

An MEG Study of Picture Naming

Willem J. M. Levelt, Peter Praamstra, and Antje S. Meyer

Max Planck Institute for Psycholinguistics, The Netherlands

Päivi Helenius and Riitta Salmelin

Helsinki University of Technology

Abstract

■ The purpose of this study was to relate a psycholinguistic processing model of picture naming to the dynamics of cortical activation during picture naming. The activation was recorded from eight Dutch subjects with a whole-head neuromagnetometer. The processing model, based on extensive naming latency studies, is a stage model. In preparing a picture's name, the speaker performs a chain of specific operations. They are, in this order, computing the visual percept, activating an appropriate lexical concept, selecting the target word from the mental lexicon, phonological encoding, phonetic encoding, and initiation of articulation. The time windows for each of these operations are reasonably well known and could be related to the peak activity of dipole sources in the individual

magnetic response patterns. The analyses showed a clear progression over these time windows from early occipital activation, via parietal and temporal to frontal activation. The major specific findings were that (1) a region in the left posterior temporal lobe, agreeing with the location of Wernicke's area, showed prominent activation starting about 200 msec after picture onset and peaking at about 350 msec, (i.e., within the stage of phonological encoding), and (2) a consistent activation was found in the right parietal cortex, peaking at about 230 msec after picture onset, thus preceding and partly overlapping with the left temporal response. An interpretation in terms of the management of visual attention is proposed. ■

INTRODUCTION

Salmelin, Hari, Lounasmaa, and Sams (1994) reported the first successful magnetoencephalograph (MEG) study of picture naming. The study demonstrated that picture naming is a feasible task in MEG studies of language production and also that a distinct dynamics of cortical activation is obtained that progresses bilaterally from the occipital visual cortex toward the temporal and frontal lobes. This dynamics must relate to the time course of processes underlying the naming of a picture, ranging from the visual analysis of the picture to the motor execution of the picture's name. Picture naming is an excellent choice as a language production task. It is a natural task, it involves all stages of language production at least to some extent, and importantly, it is one of the best-studied tasks in language production research. Picture naming is widely used to probe language function during brain surgery, as pioneered by Penfield and Roberts (1959) and Ojemann (1983). Preceding resective surgery for epilepsy or a cerebral tumor, the patient, under local anaesthesia, is tested for "language sites" in the relevant areas of the cortex. This is done by stimulation mapping (i.e., electrical stimulation of a particular brain area while the patient names a picture). If the naming response is erroneous or fully blocked, the area

is mapped as a language site. Although phonemic and semantic errors do occur during stimulation mapping, the typical result of stimulating a language site is full omission of the naming response. The two major reviews of this work (Ojemann, Ojemann, Lettich, & Berger, 1989; Haglund, Berger, Shamseldin, Lettich, & Ojemann, 1994), encompassing a total of 126 epilepsy and glioma patients, report an average of 2 to 3 detected language sites per patient (when 10 to 20 sites were tested). Although these sites were largely concentrated in the superior temporal gyrus, the motor strip, and premotor Broca's area of the (tested) left hemisphere, language sites also appeared in the medial temporal gyrus and the parietal and prefrontal cortices. Patients differed substantially in their "language maps." No single cortical area was *always* involved for *all* patients. These findings make it unlikely that picture naming has a uniform cortical organization across speakers. Although this is an important conclusion by itself, one would like to see it further qualified. The failure to name under stimulation (which defines a language site) may, after all, be due to interference with any of the processing stages involved in picture naming. Failing to recognize the object, categorize it semantically, retrieve the target word or its phonological form from the lexicon, or program the word's articulatory gesture will block the naming response. It is most

unlikely that a single language site performs all of these functions.

One would like to know how, for individual subjects, these components of the picture naming process are distributed over the brain. Only then we will begin to discover what is uniform across speakers and what is variable. The obvious first hunch is that the input and output processes in picture naming (perceiving the object, articulating the word) are much less variable in their cerebral localization than the mediating processes of lexical access and of phonological encoding. In particular, one should expect to find occipital lobe activation in a picture naming task, as well as activation in the face area of the sensory-motor cortex. But what about the more endogenous responses? In a positron emission tomography (PET) study of object naming and recognition, Price, Moore, Humphreys, Frackowiak, and Friston (1996) found the following areas involved with object recognition: ventral and dorsal regions of the left middle occipital lobe, bilateral anterior temporal lobes, and the left cerebellum. Martin, Wiggs, Ungerleider, and Haxby (1996), also in a PET study, found the left middle occipital lobe only active in the naming of animals but not of tools. This study as well as the one by Damasio, Grabowski, Tranel, Hichwa, and Damasio (1996) suggests that the areas involved in the semantic processing of pictures are different for different categories of pictures to be named (but the studies disagree on what areas are specifically involved in naming tools). It is one reason for us to publish the pictures used in our experiment (see Figure 7). Price et al. (1996) found that the areas involved in the naming part of the process were in the left perisylvian area, left anterior insula, and right cerebellum. Dronkers (1996) in a study of 25 stroke patients with articulatory motor planning deficits found the left insula, in particular its precentral gyrus, damaged in all of these patients (and in none of the control patients). Recently, Abdullaev and Posner (1997) reported a high-density (64 electrode) event-related potential (ERP) study of the well-known (but ill-understood) word generation task. In this time-course study of semantically generated words a left temporo-parietal activation, probably involving Wernicke's area, was observed around 650 msec after presentation of the stimulus word (in addition to the early left frontal response that is so typical for the verb generation task). These and other findings provide as many hypotheses about the cerebral areas involved in the various operations that mediate between the visual presentation of an object and the articulation of the object's name.

There are, essentially, two ways in which the cortical localization of component processes in picture naming can be pursued by means of MEG. The first one is to carefully map the psychological time course of picture naming (i.e., its staging in time) onto the speaker's recorded magnetic brain activity over time during picture

naming. In other words, one could try to localize the cortical areas that are active during successive stage of processing. The second one is more focused but also more risky. One can try to affect the duration of a single stage of processing (for instance, phonological word encoding) by means of an independent experimental variable and observe which sensors register a covariate magnetic response. It is more risky because factors affecting a single processing stage in picture naming usually have relatively small reaction time effects. Both methods, source analysis and the "single factor" method, require detailed knowledge about the time course of picture naming and the processing stages involved in that task. Fortunately, the psychology of picture naming is well developed.

A Process Analysis of Picture Naming

Any sophisticated use of picture naming in localization studies must be based on a processing model of the task itself. Luckily, picture naming has been the subject of myriad experimental studies (see Glaser, 1992, for a review) and we can now, with some confidence, distinguish the stages in the naming of a picture (see Figure 1) outlined in the following subsections.

Recognizing the Visual Object

During this stage a visual representation of the object—which we will call the *percept*—is computed from the visual image. The percept is entirely alinguistic but does involve such abstract properties as the object's typical size (an elephant is big and a mouse is small) and color (cf. Hart & Gordon, 1992). The object representation also has a particular orientation with respect to the perceiver. There is a "point of view" (Bülthoff & Edelman, 1992).

Accessing a Lexical Concept

To refer to the percept, the speaker will activate an appropriate *lexical concept*. Which concept is appropriate depends on the speaker's *perspective* in the communicative situation (Levitt, 1996). The same percept can be conceptualized as a VEHICLE, a CAR, a MERCEDES, etc. But there tend to be "basic level" conceptualizations, such as CAR, APPLE, FORK, COAT, VIOLIN (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). A subject is easily induced to operate on this basic level in picture naming. Lexical concepts can be conceived of as nodes in a semantic network with labeled arcs (Roelofs, 1992). For instance, VIOLIN has an *is-a* relation to MUSICAL INSTRUMENT, a *has-part* relation to STRING, etc. There is always some activation spreading from the target concept to semantically related concepts.

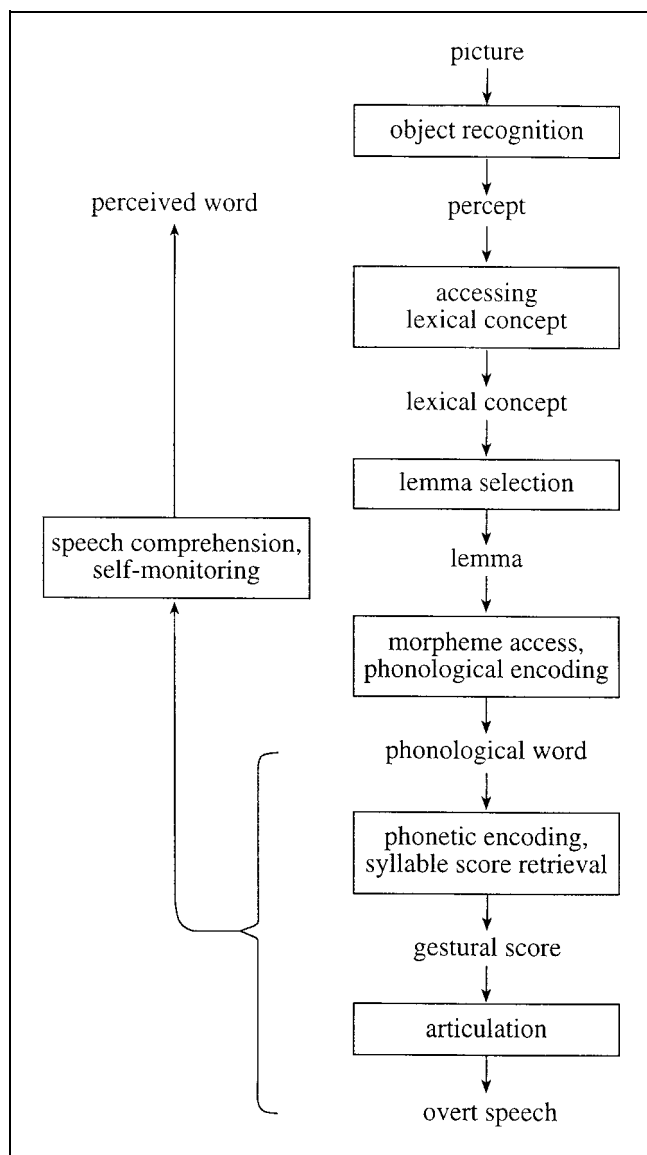


Figure 1. A stage model of picture naming. Preparing a name proceeds through stages of visual processing, activating a lexical concept, word selection, phonological encoding, phonetic encoding, and the initiation of articulation. Self-monitoring refers to phonological codes and overt speech.

Selecting the Lemma

When a lexical concept, such as VIOLIN, is active, activation spreads to the corresponding word in the mental lexicon, or more specifically, to the corresponding *lemma*. This is the word's *syntactic* representation, involving the word's syntactic category (noun for *violin* or *canoe*), its gender if it is a noun—in gender-marking languages (in German: feminine for *Geige*, neuter for *Kanu*), and its subcategorization if it is a verb (transitive, intransitive, etc.). These are the properties of a word that are essential for grammatical encoding, for example, embedding it in the appropriate syntactic environment (*Geige* will need a feminine article, become head of a

noun phrase, etc.). All active lexical concepts spread activation to their lemmas in the lexicon, but eventually only one lemma will get selected, following a simple mathematical rule: At any one moment the probability of selecting the target lemma is the ratio of its activation and the total activation of all lemmas involved. This “hazard rate” determines the expected selection duration (Roelofs, 1992).

Accessing the Morpheme(s) and Generating the Phonological Word

The selected lemma (and *only* the selected lemma) spreads its activation to its word form representation, or *morpheme(s)*. The morpheme(s) represents the word's phonological shape, in particular its segments (for <canoe> these are the phonemes /k/, /ə/, /n/, /v:/) and its metrical structure (for <canoe> it is the iambic foot $\sigma\sigma'$). These two sources of information are independently retrieved (Levelt, 1992a; Roelofs & Meyer, 1997) and used to incrementally create a so-called “phonological word” (ω), a fully syllabified representation of the word, such as

$$\begin{array}{c} \omega \\ / \quad \backslash \\ \sigma \quad \sigma' \\ \wedge \quad \wedge \\ k \quad \theta \quad n \quad v: \end{array}$$

In the process, the phonemic segments are one by one (from left to right) associated with the metrical frame of the phonological word, successively creating the word's phonological syllables (/kə/ and /nv:/ in the example).

Retrieving Syllabic Gestural Scores

As soon as a syllable, such as /nv:/, has been programmed, its gestural score is retrieved from the mental syllabary (Levelt & Wheeldon, 1994). The syllabary is the repository of abstract motor representations for the frequently used syllables in the language. These syllabic scores are retrieved one by one as the phonological syllables become incrementally available. Accessing the syllabary is again due to activation spreading. In the example, syllable-initial /n/ activates all syllables with /n/ as onset or offset and all syllables with the vowel /v:/ as nucleus. The target syllable's hazard rate determines the expected selection duration (Roelofs, 1997).

Articulation

The word's articulation is initiated as soon as all of its syllabic scores have been retrieved (Levelt & Wheeldon, 1994). The articulatory system computes least-effort solutions for the execution of the (still rather abstract) gestural scores. These solutions provide the motor instructions for the respiratory, the laryngeal, and the su-

pralaryngeal systems involved in fluent articulation of speech (Levelt, 1989).

Self-Monitoring

The overt speech is self-perceived, which activates the speaker's speech comprehension system. In addition, the speaker can monitor a prearticulatory internal speech code, probably the phonological word representation (Wheeldon & Levelt, 1995). If any communicatively serious error is detected, the speaker can halt and make a self-repair (Levelt, 1983). This "late" feedback loop is probably the only one involved in the system, in spite of claims to the contrary in the connectionist literature (Levelt, 1992b; MacKay, 1992).

A comprehensive review of this stage theory, with all the experimental evidence on which it is based, can be found in Levelt, Roelofs, and Meyer (in press).

The Timing of Stages

Apart from this late self-perceiving feedback, picture naming is completely or almost completely a feedforward process. This makes picture naming particularly attractive in the study of the cortical localization of language-processing components. In analyzing functional imaging data with high temporal resolution, such as evoked MEG responses, one can assume that each component is active just once. We are, of course, aware of diverging claims in the literature (among them Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Dell & O'Seaghdha, 1991; Humphreys, Riddock, & Quinlan, 1988; 1995; Martin, Gagnon, Schwartz, Dell, & Saffran, 1996; Peterson & Savoy, 1998), but there is an almost universal agreement that picture naming is a staged process, where feedback, if any, is so local that there will not be measurable activation-deactivation-reactivation patterns for any processing component involved.

In the MEG experiments to be reported, the average picture naming latency (i.e., from picture onset to the onset of articulation), was 538 msec. How is this time distributed over the above processing stages? The experimental data in the literature do not allow us to distinguish between the first two stages (i.e., deriving the percept and accessing the lexical concept). The most recent estimate is from Thorpe, Fize, and Marlot (1996). In an ERP go/no-go study in which subjects had to categorize a novel picture as displaying an animal or not, the authors showed that a frontal negativity developed around 150 msec post-picture onset only on the no-go trials. At that moment, apparently, the subject recognized that the picture contained no animal. This picture recognition time estimate agrees reasonably well with estimates based on the data in Potter (1983), the concept beginning to be accessed some 100 msec after picture onset. One should keep in mind, though, that much depends on the quality of the picture. The pictures in

our MEG experiment were high-quality line drawings (see Figure 7). Thorpe et al. used color photographs, probably also high-quality. An additional reason for assuming that Thorpe et al.'s estimate holds well for our materials is the following. Their average animateness decision time was 445 msec. One of our pretests (see "Methods" section) was an object/non-object decision on the experimental picture. Average object decision time was 479 msec. Most of our pictures had earlier been used in the Jescheniak and Levelt (1994) experiments. One of these used a picture recognition task. The subject saw a word on the screen (for instance *dog*), then saw the picture, and had to press a "yes" button if the two agreed, a task very close to Thorpe et al.'s. The average "yes" response took 439 msec. Because *dog* (etc.) and Thorpe's et al.'s "animate" are lexical concepts, it is likely that the 150-msec estimate includes accessing the lexical concept (i.e., it probably covers the first two stages of the naming process).

Turning now to lexical access proper, a first estimate of the duration of lemma selection in naming tasks is presented in Levelt et al. (1991): 115 msec. That number is derived from data obtained in a picture naming/lexical decision experiment. The primary task in the experiment was picture naming. But on one-third of the trials the subject heard an auditory probe, a word or a nonword. It could be presented at one of three moments after picture onset but before the initiation of naming (i.e., at three different stimulus onset asynchronies, or SOAs). On these trials the subject had to perform a secondary task before naming the picture, pushing a "yes" or "no" button depending on whether the auditory probe was a word or not. On all relevant items the probe was a word, and it could be semantically or phonologically related to the target picture or be unrelated to it (for instance, if the picture was one of a sheep, the probe word could be *goat*, *sheep*, or *house*). At some SOAs the lexical decision latency to a semantically related probe (*goat* in the example) was slower than to an unrelated probe (such as *house* in the example). The obvious interpretation for that finding is that the picture naming had advanced so far at that SOA that the target lemma was active and could begin interfering with the probe lemma. The computational model that provided an optimal fit for these lexical decision data yielded the mentioned 115-msec estimate for lemma access.

This estimate is in good agreement with the one provided by the computational model of lemma selection in Roelofs (1992). Lemma access was the very target of that paper. Roelofs used a picture/word interference task in his experiments. Subjects named pictures, and naming latency was the dependent measure in these experiments. At some moment during the task (i.e., at different SOAs) a visual distracter word was flashed in the picture and the subject was instructed to ignore it. The distracter could be semantically related to the target picture name or unrelated to it (just as in the

above experiment). Typically, naming latencies were slower when the distracter was semantically related (such as *goat* when a sheep was displayed) than when it was an unrelated distracter (such as *house*), at least when picture and distracter presentation were closely synchronized. The computational model based on these data sets gives lemma selection durations in the range of 100 to 150 msec. This, finally, corresponds well with what Potter (personal communication) derives from the data in Potter and Faulconer (1975), an estimate of 150 msec.

Duration estimates for the next stage of lexical encoding, word form access and phonological encoding, can be derived from the following three sources. The first one is Roelofs's (1997) WEAVER model of phonological encoding. It is based on myriad picture/word interference experiments in which auditory phonological distracters were presented during picture naming. Based on these data, Roelofs gives an estimate of 265 msec for the interval between lemma selection and accessing the syllable score of a monosyllabic word. Because this involves selecting the syllable node (although not "unpacking") it, the time for attaching phonemic segments to the metrical frame should be less. An estimate for the latter comes from a second study, by Wheeldon and Levelt (1994). In their experiments the subject was provided with a target phoneme, for instance an /l/ or an /r/. The (Dutch) subject would then hear an English word (for example *bitchhiker*) and decided whether the Dutch translation of that word (*lifter* in the example case) contained the target phoneme. The Dutch word was not overtly pronounced. The dependent measure was the push button (yes) latency. For bisyllabic items such as *lifter*, the average difference in phoneme monitoring speed between the word initial and the word final phoneme (/l/ and /r/ for this example) turned out to be about 125 msec. This can be taken as an estimate of the duration to (internally) encode the whole word phonologically. Finally, van Turenout, Hagoort, and Brown (1997), in an ERP study not unlike Thorpe et al.'s, where word-initial and word-final phonemes provided go/no-go cues for a semantic picture categorization, could estimate a duration of about 120 msec for phonological encoding.

Left over, then, are phonetic encoding and the initiation of articulation. If the above estimates are approximately correct, subtraction can provide an estimate for these final stages. For an average naming latency of 538 msec, the distribution over the stages would roughly be as follows: visual processing plus accessing the lexical concept: 150 msec; lemma selection: 125 msec; phonological encoding: 125 msec; and phonetic encoding and initiation of articulation: 138 msec. This latter number agrees well with a final estimate that we have. Levelt et al. (1991) provide a model estimate for the duration of phonological plus phonetic encoding, based on their picture naming/lexical decision experiments; it is 270

msec. If we subtract 125 msec for phonological encoding, we are left with 145 msec for the duration of phonetic encoding.

In the present study, we will group activation data according to the following time windows post-picture onset:

0–150 msec	Visual processing and accessing the lexical concept
150–275 msec	Lemma selection
275–400 msec	Phonological encoding
400–600 msec	Phonetic and articulatory processing

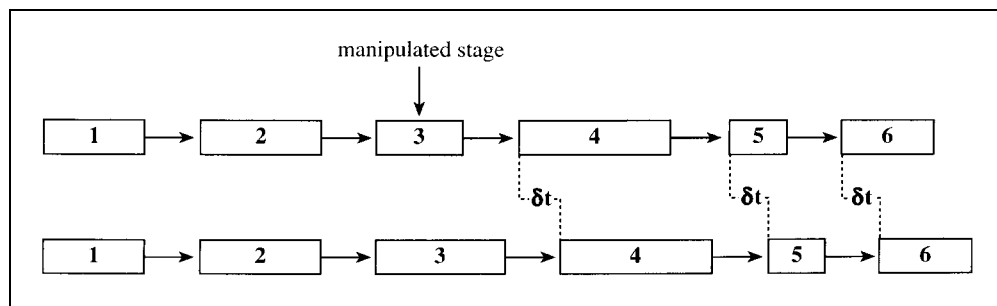
The last window extends somewhat beyond the average naming response to include most of the factual naming responses (but, of course, for each response only pre-onset activation data were analyzed). Although these windows capture all of the data in the above meta-analysis, these data do not allow us to estimate window variances. However, our selection of experimental materials (see "Methods") aimed at minimizing variances in object recognition and in total naming latency.

Singling Out Phonological Encoding

As mentioned previously, one can, in principle, study the cerebral localization of a single processing stage by introducing an independent variable that specifically affects the duration of that stage. Any dependent effect in the MEG trace should be due to the corresponding processing component. In the present study, we decided to try this out on the stage of phonological encoding. Jescheniak and Levelt (1994) discovered that the well-known word frequency effect in picture naming (Oldfield & Wingfield, 1965) is due to accessing the word's form. Pictures whose names are highly frequent (HF) in the speaker's language (such as *boat*) are named faster (by some 30 to 50 msec) than pictures whose names have low frequency (LF) (such as *broom*). The authors could show that this is not due to forming the percept or to retrieving the lemma or the word's syllables. The effect arises in accessing the word's form information.

This fact can now, in principle, be used to localize the process of word form access in the brain if one has access to a temporal high-resolution method, such as MEG. The idea is this (see Figure 2): When the independent variable in the picture naming task is target word frequency (i.e., HF versus LF picture names), we should expect that the initial processing stages up to the stage of form access are the same for HF and LF pictures. But from morpheme access on, the time course for HF pictures will divert from the time course for LF pictures. More precisely (see Figure 2), the critical stage will be relatively short for the HF pictures and relatively long for the LF pictures (the time difference δt will be about 30 to 50 msec). All subsequent stages will not differ in duration but will be desynchronized by δt . One can now

Figure 2. The single-factors method. Affecting the duration of a single stage does not affect the cortical dynamics before that stage but shifts all activity following that stage by a constant δt .



trace the activation pattern over the MEG sensors and determine whether such desynchronization arises and in particular where it begins. Inversely, one can response-lock the activation patterns and determine where desynchronization begins running backward in time. In the ideal case, one can “squeeze in” the critical process in time and place. But, as already mentioned, 30 to 50 msec is not much and is easily drowned in experimental noise.

Whatever it is worth, this *single factors* or *stage focusing* method (or SF-method) differs essentially from the subtraction method, which is now commonly used in positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies. There one tries to *eliminate* a stage completely; here all stages are always present. There is reason to have serious doubts about the possibility of eliminating a stage entirely (Friston et al., 1996). In his classical paper, Sternberg (1969) argued for a different approach, the additive factors method, which leaves all stages intact. The SF-method is a derivative of Sternberg’s method (see also Dehaene, 1996).

In the present study, eight Dutch subjects repeatedly named 24 carefully preselected pictures, 12 with LF names and 12 with HF names, while their magnetic brain activity was recorded in the Neuromag-122 whole-head magnetometer. This was done in two identical and immediately adjacent sessions of six randomly dispersed repetitions of all 24 pictures. In a prerun of exactly the same experiment, but without MEG recording and on different subjects, a 37-msec word frequency effect had been obtained (see “Results”).

RESULTS

In the following we will first, in summary fashion, describe the unsuccessful single-factors analysis and then turn, in detail, to the more revealing source/time analysis.

Single Factors Analysis

Reaction Times

As mentioned, we obtained a solid 37-msec word frequency effect in the pre-experiment. Figure 3 shows the reaction time curves for high- and low-frequency word responses. The word frequency effect was stable over

the 12 repetitions (see “Methods”). This substantial, stable word frequency effect was not replicated in the main experiment (run on different subjects). Across all subjects, mean reaction times for high- and low-frequency items differed by only 3 msec (535 versus 538 msec). The effects per subject ranged from 20 msec favoring the low-frequency items to 15 msec favoring the high-frequency items. In analyses of variance within subjects, none of these effects reached statistical significance. Separate analyses were carried out for each subject and each of the two test sessions (i.e., before versus after the short break). A significant ($p < 0.05$) frequency effect favoring high-frequency items was only obtained for one subject in one test session.

Should this null effect be added to the new set of subjects we tested or to the experimental situation in the MEG lab? To find out, we retested the same subjects 6 months after the main experiment on the same materials, but in the Max Planck Institute (i.e., without MEG recording). The instructions stressed that it was important to name the targets correctly—speed was of secondary importance. Now a significant word frequency effect of 33 msec was obtained for the 24 relevant test

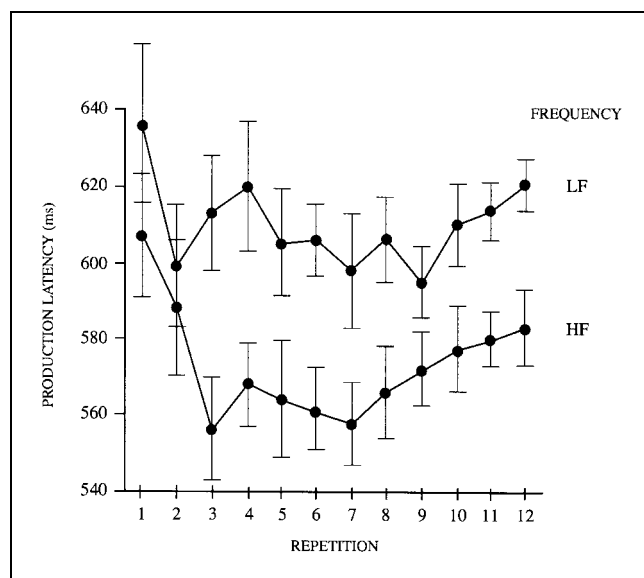


Figure 3. The word frequency effect in the naming latency for the ultimate 12 HF and 12 LF stimuli, as measured in the preexperiment over 12 repetitions of the test pictures.

items ($F(1, 7) = 13.76$, $MSE = 3771$; $F(1, 22) = 9.68$, $MSE = 7965$, both $p < 0.01$). Six subjects showed frequency effects favoring the high-frequency items in both halves of the experiment, and two subjects showed them only in one-half of it. In other words, during the retest, the subjects of the main experiments behaved very much like the subjects of the pretest. Hence, the null effect in the main experiment was not due to the set of subjects. Was it due to the testing situation? The most noticeable difference was that naming latencies were very short in the MEG experiment, on average 538 msec. These latencies were manually determined from the visually displayed acoustic wave form. In the preexperiment (different subjects) average naming latency had been 592 msec, measured by voice key. Because the voice key is triggered at some acoustic threshold, voice key latencies are always slower than latencies measured on the wave form display; this amounts to about 20 msec on average. Even with this correction, the latencies in the MEG experiment were faster by 42 msec. Is it the case that our young Dutch subjects, flown into Helsinki for a weekend, had been so highly motivated during the MEG session that their naming speeds reached ceiling levels, obliterating any difference between HF and LF words? This is unlikely. First, in their post-test their average voice key latency was 575 msec. Implementing the 20-msec correction, this differs by no more than 17 msec from the latencies in the MEG experiment. Second, we tested a new set of 10 subjects, who did the experiment twice (in two sessions 1 week apart), once under strict speed conditions and once under the instruction to avoid errors and not to worry about time. Although the voice key averages differed substantially between these two conditions (542 msec versus 618 msec), the frequency effect was exactly the same: 27 msec. Notice that the speeded condition was faster by 16 msec than in the MEG experiment (correcting for the 20 msec). So, a word frequency effect can be obtained even with very fast naming responses. In short, we have no explanation for the tragic disappearance of the effect in the MEG experiment. It did, however, undermine our single-factors analysis.

Evoked Magnetic Fields

Although the null effect for word frequency eliminated the single-factors analysis as displayed in Figure 2, there might still be a word frequency effect in the evoked magnetic data. For all subjects we visually checked the HF and LF responses recorded by the 122 sensors. The spatio-temporal patterns of evoked magnetic fields differed considerably between subjects, except for the relatively uniform exogenous visual evoked fields recorded at occipital sensors. But none of the subjects showed any systematic difference between HF and LF activation patterns. Nevertheless, we performed analyses of variance over HF versus LF curves for all sensors and all subjects,

by partitioning all curves in 50-msec windows, ranging from 0 to 600 msec post-onset. Although we found more significant word frequency effects than could be expected by chance, they were not ubiquitous and showed no interpretable patterns within or over subjects. This was further confirmed in the dipole source analysis (see Figure 4).

Dipole Source Analysis

The main aim of our dipole source analysis was to relate peak activity of distinct cortical areas to the time windows of the stage model of picture naming. Therefore, dipole source analysis was performed separately for each subject's data, using the signals from subjects' first testing session and averaged over the two frequency conditions (see "Methods" section). Figure 4 shows the original whole-head MEG responses for HF and LF words in one subject. The data were modeled by eight equivalent current dipoles (ECD), identified at distinct latencies and shown on the surface rendition of the subject's MRI. The source strengths as a function of time are plotted on the right.

The source localization results over all subjects are displayed in Figure 5. The four time windows discussed above correspond to four colors in the diagrams. Let us go over these windows in temporal order.

Visual Processing and Accessing the Lexical Concept: 0 to 150 msec

The visual evoked magnetic fields were adequately modeled by about two to three sources in each subject. These sources were located in the occipital cortex (yellow sources), with a preponderance of sources in the right hemisphere. Note that the model sphere for the source analyses was fitted in such a way that central, parietal, and temporal areas were covered optimally. Therefore, localization errors are likely to be largest in the frontal and occipital areas. Because the modeling of visual-perceptual activity was not the goal of this study, no further attempt was made to determine the location of these sources more precisely. The findings do agree, however, with the occipital lobe activation adduced to object recognition in Price et al.'s (1996) PET study of picture naming.

Lemma Selection: 150 to 275 msec

In this time window magnetic field patterns were more variable. The sources identified in this time frame (three to four per subject—blue dots in the second panel) were largely found in the occipital area (50%) and in the parietal and occasionally temporal areas. However, there was a remarkable clustering of sources (seven of eight subjects) in the right parietal cortex, along the posterior end of the superior temporal sulcus.

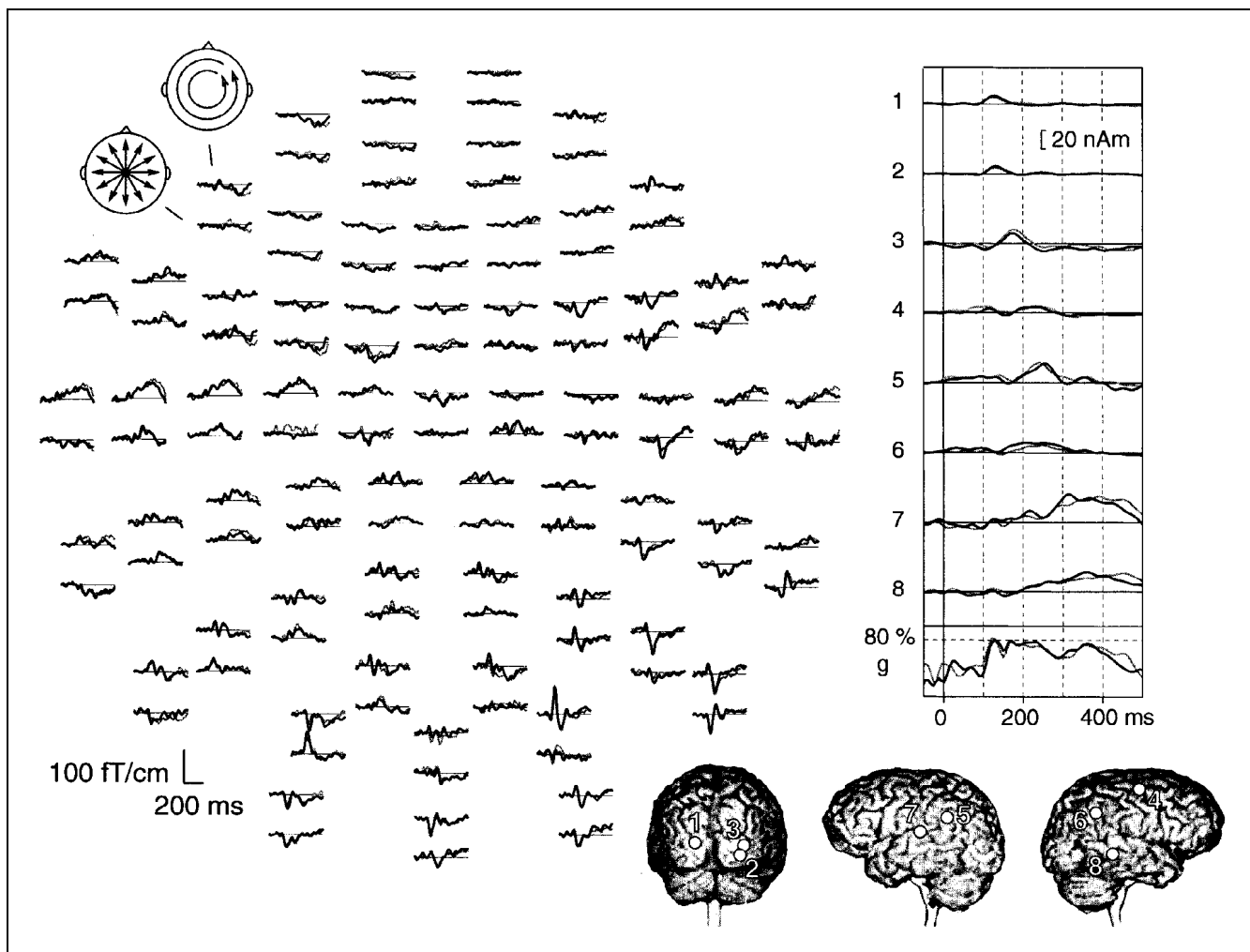


Figure 4. Picture-evoked cortical activity over all 122 sensors for one subject. At each measurement site, the variation of magnetic fields measured in two orthogonal directions, indicated on the schematic heads (top left). The curves show the original waveforms in each sensor for HF (black curves) and LF (gray) words. The source activation patterns in a multipole model of eight dipoles are presented in the right panel. The waveform marked “g” indicates the goodness of fit of the multipole model. The lower right panel shows the eight dipole sources in the model on a surface rendering of the subject’s MRI.

Phonological Encoding: 275 to 400 msec

The sources fitted in this time window are the red dots in the third panel of Figure 5. A clear source cluster (six of eight subjects) was located close to the posterior third of the superior temporal gyrus and the temporo-parietal junction, agreeing with the site of the classical Wernicke’s area. In contrast, the right-hemisphere sources in the same time frame were at remote areas, with no single source in the right-hemisphere homologue of Wernicke’s area.

Phonetic and Articulatory Processing: 400 to 600 msec

Another 14 sources were identified in the last time frame from 400 to 600 msec (green dots). These were quite scattered, with the largest concentration in the sensory-

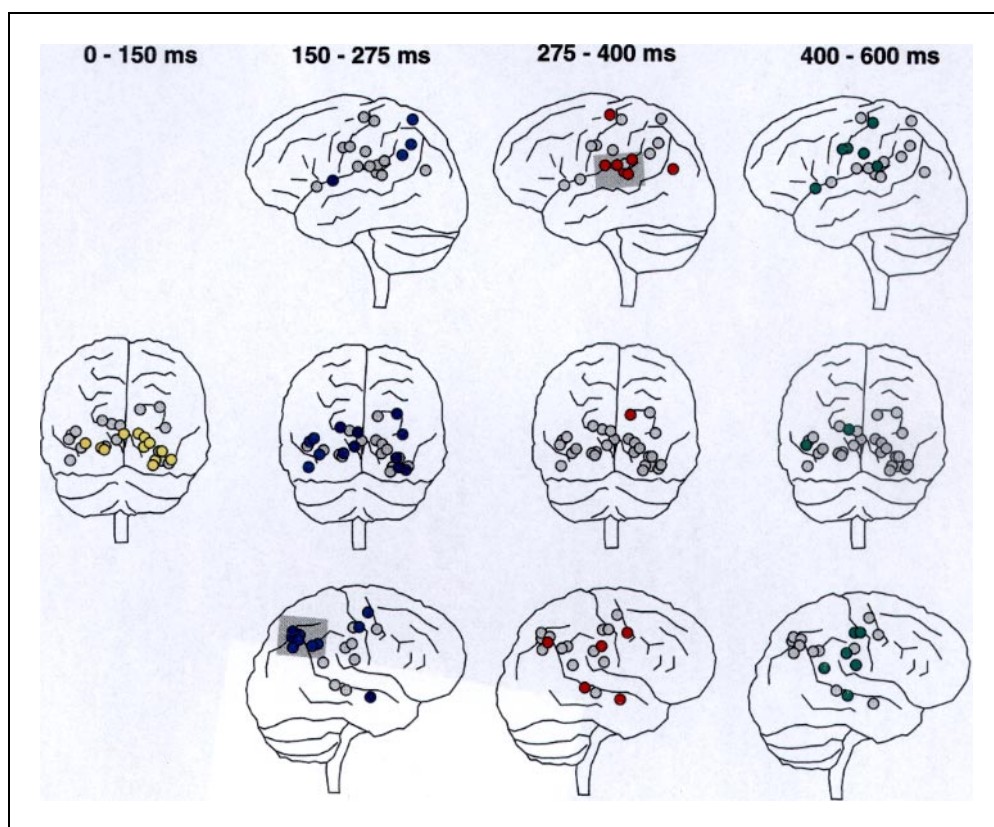
motor cortex (in particular the vicinity of the face area) and in the parietal and temporal lobes. Some of these sources may represent activity related to self-monitoring rather than phonetic encoding.

Figure 6 displays the average time course of activation for the two prominent source clusters in the right parietal cortex cluster in the 150- to 225-msec time frame (blue curve) and around Wernicke’s area in the 275- to 400-msec time frame (red curve). We will return to the activation of these areas, as well as to the potential role of self-monitoring, in the “Discussion.”

DISCUSSION

The aim of this paper was to relate a detailed psychological model of spoken word generation to the dynamics of cortical activation during picture naming. This, we hoped, would not only tell us what network of cortical

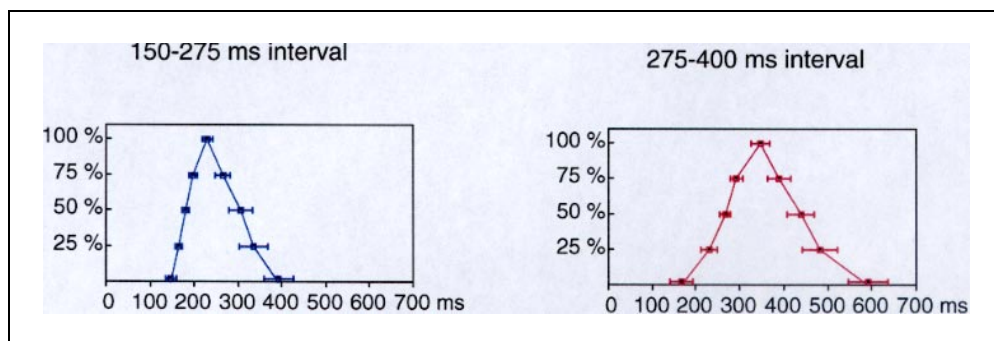
Figure 5. Dipole sources of evoked responses, collected over eight subjects. Yellow dots: Visual processing and accessing the lexical concept; 0–150 msec. Blue dots: Lemma selection; 150–275 msec. Red dots: Phonological encoding; 275–400 msec. Green dots: Phonetic and articulatory processing; 400–600 msec. The gray dots represent other sources (i.e., those that do not have their peak activation in the relevant time frame).



areas is involved in this production task but also and more interestingly, which loci are involved in successive functional stages in word generation. During each of these stages a different, highly specific computation is performed, involving different knowledge sources (visual, conceptual, lexical, phonological, phonetic, articulatory). Each of these successive computations is indispensable; if any of them fails, naming will deteriorate or block altogether. But because the operations are so wildly different, it is a reasonable working hypothesis that they all involve different, dedicated neural populations.

The two methods we applied to localize these dedicated regions were not equally successful. The first one, the single-factors method, failed to work because of an unhappy complication in the experiment: The independent variable, word frequency, was not effective for reasons that evade us in spite of careful pre- and post-tests. Still, we believe the logic of the method, portrayed in Figure 2, to be correct. To relate any stage-type processing model to high temporal-resolution imaging data, the single-factors method is to be preferred over the usual subtraction methodology. Notice that, generally, it is not required to have reliable estimates of the precise

Figure 6. The average time course of activation for the cluster of right parietal sources in the second time window (blue curve) and for the cluster of Wernicke area dipoles in the third time window (red curve). Plots were obtained by reading the latencies at which the activations were 0, 25, 50, 75%, and 100% of the maximum in the dipole wave forms of the individual subjects and then averaging the latencies over those subjects who showed activity in the relevant area.



time course (over stages) of the process. To apply the SF-method, one only needs some independent experimental variable that can sufficiently affect the duration of a single stage. That can always be pretested in a reaction time experiment.

The second approach was to relate a dipole source analysis to the precise temporal staging of the process, as measured in reaction time experiments. This type of analysis is more demanding of the psychological model. It requires rather solid estimates of the durations of the successive stages. This is not normally given for psychological stage models. But for the task at hand, picture naming, so much is available in terms of temporal measurements that reasonable estimates could in fact be made of the time windows for successive stages. That allowed us to detect stage-bound dipole sources. The results were encouraging. As a general pattern, our findings conform to Salmelin et al.'s (1994)—a steady progression of activation from the occipital to the parietal-temporal and frontal areas of the brain. But there were substantial differences in detail. First, Salmelin et al.'s data showed stronger involvement of lateral motor/premotor areas in both hemispheres than the present data. Second, only the present data show concentrated right-parietal activity (second time frame). And, third, only the present data show a definite clustering of dipoles in the left posterior sylvian region (third time frame). Still, both experiments involved rather straightforward picture naming tasks. What could have caused these and other differences in the patterns of evoked brain activation? The present experiment was not meant as a replication of the original study; it served its own purpose. In particular, the single-factors word frequency analysis put specific requirements on the task and the stimulus materials. It was, for instance, specific to the present experiment that (1) all picture names were monomorphemic, (2) all picture names were monosyllabic, (3) pretests had selected for pictures with small variance in recognition times, (4) pretests had removed all pictures that showed naming uncertainty, (5) pictures were repeatedly presented during a session (six times), (6) a limited number of different pictures was used (42 test plus filler stimuli as opposed to 200 pictures in the original study), and (7) speed instructions were given to the subjects. It is hard to adduce differences in results specifically to any one of these procedural differences between the experiments; they have probably been working in concert. The lesson is rather that even subtle differences in tasks and experimental materials may cause substantial variation in evoked brain activation patterns. Here we analyze the present results in their own right.

Let us first return to the right parietal lobe clustering of dipoles, which are active during the second time period (150 to 275 msec). We had no a priori reason to expect this area to be involved in lemma selection, but it still may be in a task where lemma selection is based

on the recognition of visual objects. There are two sets of relevant findings in the literature. The first concerns the role of the right parietal cortex in working memory tasks. In a PET study, Klingberg, Kawashima, and Roland (1996) found evidence for right inferior parietal activation (on the border of angular and supramarginal gyri) involved in auditory, visual, and tactile short-term memory tasks, with detection tasks as controls. Salmon et al. (1996), also in a PET study, found activation in the right inferior parietal region during a verbal memory task in which subjects apparently used a visual imagery strategy (see Ungerleider, 1995, for a review of cortical mechanisms for memory). In our experiment the subject saw the same picture repeatedly, in fact six times during the session on which our dipole analyses are based. Although, on average, 41 other pictures intervened between two presentations of the same picture, some subjects may have tried to retrieve the response they produced earlier for the same, now recognized picture. The problem, though, with such a strategy is that it will increase rather than decrease naming latency (cf. Levelt & Wheeldon, 1994). If such an involvement of short-term visual-verbal retention would nevertheless account for right parietal activation in our task, it should entirely disappear in a task where pictures are presented just once. A more likely interpretation of the right parietal cluster in our data derives from a second set of studies on the management of visual attention. Faillenot, Toni, Decety, Grégoire, and Jeannerod (1997), also in a PET study, observed activation in the right hemisphere at the border of Brodmann's areas (BA) 19 and 7 during a visual shape-matching task. They attributed this activation to the shifting of attention to the location of an expected object part or property. This is in good agreement with a PET study by Fink et al. (1996) in which the right parietal region (in particular BA 19 and 39) is involved in visually attending to global features of a stimulus. Husain, Shapiro, Martin, and Kennard (1997) showed that damage to this area affects the patient's ability to disengage attention from one visual stimulus to the next. Lemma access in our task may gain by active management of visual attention. Global features of the picture will often be informative of its semantic category; attending to them may speed up the zooming in on the appropriate lemma. In addition, a strategy of rapidly disengaging visual attention from the picture just named in order to be ready for the next one may have contributed to the short naming latencies that we observed in the present experiment.

The sources showing the remarkable clustering in the left temporal cortex were active during the stage of phonological encoding. It strongly suggests Wernicke's area to be a favorite site for phonological encoding in word production. This finding certainly agrees with the Ojemann et al. (1989) and Haglund et al. (1994) statistics, which show a high concentration of language sites in just this area. Our results can now interpret these par-

ticular sites as ones probably dedicated to phonological encoding. From other recent work we know that phonological encoding, not lexical selection, is the troublemaker in both the tip-of-the-tongue phenomenon (Vigliocco, Antonioni, & Garrett, 1997) and in at least one type of anomia (Badecker, Miozzo, & Zanuttini, 1995). Our results point to Wernicke's area as the cortical site of these kinds of trouble. These results also complement existing evidence for the involvement of the left superior temporal gyrus in auditory phonological processing (Boatman, Lesser, & Gordon, 1995; Démonet et al., 1992). It should be noticed though, that in their PET study of picture and color naming Price et al. (1996) did not observe any activation in Wernicke's area for phonological retrieval. But that negative result may have been due to the subtraction used. In their experiment the control task for picture or color naming was just viewing the picture and saying "yes." Preparing the word "yes" needs phonological encoding just as well as preparing a picture or color name. Finally, the present data agree with the already mentioned ERP results of Abdullaev and Posner (1997), a Wernicke area activation in the verb generation task. We can now interpret that finding as due to phonological encoding of the response word.

Still, one should be careful not to draw the reverse conclusion (i.e., to argue from our data that the sole function of Wernicke's area in word production is to phonologically encode the word). The red curve in Figure 6 shows that, on average, dipoles in the critical cluster show an activation pattern that transcends the time window for phonological encoding, which could signal the involvement of Wernicke's area in temporally adjacent processes. To be precise, the rise and fall to 25% levels of activation in the cluster encompass a time span of 250 msec, quite a bit more than the estimated 125-msec duration of phonological encoding. Although this may be due to normal variability in the data over trials and subjects, or to some temporal overlap of phonological encoding in Wernicke's area with pre- and post-phonological encoding in nearby cortical areas, one shouldn't exclude the possibility that Wernicke's area is also involved with other aspects of the production process. A particularly relevant aspect is self-monitoring (see Figure 1), which is indeed temporally contiguous with phonological encoding. It is known that self-monitoring takes self-produced overt speech as input as well as "internal speech," more precisely the phonological code (Wheeldon & Levelt, 1995). Self-monitoring supposedly involves the speaker's normal speech comprehension system (Levelt, 1989). Recent PET studies by McGuire, Silbersweig, and Frith (1996) have indeed shown that self- and other-monitoring of overt speech involve the same temporal lobe areas. Monitoring of self-produced overt speech cannot be discerned in the data analyzed here because we didn't include the stage of overt articulation. But internal speech monitoring can begin during the phase of phonological encoding (i.e., as early as 275

msec after picture onset). This may involve Wernicke's area, which is known to be essential to speech comprehension (Démonet et al., 1992). In addition, McGuire et al. found right temporal lobe activation in self-monitoring. That may also have been the case for our subjects, involving some or all of the red- and green-marked right temporal lobe sources in Figure 5.

More generally, these considerations show that active sources cannot be univocally assigned to one component process or another when these processes overlap in time. However, using as paradigms well-defined tasks such as picture naming, for which explicit, extensively tested psycholinguistic processing theories are available, it becomes possible to suggest correlations between distinct stages of cognitive processing and cortical dynamics.

METHODS

Subjects

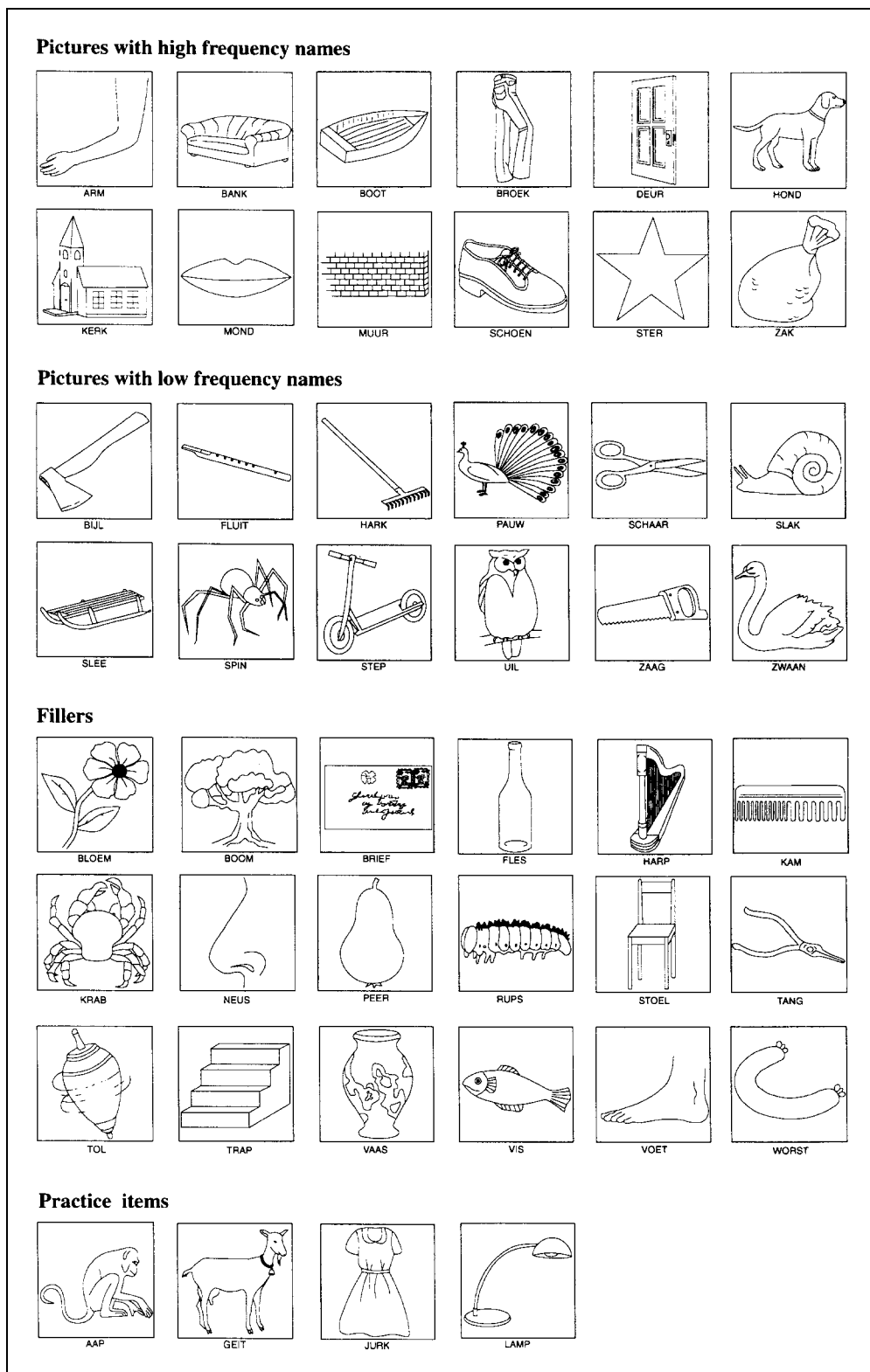
The subjects were eight healthy, right-handed students (four male, four female) between 21 and 30 years of age (mean 24). All subjects had participated in earlier psycholinguistic experiments using reaction time methods and EEG recordings. They were all native speakers of Dutch and had normal or corrected-to-normal eyesight.

Materials

The experimental materials included 4 practice and 42 experimental pictures (see Figure 7). Half the experimental pictures had high-frequency and the others low-frequency names (mean word form frequencies according to CELEX: 99.96 versus 4.08 per million). All picture names were monosyllabic. The items constituted a selection from the set used in Jescheniak and Levelt's (1994) study of word frequency effects. The pictures were presented as white line-drawings on a black background. All pictures were drawn to fit into a quadratic frame with sides corresponding to a visual angle of 6° on the subject's screen.

With these materials two pretests were carried out. The goal of the first pretest was to establish that the frequency effect was stable across a fairly large number of item repetitions. Ten subjects named each of the experimental pictures 12 times. At the beginning of the experiment, the subjects studied a booklet that included the drawings and the names they should use to refer to them. For each subject, a different test list of 12 blocks was created. In each block, all experimental pictures occurred once. The order of the pictures within a block was random and different for each block and subject. The first test block was preceded by presentation of the practice items. There was a short break after the sixth block. The seventh block was again preceded by the four practice items. On each trial, a fixation cross

Figure 7. The 24 experimental pictures with their high- or low-frequency Dutch names, the 18 fillers, and the 4 practice items that were used in the main experiment and the first pretest.



was presented for 200 msec. After a blank interval of 600 msec, a drawing was presented until a voice key was triggered by the onset of the subject's naming response. Maximum picture presentation time was 1300 msec. The next trial began 1300 msec after speech onset or after

2600 msec if no response was made within the response interval.

A highly significant frequency effect of 31 msec was obtained (means: 607 versus 576 msec; $F(1, 9) = 40.84$, $MSE = 1384$; $F(2(1, 40) = 12.39$, $MSE = 9697$, both $p <$

0.01). The interaction of blocks and frequency was not significant. Thus, the frequency effect was stable across 12 picture repetitions (see Figure 3).

The goal of the second pretest was to rule out that the frequency effect was due to differences in visual and/or conceptual processing of the pictures. Ten subjects performed an object decision task on the experimental pictures and a set of 42 drawings of nonobjects, similar to those used in Kroll and Potter's (1984) study of object naming. On each trial the subject indicated by pressing one of two response buttons whether or not the drawing represented an existing object.

The experiment included 12 test blocks in each of which all objects and nonobjects appeared once. Decision times were faster by 7 msec for high-frequency than for low-frequency pictures. This difference was significant by subjects but not by items (means: 475 versus 482 msec; $F(1, 9) = 7.13$, $MSE = 429$, $p < 0.05$; $F(1, 40) = 1.01$, $MSE = 6903$). The interaction of frequency and blocks was not significant. Thus, the pictures with high-frequency names were slightly easier to recognize than those with low-frequency names. But it is also clear that the 31-msec word frequency effect obtained in the pretest was largely independent of visual/conceptual processing.

The main experiment included the 42 experimental items of the pretest. However, just to be sure not to involve any frequency-dependent visual/conceptual preparation, we identified a subset of 12 high- and 12 low-frequency items that were matched for decision times in the second pretest, and all analyses were confined to these pictures (see Figure 7). (To identify the matched subset, pairs of high- and low-frequency pictures were formed that did not differ by more than 5 msec in average decision time. Those pictures for which no decision-time matched partner could be found were excluded.) The mean word form frequencies for the high- and low-frequency items in the subset were 115.54 and 4.12 per million. Reanalyses of the results of the first pretest showed that the frequency effect for these items was 37 msec (means: 573 versus 610 msec, $F(1, 9) = 34.03$, $MSE = 2380$; $F(1, 22) = 13.18$, $MSE = 7306$; both $p < 0.01$). The interaction of frequency and blocks was not significant (see Figure 7). In analyses per subject, seven out of ten subjects showed a significant ($p < 0.05$) frequency effect.

Task Procedure

Before the start of the measurements subjects studied the pictures and their names from a sheet of paper, as the subjects in the first pretest had done. They were instructed to name the pictures as fast as possible. Subsequently they were brought into the shielded MEG room and seated before a white screen (distance 80 cm) on which the stimuli were projected by a Sony VPL-350 QM LCD data projector. The trial structure was exactly

the same as in the first pretest. Each subject received one training block to get acquainted with the procedure and to ensure a clear pronunciation without too much movement of jaw and mouth. The experimental session included 12 blocks and was divided in two sessions of 6 blocks each, separated by a short break of about 10 min. Each block had a duration of about 2 min and 20 sec.

Recording Procedure

The subjects were tested individually in a magnetically shielded room. Magnetic signals were recorded with a 122-channel whole-head magnetometer (Neuromag-122; Ahonen et al., 1993), using a bandpass of 0.03 to 90 Hz. The signals were recorded continuously, digitized at a rate of 297 Hz, and stored on magneto-optical disks for off-line analyses. Along with the magnetic signals the electrooculogram (EOG) was recorded, using bipolar horizontal and vertical EOG derivations.

For the alignment of MEG and magnetic resonance image (MRI) data three coils were attached to the subject's head. Coil positions relative to nasion and preauricular points were measured with a 3-D digitizer (Isotrak 3S1002, Polhemus Navigation Sciences, Colchester, VT). The coil positions with respect to the magnetometer were determined from the magnetic signals produced by the coils. MRIs were obtained with a 1.5 T Siemens Magnetom.

MEG Data Analysis

Evoked magnetic fields following picture presentation and preceding the naming response were averaged stimulus-locked relative to picture onset. Trials with EOG activity exceeding 150 μV were excluded from the average (~10%). The averaged epochs encompassed -100 to 900 msec. The data consisted of averages over 110 to 144 trials, yielding waveforms with good signal-to-noise ratio (which were also highly reproducible between the two measurement sessions). The procedure involved the identification of the dipolar field patterns by scanning through the data (i.e., the magnetic field patterns estimated from the measured values at each sensor, cf. Hämäläinen, 1995). Stable field patterns were modeled at the time instant at which they were clearest and showed least interference with other source areas by fitting an equivalent current dipole (ECD) to the local field pattern measured by a subset of sensors. The 3-D location of an ECD represents the center of gravity of the active area, and its amplitude and orientation estimate the magnitude and average direction of the postsynaptic currents within the area. The ECDs thus obtained were integrated in a multidipole source model, derived by fitting the dipoles to the entire spatiotemporal field pattern, keeping the locations and orientations of the sources fixed (Hämäläinen, Hari, Ilmoniemi, Knuutila, &

Lounasmaa, 1993). The source analyses employed a spherical head model of homogeneous conductivity. The resulting source models typically explained 80 to 90% of the variance in the data, with the best fit usually in the time range between 200 and 500 msec. Occasionally, fitting the dipoles to the entire field pattern resulted in more than a single peak activation for a dipole. We took this seriously because there is no a priori theoretical reason why a particular cortical area would not be reactivated at a later stage in the process; the serial model only excludes reactivation through feedback. We used the following criterion for multiple activation of a dipole. A second moment of activation is only reported if the relevant peak in the waveform exceeds 75% of the first maximum, and the two peaks are separated by at least 100 msec, and the activity returns to baseline level during this interval. Using this criterion, four source waveforms with two distinct maxima were identified in our data.

Acknowledgments

This study was supported by the Academy of Finland and by the EU's Human Capital and Mobility program through the Large-Scale Facility BIRCH at the Low Temperature Laboratory, Helsinki University of Technology. We are grateful to the late Eddy Roskam for advice on the analyses of variance; to Peter Wittenburg for preparing the data for the analyses of variance; to Ger Desserjer and Godelieve Hofstee for assistance in preparing the stimulus materials, for running the pre- and post-experiments, and for assistance in the data analysis; to Kimmo Uutela for preparing the stimulus sequence in Helsinki; to Riitta Hari for comments on the manuscript; and to Chris Frith for comments on the right parietal lobe cluster.

Reprint requests should be sent to Willem J. M. Levelt, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands or via e-mail: pim@mpi.nl.

REFERENCES

Abdullaev, Y. G., & Posner, M. I. (1997). Time course of activating brain areas in generating verbal associations. *Psychological Science*, 8, 56-59.

Ahonen, A. L., Hämäläinen, M. S., Kajola, M. J., Knuutila, J. E. T., Laine, P. P., Lounasmaa, O. V., Parkkonen, L. T., Simola, J. T., & Tesche, C. D. (1993). 122-Channel SQUID instrument for investigating the magnetic signals from the human brain. *Physica Scripta*, T49, 198-205.

Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, 57, 193-216.

Boatman, D., Lesser, R. P., & Gordon, B. (1995). Auditory speech processing in the left temporal lobe: An electrical interference study. *Brain & Language*, 51, 269-290.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 60-64.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature*, 380, 499-505.

Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience*, 8, 47-68.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. *Psychological Review*, 98, 604-614.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in normal and aphasic speech. *Psychological Review*, 104, 801-838.

Démonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulos, J. L., & Wise, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115, 1753-1768.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384, 159-161.

Faillenot, I., Toni, I., Decety, J., Grégoire, M.-C., & Jeannerod, M. (1997). Visual pathways for object-oriented action and object recognition: Functional anatomy with PET. *Cerebral Cortex*, 7, 77-85.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees? *Nature*, 382, 626-628.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *Neuroimage*, 4, 97-104.

Glaser, W. R. (1992). Picture naming. *Cognition*, 42, 61-105.

Haglund, M. M., Berger, M. S., Shamseldin, M., Lettich, E., & Ojemann, G. A. (1994). Cortical localization of temporal lobe language sites in patients with gliomas. *Neurosurgery*, 34, 567-576.

Hämäläinen, M. (1995). Functional localization based on measurements with a whole-head magnetometer system. *Brain Topography*, 7, 283-289.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65, 413-497.

Hart, J., & Gordon, B. (1992). Neural subsystems for object knowledge. *Nature*, 359, 60-64.

Humphreys, G. W., Lam, C., & Lloyd-Jones, T. J. (1995). An interactive activation approach to object processing: Effects of structural similarity, name frequency, and task in normality and pathology. *Memory*, 3, 535-586.

Humphreys, G. W., Riddock, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, 5, 67-103.

Husain, M., Shapiro, K., Martin, J., & Kennard, C. (1997). Abnormal temporal dynamics of visual attention in spatial neglect patients. *Nature*, 385, 154-156.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 824-843.

Klingberg, T., Kawashima, R., & Roland, P. E. (1996). Activation of multi-modal cortical areas underlies short-term memory. *European Journal of Neuroscience*, 8, 1965-1971.

Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, 23, 39-66.

- Levitt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.
- Levitt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levitt, W. J. M. (1992a). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*, 1–22.
- Levitt, W. J. M. (1992b). The perceptual loop theory not disconfirmed: A reply to MacKay. *Consciousness & Cognition*, *1*, 226–230.
- Levitt, W. J. M. (1996). Perspective taking and ellipsis in spatial descriptions. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 77–107). Cambridge, MA: MIT Press.
- Levitt, W. J. M., Roelofs, A., & Meyer, A. S. (in press). A theory of lexical access in speech production. *Behavioral and Brain Sciences*.
- Levitt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*, 122–142.
- Levitt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239–269.
- MacKay, D. A. (1992). Awareness and error detection: New theories and research paradigms. *Consciousness & Cognition*, *1*, 199–225.
- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *381*, 649–652.
- Martin, N., Gagnon, D. A., Schwartz, M. F., Dell, G. S., & Saffran, E. M. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes*, *11*, 257–282.
- McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, *119*, 907–917.
- Ojemann, G. A. (1983). Brain organization for language from the perspective of electrical stimulation mapping. *Behavioral and Brain Sciences*, *6*, 189–206.
- Ojemann, G. A., Ojemann, J., Lettich, E., & Berger, M. (1989). Cortical language localization in left, dominant hemisphere. *Journal of Neurosurgery*, *71*, 316–326.
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273–281.
- Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.
- Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 539–557.
- Pollen, D. A., & Trachtenberg, M. C. (1972). Some problems of occipital alpha block in man. *Brain Research*, *41*, 303–314.
- Potter, M. C. (1983). Representational buffers: The see-mind hypothesis in picture perception, reading, and visual search. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 423–437). New York: Academic Press.
- Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words: Understanding without words. *Nature*, *253*, 437–438.
- Price, C. J., Moore, C. J., Humphreys, G. W., Frackowiak, R. S. J., & Friston, K. J. (1996). The neural regions sustaining object recognition and naming. *Proceedings of the Royal Society London, B* *263*, 1501–1507.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.
- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–284.
- Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1–18.
- Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.
- Salmelin, R., Hari, R., Lounasmaa, O. V., & Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature*, *368*, 463–465.
- Salmon, E., Van der Linden, M., Collette, F., Delfiore, G., Maquet, P., Degueldre, C., Luxen, A., & Franck, G. (1996). Regional brain activity during working memory tasks. *Brain*, *119*, 1617–1625.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance, Vol. 2* (pp. 276–315). Amsterdam: North-Holland.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- Ungerleider, L. G. (1995). Functional brain imaging studies of cortical mechanisms for memory. *Science*, *270*, 769–775.
- van Turennout, 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*, *23*, 787–806.
- Vigliocco, G., Antonioni, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, *8*, 314–317.
- Wheeldon, L., & Levitt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*, 311–334.