Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop, 5–8 September 2003, Göteborg University, Sweden. Robert Eklund (ed.), Gothenburg Papers in Theoretical Linguistics 90, ISSN 0349–1021, pp. 35–37.

Perception of disfluency in people who stutter and people who do not stutter: Results from magnitude estimation

Robert J. Hartsuiker† , Martin Corley‡, Robin Lickley§ & Melanie Russell§

† Ghent University, Belgium
‡ University of Edinburgh, UK
§ Queen Margaret University College, Edinburgh, UK

Abstract

Recent accounts of stuttering [7, 15] consider disfluencies the result of an interaction between speech planning and self-monitoring, emphasizing the continuity between errors made in everyday speech and those made by people who stutter. On Vasiç & Wijnen's [14, 15] account, the monitor is hypervigilant for upcoming problems and interrupts and restarts the speech signal, resulting in disfluent speech. Crucially, on this account, self-monitoring is a perceptual function. Therefore, this account makes two predictions (1) people who stutter are also hypervigilant in perceiving another person's speech. (2) the quality of disfluencies made by people who stutter and those who do not will be comparable. We tested these hypotheses using a magnitude estimation judgment task. Twenty participants who stutter and 20 controls were asked to rate the fluency of excerpted fluent and disfluent fragments from recorded dialogues, either between people who stutter or between non-stutterers. In line with the first hypothesis, people who stutter tended to rate all fragments as more disfluent than controls did. However the second hypothesis was not confirmed: across judges, fluent and disfluent fragments excerpted from recordings of people who stutter were rated as less fluent than those excerpted from control dialogues, suggesting that there are perceptually relevant differences between the speech of PWS and PWDNS, independent of number and type of disfluencies.

1. Introduction

There is increasing attention for the hypothesis that the disfluencies typically occurring in stuttering (e.g., blocks, prolongations, hesitations, (part-)word repetitions, and self-corrections) are related to self-monitoring processes, the processes with which speakers inspect the quality of their own speech (see [10], for a recent review of monitoring theories). In a nutshell, this hypothesis entails that persons who stutter (PWS) detect many planning problems in their internal speech, and that disfluencies result from attempts to correct these problems ([11, 15]). Monitoring accounts generally assume continuity between the speech of PWS and people who do not stutter (PWDNS): disfluencies in both groups result from the same mechanisms, which tend to come into play more often in PWS. The aim of this study is to evaluate a specific aspect of a monitoring account proposed by Vasiç & Wijnen [15] and to put the continuity hypothesis to the test.

The first monitoring account was Postma & Kolk's [7], [11] Covert Repair Hypothesis, which localizes the difference between PWS and PWDNS at the processing level where the segmental content of words is determined, i.e., phonological spell-out [5]. Because of the phonological impairment, PWS produce many phonological speech errors internally, which are subsequently detected and edited out by the self-monitor. The editing phase (interrupting and restarting) would result in disfluencies, and the type of disfluency would depend on the moment of interruption. However, evidence for the covert repair hypothesis is mixed (see [6, 15] for reviews). In particular, there is little evidence that PWS produce excessive rates of phonological speech errors internally. Additionally, a recent study [4] found no group difference on an implicit priming task, a paradigm that is assumed to tap into phonological encoding [9].

More recently, Vasiç & Wijnen [14, 15] presented a variant of the covert repair hypothesis which no longer assumes a phonological encoding deficit. Instead, their 'vicious circle hypothesis' directly implicates the self-monitor. In particular, the self-monitor would be hypervigilant so that internal speech is more often considered as discrepant – and thus in need of covert repair – than is the case for PWDNS. They argued that three parameters of monitoring might be responsible for this hypervigilance, on Levelt's [8] theory in which monitoring is a perceptual function. The first monitoring parameter is effort. PWS might invest so much effort in monitoring their speech, that they detect problems that PWDNS tend to miss. The second parameter is focus, or in other words the set of those aspects of speech to which the monitor attends. The focus in PWS may be maladaptive (i.e., paying to much attention on aspects of speech that frequently deviate but which are unimportant, such as slight variations in the timing of speech plan delivery). The third parameter is threshold. PWS may set the threshold for accepting a speech plan as well-formed too high, leading to more rejections (and hence attempts at repair) than PWDNS.

Vasiç & Wijnen's study concentrated mainly on effort and focus. While participants spoke, they simultaneously performed a secondary task: a visuo-spatial tasked aimed at decreasing the amount of effort that could be invested in monitoring, or a word-spotting task, aimed at changing the focus of the monitor. Both manipulations decreased the rate of disfluencies in PWS (in particular blocks). However, in PWDNS, the visuo-spatial task decreased the number of disfluencies, but the word spotting task increased that number (in particular, of word repetitions). The data thus confirmed Vasiç & Wijnen's two predictions concerning effort and threshold. However, it is less clear whether these data are in agreement with the continuity hypothesis.

The present study evaluates the third parameter (threshold) and reassesses the continuity hypothesis. Since the vicious circle hypothesis is based on the assumption that speech is monitored by perceiving it, we chose to directly assess it in a speech perception paradigm. A group of PWS and a control group listened to short fragments of speech and judged 'how fluent they sounded'. The fragments were spoken by either PWS or PWDNS and they were either fluent or disfluent. The hypothesis that PWS set the threshold too high predicts that PWS judge fragments as more disfluent than the controls would. Additionally, the continuity hypothesis predicts that judges do not discriminate between equivalent disfluencies produced by PWS and those produced by PWDNS.

2. Method

2.1. Participants

In order to obtain the stimulus materials, 8 PWS, all males and all native speakers of (Scottish) English, participated in pairs in a dialogue task (the Map Task, [2]). 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 occasions for discussion and negotiation.

Twenty PWS and 20 age- and gender-matched controls participated in the perception experiment. 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 'stammers'.

2.2. Materials

We excerpted 50 fragments (short segments of speech, typically less than 2 seconds long) from the recorded dialogues between PWS. Of these fragments, 25 were disfluent, containing single word-onset repetitions. The remaining 25, matched for onset, were fluent. A further 50 fragments were excerpted from dialogues between male PWDNS available in the Map Task Corpus [2]. Again, 25 fragments were disfluent, and 25 matched fragments were fluent. As far as possible, pairs of fragments obtained from PWS were matched to pairs from PWDNS (of 25 matched pairs, only one differed in onset phoneme). To the resulting 100 fragments we added a further 100 filler fragments, varying in phonology and fluency, excerpted from dialogues between male speakers in the Map Task Corpus. None of the speakers used for fillers were used for experimental items. Finally, a further 10 filler fragments were selected as 'practice' fragments.

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

2.3. 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 fragment that they heard. They wrote their ratings of each fragment in the corresponding box on a prepared scoring sheet. The rating paradigm use was Magnitude Estimation ([3, 13]). This psychophysical technique requires participants 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 10 cm would be assigned the arbitrary number 100, then a veridical judgment of a line of 20 cm would be 200).

In order to explain this procedure to the participants, a first practice phase involved 5 judgments of line lengths. When the experimenter was convinced the participant understood that procedure, a second practice phase involved 10 judgments of disfluency. Instructions emphasized that the judgment should not be based on considerations of gender or accent of speaker, and neither on the content, grammatical structure, or length of the fragment. After each practice fragment, the experimenter provided a prepared comment on that fragment (e.g., 'nothing wrong with this, there is only some background noise on the tape, so this rating should be close to the reference').

When it was clear that the participant understood the task, the experimental phase began. Each trial began with a single beep, followed by the fragment. There was an interval of several seconds, to allow participants to write down each rating, between trials. The reference stimulus was always proceeded by two beeps. The experimental phase consisted of two blocks of approximately 25 minutes each.

3. Results

The raw ratings were standardized by dividing them by the reference rating. Since the data were ratios (how much more or less fluent than the reference) 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 modulus; scores less than zero indicated increased disfluency, and scores greater than zero indicated that the stimulus had been rated as relatively fluent.

The mean standardized ratings per condition are shown in Table 1.

Table 1: Mean standardized rating per condition (fluent or disfluent fragments spoken by PWS or PWDNS) and juge (PWS or PWDNS).

Judge	pws-fluent	pws-disfluent	pwdns-fluent	pwdns-disfluent
PWS	-0.07	-0.39	0.01	-0.31
PWDNS	0.06	-0.26	0.11	-0.20

The data were subjected to two analyses of variance, one with subjects (F_1) and one with items (F_2) as the random variable. We set the alpha-level at 0.05.

There were additive effects of fluency of fragment (fluent or disfluent), speaker of fragment (PWS or PWDNS) and of judge (PWS or PWDNS). Fluent fragments were judged as more fluent than the disfluent fragments (0.03 vs -0.29; $F_1(1, 38) = 212.9$; $F_2(1, 24) = 178.6$). Fragments produced by PWDNS were judged as more fluent than fragments produced by PWS (-0.10 vs –0.16; $F_1(1, 38) = 32.81$; $F_2(1, 24) = 8.33$). Finally, PWDNS provided more lenient judgments overall. This effect was highly signficant by-items, but only marginally significant in the by-subjects analysis (-0.07 vs –0.19; $F_1(1,38) = 3.02$; $F_2(1,25) = 190.13$). No second-order or third-order interaction reached significance.

The additive effects of source and fluency of fragment surprisingly suggested that PWS were always rated more disfluent, even if the fragment was fluent. This was confirmed in a post-hoc test, restricted to fluent fragments only (PWS: 0 vs PWDNS: 0.06; $F_1(1, 39) = 20.20$; $F_2(1, 48) = 4.89$).

4. Discussion

Taken together with the study reported by Vasiç & Wijnen [14, 15] the current results converge to implicate the self-monitor in stuttering. In a direct test of sensitivity to disfluency, PWS proved more likely to consider speech disfluent, and this did not depend on whether the speech was produced by a PWS, or whether we had classified it as disfluent. This result complements Vasiç & Wijnen's findings: whereas their results suggested that cognitive effort and a maladaptive focus play a role in the production of disfluencies, our study, which most likely holds the other two parameters constant, suggests that the third monitoring parameter, threshold, is set higher in PWS than in the control group. Thus, the overall picture that appears from this line of research is that all three monitoring parameters are affected: PWS invest too much effort in monitoring, they focus too much on whether upcoming speech will sound fluent, and are more likely to consider speech as disfluent.

Of course, this interpretation needs to be treated with caution. In particular, we have not considered individual differences within either group. It is possible however that there is large individual variation and in fact, this may the reason why there was a discrepancy between the significance levels in the subject and item analyses on judge: whereas the items were relatively homogenous, there is likely to be substantial individual variation with respect to judges.

The continuity hypothesis was not supported: excerpts from dialogues between PWS were rated as worse than those from PWDNS, regardless of whether they were fluent or not and regardless of who was doing the rating. Indeed, a post-hoc analysis confirmed that not only the disfluent fragments, but also the fluent fragments were rated as worse if they had been produced by a PWS. This corroborates some earlier reports, showing abnormal motor activity in the speech of PWS ([1], [12, 16]). Before rejecting the continuity hypothesis, however, follow-up research will have to address an alternative explanation. The disfluencies on each tape were generated by a limited number of speakers. It is conceivable that the judges classified a certain speaker as a PWS based on a disfluent fragment. Upon hearing a fluent fragment by the same speaker, the judge may have recognized the speaker and showed a bias to judge PWS as more disfluent. We plan to test that explanation in a follow-up study.

Even if the continuity hypothesis turns out to be false, however, it does not necessarily contradict a monitoring explanation. Although (perceptual) abnormalities in speech motor activity may be an aspect of stuttering, they do not explain what we regard as the primary symptom of stuttering: the occurrence of disfluencies. Whereas monitoring hypotheses have no straightforward account for abnormalities in speech motor programming, they do provide an explanation of disfluencies. As this study, along with other studies, has demonstrated, this explanation is testable and has survived the tests to date.

5. Acknowledgements

This research was sponsored by the British Academy.

6. References

[1] Adams, F. R., Freeman, F. J. & Edward Conture. 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.

[2] Anderson, A. H., Marcus Bader, Ellen Bard, E. Boyle, G. Doherty, Simon Garrod, S. Isard, J. Kowtko, McAllister, J. Miller, C. Sotillo, C., H. Thompson & R. Weinert. 1991. The HCRC Map Task Corpus. *Language and Speech*, vol. 34, pp. 351–366.

[3] Bard, Ellen, D. Robertson & Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language*, vol. 72, no. 1, pp. 32–68.

[4] Burger, Remca & Frank Wijnen. 1999. Phonological encoding and word stress in stuttering and non-stuttering subjects. *Journal of fluency disorders*, vol. 24, pp. 91–106.

[5] Dell, Gary. 1986. A spreading activation of retrieval in sentence production. *Psychological Review*, vol. 93, pp. 283–321.

[6] Hartsuiker, Robert, Herman Kolk & Robin Lickley. In press. Stuttering on function words and content words: A computational test of the Covert Repair Hypothesis. To appear in: R. J. Hartsuiker, R. Bastiaanse, A. Postma & F. Wijnen (eds.), *Phonological encoding and monitoring in normal and pathological speech*. Hove (East Sussex): Psychology Press.

[7] Kolk, Herman & Albert Postma. 1997. Stuttering as a covert-repair phenomenon. In: R. Corlee & G. Siegel (eds.), *Nature and treatment of stuttering*: New directions. Boston: Allyn & Bacon, pp. 182–203

[8] Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, Massachusetts: MIT Press.

[9] Meyer, Antje. 1991. The time course of phonological encoding in language production: phonological encoding inside a syllable. *Journal of Memory and Language*, vol. 30, pp. 69–89.

[10] Postma, Albert. 2000. Detection of errors during speech production: a review of speech monitoring models. *Cognition*, vol. 77, pp. 97–131.

[11] Postma, Albert & Herman Kolk. 1993. The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, vol. 36, pp. 472–487.

[12] Van Lieshout, Pascal, Herman Peters, C. Woodruff Starkweather & Wouter Hulstijn. 1993. Physiological differences between stuttererers and nonstutterers in perceptually fluent speech: EMG amplitude and duration. *Journal of Speech and Hearing Research*, vol. 36, pp. 55–63.

[13] Stevens, Stanley. 1975. *Psychophysics*. New York: Wiley.

[14] Vasiç, Nada & Frank Wijnen. Stuttering and speech monitoring.*Proc DISS 2001* 29–31 August 2001, Edinburgh, United Kingdom, vol. 1, pp. 13–17.

[15] Vasiç, Nada & Frank Wijnen. In press. Stuttering as a monitoring deficit. To appear in: R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (eds.), *Phonological encoding and monitoring in normal and pathological speech*. Hove (East Sussex): Psychology Press.

[16] Wood, Sarah. 1995. An electropalatographic analysis of stutterers' speech. *European Journal of Disorders of Communication*, vol. 30, pp. 226–236.