



Electrophysiological estimates of the time course of semantic and phonological encoding during listening and naming

Antoni Rodriguez-Fornells^a, Bernadette M. Schmitt^b, Marta Kutas^c, Thomas F. Münte^{a,*}

^a Department of Neuropsychology and Department of Neurologie II, Otto von Guericke University, Universitätsplatz 2, Gebäude 24, 39112 Magdeburg, Germany

^b Department of Psychology/Neurocognition, Maastricht University, Maastricht, The Netherlands

^c Departments of Cognitive Science and Neurosciences, UCSD, San Diego, CA, USA

Received 9 June 2000; received in revised form 24 August 2001; accepted 3 September 2001

Abstract

Current psycholinguistic models suggest that we know what we want to say before we decide how we are going to say it: in other words, for speaking, word meaning is activated prior to information about syntax and phonology. Listening likely involves the reverse order of processes: phonological processing before meaning activation. We examined the relative time courses of phonological and semantic processing during language production and comprehension using event-related brain potentials (ERPs). Participants viewed a series of pictures (with the instruction to covertly name the depicted item), or heard a series of words, and made dual choice Go/noGo decisions based on each item's conceptual (whether the item was an animal or an object) and phonological features (whether the item's German name started with a vowel or a consonant). During picture naming, the N200 component (related to response inhibition) indicated that conceptual processing preceded phonological processing by about 170 ms. During auditory word processing, on the other hand, the brain activity related to these two aspects of comprehension indicated some temporal overlap with the N200 to phonological processing preceding that to semantic processing by only about 85 ms. In sum, the data are compatible with current psycholinguistic models of speech production and comprehension and argue for serial or widely spaced cascaded processing during production but more parallel processing of information during comprehension. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Phonology; Semantic; Event-related brain potentials (ERPs); Comprehension; Production

1. Introduction

Speaking and listening to other people requires a complex processing system capable of producing and understanding words in an efficient manner. A very fast rate of processing is crucial both for selecting and articulating proper utterances and for determining the meaning of an incoming utterance. In broad stroke, psycholinguists agree that language production, i.e. the mapping of a concept to a sound, requires at least three different kinds of representations: conceptual, syntactic, and phonological [25]. It seems reasonable to suppose that for language comprehension, i.e. mapping sounds onto meaning, the relevant phonological and conceptual/semantic processes might occur in the reverse order [2]. The present study capitalizes on the high temporal resolution of on-line electrophysiological measures, namely event-related brain potentials (ERPs) to delineate the rela-

tive time courses of conceptual and phonological processes during language production and comprehension in a group of neurologically-intact young adults.

1.1. From meaning-to-sound: word production

According to a prevailing view, when people wish to name an object, they must first access the name's conceptual representation, and only then can they retrieve the syntactic information and phonological form. There are thus at least two distinct stages enroute from concept to articulation [4,22,41]. In the first stage, lexical access, concept activation drives the selection of abstract lexical candidates (*lemmata*) containing the name's syntactic features [16,22]. In the second stage, the *lemma* is used to retrieve the correct phonological form of the target word (i.e. the object's name). On this analysis of word production, for any given word three different kinds of representations, *semantic*, *syntactic*, and *phonological*, exist.

The temporal course of these processes in speech production has been studied using a variety of chronometric tasks,

* Corresponding author. Tel.: +49-391-671-8469;

fax: +49-391-671-1947.

E-mail address: thomas.muente@medizin.uni-magdeburg.de (T.F. Münte).

with the preferred paradigm being the picture naming task or PNT (see for recent reviews, [24,25]). From these reaction time (RT) studies it seems that a word's semantic properties become available before its phonological form [4,23,36,41]. Specifically, the first stage of conceptual analysis is estimated to take approximately 100–150 ms [38,43]. Lemma selection is presumably completed by 275 ms after picture onset [23,24], and the phonological stage has been placed between 275 and 400 ms after picture onset [23,39].

This sequence of events during language production has received support from ERP measurements. Specifically, in two studies employing a brain potential linked to response preparation called the lateralized readiness potential (LRP), Van Turennout et al. found that semantic information is encoded about 120 ms before phonological information [46], and that syntactic properties are encoded about 40 ms before phonological information [47]. Schmitt et al. [42] replicated Van Turennout's LRP results and observed further that a brain potential linked to response inhibition, the N200, similarly indicated that semantic encoding precedes phonological encoding during picture naming.

1.2. *From sound-to-meaning: word comprehension*

From research in auditory psychophysics it is known that a listener can accurately identify at least 15 phonemes per second. To accomplish this, the auditory system must convert fluctuations in air-pressure into a representation that allows the listener to access stored representations of words. First, the speech signal is selected from the acoustic background. Then, some segmentation procedure extracts phonetic features from the continuous speech signal. This information is used by the recognition process which includes the retrieval of lexical information and the activation of lexical candidates. Activated lexical candidates provide access to their meaning and to syntactic information (referred to here as "lexical access"). This in turn leads to the interpretation of the heard utterance as it is integrated with the ongoing discourse (for a review see [2]).

With regard to the time course of lexical access during auditory comprehension, there is a strong consensus that various lexical entries are activated, at least to some extent, as soon as the first features or phonemes of an incoming word are identified [6,31]. For instance, according to the cohort model [28,30] an incoming phoneme activates all lexical entries that begin with that phoneme (i.e. selecting the "word-initial cohort"). As more information comes in, this size of this cohort is progressively reduced, until eventually only one candidate remains. Competition among the different activated candidates during the selection process is inherent in this model (see [9,33]).

The first clear evidence in favor of the cohort model was reported by Marslen-Wilson [27], who found that fast shadowers were able to repeat tape-recorded speech passages with delays of less than 200 ms, i.e. well before the utterance ended and thus before sufficient sensory information

was available for unambiguous word identification [29]. Additional evidence comes from studies using the gating technique [11,12,44] wherein people are asked to guess what word they hear from speech presented in segments (e.g. 50 ms) of increasing duration, as well as from cross-modal priming experiments using word-initial, partial primes [49].

Again, several ERP studies have been performed to assess the time course of information processing during comprehension. One of the most widely used ERP components in language research is the N400, a negative component sensitive to meaning integration and semantic processing. For example, when a semantically incongruous word is presented at the end of a sentence a large N400 component is elicited compared to the response to congruous final words (for review see [20,21]). Measured in natural speech contexts, the N400 semantic incongruity effect appears about 200 ms after final word onset [13]. Thus, as the average duration of words exceeds 200 ms, the onset of the N400 precedes the offset of the eliciting word. O'Rourke and Holcomb [35] used this effect in a lexical decision task to compare pseudowords which differed from real words either early or late. Consistent with the cohort model they found that the onset latency of the N400 was delayed for pseudowords that differed from real words at a later time point.

Van Petten et al. [45] likewise used the latency of the N400 congruity effect to sentence final anomalous words for which they had determined recognition points via a gating procedure to examine the timing of auditory comprehension processes. Specifically, subjects heard sentences (e.g. most marine mammals have some sort of legal protection, but fishermen continue to kill . . .) that were completed by four different types of sentence final words: (a) cohort congruous (expected word, e.g. dolphins), (b) semantically anomalous endings with no obvious phonological relationship to the congruous completions of those sentences, (c) semantically incongruous words that shared three initial phonemes with the expected congruous completions (e.g. dollars) and (d) semantically incongruous words that shared final phonemes with the congruous completions (e.g. muffins). The ERPs for congruous and anomalous endings diverged around 200 ms. The N400 to semantic anomalies that shared initial phonemes with the expected congruous endings was delayed relative to the N400s to the plain incongruous endings such that the N400 congruity effect for these began around 382 ms. The N400 to semantic anomalies that rhymed with the expected congruous endings, on the other hand, was not delayed. The results indicate that semantic integration during speech comprehension begins before word recognition is complete (i.e. on the basis of partial, incomplete information).

Thus, in contrast to the findings on speech production which support serial or cascade processing models, experiments on auditory comprehension suggest that phonological analysis of an incoming word and its mapping to meaning proceed in a largely parallel fashion. The present study was designed to examine these potential differences in timing

and cognitive architecture of spoken language comprehension and speech production in the same subjects by using the N200 noGo effect.

1.3. N200 noGo component

In a classic Go/noGo paradigm, a participant is asked to respond to one class of stimuli (Go) and to withhold responding to another (noGo). Pfefferbaum et al. [37] observed a large negativity around 275 ms over frontal sites to noGo responses compared to Go responses [5,15,17,43]. One interpretation of the N200 to noGo stimuli that has received some empirical support is that it reflects an inhibitory process emanating from structures in the prefrontal cortex. Destruction of prefrontal cortex in animals has been found to lead to a profound disturbance of performance in delayed response tasks [9], and to disinhibition and impulsive behavior [26]. Also, a similar “noGo” potential has been observed in surface and chronically-implanted depth electrodes from the prefrontal cortex of monkeys performing a color discrimination task [40]. The results of a functional magnetic resonance imaging (fMRI) study [1] indicated that individuals with greater orbital frontal activation (noGo versus Go comparison) committed fewer false alarms, consistent with greater response inhibition. Garavan et al. [10] noted right-hemisphere dominance in a response inhibition task suggesting a more widespread cortical network including the dorsolateral and inferior frontal lobe [18].

1.4. N200 noGo component used as a chronometric index in language research

The temporal course of a number of cognitive processes has been addressed using the N200 component. Thorpe et al. [43], for instance, employed the N200 noGo effect to a visual categorization task with the aim of determining the minimum time needed for conceptual processing of pictures. As they observed a frontal N200 noGo component approximately 150 ms after picture onset and took this to be the minimum time needed to inhibit the response to a picture that was not a member of the desired (“Go”) category. Schmitt et al. [42] compared the latencies of N200s contingent on semantic versus phonological information in a dual choice Go/noGo RT paradigm in a modified PNT. In this task, one type of information determines the response hand (right or left hand), while the other determines whether or not a response has to be executed (Go/noGo decision). As mentioned previously, the semantic N200 effect preceded the phonological one in this study.

In the present investigation we use the N200 noGo effect to compare the temporal course of the processing stages involved in silent speech production and auditory word comprehension using a PNT and an auditory processing task, respectively. Based on previous work we expect that during implicit picture naming the semantic N200 effect will precede the phonological N200 whereas during the analysis of

a spoken word the phonological N200 will precede the semantic N200. However, in view of the compelling evidence of at least partially parallel processing of phonological and semantic properties of auditorily presented words, the N200 timing difference in the auditory task was expected to be much smaller than in the production task.

1.5. The LRP

The LRP is derived from the readiness potential (RP) that precedes voluntary hand movements. In contrast to the RP, the LRP is usually computed time-locked to the stimulus to which the response is given. By averaging the lateralized activity to responses made with the left and right-hand (given contralateral versus ipsilateral recording), any asymmetry that is not related to response preparation cancels out. The LRP is presumed to reflect the average amount of lateralization specifically related to the motor preparation of the responding hands, with larger amplitudes seen over the motor cortex of the hemisphere contralateral to the response (e.g. [19]). The LRP allows researchers to see motor-related brain activity prior to an overt or Go response [34]. Perhaps more importantly, the LRP also reveals motor activation in the absence of any overt response. That is, the LRP reflects preparation to respond, even when the response is not executed as in the case of noGo trials. These characteristics make the LRP a suitable brain measure for studying the time course of information encoding, such as during speech planning and production [42,46,47].

In the current experiment, a dual choice Go/noGo paradigm was used in conjunction with a PNT and an auditory word comprehension task (AWT). Subjects were presented with a series of either pictures or words depicting or denoting common objects or animals, respectively. They were asked to make a semantic and a phonological decision about each stimulus. For example, in one condition (hereafter referred to as Go/noGo = semantics condition) participants had to respond (Go trials) or to refrain from responding (noGo trials) depending on the semantic characteristics of the stimulus (e.g. Go = animal, noGo = object). In this case, the responding hand was contingent on the phonological properties of the stimulus (e.g. first letter = vowel = right hand, first letter = consonant = left hand). In the second condition (hereafter referred to as Go/noGo = phonology condition) the response contingencies were reversed, i.e. the response hand was based on semantic information and the Go/noGo decision was contingent on phonological information.

2. Method

2.1. Subjects

Eighteen right-handed, neurologically-intact, native speakers of German (14 women, mean age 23) were paid

to participate in the experiment (14 women, mean age 23) after giving written informed consent.

2.2. Stimuli

2.2.1. Picture naming task (PNT)

The stimuli used for this task have previously been described in Schmitt et al. [42]. Briefly, there were 100 simple black-on-white line drawings including 50 animals and 50 objects. In each semantic category, the names of half of the items started with a vowel (such as ‘elephant’ or ‘igloo’ and the other half with a consonant (such as ‘tiger’ or ‘pipe’, see Appendix B in Schmitt et al. [42]). Twenty different pictures were used for practice trials. Pictures subtended about 8° of visual angle and 8° in width at a viewing distance of 1.1 m.

2.2.2. Auditory word processing task (AWT)

The auditory stimuli consisted of the same items as in the PNT spoken by a female speaker, recorded and digitized at 22.05 kHz. After A/D conversion each word was edited using a special purpose sound editing system. Silent periods at the onset of the stimuli were eliminated. The duration of the stimuli (from word onset to word offset) did not differ among the conditions ($F(3, 96) = 1.18$: animal/vowel $954 \text{ ms} \pm 155$, animal/consonant 908 ± 220 , object/vowel 992 ± 189 and object/consonant 995 ± 175). The stimuli were presented binaurally via headphones at a comfortable intensity.

2.3. Procedure

Each participant was tested with eight different instruction sets in each task, PNT and AWT ([Go/noGo = semantic versus Go/noGo = phonology] \times [left versus right-hand assignment] \times [Go versus noGo response]) in two sessions lasting approximately 2.5 h each separated. Sessions were separated between 3 days and 2 weeks. Table 1 illustrates the eight different instruction sets and provides an example of the different responses to the same item, in this case “feather”. Each stimulus was presented eight times to each participant, i.e. once with each instruction. For four of the instruction sets, the Go/noGo decision was contingent on semantic information (Go/noGo = semantics); e.g. respond if the item is an animal (Go) but not if it is an object (noGo),

respond with the right-hand, if the word starts with a vowel, respond with the left hand, if the initial letter is a consonant. In the other four instruction sets the Go/noGo decision was contingent on the phonological information (Go/noGo = phonology) and the responding hands were determined by the semantic information.

The order of instructions was systematically varied across participants. For data analysis, the ERPs obtained with all four instructions that based the Go/noGo decision on semantic information were averaged together, likewise the ERPs from all instructions that based Go/noGo decisions on phonology. The order of PNT and AWT tasks was counterbalanced across participants. Also, the phonological and semantic Go/noGo tasks were counterbalanced: half of the subjects performed the four Go/noGo = phonology conditions first and the other half the Go/noGo = semantics conditions.

Each condition began with 20 practice trials, followed by 100 experimental trials, lasting about 10 min. The sequences of pictures were randomized in every block and for each subject.

2.3.1. Picture-naming task (PNT)

In a first training block participants were familiarized with the pictures which were shown with their name printed below. Participants were asked to study the pictures and their names. In a second training block, the pictures were presented again but without their names, and participants were asked to name each one aloud as fast and as accurately as possible. This was done to ensure that subject would in fact use the intended name.

For the experimental blocks participants were asked to carry out the assigned dual choice Go/noGo task in each block without overtly naming the picture. Each trial had the following structure: a fixation point (warning stimulus) appeared in the center of a video screen and after a random inter-trial interval (range 1500–3000) a picture appeared for 1500 ms followed by the fixation point.

2.3.2. Auditory word processing task (AWT)

To familiarize them with the words, subjects were given a written list of words which they could follow as they heard the 100 words over headphones. As in the PNT, 20 different

Table 1
Illustration of the eight different instructions exemplified for the stimulus “Feder” (feather)

	Eight different instructions (1–4: Go/noGo = phonology; 5–8: Go/noGo = semantics)	Executed response for stimulus ‘feather’: (object, consonant)
1	Press left if animal, press right if object, press only if its name starts with a consonant	Semantics right, phonology Go
2	Press right if animal, press left if object, press only if its name starts with a consonant	Semantics left, phonology Go
3	Press left if animal, press right if object, press only if its name starts with a vowel	Semantics right, phonology noGo
4	Press right if animal, press left if object, press only if its name starts with a vowel	Semantics left, phonology noGo
5	Press right if consonant, press left if vowel, press only if object	Phonology right, semantics Go
6	Press right if vowel, press left if consonant, press only if object	Phonology left, semantics Go
7	Press left if vowel, press right if consonant, press only if animal	Phonology right, semantics noGo
8	Press left if consonant, press right if vowel, press only if animal	Phonology left, semantics noGo

trials were used for practice before each experimental block. An auditory warning stimuli (Achtung/Attention!) was presented at the beginning of each block, followed by a random sequence of the 100 target words. The ISI between auditory words stimuli was varied randomly (1500–3000 ms, rectangular distribution).

2.4. Electrophysiological recording

The ERPs were recorded from the scalp using tin electrodes mounted in an elastic cap and located at 29 standard positions (Fp1/2, F7/8, F5/6, F3/4, C3/4, C5/6, T3/4, Fpz, Fz, Cz, Pz, Cp3/4, Cp5/6, Tp7/8, P3/4, T5/6, O1/2). Biosignals were re-referenced 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 vertical EOG). Electrode impedances were kept below 5 k Ω .

The electrophysiological signals were filtered with a bandpass of 0.01–70 Hz (half-amplitude cutoffs) and digitized at a rate of 250 Hz. Trials on which base-to-peak electro-oculogram (EOG) amplitude exceeded 75 μ V, amplifier saturation occurred, or the baseline shift exceeded 200 μ V/s were automatically rejected off-line. The number of rejections did not differ for the two tasks ($M = 25.6\%$ for PNT and $M = 28\%$ for AWT).

Artifact free and correct trials (minimum: 125 trials per average and subject) were averaged separately for each stimulus type and condition over epochs of 1600 ms starting 100 ms prior to the stimulus. ERPs from the different conditions were later combined to yield four basic conditions.

For all statistical effects involving two or more degrees of freedom in the numerator, the Greenhouse–Geisser epsilon was used to correct possible violations of the sphericity assumption [14]. Exact P -value after correction will be reported. Tests involving electrode \times condition interactions (e.g. factors as hemisphere or anterior–posterior electrode location) were carried out on data corrected using the vector normalization procedure described by McCarthy and Wood [32].

3. Results

3.1. Overt behavior

Incorrect responses, and responses faster than 200 ms and slower than 1700 ms were excluded from further analyses. An ANOVA on the remaining trials was computed with task (PNT or AWT) and condition (Go/noGo = phonological versus Go/noGo = semantics) as repeated measures factors. The mean RT in the PNT was 1006 ms \pm 163 (S.D.) for Go/noGo = phonology decisions and 973 ms \pm 153 for Go/noGo = semantics. In the AWT, mean RT decisions was 1090 ms \pm 107 for Go/noGo = phonology and 1116 ms \pm 105 for the Go/noGo = semantics. This difference between tasks

was significant ($F(1, 17) = 17.51, P < 0.001$; mean RT for PNT was 989 \pm 157 and for AWT was 1103 \pm 106), with Go responses in the PNT being about 114 ms faster than in the AWT. There was no main effect of condition ($F < 1$), but a significant interaction between task and condition ($F(1, 17) = 6.55, P < 0.020$). Specifically, in the PNT a decision based on semantics is about 33 ms faster than one on phonology and in the AWT, phonology Go decisions are about 26 ms faster than the semantic Go decisions.

Responses on noGo trials (failure to inhibit trials) are called false Gos. In the auditory task there were more false Gos when the Go/noGo decision was based on semantics (4.2%) than when the decision was based on phonology (2%). In the PNT, the reverse pattern was observed, namely, more false Gos based on Phonology (5.9%) than on semantics (4%). For statistical analysis, error rates were replaced by estimates using the inverse sine transformation of the square-root of the error proportions [48]. A two way ANOVA (task \times condition) revealed a significant main effect of Task ($F(1, 17) = 5.75, P < 0.028$) reflecting more “false Go” errors in the PNT. There was no main effect of condition ($F < 1$), but the interaction between task \times condition was significant ($F(1, 17) = 11.4, P < 0.004$).

The number of errors on Go trials (wrong hand) was greater for the PNT than for the AWT (task effect, $F(1, 17) = 27.6, P < 0.001$; 3.1 versus 2%, respectively). No significant differences were found for the factor condition, or the task \times condition interaction (both $P > 0.17$).

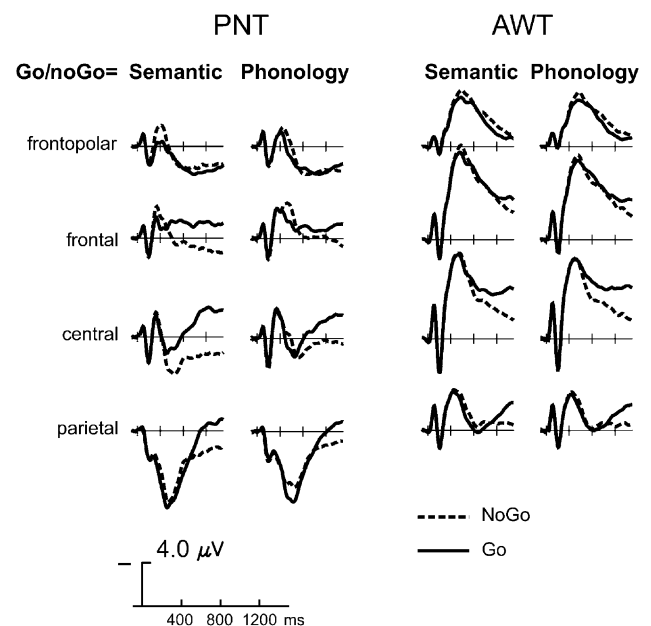


Fig. 1. Grand averaged ERPs ($n = 18$) for several midline electrodes elicited on Go and noGo trials in the Go/noGo = phonology and Go/noGo = semantics conditions. Both conditions are associated with a more negative waveform for the noGo trials, i.e. an N200 effect, in the anterior derivations. In addition, a latency shift in the late positive component is seen for the posterior channels.

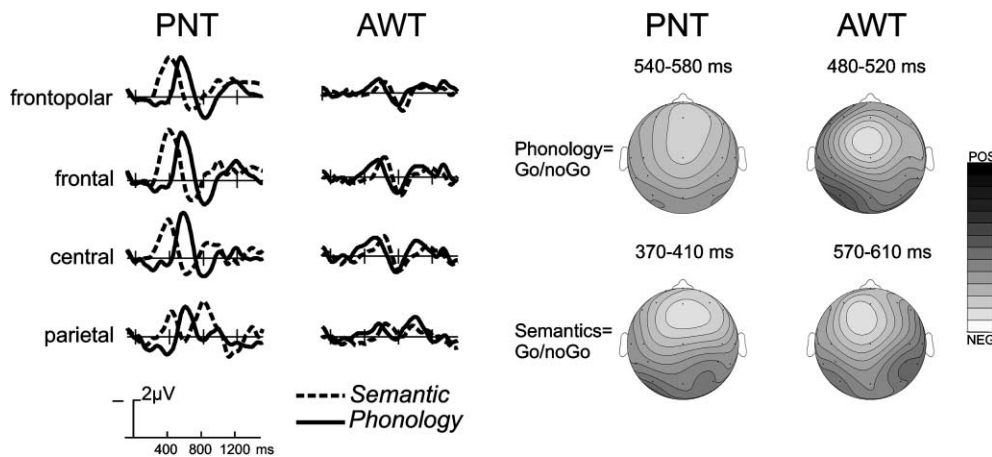


Fig. 2. Left: difference waves obtained by subtracting the ERPs to the Go trials from those to the noGo trials. A reversal of the order of N200 effects is seen between the tasks (PNT: semantic condition first, auditory task: phonological condition first). Right: spline interpolated isovoltage maps. Depicted is the mean voltage in 40 ms time-windows centered upon the peak in the grand average (scale AWT most positive 0.6 μ V, most negative -0.9μ V, PNT most positive 1.4 μ V, most negative 2.8 μ V). All 4 N200 effects show a very similar scalp topography.

3.2. ERP data

The ERPs elicited in both tasks (PNT and AWT) and for each type of decision (Go/noGo = phonology and Go/noGo = semantics) are illustrated in Fig. 1. In the PNT an early negative deflection (N100) is followed by a parieto-occipital positive wave (P200). After these early sensory components, a negative deflection (maximal over frontal and frontopolar locations and larger for noGo conditions) occurs between 300 and 400 ms depending on the type of decision to be executed. In the auditory task, a centro-temporal N100–P200 is followed by a widespread negativity peaking at about 500–600 ms at central and frontal sites. This negative component is larger for noGo trials, especially at frontopolar sites. Maximal at central sites Go trials were associated with a ramp-shaped negativity starting at 600–700 ms. This negative shift is probably related to the preparation of the motor response in these trials.

3.3. N200 Go/noGo effect

The noGo minus Go difference waves, i.e. the N200 effects, were calculated for both tasks (see Fig. 2). Onset and peak latency were measured and subjected to statistical analyses.¹ In the PNT, the N200 effect in the Go/noGo = semantics condition had earlier onset and peak latencies than that in the Go/noGo = phonology condition. This pattern was reversed in the AWT, however, the overlap of the phonological and semantic N200 effects was greater in the AWT compared to the PNT.

¹ For the analysis of N200 differences waves (noGo — Go trials), a bandpass filter (1–5 Hz) was applied in order to stabilize possible drifts observed in some subjects after the imperative stimuli.

3.4. Onset latencies

The N200 difference waves were quantified by mean amplitude measures relative to an extended baseline (-100 to 100 ms). Onset latencies were determined via a stepwise series of one-tailed serial t -tests (step size = 4 ms). For each test, data from a time-window of 40 ms were averaged (i.e. point of measure ± 20 ms), from 200 to 1000 ms after picture onset. Onset latency was defined as the point at which four consecutive t -tests showed a significant difference from zero ($P < 0.05$).

In the PNT, the difference N200 at Fz for the Go/noGo = semantics decisions diverged from the baseline between 264 and 532 ms (i.e. for a duration of 268 ms, $-2.24 < t(17) < -7.68$). The difference N200 at Fz for the Go/noGo = phonology decision diverged from the baseline between 456 and 684 ms (228 ms duration, $-2.51 < t(17) < -5.75$; see Fig. 3 for a depiction of the t -tests). Similar patterns were observed at other anterior sites. A serial paired t -test comparing the two N200 effects showed significant differences between 248 and 468 ms ($2.1 < t(17) < 9.4$) and between 528 and 712 ms ($2.4 < t(17) < 5.9$).

In the AWT, the N200 effect at Fz based on semantic decisions was significant between 568 and 700 ms (duration 132 ms, $-2.26 < t(17) < -3.47$), while that based on phonological decisions was significant between 408 and 612 ms (duration 204 ms, $-2.27 < t(17) < -3.12$). Significant differences between the conditions (paired t -test) were found between 456 and 540 ms ($2.17 < t(17) < 2.76$) and between 664 and 732 ms ($2.19 < t(17) < 2.52$). Similar results were obtained for the data from the other frontopolar and frontal sites.

To assess the degree of overlap between the two N2 components, the area under the curve (negative direction) was determined for both N2 components. In addition, the

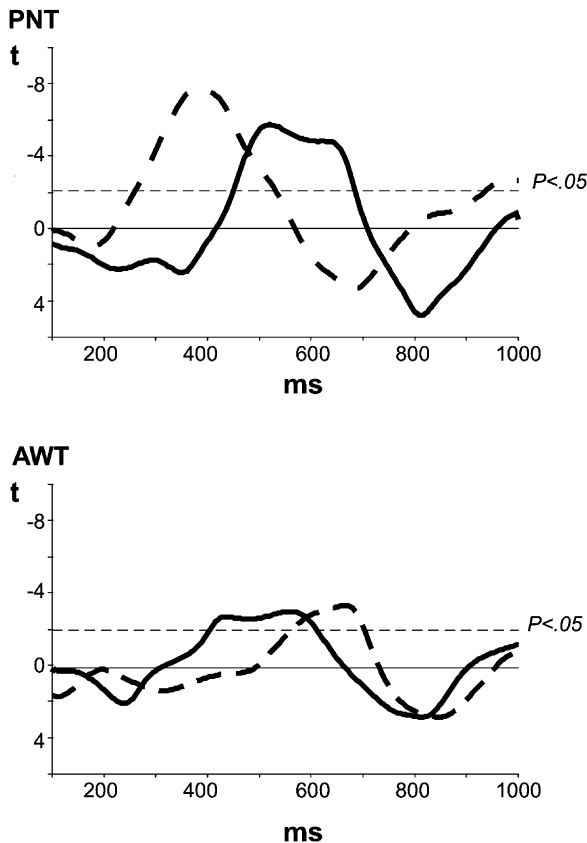


Fig. 3. One-tailed t -values (y-axis) of the N200 difference waves at Fz for Go/noGo = phonology, and Go/noGo = semantics conditions in each task (PNT and AWT) are graphed against time (x-axis, ms). Mean amplitude measurements were taken in consecutive 40 ms time-windows centered on the time-points in the figures (e.g. 300 ms represent the mean amplitude between 280 and 320 ms).

overlapping area was calculated. From these values the percentage of overlap was obtained (see Fig. 4 for an illustration). Five subjects had to be dropped from this analysis, as their N2 component did not have a monophasic morphology. The overlap was 23% (S.D. 17) in the PNT and 43% (S.D. 18) in the AWT (Wilcoxon matched pairs, $P < 0.01$).

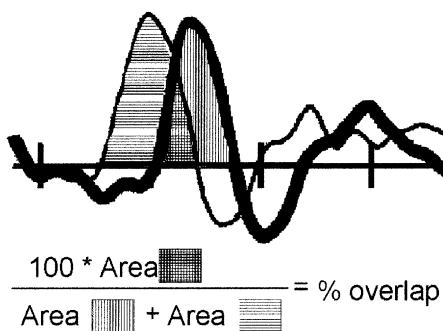


Fig. 4. Illustration of the calculation of the percentage of overlap between the N2-effects.

3.5. Peak analysis

The N200 effect peak latencies and mean amplitudes were measured at Fz. For peak latencies, there was a significant main effect of task reflecting the earlier (~ 69 ms) peak in the PNT than AWT ($F(1, 17) = 11.6$, $P < 0.003$). The condition factor was not significant ($P > 0.142$). However, the task \times condition interaction was significant ($F(1, 17) = 46.7$, $P < 0.0001$). In the PNT, the mean peak latency for the Go/noGo = phonology condition was delayed by 173 ms compared to that in the Go/noGo = semantics condition ($563 \text{ ms} \pm 73$ and $390 \text{ ms} \pm 76$, respectively). In the AWT, the reverse pattern was obtained: the N200 effect in the Go/noGo = phonology condition peaked 92 ms before that in the Go/noGo = semantics condition ($499 \text{ ms} \pm 113$ and 591 ± 117 , respectively). In the ANOVA performed on N200 difference peak amplitudes, only the main effect of task was significant ($F(1, 17) = 16.6$, $P < 0.001$), with a larger mean amplitude in the PNT ($-3.5 \mu\text{V} \pm 1.8$) than the AWT ($-1.9 \mu\text{V} \pm 0.95$).

To rule out the possibility that the pattern of effects was due merely to the subtraction procedures, a similar peak analysis was performed on the four noGo waveforms (Fz location). In the PNT, the mean peak latency for the noGo = phonology was $484 \text{ ms} \pm 33$ while that for the noGo = semantics condition was $384 \text{ ms} \pm 28$. In the AWT, the reverse pattern was obtained: N200 latency in the noGo = phonology condition was $546 \text{ ms} \pm 14$ and that in the noGo = semantics condition was $592 \text{ ms} \pm 13$, thus corroborating the pattern obtained for the difference waves. Statistically, this was reflected in a main effect of task ($F(1, 17) = 21.1$, $P < 0.001$). The main effect of condition was significant ($F(1, 17) = 5.15$, $P < 0.05$) as was the task \times condition interaction ($F(1, 17) = 18.4$, $P < 0.001$).

3.6. LRP

The stimulus-locked and response-locked LRPs are shown in Fig. 5. A typical lateralization pattern was observed for the Go trials for PNT and AWT. Analyses of the onset latencies of LRPs assessed using the same procedure as for the N200 effects indicated that in the PNT the onset latency for the Go LRP was 552 ms in the Go = semantics condition and 448 ms in the Go = phonology condition. For the AWT, the onset latency for the Go LRP was 536 ms in the Go = semantics and 648 ms in the Go = phonology condition. Thus, the onset of the Go LRPs reflects the expected order of availability of semantic and phonological information in the two tasks.

Inspection of the noGo waveforms suggests that with the possible exception of the Go/noGo = phonology condition of the PNT no LRP is visible. Statistically, the noGo LRP from that condition diverged from baseline between 320 and 364 ms. While this condition is the one that would be expected to give rise to a noGo LRP, the latency of the effect seems too early in the light of the onset latencies in the

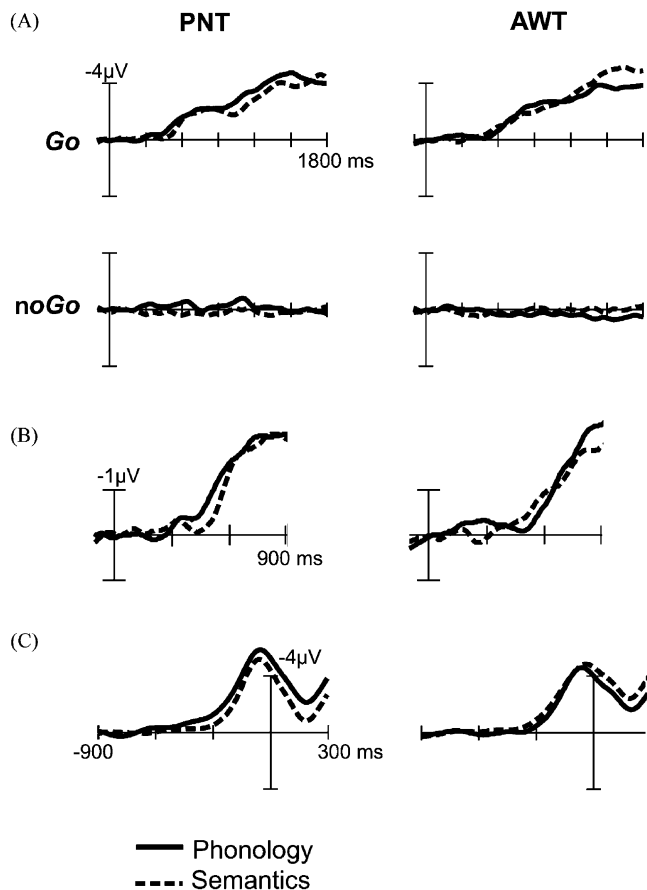


Fig. 5. (A) Stimulus-locked lateralized readiness potentials (LRPs); (B) stimulus-locked LRP for the Go trials plotted on another scale to illustrate the differences in onset latency; (C): response-locked LRP. Labels phonology/semantics refer to the information type the Go/noGo response was based on. Because of the dual task nature of the experiment, this means that if the Go response was determined by phonology, the choice of response hand was determined by semantics and vice versa.

Go LRP of the same condition (448 ms) and reported noGo LRP latencies in related studies. We therefore are reluctant to interpret this effect any further. On visual inspection, there appear to be two additional divergences from baseline in this condition (between 500–700 and 1000–1200 ms, respectively). These fail to reach significance, however.

4. Discussion

The main question of the present study concerns the temporal organization of language processing during production, the meaning-to-sound route, and during comprehension, the sound-to-meaning route. Of particular interest is the relative order of the conceptual and phonological stages and the extent to which they might overlap or proceed in parallel.

For the route from meaning-to-sound, the N200-noGo results from our PNTs replicated previous findings on the time

course of semantic and phonological encoding based on the LRP [42,46,47]. The mean peak and onset latencies for the Go/noGo = semantics condition occurred earlier than those for the Go/noGo = phonology condition, supporting the hypothesis that semantic information was processed prior to phonological information. In line with previous studies we therefore conclude that these data are most compatible with serial (e.g. [25]) or cascading models [3] of speech production. As in those previous studies, electrophysiological measures revealed a much larger timing difference than the overt RT measure (N200 peak latency: 173 ms difference, RT: 33 ms difference). Note also, that RT in the present dual choice design necessarily reflects the final decision after both types of information, phonological and semantic, are processed. Thus, pronounced timing differences for the different types of information at intermediate stages of analysis are diluted during further processing in the present design with respect to RT.

Based on models of spoken word recognition, e.g. the cohort model [28], we expected less of a separation between the processing of phonological and semantic information. Still, in as far as one type of processing did precede the other, we expected the processing of phonological information to occur earlier. This temporal order of processes was in fact what we observed as the N200 in the Go/noGo = phonology condition peaked at 499 ms and the N200 in the Go/noGo = semantics condition had its maximum at 591 ms. The two conditions differed significantly for about 84 ms. This index of the precedence of phonological information in the auditory task can be contrasted with the 173 ms precedence of semantic information in the production task.²

Thus, while production and comprehension showed the expected inverse order of semantic and phonological encoding, the two language tasks differed in the temporal distance between the engagement of the two encoding processes, as reflected in the extent of overlap between the associated N200 effects. Whereas the phonological and semantic N200 effects during implicit picture naming were clearly separated in time, those in the auditory task overlap to a much larger extent. Note, that the further interpretation of these data relies on the reasonable assumption that phonological and se-

² The difference in amplitude of the N2 component in the AWT and PNT is puzzling, because it should show modality independence, if it indeed reflects inhibitory processes in the frontal lobe. Such modality differences have been reported before (e.g. [7]) and have been addressed systematically in a recent study by Falkenstein et al. [8]. While the auditory N2 component was smaller than the visual one, it showed the same difference between good and poor subjects, i.e. it varied as a function of the number of false alarms. The authors therefore concluded that the N2 is indeed an index of motor inhibition in both modalities. With regard to the modality difference in amplitude, two suggestions were made: (a) inhibition might have different strengths in the two modalities, (b) the inhibition mechanism reflected in the N2 may arise from modality specific generators. As the false alarm rates were similar in both tasks in the current studies and the scalp maps (see Fig. 2) showed a similar distribution of the N2 effects, we are unable to decide, which of the explanations applies.

semantic encoding take about the same time in the two modalities. On the basis of this assumption, we take the difference in overlap of the N200 effects as a reflection of a difference in the temporal and thus probably also of the functional organization of the meaning-to-sound and sound-to-meaning routes. Specifically, the wide temporal separation of the N200 effects during picture naming implicates serial access to semantic and phonological information, in line with current models of speech production [6,25]. In contrast, the more substantial overlap of the waveforms in the auditory task suggests that phonological and semantic information are processed close in time and that processing is carried out at least partially in parallel. The data are thus compatible with psycholinguistic and modeling evidence regarding the timing of information during speech comprehension [2,6,31].

In summary, this study is the first to compare the electrophysiological measures of the use and timing of semantic and phonological information in preparation for speaking (implicit picture naming) and during auditory word comprehension. Once again, the N200 effect in combination with a Go/noGo paradigm was shown to be a powerful tool for the temporal dissection of language processing. Further ERP experiments using manipulations of task difficulty in the two information domains (semantic, phonological) will provide a more stringent test of the serial/parallel processing distinction.

Acknowledgements

ARF was supported by a Post-doctoral Fellowship from the Spanish Government. Supported by DFG grant MU1311/7-1 to TFM and HD22614 and AG08313 to MK.

References

- [1] Casey BJ, Trainor RJ, Orendi J, Schubert A, Nystrom L, Giedd J, et al. A developmental functional MRI study of prefrontal activation during performance of a Go–noGo task. *J Cognitive Neuroscience* 1998;9:835–47.
- [2] Cutler A, Clifton C. Comprehending spoken language: a blueprint of the listener. In: Brown CM, Hagoort P, editors. *The neurocognition of language*. Oxford: Oxford University Press 1999;123–166.
- [3] Dell GS. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 1986;93:283–321.
- [4] Dell GS, O'Seaghdha PG. Stages of lexical access in language production. *Cognition* 1992;42:287–314.
- [5] Eimer M. Effects of attention and stimulus probability on ERPs in a Go/noGo task. *Biological Psychology* 1993;35:123–38.
- [6] Elman JL, McClelland JL. Speech perception as a cognitive process: the interactive activation model. In: Lass N, editor. *Speech and language*, vol. 10. New York: Academic Press 1984;337–374.
- [7] Falkenstein M, Koshlykova NA, Kiroj VN, Hoorman J, Hohnsbein J. Later ERP components in visual and auditory Go/noGo tasks. *Electroencephalography and Clinical Neurophysiology* 1995;96:36–43.
- [8] Falkenstein M, Hoorman J, Hohnsbein J. ERP components in Go/noGo tasks and their relation to inhibition. *Acta Psychologica* 1999;101:267–91.
- [9] Fuster JM. *The prefrontal cortex. Anatomy, physiology and neuropsychology of the frontal lobe*. New York: Raven Press 1989.
- [10] Garavan H, Ross TJ, Stein EA. Right-hemispheric dominance of inhibitory control: an event-related functional MRI study. *Proceedings Natural Academy of Science* 1999;96:8301–6.
- [11] Goldinger SD, Luce PA, Pisoni DB, Marcario JK. Form-based priming in spoken word recognition: the roles of competition and bias. *J Experimental Psychology: Learning, Memory and Cognition* 1992;18:1211–38.
- [12] Grosjean F. Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 1980;28:267–83.
- [13] Holcomb PJ, Neville H. Natural speech processing: an analysis using event-related brain potentials. *Psychobiology* 1991;19:286–300.
- [14] Jennings JR, Wood CC. The e-adjustment procedure for repeated-measures analysis of variance. *Psychophysiology* 1976;13:277–8.
- [15] Jodo E, Kayama Y. Relation of a negative ERP component to response inhibition in a Go/noGo task. *Electroencephalography and Clinical Neurophysiology* 1992;82:477–82.
- [16] Kempen G, Huijberts P. The lexicalization process in sentence production and naming: indirect election of words. *Cognition* 1983;14:185–209.
- [17] Kok A. Effects of degradation of visual stimuli on components of the event-related potential (ERP) in Go/noGo reaction tasks. *Biological Psychology* 1986;23:21–38.
- [18] Konishi S, Nakajima K, Uchida Y, Kikyo H, Kameyama M, Miyashita Y. Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain* 1999;122:981–91.
- [19] Kutas M, Donchin E. Studies of squeezing: Handedness, responding hand, response force, and asymmetry of readiness potential. *Science* 1974;186:545–8.
- [20] Kutas M. Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology* 1997;34:383–98.
- [21] Kutas M, van Petten CK. Psycholinguistics electrified: event-related brain potential investigations. In: Gernsbecher MA, editor. *Handbook of psycholinguistics*. San Diego CA: Academic Press, 1994;83–143.
- [22] Levelt WJM. *Speaking: from intention to articulation*. Cambridge, MA: MIT Press 1989.
- [23] Levelt WJM, Schriefers H, Vorberg D, Meyer AS, Pechmann T, Havinga J. The time course of lexical access in speech production: A study of picture naming. *Psychological Review* 1991;98:122–42.
- [24] Levelt WJM, Praamstra P, Meyer A, Helenius P, Salmelin R. An MEG study of Picture naming. *J Cognitive Neuroscience* 1998;10:553–67.
- [25] Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 1999;22:1–75.
- [26] Luria AR. *The Working Brain*. London: Penguin 1973.
- [27] Marslen-Wilson W. Linguistic structure and speech shadowing at very short latencies. *Nature* 1973;244:522–3.
- [28] Marslen-Wilson W. Function and process in spoken word recognition: a tutorial review. In Bouma H, Bouwhuis DG, editors. *Attention and performance X: control of language processes*. Hillsdale, NJ: Erlbaum 1984. p. 125–150.
- [29] Marslen-Wilson W. Functional parallelism in spoken word-recognition. *Cognition* 1987;25:71–102.
- [30] Marslen-Wilson W, Welsh A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 1978;10:29–63.
- [31] Marslen-Wilson W, Tyler LK. The temporal structure of spoken language understanding. *Cognition* 1980;8:1–71.
- [32] McCarthy G, Wood C. Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology* 1985;62:203–8.
- [33] McQueen JM, Norris DG, Cutler A. Competition in spoken word recognition: spotting words in other words. *J Experimental Psychology: Learning, Memory and Cognition* 1994;20:621–38.

- [34] Miller J, Riehle A, Requin J. Effects of preliminary perceptual output on neuronal activity of the primary motor cortex. *J Experimental Psychology: Human Perception and Performance* 1992;18:1121–38.
- [35] O'Rourke TB, Holcomb PJ. Electrophysiological evidence for the efficiency of spoken word processing. *Language and Cognitive Processes*, in press.
- [36] Petersen RR, Savoy P. Lexical selection and phonological encoding during language production. Evidence for cascaded processing. *J Experimental Psychology: Learning, Memory and Cognition* 1998;24:539–57.
- [37] Pfefferbaum A, Ford J, Weller B, Kopell B. ERPs to response production and inhibition. *Electroencephalography and Clinical Neurophysiology* 1985;60:423–34.
- [38] Potter MC. Representational buffers: the see-mind hypothesis in picture perception, reading, and visual search. In: Rayner K. editor. *Eye Movements in Reading: Perceptual and Language Processes*. New York: Academic Press, 1983:423–437.
- [39] Roelofs A. A spreading-activation theory of lemma retrieval in speaking. *Cognition* 1992;42:107–42.
- [40] Sasaki K, Gemba H. Electrical activity in the prefrontal cortex specific to noGo reaction of conditioned hand movement with color discrimination in the monkey. *Experimental Brain Research* 1986;64:603–6.
- [41] Schriefers H, Meyer AS, Levelt WJM. Exploring the time course of lexical access in language production: Picture-word interference studies. *J Memory and Language* 1990;29:86–102.
- [42] Schmitt BM, Münte T, Kutas M. Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology* 2000;37:473–84.
- [43] Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. *Nature* 1996;381:520–2.
- [44] Tyler LK, Wessels J. Is gating an on-line task? Evidence from naming latency data. *Perception and Psychophysics* 1985;38:217–22.
- [45] Van Petten C, Coulson S, Rubin S, Plante E, Parks M. Time course of word identification and semantic integration in spoken language. *J Experimental Psychology: Learning, Memory and Cognition* 1999;25:394–417.
- [46] Van Turennout M, Hagoort P, Brown CM. Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *J Experimental Psychology: Learning, Memory and Cognition* 1997;23:787–806.
- [47] Van Turennout M, Hagoort P, Brown CM. Brain activity during speaking: From syntax to phonology in 40 ms. *Science* 1998;280:572–4.
- [48] Winer BJ, Brown DR, Michels KM. *Statistical principles in experimental designs*. 3rd ed. New York: McGraw-Hill 1991.
- [49] Zwitserlood P. The locus of the effects of sentential-semantic context in spoken word processing. *Cognition* 1989;32:25–64.