EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2009, 21 (5), 786–800

Psychology Press
Taylor & Francis Group

The role of perceptual salience during the segmentation of connected speech

Juan M. Toro

International School for Advanced Studies, Trieste, Italy, and GRNC, Parc Científic de Barcelona & Departament de Psicologia Básica, Universitat de Barcelona, Barcelona, Spain

Núria Sebastián-Gallés

GRNC, Parc Científic de Barcelona & Departament de Psicologia Básica, Universitat de Barcelona, Barcelona, Spain

Sven L. Mattys

Department of Experimental Psychology, University of Bristol, Bristol, UK

Statistical learning allows listeners to track transitional probabilities among syllable sequences and use these probabilities for subsequent speech segmentation. Recent studies have shown that other sources of information, such as rhythmic cues, can modulate the dependencies extracted via statistical computation. In this study, we explored how syllables made salient by a pitch rise affect the segmentation of trisyllabic words from an artificial speech stream by native speakers of three different languages (Spanish, English, and French). Results showed that, whereas performance of French participants did not significantly vary across stress positions (likely due to language-specific rhythmic characteristics), the segmentation performance of Spanish and English listeners was unaltered when syllables in word-initial and word-final positions were salient, but it dropped to chance level when salience was on the medial syllable. We argue that pitch rise in word-medial syllables draws attentional resources away from word boundaries, thus decreasing segmentation effectiveness.

Correspondence should be addressed to Juan M. Toro, Departament de Psicologia Bàsica, Universitat de Barcelona, Passeig de la Vall d'Hebrón, 171, CP 08035, Barcelona, Spain. E-mail: jmtoros@ub.edu

This research was supported by grants from the Human Frontier Science Program (HFSP) RGP68/2002, the OMLL-ESF BFF2002-10379-E, the MEC 2007-60751, and Consolider Ingenio 2010 CSD2007-012. We would like to thank the invaluable collaboration of Elena Azañon, Jinnam Choi, Antonia Hamilton, James Melhorn, Mikel Santesteban, and Scott Sinnett for their help in preparing and running the experiments, and thank the reviewers for helpful insights.

http://www.psypress.com/ecp DOI: 10.1080/09541440802405584

Keywords: Statistical learning; Attention; Speech segmentation; Perceptual salience.

The finding that human infants can track transitional probabilities in speech (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996a) has provided the impetus for a considerable amount of research because it suggests that at least some aspects of language learning might arise from relatively simple statistical computations. Results suggesting the involvement of statistical learning in the acquisition of linguistic structures include phonetic variations (Maye, Werker, & Gerken, 2002), phonotactics (Onishi, Chambers, & Fisher, 2002), nonadjacent dependencies (Newport & Aslin, 2004; Peña, Bonatti, Nespor, & Mehler, 2002), and phrase structures (Saffran, 2001). Statistical learning mechanisms are also involved in the segmentation of speech, extracting lexical candidates that are more easily used to label new objects (Mirman, Magnuson, Graf Estes, & Dixon, 2008), and that elicit electrophysiological signatures traditionally associated with lexical processing (as the N400; see Sanders, Newport, & Neville, 2002).

However, recent research has shown that statistical computations in a naturalistic setting also benefit from being constrained by other cues (Yang, 2004), mainly because the combinatorial explosion that such computations promote makes the number of possible outcomes for a given data set virtually infinite. That is, in order to provide realistic solutions to the problem of structure detection, additional information that bias statistical learning towards a plausible solution should be taken into account. To explore this limit on the effectiveness of statistical learning, researchers have tried to identify the way in which statistical learning interacts with, and is constrained by, other sources of information present in the speech signal.

Several studies with infants have tackled this issue. Johnson and Jusczyk (2001) provided evidence for an interaction between statistical and segmental cues, showing that English-learning 8-month-olds relied more heavily on stress than statistical cues when stress fell on the initial syllable of words. Replicating Johnson and Jusczyk's results, Thiessen and Saffran (2003) demonstrated that 9-month-old English learners relied more heavily on trochaic stress patterns than on statistical cues (see also Mattys, Jusczyk, Luce, & Morgan, 1999). However, Thiessen and Saffran also showed that this preference changed with age, as 7-month-old infants relied more heavily on statistical cues, pointing towards a change in relative cue weights in the course of early language development.

Interestingly, few studies have explored the interaction between statistical and stress-related cues in adulthood. Recently, however, prosodic markers such as intonational phrases and coarticulatory cues have been shown to affect the outcome of speech segmentation. For example, Shukla, Nespor,

and Mehler (2007) investigated the interaction between statistical learning and prosody by imposing intonational variations mimicking the prosody of intonational phrases over an artificial speech stream. The authors reported that statistically coherent sequences straddling prosodic boundaries were not remembered as well as those within prosodic constituents, which pointed to a form of "prosodic filtering" of statistical computations over syllables. Similarly, Fernandes, Ventura, and Kolinsky (2007) have shown that listeners rely more on coarticulatory than statistical cues under optimal listening conditions. Under degraded listening conditions, the predominance of coarticulation over statistical information tended to disappear. These studies point to a dynamic, interactive, and flexible use of segmentation cues. In fact, building a theoretical framework on how these different cues are integrated during adult speech segmentation, Mattys, White, and Melhorn (2005) highlighted a hierarchy of cue weights, from lexical to segmental to suprasegmental, with reliability modulated by the interpretive conditions (sentential context, signal degradation, etc.).

Saffran, Newport, and Aslin (1996b) directly addressed the potential interaction between statistical and stress-related cues in adults. Using a continuous speech stream composed of six trisyllabic nonsense words, the authors showed that participants could use the statistical "dips" between the words to extract them from the stream. Participants' performance was compared across three conditions, one in which only statistical cues marked word boundaries, and two in which a prosodic cue (vowel lengthening) was added to some words in the stream. Lengthening occurred in either the initial or the final syllable of three out of the six words of the stream. Results showed that performance was not different between the no-lengthening and the initial-lengthening conditions. In contrast, performance improved in the final-lengthening condition. The authors attributed this effect to the fact that, in natural utterances, word-final syllables tend to be lengthened, a process thought to be language-general (Hoequist, 1983).

There are, however, several aspects of those studies that leave the door open for further investigations of the effects that prosodic segmentation cues have on the extraction of statistical regularities. For example, in the experiments by Saffran et al. (1996b), the prosodic manipulation of words forming the stream was not systematic (only three out of six words composing the stream were lengthened), and the experimenters did not include a condition in which lengthening did not align with one of the word boundaries (only the first or the third syllable of words was lengthened, but never the second). More importantly, no study has fully explored how the participants' linguistic background may influence such effects.

In the present study, we aimed to explore how prosody influences the computation of statistical regularities using a cue that is not present in all natural languages as a marker to word boundaries, namely, pitch rise. Here,

pitch rise was applied to the initial, medial, or final syllable of trisyllabic words in a speech stream and compared to a flat-pitch condition. To account for a possible influence of linguistic background, we compared the performance on a speech segmentation task across speakers of three different languages: Spanish, English, and French.

Both English and Spanish have contrastive lexical stress. However the predominant position of stress in the two languages differs: It is mostly word-initial in English, (e.g., Cutler & Carter, 1987) and mostly penultimate in Spanish, (e.g., Navarro, 1966). French does not exhibit contrastive lexical stress (Banel & Bacri, 1994; Fletcher, 1991).

As an increase in pitch has been shown to be a strong correlate of lexical stress in many languages, including English and Spanish (e.g., Hirst & di Cristo, 1998; Llisterri, Machuca, de la Mota, Riera, & Rios, 2002; Spitzer, Liss, & Mattys, 2007), and lexical stress is thought to be a useful cue for word boundaries in at least some languages (Cutler, Dahan, & van Donselaar, 1997), the present experiment aims to reveal whether the native language of a listener can modulate the effect of pitch rise on the segmentation of an artificial language. Linguistic background has indeed been shown to influence speech segmentation under naturalistic conditions (Vroomen, Tuomainen, & de Gelder, 1998). There is also evidence that listeners perceive acoustic modifications in speech by reference to the prosodic properties of their native language (e.g., Cutler, Mehler, Norris, & Segui, 1986; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001).

Therefore, different hypotheses can be put forward regarding the outcome of the present experiment, depending on how pitch rise interacts with the participants' linguistic background. If pitch rise is a reliable correlate of lexical stress, and lexical stress is useful for segmentation, an interaction between pitch-rise position and language should be observed. That is, when aligned with the dominant stress position of each language (i.e., initial in English and medial in Spanish), pitch rise should improve word segmentation. Alternatively, our pitch manipulation might not be an adequate correlate of stress in English, Spanish, and French, in which case we would not expect language-specific effects. A language-general pattern of results might emerge instead (cf. Hay & Diehl, 2007). As pitch rise might help signalling the edges of words in the stream, performance should improve in the conditions where the initial or final syllable is highlighted, independently of language, provided that the effects of statistical and pitch cues are additive (as suggested by Fernandes et al., 2007). If the effects are not additive, pitch rise may not have any visible effect on word segmentation for neither of the language groups.

METHOD

Participants

The participants in the Spanish group were 80 native speakers of Spanish. They were recruited from the University of Barcelona. Participants in the English group were 68 native English speakers recruited from University College London, Oxford University, and the University of Bristol. Participants in the French group were 68 native French speakers recruited from the University of Paris and University of Marseille. Participants within each language group were randomly assigned to one of four pitch conditions and were either paid or given course credits for their participation in the study. None of the participants reported hearing deficits or problems of language acquisition.

Stimuli

Four artificial languages, identical in structure but differing in the salient syllable of their component words, were created. The languages consisted of four trisyllabic (CVCVCV) nonsense words: *tupiro*, *golabu*, *bidaku*, *padoti*—the same words as in the study by Saffran, Aslin, and Newport (1996). Following Saffran et al.'s procedure, the four words were concatenated in a semirandom order, such that there were no adjacent repetitions, and ensuring that each word was followed by each of the other three an equal number of times. In these streams, the average transitional probability between the syllables of a word was 1.00, whereas it was 0.33 between the syllables spanning a word boundary. Each word appeared a total of 605 times. The resulting stream was synthesised using the text-to-speech MBROLA software (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996) with a Spanish male (es2) diphone database at 16 kHz. Each syllable was 232 ms long. A 5 s rise and fall intonation contour was introduced at the beginning and end of the stream. In the flat condition, the fundamental frequency (*F*0) of all syllables was set to 200 Hz, and there were no acoustic markers between the words of the stream, so that the only cues available for segmentation were statistical troughs at word boundaries.

To create the three other versions of the stream, we increased the *F*0 of word-initial, word-medial, and word-final syllables, respectively. The rise consisted of a 20 Hz steady increase over the entire syllable of the word, resulting in words like *TUpiro* (word-initial), *tuPIro* (word-medial), and *tupiRO* (word-final), with capital letters indicating *F*0 rises. The *F*0 of all other syllables was kept at 200 Hz. For the test phase in all four conditions, eight items were created, four "words" and four "part-words"'. The test words consisted of the four original trisyllabic words that formed the

streams, whereas the part-words were made by concatenating the third syllable of a word and the first two syllables of another word, yielding *tibida*, *kupado*, *rogola*, *butupi*. Consistent with prior similar studies, all stimuli in the test phase were presented in their neutral format, i.e., with flat *F*0.

Procedure

Participants were told that they would hear a nonsense language, and that their task was simply to pay attention to it. They were not told that the language was composed of trisyllabic units, nor were they explicitly asked to listen for structural regularities. After 7 minutes of listening to the uninterrupted language, participants were given a two-alternative forced-choice test. They were presented with eight pairs of spoken trisyllabic stimuli, each containing a word and a part-word. The stimuli of each pair were played 500 ms apart from each other. For each pair, participants had to indicate which of the two stimuli they thought was a word of the language they had just heard by pressing either "1" or "2" on a response box. The next test pair was played when the participant pressed a response key or when 5 s had elapsed after the presentation of the test pair. The eight test pairs originated from four word/part-word pairings, in which the order of presentation (word/part-word vs. part-word/word) within the pair was counterbalanced (4×2) within each participant.

RESULTS

Percentage of correct responses in the forced-choice test was calculated and averaged across participants for each condition (see Figure 1). An analysis of variance with language (Spanish, English, French) and condition (flat, initial, medial, final) as between-subjects variables revealed no significant differences between languages, $F(2, 204) < 1$, $MSE = 338.75$ but a condition effect was observed, $F(3, 204) = 18.98$, $p < .001$. Importantly, the interaction between language and conditions was significant, $F(6, 204) = 3.28$, $p < .005$. To explore the source of this interaction, the results were analysed separately for each language group. For Spanish participants, a significant condition effect was observed, $F(3, 76) = 13.97$, $p < .001$, $MSE = 380.79$. Pairwise comparisons showed that the flat, initial, and final conditions did not significantly differ from each other (flat vs. initial, $p = .80$; flat vs. final, $p = .17$; initial vs. final, $p = .11$), but all three showed greater accuracy than the medial condition ($p < .001$, for all comparisons). Further analyses indicated that the percentage of correct responses in all of the nonmedial conditions departed significantly from chance (50%; see Table 1 for values of each condition): flat, $t(19) = 5.00$, $p < .001$; initial, $t(19) = 4.53$, $p < .001$;

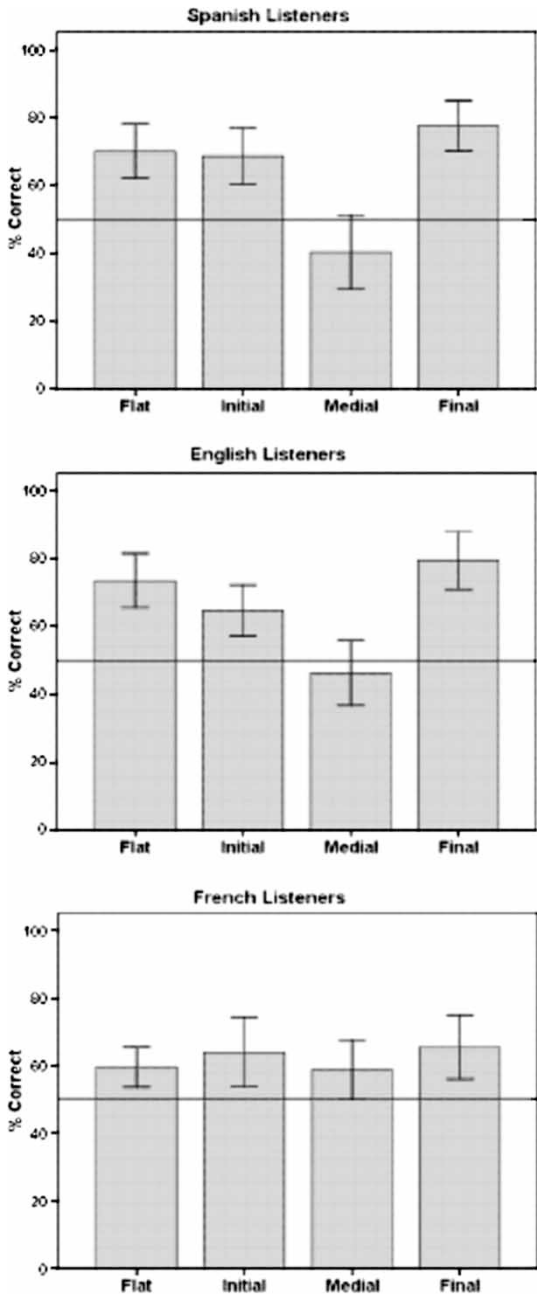

Figure 1. Group means and standard error bars for Spanish, English, and French participants.

TABLE 1
Average percentage of correct responses and standard deviation (in parentheses) for Spanish, English, and French participants in the four experimental conditions

	Conditions			
	Flat	*Initial*	*Medial*	*Final*
Spanish	70.0 (17.8)	68.6 (18.3)	40.3 (24.4)	77.5 (16.5)
English	73.5 (16.5)	64.7 (15.4)	46.3 (19.6)	79.4 (17.6)
French	59.6 (12.1)	64.0 (21.1)	58.8 (18.1)	65.4 (19.5)

final, $t(19) = 7.44$, $p < .001$; medial, $t(19) = -1.76$, $p = .09$. These results suggest that Spanish speakers were able to segment words from the speech stream using statistical cues. Interestingly, they performed less well in the medial condition than in the flat, initial, and final conditions.

For the English participants, we found an effect of condition, $F(3, 64) = 11.74$, $p < .001$, $MSE = 301.87$. Similar to the results in the Spanish group, pairwise comparisons showed that flat, initial, and final conditions did not differ from each other (flat vs. initial, $p = .54$; flat vs. final, $p = .81$; initial vs. final, $p = .12$), but they all differed from the medial condition ($p < .001$, for all comparisons). The percentage of correct responses in the flat condition was significantly higher than chance, $t(16) = 5.89$, $p < .001$. So were the initial, $t(16) = 3.92$, $p < .001$, and the final conditions, $t(16) = 6.87$, $p < .001$. Like Spanish speakers, English speakers performed at chance level in the medial condition, $t(16) = -0.77$, $p = .45$. An analysis of variance combining the results of the Spanish and English participants revealed no difference between languages, $F(1, 140) < 1$, $MSE = 344.71$, a significant condition effect, $F(3, 140) = 24.83$, $p < .001$, and no interaction between language and condition, $F(3, 140) < 1$. Thus, both Spanish and English participants, despite performing equally well in the flat, initial, and final conditions, had more difficulty segmenting the stream when *F*0 was increased in word-medial position.

A different pattern of results emerged with the French participants. In particular, there was no condition effect, $F(3, 64) < 1$, $MSE = 325.71$. However, as with the Spanish and English participants, performance was significantly better than chance in the flat, initial, and final conditions: flat, $t(16) = 3.25$, $p < .005$; initial, $t(16) = 2.72$, $p < .05$; final, $t(16) = 3.26$, $p < .005$. Difference from chance in the medial condition was only marginal, $t(16) = 2.01$, $p = .06$. Nevertheless, the relatively low level performance in the flat, initial, and final conditions may have made it more difficult to observe a dip in performance in the medial condition. Indeed, although the average performance of the French participants did not significantly differ from that of the English and Spanish participants (French = 61.9%, English = 66.0%,

Spanish = 64.1%), performance restricted to the flat condition was significantly lower in the French group than in the other two groups (French vs. English, $p < .05$; French vs. Spanish, $p < .05$). The factors that could have lead to this numerically low performance are considered in the Discussion. An analysis of variance between Spanish and French participants yielded no significant language effect, $F(1, 140) < 1$, $MSE = 355.61$, a significant condition effect, $F(3, 140) = 9.14$, $p < .001$, and, critically, a significant interaction between language and condition, $F(3, 140) = 5.15$, $p < .005$. Likewise, a comparison between English and French participants showed no language effect, $F(3, 128) = 1.77$, $p = .18$, $MSE = 313.79$, a significant condition effect, $F(3, 128) = 7.52$, $p < .001$, and an interaction between language and condition, $F(3, 128) = 4.35$, $p < .01$. Thus, although the pattern of results was highly similar for the Spanish and English listeners, with both groups showing worse performance in the medial condition, the French group did not show significant differences across the *F*0 conditions.

DISCUSSION

The discovery that listeners are able to extract information about the linguistic structure of a speech input via statistical mechanisms has attracted much attention among psycholinguists. At the same time, efforts have been made to characterise how these mechanisms may interact with other sources of information, such as prosodic cues, e.g., intonational phrases (Shukla et al., 2007), coarticulation (Fernandes et al., 2007), and vowel lengthening (Saffran, Newport, & Aslin, 1996). The present study shows that, when the initial or final syllable of trisyllabic words is made salient by increasing its *F*0, segmentation performance based on statistical regularities is unchanged compared to a stream with flat *F*0. Surprisingly, a pitch rise in the medial position makes performance drop significantly. This pattern was present in both Spanish and English listeners, even though the dominant stress placement in their language is word-initial and penultimate, respectively.

Results differed for the French participants. French participants performed rather homogeneously across the four experimental conditions and the results in the medial condition only slightly departed from chance level. Two factors might account for this cross-linguistic difference. First, in the present study, the acoustic modification was restricted to an increase in *F*0. *F*0 changes are an important component of lexical stress in English and Spanish (Hirst & di Cristo, 1998; Llisterri et al., 2002), but they are less critical in French and are, in fact, difficult to perceive by French listeners (a phenomenon known as "stress deafness"; see Dupoux, Christophe, Sebastián-Gallés, & Mehler 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008). Lexical

rhythmic patterns in French are best characterised by word-final lengthening (Fletcher, 1991). Thus, the absence of an effect of *F*0 in the French group could be due to the relative unimportance of this parameter for lexical rhythm in that language. Second, the words of this study consisted of CV syllables. CV syllables are permissible content words in French (e.g., [ʃa], [me], [mo], [ri]) but rarely in Spanish or English. Thus, French listeners could have engaged in a lexical search for possible monosyllabic CV words, which would have interfered with the processing of distributional properties of the stream, including the recurring position of the *F*0 rise. The fact that French participants' performance in the flat condition was lower than the performance of English and Spanish participants supports the idea that the former may have had difficulties extracting statistical regularities from the stream. This, and the difficulty to perceive stress changes in French listeners, could have overridden any effect of *F*0 rises on speech segmentation.

One of the predictions originally made in this study was that the acoustic manipulation we used (*F*0 rise) should interact with the participants' native language. Indeed, several experiments (e.g., Cutler et al., 1986; Soto-Faraco et al., 2001; Vroomen et al., 1998) have shown that linguistic background strongly modulates speech segmentation and lexical processing. Results from the English and Spanish groups in the present experiment did not support this prediction. A possible implication of this finding is that the assumption that stress-based segmentation mirrors a language's typical stress placement—a recurrent idea in the literature—might derive from the fact that most studies were conducted with languages in which stress occurs at the edges of words (English, Finnish, Dutch, and French). Testing a language that stresses penultimate syllables, like Spanish, might have unexpectedly allowed us to assess how stress and statistical regularities interact during speech segmentation, independent of language-specific constraints. As we will argue later, a lack of interaction between *F*0 rise and the participants' native language may be due to a strong modulation of segmentation processes by attentional manipulations.

An important question remains as to why a word-medial pitch rise caused the performance to drop in the English and Spanish groups. One possibility is that participants learned only the stressed syllables during familiarisation, together with some positional information (i.e., whether the stressed syllable was located in the initial, middle or final position of the word). As a result, participants' performance during the test may have been based on whether the learned syllable was located at the boundaries of test items or not. Because of the structure of part-words, this could have lead to some confusion in the middle condition, where the stressed syllable was second in the words, but located at a boundary in the part-words. Nevertheless, the fact that there were no significant differences between the initial and final

conditions that include part-words with target syllables in different positions (initial and medial), forces us to consider this possibility with some caution. Alternatively, the *F*0 manipulation may have filtered the syllables over which statistics were performed, in the sense that it may have allowed computations only over given, stressed syllables. Shukla et al. (2007) showed a filtering effect of prosody (in their case, intonational contours) over statistical computations. Specifically, this effect depended on the position of a target word with respect to the boundaries of an intonational contour. Participants could not recall words that "straddled" the intonational contour, whereas they did not have any problems distinguishing these words from foils when words were located within the contour. However, it is difficult to see how our manipulation may have had parallel effects on how statistics are computed, since the *F*0 rise in the present study marked a single element in the sequence, and not a series of elements that are internal or straddle an entire prosodic contour. It is more likely that the 20 Hz rise highlighted certain elements of the stream, and therefore directed attention towards those elements, as explained later. Another possibility is that stress cues may not interact at all with statistical regularities, and may be the only source of information the participants are using for segmenting out the words. Nevertheless, previous experiments have demonstrated that when no statistical regularities are present (i.e., when syllables composing the stream are scrambled), acoustic cues alone do not lead to successful segmentation of the stream (Toro, Rodriguez-Fornels, & Sebastián-Gallés, 2007).

Yet, one could expect that when syllables at the boundaries of words are made salient by an *F*0 increase, an *improvement* in performance should be observed. Indeed, Fernandes et al. (2007) showed that the effects of statistics and coarticulation are additive. On the other hand, Saffran et al. (1996b) did not find an increase in performance when the initial syllable of words in their stream was lengthened. Therefore, it might be the case that our participants (and the participants in the lengthening condition in the Saffran et al., 1996b, experiment) were not processing the acoustic manipulation as a reliable correlate of lexical stress. But it is still an open issue as to whether the use of more naturalistic acoustic manipulations would result in an improvement in the segmentation performance (see Tyler, Perruchet, & Cutler, 2006).

The fact that performance tended to drop in the medial condition could be related to how perceptual saliency helps to allocate processing resources to the stream. In previous studies, saliency has been shown to be an important factor in the modulation of perceptual processes. Both in vision (Theeuwes, 1994) and in memory (Wallace, 1965), items that stand out in a perceptual dimension are processed more readily and remembered better. Saliency is also a strong grouping cue in the auditory domain. Creel, Newport, and Aslin (2004) showed that tracking nonadjacent relations among tones in a continuous stream was greatly enhanced by the perceptual

similarity of the tones. More recently, Hay and Diehl (2007) found that, regardless of their linguistic experience, participants tend to form iambic groupings (i.e., groups that begin with a weak element followed by a strong element) for stimuli varying in duration, whereas they form trochaic groupings (groups that begin with a strong element followed by a weak element) for stimuli varying in intensity, thus adhering to what is known as the iambic/trochaic law. Similarly, a series of recent studies suggest that the stimuli at edges of perceptual groups are more readily processed than the stimuli in their middle. For example, results with infants show that words at the edges of utterances are more readily segmented than words in their middle (Seidl & Johnson, 2006). Studying word learning in adults, Creel, Tanenhaus, and Aslin (2006) demonstrated that the position of a syllable marked for lexical stress influences word recognition, as English-speaking participants had more difficulty processing words with medial than initial stress. Endress, Scholl, and Mehler (2005), showed that the extraction of simple algebraic rules is hindered by positional information when the relevant repetitions are located in the middle, but not at the edges of words. Thus, elements in medial position tend to be processed less effectively than elements at the edges, even if they are highlighted by acoustic cues. In sum, various lines of research on speech perception have shown that perceptually salient items (either by being acoustically different or by being positioned at the edges of sequences) are more readily processed than others.

Furthermore, learning distributional regularities is less effective when attention is diverted away from relevant statistical information. Toro, Sinnett, and Soto-Faraco (2005) demonstrated that when participants' attention is focused on a secondary task, the segmentation of a speech stream using distributional information can be severely disrupted, making performance drop to chance level. A similar finding was also reported with sequences of visual stimuli (Turk-Browne, Jungé, & Scholl, 2005). More recently, it has been claimed that attentional focus on relevant structures is fundamental for the extraction of structural dependencies in tasks similar to the one described here (Pacton & Perruchet, 2008). So there is an important modulation of statistical processes by the allocation of attention. This is not to say that secondary tasks always impair the extraction of statistical regularities from a speech stream. Saffran, Newport, Aslin, Tunick, and Barrueco (1997) demonstrated that participants could segment a speech stream using the statistical dependencies between its elements while drawing, and without the need to explicitly direct attention towards the stream. Thus, statistical learning can be incidental, a result that has been widely replicated. The point here is different. We suggest that when attention is diverted away from relevant regularities by means of a secondary task, or as in the present experiment, by increasing the pitch of syllables located in the middle of words, the task of extracting statistical information is severely disrupted.

It is thus our claim that the saliency of medial syllables in the present experiment may have driven processing resources away from the statistically marked word boundaries, and, hence, may have disrupted the segmentation of the stream. The present set of results build upon other studies that have tackled the interaction between statistical and acoustic cues during language processing. Our results suggest a complex interplay between language-general and language-specific effects. Some constraints over statistical learning may not be necessarily linked to specific aspects of the native language, as was the case with our Spanish and English participants. That is, even though these languages differ across several dimensions (importantly, in their predominant stress pattern), our *F*0 manipulation affected them similarly. In contrast, the performance of French speakers was not affected by our manipulation, most probably because of so-called stress deafness. The present study points to an interaction between saliency (implemented here by a rise in *F*0) and statistical processes that guides processing resources during speech segmentation.

Original manuscript received March 2008
Revised manuscript received July 2008
First published online September 2008

REFERENCES

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321–324.

Banel, M., & Bacri, N. (1994). On metrical patterns and lexical parsing in French. *Speech Communication*, *15*, 115–126.

Creel, S., Newport, E., & Aslin, R. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1119–1130.

Creel, S., Tanenhaus, M., & Aslin, R. (2006). Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 15–32.

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

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

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.

Dupoux, E., Christophe, P., Sebastián-Gallés, N., & Mehler, J. (1997). A distressing 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.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition*, *106*, 682–706.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). *The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes.* Philadelphia: ICSLP.

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

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception and Psychophysics*, *69*, 856–864.

Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics*, *19*, 193–212.

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

Hirst, D., & di Cristo, A. (1998). *Intonation systems*. Cambridge, UK: Cambridge University Press.

Hoequist, C. (1983). Syllable duration in stress-, syllable- and mora-timed language. *Phonetica*, *40*, 203–237.

Johnson, E., & Jusczyk, P. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*, 548–567.

Llisterri, J., Machuca, M., de la Mota, C., Riera, C., & Rios, A. (2002). The role of F0 peaks in the identification of lexical stress in Spanish. In A. Braun & H. Masthoff (Eds.), *Phonetics and its applications: Festschrift for Jens-Peter Köster on the occasion of his 60th birthday* (pp. 350–361). Stuttgart, Germany: Franz Steiner Verlag.

Mattys, S., Jusczyk, P., Luce, P., & Morgan, J. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465–494.

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

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can effect phonetic discrimination. *Cognition*, *82*, B101–B111.

Mirman, D., Magnuson, J., Graf Estes, E., & Dixon, J. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*, 271–280.

Navarro, T. (1966). *Estudios de fonología española [Studies of Spanish phonology]*. New York: Las Américas.

Newport, E., & Aslin, R. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127–162.

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

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 80–96.

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

Saffran, J. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*, 493–515.

Saffran, J., Aslin, R., & Newport, E. (1996a). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J., Newport, E., & Aslin, R. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

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

Sanders, L., Newport, E., & Neville, H. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*, 700–703.

Seidl, A., & Johnson, E. (2006). Infant word segmentation: Edge alignment facilitates target extraction. *Developmental Science*, *9*, 565–573.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*, 1–32.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*, 412–432.

Spitzer, S., Liss, J., & Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *Journal of the Acoustical Society of America*, *122*, 3678–3687.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 799–806.

Thiessen, E., & Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*, 706–716.

Toro, J., Rodriguez-Fornels, A., & Sebastián-Gallés, N. (2007). Stress placement and word segmentation by Spanish speakers. *Psicológica*, *28*, 167–176.

Toro, J., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*, B25–B34.

Turk-Browne, N., Jungé, J., & Scholl, B. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*, 552–564.

Tyler, M., Perruchet, P., & Cutler, A. (2006). A cross-language comparison of the use of stress in word segmentation. *Journal of the Acoustical Society of America*, *120*, 3087.

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.

Wallace, W. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, *63*, 410–424.

Yang, C. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, *8*, 451–456.