Phonological manipulation between speech perception and production activates a parieto-frontal circuit

Claudia Peschke a,b, Wolfram Ziegler c, Juliane Eisenberger c, Annette Baumgaertner a,d,*

a University Medical Center Hamburg-Eppendorf, Department of Systems Neuroscience, Martinistraße 52, 20246 Hamburg, Germany
b Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany
c City Hospital Bogenhausen, Clinic for Neuropsychology, Clinical Neuropsychology Research Group, Dachauer Str. 164, D-80992 München, Germany
d Fresenius University of Applied Sciences, Department of Speech and Language Pathology, Hamburg, Germany

ABSTRACT

Repetition has been shown to activate the so-called ‘dorsal stream’, a network of temporo-parieto-frontal areas subserving the mapping of acoustic speech input onto articulatory-motor representations. Among these areas, a region in the posterior Sylvian fissure at the temporo-parietal boundary (also called ‘area Spt’) has been suggested to play a central role particularly with increasing computational demands on phonological processing. Most of the relevant evidence stems from tasks requiring metalinguistic processing. To date, the relevance of area Spt in natural phonological operations based on implicit linguistic knowledge has not yet been investigated. We examined two types of phonological processes assumed to be lateralized differently, i.e., the processing of syllabic stress versus subsyllabic segmental processing. In two ways, subjects modified an auditorily presented pseudoword before reproducing it overtly: (a) by a prosodic manipulation involving a stress shift across syllable boundaries, (b) by a segmental manipulation involving a vowel substitution. Manipulation per se was expected to engage area Spt. Segmental compared to prosodic processing was expected to reveal predominantly left lateralized activation, while prosodic compared to segmental processing was expected to result in bilateral or right-lateralized activation. Contrary to expectation, activation in area Spt did not vary with increased phonological processing demand. Instead, area Spt was engaged regardless of whether subjects simply repeated a pseudoword or performed a phonological manipulation before reproduction. However, for both segmental and prosodic stimuli, reproduction after manipulation (compared to repetition) activated the left intraparietal sulcus and left inferior frontal cortex. We propose that these parieto-frontal regions are recruited when the task requires phonological manipulation over and above the more automated transfer of auditory into articulatory verbal codes, which appears to involve area Spt.

When directly contrasted with prosodic manipulation, segmental manipulation resulted in increased activation predominantly in left inferior frontal areas. This may be due to an increased demand on phonological sequencing operations at the subsyllabic phoneme level. Contrasted with segmental manipulations, prosodic manipulation did not result in increased activation, which may be due to a lower degree of morphosyntactic and to syllable-level processing.

© 2011 Elsevier Inc. All rights reserved.

Introduction

For decades, the link between acoustic speech information and the conceptual system has been central to research on the processing of auditory–verbal information. In contrast, the interface between auditory speech perception and the speech motor system has only recently become the focus of intensive research. Recent studies show that an auditory–motor link develops and is strengthened during a period of speech acquisition in childhood, when knowledge is acquired about how sound translates into articulation (Kuhl, 2000). There is strong evidence that the link between speech perception and production persists into adulthood. Experiments using auditory feedback perturbations have shown that adults subconsciously modify their own speech productions to counteract artificially induced shifts in pitch or formant frequency (Houde and Jordan, 2002; Larson et al., 2000; Purcell and Munhall, 2006; Tourville et al., 2008), suggesting that perceptual input may play a crucial role in guiding speech motor programming (Guenther, 2006). In support of...
this notion, adult speakers have been shown to unintentionally imitate incidental acoustic properties of linguistic stimuli during repetition (Kappes et al., 2009).

The auditory–motor integration of speech is thought to involve temporoparietal and frontal regions, usually referred to as the “dorsal stream” (Hickok and Poeppel, 2004). Within the dorsal stream, a special role is ascribed to a region in the posterior Sylvian fissure at the temporoparietal boundary (i.e., area Spt). As part of the left posterior planum temporale, area Spt has been shown to activate both during speech perception and production, suggesting that its specific role is to map acoustic speech signals onto articulatory representations (Buchsbaum et al., 2001; Hickok and Poeppel, 2004, 2007; Okada and Hickok, 2006; Papanthassious et al., 2000). Likewise, posterior superior temporal cortex has been suggested to map incoming auditory information onto stored templates derived from auditory experience, which are then used to generate a program for a motor response (Warren et al., 2005).

The strong link between sensory and motor functions implicated in speech processing may be licensed by distinct temporoparietal fiber pathways, as suggested by several structural imaging studies. The arcuate fascicle, a fiber bundle connecting posterior temporal with inferior frontal structures, shows stronger structural maturation in the left hemisphere relative to the right hemisphere across childhood and adolescence (Paus et al., 1999). According to the authors, this structural superiority may facilitate the fast bidirectional transfer of information between auditory and motor regions in the left hemisphere. A strongly left lateralized tempo-parietal white matter pathway was also shown by several diffusion tensor imaging (DTI) studies in adults (Barrick et al., 2007; Buchel et al., 2004; Parker et al., 2005).

One of the DTI studies tracking the arcuate fascicle is consistent with the assumption that Spt is part of this dorsal stream system (Catani et al., 2005). In addition to the direct tempo-frontal tract, this study reported a second indirect tempo-frontal pathway, consisting of two tracts with a connecting relay in inferior parietal cortex (i.e., BA39/BA40). According to the authors, the indirect pathway is used whenever an intervening stage, such as phonological recoding, occurs between auditory input and articulatory output. Other fiber tracking studies have independently shown structural connectivity between the supramarginal gyrus (Parker et al., 2005) or inferior parietal lobule/intraparietal sulcus (Frey et al., 2008), and inferior frontal as well as superior parietal regions, respectively.

The left tempo-parietal junction and specifically area Spt (e.g. Hickok, 2009) have been proposed to be involved in tasks which require temporary storage of phonological information (Buchsbaum and D’Esposito, 2008; Jacquemot and Scott, 2006). This is because area Spt is assumed to “act as an auditory–motor interface that serves to bind acoustic representations of speech with articulatory counterparts” (Buchsbaum and D’Esposito, 2008, p. 13). In fact, several functional imaging studies have revealed an increased involvement of Spt in tasks with increased phonological processing demands. Activation in the dorsal posterior temporal plane is influenced by word length, showing greater BOLD signal changes for multisyllabic compared to monosyllabic words (as in China versus Chinese) or vowel changes (as in woman versus women), operations which are usually performed according to implicit linguistic knowledge. Thus, the primary goal of the present study was to examine the involvement of area Spt in performing phonological manipulations, using naturalistic phonological processing tasks requiring relatively low verbal working memory resources.

Auditory–motor integration generally is assumed to rely on left-hemispheric dorsal stream structures (e.g. Hickok and Poeppel, 2007). While prelexical auditory processing involves the superior temporal gyri of both hemispheres (Hickok, 2009; Hickok and Poeppel, 2004; Hickok et al., 2008), later auditory–verbal processing stages are presumed to involve left lateralized pathways (e.g. Hickok and Poeppel, 2000). However, the lateralization of dorsal stream activation during auditory–motor integration may depend on the nature of the phonological process. Available evidence suggests differential lateralization of segmental versus prosodic processing. Sublexical processes requiring the sequencing of phonemic information primarily involve left hemispheric regions (Ashtari et al., 2004; Burton and Small, 2006; Burton et al., 2000; Gelfand and Bookheimer, 2003; Heim et al., 2003; Jacquemot et al., 2003; Zaehe et al., 2008). In contrast, findings regarding the lateralization of linguistic–prosodic processes have been less consistent. Some lesion studies provide evidence consistent with the processing of prosodic units in the right hemisphere (Bradvik et al., 1991; Weintraub et al., 1981), whereas other studies suggest a left-hemispheric dominance for linguistic prosody (Arciuli and Slowiaczek, 2007; Emmorey, 1987; Van Lancker, 1980). Neuroimaging evidence, comparing the production of rhythmic to isochronous syllable sequences, points to an involvement of right hemisphere regions in linguistic–prosodic processing at the supra-syllabic level (Riecker et al., 2002). Therefore, a secondary goal was to examine the lateralization of dorsal stream activation during segmental versus prosodic phonological manipulation.

To investigate the involvement of area Spt in phonological processing and to determine the role of left versus right hemisphere structures, we used a segmental and a prosodic manipulation task, both based on naturalistic phonological regularities. Subjects were asked to manipulate auditory pseudoword stimuli phonologically according to implicit linguistic knowledge (i.e. from ‘woman’ to ‘women’ or from ‘Chinese’ to ‘Chinese’) before overt reproduction. Verbatim repetition of the stimulus served as the control condition. In task construction we proceeded from the assumption that in order to arrive at the correct reproduction after phonological manipulation, a stimulus first has to be analyzed sequentially, then the phonological information has to be manipulated at the prosodic or the segmental level, respectively, and finally the stimulus has to be reassembled in order to generate the target utterance. In contrast, in the verbatim repetition condition a stimulus only has to be analyzed sequentially and then reassembled for production.

Our main question was whether additional phonological manipulations between speech perception and speech production would preferentially engage area Spt when compared to verbatim repetition. Our secondary question was whether the nature of the employed phonological operation, namely segmental versus prosodic manipulation, would result in preferential engagement of the left and/or the right hemisphere.

We hypothesized that repetition as well as reproduction after manipulation involves structures associated with the dorsal stream,
predominantly superior temporal and inferior frontal areas. Reproduction after manipulation, as compared to verbatim repetition, was expected to reveal increased activation of Spt due to the additional phonological manipulation required. The comparison between the two phonological processes was expected to show predominantly left lateralized activation for segmental and more bilateral or right lateralized activation for prosodic manipulation.

Materials and methods

Participants

Twenty-three healthy right-handed subjects participated in the study. One of the subjects had to be excluded from further analysis, as he misunderstood the task of one of the sessions. All subjects were native speakers of German without any history of serious medical, neurological or psychiatric illness, or hearing loss (mean age 26.8 years, range 21–36 years, eleven females). Hand preference was tested with the 10-item version of the Edinburgh Handedness Inventory (Oldfield, 1971). Subjects had an average laterality quotient of 0.85 (range 0.5–1.0).

The study was performed according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the University Medical Center Hamburg-Eppendorf. All subjects gave written informed consent prior to the experiment.

Stimuli

Pseudowords were used to discourage semantic processing, since we were primarily interested in phonological processing. The materials included a total of 144 pseudowords, and consisted of two different types of speech stimuli for the two phonological processes to be examined. The first type of stimuli, designed to examine prosodic processing (hereafter called ‘PROS’), consisted of 72 two- and three-syllabic pseudowords. The items were modeled after German bisyllabic names of countries (e.g., /’ku:.ba/ (English: Cuba – Cuban). Thus, the materials consisted of 36 bisyllabic pseudocountries and 36 associated trisyllabic pseudolanguages. Notably, these pairs differed by the location of word stress relative to segmental content, i.e., stress on /ku:/ versus /ba:/ in the above example. A pseudoword example occurring in the experimental materials is /’do:.ga/ (pseudo-country) and /’do:.ga:n/ (pseudo-language).

The second type of stimuli was designed to predominantly require segmental processing (hereafter called ‘SEGM’). The materials consisted of 36 monosyllabic pseudonouns preceded by the article “der” (as in /’de: ‘bax/, English: the stream), and 36 associated bisyllabic pseudodiminutives preceded by the article “das” (as in /’das ‘bax:lai/, English: the streamlet). These pairs differed by the frontness of the stressed vowel (umlaut). A pseudoword example of the SEGM materials is /’de: ‘tu:mi/ versus /’da:. ‘tu:mi:lai/.

The speech materials were spoken by an experienced male speaker and recorded in a sound treated booth. Mean durations of the stimuli are listed in Table 1. There was a significant difference between the two types of speech stimuli, with longer durations for the SEGM materials as compared to the PROS materials (t = 4.49; DF = 142; p < 0.001). This difference is explainable by the fact that the model speaker made short pauses between the prefixed article and the noun in the SEGM stimuli. The intensity of all stimuli was set to 80 dB.

Tasks and procedures

We chose to use overt speech production, since a recording of overt verbal responses enabled us to monitor and analyze subjects’ productions with respect to accuracy and response latency.

Subjects performed two different tasks: (1) In the repetition task (hereafter called ‘REPEAT’) participants had to repeat the speech stimuli as accurately as possible; (2) in the manipulation task (hereafter called ‘MANIP’), items had to be manipulated phonologically prior to overt reproduction. For the PROS stimuli, a pseudocountry (e.g., /’do:.ga/) had to be converted into its respective pseudolanguage (and vice versa) with the concomitant stress shift (i.e., /’do:.ga:n/). For the SEGM stimuli, a pseudounoun (e.g., /’de:. ‘tu:mi/) had to be converted into its pseudodiminutive counterpart (and vice versa) with the concomitant vowel change in the stressed syllable (i.e., /’das ‘tu:mi:lai/). Collapsed across items, input and output were identical in the experimental and the control condition. During practice, the manipulation tasks were introduced through implicit learning procedures. Specifically, for the manipulation condition no explicit description was given of the rule underlying the manipulation. Instead, subjects were required to infer the task through implicit learning. The learning procedure started with German real word examples presented by the examiner (e.g., /’ku:.ba/) — (/’ku:.ba:n/) and subjects were then asked to complete a pair when only one stimulus was given. When a participant had given several correct answers in a row, learning was extended to pseudoword examples.

To alert subjects to the beginning of an auditory stimulus, each trial started with the presentation of a visual cue (i.e., a symbol of a loudspeaker) which appeared 500 ms before the onset of each auditory stimulus, and stayed on until its offset. Five hundred ms after the offset of the auditory stimulus, a green dot was presented for 300 ms. Subjects were instructed to begin speaking as soon as the green dot had disappeared. During the rest of the trial the screen remained dark. The SOA was jittered between 5.5 and 10.5 s (see Fig. 1).

Stimulus presentation

The visual cues (a gray symbol of a loudspeaker, a green dot) were presented centrally on a dark background. They were projected onto a translucent screen located behind the head coil, which was viewed by the subjects via a mirror mounted on top of the coil. The auditory stimuli were presented via an MR-compatible electrodynamic headphone with a built-in dual-channel microphone (MR Confon GmbH, Magdeburg, Germany, http://www.mr-confon.de) for the combined presentation and recording of speech. Subjects’ responses were recorded with the sound recording software PhonOr implemented in the ConFon system which automatically preprocessed the dual channel recordings by reducing the whole frequency spectrum of the scanner noise by 20 dB in relation to the speech signal. Vocal response recordings had to be synchronized with the recordings of scanner pulses and stimulus onsets. To this end, scanner pulses were recorded in two different channels: one was routed to the standard presentation software which also recorded stimulus onset times, and the other was fed into the vocal response recording channel as an additional signal. This procedure ensured perfect synchronization of the general presentation parameters with the recording of vocal responses. Vocal responses were recorded by ConFon software running on a Samsung R65-T5500 Canpiro laptop situated outside the scanner room, and were saved as wavfiles. The volume for the auditory presentation was set to a comfortable loudness level individually for each subject in a preceding test scan. The task

<table>
<thead>
<tr>
<th>Stimulus group</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>All stimuli (n = 144)</td>
<td>889</td>
<td>185</td>
<td>544–1288</td>
</tr>
<tr>
<td>PROS stimuli (n = 72)</td>
<td>824</td>
<td>177</td>
<td>544–1136</td>
</tr>
<tr>
<td>Pseudocountries (n = 36)</td>
<td>660</td>
<td>71</td>
<td>544–810</td>
</tr>
<tr>
<td>Pseudolanguages (n = 36)</td>
<td>988</td>
<td>59</td>
<td>859–1136</td>
</tr>
<tr>
<td>SEGM stimuli (n = 72)</td>
<td>954</td>
<td>170</td>
<td>639–1288</td>
</tr>
<tr>
<td>Pseudonouns (n = 36)</td>
<td>802</td>
<td>76</td>
<td>639–962</td>
</tr>
<tr>
<td>Pseudodiminutives (n = 36)</td>
<td>1105</td>
<td>73</td>
<td>970–1288</td>
</tr>
</tbody>
</table>
sequence was controlled by a PC running "Presentation" software (Neurobehavorial Systems, http://www.neurobs.com/).

Experimental design

The study had a one-factorial within-subject design with the factor PHONOLOGICAL PROCESS (PROS versus SEGm). Verbatim repetition (REPEAT) served as the baseline condition. The experiment contained eight runs, four for each task. There were two runs for the manipulation of the prosodic and two runs for the manipulation of the segmental materials; analogously, to collect corresponding baseline data, there were four runs for the REPEAT task. Task and phonological process remained the same within a run. Instructions regarding task and process were visually presented for three seconds at the beginning of each run. Each run consisted of 36 experimental trials and took about 5.5 min to complete. Half of the 36 trials contained bisyllabic and the other half trisyllabic pseudowords. The sequence of runs was pseudorandomized, with the restriction that a task or phonological process could occur maximally two times in a row and that, to the degree possible, each run appeared in all positions of potential run sequences across subjects. The order of pseudowords within a run was pseudorandomized in a way that maximally three bi- or trisyllabic stimuli, respectively, occurred in a row. The trial sequence was fixed within runs for all subjects. The experiment took about 50 min to complete.

Data acquisition and image preprocessing

Imaging was conducted using a 3 T Siemens magnetic resonance imaging system, acquiring a total of around 1520 volumes for each subject. Functional T2-weighted gradient-echo echo-planar images were obtained from 26 axial slices (voxel size 3 × 3 × 3 mm, no gap, TR 1720 ms, TE 30 ms, flip angle 80°, field of view 216 × 216 mm², matrix 72 × 72) aligned to the anterior and posterior commissure. Additionally, a high-resolution (1 × 1 × 1 mm voxel size) T1-weighted structural MRI (MPRAGE) was acquired for each subject. Head movement was restrained within the head coil by circumaural headphones fitting tightly into the head coil and by tight foam padding. Additionally, subjects were instructed to minimize head movement. The microphone was placed centrally to the mouth as closely as possible without touching the lips.

Processing and analysis of imaging data was performed with SPM5 (http://www.fil.ion.ucl.ac.uk/spm/). Preprocessing included slice timing (correction for differences in slice acquisition time), realignment and unwarping (motion correction), coregistration between the individual structural T1-weighted image and the EPI-images, and segmentation of the structural images. The resulting estimated spatial normalization parameters were then applied to the series of functional images, which were subsequently resampled to a voxel size of 3 × 3 × 3 mm and finally smoothed using a 6 mm full-width at half-maximum isotropic Gaussian kernel. Speech movement artifacts were controlled by using an unwarp mechanism in the preprocessing of the data, thereby correcting for geometric distortions due to magnetic field inhomogeneity (Andersson et al., 2001; Hutton et al., 2002). This preprocessing step removes variance of movement-by-susceptibility—distortion interaction and is therefore useful for correcting artifacts due to stimulus-correlated movements which are unavoidable in tasks with overt speech production.

Statistical analyses

Behavioral data

Response accuracy was rated for two different parameters. First, phonemic accuracy was examined across tasks to verify that the experimental tasks could be accomplished in the scanner environment, and that subjects were able to repeat the stimuli accurately despite the scanner noise. Secondly, and more importantly, manipulation accuracy was evaluated as an index of level of performance.

Thus, the first step in vocal response analysis was to determine the phonemic accuracy of responses in the REPEAT and MANIP tasks, using the sound editing software 'Audacity' (http://audacity.sourceforge.net/). Accuracy of verbal responses was evaluated independently by two raters (CP and a second rater who was not familiar with the goal of the study). Mean concordance between raters was 0.8 (Cohen's Kappa), which constitutes substantial interrater agreement by convention (Landis and Koch, 1977). Whenever the raters differed, the respective tokens were rated by a third person (instructed by CP). The inclusion criterion was set to a minimum of 80% correct responses overall, across tasks and phonological processes.

In a second step, the accuracy of the manipulation process per se was analyzed for the MANIP tasks. Specifically, the manipulation for PROS stimuli was considered accurate whenever word stress had been shifted from the first to the second syllable or vice versa, depending on the model stimulus. The manipulation for SEGm stimuli was considered accurate whenever a vowel replacement had occurred, independently of the accuracy of the vowel per se.

Next, duration (in ms) of verbal responses was measured in two ways. The first analysis included only responses not containing syllable iterations, self-corrections, or dysfluencies and, for the MANIP task, only accurate responses. This analysis was expected to provide an unbiased estimate of response duration differences between conditions. On- and off-set of verbal responses were determined by visual and auditory inspection of the speech wave. The second analysis addressed the duration of the complete process of repetition and manipulation, respectively, measured for each event from model stimulus onset to response offset, for all stimuli regardless of dysfluencies or manipulation accuracy. The latter analysis was used to model onset and duration of events with respect to the functional imaging data.

Furthermore, response latencies were computed by calculating, for each item, the time interval between offset of the green dot to vocal response onset. In a final step, the behavioral parameters (accuracy, response duration, response latency) were correlated with each other in order to test for trade-off effects between speed and accuracy or between response duration and response latency.

Functional imaging data

Statistical analyses of the functional imaging data were performed in two steps. In a first-level analysis, a statistical model was computed...
for each subject. To this end, the processes of repetition and manipulation, respectively, were defined as one single event lasting from stimulus onset to vocal response offset, and entered as one variable into the model (since the perception and production of the pseudowords were temporally correlated, they could not meaningfully be modeled separately as two regressors). This regressor, containing the whole process of repetition or reproduction after manipulation, was then convolved with a canonical hemodynamic response function (HRF) as implemented in SPM5. Since innsutrasubject stimulus-correlated motion was corrected by using the SPM5-function ‘unwarp’ for the preprocessing of the images, the realignment parameters were not inserted into the model. Voxel-wise regression coefficients for the variable of interest were estimated using the least-squares method within SPM5, and statistical parametric maps of the \( t \) statistic (SPM\( t \)) were generated. Next, we computed the simple main effects of each of the four conditions in our design (REPEAT_PROS, REPEAT_SEGM, MANIP_PROS, MANIP_SEGM).

In a second-level analysis, the four contrast images of the first-level analysis for each subject were used to perform a group analysis. To this end, the contrast images of the four conditions were entered into a within-subject ANOVA model, including a correction for non-sphericity. All of the following analyses were computed within this model.

To identify activation due to manipulation of pseudowords across phonological processes, the contrast images of the main effects of the MANIP conditions (i.e., MANIP_PROS, MANIP_SEGM) were submitted to a conjunction analysis under a conservative conjunction null hypothesis (Nichols et al., 2005). The same analysis was performed for the REPEAT conditions.

The next step aimed to identify areas sensitive to the specific demands of segmental versus prosodic manipulation. To this end, MANIP was contrasted with REPEAT separately for both phonological processes (i.e., MANIP_SEGM > REPEAT_SEGM and MANIP_PROS > REPEAT_PROS). The reverse contrasts (i.e., REPEAT > MANIP for segmental and prosodic processing, respectively) were computed analogously. To examine whether segmental and prosodic manipulations preferentially engage the left and/or the right hemisphere we computed a lateralization index (LI), respectively, by using the LI-Toolbox (Wilke and Liddle, 2007; Wilke and Schmithorst, 2006).

We applied the bootstrapping method using weighted means for the whole brain except a sagittal section of 5 mm to the left and right of the midline. This analysis results in values between 1 (completely left lateralized) and −1 (completely right lateralized).

Finally, we computed an interaction between PHONOLOGICAL PROCESS and the two tasks (reproduction with manipulation, repetition) in order to examine the impact of manipulation while controlling for differences in stimulus durations (and, subsequently, trial duration) between the segmental and the prosodic stimuli (see Table 1). With this analysis we intended to examine whether the two phonological processes differed with respect to amount or localization of the expected additional increase in activation in the MANIP relative to the REPEAT condition.

The analyses described above were computed for the entire volume of 26 slices. Since the location of area Spt varies considerably across individuals (Westbury et al., 1999; see also Hickok, 2009), we additionally established a region of interest (ROI) for area Spt whose extent was defined by the coordinates derived from nine previous studies which reported Spt activation during auditory–motor integration (Buchsbaum et al., 2005a,b, 2011; Callan et al., 2006; Hickok et al., 2003; Okada and Hickok, 2006; Okada et al., 2003; Pa and Hickok, 2008; Wilson and Iacoboni, 2006). To determine the center of this ROI, we first transformed the Spt coordinates wherever necessary into MNI space using the function provided by Matthew Brett at http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach, and then computed the average of the respective minimum and maximum \( x \)-, \( y \)- and \( z \)-coordinates given in the nine studies. These averages yielded the central ROI coordinate (i.e., \([−54 −39 15]\)). Around this central coordinate we then defined a sphere with a radius large enough (i.e., 9 mm) to cover all of the peak coordinates of the nine studies contributing to the ROI definition.

We wanted to ensure that the observed activations did not result from deactivation in the comparison condition. Therefore, all differential contrasts were inclusively masked by the corresponding main effect of the minuend, as stated in the figure legends (voxelwise \( p < 0.05 \)). The statistical threshold was set to \( p < 0.05 \) (FWE-corrected) for simple main effects and to \( p < 0.05 \) (FDR-corrected) for differential contrasts and the interaction analyses. Anatomical localization of activation peaks was done by using the SPM Anatomy Toolbox Version 1.6 (Eickhoff et al., 2005). Spatial references are given in Montreal Neurological Institute (MNI) coordinates.

**Results**

**Behavioral data**

**Phonemic accuracy across tasks and phonological processes**

All participants reached a satisfactory level of phonemic accuracy, with a group mean of 86.8% correct responses across tasks and phonological processes (range 80.6–93.1%, SD 3.7%, see Table 2). Thus, none of the subjects had to be excluded from further analysis. Apart from the main factor PHONOLOGICAL PROCESS (i.e., segmental or prosodic processing), TASK was included as a second within-subject factor in a repeated-measures MANOVA. This analysis revealed significant main effects of TASK (\( F(1,21) = 51.24; p < 0.001 \)) and of PHONOLOGICAL PROCESS (\( F(1,21) = 9.91; p < 0.01 \)), and a significant TASK×PROCESS interaction (\( F(1,21) = 10.48; p < 0.01 \)). Responses in the REPEAT task were more accurate (mean: 91.8% correct) than in the MANIP task (mean: 81.8% correct). Across tasks, the SEGm stimuli showed higher phonemic accuracy (mean 88.9% correct) than the PROS stimuli (mean: 84.7%). The interaction is explained by a larger difference in accuracy between MANIP and REPEAT tasks in the SEGm compared to the PROS stimuli.

**Manipulation accuracy across phonological processes**

This analysis exclusively considered the success of the manipulation, regardless of phonemic errors. Manipulation per se revealed a high mean accuracy of 96.2% across phonological processes (range of individual averages across tasks and phonological processes 80.6–100%, SD 5.3%). For the PROS and SEGm stimuli, the means were 95.9% (range 66.7–100%) and 96.5% (range 72.0–100%), respectively. A paired-samples t-test failed to reveal a significant difference in manipulation accuracy between the two phonological processes (\( t(21) = −0.30; p > 0.05 \)).

**Response latency**

Average response latency, computed from the offset of the green dot, was 149 ms (range 36–325 ms; SD 88 ms). A repeated-measures ANOVA with TASK and PHONOLOGICAL PROCESS as factors revealed no significant main effects of TASK (\( F(1,21) = 2.16; p > 0.05 \)) or of PHONOLOGICAL PROCESS (\( F(1, 21) = 2.58; p > 0.05 \)) and no TASK×PHONOLOGICAL PROCESS interaction (\( F(1, 21) = 0.64; p > 0.05 \)).

**Table 2**

<table>
<thead>
<tr>
<th>Task/phonological process</th>
<th>MANIP</th>
<th>REPEAT</th>
<th>Across tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEGm</td>
<td>81.8</td>
<td>95.9</td>
<td>88.9</td>
</tr>
<tr>
<td></td>
<td>(68.1–97.2)</td>
<td>(88.9–100)</td>
<td>(81.3–97.2)</td>
</tr>
<tr>
<td>PROS</td>
<td>81.7</td>
<td>87.8</td>
<td>84.7</td>
</tr>
<tr>
<td></td>
<td>(66.7–91.7)</td>
<td>(76.4–94.4)</td>
<td>(75.0–91.7)</td>
</tr>
<tr>
<td>Across processes</td>
<td>81.8</td>
<td>91.8</td>
<td>86.8</td>
</tr>
<tr>
<td></td>
<td>(70.8–93.1)</td>
<td>(86.8–95.8)</td>
<td>(80.6–93.1)</td>
</tr>
</tbody>
</table>
Response and trial duration

In the analysis of response durations we excluded responses which contained iterations, self-corrections, delays or manipulation errors (3% of all responses). The mean response duration, computed from vocal response onset to offset across both tasks and phonological processes, was 913 ms (range 743–1248 ms, SD 113 ms). A repeated-measures ANOVA with TASK and PHONOLOGICAL PROCESS as within-subject factors revealed a significant main effect of PHONOLOGICAL PROCESS ($F(1,21)=43.63; p<0.001$). A post-hoc analysis showed a significantly longer mean response duration for the SEGM stimuli (966 ms) than for the PROS stimuli (881 ms; paired-samples t-test: $t(21)=7.16; p<0.001$). There was no significant main effect of TASK ($F(1,21)=2.80; p>0.05$) or TASK×PHONOLOGICAL PROCESS interaction ($F(1,21)=0.27; p>0.05$).

Next, trial duration, computed from stimulus onset to response offset, was determined for all responses. Across stimuli, the analysis revealed a mean trial duration of 4010 ms (range 3570–4770 ms, SD 260 ms). A repeated-measures ANOVA with TASK and PHONOLOGICAL PROCESS as within-subject factors revealed a significant main effect of PHONOLOGICAL PROCESS ($F(1,21)=210.72; p<0.001$). A post-hoc analysis showed a significantly longer trial duration for the SEGM stimuli (4190 ms) than for the PROS stimuli (3830 ms; paired-samples t-test: $t(21)=6.84; p<0.001$). There was no significant main effect of TASK ($F(1,21)=0.53; p>0.05$) and no TASK×PROCESS interaction ($F(1,21)=0.02; p>0.05$).

Correlations between behavioral variables

None of the behavioral parameters (accuracy, response duration, response latency) correlated significantly with any other parameter (all $r$ between $-0.19$ and 0.35; $p>0.05$).

Functional imaging data

Areas involved in repetition and reproduction with manipulation

The imaging data of all participants ($n=22$) were valid and could be used for functional data analyses. First, we performed a conjunction analysis, examining conjoint BOLD signal changes across phonological processes for the simple main effects of MANIP and REPEAT, respectively. Computed across PROS and SEGM stimuli, repetition resulted in an extensive pattern of perisylvian activation with bilateral changes in the post- and precentral gyri, the thalami, the middle cingulate gyri, primary motor cortices, superior temporal gyri, and the cunei. There were also changes in right middle temporal cortex, and in the left middle frontal gyrus (Figs. 2a/d). At the very conservative level of a Bonferroni correction across the whole brain, we observed relatively little activation in inferior frontal gyrus (IFG). However, IFG activation was observed when the statistical correction threshold was lowered to false discovery rate (across the whole brain).

For reproduction with manipulation, the conjunction analysis revealed a similar activation pattern (Figs. 2b/e). However, in addition to the pattern observed for repetition, there were larger clusters of activation in IFG and left middle frontal gyrus (MFG) as well as in left intraparietal sulcus (IPS).

In left posterior superior temporal gyrus (pSTG), the conjunction analyses for repetition as well as reproduction with manipulation resulted in strong and circumscribed activation in area Spt. In order to examine the strength of Spt activation for all conditions (i.e., MANIP_PROS, MANIP_SEGM, REPEAT_PROS, REPEAT_SEGM) we first extracted the respective parameter estimates from each subject’s individual peak voxel within the Spt-ROI (constructed as described above in Materials and methods). We then averaged the estimates across subjects and plotted the BOLD signal changes in area Spt separately for each condition. Parameter estimates computed within this ROI show comparable changes of activation in all four conditions (see Fig. 2c).

Areas involved in phonological manipulation

To explore the areas involved in segmental and prosodic manipulation we contrasted the MANIP with the REPEAT task separately for both phonological processes. In the segmental condition, this contrast revealed a large connected cluster of activation covering most of left inferior frontal cortex, including Broca’s area and extending into BA44 and the middle frontal gyrus (peak MNI coordinate: $[-48 12 27]$), and a large cluster of activation in the left IPS (peak MNI coordinate: $[-42 -42 42]$; Table 3, Fig. 3a). Activation in IFG and the inferior parietal lobule was also found in the right hemisphere. Additionally, there was activation in the left middle frontal gyrus (peak MNI coordinate: $[48 -34 33]$, SD 113 ms). A repeated-measures ANOVA with TASK and PHONOLOGICAL PROCESS as within-subject factors revealed a significant main effect of PHONOLOGICAL PROCESS ($F(1,21)=210.72; p<0.001$). A post-hoc analysis showed a significantly longer trial duration for the SEGM stimuli (4190 ms) than for the PROS stimuli (3830 ms; paired-samples t-test: $t(21)=6.84; p<0.001$). There was no significant main effect of TASK ($F(1,21)=2.80; p>0.05$) or TASK×PHONOLOGICAL PROCESS interaction ($F(1,21)=0.27; p>0.05$).

Next, trial duration, computed from stimulus onset to response offset, was determined for all responses. Across stimuli, the analysis revealed a mean trial duration of 4010 ms (range 3570–4770 ms, SD 260 ms). A repeated-measures ANOVA with TASK and PHONOLOGICAL PROCESS as within-subject factors revealed a significant main effect of PHONOLOGICAL PROCESS ($F(1,21)=210.72; p<0.001$). A post-hoc analysis showed a significantly longer trial duration for the SEGM stimuli (4190 ms) than for the PROS stimuli (3830 ms; paired-samples t-test: $t(21)=6.84; p<0.001$). There was no significant main effect of TASK ($F(1,21)=0.53; p>0.05$) and no TASK×PROCESS interaction ($F(1,21)=0.02; p>0.05$).

Correlations between behavioral variables

None of the behavioral parameters (accuracy, response duration, response latency) correlated significantly with any other parameter (all $r$ between $-0.19$ and 0.35; $p>0.05$).

Functional imaging data

Areas involved in repetition and reproduction with manipulation

The imaging data of all participants ($n=22$) were valid and could be used for functional data analyses. First, we performed a conjunction analysis, examining conjoint BOLD signal changes across phonological processes for the simple main effects of MANIP and REPEAT, respectively. Computed across PROS and SEGM stimuli, repetition resulted in an extensive pattern of perisylvian activation with bilateral changes in the post- and precentral gyri, the thalami, the middle cingulate gyri, primary motor cortices, superior temporal gyri, and the cunei. There were also changes in right middle temporal cortex, and in the left middle frontal gyrus (Figs. 2a/d). At the very conservative level of a Bonferroni correction across the whole brain, we observed relatively little activation in inferior frontal gyrus (IFG). However, IFG activation was observed when the statistical correction threshold was lowered to false discovery rate (across the whole brain).

For reproduction with manipulation, the conjunction analysis revealed a similar activation pattern (Figs. 2b/e). However, in addition to the pattern observed for repetition, there were larger clusters of activation in IFG and left middle frontal gyrus (MFG) as well as in left intraparietal sulcus (IPS).

In left posterior superior temporal gyrus (pSTG), the conjunction analyses for repetition as well as reproduction with manipulation resulted in strong and circumscribed activation in area Spt. In order to examine the strength of Spt activation for all conditions (i.e., MANIP_PROS, MANIP_SEGM, REPEAT_PROS, REPEAT_SEGM) we first extracted the respective parameter estimates from each subject’s individual peak voxel within the Spt-ROI (constructed as described above in Materials and methods). We then averaged the estimates across subjects and plotted the BOLD signal changes in area Spt separately for each condition. Parameter estimates computed within this ROI show comparable changes of activation in all four conditions (see Fig. 2c).

Areas involved in phonological manipulation

To explore the areas involved in segmental and prosodic manipulation we contrasted the MANIP with the REPEAT task separately for both phonological processes. In the segmental condition, this contrast revealed a large connected cluster of activation covering most of left inferior frontal cortex, including Broca’s area and extending into BA44 and the middle frontal gyrus (peak MNI coordinate: $[-48 12 27]$), and a large cluster of activation in the left IPS (peak MNI coordinate: $[-42 -42 42]$; Table 3, Fig. 3a). Activation in IFG and the inferior parietal lobule was also found in the right hemisphere. Additionally, there was activation in the left middle
Table 3

<table>
<thead>
<tr>
<th>Region (peak activation)</th>
<th>Side</th>
<th>Voxels in cluster</th>
<th>MNI coordinates</th>
<th>T-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MANIP-REPEAT for SEGM(b)</td>
<td>L</td>
<td>1217</td>
<td>−42 −42 42</td>
<td>9.12</td>
</tr>
<tr>
<td>inferior parietal lobe (IPS)</td>
<td>L</td>
<td>1698</td>
<td>−48 12 27</td>
<td>9.11</td>
</tr>
<tr>
<td>inferior frontal gyrus (BA44/BA45)</td>
<td>L</td>
<td>2076</td>
<td>−3 21 42</td>
<td>8.63</td>
</tr>
<tr>
<td>medial superior frontal gyrus (BA32)</td>
<td>L</td>
<td>24</td>
<td>−45 −54 −6</td>
<td>5.13</td>
</tr>
<tr>
<td>superior temporal gyrus</td>
<td>L</td>
<td>13</td>
<td>27 3 45 4.44</td>
<td></td>
</tr>
<tr>
<td>supramarginal gyrus</td>
<td>R</td>
<td>21</td>
<td>60 −15 24</td>
<td>3.75</td>
</tr>
<tr>
<td>Calcarine gyrus</td>
<td>R</td>
<td>85</td>
<td>21 −57 6</td>
<td>3.57</td>
</tr>
<tr>
<td>Lingual gyrus</td>
<td>L</td>
<td>84</td>
<td>−15 −48 −3</td>
<td>3.49</td>
</tr>
<tr>
<td>Superior temporal gyrus</td>
<td>R</td>
<td>15</td>
<td>48 −45 15</td>
<td>3.48</td>
</tr>
<tr>
<td>Insular lobe</td>
<td>L</td>
<td>23</td>
<td>−42 −3 0</td>
<td>3.40</td>
</tr>
<tr>
<td>Middle temporal gyrus</td>
<td>L</td>
<td>23</td>
<td>−48 −42 3</td>
<td>3.26</td>
</tr>
<tr>
<td>Cuneus</td>
<td>R</td>
<td>17</td>
<td>0 −75 21</td>
<td>3.05</td>
</tr>
<tr>
<td>Calcarine gyrus</td>
<td>R</td>
<td>22</td>
<td>12 −72 15</td>
<td>2.83</td>
</tr>
</tbody>
</table>

Note: IPS = intraparietal sulcus.
\(a\) Cluster level extent > 10; inclusively masked with main effect of MANIP_SEGM.
\(b\) Cluster level extent > 10; inclusively masked with main effect of MANIP_PROS.

and inferior temporal gyri, the superior medial frontal gyrus and the caudate nucleus. The weighted mean LI (lateralization index) was 0.69, showing a moderate left lateralization of activation for segmental processing.

In the prosodic condition, the contrast between the manipulation and repetition task revealed considerably less activation compared to the segmental condition, but a comparable activation pattern. That is, activation was predominantly found in the left IFG (peak MNI coordinates: \([-45 12 27]\)) and left IPS (peak MNI coordinates: \([-39 −42 39]\)). Similarly to the segmental condition, the activation was moderately left lateralized (weighted mean LI = 0.65).

Because we had expected reproduction after manipulation (compared to repetition) to lead to an increase of activation in area Spt due to the additional phonological manipulation required, we performed an ancillary analysis exploring the effects of manipulation within the Spt-ROI. This analysis was performed in two ways, by (1) examining the two phonological processes separately (i.e., MANIP-REPEAT for segmental and for prosodic processing, respectively) and by (2) combining the two phonological processes in a single analysis (i.e., MANIP-REPEAT for segmental and prosodic processing). Neither analysis resulted in any significantly activated voxels, even when performed without masking and when the statistical threshold was lowered to an uncorrected level of p < 0.001. Because an analysis at the group level may have obliterated significant findings for individual subjects due to individual variability in the location of area Spt, we also examined activation within the Spt-ROI for each of the 22 subjects individually (using a threshold of p < 0.001 uncorrected for the Spt-ROI). This analysis revealed that 7 out of 22 subjects showed activation within the Spt-ROI. Of those 7 subjects, 3 showed clusters of between 1 and 5 voxels, and 4 showed clusters of 8, 9, 18 and 32 voxels, respectively, when manipulation was contrasted with verbatim repetition across phonological processes.

For SEGM as well as PROS, the reverse contrast (REPEAT > MANIP) showed activation changes predominantly in the angular gyrus bilaterally, the posterior cingulate cortex, and the medial prefrontal cortex.

Finally, we computed the interaction of PHONOLOGICAL PROCESS with the two tasks (repetition, reproduction after manipulation) in order to examine the differential effects of an additional task (reproduction...
with manipulation versus verbatim repetition) independently of stimulus-inherent differences between phonological processes such as stimulus duration (and, subsequently, response duration). This analysis revealed a stronger effect of reproduction with manipulation (relative to repetition) for SEG compared to PROS particularly in the left postcentral gyrus extending along the postcentral sulcus, in medial superior frontal gyri, and in the inferior frontal gyrus bilaterally as well as in the left insula and the right caudate nucleus (Fig. 4 and Table 4).

The reverse interaction (i.e., the test for a stronger effect of prosodic than segmental processing for manipulation relative to repetition) did not yield any significant results at this threshold.

**Discussion**

The present fMRI study aimed to examine whether areas associated with the dorsal stream are involved in perception-production tasks requiring relatively few verbal working memory resources. Additionally, the present study sought to examine the relative contributions of the left and right hemispheres, depending on the segmental versus prosodic nature of a phonological manipulation.

Subjects performed two different phonological tasks which required either a segmental or a prosodic manipulation of heard pseudowords. A verbatim repetition task served as control task. Based on previous findings, both reproduction with manipulation and repetition were expected to activate a temporoparietal network. Additional activation in area Spt, defined as a region of interest in the posterior Sylvian fissure at the temporoparietal boundary, was expected specifically for the manipulation task, which – unlike verbatim repetition – required segmental or prosodic manipulation of the perceptual input. We expected left-lateralized activation for segmental, and bilateral or right-lateralized activation for prosodic processing.

Behavioral analyses demonstrated that subjects successfully performed the tasks. All participants reached a satisfactory level of phonemic accuracy. Manipulation accuracy was high and did not differ between the segmental and the prosodic task.

**Cortical areas involved in auditory-motor integration**

Both repetition and reproduction with manipulation evoked cortical activation predominantly in the postcentral gyri bilaterally, the STG bilaterally, and the left MFG. Furthermore, left IFG was strongly activated in the manipulation task and less so in repetition. This perisylvian network resembles the activation pattern found in a shadowing paradigm which examined the relatively automated, immediate repetition of pseudowords (Peschke et al., 2009). Bilateral STG activation corroborates existing evidence regarding the involvement of these regions in speech perception. The strong activation of the postcentral gyri was somewhat surprising, but replicates the activation pattern found in our previous study (Peschke et al., 2009) as well as in prior reports of sensorimotor involvement during speech production (e.g., Dahanal et al., 2008). In tasks requiring overt articulation, the postcentral gyrus appears to represent a significant component of the speech motor network (Dogil et al., 2002; Lotze et al., 2000; Riecker et al., 2005); this area may also be involved in tasks requiring articulation without phonation (Pulvermuller et al., 2006). Somatosensory cortex has been proposed to play a crucial role in speech motor control as part of a somatosensory feedback system (Guenther et al., 2006). In our study, subjects may have relied on the tactile-sensory feedback control system in order to compensate for an impoverished auditory feedback due to ambient scanner noise and tight-fitting headphones.

Activation of left MFG in the present study was unexpected and had not been detected in our former shadowing task (Peschke et al., 2009). One explanation is that despite the use of pseudowords our paradigms may have triggered word retrieval mechanisms which are known to involve left MFG (review by Price, 2010, and references therein). As an alternative explanation, MFG has been argued to be part of an inhibitory neurocognitive network (Liddle et al., 2001; Pedersen et al., 1998; Rubia et al., 2001). In the present study, MFG activation may have resulted from the fact that subjects had to hold off their response until cued by the offset of a visual stimulus.

Activation in the left IFG likely reflects speech production processes. Remarkably, this region was more active in the manipulation condition than in repetition, suggesting that the phonological manipulation examined here recruited this area in a specific way. This will be discussed in greater detail in the next section.

**Areas involved in phonological manipulation**

Our major objective of the present study was to determine the neuronal correlates of naturally occurring phonological manipulations. Our tasks required relatively few verbal working memory...
resources, since only two to three syllables had to be processed and responses occurred at relatively short delays. Contrary to our predictions, reproduction with manipulation compared to verbatim repetition did not result in predominant activation of area Spt. Instead, phonological manipulation revealed large clusters of activation in the left dorsal IFG and in IPS.

Despite our efforts to keep working memory demands low our manipulation tasks activated regions which have previously been associated with verbal working memory functions. Specifically, the inferior parietal lobule has been argued to be associated with the phonological store, whereas the IFG has been argued to be involved in articulatory rehearsal (Awh et al., 1996; Paulesu et al., 1993; Smith et al., 1996; Strand et al., 2008). Additionally, both regions have been reported to be sensitive to verbal working memory load, showing increasingly more activation as the number of items to be rehearsed increases (Martin et al., 2003; Ravizza et al., 2004; Rypma et al., 1999). On this background our results may suggest that the task of manipulating a heard verbal stimulus prior to reproduction required working memory resources to a significant extent, even though stimulus length and response delays did not place high demands on temporal storage capacities. Conceivably, there is a higher demand on verbal working memory when part of an auditory stimulus must be modified before reproduction than when the stimulus is simply repeated. In verbatim repetition, the incoming auditory stimulus may be transferred to the speech programming and articulation network without any further modifications and with only minor verbal working memory expenses. In the manipulation task, however, the auditory information has to be maintained in a more abstract form until the manipulation process has been completed, thereby delaying the start of speech motor programming.

However, the parieto-frontal activation found in the present study may not solely be explainable by verbal working memory processes, since the phonological manipulation required more than the temporary storage of a stimulus. Rather, this task called for an explicit phonological manipulation of the perceived pseudowords. We propose that left IPS as well as left IFG is involved in these manipulation processes, and that activation in these areas is independent of whether a segmental or a prosodic manipulation is performed. Comparable to our finding, sequential manipulation of syllabic units compared to a simple matching task resulted in activation of Broca’s area and the left supramarginal gyrus as well as the superior parietal lobule around the IPS (Gelfand and Bookheimer, 2003). Gelfand and Bookheimer’s (2003) study may be argued to involve a high degree of verbal working memory compared to the task used here, since a string of three syllables had to be maintained until a second syllable sequence appeared. However, involvement of frontal and parietal areas has also been reported in phonological tasks with relatively low verbal working memory demands, e.g., during phoneme monitoring (IFG and inferior parietal lobe; Newman and Twieg, 2001) or in phoneme detection (IFG; Heim et al., 2008). Moreover, the pseudodiminutives in this condition contain a real German definite article (e.g., “der”), thereby constituting a syntactic phrase and indicating that the pseudoword-items, in a grammatical sense, are nouns. Additionally, the pseudodiminutives in this condition contain a real German suffix (e.g., “chen”), perhaps allowing unwanted associative imagery of larger versus smaller objects to infiltrate segmental processing. Hence, the definite article as well as the suffixes may have added implicit ‘wordiness’ to the pseudowords in the segmental condition, thereby involving structures of a ventral lexical-semantic processing stream.
The direct comparison of repetition with manipulation, computed separately for segmental and prosodic processing, showed higher BOLD signal changes in the angular gyrus bilaterally, the posterior cingulate cortex and the medial prefrontal cortex. These signal changes resulted from a higher deactivation in these areas during manipulation than during repetition. This network of areas fits a formation of activated areas previously described as a ‘default mode network’ (Raichle et al., 2001), which has been proposed to be active during resting state or during tasks with fewer cognitive requirements, and to be deactivated or suppressed in tasks with higher cognitive demands (Damoiseaux et al., 2006; Greicius and Menon, 2004; McKiernan et al., 2003; Persson et al., 2007; Singh and Fawcett, 2008). In the present study, manipulation likely required more cognitive capacity than repetition and therefore may have led to a higher deactivation of the default mode network than repetition.

**Lateralization of segmental versus prosodic processing**

In our secondary research question we asked whether specific phonological processes preferentially engage the left and/or the right hemisphere, depending on whether segmental or prosodic operations are involved. We hypothesized that segmental processing is lateralized to the left hemisphere, whereas prosodic processing may be mediated bilaterally or preferentially by the right hemisphere.

Both phonological processes evoked a predominantly left lateralized activation pattern. A left lateralization for segmental processing (as required when altering the quality of a vowel at a subsyllabic level) is perfectly consistent with previous studies examining segmental phonological processing (e.g., Burton and Small, 2006; Heim et al., 2003). Our expectation of a bilateral or right-lateralized distribution of activation for prosodic processing, however, was not confirmed, at least for the given tasks which required a stress shift across a syllable boundary. Instead, we observed activation increases predominantly in left parietal and inferior frontal regions. The relative involvement of the left hemisphere in processing the prosodic stimuli in the present study might be explained by the specific requirements of our prosodic manipulation task: subjects had to identify and modify the temporal characteristics of the (stressed and unstressed) vowels of the stimuli in order to achieve a shift of stress across the syllable boundary. In addition, reproducing a stimulus after implementing a stress shift may be argued to impose a higher demand on auditory–motor integration, thereby leading to an increased involvement of left hemispheric areas. Cortical involvement has been reported to shift from a right to a more left-lateralized pattern when subjects are asked to evaluate prosodic features such as pitch range or vowel duration while repeating the stimuli using inner speech, as opposed to a prosodic evaluation without any explicit articulatory demands (Pihan, 2006).

We further examined specific effects of segmental versus prosodic processing by computing a phonologic process by task interaction. By way of this interaction we controlled for the longer trial durations in the segmental compared to the prosodic condition – thus, differences in stimulus length should not contribute to the observed activation. Nevertheless, segmental versus prosodic processing resulted in a large and highly significant cluster of activation in the dorsolateral IFG predominantly in the left hemisphere, as well as in activation in the left postcentral gyrus (extending to the postcentral sulcus). Among other functions, inferior frontal cortex has been associated with articulatory processes (Bonilha et al., 2006; Hillis et al., 2004). Furthermore, increased dorsolateral IFG activation during segmental processing may be related to morphosyntactic computations. This is because in German, when transforming a noun into its respective diminutive, the article has to be adapted and the appropriate suffix has to be appended to form a grammatical (diminutive) noun phrase. Such morphosyntactic processes have been associated with Broca’s area (Embick et al., 2000; Indefrey et al., 2001). In particular, several functional imaging studies have shown that grammatical gender processing activates part of the IFG (Heim et al., 2002; Hernandez et al., 2004; Padovani et al., 2005).

Thus, the segmental manipulation task may have induced more complex linguistic operations. Nearly all subjects took more time to become acquainted with the segmental compared to the prosodic manipulation during the practice trials preceding the experiment. Probably, different linguistic units serve as the basis for performing the two types of manipulation. A stress shift, as required for the prosodic manipulation, is based on the whole prosodic word (Mary and Yegnanarayana, 2008) and leaves the segmental make-up of each syllable invariant, whereas a vowel replacement, as required for the segmental manipulation, affects the nucleus of a syllable and requires a re-programming of syllabic units. Since syllables are considered the basic units of speech and are represented as pre-compiled chunks including all relevant phonetic features (Levet and Wheeldon, 1994), the re-organization of such units in the prosodic manipulation task may have occurred at more superficial processing levels than the re-organization of syllabic plans. This suggestion is substantiated by an earlier observation that processing of phonemes led to higher BOLD signal changes in the left IFG than the processing of syllables (Siol et al., 2003), and may explain our finding of more inferior frontal activation for the segmental as compared to the prosodic manipulation.

As in our previous study (Peschke et al., 2009), the present tasks revealed strong involvement of the postcentral gyri. Involvement of the left postcentral gyrus (and sulcus) was observed to be even stronger when subjects had to replace a vowel during reproduction, compared to when they had to shift the word stress from the first to the second syllable (and vice versa). We previously suggested that postcentral activation may indicate increased reliance on a tactile-sensory feedback control system in order to compensate for the impoverished auditory feedback in the noisy scanner environment (Peschke et al., 2009). Reliance on tactile-sensory feedback may be especially important in the segmental manipulation condition because segmental auditory information is probably more susceptible to scanner noise masking than prosodic information.

Finally, testing for stronger effects of prosodic versus segmental manipulation (while controlling for potential differences between the two processes during repetition) revealed no specific effects of prosodic (compared to segmental) processing. As discussed in the preceding section, this may be due to the fact that during prosodic manipulation, reprogramming could proceed at the syllabic level, without the need for further sequencing operations at the sub-syllabic level. This suggests that prosodic manipulation, at least as required by the present task of stress shifts across syllable boundaries, is mediated by inferior parietal and inferior frontal ( pars triangularis) areas, without necessary recourse to right hemispheric regions.

**Summary**

The present study was designed to examine the areas involved in phonological manipulations during auditory–motor integration. We demonstrated that the left IPS and the left dorsal IFG are involved in the manipulation of phonological representations independent of the linguistic level (i.e. segmental versus prosodic) at which the manipulation is performed. We propose that these parieto-frontal regions are co-recruited when the task requires explicit (phonological) manipulation, in addition to the more automated transfer of auditory into articulatory verbal codes which appears to involve area 5pt. As predicted, segmental manipulation resulted in predominantly left hemispheric activation, involving left parietal and large aspects of left inferior frontal cortex. Prosodic manipulation did not evoke bilateral or right lateralized activation as expected, but a clearly left lateralized pattern, perhaps due to a task-dependent focus on temporal stimulus characteristics. Segmental compared to prosodic operations resulted in a left lateralized activation pattern probably resulting from the fact that...
segmental processes are performed at the sublexical phoneme level, which may require a higher number of sequencing operations than prosodic processing. Such sequencing operations have been primarily associated with inferior frontal areas. Contrary to our predictions, a direct comparison of prosodic with segmental manipulation did not result in increased activation, which may be due to the specific requirements in the prosodic condition, namely a lower degree of morphosyntactic operations and processing at the syllabic level.

Acknowledgments

This work was supported by a grant from the German Federal Ministry of Education and Research (BMBF-01GW0572) to WZ and AB and was carried out as part of the collaborative BMBF research project “From dynamic sensorimotor interaction to conceptual representation: Deconstructing apraxia”. We thank two anonymous reviewers for their constructive comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.neuroimage.2011.07.025.

References

Buchbauer, B., Odin, O., Hutton, C., Bork, A., K                                                            


