THE EFFECT OF (NON-)NATIVE LANGUAGE AND TASK COMPLEXITY ON SPEECH ENTRAINMENT

Lucia Mareková[1], Joanna Kruyt[2,3], Štefan Beňuš[1,2]

[1]Constantine the Philosopher University in Nitra, Slovakia
[2]Institute of Informatics, Slovak Academy of Sciences in Bratislava, Slovakia
[3]Slovak Technical University, FIIT in Bratislava, Slovakia
lucia.marekova@ukf.sk, joanna.kruyt@savba.sk, sbenus@ukf.sk

ABSTRACT

Speech entrainment (also known as alignment or accommodation) has been documented on various linguistic and paralinguistic levels but often with complex and sometimes conflicting outcomes. Hence, the understanding of the mechanism(s) underlying this complex behaviour in dyadic conversations is still limited. In an effort to increase this understanding, the current study tests the effect of language (L1 Slovak vs. L2 English) and task complexity (easy, medium, difficult) on local and global entrainment in f0 and intensity in a within-subject design. Pairs of undergraduates played a collaborative game of giving directions structured into three levels gradually rising in complexity. The results corroborate other recent findings in yielding complex and unexpected patterns, particularly failing to show the assumed link between cognitively easier tasks and greater entrainment.

Keywords: entrainment, task complexity, L1 vs. L2, cognitive load.

1. INTRODUCTION

Entrainment (also referred to as alignment, convergence or accommodation) is the tendency of communicative partners to get closer to each other in verbal or non-verbal aspects of their communication. The patterns of entrainment in speech and language characteristics have been extensively analysed in native oral interactions, bringing insights on the cognitive aspects of inter-personal entrainment on various linguistic and para-linguistic levels, such as semantics [1], syntax [2, 3, 4], prosody, and phonetic realizations [5, 6, 7], or focusing on diverse social and psychological factors underlying entrainment [e.g. 8, 9]. These studies show that entrainment is a complex phenomenon occurring on multiple levels of communication, depending on diverse underlying factors whose character is still not well understood. Given this complexity and inconclusive results of previous studies, various questions regarding the patterns of speech entrainment on multiple levels of communication are pending a satisfactory address.

For instance, entrainment on the interface of native (L1) and non-native (L2) is not well understood due to conflicting findings. On the one hand, L2 speakers were likely to imitate L1 speakers when sharing certain social identities [10], or they converged in vowel pronunciation towards those about whom they assumed were L1 speakers, while diverged from those about whom they assumed were L2 speakers [11]. On the other hand, [12] in a longitudinal study found that native English speakers converged in rhythm to L2 English speakers, while L2 speakers did not converge. A similar pattern was shown in [13]: L1 English speakers aligned towards Spanish-accented speakers in a shadowing task. Yet, [14] showed greater entrainment in matched language dyads than in mismatched (L1 vs. L2) dyads.

It is believed that metalinguistic factors play a significant role in these complex outcomes of L1 vs. L2 entrainment studies. For instance, the level of entrainment in L1-L2 dialogues may be influenced by the L2 speaker´s metalinguistic awareness of L1 speaker´s knowledge of the language [15], or by a set of other L2-specific issues such as insecurity in pronunciation patterns preventing the L2 speaker to lexically converge toward the L1 speaker, or linguistic similarity of interlocutors´ L1 and L2 affecting sensitivity to automatic priming [16].

To the best of our knowledge, the analysis of the patterns of entrainment on the interface of L1 vs. L2 as a within-subjects factor in semi-spontaneous conversations warrants further exploration. There are three relevant questions in this regard. First, is the speech entrainment language dependent? Second, do the same speakers align differently when interacting in their L1 compared to interacting in their L2? Third, is the level of entrainment dependent on the cognitive demand of the conversation? For example, acoustic analysis in [17] showed that entrainment was more significant in easier conditions with fewer errors and shorter completion time. A holistic analysis in [17] brought the same results.

Entrainment covers a wide range of behaviours and literature offers many methods for assessing it; different methods presumably tap into different aspect of this behaviour. One of several dimensions

that differentiates the methods is locality: some try to capture entrainment globally at the level of a conversation, for example comparing two halves of each dialogue, and some evaluate speakers' adjustments more locally at each turn-exchange, e.g. [18]. Previous studies showed that this dimension might show differences among languages and capture individual differences, e.g. [19,20].

In the present study, we test the relationship of entrainment in L1 (Slovak) vs. L2 (English) spoken dyadic communication between undergraduate students playing an interactive map game of giving directions sequenced into three levels of task complexity: easy, medium and hard.

Regarding the language effect on entrainment, theoretical accounts lead to contradictory predictions. On the one hand, the patterns of entrainment may be similar for L1 as for L2, mainly if we consider its dependence on various social mechanisms (dominance and roles), which are in our design the same in L1 as in L2. On the other hand, speakers´ perception of each other's L2 proficiency may cause differences in how the same speakers entrain towards each other in L1 vs. in L2 due to changes in social dynamics. For the task complexity effect on entrainment, we expect more entrainment in the easy and medium tasks than in the hard task, which is based on previous findings [17].

Understanding entrainment in L2 spoken interactions and its links to L1 speech entrainment has potential for 1) better understanding and ultimately formally modelling the cognitive nature of this process, and 2) developing communicative techniques for L2 learning informed by the patterns and functions of speech entrainment since more and more interpersonal spoken interactions takes place in L2. Specifically, our main hypothesis is that we observe more entrainment in L1 and easier tasks compared to L2 and harder tasks since lower cognitive demands might facilitate managing and navigating social space and entrainment as a means for doing that [17]. Alternatively, more prevalent entrainment observed for complex tasks and L2 might support the theories advocating an automatic, possibly priming-based, account of entrainment [21].

2. METHODS

Fifteen pairs (8 females, 3 males, 4 mixed) of native Slovak undergraduates studying English took part in the experiment. The subjects used a standardized test to self-evaluate their L2 proficiency, with results ranging from the B2 to C1 levels.

Each pair completed six collaborative communicative tasks involving giving directions based on an information gap. The students were randomly assigned the role of the guided person or the guide and kept the roles for all tasks. The main factors under investigation were language, L1 Slovak (SK) and L2 English (EN), and task complexity (easy, medium, difficult). The easy task, top of Fig. 1, included giving directions while the route was indicated. In the medium task, middle of Fig. 1, the route was not indicated, but the number of elements in the map and possibilities of route choice was low. The map of the difficult task, bottom of Fig. 1, included multiple elements and afforded several possibilities for the subjects to complete the task so that they had to make several collaborative decisions. This design resulted in 90 conversations (15 pairs x 3 complexity levels x 2 languages). The tasks were blocked by the language, i.e. all three complexity levels together, but the order of the language (SK first vs. EN first) and the order of the tasks within a language block were randomised.

Figure 1: Maps used for three complexity levels: easy (top), medium (middle), and difficult (bottom).

Our task design brought together a variety of previous approaches. Firstly, following [22], our tasks contain three basic elements: there is some kind of a gap, subjects rely on their own resources (linguistic and non-linguistic), and there is an apparent communicative outcome. Next, we distinguished task complexity from task conditions/communicative stress: gradually rising task complexity is not determined by planning time, time limits for completing the tasks, or the number of

participants [23,24]. Task complexity is rather created by manipulating the character of the tasks: In line with [24], our tasks differed in the number of elements occurring in the map and by the reasoning demand (in the difficult task, time labels were included urging the student to pick the fastest route and provide arguments to support their choice). The difference is also in the prior knowledge (in line with [23]): In the easy task, the students knew the area well, as it is the centre of a city in which their university is located. The medium task is situated in the same city but near the suburbs. The difficult task is situated in a distant city. The factor of prior knowledge was verified by a post-test questionnaire.

We used an automatic tool available in Praat to separate speaking time from silent time. The silence threshold was set to -25dB and minimum silence duration to 100ms. Manual correction of the silent vs. speaking intervals boundaries for samples with consistent errors was done. We then employed Praat for extracting the raw values for f0 medium, f0 range, and mean intensity from each speaking interval. We minimised the effect of outliers and octave jumps in f0 extraction by initially setting the floor and ceiling by the speaker biological sex and following the two-pass approach in [25] for floor/ceiling adjustment.

We analysed global, task-based entrainment following [26] that compares value means extracted from the initial and final thirds of the conversation respectively (using IPU numbers for the splitting into thirds). The local entrainment was assessed at each turn-exchange using the approach in [7] that operationalizes entrainment with local proximity, convergence, and synchrony; see both papers for definitions and formulas. Our implementation of both global and local methods is described in [27].

With values for global and local entrainment we then employed standard statistical modelling using linear mixed effects in R (package lmerTest) exploring the effect of LANGUAGE (L1 Slovak, L2 English), task COMPLEXITY (easy, medium, difficult), and their interaction on f0 median, f0 range, and intensity specifying random slopes for the dyad and the order of the language/tasks. We first explored the *anova* function on the model for the main effects and then the function *emmeans* in the package of the same name for the post-hoc comparisons in case the anova results returned significant values.

3. RESULTS

The effect of language and task complexity on the global entrainment in f0 median, f0 range and intensity is illustrated in Fig. 2. Positive values indicate entrainment and negative disentrainment. Pooling over the two factors, f0 range displays weak entrainment ($t(89) = 2.36$, $p = 0.02$) and intensity weak disentrainment ($t(98) = -2.21$, $p = 0.03$). Linear mixed effects modelling shows that neither the effects of LANGUAGE (SK vs. EN), COMPLEXITY, nor their interaction, on entrainment in f0 are significant. For intensity, there is a marginal interaction between LANGUAGE and COMPLEXITY ($F = 3.1$, $p = 0.05$), which is related to the fact that in EN, subjects dis-entrain more in medium tasks than in difficult tasks, while in SK it is vice versa.

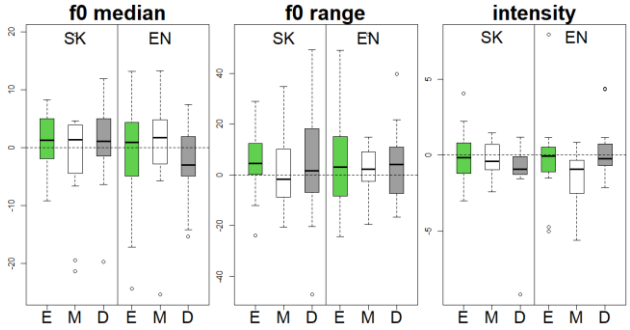

Figure 2: Global entrainment in f0 median (left), f0 range (middle), and intensity (right) divided by Language (leftmost 3 bars in each plot correspond to L1 Slovak, rightmost to L2 English), and Complexity (E[asy] - green, M[edium] - white, and D[ifficult] - grey).

The results pertaining to the local measures of entrainment are summarised in Table 1. For each of the three types of local entrainment (proximity, convergence, synchrony) and the three features (f0 median, f0 range, intensity), the table shows in the leftmost cell of each of the three columns if significant entrainment (ent) or disentrainment (dis) occurred when pooling over both factors and using a one-sample t-test. The rightmost cells show the significant patterns regarding LANGUAGE and COMPLEXITY.

	Proximity		Convergence		Synchrony	
f0 med	ent***	–	–	E_{EN}>M,D_{EN} M,D_{SK}>E_{SK}	dis***	–
f0 range	ent***	E_{EN}>E_{SK}	ent*	–	–	–
intensity	ent*	–	–	all>E_{SK}	dis***	–

Table 1: Local measures of entrainment with proximity, convergence, and synchrony. Task complexity in caps (E[asy], M[edium], D[ifficult]) and language (SK, EN) in subscript. "***" correspond to $p < 0.001$, and "*" to $p < 0.05$. See text for a detailed description.

In local proximity assessed on f0 range, the interaction between COMPLEXITY and LANGUAGE

returned a significant effect ($F = 3.84$, $p = 0.027$). The post-hoc exploration suggests that entrainment in English easy conversations was greater than in Slovak easy ones. Local proximity on f0 median and intensity did not show any significant effects of the two factors or their interaction despite the overall entrainment when both of these factors were pooled.

For local convergence in f0 median, dialogues in English were more entrained than in Slovak ($F = 10.13$, $p = 0.002$) the interaction with task complexity was also highly significant ($F = 11.17$, $p < 0.001$) stemming primarily from the difference between languages in the easy tasks ($F = 5.6$, $p < 0.001$): entrainment occurred in English and disentrainment in Slovak, which contrasted with the lack of (dis)entrainment in medium and difficult tasks.

No significant result was found in local convergence in f0 range. In intensity, however, a tendency for COMPLEXITY ($F = 3.0$, $p = 0.056$) and the significant interaction with LANGUAGE ($F = 6.43$, $p = 0.003$) were found. The pattern is partially similar to the one in f0 median: The disentrainment in Slovak easy tasks was significantly greater than the lack of entrainment in all other conditions.

Local synchrony yielded no significant effects of LANGUAGE or COMPLEXITY on the three dependent variables despite overall disentrainment in f0 median and intensity.

4. DISCUSSION AND CONCLUSION

The current study analysed the effect of language, native vs. non-native, and task complexity on local and global prosodic entrainment in f0 and intensity. The results do not yield a robust overall pattern and provide only several indicative systematic findings.

First, the effects of (non-)native language and task complexity are relatively small. We used one measure for global entrainment and three for the local one on three prosodic features. This gives 12 potential chances for these factors to affect inter-personal acoustic-prosodic entrainment. Yet, only 4 measures showed a significant effect (global intensity, local proximity f0 range, local convergence f0 median and intensity). Thus, we conclude that the patterns of speech entrainment tend to work similarly irrespectively of the language (native or non-native) or task complexity. This result might also suggest that in a learning environment, strategies positively influencing the link between speech entrainment and positive social aspects (social attractiveness, trust, task effectiveness, etc.) might provide comparable outcomes irrespective of the language of the interaction or the complexity of the task at hand.

Second, if the effect of language/complexity occurred, the easier tasks and tasks in the L1showed less entrainment (or showed disentrainment) than more difficult tasks and tasks in the L2. This was the case for three of the four significant patterns (local proximity f0 range, local convergence f0 median and intensity) where the easy tasks in Slovak showed the least entrainment. Hence, we might speculate that acoustic-prosodic entrainment in task-based spoken interactions is associated with coping with tasks demanding greater cognitive resources. The alternative possibility, that tasks with low cognitive demands free up resources for speech entrainment being used in managing and navigating social space, received little support despite findings of [17] where easier tasks yielded greater entrainment than more complex ones. Naturally, this speculation requires further rigorous testing.

Third, while the effect of language/complexity described above is relatively consistent across the selected entrainment measures, the assessment of entrainment itself is much less consistent. While f0 range showed entrainment both globally and as local proximity and convergence, intensity showed disentrainment globally and as local synchrony but entrainment as local proximity. This low consistency across entrainment measures corroborates recent studies, e.g. [28, 29], and supports efforts at researching how particular methods capture potentially different types of entrainment along potentially varied dimensions.

The sequencing of the tasks within one language was randomised as we wanted to prevent a possible uncontrollable effect of the task order. This might be a limitation of the study since sequencing tasks from simple to complex might be more relevant as the easier tasks gradually prepare the subjects for the harder tasks. Such sequencing is reasonable from a methodological point of view as a tool for increasing students´ oral competence. On the other hand, [23] showed that sequencing tasks from simple to complex increases accuracy, while sequencing tasks from complex to simple increases fluency. By randomising the order of the tasks, we aimed to prevent these diverse effects on subjects´ oral production. Furthermore, our descriptive analyses did not reveal any substantive systematic patterns regarding the order of the maps in our data.

6. ACKNOWLEDGMENT

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 and in part by the Slovak Granting Agency grant VEGA 2/0165/21, Slovak Research and Development Agency grant APVV-21-0373, and the Slovak University Granting Agency.

7. REFERENCES

[1] Boland, J. 2019. Conversation Transition Times: Working Memory & Conversational Alignment. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 159-165.

[2] Lee, J., Man, G., Ferreira, V., Gruberg, N. 2019. Aligning sentence structures in dialogue: evidence from aphasia. *Language, Cognition and Neuroscience* 34(6), 720-735.

[3] Branigan, H. P., Pickering M. J., Cleland, A. A. 2000. Syntactic coordination in dialogue. *Cognition* 75, B13-B25.

[4] Levelt W. J. M., Kelter, S. 1982. Surface form and memory in question answering. *Cognitive Psychology* 14, 78-106.

[5] Giles H., Coupland N., Coupland J. 1991. Accommodation theory: Communication, context, and consequence. In: Giles H. Coupland N. Coupland J, (eds). Contexts of accommodation: Developments in applied sociolinguistics. Cambridge University Press, Cambridge, 1–68.

[6] Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119, 2382-2393.

[7] Levitan, R., Hirschberg, J. B. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Interspeech 2011, 12th Annual Conference of the International Speech Communication Association*, 3081-3084

[8] Denby, J., Yurovsky, D. 2019. Parents' linguistic alignment predicts children's language development. *Cogsci*, 1627–1632.

[9] Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., Brown, A. 2011. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* 121(1), 41-57.

[10] Zając, M. 2013. Phonetic imitation of vowel duration in L2 speech. *Research in Language* 11(1), 19-29.

[11] Jiang, F., Kennison, S. 2022. The Impact of L2 English Learners' Belief about an Interlocutor's English Proficiency on L2 Phonetic Accommodation. *J Psycholinguist Res* 51, 217–234.

[12] Quené, H., Orr, R. 2014. Long-term convergence of speech rhythm in L1 and L2 English. *Proceedings of the International Conference on Speech Prosody*, 342-345.

[13] Lewandowski, E. M., Nygaard, L. C. 2018. Vocal alignment to native and non-native speakers of English. *Journal of the Acoustical Society of America*, 144, 620.

[14] Kim, M., Horton, W. S., Bradlow, A. R. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance, *Lab. Phonol.* 2(1), 125–156.

[15] Bortfeld, H., Brennan, S. 1997. Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes* 23, 119-147

[16] Costa, A., Pickering, M. J., Sorace, A. 2008. Alignment in second language dialogue. *Language and Cognitive Processes* 23(4), 528-556.

[17] Abel, J., Babel, M. 2017. Cognitive Load Reduces Perceived Linguistic Convergence Between Dyads. *Language and Speech* 60(3), 479-502.

[18] Wynn, C. J., Borrie, S. A. 2022. Classifying conversational entrainment of speech behavior: An expanded framework and review. *Journal of Phonetics* 94, 101173.

[19] Levitan, R., Beňuš, Š., Gravano, A., Hirschberg, J. 2015. Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison. Proceedings of SIGDIAL, 325-334.

[20] Menshikova, A., Kocharov, D., Kachkovskaia, T. 2020. Phonetic Entrainment in Cooperative Dialogues: A Case of Russian. Proc. Interspeech 2020, 4148-4152.

[21] Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences* 36(4), 329-347.

[22] Ellis, R., Skehan, P., Li, S., Shintani, N., Lambert, C. 2020. Task-Based Language Teaching: Theory and Practice. Cambridge University Press.

[23] Robinson, P. 2001. Task complexity, task difficulty, and task production: Exploring Interaction in a Componential framework. *Applied Linguistics* 22, 27-57.

[24] Robinson, P. 2005. Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching* 43, 1-32.

[25] De Looze, C., Rauzy, S. 2009. Automatic detection and prediction of topic changes through automatic detection of register variations and pause duration. In *Interspeech 2009, 10th Annual Conference of the International Speech Communication Association.*

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

[27] Kruyt, J., de Jong, D., D'Ausilio, A., Beňuš, Š. Submitted. Replicability in prosodic entrainment research: a methodological comparison.

[28] Weise, A., Levitan, R. 2018. 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), pp. 297-302.

[29] Weise, A., Levitan, S. I., Hirschberg, J., Levitan, R. 2019. Individual differences in acoustic-prosodic entrainment in spoken dialogue. *Speech Communication* 115, 78-87.