Cerebral Cortex February 2005;15:221–228
doi:10.1093/cercor/bhh124
Advance Access publication July 21, 2004

The Functional Neuroanatomy of Metrical Stress Evaluation of Perceived and Imagined Spoken Words

André Aleman[1,2], Elia Formisano[3], Heidi Koppenhagen[3], Peter Hagoort[4], Edward H.F. de Haan[2] and René S. Kahn[1]

[1]Department of Psychiatry, Rudolf Magnus Institute for Neuroscience, University Medical Center, Utrecht, The Netherlands, [2]Psychological Laboratory, Helmholtz Research Instituut, Utrecht University, Utrecht, The Netherlands, [3]Department of Cognitive Neuroscience, Faculty of Psychology, Maastricht University, Maastricht, The Netherlands and [4]FC Donders Center for Cognitive Neuroimaging, Nijmegen University, Nijmegen, The Netherlands

We hypothesized that areas in the temporal lobe that have been implicated in the phonological processing of spoken words would also be activated during the generation and phonological processing of imagined speech. We tested this hypothesis using functional magnetic resonance imaging during a behaviorally controlled task of metrical stress evaluation. Subjects were presented with bisyllabic words and had to determine the alternation of strong and weak syllables. Thus, they were required to discriminate between weak-initial words and strong-initial words. In one condition, the stimuli were presented auditorily to the subjects (by headphones). In the other condition the stimuli were presented visually on a screen and subjects were asked to imagine hearing the word. Results showed activation of the supplementary motor area, inferior frontal gyrus (Broca's area) and insula in both conditions. In the superior temporal gyrus (STG) and in the superior temporal sulcus (STS) strong activation was observed during the auditory (perceptual) condition. However, a region located in the posterior part of the STS/STG also responded during the imagery condition. No activation of this same region of the STS was observed during a control condition which also involved processing of visually presented words, but which required a semantic decision from the subject. We suggest that processing of metrical stress, with or without auditory input, relies in part on cortical interface systems located in the posterior part of STS/STG. These results corroborate behavioral evidence regarding phonological loop involvement in auditory–verbal imagery.

Keywords: auditory–verbal imagery, imagined speech, metrical stress, phonological processing, temporal lobe

Introduction

Auditory mental imagery has been defined as 'the introspective persistence of an auditory experience, including one constructed from components drawn from long-term memory, in the absence of direct sensory instigation of that experience' (Intons-Peterson, 1992). Auditory imagery might play an important role in numerous cognitive operations, such as the 'inner language' most people report to experience subjectively on a daily basis, the rehearsal processes of working memory, music perception and cognition, and even the auditory hallucinations characteristic of schizophrenia (Reisberg, 1992; Smith *et al.*, 1995; Aleman *et al.*, 2003). Research on the neural basis of mental imagery has been predominantly focused on visual imagery (Mellet *et al.*, 1998; Kosslyn *et al.*, 1999; Trojano *et al.*, 2000). For example, one of the main findings in the realm of spatial visual imagery is the evidence for a specific convergence of the pathways of imagery and perception within the parietal lobes (Aleman *et al.*, 2002; Formisano *et al.*, 2002; Sack *et al.*, 2002). With regard to auditory imagery, however, it has been argued that we cannot presume that insights about visual imagery will simply generalize (Smith *et al.*, 1995). Not only do sound and light have different sensory characteristics, hearing and vision also clearly differ in nature and extent of subserving neuroanatomical systems and associated information processing characteristics across different species (Gazzaniga, 2000).

Functional neuroimaging investigations of auditory–verbal imagery have been limited to studies using tasks that were not behaviorally controlled. For example, subjects were asked to imagine talking themselves (inner speech) or to imagine somebody talking in the second or third person (e.g. Shergill *et al.*, 2001). Notably, however, such tasks do not require a behavioral response that allows the investigator to verify actual execution of the imagery task online. Hence, the researcher cannot be certain whether or not the targeted cognitive processing actually occurred.

Behavioral measurement of auditory imagery starts from a number of assumptions (Baddeley and Logie, 1992): (i) that an auditory image involves a conscious experience; (ii) that this resembles in certain, as yet unspecified, ways the experience of hearing the sound in question directly; but (iii) the image can be present in the absence of any auditory signal; and (iv) it can be evoked intentionally by the subject. Mental imagery is mediated by working memory, as it involves temporary ('online') storage and manipulation of information. The working memory model initially proposed by Baddeley and Hitch (1974) implies two components involved in auditory processing: an attentional control system (the central executive) and a temporary storage and rehearsal system, the phonological loop. Indeed, there is evidence that, in behavioral tasks of auditory–verbal imagery, subjects subvocally rehearses the imagery material, which places the material in a phonological store that allows the imagery judgement (Reisberg *et al.*, 1989; Reisberg, 1992). In an important series of experiments, Smith *et al.* (1995) have demonstrated that concurrent articulatory suppression (e.g. repeating the word 'Suzie' while performing the imagery task) affected task performance significantly. An example of an imagery task used by these authors is a task in which the participants were presented with words ending in 's' or 'z' and had to indicate the ones that sound like they end in 'z' (e.g. 'cats' ends in an unvoiced /s/ sound, but 'dogs' ends in a voiced /z/ sound).

The aim of the present study was to investigate the neural correlates of auditory–verbal imagery, with the use of a novel task that requires subjects to encode words phonologically and that allows quantification of performance in order to verify actual execution of the imagery task online. We hypothesized that areas in the temporal lobe that have been implicated in the phonological analysis of spoken words would also be activated during the generation and phonological processing of imagined

speech. Therefore we compared a task condition that required a metrical stress evaluation of actually spoken words to a task condition that required a metrical stress evaluation of visually presented words whose phonological representation had to be imagined. Subjects were asked to indicate, for bisyllabic words, the syllable that carries the stress. English and Dutch are so-called stress-timed languages, i.e. they distinguish between strong and weak syllables. The former contain full vowels, whereas the latter contain reduced vowels. Metrical stress is one of the rhythmic properties of language that enables a listener to segment the continuous speech stream to isolate word candidates (Cutler and Norris, 1988). For individual words, the stress pattern is part of the lexical knowledge, presumably represented by means of a so-called metrical grid (Levelt, 1989). A speaker has to retrieve the metrical grid information for a particular word from the mental lexicon if he wants to produce it with the correct stress pattern. Although previous research has established an electrophysiological correlate of stress discrimination, using a metrical stress evaluation task in Dutch native speakers (Böcker *et al.*, 1999), the functional neuroanatomy of metrical stress evaluations remains unclear.

With regard to the processing components of the present stress evaluation task, we assume that, for the auditorily presented words, metrical stress is extracted from the acoustic signal and the subsequent rehearsal of the phonological form in working memory, and is eventually mapped onto the lexically specified stress pattern. In contrast, for the visually presented words, the stress pattern has to be internally generated through the retrieval and activation of the phonological form (including stress) on the basis of the orthographic input. Thus, this condition involves a stronger production component. In this case, the stress assignment task can only be performed by 'listening' to the 'inner voice', whereas for the auditorily presented words subjects can partly rely on the external voice. We predicted that, regardless of input modality, the task would activate the speech production areas, Broca and supplementary motor area (SMA) (the 'inner voice'), and would in addition activate the superior temporal gyrus (STG) and superior temporal sulcus (STS) regions (the 'inner ear').

We also predicted that metrical stress processing would be mediated predominantly by the left hemisphere (Gandour and Baum, 2001).

Materials and Methods

Subjects

Six healthy, right-handed volunteers (four male, two female) without any history of neurological or psychiatric disease participated in this study. All subjects were undergraduate university students at Nijmegen University and were native Dutch speakers. All subjects gave their written informed consent. The study was approved by the Ethical Committee of the University Medical Center Nijmegen.

Experimental Procedure

Our main interest was the comparison of two task conditions: metrical stress evaluation with and without auditory input. In the first condition (the Perception condition) bisyllabic words were presented auditorily via headphones. A digital recording was made of the words, as produced by a native Dutch female speaker. Subsequently, the subjects were required to discriminate between weak-initial words and strong-initial words. Every 5 s a stimulus word was presented auditorily, and participants used the index finger of the right hand to make a 'strong-initial' response on a keypad or the middle finger of their right hand to make a 'weak-initial' response. In the other condition (the Imagery condition) the words were presented on a screen (2000 ms on, 3000 ms off) and subjects were instructed to imagine hearing the word being spoken clearly by another person. In all other respects, the task was identical to the Perception condition. Thus, subjects were required to discriminate between weak-initial words and strong-initial words. Half of the stimuli concerned weak-initial words and half of the stimuli concerned strong-initial words. The vast majority (81%) of the words were monomorphemic. Caution was taken that subjects could not easily derive their response from the visual appearance of the letters in the word, e.g. words with ending '-en' in Dutch always almost are strong-initial words. A behavioral pilot study in our laboratory revealed that the Imagery condition was significantly more affected by concurrent articulatory suppression (longer RTs) than a visual imagery condition with the same stimuli, in accordance with the notion that the former relies on phonological encoding. That is, the metrical stress evaluation task was contrasted with a visual word-length evaluation task using the same stimuli. In the visual word-length evaluation task, subjects ($n = 30$) had to determine visual word length (longer or shorter than a given horizontal line). Two interference conditions were included: (i) articulatory suppression (interferes with phonological processing) and (ii) finger tapping of a spatial pattern (interferes with visuo-spatial processing). Subjects performed worse in the dual-task conditions than when performing the stress evaluation or the word-length task only. A specific effect of type of interference was observed in subjects' reaction time, depending on whether the task was auditory–verbal or visual. Reaction times increased significantly due to articulatory suppression for the stress evaluation task [$F(1,29) = 4.5$, $P = 0.04$], but not for the visual word-length evaluation task ($P = 0.19$). In contrast, a significant effect of concurrent finger tapping on reaction time was only observed in the visual word-length evaluation task [$F(1,29) = 70.0$, $P < 0.0001$] and not for the stress evaluation task ($P = 0.38$). These findings are clearly indicative of involvement of phonological processing in the stress evaluation task, which is not observed when subjects rely on a visual strategy.

During the imaging sessions (see below), the Imagery and Perception conditions were separated by a passive condition (Fixation), to allow for the hemodynamic response to return to baseline levels. We also included a fourth condition designed to control for brain activation that would be merely due to processing visually presented words and reacting with a motor response. The Semantic Decision condition is identical to the Imagery condition in terms of both the visual word input and the two-choice task configuration. However, in contrast to the Imagery condition, performing this task does not require the retrieval of phonological aspects such as word stress. By comparing the blood oxygenation level-dependent (BOLD) responses of these two conditions, we might thus be able to identify the area(s) with a specialization for phonological word form characteristics. Indeed, neuroimaging studies have suggested that semantic tasks do not activate areas associated with phonological processing (Noppeney and Price, 2002). In this condition (Semantic Decision), identical bisyllabic words were again presented on a screen, but subjects were instructed to make a semantic decision, i.e. whether the word had a positive or a negative connotation. The words were selected from word lists that had been rated as positive (50%) or negative (50%) words in pilot studies (Hermans and de Houwer, 1994).

Magnetic Resonance Imaging Procedure

A 1.5 T Siemens SONATA system (Siemens, Erlangen, Germany) was used to acquire both anatomical and functional volumes. Anatomical volumes were collected using a three-dimensional (3-D) T1-fast-low-angle shot (FLASH) sequence (180 slices, $T_R = 30$ ms; $T_E = 5$ ms; FA = 40°; FoV = 256 × 256 mm^2; matrix = 256 × 256; voxel size = 1 × 1 × 1 mm^3). Functional volumes were collected using a T2*-weighted echoplanar sequence with BOLD contrast (18 slices; $T_R = 2.5$ s; $T_E = 40$ ms; FA = 90°; FoV = 220 × 220 mm^2; slice thickness = 3 mm, matrix = 64 × 64; voxel size = 3.4 × 3.4 × 3 mm^3).

The tasks were performed in two functional runs. Each run consisted of six active task blocks (30 s each), alternated with six fixation blocks (15 s each). The active task blocks consisted of three imagery blocks

and three perception (or semantic decision) blocks, which were alternated. The stimulus was presented in between the scanned volumes, i.e. in a silent period. Each run commenced with the acquisition of six dummy volumes, allowing tissue magnetization to achieve a steady state.

Data Analysis

Data were processed using BrainVoyager 2000 version 4.8 software (www.BrainVoyager.com; Brain Innovation, Maastricht, The Netherlands). Preprocessing of functional volumes included interslice scan correction, 3-D head movement assessment and correction (scans with head movement >1.5 mm were rejected), removal of linear trends and high frequency temporal filtering. For the group analysis (see below), functional scans were additionally smoothed spatially using a Gaussian kernel (FWHM = 2 voxels).

Preprocessed functional time-series were coregistered with the within-session anatomical 3-D data-set using the position parameters of the scanner and transformed in the Talairach space (Talairach and Tournoux, 1988). The Talairach transformation was performed in two steps. The first step consisted of rotating the anatomical 3-D data set of each subject to be aligned with the stereotaxic axes. For this step, the location of the anterior commissure, the posterior commissure and two rotation parameters for mid-saggital alignment had to be specified manually in the 3-D dataset. In the second step, the extreme points of the cerebrum were specified. The 3-D anatomical data sets and the functional time-courses were then scaled into Talairach space using a piecewise affine and continuous transformation for each of the defined subvolumes. Preprocessed and Talairach-normalized functional time series were used for the statistical analysis (see below). Statistical results were visualized through projecting 3-D statistical maps on a Talairach-normalized unfolded brain (cf. Formisano *et al.*, 2002).

Statistical analysis was based on a voxel-by-voxel multiple regression analysis of the time-courses (Friston *et al.*, 1995). Both in the case of group analysis (Figs 1 and 2) and in the case of single-subject analysis (Table 1), a general linear model of the experiment was computed. The design matrix included one positive regressor for each task condition; the time-course of each regressor was obtained by using a linear model of the hemodynamic response (Boynton *et al.*, 1996). The activation during each condition was assessed using an *F*-statistic based on the corresponding regressor (Table 1, Fig. 1). The overall pattern of activation during our two main conditions (Perception, Imagery) was visualized using a relative contribution (RC) map (see Fig. 1, $P < 0.001$, corrected). The subset of activation that was significantly activated during both the Perception and the Imagery condition was selected by computing a conjunction map (Perception > baseline) and (Imagery > baseline) green map in Figure 2 ($P < 0.01$, corrected).

Finally, to isolate the voxels that responded to both the perceptual and the imagery task and whose activation level was higher during these conditions than during the Semantic Decision task, we repeated the conjunction analysis by including the additional constraints (Perception > Semantic Decision) and (Imagery > Semantic Decision). These conjunction analyses are very conservative in that an effect is considered significant and a voxel is color-coded only if all the involved contrasts are simultaneously significant. For this reason, we selected a less stringent, yet adequate, threshold ($P < 0.01$, corrected) than the one used to assess the overall pattern of activation. Because of technical difficulties during the scanning session of one subject in the Semantic Decision condition, data of five subjects could be included in this latter analysis. In all the analyses, significance levels were corrected for multiple comparisons using a cortex-based Bonferroni adjustment (Formisano *et al.*, 2002).

Results

Behavioral Results

Upon debriefing after the scanning session, all subjects reported they had been able to imagine hearing the words in the Imagery condition. Subjects performed the tasks without difficulty. Mean

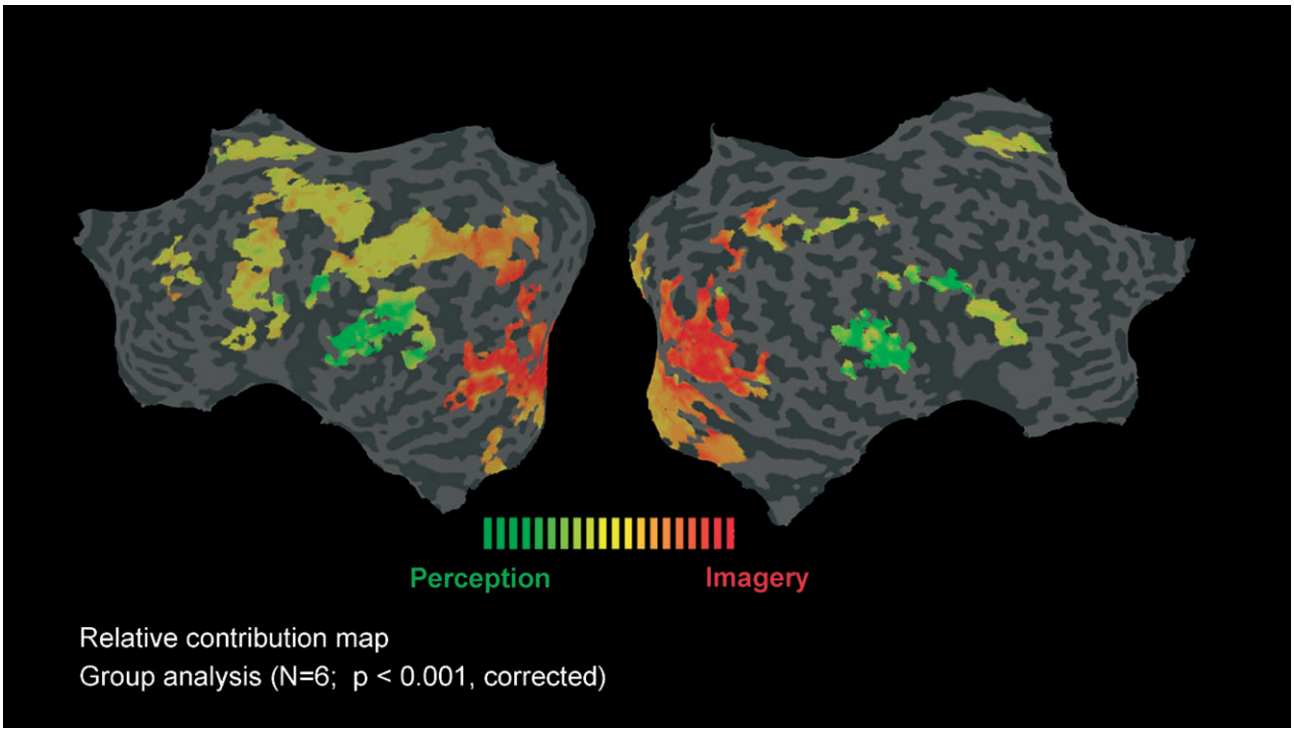

Figure 1. Color-coded group statistical maps of BOLD signal increase for task versus baseline. In this map, for each significantly activated ($P < 0.001$, corrected) voxel, the relative contribution (RC) of each condition is visualized with a red-yellow-green pseudocolour scale. An RC value of 1 (red) indicates that a voxel is solely responding during the Imagery condition, whereas an RC value of −1 (green) indicates that a voxel is solely responding during the Perception condition. An RC value of 0 (yellow) indicates that a voxel is responding during both the conditions. The maps are superimposed on a Talairach-normalized unfolded brain. From left to right: unfolded left hemisphere (frontal cortex at the left and occipital cortex at the right) and unfolded right hemisphere (frontal cortex at the right and occipital cortex at the left).

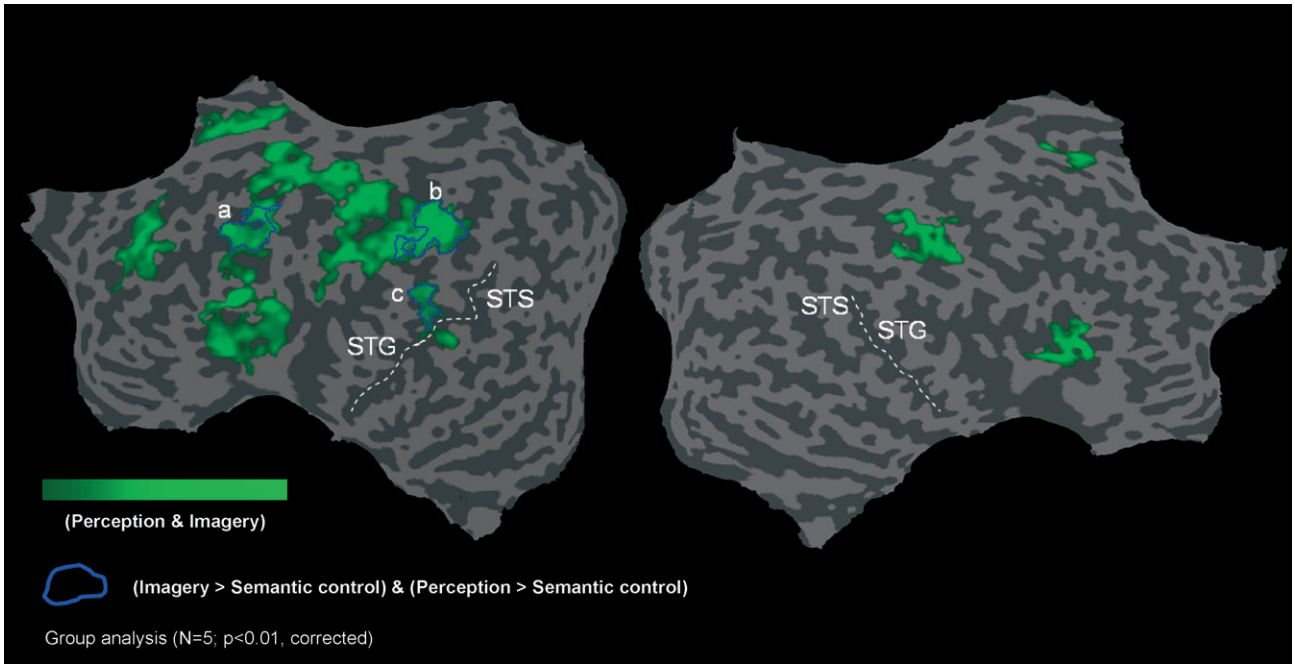

Figure 2. Conjunction analysis of perception and imagery conditions. Areas with a significantly larger BOLD response relative to the semantic decision condition are highlighted: (*a*) left precentral gyrus (−35, 4, 36), (*b*) left superior parietal lobule (−35, 40, 39) and (*c*) left STS (−52, −44, 16).

Table 1
Talairach coordinates of significantly activated regions (and number of subjects showing significant activation) for the three experimental conditions versus baseline

Region	Side	Perception					Imagery					Semantic Decision				
		x	*y*	*z*	*Z*-score	*n*/6	*x*	*y*	*z*	*Z*-score	*n*/6	*x*	*y*	*z*	*Z*-score	*n*/5
GTT/AC	L	−46	−18	14	6.2	6										
STS/STG	L	−54	−21	10	6.5	6	−43	−44	4	6.2	5					
GFd/SMA	L	−1	−1	55	7.0	6	−2	0	55	6.3	6	−6	−4	50	6.2	5
Pre-CG	L	−40	−20	52	7.2	6	−44	−10	47	6.4	6	−37	−27	53	5.3	4
Post-CG	L	−37	−31	55	6.6	5	−48	−27	47	7.0	6	−38	−32	46	5.3	5
IPS/SPL	L	−36	−42	43	6.1	6	−22	−67	34	6.0	6	−32	−55	40	5.3	4
IFG/Broca	L	−48	4	27	5.9	6	−58	2	33	5.4	5	−40	35	5	5.3	3
SFG	L	−27	45	29	5.4	6	−29	47	27	6.0	5	−4	44	34	5.5	3
Ins	L	−37	1	10	6.6	5	−32	0	11	6.6	5	−33	−2	15	5.3	4
GTT/AC	R	49	−16	5	7.6	6										
STS/STG	R	51	−23	7	6.0	6										
GFd/SMA	R	8	−1	41	6.2	5	7	6	36	5.5	5	4	20	42	5.7	5
Post-CG	R	39	−33	45	5.5	4	51	−23	45	5.5	4	42	−26	40	5.3	3
IPS/SPL	R						29	−59	40	5.5	4	41	−55	38	5.3	3
IFG	R	38	9	11	6.5	5										
Ins	R	37	8	11	6.0	5	38	9	11	5.6	4	35	13	16	5.6	3
OT cortex	L						−41	−61	−3	9.3	6					
OT cortex	R						33	−68	−5	8.5	6					
MOG	L						−18	−84	3	6.7	6	−16	−91	3	6.5	5
MOG	R						12	−80	3	7.3	6	22	−87	1	5.7	5
Amygdala	L											−15	−5	12	7.2	4

The position of each region is given as the Tailarach coordinates of the centre of mass of suprathreshold clusters ($P < 0.01$, corrected) of the group analysis. *Z*-score indicates the peak statistical value in the cluster. *n* indicates the number of subjects in which the activation was significant ($P < 0.05$, corrected) at individual level. GFd/SMA = gyrus frontalis medialis/supplementary motor area; Pre-CG = precentral gyrus; Post-CG = postcentral gyrus; IPS/SPL = intraparietal sulcus/superior parietal lobule; IFG = inferior frontal gyrus; STS/STG = superior temporal sulcus/superior temporal gyrus; MTG = middle temporal gyrus; GTT/AC = transverse temporal gyrus/auditory cortex; Ins = insula; OT cortex = occipitotemporal cortex; MOG = middle occipital gyrus.

percent correct responses was 92% in the imagery condition and 86% in the Perception condition. The difference in accuracy between the two tasks was not significant. Mean (± SE) reaction time was 1677 ± 171 ms in the Imagery condition and 2265 ± 168 ms for the Perception condition.

fMRI Results

Figure 1 illustrates the overall pattern of brain activation during the two main conditions of our experiment (Imagery and Perception). As expected, the Perception condition, which included the auditory presentation of words, yielded activation of primary and association auditory areas (green color in Fig. 1). Analogously, the Imagery condition, which included the visual display of words, yielded activation of primary and association visual areas (red color in Fig. 1). Both the conditions activated left SMA, pre- and postcentral gyrus, inferior frontal gyrus, insula, and right SMA and insula (yellow color in Fig. 1). It is important to note that only a subset of the left STS/STG seems to

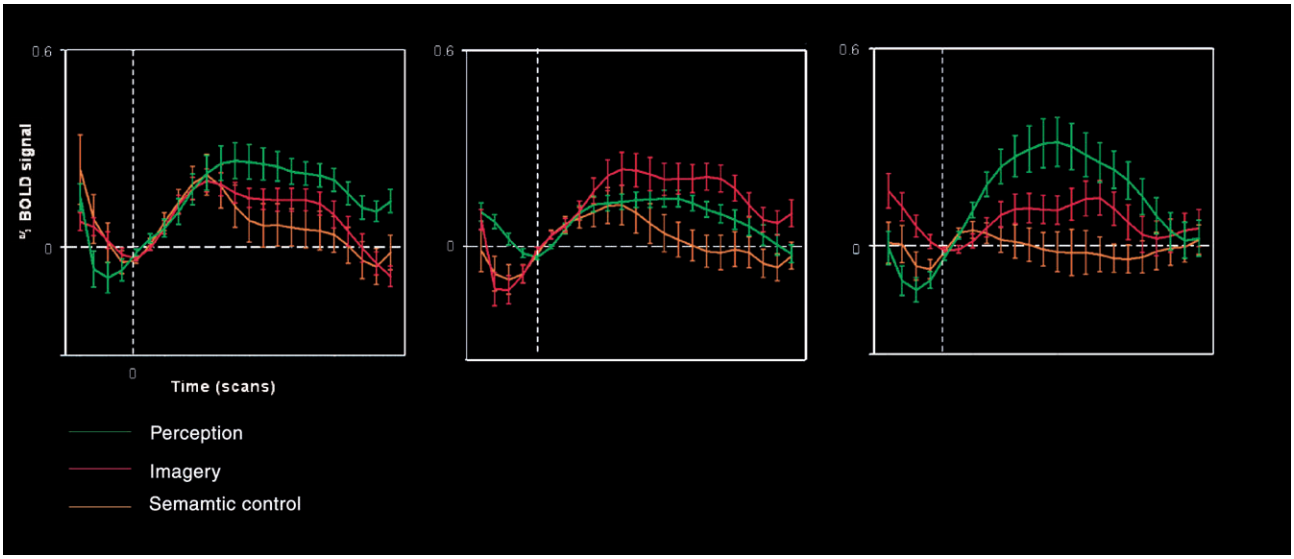

Figure 3. Average time-courses of the BOLD responses during the perception, imagery and semantic decision conditions in the (*a*) left precentral gyrus (−35, 4, 36), (*b*) left superior parietal lobule (−35, 40, 39) and (*c*) left STS (−52, −44, 16). The average time-courses were obtained by averaging over subjects the responses of the voxels belonging to the clusters highlighted in Figure 2.

Table 2
Talairach coordinates of significantly activated regions in the conjunction analyses

Region	Side	Perception and Imagery				(Perception > Semantic Decision) and (Imagery > Semantic Decision)			
		x	*y*	*z*	*Z*-score	*x*	*y*	*z*	*Z*-score
STS/STG	L	−47	−40	13	6.2	−52	−44	16	6.2
GFd/SMA	L	−2	−1	55	6.1				
Pre-CG	L	−45	−9	45	6.1	−35	−4	36	5.4
Post-CG	L	−43	−36	51	5.7				
IPS/SPL	L	−34	−44	40	6.2	−35	−40	39	5.3
IFG/Broca	L	−50	3	30	5.3				
SFG	L	−28	45	28	5.5				
Ins	L	−37	1	10	5.3				
GFd/SMA	R	8	5	40	5.3				
PostCG	R	47	−25	45	5.3				
Ins	R	40	7	11	5.3				

The position of each region is given as the Tailarach coordinates of the centre of mass of suprathreshold clusters ($P < 0.01$, corrected) of the group analysis. Z-score indicates the peak statistical value in the cluster. GFd/SMA = gyrus frontalis medialis/supplementary motor area; PreCG = precentral gyrus; PostCG = postcentral gyrus; IPS/SPL = intraparietal sulcus/superior parietal lobule; IFG = inferior frontal gyrus; STS/STG = superior temporal sulcus/superior temporal gyrus; Ins = insula.

respond to both the experimental conditions (see below). Table 1 shows anatomical locations and inter-subject reproducibility of significant clusters of activation during the imagery and perception conditions relative to the baseline fixation. Individual results confirmed that activation of these areas was robust across individuals.

The conjunction analysis of Perception and Imagery conditions confirmed involvement of SMA, precentral and postcentral gyrus, inferior frontal gyrus, insula, superior temporal sulcus, and intraparietal sulcus/superior parietal lobule for the left hemisphere, and of SMA, postcentral gyrus and insula for the right hemisphere. Thus, these areas responded to metrical stress evaluation of auditory perceived as well as imagined spoken words, irrespective of modality of stimulus presentation (green areas in Fig. 2). Regions that responded stronger to both the Perception and Imagery conditions relative to the Semantic Decision condition concerned left precentral gyrus, left postcentral gyrus/SPL and left STS (Fig. 2*a*,*b*,*c*, respectively). Figure 3 shows the BOLD response time courses for each of these areas, with a higher left precentral gyrus response of Perception relative to Imagery and Semantic Decision and a higher left postcentral gyrus/SPL response for Imagery relative to Perception and Semantic Decision. In the STS region, the Perception condition yielded the largest BOLD response after only a few scans. The BOLD response for the Imagery condition, although significantly smaller in magnitude than for the Perception condition, was significantly larger than for the Semantic Decision condition, which did not show any increase in BOLD response in this region.

Discussion

The results of this study show that areas in the left frontal and temporal lobe that have been implicated in the phonological analysis of spoken words are also activated during the generation and phonological analysis of imagined speech. This was accomplished with the use of a novel task that allowed the neural assessment of auditory–verbal imagery (internally generating and 'perceiving' imagined speech) in a behaviorally controlled design. Conditions were compared in which the subjects (i) actually heard spoken words or (ii) imagined hearing spoken words, and subsequently discriminated between weak-initial words and strong-initial words (metrical stress evaluation). Extensive activation was observed in language production areas in the left hemisphere: SMA, inferior frontal gyrus (Broca's area) and insula. These regions have typically been shown to be activated also in tasks of silent verbal fluency (Cuenod *et al.*, 1995; Ojemann *et al.*, 1998; Lurito *et al.*, 2000), and silent reading of words and pseudowords (Hagoort *et al.*, 1999). Consistent with previous language-related studies on phoneme monitoring, phoneme/syllable counting and word rhyming (for a review, Poldrack *et al.*, 1999), it is of interest to note the implication of the inferior frontal gyrus in phonological processing and not merely speech production (Wise *et al.*, 1999). The strong lateralization of activation associated with metrical

stress evaluation to the left hemisphere is in accordance with findings by Gandour and Baum (2001) in patients with unilateral lesions. In addition, this may imply that the present language task might be particularly suited for non-invasive determination of language localisation in clinical settings (e.g. pre-surgical planning for patients with epilepsy or brain neoplasms).

Three areas showed more activation in both metrical stress evaluation conditions compared with the semantic decision condition: the left precentral gyrus, left superior parietal lobule and left STS. As revealed by the time-course analyses, the semantic decision condition activated, albeit to a lesser extent, the precentral gyrus and superior parietal lobule. The gradual difference with the other conditions may be related to higher processing demands associated with these conditions, which may increase levels of motor preparation and involvement of attentional resources. Several studies of silent word generation, verbal working memory, imagining speech and speech perception have shown that cerebral areas thought to be devoted to motor aspects of speech planning and execution could also be activated even in the absence of overt speech (e.g. Yetkin *et al.*, 1995; Fiez *et al.*, 1996; Fadiga *et al.*, 2002; Shergill *et al.*, 2002; Watkins *et al.*, 2003). In the present study, this activation persists in the second conjunction analysis relative to the semantic condition which also involves motor response, suggesting that the activation is rather related to the covert rehearsal of the stimuli. Interestingly, the hemodynamic response time courses in these areas showed a stronger left precentral gyrus response for the perception condition relative to the others and a stronger left parietal response for the imagery condition. The stronger left precentral gyrus response for the perception condition might be due to a stronger rehearsal component for this condition: in contrast to the imagery condition, in which the phonological representation is internally generated in response to a visual input, in the perception condition the phonological representation is derived from the acoustic speech signal and its active rehearsal in working memory. Rehearsal of verbal material has been shown to activate the precentral gyrus (Paulesu *et al.*, 1993; Shergill *et al.*, 2002). In addition, the longer reaction times in the perception condition (due to the auditory modality of presentation) might contribute to increased levels of motor preparation and hence left precentral gyrus activation. The stronger activation of the left parietal area in the imagery condition might point to larger contribution of an imagery strategy in which subjects recover positional information of the visually presented words in determining whether the first or second syllable carried the stress. It is interesting in this regard, that Halpern and Zatorre (1999) also observed left parietal activation during auditory imagery (versus perception) of familiar melodies, in which a decision was required regarding the position of a tone in the tone sequence of the first line of a familiar song.

Crucially, however, the semantic decision condition did not activate the left STS/STG region, whereas the two stress evaluation conditions (perception and imagery) did. The semantic decision condition was identical to the imagery condition in terms of both the visual word input and the two-choice task configuration, but did not require phonological processing. Thus, the STS/STG response showed specificity for the task conditions involving phonological processing. Indeed, this activation of speech perception areas during speech imagery, is consistent with our prediction that auditory–verbal imagery would activate brain areas involved in phonological decoding. Whereas a large part of cortex in STS/STG was activated only during the auditory presentation condition, a small region located in the posterior left STS/STG was activated in both conditions. The posterior left STG/STS has been extensively documented to be involved in speech perception (Hickok and Poepel, 2000; Scott and Johnsrude, 2003). Although activation of the STS may not be speech specific, there is ample evidence that activation of the STS is involved in the analysis of the complex acoustic characteristics of the human voice and, more specifically, the phonological analysis of speech (Belin *et al.*, 2000; Jäncke *et al.*, 2002). Moreover, studies of patients with speech comprehension deficits reveal consistent damage to the posterior STS (Anderson *et al.*, 1999). Functional neuroimaging studies of phonetic perception have also shown involvement of the posterior STG and temporoparietal junction (Petersen *et al.*, 1988; Zatorre *et al.*, 1996). Finally, consistent with our finding of posterior STG/STS involvement in the processing of sound-based representations in the absence of auditory input, Halpern and Zatorre (1999) observed STG activation in a PET study of auditory imagery for familiar melodies.

Lurito *et al.* (2000) compared fMRI activation during a verbal fluency task with activation during a rhyming task. The latter task might be hypothesized to depend on auditory–verbal imagery, as visually presented words have to be matched on their sound characteristics, which are not always apparent from the visual form of the word. Indeed, behavioral studies of auditory–verbal imagery have included similar rhyming tasks. Interestingly, Lurito *et al.* (2000) report activation of Broca's area and also of the left superior temporal gyrus, including the STS. However, in contrast to the present study, Lurito *et al.* (2000) did not include a perceptual condition, nor a control condition using identical words as in the inner speech condition but requiring a different decision (in order to control for areas that are involved in processing and visually presented word and subsequent response preparation and execution).

Auditory–verbal imagery has been conceptualized to depend on verbal working memory processing (Baddeley and Logie, 1992), and as such, we expected brain areas involved in verbal working memory to be activated by our task. Particularly, phonological recoding has been suggested to activate left prefrontal areas BA 44 and 45, silent articulation to be mediated by SMA (and possibly lateral cerebellum), phonological decoding to be by posterior STG/STS or the temporo-parietal junction (Wernicke's area), and the phonological store to depend on inferior parietal areas/supramarginal gyrus (Paulesu *et al.*, 1993; Fiez *et al.*, 1996; Thierry *et al.*, 1999). Indeed, activation of areas that have been implied in phonological recoding, silent articulation and phonological decoding was observed in the present study. In contrast, no activation was observed of the left supramarginal gyrus, which has been implicated in phonological storage in several studies (Paulesu *et al.*, 1993; Henson *et al.*, 2000). However, it is important to note that the present task did not require subjects to store the information across a delay, thus not posing significant demands on storage mechanisms. Recently, Hickok *et al.* (2003) observed robust responses to both a sensory (listening to speech) and a motor (subvocal rehearsal) component of a verbal working memory task in a left lateralized region in the posterior Sylvian fissure at the parietal-temporal boundary. This finding is consistent with the activation of posterior STG/STS in the present study.

Notably, activation of the left STG/STS in tasks involving speech stimuli has been hypothesized to underlie a verbal

monitoring system (McGuire *et al.*, 1996; Shergill *et al.*, 2002). Indeed, verbal monitoring can be considered an important component of the metrical stress evaluation task we used, as the subjects are explicitly required to focus on phonological characteristics of their 'inner speech'. It will prove to be hard to design an auditory–verbal imagery task without a verbal monitoring component, although implicit processing paradigms might be helpful in this regard. Disentangling such components and their neural basis may be of relevance for understanding the neurogenesis of auditory–verbal hallucinations in schizophrenia (Seal *et al.*, 2004).

In conclusion, we used a behaviorally controlled paradigm to investigate the neural basis of auditory–verbal mental imagery in the absence of auditory stimulation. Speech perception has been shown to result from multiple, complementary representations of the input, which operate in both acoustic–phonetic feature-based (STS/STG) and articulatory–gestural (Broca, SMA) domains (Scott and Johnsrude, 2003). Our findings indicate that this extends to speech imagery, and suggest a common neural basis of speech perception and imagery in the left frontal and temporal lobe. This result parallels earlier findings of a convergence of visuospatial imagery and perception in the posterior parietal cortex (Trojano *et al.*, 2000; Formisano *et al.*, 2002).

Notes

The authors wish to thank Paul Gaalman for technical assistance during MR scanning. A.A. was supported by a VernieuwingsImpuls grant (no. 016.026.027) from the Netherlands Organization for Scientific Research (NWO).

Address correspondence to André Aleman, Department of Psychiatry, A01.126, University Medical Center, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands. Email: a.aleman@azu.nl.

References

Aleman A, Schutter DLJG, Ramsey NF, van Honk J, Kessels RPC, Hoogduin JH, Postma A, Kahn RS, De Haan EHF (2002) Functional neuroanatomy of top-down visuospatial processing in the human brain: evidence from rTMS. Cogn Brain Res 14:300–302.

Aleman A, Böcker KBE, Hijman R, de Haan EHF, Kahn RS (2003) Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. Schizophr Res 64:175–185.

Anderson JM, Gilmore R, Roper S, Crosson B, Bauer RM, Nadeau S, Beversdorf DQ, Cibula J, Rogish M 3rd, Kortencamp S, Hughes JD, Gonzalez Rothi LJ, Heilman KM (1999) Conduction aphasia and the arcuate fasciculus: a reexamination of the Wernicke–Geschwind model. Brain Lang 70:1–12.

Baddeley AD, Hitch GJ (1974) Working memory. In: Recent advances in learning and motivation (Bower G, ed.), vol. VIII, pp. 647–667. New York: Academic Press.

Baddeley A, Logie R (1992) Auditory imagery and working memory. In: Auditory imagery (Reisberg D, ed.), pp. 179–198. Hillsdale, NJ: Lawrence Erlbaum.

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

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.

Böcker KBE, Bastiaansen MCM, Vroomen J, Brunia CHM, De Gelder B (1999) An ERP correlate of metrical stress in spoken word recognition. Psychophysiology 36:706–720.

Cuenod CA, Bookheimer SY, Hertz-Pannier L, Zeffiro TA, Theodore WH, Le Bihan D (1995) Functional MRI during word generation, using conventional equipment: a potential tool for language localization in the clinical environment. Neurology 45:1821–1827.

Cutler A, Norris DG (1988) The role of strong syllables in segmentation for lexical access. J Exp Psychol Hum Percept Perform 14: 113–121.

Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002) Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur J Neurosci 15:399–402.

Fiez JA, Raife EA, Balota DA, Schwartz JP, Raichle ME, Petersen SE (1996) A positron emission tomography study of short-term maintenance of verbal information. J Neurosci 16:808–822.

Formisano E, Linden DE, Di Salle F, Trojano L, Esposito F, Sack AT, Grossi D, Zanella FE, Goebel R (2002) Tracking the mind's image in the brain. I. Time-resolved fMRI during visuospatial mental imagery. Neuron 35:185–194.

Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turener R (1995) Analysis of fMRI time-series revisited. Neuroimage 2:45–53.

Gandour J, Baum SR (2001) Production of stress retraction by left- and right-hemisphere-damaged patients. Brain Lang 79:482–494.

Gazzaniga MS, ed. (2000) The new cognitive neurosciences. Cambridge, MA: MIT Press.

Halpern AR, Zatorre RJ (1999) When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. Cereb Cortex 9:697–704.

Hagoort P, Indefrey P, Brown C, Herzog H, Steinmetz H, Seitz RJ (1999) The neural circuitry involved in the reading of German words and pseudowords: a PET study. J Cogn Neurosci 11:383–398.

Henson RN, Burgess N, Frith CD (2000) Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. Neuropsychologia 38:426–440.

Hermans D, de Houwer J (1994) Affective and subjective familiarity ratings of 740 Dutch words. Psychol Belg 34:115–139.

Hickok G, Poeppel D (2000) Towards a functional neuroanatomy of speech perception. Trends Cogn Sci 4:131–138.

Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003) Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area spt. J Cogn Neurosci 15:673–682.

Intons-Peterson MJ (1992) Components of auditory imagery. In: Auditory imagery (Reisberg D, ed.), pp. 45–72. Hillsdale, NJ: Lawrence Erlbaum.

Jäncke L, Wüstenberg T, Scheich H, Heinze H-J (2002) Phonetic perception and the temporal cortex. Neuroimage 15:733–746.

Kosslyn SM, Pascual-Leone A, Felician O, Camposano S, Keenan JP, Thompson WL, Ganis G, Sukel KE, Alpert NM (1999) The role of area 17 in visual imagery: convergent evidence from PET and rTMS. Science 284:167–170.

Levelt WJM (1989) Speaking: from intention to action. Cambridge, MA: MIT Press.

Lurito JT, Kareken DA, Lowe MJ, Chen SH, Mathews VP (2000) Comparison of rhyming and word generation with FMRI. Hum Brain Mapp 10:99–106.

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

Mellet E, Petit L, Mazoyer B, Denis M, Tzourio N (1998) Reopening the mental imagery debate: lessons from functional anatomy. Neuroimage 8:129–139.

Noppeney U, Price CJ (2002) A PET study of stimulus- and task-induced semantic processing. Neuroimage 15:9279–35.

Ojemann JG, Buckner RL, Akbudak E, Snyder AZ, Ollinger JM, McKinstry RC, Rosen BR, Petersen SE, Raichle ME, Conturo TE (1998) Functional MRI studies of word-stem completion: reliability across laboratories and comparison to blood flow imaging with PET. Hum Brain Mapp 6:203–215.

Paulesu E, Frith CD, Frackowiak RSJ (1993) The neural correlates of the verbal components of working memory. Nature 362:342–344.

Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1988) Positron emission tomography studies of the functional anatomy of single word processing. Nature 331:585–589.

Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage 10:15–35.

Reisberg D, ed. (1992) Auditory imagery. Hillsdale (NJ): Lawrence Erlbaum Assoc.
Reisberg D, Smith JD, Baxter AD, Sonenshine M (1989) Enacted auditory images are ambigous; pure auditory images are not. Q J Exp Psychol 41A:619–641.
Sack AT, Sperling JM, Prvulovic D, Formisano E, Goebel R, Di Salle F, Dierks T, Linden DE (2002) Tracking the mind's image in the brain. II. Transcranial magnetic stimulation reveals parietal asymmetry in visuospatial imagery. Neuron 35:195–204.
Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. Trends Neurosci 26: 100–107.
Seal ML, Aleman A, McGuire PK (2004) Compelling imagery, unanticipated speech and deceptive memory: neurocognitive models of auditory verbal hallucinations in schizophrenia. Cogn Neuropsychiatry 9:43–72.
Shergill SS, Bullmore ET, Brammer MJ, Williams SC, Murray RM, McGuire PK (2001) A functional study of auditory verbal imagery. Psychol Med 31:241–253.
Shergill SS, Brammer MJ, Fukuda R, Bullmore E, Amaro E Jr, Murray RM, McGuire PK (2002) Modulation of activity in temporal cortex during generation of inner speech. Hum Brain Mapp 16: 219–227.
Smith JD, Wilson M, Reisberg D (1995) The role of subvocalization in auditory imagery. Neuropsychologia 33:1433–1454.
Talairach J, Tournoux P (1988) A coplanar stereotactic atlas of the human brain. Stuttgart: Thieme Verlag.
Thierry G, Boulanouar K, Kherif F, Ranjeva JP, Demonet JF (1999) Temporal sorting of neural components underlying phonological processing. Neuroreport 10:2599–2603.
Trojano L, Grossi D, Linden DE, Formisano E, Hacker H, Zanella FE, Goebel R, Di Salle F (2000) Matching two imagined clocks: the functional anatomy of spatial analysis in the absence of visual stimulation. Cereb Cortex 10:473–481.
Yetkin FZ, Hammeke TA, Swanson SJ, Morris GL, Mueller WM, McAuliffe TL, Haughton VM (1995) A comparison of functional mr activation patterns during silent and audible language tasks. AJNR Am J Neuroradiol 16:1087–1092.
Watkins KE, Strafelle AP, Paus T (2003) Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 41:989–994.
Wise RJ, Greene J, Buchel C, Scott SK (1999) Brain regions involved in articulation. Lancet 353:1057–1061.
Zatorre RJ, Meyer E, Gjedde A, Evans AC (1996) PET studies of phonetic processing of speech: review, replication, and reanalysis. Cereb Cortex 6:21–30.