Title:

Language dominance shapes non-linguistic rhythmic grouping in bilinguals

Authors:

Monika Molnar[a) 1], Manuel Carreiras [1,2], Judit Gervain[3,4]

[1]Basque Center on Cognition, Brain, and Language (BCBL), Paseo Mikeletegi 69, Donostia, SPAIN 20009

[2]IKERBASQUE. Basque Foundation for Science.

[3]Laboratoire Psychologie de la Perception, Université Paris Descartes, Sorbonne Paris Cité

[4]CNRS, Paris, FRANCE

[a)] Author to whom correspondence should be addressed.

Electronic mail: m.molnar@bcbl.eu

Abstract

To what degree non-linguistic auditory rhythm perception is governed by universal biases (e.g., Iambic-Trochaic Law; Hayes, 1995) or shaped by native language experience is debated. It has been proposed that rhythmic regularities in spoken language, such as phrasal prosody affect the grouping abilities of monolinguals (e.g., Iversen et al., 2008). Here, we assessed the non-linguistic tone grouping biases of Spanish monolinguals, and three groups of Basque-Spanish bilinguals with different levels of Basque experience. It is usually assumed in the literature that Basque and Spanish have different phrasal prosodies and even linguistic rhythms. To confirm this, first, we quantified Basque and Spanish phrasal prosody (Experiment 1a), and duration patterns used in the classification of languages into rhythm classes (Experiment 1b). The acoustic measurements revealed that regularities in phrasal prosody systematically differ across Basque and Spanish; by contrast, the rhythms of the two languages are only minimally dissimilar. In Experiment 2, participants' non-linguistic rhythm preferences were assessed in response to non-linguistic tones alternating in either intensity (Intensity condition) or in duration (Duration condition). In the Intensity condition, all groups showed a trochaic grouping bias, as predicted by the Iambic-Trochaic Law. In the Duration Condition the Spanish monolingual and the most Basque-dominant bilingual group exhibited opposite grouping preferences in line with the phrasal prosodies of their native/dominant languages, trochaic in Basque, iambic in Spanish. The two other bilingual groups showed no significant biases, however. Overall, results indicate that duration-based grouping mechanisms are biased toward the phrasal prosody of the native and dominant language; also, the presence of an L2 in the environment interacts with the auditory biases.

Introduction

Humans automatically tend to group successive auditory events into higher-level patterns, which gives rise to the sensation of rhythm. For instance, the isochronous and identical ticks of a clock are usually perceived as a unit of two sounds like "tick tock" (e.g., Bolton, 1894). It is currently debated whether the mechanism(s) that govern rhythm perception, specifically rhythmic grouping in spoken language and in non-linguistic sounds are overlapping or distinct (e.g., Bhatara, Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013; Crowhurst & Teodocio, 2014; Hay & Diehl, 2007; Hay & Saffran, 2012; de la Mora, Nespor, & Toro, 2012; Iversen, Patel, & Ohgushi, 2008; Nespor, Shukla, Vijver, Avesani, Schraudolf, & Donati, 2008; also see Peña, Bion, & Nespor, 2011 for the visual modality). The current study addresses this question by focusing on *non-linguistic tone grouping biases* in monolingual and bilingual listeners of Spanish and Basque.

One proposal holds that the auditory system is governed by a universal bias to perceive a sequence of non-linguistic items (e.g., pure tones) contrasting in duration as an *iambic*, i.e. prominence-final, rhythmic pattern (short-long or weak-strong), and a sequence of items contrasting in pitch or intensity as a *trochaic*, i.e. prominence-initial, rhythmic pattern (loud-soft or strong-weak) – a principle known as the Iambic-Trochaic Law (ITL; Hayes, 1995; Hay & Diehl 2007). However, recent findings may suggest otherwise. It has been demonstrated that non-linguistic duration-based rhythm preferences differ as a function of language background (Iversen et al., 2008; de la Mora et al., 2012; Yoshida, Iversen, Patel, Mazuka, Nito, Gervain, & Werker, 2010). For instance, when American English and Japanese listeners were presented with non-linguistic tones alternating in intensity, both groups showed a preference for grouping the tones into a loud-soft trochaic pattern as predicted by the ITL. However, when presented with tones alternating in duration, American English listeners preferred grouping the sounds into *short-long* rhythmic units, whereas Japanese listeners preferred to group the same tone sequences into the opposite *long-short* pattern (Iversen et al., 2008).

Why do Japanese and American English listeners exhibit different tone grouping biases? It has been proposed that the differences between the two languages' syntactic phrase structures and phrasal prosodies can induce such biases (e.g., Iversen et al., 2008). For instance, in English, a functor-initial language, function words (articles, prepositions, conjunctions, etc.) normally precede their syntactically related content words (nouns, verbs, adjectives). By contrast, in Japanese, a functor-final language, content words typically precede function words. For instance, the English phrase 'to Tokyo' has an opposite word order than the corresponding Japanese phrase 'Tokyo ni', in which the noun precedes the case marker. Relevant to the current study, phrase structures have acoustic consequences at the level of phrasal prosody in spoken language (Nespor et al. 2008, Gervain & Werker 2013). Functor-initial phrases (typical in functor-initial, also called Head-Complement[1] languages: English, Spanish, Dutch, etc.) are produced with a weak-strong (iambic) prosody because the function words are less prominent than the content words. Further, this phrase-final prominence is physically realized using a durational contrast (short-long). In reverse, functor-final phrases (typical in functor-final, Complement-Head languages, such as Japanese, Basque, etc.) show a strong-weak (trochaic) pattern, because the content words are more prominent than the function words. This phrase-initial prominence is realized as a pitch or less often an intensity contrast (high-low and loud-soft, respectively). Therefore, English listeners hear a shorter vowel in the frequently occurring function word followed by a longer vowel in the content words, for instance. It has been proposed that regular exposure to these

[1] According to the generative formalism, in a syntactic phrase, the Head is the word that defines the syntactic properties of the phrase (e.g., the preposition in a prepositional phrase, a verb in a verb phrase), and it selects its argument(s), i.e. the Complement (the noun in the prepositional phrase, the object in the verb phrase).

phrasal prosody patterns influences the automatic grouping biases of non-linguistic sounds (e.g., Iversen et al., 2008); hence English listeners' preference for the short-long pattern. Recent infant studies also support the idea that experience with native phrasal prosody affects the grouping preferences of linguistic or non-linguistic auditory information (e.g., Bion, Benavides-Varela, & Nespor., 2010; Yoshida et al., 2010).

However, other studies (e.g. Hay and Diehl, 2007) have found similar grouping preferences between speakers of prosodically dissimilar languages, e.g. French and English, which differ in their lexical stress patterns. How and to what extent language experience influences non-linguistic prosodic grouping is thus still not fully understood. Accordingly, the goal of the current study is to test whether supra-segmental information, specifically the duration patterns of consonantal and vocalic intervals associated to linguistic rhythm classifications and/or the prosody of syntactic phrases shape general auditory processing (e.g., non-linguistic tone grouping biases). This question is relevant not only in terms of the relationship between linguistic and non-linguistic auditory processing, but also because prosody is known to play a role in language processing and language acquisition. Previous studies assessing rhythmic grouping of non-linguistic tones focused on American English and Japanese listeners (Iversen et al., 2008; Yoshida et al., 2010).

Even though these two studies have raised the possibility that native phrasal prosody shapes general auditory processing, the evidence they provided to support this hypothesis is limited for two reasons. First, no acoustic measurements were provided to support the theoretical claim that phrase-final prominence is realized by different phrasal prosodies across the two languages. Second, the participants were recruited from and assessed in two different countries (United States and Japan) and with native languages with very different linguistic rhythms (stress-timed vs. mora-timed). For this reason, first, it is unclear to what degree cultural or other differences between the countries (e.g., exposure to different musical traditions or educational and writing systems) contributed to the observed results. Second, thus far only monolingual speakers of stressed-timed and syllable-timed languages (e.g., English, Dutch, and French) have showed preference for short-long grouping (e.g., Bell, 1977; Hay & Diehl, 2007; Hayes, 1995; Vos, 1977), while speakers of a mora-timed language (Japanese) have demonstrated the opposite long-short preference (Iversen et al., 2008; Yoshida et al., 2010). While it is well established that Spanish is a syllable-timed language, the rhythmic properties of Basque have been up for debate, as it exhibits the characteristics of both syllable-timed and mora-timed languages (Hurch, 2013; Mehler, Sebastian-Galles, & Nespor , 2004). Therefore, to understand more about the influence of phrasal prosody, linguistic rhythm, and perceptual grouping biases, acoustic measurements of Basque is also necessary.

In the current study, speakers of Spanish and Basque are considered. These two languages provide an ideal combination, because speakers of both languages are part of the same cultural community found in Northern Spain. At the same time, Basque and Spanish have important grammatical differences, as Basque is functor-final, but Spanish is functor-initial. To what extent their prosodic and rhythmic patterns are different has not yet been fully explored. In terms of rhythm, they have been primarily characterized as syllable-timed (e.g. Hurch, 2013), i.e. similar. But no acoustic measurements exist as to their phrasal prosody, which is predicted to be different on the basis of their grammatical differences (Nespor et al., 2008; Gervain and Werker, 2013).

Moreover, how other aspects of linguistic experience, like bilingual exposure interacts with auditory grouping biases is unclear. It is a relevant question, because more than half of the world's population is exposed to at least one second language (L2) on a regular basis (e.g., Grosjean, 1982), and whether regular exposure to an L2 affects auditory biases induced by experience with the prosodic characteristics of native speech (L1) has not been assessed before. Although it has been illustrated that bilinguals rely on strategies based on their dominant language to segment linguistic stimuli into lexical units (e.g., Cutler et al., 1992; Dupoux et al., 2009), the degree to which the segmentation of non-linguistic auditory material into rhythmic units is shaped by the dominant input has not been addressed. In the current study, four different types of listeners' auditory tone grouping preferences were tested (**Table 3**): *Spanish monolinguals*, who reported to use no L2 on a regular basis; *Spanish-dominant* bilinguals (L1=Spanish) with medium Basque proficiency; *Basque-proficient bilinguals* (L1= Spanish) with high Basque proficiency and frequent exposure to Basque; and *Basque-dominant* bilinguals (L1= Basque) who are also expert Spanish users. Given the geo-political status of the Basque Country, virtually no adult monolingual speakers of Basque are available; the majority of the Basque speakers use Spanish as a second language. The auditory grouping biases of Spanish monolinguals are expected to show a similar pattern to those of American English listeners (e.g., Iversen et al., 2008), given that the phrasal structures of the two languages are similar. If language dominance shapes auditory grouping biases, then a clear difference at least between the grouping preferences of the Spanish monolingual and the Basque-dominant listeners is expected. Moreover, if the amount of exposure to an L2 does not interfere with the development of auditory grouping mechanisms, then the other two bilingual groups (Spanish-dominant and Basque proficient) should exhibit biases based on their L1s, Spanish, at the same level as shown in Spanish monolinguals.

In order to test whether there is a link between basic auditory grouping and experience with native or dominant language rhythm, two experiments were designed. First, to confirm previous theoretical predictions, the phrasal prosody and the duration patterns used in the classification of languages into rhythm classes of the two target languages were quantified (Experiment 1). Because Spanish is a functor-initial language, Spanish phrases (e.g., *en la casa* ‘in the house’) are predicted to be accompanied by iambic prosody, realized as a durational contrast (i.e. short-long). By contrast, Basque, which is a functor-final language, is predicted to contain phrases (e.g., *garren gainean* lit.: fire on top of, i.e. “on top of the fire”) with the opposite, trochaic prosodic pattern, carried by a pitch and/or intensity contrast (high-low and/or loud-soft). To date, however, no empirical data have been presented to support this hypothesis. To test whether the two languages are indeed spoken with different phrasal prosodies, the prosodies of Basque and Spanish adpositional noun phrases (e.g., prepositional phrases for Spanish and postpositional phrases for Basque), in addition to verb phrases were quantified (Experiment 1a). In addition, in Experiment 1b the linguistic rhythms (as measured by the vocalic and consonantal duration patterns used in the classification of languages into rhythm classes*)* of the two languages were also measured (Ramus et al., 1999). While it is firmly established that Spanish is a syllable-timed language, Basque has been hypothesized to be both a syllable-timed and a mora-timed language due to its syllabic and syntactic structure (e.g., Hurch, 2013; Mehler et al., 2004). Further, different Basque dialects might be spoken with different rhythms. To clarify this potential issue, we measured the rhythmic characteristics of the Basque and Spanish dialects spoken in the bilingual Guipuzcoa region of the Basque Country, where the current research took place. In Experiment 2, the perceptual tone grouping preferences of Spanish monolinguals, and Spanish-dominant, Basque-proficient, Basque-dominant bilinguals were assessed in response to tones varying in duration (*duration condition*), and tones varying in intensity (*intensity condition*) using a procedure similar to Iversen and colleagues’ (2008).

Experiment 1: Acoustic measurements

Experiment 1a: Phrasal Prosody

The phonetic realization of prosodic cues in phrases has been measured in a limited number of languages only (e.g., French, Turkish, and German). Importantly to the present study, no such measurements for Basque and Spanish are currently available. Therefore, following the method described in Nespor et al. (2008) and Gervain & Werker (2013), phrasal prosodies of these languages have been quantified based on the relative pitch, intensity, and duration of the stressed vowels in adpositional phrases. Additionally, the same measures have been carried out in Basque verb phrases.

Method

Material

The phrases for the acoustic analyses were produced by Spanish-dominant and Basque-dominant speakers. Spanish phrases were produced by 4 Spanish-dominant female speakers, who acquired Spanish (L1) from birth, and learnt Basque (L2) in school settings not earlier than 7 years of age. The Basque material were elicited from 4 Basque-dominant females, who acquired Basque (L1) from birth, and learnt Spanish after 5 years of age. The speakers have used their dominant language as a primary mean of communication with family and friends. Sentences were recorded in a news-like declarative statement style in a soundproof booth (digitized at 14 kHz) using the Praat software (Boersma and Weenink, 2009).

The Basque and Spanish phrases for the acoustic analyses were selected based on previous work on phrasal prosody realizations across VO and OV languages by Nespor et al. (2008) and Gervain & Werker (2013). Fourteen adpositional phrases were chosen for Basque and Spanish separately because they reflect the typical word orders of the two languages (prepositions in the functor-initial language Spanish, and postpositions in the functor-final language Basque). Basque prepositions are not fully analogous to Spanish prepositions for several reasons. First, they are morphologically complex. For instance, the postposition *atzean* in the postpositional phrase *garren atzean* (lit.: ‘flame behind’, i.e. "behind the flame") can be decomposed as *atze-an* 'rear/back at' (cf. the Spanish preposition *en* in *en la casa* 'in the house', which is monomorphemic). Second, and as a consequence of the first property, these postpositions are phonologically heavy (e.g. they bear stress etc.). Consequently, they might behave differently from simple function words when their prosodic properties are considered. Because postpositions are morphologically complex in Basque and not fully analogous to the prepositions of Spanish, we also tested an additional 14 verb phrases in Basque in which a bare noun and a ‘light’ verb (semantically empty verb required by grammatical structure) appeared in the same prosodic phrase (e.g., *barre egin*, lit.: ‘laugh(ing) do’, i.e. “to laugh”). Following the procedure for phrasal prosody measurements described in Nespor et al. (2008), the target phrases were embedded in a single carrier sentence to ensure natural and uniform utterance-level intonation for each phrase.

The beginning and endpoint of each phrase within the sentences were marked using PRAAT. Stressed vowels in the phrases were coded in the order of appearance (marked by bold type and underscore in the example phrases), as **V1** and **V2** (e.g., Spanish noun phrase *p**o**r la n**o**che*, ‘for the night’; or Basque noun phrase *g**a**rren **a**tzean* lit.: ‘the flame behind’). It is important to note that in Spanish V1 corresponds to the syntactic Head, and V2 to the Complement; meanwhile in Basque, V1 corresponds to the Complement, and V2 to the Head. The complete list of the phrases and an example of the acoustic analyses (e.g., how vowel boundaries were marked) can be found in the Appendix.

Analysis

Similarly to Nespor et al. (2008), the following variables were measured for the entire phrases and the stressed vowels of the two constituents (V1 and V2) using a PRAAT script: (1) duration, (2) mean pitch, (3) pitch peak, and (4) intensity. The four values of V1 and V2 were then normalized by the corresponding values of the entire phrase in order to account for talker variability across sentences. Average normalized values for the four variables by language and phrase type (Spanish PreP, Basque PostP, Basque VP) and vowel type (V1 vs. V2) are presented in **Table 1**.

First, to verify whether the two types of phrases within Basque are accompanied by different prosodies, the average values obtained for the Basque postpositional and verb phrases were compared in a linear mixed effect model with the fixed effects of phrasal type (noun phrase vs. verb phrase) and vowel type (V1 vs. V2), and speakers were included as a random effect and intercept. None of the four models (for Intensity, Duration, Mean Pitch, Pitch Peak) revealed a significant effect of phrase type or speaker (all $F<1$; all $p>.1$). Effects of vowel position were observed in the Intensity ($F_{(1, 55)}=17.5$, $p=.001$; $B=.038$, $t=2.9$) and Duration ($F_{(1, 55)}=6.4$, $p=.013$; $B=.018$, $t=1.7$) conditions, indicating that the stressed vowel of the Complement, V1, carries prominence in these two dimensions in both phrase types. No other effects of vowel type were found (all $F<1$; $p>.1$; $t<1$).

As no effect of or interaction with phrasal type were observed across the Basque postpositional and verb phrases, the two types of phrases were pooled together in the second analysis contrasting Basque and Spanish phrases as illustrated in **Figure 1**. In order to test whether the V1 and V2 productions vary as a function of language, we ran linear mixed effect models on the normalized values of Intensity, Duration, Pitch Peak, and Mean Pitch separately, with the fixed effects of Language (Basque and Spanish) and Vowel Position (V1 and V2). Speakers were included as a random effect and intercept. All post-hoc analyses reported are Bonferroni corrected.

Results

Table 1. Average normalized values (and *standard deviation*s) for the two stress vowels (V1 and V2) of the phrases for intensity, duration, mean pitch, and pitch peak by language.

	Intensity		Duration		Mean Pitch		Pitch Peak	
	V1	**V2**	**V1**	**V2**	**V1**	**V2**	**V1**	**V2**
Basque VP	1.022 (*.0226*)	.977 (*.0484*)	1.32 (*.0371*)	1.13 (*.0321*)	.989 (*.0503*)	.996 (*.0573*)	.827 (*.0905*)	.838 (*.1218*)
Basque PostP	1.012 (*.0323*)	.979 (*.0679*)	1.32 (*.0543*)	1.14 (*.0386*)	.967 (*.0703*)	.984 (*.0664*)	.811 (*.1224*)	.845 (*.1601*)
Spanish PreP	1.003 (*.0301*)	1.005 (*.0414*)	.59 (.0*127*)	.81 (.03*7*)	.978 (*.0485*)	.896 (*.0889*)	1.08 (*.1131*)	.907 (*.1355*)

The mixed effect model conducted for Intensity has yielded a significant interaction of language and vowel position ($F_{(1,83)}$=4.6, *p*=.035; B=-.004, t=3.5). No other effects reached significance (all F<1; p>.1; t<1). Post-hoc analysis has revealed a nearly significant difference between the intensity of Basque vowels in V1 position as compared to the V2 position (p=.058), therefore Basque phrasal prosody (but not Spanish) is marked by higher intensity on the Complement.

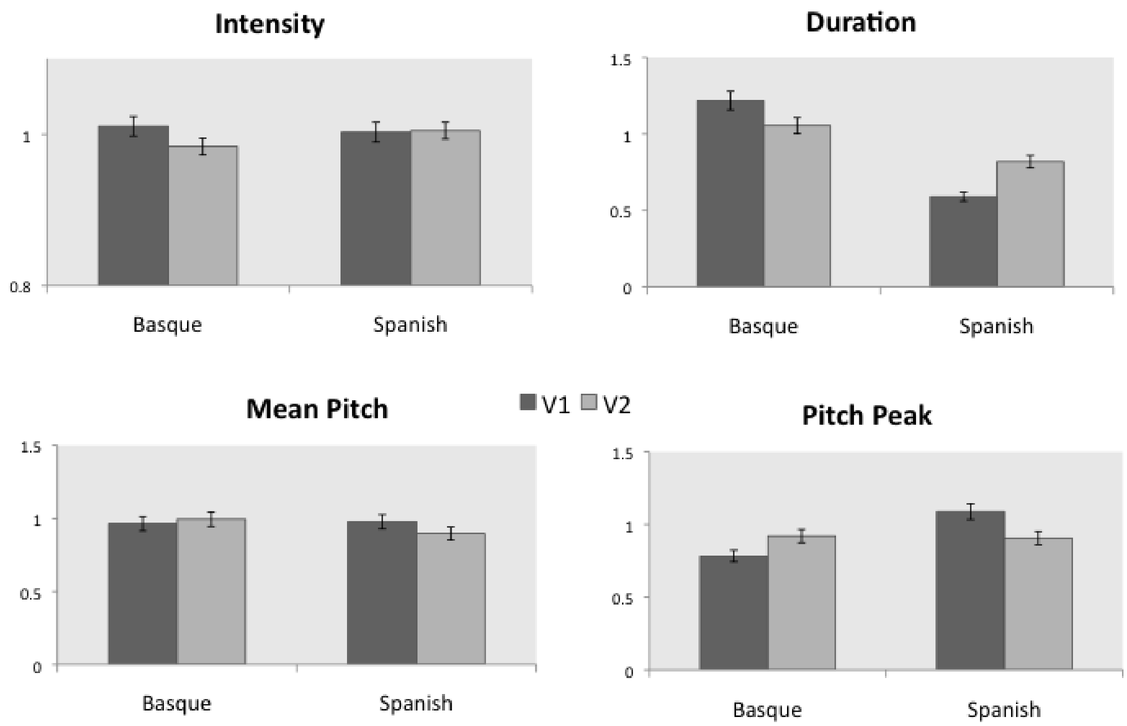


Figure 1. Normalized mean values for Intensity (dB), Duration (second), Mean Pitch (Hz), and Pitch Peak (Hz) as measured in Spanish and Basque phrases (VP and PostP phrases pooled together). V1 (dark grey bars) refers to the first stressed vowel in the phrase (Complement position in Spanish, and Head position in Basque), and V2 refers to the second stressed vowel in the phrase (Head position in Spanish, and Complement position in Basque).

A similar mixed effect model conducted for Duration has revealed a significant main effect of language ($F_{(1,83)}$=29.5, *p*=.001; B=-.08 (Spanish), t=7.2; and B=-.11 (Basque), t=15.1) due to longer vowel durations overall in Basque than in Spanish, and a significant interaction between language and vowel position ($F_{(1,83)}$=6.7, *p*=.012; B=-.042, t=-2.5). The random effect of speaker was not significant. Post-hoc analyses have yielded a significant difference between the Spanish and Basque values of V2 (p<.05), and the Spanish and Basque values of V1 (p<.01); also between the Spanish V2 and Basque V1 positions (p<.01), and the Spanish V2 and Basque V1 (p<.01). Consequently, in addition to the generally longer duration of vowels in Basque in this context, both languages use lengthening on the Complement (V1 in Basque, V2 in Spanish) as a cue to prosodic prominence.

The mixed effect model conducted for Mean Pitch has revealed no significant main effects or interactions. The analyses of Pitch Peak, however, has yielded a significant interaction between language and vowel position ($F_{(1,83)}$=15.15, p=.001; Β=-.21, t=3.9) without a significant effect of speaker. Bonferroni corrected post-hoc analyses have indicated a significant difference between vowel position within the Spanish phrases (p<.05), and importantly between the Basque and Spanish vowels in V1 positions (p<.01). Thus, this interaction is carried by the different pitch peaks of V1 in the two languages.

Discussion of Experiment 1a

Considering the overall findings of Experiment 1a, predictions regarding the phrasal prosody of Basque and Spanish are supported. The intensity values of stressed vowels in the Basque phrases follow the pattern predicted for functor-final, Complement-Head languages, that is the presence of trochaic (loud-soft) prosody within a phrase (Nespor et al., 2008). In addition, the duration values of the stressed vowels within both the Spanish and Basque phrases have produced significant differences. In Spanish, in line with the prosodic pattern of Head-Complement languages, the stressed vowels were longer in the Complements (V2), as opposed to in the Heads (V1) producing an iambic (short-long) prosodic pattern. In contrast, we also found lengthening on the stressed vowels in the Complement (V1) as opposed to the Head (V2) in Basque. A duration-based marking of phrasal prominence in Complement-Head languages was not predicted by previous accounts (Nespor et al., 2008). Nevertheless, the position of this prominence cue is on the Complement (V1) in Basque, therefore producing the expected trochaic pattern. No relevant pitch contrasts were observed between the prominent and non-prominent vowels in either of the languages. This is in line with other recent results from similar acoustic measures in Complement-Head languages (e.g. Japanese, Farsi, Korean etc., Gervain & Werker 2013), suggesting that a language typically only uses one of the trochaic acoustic cues, i.e. only intensity or only pitch, but not both. It is thus not surprising that Basque only shows intensity, but not a pitch contrast.

Whether the observed prosodic differences between Basque and Spanish alter auditory tone grouping biases in adult speakers is tested in Experiment 2. Before testing this, however, we also measured the rhythmic properties of the two languages, as it has been not ruled out thus far whether linguistic rhythm (also) interacts with grouping biases. If the language rhythm of Spanish and Basque are similar (e.g., syllable-timed) and different perceptual grouping biases are observed across Basque and Spanish listeners, then the grouping patterns are likely due the different phrasal prosodies of the two languages. If Spanish and Basque, however, belong to two distinct rhythm groups and perceptual grouping biases are observed, then further experiments are necessary to understand the link between grouping biases and experience with the rhythm and prosody of spoken language in the environment.

Experiment 1b: Language Rhythm

Material

Forty-nine sentences (25 in Basque, and 24 in Spanish) were recorded in a news-like declarative statement style by the same Spanish and Basque speakers described in Experiment 1a. Across languages, sentences were matched for the number of syllables (between 15 and 19 syllables), and approximately for meaning, following Ramus et al. (2000).

Measurement of rhythm

Among several measurement schemes developed for quantifying linguistic rhythm (e.g., Grabe and Low, 2002; Dellwo, 2006; Loukina et al., 2011; Ramus et al., 2000; Wiget et al., 2010), we opted for the proportion of vocalic intervals and the standard deviations of vocalic and consonantal interval durations (%V, ΔV, ΔC; Ramus et al., 2000), as well as the speech-rate normalized vocalic and consonantal intervals (VarcoV, VarcoC; Dellwo, 2006). As suggested by Wiget et al. (2010), it is best to use a combination of the above-mentioned two different types of metrics (e.g., to use %V in combination with VarcoV).

The steps of data processing were identical to the ones described in Ramus et al. (2000). First, the start and end points of each phoneme in the sentences were marked using PRAAT (Boersma and Weenink, 2009). Then, within each sentence the durations of vocalic and consonantal intervals were extracted. Based on these values, three variables per sentence were calculated: (1) the proportion of vocalic intervals (%V); (2) the standard deviation of the duration of vocalic intervals (ΔV); (3) the standard deviation of the duration of consonantal intervals (ΔC). Because it has been demonstrated that measuring speech across different language samples might skew the above-described variables, we have derived two additional measurements from the interval values that are normalized for speech rate (following the steps described in Dellwo, 2006; and White & Mattys, 2007): (1) the standard deviation of vocalic interval duration divided by the mean vocalic interval duration (VarcoV); (2) the standard deviation of consonantal interval duration divided by the mean consonantal interval duration (VarcoC).

Results

The summary of all five measurements are presented in **Table 2**. Different mixed effect model analyses (including speakers as a random factor and intercept) were conducted to compare the measures of %V, ΔV, ΔC, VarcoV, and VarcoC separately across the two languages. There was a significant difference between the scores for %V ($F_{(1,48)}$=14.68, p = .011; Β=6.28, t=3.83), and ΔV ($F_{(1,48)}$=6.47, p = .046; Β=1.29, t=5.71), without an effect of speaker. Further, no statistical significance was observed for the ΔC, VarcoV, and VarcoC conditions (all p > .1)

Language	**V inter's**	**C inter's**	**%V**	**ΔV**	**ΔC**	**VarcoV**	**VarcoC**
Basque	354	312	48 (*4.7*)	4.3 (*.76*)	4.41 (*.94*)	44.28 (*9.8*)	48.8 (*10.4*)
Spanish	335	340	41.4 (*5.5*)	2.7 (*1.3*)	4.46 (*1.05*)	41.93 (9.5)	48.9 (9.75)

Table 2. Measures of rhythm for Basque and Spanish: means and standard deviations (SD) for the proportion of vocalic intervals (V%), standard deviations of vocalic (ΔV) and consonantal intervals (ΔV), in addition to standard deviation of the speech-rate normalized values of the vocalic (VarcoV) and consonantal inervals (VarcoC). Average values in the last two columns (including SD values) are multiplied by 100 in order to simplify comparisons.

Figure 2 illustrates the location of Spanish and Basque, in addition to English, Spanish, Catalan, Japanese (Ramus et al., 2000), Tagalog (Bird et al., 2005), and Basque (Mehler, Sebastian-Galles, & Nespor , 2004) in the %V - ΔC plane. It has been proposed (e.g., Ramus et al., 2000) that arranging languages in the %V - ΔC plane appears to fit best with the traditional rhythm classes defined in linguistic theory. Two measurements are plotted for Spanish: 'SP' that represents the measurement of Ramus et al. (2000) collected from native monolingual Spanish speakers; and 'SP' underscored that illustrates the mean values obtained from Spanish-dominant bilinguals (L2=Basque) from the Basque Country in the current study. Similarly, two measurements are plotted for Basque: 'BQ' that represents the measurement presented in Mehler et al. (2004); and

'BQ' underscored that illustrates the mean values obtained from Basque-dominant bilinguals (L2=Spanish) from the Basque Country in the current study.

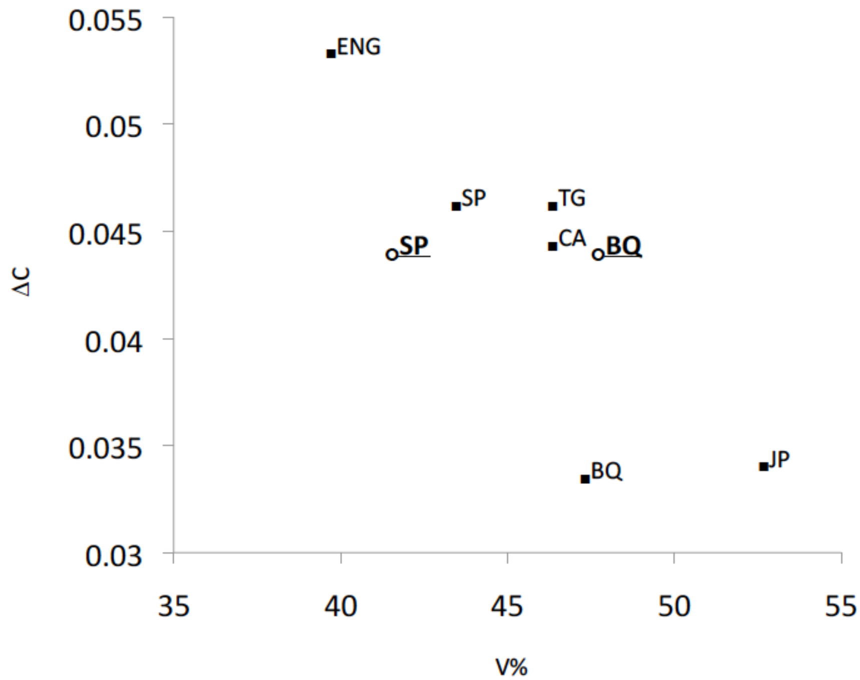


Figure 2. Locations of languages in the %V - ΔC plane. Data points from the present study are typed in bold. '**SP**' signals the mean values collected from Spanish-dominant bilinguals, and '**BQ**' represent the mean values obtained from Basque-dominant bilinguals. The rest of the languages are examples included from other studies: BQ (Mehler et al., 2004); ENG (English), SP (Spanish), CA (Catalan), JP (Japanese) from Ramus et al., 1999; and TG (Tagalog) is from Bird et al. (2005).

Discussion of Experiment 1b

The measures of linguistic rhythm performed in this experiment have placed Basque and Spanish in approximately the same space of the %V - ΔC plane, where languages traditionally classified as *syllable-timed* are found (Ramus et al., 2000). It is noteworthy, however, that the two languages are relatively far away from one another along the %V dimension, at the two extremes of the syllable-timed category, so they show statistically significant within-class differences along this dimension (see also Molnar et al. 2013 for a discussion of this point). In terms of Spanish measurements, the current values are similar to the ones obtained previously by Ramus et al., 2000. While highly similar for %V, the current measurements of Basque are inconsistent with previous measurements (Mehler et al., 2004) in terms of ∆C. Importantly, however, ∆C is a measure that correlates less well with other linguistic properties of languages. Indeed, %V has been more systematically used to determine rhythm class membership and correlations between linguistic rhythm and other language properties, such as word order and morphosyntactic type (Mehler et al. 2004).

Inconsistencies in rhythmic scores of the same language have been reported before (e.g., Grabe and Low, 2002; Ramus, 2002). Discrepancies across studies might be due to different speakers, linguistic materials, or protocols used for measuring the intervals (Wiget et al., 2010). In the present study, to avoid such potential sources of variations in rhythm scores across Basque and Spanish, the methodological recommendations of Wiget et al., 2010 were followed (e.g., combination of several metrics; multiple speakers; large sample of sentences, etc.). Another potential source of the inconsistency between the ∆C values across the current and previous Basque measurements might be that different dialects of Basque were recorded across the two studies, and/or the bilingual status of the speakers affected the current measurements. Importantly, however, in the current study the measurements of Basque and Spanish followed identical protocols, and provided evidence that the two languages represent rhythmically similar (syllable-timed) inputs, but with significant within-class differences (Molnar et al., 2013).

Taken together, Experiments 1a and 1b show that Basque and Spanish exhibit the predicted differences in terms of the position of prosodic prominence in their phonological phrases (Basque: trochaic, Spanish: iambic), even though the acoustic realization of this prominence involves not only intensity in Basque, but duration, as well. Spanish, as predicted, only uses duration as a cue to mark phrasal prosody. In terms of rhythm, the two languages fall within the same class, with some within-class differences. Whether these different phonological patterns result in different auditory grouping biases in the speakers of the two languages is tested below in Experiment 2.

Experiment 2: Perceptual tone grouping

Method

Stimuli

The stimuli consisted of 10-second-long tone sequences, constructed with similar parameters to the previous study assessing the non-linguistic rhythmic grouping preferences of English and Japanese listeners (Iversen et al., 2008). In the intensity condition, each sequence was built using either a 150 ms-long or a 250 ms-long complex base tone consisting of the fundamental (500 Hz) and the first three harmonics (1200, 1500, 2500 Hz). Each tone had a 15 ms period for rising in intensity at the beginning, and a 15 ms period for falling in intensity at the end. In each sequence, one of the base tones alternated with a second tone. Depending on the sequence, the second tone was either 1.5 or 2 times greater in amplitude than the base tone. In this way, four intensity sequences with the different tone ratios were created. In each sequence, the tones were separated by a short gap of 20 ms that produced either a 170 ms or a 270 ms stimulus onset asynchrony (SOA), depending on the base tone used in the sequence. For the duration condition, the sequences were created similarly to the intensity condition, with the exception that the alternating tones were constructed by multiplying the duration of the basic tones by 1.25, 1.5 and 3 (following the stimuli specifications of Iversen et al., 2008). This yielded 6 duration sequences with different tone ratios.

To assure that the sequences were fully ambiguous in terms of a iambic or a trochaic organization, the beginning and the end of each intensity and duration tone sequence faded in and out (at least over 2.5 s), in order to mask the starting tone order. In addition, for both conditions, each sequence was played forward (i.e. starting with the basic tone) and backward (i.e. starting with the alternating tone). This yielded 8 different intensity and 12 different duration sequences in total (similarly to Iversen et al., 2008). As further described in the procedure section below, in total 60 intensity sequences (seven times each sequence = 56, plus each of the original not reversed sequence once more = 4), and 60 duration sequences (five times each of the 12 sequence) were played separately in the intensity and duration conditions. In addition to the 60 target trials, 10 filler sequences were also inserted within each condition. The filler sequences were constructed of the same alternating tones as the target sequences, however the gaps in-between the tones varied between 20 and 120 ms, in such way that they produced both iambic and trochaic grouping of the tones within each sequence. The purpose of the filler sequences was to break the monotony of the difficult and repetitive target sequences. Responses to the filler sequences were not taken into account during analyses.

Procedure

Stimuli were presented through headphones using PRAAT presentation scripts. Half of the participants completed the Duration condition first, the other half of the participants completed the Intensity condition first. Within each condition, participants were familiarized first with the experiment by listening to three example sequences, then they heard the experimental sequences presented in a random order. In total, participants heard 60 test sequences and 10 filler sequences within each condition (duration and intensity). After listening to each sequence, participants were asked to indicate their perceived grouping of the tones (short-long or long-short in the Duration condition and loud-soft or soft-loud in the Intensity condition) by clicking on an icon schematically depicting the grouping types. The participants only received instructions at the very beginning of the task (in the language of their preference), and no verbal or written instructions appeared during the experiment. The testing session lasted for approximately 25 minutes. Participants were tested individually in a quiet room using headphones.

Participants

In total, 64 participants completed the experiment (mean age: 27.4 years, range: 21-35, 33 females). The participants formed 4 groups: Spanish Monolinguals, Spanish-dominant Bilinguals, Basque-proficient bilinguals, and Basque-dominant bilinguals. Considering their Spanish and Basque knowledge, the linguistic profile of each group is presented in **Table 3**. In addition, most participants reported to have learned a third language in school settings (e.g., English, French, Galician, and Catalan). The proficiency levels for each language were evaluated based on self-reported scores on a scale from 0 (= no knowledge) to 10 (=native proficiency). The exposure measures reflect the participants' self-report of the average exposure to the given language at the time of testing. The Spanish monolinguals were born in other regions of Spain than the Basque Country and moved to the Basque Country for study or work. The Spanish-dominant and Basque-proficient bilinguals lived in a Basque-Spanish bilingual region, and the Basque-dominant bilinguals lived in a Basque-speaking region of the Basque Country.

Group (N=64; n=16 per group)	**Average age (years) & gender distribution**	**L1**	**L2**	**L2 AoA (years)**	**Basque Proficiency (0-10)**	**Spanish Proficiency (0-10)**	**Basque exposure (%)**	**Spanish exposure (%)**
Spanish Monolinguals	27.6 (7 females)	Spanish	-	-	-	10	-	100
Spanish-dominant Bilinguals	27.2 (10 females)	Spanish	Basque	6.2	5.5	10	26	74
Basque-proficient Bilinguals	25.2 (9 females)	Spanish	Basque	4.8	9.1	10	58	42
Basque-dominant Bilinguals	29.5 (7 females)	Basque	Spanish	6	10	8.4	85	15

Table 3. The linguistic profile of the participants in Experiment 2. All average values reported here are based on participants' self-reports. The proficiency levels for each language are represented on a scale from 0 (= no knowledge) to 10 (= native proficiency). The exposure measures reflect the participants' average exposure to the given language at the time of testing.

Results

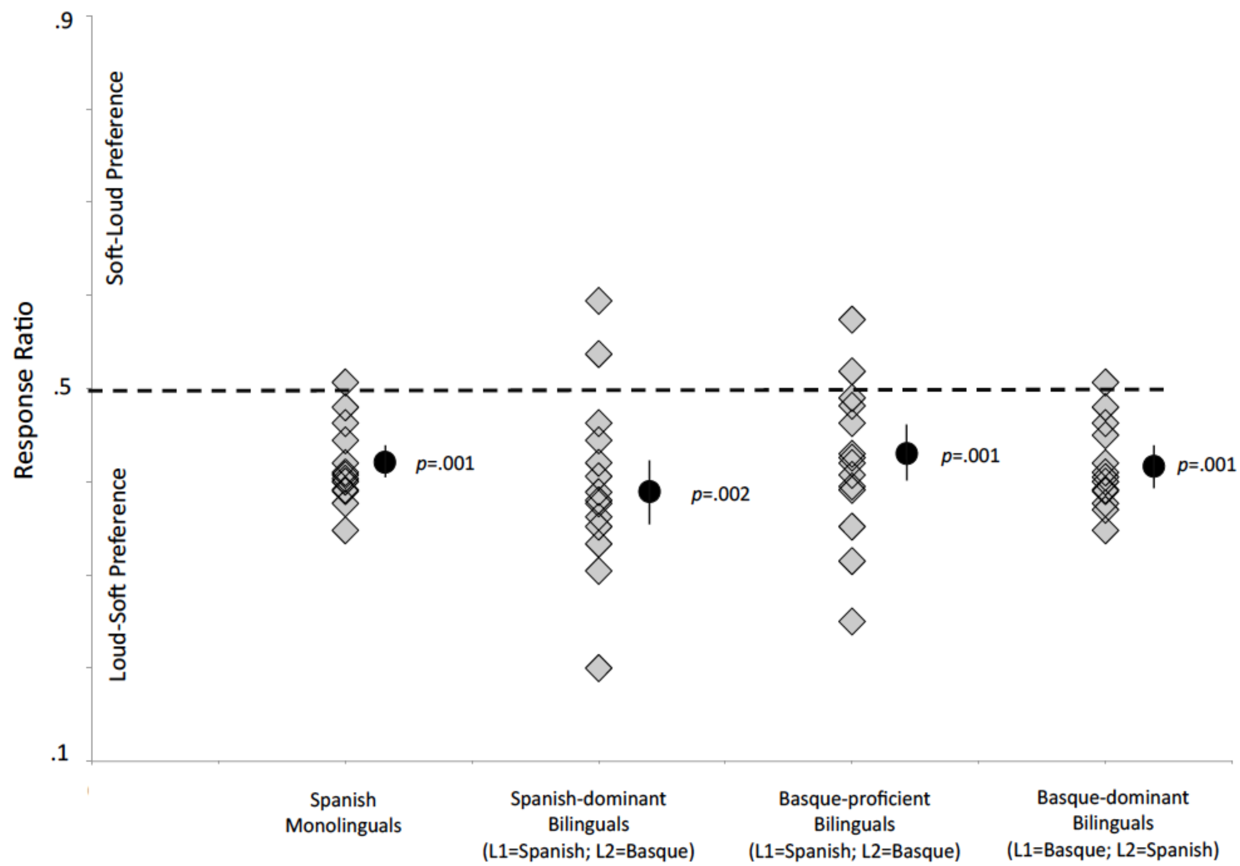
Responses were analyzed separately for the intensity and duration conditions. Because neither the basic tone duration (150 ms vs. 250 ms) or the difference ratio between the basic and the alternating tones (for the intensity condition: 1.5 and 2; for the intensity condition: 1.25, 1.75, and 3) had any effect on the participants' responses, responses across these stimulus dimensions were pooled together. Responses to the filler sequences were discarded. Based on the Shapiro-Wilk test of normality, all the data sets (with the exception of the Basque-dominant bilinguals' short-long preference scores) are normally distributed. Because of this, we opted for parametric statistical analyses.

Figure 3 illustrates the proportion of iambic responses of each participant within each language group for the Intensity (panel A), and the Duration (panel B) conditions. In the Intensity condition (**Figure 3A**), the number of soft-loud (iambic) vs. loud-soft (trochaic) responses were summed, and the proportion of iambic responses over the 60 experimental trials was calculated. The average responses of all groups' reflected a loud-soft bias (Spanish monolinguals: M=.41; SE=.01; Spanish-dominant bilinguals: M=.39; SE=.022; Basque-proficient bilinguals: M=.43; SE=.016; Basque-dominant bilinguals: M=.41; SE=.011). To assess whether there was a statistically significant bias toward the loud-soft grouping, t-tests on the response ratios per group was performed against chance level. Each group exhibited a significant bias toward the loud-soft grouping (Spanish monolinguals: $t_{(15)}$=8.1, p=.001; Spanish-dominant bilinguals: $t_{(15)}$=3.7, p=.001; Basque-proficient bilinguals: $t_{(15)}$=4.2, p=.001; Basque-dominant bilinguals: $t_{(15)}$=8.1, p=.001.) The average proportion of iambic responses per group was submitted to a one-way ANOVA with language group (*Spanish-monolinguals, Spanish-dominant bilinguals, Basque-proficient bilinguals, and Basque-dominant bilinguals*) as a between-subject factor. No main effect of group emerged ($F_{(3,60)}$=.81; p=.49).

In the Duration condition, the average proportion of short-long responses was analyzed the same way as the soft-loud proportion in the Intensity condition (**Figure 3B**). The average short-long responses for each group: Spanish monolinguals: M=.59; SE=.036; Spanish-dominant bilinguals: M=.52; SE=.027; Basque-proficient bilinguals: M=.47; SE=.027; Basque-dominant bilinguals: M=.4; SE=.034). The t-tests assessing each group's preference bias (short long vs. long-short) indicated that Spanish monolinguals exhibited a bias toward short long grouping, the Spanish-dominant and Basque-proficient bilinguals' biases reached no significance, and the Basque-dominant bilinguals exhibited a significant bias toward long short grouping (Spanish monolinguals:

$t_{(15)}$=8.1, p=.001; Spanish-dominant bilinguals: $t_{(15)}$=.64, p=.28; Basque-proficient bilinguals: $t_{(15)}$=1, p=.15; Basque-dominant bilinguals: $t_{(15)}$=2.9, p=.005.). Moreover, the ANOVA yielded a significant group effect ($F_{(3,60)}$=6.7; p=.001). Post-hoc analysis (Tukey HSD) revealed that the responses of the Spanish Monolingual group was significantly different from the responses of the Basque-proficient (p<.05) and the Basque-dominant bilingual groups (p<.01). No other group differences reached significance.

(A)



(B)

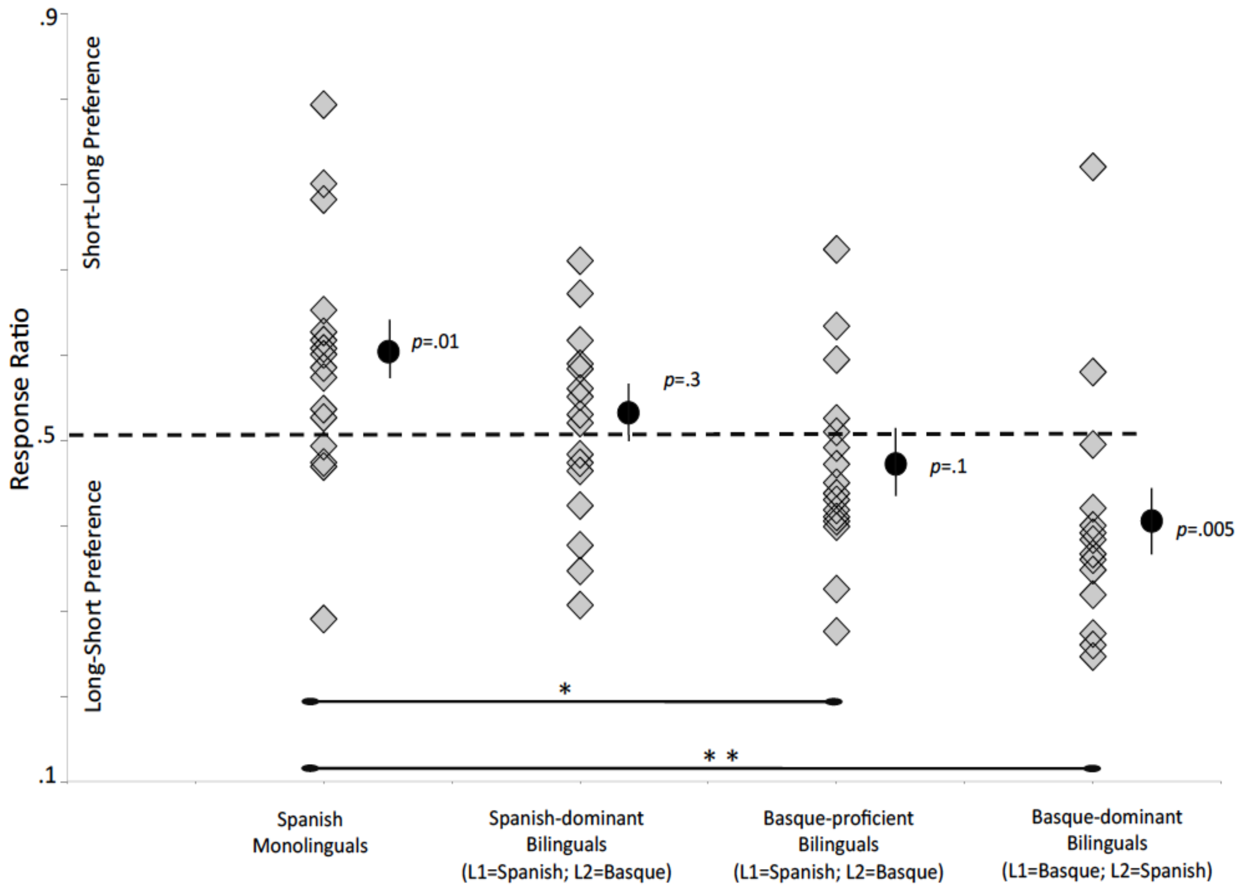


Figure 3. The results of Experiment 2. In each panel individuals' preference scores are depicted within each group. In Panel A, the proportion of Soft-Loud vs. Loud-Soft responses are displayed. In Panel B, the participants' Short-Long vs. Long-Short responses are illustrated in the Duration condition. Group averages and standard errors are marked by black dots. P-values indicate the significance of the difference from chance level (.5). Significant group differences are indicated by asterisks (* $p<.05$; ** $p<.01$).

Discussion of Experiment 2

In line with previous findings on tone grouping biases (e.g., Iversen et al., 2008; Vos, 1977), participants in the current experiment, regardless of their linguistic background, have displayed a trochaic (or loud-soft) grouping preference when listening to tones that differed in intensity. The Spanish listeners' trochaic bias is particularly relevant, because their native language utilizes no intensity contrast to mark phrasal prominence as shown in Experiment 1a. This and prior studies provides support for a universal trochaic grouping preference (in line with Hayes, 1995). Previously, perceptual grouping biases of adults (e.g., Iversen et al. 2008), infants (e.g., Bion et al., 2011) and rats (de la Mora et al. 2012) have all confirmed the universal presence of a trochaic bias, for the intensity or the pitch cue – even when the relevant cues are not extensively used by the native language.

By contrast, in the duration condition, native and dominant language-specific biases have been observed in two populations from similar cultural backgrounds and with overlapping linguistic rhythms (syllable-timed), but with clearly different phrasal prosodies across their L1s. Native Spanish listeners preferred iambic grouping, as opposed to native Basque and Basque-dominant listeners, who preferred the trochaic pattern. Even though the pattern exhibited by the Basque-dominant group is different than the Spanish monolingual pattern, the Basque-proficient group's trochaic bias did not reach significance. In addition, the Spanish-dominant group's pattern overall showed no difference from the Spanish monolingual pattern, however only the Spanish monolinguals' bias reached statistical significance when compared to chance level. Overall, these results suggest that listeners develop a trochaic grouping bias for duration-based alternating tones when their linguistic environment is Basque-dominant (e.g., L1=Basque). When both languages are present in the listener's environment in a more balanced way, no clear grouping preferences are observed at the group level. This suggests that regular exposure to an L2 influences auditory grouping biases. Whether it is the amount of exposure, or the proficiency level that primarily determines bilingual individuals' grouping biases is unclear based on the current data set, because no detailed and/or objective measures of the participants' proficiency and exposure was collected. Only self-reported proficiency and exposure values are available with a relatively small range within the groups, therefore no reliable correlation analyses could be conducted.

In either case, these results are the first to show such a fine-grained effect of the language environment, with a gradual change from a trochaic to an iambic preference as exposure/proficiency change from the exhaustive presence of an iambic language to increasing levels of dominance in a trochaic language.

General Discussion

The current study demonstrated that native Spanish and Basque listeners (from the same cultural background and geographical region), exposed to two languages with similar linguistic rhythms (syllable-timed) exhibit both universal and language-specific biases when their perceptual non-linguistic tone grouping preferences are observed. Specifically, while both groups showed a universal trochaic preference for tones contrasting in intensity, native Basque listeners had a trochaic preference, whereas Spanish listeners preferred the iambic grouping for tone sequences alternating in duration. Considering previous findings related to phrasal prominence and perceptual grouping (e.g., Bhatara et al., 2013; Bregman, 1990; de la Mora et al., 2012; Hay & Diehl, 2007; Hayes, 1995; Höhle et al., 2009; Iversen et al., 2008; Jakobsen et al., 1952; Nespor et al., 2008; Woodrow, 1909; Yoshida et al., 2010), we argue that these differences are observed because the two groups of listeners are exposed to different phrasal prosodies in their dominant languages, as demonstrated in Experiment 1a.

It should be noted that besides phrasal prosody another rhythmic regularity, lexical stress, is also present in spoken language. However, it is unlikely that the cross-linguistic rhythmic grouping biases observed in the current study were induced by differences in Spanish and Basque lexical stress. First, it has been suggested that phrasal prosody is perceived as a more dominant cue than lexical stress in continuous spoken language (Nespor et al, 1998). Second, most Spanish words are produced with a penultimate stress pattern, while Basque lexical stress is typically placed on either the second or the last syllable, but not on the first one. Therefore, the lexical items of both languages can be characterized by a weak-strong (iambic) pattern. This might explain the Spanish listeners' grouping biases, however cannot account for the Basque listeners' preference for grouping tones into a trochaic pattern.

In terms of the relationship between bilingual experience and perceptual tone grouping biases, several conclusions are apparent. First, as discussed above, the L1 plays a relevant role in the emergence of the trochaic duration-based bias, because at the group level, only the native Basque group (Basque-dominant bilinguals) exhibited a significant bias toward trochaic grouping. Second, the group differences partially suggest bilingual listeners' perceptual tone grouping is shaped by their dominant language (e.g., Basque-dominant bilingual group, or Basque-proficient group vs. Spanish monolingual group). This is in line with previous findings demonstrating that bilingual adults, when segmenting linguistic material, rely on strategies predicted based on their dominant (e.g., Cutler et al., 1992; Dupuox et al., 2009), or context-appropriate languages (de la Cruz-Pavía et al., 2014). However, at the group level, language-dominance did not clearly predict the performance of the Spanish-dominant group. Even though this group's preference pattern was not significantly different form the one of the Spanish monolingual group, the Spanish-dominant group exhibited no clear iambic bias in the duration condition. To understand more about how certain aspects of bilingual linguistic profiles (such as amount of exposure to L1 and L2 or proficiency level in L2) shape auditory biases, should be the focus of future research. Additionally, the role of the ongoing linguistic context (e.g., active language mode induced by linguistic material) should also be the interest of future research. In the current study, grouping biases were measured in a non-specific linguistic context, as no linguistic material was presented during the task, and the participants could interact in their preferred language with the experimenter. It is a possibility that inducing a specific linguistic context (Basque or Spanish) during the experiment would shift the Spanish-dominant and the Basque-proficient group responses toward the pattern predicted based on the context language. During the processing of linguistic material, such context effects have been observed in Spanish-Basque listeners previously (de la Cruz-Pavía et al., 2014).

Overall, non-linguistic perceptual grouping thus appears to show some universal biases. However these biases are readily modulated by language-specific patterns. This double-faceted nature of rhythmic grouping has been observed in the current study as well as in some of the previous studies (Iversen et al. 2008, Yoshida et al. 2010, Bhatara et al. 2013). On the basis of these results, we suggest here that the relevant question is not the simple dichotomy of whether auditory grouping biases are universal or language-specific, as both mechanisms seem to affect auditory grouping. Rather a more nuanced inquiry into the factors that influence how the two mechanisms interact within monolinguals and bilinguals should be the focus of future studies. For instance, certain stimulus properties or task demands might preferentially recruit one or the other mechanism or require interplay of the two. Future studies are therefore needed to systematically address these issues.

Furthermore, considering the current results from a broader perspective, it is in line with the idea of a shared and interactive auditory skills account of speech, sound, and music processing (for review, see Asaridou & McQueen, 2013; Kraus & Banai, 2007). It has been illustrated that low-level auditory skills (e.g., pitch or duration perception) might be enhanced by linguistic experience. For instance, native listeners of tone languages (e.g., Mandarin, in which pitch provides phonemically relevant information) outperform native listeners of non-tone languages (e.g., English) in non-linguistic pitch discrimination tasks (e.g., Guiliano et al., 2011; Bidelman et al., 2013; also see, Bent, Bradlow, & White, 2006). Similarly, native listeners of languages that use durational cues to distinguish between phonemes (e.g., Finnish, Japanese) exhibit enhanced skills in discriminating the duration of non-speech sounds (Marie et al., 2012; Tervaniemi et al., 2006). As reviewed earlier, previous studies regarding rhythmic grouping (e.g. Höhle et al., 2009; Bhatara et al., 2013) also found differences between speakers of different languages on the basis of diverging lexical stress patterns when phrasal prosody is otherwise similar. Therefore, linguistically relevant acoustic differences (e.g., pitch and duration of phonemes) appear to affect sound processing even in a non-linguistic context. In the present study, it has been demonstrated that auditory malleability, as a function of experience with native speech, extends beyond the boundaries of phonetic processing, and it is possible that speech perception interacts with automatic auditory biases related to prosody as well.

In sum, the Spanish monolingual and the Basque-dominant bilingual group exhibited opposite grouping preferences for durational contrasts in line with the phrasal prosodies of their native/dominant languages, trochaic in Basque, iambic in Spanish. These results imply that L1 listeners of Head-Complement vs. Complement-Head languages exhibit different preferences in prosodic grouping in the auditory dimension of duration, and the different preferences are likely due to listeners' experience in L1 (and L2) processing. The Spanish-dominant and Basque-proficient bilingual groups showed no significant biases, suggesting further that the amount of exposure and level of proficiency in an L2 with a different phrasal prosody than the L1 can affect non-linguistic auditory grouping preferences in gradual and subtle ways.

References

Abercrombie, D. (1967). *Elements of general phonetics* (Vol. 203). Edinburgh, Scotland: Edinburgh University Press, p. 97

Appel, R., & Muysken, P. (2006). *Language contact and bilingualism*. Amsterdam, Netherlands: Leiden University Press, pp. 153-163.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1-2), 46-63.

Bird, S., Fais, L., & Werker, J. (2005). The phonetic rhythm/syntax headedness connection: Evidence from Tagalog. *Acoustical Society of America Journal*, *117*, 2457-2457.

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language affects rhythmic grouping of speech. *The Journal of the Acoustical Society of America*, *134*(5), 3828-3843.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 97.

Bernard, C., & Gervain, J. (2012). Prosodic cues to word order: what level of representation? *Frontiers in Psychology*, *3*(451), 1-6.

Bion, R. A., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and speech*, *54*(1), 123-140.

Bolton, T. L. (1894). Rhythm. *The American Journal of Psychology*, *6*(2), 145-238.

Bosch, L., & Sebastián-Gallés, N. (2001a). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*(1), 29-49.

Bosch, L., & Sebastián-Gallés, N. (2001b). Early language differentiation in bilingual infants. *Trends in bilingual acquisition*, 71-93.

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*,*65*(1), 33-69.

Boersma, P., & Weenink, D. (2009). Praat. *Doing phonetics by computer.[computer program]. Retrieved March*, *31*, 2009.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, *21*(3), 343-348.

Bregman, A. S. (1990). Auditory scene analysis: the perceptual organization of sound. 1990.

Christophe, A., Nespor, M., Teresa Guasti, M., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head- direction parameter. *Developmental Science*, *6*(2), 211-220.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*,*40*(2), 141-201.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, *92*, 81-104.

Crowhurst, M. J., & Teodocio, A. (2014). Beyond the Iambic-Trochaic Law: the joint influence of duration and intensity on the perception of rhythmic speech. *Phonology*, *31*(01), 51-94.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for ΔC. *Language and language-processing*, 231-241.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress “deafness”. *The Journal of the Acoustical Society of America*,*110*, 1606.

de la Cruz-Pavía, I., Elordieta, G., Sebastián-Gallés, N., & Laka, I. (2014). On the role of frequency-based cues in the segmentation strategies of adult OV-VO bilinguals. *International Journal of Bilingual Education and Bilingualism*, *18*(2), 225-241.

de la Mora, D. M., Nespor, M., & Toro, J. M. (2012). Do humans and nonhuman animals share the grouping principles of the iambic–trochaic law? *Attention, Perception, & Psychophysics*, 1-9.

Gervain, J., & Werker, J. F. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass*, *2*(6), 1149-1170.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology*, *7*(515-546).

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Harvard University Press.

Hannon, E. E. (2009). Perceiving speech rhythm in music: Listeners classify instrumental songs according to language of origin. *Cognition*, *111*(3), 403-409.

Hay, J. S., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Attention, Perception, & Psychophysics*, *69*(1), 113-122.

Hay, J. F., & Saffran, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy*, *17*(6), 610-641.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago, USA: University of Chicago Press, pp. 75-86.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*,*32*(3), 262-274.

Hurch, B. (2013). Is Basque a syllable-timed language?. *Anuario del Seminario de Filología Vasca" Julio de Urquijo"*, *22*(3), 813-825.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America*, *124*, 2263-2271.

Jacobsen, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis: the distinctive features and their correlates*.

Lerdahl, F. A., & Jackendoff, R. S. (1983). *A generative theory of tonal music*. The MIT Press.

Lin, H., & Wang, Q. Mandarin Rhythm: An Acoustic Study. *Journal of Chinese Language and Computing*, *17*(3), 127-140.

Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *The Journal of the Acoustical Society of America*, *129*, 3258-3270.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477.

Marie, C., Kujala, T., & Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex*, *48*(4), 447-457.

Mehler, J., & Christophe, A. (2000). Acquisition of languages: Infant and adult data. *The Cognitive Neurosciences*, 879-908.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 101-116.

Mehler, J., & Nespor, M. (2004). Linguistic rhythm and the acquisition of language. In *Structures and Beyond: Volume 3: The Cartography of Syntactic Structures, ed. A. Belletti*. Vol. 3. Oxford, UK: Oxford University Press, pp. 213-223.

Mehler J, Sebastian-Galles N, Nespor M. (2004). Biological foundations of language: language acquisition, cues for parameter setting and the bilingual infant. In The New Cognitive Neuroscience, ed. M Gazzaniga. Cambridge, MA (USA): MIT Press, p. 825-836.

Mok, P. P. K. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics*, *2*, 148-154.

Molnar, M., Gervain, J., & Carreiras, M. (2013). Within- rhythm Class Native Language Discrimination Abilities of Basque- Spanish Monolingual and Bilingual Infants at 3.5 Months of Age. *Infancy*.

Morgan, J. L., & Demuth, K. (1996). Signal to syntax: An overview. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 1-22.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, *41*(1), 233-243.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance*, *24*(3), 756.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e linguaggio*, *7*(2), 139-168.

Nespor, M., Guasti, M. T., & Christophe, A. (1996). Selecting word order: The rhythmic activation principle. *Interfaces in phonology*, 1-26.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35-B45.

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, *119*(5), 3034-3047.

Peperkamp, S., Dupoux, E., & Sebastián-Gallés, N. (1999). Perception of stress by French, Spanish, and bilingual subjects. In *EUROSPEECH*.

Pena, M., Bion, R. A., & Nespor, M. (2011). How modality specific is the iambic–trochaic law? Evidence from vision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1199.

Ramus, F., Nespor, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition*, *75*(1), AD3-AD30.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America*, *105*, 512-521.

Shukla, M., & Nespor, M. (2010). Rhythmic patterns cue word order. *The sound patterns of syntax*, 174-188.

Sebastian-Galles, N. (2010). Bilingual language acquisition: where does the difference lie?. *Human Development*, *53*(5), 245-255.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition*, *115*(2), 356-361.

Vos, P. G. (1977). Temporal duration factors in the perception of auditory rhythmic patterns. *Scientific Aesthetics*, *1*(3), 183-199.

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in cognitive sciences*, *12*(4), 144-151.

Werker, J. F., Byers-Heinlein, K., & Fennell, C. T. (2009). Bilingual beginnings to learning words. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3649-3663.

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How Do Infants Become Experts at Native-Speech Perception?. *Current Directions in Psychological Science*, *21*(4), 221-226.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm?. *The Journal of the Acoustical Society of America*, *127*, 1559-1569.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501-522.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665-679.

Woodrow, H. H. (1909). *A quantitative study of rhythm: The effect of variations in intensity, rate and duration* (No. 14). Science Press.