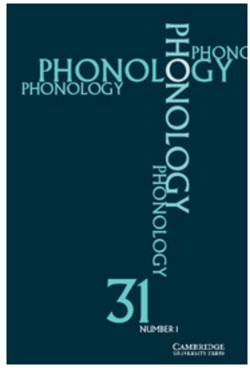

Beyond the Iambic-Trochaic Law: the joint influence of duration and intensity on the perception of rhythmic speech

Megan J. Crowhurst and Amador Teodocio Olivares

Phonology / Volume 31 / Issue 01 / May 2014, pp 51 - 94
DOI: 10.1017/S0952675714000037, Published online: 20 May 2014

Phonology **31** (2014) 51–94.
doi:10.1017/S0952675714000037

*Beyond the Iambic-Trochaic Law: the joint influence of duration and intensity on the perception of rhythmic speech**

Megan J. Crowhurst
University of Texas at Austin

Amador Teodocio Olivares
Center for the Study and Development of the Indigenous Languages of Oaxaca

The Iambic-Trochaic Law (ITL) asserts that listeners associate greater acoustic intensity with group beginnings and greater duration with group endings. Some researchers have assumed a natural connection between these perceptual tendencies and universal principles underlying linguistic categories of rhythm. The experimental literature on ITL effects is limited in three ways. Few studies of listeners' perceptions of alternating sound sequences have used speech-like stimuli, cross-linguistic testing has been inadequate and existing studies have manipulated intensity and duration singly, whereas these features vary together in natural speech. This paper reports the results of three experiments conducted with native Zapotec speakers and one with native English speakers. We tested listeners' grouping biases using streams of alternating syllables in which intensity and duration were varied separately, and sequences in which they were covaried. The findings suggest that care should be taken in assuming a natural connection between the ITL and universal principles of prosodic organisation.

* E-mail: MCROWHURST@AUSTIN.UTEXAS.EDU, TEODOCIO64@YAHOO.COM.MX.
This research was supported by National Science Foundation award 1147959 to the first author. The second author thanks the Center for the Study and Development of the Indigenous Languages of Oaxaca for release time to conduct the research. We have appreciated useful comments from Keith Johnson, Mark Liberman, Björn Lindblom, Richard Meier, Gayatri Rao, Rajka Smiljanic and Caroline Smith at various points. We are especially indebted to Colin Bannard, Randy Diehl, Matthew Gordon, Niamh Kelly, Scott Myers and four *Phonology* reviewers for their generous contributions to the design, statistical analysis and earlier versions of the manuscript. Finally, we thank the editors for their gracious and meticulous editing. Naturally, the responsibility for any limitations is ours alone. Portions of this research were presented at the 2013 meeting of the Society for the Study of Indigenous Languages of the Americas and at M@90, a workshop in honour of Morris Halle's 90th birthday in Cambridge, Mass. in September 2013.

1 Overview and objectives

Humans so naturally organise auditory events into rhythmic units that we tend to manufacture rhythm even when there are no physical cues to its presence. As an example, listeners tend to experience the clicking of a chronograph as recurrent groups of beats distinguished by volume or length, even when the sounds and their spacing in sequence are identical (Bolton 1894: 185, Handel 1989). Not surprisingly, then, many human languages are characterised by rhythmic patterns expressed as the recurrent pairwise alternation of stressed and unstressed syllables. An issue of historical importance to phonologists has been whether stress rhythms provide evidence of prosodic feet, i.e. abstract rhythmic constituents assumed to be linguistic categories (Selkirk 1980, 1984, Hayes 1985, 1995, Halle & Vergnaud 1987, among others), or whether the impression of structure is a surface phenomenon, on which view the groupings should have no status in theory (Prince 1983). The constituentist position has predominated, at least among phonologists, and the theoretical status of the foot is secure, in large part because functions independent of stress have been identified for feet in an extensive literature on prosodic morphology (McCarthy & Prince 1990, 1996 and work following in that tradition). Foot-sized units are used in 'templatic' morphological operations even though they may not be associated with stress (for example, the habilitative construction in Cupeño; Hill 1970, McCarthy 1984, Crowhurst 1994).

As to the formal characteristics of feet, one prominent proposal holds that the prosodic foot typology is that in (1) (Hayes 1987, 1995, McCarthy & Prince 1996), where 'L' represents a light, monomoraic syllable, 'H' a heavy, bimoraic syllable and 'ˈ' identifies the stress-bearing syllable.

(1) *The prosodic foot typology*

Balanced	Syllabic trochee	(ˈLL)	Quantity-insensitive
	Moraic trochee	(ˈLL)(ˈH)	Quantity-sensitive
Asymmetric	Iamb	(LˈH)	

The typology in (1) represents a particular proposal about the interdependence of three parameters that determine foot form: sensitivity to syllable-quantity, quantitative symmetry and trochaic *vs.* iambic headedness.[1] In quantity-sensitive stress systems, heavy (bimoraic) syllables preferentially receive stress, in opposition to light (monomoraic) syllables, and heavy syllables are predicted to occur in foot-final position. In quantity-insensitive stress systems, the foot is balanced, in that any two syllables may combine, whatever their content. Quantity-sensitive stress

[1] For influential theoretical discussions of syllable quantity, see Hyman (1985), Hayes (1989) and McCarthy & Prince (1996). A discussion of phonetic factors that influence syllable weight is found in Gordon (2002).

systems may also use balanced feet, in which case two light syllables may combine, while a heavy syllable stands alone. In other quantity-sensitive stress systems, a heavy syllable may combine with a light syllable in an asymmetric foot. An innovative claim represented by (1) is that stress systems based on balanced feet tend strongly to have trochaic rhythm, while those based on asymmetric feet are almost exclusively iambic, especially in systems where feet are assigned iteratively.

The typology in (1) has been understood in different ways by different phonologists. A strong version holds that only the feet in (1) are specified by UG. On this view, other shapes are defective, and are the product of phonological rules (see case studies in Hayes 1995). A more moderate view is that (1) represents typologically unmarked foot structures, although marked options also occur. As examples, quantity-insensitive iambs are argued to occur in Osage (Altshuler 2009) and Southern Paiute (Sapir 1930), and some trochaic systems enforce a quantitative requirement on stressed syllables (e.g. trochaic lengthening in languages such as Ancient Greek (Revithiadou 2004) and Swedish (Riad 1992)). The majority (though not the consensus) position may be that some version of the weaker position is defensible.[2]

Theoretical arguments for particular foot structures have generally relied on the distribution of stress in languages with iterative metrical systems. However, it often happens that more than one theoretical analysis is compatible with transcriptions provided in original sources. The extent of the problem can be illustrated with the polysyllabic word [ˈwitaŋˈhataˈlaːbaˈtsu] 'away from the Tejon Indians' in Tübatulabal (Northern Uto-Aztecan; Voegelin 1935). Some theorists (e.g. Hayes 1985, Crowhurst 1991) have argued that Tübatulabal assigns asymmetric iambs from right to left, but that non-canonical feet are allowed so that syllables may be exhaustively parsed to feet, (ˈwi)(taŋ.ˈha)(ta.ˈlaː)(ba.ˈtsu). Arguing from a different theoretical perspective, Kager (1989, 1991) treats Tübatulabal as a language with right-to-left moraic trochees, barring the final syllable, which receives word stress (by End Rule Right) without being parsed in a foot. Assuming a ban on degenerate feet, surface forms allow unparsed syllables, (ˈwi.taŋ)(ˈha.ta)(ˈlaː).ba.ˈtsu. A third analysis that relies on moraic (balanced) iambs is proposed in Kager (1993). The point is that evidence traditionally used to support theoretical arguments for metrical structure (symbols on a printed page) may not be sufficient to choose among competing analyses.

A better understanding of more concrete, physical cues that can mark transitions between feet in languages with foot-dependent phonological patterns might provide useful information bearing on typological claims about the shapes that feet can have. An important proposal advanced by Hayes (1987, 1995) in support of the prosodic foot typology invokes the principles stated in (2), which emerged from early research on the psychology of auditory grouping (e.g. Bolton 1894: 232,

[2] Critiques of (1) are discussed in greater detail in §7.

Woodrow 1909: 69, 77). These principles are known collectively as the Iambic-Trochaic Law (ITL).

(2) *The Iambic-Trochaic Law* (adapted from Hayes 1995: 80)
   a. *Intensity Principle*
      Elements contrasting in intensity naturally form groupings with initial prominence (trochees).
   b. *Duration Principle*
      Elements contrasting in duration naturally form groupings with final prominence (iambs).

The ITL is not a typological statement of distribution, as is (1), but rather concerns the nature of perceptual precursors for the foot inventory. Hayes (1995: 81) views the grouping principles in (2) as 'an *extrasystemic* motivation, in a law of rhythm, for *internal* formal principles of the linguistic system' (emphasis in the original). Therefore, a theory that includes the ITL implicitly claims, if not universality, then at least broad generality for the principles in (2).[3]

Decades of modern research in phonetics have also identified acoustic intensity and phonetic duration as important physical cues associated with the production and identification of stressed syllables in many languages (see below). Given these findings, several related questions present themselves for investigation. First, do intensity and duration function in the same way as cues to grouping affinity as they do as cues for stress? Painting in broad strokes, the picture that has emerged from the phonetic literature is that the most reliable physical correlates of (non-accent) stress in English speakers' productions are increased vowel duration and greater intensity (Fry 1955, Lehiste 1970, Beckman 1986). Published reports have claimed that vowel duration is the cue that listeners use most reliably in identifying word stress (Fry 1955, 1958, Adams & Munro 1978, Beckman & Edwards 1994, Ortega-Llebaria *et al.* 2010), with intensity playing a more limited and elusive role, not only in well-studied languages such as English, Dutch, Spanish and Catalan (see in addition Fry 1965, Sluijter & van Heuven 1996, Sluijter *et al.* 1997, Ortega-Llebaria *et al.* 2010), but also in Thai (Potisuk *et al.* 1996) and the Quiaviní dialect of Valley Zapotec (Chávez-Peón 2008). However, since a stressed syllable may occupy different positions within the foot, it should not be assumed that a stress judgement is functionally equivalent to a grouping judgement. A better understanding of perceptual influences on grouping can only be achieved through a programme of research that teases the two apart.

The early psycho-acoustic studies that inspired the ITL have been followed up by more rigorously controlled modern experiments that have also studied the separate influence of intensity and duration on the subjective grouping of patterned non-speech elements such as tones and

[3] The ITL has not been universally embraced by phonologists. See §7 for further discussion.

beeps (e.g. Bell 1977, Rice 1992, Kusumoto & Moreton 1997, Hay & Diehl 2007, Iversen *et al.* 2008), and alternating synthesised syllables (Hay & Diehl 2007, Bhatara *et al.* 2013). Bolton (1894) also studied the joint effect on subjective grouping of simultaneously varying intensity and duration in the same patterned sequences of tones.[4] To date, however, no research has investigated the interactive influence of these cues on the subjective grouping of speech-like sequences. Because the phonetic literature on stress has shown that intensity and duration are intricately and symbiotically associated, this is the next step that must be taken if we are to connect the ITL to categories in language.

A third point concerns the equivalent status of intensity and duration implied by the ITL and how these features might or might not be connected to the formal features used by phonologists to describe underlying and/or surface contrasts in languages. While phonetic duration is not directly correlated with syllable weight, it has a phonological analogue in vowel length, the most basic determinant of syllable quantity. Phonetic duration has another important role in demarcating the ends of phonological constituents from the level of the foot upward to the highest levels of the prosodic hierarchy (Nespor & Vogel 1986, Gussenhoven & Rietveld 1992, Byrd 2000). Given this connection, the proposal that the Duration Principle in (2b) might be a phonetic precursor for the constituent-final position of heavy syllables within feet, in particular the asymmetric iamb, seems reasonable. However, it is less obvious why the Intensity Principle in (2a) should be a good candidate as a precursor for balanced feet: intensity is not a feature that patterns contrastively in the phonological systems of human languages.[5] Given the state of our knowledge on these points, we might expect duration to be a more important influence than intensity on judgements of grouping. Finally, any claim of broad generality for the grouping principles in (2) must be supported by cross-linguistic evidence – the more the better – and evidence must be sought from speakers of languages beyond those usually studied.

Modern tests of the ITL have yielded interesting, though mixed, results for both non-speech and speech-like stimuli. In aggregate, these studies do not provide satisfying answers to the questions identified in the preceding paragraphs. Existing reports describe studies that have tested the ITL with native speakers of only a few languages, and the key published studies using speech-like stimuli (Hay & Diehl 2007, Bhatara *et al.* 2013, Morgan *et al.* 2014) have tested only speakers of western European languages. A limitation of all modern, rigorously controlled studies of subjective grouping is that they have only studied the influence of intensity and duration when varied separately. None has explored the extent to which

[4] Bolton (1894) does not provide detailed information about the characteristics of his subjects, but his experiments were conducted at an American university.

[5] Recent research suggests that varying vowel glottalisation, contrastive in some languages and subphonemic in others, may also influence listeners' perceptions of syllable grouping (Kelly *et al.* 2014).

intensity and duration affect subjective grouping when both are varied in the same sequences.

The goal of the research described in this paper was to take the next step, by studying the interactive influence of vowel duration and intensity on subjective grouping, using rhythmic speech sequences in which both features are varied. Three grouping experiments were conducted with native speakers of a northern Zapotec language spoken in San Melchor Betaza, Oaxaca, Mexico. For comparison, a fourth study was conducted with native speakers of a southern variety of American English. Betaza Zapotec was chosen because we had access to that community (the second author is a native speaker), and because it offered an opportunity to explore the questions of interest with speakers of a non-European language whose sound structure is unlike those of languages represented in the existing subjective grouping literature. The role of phonetic duration cues in signalling prosodic boundaries and contrasts in Betaza Zapotec and English are in some ways similar and in others different. It is not obvious, then, that Zapotec speakers will use duration cues to access rhythm in the way that has been described for speakers of English. Zapotec is also unlike other languages whose speakers have been tested to date, in employing a complicated set of phonation and tonal contrasts, which also interact with prosodic structure. Overall, the findings for both the Zapotec and English-speaking groups studied provided clear confirmation of the Intensity Principle in (2a). However, there were similarities and differences in how duration influenced subjective grouping, depending on whether and how duration and intensity were varied together in the sequences presented to listeners. Going forward, §2 provides a review of the most relevant experimental research on the grouping of rhythmic sequences. Relevant details about the study's setting and the Zapotec language spoken by participants are provided in §3. The three studies conducted with Zapotec speakers are described in §4, and the experiment conducted with English speakers is described in §5. §6 and §7 present a general discussion of the outcomes of the study, broader implications and directions for further research.

2 Experimental background

The terms RHYTHM and RHYTHMIC have been used in different ways by researchers, whose perspectives and goals are often distinct and varied. In the psycholinguistic and much of the phonetic literature, rhythm is most commonly conceived as sentence or discourse-level fluctuations in prominence resulting from the distribution of syllables bearing the main stresses of content words throughout syntactic constituents. An extensive psycholinguistic literature has focused on the strategic use that listeners make of these stressed syllables to identify the beginnings of words in segmenting speech (e.g. Cutler & Carter 1987, Cutler & Norris 1988, Cutler & Butterfield 1992). A related body of work has argued that infants become attuned early to language-specific 'rhythms' (in this larger sense), as reflected in the characteristic patterns of the adult language spoken around them (Jusczyk *et al.* 1993, Morgan 1996, Echols *et al.* 1997, Echols & Crowhurst 1998). Young children learning English use the characteristic stress patterns of the ambient language to extract words from speech (Echols 1993, Echols *et al.* 1997). Patterns in which function words and affixes in the speech of young English learners are omitted reflect a bisyllabic, trochaic (strong–weak) template (Gerken 1991, 1994, Demuth 1996), and there is some evidence that young learners of K'iche' Maya employ a similar strategy (Pye 1983). In this literature, the concepts of trochaic and iambic rhythm are often encountered as labels to distinguish bisyllabic words with initial word stress from those with final word stress in languages like English. The term rhythm, in this broader sense, has also been used by phoneticians concerned with identifying correlates to rhythm in the speech signal, often with the aim of clarifying differences among languages belonging to different 'rhythm classes' (stress-timed, syllable-timed and mora-timed languages; Ramus *et al.* 1999, Grabe & Low 2002). This perspective is also evident in the work of scholars who have studied connections between language rhythms and music, sometimes with the aim of positing evolutionary links between rhythm in these domains (Iversen *et al.* 2008, Patel 2008). None of these research traditions has been much concerned with cues that might influence subjective judgements of grouping at the level of the foot. Our use of the terms rhythm and rhythmic reflects a tradition common in the phonological literature to refer to sequences in which stressed and unstressed syllables alternate in an (often strictly, but sometimes roughly) iterative fashion. The perception literature most relevant for this study is therefore limited to research on the subjective grouping of patterned sequences of this type.

In early experimental studies of the psychology of auditory grouping, Bolton (1894) and Woodrow (1909) explored listeners' subjective groupings of rhythmic sequences in which relatively louder tones were separated by one or more softer tones, and in which longer tones were separated by at least one shorter tone. Their findings provided the basis for the ITL in (2).[6] In addition to single-parameter conditions in which only intensity or duration was varied, Bolton tested sequences in which both parameters were varied in the same strings. He reports that when intensity and duration were varied out of phase, such that a long soft tone was alternated with a short loud one in pairs, his subjects tended to prefer pairings in which the louder tone came first and the longer tone came last. On the other hand, when intensity and duration were varied in phase,

[6] This brief description simplifies the scope and findings of these early studies of rhythm. Both Bolton and Woodrow studied the effects of rate on the perception of rhythm as well as the effect of varying the duration of the silent intervals separating tones in their sequences. Bolton, in particular, studied the perception of rhythm with patterned series more complex than sequences comprised of recurrent pairs of tones.

Bolton reports a preference for groupings in which the longer, louder syllable came last. This loud-initial, long-final grouping is doubly consistent with the ITL. As the grouping principles of the ITL were in competition in Bolton's in-phase sequences, the latter finding seems to suggest that contrasts in duration provide a more salient cue to grouping affinity than contrasts in intensity. Although of great interest, these studies are limited, insofar as they do not measure up to the methodological standards of modern experimental research and their findings were not supported by statistical analyses.

Modern studies that have tested the ITL using non-speech sequences in which intensity and duration were varied separately (in different sequences) have yielded mixed results overall. Bell (1977) reports a tendency for English-speaking adults to perceive louder sounds as being group-initial in rhythmic strings that contrasted intensity. Results obtained by Vos (1977) confirm the Duration Principle in (2b) (sounds of greater duration are perceived to be group-final) for adult native speakers of Dutch, and Rice (1992) reports similar preferences for English speakers.[7] Trainor & Adams (2000) found support for the Duration Principle, but not the Intensity Principle, in a gap-detection study in which English-speaking adults and 8-month-old English-learning infants were presented with tonal sequences of a louder or longer tone separated by two softer or shorter tones respectively. For duration-varying sequences, these researchers found that both adults and infants were better able to detect gaps occurring after a short than after a longer tone. Reasoning that gaps should be more noticeable when they interrupt the integrity of natural groupings, Trainor & Adams concluded that subjects' responses were consistent with short–short–long groupings. Their study did not support the Intensity Principle in (2a), they concluded, due to inadequate controls in their intensity condition. Another study, Yoshida *et al.* (2010), tested sensitivity to variations in duration with infants whose ambient language was either English or Japanese. These authors found that by eight months, the English-learning infants showed a short–long grouping bias for duration-varying sequences. In contrast, the Japanese-learning infants showed no clear preference.

Although no conclusive grouping preference was associated with intensity in Trainor & Adams' (2000) study, other studies using non-speech sequences have found strong support for the Intensity Principle in groups of monolingual adult speakers of Japanese, English and French. Kusumoto & Moreton (1997) and Iversen *et al.* (2008) report on paired studies in which native speakers of English and Japanese were tasked with grouping tonal sequences in which every other tone was longer or louder and in which intensity and duration were varied at different ratios.[8] In both

[7] In addition to duration, Rice tested the effect of varying pitch and of varying the silence interval between sounds on listeners' grouping decisions.

[8] Kusumoto & Moreton included a third condition, in which pitch was varied. They found a weakly significant high–low parsing preference for their English-speaking group and for a Japanese-speaking group tested in one dialect region (Kansai). A

studies, all groups showed a significant loud–soft parsing preference when intensity was varied. Iversen *et al.* (2008) found this bias to be significantly stronger for the English speakers than for the Japanese speakers. Neither study found a significant effect of the ratio at which a parameter was varied (disparity), although both studies reported a positive correlation between trochaic parsing and magnitude in the case of intensity. Results reported for grouping behaviour with duration-varying sequences have been less consistent. Both Kusumoto & Moreton and Iversen *et al.* found a strong iambic (short–long) parsing preference in the duration condition among their English-speaking participants. By contrast, neither study found a consistent iambic parsing preference among their Japanese listeners when duration was varied. In fact, in both studies the grouping preferences of many of their Japanese participants contradicted the Duration Principle: some of those tested showed a strong short–long grouping bias, but many instead showed a long–short grouping pattern. Iversen *et al.* (2008) conclude that the differences they found between the English and Japanese speakers in their study are culturally conditioned, and speculate that the cultural difference most likely to be the influencing factor is language, even though the sequences they tested used tones rather than speech.

Hay & Diehl (2007), who conducted a methodologically similar subjective grouping study with native speakers of American English and French, report stronger results validating both of the principles in (2). Participants in their study were tested with rhythmic non-speech sequences (alternating square wave segments) in which either intensity or duration was varied at four different ratios. Both the English and French-speaking groups showed a significant overall trochaic grouping preference in the intensity-varying condition, and a significant overall iambic bias in the duration-varying condition, as predicted by the ITL. In the intensity-varying condition, moreover, Hay & Diehl found a significant effect of magnitude of intensity in both language groups: increases in magnitude produced proportional increases in loud-soft grouping decisions. However, the relationship between magnitude of duration and parsing preferences was less linear for their French-speaking than for the English-speaking listeners.

Remarkably, given the importance of the Iambic-Trochaic Law to phonologists, the published literature contains relatively few reports of carefully controlled perception studies parallel to those just described, but using more speech-like sequences. For the methodology used in this paper, in which subjects are tasked with grouping lengthy sequences of alternating syllables in which either duration or intensity was varied, the key studies for speech are Hay & Diehl (2007) and Bhatara *et al.* (2013). Hay & Diehl, in addition to their tests with rhythmic non-speech sequences, included parallel speech conditions in which speech-like, recombinant sequences of a synthesised syllable *ga* were tested with

---

Japanese-speaking group tested in a different dialect region (Tokyo) showed no preference.

matched groups of native English and French-speaking listeners. The outcomes in Hay & Diehl's speech conditions validated both of the grouping principles in (2): both the English and French-speaking groups showed a significant trochaic grouping bias in the intensity-varying condition and a significant iambic bias (more short–long groupings) in the duration-varying condition. Moreover, Hay & Diehl found a strong effect of magnitude for both groups of subjects, in both the intensity- and duration-varying conditions: increasing the ratio at which each parameter was varied produced increases in the proportion of responses predicted by the ITL. Given that European French does not have rhythmic stress in the way that English does, Hay & Diehl interpret their findings as evidence that subjective grouping biases are based in part on general perceptual biases that are not wholly accounted for by the specific patterns of individual languages. The outcomes of Hay & Diehl's study are important for at least two reasons. They are the first to have conducted carefully controlled grouping studies with speech. Second, differences in the processing of duration-varying sequences in the speech and non-speech conditions among Hay & Diehl's French speakers suggest that we should not assume that the findings of grouping studies conducted with non-speech stimuli can be generalised to speech.

In a methodologically similar study, Bhatara *et al.* (2013) explored the grouping preferences of French speakers and German speakers. The duration-varying and intensity-varying sequences used in Bhatara *et al.*'s experiments, although synthesised, were more speech-like in two ways: they used syllables that combined vowels and consonants of several different qualities, and transitional cues between adjacent syllables were manipulated to simulate coarticulation. As in Hay & Diehl's study, participants were tasked with indicating whether sequences were comprised of strong–weak or weak–strong syllable pairs (where a strong syllable was defined as either louder or longer than its complement). In the German-speaking group, the proportion of trochaic responses was significantly higher than chance when intensity was varied, and significantly lower than chance when duration was varied. Of special interest was that in this study, when judging control sequences in which intensity and duration were held constant, listeners thought they heard trochaic groupings almost as often as they did when intensity was varied. This outcome suggests that German-speaking listeners may have a general trochaic bias that cannot be attributed to the effect of varying intensity, but which can be overridden when cues for duration contrasts are present. In this respect (among others), Bhatara *et al.* found significant differences between the French and German-speaking groups they tested. Outcomes in the control, intensity and duration conditions for the French group in Bhatara *et al.*'s study were better differentiated, in that there were more trochaic groupings in the intensity than in the control and duration conditions, and fewer trochaic groupings in the duration than in the other two conditions. However, the magnitudes of the observed effects were smaller in the French group, indicating that these French speakers showed less sensitivity to variations in the two parameters tested than did the German-speaking listeners. Bhatara *et al.*'s findings suggest that listeners' subjective grouping biases are influenced not only by the nature of the parameters varied, but also by their language background.[9]

The grouping studies described above, in particular those that used speech-like sequences (Hay & Diehl 2007, Bhatara *et al.* 2013), represent an important first step toward a better understanding of the issues outlined in §1. However, what we have learned from grouping studies that vary duration and intensity separately has limited applicability to speech, because in actual speech, as noted earlier, vowel-duration and intensity cues interact. Before we report on the research we conducted as a first step toward addressing this gap, we present in §3 a brief overview of the relevant characteristics of the Zapotec language spoken by the participants in our study.

3 Betaza Zapotec

San Melchor Betaza is a Zapotec municipality located in the Sierra Madre de Oaxaca, a 4–5 hour bus trip northwest of Oaxaca City, Oaxaca, Mexico. The village is situated in a linguistically Zapotec-dominant zone in the Villa Alta district, and the nearest municipalities are also Zapotec-dominant. The population of Betaza was given as 919 in the 2005 *Instituto Nacional de Estadística y Geografía e Informática* census.[10] All inhabitants of Betaza are ethnically Zapotec, and Zapotec is the native and preferred language of almost all Betazans of all ages in daily life at home and in the community. At any given time, the majority of adult village dwellers are women, as a large proportion of the adult men have migrated temporarily in search of employment, and this asymmetry was reflected in the groups sampled in our studies.

The primary source on Betaza Zapotec is Teodocio Olivares (2009). Betaza Zapotec belongs to a cluster of closely related Xhon dialects that also includes Zapotec varieties spoken in Yalalag (Avelino 2004),

[9] Recent psycholinguistic studies have also investigated the predictions of the ITL using methodologies that relied on short-term memory (see the studies cited for details). For example, Bion *et al.* (2011) found that Italian-speaking adults' recall for syllable pairs on test trials was better when these pairs matched bisyllabic short–long chunks (as compared with long–short chunks) that they had heard during a familiarisation phase. Morgan *et al.* (2014) trained English-speaking adults with lists of six syllables in which every other syllable was louder and higher-pitched (i.e. three loud–soft pairs or three soft–loud pairs). In a control condition, all syllables had the same pitch and intensity. These researchers found that participants were able to remember significantly more syllables in a list when the syllables had been presented as three loud–soft pairs than when they were presented as soft–loud pairs, or when the syllables in the list were not distinguished by differences in amplitude and pitch. The investigators concluded that increased intensity can be an effective prompt for trochaic grouping, but only when syllables are arranged in a trochaic pattern.

[10] Available at http://www.oeidrus-oaxaca.gob.mx/fichas/tomoI/distrito13.pdf.

San Pedro Cajonos (Nellis & Hollenbach 1980), Yatzachi el Bajo (Pike 1948, Leal 1950, Butler 1980), Rincon Zapotec (Earl 1968), Yatee (Jaeger 1983) and San Bartolome Zoogocho (Long & Cruz 1999, Sonnenschein 2004). In common with other Xhon varieties, Betaza Zapotec contrasts short vowels with modal voicing, /i o e a/, checked vowels, /iˀ oˀ eˀ aˀ/, and rearticulated (medially glottalised) vowels, /iˀi oˀo eˀe aˀa/, and lacks a phonological length contrast (Pike 1948, Nellis & Hollenbach 1980, Avelino 2004). Xhon and all other Zapotec varieties have a highly salient set of contrasts between fortis and lenis consonants. Detailed phonetic descriptions of fortis and lenis consonants in Zapotec languages are found in Avelino (2004) for Yalálag, Jaeger (1983) for Yatee, Chávez-Peón (2010) for Quiaviní, a Valley Zapotec variety, and Leander (2008) for San Francisco Ozolotepec Zapotec, a Southern Zapotec variety. For our purposes, it is sufficient to note that in Betaza Zapotec, fortis non-sonorants are phonetically realised as voiceless obstruents, and the lenis stops are voiced and spirantised. All fortis consonants are phonetically longer than their lenis counterparts.

Primary stress in Betaza Zapotec occurs on the first syllable of the root in native, non-compounds, and is unaffected by simple affixation (prefixes and suffixes).[11] Stress is therefore morphologically, not phonologically, determined: a trochaic foot is assigned at the left edge of the morphological root. All Zapotec languages have contrastive tone, and tone structure varies among different varieties. Betaza Zapotec has four contrastive tones: high, low and falling tone contrast robustly, while rising tone is less frequent. In addition to participating in the four-way opposition among tones, the low tone is also the default tone in Betaza Zapotec. As far as we are aware, tone and stress do not interact. Some varieties of Oto-Mangue languages (for example, Ayutla Mixtec; de Lacy 2002, citing Pankratz & Pike 1967) are said to promote higher tones in stressed and lower tones in unstressed positions, but this does not seem to be the case in Betaza Zapotec (for example, low–high trochees such as [ˈjì.ʧé] 'paper' are common). Superficially, all tones are possible on stressed and on unstressed syllables, with the restriction that contour tones do not occur adjacently.[12]

Despite the relatively robust literature on Xhon varieties, there appear to be no phonetically detailed descriptions of physical correlates of stress. According to Chávez-Peón's (2008) study of stressed vowels before lenis consonants in Quiaviní Zapotec (a valley Zapotec language not closely related to Betaza Zapotec), the most important phonetic correlate of stress before a lenis consonant is vowel duration. Chávez-Peón's results, based on measurements of stressed and unstressed vowels before lenis

11 In this, Betaza Zapotec differs from some other Xhon varieties. In Cajonos Zapotec, for example, primary stress occurs on the penultimate syllable of a native root of two or more syllables which ends in an open syllable with a modal or checked vowel, otherwise on the root-final syllable (Nellis & Hollenbach 1980: 100).

12 An exploration of the extent to which tonal cues influence subjective grouping is beyond the scope of the present study, but this is an obvious direction in which this programme of research could be extended.

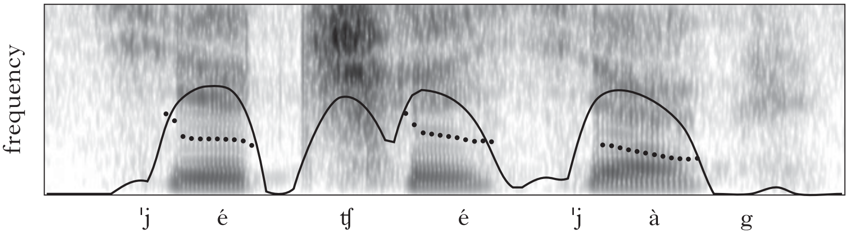

Figure 1
Spectrogram of *yeche yag* 'bush' (female speaker). The solid line is the intensity contour; the dotted lines are pitch contours.

consonants in the productions of four native speakers, suggest that stressed modal and breathy vowels before lenis consonants are 2·5 times longer than corresponding unstressed vowels in Quiaviní. While we have not conducted a formal study, spectrographic evidence suggests that stressed vowels are longer than comparable unstressed vowels in Betaza Zapotec. However, stress is not the only factor that conditions increased vowel duration in Xhon varieties, including Betaza. Pike (1948) reports a number of contexts that promote greater vowel duration in Yatzachi Zapotec: vowels may be longer before a high tone, before lenis consonants and before sonorant consonants (whether fortis or lenis). Vowels may also be longer when 'under emphasis', and 'vowels toward the front of a long word are frequently shorter than vowels in the last two syllables of a word or phrase' (Pike 1948: 167). Pike's comments suggest that increased vowel length can be both stress-marking and demarcative in Yatzachi. In Cajonos Zapotec, stressed modal vowels are lengthened in open syllables and before lenis consonants (Nellis & Hollenbach 1980: 99). Checked vowels (though apparently not rearticulated vowels) are also lengthened when stressed. Importantly, all sources agree that increased vowel duration is a marker of stress only before lenis consonants in Xhon (and other) Zapotec varieties. Before fortis consonants, stressed vowels remain short and the consonant is lengthened instead (Jaeger 1983, Avelino 2004, Leander 2008, Chávez-Peón 2010).

As in other varieties, stressed vowels in Betaza Zapotec are characterised by greater duration before lenis consonants. While we have not conducted a formal study, the distinction between stressed and unstressed vowels does not seem to be strongly cued by differences in intensity. As an illustration, Fig. 1 contains the spectrogram for *yeche yag* [ˈjéʧé ˈjàg] 'bush'. It can be seen from the intensity contour that the intensities of the first vowel (average 76·92 dB) and the second vowel (average 75·12 dB) are comparable. (The stressed vowel in [ˈjé] is relatively short, because it precedes a fortis consonant.)

In summary, the foot in Betaza Zapotec can be analysed as a quantity-insensitive trochee, increased vowel duration is a more salient cue to stress before lenis consonants and increased intensity is not a strong marker of stressed syllables. The sequences used in this study, which we describe in the next section, consist of recurring CVCV pairs in which all intervocalic consonants are lenis stops. Given these facts, if our Zapotec-speaking participants' subjective impressions of grouping affinity align with phonetic markers of stress in their language as well as with its phonological structure, we would expect two general outcomes. First, we should find a stronger preference in general for long–short groupings, an outcome not predicted by the ITL. Second, when listeners are tasked with subjectively grouping sequences in which both intensity and duration are varied, duration cues should be a more robust predictor of listeners' preferences than intensity. Anticipating the discussion to follow, the outcomes of the studies described in the sections to come were not consistent with these expectations.

4 The experiments

In this section we report on three experiments exploring the influence of intensity and duration on subjective grouping among native speakers of Betaza Zapotec. Experiment 1, described in §4.1, was a basic test of the ITL, using speech-like sequences of syllables in which either vowel intensity or vowel duration was varied. This was intended as a baseline study against which to compare the results of the subsequent experiments. Surprisingly, in light of reported outcomes of similar studies with speakers of other languages (see the references cited above), our baseline study with Zapotec speakers produced no clear evidence of grouping asymmetries. In Experiment 2 (§4.2), listeners' grouping preferences were tested using sequences in which vowel intensity and duration were both varied (i.e. covaried). In contrast to the first study, in Experiment 2 both intensity and duration cues were found to be significant predictors of listeners' grouping patterns. Experiment 3 (§4.3) was conducted to further explore the apparently paradoxical findings of the first two experiments. Because they form a tight series of studies, a general discussion of the results will be deferred until after the details of the three experiments have been presented.

All of the experiments described here employed methodological techniques similar to those in earlier studies, but they differed in several important ways. First, in some previous studies (e.g. Hay & Diehl 2007, Bhatara *et al.* 2013), subjects were required to indicate whether they had heard strong–weak or weak–strong groupings, where 'strong' was defined as longer or louder. Our sequences alternated two syllables, *de* and *ge*, and subjects were instructed to indicate whether they heard recurrent *dege* or *gede* pairings. This methodology made it possible to access subjects' intuitions about grouping without referring to notions such as stress or relative syllable strength and, we hoped, without directly tapping listeners' intuitions about the same. Second, some previous studies segregated their intensity-varying and duration-varying stimuli into different blocks, and assigned matched groups of subjects to different conditions (e.g. Hay & Diehl 2007, Bion *et al.* 2011). In the studies described here, stimuli for different conditions were intermixed in blocks that were presented to all participants, as we reasoned that this procedure would make for a stronger test of our hypotheses (see also Bhatara *et al.* 2013). Finally, while other studies (e.g. Hay & Diehl 2007, Bion *et al.* 2011, Bhatara *et al.* 2013) tested subjects individually or in small groups under laboratory conditions, our participants were tested in groups and heard stimuli in free field, played over a portable loudspeaker. We employed this method because it was a practical and culturally appropriate manner in which to test participants in the field. This method had also been used successfully in a similar study (Iversen *et al.* 2008).

4.1 Experiment 1

As noted, Experiment 1 was a basic test of the ITL, and examined the influence of varying either vowel intensity or vowel duration, but not both, on subjective grouping preferences among native speakers of Zapotec. As the results of Experiment 1 were inconclusive, our discussion of the study's outcomes in this section will be brief. However, as the basic materials used in Experiment 1 were used again in Experiments 2, 3 and 4, and as the experimental procedures were largely the same, the methods section for Experiment 1 is more detailed than might otherwise seem warranted. To facilitate the exposition of the main thread, many technical details have been moved to appendices in the supplementary online materials.

4.1.1 *Method*

Overview. Participants in all of the experiments reported here were presented with sequences of roughly ten seconds, formed by alternating coarticulated syllables *ge* and *de*. These syllables were chosen because neither *dege* nor *gede* is a word in Betaza Zapotec when both syllables are low-toned, as they were in our materials. The sequences used in Experiment 1 were made to alternate rhythmically, by decreasing the duration or intensity of the vowel in one base syllable (the syllables used in the control sequences, either *de* or *ge*) relative to the other, in one, two or three fixed increments. Varying intensity and duration in this way produced a magnitude scale on which the intensity or duration of *de* relative to *ge* increased in seven steps, as illustrated in Fig. 2.[13] The zero point on the scale represents a control condition in which there was no disparity between *de* and *ge*, and the expansion of the triangle to the left and

[13] The point of reference in Fig. 2 is the unit *dege* (as opposed to *gede*), because *dege* was arbitrarily coded as the positive response in the statistical analysis (see §4.1.2).

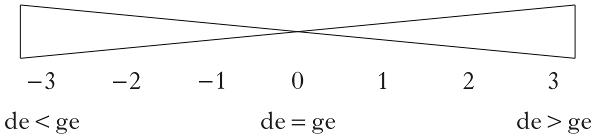

Figure 2
Seven-point magnitude scale for duration and intensity.

right represents increasing disparity in intensity or duration between the alternating syllables.

Subjects in all experiments were instructed simply to indicate whether they thought any given sequence was comprised of recurrent *dege* or *gede* groupings by marking a response sheet, as described below. A *dege* response represented a soft–loud or short–long grouping at the negative end of the scale, and a loud–soft or short–long grouping at the positive end. We did not assume that Zapotec speakers would necessarily respond to varied intensity and varied duration in the ways predicted by the ITL – but if they did, then the proportion of responses representing loud–soft and short–long groupings should increase as a function of increasing disparity between the syllables (i.e. the ratio at which intensity and duration were varied). Using this design, therefore, our specific ITL-inspired hypotheses were as stated in (3).

(3) a. *Intensity*
The proportion of *dege* responses will rise in proportion to increasing magnitude of intensity (MOI).

b. *Duration*
The proportion of *dege* responses will decrease in proportion to increasing magnitude of duration (MOD).

Although details of design differ, the magnitude scale and the specific hypotheses in (3) apply to all experiments described in this paper.

Stimulus preparation and design. Stimulus preparation began with naturally produced, coarticulated sequences …***de***-*ge*-***de***-*ge*… and …***ge***-*de*-***ge***-*de*… (in which boldface represents syllables that were emphasised naturally by the speaker), which were recorded by the second author, a native speaker of Betaza Zapotec. An important objective of this series of experiments was to study cues that might influence listeners' judgements of grouping affinity, while factoring out information that might be provided by cues to stress. Therefore, clear tokens of *de* and *ge* (or more properly [ðe] and [ɣe], since the onsets were spirantised) were selected from unemphasised positions for further manipulation in Praat (Boersma & Weenink 2011). These precursor syllables were closely matched for key phonetic parameters described in Appendix A in the online supplementary materials.[14] A control sequence of roughly ten seconds was constructed by alternating *ge* and *de* tokens whose vowels measured 255·6 and 255·1 ms in duration respectively, and 62 dB in average intensity. These values were also used as the maximum values for alternating syllables in the duration and the intensity conditions. A series of duration-varying sequences was created by alternating the longest syllable with one of three shorter syllables. The differences in duration represented in the series were 40, 80 and 120 ms.[15] Two counterbalanced series were manufactured, one with *ge* and the other with *de* as the shorter syllable. A similar series of intensity-varying sequences was produced by alternating the loudest syllable with one of three softer syllables, such that the differences in intensity represented in the series were 4, 8 and 12 dB. The maximum difference of 12 dB was chosen because this range had been used with positive results in a previous study (Hay & Diehl 2007), and we reasoned that using the same range would make it easier to compare results across studies. As called for by the design, two series were manufactured, one with *ge* and the other with *de* as the louder or longer syllable. To mask the sequences' onsets, a segment of white noise was faded out during the first five seconds, and the amplitudes of the alternating syllables were faded in over the same period (following Hay & Diehl 2007). The ends of sequences were backwards masked by a 500 ms segment of noise. The more technical details of stimulus preparation are provided in the online supplementary materials (App. A).

Subjects. The participants were 32 (26 female and 6 male) native speakers of Betaza Zapotec, whose ages ranged between (approximately) 20 and 60 at the time of the study. (Two of these were ultimately excluded from the statistical analysis, because it was clear from their responses that they had not followed instructions.) Participants were recruited through the second author's personal and professional network and were paid 100 Mexican pesos at the end of each session.

Experimental procedure. Participants in Experiment 1 were tested in a private home and in a municipal building in San Melchor Betaza.[16] Sequences were presented in free field at full volume over a portable Bose speaker positioned in front of the group, connected by cable to a Macintosh PowerBook Pro used to run the experiment. The Zapotec language environment was reinforced in the following ways: informed consent procedures, instructions and any discussion prior to testing were carried out verbally in Zapotec by the second author, who conducted

[14] Importantly, the syllables were closely matched for vowel quality, as vowel quality can be strongly correlated with stress or the lack of stress. See the detailed description in Appendix A, available in the supplementary online materials at http://journals.cambridge.org/issue_Phonology/Vol31No01.

[15] The increment size of 41 used in varying duration represents five whole period cycles in the sound wave produced by the speaker, which was just over 8 ms. We round down to 40 ms. Therefore, duration disparities reported as 40, 80 and 120 Hz were in actual fact 41, 82 and 123 ms.

[16] All experimental procedures were carried out under a protocol approved by the Institutional Review Board of the University of Texas at Austin.

the experiment. The verbal instructions provided by the experimenter were reinforced by diagrams drawn with markers on large sheets of paper, taped to a wall at the front of the room. For its symbolic value, all verbal information on instruction and response sheets given to participants was in Zapotec, even though most Betazans do not read in Zapotec. During the experiment, the second author intermittently kept pace by calling out the number of each trial in Zapotec. The first author, an outsider in this context, was not present during testing. An information sheet written in Spanish was given to participants at the conclusion of the experiment. Stimuli were presented to subjects in six test blocks, which were preceded by four practice trials. In each block, the twelve intensity-varying and twelve duration-varying sequences and two iterations of each control string (a total of 28 sequences) were intermingled. This design (28 sequences $\times$ 6 blocks) provided for a maximum of 5040 possible data points. Sequence order within blocks was randomised each time a block was run by the SuperLab program used to control the experiment. The software was managed by a member of the community who had been trained for this purpose. The verbal instructions given to participants before testing described the nature of the sequences they would hear, and instructed them to indicate whether each sequence consisted of recurrent *gede* or *dege* pairings by circling an item on response sheets (one for each block) they were given. The entire experiment took approximately 50 minutes to complete. Further details concerning the verbal instructions, printed materials and assistance provided to participants who required it are provided in the online supplementary materials (App. B).

4.1.2 *Results.* The response data for each cell in the design are presented in Table X in the online supplementary materials (App. C). The dependent variable Response measured *dege* groupings, the positively coded response in the statistical analysis. The proportion of *dege* responses as a function of increases in MAGNITUDE OF INTENSITY (MOI) and MAGNITUDE OF DURATION (MOD) is charted in Fig. 3. It is clear that the specific hypotheses for intensity and duration stated in (3) were not confirmed by the results of Experiment 1.

Statistical confirmation of this outcome was sought by fitting a logistic regression mixed effects model to the response data (1680 valid responses) using the *lmer* function in the *Matrix* library of the statistical software package R (R Development Core Team 2012). This method estimated the maximum likelihood of a positive response (*dege*, coded as 1), based on the Laplace approximation. Subject was treated as a random effect. The predictor variables Intensity and Duration were coded as scales with seven levels, −3 to 3. There were two control variables, Block (1–6) and Order (whether the sequence began with *de*, coded as 'a', or *ge*, coded as 'b').

As Intensity and Duration did not interact in Experiment 1, two data sets were constructed, consisting of the response data for the control sequences and either the intensity (Set A) or the duration sequences (Set B). The model that best fitted the data for Set A had the two terms

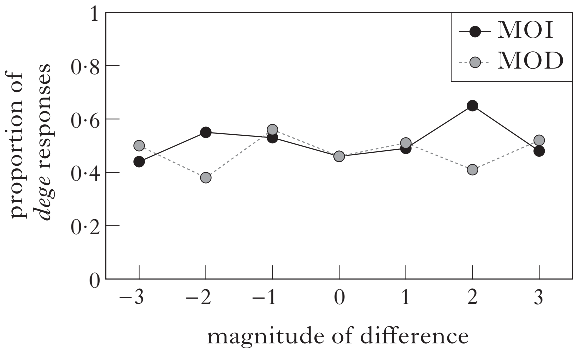

Figure 3

Response data for intensity-varying and duration-varying sequences in Experiment 1, expressed as the proportion of *dege* responses.

whose parameter estimates, standard error (SE), z score and significance levels are shown in Table I. This model provided a marginally better fit for the data than a model that included only the intercept ($\chi^2 = 3{\cdot}174$, $df = 1$, $p\ (>\chi^2) = 0{\cdot}0748$). Neither control variable contributed significantly to goodness of fit.

	coefficient	SE	z	$p\ (>\lvert z\rvert)$
intercept	0·036	0·058	0·620	0·5354
Intensity	0·037	0·021	1·786	0·0741

Table I

Statistical model measuring the likelihood of *dege* responses for intensity-varying sequences (Set A) in Experiment 1.

The output of the model that provided the best fit for the Set B data is shown in Table II. This model, which contained terms for the intercept and for the predictor variable Duration, did not provide a significantly better fit than the model that contained only the intercept ($\chi^2 = 0{\cdot}229$, $df = 1$, $p\ (>\chi^2) = 0{\cdot}633$).

It can be seen that no significant effect was associated with either Intensity or Duration in the statistical models represented by Tables I and II, and our hypotheses were therefore not validated by the results of Experiment 1: neither Intensity nor Duration was a consistent predictor of listeners' responses. This outcome stands in contrast to findings of positive ITL effects reported for similar studies that have used speech-like

	coefficient	SE	*z*	*p* (>|*z*|)
intercept	−0·096	0·045	−2·152	0·031
Duration	0·010	0·021	0·478	0·633

Table II
Statistical model measuring the likelihood of *dege* responses for duration-varying sequences (Set B) in Experiment 1.

stimuli with speakers of English, French, German and Italian (see references cited in §2). Although inconclusive, Experiment 1 served as a baseline against which to compare the results of Experiments 2 and 3, and for this reason, further comments are reserved for the general discussion following the exposition of those studies.

4.2 Experiment 2

Experiment 2 was designed as a companion study to Experiment 1, and was conducted with members of the same Zapotec community on the following day. The purpose of Experiment 2 was to study any influence on listeners' subjective grouping behaviour that vowel intensity and duration might have when both were varied in the same sequences. To this end, listeners were presented with rhythmic sequences in which syllables prepared for Experiment 1 were combined according to a new design.

4.2.1 *Methods*

Overview: design and hypotheses. A crucial difference between Experiments 1 and 2 was that the design of the second experiment allowed for the possibility of competition between the specific hypotheses stated earlier in (3). In the new design, different settings of both vowel duration and vowel intensity were varied in one of two ways. In a Cooperating condition, a softer, longer *dee* or *gee* was alternated with a louder, shorter complement (***ge*** or ***de*** respectively). The label Cooperating reflects the fact that in this condition, a loud–soft grouping (e.g. ***de****gee*), which satisfied the Intensity Principle of the ITL, was simultaneously a short–long grouping, which satisfied the Duration Principle. As a ***de****gee* grouping is doubly consistent with the ITL, our general hypothesis was that listeners' responses would trend in favour of this outcome (and not *gee****de***). In a Competing condition, on the other hand, the louder and longer syllables were the same: duration and intensity were varied so that a louder, longer ***dee*** or ***gee*** was paired with a softer, shorter complement, *ge* or *de*. In this condition, any pairing (e.g. ***gee****de* or *de****gee***) would satisfy either the Intensity or the Duration Principle, but never both. A higher proportion of responses consistent with ***gee**de* (or ***dee**ge*) pairings overall would suggest that varied intensity was the stronger predictor of the listeners' grouping preferences. Alternatively, a bias toward *de**gee*** (or *ge**dee***) pairings would suggest a greater role for duration. The questions of interest in Experiment 2 concerned the relative contributions of varying vowel intensity and duration to any grouping biases in the Competing condition. We expected that trends in the response data would consistently indicate that one parameter (varied intensity or duration) was a better predictor of listeners' responses than the other. As 'better' can be measured as a function of effect size, we had two additional competing hypotheses, stated in (4), for the Competing condition.

(4) a. The size of the effect associated with Intensity will be larger than the size of the effect associated with Duration.
b. The size of the effect associated with Duration will be larger than the size of the effect associated with Intensity.

Although the additional hypotheses in (4) could also be applied to the Cooperating condition, we had no reason to expect one outcome as opposed to the other. Regardless of which emerged as the dominant predictor of listeners' grouping patterns, we hypothesised that the strength of any preference would increase in proportion to increases in the magnitude of difference for the relevant parameter in the direction consistent with the ITL, as stated earlier in (3). Apart from the new design features and hypotheses, the procedures used in Experiment 2 were largely as described for Experiment 1 (see the descriptions in §4.1 and the online supplementary materials), so only the differences will be reported here.

Stimuli. The control sequences prepared for Experiment 1 were used again in Experiment 2. Test sequences were produced by manipulating and alternating the syllables created for Experiment 1 in new combinations according to the design in Table III in §4.2.2. The magnitude scales for intensity and duration introduced in Fig. 2 were also critical in Experiment 2, and the same increments of variation for intensity and duration were used, as described earlier. In Experiment 2, however, the set of test sequences was much larger than for Experiment 1. This was a fully counterbalanced design which called for the different levels on the MOI and MOD scales to be crossed, producing 72 new test sequences.

Subjects and experimental procedures. Twenty-seven adults aged between 14 and roughly 70 participated in Experiment 2. Most had taken part in Experiment 1 on the previous day. Testing was done in similar locations and under the same conditions as described for Experiment 1. Because the stimulus set was large, two blocks of 76 trials (four control and 72 test sequences) were presented to participants in four balanced sets of 38, each followed by a short rest. Accordingly, each participant was given a booklet consisting of an instruction page and four response sheets printed with 38 numbered, underlined blanks. Participants were instructed to write the letter 'd' in a blank if they thought the corresponding

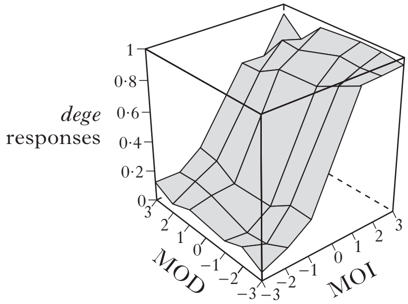

Figure 4
Three-dimensional graph showing the proportion of *dege* responses (*z*-axis) as a function of the interaction of MOI (*x*-axis) and MOD (*y*-axis) in Experiment 2. The proportion 0·5 represents even odds of a response indicating a loud–soft or soft–loud grouping. Proportions lower than 0·5 indicate increases in loud–soft groupings in which *ge* was the louder syllable, and proportions greater than 0·5 loud–soft groupings in which *de* was louder. The Competing condition is represented by the points at the top left corner of the display (positive values for both MOI and MOD) and the bottom right corner (negative values for MOI and MOD). The Cooperating condition is represented by points at the top right (positive MOI with negative MOD) and bottom left (negative MOI with positive MOD). The absence of bisecting lines describing magnitude 0 for intensity and duration reflects the fact that no strings in which only one parameter was varied were used.

sequence consisted of recurring *dege* pairings, or ‘g’ for *gede* pairings. With time for all the necessary components, the experiment took just over an hour to complete.

4.2.2 *Results.* The design and response data for Experiment 2 are presented in Table XI in the online supplementary materials (App. C). The wire graph in Fig. 4 presents this information as the proportion of *dege* responses in all test conditions as a function of increases in MOD and MOI. The graph shows that the proportion of *dege* responses was positively correlated with increases in MOI overall, consistent with hypothesis (3a). By and large, the trends for intensity across MOD levels conform to an S-shaped curve, indicating that the influence of intensity was concentrated in the centre of the scale (MOI −1 to 1), and was attenuated toward its extremes. There were some trend reversals, especially at the extremes of the MOI scale (for example, MOD −2, MOD −1 and MOD 3). An intensity disparity of 12 dB, the greatest difference between alternating syllables at MOI |3|, is considerably larger than disparities found in normal speech. Given this, it is possible that these reductions in loud–soft groupings reflect listeners’ sensitivities to a threshold which

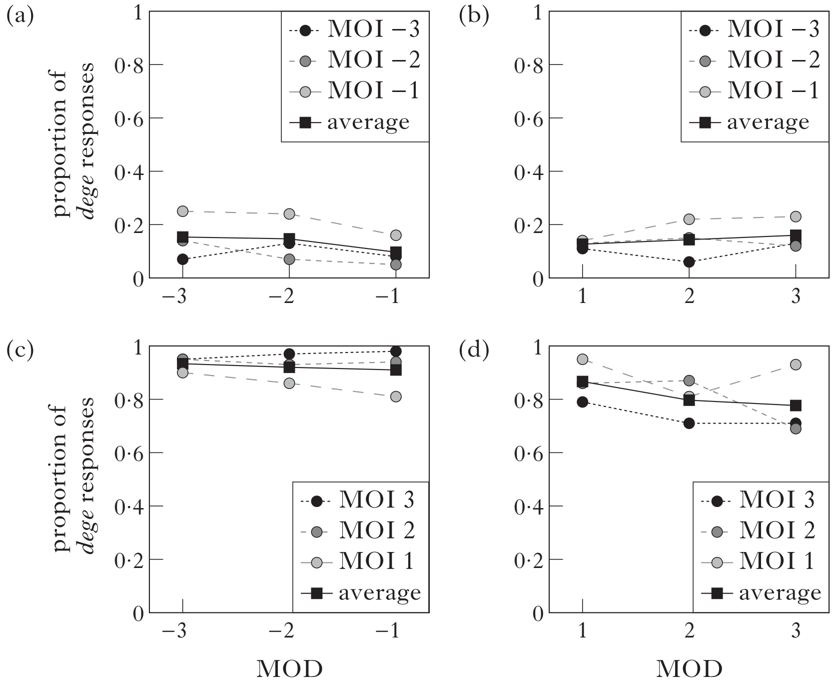

Figure 5

Disaggregated trends associated with MOD (*x*-axis) for different MOI levels (*y*-axis) in Experiment 2. The disparity in intensity for MOI |1| was 4 dB, for MOI |2| 8 dB and for MOI |3| 12 dB. The solid lines with squares represent the averages.

separated sequences that listeners might have judged as more like speech from those which were less acceptable.[17] Despite deviations from a fully monotonic trend, it is evident that increasing MOI was a stronger influence on listeners' grouping choices than increasing MOD. In comparison to the control condition in which the proportion of *dege* responses was 0·4 (not represented in Fig. 4), the proportion of loud–soft groupings in all conditions where intensity was varied was greater than 0·7 (see App. C) and, on average, exceeded 0·8 when the intensity disparity between alternating syllables was larger than 4 dB (MOI |1|).

Trends associated with duration were less consistent. These can be seen more clearly in the disaggregated line charts in Fig. 5. The proportion of *dege* responses is represented only for levels at which there was a duration

[17] We thank Matthew Gordon for this observation.

disparity in Fig. 5: changes in the proportion of *dege* responses from levels 0 to |1| reflect the influence of intensity, so that actual effects of duration are associated with levels |1|–|3| on the MOD scale. In Fig. 5, the light grey circles represent the condition in which the disparity in intensity was least (4 dB, MOI |1|), the dark grey circles represent an intensity disparity of 8 dB (MOI |2|) and the black circles represent the condition in which the intensity disparity was greatest (MOI |3|). The black squares represent the averages. The greatest influence of increasing the duration disparity occurred when the intensity disparity was least. The effect of duration at this level was consistent, but trends in the Competing and Cooperating conditions differed in a surprising way. In the Competing condition, the trends associated with increasing the duration disparity were consistent with the ITL: the increase in *dege* responses in Fig. 5a and the decrease in Fig. 5d are both compatible with increases in short–long groupings, *de**gee*** and *ge**dee***, as predicted by hypothesis (3b). However, the outcome in the Cooperating condition was not consistent with the prediction of the ITL for duration: the trends represented in Figs 5b and 5c indicate increases in long–short *dee**ge*** and *gee**de*** groupings. Outcomes associated with duration were more variable when the intensity disparity was greater (8 or 12 dB), but the trends noted here are reflected in the averages.

As before, a logistic regression mixed effects model was fitted to the participants' response data to test the significance of the trends described above and to probe for more subtle trends involving the interaction of Intensity and Duration. Subject was again treated as a random effect; the other variables were the same, and were coded as they had been for Experiment 1. The model that best fitted the response data for Experiment 2 included the four terms whose parameter estimates, standard errors, z scores and significance levels are presented in Table III. This model provided a significantly better fit for the data than models that included only the intercept ($\chi^2 = 2387{\cdot}3$, $df = 3$, p $(>\chi^2) < 0{\cdot}00001$), only Intensity (without Duration; $\chi^2 = 61{\cdot}938$, $df = 2$, p $(>\chi^2) < 0{\cdot}00001$) or only Duration (without Intensity; $\chi^2 = 2374{\cdot}9$, $df = 2$, p $(>\chi^2) < 0{\cdot}00001$). Neither control variable (Block, Order) contributed significantly to goodness of fit.

	coefficient	SE	z	p $(>\lvert z \rvert)$
intercept	0·022	0·062	0·35	0·728
Intensity	2·113	0·060	35·33	< 0·00001
Duration	−0·228	0·047	−4·88	< 0·00001
Intensity × Duration	−0·346	0·057	−6·03	< 0·00001

Table III

Statistical model for positively coded *dege* responses in Experiment 2.

The fixed effects of Intensity and Duration were both highly significant. Exponentiating the estimated coefficient (2·113) for Intensity to produce an odds ratio revealed that a unit of increase in Intensity raised the odds of a *dege* response by a factor of 8·27. The effect of Duration revealed by the statistical analysis, while also highly significant, was weaker than that of Intensity: the odds ratio for duration was 0·80, signifying that a unit of increase in Duration lowered the odds of a *dege* response by 20% on average. Finally, the interaction Intensity × Duration was highly significant. The odds ratio for the interaction term was 0·70, signifying that the effect of Intensity on the probability of a *dege* response was 30% less as MOD increased. Given this statistic, it is possible that the diminishing effect of Intensity when the intensity disparity was greatest reflects the mitigating influence of Duration. As the odds ratio associated with a variable is a measure of effect size, the difference in the odds ratios for Intensity (odds ratio = 8·27) and Duration (odds ratio = 0·80) indicated that Intensity was a much more robust predictor of Response overall than Duration, consistent with (4a).[18]

4.2.3 *Discussion.* The primary and most consistent finding of Experiment 2 was that varying intensity was a highly robust predictor of loud–soft groupings. Varying duration significantly influenced listeners' grouping patterns as well, although its specific effect differed in the Competing and Cooperating conditions. Finally, although the effect of duration was less robust, increasing duration attenuated the effect of intensity, especially when the intensity disparity was lower.

4.3 Experiment 3

Taken together, the outcomes of Experiments 1 and 2 seemed contradictory: why were the effects of varying intensity and duration consistent and significant in Experiment 2, whereas neither parameter had been a consistent predictor of listeners' grouping preferences in Experiment 1? We considered two explanations. Because most of the subjects in Experiment 2 had also participated in Experiment 1, the more consistent outcomes of Experiment 2 might reflect a learning benefit: the subjects may simply have been better at responding on the second day. An alternative (though not incompatible) explanation was that participants may have been more comfortable with the sequences in which both intensity and duration were varied. Because intensity and duration vary together in natural speech, the covaried sequences might have seemed more realistic than the sequences in which only one parameter was varied. If the second interpretation were correct, we reasoned, then we might

[18] An odds ratio of 1·0 indicates the absence of any effect. In this study, the predictor variable Intensity increases the odds of a particular outcome, while the other predictor variable, Duration, decreases the odds of the same outcome. In this case, an odds ratio of 2·0 for Intensity and an odds ratio of 0·5 for Duration represent equivalent effect sizes.

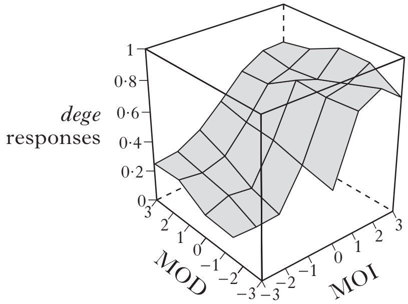

Figure 6

Three-dimensional graph showing the proportion of *dege* responses (*z*-axis) as a function of the interaction of MOI (*x*-axis) and MOD (*y*-axis) in Experiment 3. The duration-only condition is represented by the line bisecting the graph at MOI level 0, and the intensity-only condition is represented by the upward arc described by the four points that define the right edge of the graph at MOD level 0. The Competing condition is represented by all other points at the positive end of the MOI scale, and the Cooperating condition by all points at the negative end of the MOI scale.

find more consistent effects with the intensity-only and duration-only sequences if these were presented to listeners along with the (possibly) more natural covaried sequences. To further explore these questions, Experiment 3 was conducted as a follow-up study. A different group of Zapotec-speaking participants was tested with a stimulus set that included single-parameter sequences from Experiment 1 and covaried sequences from Experiment 2.

4.3.1 *Method*

Subjects and setting. The participants for Experiment 3 were 17 adult speakers of Betaza Zapotec who had not participated in either of the first two experiments. The setting was a private home in Betaza and, as before, participants were paid 100 pesos for their time.

Stimuli. To simplify the design, a reduced set of the test sequences from Experiments 1 and 2 were presented to participants in Experiment 3. This set included only the single-parameter strings from Experiment 1 in which *de* was the louder or the longer syllable (12 sequences), and the sequences in which *de* was the longer syllable in the Competing and Cooperating conditions in Experiment 2 (36 sequences). To further probe the effect of duration, which seemed to be the weaker predictor of listeners' responses in Experiment 2, sequences representing a fourth MOD level were used (an absolute difference of 160 ms between alternating syllables). This added 14 sequences to the test set. Including the two control sequences, a total of 64 sequences were used in Experiment 3.

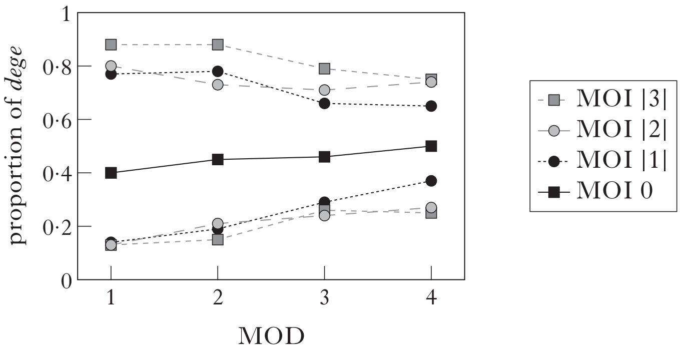

Figure 7
Disaggregated trends associated with MOD (*x*-axis) for different MOI levels (*y*-axis) in Experiment 3. MOI 0: duration only; MOI |1|: 4 dB disparity; MOI |2|: 8 dB disparity; MOI |3|: 12 dB disparity.

Experimental procedure. The stimuli were presented in three blocks of 64 stimuli, with a rest at the midpoint of each block. The response booklets provided to participants were similar to those used in Experiment 2, except that they consisted of six pages, each with 32 numbered blanks (i.e. each block was divided over two pages). The experiment was conducted in Zapotec by a member of the community trained for this purpose, and the software program was managed by the first author.

4.3.2 *Results.* The design and response data for Experiment 3 are presented in Table XII in the online supplementary materials (App. C). The trends in the data are represented by the wire graph in Fig. 6, which represents the interaction of MOI and MOD in three-dimensional space. Overall, the proportion of *dege* responses was positively correlated with MOI in all conditions in which intensity was varied, a trend which again favoured loud–soft groupings. A number of reversals may be seen in the graph, largely at the extremes of the MOI scale. Despite these reversals, the proportion of *dege* responses was still well above chance for all sequences in which there was an intensity disparity.

In order to more clearly represent them, the less robust trends associated with duration at different MOI levels are disaggregated in Fig. 7. In conditions where duration was varied, *de* was always the long syllable and a *dege* response therefore represented a long–short grouping (*deege*, ***dee**ge* or *dee**ge***). In the duration-only condition (the black squares and solid line in Fig. 7), MOI and *dege* responses were positively and monotonically associated, indicating a preference for long–short groupings, the trend counterpredicted by hypothesis (3b). The Cooperating condition is represented by the lines above the solid line, and the Competing condition by the lines below. Intensity had its greatest effect when the duration disparity was smallest (MOD 1), and its smallest effect when the duration disparity was greatest (MOD 4). This is reflected in Fig. 7 by visible differences in the deviations of the lines for the covaried conditions from the solid line representing the duration-only condition. In the Cooperating condition, increasing MOD consistently increased *dege* responses (long–short *dee**ge*** groupings), as was the case in the duration-only condition. In the Competing condition, increasing MOD generally decreased *dege* responses, reflecting an increase in short–long *ge**dee*** groupings. The trends associated with duration in the covaried conditions replicated the findings for the corresponding conditions in Experiment 2.

The procedures used in the statistical analysis of the response data were as described for Experiments 1 and 2. The model that best fitted the response data for the control and intensity-only sequences included the two terms whose parameter estimates, standard error, *z* score and significance levels are reported in Table IV. This model provided a significantly better fit than the model that included only the intercept ($\chi^2 = 44{\cdot}781$, $df = 1$, $p\ (>\chi^2) < 0{\cdot}00001$). Neither of the control variables (Block, Order) significantly improved goodness of fit in this or any other statistical model described in this section. The effect of Intensity was highly significant: on average, the odds of a *dege* response were 2·27 times greater per unit of increase, as compared with the baseline control category.

	coefficient	SE	*z*	*p* (>|*z*|)
intercept	1·064	0·231	4·604	< 0·00001
Intensity	0·819	0·129	6·348	< 0·00001

Table IV

Statistical model for intensity-only sequences in Experiment 3 (403 valid responses).

The output of the best-fitting model for the response data for the control and duration-only sequences is provided in Table V. This model, which included terms for the intercept and the predictor variable Duration, provided a significantly better fit than the model which included only the intercept ($\chi^2 = 5{\cdot}203$, $df = 1$, $p\ (>\chi^2) = 0{\cdot}023$). The solo effect of Duration evident in Fig. 6 was significant. In comparison with the baseline control category, the odds of a *dege* (long–short) response were on average 1·23 times greater per unit of increase in Duration. The effect size associated with Duration was lower than that for Intensity.

Finally, we tested for interactions between Intensity and Duration in the Competing and Cooperating conditions by fitting a logistic regression model to the full data set that included the response data set for these and for the control sequences. Table VI presents the output of

	coefficient	SE	*z*	*p* (>|*z*|)
intercept	−0·272	0·130	−2·091	0·037
Duration	0·210	0·092	2·289	0·022

Table V
Statistical model for duration-only sequences in Experiment 3 (505 valid responses).

the best-fitting model, which provided a significantly better fit than either the intercept-only model ($\chi^2 = 778{\cdot}09$, $df = 3$, $p\ (>\chi^2) < 0{\cdot}00001$) or a model in which Intensity and Duration were included with no interaction term ($\chi^2 = 35{\cdot}911$, $df = 1$, $p\ (>\chi^2) < 0{\cdot}00001$).

	coefficient	SE	*z*	*p* (>|*z*|)
intercept	−0·093	0·082	−1·129	0·259
Intensity	1·310	0·055	23·810	< 0·00001
Duration	0·103	0·046	2·229	0·0258
Intensity × Duration	−0·339	0·057	−5·906	< 0·00001

Table VI
Statistical model for Competing and Cooperating sequences in Experiment 3 (2512 valid responses).

Table VI reveals that the effect of Duration was significant and that the effect of Intensity was highly significant. The interaction Intensity × Duration was also highly significant. The odds ratios associated with the term Intensity revealed that overall, a unit of increase in Intensity improved the odds of a *dege* response by a factor of 3·71. A *dege* (long–short) responses was 1·11 times more likely per unit increase in Duration, again indicating a weaker trend, consistent with the hypothesis in (4a). The odds ratio for the interaction term, 0·71, reveals that the effect of Intensity on the probability of a *dege* response was 29 % less as magnitude of duration increased.

4.3.3 *Discussion.* Although Experiment 1 produced no support for clear grouping preferences when listeners were tested with sequences in which only a single parameter (either intensity or duration) was varied, clear preferences were found in Experiment 3: varied intensity was associated with a loud–soft and varied duration with a long–short grouping preference. The difference was that in Experiment 3 the single-parameter strings were intermingled with more complex covaried sequences that may have seemed more natural to listeners. In this context, we suggest that listeners' greater comfort with the covaried sequences may have transferred to the arguably less natural single-parameter sequences. In other respects, the findings of Experiment 3 broadly replicated those of Experiment 2. As Experiment 3 was conducted with new listeners who had not participated in an earlier study, it does not seem that the significant outcomes of Experiment 2, in contrast to Experiment 1, can be explained as an effect of learning. As a final point, the results of Experiment 3 confirmed the grouping biases found in Experiment 2, in which increasing MOD was associated with a short–long grouping trend in the Competing condition but with a long–short grouping trend in the Cooperating condition. We return to this point in the general discussion.

5 Experiment 4: a parallel test with English-speaking listeners

Experiment 4 repeated Experiment 3 with English-speaking listeners, in order to provide a cross-linguistic comparison. English speakers were used because they are readily available for use in perception experiments and because English-speaking listeners' grouping preferences with intensity-varying and duration-varying sequences have been studied before (Hay & Diehl 2007). To the extent that the response patterns of the English speakers differed from those of the Zapotec-speaking participants of Experiments 2 and 3, we would have evidence that language background influenced the listeners' grouping biases.

5.1 Methods

Thirty English-speaking undergraduates at the University of Texas at Austin participated in Experiment 4, and were paid $10 for their time. None reported special fluency in, or having been exposed to, a language other than English at home. Participants were tested in small groups in a phonetics laboratory at the University of Texas at Austin, where they heard the alternating sequences over a loudspeaker system (the methodology used in Experiments 1–3). The stimulus set for Experiment 4 was the full set of sequences used in Experiment 3. All phases of the study were conducted in English by an English-speaking research assistant. In all other crucial respects, the procedures followed in Experiment 4 were as described for Experiment 3.

5.2 Results

The design and response data for Experiment 4 are presented in Table XIII in the online supplementary materials (App. C). As before,

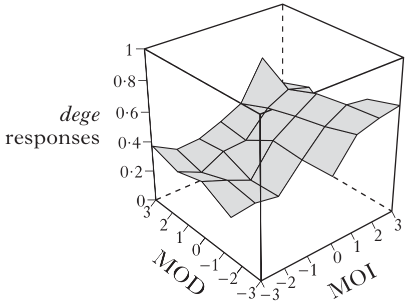

Figure 8

Three-dimensional graph showing the proportion of *dege* responses (*z*-axis) as a function of the interaction of MO I (*x*-axis) and MOD (*y*-axis) in Experiment 4. The organisation is as for Fig. 6 above.

the response data are represented by a wire graph (Fig. 8) as the proportion of *dege* responses for each cell in the design. Overall, although there are similarities, the outcomes of Experiment 4 are less robust than the findings of Experiments 2 and 3. In the intensity-only condition (the four points that describe an upward arc at MOD level 0), the correlation between *dege* responses and MOI was positively monotonic, a trend which again favoured loud–soft groupings. In the duration-only condition (the line bisecting the graph at MOI level 0), increasing MOD decreased *dege* responses. This trend, which favoured short–long groupings, differs from the trend found for the Zapotec speakers in Experiment 3, but replicates the finding for English-speaking listeners reported in Hay & Diehl (2007). In Experiment 3, with Zapotec speakers, there was a tendency for the generally positive correlation between MOI and *dege* responses to be reversed at the extremes of the MOI scale (a trend which was not prominent in Experiment 2). This effect was even more pronounced in the Competing condition (positive values on the MOI scale) in Experiment 4, and was present to some extent in the Cooperating condition (negative MOI values).

Trends associated with duration were more variable. As for Experiment 3, the disaggregated line graphs in Fig. 9 (based on Fig. 8) show how the proportion of *dege* responses varies as a function of MOD (horizontal) at different MOI levels (vertical). Again, the most consistent trends for duration were found at MOI levels −1 and 1, when the intensity disparity was least. These trends are consistent with the corresponding trends found in Experiments 2 and 3: increasing MOD increased *dege* responses in the Cooperating condition (lines below the solid line, which represents the duration-only condition), and decreased *dege* responses in the Competing condition (lines above the solid line). The latter trend also occurred at MOI level 2, except for the trend reversal seen at MOD level 4). Beyond this, no clear pattern was apparent.

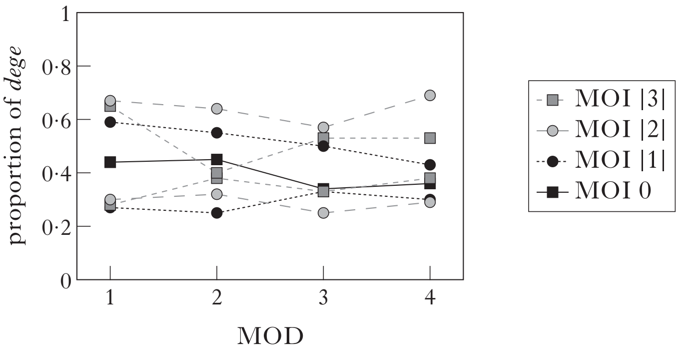

Figure 9

Disaggregated trends associated with MOD (*x*-axis) for different MOI levels (*y*-axis) in Experiment 4. MOI 0: duration only; MOI |1|: 4 dB disparity; MOI |2|: 8 dB disparity; MOI |3|: 12 dB disparity.

The statistical tests and variable coding used in Experiment 4 were carried out as described for Experiment 3. The output of the best-fitting statistical model for the response data for the control and intensity-only sequences appears in Table VII. This model provided a significantly better fit than the model which included only the intercept ($\chi^2 = 95{\cdot}279$, $df = 1$, $p\ (>\chi^2) < 0{\cdot}00001$). No other variable contributed significantly to the model. The effect of Intensity was highly significant, and the odds ratio indicated that a *dege* response was 2·32 times more likely per unit of increase in Intensity as compared with the baseline control category.

	coefficient	SE	z	$p\ (>\|z\|)$
intercept	0·824	0·227	3·624	0·0003
Intensity	0·842	0·093	9·064	< 0·00001

Table VII

Statistical model for intensity-only sequences in Experiment 4 (895 valid responses).

The output of the model that best fitted the response data for the control and duration only sequences is shown in Table VIII. This model provided a significantly better fit than the model that included only the intercept ($\chi^2 = 7{\cdot}509$, $df = 1$, $p\ (>\chi^2) = 0{\cdot}006$), and a better fit than a model that included any other variable. The negative trend associated with Duration was significant: a *dege* response was only 0·84 times as likely per unit of increase in Duration as compared with the control category. Again using the odds ratio as a measure of effect size, we see that Intensity was a more

	coefficient	SE	z	p (>|z|)
intercept	−0·246	0·130	−1·889	0·059
Duration	−0·176	0·064	−2·758	0·006

Table VIII

Statistical model for duration-only sequences in Experiment 4 (1075 valid responses).

robust predictor of response than Duration for our English-speaking participants, as was the case for the Zapotec-speaking participants in Experiments 2 and 3.

The final analysis was carried out with the response data for the control and the covaried sequences. The output of the best-fitting model is presented in Table IX. This model provided a significantly better fit than both the intercept-only model ($\chi^2 = 625{\cdot}07$, $df = 3$, $p\ (>\chi^2) < 0{\cdot}00001$) and a model that included the intercept and either Intensity ($\chi^2 = 39{\cdot}282$, $df = 2$, $p\ (>\chi^2) < 0{\cdot}00001$) or Duration ($\chi^2 = 612{\cdot}09$, $df = 2$, $p\ (>\chi^2) < 0{\cdot}00001$) alone. No other variable significantly improved goodness of fit.

	coefficient	SE	z	p (>|z|)
intercept	−0·306	0.117	−2·613	0·009
Intensity	0·816	0.035	23·153	< 0·00001
Duration	−0·116	0.032	−3·618	0·0003
Intensity × Duration	−0·187	0.039	−4·848	< 0·00001

Table IX

Statistical model for Competing and Cooperating response data in Experiment 4 (4659 valid responses).

Table IX reveals that the main effects of Intensity and Duration were both highly significant, and that the interaction Intensity × Duration was also highly significant. Overall, a unit of increase in Intensity raised the odds of a *dege* response by a factor of 2·26, whereas for Duration, the odds of a *dege* (long–short) response were 11 % less (odds ratio = 0·89) per unit of increase. Finally, the odds ratio of 0·83 for the interaction term indicates that as MOD increased, the effect of Intensity on the probability of a *dege* response was diminished by 17 %. As for the Zapotec-speaking listeners in Experiments 2 and 3, Intensity was a stronger predictor of Response than Duration for the English speakers tested, although the effect sizes were smaller for the English speakers.

5.3 Discussion

The outcomes of Experiment 4 differed from those of Experiment 3 in that the English speakers showed a short–long grouping pattern in the duration-only condition, as predicted by the ITL, in contrast to the long–short pattern found for the Zapotec speakers. In other respects, the results of Experiment 4 were broadly compatible with those of Experiment 3. The English speakers showed a consistent loud–soft grouping preference in the intensity-only condition, and somewhat less consistently in the covaried condition, where there were trend reversals at the extremes of the MOI scale. The results associated with duration in the Competing and Cooperating conditions were broadly compatible with the comparable results of Experiments 2 and 3 (more short–long groupings for the Competing sequences and more long–short groupings for the Cooperating sequences). The results associated with duration in the co-varied conditions were less consistent for the English speakers than for the Zapotec speakers, however, and the trends found with the English speakers were generally weaker. One interpretation of the weaker trends found in Experiment 4 might be that Zapotec 'voice' was novel to the participants.

6 General discussion of experimental outcomes and directions for future research

The results of our experiments confirm findings reported in the previous literature in some respects, and in other respects add to what we know about factors that help human listeners to segment streams of syllables into bisyllabic chunks. The loud–soft and short–long grouping biases found for the English speaking listeners are consistent with the ITL and with results reported for English speakers by Hay & Diehl (2007) and Crowhurst (2013).

We can report several novel findings based on Experiments 2–4. First, the loud–soft preference of the Zapotec speakers for sequences in which only intensity was varied is consistent with the ITL and with results reported for English, French, German and Spanish (Hay & Diehl 2007, Bhatara *et al*. 2013, Crowhurst 2013). One of the most interesting findings was that the Zapotec-speaking participants in Experiment 3 showed a clear preference for long–short groupings in the condition where only duration was varied. To our knowledge, this result provides the first clear exception to the Duration Principle of the ITL among studies that have tested subjects using alternating non-speech or speech-like sequences. A potential explanation for the Zapotec speakers' long–short preference in the duration-only condition may have to do with the distribution of stress in Betaza Zapotec and the language-specific relationship between stress and vowel duration in sequences that have certain properties in common with the syllables used in our studies. No detailed study of the perceptual correlates of stress among Zapotec speakers exists. Nor is extensive information available about the physical correlates of stress in the productions of Zapotec speakers, beyond that provided by Chávez-Peón (2008, 2010) and informal measurements based on our own elicited field data. However, the literature on Xhon varieties of Zapotec shows that stressed vowels are longer than unstressed vowels before lenis consonants (see references cited in § 3), and our fieldwork confirms that this is also the case for Betaza Zapotec.[19] It is also clear from our field studies that word stress falls on root-initial syllables in Betaza Zapotec. Our Zapotec-speaking participants' long–short grouping pattern would make sense if they were processing recurrent *dege* or *gede* syllable pairs as they would bisyllabic unaffixed words. Therefore our finding may reflect the influence of prosodic characteristics of Betaza Zapotec on speakers' perceptions of grouping affinity, but we can make no specific claims as to whether these characteristics relate to units of stress or rather to word structure.

The other novel finding produced by this series of studies is that intensity was a more robust predictor of grouping choices than duration for all groups of participants, both Zapotec and English-speaking. We are unable to fully evaluate this outcome in the context of the existing grouping literature, because to our knowledge there are no reports of studies that have explored the interaction of intensity and duration in the manner we have described. Any interpretation of the dominant effect of intensity would also be limited, in that two of the procedures used in manipulating intensity may have had unknown consequences, which should be explored in future work. The first is that we manipulated intensity by changing overall gain. The effect of varying intensity in this way may be compared to the effect of adjusting the volume setting on a radio. This is quite different from how syllables of different loudnesses sound in natural speech, in which intensity variations are associated with differences in vocal effort.[20] Follow-up research might usefully study the effect of varying intensity in these different ways.

The second issue was that the maximum intensity disparity of 12 dB we used is much larger than the difference normally found between any two adjacent syllables in naturally occurring speech, and this may have contributed to the attenuation or even reversal of loud–soft grouping trends at the edges of the MOI scale. It is possible that listeners' performance (as reflected by trend reversals in particular) might be linked to their sensitivity to relationships among phonetic properties that tend to characterise adjacent syllable pairs in the native language. The point at which the loud–soft trend tended to reverse may have corresponded to a threshold between more and less likely bisyllabic units. The same point could be made about the absence of consistent results in Experiment 1: it might

[19] Stressed vowels are not longer than unstressed vowels before fortis consonants, where the increased length associated with stress is associated with the following consonant. An interesting line of follow-up research might explore whether Zapotec-speaking listeners' subjective grouping preferences differ depending on this variable.

[20] We thank Mark Liberman and Björn Lindblom for pointing this out to us.

have been the case that if the sequences in which only one parameter was varied seemed unrealistic to participants, they might not have known how to respond to them. While this series of studies has been a beginning, a better understanding of what listeners might be attending to would require specific information about the ways in which features present in speakers' productions might influence their decisions in speech-perception tasks. Future research might include a production study, whose results would then be used to design a more finely tuned perception study that asks questions similar to those we tested.[21] Even if increased intensity turns out not to be a reliable physical correlate of stress in Betaza Zapotec, we suspect that it might play a role in demarcating the beginnings of words. This would be consistent with our speculation that the Zapotec speakers who took part in our study were processing recurrent syllable pairs as short words. If increased intensity does turn out to signal word beginnings, this would be consistent with a morphological and not a prosodic origin for the loud–soft grouping asymmetry we have found. Given that this asymmetry has also been found with speakers of other languages, researchers might consider the possibility of a demarcative role for intensity in those cases as well.[22]

Disparities in duration were also associated with clear grouping preferences, although these differed depending on the specific condition. In the Competing condition, increasing the duration disparity increased responses representing the short–long groupings predicted by the ITL. In contrast, duration disparities in the Cooperating condition were associated with increases in the counterpredicted long–short groupings for both English and Zapotec-speaking subjects, an outcome that seems incongruent with the result in the Competing condition. The long–short trend in the Cooperating condition might not be surprising, given how the auditory system integrates intensity over time (temporal summation): listeners tend to perceive longer sounds as louder relative to shorter sounds of the same intensity (Gordon 2005 and references cited there). If increasing duration in the Cooperating condition had the effect of increasing the perceived loudness of the longer syllables for our Zapotec-speaking study participants, this might have reduced the effect of increasing the actual intensity of the shorter syllables. At this point we do not see a way of reconciling the outcomes associated with duration in the covaried conditions: an ITL explanation does not account for the effect found in the Cooperating condition, and a temporal summation explanation does not seem compatible with the result in the Competing condition.

[21] We thank Matthew Gordon for pointing out the issues discussed above in this paragraph and for suggesting this line of investigation.

[22] Caroline Smith has commented that greater intensity is associated with word beginnings in French, and that this might provide a partial explanation for the loud–soft result reported for French-speaking listeners in previous grouping studies (i.e. Hay & Diehl 2007, Bhatara *et al.* 2013).

A temporal summation explanation could potentially explain the long–short grouping pattern found for the Zapotec participants in Experiment 3.[23] That the English speakers did not display the long–short bias in the duration condition might have had to do with different sensitivities to the type and degree of acoustic disparities between alternating syllables in our sequences. In particular, the Zapotec speakers may have been more sensitive to perceptual effects associated with temporal summation. Bhatara *et al.* (2013) discuss the issue of differing sensitivities to acoustic disparities in connection with the response patterns of their French and German-speaking subjects in another basic test of the ITL. These authors found that the grouping behaviour demonstrated for both groups indicated the same preferences (loud–soft and short–long pairings). However, the effect size associated with duration in particular was larger for the German than the French-speaking group, suggesting that the German speakers may have been more sensitive to duration cues. A follow-up study might usefully test the influence of duration on grouping asymmetries in a more nuanced way: instead of controlling for average acoustic intensity between alternating syllables, the preparation of duration-varying syllables could be guided by listeners' perceptions of loudness when asked to evaluate CV syllables with vowels of different durations.

7 Implications for phonology

The finding that different grouping patterns were associated with duration for our English and Zapotec-speaking participants indicates that ITL effects can vary depending on listeners' language backgrounds and that phonologists should be wary of assuming a natural connection between the ITL and generalisations about prosodic organisation. That the differences in listeners' grouping preferences were associated with duration is particularly significant. Formally, the foot inventory in (1) may be understood as an assertion about how quantity-sensitivity and quantitative symmetry shape prosodic categories. The syllabic and moraic trochees are constrained by symmetry; what distinguishes them is sensitivity to syllable weight. Iambic form is characteristic of stress systems in which rhythm is quantity-sensitive and optimally alternates at the level of the syllable. Within this formal system, trochaic rhythm seems to be viewed as a default typological option for systems lacking one or both of these requirements.[24] The canonical unit of syllable weight is phonological

[23] A temporal summation explanation would not be inconsistent with the language-specific explanation suggested above.

[24] Support which has been claimed for the uneven iamb as a prosodic category rhythm is the existence of metrical systems in which rhythm is iambic and vowels in foot-final syllables are phonologically lengthened, producing heavy syllables (iambic lengthening (Hayes 1995, Buckley 1998), as in Hixkaryana (Derbyshire 1985) and Tiriyó (Meira 1998), for example). The evidence for uneven iambs from prosodic

vowel length, which is naturally associated with increased phonetic duration. According to (1), when asymmetric groupings are allowed, heavy syllables stand at the right edges of constituents. As far as the foot typology is concerned, therefore, the appeal of the ITL may be that it seems to suggest that the typology in (1) puts the heavy syllable in the right place. However, a natural connection between the ITL and (1) is less clear if it turns out that duration is not as strongly connected to grouping choices as the ITL implies. This paper has provided two types of evidence against such a connection. The first is the finding that speakers of at least one language may prefer the counterpredicted long–short grouping pattern. A better understanding of how common the long–short preference is will depend on having solid results from ITL studies conducted with speakers of other languages, especially non-European ones. The finding that intensity was a stronger predictor of grouping choices than duration (and by a large margin) provides further reason for scepticism, as this finding is poorly aligned with the dominant role that proponents of (1) attribute to syllable quantity as a crucial factor underlying the foot typology. On the view that trochaic structure is merely a default in the absence of quantitative asymmetry (typologically speaking), the dominant influence of intensity on grouping preferences for sequences in which duration asymmetries are also present should be unexpected.

One of our broader goals is to show that being able to perceive natural groupings does not necessarily depend on being able to locate stressed syllables. This is not surprising, given that not all feet are rhythmic in nature (see note 24). Rhythmic and non-rhythmic feet may be present in the same forms. As an example, Crowhurst (1994) argues that in Cupeño, a language with non-iterative stress, an unstressed, prosodically circumscribed bisyllabic foot is positioned immediately after the stress foot in the habilitative construction (Hill 1970). To satisfy the stated objective, it must be demonstrated that cues that influence listeners' perceptions of stress and of grouping are different, or are at least used differently, depending on the task. The finding that intensity trumps duration as the dominant predictor of grouping choices is suggestive in this regard for the English speakers at least. The English stress foot is a quantity-sensitive trochee (e.g. Selkirk 1980, Hayes 1982). English speakers have been shown to rely primarily on duration and pitch in identifying stressed syllables, and less on intensity (see references cited above). Therefore, if English speakers use their knowledge of the properties of stressed syllables in making grouping choices, we might not expect intensity cues to be especially salient. Rather, given that greater phonetic duration characterises all stressed vowels (whether in light or heavy syllables), we might

---

morphology is more limited than the evidence for bimoraic and bisyllabic feet. Uneven iambs are associated with prosodic shape requirements on morphological units in languages with non-concatenative morphologies, such as Arabic (see McCarthy & Prince 1990). However, the iambic foot does not seem to occur as a unit in reduplicative systems, which provide some of the strongest extrarhythmic evidence for foot structure.

expect English speakers to prefer long–short, not short–long, groupings. In fact, the short–long grouping preference repeatedly found for English speakers suggests that they use duration demarcatively. As a cue to demarcation, duration may well be a precursor to iambic foot structure, but not a foot in which the long syllable is necessarily stressed, as claimed by (1).

This is not the first work to question the assumption that generalisations about foot structure can be naturally derived from the Iambic-Trochaic Law. Nor has the foot typology itself been universally accepted among phonologists. The problem for phonologists has been the claimed role of syllable weight as it relates to asymmetric feet. For Kager (1991, 1993), there are no weight asymmetries in the basic foot inventory. Noting that moraic iambs are common (e.g. Osage and Southern Paiute, mentioned in §1), Kager argues for a basic inventory of balanced bimoraic or bisyllabic feet with either trochaic or iambic headedness. In Kager's theory, these basic symmetrical feet are mapped to surface feet in a process that is guided by the requirements of prosodic layering and by rhythmic principles that disfavour stress clashes and lapses. Kager recognises a surface asymmetry insofar as the trochaic inventory is concerned – asymmetric iambs clearly represent a more common prosodic category than asymmetric trochees. In his view, this is due to a rhythmic principle that favours metrical constituents ending in a strong–weak sequence. Adding a weak mora to a stressed CV syllable optimises an iambic foot, but decreases the goodness of a trochaic foot. Revithiadou (2004) argues that the ITL makes the wrong predictions for trochaic systems. She claims that quantitative adjustments to stressed and unstressed vowels that should be restricted to iambic systems, according to (1), are common in trochaic systems as well. Revithiadou's approach to rhythmic structure de-emphasises the ITL and focuses instead on positional constraints. She favours a foot inventory that includes uneven trochees (2004: 59), although she does not fully explore the consequences of this modification.

It may be that the typology in (1) is not sufficiently representative. In particular, the distribution of iterative systems based on the asymmetric iamb seems to be geographically skewed – the best-known cases are found in languages of the Americas. However, the existence of exceptions does not falsify the typology (under the weaker interpretation), although it may be an indication that any claims about the formal properties of feet should be more rigorously supported by quantitative information. A weakness of the various foot typologies that have been put forward (e.g. Hayes 1985, 1987, 1995, Halle & Vergnaud 1987, McCarthy & Prince 1996) is that they have relied for evidence on non-native impressions of stress, often in underdocumented languages for which no corroborating reports may be available. Abercrombie (1967: 97, cited in Fletcher 2010: 558) observes that 'in order to have this immediate and intuitive apprehension of speech rhythm it is necessary … that speaker and hearer should have the same mother-tongue'. Fletcher (2010: 558) observes that 'judgments of stress or prominence can be difficult when confronted with an unfamiliar or less well studied language, and native-speaker influences or intuitions can result in premature categorization of a language into one rhythm class or another'.[25]

Although labels such as 'law' and 'principle' have been applied to the ITL in the literature (including this paper, for convenience), the statements in (2) are in fact sparsely tested generalisations about perceptual tendencies. As such, they are not claims about organising principles that may determine prosodic structure in the synchronic sound grammars of languages. Based on our study, we now know that it is not the case that speakers of all languages have an iambic short–long bias, when duration as a cue to grouping is tested alone. The fact that this is possible is an indication that asymmetric long–short units may be perceptually quite natural, in which case we might expect this type of unit to be grammaticalised in some languages. What we don't know is how widespread the tendency might be. In order to posit a connection between a perceptual tendency and natural units of prosody cross-linguistic research on grouping preferences is urgently needed from studies using both speech-like *and* non-speech stimuli. First we need to know how general the tendency is, independently of language. An accumulation of information from studies with non-speech stimuli can add to our knowledge of this issue. The two non-speech studies we know of that have been done with a non-European language (Kusumoto & Moreton 1997, Iversen *et al*. 2008) produced mixed results for duration with groups of Japanese speakers. Both studies found a strong short–long grouping bias for some Japanese speakers and a strong long–short bias for others. This shows that the short–long asymmetry stated in (2) is by no means universal, but more work needs to be done, especially with speakers of non-European languages, which are scarcely represented in the grouping literature. If, on balance, information from non-speech studies supports generalisations, perhaps those stated in the Iambic-Trochaic Law, about how humans perceive rhythm when different cues are varied, similar studies using more speech-like stimuli supply information about the extent to which whether humans' subjective grouping biases, if any, can be modified by linguistic experience.

REFERENCES

Abercrombie, David (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Adams, Corinne & R. R. Munro (1978). In search of the acoustic correlates of stress: fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English. *Phonetica* **35**. 125–156.

[25] We thank one of the anonymous reviewers in particular for the specific point and two quotes provided in the preceding three sentences. Although Fletcher's 'rhythm class' in this quote refers to the distinction between syllable-timed and stress-timed languages, a categorisation we have mentioned only in passing, her point can be generalised.

Altshuler, Daniel (2009). Quantity-insensitive iambs in Osage. *IJAL* **75**. 365–398.

Avelino, Heriberto (2004). *Topics in Yalálag Zapotec, with particular reference to its phonetic structures*. PhD dissertation, University of California, Los Angeles.

Beckman, Mary E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.

Beckman, Mary E. & Jan Edwards (1994). Articulatory evidence for differentiating stress categories. In Patricia A. Keating (ed.) *Phonological structure and phonetic form: papers in laboratory phonology III*. Cambridge: Cambridge University Press. 7–33.

Bell, Alan (1977). Accent placement and perception of prominence in rhythmic structures. In Larry M. Hyman (ed.) *Studies in stress and accent*. Los Angeles: Department of Linguistics, University of Southern California. 1–13.

Bhatara, Anjali, Natalie Boll-Avetisyan, Annika Unger, Thierry Nazzi & Barbara Höhle (2013). Native language affects rhythmic grouping of speech. *JASA* **134**. 3828–3843.

Bion, Ricardo A. H., Silvia Benavides-Varela & Marina Nespor (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech* **54**. 123–140.

Boersma, Paul & David Weenink (2011). *Praat: doing phonetics by computer* (version 5.2.17). http://www.praat.org.

Bolton, Thaddeus L. (1894). Rhythm. *American Journal of Psychology* **6**. 145–238.

Buckley, Eugene (1998). Iambic lengthening and final vowels. *IJAL* **64**. 179–223.

Butler, Inez M. (1980). *Gramática zapoteca: zapoteco de Yatzachi el Bajo*. México, D.F.: Instituto Lingüístico de Verano.

Byrd, Dani (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* **57**. 3–16.

Chávez-Peón, Mario E. (2008). Phonetic cues to stress in a tonal language: prosodic prominence in San Lucas Quiaviní Zapotec. *Proceedings of the 2008 Annual Conference of the Canadian Linguistic Association*. Available (January 2014) at http://homes.chass.utoronto.ca/~cla-acl/actes2008/CLA2008_Chavez-Peon.pdf.

Chávez-Peón, Mario (2010). *The interaction of metrical structure, tone, and phonation types in Quiaviní Zapotec*. PhD dissertation, University of British Columbia.

Crowhurst, Megan J. (1991). *Minimality and foot structure in metrical phonology and prosodic morphology*. PhD dissertation, University of Arizona, Tucson.

Crowhurst, Megan J. (1994). Foot extrametricality and template mapping in Cupeño. *NLLT* **12**. 177–201.

Crowhurst, Megan J. (2013). Perceptual influences on rhythmic grouping preferences. Paper presented at M@90: a workshop on metrical structure in honor of Morris Halle's 90th birthday, MIT.

Cutler, Anne & Sally Butterfield (1992). Rhythmic cues to speech segmentation: evidence from juncture misperception. *Journal of Memory and Language* **31**. 218–236.

Cutler, Anne & David M. Carter (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* **2**. 133–142.

Cutler, Anne & Dennis Norris (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* **14**. 113–121.

de Lacy, Paul (2002). The interaction of tone and stress in Optimality Theory. *Phonology* **19**. 1–32.

Derbyshire, Desmond C. (1985). *Hixkaryana and linguistic typology*. Dallas: Summer Institute of Linguistics & University of Texas at Arlington.

Demuth, Katherine (1996). Alignment, stress, and parsing in early phonological words. In Barbara H. Bernhardt, John Gilbert & David Ingram (eds.) *Proceedings of the UBC International Conference on Phonological Acquisition*. Somerville, Mass.: Cascadilla. 113–125.

Earl, Robert (1968). Rincon Zapotec clauses. *IJAL* **34**. 269–274.
Echols, Catharine H. (1993). A perceptually-based model of children's earliest productions. *Cognition* **46**. 245–296.
Echols, Catharine H. & Megan J. Crowhurst (1998). Developing knowledge of metrical rhythm in infancy. *CLS* **34:2**. 207–212.
Echols, Catharine H., Megan J. Crowhurst & Jane B. Childers (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language* **36**. 202–225.
Fletcher, Janet (2010). The prosody of speech: timing and rhythm. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.) *The handbook of phonetic sciences*. 2nd edn. Malden, Mass.: Wiley-Blackwell. 523–602.
Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *JASA* **27**. 765–768.
Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech* **1**. 126–152.
Fry, D. B. (1965). The dependence of stress judgments on vowel formant structure. In Eberhard Zwirner & Wolfgange Bethge (eds.) *Proceedings of the 5th International Congress of Phonetic Sciences, Münster, 1964*. Basel & New York: Karger. 306–311.
Gerken, LouAnn (1991). The metrical basis of children's subjectless sentences. *Journal of Memory and Language* **30**. 431–451.
Gerken, LouAnn (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language* **21**. 565–584.
Gordon, Matthew (2002). A phonetically driven account of syllable weight. *Lg* **78**. 51–80.
Gordon, Matthew (2005). A perceptually-driven account of onset-sensitive stress. *NLLT* **23**. 595–653.
Grabe, Esther & Ee Ling Low (2002). Durational variability in speech and the Rhythm Class Hypothesis. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory phonology 7*. Berlin & New York: Mouton de Gruyter. 515–546.
Gussenhoven, Carlos & A. C. M. Rietveld (1992). Intonation contours, prosodic structure, and preboundary lengthening. *JPh* **20**. 283–303.
Halle, Morris & Jean-Roger Vergnaud (1987). *An essay on stress*. Cambridge, Mass.: MIT Press.
Handel, Stephen (1989). *Listening: an introduction to the perception of auditory events*. Cambridge, Mass.: MIT Press.
Hay, Jessica S. F. & Randy L. Diehl (2007). Perception of rhythmic grouping: testing the iambic/trochaic law. *Perception and Psychophysics* **69**. 113–122.
Hayes, Bruce (1982). Extrametricality and English stress. *LI* **13**. 227–276.
Hayes, Bruce (1985). *A metrical theory of stress rules*. New York: Garland.
Hayes, Bruce (1987). A revised parametric metrical theory. *NELS* **17**. 274–289.
Hayes, Bruce (1989). Compensatory lengthening in moraic phonology. *LI* **20**. 253–306.
Hayes, Bruce (1995). *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.
Hill, Jane H. (1970). A peeking rule in Cupeño. *LI* **1**. 534–539.
Hyman, Larry M. (1985). *A theory of phonological weight*. Dordrecht: Foris.
Iversen, John R., Aniruddh D. Patel & Kengo Ohgushi (2008). Perception of rhythmic grouping depends on auditory experience. *JASA* **124**. 2263–2271.
Jaeger, Jeri J. (1983). The fortis/lenis question: evidence from Zapotec and Jawoñ. *JPh* **11**. 177–189.
Jusczyk, Peter W., Anne Cutler & Nancy J. Redanz (1993). Infants' preference for the predominant stress patterns of English words. *Child Development* **64**. 675–687.
Kager, René (1989). *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

Kager, René (1991). The moraic iamb. *CLS* **27:1**. 291–305.
Kager, René (1993). Alternatives to the iambic-trochaic law. *NLLT* **11**. 381–432.
Kelly, Niamh, Megan J. Crowhurst & Crystal Cobb (2014). The effect of duration and glottalization on the perception of rhythm among English speakers. Paper presented at the 40th Annual Meeting of the Berkeley Linguistics Society.
Kusumoto, Kiyomi & Elliott Moreton (1997). Native language determines parsing of nonlinguistic rhythmic stimuli. *JASA* **102**. 3204.
Leal, Mary (1950). Patterns of tone substitution in Zapotec morphology. *IJAL* **16**. 132–136.
Leander, Anita (2008). *Acoustic correlates of fortis/lenis in San Francisco Ozolotepec Zapotec*. MA thesis, University of North Dakota.
Lehiste, Ilse (1970). *Suprasegmentals*. Cambridge, Mass.: MIT Press.
Long, Rebecca & Sofronio Cruz (1999). *Diccionario zapoteco de San Bartolomé Zoogocho, Oaxaca*. México, D.F.: Instituto Lingüístico de Verano.
McCarthy, John J. (1984). Prosodic organization in morphology. In Mark Aronoff & Richard T. Oehrle (eds.) *Language sound structure*. Cambridge, Mass.: MIT Press. 299–317.
McCarthy, John J. & Alan Prince (1990). Foot and word in prosodic morphology: the Arabic broken plural. *NLLT* **8**. 209–283.
McCarthy, John J. & Alan Prince (1996). *Prosodic morphology 1986*. Ms, University of Massachusetts, Amherst & Brandeis University.
Meira, Sérgio (1998). Rhythmic stress in Tiriyó (Cariban). *IJAL* **64**. 352–378.
Morgan, James L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language* **35**. 666–688.
Morgan, Jane L., Stephanie Edwards & Linda R. Wheeldon (2014). The relationship between language production and verbal STM: the role of stress grouping. *Quarterly Journal of Experimental Psychology* **67**. 220–246.
Nellis, Donald G. & Barbara E. Hollenbach (1980). Fortis versus lenis in Cajonos Zapotec phonology. *IJAL* **46**. 92–105.
Nespor, Marina & Irene Vogel (1986). *Prosodic phonology*. Dordrecht: Foris.
Ortega-Llebaria, Marta, Maria del Mar Vanrell & Pilar Prieto (2010). Catalan speakers' perception of word stress in unaccented contexts. *JASA* **127**. 462–471.
Pankratz, Leo & Eunice V. Pike (1967). Phonology and morphotonemics of Ayutla Mixtec. *IJAL* **33**. 287–299.
Patel, Aniruddh D. (2008). *Music, language, and the brain*. Oxford: Oxford University Press.
Pike, Eunice V. (1948). Problems in Zapotec tone analysis. *IJAL* **14**. 161–170.
Potisuk, Siripong, Jack Gandour & Mary P. Harper (1996). Acoustic correlates of stress in Thai. *Phonetica* **53**. 200–220.
Prince, Alan (1983). Relating to the grid. *LI* **14**. 19–100.
Pye, Clifton (1983). Mayan telegraphese: intonational determinants of inflectional development of Quiché Mayan. *Lg* **59**. 583–604.
R Development Core Team (2012). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at http://www.r-project.org.
Ramus, Franck, Marina Nespor & Jacques Mehler (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* **73**. 265–292.
Revithiadou, Anthi (2004). The Iambic/Trochaic Law revisited: lengthening and shortening in trochaic systems. In Boban Arsenijevic, Noureddine Elouazizi, Martin Salzmann & Mark de Vos (eds.) *Leiden Papers in Linguistics* 1. Leiden: University of Leiden. 37–62.
Riad, Tomas (1992). *Structures in Germanic prosody: a diachronic study with special reference to the Nordic languages*. PhD dissertation, Stockholm University.

Rice, Curt (1992). *Binarity and ternarity in metrical theory : parametric extensions*. PhD dissertation, University of Texas at Austin.

Sapir, Edward (1930). Southern Paiute, a Shoshonean language. *Proceedings of the American Academy of Arts and Sciences* **65**. 1–296.

Selkirk, Elisabeth (1980). The role of prosodic categories in English word stress. *LI* **11**. 563–605.

Selkirk, Elisabeth (1984). *Phonology and syntax : the relation between sound and structure*. Cambridge, Mass : MIT Press.

Sluijter, Agaath M. C. & Vincent J. van Heuven (1996). Spectral balance as an acoustic correlate of linguistic stress. *JASA* **100**. 2471–2485.

Sluijter, Agaath M. C., Vincent J. van Heuven & Jos J. A. Pacilly (1997). Spectral balance as a cue in the perception of linguistic stress. *JASA* **101**. 503–513.

Sonnenschein, Aaron Huey (2004). *A descriptive grammar of San Bartolomé Zoogocho Zapotec*. PhD dissertation, University of Southern California.

Teodocio Olivares, Amador (2009). *Betaza Zapotec phonology : segmental and suprasegmental features*. MA report, University of Texas at Austin.

Trainor, Laurel J. & Beth Adams (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception and Psychophysics* **62**. 333–340.

Voegelin, Charles F. (1935). *Tübatulabal grammar*. Berkeley : University of California Press.

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

Woodrow, Herbert (1909). *A quantitative study of rhythm : the effect of variations in intensity, rate and duration*. New York : Science Press.

Yoshida, Katherine A., John R. Iversen, Aniruddh D. Patel, Reiko Mazuka, Hiromi Nito, Judit Gervain & Janet F. Werker (2010). The development of perceptual grouping biases in infancy : a Japanese-English cross-linguistic study. *Cognition* **115**. 356–361.