HAL open science

Prosodic and lexical entrainment in adults with and without schizophrenia

Joanna Kruyt, Stefan Benus, Catherine Faget, Christophe Lançon, Maud Champagne-Lavau

To cite this version:

Joanna Kruyt, Stefan Benus, Catherine Faget, Christophe Lançon, Maud Champagne-Lavau. Prosodic and lexical entrainment in adults with and without schizophrenia. Speech Prosody, May 2022, Lisbonne, Italy. hal-03759753

HAL Id: hal-03759753

https://hal.science/hal-03759753v1

Submitted on 24 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prosodic and lexical entrainment in adults with and without schizophrenia

Joanna Kruyt [1 2], *Štefan Beňuš* [1 3], *Catherine Faget* [4], *Christophe Lançon* [4], *Maud Champagne-Lavau* [5]

[1] Institute of Informatics, Slovak Academy of Sciences in Bratislava, Slovakia
[2] Slovak Technical University, FIIT in Bratislava, Slovakia
[3] Constantine the Philosopher University in Nitra, Slovakia
[4] Pôle de Psychiatrie, CHU Conception, APHM in Marseille, France
[5] Aix-Marseille University, CNRS, LPL in Aix-en-Provence, France
joanna.kruyt@savba.sk

Abstract

Entrainment refers to the tendency people have to speak more similarly during a conversation. Although entrainment has been observed frequently, the underlying mechanisms of the phenomenon are debated. A specific point of disagreement is the role of social or higher-order cognitive factors in entrainment. The present study aimed to explore prosodic and lexical entrainment in small groups of individuals with schizophrenia, a disorder that has been associated with theory of mind impairments and social difficulties, and a control group without schizophrenia. All participants completed a referential communication task with an experimenter. To determine prosodic entrainment, the measures proposed by Levitan and Hirshberg [1] were used. Results seem to suggest that the effect of task role on prosodic entrainment was larger than any possible effects of group, suggesting that social factors affect prosodic entrainment behaviour more than individual differences in cognition or other factors. Conversely, lexical entrainment was not affected by task role or group. Importantly, no clear patterns in entrainment on different dimensions, levels, or features could be observed, highlighting the complex and multifaceted nature of entrainment.

Index Terms: entrainment, convergence, alignment, schizophrenia, prosody

1. Introduction

People have a tendency to behave more similarly to each other during an interaction. One of the domains in which this happens is language. Conversational partners tend to adopt similar speaking patterns and use similar words or grammatical structures (e.g. [2, 3, 1]). This frequently observed pattern of linguistic similarity in a conversation is referred to as *entrainment*, though other terms such as *accommodation, convergence, synchrony*, or *alignment* are also commonly used.

There is an ongoing debate about the underlying mechanisms of entrainment, specifically regarding the extent to which entrainment requires higher-order cognitive processing. Pickering and Garrod's influential theory of *interactive alignment* [4, 5] states that entrainment is an automatic process caused by priming. Pickering and Garrod [4, 5] theorise that entrainment "percolates" up to different levels of language. Though this theory does not explicitly hypothesise about entrainment of prosody, it is plausible that entrainment on this level is also caused by priming (e.g. [6]).

Other theories emphasise the role of social or cognitive factors in entrainment. According to the Communication Accommodation Theory (CAT, [7, 8]), entrainment is mediated by social distance: one can alter their behaviour and language to emphasise or minimise the differences between themselves and their interlocutor.

The common ground theory highlights the influence of higher-order cognition (e.g. [9, 10, 11]): interlocutors are aware of the common ground, i.e. information that is shared and known to be shared between interlocutors, and adjust their output accordingly. The common ground account thus states that entrainment relies on perspective-taking and theory of mind (ToM) [11, 9] which is the ability to infer someone else's mental states such as intentions, beliefs, and knowledge. Unlike the interactive alignment account, there is no strong relationship between entrainment at different linguistic levels in this theory. Common ground may affect the prosodic marking of shared or new knowledge, but whether it can influence prosodic entrainment is unclear.

Indeed, some individuals with a ToM impairment, such as those with schizophrenia (SZ) [12, 13], appear to access and mark information in the common ground differently to individuals without such an impairment (healthy controls, HC). Champagne-Lavau et al. [14] recruited SZ and HC participants to complete a referential communication task with an experimenter and found that after the initial trial, participants in the HC group used more definite and less indefinite references to describe previously presented objects than SZ participants. Additionally, the group with schizophrenia required more words and turns to complete a trial. Champagne-Lavau et al. [14] concluded that SZ individuals did not attribute knowledge to their addressee to the same extent as HC participants, in relation to their ToM capacity as assessed by a standardised task.

1.1. The present study

The present study follows a similar procedure to the study by Champagne-Lavau et al. [14]. In the present study, SZ and HC individuals participated in a referential communication task with an experimenter. The collaborative task used was the Tangram task, in which participants have to repeatedly refer to a set of abstract figures without a priori names, in order to place them in a given order. The aim of the current study is to explore and describe entrainment at the prosodic and lexical levels during this collaborative task.

Entrainment has often been associated with positive social measures such as feelings of closeness and interpersonal rapport (e.g. [15, 16, 17]). On the contrary, schizophrenia is associated with social difficulties. This suggests that individuals with schizophrenia may show aberrant entrainment patterns.

A difference in entrainment between groups would suggest that entrainment is not simply the effect of priming, as suggested by [4], but rather is mediated by social factors, such as task role or social distance, or cognitive factors, such as ToM ability. Additionally, an absence of clear association between entrainment behaviours on both levels of language (i.e. prosodic and lexical) would not support the theory that entrainment at different levels "percolates up" [4], but rather would suggest that entrainment on different levels is mediated by different mechanisms. The available sample size in the present study is small and thus the approach to the aforementioned theoretical predictions is primarily descriptive and exploratory.

2. Methodology

2.1. Participants

31 SZ and 30 HC participants matched on age and education level participated in a referential communication task. All SZ individuals were diagnosed by clinicians according to DSM-IV criteria and were outpatients recruited from Hôpital Sainte-Marguerite in Marseille. They were stable and on antipsychotic medication. In the present study, data from 8 SZ and 6 HC participants will be analysed. All participants in this subset were male native French speakers. Written consent forms were obtained from all participants following ethics guidelines set by Aix-Marseille University.

The mean duration of illness in the SZ participants at time of testing was 13.6 years (SD±11.8). The mean Positive and Negative Syndrome Scale [18] scores for this group were 10.5 (SD±3.1) for positive symptoms, 14.9 (SD±4.9) for negative symptoms, and 57 (SD±10.8) overall.

Years of education, age, and Hinting task score [19] were compared between groups using the non-parametric Mann-Whitney U test. Mean years of education was 10.8 (SD±2.7) years in the SZ group and 13.7 (SD±2.8) years in the HC group. The two groups did not differ significantly in age (W=27.5, p=0.70), nor in years of education (W=27.5, p=0.090). The groups differed significantly on their performance on the Hinting task (W=40.5, p=0.036), which is a test to assess ToM ability: participants are told a story and are asked questions about the characters' intentions, beliefs, or desires. Answering correctly thus requires the attribution of a mental state to the characters in the stories. The mean score of the SZ group was 14/20 (SD±3.7) while that of the HC group was 17/20 (SD±1.7). Only the SZ participants in dyads P18, P29, and P31 obtained a Hinting task score that suggests a ToM impairment.

2.2. Procedure

All participants also completed the Tangram task [20, 21] with an experimenter. The "director" was given a set of ordered cards and had to communicate this order to the "addressee". Sixteen dialogues were included in the present study. Half of the interactions involved SZ participants while the other half involved HC participants. Participant task role was counterbalanced across groups. All dyads were mixed gender male-female. Importantly, two participants in the HC group participated in both conditions, i.e. as addressee and as director.

2.3. Analysis

2.3.1. Prosodic entrainment

Mean intensity, median fundamental frequency (f0), f0 range, and maximum f0 (max f0) were extracted from segments of speech that are separated by silences of at least 50ms, or interpausal units (IPUs, following [1]). Overlapping speech (including backchannels) was excluded from analysis. Feature extraction was done using a Praat script [22]. F0 values were extracted using pitch floors and ceilings adjusted for gender and speaker.

Features were normed by gender following a z-score based norming procedure, as proposed by [1]. Prosodic entrainment was measured following the methods in [1]. Specifically, the measures of "local" entrainment, i.e. entrainment that occurs within adjacent IPUs, were used. This is because some of the global measures involve comparing speakers between different dyads. In the present study, several speakers were part of different dyads, which skews such comparisons.

Local proximity is calculated by defining a "partner difference" and "other difference". The difference in feature value between the target IPU and its adjacent IPU (uttered by the other speaker) is the partner difference. The difference between the target IPU and the mean of 10 randomly selected IPUs from the other speaker constitutes the other difference. A paired t-test is conducted to compare the partner and other differences. Local convergence is reflected by the Pearson correlation between partner differences of target IPUs and their IPU number (as an indication of time). To determine local synchrony, feature values from a target IPU and its adjacent IPU are extracted. The Pearson correlation between the features indicates local synchrony.

Following [1], for local synchrony and local convergence the calculations were repeated 10 times with randomly shuffled data (shuffled such that turn-initial IPUs were always matched to turn-final IPUs). Results were only considered valid and significant if $p<0.05$ and no more than 1/10 randomly shuffled correlations was significant. All these analyses were done using R 4.1.0. All code used for these analyses (Praat, R, Python) will be made available on the first author's Github. Because of the exploratory nature of the present study, the small sample sizes, and the non-independence of the data caused by two participants who acted as both director and addressee, additional statistical analyses will not be conducted. Instead, results will be described to see whether any clear patterns can be observed.

2.3.2. Lexical entrainment

Measures of lexical entrainment were obtained using Spacy 3.1.3, specifically the `"fr_core_news_sm"` pipeline, in Python 3.8.0. Utterances of each speaker were lemmatised and stop words were removed. Lexical entrainment was determined on a global dimension (i.e. over the whole interaction) rather than at the turn-level, since some turns were very short (e.g. only one word such as *d'accord* (okay) or *oui* (yes)).

3. Results

Results of the analyses are summarised in two tables, where significant results (i.e. $p<0.05$) are indicated in the "entr." column. A "+" indicates significant entrainment while a "-" indicates significant disentrainment. Marginally significant results (i.e. $p>0.05$ and $p<0.06$) are presented in brackets. Note that for correlations, for a result to be significant and valid, $p<0.05$ and no more than 1 of the random permutations must return a significant result.

Table 1 presents local proximity results. Many dyads showed significant entrainment on at least one feature. Entrainment on one feature does not always mean that the same dyad will show entrainment on another feature, even when two fea-

Table 1: *Results of the paired t-tests for local proximity. A "+" in the "entr." column indicates significant entrainment while a "-" indicates significant disentrainment. Marginally significant results (i.e. p>0.05 and p<0.06) are presented in brackets.*

condition			median f0				range				max				intensity			
dyad	group	role	t	df	p	entr.	t	df	p	entr.	t	df	p	entr.	t	df	p	entr.
P02	SZ	addressee	2.10	75	0.039	+	1.74	75	0.086		1.58	75	0.118		-3.51	75	0.001	+
P10	SZ	addressee	-2.37	97	0.020	+	-1.17	97	0.245		-2.26	97	0.026	+	-3.03	97	0.003	+
P18	SZ	addressee	-0.08	49	0.937		-0.56	49	0.578		-0.64	49	0.527		-0.37	49	0.712	
P21	SZ	addressee	0.61	81	0.547		2.42	81	0.018	+	1.96	81	0.053	(+)	-0.97	81	0.336	
T06	HC	addressee	-0.98	91	0.331		-1.67	91	0.098		-2.07	91	0.041	+	-2.20	91	0.030	+
T09	HC	addressee	-7.09	55	<0.001	+	-1.51	55	0.138		-4.00	55	<0.001	+	-2.73	55	0.009	+
T13	HC	addressee	-5.68	59	<0.001	+	-2.05	59	0.045	+	-3.86	59	<0.001	+	0.69	59	0.496	
T15	HC	addressee	-4.32	49	<0.001	+	-0.89	49	0.376		-2.14	49	0.038	+	-3.41	49	0.001	+
P13	SZ	director	2.45	68	0.017	+	0.44	68	0.663		1.17	68	0.248		-2.66	68	0.010	+
P26	SZ	director	-0.16	35	0.875		4.33	35	<0.001	+	1.11	35	0.273		10.56	35	<0.001	+
P29	SZ	director	3.14	67	0.003	+	9.25	67	<0.001	+	7.86	67	<0.001	+	7.94	67	<0.001	+
P31	SZ	director	3.14	7	0.016	+	1.81	7	0.114		2.70	7	0.031	+	2.30	7	0.055	(+)
T05	HC	director	0.01	64	0.989		-1.97	64	0.053	(+)	-0.56	64	0.574		-1.46	64	0.148	
T08	HC	director	-0.21	27	0.834		0.38	27	0.705		0.16	27	0.875		3.82	27	0.001	+
T13	HC	director	1.87	49	0.068		-2.35	49	0.023	+	-0.56	49	0.577		-0.16	49	0.871	
T15	HC	director	-5.55	43	<0.001	+	-3.22	43	0.002	+	-3.35	43	0.002	+	-3.64	43	0.001	+

tures are closely related (e.g. f0 range and max f0). Furthermore, there is no systematic effect of role or group. Nonetheless, there are some patterns. Firstly, more entrainment on f0 range was observed in director condition (4(+1)/8) than in the addressee condition (2/8). Secondly, slightly more max f0 entrainment can be seen in the addressee condition (5(+1)/8) than in the director condition (3/8). Thirdly, more max f0 entrainment can be seen in the HC group (5/8) than the SZ group (3(+1)/8). The participants that participated twice (T13 and T15) did not show the same entrainment behaviour in the different conditions. Finally, there are no clear trends in entrainment behaviour in the dyads with an individual with a ToM impairment (i.e. P18, P29, and P31): for example, while dyad P18 showed no significant entrainment, P29 showed entrainment on every feature. The three dyads also do not appear to exhibit different patterns than the other dyads, suggesting that ToM ability may not affect prosodic entrainment on the dimension of local proximity.

Table 2 shows that on the dimension of local synchrony, all but two of the significant results indicate disentrainment. However, no clear pattern can be seen between disentrainment on the different features: some dyads disentrain on only one feature, while others disentrain on several. Additionally, a dyad may show disentrainment on one feature, but entrainment on another (e.g. P13). Interestingly, the only two instances of entrainment occured in the director condition. In the addressee condition, only disentrainment was found. Conversely, more disentrainment on mean f0 was found in the director condition (5/8) than in the addressee condition (2/8). There seem to be no clear differences in entrainment between groups, though more disentrainment on max f0 can be seen in the SZ (3/8) than HC condition (1/8). The dyads T13 and T15 appear to follow the same pattern of entrainment in both conditions: no entrainment is observed in interactions with T13, while disentrainment on f0 range is found in both conditions for dyad T15. Finally, results from the dyads P18, P29, and P31, show no systematic patterns or differences.

Results of the lexical entrainment analysis are presented in the final column of Table 2. Note that the mean of the SZ group (M=0.74, SD±0.15) and HC group (M=0.71, SD±0.09) are practically identical. Similarly, the mean of the interactions in the addressee (M=0.77, SD±0.06) and director (M=0.68, SD±0.14) conditions are very similar. Note that dyads P18, P29, and P31 do not seem to lexically entrain differently to the other dyads. Within the three dyads, no pattern can be seen either: for example, dyad P29 exhibited the highest degree of lexical entrainment of all dyads, while dyad P31 exhibited the lowest cosine similarity.

4. Discussion

The aim of this paper was to conduct an exploratory investigation of prosodic and lexical entrainment in SZ and HC groups. Minimal differences in prosodic entrainment between HC and SZ individuals were detected, but results suggest that the influence that social factors such as task role have on prosodic entrainment may be more pronounced than any possible effects of group. An example of this can be seen in Table 2: three dyads showed disentrainment in the addressee condition (2 SZ, 1 HC) and two dyads showed entrainment in the director condition (1 SZ, 1 HC).

It would not be surprising if task role affects prosodic entrainment: previous research has shown that social factors such as and social dominance (e.g. [23]) and gender, task role, and the interaction between the two (e.g. [24]) can influence prosodic entrainment. The finding may be further explained by considering that the analysed conversations involve an experimenter who participated in multiple interactions, and that the addressee tends to speak less than the director. The entrainment behaviour of the experimenter may thus influence the results more than that of the participants in the addressee conditions. Furthermore, task role could be linked to feelings of dominance, since the director possesses more information than the addressee. This may increase feelings of social distance and could thus, following CAT, influence entrainment.

The dyads including the participants that acted as both director and addressee did not always show the same entrainment behaviours: for example, on the dimension of local proximity, T13 showed entrainment on all but one feature in the addressee condition, and on only one feature in the director condition. This suggests that the effect of role on prosodic entrainment may have a larger effect than individual differences. On the contrary, on the dimension of local synchrony, dyad T15 showed the same pattern of disentrainment in both conditions. These findings could be interpreted as suggesting that entrainment is

Table 2: *Results of the analyses for local synchrony on prosodic features and for global lexical entrainment. Pearson's correlation for local synchrony measures are presented per feature. A "+" in the "entr." column indicates significant entrainment (p<0.05) while a "-" indicates significant disentrainment. Marginally significant results (i.e. p>0.05 and p<0.06) are presented in brackets. Cosine similarity presented in the final column.*

condition			median f0			range			max			intensity			lexical entrainment
dyad	SZ	role	r	p	entr.	r	p	entr.	r	p	entr.	r	p	entr.	cosine similarity
P02	SZ	addressee	-0.58	<0.001		-0.35	0.002		-0.47	<0.001	-	-0.48	<0.001		0.84
P10	SZ	addressee	-0.32	0.001	-	0.00	0.999		0.06	0.568		-0.16	0.123		0.82
P18	SZ	addressee	-0.82	<0.001		0.05	0.738		-0.21	0.151		-0.55	<0.001	-	0.77
P21	SZ	addressee	-0.55	<0.001		-0.09	0.443		-0.35	0.001	-	-0.37	0.001	-	0.81
T06	HC	addressee	-0.17	0.099		0.06	0.556		0.00	0.983		-0.23	0.026	-	0.82
T09	HC	addressee	0.05	0.717		0.00	0.977		-0.03	0.853		0.03	0.839		0.76
T13	HC	addressee	0.03	0.807		0.08	0.559		0.07	0.588		0.23	0.075		0.67
T15	HC	addressee	-0.32	0.024	-	0.01	0.932		-0.01	0.933		-0.15	0.299		0.68
P13	SZ	director	-0.36	0.003		-0.39	0.001	-	-0.57	<0.001	-	0.48	<0.001	+	0.68
P26	SZ	director	-0.68	<0.001	-	0.24	0.153		-0.66	<0.001		0.31	0.065		0.57
P29	SZ	director	-0.35	0.004	-	0.03	0.836		-0.07	0.584		0.07	0.546		0.94
P31	SZ	director	-0.19	0.657		-0.43	0.285		-0.39	0.343		-0.68	0.066		0.50
T05	HC	director	-0.33	0.007	-	-0.21	0.087		-0.25	0.041	-	0.37	0.003	+	0.73
T08	HC	director	-0.67	<0.001	-	-0.23	0.234		-0.74	<0.001		-0.20	0.304		0.58
T13	HC	director	-0.03	0.839		-0.26	0.070		-0.08	0.574		0.24	0.093		0.61
T15	HC	director	-0.57	<0.001	-	-0.07	0.636		-0.27	0.072		-0.64	<0.001		0.81

mediated by both social factors and individual differences, and that the balance between the two may differ from person to person.

While the comparison of these two dyads in different conditions is interesting, the fact that these individuals participated in both condition is a limitation of the study. The non-independence of the data that results from this means that no statistical comparisons between roles can be done. Additionally, this limitation makes the findings less generalisable and may have skewed the results.

Contrary to the results regarding prosodic entrainment, no clear effect of role, group, or ToM ability can be observed on lexical entrainment. It is possible that no significant differences were found because of the task: participants were to discuss specific objects; conversation topics were thus highly constrained. An alternative explanation is the small sample size included in the present study. Alternatively, it is possible that lexical entrainment simply is not affected by social or cognitive factors during this task. However, there is substantial variation in the cosine similarities between dyads (range from 0.50-0.94). It is thus possible that other social factors or individual differences affect lexical entrainment.

Importantly, there is no apparent relationship between prosodic and lexical entrainment: they are not similarly affected by social or cognitive factors, and dyads who exhibited entrainment on prosodic features do not necessarily show more lexical entrainment. See for example dyad P29 in Table 2, who showed the most lexical entrainment, but disentrainment on median f0 on the dimension of local synchrony. This suggests that entrainment at the prosodic and lexical level may not be mediated by the same underlying mechanism. This is not in line with the interactive alignment account [4]. In general, the exploratory results of the present study seem to be more in line with CAT [8] than the other theories.

Moreover, results of the present study highlight the lack of structure or patterns in entrainment behaviour: no pattern can be observed in entrainment on different features, dimensions, or levels. For example, all signficant results on the dimension of local proximity reflected entrainment, while all but two of the significant results on the dimension of local synchrony indicated disentrainment. Furthermore, prosodic and lexical entrainment seemed unrelated. This lack of structure or pattern is corroborated by recent research [25, 26] that investigated whether entrainment on different levels, dimensions, and features is correlated, but failed to find an underlying structure. The authors write that "it appears that entrainment, rather than a single behavior or a structured collection of behaviors, is a set of behaviors which are only loosely linked and perhaps independently explained by the competing theories." (p.301, [25]). Entrainment on different levels, dimensions, and features may serve different functions and may be governed by different mechanisms. Different forms of entrainment may be affected by social, cognitive, physiological, attentional, or contextual factors, and some of these factors may weigh more heavily than others. Additionally, this weighting may differ from person to person.

In other words, the present study provides an initial exploration of prosodic and lexical entrainment in SZ and HC individuals. Results seem to suggest minimal effects of group or ToM ability, but suggest that task role may affect entrainment. While data from all participants recruited should and will be analysed to confirm and extend these initial findings, it is important to keep in mind the complex, multifaceted nature of entrainment. Although the results of the present study come from a small sample size and are primarily descriptive, the findings are nonetheless informative and valuable, especially considering the challenges of collecting such data from clinical populations.

5. Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 859588. Collection of the data analysed in this study was supported by an Agence Nationale de la Recherche grant (ANR-11-BSH2-006-01 MINDPROGEST) to MCL.

6. References

[1] R. Levitan and J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[2] H. P. Branigan, M. J. Pickering, and A. A. Cleland, "Syntactic coordination in dialogue," *Cognition*, vol. 75, no. 2, pp. B13–B25, 2000.

[3] 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.

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

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

[6] A. Ward and D. J. Litman, "Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora," 2007.

[7] H. Giles, "Accent mobility: A model and some data," *Anthropological linguistics*, pp. 87–105, 1973.

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

[9] H. H. Clark and C. Marshall, "Reference diaries," in *Theoretical issues in natural language processing-2*, 1978.

[10] H. H. Clark, *Using language*. Cambridge [England] ; New York: Cambridge University Press, 1996.

[11] H. H. Clark and G. L. Murphy, "Audience design in meaning and reference," in *Advances in psychology*. Elsevier, 1982, vol. 9, pp. 287–299.

[12] A. Michelas, C. Faget, C. Portes, A.-S. Lienhart, L. Boyer, C. Lançon, and M. Champagne-Lavau, "Do patients with schizophrenia use prosody to encode contrastive discourse status?" *Frontiers in Psychology*, vol. 5, p. 755, 2014.

[13] D. Dickinson, A. S. Bellack, and J. M. Gold, "Social/communication skills, cognition, and vocational functioning in schizophrenia," *Schizophrenia bulletin*, vol. 33, no. 5, pp. 1213–1220, 2007.

[14] M. Champagne-Lavau, M. Fossard, G. Martel, C. Chapdelaine, G. Blouin, J.-P. Rodriguez, and E. Stip, "Do patients with schizophrenia attribute mental states in a referential communication task?" *Cognitive Neuropsychiatry*, vol. 14, no. 3, pp. 217–239, 2009.

[15] J. Michalsky, H. Schoormann, and O. Niebuhr, "Conversational quality is affected by and reflected in prosodic entrainment," *Proceedings of Speech Prosody 9, Poznan, Poland, Proceedings*, 2018.

[16] J. S. Pardo, R. Gibbons, A. Suppes, and R. M. Krauss, "Phonetic convergence in college roommates," *Journal of Phonetics*, vol. 40, no. 1, pp. 190–197, 2012.

[17] T. L. Chartrand and J. A. Bargh, "The chameleon effect: the perception–behavior link and social interaction." *Journal of personality and social psychology*, vol. 76, no. 6, p. 893, 1999.

[18] S. R. Kay, A. Fiszbein, and L. A. Opler, "The positive and negative syndrome scale (panss) for schizophrenia," *Schizophrenia bulletin*, vol. 13, no. 2, pp. 261–276, 1987.

[19] R. Corcoran, G. Mercer, and C. D. Frith, "Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia," *Schizophrenia research*, vol. 17, no. 1, pp. 5–13, 1995.

[20] H. H. Clark and D. Wilkes-Gibbs, "Referring as a collaborative process," *Cognition*, vol. 22, no. 1, pp. 1–39, 1986.

[21] R. M. Krauss and S. Weinheimer, "Concurrent feedback, confirmation, and the encoding of referents in verbal communication." *Journal of personality and social psychology*, vol. 4, no. 3, p. 343, 1966.

[22] P. Boersma, "Praat: doing phonetics by computer," *http://www.praat. org/*, 2006.

[23] Š. Beňuš, A. Gravano, R. Levitan, S. I. Levitan, L. Willson, and J. Hirschberg, "Entrainment, dominance and alliance in supreme court hearings," *Knowledge-Based Systems*, vol. 71, pp. 3–14, 2014.

[24] U. D. Reichel, Š. Beňuš, and K. Mády, "Entrainment profiles: Comparison by gender, role, and feature set," *Speech Communication*, vol. 100, pp. 46–57, 2018.

[25] A. Weise and R. Levitan, "Looking for structure in lexical and acoustic-prosodic entrainment behaviors," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 2018, pp. 297–302.

[26] R. Ostrand and E. Chodroff, "It's 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.