EXPLORING GLOBAL AND LOCAL ARTICULATION RATE ENTRAINMENT IN TYPICAL AND ATYPICAL SPEAKERS

Lotte Eijk[a)], Stefany Stankova[b)], Sophie Meekings[c)]
Department of Psychology, University of York, United Kingdom
[a)]lotte.eijk@york.ac.uk [b)]ss3345@york.ac.uk [c)]sophie.meekings@york.ac.uk

ABSTRACT

Entrainment, in which talkers adapt their speech to become more similar to their interlocutor's, seems to be a persistent phenomenon among typical speakers. However, it is not clear how this behaviour is affected when one speaker has difficulty controlling their speech production, for example because they stutter. We explore whether typical speakers entrain their Articulation Rate differently to people who stutter than to fluent speakers. Twenty participant pairs (10 typical-typical, 10 typical-atypical) participated over Zoom, completing a solo picture description, and three rounds of a spot-the-differences Diapix task in interaction. Local entrainment was defined as turn-by-turn entrainment. Global entrainment was investigated by comparing the Articulation Rate of speakers of a pair between Diapix rounds, and the solo picture description. Results indicate that speakers entrained locally. No evidence for global entrainment nor differences between the two groups were found. Overall, this suggests speakers entrain locally to both typical and atypical speakers.

Keywords: entrainment; articulation rate; stuttering; interaction.

1. INTRODUCTION

Speakers in conversation tend to change their speech to become more like that of their interlocutor. This phenomenon is, among other things, referred to as entrainment, alignment, or accommodation. Here we will use the term *entrainment* to refer to speakers changing their speech to match their interlocutor's speech more closely. Entrainment has been thoroughly investigated on different linguistic levels in typical speakers (e.g., phonetic entrainment [1], syntactic entrainment [2], lexical entrainment [3]), and can occur over the course of the whole interaction (global entrainment) or turn-by-turn (local entrainment). However, studies on entrainment are less prevalent in atypical populations. It is thus not clear whether speakers entrain to atypical speech to the same extent as they do to typical speech.

Entrainment has been found to relate to a range of positive aspects, such as greater task success [4] or higher likeability [5]. Therefore, it is important to understand whether disabilities that affect speech production also influence conversational entrainment, as this may affect how speakers are perceived, with implications for their general quality of life. As hypothesised by Borrie and Liss [6], impaired entrainment abilities may partially contribute to intelligibility difficulties in conversation for this group of speakers.

This study explores Articulation Rate (AR) entrainment between typical speakers and people with persistent developmental stuttering. Stuttering is a neurological speech condition that causes involuntary syllable repetitions, prolongations, and 'blocks' where speakers are unable to produce sounds. This affects speaking and articulation rate [7, 8]. In addition, people who stutter demonstrate difficulty with speech motor skills, which may affect their ability to adapt their speech to different situations [9]. Stuttering is also associated with negative listener perceptions and influences listener's recall of spoken information [10], meaning that the condition presents several challenges to successful spoken interaction.

In this study, we aimed to investigate whether these difficulties affect Articulation Rate entrainment in conversation. Specifically, we asked whether typical speakers, when partnered with a person who stutters, would adapt their own speech more to compensate for their interlocutor's reduced speech motor control. We also aimed to investigate whether entrainment in these interactions occurs at the local or global level, both, or neither.

Studies on entrainment usually investigate local and/or global entrainment in different manners. Even though definitions of what a turn is can differ between studies, local entrainment is commonly investigated as turn-by-turn entrainment [e.g., 2]. There are a larger range of possible ways to measure global entrainment. Most definitions of global entrainment are, however, similar, where global entrainment is defined as speakers becoming more similar over a conversation. Within this definition there is variety in how this measure is employed. Examples of different measures are comparing one half of a conversation to a different half [11] or comparing real pairs to pairs who did not interact [4]. Entrainment is thus investigated in a variety of ways.

Articulation Rate (AR) entrainment has been investigated in many studies in typical speakers [e.g., 5, 11, 12]. These studies show mixed results. Where

Eijk and colleagues [12] find no evidence for local entrainment of AR (number of syllables divided by phonation time – speaking time excluding pauses), and indications of global entrainment, Levitan and Hirschberg [11] find both local and global entrainment in Speaking Rate (the same as AR, but including pauses), and Schweitzer and Lewandowski [5] find overall divergence in AR (but reversed or weakened depending on mutual likeability). Another study, exploring entrainment in an atypical population, investigated Speaking Rate to atypical populations of speakers with dysarthria [6]. The authors found global Speaking Rate entrainment to both typical and atypical speakers in a sentence reading paradigm. However, speakers entrained more to typical speakers.

To get further insight into entrainment to different atypical populations in a semi-naturalistic setting, the research question was studied in a spot-the-differences task with typical speakers in interaction with speakers who stutter over Zoom, an online meeting application. AR is defined as the number of syllables in an utterance divided by the phonation time (time excluding pauses). An utterance is defined as an inter-pausal unit, separated by at least 0.5s of silence [e.g., 11]. We define local entrainment as turn-by-turn entrainment and global entrainment as similarity between speakers of a pair in their AR over different pictures.

2. METHODS

2.1. Participants

Forty native English participants participated in this study. The participants were paired into two different groups, resulting in ten pairs of typical-typical speakers (age: *M* = 31.3, *SD* = 12.6) and ten pairs of typical-stuttering speakers (age: *M* = 32.8, *SD* = 12.2). Participants who stuttered were self-identified stutterers. The typical-typical pairs consisted of 3 female-male, 6 female-female, and 1 male-male pairs, and the typical-stuttering pairs were 7 female-male, 2 female-female, and 1 male-male pairs.

2.2. Tasks

Speakers completed multiple experimental parts over Zoom. The first day, they completed different tasks by themselves, including a solo picture description. In this task, participants were asked to describe a picture from a spot-the-differences task - the Diapix task [13] - by themselves, starting from the top left corner and continuing clockwise.

On a different day, participants joined over Zoom again to participate in the Diapix task. In this task, participants were asked to turn off their camera to only provide the interlocutor with information via speech. They were furthermore asked to use a separate microphone and headphones. The interactional task had a mean duration of 29.84 min (*SD* = 2.90 min) over all participants.

2.2. Procedure

After doing the solo picture description on the first day, on a different day participants completed three rounds of the Diapix task, and thus described three different pictures to each other (Beach, Farm and Street). Pairs were divided into a participant *A* and a participant *B*. Participant A was the leader in Round 1 and 3, and participant B was the leader in Round 2. In the pairs with a participant who stuttered, this participant was always participant A. Being the leader entailed starting the picture description. The other participant was the follower in that round and was instructed to notify the other participant when something seemed different in their picture and together find the differences. There was a time limit of 10 minutes per picture and each picture contained 12 differences to be identified.

2.3. Pre-processing

2.3.1. Audio pre-processing

Since data was collected online and varied quite drastically in quality, pre-processing was necessary to conduct analyses. This process consisted of six steps. The first step was to use dynamic time warping with a custom Python (version 3.9.12) script to align the separate mono wav files from the participants to be time aligned to a wav file containing both participants' speech. These files were then combined to a stereo file with one channel for each speaker and cut up into separate files per described picture. These were split into mono again to reduce the noise per participant with a custom script in Praat [14] using the "Reduce noise…" function. Next, we used the "To TextGrid (silences)…" function in Praat with individual optimal settings per participant to separate the utterances. Minimum pause duration was set to 0.5s for all participants to divide the audio into inter-pausal units (IPUs; commonly divided by a minimum pause of 0.5s [e.g., 11]). The resulting TextGrid files were manually checked by two trained researchers. Stutters on a word were annotated to be included in the same IPU as the word itself. The last step was to extract all separate IPUs, taking into account the starting time within the file.

2.3.2. Articulation Rate extraction

Articulation Rate (AR) was calculated per IPU with a script [15] in Praat. The standard settings for the script were used (Silence threshold (dB): -25, Minimum dip near peak (dB): 2, and Minimum pause duration (s): 0.3). After extraction of the AR, IPUs with less than 5 syllables were removed to avoid including utterances too short for a reliable AR (due to e.g., filled pauses or phrase final lengthening) in the analyses. This resulted in 6971 IPUs for all participants together (about 58 percent of the total).

2.4. Analyses

2.4.1. Global entrainment

To investigate global entrainment, we calculated a single AR per picture description. This score was calculated by dividing all syllables from the different utterances used to describe a picture by the total phonation time per picture. This resulted in AR values per participant per picture. Within pairs, these scores were then subtracted from each other, after which we took the absolute value to calculate a difference score per participant pair (see (1) below). These scores represent how similar a participant A's AR was to that of participant B in the solo picture description and in the different Diapix rounds. A score closer to 0 means that participants' ARs in a round are more similar.

(1) $$\text{Difference score} = |\text{AR}_\text{A} - \text{AR}_\text{B}|.$$

These difference scores were then used to perform two linear mixed-effects regression (LMER) models in R version 4.2.1 [16], using the lme4 version 1.1.31 [17] and car package version 3.1.1 [18]. Ggplot2 version 3.4.0 [19] was used for visualisation. The first LMER model tested for differences between groups (typical-typical pairs versus typical-stuttering pairs) between the solo picture description and the different Diapix rounds and included a random intercept of participant. The second model was the same, but excluding the interaction with group to test for overall differences between the rounds in the whole group.

2.4.1. Local entrainment

After deletion of utterances with less than 5 syllables, utterances of participant B that were preceded by an utterance of participant A were identified. This was counted as consecutive turns and used in the statistical models. This led to a total of 1297 data points for the statistical analyses of local entrainment.

We tested two LMER models in R version 4.2.1 [16]. In the first model, we tried to predict the AR of participant B with the AR of participant A, in interaction with group (typical-typical versus typical-stuttering). This model included a random intercept of participant. The second model was the same model, excluding the interaction with group to test for an overall effect within both groups together.

3. RESULTS

Figure 1 shows the difference scores per participant pair per round of the experiment. Descriptively, the rounds seem to differ from each other. Figure 2 shows the same data points, but now connected with a line for each pair. This shows that the entrainment patterns considerably differ between the different pairs.

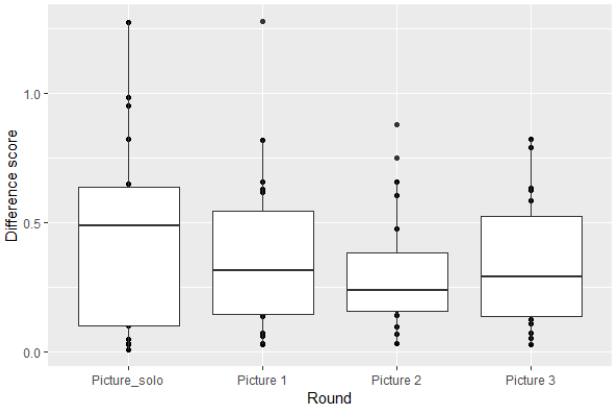

Figure 1: Difference scores between participants' AR in different parts of the experiment.

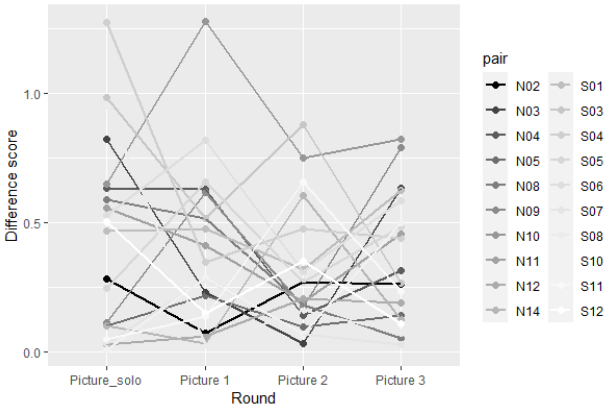

Figure 2: Difference scores between participants' AR in different parts of the experiment. Lines connect each pair. N = typical-typical pair, S = typical-stuttering pair.

3.1. Global entrainment

We performed an LMER testing for differences between the groups between the different rounds (Solo Picture, Diapix Round 1, Diapix Round 2, and Diapix Round 3). These analyses showed no significant effects, as indicated by running an ANOVA from the car package on this model. We then tested the same model, but taking out the interaction with group to test for global entrainment

in the whole group. This showed no significant difference either. This is not too surprising given the diversity of patterns shown in Figure 2.

3.2. Local entrainment

Results of the first LMER predicting the Articulation Rate (AR) of participant B with that of participant A and checking for group effects showed no significant effects. We thus do not find evidence for differences in local entrainment between the two groups.

We then tested a similar model, without the interaction. This model showed a significant effect of the AR of participant A on the AR of participant B (see Table 1). This indicates that over the total group of twenty pairs, we find evidence of local entrainment.

Variable	β	SE	T Value
Intercept	3.55	0.12	29.50
AR_A	0.07	0.03	2.46

Table 1: Output from the local entrainment model.

4. DISCUSSION AND CONCLUSION

This study investigated whether typical speakers entrained more to atypical speakers than to typical speakers, and whether entrainment in Articulation Rate (AR) was reflected in local or global measures. We studied AR in a solo picture description task and in speakers in interaction over Zoom performing three rounds of a spot-the-differences task.

No evidence was found to suggest that speakers entrain their AR differently to speakers who stutter as compared to typical speakers. No entrainment was found globally. However, over the whole group, we found evidence for local entrainment, indicating that speakers changed their AR to more closely match the AR of their interlocutor in the turn prior to theirs.

These findings are in line with Levitan and Hirschberg [11] who found local entrainment of speaking rate. However, the authors also found global entrainment (like [6, 12]). One may assume that speakers must also entrain globally when local entrainment takes place. However, the measure used in this study may have not been optimal to detect global entrainment as found in other studies. Here, we used one AR over a whole picture to get an indication of possible global entrainment. Other ways of measuring global entrainment such as the measurement by Levitan and Hirschberg, who measured the degree to which Speaking Rate became more similar over time between two speakers, may shed more light on global entrainment. Global entrainment will thus be further explored in future studies using the same dataset. Nevertheless, since there is no standard way of measuring global entrainment, this study shows a first insight into this measure in a population of speakers who stutter.

Another reason why our results may differ partly from the literature, and why findings in the literature may be rather divergent, is the possible effect of the task. Since entrainment has been shown to relate to e.g., task success [4] or social factors [11], it is likely that the kind of task influences the outcome.

Moreover, different studies may investigate slightly different measures. As mentioned before, some of the studies cited here investigate Speaking Rate [e.g., 11] instead of Articulation Rate. Another difference in the measurement of both Articulation Rate and Speaking Rate is that one can either use the actually realised syllables found in the speech signal (excluding reduced syllables) or base the rate on the number of syllables that the canonical words contain. The possible differences and the possible effect of these different measures of speech tempo on entrainment patterns should be further investigated to gain insight into what speakers actually entrain to.

Different pairs of speakers showed various entrainment patterns. Stuttering is a heterogeneous disorder [20], and the amount of stuttering in an affected person's speech can vary widely depending on the situation [21, 22]. Differences in overall stuttering severity as well as variation in the amount of stuttering may have affected a pair's ability to entrain; additionally, the online nature of the experiment may have introduced additional challenges to entrainment such as environmental noise or connection problems.

In conclusion, this study advances our understanding of how speech patterns align during communication, particularly for individuals who stutter. A key finding is that local entrainment of articulation rate occurs in both typical-typical and typical-stuttering pairs. This suggests that stuttering may not affect rhythmic conversational entrainment when people who stutter converse with typical talkers. Local entrainment was observed despite the relatively poor audio quality of Zoom compared to in-person conversation. This study thus shows the potential of using Zoom data to test populations that may be difficult to assess in the lab, even though these data pose audio quality challenges. Although no significant global entrainment was found, the individual variation in entrainment patterns suggests that factors such as stuttering severity may play a role in how pairs align to each other. These findings could have implications for how clinicians and researchers assess and treat individuals with stuttering.

6. REFERENCES

[1] Pardo, J.S., 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), pp.2382-2393.

[2] Branigan, H.P., Pickering, M.J. and Cleland, A.A., 2000. Syntactic co-ordination in dialogue. *Cognition*, *75*(2), pp.B13-B25.

[3] Brennan, S.E. and Clark, H.H., 1996. Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, memory, and cognition*, *22*(6), p.1482.

[4] Reitter, D., and Moore, J. D., 2007. Predicting success in dialogue. In A. Zaenen & A. van den Bosch (Eds.), *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 808–815). Prague, Czech Republic: Association for Computational Linguistics.

[5] Schweitzer, A. and Lewandowski, N., 2013. Convergence of articulation rate in spontaneous speech. In *INTERSPEECH* (pp. 525-529).

[6] Borrie, S.A. and Liss, J.M., 2014. Rhythm as a coordinating device: Entrainment with disordered speech. *Journal of Speech, Language, and Hearing Research*, *57*(3), pp.815-824.

[7] Hall, K., Amir, O., and Yairi, E. 1999. A longitudinal investigation of speaking rate in preschool children who stutter. *Journal of Speech, Language, and Hearing Research*, *42*, 1367–1378.

[8] de Andrade C.R., Cervone L.M., Sassi F.C., 2003. Relationship between the stuttering severity index and speech rate. *Sao Paulo Med J. 121*(2):81-4. doi: 10.1590/s1516-31802003000200010.

[9] Namasivayam, A.K. and van Lieshout, P., 2011. Speech Motor Skill and Stuttering. *Journal of Motor Behavior*, *43*(6), 477-489.

[10] Healey, E., 2010. What the Literature Tells Us about Listeners' Reactions to Stuttering: Implications for the Clinical Management of Stuttering. *Seminars in Speech and Language*, *31*(04), 227–235.

[11] Levitan, R. and Hirschberg, J.B., 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland, OR.

[12] Eijk, L., Ernestus, M., Schriefers, H., 2019. Alignment of pitch and articulation rate. *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia (2019), pp. 2690-2694.

[13] Baker, R. and Hazan, V., 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior research methods*, *43*(3), pp.761-770.

[14] Boersma, P. and Weenink, D., 2022. Praat: Doing phonetics by computer (Version 6.2.23).

[15] De Jong, N.H., Pacilly, J., & Heeren, W., 2021. PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically, *Assessment in Education: Principles, Policy & Practice*, 28:4, 456-476, DOI: 10.1080/0969594X.2021.1951162

[16] R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[17] Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01.

[18] Fox, J. and Weisberg, S., 2019. An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

[19] Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

[20] Stager, S. V., 1990. Heterogeneity in stuttering. *Journal of Fluency Disorders*, *15*(1), 9–19. doi:10.1016/0094-730x(90)90028-q

[21] Schwartz, H. D., & Conture, E. G. 1988. Subgrouping Young Stutterers. *Journal of Speech Language and Hearing Research*, *31*(1), 62.

[22] Bloodstein, O., 1950. Hypothetical conditions under which stuttering is reduced or absent. *Journal of Speech & Hearing Disorders*, *15*, 142–153.