available at www.sciencedirect.com

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study

Toni Cunillera[b,c], Juan M. Toro[d], Nuria Sebastián-Gallés[e], Antoni Rodríguez-Fornells[a,b,*]

[a]Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain
[b]Dep. Psicologia Bàsica, Facultat de Psicologia, Universitat de Barcelona, Spain
[c]Dep. Psicologia. Universitat Illes Balears, Spain
[d]International School for Advanced Studies (SISSA/ISAS), Trieste, Italy
[e]GRNC, Parc Científic de Barcelona-UB; Hospital de Sant Joan de Dèu. Barcelona, Spain

ARTICLE INFO

Article history:
Accepted 7 September 2006
Available online 24 October 2006

Keywords:
Speech segmentation
Event-related brain potentials (ERPs)
Lexical stress
P2 component
Language learning

ABSTRACT

The study of the processes involved in speech segmentation has gained special relevance in recent years by trying to establish what type of information listeners use to segment the speech signal into words. An event-related brain potential experiment was conducted in order to understand how two of these cues (statistical and stress cues) interact. The experiment consisted of the presentation of artificial speech streams in which words were marked either by statistical cues alone, or by a combination of statistical and stress cues. As a baseline, comparison streams were also created with the same syllables but organized in random order. Results showed an N400 component that marks the on-line segmentation of speech into words, and an increased positivity (P2 component) for languages that include both types of cues. Possible implications of these results for the process of speech segmentation are discussed.

1. Introduction

In order to acquire language and build up a vocabulary, infants and second language learners must first learn to identify units (words) from fluent speech. The difficulty in segmenting units from a spoken utterance is accentuated by the lack of clearly marked word boundaries. That is, learners of foreign languages hear a continuous flow of speech that seems to be composed of a limited number of very long units. It is not until a certain degree of familiarity with a second language is gained that learners begin to identify clear word units. Eventually, parsing the speech stream is possible by exploiting different sources of information. However, invariant acoustic cues that mark word boundaries do not exist across languages and therefore learners have to rely on different strategies in order to segment speech. Previous research has shown that learners choose the strategy which is most appropriate to the particular language (see for a review, Cutler and Clifton, 1999).

Behavioral studies have pointed out the importance of different types of cues in speech segmentation (see Johnson and Jusczyk, 2001; Peña et al., 2002; Saffran et al., 1996a) were the first to demonstrate that infants can use statistical regularities (most likely transitional probabilities, Aslin et al., 1998) available in linguistic input to discover word boundaries, and subsequently segment continuous speech. In the study by Saffran et al. (1996a) a continuous 2-min synthetic stream

[*] *Corresponding author*. Dept. Psicologia Bàsica, Facultat de Psicologia, Universitat de Barcelona, Passeig de la Vall d'Hebron 171, Barcelona 08035, Spain. Fax: +34 93 4021363.
E-mail address: antoni.rodriguez@icrea.es (A. Rodríguez-Fornells).

0006-8993/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2006.09.046

composed of four tri-syllabic nonsense words was presented, with no pauses or any other acoustic information about possible word boundaries (e.g., "*bidakupadotibidakugolabu…*", where "bidaku", "padoti" and "golabu" are possible words). The only available cue for segmenting words was the distributional properties of the syllable sequence: in each word the likelihood of one syllable following another is three times greater than the likelihood of one syllable following a syllable from another word. The results of their experiment clearly showed that 8-month-old infants were able to segment these speech units based on statistical cues.

In addition to statistical properties, prosodic markers constitute another important source of information that listeners may use in word segmentation. In adults, Cutler and Norris (1988) posited that listeners use a metrical segmentation strategy in which they take each stressed syllable to mark the onset of a new word (see Norris et al., 2000). This strategy is potentially valid if a predominant trochaic (strong–weak) stress pattern is present in the corresponding language, as in stress-timed languages like English, but is less effective with languages that show different patterns (such as Spanish, see Sebastián-Gallés and Costa, 1997). Although the segmentation units appear to differ across languages (Cutler et al., 1983, 1986; Otake et al., 1993), a universal prosodic strategy for the segmentation of continuous speech has been proposed that exploits the rhythmic structure of speech input (Cutler and Clifton, 1999; Cutler and Norris, 1988). During language development, rhythm-based segmentation strategies adapted to the phonological structure of each language would be exploited and help in the acquisition of the initial lexicon. Indeed, several infant studies have provided evidence suggesting that native-language prosodic patterns are used for speech segmentation (Cutler, 1994; Echols et al., 1997; Jusczyk et al., 1993; Jusczyk et al., 1999). These results clearly demonstrate that infants use prosodic cues (lexical stress) in the early stages of creating a lexicon. Furthermore, other language-specific cues could also be involved in speech segmentation. For example, in Finnish, vowel harmony has been observed to facilitate speech segmentation in adults (Suomi et al., 1997).

One of the interesting issues regarding speech segmentation is how the multiple cues present in the speech signal are integrated in order for segmentation to proceed. By providing conflicting information regarding word boundaries one can explore this issue (see for example, Johnson and Jusczyk, 2001; Mattys et al., 1999). Using natural speech in 8-month-old infants, Johnson and Jusczyk (2001) demonstrated that infants rely more heavily on stress cues when both statistical and stress information conflict. In a similar study (Thiessen and Saffran, 2003), 9-month-old infants preferred stress as a segmentation cue rather than statistical information when confronted with both sources of information. However, the opposite pattern was observed in a younger group of 7-month olds. In a study with adults in which prosodic and statistical cues were combined (Saffran et al., 1996b) participants' performance improved when the prosodic cue coincided with the end of the word, but no interaction occurred when the prosodic cue indicated word onset. Thus, and in contrast with the results obtained using conflicting cues, the combination of statistical and prosodic cues yielded an improvement in the identification of word boundaries depending on the location of the prosodic cue within the word. Taken together, these results suggest that different strategies may be elicited when both sources of information indicating word boundaries are available, and therefore, that possible conflicts may arise when stress cues and statistical information do not match.

One way of addressing this question is with scalp recorded ERPs, which are used as an on-line direct measurement of brain activity with millisecond temporal resolution (Münte et al., 2001) and allow a chronometric analysis of language processes (Kutas, 1997; Kutas and Federmeier, 2000; Osterhout et al., 1997). Recently, ERPs have been used to investigate language segmentation (Sanders et al., 2002). A specific component, the N1, seems to be sensitive to word onset perception. For example, in a study performed with native English speakers, word onsets in continuous speech elicited a larger N1 than word-medial syllable onsets matched for physical characteristics (Sanders and Neville, 2003). A similar N1 effect has been observed when stressed and unstressed syllables were compared. However, in this same study, N1 effects were not observed for late English learners (native Japanese speakers), suggesting that non-native speakers do not use stress information in order to segment speech as demonstrated by native speakers. Sanders et al. (2002) also reported evidence that nonsense words elicited a larger N400 after training, a finding consistent with the interpretation of the N400 as an index of lexical search. This study clearly demonstrated that the N400 component is involved in the learning process of nonsense words and regarded it as an indicator of the fact that speech has been segmented.

To assess the contributions of stress and statistical cues in speech segmentation two experimental conditions were created (see Fig. 1A). In the first condition, nonsense language streams (hereafter referred to as word streams) were constructed with the same structure as the languages used in Saffran et al. (1996a). In the second condition, tested in a separate ERP session, an acoustic cue was added in order to facilitate the segmentation of the word stream (Johnson and Jusczyk, 2001). The pitch of the first syllable of each word was increased, thereby creating an artificial stress (cue) at the beginning of each word. In addition, non-word streams were created for each condition. They were formed by the same syllables as the word streams, but randomly concatenated. These non-word streams were used as control streams for the two word conditions.

The on-line measure during stream presentation allowed for a direct comparison between word and non-word streams, and of how word segmentation is performed on the basis of purely statistical information or when convergent statistical and stress information are presented. A critical difference between this study and the previous one (Sanders et al., 2002) was that ERPs were recorded during an on-line speech segmentation of possible words that had not been previously learned, that is, participants were exposed to new items each time they heard a different language streams. As a working hypothesis, since the N1 has been proposed as a word segmentation index (Sanders et al., 2002), differences in the amplitude of the N1 were expected between possible words

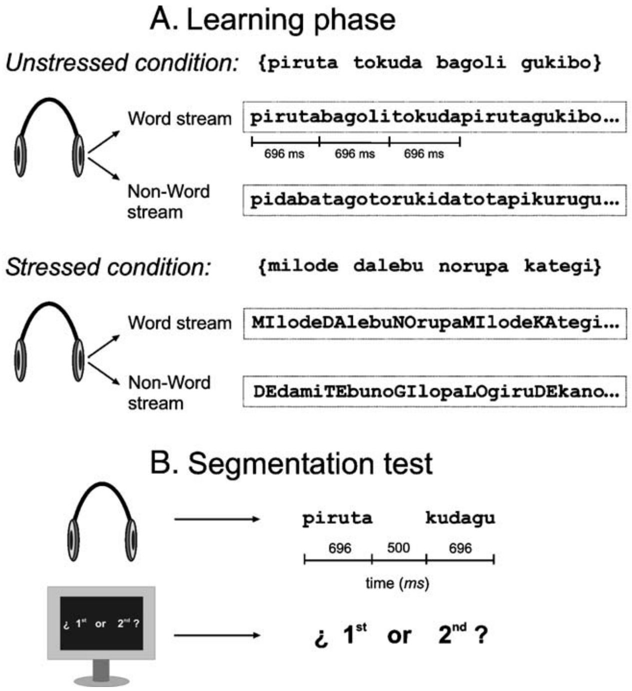
A. Learning phase
Unstressed condition: {piruta tokuda bagoli gukibo}
Word stream
pirutabagolitokudapirutagukibo...
696 ms
696 ms
696 ms
Non-Word stream
pidabatagotorukidatotapikurugu...
Stressed condition: {milode dalebu norupa kategi}
Word stream
MIlodeDAlebuNOrupaMIlodeKAtegi...
Non-Word stream
DEdamiTEbunoGIlopaLOgiruDEkano...
B. Segmentation test
piruta
kudagu
696
500
696
time (ms)
¿ 1st or 2nd ?
¿ 1st or 2nd ?

Fig. 1 – Illustration of the procedure used in the language learning and segmentation test phases in both unstressed and stressed conditions. (A) In the unstressed condition, word streams were created by randomly concatenating, without any pause, each of the four words, which were created combining the initial pool of 12 syllables (see Appendix 1). In the first stream, the defined words were: "piruta", "tokuda", "bagoli", "gukibo" (in between brackets). In contrast, in the non-word streams, the syllables were randomly ordered. In the stressed condition, the initial syllable of each word, or every third syllable in the non-word streams, was stressed (stressed syllables are represented in capital letters). (B) An example of a trial in the segmentation auditory two-alternative-forced-choice test is shown. A word and a part-word from the stream were presented and the participant had to decide which one of them had been previously presented in the stream. Part-words were created by concatenating the last two syllables of a word and the first one of another (e.g., kudagu) or the last syllable of a word with the two first syllables of another word.

(language streams) and non-words (random streams) in both experimental conditions. According to Sanders and Neville (2003), a similar effect of the amplitude of the N1 might be observed when comparing stressed and unstressed words. We also explored the possibility that another ERP component instead of the N1 might be related to a prosodic cue.

Participants' performance in the behavioral word recognition test (see Fig. 1B) is expected to be better in the stress condition, in which a combination of prosodic and statistical cues should help to identify word boundaries (Saffran et al., 1996b). Finally, for the other ERP word segmentation index proposed, that is, the N400, a larger amplitude might be observed for possible words than for non-words, which may show the N400's typical lexical search effect.

2. Results

2.1. *Behavioral performance*

In the unstressed condition, the percentage of correctly detected words was 67.5%, which was significantly better than chance (50%) ($t(14)=6.6$, $P<0.001$). In the stressed condition, the mean percentage was 62.5% (significantly different from chance, $t(14)=3.9$, $P<0.01$). There were no differences in performance in the two conditions ($t(14)=1.2$, $P>0.2$).

2.2. *ERPs*

Grand average ERPs for the stressed and unstressed conditions, and for word streams vs. non-word streams are shown in Figs. 2 and 3. The following auditory-evoked components are clearly identified at central and frontal sides: after a small positive peak identified as the P1, the N1 component is present, peaking at about 90 ms and followed by a positivity (the P2) with a peak at approximately 170 ms. These early components are followed by a broadly distributed negativity (range 300–500 ms) in the word conditions which is not present in the non-word conditions. In the non-word streams, the P1–N1–P2 complex is repeated for each syllable presented and easily identified in the corresponding figures. In the word streams, this pattern overlaps with the N3–N4 component in the stressed and unstressed conditions. In the stressed conditions, a clearly enhanced positivity with an onset at about 100 ms differentiated words from non-words. When comparing words in the stressed and unstressed languages (Fig. 3), this enlarged positivity, which encompasses part of the P2 component, is clearly observed. Note, however, that the N3–N4 component showed a similar increase in both word conditions. Finally, the only difference seen for the N1 component was found between words and non-words in the stressed condition. No difference was found for the N1 either between words or between non-words from both stressed and unstressed conditions.

The results of the omnibus analysis on the mean amplitude measures of the previous components are shown in Table 1. Although a main effect of stress was not observed in any time window, a clear interaction between Word/Non-word and Stress was present at the P2 and N3–N4 components, as well as for the interaction between these factors and electrode. These interactions suggest differences in the topographical distribution of the response elicited by stress in the word and non-word streams. To better characterize the mentioned effects, follow-up analyses were conducted for stressed and unstressed conditions as a function of words and non-words. Notice that, when considering the results of the omnibus ANOVA, as the interaction between Stress × Word/Non-word was not significant in the first two time windows, the decomposition is not permitted. However, we nevertheless performed the pairwise ANOVAs in all time windows because of the rather exploratory characteristics of the present investigation. Furthermore, the lack of significance in the interaction between the Stress × Word/Non-word condition at the N1 time window, is most likely due to the

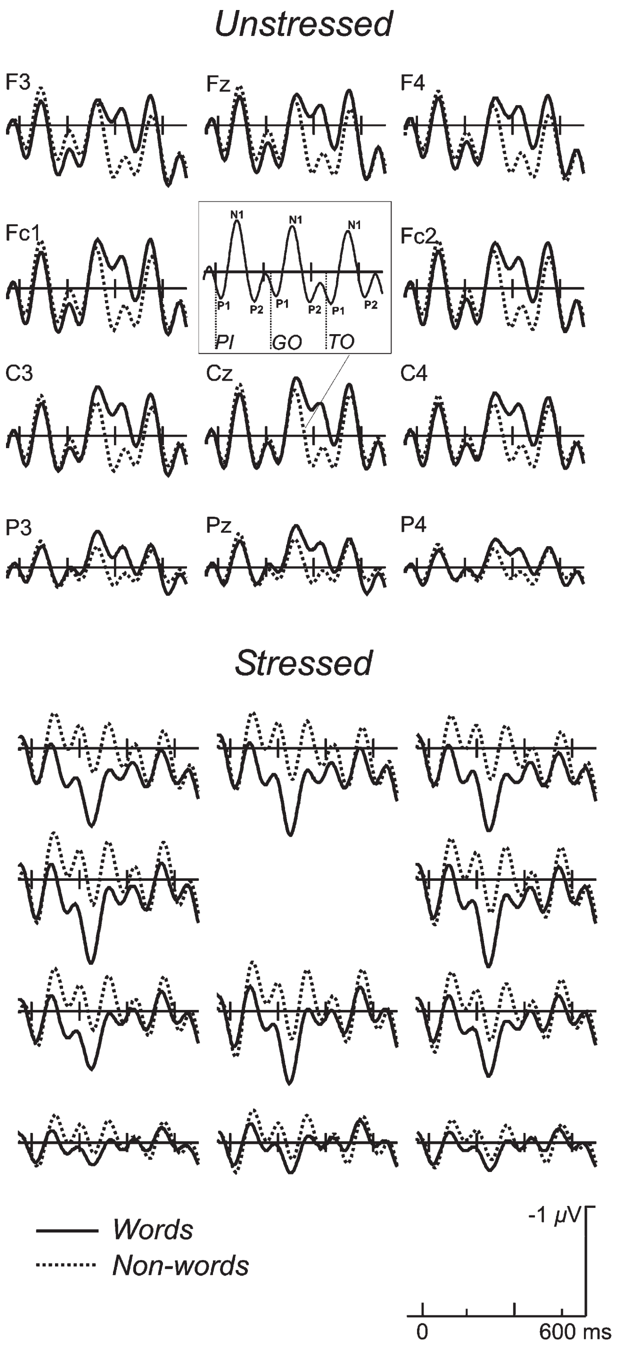

Fig. 2 – Grand average potentials for the different word and non-word streams are depicted for the unstressed (upper panel) and the stressed (lower panel) conditions. Shown are the ERPs to words (solid line) and non-words (dotted line). Different electrode positions at central and parasagittal locations are shown.

fact that in both conditions (stressed and unstressed) non-words tend to show a larger amplitude (see Fig. 2).

Table 2 shows the effects between words and non-words in the stressed and unstressed conditions. In the unstressed condition, words differed from non-words only in the N3–N4 time range. As shown in Fig. 2, this effect is due to the increased negativity observed in the words after 300 ms (mean amplitude words −0.04±0.26 μV; non-words 0.15±0.2).

The interaction with electrode and condition was further studied with the topographical analysis (see also the topographical map in Fig. 4). Scalp distribution analysis after vector normalization showed that this N400 component was larger at medial, central, and frontocentral locations for words (Word/Non-word and Laterality, $F(1,14)=12.2$, $P<0.01$, Word/Non-word × Anterior–posterior, $F(2,28)=11.7$, $P<0.01$). Specifically, it was more prominent at central medial locations in the left hemisphere (Word/Non-word × Hemisphere × Laterality, ($F(2,28)=18$, $P<0.001$)).

In the stressed condition (Table 2), words and non-words differed very early, with an onset at approximately 70 ms (see the difference in waveforms shown in Fig. 4). Statistically, this effect is maximal at the P2 time window where words showed a large increase in positivity (mean amplitude words 0.25±0.26 μV; non-words −0.04±0.18). The interactions with electrode suggest a different topographical distribution (see Fig. 4). Distributional analysis showed that this positivity is more prominent for words at the frontal and medial locations (at 170–250 ms, Word/Non-word × Anterior–posterior, $F(2,28)=18.1$, $P<0.001$ and Word/Non-word × Laterality, $F(1,14)=11.5$, $P<0.05$). These interactions were also significant at 70–130 ms (N1) but not at the 300–500 ms (N3–N4) time window.

Table 3 shows the effect of stress for word and non-words stream (see also Fig. 3). In the word streams, stress showed an increased positivity after the P2 component, which is reflected by the interactions between Stress × Electrode and the main effect of stress at the N3–N4 interval. A further analysis restricted to the 200- to 300-ms interval, showed a main effect of stress in words ($F(1,14)=10.4$, $P<0.01$; unstressed words, 0.17±0.25 μV; stressed words 0.36±0.32) as well as an interaction with Electrode ($F(14,196)=8.7$, $P<0.001$). Further distributional analysis showed the central distribution of this positivity (Stress × Laterality, at 70–130, $F(1,14)=15.5$, $P<0.01$; at 170–250 ms, $F=7.34$, $P<0.02$; at 300–500, $F=24.5$, $P<0.001$). In contrast, a different pattern was observed in the non-word streams (see Fig. 3), which showed a small increase in the negativity in the range of the P2–N2 components of each

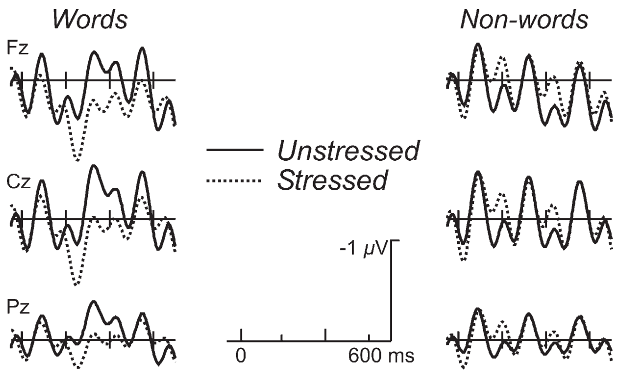

Fig. 3 – Grand average potentials at midline electrode locations directly comparing unstressed vs. stressed conditions in the word (left side) and non-word (right) streams. Stressed words showed an enhanced P2 component immediately followed by the negative increase at the 300–400 ms time range.

Table 1 – Overall results of the omnibus ANOVA

Time window (ms)	St *F*(1,14)	St × *E* *F*(14,196)	WnW *F*(1,14)	WnW × *E* *F*(14,196)	St × WnW *F*(1,14)	St × WnW × *E* *F*(14,196)
20–80 (P1)	0.30	0.95	0.45	0.95	3.32	1.14
70–130 (N1)	1.09	1.95	**18.9**$^{+++}$	**4.01**$^{+}$	2.54	2.06
170–250 (P2)	1.80	0.72	**39.6**$^{+++}$	**12.9**$^{+++}$	**19.0**$^{+++}$	**6.52**$^{+++}$
300–500 (N3–N4)	1.52	2.15	2.73	1.21	**10.6**$^{++}$	**7.81**$^{+++}$

Notes. Bold numbers indicate significant effects. 'St' = stress (unstressed vs. stressed conditions); 'WnW' = word/non-word; 'E' = electrode. $^{+}P<0.05$, $^{++}P<0.01$, $^{+++}P<0.001$.

syllable for the stressed conditions. This effect was significant at the P2 time window with no differences between conditions in the topographical distribution (unstressed non-words, 0.13 ± 0.15 μV; stressed non-words, −0.04 ± 0.18).

To obtain an estimate of the neural generators underlying the specific P2 stress component, a brain-electric source analysis (BESA) was performed on the grand average (words minus non-words) difference waveform. The P2 component was explained by a two-source model with two symmetrical sources located in the superior temporal gyrus, near the primary/secondary auditory cortex (see Fig. 4C). Source waveforms originating from both hemispheres are depicted in Fig. 4D. The Talairach coordinates for the left hemisphere source were −47, −18, and 8. The residual variance not accounted for by the two-source model in the P2 time window (175–250 ms) was only 1.3%, showing a very good fit for the scalp distribution studied.

3. Discussion

The present study evaluated the effect of prosodic cues on speech segmentation using on-line ERP measures. In one condition, speech segmentation could only be accomplished using statistical cues. In the other condition, two types of information could be used: statistical information (transitional probabilities between syllables) and a prosodic cue (pitch increase in the first syllable) that marked the onset of each word.

Word and non-word streams differed in the unstressed condition at the N3–N4 component. This result replicates those of Sanders et al. (2002). In that study, the authors interpreted the word effect on the N400 component as a possible lexical search process triggered by segmented words. Note however that the scalp distribution of the N400 component in our study shows a maximum at frontocentral sites, which does not agree with the more posterior distribution of the N400 component described in Sanders et al. (2002). Differences in the scalp distribution of the N400 component are probably due to the differences in the paradigms used in the two studies. Sanders et al. trained the participants to recognize six nonsense words presented in a continuous auditory stream. Possible nonsense words were compared before and after training. In contrast, our paradigm focused directly on *on-line segmentation*, because nonsense words are discovered during exposure to the continuous auditory streams without any previous training. Thus, differences in the distribution of the N3–N4 component between the studies might be attributed to (i) the on-line segmentation of nonsense words that had never been heard before and (ii) the segmentation of the auditory stream using lexical recognition of previously learned nonsense words (as in Sanders et al., 2002). In the present study, it is impossible to distinguish between both types of mechanisms. This latter process – segmenting speech based on newly learned words – also deserves mentioning. For example, in the unstressed condition, listeners could use their familiarity with a recently segmented new item or word to predict that whatever comes next must be the onset of a new lexical item. New learned words are very useful for speech segmentation in the early stages of learning a new language. In fact, it is a process used to segment speech independently of how familiar a listener is with particular prosodic features of a language. In order to disentangle the issue of the specificity of the N3–N4 component as a segmentation index (and its scalp distribution), further studies are needed to evaluate both segmentation mechanisms.

A very important issue is the degree to which the mental representation elicited during this learning process has any lexical or pre-lexical status. It has been shown that infants'

Table 2 – Effects of word/non-word separated for unstressed and stressed conditions

Time window (ms)	Unstressed		Stressed	
	Word/Non-word *F*(1,14)	Word/Non-word × Electrode *F*(14,196)	Word/Non-word *F*(1,14)	Word/Non-word × Electrode *F*(14,196)
20–80 (P1)	2.63	0.42	1.3	1.54
70–130 (N1)	2.07	0.50	**17.8**$^{+++}$	**5.78**$^{++}$
170–250 (P2)	3.58	2.61	**45.2**$^{+++}$	**17.4**$^{+++}$
300–500 (N3–N4)	**20.45**$^{+++}$	**5.59**$^{+++}$	1.66	**3.45**$^{+}$

Notes. $^{+}P<0.05$, $^{++}P<0.01$, $^{+++}P<0.001$.

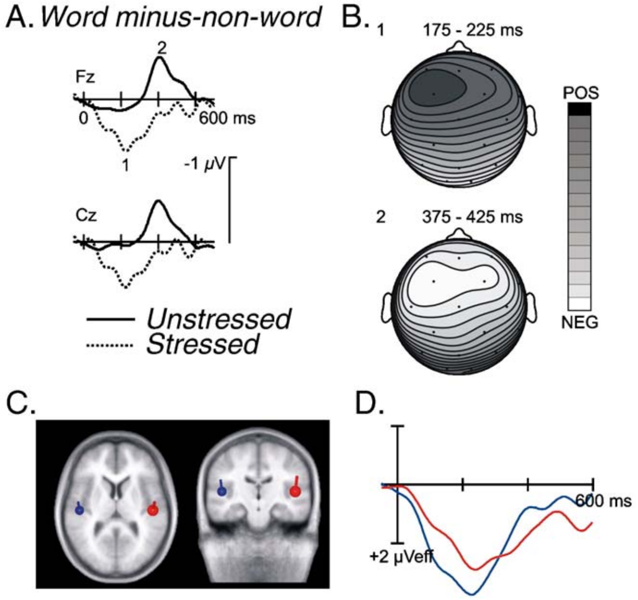

Fig. 4 – (A) Difference waveforms (word minus non-word streams) in the unstressed and stressed conditions for frontocentral locations. A very different morphology of the difference waveforms is found for both conditions, reflecting the enhanced P2 component in the stressed condition and the N3–N4 component in the unstressed condition. (B) Spline-interpolated isopotential maps derived from the word minus non-word difference waveform for the P2 component (time window 175–225 ms) in the stressed condition. Below, the corresponding map for the negative component (N3–N4, time window 375–425 ms) in the unstressed condition (word minus non-word). Both sets of maps are scaled to encompass the minimal and maximal voltages for these time windows (minimum/maximum values for the P2 map 0.02/0.57 μV and for the N3–N4 map −0.51/−0.07 μV). (C) Dipole solutions for the P2 component in the word minus non-word difference waveform (stressed condition). Estimated anatomical locations (in Talairach coordinates) of both dipolar sources are shown projected onto a coronal and axial sections from a standard T1 MRI image. (D) Time-varying source waveforms for each of the dipoles used in the forward model. This model explained 98.7% of the variance from the difference waveforms in the 175–250 ms time window.

listening times increase when a segmented word is embedded in a real language context (e.g., *I want to see a "golabu"*) compared to words embedded in nonsense frames (e.g., *Te gatesi tu ner e "golabu"*) (Saffran, 2001). This result shows that infants consider the segmented words as a possible candidate word in their native language. Based on this, Saffran (2001) proposed that the output of statistical learning is the creation of a linguistic or word-like representation for segmented items. The N400 results (in our study and in Sanders et al., 2002) also point to the construction of a possible linguistic trace in the lexicon that might be prepared to be further mapped to a referent. These electrophysiological results also lend some support to the hypothesis that statistical learning, as a domain-general mechanism, is used in first and second language learning (see Gomez and Gerken, 2000).

In the unstressed condition, no differences were observed between words and non-words in the N1 component, although an amplitude decrease is encountered in this time range for stressed words. The reduction of the N1 could be due to the fact that these components are superimposed on a larger positivity, the P2 specifically seen for stressed words. The lack of an effect in the N1 component for the unstressed condition, which was expected in view of previous studies (Sanders et al., 2002; Sanders and Neville, 2003), may also have been due to the differences in study design (see above). In fact, in Sanders et al. (2002), the N1 effect was not significant across all subjects. When the sample was divided considering a post-training accuracy test, listeners whose performance was in the range of the current study (~67%) showed only an N400 effect whereas those with better performance showed both an N1 and N400 effect. In the present study, as well as in the low-performance group in Sanders et al., the N1 component may reflect only the detection of syllables rather than word-initial onset. The N1 effect observed in Sanders et al. (2002) may therefore be associated with strategic individual differences in learning new words. For example, differences in the N1 component may appear if listeners use predictive segmentation cues (e.g., word-final syllable lengthening) and not when listeners base their segmentation on information available from word-initial syllables.

In contrast, in the stressed conditions, word/non-word differences were already observed at the P2 component time range, with a large increase in positivity and statistically affecting all subsequent ERP components. This effect was greatest at about 225 ms after the onset of the first syllable, which was identified in the present study as an increased P2 component. After this positivity, the N3–N4 modulation was also observed in the stressed condition. The topographical distribution of the P2 component was frontocentral showing a polarity reversal at occipitotemporal locations, replicating a previous report (Potts et al., 1998). The neural generators of this component have been attributed to multiple sources near the primary and secondary auditory regions in the superior temporal lobe (Brodmann's area 22 and planum temporale, Bosnyak et al., 2004; Godey et al., 2001; Hari et al., 1987; Liegeois-Chauvel et al., 1994; Pantev et al., 1996; Picton et al., 1999; Scherg and von Cramon, 1986). In our study, the results of the forward solution using dipole modeling showed a nearly perfect fit using two sources located in the superior temporal gyrus. Despite the spatial resolution limitations of the neural source analysis, the changes observed in the P2 component can be attributed to neural generators seeded in auditory regions along the Sylvian fissure.

The increased positivity in the P2 time range may be related to the strategy that listeners used to segment and recognize the nonsense words. For example, listeners may use the repetition of two lower pitch syllables to predict that the following syllable would be a word-initial syllable. However, a non-predictive segmentation cue could also be used. Listeners may hear a higher pitched syllable in the stressed word condition and then use that information as indicating the word-initial syllable. Note that both strategies are plausible based on the acoustic properties of syllables

Table 3 – Effects of stress separated for words and non-words conditions

Time window (ms)	Words		Non-words	
	Stress *F*(1,14)	Stress × Electrode *F*(14,196)	Stress *F*(1,14)	Stress × Electrode *F*(14,196)
20–80 (P50)	0.26	0.24	1.73	1.95
70–130 (N1)	4.31	**3.57^{+}**	0.08	0.41
170–250 (P2)	1.15	**3.21^{+}**	**12.7^{++}**	2.72
300–500 (N3–N4)	**9.37^{++}**	**5.78^{+++}**	2.97	**2.86^{+}**

Notes. $^{+}P<0.05$, $^{++}P<0.01$, $^{+++}P<0.001$.

(pitch differences). Likewise, this difference between predictive and initial-word pitch recognition may explain differences observed in the ERP components. In view of these data, we favor the idea that higher pitches elicit larger P2s, especially if the specific syllables that received this higher pitch are predictable, as in the words streams. Interestingly, items in the non-word streams of the stressed condition did not show the increased positivity observed in stressed-words. However, an enlarged N2 component when compared to unstressed non-words was observed. This interaction between word/non-word and stress probably reflects the differential effect of prosodic cues in the words or non-word streams. Additionally, the present neurophysiological data support previous studies showing that word stress – which is a component of prosody in speech – can be used to discover word boundaries (Cutler and Clifton, 1999; Jusczyk et al., 1999; Mattys et al., 1999).

Although the P2 component has received little attention in auditory perception, it has recently been proposed as an ERP index of neuroplasticity in the auditory cortex. Intensive training on complex auditory discrimination has been associated with increases in the P2 component in different studies (Atienza et al., 2002; Bosnyak et al., 2004; Reinke et al., 2003; Shahin et al., 2003; Tremblay and Kraus, 2002; Tremblay et al., 2001). The increased amplitude of the P2 component after cochlear implant in congenitally deaf patients (Purdy et al., 2001) supports this idea. Based on these studies, two different explanations have been proposed to account for the neurophysiological bases underlying the learning-induced changes in the P2 component. Tremblay et al. (2001) proposed that the improved auditory perception is related to a larger neural synchronicity of the auditory regions that respond to specific learned acoustic components (see also Tremblay and Kraus, 2002). This view is based on the idea that ERPs reflect synchronized postsynaptic potentials in clusters of cortical pyramidal neurons (Allison et al., 1986). The other view considers that the P2 learning effect indicates the recruitment of larger auditory neural populations involved in perceptual learning (Reinke et al., 2003). This idea is supported by studies (Recanzone et al., 1993) observing an increase in the number of neurons firing to the specific acoustic feature when using an auditory frequency discrimination task in owl monkeys.

In a sense, our findings corroborate the importance of the P2 component as an ERP index of auditory discrimination and its involvement in fast auditory learning. We found the enhancement of the P2 to be selective for the combination of statistical information and prosodic markers, which probably induced an increase in the temporal synchrony or recruitment of neural populations in the auditory cortex. One possible interpretation is that prosodic markers act as attentional cues that prime language segmentation. Since speech segmentation is basically a learning task, attention has been shown to play a role for successful segmentation (Toro et al., 2005), and may be critical in the neurophysiologic changes observed during speech segmentation. However, no statistical behavioral differences were observed between the stressed and unstressed conditions. Interestingly, we did not find a facilitatory effect of stress in this study, as reported by other authors (Saffran et al., 1996b). One possible explanation might be related to the fact that in Spanish, stress placement is not fixed (Navarro Tomás, 1965). Although penultimate stress is predominant, about one quarter of polysyllabic words has final or antepenultimate stress. It is important to note that the effect of stress in speech segmentation is highly language dependent. For example, some languages have fixed stress, as in Finnish, where stress is always placed on the initial syllable. Vroomen et al. (1998) have reported that speech segmentation for Finnish speakers is facilitated by word-initial stress cues. Finally, it is important to note that the mean percentage of correct recognition in these studies is around 65–70% (Newport and Aslin, 2004; Peña et al., 2002; Saffran et al., 1996a,b), which may also point to a ceiling effect in this type of task.

To our knowledge, this is the first study to provide different objective brain indexes of the time course associated with word segmentation using prosodic and statistical cues. Moreover, the proposed ERP design might be very useful for addressing further aspects related to the problem of how segments of natural and artificial synthetic speech are isolated. We recorded ERPs to word and non-word streams with the presence or absence of stress cues. Our electrophysiological results demonstrate a clear interaction between statistical learning and stress cues. Possible words compared to non-words showed a clear N3–N4 component. In contrast, stressed words elicited an increased positivity in the range of the P2 component with a possible generator in the non-primary auditory cortex. The effect was reversed for the non-word streams.

4. Experimental procedure

4.1. *Participants*

Eighteen healthy adult undergraduate psychology students (5 males) at the University of Barcelona participated in the experiment. All participants [mean age 23.4±4.9 (SD)] were right-handed and were native speakers of Spanish. Three participants' data were discarded from the analysis due to excessive eye movement. The data from the remaining 15 participants was analyzed. The experiment was approved by the local ethical committee. Written consent was obtained from each subject prior to the experiment. All participants were paid at the end of the experiment or were given extra course credit for their participation. No participants reported hearing deficits or language learning impairment.

4.2. *Stimuli*

Five words streams were created for each of the stressed and unstressed conditions. The word streams had the same structure as the ones created by Saffran et al. (1996a) (see Fig. 1). Each stream consisted of 4 different tri-syllabic nonsense words (referred in the text as words). Each word was repeated 192 times, thereby resulting in a total of 3840 items per condition [5 (languages) × 4 (words/Non-words) × 192 (repetitions)]. Words were concatenated to form a text-stream which was then transformed as a whole into an acoustic-stream using the speech synthesizer MBROLA software which is based on concatenation of diphones (Dutoit et al., 1996). Cooledit software was used to equate the length of the different streams into millisecond precision, which was necessary for posterior ERPs analysis. The exact duration of each stream was 8 min 54 s and 528 ms. Crucially, the streams were constructed with no acoustic pauses between items. Because only 59 syllables can be used for the construction of the five streams, only one syllable was repeated in 2 streams. In all streams the transitional probability within syllables forming a word was 1.0, while syllables spanning word boundaries had a transitional probability of 0.33. The same pool of syllables was used for the construction of the languages in both stressed and unstressed conditions, but concatenated in a different order (see Appendix 1).

The resulting word streams in the unstressed condition contained no pauses or other acoustic indication of word onset. In contrast, in the stressed condition, every stream contained an acoustic indication of word onset. The pitch of the first syllable of each word in the stream was increased by 20 Hz, creating an artificial stress at the beginning of each possible word (Johnson and Jusczyk, 2001). Although stressed syllables are also characterized by an increase in length, we maintained the duration stable across syllables within a word in order to avoid segmentation being based on syllable lengths rather than on pitch. The fact that all syllables across streams are matched in length permits a direct comparison between conditions.

Furthermore, 10 different streams (five for each condition, stressed and unstressed) were created as a baseline (*non-word streams*) by using the same syllables presented in each one of the word streams but concatenated in random order. That is, each syllable in the stream could be followed by any of the other eleven syllables composing the stream. Thus, the transitional probability across syllables was 0.09. The low transitional probability should create a condition where the extraction of a clearly segmented word should be impossible, as the probability could not be used in order to identify possible words. In the unstressed condition, non-word streams contained no acoustic information. In contrast, in the stressed condition, the first syllable of each syllable triplet was stressed regardless of which syllable fell in the stressed positions.

In order to test segmentation of the streams, test items were created. For the stressed and unstressed conditions items consisted of the four words forming each stream, plus part-words that consisted in the concatenation of the two last syllables of a word and the first one of another, or the last syllable of a word and the first two syllables of another. For non-words conditions, items comprised sixteen tri-syllabic sequences selected from the streams.

4.3. *Procedure*

Subjects participated in two ERP recording sessions held with an interval of at least two weeks in between measures. Five unstressed word and 5 unstressed non-word streams were presented during the first session, and 5 stressed word and 5 stressed non-word streams were presented during the second (see Fig. 1). In both sessions participants were always required to listen carefully to a stream of sounds and asked to identify the words that formed it. They were not informed of the type of stream to which they were being exposed to in each session. The first stream was always one of the languages that contained possible words. The presentation of the streams was constrained, so that no more than two streams of each type (word or non-word) followed each other.

A two alternative forced choice (2AFC) auditory behavioral test was administered immediately after exposure to each stream in order to determine whether participants were able to identify the words from the word stream. ERP recordings were not measured during this phase. The test comprised of 8 pairs of test items (word vs. part-word). For non-words streams, test items were composed by 16 different tri-syllabic groupings. The task was the same for the words and non-words conditions. It should be noted, however, that correct responses were not possible in the non-word condition, but the same test was delivered to the participants in both conditions, mainly to keep them uninformed about the impossibility of segmenting the non-word streams and consequently, to ensure that subject motivation was the same across all streams. After the presentation of each test pairing, participants had to press a response button indicating whether it was the first or the second word in the pair that belonged to the stream they had just heard. The order of presentation of words and part-words in the test pairs was balanced. In the stressed condition, pitch was removed from the test items in order to match words and part-words.

Subjects were able to rest after each stream. The next stream began after listeners felt comfortable enough to maintain their attention for the next 9 min.

4.4. *Electrophysiological recording*

The ERPs were recorded from the scalp using tin electrodes mounted in an electrocap (Electro-Cap International) and located at 29 standard positions (Fp1/2, Fz, F7/8, F3/4, Fc1/2 Fc5/6, Cz, C3/4, T3/4, Cp1/2, Cp5/6, Pz, P3/4, T5/6, Po1/2, O1/2). Biosignals were referenced on-line to an electrode placed in the outer canthus of the right eye and then re-referenced off-line to the mean of the activity at the two mastoids processes. Vertical eye movements were monitored with an electrode at the infraorbital ridge of the right eye. Electrode impedances were kept below 3 kΩ.

The electrophysiological signals were filtered with a bandpass of 0.01–50 Hz (half-amplitude cutoffs) and digitized at a rate of 250 Hz. Trials with base-to-peak electro-oculogram (EOG) amplitude of more than 50 μV, amplifier saturation, or a baseline shift exceeding 200 μV/s were automatically rejected off-line. No significant differences were observed for the percentage of rejected trials in both sessions ($t(14)=0.149$, $P>0.8$; unstressed 15.9% and stress 15.5%).

4.5. *Data analyses*

Stimulus-locked ERPs (word/non-word onset) for artifact free trials corresponding to all words and non-words from the streams were averaged for epochs of 1024 ms starting 100 ms prior to the stimulus. This was performed separately for the 4 different conditions. Mean amplitude measures were taken in four time windows which encompass the major ERP components of this study and based on previous studies (Sanders and Neville, 2003; Sanders et al., 2002): P1 (20–80 ms), N1 (70–130 ms), P2 (170–250), N3–N4 (300–500). These mean voltage measures were submitted to an omnibus repeated measures analysis of variance (ANOVA) including three within-subject factors: *Stress* (stressed vs. unstressed conditions), *Word/Non-word* (words vs. non-words streams) and 15 levels of electrode. Twelve of the 15 selected electrodes were used for topographical analysis and the decomposition of the interactions only when the Factor × Electrode interaction was significant. This analysis was carried on data corrected using the vector normalization procedure described by McCarthy and Wood (1985) (see also Urbach and Kutas, 2002). The 12 selected electrodes (F7, F3, F4, F8, T3, C3, C4, T4, T5, P3, P4, T6) were divided according to three factors: anterior–posterior (anterior, central, posterior), lateral (medial and lateral) and hemisphere (right–left). In the omnibus ANOVA these 12 electrodes plus midline sides (Fz, Cz, Pz) were included. For all statistical effects involving two or more degrees of freedom in the numerator, the Greenhouse–Geisser epsilon was used to correct for possible violations of the sphericity assumption Jennings and Wood (1976). The exact p-value after the correction is reported.

The dipolar sources of the P2 component related to stress were modeled using brain electric source analysis algorithm (BESA, v. 4.2, Scherg, 1990). The difference waveform obtained subtracting non-words minus words in the stress condition was analyzed. Following previous descriptions of the source structure of the P2 component, two symmetrical regional sources were first fitted for location and orientation near the supratemporal plane, next to the primary and secondary auditory cortex (Scherg and von Cramon, 1985, 1986). Then, the regional sources were transformed to single dipole sources. The BESA algorithm estimates the location and the orientation of multiple equivalent dipolar sources by calculating the scalp distribution that would be obtained for a given dipole model (forward solution) and comparing it to the original scalp distribution. Interactive changes in the location and in the orientation in the dipole sources lead to a minimization of the residual variance between the model and the spatio-temporal distributions observed. BESA analysis was conducted using an idealized four-shell ellipsoidal head model with relative conductivities of 0.33, 1.0, 0.0042 and 0.33 for the brain, cerebrospinal fluid, skull and scalp. The thickness of the head, scalp, bone, and cerebrospinal fluid was 85, 6, 7, and 1 mm, respectively (Berg and Scherg, 1994). The final locations of each dipole in the group-average BESA model were projected on mean structural T1 MRI image of 24 individuals and converted into Talairach and Tournoux (1988) coordinates in the standard Montreal brain.

Acknowledgments

This research was supported by research grants from the Spanish Government (MCYT) to A.R.F. (with EC Funds FEDER SEJ2005-06067/PSIC) and N.S.G. (with EC Funds FEDER SEJ2004-07680-C02-01/PSIC), the Human Frontier Science Program (HFSP, grant RGP68/2002), the James S. McDonnell Foundation (Bridging Mind, Brain and Behavior), and the European Science Foundation (EUROCORES: The origin on man, language and languages). J.M.T. was supported by a fellowship of the Spanish Ministerio de Educación y Cultura (AP2000-4164). A.R.F. was supported by the Ramon y Cajal research program. We are grateful to the two anonymous reviewers for their valuable comments and suggestions on an earlier version of the manuscript. Special thanks to Sergio Romero, Jordi Riba, Jürn Müller, Ruth de Diego Balaguer, Xavier Mayoral, Scott Sinnett, and other members of the GNRC group for the technical help and comments at various stages of the project.

Appendix A

Artificial languages used in stress and unstressed conditions. Part-word construction in each language followed the same structure as exemplified in the first language.

A.1. *Unstressed*

Language 1: *Words*: PIRUTA, BAGOLI, TOKUDA, GUKIBO
Part-words: RUTABA, TABAGO, GOLITO, LITOKU, KUDAPI, DAPIRU, GOLIGU, LIGUKI, KIBOBA, BOBAGO, RUTAGU, TAGUKI, KIBOTO, BOTOKU, KUDABA, DABAGO, GOLIPI, LIPIRU, RUTATO, TATOKU, KUDAGU, DAGUKI, KIBOPI, BOPIRU
Language 2: *Words*: PABELA, DINEKA, LUFARI, JISODU
Language 3: *Words*: MAJUPE, JEROGA, DEMUSI, FOLETI
Language 4: *Words*: PUKEMI, RAFINU, BINAPO, MEDOGI
Language 5: *Words*: NONIGE, BULOTE, REMOFU, KOTUSA

A.2. *Stressed*

Language 1: *Words*: MILODE, DALEBU, NORUPA, KATEGI
Language 2: *Words*: NEDOLI, RIFONU, BATOGU, KIRAPU
Language 3: *Words*: GONABE, MUDILA, RONIGE, PIKUSA

Language 4: *Words*: FUBIRE, JETUSI TAFIKO, KEMAPO
Language 5: *Words*: TIFAJU, SODUJI, MELUBO, GANIPE

REFERENCES

Allison, T., Wood, C.C., McCarthy, G., 1986. The central nervous system. In: Coles, M.G.H., Donchin, E., Porges, S. (Eds.), Psychophysiology: Systems, Processes, and Applications. Guilford Press, New York, pp. 5–25.

Aslin, R.N., Saffran, J.R., Newport, E.L., 1998. Computation of conditional probability statistics by 8-month-old infants. Psychol. Sci. 9, 321–324.

Atienza, M., Cantero, J.L., Dominguez-Marin, E., 2002. The time course of neural changes underlying auditory perceptual learning. Learn. Mem. 9, 138–150.

Berg, P., Scherg, M., 1994. A fast method for forward computation of multiple-shell spherical head models. Electroencephalogr. Clin. Neurophysiol. 90, 58–64.

Bosnyak, D.J., Eaton, R.A., Roberts, L.E., 2004. Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. Cereb. Cortex 14, 1088–1099.

Cutler, A., 1994. The perception of rhythm in language. Cognition 50, 79–81.

Cutler, A., Clifton, C., 1999. Comprehending spoken language: a blueprint of the listener. In: Hagoort, P., Brown, C.M. (Eds.), The Neurocognition of Language. Oxford University Press, New York, pp. 123–166.

Cutler, A., Norris, D., 1988. The role of strong syllables in segmentation for lexical access. J. Exp. Psychol. Hum. Percept. Perform. 14, 113–121.

Cutler, A., Mehler, J., Norris, D., Segui, J., 1983. A language-specific comprehension strategy. Nature 304, 159–160.

Cutler, A., Mehler, J., Norris, D., Segui, J., 1986. The syllables differing role in the segmentation of French and English. J. Mem. Lang. 25, 385–400.

Dutoit, T., Pagel, N., Pierret, F., Bataille, O., van der Vreken, O., 1996. The MBROLA project: towards a set of high-quality speech synthesizers free of use for non-commercial purposes. 3, 1393–1396. Philadelphia. Conference Proceeding ICSLP'96.

Echols, C.H., Crowhurst, M.J., Childers, J.B., 1997. The perception of rhythmic units in speech by infants and adults. J. Mem. Lang. 36, 202–225.

Godey, B., Schwartz, D., de Graaf, J.B., Chauvel, P., Liegeois-Chauvel, C., 2001. Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. Clin. Neurophysiol. 112, 1850–1859.

Gomez, R.L., Gerken, L., 2000. Infant artificial language learning and language acquisition. Trends Cogn. Sci. 4, 178–186.

Hari, R., Pelizzone, M., Makela, J.P., Hallstrom, J., Leinonen, L., Lounasmaa, O.V., 1987. Neuromagnetic responses of the human auditory-cortex to onsets and offsets of noise bursts. Audiology 26, 31–43.

Jennings, J.R., Wood, C.C., 1976. Epsilon-adjustment procedure for repeated-measures analyses of variance. Psychophysiology 13, 277–278.

Johnson, E.K., Jusczyk, P.W., 2001. Word segmentation by 8-month-olds: when speech cues count more than statistics. J. Mem. Lang. 44, 548–567.

Jusczyk, P.W., Cutler, A., Redanz, N.J., 1993. Infants preference for the predominant stress patterns of English words. Child Dev. 64, 675–687.

Jusczyk, P.W., Houston, D.M., Newsome, M., 1999. The beginnings of word segmentation in English-learning infants. Cogn. Psychol. 39, 159–207.

Kutas, M., 1997. Views on how the electrical activity that the brain generates reflects the functions of different language structures. Psychophysiology 34, 383–398.

Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn. Sci. 4, 463–470.

Liegeois-Chauvel, C., Musolino, A., Badier, J.M., Marquis, P., Chauvel, P., 1994. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr. Clin. Neurophysiol. 92, 204–214.

Mattys, S.L., Jusczyk, P.W., Luce, P.A., Morgan, J.L., 1999. Phonotactic and prosodic effects on word segmentation in infants. Cogn. Psychol. 38, 465–494.

McCarthy, G., Wood, C.C., 1985. Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr. Clin. Neurophysiol. 62, 203–208.

Münte, T.F., Urbach, T.P., Düzel, E., Kutas, M., 2001. Event-related brain potentials in the study of human cognition and neuropsychology. In: Boller, F., Grafmann, J., Rizolatti, G. (Eds.), Handbook of Neuropsychology. Elsevier Science, Amsterdam, pp. 139–235.

Navarro Tomás, T., 1965. Manual de pronunciación española. Consejo Superior de Investigaciones Científicas, Madrid.

Newport, E.L., Aslin, R.N., 2004. Learning at a distance: I. Statistical learning of non-adjacent dependencies. Cogn. Psychol. 48, 127–162.

Norris, D., McQueen, J.M., Cutler, A., 2000. Merging information in speech recognition: feedback is never necessary. Behav. Brain Sci. 23, 299–325.

Osterhout, L., Bersick, M., McKinnon, R., 1997. Brain potentials elicited by words: word length and frequency predict the latency of an early negativity. Biol. Psychol. 46, 143–168.

Otake, T., Hatano, G., Cutler, A., Mehler, J., 1993. Mora or syllable—Speech segmentation in Japanese. J. Mem. Lang. 32, 258–278.

Pantev, C., Elbert, T., Ross, B., Eulitz, C., Terhardt, E., 1996. Binaural fusion and the representation of virtual pitch in the human auditory cortex. Hearing Res. 100, 164–170.

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

Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., Trujillo, N.J., 1999. Intracerebral sources of human auditory-evoked potentials. Audiol. Neuro-Otol. 4, 64–79.

Potts, G.F., Dien, J., Hartry-Speiser, A.L., McDougal, L.M., Tucker, D.M., 1998. Dense sensor array topography of the event-related potential to task-relevant auditory stimuli. Electroencephalogr. Clin. Neurophysiol. 106, 444–456.

Purdy, S.C., Kelly, A.S., Thorne, P.R., 2001. Auditory evoked potentials as measures of plasticity in humans. Audiol. Neuro-Otol. 6, 211–215.

Recanzone, G.H., Schreiner, C.E., Merzenich, M.M., 1993. Plasticity in the frequency representation of primary auditory-cortex following discrimination-training in adult owl monkeys. J. Neurosci. 13, 87–103.

Reinke, K.S., He, Y., Wang, C.H., Alain, C., 2003. Perceptual learning modulates sensory evoked response during vowel segregation. Cogn. Brain Res. 17, 781–791.

Saffran, J.R., 2001. Words in a sea of sounds: the output of infant statistical learning. Cognition 81, 149–169.

Saffran, J.R., Aslin, R.N., Newport, E.L., 1996a. Statistical learning by 8-month-old infants. Science 274, 1926–1928.

Saffran, J.R., Newport, E.L., Aslin, R.N., 1996b. Word segmentation: the role of distributional cues. J. Mem. Lang. 35, 606–621.

Sanders, L.D., Neville, H.J., 2003. An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. Cogn. Brain Res. 15, 228–240.

Sanders, L.D., Newport, E.L., Neville, H.J., 2002. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nat. Neurosci. 5, 700–703.

Scherg, M., 1990. Fundamentals of dipole source analysis. In: Grandori, F., Hoke, M., Roman, G.L. (Eds.), Auditory Evoked Magnetic Fields and Electric Potentials. . Advances in Audiology, vol. V. Karger, Basel, pp. 40–69.

Scherg, M., von Cramon, D., 1985. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalogr. Clin. Neurophysiol. 62, 32–44.

Scherg, M., von Cramon, D., 1986. Evoked dipole source potentials of the human auditory cortex. Electroencephalogr. Clin. Neurophysiol. 65, 344–360.

Sebastián-Gallés, N., Costa, A., 1997. Metrical information in speech segmentation in Spanish. Lang. Cogn. Processes 12, 883–887.

Shahin, A., Bosnyak, D.J., Trainor, L.J., Roberts, L.E., 2003. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J. Neurosci. 23, 5545–5552.

Suomi, K., McQueen, J.M., Cutler, A., 1997. Vowel harmony and speech segmentation in Finnish. J. Mem. Lang. 36, 422–444.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme Medical Publishers, New York.

Thiessen, E.D., Saffran, J.R., 2003. When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. Dev. Psychol. 39, 706–716.

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

Tremblay, K.L., Kraus, N., 2002. Auditory training induces asymmetrical changes in cortical neural activity. J. Speech Lang. Hear. Res. 45, 564–572.

Tremblay, K., Kraus, N., Mcgee, T., Ponton, C., Otis, B., 2001. Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. Ear Hear. 22, 79–90.

Urbach, T.P., Kutas, M., 2002. The intractability of scaling scalp distributions to infer neuroelectric sources. Psychophysiology 39, 791–808.

Vroomen, J., Tuomainen, J., de Gelder, B., 1998. The roles of word stress and vowel harmony in speech segmentation. J. Mem. Lang. 38, 133–149.