Title: Native language influence in the segmentation of a novel language

Running Had: Prosody and statistics in segmentation

Authors: Mikhail Ordin[1, 2], Marina Nespor[3]

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

[2.] Ikerbasque – The Basque Foundation for Science, Spain

[3.] International School for Advanced Studies of Trieste, Italy

ACKNOWLEDGEMENT

The research leading to these results has been supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC Grant Agreement Number 269502 (PASCAL), by the Basque Foundation of Science, Spain, and by the Alexander von Humboldt Foundation, Germany. Special gratitude goes to three reviewers and to the editor, whose extensive suggestions, comments and constructive critique resulted in substantial improvements in the final version of the article.

ABSTRACT

A major problem in second language acquisition (SLA) is the segmentation of fluent speech in the target language, i.e. detecting the boundaries of phonological constituents like words and phrases in the speech stream. To this end, among a variety of cues, people extensively use prosody and statistical regularities. We examined the role of pitch, duration, and transitional probabilities (TPs) between adjacent syllables in the segmentation of a novel language by native speakers of German and compared their responses with the segmentation by the listeners with a phonologically different native language: Italian. We used an artificial language with different prosodic cues marking the boundaries of statistically defined words. In artificial-language learning experiments, we compared how Germans and Italians use prosodic and statistical cues for segmenting continuous speech. We show that native phonology modulates the processing of prosodic cues in novel languages. While native speakers of Italian interpret prosodic cues at both the word and the phrasal level, native speakers of German interpret them exclusively at the phrasal level. Phrasal prosody can facilitate the segmentation of a novel language when prosodic and statistical cues lead to the same segmentation solution. Word-level prosody does not necessarily facilitate segmentation, but it can disrupt it when statistical and prosodic cues lead to different segmentation solutions.

Introduction

It is a common experience that segmentation of continuous speech into discrete words is difficult in an unknown language, while it is quite a trivial task in the native language of the listener. When people extract words from their native language or from any language they speak fluently, they exploit a wide variety of cues to detect word boundaries, and they usually allocate different weights to these cues. As shown in Mattys, White and Melhorn (2005), the cues are hierarchically integrated with decreasing power assigned to lexical (semantic and syntactic), segmental (phonotactic, statistical, allophonic) and suprasegmental (word stress) cues.

When listeners segment an unknown language, they do not have access to the most powerful and informative cue to segmentation, that is, knowledge of the lexicon. Yet adults are, to a certain extent, able to identify words and phrases in a natural language unknown to them without benefit of meaning (Doughtie & Yom (1974; Pilon, 1981; Endress & Hauser, 2010). For extracting discrete constituents from continuous speech in a novel language, lower-level cues including phonotactic, prosodic and statistical cues are available to listeners. The artificial language learning paradigm is frequently employed to explore the use of lower-level cues for the segmentation of an unknown language. The artificial language is used as a proxy to investigate extracting and processing novel patterns in natural languages. Multiple studies within this paradigm clearly demonstrated that listeners exposed to an artificial language tend to use the segmentation cues available in their native language and employ the segmentation strategies they have developed to process their native language (Vroomen, Tuomainen, & de Gelder, 1998; Toro, Pons, Bion, & Sebastian-Galles, 2011; Cutler, Mehler, Norris, & Segui, 1986; Murty, Otake, & Cutler, 2007; Kim, Davis, & Cutler, 2008; Tremblay, Coughlin, Bahler, & Gaillard, 2012).

Cues for the Segmentation of an Unfamiliar Language

A large body of research has been devoted to identifying the specific cues and strategies used for segmentation purposes. In our study, we concentrate on statistical and prosodic cues, which will be discussed separately in the next sections.

Statistical cues. Many studies have emphasized that people are able to track simple statistical regularities and use them for segmentation. Hayes and Clark (1970) used a limited inventory of phonemes and applied artificial phonotactic rules to combine phonemes into words so that certain phoneme pairs were more frequent when they were within words than when they were straddling word boundaries. After 45-minute familiarisation with this artificial language, participants judged the pause between words as more natural than that within words. The researchers hypothesized that participants calculated the transitional probability (TP) from phoneme A to phoneme

B by dividing the frequency of an AB phoneme pair by the frequency of the A phoneme in the speech stream. These TPs are calculated for each phoneme pair, and the troughs in TPs – i.e. the pairs with lower TPs – are used to detect word boundaries. Subsequent research on the role of statistical cues has provided evidence that listeners are able to track TPs between adjacent syllables for the purposes of speech segmentation in an artificial language (Saffran, Newport, & Aslin, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997).

Although TPs between adjacent syllables, in the absence of other cues, are sufficient for the successful segmentation of an artificial language (Saffran, et al., 1996), they are not the only statistical cues available to language learners for segmentation. Among other statistical cues that are successfully exploited are distributional properties of phonemes in words or syllables (Onishi, Chambers, & Fisher, 2002), non-adjacent TPs, i.e., TPs between initial and final syllables in trisyllabic words (Peña, Bonatti, Nespor, & Mehler, 2002), the relative frequency of speech constituents, that is assumed to be determined by the order of functors and lexical items (De la Cruz Pavia, Elordieta, Sebastian-Galles, & Laka, 2015; Gervain, Sebastian-Galles, Diaz, Laka, Mazuka, Yamane, Nespor, Mehler, 2013), as well as TPs between consonants (Bonatti, Peña, Nespor, & Mehler, 2005).

Prosodic cues. Besides statistical cues, also prosodic information (pitch, duration, intensity) can be used for segmentation. Prosodic parameters vary to manifest lexical stress, phrasal prominence and edges of phonological phrases (PP) and intonational phrases (IP), and listeners use prosodic cues to identify the boundaries of speech constituents.

Word-level prosody. Fluctuations of pitch (F0) and duration make one syllable acoustically more prominent and perceptually more salient than adjacent syllables, and thus mark lexical stress. Experimental studies indicate that listeners can use lexical stress at a word edge - right or left - for the segmentation of continuous speech in an unfamiliar language, if the stress location in the language of exposure matches that of the participants' native language (Tyler and Cutler, 2009; Cutler & Norris, 1988; Cutler, Mehler, Norris, & Segui, 1992). The basic principle is that language-specific biases for stress placement might provide reliable cues to the beginning or end of the word. The role of lexical stress in the segmentation of languages where stress is not aligned with one of the edges of words is underresearched. For example, we do not know if stress in a novel language that does not align lexical stress with word boundaries facilitates segmentation by listeners whose native language places lexical stress on the word-initial or word-final syllable.

Another factor that adds complexity to the issue of the use of word prosody in segmentation is the nature of stress. Acoustically, stress is manifested by duration, overall intensity and F0 (Gussenhoven, 2003, pp. 12-19). However, the precise acoustic manifestation of stress varies cross-linguistically. For example, in Dutch, compared

to English, stress coincides with a smaller reset of pitch (Johnson & Seidl, 2008). In French, lexical prominence is not manifested acoustically except in phrase-final positions (Dell, 1984). Differences in the acoustic correlates of stress have also been found between Catalan and Spanish (Ortega-Llebaria & Prieto, 2011), Italian and Spanish, German and Italian (Bissiri, Pfitzinger, & Tillman, 2008), Spanish and Italian (Alfano, Llisterri, & Savy, 2007), Welsh and English (Williams, 1985) and several other language pairs (Dogil & Williams, 1999). As the contribution of each acoustic correlate to word prominence varies across languages, it is not surprising that the weight of the acoustic correlates of stress perception also varies across speakers of different languages (Ortega-Llebaria, del Mar Vanrell, & Prieto, 2010, Peperkamp, Dupoux, & Sebastian-Galles, 1999, Alfano et al., 2007; Ortega-Llebaria & Prieto, 2009). We do not know for certain whether lexical stress will assist segmentation when it is manifested differently in the native and a novel language of the listener.

Phrase-level prosody. Experimental evidence regarding the use of PP-level and IP-level prosody for segmentation is more consistent. Tyler and Cutler (2009) showed that English, Dutch and French listeners exposed to an artificial language perceive the lengthening of a syllable as final lengthening, and benefit from the increase in duration aligned with the right-edge syllable in TP-defined words. Kim, Broersma, and Cho (2012) as well as Saffran et al. (1996) also report the facilitatory effect of final lengthening in the segmentation of an artificial language.

F0 contours mark the boundaries of IPs. As the edges of higher-level prosodic constituents always coincide with the edges of lower-level prosodic constituents (Nespor & Vogel, 2007), IP boundaries coincide with word boundaries and might thus also be relevant for the segmentation of speech into words. Shukla, Nespor, and Mehler (2007) showed that adults are able to exploit the prosodic markers of IP edges for word segmentation. They found that trisyllabic statistically defined words of an artificial language were segmented better when aligned with IP boundaries than when they occurred in the middle of IP contours. They also found that if a statistical word straddles two F0 contours, it is not recognized. Shukla et al. (2007) have also shown that participants can exploit F0 contours extracted from a foreign language and imposed on an artificial speech stream, for the purposes of segmentation. Langus, Marchetto, Bion, and Nespor (2012) showed that adults rely on prosodic information to discover hierarchy in continuous speech and use pitch declination to segment IPs, and lengthening cues to detect PP boundaries. IP-boundary cues (Kjelgaard & Speer, 1999; Warren, Grabe, & Nolan, 1995; Shukla et al., 2007) and PP-boundary cues (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Millotte, Rene, Wale, & Christophe, 2008; Langus et al., 2012) constrain the online syntactic analysis of speech and restrict lexical access. This suggests that final lengthening and F0 downdrift in speech processing by adults may be universal.

Research Interests

In the present study, we wanted to explore the influence of the native language on how the listener uses prosodic cues for the segmentation of an artificial language. For this purpose, we set up perception experiments with native speakers of German using the material and the research paradigm applied to study the use of prosody and TPs by native speakers of Italian (Ordin & Nespor, 2013). We then performed cross-linguistic analyses to reveal the differences in segmentation strategies employed by German and Italian listeners.

Ordin and Nespor (2013) systematically implemented lengthening, pitch raising, or a combination of both on one of the syllables of 3-syllabic statistical words in a continuous speech stream. The stream was presented to native Italian listeners, followed by the dual forced-choice task to recognize which of the two three-syllabic sequences was a word from the language they had heard. The results showed that lengthening, which is an acoustic correlate of penultimate lexical stress in Italian – the unmarked stress location - (Bertinetto, 1980), disrupts segmentation in an artificial language, if it does not fit the unmarked location of lexical stress in the native language of the participants. Otherwise segmentation is successful, but it is not facilitated by lengthening cues. Pitch cues impede segmentation when they are not aligned with the edges of statistical words. Pitch cues associated with the word boundaries neither facilitate, nor hinder segmentation. The combination of pitch raising and lengthening – a cue to prominence at the phrasal level – facilitates segmentation if the phonetic realization of pitch accents corresponds to that in the native language of the listener and is aligned with the word edges.

German and Italian represent two phonological systems that differ in the location and realization of prominence at the word and phrasal levels, in their phonotactics, in their phonetic inventory and in other aspects of phonology. If processing of an artificial language is indeed filtered through the L1 phonological systems for the purposes of segmentation, then we should expect differences in the use of prosodic cues that manifest prominence in the native languages of the participants. The majority of Germanic words have lexical stress on the first syllable of the stem, but a large size of German vocabulary consists of lexical items of foreign origin that are not stress-initial (Dogil & Williams, 1999; Jessen, 1999). Corpus analysis of Delattre (1965) revealed that the frequency of 3-sylalbic words with penultimate (51%) and antepenultimate (49%) stress in German does not differ significantly, while the 3-syllabic words with final stress are almost absent in the analysed texts. Consequently, we can say that Germans are equally familiar with both penultimate and antepenultimate stress patterns.

Lexical stress in German is signalled by F0 movements above a critical threshold (Isachenko & Schädlich, 1966). However, Isachenko & Schädlich investigated words in isolation and in short carrier phrases, in which the

stressed syllable also bears phrasal accent. That is why these results could not reliably demonstrate whether the F0 movement was associated with lexical stress per se, or with phrasal prominence. Later studies tried to disentangle the effect of phrase-level and word-level prominence on acoustic parameters. Dogil and Williams (1999) validated that in German, duration can be a reliable perceptual correlate of word stress in unaccented syllables, i.e., syllables that bear lexical stress but not phrasal prominence. Kohler (2012) argued that there is no fixed ranking of variables which signal stress in German. Depending on their phonetic realization, some cues can outweigh other cues, and both duration and pitch are important perceptual correlates of stress. Therefore, native speakers of German are familiar with a wider range of stress patterns – in terms of location and realization of lexical prominence – than native speakers of Italian. Relevant details about Italian phonology and phonetics are presented in Ordin and Nespor (2013), both in the introduction and in the supplementary material.

It should also be noticed that German, unlike Italian, has a phonological opposition of long and short vowels. That is, words with stress on the first syllable can have a longer vowel in the following unstressed syllable, if the unstressed vowel is phonemically long and the stressed vowel is phonemically short. Thus, length in German is also involved in coding phonemic contrasts. German listeners may perceive the lengthening of a vowel in a novel language not only as a marker of lexical stress in unaccented position, but also as a phonemic feature. In Italian vowel lengthening is never phonemic, and Italians are more likely than Germans to interpret the increase in duration as a prosodic correlate of stress. Italians may perceive unstressed penultimate syllables with a long vowel in German trisyllabic phonological words (PWs) as stressed (Bissiri, 2008; Bissiri et al., 2008). Bissiri et al. (2008) used the method developed by Dupoux, Peperkamp and Sebastian-Galles (2001) to study the perception of stress in German words by Italian listeners. They showed that in a sequence repetition task with German words, Italians performed very well on phoneme contrast detection (i.e. they could easily differentiate between long and short vowels), but had a significantly poorer performance on stress contrast detection (had difficulties in differentiating trisyllabic words with long penultimate vowels and either penultimate or antepenultimate stress). The authors suggest that Italians perceive German complex words as always having the same stress pattern because their stress perception is influenced by the phonology of their native language, which does not have phonemic contrasts in vowel duration and exhibits the unmarked stress location – with duration as the main correlate of lexical stress – on the penultimate syllable.

In sum, German and Italian differ in the unmarked location of lexical stress (aligned with the penultimate syllable in Italian and more variable in German), in the relative weights of the acoustic and perceptual correlates of

stress, and in the presence/absence of long and short vowels in the phonemic inventory. German and Italian have sufficiently different phonological systems to reveal the influence of L1 phonology on segmentation.

Experiment 1

In the first experiment, we investigated the influence of lengthening cues on the segmentation of an artificial language by Germans, and then compared the performance of native German listeners with that of native Italian listeners. Durational cues can potentially be processed in three different ways. First, the increase in duration can be interpreted as PP-final lengthening and thus as a cue to the right edge of the phonological phrase and consequently also the end of the word. Second, duration can be interpreted at the word level as a correlate of lexical stress. Third, lengthening can be perceived as a phonemic distinction between long and short vowels.

Participants

Twenty-four speakers of German were recruited from the pool of undergraduates at Bielefeld University in North Rhein Westfalia. None either reported or manifested any speech or hearing disorders. Each participant received monetary compensation. All speakers were from monolingual families and were raised in monolingual environments, none were regularly using foreign languages, but all had learned English as a compulsory school subject.

Stimuli

The artificial language for the experiment was the same as that used in Ordin and Nespor (2013). For the readers' convenience, we provide here the necessary details regarding the material. The language was created with an inventory of five vowels and 11 consonants: /i/, /e/, /a/, /o/, /u/, /k/, /m/, /p/, /b/, /l/, /t/, /g/, /v/, /n/, /f/, /d/. Concatenations of the phonemes resulted in a set of 12 three-syllabic words consisting of consonant-vowel (CV) syllables (*komipa, bolatu, kupige, vunelu, bamofe, defida, bukite, vifole, dubipo, vaputa, donume, ginefa*). These words were split into two equal groups to create two speech streams. The 166 tokens of each word within each stream were concatenated randomly with the only restriction that the adjacent repetition of words was prohibited. Two speech streams had TPs between adjacent syllables within words of 1.0 and TPs between adjacent syllables straddling word boundaries of .15. Speech synthesis was done using the MBROLA speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996) and a female German voice (de5 diphone database).

Following Peña et al. (2002) and Tyler and Cutler (2009), we assigned an equal base length of 100 ms to each vowel and consonant and flat F0 of 200 Hz with 5-second fade-in and fade-out at the edges of the streams. The resulting streams were presented auditorily to three German listeners who were asked to listen to any part of the stream that sounded like a real German word. The listeners, who were aware of the purpose of the experiments, did not find any German words in the speech streams during the stream evaluation. As these streams contained TPs between adjacent syllables as the only cues for segmentation, we refer to them as TP-only.

The streams were then modified prosodically. The first, the second, or the third vowel in each TP-defined word was lengthened by 80 ms.

The present research investigates the mechanisms that allow people to segment continuous speech despite the fact that the acoustic implementation of the same linguistic event may differ across languages. When people learn to segment speech in an unknown language, they apply the phonological and phonetic regularities of their native language to the incoming speech stream. The artificial language in our study is used as a proxy to explore the extraction, learning and use of novel patterns of natural languages that are not necessarily similar to the patterns of the native language of the listener. Therefore, we avoided implementing German-specific phonetic patterns into the artificial language and rather concentrated on ensuring compatibility with other studies on segmentation of an artificial language. The values for lengthening were also chosen in an attempt to ensure optimal compatibility across different studies. This approach could enable us to investigate whether and how participants use their native phonology when segmenting an artificial language.

Procedure

Each participant came for the experiment twice with a two-week interval between sessions. In the first session, s/he was exposed to stream 1 and stream 2 in two different conditions, and in the second session – to stream 1 and stream 2 in the other two conditions. The combination of stream, condition, and order of presentation was randomized for each participant. Ordin and Nespor (2013) showed that for Italian participants, the segmentation performance did not differ either between sessions, order of presentation within sessions, or between streams (i.e., artificial languages). The listeners' ability to easily discriminate between two subsequently presented artificial languages has also been shown in a more recent study (Franco, Cleermans, & Destrebecqz, 2011). Gebhart, Aslin and Newport (2009) also demonstrated that interference between two subsequently presented languages takes place only when the familiarisation exposure is short, when two languages are not separated by the pause, and when the instructions do not specify the presence of different languages. If any of these three conditions are not met,

interference disappears. We thus assume that there was no influence of one stream on the other during familiarization or test in our experiments.

After the exposure to one of the streams, we set up a dual forced-choice task. We asked participants to listen to pairs of imaginary words and decide which of the two they thought had been presented in the familiarization language. In the test stimuli, prosodic cues were removed and listeners could use only segmental and statistical information. Six words were pitted against six partwords. Words and partwords in each of the 36 pairs were separated by a 500-ms pause. The order of words and partwords in the pairs was counterbalanced. Participants had to click button "1" if they thought that the first item in the pair was a real word from the language they had listened to, and button "2" if they thought that the second item in the pair was a real word. Upon completing the test, participants had a 5-minutes pause before the procedure was repeated with a different stream*condition.

The number of correctly chosen words represents segmentation performance. Comparing segmentation performance with chance level (50%) shows whether segmentation is overall successful. For the segmentation to be successful, the number of correct responses should be significantly above chance. Previous research allows us to predict that the segmentation is successful when only TPs between adjacent syllables are available (Saffran, et al., 1996). Planned comparisons were performed between segmentation performance in the TP-only condition and that in conditions with implemented prosodic cues, in order to explore if prosody facilitates segmentation (the number of correct responses is significantly higher in a condition with prosody than in the TP-only condition) or impedes segmentation (the number of correct responses drops to the chance level or is significantly lower than in the TP-only condition). If prosodic cues neither facilitate nor impede segmentation, performance is significantly above chance but does not differ from performance in the TP-only condition. For cross-linguistic analysis, we used mixed ANOVA tests with *the presence and location* of the prosodic cues within statistical word as a within-subject factor, and with *L1* of the listener as a between-subject factor, followed by the planned comparisons with the Bonferroni-Holm correction for multiple comparisons. This analysis is different from the one undertaken in Ordin and Nespor (2013) on Italian data, where a series of one-sample t-tests (against 50% of the correct responses expected by chance) was carried out to identify in which conditions the segmentation by Italians was successful. In this study, we focused on impeding and facilitatory effect of prosodic cues compared to TP-only condition. The different objectives of this study and the need to add the analysis of cross-linguistic differences explains a more sophisticated approach to the data analysis.

Results

For the cross-linguistic analysis, we first constructed a mixed ANOVA model with *lengthening presence and location* (TP-only, word-initial, penultimate or word-final lengthening) as a within-subject factor, and the *L1* of the listener (Italian vs. German) as a between-subject factor, and *segmentation performance* (i.e., the number of correct responses) as a dependent variable. The data for the *segmentation performance* by Italian listeners is taken from Ordin and Nespor (2013). The analysis showed that the effect of *L1* was significant, $F(1, 46) = 4.497$, $p = .039$, $\eta_p^2 = .089$: the overall segmentation performance of German listeners was higher than that of Italian listeners. The effect of the *lengthening presence and location* was not significant, $\lambda = .905$, $F(3, 44) = 1.538$, $p = .218$, $\eta_p^2 = .095$. However, the model revealed a significant effect of interaction between *L1* and *lengthening presence and location* on *segmentation performance*, $\lambda = .681$, $F(3, 44) = 6.877$, $p = .001$, $\eta_p^2 = .319$. This shows that the influence of the *lengthening location* differs depending on the native language of the listener. Thus the influence of the *lengthening location* should be investigated separately for the group of German and Italian listeners.

We performed the ANOVA analysis to assess the effect of the *lengthening presence and location* on the number of correct responses only by Germans. The effect of the *lengthening presence and location* on segmentation is significant, $\lambda = .557$, $F(3, 21) = 5.578$, $p = .006$, $\eta_p^2 = .443$. Contrasts reveal that performance by German listeners is significantly better compared to the TP-only condition when the final syllable is marked by duration, $F(3, 23) = 8.145$, $p = .009$ (0.017)[1], $\eta_p^2 = .262$. Performance in segmentation by German listeners does not differ from the TP-only condition when either the penultimate syllable, $F(3, 23) = 1.138$, $p = .263$, $\eta_p^2 = .054$, or the initial syllable is lengthened $F(3, 23) = .325$, $p = .574$, $\eta_p^2 = .014$. Thus the lengthening of the final syllable facilitates segmentation, while the lengthening of either the penultimate or the initial syllable does not affect segmentation by German listeners. A comparison of the performance of German listeners with the chance level shows that German participants segmented reliably above chance in all conditions ($t(23) = 3.77$, p=.001 (.017), r=.62 for TP-only condition, $t(23) = 2.944$, p=.007 (.025) for initial lengthening, r=.52, $t(23) = 2.569$, p=.017 (.05) for penultimate lengthening, r=.47, $t(23) = 9.466$, p<.0005 (.0125), r=.89 for final lengthening).

Ordin and Nespor (2013) showed that segmentation by Italians is significantly above chance for the TP-only condition and for penultimate lengthening. However, penultimate lengthening does not facilitate segmentation. Segmentation fails when word-final and word-initial syllables were lengthened. Figure 1 provides

[1] For each significant contrast, the actual p value is reported first, followed by the corrected alpha (i.e., maximum p value for the contrast to remain significant with 5% error I threshold). Corrections are performed by the Bonferroni-Holm method. That is, $p = .009$ (.017) means that the actual p value is *.009, but for the contrast to remain significant* after correcting for the multiple comparisons, the p value should not exceed .017. As 0.009 < .017, the contrast remains significant after the correction.

the mean number of correct answers for each condition (location of lengthening) for German and Italian participants.

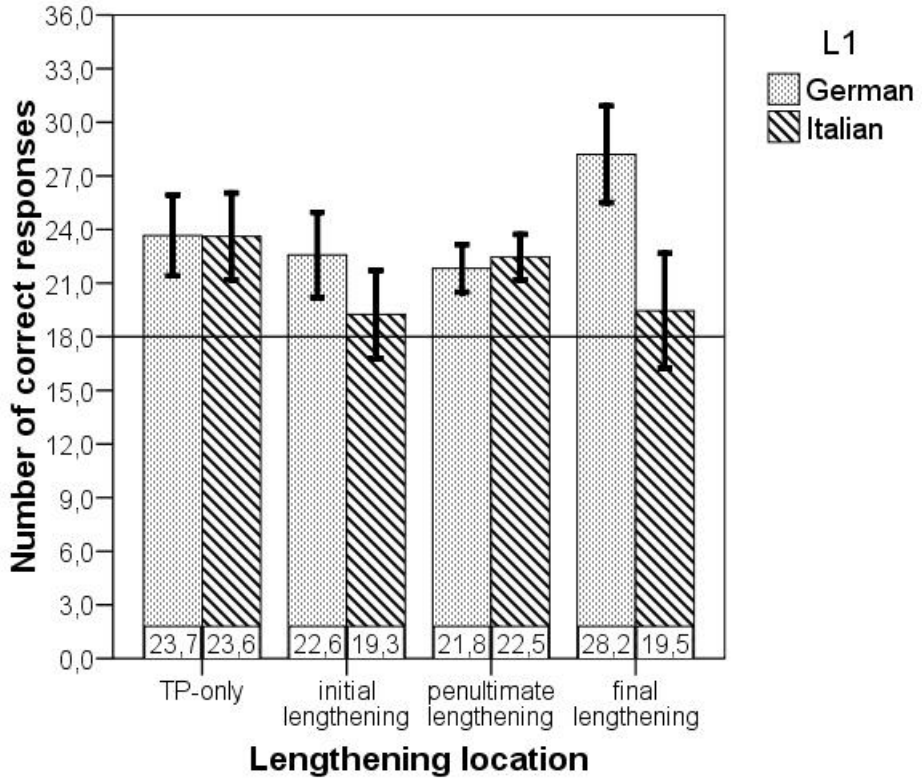

Figure 1. Mean number (±2 standard errors) of correct answers for each lengthening location and native language of the participants. The horizontal line stands for the chance level (18 correct answers).

Discussion

TPs between adjacent syllables provide sufficient cues for successful segmentation, which is in line with the large body of previous research (e.g., Saffran, et al., 1996; Shukla et al., 2007; Ordin & Nespor, 2013, etc.). When prosodic cues are added, they interact with TPs and may affect the segmentation performance. German listeners benefit when the word-final syllable is lengthened, in line with the result patterns obtained with native Korean, English, Dutch and French listeners (Tyler & Cutler, 2009; Kim, Broersma, & Cho, 2012). It has been assumed that the processing of lengthening cues in an artificial language is universal because longer syllables are interpreted as a right-edge boundary signal. According to that line of interpretation, we assume that Germans process the lengthening cues as the markers of the right edge of the PP, and the lengthening of the word-final syllables facilitates segmentation.

The increase in vowel duration of the lengthened syllables amounted to 80ms. This increase roughly corresponds to the averaged increase of vowel duration in the vicinity of a phrase boundary (Wellman, Holzgrefe, Truckenbrodt, Wartenurger, & Hoehle, 2012; Peters, 2005; Peters, Kohler, & Wesener, 2005). The differences in

duration between corresponding long and short vowels in German are smaller than 80ms. However, the durational ratio between tense vowels in stressed and unstressed syllables may even exceed 1:2 (Mooshammer, Fuchs, & Fischer, 1999), and easily reaches the increase from 100 to 180 ms. – the values we implemented in our artificial languages. Also Mixdorf, Hayashi and Ushiyama (2015) showed that the differences in duration between pre- and post-boundary syllables in German are similar to the durational differences between stressed and unstressed syllables. Consequently, Germans could potentially perceive lengthening in the artificial language as a cue to lexical stress. However, even when the lengthening in the artificial language corresponded to the most frequent location of lexical stress in 3-syllabic German words (either penultimate or antepenultimate), segmentation was not improved. Therefore we conclude that word-level prosody (i.e., lexical stress) does not facilitate segmentation, or that German listeners do not interpret lengthening cues as acoustic correlates of lexical stress. The first alternative would correspond to the interpretation proposed by Ordin and Nespor (2013) who suggested that word-level prominence does not facilitate segmentation, but it can disrupt it if the location of the lexical stress in the native language of the learner does not correspond to that in the artificial language. Both penultimate and antepenultimate syllables of 3-syllabic words can bear stress in German, thus word-initial and penultimate lengthening in the statistical words of the artificial language do not disrupt segmentation. However, only final lengthening, being a PP-level prosodic cue, facilitates segmentation.

Tyler and Cutler (2009) and Kim et al. (2012) claimed that lengthening on a final syllable always facilitates segmentation in a novel language, whatever the native language of the participant is. Listeners are likely to interpret word-final lengthening as a PP-final cue because final lengthening is a universal phenomenon. However, word-level prominence in the native languages of participants (English, Dutch, French, Korean) in their experiments was aligned with the word edges. In addition, duration has less weight in the manifestation of lexical stress in the languages selected by Tyler and Cutler (2009) and Kim et al. (2012) compared to Italian. Therefore, the results we obtained with Italian participants (Ordin & Nespor, 2013) neither contradict earlier studies, nor our results with German participants, but rather show that under certain conditions, prosodic lengthening in a novel language can be filtered through the phonology of the participants' native language, and can be reinterpreted as a correlate of lexical stress.

Experiment 2

An alternative explanation for the result pattern of our first experiment is that Germans do not perceive pure lengthening as a cue to lexical stress, and therefore the potential facilitatory effect of stress did not emerge. Unlike

Italian, where lengthening is a major acoustic correlate of lexical stress, German exhibits a more complex interplay between F0 and duration in prominence at the word level. In the second experiment, we investigated the influence of a combination of lengthening and pitch raising in the segmentation of an artificial language by Germans, and then compared the performance of native German listeners with that of native Italians.

Participants

Twenty-four native speakers of German who did not participate in the previous experiment were recruited and received a monetary compensation. We used the same pool and the same criteria to recruit participants as in experiment 1.

Stimuli

The speech stimuli from experiment 1 were also used in experiment 2, but simultaneous lengthening and F0 increase were used in the familiarization streams as prosodic cues. Either the first, the second or the third vowel in each word was lengthened by 80 ms, and the F0 was increased to 240 Hz forming an inverted parabola shape of F0 contour on the lengthened syllable. The same F0 contour was used in other experiments, e.g., Tyler and Cutler (2009) and Thiessen and Saffran (2003). The values of F0 fluctuations (in semitones), like the values of duration in this experiment as well as in experiment 1, were chosen in an attempt to ensure optimal compatibility across studies. Please refer to Ordin and Nespor (2013) for a more detailed description of the experimental material.

Procedure

The procedure was the same as in experiment 1.

Results

For the cross-linguistic analysis, we first constructed a mixed ANOVA model with *prominence presence and location* (TP-only, word-initial, penultimate or word-final prominence) as a within-subject factor, the *L1* of the listener (Italian vs. German) as a between-subject factor, and *segmentation performance* (i.e., the number of correct responses) as a dependent variable. The data for segmentation performance by Italian listeners is taken from Ordin and Nespor (2013). The analysis showed that the effect of the *L1* was not significant, $F(1, 46) = 2.393$, $p = .129$, $\eta_p^2 = .049$, with Germans and Italians giving on average the same percentage of correct responses. The effect of *prominence presence and location* was significant, $\lambda = .559$, $F(3, 44) = 11.567$, $p < .0005$, $\eta_p^2 = .441$. Interaction between *L1* and *prominence presence and location* was not significant, $\lambda = .976$, $F(3, 44) = .368$, $p = .776$, $\eta_p^2 =$

.024. This shows that the influence of *prominence presence and location* on the number of correct responses is similar cross-linguistically.

Ordin and Nespor (2013) showed that native speakers of Italian benefit in segmentation when the word-final syllable is marked by a simultaneous increase in duration and a raise in pitch. The ANOVA analysis on the number of correct responses by German listeners as a function of *prominence presence and location* also revealed a significant effect, λ = .579, $F(3, 21) = 5.084$, p = .008, η_p^2 = .421. Contrasts revealed that final prominence has a facilitation effect on segmentation by native speakers of German, $F(1,23) = 5.847$, p = .024 (.025), η_p^2 = .203. Prominence on the penultimate syllable impedes segmentation, $F(1,23) = 4.507$, p = .045 (.05), η_p^2 = .164. Prominence on the word-initial syllable in 3-syllabic words does not affect segmentation, $F(1,23) = .454$, p = .507, η_p^2 = .019.

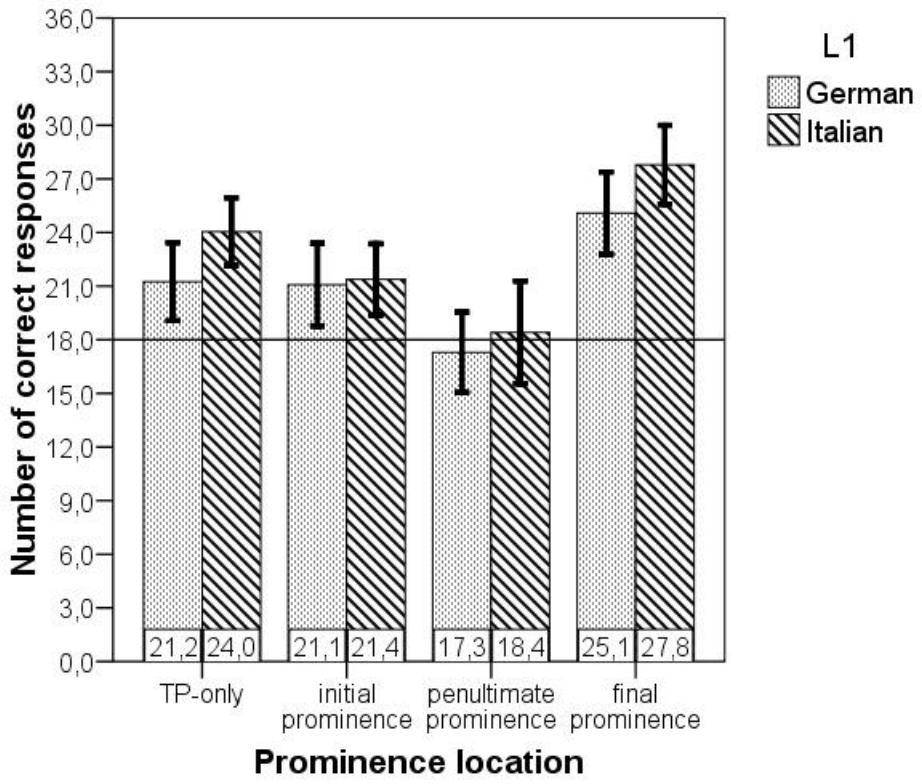

Figure 2. Mean number (±2 standard errors) of correct answers for each *prominence location* and native language of the participants. The horizontal line stands for the chance level (18 correct answers).

One-sample *t*-tests were performed to compare the number of correct answers in each condition with the chance level (50%). The tests showed that in the middle-prominence condition the segmentation was not only impeded, but it was disrupted (not significantly different from chance), t(23)=-.558, p=.582, r=.12. In other conditions participants' segmentation was significantly above chance, t(23)=2.806, p=.01 (.025), r=.5 for TP-only condition, t(23)=2.266, p=.033 (.05), r=.18 for word-initial prominence, t(23)=5.041, p<.0005 (.0125), r=.72 for word-final

prominence. The same result pattern was revealed for the segmentation of the same artificial language by Italians (Ordin & Nespor, 2013), which can also be seen in figure 2.

Discussion

The results showed that Germans, like Italians, benefit for segmentation if the word-final syllable is marked by the simultaneous increase of duration and F0. This challenges the alternative explanation of the results of the first experiment: word-level prominence does facilitate segmentation, but Germans do not perceive pure lengthening as a correlate of stress in the artificial language, and need an additional F0 manipulation. If that were the case, we should have expected a facilitatory effect of lengthening accompanied by pitch raising on the word-initial and penultimate syllables in the artificial language. However, no facilitatory effect has been detected. Moreover, placing prominence on the penultimate syllable disrupted segmentation. Thus we rejected the alternative interpretation of the first experiment's results and confirmed the proposed interpretation of the results for the first experiment: word-level prominence does not facilitate segmentation.

We concluded that simultaneous increase in duration and pitch is not perceived by German listeners as lexical stress, and is rather processed at the PP-level. Introducing the cross-linguistically universal feature of final lengthening can indeed facilitate the segmentation of an artificial language, since listeners interpret lengthening as a boundary signal for the right edge of the PP. Germans perceive duration on the word-final syllable as PP-final lengthening, both if it is pure lengthening and if lengthening is accompanied by simultaneous pitch raising, which tends to attract listener's attention to the edge of constituents. Italians perceived lengthening on the word-final syllable as PP-final lengthening only when duration is accompanied by a pitch cue. When duration is not accompanied by a pitch cue, Italians process it at a lower hierarchical level, i.e. as a cue to lexical stress. However, if lexical stress is on the final syllable of a word in Italian, the vowel is not lengthened (Rogers & d'Arcangeli, 2004). Thus, segmentation by Italians in the lengthening-final condition is disrupted. The interpretation that Italians may process lengthening cues at different levels of the prosodic hierarchy while Germans are restricted to processing lengthening only as a PP-final marker calls for an additional series of experiments. Therefore, we conducted the experiments described below to support or refute this proposal.

Experiment 3

The third experiment is aimed at exploring whether Italians can indeed process lengthening at different hierarchical levels, and switch between processing durational increase as a correlate of lexical stress and as a PP-boundary

marker, while Germans are more constrained in their interpretation of prosodic lengthening in continuous speech.

Participants

We recruited eighteen German listeners from the same pool of participants of the previous experiments. Eighteen Italian listeners were recruited from the pool of participants reported in Ordin and Nespor (2013), namely, monolingual Italian first- and second-year students from the University of Trieste without broad experience of living abroad. None used any foreign language on a regular basis and none had been exposed to any second language in childhood, until starting to learn English as a compulsory school subject.

Stimuli

We have used the same inventory of syllables as in the previous experiments. Six three-syllabic sequences that defined the statistical words for the first experiment were repeated 166 times in the same order so that the TPs between syllables were set to 100% throughout. A stretch from the generated stream is given below:

…**FE** DE FI DA KO MI **PA** BO LA TU **KU** PI GE VU **NE** LU BA MO **FE** DE FI DO KO MI **PA** BO LA…

The duration of each syllable was 200 msec, duration of bold underlined syllables was increased by 80 msec by lengthening the vowel. The total duration of the stream is 10.85 minutes.

For the dual forced-choice task, we selected four three-syllabic sequences with lengthening on the final syllable (e.g., *komi**pa***), four sequences with penultimate lengthening (e.g., *ma**pi**bo*), and also four sequences with no lengthening (*e.g., bolatu*). Each sequence with lengthening on the third syllable was paired against two possible sequences with penultimate lengthening. Each pair was created in two versions: the sequence with final lengthening followed by the sequence with penultimate lengthening, and the two sequences in the reverse order. This amounted to 16 pairs of sequences. Similarly 16 pairs were created, in which one sequence with lengthening on the final syllable in the stream was paired with a sequence with no lengthening. Finally, 16 pairs were created to pit sequences with penultimate lengthening against sequences with no lengthening. All syllables in the pairs were made equal in duration (200 msec). The sequences in pairs were separated by a 500 ms. pause. Similar streams and pairs were created with the syllables used for the second stream in the previous experiments.

Procedure

Each participant had to come for the experiment once. S/he was exposed to one of the two streams (9 listeners of each L1 group per stream). After familiarization, they had to perform a dual forced-choice test. 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. We asked them to give the first answer that comes to mind. The order of the pairs in the test was randomized for each participant. As the streams did not provide any statistical or other cues for segmentation, the only signals people could exploit were the lengthening cues.

Participants did a dual forced-choice task with 3 types of pairs, each type consisting of 16 pairs. If the listener did not have preference for or against any prosodic pattern (final vs. penultimate lengthening, final vs. no lengthening, penultimate vs. no lengthening), we would expect them to choose the paired sequences within each type with equal frequency. If, for example, participants have no preference choosing between sequences with final and penultimate lengthening, both sequences within final- vs. penultimate lengthening pair-type should be chosen an equal number of times. If we compare for this pair-type how many times the listeners select final-lengthening sequences with chance (50%), the lack of significant difference will indicate the lack of preference. If the number of times final-lengthening sequences are chosen is significantly higher than 50%, this will indicate preference for final lengthening against penultimate lengthening. If the number of times final-lengthening sequences are chosen is lower than 50%, this will indicate preference for penultimate lengthening against final lengthening. A similar argumentation is applied to the other pair-types.

Results

Figure 3 presents the results indicating how many times participants chose sequences with final lengthening instead of those with no lengthening, final lengthening instead of penultimate lengthening, and penultimate lengthening instead of no lengthening. One-sample t-tests were run to compare the number of listeners’ responses within each pair-type with the chance level (table 1).

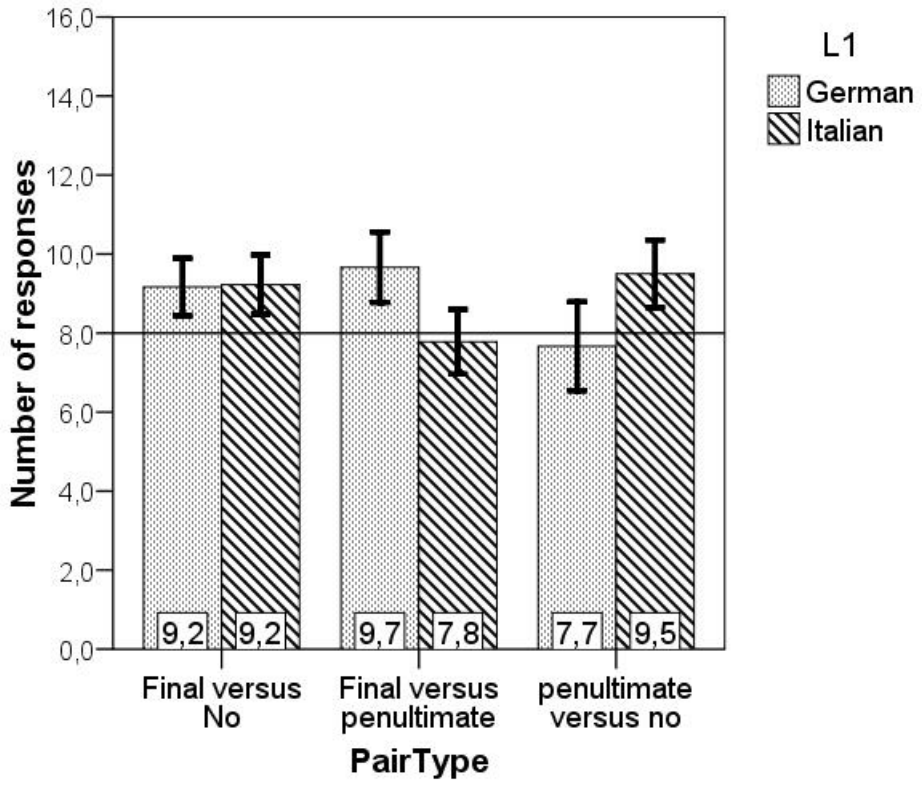

Figure 3. Preferences of German and Italian listeners for different prosodic patterns (see text for explanation). The horizontal line stands for the chance level (8 responses per condition). Error bars represent ±2SE.

Table 1. Statistic and significance values for one-sample t-tests in experiment 3.

		Final lengthening vs. no lengthening	**Final lengthening vs. penultimate lengthening**	**Penultimate lengthening vs. no lengthening**
German speakers	**t-value**	$t(17) = 3.226$	$t(17) = 3.748$	$t(17) = -.594$
	Significance	$p = .005$ (.025)	$p = .002$ (.017)	$p = .56$ (.05)
	Effect size	r=.399	r=.425	r=.184
Italian speakers	**t-value**	$t(11) = 4.618$	$t(11) = .662$	$t(17) = 2.633$
	Significance	$p = .001$	$p = .552$	$p = .023$
	Effect size	r=.544	r=.238	r=.439

The results show that Germans have preferences for the sequences with final lengthening when they have to choose either between sequences with final lengthening vs. penultimate lengthening, or between sequences with final lengthening vs. no lengthening. This suggests that Germans are likely to process an increase in duration only as final lengthening, i.e., as a cue to the right edge of a prosodic constituent. As final lengthening is a prosodic phenomenon that operates at a PP level, we conclude that Germans process the implemented increase in duration at the PP level, and the segmented speech constituents roughly correspond to PPs. However, when Germans have to choose between sequences with penultimate lengthening vs. sequences with no lengthening, they make their choice

randomly and respond at chance. Natural languages make abundant use of durational contrasts at different levels of the prosodic hierarchy, and the sequences with no durational contrasts are dispreferred because they are not plausible in speech. As German listeners do not have preferences for penultimate lengthening against no lengthening, we concluded that Germans interpret both types of sequences as not legitimate candidates for structural speech constituents, and respond randomly when they have to choose between two dispreferred candidates.

Italians, instead, respond at chance when they have to choose between sequences with final lengthening vs. sequences with penultimate lengthening. However, they have a significant preference for sequences with final lengthening to sequences without lengthening. Choosing between sequences with penultimate lengthening vs. no lengthening, Italians go for penultimate lengthening significantly more often than for no lengthening. Sequences without lengthening are supposedly dispreferred candidates for segmented units because natural languages do use durational cues within speech constituents. As sequences with penultimate lengthening are preferred to those without lengthening, and sequences with final lengthening are selected as frequently as those with penultimate lengthening, we assume that Italians interpret penultimate lengthening as a legitimate pattern for segmented speech constituents. This indicates that both types of lengthening can be used by Italians for segmentation. As penultimate stress-induced lengthening is the most frequent pattern at the PW level, we assume that penultimate lengthening is processed by Italians at the PW level, and final lengthening at the PP level.

Discussion

The results of this additional experiment confirm cross-linguistic differences in interpreting syllabic lengthening in artificial languages. Final lengthening is interpreted as a legitimate pattern both by Germans and Italians, but only Italians perceive penultimate lengthening as a legitimate prosodic pattern. The processing of lengthening cues by Italians depends on the availability of other cues (e.g., TPs, pitch) that lead the listener to conclude whether the segmented constituent is a PP or a PW. In Ordin and Nespor (2013), we proposed that final lengthening disrupts segmentation by Italians because listeners processed the segmented constituents as words, and therefore processed lengthening as a correlate of lexical stress. In the reported experiment (3) of the current study, TPs were not informative for segmentation, and that is why Italian listeners could not decide whether lengthening had to be processed at the word or the PP level, thus both patterns were accepted. This proposal needs further empirical support in future research.

Experiment 4

In order to investigate how listeners use prosodic cues that do not exist in their native languages, we set up the last series of experiments. We used a local F0 peak to mark one of the syllables in statistical trisyllabic words. In intonational languages, local F0 fluctuations are associated with accented syllables that bear lexical stress, and the stressed syllable is also marked by an increased duration. In our experimental stimuli we added F0 cues without lengthening of the corresponding syllable.

Participants

Twenty-four native speakers of German who did not participate in the previous experiments were recruited and received a monetary compensation. We used the same pool and the same criteria to recruit the participants as in the previous experiments.

Stimuli

The stimuli from experiment 1 were also used in experiment 4, but the prominence in the familiarization speech stream was manifested only by an F0 inverted parabola contour either on the final, the penultimate or the initial syllable, with F0 peak reaching 240 Hz. Duration was kept constant on all syllables. We generated the streams in four different conditions, TP-only, initial-pitch, middle-pitch and final-pitch.

Procedure

The procedure was the same as in experiments 1 and 2.

Results

During our cross-linguistic analysis, a mixed ANOVA with *pitch presence and location* as a within-subject factor and the *L1* of the listener (German vs. Italian) as a between-subject factor (the data for segmentation performance by Italian listeners is taken from Ordin & Nespor, 2013) showed insignificant effect of the native language, $F(1, 46) = 0.14$, $p = .906$, $\eta_p^2 < .0005$, insignificant effect of interaction between *L1* and *pitch presence and location*, $\lambda = .97$, $F(3, 44) = .459$, $p = .712$, $\eta_p^2 = .03$, and significant effect of *pitch presence and location*, $\lambda = .7967$, $F(3, 44) = 9.761$, $p = .017$, $\eta_p^2 = .204$. The mean numbers of correct answers for each *pitch presence and location* and *L1* are presented in Figure 4. These results indicate that there are differences in the number of correct responses between conditions, and the differences between conditions are similar in both language groups.

However, a separate ANOVA performed only on the number of correct responses by German listeners revealed that *pitch presence and location* does not significantly affect the segmentation of an artificial language by

Germans, if prominence is manifested only by pitch, λ = .841, $F(3, 21) = 1.323$, p = .293, η_p^2 = .159. Planned contrasts comparing performance by German listeners in the TP-only condition with pitch-initial, F(1,23) = .104, p = .75, η_p^2 = .005 pitch-middle, $F(1,23) = .496$, p = .488, η_p^2 = .021 or pitch-final, $F(1,23) = .782$, p = .386, η_p^2 = .033 conditions are also not significant. Comparing the *number of correct responses* for each *pitch location* with the chance level shows that segmentation is above chance is all conditions, t(23)=3.65, p=.001 (.025), r=.61 for the TP-only condition, t(23)=4.19, p<.0005 (.0125), r=.66 for the pitch-initial, t(23)=3.04, p=.006 (.05), r=.31 for the pitch-penultimate, t(23)=4.77, p<.0005 (.017), r=.71 for the pitch-final conditions. This shows that solely pitch cues neither impede nor facilitate segmentation by German listeners. This contradicts the reported results of the mixed ANOVA showing that the effect of *pitch location* was significant, while there was no significant effect of *L1*. This contradiction encouraged us to analyse the data for the Italian participants in a separate RM ANOVA. The effect of *pitch location* turned out to be significant for Italian participants, λ = .796, $F(3, 21) = 3.073$, p = .05, η_p^2 = .3.5. Contrasts revealed that when pitch marks the penultimate syllable in three-syllabic words, segmentation is impeded by Italian listeners, $F(1,23) = 4.41$, p = .047, η_p^2 = .161. Segmentation is not affected - neither facilitated nor impeded - when the initial syllable, $F(1,23) = .402$, p = .532, η_p^2 = .017, or the final syllable, $F(1,23) = 1.224$, p = .28, η_p^2 = .051 are marked by a pitch increase.

Ordin and Nespor (2013) reported that Italians – unlike Germans in the present study – failed to segment when the pitch cue was aligned with the penultimate syllable of the statistical words. The number of correct responses by Italians in the pitch-penultimate condition is significantly lower than that in the TP-only condition, and this is sufficient to make the effect of *pitch presence and location* significant in the mixed analysis. Yet the differences in the number of correct responses between conditions go in a similar direction (figure 4) in both language groups: both Germans and Italians segment less successfully in the pitch-penultimate condition. Thus the interaction effect was not significant. The differences were big enough to reach significance in the sample of Italian listeners, but not in the sample of German listeners.

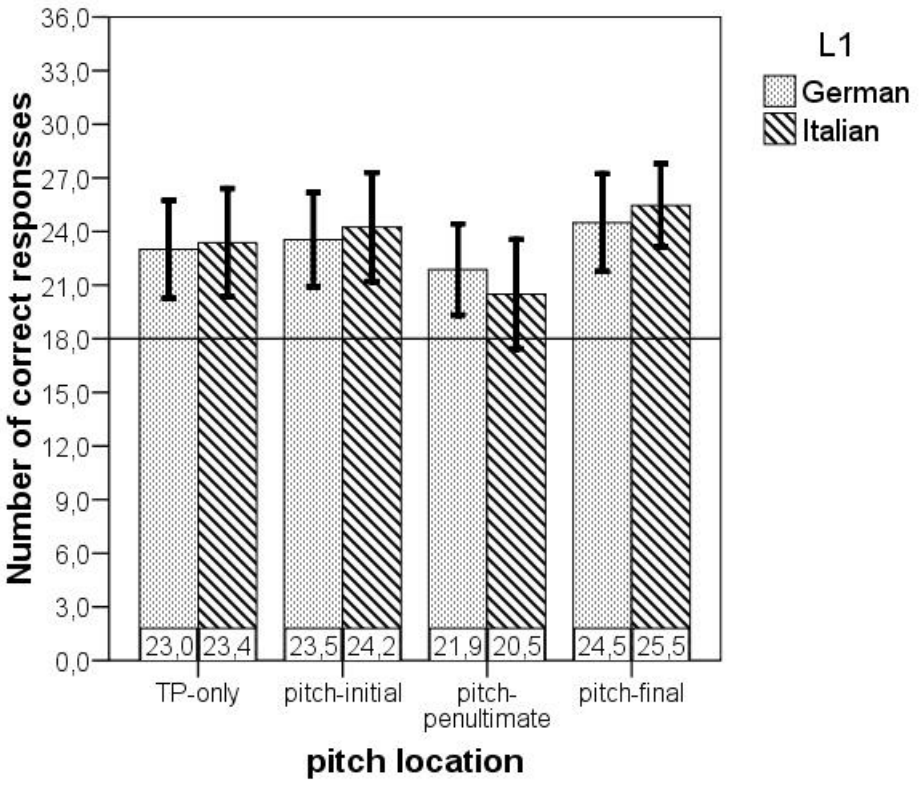

Figure 4. Mean number (±2 standard errors) of correct answers for each *pitch location* and native language. The horizontal line stands for the chance level (18 correct answers).

Discussion

Segmentation by German participants is successful in all conditions, but in none of the conditions it differs from that in the TP-only condition. Segmentation by Italian listeners is disrupted when the F0 peak is on the penultimate syllable of trisyllabic words. In the other conditions segmentation is successful, but is not different from the TP-only condition. We conclude that neither German nor Italians could benefit from the pitch cues for the segmentation of a novel language. This is not surprising because pitch cues, in the absence of statistical regularities, cannot be used for segmentation (Toro, Rodrigez-Fornels, & Sebastian-Galles, 2007).

The failure to segment trisyllabic words with pitch cues on the penultimate syllable was also confirmed for English and Spanish listeners (Toro, Sebastian-Galles, & Mattys, 2009). Probably, pitch cues on the word-initial and word-final syllables attracts listeners' attention to the statistical words that are independently extracted on the basis of TPs. Local pitch fluctuations in the middle of the word divert listeners' attention from the edges. Syllables at the edges of the extracted words are more readily remembered. Possibly, words are recognized during the test phase by their edge syllables, and when processing resources are diverted from the edges, listeners find it more difficult to remember the word they have extracted automatically on the basis of TPs. This suggestion is in agreement with a body of results showing that words are segmented more easily at the edges than in the middle of

utterances (Shukla et al., 2007), and that repetitions are detected and remembered better at word edges than in the middle of words (Endress, Scholl, & Mehler, 2005). Paying attention to the edges of discrete constituents appears to be a universal feature of perception: thus placing the pitch cue on the middle syllable diverts listeners' attention from the edges and makes the segmentation task harder (Endress & Hauser, 2010). F0 peaks on the penultimate syllable in 3-syllabic words do not disrupt TPs computation, but hinder remembering the words heard during familiarization, and hampers later retrieval from memory and recognition of the words during the test phase.

The reason why segmentation by German listeners was not disrupted by placing the pitch in the middle of the word is not quite clear. We do see that Germans perform worse in the pitch-medial condition compared to the TP-only condition, though the drop in performance is not significant. Toro et al. (2009) had the same results with French listeners who performed above chance regardless of which syllable in 3-syllabic words was marked by pitch. Performance was slightly poorer – but not significantly so - when the pitch cues were placed on penultimate syllables in trisyllabic words. They concluded that French listeners find it hard to discriminate between two items that differ only in the location of the stressed syllable – the so-called stress deafness effect (Dupoux, Pallier, Sebastian-Galles, & Mehler, 1997). However, German speakers have never been reported to be stress deaf. We might tentatively hypothesize that segmentation can be successful even when the local F0 fluctuations are not aligned with the word edges if participants have a high overall level of attention.

Although the results of Toro et al. (2009) and of our study are similar, they are in conflict with Tyler and Cutler (2009) who found language-specific differences in processing pitch cues for segmentation purposes by English, Dutch and French listeners. Tyler and Cutler (2009) found that native speakers of English segment better when the initial syllable of the word is marked by pitch, while French segment better when final syllable is marked by pitch. However, Tyler and Cutler (2009) included 2, 3 and 4-syllabic words and mixed them in their artificial language stream, and they did not introduce pitch cues on the non-boundary syllables in longer words. Pitch was always associated with only one of the edge syllables of the statistical words. In addition, adding 2-syllabic words to the vocabulary triggers other perceptual mechanisms that are not activated with longer words, for example, the iambic-trochaic law (Hay & Diehl, 2007). Moreover, in a 2-syllabic word, a pitch cue on any of the syllables provides a dichotomy between a cued (marked) and an uncued (unmarked) syllable. Consequently, in a 2-syllabic word, one boundary is signalled by marking the syllable, and the other boundary is signalled by the fact that the syllable is not marked. Therefore, adding 2-syllabic words possibly makes the segmentation task easier. Therefore the results in our study and in the study by Tyler and Cutler (2009) are not directly comparable, and this makes estimating the degree of disagreement between these two studies not very straightforward.

General Discussion

Our results confirm that TPs between adjacent syllables provide sufficient information for segmentation. Both German and Italian listeners in our experiments performed reliably above chance in the TP-only conditions, and there was no difference between the speakers of these two languages. Adding prosodic cues may facilitate (improve performance beyond that in the TP-only condition) or disrupt segmentation. Prosodic cues in a novel language are indeed filtered through the phonology of the native languages of the participants. Sometimes, performance is not affected by adding prosodic cues. Table 2 provides the overview of the results summarized across the experiments with the German and Italian participants reported here and in Ordin and Nespor (2013). The effect of prosodic cues on segmentation depends on: 1) whether a prosodic cue is processed at the level of the word or at the level of the phrase; 2) whether the cues in the native and the novel language of the speaker match in location and phonetic implementation. We have found that word-level prosody does not facilitate segmentation, but it can disrupt it, when the prosodic pattern on the word level in the native language of the listener does not match the corresponding pattern in the novel language. If the prosodic cue is processed as a marker of lexical stress, then segmentation will be disrupted if the location of stress in a novel language does not match the unmarked location of stress in the native language of the participants. If the stress locations in the native and the target languages match, then segmentation is successful, but is not facilitated. Phrase-level prosody, e.g., PP-final lengthening, can facilitate segmentation, i.e., improve segmentation performance beyond that in the TP-only condition for both groups. Prosodic patterns in the artificial language that cannot be mapped onto any of the existing cues in the native language are not filtered via native phonology. An example of such pattern in intonational languages can be a local pitch raising associated with a single syllable, without simultaneous syllabic lengthening. Therefore, processing of such patterns is not affected by the listener's L1.

Table 2. Summary of results showing the influence of location and the type pf prosodic cue on segmentation performance.

Type of prosodic cue	**Location of a prosodic cue in a trisyllabic word**					
	Initial syllable		**Penultimate syllable**		**Final syllable**	
	Segmentation by Germans	**Segmentation by Italians**	**Segmentation by Germans**	**Segmentation by Italians**	**Segmentation by Germans**	**Segmentation by Italians**
Duration	Segmentation is disrupted	Segmentation is disrupted	Segmentation is disrupted	Segmentation is successful but not facilitated	Segmentation is facilitated	Segmentation is disrupted
Pitch	Segmentation is successful but not facilitated	Segmentation is successful but not facilitated	Segmentation is disrupted	Segmentation is disrupted	Segmentation is successful but not facilitated	Segmentation is successful but not facilitated
Combination	Segmentation is	Segmentation is	Segmentation is	Segmentation is	Segmentation is	Segmentation is

of duration and pitch	successful but not facilitated	successful but not facilitated	disrupted	disrupted	facilitated	facilitated

An interesting question to discuss is when prosody and statistics begin to interact, considering that segmentation is possible using only TPs. When is segmentation facilitated or impeded by the listener? Shukla et al. (2007) suggested that all statistical words are segmented and extracted online, during familiarization, but not all word candidates are retained. Prosody defines the frames for discrete constituents and acts as a filter, allowing the syllabic sequences with high internal TPs that fit the frames, and suppressing the sequences that straddle the frame boundaries. Shukla et al. (2007) used declining F0 contour spanning over ten syllables to signal IP boundaries. 3-syllabic statistical words either straddled the boundaries of intonational IPs, or occurred within the IPs. After the familiarization exposure, adult Italians were asked to listen to pairs of words versus non-words[2]. Participants discriminated contour-internal words from non-words, but failed to discriminate contour-straddling words from non-words. Another group of participants was presented with the same pairs of words versus non-words visually. These participants successfully differentiated both contour-straddling and contour-internal words from non-words. The authors suggest that during test, upon reading the test items, participants generated representations of the visually presented words that were equivalent to the stored representations of statistically defined words segmented during familiarization. As the visual input contains no acoustic information, a representation as a string of syllables is dominant and the effect of prosody, which prevents recognizing contour-straddling words, is overcome. Therefore, Shukla et al. (2007) concluded that all statistical words are segmented and extracted during familiarization, but the words that do not fit the prosodic frames are then suppressed and not recognized during the test phase. In a series of experiments Langus et al. (2012) included final lengthening to mark the right edge of PPs, and intonation contours with pitch declination to signal IP boundaries. Adding prosodic cues at the PP-level allowed Langus et al. (2012) to further extend Shukla et al.'s (2007) model and to suggest that prosody filters statistical words both at the level of the IP and at the level of the PP.

Toro et al. (2011) also found that TPs are calculated independently from prosody and statistical words are extracted on-line by Catalan participants. Middle vowels in Catalan are always stressed, and therefore the presence of two middle vowels in the same word is not allowed. Toro et al. (2011) found that Catalan listeners successfully segmented an artificial language and recognized statistically defined words with two middle vowels in a novel language, if both words and non-words were presented visually for discrimination after familiarization. However, when Catalan listeners had to discriminate the same pairs of words and non-words auditorily, they failed to

[2] Non-words are combinations of syllables from the speech stream that never occurred sequentially during familiarization.

recognize the words with two middle vowels. Toro et al. (2011) argued that TP-computations were performed on-line irrespective of language-specific phonological constraints on the vocalic structure of words, and the word-level constraints filter unsuitable candidates later.

Considering that language-specific constraints at the word level are imposed on the already segmented statistical words, we can further extend the Shukla et al. (2007) and Langus et al. (2012) model from the IP and PP to the word level. Prosody builds the frames for the phonological words, phonological phrases, and intonational phrases, and suppresses those candidates that do not fit the prosodic frames. Prosodic cues are processed at different levels of the prosodic hierarchy. Germans, for example, interpreted lengthening as a PP-boundary marker and constructed the frames for the PPs with the longer slots for the right-edge syllable. Italians can use lengthening to construct frames both for PWs and PPs, depending on whether the increase in duration is processed at the word- or the PP level. The hierarchical level at which lengthening is processed by Italians depends on the availability of other prosodic cues, access to statistical information, and possibly other factors. Further cross-linguistic research is needed to fully understand the mapping between the prosodic hierarchy of natural languages and the nature of the segmented speech constituents in artificial languages.

We aimed at investigating how native speakers of phonologically different languages segment discrete constituents from continuous speech in a novel language. We have shown that processing novel prosodic patterns depends on the phonology of the listeners' native language. However, with proficiency growth, segmentation strategies might adapt to the language of exposure. A more ecologically plausible experimental design is needed to investigate how segmentation strategies change when mastery in the target language increases. This line of research will be pursued in further studies.

References

Alfano, I., Llisterri, J., & Savy, R. (2007). The perception of Italian and Spanish lexical stress: A first cross-linguistic study. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1793-1796). Saarbrücken, Germany, 6-10 August.

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

Bissiri, M.P., Pfitzinger, H.R., & Tillman, H.G. (2008). Lexical stress perception of German morphologically complex words by Italian speakers. *Speech Prosody 2008* (pp. 639-642). Campinas.

Bissiri, M.P. (2008). *Experentalle Untersuchungen zum computer-gestützten Erlernen der Prosodie des Deutschen für italienischsprachige Lernende.* PhD Thesis, Lüdwig-Maximillian Universität, München.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on Statistical Computations. *Psychological Science*, 16(6), 451-9.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological Phrase Boundaries Constrain Lexical Access I. Adult Data. *Journal of Memory and Language*, 51(4), 523-547.

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.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385-400.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology, 24*, 381–410.

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.

Delattre, P. (1965). Comparing the Phonetic Features of English, French, German and Spanish. Heidelberg: J. Groos Verlag.

Dell, F. (1984). L'accentuation dans les phrases en français. In F. Dell, J.-P. Vergnaud, & D. Hirst (Eds.)., *Les representations en phonologie* (pp. 65-112). Paris: Hermann.

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

Dupoux, E., Pallier, Ch., Sebastian-Galles, N., & Mehler, J. (1997). Destressing 'deafness' in French. *Journal of Memory and Language 36*, 406-421.

Dupoux, E., Peperkamp,S., & Sebastián-Gallés, N. (2001). A robust method to study stress ''deafness''. *Journal of the Acoustical Society of America, 110*, 1606–1618.

Dutoit, T., Pagel, N., Pierret, F., Bataille, O., & van der Vrecken, O. (1996). The MBROLA Project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. *Proceedings of the Fourth International Conference on Spoken Language Processing*, 1393–1396.

Endress, A., & Hauser, M. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology, 61*, 177-199.

Endress, A.D., Scholl, B.J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General* 134 (3), 406-419.

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: How conscious? *Frontiers in Psychology 229*, 1–12.

Gebhart, A., Aslin, R., & Newport, E. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science 33*, 1087–1116.

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.

Gussenhoven, C. (2003). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

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

Hayes, J.R. & Clark, H.H. (1970). Experiments in the segmentation of an artificial speech analog. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

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

Jessen, M. (1999). German. In H. van der Hulst (Ed.), *Word Prosodic Systems in the Languages of Europe* (pp. 515-545). Berlin: Mouton de Gruyter.

Johnson, E. K., & Seidl, A. (2008). A cross-linguistic perspective on the detection of juncture in speech. *Infancy 13*(5), 440–455.

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech 51*(4), 343-359.

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.

Kjelgaard, M., & Speer, S. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language 40*(2), 153-194.

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.

Ladd, R. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press.

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.

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

Millotte, S., Rene, A., Wales, R., & Christophe, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 874–885.

Mixdorff, H., Hayashi, R., & Ushiyama, S. (2014). Identification of Word Boundaries and Accented Syllables in German by German and Non-German Listeners. In A. Leemann, Kolly, M.-J., Schmid, S., & Dellwo, V. (Eds.) Trends in Phonetics & Phonology in German-speaking Europe. (pp. 153-167). Zurich: Peter Lang.

Murty, L., Otake, T., & Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. Mora Rhythm. *Language and Speech 50*(1), 77-99.

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

Onishi, K., Chambers, K., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*, B13-B23.

Ordin, M., Nespor, M. (2013). Transition Probabilities and Different Levels of Prominence in Segmentation. *Language Learning 6(4)*, 800-834.

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

Ortega-Llebaria, M., Vanrell, M. M., & Prieto, P. (2010). Catalan speakers' perception of word stress in unaccented contexts. *Journal of the Acoustical Society of America, 127*(1), 462–471.

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

Peperkamp, S., Dupoux, E., & Sebastián-Gallés, N. (1999). Perception of stress by French, Spanish, and bilingual subjects. In *Proceedings of EUROSPEECH'99* (pp. 2683-2686). Budapest, Hungary, 5-9 September.

Peters, B. (2005). Weiterführende Untersuchungen zu prosodischen Grenzen in deutscher Spontansprache. In K. Kohler, F. Kleber & B. Peters (Eds.). Prosodic Structures in German Spontaneous Speech (AIPUK 35a) (pp. 203–345). Kiel: IPDS.

Peters, B., Kohler, K. J., & Wesener, T. (2005). Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In K. Kohler, F. Kleber & B. Peters (Eds.). Prosodic Structures in German Spontaneous Speech (AIPUK 35a) (pp. 143–184). Kiel: IPDS.

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

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

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.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101–105.

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.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Statistical and stress cues in infant word segmentation. *Developmental Psychology, 39,* 706–716.

Toro, J. M., Rodriguez-Fornells A., & Sebastian-Galles N. (2007). Stress placement and word segmentation by Spanish speakers. *Psicologica 28,* 167-176.

Toro, H. M., Sebastián-Gallés, N., & Mattys, S. (2009). The role of perceptual salience during the segmentation of connected speech. *European Journal of Cognitive Psychology 21,* 786–800.

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.

Tremblay, A., Coughlin, C., Bahler, C., & Gaillard, S. (2012). Differential contribution of prosodic cues in the native and non-native segmentation of French speech. *Laboratory Phonology 3*(2), 358-423.

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

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38,* 133–149.

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

Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Language and Cognitive Processes 10*(5), 457–486.

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology*, 3(580).

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