UNIVERSITEIT GENT

FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN

Verbal monitoring in production and perception: A cognitive neuroscience approach

Hanna S. Gauvin

Promotor: Prof. Dr. Robert J. Hartsuiker

Copromotor: Prof. Dr. Peter Mariën

Proefschrift ingediend tot het behalen van de academische graad van

Doctor in de Psychologie

2015

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Eind mei is hier dan opeens mijn proefschrift. Terwijl ik het gevoel kreeg net goed op dreef te zijn, moest er een afrondend document gemaakt worden. Het voelt bijgevolg ook niet als een afsluitend document, maar meer als een tussentijds verslag. Ik hoop dat mijn gevoel juist is, en dat dit het begin zal zijn van een academische carriere, en zeker niet het einde.

De eerste persoon die ik wil bedanken is Rob Hartsuiker. Bedankt dat jij mij de kans hebt gegeven om te komen doctoreren in Gent. De afgelopen jaren heb ik duidelijk gezien wat het effect is van een promotor op hoe het doctoraat wordt ervaren. Voor mij is het een super ervaring geweest. Jouw deur was altijd open, en jouw feedback relevant. Daarnaast heb je mij ook veel vrijheid gegeven, waardoor ik ook naast het doctoraat van alles heb kunnen organiseren (o.a. PhD meetings, labmeetings, een conflict modellen clubje, taalgroepuitjes en een mini-symposium). En natuurlijk al die congressen waar jij mij heen hebt laten gaan. Wat een feest!

Peter Mariën, wat een warme, vriendelijke en enthousiate persoon bent u! Het was steeds motiverend om contact met u te hebben. Ook de overige leden van mijn doctroraatscommissie wil ik bedanken voor hun begeleiding. Martin Pickering, thank you for your valuable feedback and interesting discussions over the past three years. Marcel Brass, bedankt voor uw kennis en begeleiding rondom de fMRI en AC studie. En ook zeker voor het in aanraden van Wouter de Baene en Lara Bardi. Patrick Santens, hartelijk dank voor uw hulp rondom de Parkinson patiënten studie. De uitzonderlijk grote bereidwilligheid van deze patiënten om deel te nemen reflecteert zeker hoe waardevol u voor hen bent.

Wouter de Baene, ik snap nog altijd niet dat ze jou uit Gent hebben laten gaan. Voor mij ben je goud waard geweest, en ik was perfect tevreden mocht ik ieder experiment voortaan met jou samen doen. Lara Bardi, thank you for your ad-hoc advice and supervision! Elger, terwijl iedereen het rustig aan

deed in de zon in de Provence kwam jij opeens met super interessante en relevante ideeën. En zonder dat je het door hebt gehad, vermoed ik, ben je een soort van mentor geworden.

Daarnaast zijn er natuurlijk mensen buiten het onderzoek die een gigantische bijdrage hebben geleverd. Teveel om hier te bedanken, dus ik kies er een paar uit die er uit springen. Maar een ieder die een pintje met mij heeft gedronken valt daar zeker onder!

Thijs, uiteraard! Nummer één. De tijd in Gent zit er eindelijk op! Of toch niet? De afgelopen jaren is bijna alles altijd onduidelijk geweest. Maar waarover ik niet heb hoeven twijfelen, dat ben jij! Bedankt voor je constante vertrouwen en liefde.

Pappa en mama, jullie zijn schatten. Hoewel het misschien niet altijd even duidelijk was wat nu precies de bedoeling was, zijn jullie mij altijd blijven steunen. Bedankt voor die steun en het blinde volste vertrouwen.

Han en Nicky, bedankt voor jullie niet aflatende interesse, alle meelevendheid, het fijne rustpunt met haring en heerlijke biertjes, en het bieden van een perfecte werkplek voor de laatste loodjes.

Debbie, zonder jou was ik hier nooit gekomen, letterlijk. Dus bedankt! Jij en Farhang hebben mij altijd een tweede huiskamer in Gent geboden, met altijd een heerlijk hapje en lekker biertjes. Geweldig!

Uschi, zonder jou was ik nooit gebleven. Je hebt me een thuis gegeven en de leukste plekken van Gent laten zien. Hanne, ik geniet zo van jouw humor! Wat een geluk dat ik samen met jullie twee heb mogen starten hier. En nu dus samen doctoreren. (Hoera! We hebben het overleefd!) Aster en Eowyn, wat jammer dat jullie er pas zo laat bij zijn gekomen – jullie zijn zo'n plezier en een enorme toevoeging! Eliana, you were always there with advice and to discuss the issues of having a boyfriend who isn't living in Belgium. Thank you all for bringing me pleasure!

Sarah, Joris en Nele, mijn lieve kantoorgenootjes! Bedankt voor jullie gezelschap de afgelopen jaren. Nele, zonder jou had ik nog altijd niet

geweten hoe onze vakgroep werkt. Sarah, ik ben blij we toch mijn hele doctoraat samen hebben mogen werken– de gezellige vakgroep historicus van ons kantoor. Joris, je bent zo'n lieve kantoorgenoot geweest al die tijd. Ik hoop dat je volgende kantoor genoot zo lief is voor jou als jij voor mij bent geweest.

Tot slot Martijn en Stijn, bedankt voor het rustige huis waar we in wonen. Ik had het niet beter kunnen hebben in dit laatste, gigantisch, drukke jaar!

CHAPTER 1
INTRODUCTION

We are humans: social animals that love to communicate, be it via facial expression, pheromones, emoticons, social media or via speech. Interestingly, we use the expression 'I am only human' to explain that we are not flawless, we make mistakes. Apparently, making mistakes is something we find extremely human. The current thesis centers around the combination of these two, apparently very human, properties, namely speech errors. More specifically, this thesis investigates how we prevent the production of speech errors.

One first question that needs answering, before we can go into the prevention of speech errors, is 'how do we detect our speech errors'? Of course once we have produced an error, we can hear it. But if we could only rely on hearing our errors to detect them, we would never be able to prevent an error from happening. It would be like driving a car, and only being able to see through the rear view mirror. You would only have information about where you have been, and not about where you are and where you are going (Hickok, 2012, p.135). We have reason to believe that we can also see through the windscreen; that we can detect speech errors before we have actually produced them. For instance when asked to repeat a tongue twister sentence, such as 'a proper copper coffee pot', several times internally (covert), we are still able to report that we have made an error in these repetitions. Also when we speak out loud (overt), but we cannot hear ourselves speak because our external speech is masked by a loud noise, we are still able to report when we have made an error. And when we do make a speech error, sometimes the production of the erroneous utterance is halted just after the first phoneme is produced. This is much too fast for an error detection mechanism that relies on hearing only, where we would first have to hear what we have said, then realize that it is wrong, and then stop the production of the word. Additionally, a larger P2 component is observed in EEG studies into word production, when later on the actual production lead to a speech error (Trewartha & Philips, 2013), suggesting that the error is detected before production. Taken together, these data argue for an ability to not only monitor our external speech, but that we are able to monitor internally.

How exactly this internal monitoring is performed is a matter of debate. We can roughly make a distinction between two classes of internal monitoring theories, based on what part of speech production and comprehension are involved in the monitoring process. Perception-based theories, of which the Perceptual Loop Theory (Levelt, 1983, 1989; Indefrey & Levelt 2004; Indefrey 2011) is the most influential, assumes that just before articulation a copy of the constructed speech is sent to the speech perception system. The speech perception system thus perceives the speech before it is spoken, in the same way as it perceives the speech after articulation. Production-based theories, (e.g., conflict monitoring by Nozari, Dell, & Schwartz, 2011; forward model theories by Pickering & Garrod, 2013, 2014; Hickok, 2012) assume that speech is constantly monitored for errors during the different stages of speech construction. According to these theories, internal speech monitoring is thus independent of speech perception. The production based monitoring theories can be divided in conflict monitoring and forward model theories. Conflict monitoring theories assume that during the selection of an item at a processing stage, multiple candidates become active, leading to a conflict between the competing items. When there is a high amount of conflict, this is detected by a monitoring system. Forward model theories assume that for each production stage a prediction is made of the outcome (a forward model), which is then compared to the actual outcome. When the prediction and the actual outcome do not match, this is detected as an error.

The question of how we detect our errors can, on the basis of existing theories of internal monitoring, be reformulated as: do we detect our errors via our perception system, similar to when we hear someone else make an error? This would be in line with perception-based monitoring. Or is the detection of our own errors during production independent of the speech perception system? This would by hypothesized by a production-based account of verbal monitoring. This distinction is investigated first via a literature review in Chapter 2. In Chapter 3 we investigate the role of the perception system during speech production on the basis of behavioral data, by measuring the effects of auditory perception on visual attention. In Chapter 4 we investigate the hypothesized neuroanatomical substrates of perceptual error detection with neuroimaging data. In Chapter 5 we investigate the relative contributions of the internal and external monitoring channels in self-monitoring of Parkinson's patients. In all the empirical studies we compare self-monitoring with monitoring of speech produced by someone else, to investigate if these two modalities are monitored the same, as the perception based account of self-monitoring proposes, or not. In Chapter 6 we discus how the production of speech errors is prevented.

In Chapter 2 an extensive review of the current models of internal verbal monitoring is provided. First we discuss the perception based monitoring model: the Perceptual Loop Theory (Hartsuiker & Kolk, 2001; Levelt, 1983, 1989; Indefrey & Levelt 2004; Indefrey 2011). Next we discuss the production-based monitoring models in the following order: the Forward Model (Pickering & Garrod, 2013, 2014), the Hierarchical State Feedback Control Model (Hickok, 2012) and the Conflict Monitoring Model (Nozari et al., 2011). Each model is explained, as well as the evidence on which the model is built. Further evidence in support of the models, or opposing the models is also discussed, as well as caveats. From this review we conclude that currently there is no model that is able to give a full account of error detection during production and perception, and to explain how error production is prevented.

In Chapter 3 and Chapter 4 we investigate whether we find behavioral or neuroimaging evidence for a role for the speech perception system during speech production, as hypothesized by the perceptual loop theory. In Chapter 3 we investigate whether there is evidence for perceptual effects in internal speech monitoring, as hypothesized by the perceptual loop theory, by measuring eye-movements during speech production. Previous research has demonstrated that during speech perception, participants have a preference to look at phonologically related items presented on the screen, compared to other items (McQueen and Viebahn, 2007; Huettig and McQueen, 2007, 2011). The same is observed when the participants produce speech; eye-movements are directed to items that are phonologically related to what they are saying (Huettig & Hartsuiker, 2010). We investigate whether during speech production, the eye-movements occur sooner than during speech perception. If indeed speakers hear their own speech via an internal route before they start speaking, as proposed by the perceptual loop theory, we would expect that eye-movements to the phonologically related item would occur earlier than eye-movements driven by the speech of someone else. However, our data show that eye-movements to the phonological item on display are observed in the same time frame during speech production as when listening to someone else's speech. This suggests that people do not hear themselves speak before production, as the perceptual loop theory proposes, but rather that internal speech monitoring occurs via a different mechanism.

In Chapter 4 we investigate the neural correlates of error detection. With the use of fMRI we investigate what brain areas are involved in the detection of an error in self produced speech, and in the detection of an error in speech produced by someone else. Participants repeat tongue twister sentences in the scanner, or listen to the production of someone else, and indicate via a button press whether the repetition is flawless or contains an error. By measuring the brain activation during the sentences in which an error is detected, and subtracting the brain activations during correct sentences, we can see what brain activations are specific to verbal error detection. This comparison is made for both speech production and for speech perception. The areas involved during error detection turned out to be highly similar during speech production and during speech perception. This suggests that there might be one error detection mechanism, and that error detection happens similarly during production and perception. However, the involved areas did not involve the speech perception system. Instead a network of areas was found that has often been associated with domain-general processing of errors; this included the ACC and IFG. Therefore, the results do not offer support for the perceptual loop theory, but are rather in favor of the domain general monitoring account, such as the conflict monitoring theory.

In Chapter 5 we test the language performance of Parkinson's disease patients, to investigate to what extent verbal monitoring is impaired, and what the relative contributions are of the internal and external monitoring channel in Parkinson's disease patients. Previous research into self-monitoring skills of Parkinson's patients found that self-monitoring was impaired, despite good comprehension skills (McNamara et al., 1992). A group of Parkinson's patients and a group of age matched controls performed nine tasks that measured language performance, and two tasks that measured cognitive performance. Part of these language tasks measured covert, internal, speech processing, and the other part measured overt, external, speech processing. They also performed one task that measured monitoring performance of the internal monitoring channel (during noise-masked feedback) and of the external monitoring channel during normal feedback and the perception of someone else's speech. By comparing the data of the Parkinson's patients with the control group, we investigate whether monitoring is impaired in the patient group. A comparison of internal speech monitoring with external speech monitoring is performed to investigate if Parkinson's patients rely more heavily on their internal monitoring channel, as observed in other patient populations, or if they use the internal and external channel similarly to healthy controls. We observed an effect of feedback condition on repair rate; overall both Parkinson's patients and controls repair more errors during normal feedback than during noise-masked feedback. However, in case of semantic errors we observe that the Parkinson's patients are not significantly affected by the noise masking, while the healthy controls are. Finally, we tested whether internal (external) speech performance was predictive of internal (external) monitoring, as the PLT would predict. Internal monitoring behavior was not predicted by internal speech task performance, and external monitoring behavior was not predicted by external speech task performance. This suggests that monitoring occurs relatively independent of speech perception. This calls into question the role of the perception system for verbal monitoring, especially for internal monitoring. However, the external speech performance tasks also did not predict speech monitoring during normal or noise-masked feedback, which is not concurrent with production-based accounts of error detection.

In Chapter 6 we propose a new model for conflict monitoring. The literature review in Chapter 2, and the empirical studies in Chapters 3, 4 and 5, provide converging evidence that there is not much support for the perceptual loop theory. These data rather support a conflict-monitoring model. In Chapter 6 we propose an extension of the existing conflict monitoring account from error detection in production to error detection in perception. We extend the model further, by going into monitoring for errors in others' speech and by going into the process of what happens after an error is detected. From the conflict monitoring literature (Verguts & Nootebaert, 2008; 2009) we borrow the idea that upon the detection of high conflict (which arguably takes place in the ACC), a boost of activation is given to input representations. At the neural level, one might speculate that such a boost corresponds to the secretion of a neurotransmitter. This neurotransmitter leads to an aspecific activation of all active nodes, which allows for selection of the correct node in high conflict situations, and thereby prevents the production of an error in the majority of trials.

In sum, the goal of this dissertation is to investigate the process of verbal monitoring. Specifically, this thesis investigates whether internal and external monitoring proceeds via the same, perception-based process, as proposed by the perceptual loop theory. We compare verbal internal and external monitoring with the use of eye-tracking, fMRI and Parkinson patient data. The data obtained suggest that verbal monitoring is not perception based, and that is a domain general process. We therefore propose the improvement of current monitoring models by describing a domain general monitoring mechanism for internal monitoring and external monitoring, by which conflict is resolved in a process-independent manner.

REFERENCES

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13, 135-145

Huettig, F., Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Processes*, 3, 347–374.

Huettig, F., McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language*, 54, 460–482.

Huettig, F., McQueen, J. M. (2011). The nature of the visual environment induces implicit biases during language-mediated visual search. *Memory & Cognition*, 39, 1068–1084.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, 2, 255.

Indefrey, P., Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* MIT Press.

McNamara, P., Obler, L. K., Au, R., Durso, R., Albert, M. L. (1992). Speech monitoring skills in Alzheimer's disease, Parkinson's disease, and normal aging. *Brain and Language*, 42, 38–51.

McQueen, J. M., Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, 60, 661–671.

Nozari, N., Dell, G. S., Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology,* 63, 1–33.

Pickering, M.J., Garrod, S. (2013). An integrated theory of language production and comprehension. Behavioral and Brain Sciences 36, 329-392.

Pickering, M.J., Garrod, S. (2014). Self- Other and joint monitoring using forward models. *Frontiers in Human Neuroscience*, 8, 132-143.

Trewartha, K. M., Phillips, N. A. (2013). Detecting self-produced speech errors before and after articulation: an ERP investigation. *Frontiers in human neuroscience*, *7*.

Verguts, T., Notebaert, W. (2008). Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. *Psychological Review*, 115, 518-525.

Verguts, T., Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. *Trends in Cognitive Sciences*, 13, 252-257.

CHAPTER 2
MONITORING SELF AND OTHERS: A LITERATURE REVIEW

***Manuscript submitted for publication*[1]**

Speakers can detect errors in their own speech and listeners can detect errors in somebody else's speech. Theories about such monitoring processes differ in whether they assume that self-monitoring in production uses comprehension or whether they are based on production-internal mechanisms (for instance one that would detect conflict between competing representations within the production system). Theories similarly differ in whether they assume that monitoring of production and comprehension employ the same mechanisms or not. Here, we review and discuss these models in light of the key empirical findings in this domain. We point out that current theories have important gaps in their coverage of monitoring phenomena, in particular with respect to the processes that are subsequent to error detection.

[1] This paper was co-authored by Robert Hartsuiker. A shorter version of this chapter together with chapter 6 is submitted for publication.

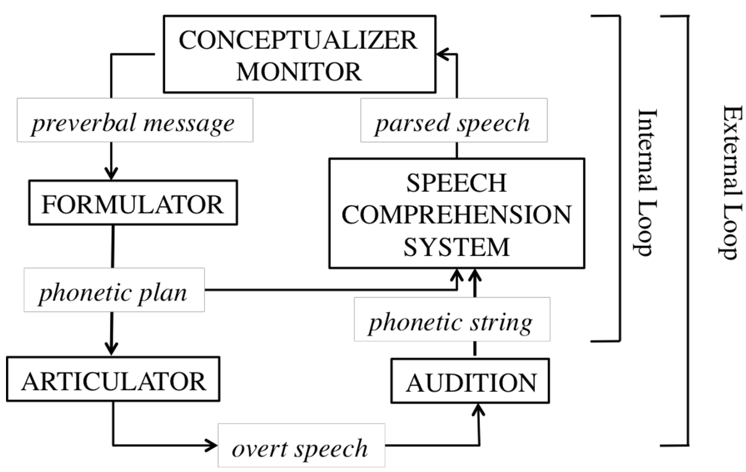
INTRODUCTION

During speech production our speech is constantly and automatically monitored for errors. As a result approximately one out of every ten utterances in naturalistic speech undergoes some form of revision (Nakatani & Hirschberg, 1994). Corpus analyses by Meringer (1908), reanalyzed by Nooteboom (1980; 2005), revealed that 70 – 80% of phonological errors and 50 - 63% of lexical errors are corrected. Similarly, when we listen to somebody else speak, we can also detect errors. Either because what we hear is incompatible with our internal knowledge of the world, such as intention (e.g. a politician says: 'It will take time to restore chaos and order'), or linguistic criteria are violated, such as grammatical agreement (that same politician: 'Rarely is the question asked: Is our children learning?'), or when the speech does not map unto any meaningful message (e.g. 'I felt so strap'). Indeed, without this ability it would not have been possible to collect corpora of speech errors, which have been highly influential in shaping our ideas about language production (e.g., Garrett, 1975). Thus, speech monitoring is a process that takes place in both speech production and comprehension and it is possible therefore that both modalities use similar or identical mechanisms to implement this process. Indeed, as we can listen to ourselves as we talk, it is possible that we use our comprehension system to monitor our production (Levelt, 1983). Monitoring therefore seems a very promising candidate to evaluate the relationship between language production and comprehension. Below we review the properties of error detection, several influential models of speech-monitoring detection, and evaluate their support given empirical findings and theoretical considerations. For any theory of speech-monitoring to be complete, it should account for how errors are detected in speech-production, how errors are detected in speech-perception, and how the detected error is consequently resolved, and this all in accordance with experimental findings and patient data.

THE PERCEPTUAL LOOP THEORY

A highly influential account of speech monitoring is provided by the perceptual loop theory (PLT) (Levelt, 1983, 1989; Indefrey & Levelt 2004; Indefrey 2011; see also Hartsuiker & Kolk, 2001 for an implemented computational version of this account). A speaker hears herself speak, and by perceiving her speech she is able to correct errors. Several observations lead to the additional postulation of an *inner* loop in addition to the external loop (perception of the speech after production), which allows the producer to monitor the utterance before actual production (see Figure 1). A primary observation for postulating the inner loop is that of extremely fast self-corrections. The processes of production, interruption, and repair are too fast for monitoring to take place through the external route of perception. If the external route is used for monitoring, the process of hearing, recognition and interruption are estimated to take between 350 and 400 ms (Levelt, 1989, Marslen-Wilson and Tyler, 1980, Hartsuiker & Kolk, 2001). Measurements of actual interruptions reveal that interruption and onset of the repair happened within 100 ms in almost half of the overt repairs (Blackmer & Mitton, 1991). Even extremely short error-to-cutoff intervals are observed, in which the erroneous item is cutoff almost immediately after initiation ('v-horizontal', Levelt, 1989). Clearly the interruption follows the erroneous production too soon for the interruption to have been processed via production of the phoneme, hearing and processing the phoneme, realization that it is an incorrect phoneme and interrupting production of the incorrect word.

Figure 1. *The Perceptual Loop Model of Self-Monitoring*



The existence of an internal monitor is further supported by studies demonstrating that participants are able to detect produced errors when external speech was not available, as it was masked by a loud noise, in essence forcing participants to use internal monitoring (Lackner & Tuller, 1979, Postma & Kolk 1992a, b). These detections were faster than those with normal feedback, as the PLT would predict. In monitoring via the internal loop, the formulated word is sent to the perception system before articulation. The external loop additionally requires the articulation and perception of the word, which is a longer process compared to internal monitoring. It is thus assumed that monitoring speech via the internal route is faster, leading to faster detections, compared to monitoring via the external route. Additionally, when participants perform a task only using internal speech, they still report the production of errors, demonstrating that indeed internal speech is monitored (Dell & Repka 1992; Oppenheim & Dell, 2008). Also, when presented with a SLIP task in which certain slips would result in taboo utterances (e.g., TOOL – KITS), participants produced fewer of these slips compared with neutral slip utterances (Motley, Camden, & Baars, 1981, 1982). This indicates that the participant made the SLIP internally, and was able to prevent production with a process of covert editing (Motley et al., 1982). It also suggests that top-down influence can be exerted over the monitoring system. The participant really wants to avoid producing taboo utterances, and is indeed able to intercept and repair the taboo utterance quicker compared to neutral slips. This is further supported by an elevated galvanic skin response that was measured in the taboo trials, even when no slip was made. Similarly an fMRI study investigating the neural correlates of inhibition of taboo utterances found increased right inferior frontal gyrus (rIFG) activation on taboo trials compared to neutral trials (Severens, Janssens, Kühn, Brass, & Hartsuiker, 2011); the rIFG is an area of the brain that is thought to play a role in the inhibition of action (Xue, Aaron & Poldrack 2008). However, the amount of top-down control that is exerted over the monitoring system seems to be limited: in spontaneous speech (Meringer, 1908) and in experiments with task relevant speech (Levelt, 1983) the correction rate is similar, while one might expect the participants to want to exert more control in the formal experimental setting compared to spontaneous speech.

Another property of the monitoring system is revealed by the lexical bias effect (LBE). The lexical bias is the tendency for phonological slips to result in an existing word, rather than a non-word (Baars et al., 1975; Dell, 1986; Dell, 1990; Humphreys, 2002; Costa, Roelstraete & Hartsuiker, 2006 and Nooteboom, 2005). This effect is modulated by context; in a non-word context, the LBE disappears (Baars et al., 1975, Hartsuiker, Corley & Martensen, 2005). Thus, like the taboo word effects, one can consider the LBE as the result of covert editing based on a monitoring criterion that is sensitive to the context. Note that the LBE can be viewed as a detection bias, where it would be easier to detect a non-word as erroneous, but also as a correction bias, where some detected errors are corrected more often than others; however, there is no evidence that supports the latter theory (see Hartsuiker, 2006).

The perceptual loop theory by Levelt (1989) thus assumes that speakers can use both external and internal monitoring. In this theory the participant monitors speech by listening to the produced speech (the external route), or via perception of the to be produced speech (the internal route) before production. In the internal loop, a phonemic representation is fed into the speech comprehension system (Wheeldon & Levelt, 1995). This assumption makes it a highly parsimonious account; the model assumes one system, the perception system, which is necessarily there, by which error detection takes place after production and during comprehension. No system outside the language system is needed to detect language errors.

The perceptual loop theory assumes that after detection of an error, via the internal or external route, speech is immediately halted and a restart is initiated (Nooteboom, 1980). This was based on the observation that interruptions do not follow word boundaries but seem to be instantiated immediately after error detection (an exception to this observation are “appropriateness repairs” (e.g. ‘a glass’ followed by the repair ‘a tall glass’), which are often delayed until the end of a word, Levelt, 1983). Further support for this ‘main interruption rule’ comes from a study by Brédart (1991) showing that short words are more often completed before interruption than long words.

A computational test of the theory was performed by Hartsuiker and Kolk (2001). With these simulations Hartsuiker and Kolk tested whether the observed short error-to-cutoff and cutoff-to-repair intervals were possible in a model using the perceptual loop for monitoring. This was indeed possible if interruptions and repairs were assumed to immediately follow the detection of errors and if they were planned in parallel. The computational model was even able to simulate the effect of speech rate on error-to-cutoff and cutoff-to-repair intervals. These simulations showed that error correction via perception is fast enough to explain the short error-to-cutoff intervals, but only with a working inner loop. Importantly, when the inner loop in the model was lesioned, the error-to-cutoff intervals were much longer than in the empirical data.

Evidence in support of similar monitoring for internal and external speech comes from experiments showing similar distributions in detecting semantic and phonological errors in overt and covert speech (Dell 1978; Dell & Repka 1992; Postma & Noordanus, 1996). Further support for the link between internal and external monitoring is provided by the phoneme-monitoring task (Özdemir, Roelofs, and Levelt, 2007). In this inner speech task, an influence of the (perception-specific) uniqueness point effect was observed. The uniqueness point is the phoneme in a word where it diverges from all other words in the language. This uniqueness point influences speed of word recognition: words with an early uniqueness point are recognized faster compared to words with a later uniqueness point (Marslen-Wilson, 1990). In the Özdemir et al. (2007) experiment participants were presented with a phoneme, for instance 'd', followed by the presentation of a picture, e.g. 'panda'. The participants were asked to press a button when the name of the depicted object contained the presented phoneme, without speaking out loud. The target phoneme was varied with respect to serial position (initial, medial, final - /s/ in 'sigaar' 'hamster' and 'cactus') and distance to uniqueness point (no, short, long distance to the italicized uniqueness point – /r/ in 'tijge*r*' 'mot*o*r' and 'do*k*ter'). There was an effect of serial position and of uniqueness point; RTs were smallest to initial phonemes compared to medial and final phonemes and RTs were smallest in the long distance condition compared to short and no distance (once the uniqueness point had passed, the words were processed faster). As there was an effect of a perceptual variable on the self-monitoring of inner speech, one might consider this finding as evidence for the perceptual loop theory. Note however that there are problems with viewing the experiment as an operationalization of the perceptual loop theory, as only inner speech in the absence of external speech was tested (in contrast to normal speech production).

Neuroimaging evidence for the perceptual loop theory comes from studies showing increased activation in neural structures active in speech perception, such as the superior temporal gyrus (STG), in response to a manipulation in which speech feedback is altered in dimensions such as frequency or time (McGuire et al, 1996; Hirano et al, 1997; Hashimoto & Sakai 2003; Christoffels et al, 2007, 2011; Tourville et al, 2008; Zheng et al, 2010; Takaso et al, 2010, Shergill et al., 2002). This has been taken to be evidence of the involvement of the perception system in speech monitoring during production (Indefrey & Levelt, 2004; Indefrey, 2011). It is, however, debatable if these data are evidence for the perceptual loop theory. Specific about the perceptual loop theory is that it assumes that *internal* speech is processed via the perceptual system. The neuroimaging studies above merely point out that altered *external* speech is processed via the perceptual system. In our view, this is not compelling evidence for the perceptual loop theory, as it does not clarify anything about the internal route.

Two main arguments for the perceptual loop theory are formed by the parsimoniousness of the theory, and the computational simulations performed by Hartsuiker and Kolk (2001). First, the model assumes one system, the perception system, which is necessarily there, by which error detection takes place after production and during comprehension. No system outside the language system is needed to detect language errors, which makes the system very parsimonious. The second argument is a computational test of the theory performed by Hartsuiker and Kolk (2001). With these simulations Hartsuiker and Kolk tested whether the observed short error-to-cutoff and cutoff-to-repair intervals were possible in a model using the perceptual loop for monitoring. This was indeed possible if interruptions and repairs were assumed to immediately follow the detection of errors and if they were planned in parallel. The computational model was even able to simulate the effect of speech rate on error-to-cutoff and cutoff-to-repair intervals. These simulations showed that error correction via perception is fast enough to explain the short error-to-cutoff intervals, but only with a working inner loop. Importantly, when the inner loop in the model was lesioned, the error-to-cutoff intervals were much longer than in the empirical data.

COMMENTS ON THE PERCEPTUAL LOOP THEORY

Several criticisms can be raised against this form of perception-based monitoring, both theoretical and empirical. Large portions of the empirical findings that argue against the perceptual loop theory come from patient data. These findings are discussed in a separate section.

Theoretical issues

First of all, both the inner and outer loop recruit the perception system so that this system deals with two versions of the same signal with a small temporal delay. Nevertheless, speakers do not report the perception of overt speech as an "echo" of inner speech (Vigliocco & Hartsuiker, 2002; Nozari et al., 2011). One theoretical solution would be to assume that one of the channels remains unperceived as a result of selective attention. However, this idea is not supported by data of error detection rates. Error detection rates are consistently found to be higher in one's own overt (normal) speech, compared to masked feedback (where the participant can only monitor internal speech) and compared to the detection of errors in speech produced by others (where only the external monitoring route can be used), suggesting that in normal speech both the internal and external route are attended.

Second, the PLT leaves the process of comparison rather underspecified. That is, it assumes that the output of the comprehension system, "parsed speech", is fed back into the system that created the message for production (the "conceptualizer") and that comparison takes place at that level. It is unclear, however, what kind of representation of intended speech can be compared with what kind of perceived speech. It is unclear, for instance, whether "parsed speech" is still phonologically specified, and if so whether that would be useful for comparison at the message level.

Empirical evidence again the PLT

There is no proper evidence for perceptual effects in internal speech monitoring in the presence of external feedback. In a series of experiments in which perception specific effects in inner speech were tested in the presence of external speech, no inner speech effects were observed (Huettig & Hartsuiker 2010; Gauvin, Hartsuiker & Huettig, 2013; see Chapter 3). In these visual world eye tracking experiments participants were presented with a display with four elements that related in a different way to the target word they had to produce or heard: one was the target that had to be named or which they heard, one was semantically related, one was phonologically related, and one was unrelated. Around the onset of the target word production, eye-movements were driven to the item on the display identical to the target, but after a while a substantial amount of the eye-movements was driven to the phonological competitor in the display (more so than to the semantic and unrelated items). Self produced speech drove eye-movements to a phonological competitor in the display in the same time frame as speech produced by others, both between 200 and 400 ms. If indeed participants were able to perceive an internal production before an external one, as hypothesized by the perceptual loop theory, earlier eye-movements to the phonological competitor should have been observed.

A second empirical issue is that there is abundant evidence that during perception expectancies are created. During speech perception, anticipatory eye-movements are observed to relevant or related items (Altmann & Kamide 1999; Kamide, Altmann & Haywood, 2003). And the N400 ERP component is known to show amplitude alternations directly related to the expectancy of the heard word (for an extensive review, see Kutas and Federmeier, 2011). This indicates that speech perception entails more than extracting the meaning out of perceived speech, but also entails a predictive component. The perceptual loop theory does not assume a role for production during perception, but only for perception during production.

A third criticism comes from research showing a big role for the external route in monitoring, calling into question the importance of an internal monitoring route (Lind, Hall, Breidegard, Balkenius, & Johansson, 2014a,b). In an experiment real-time speech exchanges of productions in a Stroop task were auditorily presented through headphones (e.g., presentation of 'green' through headphones when the participant produced 'grey') (Lind et al., 2014a,b). In some cases the replacement words that the participants heard, were accepted by the subject as their own production. This led the authors to conclude that we use our external channel to specify the meaning of what we are saying. While this conclusion may be formulated a bit strongly, see Meekings et al. (in press) for a critical review, the findings of this paper suggest that he external monitoring channel is very important and might sometimes overrule the internal monitoring channel.

Fourth, the PLT relies on comprehension for error detection. Nevertheless, there is ample evidence for a dissociation between error detection in language production and perception. Particularly studies with brain-damaged patients, which we turn to in the next section, provide a considerable amount of data with differences in performance for error detection in production and perception.

Verbal monitoring in patients

A number of studies have shown patients with a combination of defective self-monitoring during production with intact comprehension. Butterworth and Howard (1987) describe three patients that show neologistic speech (frequent use of phonologically related and unrelated non-words), but who have good comprehension. Here internal self-monitoring was impaired in spite of good comprehension. In a study with Parkinson's patients, Alzheimer patients and healthy controls (McNamara, Obler, Au, Durso & Albert, 1992) it turned out that the patients with dementia had poor comprehension and showed poor self-monitoring, and healthy controls showed good self-monitoring and have good comprehension. Interestingly, the Parkinson's patients showed relatively poor self-monitoring, despite intact comprehension skills (but see Chapter 5 of this thesis).

Other patient studies have demonstrated a dissociation between self-monitoring performance and performance on perceptual tasks. Miceli, Gainotti and Caltagirone et al. (1980) studied 69 aphasics, and found no relationship between the degree of phonemic output disorder and the number of phonemic discrimination errors as tested by a phoneme discrimination test and a standard aphasia battery comprising of a verbal sound and meaning discrimination test and a sentence comprehension test. Some of the patients with the most severe output disorder had no discrimination problems. And some patients with a less severe output disorder were incapable of performing the phonemic discrimination in the perception task. Nickels and Howard (1995) examined a group of 15 aphasic patients with phonological errors in production, and found no correlation between the proportion of phonological errors in naming and their performance on a series of comprehension tasks. Also self-monitoring behavior, proportion of attempted error corrections, showed no relation with their performance on auditory comprehension. However, reanalysis of these data by Roelofs (2005) showed that performance on the homophone task was negatively correlated with the number of semantic errors, and positively correlated with phonological self-corrections and false starts. So at least for phonological processing production and perception skills were correlated. The patient studies described above demonstrate that intact perception and monitoring of someone else are possible when self-monitoring is impaired, and that intact comprehension is not sufficient for self-monitoring.

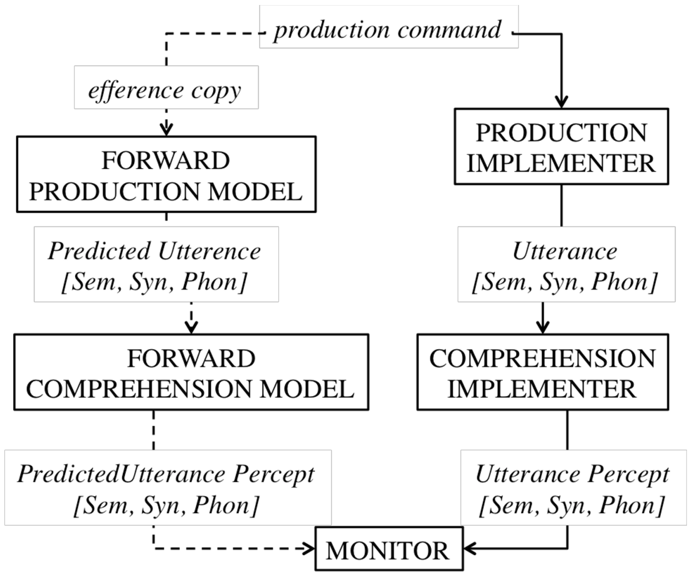
One particularly interesting case of a dissociation between monitoring in production and perception is described by Marshall, Rappaport, and Garcia-Bunuel (1985). They reported the case of a woman with severe auditory agnosia. There was a near-total loss of the ability to understand speech and non-speech sounds even though her hearing was physically intact. Despite her loss of the ability to understand speech, she corrected and attempted to correct many of her phonemic errors, while she ignored her semantic errors. These findings suggest that a) self-monitoring is independent of sound perception and b) semantic and phonemic monitoring can occur independently. Relatedly, Oomen, Postma, and Kolk (2005) described a patient with Broca's aphasia, G., who relied heavily on an internal channel for self-monitoring (when external feedback was masked by white noise, self-monitoring performance remained the same, whereas in the healthy controls self-monitoring decreased). Furthermore G. produced many phonological errors, after which often multiple attempts for repair were made and that were only successfully repaired 38% of the time. Semantic errors were produced far less frequently, and these were successfully repaired in 64% of the trials. In the perception task, G. repaired fewer semantic errors (60%) compared to controls (89%). Interestingly, the percentage of phonologic errors repaired was similar to controls (84% vs. 86%). So while semantic monitoring is impaired in both production and perception, phonological monitoring is only impaired in production. Importantly, this finding further suggests that monitoring can be impaired separately for semantic and phonological processing. Marshall et al. (1998) observed subjects who had preserved comprehension, but impaired self-monitoring. Most interestingly, some patients showed, despite their deficit in self-monitoring, successful monitoring of someone else's speech. This result suggests that self- and other-monitoring occurs via different processing routes. Taken together these patient data show that self-monitoring and other-monitoring can be selectively impaired at the semantic and phonological level, and that intact comprehension and intact other monitoring are not sufficient for correct self-monitoring.

These data are problematic for the Perceptual Loop Theory, as internal self-monitoring, external self-monitoring, and comprehension would all be performed by the comprehension system, thereby assuming a tight link between these processes. In this model it is unclear how self-monitoring would be impaired if comprehension is good. The finding that semantic and phonological monitoring can be impaired independently is also problematic for the perceptual loop theory, as these two processing steps would not be monitored separately. In the PLT either the phonetic plan is monitored via the internal loop, or the spoken word is monitored via the external loop. There is no separate monitoring at the semantic and phonological level.

MONITORING BY FORWARD MODELS

A different model for monitoring that also proposes a strong link between production and perception is the forward modeling account that has recently been proposed by Pickering and Garrod (2013a, 2014), and is based on Wolpert's proposal from computational neuroscience (Davidson & Wolpert, 2005; Wolpert, 1997). Wolpert's theory was designed to explain movement in motor theory, and considers forward internal models that predict the consequences of actions as a central aspect of motor control and learning. In Pickering and Garrod's forward modeling theory of language production, a "prediction of the production" is created at each step of the production process, at the semantic, lexical, and phonological level, based on one's intentions and production outcomes in the past. Each utterance starts with an action command. From this command two processing streams start. The first goes through an action implementer to create a speech act. Next, this act goes through a perception implementer to create a percept. The second stream goes through a forward action model to create a predicted act. This predicted act goes through a forward perceptual model to create a predicted percept. The percept and predicted percept are then compared in a comparator. Comparison takes place sequentially for each level of language production; semantic representations can therefore be compared earlier than phonological ones (see Figure 2). Small differences between the two, could be resolved by updating the prediction whereas big differences between predicted and actual utterance percept, would require an adjustment of the production. Importantly, this mechanism is similarly applied to speech produced by others. We use prediction-by-simulation to predict upcoming words via our own speech production system. Similar to our own speech production the predicted utterance percept, created internally by the listener, is compared to the actual utterance percept. Any deviations will lead to an updating of the prediction of the upcoming utterance.

Figure 2. *The Forward Model account of Self-Monitoring. Sem is the semantic representation, Syn is the syntactic representation, and Phon is the phonological representation.*



There is indeed abundant evidence that prediction plays a central role in in language processing. For instance, Altmann and Kamide investigated anticipatory eye-movements during sentence perception (Altmann & Kamide 1999; Kamide, Altmann & Haywood, 2003). When presented with a visual display, eye-movements are directed towards the picture describing a predicted sentence ending. For instance the sentence fragment 'The man wanted to ride' elicited eye-movements towards the picture of a bike, whereas 'The girl wanted to ride' elicited eye-movements towards a picture of a carousel. In an EEG study, Delong et al (2005) showed that violation of expectancies can already be observed in response to a determiner when it does not match the expected noun. For instance the sentence 'The day was breezy, so the boy went out to fly an ...', with a high cloze probability for 'a' (86%) 'kite' (89%), elicits an N400 in response to the incorrect determiner 'an' for the expected noun 'kite'. For an overview of N400 effects to expectancy violations, see Kutas and Federmeier (2011). Similarly, Van Berkum et al. (2005) reported EEG evidence showing that Dutch listeners respond to the grammatical gender-incongruency of an adjective with a noun they expect in the context.

A second assumption of the forward modeling account is that production is the source of predictions in language comprehension. While for motor control there is abundant evidence to suggest that indeed predictions are used to detect errors in the movement, so far relatively little evidence exists to support this claim for language. One piece of evidence comes from Mani and Huettig (2012) who showed that the productive vocabulary size in young children was correlated with prediction skills in sentence perception. However, there is no direct evidence to support the claim that predictions (the forward model) are used for error detection in perception of speech produced by others.

COMMENTS ON THE FORWARD MODEL THEORY

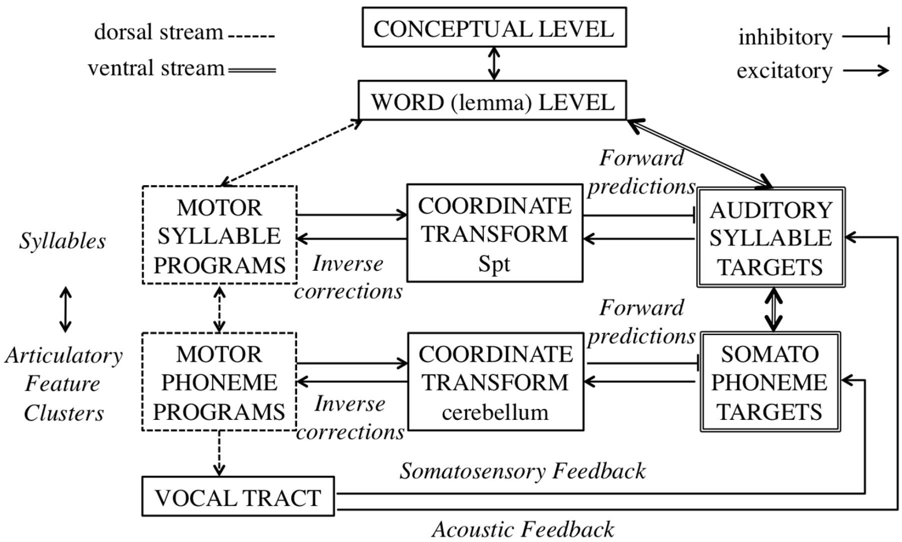
The main criticism on the Forward Model Theory is that it is not very parsimonious. The utterance is produced twice; once as the intended product and once as a prediction (forward model). It is unclear what the advantage is of producing the same utterance twice, especially as the forward model is an impoverished version of the utterance (e.g., Bowers, 2013; Strijkers et al, 2013; Meyer & Hagoort 2013). If in the monitor the utterance percept is compared to the predicted utterance percept (forward model) for error detection, the forward model needs to be correct. To allow for speedy processing, the forward model needs to be a reduced form of the full utterance. But if the forward model is always able to correctly and extremely quickly predict the outcome, why does there need to be a slow construction process following it (e.g., Hartsuiker, 2013)? It is also unclear why the forward model would be better than the utterance percept. The assumption is that the forward model is an impoverished form of the percept, which makes it unclear how aspects of the percept are corrected that are incorrect but not part of the forward model. If the forward model, for instance, does not completely specify all the phonological details, perhaps voicing is not specified, than how can an error in voicing be detected by a comparison between the percept and the forward model? Meyer and Hagoort (2013) specify a range of related issues regarding the forward modeling account. For instance, while the role for prediction in comprehension is quite clear, this is not so for production. Construction of a prediction is an especially useful tool to plan upcoming events that have some degree of independence. The question is whether it is still a useful tool if the prediction in the form of a forward model and the construction come from the same mechanism. Pickering and Garrod reply (2013b) that prediction in the form of a forward model is useful in production, because they assume that the processes that take place in the production implementer are subject to internal noise or priming. These processes are hypothesized not to affect the forward modeling process.

Another potential problem for the forward model theory is its difficulty to account for differences in error detection during production and perception of someone else's speech, as found in some of the patient studies discussed above (see also Hartsuiker 2013). Such differences seem incompatible with a system predicting upcoming productions that is identical during production and perception.

THE HIERARCHICAL STATE FEEDBACK CONTROL MODEL

An alternative to Pickering and Garrod's forward modeling theory is the hierarchical state feedback control model (HSFC) proposed by Hickok (2012). This model deals specifically with the interplay between phonemes, the articulatory system and the auditory system. Like Pickering and Garrod's account, it assumes an important role of forward models. The HSFC assumes a lexical-conceptual system and a phonological system, which is split into components; a sensory input system and a motor output system (see Figure 3). The sensory input is processed via the ventral stream, which uses the superior and middle temporal lobe, and processes the signal for comprehension. This stream is an interface between sensory-motor representations. The motor output is processed via dorsal stream, situated in the posterior planum temporale and posterior frontal lobe, which translates acoustic speech signals into articulatory representations, and forms an interface between auditory and motor representations of speech. These two systems each have their own forward prediction.

Figure 3. *The Hierarchical State Feedback Model of Self-Monitoring*



Furthermore, these two streams are divided into two levels; a higher level that codes speech at a syllable level, and a lower level which codes speech at the phoneme/ feature cluster level. A sensory motor translation system is instantiated for both levels; at the lower level the cerebellum mediates between the two processing streams, at the higher level mediation between the dorsal and ventral stream is performed in the sylvian parietal temporal (Spt) area, located within the Sylvian fissure at the parietal-temporal boundary.

Activation of an auditory speech form automatically activates the corresponding motor program, regardless of whether there is an intention to speak. The lexical level activates the target of a speech act, and the associated motor phonological representation. To make sure that the activated motor representation will hit the auditory target, interactions between the two streams occur. The auditory target then activates the motor representation, which further increases motor activation. The activated motor representation sends an inhibitory signal to the auditory target. When the prediction and the detection match, so no correction is needed, the inhibitory motor-to-sensory efference signal turns off the sensory representation, so that it no longer functions as a correction signal. If an incorrect motor program is selected, the correction signal remains active (as a non-target in the sensory system if activated) and will continue to work towards activating the correct motor representation. Internal monitoring takes place in an early phase; errors in motor planning fail to inhibit the correction signal of the sensory representation. External monitoring takes place in a later phase; suppression of the sensory representation enhances the detection from deviation from expectation.

Evidence for the HSFC model comes from studies showing efference copy effects during mental imagery. Tian and Poeppel (2010) made MEG recordings while participants covertly produced. Around 170 ms after motor estimations by the participants, a response in the auditory cortex was recorded. In a follow-up study Tian and Poeppel (2013) demonstrated context dependent modulations of the auditory cortex to internal simulation. Participants produced speech, imagined speaking and imagined hearing. After this they perceived an auditory syllable probe. The neural responses to the probe, as measured by MEG, varied systematically as a function of the preceding process; repetition suppression was observed in response to (overt or imagined) perception, and an enhancement was observed in response to (overt or covert) production. These data are highly suggestive of internally generated representations that guide subsequent perception in a functionally specific manner.

COMMENTS ON THE HIERARCHICAL STATE FEEDBACK CONTROL MODEL

The hierarchical state feedback control model also suffers shortcomings. While the HSFC model provides a very elegant and neurological based description of phonological processing and motor processing, it does not take semantic and syntactic processing into account. Because the scope of the theory is restricted to phonological processing, it is unclear whether and how an extended version of model could apply to the earlier processing stages of speech production. Especially for the selection of grammatical structure or semantic items it is difficult to imagine how the model would apply, as no sensory feedback is available. One possibility is that these levels operate independently. There is ample evidence to suggest a dissociation between semantic and phonological processing, see for instance the patient data above.

Second, the model only handles speech production but offers no account for monitoring perception. One possibility would be that during perception a prediction is made of the upcoming words, as proposed in the forward model theory by Pickering and Garrod (2013a, 2014). Indeed, a suggestion is made to the application in perception: „it [the inhibitory input to sensory systems] provides a mechanism for explaining the influence of the motor system on the perception of others' speech" (p.8 of Hickok 2012). If indeed motor representations were activated in speech perception, the model would be subject to the same criticisms that Hickok himself has on the forward model account of Pickering and Garrod (Hickok, 2013); sensitivity would be decreased for the perception of someone else's speech.

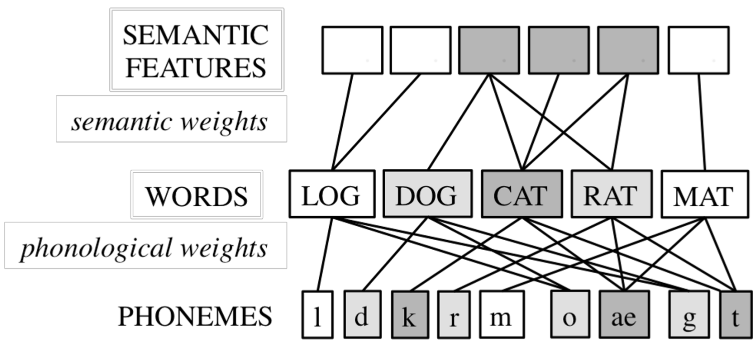
A third caveat is that this model also leaves implicit what happens after an error is detected. It explains how a correction signal arises, but there is no indication of how the correction signal contributes to the interruption of the incorrect utterance and the instantiation of the repair.

Furthermore, self-induced auditory suppression is observed in response to simple utterances, as predicted by forward model theories of self-monitoring. However, some research suggests that the amplitude of the auditory suppression decreases in the presence of more complex feedback (Ventura, Nagarajan & Houde, 2009). Greatest auditory suppressions were observed for simple utterances, such as /a/, in comparison dynamic utterances, such as /a-aa-a/. The authors contribute this difference to the presence of an internal representation for the simple utterance type, which might be absent for the complex utterance type. How this outcome relates to real words remains to be rested.

CONFLICT MONITORING

A final type of account is the conflict monitoring account of Nozari, Dell and Schwartz (2011). In contrast to the models above, it does not necessarily assume a parallel between production and perception. The model builds on domain-general theories of error detection and conflict resolution (e.g., Botvinick et al. 2001; Mattson & Baars 1992; Yeung et al., 2004). The conflict monitoring model proposes that speech monitoring takes place by determining the conflict between response options in a representational system, where conflict can be seen as a function of the activation levels of units representing these options. In case of an error, there are typically multiple units with high activation whereas correct productions are characterized by only a single highly active unit. The conflict information is then relayed to a domain-general executive center (which might well be localized in the anterior cingulate cortex [ACC]). Nozari and colleagues extended Dell and colleagues' (e.g. Dell et al., 2007; Foygel & Dell, 2000) production model with assumptions about conflict monitoring. The model assumes a layer of lexical nodes and a layer of phoneme nodes that are connected via reciprocal connections (see Figure 4). Because of noise in the system and an interplay of different nodes sending and receiving activation, other units than the target one may be highly active (i.e., there is high conflict). Simulations with the model showed that on trials in which the model produced an error, a measure of this conflict was typically much higher than on trials in which the model was correct, suggesting that conflict is a useful measure for error detection. In other words, a high amount of conflict is diagnostic for the occurrence of an error. As conflict is a layer specific mechanism, conflict detection can also be layer specific, and so it is also clear where (in which layer) there is need for conflict resolution.

Figure 4. *Conflict Monitoring Model of Self-Monitoring*



A strength of the conflict monitoring model is that the model can also be extended to explain how it would function in sentence context, thereby increasing its ecological validity compared to models that only model single word production. How the model is able to deal with this form of conflict is extensively addressed by Dell, Oppenheim and Kitteredge (2008).

Nozari et al.'s (2011) simulations demonstrated a strong correlation between patients error-detection ability in a picture naming task and how the model characterized their production skill in terms of lesioning of the semantic and phonological weights. In other words, the strength of the semantic and phonological weights in production, predicted accurately what errors were corrected. The fact that the model so accurately simulated these error detection data brings up two interesting points. It first of all calls into question a role for the perception mechanism in monitoring during production, as these simulations - without a role for perception - worked well. Second, lesions at different steps of the model led to different error detection patterns. Therefore there must be monitoring at several stages, and monitoring can also be impaired at those several stages.

In an attempt to distinguish between the different models for error detection, (Nooteboom & Quené, 2013) investigated the relation between the perceptibility of errors in production and perception. Participants performed an identification task on consonants from a large corpus of SLIP-task speech errors (taken from Nooteboom & Quené 2008). This study had two very interesting findings: the first interesting finding is that early detected errors did not suffer from articulatory blending, as reflected in the same RT's for early detected errors for as correct controls. This suggests that in the case of early-detected errors a wrong word is selected for production, which is quickly interrupted and followed by the correct utterance. The detection of these early errors apparently do not result from perceptual unclarity, but rather a delayed conflict resolution. A second interesting finding was that late undetected errors had smaller RT's in the identification task than undetected errors, while on the basis of the perceptual monitoring account we would predict that undetected errors would be harder to identify than late detected errors. Together these findings argue against perception based verbal monitoring.

An important strength of conflict-based monitoring accounts is that they are compatible with monitoring accounts outside the domain of language and speech, in fact outside of the domain of speech error monitoring, there is much evidence for a domain-general conflict monitoring account of human performance. For instance, neural correlates of error detection, a topic we turn to now, show a high degree of overlap between linguistic and non-linguistic processing.

NEURAL CORRELATES OF DOMAIN GENERAL CONFLICT MONITORING

Conflict monitoring studies typically show an ERN/Ne component in EEG research and ACC/pre-SMA activation in fMRI research during both error production and in high-conflict situations. The Ne component is a response-locked error related negativity that is observed 50-150 ms after the initiation of an incorrect response in non-linguistic tasks (Vidal et al., 2000; Falkenstein et al 1990, 1991, 1995; Gehring et al 1993) and in linguistic tasks with a manual response (Ganushchak & Schiller 2006, 2008b) and 50 and 100 msec after vocal onset with overt speech errors (Ganushchak & Schiller, 2008a; Masaki et al., 2001; Riès et al., 2011). The Ne component is observed independently of whether the participant was aware of the error (Endrass, Franke & Kathmann 2005; Nieuwenhuis et al. 2001; Postma 2000; Ullsperger & von Cramon, 2006), and is independent of response modality (Holroyd et al. 1998). The Ne is also observed in response to situations with high amounts of conflict, such as the Stroop and Eriksen flanker task (for an overview see Botvinick et al., 2001), semantic blocking during picture naming (Ganushchak & Schiller, 2008a), in a language decision task with homographs (Van Heuven et al., 2008) and in potentially taboo-eliciting trials in a SLIP task (e.g. ‘tool kits’, where a slip would result in the production of ‘cool tits’) (Severens et al., 2011). After correct trials the negative component is also observed in both non-linguistic tasks (Bartholow et al., 2005; Roger et al., 2010; Vidal et al., 2000, 2003) and in linguistic tasks (Acheson et al., 2012; Ganushchak & Schiller, 2009; Riès et al., 2011; Sebastian Gallés et al., 2006). This negative wave after correct trials peaks earlier and is somewhat smaller than the negative wave after errors, but it is often not observed, perhaps due to technical issues in preprocessing of the EEG data (Acheson et al., 2012; Riès et al., 2011). The amplitude of the Ne is affected by error rate and time pressure; a low error rate induces a larger Ne after incorrect responses in both linguistic and non-linguistic experiments, and time pressure decreases the amplitude (Falkenstein et al., 1996; Gehring et al., 1993; Ganushchak & Schiller, 2006, 2009). The amplitude of the Ne after a correct response is affected by time pressure and certainty of the response (Ganushchak & Schiller, 2009; Sebastian Gallés et al., 2006).

Source localization has localized the ACC/ pre-SMA region as the origin of the ERN component. The ACC region is broadly connected to motor planning and control systems, and has consistently been observed to be active in neuroimaging research during error production and in high-conflict situations (Coles et al. 1998; Dehaene et al. 1994; Falkenstein et al. 1991; Holroyd et al., 1998; Miltner et al. 2003; Van Veen & Carter 2002; Roger et al. 2010; Debener et al., 2005).

The ACC has been shown to be active in a wide variety of tasks, including language, learning and memory, motor control, imagery, dual task performance (for an overview of experiments, see Botvinick et al., 2001). Most of the studies are consistent with the idea that ACC responds to conflict, and there is broad support for the idea that ACC is involved in cognitive control (D'Esposito et al. 1995; LaBerge 1990; Mesulam 1981; Posner & DiGirolamo 1998). Also in language tasks where participants can freely select from multiple responses, in which a high number of items compete for selection, there is a consistent report of increased ACC activation. Thus, the ACC is seen to be more activated in verb generation, both overt and covert, compared to verb repetition or verb reading

(Andreason, et al., 1995; Petersen et al., 1988; 1989; Warburton, et al., 1996; Wise et al., 1991). Phonological and semantic fluency tasks, in which multiple candidates compete for selection, also elicit higher ACC activation, both when produced overtly and covertly, in comparison to repeating the letter-name cue, repetition of auditory presented words and a lexical decision task the response is pre-determined, and thus low in terms of competition for selection (Friston, Frith, Liddle, & Frackowiak, 1993; Frith, Friston, Liddle, & Frackowiak, 1991A; 1991B; Yetkin, et al., 1995). The same effect is observed for stem completion (Buckner et al., 1995).

If conflict monitoring is indeed a domain-general mechanism, one might expect it to have a similar neural signature in error monitoring in both language and non-language tasks eliciting conflict (Nozari et al., 2011; Acheson & Hagoort 2014). However, the ACC is a highly somatotopically organized structure (Chainay et al., 2004) which elicits different signals in response to different tasks. Fan et al (2003) studied the relationship between the types of conflict in various Stroop and Flanker tasks. In the study of Fan et al. (2003) all tasks, Stroop, Flanker, and "hybrid", do lead to prefrontal and ACC activations, but no interaction effects were found. If one assumes a single unified network for processing conflict, effects of error detection (such as the Gratton effect) would not only be observed within tasks, but also between tasks. The lack of such interaction effects suggests that there are either distinct networks for each conflict tasks, or a single network that uses different sites to resolve the conflict. Another possible solution, as suggested by Botvinick et al (2001), is that the ACC only monitors for conflict, but does not execute operations that resolve the conflict.

A recent study (Acheson & Hagoort, 2014) investigated whether indeed cross-task correlations of error detection could be found in the EEG signal acquired during three conflict tasks; the Eriksen flanker task, the Stroop task, and a tongue twister task. No cross-task correlations with the tongue twister task were found. This led the authors to conclude that the different signatures probably did not arise from a single domain-general conflict monitor. This conclusion, however, may be premature for several reasons. Acheson and Hagoort (2014) used two tasks, which both require conflict resolution from the participants, but differ in quite some other aspects. The flanker task consists of congruent (<<<<<) and incongruent (<<><<) trials. In this task a manual response selection conflict needs to be resolved by the participants. The two conditions not only differ with respect to whether response conflict needs to be resolved, but also in terms of the complexity of the visual display; the incongruent condition has a complex visual display, while in the congruent condition the display is always visually easy to process (Hazeltine et al., 2000). Furthermore, the tonguetwister task requires repetition of a list of non-words. These non-word repetitions lead to conflict in phoneme selection and conflict in the execution of articulatory gestures. As each list of four non-words is repeated in the same order three times in a row, memorizing the items might be a profitable strategy for the participants. Memorizing the items allows preparation of the response, thereby decreasing the amount of conflict. If indeed the participants did memorize the items, a memory component is part of this task as well, contrasting with the flanker task. Indeed, it is no surprise that a highly somatotopically organized structure elicits different signals in response to two tasks that differ on quite a few aspects. The findings of Acheson and Hagoort (2014) are very much in line with the study by Fan et al (2003) who also found no significant relationship between the types of conflict in various Stroop and Flanker tasks.

A recent fMRI study investigated the neural correlates of error detection in speech production and perception (Gauvin, De Baene, Brass & Hartsuiker, submitted; see Chapter 4). If indeed error detection during speech production is mediated by the speech perception system as hypothesized by the PLT, STG activation would be observed in error detection during both production and perception. The conflict monitoring account, on the other hand, predicts a role for the ACC as a primary neural correlate of monitoring. Gauvin et al., presented a tongue twister task, which led to the production of a broad range of error types; phonologic, syntactic, and semantic errors were produced. Analysis of the data revealed a network of areas that was active during error detection for both production and perception. There was no evidence for a role of the perception system in error detection during speech production. The observed network, consisting of pre-SMA, dorsal ACC, bilateral IFG, and anterior insula, is one that has consistently been found to be active for error monitoring in the action domain (Rizzolatti et al., 2001; Wicker et al., 2003; Keysers et al., 2004; Iacoboni, 2005; Botvinick et al., 2005, Shane et al., 2008; de Bruijn et al., 2009; Newman-Norlund et al., 2009; Desmet et al., 2013). These results confirm the predictions of the conflict monitoring account of error detection, and are not directly compatible with a perception based account.

COMMENTS ON THE CONFLICT MONITORING THEORY

This conflict monitoring model has several shortcomings. As the theory is production specific, error detection during perception is not within the scope of the theory. It only deals with conflict monitoring during response selection in a production task.

Another theoretical issue that can be raised against the conflict monitoring account is that it leaves many aspects underspecified. For instance, it is underspecified what happens once the conflict is detected. It is unclear whether the ACC responds to an overall level of conflict, or whether it comes into play once the conflict reaches a certain level. Furthermore, it is unclear what exactly the role is of the ACC, and how the conflict can be resolved. Also unclear at this moment is whether there is any neurological plausibility to the way in which the amount of activation of the items is determined, and subsequently how the amount of conflict is calculated. No neurological suggestion is given of how the ACC and the speech production process communicate.

GENERAL DISCUSSION

From the review above, we can conclude that each model has its own specific strengths and weaknesses. None of the current model presented above is able to account for speech monitoring during both production and perception, and give an adequate explanation of how the error is resolved.

Main arguments in support of the perceptual loop theory (Levelt, 1989) are the parsimoniousness of the theory, and the computational simulations by Hartuiker & Kolk (2001) of the main interruption rule. However, there is no real evidence to support a role of the perception system during internal verbal monitoring. On the contrary, there is quite some evidence from patient data to suggest a distinction between monitoring during production and perception.

The forward model theory by Pickering and Garrod (2013a, 2014) is supported by prediction during language perception. However, there is not a lot of evidence to support a role for production during language comprehension. There are several problems with the idea of a forward model, such as the specificity of the forward model, that need to be clarified. At this moment it is also unclear what exactly the benefit is of the forward model. And as the model also proposes a tight link between production and perception, it is also contradicted by the patient data that show a clear dissociation between monitoring production and perception.

The hierarchical state feedback model receives support from neurological data, showing brain responses to expected auditory responses in the absence an actual auditory input. What sets this model apart is the high amount of neurological specific formulation. Somewhat problematic for the hierarchical state feedback control model, is that the scope of the model is quite limited. Only the production of speech falls under the scope. And within the production process, the theory is limited to the selection of phonemes, articulating the sounds and perceiving these. It also provides no clear hypothesis about how the detected error can be resolved.

The conflict monitoring account proposes a domain general, and computational testable model for error detection. The model is able to simulate word production data of patients with different speech pathologies. Furthermore it receives support from the increasing number of publication that show evidence for a domain general monitoring process. However, no explanation is given about how an error can be resolved once it is detected. And in its current form the conflict monitoring account does not explain error monitoring during speech perception. Furthermore, no explanation is given about how an error can be resolved once it is detected.

In the following chapters, chapter 3 and chapter 4, further research into weather we can find evidence for a role of the perception system during verbal monitoring in overt speech production is presented. In chapter three eye-movements in response to self produced speech and speech produced by others is investigated. In chapter 4 the neural correlates of error detection during speech production and speech perception are investigated using fMRI. Chapter 5 investigated the contributions of the relative monitoring channels in a group of Parkinson's patients.

As none of the models presented above can fully explain error monitoring during production and perception, we present a new model for error detection in chapter 6. The model we propose in chapter 6 is an extension of the conflict monitoring account into perception. We also extend the model by proposing an account of what mechanisms come into play to resolve the conflict after it is detected.

REFERENCES

Acheson, D. J., Hagoort, P. (2014). Twisting tongues to test for conflict-monitoring in speech production. *Frontiers in Human Neuroscience*, 8, 206.

Acheson, D. J., Ganushchak, L. Y., Christoffels, I. K., Hagoort, P. (2012). Conflict monitoring in speech production: Physiological evidence from bilingual picture naming. *Brain and Language*, 123, 131–136.

Altmann, G. T. M., Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition,* 73, 247–264.

Andreasen, N. C., O'Leary, D. S., Cizadlo, T., Arndt, S., Rezai, K., Watkins, G. L., ... & Hichwa, R. D. (1995). Remembering the past: two facets of episodic memory explored with positron emission tomography. *The American journal of psychiatry.*

Baars, B. J., Motley, M. T., MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behaviour*, 14, 382–391.

Bartholow, B. D., Pearson, M. A., Dickter, C. L., Sher, K. J., Note Fabiani, M., Gratton, G. (2005). Strategic control and medial frontal negativity: Beyond errors and response conflict. *Psychophysiology*, 42, 33–42.

Blackmer, E. R., Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39, 173–194.

Botvinick M., Jha A.P., Bylsma L.M., Fabian S.A., Solomon P.E., Prkachin K.M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage*, 25, 312-319.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review* 108, 624–652.

Bowers, J. (2013). How do forward models work? And why would you want them? *Behavioral and Brain Sciences*, 36, 349-350.

Bredart, S. (1991). Word interruption in self-repairing. *Journal of Psycholinguistic Research*, 20,123-138.

Buckner, R.L., Raichle, M.E., Petersen, S.E., (1995). Dissociation of human prefrontal cortical areas across different speech production tasks and gender groups. *Journal of Neurophysiology*, 74, 2163–2173.

Butterworth, B., Howard, D. (1987). Paragrammatism. *Cognition,* 26, 1-37.

Chainay, H., Krainik, A., Tanguy, M.L., Gerardin, E., Le Bihan, D., Lehericy, S. (2004). Foot, face and hand representation in the human supplementary motor area. *NeuroReport,* 15, 765–769.

Christoffels, I.K., Formisano, E., Schiller, N.O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Human Brain Mapping*, 28, 868–879.

Christoffels, I.K., Van De Ven, V., Waldorp, L.J., Formisano, E., Schiller, N.O. (2011). The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. *PLoS One*: 6.

Coles, M.G.H., Scheffers, M.K., Holroyd, C. (1998). *Berger's dream? The error-related negativity and modern cognitive psychophysiology*, in: Witte, H., Zwiener, U., Schjack, B., Döring, A. (Eds.), Quantitative and Topological EEG and MEG Analysis. Druckhaus Mayer Verlag, Jena-Erlangen, pp. 96–102.

Costa, A., Roelstraete, B., Hartsuiker, R. J. (2006). The lexical bias effect in bilingual speech production: Evidence for feedback between lexical and sublexical levels across languages. *Psychonomic Bulletin & Review*, *13*, 972-977.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature*, 378, 279 – 281.

Davidson, P.R. Wolpert, D.M. (2005). Widespread access to predictive models in motor system: a short review. *Journal of Neural Engineering,* 2, 313-319.

de Bruijn, E.R.A., de Lange F.P., von Cramon D.Y., Ullsperger M. (2009). When errors are rewarding. *Journal of Neuroscience* 29, 12183-6.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent EEG and fMRI identifies the dynamics of performance monitoring. *Journal of Neuroscience*, 25, 11730–11737.

Dehaene, S., Posner, M. I., Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, 303-305.

Dell, G. S. (1978). *Slips of the mind.* The fourth Lacus forum (pp. 69–75). Columbia, S.C.: Hornbeam Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5, 313–349.

Dell, G. S., Martin, N., Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, 56, 490-520.

Dell, G. S., Oppenheim, G. M., Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and cognitive processes*, 23, 583-608.

Dell, G.S. Repka, R.J. (1992). *Errors in inner speech*. B.J. Baars (Ed.), Experimental slips and human error: Exploring the architecture of volition (pp. 237–262). Plenum, New York.

DeLong, K.A., Urbach, T.P., Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117–1121.

Desmet, C., Deschrijver, E., Brass, M. (2013). How social is error observation? The neural mechanisms underlying the observation of human and machine errors. *Social Cognitive and Affective Neuroscience* 9,427-435.

Endrass, T., Franke, C., Kathmann, N. (2005). Error awareness in a saccade countermanding task. *Journal of Psychophysiology*, 19, 275–280.

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L. (1990). *Effects of errors in choice reaction tasks on the ERP under focused and divided attention*. In C. Brunia, A. Gaillard, & A. Kok (Eds.), Psychophysiological brain research (pp. 192–195). Tilburg, the Netherlands: Tilburg University Press.

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components: II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, 78, 447–455.

Falkenstein, M., Koshlykova, N.A., Kiroj, V.N., Hoormann, J., Hohnsbein, J. (1995). Late ERP components in visual and auditory Go/Nogo tasks. *Electroencephalography and Clinical Neurophysiology*, 96, 36–43.

Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., Posner, M. I. (2003). Cognitive and brain consequences of conflict. *NeuroImage*, 18, 42–57.

Friston, K. J., Frith, C. D., Liddle, P. F., Frackowiak, R. S. J. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *Journal of cerebral blood flow and metabolism*, 13, 5-5.

Frith, C. D., Friston, K. J., Liddle, P. F., Frackowiak, R. S. J. (1991a). Willed action and the prefrontal cortex in man: a study with PET. *Proceedings of the Royal Society of London.* Series B: Biological Sciences, 244, 241-246.

Frith, C. D., Friston, K. J., Liddle, P. F., Frackowiak, R. S. J. (1991b). A PET study of word finding. *Neuropsychologia*, 29, 1137-1148.

Ganushchak, L. Y., Schiller, N. O. (2006). Effects of time pressure on verbal self-monitoring: An ERP study. *Brain Research*, 1125, 104–115.

Ganushchak, L. Y., Schiller, N. O. (2008a). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *Neuroimage*, 39, 395–405.

Ganushchak, L. Y., Schiller, N. O. (2008b). Brain error-monitoring activity is affected by semantic relatedness: An event-related brain potentials study. *Journal of Cognitive Neuroscience*, 20, 927–940.

Ganushchak, L. Y., Schiller, N. O. (2009). Speaking one's second language under time pressure: An ERP study on verbal self-monitoring in German–Dutch bilinguals. *Psychophysiology*, 46, 410-419.

Garrett, M. F. (1975). *The analysis of sentence production*. In G. H. Bower (Ed.), The psychology of learning and motivation (pp. 133–177). New York: Academic Press.

Gauvin H. S., Hartsuiker R. J., Huettig F. (2013). Speech monitoring and phonologically-mediated eye gaze in language perception and production: a comparison using printed word eye-tracking. *Frontiers in Human Neuroscience*, 7, 818.

Gauvin, H.S. De Baene, W. Brass, M., Hartsuiker, R.J. Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. Manuscript submitted for publication, 2015.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, 4, 385–390.

Hartsuiker, R. J. (2006). Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes*, 21, 856–891.

Hartsuiker, R. J. Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the Perceptual Loop Theory. *Cognitive Psychology*, 42, 113-157.

Hartsuiker, R., Corley, M., Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related Beply to Baars, Motley, and MacKay (1975). *Journal of Memory and Language*, 52, 58–70.

Hartsuiker, R. J. (2013). Are forward models enough to explain self-monitoring? Insights from patients and eye movements. *Behavioral and Brain Sciences*, 36, 357-358.

Hashimoto, Y., Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human brain mapping*, 20, 22-28.

Hazeltine, E., Poldrack, R., Gabrieli, J. D. E. (2000). Neural Activation During Response Competition. *Journal of Cognitive Neuroscience*, 12, 118–129.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience,* 13, 135-145.

Hickok, G. (2013). Predictive coding? Yes, but from what source? *Behavioral and Brain Sciences*, 36, 358-358.

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

Holroyd, C. B., Dien, J., Coles, M. G. H. (1998). Error-related scalp potentials elicited by hand and foot movements: Evidence for an output independent error-processing system in humans. *Neuroscience Letters*, 242, 65–68.

Huettig, F., and Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Process*. 3, 347–374.

Humphreys, K. (2002). Lexical bias in speech errors. Unpublished doctoral dissertation, University of Illinois at Urbana-ChampaignIacoboni M (2005) Neural mechanisms of imitation. *Current Opinion in Neurobiology*, 15, 632-637.

Iacoboni, M. (2005). Neural mechanisms of imitation. *Current opinion in neurobiology*, 15, 632-637.

Indefrey, P. (2011) The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, 2, 255.

Indefrey, P., Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144.

Kamide, Y., Altmann, G., Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133-156.

Keysers, C., Wicker, B., Gazzola, V., Anton, J.L., Fogassi, L., Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, 42, 335-46.

Kutas, M., Federmeier, K. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621.

LaBerge, D. (1990). Thalamic and cortical mechanisms of attention suggested by recent positron emission tomographic experiments. *Journal of Cognitive Neuroscience*, 2, 358-372.

Lackner, J. R., Tuller, B. H. (1979). *Role of efference monitoring in the detection of selfproduced speech errors.* In W. E. Cooper & E. C. T. Walker (Eds.), Sentence processing: Studies dedicated to Merrill Garrett (pp. 281–294). Hillsdale, NJ: Earlbaum.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., Johansson, P. (2014a). Speakers' acceptance of Real-Time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science*, 25, 1198-1205.

Lind, A., Hall, L., Breidegard, B., Balkenius, C., Johansson, P. (2014b). Auditory feedback of one's own voice is used for high-level semantic monitoring: the "self-comprehension" hypothesis. *Frontiers in Human Neuroscience*, 8, 166.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behaviors. *Psychological Review*, 89, 483–506.

Mani, N., Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance,* 38, 843.

Marshall, J., Robson, J., Pring, T., Chiat, S. (1998). Why does monitoring fail in jargon aphasia? Comprehension, judgment, and therapy evidence. *Brain and Language*, 63, 79-107.

Marshall, R. C., Rappaport, B. Z., Garcia-Bunuel, L. (1985). Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language*, 24, 297–313.

Marslen-Wilson, W. (1990). *Activation, competition, and frequency in lexical access.* In G. Altmann (Ed.), Cognitive models of speech processing (pp. 148–172). Cambridge: MIT Press.

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

Mattson, M., Baars, B. J. (1992). *Error-minimizing mechanisms: Boosting or editing?* In Experimental slips and human error: Exploring the architecture of volition, ed. B. J. Baars , pp. 263–87.

McGuire, P.K., Silbersweig, D.A., Frith, C.D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, 119,907–917.

McNamara, P., Obler, L. K., Au, R., Durso, R., Albert, M. L. (1992). Speech monitoring skills in Alzheimer's disease, Parkinson's disease, and normal aging. *Brain and Language*, 42, 38–51.

Meekings, S., Boebinger, D., Evans, S., Lima, C., Chen, S., Ostarek, M. Scott, S. (in press) Do we know what we're saying? A response to Lind et al 2014. *Psychological Science*.

Meringer, R. (1908). *Aus dem Leben der Sprache*. Berlin: V. Behr's Verlag.

Mesulam, M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of neurology*, 10, 309-325.

Meyer, A. S., Hagoort, P. (2013). What does it mean to predict one's own utterances?. *Behavioral and Brain Sciences*, 36, 367-368.

Miceli, G., Gainotti, G., Caltagirone, C., Masullo, C. (1980). Some aspects of phonological impairment in aphasia. *Brain and language*, 11, 159-169.

Miltner, W. H. R., Lemke, U., Weiss, T., Holroyd, C., Schevers, M. K., Coles, M. G. H. (2003). Implementation of error-processing in the human anterior cingulate cortex: A source analysis of the magnetic equivalent of the error-related negativity. *Biological Psychology*, 64, 157–166.

Motley, M.T., Camden, C.T. Baars, B.J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 21, 578-594.

Motley, M.T., Camden, C.T., Baars, B.J. (1981). Toward verifying the assumptions of laboratory-induced slips of the tonguethe output-error and editing issues. *Human Communication Research*, 8, 3–15

Nakatani, C.H., Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, 95, 1603–1616

Newman-Norlund, R.D., Ganesh, S., van Schie, H.T., De Bruijn, E.R.A., Bekkering, H. (2009). Self-identification and empathy modulate error-related brain activity during the observation of penalty shots between friend and foe. *Social Cognitive and Affective Neuroscience*, 4, 10-22.

Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex*, 31, 209–237.

Nieuwenhuis, S., Ridderinkhof, K. R., Blow, J., Band, G. P. H., Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, 38, 752–760.

Nooteboom, S. G. (1980). *Speaking and unspeaking: Detection and correction of phonological and lexical errors of speech.* In V. A. Fromkin (Ed.), Errogarrett

Nooteboom, S. G. (2005). *Listening to one-self: Monitoring speech production.* In R. Hartsuiker, Y. Bastiaanse, A. Postma, & F. Wijnen (Eds.), Phonological encoding and monitoring in normal and pathological speech (pp. 167–186). Hove: Psychology Press.

Nooteboom, S. G., Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language,* 58, 837–861.

Nooteboom, S.G. (1969). *The tongue slips into patterns*. In: A. Sciarone et al (eds.). Nomen: Leyden studies in linguistics and phonetics. The Hague: Mouton, pp. 114-132. Also in V.A.Fromkin (ed.), Speech Errors as linguistic Evidence (1973), The Hague: Mouton. pp. 144-156.

Nozari, N., Dell, G. S., Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63,1–33.

Oomen, C. C. E., Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, 30, 163–184.

Oomen, C. C. E., Postma, A., Kolk, H. H. J. (2005). Speech monitoring in aphasia: Error detection and repair behaviour in a patient with Broca's aphasia. *Phonological encoding and monitoring in normal and pathological speech*, 209.

Oppenheim, G.M., Dell, G.S., Schwartz, M.F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition*, 114, 227–252.

Oppenheim, G. M., Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition,* 106, 528-537.

Özdemir, R., Roelofs, A., Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition*, 105, 457–465.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., Raichle, M. (1989). Positron emission tomography studies of the processing of single words. *Journal of Cognitive Neuroscience*, 1, 153-170.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331, 585-589.

Pickering, M.J., Garrod, S. (2013a). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36, 329-392.

Pickering, M. J., Garrod, S. (2013b). Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences*, 36, 377-392.

Pickering, M.J., Garrod, S. (2014). Self- Other and joint monitoring using forward models. *Frontiers in Human Neuroscience,* 8,132-143.

Riès, S., Janssen, N., Dufau, S., Alario, F. X., & Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience*, 23, 1419-1436.

Rizzolatti, G., Fogassi, L., Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience,* 2, 661-670.

Roelofs, A. (2005). *Spoken word planning, comprehending, and self-monitoring: Evaluation of WEAVER++.* In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), Phonological encoding and monitoring in normal and pathological speech (pp. 42–63). Hove, UK: Psychology Press.

Roger, C., Bénar, C. G., Vidal, F., Hasbroucq, T., Burle, B. (2010). Rostral Cingulate Zone and correct response monitoring: ICA and source localization evidences for the unicity of correct-and error-negativities. *Neuroimage*, 51, 391-403.

Postma, A., Kolk H.H.J. (1992a). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech and Hearing Research*, 35, 537–544.

Postma, A., Kolk, H. H.J. (1992b). Error Monitoring in People Who StutterEvidence Against Auditory Feedback Defect Theories. *Journal of Speech, Language, and Hearing Research,* 35, 1024-1032.

Sebastian-Galles N, Rodriguez-Fornells A, Diego-Balaguer R, Diaz B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience,* 18, 1277–1291.

Severens, E., Janssens, I., Kuhn, S., Brass, M., Hartsuiker, R. J. (2011). When the brain tames the tongue: covert editing of inappropriate language. *Psychophysiology,* 48, 1252–1257.

Shane, M.S., Stevens, M., Harenski, C.L., Kiehl, K.A. (2008) Neural correlates of the processing of another's mistakes: a possible underpinning for social and observational learning. *Neuroimage,* 42, 450-459.

Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E. Jr., Murray, R. M., and McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping.* 16, 219–227.

Strijkers, K., Runnqvist, E., Costa, A., Holcomb, P. (2013). The poor helping the rich: how can incomplete representations monitor complete ones? *Behavioral and Brain Sciences*, 36, 374-375.

Takaso, H., Eisner, F., Wise, R.J., Scott, S.K. (2010). The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study. *Journal of Speech Language and Hearing Research* 53,226–236.

Tian, X., Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology,* 1, 166.

Tian, X., Poeppel, D. (2013). The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of cognitive neuroscience*, 25, 1020-1036.

Tourville, J.A., Reilly, K.J., Guenther, F.H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, 39,1429–1443.

Ullsperger, M., Von Cramon, D. Y. (2006). How does error correction differ from error signaling. An event-related potential study. *Brain Research*. 1105, 102–109.

Van Berkum, J.J.A., Brown, M.C., Zwitserlood, P., Kooijman, V., Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443–467.

Van Hest, E. (1996). *Self-repair in L1 and L2 production.* Tilburg: Tilburg University Press.

van Heuven, W. J. B., Schriefers, H., Dijkstra, T., Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex*, 18, 2706–2716.

van Veen, V., Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, 14, 593–602.

Ventura, M. I., Nagarajan, S. S., Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC neuroscience*, 10, 58.

Vidal, F., Burle, B., Bonnet, M., Grapperon, J., Hasbroucq, T. (2003). Error negativity on correct trials: A reexamination of available data. *Biological Psychology,* 64, 265–282.

Vidal, F., Hasbroucq, T., Grapperon, J., Bonnet, M. (2000). Is the "error negativity" specific to errors? *Biological Psychology*, 51, 109–128.

Vigliocco, G., Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin,* 128, 442–472.

Warburton, E., Wise, R. J., Price, C. J., Weiller, C., Hadar, U., Ramsay, S., Frackowiak, R. S. (1996). Noun and verb retrieval by normal subjects studies with PET. *Brain*, 119, 159-179.

Wheeldon, L. R., Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, 311–334.

Wicker, B., Keysers, C., Plailly, J., Royet, J.P., Gallese, V., Rizzolatti, G. (2003). Both of us disgusted in My Insula: the common neural basis of seeing and feeling disgust. *Neuron*, 40, 655-64.

Wise, R., Chollet, F., Hadar, U. R. I., Friston, K., Hoffner, E., Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain*, 114, 1803-1817.

Wolpert, D.M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1, 209-216.

Xue, G., Aron, A. R., Poldrack, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex*, 18, 1923–1932.

Yetkin, F. Z., Hammeke, T. A., Swanson, S. J., Morris, G. L., Mueller, W. M., McAuliffe, T. L., Haughton, V. M. (1995). A comparison of functional MR activation patterns during silent and audible language tasks. *American Journal of Neuroradiology*, 16, 1087-1092.

Yeung, N., Botvinick, M. M., Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111, 931–959.

Zheng, Z.Z., Munhall, K.G., Johnsrude, I.S. (2010). Functional Overlap between Regions Involved in Speech Perception and in Monitoring One's Own Voice during Speech Production. *Journal of Cognitive Neuroscience,* 22, 1770–178.

CHAPTER 3
SPEECH MONITORING AND PHONOLOGICALLY-MEDIATED EYE GAZE IN LANGUAGE PERCEPTION AND PRODUCTION: A COMPARISON USING PRINTED WORD EYE-TRACKING

Frontiers in Human Neuroscience 7, 818.[1]

The Perceptual Loop Theory of speech monitoring assumes that speakers routinely inspect their inner speech. In contrast, Huettig and Hartsuiker (2010) observed that listening to one's own speech during language production drives eye-movements to phonologically related printed words with a similar time-course as listening to someone else's speech does in speech perception experiments. This suggests that speakers use their speech perception system to listen to their own overt speech, but not to their inner speech. However, a direct comparison between production and perception with the same stimuli and participants is lacking so far. The current printed word eye-tracking experiment therefore used a within-subjects design, combining production and perception. Displays showed four words, of which one, the target, either had to be named or was presented auditorily. Accompanying words were phonologically related, semantically related, or unrelated to the target. There were small increases in looks to phonological competitors with a similar time-course in both production and perception. Phonological effects in perception however lasted longer and had a much larger magnitude. We conjecture that this difference is related to a difference in predictability of one's own and someone else's speech, which in turn has consequences for lexical competition in other-perception and possibly suppression of activation in self-perception.

---

[1] This paper is co-authored by Robert Hartsuiker and Falk Huettig

INTRODUCTION

It has been estimated that during speech production, in both conversation and monolog, one out of ten utterances are subject to revision (Nakatani and Hirschberg, 1994). These revisions partly take place after articulation, but there is reason to believe that speech error monitoring also takes place before articulation. Evidence for such a pre-articulatory speech production monitor comes from our capacity to produce extremely fast corrections, even before the error is fully produced (Levelt, 1989). Additionally, corrections are still made when auditory feedback is disrupted, for instance by masking overt speech (Postma and Noordanus, 1996). And even when speech is only produced internally, production errors are still reported (Oppenheim and Dell, 2008). Such pre-articulatory monitoring might affect patterns of speech errors, as shown in studies where participants produce fewer word slips when this slip would result in a taboo utterance or a nonsense word (Baars et al., 1975; Motley et al., 1982; Hartsuiker et al., 2005; Nooteboom and Quené, 2008; Dhooge and Hartsuiker, 2012). In sum, there appears to be an external monitoring channel that monitors speech after articulation, and an internal monitoring channel that monitors speech before articulation.

There are several theories on how the internal speech monitoring mechanism works. One influential theory, the Perceptual Loop Theory (Levelt, 1989) holds that during speech production copies of the created speech plan are sent via internal loops to the speech comprehension system. This takes place at two levels of production, namely the preverbal message (conceptual loop) and the articulatory buffer (inner loop). Wheeldon and Levelt (1995) further suggested that a phonemic representation is fed back to the comprehension system. In essence, the perceptual loop theory assumes one speech monitoring mechanism for both internally and externally produced speech that is based on speech perception. On the other hand, production-based approaches assume monitoring systems that are extrinsic to the perception system (see Postma, 2000, for a review). For instance, several authors have recently suggested monitoring systems based on *forward models* (e.g., Hickok, 2012; Pickering and Garrod, 2013); the speaker creates a prediction (or forward model) of the expected utterance and compares it to the actual produced speech. Additionally, Nozari et al. (2011) argue that a monitor that assesses the amount of conflict within representational layers (i.e., whether only a single representational unit is highly active or whether several units are highly active) would be diagnostic of error trials. All such production monitoring accounts have in common that internal monitoring would be based on mechanisms that are internal to the production system, rather than on the perception of inner speech.

Empirical evidence on the systems responsible for inner monitoring is scarce and inconsistent. Studies with brain-damaged patients have shown dissociations between comprehension abilities and self-monitoring abilities, a finding that appears inconsistent with the perceptual loop theory. A particular striking study (Marshall et al., 1985) reports the case of a patient with the inability to ascribe meaning to spoken words (and even everyday sounds) indicating a profound disorder of comprehension, who nevertheless initiated self-corrections in her speech. On the other hand, a reaction time study with healthy young adults (Özdemir et al., 2007) reported that response times to phoneme monitoring (e.g., push the button if the name of a target picture contains a particular phoneme) depended on the so-called perceptual uniqueness point, a variable that affects speech perception but not production. However, neither type of evidence is fully convincing: it is possible that patients with good monitoring despite poor comprehension have a comprehension deficit at a relatively early perceptual stage and so accurately perceive inner speech (Hartsuiker and Kolk, 2001). Moreover, it is possible that the phoneme-monitoring task is a very poor model of monitoring in overt production where perception of inner speech might interact with the perception of overt speech (Vigliocco and Hartsuiker, 2002).

In a more recent test of the role of the speech perception system in the internal channel of speech monitoring, Huettig and Hartsuiker (2010) conducted an object naming study using a printed word version of the visual world paradigm. In this version of the visual world paradigm, participants view a display of printed words (typically four words, one in each corner) and listen to spoken language. Looks to each of the words are recorded as a function of the spoken stimuli. For instance, McQueen and Viebahn (2007) showed that participants are more likely to look at printed words with names matching the onset of the concurrent spoken word (e.g., the Dutch word *buffer* when hearing *buffel* “buffalo”) than to printed words that are phonologically different or which match at word offset (e.g., *motje* “moth” for *rotje* “firecracker”). These results are consistent with experiments using the picture version of the visual world paradigm (e.g., Allopenna et al., 1998) and several other methods (phoneme monitoring, Connine et al., 1997; cross-modal priming, Marslen-Wilson and Zwitserlood, 1989). In Huettig and Hartsuiker's production study, participants named visual objects that were presented together with three printed words. These printed words were phonologically related, semantically related, or unrelated to the target. Consistent with earlier perception studies using printed words, there were no increased looks to semantic competitors when phonological competitors were co-present in the display (cf. Huettig and McQueen, 2007, 2011). However, similar to perception studies, phonological competitors received significantly more looks than phonologically unrelated distractors.

Importantly, the perceptual loop theory hypothesizes that the internal channel bypasses articulation and low-level auditory analysis. This allows for speech monitoring even before external speech. By skipping articulation, the target reaches the perceptual system between 250 ms before speech onset (Levelt, 1989; Hartsuiker and Kolk, 2001) and 145 ms before speech onset

(Indefrey and Levelt, 2004). In other words, the perceptual loop theory predicts eye-fixations on printed phonological competitor words before participants produce their own speech. If we assume programming and eye-movements to take about 200 ms (Saslow, 1967), one expects the following results: eye-movements to the phonological competitor driven by internal speech should start between 50 ms before speech onset and 55 ms after; looks to the phonological competitor driven by external speech should start from 200 ms after speech onset. Huettig and Hartsuiker's (2010) results showed a phonological competitor effect in the same time range (300 ms post-articulation) as had been found in earlier perception studies (Huettig and McQueen, 2007), leading to the conclusion that listening to your own overt speech is the same as listening to someone else's speech. Because there was no indication that participants listen to their internal speech in overt speech production, their results argue against the perceptual loop theory for speech monitoring.

CURRENT STUDY

There are both theoretical and practical reasons to revisit the claim that listening to self-produced speech is similar to listening to someone else's speech. First, while the similarity of the findings in Huettig and Hartsuiker (2010) and Huettig and McQueen (2007) is striking, they do not constitute a direct comparison between the modalities. Huettig and Hartsuiker only had a production condition, which was compared to results from a perception condition of an experiment with a different setup (Huettig and McQueen, 2007). For example, the former had a display with one target picture and three written competitors and the latter had a visual display with four printed words. Also target words were embedded in a sentence context in the perception condition, while the production experiment required only production of the target word. Thus, we believe a more direct comparison between production and comprehension is needed to establish whether listening to one's own production is really based on one's overt speech only[1].

Second, and perhaps more interestingly, even if we take for granted that listening to one's own speech production is based on overt speech, there might still be differences between listening to one's own overt speech and to the overt speech of somebody else. This is because of an important difference between speech production and speech perception, namely that in speech production one can make much more accurate predictions of what speech will be produced than in perception. Pickering and Garrod (2013), for instance, hypothesized a role for prediction in both comprehension and production. By predicting upcoming words in speech perception, perception processes can take place much faster than if it were dependent on bottom-up processes only. However, predicting someone else's speech is of course associated with more uncertainty about upcoming speech than predicting one's own utterance, which is likely to affect patterns of overt visual attention in a visual world paradigm. In sum, given the sensitivity of eye-movements to linguistic predictions in visual world studies (e.g., Altmann and Kamide, 1999; Weber et al., 2006; Kamide, 2008; Kukona et al., 2011; Mani and Huettig, 2012), one might expect differences between eye-movements driven by hearing one's own voice vs. someone else's voice.

Thus, the current experiment investigated whether there is a role for the internal monitoring channel in speech production. By directly comparing phonologically-driven eye-movements in a visual world paradigm using matched speech perception and production conditions, we tested whether listening to one's own overt speech has similar perceptual effects to listening to someone else's overt speech. In both conditions participants were presented with a display with four written words and auditory stimuli consisting of only a noun in both the perception and production conditions.

METHOD

Participants

Forty participants (8 males, 32 females, aged 17–35) took part in exchange for course credits. Participants were recruited at the psychology department of Ghent University and were all native speakers of Dutch. All reported to have no dyslexia, no hearing problems, and correct or corrected to normal vision.

Materials

We created 72 sets of visual displays. Each display showed four printed words (Font MS Trebuchet, size 20), each in a different quadrant of the screen (Figure 1). Each display consisted of one target word, one competitor word and two unrelated filler words. The words were randomly assigned to a quadrant per trial.

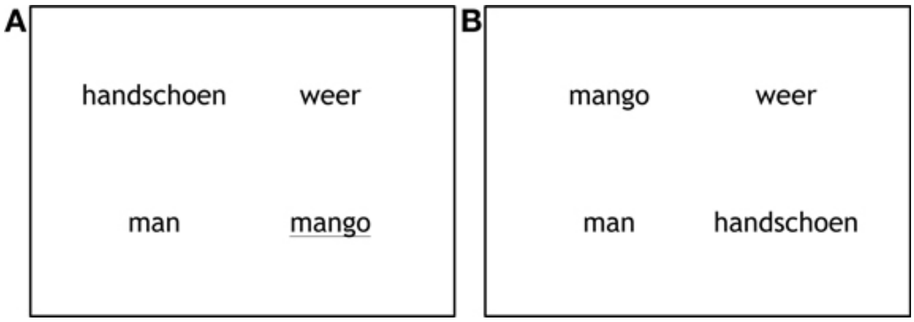


Figure 1. *Examples of the display in the production condition (A) and the perception condition (B).*

There were three conditions: in the semantic condition, the competitor was semantically related to the target; in particular, it came from the same category. In the phonological condition, the competitor shared the onset (from 1 up to 3 phonemes) with the target. In the neutral condition, the competitor was unrelated to the target. Each item was presented as target or competitor in one display, and presented as unrelated filler in another display; for example, targets and competitors in the phonological condition occurred as unrelated items in the neutral condition. Differences in looks to targets and competitors compared to unrelated items can therefore not be the result of intrinsic properties of the items. There were 24 trials in each condition (semantic, phonological, and neutral). The order of the trials was determined randomly. All stimuli were presented once in the production condition and once in the perception condition.

Most words in the phonological and semantic conditions were taken from Huettig and Hartsuiker (2010). For the phonological condition eight word pairs were created that shared a higher CV overlap between target and competitor compared to the original word pairs. In the semantic condition 11 new word pairs were created that (subjectively) had a stronger semantic relatedness. An overview of the stimuli can be found in Appendix A.

Participants filled out a questionnaire on their reading and auditory skills and signed a written consent form. The participants received written task instructions. Next the eye-tracking device was adjusted for each participant and calibrated. The experiment consisted of 12 blocks. At the beginning of each block a calibration of the eye-tracker was performed. During each block six trials of the production condition and six trials of the perception condition were presented consecutively. A picture of an ear (perception) or a mouth (production), displayed for 2000 ms, signaled the task. Each trial started with a fixation cross, followed by a 3000 ms display of the four written words. Displays were randomly assigned to each trial.

In the production trials the target word was underlined and was read out loud by the participant. In the perception condition participants heard the target word after a 200 ms preview of the display. After the experiment participants filled out similarity ratings for the semantic (how well do the words match?) and phonological word pairs (how similar are the word onsets?) on a 5-point scale, with 1 being “not at all” and 5 being “very much.” On average both semantic and phonological word pairs were rated as being between neutral and fairly similar. Semantic word pairs were rated as more similar (*M* = 3.52, *SE* = 0.077) than phonological word pairs (*M* = 2.96, *SE* = 0.083).

Apparatus. Experiments were created in Experiment Builder 1.10.1 (SR Research Ltd. 2004–2010). Eye movements were recorded using an EyeLink 1000 eye-tracker. Speech in the production trials was recorded and speech onsets were measured manually in Praat (Boersma and Weenink, 2012).

RESULTS

Responses in which there was a hesitation (e.g., ‘eh’) or in which the response was produced after the display had disappeared (i.e., after 3000 ms) were excluded from analysis. No other outliers were excluded. This led to a total loss of 1.4% of all the production trials. Errors were fairly equally distributed among the three conditions. Naming latencies were around 1100 ms for all three conditions (Table 1).

Table 1. *Number of errors and mean speech onset.*

Condition	**Errors (%)**	**Onset (ms)**	***SD* onset**
Neutral	1.1	1160	315
Phonological	1.5	1118	296
Semantic	1.7	1169	302

ANALYSIS OF FIXATION DATA

Fixation proportions to targets and related competitors were compared to an average of unrelated neutral competitor words in the respective conditions.

The fixation proportions were calculated for 200 ms timeframes, until 1000 ms after speech onset for both production and perception conditions. To test whether visual attention in the production condition indeed precedes production analysis starts 200 ms before speech onset. Fixation proportions were normalized before averaging per condition using a log 10 transformation, and (as hypotheses were directional) compared in one-tailed paired *t*-tests, with alpha set at 0.05. Effect sizes reported are Cohen's *d.*

For ease of interpretation, reported means, and standard deviations in text, figures, and appendix, are the untransformed values.

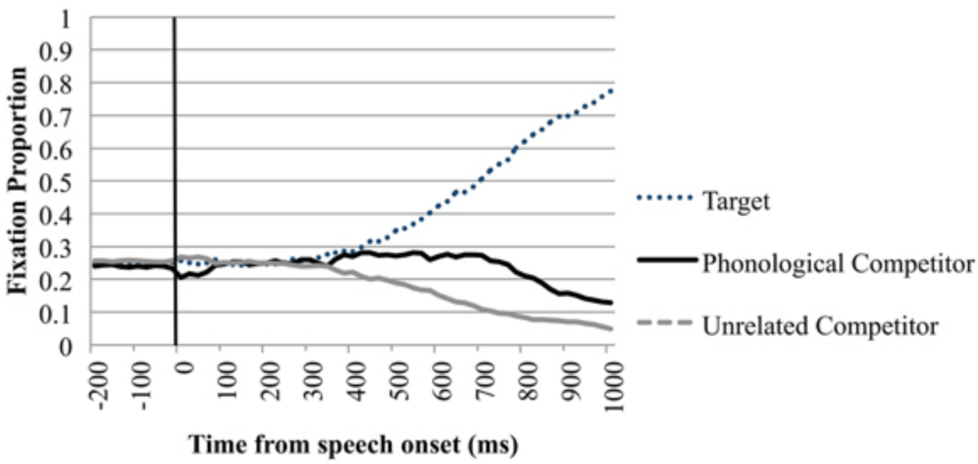
Perception

Fixations on target items. In the perception trials, eye-movements started to diverge toward the target words from around 200 ms after speech onset. Looks to the target in the neutral condition (M = 0.287 SD = 0.398) differentiated significantly from looks to the unrelated words (M = 0.232, SD = 0.360) between 200 and 400 ms after speech onset $t1(39) = 3.420$, $p < 0.001$; $t2(23) = 4.34$, $p < 0.001$; $d = 0.145$. Looks to the target in the phonological condition (M = 0.365, SD = 0.429) diverged significantly from the unrelated items (M = 0.182, SD = 0.333) between 400 and 600 ms after speech onset $t1(39) = 8.753$, $p < 0.001$;$t2(23) = 6.794$,$p < 0.001$;$d = 0.476$. Looks to th target in the semantic condition (M = 0.409, SD = 0.432) also diverged significantly from the unrelated items (M = 0.190, SD = 0.335) between 400 and 600 ms after speech onset $t1(39) = 10.308$, $p < 0.001$; $t2(23) = 8.923$, $p < 0.001$; $d = 0.566$.

Fixations on competitor items. Analysis of fixations to the neutral competitors revealed no significant differences at any time interval. For the semantic condition there was also no significant difference between looks toward the competitor and unrelated items in any of the timeframes.

In the phonological condition there was a similar proportion of looks toward the phonological competitor and the unrelated items −200ms until speech onset and in the timeframe from speech onset until 200ms after this onset. Between 200 and 400 ms after speech onset we observed a 2.3% difference between looks toward the phonological competitor and the unrelated items. This difference did not reach statistical significance; the following timeframes however showed a more robust increase of this difference. This suggests that the phonological competitor effect started to arise in the 200–400 ms timeframe. Looks toward the phonological competitor (M = 0.270, SE = 0.385) diverged significantly from looks toward the unrelated items (M = 0.182, SD = 0.333) between 400 and 600 ms after speech onset $t_{1(39)} = 5.743$, $p < 0.001$; $t_{2(23)} = 4.591$, $p < 0.001$; $d = 0.244$. This effect increased and remained significant throughout the 1000 ms after speech onset. Figure 2 shows the time course probability plots for the phonological condition in the perception trials.

Figure 2. Eye-movements in the phonological trials of the perception condition**.** Proportion of fixations are sorted per quadrant and plotted as a function of time. Time point −200 is the onset of the display. Speech onset of the target word is at 0 ms, as indicated by the black line.



Production

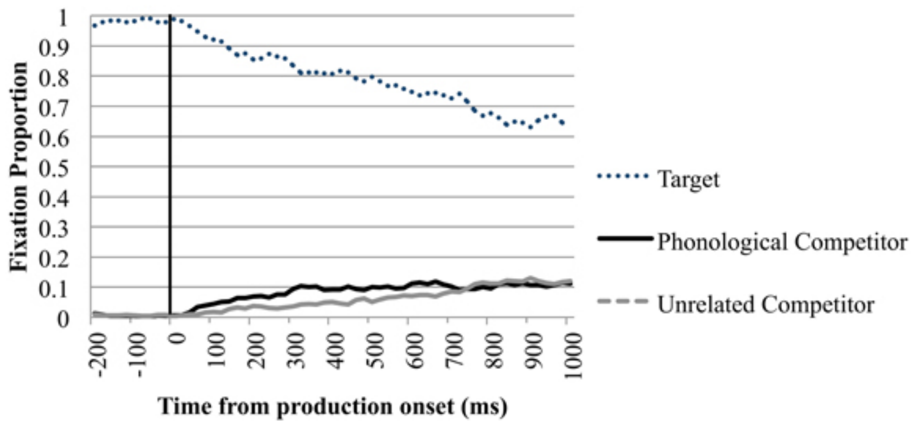
Fixations on target items. At the start of the analysis, 200ms before speech onset, fixations in the production condition were directed significantly more toward the target than to unrelated items in all three con- ditions. The difference between looks to the target items and unrelated items remained significant ($p < 0.001$) throughout the later timeframes.

Fixations on competitor items. In both the neutral and semantic condition there was never a significant difference of fixations between the competitors and unrelated items in all time bins.

Of main interest are the fixations on the phonological competitor. Importantly, between 200ms before and 200ms after speech onset, looks toward the phonological competitor did not differ significantly from looks to the unrelated items. Between 200 and 400ms after speech onset looks to the phonological competitor (M = 0.076, SD = 0.247) diverged from the unrelated items (M = 0.054, SD = 0.214), $t_{1(39)} = 1.717$, $p = 0.047$; $t_{2(23)} = 2.445$, $p = 0.012$; $d = 0.095$. This timeframe is comparable to the 350–500ms after speech onset in which eye-movements to phonological competitors were observed in Huettig and Hartsuiker (2010). In the 400–600ms after speech onset there were more looks to the competitor in the phonological condition (M = 0.096, SD = 0.277) than to the unrelated items (M = 0.080, SD = 0.254) significant in the by- participant analysis $t_{1(39)} = 2.110$, $p = 0.021$; $t_{2(23)} = 1.300$, $p = 0.104$; $d = 0.060$. In the timeframes after 600 ms the phonological competitor did not attract more fixations than the unrelated items. Figure 3 shows the time course of fixation proportions for the phonological condition in the production trials.

In sum, in both production and perception trials eye-movements were directed more to phonological competitors than to unrelated printed words shortly after the critical onset of word perception or word production. The magnitude of the phonological competition effect however differed considerably between production and perception trials. An overview of the fixation proportions in the phonological condition can be found in Appendix B.

Figure 3. Eye-movements in the phonological trials of the production condition. Proportion of fixations are sorted per quadrant and plotted as a function of time. Time point 0 is the speech onset, as indicated by the black line.



GENERAL DISCUSSION

In the present experiment using the printed word version of the visual world paradigm, we observed more looks toward a phonological competitor in both speech perception and speech production. In line with previous studies (Huettig and McQueen, 2007; Huettig and Hartsuiker, 2010), there were no semantic effects using this version of the paradigm. The experiment allows for two main conclusions. First, phonological competition effects in production did not occur in the timeframe predicted by perceptual loop theory. Second, the magnitude and longevity of the effect was considerably larger in perception than in production. This suggests that overt speech is processed differently if it is produced by someone else rather than by oneself.

The speech production condition did not show a robust increase of eye-movements toward the phonological competitor shortly before or around speech onset. Thus, consistent with the results of Huettig and Hartsuiker's (2010), these findings do not support a role of speech perception for the monitoring of inner speech. Given the small effects in this study, however, we must acknowledge the possibility that the absence of early phonological competition effects reflects a lack of sensitivity to internal monitoring processes of our method. Future work should ideally use additional methods to provide converging evidence.

The difference in magnitude of looks toward the phonological competitor between speech production and speech comprehension is, in our view, the result of at least two processes that are both related to a key difference between the processing of speech produced by oneself and by somebody else, namely predictability. The speaker knows what word she is about to say and can thus anticipate hearing a particular word (arguably, one could consider this prediction a forward model). The listener, in contrast, cannot make such reliable predictions. In the specific context of our task, each word on the screen has a 0.25 probability of being spoken, which means that any prediction in perception is likely to be correct on only a minority of trials. In realistic situations, depending on context, the listener may sometimes have much better odds of predicting somebody's speech, but many other times the odds will be much worse. This difference between comprehension and production may have played out in our experiments in two ways.

We conjecture that one factor that contributes to a difference in visual attention in speech production and perception is a suppression of activation of the target in speech production.

Evidence for such a suppression of activation comes for instance from MEG studies of word production (Heinks-Maldonado et al., 2006; Tian and Poeppel, 2011). During production upcoming words are predicted, followed by a suppression of activation of the predicted word. An interesting possibility is that such predictions result in early eye-movements to the phonological competitors (prediction) followed by a lack of activation-related effects (suppression). However, as noted above, there was no evidence for such early competition effects. But we do find a difference in magnitude of effects in production and perception. The suppression of activation of the target word could lead to decreased priming in production compared to perception, reflected by the decreased fixation proportions to the phonological competitor in production compared to perception.

In addition, lack of predictability in perception as compared to production may also affect phonological competition between cohort members. As in previous studies, results in the current experiment show that in perception trials the listener looks at the word with the closest correspondence to the word they hear. In the trial in which there is a phonological competitor word, which shares the onset, participants cannot be sure about which word will be the target until the uniqueness point has passed. The phonological competition is thus at least partly driven by uncertainty about the target word. In contrast, in the production trials the participant can predict which words she will hear herself say with almost complete certainty. Therefore any evidence for phonological competition in the eye gaze pattern is unlikely to reflect uncertainty about the target. Instead, we suggest that phonological competitor effects in production are due to activation spreading in the representational conglomerate that binds together the visuospatial and linguistic elements (Huettig et al., 2011): producing the phonology of one object's name at a particular spatial location primes the phonological representation of a related object as well as a pointer to its spatial location.

To conclude, the present results are most compatible with the view that eye-movements are driven similarly by the perception of one's own speech production and by the perception of someone else's speech. In both cases, overt, and not inner speech, drives the observed eye-movements. However, phonological competition effects in speech production and speech perception are also influenced by distinct processes as evidenced by much larger and much more long-lived phonological competitor effect in perception. In production trials, the target is strongly predicted and target activation is suppressed, whereas in perception, decreased target predictability results in phonological cohort competition. Thus, while it is correct that the channel we use to listen to ourselves and to someone else is the same (i.e., overt speech), because of a profound difference in predictability, the way we listen is different.

REFERENCES

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439. doi: 10.1006/jmla.1997.2558

Altmann, G. T. M., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1

Baars, B. J., Motley, M. T., and MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *J. Verb. Learn. Verb. Behav.* 14, 382–391. doi: 10.1016/S0022-5371(75)80017-X

Boersma, P., and Weenink, D. (2012). *Praat: doing phonetics by computer [computer program]*. Version 5.3.11, (Retrieved 27 March 2012). Avilable online at http://www.praat.org/

Connine, C. M., Titone, D., Deelman, T., and Blasko, D. (1997). Similarity mapping in spoken word recognition. *J. Mem. Lang.* 37, 463–480. doi: 10.1006/jmla.19 97.2535

Dhooge, E., and Hartsuiker, R. J. (2012). Response exclusion as verbal self-monitoring: effects of lexicality, context, and time pressure in the picture-word interference task. *J. Mem. Lang.* 66, 163–176. doi: 10.1016/j.jml.2011.08.004

Hartsuiker, R. J., Corley, M., and Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: related beply to Baars et al. (1975). *J. Mem. Lang.* 52, 58–70. doi: 10.1016/j.jml.2004. 07.006

Hartsuiker, R. J., and Kolk, H. H. J. (2001). Error monitoring in speech production: a computational test of the perceptual loop theory. *Cogn. Psychol.* 42, 113–157. doi: 10.1006/cogp.2000.0744

Heinks-Maldonado, T. H., Nagarajan, S. S., and Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport* 17, 1375–1379. doi: 10.1097/01.wnr.0000233102.43526.e9

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* 13, 135–145. doi: 10.1038/nrn3158

Huettig, F., and Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Lang. Cogn. Process.* 3, 347–374. doi: 10.1080/01690960903046926

Huettig, F., and McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *J. Mem. Lang.* 54, 460–482. doi: 10.1016/j.jml.2007.02.001

Huettig, F., and McQueen, J. M. (2011). The nature of the visual environment induces implicit biases during language-mediated visual search. *Mem. Cogn.* 39, 1068–1084. doi: 10.3758/s13421-011-0086-z

Huettig, F., Olivers, C. N. L., and Hartsuiker, R. J. (2011). Looking, language, and memory: bridging research from the visual world and visual search paradigms. *Acta Psychol.* 137, 138–150. doi: 10.1016/j.actpsy.2010.07.013

Indefrey, P., and Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition. 2002.06.001

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Lang. Ling. Compass* 2, 647–670. doi: 10.1111/j.1749-818X.2008.00072.x

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., and Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition* 119, 23–42. doi: 10.1016/j.cognition.2010.12.002

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Mani, N., and Huettig, F. (2012). Prediction during language processing is a piece of cake - but only for skilled producers. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 843–847. doi: 10.1037/a0029284

Marshall, R. C., Rappaport, B. Z., and Garcia-Bunuel, L. (1985). Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain Lang.* 24, 297–313. doi: 10.1016/0093-934X(85)90137-3

Marslen-Wilson, W. D., and Zwitserlood, P. (1989). Accessing spoken words: the importance of word onsets. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 576–585. doi: 10.1037/0096-1523.15.3.576

McQueen, J. M., and Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Q. J. Exp. Psychol.* 60, 661–671. doi: 10.1080/17470210601183890

Motley, M. T., Camden, C. T., and Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: evidence from experimentally elicited slips of the tongue. *J. Verb. Learn. Verb. Behav.* 21, 578–594. doi: 10.1016/S0022- 5371(82)90791-5

Nakatani, C. H., and Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *J. Acoust. Soc. Am.* 95, 1603–1616. doi: 10.1121/1.408547

Nooteboom, S., and Quené, H. (2008). Self-monitoring and feedback. A new attempt to find the main cause of lexical bias in phonological speech errors. *J. Mem. Lang.* 58, 837–861. doi: 10.1016/j.jml.2007.05.003

Nozari, N., Dell, G. S., Schwartz, M. F. (2011). Is comprehension necessary for error detection. A conflict-based account of monitoring in speech production. *Cogn. Psychol.* 63, 1–33. doi: 10.1016/j.cogpsych.2011.05.001

Oppenheim, G. M., and Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* 106, 528–537. doi: 10.1016/ j.cognition.2007.02.006

Özdemir, R., Roelofs, A., and Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition* 105, 457–465. doi: 10.1016/ j.cognition.2006.10.006

Pickering, M. J., and Garrod, S. (2013). An intergrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 1–64. doi: 10.1017/S0140525X1 2001495

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition* 77, 97–132. doi: 10.1016/S0010-0277(00)00090-1

Postma, A., and Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Lang. Speech* 39, 375–392.

Saslow, M. G. (1967). Latency for saccadic eye movement. *J. Opt. Soc. Am.* 57, 1030–1033. doi: 10.1364/JOSA.57.001030

Tian, X., and Poeppel, D. (2011). "Dynamics of self-monitoring in speech production: evidence from a mental imagery MEG paradigm," in *41th Annual Meeting of Society of Neuroscience* (Washington, DC).

Vigliocco, G., and Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychol. Bull.* 128, 442–472. doi: 10.1037/0033- 2909.128.3.442

Weber, A., Grice, M., and Crocker, M. (2006). The role of prosody in the interpretation of structural ambiguities: a study of anticipatory eye movements. *Cognition* 99, B63–B72. doi: 10.1016/j.cognition.2005.07.001

Wheeldon, L. R., and Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *J. Mem. Lang.* 34, 311–334. doi: 10.1006/jmla.1995.1014

CHAPTER 4
CONFLICT MONITORING IN SPEECH PROCESSING: AN fMRI STUDY OF ERROR DETECTION IN SPEECH PRODUCTION AND PERCEPTION

***Manuscript submitted for publication*[1]**

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 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 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 conflict between response options, and

---

[1] This paper was co-authored by Wouter de Baene, Marcel Brass and Robert J. Hartsuiker

subsequent conflict resolution by a domain general executive center (e.g., the ACC).

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 (Levelt, 1989; Blackmer & Mitton, 1991), 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 proposed 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 inner 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 and Hochberg, 1965).

Presently there are roughly three classes of theories on monitoring inner 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), which combine properties of the other two classes of theory. Perception-based theories assume that inner 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. As a consequence monitoring your own speech, be it internal or external, and monitoring someone else's speech, all take place via the same monitoring mechanism, namely the perception system, which is located in the superior temporal lobe (Price, 2012).

Production-based theories do not necessarily assume the same monitoring system for production and perception, and assume that 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 conflict between response options, which is subsequently resolved by a domain-general cognitive control unit (located in the anterior cingulate cortex).

Forward modeling accounts of speech monitoring assume that during production a prediction, or forward model, of the expected outcome is made. The 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 observations 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 functions as a corrective signal (Curio et al., 2000; Heinks-Maldonado et al., 2005; 2006; Numminen et al., 1999, Eliades & Wang, 2003; 2005). Note, however, that forward modeling theories can differ in important aspects, such as the exact effect a discrepancy between sensory information and the forward model have on further processing.

Most forward model theories, however, rely on sensory feedback for monitoring. Consequently, internal speech monitoring, which is investigated in the current study, is outside the scope of these theories. Pickering and Garrod's model (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 perceived 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.

Because production- and perception-based monitoring make distinct predictions about the functional neuroanatomy of speech monitoring, fMRI can be used to distinguish between these competing theories. Perception-based monitoring assumes 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. Such an account predicts 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 inner 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).

So far few studies have applied fMRI to investigate inner 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 directly compares error detection in speech production and perception to investigate whether internal pre-articulatory monitoring recruits the speech perception system similarly for error detection during perception and production as hypothesized by perception-based monitoring theories of speech processing. Participants performed a tongue twister task in which they repeated tongue twisters, or listened to a recording of a repetition, after which they judged the repetition on correctness. The percentage of errors in the perception condition is 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. 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. Also by having the participant produce overt speech, unlike covert speech, the speech plan is fully formed (Barch et al., 1999; Huang et al., 2001; Gracco et al., 2005). A downside to this approach 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. We therefore had participants produce noise-masked speech that excluded external monitoring, and allowed for verification of their production.

If internal monitoring takes place via (a subpart of) the perceptual system, as hypothesized by perception-based theories of error monitoring, we expect to find activation patterns in perceptual areas similarly for error detection in internal speech in production and error detection in external speech in perception. If, however, internal monitoring does not take place via the perceptual system, as hypothesized by production-based theories of error monitoring, we expect to find perception-independent activations for error detection of internal speech during production.

METHOD

Participants

Twenty-four participants were recruited from Ghent University, of which 3 were discarded; one due to excessive motion, one because of too many errors (>80%) and one due to too few errors (<10%). Final analyses included 21 participants (15 females, 6 males, mean age: 21, ranging from 19 to 30). All reported to be native speakers of Dutch, have no dyslexia or other speech or language impairments, no hearing problems, and normal or corrected-to-normal vision. No subject had a history of neurological, psychiatric or major medical disorder. All subjects were right handed as assessed by the Edinburgh Handedness inventory (Oldfield, 1971) (n=21, EHI score M=90.4, SD=15.8, range = 41.2 to 100, mode = 100). A monetary reward was received for participation. The study was approved by the local ethical committee of Ghent University's Medical Department and was conducted

Materials and design

Stimuli were selected on the basis of a pilot study in which 56 tongue twister 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. Audio files were created in which these 22 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. Actual recordings of 4 female participants producing the sentences in the pilot study, correctly and incorrectly, were used as auditory stimuli. Pitch was adjusted (increased with 50 or 20 Hz) for 3 of the 4 participants to facilitate auditory perception in the 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 consecutively 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. Although this lack of counterbalancing may have disadvantages we felt these were outweighed greatly by the advantage of being able to match the error percentages in perception to that in production. This is of course only possible with a fixed order of the conditions, and allows for a direct comparison could be made between production and perception. An unbalanced distribution of error percentages in the production and perception condition would severely impair the validity of a comparison between error detection in the production and perception condition. The total duration of the experiment was approximately 45 minutes.

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. After a familiarization phase of 3 trials, three target blocks were presented that each constituted of the 22 tongue twister sentences in random order. Between trials a varying ISI of between 1250 and 5500 ms occurred (mean 2867 ms).

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). 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.

Although the button press response for error detection makes the task somewhat less naturalistic, and focuses the attention of the participants on error detection, it was included in the paradigm to measure whether the participant was aware of the error or not. People do not correct all their speech errors (e.g. Nooteboom, 1980), but it is unclear whether uncorrected errors are ones the producer was unaware of or ones where the producer was aware of the error but did not bother to correct it (Berg, 1986). Also, if large numbers of both conscious and unconscious errors had been made, it would have been interesting to investigate whether there is a difference in brain activations between conscious and unconscious error production. However, too few unconscious errors were produced to make this comparison.

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, distance factor = 17%, voxel size 3.5 × 3.5 × 3.51 mm^3, 34 axial slices). Specific care was taken to ensure that frontal areas and (near) complete cerebellum were included in the imaging volume. A varying number of images were acquired per run due to the self-paced ending of trials.

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

While it is generally assumed that overt speech in the scanner will cause large motion and signal artifacts (see Gracco, Tremblay & Pike, 2005 for an overview) we did not find this to be the case in our specific set-up. Instead of using a special scanning procedure (e.g. Eden et al., 1999; Huang et al., 2001; Menenti et al., 2011) or limit volume acquisition to the time interval after speech production, we applied a common acquisition procedure. Nevertheless, motions were well within the 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.

DATA ANALYSIS

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.

General GLM analysis

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

Region of interest analysis

A region of interest (ROI) analysis was performed for each of the brain regions that were identified in the whole brain analysis to investigate the patterns of brain activation. Note that the ROI analysis was not used to determine significance of activation clusters (double dipping), but only to further explore the activation patterns. 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 to further explore the data.

RESULTS

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 her performance were taken into analysis; the incorrectly judged items were discarded from all analyses. Over all 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). In the perception condition participants correctly identified 53% as correct and 27% as incorrect repetitions of the tongue twister. In 19% of the trials the participants made an incorrect judgment (40% misses, 60% false alarms). The striking similarity between the two conditions is the result of online scoring of the production trials, to which the perception trials were matched in percentage of errors.

FMRI DATA

Conjunction analysis.

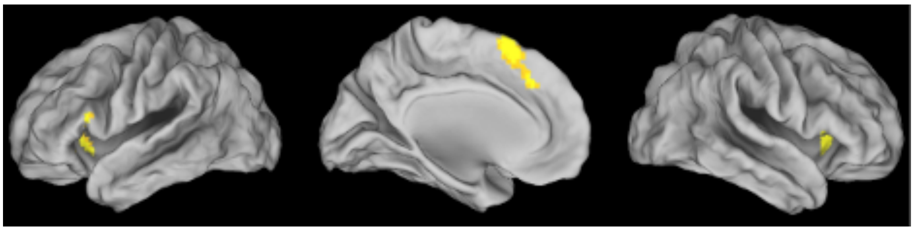
A conjunction analysis was used to investigate the areas underlying error detection that are common to speech production and speech perception. In this analysis, we tested for a rejection of the conjunction null hypothesis (i.e., only those voxels were reported as active which proved to be significant for speech production and speech perception). The conjunction analyses revealed several clusters that were commonly more active in erroneous compared to correct trials (Table I; Figure 1). Clusters of activation were found in the pre-SMA extending into the dACC, the left AI and IFG, and the right IFG extending into AI.

ROI analysis confirmed a main effect of accuracy (all p's <.001), with an increase in BOLD response for erroneous compared to correct trials in all areas (all p's <.001). Only the left IFG showed a significant effect of modality (*p* <.05, all other *p*'s >.10), with higher activation for perception compared to production. The left insula and SMA showed a significant interaction effect (p <.05). The interaction is driven by a larger activation difference between erroneous and correct trials in production compared to perception.

Table 1. *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 1. Activation map averaged across 21 subjects (p<.05, familywise error corrected) of the conjunction analysis error trials production and error trials perception.



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 II, Figure 2) in the left temporal pole, pre-SMA and dACC and BA 48.

The ROI analysis of production specific areas revealed a significant effect of accuracy in the production trials (all p's <.001) for all areas. A main effect of modality was found in the SMA and dACC (all p's <.001), with a higher activation in production compared to comprehension. All areas showed significant interaction effects (all p's <.005) driven by a larger difference between erroneous and correct trials in production compared to perception.

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 II, Figure 3).

A main effect of accuracy was observed for all areas (all p's <.05). Activation in erroneous trials was increased compared to correct trials. A main effect of modality was observed in the right middle frontal gyrus, bilateral middle temporal gyrus, right inferior frontal gyrus, thalamus, and left insula (all p's <.05). A significant interaction was observed in all area's (left insula p<.05, other regions p<.005) apart from a cluster in the right medial frontal area (coordinates: 30 11 55, p=.162). This interaction was driven by significant higher activation in erroneous trials compared to correct trials in perception, but not in production.

Figure 2. *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.*

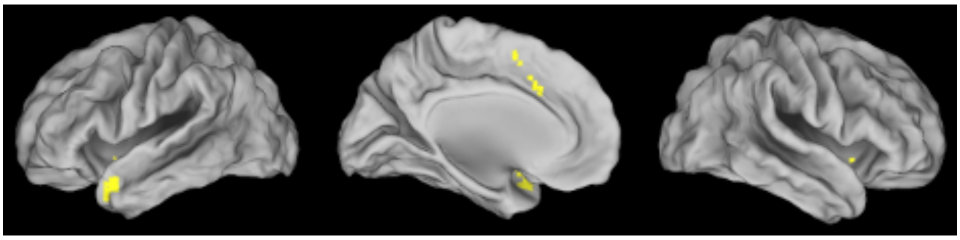


Figure 3. *Activation map averaged across 21 subjects (p<.05, familywise error corrected) of the disjunction analysis error trials perception masked by error trials production, revealing activation specific for error detection in perception.*

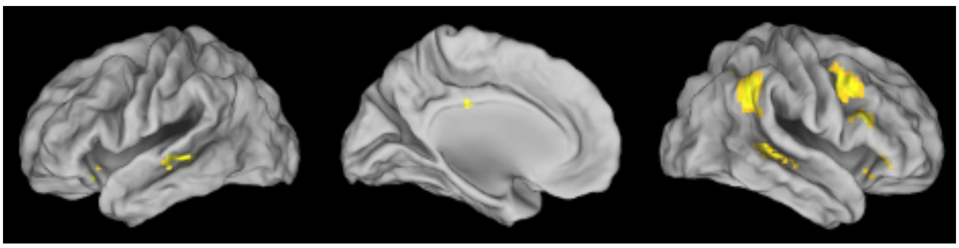


Table 2. *Peak Clusters of Activation revealed by Disjunction Analysis Error Trials Production and Perception*

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	33 23 -14	5.61	10

Frontal Gyrus			
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

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 and 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 show 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). This interaction was 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 inconsistent 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. 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 and correct trials (increases in area -58 12 4, and decrease in area -60 -18 4) in production, but not in perception.

Activation differences between erroneous and correct trials during production and perception are presented in table III. Essentially, this ROI analysis shows that the bilateral STG are stronger activated during speech perception compared to production. With respect to error processing, however, no consistent pattern was found.

Table 3. *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**

GENERAL DISCUSSION

Our goal was to investigate whether internal speech monitoring during speech production and speech monitoring during speech perception recruit similar neuronal structures. Perception-based theories of self-monitoring in speech 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 speech production and perception. Crucially, however, no consistent pattern of activation related to error detection was observed in the bilateral superior temporal sulcus (Indefrey & Levelt, 2004; McGuire et al. 1996; Hirano et al. 1997). If indeed the STS/STG were the main locus for error detection in speech, activation increases would be expected for erroneous trials compared to correct trials, independent of production or perception. The current findings therefore 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). According to this theory, the speaker generates a production command that feeds into a "production implementer," which contains processing steps involved in production. The output of this stage feeds into a "comprehension implementer" to construct an utterance percept. At the same time, an efference copy of the production command feeds into the forward production model that then feeds into the forward comprehension model, thereby creating a predicted utterance percept. Self-monitoring takes place by comparing the utterance percept and predicted utterance percept. This mechanism is similarly applied to speech produced by others. To understand the production of other people, listeners make use of prediction-by-simulation: we predict upcoming words through our own production system. Similar as in one's own speech production the comprehender's predicted utterance percept is compared to the actual utterance percept. The comparison between these two constructs does not necessarily take place in perception systems. This theory is consistent with the action monitoring literature, where it has been claimed that simulation is used to interpret each other's behavior. Similar brain regions have been shown to respond when we perform and observe actions, emotions and touch (e.g. Rizzolatti et al., 2001; Wicker et al., 2003; Keysers et al., 2004; Iacoboni, 2005; Botvinick et al., 2005). In the same vein, error observation leads to a simulation of errors (Shane et al., 2008; de Bruijn et al., 2009; Newman-Norlund et al., 2009; Desmet et al., 2013).

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). 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 outcome (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.

Apart from the domain-general activations, as observed in the conjunction analysis, error detection in speech perception and production showed process-specific activations. Self-monitoring during noise-masked speech production recruited the left temporal pole and pre-SMA and ACC, of which the latter two showed a stronger activation during production compared to perception. The pre-SMA is known to have a somatotopic organization (Chainay et al, 2004; Alario et al, 2006), resulting in process-specific activations. Furthermore differences in recruitment of the control system required in production, compared to comprehension, might also play a role in the observed difference. Left temporal pole activations are observed in tasks requiring the composition of sentence meaning, and more specifically in the processing of syntactic structure (Vandenberghe, 2002; Grodzinski & Friederici 2006; Humphries, 2006). A higher recruitment of this area in incorrect trials compared to correct trials might be indicative of an increased processing cost related to error detection and repair in erroneous production.

Error detection in speech perception revealed process-specific activations in a few clusters in the left hemisphere, and a more extensive pattern of activation clusters in the right hemisphere. Left hemisphere activations include anterior insula and posterior middle temporal gyrus. Activations in the pSTS/MTG are observed bilaterally in response to (noisy) auditory stimuli (Bates, 2003; Boatman 2004; Fu et al. 2006), and in integration of auditory and visual information (Beauchamp et al., 2004). Left MTG has also been linked to semantic processing (e.g. Demonet et al., 1992, 1994; Vandenberghe et al, 1996; Stromswold et al, 1996; Binder et al, 2009; Diaz and McCarthy, 2009; but see Price, 2012). In the right hemisphere large clusters are observed in the posterior middle frontal gyrus, precentral gyrus, in the supra marginal gyrus, in the IFG/AI, and in the pSTS/MTG. The supra marginal gyrus is involved in phonological perception and decision making (Hartwigsen et al. 2010; Buchsbaum et al. 2008; McDermott et al. 2003; Price et al. 1997) although it typically does not show up in speech comprehension tasks (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009).

The current findings are in line with preceding research into language control and monitoring of altered feedback, which also consistently reported activations in the ACC, SMA and frontal areas (e.g. Fu et al, 2006;

Christoffels et al, 2007; Tourville et al, 2008; Piai et al. 2013). One interesting difference between the before-mentioned studies of Fu et al. (2006) and Christoffels et al. (2007) into feedback monitoring and our findings is that we did not find increased activations in the cerebellum. These cerebellar activations during feedback processing have also been related to error detection in perception-based models, as it is hypothesized to drive corrective motor commands to the motor cortex after receiving input from somatosensory and auditory areas (Ito, 2008; Tourville & Guenther, 2011; Hickok, 2012). While the studies above specifically looked at the effect of manipulating external feedback, we have excluded external feedback by noise masking. This hints that the role of the cerebellum might be more closely related to external feedback instead of monitoring proper.

In line with our findings are recent studies in which fMRI was used to study conflict resolution in language processing. Wittfoth et al (2009) investigated emotional conflict processing in speech perception, and Piai at al. (2014) investigated attentional conflict in language and non-language processing. Processing of emotional conflicting information (e.g., a semantically positive sentence with a negative prosody) also showed an increase in BOLD response in the posterior medial prefrontal cortex extending into ACC, bilateral insula and IFG, posterior cingulate and inferior parietal lobule. Processing of attentional conflict in a Stroop Task (color word is printed in an incongruent ink color), a Picture-Word Interference Task (picture and distractor are semantically related), and a Simon Task (press a left or right button to a visual stimulus presented on the opposite side) all elicited ACC activation. So what we observe in speech error detection are activations consistent with a domain-general error detection mechanism, through performance monitoring and adjustment of cognitive control.

In summary, our results suggest that error detection in speech processing takes place through a domain-general conflict monitoring system, which comprises the dorsal anterior cingulate cortex, supplementary motor area, bilateral anterior insula, and inferior frontal gyrus. This network, which has been consistently observed in non-linguistic conflict, is recruited for both speech perception and speech production. The lack of evidence for the involvement of the superior temporal gyrus, does not offer support for perceptual theories of error monitoring. The involvement of the conflict-monitoring network rather argues for a conflict monitoring account of error detection in speech.

REFERENCES

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

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

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

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

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

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

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

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

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

Blackmer ER, Mitton JL (1991) Theories of monitoring and the timing of repairs in spontaneous speech. Cognition 39:173-194.

Boatman D (2004) Cortical bases of speech perception: evidence from functional lesion studies. Cognition 92:47–65.

Bonini, F, Burle, B, Liégeois-Chauvel, C, Régis, J, Chauvel, P, Vidal, F (2014) Action monitoring and medial frontal cortex: leading role of supplementary motor area. Science 343:888-891.

Botvinick MM, Braver TS ,Barch DM ,Carter CS, Cohen JD (2001) Conflict Monitoring and Cognitive Controle. Psychological Review 108:624-652.

Botvinick M, Jha AP, Bylsma LM, Fabian SA, Solomon PE, Prkachin KM (2005) Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. Neuroimage 25:312-9.

Buchsbaum BR, D'Esposito M (2008) The search for the phonological store: from loop to convolution. Journal of Cognitive Neuroscience 20:762–778

Chainay H, Krainik A, Tanguy ML, Gerardin E, Le Bihan D, Lehericy S. (2004) Foot, face and hand representation in the human supplementary motor area. Neuroreport 15:765–769.

Christoffels IK, Formisano E, Schiller NO (2007) Neural correlates of verbal feedback processing: an fMRI study employing overt speech. Human Brain Mapping 28:868–879.

Christoffels IK, Van De Ven V, Waldorp LJ, Formisano E, Schiller NO (2011) The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. PLoS One: 6.

Craig AD (2010) Once an island, now the focus of attention. Brain Structure and Function 214:395–396.

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

de Bruijn ERA, de Lange FP, von Cramon DY, Ullsperger M (2009) When errors are rewarding. Journal of Neuroscience 29:12183-12186.

Demonet JF, Chollet F, Ramsay S, Cardebat D, Nespoulous JL, Wise R, Rascol A, Frackowiak R (1992) The anatomy of phonological and semantic processing in normal subjects. Brain 115:1753–1768.

Demonet JF, Price C, Wise R, Frackowiak RS (1994) Differential activation of right and left posterior sylvian regions by semantic and phonological tasks: a positron-emission tomography study in normal human subjects. Neuroscience Letters 182:25–28.

Desmet C, Deschrijver E, Brass M (2013) How social is error observation? The neural mechanisms underlying the observation of human and machine errors. Social Cognitive and Affective Neuroscience 9:427-435.

Diaz, MT, McCarthy G, (2009) A comparison of brain activity evoked by single content and function words: an fMRI investigation of implicit word processing. Brain Research 1282:38–49.

Eden GE, Joseph JE, Brown HE, Brown CP, Zeffiro TA (1999) Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. Magnetic Resonance in Medicine 41:13-20.

Eliades SJ, Wang X. (2003) Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. Journal of Neurophysiology 89:2194–2207.

Eliades SJ, Wang X. (2005) Dynamics of auditory-vocal interaction in monkey auditory cortex. Cerebral Cortex 15:1510–1523

Fu CH, Vythelingum GN, Brammer MJ, Williams SC, Amaro Jr E, Andrew CM, Yaguez L, van Haren NE, Matsumoto K, McGuire PK (2006) An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. Cerebral Cortex 16:969–977.

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

Grodzinsky Y, Friederici AD (2006) Neuroimaging of syntax and syntactic processing. Current Opinion in Neurobiology 16:240–246.

Hartwigsen G, Baumgaertner A, Price CJ, Koehnke M, Ulmer S, Siebner HR (2010) Phonological decisions require both the left and right supramarginal gyri. Proceedings of the National Academy of Sciences U.S.A. 107:16494–16499.

Hartsuiker RJ, Kolk HHJ (2001) Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology* 42:113-157.

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

Heinks-Maldonado TH, Nagarajan SS, Houde JF. (2006) Magnetoencephalographic evidence for a precise forward model in speech production. Neuroreport 17:1375–1379.

Herrmann MJ, Rommler J, Ehlis AC, Heidrich A, Fallgatter AJ (2004) Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). Cognitive Brain Research 20:294–299

Hickok G, Poeppel D. (2007) The cortical organization of speech processing. Nature Reviews Neuroscience 8:393–402

Hickok G (2012) Computational neuroanatomy of speech production. Nature Reviews Neuroscience 13:135-145

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

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

Humphries C, Binder JR, Medler DA, Liebenthal E (2006) Syntactic and semantic modulation of neural activity during auditory sentence comprehension. Journal of Cognitive Neuroscience 18:665–679.

Iacoboni M (2005) Neural mechanisms of imitation. Current Opinion in Neurobiology 15:632-637.

Indefrey P (2011) The spatial and temporal signatures of word production components: a critical update. Frontiers in Psychology 2:255.

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

Ito M (2008) Control of mental activities by internal models in the cerebellum. Nature Reviews Neuroscience 9:304–313.

Keysers C, Wicker B, Gazzola V, Anton JL, Fogassi L, Gallese V (2004) A touching sight: SII/PV activation during the observation and experience of touch. Neuron 42:335-46.

Lackner JR, Tuller BH (1979) Role of efference monitoring in the detection of self- produced speech errors. In: Sentence processing (Cooper WE, Walker ECT, ed), pp. 281–294. Hillsdale, N.J.: Erlbaum.

Laver J (1980) Monitoring systems in the neurolinguistic control of speech production. In: Errors in linguistic performance: Slips of the tongue, ear, pen, and hand (Fromkin VA, ed), pp. 287-305. New York: Academic Press.

Levelt WJM (1983) Monitoring and self-repair in speech. Cognition 14:41–104.

Levelt WJM (1989) Speaking: From intention to articulation. Cambridge, MA: MIT Press.

Matlab and SPM8 software Wellcome Department of Cognitive Neurology, London, UK.

McDermott KB, Petersen SE, Watson JM, Ojemann JG (2003) A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. Neuropsychologia 41:293–303

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

Menenti L, Gierhan SME, Segaert K, Hagoort P (2011) Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by fMRI. Psychological Science 22:1173–82.

Monfardini E, Gazzola V, Boussaoud D, Brovelli A, Keysers C, Wicker B (2013) Vicarious neural processing of outcomes during observational learning. PLOS One 8.

Newman-Norlund RD, Ganesh S, van Schie HT, De Bruijn ERA, Bekkering H (2009) Self-identification and empathy modulate error-related brain activity during the observation of penalty shots between friend and foe. Social Cognitive and Affective Neuroscience 4:10-22.

Nooteboom SG (1980) Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech In: Errors in linguistic performance: Slips of the tongue, ear, pen, and hand (Fromkin VA, ed), pp. 87-95. New York: Academic Press.

Nozari N, Dell GS, Schwartz MF (2011) Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. Cognitive Psychology 63:1-33.

Numminen J, Salmelin R, Hari R. (1999) Subject's own speech reduces reactivity of the human auditory cortex. Neuroscience Letters 265:119–122.

Oppenheim GM, Dell GS (2008) Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. Cognition 106:528-537.

Piai V, Roelofs A, Acheson DJ, Takashima A (2013) Attention for speaking: Neural substrates of general and specific mechanisms for monitoring and control. Frontiers in Human Neuroscience 7:832.

Pickering MJ, Garrod S (2013) An integrated theory of language production and comprehension. Behavioral and Brain Sciences 36:329-392.

Pickering MJ, Garrod S (2014) Self- Other and joint monitoring using forward models. Frontiers in Human Neuroscience 8:132-143.

Postma A, Kolk HHJ (1992) The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. Journal of Speech and Hearing Research 35:537–544.

Postma A, Noordanus C (1996) Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. Language and Speech 39: 375–392.

Price CJ, Moore CJ, Humphreys GW, Wise RJS (1997) Segregating semantic from phonological processes during reading. Journal of Cognitive Neuroscience 9:727–733

Price CJ (2012) The anatomy of language: a review of 100 fMRI studies published in 2009. Annals of the New York Academy of Sciences 1191:62–88

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nature Neuroscience 12:718–724.

Ridderinkhof RK, Van den Wildenberg WPM, Segalowitzd SJ, Cartere CS (2004) Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. Brain and Cognition 56:29–140.

Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. Nature Reviews Neuroscience 2:661-70.

Shane MS, Stevens M, Harenski CL, Kiehl KA (2008) Neural correlates of the processing of another's mistakes: a possible underpinning for social and observational learning. Neuroimage 42:450-459.

Severens E, Kühn S, Hartsuiker RJ, Brass M (2011). Functional mechanisms involved in the internal inhibition of taboo words. *Social cognitive and affective neuroscience*, nsr030.

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

Stromswold K, Caplan D, Alpert N, Rauch S (1996) Localization of syntactic comprehension by positron emission tomography. Brain and Language 52:452–473.

Takaso H, Eisner F, Wise RJ, Scott SK (2010) The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study. Journal of Speech Language and Hearing Research 53:226–236.

Tops M, Boksem MA (2011) A potential role of the inferior frontal gyrus and anterior insula in cognitive control, brain rhythms, and event-related potentials. Frontiers in Psychology 2:330.

Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory feedback control of speech. Neuroimage 39:1429–1443.

Tourville JT, Guenther FH (2011) The DIVA model: A neural theory of speech acquisition and production. Language and Cognitive Processes 25:952-981.

Ullsperger M, von Cramon DY (2006) The role of intact frontostriatal circuits in error processing. Journal of Cognitive Neuroscience, 18:651-664.

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

Vandenberghe R, Nobre AC, Price CJ (2002) The response of left temporal cortex to sentences. Journal of Cognitive Neuroscience 14:550–560.

Van Veen V, Carter CS (2002) The Timing of Action-Monitoring Processes in the Anterior Cingulate Cortex. Journal of Cognitive Neuroscience 14:593-602.

Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G (2003) Both of us disgusted in My Insula: the common neural basis of seeing and feeling disgust. Neuron 40:655-664.

Wittfoth M, Schröder C, Schardt DM, Dengler R, Heinze HJ, Kotz S A (2009). On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. Cerebral Cortex, 20:383-392.

Xue G, Aron AR, Poldrack, RA (2008). Common neural substrates for inhibition of spoken and manual responses. Cerebral Cortex 18:1923-1932.

Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. Psychological Review 11:931–959.

Zheng ZZ, Munhall KG, Johnsrude IS (2010) Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of cognitive neuroscience,* 22(8), 1770-1781.

CHAPTER 5
VERBAL MONITORING IN PARKINSON'S DISEASE: A COMPARISON BETWEEN INTERNAL AND EXTERNAL MONITORING

Manuscript in preparation[1]

Parkinson's disease patients display a variety of impairments in linguistic processes; speech is decreased in amplitude, and temporal aspects, such as prosody and speed, are affected. The present study investigated whether Parkinson's patients verbal monitoring is impaired, and what the relative contributions of the internal and external monitoring route are on verbal monitoring. Furthermore the data were used to investigate whether internal monitoring performance could be predicted by internal speech perception tasks, as perception based monitoring assumes. Performance of 18 Parkinson's disease patients was measured on two cognitive performance tasks and a battery of 11 linguistic tasks, which measured performance on internal and external monitoring. Results were compared with those of 16 age matched controls, and (where available) normative data. A comparison of monitoring performance on internal speech, during noise-masked feedback, and external monitoring, during normal feedback, revealed that Parkinson's patients relied more on their internal monitoring route for verbal monitoring. A regression analysis with monitoring behavior as

---

[1] This chapter was written in collaboration with Jolien Mertens, Peter Mariën, Patrick Santens, Barbara A. Pickut & Robert J. Hartsuiker

dependent variable showed that internal monitoring behavior was independent of internal speech tasks performance, suggesting that internal monitoring occurs independent of internal speech perception.

INTRODUCTION

During speech production, our speech is constantly monitored for errors. After speech production we can monitor the produced speech via perception. This route is called the external monitoring route. Additionally we can monitor our speech internally, before production. Previous research into verbal monitoring has shown alterations in verbal monitoring behavior as a result of brain damage (Schlenck et al., 1987; Liss 1998; Oomen, Postma & Kolk 2001). More specifically, these patients show a greater reliance on the internal monitoring route for self-monitoring, compared to healthy adults. The current study investigates if Parkinson's disease patients have altered verbal monitoring, by specifically investigating the contribution of each of the verbal monitoring routes.

Evidence for an internal monitoring route has come from extremely fast self-corrections (50% of errors are repaired within 100 ms) that have been observed in both spontaneous speech and in experimentally elicited speech (Blackmer & Mitton, 1991; Levelt, 1989), the ability of people to report errors in internal speech (covert speech production) (Dell & Repka 1992; Oppenheim & Dell, 2008), or when they cannot hear themselves speak as their speech is masked by a loud noise (Lackner & Tuller, 1979, Postma & Kolk 1992a, b). Also when in a speech error elicitation task the production of a slip of the tongue would lead to an inappropriate utterance (e.g. TOOL – KITS), the galvanic skin is elevated, and slips occur less frequently (Motley, Camden, & Baars, 1981, 1982). This suggests that internally the SLIP is made, but repaired before production as a result covert editing. It also suggests that top-down influence can be exerted over the monitoring system.

How internal monitoring works is still under debate. Roughly there are two classes of theories with respect to internal verbal monitoring; perception-based monitoring theories and production-based monitoring theories. The perception-based monitoring theory, the Perceptual Loop Theory (Levelt, 1983, 1989; Indefrey & Levelt, 2004; Indefrey 2011), assumes that internal monitoring is performed via the perception of a phonemic representation. According to this theory, monitoring internal speech, monitoring one's own external speech, and monitoring someone else's speech, all occur via the same perception mechanism. The perceived speech is then compared to the communicative intention by a central monitor. There are also a number of theories that assume production-based verbal monitoring, such as the conflict monitoring theory (Nozari, Dell & Schwartz, 2011), forward model theory (Pickering and Garrod 2013a, 2014), the hierarchical state feedback control model (Hickok, 2012), and DIVA model (Guenther, Hampson, & Johnson, 1998; Golfinopoulos, Tourville, & Guenther, 2010) While each of these theories differ with respect to how exactly the error is detected, all assume that monitoring takes place during speech production, and so not rely on the perception of internal speech similar to external speech.

Studies of monitoring in patients with acquired or progressive brain damage has provided a unique insight into verbal monitoring and the relationship between verbal monitoring during production and comprehension. For instance, several studies, as discussed in Chapter 2, have investigated whether verbal self-monitoring is related to language comprehension skills, as hypothesized on the basis of the perceptual loop theory. One study indeed found that patients with good comprehension also showed a high number of self-repairs (Marshall et al., 1994). However, most other patient studies that investigated the relation between comprehension and monitoring found a dissociation between comprehension and self-repair skills. For instance McNamara, Obler, Au, Durso, and Albert (1992) investigated three groups; Parkinson's Disease (PD) patients, Alzheimer patients, and healthy controls. This study found indeed that the patients with dementia, who had poor comprehension also showed poor self-monitoring, and healthy controls showed good self-monitoring and good comprehension. But interestingly, the Parkinson's patients showed relatively poor self-monitoring, despite intact comprehension skills; 75% of the produced errors remained unrepaired, despite good verbal perception. The PD patients made 3 times more errors than healthy controls, and only 25% of the errors were corrected, compared to 72-92% error corrections in healthy adults. In a study investigating 69 aphasic patients, Miceli, Gainotti, Caltagirone and Masullo (1980) found no correlation between the performance on a phonemic discrimination test and the degree of phonemic output disorder, which suggests a dissociation between phonemic processing during production and perception. A comparison of 15 aphasic patients with phonological production errors by Nickels and Howard (1995), found no correlation between the proportion of naming errors and performance on a comprehension task. However, a reanalysis by Roelofs (2005) showed a positive correlation between performance on a homophone task and phonological self-corrections and false starts, but a negative correlation with the number of semantic errors. Marshall, Rappaport, and Garcia-Bunuel (1985) describe a woman with such a severe auditory agnosia that she had a near-total loss of the ability to understand speech and non-speech sounds (despite of intact hearing), who nevertheless corrected, and made many attempts to correct, the phonemic errors she produced. Interestingly, she ignored semantic errors. Finally, Oomen, Postma, and Kolk (2005) reported the interesting case of a Broca's aphasic who made many phonological errors during production, which he failed to repair in most attempts (only 38% of attempts was successful). Semantic errors on the other hand, were produced much less frequently, and they were successfully repaired most of the time. However, during perception he repaired as many phonological errors as the healthy controls, but repaired fewer semantic errors than the controls. Taken together these patient data show that self-monitoring and other-monitoring can be selectively impaired at the semantic and phonological level, and that intact comprehension and intact other-monitoring are not sufficient for correct self-monitoring.

One interesting finding in patient studies investigating verbal monitoring, has been the finding that patients use their monitoring channels differently from healthy controls. Broca's aphasics have been found to strongly rely on the internal route for speech monitoring in several studies (Schlenck et al., 1987; Kolk, 1995; Oomen, Postma & Kolk, 2001). Also Wernicke's patients rely strongly on the internal monitoring route, as suggested by the production of few repairs and many disfluencies (Slenck et al. 1987). A study investigating apraxic speakers found that these speakers were quick to interrupt an error (Liss, 1998), but slow to restart, suggesting that internal monitoring was relatively intact but that planning a repair was effortful. It is unclear at this moment why brain damaged patients would rely more on their internal monitoring route for verbal monitoring. However, we have two obvious suggestions. First of all, the patient has a monitoring deficit in the external monitoring route, and therefore has to rely on internal verbal monitoring. Secondly, the patient suffers from an attention deficit, and can therefor only monitor one channel.

Research into language functioning processing of Parkinson's patients have uncovered an array of problems, which we discuss more extensively below. PD patients for instance show deficits in verb inflection and generation (Longworth et al., 2005; Ullman et al., 1997; Crescentini et al., 2008; Péran et al., 2003) and impaired semantic priming (Castner et al., 2007; Copland, 2003). Also comprehension is impaired; processing long and complex sentences is performed slower and less accurately than controls (Grossman et al., 1992; Lieberman et al., 1990), and PD patients often show difficulty in understanding metaphoric meaning, and distinguishing between jokes and lies (Monetta & Pell, 2007; Monetta, Gindrod & Pell, 2009). To date there have been only a small number of studies that investigated whether verbal monitoring is impaired in Parkinson's disease patients, and no studies, to the best of our knowledge, specifically investigating what the relative contributions of the internal and external monitoring routes are in Parkinson's disease patients.

PARKINSON'S DISEASE

Parkinson's disease (PD) is the second most prevalent age-related neurodegenerative disease. PD is caused by the death of dopaminergic neurons in the substantia nigra, situated in the basal ganglia, leading to a variety of symptoms of which movement related symptoms are the most well known. This progressive neurodegenerative disease has a mean age of onset at 55. Within 5 to 10 years after onset, independent of medicinal treatment, participants develop severe motor disability. Before the symptoms emerge, it is estimated that the disease will have progressed over a period of 5 to 15 years, in which a loss of 60-80% of the dopaminergic neurons in the substantia nigra is observed (Miller & O'Callaghan, 2014). As a result medicinal treatment has limited effects.

From a neurological perspective, characteristic about Parkinson's disease is the loss of of dopamineric cells in the substantia nigra, as well as the formation of 'Lewy Bodies' (LB; abnormal aggregation of proteins inside neurons) in this area (Olanow & Obeso, 2012). Neurodegeneration and LB formation is also observed in other areas throughout the midbrain (e.g. the locus coeruleus, nucleus basalis, raphe nucleus) as well as in the cerebral cortex (specifically the cingulate and entorhinal cortex), the olfactory bulb, and autonomic nervous system (see review by Hornykiewicz and Kish, 1986). Disruptive functioning of the basal ganglia results in an impoverished ability to select actions, which is manifested first and most clearly in the motor domain. Parkinson's disease is characterized by tremor at rest (but not during voluntary movements), rigidity (stiffness/ resistance to passive movements of a limb), freezing (inability to initiate voluntary movement), bradykinesia (slowness of movement), hypokinesia (amplitude reduction in movement), and akinesia (absence of normal unconscious movements, such as arm swing in walking). Often a stooped posture is developed and a patient may lose normal postural reflexes, leading to falls. Overall cognitive processes are observed to be slowed (bradyphrenia), which for instance results in a delayed responding to questions. For an extensive overview of the syndrome, pathogenesis, and pathophysiolosy of Parkinson's disease, see for instance Bartels and Leenders (2009).

LANGUAGE PROBLEMS IN PARKINSON'S DISEASE

Communication through speech and writing is compromised in PD patients as the disease progresses. Handwriting is impaired as both writing speed and the size of the handwriting are decreased (micrographia). Speech production is impaired, as typically PD patients develop a hypokinetic dysarthria, which results in a deficient respiratory control, articulatory imprecision, poor control of voice onset and offset decreased voice volume (hypophonia), and defective prosody (absence of melody and stress) (Critchley. 1981; Darley et al., 1969; Darley, Aronson & Brown, 1975; Gallina, Smith, Zeffiro & Ludlow, 2001; Logeman, Fisher, & Bowler, 1978; Sanabria et al., 2001; Solomon & Hixon. 1993). Speech rate can be affected in a variety of manners; normal speech rate is observed for mild cases of PD (Metter & Hanson, 1986), but speech rate is also observed to be increased (Flint et al., 1992) or decreased (Ludlow, Connor, & Bassich, 1987; Hammen, 1990). Speech fastination can also be observed, in which the speech rate increases during production to a rate at which intelligibility is severely affected (Critchley, 1981). Stuttering is also commonly observed in PD patients. The alterations in speech production affect the listener's perception of PD patients in terms of linguistic and social competence (McNamara & Durso, 2003). Inaccurate temporal processing and prosody lead to a characterization of PD patients with terms as 'cold', 'anxious' and 'unhappy' as compared to healthy adults (Pitcairn, Clemie, Gray, & Pentland, 1990). Admission of levodopa gives inconsistent results with respect to speech perception. Some studies have reported an improvement in performance on articulation, loudness, and persistence of phonation (Critchley, 1981; Wolfe, Garvin, Bacon, & Waldrop, 1975; Sanabria et al., 2001; De Letter et al., 2007), whereas other studies did not find such improvements (Gentil, Tournier, Pollack, & Benabid, 1999; Poluha, Teulings, & Brookshire, 1998), or found mixed results (De Letter, Santens, & Van Borsel, 2005), or even a decrease in performance after treatment with levodopa (Louis, 2001).

A large part of the communication problems described above can be explained as a result of problems in the motor domain, such as micrographia, deficient respiratory control, decreased voice volume and articulatory imprecision. However, an additional number of linguistic processes have been observed to be disrupted that are independent of motor processes.

One consistent finding is that semantic associations are impaired in PD patients. For instance in a semantic association task, semantic activations are delayed in PD patients compared to healthy controls. This delay is specifically observed during controlled semantic association, and to a lesser extent for automatic semantic association (priming) (Angwin et al., 2005; Angwin et al., 2009; Grossman, Zurif & Lee, 2002; Castner et al., 2007; Copland, 2003). Automatic semantic associations were not observed in PD patients that did not use levodopa, compared to the on-levodopa condition (Arnott et al., 2011). Semantic ambiguity resolution is less accurate and slower in PD patients compared to healthy controls (Copland et al., 2009; Ketteler et al., 2014). Also in discourse context PD patients have shown to have difficulty disambiguating (Copland et al., 2001; Copland, 2003). A vast body of literature describes problems with grammatical processing; during production sentences are often short and semantically limited (Illes, 1989; Illes et al., 1988). Comprehension problems are found for sentences that are long or syntactically complex (e.g., Grossman et al., 1991, 1992, 1993; Hochstadt et al., 2006; Lieberman et al., 1990, 1992; Natsopoulos et al., 1991, 1993). A study investigating morphological grammatical processing found the semantic and morphological processing in perception was normal, while during a production task in which novel verbs were inflected, participants displayed difficulty in suppressing semantic alternatives (Longworth et al., 2005.). Verb generation is also impaired in PD patients. Problems have been demonstrated in the domains of verb learning (Grossman et al., 1994) and in verb production in sentence context (Ullman et al., 1997; Colman et al., 2009).

Many of the language processing problems that arise in Parkinsons' disease also arise in patients with Broca's aphasia. These patients, for instance, also show impaired verb production (Bastiaanse 2008; Bastiaanse et al., 2002), inflectional morphology (Penke et al., 1999; 2006), and complex sentence processing (Bastiaanse and Van Zonneveld, 2005, 2006; Grodzinsky, 1995; Lee & Thompson, 2004). This parallel in results is not surprising if we consider the origin of the deficits. Parkinsons' disease arises as the result of a loss of dopaminergic cells in the substantia nigra, which leads to a defective functioning in the basal ganglia. Broca's patients have brain lesions in the vicinity of Broca's area, which includes, or is sometimes even restricted to subcortical connections to the basal ganglia (Bastiaanse & Leenders, 2009). So there is reason to assume that problems in linguistic processing of both patient groups arise as a result of the same underlying deficit. A special issue of *Cortex* investigating language problems in PD patients concluded that there is substantial overlap in symptoms between PD patients and Broca'a aphasics: impaired semantic activations, poor verb production and decreased sentence comprehension are observed for both patient groups (Bastiaanse & Leenders, 2009). However, the observed deficits are less severe in PD patients compared to Broca's aphasia.

One aspect that has not received much attention from studies investigating language-processing deficits in PD patients is verbal monitoring. Verbal monitoring has been studied more intensively in Broca's aphasics (Schlenck et al., 1987; Kolk, 1995; Oomen, Postma & Kolk, 2001). By comparing linguistic monitoring performance during a speech production task under normal feedback, where participants can use both the internal and external monitoring route, and under noise-masked feedback, where participants have to rely on their internal monitoring route only, Oomen et al., (2001) found that Broca's aphasics rely heavily on their internal monitoring route. Contrary to controls, monitoring performance in the aphasic patients was not decreased when the external route was not available. Based on the large body of literature that show similarities between language performance in Broca's aphasics and PD patients, we expect PD patients to demonstrate similar verbal monitoring skills as the Broca's aphasics in Oomen et al. (2001). If PD patients are affected similarly as Broca's aphasics, but too a lesser degree as the literature suggests, then we expect to find a stronger reliance on internal verbal monitoring compared to external verbal monitoring in Parkinson's patients. We use the same task as Oomen et al., (2001), the network task with a normal feedback and a noise-masked feedback condition, which allows for a direct test of reliance on the internal and external monitoring route. If indeed patients rely more heavily on internal monitoring, we expect to find no difference in performance during noise masked-feedback compared to performance during normal feedback. Additionally we wish to investigate whether internal verbal monitoring during noise-masked feedback is correlated with performance on internal speech tasks. If self-monitoring is performed by perception of internal speech, as the perceptual loop theory assumes, we would expect that internal monitoring is directly related to performance on internal speech tasks. Production based accounts of self-monitoring do not necessarily assume such a direct relation.

MATERIALS AND METHODS

PARTICIPANTS

In this study 21 Parkinson's patients and 21 age-matched controls were tested on a series of language tasks. Participants were recruited ad hoc during examination at the ZNA for adjustments of their medication scheme, or were contacted beforehand to coordinate the experiment with their regular examination for adjustments of their medication scheme at Ghent University Hospital. They were all treated with levodopa combined with various schemes of dopamine receptor agonists and amantadine. These medication schemes were all individualized resulting in highly variable dose regimens, as is characteristic in PD. Their partners were recruited as control participants, as they matched on age and SES. For those patients of whom the partner did not want to participate, or who did not have a partner, age-matched controls were recruited. As stated above, participants were recruited ad hoc; either after their examination or before patients and their partners were asked by the medical examiner to participate in this study. As testing took quite long, ofter over 1 hour, a few patients and controls ended their participation before completing all tasks. These patients (n=2) and controls (n=2) were left out of the final analysis. One patient was discarded as during testing short-term memory problems became apparent. Three further controls, partners of patients, were excluded as one suffered from Alzheimer's disease, one had suffered a stroke from which she had not fully recovered, and one as she was not a native speaker of Dutch.

The data of 18 Parkinson's patients (aged between 44 - 80 years, mean 65.72. SD 8.47) were taken into analyses, and those of 16 age matched controls (range 43 – 82, mean 64.19. SD 8.33). All participating Parkinson's patients were clinically diagnosed idiopathic PD. All diagnoses were made by the fourth (P.S.) and fifth (B.P.) author of this paper according to the Unified Parkinson's Disease Rating Scale (UPDRS) (Fahn, Elton, &

members of the UPDRS Development Committee, 1987). Patients reported a mean onset of the disease of 10.5 years (SD 6.09 range 3-20) since onset of movement related problems. For one Parkinson's patient and four of the control patients, the tasks were performed at home. For all others testing took place in a room of the respective hospitals.

DESIGN

Each participant performed 11 tasks. The tasks were chosen so that they measured cognitive performance, speech production, internal speech perception, external speech perception, and verbal monitoring. Of the speech tasks, two tasks were speech perception tasks, and all other tasks measured overt or covert speech production. Additionally the handedness was measured via the Edinburgh handedness inventory (Oldfield 1971) and a questionnaire was given that asked about their language history, and any language impairments, and issues related to hearing or sight.

COGNITIVE PERFORMANCE TASKS

Mini Mental State Examination

The mini mental state examination (Folstein, Folstein & McHugh 1975) is a short test suited to acquire a global idea of cognitive performance in a short amount of time. The task takes up to 10 minutes, and measures both mental and motor functioning. Braak et al (2005) found a correlation between the degree of Parkinson's disease and score on the mini mental state examination (MMSE). This study found that as the disease progresses, the chance of developing dementia increases.

Ravens Progressive Matrices

Ravens Progressive Matrices (Raven, Court, & Raven, 1998) is a non-verbal intelligence test based on a multiple-choice questionnaire, which is generally accepted as a good measure of fluid intelligence (Daley, Whaley, Sigman, Espinosa, & Neumann, 2003; Mani, Mullainathan, Shafir, & Zhao, 2013). The participant is presented with an array of visual displays of which one element is missing. The participant is asked to indicate which of the options would best suit the place of the missing element. The complexity of the visual display increases with each trial. Participants under 65 performed set B, C, and D of the Standard Progressive Matrices (SPM). For participants over 65 the Coloured Progressive Matrices (CPM) are commonly used, as they give a better estimation of intelligence for people of this age group (O'Leary et al., 1991; Lindeboom et al., 1999). Therefore participants over 65 years of age performed set A, AB and B of the CPM. The task was presented on a computer screen, and was preformed in a self-paced manner. The duration of the task was approximately 25 minutes.

SPEECH PRODUCTION TASKS

Boston Naming Task

The Boston naming (BNT) (Kaplan, Goodglass en Weintraub, 1983) is a measure of confrontational word retrieval. Participants name 60 black and white outline drawings of objects and animals. The order of the items is constructed so that with each picture, frequency decreases; item 1 is 'bed', and the final item is 'abacus'. The BNT is a widely used tool to assess damage in word retrieval capacities in both adults and children, and with a wide variety of cerebral pathologies (see for instance Mariën, Mampaey, Vervaet, Saerens & De Deyn, 1998 for an overview). This test has found decreased performance in patients with dementia of the Alzheimer type (Bowles, Obler, Albert, 1987), as well as with general cognitive decline as seen in normal ageing of healthy individuals (Connor, Spiro, Obler & Albert, 2004). In the current experiment, as concurrent with standard administration, the PD patients were shown all 60 pictures, starting with picture 1. No time pressure was applied, and no cues were given. In the control group, the participants started with naming at item 30. If an error was made before item

38, the test returned to item 29 and was continued backwards until 8 consecutive items were named correctly. Once 8 consecutive items were named correctly, naming resumed at item 39. If in the control group the picture was interpreted incorrectly, a semantic cue was given. If after 20 seconds no response was given, a phonological cue was given. Responses were tape recorded and transcribed. Results are compared against norms of Dutch elderly (Mariën, 1998). The duration of the BNT was approximately 5 minutes.

Controlled Word Association Task

The Controlled Word Association Task (COWAT) (Benton & Hamsher, 1976) measures the participants' phonological and semantic verbal fluency. Performance on this task is affected by brain damage (e.g. Stuss et al., 1998 for an overview). In both age-related cognitive decline and Alzheimer's disease category fluency is impaired, while letter fluency remains spared (Monsch et al., 1995; Tombaugh, Kozak, & Rees, 1999).

In 60 seconds participants are required to name as many words as possible according to a given instruction. The current experiment used a version for which norms for Dutch-speaking elderly exist (Miatton, Wolters, Lannoo & Vingerhoets, 2004). In the phonological fluency task the participant is asked in three trials of 60 seconds to name as many words as possible starting with the letters 'N' 'A' or 'K'. In the semantic fluency task the participant is asked in two trials of 60 seconds to name as many words as possible belonging to the category 'animals' or 'professions'. The category animals and the total sum of the letter categories are age and education dependent. The category professions is only affected by educational level. Responses were tape recorded and transcribed. Performance on this task will be compared against normative data of Dutch elderly (Miatton et al., 2004). The duration of the COWAT was approximately 6 minutes.

INTERNAL SPEECH PERCEPTION TASKS

Homophone Decision Task

The homophone decision task is taken from the Dutch version of the Psycholinguistic Assessments of Language processing in Aphasia test battery (Bastiaanse, Bosje & Visch-Brink, 1995) (PALPA test 27). In this task participants are presented with 2 written words on a display screen. The participants task is to indicate with a button press whether the 2 words sound the same (green button on the right) or whether they do not (red button on the left), independent of whether it is a real word or not. The instructions specifically state that the task needs to be performed silently so that the participant is dependent on his internal speech to determine whether the words sound the same. In total 50% of the words are homophones. Of the homophones, 50% are words ('zij' *she* - 'zei' *said*) and 50% are non-words ('klicht' – 'kligt'). Of the non-homophones, also 50% are words ('zien' *see* - 'ziek' *sick*) and 50% are non-words that follow the rules for Dutch word composition ('klicht' – 'klogt'). The experiment begins with 3 practice trials in which the participant receives feedback, followed by 44 experimental word pairs. The total duration of the homophone task was approximately 5 minutes.

Phoneme Monitoring Task

The phoneme-monitoring task (Özdemir, Roelofs & Levelt, 2007) was adjusted for this patient study, and built up from three tasks: a familiarization task, a naming task, and a phoneme-monitoring task. The participant is first presented with pictures and their corresponding names in a familiarization task. The duration of the presentation is self-paced. After this part the participant is again presented with the pictures and has to name the pictures out loud. When the picture is named incorrectly, the experimenter corrects the participant. This is followed by the experimental phoneme-monitoring task in which the participant is presented with a phoneme and a picture, and the participant has to indicate with a button press whether the presented phoneme is part of the name of the picture (green button on the right side) or not (red button on the left side). Each picture is presented twice; once with a phoneme that is part of the name, and once with a phoneme that is not part of the name. The procedure (familiarization, naming, and phoneme monitoring) was applied to blocks of four words. In total 46 items were presented. The total duration of the homophone task was approximately 25 minutes.

Visual Rhyme Judgment Task

The rhyme judgment task is taken from the Psycholinguistic Assessments of Language processing in Aphasia test battery (Kay, Lesser & Coltheart 1996)(PALPA test 14). In the visual rhyme judgment task participants are presented with 2 written words on a display screen. The participant's task is to indicate with a button press whether the 2 words rhyme (green button on the right) or whether they do not (red button on the left). The instructions specifically state that the task needs to be performed silently; so that the participant is dependent on his internal speech to determine whether the words rhyme. Of the presented word pairs 50% rhyme. In Appendix A a more elaborate overview of the conditions and the outcomes per conditions is given. The experiment begins with three practice trials in which the participant receives feedback, followed by 60 experimental word pairs.

EXTERNAL SPEECH PERCEPTION TASKS

Auditory Rhyme Judgement Taks

In the auditory rhyme judgment task the participant hears 2 words spoken by a native speaker of Belgian Dutch through headphones. The participants' task is to indicate with a button press whether the 2 words rhyme (green button on the right) or whether they do not (red button on the left). During this task the participant relies on the external monitoring route to determine whether the 2 words rhyme or not. Of the presented word pairs 50% rhyme.

The experiment begins with three practice trials in which the participant receives feedback, followed by 60 experimental word pairs. The word pairs are the same in the visual presentation condition as in the auditory presentation condition. The total duration of the rhyme task was approximately 10 minutes.

Network Perception Task

In the perception condition of the network task (see below) the participant hears a native speaker of Dutch explain the route of the dot through the network. At the same time the participant sees the visual display of the dot moving through the network. In total four networks were presented in the perception condition. In total 22 errors were distributed over the four networks. The errors were scripted, and constructed in such a way so they varied with respect to the origin of the error, and to have an equal distribution of errors over the 4 networks. Of the errors 3 were phonological, all others were of a semantic nature. Of the phonological errors, 2 were repaired. Of the semantic errors 5 related to the picture, 4 related to the shape of the line, 4 errors related to the location of line or the picture, 5 were related to the direction of the dot movement, and one was related to the color of the dot. Of the semantic errors, one error relating to the direction of the movement, was corrected. During the perception task the network descriptions continued while the participant reported an error, therefore the participants were instructed to report the detection of an error by saying 'yes'. The total duration of the network perception task was approximately 3 minutes.

VERBAL MONITORING TASKS

Network Task

In the network task (Oomen & Postma, 2001) the participant is presented with a display in which 5 simple black and white drawing of everyday objects or animals are presented. The objects are connected through lines. At the sound of a beep a red dot appears in the network, and starts to move through the network, following the lines and pictures. The participant's task is to describe the trajectory of the dot through the network, making sure to name the picture, the direction of the movement, the curvature of the line (straight or curved), and the orientation of the line with respect to the other lines (left, right, or middle). The speed of the dot was determined by presenting the participants with networks with increasing speed. The fastest speed at which the participant was still able to give an adequate description was chosen for subsequent presentation. After speed selection, the participants were presented with 16 networks. During the presentation of eight of these networks, the participant heard a loud white noise (89.6 dB). The loudness of the white noise was similar to that in Oomen et al. (2001) (90 dB). During these trials, the participant canot use their auditory feedback for verbal monitoring, and therefor have to rely on internal speech for monitoring. Although we cannot guarantee that participants did not hear their external speech, we are sure that the perception, if any, was heavily impaired. The order of the networks and the presentation with or without noise was counterbalanced over participants. The total duration of the network description task was between 10 and 15 minutes.

From the descriptions of the participants we first calculated how complete the descriptions were; whether the picture was named, the direction of the ball was named, the form of the line was named (straight / curved) and the orientation with respect to the other lines was named (top / bottom / middle). Monitoring performance was measured by counting the number of errors that are made and, perhaps more importantly, the percentage of those errors that were repaired. Per error, we report the numbers of syllables involved. We differentiate between a number of error types. First of all, there are several types of disfluencies. We counted the number of reformulations, which can be seen as a repair without the initial production of an error (e.g.

'via een – naar rechts' *via the- to the right*). The number of repetitions, in which the participant repeats the same words without adjustment (e.g. 'naar het – naar het potlood' *to the – to the pencil*) were also counted. Finally, we counted the number of filled pauses, in which we incluided prolongations (e.g. 'kromme lijn eeeeh naar rechts' *curved line eeeh to the right*). Other error types differentiated were semantic errors, phonological errors and grammatical errors. Per error type, the total number of errors is reported, and the percentage of errors which were repaired. Examples of repaired errors of these types are: grammatical repaired errors 'naar het – de deur' (to $\text{the}_{\text{neuter}}$ – $\text{the}_{\text{common}}$ $\text{door}_{\text{common}}$), phonological repaired errors 'trof- tros druiven' (a buns- a bunch of grapes), semantic errors 'vaatwasser, nee, wasmachine' (dishwasher no washing machine).

PROCEDURE

All participants performed the tasks in a similar order. First the questionnaire and the mini mental state examination were filled out, followed by the network production task, and the network perception task. Then the homophone task, the auditory and visual rhyme task, and the phoneme-monitoring task were carried out. Next the COWAT, BNT and finally Ravens Matrices were performed.

DATA ANALYSIS

All data were analyzed using SPSS 22. Comparisons between the two groups of participants were performed with an independent samples t-test for normally distributed data, or with the Mann-Whitney Test if normality could not be assumed. Within group comparisons for task performance were compared with a paired t-test or an ANOVA for normally distributed data, or with the Wilcoxon Signed-rank Test if normality could not be assumed.

RESULTS

We report the results of the tasks in the following order: control variables (dexterity, MMSE, intelligence as measured by RAVEN), speech production variables (COWAT, BNT, production speed during network task), internal speech perception (visual rhyme task, homophone task and phoneme monitoring task), external speech perception tasks (auditory rhyme task and network perception task) and finally the verbal monitoring variables (which were obtained in the network task). All trials that deviated more than 3 SD from the mean were removed from the data set. Per task we report how much data were lost.

To check whether the different tasks in each measure showed internal consistency, we computed correlations between the scores on the different tasks, which we report at the end of each section. For this correlational analysis we used the scores of the two groups together. Only significant correlations are reported, or the lack of any significant correlations.

Finally we report the result of a regression analysis, where we test whether verbal monitoring behavior is related to performance on any of the control variables, speech production variables, internal speech perception and external speech perception tasks.

CONTROL VARIABLES

An overview of the scores on the control variables is presented in Table 1.

Edinburgh Handedness Inventory

In the PD group, one of the participants was left handed (score of -58.3), and all others were right handed (score of 75 or above). In the control group also one of the participants was left handed (score of -58.3), and all others were right handed (score of 83.3 or above).

Mini Mental State Examination

The mini mental state exam was performed to have an indication of the cognitive health of the participants. In the Parkinson's patients group, 2 participants scored 23, indicating a mild cognitive impairment, and the remaining 16 patients scored between 24 and 30, indicating no cognitive impairment. In the control group, 1 participant scored 23, indicating a mild cognitive impairment and the remaining 15 participants scored between 24 and 30, indicating no cognitive impairment. An independent samples t-test determined that the groups were matched with respect to cognitive health, as the scores did not differ significantly ($t_{(32)}$= 1.210, p=.235).

Ravens Progressive Matrices

Ravens progressive matrices were measured to have an indication of the non-verbal intelligence of our participants. An independent samples t-test determined that the performance on the Ravens Progressive Matrices did not differ significantly between the two groups ($t_{(32)}$= .772, p=.446), and we can therefore assume that the patient and control groups are matched on non-verbal intelligence. A comparison with the norms, as presented in Appendix 5A, shows that our participants fit nicely within the mean.

Table 1 *Mean score and standard deviation per group on age, gender, the Edinburgh Handedness Inventory, Mini Mental State Exam and Ravens Progressive Matrices*

	Patients	Controls
Age	65.72 (8.47)	64.19 (8.32)
Gender	4 female 14 male	5 female 11 male
Handedness (SD)	85.42 (36.60)	82.03 (41.82)
MMSE (SD)	26.83 (2.33)	27.75 (2.05)
Ravens Matrices	25.83 (4.96)	27.13 (4.76)

Age and performance on the MMSE were negatively correlated (r=-.343, p=.047).

SPEECH PRODUCTION VARIABLES

Controlled Word Association Task

The COWAT measures verbal fluency. The mean scores on this task are reported in Table 2. Both groups generated approximately the same number of items in both the phonological and semantic fluency task. An independent samples t-test determined that there was no significant difference between performance on the two groups for the phonological fluency ($t_{(32)}$= 1.267, p=.214), nor for the semantic fluency ($t_{(32)}$= 1.162, p=.254).

Table 2. *Phonological and Semantic fluency performance on the COWAT task*

	Phonological		Semantic	
	Patients	Controls	Patients	Controls
N items	29.22 (10.65)	33.75 (10.12)	30.72 (7.79)	34.13 (9.29)

Performance of the two groups separately for the phonological fluency categories (N, A, K) and the semantic fluency categories (animals, professions), as well as a comparison with normative data from Miatton et al. (2004) is given in Appendix 5B.

Boston Naming Task

The Boston naming task measures confrontational word retrieval. The mean scores on this task are reported in Table 3. The Mann-Whitney test revealed that there was no significant difference between performance of the two groups (U=104.00, z=1.385, p=.085).

Table 3 *Mean score per group on the Boston naming task*

	Patients	Controls
Items named correctly	52.78 (4.04)	51.13 (4.27)
Normative data	56.32	57.84

Compared to the normative data for Dutch speaking Belgian elderly from Marien, Mampaey, Vervaet, Saerens and De Deyn (1998), which were constructed on the basis of 200 native Dutch speaking elderly, both groups perform below normal. The norms are calculated per age group. As our groups have slightly different ages, the normative score per group also differs. The difference in performance as calculated with the Wilcoxon signed ranks test, was significant for the patient group (z=2.201, p=.026, r=.52) and the control group (z=3.363, p<.0013, r=.84). Interestingly, the difference between the normative data and the scores of the current experiments is largest for the control group. This is somewhat surprising, as the control group is expected to perform similar to the normative data.

Network Task Production Speed

For each participant the production speed was calculated. From two networks with noise (Network 9 and Network 11) and two networks without noise (Network 8 and Network 10) the number of produced syllables was divided by the duration of the description (counted from speech onset of the first word until speech offset of the final word). A summary of these data are provided in Table 4. These networks were chosen as the participants were accustomed to the task and the noise condition, but they were not as fatigued as at the end of testing.

A condition x group repeated measures ANOVA for the length of the descriptions, measured in seconds from the first word until the end of the last word, yielded a significant main effect of condition (F (1,32) = 13.35, p = .001), and a significant effect of group (F (1,32) = 7,15), p=.012), but no significant interaction with group (F (1,32) = .06, p = .814). Post-hoc analyses were performed with a Wilcoxon Signed-rank test and the Mann-Whitney test, as the data were not normally distributed. Between groups, there was a significant difference between the durations of the descriptions for the normal feedback condition (U=73.00, z=2.458, *p*=.006, *r*=.42) and the noise condition (U=77.50, z=2.306, *p*=.010, *r*=.40), as the controls finished their descriptions faster in both conditions than the PD patient group. Within groups, the network descriptions under noise were significantly longer than those under the normal feedback condition for both the patient group (*z*=2.540, *p*=.005, *r*=.60) and the control group (*z*=1.831, *p*=.033, *r*=.46).

A condition x group repeated measures ANOVA for the production speed, as measured by the number of syllables produced per second, yielded a significant main effect of condition (F(1,32) = 11.39, p = .002), but no significant effect of group (F(1,32) =.73, p= .399). There was also no interaction of condition x group (F(1,32) = .20, p= .655). Between groups, the production speed as measured by the number of syllables produced per second, did not differ significantly between the normal feedback condition (U=127.50, *z*=.569, *p*=.290) and the noise condition (U=126.00, *z*=.621, *p*=.272). Within groups there was an effect of condition: production speed was significantly higher in the normal condition compared to the noise condition for the patient group (*z*=2.604, *p*=.003, *r*=.61) and control group (*z*=1.903, *p*=.029, *r*=.48).

A comparison with the data from Oomen et al. (2001) reveals that our participants were much faster on this task. The duration of the descriptions for their aphasic patients was on average 84 seconds, while the average duration of the descriptions of the PD patients of this study was 32 seconds. The controls in the study by Oomen et al. cannot be directly compared, as Oomen et al.'s controls described a larger network (8 objects) than our controls (5 pictures; their agrammatic patients and our PD patients also described 5 pictures). Average duration per object (i.e., total duration divided by number of objects) was similar for our controls (5.6 s) and Oomen et al.'s controls (6.8 s). The difference in description duration between the patient groups is likely the result of a considerably less impaired speech production for PD patients compared to Broca's aphasics.

Table 4. *Speed and duration of the network descriptions by the patient and control group.*

	PD Patients		Controls	
	Normal	Noise	Normal	Noise
Mean syllables per second (SD)	2.61 (0.64)	2.45 (0.67)	2.77 (0.59)	2.64 (0.58)
Mean duration of the network description in seconds (SD)	31.02 (5.21)	32.16 (5.05)	27.56 (2.06)	28.56 (1.49)

Previous research investigating how speech rate is affected in Parkinson's disease has had some mixed results. In the current study we found a difference in overall durations of the descriptions of the networks;

PD patients took longer to complete the descriptions. However, speech rate did not differ from that of the control group. This is a result of the number of syllables produced per description, which was higher for PD patients (normal feedback mean = 79.9 syllables, noise-masked feedback mean = 77.5 syllables) than the control group (normal feedback mean = 76.3 syllables, noise-masked feedback mean = 76.1 syllables), but not significantly so (normal feedback condition U=124.0.0, z=.690, p=.506 and the noise condition U=130.50, z=.466, p=.646). The finding that speech rate was not different for the PD group compared to controls is in line with a study by Metter and Hanson (1986), who found a normal speech rate for mild cases of PD.

When correlations were calculated between the different measures of speech production, we found significant correlations between the phonologic fluency and semantic fluency, as measured by the COWAT (r=.388, p=.023). There was also a significant correlation between performance on the BNT and the semantic fluency as measured by the COWAT (r=.440, p=.009). Performance on the BNT further correlated with the speed of the network descriptions under normal feedback (r=.387, p=.024) and under noise (r=.448, p=.008). Production speed under noise correlated significantly with production speed under normal feedback. (r=.920, p<.001). Network description duration under noise correlated significantly with network description duration under normal feedback. (r=.924, p<.001). These correlations show a nice internal consistency for the speech production variables.

INTERNAL SPEECH PERCEPTION

Visual Rhyme Judgment Task

In the visual rhyme judgment task, participants silently judged whether two words presented on a computer screen rhyme or not, which allows the participant to rely on internal speech only to complete the task. Trials of which the RT deviated more than 2 standard deviations from the mean were excluded from analysis. Of the control group data 8 trials were deleted (<1%) and of the PD group 43 trials were deleted (4%). A summary of the performance on the task is given in Table 5. An extensive overview of the task performance on the rhyme task per category, as well as the normative data, is presented in appendix 5C.

Between groups there were no significant differences in performance for overall accuracy and RTs. A comparison between the rhyming word pairs and non-rhyming word pairs also revealed no between-group differences on accuracy or RTs.

There was no significant difference within the two groups on the judgment accuracy for the rhyming word pairs compared to the accuracy of the non-rhyming word pairs. RTs also did not differ significantly between the two conditions.

Table 5. *Mean accuracy and mean reaction times in milliseconds in the visual rhyme judgment task. Standard deviations are in brackets.*

	Patients		Controls	
	Rhyme	Non-rhyme	Rhyme	Non-rhyme
Accuracy	83.08 (37.53)	82.21 (38.28)	82.60 (37.95)	84.42 (36.30)
RT (ms)	2634 (1493)	2674 (1603)	2538 (1247)	2689 (1450)

Homophone Decision Task

In the homophone decision task participants rely on internal monitoring to decide whether two words, or two non words, are homophones or not. Trials of which the RT deviated more than 2 standard deviations from the mean were excluded from analysis. Of the PD group 38 trials were deleted (4.8%) and of the control group data 26 trials were deleted (3.7%). A summary of the performance on the task is given in Table 6. A more extensive table with the data per condition and the normative data can be found in 5D.

Table 6. *Mean accuracy and reaction times and SDs in the homophone task*

	Patients		Controls	
	Homophone	Non-homophone	Homophone	Non-homophone
Accuracy	84.31 (36.42)	83.86 (36.84)	79.10 (40.72)	85.71 (35.04)
RT (ms)	2789 (1308)	3054 (1349)	2794 (1275)	2823 (1213)

There were no significant differences between groups with respect to accuracy or RTs for the variables judgment of homophone word-pairs, judgment of non-homophone word-pairs, and overall performance on the homophone task.

Separate analyses for each group revealed that the PD group showed no difference between accuracy scores of the homophones and non-homophones, but there was a significant difference between the RTs; the homophones were judged faster (M = 3168 ms) compared to the non-homophones (M = 3168 ms), $t_{(17)}$= 2.910, p=.010. There was no significant difference between the accuracy or the RTs between the homophone and non-homophone condition for the control group.

Phoneme Monitoring Task

In the phoneme monitoring task, participants rely on internal speech to determine whether a letter is part of the name of a picture or not. Trials of which the RT deviated more than 2 standard deviations from the mean were excluded from analysis. Of the PD group 58 trials were deleted (3.6%) and of the control group data 49 trials were deleted (3.4 of %). A summary of the performance on the task is given in Table 7. A comparison with the data of Özdemir, Roelofs, and Levelt (2007) is given in appendix 5E.

A phoneme position x group repeated measures ANOVA revealed a main effect of phoneme position on the accuracy scores (F (2, 64) = 5.124, p=.009), as the phonemes were judged significantly more accurate in initial position (96%) compared to the medial (91.5%) and final (91.4%) position. There was no significant effect of group (F (1,32) = .823, p=.371). Accuracy scores showed no interaction with the group. The RTs also showed main effect of phoneme position (F (2, 64) = 47.550, p<.001), with the phoneme being judged significantly faster in initial position (1821 ms) compared to in medial (2189 ms) and final (2173 ms) position. There was no significant RT x group interaction.

Table 7. *Mean accuracy and mean reaction times in milliseconds on the phoneme monitoring task, with standard deviations in brackets. Results are displayed per location of the target phoneme. For example. the target phoneme /N/ corresponds to the following conditions: initial 'naald', medial 'panda', final in 'ballon'.*

	Patients			Controls		
	Initial	Medial	Final	Initial	Medial	Final
Accuracy	95.08 (21.68)	87.45 (33.19)	87.11 (33.78)	97.02 (17.04)	95.65 (20.44)	96.12 (19.35)
RT (s)	1844 (721)	2260 (799)	2225 (758)	1795 (538)	2099 (639)	2089 (600)

In sum, the internal speech tasks show no significant difference between the PD patient group performance and the control group performance. Comparison of the data with the result from the study by Özdemir et al., (2007) revealed a similar performance on accuracy between our groups and the participants of the Özdemir et al. However, RTs are much slower for our groups. This is not surprising, as our groups are elderly, and the participants in the study by Özdemir et al. are not.

There were significant correlations between all three internal speech measures. Accuracy on the visual rhyme task correlated significantly with the accuracy on the homophone task (*r*=.487, p=.003), and with accuracy on the phoneme monitoring task (*r*=.601, p<.001). Accuracy on the homophone task correlated with accuracy on the phoneme monitoring task (*r*=.425, p=.012). These correlations show internal consistency for the internal speech perception tasks.

EXTERNAL SPEECH PERCEPTION TASKS

Auditory rhyme task

In the auditory rhyme judgment task, participants judged whether two words presented auditorily via headphones rhyme or not, which allows the participant to rely on external speech monitoring only to complete the task. Trials of which the RT deviated more than 2 standard deviations from the groups' mean were excluded from analysis. Of the control group data 35 trials were deleted (3.6%) and of the PD group 6 trials were deleted (<1%). A summary of the performance on the tasks is given in Table 8. A detailed overview of the task performance on the rhyme task per category, and normative data, is presented in appendix 5C.

Table 8. *Mean accuracy and mean reaction times in milliseconds in the auditory rhyme judgment task. Standard deviations are in brackets.*

	Patients		Controls	
	Rhyme	Non-rhyme	Rhyme	Non-rhyme
Accuracy	88.87 (31.48)	79.33 (40.53)	88.36 (32.10)	88.72 (31.67)
RT (ms)	2792 (1297)	2995 (1298)	2564 (626)	2647 (605)

The groups did not differ in their overall accuracy and RTs. A comparison between the rhyming word pairs and non-rhyming word pairs also revealed no between group differences on accuracy or RTs.

There was no significant difference within the two groups on the judgment accuracy for the rhyming word pairs compared to the accuracy of the non-rhyming word pairs. RTs also did not differ significantly between the two conditions.

A comparison between the performance of the auditory rhyme task and the visual rhyme task showed a significant difference in accuracy of the rhyming word pairs, with the auditory presented rhyming word pairs being judged more accurate than the visual presented rhyming word pairs for both the PD patient group ($t_{(17)}$= 2.655. p=.017), and the control group ($t_{(15)}$= 2.207. p=.043). However, from a more detailed analysis, presented in appendix 5C, it is clear that the orthographic transparency of the word pairs drive this difference in the rhyming conditions; the orthographically untransparent word pairs decreased accuracy in the visual rhyme task, but not decrease accuracy in the auditory rhyme task. All other conditions show no significant difference between the visual and auditory rhyme task.

Network perception task

In the network perception task, participants listened to the description of four networks, similar to the descriptions the participants gave in the production network task. The participants were instructed to immediately report an error if they heard one. A summary of performance on the task is found in Table 9. A independent samples t-test revealed that there was no significant difference between the performance of the two groups on the number of semantic errors detected ($t_{(32)}$= 0.326. p=.746). The small number of phonological errors that were presented (n=3) were not enough to make a statistical comparison on.

Table 9. *Performance on the network description error detection task for PD patients and the control group.*

	PD Patients		Controls	
Errors detected	N	%	N	%
Semantic	11.50 (3.38)	55.63 (16.52)	11.13 (3.30)	57.50 (16.91)
Phonological	0.67 (0.69)	22.22 (22.87)	0.94 (0.77)	31.25 (25.73)

The percentage of errors detected contrasts markedly with the findings of Oomen et al., (2001) who report an error rate of around 85% for controls and 65% for aphasics. Both our groups detect just over half of the errors (both groups 55%). Closer inspection of the network task reveals that our network task was probably rather difficult. Four of our semantic errors were replacements by an item that was semantically close (e.g., car by truck, rhino by elephant, etc.). Furthermore, two errors followed very quickly upon the previous error, possibly resulting in an attentional 'deafness' to detect the second one. Interestingly, there is a big difference in the report rate of the two items that were repaired. While both items had a similar timing pattern, the first error is only detected 8 times (PD group reported this error 1 time), and the second error is detected 14 times (PD group reported this error 7 times).

Repaired item 1. „mus [silence 240 ms][eh 330]muts'' *sparrow, uhm, cap*

Repaired item 2. „boek[silence 251 ms][eh 273] bloem'' *book, uhm, flower*

The external speech tasks, in sum, did not reveal any significant differences between performance of the patient group and the control group. Finally, there was a significant correlation between accuracy of the performance on the auditory rhyme task and accuracy of performance on the network perception task (r=.358, p=.038).

VERBAL MONITORING VARIABLES

Recordings of the network descriptions were transcribed. Errors were coded for type of error and whether they were repaired. Early interrupted errors were errors that were interrupted within the first syllable. As there is not enough information to further classify the error (as semantic, phonological, or grammatical), early interrupted errors were counted as a separate group that was always repaired.

A summary of the data is provided in Table 10. Under the normal feedback condition participants can use both their internal and external monitoring route to detect errors. In healthy participants more errors are detected under normal feedback, in which case the participant can monitor both internal and external speech. In the noise masked feedback condition, external speech is masked by a loud, white noise, allowing only internal speech to be monitored. Under noise masked speech production by healthy participants, fewer speech errors are detected, indicating the relative contribution of the external monitoring channel (e.g. Postma and Noordanus, 1996). If internal monitoring is impaired compared to external monitoring, we expect the noise masking to affect monitoring behavior. Under noise masking the participant can only monitor internal speech, so if internal monitoring is impaired, we would expect the participants to make more errors and repair fewer of them under the noise masking condition, unless participants make less errors under noise masking.

A condition x error type x group repeated measures ANOVA demonstrated a significant main effect of error type ($F(2.365, 75.69) = 76.33$, $p<.001$). The number of errors of all types (early interrupted errors, semantic errors, phonological errors, grammatical errors and disfluencies) all differed significantly from another ($p<.001$), except for the grammatical and phonological errors. No significant main effects were observed for group nor of condition. Furthermore, no significant interactions were observed.

Table 10. *Summary of the total number of errors in the network descriptions under normal and noise-masked feedback by the PD patients and control group.*

	PD Patients		Controls	
	Normal	Noise	Normal	Noise
Early interrupted errors	44	34	40	47
Semantic	119	119	93	83
Phonological	15	17	8	7
Grammatical	17	15	16	3
Disfluencies	169	156	168	197
Total Number of errors	364	341	325	337

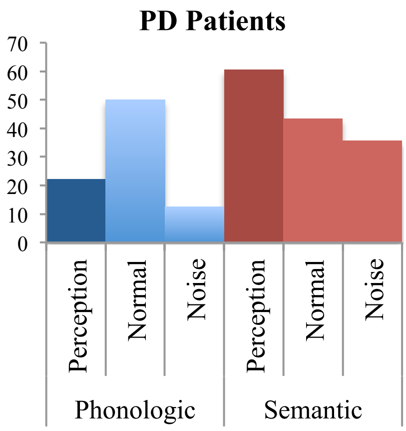
An overview of the repair behavior of the participants is provided in Table 11. Figure 1 shows the percentage of phonological and semantic errors that were repaired. To test the effect of noise masking on repair behavior (percentage of errors repaired) a condition x error type x group repeated measured ANOVA was performed on the data. Note that we did not take disfluencies into account here, as it is unclear whether a disfluency is the repair of an error. It might, for example, also be indicative of search behavior. Below disfluencies are extensively addressed. As the phonological errors and the grammatical errors did not have enough values for a relevant analysis, the analysis was performed on the total number of errors repaired in the normal and noise-masked feedback, and on the semantic errors. For the total percentage of errors repaired, so significant main effect or interactions were observed. Under normal feedback a mean of 49.8% of the errors was repaired, while under the noise-masked feedback condition 25.2% of the errors was repaired. However, for the semantic errors a significant effect condition was found ($F(1, 25) = 5.726$, $p = .025$). Under noise masking

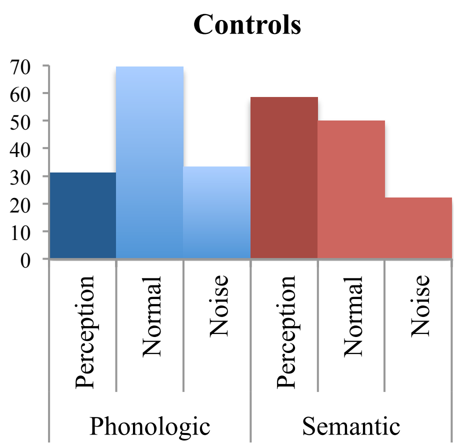
30.1% of the errors was repaired, while under normal feedback 46.6% of the errors was repaired. A post-hoc analysis of the data revealed that for the patient group there was no significant difference between the percentage of errors repaired under normal feedback and under noise masked feedback. In the control group there was a significant difference between the percentage of semantic errors repaired under normal and noise-masked feedback ($t(11) = 3.618$, p=.004. For the total percentage of errors repaired, no significant difference was observed between the two feedback conditions.

Table 11. *Summary of the percentage of corrections per error type in the network descriptions by the PD patients and control group.*

	PD Patients		Controls	
	Normal	Noise	Normal	Noise
Semantic	43.37	35.64	50.10	22.22
Phonological	50.00	12.50	69.44	33.33
Grammatical	57.43	35.83	43.33	100
Total	47.71	31.39	48.08	30.28

Figure 1. *Percentage of errors repaired per category. In blue are phonological errors, in red the semantic errors. In dark colors are the detection rates during the network perception task*

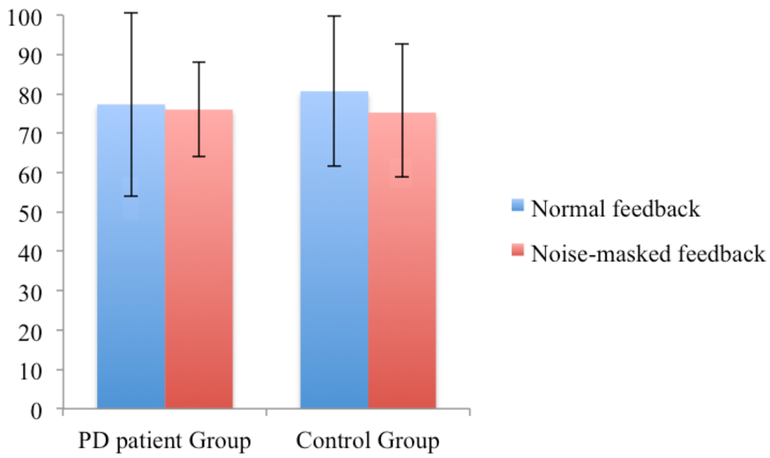




Disfluencies

In the verbal monitoring behavior analysis of Oomen et al., (2001) the disfluencies are also counted as repaired errors. In order to compare our data with those of Oomen et al., (2001), the percentage of errors repaired was calculated with the semantic, phonologic and grammatical repaired errors, plus the early interrupted errors and disfluencies. See Figure 2. The percentage of errors repaired under normal feedback is similar to that of the controls in the study by Oomen et al., (2001). The number of errors repaired under noise masked feedback is much higher for both our groups than for the control group in the study by Oomen, who repaired around 60% of their errors under noise masked feedback. The Broca's aphasics repair between 50 and 60% of their errors, while our PD patients repair between 75 and 80% of their errors. A condition x group repeated measures analysis revealed no significant main effect of condition, nor of group. Also the interaction of condition x group was not significant.

Figure 2. *Percentage of overt and covert errors repaired per feedback condition*



Disfluencies, called prepairs by Schlenck et al., (1987) or covert repairs by Levelt (1983; 1989) have sometimes been considered repair behavior in the absence of an overt error (Postma & Kolk, 1993). It is often taken to be the result of pre-articulatory, internal speech monitoring. It is therefore an interesting category of errors to investigate. Here we distinguish between reformulations, filled pauses, and repetitions. An overview of the disfluencies per type is given in table 12.

A disfluency type x condition x group ANOVA was performed on the data. No main effects were observed, but there was a significant interaction of condition x group ($F(1,32) = 4.35$, $p = .045$), with the Parkinson patients making more disfluencies under noise masking than under normal feedback, and the controls making more disfluencies under normal feedback than under noise. Post-hoc analysis revealed no difference between the number of disfluencies between the normal and noise-masked feedback condition for both the PD and control group. There was also an interaction of disfluency type x group ($F(2, 64) = 3.54$, $p = .035$), reflecting the different error patterns; PD patients made mostly reformulations ($M = 4.67$), followed by (repetitions $M = 3.33$), and least filled pauses ($M = 3.25$). The control group made mostly filled pauses ($M = 4.78$), followed by reformulations ($M = 2.47$), and least repetitions ($M = 1.68$).

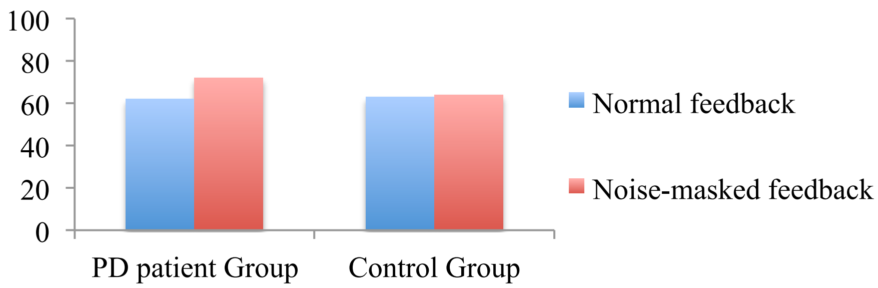
In the study of Oomen et al., (2001) the disfluencies account for around 65% of the repaired errors in both normal and noise-masked monitoring. This is comparable to the PD patients under the normal feedback. For their controls, however, the disfluencies only accounted for 30% of the repaired errors, while in our study the controls actually perform very similar to the PD patients. In fact, the only difference is that in the PD patient group disfluencies account for slightly more repaired errors in the noise masked feedback condition.

Table 12. *Total number of disfluencies per error type in the network descriptions by the PD patients and control group.*

	PD Patients		Controls	
	Normal	Noise	Normal	Noise
Reformulation	89	79	41	38
Filled Pause	49	68	89	64
Repetition	40	80	29	25
Total	178	227	159	127

Figure 3 shows the percentage of disfluencies as a percentage of the total number of overt and covert repaired errors, demonstrating that they count for a large proportion of the repaired errors in all conditions.

Figure 3. *Disfluencies as a percentage of the total number of (overt and covert) repaired errors.*



A correlational analysis of the data revealed a significant correlation between the number of errors repaired between the normal feedback and the noise masked feedback condition (r=.439, p=.009). The number of errors made correlated between the normal feedback and the noise masked feedback condition (r=.612, p<.001). Also the number of semantic errors correlated between the normal feedback and the noise masked feedback condition (*r*=.422, p=.013).

Unsurprisingly, the number of errors produced during normal feedback correlated with the number of errors repaired during both normal feedback (*r*=.853, p<.001), and a bit less with noise masked feedback (*r*=.529, p=.001). The number of errors produced during noise masked feedback correlated with the number of errors repaired during both normal feedback (*r*=.495, p=.003), and more strong with noise masked feedback (*r*=.942, p<.001).

The number of errors repaired during normal feedback correlated with the number of disfluencies during both normal feedback (*r*=.901, p<.001), and a bit less with noise masked feedback (*r*=.352, p=.041). The number of errors repaired during noise masked feedback correlated with the number of disfluencies during both normal feedback (*r*=.440, p=.009), and more strongly with noise-masked feedback (*r*=.894, p<.001), and with the number of semantic errors produced under noise (*r*=512, p=.002) and the number of grammatical errors produced under noise (*r*=344, p=.046).

The total number of errors produced under normal feedback correlated highly with the number of disfluencies produced under normal feedback (*r*=813, p<.001), the number of semantic errors produced under normal feedback (*r*=584, p<.001) and to a lesser extent with disfluencies produced under noise-masked feedback (*r*=459, p=.006), and semantic errors produced under noise masked feedback(*r*=482, p=.004).The total of errors produced under noise-masked feedback correlated highly with the number of disfluencies produced under noise-masked feedback (*r*=829, p<.001), the number of semantic errors produced under noise masked feedback (*r*=707, p<.001), and to a lesser extent with disfluencies produced under normal feedback (*r*=437, p=.010), semantic errors produced under normal feedback(*r*=429, p=.011), phonological errors produced under noise-masked feedback (*r*=443, p=.009) and grammatical errors produced under noise-masked feedback (*r*=342, p=.048). An unexpected and presumably spurious correlation is that disfluencies produced under noise correlate with semantic errors under normal feedback (*r*=353, p=.041).

MULTIPLE REGRESSION

A forward multiple regression analysis was used to asses whether verbal monitoring behavior can be predicted by the measured variables of cognitive performance, speech production, external speech perception and internal speech perception. To avoid multicollinearity between predictors, average production speed and description duration were calculated from the measures during the two feedback conditions. Independent variables in the regression were formed by the variables: group, age, handedness, MMSE score, Raven's matrices score, phonologic fluency, semantic fluency, BNT score, homophone task accuracy, phoneme monitoring task accuracy, visual rhyme task accuracy, auditory rhyme task accuracy, semantic errors detected during perception, phonologic errors detected during perception, production speed and duration of the network descriptions. Dependent variables were the percentage of errors produced during normal feedback, the percentage of errors produced under noise-masked feedback, the number of errors corrected under normal feedback and the number of errors corrected under noise masked feedback, semantic errors made during normal feedback, semantic errors made during noise-masked feedback, phonological errors made under normal feedback and phonological errors made under noise-masked feedback. Results of the forward regression are presented in table 13.

Table 13. *Results of the regression analysis with monitoring variables as dependent variables, and measured variables of cognitive performance, speech production measures, external speech perception, internal speech perception as independent variables.*

	B	SE B	β	R^2	Sig.
Number errors produced under normal feedback					
Constant	-10.38	10.91			
Age	0.47	0.17	.45	.20	<.001
Number of errors produced under noise-masked feedback					
Constant	3.35	13.46			
Age	.428	0.18	.36	.12	.022
Phonologic fluency	-.358	0.14	-.38	.23	.004
Percentage of errors repaired under normal feedback	-				
Percentage of errors repaired under noise-masked feedback					
Constant	-.145	.217			
Production speed under noise	.179	.083	.37	.14	.039
Percentage semantic errors repaired under normal feedback	-				
Percentage semantic errors under noise-masked feedback	-				

From the regression analysis, we expected to find that monitoring behavior dependent on the internal monitoring channel, as measured during noise-masked feedback, would be predicted by performance on the internal speech tasks. External monitoring behavior was expected to correlate with external monitoring task performance. That the production of errors is best predicted by age (in both feedback conditions and specifically semantic errors under noise-masked feedback), is concurrent with the consistent lack of differences between the performance of the two groups. The number of errors produced was in both feedback conditions predicted by age. The number of errors produced under noise masked feedback was additionally predicted by a negative correlation with phonologic fluency as measured by the COWAT, and not an internal speech task. The percentage of errors repaired under noise-masked feedback was predicted by the production speed under noise; it is possible that repair behavior increased production speed.

To assess the effect of the Parkinson on the variables measured in this study, a separate multiple regression analysis was performed in which the reported length of the PD formed the dependent variable, and independent variables were fluency, semantic fluency, BNT score, homophone task accuracy, phoneme monitoring task accuracy, visual rhyme task accuracy, auditory thyme task accuracy, semantic errors detected during perception, phonologic errors detected during perception, production speed and duration of the network descriptions, the percentages and total number errors produced, the number of grammatical errors, semantic errors, phonological errors and disfluencies, all separately for the two feedback condition. From this regression analyses, we found that length of PD was only significantly predicted by a negative relation with the production speed under normal feedback as presented in table 14.

Table 14. *Results of the regression analysis with duration of PD as dependent variable, and task performance as independent variables.*

	B	SE B	β	R^2	Sig.
Errors produced under normal feedback					
Constant	27.88	4.98			
Production speed under normal feedback	-6.43	1.91	-.68	.43	.005

The outcome of this analysis suggests that from the measures taken in this study, we only found our PD patients to be affected in the production speed. The longer a patient was affected by PD, the slower their production speed.

DISCUSSION

The current study set out to investigate whether verbal monitoring is impaired in PD patients, compared to healthy controls. More specifically, the reliance on the internal and external monitoring channel was investigated. A third point of interest was the question whether internal verbal monitoring behavior could be predicted from internal speech monitoring tasks. We measured performance on several variables: cognitive performance, speech production performance, external speech perception, internal speech perception and verbal monitoring behavior. Each of these variables was measured via a number of tasks.

Within each of the variables, the tasks demonstrated a good coherence, as measured via a correlational approach. Cognitive performance tasks correlated with age of the participant. Speech production performance, as measured with the BNT, COWAT, and the production speed of the network task, all correlated well. The internal speech perception tasks, phoneme monitoring, homophone judgment, and visual rhyme judgment, all correlated significantly with each other. External speech perception, as measured with the auditory rhyme task and network perception task, also correlated well. Verbal monitoring, as measured by the errors produced and repaired during normal feedback and noise masked feedback, also showed a high inter variable consistency.

To investigate if PD patients have impaired verbal monitoring, we compared performance of PD patients with the performance of healthy controls on a number of tasks. On the 11 measured tasks we only found a few differences in performance between the PD patient group and the healthy controls; the duration of the descriptions differed and the percentage of semantic errors repaired. We also found a difference between performance of our group, and the normative data when they were available. This suggests that language performance in our groups is affected, but not by PD. This leaves age as the most likely variable to affect the outcome, although compared to normative data of the same age cohort, our participants scored relatively bad on the BNT.

Contrary to a vast body of literature demonstrating impaired speech perception in PD patients, we found no difference in performance between the PD and control group on most of the tasks administered. The lack of difference between the PD group results and control group's results on almost all of the tasks suggests that perhaps the patient population of this study was only mildly affected by PD. Also the fact that the speech rate between PD patients and controls did not differ, suggest that our participants might be only mildly affected by Parkinson's disease. This is in line with previous research showing no decrease or increase in production speed for mild PD patients (Metter & Hanson, 1986). A comparison between the self-corrections in this study with a previous study investigating PD patients (McNamara et al., 1992) shows almost contradictory findings. In McNamara et al., (1992) PD patients produced thrice the number of errors compared to healthy controls, and only repaired 25%. In the current study, PD patients produce and repair as many errors as the control group.

The second question of interest was whether PD patients, similar to Broca's aphasics, rely more on their internal monitoring route for speech monitoring, or whether they use their internal and external monitoring route similarly to healthy adults. This question was investigated by comparing verbal monitoring performance during noise-masked speech, when only the internal monitoring route is available for monitoring, and during normal feedback when also the external monitoring route can be used. Most interestingly, for the semantic errors there was an effect of condition; the control group repaired significantly more semantic errors under normal feedback than under noise-masking, while the PD group did not. A comparison of the total percentage of errors repaired in the two conditions revealed no difference in performance, suggesting that PD patients indeed do rely more on their internal monitoring route for verbal monitoring. However, the same result was obtained for the control group; when the percentage of all errors repaired was taken into account, there was no effect of condition. That noise masked feedback only affected semantic error monitoring suggests that different error types might be monitored by different monitoring routes. It is not very surprising that different errors could use different monitoring channels; a lot of patient data demonstrate a difference in semantic and phonological processing.

A possible explanation for the lack of a main effect of noise-masking on monitoring performance of these groups, could be that the manipulation was not successful. While indeed it is possible that the noise did not completely drown out the speech produced by the participants, it is unlikely that the white noise produced at 90 dB would not have interfered with the perception of speech via the ears. Even more so as the same noise at the same volume was successful in manipulating monitoring performance in a previous study (Oomen et al., 2001).

Our third question of interest was to investigate whether internal speech monitoring could be predicted by performance on internal speech tasks. With a regression analysis we tested whether speech-monitoring behavior could be predicted by the performance on the internal and external speech tasks. According to the perceptual loop theory, internal speech monitoring takes place via perception of internal speech. This theory thus assumes a tight link between internal speech monitoring and internal speech perception. In the current study we did not find evidence to support monitoring via internal speech perception, as internal monitoring was not significantly predicted by performance on any of the internal or external speech tasks. Also external speech monitoring was not significantly predicted by any of the internal or external speech tasks. The outcome of this analysis suggests that verbal monitoring might be occur largely independent of speech perception.

The high degree of overlap in performance between the two groups, and the comparison with normative data brings us to two caveats of this study. First of all, we have no proper measure of how far the disease has progressed in our PD group. We only have the self-reports of the onset of the disease. A professional estimation, such as the score on the UPDRS, could give insight into whether these patients are only mildly affected. Without this measurement it remains unclear whether the lack of difference between the PD patients is the result of testing patients that are mildly affected by PD, or whether PD has not affected the processes underlying the variables measured in this study.

A second caveat is that demographic data on educational levels was not gathered for these participants. This makes a comparison of the data with normative data troublesome, as educational level is used as a discriminating factor between some of the normative data available (BNT and COWAT). The low scores on the BNT and COWAT could possibly be the result of low educational levels. However, as we recruited the partners of the patients for our control group, it is reasonable to assume that the two groups will have a comparable educational level.

In sum, this study investigated whether verbal monitoring is affected in PD patients. Furthermore, it investigated whether PD patients rely more on the internal route for monitoring similar to Broca's aphasic, or whether the internal and external monitoring route are used as in healthy controls. We found that monitoring performance increased in the presence of external feedback, for both the healthy control group and the PD patient group. This difference in monitoring performance between normal and noise-masked feedback was statistically significant for the percentage of repaired semantic errors of the control group. However, there was no statistical significant effect of feedback when the overall percentage of errors was taken into account. This result suggests that different error types might be monitored trough different monitoring routes. The fact that there was no significant difference between the two feedback conditions for the PD group suggests that that indeed PD patients rely more on their internal route for monitoring speech, as masking the external route by a loud white noise did not affect verbal monitoring performance in any condition. However, a lack of performance difference between the control group and the PD patient group on nearly all tasks leaves open several options. First of all, it is possible that the PD patients investigated in the current study are only mildly affected by PD, and as a result their verbal monitoring skills are not affected. Secondly, PD might not affect verbal monitoring behavior, and the observed reliance on the internal monitoring route is the effect of cognitive decline associated with age. A third question investigated in the current experiment is whether internal speech perception is predictive of internal speech monitoring. We found that internal monitoring behavior was not predicted by performance on internal speech monitoring tasks, suggesting that internal speech monitoring might be a process independent of speech perception.

REFERENCES

Angwin. A. J., Arnott. W. L., Copland. D. A., Haire. M. P., Murdoch. B. E., Silburn. P. A., Chenery. H. J. (2009). Semantic activation in Parkinson's disease patients on and off levodopa. *Cortex*, 45, 950-959.

Angwin. A. J., Chenery. H. J., Copland. D. A., Murdoch. B. E., Silburn. P. A. (2005). Summation of semantic priming and complex sentence comprehension in Parkinson's disease. *Cognitive Brain Research*. 25, 78-89.

Arnott. W. L., Copland. D. A., Chenery. H. J., Murdoch. B. E., Silburn. P. A., & Angwin. A. J. (2011). The Influence of Dopamine on Automatic and Controlled Semantic Activation in Parkinson's Disease. *Parkinson's Disease*.

Aronson. A. E., Brown. J. R. (1975). *Motor speech disorders*. WB Saunders Company. Philadelphia.

Bartels, A. L., Leenders, K. L. (2009). Parkinson's disease: the syndrome, the pathogenesis and pathophysiology. *Cortex*, 45, 915-921.

Bastiaanse, R., Van Zonneveld, R. (2005). Sentence production with verbs of alternating transitivity in agrammatic Broca's aphasia. *Journal of Neurolinguistics,* 18, 57–66.

Bastiaanse, R., Hugen, J., Kos, M., van Zonneveld. R. (2002). Lexical, morphological, and syntactic aspects of verb production in agrammatic aphasics. *Brain and Language*, 80, 142–159.

Bastiaanse, R. (2008). Production of verbs in base position by Dutch agrammatic speakers: inflection versus finiteness. *Journal of Neurolinguistics,* 21, 104–119.

Bastiaanse, R., Leenders, K. L. (2009). Language and Parkinson's disease. *Cortex*, 45, 912-914.

Benton. A.L., Hamsher. K. (1976). *Multilingual Aphasia Examination*. University of Iowa. Iowa City

Berg, E., Bjornram, C., Hartelius, L., Laakso, K., Johnels, B. (2003). High-level language difficulties in Parkinson's disease. *Clinical Linguistics and Phonetics,* 17*,* 63-80.

Blackmer, E. R., Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39, 173–194.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review* 108, 624–652.

Bowles. N. L., Obler. L. K., Albert. M. L. (1987). Naming errors in healthy aging and dementia of the Alzheimer type. *Cortex,* 23, 519-524

Braak, H., Rüb, U., Steur, E. J., Del Tredici, K., De Vos, R. A. I. (2005). Cognitive status correlates with neuropathologic stage in Parkinson disease. *Neurology*, 64, 1404-1410.

Brown J.R. (1975). Motor speech disorders. In: WB Saunders, Philadelphia

Cameron, I. G. M., Pari, G., Alahyane, N., Brien, D. C., Coe, B. C., Stroman, P. W., Munoz, D. P. (2012). Impaired executive function signals in motor brain regions in Parkinson's disease. *NeuroImage*, 60, 1156–1170.

Castner J.E., Chenery H.J., Copland D.A., Coyne T.J., Sinclair F., Silburn P.A. (2007). Semantic and affective priming as a function of stimulation of the subthalamic nucleus in Parkinson's disease. *Brain*, 130: 1395–1407.

Colman, K. S., Koerts, J., van Beilen, M., Leenders, K. L., Post, W. J., Bastiaanse, R. (2009). The impact of executive functions on verb production in patients with Parkinson's disease. *Cortex*, 45, 930-942.

Connor. L. T., Spiro. A., Obler. L. K., Albert. M. L. (2004). Change in object naming ability during adulthood. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 59, 203-209.

Copland, D.A. (2003). The basal ganglia and semantic engagement: potential insights from semantic priming in individuals with subcortical vascular lesions, Parkinson's disease and cortical lesions. Journal of the International Neuropsychological Society, 9, 1041–1052.

Copland, D. A. (2003). The basal ganglia and semantic engagement: potential insights from semantic priming in individuals with subcortical vascular lesions. Parkinson's disease. and cortical lesions. *Journal of the International Neuropsychological Society*, 9, 1041-1052.

Copland. D. A., Chenery. H. J., Murdoch. B. E. (2001). Discourse priming of homophones in individuals with dominant nonthalamic subcortical lesions. cortical lesions and Parkinsons disease. *Journal of Clinical and Experimental Neuropsychology*, *23*, 538-556.

Crescentini, C., Mondolo, F., Biasutti, E., Shallice, T. (2008). Supervisory and routine processes in noun and verb generation in nondemented patients with Parkinson's disease. *Neuropsychologia*, 46, 434–447.

Critchley. E. M. (1981). Speech disorders of Parkinsonism: a review. *Journal of Neurology. Neurosurgery & Psychiatry*. 44, 751-758.

Daley, T. C., Whaley, S. E., Sigman, M. D., Espinosa, M. P., Neumann, C. (2003). IQ on the rise. The Flynn Effect in rural Kenyan Children. *Psychological Science,* 14, 215-219.

Darley. F. L., Aronson. A. E., Brown. J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech, Language and Hearing Research*. 12, 246-269.

De Letter, M., Santens, P., Van Borsel, J. (2005). The effects of levodopa on word intelligibility in Parkinson’s disease. *Journal of Communication Disorders*, 38, 187–96.

De Letter, M., Santens, P., Estercam, I., Van Maele, G., De Bodt, M., Boon, P., Van Borsel, J. (2007). Levodopa-induced modifications of prosody and comprehensibility in advanced Parkinson’s disease as perceived by professional listeners. *Clinical Linguistics & Phonetics*, 21, 783–91.

Dell, G.S. Repka, R.J. (1992). *Errors in inner speech.* B.J. Baars (Ed.), Experimental slips and human error: Exploring the architecture of volition (pp. 237–262). Plenum, New York.

Elble. R. J. (2002). Tremor and dopamine agonists. *Neurology*. *58*, S57-S62.

Fahn. S., Elton. R. Members of the UPDRS Development Committee Unified Parkinson’s Disease rating scale (1987). Recent developments in Parkinson’s disease. *Fahn SMC. Colne DB. Goldstein (eds) Florham Park NY Mac Millan Healthcare Information*. 153-163.

Flint, A. J., Black, S. E., Campbell-Taylor, I., Gailey, G. F., Levinton, C. (1992). Acoustic analysis in the differentiation of Parkinson's disease and major depression. *Journal of Psycholinguistic Research*, *21*(5), 383-399.

Folstein, M.F., Folstein, S.E., McHugh, P.R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 12. 189-198.

Gallena. S., Smith. P. J., Zeffiro. T., Ludlow. C. L. (2001). Effects of levodopa on laryngeal muscle activity for voice onset and offset in Parkinson disease. *Journal of Speech. Language. and Hearing Research* 44, 1284-1299.

Gauvin, H.S., Hartsuiker, R.J. (submitted) Monitoring self and others: error detection and its aftermath in production and perception.

Gentil. M. E. L., Tournier. C. L., Pollak. P., Benabid. A. L. (1998). Effect of bilateral subthalamic nucleus stimulation and dopatherapy on oral control in Parkinson's disease. *European neurology*. 42, 136-140.

Poluha. P., Teulings. H. L., Brookshire. R. (1998). Handwriting and speech changes across the levodopa cycle in Parkinson's disease. *Acta psychologica*, 100, 71-84.

Goldman, W.P., Baty, J.D., Buckles, V.D., Sahrmann, S., Morris, J.C. (1998). Cognitive and motor functioning in Parkinson disease – Subjects with and without questionable dementia. *Archives of Neurology, 55,* 674-680.

Golfinopoulos, E., Tourville, J. A., Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage* 52, 862–874.

Golfinopoulos, E., Tourville, J. A., Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage,* 52, 862–874.

Grodzinsky, Y. (1995). A restrictive theory of agrammatic comprehension. *Brain and Language*, 51, 26–51.

Grossman, M., Carvell, S., Gollomp, S., Stern, M.B., Reivich, M., Morrison, D., Alavi, A., Hurtig, H.I. (1993). Cognitive and physiological substrates of impaired sentence processing in Parkinson's disease. *Journal of Cognitive Neuroscience*, 5, 480–498.

Grossman, M., Carvell, S., Gollomp, S., Stern, M.B., Vernon, G., Hurtig, H.I. (1991). Sentence comprehension and praxis deficits in Parkinson's disease. *Neurology*, 41, 1620–1626,

Grossman, M., Carvell, S., Stern, M.B., Gollomp, S., Hurtig, H.I. (1992). Sentence comprehension in Parkinson's disease: the role of attention and memory. *Brain and Language,* 42, 347–384.

Grossman, M., Stern, M.B., Gollomp, S., Vernon, G. (1994). Verb learning in Parkinson's disease. *Neuropsychology*, 8, 413–423.

Grossman. M., Zurif. E., Lee. C., Prather. P., Kalmanson. J., Stern. M. B., Hurtig. H. I. (2002). Information processing speed and sentence comprehension in Parkinson's disease. *Neuropsychology*. *16*, 174.

Guenther, F. H., Hampson, M. Johnson, D. (1998). A Theoretical investigation of reference frames for the planning of speech movements. *Psychological Review,* 105, 611–633

Hammen. V. L. (1990). *The effects of speaking rate reduction in parkinsonian dysarthria*. Dissertation Abstracts International. 51/09B. 4304 (University Microfilms No AAC 9104243).

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscienc*e 13,135-145

Hochstadt, J., Nakano, H., Lieberman, P., Friedman, J. (2006). The roles of sequencing and verbal working memory in sentence comprehension deficits in Parkinson's disease. *Brain and Language,* 97, 243–257.

Hornykiewicz, O., Kish, S. J. (1986). Biochemical pathophysiology of Parkinson's disease. *Advances in neurology*, *45*, 19-34. Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). The Boston naming test, 2nd. *Philadelphia: Lea & Febiger*.

Illes, J., Metter, E.J., Hanson, W.R., Iritani, S. (1988). Language production in Parkinson's disease: acoustic and linguistic considerations. *Brain and Language*, 33, 146–160.

Illes, J. (1989). Neurolinguistic features of spontaneous language production dissociate three forms of neurodegenerative disease: Alzheimer's, Huntington's, and Parkinson's. *Brain and Language*, 37, 628–642.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. Frontiers in Psychology 2:255.

Indefrey, P., Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition,* 92, 101–144.

Katzenschlager, R., Lees, A. J. (2002). Treatment of Parkinson's disease: levodopa as the first choice. *Journal of neurology*, 249, ii19-ii24.

Ketteler, S., Ketteler, D., Vohn, R., Kastrau, F., Schulz, J. B., Reetz, K., Huber, W. (2014). The processing of lexical ambiguity in healthy ageing and Parkinson′s disease: role of cortico-subcortical networks. *Brain Research*, 1581, 51–63.

Kolk, H. H. J. (1995). A time-based approach to agrammatic production. *Brain and Language, 50,* 282-303.

Lackner, J. R., Tuller, B. H. (1979). *Role of efference monitoring in the detection of selfproduced speech errors*. In W. E. Cooper & E. C. T. Walker (Eds.), Sentence processing: Studies dedicated to Merrill Garrett (pp. 281–294). Hillsdale, NJ: Earlbaum.

Lee, M., Thompson, C.K. (2004). Agrammatic aphasic production and comprehension of unaccusative verbs in sentence contexts. *Journal of Neurolinguistics*, 17, 315–330.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* MIT Press.

Lieberman, P., Friedman, J., Feldman, L.S. (1990). Syntax comprehension deficits in Parkinson's disease. *The Journal of Nervous and Mental Disease,* 178, 360–365.

Lieberman, P., Kako, E., Friedman, J., Tajchman, G., Feldman, L.S., Jiminez, E.B. (1992). Speech production, syntax comprehension, and cognitive deficits in Parkinson’s disease. *Brain and Language*, 43, 169–189.

Liss, J.M. (1998). Error revision in the spontaneous speech of apraxic speakers. *Brain and Language, 62,* 342- 360.

Logemann. J. A., Fisher. H. B., Boshes. B., Blonsky. E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders*. 43, 47-57

Longworth. C. E., Keenan. S. E., Barker. R. A., Marslen-Wilson. W. D., Tyler. L. K. (2005). The basal ganglia and rule-governed language use: evidence from vascular and degenerative conditions. *Brain*, 128, 584-596.

Louis. E. D., Winfield. L., Fahn. S., Ford. B. (2001). Speech dysfluency exacerbated by levodopa in Parkinson's disease. *Movement Disorders*, 16, 562-565.

Ludlow, C. L., Connor, N. P., Bassich, C. J. (1987). Speech timing in Parkinson's and Huntington's disease. *Brain and language*, 32, 195-214.

Marien, P., Mampaey, E., Vervaet, A., Saerens, J., De Deyn, P. P. (1998). Normative data for the Boston Naming Test in native Dutch-speaking Belgian elderly. *Brain and language*, 65, 447-467.

Marschall, R., Neuburger, S., Phillips, D. (1994). Verbal self-repair and improvement in treated aphasic clients. *Aphasiology, 8,* 535-547.

Marshall, R. C., Rappaport, B. Z. Garcia-Bunuel, L. (1985). Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language,* 24, 297–313.

McNamara, P., Obler, L. K., Au, R., Durso, R., Albert, M. L. (1992). Speech monitoring skills in Alzheimer's disease, Parkinson's disease, and normal aging. *Brain and Language*, 42, 38–51.

McNamara, P., Obler, L. K., Au, R., Durso, R., Albert, M. L. (1992). Speech monitoring skills in Alzheimer's disease, Parkinson's disease, and normal aging. *Brain and Language*, 42, 38–51.

Metter, E. J., Hanson, W. R. (1986). Clinical and acoustical variability in hypokinetic dysarthria. *Journal of Communication Disorders*, 19, 347-366.

Miatton. M., Wolters. M., Lannoo. E., Vingerhoets. G. (2004). Updated and extended Flemish normative data of commonly used neuropsychological tests. *Psychologica Belgica*. 189-216.

Miceli, G., Gainotti, G., Caltagirone, C., Masullo, C. (1980). Some aspects of phonological impairment in aphasia. Brain and language, 11, 159-169.

Miller, D. B., O'Callaghan, J. P. (2014). Biomarkers of Parkinson's Disease (Pd): Present and Future. *Metabolism*, 64, 40–46.

Molloy, S. A., Rowan, E. N., O'brien, J. T., McKeith, I. G., Wesnes, K., & Burn, D. J. (2006). Effect of levodopa on cognitive function in Parkinson's disease with and without dementia and dementia with Lewy bodies. *Journal of Neurology, Neurosurgery & Psychiatry*, 77, 1323-1328.

Monetta, L. and Pell, M.D. (2007). Effects of verbal working memory deficits on metaphor comprehension in patients with Parkinson's disease. *Brain and Language*, 101, 80–89.

Monetta, L., Grindrod, C. M., Pell, M. D. (2009). Irony comprehension and theory of mind deficits in patients with Parkinson's disease. *Cortex*, 45, 972-981.

Monsch. A. U., Bondi. M. W., Salmon. D. P., Butters. N., Thal. L. J., Hansen. L. A., Klauber. M. R. (1995). Clinical validity of the Mattis

Dementia Rating Scale in detecting dementia of the Alzheimer type: a double cross-validation and application to a community-dwelling sample. *Archives of Neurology*. 52, 899-904.

Motley, M. T., Camden, C. T., Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 21, 578-594.

Natsopoulos, D., Grouios, G., Bostantzopoulou, S., Mentenopoulos, G. (1993). Algorithmic and heuristic strategies in comprehension of complement clauses by patients with Parkinson's disease. *Neuropsychologia*, 31, 951–964.

Natsopoulos, D., Katsarou, Z., Bostantzopoulou, S., Grouios, G., Mentenopoulos, G., Logothetis, J. (1991). Strategies in comprehension of relative clauses by parkinsonian patients. *Cortex*, 27, 255–268.

Nickels, L., Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex*, 31, 209–237.

Nozari, N., Dell, G.S., Schwartz, M.F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology,* 63, 1-33.

Nozari, N., Dell, G. S. Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology,* 63, 1–33.

O'Leary, U. M., Rusch, K. M., Guastello, S. J. (1991). Estimating age-stratified WAIS-R IQS from scores on the raven's standard progressive matrices. *Journal of Clinical Psychology*, 47, 277-284.

Olanow. C. W., Obeso. J. A. (2012). The significance of defining preclinical or prodromal Parkinson's disease. *Movement Disorders*, 27, 666-

669.Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburg inventory. *Neuropsychologia,* 9*,* 97-113.

Oomen, C. C. E., Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, 30, 163–184.

Oomen, C. C. E., Postma, A., Kolk, H. H. J. (2005). Speech monitoring in aphasia: Error detection and repair behaviour in a patient with Broca's aphasia. *Phonological encoding and monitoring in normal and pathological speech*, 209.

Oppenheim, G. M., Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition,* 106, 528-537.

Özdemir, R., Roelofs, A., Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition,* 105, 457–465.

Penke, M., Westermann, G. (2006). Broca's area and inflectional morphology: evidence from broca's aphasia and computer modeling. *Cortex*, 42, 563–576.

Penke, M., Janssen, U., and Krause, M. (1999). The representation of inflectional morphology: evidence from Broca's aphasia. *Brain and Language*, 68, 225–232.

Péran, P., Rascol, O., Demonet, J.F., Celsis, P., Nespoulous, J.L., Dubois, B., and Cardebat, D. (2003). Deficit of verb generation in nondemented patients with Parkinson's disease. *Movement Disorders*, 18, 150–156.

Pickering, M.J., Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36, 329-392.

Pickering, M.J., Garrod, S. (2014). Self- Other and joint monitoring using forward models. *Frontiers in Human Neuroscience* 8, 132-143.

Pitcairn, T. K., Clemie, S., Gray, J. M., Pentland, B. (1990). Impressions of parkinsonian patients from their recorded voices. *International Journal of Language & Communication Disorders*, 25, 85-92.

Poluha, P., Teulings, H. L., Brookshire, R. (1998). Handwriting and speech changes across the levodopa cycle in Parkinson’s disease. *Acta psychologica*, 100, 71-84.

Postma A, Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*, 39, 375–392.

Postma, A, Kolk, H.H.J. (1992a). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech and Hearing Research,* 35, 537–544.

Postma, A., Kolk, H. H. J. (1992b). Error Monitoring in People Who StutterEvidence Against Auditory Feedback Defect Theories. *Journal of Speech, Language, and Hearing Research,* 35, 1024-1032.

Praamstra, P., Stegeman, D.F., Cools, A.R., Horstink, M.W. (1998). Reliance on external cues for movement initiation in Parkinson’s disease: evidence from movement-related potentials. *Brain,* 121,167 – 177.

Raven, J. C., Court, J. H., Raven, J. (1998). *Progressive Coloured Matrices*. Oxford: Oxford Psychologists Press.

Roelofs, A. (2005). *Spoken word planning, comprehending, and self-monitoring: Evaluation of WEAVER++*. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), Phonological encoding and monitoring in normal and pathological speech (pp. 42–63). Hove, UK: Psychology Press.

Sanabria. J., Ruiz. P. G., Gutierrez. R., Marquez. F., Escobar. P., Gentil. M., Cenjor. C. (2001). The effect of levodopa on vocal function in Parkinson's disease. *Clinical neuropharmacology*. 24, 99-102.

Schlenck, K. J., Huber, W., Willmes, K. (1987). "Prepairs" and repairs: Different monitoring functions in aphasic language production. *Brain and Language*, 30, 226-244.

Seiss E. Praamstra P. (2004). The basal ganglia and inhibitory mechanisms in response selection: evidence from sub- liminal priming of motor responses in Parkinson's disease. *Brain,* 127, 330 – 339.

Smits, C. H., Smit, J. H., van den Heuvel, N., Jonker, C. (1997). Norms for an abbreviated Raven's Coloured Progressive Matrices in an older sample. *Journal of clinical psychology*, 53, 687-697.

Solomon. N. P., Hixon. T. J. (1993). Speech breathing in Parkinson's disease. *Journal of Speech, Language, and Hearing Research*. 36, 294-310.

Stuss. D. T., Alexander. M. P., Hamer. L., Palumbo. C., Dempster. R., Binns. M. Izukawa. D. (1998). The effects of focal anterior and posterior brain lesions on verbal fluency. Journal of the International Neuropsychological Society. 4, 265-278.

Tombaugh. T. N., Kozak. J., Rees. L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*. 14,167-177.

Ullman, M.T., Corkin, S., Coppola, M., and Hickok, G. (1997). A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience,* 9, 266–276.

Wolfe. V. I., Garvin. J. S., Bacon. M., & Waldrop. W. (1975). Speech changes in Parkinson's disease during treatment with L-dopa. *Journal of communication disorders*. 8, 271-279.

Wylie, S.A., van den Wildenberg, W.P.M., Ridderinkhof, K.R., Bashore, T.R., Powell, V.C., Manning, C.A., Wooten. G. F. (2009). The effect of speed-accuracy strategy on response interference control in Parkinson's disease. *Neuropsychologia,* 47, 1844–1853.

CHAPTER 6
TOWARDS A NEW MONITORING MODEL: CONFLICT MONITORING AND RESOLUTION DURING PRODUCTION AND PERCEPTION

***Manuscript submitted for publication*[1]**

Currently none of the models of verbal monitoring is able to give a full account for verbal monitoring during production and perception. Especially patient data have been very problematic. In the current chapter we propose a model of verbal monitoring. We extend the conflict-monitoring model of Nozari, Dell & Schwartz (2011) from a speech production based model, into a speech production and perception monitoring model. In the new model monitoring during both production and perception occur via conflict monitoring. The production and perception streams are connected via links at the semantic layer, word layer and phonological layer, that allow spreading activations. This allows production and perception to influence each other, without assuming interdependency. Furthermore, we propose a domain general conflict resolution mechanism from the cognitive control literature.

---

[1] This paper was co-authored by Robert Hartsuiker, but it could not have been written without the suggestions and discussions with Elger Abrahamse, Massimo Silvetti, and Tom Verguts. A version of this chapter together with a shorter version of chapter 2 is submitted for publication.

INTRODUCTION

From the review in chapter 2, we concluded that each of the current models has its own specific strengths and weaknesses. However, none the current model discussed in chaper 2 is able to account for speech monitoring during both production and perception, and give an adequate explanation of how the error is resolved. Perhaps the most problematic finding for the theories that try to account for error detection in production and perception, the PLT and forward model theory, are the patient data that we reviewed. Chapter 2 discussed patient data that clearly show that verbal monitoring can be affected differently in production and perception. They further showed that verbal monitoring in semantic processing and phonological processing can be affected independently. Most specifically, patient G. from Oomen et al (2005), who shows a dissociation between semantic and phonologic error detection in production, and a dissociation between semantic and phonological error detection in perception. The PLT nor the Forward Model Theory can explain these, as both assume a high degree of overlap between production and perception. The conflict monitoring account is a promising candidate to explain these findings; by altering the strengths of the connections between the processing levels, Nozari et al. (2011) were able to simulate actual aphasic patient speech production data.

However, the conflict monitoring model by Nozari et al. (2011) is restricted to speech production. Interestingly, the proposed monitoring mechanism, measuring conflict between response options and relaying this to an executive center, is very straightforward. It is not difficult to imagine how this mechanism might be applied to monitoring during perception.

In the current chapter we propose an extension of the conflict monitoring during speech production model by Nozari et al. (2011) into a speech conflict based monitoring model for speech production and perception. The main issue with previous monitoring models that account for verbal monitoring in production and perception, is that they have made the two processes so dependent, that a separate lesioning of one of the two modalities, would make verbal monitoring in both modalities impossible. However, predictive processing during perception, as observed in anticipatory eye movements (Altmann & Kamide 1999; Kamide, Altmann & Haywood, 2003) expectancy effects (Kutas & Federmeier, 2011) and perceptive processes during production, such as the integration of perceived sounds into our production (Delvaux & Soquet, 2007; Pickering & Garrod, 2004) indeed suggest that the two modalities interact. Therefore, we propose to connect verbal monitoring during speech production and perception in a manner that allows for an interaction between production and perception, but without making the two processes dependent on each other.

However, if we extend the conflict-monitoring model into perception, the model still does not provide a full account of verbal monitoring; the model addresses how an error is detected, but not how the error is resolved. An interesting solution as to how an error is resolved comes from the cognitive control literature. Verguts and Notebaert (2008, 2009) propose that the detection of conflict can lead to adaptation via an aspecific boost of activation. This aspecific boost could be implemented at a neural level by a noradrenergic (NA) response. The release of this neurotransmitter into the cortex boosts all the active units. If such a boost comes into the speech production system, the items under competition will see an exponential increase in their activation levels, which allows the correct item to increase sufficient in activation to be selected.

Currently there is no model of verbal monitoring that is able to adequately account for the full process of verbal monitoring, especially if patient data are taken into account. If the conflict monitoring account for speech production, as proposed by Nozari, Dell & Schwartz (2011), can be successfully extended to speech perception with addition of a conflict resolution mechanism, this would be, in my opinion, an important step forward. Furthermore, as it is a computational model, an extension of the model lends itself for direct testing.

A CONFLICT MONITORING ACCOUNT FOR PRODUCTION AND PERCEPTION

We propose an extension of the conflict-monitoring theory of Nozari et al., (2011) into perception, with a domain general conflict resolution system. Below we first discuss the architecture of the model speech production, speech perception, and how these two are linked. Second, we discuss the parallels with forward models.

ARCHITECTURE OF THE MODEL

For production we assume the same mechanism as Nozari et al., (2011); we propose an interactive feedback model in which the semantic features in a semantic layer are connected to lemmas in a word layer. The lemmas are connected to phonemes in a phoneme layer. The semantic weight is the strength of the connection between the semantic and the word layer. The phonological weight is the strength of the connection between the word and the phoneme layer. The value of these weights determines how strongly the information is transferred between those layers, and thus the strength of the connection. In order to simulate the patient data Nozari et al., (2011) decreased the value of these semantic or phonological weights. Speech production happens in two steps. First the semantic features of the target become active. The activation spreads trough the network, activating the target lemma, for instance ‘tea’, but also activating the competitors at the word layer, such as ‘coffee’. Via cascading, the activation is further spread down to the phoneme layer. As the model is interactive, the nodes in the lower layer send activation back to the higher layers (feedback). The activation of each node is the sum of activations the node receives from connecting nodes, and this activation is subject to decay and random noise. After a certain amount of time, the highest activated node becomes selected. In the second step a boost of activation is sent from the selected node (Nozari et al., 2011; Foygel & Dell, 2000) to the phoneme layer below. After a certain amount of time the most active node at each phoneme cluster is selected for the final response (e.g. onset [t], vowel [i:], coda ø). The amount of conflict is predictive of the occurrence of an error, as demonstrated by simulations. And conflict at each layer specifically predicts what error arises from that layer; conflict at the word layer can lead to a semantic, but not to a phonological error. And conflict at the phoneme level can lead to a non-word, but not a semantic error. This layer specificity was also simulated successfully.

We propose to extend the model with a perception system parallel to the production system, with similar representations in the systems and a tight link between the representations (see Figure 5). The assumption that speech production and perception have two distinct systems receives strong support from the brain damaged patient data, as discussed in chapter 2, which show dissociations between verbal monitoring in production and perception. The assumption of distinct lemma representations for production and perception is further supported by research into language acquisition (Gupta & MacWhinney, 1997), semantic errors in language production (Caramazza, 1997), and tip-of-the-tongue states (Miozzo & Caramazza, 1997). Additionally we assume distinct representations at the semantic and at the phoneme level. The tight link between the production and the perception system are motivated by the finding of cross-modal priming. For instance properties of perceived sound can be integrated in our speech production (Delvaux & Soquet, 2007; Pickering & Garrod 2004) and (especially semantically related) speech we are trying to ignore can be confused with attended speech during perception (Brungart 2001; Gray & Wedderburn

1960). Evidence of production processes during perception are observed in anticipatory eye movements (Altmann & Kamide 1999; Kamide, Altmann & Haywood, 2003) and expectancy effects (Kutas & Federmeier, 2011).

The process of perception starts at the phoneme level. Upon hearing speech the relevant phonemes become active, which in turn send activation to the word layer via spreading, which in turn sends activations to the semantic layer and so on. The word layer also sends activation back to the phoneme layer.

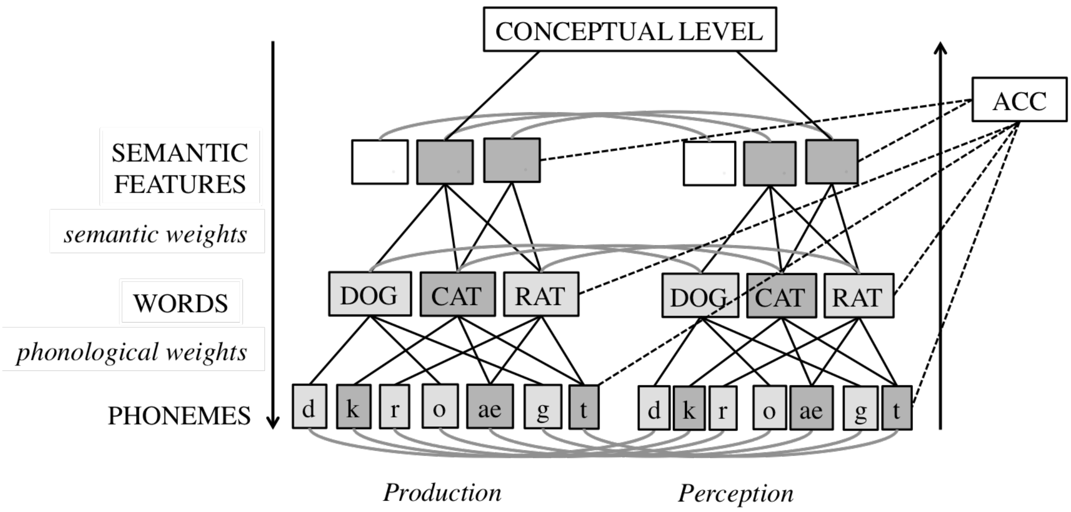
The production and perception representations are tightly linked, and activations flow automatically via spreading between the production node and the perception node of a representation. If a node in production becomes active, the same node is activated in the perception system, and vice versa. When a word is produced, the representations of the production system are consequently active in the perception system. For instance if at the semantic production level the semantic features of the target word 'tea' are active, the same nodes in the parallel semantic perception layer increase somewhat in activation. In the production system the activation spreads, and at the production word level the word 'tea' becomes active, as well as the semantic competitor 'coffee'. Via the interconnections between the two processing streams, and via spreading of activation in the perception level, the words become parallel activated in the perception word level. In the production system the target word 'tea' is selected, and a jolt of activation is sent to the production phoneme layer. Via the interconnections, and spreading from the perception word level, the phonemes in the perception layer also become active. Now the word 'tea' is produced, and the perception system is already fully prepared to perceive this word. These activations act as a prediction of the upcoming percept, and can thereby perform external self-monitoring. In the perception system the phonemes [t] and [i:] increase more in activation, and via cascading the word increases in activation as well. After some time the most highly active word is selected, and received a jolt of activation.

This increases the activations at the semantic level, leading to comprehension at the conceptual level.

If an incorrect phoneme were to be selected at the phoneme production level, and an incorrect word would be produced, the perception system would benefit from the spreading activations to recognize the incorrect word. For instance the incorrect word ‘cap’ is produced, instead of the target word

‘cat’. In the perception system the word ‘cat’ would allready be active at both the semantic and the word layer. At the phoneme layer the phonemes [k] [æ] [p] [t] are active. Now at the word level competition emerges between the word /cat/ and /cap/. Only /cat/ receives increased activation from the production and from the semantic perception level, and therefore becomes the most highly activated node that becomes selected for comprehension.

Figure 5. *Hierarchical Conflict Model for Self- and Other Monitoring. Speech production and perception have separate semantic features, words and phonemes, which are tightly connected via links. Arrows indicate the direction of processing.*



When listening to someone else speaking, a top-down process driven by speech production and a bottom up process driven by the incoming speech are started. During listening, the perceived words become active first in the perception system, and via the tightly interconnections the representations in the production system also become active. On the basis of the incoming speech, nodes become active in the perception system, and consequently in the production system. Based on past experience, the production system activates related nodes, which in turn become active in the perception system, thereby creating a prediction of the upcoming percept. For instance perceiving the utterance 'I just ate a' would lead to the activation of items edible in the semantic system of the production level. Additional experience, for instance time of the day, would contribute to an increase in activation of specific items, such as 'sandwich'. The higher the close probability of the word, the higher the activation the item. In this case, a semantically related word like 'salad' will also have high activations. The activation of the items in the production system leads to parallel activations in the perception system, thereby preparing the perception system for this item/ these items, and thus creating expectancies. When the predicted utterance is met, the active nodes in the perception system increase in activation, until after an amount of time the word is selected for comprehension. When perceived words match the predicted percept, speech perception thus becomes a low-effort process, as the perception of that word is perpaired.

Predictions also include para-linguistic aspects, such as speed, intonation and voice pitch. As these predictions are more accurate for self produced speech than for speech produced by others, there is more conflict in perceiving someone else's speech compared to self-perception, consistent with neuroimaging data on feedback processing. One prediction of this model is that more conflict is observed in response to an unfamiliar voice compared to listening to a highly familiar voice, for instance a family member.

The model as we propose here has two separate processing streams for production and perception, which can function separately and consequently be lesioned separately. The way conflict is operationalized here, as the increase of activation over multiple selected nodes, means that error detection is a byproduct of language production and perception. Subsequent detection of the conflict is done by a domain general mechanism. This process functions in exactly the same way during production and perception.

In the current model errors can be detected via three mechanisms. The first is via conflict in response selection between highly active nodes in the production layer. The second mechanism is via conflict between highly active nodes in the perception layer. This type of error detection is used in the perception of someone else's speech. A third method of error detection is via conflict between the perception of external speech, and the activations in the perception system that result from spreading activations in the production system, such as when listening to your own speech or when listening to speech in a predictable context.

A PARALLEL WITH FORWARD MODELS

In the current model we consider cascading and feedback from lower levels as a form of forward modeling, as also suggested by Dell (2013). In forward model theories, the forward model is an impoverished version of the production command. In a hierarchical feedback model the activation cascading down also increases the activation of connected nodes in the next layer, but to a much lesser extent than if the nodes were committed parts of a representation. The goal of the forward model is to ease the selection process, thereby speeding it up. The cascading of activation fulfills the same function. By already increasing activation of the connected nodes, the construction of representation at those levels is prepared. And finally, the feedback from the lower layers to the higher layers in a cascading model allows information of the lower layers to become available to the higher layers, so these different layers can be coordinated, much like the forward model does for the production command.

CONFLICT RESOLUTION

In the current conflict-monitoring model, the story ends at the moment conflict is detected. There is no complete account of how the detection of the error is handled. How does the monitor decide which errors to handle? Does an error signal lead to an interruption? Is the interruption followed by a restart? Or is correct selection of a target sufficient?

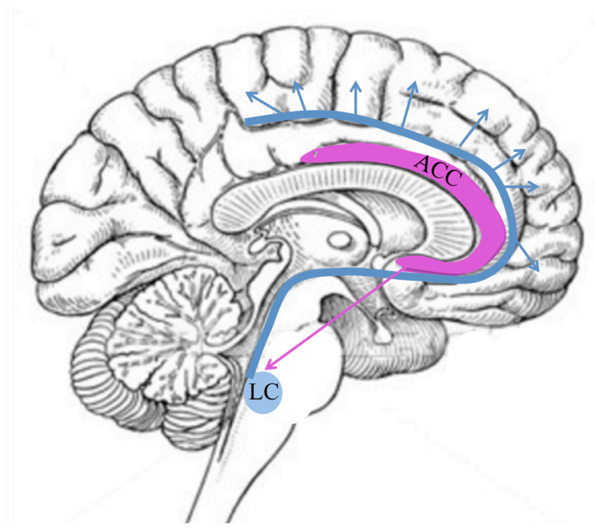
The question of the aftermath of error detection has been studied, but these studies do not answer all questions stated above. Studies investigating the aftermath of error production have often focused on the temporal coordination between interruption and repair and to what extent strategic components are used, such as postponing the interruption until the repair is planned (Hartsuiker, Catchpole, De Jong, & Pickering, 2008; Seyfeddinipur, Kita, & Indefrey, 2008; Tydgat, Stevens, Hartsuiker, & Pickering, 2011, Gambi, Cop, & Pickering, 2014). Another question concerns the mechanisms of repairing; do you start with a clean slate once you've interrupted an incorrect utterance? These studies showed that planning of a new word is affected by residual activation of representations pertaining to the abandoned word. Specifically, the experiments showed semantic facilitation and phonological interference effects of abandoned words on repair words (Hartsuiker, Pickering, & De Jong, 2005; Tydgat, Diependaele, Hartsuiker, & Pickering, 2012). A study where participants had to quickly adapt their utterance to make it appropriate for a new context, suggested that utterances can sometimes be repaired by revising the speech plan, rather than plan from scratch (Boland, Hartsuiker, Pickering, & Postma, 2005).

At this moment, however, it is unclear what happens between the detection of an error and the production of the repair. „Existing theories portray the relevant mechanisms as coming into play when the participation is required, but without an account of how the need for intervention is detected or how the intervention itself is triggered. Without a good theory, control remains a sort of homunculus that 'just knows' when to intercede. For any theory on cognitive control to be complete, it will need to offer an account of how the system determines when control is required.'' (Botvinick et al., 2001, p. 624). Although the conlfict monitoring account does provide an answer to the question 'when' intervention is needed (when conflict surpasses a certain threshold), it does not explain how the intervention is performed.

A possible answer to the question of how an intervention is performed might be found in the cognitive control literature. Verguts and Notebaert (2008, 2009) suggest that the detection of conflict can lead to adaptation via an aspecific boost of activation. A probable candidate for this functional boost is noradrenaline, delivered via the LC. Anatomically the ACC is connected to brainstem neuromodulatory centers, including the locus coeruleus (LC). Previous research has shown that stimulation of the ACC leads to activation changes in the LC (e.g., Jodoj, Chiang, & Aston-Jones, 1998), which plays an important role in attention, response selection and task engagement (Aston- Jones, Chiang, & Alexinsky, 1991; Aston-Jones and Cohen 2005). However, another probable candidate for this role, which cannot be excluded at this time, is dopamine.

Below we integrate the noradrenergic response for conflict resolution for error detection in both production and perception in the conflict monitoring model. Note, however, the activation boost can in general also be applied to the models above, in the following way: on an erroneous trial an error is signaled, by any of the mechanisms described in the models above. The error signal is picked up by the ACC, which sends a signal to the locus coeruleus (LC), which triggers the aspecific boost of activation, thereby increasing activation of all the active neurons (see Figure 6). The top-down boost of activation that is sent into the language processing system causes all active items to increase in activation exponentially, leading to an improved signal to noise ratio. This boost serves two functions: strengthen active connections, which are task relevant. And improve signal-to-noise ratio, leading to a faster selection of the correct item.

Figure 6 *Schematic overview of the LC, ACC and the projections of noradrenaline into the cerebral cortex.*



CONFLICT RESOLUTION AFTER ERROR DETECTION

Once the ACC picks up on the conflict, some time is needed to signal to the LC, and for the NA response to reach the process. During this time competition between the competing items will have continued. The incremental boost increases the activations exponentially from the top down, so that the intended target was only marginally higher activated than its competitor, it will now be much higher activated. The correct node is now easily selected. And the connections between the active nodes are strengthened, decreasing the error chance for a next time.

STAGES OF CONFLICT RESOLUTION DURING PRODUCTION

If during the selection process multiple nodes compete for selection, there are two stages at which the conflict can be resolved. The conflict is resolved either before selection of the target, leading to steering, or after selection of the target, leading to error correction. A third function of the model is learning; via the NA boost the connection is strengthened.

Steering

We assume that if during the selection stage two nodes become highly active, the conflict is detected by the ACC and a noradrenaline boost is released. Between the occurrence of the conflict, and the resolution via the boost, some time passes. If conflict is resolved before target selection, the conflict is resolved within the level, and resolution is part of the selection process. This form of correction, where the error is resolved before it is produced, is also called prepairs (Schlenck et al., 1987; Kolk, 1995) or covert repairs (Levelt, 1983, 1989; Postma, 2000).

Correction

If during the selection process a high conflict arises, the highest active node might become selected before the boost response reaches the conflict. As stated above, some time passes between the detection of the conflict and the moment the boost reaches the conflict site. Once the boost reaches the conflict, the boost will increase activation of the items exponentially, and again an item is selected. Both the correct and the incorrect target might reach the threshold for selection before conflict is resolved. If a correct target is selected first, it will be selected again the second time as it receives activation from the conceptual level and it receives activation via feedback from the lower levels. If the incorrect target node was selected first, the second time the competition needs to be resolved, the correct target will now be selected as it receives information from the conceptual level the incorrect target node does not. One assumption here is that the feedforward activation is stronger than the feedback (see also Dell, 1980; 1986).

If indeed the boost of activation leads to re-selection of a target, this means that now two items are being processed for production in the same slot. The final result of this situation is dependent on the amount of time that passes between initial target selection (possible error) and secondary target selection (possible repair). If the second target is selected fast enough after the first target, the high amount of activation resulting from the boost will as a result increase activations the connected nodes in the next processing layer. Essentially 'overruling' the activation that came with the first item that was selected. However, if there is a bigger lag between selection of the first and the second target, the first target will have advanced in processing quite a bit, leading to production of (part of) the incorrect utterance, requiring a repair. Production of the incorrect utterance might be halted as the result of competing activations at the phoneme level. Once the conflict is resolved, production continues with the production of the correct item. Note that the same situation can arise in the case of correct target of selection from a highly competitive environment; the correct target will be selected twice. This results in a restart of the same utterance.

STAGES OF CONFLICT RESOLUTION DURING PERCEPTION

As in production, when multiple nodes become highly active in perception, a signal is sent to the ACC, which triggers a noradenalinergic response, which increases the signal to noise ratio, thereby resolving the conflict.

Unconnected speech

In unconnected speech, so in the perception of a single word in for instance an experimental setting, conflict would only arise if the perceived word contains enough correct phonemes to activate a word at the semantic layer via the reciprocal connections. Upon perceiving ‘cactut’, ‘cactu’ leads to activation of the lexical item of ‘cactus’, which leads to competition between the expected ‘s’ and the perceived ‘t’. Upon perceiving ‘cap’, where ‘cat’ is intended, no error is detected, as ‘cap’ is a valid entry that does not lead to competition.

Connected speech

In connected speech, or speech within a context, the production system comes into play. When the predicted utterance is wrong, the incoming speech will activate different nodes than those selected by the production system. The subsequent competition is resolved by the ACC, LC system. For instance, you hear the utterance ‘Last night before I went to bed, I was brushing ...’. While listening, you predict that the upcoming words will probably be ‘my teeth’ or ‘my hair’. In fact, the sentence ends with ‘my teef’. In this case the nodes belonging to ‘my’ are highly active, the node is selected, and the activation decays. ‘teef’ activates the phonemes ‘t’ ‘e’ and ‘f’. As the production system has already activated the semantics and phonology of ‘teeth’, the phonemes ‘t’ and ‘e’ will increase the activation of the semantic node. Competition between the ‘th’ and ‘f’ are resolved by the ACC, LC system, selecting the already highly active node ‘teeth’. Alternatively, the listener might know that the speaker is a Dutch person, and might have predicted that the upcoming utterance ‘my teeth’ will be produced as ‘my teef’. As a listener becomes familiar with a specific speaker, a speaker-specific representation is built, analogous to a mini-grammar (Warker & Dell 2006). This mini-grammar is subsequently used to make speaker-specific predictions.

ADVANCES AND PREDICTIONS OF THE MODEL

In sum, we propose a hierarchical feedback model with conflict resolution by a domain general monitor. Conflict arises as multiple nodes increase in activation, thereby competing for selection. This conflict is picked up by the domain general conflict monitor in the ACC, and consequently resolved by noradrenaline response which boosts all active nodes. As a result of the boost, the signal-to-noise ratio of competing items increases and activations are strengthened.

This model is able to explain how patients with faulty self-monitoring exhibit intact other monitoring, unlike the perceptual loop theory or forward models theory. It also explains how intact self-monitoring can occur without using the external route for self-monitoring, as suggested by the patient data discussed in chapter 2. If the connections between the production and perception representations are lesioned, the person can still exhibit intact self-monitoring and intact other-monitoring, but is impaired in predicting upcoming speech. This last theory might be interesting to investigate in the light of schizophrenic patients.

Predictions of the model

This model makes several testable predictions, which we have sketched below. As the current account makes a direct link between NA release and conflict resolution, a clear prediction is that conflict leads to an NA release, with the amount of NA release related to the amount of conflict. Furthermore, as the NA release leads to a strengthening of the connections, we also have predictions with respect to the aftermath of conflict resolution.

Pupil dilation.

A tight link exists between the neurotransmitter noradrenaline and pupil dilation (Rajkowski et al., 1994; Philips et al., 2000; Samuels & Szabadi 2008; Sterpenich et al., 2006). Under constant illumination the norepinephrine levels are reflected in the dilation of the pupil. If indeed NA is responsible for resolving conflict in linguistic processing, than the amount of conflict should be reflected in the dilation of the pupil. In both visual and auditory ambiguity resolution an increase in pupil diameter is measured just before a perceptual switch was reported (Einhäuser, Stout, Koch & Carter, 2008). The magnitude of the observed dilation was indicative of the subsequent duration of perceptual stability. Pupil dilation has also been measured during discourse processing, and in this study it was found that correct prosodic cues were indeed related to the smallest dilations compared to uninformative prosodic cues (Zellin, 2011). Similarly Zekveld et al., (2014) found pupil responses to different degrees of audibility of speech. When listening to someone speak, largest dilations were observed when the participant heard someone else speaking simultaneously. Smaller dilations were observed when the speech was masked by random noise. And the smallest dilations were observed for noise-vocoded speech. Pupil dilation has also been reported to be a reflection of word retrieval effort in bilinguals (Schmidke, 2014); low word frequency and high neighborhood density were related to high pupil dilation, as predicted on the basis of our model. High proficient bilinguals showed, in comparison to low proficient bilinguals, an earlier pupil response and a smaller effect of neighborhood density and frequency. These findings are very much in line with our model, in which NA response is released for conflict resolution.

Drugs.

Alpha-blockers inhibit the firing of cells in the LC, thereby reducing the release of norepinephrine. Alpha-blockers are used for the treatment of anxiety, panic disorders an PTSD. A direct effect of Alpha-blockers on self-monitoring is expected, as conflict resolution will be heavily impaired. The production of errors will increase and fewer corrections will be made.

Neurological Disorders.

Two patient groups typically associated with abnormal noradrenaline functioning are schizophrenic patients and patients with Alzheimer's disease. In schizophrenic patients increased NA levels are measured in the cerebro spinal fluid (CSF) compared to age matched controls (Kemali et al., 1982; Lake et al,. 1980). Treatment of these patients with clondine or guanfacine (α_2 adrenergic agonist) causing reduced functioning of NA receptors, improved cognitive functioning as measured by learning, delayed recall and the Trail B task (Fields et al., 1988; Friedman et al., 1999). Deficits in self-monitoring have been hypothesized to be the cause of the auditory illusions in some schizophrenic patents; what exactly the effects are of abnormally high NA levels remains to be investigated (e.g. Wilkinson, 2014; Hommes et al., 2012).

Alzheimer's disease is associated with a loss of up to 70% of norepinephrine projecting cells in the LC (Heneka et al., 2010). However, Alzheimer's disease also leads to loss of neurons and synapses in the cortex and sub-cortical regions, including the frontal lobe, parietal lobe, temporal lobe and cingulate gyrus (Wenk, 2003), thereby making this disease a less ideal candidate to investigate the role of NA in language processing.

Aftermath of conflict resolution

During conflict resolution the aspecific boost of activation increases the strength between the connections according to Verguts and Notebaert (2008;2009). This is a form of learning. As the boost of activation is in function of the amount of conflict, learning is also in function of the amount of conflict. So once conflict is resolved, a subsequent encounter of the situation should not lead to as much conflict. If, for instance, conflict arises between 'tea' and 'coffee' at a lexical level, the boost of activation will strengthen the connections between the semantic representation of 'tea', and the lexical level representation of 'tea'.

As the aspecific boost of NA strengthens the connections between the active nodes, conflict between nodes in the trial decreases the conflict of the active items the next time they need to become active (independent of whether the conflict was at the phonological or sematic level). So conflict resolution between two semantic items ('dog' and 'cat') or conflict between phonological items ('bed' and 'bad') should facilitate the production of these items in following trials. Certainly the production of the correct items should be facilitated, and possibly the production of the competing item (compared to unrelated items) as well as boosting is aspecific.

If on trial n 'chest' and 'brest' compete, the production of 'torso' in trial n+1 is subject to interference as the semantically related item 'chest' was strengthened in the previous trial. This latest prediction is confirmed by the finding of cumulative semantic interference e.g., Brown, 1981; Costa, Strijkers, Martin, & Thierry, 2009; Howard et al., 2006; Navarrete et al., 2010; Oppenheim et al., 2010; Runnqvist, Strijkers, Alario, & Costa, 2012): naming latencies for items of the same semantic category increase at each consecutive trial.

The hierarchy of the model predicts that during production, a conflict at the semantic level with a late conflict resolution will also lead to a conflict at the phonemic level. Thus, shortly before the misselection of 'dog' for 'cat', both the lexical items 'dog' and 'cat' will both cascade activation to phonological coding systems. This leads to high conflict at the phonological level. A conflict at the phonemic level will, however, not lead to conflict at the semantic level. The feedback activation, here from the phonological level to the semantic level, is not strong enough to create competition at a higher level. As a result competition at the phonemic level does not boost activation of competing semantic items, but competition at the semantic level does boost activation at the phonological level of phonemic items. So competition between 'cat' and 'dog' should facilitate the production of 'log' in the subsequent trial, but competition between 'dog' and 'log' should not facilitate the production of 'cat' in a subsequent trial.

GENERAL DISCUSSION

This chapter describes a new verbal monitoring account via a domain general conflict detection mechanism, in combination with a domain general conflict resolution mechanism. This model proposes that verbal monitoring takes place during production and perception with domain general mechanisms.

The model we have described above has several important advantages compared to the existing models discussed in chapter 2. First of all, our model described not only how errors are detected, but also what mechanisms come in to play to resolve the conflict. We propose that the repair of an error is mediated by conflict detection by the domain general ACC and a subsequent noradrenaline boost from the locus coeruleus. This NA boost exponentially increases activation of all active items, thereby increasing the signal-to-noise ratio, allowing for a fast (re)selection of the correct item. Additionally the NA boost strengthens the connections between the active items.

A second strength of the model above is that we specify for both speech production and speech perception how error detection and conflict resolution take place. In both modalities, conflict arises as two or more items compete for selection. This conflict is in both cases resolved by a domain-general conflict monitor as described above.

A third strength of the model is that it proposes a tight link between production and perception, without reduplicating the production process. By assuming a cascade of activation through tight links between the nodes in the production system and the perception system, the production system can be involved in predicting upcoming utterances. And the perception system can use the cascading information from the production system for verbal self-monitoring via the external loop. It also gives a natural explanation of how perceived sounds can be integrated in our production.

Importantly, while the model assumes an interplay between production and perception, it also allows for a separate lesioning of the two processes. As a result, this model might be able to account for the patient data discussed above. Although production and perception are connected via tight links, neither module (intact production or perception) is a prerequisite for monitoring in the other module. Monitoring during production and perception can take place independently, and can be lesioned separately. Whether this assumption holds, remains to be tested via computational simulations.

As the model is a direct adaptation of the conflict model of Nozari, Dell, and Schwartz (2011), semantics and phonology can be lesioned separately. The strength of the links between the different layers can be weak or strong, independent of the strength of connections to other layers. Computational simulations by Nozari et al. (2011) has proven the merits of the model for the production domain. Whether the strength of the connections can explain perceptual observations of patient data remains to be tested.

In sum, we proposed a conflict-monitoring model for error detection in production and perception. We acknowledge the role of the production system in the perception of language via reciprocal connections through which activation cascades. As a result perceptual representations become active during production, allowing for external self-monitoring. And during perception, activations become active in the production system, whereby prediction of the upcoming percept can be made. Further research is needed to test the claims of the proposed model for error detection and resolution.

REFERENCES

Altmann, G. T. M., Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition,* 73, 247–264.

Aston-Jones, G., Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403-450.

Boland, H. T., Hartsuiker, R. J., Pickering, M. J., Postma, A. (2005). Repairing inappropriately specified utterances: Revision or restart? *Psychonomic bulletin & review*, 12, 472-477.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review* 108, 624–652.

Brown, A. S. (1981). Inhibition in cued retrieval. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 204–215.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America,* 109, 1101–1109.

Caramazza, A. (1997). How many levels of processing are there in lexical access?. *Cognitive neuropsychology*, 14, 177-208.

Caramazza, A., Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the "tip-of- the-tongue" phenomenon. *Cognition*, 64, 309–343. Comrie,

Costa, A., Strijkers, K., Martin, C., Thierry, G. (2009). The time-course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences*, 106, 21442–21446.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5, 313–349.

Dell, G. S. (2013). Cascading and feedback in interactive models of production: A reflection of forward modeling? *Behavioral and Brain Sciences*, 36, 351-352.

Delvaux, V., Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica- Basel*, 64, 145.

Einhäuser, W., Stout, J., Koch, C., Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences*, 105, 1704-1709.

Fields, R. B., Van Kammen, D. P., Peters, J. L., Rosen, J., Van Kammen, W. B., Nugent, A., ... & Linnoila, M. (1988). Clonidine improves memory function in schizophrenia independently from change in psychosis: preliminary findings. *Schizophrenia research*, 1, 417-423.

Friedman, J. I., Adler, D. N., Davis, K. L. (1999). The role of norepinephrine in the pathophysiology of cognitive disorders: potential applications to the treatment of cognitive dysfunction in schizophrenia and Alzheimer's disease. *Biological psychiatry*, 46, 1243-1252.

Foygel, D., Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, 43, 182-216.

Gambi, C., Cop, U., Pickering, M. J. (2014). How do speakers coordinate? Evidence for prediction in a joint word-replacement task. *Cortex*.

Gray, J. A., Wedderburn, A. A. I. (1960). Shorter articles and notes grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology,* 12, 180-184.

Gupta, P., MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and language*, 59, 267-333.

Hartsuiker, R. J., Catchpole, C. M., de Jong, N. H., Pickering, M. J. (2008). Concurrent processing of words and their replacements during speech. *Cognition,* 108, 601–607.

Hartsuiker, R. J., Pickering, M. J., de Jong, N. H. (2005). Semantic and phonological context effects in speech error repair. Journal of Experimental Psychology. *Learning, Memory, and Cognition*, 31, 921–932.

Heneka, M.T., Nadrigny, F., Regen, T., Martinez-Hernandez, A., Dumitrescu-Ozimek, L., Terwel, D., Jardanhazi-Kurutz, D., Walter, J., Kirchhoff, F., Hanisch, U.K., Kummer, M.P. (2010). Locus ceruleus controls Alzheimer's disease pathology by modulating microglial functions through norepinephrine. *Proceedings of the National Academy of Sciences of the United States of America,* 107, 6058–6063.

Hommes, J., Krabbendam, L., Versmissen, D., Kircher, T., van Os, J., van Winkel, R. (2012). Self-monitoring as a familial vulnerability marker for psychosis: an analysis of patients, unaffected siblings and healthy controls. *Psychological medicine*, 42, 235-245.

Howard D, Nickels L, Coltheart M, Cole-Virtue J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, 100, 464–482.

Jodoj, E., Chiang, C., Aston-Jones, G. (1998). Potent excitatory influence of prefrontal cortex activity on noradrenergic locus coeruleus neurons. *Neuroscience,* 83, 63-79.

Kamide, Y., Altmann, G., Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language,* 49, 133-156.

Kemali, D., del Vecchio, M., Maj, M. (1982). Increased noradrenaline levels in CSF and plasma of schizophrenic patients. *Biological psychiatry*.

Kolk, H.H.J. (1995). A time-based approach to agrammatic production. *Brain and Language,* 50*,* 282-303.

Kutas, M., Federmeier, K. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 621.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition,* 14, 41-104.

Levelt, W.J.M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Navarrete, E., Mahon, B. Z., Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica*, 134, 279–289.

Oomen, C. C. E., Postma, A., Kolk, H. H. J. (2005). *Speech monitoring in aphasia: Error detection and repair behaviour in a patient with Broca's aphasia*. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), Phonological encoding and monitoring in normal and pathological speech. Hove: Psychology Press.

Oppenheim G.M., Dell G.S., Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition,* 114, 227–252.

Phillips M. A., Szabadi E., Bradshaw C. M. (2000). Comparison of the effects of clonidine and yohimbine on spontaneous pupillary fluctuations in healthy human volunteers. *Psychopharmacology (Berl.) 150, 85–89*

Pickering, M. J., Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169-226.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition,* 77, 97–132.

Rajkowski, J., Kubiak, P., Aston-Jones, G. (1994). Locus coeruleus activity in monkey: phasic and tonic changes are associated with altered vigilance. *Brain Research Bulletin*, 35, 607–616

Runnqvist, E., Strijkers, K., Alario, F. X., Costa, A. (2012). Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language*, 66, 850–869.

Samuels E. R., Szabadi E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Current Neuropharmacology*, 6, 254–285

Schlenck, K.J., Huber, W., Willmes, K. (1987). "Prepairs" and repairs: Different monitoring functions in aphasic language production. *Brain and Language,* 30*,* 226-244.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology.* 5, 137.

Seyfeddinipur, M., Kita, S., Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. Cognition, 108, 837–842.

Sterpenich V., D'Argembeau A., Desseilles M., Balteau E., Albouy G., Vandewalle G., Degueldre C., Luxen A., Collette F., Maquet P. (2006). The locus ceruleus is involved in the successful retrieval of emotional memories in humans. *Journal of Neuroscience*, 26, 7416–7423

Tydgat, I., Diependaele, K., Hartsuiker, R. J., Pickering, M. J. (2012). How lingering representations of abandoned context words affect speech production. *Acta Psychologica*, 140, 218–229.

Tydgat, I., Stevens, M., Hartsuiker, R. J., Pickering, M. J. (2011). Deciding where to stop speaking. *Journal of Memory and Language*, 64, 359–380.

Verguts, T., Notebaert, W. (2008). Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. *Psychological Review*, 115, 518-525.

Verguts, T., Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. *Trends in Cognitive Sciences*, 13, 252-257.

Warker, J. A., Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 387.

Wenk, G.L. (2003) Neuropathologic Changes in Alzheimer's Disease. *The Journal of Clinical Psychiatry*. 64 Supplement 9, 7–10.

Wicker, B., Keysers, C., Plailly, J,. Royet, J.P., Gallese, V., Rizzolatti, G. (2003). Both of us disgusted in My Insula: the common neural basis of seeing and feeling disgust. *Neuron*, 40, 655-64.

Wilkinson, S. (2014). Accounting for the phenomenology and varieties of auditory verbal hallucination within a predictive processing framework. *Consciousness and cognition*, 30, 142-155.

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, 101, 76-86.

Zellin, M., Pannekamp, A., Toepel, U., van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology*, 81, 133-141.

CHAPTER 7

GENERAL DISCUSSION

The aim of the studies presented in this doctoral dissertation was to further investigate how verbal monitoring takes place. More specifically, we investigated the role of perception for internal verbal monitoring. In this final chapter the main findings of this thesis are summarized and discussed. This chapter concludes with directions for future research on verbal monitoring.

RESEARCH OVERVIEW AND THEORETICAL IMPLICATIONS

The research in this thesis dealt with the question ‘how do we detect our speech errors?’. We know that there are two possible stages at which verbal monitoring can occur; pre-articulatory and post-articulatory. Of special interest in the current thesis is pre-articulatory monitoring, as it is at this stage that we can prevent errors from occurring. Current theories of internal verbal monitoring can be roughly divided into two classes: one class of theories that assume error monitoring to take place in the perception system, via the perception of a copy of the speech plan (Levelt, 1983, 1989; Indefrey & Levelt 2004; Indefrey 20110). The second class of theories assume that error monitoring takes place during the stages of speech production (Nozari, Dell, & Schwartz, 2011; Pickering & Garrod, 2013, 2014; Hickok, 2012). Based on this the distinction between the two classes, we can investigate if people detect their errors via the perception system, as proposed by perception based monitoring theories, or if the detection of speech errors occurs independently of the speech perception system, as proposed by production-based monitoring theories.

In Chapter 2 we provide and extensive overview of the current models for verbal monitoring. Each of the models, as well as evidence supporting or opposing the theories is discussed. In addition to a number of specific problems that are specific to each of these models we noted a number of common problems for all the theories. These are summarized below.

A first problem for all monitoring theories face is that they are underspecified. In both the perceptual loop theory and the forward model theory, error detection in based on a comparison. In the Perceptual Loop Theory (PLT) it is unclear what kind of representation of intended speech can be compared with what kind of perceived speech in the conceptualizer. In the forward model theory the outcome of a production stage is compared with a forward model that was created for that processing stage. What exactly the form is of this forward model is underspecified, and faces a theoretical problem. For the forward model to be fast, it needs to be underspecified, but for a meaningful comparison that allows the detection of an error, it needs to be quite specific. The hierarchical state feedback control model (HSFC) (Hickok, 2012) and the conflict monitoring model (Nozari, Dell, & Schwartz, 2011) both give a nice description of how an error signal can arise, but do not describe how this error signal leads to the subsequent repair of an utterance.

A second problem is that none of the theories is able to provide an adequate account of how error monitoring takes place during production and perception. This is the case for some theories because they only describe part of the speech process; the HSFC only deals with phonemes, articulation, and auditory perception, and the conflict monitoring model only deals with speech production. In contrast, the forward model theory and the perceptual loop theory give a full account of how speech errors are detected during production and perception of speech; however, we argue at length in Chapter 2 that neither of these theories accounts for verbal monitoring data gathered from patient studies.

From the review in Chapter 2 we concluded that each of the models discussed has their own strengths and weaknesses, and that none of the current models of verbal monitoring provides an adequate account of how errors are detected during the production and perception of speech, and how the error is consequently resolved.

In Chapter 3 we further investigated evidence for perception-based monitoring. Previous research has found that listening to one's own speech drives eye-movements to phonological competitors in a display, similar to when one hears someone else produce the speech (Huettig & Hartsuiker 2010; Huettig & McQueen, 2007). This parallel between production and perception is prediced by the perceptual loop theory. However, the PLT would predict that the internal loop would allow monitoring of one's own speech via perception to occur before the perception of someone else's speech, which critically was not the case. The finding that one's own speech drives eye-movement similar to that of someone else was based upon two separate studies. The first of these studies investigated the perception of a word in the context of a sentence spoken by someone else, while looking at a display with four printed words on them (Huettig & McQueen, 2007). The second of these studies investigated the production of a single word, while looking at a display with three written words and one target picture (Huettig & Hartsuiker 2010). In the current study we minimized the difference between the two conditions; the perception and production condition both consisted of a display with four written words, with the minimal difference that in the production condition the target word was underlined. In the production condition a single word was produced, and in the perception condition a single word was perceived. We observed that in speech production and speech perception, eye-movements were indeed driven to the phonological competitor words presented on the screen. Crucially, the eye-gazes were observed in the same time frame after speech onset during production and during perception. The result also showed that distinct processes influence phonological competition effects in speech production and speech perception. The competitor effect in speech perception was larger and lasted longer. This possibly results from the predictability of the target during speech production, and from suppression after target selection during speech production. In conclusion, we perceive our own and someone else's speech similarly via the external monitoring route, but because of a profound difference in predictability, the way we listen is different.

The study in Chapter 3 replicated the earlier finding that listening to one's own speech drives eye-movements similarly to phonological competitors, and it does not support the idea that we attend our internal speech, similar to the way we attend external speech. If internal speech is attended similarly to external speech during speech production, it did not drive the eye-movements in a similar fashion.

The results of this first empirical study call into question whether it is possible to attend to our own internal speech during overt speech production, as is suggested in the perceptual loop theory. According to the perceptual loop theory, we perceive our own internal speech before speech production, and can monitor our internal speech through this perception. The results of this study imply that our internal speech is not perceived by ourselves similarly to our external speech. Previous studies have shown that we are able to attend to internal speech in covert production tasks, but the current study calls into question whether we also do so during overt speech production. Consequently, the results of this study argue against a perceptual loop theory for verbal monitoring.

In Chapter 4 we used fMRI to investigate whether verbal error detection takes place through the speech perception system, as perception based theories of verbal monitoring assume, or whether verbal error detection is independent of perception, as production based theories of verbal monitoring assume. In this fMRI experiment, participants performed a tongue twister task in a perception and a production condition. In the production condition participants repeated tongue twister sentences overtly while their external speech was noise-masked by a loud white noise presented over headphones (additionally to the noise of the scanner). After repetition of a tongue twister sentence, participants indicated via a button press whether they had made an error in the repetition. In the perception condition participants heard someone else repeat the tongue twister sentence, and again indicated via a button press whether the repetition contained an error. By subtracting activations of correct repetitions from activations of incorrect repetitions, we obtained neurofunctional data of neuronal structures involved in verbal error detection during production and verbal error detection during perception. A conjunction analysis revealed that verbal error detection during production and perception share a neural network, which includes pre-supplementary motor area (pre-SMA), dorsal anterior cingulate cortex (dACC), anterior insula (AI), and inferior frontal gyrus (IFG). This same network has been found to be active in error making and error observation in the action domain (Newman-Norlund et al. 2009; Desmet et al. 2013, Monfardini et al. 2013). Crucially, a ROI analysis of areas in the superior temporal cortex revealed that the STG was more active in speech perception compared to speech production, but no activations consistent with error detection were found. The findings of this study therefore do not offer support for perceptual monitoring theories, but support production-based monitoring with a domain general error-detection mechanism, such as the conflict monitoring theory proposed by Nozari et al. (2011) or forward models as proposed by Pickering and Garrod (2013, 2014).

In Chapter 5 we investigated whether verbal monitoring is impaired in Parkinson's disease patients, compared to healthy age-matched controls. Furthermore we investigated the contributions of the internal and external monitoring route in verbal monitoring, by comparing verbal monitoring behavior during noise-masked and normal feedback. Participants performed 11 tasks, which measured cognitive performance, internal speech processing, external speech processing, internal verbal monitoring (during noise-masked feedback) and external verbal monitoring. Verbal monitoring showed a significant effect of the feedback conditions for semantic errors; the control group repaired relatively more errors during normal feedback than during normal feedback. This effect was not observed when the same analysis was performed on the percentage of all errors that were repaired. Additionally, we used the obtained data to test whether internal verbal monitoring performance could be predicted on the basis of internal speech tasks. The perceptual loop theory would assume so, as according to this theory internal speech monitoring takes place via the perception of internal speech. Results of the regression analysis revealed that internal speech monitoring could not be significantly predicted by performance on any of the internal speech tasks. Neither could the monitoring performance under normal feedback be predicted by performance on any of the language tasks. This suggests that speech monitoring might take place relatively independently of speech perception.

To summarize, the results reported in Chapter 3, Chapter 4, and Chapter 5 call into question the validity of a perceptual based account for verbal monitoring. In Chapter 2 we did not find any evidence for the perception of internal speech during overt production. If internal speech were perceived similar to overt speech, we would have expected to find perceptual effects driven by this internal speech. In Chapter 4 we did not find an increased neural activation in the auditory perceptual area (STG) during error detection. Instead, we found increased activation in a network of areas that has been widely demonstrated to be active during error detection in the action domain, suggesting that a domain general mechanism is active during verbal monitoring. In Chapter 5 we found that internal speech processing did not predict internal monitoring behavior, suggesting that speech monitoring might occur independent of speech perception. Taken together, these studies argue against the perception of internal speech as the mechanism for verbal monitoring, and the empirical fMRI study provide positive evidence in favor of a domain general conflict-monitoring network.

In Chapter 6 we proposed an extension to the production-based conflict monitoring account of verbal monitoring (Norzari et al., 2011). In this chapter we propose conflict monitoring as an account for monitoring during speech production and perception. Furthermore, we propose that at the neural level the conflict is detected by the ACC, which sends a signal to the LC and triggers the release of a neurotransmitter. This neurotransmitter boosts all ongoing activations, thereby increasing the signal-to-noise ratio, and their connections, whereby learning takes place. The function of this boost of activation would be to help selection of the correct item for production or perception. Depending on the timing of the boost, it might steers the selection of the correct item (preventing an error), or it might select a repair (reacting to an error).

In this dissertation we used a multitude of approaches to investigate verbal monitoring, which has given a very interesting insight into how internal verbal monitoring takes place. A disadvantage of this approach is that it also leaves a lot of questions, as none of the methods is used exhaustively. It would, for example, have been interesting to see whether the observed brain activations in response to tongue twister repetitions are also observed in more naturalistic speech. Furthermore, the current dissertation has focused on the role of the perception system for error monitoring. While the results with regard to this question is quite clear, the other theories of verbal monitoring are not really put to the test in the current dissertation.

DIRECTIONS FOR FUTURE RESEARCH

The chapters presented in the current thesis leave several questions unanswered. First of all, we propose a new model in Chapter 6, of which the assumptions need verification. Second, Chapters 2-5 argue against a perception-based account for verbal monitoring, but do not argue against any of the production based monitoring theories. We should therefore now ask the question: is speech monitoring forward model based, or conflict based? And how domain general is verbal monitoring? Is monitoring at least the same for different modalities of language production? And then there are two domains that have received little attention from research into error monitoring and repair, but which are quite interesting: the domains of acquisition and bilingualism.

IS THERE EVIDENCE FOR THE RELEASE OF A NEUROTRANSMITTER SUCH AS DOPAMINE OR ADRENALINE DURING SPEECH PRODUCTION AND PERCEPTION?

In Chapter 6 we provided a theoretical account of how errors can be detected and resolved via a domain general conflict monitoring mechanism. A claim that we make is that the release of a neurotransmitter, such as dopamine or adrenaline, leads to a (rather aspecific) boost of activation, which in turn helps the selection of the correct item for production or perception. This claim can be tested in several ways. First of all, a computational simulation of the boost of activation in the model proposed by Norazi et al. (2011) can test if there is credence to this claim. Note that Nozari et al.'s model shows that in the two-step model conflict is a clear indication of an error. A further step would be to demonstrate that a system that detects conflict and responds to it (i.e., the booster mechanism we propose) has the dual effect of reducing the number of errors (the steering function) and selecting successful repairs. Second, our proposal can be tested empirically using measurements of this neurotransmitter in situations of high conflict. This can be done directly or indirectly, such as via pupil dilations (noradrenaline) or in patient populations. Abnormal noradrenaline functioning is observed in schizophrenic patients and patients with Alzheimer's disease. Dopamine functioning is for instance impaired in Parkinson's patients, ADHD, and Schizophrenic patients (Jensen et al., 2007). In Chapter 5 we did not find big effects on verbal monitoring in Parkinson's patients, which might partially result from the levodopa medication. By testing the patients without levodopa medication we could perform a more direct test of the effect of dopamine during monitoring.

IS MORNITORING SIMILIAR FOR LANGUAGE PROCESSING TASKS?

While the current research has found a network for verbal error detection that has been observed to serve the same function during in the action domain (Newman-Norlund et al. 2009; Desmet et al. 2013, Monfardini et al. 2013), the current thesis has not investigated verbal monitoring outside of the speech domain. An interesting combination of speech and action, which deserves more research, is sign language. Some fundamental research on monitoring has been performed, such as the role of visual feedback in self-monitoring (Emmory et al, 2007, 2009), but this area is relatively unexplored. Another obvious direction to study verbal monitoring is via writing or typing. Especially the last method is very suited to use in techniques that require participant to move as little as possible (e.g. fMRI).

DO WE FIND SIMILAR EFFECTS IN VERBAL MONITORING AS IN CONFLICT MONITORING?

Within the domain of conflict monitoring there are a few well known cognitive adaptation effects after conflict; post-error slowing and the Gratton effect. Post-error slowing entails that the participant responds slower on a trial if it follows a conflict trial (post-error slowing; Rabbitt & Phillips, 1967). The Gratton effect entails that after an incongruent trial, a reduced congruency effect is observed (see Gratton, Coles, & Donchin, 1992). Off course language processing is in general much more complex than any of the tasks normally used in conflict monitoring studies, such as the Simon task or Flanker task, however, it would be worthwhile to investigate whether the same post-conflict effects are observed. It is difficult to operationalize these effects for speech; at what level should we observe the effect? At the phoneme level, syllable level or word level? Systematic research into this question might provide this insight. If indeed we find these effects, they would provide strong evidence for domain general monitoring process.

IS DOMAIN GENERAL MONITORING CONFLICT BASED, OR FORWARD MODEL BASED?

The current dissertation, especially the fMRI study presented in Chapter 4, provides evidence for a domain general, rather than a language specific, verbal monitoring. However, there is not much empirical evidence that suggests that monitoring is conflict based, rather than forward model based. Arguments against a forward modeling account are mostly theoretical. One empirical finding that argues against a forward model account for verbal monitoring, is the observation that the cerebellum was not more active in erroneous trials compared to correct trials. The cerebellum has been proposed as the critical structure for the storage and construction of forward models (Ito, 2008). However, neither Pickering and Garrod, nor Hickok assumes that the role of the cerebellum would be to detect errors. Contrastingly, Indefrey and Levelt do assume that the detection of errors would occur in the STG. Therefore the lack of cerebellar activation is no direct evidence against the forward models, but the lack of STG activation is direct evidence against the perceptual loop theory. So while the evidence from our fMRI study slightly favors a conflict approach for monitoring, it does not provide a strong argument against forward model theories.

One interesting possibility to test the conflict monitoring vs. forward modeling distinction would be to test schizophrenic patients. The idea of forward models for speech production is based on the success of forward models to be able to explain proprioceptive effects, and motor control (Wolpert, 1997). Several studies have suggested that schizophrenic patients have deficient forward model mechanisms (Feinberg, 1978, Ford and Mathalon, 2005, Frith, 1987, Ferri et al., 2014). If indeed monitoring is domain general, and forward model based, then these schizophrenic patients should display severe verbal monitoring problems. These deficient forward model mechanisms have been proposed as a means to explain auditory hallucinations in verbal monitoring (Feinberg, 1978; Frith, 1992; Jones &

Fernyhough, 2007; Seal et al., 2004). But a thorough investigations of how exactly verbal monitoring is impaired in schizophrenic patients is lacking at this moment.

Interestingly, as noted above, in schizophrenic patients the neurotransmitters dopamine and noradrenaline also function abnormally. By investigating the relationship between the neurotransmitters (dopamine and noradrenaline), forward models and verbal monitoring, we can start to investigate the contribution made by the neurotransmitters and forward models on verbal monitoring.

Another possible outcome from the research contrasting conflict monitoring with forward model theories might be that both of the theories are correct. At the end of page 28 in the paper by Nozari et al (2001) and as suggested by Dell (2013), it is possible to assume a merge of the two theories. Both theories assume that a conflict signal arises, which leads to the subsequent repair. And both theories assume that comparisons take place without the involvement of the perception system. Cascading of activation and forward models seem to serve similar functions; preparation of the response

HOW IS VERBAL MONITORING ACQUIRED?

Research into the acquisition of error monitoring so far, suggests that monitoring is gradually acquired during development. Jaeger (1992) kept a diary of the speech errors and repairs produced by three children from the ages one to five. Produced errors are categorized into phonological, word-based, or phrase-based errors. Between the ages 2 and 4 a reasonable number of data for each category are collected (n>25). Over all categories there is a steady decrease of the percentage of uncorrected errors and hence an increase of the percentage of self-corrected errors. Phonological errors have a self-repair rate of 36% at age 2, 48% at age 3, and 62% at age 4. Adults show a self-repair rate of phonological errors of 75% (Nooteboom, 1980). Both word errors and phrasal errors have an increasing rate of self-repair between ages 2 and 4, even exceeding the self-repair rate of adults. The results show a clear increase in self-monitoring performance. Some caution is warranted for these data, as they are diary studies and therefore prone to an observer bias. Additionally the sample size of the study is very small (n=3). One prediction that our model makes, is that the acquisition of verbal self-monitoring is correlated to self-monitoring in other domains (such as action monitoring). Another hypothesis is that the acquisition of monitoring will be correlated to the maturation of certain brain structures, in particular the LC and ACC.

IS MONITORING THE SAME IN L2 AS IN L1?

Self-monitoring in a second language studies have largely focused on the development of monitoring throughout L2 acquisition (for an overview, see Kormos, 2006). This research has revealed that the global frequency of errors does not depend on the level of proficiency, but that the nature of the self-repairs changes with proficiency; discourse level and appropriateness repairs are associated with a high competence, whereas repairs at the structural, semantic and phonological level are associated with lower proficiency. (Evans 1985; O'Connor 1988; Lennon 1990; Kormos 2006, Van Hest 1996). A comparison between self-monitoring in L1 and L2 revealed a similar distribution of self-repairs (Van Hest, 1996). But many open questions remain. For instance, it is not clear whether monitoring is as automatic in L1 as in L2, and whether monitoring interacts with for instance language switching is still open for investigation. One prediction our model makes with respect to bilingual monitoring is that monitoring in production and perception can develop independently (Note, however, that this prediction does not seem to be specific for bilingualism). It is for instance possible that one is unable to produce a phoneme correctly, but can very well detect the incorrect production in the speech produced by someone else as the perceptual levels are developed further than the corresponding production levels.

CONCLUSIONS

The research reported in this dissertation provides converging evidence for a domain-general verbal monitoring system. Taken together the research in this thesis argues against perception-based verbal monitoring; there is no behavioral evidence that supports the perception of internal speech similar to external speech, and the neurological data show that the STG is active in speech perception, but not as a monitoring mechanism. The current research clearly supports the existence of a domain-general monitor.

REFERENCES

Desmet, C., Deschrijver, E., Brass, M. (2013). How social is error observation? The neural mechanisms underlying the observation of human and machine errors. *Social Cognitive and Affective Neuroscience* 9,427-435.

Emmorey, K., Mehta, S., Grabowski, T. J. (2007). The neural correlates of sign versus word production. *Neuroimage*, *36*, 202-208.

Emmorey, K., Bosworth, R., Kraljic, T. (2009). Visual feedback and self-monitoring of sign language. *Journal of memory and language*, *61*(3), 398-411.

Evans, M. (1985). Self-initiated speech repairs: A reflection of communicative monitoring in young children. *Developmental Psychology*, 21, 365–371.

Feinberg, I. (1978). Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin*, 4, 636–640.

Ferri, F., Costantini, M., Salone, A., Di Iorio, G., Martinotti, G., Chiarelli, A., ... & Gallese, V. (2014). Upcoming tactile events and body ownership in schizophrenia. *Schizophrenia research*, *152*(1), 51-57.

Ford, J.M., Mathalon D.H. (2005). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *International Journal of Psychophysiology*, 58, 179–189.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia.* Psychology Press.

Frith, C. D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine,* 17,. 631–648

Gratton, G., Coles, M. G. H., Donchin, E. (1992). Optimizing the use of information – Strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121, 480–506.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13, 135-145

Huettig, F., Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language and Cognitive Process.* 3, 347–374.

Huettig, F., McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language,* 54, 460–482

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, 2, 255.

Indefrey, P., Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144.

Jaeger, J. J. (1992). Phonetic features in young children's slips of the tongue. *Language and Speech*, 35, 189-205.

Jensen, J., Smith, A. J., Willeit, M., Crawley, A. P., Mikulis, D. J., Vitcu, I., Kapur, S. (2007). Separate brain regions code for salience vs. valence during reward prediction in humans. *Human Brain Mapping.* 28, 294–302.

Jones, S. R., Fernyhough, C. (2007). Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition,* 16, 391–399.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Levelt, W. J. M. (1983) Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* MIT Press.

Monfardini E, Gazzola V, Boussaoud D, Brovelli A, Keysers C, Wicker B (2013). Vicarious neural processing of outcomes during observational learning. *PLOS One*, 8.

Newman-Norlund, R.D., Ganesh, S., van Schie, H.T., De Bruijn, E.R.A., Bekkering, H. (2009). Self-identification and empathy modulate error-related brain activity during the observation of penalty shots between friend and foe. *Social Cognitive and Affective Neuroscience*, 4, 10-22.

Nooteboom, S. G. (1980). *Speaking and unspeaking: Detection and correction of phonological and lexical errors of speech*. In V. A. Fromkin (Ed.), Errogarrett

Nozari, N., Dell, G. S. Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology,* 63,1–33.

O'Connor, N. (1988). *Repairs as indicative of interlanguage variation and change.* Georgetown University round table on languages and linguistics. ed. T. J. Walsh, 251–259. Washington, DC: Georgetown University Press. Lennon, P. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–417.

Pickering, M. J., Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36, 329-392.

Pickering, M. J., Garrod, S. (2014). Self- Other and joint monitoring using forward models. *Frontiers in Human Neuroscience,* 8,132-143

Rabbitt, P. M., Phillips, S. (1967). Error-detection and correction latencies as a function of S-R compatibility. *The Quarterly Journal of Experimental Psychology*, 19, 37–42.

Seal, M. L., Aleman, A., McGuire, P. K. (2004). Compelling imagery, unanticipated speech and deceptive memory: Neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry*, 9(1–2), 43–72.

Van Hest, E. (1996). *Self-repair in L1 and L2 production.* Tilburg: Tilburg Univ. Press.

Wilkinson, S. (2014). Accounting for the phenomenology and varieties of auditory verbal hallucination within a predictive processing framework. *Consciousness and cognition*, *30*, 142-155.

Wolpert, D. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6), 209–216.

NEDERLANDSE SAMENVATTING

Om de communicatie vlot te laten verlopen is het belangrijk om zo min mogelijk fouten te produceren tijdens het spreken. In het huidige doctoraat wordt onderzocht op welke wijze we onze spraak controleren op fouten, het zogenaamde monitoren van de spraak. Deze dissertatie richt zich specifiek op de vraag in hoeverre spraakperceptie een rol speelt bij het monitoren van de eigen spraak.

Als we willen weten hoe we fouten voorkomen, zullen we eerst moeten weten hoe we fouten ontdekken. Zodra we een fout hebben geproduceerd kunnen we hem natuurlijk horen. Deze manier om fouten te ontdekken noemen we de externe route. Maar als we pas een fout zouden kunnen ontdekken nadat we hem produceren, zouden we nooit een fout kunnen voorkomen.

Er zijn verschillende redenen om aan te nemen dat we versprekingen kunnen ontdekken voor we ze hebben geproduceerd. De eerste observatie is dat mensen extreem snel zijn met het stoppen van het uitspreken van het verkeerde woord, en dat ze dan soms vrijwel direct het juiste woord beginnen uitspreken, bijvoorbeeld ‘v-horizontaal’ (Levelt, 1989). De tweede observatie is dat wanneer je jezelf niet kunt horen spreken, omdat je bijvoorbeeld alleen in gedachten spreekt, je toch kunt aangeven dat je een fout hebt gemaakt (Dell & Repka 1992; Oppenheim & Dell, 2008). Dit is bijvoorbeeld het geval bij het intern herhalen van een tongbreker, zoals ‘Gijs grijpt de grijsgrauwe gans graag gauw’. En ook wanneer de externe spraak gemaskeerd wordt door een luid geluid, zodat je niet meer kunt horen wat je zelf zegt, kun je toch aangeven wanneer je een verspreking hebt gemaakt (Lackner & Tuller, 1979, Postma & Kolk 1992a, b). Een derde observatie is dat wanneer een verspreking wordt uitgelokt, er minder fouten worden gemaakt in die gevallen wanneer een spreekfout zou leiden tot een gênante uiting (vb. geen stijl – steen geil) (Motley, Camden, & Baars, 1981, 1982). In de trials waarin zulke woordparen worden gepresenteerd wordt een huidreactie gemeten, de electrodermale respons, wat er op wijst dat de persoon zich bewust is van de potentiele gênante uiting. Blijkbaar heeft de persoon de verspreking intern gemaakt, maar kan zij voorkomen dat de verspreking ook daadwerkelijk wordt geuit. Deze punten samengenomen wijzen er dus op dat er naast de externe route ook een interne route is om fouten te detecteren.

Hoe precies het interne monitoren plaatsvind is nog niet volledig duidelijk. Er zijn een aantal theorieën rondom interne verbale monitoring, die we grofweg in twee categorieën kunnen opdelen, op basis van de manier waarop de fout zou worden ontdekt. De eerste groep van theorieën gaan uit van een perceptiegebaseerde monitoring, de tweede groep veronderstelt dat verbale monitoring plaatsvind tijdens het productieproces.

Van de perceptiegebaseerde monitoring theorieën is de perceptuele loop theorie (PLT) de belangrijkste (Levelt, 1983, 1989). Deze theorie veronderstelt dat vlak voor het uitspreken een tussenproduct in de productie, namelijk het fonetisch plan, van hetgeen je wilt zeggen naar het perceptie systeem wordt gestuurd. Via deze interne route hoor je dan je eigen interne spraak op eenzelfde wijze op de manier waarop je je eigen externe spraak hoort. Vervolgens wordt intern een vergelijking gemaakt van de intentie en hetgeen je hebt gezegd. De belangrijkste neurologische structuur verantwoordelijk voor het verbale monitoren is volgens deze theorie de gyrus temporalis superior (GTS) (Indefrey & Levelt, 2004; Indefrey, 2011). Een pluspunt van de PLT is dat de theorie zeer zuinig is; ze veronderstelt dat monitoring kan plaatvinden via het perceptiesysteem, waarvan we zeker weten dat het aanwezig is. Bewijs voor de theorie komt onder meer van computersimulaties van het model, die de korte interruptietijden die zijn geobserveerd na het produceren van een fout konden repliceren (Hartsuiker & Kolk, 2001). Overig bewijs voor de PLT is niet zo sterk. Zo is er bewijs gevonden voor perceptuele effecten tijdens het monitoren van interne spraak (Özdemir, Roelofs en Levelt, 2007). Dit effect werd echter waargenomen in een proefopzet waarbij interne spraak werd gemonitord in afwezigheid van externe spraak. In een experiment waarbij de externe spraak wel aanwezig was, werd er geen perceptueel effect waargenomen van het monitoren van interne spraak (Huettig & Hartsuiker, 2010). Bewijs voor de rol van de GTS tijdens het monitoren van de spraak komt van studies waarbij de spraak van een proefpersoon werd veranderd in toonhoogte of timing voordat de proefpersoon deze hoorde ('altered auditory feedback'). In dergelijke experimenten werd telkens een verhoogde activatie in de GTS gemeten, ten opzichte van het luisteren naar onveranderde feedback (McGuire et al, 1996; Hirano et al, 1997; Hashimoto & Sakai 2003; Christoffels et al, 2007, 2011; Tourville et al, 2008; Zheng et al, 2010; Takaso et al, 2010, Shergill et al., 2002). Wat hier echter wordt gemeten is hersenactiviteit als reactie op veranderde feedback, en niet als reactie op een fout. Bovendien wordt in bovengenoemde studies alleen de externe spraak veranderd, terwijl de vraag juist is 'hoe wordt interne spraak gemonitord?' Daarnaast zijn er veel patiëntenstudies die dissociaties aantonen tussen verbale monitoring tijdens productie en perceptie, en dat zowel het fonologische als het semantische niveau tijdens productie apart kunnen zijn aangetast (vb. Oomen, Postma & Kolk, 2005). In de PLT is monitoring tijdens productie en perceptie afhankelijk van perceptie. Het is dus in strijd met deze theorie dat een van beide processen onafhankelijk zou zijn aangetast.

Productietheorieën van verbale monitoring veronderstellen dat tijdens de constructie van de spraak er al steeds gecontroleerd wordt op fouten. Dit kan op twee manieren. Ten eerste is het mogelijk dat sprekers tijdens elke constructie stap een *voorspelling* van de uitkomst vergelijken met de daadwerkelijke uitkomst, zoals theorieën op basis van zgn. voorwaartse modellen veronderstellen (Pickering & Garrod, 2013, 2014; Hickok, 2012). Ten tweede zou de spreker het *conflict* kunnen monitoren tijdens de selectie van een item voor productie (Nozari, Dell, & Schwartz, 2011), waarbij conflict (meerdere items hebben een hoge activatie) diagnostisch zou zijn voor fouten.

Een belangrijke aspect van voorwaartse modellen theorieën is de aanname dat sprekers (en actoren in het algemeen) voorspellingen maken over de uitkomst van hun acties. Wanneer ik bijvoorbeeld zelf op de onderkant van een ketchupfles sla, zal deze nauwelijks bewegen; maar als ik de ketchupfles vasthoud en jij slaat op deze fles, zal de fles sterk naar beneden bewegen. Ik kan immers nauwkeurig de timing en kracht voorspellen van mijn eigen slag op de fles, maar niet van de jouwe. Veel onderzoek heeft aangetoond dat mensen tijdens het verwerken van taal ook voorspellingen maken. In oogbewegingsstudies is gevonden dat tijdens het luisteren naar spraak een voorspelling wordt gemaakt van wat gezegd gaat worden; de blikken gaan naar het plaatje waarvan het het meest waarschijnlijk is dat het zal worden genoemd (Altmann & Kamide 1999; Kamide, Altmann & Haywood, 2003). EEG studies, waarin hersenactiviteit wordt gemeten, laten een piek in hersenactiviteit zien wanneer een woord wordt gebruikt dat onverwacht is (Delong et al., 2005; Kutas & Federmeier, 2011; Van Berkum et al., 2005). Daarnaast zijn in twee MEG studies effecten gevonden van voorwaartse modellen tijdens het inbeelden van een spreken en luisteren (Tian & Poeppel, 2010, 2013). In deze studies vond men dat wanneer je inbeeld dat je iets zegt, er ook hersenactivatie volgt in de perceptuele gebieden.

Theorieën op basis van voorwaartse modellen hebben echter een aantal theoretische problemen. Enerzijds moet het voorwaartse model sneller worden geconstrueerd dan de daadwerkelijke uiting, maar anderzijds moet het zeer specifiek zijn om een zinvolle vergelijking te kunnen maken die toelaat om fouten te kunnen opmerken. Als het voorwaartse model snel, accuraat en zeer gespecificeerd is, waarom zouden we dan nog een traag constructieproces nodig hebben dat er op volgt? (e.g., Hartsuiker, 2013; Bowers, 2013; Strijkers et al, 2013; Meyer & Hagoort 2013) Daarnaast veronderstelt de voorwaartse model theorie van Pickering en Garrod (2013, 2014) ook een wisselwerking tussen productie en perceptie die dusdanig is, dat de dissociaties uit de patiëntenstudies niet verklaard kunnen worden. De theorie van Hickok (2012) veronderstelt dit niet, maar deze theorie beperkt zich tot foneemselectie, articulatie en de perceptie daarvan.

Conflictmonitoring theorieën gaan er vanuit dat tijdens het selectieproces meerdere items actief zijn, en in competitie om geselecteerd te worden. Deze competitie leidt tot als maar hoger wordende activatie van meerdere items. In zo'n geval spreken we van een conflict. Wanneer dit conflict een bepaalde grens overschrijdt wordt het opgemerkt door een algemeen monitoringmechanisme. Het eerste argument voor dit model is dat van dit model een computersimulatie is gemaakt waarmee men de foutenpatronen heeft kunnen nabootsen van patiënten met hersenbeschadigingen, door op bepaalde plaatsen in het model de verbindingen zwakker te maken. Een tweede argument voor dit model is dat zij aansluit op theorieën buiten de taalverwerking. In de cognitieve controle literatuur is al veel ondersteuning gevonden voor conflictmonitoring als een controlemechanisme. Bewijs voor een onderliggend monitoring mechanisme dat op conflict reageert is de bevinding dat fMRI studies waarin conflict tussen responsopties optreedt altijd een verhoogde activatie in de cortex cingularis anterior (ACC) vinden (zie bijvoorbeeld Botvinick et al. 2001 voor een overzicht). En in EEG-studies vindt men altijd de ERN/Ne component wanneer er conflict tussen responsopties is (e.g., Vidal et al., 2000; Falkenstein et al 1990, 1991, 1995; Gehring et al 1993; Ganushchak & Schiller 2006, 2008). Deze twee bevindingen zijn gerepliceerd binnen zowel taalonderzoek als cognitieve controle onderzoek. Een commentaar op de conflict monitoring theorie zoals voorgesteld door Nozari, Dell & Schwartz (2011) is dat zij niet duidelijk maakt hoe precies het verhoogde conflict leidt tot het voorkomen van de foutproductie. Daarnaast is het model alleen omschreven voor verbale monitoring tijdens productie, maar niet tijdens perceptie.

Samengevat heeft ieder van deze modellen diverse sterke en zwakke punten. Geen van de huidige modellen is in staat om een volledige en adequate beschrijving te geven van verbale monitoring tijdens productie en perceptie.

DE ROL VAN PERCEPTIE TIJDENS INTERNE VERBALE MONITORING

In deze thesis onderzoeken we verder of we een rol kunnen vinden voor perceptie in verbale monitoring tijdens productie. De perceptuele loop theorie veronderstelt dat je het fonetisch plan hoort, op een zelfde manier als wanneer je iemand anders zou horen spreken, en dat je aan de hand daarvan kunt controleren op fouten. Je zou dus wanneer je zelf spreekt eerst via de interne route het fonetisch plan horen, en daarna via de externe route nogmaals de eigen spraak. In deze thesis onderzoeken we of we een effect van die perceptie van de interne spraak kunnen vinden tijdens de productie van woorden. Eerdere studies hebben gevonden dat wanneer op een scherm een woord stond wat sterke fonologische overeenkomt had met een woord wat moest worden uitgesproken of wat werd gehoord (je hoort bijvoorbeeld 'man' en in beeld staat ook het woord 'mango'), er zeer veel visuele aandacht was voor dit woord ten opzichten van andere woorden op het scherm. Verreweg de meeste aandacht ging uit naar het woord dat identiek was (dus uitgesproken of gehoord werd), en daarna naar het woord dat sterke fonologische overeenkomsten had. Opmerkelijk was dat de oogbewegingen naar het fonologisch gerelateerde woord in vrijwel hetzelfde tijdsframe plaatsvonden wanneer men zelf sprak als wanneer men een ander hoorde spreken (Huettig & McQueen, 2007; Huettig & Hartsuiker, 2010). Dit komt niet overeen met de PLT, die immers veronderstelt dat je vlak voor je gaat spreken al het fonetisch plan hoort van hetgeen je gaat zeggen via de interne route. Dit is echter gemeten in twee verschillende studies, die qua opzet vrij veel van elkaar verschilden. In de productie studie zag men een scherm met een afbeelding en drie geschreven woorden, waarbij de afbeelding moest worden benoemd (Huettig & Hartsuiker, 2010). In de perceptiestudie zag men vier geschreven woorden, en hoorde men het doelwoord in de context van een zin (Huettig & McQueen, 2007). Om te testen of we inderdaad onze visuele aandacht laten sturen door externe spraak, en niet door interne spraak wanneer wij zelf produceren, hebben we een soortgelijke studie uitgevoerd waarin het design is verbeterd. In deze studie werden de productie- en perceptieconditie uitgevoerd door dezelfde proefpersonen, en in beide gevallen zag men op het scherm vier geschreven woorden met als minimaal verschil dat in de productieconditie het woord dat moest worden uitgesproken was onderstreept. Men hoorde of produceerde slechts een woord. Deze studie toonde aan dat ook in deze gelijke condities de oogbewegingen werden gedreven door externe spraak; zowel wanneer men zelf sprak als wanneer men luisterde begon men tussen de 400 en 600 ms nadat het woord werd waargenomen te kijken naar het fonologisch gerelateerde woord. De oogbewegingen tijdens spraakproductie werden dus niet gestuurd door de perceptie van interne spraak, maar uitsluitend door de externe spraak.

In een fMRI studie hebben we onderzocht welke hersenstructuren betrokken zijn bij het detecteren van spraakfouten. De vraag hierbij was of we inderdaad verhoogde activiteit zouden vinden in de auditieve perceptie gebieden in de STG, zoals de PLT veronderstelt, of dat we activiteit in andere gebieden zouden vinden. De proefpersonen herhaalden tongbrekerzinnen (bijvoorbeeld 'Pappa pakt de platte blauwe bakpan') terwijl ze in een 3T MRI scanner lagen. Tijdens het luidop spreken hoorde de proefpersoon via een koptelefoon zeer luid een witte ruis geluid, waardoor ze, zeker in combinatie met het lawaai van de scanner, zichzelf niet meer konden horen spreken. De proefpersoon was dus afhankelijk van interne spraak om haar spraak te kunnen monitoren. Na de herhaling van de tongbreker gaf de proefpersoon aan door een druk op de knop of ze de zin foutloos had herhaald, of dat ze een fout had gemaakt. Dezelfde taak werd herhaald in een perceptieconditie, waarbij de proefpersoon een geluidsopname hoorde van iemand anders die de tongbreker herhaalde. Ook hier moest de proefpersoon aangeven of de herhaling van de tongbreker foutloos was, of dat er een fout in voorkwam. Vervolgens hebben we de hersenactiviteit genomen van alle gevallen waarin een fout was gemaakt en de proefpersoon deze had opgemerkt, en daar de activatie vanaf gehaald tijdens de correcte trials. De resterende activatie geeft nu weer welke hersengebieden precies actief zijn bij het detecteren van een fout. Dit is apart voor productie, waarbij we interne spraak monitoring hebben gemeten, en voor perceptie, waarbij we externe spraak monitoring hebben gemeten. Daarna gingen we na welke gebieden actief werden bij zowel foutdetectie in productie, als in perceptie. Uit deze vergelijking kwam een netwerk van hersengebieden naar voren, bestaande uit de area premotoria supplementaria (pre-SMA), cortex cingularis anterior dorsalis (dACC), insula anterior (AI), en de gyrus frontalis inferior (IFG). Van dit netwerk was reeds bekend dat zij actief was bij het detecteren van fouten vanuit studies naar het oplossen van conflict in het actie domein (Newman-Norlund et al. 2009; Desmet et al. 2013, Monfardini et al. 2013). Daarnaast hebben we nog specifiek de activaties in de STG onderzocht, en hieruit vonden we geen patroon wat overeenkwam met het detecteren van fouten. Wat we dus hebben vastgesteld is dat niet de auditieve cortex verantwoordelijk is voor het vaststellen van fouten, maar eerder een algemeen monitorings mechanisme.

Patiëntendata zijn onmisbaar geweest voor het verkrijgen van inzicht in verbale monitoring. Parkinson patiënten vertonen een verscheidenheid aan taalproblemen. Wij hebben onderzocht of verbale monitoring is verstoord in Parkinson patiënten. Een groot deel van de taalproblemen bij Parkinson patiënten vertoont grote overeenkomsten met de problemen bij patiënten met een afasie van Broca. Bij patiënten met een afasie van Broca is reeds vastgesteld dat zij bovenmatig afhankelijk zijn van de interne route voor monitoring. Het aantal fouten dat werd gerepareerd was even groot wanneer zij alleen interne spraak konden monitoren omdat de externe spraak door een harde ruis was gemaskeerd, als wanneer zij zowel intern als extern konden monitoren in normale spraak. Met behulp van dezelfde taak en condities hebben wij vastgesteld dat ook Parkinson patiënten in grotere mate afhankelijk zijn van hun interne monitoring route in vergelijking met gezonde proefpersonen van dezelfde leeftijd. Dit was in het bijzonder het geval voor semantische fouten. Daarnaast hebben wij met de verkregen data onderzocht of het interne monitoring gedrag (vb., percentages gecorrigeerde fouten tijds spraak onder maskering met luide ruis) kon worden voorspeld op basis van de prestaties op een aantal interne spraak taken zoals bijvoorbeeld het zeggen of een bepaalde klank voorkomt in de naam van een plaatje. Dit bleek echter niet het geval. Het feit dat interne monitoring niet samenhangt met prestaties op interne spraak taken suggereert dat spraakmonitoring onafhankelijk is van interne spraak perceptie.

Samengevat wijzen de empirische studies in dit proefschrift er op dat er weinig evidentie is voor een rol van perceptie tijdens interne verbale monitoring. Onze resultaten wijzen eerder in de richting van een algemeen monitoringmechanisme dat onafhankelijk is van spraakperceptie.

EEN NIEUW MODEL VOOR VERBALE MONITORING

Op basis van eerdere onderzoeken en de studies in het huidig proefschrift hebben wij geconstateerd dat het onwaarschijnlijk is dat het monitoren van interne spraak plaatsvindt via de perceptie van deze interne spraak. Het is waarschijnlijker dat monitoring plaatsvindt tijdens het productieproces zelf. Het conflictmonitoring model (Nozari et al., 2011) omschrijft een algemeen monitoring mechanisme waarin de verbale monitoring plaatsvindt via de detectie van conflict tussen responsopties in ieder van de productiestappen tijdens de spraakproductie. Wij stellen voor om ditzelfde mechanisme toe te passen op spraakperceptie. Omdat de theorie zeer algemeen is, is zij eenvoudig toe te passen op spraakperceptie. Daarbij is deze theorie reeds computationeel geïmplementeerd, dus een uitbreiding naar perceptie zou direct getoetst kunnen worden door middel van computationele simulaties. Via een automatische spreiding van activatie tussen gelijkaardige niveaus in het productie en perceptie proces staan de twee processen in verbinding. Op deze manier kunnen productie-effecten tijdens perceptie en perceptie-effecten tijdens productie verklaard worden zonder de twee systemen afhankelijk van elkaar te maken. Wanneer de systemen afhankelijk van elkaar zijn is het model immers incompatibel met de patentendata die onafhankelijk functioneren van de twee modaliteiten demonstreren.

Bovendien lenen we een idee vanuit de cognitieve controle literatuur over hoe het gedetecteerde conflict kan worden opgelost. Verguts en Nootebaert (2008, 2009) stellen voor dat conflict tussen responsopties wordt gedetecteerd door de ACC. Op zijn beurt zal de ACC een signaal sturen aan de locus caeruleus (LC), die vervolgens een neurotransmitter de cortex in zend. De neurotransmitter zorgt voor een boost van activatie in alle actieve processen. Op deze wijze kan het conflict tussen verschillende responsopties eenvoudig worden opgelost. Afhankelijk van het moment waarop de activatieboost plaatsvindt tijdens responsselectie, zal de boost zorgen voor bijsturing, of voor de selectie van het juiste item (als reparatie van de fout). Dit mechanisme zou wellicht zowel kunnen werken voor responsselectie tijdens spraakproductie en perceptie.

CONCLUSIE

In dit doctoraatsonderzoek werd onderzocht of er een empirische basis is voor perceptie gebaseerde verbale monitoring van de interne spraak productie. Via verschillende technieken, het meten van oogbewegingen, het meten van hersenactiviteit via fMRI en via een patiëntenstudie, hebben wij getoetst of wij evidentie kunnen vinden voor perceptie gebaseerde monitoring. Op basis van de resultaten van de verschillende onderzoeken moeten wij echter concluderen dat het zeer onwaarschijnlijk is dat interne verbale monitoring plaatsvindt via perceptie van de spraak. De studies wijzen eerder in de richting van spraakmonitoring tijdens het productieproces. Verder onderzoek is nodig om te onderzoeken of verbale monitoring plaatsvindt via voorwaartse modellen, of eerder via conflict monitoring.

REFERENTIES

Altmann, G. T. M., Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition,* 73, 247–264.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review* 108, 624–652.

Bowers, J. (2013). How do forward models work? And why would you want them? *Behavioral and Brain Sciences*, 36, 349-350.

Christoffels, I.K., Formisano, E., Schiller, N.O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Human Brain Mapping*, 28, 868–879.

Christoffels, I.K., Van De Ven, V., Waldorp, L.J., Formisano, E., Schiller, N.O. (2011). The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. *PLoS One*: 6.

Dell, G.S. Repka, R.J. (1992). Errors in inner speech. B.J. Baars (Ed.), Experimental slips and human error: Exploring the architecture of volition (pp. 237–262). Plenum, New York.

DeLong, K.A., Urbach, T.P., Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117–1121.

Desmet, C., Deschrijver, E., Brass, M. (2013). How social is error observation? The neural mechanisms underlying the observation of human and machine errors. *Social Cognitive and Affective Neuroscience* 9, 427-435.

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L. (1990). *Effects of errors in choice reaction tasks on the ERP under focused and divided attention*. In C. Brunia, A. Gaillard, & A. Kok (Eds.), Psychophysiological brain research (pp. 192–195). Tilburg, the Netherlands: Tilburg University Press.

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components: II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, 78, 447–455.

Falkenstein, M., Koshlykova, N.A., Kiroj, V.N., Hoormann, J., Hohnsbein, J., (1995). Late ERP components in visual and auditory Go/Nogo tasks. *Electroencephalography and Clinical Neurophysiology*, 96, 36–43.

Ganushchak, L. Y., Schiller, N. O. (2006). Effects of time pressure on verbal self-monitoring: An ERP study. Brain Research, 1125, 104–115.

Ganushchak, L. Y., Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *Neuroimage*, 39, 395–405.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, 4, 385–390.

Hartsuiker, R. J. Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the Perceptual Loop Theory. *Cognitive Psychology,* 42, 113-157.

Hartsuiker, R. J. (2013). Are forward models enough to explain self-monitoring? Insights from patients and eye movements. *Behavioral and Brain Sciences*, 36, 357-358.

Hashimoto, Y., Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human brain mapping*, 20, 22-28.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13, 135-145.

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

Huettig, F., Hartsuiker, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. Language and Cognitive Process. 3, 347–374.

Huettig, F., McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language,* 54, 460–482.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, 2, 255.

Indefrey, P., Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144.

Kamide, Y., Altmann, G., Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133-156.

Kutas, M., Federmeier, K. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, 62, 621.

Lackner, J. R., Tuller, B. H. (1979). *Role of efference monitoring in the detection of selfproduced speech errors.* In W. E. Cooper & E. C. T. Walker (Eds.), Sentence processing: Studies dedicated to Merrill Garrett (pp. 281–294). Hillsdale, NJ: Earlbaum.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

McGuire, P. K., Silbersweig, D. A., Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, 119, 907–917.

Meyer, A. S., Hagoort, P. (2013). What does it mean to predict one's own utterances?. *Behavioral and Brain Sciences*, 36, 367-368.

Monfardini, E., Gazzola, V., Boussaoud, D., Brovelli, A., Keysers, C., Wicker, B. (2013). Vicarious neural processing of outcomes during observational learning. *PLOS One*, 8.

Motley, M.T., Camden, C.T., Baars, B.J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 21, 578-594.

Newman-Norlund, R.D., Ganesh, S., van Schie, H.T., De Bruijn, E.R.A., Bekkering, H. (2009). Self-identification and empathy modulate error-related brain activity during the observation of penalty shots between friend and foe. *Social Cognitive and Affective Neuroscience*, 4, 10-22.

Nozari, N., Dell, G. S., Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63, 1–33.

Oppenheim, G. M., Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition,* 106, 528-537.

Özdemir, R., Roelofs, A., Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition*, 105, 457–465.

Pickering, M. J., Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36, 329-392.

Pickering, M. J., Garrod, S. (2014). Self- Other and joint monitoring using forward models. *Frontiers in Human Neuroscience,* 8,132-143.

Riès, S., Janssen, N., Dufau, S., Alario, F. X., Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience*, 23, 1419-1436.

Postma, A., Kolk H.H.J. (1992a). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech and Hearing Research*, 35, 537–544.

Postma, A., Kolk, H. H.J. (1992b). Error Monitoring in People Who StutterEvidence Against Auditory Feedback Defect Theories. *Journal of Speech, Language, and Hearing Research,* 35, 1024-1032.

Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E. Jr., Murray, R. M., McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping.* 16, 219–227.

Strijkers, K., Runnqvist, E., Costa, A., Holcomb, P. (2013). The poor helping the rich: how can incomplete representations monitor complete ones? *Behavioral and Brain Sciences*, 36, 374-375.

Takaso, H., Eisner, F., Wise, R.J., Scott, S.K. (2010). The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study. *Journal of Speech Language and Hearing Research*, 53, 226–236.

Tian, X., Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology,* 1, 166.

Tian, X., Poeppel, D. (2013). The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of cognitive neuroscience*, 25, 1020-1036.

Tourville, J. A., Reilly, K. J., Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, 39,1429–1443.

Van Berkum, J. J. A., Brown, M. C., Zwitserlood, P., Kooijman, V. Hagoort, P. (2005) Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443–467.

Verguts, T., Notebaert, W. (2008). Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. *Psychological Review*, 115, 518-525.

Verguts, T., Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. *Trends in Cognitive Sciences*, 13, 252-257.

Vidal, F., Hasbroucq, T., Grapperon, J., Bonnet, M. (2000). Is the "error negativity" specific to errors? *Biological Psychology*, 51, 109–128.

Zheng, Z. Z., Munhall, K. G., Johnsrude, I. S. (2010). Functional Overlap between Regions Involved in Speech Perception and in Monitoring One's Own Voice during Speech Production. *Journal of Cognitive Neuroscience,* 22, 1770–1778.

APPENDIX 3A

Items in the display in the semantic, semantic and neutral conditon.

Semantic Target	**Competitor**	**Filler 1**	**Filler 2**
mes *knife*	kopje *cup*	fiets *bicycle*	vinger *finger*
been *leg*	pols *wrist*	zeef *sieve*	harp *harp*
duim *thumb*	oog *eye*	schaal *bowl*	bok *goat*
teen *toe*	arm *arm*	water *water*	lampion *lamp*
nek *neck*	lippen *lips*	kam *comb*	appel *apple*
steen *stone*	kei *boulder*	wind *wind*	computer *computer*
bed *bed*	stoel *chair*	oorlog *war*	lippenstift *lipstick*
berg *mountain*	rots *rock*	netje *net*	selder *celery*
tent *tent*	caravan *caravan*	ananas *pineapple*	saxofoon *saxophone*
onweer *lightning*	regen *rain*	sokken *socks*	broer *brother*
trein *train*	helikopter *helicopter*	printer *printer*	rog *ray*
pan *pan*	oven *oven*	bank *sofa*	vlag *flag*

appelsien *orange*	peer *pear*	heuvel *hill*	competitie *competition*
perzik *peach*	aardbei *strawberry*	kano *canoe*	broek *trousers*
zon *sun*	wolk *cloud*	hoofd *head*	radio *radio*
hamer *hammer*	boor *drill*	jas *coat*	mango *mango*
glas *glass*	kan *pitcher*	hok *pen*	batterij *battery*
hark *rake*	kruiwagen *wheelbarrow*	droom *dream*	lever *liver*
nagel *nail*	hand *hand*	fles *bottle*	salade *salad*
ring *ring*	horloge *watch*	vis *fish*	schilderij *painting*
ketting *necklace*	armband *bracelet*	grond *soil*	bot *bone*
touw *rope*	draad *thread*	hemd *shirt*	lam *lamb*
tuinslang *hose*	emmer *bucket*	dag *day*	servet *napkin*
stropdas *tie*	handschoen *glove*	neef *nephew*	plant *plant*

Phonological Target	**Competitor**	**Filler 1**	**Filler 2**
lamp *lamp*	lam *lamb*	kopje *cup*	druif *pigeon*

vinger *finger*	vin *fin*	caravan *caravan*	poes *cat*
hart *heart*	harp *harp*	appelsien o*range*	nest *nest*
bot *bone*	bok *goat*	perzik *peach*	klok *clock*
lat *latch*	lampion *lamp*	peer *pear*	neus *nose*
appel *apple*	anker *anchor*	hark *rake*	bil *buttock*
banaan *banana*	batterij *battery*	tuinslang *hose*	raam *window*
lepel *spoon*	lever *liver*	stoel *chair*	gordijn *curtain*
saxofoon *saxophone*	salade *salad*	helikopter *helicopter*	kleed *rug*
vlieger *kite*	vlierstruik *elder*	duim *thumb*	spons *sponge*
lippenstift *lipstick*	lift *elevator*	trein *train*	hond *dog*
rok *skirt*	rog *ray*	glas *glass*	borstel *brush*
riet *reed*	riem *belt*	tent *tent*	kiwi *kiwi*
broek *trousers*	broer *brother*	arm *arm*	ketel *kettle*
aardappel *potato*	aap *monkey*	rots *rock*	haven *harbour*
selder *celery*	servet *napkin*	teen *toe*	vork *fork*
gitaar *guitar*	gieter *water can*	horloge *watch*	trap *stairs*

vlag *flag*	vlam *flame*	hamer *hammer*	telefoon *telephone*
computer *computer*	competitie *competition*	kan *pitcher*	maan *moon*
radio *radio*	radar *radar*	ketting *necklace*	mok *mug*
schilderij *painting*	schil *peel*	touw *rope*	kin *chin*
man *man*	mango *mango*	handschoen *glove*	weer *weather*
plant *plant*	plak *slice*	wolk *cloud*	zebra *zebra*
bakker *baker*	bakfiets *tricycle*	aardbei *strawberry*	zaag *saw*

Neutral Target	**Filler 1**	**Filler 2**	**Filler 3**
hond *dog*	dag *day*	steen *stone*	lamp *lamp*
poes *cat*	droom *dream*	boor *drill*	vin *fin*
vis *fish*	grond *soil*	onweer *thunderstorm*	hart *heart*
fles *bottle*	maan *moon*	kruiwagen *wheelbarrow*	lat *latch*
neus *nose*	hemd *shirt*	ring *ring*	anker *anchor*
kin *chin*	hok *pen*	zon *sun*	banaan *banana*

hoofd *head*	jas *coat*	bed *bed*	lepel *spoon*
bil *buttock*	kano *canoe*	regen *rain*	vlieger *kite*
raam *window*	kiwi *kiwi*	lippen *lips*	bakker *baker*
printer *printer*	klok *clock*	oven *oven*	rok *skirt*
gordijn *curtain*	haven *harbour*	pan *pan*	vlierstruik *elder*
kleed *rug*	mok *mug*	hand *hand*	riet *reed*
bank *sofa*	neef *nephew*	oog *eye*	riem *belt*
heuvel *hill*	nest *nest*	emmer *bucket*	aardappel *potato*
wind *wind*	netje *net*	mes *knife*	aap *monkey*
ananas *pineapple*	oorlog *war*	draad *thread*	gitaar *guitar*
druif *pigeon*	sokken *socks*	berg *mountain*	gieter *water can*
spons *sponge*	telefoon *telephone*	armband *bracelet*	vlam *flame*
borstel *brush*	trap *stairs*	nagel *nail*	radar *radar*
kam *comb*	water *water*	stropdas *tie*	bakfiets *tricycle*
ketel *kettle*	zaag *saw*	pols *wrist*	man *man*
schaal *bowl*	weer *weather*	nek *neck*	plak *slice*
zeef *sieve*	vork *fork*	been *leg*	lift *elevator*

fiets *bicycle*	zebra *zebra*	kei *boulder*	schil *peel*

APPENDIX 3B

Proportion of fixations in the phonological condition, measured in 200 ms timeframes, and statistical values of the comparison of fixation proportions to target and competitor against the unrelated items.

		Perception				
Start time of measurement relative to speech onset (ms)		Mean	t_1	*p*	t2	*p*
-200	Target	0.246	-.647	.261	-.609	.275
	Competitor	0.245	-.551	.293	-.509	.308
	Unrelated	0.254				
0	Target	0.248	-.129	.449	-.103	.469
	Competitor	0.251	.068	.473	.033	.485
	Unrelated	0.250				
200	Target	0.268	1.388	.087	1.076	.147
	Competitor	0.259	1.182	.122	1.071	.148
	Unrelated	0.236				
400	Target		8.753	.000	6.794	.000

		0.365				
	Competitor	0.270	5.743	.000	4.591	.000
	Unrelated	0.182				
600	Target	0.524	22.674	.000	19.134	.000
	Competitor	0.247	7.904	.000	5.832	.000
	Unrelated	0.114				
800	Target	0.701	23.953	.000	25.335	.000
	Competitor	0.148	6.878	.000	3.995	.001
	Unrelated	0.075				

Start time of measurement relative to speech onset (ms)		*Production*				
		Mean	t_1	*p*	t2	*p*
-200	Target	0.957	93.100	.000	81.338	.000
	Competitor	0.014	-.114	.455	-.122	.452
	Unrelated	0.015				
0	Target	0.897	47.804	.000	55.878	.000
	Competitor	0.041	1.775	.042	1.089	.144
	Unrelated	0.031				
200	Target	0.815	23.035	.000	36.292	.000
	Competitor	0.076	1.717	.047	2.445	.012
	Unrelated	0.054				
400	Target	0.742	14.591	.000	33.059	.000
	Competitor	0.096	2.110	.021	1.300	.104
	Unrelated	0.080				
600	Target	0.700	12.533	.000	32.661	.000
	Competitor	0.100	.018	.493	.056	.478
	Unrelated	0.099				
800	Target	0.666	11.531	.000	27.399	.000
	Competitor	0.106	-.490	.314	-.481	.318
	Unrelated	0.113				

APPENDIX 3B

APPENDIX 5A

Ravens matrices

Participants below 65 years performed set B, C and D from the standard progressive matrices. From the Raven Manual we used the normal distribution of scores to estimate their score for the full set. This estimation was used to compare with the norms for adults in Belgium to calculate at what percentile their score was.

Participants +65 years of age performed test A, AB, B of the colored progressive matrices. From the Raven Manual used the norms for elderly people in the Netherlands to calculate at what percentile their score was.

PD Patients				Control group			
Age in years	Score	Estimation SPM total	Percentile	Age in years	Score	Estimation SPM total	Percentile
69	21		50	58	30	47	75
44	22	34	25	70	33		95
62	33	53	95	82	28		75
75	23		50	66	23		25
73	24		95	60	29	45	75
70	31		90	58	20	32	25
60	24	37	50	62	26	39	50
65	27		25	68	32		50
65	35		25	66	34		25

72	25		95	64	29	45	75
64	23	35	50	68	34		50
50	25	38	25	67	23		50
69	32		75	69	21		75
80	22		50	69	24		90
67	25		50	57	27	41	50
62	15	25	10	43	21	33	10
67	30		50				
69	28		75				
		Mean	54,7			Mean	55,9

Number of observations per percentile

Percentile	PD Group	Control Group
95	3	1
90	1	1
75	2	5
50	7	5
25	4	3
10	1	1

APPENDIX 5B

COWAT

Performance on the COWAT of the PD group and the control group were compared with normative data taken from Miatton et al., (2004) and tested for significance with a t-test. Normative scores of Miatton et al. (2004) were reported per age bin, which we used to compare the performance of out participants with. We therefor obtained different normative scores per group. A summary of these data is presented in the table below. Significant differences between the group performances compared to the normative data ($p<.05$) are signaled with an asterisk. As we did not inquire out participants about their education, it is not certain which of the normative data suits our group best. We therefore report the results for had the participants had less than 12 years of education (left of the dash), and had the participant had more than 12 years of education (right of the dash). For the semantic verbal fluency in naming professions, there no difference between the number of years of education.

Performance on the COWAT task compared to normative data (Miatton et al., 2004), calculated for the age distribution within the respective groups. On the left are the data for an education <12 years, and right of the / are the data for an education >12 years.

	N	A	K	Animals	Professions
PD group	8.8	8.9	11.5	18.9	11.8
Normative for PD group	9.0 /*12.2	9.5 /*12.0	12.1 /*15.5	21.1 /*21.9	*15.9

Control group	11.3	9.3	13.3	19.7	14.4
Normative for Control group	9.5/12.4	9.8 /*12.1	12.4/15.8	21.8/*22.7	*16.6

Performance on the COWAT task of the PD patient group differed significantly from the normative data for an education of more than 12 years, but not for the normative data for an education of less than 12 years. The control group's performance on the COWAT task differed significantly from the normative data for an education of more than 12 years on the semantic fluency categories, and for phonological fluency of the letter A, but not from the other phonological fluency categories, nor from the normative data for an education of less than 12 years.

APPENDIX 5C

Rhyme task

Of the rhyming word pairs 50% has a transparent rhyming orthography (‘bank’ *bank* – ‘mank’ *limp)*, and in 50% of the cases the rhyme is orthographically untransparent (O) (‘krijt’ *crayon* – ‘meid’ *girl*). The word pairs that don’t rhyme, 50% has a different medial vowel (NM) (‘grof’ *coarse* – ‘graaf’ *count*) and 50% has a different final consonant (NF) (‘pen’ *pen* – ‘pek’ *pitch*).

Accuracy and reaction times in the rhyme task. R are rhyming word pairs with transparent orthography ‘lamp – damp’. O are rhyming word pairs with untransparent orthography ‘hoes – blouse’. NM are non-rhyming word pairs with different vowels ‘grof – graaf’. NF are non-rhyming word pairs with different final consonants ‘raam – raaf’.

	Patients				Controls			
	Rhyme		Non-rhyme		Rhyme		Non-rhyme	
	R	O	NM	NF	R	O	NM	NF
Visual								
Accuracy	96.80 (17.64)	66.95 (47.14)	88.76 (31.65)	75.68 (42.99)	98.44 (12.43)	64.25 (48.03)	90.72 (29.01)	78.15 (41.41)
RT’s	2.19 (1.35)	3.16 (1.49)	2.55 (1.47)	2.79 (1.72)	2.08 (0.95)	3.07 (1.34)	2.62 (1.44)	2.76 (1.46)
Auditory								

Accuracy	94.06 (23.69)	82.79 (37.83)	83.96 (36.77)	74.72 (43.54)	95.30 (21.22)	79.02 (40.81)	87.08 (33.61)	87.50 (33.14)
RT's	2.65 (1.10)	2.96 (1.48)	2.85 (1.17)	3.14 (1.40)	2.53 (0.72)	2.83 (0.99)	2.73 (1.14)	2.86 (9.02)

Performance of the two groups on the task was highly comparable: a comparison of the two groups' performance on the accuracy and RT's of the four conditions remained insignificant.

Performance on the visual and auditory rhyme task was compared for all conditions, per group. In the PD patient group there was a significant difference in performance between the accuracy in the untransparent orthography condition of the visual and auditory rhyme task ($t_{I(17)}$= 3.370, p=.004). For the untransparent orthography the auditory rhyme task was performed better compared to the visual rhyme task. The RT's between the two tasks did not differ significantly. None of the other conditions showed a significant difference in performance between the auditory and visual task.

In the control group we get a very similar result; there was a significant difference in performance between both the accuracy and additionally the RT's in the untransparent orthography condition of the visual and auditory rhyme task (Accuracy $t_{I(15)}$= 2.854, p=.012, RT $t_{I(15)}$= 2.122, p=.051). For the untransparent orthography the auditory rhyme task was performed better and faster compared to the visual rhyme task. None of the other conditions showed a significant difference in performance between the auditory and visual task.

Mean accuracy on the rhyme task for PD patients, the control group and normative data

	Visual				Auditory			
	Rhyme		Non-rhyme		Rhyme		Non-rhyme	
	R	O	NM	NF	R	O	NM	NF
Patient Group	96.80	66.95	88.76	75.68	94.06	82.79	83.96	74.72
Control group	98.44	64.25	90.72	78.15	95.30	79.02	87.08	87.50
Normative data	99.67	89.53	96.87	97.67	99.33	94.00	95.33	92.87

APPENDIX 5D

Homophone task

Accuracy on the homophone task for PD patients, the control group and normative data

	Homophone		Non-homophone	
	Word	Non Word	Word	Non-Word
Patient Group	82.42	88.33	85.71	79.83
Control group	76.65	84.26	87.76	81.13
Normative data	94.67	92.86	95.67	90.29

Accuracy and reactiontimes in the homophonetask. W are homophone words 'slib - slip'. NW are homophone non-words 'mucht - mugt'. CW are non-homophone words 'lap – lat'. CNW are non-homophone non-words 'mub - nup'

	Patients		Controls	
	Homophone		Homophone	
	W	NW	W	NW
Accuracy	82.42 (38.14)	88.33 (32.24)	76.65 (42.40)	84.26 (36.59)
RT (s)	2661 (1242)	3060 (1407)	2698 (1239)	2997 (1331)
	Non-homophone		Non-homophone	
	CW	CNW	CW	CNW
Accuracy	85.71 (35.06)	79.83 (40.30)	87.76 (32.84)	81.13 (39.31)
RT (s)	2920 (1270)	3345 (1469)	2713 (1205)	3069 (1202)

APPENDIX 5E

Phoneme monitoring task

Performance on the phoneme monitoring task for PD patients, the control group and participants in the study Özdemir, Roelofs, and Levelt (2007).

	Accuracy			RTs		
	Initial	Medial	Final	Initial	Medial	Final
PD patients	95.08	87.45	87.11	1844	2260	2225
Controls	97.02	95.65	96.12	1795	2099	2089
R, O & L	97.9	92.7	89.5	846	1083	1092

DATA STORAGE FACT SHEET CHAPTER 3

Author: Hanna Gauvin

Date: 02/06/2015

1. Contact

==

1a. Main researcher

- name: Hanna Gauvin
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: Hanna.Gauvin@Ugent.be

1b. Responsible ZAP (if different from the main researcher)

- name: Robert Hartsuiker
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail:Robert.Hartsuiker@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies

===

* Reference of the publication in which the datasets are reported:

Gauvin H. S., Hartsuiker R. J., Huettig F. (2013). Speech monitoring and phonologically-mediated eye gaze in language perception and production: a comparison using printed word eye-tracking. Frontiers in Human Neuroscience, 7, 818.

Gauvin, H. (2015). Verbal Monitoring in production and perception: A cognitive neuroscience approach (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:

All datasets in Frontiers publication, and all datasets reported in Chapter 3 of the doctoral dissertation.

3. Information about the files that have been stored

===

3a. Raw data

----

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

[x] researcher PC

[x] research group file server

[ ] research group file server via DICT

* Who has direct access to the raw data (i.e., without intervention of another person)?

[x] main researcher

[ ] responsible ZAP

[x] all members of the research group

[ ] all members of UGent

[ ] other (specify): ...

3b. Other files

----

* Which other files have been stored?

- [ ] file(s) describing the transition from raw data to reported results. Specify:

- [ ] file(s) containing processed data. Specify:

- [ ] file(s) containing analyses. Specify:

- [ ] files(s) containing information about informed consent. Specify: ...

- [ ] a file specifying legal and ethical provisions. Specify: ...

- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify:

- [ ] other files.

* On which platform are these other files stored?

- [X] individual PC

- [ ] research group file server

- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?

- [X] main researcher

- [ ] responsible ZAP

- [ ] all members of the research group

- [ ] all members of UGent

- [ ] other (specify): ...

4. Reproduction

===

* Have the results been reproduced?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):

DATA STORAGE FACT SHEET CHAPTER 4

Author: Hanna Gauvin

Date: 02/06/2015

1. Contact

===

1a. Main researcher

- name: Hanna Gauvin
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: Hanna.Gauvin@Ugent.be

1b. Responsible ZAP (if different from the main researcher)

- name: Robert Hartsuiker
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail:Robert.Hartsuiker@Ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies

===

* Reference of the publication in which the datasets are reported:

Gauvin, H.S. De Baene, W. Brass, M. & Hartsuiker, R.J. (in press) Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. NeuroImage.

Gauvin, H. (2015). Verbal Monitoring in production and perception: A cognitive neuroscience approach (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:

All datasets in the NeuroImage publication, and all datasets reported in Chapter 4 of the doctoral dissertation.

3. Information about the files that have been stored

===

3a. Raw data

----

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

[X] researcher PC

[X] research group file server

[ ] research group file server via DICT

* Who has direct access to the raw data (i.e., without intervention of another person)?

[X] main researcher

[ ] responsible ZAP

[X] all members of the research group

[ ] all members of UGent

[ ] other (specify): ...

3b. Other files

----

* Which other files have been stored?

- [X] file(s) describing the transition from raw data to reported results. Specify: Matlab files .m in Hanna/Tongbrekers/Functions and Hanna/Tongbrekers/Analyses

- [X] file(s) containing processed data. Specify: Hanna/Tongbrekers/Converted

- [X] file(s) containing analyses. Specify: Matlab files .m in Hanna/Tongbrekers/Functions and Hanna/Tongbrekers/Analyses

- [ ] files(s) containing information about informed consent. Specify: ...

- [x] a file specifying legal and ethical provisions. Specify: Ethical committee approval + screening forms, Word documents

- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify:
- [ ] other files.

* On which platform are these other files stored?
- [X] individual PC
- [ ] research group file server
- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?
- [X] main researcher
- [ ] responsible ZAP
- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction
===

* Have the results been reproduced?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):

DATA STORAGE FACT SHEET CHAPTER 5

Author: Hanna Gauvin

Date: 02/06/2015

1. Contact

===

1a. Main researcher

- name: Hanna Gauvin
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: Hanna.Gauvin@UGent.be

1b. Responsible ZAP (if different from the main researcher)

- name: Robert Hartsuiker
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail:Robert.Hartsuiker@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies

===

* Reference of the publication in which the datasets are reported:

Gauvin, H.(2015). Verbal Monitoring in production and perception: A cognitive neuroscience approach (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:

All datasets in Frontiers publication, and all datasets reported in Chapter 5 of the doctoral dissertation.

3. Information about the files that have been stored

===

3a. Raw data

----

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

[X] researcher PC

[ ] research group file server

[ ] research group file server via DICT

* Who has direct access to the raw data (i.e., without intervention of another person)?

[X] main researcher

[ ] responsible ZAP

[ ] all members of the research group

[ ] all members of UGent

[ ] other (specify): ...

3b. Other files

----

* Which other files have been stored?

- [X] file(s) describing the transition from raw data to reported results. Specify: Excel sheets with transcriptions of recorded errors

- [X] file(s) containing processed data. Specify: Specify: SPSS .sav files

- [X] file(s) containing analyses. Specify: SPSS syntax and output files

- [ ] files(s) containing information about informed consent. Specify: ...

- [x] a file specifying legal and ethical provisions. Specify: ethical committee approval, Word document

- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify:

- [ ] other files.

* On which platform are these other files stored?

- [X] individual PC

- [ ] research group file server

- [ ] other: ...

* Who has direct access to these other files (i.e., without intervention of another person)?

- [X] main researcher

- [ ] responsible ZAP

- [ ] all members of the research group

- [ ] all members of UGent

- [ ] other (specify): ...

4. Reproduction

===

* Have the results been reproduced?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):