

Meno a priezvisko/Name and Surname

Joanna Alexandra Elizabeth Kruyt

študenta doktorandského študijného programu/
of the student of the doctorate degree study programme)

Autoreferát dizertačnej práce/Dissertation Thesis Abstract

Conversation coordination and Theory of Mind

na získanie akademického titulu/to obtain the Academic Title of
„doktor“ („philosophiae doctor“, v skratke/abbreviated as „PhD.“

v doktorandskom študijnom programe/in the doctorate degree study programme:
9.2.9. Applied Informatics

v študijnom odbore/in the field of study:
18. Computer Science

Forma štúdia/Form of Study:
part-time study

Miesto a dátum/Place and Date: Bratislava, 04-09-2023



Dizertačná práca bola vypracovaná na/Dissertation Thesis has been prepared at
The Institute of Informatics, Slovak Academy of Sciences in Bratislava, Slovakia

Predkladateľ/Submitter:

Joanna Alexandra Elizabeth Kruyt
Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9
845 07 Bratislava 45
Slovak Republic

Školiteľ/Supervisor:

prof. Mgr. Štefan Beňuš, PhD.
Institute of Informatics, Slovak Academy of Sciences

Konzultant/Consultants:

NA

Autoreferát bol rozoslaný/Dissertation Thesis Abstract was sent:

30-10-2023

Obhajoba dizertačnej práce sa bude konať dňa/Dissertation Thesis Defence will be held on
13.11.2023 o/at 14:30 h / 2:30 p.m. (am/pm)

na/at Ústave informatiky Slovenskej akadémie vied, Dúbravská cesta 9, 845 07 Bratislava 45

.....
rektor STU alebo dekan fakulty,
ak sa doktorandský študijný program uskutočňuje na fakulte
prof. Ing. Ivan Kotuliak, PhD.
Rector of STU or Dean of Faculty of STU,
if studies of the doctorate degree study programme were carried out at the Faculty

Table of Contents

Annotations 4

 Annotation (English) 4

 Annotation (Slovak) 4

Summary of thesis 5

 Introduction 5

 Goals of thesis 5

 Contributions and impact 11

 Theory of Mind and entrainment 11

 Methodologies in entrainment research 12

 Conclusions 13

 References 14

Research output 16

 Publications based on work in the thesis 16

 Other publications 16

 Invited talks 16

Annotations

Annotation (English)

Interlocutors' tendency to behave similarly in conversation is called *entrainment* and is linked to positive social outcomes. Interest in implementing entrainment into technology has grown; however, the underlying psychological mechanisms of the phenomenon, specifically the role of "theory of mind" (ToM), are debated.

The first aim of this thesis is to characterise the role of ToM in linguistic entrainment, with the goal of advancing the state-of-the-art and facilitating future implementation of entrainment in technological applications. Lexical and prosodic entrainment was measured in people with ToM impairments (autism, schizophrenia) and control groups. Results showed that ToM had no systematic effect on entrainment; ToM is thus likely not crucial for entrainment implementations. However, inconsistent patterns of entrainment were found, which may be attributed to methodological differences that are also present in previous studies.

The second aim of this thesis is to evaluate the replicability of entrainment research. One dataset was analysed using 12 different methods; their results did not always correspond. A follow-up study showed that differences in prosodic norming accounted for some, but not all of the variance. The practical and theoretical implications for entrainment research and implementation were discussed.

Annotation (Slovak)

Pri vzájomnom *prispôsobovaní* (angl. *entrainment*) má podobnosť rečníkov tendenciu sa zvyšovať, čo pozitívne vplyva na rôzne aspekty rozhovoru. Záujem o implementáciu prispôsobovania do interakcií človek-stroj vzrastá, ale na mechanizmoch tohto fenoménu, konkrétne na úlohe "teórie mysle" (TM), sa konsenzus ešte nedosiahol.

Prvým cieľom práce je charakterizovať úlohu TM v lingvistickom prispôsobovaní a uľahčiť tak budúcu implementáciu prispôsobovania v automatických dialógových systémoch. Meralo sa lexikálne a prozodické prispôsobovanie u ľudí s poruchami TM (autizmus, schizofrénia) a u kontrolných skupín. Výsledky ukázali, že TM nemala systematický vplyv na prispôsobovanie a teda nie je asi pre implementáciu prispôsobovania rozhodujúca. Zistili sa však nekonzistentnosti v pozorovanom prispôsobovaní, čo možno pripísať metodologickým rozdielom, ktoré sú prítomné aj v predchádzajúcich štúdiách.

Druhým cieľom je zhodnotiť replikovateľnosť výskumu prispôsobovania: rovnaký rečový korpus bol analyzovaný pomocou 12 rôznych metód, ktorých výsledky sa nie vždy zhodovali. Následná štúdia ukázala, že rozdiely v normovaní akusticko-prozodických charakteristík reči rečníkov vysvetlili niektoré, ale nie všetky rozdiely medzi metódami. Tieto zistenia majú podstatné praktické a teoretické dôsledky pre budúci výskum a implementáciu rečového prispôsobovania.

Summary of thesis

Introduction

The coordination of conversation is more complex and subtle than it might appear. Humans employ many different strategies for coordinating a conversation, including providing feedback to interlocutors (e.g. [1, 2, 3]), or adapting to interlocutors to become more similar in behaviour (e.g. [4, 5, 6, 7]).

An increasing proportion of our day-to-day communication is not between two humans, but between a human and a machine. Such interactions are likely to become more frequent, both in text-based and spoken modalities. Most dialogue systems and robots are modelled after human behaviour. A better understanding and ability to model human behaviour during conversations can improve our conversational agents and dialogue systems, and will likely lead to more effective, smooth, and satisfying communication with these machines.

This thesis focused on a specific communication-related phenomenon that is referred to as *entrainment*: the tendency of people to behave more similarly during an interaction. Entrainment has been studied by both psychologists and computer scientists, and was found to be associated with a range of positive social measures, such as for example feelings of closeness, and effective, smooth, and natural conversation (e.g. [8, 9, 10, 7, 11, 12, 13, 14, 15]). Because of its associations with social affiliation, the interest in entrainment and its potential implementations into dialogue systems and other conversational agents has grown immensely.

Despite its apparent strong ties to social processes and promising potential for the improvement of human-machine interaction, some important questions regarding entrainment remain unanswered. Specifically, the underlying cognitive mechanisms of entrainment are debated. Insights into these mechanisms have implications in a number of areas: a better understanding of how entrainment occurs in human-human interaction can for example inform the development of smoother human-machine interaction. After all, current implementations of entrainment in spoken dialogue systems are scarce and often not very sophisticated, which highlights the need for better understanding and modelling of entrainment behaviour and its underlying mechanisms.

The most prominent debate concerning the underlying mechanisms of entrainment is centred around the role of higher-order cognitive processes, and specifically on the extent to which the ability to infer our interlocutor's mental states, also known as "mind-reading" or *Theory of Mind* (ToM), is required for entrainment. The aim of the present project was to take an interdisciplinary approach to characterise the role of ToM in entrainment, with the goal of advancing our understanding of this complex phenomenon and its potential for implementation in various technological settings. Understanding the role of ToM in entrainment will for example elucidate whether spoken dialogue systems require some form of ToM modelling in order to "properly" entrain, and can inform the temporal dimensions of entrainment.

Importantly, the topic is multi- and inter-disciplinary in nature, which not only means that it can have far-ranging implications in a number of fields, but also that in order to properly understand and investigate the link between entrainment and ToM, information and methodologies from various disciplines must be employed and synthesised. This interdisciplinary approach was central to this work and was maintained throughout the thesis.

Goals of thesis

This thesis project was concerned with (para)linguistic entrainment, specifically at the prosodic and lexical levels. This type of entrainment has been associated with a number of positive social measures, such as

feelings of closeness and smooth, effective communication. A thorough understanding of the role of entrainment in social interaction is important for applications using automatic speech processing tools in both clinical and non-clinical environments.

This thesis aimed to provide a theoretical and practical scaffolding upon which future, more applied research and work can build. In order to implement entrainment in human-machine interactions, an in-depth understanding of its role and of its underlying mechanisms in human-human interactions is essential, as this can elucidate its functioning and functions in social communication. Current implementations of entrainment in spoken dialogue systems have been rather rudimentary (e.g. [16, 17]) and a substantial amount of theoretically-oriented, multidisciplinary research is needed before more sophisticated models can be implemented.

Some unanswered questions and open issues arise from the existing literature from different disciplines, and this thesis aimed to address several of these in the following ways:

Issue 1: The role of theory of mind in entrainment is debated

The role of theory of mind (ToM) in entrainment is not sufficiently understood for future implementations. Different theories disagree on the extent to which ToM is required for entrainment. As mentioned before, insight into the underlying mechanisms of entrainment must inform the design and development of automatic dialogue systems and the deployment of such systems for future human-machine spoken interactions. For example, if entrainment is caused by priming, dialogue systems should simply repeat phrases used by humans, and should do so relatively shortly after the human's mentions. If, on the other hand, entrainment relies on higher-order cognitive processes like ToM, it may be better for referring expressions to be generated with the consideration of contextual information and perhaps even computationally modelling ToM, and entrainment should develop more slowly and gradually than in the priming scenario. In such ways, a deeper theoretical understanding of the role that ToM plays in entrainment is critical for the development of spoken dialogue systems and human-machine interaction.

In concrete terms, this thesis aimed to address these unanswered questions and to further our understanding of the role of ToM in entrainment. The link between ToM and entrainment was primarily explored in two experiments that compare entrainment in neurotypical populations to that in groups of people who have ToM impairments, namely children diagnosed with autism and adults diagnosed with schizophrenia. If we find significant differences in entrainment between these clinical and neurotypical populations, this would support the necessity of including ToM in efforts for implementing entrainment in human-machine spoken dialogue systems. If, on the other hand, differences are minimal or non-existent, the need for modelling and implementing ToM would be reduced. Besides providing information on the relationship between ToM and entrainment, understanding whether such clinical populations entrain similarly to healthy populations may reveal whether entrainment can be used as a proxy to measure social difficulties, or perhaps as a therapeutic target for intervention strategies, potentially even implemented using spoken dialogue systems or social robots. In order to uncover the therapeutic potential of entrainment in clinical populations, we need to know whether clinical populations entrain differently to healthy populations, and whether this is associated with certain clinical traits or symptoms.

Goal and contribution 1a: measure entrainment at the prosodic and lexical levels in adults with and without schizophrenia

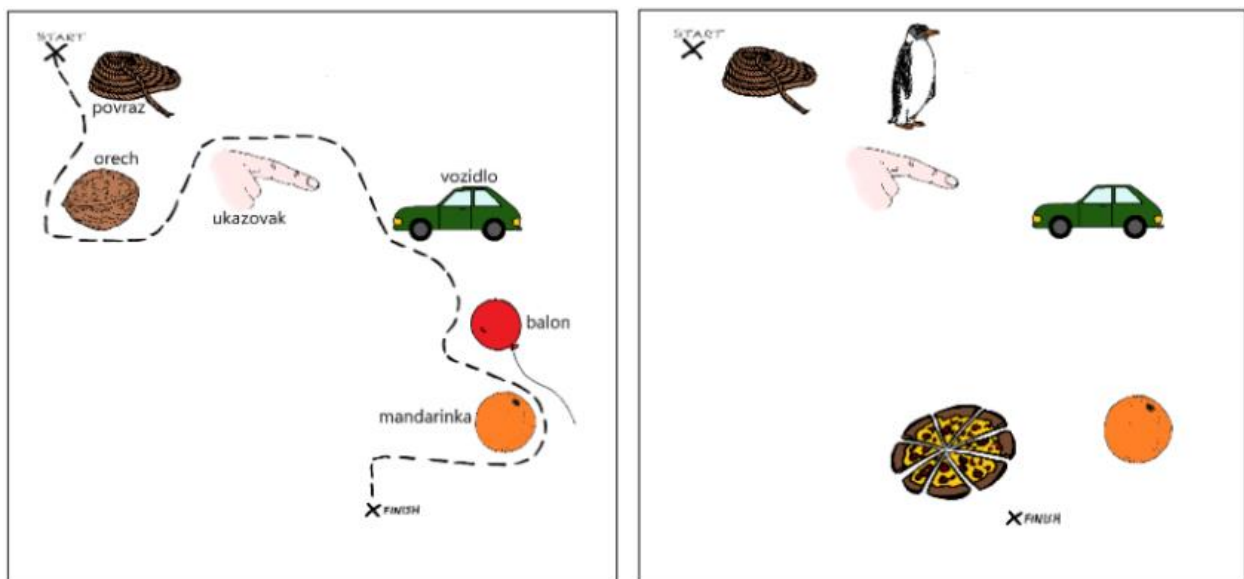
In this first exploratory study, an existing dataset of speech recordings was analysed. The dataset consisted of recordings of interactions between participants (with or without schizophrenia) and an

experimenter, who together played a cooperative game in which they had to collaborate to arrange a set of objects in a particular order. Prosodic entrainment in these conversations was measured using three different methods proposed by [18], and lexical entrainment was measured using cosine similarity calculations. Results suggested no systematic differences between groups on either level of entrainment, but suggest that social factors, such as the role a participant played in the task, may have a minimal effect on entrainment at the prosodic level.

Goal and contribution 1b: quantify entrainment at the prosodic and lexical levels in children with and without autism spectrum disorder

In the second study, children with and without autism were recruited and asked to participate in a novel task designed especially for this experiment. Their conversations during this task were recorded, transcribed, and manually time-aligned in order to create a large, shareable database that can be used for future research. The developed task was a collaborative navigation task, during which participants had to cooperate and describe routes on maps to one another. Importantly, these maps contained “landmarks”, some of which could be referred to using multiple names (e.g. mandarin vs orange). The experimenter in the task always used the “dispreferred” word for these objects, and I operationalised lexical entrainment by measuring how often the child participant also used to dispreferred word rather than the preferred word. I measured entrainment at the prosodic level using three different methods proposed by [18] and a geometric approach developed by [19]. Results from these analyses suggest that neither group nor ToM score significantly predicts entrainment on either level. Social factors, such as gender, and other cognitive factors, such as executive functioning ability, may significantly predict lexical entrainment. No clear patterns were found in the various measures of prosodic entrainment.

Figure 1. Example of maps designed for the navigation task. The left map was used by the adult experimenter, and the landmarks on her map were labelled with the word she was supposed to use to describe the landmarks.



Overall, the results of these first two experiments suggest that ToM is not required for entrainment at either the lexical or prosodic level, which means that this is likely not a fruitful direction for clinical

applications. Additionally, results imply that ToM modelling is not required for the implementation of entrainment into spoken dialogue systems or other forms of human-machine interaction.

Taken together, no clear patterns emerged from the prosodic entrainment results of these two studies. The results cannot be explained by merely one of the major theories of entrainment, which is in line with existing research that often reported unclear or contradictory results. A possible explanation for this lack of clear patterns in these two experiments as well as in existing research, is the fact that different methodologies were used in almost every study.

Issue 2: scarcity of data of clinical populations, especially in underrepresented languages such as Slovak

Importantly, one of the reasons that the link between ToM and entrainment is debated, is that datasets that can be used to study these phenomena are scarce. Many of the clinical groups with ToM impairments are small, meaning it is difficult to recruit enough participants during data collection. Additionally, speech recordings from clinical groups such as people with ASD or schizophrenia are subject to more stringent ethical restrictions, which makes it difficult to share these datasets between researchers. While there are a few available corpora, most of these are in English or other Germanic languages, and practically no shareable data exists in Slavic languages. This means that this group of languages is underrepresented in this field of research. Finally, most available corpora consist of either highly structured, predictable, task-oriented conversations, or of entirely unstructured, free conversations. The corpus collected as part of this thesis consists of semi-structured, semi-naturalistic task-oriented interaction, and is in this way thus also a valuable contribution to the state-of-the-art.

Goal and contribution 2: create a shareable database of interactions between Slovak children with and without ASD and an experimenter

I collected a database consisting of speech recorded from children with and without autism spectrum disorder (ASD), which is a disorder that is associated with ToM impairments. The database consists of semi-naturalistic, task-oriented speech that was recorded during an interactive, collaborative task performed with a clinician. In total, the database includes interactions with 76 children: 41 with ASD and 35 neurotypical children. The total corpus contains over 17 hours of speech. The audio recordings as well as time-aligned transcripts can be shared with other researchers who can investigate conversation coordination in ASD, or who want to research the application of automatic speech processing tools for potential diagnostic, monitoring, or other therapeutic purposes.

Importantly, this database can not only be used to study entrainment, it can also be used to investigate other conversation coordination mechanisms such as turn-taking, feedback, or misunderstanding and conversational repair. Any findings relating to such mechanisms in this corpus can provide insight into the psychological underpinnings of these mechanisms, which in turn are useful for the development of dialogue systems. Additionally, the fact that our collected database contains both children with ASD as well as children without ASD, allows for the comparison of conversation coordination strategies between groups. This makes the corpus suitable for the testing of automatic screening tools for ASD or other communication difficulties. Furthermore, the corpus consists of child speech and could thus be used to test or validate automatic speech recognition for (atypical) child speech, which can improve such speech processing technologies.

Several psychological test scores, such as IQ scores and ToM scores, are also included in the database. Naturalistic speech data of children with ASD is scarce, especially in languages other than

English, so this will also be a valuable contribution to the fields of (developmental) linguistics and psychology more generally.

Issue 3: plurality of methods to measure entrainment, lack of consensus on methodology

What is essential to keep in mind when discussing the various applications and implications of entrainment research, is how entrainment is measured. If one wants to incorporate entrainment in a spoken dialogue system, the actual entrainment in conversations with this implementation needs to be validated with a reliable method in order to assess the success of the implementation. Similarly, if entrainment is going to be used in any clinical contexts, either for screening or as a therapeutic intervention, it is essential that measurements of the degree of entrainment are reliable. When surveying the literature, it becomes clear that existing entrainment research is sometimes difficult to interpret because of the large number of different methodologies that are employed. This opened up a second line of investigation in this thesis: the evaluation of methodologies used in entrainment research.

Goal and contribution 3a: compare and validate different methods that are used to measure entrainment

Validating measures of entrainment is crucial for the comparison of existing studies: if different methods measure different things, existing research can no longer simply be compared, and applications that rely on findings from previous entrainment research would be difficult to implement. Additionally, if one manages to implement entrainment into a spoken dialogue system, it is crucial that any measurements of resulting entrainment are reliable and valid.

In this major methodological study, the same dataset was analysed using different methods, and the results of all methods were compared to one another. Prosodic entrainment was measured in conversations in an open-access corpus that is publicly available online. Twelve different methods (taken from the following papers: [18, 19, 20, 21, 22, 23, 24], see Table 1) for measuring prosodic entrainment were implemented. These methods differed in the time-frame in which they measured entrainment, as well as the (sub)types of entrainment they captured. Previously developed frameworks from existing literature were extended to categorise the twelve different methods, in an attempt to offer an initial investigation into how the results of different methods used to measure entrainment compare to one another. In the thesis, each method is explained and its characteristics, advantages, and disadvantages are discussed. The results of all methods are also presented and discussed in detail.

Table 1. Categorisation of methods described and used in the methodological comparison study. Please note that following the framework by [27], all time-series based methods are considered to measure entrainment globally.

		Time-scale		
		Local	Time-series	Global
D i m	P r o	Static local proximity local proximity, Levitan & Hirschberg (2011)	Static global proximity CRQA, Fusaroli & Tuyen (2016)	Static global proximity global proximity, Levitan & Hirschberg (2011)

e n s i o n	x i m i t y	Dynamic local proximity local convergence, Levitan & Hirschberg (2011)		Dynamic global proximity global convergence, Levitan & Hirschberg (2011) geometric approach, Lehnert-Lehouillier et al. (2020)
	S y n c h r o n y	Static local synchrony local synchrony, Levitan & Hirschberg (2011) linear mixed effects models, Schweitzer & Lewandowski (2013)	Static global synchrony TAMA, Kousidis et al. (2008) HYBRID, Looze et al. (2014) WLCC, Boker et al. (2002)	
			Dynamic global synchrony HYBRID, Looze et al. (2014)	

This study revealed that different methods do not necessarily produce the same results, even when they are used to analyse the exact same dataset. Various potential sources of this variation are discussed in the thesis, and one possible source of this variation in results is addressed in a follow-up study. In this follow-up study, I once again quantified entrainment in one dataset using different methods to compare these different approaches, but used three different norming procedures on the same raw data to investigate whether different norming procedures explain some of the variance in results that I found in my previous methodological study. To sum up, results suggested that the norming procedure accounts for some, but not all of the variance in results found in the study presented in the original methodological comparison.

The findings of these two studies have substantial theoretical and practical implications. Theoretically speaking, the results beg the question whether entrainment is one phenomenon, or whether it is a set of behaviours that may or may not be related. Practically speaking, the differences between methods make certain methods more suitable for specific types of datasets or research questions. Both these theoretical and practical implications are discussed in detail in the thesis.

Goal and contribution 3b: make entrainment research more accessible by sharing code to measure entrainment

An additional goal of the methodological evaluation is to make it easier for other researchers to measure entrainment in their data by making our analysis code available to use. To this end, all the code we used to measure and evaluate the different methods will be made publicly available. This can facilitate replication of existing research, which is a critical part of the scientific method, and can make entrainment research more accessible to other researchers, including those in other disciplines. We hope that by making our code available as tools to measure entrainment, a more complete picture of the phenomenon can be painted; and the better we understand entrainment, the better we will eventually be able to implement it.

In essence, this thesis aimed to provide contributions to the field in two major ways, that are both practical and theoretical:

- an initial investigation into the role of ToM in entrainment, including the creation of a clinical speech database, and
- an in-depth evaluation of the methods that are used to measure entrainment at the prosodic level.

Contributions and impact

First, the impact of the main results pertaining to ToM and entrainment will be presented, before the main results of the methodological part of the thesis will be discussed.

Theory of Mind and entrainment

Summary of goals and contributions

One of the primary goals of this thesis project was to investigate and characterise the role of theory of mind (ToM) in entrainment at the lexical and prosodic level (issue 1), in order to facilitate future research and implementation of entrainment in spoken dialogue systems. This goal was achieved by measuring entrainment at these two levels in adults with and without schizophrenia (goal 1a) and in children with and without autism spectrum disorder (ASD, goal 1b), two neurodevelopmental disorders that are associated with ToM impairments. Results from the first experiment suggest that neither group (with vs without schizophrenia) nor ToM score seemed to influence entrainment at either level, but suggest that task role may have an effect on entrainment. Results from the second experiment suggest that there was no difference in entrainment at either level between groups (with vs without ASD), but suggest that a different set of higher-order cognitive functions called executive functioning may influence entrainment at the lexical level, and that gender may influence entrainment at the lexical level as well. Taken together, the results from these two experiments seem to suggest that ToM is not required for entrainment.

The contributions provided by these experiments are novel for several reasons. First of all, little to no research has been done into entrainment in adults with and without schizophrenia, so even the initial exploratory investigation that I conducted, is, despite the relatively small sample size, a novel and relevant contribution. Additionally, the study investigating entrainment in children with and without ASD implemented an experimental paradigm that was less predictable and required more naturalistic turn-taking than tasks used in many other studies where lexical entrainment on target words was investigated in children with and without ASD. While less restricted, more naturalistic tasks limit the amount of control a researcher has over the data and thus introduces noise, such semi-naturalistic conversations more closely mimic real life and can thus provide insight into the mechanics of everyday human-human conversation, which in turn can inform the development human-machine interaction (HMI).

Impact

Results from the studies presented in this thesis seem to suggest that theory of mind (ToM) is not required for entrainment. This means that any clinical applications that assume a strong and direct link between entrainment and ToM, such as automatic screening tools or entrainment as a therapeutic target for interventions, are likely not the most effective targets. Nonetheless, the link between entrainment and other symptoms, such as executive functioning difficulties, can be explored in future research, as this may be a more promising therapeutic approach.

In the context of HMI, these findings suggest that modelling of ToM is probably not necessary for the implementation of natural-sounding entrainment behaviour. Further research can elucidate how

executive functioning shapes entrainment, and how gender influences entrainment behaviours, particularly in an HMI setting.

Importantly, the speech recordings of the children with and without ASD were compiled into a shareable database (goal 2) that can be used for future research. This database can be used to replicate and extend our findings on entrainment in individuals with and without ASD, to confirm whether ToM is not related to entrainment, but can also be used to study other conversation coordination mechanisms. The corpus is relatively large (73 interactions), so it can also be used to train and test various machine learning algorithms. The shareable database of speech recording and several psychometric test scores will be a valuable resource for research in various academic disciplines.

Methodologies in entrainment research

Summary of goals and contributions

The literature review conducted for the prior two experiments revealed that a large number of different methods is used to measure entrainment (issue 3). This opened up a second line of investigation for this thesis: to compare and evaluate a number of methods that are used to measure prosodic entrainment (goal 3). The same corpus was analysed using 12 different methods, and the outcome of the study was that there are discrepancies between the results of different methods. A follow-up study was conducted to investigate norming procedures as a potential source of this variation in results, and results of this study suggest that norming procedures account for some, but not all of the observed variability.

Impact

The findings of the studies into methods in prosodic entrainment research highlighted that different methodological decisions can influence the outcome of a study, and that different methods can lead to different results. In terms of direct impact, this finding means that results from existing studies that relied on different methods cannot be directly compared, as different methods may measure different subtypes of entrainment.

For the implementation of entrainment into spoken dialogue systems or other HMI applications, these findings highlight the importance of frameworks: different subtypes of entrainment should be specified following such frameworks, and different algorithms should be used to implement different subtypes of entrainment. Implementing different subtypes of entrainment into HMI applications may even provide opportunities for investigating the relationship between these different subtypes and the various potential social functions of entrainment.

An additional goal of these studies was to make entrainment research more accessible by sharing all the code we wrote for the different methods. Part of the code has already been made publicly accessible, and several of the methods have already been used by researchers with whom I collaborated on two papers that will both be published in the proceedings of the International Congress of Phonetic Sciences in 2023. The other code will also be made publicly available, and can be used by anybody to measure entrainment in their datasets, in hopes that more entrainment research will further our understanding of this complex, multifaceted phenomenon. This is another valuable contribution of this thesis.

More generally, the findings from these two methodological studies emphasise the importance of replicability in research, and raise a number of practical and theoretical points. Practically speaking, the studies outlined (dis)advantages for the different methods, and findings highlight the importance of detailed descriptions of any methods used in research. Moreover, these experiments illustrate the

importance of open science principles in the context of replication: if researchers make their analysis code available, analyses can more easily be replicated without the introduction of more methodological decisions and researchers' degrees of freedom.

Theoretically speaking, these studies raise several questions, such as how findings from existing studies can be compared, and whether "entrainment" is one construct, or a set of (potentially related) behaviours.

Conclusions

In conclusion, this thesis aimed to shed light on the possible relationship between theory of mind (ToM) and entrainment at two different linguistic levels, in order to scaffold future research and work into the implementation of entrainment in spoken dialogue systems. Results of two experiments and several different analyses suggest that there is no direct, systematic link between ToM and entrainment. However, a synthesis of the literature highlighted the large number of different methods that have been used to measure entrainment.

The second part of this thesis was dedicated to comparing and evaluating different methods for quantifying entrainment. Results suggest that there is little agreement between results from different methods, even when these methods are used to analyse the exact same dataset. Based on this, I argue that we need a more in-depth, interdisciplinary, and comprehensive understanding of what entrainment is and how it should be measured, before we focus on understanding the specifics of its psychological underpinnings or implementing entrainment into various applications.

Concretely, the work of this thesis resulted the creation of a shareable speech database, a pipeline for feature extraction, and a number of different shareable scripts used to measure entrainment, as well as valuable theoretical insights into the psychological mechanisms underlying entrainment and methodological considerations that could have far-reaching implications for the field.

References

- [1] J. Allwood, J. Nivre, E. Ahlsen, On the semantics and pragmatics of linguistic feedback, *Journal of Semantics* 9 (1992) 1–26.
- [2] H. Clark, *Using language*, Cambridge University Press, 1996.
- [3] E. Schegloff, Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences, *Analyzing discourse: Text and talk* 71 (1982)
- [4] H. Branigan, M. Pickering, A. Cleland, Syntactic co-ordination in dialogue, *Cognition* 75 (2000) 13– 25.
- [5] S. Brennan, H. Clark, Conceptual pacts and lexical choice in conversation, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22 (1996) 1482.
- [6] M. Pickering, S. Garrod, Toward a mechanistic psychology of dialogue, *Behavioral and brain sciences* 27 (2004) 169–190.
- [7] R. Levitan, A. Gravano, L. Willson, Beňuš, J. Hirschberg, 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, p. 11–19.
- [8] S. Borrie, N. Lubold, H. Pon-Barry, Disordered speech disrupts conversational entrainment: a study of acoustic-prosodic entrainment and communicative success in populations with communication challenges, *Frontiers in psychology* 6 (2015) 1187.
- [9] T. Chartrand, J. Bargh, The chameleon effect: the perception–behavior link and social interaction, *Journal of personality and social psychology* 76 (1999) 893.
- [10] R. Fusaroli, B. Bahrami, K. Olsen, A. Roepstorff, G. Rees, C. Frith, K. Tuyen, Coming to terms: Quantifying the benefits of linguistic coordination, *Psychological science* 23 (2012) 931–939.
- [11] A. Nenkova, A. Gravano, J. Hirschberg, High frequency word entrainment in spoken dialogue, in: *Proceedings of ACL-08: HLT, Short Papers*, 2008, p. 169–172.
- [12] D. Reitter, J. Moore, Alignment and task success in spoken dialogue, *Journal of Memory and Language* 76 (2014) 29–46.
- [13] C. Lee, M. Black, A. Katsamanis, A. Lammert, B. Baucom, A. Christensen, S. Narayanan, Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples, in: *Eleventh Annual Conference of the International Speech Communication Association*, 2010
- [14] J. Pardo, R. Gibbons, A. Suppes, R. Krauss, Phonetic convergence in college roommates, *Journal of Phonetics* 40 (2012) 190–197.
- [15] N. Lubold, H. Pon-Barry, Acoustic-prosodic entrainment and rapport in collaborative learning dialogues, in: *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, 2014, p. 5–12.

- [16] R. Levitan, S. Beňuš, R. H. Galvez, A. Gravano, F. Savoretti, M. Trnka, A. Weise, J. Hirschberg, Implementing acoustic-prosodic entrainment in a conversational avatar., in: *Interspeech*, volume 16, San Francisco, CA, 2016, pp. 1166–1170.
- [17] Beňuš, M. Trnka, E. Kuric, L. Martak, A. Gravano, J. Hirschberg, R. Levitan, Prosodic entrainment and trust in human–computer interaction, in: *Proceedings of the 9th International Conference on Speech Prosody*, International Speech Communication Association, Baixas, France, 2018, p. 220–224.
- [18] R. Levitan, A. Gravano, J. Hirschberg, Entrainment in speech preceding backchannels, in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011-06, p. 113–117.
- [19] A. Schweitzer, N. Lewandowski, Convergence of articulation rate in spontaneous speech, in: *INTERSPEECH 2013*, 2013-08, p. 525–529.
- [20] H. Lehnert-LeHouillier, S. Terrazas, S. Sandoval, Prosodic entrainment in conversations of verbal children and teens on the autism spectrum, *Frontiers in Psychology* 11 (2020) 2718.
- [21] S. Kousidis, D. Dorrán, Y. Wang, B. Vaughan, C. Cullen, D. Campbell, E. Coyle, Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues, in: *Ninth Annual Conference of the International Speech Communication Association*, 2008.
- [22] C. Looze, S. Scherer, B. Vaughan, N. Campbell, Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction, *Speech Communication* 58 (2014) 11–34.
- [23] S. Boker, J. Rotondo, M. Xu, K. King, Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series, *Psychological methods* 7 (2002) 338.
- [24] R. Fusaroli, K. Tuyen, Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance, *Cognitive science* 40 (2016) 145–171.
- [25] H. Clark, C. Marshall, Reference diaries, in: *Theoretical issues in natural language*, 1978, p. 2.
- [26] H. Clark, G. Murphy, Audience design in meaning and reference, in: *Advances in Psychology*, volume 9, North-Holland, 1982, p. 287–299.
- [27] C. Wynn, S. Borrie, Classifying conversational entrainment of speech behavior: An expanded framework and review, *Journal of Phonetics* 94 (2022) 101173.
- [28] M. Rasenberg, A. Özyürek, M. Dingemans, Alignment in multimodal interaction: An integrative framework, *Cognitive Science* 44 (2020), 12911.
- [29] H. Branigan, M. Pickering, J. Pearson, J. McLean, A. Brown, The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers, *Cognition* 121 (2011) 41–57.

Research output

Publications based on work in the thesis

Kruyt, J., & Beňuš, Š. (2021). Prosodic entrainment in individuals with autism spectrum disorder. *Topics in Linguistics*, 22(2), 47-61.

Cited in:

Patel, Shivani P., Cole, Jennifer, Lau, Joseph C. Y., Fragnito, Gabrielle, Losh, Molly. Verbal entrainment in autism spectrum disorder and first-degree relatives. In *Scientific Reports*, 2022, vol. 12, no. 1, pp. ISSN 2045-2322.

Weise, A., McNeill, M., Levitan, R. The Brooklyn multi-interaction corpus for analyzing variation in entrainment behavior. In *Proceedings of the 13th Conference on Language Resources and Evaluation Conference (LREC 2022)*. 2022, pp. 1721-1731.

Weise, A. (2022). *Towards Explaining Variation in Entrainment* (Doctoral dissertation, City University of New York).

Kruyt, J., Beňuš, Š., Faget, C., Lançon, C., & Champagne-Lavau, M. (2022, May). Prosodic and lexical entrainment in adults with and without schizophrenia. In *Speech Prosody 2022*.

Kruyt, J., Polónyiová, K., Ostatníková, D., & Beňuš, Š. (2023). Lexical entrainment on target words during task-oriented interaction in children with and without autism spectrum disorder. In *Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue* (in press).

Kruyt, J., Polónyiová, K., Ostatníková, D., & Beňuš, Š. (2023). Global and local prosodic entrainment in a task-oriented interaction in autistic and neurotypical children. In *Proceedings of the Connecting Multiple Disciplines to AI Techniques in Interaction-centric Autism Research and Diagnosis workshop at SigDial-INLG* (in press).

Kruyt, J., De Jong, D., D'Ausilio, A. & Beňuš, Š. (2023). Measuring prosodic entrainment in conversation: a review and comparison of different methods. In *Journal of Speech, Language, and Hearing Research* (in press).

Kruyt, J., Trnka, M., R. Sabo., R., & Beňuš, Š. (2023). Presenting the Slovak Autistic and Non-Autistic Child Speech (SANACS) corpus. (in preparation).

Other publications

Kruyt, J., Huttner, L., & O'Mahoney, J. (2023). Investigating the relationship between prosodic entrainment and interaction style. In *Proceedings of International Congress on Phonetic Sciences 2023* (in press).

Marekova, L., **Kruyt, J., & Beňuš, Š.** (2023). The effect of non-native language and task complexity on speech entrainment. In *Proceedings of International Congress on Phonetic Sciences 2023* (in press).

Kruyt, J., Huttner, L., & Gandolfi, G. (2023). Misalignment in Alignment Research: a descriptive, cross-disciplinary investigation into terminological preferences. (in preparation).

Invited talks

“Designing an experiment to measure speech convergence in children with and without autism” at the LingPril workshop in April 2022 in Nitra, Slovakia

“Do methods matter? Insights from research into entrainment and autism spectrum disorder” at the Language in Interaction workshop in July 2023 in Wuppertal, Germany