

Explorations in the Speakers' Interaction Experience and Self-assessments

Kristiina JOKINEN

UNIVERSITY OF TARTU, J.Liivi 2, 50409 Tartu

kjokinen@ut.ee

and

UNIVERSITY OF HELSINKI, PO Box 9 (Siltavuorenpenger 1 A)

Kristiina.Jokinen@helsinki.fi

ABSTRACT

The paper focuses on the interlocutors' self-evaluation in Finnish and Estonian first encounter dialogues. It studies affective and emotive impressions of the participants after they have met the partner for the first time, and presents comparison of the evaluation along the gender, age and education parameters. The results bring forward some statistically significant differences between the two groups, and point to different, culturally determined evaluation scales. The paper discusses the impact of the findings on the complex issues related to the evaluation of automatic interactive systems, and carries over to such applications as intelligent training and tutoring systems, and interactions with robots, encouraging further studies on the interlocutors' engagement in interaction and their evaluation of the success of the interaction.

KEYWORDS : dialogue, conversational engagement, self-assessment, cross-cultural evaluation

Kõnelejate suhtluskogemuse ja enesehinnangute uuringud

KOKKUVÕTE

Artikkel keskendub vestluskaaslaste enesehinnangutele esmakohtumisel peetud dialoogides soome ja eesti keeles. Uuritakse osalejate afektiivseid ja emotiivseid muljeid pärast seda, kui nad on kohtunud partneriga esmakordselt, ja esitatakse hinnangute võrdlus soo, vanuse ja hariduse parameetrite alusel. Tulemused toovad esile statistiliselt olulised erinevused kahe rühma vahel ja viitavad erinevatele, kultuuriliselt determineeritud hinnanguskaaladele. Artikkel analüüsib nende tulemuste mõju keerulistele probleemidele, mis on seotud automaatsete interaktiivsete süsteemide evalveerimisega, ja arendab edasi selliseid rakendusi nagu intelligentsed treenimis- ja õpetamissüsteemid ning suhtlus robotitega, pannes aluse edasistele uuringutele suhtlejate vestlusesse lülitumise ja vestluse edukuse hindamise kohta.

VÖTMESÕNAD : dialoog, vestlusesse lülitumine, enesehinnang, kultuuridevaheline evalveerimine

1 Introduction

A growing number of studies concerns how the interlocutors' affective state influences their experience and the interaction as a whole (e.g. Bavelas et al. 1986; Schroeder 2004; Lee et al. 2007; Mancini 2007; Nakano and Nishida, 2007), and how the speakers synchronise and coordinate their actions with each other (e.g. Goodwin, 2000; Kraemer and Swerts 2007; Heldner et al. 2010; Pickering and Garrod 2004; Battersby 2011). These aspects are regarded as signs of the speakers' cooperation and engagement: the speakers align and synchronise their behaviour, and show willingness to listen to their partner and provide coherent contributions. They also provide an important motivation to the design and evaluation of intelligent interactive agents, where user engagement is one of the core issues ranging from service oriented applications to amusing companions. Besides the traditional task completion, user's positive experience with the system is considered important for more natural interaction (Jokinen, 2009; Carlson et al. 2006).

It can be hypothesised that the more active the interlocutors appear to be in their communication, the more engaged they are in the conversation: their speaking frequency, tone of voice and body posture indicate interest and commitment to the topic of the conversation. Previous work has used such measures as utterance density (Campbell and Scherer, 2010; Jokinen, 2011), silence duration (Edlund et al. 2009), speech prosody (Levitan et al. 2011), lexical elements (Pickering and Garrod 2004), and eye-gaze (Levitski et al. 2012), among others. The interlocutors themselves usually describe such interactions afterwards as pleasant, enjoyable, and interesting.

This paper studies engagement and the interlocutors' experience in Finnish and Estonian interactions, and focuses especially on the interlocutor's self-evaluations. Self-evaluations are quick estimates of the interlocutors' conversational experience and provide complementary information to the studies that focus on studying engagement from the point of view of the interlocutors' verbal and non-verbal communicative signalling. It is hypothesised that engagement is related to the interlocutors' experience of the interaction in general, and the more engaged the interlocutors are in the conversation, the more positive their experience is. Such engaging events are described by positive affective adjectives, and thus the speakers' self-evaluation can reveal how they monitored the interaction on the affective level. By comparing the self-evaluations of two linguistically related participant groups, the paper also explores intercultural differences, and draws some interesting results concerning culturally determined interaction evaluation scales.

The paper first introduces the data and the questionnaire, then presents results from the comparison studies, and finishes with discussion and future research topics.

2 The data

2.1 First encounters video corpora

The corpus consists of first encounter dialogues collected in the Estonian MINT project (Jokinen and Tenjes, 2012) and in the Nordic collaboration project NOMCO (Navarretta et al. 2012). The projects aim at providing comparable databases for multimodal interaction studies in Nordic and Baltic languages. Dialogues are conducted between two people who are unfamiliar with each other. They are not given any specific topic to discuss, nor is there any external task to be solved. Each participant took part in two interactions, with a different partner. The Estonian corpus consists of 23 dialogues (12 male and 11 female participants, age range 21-61 years) while the Finnish data consists of 16 dialogues (8 male and 8 female participants, age range 20-59 years).

2.2 Questionnaire

Self-evaluation was conducted via a questionnaire that aims to measure naturalness and homogeneity of the interaction in terms of the speakers' experience. The descriptive features with negative and positive values were chosen following Nezlek (2010). The participants had to describe their experience of the interaction with respect to the given features using a 5-point scale, with 1 indicating that the adjective did not describe their experience at all, and 5 that did so very much. The web-based questionnaire was filled right after the person had had a videotaped interaction, and the participants were asked to fill it in quickly, so as to encourage first impressions rather than carefully considered responses. The adjectives were presented in the participants' mother tongue, and they are translated into English in Table 1.

The questionnaire also asked demographic information like gender, age-group, and education, as well as the person's familiarity with computers and video cameras. To find out how the descriptive features correlate within the two linguistic groups, a series of Student's t-tests were conducted for the two independent samples. Below we briefly go through the correlations along the gender, age, and education.

Descriptive feature	Estonian	Finnish
enjoyable	4.1	3.7
friendly	3.0	2.5
impressive	3.7	2.1
nice	4.1	4.0
interesting	4.1	3.8
relaxed	3.6	3.0
anxious	2.3	1.9
natural	3.4	2.6
happy	4.2	2.6
tense	2.0	1.9
awkward	1.9	2.1
angry	1.0	1.3
Average	3.1	2.6

TABLE 1 – Mean values for the descriptive features of the interaction. Boldface marks statistically significant differences ($p < 0.01$).

3 Self-assessment and experience

As seen in Table 1, Estonian speakers seem to provide more positive evaluations of their first encounters than the Finnish speakers (mean = 3.1, std deviation = 1.0, std error = 0.3 for Estonian, and mean = 2.6, std deviation = 0.8, std error = 0.2 for Finnish). Differences between three features (impressive, natural, and happy) are statistically significant ($p < 0.01$), and other big differences also occur concerning friendly, interesting and relaxed interaction, although these are not statistically significant. It is also interesting that Estonians ranked all 8 positive features over the neutral value 3, while Finns had only 4/8 features ranked so high. The positive views are further supported by the low evaluation of the negative aspects such as awkward and angry. As for the negative impressions in general, Estonians seem to rate their interactions slightly more anxious and tense, while Finns considered interactions slightly more awkward and angry, but these differences are not statistically significant.

3.1 Gender

Gender differences were not significant within either linguistic-cultural group. When the self-evaluation data is analysed across the cultures and languages, however, we find some statistically significant differences. Estonian male participants differ with respect to the features impressive, natural, and happy, from the Finnish male participants ($p < 0.05$). Also the Estonian female participants evaluate their interactions higher than the Finnish female participants with respect to these features, but also consider their interactions more interesting ($p < 0.05$). In general, Finnish and Estonian male participants seem to provide more similar self-evaluations along the different features, whereas the Finnish female participants were more critical of their interactions than the Estonian ones. Finnish female participants tend to rate their interactions lower concerning such features as friendly, interesting, and relaxed, yet also gave lower rates to negative aspects like anxious and tense. The distribution of self-assessment values with respect to male and female participants is given in Figure 1 and Figure 2, respectively.

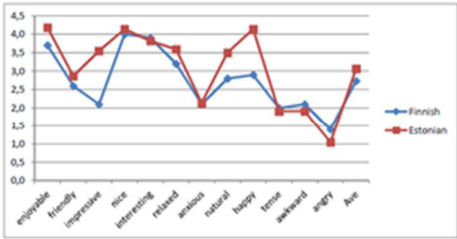


FIGURE 1 – Significant differences between Finnish and Estonian male evaluations concern impressive, natural, and happy interactions.

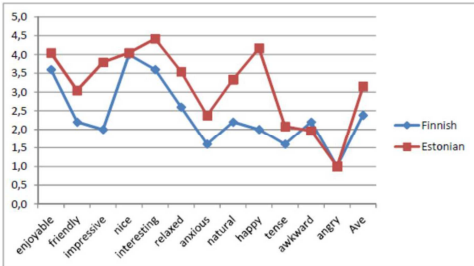


FIGURE 2 – Significant differences between Finnish and Estonian female speakers concern impressive, interesting, natural, and happy interactions.

3.2 Age

The participants were mostly young adults although both populations had one over 55 years of age. Since the two linguistic groups were not balanced with respect to age, it is not possible to draw differences in this respect. However, if we look at the differences across the languages, we can distinguish two age groups, those under 30 years and those above, and we find that the younger Estonians tend to evaluate the interactions significantly ($p < 0.01$) more impressive and happier than the younger Finns, and also more relaxed (significance level $p < 0.05$). Following the same tendency, younger Finns also considered interactions slightly more tense, awkward, and

angry than the younger Estonians (Figure 3). Concerning older participants, Figure 4 shows that the older Estonians had stronger experience of their conversations being friendlier, more natural, and happier than what the older Finns do ($p < 0.01$). However, it is interesting, that the older Finns describe their interactions slightly nicer and more relaxed than the old Estonians, and concerning negative impressions, they also find interactions less anxious and tense than the older Estonians, although as mentioned, the differences are not statistically significant. As a summary, it seems that in our first encounter data, the younger and older participants had opposite experiences: the younger Estonians rate their interactions more relaxed and less tense than the younger Finns, while the older Estonians rate interactions less relaxed and more tense than the older Finns.

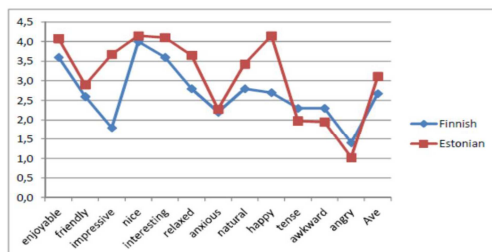


FIGURE 3 – Significant differences between Finnish and Estonian younger interlocutors (age < 30) concern impressive, relax, and happy interactions.

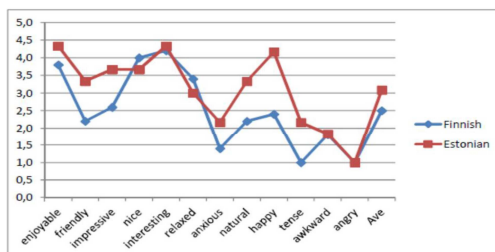


FIGURE 4 – Significant differences between Finnish and Estonian older interlocutors (age > 30) concern friendly and happy interactions.

3.3 Education

The participants' education ranged from undergraduate students to those who had completed Master's degree. It is interesting that this parameter draws statistically significant differences also within the linguistic groups: Estonian students found interactions more relaxed and less anxious than those who had completed their degree, yet also less interesting, while the Finnish students tend to consider interactions less friendly and interesting, yet more anxious and tense than those who had completed their degree.

Across the linguistic groups, Estonian students regarded interactions as more impressive and happier than the Finnish students (Figure 5), while those with a degree, considered interactions also more friendly and natural than the Finns (Figure 6). It is interesting that the student

descriptions in both language groups seem to be rather similar, in particular considering the negative features, but the differences grow bigger between those with a degree. This seems to be due to the change in the Finnish participants' evaluation: the Finnish degree holders tend to rate their interactions less anxious, tense and awkward than the Finnish students, while the difference between Estonian students and Estonian degree holders does not vary as much.

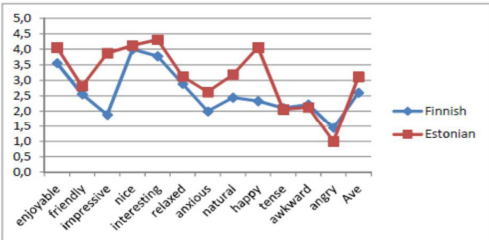


FIGURE 5 – Differences between Finnish and Estonian undergraduate students.

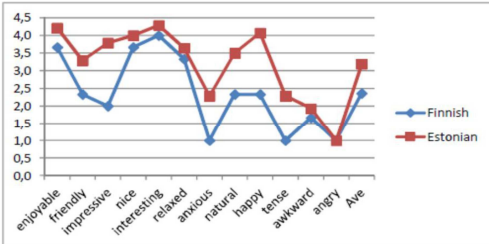


FIGURE 6 – Differences between Finnish and Estonian participants with a master's degree.

Conclusion and perspectives

This paper has focussed on the interlocutors' self-evaluation and impressions on Finnish and Estonian first encounter interactions, and compared the evaluations between two linguistically related participant groups. Statistically significant differences were found between the groups with respect to gender, age, and education, and concerning the evaluation parameters impressive, natural and happy, but due to a small dataset, we refrain from drawing conclusions concerning cultural characteristics.

However, it is noticeable that the Estonian groups consistently give higher evaluation scores for their interactions, and this is consistent regarding all descriptive categories, i.e. also their negative impressions were evaluated more strongly, although the differences were not as big as with the positive ones. We can assume that the differences are due to the different evaluation scales the participants use within their linguistic and cultural contexts. The Estonians' higher scores in all descriptive categories seem to suggest that the starting point of their evaluation scale was set high whereas the Finns tend to describe their impressions of the interaction using less extreme ends of the evaluation scale, thus giving lower scores to their experience compared with the Estonians.

However, even though the assessment scores may be low, the interaction itself may be considered “normal” and typical within the cultural group. In our case, it is impossible to say whether, in absolute terms, the Finns experience their interactions less impressive, natural or happy than the Estonians, or if the Estonians experience their conversations in more extreme terms than the Finns. Nevertheless, there is a clear difference in the way the Finns and Estonians describe their interactive behaviour, and we can only conclude that even within neighbouring countries, with closely related languages, the assessment of one’s engagement in interactive situations becomes a complex issue that is not necessarily related to straightforward measurements of overt behaviour. The cultural context deals with social norms and relationships which constrain the appropriateness of interactive behaviour, and also affects the way the behaviour is presented and assessed within the group.

Consequently, this study has impact on the evaluation and annotation methodology for various NL related systems that concern the speakers’ attitudes, affection, and emotions: the speaker’s judgements do not depend only on the personal characteristics of the users, but also on the different evaluation scales, which seem to depend on a larger cultural context. Annotations dealing with such issues as sentiment analysis, emotion categorisation and intercultural communication, thus need to take into consideration the variation that stems from culturally bound subjective evaluation scales. The design and evaluation of interactive systems also includes modelling of the partner’s affective and emotional state, and the correct interpretation of subtle behavioural signals thus requires exposure to the cultural context in which the interaction takes place. For instance, a tutoring system that tries to assess the student’s level of interest in intercultural context may reach a wrong conclusion that the student is not interested if the student’s self-assessment is not high enough, and this may lead to inappropriate strategies on how to continue the interaction.

Future work concerns more detailed studies of the notion of engagement and its relation to the subjective impressions of the success of interaction. The results can be applied e.g. to robot interaction where the robot tries to engage human in conversations about interesting Wikipedia topics (Jokinen and Wilcock, 2012). It is also important to work on more discriminating self-assessment questionnaires and methods to assess interactions in a more detailed manner, and to refine the set of descriptive parameters to cover the crucial aspects. For instance, it is useful to distinguish parameters that deal with different view-points: the speaker, the partner, and the general view of the interaction as a whole. Related to this it is also possible to extend the work with multimodal features and to correlate self-evaluations with the speakers’ multimodal activity, automatically extracted from the video. Finally, to better understand human communication in multicultural contexts, it is necessary to investigate differences in a larger population.

Acknowledgments

Thanks go to all the participants who took part in the video recordings, and to Matti Hocksell and Joonas Toivio in Helsinki and Sven Laater and Anti Torp in Tartu for their assistance in the data collection and analysis. Thanks go also to Mare Koit for her support of the research, and for the translation of the abstract into Estonian.

References

- Battersby, S. (2011). Moving Together: the organization of Non-verbal cues during multiparty conversation. PhD Thesis, Queen Mary, University of London.
- Bavelas, J. B., A. Black, C. R. Lemery and J. Mullett (1986). "I show how you feel": Motor mimicry as a communicative act. *Journal of Personality and Social Psychology* 50(2): 322-329.
- Campbell, N. and Scherer, S. (2010). Comparing Measures of Synchrony and Alignment in Dialogue Speech Timing with respect to Turn-taking Activity. *Proceedings of Interspeech 2010*. Makuhari, Japan
- Carlson, R., J. Edlund, M. Heldner, A. Hjalmarsson, D. House and G. Skantze (2006). Towards human-like behaviour in spoken dialog systems. Swedish Language Technology Conference (SLTC). Gothenburg, Sweden.
- Edlund, J., M. Heldner and J. Hirschberg (2009). Pause and gap length in face-to-face interaction. *Proceedings of Interspeech 2009*. Brighton, UK.
- Goodwin, C. (2000). Action and embodiment within Situated Human Interaction. *Journal of Pragmatics*, 32:1489-1522
- Heldner, M., Edlund, J. and Hirschberg, J. (2010). Pitch similarity in the vicinity of backchannels. *Proceedings of Interspeech 2010*, Makuhari, Japan.
- Jokinen, K. (2009). *Constructive Dialogue Modelling – Speech Interaction and Rational Agents*. John Wiley & Sons, Chichester, UK.
- Jokinen, K. (2011). Turn taking, Utterance Density, and Gaze Patterns as Cues to Conversational Activity. *Proceedings of The International Conference on Multimodal Interaction (ICMI-2011) Workshop on Multimodal Corpora for Machine Learning (MMC)*, Alicante, Spain.
- Jokinen, K. and Tenjes, S. (2012). Investigating Engagement – Intercultural and Technological Aspects of the Collection, Analysis, and Use of Estonian Multiparty Conversational Video Data. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. European Language Resources Association (ELRA).
- Jokinen, K. and Wilcock, G. (2012). Multimodal open-domain conversations with the Nao robot. *Proceedings of the 4th International Workshop on Spoken Dialogue Systems (IWSDS 2012)*, Paris.
- Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57 (3), 396-414.
- Lee, J., Marsella, T., Traum, D., Gratch, J., and Lance, B. (2007). The Rickel Gaze Model: A Window on the Mind of a Virtual Human. In C. Pelachaud et al. (Eds.) *Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA-2007)*. Springer Lecture Notes in Artificial Intelligence 4722, pp. 296-303. SpringerVerlag Berlin Heidelberg.
- Levitan, R., Gravano, A. and Hirschberg, J. (2011). Entrainment in Speech Preceding Backchannels. *Proceedings of ACL 2011*, pp. 113-117.

- Levitski, A., Radun, J., and Jokinen, K. (2012). Visual interaction and conversational activity. *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Eye Gaze and Multimodality*. Santa Monica, USA
- Mancini, M., Castellano, G., Bevacqua, E. and Peters, C. (2007). Copying Behaviour of Expressive Motion. *Lecture Notes in Computer Science, Computer Vision/Computer Graphics Collaboration Techniques*, 4418:180-191.
- Nakano, Y. and Nishida, T. (2007). Attentional Behaviours as Nonverbal Communicative Signals in Situated Interactions with Conversational Agents. In Nishida, T. (Ed.) *Engineering Approaches to Conversational Informatics*, John Wiley.
- Navarretta, C., Ahlsén, E., Allwood, J., Jokinen, K., and Paggio, P. (2012). Feedback in Nordic first-encounters: a comparative study. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. European Language Resources Association (ELRA).
- Nezlek, J. B. (2010). Multilevel modeling and cross-cultural research. In D. Matsumoto and A. J. R. van de Vijver (Eds.) *Cross-Cultural research methods in psychology*. Oxford.
- Pickering, M. and Garrod, S. (2004). Towards a mechanistic psychology of dialogue, *Behavioral and Brain Sciences* 27:169– 226.
- Schroeder, M. (2001). Emotional Speech Synthesis: A review. *Proceedings of the 7th European conference on Speech Communication and Technology (Eurospeech'01/Interspeech)*, Aalborg.

