CROSS-LINGUISTIC DIFFERENCES IN THE USE OF DURATIONAL CUES FOR THE SEGMENTATION OF A NOVEL LANGUAGE

SHORT TITLE: CROSS-LINGUISTIC DIFFERENCES IN PROCESSING OF LENGTHENING

MIKHAIL ORDIN[1,2*], LEONA POLYANSKAYA[3], ITZIAR LAKA[4] MARINA NESPOR[5]

[1]Basque Centre on Cognition, Brain and Language, Spain

[2]IKERBASQUE – Basque Foundation for Science, Spain

Donostia, 20009, Paseo Mikeletegi 69, BCBL

+34-943309300, ext. 228

m.ordin@bcbl.eu

[3]Università degli Studi di Padova, Italia

Dipartimento di Studi Linguistici e Letterari (DiSLL)

via B. Pellegrino 26, 35137 Padova

leona.polyanskaya@gmail.com

[4]University of the Basque Country UPV/EHU, Spain

Department of Linguistics and Basque Studies (HEIS)

Paseo de la Universidad, 5, 01006 Vitoria-Gasteiz, Álava

itziar.laka@ehu.eus

[5]SISSA, Cognitive Neuroscience Sector, Laboratory of Language, Cognition and Development, Italy

Via Bonomea 265, 34136, Trieste, Italia

(+39) 040 3787624

marina.nespor@sissa.it

[*]Corresponding author (m.ordin@bcbl.eu)

Abstract

It is widely accepted that duration can be exploited as phonological phrase final lengthening in the segmentation of a novel language, i.e., in extracting discrete constituents from continuous speech. The use of final lengthening for segmentation and its facilitatory effect has been claimed to be universal. However, lengthening in the world languages can also mark lexically stressed syllables. Stress-induced lengthening can potentially be in conflict with right edge phonological phrase boundary lengthening. Thus the processing of durational cues in segmentation can be dependent on the listener's linguistic background, e.g. on the specific correlates and unmarked location of lexical stress in the native language of the listener. We tested this prediction and found that segmentation by both German and Basque speakers is facilitated when lengthening is aligned with the word final syllable and is not affected by lengthening on either the penultimate or the antepenultimate syllables. Lengthening of the word final syllable, however, does not help Italian and Spanish speakers to segment continuous speech, and lengthening of the antepenultimate syllable impedes their performance. We have also found a facilitatory effect of penultimate lengthening on segmentation by Italians. These results confirm our hypothesis that processing of lengthening cues is not universal, and interpretation of lengthening as a phonological phrase final boundary marker in a novel language of exposure can be overridden by the phonology of lexical stress in the native language of the listener.

Introduction

One of the key issues in language acquisition research is the identification of the mechanisms people exploit to segment continuous speech into discrete sequential constituents, like words, phrases and sentences. The parsing involves exploiting a wide range of cues. As the segmentation cues are integrated hierarchically, listeners have to assign different weights to each cue. When segmenting speech in their native language, adult listeners assign the highest weights to lexical-semantic and syntactic information (Mattys, White, & Mehlhorn, 2005). When listeners are processing speech in a novel language, or in a language in which they are not very fluent, this information is not always available. A whole body of studies have shown that although the most powerful and informative cues are not available to people segmenting speech in a novel language, they can nevertheless successfully cope with segmentation tasks (Wakefield, Doughtie, & Yom, 1974; Pilon, 1981). In the absence of higher-level linguistic information, listeners rely on other cues, including segmental (phonotactic, allophonic) and prosodic (duration, intensity, pitch) cues, which signal lexical stress, as well as other levels of prominence, and phrase boundaries (Vroomen, Tuomainen, & de Gelder, 1998; Toro, Pons, Bion, & Sebastián-Gallés, 2011; Langus, Marchetto, Bion, & Nespor, 2012; Ordin & Nespor, 2013). Differences in transitional probabilities (TPs) between adjacent syllables within words or straddling the word boundaries are also used to segment words from an artificial language (Saffran, Newport, & Aslin, 1996), as well as frequency distribution of more- and less frequent speech constituents (de la Cruz-Pavia, Elordieta, Sebastián-Gallés, & Laka, 2014; Gervain, Sebastian-Galles, Diaz, Laka, Mazuka, Yamane, Nespor, & Mehler, 2013).

Among the different prosodic boundary cues for segmentation purposes, much attention has been paid to investigating the use of duration. Duration is one of the most reliable and consistent boundary cues that mark the end of a phrase. Final lengthening - the increase in duration of syllables and segments in the vicinity of the right edge boundary with lengthening proportional to the boundary strength (Turk & Shattuck-Hufnagel, 2007; Whightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992) – has been found in many languages and is deemed to be universal. Christophe, Peperkamp, Pallier, Block, and Mehler (2004) and Saffran et al. (1996) found that adults and infants are sensitive to final lengthening and are likely to use it for segmentation purposes. This was further verified in a number of studies (Tyler & Cuter, 2009; Kim, Broersma, & Cho, 2012; Langus et al., 2012 among others). In these studies, participants had to segment streams of an artificial language. The words in the

language were constructed using a limited inventory of syllables and concatenated without inserting pauses so that TPs between adjacent syllables within words were the only cue to word boundaries, being higher than TPs between adjacent syllables straddling word boundaries. Therefore, dips in TPs marked the boundaries between statistical words. As shown in Aslin, Saffran, and Newport (1998), this information is sufficient for the purposes of segmentation of an artificial language when no other cues are implemented into the speech stream. Adult listeners, as it was expected, could reliably segment speech into statistical words. When the final syllable of a word received lengthening, segmentation was facilitated compared to the TPs-only – i.e. the no prosody - condition. The facilitatory effect of final lengthening has been documented for speakers of Dutch, English, French and Korean (Kim, Broersma, & Cho, 2012; Tyler & Cutler, 2009). It was suggested that final lengthening facilitates segmentation because it is universal and easily detectable. In addition, lengthening associated with the right edge of a constituent is not only a linguistic phenomenon. It is also observed in processing non-linguistic streams, e.g., music (Palmer, 1997), and it is linked to a more general mechanism known as the iambic-trochaic law (ITL) that defines the preference to group elements of continuous streams into units with the longer element in final position (Nespor, Shukla, van de Vijver, Avesani, Schraudolf, Donati, 2008; Hay & Diehl, 2007; Bion, Benavides-Varela, & Nespor, 2011). This grouping preference is based on general auditory mechanisms, and is even present in the visual modality (Peña, Bion, & Nespor, 2011). Due to its universal nature, final lengthening was hypothesized to be exploited for the segmentation of a novel language regardless of the first language of the listener (Kim et al., 2012; Tyler and Cutler, 2009).

However, a number of cross-language phonetic and phonological differences suggest that lengthening also exhibits cross-language differences in functionality, and these differences suggest that durational cues may be processed differently cross-linguistically. Lengthening is used to signal syntagmatic prominence at the word as well as at the phonological phrase level (Gussenhoven, 2004). The use of prominence for segmentation has been very well documented (e.g., Cutler and Norris, 1988; Cutler, Mehler, Norris, & Segui, 1992), and languages differ in 1) how prominence is manifested cross-linguistically, and 2) in the most frequent (i.e., unmarked) location of prominent syllables within words and within phonological phrases. In addition, durational contrasts are used in some languages to make phonemic distinction between short and long vowels as well as short and long consonants (Ladefoged & Maddieson, 1995). Finally, it is worth taking into account that although the presence of final lengthening has been attested cross-linguistically, its phonetic implementation

and the domain over which final lengthening operates is language-specific (Nakai et al., 2012 for Finnish; White & Turk, 2010; Turk & Shattuck-Hufnagel, 2007; Whightman et al., 1992 for English; Cambier-Langeveld, Nespor, & van Heuven, 1997 for Dutch; Frota, 2000 for Portuguese; Elodrieta, Frota, & Vigário, 2005 for Spanish and Portuguese; D'Imperio, Elordieta, Frota, Prieto & Vigário, 2005 for other Romance languages). Such cross-linguistic differences in the functional load and acoustic manifestation of lengthening cues lead some researchers to suggest that lengthening cues for the segmentation of an unknown language might be language-specific and, at least to some extent, depend on the first language (L1) of the listener (Ordin & Nespor, 2013; Ordin & Nespor, 2016; de la Mora, Nespor, Toro, 2013; Toro & Nespor, 2015; Bhatara, Boll-Avetisyan, Unger, Nazzi & Hoehle, 2013). Ordin and Nespor (2013; 2016) showed that Germans indeed use lengthening as the phrase boundary marker, and when the lengthened syllable in a novel language marked the right edge of the discrete constituent, the segmentation performance improves. However, Italians do not conform to this general pattern. Lengthening of the word-final syllables, as well as the word-initial lengthening in the artificial language impeded the segmentation by the Italian listeners, contrary to what should have been expected, if final lengthening were universally used as a right-edge boundary marker facilitating the parsing. The authors argued that the facilitatory effect of the final lengthening had only been detected for the languages in which prominence (e.g., lexical stress) tends to align with the constituent edges. As in Italian prominence is not aligned with the word edges, Italian participants could not unambiguously interpret lengthening as the right-edge cue. Lengthening produced a confounding effect for Italian listeners who were more accustomed to process lengthening as a correlate of lexical stress, which is aligned with the penultimate syllable in their native language. Iversen, Patel and Ohgushi (2008) questioned the universality of the ITL by revealing the differences between English and Japanese listeners in rhythmic grouping of tone sequences into chunks, suggesting that the basic auditory processes might not be universal. The same conclusion was reached by Bhatara et al. (2013), who used linguistic stimuli (syllabic sequences) to demonstrate significant differences in perception of durational cues for chunking the continuous acoustic stream by French and German listeners. German listeners reliably used lengthening as a right-edge boundary cue, while French listeners showed inconsistent grouping patterns. This result indicated that the processing of lengthening cues is at least partially modulated by the linguistic experience. Toro et al. (2015) demonstrated that the rats also develop the processing bias for parsing the continuous speech into discrete chunks coherent with the exposure they had. Moreover, de la Mora et al.

(2013) compared the use of pitch and durational cues for the parsing of continuous acoustic streams by humans and rats and concluded that the trochaic rhythmic grouping based on pitch is universal, while the iambic grouping based on duration can be modulated by the linguistic experience (for humans) or exposure (for rats).

Therefore, it is not yet clearly answered whether lengthening in a novel language is processed universally by speakers with different native languages, as suggested by Tyler and Cutler (2009) or Kim et al (2012), or whether it is coherent with the linguistic input, as suggested by Ordin and Nespor (2013; 2016), Toro et al. (2015) or Bhatara (2013). Considering inconsistent results reported in the literature regarding the use of lengthening for segmentation, we decided to test and refine the hypothesis of L1-specific use of lengthening cues with a larger pool of languages that exhibit more variety in the role lengthening plays in the manifestation of linguistic structure (see table 1). To test the hypothesis, we adopted the artificial language learning paradigm (Saffran et al., 1996). We invited native monolingual Italian, German and Spanish speakers and Basque-dominant bilinguals to participate in the experiments. These languages exhibit important differences in regard to lengthening (the summary of the differences is presented in table 1). Within a single study, we test the hypothesis with languages that (a) tend to align stress with the edges of the constituents (German), (b) tend to place stress inside the constituents (Spanish and Italian), and (c) have non-contrastive movable stress at the lexical level (Gipuzkoan Basque). This selection of languages allows formulating clear predictions to test our hypothesis and to clarify the differences in the result patterns in previous experiments. Moreover, previous studies also differed in the type of instructions given to the participants. In some studies the instructions were incidental, i.e., participants were told that they should listen to an imaginary language mimicking the attitude they may have when listening to music or an unknown language. In other studies the instructions were intentional, i.e., the participants were told to listen to an imaginary language and detect and remember the words from this language. Potentially, differences across studies in the obtained results with listeners of various languages could be affected, among other things, by incidental vs. intentional instructions. In this study, we decided to give the listeners intentional instructions. One advantage of this approach is that we have comparable results from speakers of different native languages due to the consistency in the procedure. Secondly, we can compare the performance of Italian and German participants who receive intentional instructions in the current study with the performance reported in Ordin and Nespor (2013; 2016), where the listeners received incidental instructions. This will provide insight about the influence of different instructions

given to the participants, and a deeper understanding of how prosodic features interact with statistical cues in word segmentation.

We tried to maximize the differences in prosodic structure between languages, paying particular attention to the role of lengthening in manifesting linguistic structure in the acoustic speech stream. Cross-linguistic differences should permit building testable hypotheses regarding how lengthening in the L1 of the listener affects the segmentation strategies applied to a novel language, how lengthening is used to detect the discrete constituents in continuous acoustic stream, if the native language indeed influences the use of lengthening for the purposes of segmentation of an unfamiliar language.

The main perceptual correlate of lexical stress in Italian is duration (Bertinetto, 1980), especially when the stressed syllable is open and in penultimate position, which is the most frequent stress location in Italian (Krämer, 2009; Nespor, 1993). Stressed syllables in word-final position receive no lengthening except when this syllable is also phrase-final (Rogers & d'Arcangeli, 2004). Vowels in open stressed penultimate syllables are significantly and substantially longer than in stressed antepenultimate syllables because of the cumulative effect of phonetic and phonological lengthening in penultimate syllables and only phonetic lengthening in antepenultimate syllables (D'Imperio & Rosenthall, 1999). Thus, lengthening is a particularly important correlate of stress for Italian speakers in penultimate syllables. Italian phrasal prominence is aligned with the stressed syllables of the word (Nespor & Vogel, 2007), and adds additional degree of lengthening. Therefore, we expect that Italians are very likely to interpret prominence in a novel language as a feature of a prominent syllable, and therefore word-penultimate lengthening might have a facilitatory effect for the segmentation of a novel language.

The unmarked location of lexical stress in Spanish is also the penultimate syllable (Roca, 1999; Delattre, 1965). Delattre (1965) says that Spanish reveals a strong tendency to locate word stress on the penultimate syllable, and 74% of tri-syllabic words have stress on the penultimate syllable, while only 6% of words have stress on the antepenultimate syllable, the rest 20% of words exhibit word-final stress. Like in Italian, the most stable acoustic correlate of stress is lengthening, which is true even for unaccented positions (Ortega-Llebaria & Prieto, 2011; 2009). Therefore, Spanish and Italian are similar in regard to the most important acoustic and perceptual correlate of lexical stress and the unmarked location of lexical stress. Consequently, we expect the Spanish and Italian listeners to behave in a similar manner and to interpret

lengthening as a correlate of lexical stress. However, unlike in Italian, there is no evidence of phonological lengthening or different degrees of stress-induced lengthening depending on the location of stress in Spanish words, which could reduce the degree of the lengthening aid, compared to that of Italian listeners.

The results regarding the main acoustic correlates of word-level prominence in German are inconclusive (Isachenko and Schädlich, 1966 say that F0 is a stronger acoustic manifestation of lexical stress in German, while Dogil and Williams, 1999 say that increase in duration matters more in acoustic manifestation of stress). As for perceptual correlates, most studies indicate that the effect of acoustic lengthening on the perception of prominence can be easily overridden by pitch movements and vowel quality (Kohler, 2012). The perceptual distinction between prominent (both at word- and at phrase-levels) and non-prominent syllables is based on pitch fluctuations above a certain threshold, and duration plays a minor role in the perceptual domain (Fery, Hoerning, Pahaut, 2011; Nespor, Shukla, van de Vijver et al., 2008; Isachenko & Schädlich, 1966). Moreover, the perception of prominent syllables in German is more linked to the presence of full vowels instead of reduced vowels (Kohler, 2012), while in Italian and Spanish the importance of pitch as a perceptual correlate of lexical stress is less important, and qualitative reduction (reduction of a vowel to schwa in unstressed syllables) is almost non-existent (Bertinetto, 1980; Ortega-Llebaria & Prieto, 2011). Unlike Spanish and Italian, German exhibits phonemic opposition between long and short vowels, and therefore the length contrasts between stressed and unstressed syllables are longer than in Romance languages (stress-induced lengthening should be stronger than phonological length phonemic contrasts). This, in turn, makes German less sensitive to smaller differences in durational ratios due to stress contrasts (Kohler, 2012). Another major difference in stress phonology between German and Spanish/Italian is the most frequent location of lexical stress. In words of Germanic origin, stress is word-initial. In words of foreign origin - which compose a large part of German vocabulary - stress location is influenced by heavy syllables: long vowels and complex coda's consonantal clusters tend to attract stress (Dogil & Williams, 1999). Wiese (1996) says that there is a preference for antepenultimate stress if the penultimate syllable is open, and a preference for penultimate stress in case of closed penultimate syllable, but he rejects the hypothesis that stress is quantity-sensitive, i.e., attracted by long vowels, or bi-moraic nuclei. Delattre (1965) says that although German reveals a clear tendency to word-initial stress in general, in tri-syllabic words frequency of penultimate and antepenultimate lexical stress does not differ statistically. 51% of tri-syllabic words have penultimate stress, and 49% exhibit antepenultimate stress.

This shows that the location of stress is more variable in German than it is in Spanish and Italian, which weakens the potential of prominent syllables to mark either the edges of the linguistic units, or to consistently mark a certain position within the constituents. Consequently, if Germans will process lengthening in a novel language via the filter of their native language phonology, they will be less likely to interpret lengthening as a phrase-final boundary marker than as a manifestation of prominence.

The accentual systems of the Basque language vary a lot depending on the geographical variety. We present below a brief overview of some relevant features present in Gipuzkoan Basque dialects (the geographical dialectal area is defined by Hualde, 1999). Elordieta and Hualde (2014: 408) say that word stress is lexically contrastive only in the easternmost French region of Zuberoa, outside Gipuzkoa; in all the areas of Gipuzkoan dialects there is no contrast in word-level prosody at all. Usually the word-second syllable bears stress, but occasionally stress may also non-contrastively fall on the word-initial or word-final syllable even in the same phrasal environment (Elodrieta & Hualde, 2014: 463, 440). In addition, the location of the most prominent syllable within a word is often influenced by inflectional suffixation and in some central Gipuzkoan dialects can be attached to the position of the syllable in the phonological phrase, not in the word (Hualde, 1999), with lengthening on the phrase-second and phrase-final syllable. Thus, word-level prominence can hardly be used to detect the boundaries of lexical items. However, phrase-final lengthening can be a much more reliable correlate to detect the right edge of the phonological phrases because phrase-final syllable receives phonological as well as phonetic lengthening.

INSERT TABLE 1 SOMEWHERE HERE

Based on the overview of phonetic and phonological distinctions between the selected languages, we assume cross-linguistic differences in the processing of lengthening in *native* languages of Basque, German, Spanish and Italian speakers. If adult listeners indeed process the prosodic cues in a *novel, unfamiliar* language via the filter of their native phonology, we might expect to find cross-linguistic differences in the use of lengthening for the segmentation of a novel language. As we use intentional instructions and straightforwardly inform participants that they need to detect the words of an imaginary language, we assume they recourse to word-level phonology when interpreting lengthening as prominence correlate. These assumptions allow us to build the following predictions for the experimental outcome. We expect that native Italians might benefit from the penultimate lengthening, i.e., the increase in duration of penultimate syllables in statistical words should

improve segmentation performance compared to the condition without lengthening cues. In Spanish, prominence is aligned with penultimate syllables, like in Italian, and we expect the performance pattern of Spanish and Italian listeners to be similar, thus confirming the effect of linguistic experience on the interpretation of durational cues in a novel language. However, the facilitatory effect of penultimate lengthening on segmentation of a novel language by native Spanish listeners might be smaller than that by native Italian listeners, because the stress-induced lengthening in penultimate position in Italian is greater than in antepenultimate or word-final position, and it is substantially greater than in penultimate position in Spanish. Native German listeners are less likely to perceive increase in duration as a word-level prominence because the role of duration in manifesting stress in their native language in inferior to that of F0 fluctuations. We expect that in our material lengthening should be interpreted as a tight-edge boundary signal by native German listeners, who are likely to perceive the increase in duration as a final lengthening cue, rather than as a stress correlate. Native Basque listeners are expected to behave more like German listeners, i.e., benefit when the final syllables of sequential constituents are marked by lengthening. However, the effect will be weaker than that for German listeners due to confounding phrase-second syllable lengthening in their native language.

Methods

To verify our predictions, we exposed participants to artificial languages with statistical words bearing lengthening either on the final, the penultimate or the antepenultimate syllable, and evaluated the segmentation performance of listeners with different native languages in each of these conditions. We also created artificial languages without implemented prosodic cues and used segmentation performance in this condition as the reference baseline. Comparing performance in this reference condition with performance in a condition with implemented durational cues will reveal facilitatory or impeding effect of lengthening on different syllables for speakers with different native languages.

Participants

We invited monolingual Italian, Spanish, German and bilingual Basque speakers (24 participants per language group) who received monetary contribution for taking part in the experiment. None of the participants either reported or showed any speech or hearing disorders. None of the listeners had participated in the experiments reported in Ordin and Nespor (2013; 2016). Italian participants came from families with

monolingual parents, were exposed only to Italian from birth, and started learning English as a compulsory subject at school for the first time. We took care to recruit speakers of north-eastern Italian varieties without strong dialectal influences (e.g., Friulano or Veneto speakers were not in the sample). Approximate age was 19-21 years. All were students from Trieste University. The experiment was carried out in Italy, Trieste.

German participants were recruited among the students of Bielefeld University, all came from monolingual families, were raised in or around the city of Bielefeld, all were standard northern German speakers, were exposed only to German from birth till they started learning English as a compulsory subject at school. Approximate age was 19-22 years. The experiment was carried out in Bielefeld, Germany.

Basque speakers were Basque-Spanish bilinguals, all reported to be Basque dominant, with Basque being the only family language. We did not include into our sample participants, if one of the parents or grandparents was not a Basque speaker from birth. Our participants have reported to be using Basque predominantly and much more than Spanish, and they all lived in Basque-dominant towns. We selected speakers from the geographical area that encompasses Gipuzkoan dialects only. Spanish speakers came from monolingual families, north of Spain (Asturais, Cantaqbria, Burgos, La Rioja), reported to have little or no contact with other languages, were exposed only to northern varieties of Spanish from birth. All Spanish and Basque participants were students at University of the Basque Country, approximate age 20-27 years. The monolingual Spanish participants were selected among those who had recently come to the Basque country to study in Vitoria, therefore they had not had extensive exposure to the Basque language. Moreover, Vitoria and Araba are Spanish-speaking town and province, where Basque is not frequently used by the inhabitants. Majority of residents in Vitoria are also L1 Spanish speakers. The experiment was carried out in Vitoria-Gasteiz, Spain.

Stimuli

The same material as in Ordin and Nespor (2013) were used. We created 12 statistically-defined words using CV syllables (*komipa, bolatu, kupige, vunelu, bamofe, defida, bukite, vifole, dubipo, vaputa, donume, ginefa*). Two artificial language streams were synthesized, each consisting of 166 repetitions of six randomly concatenated words, with the TPs between the syllables within the words 100%, and the TPs between the syllables straddling the word boundaries around 16%. Following the test (see procedure below), we tested that

none of the statistical words is recognized significantly better or worse than the other statistical words within the same language stream. Each word consisted of three consonant-vowel syllables; duration of each sound was set to 100ms. The concatenated sequences of words were synthesized with monotonized F0 set to 200Hz. Each stream was then modified to implement lengthening cues, with either the first (word-initial), the second (penultimate) or third (word-final) syllables lengthened by increasing the vowel duration by 80ms. In the end each stream was prepared in four different conditions: TP-only, initial-lengthening, middle-lengthening, and final-lengthening. We used French voice in MBROLA for speech synthesis. French voice was used because we wanted to use the same material for speakers of four different languages and to avoid giving advantage to any group of participants by using native phonemes. Therefore, we chose the French voice.

Procedure

Participants had to come for the experiment twice. In the first session they were exposed to stream 1 and stream 2 in two different conditions, and in the second session – at least one week later – to stream 1 and stream 2 in the other two conditions. The combination of stream × condition × order of presentation (order in which the conditions were presented to individual listeners) was randomized (24 unique combinations), and one participant per language (L1) was assigned to one unique combination.

Participants were instructed to listen to an imaginary language. We told them that the language does not contain words from real languages; it has its own vocabulary. Before the language exposure phase, participants were informed that after listening to the imaginary language, they would hear pairs of possible words from this unknown language. Only one word in each pair would be a real word from the language, and they will have to choose which one of the two possible candidates is a real word.

After exposure, participants had to do a dual forced-choice task. In the test phase we pitted words against partwords, i.e. syllable sequences that were present in the stream but had a TP trough between two syllables. We made six partwords for each artificial language. Three partwords were formed from the third syllable of one statistically-defined word and the first and second syllables of the following word, and three partwords were formed from second and third syllables of one word and the first syllable of the next word. Pitting all possible words against all possible partwords gave 36 pairs, each containing one word and one partword. The order of words and partwords in the pairs was counterbalanced. The order of the pairs was

randomized for each participant. The items in the pair were separated by a 500-ms pause. Participants were instructed to listen to the pair and to click either button 1 or button 2, depending on whether they considered the first or the second item in the pair a word in the language they had just listened to.

The stream and the test items were presented via headphones. Participants were instructed and tested individually. After the test was over, participants had a 5-minutes pause before the second stream was presented, followed by a new test. The same procedure but with the streams of the other two conditions was used in the second session. Franco, Cleeremans, and Destrebecqz (2011) showed that people are able to learn two artificial languages sequentially and to easily differentiate between them. Gebhart, Aslin, and Newport (2009) found interference between statistically coherent languages when they were presented sequentially, but the interference disappears if either the exposure to the second language was long enough, or the presence of two different structures was marked explicitly in the instructions, or when the two subsequent languages were separated by a pause. All three conditions were fulfilled in our experiments. We thus assume that one stream did not influence the other during either familiarization or test. Following Ordin and Nespor (2013; 2016), we have also carried out t-tests comparing the number of correct responses for stream 1 against that for stream 2, comparing the number of correct responses in the first and in the second session, as well as the number of correct responses in the first and second stream presented within each session. None of the comparisons was significant. This confirmed the assumption made on previous empirical findings, that in our participant sample and material the two languages do not interfere, and that neither the order of presentation nor the session have a significant effect on the segmentation performance.

Results

Segmentation performance has been assessed by the number of correct responses. Figure 1 provides the mean number of correct answers for each condition and language group and the bars show ±2 standard errors.

INSERT FIGURE 1 SOMEWHERE HERE

Statistical tests were applied to assess the effect of the *native language* of the participant (German vs. Italian vs. Spanish vs. Basque) and *lengthening presence and location* (on the word-antepenultimate, penultimate or final syllable, or lack of lengthening) on the segmentation. As in most cases with repeated measures, the observations, although perfectly counter-balanced in our study, are not independent. Therefore,

for a proper application of the repeated-measures ANOVA, additional tests are to be performed (checking for sphericity and for the symmetry of compounds) and, if necessary, correcting the degrees of freedom in the main analysis. To avoid additional tests on the same data set, it is often recommended to use a multivariate approach, especially for cases with more than two levels of a dependent variable (i.e., with more than two measurements per subject). We follow the recommendations and procedures described in O´Brien and Kaiser (1985) and Max and Onghena (1999) how to do repeated-measures analysis with MANOVA approach. The number of correct answers out of a total of 36 responses represents segmentation performance. Planned contrasts for comparing the performance in the TP-only condition with the conditions with implemented lengthening reveal whether the lengthening of a particular syllable facilitates or impedes segmentation. Comparing segmentation performance with chance level (50%) shows whether segmentation is overall successful. If the lengthening of a particular syllable leads to a significant drop in performance compared to the TP-only condition to the degree that the difference with the chance is no longer significant, we say that the segmentation is disrupted. If the number of correct responses after a significant drop in performance is still above chance, we say that the segmentation is impeded. The stepwise Bonferroni correction was applied for multiple comparisons.

We have detected significant interaction of *native language* and *lengthening location,* $\lambda = .689$, $F(9, 219) = 4.031$, $p < .0005$, $\mu^2 = .117$. This shows that listeners with different native languages indeed react differently to duration cues. Figure 1 shows that segmentation by Italian participants in middle-lengthening condition is better than in TP-only condition, segmentation by German and Basque participants is better in final-lengthening than in TP-only condition, and segmentation by Italian and Spanish listeners in initial-lengthening condition is worse compared to TP-only condition. We therefore decided to run separate analyses to assess the influence of lengthening on the segmentation of a novel language by listeners from different language groups. Each test was followed by planned contrasts in order to find out whether the differences in performance across conditions within each language group are significant and statistically substantial.

The effect of *lengthening location* on segmentation by Germans is significant, $\lambda = .535$, $F(3, 21) = 6.079$, $p = .004$, $\mu^2 = .465$. In order to detect the facilitatory or impeding effects of prosody on segmentation, planned contrasts were made to compare segmentation performance in the TP-only condition with that in the other conditions. Contrasts reveal that performance by German listeners is significantly better when the final

14

syllable is marked by duration, $F(1, 23) = 7.791$, $p = .01$, $\mu^2 = .253$. Performance in segmentation by German listeners does not differ from the TP-only condition when either the penultimate syllable, $F(1, 23) = 1.889$, $p = .183$, $\mu^2 = .076$, or the antepenultimate syllable is lengthened $F(1, 23) = .8$, $p = .38$, $\mu^2 = .034$. This means that the increase of duration on the final syllable facilitates segmentation, while increase of duration on the penultimate or antepenultimate syllable does not affect segmentation by German listeners.

The test shows that the effect of *lengthening location* on segmentation by Italians is also significant, $\lambda = .299$, $F(3, 21) = 16.385$, $p < .0005$, $\mu^2 = .701$. Planned comparisons revealed that participants' performance was significantly worse when duration in a three-syllabic word marked the antepenultimate syllable, $F(3, 23) = 9.948$, $p = .004$, $\mu^2 = .302$. Segmentation is significantly better when the penultimate syllable in the word is lengthened, $F(3, 23) = 8.653$, $p = .007$, $\mu^2 = .273$. The difference in the number of correct responses in TP-only and final-lengthening conditions is not significant, $F(3, 23) = .002$, $p = .964$, $\mu^2 < .0005$. This means that segmentation of a novel language by Italian listeners is facilitated by word-penultimate lengthening and is impeded by antepenultimate lengthening, while word-final lengthening does not affect the segmentation performance.

Segmentation performance by Spanish listeners is also affected by *lengthening location*, $F(3, 23) = 4.247$, $p = .017$, $\mu^2 = .378$. Planned contrasts showed that the number of correct responses in initial-lengthening condition is significantly lower than in TP-only condition, $F(3, 23) = 8.285$, $p = .008$, $\mu^2 = .265$. Performance in segmentation by Spanish listeners does not differ in the TP-only and the middle-lengthening conditions, $F(3, 23) = 1.864$, $p = .185$, $\mu^2 = .075$, and in the TP-only and the final-lengthening conditions, $F(3, 23) = .075$, $p = .787$, $\mu^2 = .003$. These results show that segmentation is impeded by lengthening on the antepenultimate syllable, and is not affected, i.e., neither impeded, nor facilitated, by word-medial and word-final lengthening.

The effect of *lengthening location* on segmentation by Basque speakers is not significant, $\lambda = .805$, $F(3, 21) = 1.697$, $p = .198$. Still, a large effect size, $\mu^2 = .195$ encouraged us to explore the planned contrasts on segmentation performance in the TP-only condition with that in initial-, middle-, and final-lengthening conditions. The comparisons reveal no significant difference between the number of correct responses in TP-only and initial-lengthening, $F(3, 23) = .134$, $p = .718$, $\mu^2 = .006$ and between TP-only and middle-lengthening $F(3, 23) = .002$, $p = .964$, $\mu^2 < .0005$ conditions. The number of correct responses is higher in final-lengthening compared to TP-only conditions, and the difference is on the verge of significance, with moderate effect size,

$F_{(3, 23)} = 3.402$, $p = .078$, $\mu^2 = .129$. This allows us to tentatively suggest that final lengthening probably facilitates segmentation of a novel language by Basque speakers. However, given the non-significant effect of the full model, and a very vague statistical evidence of final-lengthening facilitation, this conclusion is tentative and preliminary. We consider it as an indication that if the Basque listeners benefit from lengthening at all, it can only be from final lengthening.

As the participants had to do a dual forced-choice task, they could score 50% of correct answers with random responses, which is the chance level. If participants successfully segmented the speech stream, then we should expect preference for words over partwords, i.e. the number of correct responses to be significantly above chance. One-sample *t*-tests were performed to compare the number of correct answers in each condition with the chance level.

Comparing the performance of listeners with different native languages with the chance level (table 2), we can see that German and Basque listeners reliably segment continuous speech regardless of the lengthening location (the number of correct responses is always above chance). Segmentation by Italian and Spanish listeners fails in the initial-lengthening condition (the number of correct responses does not differ from what might be expected by chance). These results show that lengthening of antepenultimate syllables disrupts – not merely impedes – segmentation by Spanish and Italian participants; no disrupting effect of lengthening location has been detected with the German and Basque participants.

INSERT TABLE 2 SOMEWHERE HERE

Discussion

The results show that speakers with different native languages indeed process lengthening cues differently. Word-final lengthening is beneficial for German participants and probably also for Basque participants. Lengthening of penultimate syllables facilitates segmentation by Italians, and antepenultimate lengthening impedes and disrupts segmentation by Spanish and Italian listeners (figure 1 and table 2).

TPs between syllables are computed online, during exposure (Gomez, Bion, & Mehler, 2011), and provide sufficient cues for extraction of discrete sequential constituents from continuous acoustic streams. Prosody is not essential for segmentation, but may affect the segmentation performance. We should try to understand at what stage prosody intervenes with statistical cues. We suggest that TP computation and

16

extraction of prosodic regularities, i.e., prosodic structures, are parallel and independent processes. This suggestion is based on neuroimaging, theoretical and behavioural evidence. Neuroimaging studies provide evidence that segmental and prosodic information is processed in different hemispheres and at different timescales, but in parallel (Telkemeyer, Rossi, Koch, Nierhaus, Steinbrink, & Poeppel, 2009). This also agrees with the proposal by Christiansen and Charter (2016) that the processing of incoming speech happens across multiple levels of linguistic representations (e.g., syllabic sequences, words, phrases, etc.), each involving parsing within different time-windows. Wheeldon and Lahiri (1997) demonstrated that the time needed to initiate articulation depends on prosodic structure, and more complex prosodic structures lead to longer planning, even when syllables are held constant. This also suggests that the planning of prosodic structure and segmental material happen at different timescales. Our data suggests that TP computations happen at the timescale of the syllable (contrasting the TPs between the syllable pairs within the words and the syllable pairs straddling the word boundaries), while extraction of prosodic regularities happens simultaneously at a longer timescale (a timescale of statistical words, within which the stress-assignment rules are operating). Shukla et al. (2007) and Toro et al. (2011) showed that the impeding effect of language-specific constraints disappears if the test is performed in the visual modality (i.e., the word candidates are presented visually, not auditorily during the test), thus indicating that the sequential TP-based constituents are successfully extracted from continuous streams, irrespective of whether regularities of a native language and a novel artificial language match or mismatch. Violations of prosodic regularities do not prevent TP-based parsing. We further suggest that prosody is used to construct frames. The syllabic sequences that correspond to statistical regularities are used to fill in these frames (see also Ordin & Nespor, 2016). If the segmented syllabic sequence does not fit the frame, it is suppressed as a possible word candidate. That would mean that prosody intervenes later, when the constituents are already extracted, and filters out possible word candidates. Candidates that do not fit the prosodic constraints are filtered out (Shukla et al., 2007). Therefore we assume that the segmentation mechanism is based on splitting the incoming continuous speech into syllabic sequences – using TPs – embedded into constructed prosodic frames. Prosody and statistical cues can interact at the stage of recognition, when the inventory of the recognized and retained syllabic strings is updated. When a word candidate is remembered, it is used as "an anchor word", to facilitate further extraction of statistical word candidates from A continuous acoustic stream. When the listener recognizes the syllabic stream as a discrete constituent, he can

process the syllables following and preceding this constituent (the recognized syllabic sequence) as the edge syllables for other constituents, thus using the recognized unit as an anchor word for further segmentation. The facilitatory effect of anchor words has been clearly demonstrated using behavioural as well as electrophysiological measures (Cunillera, Laine, & Rodrigues-Fornells, 2016). Constant update of the inventory of recognized and retained constituents allows for the time-varying continuous processes at each level are modulated by processes at the level above and at the level below, as specified by the predictive coding framework on speech processing (Clark, 2013; Lupyan & Clark, 2015 Christiansen & Charter, 2016).

The proposed mechanism also appears to be at work during the segmentation of natural languages. Salverda, Dahan, Tanenhaus, Crosswhite, Masharov, & McDonough (2007) showed that prosodic structure affects the degree at which lexical candidates compete in speech decoding. If there are several lexical competitors, "prosodically matching" candidates compete more strongly than "prosodically mismatching" candidates, even when the latter exhibit greater segmental overlap. In other words, word candidates that fit the prosodic structure are retrieved earlier than the lexical candidates that do not fit the prosodic structure.

The proposed mechanism also agrees with a number of frame-filler models of phonological encoding (Dell, 1986; 1988; Shattuck-Hufnagel, 1992; Levelt, 1989; 1992). In speech production, prosody is also encoded separately from segments (Ferreira, 1993). More recent studies also suggest the independence of segmental and prosodic representations in speech perception (Schild, Becker, & Friedrich, 2014), with shared neural networks and mechanisms underlying both encoding and decoding processes (Silbert, Honey, Simony, Poeppel, & Hasson, 2014). Prosodic information at the word level is used to construct the prosodic frames for phonological words. The segmental information is accessed separately and the phonological segments are combined in a string that is fitted into the prosodic frame sequentially from left to right.

Our results indicate that German and Basque listeners construct frames for tri-syllabic sequences with a lengthened "slot" for the final syllable, Italians construct frames with the lengthened penultimate syllable, and the frames that Spanish listeners construct suppress the word candidates with antepenultimate lengthening and retain the candidates with penultimate and final lengthening. Why do German and Basque listeners construct the frames with the word-final lengthening, while the final lengthening operates at a level of the phonological phrase, not of the phonological word? Shukla et al. (2007) and Endress and Mehler (2009) showed that the units at a lower level of the hierarchy are more easily detected and remembered when they are at the edges of larger

units of a higher hierarchical level. That is, the segmentation performance is higher for the statistical words that are aligned with the edges of phrases than for the statistical words inside the phrases. Endress, Nespor and Mehler (2009) and Hochmann, Langus and Mehler (2016) explained this effect by the memory and perceptual constraint, which lead to enhanced encoding of units located at the edges of larger units. We can assume that the prosodic frames can be created for the statistical words at the edges of phrases, and these frames might differ from those created for the statistical words in the middle of phrases. This requires that listeners use not only word-level, but also phrase-level prosody to create prosodic frames. This also requires that listeners differentiate between lengthening cues at different levels of the prosodic hierarchy. The proposed hypothetical explanation for the observed result pattern requires further empirical tests. However, some initial support for this proposal can be grounded in the work by Wheeldon and Lahiri (2002), who showed that the properties of phrasal prosody can also be important for phonological encoding, and proposed that the processes sensitive to phrasal prosody for phonological decoding are blind to word-level prosody, (this requires that the syllabic sequences segmented as a unit are then transferred as a single chunk to a phrasal level). Moreover, in natural languages that exhibit both stress-induced and phrase-final lengthening, speakers provide other cues for the listener to adequately assign durational information to lexical stress or to signalling finality (Monaghan, White, & Merkx, 2013).

Italians more readily decode lengthening as a correlate of stress because lengthening is the main acoustic and perceptual correlate of stress in their native language. Italian listeners probably constructed frames with the slot for the longer syllable in penultimate position, corresponding to the unmarked location of lexical stress. Germans are constantly re-ranking stress correlates depending on the context and rely on the complex of F0 and duration fluctuations and spectral differences of vowels in stressed and unstressed syllables (Kohler, 2012), and therefore do not associate the stable lengthening pattern in the artificial language as a correlate of stress. Instead, they probably perceive it as a phrase-final cue. In the Gipuzkoan varieties of Basque, the word-final syllable and the stressed syllable are lengthened, but the position of the stressed syllable in a word is variable and may even be shifted by inflexional morphemes and differ between repetitions of the same word, thus leaving the word-final syllabic slot as a reliable anchor as to where lengthening can occur. This slightly improves segmentation by Basque native speakers in the final-lengthening condition compared to TP-only condition. German and Basque listeners construct frames with the slot for the longer syllable in final position.

As there is no unmarked location of lexical stress in German and Basque, word candidates with penultimate and antepenultimate lengthening are not suppressed, but the segmentation performance for these words by German and Basque listeners is not facilitated either. The frames that Spanish listeners construct suppress the word candidates with antepenultimate lengthening and retain the candidates with penultimate and final lengthening, which indicates that the Spanish might have used prosody to create two possible frames, using either phrasal or word prosody.

An interesting question is the difference in the segmentation performance by Italian and Spanish participants in the middle-lengthening condition in this experiment (same material, same procedure, similar conditions). Spanish listeners did not benefit from penultimate lengthening. They successfully segmented in the middle-lengthening condition, but their performance did not differ from that in TP-only condition, although the unmarked location and the main correlates of lexical stress match in Spanish and Italian. A possible explanation is the combination of phonetic and phonological lengthening in open stressed penultimate syllables in Italian, while Spanish stress-induced lengthening is only phonetic. There is no evidence that antepenultimate stressed vowels in open syllables in Spanish are shorter than corresponding penultimate vowels, while lengthening of penultimate vowels in Italian is more substantial than in antepenultimate positions.

We need to explain the discrepancy in the result patterns reported here and in those reported in Ordin and Nespor (2013) regarding segmentation of the same material in the same environment by Italians. Ordin and Nespor (2013) showed that Italians failed to segment the same streams in initial- and final-lengthening conditions. Segmentation was successful in middle-lengthening condition, but not above TP-only condition, thus middle lengthening had neither beneficial, nor impeding effect on segmentation performance. In this study, however, middle lengthening exercised a facilitatory effect, raising the segmentation performance in middle-lengthening condition above TP-only condition. Contrary to what was reported in Ordin and Nespor (2013), final lengthening in this experiment did not show impeding effect on segmentation, neither did facilitate segmentation by Italian listeners.

The difference might potentially stem from a better control for the participants' background in this experiment. We were very strict to select only FVG (Friuli-Venezia-Giulia region) speakers, and most of them were speakers from Trieste. Ordin and Nespor (2013) had Italian speakers of other dialects: although most of them were from the North of Italy, not all of them grew up in Trieste. Yet, it is very unlikely that idiosyncratic

characteristics of the participant sample had any effect on the segmentation because all Italian dialects share the lengthening features that are described in the introduction and assumed to have an influence on selecting the word candidates.

Instead, we propose that the difference appeared due to different instructions given to the participants. In previous experiments, participants were instructed to listen to the language attentively mimicking the attitude they would have when listening to real speech, because after listening they would have to answer some questions about the novel language. In the experiment reported in this manuscript, participants were told to detect and learn the words of the novel language, because after listening they would hear pairs of possible words and they would have to choose which candidate in each pair is a real word from the novel language they were about to listen to. The instructions that encourage intentional learning raises listener's awareness of word-level prosody, and prosodic representations (prosodic frames) are formed faster and with higher precision than in the case of incidental learning, thus updating representations at the segmental level (sequences of syllables that are used as content for the frames). This explanation fits the predictive coding framework (Clark, 2013; Lupyan & Clark, 2015; Sohoglu, Peelle, Carlyon, & Davis, 2012), which suggests that an incoming acoustic speech stream is simultaneously processed at different levels. Predictions for the higher-level representations (prosodic frames) are generated faster than lower-level representations (syllabic sequences), and constrain lower-level processing of sensory information before it even occurs (Christiansen & Charter, 2016). The sensory input in the current study is modulated by higher-level expectations regarding the lengthening positioning in the native language of the participant, and by the level of attention, which is modulated by the type of instructions. Thus, giving direct instructions for intentional learning primes the participant to use their native language as a tool for tuning sensory input. Segmentation is facilitated by the existing processing skills honed for the native language of the participant. The existing processing skills and knowledge of L1 phonology predict and enable rapid decoding of the future input. This predictive mechanism can also explain why the durational cues to the word boundary placement are being re-ranked as speech processing continues with more input, and lengthening can play more or less significant role at different times even for the speakers of the same language (Heffner, Dilley, McAuley & Pitt, 2012). We believe that this mechanism explains why speakers of different languages, or even speakers of the same language at different times sometimes reconstruct different representations from the same sensory input.

The obtained results confirm the original hypothesis that, to some extent, the use of durational cues in the segmentation of a novel language is L1-specific. Processing of lengthening cues in a novel language of exposure is not universal and interpretation of lengthening as a universal phrase-final boundary marker in a novel language can be overridden by language-specific phonology of lexical stress in the native language of the listener, and by the attentional factors. An interesting question for further research is to address the issue of how much exposure is needed to overcome the universal bias to interpret the lengthening as the right-edge boundary marker in L1 acquisition, and in L2 acquisition by learners whose native and target languages encourage different processing of lengthening cues. Also, it would be interesting to set up experiments with non-human animals that do not have linguistic abilities. This work could reveal to what extent the linguistic structure and the universal processing bias can be identified as having a non-linguistic basis.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324. doi: 10.1111/1467-9280.00063.

Bertinetto, P.M. (1980). The perception of stress by Italian listeners. *Journal of Phonetics, 8*, 385-395.

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language affects rhythmic grouping of speech. *Journal of the Acoustical Society of America* 134(5), 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 (1), 123-140. doi: 10.1177/0023830910388018.

Cambier-Langeveld, T., Nespor, M., & van Heuven, V., J. (1997). The domain of final lengthening in production and perception in Dutch. *EUROSPEECH-1997*, 931-934.

Christiansen, M.H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences* 39. doi: 10.1017/S0140525X1500031X

Christophe, A., Peperkamp, Sh., Pallier, Ch., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language 51(4)*, 523–547. doi: 10.1016/j.jml.2004.07.001.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences 36(03)*, 181-204.

Crystal, T. H., & House, A. S. (1987). Segmental duration in connected speech syllables: Syllabic stress. *Journal of the Acoustical Society of America, 83*, 1574–1585. doi: 10.1121/1.395912.

Cunillera, T., Laine, M., & Rodrigues-Fornells, A. (2016). Headstart for speech segmentation: a neural signature for the anchor word effect. *Neuropsychologia ,* 189-199. DOI: 10.1016/j.neuropsychologia.2016.01.011

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113–121. doi:10.1037/0096-1523.14.1.113.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology, 24*, 381–410. DOI: 10.1016/0010-0285(92)90012-Q

D'Imperio, M., & Rosenthall, S. (1999). Phonetics and phonology of main stress in Italian. *Phonology, 16*(1), 1–28. Retrieved from: http://journals.cambridge.org.

D'Imperio, M., Elordieta, G., Frota, S., Prieto, P., & Vigário, M. (2005) Intonational Phrasing in Romance: The role of prosodic and syntactic structure. In S. Frota, M. Vigário & M. J. Freitas (eds) Prosodies. Phonetics & Phonology Series. Berlin: Mouton de Gruyter, pp.59-97.

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,* 225-241. doi: 10.1080/13670050.2014.904839.

de la Mora, D. M., Nespor M., & Toro J. M. (2013). Do humans and non-human animals share the grouping principles of the Iambic-Trochaic Law? *Attention, Perception & Psychophysics 75,* 92-100.

Dell, G.S. (1986). A spreading activation theory of retrieval in sentence production. Psychological Review 93, 283-321. doi: 10.1037/0033-295X.93.3.283.

Dell, G.S. (1988). The retrieval of phonological forms in production: Tests of production from a connectionist model. Journal of Memory and Language 27, 124-142. doi: 10.1016/0749-596X(88)90070-8.

Dogil, G., & Williams, B. (1999). The phonetic manifestation of word stress. In H. van der Hulst (Ed.), Word Prosodic Systems in the Languages of Europe (pp. 273–311). Berlin: Mouton de Gruyter.

Elordieta, G. & Hualde, J. (2014). Intonation in Basque. In Sun-Ah Jun (Ed.). *Prosodic Typology II* (pp. 405-464). Oxford: Oxford University Press.

Elordieta, G., Frota, S., & Vigário, M. (2005) Subjects, objects and intonational phrasing in Spanish and Portuguese. Studia Linguistica 59 (2-3), 110-143. doi: 10.1111/j.1467-9582.2005.00123.x.

Endress, A. D., & Mehler, J. (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology 62(11)*, 2187–2209.

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences 13(8)*, 348–353.

Ferreira, F. (1993). The creation of prosody during sentence production. *Psychological Review 100*, 233-253. doi: 10.1037/0033-295X.100.2.233.

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics 36*, 680–703.

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: how conscious? *Front. Psychology (Language Sciences) 229*, 1-12. doi: 10.3389/fpsyg.2011.00229.

Frota, S. (2000). Prosody and focus in European Portuguese. Phonological phrasing and intonation. New York: Garland Publishing.

Gebhart, A., Aslin, R., & Newport, E. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science, 33(6)*, 1087-1116. doi: 10.1111/j.1551-6709.2009.01041.x.

Gervain, J.; Sebastian-Galles, N.: Diaz, B.; Laka, I.; Mazuka, R.; Yamane, N.; Nespor, M.; Mehler, J. (2013). Word Frequency Cues Word Order in Adults: Cross-Linguistic Evidence. *Frontiers in Psychology* 4:689, 1-12. doi:10.3389/fpsyg.2013.00689.

Gomez, D.M., Bion, R., & Mehler, J. (2011). The word segmentation process as revealed by click detection. *Language and Cognitive Processes 26(2)*, 212-223.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge, UK: Cambridge University Press.

Gussenhoven, C. (2009). Vowel duration, syllable quantity, and stress in Dutch. In K. Hanson & Sh. Inkelas (Eds.). The Nature of the Word. Essays in Honor of Paul Kiparsky (pp. 181-198). Cambridge, MA: MIT Press.

Hay, J., & Diehl, R. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics 69(1)*, 113-22. doi: 10.3758/BF03194458.

Heffner, C., Dilley, L., McAuley, D., & Pitt, M. (2013). When cues combine: how distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes 28*, 1275-1302.

Hochmann, J.-R., Langus, Al., & Mehler, J. (2016). An Advantage for Perceptual Edges in Young Infants' Memory for Speech: An Advantage for Perceptual Edges. *Language Learning 66 (S2)*, 13-28.

Hualde, J. (1999). Basque accentuation. In van der Hulst, H. (Ed.), *Word Prosodic Systems in the Languages of Europe* (pp. 947-993). Berlin: Mouton de Gruyter.

Isachenko, A.V., & Schädlich, H.-J. (1966). Untersuchungen über die deutsche Satzintonation. *Studia Grammatica 7*, 7-64.

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

Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Studies in Second Language Acquisition, 34(3)*, 415-444. doi:10.1017/S0272263112000137.

Kohler, K. (2012). The perception of lexical stress in German: Effects of segmental duration and vowel quality in different prosodic patterns. *Phonetica, 69(102)*, 68-93. doi: 10.1159/000342126.

Krämer, M.(2009). *The Phonology of Italian*. Oxford: Oxford University Press.

Ladefoged, P., & Maddieson, I. (1995). *The Sounds of the World's Languages*. Cambridge: Wiley-Blackwell.

Langus, A., Marchetto, E., Bion, R., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language, 66*, 285-306. doi: 10.1016/j.jml.2011.09.004.

Levelt, W. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levelt, W. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition 42*, 1-22. doi: 10.1016/0010-0277(92)90038-J.

Lupyan, G., & Clark, A. (2015). Words and the world predictive coding and the language-perceptioncognition interface. *Current Directions in Psychological Scienc, 24(4)*, 279-284.

Manolescu, A., Olson, D., & Ortega-Llebaria, M. (2009). Cues to contrastive focus in Romanian. In M. Vigário, S. Frota, and M. J. Freitas (Eds.), *Phonetics and Phonology: Interactions and Interrelations. Current Issues in Linguistic Theory 306* (pp. 71–90). Amsterdam: John Benjamins.

Mattys, S., White, L., & Melhorn, J. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477-500. doi: 10.1037/0096-3445.134.4.477.

Monaghan, P., White, L., & Merkx, M. (2013). Disambiguating durational cues for speech segmentation. *Journal of Acoustical Society of America 134(1)*, EL45-EL57. doi: 10.1121/1.4809775.

Nakai, S., Turk, A., Kari, S., Granlund, S., Ylitalo, R., & Kunnari, S. (2012). Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics 40(6), 796-807*. doi: 10.1016/j.wocn.2012.08.003.

Nespor, M. (1993) *Fonologia*. Bologna: il Mulino. pp. 348.

Nespor, M., & Vogel, I. (2007). *Prosodic Phonology*. Berlin, Germany: Walter de Gruyter.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue e Linguaggio* 7(2), 1-28. doi: 10.1418/28093.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue e linguaggio 7(2)*, 1–28.

Ordin, M., & Nespor, M. (2013). Transition probabilities and different levels of prominence in segmentation. *Language Learning 63(4)*, 800-834. doi: 10.1111/lang.12024.

Ordin, M., & Nespor, M. (2016). Native Language Influence in the Segmentation of a Novel Language. *Language Learning and Development*, DOI: 10.1080/15475441.2016.1154858.

Ortega-Llebaria, M., & Prieto, P. (2009). Perception of word stress in Castilian Spanish. In M. Vigário, S. Frota & M. J. Freitas (Ed.). *Phonetics and Phonology: Interactions and Interrelations* (pp. 35-50). Amsterdam, The Netherlands: John Benjamins.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic Correlates of Stress in Central Catalan and Castilian Spanish. *Language and Speech, 54*(1), 73-97. doi: 10.1177/0023830910388014.

Palmer, C. (1997). Music performance. *Annual Review of Psychology 48*, 115-138. doi: 10.1146/annurev.psych.48.1.115.

Peña, M., Bion, R. A. H. & Nespor, M. (2011) How modality specific is the iambic–trochaic law? Evidence from vision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*; Vol 37(5), pp. 1199-1208. doi: 10.1037/a0023944.

Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604 -607. doi: 10.1126/science.1072901.

Pilon, R. (1981). Segmentation of speech in a foreign language. *Journal of Psycholinguistic Research 10(2)*, 113-122.

Potisuk, S., Gandour, J., & Harper, M. P. (1996). Acoustic correlates of stress in Thai. *Phonetica, 53*, 200–220. doi: 10.1159/000262201.

Rietveld, T., Kerkhoff, J., Gussenhoven C. (2004). Word prosodic structure and vowel duration in Dutch. *Journal of Phonetics 32(3)*, 349-371. doi: 10.1016/j.wocn.2003.08.002.

Roca, I. (1999). Stress in the Romance Languages. In van der Hulst, H. (Ed.), *Word Prosodic Systems in the Languages of Europe* (pp. 659-811). Berlin: Mouton de Gruyter.

Rogers, D., & d'Arcangeli, L. (2004). Italian. *Journal of International Phonetic Association 34(1)*, 117-121. doi: 10.1017/S0025100304001628.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621. doi: 10.1006/jmla.1996.0032.

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically-modulated sub-phonetic variation on lexical competition. *Cognition 105*, 466-476.

Schild, U., Becker, A. B., & Friedrich, C. K. (2014). Phoneme-free prosodic representations are involved in pre-lexical and lexical neurobiological mechanisms underlying spoken word processing. *Brain and Language 136*, 31-43.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition 42*, 213-259. doi: 10.1016/0010-0277(92)90044-I.

Shukla M., Nespor M., & Mehler J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology, 54(1)*, 1-32. doi: 10.1016/j.cogpsych.2006.04.002.

Silbert, L., Honey, Ch., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences 111(43)*, E4689-E4696.

Sluijter, A.M., & Van Heuven, V. (1996). Spectral tilt as an acoustic correlate of linguistic stress. *Journal of Acoustical Society of America, 100*(4), 2471-2485. doi: 10.1121/1.417955.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *The Journal of Neuroscience 32(25)*, 8443-8453.

Telkemeyer, S., Rossi, S., Koch, S., Nierhaus, T., Steinbrink, J., Poeppel, D., et al. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *The Journal of Neuroscience 29*, 14726–14733.

Toro, J. M., & Nespor M. (2015). Experience-dependent emergence of a grouping bias. *Biology Letters 11(9)*, 20150374. DOI: 10.1098/rsbl.2015.0374

Toro, J., Pons, F., Bion, R., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language, 64*, 171-180. doi: 10.1016/j.jml.2010.11.005.

Turk, A., & Shattuck-Hufnagel, S. (2007). Multiple targets for phrase-final lengthening in American English words. *Journal of Phonetics 35*, 445-472. doi: 10.1016/j.wocn.2006.12.001.

Tyler, M., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of Acoustical Society of America, 126*(1), 367-376. doi: 10.1121/1.3129127.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The role of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*, 133-149. doi: 10.1006/jmla.1997.2548.

Wakefield, J., Doughtie, E., & Yom, B. (1974). The identification of structural components of an unknown language. *Journal of Psycholinguistic Research 3*(3), 261-269.

Wheeldon, L. & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language 37*, 356-381. doi: 10.1006/jmla.1997.2517.

Wheeldon, L. & Lahiri, A. (2002). The minimalunit of phonological encoding: prosodic or lexical word. *Cognition 85*, B31-B41.

White, L., & Turk, A. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics 38*, 459-471. doi: 10.1016/j.wocn.2010.05.002.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of Acoustical Society of America 91*, 1707-1717. doi: 10.1121/1.402450.

Williams, B. (1985). Pitch and duration in Welsh stress perception: The implications for intonation. *Journal of Phonetics, 13*, 381–406.

Table 1. The overview of phonetic and phonological factors affecting lengthening in the German, Italian, Spanish and Gipuzkoan Basque languages.

	German	Italian	Spanish	Basque (Gipuzkoan dialect)
Main correlate of lexical prominence	F0, duration plays an inferior role	Duration, lengthening is the largest in penultimate syllable, no stress-induced lengthening in word-final syllables	Duration, but lengthening is less than in Italian in penultimate syllable	Duration
Unmarked location of stress in 3-syllabic words	Antepenultimate and penultimate	Penultimate	Penultimate	Changeable within a word depending on the position of the word in a phrase. Phrase-second and phrase-final syllables tend to be lengthened.
Presence of vowel phonemic durational contrast	Yes	No	No	No
Facilitatory effect on segmentation	Final lengthening	Penultimate lengthening	Penultimate lengthening	Final lengthening

Table 2. *t*-tests (2-tailed) comparing segmentation performance by participants with difference native languages with the chance level.

|		TP-only	Initial lengthening	Middle lengthening	Final lengthening
German speakers	t-value	t(23) = 3.718	t(23) = 2.731	t(23) = 2.391	t(23) = 9.335
	Significance	p = .001	p = .012	p = .025	p < .0005
	Effect size	r = .61	r = .49	r = .45	r = .89
Italian speakers	t-value	t(23) = 3.803	t(23) = 0.867	t(23) = 9.156	t(23) = 4.548
	Significance	p = .001	p = .359	p < .0005	p < .0005
	Effect size	r = .62	r = .18	r = .89	r = .69
Spanish speakers	t-value	t(23) = 4.239	t(23) = 1.716	t(23) = 2.376	t(23) = 4.554
	Significance	p < .0005	p = .1	p = .026	p < .005
	Effect size	r = .66	r = .34	r = .44	r = .69
Basque speakers	t-value	t(23) = 3.842	t(23) = 2.892	t(23) = 3.322	t(23) = .6.86
	Significance	p = .001	p = .008	p = .003	p < .0005
	Effect size	r = .63	r = .52	r = .57	r = .82
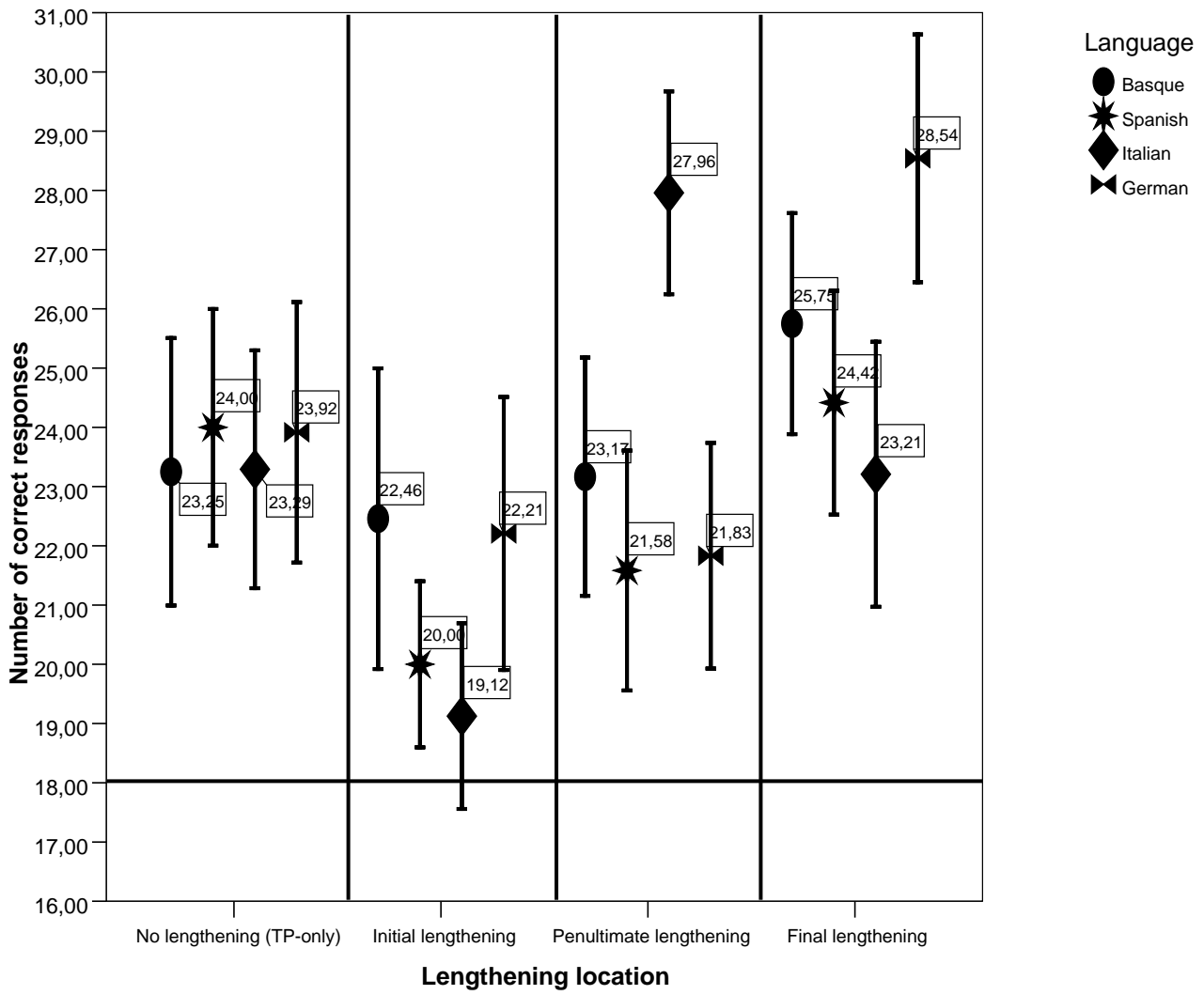
Figure 1. Segmentation performance (±2 standard errors) in the test phase for each condition and language group. Horizontal line shows chance level.