TITLE

Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception

ABBREVIATED TITLE

Error detection in speech production and perception

Hanna S. Gauvin[1], Wouter De Baene[1, 2], Marcel Brass[1] and Robert J. Hartsuiker[1]

1 Department of Experimental Psychology, Ghent University, 9000 Ghent, Belgium

2 Department of Cognitive Neuropsychology, Tilburg University, 5000 LE Tilburg, The Netherlands

Correspondence:

Hanna S. Gauvin

Department of Experimental Psychology

Ghent University

Henri Dunantlaan 2

9000 Ghent, Belgium

Hanna.Gauvin@ugent.be

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgements: This research was supported by the Fund for Scientific Research Flanders (FWO), project “Internal self-monitoring: speech perception or forward models?” no. G.0335.11N

Abstract

To minimize the number of errors in speech, and thereby facilitate communication, speech is monitored before articulation. It is, however, unclear at which level during speech production monitoring takes place, and what mechanisms are used to detect and correct errors. The present study investigated whether internal verbal monitoring takes place through the speech perception system, as proposed by perception-based theories of speech monitoring, or whether mechanisms independent of perception are applied, as proposed by production-based theories of speech monitoring. With the use of fMRI during a tongue twister task we observed that error detection in internal speech during noise-masked overt speech production and error detection in speech perception both recruit the same neural network, which includes pre-supplementary motor area (pre-SMA), dorsal anterior cingulate cortex (dACC), anterior insula (AI), and inferior frontal gyrus (IFG). Although production and perception recruit similar areas, as proposed by perception-based accounts, we did not find activation in superior temporal areas (which are typically associated with speech perception) during internal speech monitoring in speech production as hypothesized by these accounts. On the contrary, results are highly compatible with a domain general approach to speech monitoring, by which internal speech monitoring takes place through detection of conflict between response options, which is subsequently resolved by a domain general executive center (e.g., the ACC).

48 Keywords: self-monitoring, speech production, speech perception, conflict
49 monitoring, verbal monitoring, fMRI

1. Introduction

In the domain of language production there is consensus about the existence of an internal speech monitoring system, which monitors speech before production, in addition to an external monitoring system (i.e., hearing one's own speech). Evidence for an internal monitoring system comes from research showing extremely fast self-corrections for which the external monitoring system would simply be too slow (Levelt, 1989; Blackmer & Mitton, 1991; Hartsuiker & Kolk 2001), the report of errors when silently performing a speech task (Oppenheim & Dell, 2008), and the report of errors when overt speech is masked by loud noise (Lackner & Tuller, 1979; Postma & Kolk, 1992). However, there is currently no consensus on the underlying nature of such an internal speech error monitoring mechanism. In a review of verbal monitoring models, Postma (2000) discusses eleven possible locations during the process of speaking at which monitoring has been proposed to take place. Most of the proposed models are directed at monitoring internal speech. Additionally, external speech can be monitored via perception of the speech and via perception of the articulators and muscles (proprioceptive feedback) (Abbs & Gracco, 1983; Abbs et al., 1984; Siegenthaler & Hochberg, 1965).

Presently there are roughly three classes of theories on monitoring internal speech: perception-based accounts (Perceptual Loop Theory, Hartsuiker & Kolk, 2001; Levelt, 1989; Indefrey, 2011), production-based accounts (Local Monitors, Laver 1980; Conflict Monitors, Nozari et al., 2011), and forward modeling accounts (e.g., Hickok, 2012; Pickering & Garrod, 2013; Tourville & Guenther, 2011). In the current study we investigated whether the neural structures involved in verbal monitoring lend support for any of these three classes of theories. Below we will first

outline the theories in more detail, including their neuro-anatomical hypotheses, after which we outline how the fMRI data can be used to dissociate these theories.

Perception-based theories assume that internal speech monitoring takes place in the speech perception system, during both production and perception. During production, the phonetic plan is sent directly to the perception system (i.e., before articulation) for internal monitoring. Essentially the same monitoring mechanism would be used for both internal and external monitoring according to the perception-based theories of monitoring, with internal monitoring using part of the external monitoring route. Monitoring your own internal and external speech and monitoring someone else's speech, all rely on the perception system, for which the superior temporal gyrus is the main neural substrate (e.g. Price, 2012).

Production-based theories do not necessarily assume the same monitoring system for production and perception, and assume that internal monitoring during production takes place independently of speech perception systems. A recently proposed production monitoring account uses conflict within the production system as a basis for monitoring (Nozari et al., 2011). Analogous to domain-general theories of error detection (e.g., Botvinick et al., 2001; Yeung et al., 2004), monitoring rather takes place through detection of conflict between response options, which is subsequently resolved by a domain-general cognitive control unit located in the Anterior Cingulate Cortex (ACC). Support for these domain-general theories comes from the vast, and increasing, body of literature in which in response to conflict an ERN component is found in EEG studies and ACC activation in fMRI studies. Source localization has traced the ERN component to originate from the ACC (e.g. Van Veen

& Carter, 2002; Herrmann et al., 2004). The ERN component and ACC activation are similar across cognitive domains, suggesting the existence of a domain general conflict response. For a more detailed overview of the conflict monitoring literature in relation to the ERN and ACC, see for instance Van Veen and Carter (2006).

Another type of production-based monitoring that has been proposed is monitoring via forward models. Forward modeling accounts of speech monitoring assume that during production a prediction, or forward model, of the expected outcome is made. The actual outcome is compared to the predicted outcome, and if a mismatch between these two is detected, a corrective signal arises. Forward model theories of speech production are supported by the observation of auditory response suppression during speech production; based on the prediction of the sensory feedback of the upcoming event, the sensory cortex is inhibited. When the sensory feedback is not in accordance with the prediction, an increase in activation is observed, which might function as a corrective signal (Curio et al., 2000; Heinks-Maldonado et al., 2005, 2006; Numminen et al., 1999; Eliades & Wang, 2003, 2005). Direct evidence in support of forward models during speech production comes from a series of MEG experiments, showing context dependent activation changes in the auditory cortex in response to imagined speech production (so in the absence of actual auditory stimulation) at a same time frame as observed after normal speech production (Tian & Poeppel, 2010, 2013). Most forward model theories rely on sensory feedback for monitoring. Consequently, internal speech monitoring, which is investigated in the current study, is outside the scope of these theories. However, Pickering and Garrod's forward model theory (2013, 2014) does make predictions

about monitoring in internal speech production and speech perception. According to their theory, both during production and perception, predictions are made and compared to the actual utterance. These comparisons are made in a comparator, which is a speech-modality (production / perception) independent system. So a difference between correct and incorrect sentences is expected to lead to differences in activation in the comparator, which is separate of the perception system. However, no anatomical predictions are made with respect to this comparator in Pickering and Garrod's forward model theory.

Because production- and perception-based monitoring theories make distinct predictions about the functional neuroanatomy of speech monitoring, fMRI is a useful tool to distinguish between these competing theories. Perception-based monitoring accounts assume that, as the bilateral superior temporal gyri (STG) are involved in monitoring external speech, internal speech must be monitored via (a subpart of) the same neuronal structures (Indefrey, 2011). So if (pre-verbal) internal monitoring is perception-based, we expect superior temporal gyrus (STG) activation for error detection in both production and perception, even when auditory feedback is unavailable during production. If monitoring is production-based, however, we expect to find error monitoring independently of perceptual areas during production. Production-based monitoring accounts predict activation in areas associated with subcomponents of the production process, as well as domain-general areas associated with conflict monitoring in the medial frontal areas, such as the ACC. In the experiment reported below, we compared internal speech monitoring during production with external speech monitoring during perception, in order to investigate whether all monitoring is indeed performed by the perceptual

system (as proposed in perception-based theories of monitoring such as the perceptual loop theory by Levelt, 1989; Indefrey, 2011) or whether monitoring is performed independently of the perceptual system (as proposed by production-based theories of monitoring such as the conflict monitoring theory by Nozari et al., 2011, and the forward model theory by Pickering & Garrod, 2013, 2014).

At this moment there are no publications that describe the neuronal structures involved in internal and external verbal monitoring and their differences. Only few studies have applied fMRI to investigate internal speech monitoring and none have compared monitoring in speech production with monitoring in speech perception. Monitoring of external speech has been investigated predominantly by manipulating acoustic feedback in the dimensions of frequency or time (McGuire et al, 1996; Hirano et al, 1997; Fu et al, 2006; Christoffels et al, 2007; Tourville et al, 2008). Perception of altered feedback led to increased activation in the superior temporal lobe compared to unaltered feedback. Note that these are externally induced 'errors'; the participant made no error during production, but via manipulation of the feedback the perception of the speech is changed. There is only one published fMRI study on error production in language processing that targets errors made by the producer herself (Abel et al, 2009). In this experiment, participants overtly named pictures during scanning, and resulting activations during correct production, incorrect production, and a rest baseline were compared. This study found increased activations during error production in the ACC, prefrontal and premotor regions, basal ganglia, thalamus, SMA and precentral gyrus. This experiment had, however, several limitations: few errors were made, and the

reported errors were not very naturalistic as some were merely errors against the instructed label for each picture (e.g., call this picture ‘flower’ and the participant responds with ‘sunflower’). There have been no published studies, to the best of our knowledge, with a direct comparison between fMRI data of speech error detection in production and perception.

The current study therefore aims to investigate the neural underpinnings of internal verbal monitoring and external verbal monitoring. These data can be used to distinguish between several highly influential theoretical models of verbal monitoring, as these theories have neuroanatomically specific predictions. Below we outline the experimental setup, and we discuss the hypotheses the monitoring models make with respect to the neuronal functional data.

Participants performed a tongue twister task in which they repeated tongue twister sentences, or listened to a recording of a tongue twister repetition, after which they judged the repetition on correctness. The percentage of errors in the perception condition was matched to the number of errors in the production condition, to allow for a comparison of the areas involved in error detection in both modalities. In order to test the involvement of the speech perception system in internal speech monitoring during production, normal feedback was precluded, as auditory feedback would necessarily involve external monitoring via the speech perception system. Perceptual-based monitoring is only supported if we find a role for the perception system in internal speech monitoring.

The theories of internal verbal monitoring make the following predictions with respect to the neural structures involved in verbal monitoring; perception-based monitoring assumes a major role for the auditory perceptual system, located

191 in the bilateral STG, during both internal and external verbal monitoring. Production-
192 based monitoring assumes involvement of a domain general monitoring mechanism
193 located in the ACC, and crucially it assumes no role for the auditory perceptual
194 system during internal verbal monitoring.
195
196 2. Materials and Methods
197
198 2.1 Participants
199 Twenty-four participants were recruited from Ghent University, of which 3
200 were discarded: one due to excessive motion, one because of too many errors
201 (>80%) and one due to too few errors (<10%). Final analyses included 21 participants
202 (15 females, 6 males; mean age: 21, ranging from 19 to 30). All reported to be native
203 speakers of Dutch, have no dyslexia or other speech or language impairments, no
204 hearing problems, and normal or corrected-to-normal vision. No subject had a
205 history of neurological, psychiatric, or major medical disorder as assessed by a pre-
206 scanning questionnaire. All subjects were right handed as assessed by the Edinburgh
207 Handedness inventory (Oldfield, 1971) *(n*=21, EHI score *M*=90.4, *SD*=15.8, range =
208 41.2 to 100, mode = 100). A monetary reward was received for participation. The
209 study was approved by the local ethical committee of Ghent University's Medical
210 Department and was conducted in accordance to the Declaration of Helsinki.
211
212 2.2. Stimulus material and task design
213 Stimuli were selected on the basis of a pilot study in which 56 tongue twister
214 sentences were tested. For the production condition sentences were selected that

elicited linguistic errors (phonological slips, semantic substitutions, and syntactic errors). From these, we selected 17 sentences with high error production rates (30% - 60% of repetitions contained an error). An additional 5 sentences were selected that were relatively easy (10% - 25% of repetitions contained an error), in order to prevent discouragement among the participants. Each sentence was presented 3 times per condition. We used tongue twister sentences of the type 'A proper copper coffee pot', and 'How can a clam cram in a clean cream can?'. A full overview of the tongue twister sentences is provided in Appendix A. The sentences were selected on the basis of the errors they elicited, and were not matched for frequency or length[1]. Audio files for the instruction phase were created in which these sentences were clearly pronounced at a normal speech rate by a male native speaker of Dutch. These audio files were presented together with a visual presentation of the tongue twister at the beginning of each trial.

For the perception condition 22 different tongue twister sentences were selected in order to decrease repetition effects and minimize attention loss. As in the production condition, each sentence was presented 3 times during the perception condition. Actual recordings of 4 female participants producing the sentences in the pilot study, correctly and incorrectly, were used as auditory stimuli. The errors selected for the perception condition highly resembled those produced in the production condition, and were all linguistic errors. Pitch was adjusted (increased with 50 or 20 Hz) for 3 of the 4 participants to facilitate auditory perception in the

[1] Note that very long sentences that were difficult to remember were not selected from the pilot study. Only sentences that elicited linguistic errors were selected.

scanner. Experiments were created in E-prime 2.0 (Psychology Software Tools, Inc, Pittsburg, PA).

Before entering the scanner participants were briefed on the task. After entering the scanner the participants again received instructions on the production task, followed by a familiarization phase and successively the actual experiment. Participants were instructed to speak normally, while keeping their heads fixed. To minimize movements, foam pads were placed between the head and head coil. Once the participant was set up to enter the scanner bore, the participant was asked to speak, and once again the experimenters stressed to the participants to speak normally and avoid any head movements, as motion artifacts are often observed with speech production during acquisition (see below). During the production condition the experimenter scored the number of incorrectly produced sentences, which allowed for an error percentage match with the perception condition. After completing the production condition, participants received instructions for the perception condition, followed by a familiarization phase and consecutively the perception condition of the tongue twister task. The total duration of the experiment was approximately 45 minutes.

2.2.1 Production Condition

Each trial consisted of a visual presentation of the target sentence with a simultaneous auditory presentation, followed by a blank screen and after 200 ms a repetition of the auditory presentation. After a pause of 250 ms a visual cue was presented (*) to signal to the participant to start producing the target sentence. After producing the sentence the participant pushed a button to indicate whether

the sentence was correct (right hand) or incorrect (left hand). From cue onset until a correctness judgment was made, after which the cue disappeared, the participant heard a white noise at maximum volume over the headphones to mask auditory feedback. An illustration of this task is provided in Figure 1. After a familiarization phase of 3 trials, three target blocks were presented that each consisted of the 22 tongue twister sentences in random order. Between trials a varying ISI of between 1250 and 5500 ms occurred (mean 2867 ms).

2.2.2 Perception Condition

In the perception trials the participants were presented with a visual presentation of the target sentence with simultaneous auditory presentation, exactly as in the production condition. After a pause of 200 ms the participants heard a recording of a person producing the sentence. The participant pushed a button to indicate whether the sentence was repeated correctly (right hand) or incorrectly (left hand). An illustration of this task is provided in Figure 1. After a familiarization phase of 3 trials, three target blocks were presented that each consisted of the 22 tongue twister sentences in random order. Between trials a varying ISI of between 1250 and 5500 ms occurred (mean 2867 ms), similar to the production condition. We constructed 8 versions of the perception condition with different error rates, ranging from 10% to 45% errors (with 5% intervals), to approximate the number of errors produced in the production condition.

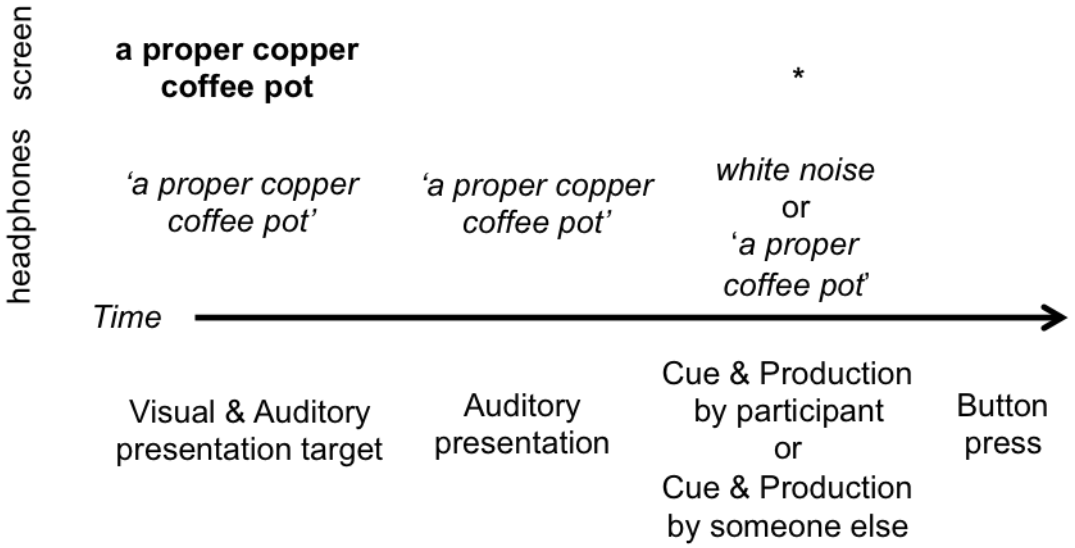

Figure 1. *Overview of the tongue twister task. The production and perception condition only differ in the production part of the sequence; in the production condition a white noise is presented over headphones during which the participant produces the sentence, while in the perception condition a pre-recorded production of the tongue twister sentence is played over headphones.*

2.3 Scanning procedure

Images were collected with a 3T Magnetom Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany), using a standard 32-channel radio-frequency head coil. A 3D high-resolution anatomical image of the whole brain was acquired first, for co-registration with the functional images using a T1-weighted 3D MPRAGE sequence (TR = 2530 ms, TE = 2.58 ms, TI = 1100 ms, acquisition matrix = 256 × 256 × 176, sagittal FOV = 220 mm, flip angle = 7°, voxel size = .90 × .86 × .86 mm3 (resized to 1 × 1 × 1 mm3)). Whole brain functional images were collected using a T2*-weighted EPI sequence, sensitive to BOLD contrast (TR= 2000 ms, TE=28 ms, image matrix=64×64, FOV=224 mm, flip angle = 80°, slice thickness = 3 mm,

298 distance factor = 17%, voxel size 3.5 × 3.5 × 3.51 mm^3, 34 axial slices). Specific care
299 was taken to ensure that frontal areas and (near) complete cerebellum were
300 included in the imaging volume. A varying number of images were acquired per run
301 due to the self-paced ending of trials. In the production condition the number of
302 images per run ranged from 450-528, in the perception condition the number of
303 images per run ranged from 334-382.

304 Participants went head first and supine into the magnetic bore. They were
305 instructed to speak normally but to avoid movements of their heads in order to
306 avoid motion artifacts. Foam pads were placed between the head and head coil to
307 minimize movement. Auditory stimuli were presented through MR-compatible
308 headphones with noise-cancellation (OptoACTIVE). An audio recording of the
309 participant's response was made with an fMRI compatible microphone (OptoACTIVE
310 FOMRI-III) attached to the headset, which was used to verify the correctness of the
311 produced sentence. At debriefing participants reported that during production they
312 were unable to hear themselves speak, confirming that the noise masking of
313 auditory feedback was successful.

314 While it is generally assumed that overt speech in the scanner will cause large
315 motion and signal artifacts (see Gracco, Tremblay & Pike, 2005 for an overview) we
316 did not find this to be the case in our specific set-up. Instead of using a special
317 scanning procedure (e.g. Eden et al., 1999; Huang et al., 2001; Menenti et al., 2011)
318 or limit volume acquisition to the time interval after speech production, we applied a
319 common acquisition procedure. Nevertheless, motions were well within the
320 boundaries of acceptability (no movement in any direction exceeding the voxel

dimensions of 3.5 mm), and no signal artifacts were found. Of the total group only data of one participant had to be discarded due to excessive motion artifacts.

2.4 Data analysis

2.4.1 fMRI data pre-processing

Data processing and analyses were performed using Matlab and SPM8 software (Wellcome Department of Cognitive Neurology, London, UK). The first nine scans of all EPI series were excluded from the analysis to minimize T1 relaxation artifacts and to allow for an optimization of the noise-cancellation. Data processing started with slice time correction and realignment of the EPI datasets. A mean image for all EPI volumes was created, to which individual volumes were spatially realigned by rigid body transformation. The high-resolution structural image was co-registered with the mean image of the EPI series. The structural image was normalized to the Montreal Neurological Institute (MNI) template. The normalization parameters were then applied to the EPI images to ensure an anatomically informed normalization. Motion parameters were estimated for each session separately. A commonly applied filter of 8 mm FWHM (full-width at half maximum) was used. The time series data at each voxel were processed using a high-pass filter with a cut-off of 128 s to remove low-frequency drifts.

2.4.2 General GLM analyses

The subject-level statistical analyses were performed using the general linear model (GLM). All events of interest were time-locked to the correctness judgments. We time-locked to judgments rather than to speech errors themselves for several

345 reasons. First, it was not uncommon for participants to produce multiple errors per
346 sentence. In this case it is unclear which error the activation needs to be time-locked
347 to. Second, there presumably is high variation in timing between the production of
348 an error and the detection of that error (e.g. Hartsuiker & Kolk, 2001). So time-
349 locking to the production is still not time-locking to the error detection. And as the
350 BOLD response is quite slow and broad (peaks at 5-6 seconds after stimulus onset
351 and declines slowly until about 10 seconds after stimulus onset), time-locking to the
352 correctness judgment will still capture relevant activations. For this analysis the
353 events of interest were Correct trials (where the sentence production was correct)
354 and Incorrect trials (where the repetition contained an error). Trials where the
355 participant had given an incorrect judgment formed a separate regressor of no
356 interest (data loss: 16% in the production condition, 19% in the perception
357 condition). Vectors containing the event onsets were convolved with the canonical
358 hemodynamic response function (HRF) to form the main regressors in the design
359 matrix (the regression model). The vectors were also convolved with the temporal
360 derivatives and the resulting vectors were entered into the model. In the model, we
361 also included regressors to account for variance associated with head motion. The
362 statistical parameter estimates were computed separately for each voxel for all
363 columns in the design matrix. Separately for the production and perception
364 condition, one main contrast was calculated for each single subject: erroneous trials
365 vs. correct trials. These contrasts from the single subject analyses were submitted to
366 a factorial design with condition (production vs. perception) as factor.

Only results significant at the familywise peak-level threshold of $p < .05$ are reported. The resulting maps were overlaid onto a structural image of a standard MNI brain and the coordinates reported correspond to the MNI coordinate system.

2.4.3 Region of interest analysis

To specifically test the involvement of the STS/STG in verbal monitoring, a region of interest (ROI) analysis was performed for brain regions in the STS/STG that were previously identified to be involved in verbal monitoring. For ROI analysis spheres with a radius of 6 mm were created at the peaks of activation clusters with the use of MarsBar tool for SPM (http://marsbar.sourceforge.net/). Resulting percent signal changes were analyzed in a 2 x 2 (Condition and Accuracy) repeated measures ANOVA.

3. Results

3.1 Behavioral data

During scanning of the production trials, the repetitions of the tongue twisters were recorded. The experimenter later checked these sound files for correctness of production and judgment. Only the items in which the participant had correctly identified his or her performance were included in the analysis; the incorrectly judged items were discarded from all analyses. Overall the participants repeated 56% of the tongue twisters correctly and produced errors in 28% of the trials. In the remaining 16% of the trials, the productions were judged incorrectly (68% misses, 32% false alarms). Frequently produced errors were phonological slips (of the type '*a proper cropper*…'), word order errors (of the type '*How a clam can*

391 *cram*...'), adjective omissions (of the type '*a proper coffee pot'*), and semantic
392 substitutions that specifically seemed to be targeted at a circumvention of the
393 troublesome syllables (similar to '*How can a clam cram into a tidy cream can'*). In
394 the perception condition participants correctly identified 53% of the items as correct
395 and 27% as incorrect repetitions of the tongue twister. In 19% of the trials the
396 participants made an incorrect judgment (40% misses, 60% false alarms). The striking
397 similarity between the two conditions is the result of online scoring of the
398 production trials, to which the perception trials were matched in percentage of
399 errors. An overview of accuracy scores and response times measured from cue until
400 judgment is provided in Table 1.
401 In the production condition a significant learning effect was observed ($F(2, 882) = 38.16$, $p<.001$). In the first block 42% of the sentences was produced correctly, in the
402 second block 60% of the trials was produced correctly, and in the third block 68% of
403 the repetitions was produced correctly.
404 Table 1. *Accuracy and response time in milliseconds for the correct and incorrect*
405 *trials in the production and perception condition of the tongue twister task.*

	Production		Perception	
	Score	RT	Score	RT
Correct	M=56%	3693	M=53%	3554
	Range 35%-70%	(SD 1061)	Range 39-74%	(SD 913)
Incorrect	M=28%	4730	M=27%	3641
	Range 15%-55%	(SD 1357)	Range 17-41%	(SD 1134)

406

407 3.2 fMRI data

409 The contrasts made with the fMRI data were the following:

410 1. For each condition separately we contrasted erroneous trials > correct trials

411 Results from contrast 1 were used to make the following contrasts:

412 2. Conjunction analysis comparing similarities between activations during production

413 and perception.

414 3. Disjunction analysis comparing the activations that are independent for the

415 production and perception condition.

416 3a. Production > Perception.

417 3b. Perception > Production.

418

419 3.2.1. Conjunction analysis

420 A conjunction analysis was used to investigate the areas underlying error

421 detection that are common to speech production and speech perception. In this

422 analysis, we tested for a rejection of the conjunction null hypothesis (i.e., only those

423 voxels were reported as active which proved to be significant for speech production

424 and speech perception). The conjunction analyses revealed several clusters that

425 were commonly more active in erroneous compared to correct trials (Table 2; Figure

426 2). Clusters of activation were found in the pre-SMA extending into the dACC, the

427 left AI and IFG, and the right IFG extending into AI.

428 Table 2.

Peak Clusters of Activation revealed by Conjunction Error Trials Production and Perception

Structure	Peak coordinates (MNI)	Z-score	Extent

Pre-Supplementary Motor Area	-6 17 58	5.92	158
Left Insula	-33 20 5	5.95	63
Right Inferior Frontal Pars Triangularis	45 23 1	5.58	62
Left Inferior Frontal Opercularis	-45 20 13	5.05	15

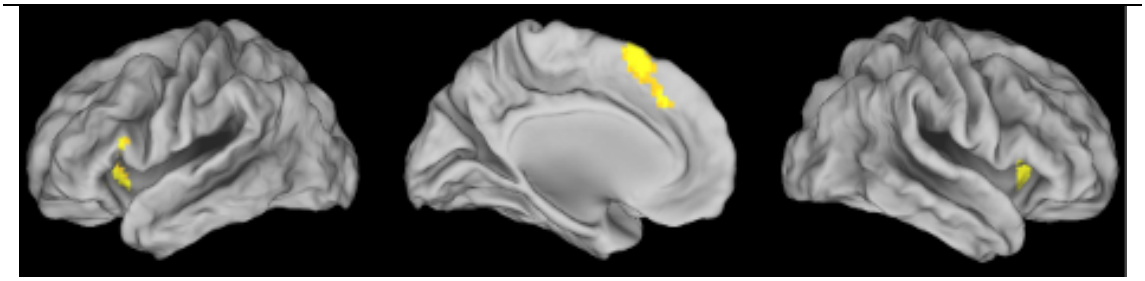

Figure 2. Activation map averaged across 21 subjects (p<.05, familywise error corrected) of the conjunction analysis error trials production and error trials perception.

3.2.2 Disjunction analysis

To investigate process-specific activations, namely production- and perception-specific error detection activation patterns, both an interaction analysis and a disjunction analysis can be applied. Both approaches were used to analyze the data. As the two analysis methods roughly yielded the same results, we chose to report only the results from the disjunction analysis, as they are more straightforward to interpret.

A disjunction analysis was used to investigate areas active in error detection specific for the two modalities, production and perception. Error detection in production was masked by error detection in perception to reveal what areas are specific for error detection in production. This analysis revealed clusters of activation (Table 3, Figure 3) in the left temporal pole, pre-SMA and dACC and BA 48.

Error detection in perception was masked by error detection in production to reveal areas specific to error detection in perception. This analysis revealed an array of clusters, including bilateral posterior superior temporal sulcus / middle temporal gyrus (pSTS/MTG), left AI and IFG, right supra marginal gyrus, middle frontal gyrus and precentral gyrus, extending into IFG (Table 3, Figure 4).

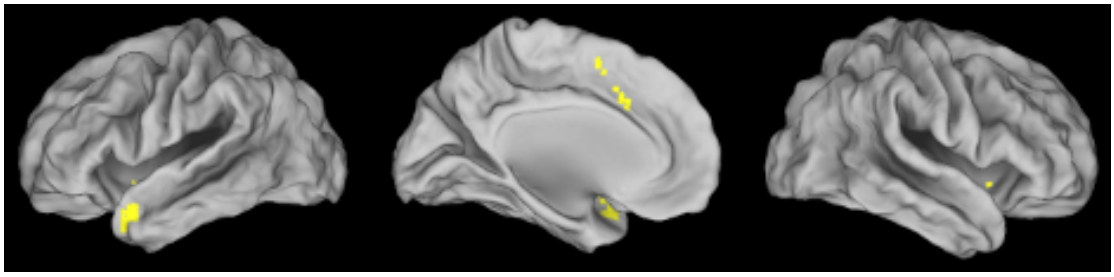

Figure 3. *Activation map averaged across 21 subjects (p<.05, familywise error corrected) of the disjunction analysis error trials production masked by error trials perception, revealing activation specific for error detection in production.*

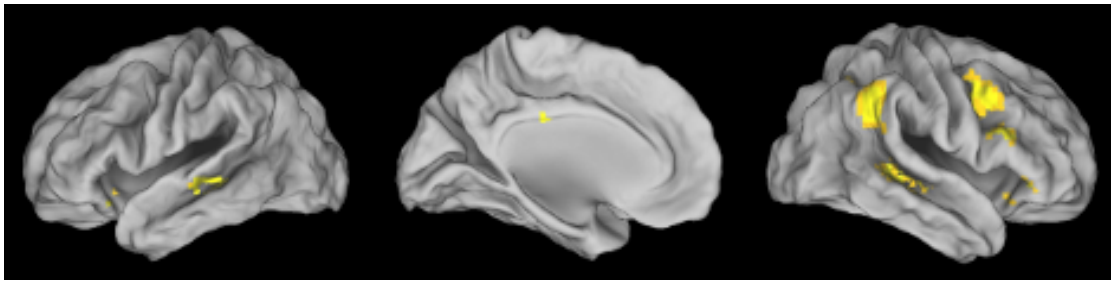

Figure 4. *Activation map averaged across 21 subjects (p<.05, familywise error corrected) of the disjunction analysis error trials perception masked by error trials production, revealing activation specific for error detection in perception.*

Table 3.

Peak Clusters of Activation revealed by Disjunction Analysis Error Trials Production and Perception

Structure	Peak coordinates (MNI)	Z-score	Extent

Production Errors –Perception Errors			
Left Temporal Pole	-42 11 -17	5.67	68
ACC	-6 20 34	5.36	18
pre-SMA	-6 8 49	5.36	13
White matter	33 11 -8	5.09	8
Perception Errors –Production Errors			
Right Middle Frontal Gyrus	45 11 43	6.23	277
Right Middle Temporal Gyrus	54 -37 1	5.75	125
Right Supra Marginal Gyrus	60 -46 31	6.03	173
Left Middle Temporal Gyrus	-57 -28 -5	6.20	56
Right Frontal Inferior Orb	45 35 -5	5.92	18
Right Thalamus	9 -16 10	5.77	10
Right Orbital Inferior Frontal Gyrus	33 23 -14	5.61	10
Corpus Callosum	-3 -25 28	5.06	10
Left Insula	-30 20 -11	5.15	7
Right Middle Frontal Gyrus	30 11 58	4.88	7

3.2.3 ROI analysis of the Superior Temporal Gyrus

The perceptual monitoring theories hold that speech monitoring takes place through the speech perception system. Many studies have pointed to the STG as a main locus for speech perception (see Price, 2012) and a possible candidate for perception-based error detection as it has been observed to respond to feedback alterations (McGuire et al, 1996; Hirano et al, 1997; Indefrey & Levelt, 2004; Christoffels et al,

2007, 2011; Tourville et al, 2008; Zheng et al, 2010; Takaso et al, 2010). To further examine the role of STS and STG, the hypothesized locus of the perceptual route for error detection, additional ROI analyses were conducted. From McGuire, Silbersweig, and Frith (1996) and Hirano et al. (1997) the clusters that increased for distorted feedback were selected for ROI analysis, as they are the basis of the hypothesis for perceptual monitoring through the STS/STG. Nine clusters were selected, four in the right hemisphere (all STG) and five in the left hemisphere (one in the STS, four in the STG). In the right hemisphere all selected areas showed a main effect of modality (all *p*'s <.005), with higher activation in perception compared to production. A main effect of accuracy was only significant for one ROI (coordinates: 62 -30 12, $p < .05$), which showed an activation decrease in erroneous trials compared to correct trials. Significant interactions were observed for three out of four ROIs (all *p*'s <.05) (not for 46 -20 4). These interactions were driven by significant lower activation in erroneous trials compared to correct trials in production, but not in perception. Results in the left hemisphere gave a more heterogeneous pattern; all areas showed a significant main effect of modality (all *p*'s <.05), with four out of five areas showing higher activation for perception compared to production (coordinates -58 12 4 showed the reverse pattern). With respect to accuracy an inconsistent and insignificant pattern of activations was observed, with only a significant main effect in one ROI (coordinates -52 -36 16, $p < .05$), which showed decreases in erroneous trials compared to correct trials for both production and perception. A significant interaction was observed in two area's (*p*'s <.005). This interaction was driven by significant activation differences between erroneous trials

491 and correct trials (increase in area -58 12 4, and decrease in area -60 -18 4) in
492 production, but not in perception.

493 Activation differences between erroneous and correct trials during
494 production and perception are presented in table 4. Essentially, these ROI analyses
495 show that the bilateral STG are stronger activated during speech perception
496 compared to production. With respect to error processing, however, the pattern of
497 activations observed was inconsistent with respect to the hypothesis that the STG
498 plays a primary role in internal verbal monitoring.

499 Table 4.

Percentage signal change in bilateral STG in erroneous trials compared to correct trials. Significant signal change is indicated by an asterisk (p<.05, ** p<.005)*

Structure	coordinates (MNI)	Perception	Production
Left STS	-50 -10 0	0.014	0.020
Left STG	-52 -36 16	-0.013	-0.050*
Left STG	-56 -8 0	0.014	-0.013
Left STG	-58 12 4	-0.005	0.109**
Left STG	-60 -18 4	0,031	-0.097**
Right STG	46 -20 4	0.014	-0.015
Right STG	54 -26 8	0.032	-0.070**
Right STG	52 -26 4	0.046*	-0.045*
Right STG	62 -30 12	0.030	-0.108**

500

501 4. Discussion

The goal of the current study was to investigate the neuronal structures underlying internal speech monitoring during production and speech monitoring during perception, and to use these functional neuroimaging data to distinguish between current theories of verbal monitoring. Perception-based verbal self-monitoring theories assume that error detection during speech production and speech perception both use similar, perceptual routes for error detection. Production-based theories of self-monitoring do not assume a role for the speech perception system in internal speech monitoring during production. We observed that error detection in noise-masked speech production and in speech perception both recruit the pre-supplementary motor area (pre-SMA), dorsal anterior cingulate cortex (dACC), bilateral anterior insula (AI), and inferior frontal gyrus (IFG). These observations suggest that error detection indeed recruits similar neural substrates and therefore might apply similar mechanisms for monitoring speech during production and perception. Crucially, no consistent pattern of activation related to error detection was observed in the bilateral superior temporal sulcus (Hirano et al. 1997; Indefrey & Levelt, 2004; McGuire et al. 1996), suggesting that verbal monitoring occurs largely independent of speech perception systems.

The findings of the activation of a perception-independent monitoring network, and the inconsistent finding with respect to STG activation, taken together do not offer support for the perceptual monitoring theories, which assume error detection in internal speech to take place through speech perception processes (Hartsuiker & Kolk, 2001; Levelt, 1983, 1989; Indefrey and Levelt, 2004; Indefrey 2011; Hickok 2012), but rather supports a conflict monitoring model of error detection in speech, as proposed by Nozari et al. (2011). This conflict monitoring

theory builds on domain-general theories of error detection and conflict resolution (e.g., Botvinick et al, 2001; Yeung et al, 2004) and proposes that speech monitoring takes place by measuring conflict in a processing layer, which is sent to a domain-general executive center, such as the ACC, which increases control in order to resolve the conflict. Note, however, that our findings are also compatible with the forward modeling theory for monitoring as proposed by Pickering and Garrod (2013, 2014), which also assumes a perception independent monitor.

4.1 The role of the STG in verbal monitoring

If the STS/STG were the main locus for error detection in speech, activation increases would be expected for erroneous trials compared to correct trials, in both speech production and perception. Instead we found increased activations in the STG in production, compared to comprehension, and an inconsistent pattern of activation with respect to erroneous compared to correct trials. In both hemispheres one cluster showed a main effect of accuracy, with decreased activation in erroneous compared to correct trials. Additionally in the right hemisphere we found an interaction of accuracy and modality, with lower activations in erroneous trials compared to correct trials in production, but not in perception. This finding is surprising, and not easy to interpret. At least the finding suggests a role for the right STG during speech production, related to verbal monitoring. However, we must be cautious in interpreting this finding, as it is a finding from a post-hoc analysis, and the direction of the effect does not conform to any of our predictions.

4.3 Domain general conflict monitoring

The conflict monitoring literature supports an explanation of the current findings within a framework of a domain general monitoring mechanism. The

structures found to be active in monitoring during speech perception and internal speech monitoring during speech production (the pre-SMA, ACC, IFG, and AI) are all regions that have been related to conflict processing in numerous tasks that require conflict resolution. The same network has been found to be active in both error making and error observation in the action domain: error detection increased activity in the ACC, SMA, pre-SMA, and AI (Newman-Norlund et al. 2009; Desmet et al. 2013, Monfardini et al. 2013). In the literature this network is also described as the cingulo-opercular network, which has been related to task maintenance (e.g. Dosenbach et al., 2008). The pre-SMA and ACC play a critical role in performance monitoring and adjustment of cognitive control (e.g., Botvinick et al., 2001; Ullsperger & Von Cramon, 2006; Bonini et al., 2014). The ACC has consistently been found to be activated after response conflict detection, errors, and unfavorable outcomes (see Ridderinkhof, 2004 for an overview). Also the dorsal ACC has been localized as the primary generator of the ERN component (e.g., Van Veen & Carter, 2002; Herrmann et al., 2004). The IFG / AI has also frequently been observed in cognitive control tasks and tasks engaging attentional processes (e.g., Craig, 2010), and is hypothesized to be responsible for signaling awareness and in regulating response selection (see Tops & Boksem, 2011 for an overview). Increased right IFG activation is often observed in tasks involving stopping one's actions, including stopping speech (Xue et al., 2008). Increased right IFG activation was also observed in preparation of word pairs that were primed to lead to embarrassing vs. neutral speech errors, showing its involvement in increased control during language processing (Severens et al., 2011). Together these areas form a domain-general network for conflict resolution.

574 4.4 Process specific activations

575 Apart from the domain-general activations, as observed in the conjunction
576 analysis, error detection in speech perception and production showed process-
577 specific activations. Self-monitoring of internal speech during noise-masked speech
578 production recruited the left temporal pole and pre-SMA and ACC. The pre-SMA is
579 known to have a somatotopic organization (Chainay et al, 2004; Alario et al, 2006),
580 resulting in process-specific activations. Left temporal pole activations are observed
581 in tasks requiring the composition of sentence meaning, and more specifically in the
582 processing of syntactic structure (Vandenberghe, 2002; Grodzinsky & Friederici
583 2006; Humphries, 2006

584 Error detection in speech perception revealed process-specific activations in
585 a few clusters in the left hemisphere, and a more extensive pattern of activation
586 clusters in the right hemisphere. Left hemisphere activations include anterior insula
587 and posterior middle temporal gyrus. The left insula has interestingly been
588 demonstrated to play a crucial role in phonological retrieval and articulation (Shafto
589 et al., 2010). Activations in the pSTS/MTG are observed bilaterally in response to
590 (noisy) auditory stimuli (Bates, 2003; Boatman 2004; Fu et al. 2006), and in
591 integration of auditory and visual information (Beauchamp et al., 2004). Left MTG
592 has also been linked to semantic processing (e.g. Demonet et al., 1992, 1994;
593 Vandenberghe et al, 1996; Stromswold et al, 1996; Binder et al, 2009; Diaz and
594 McCarthy, 2009; but see Price, 2012). In the right hemisphere large clusters are
595 observed in the posterior middle frontal gyrus, precentral gyrus, in the
596 supramarginal gyrus, in the IFG/AI, and in the pSTS/MTG. The supramarginal gyrus is
597 involved in phonological perception and decision making (Hartwigsen et al. 2010;

598 Buchsbaum et al. 2008; McDermott et al. 2003; Price et al. 1997) although it typically
599 does not show up in speech comprehension tasks (Hickok and Poeppel, 2007;
600 Rauschecker and Scott, 2009).

601 4.5 Similar findings in monitoring language processing

602 A highly similar pattern of activation for error perception processing was
603 found in a study in which participant detected semantic errors during reading
604 (Raposo & Marques, 2013). Compared to correct sentences, sentences with
605 semantic anomalies increased attention in the right precentral gyrus, right marginal
606 gyrus, and the ACC. The same areas are observed to be increased in activation in the
607 perception condition of the current experiment. The fact that monitoring in this
608 different modality, namely reading, shows similar results further supports a domain
609 general monitoring mechanism.

610 The current findings are also in line with preceding research into language
611 control and altered feedback monitoring, which consistently reported activations in
612 the ACC, SMA and frontal areas (e.g. Fu et al, 2006; Christoffels et al, 2007; Tourville
613 et al, 2008; Piai et al. 2013). One interesting difference between the before-
614 mentioned studies of Fu et al. (2006) and Christoffels et al. (2007) into feedback
615 monitoring and our findings is that we did not find increased activations in the
616 cerebellum. These cerebellar activations during feedback processing have also been
617 related to error detection in perception-based models, as it is hypothesized to drive
618 corrective motor commands to the motor cortex after receiving input from
619 somatosensory and auditory areas (Ito, 2008; Tourville & Guenther, 2011; Hickok,
620 2012). While the studies above specifically looked at the effect of manipulating
621 external feedback, we have excluded external feedback by noise masking. This hints

622 that the role of the cerebellum might be more closely related to external feedback
623 instead of monitoring proper.

624 In line with our findings are recent studies in which fMRI was used to study
625 conflict resolution in language processing. Wittfoth et al (2009) investigated
626 emotional conflict processing in speech perception, and Piai at al. (2013)
627 investigated attentional conflict in language and non-language processing.
628 Processing of emotional conflicting information (e.g., a semantically positive
629 sentence with a negative prosody) also showed an increase in BOLD response in the
630 posterior medial prefrontal cortex extending into ACC, bilateral insula and IFG,
631 posterior cingulate and inferior parietal lobule. Processing of attentional conflict in a
632 Stroop Task (color word is printed in an incongruent ink color), a Picture-Word
633 Interference Task (picture and distractor are semantically related), and a Simon Task
634 (press a left or right button to a visual stimulus presented on the opposite side) all
635 elicited ACC activation. So what we observe in speech error detection are activations
636 consistent with a domain-general error detection mechanism, through performance
637 monitoring and adjustment of cognitive control.

638 The finding of the current study, that of a conflict monitoring system which
639 operates during both production and perception, and which has been observed to
640 perform the same task in non-linguistic processes is important for three reasons.
641 First of all, these results have provided a preliminary answer to the question
642 whether verbal monitoring in production is perception-based. Clearly verbal
643 monitoring can occur largely independent of perception systems, and is therefore
644 production-based.

Second, as the network described here for verbal monitoring already has been studied much more extensively in relationship to conflict monitoring, we can now further investigate whether conflict monitoring mechanisms can apply similarly to verbal monitoring. This could hugely increase our understanding of verbal monitoring, and how this could lead to monitoring deficits, which presumably underlie speech pathologies such as stuttering, and auditory verbal hallucinations such as observed in schizophrenia.

Third, most current theories of production-based monitoring are limited to speech production. The current findings provide insight into how verbal monitoring might occur during speech perception; namely highly similar as during production.

4.6 Limitations

The current study has some limitations, of which a few pertain to the use of noise masking. The first issue regarding the presentation of noise masking during production is that it might have induced activations (e.g. Scott & McGettigan, 2013; Scott et al., 2004) and increased the cognitive load for the participants. However, as this would have equally affected the erroneous and the correct trials, which we contrasted to see the neural basis of error detection in verbal monitoring, the activations we report in the current paper are not noise-induced activations. If indeed the presence of noise did increase the cognitive load, it might have resulted in the production of more errors, which would have been beneficial for the current study.

A second comment related to noise masking during production is that proprioception and bone conduction of the produced speech cannot be excluded as a monitoring channel. Lackner and Tuller (1979) hypothesized that word selection

errors could be detected on the basis of tactile feedback. However, a more recent study by Postma and Noordanus (1996) contradicts this claim. In their study errors were reported during four production conditions: silent, mouthed, noise-masked and normal feedback. The number of reported errors were the same for the first three conditions, but increased in the fourth. Only the feedback from the external channel after production provides additional information for error detection on top of internal channel monitoring. If proprioception and bone conduction were channels by which monitoring can take place on top of internal speech, one would expect to see an increase in number of errors reported in the mouthed (proprioception) and noise-masked condition (proprioception and bone conduction) compared to the silent condition. But since proprioception and bone conduction did not contribute to the detection of more errors compared to the silent speech task, we cannot assume these channels to be of significant value for monitoring.

Despite these limitations resulting from noise masking during speech production, we opted for noise-masked feedback. By noise masking the overt speech with headphones, the participant could not hear his or her auditory feedback and would thus have to monitor their internal speech, and the experimenter could use the produced overt speech to verify the correctness of the repetition. Another benefit is that by having the participant produce overt speech, unlike covert speech, it is certain beyond doubt that the speech plan is fully formed (Barch et al., 1999; Huang et al., 2001; Gracco et al., 2005).

The use of a button press response for error detection can also be seen as a limitation, as it makes the task somewhat less naturalistic, and focuses the attention of the participants on error detection. The button press was included in the

693 paradigm to measure whether the participant was aware of the error or not. People
694 do not correct all their speech errors (e.g. Nooteboom, 1980), but it is unclear
695 whether uncorrected errors are ones the producer was unaware of, or ones where
696 the producer was aware of the error but did not bother to correct it (Berg, 1986). So
697 the only way to be sure that a participant was aware of the error was to directly ask
698 the participants. Also, if large numbers of both conscious and unconscious errors had
699 been made, it would have been interesting to investigate whether a difference exists
700 in brain activations between conscious and unconscious error production. However,
701 too few unconscious errors were produced to make this comparison.
702 A further limitation of this study is that there was no counterbalancing of the
703 order of the production and the perception condition; each participant first
704 performed the production condition and then the perception condition. Although
705 this lack of counterbalancing may have disadvantages, we felt these were
706 outweighed greatly by the advantage of being able to match the error percentages
707 in perception to that in production. This is of course only possible with a fixed order
708 of the conditions, and allows for a direct comparison between production and
709 perception. An unbalanced distribution of error percentages in the production and
710 perception condition would severely impair the validity of a comparison between
711 error detection in the production and perception condition.
712
713 In summary, our results suggest that error detection in speech processing
714 takes place through a domain-general conflict monitoring system, which comprises
715 the dorsal anterior cingulate cortex, supplementary motor area, bilateral anterior
716 insula, and inferior frontal gyrus. This network, which has been consistently

717 observed in non-linguistic conflict, is recruited for both speech perception and
718 speech production. The lack of evidence for the involvement of the superior
719 temporal gyrus does not offer support for perceptual theories of error monitoring.
720 The involvement of the conflict-monitoring network rather argues for a conflict
721 monitoring account of error detection in speech.
722

723 References

724 Abel S, Dressel K, Kümmerer D, Saur D, Mader D, Weiller C, Huber W (2009) Correct
725 and erroneous picture naming responses in healthy subjects. Neuroscience Letters
726 463:167-171.

727 Abbs JH, Gracco VL (1983). Sensorimotor actions in the control multi-joint speech
728 gestures. Trends in Neuroscience 6:391-395.

729 Abbs JH, Gracco VL, Cole KJ (1984). Control of multimovement coordination:
730 sensorimotor mechanisms in speech motor programming. Journal of Motor Behavior
731 16:195-231.

732 Alario F-X, Chainay H, Lehericy S, Cohen L (2006) The role of the supplementary
733 motor area (SMA) in word production. Brain Research 1076:129-143.

734 Barch DM, Carter CS, Braver TS, Sabb FW, Noll DC, Cohen JC (1999) Overt verbal
735 responding during fMRI scanning: empirical investigations of problems and potential
736 solutions. Neuroimage 10:642-657.

737 Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF (2003)
738 Voxel-based lesion-symptom mapping. Nature Neuroscience 6:448–450.

739 Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual
740 information about objects in superior temporal sulcus. Neuron 41: 809-823

741 Berg T (1986) The aftermath of error occurrence: Psycholinguistic evidence from cut-
742 offs. Language and Communication 6:195–213.

743 Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic
744 system? A critical review and meta-analysis of 120 functional neuroimaging studies.
745 Cerebral Cortex 19:2767–2796.

746 Blackmer ER, Mitton JL (1991) Theories of monitoring and the timing of repairs in
747 spontaneous speech. Cognition 39:173-194.
748 Boatman D (2004) Cortical bases of speech perception: evidence from functional
749 lesion studies. Cognition 92:47–65.
750 Bonini F, Burle B, Liégeois-Chauvel C, Régis J, Chauvel P, Vidal F (2014) Action
751 monitoring and medial frontal cortex: leading role of supplementary motor area.
752 Science 343:888-891.
753 Botvinick MM, Braver TS ,Barch DM ,Carter CS, Cohen JD (2001) Conflict Monitoring
754 and Cognitive Controle. Psychological Review 108:624-652.
755 Buchsbaum BR, D'Esposito M (2008) The search for the phonological store: from
756 loop to convolution. Journal of Cognitive Neuroscience 20:762–778
757 Chainay H, Krainik A, Tanguy ML, Gerardin E, Le Bihan D, Lehericy S. (2004) Foot, face
758 and hand representation in the human supplementary motor area. Neuroreport
759 15:765–769.
760 Christoffels IK, Formisano E, Schiller NO (2007) Neural correlates of verbal feedback
761 processing: an fMRI study employing overt speech. Human Brain Mapping 28:868–
762 879.
763 Christoffels IK, Van De Ven V, Waldorp LJ, Formisano E, Schiller NO (2011) The
764 sensory consequences of speaking: parametric neural cancellation during speech in
765 auditory cortex. PLoS One: 6.
766 Craig AD (2010) Once an island, now the focus of attention. Brain Structure and
767 Function 214:395–396.
768 Curio G, Neuloh G, Numminen J, Jousmaki V, Hari R. (2000) Speaking modifies voice-
769 evoked activity in the human auditory cortex. Human Brain Mapping, 9, 183–191.

770 Demonet JF, Chollet F, Ramsay S, Cardebat D, Nespoulous JL, Wise R, Rascol A,
771 Frackowiak R (1992) The anatomy of phonological and semantic processing in
772 normal subjects. Brain 115:1753–1768.
773 Demonet JF, Price C, Wise R, Frackowiak RS (1994) Differential activation of right and
774 left posterior sylvian regions by semantic and phonological tasks: a positron-
775 emission tomography study in normal human subjects. Neuroscience Letters
776 182:25–28.
777 Desmet C, Deschrijver E, Brass M (2013) How social is error observation? The neural
778 mechanisms underlying the observation of human and machine errors. Social
779 Cognitive and Affective Neuroscience 9:427-435.
780 Diaz MT, McCarthy G (2009) A comparison of brain activity evoked by single content
781 and function words: an fMRI investigation of implicit word processing. Brain
782 Research 1282:38–49.
783 Dosenbach NU, Fair DA, Cohen AL, Schlaggar BL, Petersen SE (2008) A dual-networks
784 architecture of top-down control. Trends in cognitive sciences 12:99-105.
785 Eden GE, Joseph JE, Brown HE, Brown CP, Zeffiro TA (1999) Utilizing hemodynamic
786 delay and dispersion to detect fMRI signal change without auditory interference: the
787 behavior interleaved gradients technique. Magnetic Resonance in Medicine 41:13-
788 20.
789 Eliades SJ, Wang X (2003) Sensory-motor interaction in the primate auditory cortex
790 during self-initiated vocalizations. Journal of Neurophysiology 89:2194–2207.
791 Eliades SJ, Wang X (2005) Dynamics of auditory-vocal interaction in monkey auditory
792 cortex. Cerebral Cortex 15:1510–1523

793 Fu CH, Vythelingum GN, Brammer MJ, Williams SC, Amaro Jr E, Andrew CM, Yaguez
794 L, van Haren NE, Matsumoto K, McGuire PK (2006) An fMRI study of verbal self-
795 monitoring: neural correlates of auditory verbal feedback. Cerebral Cortex 16:969–
796 977.
797 Gracco VL, Tremblay P, Pike B (2005) Imaging speech production using fMRI.
798 Neuroimage 26:294–301.
799 Grodzinsky Y, Friederici AD (2006) Neuroimaging of syntax and syntactic processing.
800 Current Opinion in Neurobiology 16:240–246.
801 Hartwigsen G, Baumgaertner A, Price CJ, Koehnke M, Ulmer S, Siebner HR (2010)
802 Phonological decisions require both the left and right supramarginal gyri.
803 Proceedings of the National Academy of Sciences U.S.A. 107:16494–16499.
804 Hartsuiker RJ, Kolk HHJ (2001) Error monitoring in speech production: A
805 computational test of the perceptual loop theory. *Cognitive Psychology* 42*:*113-157*.*
806 Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM (2005) Fine-tuning of auditory
807 cortex during speech production. Psychophysiology 42:180–190.
808 Heinks-Maldonado TH, Nagarajan SS, Houde JF (2006) Magnetoencephalographic
809 evidence for a precise forward model in speech production. Neuroreport 17:1375–
810 1379.
811 Herrmann MJ, Rommler J, Ehlis AC, Heidrich A, Fallgatter AJ (2004) Source
812 localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe).
813 Cognitive Brain Research 20:294–299
814 Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nature
815 Reviews Neuroscience 8:393–402

816 Hickok G (2012) Computational neuroanatomy of speech production. Nature
817 Reviews Neuroscience 13:135-145
818 Hirano S, Kojima H, Naito Y, Honjo I, Kamoto Y, Okazawa H, Ishizu K, Yonekura Y,
819 Nagahama Y, Fukuyama H, Konishi J (1997) Cortical processing mechanism for
820 vocalization with auditory verbal feedback. Neuroreport 8:2379–2382.
821 Huang J, Carr TH, Cao Y (2001) Comparing cortical activations for silent and overt
822 speech using event- related fMRI. Human Brain Mapping 15:39–53.
823 Humphries C, Binder JR, Medler DA, Liebenthal E (2006) Syntactic and semantic
824 modulation of neural activity during auditory sentence comprehension. Journal of
825 Cognitive Neuroscience 18:665–679.
826 Indefrey P (2011) The spatial and temporal signatures of word production
827 components: a critical update. Frontiers in Psychology 2:255.
828 Indefrey P, Levelt WJM (2004) The spatial and temporal signatures of word
829 production components. Cognition 92:101–144.
830 Ito M (2008) Control of mental activities by internal models in the cerebellum.
831 Nature Reviews Neuroscience 9:304–313.
832 Lackner JR, Tuller BH (1979) Role of efference monitoring in the detection of self-
833 produced speech errors. In: Sentence processing (Cooper WE, Walker ECT, ed), pp.
834 281–294. Hillsdale, N.J.: Erlbaum.
835 Laver J (1980) Monitoring systems in the neurolinguistic control of speech
836 production. In: Errors in linguistic performance: Slips of the tongue, ear, pen, and
837 hand (Fromkin VA, ed), pp. 287-305. New York: Academic Press.
838 Levelt WJM (1983) Monitoring and self-repair in speech. Cognition 14:41–104.

839 Levelt WJM (1989) Speaking: From intention to articulation. Cambridge, MA: MIT
840 Press.
841 Matlab and SPM8 software Wellcome Department of Cognitive Neurology, London,
842 UK.
843 McDermott KB, Petersen SE, Watson JM, Ojemann JG (2003) A procedure for
844 identifying regions preferentially activated by attention to semantic and
845 phonological relations using functional magnetic resonance imaging.
846 Neuropsychologia 41:293–303
847 McGuire PK, Silbersweig DA, Frith CD (1996) Functional neuroanatomy of verbal self-
848 monitoring. Brain 119:907–917.
849 Menenti L, Gierhan SME, Segaert K, Hagoort P (2011) Shared language: Overlap and
850 segregation of the neuronal infrastructure for speaking and listening revealed by
851 fMRI. Psychological Science 22:1173–82.
852 Monfardini E, Gazzola V, Boussaoud D, Brovelli A, Keysers C, Wicker B (2013)
853 Vicarious neural processing of outcomes during observational learning. PLOS One 8.
854 Newman-Norlund RD, Ganesh S, van Schie HT, De Bruijn ERA, Bekkering H (2009)
855 Self-identification and empathy modulate error-related brain activity during the
856 observation of penalty shots between friend and foe. Social Cognitive and Affective
857 Neuroscience 4:10-22.
858 Nooteboom SG (1980) Speaking and unspeaking: detection and correction of
859 phonological and lexical errors in spontaneous speech In: Errors in linguistic
860 performance: Slips of the tongue, ear, pen, and hand (Fromkin VA, ed), pp. 87-95.
861 New York: Academic Press.

862 Nozari N, Dell GS, Schwartz MF (2011) Is comprehension necessary for error
863 detection? A conflict-based account of monitoring in speech production. Cognitive
864 Psychology 63:1-33.
865 Numminen J, Salmelin R, Hari R (1999) Subject's own speech reduces reactivity of the
866 human auditory cortex. Neuroscience Letters 265:119–122.
867 Oppenheim GM, Dell GS (2008) Inner speech slips exhibit lexical bias, but not the
868 phonemic similarity effect. Cognition 106:528-537.
869 Piai V, Roelofs A, Acheson DJ, Takashima A (2013) Attention for speaking: Neural
870 substrates of general and specific mechanisms for monitoring and control. Frontiers
871 in Human Neuroscience 7:832.
872 Pickering MJ, Garrod S (2013) An integrated theory of language production and
873 comprehension. Behavioral and Brain Sciences 36:329-392.
874 Pickering MJ, Garrod S (2014) Self- Other and joint monitoring using forward models.
875 Frontiers in Human Neuroscience 8:132-143.
876 Postma A, Kolk HHJ (1992) The effects of noise masking and required accuracy on
877 speech errors, disfluencies, and self-repairs. Journal of Speech and Hearing Research
878 35:537–544.
879 Postma A, Noordanus C (1996) Production and detection of speech errors in silent,
880 mouthed, noise-masked, and normal auditory feedback speech. Language and
881 Speech 39: 375–392.
882 Postma A (2000) Detection of errors during speech production: a review of speech
883 monitoring models. Cognition 77:97–132.
884 Price CJ, Moore CJ, Humphreys GW, Wise RJS (1997) Segregating semantic from
885 phonological processes during reading. Journal of Cognitive Neuroscience 9:727–733

886 Price CJ (2012) The anatomy of language: a review of 100 fMRI studies published in
887 2009. Annals of the New York Academy of Sciences 1191:62–88
888 Raposo A Marques JF (2013) The contribution of fronto-parietal regions to sentence
889 comprehension: Insights from the Moses illusion. NeuroImage 83:431-437.
890 Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman
891 primates illuminate human speech processing. Nature Neuroscience 12:718–724.
892 Ridderinkhof RK, Van den Wildenberg WPM, Segalowitzd SJ, Cartere CS (2004)
893 Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in
894 action selection, response inhibition, performance monitoring, and reward-based
895 learning. Brain and Cognition 56:29–140.
896 Scott S K, McGettigan C (2013) The neural processing of masked speech. Hearing
897 research, 303:58-66.
898 Scott SK, Rosen S, Wickham L, Wise RJS (2004) A positron emission tomography
899 study of the neural basis of informational and energetic masking efforts in speech
900 perception. The Journal of the Acoustical Society of America 115:813-821.
901 Shafto MA, Stamatakis EA, Tam PP, Tyler LK (2010) Word retrieval failures in old age:
902 the relationship between structure and function. Journal of Cognitive Neuroscience
903 22:1530-1540.
904 Severens E, Kühn S, Hartsuiker RJ, Brass M (2011). Functional mechanisms involved
905 in the internal inhibition of taboo words. *Social cognitive and affective neuroscience*,
906 nsr030.

907 Siegenthaler BM, Hochberg I (1965). Reaction time of the tongue to auditory and
908 tactile stimula- tion. Perceptual and Motor Skills 21:387-393.

909 Stromswold K, Caplan D, Alpert N, Rauch S (1996) Localization of syntactic
910 comprehension by positron emission tomography. Brain and Language 52:452–473.
911 Takaso H, Eisner F, Wise RJ, Scott SK (2010) The effect of delayed auditory feedback
912 on activity in the temporal lobe while speaking: a positron emission tomography
913 study. Journal of Speech Language and Hearing Research 53:226–236.
914 Tian X, Poeppel D (2010) Mental imagery of speech and movement implicates the
915 dynamics of internal forward models. Frontiers in Psychology 1: 166.
916 Tian X, Poeppel D (2013) The effect of imagination on stimulation: the functional
917 specificity of efference copies in speech processing. Journal of cognitive
918 neuroscience 25: 1020-1036.
919 Tops M, Boksem MA (2011) A potential role of the inferior frontal gyrus and anterior
920 insula in cognitive control, brain rhythms, and event-related potentials. Frontiers in
921 Psychology 2:330.
922 Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory
923 feedback control of speech. Neuroimage 39:1429–1443.
924 Tourville JT, Guenther FH (2011) The DIVA model: A neural theory of speech
925 acquisition and production. Language and Cognitive Processes 25:952-981.
926 Ullsperger M, von Cramon DY (2006) The role of intact frontostriatal circuits in error
927 processing. Journal of Cognitive Neuroscience, 18:651-664.
928 Van Veen V, Carter CS (2006) Error detection, correction, and prevention in the
929 brain: a brief review of data and theories. Clinical EEG and neuroscience, 37:330-335.
930 Van Veen V, Carter CS (2002) The timing of action-monitoring processes in the
931 anterior cingulate cortex. Journal of cognitive neuroscience, 14:593-602.

932 Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RS (1996) Functional
933 anatomy of a common semantic system for words and pictures. Nature 383:254–
934 256.
935 Vandenberghe R, Nobre AC, Price CJ (2002) The response of left temporal cortex to
936 sentences. Journal of Cognitive Neuroscience 14:550–560.
937 Wittfoth M, Schröder C, Schardt DM, Dengler R, Heinze HJ, Kotz SA (2009). On
938 emotional conflict: interference resolution of happy and angry prosody reveals
939 valence-specific effects. Cerebral Cortex, 20:383-392.
940 Xue G, Aron AR, Poldrack RA (2008). Common neural substrates for inhibition of
941 spoken and manual responses. Cerebral Cortex 18:1923-1932.
942 Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: conflict
943 monitoring and the error-related negativity. Psychological Review 11:931–959.
944 Zheng ZZ, Munhall KG, Johnsrude IS (2010) Functional overlap between regions
945 involved in speech perception and in monitoring one's own voice during speech
946 production. Journal of Cognitive Neuroscience 22:1770-1781.
947

Appendix A

Dutch tongue twister sentences with rough translations to English in italics.

Tongue twister used in the production condition.

1. Ruud rups raspt rap rode ronde radijsjes.

   Cathy catterpillar quicky grates round red radishes.

2. De meid sneed zeven scheve sneden brood.

   The maid cut seven skew slices of bread.

3. Als apen apen naapen, apen apen apen na.

   When monkeys mimic monkeys, monkeys mimic monkeys.

4. Wiske mixt whisky in de whisky mixer.

   Wilma mixes whisky in the whisky mixer.

5. Gijs grijpt de grijsgrauwe gans graag gauw.

   Gordon gladly grabs the grey goose swiftly.

6. Baardige artsen helpen aarzelende bedelaars.

   Bearded doctors help hesitant beggars.

7. Als een potvis in een pispot pist, zit de pispot vol met potvispis.

   If a sperm whale pisses in a pissjar, the pissjar is filled with sperm whale piss.

8. Een pet met een platte klep is een platteklep.

   A cap with a flat flap is a flat flap cap.

9. Vaders vader vond vier vuile vesten van vier vuile venten.

   Father's father found four filthy cardigans of four filthy blokes.

10. Sluwe feministen foeteren op flemende sloeries.

    Sly feminists grumble about flanneling floozies.

11. Jeukt jouw jeukende neus zoals mijn jeukende neus jeukt?

Does your itchy nose itch like my itchy nose itches?

12. De koetsier poetst de postkoets met postkoetspoets.

The coachman polishes the coach with coach polish.

13. Pappa pakt de platte blauwe bakpan.

Daddy grabs the flat blue frying pan.

14. Pseudo-psychologen sporten als speren.

Pseudo-psychologists sport like crazy.

15. Aaibare kraaien leggen kale kraaie-eieren.

Cuddly crows lay bald crow eggs.

16. Ping pingpongde de pingpongbal naar Pong.

Ping ping-ponged the ping pong ball to Pong.

17. Krakende krekels trippelen op tegels.

Creaking crickets patter on the tiling.

18. De kat krabt de krullen van de trap.

The cat scratches shavings of the stairs.

19. Knappe kappers kappen knap.

Handsome hairdressers cut hansomely.

20. Achtentachtig achterdochtige doktersdochters.

Eighty-eight suspicious doctor's daughters.

21. Zeven zotten zullen zes zomerse zondagen zwemmen zonder zwembroek.

Seven fools will swim six Sundays without swimming trunks.

22. Piet's priesterpij is piepklein.

Pete's priests frock is very tiny.

Tongue Twisters used in the perception condition

1. Trillend trippelde tante Tiny tandloos naar de treiterende tandarts toe.

Aunt Tilly tremblingly toddled toothless to the harassing dentist.

2. Tijdens de afwas viel de asbak in de afwasbak.

During the wash up the ashtray fell into the kitchen sink.

3. Liesje leerde Lotje lopen langs de lange Lindenlaan.

Lacey learned Laney how to walk along the long Linden lane.

4. Knappe slakken snakken naar slappe sla.

Pretty snakes yearn for limp lettuce.

5. Toen Lotje niet wou lopen, liet Liefje Lotje staan.

As Lany would not walk, Lacy left Lany.

6. Vissende vissers die vissen naar vissen, maar vissende vissers die vangen vaak bot.

Fishing fishermen fish for fish, but fishing fishermen often catch zilch.

7. Dikke drilboren drillen door dikke deuren.

Large drills drill trough thick doors.

8. Kriegelig kocht Krelis kilo's kruimige krieltjes.

Grumpily Gary bought kilo's of floury spuds.

9. De dunne dokter duwde de dikke dame door de draaiende draaideur.

The thin doctor pushed the fat lady through the spinning revolving door.

10. Trollen rollebollen als dollen in de drollen.

Trolls horse around in the turds like crazy.

11. Ezels eten netels niet en netels eten ezels niet.

Donkeys don't eat nettles, and nettles don't eat donkeys.

12. De pasgewassen was was pas gewassen nadat de pasgewassen was gewassen was.

The freshly washed laundry was only washed after the freshly washed laundry was washed.

13. De magere marktskraamvrouw kookte veel makreel.

The skinny stall woman cooked lots of mackerel.

14. De toetsenist test het toetsenbord.

The keyboardist tests the keyboard.

15. De grommende beer bromt beestachtig geestig.

The growling bear grumbles mightily funny.

16. Kniezende kneuzen kiezen kale keukens.

Moping misfits choose bare kitchens.

17. Babbelende baby's dromen van dommelende bosduifjes.

Babbling babies dream of dozy wild pigeons.

18. Pinnige dikke piloten drinken prille pils.

Stingy fat pilots drink early bears.

19. Nukkige nuchtere Nellie is niet nuttig.

Crancky sober Nelly is not usefull.

20. Gerooide woudreuzen groeien in mooie wouden.

Cleared wood giants grow in pretty woods.

21. Slome slavinnen lopen in sombere lompen.

Slow slaves walk around in dreary duds.

22. De stille prinses at knisperende spritsen.

The princess ate crackling cookies.