

## Time course and functional neuroanatomy of speech segmentation in adults

Toni Cunillera<sup>a</sup>, Estela Càmara<sup>b,c</sup>, Juan M. Toro<sup>a</sup>, Josep Marco-Pallares<sup>b</sup>, Nuria Sebastián-Galles<sup>d</sup>, Hector Ortiz<sup>e,f</sup>, Jesús Pujol<sup>e,f</sup>, Antoni Rodríguez-Fornells<sup>b,g,\*</sup>

<sup>a</sup> Department of Basic Psychology, Faculty of Psychology, University of Barcelona, 08035, Barcelona, Spain

<sup>b</sup> Department of Physiology, Faculty of Medicine, Campus de Bellvitge – IDIBELL, University of Barcelona, 08907, Barcelona, Spain

<sup>c</sup> Department of Neuropsychology, Otto-von Guericke University, 39106, Magdeburg, Germany

<sup>d</sup> Brain and Cognition Unit, Universitat Pompeu Fabra, 08018 Barcelona, Spain

<sup>e</sup> Institut d'Alta Tecnologia-PRBB, CRC Corporació Sanitària, Barcelona, Spain

<sup>f</sup> Networking Research Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Barcelona, Spain

<sup>g</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain

### ARTICLE INFO

#### Article history:

Received 21 January 2009

Revised 2 June 2009

Accepted 28 June 2009

Available online 4 July 2009

### ABSTRACT

The present investigation was devoted to unraveling the time-course and brain regions involved in speech segmentation, which is one of the first processes necessary for learning a new language in adults and infants. A specific brain electrical pattern resembling the N400 language component was identified as an indicator of speech segmentation of candidate words. This N400 trace was clearly elicited after a short exposure to the words of the new language and showed a decrease in amplitude with longer exposure. Two brain regions were observed to be active during this process: the posterior superior temporal gyrus and the superior part of the ventral premotor cortex. We interpret these findings as evidence for the existence of an auditory–motor interface that is responsible for isolating possible candidate words when learning a new language in adults.

© 2009 Elsevier Inc. All rights reserved.

### Introduction

The present study was focused on understanding one of the initial processes in language learning: speech segmentation. The difficulty of the task lies in the lack of reliable acoustic cues that indicate where a word begins and ends. Thus, unlike the blank spaces that appear between printed words, the spoken acoustic signal could be considered, in many respects, continuous. Listeners, therefore, must parse the speech signal in order to start learning the new language. Notice that this initial process of isolating words is mandatory for subsequent language processes; for example, associating the possible lexical trace with a specific meaning.

To segment the auditory stream into words, the learner could exploit different acoustical cues such as allophonic variation, stress patterns, prosody, or/and distributional cues such as phonotactic regularities and transitional probabilities of syllable combinations (see for a review Jusczyk et al. (1999)). In relation to the distributional cues, several models have proposed the existence of a powerful statistical learning mechanism in order to explain speech segmentation (Brent, 1999). Supporting this idea, several experiments have shown that listeners are able to exploit the distributional properties of the input, regardless of whether it consists of syllables, tones or visual

information (Fiser and Aslin, 2001; Saffran et al., 1999), and it likely proceeds in an incidental fashion (Saffran et al., 1997; Toro et al., 2005; Turk-Browne et al., 2005). Thus, statistical learning is understood as a domain-general mechanism that profits from the regularities of the environment to drive learning.

The computational implementation of this hypothesis has been successfully applied by connectionist models (Christiansen et al., 1998; Elman, 1990; for a different account see Brent, 1997). The underlying idea is that word boundaries are in locations where the transitional probabilities between two sounds are low. In other words, word boundaries can be inferred based on the fact that transitional probabilities are higher for word-internal than for word-external pairs of syllables. For instance, in the sentence “look, a balloon” the string “ba/lloon” is more likely to occur together across other sentences than the string “a/ba” as the latter string would not be heard in phrases such as “the balloon”, “the red balloon”, etc. In fact, it has been demonstrated in infants and adults that computing the transitional probabilities between syllables is sufficient for isolating new words that are embedded in an artificial continuous speech stream when no acoustical cues are available (Saffran et al., 1996a,b). Remarkably, this process is also accomplished by other animal species (Hauser et al., 2001; Toro and Trobalon, 2005), highlighting its generality.

Interestingly, an alternative computational proposal suggests that the formation of syllabic chunks might be the only process required to isolate possible words (Perruchet and Vinter, 1998). Accordingly to the authors, a chunk can be considered that brings together the elements that are at the attentional focus at a particular moment. In subsequent

\* Corresponding author. Department of Ciencias Fisiològiques, Faculty of Medicine, Campus de Bellvitge, IDIBELL, Feixa Llarga s/n, 08907. L'Hospitalet (Barcelona), Spain. Fax: +34 934024268.

E-mail address: [antoni.rodriguez@icrea.es](mailto:antoni.rodriguez@icrea.es) (A. Rodríguez-Fornells).

encounters with the initially formed chunks, those are forgotten or strengthened according to the associative memory laws. In this regard, recently [Pacton and Perruchet \(2008\)](#) have proposed a general associative learning model that asserts attention as the necessary and sufficient condition for associative learning to occur and in which is included the formation of linguistic chunks. From this perspective, statistical computations would come into play solely after chunks are selected, and only for reinforcing or weakening the formed chunks that are repeatedly selected or varied, respectively ([Perruchet and Pacton, 2006](#); [Perruchet and Vinter, 1998](#)).

During the last decade, the use of ERP measures has provided important data about the cognitive mechanisms involved in speech segmentation ([Cunillera et al., 2006, 2008](#); [Sanders et al., 2002](#); [Sanders and Neville, 2003a,b](#)). Specifically, the N100 and the N400 ERP components have been proposed as speech segmentation indexes ([Abla et al., 2008](#); [Cunillera et al., 2006](#); [Sanders et al., 2002](#)). The N100 seems to be sensitive to word onset perception, whereas the N400 appears to indicate the identification of recently segmented words ([Abla et al., 2008](#); [Cunillera et al., 2006](#); [Sanders et al., 2002](#)). In this vein, the N400 component has been related to the learning process of nonsense words and has been regarded as an indirect electrophysiological brain signal that indicates that speech segmentation has been accomplished. Recently, [Abla et al. \(2008\)](#) have explored the time-course of the on-line segmentation process using non-linguistic items. Specifically, participants listened to a continuous stream of tones composed by groups of tri-tones organized randomly in a continuous sequence. Participants were classified as high, middle and low learners as a function of their behavioral performance in a target recognition test that followed the segmentation phase. The authors measured ERP responses corresponding to the average signal recorded separately in three consecutive 6-min periods of the learning phase. For high learners, the results showed enhanced amplitudes of the N400 component in the early session, whereas for the group of middle learners the same enhancement was observed in a latter session. Low learners did not show any ERP effect along the exposure to the tones stream. The authors concluded that their results revealed the on-line statistical learning process and more important, that the N400 indicates the degree of on-line statistical learning (see also [Buiatti et al., 2009](#)).

In the neuroscientific literature no clear hypothesis has been raised about which brain mechanisms and regions might be involved in statistical learning and, particularly, while segmenting the speech signal into potential new words. Thus far, the only brain imaging experiment that has studied speech segmentation ([McNealy et al., 2006](#)) used functional Magnetic Resonance Imaging (fMRI) in order to investigate the neural correlates of speech segmentation. The authors showed that a left frontotemporal network was activated during the on-line speech segmentation process. In particular, while listening to the language streams composed of nonsense words, the activation of the superior temporal gyrus (STG) was found to increase over time correlating with an independent word discrimination task.

Several proposals in the context of language processing converge with the idea that an auditory–motor interface that directly links the auditory input with the motor representations of speech might support the decoding of speech ([Hickok and Poeppel, 2000, 2004, 2007](#); [Scott et al., 2000](#); [Scott and Wise, 2004](#); [Wise et al., 2001](#)). This auditory–motor interface might engage the left posterior temporal regions and the frontal regions, sustaining motor speech representations. Furthermore, it has been proposed that this dorsal circuit of speech processing comes into play when it is necessary to keep auditory representations in an active state during the performance of a task (i.e., phonological short-term memory, [Aboitiz et al., 2006](#); [Buchsbbaum et al., 2005](#); [Jacquemot and Scott, 2006](#); [Scott et al., 2006](#)). More important to the present study, [Hickok and Poeppel \(2007\)](#) have recently suggested that this dorsal network could also be involved in the acquisition of new vocabulary, that is, in language learning. This process might involve generating a new sensory representation of the

novel word by linking a sequence of syllables or segments while, at the same time, the newly created traces guide the motor articulatory sequences. This auditory–motor interface hypothesis is reminiscent of the early motor theory of speech perception ([Liberman and Mattingly, 1985](#)) and the hypothesis that word learning is mediated by the phonological working memory loop ([Baddeley et al., 1998](#)). Indeed, this auditory–motor interface is postulated to be very important in speech development because speaking inherently requires fine-tuned motor learning. This mechanism is proposed to be active in adults as well ([Hickok and Poeppel, 2000](#)).

### Objectives

The present study was devoted to exploring the time-course of the on-line process of speech segmentation. Thus, the combination of ERPs (Experiment 1) and fMRI (Experiment 2) allowed us first, to closely inspect the time required for isolating novel words from fluent speech, and in a second experiment to unravel the brain architecture underlying the detection of word boundaries. Based on previous findings, in the ERP experiment, we predicted a time-dependent modulation of the N100 and N400 components associated with the process of segmenting and, afterwards, recognizing the isolated words (see [Cunillera et al., 2006](#); [Sanders et al., 2002](#)).

Based on the hypothesis of an auditory–motor interface involved in word learning, in the fMRI experiment, we predicted that this left hemisphere auditory–motor network would be recruited during the speech segmentation process. The reasons are the following: (i) the process of speech segmentation involves keeping the incoming syllable information active, (ii) this information is required either to compute transitional probabilities ([Cunillera et al., 2006](#); [Saffran et al., 1996b](#)) or to create possible new word-templates or chunks ([Perruchet and Vinter, 1998](#)), and (iii) the information from the isolated word units might be feed-forwarded to the frontal cortex (premotor and motor regions) in order to create corresponding motor articulatory sequences. Most importantly, the activation of this network has to be *time-dependent*: increased activation should be observed only during the initial learning period (isolation of the word-like units) but not after the new, possible token has been identified (time-dependent hypothesis; [Poldrack, 2000](#); [Raichle et al., 1994](#)).

### Experiment 1

In the first experiment of this study, event-related brain potentials (ERPs) were used to investigate the time-course underlying speech segmentation in adult listeners who were exposed to language and random streams.

#### Methods

##### Participants

After giving informed consent, fifteen healthy adult (5 males) undergraduate Psychology students at the University of Barcelona with normal hearing and no neurological impairment participated in the experiment. All participants [mean age  $23.4 \pm 4.9$  (SD)] were right-handed and were Spanish–Catalan native speakers. The experiment was approved by the local ethical committee of the University of Barcelona. All participants were paid at the end of the experiment or received extra course credit for their participation.

##### Materials and procedure

Five language streams were developed with the same structure as those originally created by [Saffran et al. \(1996a\)](#). Each stream consisted of 4 different trisyllabic nonsense words (e.g., *pirutabagolitokudapirutagukibo...*) with an exact duration of 232 ms per syllable. Nonsense words were concatenated to form a text stream that was subsequently transformed into an acoustic stream using the MBROLA speech

synthesizing software (Dutoit et al., 1996; Dutoit and Gosselin, 1996), which is based on concatenation of diphones and includes a Spanish male diphone database at 16 kHz (available at <http://tcts.fpmo.ac.be/synthesis/mbrola.html>). Afterwards, the Cooledit software was used to equate the length of all the streams into millisecond resolution, which was necessary for the exact ERP triggering. Words were concatenated in a pseudo-random order with the restriction that there were no immediate repetitions of words. All phonemes had the same duration (116 ms) and pitch (200 Hz; equal pitch rise and fall, with pitch maximum at 50% of the phoneme) in the language streams. Thus, the flat acoustic stream that was created had no acoustic markers between words and, therefore, only statistical cues were available for segmenting the possible words. The transitional probability between the syllables forming a word was 1.0, while the transitional probability between syllables spanning word boundaries was 0.33. New syllables were used for every language stream (see appendix). Because there were not enough syllables for the 10 languages to be created, only one syllable was repeated in two streams. The stimuli were displayed using the Presentation 0.52 software (Neurobehavioral Systems). In 1 min of a language stream, each word was repeated 26 times. Based on a previous behavioral experiment, the exact duration of all streams was set to 8 min and 48 s in which the four words were repeated each one 208 times along the stream.

Five different auditory random streams were created as a Random condition (e.g., “tatopidalirubagutarugopikubo...”). The idea underlying this condition is that participants could not extract or segment any word because transitional probabilities could not be used in order to identify clear word boundaries. With this aim, the transitional probability across syllables was maintained equal; therefore, each syllable within the random stream had the same probability to be followed by any syllable of the corresponding language stream (the transitional probability across all syllables was always 0.09).

During the experiment, participants were required to listen carefully to the 10 streams of sounds and were told that their main task would be to discover the words of those “alien” languages. The first stream was always one of the languages that contained possible words. The presentation order of the streams was randomized for each participant with the only constrain that no more than two streams of each type (language or random) followed each other.

After exposure to each stream, a two alternative forced-choice (2AFC) behavioral test was administered to determine whether the participants were able to identify the words heard previously. No ERPs were recorded during this phase. The test comprised eight pairs of auditory test items (a word and a part-word) that were randomly presented. Part-words were made by the concatenation of the third syllable of a word and the first two syllables of another word (3-1-2 part-words; e.g., *rutaba*, *dapiru*, *bopiru*, *litoku*) or the last two syllables of a word and the first syllable of another word (2-3-1 part-words; e.g., *tabago*, *golito*, *kudagu*, *kibopi*). For random streams, test items were composed of 16 different trisyllabic groupings. 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 just heard. The presentation of the next test pair was not delivered until a response was produced. The order of presentation of words and part-words in the test pairs was balanced. Brief rests were allowed after each stream. The next stream began after listeners felt comfortable enough to maintain their attention for the next stream.

#### Data acquisition and analysis

ERPs were recorded from the scalp using tin electrodes mounted in an electrocap (Electro-Cap International) and located at 29 standard positions. Biosignals were rereferenced off-line to the mean of the activity at the two mastoid processes. Vertical eye movements were monitored with an electrode at the infraorbital ridge of the right eye. Electrode impedances were kept below 3 k $\Omega$ . The electrophysiological signals were filtered on-line with a bandpass of 0.01–50 Hz (half-

amplitude cutoffs) and digitalized at a rate of 250 Hz. Trials with base-to-peak electro-oculogram (EOG) amplitude of more than 50  $\mu$ V, amplifier saturation, or a baseline shift exceeding 200  $\mu$ V/s were automatically rejected off-line. ERPs were averaged for epochs of 1024 ms, starting 50 ms prior to the stimulus.

A mean amplitude measure of the 70–130 and 350–550 ms time windows were taken in order to encompass the N100 and the N400 component, respectively. These mean voltage measures were submitted to an omnibus repeated measures analysis of variance (ANOVA) including two within-subject factors: *Word/Non-word* (language vs. random streams) and 15 levels of *Electrode* (F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, and T6). This analysis was carried out on data that was corrected using the vector normalization procedure (Mccarthy and Wood, 1985). 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 corrected *P*-value is reported.

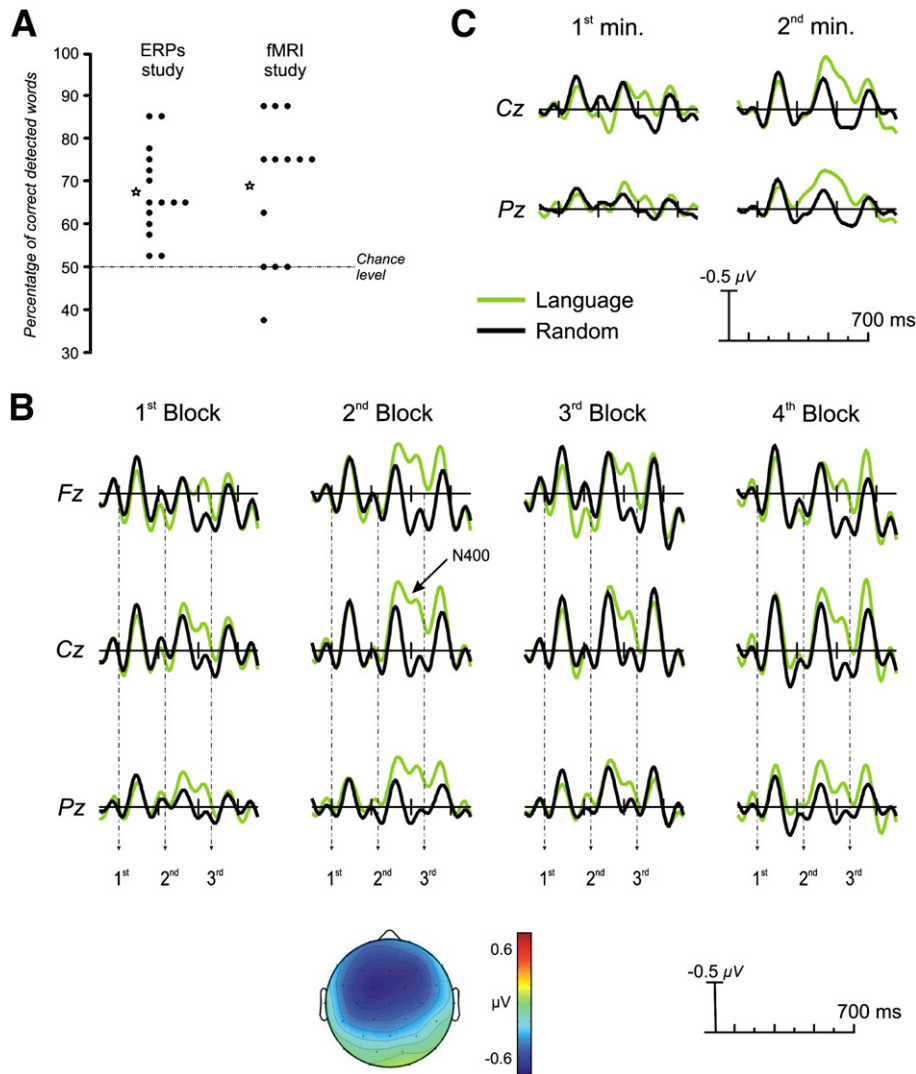
#### Results

After the presentation of each language stream, a behavioral test was performed in which participants were required to identify possible words against a set of part-words. The mean percentage of recognition of possible words was  $67.3 \pm 10.2\%$  (see Fig. 1A), which was significantly different from chance (one-sample *t*-test with chance level at 50%,  $t(14) = 6.58$ ,  $P < 0.001$ ). Thus, the listeners were able to discover possible words from the speech stream using only statistical cues.

Fig. 1B shows the ERP signature in each of the 2-minute blocks pooled across the five languages and the five random streams composing both conditions. Notice that the presentation of each syllable (one every 232 ms) elicited a clear evoked potential auditory pattern, showing the P50, N100, and P200 components. At the first block (first 2 min), a negative increase was observed in the 350–550 ms range when comparing the Language against the Random condition; however, this difference was not statistically significant ( $F(1,14) = 3.06$ ,  $P > 0.05$ ). This difference reached significant values in the second block (2–4 min;  $F(1,14) = 17.6$ ,  $P < 0.001$ ), and remained significant in the third and fourth block [third block (4–6 min): Word/Non-words,  $F(1,14) = 9.7$ ,  $P < 0.01$ ; fourth block (6–8 min): Word/Non-words,  $F(1,14) = 7.0$ ,  $P < 0.05$ ]. The observed N350–N550 component resembles the well-known N400 component, which has been associated with lexical-semantic processing (Kutas and Federmeier, 2000). No differences were found in the N100 component across blocks or when comparing language streams against random streams (all  $P > 0.1$ ).

Interestingly, in our study, the N400 component began to reduce its amplitude in the third and fourth blocks, although remained highly significant as it was expected based on the time-dependent hypothesis (see Table 1). At Fz location, a significant quadratic trend was encountered for the interaction between Word/Non-word and Block ( $F(1,14) = 4.84$ ;  $P = 0.045$ ). In order to further investigate whether the N400 component had already developed during the first 2 min, we decomposed the first block into two 1-minute averages (see Fig. 1C). As could be observed, a clear N400 component appeared in the second minute of exposure (Word/Non-words,  $F(1,14) = 8.6$ ,  $P < 0.01$ ), while it was not present in the first minute ( $F < 1$ ).

A subsequent behavioral experiment was conducted with 48 new participants in order to further investigate the performance of participants exposed to language streams of different durations (1, 2, 4, and 8 min, each language composed with different syllables) using a Latin-square design. In convergence with the previous ERP experiment, after just 1 min of exposure to the new language subjects were able to isolate its constituent words (segmentation rate  $69.6 \pm 19.6\%$ ;  $t(47) = 6.95$ ,  $P < 0.001$ ). Interestingly, performance was not improved as a function of time ( $F(3,141) < 1$ ,  $P > 0.5$ ; 1-minute:  $69.6 \pm 19.6\%$ , 2-minute:  $71.7 \pm 22.4\%$ , 4-minute:  $72.7 \pm 18.8\%$ , and 8-minute:



**Fig. 1.** (A) Percentage of correct segmented novel words in the behavioral test performed after the auditory presentation of streams for the ERP ( $n = 15$ ) and the fMRI experiment ( $n = 13$ ). Circles represent individual values and stars correspond to the mean values of each condition. (B) Grand average ERP potentials at midline electrode locations (frontal—Fz, central—Cz, and parietal—Pz) for the different Language (words) and Random (non-words) conditions for the four 2-minute blocks of exposure. Notice that words in the Language stream developed a N400 component that showed its maximum during the second block. A significant interaction appeared in this time window when comparing the first and the second block, showing the development of the N400 component in the second block. At the bottom, the scalp distribution (isovoltage spline interpolation) of this N400 component is depicted for the difference (waveform word minus non-word) in the second block. Notice the frontocentral distribution of this component. (C) The first block was further decomposed into two 1-minute blocks in order to observe the incremental pattern of the N400 component. As it is clear, during the first minute, no significant N400 word-non-word differences appeared. In contrast, in the second minute, a well developed N400 component showed a large difference between words (Language streams) and non-words (Random streams).

$74.9 \pm 19.7\%$ ; 1-minute vs. 8-minute:  $t(47) = -1.44$ ,  $P > 0.1$ ), reflecting a ceiling effect ( $\sim 70\%$ ) that was rapidly achieved.

## Discussion

The results of the ERP experiment and the complementary behavioral experiment clearly showed that after just a short exposure

**Table 1**  
Effects of Word/non-Word (language vs. random streams).

Block	Word/Non-word $F(1,14)$	Word/Non-word $\times$ Electrode $F(14,196)$
1st	3.06	1.42
2nd	<b>17.6<sup>+++</sup></b>	<b>6.65<sup>+++</sup></b>
3rd	<b>9.7<sup>++</sup></b>	<b>3.31<sup>++</sup></b>
4th	<b>7.0<sup>+</sup></b>	<b>3.37<sup>++</sup></b>

Results of the omnibus ANOVA for the N400 effect (Word/Non-words, Language vs. Random streams) at the time-window of 350–550 ms.

Notes. <sup>+</sup> $P < 0.05$ , <sup>++</sup> $P < 0.01$ , <sup>+++</sup> $P < 0.001$ .

15 electrodes: F3/4, Fz, F7/8, C3/4, Cz, T3/4, T5/6, Pz, P3/4.

(1 min) to the language stream speech segmentation was rapidly achieved. These results converge with those showing that infants succeed in segmenting words after only 2 min of exposure to similar language streams (Saffran et al., 1996a). Thus, as segmentation is an obligatory step in word learning, the N400 might reflect the initial protolexical trace that is created after isolating a given group of sounds (word candidates). Consequently, after accomplishing this process, listeners might begin to match the possible candidates with an already-existing word or a meaningful reference in the world. This hypothesis has been supported by a recent study in which 17-month-old infants preferred to associate a visual referent to a novel segmented word rather than to a part-word (Graf-Estes et al., 2007). Both results point to the idea that new segmented words are represented and stored as possible lexical candidates.

The involvement of the N400 in fast word learning has also been reported in other studies addressing different aspects of word learning (De Diego-Balaguer et al., 2007; McLaughlin et al., 2004; Mestres-Misse et al., 2007; Mueller et al., 2008; Perfetti et al., 2005). In infants' studies, a similar ERP negativity modulation with a frontal distribution



was found in 14-month-old infants performing a fast learning object-word mapping task (Friedrich and Friederici, 2008) and for known compared to unknown words in 19- to 22-month-old infants (Conboy and Mills, 2006). Likewise, in a recent ERP study conducted with 20-month-olds infants, Mills et al. (2005) observed a larger N400 to trained words paired previously with an object compared to untrained words.

Interestingly, in our study, the N400 component began to reduce in amplitude in the third and fourth blocks as was expected based on the time-dependent hypothesis, although it remained highly significant. This dynamic pattern of the N400 is similar to the one reported in Abla et al. (2008) for a group of high learners using streams composed by tones. In the present study, the inverted U-shape amplitude pattern might reflect the interplay between the initial and mandatory speech segmentation process (2–4 min of the learning phase) and the later recognition process of the already-segmented words. Repeated words develop a reduced N400 component (Rugg and Coles, 1995), which might explain why the N400 component tends to be reduced with longer exposure times. However, this hypothesis has to be confirmed in further experiments.

Our results concur with and extend the interesting study of Sanders et al. (2002) in which the comparison between trained vs. untrained words of an artificial language showed a posteriorly distributed N400 component. However, as shown in Fig. 1C, the scalp distribution of our N400 component was frontocentral. The differences observed in these experiments might be attributed to the differences in the design. In the present study, we specifically focused on the on-line learning process of word segmentation while in the work by Sanders et al., the N400 component most probably reflected the effect of word-recognition (of previously learned words) in a segmented language stream. In the same vein, the authors proposed that changes in the amplitude of the N100 component could index word segmentation. However, our results point to the fact that when speech segmentation is explored on-line (without previous training with the new possible words), the N100 seems to maintain its amplitude across time, and, most importantly, it does not seem to distinguish between a string of random ordered syllables and a unitary group of syllables.

It seems so that in adult and infant studies a negative polarity increase in the time range of 200–500 ms appears to be related to the word learning process. In the present study, the N400 modulation observed may reflect the learners' ability to extract the co-occurrence statistics found in the syllabic language streams. Our interpretation of the observed modulation of the N400 differs from the typical lexical-semantic interpretation associated with it. While, for the latter, the amplitude of the component decreases when lexical-semantic integration demands are reduced (e.g., with repetition or semantic contextual priming), the learning-related N400 shows the opposite pattern, with progressive amplitude enhancement as a function of increased exposure to the new-word (Mueller et al., 2008). In addition, the topographic distribution of this learning-related N400 component is more frontocentral, whereas the classical semantic N400 typically shows a right centro-parietal distribution (Kutas and Federmeier, 2000). These differences in the dynamics and topography of the N400 may indicate that these ERP modulations might not share the same cognitive processes and neural generators.

## Experiment 2

Using functional neuroimaging (fMRI) we evaluated the involvement of the dorsal auditory-motor network in speech segmentation (Hickok and Poeppel, 2007). An important aim of fMRI learning designs is to observe increases and decreases in activation over the training period (Poldrack, 2000). An increased activation of a brain region concomitant to learning is assumed to indicate the engagement of a particular cognitive process, or the development of new learning-related representations (Friston et al., 1992; Poldrack et al., 1998;

Raichle et al., 1994). Thus, a particular region might remain active until learning is accomplished and then decrease its activation (time-dependent hypothesis). Based on the previous behavioral and ERP results, the exposure to the languages and random streams we employed in Experiment 1 was limited to only two blocks of 2 min using a block fMRI design.

## Methods

### Participants

Thirteen new right-handed healthy native speakers of Spanish (mean age  $21.1 \pm 2.9$  years; 12 women) participated in this experiment. The experiment was approved by the local ethical committee of the University of Barcelona. Written consent was obtained from all participants. Participants were either paid or received extra course credit for their participation at the end of the experiment.

### Materials and procedure

We used four artificial languages, as described previously in Experiment 1. The structure was the same as that in Saffran et al. (1996a). Language and random streams were identical to those described in the ERP experiment (see Appendix), but only two language and two random streams were used in the present experiment. Each participant was exposed to one language and one random stream (different syllables formed the streams in each condition). Half of the subjects were exposed to Language A and Random B and the other half were exposed to Language B and Random A. Each functional run corresponded to one experimental condition and lasted 8 min. Within each run, eight 30 s blocks of rest (OFF) were interleaved with eight 30 s task blocks (ON). The order of presentation of the two experimental conditions was counterbalanced across participants.

We used the Presentation 0.52 software (Neurobehavioral Systems) for stimulus presentation. Auditory stimuli were delivered using MRI-compatible headphones with pneumatic sound transmission, which also attenuated the scanner noise. At the end of each stream, the two alternative forced-choice test (2AFC) was presented exactly as described in the ERP experiment.

Before the scanning protocol began, participants were instructed to close their eyes throughout the whole run and were encouraged to keep still and try to identify the words of the "alien" language that were going to be presented. Participants were also informed, before entering in the scanner, about the behavioral task presented at the end of each run. At the end of the session the participants were debriefed.

### Data acquisition and analysis

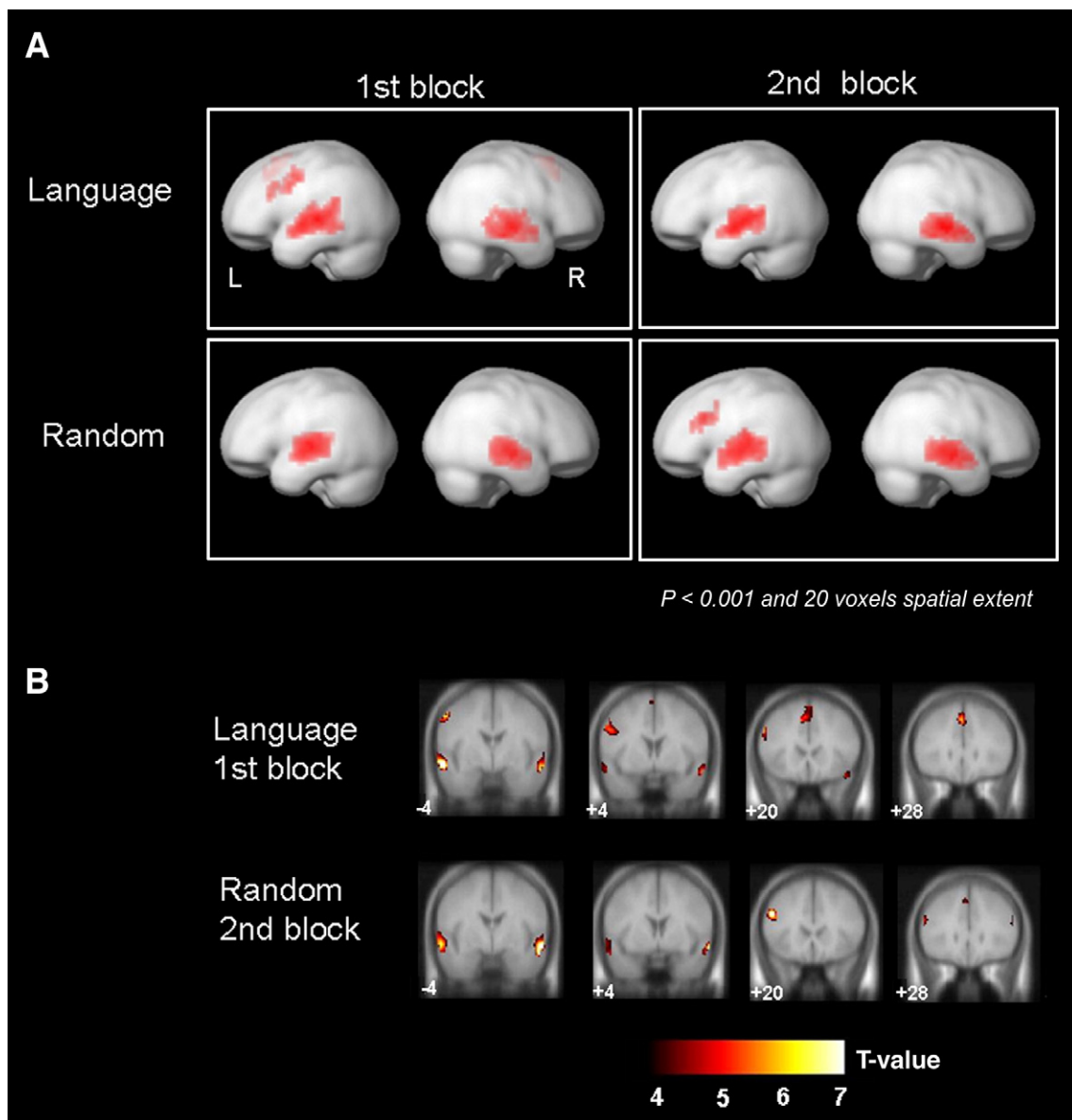
We collected fMRI images while the participants listened to 8 min of alternating periods of 30 s of an artificial language, separated with 30 s of rest periods. In a different run, a Random stream was presented using the same on-off distribution. We restricted the amount of exposure to the streams to 4 min (eight periods of 30 s) in accordance with the previous ERP results, which showed that the critical time window for speech segmentation occurs during the first 4 min.

Imaging was performed using a 1.5 T whole body MRI scanner (General Electric Cardio Vascular System) with standard head coil. A pillow and a restraining belt were used to minimize head movement. First, an axial anatomical localizer image was acquired covering the whole brain parallel to the anterior-posterior commissure plane using T1-weighted sequence (slice thickness = 5 mm; gap = 1 mm; number of slices = 20; repetition time (TR) = 460 ms; echo time (TE) = 14 ms; matrix =  $256 \times 192$ ; field of view FOV = 26 cm). Subsequently, functional images were obtained by using a single-shot T2\*-weighted gradient-echo EPI sequence (slice thickness = 5 mm; gap = 1 mm; number of slices = 20; TR = 3000 ms; TE = 50 ms; flip angle =  $90^\circ$ ; matrix =  $96 \times 64$ ; field of view FOV = 26 cm). Each functional run consisted of 160 sequential whole-brain acquisitions.

**Preprocessing.** Preprocessing steps were implemented using Statistical Parametric Mapping (SPM99, Friston et al., 1995a,b, 1998). First, for each participant, functional volumes were phase-shifted in time with reference to the first slice in order to minimize purely acquisition-dependent signal variations across slices. Second, head movement artifacts were corrected based on an affine rigid body transformation with reference to the first image of the first run. Third, structural and functional data were coregistered and the mean functional image was normalized to a standard stereotactic space using the EPI derived MNI template (ICBM 152, Montreal Neurological Institute) provided by SPM (Cocosco et al., 1997), using a 12-parameter affine transformation along with a nonlinear transformation using cosine basis functions. Functional EPI volumes were resampled into 4 mm cubic voxels and then spatially smoothed with an 8 mm full-

width half-maximum (FWHM) isotropic Gaussian kernel to accommodate residual anatomical differences across participants and to increase the validity of the parametric statistical test (increasing the fit of the data to a Gaussian field model).

**Data analysis.** The statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations after normalization for global effect by proportional scaling (Friston et al., 1995a,b, 1998; Worsley and Friston, 1995). The design matrix was generated with a boxcar function that included a convolved hemodynamic response function, and was temporally smoothed by convolution with a Gaussian kernel of dispersion of 4 s FWHM. Each block comprised the four ON and four resting (OFF) blocks. Contrast images were calculated for each



**Fig. 2.** (A) Rendered views in standard stereotactic space of the group average comparisons (ON–OFF periods) between the Language and Random conditions at each block for Left (L) and Right (R) hemispheres (for all activated regions,  $P < 0.001$ , corrected for spatial extent, 20 voxels cluster). Notice that during the first 2 min (first block), when the speech segmentation process is occurring (see previous ERP figure), an increase of activation was noticeable in the posterior superior temporal region, middle frontal gyrus, and ventral premotor area (BA 6) in the Language condition (compare these activations with the Random and Language conditions in the second block; see Tables 2–3). (B) Coronal views showing the spread of the posterior–anterior activation gradient in the left prefrontal cortex of the Language (first block) and Random (second block) conditions. Notice that in the first Language block, a more posterior region is recruited (ventral premotor, BA6, slices  $y = -4$  and  $y = 4$ ) and this same region is not recruited in the Random condition. However, in the Random condition, a clear activation is shown in the middle frontal gyrus (slices  $y = +20$  and  $y = +28$ ), which is also active in the Language condition (first block). These coronal views presented were superimposed on the canonical T1 MNI template. Values in the color scales refer to the  $t$  values of the corresponding contrast.

**Table 2**

Brain regions showing changes in activity comparing language/random vs. rest in the first block.

Brain region	~BA	n. voxels	Stereotactic coordinates				P-value
			x	y	z	T peak	
<i>Language vs. rest</i>							
<b>R PT (STG)</b>	21/22	244	56	-20	-8	14.9	<0.001
pSTG	22		64	-44	-4	7.8	<0.001
MTG	21		60	-8	-20	6.9	<0.001
<b>L PT (STS)</b>	21/22	282	-64	-24	0	12.4	<0.001
STS	21/22		-48	-28	0	10.8	<0.001
MTG	21		-56	-8	-8	9.6	<0.001
pSTG	22		-60	-40	-4	7.2	<0.001
			-64	-44	24		
<b>L Precentral G (PMC)</b>	6	50	-56	0	44	6.8	<0.001
PMC	6		-52	-4	48	6.4	<0.001
Inferior frontal sulcus	44/9		-52	20	32	6.1	<0.001
PMC	6		-60	0	36	5.6	<0.001
IFG	44		-52	4	32	5.8	<0.001
<b>SMA/anterior cingulate</b>	6/8	51	0	28	44	6.0	<0.001
	6		0	16	56	4.8	<0.001
	6		0	8	60	4.7	<0.001
<i>Random vs. rest</i>							
<b>R PT (STS)</b>	21/22	190	60	-20	-8	16.1	<0.001
MTG	21		56	-4	-16	8.3	<0.001
Posterior MTG	21		64	-36	-12	6.8	<0.001
<b>L PT (STS)</b>	21/22	207	-64	-32	4	9.7	<0.001
MTG	21		-52	-16	-12	8.7	<0.001
pSTG	22		-56	-40	4	5.5	<0.001
STG	22		-44	-36	8	4.5	<0.001

MNI coordinates and T value for the peak location in a particular identified anatomical cluster ( $P < 0.001$ ; 20 voxels spatial extent) for the statistically significant differences of the corresponding activated regions. Reported also the P-value for the peak of activation and the number of voxels in each cluster (n. voxels). BA=approximate Brodman's area; L=Left hemisphere, R=Right hemisphere, IFG=inferior frontal gyrus; MFG=middle frontal gyrus; IPL=inferior parietal lobe; STG=superior temporal gyrus; pSTG=posterior superior temporal gyrus; STS=superior temporal sulcus; PMC=Premotor cortex; PT=Planum Temporale; G=Gyrus.

participant between the two conditions (Language and Random) and the baseline condition through the block (the complete 8 blocks) and within blocks (the four first blocks and the four ending blocks). The individual contrast images were entered into a second-level analysis using a one-sample t-test. Unless mentioned otherwise, the threshold of the contrasts was  $P < 0.001$ , and only clusters with a  $P < 0.001$  (corrected for multiple comparisons) were reported (Worsley and Friston, 1995). The maxima of suprathreshold regions were localized by rendering them onto normalized T1 structural

**Table 3**

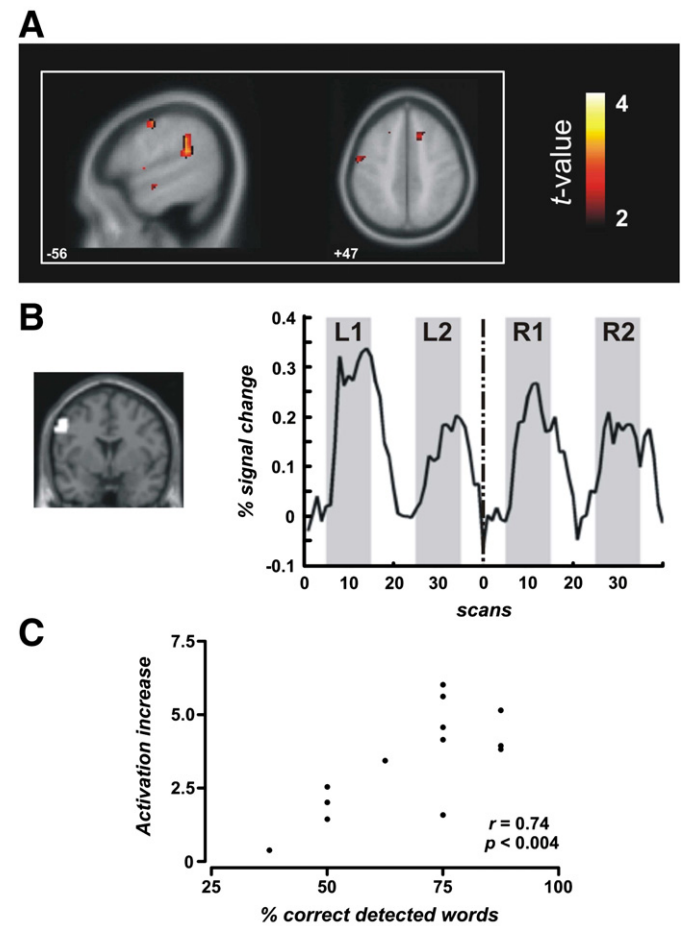
Brain regions showing changes in activity for Language / Random vs. Rest contrasts in the second block.

Brain region	~BA	n. voxels	Stereotactic coordinates				P-value
			x	y	z	T peak	
<i>Language vs. rest</i>							
<b>R PT (STS)</b>	21/22	233	52	-20	-4	12.7	<0.001
MTG	21		64	-32	-4	8.1	<0.001
MTG	21		64	-28	-16	6.6	<0.001
<b>L PT (STG)</b>	21/22	258	-60	-20	0	13.9	<0.001
STS	21/22		-48	-28	4	7.5	<0.001
pSTG	22		-64	-36	8	6.1	<0.001
<i>Random vs. rest</i>							
<b>R PT (STS)</b>	21/22	239	60	-20	-8	11.5	<0.001
<b>L PT (STS)</b>	22	251	-64	-24	8	11.7	<0.001
pSTG	22		-64	-36	-4	6.9	<0.001
MTG	21		-60	-4	-8	6.6	<0.001
STS	21/22		-48	-36	4	5.1	<0.001
<b>L MFG (inferior frontal sulcus)</b>	9/44	44	-48	20	28	8.9	<0.001
MFG	46		-52	32	24	6.0	<0.001
IFG	44		-48	12	28	5.8	<0.001

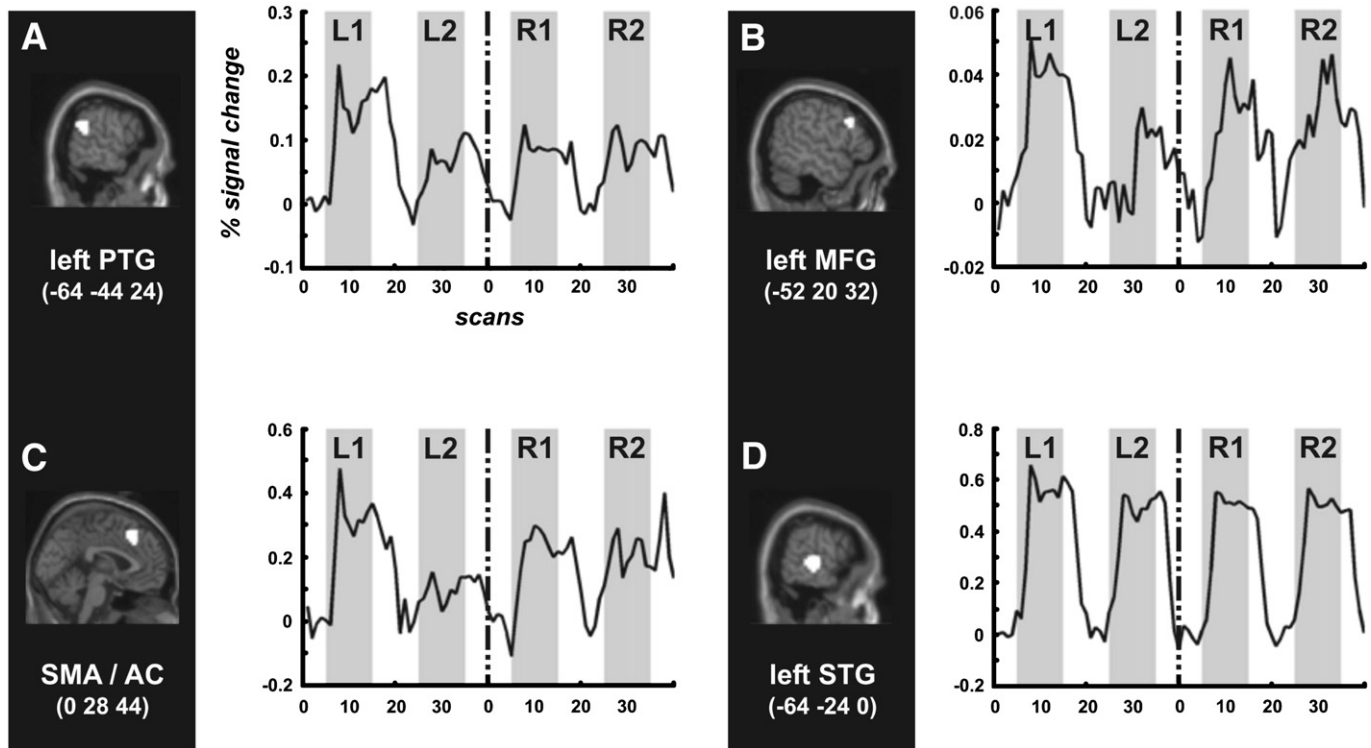
images on the MNI reference brain (Cocosco et al., 1997). Maxima and all coordinates are reported in MNI coordinates, and labeled following the Talairach atlas.

**Results**

The mean percentage of segmented words was  $68.3 \pm 16.6\%$ , which was different from chance [(50%),  $t(12) = 3.96$ ,  $P < 0.01$ ], and indicates that participants were able to segment the language stream (see Fig. 1A). The fMRI analysis was divided into two blocks of 2 min each. Main activation contrasts (Language vs. Rest and Random vs. Rest) in both blocks are depicted in Fig. 2A (see activated regions in Tables 2 and 3). During the first block, language segmentation clearly activated the left superior temporal cortex, extending to the posterior temporal part (pSTG). Furthermore, left superior ventral premotor cortex activation (svPMC, BA6) was observed in this contrast, which also comprised part of the inferior frontal cortex (opercular BA 44), and was not present in the Random vs. Rest contrast (see Fig. 2B). During the second block, these regions did not show differential activation when compared to the rest conditions. This idea was confirmed studying the interaction term [(Language 1st block > Language 2nd block) -



**Fig. 3.** (A) Sagittal and axial views showing the interaction term [(Language 1st block > Language 2nd block) - (Random 1st block > Random 2nd block)]. Both regions, the svPMC and the pSTG showed larger activation in the language condition during the first block (PMC, x, y, z = -56, -8, 5; pSTG, -52, -44, 20). (B) Hemodynamic responses for the region of interest reconstructed in the left ventral premotor cortex (PMC, BA 6, center -56, 0, 44); L1 = Language first block; L2 = Language second block; R1 = Random first block; R2 = Random second block. Notice the larger increase in activation during the first block of exposure to the language. (C) Brain behavior correlation analysis showing the correlation between the behavioral performance (percentage of words recognized in the test-phase) across subjects and the activation increase (individual T values) in the ventral PMC (peak -56, 0, 44) in the first block (2 min) of Language exposure.



**Fig. 4.** Hemodynamic responses for four selected regions of interest. L1 = Language first block; L2 = Language second block; R1 = Random first block; R2 = Random second block. Regions of interest (left to right): (A) left pSTG (BA22, center coordinates  $-64, -44, 24$ ), (B) left anterior MFG (BA 45/46, center  $-52, 20, 32$ ), (C) supplementary motor area (SMA, BA6/8, center  $0, 28, 44$ ); (D) left middle temporal gyrus (BA 21, center,  $-64, -24, 0$ ), depicted as a control non-task related area (no differences across blocks or conditions). Notice the different pattern of activation in the first block for the Language condition in the pSTG when compared to the other condition and block. Horizontal axis: time [repetition time (TR) units]. Vertical axis: mean percentage signal change.

(Random 1st block > Random 2nd block)]. As it is shown in Fig. 3A, both regions the svPMC and the pSTG were significant (PMC,  $x, y, z = -56, -8, 5$ ;  $t$ -value = 2.9,  $P < 0.007$ ; pSTG, coordinates  $-52, -44, 20$ ;  $t$ -value = 3.4,  $P < 0.003$ ), showing larger activation in the language condition during the first block. Furthermore, the activation in the svPMC was clearly observed in the direct contrast between Language and Random conditions [(Language – Rest) – (Random – Rest)], but only in the first block of exposure [peak coordinate: ( $-44, 8, 52$ );  $t = 3.35$ ,  $P < 0.02$ ].

*Region-of-interest (ROI) analysis* was performed in order to confirm these language segmentation-related findings. All ROI were functionally defined in the first block based on the peak activations in the second level of the group activation maps found in the Language vs. Rest contrast (Table 2). These regions covered the left posterior superior temporal gyrus [BA 22, peak center coordinates ( $-64, -44, 24$ )], left superior ventral premotor region [BA 6, coordinates ( $-56, 0, 44$ )], left anterior middle frontal gyrus [BA 45/46, center coordinates ( $-52, 20, 32$ )], supplementary motor area [SMA, BA6/8, center coordinates ( $0, 28, 44$ )] and left middle temporal gyrus (BA 21) [center coordinates ( $-64, -24, 0$ )].

The hemodynamic response function (HRF) was estimated in the five selected regions by plotting the percent signal change for each

condition and block. In order to draw statistical differences in BOLD signal changes random effects repeated measures analysis of variance were conducted between Language and Random condition in each block and defined region. Specifically, ROI averaged time series were baseline corrected for each condition and then averaged across participants. Afterwards, the HRF was estimated in each task period for the (ON) and (OFF) condition.

Indeed, as is clearly shown in Figs. 3B and 4 (see the statistical results in Table 4), a significant interaction was found between Block and Type of stream (Language vs. Random) for the svPMC (Fig. 3B) and the pSTG (Fig. 4A), indicating that the activation in the svPMC and the pSTG was larger in the first block of exposure when compared to the pattern observed in the Random condition. The interaction Block  $\times$  Type of stream was not significant for the supplementary motor area (SMA) or the left middle frontal gyrus (LMFG) (Figs. 4B, C). No differences were observed in the control non-task related region, the left middle temporal gyrus, which showed no modulation for the experimental conditions across blocks (Fig. 4D).

The crucial role of the superior ventral PMC in speech segmentation during the first block was also supported by the strong correlation encountered between the activation of this region in the first block and the percentage of segmented words at the end of the exposure to

**Table 4**  
Brain regions showing the alternating pattern of task-rest blocks changes in activity for Language / Random vs. Rest contrasts.

	<i>df.</i>	L PMC <i>F=, P&lt;</i>	L pSTG <i>F=, P&lt;</i>	R pSTG <i>F=, P&lt;</i>	Anterior cingulate <i>F=, P&lt;</i>	L MFG <i>F=, P&lt;</i>	L STG <i>F=, P&lt;</i>
Type	1,12	0.6, 0.430	3.9, 0.069	2.1, 0.177	0.4, 0.533	0.3, 0.617	1.2, 0.294
Block	1,12	0.6, 0.107	3.9, 0.070	7.2, 0.020	2.4, 0.144	0.7, 0.426	0.3, 0.580
Type $\times$ Block	1,12	4.9, 0.045	20.2, 0.001	0.9, 0.359	4.4, 0.058	2.9, 0.111	2.9, 0.114
Type $\times$ time( <i>t</i> )	19,228	1.8, 0.124	3.4, 0.006	0.9, 0.504	1.1, 0.375	0.5, 0.766	1.2, 0.332
Block $\times$ time	19,228	1.9, 0.088	2.3, 0.042	2.7, 0.027	2.7, 0.029	1.3, 0.243	1.5, 0.197
Type $\times$ block $\times$ time	19,228	1.1, 0.385	2.8, 0.014	0.5, 0.847	1.2, 0.298	1.9, 0.100	0.7, 0.641



the language streams ( $r = 0.74$ ,  $P < 0.004$ , see Fig. 3C). This correlation was not significant for either the second-language block ( $r = 0.4$ ,  $P > 0.1$ ) or for the same computations performed for the pSTG in both the first and the second block.

#### Source reconstruction analysis of the ERP data

In order to study the possible cortical sources involved in the generation of the N400 effect observed for the language conditions in Experiment 1, we applied a source reconstruction analysis in the second block, where this effect was observed to be maximum. The difference waveform (word minus non-word) was analyzed using the mean value of the grand average at the 350–550 ms time window. Computations were performed under SPM 8. First, we coregistered the 29 electrodes in a standard MNI head scalp. After that, we computed the forward model in a 5294 vertices by using a three shell spherical model (Ary et al., 1981). Finally, we applied the Multiple Sparse Priors method (Friston et al., 2008) to compute sources responsible for the generation of the N400 evoked potential. This model automatically selects between either sparse or distributed model on the bases of a Bayesian approach.

The results of the source reconstruction analysis showed very similar results to the ones observed in the fMRI study. Possible neural sources were observed at the premotor regions and also the superior temporal gyrus, extending to posterior STG and inferior parietal cortex. This pattern of activation was larger in the left hemisphere (see Fig. 5).

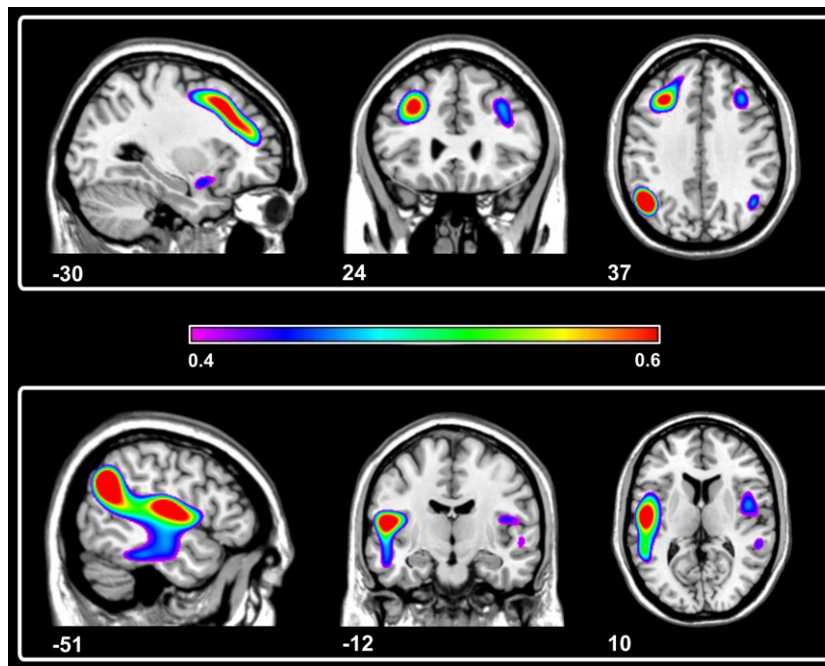
#### Discussion

The brain network related to successful word segmentation identified in the present study comprises the pSTG and the superior ventral PMC. This auditory–motor network might allow listeners of a new language to segment discrete sound units or possible “words” from a continuous speech signal. The posterior superior temporal gyrus (pSTG) has been related to phonological processing (Burton et al., 2000; Celsis et al., 1999; Heim and Friederici, 2003; Hickok and Poeppel, 2004; LoCasto et al., 2004; Zatorre et al., 1996). This region has been found to be also activated when processing phonetic stimuli

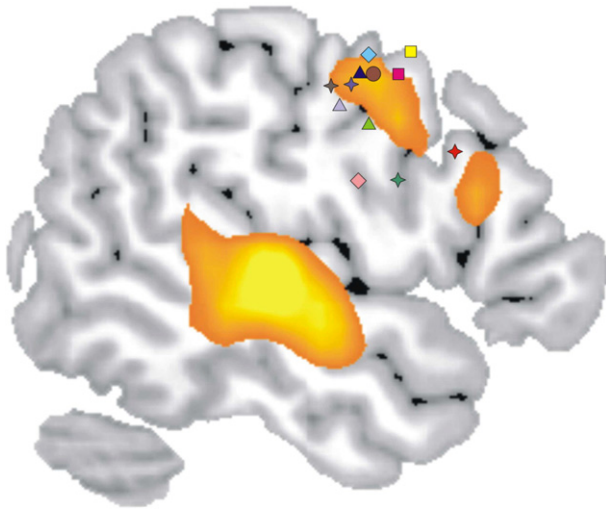
that are not necessarily intelligible (Scott et al., 2000). Therefore, it seems that upon hearing a word or a syllable, the mapping of acoustic properties into phonemes is carried out in the left middle and posterior STG, a region which has been involved in transiently storing phonological information of perceived words (Wise et al., 2001).

The exact role of the inferior frontal gyrus in phonological processing is still under debate (Bookheimer, 2002). However, several studies have observed activation in the posterior superior region of the IFG (BA44/6), extending to the PMC and, anteriorly, to the middle frontal gyrus in phonological processing (Burton et al., 2000; Demonet et al., 1992; Fiez et al., 1995; Heim and Friederici, 2003; Poldrack et al., 1999; Zatorre et al., 1992, 1996). The role of the PMC in speech perception has also been evidenced by various functional neuroimaging studies and single-pulse transcranial magnetic stimulation (TMS) studies, which have shown that this region as well as the motor cortex is activated in passive speech perception (Fadiga et al., 2002; Meister et al., 2007; Watkins et al., 2003; Wilson et al., 2004). For the sake of comparison with the premotor region observed in the present study, in Fig. 6 we have depicted the activations in the PMC reported in various studies. These studies are related to artificial language learning (Berns et al., 1997; Tettamanti et al., 2002), speech segmentation (McNealy et al., 2006), phonemic and speech discrimination (Burton et al., 2000; Meister et al., 2007; Zatorre et al., 1992, 1996), new-word learning (Mestres-Misse et al., 2008, in press; Wong et al., 2007), and passive listening (Pulvermuller et al., 2006; Wilson and Iacoboni, 2006; Wilson et al., 2004). The activation in the PMC region observed in the present study (see Fig. 6) clearly intersects with the regions reported in these language related studies. These phonological activations contrast with those elicited in the left IFG for semantic processing, which encompass anterior IFG regions (at the border of BA45 and 47, Poldrack et al., 1999).

The current pattern of activations is very similar to that found by McNealy et al. (2006) with an analogous artificial language segmentation task. Taking together the results of both studies, speech segmentation seems to be clearly mediated by a frontotemporal network. The main difference between these studies is the fact that in the McNealy et al. study, the participants were not able to recognize the nonsense words in a subsequent behavioral test. In contrast, in our



**Fig. 5.** Possible neural sources of activation of the grand ERP, word minus non-word conditions between 350 and 550 ms in the second block (2–4 min). At the coronal slice it is possible to observe bilateral activation of the premotor cortex, superior frontal gyrus and left posterior temporal and inferior parietal gyrus. Notice the higher activation of the left hemisphere compared to the right one. Units of the colorbar are proportional to  $\text{mV}^2$ .



**Fig. 6.** Review of studies in which PMC activation is reported in relation to speech processing, specifically dealing with artificial language learning (square), phonemic and speech discrimination (stars), speech segmentation (circle), new-word learning (rhombus), and passive listening (triangles). All peaks of activations were placed in the MNI stereotactic space and were projected using MRICron on a sagittal MNI single-subject slice ( $X = -52$ ). Peaks of activation depicted coming from the studies of **artificial language learning:** ■ Berns et al., 1997; ■ Tettamanti et al., 2002; **phonemic and speech discrimination:** ★ Burton et al., 2000; ○ Meister et al., 2007; ○ Zatorre et al., 1992; ○ Zatorre et al., 1996; **speech segmentation:** ● McNealy et al., 2006; **new-word learning:** ◆ Mestres-Misse et al., 2008; ◆ Wong et al., 2007; and **passive listening:** ▲ Pulvermuller et al., 2006; ▲ Wilson and Iacoboni, 2006; ▲ Wilson et al., 2004. Activation of the contrast Language (first + second block) vs. Rest ( $P < 0.001$ , spatial extent,  $n = 20$  voxels, uncorrected) is overlaid on the same slice for the sake of comparison with the peaks observed in the other studies.

study, participants were able to segment the speech stream as determined by the behavioral test and, moreover, segmentation performance correlated with the activity in the PMC. It is likely, however, that the differences in the behavioral results arose mainly due to the different type of test administered to participants in both studies. McNealy et al. employed a word discrimination task in which in each test trial participants listened to a word or a part-word and responded yes or no as to whether they considered each trisyllabic combination a word in the artificial language stream. In contrast, in our study we used a forced-choice test in which a pair composed by a word and a part-word was delivered to participants in each test trial. Noteworthy, with a forced-choice paradigm, it is not possible to discern whether a target (a word) is recognized or on the contrary, whether is the non-target (a part-word) which is rejected.

Our pattern of results converges with the idea that transient representations of the sequence of sounds of potential words are formed in the pSTG during auditory perception of syllabic streams. As predicted, this region might be acting as an auditory–motor interface (Hickok and Poeppel, 2004) that transmits the segmented sequence of isolated sounds as potential pronounceable words to the PMC (BA6/superior–posterior IFG). It has also been proposed (Warren et al., 2005) that a template-matching algorithm in the pSTG might allow for the detection of coincidences between the stored auditory memory representations or “phonological templates” derived from previous exposures and the new incoming stimuli. The output of this process should be an ordered sequence of auditory representations that would be forwarded to the PMC. In the premotor areas, the encoded sequence of sounds might be mapped into a sequence of articulatory gestures, which would keep the segmented words active through a rehearsal mechanism. In fact, the PMC has been associated with the active maintenance of non-meaningful verbal representations through the articulatory subvocal rehearsal component of the phonological working memory (Paulesu et al., 1993; Smith and Jonides, 1998). This rehearsal mechanism might be crucial in the first

stages of learning unfamiliar new words (Baddeley et al., 1998), reflecting the involvement of the phonological working memory (rehearsal component) while segmenting possible words based only on the computation of transitional probabilities.

More importantly, this idea is clearly supported in the present study by the strong correlation observed between the PMC activation in the first block and the percentage of words recognized at the end of the learning phase (see Fig. 4B). Our speech segmentation task demanded the tracking of the predictability of upcoming attended stimuli (as possible words are repeated constantly and a specific syllable in a word is always followed by another specific syllable) and the premotor region may be involved in keeping this process active. In fact, the premotor cortex responds to the prediction of auditory events in a structured sequence (Schubotz et al., 2003). As reciprocal connections exist between the PMC and the pSTG, the PMC might be acting as a “top-down” mechanism that regulates the template-matching process engaged in the pSTG. In fact, direct anatomical connections between the PMC and the posterior superior temporal regions exist through the arcuate fasciculus (Catani et al., 2005; Hackett et al., 1999; Romanski et al., 1999).

Noteworthy, in the Language condition the activation of the left PMC diminished after the first 2 min of exposition to the language stream, which closely correspond to the time when we began to observe in the ERP experiment a significant amplitude enhancement of the N400 component. A possible explanation, in regard to the time-dependent hypothesis, could be that words are rapidly segmented out and afterwards repeated exposure to those words caused the activation in the PMC to decrease. An alternative explanation would be that the auditory–motor interface acts as an initial word learning neural circuit which would be recruited only in the initial stages of speech segmentation.

It is worth mentioning the convergence observed in the present study between the neural network identified using fMRI and the one provided from the ERPs using a new source reconstruction method (Friston et al., 2008). Friston et al. (2008) have proposed that the multiple sparse priors method, which uses no a priori constraints, presented more than the 90% of explained variance of synthetic and real data, and showed better spatial accuracy and less localization error than two comparative methods (weighted minimum norm and realistic model with coherence prior). In addition, sources found in the analysis of real data (e.g., face processing) were located in a more plausible ventral occipito-parietal physiological locations than the ones obtained with the other two methods studied. Using this source reconstruction method we showed as possible neural sources the dorsal auditory–motor network, very similar to the observed fMRI pattern. Although the complexity of linking electrophysiological and hemodynamic responses is inherent (Friston et al., 2008; Kilner et al., 2005; Logothetis, 2002), the high degree of convergence between both results in the present study further support our interpretation about the involvement of the auditory–motor network in speech segmentation.

Previous studies using source reconstruction in MEG experiments have observed a fairly broad distribution of the scalp-recorded N400 component. Possible sources has been detected at the superior temporal cortex for spoken (Helenius et al., 2002; Kwon et al., 2005) or printed sentences (Helenius et al., 1998), including the middle temporal gyrus and the temporoparietal junction, being these sources more consistent in the left hemisphere. In a similar vein, using distributed source modelling methods in MEG, Halgren et al. (2002) showed the time-course of several N400 effects, evolving from the posterior half of the left STG (at 250 ms), and spreading forward and ventrally to the temporal lobe (at 365 ms). Afterwards, during the N400 time window large activation was observed in the frontal lobe bilaterally, especially in the incongruence conditions. Similar results using the N400 priming effect have showed activation in the left STG and the left IFG (Marinkovic et al., 2003). Overall, the neuropsychological, intracranial, and MEG results converge to suggest that the neural generators of the N400 component

are located in a large portion of the temporal lobe as well as the inferior prefrontal cortex (Halgren et al., 1994; Nobre et al., 1994; see also van Petten and Luka, 2006).

Considering these N400 related studies, the source reconstruction solution encountered in the present study seems very similar, pointing to the involvement of the superior temporal gyrus. However, the distribution of the N400 effect modelled (difference between non-word and words) was more frontal than the standard parietocentral distribution of the N400 component. This topographical difference might be explained by the large involvement of the premotor regions and probably more anterior prefrontal regions. The frontocentral N400 component encountered in our speech segmentation experiment is in agreement with the idea of the involvement of the inferior frontal gyrus in pseudoword processing (Clark and Wagner, 2003). Activation in the precentral and inferior frontal gyrus has also been observed when learning new words from congruent and incongruent contexts (Mestres-Misse et al., 2008, in press), in phonological new-word learning (Gronholm et al., 2005), in second-language learning of novel pitch patterns of words (Wong et al., 2007) and statistical segmentation of tone sequences (using near-infrared spectroscopy, Ablá and Okanoya, 2008). Similar frontal shifts in the N400 component have been observed in other speech segmentation tasks using ERPs (see De Diego-Balaguer et al., 2007; Ablá et al., 2008). This involvement of precentral and inferior frontal regions in word learning might explain the different topography observed in several studies for the N400 in new learned words. It would be interesting to investigate in follow-up EEG–fMRI combined recordings, in which degree the neural hemodynamic activations could be used as spatial priors to constrain equivalent dipole or distributed estimates of the N400 component observed during speech segmentation (Dale et al., 2000).

All in all, the auditory–motor brain network identified for speech segmentation might be essential for second-language learning and, most probably, for language acquisition in infants (Doupe and Kuhl, 1999; Warren et al., 2005). Language perception and production in the developing infant brain requires a specific tuning to the language sounds encountered during the first year of life. First words imitated by a child are guided by the “gestural” features of the sound, i.e. by the actions of the mouth rather than by a sound’s acoustic features (Studdert-Kennedy, 1987). Because perception and production of sounds are so closely tied from the first years of life, it is possible that the dorsal auditory perception stream might play an important role during language acquisition (Hickok and Poeppel, 2007). This dorsal stream might recruit the posterior superior temporal regions for encoding and storing sequences of sounds and acting as a sensory–motor interface via the participation of the PMC. Importantly, the identified pathway must also be related to the brain network subserving imitation of simple movements (Iacoboni et al., 1999). In fact, the ability to mimic sounds is essential for learning a new language. This idea has been revitalized by the discovery of mirror neurons, recorded in macaques in the homologue of the ventral PMC region (including the superior part of Broca’s region) and in humans (Fadiga and Craighero, 2006; Rizzolatti and Arbib, 1998). These specific audiovisual mirror neurons discharge not only when performing and observing a specific action, but also when hearing a specific sound representative of the observed action (Kohler et al., 2002). Mirror neurons also provide a mechanism for integrating perception and action at the neuronal level, which, at the same time, might contribute to various developmental processes such as the imitative behavior of infants, the necessity to integrate perceived and performed actions (Meltzoff and Decety, 2003) and communicative acts (Rizzolatti and Arbib, 1998).

## Conclusions

Recently, much progress has been made in the identification of the brain mechanisms involved in language learning, but much still remains to be done (Gullberg and Indefrey, 2006; Kuhl, 2004). In the

present study we investigated the time-course and brain regions involved in segmenting a new language. The brain network involved in this process recruited, selectively, the pSTG gyrus and the superior part of ventral PMC (superior IFG). After a short exposure to words of a new language, the brain appears to elicit a protolinguistic trace, evidenced by the appearance of the N400. As speech segmentation is pivotal at the early stages of language learning, the described auditory–motor network for speech segmentation might be essential for second-language learning and for language acquisition in infants. Further studies will be needed to disentangle the exact role of the different brain structures previously described, how they interact, and how exactly they contribute to statistical learning, speech segmentation, and language learning in general.

## Acknowledgments

Special thanks to D. Cucurell, R. de Diego Balaguer, T. Gomila, M. Laine, J. Müller, and M. Guxens for their support and comments during various stages of the project. We also thank J. Riba, S. Romero, X. Mayoral, and C. Soriano for their technical help. This research was supported by the Spanish Government (MCYT) to ARF (SEJ2005-06067/PSIC; Ramon y Cajal program) and NS/JT (Consolider Ingenio 2010, CSD2007-00012 and SEJ 2007-60751). Support for the research of N.S. was received through the prize “ICREA Academia” for excellence in research, Generalitat de Catalunya.

## Appendix

Artificial languages used in the ERP experiment. Part-word construction in each language followed the same structure as exemplified in the first language.

\*Language streams used in the fMRI experiment. Random stream were constructed as a result of combining the syllables of these languages streams.

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

## References

- Ablá, D., Okanoya, K., 2008. Statistical segmentation of tone sequences activates the left inferior frontal cortex: a near-infrared spectroscopy study. *Neuropsychologia* 46, 2787–2795.
- Ablá, D., Katahira, K., Okanoya, K., 2008. On-line assessment of statistical learning by event-related potentials. *J. Cogn. Neurosci.* 20, 952–964.
- Aboitiz, F., García, R., Brunetti, E., Bosman, C., 2006. The origin of Broca’s area and its connections from an ancestral working/active memory network. In: Amunt, K., Grodzinsky, Y. (Eds.), *Broca’s Region*. Oxford University Press, Oxford, pp. 3–16.
- Ary, J.P., Klein, S.A., Fender, D.H., 1981. Location of sources of evoked scalp potentials: corrections for skull and scalp thicknesses. *IEEE Trans. Biomed. Eng.* 28, 447–452.
- Baddeley, A., Gathercole, S., Papagno, C., 1998. The phonological loop as a language learning device. *Psychol. Rev.* 105, 158–173.
- Berns, G.S., Cohen, J.D., Mintun, M.A., 1997. Brain regions responsive to novelty in the absence of awareness. *Science* 276, 1272–1275.
- Bookheimer, S., 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Ann. Rev. Neurosci.* 25, 151–188.
- Brent, M.R., 1997. Toward a unified model of lexical acquisition and lexical access. *J. Psycholinguist. Res.* 26, 363–375.
- Brent, M.R., 1999. Speech segmentation and word discovery: a computational perspective. *Trends Cogn. Sci.* 3, 294–301.
- Buchsbaum, B.R., Olsen, R.K., Koch, P., Berman, K.F., 2005. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48, 687–697.



- Buiatti, M., Pena, M., haene-Lambertz, G., 2009. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* 44, 509–519.
- Burton, M.W., Small, S.L., Blumstein, S.E., 2000. The role of segmentation in phonological processing: an fMRI investigation. *J. Cogn. Neurosci.* 12, 679–690.
- Catani, M., Jones, D.K., Ffytche, D.H., 2005. Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.
- Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J.P., Berry, I., Nespoulous, J.L., Chollet, F., 1999. Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage* 9, 135–144.
- Christiansen, M.H., Allen, J., Seidenberg, M.S., 1998. Learning to segment speech using multiple cues: a connectionist model. *Lang. Cogn. Processes* 13, 221–268.
- Clark, D., Wagner, A.D., 2003. Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia* 41, 304–317.
- Coccosco, C.A., Kollokian, V., Kwan, R.K.S., Evans, A.C., 1997. BrainWeb: online interface to a 3D MRI simulated brain database. *Neuroimage* 5 (4), S425.
- Conboy, B.T., Mills, D.L., 2006. Two languages, one developing brain: event-related potentials to words in bilingual toddlers. *Dev. Sci.* 9, F1–12.
- Cunillera, T., Toro, J.M., Sebastian-Galles, N., Rodriguez-Fornells, A., 2006. The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Res.* 1123, 168–178.
- Cunillera, T., Gomila, A., Rodriguez-Fornells, A., 2008. Beneficial effects of word final stress in segmenting a new language: evidence from ERPs. *BMC Neurosci.* 9, 23.
- Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., et al., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67.
- De Diego-Balaguer, R., Toro, J.M., Rodriguez-Fornells, A., Bachoud-Levi, A.C., 2007. Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE* e1175, 2.
- Demonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., Rascol, A., Frackowiak, R., 1992. The anatomy of phonological and semantic processing in normal subjects. *Brain* 115 (Pt 6), 1753–1768.
- Doupe, A.J., Kuhl, P.K., 1999. Birdsong and human speech: common themes and mechanisms. *Ann. Rev. Neurosci.* 22, 567–631.
- Dutoit, T., Gosselin, B., 1996. On the use of a hybrid harmonic/stochastic model for TTS synthesis-by-concatenation. *Speech Commun.* 19, 119–143.
- 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. 3rd ed., Philadelphia, pp. 1393–1396.
- Elman, J.L., 1990. Finding structure in time. *Cogn. Sci.* 14, 179–211.
- Fadiga, L., Craighero, L., 2006. Hand actions and speech representation in Broca's area. *Cortex* 42, 486–490.
- Fadiga, L., Craighero, L., Buccino, G., Rizzolatti, G., 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402.
- Fiez, J.A., Raichle, M.E., Miezin, F.M., Petersen, S.E., Tallal, P., Katz, W.F., 1995. PET studies of auditory and phonological processing – effects of stimulus characteristics and task demands. *J. Cogn. Neurosci.* 7, 357–375.
- Fiser, J., Aslin, R.N., 2001. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* 12, 499–504.
- Friedrich, M., Friederici, A.D., 2008. Neurophysiological correlates of online word learning in 14-month-old infants. *Neuroreport* 19, 1757–1761.
- Friston, K.J., Frith, C.D., Passingham, R.E., Liddle, P.F., Frackowiak, R.S., 1992. Motor practice and neurophysiological adaptation in the cerebellum: a positron tomography study. *Proc. Biol. Sci.* 248, 223–228.
- Friston, K.J., Frith, C.D., Turner, R., Frackowiak, R.S.J., 1995a. Characterizing evoked hemodynamics with fMRI. *Neuroimage* 2, 157–165.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R., 1995b. Statistical Parametric Maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210.
- Friston, K.J., Josephs, O., Rees, G., Turner, R., 1998. Nonlinear event-related responses in fMRI. *Magn. Reson. Med.* 39, 41–52.
- Friston, K.J., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. *Neuroimage* 39, 1104–1120.
- Graf-Estes, K., Evans, J.L., Alibali, M.W., Saffran, J.R., 2007. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260.
- Gronholm, P., Rinne, J.O., Vorobyev, V., Laine, M., 2005. Naming of newly learned objects: a PET activation study. *Brain Res. Cogn. Brain Res.* 25, 359–371.
- Gullberg, M., Indefrey, P., 2006. Cognitive Neuroscience of Second Language Acquisition. Oxford.
- Hackett, T.A., Stepniewska, I., Kaas, J.H., 1999. Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* 817, 45–58.
- Halgren, E., Baudena, P., Heit, G., Clarke, J.M., Marinkovic, K., Clarke, M., 1994. Spatiotemporal stages in face and word processing. I. Depth-recorded potentials in the human occipital, temporal and parietal lobes [corrected]. *J. Physiol. Paris* 88, 1–50.
- Halgren, E., Dhond, R.P., Christensen, N., Van, P.C., Marinkovic, K., Lewine, J.D., Dale, A.M., 2002. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage* 17, 1101–1116.
- Hauser, M.D., Newport, E.L., Aslin, R.N., 2001. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78, B53–B64.
- Heim, S., Friederici, A.D., 2003. Phonological processing in language production: time course of brain activity. *Neuroreport* 14, 2031–2033.
- Helenius, P., Salmelin, R., Service, E., Connolly, J.F., 1998. Distinct time courses of word and context comprehension in the left temporal cortex. *Brain* 121 (Pt 6), 1133–1142.
- Helenius, P., Salmelin, R., Service, E., Connolly, J.F., Leinonen, S., Lyytinen, H., 2002. Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *J. Neurosci.* 22, 2936–2944.
- Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138.
- Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99.
- Hickok, G., Poeppel, D., 2007. Opinion – the cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., Rizzolatti, G., 1999. Cortical mechanisms of human imitation. *Science* 286, 2526–2528.
- Jacquemot, C., Scott, S.K., 2006. What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.* 10, 480–486.
- Jennings, J.R., Wood, C.C., 1976. Epsilon-adjustment procedure for repeated-measures analyses of variance. *Psychophysiology* 13, 277–278.
- Jusczyk, P.W., Houston, D.M., Newsome, M., 1999. The beginnings of word segmentation in English-learning infants. *Cogn. Psychol.* 39, 159–207.
- Kilner, J.M., Mattout, J., Henson, R., Friston, K.J., 2005. Hemodynamic correlates of EEG: a heuristic. *Neuroimage* 28, 280–286.
- Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848.
- Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843.
- Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470.
- Kwon, H., Kuriki, S., Kim, J.M., Lee, Y.H., Kim, K., Nam, K., 2005. MEG study on neural activities associated with syntactic and semantic violations in spoken Korean sentences. *Neurosci. Res.* 51, 349–357.
- Liberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised. *Cognition* 21, 1–36.
- LoCasto, P.C., Krebs-Noble, D., Gullapalli, R.P., Burton, M.W., 2004. An fMRI investigation of speech and tone segmentation. *J. Cogn. Neurosci.* 16, 1612–1624.
- Logothetis, N.K., 2002. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1003–1037.
- Marinkovic, K., Dhond, R.P., Dale, A.M., Glessner, M., Carr, V., Halgren, E., 2003. Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* 38, 487–497.
- 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.
- McLaughlin, J., Osterhout, L., Kim, A., 2004. Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* 7, 703–704.
- McNealy, K., Mazziotta, J.C., Dapretto, M., 2006. Cracking the language code: neural mechanisms underlying speech parsing. *J. Neurosci.* 26, 7629–7639.
- Meister, I.G., Wilson, S.M., Deblieck, C., Wu, A.D., Iacoboni, M., 2007. The essential role of premotor cortex in speech perception. *Curr. Biol.* 17, 1692–1696.
- Meltzoff, A.N., Decety, J., 2003. What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 491–500.
- Mestres-Misse, A., Rodriguez-Fornells, A., Munte, T.F., 2007. Watching the brain during meaning acquisition. *Cerebral Cortex* 17, 1858–1866.
- Mestres-Misse, A., Camara, E., Rodriguez-Fornells, A., Rotte, M., Munte, T.F., 2008. Functional neuroanatomy of meaning acquisition from context. *J. Cogn. Neurosci.* 20, 2153–2166.
- Mestres-Misse, A., Munte, T.F., Rodriguez-Fornells, A., in press. Functional neuroanatomy of contextual acquisition of concrete and abstract words. *J. Cogn. Neurosci.* doi:10.1162/jocn.2008.21171.
- Mills, D.L., Plunkett, K., Prat, C., Schafer, G., 2005. Watching the infant brain learn words: effects of vocabulary size and experience. *Cogn. Dev.* 20, 19–31.
- Mueller, J.L., Bahlmann, J., Friederici, A.D., 2008. The role of pause cues in language learning: the emergence of event-related potentials related to sequence processing. *J. Cogn. Neurosci.* 20, 892–905.
- Nobre, A.C., Allison, T., Mccarthy, G., 1994. Word recognition in the human inferior temporal lobe. *Nature* 372, 260–263.
- Pacton, S., Perruchet, P., 2008. An attention-based associative account of adjacent and nonadjacent dependency learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 80–96.
- Paulesu, E., Frith, C.D., Frackowiak, R.S.J., 1993. The neural correlates of the verbal component of working memory. *Nature* 362, 342–345.
- Perfetti, C.A., Wlotko, E.W., Hart, L.A., 2005. Word learning and individual differences in word learning reflected in event-related potentials. *J. Exper. Psychol.-Learn. Mem. Cogn.* 31, 1281–1292.
- Perruchet, P., Vinter, A., 1998. PARSER: a model for word segmentation. *J. Mem. Lang.* 39, 246–263.
- Perruchet, P., Pacton, S., 2006. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci.* 10, 233–238.
- Poldrack, R.A., 2000. Imaging brain plasticity: conceptual and methodological issues – a theoretical review. *Neuroimage* 12, 1–13.
- Poldrack, R.A., Desmond, J.E., Glover, G.H., Gabrieli, J.D., 1998. The neural basis of visual skill learning: an fMRI study of mirror reading. *Cereb. Cortex* 8, 1–10.



- Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., Gabrieli, J.D., 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10, 15–35.
- Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado, M.F., Hauk, O., Shtyrov, Y., 2006. Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7865–7870.
- Raichle, M.E., Fiez, J.A., Videen, T.O., MacLeod, A.M., Pardo, J.V., Fox, P.T., Petersen, S.E., 1994. Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb. Cortex* 4, 8–26.
- Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. *Trends Neurosci.* 21, 188–194.
- Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., Rauschecker, J.P., 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136.
- Rugg, M.D., Coles, M.G.H., 1995. *Electrophysiology of Mind. Event-Related Brain Potentials and Cognition*. Oxford University Press, Oxford.
- 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.
- Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A., Barrueco, S., 1997. Incidental language learning: listening (and learning) out of the corner of your ear. *Psychol. Sci.* 8, 101–105.
- Saffran, J.R., Johnson, E.K., Aslin, R.N., Newport, E.L., 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.
- Sanders, L.D., Neville, H.J., 2003a. An ERP study of continuous speech processing I. Segmentation, semantics, and syntax in native speakers. *Cogn. Brain Res.* 15, 228–240.
- Sanders, L.D., Neville, H.J., 2003b. An ERP study of continuous speech processing II. Segmentation, semantics, and syntax in non-native speakers. *Cogn. Brain Res.* 15, 214–227.
- 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.
- Schubotz, R.I., von Cramon, D.Y., Lohmann, G., 2003. Auditory what, where, and when: a sensory somatotopy in lateral premotor cortex. *Neuroimage* 20, 173–185.
- Scott, S.K., Wise, R.J.S., 2004. The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92, 13–45.
- Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123 (Pt 12), 2400–2406.
- Scott, S.K., Rosen, S., Lang, H., Wise, R.J.S., 2006. Neural correlates of intelligibility in speech investigated with noise vocoded speech – a positron emission tomography study. *J. Acoust. Soc. Am.* 120, 1075–1083.
- Smith, E.E., Jonides, J., 1998. Neuroimaging analyses of human working memory. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12061–12068.
- Studdert-Kennedy, M., 1987. The phoneme as a perceptuomotor structure. In: Allport, D., MacKay, D., Prinz, W., Scheerer, E. (Eds.), *Language Perception and Production*. Academic Press, London, pp. 67–84.
- Tettamanti, M., Alkadhi, H., Moro, A., Perani, D., Kollias, S., Weniger, D., 2002. Neural correlates for the acquisition of natural language syntax. *NeuroImage* 17, 700–709.
- Toro, J.M., Trobalon, J.B., 2005. Statistical computations over a speech stream in a rodent. *Percept. Psychophys.* 67, 867–875.
- Toro, J.M., Sinnott, S., Soto-Faraco, S., 2005. Speech segmentation by statistical learning depends on attention. *Cognition* 97, B25–B34.
- Turk-Browne, N.B., Junge, J.A., Scholl, B.J., 2005. The automaticity of visual statistical learning. *J. Exp. Psychol.-Gen.* 134, 552–564.
- van Petten, C., Luka, B.J., 2006. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang.* 97, 279–293.
- Warren, J.E., Wise, R.J.S., Warren, J.D., 2005. Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci.* 28, 636–643.
- Watkins, K.E., Strafella, A.P., Paus, T., 2003. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41, 989–994.
- Wilson, S.M., Iacoboni, M., 2006. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* 33, 316–325.
- Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702.
- Wise, R.J.S., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., Warburton, E.A., 2001. Separate neural subsystems within 'Wernicke's area'. *Brain* 124, 83–95.
- Wong, P.C., Perrachione, T.K., Parrish, T.B., 2007. Neural characteristics of successful and less successful speech and word learning in adults. *Hum. Brain Mapp.* 28, 995–1006.
- Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited – again. *Neuroimage* 2, 173–181.
- Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256, 846–849.
- Zatorre, R.J., Meyer, E., Gjedde, A., Evans, A.C., 1996. PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cereb. Cortex* 6, 21–30.