Figure 4a**, the graph for English, because differences in the baseline duration condition allowed the plotted lines to be well differentiated. Here, we see that introducing a duration disparity weakened the correlation between MOI and *baga* responses; this is revealed in the flatter slopes of the intensity functions.

The Competing condition is represented by points labeled with squares (the blue lines at negative MOIs and the red lines at positive MOIs), and the Co-operating condition by the points labeled with triangles (the blue lines at positive MOIs and the red lines at negative MOIs). In **Figure 4a**, we see that in the English group, the lines representing the 60 and 120 ms disparities diverge in the Competing condition, and converge in the Co-operating condition. This indicates that in the Competing condition, as a syllable got longer, there were fewer loud-first (long-short) and more long-last groupings (more *baga* responses when *ga* was longer, fewer *baga* responses when *ba* was longer) as the same syllable got louder. In the Co-operating condition, by contrast, lengthening the softer syllable produced more long-short groupings, regardless of how loud the short syllable was. An explanation for this finding is suggested in the general discussion.

The information presented in **Figure 4** is rearranged in **Figure 5** to show the effect of Intensity on Duration. In the English group, we see that effects of duration were most robust in the co-varied conditions when *ba* was louder than *ga* (the red lines), and when *ga* was 12 dB louder than *ba* (blue, leaving aside the point representing MOD 0). In the Spanish group, the influence of duration in the co-varied conditions was clear only when *ba* was 6 dB louder than *ga* (dashed red line). Otherwise, varied duration had no consistent influence on the Spanish speakers' grouping choices.

4.2.2 Statistical analysis

Mixed effects logistic regression models were fit to the response data set using the *glmer* function in R and as before, the dependent variable measured *baga* responses. The variables Subject, Block, and Order (coded as for Experiment 1) were again treated as random effects. The predictor variables Intensity and Duration were coded as 5-point scales (from -2 to 2). Confidence was treated as a factor with three levels (sure, not sure, no response), as described for Experiment 1. For this purpose, "sure" and "reasonably sure" responses in the English group were conflated into a single category. The output of the most parsimonious model that best fit the full data set is shown in **Table 7**. Confidence did not contribute significantly to goodness of fit ($\chi^2 = 1.465, df = 2, p = 0.48$) and was therefore not included in subsequent analyses. (The output of the full model which included Confidence is presented in Table D in Appendix C.)

Table 7 reveals that fixed effects of Intensity and Duration were highly significant in the data set overall. As in Experiment 1, Intensity was positively and Duration negatively correlated with *baga* responses. The values associated with the Language term and with the interactions Intensity*Language and Duration*Language in **Table 7** indicate that there were significant differences between the English and Spanish groups, as there had been in Experiment 1. No significance was associated with the interactions Duration*Intensity and Duration*Intensity* Language in the overall analysis.

As the preliminary analysis indicated that significant language-specific differences were associated with varied intensity and duration, additional models were constructed in which the response data for the two language groups were analysed separately. The outputs of the best fitting models are shown in **Table 8**.

(n = 7,206)	Coefficient	OR	SE	z	p	
Intercept	-0.240	.79	0.068	-3.540	0.0004	***
Duration	-0.408	.66	0.043	-9.425	<0.00001	***
Intensity	1.037	2.82	0.047	22.209	<0.00001	***
Language (Span)	0.247	1.28	0.064	3.829	0.0001	***
Duration*Intensity	0.050	1.05	0.046	1.100	0.2713	
Duration*Language (Span)	0.424	1.53	0.061	6.937	<0.00001	***
Intensity*Language (Span)	1.102	3.01	0.077	14.362	<0.00001	***
Duration*Intensity *Language (Span)	-0.010	.99	0.075	-0.131	0.8960	

Table 7: Output of logistic model predicting *baga* responses in the data overall in Experiment 2.

(a) Spanish						
(n = 4,269)	Coefficient	OR	SE	z	p	
Intercept	0.006	1.00	0.072	0.08	0.936	
Duration	0.017	1.01	0.043	0.38	0.701	
Intensity	2.153	8.61	0.062	34.66	<0.00001	***
Duration*Intensity	0.042	1.04	0.060	0.70	0.485	

(b) English						
(n = 2,937)	Coefficient	OR	SE	z	p	
Intercept	-0.209	.81	0.090	-2.335	0.0195	*
Duration	-0.411	.66	0.043	-9.450	<0.00001	***
Intensity	1.044	2.84	0.047	22.172	<0.00001	***
Duration*Intensity	0.050	1.05	0.046	1.098	0.272	

Table 8: Output of logistic model predicting *baga* responses in Experiment 2 by language.

4.2.2.1 Results for duration and intensity varied singly

Table 8 reveals that the fixed effect of Intensity was highly significant in both language groups. As expected based on the trends evident in **Figure 3**, the fixed effect of Duration was highly significant in English and was not significant in the Spanish group. The values associated with Duration*Intensity suggest that there were no significant interactions between these variables. This outcome does not tell the full story, and a fuller discussion of interactions between intensity and duration is undertaken in Section 4.2.2.2.

As before, pairwise comparisons in which Duration and Intensity were treated as factors were made using the procedure described for Experiment 1. The outcomes of these tests appear in **Tables 9** and **10**. **Table 9** indicates that there were no significant differences between manipulation levels at the same end of the MOD scale (i.e., from MOD 0 to 2 and from MOD -2 to 0, **Table 9.a.ii** and **9.a.i**) in the Spanish group, as expected, or in the English group, in spite of the significant trend revealed by **Table 8b** and charted in **Figure 3a**. For Intensity, **Table 10** reveals that all comparisons with MOI 0 (**Table 10.a.i-iv** and **Table 10.b.i-iv**) were highly significant in both language groups. This means that in comparison with the control sequence, MOI 0, increasing the intensity disparity to 6 and 12 dB significantly increased responses representing loud-first groupings. The differences between MOI -2/MOI -1 and MOI 1/MOI 2 were not significant in either group

(**Table 10.a.v–vi** and **Table 10.b.v–vi**). The latter result, along with the significant differences between MOI -1/MOI 0 and MOI 0/MOI 1 confirm the observation that the greatest effects of Intensity were again clustered in the centre of the magnitude scale.

Comparison	Diff. (in MOD units)	Coef.	SE	z	p
(a) Spanish					
i. MOD -2 / MOD 0	2	0.108	0.193	0.560	1.0
ii. MOD 0 / MOD 2	2	-0.519	0.195	-2.662	0.5675
iii. MOD -1 / MOD 0	1	0.059	0.193	0.303	1.0
iv. MOD 0 / MOD 1	1	-0.046	0.193	-0.239	1.0
v. MOD -2 / MOD -1	1	0.049	0.223	0.222	1.0
vi. MOD 1 / MOD 2	1	-0.473	0.225	-2.103	0.9240
(b) English					
i. MOD -2 / MOD 0	2	0.700	0.237	2.957	0.3384
ii. MOD 0 / MOD 2	2	0.689	0.255	2.706	0.5320
iii. MOD -1 / MOD 0	1	0.333	0.234	1.421	0.9996
iv. MOD 0 / MOD 1	1	0.248	0.242	1.025	1.0
v. MOD -2 / MOD -1	1	0.367	0.271	1.352	1.0
vi. MOD 1 / MOD 2	1	0.441	0.294	1.503	0.9990

Table 9: Pairwise comparisons in the singly-varied duration condition (MOD levels at MOI 0) by language group in Experiment 2.

Comparison	Diff. (in MOI units)	Coef.	SE	z	p
(a) Spanish					
i. Moi -2 / Moi 0	2	-3.066	0.404	-7.582	<0.0001 ***
ii. Moi 0 / Moi 2	2	-2.215	0.275	-8.045	<0.0001 ***
iii. Moi -1 / Moi 0	1	-2.239	0.294	-7.606	<0.0001 ***
iv. Moi 0 / Moi 1	1	-3.425	0.431	-7.937	<0.0001 ***
v. Moi -2 / Moi -1	1	-0.827	0.474	-1.743	0.9913
vi. Moi 1 / Moi 2	1	1.209	0.487	2.485	0.7073
(b) English					
i. Moi -2 / Moi 0	2	-0.999	0.269	-3.714	0.0413 *
ii. Moi 0 / Moi 2	2	-2.051	0.293	-6.995	<0.0001 ***
iii. Moi -1 / Moi 0	1	-0.981	0.274	-3.582	0.0642 *
iv. Moi 0 / Moi 1	1	-1.536	0.262	-5.872	<0.0001 ***
v. Moi -2 / Moi -1	1	-0.018	0.331	-0.054	1.0
vi. Moi 1 / Moi 2	1	-0.515	0.342	-1.506	0.999

Table 10: Pairwise comparisons between singly-varied intensity condition (MOI levels at MOD 0) by language group in Experiment 2.

4.2.2.2 Interactions between varied intensity and duration in the co-varied conditions

The outputs of the models shown in **Tables 7** and **8** attached no significance to the interactions Duration*Intensity and Duration*Intensity*Language; however, the trends charted in **Figures 4** and **5** clearly suggest interactions between varied intensity and

duration. Reprising, each line in **Figures 4a** and **4b** represents the effect of varying intensity (on the x-axis) at a different MOD level. What is needed is an assessment of the extent to which changes in the effect size for Intensity are produced by increasing the duration disparity, compared with the baseline condition in which duration did not vary. For this analysis, the data for MODs –2 and 2 and the data for MODs –1 and 1 were conflated. In the logistic model, Duration was coded as a new predictor variable, Duration.Disp, a scale with 3 levels representing disparities of 0, 60, and 120 ms, regardless of which alternating syllable was longer. Table C in Appendix B presents cross-tabulations for the response data, reorganized in this way.

Table 11 presents the outputs of the best-fitting models for English and for Spanish. As before, the fixed effect of Intensity was highly significant in both language groups. There was no significant fixed effect of Duration.Disp in either language group, a finding that appears to conflict with the significant outcome (in the English group, at least) when Duration was coded as a 5-point scale (see **Table 8b**). The reason for this finding is that in conflating disparity levels, summing *baga* and *gaba* responses over MODs –1 and 1 and over MODs –2 and 2 at MOI 0 (the baseline duration category) obscures the effect of the disparity when no intensity disparity is present. (This can be verified by performing these summations with the data in Table B, Appendix B.) Summation produces a similar effect when MODs –1 and 1 and MODs –2 and 2 are conflated. Importantly, however, it can be seen in **Figure 3** that the slopes of the functions from MOI to *baga* responses are unaffected by this procedure. What the analysis tests for are changes in the slopes of functions associated with intensity as the duration disparity increases from 0 to 60 ms and 120 ms of disparity.

The values associated with Duration.Disp*Intensity in **Table 11** reveal that the interaction was significant in both language groups. This meant that when *ga* was longer or shorter than *ba*, increasing MOI produced fewer *baga* responses than when there was no duration disparity, and increasing the duration disparity increased its effect on *baga* responses. The magnitude of the effect of Duration.Disp on Intensity was roughly equivalent in both language groups. The odds ratios for the interactions in **Table 11** reveal that the effect of Intensity on *baga* responses was 20% less per unit increase in Duration.Disp in the Spanish group (OR = .80), and 21% less in the English group (OR = .79), in comparison with the baseline intensity condition (MOD 0).

(a) Spanish					
(n = 4,269)	Coefficient	OR	SE	z	p
Intercept	0.005	1.0	0.072	0.070	0.9408
Duration.Disp	0.030	1.03	0.042	0.710	0.4790
Intensity	2.186	8.89	0.064	34.100	<0.00001 ***
Duration.Disp*Intensity	-0.218	.80	0.066	-3.320	0.00090 ***

(b) English					
(n = 2,937)	Coefficient	OR	SE	z	p
Intercept	-0.193	.82	0.087	-2.226	0.026 *
Duration.Disp	-0.048	.95	0.042	-1.135	0.256
Intensity	1.036	2.81	0.047	22.036	<0.00001 ***
Duration.Disp*Intensity	-0.227	.79	0.049	-4.667	<0.00001 ***

Table 11: Outputs of the best-fitting logistic models testing for interactions between varied intensity and the size of the duration disparity.

The results of pairwise comparisons which tested for significant differences between individual MOI levels in the co-varied conditions appear in **Table 12**. Recall that in the baseline (singly-varied) intensity condition, increasing the intensity disparity from 0 to 6 dB and 12 dB significantly increased responses representing loud-first groupings. The statistics in **Table 12a** reveal the same pattern of highly significant results when a duration disparity (whether 60 or 120 ms) was also present in the Spanish group. **Table 12b** reveals that in the English group, there were fewer significant comparisons. When *ba* was 120 ms longer than *ga*, a 12 dB increase in the intensity disparity (from 0 to 12 dB) made a significant difference in comparison **12.b.xxii**. (Comparison **12.b.i** approached significance.) When there was a duration disparity of 60 ms, regardless of which syllable was louder, comparisons between categories representing an increase in the intensity disparity from 0 to 12 dB were significant (comparison **12.b.xiii**) or highly significant (comparisons **12.b.vii**, **12.b.x**, and

(a) Spanish

Comparison (MOI, MOD/MOI, MOD)	Diff. (in MOI units)	Coef.	SE	z	p
MOI AT MOD -2					
i. -2,-2 / 0,-2	2	-2.492	0.330	-7.541	<0.0001 ***
ii. -2,-2 / -1,-2	1	-0.760	0.361	-2.103	0.9238
iii. -1,-2 / 0,-2	1	-1.732	0.267	-6.492	<0.0001 ***
iv. 0,-2 / 2,-2	2	-2.400	0.322	-7.452	<0.0001 ***
v. 0,-2 / 1,-2	1	-1.542	0.257	-5.997	<0.0001 ***
vi. 1,-2 / 2,-2	1	-0.858	0.347	-2.475	0.7150
MOI AT MOD -1					
vii. -2,-1 / 0,-1	2	-2.541	0.340	-7.471	<0.0001 ***
viii. -2,-1 / -1,-1	1	-0.907	0.368	-2.463	0.7236
ix. -1,-1 / 0,-1	1	-1.634	0.264	-6.188	<0.0001 ***
x. 0,-1 / 2,-1	2	-2.691	0.351	-7.665	<0.0001 ***
xi. 0,-1 / 1,-1	1	-2.614	0.340	-7.686	<0.0001 ***
xii. 1,-1 / 2,-1	1	-0.078	0.435	-0.178	1.0
MOI AT MOD 1					
xiii. -2,1 / 0,1	2	-2.947	0.397	-7.426	<0.0001 ***
xiv. -2,1 / -1,1	1	-0.803	0.446	-1.803	0.9865
xv. -1,1 / 0,1	1	-2.143	0.302	-7.088	<0.0001 ***
xvi. 0,1 / 2,1	2	-2.715	0.351	-7.729	<0.0001 ***
xvii. 0,1 / 1,1	1	-2.566	0.331	-7.763	<0.0001 ***
xviii. 1,1 / 2,1	1	-0.149	0.427	-0.348	1.0
MOI AT MOD 2					
xix. -2,2 / 0,2	2	-2.692	0.310	-8.695	<0.0001 ***
xx. -2,2 / -1,2	1	-0.084	0.369	-0.229	1.0
xxi. -1,2 / 0,2	1	-2.608	0.304	-8.584	<0.0001 ***
xxii. 0,2 / 2,2	2	-1.860	0.309	-6.017	<0.0001 ***
xxiii. 0,2 / 1,2	1	-1.343	0.268	-4.862	0.0003 ***
xxiv. 1,2 / 2,2	1	-0.556	0.341	-1.633	0.9965

(contd.)

(b) English

Comparison (MOI, MOD/MOI, MOD)	Diff. (in MOI units)	Coef.	SE	z	p
MOI AT MOD -2					
i. -2,-2 / 0,-2	2	-0.979	0.277	-3.532	0.0752 .
ii. -2,-2 / -1,-2	1	-0.229	0.277	-0.828	1.0
iii. -1,-2 / 0,-2	1	-0.750	0.273	-2.742	0.5023
iv. 0,-2 / 2,-2	2	-1.003	0.302	-3.323	0.1394
v. 0,-2 / 1,-2	1	-0.510	0.284	-1.800	0.9868
vi. 1,-2 / 2,-2	1	-0.493	0.311	-1.584	0.9977
MOI AT MOD -1					
vii. -2,-1 / 0,-1	2	-1.491	0.309	-4.820	0.0004 ***
viii. -2,-1 / -1,-1	1	-0.572	0.322	-1.772	0.9892
ix. -1,-1 / 0,-1	1	-0.920	0.285	-3.233	0.1774
x. 0,-1 / 2,-1	2	-1.728	0.322	-5.375	<0.0001 ***
xi. 0,-1 / 1,-1	1	-1.103	0.289	-3.812	0.0293 *
xii. 1,-1 / 2,-1	1	-0.625	0.338	-1.849	0.9815
MOI AT MOD 1					
xiii. -2,1 / 0,1	2	-1.323	0.343	-3.859	0.0247 **
xiv. -2,1 / -1,1	1	-0.403	0.370	-1.089	1.0
xv. -1,1 / 0,1	1	-0.920	0.315	-2.922	0.3634
xvi. 0,1 / 2,1	2	-2.121	0.316	-6.721	<0.0001 ***
xvii. 0,1 / 1,1	1	-1.114	0.280	-3.980	0.0158 **
xviii. 1,1 / 2,1	1	-1.007	0.313	-3.222	0.1828
MOI AT MOD 2					
xix. -2,2 / 0,2	2	-0.667	0.336	-1.984	0.9584
xx. -2,2 / -1,2	1	0.051	0.370	0.138	1.0
xxi. -1,2 / 0,2	1	-0.718	0.341	-2.104	0.9235
xxii. 0,2 / 2,2	2	-1.517	0.292	-5.204	0.0001 ***
xxiii. 0,2 / 1,2	1	-0.840	0.288	-2.918	0.3667
xxiv. 1,2 / 2,2	1	-0.677	0.274	-2.467	0.7209

Table 12: Pairwise comparisons between MOI levels in the co-varied conditions (at MODs -2, -1, 1, and 2) for the Spanish and English groups in Experiment 2.

12.b.xvi). An increase in the intensity disparity from 0 to 6 dB was associated with significant differences only at the positive end of the MOI scale (comparisons **12.b.xi** and **12.b.xvii**). This pattern of results, consistent with the outcome of the logistic regression, confirms that incrementally increasing the duration disparity lessened the effect of Intensity on responses.

5 Discussion of specific experimental outcomes

5.1 Intensity and duration taken separately

At the most general level, outcomes for intensity in Experiments 1 and 2 showed that the presence of an intensity disparity between alternating syllables, whatever its magnitude, increased responses representing loud-first syllable pairings relative to a control

sequence in which no disparity was present. This overall preference, expressed as more *baga* responses when *ba* was louder and fewer when *ga* was louder, was significant in both language groups, in both experiments, and in all conditions in which intensity was varied. The basic finding is consistent with the basic prediction of the Loud-Soft Principle of the ITL in (1).

More granular predictions were made by the Incremental Loud-Soft and Intensity Attenuation hypotheses in (2a) and (2b) (see Section 2.4). The Incremental Loud-Soft hypothesis predicted a positive association between increases on the MOI scale and *baga* responses, creating the expectation of a monotonic function. According to the Intensity Attenuation hypothesis, changes between levels were not expected to be proportional. Rather, the effect of increasing the intensity disparity by the same constant (an increment of 3 dB in Experiment 1, and 6 dB in Experiment 2) would be greatest between MOI 0 and MOIs -1 and 1 and should taper off at successively higher levels. This means that the function associated with intensity should be curvilinear at each end of the scale, producing an overall S-shaped function. These hypotheses were generally borne out, with the exception of several unanticipated decreases in *baga* responses (to be discussed momentarily): in the baseline intensity condition in Experiment 2, the increase from 0 to 6 dB of disparity produced proportionally more loud-first groupings than the increase from 6 dB to 12 dB of disparity, with one exception in the Spanish group. A similar pattern was found in Experiment 1. The Intensity Attenuation hypothesis received weaker support in the co-varied conditions in Experiment 2, due to interactions with duration. This outcome was predicted and is discussed later in the section.

Focusing on the baseline intensity conditions, at least two interesting differences were observed across language groups and across experiments. First, the magnitude of the fixed effect of Intensity was greater in Experiment 2 than in Experiment 1. The ORs associated with Intensity indicate that a unit of MOI increase doubled the odds in Experiment 1 (OR = 2.0), and nearly tripled the odds (OR = 2.82) in Experiment 2 of a *baga* response overall. Second, the effect of Intensity was greater in the Spanish than in the English group in both experiments. In Experiment 1, the ORs for Intensity were 2.92 for Spanish and 2.01 for English, and in Experiment 2, the ORs were 8.61 for Spanish and 2.84 for English. These measures reveal large differences across experiments: the odds of a *baga* response were approximately a third higher in the Spanish group than in the English group in Experiment 1; but the odds were three times greater in the Spanish group in Experiment 2. The reason for the between-group difference is unclear, but it suggests that the Spanish speakers were more sensitive to intensity variations than the English speakers tested. A similar between-language difference in a cross-linguistic ITL study was observed by Bhatara et al. (2013). These investigators report a loud-first grouping preference among both monolingual German and French speakers they tested, but this bias was more robust among the German speakers.

The between-experiment differences in the effect of intensity in the current research may be attributable to differences in the design of the stimulus set: in Experiment 1, singly-varied intensity sequences were presented together only with singly-varied duration sequences, whereas in Experiment 2, co-varied sequences were added to the mix. It is difficult to identify other factors that might have contributed to the observed outcome: within each language group, participants were drawn from the same populations and were tested in the same locations; the experimenters, procedures, and equipment were the same. Given these factors, the most reasonable interpretation of the finding may be that listeners' perception of the simpler singly-varied intensity sequences was heightened in comparison with the more complex co-varied sequences.

The S-shaped function associated with intensity, especially in the singly-varied conditions, may have an auditory explanation. When listeners with normal hearing are asked to rate the loudness of a pure tone on a scale from “soft” to “too loud”, loudness growth is roughly linear up to a subjective “loud” rating, at about .4 subjective loudness units per dB (Allen et al., 1990; Olsen et al., 1999). After about 85 dB, the slope is steeper, about 1 subjective loudness unit per dB (Olsen et al., 1999, p. 303). However, Cox et al. (1997) report that loudness growth is different for speech than for tones. These authors conducted a study of loudness growth effects among normally hearing subjects who were asked to subjectively rate the loudness of speech stimuli varying in mean-level intensity (dB HL). These investigators report that at intensity levels characteristic of speech, the loudness growth function was more curvilinear than for tones: increasing intensity (dB HL) was associated with increasingly smaller increases in subjective loudness units between approximately 20 and 70 dB HL. (After approximately 70 dB, the relationship between unit increases in intensity and subjective loudness began to reverse.) In light of Cox et al. (1997), the “tapering off” of loud-first judgments in the current studies suggests that equal increments in the intensity of the more intense syllables may not have produced equal differences in the perception of their loudness.

The outcomes of particular interest for the main research questions were those associated with varied duration. In the singly-varied duration conditions in Experiments 1 and 2, significant and straightforwardly interpretable outcomes for duration were found only in the English group. With one exception (at MOD 2), the presence of any duration disparity in Experiment 1 was associated with more long-last groupings compared with the control sequence, MOD 0 (more *baga* responses when *ga* was longer and fewer when *ba* was longer). This general result was significant and provided overall support for the Long-Last Principle in (1). At a more granular level, the Incremental Long-Last hypothesis in (3b) predicted that proportional increases in MOD would be associated with decreases in *baga* responses. However, this hypothesis was not fully supported in Experiment 1 because changes in *baga* responses between MOD levels were often not proportional and there were irregular deviations from the expected linear function (see **Figure 2a**). In contrast to Experiment 1, the negative trend associated with duration for singly-varied sequences in Experiment 2 was linear and therefore more consistent with the Incremental Long-Last hypothesis. However, the size of the effect was smaller: a unit of MOD increase reduced the odds of a *baga* response by 44% on average (OR .56) in Experiment 1, whereas in Experiment 2, the odds were only 34% lower (OR .66).

In the Spanish group, there was a small but significant negative association between MOD and *baga* responses in Experiment 1. However, changes between levels were very small and only one difference between individual MOD levels (between MODs 0 and 3) was significant, in the direction predicted by the ITL. That the negative correlation was significant suggests a marginal trend supporting the Incremental Long-Last hypothesis. This was the outcome predicted by the ITL, and not what we would expect if listeners were attending to increased duration as a stress-marking cue. However, because the changes were so small, it is difficult to be confident of this interpretation. In Experiment 2, increasing MOD did not affect listeners’ responses in the singly-varied duration condition, with one exception: an increase in *baga* responses at MOD 2, consistent with a long-short grouping bias, was in the direction expected if listeners were attending to varied duration as relating to stress. It is difficult to assign a meaningful interpretation to this result in the absence of a clear trend, and given that outcomes across the two experiments were in conflict.

5.2 Intensity and duration together

An interesting outcome, consistent across experiments and across language groups, was that the fixed effect of Duration was less robust than that of Intensity. The ORs associated with Duration were .56 in Experiment 1 and .66 in Experiment 2, overall, and fixed effects of Duration were more robust in the English group.

Outcomes in the co-varied conditions in Experiment 2 provide two important new pieces of information. First, just as the fixed effect of Intensity was more robust than the fixed effect of Duration (the singly-varied sequences), Intensity was also the more dominant predictor of listeners' responses in the co-varied conditions. These outcomes confirmed the Intensity Wins hypothesis in (2c). Perhaps the most interesting outcome was that even though its influence was less robust, manipulating duration reduced the size of the intensity effect: When there was a fixed duration disparity between alternating syllables, increasing intensity produced fewer responses representing loud-soft groupings (more *baga* groupings when *ga* was longer, fewer when *ba* was longer), and this effect generally strengthened as the duration disparity increased. This finding was consistent with the Incremental Long-Last hypothesis. Based on the outcomes of Experiment 1, this result in the English group was not surprising. However, it was also true in the Spanish group, even though the fixed effect of Duration was not robust in Experiment 1 and not significant in Experiment 2. Surprisingly, given the differences in the fixed effects, the effect on Intensity of increasing the duration disparity was comparable in the two language groups in the co-varied condition (see the ORs associated with the interaction Duration. disp*Intensity in **Table 11**).

The finding that increasing the duration disparity lessened the effect of manipulating intensity on listeners' responses meant different things in the Competing and Co-operating conditions. In **Figures 4** and **5**, the Competing condition is represented by points labeled with squares (the blue lines at negative MOIs and the red lines at positive MOIs), and the Co-operating condition by the points labeled with triangles (the blue lines at positive MOIs and the red lines at negative MOIs). Trends are clearest in **Figure 4a**; there, functions associated with intensity are well differentiated due largely to differences in the baseline duration condition (the stacked circles at MOI 0). Taking **Figure 4a** as the point of reference, it can be seen that the functions associated with co-varied sequences converge with the baseline intensity function in the Co-operating condition, and diverge in the Competing condition. In the Co-operating condition (where one alternating syllable was long and the other loud), the interaction between the predictor variables meant that increasing the duration disparity produced increases in long-short responses. As noted in Section 2.3 (the discussion of Crowhurst and Teodocio, 2014), a possible explanation for this superficially incongruent effect, given listeners' responses to varying duration in other conditions, may be related to the way in which the auditory system integrates intensity over time. When sounds of the same intensity have different lengths, the longer sound is perceived to be louder (see Gordon, 2005, and references cited there). This suggests that there may have been a tendency for listeners to perceive longer, low-intensity syllables as louder than the high-intensity short syllables in the Co-operating condition. It should be noted that the same auditory explanation might be proposed for Crowhurst and Teodocio's (2014) finding of a long-last bias among Zapotec speakers for singly-varied duration sequences. However, that Crowhurst and Teodocio found different grouping biases in the Zapotec and English groups in their singly-varied duration condition when the same stimuli were used may indicate that listeners from different language backgrounds may respond differently, even at a purely auditory level.

For the theoretical questions explored in the current study, the outcome in the Competing condition is more interesting. Here, it can be seen in **Figure 4** that the effect of increasing the duration disparity was to decrease responses representing loud-first groupings as the intensity disparity increased. In other words, increasing a duration disparity increased long-last groupings in both language groups. In the Spanish group, this effect was observed only when duration and intensity were varied together (in competition). Implications of the outcome in the Competing condition for the issues identified in Section 1 and Section 2 are discussed in the following section.

On the whole, generalizing over the two experiments, the Incremental Loud-First and Intensity Attenuation hypotheses were confirmed to an extent that permits a degree of confidence in the relationships they assert. However, as discussed in the results sections for Experiments 1 and 2, several irregularities in the form of unanticipated deviations from the expected (curvi)linear relationships were observed in both language groups in one or both studies. The functions associated with duration in Experiment 1 were not monotonic, as expected, and changes between levels were not proportional, most dramatically in the English group. Such irregularities are not unique to the studies reported here; they have also been found in prior ITL studies that have tested participants using sequences of synthesized syllables (Bhatara et al., 2013; Hay & Diehl, 2007), and sequences produced from naturally recorded speech (Crowhurst & Teodocio, 2014). On the whole, irregularities of this nature appear to have been more prevalent in the series of studies that have based stimuli on natural speech (Crowhurst & Teodocio, 2014, and the research reported here). This difference may reflect a cost of using stimuli prepared from resynthesized natural speech, as opposed to synthesized sequences. It is not possible to manipulate naturally produced speech to precisely control all aspects of stimuli, and this limitation may account for a degree of noise in the study's outcomes. Another source of noise may have been linked to the way in which stimuli were presented to subjects. Other prominently published studies in which subjects heard stimuli in free field over loudspeakers (e.g., Iversen et al., 2008; Sluijter et al., 1997) have achieved clear results; however, a legitimate concern for some research would be that participants inevitably hear stimuli at different sound levels depending on their position in the room. The rationale for presenting stimuli in this way was that the issue of interest was whether listeners' response patterns, based on a binary decision, would generally reflect the order of steps on an ordinal scale; the dependent measure was not a continuous variable that measured the finer details of listeners' responses. I also considered that listeners naturally hear speech in free field, not through headphones. However, concerns about how the mode of presentation might affect listeners' responses could be avoided in future studies by presenting sequences in a more controlled manner under laboratory conditions.

Another factor that may have contributed to the irregularity noted was that in the current studies, intensity-varied and duration-varied sequences were not segregated into different blocks as has been done in prior ITL studies, with the exception of Crowhurst and Teodocio (2014). In the current series of studies (the two reported here, together with Crowhurst & Teodocio, 2014), control, intensity-varied, and duration-varied sequences were presented to study participants in the same blocks because the investigator reasoned that this would provide a stronger test of the ITL. That is, we may be more confident that a loud-first or long-last bias is meaningful if it is observed under conditions in which listeners could be distracted by more variation. Another way in which the investigator worked to increase the rigor of the test was in designing the task in a way that did not direct listeners to think in terms of the relative amplitudes, durations, or prominence in general of the syllables in the alternating sequences they were hearing. Orienting subjects

to the segmental properties of the alternating syllables (*which do you hear – baga or gaba groupings?*) provided an indirect measure of listeners' perceptions, while directing their attention away from the acoustic parameters being manipulated. However, the very indirect framing of the task that was intended to provide a stronger confirmation of the ITL's basic predictions may also have contributed to irregularity in listeners' response patterns.

6 General Discussion

A central debate in the growing literature on rhythmic grouping is whether rhythmic grouping preferences (possibly those described by the ITL) are innate, or whether grouping preferences emerge differently depending on listeners' language backgrounds. The nativist position is represented by Bolton (1984), Hayes (1995), and Hay and Diehl (2007). Others have argued that whether or not there are innate predispositions, rhythmic grouping behaviours show the influence of linguistic experience both in adulthood (Bhatara et al., 2013; Crowhurst & Teodocio, 2014; Iversen et al., 2007) and as early as 7–8 months (Yoshida et al., 2010).

Adult ITL studies exploring intensity-based grouping preferences, including those described here, have consistently reported a loud-first grouping bias (see Section 2.1). A contribution of the current studies has been to add Spanish to the list of languages with this result. Psycholinguistic studies using methodologies that rely on short-term memory also provide evidence that listeners organize syllabic sequences into loud-soft units. In one such study, Morgan et al. (2014) tasked English-speaking adults with memorizing lists of six syllables in which every other syllable was louder and higher-pitched. Participants' recall was better for lists organized in a trochaic pattern (three loud-soft syllable pairs) as compared with an iambic condition (lists arranged into three soft-loud pairs) and a control condition. So far then, the evidence from adult studies using a variety of approaches makes a strong and consistent case for the psychological reality of an intensity-based trochaic template for older speakers regardless of language background.

By contrast, different and sometimes conflicting outcomes for duration produced by ITL studies with adult speakers of various languages, including English, French, German, Japanese, Zapotec, and now Spanish suggest that grouping preferences based on duration are shaped at least in part by language experience (see Section 2.1). Infant studies provide persuasive evidence supporting this conclusion. Two studies stand out. Bion et al. (2011) tested the pitch-based and duration-based rhythmic grouping preferences of native Italian-speaking adults and 7-month-old Italian infants using alternating syllable sequences. These researchers found that when pitch was the feature varied, both adults and infants were better able to recognize pairs of syllables that had been presented in a high-low pitch pattern as opposed to a low-high pattern in a pre-test familiarization phase. When duration was the target feature, the adults were better able to recognize syllable pairs that had been presented in an iambic short-long pattern as opposed to a trochaic long-short pattern during familiarization. In the infant study, however, syllables' prosodic arrangement during familiarization (short-long vs. long-short) was not associated with significant differences in the dependent measure (cumulative looking time). In other words, at 7 months, Italian-learning infants had acquired the pitch-based but not the duration-based grouping preference demonstrated for Italian adults. This study's outcomes suggest that rhythmic grouping preferences may emerge or develop at different times depending on the patterned feature. Clear evidence that sensitivity to duration cues develops as language learning proceeds is reported by Yoshida et al. (2010), who exposed English-learning and Japanese-learning infants to duration-varied tone sequences. They found that at 5–6 months, neither the English nor Japanese learners demonstrated clear preferences for

either short-long or long-short groupings. At 7–8 months, the English learners but not the Japanese learners showed a preference for long-last groupings. Yoshida et al.'s findings are congruent with findings reported for English- and Japanese-speaking adults (Iversen et al., 2008; Kusumoto & Moreton, 1997).

If duration-based rhythmic grouping preferences are learned, then it would make sense for language-specific duration patterns to have a role in shaping them. The general expectation for the current research was that long-last ITL effects might be robust to the extent that constituent-final syllables are saliently marked by increased duration, whatever its origin. English and Spanish were chosen for the present studies because while they are prosodically similar in having a trochaic stress foot at the right edge of lexical words, stress-related and preboundary lengthening patterns differ in ways that are directly related to the main question. In English, stressed and domain-final syllables tend to be longer, and some degree of stress on final syllables is relatively common, especially in conversational speech (see Section 2.3). Because stress-related and preboundary lengthening effects are additive and because the duration contrast is highly salient in English, due in part to unstressed vowel reduction (e.g., Beckman & Edwards, 1994), the occurrence of longer syllables at ends of prosodic constituents should be quite natural for English speakers. It should therefore not be surprising that multiple studies (see Section 2.1), including the current research, have found a long-last grouping preference for English speakers. Final stress is less common in Spanish than in English, and duration differences that signal stress and finality are smaller in Spanish (Prieto et al., 2010). For these reasons, I expected long-last effects to be less robust among the Spanish speakers tested, if they were present at all. This prediction was confirmed by the results of Experiments 1 and 2 in the singly-varied duration conditions. Here, in fact, the Spanish group produced no convincing evidence for any grouping preference: not the long-last bias expected if the Spanish speakers were interpreting duration as a cue signaling finality, nor the long-short bias expected if they were treating duration as a cue to stress and segmenting sequences on this basis.

Outcomes across the two language groups in conditions where intensity and duration were co-varied, on the other hand, were surprisingly similar, given the different results in the singly-varied duration condition. One of the more interesting findings was that in the Competing condition (in which the same syllable was both louder and longer), when the intensity disparity was held constant, increasing the duration disparity increased long-last groupings, simultaneously reducing loud-soft groupings (the dominant bias). This pattern was found in both language groups, and the magnitude of the effect of duration on intensity was comparable across groups. The comparable robustness of the duration effect for Spanish relative to the English group was not consistent with the study's main expectation that duration effect should be weaker among speakers of languages in which the association between duration and final positions is weaker, and it was not consistent with the findings for the singly-varied duration conditions. The long-last response pattern in the Competing condition might be taken to suggest that listeners in both language groups were interpreting length as a cue to finality. Under this interpretation, one possible reason for the failure to find a grouping preference for singly-varied duration sequences for Spanish might be that these seemed less natural to listeners, perhaps because they were not multidimensional, as actual speech is.

Whatever potential explanation attaches to the findings for duration in the current experiments, they present us with four useful pieces of information. The first is that how listeners process duration is to some extent dependent on the context in which it is varied (singly or together with intensity). Second, based on the outcomes with singly-varied duration sequences, we may also conclude that because the English- and Spanish-speaking

participants were tested using the same sequences, the finding of clear differences is evidence that language background is a factor in determining listeners' perceptions of natural syllable groupings, whatever the source of the differences. While these differences were subtle, the revelation of subtle differences between speakers of prosodically similar languages, in relation to the research hypotheses, is a valuable contribution to a program of research in laboratory phonology. The third piece of information (related to the last) is that listeners' duration-based rhythmic grouping preferences are more variable across speakers of different languages than intensity, a finding that is congruent with reports for prior studies (see Section 2.1 and elsewhere). Finally, listeners' use of duration and intensity cues may be to some extent task-dependent. As discussed in Section 2.3, the literature reports that both English and Spanish speakers rely most on duration cues in stress identification tasks (Ortega-Llebaria et al., 2007; Turk & Sawusch, 1996). If listeners use stress-related cues to segment alternating sequences, then the finding that intensity is a stronger predictor of responses than duration across experiments, language groups, and conditions should be surprising. In fact, results in the Competing condition in Experiment 2 indicate that in the Spanish group, duration was a reliable predictor of grouping behaviour only when intensity differences were also present. The clear implication is that how listeners process information is to some extent dependent on the task: listeners seem to make different use of information about intensity and duration when they are assigned a preferential grouping or segmentation task than they do when they perform a stress identification task.

So, what are listeners doing when they make judgements about natural syllable groupings in an ITL study, if they are not accessing and using language-specific information related to the stress contrast? Sequences in which two syllables alternate to form recurrent *gaba* or *baga* units (in the current experiments) are less natural than alternating stress sequences in long words like *Àpalàchicóla* (a city in Florida), and it cannot be assumed that listeners are treating binary groupings as stress feet. Another possibility, however, is that they are processing recurrent bisyllabic units as though they were short words. While we cannot be certain of this, informal observations are suggestive. The speakers who recorded alternating syllable sequences for use in stimulus production tended naturally to pause after each syllable pair (e.g., *baga, baga...*) and required training to produce more fluid sequences. The second observation is that when listeners are exposed to lengthy alternating sequences such as those used here, their perceptions can shift and some have reported hearing units that resembled words. The La Salle instructor who conducted sessions in Mexico humorously commented at one point that she was hearing sequences as repetitions of the word *cabrón!* (a term of insult).

Psycholinguists who have conducted ITL studies and statistical learning studies in both adults and infants have made the connection between long-last effects (or the failure to find them) and language-specific syntactic differences that have implications for the distribution of prosodic length. Iversen et al. (2008) and Yoshida et al. (2010) attribute long-last effects among adult English speakers and English-learning infants and the absence of a consistent preference among Japanese adults and infants to language-specific differences in word order. These researchers note that English is a complement-head language in which prosodically-weak function words precede lexical words in heads of phrases. Japanese has the opposite head-complement order, with function words following syntactic heads. Of course, from a phonological perspective, this translates to a probabilistic difference in the distribution of duration: syllables in function words are shorter than prominent syllables in lexical words. The complement-head structure of English places longer syllables closer to ends of phrases than is the case in Japanese. It is not clear that

differences between Spanish and English speakers in this study can be attributed to a similar difference in syntactic organization, as Spanish is also a complement-head language. One syntactic characteristic that can increase the distance between stressed syllables and word endings in Spanish is the presence of unstressed, post-verbal clitics (e.g., *dímelo* ‘tell it to me’, *siéntate* ‘sit down’). This characteristic was mentioned in Section 2.3 as one of the factors contributing to a strong tendency for constituent-final syllables in Spanish to be unstressed in connected speech.

The robust loud-first grouping bias demonstrated for adult listeners, and differences in the magnitudes of the effects associated with intensity and duration in the current research, may have a point of connection with results of a study reported in Hay and Saffran (2012). These investigators found that 6.5-month-old English-learning infants were able to correctly segment words from longer sequences in a statistical learning and segmentation experiment (and distinguish them from partial words) when the word’s initial syllable was more intense. However, there was no evidence that the 6.5-month-olds used increased duration to identify word endings. In light of current findings relating to the relative influence of varied intensity and duration on listeners’ grouping behavior, Hay and Saffran’s study provides evidence that the strategy of associating increased intensity with word beginnings emerges early in learners of at least one language, before a similar strategy associated with duration is in evidence.

The findings of the current work and the considerations discussed in the preceding paragraphs point to clear avenues for future research. Three studies have now shown that speakers of three languages process varied duration differently depending on whether it is varied singly or together with intensity. When duration is varied singly, Zapotec speakers show a long-short preference (Crowhurst & Teodocio, 2014), English speakers a long-last preference, and Spanish speakers no clear preference (the current experiments). However, when duration and intensity are co-varied, patterns of response in the three language groups are similar. These similarities and differences suggest that future ITL-style studies of grouping preferences should include stimuli that are multidimensional in the features varied, as speech is, and that researchers should be cautious in drawing conclusions based on findings associated with more one-dimensional stimuli. Therefore, further studies to explore effects of co-varying intensity and duration on listeners’ grouping preferences could be useful. An important consideration in designing such studies must be that we don’t yet know to what extent the perceptual grouping of non-speech stimuli such as tones is distinct from the perceptual grouping of more speechlike stimuli. The studies surveyed in Section 2 can largely be sorted into studies that have tested participants using nonspeech stimuli and those that have used edited or synthesized speech. Only one published study, Hay and Diehl (2007), has used a design that included conditions testing both nonspeech and speech-like sequences. However, their groundbreaking study did not include conditions in which intensity and duration were co-varied. Results obtained in the Competing condition in Experiment 2 here were similar to results in the comparable condition in Crowhurst and Teodocio (2014), who also used speechlike sequences. But these outcomes seem to be inconsistent with Bolton’s (1894) report that when intensity and duration were increased on the same tone, subjects preferred long-last and not loud-first groupings. Bolton does not appear to have experimented with different settings for intensity and duration to the extent done here, and so it is not possible to make a nuanced comparison between the finding he reports and outcomes in the current outcomes. However, given the importance of exploring comparisons between listeners’ perceptions of rhythm in the nonspeech and linguistic domains, studies modeled on Hay and Diehl (2007) but which include co-varied conditions would be interesting and timely.

Supplementary Files

The supplementary files for this article can be found as follows:

- **Supplementary File 1: Appendix A.** <http://dx.doi.org/10.5334/labphon.42.s1>
- **Supplementary File 2: Appendix B.** <http://dx.doi.org/10.5334/labphon.42.s2>
- **Supplementary File 3: Appendix C.** <http://dx.doi.org/10.5334/labphon.42.s3>

Acknowledgements

The current research was funded by National Science Foundation of the United States of America grant BCS-1147959, awarded to the University of Texas at Austin for Megan Crowhurst (PI). The Spanish experiments could not have been conducted without the capable and generous assistance of Ms. Dinorah Fernandez Esquer (MA), who recruited and tested subjects in Obregon, Mexico, and Mr. Francisco Jesus Leyva Quintero (MA), Vice Rector of La Salle University, where the Spanish studies were conducted. Associate Editor Lisa Davidson and two anonymous *Laboratory Phonology* reviewers provided thoughtful, detailed feedback on the manuscript which contributed to significant improvements. The paper has also benefitted from comments, questions, and statistical advice offered at different points by Sally Amen, Randy Diehl, and Scott Myers. Portions of this paper have been presented to audiences at the Massachusetts Institute of Technology and the University of Texas at Austin.

Competing Interests

The author declares that she has no competing interests.

References

Alfano, I., Llisterri, J., & Savy, R. 2007. The perception of Italian and Spanish lexical stress: A first cross-linguistic study. *International Conference on the Phonetic Sciences XVI*, 1793–1796.

Allen, J. B., Hall, J. L., & Jeng, P. S. 1990. Loudness growth in $\frac{1}{2}$ -octave bands (LGOB) – a procedure for the assessment of loudness. *Journal of the Acoustical Society of America*, 88, 745–753.

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

Beckman, M. E. 1986. *Stress and non-stress accent*. Dordrecht, Holland/Riverton, USA: Foris Publications. DOI: <http://dx.doi.org/10.1515/9783110874020>

Beckman, M. E., & Edwards, J. 1994. Articulatory evidence for differentiating stress categories. In Keating, P. (Ed.), *Phonological structure and phonetic form. Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, pp. 1–33.

Bell, A. 1977. Accent placement and perception of prominence in rhythmic structures. In Hyman, L. (Ed.), *Studies in stress and accent, Southern California Occasional Papers in Linguistics 4*, pp. 1–13.

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. 2013. Native language and stimulus complexity affect rhythmic grouping of speech. *Journal of the Acoustical Society of America*, 134, 3828–3843. DOI: <http://dx.doi.org/10.1121/1.4823848>

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. 2011. Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, 54, 123–140. DOI: <http://dx.doi.org/10.1177/0023830910388018>

Boersma, P., & Weenink, D. 2011. *Praat: doing phonetics by computer* [Computer program]. Version 5.3.13. Retrieved from <http://www.praat.org/>

Bolton, T. 1894. Rhythm. *American Journal of Psychology*, 6, 145–238. DOI: <http://dx.doi.org/10.2307/1410948>

Buckley, E. 1998. Iambic lengthening and final vowels. *International Journal of American Linguistics*, 64, 179–223. DOI: <http://dx.doi.org/10.1086/466357>

Byrd, D. 2000. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57, 3–16. DOI: <http://dx.doi.org/10.1159/000028456>

Cambier-Langeveld, T. 1997. The domain of final lengthening in the production of Dutch. In de Hoop, C. (Ed.), *Linguistics in the Netherlands*. Amsterdam: John Benjamins, pp. 13–24.

Campbell, N., & Beckman, M. E. 1997. Stress, prominence, and spectral tilt. *Proceedings of the ESCA Workshop on Intonation: Theory, Models, and Applications*. Athens, Greece, 67–70.

Cox, R. M., Alexander, G. C., Taylor, I. M., & Gray, G. A. 1997. The contour test of loudness perception. *Ear and Hearing*, 18, 388–400. DOI: <http://dx.doi.org/10.1097/00003446-199710000-00004>

Crowhurst, M. J., Kelly, N. E., & Teodocio, A. 2016. The influence of vowel glottalisation and duration on the rhythmic grouping preferences of Zapotec speakers. *Journal of Phonetics*, 58, 48–70. DOI: <http://dx.doi.org/10.1016/j.wocn.2016.06.001>

Crowhurst, M. J., & Teodocio Olivares, A. 2014. Beyond the Iambic-Trochaic Law: the joint influence of duration and intensity on the perception of rhythmic speech. *Phonology*, 31, 51–94. DOI: <http://dx.doi.org/10.1017/S0952675714000037>

Cutler, A., & Butterfield, S. 1992. Rhythmic cues to speech segmentation: evidence from juncture misperception. *Journal of Memory and Language*, 31, 218–236. DOI: [http://dx.doi.org/10.1016/0749-596X\(92\)90012-M](http://dx.doi.org/10.1016/0749-596X(92)90012-M)

Cutler, A., & Carter, D. M. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–142. DOI: [http://dx.doi.org/10.1016/0885-2308\(87\)90004-0](http://dx.doi.org/10.1016/0885-2308(87)90004-0)

den Os, E. 1988. *Rhythm and tempo of Dutch and Italian: A contrastive study* (Doctoral dissertation). Rijksuniversiteit, Utrecht.

Dogil, G., & Williams, B. 1999. The phonetic manifestation of word stress. In van der Hulst, H. (Ed.), *Word prosodic systems in the languages of Europe*. Berlin: Mouton de Gruyter, pp. 273–311. DOI: <http://dx.doi.org/10.1515/9783110197082.1.273>

Domahs, U., Plag, I., & Carroll, R. 2014. Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *Journal of Computational German Linguistics*, 17, 59–96. DOI: <http://dx.doi.org/10.1007/s10828-014-9063-9>

Fletcher, J. 2010. The prosody of speech: timing and rhythm. In Hardcastle, W. J., Laver, J., & Gibbon, F. E. (Eds.), *The handbook of phonetic sciences* (2nd ed.). Malden, MA: Wiley-Blackwell, pp. 523–602. DOI: <http://dx.doi.org/10.1002/9781444317251.ch15>

Gordon, M. (2005). A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory*, 23, 595–653. DOI: <http://dx.doi.org/10.1007/s11049-004-8874-9>

Halle, M., & Vergnaud, J.-R. 1987. *An essay on stress*. Cambridge, MA: MIT Press.

Harris, J. 1983. *Syllable structure and stress in Spanish*. Cambridge, MA: MIT.

Hay, J., & Diehl, R. 2007. Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics*, 69, 113–122. DOI: <http://dx.doi.org/10.3758/BF03194458>

Hay, J. F., & Saffran, J. R. 2012. Rhythmic grouping biases constrain infant statistical learning. *Infancy*, 17, 610–641. DOI: <http://dx.doi.org/10.1111/j.1532-7078.2011.00110.x>

Hayes, B. 1982. Extrametricality and English stress. *Linguistic Inquiry*, 13, 227–276.

Hayes, B. 1985. *A metrical theory of stress rules*. New York: Garland Press.

Hayes, B. 1987. A revised parametric metrical theory. *Northeastern Linguistics Society*, 17, 274–289.

Hayes, B. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Heldner, M., & Strangert, E. 2001. Temporal effects of focus in Swedish. *Journal of Phonetics*, 29, 329–361. DOI: <http://dx.doi.org/10.1006/jpho.2001.0143>

Hosmer, D. W., Jr., & Lemeshow, S. 2004. *Applied logistic regression*. New York: John Wiley & Sons.

Hualde, J. I. 2007. Stress removal and stress addition in Spanish. *Journal of Portuguese Linguistics*, 6, 59–89. DOI: <http://dx.doi.org/10.5334/jpl.145>

Iversen, J. R., Patel, A. D., & Ohgushi, K. 2008. Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, 124, 2263–2271. DOI: <http://dx.doi.org/10.1121/1.2973189>

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. 2005. Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118, 1038–1054. DOI: <http://dx.doi.org/10.1121/1.1923349>

Kusumoto, K., & Moreton, K. 1997. Native language determines parsing of nonlinguistic rhythmic stimuli. *Journal of the Acoustical Society of America*, 105, 3204. DOI: <http://dx.doi.org/10.1121/1.420936>

Lindblom, B. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35, 1773–1781. DOI: <http://dx.doi.org/10.1121/1.1918816>

Lipsky, J. 1997. Spanish word stress: the interaction of moras and minimality. In Martínez-Gil, F., & Morales-Front, A. (Eds.), *Issues in the phonology and morphology of the major Iberian languages*. Washington, DC: Georgetown University Press, pp. 559–593.

Morales-Front, A. 1999. El acento. In Núñez-Cedeño, R., & Morales-Front, A. (Eds.), *Fonología generativa de la lengua Española*. Washington, DC: Georgetown University Press, pp. 203–230.

Morgan, J. L., Edwards, S., & Wheeldon, L. R. 2014. The relationship between language production and verbal STM: the role of stress grouping. *Quarterly Journal of Experimental Psychology*, 67, 220–246. DOI: <http://dx.doi.org/10.1080/17470218.2013.799216>

Olsen, S. Ø., Rasmussen, A. N., Nielsen, L. H., & Borgkvist, B. V. 1999. The acoustic reflex threshold: not predictive for loudness perception in normally-hearing listeners. *International Journal of Audiology*, 38, 303–307. DOI: <http://dx.doi.org/10.3109/00206099909073040>

Ortega-Llebaria, M., & Prieto, P. 2007. Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 282, 155. DOI: <http://dx.doi.org/10.1075/cilt.282.11ort>

Ortega-Llebaria, M., & Prieto, P. 2011. Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech*, 54, 73–97. DOI: <http://dx.doi.org/10.1177/0023830910388014>

Ortega-Llebaria, M., Vanrell, M., & Prieto, P. 2007. Perceptual evidence for direct acoustic correlates of stress in Spanish. In *Proceedings of the XVIth international congress of phonetic sciences*, pp. 1121–1124.

Ortega-Llebaria, M., Vanrell, M., & Prieto, P. 2010. Catalan speakers' perception of word stress in unaccented contexts. *Journal of the Acoustical Society of America*, 127, 462–471. DOI: <http://dx.doi.org/10.1121/1.3268506>

Prieto, P., Vanrell, M., Astruc, L., Pyne, E., & Post, B. 2010. Speech rhythm as durational marking of prosodic heads and edges. Evidence from Catalan, English, and Spanish. *Proceedings of Speech Prosody*.

Quilis, A. 1993. *Tratado de fonología y fonética españolas*. Madrid: Gredos.

R Development Core Team. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org/>

Rao, R. 2010. Final lengthening and pause duration in three dialects of Spanish. In Ortega-Llebaria, M., (Ed.), *Selected proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*. Somerville, MA: Cascadilla Proceedings Project, pp. 69–82.

Rietveld, T., Kerkhoff, J., & Gussenhoven, C. 2004. Word prosodic structure and vowel duration in Dutch. *Journal of Phonetics*, 32, 349–371. DOI: <http://dx.doi.org/10.1016/j.wocn.2003.08.002>

Rice, C. C. 1992. *Binarity and ternarity in metrical theory: Parametric extensions* (Unpublished doctoral dissertation). University of Texas, Austin.

Sebastian, N., & Costa, A. 1997. Metrical information in speech segmentation in Spanish. *Language and Cognitive Processes*, 12(5-6), 883–887. DOI: <http://dx.doi.org/10.1080/016909697386781>

Selkirk, E. O. 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11, 563–605.

Selkirk, E. O. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Sluijter, A. M. C., & van Heuven, V. J. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100, 2471–2485. DOI: <http://dx.doi.org/10.1121/1.417955>

Sluijter, A. M. C., van Heuven, V. J., & Pacilly, J. J. A. 1997. Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America*, 101, 503–513. DOI: <http://dx.doi.org/10.1121/1.417994>

Turk, A., & Sawusch, J. 1996. The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*, 99, 3782–3790. DOI: <http://dx.doi.org/10.1121/1.414995>

Turk, A. E., & Shattuck-Hufnagel, S. 2000. Word-boundary-related durational patterns in English. *Journal of Phonetics*, 28, 397–440. DOI: <http://dx.doi.org/10.1006/jpho.2000.0123>

Turk, A., & Shattuck-Huffnagel, S. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35, 445–472. DOI: <http://dx.doi.org/10.1016/j.wocn.2006.12.001>

Vos, P. 1977. Temporal duration factors in the perception of auditory rhythmic patterns. *Scientific Aesthetics/Sciences de l'Art*, 1(3), 183–199.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707–1717. DOI: <http://dx.doi.org/10.1121/1.402450>

Woodrow, H. 1909. *A quantitative study of rhythm: The effect of variations in intensity, rate, and duration*. Columbia University Contributions to Philosophy and Psychology, Vol. XVIII, No. 1.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. F. 2010. The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition*, 115, 356–361. DOI: <http://dx.doi.org/10.1016/j.cognition.2010.01.005>

How to cite this article: Crowhurst, M 2016 Iambic-Trochaic Law Effects among Native Speakers of Spanish and English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1):12, pp.1–41, DOI: <http://dx.doi.org/10.5334/labphon.42>

Published: 07 October 2016

Copyright: © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 