# The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax

G. Dogil*, H. Ackermann, W. Grodd, H. Haider, H. Kamp,
J. Mayer, A. Riecker, D. Wildgruber

*Institut fur Maschinelle Sprachverarbeitung, Experimentelle Phonetik,
Azenbergstrasse 12, 70174 Stuttgart, Germany*

## Abstract

In this study we will give an overview of the experimental work on the neuroanatomical correlates of language and speech production that we have done in recent years. First we will introduce the methodology of event-related functional magnetic neuro-imaging and the experimental paradigm that we employed. Then we will present and discuss the results of our experiments on (1) speech motor control, (2) articulatory complexity, (3) the neuroanatomical correlates of prosody, and (4) the neurocognitive substrates of syntactic processing. Experiments (1) and (2) show that the expected large motor speech network consisting of SMA, motor cortex and cerebellum is only active in planning and execution of simple articulatory movements. Increased articulatory complexity leads to more focused activation. Furthermore, we can show that only the execution of speech movements recruits the left anterior insula, while articulatory planning does not. The results of experiment (3) indicate that it is not the function of prosody (linguistic vs affective) that controls lateralization of prosodic processing, but that more general characteristics of the processing units like the size of the prosodic frame are responsible for the activation of different cortical regions. Finally, in experiment (4) we present first results on syntactic processing in speech production. Besides the expected activation of Broca's area we found activations in Wernicke's area and in the cerebellum. We have also found evidence for activations in other cortical areas, which are less often implicated in clinical studies on brain language correlations. The cognitive relevance of these areas and networks is still to be elucidated. © 2001 Elsevier Science Ltd. All rights reserved.

## 1. Introduction

Human language is unique as it has no ancestors in the animal world. In this sense it is

* Corresponding author. Tel.: +49-711-121-1379; fax: +49-711-121-1366.
  *E-mail address:* dogil@ims.uni-stuttgart.de (G. Dogil).

much more unique than the human brain, the development of which can be traced across the biolological evolution of species. The study of the correlation between brain and language has always lagged behind the study of other aspects of behaviour and cognition due to the lack of an animal model. Thus, the scientific community predominantly had to rely on the evidence provided by the correlation of neurological disorders with language disorders, apart from selective brain recording and stimulation techniques performed during neurosurgery. Recent advances in scanning techniques and computer technology have enabled scientists to have a closer look at the undisturbed working brain. These techniques either register/distort the electro-magnetic activity of the active neurons (EEG, MEG, TMS) or explore the relationship between blood flow and neural activity (PET, fMRI).[1] The research described in this study involves what we consider the simplest and most accurate technique of investigating the brain/language correlation—event-related functional magnetic resonance imaging [ER-fMRI]. We will begin with a short introduction to this particular method and our implementation of it.

## 2. The technique

fMRI takes advantage of the relationship between blood flow and neuronal activity.[2] Neuronal activity causes an increase in regional Cerebral Blood Flow (rCBF) to compensate for the enhanced metabolic activity. Circulation actually overcompensates for the increased metabolic activity, leading to an excess of oxyhemoglobin in active brain tissue. The brain uses up to 20% of the oxygen we breathe, however, the amount of blood supplied to a specific brain region varies only slightly between the time when it is active and when it is at rest. The self- regulating system of blood flow in the brain forces more oxygen to those areas where neuronal activity is greatest. This so-called blood oxygenation level dependent (BOLD) effect forms the physiological basis for the success of fMRI. Consider Fig. 1 for a schematic illustration of this hemodynamic effect.

The magnetic properties of the oxygenated and the deoxygenated blood are different. Deoxyhemoglobin is paramagnetic while oxyhemoglobin is diamagnetic. Paramagnetic substances (like deoxyhemoglobin) become magnetized in the presence of a magnetic field provided by an MRI scanner giving rise to distortions in the magnetic field and subsequent signal dephasing. In a region of activation there is more oxygenated blood than before the activation started (cf. Fig. 1). This results in a net decrease in paramagnetic material. Therefore, we get a signal increase in activated areas due to less dephasing.

The non-invasive character of magnetic resonance allow for the design of elegant paradigms and milestone discoveries in cognitive neuroscience [24,47]. The study of language and speech *production* within the MRI (or PET) scanner, however, has been considered to be technically difficult, if not impossible. The net increase of the signal which can be measured with fMRI is very small (3–5%) and dangerously close to the statistical signal increase ratio. Hence, the measurements which allow a statistically solid interpretation have to be sustained for a very long period of time (20–60 s). Such long-phase activation

---

[1] There are several excellent introductions to these electrophysiological and hemodynamic techniques. We refer our readers to Ref. [24], chapter 4) and to Ref. [52].

[2] This relationship has been known to neurophysiologists for over a century.

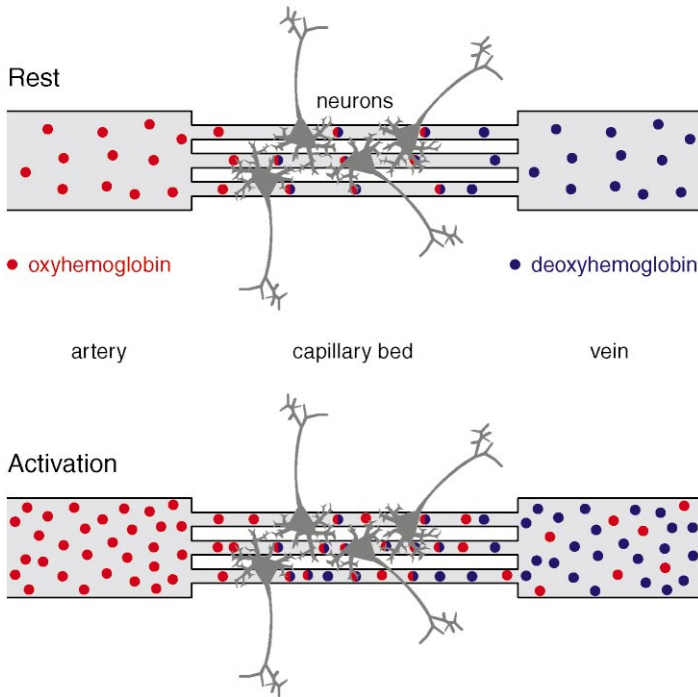## Blood Oxygenation Level Dependent (BOLD) effect



Fig. 1. Vasculature in the brain and the BOLD effect.

paradigms are known under the heading of the 'block-design', and have dominated cognitive experimentation for a long time. One can express doubts as to whether such designs are well suited for the study language and speech in general. However, the technical problems in speech production are of a different nature. Speaking involves movement, e.g. of the head, the lips, the tongue, the larynx, and these movements lead to changes of the volumes of the cavities close to the brain. Unfortunately, the brain-imagining techniques are quite susceptible to motion and volume-change artefacts. The block-design, due to its poor temporal characteristics, does not permit the separation of signal changes caused by artefacts from the signals due to the BOLD-effect. The only sensible method for studying language and speech production had to use so-called 'inner' or 'imaginative' speech or whispering [28]. The developments in the fMRI technique—stronger magnetic field strength and better interpretation of the temporal properties of the BOLD-effect—have led to the establishment of an experimental paradigm which allows for studies of overt speech. The paradigm offers, at present, optimal temporal and spatial resolution.[3]

---

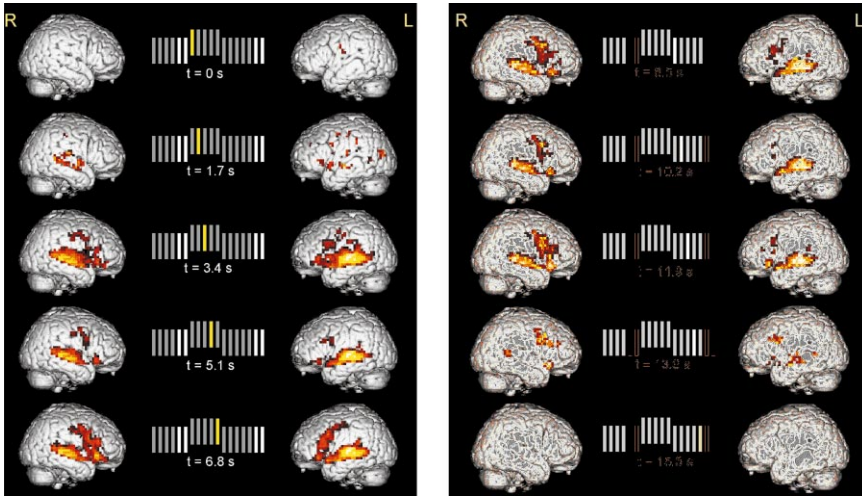[3] For technical limits of the state-of-the-art fMRI, cf. Ref. [39].

Fig. 2. Mental chronometry in a task of comparing prosodic expressiveness of sentences. The initiation (acoustic stimulus) is marked by a white bar. The hemodynamic reaction in various brain regions can be observed in steps up to 12.6 s following the task.

## 2.1. Event-related fMRI

Hemodynamic responses to neural activity are sluggish. The BOLD-effect can be registered in the cortical areas long after their activation has terminated. For example, the BOLD-effect can be registered up to 12 s after neural activity has ceased. This long-lag hemodynamic effect of brain activity allows for so-called event-related fMRI designs [7,30]. The strength of these designs obviously lies in their temporal resolution. Given the generic strength of fMRI in spatial resolution, event-related fMRI may become the method of choice in the functional neuroimaging of language. In our group we use fast image processing methods that allow us to screen the mental chronology of the activated brain. An example of such a cognitive cinematography with fMRI is provided in Fig. 2. It is taken from a study on the perception of affective prosody [59]. The BOLD-effect can be observed several seconds after the subjects have made a decision about the prosodic expressiveness of a sentence that they have listened to.

The fact that the regional cerebral blood flow change observed in fMRI is delayed several seconds relative to the stimulus onset allowed us to use this method to study speech production. Notice that we can practically disregard activity registered at the initiation of the task (when movement artefacts might occur due to speech production) and concentrate on those epochs which are not influenced by movement and consequently are not vulnerable to artefacts. The cortical response to a speaking task which is undisturbed by movement artefacts forms a basis of our scanning procedure. A schematic illustration of this experimental design is provided in Fig. 3.

The event-related method paired with a randomized experimental design guarantees that experimental manipulations are exerting stimulus-related effects. Furthermore, the data can be easily subjected to a post hoc control for possible artefacts. Much of this
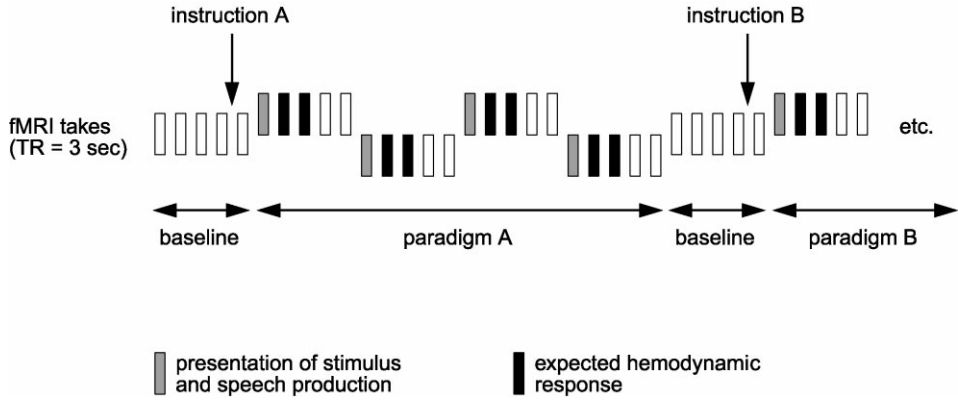
Fig. 3. Event-related design used in our experiments in speech production.

control is actually being performed by standardized data-interpretation paradigms. The software package used in our experiments is the generally available Statistical Parametric Mapping (SPM) package, versions SPM96 and SPM99 (Wellcome Department of Cognitive Neurology, London). It allows for normalization of co-registered brain images (spatial normalization and movement correction), smoothing (Gaussian filter), and exploratory data analysis. Ultimately, SPMs represent significant activity increases in specific areas of the brain. For a schematic presentation of some of the crucial SPM steps see Fig. 4.

## 2.2. Experimental procedure

The technique presented above is not a simple one, but if the technical prerequisites are
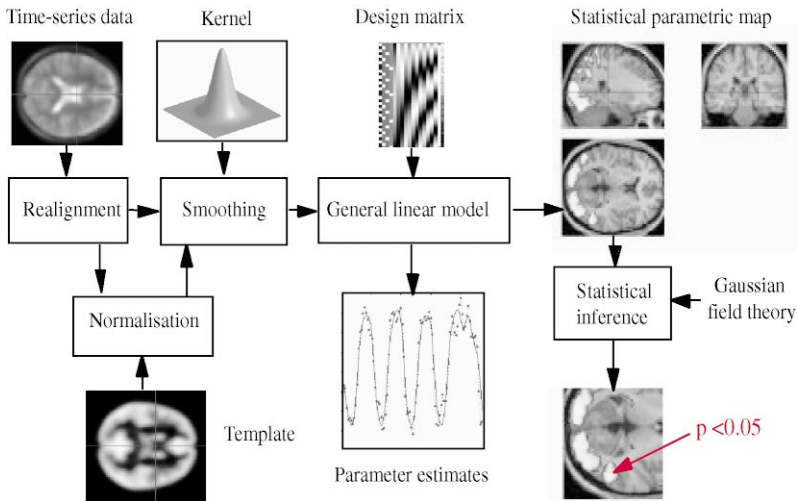


Fig. 4. Essential data transformation processes leading to the Statistical Parametric Maps of the brain activity registered by fMRI (from [22], p. 11).

given it is easy to use and its spatial resolution is high. The non-invasiveness of the technique allows it to be repeated over an over again, also from an ethical point of view. How do we go about this? What tasks do we choose? Which areas of the brain do we concentrate on?

The methodology of *paired image subtraction* has become the standard in fMRI studies of cognition. The first assumption that this methodology is based on is that the brain is equipotential, with each activity requiring the interaction of the entire structure. Hence, the whole brain—and not the specific regions of interest—is mapped during the experimentation. According to the subtraction methodology images of the entire brain are obtained in two states—the task state and the control state. These two images are then subtracted from one another to create a difference image. The subtraction images identify those areas of the brain where the activity significantly differs between the task state and the control state. The subtraction images, not the a priori assumptions, define the location of functionally significant regions and the magnitude of increased activity—"the human brain tells us how it is organized".[4] For the brain to tell us as much as possible, the experimenter has to design the tasks in such a way that they differ only minimally. In addition to this, it is assumed that when a task state is compared with a control state, the difference represents the addition of processing components unique to the task state without affecting processing components in the control state.[5] A subtraction image will then reveal the areas of increased brain activity unique to the task state. The paired image subtraction paradigm has been used in all the experiments which will be discussed in the following sections.

The Paired-Image Subtraction Design:
Task A has **n** cognitive components and **m** neural components.Task B has **n + x** cognitive components and **m + y** neural components.The subtraction B − A reveals the correlation between the cognitive component **x** and the neural component **y**.

## 3. fMRI experiments in speech and language production

Experimental task selection is guided by a model. Intuition is a bad advisor in brain imaging research. The subtraction method requires the strict decomposition of those

---

[4] Marcus E. Raichle in an interview with [24]: 114–115).

[5] This assumption, which is well known from clinically motivated neuroscience and notorious under the name of 'pure insertion assumption', has been often criticized as unrealistic, or even outright wrong. Some critics of the image subtraction methodology argue that the cognitive decomposition of tasks hardly ever leads to the identification of 'unique' differences, and that therefore the methodology is impractical. Even if this were the case (which we will show it is usually not), the slight extension of the methodology allows for highly predictable task isolation. Price and Friston [48], for example, suggest a method [somewhat misleadingly called 'cognitive conjunction'] in which a series of image subtractions among non-uniquely different tasks depicts these cognitive components which are under experimental scrutiny. Even if pure insertion is violated—i.e. when it is the case that brain areas active in the control state are not active in the task state—the substraction methodology provides interesting insights. In such a case the subtraction images reveal not only the increase of brain activity in the task state, but also the decreased activity characteristic of those areas that are used in the control state but not in the task state. The violation of the 'pure insertion assumption' is thus not a catastrophy, but a further challenge to the image subtraction methodology (cf. Raichle in Ref. [24]: 115).

processing components that are involved in a given task. Concerning language production we follow the processing model developed by Levelt and others during the last decade [32,33]. Not only is it the model with the strongest experimental support, but it is also the one that understands speech and language production as a modular process, including several self-contained tasks which can be easily tested using the image-subtraction methodology. Levelt's 'blueprint' for the speaker with its most essential components is summarized in Fig. 5. The challenge for us is to find the correlates for these processing components in the working brain. Obviously, we cannot test all of the components and have to concentrate on some particular questions. However, we will attempt to present a 'big picture' in this review. We will describe experiments concerning the correlates of ARTICULATION (Section 3.1); another experiment dealing with the generation of PROSODY (Section 3.2) and yet another one concerning the generation of SYNTAX (Section 3.3). In the general discussion we will suggest the localizations of the respective brain areas which the three arrows depicted in Fig. 5 point to. We will also refer to some dynamic processes within the uncovered network.

### 3.1. Planning and execution of speech

The 'blueprint' model in Fig. 5 distinguishes three building blocks in the phonological/ phonetic system: morpho-phonological encoding, phonetic encoding and articulation. Our particular interest has been devoted to the neural correlates of the latter two components. It is known that damage to the anterior peri-/intra-sylvian cortex of the dominant hemisphere gives rise to a syndrome of specific articulatory deficits known as Apraxia of Speech (AOS).

> Apraxia of Speech is a nosological entity in its own right which co-occurs with aphasia only occasionally. This...conviction rests on detailed descriptions of patients who have a severe and lasting disorder of speech production in the absence of any significant impairment of speech comprehension, reading or writing as well as of any significant paralysis or weakness of the speech musculature ([31]: 380)

Most of the explanatory approaches to AOS refer to higher cognitive functions in the motor system in order to describe its underlying pathomechanism [51,60]. Either motor planning or motor execution are argued to be distorted. Dogil and Mayer [16] claim that in AOS it is not the procedural components (phonological encoding, phonetic encoding, articulation—cf. Fig. 5) but the representational component of the speech generation system (phonological and/or articulatory score—cf. Fig. 5) that are impeded. In particular, they have argued that articulatory scores of apraxics are 'overspecified'—in the sense that they contain too much articulatory detail. Dogil and Mayer provide clinical data which cannot be accommodated by the standard accounts that are based on gestural complexity and motor planning.

The available clinical data are less ambiguous with respect to the relevant lesion site. Dronkers [19] conducted a study in which brain lesions of 25 patients with AOS were compared to those of 19 patients without apraxic deficits. A 100% lesion overlap identified the precentral gyrus of the left insula as a site involved in Apraxia of Speech. The lesions of the 19 patients without Apraxia of Speech all spared the insula. Dronkers ([19]: 159) concluded that the left anterior insula "seems to be specialized for the motor planning of
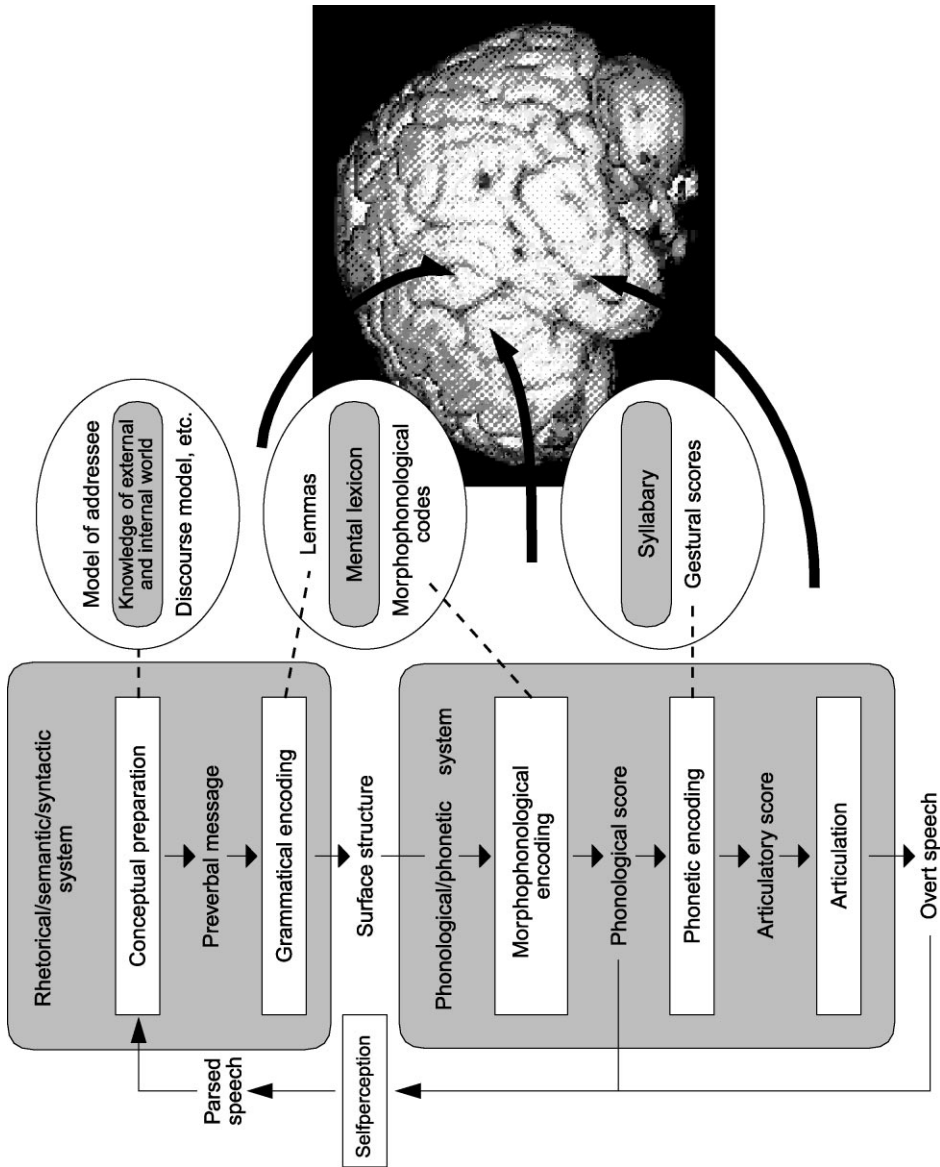
Fig. 5. The processing components of Levelt's model of speech production and its challenges for functional localization in cognitive neuroscience.
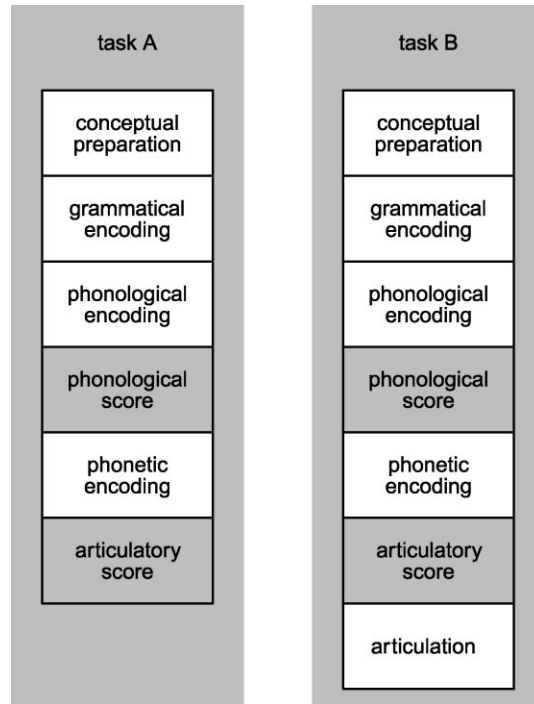
Fig. 6. Cognitive decomposition of task pairs.

speech". Does the healthy brain tell us the same story about planning, execution and control of speech production? Does it support the specific localization for "the motor planning of speech".

### 3.1.1. Experiment 1: the speech network[6]

We have already mentioned the severe restrictions which early fMRI studies of speech were confronted with. The movement artefacts forced researchers to use ways of speech production that were not quite natural (whisper or inner speech). The event-related fMRI methodology allows us to eliminate the movement artefacts, while still being able to use inner speech as a control task for the speaking task. The subtraction of these two tasks should illustrate the difference between the planning and the execution stage of speech. It shall reveal the contribution of ARTICULATION to the production of speech. Thus, we attempt to find a cognitive composition of our paired-subtraction task such as the one presented in Fig. 6.

In order to keep the conceptual and grammatical components of the tasks simple and

---

[6] This experiment is part of a larger study, investigating covert and overt singing as well, published in Ref. [49]. An overt and covert reproduction of a melody ("Eine kleine Nachtmusik"—W. A. Mozart) with a syllable 'la' was required. We will not discuss the results of the singing task. The reader will notice that the results of the singing task are an almost complete mirror image of the results of the speaking task. For details see [49].

constant we asked the subjects to continuously repeat the names of the months of the year. The highly automatized word strings they produced were expected to require minimal procedural efforts in phonological and prosodic encoding. The critical task was performed aloud and the control task was performed in a silent mode. Eighteen healthy right-handed speakers of standard German (nine females; age 22–63 years, median = 39 years) participated in the present study, after informed consent had been obtained. Prior to the experiment, subjects performed each of the tasks three times outside the scanner to get acquainted with the test materials. During fMRI, participants were lying supine in a 1.5 T magnetic resonance tomograph (Siemens Vision; Erlangen, Germany), the head being secured by a foam rubber to minimize movement artifacts. The tasks were applied 12 times each in a counter-balanced order, the onset-to-onset intervals between activation phases amounting to 24 s. Participants were asked to refrain from verbal thought during the rest periods in between. Distinct visual symbols appearing on a screen indicated which task was to be carried out during the next activation period. Subjects were asked to perform the tasks as fast as possible, while avoiding slurred production of the test materials in order to preserve the distinct character of the sounds and tones. Under these conditions, similar demands concerning motor control mechanisms were placed on the participants. The produced utterances and tunes were recorded. Twenty-eight parallel axial slices (thickness 4 mm, gap 1 mm) were obtained across the complete brain (EPI, $64 \times 64$ matrix, field of view 192 mm, TE 39 ms, TR 3 s, $\alpha = 90°$) during each measurement period ( = 3 s). A T1-weighted 3D turbo-flash sequence (MP range, 128 sagittal slices, thickness 1.5 mm, $256 \times 256$ matrix, field of view 256 mm, TE 4 ms, TR 9.7 ms) served as an anatomical reference for functional images (Lotze et al., 1999). Signal analysis including realignment of functional images, coregistration with structural images, spatial normalization, smoothing with a 10 mm Gaussian filter, and statistical analysis ($z$-value cut-off, $P < 0.001$, corrected extent threshold $P < 0.05$) were performed with the SPM96 package (Wellcome Department of Cognitive Neurology, London).

The covert speech task elicited increased hemodynamic activity at the level of the supplementary motor area [SMA] and the left motor cortex [MC], as well as contra-lateral activity in the cerebellum [CER]. In the overt speech task this network of activation was enriched by the activation of the right motor cortex and the left cerebellum. In addition to the motor speech network (MC-SMA-CER), the activation of the left anterior insula was elicited only during the overt speech task. Consider Fig. 7 and Table 1. The activation of the intrasylvian cortex (left anterior insula) appears to be a clear contribution of a task in which speech is actually activated and not only planned.

Clear cut lateralization effects, which emerged during the covert task performance at the level of the motor cortex and the cerebellum are very much in line with our previous studies [1,2,57]. Subtraction of the activation patterns obtained during the 'planning' and 'execution' tasks yielded rather identical hemodynamic differences at the level of the rostal insular cortex. The result indicates that *activation of the insular cortex is tightly connected to overt vocalization*. Given what we know about the anterior insular cortex and its involvement with Apraxia of Speech this result is rather surprising. The anterior insula is the part of the speech network which is evidently not involved in speech planning. Still its lesion leads to a syndrome which, at least due to standard interpretations, illustrates a
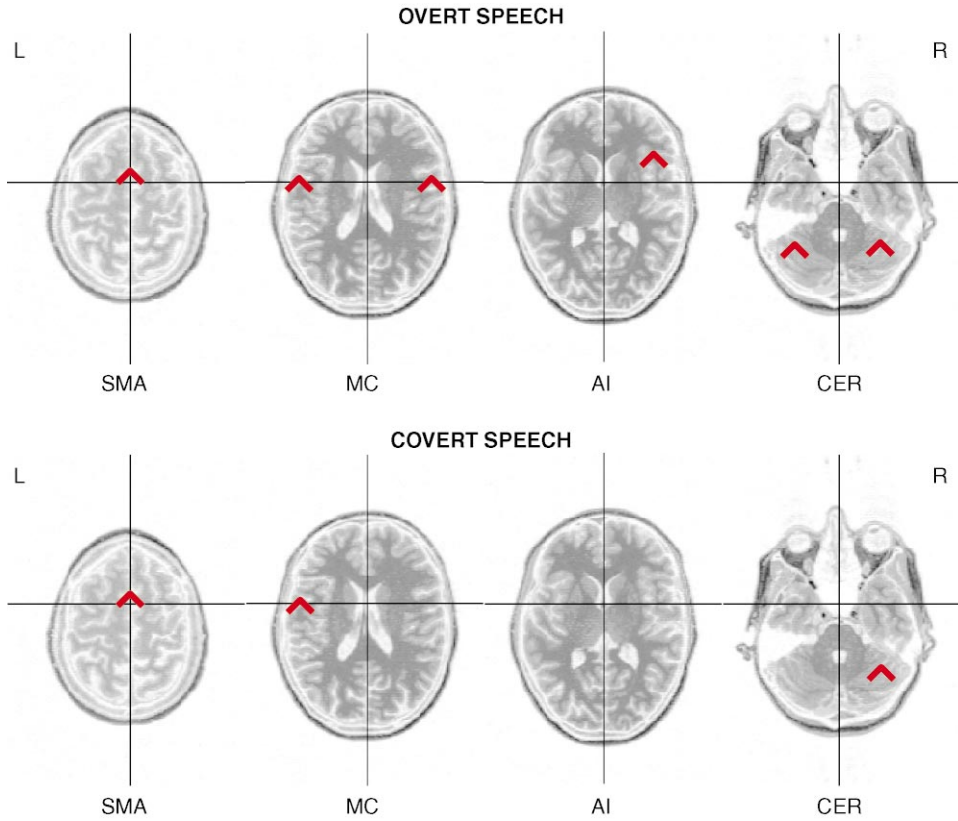
Fig. 7. Transverse slices at the level of the suplementary motor area (SMA), motor cortex (MC), anterior insula (AI), and Cerebellum (CER). Red arrows indicate the localization of activations.

Table 1
Hemodynamic responses (fMRI) at the level of supplementary motor area (SMA), motor cortex/inferior frontal gyrus (MC/IFG), anterior insula and cerebellum during overt and covert speech. Cluster-size = number of activated voxels above the selected threshold; voxel-level = individual voxel with maximum $z$-value within a cluster ($z \lessapprox 3.09$, $P < 0.05$ corrected); coordinates refer to Talairach-space

| Condition | Region | Side | Tal. X | Tal. Y | Tal. Z | Vozel-level ($z$-value) | Cluster-size (no. of voxels) |
|---|---|---|---|---|---|---|---|
| Overt speech | SMA | Medial | −3 | 8 | 47 | 6.63 | 177 |
| | MC/IFG | R | 65 | −3 | 17 | 7.43 | 198 |
| | | L | −68 | −8 | 16 | 7.87 | 385 |
| | Anterior Insula | L | −35 | 18 | −4 | 5.93 | 43 |
| | Cerebellum | R | 27 | −45 | −20 | 7.26 | 246 |
| | | L | −30 | −39 | −18 | 5.68 | 164 |
| Covert speech | SMA | Medial | −3 | 16 | 30 | 5.29 | 165 |
| | MC/IFG | L | −65 | 6 | 18 | 7.25 | 296 |
| | Cerebellum | R | 30 | −60 | −27 | 5.66 | 157 |

distortion of speech planning. There is an alternative explanation, however. One could argue (with Dogil and Mayer [16]) that the underlying impairment in Apraxia of Speech affects the checking mechanism which transfers the phonetic code into the articulation module (articulatory score in Levelt's model, cf. Fig. 5). Dogil and Mayer [16] argue that an important part of this articulatory score is the 'phonetic underspecification' of articulatory gestures. They show that the process of the implementation of the phonetic gestural score can only work, if gestures are 'underspecified' for features which are not relevant for establishing phonemic contrasts in a language. We stipulate that the insula plays a central role in this checking mechanism at the last stage of the transition between the higher cognitive (linguistic) and purely realizational motor processes.[7]

A highly promising concept in neural theories of cognition is the idea of large-scale cortical networks, where elementary functions are localized in discrete cortical and subcortical regions, and complex functions involve parallel processing of information in wide-spread networks [5,15]. In the experiment presented above the brain revealed its own wide-spread network for generating speech. The activity of this network consisting of SMA-MC/IFG-CER-AI (cf. Table 1) is a dynamic process, which allows for a smooth and robust implementation of speech production. One can consider the checking function of the specification of the articulatory score (temporal specification) as a central, but fairly late dynamic control process. The malfunction of this control process due to a specific breakdown within the anterior insula might be considered a neurological factor underlying the loss of underspecification or the problems associated with the temporal specification of the articulatory gestures. The other parts of the network might preserve their primary functions and connectivity, as was evidenced by the activation patterns in covert speech, the error patterns of Apraxia of Speech, and the realizational problems of stuttering.

### 3.1.2. Experiment 2: articulatory complexity[8]

The 'speech network' that we have discovered in the experiment described above shows a dynamic activation pattern that depends on the degree of execution of the phonological and phonetic code. Does the network activity depend in a similar way on the degree of articulatory/phonetic complexity of speech? There is neurophysiological evidence that the sensorimotor cortex itself reflects the individualized motor control and movement fractionation. In the pioneering work carried out in the 1940s at the Montreal Neurological Institute, Penfield and Jasper [44] established a functional map of the sensorimotor areas along the central sulcus. In particular, they found that electrical stimulation of the precentral gyrus (motor cortex) produced individualized movement. Stimulation of the medial portion of the gyrus led to the movement of (contralateral) toes and foot, while stimulation of the lateral surface evoked movements of the lips, the jaw and the tongue. Systematic movement of the stimulating electrode across the precentral gyrus revealed a

---

[7] It appears to be critical that the anterior insula is involved in the process of speech production at this final realizational stage, and not during the planning stages. Evidence for this comes from Fox et al. [21], who found hyperactivation of the cortex at the level of the insula (with peaks of activation in the anterior insula) in a PET study of stuttering. The hyperactivation of the anterior insula occurred during both overt and covert productions of the stutterers.

[8] Cf. Riecker et al. [50].

complete motor representation of the body. The functional map of the motor area, the s.c. 'homunculus' representation, was established.

In the fMRI experiment described in the previous section the motor cortex was activated only to a small degree in the vicinity of the 'face' area of the motor cortex. Notice that fairly undifferentiated, automatized speech was used in that experiment (repetition of the months of the year). Would the activation of the network change, if we introduced a more differentiated articulatory and phonetic material?

Ten healthy right-handed native speakers of German (five females, five males; median age = 26 years, range 21–32 years) participated in the experiment. None of them had a history of cerebral disorders or diseases of the cranial nerves. The monosyllables [ta] and [stra], the syllable sequence [pataka], the lexical item 'Tagebau' [tagebau] 'strip mining', and horizontal tongue movements served as test materials. Horizontal tongue movements constitute a relatively simple motor activity but strangely they are never used in the articulation of speech (in contrast to vertical tongue movements, which are used in every phonetic code). All the nonsense items in our experiment, i.e. [ta], [stra], [pataka] comply with the phonotactic rules of German but they systematically differ in phonetic/ gestural complexity. [ta] is a prototypical CV syllable; [stra] (CCCV) has a complex onset, the consonants of which are all realised with the same articulator (tongue blade = corona) but illustrate three different articulatory modes (fricative—plosive—approximant). The syllabic sequence [pataka] uses three different places of articulation in its onset consonants (labial—coronal—velar). The same gestural complexity is characteristic of the onsets of the lexical word [tagebau], however, the gestural score for the vowels is more complex than in the case of the [pataka], and the voicing score is different as well (voiceless— voiced—voiced in [tagebau] vs voiceless—voiceless—voiceless in [pataka].

Subjects lay supine in a 1.5 T whole body scanner (Siemens Vision), the heads were secured by means of a foam rubber in order to minimize movement. They were asked to produce the tokens in a monotonous manner, i.e. without prosodic modulation, at a self-paced comfortable speaking rate during the measurement periods of 1 min, respectively, and to refrain from verbal thought during the rest periods in between. It was assumed that this instruction would elicit the participants' habitual speaking rate. The utterances were recorded by means of a directional microphone during functional imaging. The rate of non-speech movements could not be determined during scanning. Prior to the experiment, however, the participants were asked to perform horizontal tongue movements at a tempo of their own preference.[9]

Twenty eight parallel axial slices (thickness = 4 mm, gap = 1 mm) were acquired across the complete brain volume by means of a multislice echoplanar imaging sequence (TE = 46 ms, TR = 6 s, flip angle = 90°, FOV = 192 mm, 64 × 64 matrix). A T1-weighted 3D turbo-flash sequence (MP range, 128 sagittal slices, thickness 1.5 mm, 256 × 256 matrix, field of view 256 mm, TE 4 ms, TR 9.7 ms) served as an anatomical reference. Each task included eight groups of 10 measurements, alternately performed during rest (four groups of measurements) and activation (four groups of measurements). The total scanning time per task amounted to 8 min. MRI data were transformed to an

---

[9] All our speakers performed the horizontal tongue movements with their mouth open. Maybe this is the reason why this type of movement does not occur in speech.
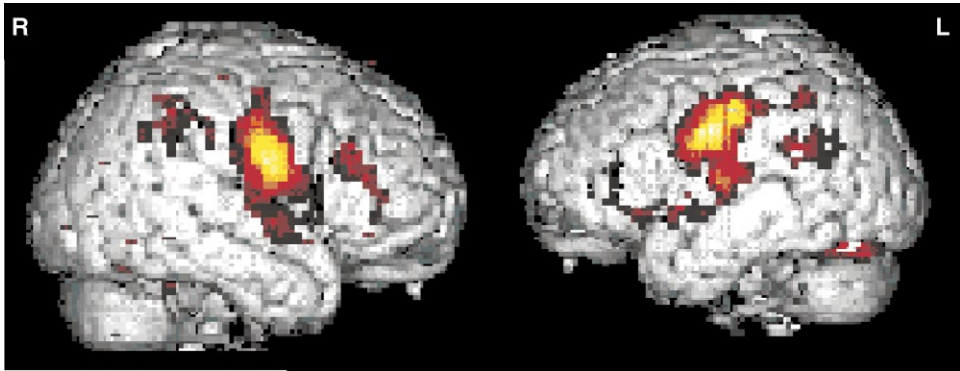
Fig. 8. SPM representation of neural activity following the repetitive production of horizontal tongue movements: paired image subtraction activation—baseline (SPM 96, $n = 10$, $P < 0.001$, corr. $P = 0.05$).

ANALYZE-compatible format and were further processed by the SPM96 software package.

Horizontal tongue movements yielded a strong bilateral hemodynamic reaction in the sensorimotor cortex. The hemodynamic activity extended from this primary localization to the premotor areas. A slight activation of the areas in the parietal lobe (at the level of the supramarginal gyrus left, and the areas dorsal to the left and right supramarginal gyri) was also registered. Weak bilateral hemodynamic activity of the cerebellar cortex was also registered. Consider the SPM representation in Fig. 8.

Repetition of the monosyllables elicited a bilateral and rather symmetric hemodynamic response at the level of the anterior and the posterior bank of the motor sulcus (primary sensorimotor cortex).

Repetition of the prototypical CV syllable [ta] led to strong activation along the central sulcus (bilateral) and some minor activation in the rostal parts of the prefrontal cortex (bilateral) and the caudal parts of the temporal cortex (left dominant), cf. the SPM representation in Fig. 9.
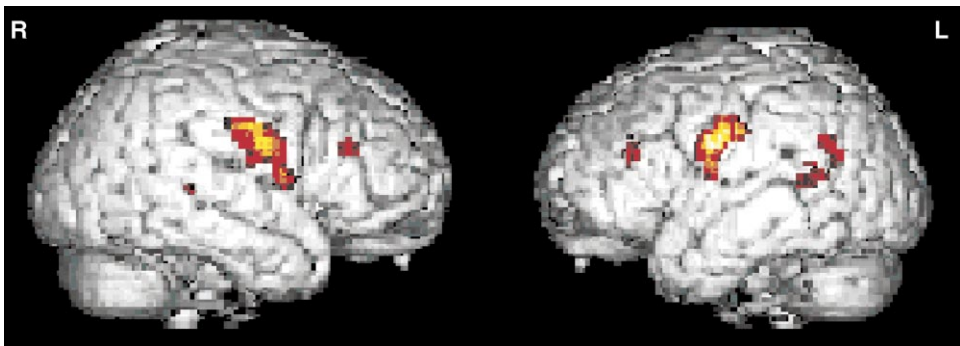


Fig. 9. SPM representation of neural activity in the repetitive production of the syllable [ta]: paired image subtraction activation—baseline (SPM 96, $n = 10$, $P < 0.001$, corr. $P = 0.05$).
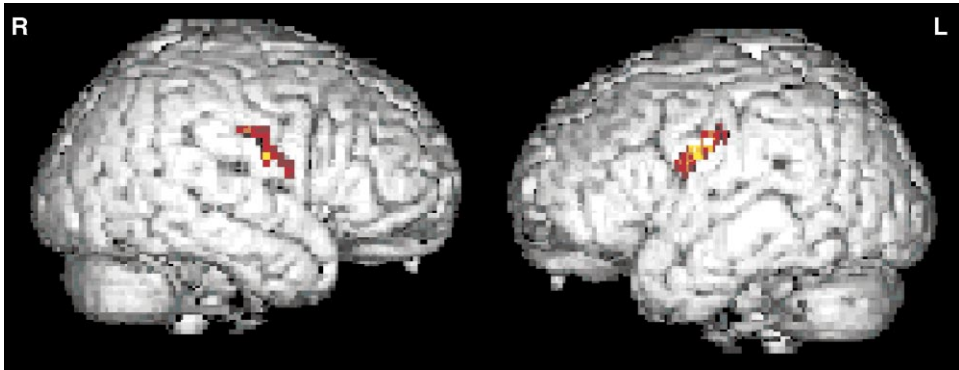
Fig. 10. SPM representation of neural activity in the repetitive production of the syllable [stra]: paired image subtraction activation—baseline (SPM 96, $n = 10$, $P < 0.001$, corr. $P = 0.05$).

Repetition of the articulatorily more complex, but phonologically very widespread monosyllable [stra] revealed a much smaller, rather bilateral activation of the primary motor cortex—consider Fig. 10.

During the production of both the lexical and non-lexical polysyllables a much more limited area of neural activity was observed. In the production of the lexical polysyllable [tagebau] the hemodynamic response was restricted to a small, central area of the motor cortex. The activity was significantly lateralized to the left hemisphere, as shown in the SPM representation in Fig. 11.

The smallest hemodynamic response, and the one which was exclusively lateralized towards the motor cortex of the left hemisphere was registered in the production of the non-lexical, articulatorily complex polysyllable [pataka]. Consider the representation in Fig. 12.

The experiment on articulatory complexity did not replicate the *speech network* that we detected in the 'simple' speech production experiment (see Section 3.1.1). In none of the *speech* tasks of the present experiment were we able to find significantly prominent
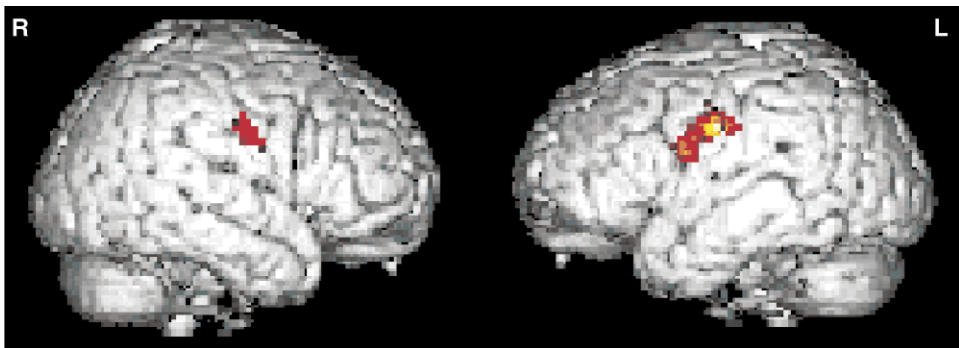


Fig. 11. SPM representation of neural activity in the repetitive production of the lexical polysyllable [tagebau] 'strip mining': paired image subtraction activation—baseline (SPM 96, $n = 10$, $P < 0.001$, corr. $P = 0.05$).
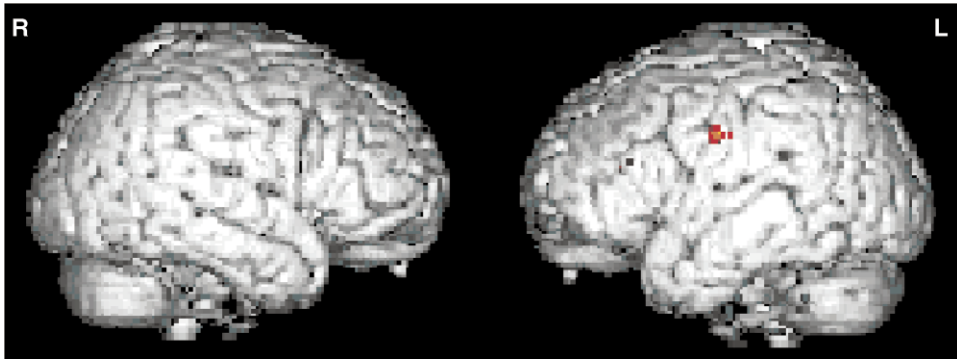
Fig. 12. SPM representation of neural activity in the repetitive production of the non-lexical polysyllable [pataka]: paired image subtraction activation—baseline (SPM 96, $n = 10$, $P < 0.001$, corr. $P = 0.05$).

hemodynamic activity in the anterior insula, the cerebellum or the SMA. The only part of the speech network which was significantly activated in all our tasks was the central sulcus[10] (primary sensorimotor cortex), but the levels of activation in individual tasks also revealed rather unexpected patterns.

There is neurophysiological evidence that the primary sensorimotor cortex mediates the 'fractionation' of movements. Having organized our task in such a way that the articulatory complexity of the required movement was steadily increased [ta]-[stra]-[pataka], we expected the greater cortical activity to correlate with the greater articulatory complexity. Our results show that the neurophysiologically observed 'individualization' of movements with respect to the particular articulators (lips, tongue, pharynx, etc.) is not the preferred mode of producing articulatorily complex speech. On the contrary, the more complex the articulatory movement that the production task required was, the more limited (and the more lateralized) was the area that was revealed in the hemodynamic activity registration. Apparently we did not test the production of the individualized speech movements but the production of the higher order movement organization units which pose small, if any, demands on 'movement fractionation'. What are these units, and how do they contribute to the production of speech?

Levelt, [32,33] following the original proposal by Crompton [11], argued that speakers have access to a *mental syllabary*. A mental syllabary is a set of highly overlearned articulatory gestures which are retrieved as fixed syllable programmes. It has been observed that out of the over 12,000 possible syllables in languages like English or German, only about 500 are systematically used in speech production. It is argued that these high-frequency syllables are not assembled on line by using the segmental and

---

[10] Only non-linguistic and non-phonetic horizontal tongue movements yielded a hemodynamic reaction extending beyond the sensorimotor cortex to the inferior frontal lobes. Since limbic projections target the inferior dorsolateral frontal lobe, the enlarged region of activation during horizontal tongue movements might reflect the larger attentional requirements of this task.

metrical information provided in the phonological representation, but that they are directly stored as complete gestural programmes in the mental syllabary.[11]

The tasks in our experiment were chosen in such a way that all the monosyllables and polysyllables that we used were clearly a part of the 'mental syllabary'. If, as we would stipulate, the representations of the units of the mental syllabary are cognitively different from the gestural representations for phonological units which are assembled on-line, we expect a basic difference between those types of cognitive processing which build the gestural representation on-line and those which retrieve stored articulatory programmes.[12] Hence, our experiment did not overlook the speech network, but rather discovered the cognitive substrate of a 'mental syllabary', a knowledge base which is as central to speech production as is the lexicon to the production of language. Additional hypothesis-driven experiments are on the way. These experiments will have the purpose of verifying the role of syllabary in the production of speech and the role of the sensorimotor cortical areas as the 'storage' device for the mental syllabary.

### 3.2. The neuroanatomical basis of prosody

Prosody is a mode of communication which provides a parallel channel to speech. Prosodic features, unlike other linguistic components, are often produced without conscious intention and are open to forms of interpretation which rely on emotional, non-cognitive processes. The communicative content of many prosodic signals parallels that of stereotypic call vocalizations characteristic of communication systems of subhuman species. It has often been argued that the neuroanatomical basis for these call vocalizations should be fundamentally different from that of the symbolic aspects of human communication.[13] But unlike calls of other species, prosodic organization of human communication is continuous and highly correlated with the semantic, syntactic, morphological and segmental organization of speech, which, as we know, is fully under neocortical control. Regardless of function, there exist only three prosodically active phonetic parameters: duration, intensity and pitch. Are prosody and its phonetic exponents controlled by the neocortex or by the midbrain and the limbic system?

### 3.2.1. Experiment 3: the prosody generator

In our research we assume a highly fractionalised and elaborated model of prosody generation ([32]: 366ff) and test its individual components with experiments designed to reveal the function of the active, healthy brain. The experiments follow the methodological spirit discussed in the first part of this study, i.e. close observation of linguistic components given by a speech production model. The main interest in the experiment

---

[11] The first 'syllabaries' are constructed by children in the very early stages of language development (at around the age of 12 months). Building on ingestion-related articulatory programmes for chewing, sucking and licking and on the simple syllables of the babbling phase, children start developing quite elaborate syllabaries which are 'tuned' to the repertoire of the high-frequency syllables of their native language [33,34]).

[12] This is another example of the much debated *stored vs rule-based* cognitive partition of language [29].

[13] Ploog [46] and Müller-Preuss and Ploog [41] have shown that primate call systems are controlled by caudal midbrain structures which are directly expressive of emotions like fear, aggression, etc. The only neocortical input to these calls is from the limbic parts of the anterior cingulate gyrus [42]).
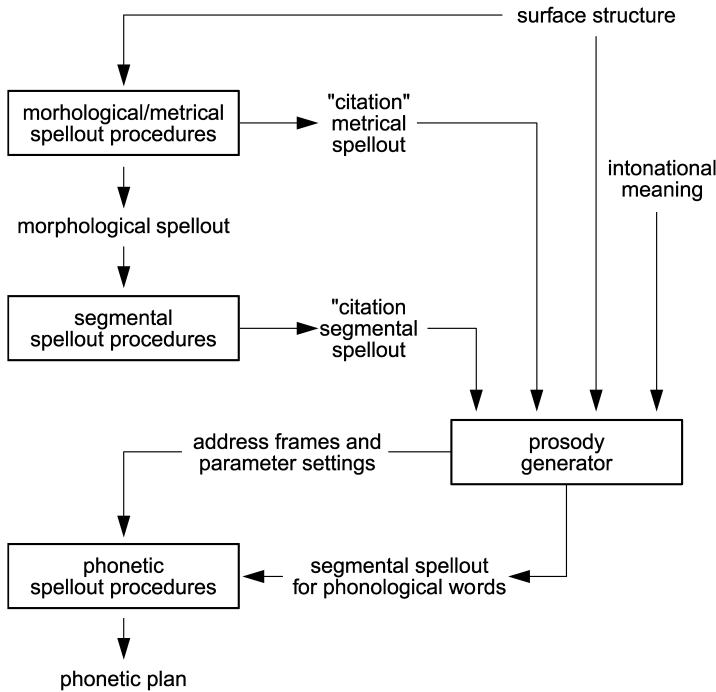
Fig. 13. The model of prosody generation with its linguistic input and phonetic output systems ([32], p. 366).

described below concerns the role of address frames in prosody generation. Fig. 13 displays the model to be tested. Prosody is considered here as an assignment of parameters to frames and nothing else. Other aspects of prosody (parsing, garden pathing, affect perception etc.) have been considered in other studies [56,59].

The study recruited healthy native German subjects (five females, four males, mean age 26.2 years, range 21–32 years) being paid for their participation in the experiment. All participants were right-handed as determined by a standardized inventory, and none of them had a history of neurological disorders. Informed consent had been obtained from each subject. Participants were asked to produce a logatom consisting of five syllables [dadadadada] with various pitch accent types and locations (the FOCUS condition), various boundary tone types (the MODUS condition), and various kinds of emotional state marking (the AFFECT condition). As a baseline for statistical analysis, subjects produced the logatomes [dadadadada], [didididididi], [dodododododo], [dududududu] in a monotonous voice (with a syllable frequency of ca. 5 Hz). The material is summarized in the Fig. 14. We used reiterant syllables and meaningless words in order to minimize the influence of syntactic, semantic, morphological and segmental factors on prosody genera-tion. The aspects of prosody being controlled in this experiment were correlated only with different address frames and parameter settings (cf. Fig. 13).

Subjects lay supine in the MR scanner (1.5 T whole body scanner; Siemens Vision, Erlangen, Germany), their heads being secured by means of a foam rubber in order to
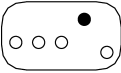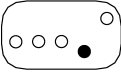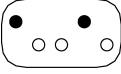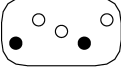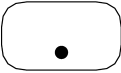
| Stimuli | Reaction paradigms | |
|---|---|---|
| | FOCUS | MONOTONOUS |
| | *dadada**da**da* <br> H*L | *didididi* |
| | *dadada**da**da* <br> L*H | *dadadadada* |
| | ***da**dada**da**da* <br> H*L    H*L | *dududududu* |
| | ***da**dada**da**da* <br> L*H    L*H | *dodododo* |
| | **MODUS** | |
| | *dadada**da**da* <br> L% | *didididi* |
| | *dadada**da**da* <br> H% | *dadadadada* |
| | **AFFECT** | |
| | *dadada**da**da* <br> [happy] | *dududududu* |
| | *dadada**da**da* <br> [sad] | *dodododo* |

Fig. 14. Schematic representation of the experimental material.

minimize movement artifacts. The stimuli were presented visually every 15 s for a period of 3 s each. Pauses between the stimuli extended to a duration of 12 s. During these pauses, subjects produced the displayed prosodic variation. Every 60 s a paradigm change was initiated by an acoustic instruction. Each stimulus was presented eight times. In four out of these eight presentations the 'prosodic' reaction was required. During the other four trials, subjects uttered the respective logatomes in a monotonous manner. The event-related design of the experiment required four complete repetitions of the whole material. Fig. 3 illustrates the data collection procedure used in the experiment. Prior to the experiment, the material and the procedure had been validated in a pilot study performed outside of the MR scanner [37].

Twenty-eight parallel axial slices (thickness = 4 mm, gap = 1 mm) were acquired

across the complete brain volume by means of multislice echoplanar imaging sequence T2*EPI (TE = 39 ms, TR = 3 s, $\alpha$ = 90° FOV = 192 mm, $64^2$ matrix). The fMRI data were processed using the SPM96 software package. Each mean image was coregistered and movement correction as well as space normalization procedures were performed. The normalized fMRI data were filtered (Gaussian filter, 6 mm full width half maximum [FWHM]). Since prior fMRI studies of speech production revealed a delay of the hemo-dynamic response extending from 3 to 6 s, only the images within this time window (cf. the takes marked in bold in Fig. 3) were considered for analysis. For an optimal localiza-tion of significantly activated areas, SPM(t)-maps were superimposed on the structural MR images averaged across all nine subjects.

Fig. 15 displays the pattern of significant neural activity correlated with the various selected experimental tasks. Monotonous speech, using the rest condition as a baseline for subtraction, was characterized by a bilateral neural response within the motor and pre-motor cortex as well as in the superior lateral hemispheres of the cerebellum. These findings accord the data obtained during continuous production of the monosyllable [ta] in the experiment on speech generation (cf. Fig. 9). By subtracting the condition 'mono-tonous speech' from task 1 (simulation of FOCUS), we registered enhanced activity at the level of the left superior gyrus. Comparison of the rest condition with task 2 (simulation of linguistic MODUS) yielded a significant hemodynamic response in the posterior part of the right superior temporal gyrus, extending to the adjacent middle temporal gyrus. Finally, subtracting the rest condition from task 3 (simulation of AFFECT) revealed neural activity in a more anterior part of the right superior temporal gyrus.

The obtained results suggest exclusively neocortical areas to be critically involved in prosody generation. Neither the limbic system nor the midbrain structures displayed activation during the model-controlled prosodic tasks. Human control of prosody, thus, seems to be substantially different from the way in which primates control their emotional calls.

Furthermore, the results support the view that both hemispheres subserve the processing of prosodic features of speech. Obviously, this type of processing is highly localized (superior temporal gyrus) and lateralized in accordance with the required prosodic task (FOCUS assignment vs MODUS assignment and AFFECT expression). The lateralization is not consistent with the distinction between the linguistic and the emotional functions of prosody: Both FOCUS and MODUS assignment are characterised as linguistic functions [18]. It is rather the case that prosodic features which require a short address frame (e.g. focused syllable) are lateralized differently as compared to prosodic features comprising a long address frame (the whole intonational phrase for linguistic modus and paralinguistic affect). Thus, prosodic frame length seems to be the basis of lateralization and not the linguistic/affective function.

There appears to be no single localization with respect to the prosody generator, and not a single specialized network in which its activity is distributed. Prosodic frame length (short, e.g. syllable or mora vs long, e.g. intonational phrase) seems to be the cognitively relevant factor in prosody generation. Converging evidence from clinical studies and more recent neuroimaging studies point to the same conclusion. For example, Behrens [4], who carried out an extensive study of prosodically disturbed clinical populations, observed disturbances in the control of prosodic processing on the syllable level in subjects who
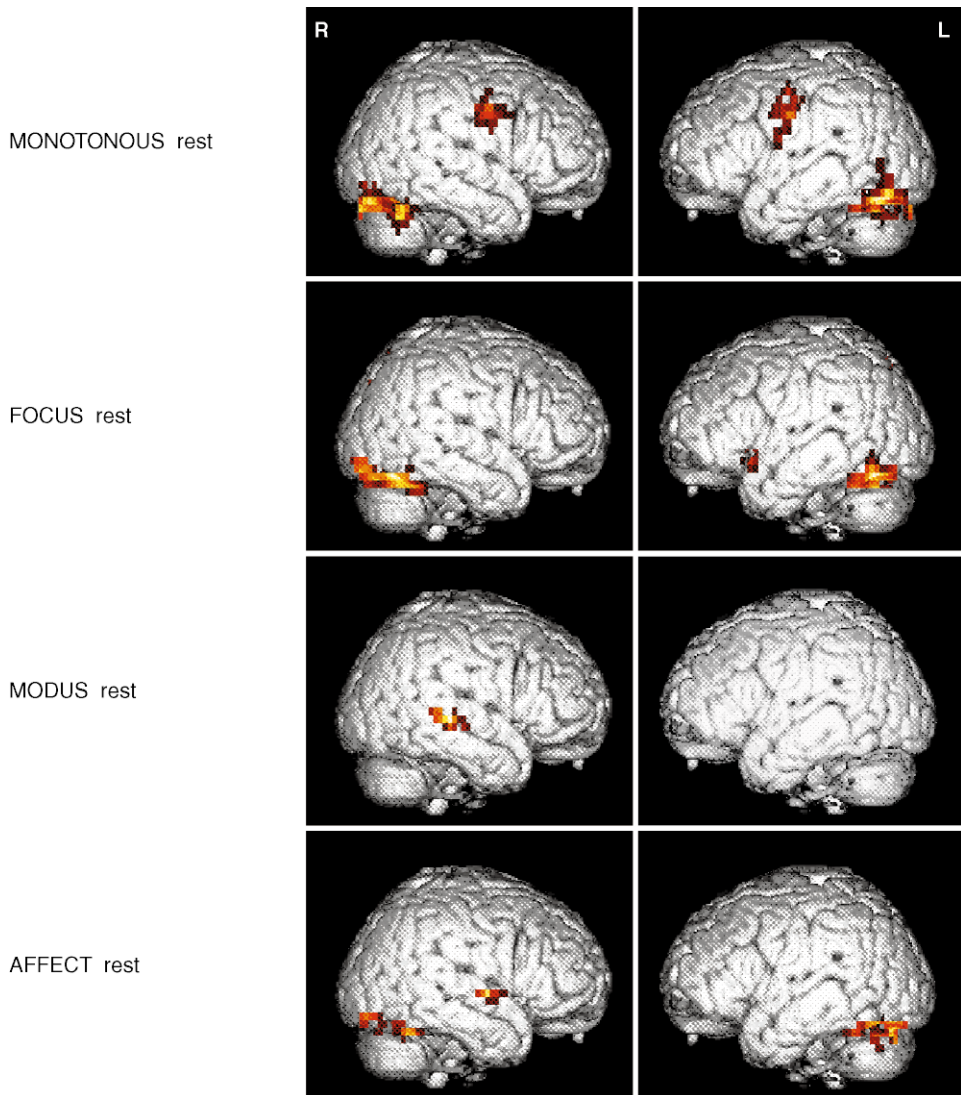
Fig. 15. Statistical Parametric Maps of prosody generation (SPM96). Significance levels for all maps (except for FOCUS-MONOTONOUS): $P < 0.001$, uncorrected (voxel level), $P < 0.05$, corrected for multiple comparisons (cluster level).

suffered damage to the left hemisphere, and problems of processing on the level of the intonational phrase in patients with dysfunction of the right hemisphere. Extensive reviews of disturbed prosody processing in clinical populations also generally support this view [3,18].

More recent neuroimaging studies of prosody perception also lend support to the localizational findings of our production study. Gandour et al. [23], in a study of the perceptual

processing of tone (tone is tied to the smallest prosodic units—syllable and mora), registered increased activity in the left superior temporal gyrus—the localization close to the one that we detected during the production of FOCUS (also tied to the syllabic constituents in our experiment). Wildgruber and co-workers [58], in an fMRI study on the perception of affective prosody, found activation in area 22/42 of the right hemisphere. This region was significantly activated during the task in which hearers were asked to assess the prosodic expressiveness of digitally edited stimuli, while individual acoustic cues (duration, intensity and pitch) were altered systematically. Individual acoustic cues had no significant effect neither on lateralization nor on activation levels, instead, they all activated the same area in the right hemisphere. The localization of this area is very similar to the one found in our experiment during the production of affective prosody. The fact that apparently the same areas are involved in both the production and perception of prosody has often been noted in clinical studies [43,57]. It may be the case that the brain simply does not care whether effective neocortical processing is further used in perception or production. As long as it is efficient, the processing of perception and production may be localized in a single area. At least prosodic processing appears to be generalized over both production and perception. It is also strictly localized to small areas of the neocortex. On top of that it appears to be particularly 'cognitively prominent'.

Reconsider the SPM representations in Fig. 15. The top panel shows the activity registered during the production of monotonous speech. What is depicted is actually the result of subtracting the rest condition from the images activated during the production of monotonous speech. We see the expected activation of the 'face' areas of the sensorimotor cortex and the bilateral activity in the cerebellum. In the middle panel we subtracted monotonous speech [dadadadada] from the prosodic focus activation [dadadaDAda]— as expected, activation of the face area and the cerebellum disappear from the SPMs. Note, however, that in all other subtractions represented in Fig. 15 the face area is not present! Whenever prosodic processing is involved, the activity of the correlated cortical areas overrides the activity of the speech areas. Although subjects speak (they produce simple [dadada..], [dididi..], [dududu..] articulatory movements) in all tasks, the articulatory production is evident only in monotonous speech. Does this mean that prosodic activity in the FOCUS, MODUS and AFFECT conditions eliminates the articulatory requirements of speech? Obviously not! Speech is present and cognitive activation of the speech areas is present too, but at a much lower level of significance. In Fig. 16 we show the SPM representation of the subtraction of the rest condition from the affective prosody condition (cf. Fig. 15—lowest panel). The only difference consists in the lowering of the significance level from $P < 0.001$ to $P < 0.005$. As expected, the activation of the face area (in the left hemisphere) and of the cerebellum occurs again. Linguistic tasks and speech tasks coincide as expected, but the language task is clearly more prominent. It still has to be examined whether this 'cognitive prominence' is a general feature of all language tasks.

### 3.3. Syntactic processing

Finding a pair of tasks that isolates prosody generation from other aspects of speech production was not at all easy. After all, prosody is always present when we use overt speech. However, it is an even bigger challenge to construct a set of tasks that isolate
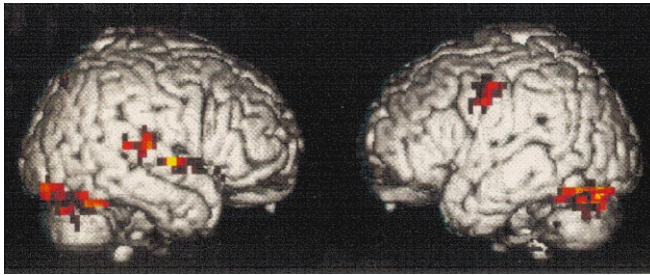
Fig. 16. SPM representation of prosodic processing (AFFECT)—baseline. SPM 96, $n = 9$, $P < 0.005$, corr. $P = 0.05$.

syntactic processing from general cognition. Syntactic processing has been considered to be absolutely central and mandatory to all types of language behaviour at least since Chomky's seminal studies in the 1950s. For some schools of linguistic research it would be theoretically impossible to isolate syntactic processing from semantic or (morpho)lexical processing.[14] Nevertheless the classical assumption of the autonomy of syntax is still a prominent feature of the generative research programme [10]. This assumption is also strongly supported by clinical data, where, following Broca's initial discovery over 100 years ago, compelling evidence for the autonomy of syntactic processing has been collected ([9] for an extensive overview and [17] for a particularly telling case). The lateralization of syntactic processing to the left hemisphere and its localization in the inferior dorsolateral region of this hemisphere ('Broca's area') is also fairly obvious from the clinical data. If we can isolate genuine syntactic processing in a language production task, can we replicate the findings which have formed a basis of (cognitive) neuroscience over the last century?

### 3.3.1. Experiment 4: structure dependency[15]

The present experiment aims to isolate the genuine syntactic processing in speech production by means of event-related fMRI technology.[16] Using the paired-image subtraction methodology, we compare a task which utilizes the phrase structure dependency of syntactic operations with a control task involving unstructured list generation.

Ten native speakers of German (five female, five male, age med. $= 27$) being paid for the participation in the experiment were examined. All participants were right-handed as determined by a standardized inventory, and none of them had a history of neurological disorders. Informed consent had been obtained from each subject.

The subjects lay supine in the MR scanner (1.5 T whole body scanner; Siemens Vision, Erlangen, Germany), their heads being secured by means of a foam rubber in order to

---

[14] E.g. Generative Semantics, LFG, HPSG, Montague Grammar ([54] for a review).

[15] This is just one of a series of experiments concerned with the isolation of the syntactic processing that we have performed. For details [38].

[16] For a general overview of neuroimaging studies of syntactic processing, cf. Hagoort, Brown and Osterhout [25]. Most of the studies in that overview are concerned with the perception of syntactic well-formedness investigated with electrophysiological (EEG, MEG) methods.

minimize movement artifacts. Stimuli were presented visually as lines of text (14 syllables in length) on a computer screen. The test items consisted of German sentences ($n = 16$) with three syntactic constituents [the test task], and German word lists ($n = 16$) containing three elements. In the test task the subjects were asked to read the sentence silently, then manipulate the word order and produce the result with overt speech. As a control, subjects read the word list silently, then moved the second element of the list to the first position and produced the modified list in overt speech. One of the word lists used in the experiment and the required manipulation is given below:

Kaffeemaschine Waschmaschine Bügelmaschine
(coffee machine washing machine ironing machine)
Waschmaschine Kaffeemaschine Bügelmaschine
(washing machine coffee machine ironing machine)

In contrast to the general cognitive task of list reordering, word order manipulation is a grammatically geared task. We take advantage of the fact that German grammar allows a fairly free choice as to which syntactic constituent is followed by the finite verb. The critical point is that only a syntactic constituent but not just any word of a sentence can be moved to the pre-verbal position. We instructed the subjects to start the test sentences differently, which to them meant that they had to replace the initial constituent. An example of the test task requirement is illustrated by the transformation below.

Eine Kundin hat sich über den Manager beschwert
A customer has complained about the manager
Über den Manager hat sich eine Kundin beschwert
About the manager has complained a customer

Both word orders are grammatical and stylistically neutral in standard German. This task is grammatically geared as it requires the identification of an alternative syntactic constituent, which in turn presupposes a syntactic analysis of the sentence, based on an established syntactic routine of German sentence production [26]. The comparison (paired-image subtraction) with the list reordering task serves to differentiate a general symbol manipulation activity (list reordering) and a syntactically grounded reordering activity. Both tasks involve other linguistic (e.g. lexical search) and cognitive subsystems (visual processing, grapheme-phoneme conversion, speech motor control) that should be neutralized by the subtraction design, with only genuine syntactic activities remaining.

The text items of both tasks were put together and presented in a randomized order. fMRI data were obtained across the whole brain volume (28 slices, 4 mm thickness, 1 mm gap) using Echo Planar Imaging (1.5 T, TR 3 s, matrix $64 \times 64$). Spatial transformation of images and statistical analysis were performed using SPM99. The results are based on a random effects analysis of the group data.

Subtracting list reserialization from the syntactic task revealed significant cortical activation, among others, in the left dorsolateral frontal lobe, extending to Broca's area, and at the level of the left temporal lobe, encroaching on Wernicke's area. Furthermore, we registered a significant hemodynamic response of the mesiofrontal region (anterior cingulate cortex [ACC]) and the cerebellum. This pattern of activity (cf. Fig. 17) emerged from a paired-image cognitive subtraction approach, isolating the structure dependency of syntactic
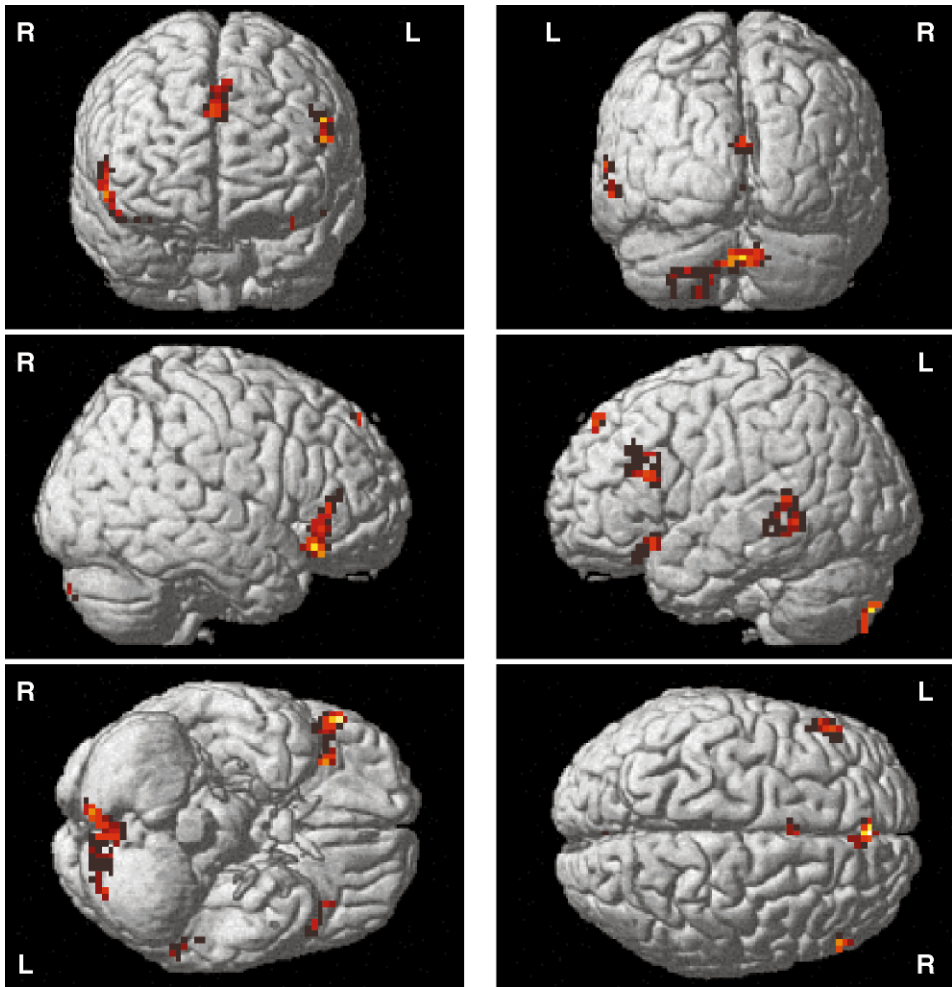
Fig. 17. SPM representation of group ($n = 10$) random efffects analysis of the paired image cognitive subtraction: syntactic transformation—list reserialization.

operations, one of the central aspects of linguistic processing. Hence, we may conclude that the structure dependency of syntactic operations is controlled by the delineated network. What is the status of the individual focal areas of the network and what is the status of the connections between these focal areas? We hope to clarify this question in our future experiments.[17] It has to be stressed, however, that the network we discovered in the present experiment makes a lot of sense taking into account what we already know from clinical and experimental studies.

---

[17] Actually, the described set of tasks is an independent subpart of a larger design that encompasses overt reading, lexical semantic processing, a combination of syntactic and lexical semantic processing, and list manipulation with expanded lexical search [38]).

The role of Broca's area in syntactic processing is beyond any doubt. The aphasiological evidence [9] clearly supports this view (at least all aphasiological evidence on syntax *production* does). Also, studies of neurocognition of language [6] stress the central role of this area in syntactic processing. The hemodynamic response registered in our experiment is particularly strong in the dorso-lateral extensions of it (dorso-caudal part of BA 44). The reason why the activity is not visible across the whole of Broca's area[18] might lie in the subtraction design. It is not unlikely that the list reserialization task is also supported by the structures of the same area and that the subtraction would 'eliminate' the activation of the ventral-lateral prefrontal cortex.

The role of the posterior STG in syntactic processing is less clear. However, aphasiological evidence points to certain syntactic deficits in Wernicke patients. These deficits are not quite clear on the morpho-syntactic level of production (more so as far as perception is considered). However, they become particularly obvious[19] when one considers a more general control of syntactic operations, such as the structure dependence of syntactic operations we tested in our experiment. Wernicke's area is also strongly interconnected with Broca's area, and discharges in both directions would by themselves give rise to the increased hemodynamic responses in both of these areas, if the activity were very strong in any single one. We will return to the neural connectivity issue later in this section.

The activation of the cerebellum in this syntax-specific cognitive processing task might be more of a surprise. The cerebellum plays an important role in the organization of speech ([2] and the experimental results discussed in Section 3.1.). However, speech is equally present in our paired subtraction design (note that we controlled the number of syllables in both tasks), and hence all the areas contributing to overt speech but not to syntactic processing should be eliminated. The activation in the cerebellum suggests the importance of this structure for syntax-specific processing. Recent clinical findings suggest involvement of the cerebellum in syntactic and other higher cortical processing [12,20,36,53]. There is also an ongoing discussion of the role of the cerebello-cerebral network in speech and language functions. There are explicit axonal connections between the cerebellum and the prefrontal cortex in non-human primates. Middelton and Strick [40] identified an afferent pathway connecting the cerebellum with area 46 of the left prefrontal cortex in primates. PET experiments of word generation also show the coactivation of the left prefrontal cortex and contralateral cerebellum [8]. The coactivated cerebral-cerebellar network has been claimed to function as a working memory for time sensitive operations [35]). It still has to be shown in what sense the *working memory* concept is involved in specifically syntactic operations. It is remarkable, however, that networks and structures usually connected with working memory are coactivated by a syntax-specific task.

The dorsal prefrontal cortex (DPF) is exactly one such area that has always been associated with memory and almost never with language. Petrides [45] argued that the DPF area activated in our task has a more general role in the higher level organization of

---

[18] The neuroanatomical definition of Broca's area is still subject to research itself (cf. Uylings et al., 1999). Whereas most researchers restrict it to BA 44, 45, others see it extend to BA 47 (cf. Harasty et al., 1997) and still others consider the whole ventral-lateral prefrontal cortex to constitute the 'language' area [15]).

[19] Cf. [27] for a seminal study of these issues. Cf. also [55] and [61] on distorted thematic relations in Wernicke's aphasia.

basic working memory maintenance. Apparently, such a higher level organization of working memory is required for structure-dependent reserialization but not in 'simple' list reorganization. The DPF is also tightly connected to the dorso-caudal part of Broca's area, a connection which will be discussed later in this section.

Another cortical structure which was activated in the syntactic task, and which is usually considered to be less relevant for linguistic processing is the mesiofrontal cortex [ACC]. Lesions of the ACC were never associated with language disorders. Buried in the depths of the frontal lobes the ACC was interpreted to be a component of the limbic system modulating automatic responses to pain and discomfort. However, as soon as neuroimaging became a daily practice, the ACC has been found to be activated in tasks such as semantic association, visual judgement, recent memory, remote memory etc. Actually, the number of studies referring to the cingulate increased by 110% since the outset of the first PET experiments (cf. [42] for a review). Why this huge increase? The methodology of paired-image subtraction used in PET and fMRI stipulates that two tasks that are used in the subtraction should consist of closely matched cognitive components (reading, speaking, reserialization), but what they do not control is the immanent 'difficulty' of the task. Paus et al. [42] in their review of 387 of such subtractions have found that the ACC was activated when the more difficult tasks were compared with the less difficult ones. Comparing task difficulty is, however, a very sensitive issue—we would not know in what sense our syntactic task was more difficult than the list reserialization task! Nevertheless, the ACC seems to establish its role as a supervisory attentional system in all behaviour ([24]: 459 ff). Our experimental results do not falsify this role.

The *syntax* network, analogously to the speech network discussed in 3.1. is widely distributed in the human brain. The focal areas of the network (Broca's area, Wernicke's area, Cerebellum, DPF, ACC) are very strongly interconnected. Actually, the connections of these focal areas are analogous to the ones found in animals. Deacon [13,14] has presented the results of in vivo axonal tracer experiments examining the connectional organization of the monkey homologue to Broca's language area. The result is a network which is astonishingly similar to the syntax (and the speech) network that we found in our fMRI study. The network, with the original description of its connectivity is given in Fig. 18.

Our network of specific syntactic processing (cf. Fig. 17) is homologous to the network of language processing based on the animal experiment (cf. Fig. 18). This brings us directly back to the beginning of this study, when we claimed that neurocognitive studies of language will always be impaired by the lack of the animal model. Well, from the neuroanatomical point of view the animal model does not look that bad! Why don't monkeys use their homologous neuroanatomy for the same function (i.e. language)? Neuronal homology is far from meaning functional identity, but how far? Before we indulge in a difficult speculation let us return to more direct discoveries about the speaking brain.

## 4. General conclusion: the vanishing networks

In the experiments described in this study we managed to discover neuroanatomical structures underlying the production of speech and structure dependent syntactic
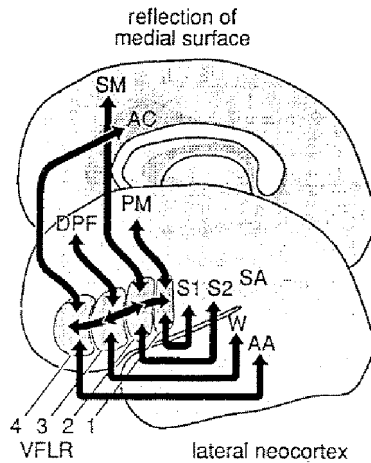
Fig. 18. Connectional organization of the ventral frontal language region (VFLR; from Deacon [15]: 56). The VFLR is subdivided into four cortical areas: the laryngeal-oral motor area (1), a transitional premotor area (2), and two ventral prefrontal areas (3, 4). The ventral prefrontal areas are connected with Wernicke's area (W), association auditory areas (AA), the anterior cingulate cortex (AC), and the dorsal prefrontal cortex (DPF). Areas 1 and 2 are connected to somatosensory areas for the face and oral cavity (S1, S2), inferior parietal sensory association area (SA), and to the supplementary motor (SM) and other premotor areas (PM). All connections are reciprocal although not symmetrical in layer of termination within cortex or topographic specificity with a cortical area. Motor area (1) is involved in the production of oral-vocal movements, the premotor area (2) in organizing complex sequences of movements, the caudal ventral prefrontal area (3) in grammatical and conditional association functions, and the rostral ventral prefrontal area (4) in word association and verbal shortterm memory functions (cf. [15]).

operations. Both of these neural substrates of language can be represented as networks of focal areas widely distributed in the brain. We were able to identified in a general speech production experiment the speech network consisting of *Motor Cortex— Supplementary Motor Area—Cerebellum—Anterior Insula* of the dominant hemisphere. We were also successful in observing some of the dynamic processing involved in speech production. In particular, we noticed that some focal areas of the general network are present only in specific speech production tasks (e.g. activation of anterior insula in overt speech). We expect this network to function as a 'mask' in which the activation will increase or decrease in correlation with the specificity and complexity of the speech tasks. The results of the specific experiments on articulatory and phonetic complexity allowed us to reconsider this prediction. Instead of verifying the network we started loosing it. The phonetically most specific functions were localized most focally. As soon as we asked the brain to perform linguistically constrained complex tasks it stopped showing us the widely distributed activity and told us that *abstract function is localized in focal areas*. This was most clear in the prosody experiment, and we expect the same result from the follow-ups of the syntax experiment. We actually predict that the distributed activity registered in our syntax experiment will become irrelevant once we start testing more lexically constrained aspects of syntactic organization. We expect that in those cases very small

cortical areas will emerge as cognitively prominent. These areas will probably be a part of the large-scale network that we discovered, but the network itself will become 'invisible'. It could be the case that this particular characteristic of neuronal processing—the possibility of breaking down the large-scale neural network and storing the function in a focal area 'disconnected' from the network—underpins the linguistic brain.

The results also show different modes of cognitive representation for on-line speech/language generation on the one hand, and the cognitive representation of the knowledge sources (the syllabary, the mental lexicon and the world knowledge) required for the task of language and speech generation. On-line generation of very general language and speech tasks involves a distributed cortical network, whereas a task demanding strong involvement of the knowledge sources shows a much more localized and modularized neural correlate. This result is particularly conspicuous in the area of the speech specific knowledge source—the syllabary—where gestural complexity correlates with focal cortical activity. We expect the same tendency to emerge in the cortical correlates of other knowledge sources. The more specific the task is, the more focal its neuro-cognitive correlates will be. We hope that this general prediction of our research will be verified by further studies of the slowly emerging picture of the speaking brain.

## References

[1] Ackermann H, Wildgruber D, Daum I, Grood W. Does the cerebellum contribute to cognitive aspects of speech production? A fMRI Study in Humans. Neuroscience Letters 1998;247:187–90.

[2] Ackermann H, Hertrich I. Cerebellum and speech processing. Journal of Neurolinguistics—Special Issue on Language and the Cerebellum 2000;13:95–116.

[3] Baum S, Pell M. The neural bases of prosody: insights from lesion studies and neuroimaging. Aphasiology 1999;13:581–608.

[4] Behrens S. Characterizing sentence intonation in a right hemisphere-damaged population. Brain and Language 1989;37:181–200.

[5] Bressler SL. Large-scale cortical networks and cognition. Brain Research Reviews 1995;20:288–304.

[6] Brown CM, Hagoort P. In: Brown CM, Hagoort P, editors. The neurocognition of language. Oxford: Oxford University Press, 1999.

[7] Buckner R, Bandettini P, O'Craven K, Savoy R, Petersen S, Reichle M, et al. Detection of cortical activation during averaged single trials of a cognitive task using fMRI. Proceedings of the National Academy of Sciences of the United States of America 1996;93:14878–83.

[8] Cabeza R, Nyberg L. Imaging cognition: an empirical review of PET studies with normal subjects. Journal of Cognitive Neuroscience 1997;9:1–27.

[9] Caplan D. Neurolinguistics and linguistic aphasiology: an introduction. Cambridge: Cambridge University Press, 1987.

[10] Chomsky N. The minimalist program. Cambridge, MA: MIT Press, 1995.

[11] Crompton A. Syllables and segments in speech production. In: Cutler A, editor. Slips of the tongue and language production. Berlin: Mouton, 1982. p. 109–62.

[12] Daum I, Ackermann H. Cerebellar contributions to cognition. Behavioural Brain Research 1995;67:201–10.

[13] Deacon T. Connections of the inferior periarcuate area in the brain of Macaca fascicularis: an experimental and comparative investigation of language circuitry and its evolution. PhD. Thesis, Harvard University, 1984.

[14] Deacon T. The neural circuitry underlying primate calls and human language. Human Evolution 1989;4:367–401.

[15] Deacon T. Brain-language coevolution. In: Hawkins J, Gell-Mann M, editors. The evolution of human languages. New York: Addison-Wesley, 1992. p. 49–83.

[16] Dogil G, Mayer J. Selective phonological impairment: a case of apraxia of speech. Phonology 1998;15:143–88.

[17] Dogil G, Haider H, Husmann R, Schaner-Wolles C. A radical autonomy of syntax: evidence from transcortical sensory aphasia. Aphasiology 1995;9:577–602.

[18] Dogil G. Understanding prosody. In: Rickheit G, Hermann Th, Deutsch W, editors. Psycholinguistics: an international handbook. Berlin: de Gruyter, 2000.

[19] Dronkers N. A new brain region for coordinating speech articulation. Nature 1996;384:159–61.

[20] Fabbro F, Moretti R, Bava A. Language impairments in patients with cerebellar lesions. Journal of Neurolinguistics—Special Issue on Language and the Cerebellum 2000;13:173–188.

[21] Fox PT, Ingham RJ, Ingham JC, Hirsch T, Downs H, Martin C, Jerabek P, Glass T, Lancaster JL. A PET study of the neural systems of stuttering. Nature 1996;328:158–62.

[22] Friston KJ. Data analysis: basic concepts and overview. In SPM Short Course Notes. Technical Report, UCL, Wellcome Department of Cognitive Neurology, 1997.

[23] Gandour J, Wong D, van Lancker D, Hutchings D. A PET investigation of speech prosody in tone languages. Brain and Language 1997;60:192–4.

[24] Gazzaniga M, Ivry R, Mangun G. Cognitive neuroscience. New York: W.W. Norton, 1998.

[25] Hagoort P, Brown CM, Osterhout L. The neurocognition of syntactic processing. In: Brown CM, Hagoort P, editors. The neurocognition of language, 1999. p. 273–316.

[26] Haider H. Deutsche syntax: generativ. Tübingen: Niemeyer Verlag, 1993.

[27] Heeschen C. Agrammatism versus paragrammatism: a ficticious opposition. In: Kean M-L, editor. Agrammatism. San Diego: Academic Press, 1985. p. 207–48.

[28] Indefrey P, Gruber O, Brown C, Hagoort P, Posse S, Kleinschmidt A. Lexicality and not syllable frequency determines lateralized premotor activation during the pronunciation of the word-like stimuli: an fMRI study. Human brain mapping 1998;7:S4.

[29] Jaeger J, Lockwood A, Kemmerer D, Van Valin R, Murphy B, Khalak H. A positron emission tomographic study of regular and irregular verb morphology in English. Language 1996;72:451–97.

[30] Josephs O, Turner R, Friston K. Event-related fMRI. Human brain mapping 1997;5:243–8.

[31] Lebrun Y. Apraxia of speech: a critical review. Journal of Neurolinguistics 1990;5:379–406.

[32] Levelt WJM. Speaking: from intention to articulation. Cambridge, MA: MIT Press, 1989.

[33] Levelt WJM. Producing spoken language: a blueprint of a speaker. In: Brown C, Hagoort P, editors. The neurocognition of language, Oxford: Oxford University Press, 1999. p. 83–122.

[34] MacNeilage P. The frame/content theory of evolution of speech production. Behavioral and Brain Sciences 1998;21:499–511.

[35] Mangels J, Ivry R, Shimizu N. Dissociable contributions of the prefrontal and neocerebellar cortex to time perception. Cognitive Brain Research 1998;7:15–39.

[36] Marien P, Engelborghs S, Pickut B, de Deyn P. Aphasia following cerebellar damage: fact or fallacy? Journal of Neurolinguistics—Special Issue on Language and the Cerebellum 2000;13:145–171.

[37] Mayer J, Dogil G, Wildgruber D, Riecker A, Ackermann H, Grodd W. Prosody in speech production: a fMRI study. San Francisco 1999;ICPhS'99:635–8.

[38] Mayer J, Haider H, Dogil G, Ackermann H, Erb M, Riecker A, Wildgruber D, Grodd W. Cognitive substrate of syntactic operations—evidence from fMRI. NeuroImage—Proceedings of the Human Brain Mapping Conference, San Antonio, Texas, 2000.

[39] Mennon R, Kim S-G. Spatial and temporal limits in cognitive neuroimaging with fMRI. Trends in Cognitive Sciences 1999;3:207–16.

[40] Middelton F, Strick P. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. Science 1994;266:458–61.

[41] Müller-Preuss P, Ploog D. Central control of sound production in mammals. In: Lewis B, editor. Bioacoustics—a comparative study. London: Academic Press, 1983. p. 125–46.

[42] Paus T, Koski L, Caramanos Z, Westbury C. Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: a review of 107 PET activation studies. NeuroReport 1998;9:37–47.

[43] Pell M, Baum S. The ability to perceive and comprehend intonation in linguistic and affective contexts by brain damaged adults. Brain and Language 1997;57:80–99.

[44] Penfield W, Jasper H. Epilepsy and the functional anatomy of the human brain. Boston: Little and Brown, 1954.

[45] Petrides M. Frontal lobes and behaviour. Current Opinion Neurobiology 1994;4:207–11.

[46] Ploog D. Neuroethnological foundations of human speech. In: Deecke J, Eccles J, Mountcastle V, editors. From neuron to action. Berlin: Springer, 1990. p. 365–74.

[47] Posner M, Raichle M. Images of mind. New York: W.H. Freeman, 1994.

[48] Price C, Friston K. Cognitive conjunctions: a new approach to brain activation experiments. NeuroImage 1997;5:261–70.

[49] Riecker A, Ackermann H, Wildgruber D, Dogil G, Grodd W. Opposite hemispheric lateralization effects during speaking and singing at motor cortex. Insula and Cerebellum. NeuroReport 2000;11:1997–2000.

[50] Riecker A, Wildgruber D, Ackermann H, Mayer J, Dogil G, Haider H, Grodd W. Articulatory/phonetic sequencing at the level of the anterior perisylvian cortex: a functional magnetic resonance imaging (fMRI) study. Brain and Language 2000 (in press).

[51] Rosenbeck J, McNeil M, Aronson A. Apraxia of speech: physiology, acoustics, linguistic management. San Diego: College Hill Press, 1984.

[52] Rugg M. Functional neuroimaging in cognitive neuroscience. In: Brown CM, Hagoort P, editors. The neurocognition of language. Oxford: Oxford University Press, 1999. p. 15–37.

[53] Schmahmann J. In: Schmahmann J, editor. The cerebellum and cognition. San Diego: Academic Press, 1997.

[54] Sells P. Lectures on contemporary syntactic theories. CSLI, Stanford, 1985.

[55] Shapiro L, Gordon B, Hack N, Killackey J. Verb-argument structure processing in complex sentences in Broca's and Wernicke's aphasia. Brain and Language 1993;45:423–47.

[56] Steinhauer K, Alter K, Friederici A. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nature Neuroscience 1999;2:191–6.

[57] van Lancker D. Cerebral lateralization of pitch cues in the linguistic signal. International Journal of Human Communication 1980;13:227–77.

[58] Wildgruber D, Kischa U, Ackermann H, Klose U, Grodd W. Dynamic pattern of brain activation during sequencing of word strings evaluated by fMRI. Cognitive Brain Research 1999;7:285–94.

[59] Wildgruber D, Pihan H, Erb M, Ackermann H, Grodd W. Differential activation patterns during discrimination of affective prosody: influence of acoustics, emotional valence, accuracy and sex. NeuroImage 1999;9:S1018.

[60] Ziegler W, von Cramon D. Die sprechapraxie—eine apraktische störung? Fortschritte Neurologischer Psychiatrie 1989;57:198–204.

[61]  Zurif E, Swinney D, Prather P, Solomon J, Bushell C. An on-line analysis of syntactic processing in Broca's and Wernicke's aphasia. Brain and Language 1993;45:448–64.


## Further reading

[62]  Deacon T. Cortical connections of the inferior arcuate sulcus cortex in the macaque brain. Brain Research 1992;573:8–26.
[63]  Deacon T. The symbolic species. New York: W.W. Norton, 1997.
[64]  Dogil G. The phonological and acoustic form of neologistic jargon aphasia. Clinical Linguistics and Phonetics 1989;3:265–79.
[65]  Frackowiak R, Friston K, Frith C, Dolan R, Mazziotta J. Human brain function. London: Academic Press, 1997.
[66]  Levelt WJM, Wheeldon L. Do speakers have access to a mental syllabary. Cognition 1994;50:239–69.
[67]  Ojemann G, Mateer Ch. Human language cortex: localization of memory syntax, and sequential motor-phoneme identification systems. Science 1979;205:1401–3.
[68]  Paus T, Perry D, Zatorre R, Worsley K, Evans A. Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. European Journal of Neuroscience 1996;8:2236–46.
[69]  Wildgruber D, Ackermann H, Klein M, Riecker A, Grodd W. Brain activity during identification of affective speech melody: influence of emotional valence and sex. NeuroImage 2000;11:S341.
[70]  Wildgruber D, Ackermann H, Klose U, Kardatzki B, Grodd W. Functional lateralization of speech production at primary motor cortex: a fMRI study. NeuroReport 1996;7:2791–5.