CHARACTERISTICS AND DISTRIBUTION OF SILENT PAUSES IN CONVERSATIONS BETWEEN AUTISTIC AND NON-AUTISTIC DYADS

Simon Wehrle[1], Kai Vogeley[2], Martine Grice[1]

[1]IfL-Phonetik, University of Cologne, Germany [2]Dept. of Psychiatry, University Hospital Cologne, Germany
{simon.wehrle, martine.grice}@uni-koeln.de; kai.vogeley@uk-koeln.de

ABSTRACT

Silent pauses are a pervasive feature of speech, but one that has not received much attention in the context of atypical populations. We investigated the use of silent pauses in conversations between dyads of autistic as compared to non-autistic adults. Previous research on silent pauses in Autism Spectrum Disorder (ASD) is limited to three studies, with contradictory results. We found many similarities in silent pause use between groups, but also a robust difference, as autistic speakers produced more long silent pauses. We also investigated the effect of filled pauses on the duration of subsequent silent pauses. While we found no group-level difference, we were able to replicate, but also qualify, the previous finding that *uhm* precedes longer silent pauses than *uh*. This study hence 1) is a substantial contribution to our understanding of conversational behaviour in ASD and 2) extends important previous findings on silent pauses in the general population.

Keywords: autism spectrum disorder; silent pause; filled pause; conversation; Bayesian modelling

1. INTRODUCTION

Silent pauses feature in the majority of spoken utterances, and they are particularly prevalent in conversational speech. While there is a solid amount of general research on the topic, and in the context of second-language speech in particular [1, 2], very little is known about the use of silent pauses in atypical populations, such as in the speech of persons diagnosed with Autism Spectrum Disorder (ASD).

Previous work on silent pause use in ASD seems to be limited to three studies comparing autistic speakers with matched controls (CTR), with contradictory results. In [3], fewer silent pauses in picture book narrations by English-speaking autistic children are reported. In contrast, [4] report a higher rate of silent pauses in interview-style conversations between experimenters and English-speaking autistic adults. Finally, [5] found equivalent silent pause rates for autistic and matched non-autistic adults in a sentence repetition task.

Crucial differences in the age of participants and/or speech material make comparison with our own data difficult in the cases of [3, 5]. As these confounding factors are less of a concern regarding the work by [4] (age range and speech data being similar to our own), we will focus on this latter study for comparison with our own findings.

The authors of [4] report a higher rate of silent pauses in autistic adults compared to non-autistic controls, but did not examine any silent pauses with a duration of under 2 seconds. The authors provide no specific reasons for using this extremely high cut-off point, only stating that “this was done in order to ensure that we excluded normal prosodic pauses” (p. 138). We can further call into question the sheer utility of such a threshold from a pragmatic–analytic point of view, as employing it entails excluding almost all silent pauses in a given data set; speakers from the control group in [4] in fact did not produce *any* silent pauses longer than 2 seconds.

For this study, we separately analysed 1) silent pauses of any duration, 2) a subset of silent pauses over 2 seconds in duration (for comparison with [4]) and 3) silent pauses of 700 ms or longer (a subset of 1) and a superset of 2); see Section 3.3 for rationale).

In addition to simply comparing the rate and duration of silent pauses, we decided to also probe for other potential group differences in the form of the distributional characteristics of silent pauses. Specifically, we investigated which effect the lexical form of a preceding filled pause (“uh” or “uhm”) had on the duration of the following silent pause.

This is chiefly inspired by the highly influential work in [6] comparing the use of *uh* und *uhm* in spontaneous speech. The authors claim that there is a considerable difference in the average duration of silences following *uh* as compared to *uhm*, with *uhm* preceding silences of at least twice the duration of silences following *uh*. In a comparison of autistic and non-autistic children, [7] confirmed this effect for their CTR, but not their ASD group.

While [6] also showed, in a binary distinction, that silences following “lengthened” productions of both *uh* and *uhm* were considerably longer overall, the duration of the *uh* vs. *uhm* tokens themselves was not controlled for. In fact, to our knowledge, none of the subsequent papers examining this phenomenon involved an analysis that systematically controlled for the inherent average duration of *uh* and *uhm*.

This is a serious concern since, in our dataset at least, *uhm* is considerably longer (521 ms) than *uh* (329 ms) on average . Thus, it is important to establish whether and to what extent the effect ascribed to a difference in filled pause type (nasal vs. non-nasal) is in fact simply due to filled pause duration, independent of whether a final nasal was present (which for simple reasons of physiology and aerodynamics increases the likelihood of longer durations). We thus attempted to replicate the relevant effect while controlling for the confound of filled pause duration, all in the context of investigating differences between the ASD and CTR groups in our data set.

2. METHODS

2.1. Material and subjects

We analysed data from a corpus of approximately 5 hours of semi-spontaneous (Map Task) speech produced by 28 German adults [8–10]. Dialogues had a mean duration of 20 min. overall (SD = 12'; range: 9'–49'), and they tended to be shorter in the ASD (mean = 15'; SD = 7'; range: 9'–30') compared to the CTR group (mean = 26'; SD = 15'; range 13'–49').

Half of the participants (ASD group) had been diagnosed with ASD, the other half (CTR group) were matched for age, gender and verbal IQ. Dialogues took place in disposition-matched dyads, that is, both interlocutors in a dyad either did or did not have a diagnosis of ASD. We collected Autism Spectrum Quotient (AQ) scores for all participants and found that all ASD subjects scored above the suggested threshold of 32 points and all CTR subjects scored below the same threshold [11, 12]. Differences between groups were confirmed using Bayesian modelling (see below).

2.2. Data and methods

All dialogues were manually transcribed using *Praat* [13], and annotations were divided into either silences or inter-pausal units (IPU) of speech with a minimum duration of 200 milliseconds [2, 14, 15]. The corpus contains a total of 3473 silent pauses.

Portions of dialogue that contained only audible breathing, clicks, and similar noises were counted as being part of silent intervals. In contrast, all other speech sounds, including filled pauses, were counted as being part of IPUs. While we acknowledge that most such "silent" pauses are not completely silent from a strictly acoustic perspective [16], we chose to adhere to the conventional definition outlined above, since the main aim of this study is to enable comparison with the (sparse) previous literature on silent pauses in ASD as well as with the more general literature on the topic.

For the analysis of silences following filled pauses, we investigated all 1027 filled pause tokens in the data set, as well as their surrounding linguistic context (see [9] for a full analysis of filled pause use in this corpus). If filled pauses were followed not by any period of silence, but instead directly by another utterance (by either of the interlocutors), we assigned the following silence a duration of 0.

We used Bayesian linear modelling for statistical analysis [17–19]. All code, scripts, model specifications and data frames are available in the *OSF* repository at https://osf.io/bph2t/.

3. RESULTS

3.1. All silent pause tokens

The mean duration of silent pauses was close to identical across groups, with means of 677 ms (SD = 563) for the ASD and 628 ms (SD = 527) for the CTR group. The mean rate of silent pauses was exactly identical across groups, with a value of 12.2 silent pauses per minute. A dyad-specific analysis of silent-pause rates also gives no indications of ASD-specific behaviour; see Fig. 1. Note that there was a considerable degree of by-dyad variability and overlap between groups, not only for this analysis, but also all of the ones described below.

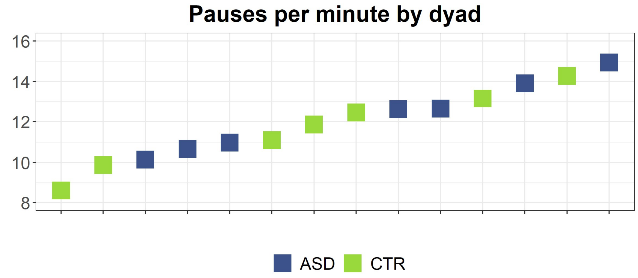

Figure 1: Rate of silent pauses by dyad and group (ASD in blue, CTR in green).

3.2. Silent pauses >2 seconds

To allow for a direct comparison with [4], we separately analysed a subset of all silent pauses with a duration of over 2 seconds. Silent pauses of this kind were very rare in our corpus (73 tokens, or 2%, of the total 3473). The number of such pauses produced by each dyad ranged from 0 to 13.

The ASD group produced a higher mean rate of long silent pauses (>2 s) per minute (0.33; n = 34) than the CTR group (0.21; n = 39); see Fig. 2. Given the low overall number of instances, a more intuitive way of stating the same observation is that a 20-minute dialogue (average duration) would typically contain 7 long pauses (>2 s) in the ASD group and 4 long pauses (>2 s) in the CTR group.

Bayesian Poisson regression suggests that this was a robust difference between groups (mean δ = -0.12; 95% CI [-0.23, -0.01]; P (δ > 0) = 0.97). However, the proximity of the higher end of the credible interval to zero and the very low overall number of observations are reasons for exercising some caution in the interpretation of these data.

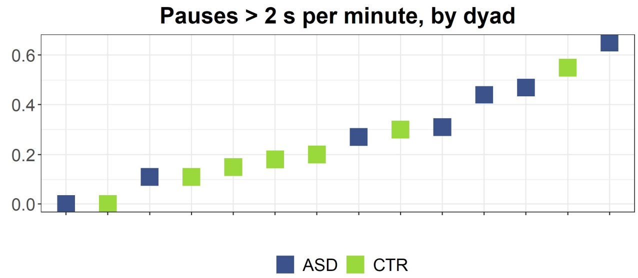

Figure 2: Rate of silent pauses >2 s in duration, by dyad and group (ASD in blue, CTR in green).

3.3. Silent pauses ≥700 milliseconds

For a more reliable and representative metric of “long pauses” in dialogue, we proceeded to use a lower cut-off value, at 700 milliseconds. We chose this particular threshold mainly because it clearly exceeds mean pause durations in the data set used here (646 ms across groups) as well as in previous work [2, 14, 20]. Additionally, the same value has been used for categorising long silent gaps between speakers [10], based on the findings that gaps of 700 ms or longer are perceived as unusual by listeners and often cue repair initiations or non-affiliating responses [21–23]. As the difference between within-speaker pauses and between-speaker gaps structurally lies only in who takes the following turn, these findings further support the use of a 700-millisecond threshold for silent pauses.

Using this cut-off point leaves far more observations for analysis (n = 1052) and will therefore also yield more robust and reliable results.

The group rate of silent pauses ≥700 ms was higher for the ASD group (4.02) than for the CTR group (3.52); see Fig. 3. Although this group difference is not very large, Bayesian negative binomial regression shows the effect to be robust, confirming that the CTR group produced a lower rate of long silent pauses (≥700 ms) than the ASD group (mean δ = -0.5; 95% CI [-0.9, -0.1]; P (δ > 0) = 0.98).

Thus, we can conclude that autistic dyads produced more long silent pauses than non-autistic dyads, independently of the exact cut-off point used to define a “long” pause, although no difference between groups was found when tokens of any duration were taken into account.

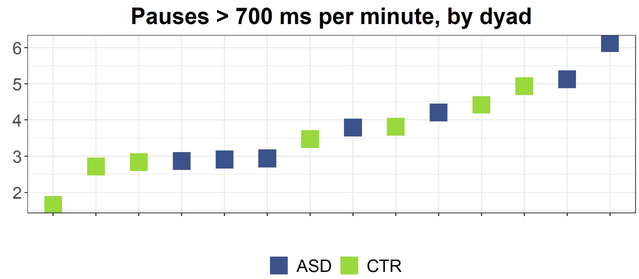

Figure 3: Rate of silent pauses ≥700 ms in duration, by dyad and group (ASD in blue, CTR in green).

3.4. Silent pauses following *uh* vs. *uhm*

The effect of *uh* and *uhm* on subsequent silent pause duration was equivalent for the ASD and the CTR group overall, as *uhm* was followed by longer silences in both groups and for all analyses. We therefore report results across groups below.

When disregarding filled pause duration (as in previous studies), we found a clear difference in the mean duration of following silence according to filled pause type: silences were on average 355 ms longer following *uhm* (mean = 541; SD = 1056) than following *uh* (mean = 186; SD = 517). We further calculated the proportion of filled pauses followed by a period of silence with a duration > 0 (i.e. not followed directly by speech). This was the case more frequently for *uhm* (69.4%) than for *uh* (45%).

As a sanity check, we next ran a Bayesian linear regression model with hurdle log-normal distribution to check whether filled pause duration, independent of filled pause type, could actually be shown to be correlated with the duration of the following silence at all. Model output unambiguously confirms this to be the case (mean δ = 0.29; 95% CI [0.16, 0.43]; P (δ > 0) = 1): longer filled pauses clearly tended to be followed by longer intervals of silence.

To conclusively establish whether differences between filled pause types were independent of the fact that *uhm* tokens in themselves were typically considerably longer than *uh* tokens, we fitted a model with log-normal distribution to the duration of the following silence, with speaker and, crucially, duration of filled pause, as random factors.

We found that silences following *uhm* were indeed longer than those following *uh*, regardless of the duration of filled pause tokens, even though the difference was quite small (150 ms). More details on statistical modelling are reported below.

In our main model, we only included observations where filled pauses were followed by at least 200 ms of silence (i.e. followed by a new, separate IPU). The difference between types is presented with *uhm* as the reference level. Model output shows the difference to be robust, even though the upper bound of the credible interval is close to zero (mean δ = -0.15; 95%

CI [-0.28, -0.02]; P ($\delta > 0$) = 0.97). A second model, including also all cases where the following silence was 0 (using a hurdle log-normal model), confirms the finding in also showing a robust effect for the difference between filled pause types (mean δ = -0.12; 95% CI [-0.19, -0.07]; P ($\delta > 0$) = 1) .

4. DISCUSSION

The analyses presented in this paper show that there were more long silent pauses in conversations between autistic dyads as compared to non-autistic control dyads. This is broadly in line with results from one of the three previously published studies on the same topic [4], but stands in contradiction to an earlier account [3]. We did not find group differences when considering all silent pauses regardless of duration, nor for mean pause duration (similarly to results in [5]). There was also no between-group difference regarding the effect of preceding filled pause type on subsequent silent pause duration. This stands in contrast to results in [7], where longer silences following *uhm* were found for non-autistic, but not autistic children.

While differences were thus rather subtle overall, the higher rate of long silent pauses in the ASD group is still likely to have a discernible effect on spoken interaction [2, 15], and might thus contribute to perceptions of a difference in communication styles. This is all the more true when considering that we found evidence for idiosyncratic behaviour by the same autistic speakers in the related domain of turn-timing, where they produced more long silent gaps *between* speakers compared to non-autistic dyads (but only in the early stages of conversations) [10].

Besides uncovering group differences in silent pause use, we were able to replicate the finding that silent pauses tend to be longer following *uhm* compared to *uh* [6]. Our study does not merely provide a replication, however, but rather extends and qualifies the original finding, as we added duration of filled pause as a random factor in a Bayesian regression model. This allowed us to explicitly show that the effect described is independent of intrinsic filled pause length (importantly, as *uhm* tends to be longer than *uh)*.

Moreover, our results suggest that the effect of longer silent pauses following *uhm* compared to *uh* is more subtle than previously described. While a two-fold difference in silence duration according to filled pause type is reported in [6], we found a difference of only 150 ms (with an average silent pause duration of 646 ms) when factoring in filled pause duration. It is not obvious how relevant such a difference might be in real-life spoken interactions. We will leave this question for future perception experiments.

There are a number of important limitations to the present study. These include the facts that we tested subjects from one extreme end of the autism spectrum (verbal, socially relatively skilled and motivated individuals with average or above-average IQ) and that conversations were task-based and limited to the spoken modality—although it could be argued that both of these limitations add to the specificity and interpretability of our findings. Additionally, the comparability between our own and previous studies is limited, obviously in the case of [3, 5], but also in the case of the more closely related work in [4]. In this study, we analysed semi-structured conversations between disposition-matched dyads of autistic speakers (ASD–ASD), whereas [4] analysed structured interactions between autistic adults and (presumably) non-autistic interviewers (ASD–CTR). While this is representative of most previous research on interactive communication in ASD, we submit that such research cannot yield any reliable insights into "autistic communication" per se. It instead only allows us to observe the particular patterns of behaviour arising in interactions between individuals with different cognitive styles [24–26].

5. CONCLUSION

We have shown evidence for a robust tendency towards a higher rate of long silent pauses in conversations between autistic compared to non-autistic dyads, but also found many similarities in the silent pause use of both groups. For instance, we replicated, extended and qualified the finding that longer pauses follow the filled pause type *uhm* compared to *uh,* across groups, showing further that the effect is subtle and independent of filled pause duration. Thus, besides extending important findings on the distribution of silent pauses in general, the current study specifically advances our understanding of silent pause use in ASD, which is especially important given that previous research has not provided any unambiguous indications in this regard. We hope that these insights might, in conjunction with related findings, ultimately serve as an important component in a multi-dimensional characterisation of conversational behaviour in ASD [8–10, 27].

6. ACKNOWLEDGEMENTS

Thanks to Martina Krüger and Alicia Janz for help with recording, annotation and processing. Funding was provided by the *Deutsche Forschungsgemeinschaft* (Project-ID: 281511265; SFB 1252) and the *Studienstiftung des deutschen Volkes* (Promotionsstipendium).

7. REFERENCES

[1] A. R. Bradlow, M. Kim, and M. Blasingame, 'Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate', *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. 886–899, 2017.

[2] N. H. De Jong and H. R. Bosker, 'Choosing a threshold for silent pauses to measure second language fluency', in *The 6th workshop on disfluency in spontaneous speech (diss)*, 2013, pp. 17–20.

[3] C. Thurber and H. Tager-Flusberg, 'Pauses in the narratives produced by autistic, mentally retarded, and normal children as an index of cognitive demand', *Journal of Autism and Developmental disorders*, vol. 23, no. 2, pp. 309–322, 1993.

[4] J. K. Lake, K. R. Humphreys, and S. Cardy, 'Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders', *Psychonomic bulletin & review*, vol. 18, no. 1, pp. 135–140, 2011.

[5] P. E. Engelhardt, O. Alfridijanta, M. E. G. McMullon, and M. Corley, 'Speaker-Versus Listener-Oriented Disfluency: A Re-examination of Arguments and Assumptions from Autism Spectrum Disorder', *J Autism Dev Disord*, vol. 47, no. 9, pp. 2885–2898, Sep. 2017, doi: 10.1007/s10803-017-3215-0.

[6] H. H. Clark and J. E. Fox Tree, 'Using uh and um in spontaneous speaking', *Cognition*, vol. 84, no. 1, pp. 73–111, May 2002, doi: 10.1016/S0010-0277(02)00017-3.

[7] R. Lunsford, P. A. Heeman, L. Black, and J. van Santen, 'Autism and the use of fillers: differences between 'um'and "uh"', in *DiSS-LPSS Joint Workshop 2010*, 2010.

[8] S. Wehrle, 'A Multi-Dimensional Analysis of Conversation and Intonation in Autism Spectrum Disorder', PhD Thesis, University of Cologne, Cologne, Germany, 2021.

[9] S. Wehrle, M. Grice, and K. Vogeley, 'Filled Pauses Produced by Autistic Adults Differ in Prosodic Realisation, but not Rate or Lexical Type', *Journal of Autism and Developmental Disorders*, accepted.

[10] S. Wehrle, F. Cangemi, A. Janz, K. Vogeley, and M. Grice, 'Turn-timing in conversations between autistic adults: Typical short-gap transitions are preferred, but not achieved instantly', *PLOS ONE*, vol. 18, no. 4, p. e0284029, Apr. 2023, doi: 10.1371/journal.pone.0284029.

[11] K. L. Ashwood *et al.*, 'Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire', *Psychological medicine*, vol. 46, no. 12, pp. 2595–2604, 2016.

[12] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, 'The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians', *Journal of autism and developmental disorders*, vol. 31, no. 1, pp. 5–17, 2001.

[13] P. Boersma and D. Weenink, 'Praat: doing phonetics by computer [Computer program]'. 2020. [Online]. Available: http://www.praat.org/

[14] H. Cho and D. Hirst, 'The contribution of silent pauses to the perception of prosodic boundaries in Korean read speech', *Proceedings of Speech Prosody 2006*, 2006.

[15] F. Goldman-Eisler, *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press, 1968.

[16] M. Belz and J. Trouvain, 'Are 'Silent'Pauses Always Silent?', in *19. International Congress of Phonetic Sciences (ICPhS)*, Melbourne, 2019, pp. 2744–2748.

[17] P.-C. Bürkner, 'brms: An R package for Bayesian multilevel models using Stan', *Journal of statistical software*, vol. 80, no. 1, pp. 1–28, 2017.

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

[19] RStudio Team, 'RStudio: Integrated Development Environment for R'. RStudio, PBC, Boston, MA, 2021. [Online]. Available: http://www.rstudio.com/

[20] B. Megyesi and S. Gustafson-ˇCapková, 'Production and perception of pauses and their linguistic context in read and spontaneous speech in Swedish', in *Seventh International Conference on Spoken Language Processing*, 2002.

[21] K. H. Kendrick, 'The intersection of turn-taking and repair: the timing of other-initiations of repair in conversation', *Frontiers in psychology*, vol. 6, no. 250, pp. 10–3389, 2015.

[22] K. H. Kendrick and F. Torreira, 'The timing and construction of preference: A quantitative study', *Discourse Processes*, vol. 52, no. 4, pp. 255–289, 2015.

[23] F. Roberts and A. L. Francis, 'Identifying a temporal threshold of tolerance for silent gaps after requests', *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. EL471–EL477, 2013.

[24] D. Milton, 'On the ontological status of autism: the "double empathy problem"', *Disability & Society*, vol. 27, no. 6, pp. 883–887, 2012.

[25] G. L. Williams, 'Theory of autistic mind: A renewed relevance theoretic perspective on so-called autistic pragmatic "impairment"', *Journal of Pragmatics*, vol. 180, pp. 121–130, Jul. 2021, doi: 10.1016/j.pragma.2021.04.032.

[26] G. L. Williams, T. Wharton, and C. Jagoe, 'Mutual (Mis)understanding: Reframing Autistic Pragmatic "Impairments" Using Relevance Theory', *Front. Psychol.*, vol. 12, 2021, doi: 10.3389/fpsyg.2021.616664.

[27] S. Wehrle, F. Cangemi, K. Vogeley, and M. Grice, 'New evidence for melodic speech in Autism Spectrum Disorder', *Proc. Speech Prosody 2022*, pp. 37–41, 2022.