INVESTIGATING THE RELATIONSHIP BETWEEN PROSODIC ENTRAINMENT AND INTERACTION STYLE

Joanna Kruyt[1 2*], Lena-Marie Huttner [3*], Johannah O'Mahony [4*]

[1] Institute of Informatics, SAS, Bratislava, [2] Faculty of Informatics and Information Technologies, Slovak Technical University, [3] LPL, CNRS, Aix-Marseille University, [4] CSTR, University of Edinburgh, [*]All authors contributed equally to this paper

joanna.kruyt@savba.sk, lena-marie.huttner@univ-amu.fr, johannah.o'mahony@ed.ac.uk

ABSTRACT

Entrainment is a phenomenon whereby speakers imitate each others' speech patterns during a conversation. It is influenced by social and individual factors. This study aims to investigate whether interaction style predicts entrainment. Following existing studies, we operationalise interaction style using features related to expressive paralinguistics, turn-taking behaviour, and topic persistence. We extracted these features from conversations in the Switchboard I corpus before reducing them into a single *interaction style* variable. We measured the overall prosodic entrainment of the conversation as well as each speaker's contribution using a geometric approach. To investigate whether speakers' involvement scores and entrainment contributions are correlated, we used an actor-partner interdependence model. Results suggest that interaction style and prosodic entrainment are not associated. Possible explanations for these findings, as well as suggestions for future research are discussed.

Keywords: entrainment, convergence, involvement, interaction style, conversation style

1. INTRODUCTION

During interaction, interlocutors' speech patterns sometimes become more similar [1, 2]. This phenomenon is referred to as *entrainment*. Why entrainment occurs is debated: while some argue it is an automatic process resulting from priming [3, 4], others hypothesise that social factors play a central role. Communication Accommodation Theory (CAT) [5] states that speakers (dis-)entrain to minimise or emphasise social differences between themselves and their interlocutor. Entrainment may be more likely to occur if it yields a communicative benefit [6]. Studies have also found that individual differences, such as openness and attention switching abilities [7], can affect entrainment. Additionally, [8] found that individuals who speak with greater variability in their acoustic and articulatory features tend to entrain to others more than individuals who speak with less variability. This suggests that the way someone speaks may relate to entrainment at the prosodic level. In other words, entrainment behaviour seems to be affected both by social and individual factors.

A factor that pertains both to individual differences and to social factors is *interaction style*. Interaction style describes how a person behaves in a conversation in terms of turn-taking behaviour, discussed topics, and paralinguistic expressiveness [9, 10]. Interaction style and entrainment are both associated with rapport [9, 1] and both influence the way people coordinate conversations. We aim to investigate whether an individual's interaction style is associated with their prosodic entrainment behaviour.

Interaction style is an abstract concept that can be operationalised in different ways, but one of the most popular frameworks is Tannen's [9, 10]. Central to her theory is that interaction styles can be considered either *high involvement* or *high considerateness*. High involvement is characterised by a preference for personal topics, quick shifts in topic, topic persistence or repetition of previously brought up topics, telling of stories and rapid turn-taking with short or even negative pauses and overlapping speech. High involvement speakers have a relatively fast speech rate and large pitch and intensity ranges [9, 10]. High considerateness, on the other hand, is associated with little overlapping speech, longer, more frequent pauses both between and within turns, relatively slow speech rate, little variation in pitch, and a preference for less personal topics [9, 10].

Measuring interaction style relies on manual annotations of conversation style, which can be expensive and time-consuming. In this paper we adopt a method by [11], which relies on automatically extracted features to quantify interaction style. While [9, 10] describes high involvement and high considerateness as two opposing styles, in [11]'s operationalisation, which we follow here, interaction style is a spectrum. To our knowledge, no research has directly investigated the association between interaction style and prosodic entrainment. However, [12] suggested that entrainment behaviour is part of linguistic style, and that speakers tend to match each other's linguistic style. [13] found that more prosodic entrainment was observed in conversations where speakers were more engaged or involved with one another. Both [12] and [13] measured their variables at the conversation level, rather than for the individual speakers. In this study, we measured entrainment and involvement per speaker to characterise the relationship between the two.

We aim to investigate whether a difference in interaction style affects entrainment in conversation: according to Tannen, having similar conversation styles facilitates conversation [9, 10], and may thus decrease the perceived social distance between speakers. Following CAT [5], this perceived social distance based on difference in interaction style may influence entrainment.

2. METHOD

2.1. Data

We analysed Switchboard I (SWB) [14] using the Mississippi State University word alignments. We excluded any recordings where a speaker spoke to someone in the background. For our analysis we used 1295 conversations in which both speakers were the same gender (so they could be considered "indistinguishable dyads" in our statistical analyses).

2.2. Measure of involvement

To quantify *interaction style*, we follow the method used in [11]. [11] selected and extracted 11 variables that reflect the characteristics of involvement – topic, pace, and paralinguistic expressiveness – according to Tannen [9, 11]. We extract the same variables as [11], but have operationalised them slightly differently to allow for cross-corpus replication. Unlike [11], who used a line of transcript from a speech-to-text system as their main conversational unit, we used inter-pausal units (IPUs) as our baseline unit. Following [15], the data were split into IPUs using a silence threshold of 180ms and a minimum IPU duration threshold of 100ms [16]. We treated non-speech markers in the SWB dataset as silences. We excluded any word transcriptions with a duration of 0s. We implemented the silence and overlap classification (SOC) of [15] to find pauses (segment of silence within a turn of a single speaker), gaps (silence between turns of two different speakers) and overlaps.

2.2.1. Acoustic Features

[11] extracted features to reflect a speaker's expressive paralinguistics, based on [9]. Pitch variance (pv) was calculated by extracting pitch for each IPU of a speaker using the autocorrelation function in Praat [17] via Parselmouth [18]. We used the two-pass method in [19] to determine the speaker-specific pitch floor and ceiling and removed pitch points which were more than two standard deviations from the speaker's mean. Pitch values were then transformed into semitones using the speaker's median, and the pitch range was calculated. Intensity variance (iv) was calculated by extracting the intensity for each IPU using Praat via Parselmouth [18] and calculating range across all IPUs. Finally, speech rate (wps) was calculated in words per second across the entire recording.

2.2.2. Turn Features

[11] selected a number of features to quantify turn-taking behaviours. We calculated mean length of between-own pauses (boplen) by taking the mean duration of all *pauses* per speaker in the SOC classification. The mean length of post-other pauses (poplen) was calculated per speaker as the mean length of all gaps. Words per utterance (wpu) was calculated as the mean number of words per IPU.

2.2.3. Textual Features

[11] also measured conversation topics and topic persistence. The rate of personal pronoun use (ppron) was calculated by dividing the number of words in a speaker's transcript by the total number of first and second personal pronouns. Following [11], we removed stopwords and filler words and used the Stanford Stanza NLP toolkit [20] to lemmatize remaining tokens. To quantify topic repetition, we calculated the mean number of terms in each IPU that were repeated from the previous IPU (rept). We also calculated the percentage of IPUs which contained a repeated item (repu) by dividing the number of IPUs with a repetition by the total number of IPUs.

2.2.4. Involvement variables

Following [11], interaction style was calculated using a principal component analysis (PCA) on the extracted variables, which were scaled to zero mean and unit standard deviation.The initial scree plot suggested that the variance in the data can be explained by 3 principal components with an eigenvalue >1. The first component explained 28% of the variance, while two less important components explained 16% and 9%, respectively. On a theoretical level, the extracted components account for a single underlying theoretical construct: involvement [9, 10, 11]. In accordance with [11] we interpreted involvement to be represented by the first component, explaining 28% of the variance. A speaker's involvement was quantified as the individual coordinates on the first factor [11]. Table 1 shows the loadings for the individual variables on the involvement component.

Table 1: Loadings of variables on the involvement component

rept	repu	wpp	wpu	olap	wps	iv	pv	ppron	boplen
0.46	0.45	0.38	0.33	0.30	0.20	0.07	0.06	-0.18	-0.29

2.3. Measure of entrainment

To measure prosodic entrainment, we calculated the mean f0 for each speaker (mean f0 per IPU, weighted by IPU duration) for the first and last third of the conversation (based on IPU number). Following [21, 22], we took a geometric approach to quantify entrainment: with speaker A's f0 on the x-axis and speaker B's f0 on the y-axis, a "matching line" is drawn on the x=y line to represent a situation in which both speakers have the same mean f0. Two points were drawn to represent both speakers' mean f0 during the first and last third of the conversation. The minimum distances between these points and the matching line were calculated. The difference between these minimum distances represents the overall entrainment. A speaker's contribution to the overall entrainment was calculated by determining the proportion of change that occurred along each axis. This relative indication of contribution to entrainment was converted to an absolute value by multiplying it with the overall entrainment.

2.4. Statistical analysis

To assess the relationship between involvement and entrainment, we used actor-partner interdependent models (APIM), implemented using the Shiny web-app by [23]. APIMs can be used to quantify actor- and partner-effects, where actor effect refers to how one individual's independent variable predicts their own dependent variable, and partner effect refers to how an individual's independent variable predicts their partner's dependent variable. Both members of the dyad have actor and partner effects. Typically, APIM is used to calculate k, the ratio between the actor and partner effects, which can be used to estimate dyadic patterns. For more information on APIM, see [23]. Here, APIM is used to investigate whether a speaker's involvement score correlates with their own absolute entrainment contribution, their interlocutor's absolute entrainment contribution, and their interlocutor's involvement score. To investigate whether a mismatch in interaction style influenced the observed entrainment, a Pearson correlation was run on the overall entrainment and the absolute difference between both speaker's involvement scores. To investigate whether speakers' combined involvement influenced entrainment, a Pearson correlation between the sum of both speaker's involvement scores and the overall entrainment was conducted.

3. RESULTS

Results of the entrainment analyses show that approximately half of the 1295 dyads (664) showed entrainment and in this group, the mean overall observed entrainment was 6.81Hz and the the average absolute entrainment contribution of both speakers was 3.40Hz. In the other 631 conversations, disentrainment was measured, with the mean overall disentrainment being -6.93Hz and the mean absolute contribution of both speakers in these conversations being -3.46Hz. Table 2 shows a summary of the APIM. The variance of the errors is 31.72 and the R squared is 0.001. The partial intraclass correlation for entrainment contribution while controlling for the other predictor variables is 0.47 and is statistically significant ($p<0.001$, 95% CI [0.41, 0.54]). The predicted entrainment contribution when involvement score equals zero (i.e. the intercept)is equal to 0.06 and is not statistically significant ($p=0.68$, 95% CI [-0.21, 0.33]). The actor effect is -0.07 ($p=0.237$, 95% CI [-0.19, 0.04]). The standardised actor effect is -0.02 (partial r=-0.02). The partner effect is -0.03, which is not statistically significant ($p=0.555$, 95% CI [-0.15,0.08]), and its overall standardised effect is -0.01 (partial r=-0.01). Since neither actor nor partner effects are significant, k cannot be estimated adequately, meaning that calculating and interpreting it will provide no insight regarding the relationship between partner and actor effects.

Table 2: Summary of APIM.

effect	estimate	lower bound CI	upper bound CI	p	beta	r
intercept	0.06	-0.21	0.33	0.68		
actor	-0.07	-0.19	0.04	0.33	-0.02	-0.02
partner	-0.03	-0.15	0.08	0.56	-0.01	-0.01

The Pearson correlation between the speakers' difference in involvement scores and the overall entrainment is not significant (r(1293)=-0.008, p= 0.764), nor is the Pearson correlation between the overall entrainment and the sum of the speakers' involvement scores (r(1293)=-0.032, p=0.256).

4. DISCUSSION

The results of the APIM suggest that there are no significant partner or actor effects, i.e. that an individual's absolute entrainment contribution is not predicted by their own or their conversation partner's involvement score. The two conducted Pearson's correlations suggest no significant correlations between the sum or difference between two speakers' involvement scores and the overall entrainment in their conversation. Taken together, these results suggest that there is no clear link between prosodic entrainment as measured using [21]'s method and interaction style as defined by Tannen [9, 10] and operationalised by [11].

Entrainment was observed in 51.2% of the conversations, whereas disentrainment was found in the other 48.7% of the conversations. The only significant correlation was the intraclass correlation between the two speakers' absolute contribution to the overall entrainment: if one speaker showed a lot of entrainment, the other speaker tended to entrain, too. In conversations where overall entrainment was high, both speakers' contributions were more likely to be high than if the overall entrainment was low.

The proportion of explained variance in the PCA is similar to the analysis reported by [11], though our loadings of variables on the first principal component differ. The differences in the loadings between our analysis and that of [11] could be because [11] analysed involvement in human-machine interaction, while SWB [14] consists of free telephone conversation between people. Different features may play different roles in these different contexts: for example, both pitch and intensity variation both play a large role in [9, 10]'s definition of involvement but may play a smaller role in phone conversations compared to face-to-face interactions, where people can rely on non-verbal cues. In addition, SWB consists of conversations between strangers on the phone. Our findings may not translate to face-to-face interactions or conversations between familiar individuals.

Our results suggest that prosodic entrainment is not associated with interaction style. There are several possible explanations for this. The most straight-forward one is that individual interaction style and prosodic entrainment are simply unrelated. Alternatively, measurement level may have affected the results: while previous studies measured both entrainment and involvement at a conversation level [12, 24], we measured both at an individual level, following [21, 22] and [11]. Possibly, the two phenomena interact differently at different measurement levels.

While we did not find any effect of interaction style on entrainment at the prosodic level, there may be a correlation between interaction style and coordinative behavior at other levels of language such as syntax or lexical choice. It is also possible that "interaction style" is not a stable individual trait, but is context-dependent and can change dynamically. A recent study suggests this: [24] calculated a measure of interaction style based on features extracted from 30 second intervals, and found that individuals exhibited great variation in their interaction styles. Our methods could not capture such dynamic changes.

5. CONCLUSIONS

Our findings suggest that interaction style as as defined by Tannen [9, 10] and operationalised by [11] is not associated with prosodic entrainment as measured with methods by [21, 22]. This could be because the two are simply unrelated, but it could also be that interaction style is not a stable individual trait but changes dynamically throughout a conversation, which our analysis did not account for. Nonetheless, Tannen's [10, 9] definition of involvement as operationalised by [11], is a robust measure of involvement style, which is now replicable over different corpora. Future studies may investigate if involvement relates to entrainment on linguistic levels other than prosody, whether entrainment and involvement dynamically vary together during a conversation, or whether our findings translate to face-to-face interactions or conversations between friends or acquaintances.

6. ACKNOWLEDGEMENTS

This work has been conducted in the framework of the Conversational Brains (COBRA) Marie Sklodowska-Curie Innovative Training Network and has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 859588.

7. REFERENCES

[1] R. Levitan, A. Gravano, L. Willson, Š. Beňuš, J. Hirschberg, and A. Nenkova, "Acoustic-prosodic entrainment and social behavior," in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, 2012, pp. 11–19.

[2] J. S. Pardo, "On phonetic convergence during conversational interaction," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, 2006.

[3] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *Behavioral and brain sciences*, vol. 27, no. 2, pp. 169–190, 2004.

[4] ——, "An integrated theory of language production and comprehension," *Behavioral and brain sciences*, vol. 36, no. 4, pp. 329–347, 2013.

[5] H. Giles, N. Coupland, and I. Coupland, "1. accommodation theory: Communication, context, and," *Contexts of accommodation: Developments in applied sociolinguistics*, vol. 1, 1991.

[6] R. Ostrand and E. Chodroff, "Itâs alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue," *Journal of phonetics*, vol. 88, p. 101074, 2021.

[7] A. C. Yu, C. Abrego-Collier, and M. Sonderegger, "Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and âautisticâ traits," *PloS one*, vol. 8, no. 9, p. e74746, 2013.

[8] Y. Lee, L. Goldstein, B. Parrell, and D. Byrd, "Who converges? variation reveals individual speaker adaptability," *Speech Communication*, vol. 131, pp. 23–34, 2021.

[9] D. Tannen *et al.*, *Conversational style: Analyzing talk among friends*. Oxford University Press, 2005.

[10] D. Tannen, "Toward a theory of conversational style: The machine-gun question." 1980.

[11] P. Thomas, M. Czerwinski, D. McDuff, N. Craswell, and G. Mark, "Style and alignment in information-seeking conversation," in *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, 2018, pp. 42–51.

[12] K. G. Niederhoffer and J. W. Pennebaker, "Linguistic style matching in social interaction," *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 337–360, 2002. [Online]. Available: https://doi.org/10.1177/026192702237953

[13] C. De Looze, C. Oertel, S. Rauzy, and N. Campbell, "Measuring dynamics of mimicry by means of prosodic cues in conversational speech," in *ICPhS 2011*, Hong-Kong, China, 2011. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01705525

[14] J. J. Godfrey, E. C. Holliman, and J. McDaniel, "Switchboard: Telephone speech corpus for research and development," in *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing - Volume 1*, ser. ICASSP'92. USA: IEEE Computer Society, 1992, p. 517â520.

[15] M. Heldner and J. Edlund, "Pauses, gaps and overlaps in conversations," *Journal of Phonetics*, vol. 38, no. 4, pp. 555–568, 2010.

[16] B. Bigi and B. Priego-Valverde, "Search for Inter-Pausal Units: application to Cheese! corpus," in *9th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznań, Poland, May 2019, pp. 289–293. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02428485

[17] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 5.1.13)," 2009. [Online]. Available: http://www.praat.org

[18] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing parselmouth: A python interface to praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 2018.

[19] C. D. Looze and S. Rauzy, "Automatic detection and prediction of topic changes through automatic detection of register variations and pause duration," in *INTERSPEECH*, 2009.

[20] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza: A python natural language processing toolkit for many human languages," 2020.

[21] H. Lehnert-LeHouillier, S. Terrazas, and S. Sandoval, "Prosodic entrainment in conversations of verbal children and teens on the autism spectrum," *Frontiers in Psychology*, vol. 11, p. 2718, 2020.

[22] H. Lehnert-LeHouillier, S. Terrazas, S. Sandoval, and R. Boren, "The relationship between prosodic ability and conversational prosodic entrainment," *Speech prosody (Urbana, Ill.)*, vol. 2020, p. 769, 2020.

[23] L. Stas, D. A. Kenny, A. Mayer, and T. Loeys, "Giving dyadic data analysis away: A user-friendly app for actor–partner interdependence models," *Personal relationships*, vol. 25, no. 1, pp. 103–119, 2018.

[24] N. Ward, "Individual interaction styles: Evidence from a spoken chat corpus," in *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2021, pp. 27–31.