Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 193 (2015) 217 – 222

Procedia
Social and Behavioral Sciences

10th Oxford Dysfluency Conference, ODC 2014, 17 - 20 July, 2014, Oxford, United Kingdom

Accuracy of articulation rate control with visual feedback in persons who do and do not stutter

Keiko Ochi*, Koichi Mori*, Naomi Sakai and Jun Aoki-Ogura

Research Institute of National Rehabilitation Center for Persons with Disabilities, Tokorozawa, Saitama, 359-8555, Japan

Abstract

The ability to control speech rate with real-time visual feedback was compared between people who do and do not stutter (PWS/PWNS). Nine PWS and 7 PWNS participated in the study. Fifteen sentences were read aloud after repeating a played-back sentence twice in each of 6 trials at 6 different target speeds. The 6 trials comprise a session, and there were 3 sessions (A1, B, A2) with only the second session (B) accompanied by real-time visual feedback of the subject's speech rate and the target speed. The speech rate excluding pauses or dysfluencies was significantly reduced in B and A2 from that in session A1. Although there was no difference in speech rate between B and A2, (a) there was an interaction between the target rate and the group in session B, and (b) the variability in the error of the PWS was larger than that of the PWNS in the retention session (A2). These results suggests (a) that at least some of the PWS use a different strategy in controlling their speech rate than PWNS, and (b) that some of the PWS were less accurate in retaining the learned speech rate in the previous session B with visual feedback than the PWNS, although they did use the visual feedback information and learned the speech rate, to a similar averaged accuracy during the feedback.
© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the Scientific Committee of ODC 2014.

Keywords: Stuttering; Speech rate; Articulation; Visual feedback; Speech control; Fluency shaping

1. Introduction

Fluency shaping as part of stuttering therapy usually includes reducing speaking rate, gentle voicing onset, and gentle contact (e.g. Guitar, 2013). It is, however, difficult for people who stutter (PWS) to recognize the appropriateness of their speech rate by themselves while they speak. Both of fast and slow fluent speaker groups

* Corresponding author. Tel.: +81- 4-2995-3100; fax: 81-4-2995-3132.
E-mail address: ochi.rh.02011@gmail.com

1877-0428 © 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the Scientific Committee of ODC 2014.
doi:10.1016/j.sbspro.2015.03.262

uttered at approximately two thirds of their normal rate, with a large individual variance, when they were instructed to speak at 50 percent of their normal speaking rate (Tsuao, Weismer & Iqbal, 2006). In an attempt to control the speaking rate (Curlee & Perkins 1969; Ingham, Martin & Kuhl, 1974), speakers were alerted when their speaking rate became too fast. There has been research involving computer-assisted home practice systems which enable them to monitor whether their speech rate is appropriate (Webster, 1980; Euler, Gudenberg, Jung & Neumann, 2009). However, this kind of research for Japanese PWS has been scarce or unsystematic. We examined the short-term effectiveness of visual feedback on the control of reading rate with automated speech recognition technology, and compared the results with people who do not stutter (PWNS).

2. Method

2.1. Participants

Nine PWS (six men and three women whose age ranged from 22 to 36 years) and seven PWNS (six men and one woman whose age ranged from 22 to 23 years) who were all native Japanese speakers, participated in the experiment after written informed consent. This study was approved by the ethical committee of the National Rehabilitation Center for Persons with Disabilities (NRCD).

2.2. Reading material

Participants read aloud a single passage consisting of 15 sentences in each trial. The length of the sentences was 29.1 morae long on average. Each trial was preceded by a learning phase of the target speed in which a sentence that was different from any of the 15 sentences was presented auditorily.

2.3. Task

The experimental design is illustrated in Figure 1. Each participant underwent three sessions (A1, B, and A2), each consisting of six trials of a reading task. Only in the second session (session B), visual feedback of the articulation rate was presented to the participant. In each trial, participants read aloud the passage trying to keep the speech rate as accurately as the instruction phrase after repeating it twice before the reading phase. The trial was repeated six

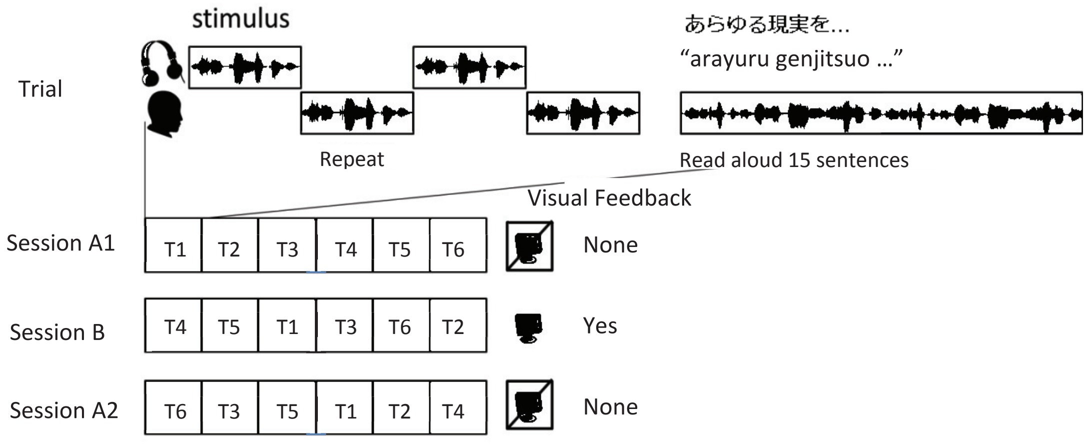

Figure 1. Experimental design
Upper panel: Each trial starts with stimulus and repetition (twice), followed by reading 15 sentences. The model speech instructed the target speech rate. Lower panel: Each session included 6 trials with different target speeds in a randomized order. Visual feedback of speech rate with a reference line showing the target speed was presented only in session B.

times with six different target rates in a randomized order in a session. The stimuli (model speech) for the instruction phrase were made by digitally extending a recorded sentence without changing its pitch or spectral contents. The measured articulation rates of the model utterances were (1)3.91, (2) 4.47, (3) 5.02, (4) 5.57, (5) 6.13, and (6) 6.73 morae/s.

2.4. Visual feedback of articulation rate

The articulation rate of the participant's speech was estimated on-line and off-line with open-source speech recognition software "Julius" (Lee & Kawahara, 2009) which is capable of determining the kinds and timings of uttered phonemes and words in running speech. The articulation rate in morae/s was calculated as the numbers of morae spoken between pauses divided by the time for the segment. The articulation rate was displayed as the number of asterisks for the visual feedback purpose. One asterisk represented 0.25 mora/s. If a row of the asterisks exceeded the vertical line of the target speed, the participant was required to read more slowly, and they needed to speak faster if the row of the asterisks were less than the target speed.

2.5. Data analysis

The articulation rate was calculated in a similar way as the procedure for obtaining the speed for the visual feedback. A partially different procedure was used in that the last mora prior to a pause was omitted from the calculation of the speech rate, because it usually stops short of a full mora length. Although this modification, only possible during the off-line analysis, should produce more accurate estimates of the speech rate, the difference from the on-line method may be small when the blocks of utterances between pauses were more than several morae long, which was mostly the case with the current study. The results were analyzed by 3-way analysis of variance (ANOVA) with factors: group, session, and target speech rate. We performed post hoc analysis with Holm's adjustment at the $p < 0.05$ level. The absolute difference from the target of each group was compared by Wilcoxon.

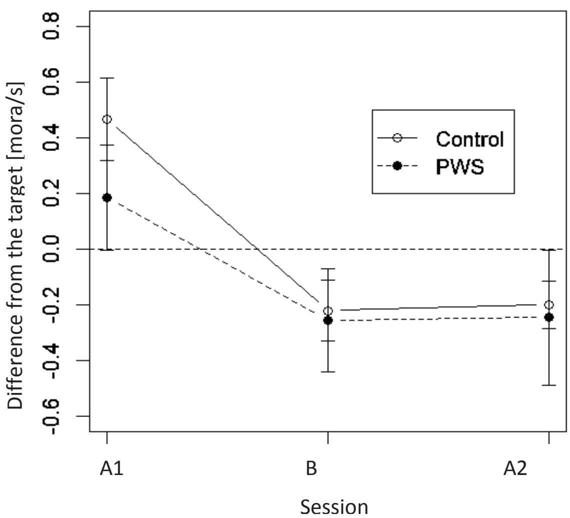

Figure 2: Averaged difference from the respective target articulation rates.
Bars indicate standard errors of the subject-wise means.

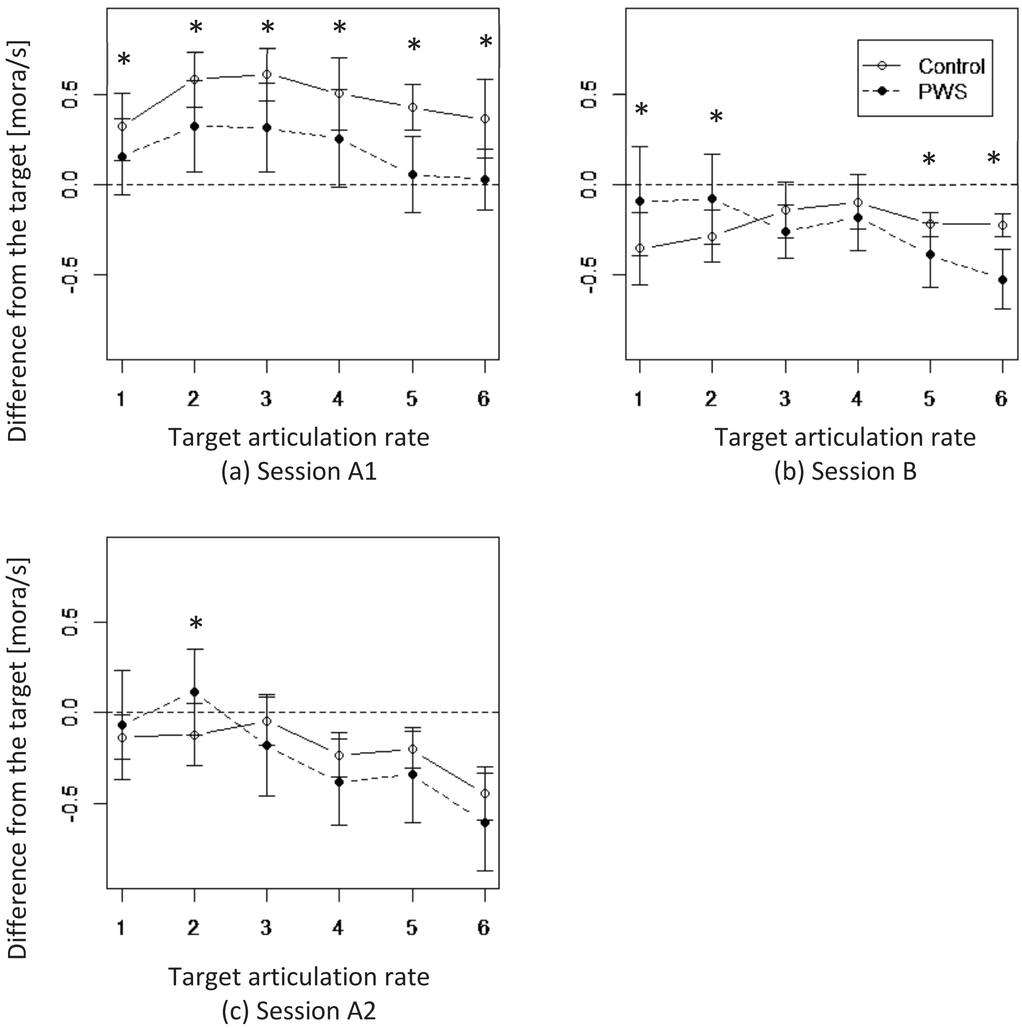

Figure 3: Group differences from the 6 respective target articulation rates in 3 sessions. Bars indicate standard errors of the subject-wise means, *: $p < 0.05$, between group post-hoc test.

3. Results

The three-way ANOVA revealed significant main effects on all of group ($F(1, 4294) = 282.19$, $p < 0.05$), session $F(2, 4294) = 120.36$, $p < 0.05$) and target speech rate ($F(1, 4294) = 18.96$, $p < 0.05$). There were significant interactions between group and session ($F(2, 4294) = 7.34$, $p < 0.05$), between group and target speech rate ($F(1, 4294) = 19.45$, $p < 0.05$), and between session and target speech rate ($F(2, 4294) = 17.67$, $p < 0.05$). There was also a significant three-way interaction effect ($F(2, 4294) = 9.58$, $p < 0.05$). Post hoc tests showed that the PWS read the texts significantly slower than the control group. The speech in sessions B and A2 was significantly slower than that in session A1 in both groups as is shown in Figure 2. Articulation rates were slowed down by the visual feedback, and the effect remained after the visual feedback was removed in session A2 in both groups.

Figure 3 demonstrates the interaction of the differences from the respective target rates with the target rates and the sessions. The PWS read the task passage significantly slower than the PWNS for all the target rates in session A1. In session B, however, the PWS read significantly faster for targets 1 and 2, but slower for targets 5 and 6 than the PWNS. PWS read relatively slower when the target was slower, whereas the PWNS read relatively more slowly when referenced to the target speed as the target rate became faster. In session A2, there was a significant difference only for target 2.

The mean differences (errors) from the respective target speeds of the individual participants are shown in Figure 4. The variance of the PWS increased in session B and A2. Four of the PWS had slower articulation rates than the slowest one in the group of PWNS in session A2, while one of the PWS was a positive outlier throughout the three sessions. The absolute value of difference from the target was significantly larger in the PWS group than in the PWNS group in session A2.

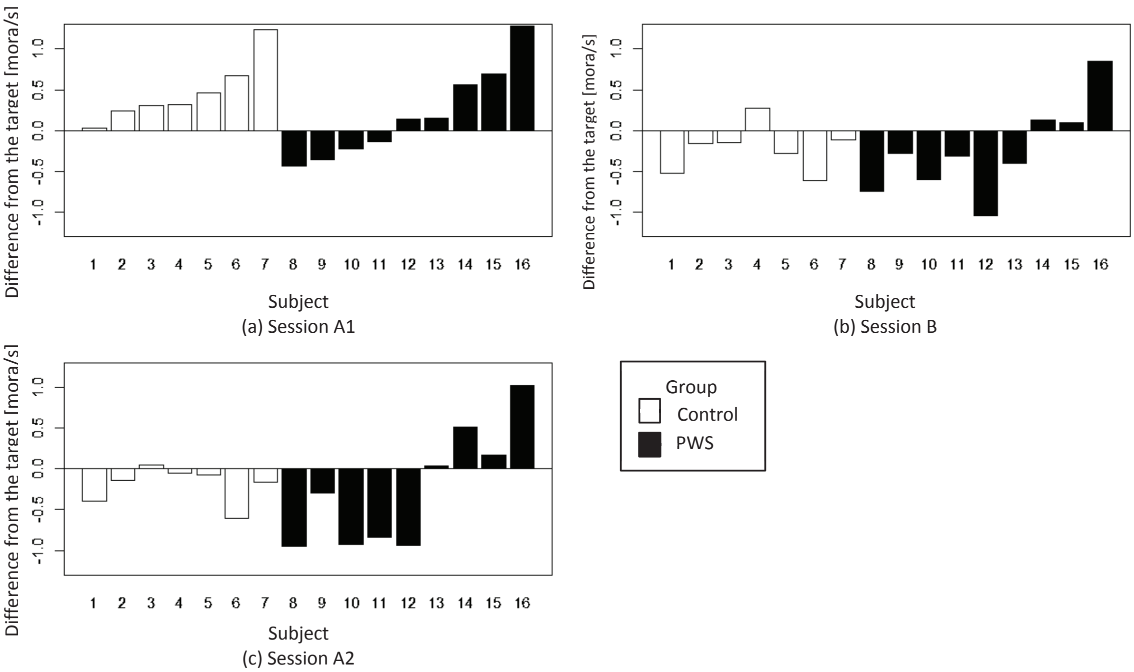

Figure 4. Difference from the target rate of individual participants.
Numbers below each panel are subject IDs that are common among the three sessions.

4. Discussion

Behavioral changes occurred after the visual feedback of reading rate in both groups, which suggests the effectiveness of visual feedback over simple repetition learning of speech rate, which was used in the A1 and later sessions. Since there was no difference in the speech rate between sessions B and A2, it is assumed that the speech rate was learned during Session B with visual feedback, at least for the short term in both groups, although the averaged achieved speech rate was somewhat different between the groups. The interaction of the accuracy of the reading rate between target rate and the group in session B (Figure 3) suggests that the PWS took different strategies from those of the PWNS when visual feedback was presented.

Although on average the PWS and PWNS performed similarly in terms of the overall error from the target speech rate, examining the individual differences revealed that some among the PWS have more difficulty in controlling their articulation rate than the PWNS even with the visual feedback. The one PWS who showed the largest positive error in the speech rate seems to have difficulty reducing the speech rate, and was an outlier even within the PWS group. The PWS on average showed as good learning ability of the reading rate as the PWNS, judging from the non-significant difference between the B and A2 sessions (Figure 2). The majority of the PWS, however, showed either more negative or more positive extreme errors than the PWNS (Figure 4). This result seems to suggest that they also use a different strategy than PWNS in adjusting their speech rate after visually learning their speech rate, and that they find it harder to retain their learnt speech rate.

References

Curlee, R. F., & Perkins, W. H. (1969). Conversational rate control therapy for stuttering. *Journal of Speech & Hearing Disorders.*

Euler, H. A., Gudenberg, A. W. V., Jung, K., & Neumann, K. (2009). Computer-assisted therapy for speech disfluency: The long-term effectiveness of the Kassel stuttering therapy (KST). *Sprache, Stimme, Gehör*, 33(04), 193-202.

Guitar, B. (2013). Stuttering: An Integrated Approach to Its Nature and Treatment. 4th edition. Baltimore, Lippincott Williams & Wilkins.

Ingham, R. J., Martin, R. R., & Kuhl, P. (1974). Modification and control of rate of speaking by stutterers. *Journal of Speech, Language, and Hearing Research*, 17(3), 489-496.

Lee A. and Kawahara, T. (2009). Recent Development of Open-Source Speech Recognition Engine Julius, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (*APSIPA ASC*), 2009.

Tsao, Y. C., Weismer, G., & Iqbal, K. (2006). Interspeaker variation in habitual speaking rate: Additional evidence. *Journal of Speech, Language, and Hearing Research,* 49(5), 1156-1164.

Webster, R. L. (1980). The precision fluency shaping program: Speech reconstruction for stutterers. *Communications Development Corporation.*