

Chapter 29

THE ELECTROPHYSIOLOGY OF SPEAKING: POSSIBILITIES OF EVENT-RELATED POTENTIAL RESEARCH ON SPEECH PRODUCTION

Peter Hagoort, Miranda van Turenhout

In this chapter we discuss problems and possibilities of using electrophysiological recordings to study speaking. First we introduce the method of recording Event Related Potentials (ERPs). More in particular we discuss the so-called Readiness Potential (RP), an ERP component that precedes voluntary movements. We conclude that contamination of the EEG signal with EMG activity of the muscles involved in articulation and inconclusive results in earlier RP studies preclude strong conclusions on the basis of ERP recordings about the areas of the brain that supervise or dominate the articulation of speech. We argue that, rather than in localizing the brain regions involved in speaking, the importance of the electrophysiological recordings is related to the high temporal resolution of the ERP signal. This characteristic allows us to map out the time course of the retrieval of lexical information preceding overt articulation. We discuss an experimental paradigm that we used for estimating the time course of the retrieval of semantic, syntactic and phonological word information. On the basis of our results we present a rough estimate of the time course of the processes involved in lexical access during speaking. We conclude that, in principle, this experimental paradigm can be helpful in determining the locus of nonfluencies in speaking, provided that this locus precedes articulation.

INTRODUCTION

One way of recording brain activity related to language processing is by recording event-related brain potentials (ERPs). Human ERPs are recorded from a number of electrodes placed on the scalp. Scalp recording of ERPs provides a non-invasive technique for evaluating electrophysiological activity related to stimulus processing and response preparation.

As can be seen in Figure 1, ERPs are usually recorded time-locked to and averaged over a number of stimulus events. In the resulting averaged ERP waveform one can see a number of distinct ERP components emerge. These ERP components are often labeled on the basis of their polarity (P for positive, N for negative) and either their latency in milliseconds (e.g., P300) or their sequential number in the series of peaks (e.g., P3). For example, the P300 (or P3) is a late positive wave with a typical peak latency around 300 ms post-stimulus, that occurs when an infrequent target stimulus is detected in a series of frequent standard stimuli.

An example of a psycholinguistically more interesting ERP component is the N400.

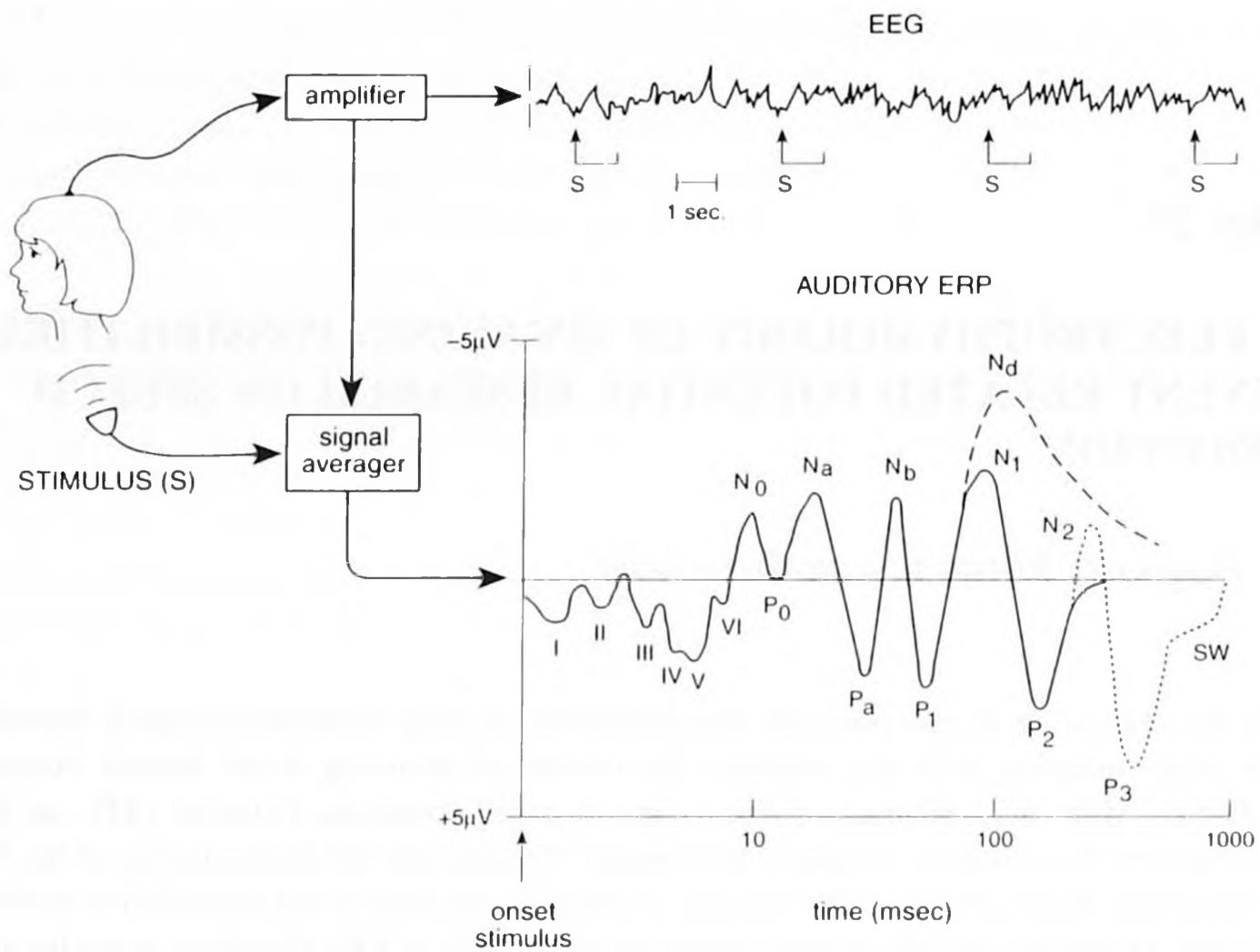


Figure 1 (after Hillyard & Kutas, 1983). Idealized waveform of a series of ERP components that become visible after averaging the EEG to repeated presentations of a short auditory stimulus. Usually, averaging over a number of stimulus tokens is required to get an adequate signal-to-noise ratio. Along the logarithmic time axis the early brainstem potentials (Waves I-VI), the midlatency components (No, Po, Na, Pa, Nb), the largely exogenous components (P1, N1, P2), and the endogenous, cognitive ERP components (Nd, N2, P300, Slow Wave) are shown. The components with a negative polarity are plotted upwards, the components with a positive polarity are plotted downwards.

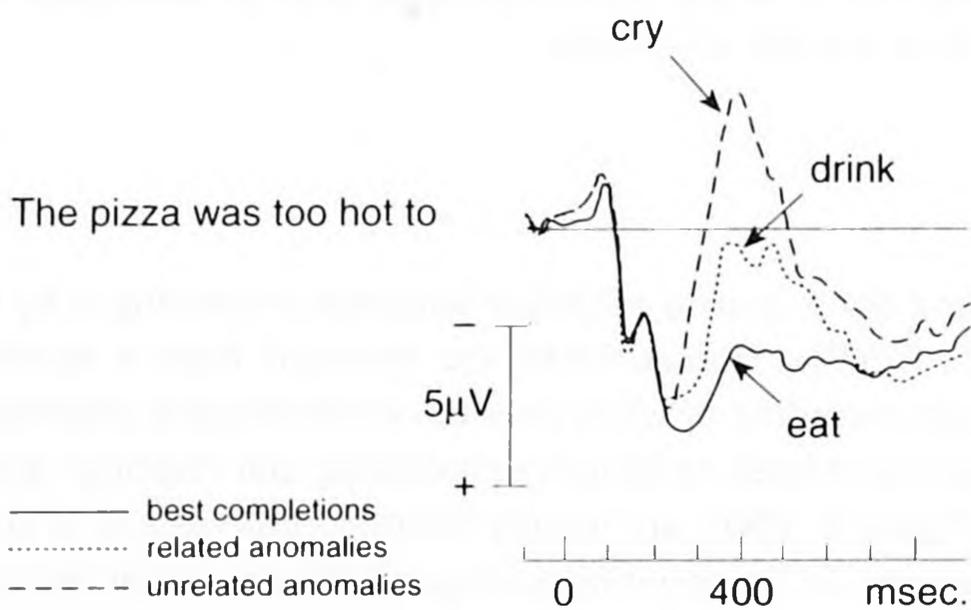


Figure 2 (after Kutas & Hillyard, 1980): ERP to the sentence-final words for visually presented sentences. The sentence-final semantic anomaly (i.e., *socks*) results in a substantial negative shift relative to the ERP elicited by the semantically expected ending (i.e., *butter*). This negative shift is known as the N400-effect.

This component can be elicited by a mismatch between the semantics of the sentence context and the lexical meaning of a particular word, as in the sentence "He spread his warm bread with *socks*." (Kutas & Hillyard, 1980). Compared to the correct sentence-final word *butter*, these semantic anomalies elicit a large negative shift that onsets at about 250 ms after the semantic anomaly and reaches its maximal amplitude at about 400 ms.

The general advantage of ERPs over brain imaging techniques such as PET and fMRI is its millisecond temporal resolution. Especially for rapid, transient processes such as speaking or listening to language, this aspect of ERPs allows the investigation of a central aspect of language processing, namely its time course. However, a disadvantage of ERPs is that especially for the later components such as P300 and N400, the localization of the neural generators that contribute to the surface potentials recorded at the scalp is still problematic.

So far, we have discussed ERPs that are elicited by and follow the occurrence of a particular stimulus event, such as the presentation of tones or words. However, it is also possible to record ERPs that precede a response. The most well-known response related potential is the so-called Readiness Potential (or *Bereitschaftspotential*). This potential was first discovered by Kornhuber and Deecke (1965). These authors recorded brain potentials over the left and right motor cortex, before and during voluntary movement of the left hand. As can be seen in Figure 3, about 1 second before the actual hand movement can be registered in the EMG, a slow negative going potential appears in the EEG signal that reaches its maximum just after movement onset. Moreover, in the final phase before movement onset this negativity becomes larger for sites contralateral to the moving hand; that is, at sites over the motor cortex that are known to be involved in the initiation of the movement. Similar larger contralateral negativities have been obtained for arm and finger movements (Vaughan et al., 1968).

Readiness Potential and speech

At first sight, the Readiness Potential (RP) seems to be appropriate for studying aspects

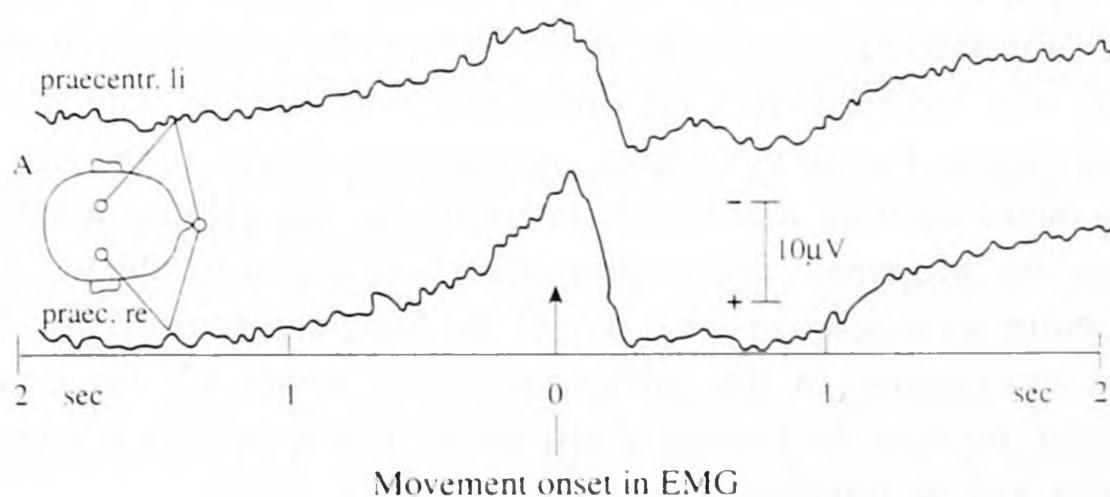


Figure 3 (after Kornhuber & Deecke, 1965). Readiness Potential or *Bereitschaftspotential* during voluntary movement of the left hand. The negative going potential during movement preparation is larger over the contralateral right hemisphere. Zero is the onset of movement in the electromyogram.

of the articulatory movements of speech, with potential applications for studying disorders of voluntary movement in for instance apraxia of speech and stuttering. However, the results of studies examining the RP in oral speech and non-speech gestures are less clear-cut (cf. Wohlert, 1993).

One serious problem that plagues ERP studies on speech production is the contamination of the ERP signal with the EMG activity of the muscles producing the articulatory gestures. Due to the proximity of the relevant orofacial muscles to the EEG recording sites, EMG activity is a much greater artifact source in the case of oral movements than in the case of finger or toe movements. Brooker and Donald (1980), for instance, observed that EEG recorded at inferior frontal lobe sites correlated with the EMG activity recorded from the masseter and temporalis muscles. Moreover, these and other muscles (e.g., the orbicularis oris) showed activity up to 500 ms before the vocalization trigger. Also it has been reported that tongue movements can contribute to scalp-recorded potentials (Szirtes & Vaughan, 1977).

The presence of major artifact sources in recording potentials related to speaking is clearly one of the main reasons why, compared to ERP studies on language comprehension, the number of ERP studies on speaking is very limited indeed.

An additional problem is that the few available studies do not provide an unequivocal pattern of results. Deecke, Engel, Lang, and Kornhuber (1986) examined the RP preceding the utterances of single words. To avoid the most problematic artifact sources, early facial EMG activity was chosen as the trigger for backwards averaging. This procedure is claimed to result in EEG activity not contaminated with articulatory EMG activity. In their study, Deecke et al. (1986) obtained an initial bilateral RP that lateralized over the left hemisphere during the last 100 msec preceding speech onset. They interpreted these results as suggesting that both motor cortices are involved in the initiation of articulation (the early bilateral RP), but that the final execution of the articulatory gestures was dominated by the left motor cortex (the lateralized component).

In contrast to the findings of Deecke and colleagues, no such lateralization of the RP was seen in a study by Wohlert (1993; see also Wohlert & Larson, 1991), who examined the RPs preceding three types of oral gestures: (i) a nonspeech gesture (lip press); (ii) a basic phonemic gesture (lip rounding, as in an unvoiced /u/; (iii) the production of the spoken word "pool". The EMG was recorded from the orbicularis oris superior, which is an important muscle for lip movements. For all three tasks, symmetrical activity in the RPs recorded over the left and right motor cortices was found. Interestingly, the greatest amplitude was seen over Cz, a midline site. Moreover, this midline effect was largest for the word task. Wohlert interpreted this finding as suggestive for a role of the Supplementary Motor Area (SMA) in rapid, continuous movements. In this interpretation the vertex RP for producing speech movements is largest, because the amount of sequenced muscle activity is greater in speaking than in the lip press and lip rounding movements (Wohlert, 1993).

In short, studies on the RP preceding overt speech activity have not resulted in consistent findings with respect to the brain areas that supervise or dominate the process of articulation in normal subjects.

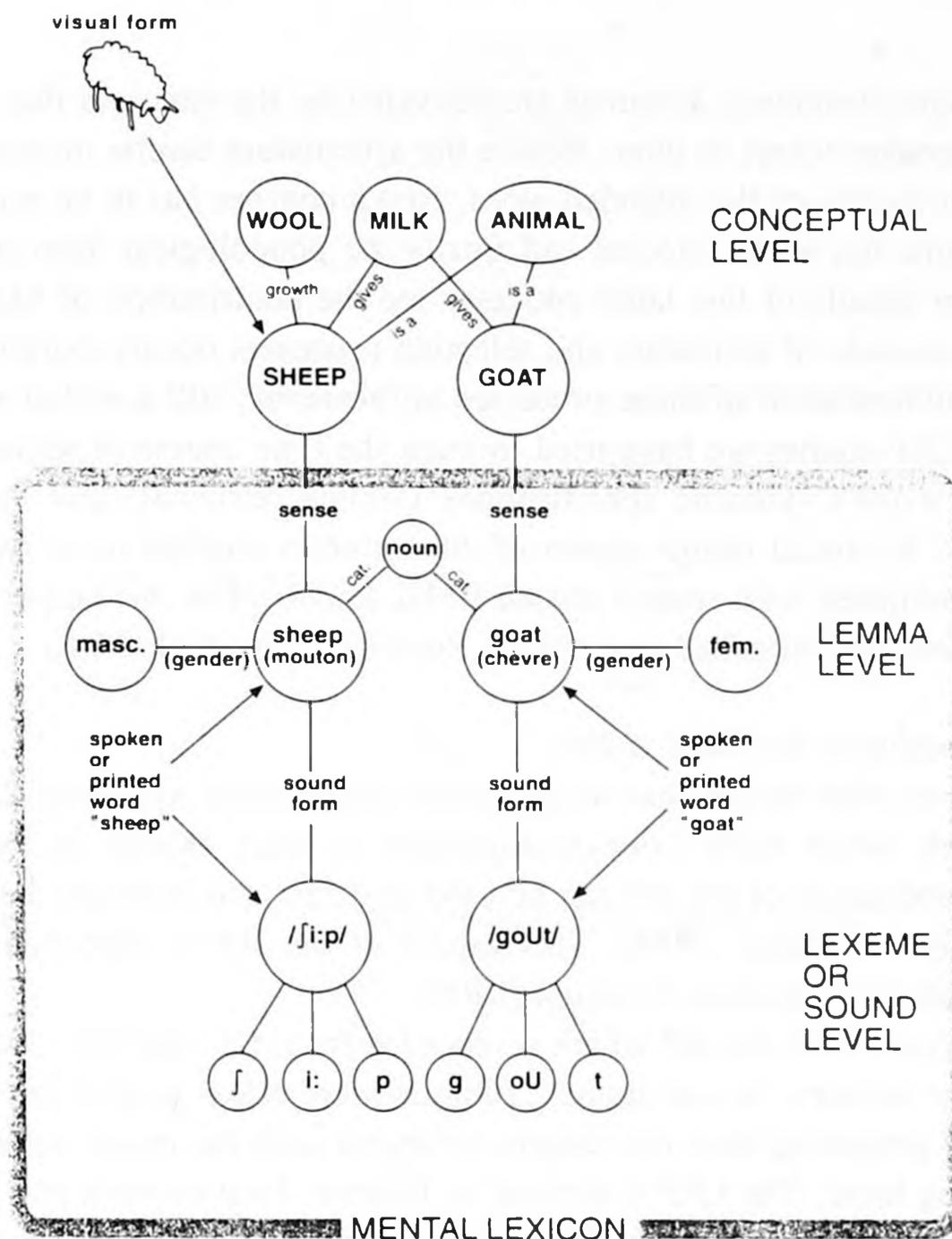


Figure 4. The Levelt and Roelofs model for speaking. Concept nodes (SHEEP) are activated on the basis of sensory and/or conceptual input. Activation from a concept node spreads to its lemma node (*sheep*) in the mental lexicon. Each concept node is linked to exactly one lemma in the lexicon. At the lemma level the syntactic word information is specified, such as grammatical gender and word class. For instance, in French the gender of the lemma *sheep* is male and the gender of the lemma *goat* is female. After the lemma has been selected, word form information is retrieved and prepared for articulation.

For the time being, this severely limits the possibilities of fruitfully using the RP in research on speech related movement disorders.

In the remainder we will discuss the approach that we developed to study the process of speaking (van Turenout et al., 1997). In our approach we were not so much interested in establishing the brain areas that are involved in speaking, but rather in using the superior temporal resolution of the ERP signal to examine the time course of speaking.

The issue: Lexical access in speech production

Figure 4 specifies the Levelt and Roelofs model for word production (cf. Levelt, 1989; Roelofs, 1992). According to this model the lexicon contains syntactic word nodes (lemmas)

and word form nodes (lexemes). Lemmas are activated by the concepts that are part of the message that the speaker wants to utter. Before the articulators can be instructed to actually produce the speech sounds of the intended word, first a concept has to be activated, then its corresponding lemma has to be selected and finally the phonological form of the word has to be retrieved (for details of this latter process, see the contribution of Meyer). In fluent speech this whole cascade of activation and selection processes occurs extremely rapid. The precise temporal orchestration of these processes is, however, still a matter of debate.

In a series of ERP studies we have tried to track the time course of semantic activation, the retrieval of a word's syntactic specifications (lemma retrieval), and the phonological encoding of words. A crucial design aspect of these studies enabled us to record ERPs that could not be contaminated with speech related EMG activity. For this purpose we used one particular ERP effect, the so-called Lateralized Readiness Potential (LRP).

The Lateralized Readiness Potential (LRP).

It has been shown that the RP that we discussed above starts to lateralize as soon as the subject knows with which hand (s)he is supposed to react (Kutas & Donchin, 1980). Therefore, the lateralization of the RP can be used to detect and measure the preparation of a specific response (cf. Coles, 1989). This aspect of the RP is exploited for using and deriving the Lateralized Readiness Potential (LRP).

The LRP is derived from the RP which is recorded from C3' and C4', located above the left and right motor cortices. In our studies, subjects were asked to give left and right hand responses. The RP preceding hand movements is largest over the motor cortex contralateral to the corresponding hand. The LRP is derived as follows: First on each trial, the difference is obtained between the potentials recorded from C3' and C4'. These difference waveforms are averaged separately for trials in which the left versus the right hand is cued. Second, to cancel out lateralized potentials that are not specifically related to response preparation, the waveform obtained for the left-hand trials is subtracted from the waveform obtained for the right-hand trials. The resulting LRP reflects the average amount of lateralization occurring as a result of the preparation of the hand response (see Figure 5). The LRP deviates from the baseline in upward direction as soon as response preparation for the cued response hand occurs.

A finding in the LRP literature that is crucial for our purposes is that the LRP starts to develop as soon as relevant stimulus information is used for response preparation (Coles, 1989; De Jong et al., 1988; Miller & Hackley, 1992; Osman et al., 1992). This makes the LRP a real-time measure of the moments in time at which different kinds of information influence the preparation of a response.

The LRP and lexical access during speaking

We will illustrate the way in which we used this paradigm for testing an assumption in most models of language production, including the Levelt and Roelofs model. This assumption is that the meaning of a word is available before its phonological form can be

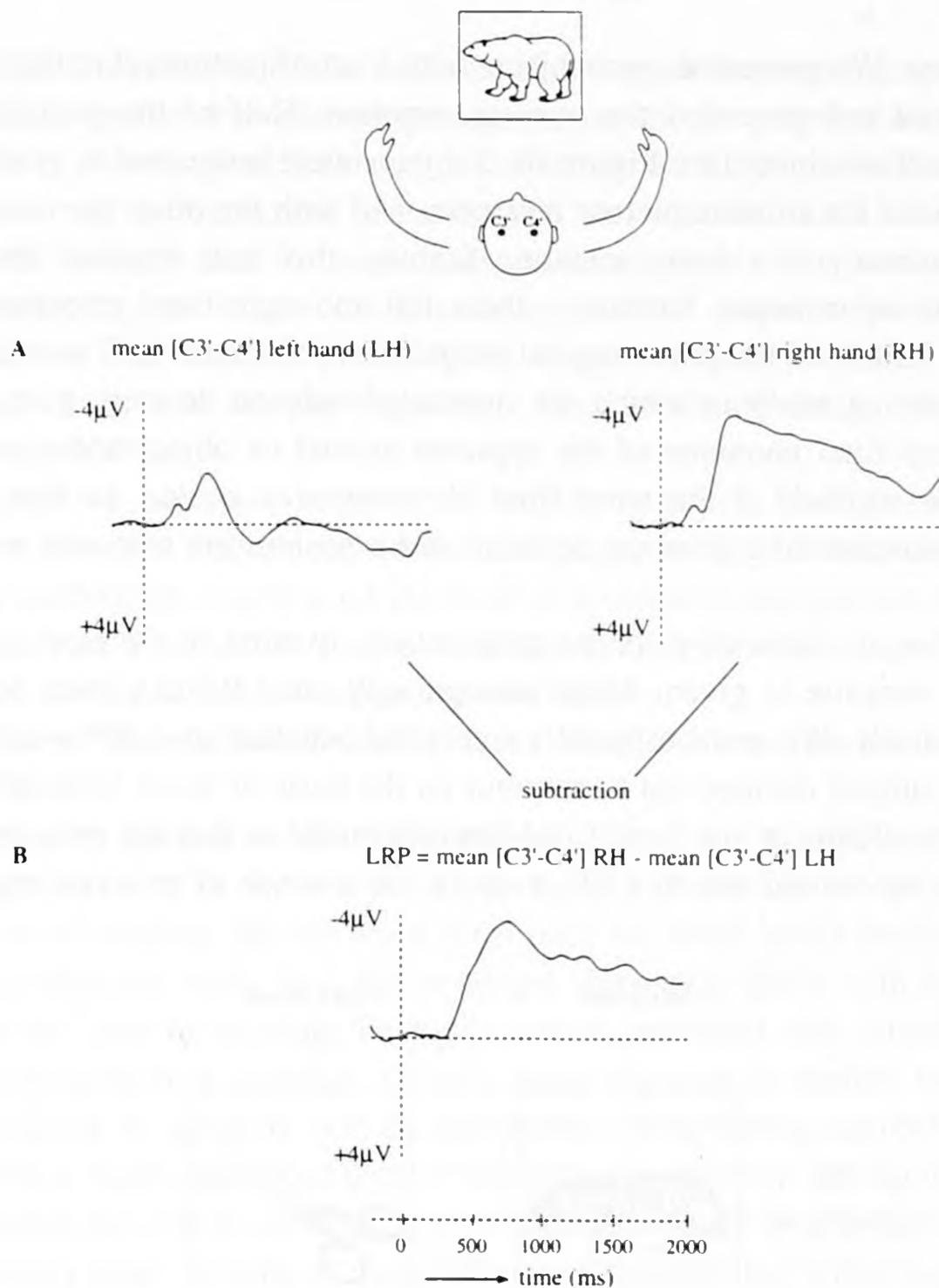


Figure 5. Derivation of the Lateralized Readiness Potential (LRP). The LRP is derived from the Readiness Potential (RP) which is recorded from C3' and C4', located above the left and right motor cortices. In this example subjects are presented with target pictures that either cue a left-hand or a right-hand response. The RP preceding hand movements is largest over the motor cortex contralateral to the corresponding hand. The LRP is derived as follows: First on each trial, the difference is obtained between the potentials recorded from C3' and C4'. These difference waveforms are averaged separately for trials in which the left versus the right hand is cued (a). Second, to cancel out lateralized potentials that are not specifically related to response preparation, the waveform obtained for the left-hand trials is subtracted from the waveform obtained for the right-hand trials (b). The resulting LRP reflects the average amount of lateralization occurring as a result of the preparation of the hand response. The LRP deviates from the baseline in upward direction as soon as response preparation for the cued response hand occurs.

retrieved. Tested in an LRP paradigm it implies that conceptual-semantic information should be transmitted earlier to the motor system than phonological information, provided that these sources of information are relevant for making a response.

The response relevance of both semantic and phonological information was guaranteed

in the following way: We presented our subjects with a set of pictures that they had to name. However, the crucial task preceded the naming response. Half of the pictures depicted an animal, the other half an object (see Figure 6). Subjects were instructed to give a pushbutton response with one hand for animate picture referents, and with the other for inanimate picture referents. Since animacy is a basic semantic feature, this task requires the retrieval of conceptual-semantic information. Crucially, these left and right hand responses were made conditional on the nature of the phonological word form information. The materials shown in Figure 6 are from a study in which we instructed subjects to only give a pushbutton response if the word-final phoneme of the depicted animal or object ended with an /r/. A response should be withheld if the word-final phoneme was an /n/. In this way the task consisted of a conjunction of a go-nogo decision and a pushbutton response with the left or right hand.

Since the LRP is an index of response preparation, it starts to develop well before the actual pushbutton response is given. More interestingly, an LRP can even occur when no response is given at all. We were especially interested whether an LRP would develop on trials in which the subject decided not to respond on the basis of word form information (the nogo trials). The prediction of the Levelt and Roelofs model is that not only on go trials but also on nogo trials we should see an LRP, even in the absence of an overt response.

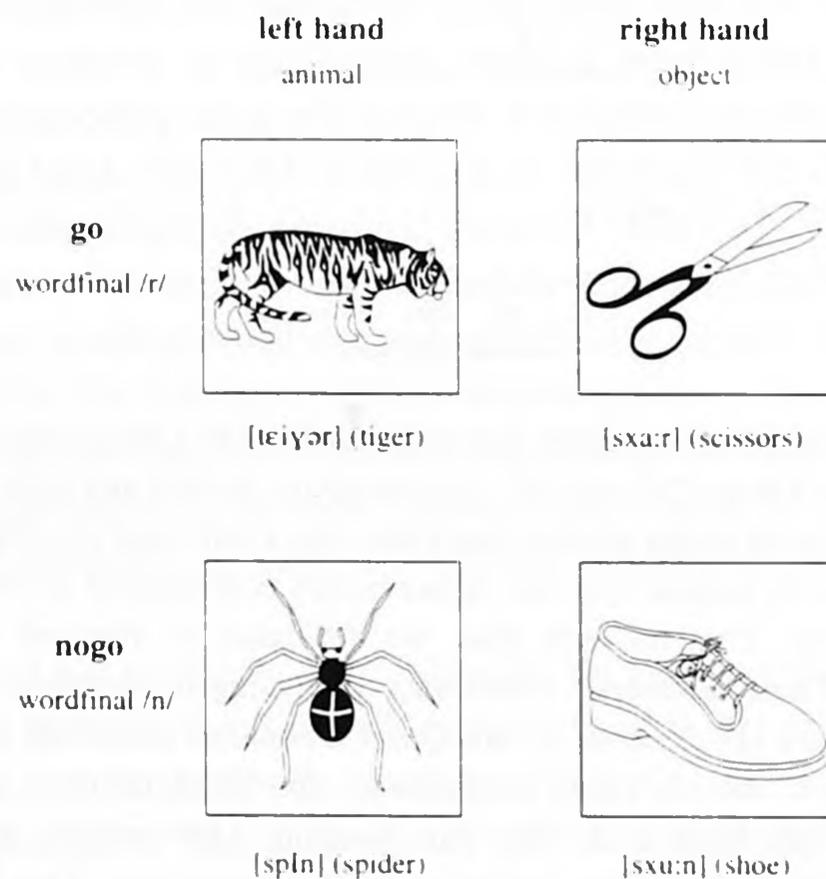


Figure 6 (after van Turenout et al., 1997). Examples of the pictures used in one of our LRP studies. In the figure, the Dutch picture names and their English translations are shown below the pictures. Dutch picture names are in the International Phonetic Alphabet. The four pictures depicted here represent separate trials for the four conditions in the experiment. In the example, an animal cues a left-hand response, and an object cues a right-hand response. The response has to be executed if the picture name ends with an /r/ (go-trials), but withheld if it ends with an /n/ (nogo-trials).

The logic behind this prediction is as follows: If during picture naming semantic information becomes available earlier than phonological information, the preparation of a hand response can start before phonological information informs the subject about whether or not to respond. This is exactly the pattern of results that we observed. An LRP developed not only for go trials, but initially also for nogo-trials, in the absence of an overt response. The early availability of semantic information enabled response preparation, but when information about the word's phonological form became available, further response preparation was overruled on the nogo-trials (for details of design and results, see van Turennout et al., 1997).

In our studies we asked our subjects not only to determine their go-nogo decision on the word-final phoneme, but also on the word-initial phoneme. In addition we asked subjects not only to give a pushbutton response on the basis of a semantic component (i.e., animacy), but also on the basis of a syntactic feature of words (i.e., grammatical gender). Together the results of these studies allow the following conclusions:

(i) The retrieval of both semantic and syntactic word features precedes the retrieval of its phonological form.

(ii) The information about a word's phonological form is not available at once, but accrues in a left-to-right order. For the words in our study (on average 4.5 phonemes) it took an additional 80 ms to retrieve the full word form once the word-initial phoneme was available.

The results obtained with the LRP paradigm show that ERPs with their high temporal resolution can be used to observe the rapid mental processes that underly speaking, even before overt articulation is initiated. Clearly more research is needed before a temporally finegrained analysis of speaking will be completed. Nevertheless current ERP data already allow us to give a rough approximation of the time course from perception to articulation.

Let us assume that while walking in your back garden all of a sudden you stand eyes in eyes with a grizzly bear. Among all your behavioral options, the following two belong to the more probable ones. One reaction is to run away. An alternative one is to say: "A bear". For our purposes we are only interested in the second behavior. A recent ERP study suggests that it takes roughly 150 ms to perceive and categorize a picture as an animal (Thorpe et al., 1996). On the basis of this finding we think it is a reasonable guess that the activation of the concept BEAR on the basis of the visual input in your garden is less than 200 ms. On the basis of our own results we know that it takes about 600 ms before articulation of the word *bear* starts. The period in between (roughly 400 ms) has to do with the cascade of activation and selection processes of lemma and word form information. Our results indicate that the retrieval of the full phonological form of a word like *bear* takes in the order of 120 ms once the conceptual-semantic information has been activated. The remaining time is necessary for lemma selection and preparing the articulatory programme on the basis of the phonological information (for a comparable estimation and more details see Levelt et al., submitted).

Perspectives for further research

Clearly the LRP paradigm that we developed seems useful for tracking the time course

of the processes of speaking that precede overt articulation. However, measuring the parameters of the actual execution of the articulatory gestures does not lend itself easily to a fruitful application of the ERP-technique. Until reliable algorithms become available for correcting EMG activity out of the EEG signal, the artifact problems seem too severe for investigating the speech motor production directly with the help of this technique. This, however, does not invalidate electrophysiological methods for studying nonfluencies in speech production. Since nonfluencies can arise and indeed are often claimed to arise at levels preceding overt articulation (Postma & Kolk, 1993; Wingate, 1988), in principle the LRP paradigm can be used to get a fairly precise estimation of the stages in speech production at which nonfluencies might arise. As such research on nonfluencies in speech production could benefit from exploiting the kind of paradigm that we discussed.

REFERENCES

- Brooker, B.H., & Donald, M.W. (1980). Contribution of the speech musculature to apparent human EEG asymmetries prior to vocalization. *Brain and Language*, 9, 256-266.
- Coles, M.G.H. (1989). Modern mind-brain reading: Psychophysiology, physiology, and cognition. *Psychophysiology*, 26, 251-269.
- Deecke, L., Engel, M., Lang, W., & Kornhuber, H.H. (1986). Bereitschaftspotential preceding speech after holding breath. *Experimental Brain Research*, 65, 219-223.
- Hillyard, S.A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology*, 34, 33-61.
- Jong, de R., Wierda, M., Mulder, G., & Mulder, L.J.M. (1988). Use of partial information in responding. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 682-692.
- Kornhuber, H.H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflüger's Archive*, 284, 1-17.
- Kutas, M., & Donchin, E. (1980). Preparation to respond as manifested by movement-related brain potentials. *Brain Research*, 202, 95-115.
- Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (submitted). A theory of lexical access in speech production
- Miller, J., & Hackley, S.A. (1992). Electrophysiological evidence for temporal overlap among contingent mental processes. *Journal of Experimental Psychology: General*, 121, 195-209.
- Osman, A., Bashore, T.R., Coles, M.G.H., Donchin, E., & Meyer, D.E. (1992). On the transmission of partial information: Inferences from movement-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 217-232.
- Postma, A., & Kolk, H.H.J. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, 36, 472-488
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107-142.
- Szirtes, J., & Vaughan, H.G. (1977). Characteristics of cranial and facial potentials associated with speech production. In J.E. Desmedt (Ed.), *Language and hemispheric specialization in man: Cerebral ERPs* (pp. 112-126). Basel: Karger.

- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520-522.
- Turenout, van M., Hagoort, P., Brown, C.M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, in press.
- Vaughan, H.G., Costa, L.D., & Ritter, W. (1968). Topography of the human motor potential. *Electroencephalography and Clinical Neurophysiology*, *25*, 1-10.
- Wingate, M.E. (1988). *The structure of stuttering: A psycholinguistic analysis*. New York, NY: Springer-Verlag.
- Wohlert, A.B. (1993). Event-related brain potentials preceding speech and nonspeech oral movements of varying complexity. *Journal of Speech and Hearing Research*, *36*, 897-905.
- Wohlert, A.B., & Larson, Ch.R. (1991). Cerebral averaged potentials preceding oral movement. *Journal of Speech and Hearing Research*, *34*, 1387-1396.