The neural correlates of verbal feedback processing: An fMRI study employing overt speech

Christoffels, I.K.; Formisano, E.; Schiller, N.O.

◆ Human Brain Mapping 28:868–879 (2007) ◆

Neural Correlates of Verbal Feedback Processing: An fMRI Study Employing Overt Speech

Ingrid K. Christoffels,[1]* Elia Formisano,[1] and Niels O. Schiller[1,2]

[1]*Department of Cognitive Neuroscience, Faculty of Psychology, Maastricht University, The Netherlands*
[2]*Leiden Institute for Brain and Cognition, Leiden University, The Netherlands*

Abstract: Speakers use external auditory feedback to monitor their own speech. Feedback distortion has been found to increase activity in the superior temporal areas. Using fMRI, the present study investigates the neural correlates of processing verbal feedback without distortion. In a blocked design, the following conditions were presented: (1) overt picture-naming, (2) overt picture-naming while pink noise was presented to mask external feedback, (3) covert picture-naming, (4) listening to the picture names (previously recorded from participants' own voices), and (5) listening to pink noise. The results show that auditory feedback processing involves a network of different areas related to general performance monitoring and speech-motor control. These include the cingulate cortex and the bilateral insula, supplementary motor area, bilateral motor areas, cerebellum, thalamus and basal ganglia. Our findings suggest that the anterior cingulate cortex, which is often implicated in error-processing and conflict-monitoring, is also engaged in ongoing speech monitoring. Furthermore, in the superior temporal gyrus, we found a reduced response to speaking under normal feedback conditions. This finding is interpreted in the framework of a forward model according to which, during speech production, the sensory consequence of the speech-motor act is predicted to attenuate the sensitivity of the auditory cortex. *Hum Brain Mapp 28:868–879, 2007.* © 2007 Wiley-Liss, Inc.

Key words: speech production; auditory feedback; self-monitoring; performance monitoring; insula; cingulate cortex; forward model; speech-motor control

INTRODUCTION

Most speakers produce numerous words per second seemingly without effort or conscious control of the speaking process. Nevertheless, we constantly monitor our own speech output on aspects such as content, grammaticality, fluency and volume. Verbal auditory feedback is very important in speech production to assure correct and proper speech output [Levelt et al., 1999]. When we are listening to music over headphones such that verbal auditory feedback is attenuated, we may accidentally speak inappropriately loud. Experimentally, the importance of verbal feedback in speaking is revealed, for example, by the profound effects of delaying auditory feedback [Lee, 1950]. When our speech is played back at a delay of a few hundred milliseconds, the fluency of speech becomes severely disrupted (see Yates [1963], for a review).

This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/1065-9471/suppmat.

Contract grant sponsor: Netherlands Organization for Scientific Research (NWO); Contract grant numbers: 453-02-006, 452-04-337.

*Correspondence to: Ingrid Christoffels, Cognitive Psychology Unit, Department of Psychology, Faculty of Social Sciences, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands.
E-mail: ichristoffels@fsw.leidenuniv.nl

Received for publication 23 November 2005; Revised 17 May 2006; Accepted 4 June 2006

DOI: 10.1002/hbm.20315

Published online 31 January 2007 in Wiley InterScience (www.interscience.wiley.com).

In word production, the processing components that are generally implicated in mapping meaning to sound include conceptual preparation, lexical and syntactic encoding, phonological encoding and articulation [e.g., Levelt, 1999]. According to one of the most influential models of speech production [Levelt, 1989; Levelt et al., 1999], speakers self-perceive their internally and overtly produced speech (internal and external monitoring). The model postulates that self-monitoring is a centrally controlled process with limited capacity, which evaluates the quality of the speech by means of the speech comprehension process. The speech comprehension system, used for understanding speech of others, also subserves verbal self-monitoring. The abstract phonological code is presumably used for internal monitoring. The acoustic speech signal of one's own voice serves as input for the external monitoring, the focus of the present study. This has the advantage that no separate speech perception component has to be postulated. Listening to one's own voice appears to rely indeed on the same areas of the temporal cortex as listening to someone else's voice [McGuire et al., 1996; Price et al., 1996; Wise et al., 1999], suggesting an important role for speech comprehension in the processing of verbal feedback. Nevertheless, the neural correlates of normal ongoing verbal self-monitoring are not well defined. The purpose of the current study is to shed more light on a crucial aspect of self-monitoring in speech: the neural basis of verbal feedback processing.

Note that self-monitoring is an intrinsic part of every verbal task we carry out and it always plays a role in studies concerning aspects of speech production [Aleman et al., 2005; Schiller et al., 2006; Wheeldon and Levelt, 1995]. Every time we overtly produce a word we present ourselves with auditory input as well, which means that we not only engage speech production processes but also auditory speech comprehension. Although these aspects are often difficult to disentangle, for external feedback they are addressed in the present study.

In their meta-analysis on the neural correlates of picture naming, Indefrey and Levelt [2004] suggested that regions were involved in the external loop of self-monitoring if they were reliably found to be activated in word listening tasks and were more strongly activated in experiments involving overt responses. This was the case for the bilateral superior temporal gyri (STG). More direct evidence on the areas involved in processing verbal feedback comes from studies, which manipulated auditory feedback. Here, authors assumed that the modulation of verbal feedback engages the self-monitoring process more strongly. For instance, McGuire et al. [1996] conducted a positron emission tomography (PET) study, in which participants read aloud single words. Increased activity was reported in the right STG, and a weaker similar left-sided activity when feedback was modulated by pitch distortion or by playing the voice of the experimenter rather than the participant's own voice. In contrast, no STG activity was found in a PET study during articulation of common sentences [Hirano et al., 1996]. However, in a later study, feedback was manipulated by filtering and delaying the verbal output, which resulted in more bilateral activation of the STG [Hirano et al., 1997]. More recently, using fMRI, Hashimoto and Sakai [2003] reported more activation in the temporo–parietal regions for delayed auditory feedback conditions than for normal feedback.

On the basis of these studies, it seems that regions in the STG are involved in processing verbal feedback and therefore these regions support the self-monitoring component of speech production. It is, however, important to consider that in these studies more activity is found for modulated feedback, i.e. when the auditory signal was distorted, replaced, or delayed, compared to normal feedback. Since this "abnormal" feedback might induce compensatory processing, it is not entirely clear whether these findings are informative to normal feedback processing. Furthermore, in a number of the above-mentioned studies, participants may have heard their own voice conducted through the skull bone together with the manipulated feedback, which may have rendered the actual auditory input very complex. Since previous studies assessed the consequence of feedback distortion rather than normal ongoing processing, the role of the STG in feedback processing remains unclear.

There is also some discrepancy with a different body of research in which auditory feedback processing has been associated with suppression of activity in the temporal cortices. A number of MEG and ERP studies showed reduced and delayed N100 responses in the auditory regions of the temporal cortices during speech production, in comparison to listening to played-back speech or altered feedback [Curio et al., 2000; Heinks-Maldonado et al., 2005; Houde et al., 2002; Numminen and Curio, 1999]. In the monkey auditory cortex, self-initiated vocalization resulted in suppression of neural discharges in single neurons [Eliades and Wang, 2003].

An important function of the ongoing monitoring process is to detect, intercept, and correct speech production errors [e.g., Hartsuiker and Kolk, 2001; Postma, 2000; Schiller, 2005; Schiller and De Ruiter, 2004]. In studies involving EEG, the error-related negativity, known from action monitoring [e.g., Holroyd and Coles, 2002], has been shown to be generated in verbal errors as well [Masaki et al., 2001; Ganushchak and Schiller, in press]. Therefore, it is likely that verbal self-monitoring is, at least partly, not a language-specific process, but relies on general performance monitoring instead. The idea that general executive functioning is important for verbal monitoring is in accordance with the assumption in Levelt's model that verbal self-monitoring relies on controlled processing [Levelt, 1989]. In action monitoring studies, anterior cingulate cortex (ACC) activity has been interpreted as reflecting error processing, monitoring of conflicting responses, or performance monitoring [e.g., Botvinick et al., 2001; Carter et al. 1998; Ridderinkhof et al., 2004a] and more general as activity monitoring [Schneider and Chein, 2003] or effortful processing [Mulert et al., 2005]. Although its exact function

is still a matter of debate, the cingulate cortex has previously been implicated in (overt) word production [e.g., Barch et al., 2000; Kan and Thompson-Schill, 2004; Shuster and Lemieux, 2005] and has been implicated in identification of self-generated speech [Allen et al., 2005]. Although the present study does not address speech production error detection and correction, even single word production, without errors or conflict, might engage performance monitoring in speech production. Therefore, we expected more activity in the ACC when feedback is present under normal circumstances than when it is not.

The aim of the present study was to investigate the neural correlates of verbal feedback in speaking. We tried to tease apart as much as possible the functional neuroanatomy of speaking, listening and auditory verbal feedback by using five experimental conditions. We used a standard picture-naming task in which the participants were asked to name the presented pictures. In the two crucial conditions, participants named the pictures aloud, but in one condition, perception of their own voice was masked by pink noise. In contrast to previous studies, auditory feedback was prevented rather than changed. Pink noise was presented at a high volume to effectively mask external feedback and speech conducted through the skull bone. As a consequence of the noise, the auditory input was not completely comparable between these conditions in terms of complexity and volume. It has been demonstrated that some areas in the temporal cortex show preference for speech-like stimuli [e.g., Belin et al., 2000; Scott et al., 2000; Vouloumanos et al., 2001]. Therefore, we presented two control conditions in which the participants listened to either their own prerecorded voice or to the pink noise. Finally, we included a covert picture-naming condition, in which no articulation takes place, to assess the comparability of covert and overt speech. A common assumption behind using covert speech is that it involves all the processes and their neural correlates of overt speech with the exception of the final motor execution. Differences between overt and covert speech have been reported, however, that may relate to linguistic or cognitive control [Munhall, 2001], and the use of covert speech as substitute for overt-naming has been questioned [e.g., Barch et al., 1999; Gracco et al., 2005; Huang et al., 2001].

The picture-naming task used in this study involves all the core components of word production and has often been used in behavioral and imaging studies [e.g., Indefrey and Levelt, 2004; Price et al., 2005; Van Turennout et al., 2003]. However, due to the problems concerning motion and other artifacts associated with speaking in the scanner, only recently have fMRI studies used overt speech in fMRI [e.g., Barch et al., 2000; De Zubicaray et al., 2001; Kan and Thompson-Schill, 2004; Palmer et al., 2001; Shuster and Lemieux, 2005]. In this study, we used overt speech in combination with a clustered acquisition protocol [e.g., De Zubicaray, 2001; Jäncke et al., 2002; Van Atteveldt et al., 2004] for stimulus presentation and speech production to take place in the silent interval between scans. By avoiding scanning during speaking, a considerable reduction in motion-related artifacts is achieved [Birn et al., 2004; Gracco et al., 2005]. An important advantage was that the auditory input (real time feedback or prerecorded voices and noise) was not interfered with EPI noise and that verbal responses could be recorded.

In summary, we aim to assess the neural correlates of auditory feedback processing by masking auditory feedback in an overt-naming paradigm.

MATERIALS AND METHODS

Participants

Fourteen healthy volunteers (5 male, 9 female, mean age 24) without any history of neurological or psychiatric disease participated in this study. They were right-handed according to the Edinburgh Handedness Inventory [Oldfield, 1971]. All participants were undergraduate or graduate students at Maastricht University, native speakers of Dutch and had no history of hearing- or language-related problems. All participants gave their written informed consent. The study was approved by the Ethical Committee of the University Medical Center of Nijmegen, The Netherlands.

Experimental Procedure and Material

In a blocked design, we presented five different task conditions, which alternated with a fixation condition. Of main interest were two overt picture-naming conditions: naming without pink noise when participants could hear their voices (PNvoice) and naming with the presentation of loud masking pink noise during the response (PNnoise) to mask the participants' own voice during picture-naming. Furthermore, a covert-naming condition (PNcovert) was presented in which participants internally generated a picture name. Finally, two control conditions were presented in which the auditory input was similar to the picture-naming conditions: both the participants' voice (LISvoice) and the pink noise (LISnoise) were presented, but participants were not required to give a naming response.

Each experimental block consisted of five trials of one condition and lasted 20 s. During the fixation blocks, a symbolic instruction was presented visually to instruct the condition of the upcoming experimental block. The duration of a fixation block was 16 s; twelve seconds after the onset of the presentation of a fixation cross, the instruction picture was presented for 2 s, followed by the fixation cross (2 s). The conditions were presented in five functional runs. Each run consisted of 15 experimental task blocks that alternated with fixation blocks. The 15 experimental blocks consisted of three repetitions of each condition (i.e. totally 15 repetitions of each condition per participant). The block order was pseudo-random, different for each run and the same for each participant. Run order was counterbalanced across participants.

In all conditions, pictures were presented for 1,000 ms, with an onset of 100 ms after the beginning of the silent delay between volume acquisitions. There were 2,250 ms of silence before the next volume was acquired in which the participants responded (Fig. 1). A fixation cross was presented immediately before picture onset and during fixation conditions. Twenty-five pictures were selected from the Max Planck Institute for Psycholinguistics database consisting of simple white-on-black line drawings (Fig. 1B and Fig. S1 in the supplementary material). The same pictures served equally often in each condition across runs. Picture names corresponded to mono- and bisyllabic words consisting of 3.7 phonemes on average. They were of relatively high word frequency (mean = 231 occurrences per one million words, CELEX database [Baayen et al., 1995]) and had high name agreement (mean = 99%; the percentage in which a given picture solicited the same name across participants, pre-tested in a pilot study).

The masking noise sound presented in PNnoise and LISnoise consisted of 1.5 s of digital mono recording of pink noise (1993, Sound Check Productions, A. Parson and S. Court). Presentation of noise started 400 ms after the picture onset. Therefore, the noise masked responses occurring between 400 and 1900 ms after picture onset. Typically, picture-naming takes 600 ms [e.g., Indefrey and Levelt, 2004].

In the two control conditions (LISvoice, LISnoise), a scrambled picture was presented on screen and the auditory stimulus was presented using the same timing as the PNnoise condition. For the LISnoise condition, the same pink noise recording was used as in the PNnoise condition.

The scrambled pictures were derived from the intact picture stimuli, which were spatially distorted by applying a distorting wave filter (Adobe Photoshop software 7.0), which transforms nonlinearly horizontal and vertical lines into sinusoidal patterns. See Figure 1B for examples. Distorted pictures were presented to prevent participants from automatically generating the name of the pictures; otherwise, processing in the listening conditions might become very similar to the speech production conditions. In the two listening conditions, the participants were instructed not to generate any names but to just passively listen to the auditory stimuli.

For the LISvoice condition unique material for each participant was recorded. During a picture-naming session that took place separately and prior to the scanning session, a digital recording was made in a soundproof booth in which participants named the pictures several times. For each word, samples were selected in which pronunciation was clear, which contained no speech errors or which obviously deviated from the other samples. Clicks were removed if necessary. This resulted in 25 audio stimuli for each subject (44.1 kHz, 16 bits, mono), with an average duration of 490 ms.

In the picture-naming conditions, participants were required to name pictures as quickly and accurately as possible. Participants were instructed not to be concerned with audibility to minimize speech-related movement; they were told not to over-articulate or speak loudly. Furthermore, they were made aware of the automatic tendency to increase the loudness of their voice in the presence of noise (Lombard effect [Lane and Tranel, 1971]) and instructed to speak at the same volume throughout the experiment. Finally, in the covert picture-naming condition (PNcovert), the participants were asked to internally generate the picture name without producing it aloud or moving their lips.

Because of individual differences in voice and sensitivity to sound, the volume for presentation of the auditory stimuli was determined separately for each participant before the actual scanning started. After being placed in the scanner, participants received trials of the PNnoise condition. The volume of the noise was gradually increased until participants reported they could no longer hear themselves (between ~92 and 108 dB SPL, 105 dB SPL on average). After a number of trials on the set level they were asked again whether they could hear themselves to check their judgment. Next, the volume of the picture names was tested by presenting the LISvoice condition. Participants were asked whether or not the volume was similar to hearing their own voice. If necessary, the attenuation of these stimuli was increased or decreased. A few trials of the remaining LISnoise, PNvoice and PNcovert conditions were presented to the participant to serve as practice. During each run, a digital audio recording was made of the participants' verbal responses.

Magnetic Resonance Imaging Procedure

Imaging was performed on a 3 T whole-body system (Magnetom Trio, Siemens Medical Systems, Erlangen, Germany), using a standard head coil. Functional volumes were acquired using a T2*-weighted echoplanar sequence with BOLD contrast (27 transversal-oblique slices, TR = 4 s, TE = 30 ms, FA = 90°, FoV = 224 mm^2, slice thickness = 4.5 mm, no interslice distance, matrix = 64 × 64 × 27, voxel size = 3.5 × 3.5 × 4.5 mm^3, phase encoding direction = A>P). Volume scanning time was 1.75 s and the interscan gap was 2.25 s. A total of 705 volumes were acquired for each participant in 5 runs of 141 volumes each. The first two volumes of every run were discarded to account for the T1 saturation effect. High-resolution anatomical volumes (voxel size 1 × 1 × 1 mm^3) were acquired using a T1-weighted 3D MP-RAGE (magnetization-prepared rapid acquisition gradient echo) sequence (192 sagittal slices, TR = 2.3 s, TE = 3.93 ms).

Participants were placed comfortably in the scanner and their heads were fixated with the headset and foam pads. Mounted on the head coil was a mirror through which participants could see the stimuli projected on a screen placed outside the scanner.

Auditory stimuli were presented with a MR-compatible Intercom Commander XG MRI Audio System from Resonance Technologies. The sound from this system is transduced by a two-way stereo headset that also serves as ear defender. The headset is air tube driven for the lower fre-

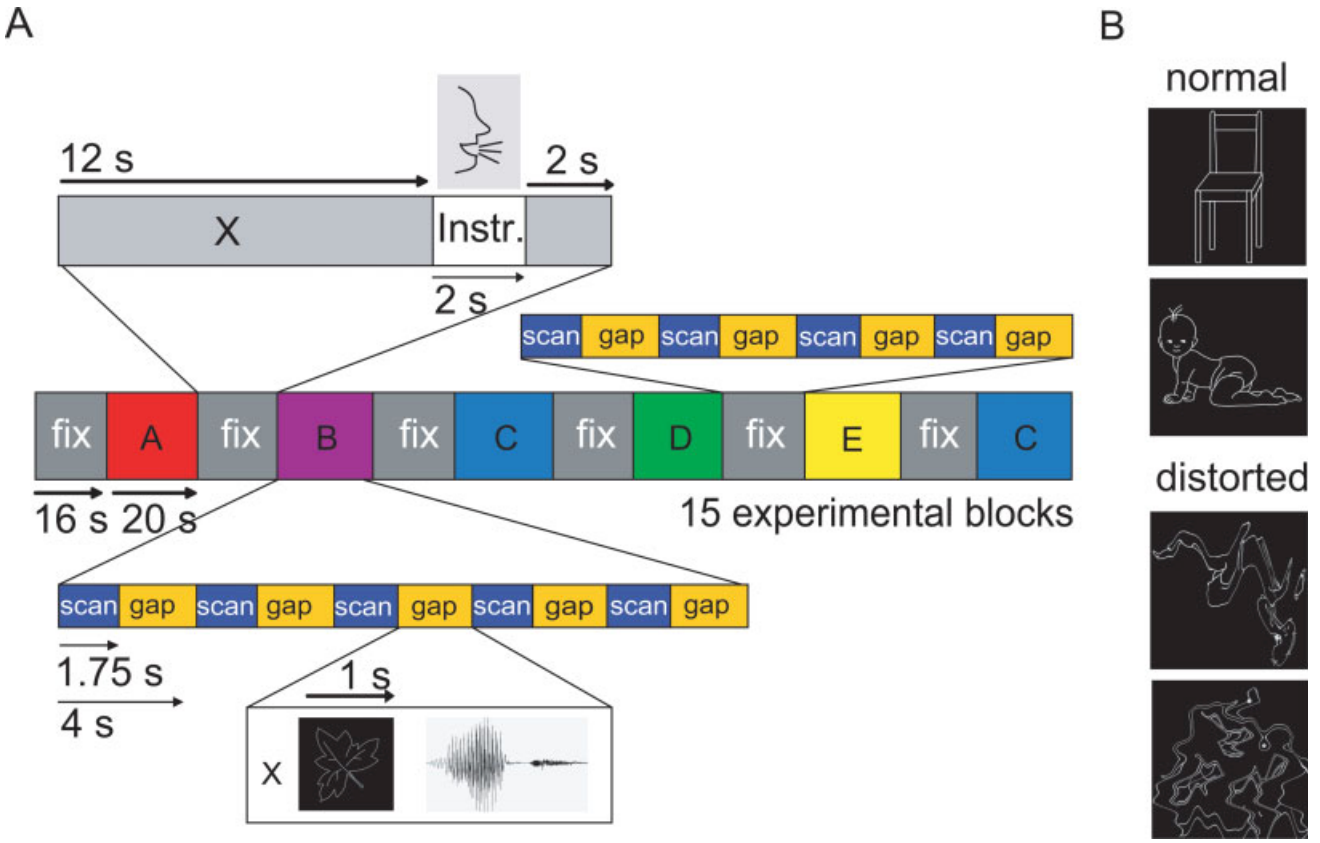

Figure 1.
Stimulation protocol. Illustrated are the timing of task and stimuli and how stimulus presentation and response generation takes place in the interval between volume acquisitions. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

quencies in combination with a nonmagnetic piezo tweeter. Albeit wearing the headset, during speaking with normal feedback, participants were able to hear their own voice. Prior to scanning, the volume of the noise was set individually via the audio system and the voice stimuli were attenuated for an individually set amount. An audio recording was made for each run with a microphone attached to the headset and a separate computer. Trial presentation was synchronized with MR data acquisition by triggering each trial with an MR pulse.

Data Analyses

Anatomical and functional images were analyzed using BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands). The preprocessing steps of the functional images were slice scan time correction, linear trend removal, temporal high-pass filtering (0.00539 Hz), 3D head-movement assessment, and correction by using rigid body transformations. The estimated translation and rotation parameters were inspected and never exceeded 1 mm or degree. Preprocessed functional time-series were coregistered with the within-session anatomical 3-D dataset using position parameters from the scanner and manual adjustment. They were then transformed into Talairach space [Talairach and Tournoux, 1988]. Preprocessed and Talairach-normalized functional volume time-series were used for the statistical analysis.

Group statistical maps were obtained with a voxel-by-voxel two level (hierarchical) random effect analysis of the BOLD-response time courses. At the first-level, a general linear model (GLM) of the experiment was computed, using separate predictors per participant and condition. The predictor time courses were adjusted for the hemodynamic response delay by convolution with a hemodynamic response function [Boynton et al., 1996]. At the second level, group contrast (t) maps were obtained based on the parameter estimates (betas) derived from the first level analysis.

The crucial contrast of interest was between PNvoice and PNnoise. In these two conditions, the only difference was whether participants could hear their own verbal output or whether their speech was masked by noise instead. Random-effect maps were thresholded based on a three-dimensional extension of the randomization procedure described in

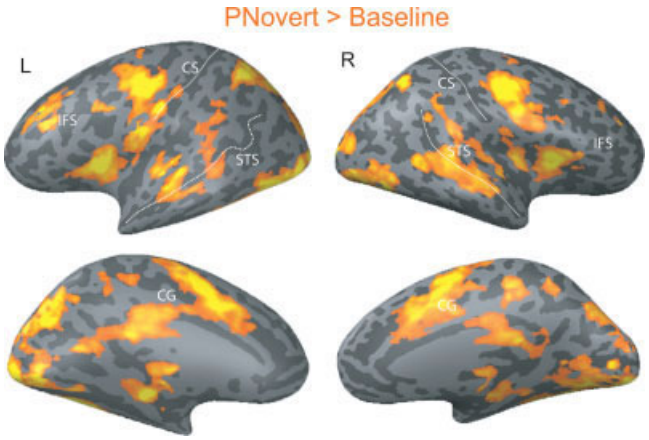

Figure 2.
Picture-naming against baseline. Group-averaged random effect activation maps are superimposed on an inflated MNI template brain. On the inflated template, light and dark gray regions indicate gyri and sulci, respectively. Color indicates *t*-value: $t(13)$ >4.5 to >8 (red to yellow), positive activity. CS = central sulcus; STS = superior temporal sulcus; IFS = inferior frontal sulcus; CG = cingulate gyrus. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Forman et al. [1995]. First, a voxel-level threshold was set at $t = 3.0$ ($P < 0.01$, uncorrected). Thresholded maps were then submitted to a whole-brain correction criterion based on the estimate of the map's spatial smoothness and on an iterative procedure for estimating cluster-level false-positive rates. After 1,000 iterations, maps were applied to the minimum cluster size threshold, which yielded a corrected cluster-level false-positive rate (α) of 5%. The same cluster size threshold was applied to the conjunction analyses, which were based on the minimum statistic. Group data are projected on the anatomical brain template of the Montreal Neurological Institute (MNI).

RESULTS

Behavioral Results

During scanning, participants' responses were recorded and checked for correctness in the overt conditions and for lack of any responses in the covert and listening conditions. None of the participants reported to have problems in following or remembering the block instructions. Very few errors were made (0.02% of the trials). Nevertheless, the recordings showed that one participant gave overt responses in one of the covert blocks. The corresponding run was discarded from further analyses (i.e. only four runs were taken into account for this participant). The audio-recordings furthermore indicated that, in the PNnoise conditions, the pink noise covered the duration of the verbal responses and that the responses were of the same volume in the PNvoice and PNnoise conditions.

During debriefing, participants were asked again whether or not they were able to hear their own voice during the PNnoise condition. Most participants reported that they had not heard their own voice at all. Four participants reported they were not absolutely sure whether or not they might have heard their own voice. However, these four participants indicated that they had difficulty in distinguishing actual perceived speech feedback from their inner speech. On the basis of noise volume and subjects' self-report, it can be concluded that the masking of both external and skull conducted speech was successful.

FMRI Results

Overt speaking

Figure 2 illustrates the overall pattern of brain activity for overt speech against baseline, the basic condition of this experiment (PNvoice). As expected, many areas in the brain are involved in object recognition and the production of single words. Areas were bilaterally activated and included the superior and middle temporal cortices, the inferior frontal gyrus (more extensive on the left), the precentral gyrus, the mid and anterior cingulate cortex, superior parietal lobe, occipital lobe, insula, the supplementary motor area (SMA), and subcortical regions including the thalamus and the cerebellum.

Verbal feedback

The contrast between picture-naming under normal and masked feedback conditions (PNvoice > PNnoise) revealed clusters of stronger activity for normal feedback in a number of different areas (Table I and Fig. 3B). Large clusters were revealed bilaterally in the insula and the ACC (Fig. 3A). The SMA and the bilateral precentral gyri were also activated more strongly when normal feedback was present. Furthermore, a number of subcortical areas responded stronger when feedback was present: the thalamus, the basal ganglia, the cerebellum and the pons. In Figure 3C, the average time courses are plotted for the ACC and the right insula. The average BOLD response indicates that, in the ACC and the insula, the response to both overt speech conditions is stronger than to the listening control conditions. However, the amount of activity is modulated in relation to the normal feedback condition. Especially in the ACC, the time courses show that for the PNnoise condition the average BOLD response initially goes up in parallel to the PNvoice condition but drops down earlier. This pattern suggests that activity is initially similar for both conditions, but it is not sustained when monitoring of auditory feedback is prevented.

Since PNvoice and PNnoise differ not only in the presence of verbal feedback but also in acoustic quality and intensity of the auditory input, one could argue that the results we found are due to differences in auditory input, i.e., between hearing noise and hearing one's own voice. We therefore tested the sensitivity of the regions that were

TABLE I. Talairach coordinates of regions involved in auditory feedback processing, revealed by contrasting PNvoice with PNnoise, vice versa, and the significance ($P < 0.05$, corrected) of these regions in the conjunction between PNvoice > PNnoise and PNvoice > LISvoice

Region	Side	X	Y	Z	t[a]	Conjunction[b]
PNvoice > PNnoise						
Anterior CC		0	8	35	5.7	Yes
CC/GFd (SMA)		−1	−9	46	5.9	Yes
Pons		2	−17	−31	5.6	Yes
Thalamus		1	−15	9	4.6	Yes
Anterior insula	R	37	7	11	6.2	Yes
Precentral gyrus	R	49	−9	30	6.2	Yes
Fusiform gyrus	R	28	−34	−14	5.7	
Cerebellum	R	29	−54	−13	6.4	
STS/MTG	R	49	−25	−2	5.4	
Anterior insula	L	−40	2	9	6.5	Yes
Precentral gyrus	L	−42	−21	41	6.0	Yes
Lingual gyrus	L	−10	−68	7	6.0	
Cerebellum	L	−15	−53	−15	5.2	
NL	L	−27	−10	5	5.5	Yes
PNnoise > Pnvoice						
STG	R	51	−18	8	5.5	
STG	R	57	−30	12	4.9	
STG	L	−41	−28	7	4.1	

CC = cingulate cortex; GFd (SMA) = gyrus frontalis medialis (supplementary motor area); STG = superior temporal gyrus; STS/MTG = superior temporal sulcus/middle temporal gyrus; NL= nucleus lentiformus.

The position of each region is given as the Talairach coordinates of the centre of mass of the suprathreshold clusters of the random effects group analyses.

[a] The *t*-value indicates the peak statistical value for the cluster.

[b] The conjunction column marks whether the cluster was significantly activated in the conjunction of [PNvoice > PNnoise ∩ PNvoice > LISvoice].

modulated by the presence of verbal feedback to differences in auditory input. In a region-of-interest analysis on the region reported in Table I, the contrast LISvoice > LISnoise was tested. We found that there were no significant differences in response to the two listening control conditions in any region, with the exception of the thalamus [$t(13) = 4.1$, $P < 0.002$] and a region in the right MTG [$t(13) = 3.6$, $P < 0.004$], discussed below (see also Fig. 3C,E). Areas modulated by the presence of feedback during speaking were generally not activated differently to auditory noise or speech input. The overall pattern of activity for the two listening conditions against baseline and the contrast map of LISvoice versus LISnoise are illustrated in the supplementary material.

Based on previous research [e.g., Indefrey and Levelt, 2004; McGuire et al., 1996], we expected a substantial role for the STG. In this area, less activity rather than more activity was revealed in the bilateral STG, when feedback was present, by the contrast PNnoise > PNvoice. This cluster is illustrated in Figure 3D. Because of properties of skull bone conduction and middle ear muscle contraction, listening to self-produced speech while speaking (PNvoice) and listening to a recording of one's own voice (LISvoice) are perceptually not completely identical. Although some caution is therefore warranted in interpreting the reduced activity in the STG, note that the average BOLD response is similarly high in all other conditions involving auditory input (PNnoise, LISvoice, LISnoise), even though the recorded speech was presented at a lower volume than the noise (Fig. 3E, top graph). Moreover, in the region-of-interest analysis, the difference between listening conditions was not significant. It is, therefore, unlikely that the reduced response is related to differences in auditory input.

Interestingly, in the right middle superior temporal sulcus/middle temporal gyrus (STS/MTG), inferior to the area associated with reduced response, a small cluster was more active for PNvoice than PNnoise (Fig. 3D, right panel). The graph in Figure 3E shows that this region responds strongly to PNvoice, and (less strongly) to PNnoise and LISvoice, but not to LISnoise. Previously, the MTG has been implicated in semantic processing, albeit especially on the left side [Vandenberghe et al., 1996]. Possibly this pattern reflects the activation of semantic representations necessary for both speech production and comprehension.

Conjunctions

Conjunction analyses may help distinguish areas involved in feedback processing that are associated strongly to speech production from those that are associated strongly to speech perception and auditory processing.

First we combined the contrast of PNvoice > PNnoise with PNvoice > LISvoice. This conjunction revealed those areas that are more strongly activated when feedback is present and more strongly when producing speech than when listening to speech. Areas that were activated in

Figure 3.

The comparison of speaking with normal and masked feedback. (**A**) Group-averaged random effect activation maps are superimposed on a MNI template brain (colors indicates *t*-value). (**B**) Group results are superimposed on an inflated MNI template brain. (**C**) Time courses of the average BOLD response (in percent signal chance) during each of the five conditions for regions in the ACC (blue) and the right insula (green). Bars denote standard errors. The average time courses represent the average response of the voxels belonging to the clusters marked in (A); (**D**) Group-averaged random effect activation maps are superimposed on a MNI template brain (color indicates *t*-value). (**E**) Time courses of the average BOLD response (in percent signal change) during each of the five condition for regions in the STG (purple) and the STS/MTG (orange). Bars denote standard errors. The average time courses represent the average response of the voxels belonging to the clusters marked in (D).

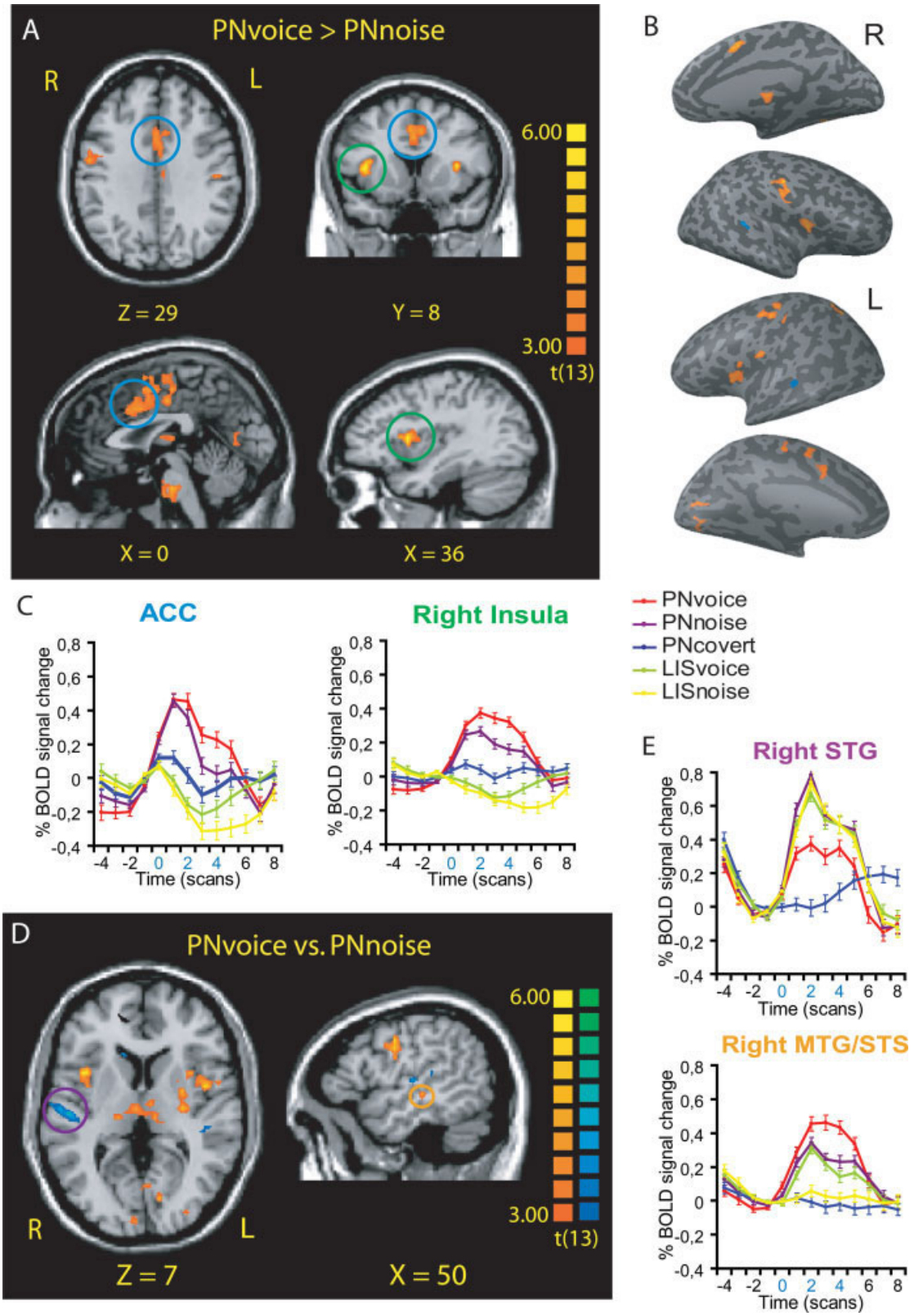
A
PNvoice > PNnoise
R
L
6.00
3.00
t(13)
Z = 29
Y = 8
X = 0
X = 36
B
R
L
C
ACC
Right Insula
% BOLD signal change
Time (scans)
PNvoice
PNnoise
PNcovert
LISvoice
LISnoise
D
PNvoice vs. PNnoise
R
L
Z = 7
X = 50
E
Right STG
Right MTG/STS

Figure 3.

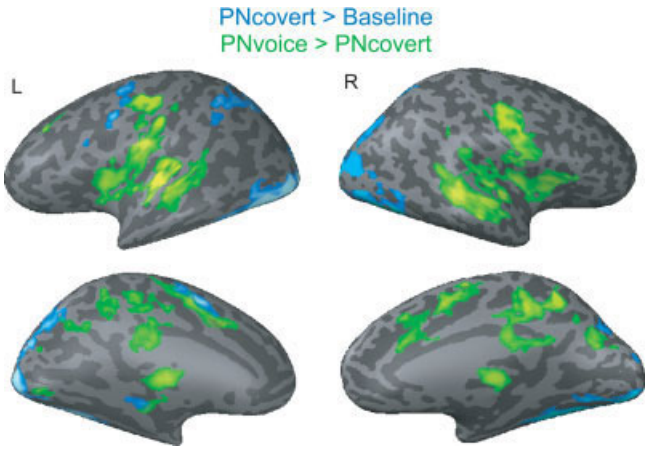

Figure 4.

Covert picture-naming against baseline (blue) and the contrast between overt- and covert-naming (green). Group-averaged random effect activation maps are superimposed on an inflated MNI template brain. Color indicates *t*-value: $t(13)$ >4.5 to >8 (light to dark), positive activity. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

speech perception only are now excluded. The analysis revealed many clusters in the same areas reported earlier, including the ACC, insulae, motor areas, and the subcortical areas. These areas are marked in Table I.

In the second conjunction, we combined the contrast of PNvoice > PNnoise with LISvoice > LISnoise. Including the latter contrast in the conjunction restricts significant clusters to those that are selectively more active to complex auditory input, such as speech, than to noise. The conjunction therefore revealed those areas that are involved in feedback processing but mainly respond to the difference in auditory input (speech versus noise), thus focusing on the comprehension part of feedback processing. In this conjunction, there were no significantly activated areas (the cluster in the STS/MTG mentioned earlier was not significant because regions of the separate simple contrasts did not completely overlap). Combined, the results of these conjunctions indicate that feedback-masking mainly modulated areas related to speech production.

Covert versus overt speaking

Overall there was a large difference in magnitude of the BOLD signal in the overt- and covert-naming conditions. Comparison of PNcovert against baseline revealed active regions in the cingulate cortex, SMA, precentral gyri, superior parietal lobule, occipital lobe and cerebellum. The contrast of PNvoice > PNcovert revealed extensive clusters in many areas including articulation-related areas such as the bilateral precentral gyri, the thalamus, and the basal ganglia extending into the bilateral insulae. Clusters in the bilateral superior and middle temporal gyri, the anterior and mid cingulate cortex extending into the SMA were also more active for overt as compared to covert-naming. There were no clusters significantly more activated for covert picture-naming. In Figure 4, the overall pattern of activity of PNcovert against baseline is shown together with the contrast of PNovert > PNcovert. The figure illustrates the large difference in responsiveness of the BOLD signal in the two conditions.

DISCUSSION

This study examined the neurocognitive correlates of verbal feedback processing by using normal and auditorily masked feedback conditions. The results show that auditory feedback processing involves a network of different areas related to general performance monitoring and speech-motor planning and control, including the cingulate cortex, the bilateral insulae, SMA, bilateral motor areas, thalamus, cerebellum and basal ganglia. As expected, we found modulation of activity in areas almost all of which are reliably found in overt speech production [Indefrey and Levelt, 2004]. In most areas, the masking of feedback (PNnoise) was associated with a reduction in activity in comparison to the normal feedback condition (PNvoice).

Although speech output was the same in PNvoice and PNnoise, the modulation of response in the cerebellum, motor cortex and SMA suggests a difference in speech-motor activity between the two conditions. As mentioned earlier, there is evidence that speakers immediately adjust their speech output, such as its volume, to external circumstances. Houde and Jordan [1998] report, for instance, that altering the quality of feedback induces compensatory changes such as in the production of vowels during alteration of perceived formants. Our data therefore indicate a network of areas that acts on auditory input to adjust speech-motor planning. When feedback processing is possible and output-adjustment takes place based on this feedback, there is more activity in areas related to speech-motor planning and programming. The conjunction analyses suggest that most areas that are modulated by external feedback are indeed involved in speech production rather than comprehension. In other words, the pattern of results we found strongly suggests that even in simple overt picture-naming, the auditory speech input is processed continuously to adjust speech-motor planning.

We furthermore found strong responses to speech with external feedback of two areas that have previously been reported to be activated in a range of different tasks, i.e., the ACC and the bilateral insulae. This suggests that speech-monitoring largely depends on language nonspecific areas. The ACC, usually associated with response inhibition and conflict monitoring [e.g., Botvinick et al., 2004; Ridderinkhof et al., 2004b], is apparently also engaged in ongoing speech-monitoring. The modulation of ACC activity is noteworthy, since it is unlikely that response conflict was present in our task, and certainly no reason for conflict to be higher in the feedback condition than in the masked-feedback condition. Therefore, our results suggest

that the ACC may serve quite a general monitoring function: the continuous monitoring of output performance. Our design allowed for the study of ongoing performance monitoring without increasing task difficulty or conflict, and the data show that the ACC is involved in this.

Other regions that respond strongly in the presence of feedback are the bilateral anterior insular cortices. The exact function of the insula is unclear, since activity has been reported in many different studies. It is involved in affective processing, pain perception [Craig, 2002], in the introspective awareness of heart beat [Critchley et al., 2004] and in monitoring the sensory environment [Linden et al., 1999]. The insula is also implicated in speech production, although its exact role there is a matter of debate [Shuster and Lemieux, 2005]. Nevertheless, the bilateral anterior insulae are reliably activated in word production [Indefrey and Levelt, 2004]. The left anterior insula furthermore has not only been related to articulatory speech programming [Dronkers, 1996] and to automatization of lexical retrieval [Van Turennout et al., 2000], but also to auditory speech processing and verbal working memory [Augustine, 1996]. Ackermann and Riecker [2004] suggested that the anterior insular cortex is involved in speech-motor control because they found no activity in covert speech. Similarly, we did not find strong activity in the insula in the covert-naming condition. Processing verbal feedback requires the integration of information from auditory areas and fine adjustment of motor processing. The insulae are well-connected with reciprocal connections to the ACC and the orbitofrontal cortex and extensive unilateral efferent connections to motor areas. Afferent input comes from both somatosensory as well as auditory areas [Augustine, 1996]. Therefore, it seems that the insulae are ideally suited for on-line adjustment of speech-motor processing. We suggest that in our study this happened on the basis of verbal feedback.

Based on previous studies, we expected that the STG would be important for the processing of feedback [e.g., Indefrey and Levelt, 2004]. Our results indicate, however, that its role might be different than previously suggested [Hirano et al., 1997; McGuire et al., 1996]. We found no evidence that the STG is more activated when the feedback was present than when it was masked. Rather, in this area, we found relatively small clusters that showed the opposite pattern. The presence of feedback resulted in reduced activity bilaterally, most notably in the right STG relative to all other conditions where auditory input was present. This is in accordance with studies that report a reduced and delayed MEG or ERP signal when speaking [Curio et al., 2000; Houde et al., 2002; Numminen and Curio, 1999]. It may be possible that we observed the BOLD equivalent of this attenuation to the response to auditory feedback when speaking. Interestingly, using fMRI we were able to show that this attenuation is relatively localized because just inferior to this region the presence of feedback during overt speech is related to a relatively stronger BOLD signal in the right STS/MTG.

A plausible framework to explain the modulation of the activity in the STG in the normal feedback condition is that the activity is attenuated by a forward or priming mechanism [e.g., Ford et al., 2005; Heinks-Maldonado et al., 2005; Houde et al., 2002; Martikainen et al., 2005; Numminen and Curio, 1999; Paus et al., 1996]. The auditory cortex is primed for a change in the input that is about to occur due to one's own speech production. An efference copy, a copy of the motor commands, is used to predict the sensory consequence and when it matches the actual input, it is used to attenuate the response. In analogy to motor control theory [e.g., Blakemore et al., 1998; Wolpert et al., 1995), it is assumed that the comparison between prediction and actual sensory feedback makes it possible to distinguish the sensory consequences of our actions from sensory signals due to changes in the outside world. In evidence, different rates of whispering affected cerebral blood flow in the secondary auditory cortex even though auditory input was masked [Paus et al., 1996]. In the absence of different auditory input, the modulation of the auditory cortex may have been mediated by a feedback mechanism in this study.

This forward model framework provides a straightforward explanation for the discrepancy between our results and some of the previous PET and fMRI studies that found strong STG activity as a consequence of feedback modulation. Since these studies all manipulated verbal feedback in one way or another, this presumably reduced the match between expected and actual feedback. Consequently, there was no response attenuation to the voice during speaking in the manipulated conditions, resulting in stronger STG activity. This explanation provides an alternative account for the lack of STG activity reported by Hirano et al. [1996] when comparing reading aloud short sentences and syllables to the resting baseline. Hirano et al. [1997] suggested that no speech-monitoring takes place for speaking of familiar sentences to explain the lack of STG activity in that study. This seems unlikely, since in our study only single high-frequency words were produced. Possibly, in their study, feedback-related response reduction and activation cancelled each other out, since only a small area of the STG showed the reduced response, and immediately inferior in the STS/MTG, at least in the right hemisphere, the opposite pattern occurs. This account also provides an alternative to the suggestion that more activity in the STG for a feedback manipulation condition reflects more self-monitoring in this condition [McGuire et al., 1996]. It is possible that external feedback was not recognized as being self-produced anymore, and therefore not attenuated. Note that even though our interpretation of the data slightly differs from McGuire et al. [1996], our data are in general compatible with theirs. We found active clusters in some of the areas in which McGuire et al. [1996] reported decreases in conditions of distorted feedback (the left ACC, the SMA) and alien feedback (trends in the left putamen and insula). Similar to our study, these areas were more active during normal feedback than when feedback could not be clearly perceived. It appears that the STG is involved in feedback-processing and that it has an important role in distinguishing between self-produced and external speech.

Functional imaging studies not only reported much greater BOLD responses during overt speech than during covert speech in many brain regions, but different patterns of activity as well [Huang et al., 2001; Shuster and Lemieux, 2005]. In our study, quite large differences between overt and covert picture-naming were observed. There was also a large difference in magnitude of the BOLD signal, which made the two conditions difficult to compare so that any qualitative differences between overt and covert conditions cannot easily be assessed. Nevertheless, these results cast further doubt on the comparability of covert to overt production [Huang et al., 2001; Shuster and Lemieux, 2005]. Moreover, our study showed that differences in auditory feedback, a fundamental aspect of speech monitoring, which is in itself a core aspect of speech production, gives rise to the modulation of activity in a large network of interconnected areas in the brain. The lack of auditory feedback is not taken into account when covert speech production is employed to study the neural correlates of speech. Taken together, the large difference in magnitude of the BOLD signal between overt and covert conditions and the continuous interaction between verbal feedback and speaking indicates that covert-naming may not address normal speech production.

To conclude, we found that in the superior temporal gyrus the response to feedback during speaking is reduced, suggesting that this area is important for distinguishing self-produced from alien speech. Our data demonstrate that feedback processing engages a network of cortical and subcortical areas, which reflects the continuous monitoring of our verbal output on factors like content, quality, pitch and intensity, and acts accordingly through continuous adjustment of the speech-motor plans. Especially the bilateral insulae and the ACC appear to be the core regions of this network.

ACKNOWLEDGMENTS

The authors wish to thank Peter Hagoort, Peter Indefrey, Lourens Waldorp, and the members of the Maastricht Brain Imaging Center for helpful discussions and Paul Gaalman for technical assistance during scanning.

REFERENCES

Ackermann H, Riecker A (2004): The contribution of the insula to motor aspects of speech production: A review and a hypothesis. Brain Lang 89:320–328.

Aleman A, Formisano E, Koppenhagen H, Hagoort P, De Haan EHF, Kahn RS (2005): The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. Cereb Cortex 15:221–228.

Allen PP, Amaro E, Fu CHY, Williams SCR, Brammer M, Johns LC, McGuire PK (2005): Neural correlates of the misattribution of self-generated speech. Hum Brain Mapp 26:44–53.

Augustine JR (1996): Circuitry and functional aspects of the insular lobe in primates including humans. Brain Res Rev 22:229–244.

Baayen RH, Piepenbrock R, Gulikers L (1995): The CELEX Lexical Database. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania. CD-ROM.

Barch DM, Sabb FW, Carter CS, Braver TS, Noll DC, Cohen JD (1999): Overt verbal responding during fMRI scanning: Empirical investigation of problems and potential solutions. NeuroImage 10:642–657.

Barch DM, Braver TS, Sabb FW, Noll DC (2000): Anterior cingulate and the monitoring of response conflict: Evidence from an fMRI study on overt verb generation. J Cogn Neurosci 12:298–309.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000): Voice-selective areas in human auditory cortex. Nature 403:309–312.

Birn RA, Cox RW, Bandettini PA (2004): Experimental designs and processing strategies for fMRI studies involving overt verbal responses. NeuroImage 23:1046–1085.

Blakemore S-J, Wolpert DM, Frith CD (1998): Central cancellation of self-produced tickle sensation. Nature Neurosci 1:635–640.

Botvinick MM, Braver DM, Barch CS, Carter JD, Cohen JD (2001): Conflict monitoring and cognitive control. Psychol Rev 108: 624–652.

Botvinick MM, Cohen JD, Carter CS (2004): Conflict monitoring and anterior cingulate cortex: An update. Trends Cogn Sci 8:539–546.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996): Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD (1998): Anterior cingulate cortex, error detection, and online monitoring of performance. Science 280:747–749.

Craig AD (2002): How do you feel? Interoception: The sense of the physiological condition of the body. Nat Rev Neurosci 3:655–666.

Critchley HD, Wiens S, Rotshtein P, Öhman A, Dolan RJ (2004): Neural systems supporting interoceptive awareness. Nat Neurosci 7:189–195.

Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R (2000): Speaking modifies voice-evoked activity in the human auditory cortex. Hum Brain Mapp 9:183–191.

De Zubicaray GI, Wilson SJ, McMahon KL, Muthiah S (2001): The semantic interference effect in the picture-naming word paradigm: An event-related fMRI study employing overt responses. Hum Brain Mapp 14;218–227.

Dronkers NF (1996): A new brain region for coordinating speech articulation. Nature 384;195–161.

Eliades SJ, Wang X (2003): Sensory-motor interaction in the primate auditory cortex during self-initiated. J Neurophysiol 98: 2194–2207.

Ford JM, Gray M, Faustman WO, Heinks TH, Mathalon DH (2005): Reduced γ-band coherence to distorted feedback during speech when what you say is not what you hear. Int J Psychophysiol 57:143–150.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995): Improved assessment of significant activation in Functional Magnetic Resonance Imaging (fMRI): Use of a cluster-size threshold. MRM 33:636–646.

Gracco ZL, Tremblay P, Pike B (2005): Imaging speech production using fMRI. NeuroImage 26:294–301.

Ganushchak L, Schiller NO (in press): Effects of time pressure on verbal selfmonitoring. Brain Res.

Hartsuiker RJ, Kolk HHJ (2001): Error monitoring in speech production: A computation test of the perceptual loop theory. Cogn Psychol 42:113–157.

Hashimoto Y, Sakai KL (2003): Brain activations during conscious self-monitoring of speech production with delayed auditory feedback. Hum Brain Mapp 20:22–28.

Heinks-Maldonado TH, Mathaldon DH, Gray M, Ford JM (2005): Fine-tuning of auditory cortex during speech production. Psychophysiology 42:180–190.

Hirano S, Kojima H, Naito Y, Honjo I, Kamoto Y, Okazawa H, Ishizu K, Yonekura Y, Nagahama Y, Fukuyama H, Konishi J (1996): Cortical speech processing mechanisms while vocalizing visual presented languages. NeuroReport 7:363–367.

Hirano S, Kojima H, Naito Y, Honjo I, Kamoto Y, Okazawa H, Ishizu K, Yonekura Y, Nagahama Y, Fukuyama H, Konishi J (1997): Cortical processing mechanism for vocalization with auditory verbal feedback. NeuroReport 8:2379–2382.

Holroyd CB, Coles MGH (2002): The neural basis of human error processing: Reinforcement learning, dopamine and the error-related-negativity. Psychol Rev 109:679–709.

Houde JF, Jordan MI (1998): Sensimotor adaption in speech production. Science 279:1213–1216.

Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002): Modulation of the auditory cortex during speech: An MEG study. J Cogn Neurosci 14:1125–1138.

Huang J, Carr TH, Cao Y (2001): Comparing cortical activation for silent and overt speech using event-related fMRI. Hum Brain Mapp 15:39–53.

Indefrey P, Levelt WJM (2004): The spatial and temporal signatures of word production components. Cognition 92:101–144.

Jäncke L, Wustenberg T, Scheich H, Heinze HJ (2002): Phonetic perception and the temporal cortex. NeuroImage 15:733–746.

Kan IP, Thompson-Schill SL (2004): Effect of name agreement on prefrontal activity during overt and covert picture naming. Cogn Affect Behav Neurosci 4:43–57.

Lane H, Tranel B (1971): The Lombard sign and the role of hearing in speech. J Speech Hear Res 14:667–709.

Lee BS (1950): Effects of delayed speech feedback. J Acoust Soc Am 22:824–826.

Levelt WJM (1989): Speaking: From Intention to Articulation. Cambridge, MA: MIT Press.

Levelt WJM (1999): Models of word production. Trends Cogn Sci 3:223–231.

Levelt WJM, Roelofs A, Meyer A (1999): A theory of lexical access in speech production. Behav Brain Sci 22:1–75.

Linden DEJ, Prvulovic D, Formisano E, Völlinger M, Zanella FE, Goebel R, Dierks T (1999): The functional neuroanatomy of target detection: An fMRI study of visual and auditory oddball tasks. Cereb Cortex 9:815–823.

Martikainen MH, Kaneko K, Hari R (2005): Suppressed responses to self-triggered sounds in the human auditory cortex. Cereb Cortex 15:299–302.

Masaki H, Tanaka H, Takasawa N, Yamazaki K (2001): Error-related brain potentials elicited by vocal errors. NeuroReport 12:1851–1855.

McGuire PK, Silversweig DA, Frith CD (1996): Functional neuroanatomy of verbal self-monitoring. Brain 119:907–917.

Mulert C, Menzinger E, Leight G, Pogarell O, Hegerl U (2005): Evidence for a close relationship between conscious effort and anterior cingulate activity. Int J Psychophys 56:65–80.

Munhall KG (2001): Functional imaging during speech production. Acta Psychol 107:95–117.

Numminen J, Curio G (1999): Differential effects of overt, covert and replayed speech on vowel evoked responses of the human auditory cortex. Neurosci Lett 272:39–32.

Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.

Palmer ED, Rosen HJ, Ojemann JG, Buckner RL, Kelly WM, Petersen SE (2001): An event-related fMRI study of overt and covert word stem completion. NeuroImage 14:182–193.

Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC (1996): Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. Eur J Neurosci 8:2236–2246.

Postma A (2000): Detection of errors during speech production: A review of speech monitoring models. Cognition 77:97–131.

Price CJ, Wise RJS, Warburton CJ, Moore D, Howard K, Patterson K, Frackowiak RSJ, Friston KJ (1996): Hearing and saying: The functional neuro-anatomy of auditory word processing. Brain 119:919–931.

Price CJ, Devlin JT, Moore JM, Morton C, Laird AR (2005): Meta-analyses of object naming: Effect of baseline. Hum Brain Mapp 25:70–85.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004a): The role of the medial frontal cortex in cognitive control. Science 306:443–447.

Ridderinkhof KR, Van der Wildenberg WPM, Segalowitz SJ, Carter CS (2004b): Neurocognitive mechanisms of cognitive control: The role of the prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. Brain Cogn 56:129–140.

Schiller NO (2005): Verbal self-monitoring. In: Cutler A, editor. Twenty-First Century Psycholinguistics: Four Cornerstones. Hillsdale, NJ: Lawrence Erlbaum Associates. pp 245–261.

Schiller NO, De Ruiter JP (2004): Some notes on priming, alignment, and self-monitoring [Commentary]. Behav Brain Sci 27:208–209.

Schiller NO, Jansma BM, Peters J, Levelt WJM (2006): Monitoring metrical stress in polysyllabic words. Lang Cogn Process 21:112–140.

Schneider W, Chein JM (2003): Controlled and automatic processing: Behavior, theory, and biological mechanism. Cogn Sci 27:525–559.

Scott SK, Blank CC, Rosen S, Wise RJS (2000): Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Shuster LI, Lemieux SK (2005): An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. Brain Lang 93:20–31.

Talairach T, Tournoux P (1988): A Coplanar Stereotactic Atlas of the Human Brain. Stuttgart: Thieme Verlag.

Van Atteveldt N, Formisano E, Goebel R, Blomert L (2004): Integration of letters and speech sounds in the human brain. Neuron 43:271–282.

Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RSJ (1996): Functional anatomy of a common semantic system for words and pictures. Nature 383:254–256.

Van Turennout M, Ellmore TL, Martin A (2000): Long-lasting cortical plasticity in the object naming system. Nat Neurosci 3:1329–1334.

Van Turennout M, Bielamowicz L, Martin A (2003): Modulation of neural activity during object naming: Effects of time and practice. Cereb Cortex 13:381–391.

Vouloumanos A, Kiehl KA, Werker JF, Liddle PF (2001): Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. J Cogn Neurosci 13:994–1005.

Wheeldon L, Levelt WJM (1995): Monitoring the time course of phonological encoding. J Mem Lang 34:311–334.

Wise RJS, Greene J, Büchel C, Scott S (1999): Brain regions involved in articulation. Lancet 353:1057–1061.

Wolpert DM, Ghahramani Z, Jordon MI (1995): An internal model for sensorimotor integration. Science 269:1880–1882.

Yates AJ (1963): Delayed auditory feedback. Psychol Bull 60:213–232.