Journal of Phonetics 66 (2018) 82–99

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

The influence of varying vowel phonation and duration on rhythmic grouping biases among Spanish and English speakers

Megan J. Crowhurst

The University of Texas at Austin, United States

A R T I C L E I N F O

Article history:
Received 1 November 2016
Received in revised form 9 August 2017
Accepted 6 September 2017

Keywords:
Speech perception
Vowel phonation
Creaky voicing
Rhythmic grouping biases
Iambic-Trochaic Law

A B S T R A C T

Native Mexican Spanish and American English speakers were presented with streams of alternating syllables in which vowel duration and/or creaky phonation were rhythmically varied. Participants' grouping biases were measured as a function of their behaviour in segmenting sequences into recurrent bisyllabic units. Results indicated a creak-last grouping bias in both language groups. Duration varied singly was associated with a weak long-first grouping bias for Spanish and no consistent trend for English. When long creaky and short modal syllables were alternated, there was a significant creak-last bias and again no effect of duration in the English group. However, in the Spanish group, the long-first trend observed for duration varied singly was reversed and the effects of duration and creak were additive. Finally, when short creaky and long modal syllables were alternated, duration effects were highly significant in both language groups (fewer creak-last, more long-last groupings). Creak has been associated with final positions in higher-order prosodic domains in English, and less prevalently in Spanish. The current results show that both English and Spanish speakers can use this cue to segment rhythmic sequences into smaller, foot or word-sized units. This study is the first to establish that creak is perceptually salient for Spanish speakers and to demonstrate that the percept associated with duration can differ depending on whether it is varied singly or together with creak. More generally, the current findings show that grouping effects extend beyond intensity, pitch and duration, the features most often manipulated in rhythmic grouping studies inspired by the Iambic-Trochaic Law.

1. Introduction

When presented with rhythmically organised sound sequences, humans naturally segment them into smaller, recurrent groupings. For example, psychoacoustic studies of rhythmic grouping biases (RGBs) as early as Bolton (1894) and Woodrow (1909, 1911) report that listeners associate greater acoustic intensity with group onsets (a loud-first RGB) and greater duration with group endings (a long-last RGB). These generalisations have become known as the Iambic/Trochaic Law (ITL; Hayes, 1995). Some researchers attribute RGBs associated with intensity and pitch to hard-wired auditory mechanisms (de la Mora, Nespor, & Toro, 2013; Hay & Diehl, 2007; Hayes, 1995). However, whatever there is of "nature" in the development of human RGBs, there is growing evidence for a "nurture" component as well. Recent findings suggest that duration-based RGBs, in particular, are malleable and may be shaped by linguistic experience. Studies with both adults and infants have contributed persuasive evidence supporting this conclusion.

The outcomes of studies with adults indicate that a long-last RGB is not universal and that sensitivities to varied duration differ by language. Some studies in which adult listeners have been exposed to alternating, duration-varied sequences of nonlinguistic tones report the ITL-predicted long-last RGB for Dutch (Vos, 1977), English (Hay & Diehl, 2007; Iversen, Patel, & Ohgushi, 2008; Kusumoto & Moreton, 1997; Rice, 1992), French (Hay & Diehl, 2007), and Peninsular Spanish (Molnar, Carreiras, & Gervain, 2016). Similar studies that have used more speech-like sequences of syllables report the same long-last RGB for English (Crowhurst, 2016; Crowhurst & Teodocio-Olivares, 2014; Hay & Diehl, 2007; Kelly, Crowhurst, & Cobb, 2014), Italian (Bion, Benavides-Varela, & Nespor, 2011), French (Bhatara, Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013; Hay & Diehl, 2007), German (Bhatara et al., 2013), and Peninsular Spanish (Molnar et al., 2016).

E-mail address: mcrowhurst@austin.utexas.edu

https://doi.org/10.1016/j.wocn.2017.09.001

However, other studies report the absence of a duration-based grouping preference or even counter-ITL effects (a long-first RGB). Studies with speakers of Japanese (Iversen et al., 2008; Kusumoto & Moreton, 1997) and Mexican Spanish (Crowhurst, 2016) report no duration-based RGB, in contrast with comparison groups of English speakers. Jeon and Arvaniti (2016) report a weak trochaic, long-first grouping bias for Greek speakers and Korean speakers who were tasked with segmenting duration-varied sequences of tones separated by a 200 ms interval. Jeon and Arvaniti observed no long-last RGB for English speakers, in contrast to findings for English reported in other studies cited above. Crowhurst and Teodocio-Olivares (2014) report a trochaic long-first RGB for Zapotec speakers when duration was varied singly in alternating streams of syllables. In one crosslinguistic non-speech study, Molnar et al. (2016) observed a long-first RGB among native Basque speakers and Basque-dominant bilingual adults, but found a long-last RGB among adult monolingual speakers of Peninsular Spanish. Molnar and her colleagues also report that proficiency in an L2 can influence outcomes: two groups of native Spanish speakers with Basque as L2 (a Spanish-dominant group and a group of more balanced bilinguals) showed no significant grouping bias. At least two studies have found the magnitude of the long-last effect to differ more granularly across language groups (see Bhatara et al., 2013, for German and French, and Hay & Diehl, 2007, for French and English), and even across experiments with speakers of the same language (Bhatara et al., 2013), leading investigators to propose that listeners' linguistic experience may be associated with differences in sensitivity to rhythmic cues.

Outcomes for duration-based rhythmic segmentation in adult studies have also differed depending on the phonetic context in which duration was manipulated, both across and within experiments. As noted above, Crowhurst and Teodocio-Olivares (2014) observed a counter-ITL long-first bias among Zapotec speakers when duration-varied sequences of syllables were presented together with intensity-varied sequences, but Crowhurst, Kelly, and Teodocio (2016) observed a long-last RGB among members of the same Zapotec community in a different study in which duration was varied together with creaky voicing. In addition to conditions in which duration and intensity were varied singly, Crowhurst and Teodocio-Olivares (2014) and Crowhurst (2016) co-varied duration and intensity in the same sequences in a pattern that presented conflicting cues. These studies report that increasing a duration disparity between alternating syllables increased long-last groupings when an intensity disparity was held constant. This finding, observed among both English and Zapotec speakers, conflicted with the long-first outcome in the Zapotec group in Crowhurst and Teodocio-Olivares (2014) and with the absence of a consistent bias among Spanish speakers in Crowhurst (2016) when duration was varied singly, without intensity. These findings suggest that the processing of rhythmic duration cues can be influenced by cues associated with another sound feature when both are present in the listening context.

The argument that duration-based RGBs have an acquired, linguistically-sensitive component finds additional support from research with infants showing that sensitivity to varied duration develops (or not) as they gain experience with the native language. Yoshida et al. (2010) exposed groups of Japanese- and English-learning infants to alternating duration-varied tone sequences at two points during their first year. Neither group showed a bias at 5–6 months, but when tested at 7–8 months, the English learners showed evidence of the adult long-last RGB. In contrast, no preference was observed for the Japanese learners at 7–8 months, consistent with findings for Japanese speaking adults (Kusumoto & Moreton, 1997; Iversen et al., 2008). Molnar, Lallier, and Carreiras (2014) argue that by 9–10 months, infants developing in a Spanish/Basque bilingual context had learnt the duration-based RGB shown by adult speakers of the more dominant language, a long-last RGB for Spanish and a long-first RGB for Basque. Duration-based RGBs may also develop differently from RGBs associated with other features. For example, Bion et al. (2011) report that Italian-learning 7 month olds had in place a trochaic high-low grouping bias for pitch (the pitch-based RGB they observed among Italian-speaking adults), but not yet the adult long-last RGB.

The premise of the current research is that if RGBs can have an acquired component that is influenced by linguistic experience, then it should be possible to find evidence of grouping biases for various phonetic features which are available to listeners in segmenting speech. The current research studied this issue by conducting a rhythmic grouping experiment in which native speakers of Spanish and of English were tasked with segmenting alternating syllable sequences in which vowel phonation (modal vs. creaky voicing) and vowel duration were systematically varied.

2. The motivation for testing phonation and duration

Rhythmic grouping studies inspired by the ITL have generally focussed on intensity, duration and pitch. Along with Kelly et al. (2014) and Crowhurst et al. (2016), discussed below, a contribution of the current study is to expand the research program to include other speech features which are associated with prosodic organisation in speech. The pairing of creaky phonation and vowel duration in a rhythmic grouping experiment is interesting, as these cues can occur in overlapping linguistic environments. Moreover, some research (e.g. Andreeva, Barry, & Steiner, 2007; Barry, Andreeva, & Koreman, 2009; Jun, 2014) puts forward that lowered F0, a canonical property of phrase-final creak, may be an even more important cue to phrase finality than duration, although the relative contribution of these cues to perception is less well understood. For these reasons, the decision to study creaky phonation and vowel duration together was taken with the goal of learning whether listeners make similar or different use of duration-based and phonation-based cues in segmenting rhythmic sequences.

Duration-based cues are associated with both group finality and stress in the languages of this study. The association of increased duration with final positions in higher-level prosodic constituents (pre-boundary lengthening) is well studied and thought to exist in all languages in some form (Fletcher, 2010). Reports for English include Beckman and Edwards (1994) and Turk and Shattuck-Huffnagel (2000, 2007), and for Spanish, Rao (2010) and Prieto, Vanrell, Astruc, Payne, and Post (2012). Lengthening is also associated with stress in both languages (for English, see Beckman, 1986; Turk & Sawusch, 1996; for Spanish, Ortega-Llebaria & Prieto, 2007, 2011). Both stress-related and boundary-related lengthening effects are reported to be more pronounced in English than Spanish (Prieto et al., 2012), and Crowhurst's (2016) finding that English speakers were more sensitive than Spanish speakers to duration-based cues in a rhythmic grouping study would seem to be consistent with this difference.

Creaky phonation is canonically associated with ends of higher-order prosodic constituents in English (Redi & Shattuck-Hufnagel, 2001). Henton and Bladon (1988) found that glottalisation was more likely to occur sentence finally than in non-final positions in both British and American English speech. Creaky voicing has also been found to mark finality at the word level and has been increasingly studied as a sociolinguistic marker of finality in the productions of American English speakers. Redi and Shattuck-Hufnagel (2001) report that the rate of glottalisation in constituent-final words in American English speech increases with dominance in the prosodic hierarchy, and they interpret creaky voicing primarily as a syntactic marker. Other recent research finds that creaky voicing, where it occurs, is more likely to be associated with utterance-final than medial words; is common in the speech of both middle-aged and college-aged women (Oliveira, Davidson, Holczer, Kaplan, & Paretzky, 2015; Wolk, Abdelli-Beruh, & Slavin, 2012); and is also present though less common in the productions of college-aged men (Abdelli-Beruh, Wolk, & Slavinn, 2014). Creaky voicing can also occur initially in phrases that begin with a vowel, especially when the initial syllable carries the phrasal accent (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Ding, Jokisch, & Hoffmann, 2006; Garellek & Keating, 2015).[1] The properties of phrase-initial and final creak differ. Garellek and Keating (2015) report that creak in initial and final positions are distinguished in that phrase-final creak is characterised by lower F0 and a lower harmonics-to-noise ratio than phrase-initial creak. Moreover, because phrase-initial creak is associated with syllable onsets, it naturally manifests in vowel-initial glottal pulses, whereas phrase-final creak naturally occurs at least at syllable endings. In contrast to the substantial literature on glottalisation in English speech, there are virtually no descriptions of creaky voicing in the productions of Spanish speakers. However, Ding et al. (2006) report the presence of both word-initial and phrase-final creaky voicing in a small-scale study with eight native Spanish speakers. Garellek and Keating (2015) also report that phrase-final creak was present in the speech of at least 9 of 12 native Spanish speakers, though less prevalent than in the speech of English speakers in the same sentence-reading study. They further note higher rates of phrase-final creak for women than men, and that this difference was greater in the Spanish than the English group. It is not clear that creaky phonation is associated with prosodic prominence in Spanish.

[1] Gibson (2017) reports that creak was more likely to occur in unstressed than stressed syllables in the speech of college-aged American English speaking women, and that the likelihood of creak increased as a function of distance from the stressed syllable. As participants read lists of 1–4 syllable non-words, it is possible that "words" were treated as distinct phrasal units. If so, then creaky unstressed syllables would be in phrase-final or prefinal position. It is not clear from Gibson's description whether stressed syllables were vowel-initial, a primary conditioning factor for prominence-related creak.

Where there is an association between creaky phonation and finality in speakers' productions, it is reasonable to expect that listeners can also use creaky phonation to locate endings of higher-order prosodic constituents in speech. Kreiman (1982) and Carlson, Hirschberg, and Swerts (2005) find this to be the case for English speakers. However, there is relatively little research on the perception of creaky voicing among English speakers (in contrast to the fairly robust production literature), and there are no published reports for Spanish. It is therefore interesting to ask whether both Spanish speaking and English speaking listeners are sensitive to creaky phonation in rhythmic grouping tasks, and whether they make similar use of this information. To date, only two published studies have explored whether listeners can use creaky voicing to segment rhythmically alternating sequences. Crowhurst et al. (2016) presented adult Zapotec speakers with streams of syllables in which creaky vowel phonation, or vowel duration, or both were rhythmically varied. These investigators report a long-last RGB in all conditions in which duration was varied, and a significant though less robust effect associating creaky phonation with group onsets (a creak-first RGB) when phonation was varied singly. However, when shorter, creaky syllables were alternated with longer, modally phonated syllables, there were fewer long-last (and more creak-first) responses than when duration was varied alone. Both effects can be related to properties of the language: laryngealisation defines contrastive vowel categories in all Zapotec varieties. In the variety studied, the manner in which creaky phonation was expressed in test stimuli (an extended creaky period in the medial portion and/or latter half of the vowel) was especially characteristic of non-final positions in connected speech and the investigators proposed that listeners used this knowledge when no duration differences were present. In connected Zapotec speech, word-final creaky vowels tend to be shorter than preceding modal vowels, and the investigators concluded that this accounted for the seemingly conflicting outcome in the co-varied condition. They noted that their results were consistent with listeners using phonetic knowledge of their language to segment sequences and not lexical knowledge about the distribution of contrastive laryngealised vowel categories in words.

Kelly et al. (2014) explored the influence of varied duration and phonation on RGBs with English speakers and report long-last and creak-last biases when duration and creaky phonation were varied singly. When these features were co-varied "co-operatively", alternating long creaky and short modal syllables, there were more long-last/creak-last groupings than when either feature was varied singly, suggesting that phonation- and duration-based effects were additive. When duration and phonation were co-varied in conflict (alternating long modal and short creaky syllables), lengthening a vowel-final creaky segment increased creak-last groupings at the expense of long-last groupings, but the tendency was not significant. For English, then, limited evidence suggests that listeners make similar use of creak-based and duration-based cues to locate unit endings in rhythmically organised syllable sequences.

Together, Kelly et al. (2014) and Crowhurst et al. (2016) suggest that as for duration, (creaky) phonation-based RGBs may differ depending on properties of the native language. However, the findings of these studies cannot be directly compared for several reasons. First, the linguistic status of laryngealisation differs in the two languages: as noted, vowel laryngealisation is contrastive in Zapotec but is a noncontrastive surface phonetic feature in English. Second, the stimuli used in the two studies were different. Importantly, creaky phonation was differently sourced (naturally produced in Crowhurst et al., 2016, and synthesised by lowering F0 in Kelly et al., 2014). Nor can Kelly et al. (2014), a preliminary study, be considered to have provided an adequate test of the influence of duration on English speakers' RGBs because duration was varied in a binary opposition and not incrementally in a series of steps. In the present study, Spanish- and English-speaking groups were tested using a single set of alternating sequences of syllables in order to provide a more direct crosslinguistic test of RGBs associated with creaky phonation and to compare phonation-based and duration-based rhythmic grouping behaviour. The current study extends Kelly et al. (2014) not only in including a second language group but also in that the design included multiple manipulation levels for both varied duration and varied phonation.

3. Hypotheses

Creak in this study was simulated by mimicking canonical properties of phrase-final creak in English, in particular, lowered F0 (Garellek & Keating, 2015; Redi & Shattuck-Hufnagel, 2001; Keating and Garellek, 2015; Garellek & Seyfarth, 2016). As phrase-final creak is reported in both languages of this study, and in light of Kelly et al.'s (2014) finding for varied phonation, I expected to find a creak-last RGB in both the English and Spanish groups, at least when phonation was varied singly. Given Garellek and Keating's (2015) report that phrasal creak is less prevalent in Spanish than English, I expected phonation-based effects to be weaker in the Spanish group.

As length has multiple roles in English and Spanish in signalling both stress and finality, competing hypotheses for duration were considered. A long-last RGB would be expected if listeners interpreted length as a boundary cue. On the other hand, as stress in both English and Spanish is canonically penultimate in words ending in a CV syllable (as were the syllables used in this study) a long-first RGB would be more consistent with listeners' interpreting length as a prominence cue. However, I considered the possibility of a long-first RGB to be stronger in the Spanish than the English group due to asymmetries in the distribution of stress in English and Spanish. According to Crowhurst (2016:7–8), nonfinal stress is arguably more prevalent in Spanish than in English, and even more so in connected speech than in the lexicon. Crowhurst (2016) found no consistent trend among Spanish speakers when duration was varied singly and suggested that this indeterminate outcome might reflect competition between opposing stress- and boundary-related interpretations of duration-based cues. As outcomes for varied duration have differed across studies (see Section 1), trends for the Spanish group in the current study might be clearer, and could go either way. In English, word-final stress (primary or secondary stress) is more prevalent so that stress- and boundary-related duration cues are more likely to converge on final syllables. This holds at the phrase level as well as the word level, given that unstressed function words right-adjoin to form phrasal constituents in English. For this reason and given the findings of prior studies (see Section 1), I expected a long-last RGB to be the more likely outcome in the English. On the whole, because prosody-related duration cues are more subtle and phrase-final creak less prevalent in Spanish than English speech, I expected the magnitude of any duration- or phonation-based grouping biases to be larger in the English group.

In addition to varying duration and phonation singly, these features were also co-varied in two patterns. In one pattern, these features were co-varied co-operatively (a short, modal syllable alternating with a long, creaky one), and I expected the influence of these features to be straightforwardly additive in the English group: there should be more long-last/creak-last groupings in this condition than when either feature was varied singly, but no evidence of an interaction. This was also expected of the Spanish speakers, provided that they interpreted duration-based cues as signaling group-finality. In the second pattern, duration and phonation were varied in opposition: a short, creaky syllable was alternated with a long, modal one. The primary reason for including this condition was to see which of the two cues would be more strongly associated with group finality when they were arranged in a conflicting pattern. Given that prosody-related lengthening is not restricted to phrase-final position in either English or Spanish, I thought it was possible that the percept for duration might shift, with listeners associating length with group onsets in this condition. I expected to find a creak-last/long-first bias for the English speakers in this condition even if evidence for a long-last RGB were present in other conditions. A creak-last/long-first bias for the Spanish speakers was anticipated in any event.

4. Method

4.1. Overview

To test the hypotheses stated at the end of Section 3, native speakers of Spanish and English were presented with alternating streams of the syllables *ba* and *ga* in which either vowel duration, vowel phonation, or both were varied in a rhythmic, binary pattern. Their task was to indicate whether they thought each sequence consisted of recurrent *baga* or *gaba* syllable pairs. The design included a control and four test conditions. In a singly-varied Duration condition, syllables with short (*ba*) and long (*ga*) modal vowels were alternated. In a singly-varied Phonation condition, a modal (*ba*) and a creaky (*ga*) syllable of the same duration were alternated. As described in Section 3, there were also two co-varied conditions. In a Co-operating condition, short modal *ba* was paired with a longer creaky *ga*. In a Competing condition, short creaky *ba* and longer modal *ga* were alternated. All procedures were carried out under a protocol approved by the Institutional Review Board of the University of Texas at Austin.

4.2. Stimulus preparation

As the procedures for constructing stimuli are described elsewhere (Crowhurst, 2016; Crowhurst & Teodocio-Olivares, 2014; Crowhurst et al., 2016), this section focuses on details specific to the current study. Coarticulated, alternating sequences of the syllables *ba* and *ga* were recorded by an adult female native speaker of English at a relaxed tempo. Modally voiced tokens of *ba* and *ga* were selected from naturally emphasised positions to avoid irrelevant position-related differences. These tokens were edited in Praat (Boersma & Weenink, 2016) to produce a series of syllables for use in stimuli. In the Control condition, the vowels in *ba* and *ga* were modal and measured 165 ms. In the Duration condition, the duration disparity was increased in increments of approximately 40 ms; *ba* with a 125 ms vowel was alternated with *ga* with a 165, 205, or 245 ms vowel. The 40 ms interval was based on seven full modal voicing cycles (approximately 5.7 ms for this speaker). In the Phonation condition, *ba* and *ga* both had 165 ms vowels, and *ga* ended in a creaky period. The most important properties of phrasal creak are lowered F0, aperiodic voicing and damping (Dilley et al., 1996; Garellek & Keating, 2015; Garellek & Seyfarth, 2016), with lowered F0 being sufficient to induce the percept (Dilley et al., 1996). Accordingly, creaky phonation was synthesised from prepared modal syllable tokens by decreasing F0 in every second or third voicing cycle to the point that no pitch track was visible in the creaky portion of the vowel. Three manipulation levels were created in which roughly the final 10%, 30% or 50% of a vowel was creaky, creating a scale. These proportions were preserved across duration levels. Fig. 1 presents the sound wave and spectrogram for a *ga-ba* sequence in the phonation condition (level 2). In both of the co-varied conditions, duration was manipulated as described for the Duration condition, with three levels. The difference was that in the Co-operating condition, the longer syllable (*ga*) was creaky, and in the Competing condition, the shorter syllable (*ba*) was creaky. Mean level intensity for all syllables was 67 dB. In modally voiced portions of vowels across the stimulus set, F0 averaged between 173.5 and 175.1 Hz.

In sequencing, syllables were separated by 100 ms of 20.5 dB noise to simulate stop closure. Sequences were 10–10.5 s in length, had between 36 and 40 syllables, were counterbalanced for initial syllable (*ba* vs. *ga*), and consisted of a whole number of syllable pairs (*baga* or *gaba*). To mask sequence onsets, alternating syllables were blended with white noise for the first five seconds. Over this period, the noise was ramped down from 67 to 0 dB and the syllables were ramped up from 0 dB. A 500 ms segment of 67 dB white noise was added to sequence endings as a distractor in case participants listened to the end. Syllables were sequenced according to a 3×3 design (three manipulations levels for phonation and three for duration). The design and values assigned to the alternating syllables are presented in Table 1(a) for the control and singly varied Duration and Phonation conditions, and Table 1(b) for the co-varied conditions). In all, there were 25 distinct combinations (one control combination, 3 combinations in each of the singly-varied conditions, and nine combinations in each of the co-varied conditions). Together with their counterbalanced counterparts, there were 50 distinct sequences in all.

4.3. Participants

Twenty-four adult native English speakers (seven men and 17 women) and twenty-eight adult native speakers of a northern Mexican variety of Spanish (14 women and 14 men) participated in the study. The English speakers were undergraduate students at The University of Texas at Austin. They were recruited through an announcement in a campus events bulletin and were tested in a campus phonetics laboratory. The Spanish speakers were recruited by an instructor at La Salle University in Obregón City, Sonora, Mexico. All were undergraduate students at La Salle University and were tested in a classroom on that campus. In response to screening questions, subjects reported normal hearing and having not lived in an environment where a language other than the native language was spoken. No participant had had early exposure (before puberty) to a language other than the native language and, although all had taken one or more foreign language courses in high school or at university, none reported having more than basic proficiency in a non-native language. Participants were paid for their time in the national currency.

4.4. Experimental procedures

Sessions were conducted in the participants' native language by the author or a trained assistant. The experiment was controlled by the software program SuperLab 4.0 (Cedrus Corporation) running on a MacBook Pro laptop computer. Participants were tested in groups and heard sequences over a high-quality Bose speaker connected by cable to the computer and positioned in front of the group. A set of 52 stimuli (the 48 test sequences and two repetitions of each control sequence) was broken into two balanced lists, which were presented to subjects 3 times (i.e. 6 sets of 26 trials). In all, every test combination was presented six times and the control sequences 12

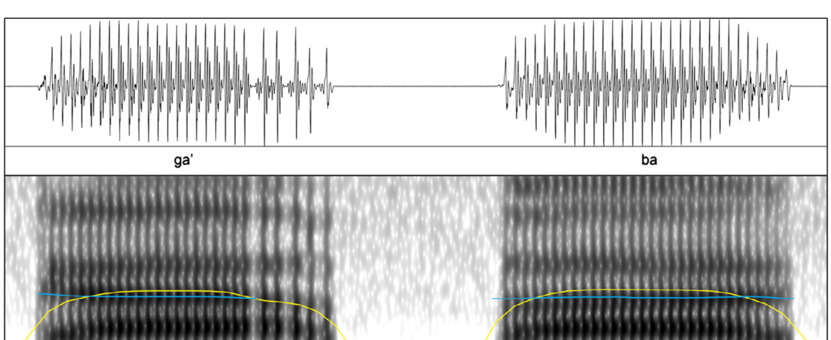

Fig. 1. Typical *ga-ba* sequence overlaid with F0 (blue) and intensity (yellow) contours. (Phonation condition, level 2.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Experimental design and values assigned to syllables in alternating sequences. (The F0 measurement represents average F0 for the modal portion of the vowel.)

	Manipulation level	**ba** Duration (ms)	Creak (%V)	Avg F0 (Hz)	**ga** Duration (ms)	Creak (%V)	Avg F0(Hz)
(a) Control, Duration and Creak conditions							
(a) Control	Level 0	165	0	173.5	165	0	175.2
(b) Duration	Level 1	125	0	173.5	165	0	175.2
	Level 2	125	0	173.5	205	0	175.1
	Level 3	125	0	173.5	245	0	175.2
(c) Phonation	Level 1	165	0	173.5	165	10	174.6
	Level 2	165	0	173.5	165	30	175.1
	Level 3	165	0	173.5	165	50	175.6

	Manipulation level	**ba** Dur (ms)	Cr (%V)	Avg F0 (Hz)	Manipulation level	**ga** Dur (ms)	Cr (%V)	Avg F0 (Hz)
(b) Competing and Co-operating conditions								
Competing					Dur 1	165	0	174.2
	Dur 0/Cr 1	125	10	173.7	Dur 2	205	0	174.1
					Dur 3	245	0	174.2
					Dur 1	165	0	174.2
	Dur 0/Cr 2	125	30	174.0	Dur 2	205	0	174.1
					Dur 3	245	0	174.2
	Dur 0/Cr 3	125	50	175.4	Dur 1	165	0	174.2
					Dur 2	205	0	174.1
					Dur 3	245	0	174.2
Co-operating					Dur 1/Cr 1	165	10	174.6
	Dur 0/Cr 0	125	0	173.5	Dur 2/Cr 1	205	10	174.5
					Dur 3/Cr 1	245	10	174.4
					Dur 1/Cr 2	165	25	175.1
	Dur 0/Cr 0	125	0	173.5	Dur 2/Cr 2	205	25	175.1
					Dur 3/Cr 2	245	25	174.4
	Dur 0/Cr 0	125	0	173.5	Dur 1/Cr 3	165	50	175.4
					Dur 2/Cr 3	205	50	174.6
					Dur 3/Cr 3	245	50	175.2

times over the course of the experiment. Presentation order within sets was automatically randomised by the software each time a block was run. There was a brief rest between sets. Participants were instructed to respond after the initial masking noise had faded away but while the sequence was still playing. Testing was preceded by four practice items and lasted approximately 35 min. The task was to indicate whether sequences divided most naturally into *gaba* or *baga* syllable pairs. Participants were provided with response booklets consisting of (i) a cover sheet requesting answers to screening questions and participants' informed consent; (ii) a page with instructions and lines for the practice trials; and (iii) six labeled response sheets. These presented 26 numbered lines with arbitrary sequences of alternating *ba* and *ga* syllables (see the example in Appendix A). Participants indicated their grouping decisions by bracketing a pair of adjacent syllables on each line. Lines with different syllable arrangements were ordered differently on successive response sheets. To the right on each line, subjects were asked to provide a confidence rating (sure vs. not sure).

4.5. Data analysis procedures

The design of the study provided for a maximum of 8112 data points (3744 in the English and 4368 in the Spanish group). There were 214 missing responses (English 21, Spanish 193) leaving a total of 7898 data points for the analysis (Spanish 4175, English 3723). For the statistical analyses, mixed effects logistic models were fit to the response data using the *glmer* function in the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) of the statistical software program R (R Core Team, 2016). This method estimated the maximum likelihood of the positively coded response, a *baga* decision. The variables input into models were the predictors LANGUAGE (English vs. Spanish), DURATION and CREAK. DURATION was coded with four levels representing the magnitude of the duration disparity (0–3). CREAK was coded with seven levels representing increases in the duration of creaky phonation in *ga* relative to *ba*: 0 for "no difference"; 1–3 for manipulation levels representing increases in creaky phonation in *ga* in the Co-operating condition; and levels −1 to −3 for increases in creaky phonation in *ba* in the Competing condition. All models included SUBJECT as a random intercept. (Confidence ratings were not analysed as it was clear that many of the Spanish-speaking participants had not used these as instructed.) The best-fitting statistical model was determined through a process of model comparison using the method of backwards elimination. In the main analysis, the full model was the maximal model (Barr, Levy, Scheepers, & Tily, 2013) in which the predictors were entered as the three-way interaction DURATION*CREAK*LANGUAGE in fixed effects, and which included a random slope for DURATION*CREAK. More parsimonious models were constructed by removing terms one by one from random slopes, then from fixed effects, creating a series of nested models. Each new model with a term removed (Fit-1) was compared with the one before (Fit) using ANOVA. This process ended when the results indicated that Fit-1 provided a significantly worse fit for the data than Fit ($\alpha = 0.05$). Given the hypotheses presented in Section 3, the following tests were planned: (i) an overall model fit to the entire data set for both languages; (ii) models for CREAK at baseline DURATION (phonation varied singly) and (iii) DURATION at baseline CREAK (duration varied singly) for each language group; (iv) a model for CREAK and DURATION varied co-operatively for each language group; and (v) a model for CREAK and DURATION varied in opposition for each language group. When interactions in models described in (iv) and (v) were significant, pairwise comparisons (0.95 confidence level, Tukey adjustment) were made using the *lsmeans* package in R (Lenth, 2016). For the overall analysis, outcomes are treated as significant if they are associated with *p*-values that meet the alpha criterion of 0.05. Outcomes with *p*-values between 0.05 and 0.1 are considered marginally significant, with confidence diminishing as the *p*-value increases. For statistical tests conducted on subsets of the response data, the more conservative alpha criterion of 0.01 is adopted, bearing in mind the increased risk of a type I error. When the more conservative alpha criterion is adopted, *p*-values falling between 0.01 and 0.06 are discussed as marginally significant.

5. Results

The response data for both language groups, organised by condition, are presented in Tables B.1 (English) and Table B.2 (Spanish) in Appendix B. As a first step in the statistical analysis, mixed effects logistic regression models were fit to the combined response data for both language groups. The best-fitting model, whose output appears in Table 2, included the three-way interaction CREAK*DURATION*LANGUAGE in fixed effects, a random slope for CREAK and a random slope for DURATION. The best-fitting model provided a significantly better fit for the response data than a model which included a random slope for the interaction CREAK*DURATION ($X^2 = 0.5063$, df = 4, $p = 0.9729$), and a model which replaced the three-way interaction with the two-way interactions CREAK*LANGUAGE and DURATION*LANGUAGE ($X^2 = 6.2871$, df = 2, $p = 0.0431$).

The most important outcomes in the overall analysis are those associated with the interactions CREAK*DURATION and CREAK*DURATION*LANGUAGE. The interaction CREAK*DURATION approached significance; this, and the negative sign on the coefficient indicate that the effect of CREAK decreased as DURATION increased in the English comparison group. As the *p*-value associated with the CREAK*DURATION term fell short of significance at the 0.05 alpha criterion, we cannot be fully confident that this outcome is not due to chance. Importantly, however, the values associated with the interaction CREAK*DURATION*LANGUAGE indicate that whatever the outcome for the CREAK*DURATION interaction in the English comparison group, the outcome in the Spanish was significantly different. Given the difficulty of interpreting effects associated with predictor variables when interactions are significant, and to explore differences by language, the planned statistical tests described in Section 4.5 were carried out. Outcomes for duration and phonation varied singly are discussed in Section 5.1, followed by the co-varied conditions in Section 5.2.

5.1. Singly varied conditions

Stated in terms of the statistical analysis, the main expectations for duration and phonation varied singly were that increases in both CREAK at DURATION 0 and DURATION at CREAK 0 would be positively correlated with increases in *baga* responses (more creak-last and long-last groupings) in both language groups. The alternative possibility of a negative correlation between DURATION and *baga* responses was considered more likely for the Spanish than the English group in light of considerations discussed in Section 3. A negative trend in *baga* responses as DURATION increased would suggest that Spanish speakers were interpreting length as stress related, and not as a cue to finality. Based on the literature cited in Section 2 (e.g. Crowhurst, 2016; Garellek & Keating, 2015), I also expected that any biases favouring long-last and creak-last responses might be stronger in the English than in the Spanish group.

The outcomes for phonation varied singly (CREAK at DURATION 0) in the English and Spanish groups are charted in Fig. 2a. We see an overall positive association between CREAK and *baga* responses in the English group, and the outcome in the Spanish group is comparable. The results shown in Table 2 reveal that fixed effect of CREAK was significant in the English comparison group, which confirms our expectation of a creak-last grouping trend, overall. Using the odds ratio (OR, the exponent of the coefficient) as an indicator of effect size (Hosmer & Lemeshow 2004:47), we see that the odds of a *baga* response were on average 1.12 times greater per unit of increase in CREAK (at DURATION 0). The values associated with the CREAK*LANGUAGE(SPAN) term in Table 2 indicate that the outcome for CREAK at DURATION 0 in the Spanish group was not significantly different from the outcome in the English comparison group. The expectation that increases in CREAK would straightforwardly increase creak-last groupings was not fully confirmed, as the function associated with CREAK was not completely linear: CREAK and *baga* responses were positively associated between CREAK 0 and 2, but this trend reversed at

Table 2
Best-fitting model predicting *baga* responses overall.

Random slopes	Var	SD	Cor	
SUBJ (intercept)	0.229	0.479		
CREAK	0.025	0.157	−0.34	
DURATION	0.042	0.205	0.55	0.11

($n = 7898$)	Coef	OR	SE	z	p
Intercept	0.261	1.30	0.119	2.183	0.0290
CREAK	0.117	1.12	0.050	2.327	0.0200
DURATION	0.078	1.08	0.054	1.433	0.1520
LANGUAGE(SPAN)	0.062	1.06	0.163	0.380	0.7036
CREAK*DURATION	−0.035	0.97	0.019	−1.825	0.0679.
CREAK*LANGUAGE(SPAN)	0.030	1.03	0.069	0.441	0.6594
DURATION*LANGUAGE(SPAN)	−0.035	0.97	0.074	−0.474	0.6355
CREAK*DURATION*LANGUAGE(SPAN)	0.068	1.07	0.026	2.601	0.0093

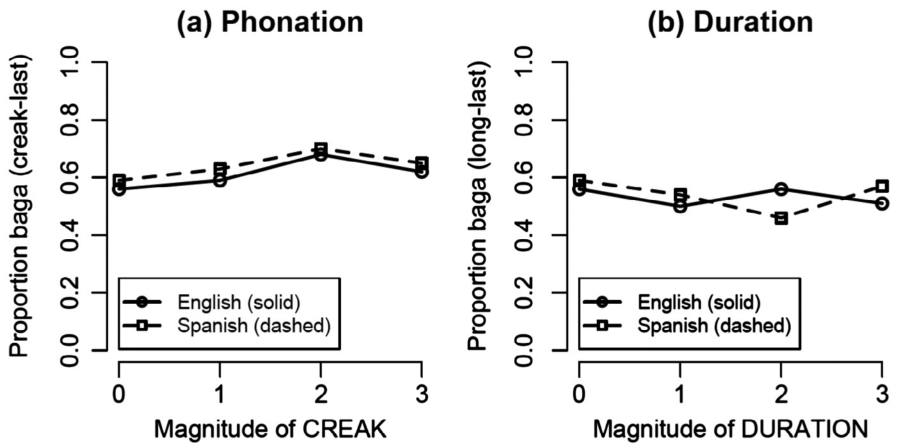

Fig. 2. Proportion of *baga* responses as a function of (a) increasing CREAK in the Phonation condition (*ga* is creaky) and (b) increasing DURATION in the Duration condition (*ga* is long) in the English and Spanish groups.

Table 3
Best-fitting factorial model predicting *baga* responses for phonation varied singly in the English and Spanish groups.

(n = 1518)	Coef	OR	SE	z	p
Intercept	0.312	1.37	0.104	2.993	0.0028
CREAK 1	0.167	1.18	0.148	1.126	0.2600
CREAK 2	0.519	1.68	0.152	3.415	0.0006
CREAK 3	0.271	1.31	0.147	1.846	0.0649

CREAK 3 in both languages. A factorial model fitted to the baseline CREAK data (control and Phonation conditions) for both languages indicated that the difference between CREAK 0 and 2 was significant. This model, whose output appears in Table 3, included only the intercept and the fixed effect CREAK and provided a better fit for the data than a model which also included a random slope for CREAK (X^2 = 14.618, df = 9, p = 0.102). The difference between CREAK 0 and 3 was not significant at the more conservative 0.01 alpha level. By language, the difference between CREAK 0 and 2 was significant in the English group (n = 719, Coef. = 0.568, SE = 0.221, z = 2.569, p = 0.0102). The same outcome was only marginally significant in the Spanish group (n = 799, Coef. = 0.475, SE = 0.209, z = 2.272, p = 0.0231). Here, the significance of the overall result for both language groups and the similarity to the outcome in the English group increase our confidence that the Spanish result is not due to chance.

Outcomes for duration varied singly are charted in Fig. 2b. In the overall analysis (Table 2), the fixed effect of DURATION was not significant. However, we see in Fig. 2b that while no clear trend was associated with baseline duration in the English group, there was a limited negative trend in the Spanish group: increases in DURATION between levels 0–2 decreased *baga* responses (fewer long-last groupings), but there was an upturn in *baga* responses at level 3. To test for the significance of this trend, a model which included a Subject intercept and DURATION as a fixed effect was constructed for the Spanish data in the Control and Duration conditions. The overall negative correlation seen in Fig. 2b was not significant at the 0.01 alpha level (Coef. = −0.115, SE = 0.062, z = −1.854, p = 0.0637). (However, see related discussion of the fixed effect of DURATION in Section 5.2.1.)

Summarising, outcomes in the singly-varied conditions confirmed the expectation of a creak-last RGB overall in both language groups: increasing the creaky segment from 0% to 10% to 30% of the vowel in *ga* was linearly associated with increases in *baga* (creak-last) groupings. Although this trend seemed to reverse at CREAK 3 in both language groups, the proportion of creak-last groupings was still higher than at CREAK 0 and 1, although the difference was not significant. Regarding the trend reversal at CREAK 3, I speculate that increasing synthesised creaky phonation to 50% of the vowel may have seemed unnatural to speakers. (This reversal was either less prominent or not present at all when CREAK and DURATION were increased together; see Section 5.2.)

Outcomes for duration varied singly did not confirm the expectation of a long-last RGB. No clear trend was observed in the English group. Although Fig. 2b suggests the presence of a contra-ITL long-first RGB in the Spanish group between DURATION 0 and 2, which would be consistent with Spanish speakers interpreting length as a stress-related cue, this trend was not statistically significant. As the proportions of *baga* responses at DURATION 0 and 3 were comparable (0.59 vs. 0.57), it might simply be that the decrease at DURATION 3 indicates that a duration disparity of 120 ms seemed unnatural to the Spanish speakers and that the DURATION 3 outcome reflects indecision. This interpretation of the results is plausible, as prosody-related duration effects are reported to be relatively small in Spanish, compared with English (Ortega-Llebaria & Prieto, 2007; Ortega-Llebaria & Prieto, 2011). Finally, the hypothesis that any trends would be weaker in the Spanish group was not supported by outcomes in the singly-varied conditions, given the similar patterning in the two language groups in the Phonation condition and the finding of a duration-based trend in the Spanish but not the English group.

5.2. Co-varied conditions

In the best-fitting overall model in Table 2, the values associated with the interactions CREAK*DURATION and CREAK*DURATION*LANGUAGE indicate that there were differences in Response when CREAK and DURATION in one or more conditions in which both CREAK and DURATION were increased together, as compared with baseline, and that there were significant between-group differences. However, these overall findings do not indicate which method of co-varying CREAK and DURATION (co-operatively, competitively, or both) might have been responsible for any significant differences. In order to better understand the effects indicated by the interactions, statistical models were constructed for the English and Spanish groups separately in each of the co-varied conditions.

5.2.1. Co-operating condition

In the Co-operating condition the syllable *ga* was both long and creaky so that a *baga* response represented both a creak-last and a long-last grouping. Under the initial expectation that creak-last and long-last RGBs would be confirmed, the purpose of this condition was to test whether the effects of increasing CREAK and DURATION in the same position would be additive. An affirmative outcome would be supported if the proportion of *baga* responses in the Co-operating condition was proportionally higher than when either feature was varied singly. If the magnitude of a positive trend in the Co-operating condition were comparable to a positive trend in either baseline condition, this would suggest that listeners were relying on either duration-based or phonation-based cues, and may have treated the other cue as redundant. If a creak-last RGB and a counter-ITL long-first RGB were found in the baseline conditions in the Spanish group, then duration- and phonation-based cues would be in conflict not co-operation. In that case, it should be possible to determine which of the two cues was more dominant.

Trends in the Co-operating condition are charted in Fig. 3. The graphs display the proportion of creak-last groupings (left), and long-last groupings (right), as a function of increasing CREAK and DURATION, respectively. Dashed lines represent singly varied phonation and duration. The graphs for the English group in Fig. 3a indicate that the functions associated with CREAK at DURATION 1–3 track baseline CREAK: we see a positive trend favouring *baga* responses from CREAK 0–2, with a downturn at CREAK 3, indicating an overall creak-last RGB in the Co-operating condition, as in the Phonation condition. As in the singly-varied Duration condition, no clear bias was associated with varied duration. In the rightmost panel we see evidence of a weak positive trend favouring long-last groupings at CREAK 3 when DURATION increased from 1 to 3, but it is difficult to interpret this result as meaningful in the absence of trends at other CREAK levels. We can be confident that in the Co-operating condition, English speakers' decisions were driven by phonation- and not by duration-based cues.

Outcomes for Spanish are graphed in Fig. 3b. Varied duration had a stronger influence in the Spanish group and we see evidence of an interaction between CREAK and DURATION: the leftmost panel shows that slopes for trends associated with

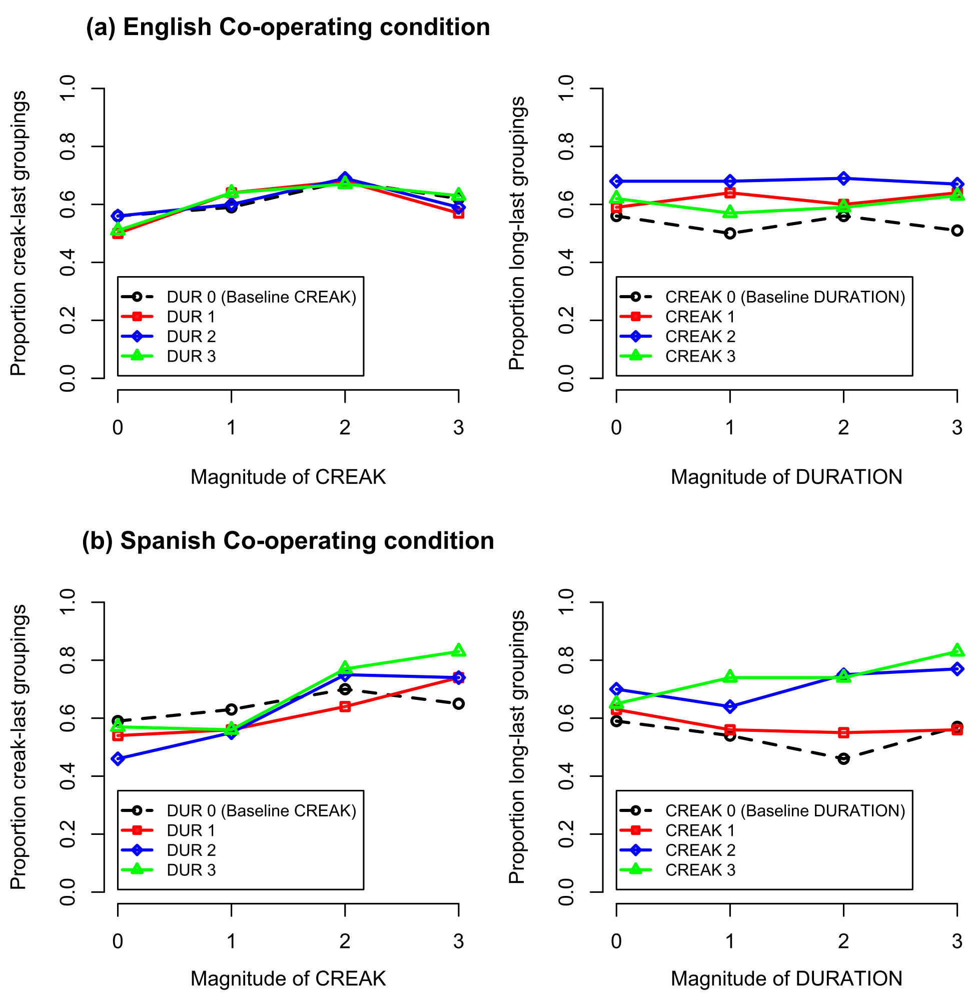

Fig. 3. Co-operating condition (*ga* is long/creaky, *ba* is short/modal). Left: Proportion of creak-last groupings as a function of increasing CREAK at different DURATION levels. Right: long-last groupings as a function of increasing DURATION at different CREAK levels.

CREAK were overall steeper between DURATION 1–3 than at baseline CREAK. The trends associated with CREAK at DURATION 1 and 3 were also more consistently positive than at DURATION 0, with no reduction in creak-last groupings at CREAK 3. In the panel on the right, we see that the slope of the function for varied duration flattens at CREAK 1 and reverses at CREAK 2 and 3 (apart from a reduction in long-last responses at level CREAK 2/DURATION 1).

To explore outcomes in the Co-operating condition, models were constructed on a subset of the data for each group that combined the data for the Co-operating condition with the baseline CREAK and DURATION data (Control, Duration and Phonation conditions). The model that provided the best fit for the English data, shown in Table 4, included only the fixed effect CREAK, a random slope for CREAK, and SUBJECT as a random intercept. This model provided a significantly better fit for the data than the model which did not include a random slope for CREAK (X^2 = 16.062, df = 2, p = 0.0003). No model that included DURATION significantly improved goodness of fit (with DURATION included as a fixed effect: X^2 = 0.088, df = 1, p = 0.7671). In Table 4 we see that the OR (= 1.13) indicates that a creak-last grouping was on average 13% more likely per unit of increase in CREAK, but that the fixed effect of CREAK was not significant at the 0.01 level when CREAK was treated as a discrete variable. As the effect was marginally significant, I cautiously interpret the result as indicating that there was a somewhat weak positive overall trend favouring creak-last groups when CREAK was increased at DURATION 0 and when DURATION and CREAK were increased together in the same syllable position. A factorial model constructed test for significant differences between CREAK 0 and CREAK 1, 2 and 3 increases confidence that the observed tendency toward creak-last groupings is due to the treatment and not to chance: the output of the factorial model, shown in Table 5, indicates that differences between CREAK 0 and 1 and between CREAK 0 and 2 were significant at the 0.01 level.

Table 4
Best-fitting model predicting *baga* responses in the Control, Duration Phonation, and Co-operating conditions in the English group.

Random slopes		Var		SD	Cor
SUBJ (intercept)		0.306		0.553	
CREAK		0.055		0.234	−0.70
(n = 2436)	Coef	OR	SE	z	p
Intercept	0.280	1.32	0.132	2.130	0.0332
CREAK	0.119	1.13	0.061	1.954	0.0506.

Table 5
Factorial model predicting *baga* responses in the Control, Duration Phonation, and Co-operating conditions in the English group.

Random slopes	Var	SD	Cor		
SUBJ (intercept)	0.305	0.552			
CREAK 1	0.105	0.324	−0.26		
CREAK 2	0.574	0.757	−0.57	−0.18	
CREAK 3	0.357	0.598	−0.47	−0.73	0.61
(n = 2436)	Coef	OR	SE	z	p
Intercept	0.159	1.17	0.137	1.165	0.2441
CREAK 1	0.362	1.44	0.136	2.658	0.0079
CREAK 2	0.655	1.93	0.197	3.316	0.0009
CREAK 3	0.276	1.32	0.170	1.625	0.1042

The best-fitting model for the Spanish data, whose output appears in Table 6, included DURATION and CREAK as fixed effects and the interaction CREAK*DURATION. This model also included random slopes for DURATION and CREAK, and SUBJECT as a random intercept. The model with the interaction CREAK*DURATION in fixed effects significantly improved goodness-of-fit over the model without (X^2 = 22.135, df = 1, p < 0.00001), but adding a random slope for the interaction CREAK*DURATION did not (X^2 = 3.1239, df = 4, p = 0.5373).

In the best-fitting model, the fixed effect of DURATION was significant, the fixed effect of CREAK was not significant at the 0.01 level, but the interaction CREAK*DURATION was highly significant. The OR (0.85) for DURATION indicates that a long-last grouping was 15% less likely per unit of increase in DURATION at CREAK 0. The interaction can be interpreted in terms of how the effect of CREAK changes across levels of DURATION. Multiplying the OR for CREAK (1.14) by the OR for the interaction term (1.17) for every unit of increase in CREAK, we find that the odds of a long-last/creak-last grouping were 1.33 times greater for each additional increase in CREAK at DURATION 1 (1.14*1.17 = 1.33), 1.56 times greater at DURATION 2 (1.14*1.17*1.17), and 1.82 times greater at DURATION 3 (1.14*1.17*1.17*1.17).

We see clearly in Fig. 3b (left panel) that the effect of CREAK was greater when CREAK and DURATION were varied together than when CREAK was varied alone (at DURATION 0). To test for significant differences between levels by CREAK and by DURATION in the Co-operating condition in the Spanish group, pairwise comparisons were made as described in Section 4.5. The comparisons that were significant or approached significance at the 0.01 alpha level are presented in Table 7. Of the comparisons based on varied duration, only the difference between DURATION 1 and 3 at CREAK 3 approached significance. For varied CREAK, differences between CREAK 0 and 2 were significant at DURATION 2 and 3. Differences between CREAK 0 and 3 were significant at DURATION 2 and 3, but only approached significance at DURATION 1. Differences between CREAK 1 and 2 were marginally significant at DURATION 2 and 3. The difference between CREAK 1 and 3 was significant only at DURATION 3, although it approached significance at DURATION 2.

Outcomes in the Co-operating condition confirm a creak-last grouping bias in both language groups, as predicted by the hypothesis for phonation. Varied phonation was clearly the dominant predictor of responses in both groups. The expectation of a long-last RGB was not confirmed in the English group, and as varied duration had no observable influence on grouping decisions among the English speakers, the prediction that the effects of CREAK and DURATION would be additive in the co-operative condition was not supported. Outcomes

Table 6
Best-fitting model predicting *baga* responses in the Control, Duration Phonation, and Co-operating conditions in the Spanish group.

Random slopes	Var		SD	Cor	
SUBJ (intercept)	0.219		0.468		
CREAK	0.059		0.243	−0.66	
DURATION	0.019		0.139	−0.29	0.36
(n = 2716)	Coef	OR	SE	z	p
Intercept	0.343	1.41	0.130	2.641	0.0083
CREAK	0.133	1.14	0.073	1.819	0.0689
DURATION	−0.159	0.85	0.061	−2.614	0.0090
CREAK*DURATION	0.155	1.17	0.033	5.681	<0.0001

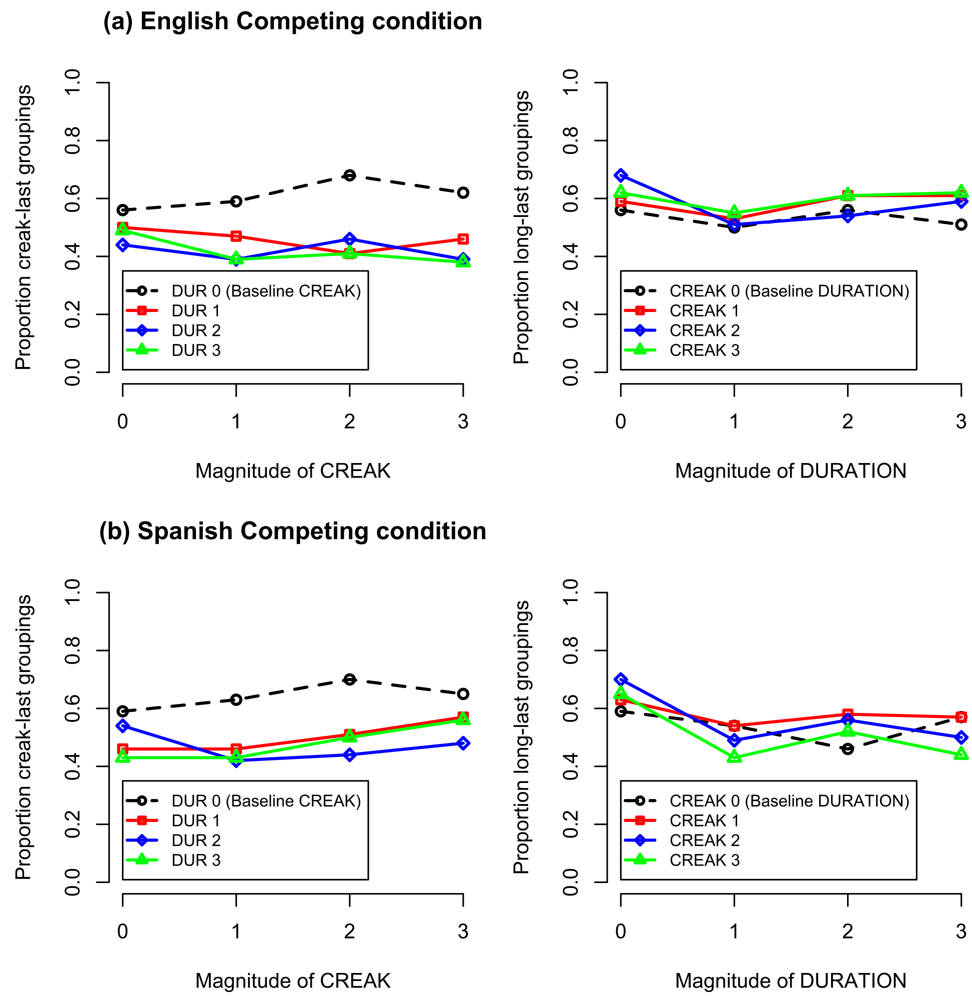

Fig. 4. Competing condition (*ga* = long/modal, *ba* = short/creaky). Left: Proportion of creak-last groupings as a function of increasing CREAK at different DURATION levels. Right: long-last groupings as a function of increasing DURATION at different CREAK levels.

Table 7
Pairwise comparisons in the Control, singly-varied and Co-operating conditions in the Spanish group.

	Comparison	Coef	SE	*z*	*p*
Varied Creak	CR 0/DUR 2–CR 2/DUR 2	−1.277	0.246	−5.187	<0.0001
	CR 0/DUR 3–CR 2/DUR 3	−1.095	0.248	−5.412	0.0011
	CR 0/DUR 1–CR 3/DUR 1	−0.899	0.245	−3.677	0.0215.
	CR 0/DUR 2–CR 3/DUR 2	−1.210	0.244	−5.963	0.0001
	CR 0/DUR 3–CR 3/DUR 3	−1.476	0.264	−5.597	<0.0001
	CR 1/DUR 2–CR 2/DUR 2	−0.938	0.246	−3.812	0.0131.
	CR 1/DUR 3–CR 2/DUR 3	−0.965	0.249	−3.882	0.0101.
	CR 1/DUR 2–CR 3/DUR 2	−0.872	0.244	−3.575	0.0306
	CR 1/DUR 3–CR 3/DUR 3	−1.346	0.264	−5.096	<0.0001
Varied Duration	CR 3/DUR 0–CR 3/DUR 3	−0.973	0.267	−3.645	0.0240.

were more interesting in the Spanish group. We saw in Section 5.1 that when only the data for DURATION varied at CREAK 0 were considered, the fixed effect of DURATION was not significant in the Spanish group at the 0.01 alpha level. At that point, we were not able to rule out the possibility that an overall trend toward long-first groupings when duration was varied singly (see Fig. 2b) might have been due to chance. However, the failure to reach significance in that analysis is attributable to the smaller number of observations. The results in Table 6 indicate that the fixed effect of DURATION was significant in a model that included a fuller set of observations. We can therefore be more confident that the negative trend associated with DURATION in Fig. 2 is a valid outcome, and may consider the possibility that the trend reflects listeners' interpreting duration-based cues as stress-related when duration was varied singly. However, the findings in the Co-operating condition suggest that combining phonation- and duration-based cues in the same syllable influenced the percept associated with varied duration. The significant interaction in this condition in the Spanish group indicated that the effects of CREAK and DURATION were not straightforwardly additive; rather, the odds of a long-last grouping increased as CREAK and DURATION increased together. Interpreting, it is worth considering the possibility that combining phonation-based cues, the dominant indicator of group finality, with duration-based cues may have incrementally shifted Spanish-speaking listeners from associating duration-based cues with stress to associating them with group finality.

5.2.2. Competing condition

Outcomes were not expected to be straightforward in the Competing condition. In this condition, short creaky *ba* was alternated with long modal *ga*, so that a creak-last grouping was also a long-first grouping. Given the prior expectation that long-last and creak-last RGBs would be observed in the singly varied conditions, the purpose of the Competing condition was to test whether one cue would be a stronger predictor of responses when duration- and phonation-based cues were at odds. As creaky phonation is canonically associated with group finality whereas duration-based cues are more ambiguous, I expected to find a stronger preference for creak-last groupings in both language groups.

Outcomes in the Competing condition are charted in Fig. 4. The primary findings for English and Spanish are most evident from the leftmost panels: in both language groups, there were fewer creak-last (and more long-last) groupings when phonation and duration were co-varied than when phonation was varied alone. In the Spanish group (Fig. 4b), varied phonation was more linearly associated with increases in creak-last groupings (although only weakly at DURATION 2) in comparison with the baseline Phonation condition. Varied duration was not linearly associated with differences in listeners' responses in the Competing condition. The graphs in Fig. 4a suggest a weak interaction in the English group: when both phonation and duration were varied, increasing DURATION did not increase creak-last groupings, as compared with the two baseline phonation condition.

The statistical analysis for the Competing condition was complicated by the fact that a different syllable was creaky in the Phonation and Competing conditions: the positively coded response (*baga*) was a creak-last grouping in the Phonation condition but a creaky-first grouping in the Competing condition. So that the Phonation condition could be used as the baseline comparison for the Competing condition, it was necessary to recode the response data. First, the negative values for CREAK in the Competing condition were changed to positive so that the CREAK scale was 1–3 in both the Phonation and Competing conditions. Second, a new dependent variable was created in which the values 0 and 1 in the Phonation condition were inverted. With this transformation, the positively coded response in the statistical analysis represented a creaky-first grouping in the Phonation and Competing conditions. As the positively coded response represented a long-last grouping in wherever duration was varied, the singly varied Duration condition still served as the baseline comparison for the Competing condition. However, with the recoding, responses in the control condition were no longer straightforwardly related to the Phonation condition, and so the control data were excluded from the analysis.

The model that best fit the data for the English group, whose output is shown in Table 8, included an intercept, the fixed effect DURATION, a random slope for DURATION and SUBJECT as a random intercept. This model provided a better fit than the model which also included CREAK as a fixed effect ($X^2 = 1.031$, df = 1, $p = 0.31$). The model's output reveals that the fixed effect of DURATION was highly significant; the OR indicates that the odds of a long-last response were 35% greater per unit increase of DURATION, on average. A factorial model was constructed to test whether differences between discrete levels were significant. The output of this model, which appears in Table 9, indicates that differences between DURATION 0 and DURATION 1, 2 and 3 were highly significant. These outcomes confirm the initial prediction (not supported in other conditions) that increasing DURATION would increase long-last groupings in the English group.

The results for the Competing condition in the Spanish group revealed that in this group as well, increasing DURATION increased long-last groupings at the expense of creak-last groupings. The best-fitting model for Spanish, shown in Table 10, included DURATION and CREAK as fixed effects and a random slope for CREAK. Adding a random slope for DURATION did not improve goodness-of-fit ($X^2 = 4.5791$, df = 3, $p = 0.2053$). No model which included the interaction DURATION*CREAK provided a better fit for the data. Table 10 indicates that the fixed effect of DURATION was highly significant. This can be seen in the left panel of Fig. 4b: adding a duration disparity of any magnitude while holding CREAK constant at levels 1 and 2 increased long-last groupings at the cost of creak-last groupings. Although the proportion of creak-last groupings was lower overall in the Competing condition in the Spanish group, Fig. 4b suggests a positive correlation between CREAK and creak-last responses between CREAK 1 and 3, at DURATION 1 and 3 (inverting the result for CREAK in Table 10; recall that the model measured creak-first responses). However, the fixed effect of CREAK was not significant.

In the factorial model constructed for the Spanish group, whose output appears in Table 11, differences between DURATION 0 and DURATION 1, 2 and 3 were highly significant, as they were in the English group.

Table 8
Output of best-fitting model predicting *long-last/creak-first* responses in the singly-varied and Competing conditions for English.

Random slopes		Var		SD	Cor
SUBJ (intercept)		0.028		0.168	
DURATION		0.069		0.263	0.90
(*n* = 2148)	Coef	OR	SE	*z*	*p*
Intercept	−0.350	0.70	0.086	−4.054	<0.0001
DURATION	0.298	1.35	0.069	4.334	<0.0001

Table 9
Factorial model predicting *long-last/creak-first* responses in the singly-varied and Competing conditions for English.

Random slopes	Var	SD	Cor		
SUBJ (intercept)	0.155	0.394			
DURATION 1	0.850	0.922	−0.75		
DURATION 2	0.777	0.882	−0.52	0.95	
DURATION 3	0.697	0.835	−0.31	0.84	0.97

(*n* = 2148)	Coef	OR	SE	*z*	*p*
Intercept	−0.547	0.58	0.130	−4.204	<0.0001
DURATION 1	0.651	1.92	0.232	2.808	0.0050
DURATION 2	0.911	2.49	0.226	4.026	<0.0001
DURATION 3	0.932	2.54	0.219	4.255	<0.0001

Table 10
Output of best-fitting model predicting *long-last/creak-first* responses in the singly-varied and Competing conditions for Spanish.

Random slopes		Var		SD	Cor
SUBJ (intercept)		0.158		0.397	
CREAK		0.050		0.223	−0.79
(n = 2411)	Coef	OR	SE	z	p
Intercept	−0.232	0.79	0.128	−1.812	0.0700
CREAK	−0.099	0.91	0.057	−1.721	0.0853
DURATION	0.194	1.21	0.039	4.970	<0.00001

Under the initial assumption that listeners would interpret both duration- and phonation-based cues as signalling group finality, these cues were expected to be in conflict in the Competing condition. I anticipated that outcomes in this condition might indicate whether one type of cue was a stronger predictor of responses than the other. The "top line" result in the Competing condition was that the presence of a duration disparity of any magnitude decreased creak-last groupings, as compared with the baseline singly-varied Phonation condition, indicating that listeners interpreted the presence of a duration disparity as a cue to group finality and that this mitigated any effects of varied phonation.

The statistical analysis did not associate significance with the fixed effect of CREAK in the Competing condition in either language group. However, a visual inspection of the charts in Fig. 4 suggest outcomes related to CREAK which deserve comment. The best-fitting model for the Spanish group included CREAK as a fixed effect. However, this effect was not significant or marginally significant at the 0.01 level in the model in Table 10, and the factorial model in Table 11 revealed no significant differences between CREAK 0 and CREAK 1, 2 or 3. Therefore, even though the direction of a small trend favouring creak-last groupings we see in the Competing condition for Spanish appears to be largely consistent with the baseline trend for varied phonation, the statistical analysis does not permit us to accept this result.

In the English group, that the best-fitting model did not include a term for CREAK*DURATION was somewhat surprising, given that the model output shown in Table 2 indicated that the interaction CREAK*DURATION approached significance (p = 0.0679) in the English comparison group in the main analysis. The visual inspection of Fig. 3a reveals no hint of an interaction

Table 11
Factorial model predicting *long-last/creak-first* responses in the singly-varied and Competing conditions for Spanish.

Random slopes	Var	SD	Cor		
SUBJ (intercept)	0.176	0.419			
CREAK 1	0.257	0.507	−0.65		
CREAK 2	0.428	0.654	−0.73	0.74	
CREAK 3	0.339	0.583	−0.39	−0.22	0.46
(n = 2411)	Coef	OR	SE	z	p
Intercept	−0.694	0.50	0.167	−4.150	<0.0001
CREAK 1	0.192	1.21	0.160	1.203	0.229
CREAK 2	−0.026	0.97	0.178	−0.145	0.885
CREAK 3	−0.160	0.85	0.169	−0.947	0.344
DURATION 1	0.702	2.02	0.131	5.350	<0.0001
DURATION 2	0.825	2.28	0.131	6.287	<0.0001
DURATION 3	0.747	2.11	0.130	5.713	<0.0001

in the English group when CREAK and DURATION were varied co-operatively, but we see the suggestion of an interaction in the Competing condition in Fig. 4a: creak-last groupings are flat or falling when CREAK and DURATION are varied in opposition, as compared with CREAK at DURATION 0. A model which included a random slope for DURATION, the fixed effects CREAK and DURATION, and the interaction DURATION*CREAK fit the data nearly as well as the model which did not include the interaction term (X^2 = 3.775, df = 1, p = 0.0520). In this model, the interaction was marginally significant, although close to the upper threshold of 0.06 for considering outcomes to be marginally significant at the 0.01 level (Coef. = 0.082, SE = 0.042, x = 1.942, p = 0.0521). However, this model was rejected as being above threshold.

6. General discussion

Findings reported in the broader phonetic literature indicate both that length and creaky phonation are associated with finality, and that listeners can use these cues to locate ends of prosodic constituents (see Sections 1 and 2). As the findings of rhythmic grouping studies described at the outset have been consistent with this, I expected to observe creak-last and long-last RGBs among English speakers in this study and that increasing the magnitude of the disparities between alternating syllables would strengthen these tendencies. Because phonation cues were synthesized to be consistent with phrase-final creak, I considered a creak-last RGB to be especially likely in both conditions whenever phonation was manipulated. In relation to duration, a long-last RGB was not the only possibility considered for the Spanish and English groups, as length is associated with stress as well as constituent endings in both languages. For reasons discussed in Section 3, I considered a long-first RGB to be a possible outcome in the Spanish group, at least when duration was varied singly, and in the competing condition. To the extent that the prediction of a long-last RGB in the singly-varied duration condition was supported, I expected the effects of varying phonation and duration to be additive in the Co-operating condition; as both types of cue converged in the same syllable position, they might be mutually enhancing. The reason for the Competing sequences was to test which of two finality-marking cues was more dominant when they were arranged in a conflicting pattern. I expected to find a creak-last RGB in this condition, and I thought that this bias (if confirmed) might affect the percept for duration, making listeners more likely to associate length with the nonfinal position, as this would be (to differing extents) consistent with the sound pattern of both English and Spanish. On the whole, I expected any duration- and phonation-based grouping biases to be stronger in the English than the Spanish group.

The results indicated at least limited support for the three predictions (long-last and creak-last RGBs, and stronger effects in the English group), but there were some surprises. Summarising, evidence of a creak-last RGB was observed in all varied phonation conditions in both language groups, except in the Competing condition in the English group. Varied duration was associated with long-last effects in the Competing condition in the English group, but there were no long-last effects in the singly-varied and Co-operating conditions.

In the Spanish group, long-last effects were observed in the Co-operating and Competing conditions in the Spanish group. This outcome conflicted with the opposite long-first bias found in the Spanish group when duration was varied singly.

More granularly, increasing creak increased creak-last groupings in the singly-varied Phonation condition, except when creaky voicing was increased to 50% of the vowel, at which level a reduction was observed in both language groups. As noted in Section 5, a plausible explanation for this reduction is that the half-creaky vowel might have seemed unnatural to listeners. As naturally produced phrasal creak can be quite extensive, I speculate that the reduction at creak 3 in this study may be related to difficulties with synthesizing naturalistic sounding creak over extended durations. The difference between the language groups was not significant in this condition, although the graph in Fig. 2a shows that the creak-last bias was a shade stronger in the Spanish group. The findings for varied phonation in this study increase our knowledge of how listeners use phonation based cues. The prior literature indicates that creaky phonation is canonically associated with endings of higher-order prosodic constituents (see Section 2). The results of this study indicate that listeners can readily use creak to segment sequences into smaller, foot or word sized units as well. The findings for varied phonation also answer a question posed by Redi and Shattuck-Hufnagel (2001) as to whether listeners interpret creaky voicing cues along a continuum or as a binary feature. Listeners' responses to incrementally lengthened creaky voicing cues in this study suggest that listeners interpret creaky voicing cues along a continuum.

Outcomes in the singly-varied Duration condition were surprising in that no consistent bias was found in the English group. Although a long-last RGB was predicted for the Spanish group, the long-first RGB actually observed was consistent with a considered alternative. These results did not confirm the prediction that grouping effects would be stronger in the English group. The failure to find a duration-based grouping effect in the English group was repeated in the Co-operating condition. As can be seen in Fig. 2a and by referring back to discussion in Sections 5.1 and 5.2, outcomes in the Co-operating and singly-varied Phonation conditions were not significantly different. This being the case, the study's results do not bear meaningfully on the expectation of additive duration- and phonation-based effects in the Co-operating condition. The outcomes in the Phonation and Co-operating conditions in the English group confirmed the prediction that phonation-based effects should be stronger than duration-based effects. In that light, it is interesting that this hierarchy was inverted in the Competing condition. Here, when creak was held constant at levels 1–3 in *ba*, lengthening *ga* increased long-last at the expense of creak-last groupings. The trends charted in the rightmost panel of Fig. 4a suggest a weak positive correlation between the magnitude of the duration disparity and long-last groupings (and a corresponding negative trend in creak-last groupings in the leftmost panel), and this trend was marginally significant in the statistical analysis.

Outcomes in the English group in this study do not fully replicate Kelly et al.'s (2014) findings for English. Those investigators observed both long-last and creak-last RGBs in their singly-varied conditions, and found that the effects of varied duration and phonation were additive in their Co-operating condition. On the other hand, the results for the Competing condition here are in line with Kelly et al.'s (2014) finding of a reduction in creaky-final groupings when the modal syllable was longer. The finding of a creak-last RGB in both studies increases our confidence that English speakers can use creaky phonation cues in segmenting rhythmic sequences into small, foot or word sized units.

Outcomes in the Spanish group were similar to those for the English group in two important ways. As noted, the pattern of responses in the Phonation condition was similar and provided evidence for a creak-last RGB of comparable magnitude. Second, introducing a duration disparity of any magnitude in the Competing condition significantly increased long-last (and reduced creak-last) groupings in both language groups. There were also interesting between-group differences. In the Spanish group, phonation-based effects were more pervasive, as they were found in the Competing as well as the Phonation and Co-operating conditions. The other differences had to do with duration. When duration was varied singly, increasing duration increased counter-ITL long-first groupings, apart from a decrease at duration level 3. The explanation offered in Section 5.1 for the long-first RGB was that listeners may have tended to associate length more with stress than with group finality in the singly-varied Duration condition. The proportion of long-first groupings at duration 3 was comparable to the control condition, and the suggested explanation for this reduction was that the disparity of 120 ms between vowels may have seemed unnatural to listeners, given that the difference was so out of scale with actual speech-based differences in Spanish (Ortega-Llebaria & Prieto, 2007, 2011).

In contrast to the singly-varied Duration condition, the outcomes in the co-varied conditions in the Spanish group were consistent with a long-last RGB, suggesting a shift in the percept associated with duration-based cues when duration and phonation were varied together in either pattern. This shift was particularly evident in the Co-operating condition, which presented one of the study's more interesting results: here, there was an interaction between the predictors such that the odds of a long-last grouping increased as creak and duration increased together. In the end, then, outcomes in the Spanish group in this condition were not additive, as predicted, but rather synergistic. Given the long-first RGB observed for singly-varied duration, we might have expected increases in duration to reduce creak-last groupings in the Co-operating condition, but this did not happen. The explanation offered in Section 5.2.1 for the opposite finding was that pairing creaky phonation – a strong indicator of group finality, with length – which *could* be associated with group finality, reinforced the latter interpretation. This type of effect has also been noted by Redi and Shattuck-Hufnagel (2001:427), who comment that "it may enhance contrasts which are more directly signaled by other cues such as duration lengthening and boundary tones". This outcome differed strikingly from the finding in the English group, where duration-based cues had no observable effect in the Co-operating condition. In the Competing condition, as noted, the presence of any duration disparity increased long-last and decreased creak-last groupings in the Spanish as in the English group. In other respects, duration-based effects differed by group in this condition. In the English group, unit increases in DURATION in the Competing condition slightly increased long-last/creak-first groupings, and the interaction CREAK*DURATION was marginally significant. There was no comparable incremental effect of DURATION in the Spanish group. Rather, it can be seen in Fig. 4b that increasing CREAK increased creak-last groupings in the Competing condition, although this trend fell short of significance in the statistical analysis. Overall, the effect of duration could be seen as subtractive in both languages in the Competing condition. The finding that outcomes for a feature can differ in the singly-varied and co-varied conditions need not be surprising: Crowhurst and Teodocio-Olivares (2014) and Crowhurst (2016) have made similar observations and suggest that humans' processing of phonetic features in rhythmic grouping tasks (as in many other tasks) is sensitive to the phonetic context. This is one of the lessons of the current study as well.

The study's outcomes raise several questions of special interest for further research. The first relates to why creaky voicing was a more robust cue than duration in English in the singly-varied and Co-operating conditions. More to the point, given that other studies have observed long-last RGBs amongst English speakers, to what might we attribute the English speakers' lack of sensitivity to duration-based cues in these conditions? One type of answer is that outcomes for varied duration have not always been consistent: Bhatara et al. (2013) report a significant long-last RGB among French speakers who grouped sequences of syllables in one study, but they did not replicate this effect in a second study with the same population. Jeon and Arvaniti (2016) did not find a strong long-last RGB for English speakers in a similar rhythmic grouping study, in which they were tasked with grouping nonlinguistic sequences of tones. In fact, in a "summation" condition in which the intensity of long tones was reduced to compensate for the perception that longer sounds are louder than comparable, shorter sounds, Jeon and Arvaniti observed an increase in counter-ITL long-first groupings. Differing outcomes for duration across studies suggest that duration-based grouping effects may be fragile and influenced by other cues in the rhythmic context.

Another potential explanation can be framed in relation to both auditory and attentional factors: it may be that listeners were attending more to phonation cues because these were most useful to them in the specific task they were given. I suggest that creaky voicing, especially as manipulated in this study, may have been a more straightforward cue to group finality than length, which is a more ambiguous cue in that it can occur in nonfinal as well as final positions. Moreover, as an anonymous Journal of Phonetics reviewer points out, the variations in voice quality employed in the study may have been more salient than the incremental differences in duration at the auditory level. If phonation-based cues were both more helpful and more noticeable, then the English speakers might have developed the strategy of listening for creaky phonation, whether consciously or not. The absence of a duration-based effect in the Co-operating condition suggests that attending to varied phonation was sufficient to identify group-final syllables, and that information about duration was not additionally helpful. In this scenario, there might have been some spillover to the Duration condition: if participants were listening for audible, helpful phonation-based cues, they might have tended to overlook more subtle duration-based cues in the singly-varied condition. That Kelly et al. (2014) observed an additive effect of varied duration and phonation in a group of English speakers from the same population (University of Texas students) indicates that we can expect to find sampling differences. Other studies associated with this research program (Crowhurst, 2016; Crowhurst & Teodocio-Olivares, 2014) have found a robust long-last RGB with English speakers, when varied duration was tested in a context that was also defined by varied intensity. However, intensity is associated with a different percept; it has been found to signal group onsets in rhythmic grouping tasks (see also Hay & Diehl, 2007 and Bhatara et al., 2013, for relevant speech-based studies). Varied together with intensity, durational differences as cues to finality might be more valuable than when duration is co-varied with creaky phonation in the co-operative pattern.

Given the proposal that English speakers may have tended to allocate their attentional resources to varied phonation and not to duration in the Co-operating condition, with a collateral effect in the Duration condition, what are we to make of the reversal in the Competing condition in which duration and not phonation had a significant influence? It seems that when they listened to sequences in which varied phonation and duration were at odds, participants were forced to shift their attention to information about duration. One question to ask is whether the modal portions of vowels might have made a greater contribution than the creaky portions of vowels to listeners' perception of duration. In the Competing condition, the modal portion of the vowel in the short syllable was shorter than in other duration-varying conditions (roughly 112 ms, 87.5 ms, and 62.5 ms, subtracting 10%, 30%, and 50% of the vowel's overall duration, 125 ms). If listeners assigned greatest weight to the modal portions of vowels in assessing relative duration, then duration disparities may have seemed more pronounced in the Competing condition than in other conditions, and these perceived differences may have been of a degree sufficient to produce a stronger result than in other duration-varying conditions. This general, "modal salience" explanation may provide additional insight into the different outcome in the Co-operating condition: if creaky, long vowels seemed shorter than they were, then the duration disparity between long, creaky vowels and short, modal vowels might not have seemed as great as in fact it was. The modal salience hypothesis proposed for the Competing condition does not explain why there was no effect of varied phonation in the Competing condition in the English group. However, it may be that when duration differences were below some unknown threshold, listeners relied primarily on phonation cues, and that their attention shifted to duration cues only when duration differences crossed this threshold. An anonymous Journal of Phonetics reviewer comments that indirect support for the idea that listeners may assign greater weight to modal than to creaky portions of vowels in assessing duration is provided by sound patterns which suggest that nonmodal phonation is not perceptually optimal (Gordon & Ladefoged 2001). As examples, the reviewer points out that vowels with nonmodal phonation tend to be longer than modally phonated vowels, and that creak tends to migrate to vowels that are phonetically longer for independent reasons, either because they carry stress or have contrastive length. Both types of pattern are richly attested among Zapotecan languages (see for example, Gerfen, 2013 on Coatzospan Mixtec and Sonnenschein, 2005 on San Bartolomé Zo'ogocho Zapotec).

The inclusion of a Spanish comparison group allowed for a test of the general hypothesis that RGBs can be shaped by the listeners' background and that speakers of different languages may respond differently to rhythmically varied stimuli. In the current study, the most interesting difference may be the different allocation of attentional resources, as indicated by the different effects associated with duration-based cues in the Spanish and English groups, and evidence for a possible shift in the percept associated with duration induced by phonation-based cues. The most similarity across the two language groups was the similar response to varied phonation. As noted at the outset, phrase-final creak does not appear to be as prevalent in Spanish and in English, and is rarely described in the phonetic literature. It was therefore interesting to find that phonation-based cues were most consistently associated with a rhythmic grouping bias in the Spanish group. As phrasal creak is less prevalent in Spanish than English, it might be that Spanish-speaking listeners tended to notice its presence even more than English speakers in the study. To my knowledge, this is the first study to show that creaky voicing is perceptually salient to Spanish speakers.

7. Concluding remarks

Finally, the results of the study must be considered in light of their contribution to a program of rhythmic grouping research inspired by the Iambic Trochaic Law, introduced at the beginning of Section 1. Psychologists have long studied the influence of rhythmically patterned acoustic features on listeners' impressions of grouping, as initial references to the historic research of Bolton and Woodrow will have signalled. The research programs of these and numerous other researchers who have explored the perception of rhythm studied not only the influence of varying the duration and intensity of sounds, but also pitch, tempo, and the duration of pauses between sound events (Bolton, 1894; Woodrow, 1909, et seq). The ITL, as conceived by Hayes (1995:80) selectively packages the generalisations that "[i]ntensity has a group-beginning effect: duration, a group-ending effect", in the words of Woodrow (1911:77), which Hayes considered to provide extra-linguistic support for his influential typology of metrical foot structures in which feet are either quantitatively balanced or asymmetric (Hayes, 1995:63–69). Asymmetric feet in Hayes' typology are iambs, which combine a light, canonically CV syllable with a foot-final heavy syllable, CVV or CVC, depending on the language. For Hayes, iambic feet are right-headed; he assumes that the primary phonological characteristic of stressed syllables in iambs is bimoraicity, which may be signalled by phonetic duration. On the other hand, balanced feet, he claims, are optimally trochaic. A literature too extensive to adequately cite here (but see Fletcher, 2010 for a review, and Patel, 2010, Chapter 3 for a review that encompasses the rhythm literature) has associated pitch as well as intensity with trochaic stress. Of potential relevance for the current study are findings regarding the influence of varied pitch on listeners' grouping preferences. Findings for pitch have been mixed (see Crowhurst et al., 2016 for a selective review). Woodrow did not observe consistent pitch-based grouping effects, and in fact, the fullest statement of his (1911:77) conclusion is: "Pitch, intensity and duration can no longer be looked upon as *stellvertretende* factors, any one of which may be substituted for either of the other two. The rôle of each in rhythm is radically different. Intensity has a group-beginning effect: duration, a group ending effect: pitch, neither a group-ending nor a group-beginning effect."

We now know that varied pitch is quite strongly associated with trochaic rhythm (higher pitch signalling group onsets) in humans and some other species that have been studied (Long-Evans rats, for example; see de la Mora et al., 2013), and some researchers classify higher pitch and increased intensity together in a reconceptualisation of the ITL (e.g. Bion et al., 2011; de la Mora et al., 2013). The distribution of pitch and pitch movements is a key determinant of linguistic rhythm, and connections can be considered between the F0 manipulations employed in this study, and linguistic tone. Specifically, as the pitch track in Fig. 1 in Section 4.2 illustrates, creaky syllables in this study begin with stable F0 at about 175 Hz, which drops radically in the final half, third, or 10% of the vowel, mimicking a falling tone. Jun (2014) builds a prosodic classification around the notion of linguistic macro-rhythm, which she defines as a tonal rhythm grounded in listeners' perceptions of changes in F0. In Jun's model, patterns are more strongly macro-rhythmic to the extent that (i) there is a regular alternation of similar tonal units, (e.g. L-H-L-H-L-H. . .L%), (ii) which is deployed over similar and highly regular sequences of sub-tonal units at the "micro-rhythmic" level (for example, feet and syllables), (iii) and when like contour tones participate in the rhythmic alternation. If a syllable with modal + creaky phonation simulated by dropping F0 can be compared to a syllable with falling tone, then the sequences used in this study would count as highly macrorhythmic in Jun's typology. If participants in this study interpreted creaky phonation as a boundary L tone so that recurrent units were perceived as small words or phrases, then the creak-last RGB observed in nearly all of the study's conditions would make sense. As F0 was higher at the onset of creak-last units, this percept could also potentially be reinterpreted in terms of pitch. So, in the end, were participating listeners using creaky voicing to locate group endings, or were they using higher pitch and/or falling pitch to identify group onsets and/or endings? It is not clear that responses to phonation cues in this study can be repackaged in this way. Phrase-final creak is not simply a correlate of low F0 at phrase boundaries, but can be independently planned. As Redi and Shattuck-Hufnagel (2001) note, evidence for this is that glottalisation is also observed where F0 is closer to the speaker's middle pitch register or even higher. To further explore these issues, future research might profitably investigate the perception of different types of glottalised phonation and the effects of their manipulation on listeners' perception of prosodic boundaries.

Acknowledgements

Thanks are due to Ms. Dinorah Fernandez Esquer (MA), who recruited and helped to test subjects in Obregon, Mexico, to Mr. Francisco Jesus Leyva Quintero (MA), Vice Rector of La Salle University, where the Spanish sessions were conducted, and to Scott Myers and Sally Ragsdale for their comments, questions, and statistical advice. I have also appreciated the thoughtful and constructive feedback provided by three anonymous Journal of Phonetics Reviewers and by Dr. Marija Tabain in her role as Associate Editor. Their insights have helped to make this a stronger paper.

Funding

The current research was funded by National Science Foundation of the United States of America grant BCS-1147959 (Megan Crowhurst, PI).

Appendix A. Response sheet

Bloque 1

1.	. . . ba ga ba ga ba . . .	Seguro/a:	si	no
2.	. . . ga ba ga ba ga ba . . .	Seguro/a:	si	no
3.	. . . ga ba ga ba ga . . .	Seguro/a:	si	no
4.	. . . ba ga ba ga . . .	Seguro/a:	si	no
5.	. . . ba ga ba ga ba ga . . .	Seguro/a:	si	no
6.	. . . ga ba ga ba . . .	Seguro/a:	si	no
7.	. . . ba ga ba ga ba . . .	Seguro/a:	si	no
8.	. . . ga ba ga ba ga ba . . .	Seguro/a:	si	no
9.	. . . ga ba ga ba ga . . .	Seguro/a:	si	no
10.	. . . ba ga ba ga ba . . .	Seguro/a:	si	no
11.	. . . ba ga ba ga ba ga . . .	Seguro/a:	si	no
12.	. . . ga ba ga ba . . .	Seguro/a:	si	no
13.	. . . ba ga ba ga ba . . .	Seguro/a:	si	no

Appendix B. Response data

B*aga* and *gaba* responses, expressed as absolute number of responses and proportion (*baga* responses) of the total. Singly-varied Duration/Creak conditions: *ga* is long/creaky.

Table B.1
English group.

	Response	(c) Competing condition			(b) Duration	(d) Cooperating condition		
3	*Baga*	89 (0.62)	85 (0.59)	86 (0.61)	74 (0.51)	92 (0.64)	96 (0.67)	90 (0.63)
(120 ms)	*Gaba*	55 (0.38)	59 (0.41)	56 (0.39)	70 (0.49)	51 (0.36)	47 (0.33)	54 (0.37)
2	*Baga*	87 (0.61)	77 (0.54)	87 (0.61)	80 (0.56)	87 (0.60)	98 (0.69)	83 (0.59)
(80 ms)	*Gaba*	56 (0.39)	66 (0.46)	55 (0.39)	64 (0.44)	57 (0.40)	45 (0.31)	57 (0.41)
1	*Baga*	79 (0.55)	74 (0.51)	75 (0.53)	71 (0.50)	91 (0.64)	98 (0.68)	81 (0.57)
(40 ms)	*Gaba*	64 (0.46)	70 (0.41)	67 (0.47)	71 (0.50)	52 (0.36)	46 (0.32)	62 (0.43)
0	*Baga*	(a) Control and Phonation condition ⇒			160 (0.56)	84 (0.59)	98 (0.68)	89 (0.62)
	Gaba				128 (0.44)	59 (0.41)	46 (0.32)	55 (0.38)
⇑ Duration	Creak ⇒	3 (50%)	2 (30%)	1 (10%)	0	1 (10%)	2 (30%)	3 (50%)
		Ga is long, *ba* is creaky				*Ga* is long and creaky		

Table B.2
Spanish group.

	Response	(c) Competing condition			(b) Duration	(d) Cooperating condition		
3	*Baga*	72 (0.44)	83 (0.50)	93 (0.57)	93 (0.57)	92 (0.56)	123 (0.77)	135 (0.83)
(120 ms)	*Gaba*	93 (0.56)	81 (0.50)	69 (0.43)	69 (0.43)	71 (0.44)	37 (0.23)	28 (0.17)
2	*Baga*	86 (0.52)	89 (0.56)	93 (0.58)	73 (0.46)	87 (0.55)	122 (0.75)	120 (0.74)
(80 ms)	*Gaba*	78 (0.48)	71 (0.44)	68 (0.42)	84 (0.54)	71 (0.45)	40 (0.25)	42 (0.26)
1	*Baga*	72 (0.43)	78 (0.49)	86 (0.54)	84 (0.54)	85 (0.56)	103 (0.64)	120 (0.74)
(40 ms)	*Gaba*	94 (0.57)	81 (0.51)	72 (0.46)	71 (0.46)	67 (0.44)	57 (0.36)	42 (0.26)
0	*Baga*	(b) Control and Phonation condition ⇒			190 (0.59)	97 (0.63)	110 (0.70)	108 (0.65)
	Gaba				132 (0.41)	56 (0.37)	48 (0.30)	58 (0.35)
⇑ Duration	Creak ⇒	3 (50%)	2 (30%)	1 (10%)	0	1 (10%)	2 (30%)	3 (50%)
		Ga is long, *ba* is creaky				*Ga* is long and creaky		

References

Abdelli-Beruh, N., Wolk, L., & Slavinn, D. (2014). Prevalence of vocal fry in young adult male American English Speakers. *Journal of Voice, 28*(2), 185–190.
Andreeva, B., Barry, W. J., & Steiner, I. (2007). Producing phrasal prominence in German. *Proceedings of the 16th Meeting of ICPhS*. (Saarbrücken, 6–10 August 2007).
Barr, D., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.
Barry, W., Andreeva, B., & Koreman, J. (2009). Do rhythm measures reflect perceived rhythm? *Phonetica, 66*(1–2), 78–94.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.
Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht, Holland/Riverton, USA: Foris Publications. https://doi.org/10.1515/9783110874020.
Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. Keating (Ed.), *Phonological structure and phonetic form. Papers in Laboratory Phonology III* (pp. 1–33). Cambridge: Cambridge University Press.
Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language and stimulus complexity affect rhythmic grouping of speech. *Journal of the Acoustical Society of America, 134*, 3828–3843.
Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54*, 123–140.
Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.21, retrieved 25 September 2016 from http://www.praat.org.
Bolton, T. (1894). Rhythm. *American Journal of Psychology, 6*, 145–238.
Carlson, R., Hirschberg, J., & Swerts, M. (2005). Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication, 46*(3), 326–333.
Crowhurst, M. (2016). Iambic-trochaic law effects among native speakers of Spanish and English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 1*(12), 1–41. https://doi.org/10.5334/labphon.42.
Crowhurst, M. J., & Teodocio-Olivares, A. (2014). The joint influence of duration and intensity on the subjective grouping of rhythmic speech among Zapotec and English speakers. *Phonology, 31*, 1–44.
Crowhurst, M. J., Kelly, N. E., & Teodocio, A. (2016). The influence of vowel glottalisation and duration on the rhythmic grouping preferences of Zapotec speakers. *Journal of Phonetics, 58*, 48–70. https://doi.org/10.1016/j.wocn.2016.06.001.
Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 24*, 423–444.
Ding, H., Jokisch, O., & Hoffmann, R. (2006, May). The effect of glottalization on voice preference. In: *Proceedings of speech prosody* (pp. 851–854).
Fletcher, J. (2010). The prosody of speech: timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (2nd ed., pp. 523–602). Malden, Mass.: Wiley-Blackwell.
Garellek, M., & Keating, P. (2015). Phrase-final creak: Articulation, acoustics, and distribution. Presented at the Annual Meeting of the Linguistic Society of America (Portland, OR, USA).

Garellek, M., & Seyfarth, S. (2016). Acoustic differences between English/t/glottalization and phrasal creak. In: *Proceedings of interspeech* (pp. 1054–1058).
Gerfen, H. (2013). *Phonology and phonetics in Coatzospan Mixtec* (Vol. 48). Springer Science & Business Media.
Gibson, T. (2017). The role of lexical stress on the use of vocal fry in young adult female speakers. *The Journal of Voice, 31*(1), 62–66.
Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics, 29*(4), 383–406.
Hay, J., & Diehl, R. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics, 69*, 113–122.
Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.
Henton, C., & Bladon, R. A. W. (1988). *Creak as a sociophonetic marker. Language, speech and mind: studies in honour of Victoria A. Fromkin*. London and NewYork: Routledge.
Hosmer, D. W., Jr, & Lemeshow, S. (2004). *Applied logistic regression*. John Wiley & Sons.
Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America, 124*, 2263–2271.
Jeon, H.-S., & Arvaniti, A. (2016). Rhythmic grouping in English, Greek and Korean: Testing the iambic-trochaic law. *Speech Prosody*. https://doi.org/10.21437/SpeechProsody. 2016-233.
Jun, S.-A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 520–539). Oxford: Oxford University Press.
Keating, P., & Garellek, M. (2015). Acoustic analysis of creaky voice. Presented at the Annual Meeting of the Linguistic Society of America (1/11/15, Portland, OR, USA).
Kelly, N. E., Crowhurst, M. J., & Cobb, C. (2014). The effect of duration and glottalization on the perception of rhythm. In: H. Leung et al. (eds.), *Proceedings of the north east linguistics society* 40 (pp. 215–232). Berkeley Linguistics Society, Inc.
Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics, 10*(2), 163–175.
Kusumoto, K., & Moreton, K. (1997). Native language determines parsing of nonlinguistic rhythmic stimuli. *Journal of the Acoustical Society of America, 105*, 3204.
Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01.
Molnar, M., Carreiras, M., & Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition, 152*, 150–159.
Molnar, M., Lallier, M., & Carreiras, M. (2014). The amount of language exposure determines nonlinguistic tone grouping biases in infants from a bilingual environment. *Language Learning, 64*(s2), 45–64.
de la Mora, D., Nespor, M., & Toro, J. M. (2013). Do humans and nonhuman animals share the grouping principles of the iambic–trochaic law? *Attention, Perception, & Psychophysics, 75*(1), 92–100.
Oliveira, G., Davidson, A., Holczer, R., Kaplan, S., & Paretzky, A. (2015). A comparison of the use of glottal fry in the spontaneous speech of young and middle-aged American women. *Journal of Voice, 30*(6), 684–687.
Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. *Amsterdam Studies in the Theory and History of Linguistic Science Series, 4*(282), 155.
Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in central catalan and castilian Spanish. *Language and Speech, 54*, 73–97.
Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.
Prieto, P., Vanrell, M. D. M., Astruc, L., Payne, E., & Post, E. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication, 54*(2012), 681–702.
R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/.
Rao, R. (2010). Final lengthening and pause duration in three dialects of Spanish. In: *Selected proceedings of the 4th conference on laboratory approaches to Spanish phonology* (pp. 69–82). Cascadilla Proceedings Project Somerville, MA.
Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics, 29*, 407–429.
Rice, C. C. (1992). *Binarity and ternarity in metrical theory: Parametric extensions* Unpublished doctoral dissertation. Austin: University of Texas.
Sonnenschein, A. H. (2005). *A descriptive grammar of San Bartolomé Zoogocho Zapotec* (Doctoral dissertation, Lincom München).
Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related durational patterns in English. *Journal of Phonetics, 28*, 397–440. https://doi.org/10.1006/jpho.2000.0123.
Turk, A., & Shattuck-Huffnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics, 35*, 445–472. https://doi.org/10.1016/j.wocn.2006.12.001.
Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America, 99*(6), 3782–3790.
Vos, P. (1977). Teporal duration factors in the perception of auditory rhythmic patterns. *Scientific Aesthetics/Sciences de l'Art, 1*(3), 183–199.
Wolk, L., Abdelli-Beruh, N. B., & Slavin, D. (2012). Habitual use of vocal fry in young adult female speakers. *Journal of Voice, 26*(3), e111–e116.
Woodrow, H. (1909). *A quantitative study of rhythm: The effect of variations in intensity, rate, and duration*. Columbia University Contributions to Philosophy and Psychology. Vol. XVIII, No. 1.
Woodrow, H. (1911). The role of pitch in rhythm. *Psychological Review, 18*(1), 54.
Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., et al. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition, 115*, 356–361.