Speech-intentions and self-monitoring

Manipulating verbal feedback in a single-word production task

Andreas Lind

MA thesis

Spring 2007

Supervisors: Kenneth Holmqvist (Centre for Language and Literature, Lund University)

& Sverker Sikström (Lund University Cognitive Science)

Abstract

It is commonly assumed that the speech production process is started and guided by a clear conception of what to say, i.e. an intention or a pre-linguistic message. This intention can also function as a standard of accuracy against which actual performance can be measured. But critique against the idea of such a centrally governed process has been offered, and propositions for how a distributed model can account for the assignment of content to speech acts have been given (Dennett, 1991). In this thesis, the role of the auditory feedback of one's own voice in the understanding of the meaning of self-produced speech is investigated. As participants performed a computerized Stroop test while hearing their own voice exclusively through earphones, certain words were covertly recorded. While the feedback of participants' own voices was blocked out, these words were played back later in the test concurrently with participants uttering another, phonetically similar but semantically dissimilar, word. Results show that, while such manipulations were almost always retrospectively detected, a majority of participants reported in post-test interviews that they had experienced confusion as to the actual source of the manipulated feedback, not being certain if it was self- or other-produced. On a minority of manipulated trials, participants acted towards the manipulated feedback as if it was self-produced.

Keywords: speech-intention, verbal self-monitoring, manipulation of verbal feedback, conceptualization, pandemonium model of speech production.

How can I tell what I think till I see what I say?
– E.M. Forster, 'Aspects of the Novel' (1927)

Introduction

As is often held, the faculty of speaking is extremely complex, involving various cognitive- and motor skills which depend on the cooperation and co-ordination of large numbers of cerebral structures the workings of which are only just beginning to be understood with the aid of neuroimaging methods. It is usually supposed that the process of speaking starts with fixed intentions, which are realised as speech through a serial process (e.g. Levelt, 1989). These intentions have the further function of providing a measure of correctness for the actual performance. But the specific nature of these intentions has not been investigated empirically to a high degree. This is, no doubt, due in part to the great difficulty of investigating such a subjective aspect of this cognitive process. But it may also be related to the fact that such an inquiry has not been necessary within the framework of existing theories as these theories are mainly interested in what happens *after* intentions are fixed. In this vein, Dennett (1991) offers some reasonable critique against viewing speaking as a centrally governed activity. Bearing some relevant aspects of this critique in mind, in this thesis we attempt to arrive at some novel means of empirically investigating the aspect of speech production which is generally called *conceptualization*.

Speech production – from intention to articulation?

The three main areas of psycholinguistic research are language comprehension, language acquisition and language production. Up until the early 1990s there was no major coherent account of speech production. Language comprehension and acquisition, on the other hand, were the focus of much research (Dell, 1986; Levelt 1989). There *was* a quite voluminous research on production, primarily using speech-errors and naming latencies (Levelt, 1999a), but this was scattered across a variety of disciplines such as conversational analysis, pragmatics, discourse semantics, phonetics and artificial intelligence. Recognizing this lack of focus on production, Levelt (1989) attempted to summarize insights from this research into a coherent account of speech production. What emerged was a highly structured model with several components deemed necessary for speech (see fig. 1). First, a preverbal message is formed in a processing system named *the conceptualizer*[1]. The formation of the preverbal message is dependent on a

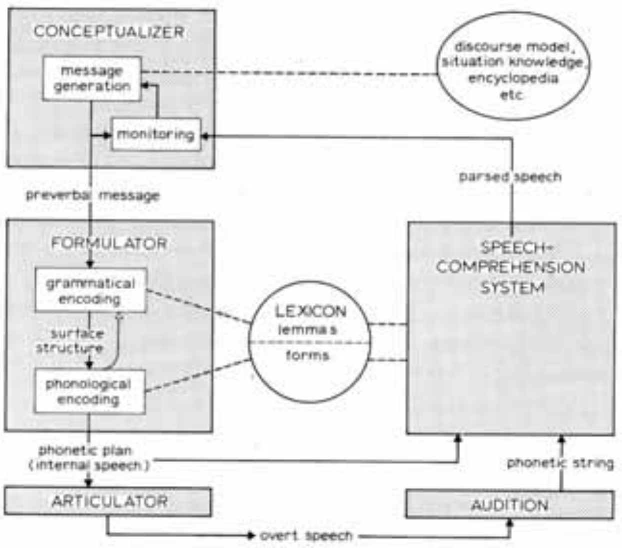

Figure 1. Flowchart model of speech production (from Levelt 1989)

[*] The present study was conducted within the framework of a larger project carried out by a research team made up of (apart from the author) Petter Johansson, Lars Hall and Sverker Sikström. Thank you to the following: Kenneth Holmqvist and Sverker Sikström, who excellently supervised my work on this thesis; Anders Holmberg, Johan Blomberg and Stefan Lindgren. Also, thank you to all the participants of the thesis-seminars, led by Mr. Holmqvist. Finally, thank you to Daniel Dennett and the organizers and participants of the workshop 'Dennett's thinking: Aspects and implications', University of Copenhagen, May 10th, 2007, where the present project was presented and discussed.

[1] In Levelt, Roelofs and Meyer (1999) this is framed in terms of processing as "conceptual preparation".

number of things such as procedural and declarative knowledge, situation knowledge, a discourse model etc. and it involves, among other things, forming an intention, selecting and ordering relevant information to express this intention and keeping track of all the things that have been said before in the conversation. All this work is performed by the conceptualizer. The preverbal message is then the input for the system called *the formulator*, which translates the conceptual structure of the preverbal message into a grammatical and phonological code or plan. This plan is sent to *the articulator*, which produces audible speech by way of the articulatory apparatus. *Self-monitoring* is active both at the level of internal speech and the level of overt speech. Using the same capacities that handle comprehension in general, the speaker listens to him- or herself talking in much the same way s/he listens to others talking, with the difference that internal speech is also monitored.[2]

While differing in details from other models of production there is an important point of agreement between Levelt's model and other accounts of production (e.g. Dell, 1986; Fromkin, 1971) in that Levelt does not speculate much about what happens before the content of a message to be conveyed is determined. While he grants us a capacity for conceptualization, he does not see it as his task to uncover what processes support it. For all intents and purposes, the conceptualizer is what gets everything started. Levelt himself remarked that the conceptualizer was a simplification that certainly would have to be investigated further (1989, p. 9). While stating that "the mother of each speech act is a communicative intention" (ibid., p. 108) he also made clear that "where intentions come from is not a concern of this book" (ibid., p. 59).[3] Still, the conceptualizer was allotted its place in the model and in many ways it has stayed in that place without being, to any large extent, investigated further.

Avoiding this question of, as Dell has put it (1986), *why* a speaker says what is said, and focussing instead on *how* it is said, can be seen as a rather conscious and pragmatic strategy. It clearly is motivated by the extreme difficulty of trying to get a grip of the processes involved in specifying the content of speech-acts. Such avoidance allows the researcher to focus on aspects of production where there is either existing empirical evidence or the potential for quantitative predictions. However, the result of this avoidance has been that the delegation of the process of conceptualization to a conceptualizer has been quite solidified and is not really questioned. Levelt's model has been very influential and many subsequent psycholinguistic models bears resemblance to it. Also, many *accounts* of speech production begin with simply stating that all speaking starts with conceptualization, and most accounts do so without really defining what goes into the conceptualization or what mechanisms support it (see e.g. Griffin, 2004; Levelt, 1999b; Levelt, Roelofs, & Meyer, 1999; Meyer, 1992; Postma, 2000 and textbooks such as Carroll, 1999 & Harley, 2001).

Dennett's pandemonium model of speech production

Positing *central executives* in psychological models is not unanimously endorsed. Dennett is one of the critics of this practice (1987, 1991). He warns against relying too much on folk-psychological concepts, such as *intention*, when building scientific models of behaviour (1987) and he specifically questions Levelt's reliance on the conceptualizer as both the point of origin and supervisor of the whole process of word-production (1991). In Levelt's model the conceptualizer produces finished intentions that guide the whole speech-production process in a serial manner. But, Dennett cautions, postulating a "thought-*thinker*

[2] The nature of 'inner speech' is controversial. Levelt (1989) suggests it is the equivalent of the phonetic plan sent from the formulator to the articulator.

[3] Furthermore, in response to a review of Levelt (1989), Levelt (1992b) notes, quite justified of course, while referring to the history of the western intellectual tradition, that the lack of a detailed elucidation of the nature of (speech) intentions is hardly unique to him.

[that] begins with a determinate thought to be expressed" only moves the problem back a step (ibid., p. 241, emphasis in original). If we want to know how the content of speech-intentions is actually specified or fixed, we must not be tempted to leave the *how* up to this "thought-*thinker*" (ibid.).

Dennett (1991) offers some interesting suggestions for how the notion of a conceptualizer can be revised in such a way as to open up for the possibility of empirically investigating the mechanisms specifying or fixing the content of speech-acts. He proposes a distributed, or pandemonium, model of production where the content of speech is opportunistically determined in a "cobbled-together collection of specialist brain circuits" that have evolved independently, but now function jointly to much greater power (ibid.). Within the conglomeration of these brain circuits, which Dennett nicknames *the Joycean machine*, recalling the stream-of-consciousness style of writing of novelist James Joyce, there is a competitive process where no one circuit is ever in control for long. Rather, power is shifted around in a "quasi-evolutionary process" where the current mind-set of the speaker functions as a constraining mechanism that "judges" all the "suggestions" for speech-acts that float around in the Joycean machine at a specific time.[4] In effect, several different variants of a proposed utterance are, often non-consciously but sometimes consciously, weighed against each other in a "collaboration [] of various subsystems none of which is capable on its own of performing – or ordering – a speech act" (ibid., p. 239). Eventually one of the proposed utterances gains strength and is spoken.[5] The weighing is carried out on a scale of microseconds and, importantly, this means that *the content of the speech act is not fully specified or fixed until it is actually spoken* (ibid.). If the situation demands it, if we are hard pressed not to say the wrong thing, we can carry out a silent, conscious rehearsal of some of the propositions (presumably this is done in the phonetic loop of Levelt's model). When speaking casually, however, there is no time for this and we learn the specific content of our speech act at the same time as our listener (ibid.). Furthermore, what was actually spoken can, probably as a result of interpretative mechanisms similar to the ones used to understand other people, work as a ratification of what we actually *meant*, and this can be further reacted to and acted upon in a continuous process of conceptualization that, therefore, is not necessarily over once an utterance has been pronounced (ibid.).

In effect, Dennett (1991) does not in any way deny the existence of a process of conceptualization, only the existence of an "*inner* conceptualizer that is a proper part of the language-producing system" (ibid., p. 251). Rather, he postulates a *global* conceptualizer which is equivalent to the whole person that does the uttering and "of which the language-producing system is itself a proper part“ (ibid., p. 251).

Inspiration for a methodology

To sum up, despite being perhaps the most intriguing challenge when building models of speech production, there is a large need of finding new methods for how to test different aspects of the process of conceptualization. Dennett's pandemonium model, while providing enthusiastic incentive to do just this, has not received much interest in studies of speech production (though see Levelt et al., 1999 for a brief mention) and attempts at trying to test it empirically have been scant.[6]

[4] Dennett defines "mind-set" as "well developed meta-habits". The vagueness of this definition, then, is Dennett's own.

[5] Dennett's model parallels models of action selection which, as Gazzaniga, Ivry and Mangun (2002, p. 489-90) writes, proposes that "processing in the cortical motor areas can be viewed as a competitive process in which candidate actions compete for control of the motor apparatus" (see also Rosenbaum, 1991).

[6] This applies to Dennett himself as well. Apart from Dennett (1991), where the pandemonium model is presented, he has not really developed it more specifically.

The contrast between a Levelt type serial model and Dennett's pandemonium model can serve as guidance when attempting to make meaningful studies. When trying to find inspiration for new methodologies to use we can also look at some other domains of inquiry. For example, some useful pointers can be obtained from experiments that make use of covert manipulations of stimuli to investigate the mechanisms behind the self-ascription of actions and intentions. Studying volition, Wegner (2002; Wegner & Wheatley, 1999) argues that our sense of will is the result of an interpretative mechanism coming out of the enormously complex and largely non-conscious processes underlying action. This mechanism, according to Wegner, depends on an *apparent* causal relationship between thought and action. Though claims such as these, involving highly subjective aspects of experience, are notoriously difficult to study empirically, Wegner's theory finds empirical support in several studies. For example, Nielsen (1963) made participants experience that they were watching their own hand perform involuntary movements. Seeing their gloved hand inside a box through a tube on top of the box, participants were told to copy a pre-printed line with a pencil and to stop copying as the light inside the box was turned off. With the aid of a mirror apparatus, a confederate's hand was sometimes presented as the participant's hand, purposely drawing slightly off-target and instilling the experience of involuntary movements (see also e.g. Fourneret & Jeannerod, 1998; Sørensen, 2005; Gallagher & Sørensen, 2006). Employing a similar experimental set-up, but this time operating in the auditory modality, Nielsen, Praetorius, and Kuschel (1965) had skilled female singers seated in a room with sound-absorbing walls. While hearing their voice exclusively through earphones they were instructed to hold a note fed by a sine wave generator. On specific trials the auditory feedback was interrupted by a brief noise period after which a confederate's singing voice was fed through the earphones, blocking out the participant's own voice, and continuously falling or rising in pitch. Participants made compensatory rises/falls in pitch, but expressed in post-experiment interviews that they felt it was their own voice they heard and that it was beyond their voluntary control.[7]

Additional indications that we sometimes fail to detect mismatches between intention and outcome come from the so called choice blindness-studies. Johansson, Hall, Sikström, and Olsson (2005) presented participants with pairs of pictures of female faces and asked them to choose the face they found the most attractive. After this the cards were laid down on the table and the chosen card was slid to the participant who picked it up and was asked to motivate his or her choice. By way of a double-card manipulation, the participant was sometimes presented with the opposite of what s/he actually chose. Only 26% of these manipulations were detected. The participants otherwise accepted the manipulated picture as their actual choice and were able to provide convincing verbal motivations for making it (Lind, 2006; Johansson, Hall, Sikström, Tärning, & Lind, 2006). By having participants taste jam and smell tea in a similar experimental set-up, Hall, Johansson, Tärning, Deutgen, and Sikström (2006) showed that the effect applies to the modalities of taste and smell as well, with the results being similar to those of Johansson et al. (2005). Counting all types of

[7] On an interesting note, Nielsen et al. (1965) mentions some preliminary results in passing at the very end of their article. Employing the same apparatus as used in the pitch-manipulation experiment, they instructed participants to "recite a verse line with a certain rhythm" (ibid., p. 208). As the participant did this, an assistant covertly exchanged a word or a syllable using his/her own voice. While very different in meaning, these words or syllables were only slightly different in pronunciation from the words the participant actually uttered. The results showed that participants sometimes were confused as to the source of the manipulated words or syllables, attributing them to him/herself. Interestingly, the manipulations also often had an effect on participants' recitation of the verse with respect to rhythm or pronunciation (the authors are not more specific than that on this point). The results they present can not be taken as reliable; for example, the specifics of their method are not presented. Efforts to find out if the authors, jointly or individually, published the results of this or a similar study at a later date have been in vain.

detection, 33.3% of the manipulated jam trials, and 32,2% of the manipulated tea trials were detected. Again, motivations for the “choices” were readily provided.

The experiments recounted above all used covert manipulation of stimuli in order to “pry apart” intentions and feedback, and thereby investigate the role of feedback in the participants’ interpretations of the origins or nature of their intentions. Taking inspiration from this strategy, we now return to speech. But before proceeding to a full description of our experiment, a few words about verbal self-monitoring are in place.

There lies an inherent difficulty in discussing the voluminous research on monitoring and feedback loops (of which Postma, 2000 provide a review) without presupposing pre-linguistic messages, because the literature on monitoring usually takes Levelt’s conceptualizer (or something similar) as the unquestioned starting point of speech. Therefore, a lot of empirical evidence has been accumulated over the years in support of a Levelt-type model, but this evidence is very dependent on the interpretation of speech production that Levelt put forth. In the model of Levelt (1989, 1999b; Levelt et al., 1999; Postma, 2000), all monitoring loops go back to the conceptualizer and the grammatical, phonetic, syntactic, articulatory, motoric etc. levels can not influence conceptualization[8]. If they did, it seems, the whole monitoring system would be put out of balance. For Dennett (1991), as we have seen, such influence is explicitly implied in the notion of a global conceptualizer.[9]

It is our hope that the present project will open up for some new possibilities and show how we can study the role of feedback and self-monitoring in the process of conceptualization without delegating the role of conceptualization to a conceptualizer that is independent of the rest of the speech production-system. It seems that if we modify the idea of a central conceptualizer, then we can start thinking about what mechanisms support the processes of conceptualization. In a sense, this means taking a step towards the “further explanation” of the reification tagged conceptualization that Levelt himself called for (1989, p. 9).

This experiment

This study aimed to investigate the role of the auditory feedback of our own voice in the process of conceptualization. Our approach to doing this was by manipulating the auditory feedback that participants receive so that they pronounce one word but concurrently hear themselves say another word. If the process of conceptualization is not to be considered to be finished until after the utterance has actually been spoken (Dennett, 1991), then it is not certain that such manipulations will be noticed by participants. If that is the case, the influence of the manipulated feedback on participants understanding of what they have said should be possible to study.

For a believable voice-exchange to occur, we must know both when the participants will say the words we want to record and when they will say the words we want to exchange with the recorded ones. We also need to know that the recorded word will match the exchanged word fairly well in delivery. Taking this into account, along with the sheer

[8] Although Levelt (1989) speculates on how such influence *could* work (see also Levelt, 1992b).

[9] Furthermore, in a hierarchical model such as Levelt’s (1989), there is not supposed to be any interaction between lexical selection and phonological and grammatical encoding. However, two robust findings question this assumption. The first is the *mixed error effect*, or the observation that semantic errors in speech are statistically more often phonetically similar than not (e.g. Dell & Reich, 1981; Martin, Gagnon, Schwartz, Dell, & Saffran, 1996; Slevc & Ferreira, 2006). The second is the *semantic bias effect*, i.e. that when phonological errors are made, the outcome is more often than not an actual word (e.g. Baars, Motley, & MacKay, 1975). Both the mixed error effect and the semantic bias effect have been taken as indications that the localist type models are too strict in their denial of interaction between the different processes (e.g. the conceptual/lexical and phonological levels) (e.g. Levelt, 1992a). For a thorough comparison between non-interactive and interactive models, see Rapp and Goldrich (2000).

technical difficulty of performing the type of manipulations we wished to perform, the Stroop test (Stroop, 1935) was chosen for implementing the present experiment. In the Stroop test you are shown colour-words; the letters of these colour-words have a specific colour that does not always match the written word. The task is to always name the colour of the letters, while ignoring the spelled word. The Stroop test has quite fixed rules and, apart from when people make mistakes in it, it is fairly easy to predict what they will say and when they will say it. They can also be instructed to use the same tone of voice throughout the test, ensuring that voice quality will not differ dramatically between what they actually say and what they hear themselves say on manipulated trials. The Stroop test is, in itself, a test of some difficulty, and the participants cognitive resources are somewhat taxed (MacLeod, 1991). Furthermore, Stroop himself quoted some interesting speculations made close to a century ago by Woodworth and Wells:

> The real mechanism here may very well be the mutual interference of the five names, all of which, from immediately preceding use, are ‘on the tip of the tongue,’ all are equally ready and likely to get in one another’s way (Stroop, 1935, p. 465).

Should this be the case, this would mean that, during the Stroop test, the five words are all highly activated and ready to be pronounced, and this in turn should make them as good candidates as any for being exchanged.[10]

Questions. An open question in this study was *how will participants react to the manipulations?* Will they immediately know that the feedback did not match what they said, or what they *meant* to say (as would be predicted in Levelt’s model)? Or will they take the manipulated feedback into consideration in the, according to Dennett, split-second process of conceptualization and perhaps adjust their understanding of what the *meaning* of their utterance was/is?

In order to attempt to distinguish between these two possibilities, we will instruct participants to signal each time they make a mistake in the test and they are to do this simply by saying “wrong!”. The signalling of detection of a mistake during a manipulated trial will be considered a measure showing that the manipulated feedback exercised influence on the process of conceptualization and therefore on the meaning eventually ascribed to the utterance by the participant. We also wanted to probe the experiential aspects of the voice exchange and this was done by asking the participants a series of questions directly after the experiment.

Method

Participants

Forty-six participants (23 females) were drawn from a student population. Their mean age was 23.5 years (sd 2.4). Only native speakers of Swedish were included, and none of the participants had any auditory or visual deficits. All were unaware of the actual purpose of the experiment, but were fully informed afterwards.

Stimulus materials

A computerized version of the Stroop test was employed (see e.g. Linnman, Carlbring, Åhman, Andersson, & Andersson, 2006). The colours blue, green, grey, red and brown were

[10] Also, Levelt (1983) showed, although not directly in relation to the Stroop task, that the use of erroneous colour words were a particularly prominent error made (see also Levelt, 1989).

used, all written in their Swedish counterparts.[11] The colours were modified so that the grey and green colours and the grey and blue colours were not too dissimilar, while still retaining their distinctiveness[12]. The use of standard colours was judged to increase the likelihood of participants getting too strong visual cues about what they had actually said. All words were preceded by 4 fixation points (****) and were presented one by one in lower case letters within a fixation-square centred on the computer screen. The words were shown for 200 ms and the interval between words was 1500 ms. The test consisted of 50 trials. Three of these trials were manipulated so that the participants said one word but, through the earphones, heard themselves say another word.

Trial sequence. To provide a basis for the construction of the test, the 25 possible word/colour-combinations were randomly distributed in the trial sequence, each combination appearing twice. It follows that 10 of the trials were in a congruent condition, i.e. where the colour of the letters matches the written word. Including congruent trials in the test has been established as a potent way of preventing participants from adopting strategies (such as peering at the words) to exclude the semantic information, and to make them attend to both dimensions (see MacLeod, 1991). As a further measure against participants adopting such strategies, five fruit and-vegetable words were inserted and a condition where participants are to pronounce the fruit- or vegetable words whenever they appear, at the same time ignoring the colour of the letters, was included. The fruit and-vegetable words (we used *apple*, *pear* and *tomato*) were inserted into the randomized sequence so that the colour would match the actual colour of the fruit or vegetable (i.e. *apple-green* etc.). Two of these words came before the first manipulated trial, and then one between each manipulation and finally one after the last manipulation.

Since we did not want the manipulated trials to stand out simply on the basis of them being mistakes, as the participant may experience them to be, it was important to make sure that the general mistake frequency was not at a minimum during the non-manipulated part of the test. In an attempt to raise the number of mistakes made by the participant, *the distractor suppression effect* (Neill, 1977) was introduced. If the to-be-suppressed, spelled word on trial *n*-1 is the same as the to-be-named colour on trial *n*, then this will increase the interference on trial *n* (MacLeod, 1991), possibly, though not certainly, also increasing the chance of mistakes. Two such sequences were among the first 20 trials and then there was one between each M. Also, we did not include any manipulated trials during the first 20 trials in order to allow the participant to make a few self-generated mistakes before the manipulations.

Finally, the manipulated trials were inserted into the randomized sequence. They were placed on trials 21, 33 and 45. The colour/word-combinations for the manipulated trials were *blue/grey*, *green/grey* and *grey/blue*. This decision was made based on phonological similarity. In Swedish, *blue* is pronounced [bloː], *grey* is pronounced either [ɡɹoː] or [ɡʁoː], depending on dialect, and *green* is, similarly, pronounced either [ɡɹøːn] or [ɡʁøːn]. The written word always matched the manipulated feedback, thus giving participants a clue as to why they may have made a mistake. In order to place these combinations at the indicated trials, some minimal rearranging was made in the randomized trial sequence, while keeping the number of each colour/word-combination intact. The same trial sequence was used for all participants.

Pre-test. A pre-test of 15-18 trials preceded the actual test. Here, all types of trials, save the distractor suppression effect-trials, were represented (i.e. the incongruent, congruent and

[11] In keeping with the colour-word version of the Stroop test it is common practice to include between two and five stimuli. Further increases in set size has been shown to increase the time for naming the colours (MacLeod, 1991), something that is neither necessary nor desirable for the present study.

[12] The colour codes were, respectively, [.3 .3 .8], [.4 .7 .5], [.5 .5 .6], [.9 .0 .0] and [.5 .3 .0].

fruit/vegetable-trials). While this pre-test was disguised as a joint practice- and calibration session, its main purpose was to provide a chance to record the three words that were later used during the manipulated trials. Given that the interval between the presentation of words to the participant was not required to be constant, we were free to edit or re-record the recorded words, should they be edited unsatisfactorily by the computer-program.

Added noise. Due to the nature of the program used to run the experiment, a low buzzing that was present in the auditory feedback on all non-manipulated trials disappeared during manipulated trials. During pilot testing, several participants remarked on this as a prime reason for detecting the manipulations. Therefore, this difference was masked by playing back a slightly louder homogenous buzzing from a separate Mp3-player through the duration of the test.

Apparatus. The Stroop test was presented on a 17" monitor connected to a Hewlett Packard Compaq nc6120 laptop, on which the program was run. The computer, in addition to an Avant MP510 Mp3-player on which the Mp3-file of recorded buzzing was constantly playing, was connected to a Velleman PROMIX 100 mixer. Leading from the mixer were two sets of earphones (one set for the participant and one for the experimenter) and a cord connecting it to a Marantz PMD660 portable recording device able to record on two separate channels. Also connected to the Marantz PMD660 was a stationary table-microphone placed in front of the stimulus-monitor.

Procedure

All instructions were given following a manuscript to ensure that they did not vary between participants. Prior to the experiment, participants were told that they could stop the experiment at any time.

The Stroop test. The participant was seated in front of the stimulus screen and the experiment was presented as a common Stroop test, i.e. s/he was told that colour-words would appear one by one on the screen and that the letters of these words would have a specific colour that did not always match the actual word. The task was always to name the colour of the letters. Participants were also told that, on a small number of trials, fruit- or vegetable-words would appear and that when they did, they were to read the word rather than say its colour. Furthermore, they were told that it was important that they speak in a similar voice and volume during the whole experiment and to pronounce words in their primary forms as 'blå', 'grå', 'röd' etc., rather than 'blått', 'grått', 'rött' etc.[13] During the pre-test, the participant did not wear headphones but solely spoke into the microphone.

After the pre-test, participants were, in addition to the microphone, equipped with earphones and hearing protection. The hearing protection was used to minimise the natural feedback of their own voice.[14] They were also told that it is very common for people to make mistakes in this test, and that it was absolutely fine if they did so. However, they were instructed to signal in case they noticed themselves making a mistake; this they were told to do by saying "wrong!" as quickly as possible before moving on to the next word.

[13] These latter forms are used in Swedish when the colour-word (functioning in such a case as an adjective) grammatically agrees with the noun that it modifies. These forms are sometimes also used to *name* colours, and it was necessary to ensure that participants did not use the two possible forms interchangeably.

[14] Ideal in this experiment would be to, in addition, equip the participants with bone conductance headphones. These are attached somewhere on the skull and sends a sound signal that sets the skull in vibration, thereby by-passing the outer ear and sending signals directly to the auditory nerve. As the vibration of the skull plays a large part in hearing ones own voice (see e.g. Christoffels, Formisano, and Schiller, 2006) bone conductance headphones would provide a way of simulating the natural resonance of one's own voice setting the skull in vibration. Unfortunately, such headphones were, due to budgetary constraints, not possible to obtain for this project.

Post-test interviews. Directly following the experiment, participants were first asked simply what they thought of the experiment. They were then asked if they had made any mistakes. To indicate that something might have been wrong during the experiment, they were then asked if they had noticed anything strange. If the participant either spontaneously directly after the experiment *or* when s/he was asked what s/he thought about the experiment *or* when asked if s/he had noted anything strange stated that s/he had noticed the manipulation(s), s/he was first asked how many times s/he had noticed them. Then s/he was asked how s/he had experienced the manipulations, keeping this question neutral. After this, follow-up questions were posed; if the participant stated that it was confusing and that s/he had initially been confused as to whether or not s/he had said the manipulated word, s/he was asked to describe in detail how s/he felt during manipulations, and also if this description applied to all the manipulated trials or only one or two of them. Finally, participants were asked if they had become suspicious after the first manipulation. Before leaving, all participants gave their written consent.

Data analysis

Each experiment, complete with post-test interviews, was recorded and organized in separate audio-files. Each audio-file has recordings on two channels. One channel (C1) contains a recording of the experiment as it was actually performed by the participant, and the other channel (C2) contains a recording of the audio input that the participant received, including the manipulated feedback and the added noise. Two independent analyses were performed, and in the few occurrences where opinions differed, these differences were settled through discussion.

Analysis of the audio-files was carried out in Audacity 1.2.6. and SoundForge by listening first to C1 in order to, firstly, make certain that participants did not make any mistakes on their own during manipulated trials, which in itself would motivate them to vocalize “wrong!” after such a trial and, secondly, to count the number of mistakes made during the whole test. Then, C1 and C2 were analyzed jointly in order to scrutinize if the manipulated feedback matched the actual vocal performances made by the participant. Comparisons were made first based on word onset. The manipulated feedback was allowed to have its onset no more than 100ms (+/-20ms) after onset of the actually spoken word. If onset fell within this range, subsequent comparisons were made based on strength, length and pitch of the relevant words. The average difference in onset was 46.4 ms. Due to an unidentified problem with the computer program, feedback to certain participants was, after a small amount of manipulated trials, shut off for between one to three non-manipulated trials. These trials, and all subsequent trials for the specific participants for whom this happened were removed from further analysis. All in all, 20 participants were removed due to differences in onset, strength, length or pitch or problems with disappearing feedback.[15] Among the remaining 26 participants, another 14 individual manipulated trials were removed from analysis due to either too large differences in onset or due to the problem of the disappearing feedback. 10 of these were manipulated trials number 2 and 3 of specific participants. The remaining 4 trials were the third manipulated trial of specific participants. For the remaining 64 manipulated trials, all trials followed by a response that indicates that the manipulated feedback played a role in participants understanding of the meaning of their utterance were noted. Such responses included, but could not be limited to, “wrong!”-pronouncements; also falling under this description was repetitions of the actually uttered

[15] It should be mentioned that the instruction to speak in as similar a voice as possible throughout the test was followed by most participants. This, of course, is no guarantee that the non-manipulated voice and the manipulated one will match well enough.

word (which can be interpreted as a correction when compared to the manipulated feedback) and exclamations of “no!”. Finally, all trials where participants signalled a detection showing they understood the manipulated feedback was externally produced were labelled as *concurrent detection*.

Next, the post-experimental interviews of the remaining participants were analyzed. Here, each participant was tagged in a separate excel-file for a number of characteristics: they were assigned the label *retrospectively claimed detection* if they either immediately after the experiment or when asked what they thought of the experiment said that they had noticed that the feedback had been manipulated. Based on their answer to the question of how they had experienced the manipulations they were tagged as either *confusion* or *certain-that-manipulation-was-external*. It was noted if they claimed to have become more suspicious after the first manipulated trial.

Results

The results are presented in two sections. The first section deals with the direct responses elicited during the actual test. The second section deals with the experiential aspects of the test and, specifically, of the manipulations which we attempted to probe in the post test interviews. Table 1 provides a summary of the results.

The Stroop test

A total of 11 manipulated trials of 6 different participants were followed by either “wrong!”, “no!” or a repetition. No actual mistakes were made during these trials.

No participants signalled detection of the feedback having been manipulated during the experiment. All participants who detected the manipulations waited until the post-experiment interviews to indicate that they had noticed something being wrong.

Across all participants, out of 1222 non-manipulated trials, 34 mistakes were made. This corresponds to 2.78%.

Post-experiment interviews

All but one participant said they had detected that the feedback had been manipulated. This one participant also responded to all three manipulations with repetitions (2 times) and a “no!”. The remaining 25 participants suggested, when asked how many times they had detected that the feedback had been manipulated, that between 2 to 5 trials had been manipulated. None were surprised to find out the exact number. The follow-up questions revealed that 18 participants had experienced a feeling of confusion as to the actual source of the utterance during manipulated trials. This confusion was present on a total of 32 manipulated trials and was described as a short bewilderment followed by a “questioning” to the effect of “did I say that!?”, after which the participants “decided” they had *not* produced the manipulated feedback. There was some diversity regarding the number of trials on which the participants claimed that this confusion had been present. This diversity is presented in a subsection to the *confusion*-measure in table 1.

14 participants said they had become suspicious after the first manipulated trial. 11 participants, on a total of 22 manipulated trials, claimed to have been absolutely certain that the feedback had been manipulated (these participants make up the category *certain*). They claimed to have ascribed the manipulations to either the computer or the experimenter.[16] 4 of these participants also fall within the category of *confusion*. In these cases this is because they felt a confusion on the first manipulated trial but then became more suspicious and felt certain that the two remaining manipulations were made by the apparatus or experimenter.

[16] In the context of this experiment, this amounts to the same thing.

Table 1.

		Nr. of NM trials	Nr. of M trials	Nr. of participants
The Stroop test	*"Wrong!", "no!" or repetition*	-	11 (17.2)	6 (23.1)
	Concurrent detection	-	0	0
	Mistakes	34 (2.8)	1 (1.6)	18 (69.2)
Post-test interviews	*Retrospectively claimed detection*	-	61 (95.3)	25 (96.2)
	Confusion	-	32 (50)	18 (69.2)
	Nr. of trials where confusion was claimed to have been experienced	-	*3*	*6 (23.1)*
		-	*2*	*2 (7.7)*
		-	*1*	*10 (38.5)*
	Certain	-	22 (34.4)	11 (42.3)
	Suspicious after first M	-	27 (42.2)	14 (53.8)
	Total	1222	64	26

Note: NM refers to non-manipulated trials, M refers to manipulated trials. % in parenthesis.

Discussion

The present study employed a computerized voice-exchange set-up in order to manipulate the auditory feedback in a Stroop test so that participants said one word but concurrently heard themselves say another, phonetically similar but semantically dissimilar, word. By instructing participants to signal whenever they made mistakes in the test, we attempted to measure the effect of the manipulated feedback upon the participants' process of conceptualization. The most important aspects of the results are briefly recounted here. 6 participants, on a total of 11 trials, reacted to the manipulated feedback by saying either "wrong!", "no!" or by repeating themselves. In the post-experiment interviews, 25 out of the 26 participants, i.e. an overwhelming majority, reported that they had detected that the feedback was manipulated on certain trials. At the same time, 18 of the 26 participants also reported that they had initially been very confused as to the source of the manipulated feedback.

Two obvious and related questions that arise out of these results regards (1) the significance of the fact that certain participants behaved, on one or more occasions, towards the manipulated feedback as if it was themselves who had produced it, and (2) the significance of the state of confusion during one or more manipulated trials that was described by a majority of participants. Even though all but one of the participants who acted towards the manipulated feedback as if it was self-produced later reported that they had detected that the feedback had been manipulated, it is important to stress that immediate reaction and retrospective reasoning must be considered separately. Our access to, as well as our ability to reason about, our own higher cognitive processes has repeatedly been shown to be fallible (Nisbett & Wilson, 1977; Johansson et al., 2005; Johansson et al., 2006; Lind, 2006).

We can see, then, how the manipulated feedback clearly influenced participants understanding of what they had said on 17.2% of all manipulated trials. This is a very interesting finding. It shows that we do rely, to some extent, on the feedback of our own voice when making inferences of what we have said. Furthermore, the observation of this influence invites to speculation about its possible relation to the reported state of confusion, which was more widespread. Accepting the manipulated feedback as self-produced is not certain to be a matter of *either-or*. Rather, there could be a continuum between, on the one hand, the state of confusion and, on the other hand, a point where the feeling of confusion gains enough strength to provoke the participant to act upon the manipulated feedback *as if* it was self-produced, only to, sometimes at least, directly afterwards (or after some time) reinterpret the situation (and the meaning of his/her utterance) and decide that "something was wrong". This interpretation is, it seems, in line with Dennett's (1991) pandemonium model of speech production.

Another interpretation of the state of confusion could be that, while we may, as Levelt's (1989) model proposes, be endowed with quite strong speech-intentions, it is the *situation itself* that provokes the confusion. After all, it is *very* rare for us to say one thing but hear ourselves say another thing. So confusion may follow, and participants may then, more or less consciously (after all they are simultaneously engaged in a quite difficult task), start eliminating the possible explanations to what just happened. After a while, and perhaps encouraged by the subsequent two manipulations, they come up with the most plausible explanation: the feedback must have been manipulated. This explanation of the state of confusion, however, can not explain why it is also observed that participants act upon the manipulated feedback as if it was self-produced. How could we act against a speech-intention that has set the whole speech-production system in motion and against which the output is supposed to be judged?

Future studies

The results of this study calls for further testing using the same basic method. The next two sections of this discussion will be devoted to discussing, firstly, some issues of ecological validity to be considered for future experiments and, secondly, some technical aspects of the experiment. The Stroop test of the present set-up will be used as a point of reference and discussing these issues also serves as a complement to the discussion of the results of the present study. As mentioned before, the Stroop test was chosen because it afforded an opportunity to control what participants will say and when they will say it. However, it can be argued that a more ideal situation to make the proposed manipulations would be in a highly ecologically valid speech situation, such as a normal conversation between two or more people. For example, Pickering and Garrod (2004) argue that dialogue should be considered the basic form of language use. Conversations often involve situations where speech flows quite fast and there is not time to internally go over the formulation. Seen this way, the present experiment can be viewed as a compromise between an ideal experiment for testing the role of the auditory feedback of our own voice in the process of conceptualization and the technical requirements that needed to be met. We believe it is important to discuss the implications of this compromise for the present results. Furthermore, the technical difficulties of the experiment as such were of course the reason why such a high number of participants had to be excluded from analysis.

Some notes on ecological validity. It is important not to make the appearance of the manipulated feedback seem inappropriate to the situation in which the experiment is carried out. When considering the Stroop test, for example, it is probably of some importance that the number of self-produced mistakes was very low during the test. Had the mistake-frequency been higher, say between 15-25% instead of the observed 2.8%, it would probably not have made the manipulated trials stand out the way we may assume they did, at least not because they were "mistakes".

The speed with which participants are speaking should not be too artificial. As mentioned before, it has been argued that the basic form of language use is dialogue and in dialogue the processes of production usually operates at a much faster pace than it does in, for example, our Stroop test.[17]

It has previously been shown that, when speaking, we do not monitor *all* aspects of our speech *all* the time (Levelt, 1989). Neither do we always pay the same general amount of

[17] In an attempt to increase the speed with which decisions about what to say must be made (and, potentially, thereby also increasing the general mistake-frequency), six participants (outside the test reported in this thesis) were tested in an identical Stroop task, but with a shorter interval between words. Two participants were tested with an interval of 1000 ms and four participants were tested with an interval of 800 ms. This, however, did not yield different results in any regard.

attention on monitoring our output (ibid.). Rather, our attention on monitoring fluctuates depending on our needs. For future experiments, it is important to strike up a balance between too little monitoring, and too much monitoring. Clearly, participants have to monitor their speech enough to detect mistakes (or whatever criteria the experimental design uses as a way of measuring participant's reactions to the manipulated feedback). At the same time they should not monitor their speech to an unnaturally high degree, as this would test the research question only under the specific circumstances that such a high degree of self-monitoring would entail. For the present experiment, it is difficult to evaluate the precise degree of monitoring that the instruction to detect and signal self-produced mistakes actually results in. But, no doubt, this is a very important factor to take under consideration when designing new experiments.

In common, everyday speech, the incentive for what to say is only very rarely displayed on a computer screen the way it was in the present experiment. This entails that, in the present test, there is an actual "objective correctness", which means that participants can measure their behaviour not necessarily against some conceptualization (as they are supposed to be doing in most models of speech production); rather, they can simply compare their performance with what was on the screen. When discussing this lack of "creative speaking" in the design of the present experiment, Dennett (personal communication) related how he had suggested an experiment to Levelt. In this experiment, participants are to be covertly primed by different cues in the environment (such as words on posters on the walls of the experiment-room or certain words in the instructions given to participants), to find out if these words, or perhaps phrases, would then turn up in participants' discussions of topics *largely unrelated* to the words or phrases. This test group is then compared to a control group who has not been subjected to the same priming (see also appendix B in Dennett, 1991). Dennett suggested that getting such priming to work in a controlled fashion in a voice-exchange experiment similar to the present one would certainly be an elegant way of meeting the technical requirements but still maintaining a very high degree of ecological validity.

Some notes on technical aspects of the experiment. While all trials in our Stroop test where it was *obvious* that the played back manipulated word was very different in strength, pitch, length or onset were removed from analysis, it is difficult to be certain that minor and perhaps not consciously detectable differences between the actually pronounced word and the played back manipulated word were not there. Such differences could of course play a large role in detection and still go unreported. For example, McGuire, Silbersweig, and Frith (1996) showed that when participants read aloud and the feedback they received was either another person's voice or their own pitch-distorted voice there was increased activation in the lateral temporal cortices as compared to just hearing their own, undistorted voice. In a study like the present one, the played back manipulated feedback could certainly be interpreted, on some level, as other-produced if there was but minor differences in aspects such as prosodic performance. If it were so, it would seem, we would not be measuring whether or not the participants accept the conceptual incongruence, but only whether or not they could judge the feedback as self- or other-produced. This is in essence a technical issue and there would appear to be two possible ways of approaching it. One is to construct the experiment in such a way as to make sure pronunciation is identical, or next to identical (save, of course, the phoneme(s) that differentiate the spoken word from the manipulated feedback). The other is to slightly distort the voice in the feedback loop so that differences in delivery are masked enough to rule out detection on their basis.

Concluding remarks

The present study represents an attempt at empirically investigating an aspect of speech production that has been dealt with largely in a theoretical manner. We have seen how the manipulation of verbal feedback has shown some interesting, and perhaps unintuitive, effects on participants understanding of the meaning of what they have said. Importantly, the results call for future studies that expand the scope of experimental situations in order to explore more fully the extent and limitations of the present results. By analyzing the design of the present experiment we have learned a great deal that will be useful in constructing future experiments using the same basic idea as was used in this study.

References

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.

Carroll, D.W. (1999). *Psychology of language (3rd ed.).* Pacific Grove, Ca.: Brooks/Cole Publishing Company.

Christoffels, I.K., Formisano, E., & Schiller, N.O. (2006). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping, 28(9),* 868-879.

Dell, G.S., & Reich, P.A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of verbal learning and verbal behaviour, 20,* 611-629.

Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review, 93(3)*, 283-321.

Dennett, D.C. (1987). *The intentional stance*. Cambridge: MIT Press.

Dennett, D.C. (1991). *Consciousness explained*. Boston: Little, Brown & Company.

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia, 36(11),* 1133-1140.

Fromkin, V.A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47(1),* 27-52.

Gallagher, S., & Sørensen, J. B. (2006). Experimenting with phenomenology. *Consciousness and Cognition, 15,* 119-134.

Gazzaniga, M.S., Ivry, R., & Mangun, G.R. (2005). *Cognitive Neuroscience – The Biology of the Mind (2nd ed.).* New York: W.W. Norton.

Griffin, Z.M. (2004). Why look? Reasons for eye movements related to language production. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 213-248). New York: Psychology Press.

Hall, L., Johansson, P., Tärning, B., Deutgen, T., & Sikström, S. (2006). Magic at the marketplace. *Lund University Cognitive Studies, 129.*

Harley, T. (2001). *The psychology of language (2nd ed.)*. New York: Psychology Press.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science, 310,* 116-119.

Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we can know: On Choice blindness and introspection. *Consciousness and Cognition, 15,* 673-692.

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*. Boston: MIT Press.

Levelt, W.J.M. (1992a). Accessing words in speech production: Stages, processes and representations. *Cognition, 42,* 1-22.

Levelt, W.J.M. (1992b). Fairness in reviewing: A reply to O'Connell. *Journal of Psycholinguistic Research, 21(5),* 401-403.

Levelt, W.J.M. (1999a). Models of word production. *Trends in Cognitive Sciences, Vol. 3, No. 6,* 223-232.

Levelt, W.J.M. (1999b). Producing spoken language: A blueprint of the speaker. In C.M. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 83-114). New York: Oxford University Press.

Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

Lind, A. (2006). On knowing and motivating one's choices – Markers of uncertainty and cognitive load in manipulated choice-reports. Unpublished BA Thesis, Centre for Language and Literature, Lund University.

Linnman, C., Carlbring, P., Åhman, Å., Andersson, H., & Andersson, G. (2006). The Stroop effect on the internet. *Computers in Human Behavior, 22*, 448-455

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109(2)*, 163-203.

Martin, N., Gagnon, D.A., Schwartz, M.F., Dell, G.S., & Saffran, E.M. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes, 11(3)*, 257-282.

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

Meyer, A.S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition, 42*, 181-211.

Neill, W.T. (1977). Inhibitory and facilitatory processes in selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 444-450.

Nielsen, T.I. (1963). Volition: A new experimental approach. *Scandinavian journal of psychology, 4*, 225-230.

Nielsen, T.I., Praetorius, N., & Kuschel, R. (1965). Volitional aspects of voice performance: An experimental approach. *Scandinavian Journal of Psychology, 6(3)*, 201-208.

Nisbett, R.E., & DeCamp Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84(3)*, 231-259.

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

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

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107(3)*, 460-499.

Rosenbaum, D.A. (1991). *Human motor control*. San Diego: Academic Press.

Slevc, R.L., & Ferreira, V.S. (2006). Halting in single word production: A test of the perceptual loop theory of speech monitoring. *Journal of Memory and Language, 54*, 515-540.

Sørensen, J.B. (2005). The alien-hand experiment. *Phenomenology and the Cognitive Sciences, 5*, 73-90.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Wegner, D.M., & Wheatley, T. (1999). Apparent mental causation – Sources of the experience of will. *American Psychologist, 54(7)*, 480-492.

Wegner, D.M. (2002). *The illusion of conscious will*. Cambridge: MIT Press.