Judgment of Disfluency in People who Stutter and People who do not Stutter: Results from Magnitude Estimation

Robin J. Lickley[1], **Robert J. Hartsuiker**[2], **Martin Corley**[3], **Melanie Russell**[1], **and Ruth Nelson**[1]

[1] *Queen Margaret University College, Edinburgh, U.K.*
[2] *Ghent University, Belgium*
[3] *University of Edinburgh, U.K.*

Key words

disfluency judgment

self-monitoring

stuttering

Abstract

Two experiments used a magnitude estimation paradigm to test whether perception of disfluency is a function of whether the speaker and the listener stutter or do not stutter. Utterances produced by people who stutter were judged as "less fluent," and, critically, this held for apparently fluent utterances as well as for utterances identified as containing disfluency. Additionally, people who stutter tended to perceive utterances as less fluent, independent of who produced these utterances. We argue that these findings are consistent with a view that articulatory differences between the speech of people who stutter and people who do not stutter lead to perceptually relevant vocal differences. We suggest that these differences are detected by the speech self-monitoring system (which uses speech perception) resulting in covert repairs. Our account therefore shares characteristics with the Covert Repair (Postma & Kolk, 1993) and Vicious Circle (Vasić & Wijnen, 2005) hypotheses. It differs from the Covert Repair hypothesis in that it no longer assumes an additional deficit at the phonological planning level. It differs from the Vicious Circle hypothesis in that it no longer attributes hypervigilant monitoring to unknown, external factors. Rather, the self-monitor becomes hypervigilant *because* the speaker is aware that his/her speech is habitually deviant, even when it is not, strictly speaking, disfluent.

1 Introduction

In the literature on stuttering, there are at least two major theoretical perspectives on the issue of the processing mechanisms that are implicated. One set of theorists points to deviations at the level of speech motor control (e.g., Adams, Freeman, &

Acknowledgments: This research was sponsored by British Academy, grant number 32045. The authors would like to thank all speakers and listeners who took part in this study, for making the study possible.

Address for correspondence: Dr. R. J. Lickley, Speech Science Research Centre, Queen Margaret University College, Edinburgh EH12 8TS, United Kingdom;
e-mail: <rlickley@qmuc.ac.uk>.

Conture, 1985; van Lieshout, Peters, Starkweather, & Hulstijn, 1993; Wood, 1995) whereas other theorists are primarily concerned with impairments at the level of linguistic planning (Conture, 1991; Kolk, 1991). There is indeed much evidence for speech motor control differences between people who stutter (PWS) and people who do not stutter (PWDNS), even when just fluent utterances are considered. However, many authors have argued that disturbances at the level of speech motor control do not suffice to explain the characteristics of stuttering (e.g., Conture, 1991; Hulstijn & Peters, 2000; Kolk, 1991; Melnick, Conture, & Ohde, 2005; Wingate, 1988). In general it has been contended that motor control theories of stuttering do not allow for a transparent mapping between underlying problem and key symptom. Additionally, motor control theories cannot easily account for the effects of linguistic variables such as grammatical complexity on stuttering frequency (e.g., Yaruss, 1999).

Taking these considerations into account, a number of authors have suggested that stuttering involves a linguistic planning impairment, and more specifically an impairment involving the mechanisms that *monitor* these planning processes (see Levelt, 1989; Postma, 2000 for reviews of self-monitoring theories). These hypotheses entail that the self-monitoring system perceives representations of speech that is not yet articulated, and that the disfluencies typically occurring in stuttering (blocks, prolongations, [part-] word repetitions) result from attempts to correct problems in internal speech (Kolk & Postma, 1997; Postma & Kolk, 1993; Vasić & Wijnen, 2005). An essential assumption of these hypotheses, for our current purposes, is that the monitor uses the *speech perception* system for monitoring internal and external speech.

The problems to which the monitor reacts in these theories are hypothesized to come from speech planning processes (i.e., before the onset of motor planning). Although these views acknowledge the existence of motor control impairments, they view them as unrelated to the key symptoms of stuttering (e.g., Wijnen, 2000, pp. 211–212).[1] However, a more parsimonious theory of stuttering would include motor control problems within a theory of stuttering rather than considering them as distinct phenomena. The possibility we will explore in this study is that the articulatory differences between the fluent speech production of PWS and PWDNS lead to acoustic differences that listeners (and therefore the self-monitoring systems of speakers) can perceive. On this account, it is the perceivable difference between intention and utterance to which the speaker is sensitive. If the account is right, it would still remain to be established how articulatory differences relate to acoustic differences: There could be a direct causal link, or there could be an indirect link, as a result of speaker strategies that minimize articulatory problems. Additionally, it would need to be established whether the acoustic differences increase the self-monitor's sensitivity to malformed speech in general, or whether the self-monitor initiates a correction each time it encounters a fragment of speech which differs from the norm in minor ways. The aim of this paper is much more modest, namely to test the assumption that apparently fluent speech of PWS is perceived as more deviant than that of PWDNS.

[1] A translation of the relevant passage is "However important these findings [concerning articulation programming deficits] may be, …, it is in my opinion fruitless to use them to account causally for the key symptoms of stuttering."

A secondary aim is to test whether PWS perceive speech as more deviant than PWDNS. In the next few paragraphs, we briefly discuss findings from the perspective of self-monitoring, before returning to the present study.

It has been suggested that PWS have deficits at the level of linguistic planning, and that the symptoms arise from interactions between such planning deficits and the self-monitoring system. An influential model has been the *Covert Repair hypothesis* (Kolk & Postma, 1997; Postma & Kolk, 1993). According to this view, the underlying deficit is linguistic in nature. In particular, there would be impairment at the processing level where the segmental content of words is determined, namely phonological spell-out (Dell, 1986; Levelt & Wheeldon, 1994). Because of this phonological impairment, PWS produce many phonological speech errors internally, which are subsequently detected and repaired by the self-monitor. The editing phase (interrupting and formulating a restart) results in disfluencies, and the type of disfluency depends on the moment of interruption. Note that according to this view the sensitivity of the monitor remains fixed, whereas the base rate of incidents eliciting correction increases.

Studies designed to test this hypothesis have focused on either the phonological encoding component (Burger & Wijnen, 1999; Wijnen & Boers, 1994) or on the monitoring component (Postma & Kolk, 1992; 1993; Postma, Kolk, & Povel, 1990). However, the evidence from these studies has been mixed (see Hartsuiker, Kolk, & Lickley, 2005; Vasić & Wijnen, 2005, for reviews). Whereas stuttering rates can be affected by experimental manipulations aimed at the monitor (e.g., dual-task paradigms), there is little evidence to support the contention of a separate phonological encoding disorder (e.g., Burger & Wijnen, 1999, observed no differences between PWS and PWDNS with respect to phonological encoding, using the so-called "implicit priming task," Meyer, 1990; 1991).

More recently, Vasić and Wijnen (2001; 2005) have presented a new monitoring hypothesis which no longer assumes a phonological encoding deficit. Instead, their "Vicious Circle hypothesis" localizes the deficit in the self-monitoring system itself. In particular, in PWS the self-monitor is hypervigilant, so that internal speech is more often considered as discrepant (and thus in need of covert repair) than is the case for PWDNS. In this model, PWS and PWDNS do not differ with respect to the speech that is initially generated as input to the monitor. Note that this claim is similar to an earlier proposal by Sherrard (1975), according to which PWS would repair "phantom" speech errors. The Vicious Circle hypothesis is attractive because it provides a parsimonious account of the variation in stuttering rates reported across periods of time and differing situations. By suggesting that the sensitivity of the monitor varies (either because of availability of resources, or because of a situation-driven awareness of speech quality), the model is easily extended to account for a number of key findings in the stuttering literature (e.g., effects of delayed, or frequency-altered auditory feedback, effects of speaking in synchrony with a metronome or with other people, singing, whispering, shouting, changing regional accent, and the probability of stuttering dependent on utterance position; cf. Wijnen, 2000).

However, there are three shortcomings with the model as it stands. First and foremost, the empirical evidence offered to support this model fails to discriminate between the Vicious Circle and Covert Repair hypotheses: Vasić and Wijnen's (2005) study showed that PWS had reduced stuttering rates when concurrently performing secondary tasks, but, as mentioned above, the Covert Repair hypothesis predicts the same result.[2] Second, there is no description of the conditions that lead the self-monitor to become hypersensitive in the first place: Because there is no difference in the numbers of speech incidents to which it might react, its hypersensitivity has to be attributed to (unexplained) external factors. And third, in common with many other psychologically-motivated models, no attempt is made to incorporate findings from the motor control literature: The model accounts for the acoustic phenomena associated with stuttering, but it does not link these to the observable articulatory phenomena.

One possible solution to the last two problems may be that the sensitivity of the monitor arises precisely because a PWS, at some point in his or her linguistic development, becomes aware that their speech is on occasion "malformed," perhaps as a result of motor differences from the norm. But this implies that PWS, and by implication their self-monitoring systems, must be able to register the effect that deviant motor control has on speech. In the following sections we describe two experiments designed to test whether such differences can be detected.

1.1
Plan of the experiments

Adults who stutter have usually stuttered most of their lives. According to the Vicious Circle hypothesis, their self-monitoring systems have become extremely sensitive to small deviations in speech, and they will often interrupt and repair themselves. This in turn leads to many of the key symptoms of stuttering. On the assumption of many monitoring theories that we monitor speech through the perceptual system (Hartsuiker & Kolk, 2001; Levelt, 1983; 1989; Nooteboom, 2005; see Postma, 2000, for a review) the sensitivity of the monitor to variations in speech fluency is testable using a perception paradigm in which participants rate excerpted recordings of speech.

Because fluency or disfluency of speech are relative qualities, it is important to find a way to allow listeners to express their judgments even in cases where the perceived differences are small. This militates against a categorical or n-point scale, such as a Likert scale (see also Finn & Ingham, 1989, for a discussion of fluency rating scales). Magnitude Estimation (ME), a technique imported into linguistics from psychophysics (see Bard, Robertson, & Sorace, 1996), allows participants to make fine-grained judgments of linguistic phenomena. ME has been demonstrated to be sensitive to a number of graded linguistic phenomena such as grammaticality (Keller & Alexopolou, 2001). In previous research on stuttering, it has been argued that ME has greater construct validity than comparable techniques (Schiavetti, Sacco, Metz, & Sitler, 1983).

A group of PWS and a group of PWDNS listened to short fragments of speech and judged "how fluent they sounded" using ME. The fragments were spoken by

[2] In contrast to the Vicious Circle hypothesis, the Covert Repair hypothesis also predicts an increase in error frequency when disfluency decreases, since errors are generated at the speech planning stage. However, there is no unequivocal evidence to support this prediction.

either PWS or PWDNS and they were classified as either fluent or disfluent. Just as the perception of speech as fluent or disfluent is not clear-cut, so the selection of fluent or disfluent stimuli for experimentation can be controversial (Finn & Ingham, 1989). In this experiment, disfluent utterances were defined as those that contained a single repetition of a word onset and no other disfluency (e.g., *more or less ben- beneath that point*). Thus, the disfluent utterances in our study were all of a kind that is common in the speech of PWDNS, rather than being typical of stuttered speech alone. Fluent utterances were defined as those that contained no prolongations, blocks, repetitions, revisions, or major hesitations such as filled pauses (*um, uh*).

This design allowed us to make two testable predictions. First, even fluent fragments of speech produced by PWS (in the sense that the fragment contains no observable stuttering incidents) should be rated as less fluent than fluent speech by PWDNS. This follows from the assumption that a reduced speech quality, perhaps as a result of motor control deviations in PWS, can be perceived. Second, if PWS and PWDNS judge the same fragments, the PWS should tend to use lower fluency ratings. This follows from Vasić and Wijnen's (2005) assumption that PWS have become hypersensitive to fluency problems in speech, and are more likely to classify a given speech fragment as "disfluent."

2 Experiment 1

2.1 *Method*

2.1.1 Participants

In order to obtain the stimulus materials, eight PWS, (7 males, 1 female) and all native speakers of (Scottish) English, participated in pairs in a dialog task (based on the Map Task, Anderson et al., 1991). In this task, one person (the instruction-giver) describes a route on a slightly different map to another person (the instruction-follower). Each participant was recorded playing each role. This task results in natural speech, since discrepancies between the maps provide ample occasions for discussion and negotiation.

Twenty PWS and 20 age- and gender-matched controls participated in the experiment proper. In each group, there were 16 males. Average age for each group was 45 years. All PWS, but none of the controls, considered themselves to have a stuttering impairment.

2.1.2 Materials

We excerpted 50 utterances (short segments of speech, typically less than 2 s long) from the recorded dialogs between PWS, using samples from five of the eight speakers, one female. Of these utterances, 25 were disfluent, containing single word-onset repetitions. The remaining 25 were fluent (see above for criteria) and matched with a disfluent utterance for onset of the disfluent word. A further 50 short utterances were excerpted from dialogs between PWDNS available in the HCRC Map Task Corpus

(Anderson et al., 1991). This corpus is fully annotated for disfluency, following Lickley (1998); the corpus of dialogs by PWS was also annotated, using a new version of this system, supplemented to include blocks and prolongations. Again, 25 utterances were disfluent, and 25 matched utterances were fluent. As far as possible, pairs of stimuli obtained from PWS were matched to pairs from PWDNS on the basis of the word containing the disfluent onset phoneme (of 25 matched pairs, only one differed in onset) and for sex of speaker. Examples of the sets of stimuli are shown orthographically in Table 1.

Table 1

Experiment 1: Examples of a set of matched stimuli, transcribed orthographically

Speaker	*Fluency of stimulus*	*Stimulus*
PWS	Disfluent	*about f- four inches south*
PWS	Fluent	*approximately four to four and a half inches*
PWDNS	Disfluent	*fo- four inches down*
PWDNS	Fluent	*four inches down*

To the resulting 100 stimuli we added a further 100 fluent filler stimuli, excerpted from dialogs between speakers in the HCRC Map Task Corpus. Recordings for the fillers came from a different set of speakers than those selected for the experimental items. Additionally, a further 10 filler stimuli were selected as "practice" items. Finally, a reference stimulus was selected (in Magnitude Estimation, all judgments are made relative to a given stimulus). The reference stimulus (1) was selected on the basis that it contained no disfluent repetition but could be viewed as moderately disfluent, as it contained a silent pause:

(1) You're still going to be abou (380ms pause) t six centimeters down.

Four lists were constructed, each containing all the stimuli in a different random order, with the restriction that each list began with the 10 practice stimuli and was followed by the reference stimulus. The reference stimulus was repeated every 10 items. The lists were recorded on DAT tapes.

2.1.3 Procedure

The experiment was administered as a paper and pencil task. Participants listened to the DAT tapes over high quality headphones and judged the fluency of each stimulus that they heard. They wrote their ratings of each stimulus in the corresponding box on a prepared scoring sheet. Participants were required to assign an arbitrary number to the reference stimulus, and judge each stimulus in comparison to the reference (e.g., if a reference line of 10cm were assigned the arbitrary number 100, then a veridical judgment of a line of 20cm would be 200).

In order to explain this procedure to the participants, a first practice phase involved five judgments of line lengths. When the experimenter was convinced the participant understood the initial procedure, a second practice phase involved 10 judgments of fluency. Participants were instructed to assign higher scores to relatively fluent items and lower scores to relatively disfluent items. Additional instructions emphasized that the judgments should not be based on considerations of gender or accent of speaker, nor on the content, grammatical structure, or length of the stimulus. After this practice session, the experimenter provided a prepared comment on each stimulus (e.g., *nothing wrong with this, there is only some background noise on the tape, so this rating should be close to the reference*). At no point were subjects informed that some of the speakers they would hear in the experiment had a stuttering impairment.

When it was clear that the participant understood the task, the experimental phase began. Each trial began with a single beep, followed by the stimulus. There was an interval of several seconds, to allow participants to write down each rating, between trials. The reference stimulus was always preceded by two beeps. The experimental phase consisted of two blocks of approximately 20 mins each.

2.2
Results

The raw ratings were standardized by dividing them by the reference rating. Because the standardized data were ratios (how much more or less fluent than the reference stimulus) they were then log-transformed. A transformed rating of zero thus indicated that the participant had judged a stimulus to be equivalently fluent to the reference stimulus; scores less than zero indicated increased disfluency, and scores greater than zero indicated that the stimulus had been rated as relatively fluent. The mean log standardized ratings per condition are shown in Table 2.

Table 2

Mean standardized rating per condition (fluent or disfluent fragments spoken by PWS or PWDNS) and judge (PWS or PWDNS)

		Stimulus			
		PWS		*PWDNS*	
		fluent	*disfluent*	*fluent*	*disfluent*
Judge	PWS	−0.07	−0.39	0.01	−0.31
	PWDNS	0.06	−0.26	0.11	−0.20

The data were subjected to two ANOVAs, one with subjects (F_1) and one with items (F_2) as the random variable.

There were additive effects of fluency of stimulus (fluent or disfluent), speaker of stimulus (PWS or PWDNS), and of judge (PWS or PWDNS). Fluent stimuli were judged as more fluent than the disfluent stimuli, 0.03 versus −0.29; $F_1(1, 38) = 212.91$, $p < .001$; $F_2(1, 24) = 178.59$, $p < .001$, confirming that participants followed the task instructions and were highly sensitive to disfluency. Second, PWDNS provided somewhat more lenient judgments overall. This effect was highly significant by items, but only marginally significant in the by-subjects analysis, −0.07 versus −0.19; $F_1(1, 38) = 3.02$, $p < .10$; $F_2(1, 25) = 190.13$, $p < .001$.[3] Finally, as predicted, stimuli produced by PWDNS were judged as more fluent than stimuli produced by PWS, −0.10 versus −0.16; $F_1(1, 38) = 32.81$, $p < .001$; $F_2(1, 24) = 8.33$, $p < .01$. No second-order or third-order interaction reached significance.

The additive effects of source and fluency of stimulus suggested that PWS were always rated more disfluent, even if the stimulus was fluent. This was confirmed in a post hoc test, restricted to fluent stimuli only, PWS: 0 versus PWDNS: 0.06; $F_1(1, 39) = 20.20$, $p < .001$; $F_2(1, 24) = 5.42$, $p < .05$.

2.3
Discussion

Experiment 1 demonstrates that judges are indeed sensitive to disfluency. Thus, the current experimental approach of magnitude estimation is justified. Interestingly, there is some evidence to suggest that PWS are more sensitive to disfluency than PWDNS (cf. Tuthill, 1940). Given this finding, there is no reason to contradict Vasić and Wijnen's (2005) assertion that the self-monitoring system is hypersensitive to malformed speech in PWS. On the assumption, then, that the self-monitoring system is implicated in stuttering, the question remains whether its sensitivity can be "bootstrapped" by perception of minor deviations in speech. Crucially, fragments produced by PWS were rated as less fluent than those produced by PWDNS, even in the case of "fluent" fragments. This suggests that small differences in the speech of PWDNS and PWS (possibly attributable to motor control differences) have perceptual consequences and thus are in principle available to the self-monitoring system. However, before embracing this conclusion, we turn to an alternative explanation of the findings. The disfluencies on each tape were generated by a limited number of speakers. It is therefore conceivable that the participants were able to assign the recorded speakers to categories: A PWS might be recognized from their vocal characteristics, even where the fragment currently being rated happened to be fluent. If participants rated disfluent fragments recorded by PWS as less fluent than those recorded by PWDNS, it is possible that knowing "who was speaking" when encountering a fluent fragment may have affected their ratings. In other words, the effect for fluent fragments alone, reported above, might be the consequence of noting differences in the way (groups of) speakers produced disfluent fragments. In order to directly test this assertion, Experiment 2 replicates Experiment 1, but includes no disfluent fragments and includes only PWDNS as listeners: Any distinction between stimuli produced by

[3] When a single PWDNS, who provided extremely low ratings, was discarded from analysis, the effect of judge did become significant, $F_1(1, 37) = 5.58$, $p < .05$; $F_2(1, 24) = 302.11$, $p < .001$.

PWS and those produced by PWDNS must be made on the basis of fluent fragments alone, and by implication, because of slight vocal differences.

3 Experiment 2

3.1 *Method*

3.1.1 Participants

Fourteen subjects (6 males, 8 females) aged 21 – 52 took part in this experiment. All were native speakers of British English and none reported any hearing disorder or speech impairment. None had taken part in the previous study.

3.1.2 Materials

All fluent stimuli used in Experiment 1 were included in this experiment recorded onto four DAT tapes in the same random orders as previously. So once again there were 25 fluent stimuli extracted from dialogs involving PWS and 25 extracted from dialogs involving PWDNS, matched as described previously. These were combined randomly with 50 filler items, produced by PWDNS whose speech was not included in the test items.

3.1.3 Procedure

As in the previous experiment, this experiment was preceded by an introduction to the task of magnitude estimation, followed by similar practice phases. In the practice phase with spoken stimuli, the set of stimuli was altered to include only stimuli with no repetition, to reflect the nature of the stimuli in the experiment proper. The experiment then continued as described previously, except that, since there were fewer stimuli, the running time was halved and no break was required. Once again, subjects were not informed in advance that some of the people that they would hear had a stuttering impairment.

3.2 *Results*

As in the previous experiment, raw ratings were standardized by dividing them by the rating for the reference stimulus and then log-transformed, so that scores above zero indicated that judges rated a stimulus as more fluent than the reference stimulus and scores below zero indicated a "more disfluent" judgment. As before, judges rated stimuli produced by PWS as more disfluent than those produced by PWDNS (−0.126 vs. 0.153). The difference was significant by subjects, $t_1(13) = 5.66$, $p < .001$, and by items, $t_2(24) = 5.48$, $p < .001$.

If listeners in Experiment 1 had been influenced in their judgments on fluent samples by recognizing speakers as PWS from the nature of their disfluent stimuli, then the ratings in Experiment 2 should reflect this: Fluent stimuli of PWS should be rated as more fluent than in Experiment 1 (the Artifact hypothesis). They were not: Two ANOVAs, one with subjects (F_1) and one with items (F_2) as the random variable show a main effect of speaker (PWS or PWDNS), − 0.019 versus 0.125; $F_1(1, 32) = 51.96$, $p < .001$; $F_2(1, 48) = 18.61$, $p < .001$. A main effect of experiment (Experiment 1 vs. Experiment 2) was only significant in the by-items analysis, 0.081 versus 0.013; $F_1(1, 32) < 1$; $F_2(1, 48) = 10.65$, $p < .01$. A significant Speaker × Experiment interaction was in the opposite direction from that predicted by the Artifact hypothesis: Listeners actually judged fluent stimuli of PWS as *more* disfluent in Experiment 2 and those of PWDNS as more *fluent*, $F_1(1, 32) = 24.97$, $p < .001$; $F_2(1, 48) = 30.33$, $p < .001$. This was probably a result of judges in Experiment 2 mapping less variable stimuli (because they were all fluent) onto a comparable scale to the one used in Experiment 1.

3.3 *Discussion*

Experiment 2 dispels some doubts about the source of the lower ratings for fluent stimuli from PWS in Experiment 1. In Experiment 2, subjects rated fluent stimuli produced by PWS as less fluent than control stimuli, even without hearing any disfluent stimuli produced by the same speakers. Not only was there no support for the artifact hypothesis, but the results pointed in the opposite direction to that predicted by the hypothesis, with more “disfluent” judgments for stimuli from PWS and more “fluent” judgments for stimuli from PWDNS. So, we must assume that audible features of the fluent speech samples themselves influenced subjects’ ratings.

4 General Discussion

Experiment 1 showed that the speech of PWS was rated as more disfluent than that of PWDNS, irrespective of whether we had classified it as fluent or not and of whether the listener was a PWS or not. Experiment 2 ruled out the possibility that the latter finding was a result of listeners recognizing the voices of PWS from their disfluent utterances and this negatively influencing their subsequent judgments of the same speakers’ fluent utterances. Additionally, Experiment 1 suggested that PWS tended to give harsher judgments of fluency, rating both fluent and disfluent stimuli as more disfluent than did PWDNS.

The main finding from these experiments, then, was that listeners are able to discriminate between the speech of PWS and that of PWDNS even in cases where there is no overt disfluency. One interpretation of this finding is that small deviations in speech brought about by motor control differences are not merely part of the phenomenology of stuttering but are directly available to the perceptual processes and by implication to the self-monitoring system. Over time PWS may come to realize that the speech they produce differs from the norm. This in turn may lead to an attempt to rectify the situation by investing more resources in self-monitoring of planned (internal) speech. Presently, we can only speculate as to whether the motor control differences are directly linked to the vocal differences, or whether this is mediated by strategic adaptation. Adults who stutter have often been through various forms of speech therapy or self-teaching and may use speech techniques in order to maintain fluency. These include slow and prolonged speech, the use of soft articulatory contacts and various breathing techniques. Such techniques may make speech sound fluent, in that it contains little repetition or blocking, but unnatural, in that articulatory and durational features deviate from normal patterns: such speech has been referred to as "tenuous fluency" (Adams & Runyan, 1981) or "pseudofluency" (e.g., Dayalu & Kalinowski, 2002).

An important implication of our perceptual studies is that there are acoustic differences between the fluent speech stimuli of our two sets of speakers. While it is beyond the scope of the present study to investigate these fully and difficult to implement a well-controlled acoustical study on the basis of this relatively small set of speech data, we were at least able to make a comparison of speech rates between the two sets of speakers. PWS were found to speak more slowly than PWDNS, 4.2 versus 5.4 syllables per second: $t(48) = 2.77, p < .01$. However, the relationship between rate of speech and fluency judgments is not clear. A significant correlation between speech rate and fluency judgments was found only for fluent stimuli produced by PWS in Experiment 2 ($r = 0.47$, $N = 25$, $p < .02$). It is important to note though that the stimuli produced by PWS were more variable in speech rate (PWS: $SD = 1.70$; PWDNS: $SD = 1.21$), and that judgments of these stimuli varied much more (PWS: $SD = 0.23$; PWDNS: $SD = 0.09$), which could account for the lack of correlation in the PWDNS data. This leaves open the possibility that the judges partially based their judgments on speech rate.

A second finding from this study was that PWS tend to judge speech as "more disfluent" and this did not depend on whether the speech was produced by a PWS, or whether we had classified it as disfluent. That result is consistent with Vasić and Wijnen's (2001; 2005) Vicious Circle hypothesis. Because these authors assume that self-monitoring in production is perceptually based, they argued that the quality of monitoring is a function of attentional parameters, specifically, *effort* (the amount of cognitive resources that are invested in monitoring), *focus* (the degree of attention to specific aspects of speech) and *threshold* (the level of perceived deviation that is necessary to trigger a repair). Whereas these authors presented their findings as supporting increased cognitive effort and maladaptive focus in PWS, our study suggests differences in threshold.

In conclusion, our account shares characteristics with both the Covert Repair and Vicious Circle hypotheses. Our data strongly suggest that PWS produce perceptually deviant *fluent* speech, and on our account, this increases the baseline rate of incidents to which the monitor reacts. The data also suggest that PWS more readily classify speech as deviant, which increases the number of reactions by the monitor. The novelty in our account is that we link speech deviancy to aspects of speech production that have indeed been shown to be deviant, namely articulation. In contrast, the Covert Repair hypothesis postulates an additional deficit, at the level of phonological encoding, for which there is no experimental evidence. Furthermore, while the Vicious Circle hypothesis assumes a hypersensitive self-monitor, it has no account for why the monitor becomes hypersensitive in the first place. While we do not dispute that the self-monitor in PWS is hypersensitive, on our account speech deviancy comes first, and results, at least initially, from articulatory deviancies. Whether articulatory deviancies would result in observable stuttering without hypervigilant self-monitoring is presently unclear.

Of course, the current study only provides evidence for a precondition of our account — namely, the condition that apparently fluent speech in PWS is perceptually deviant. As pointed out in the introduction, further research will have to establish how articulatory and perceptual deviancies are related. One suggestion from our acoustic analyses is that this relation is indirect (strategic slowing of speech rate) rather than direct. A related point concerns individual differences. Adults who stutter have, for example, different therapy histories, possibly promoting variation in the types of strategies they apply to maintain fluency. Such variation could have repercussions for the way in which their fluent speech deviates from the norm, and possibly also for the criteria they apply in monitoring their speech.

manuscript received: *11.19.2004*
manuscript accepted: *08.15.2005*

References

ADAMS, F. R., FREEMAN, F. J., & CONTURE, E. (1985). Laryngeal dynamics of stutterers. In R. F. Curlee & W. H. Perkins (Eds.), *Nature and treatment of stuttering: New directions*. San Diego, CA: College-Hill Press.

ADAMS, F. R., & RUNYAN, C. M.(1981). Stuttering and fluency: Exclusive events or points on a continuum? *Journal of Fluency Disorders*, **6**, 197–218.

ANDERSON, A. H., BADER, M., BARD, E., BOYLE, E., DOHERTY, G., GARROD, S., ISARD, S., KOWTKO, J., McALLISTER, J., MILLER, C., SOTILLO, C., THOMPSON, H., & WEINERT, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, **34**, 351–366.

BARD, E., ROBERTSON, D., & SORACE, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, **72**, 32–68.

BURGER, R., & WIJNEN, F. (1999). Phonological encoding and word stress in stuttering and non-stuttering subjects. *Journal of Fluency Disorders*, **24**, 91–106.

CONTURE, E. G. (1991). *Stuttering: Its nature, diagnosis and treatment*. Needham Heights, MA: Allyn and Bacon.

DAYALU, V. N., & KALINOWSKI, J. (2002). Pseudofluency in adults who stutter: The illusory outcome of therapy. *Perceptual and Motor Skills*, **94**(1), 87–96.

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

FINN, P., & INGHAM, R. J. (1989). The selection of “fluent” samples in research on stuttering: Conceptual and methodological considerations. *Journal of Speech and Hearing Research*, **32**, 401–418.

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., KOLK, H. H. J., & LICKLEY, R. J. (2005). Stuttering on function words and content words: A computational test of the Covert Repair hypothesis. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 261–280). Hove, U.K.: Psychology Press.

HULSTIJN, W., & PETERS, H. F. M. (2000). Stuttering: A disorder of motor control? In B. Maasssen, W. Hulstijn, R. Kent, H. F. M. Peters, & P. H. M. M. van Lieshout (Eds.), *Speech motor control in normal and disordered speech.* Proceedings of the 4th International Speech Motor Conference (pp. 316–321). Nijmegen, The Netherlands: Vantilt.

KELLER, F., & ALEXOPOLOU, T. (2001). Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition*, **79**, 301–372.

KOLK, H. H. J. (1991). Is stuttering a symptom of adaptation or of impairment? In Peters, H. F. M., Hulstijn, W., & Starkweather, C. W. (Eds.), *Speech motor control and stuttering.* Amsterdam: Elsevier/Excerpta Medica.

KOLK, H. H. J., & POSTMA, A. (1997). Stuttering as a covert-repair phenomenon. In: R. Corlee & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (pp. 182–203). Boston: Allyn and Bacon.

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, Massachusetts: MIT Press.

LEVELT, W. J. M., & WHEELDON, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, **50**, 239–269.

LICKLEY, R. J. (1998). HCRC Disfluency Coding Manual. HCRC Technical Report. HCRC/TR–100

MELNICK, K. S., CONTURE, E. G., & OHDE, R. N. (2005). Phonological encoding in young children who stutter. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 102–118). Hove, U.K.: Psychology Press.

MEYER, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, **29**, 524–545.

MEYER, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, **30**, 69–89.

NOOTEBOOM, S. G. (2005). Listening to oneself: Monitoring speech production. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 167–186). Hove, U.K.: Psychology Press.

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

POSTMA, A., & KOLK, H. H. J. (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., & KOLK, H. H. J. (1993). The Covert Repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, **36**, 472–487.

POSTMA, A., KOLK, H. H. J., & POVEL, D. J. (1990). On the relation among speech errors, disfluencies, and self-repairs. *Language and Speech*, **33**, 19–29.

SCHIAVETTI, N., SACCO, P. R., METZ, D. E., & SITLER, R. W. (1983). Direct magnitude estimation and interval scaling of stuttering severity. *Journal of Speech and Hearing Research*, **26**, 568–573.

SHERRARD, C. A. (1975). Stuttering as false alarm responding. *British Journal of Disorders of Communication*, **10**, 83–91.

TUTHILL, C. (1940). A quantitative study of extensional meaning with special reference to stuttering. *Journal of Speech Disorders*, **5**, 189–191.

Van LIESHOUT, P., PETERS, H., STARKWEATHER, C. W., & HULSTIJN, W. (1993). Physiological differences between stutterers and nonstutterers in perceptually fluent speech: EMG amplitude and duration. *Journal of Speech and Hearing Research*, **36**, 55–63.

VASIĆ, N., & WIJNEN, F. (2001). Stuttering and speech monitoring. *Proceedings of DISS 2001* 29–31 August 2001, Edinburgh, United Kingdom, pp. 13–17.

VASIĆ, N., & WIJNEN, F. N. K. (2005). Stuttering as a monitoring deficit. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 226–247). Hove: Psychology Press.

WIJNEN, F. N. K. (2000). Stotteren als resultaat van inadequate spraakmonitoring [Stuttering as the result of inadequate speech monitoring]. *Stem-, Spraak- en Taalpathologie*, **9**, 211–230.

WIJNEN, F. N. K., & BOERS, I. (1994). Phonological priming effects in stutterers. *Journal of Fluency Disorders*, **19**, 1–20.

WINGATE, M. (1988). *The structure of stuttering*. New York: Springer.

WOOD, S. (1995). An electropalatographic analysis of stutterers' speech. *European Journal of Disorders of Communication*, **30**, 226–236.

YARUSS., S. (1999). Utterance length, syntactic complexity and childhood stuttering. *Journal of Speech, Language and Hearing Research*, **42**(2), 329–344.