ISCA Archive
http://www.isca-speech.org/archive

ITRW on Disfluency in
Spontaneous Speech (DiSS'01)
Edinburgh, Scotland, UK
August 29-31, 2001

Stuttering and Speech Monitoring

Nada Vasic and Frank Wijnen

Utrecht Institute for Linguistics OTS
Utrecht University, Netherlands
Nada.Vasic@let.uu.nl

Abstract

In ths paper we would like to argue that stuttering represents inadequate monitoring of the speech production process. The model we are proposing is *the vicious circle hypothesis*. The stuttering speakers have a malfunctioning monitor whose three parameters, namely, *focus, effort* and *theshold* are inappropriately set.

In order to test our hypothesis, we tested 20 stuttering individuals in a dual task situation. The experiment consisted of three conditions: baseline where semi-spontaneous speech was elicited and two dual task conditions. First dual task was speaking and playing a computer game at the same time where the processing resources were taken away from monitoring. The second dual task was designed to shift the monitor's focus away from habitual monitoring. Subjects were asked to monitor for a particular word in their speech. The preliminary results of our experiment show that in the dual task condition the number of disfluencies decreased in relation to the number of words, which, in turn supports our prediction that distraction has a positive effect on fluency in the case of stuttering individuals.

1. Introduction

The experiment presented here yields results that support the Vicious Circle Hypothesis which assumes that the abundance of dysfluency in stuttering speakers is due to a malfunction of the monitoring mechanism of the speech production system. We propose a psychological hypothesis that can account for both the primary and secondary characteristics of stuttering, and that is based on a well-established model of the human language-production mechanism, namely the speech production model (Levelt, 1989)[1]. Specifically, we will argue that some of the basic parameters of the monitoring process are maladapted in stuttering speakers.

2. Vicious Circle Hypothesis

The hypothesis proposed here states that in stuttering individuals three 'attention' parameters of the monitor: effort, focus and threshold are inappropriately set. A number of relatively clear and testable predictions can be derived:

(1) Effort. The vicious circle hypothesis argues that stuttering speakers invest an excessive amount of resources in monitoring. A prediction is that if you take away some of the available resources, disfluency will decrease.

(2) Focus. Persons who stutter do not just invest too much energy in monitoring, they do so for a reason, namely to 'look out' for all kinds of signs of realized or imminent disfluency. Their monitoring focus is rigid and unadaptive. The prediction we make here is that if you force them to focus on something other than temporal discontinuity, given the limited resources available, the chance of detecting discontinuities or temporal fluctuations will decrease. Hence, disfluency will decrease. Note that, in principle, focus can be altered independently of effort (it is just redirecting), so the prediction is essentially different from the foregoing.

(3) Threshold. Stutterers entertain acceptability criteria that are too strict. As a consequence, they reject speech output that other, non-stuttering people would consider completely normal. Since we assume that monitoring is based on normal perceptual processing, we predict that also in listening to speech from someone else, stutterers will entertain a more conservative standard with respect to disfluency than non-stutterers.

3. Experiment

In this experiment we concentrated on the first and the second parameter. The experiment was a dual task study, aimed at supplying further evidence for the assumption that persons who stutter do benefit from distraction of attention by a secondary, and highly demanding task, and from reorienting the monitor's focus.

3.1. Subjects

We conducted an experiment on 30 individuals, 20 stutterers and 10 non-stutterers. We defined stutterers as persons who were diagnosed as such by a speech therapist, and who considered stuttering to be a problem at the time of the experiment. The subjects ranged from mild to severe stutterers.

3.2. Tasks

The experiment comprised two conditions, namely, speech only and a dual task conditions in which speaking was accompanied/combined with one of the two tasks described below. In the primary task semi-spontaneous speech was elicited. Subjects were asked to read a newspaper article prior to each condition in order to retell the article's content. In case they stopped talking, the experimenter would call out a cue word/topic, such as 'holidays', 'family', 'hobbies' etc., and the subjects had to elaborate on the topic.

The first distraction task was supposed to take away general resources so that the monitor would be left with less energy to invest into speech monitoring. This distraction task was a computer game 'PONG'; subjects had to play table tennis against the computer. Each subject was screened for 'PONG' so that the level of the game could be tailored to the level of each individual subject.

As a contrast to the first distraction task, the second distraction task was designed to alter monitor's focus. It involved monitoring for a particular word in speech production. The subjects were asked to closely monitor their speech for the word *die* (that – indexical and relative pronoun), which is a fairly frequent word in Dutch. They had to press a button each time they heard *die* in their speech. The computer recorded their response in order to check whether they were paying attention to the secondary task.

Each experimental condition was 10 minutes long. The first condition was a baseline condition (Speech only) in which the subjects were asked to retell the story they read in a newspaper article with no distraction. In the second condition (Dual difficult) they were instructed to retell the story and play PONG at the same time. This was the condition in which the distraction task was demanding and where the level of the game was constantly increasing in its level. The third condition (Dual simple) was also a double task condition, however, the distraction task was not as demanding as in the previous condition; subjects were invited to speak and play PONG at the same time, the level of PONG was kept constantly at its lowest (speed and acceleration). In both condition two and three the computer kept track of subjects' performance. Finally, the last condition was Monitoring *die*, subjects had to speak and react by pressing a button each time they heard themselves utter *die*.

3.3. Design

Subjects were tested individually and were taped on a digital audio recorder (DAT). After reading the instructions with a short description of the experimental tasks subjects were screened for PONG. This was used to determine the level for PONG in the Dual difficult conditions. The order of conditions was rotated across subjects such that for each block of four subjects four conditions were fully rotated.

3.4. Analysis

The first seven minutes of speech in each condition were transcribed transcribed and coded for disfluencies. Disfluencies were transcribed and classified as blocks, self-corrections, prologations, repetitions, senseless sound insertions, word breaks, unfilled pauses and filled pauses. Several occurrences of one type of disfluency on the same position were counted as one disfluency. However, two disfluencies of a different type that occurred on the same segment were coded twice.

Transcribing and coding was done in the CLAN program originally designed for the CHILDES database. In the case of our experiment we needed to count the number of words and the number of dysfluencies per condition. Disfluency is, therefore, determined in terms of the number of disfluencies per condition in relation to the number of words uttered in each condition. Filled pauses were coded but were not included in the analysis. These pauses are very common in the speech of fluent individuals and are therefore not treated as indicators of stutter-moments.

The null-hypothesis predicts no difference in the number of disfluencies between the three conditions across subjects relative to the number of words. Our model predicts a decrease in the mean number of disfluencies in the dual task conditions depending on the success with which the secondary tasks take away processing resources or manage to shift monitor's focus.

3.5. Results

It is well known that the population of stutterers is rather heterogeneous, stuttering ranges from very mild to extremely severe. It is conceivable that the effects of our experimental manipulations could interact with stuttering severity. However, the exact effect of heterogeneity in our sample was unknown beforehand, therefore we had no specific expectations on whether the results might be affected by subject variables. Additionally, the number of words produced by different subjects in different conditions varied. Therefore, we had to deal with two sources of variability in our experiment that could potentially affect the dependent variable. In order to find out how this construct affects the measurement of interest, the amount of disfluency in stuttering individuals, we performed a multilevel analysis (Goldstein, 1995)[2] on the data. This particular statistical analysis is optimally tailored to a full exploration of our data since it allows an explicit modelling of hierarchical and nested relationships between observations. The multilevel analysis allows a more complex structure of the error variances so that both differences between subjects and within subjects (between different conditions) can be removed from the residual error variance (te Riele, 1999)[3].

We wanted to check whether the number of disfluencies differed across conditions. To do so, the amount of disfluency had to be calculated in relation to the number of words produced in each condition. To this end, each individual word a particular subject produced was coded as fluent or as not fluent. A two-level regression model of the MLN computer program (Prosser, Rasbash and Goldstein, 1995)[4] was used. Its output consists of two parts: one is the fixed part of the model that gives estimates of proportions of responses, and the other one, which is the random part of the model that gives the estimates of the variances from the estimated means. Consequently, differences between estimated means were calculated and their significance was determined on the basis of chi-square.

Table 1:

Parameter estimates (logits) and proportions per type of dysfluency per total number of words

	Pong difficult		**Pong simple**		**Speech only**		**Monitoring *die***	
	Logit	%	Logit	%	Logit	%	Logit	%
BLK	-2.273(.268)	.09	-2.294(.268)	.09	-2.205(.268)	.10	-2.852(.269)	.05
COR	-4.921(.193)	.007	-4.997(.195)	.006	-4.597(.186)	.009	-4.548(.185)	.01
PRL	-7.167(.576)	.0007	-5.812(.511)	.002	-6.427(.531)	.002	-5.895(.514)	.002
REP:								
Cpx	-5.821(.600)	.002	-5.860(.600)	.002	-5.651(.597)	.003	-5.107(.589)	.006
Isg	-4.399(.225)	.01	-4.449(.226)	.01	-4.550(.227)	.01	-4.140(.228)	.02
Isy	-7.114(.488)	.0008	-6.718(.456)	.001	-7.158(.491)	.0007	-7.767(.565)	.0004
Wrd	-3.571(.167)	.027	-3.495(.166)	.029	-3.379(.165)	.033	-3.213(.164)	.039
Wst	-4.131(.284)	.016	-4.232(.285)	.014	-4.113(.284)	.016	-3.675(.281)	.025
SSI	-5.751(.269)	.003	-5.314(.252)	.005	-5.009(.245)	.007	-5.069(.246)	.006
UPS	-5.927(.399)	.003	-5.884(.397)	.003	-5.954(.399)	.003	-5.471(.386)	.004
WBR	-4.988(.325)	.007	-5.247(.329)	.005	-4.937(.324)	.007	-4.998(.325)	.007
TOTAL	-1.575(.124)	**.17**	-1.559(.124)	**.17**	-1.541(.124)	**.19**	-1.575(.124)	**.17**

NOTE: **BLK** = block, **COR** = self-correction, **PRL** = prolongation, **REP:cpx** = complex repetition, **REP:isg** = repetition of an initial segment, **REP:isy** = repetition of an initial syllable, **REP:wrd** = repetition of a word, **REP:wst** = repetition of a word string, **SSI** = senseless sound insertion, **UPS** = unfilled pause, **WBR** = word break.

In *Table 1*, the parameter estimates are presented per condition. More specifically, the mean number of disfluencies was calculated for each condition and is represented as totals in the table. The calculations were performed in logits, which are means expressed on a scale different than the scale in which the observations were made. The logit of a proportion P is defined as log (P/1-P). In order to convert the calculated means back to values that are interpretable, proportions calculated from the logits. If we look at the total proportions, it is obvious that the baseline condition (Speech Only) has a higher proportion of disfluencies in comparison to the other three conditions. This difference is small, nevertheless, it is significant. It is essential to note that the strength of our experiment lies in the number of observations made - words that were produced fluently or non-fluently. The total number of observations (words for all conditions for all subjects) for the whole sample that we analysed was 55177.

We found a significant difference in the mean number of disfluencies between the Pong Difficult condition and Speech Only condition (χ^2_i = 15.25, df=1 (p<. 01)). In the baseline condition (Speech Only) subjects produced more disfluencies than in the dual task difficult condition. A significant difference was also found between Pong Simple condition and Speech Only condition (χ^2_i = 11.98 df=1 (p<. 01)). Once again, there were more disfluencies in the baseline condition as opposed to the simple dual distraction condition. Finally, a significant difference was also found between the baseline condition and Monitoring *die* condition χ^2_i = 15.51 df=1 (p<. 01).

There were 11 different types of disfluencies, and as can be seen in Figure 1, some of these occurred very infrequently. The most frequent ones were **bloks** and **repetitions. Blocks** were the most frequent type of all with close to 10% occurrence per total number of words across conditions. There was also a significant difference in the mean number of blocks between all dual conditions and the baseline - Pong Difficult condition vs. baseline (χ^2_i = 4.27 (p<. 01)); Pong Simple condition vs. baseline (χ^2_I = 4.80 (p<. 01)); Monitoring *die* condition vs. baseline (χ^2_I = 189.90 (p<. 01)). In all instances the number of blocks increased in the baseline condition as opposed to other conditions

Repetitions were split into 5 sub-categories, which were analysed as different types. Both repetitions of **initial segments** of **initial syllables** were rather infrequent in all conditions. Only in the distraction condition where the subject had to pay attention to *die* we found an increase in the number of disfluencies for both types. **Complex** repetitions were also very infrequent (average .08%) and no difference was found between the conditions. **Word** repetitions were more numerous (average 13%) with a significant difference between Pong Difficult condition and the baseline condition (χ^2_I= 8.68 p< .01). It is interesting to note that there were significantly more disfluencies in the Monitoring *die* vs. all other conditions. Similar results were found for **word string** repetitions which occurred also slightly more often than word repetitions. There was an increase in the number of word string repetitions in the Monitoring *die* condition. Other types exhibited no clear pattern in relation to our predictions.

4. Discussion

Our hypothesis makes two crucial predictions with regard to the experimental study that we conducted. First one is related to the general processing resources that the monitor uses while keeping track of the speech production process. These resources can be taken away from monitoring by means of a demanding secondary task. By doing so, monitoring should become less excessive in the case of the stuttering individuals and, therefore, disfluency is expected to decrease. Second prediction is related to the focus of the monitor during speech production. We proposed that the monitor habitually over-focuses on the temporal characteristics of speech in the stuttering individuals. If we somehow push away the monitor from its habitual focus the speech should result in less disfluencies. In this case we are not taking away the resources it needs for processing but reorienting the monitor.

From the results obtained it can be concluded that it is indeed the case that when distracted, stutterers produce less disfluencies. Many previous studies failed to reproduce the same results, which many claimed to be a consequence of a secondary task that was not sufficiently engaging. In order to verify whether the effects of a dual task on speech fluency is due to the degree to which the secondary task takes away resources the first secondary task (PONG) was varied. Our study shows that even a slight distraction could be beneficial. In the case of the most simple computer game (Pong Simple condition) where the level was at its lowest and was kept constant subjects produced more fluent speech. It is also important to emphasise that the type of distraction is not crucial. Our results show that fluency increases with general distraction in the case of the computer game, but also in the experimental condition where the focus of the monitor was shifted towards attributes of the speech other than its temporal characteristics.

We found that the different types of disfluencies responded differently to the two types of secondary tasks. The frequency of blocks decreases in all three dual task conditions (Pong difficult, Pong easy, *die*-montoring). By contrast, the amount of repetitions dropped in the PONG conditions, but appears to raise in the *die*-monitoring condition. Blocks are known to be the most 'pathological' symptom of stuttering (i.e.,they are not among the disfluencies that can be considered normal). Note that in many clinical analyses of the development of stuttering in children (e.g. McDearmon, 1968; Yairi and Lewis, 1984)[5][6] all other types of disfluencies with the exception of blocks occur in both stuttering and non stuttering children. The emergence of blocks is considered to be an ominous phenomenon. Once the child starts to block, there appears to be no way back, the 'physiological' disfluency of the immature child has turned into a problem. Particularly, it has been argued (Johnson 1956)[7] that the emergence of blocks is correlated with the child's emerging awareness of his disfluency. We speculate that blocking is a learned response to the self-perception of (imminent) disfluency, which arises and remains partly under conscious (attentive) control.

Reiterating previously articulated material (repeating), on the other hand is a natural response of the language production system to trouble in planning or delivery. We may assume that part of the repetitions in stuttering are these 'normal' reactions of the language production system. Another part may be due the maladaptive monitoring process that we hypothesize. Naturally occurring disfluency is sensitive to 'amount of work' in the production system: we have seen that disfluency rises as the speech task gets more difficult.

On the one hand, we have the perceptuo-motor secondary task (PONG) which takes away processing resources across-the-board. When performing this task stutterers are prevented from following their habit. In comparison to the baseline condition (speaking only), in the perceptuo-motor task there is no additional 'pressure' within the language production system, which has a beneficial effect on the speech of stutteres. The focus-redirecting monitoring task, on the other hand, appears to have two simultaneous effects: (1) what it's supposed to do, namely drawing the stuttering person's monitor away from what it normally focuses on, and (2) by doing so increasing the load on the production system. Consider what it means to be instructed to explicitly report every occurrence of particular word in your speech output. It means continuous controlled, attentive monitoring. It is very likely that this will interfere with normal speech planning and delivery, and that, therefore, the number of "normal" disfluencies will rise, and more so than the distracting effect will suppress them. From this perspective, it is very meaningful that the decrease in this condition is in the non-normal type of disfluency, whereas the rise is in the class of disfluencies that (at least in part) can be considered normal. This observation may provide support for our interpretation that in fact two processes co-occur, one which affects the normal "healthy" part of the language production system (→ more repetitions) and one which affects the "pathological" part, what we have named maladaptive monitoring (→ less blocks).

From our results we can conclude that even the perceptuo-motor secondary task in its easy form helps in bringing down the number of disfluencies in stuttering speakers. This seems to go with the suggestion made by us and others, (Thompson, 1985)[8] that the effectiveness of a secondary task hinges on its capacity to continuously engage attentive processing. We expected that our easy task would fail to do so, but the results suggest differently. It is likely that we have underestimated the demands made by the PONG game, particularly for relatively inexperienced players. Even when played at a low level, it may very well be that PONG is highly engaging. At the very least, the game requires a certain amount of visual attention in order for it to be played successfully.

5. Conclusions

The results of our experiment support the claim that stuttering is a consequence of maladaptive monitoring during speech production. Taking away the processing resources or shifting the focus of the monitor results in a more fluent speech.

6. References

[1] Levelt, W.J.M. *Speaking:From Intention to Articulation.* Cambridge, MA:MITPress, 1989.

[2] Goldstein, H.(1995). Multilevel Statistical Models. London:Arnold.

[3] Riele, S. te, (1999). *Early Context Effects in Spoken-word Perception.* Doctoral dissertation. Utrecht University, The Netherlands.

[4] Prosser, R., Rasbash, J. and Goldstein, H. (1995) *MLN software for multilevel analysis*. London: University of London, Institute of Education.

[5] McDearmon, J.R.(1968). Primary stuttering at the onset of stuttering: A reexamination of data. *Journal of Speech and Hearing Research, 11,* 631-637.

[6] Yairi, E. and Lewis, B. (1984). Disfluencies at the onset of stuttering. *Journal of Speech and Hearing Research, 27.* 154-159.

[7] Johnson, W. (1956). Stuttering. In W.Johnson, S.J. Brown, J.J. Curtis, C.W. Edney and J. Keaster (Eds.), *Speech in handicapped school children, revised edition.* New York: Harper & Brothers.

[8] Thompson, A.H. (1985). A test of the distraction explanation of disfluency modification in stuttering. *Journal of Fluency Disorders, 10*, 35-50.