Local and global convergence in the temporal domain in Polish task-oriented dialogue

Maciej Karpiński[1], Katarzyna Klessa[1], Agnieszka Czoska[2]

[1]Institute of Linguistics, [2]Institute of Psychology,
Adam Mickiewicz University in Poznań, Poland
maciej.karpinski@amu.edu.pl, klessa@amu.edu.pl, agaczoska@gmail.com

10.21437/SpeechProsody.2014-137

Abstract

Conversational parties tend to mutually adapt their communicative behaviour in a number of dimensions, from the level of physical aspects of speech signal and gesture, utterance properties, up to the level of mental representations. In the present study, an attempt is made to track the process of convergence in the temporal domain both as a global tendency and a local phenomenon. The material under study consists of two sets of task-oriented dialogues recorded with or without eye contact (telephone conversations) between the speakers. All the recordings were segmented into syllables and analysed in terms of speech rate and syllable timing pairwise variability (s*nPVI)* for each speaker as well as for the correlations between the speakers in each pair. Global convergence tendencies were proven to be weak but some influence of dialogue settings and gender was found. The results seem to support the hypotheses that the alignment-related processes remain under the influence of many factors related to the dialogue flow and cannot be modelled as simply incremental.

Index Terms: alignment, convergence, timing, dialogue, recording scenario

1. Background and previous studies

It is generally agreed that participants of dialogue tend to mutually accommodate, align and converge in various domains [1,2] that include the phonetic level of utterances [3,4,5,6,7], lexicon and syntax [8], postures, gestures and facial expressions [9,10,11] and, ultimately, mental models of situations [2,12].

Alignment is sometimes viewed as an almost automatic process based on priming [13] but it may be well driven by conscious communicative strategies [14]. Increased alignment has been shown to predict better communication efficiency, outcomes, and, presumably, better results of common task solving [15,16,17]. On the other hand, convergence has its limits: when mimicry goes too far it may result in the impression of parody. Moreover, in many communicative situations alignment may not work as a fully symmetrical process of increasing convergence between speech parameters of dialogue participants. Due to a higher social position, more assertiveness or some other factors, one of the dialogue parties may become the leader, while the other may decide to take a more subordinate role just not to inhibit the flow of communication [18].

The temporal aspects of dialogue interaction early attracted the attention of researchers, especially on the grounds of conversational analysis, where the precision in the arrangement of turn-taking has been often stressed [19]. More recently, formal approaches to turn-taking modelling have become more influential [20,21,22] and some more light has been shed on turn-boundary phenomena [7]. A number of studies have been devoted to the convergence of phonetic and especially of prosodic parameters of speech between conversational partners [23,4,24,25,26,28]. Nevertheless, many other aspects of the prosodic and temporal dynamics of dialogue remain less explored or, although extensively analysed, they still require more empirical evidence. Speech rate interdependencies in dialogue have been analysed in a limited number of studies, e.g. [3,5]. This type of prosodic alignment may be highly sensitive to various phenomena typical of spontaneous communication (hesitations, repairs, pauses, etc.) as well as the characteristics of the communication situation (external distractors) that can easily distort or disrupt it.

2. The aim of the study

According to the results of the abovementioned studies, the parties of dialogue tend to align locally or globally, and this process may also include convergence in the temporal domains, e.g., in the domain of speech rate and speech rhythm irregularity. The aim of the present study is to explore this type of convergence in the recordings of two types of Polish task-oriented dialogues (mutual visibility vs. lack of mutual visibility). Speech rate changes and speech rhythm irregularities may reflect important aspects of dialogue flow, including various dialogue fluency stages as well as task realisation stages and, in general, the quality of dialogue interaction [29].

3. Data and analysis

3.1. Recordings, segmentation and transcription

The recordings of task-oriented dialogues come from two Polish corpora: DiaGest2 corpus [30] and the Paralingua corpus [31]. Altogether, we analysed data based on the recordings of 40 voices, 20 from each of the corpora: 15 female speakers and 5 male speakers from the DiaGest2 corpus; 12 females and 8 males from the Paralingua corpus. All the recorded sessions were either conversations of two female speakers or a male and a female (no dialogues between two male interlocutors). In the DiaGest2 corpus, the task of the participants resolved itself in re-creating a figure made of paper by one of the participants (Instruction Follower, IF) according to instructions by the other (Instruction Giver, IG).

Only IG could see the original figure. IF was provided with all the materials necessary to re-construct the figure. IG and IF could see each other. All the IFs were females while there were five females and five males among the IGs. The speakers were given a time limit of 5 minutes to solve the task. The recordings from the Paralingua corpus dialogue were obtained based on a dialogue task in which the speakers were asked to find the differences between two pictures of a room. The task was to co-operate with the interlocutor in order to find as many differences as possible in the shortest possible time. The major differences between the tasks in DiaGest2 were the asymmetry (DiaGest2) vs. symmetry (Paralingua) of speakers' roles and mutual visibility (DiaGest2) vs. the lack of mutual visibility (Paralingua).

The material used for the present study was segmented into syllables. It was transcribed orthographically on the word level of segmentation and phonemically on the syllable level. Filled pauses were marked on a separate tier. Two software tools were used for transcription and annotation: Praat [32] (annotation and transcription of the DiaGest2 corpus) and Annotation Pro [33] (annotation and transcription of the Paralingua corpus as well as further annotation-mining and analyses of both corpora).

3.2. Analysis of speech rate changes and syllable duration variability

The data were explored using a "moving time window" approach, where a selected parameter was measured and averaged for each speaker within the window that was moving along the time axis. A similar method was used in [5,34] where additionally weights were used in the formula applied to measurements of accommodation of various acoustic/prosodic features, e.g., pitch, intensity or speech rate (thus using a weighted mean, where the interval durations were the weights). Four different window sizes were tested (5, 10, 30 and 60 seconds) in order to choose the optimum one, while the step was calculated as a proportion of the window size (33% in each case). Extremely narrow ones resulted in a substantial number of empty frames (i.e., not including any speech on one or both sides of the conversation) and thus did not support time-aligned analysis of dialogue turns. Applying very wide time windows helped to avoid this issue but, on the other hand, it might have concealed small-scale phenomena. As a consequence, the time window size was a result of a compromise. For most of the below calculations, the size of 30 seconds was applied as the smallest one that allowed to avoid empty frames.

Two timing-related parameters were taken into account. A "net" measure of speech rate (henceforth *NSR*) was calculated as the number of syllables within a time window (including those partially contained in the window as the respective fractions and excluding all kinds of pauses) divided by their total duration. Additionally, using a similar "moving time window" approach, a syllable-based *nPVI* (henceforth *snPVI*) [35] was calculated for each time window and each speaker in the dialogues under study. In case of the segments partially contained in a given window, their total durations were used in the *nPVI* formula. The *nPVI* was used with syllable-sized segments; cf. [36,37,38,39] as regards the definitions and plausibility of basic units for the analysis of pairwise variability.

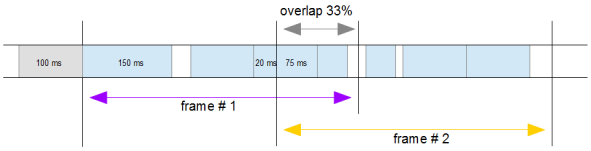

Figure 1. Segment rate moving average scheme

The transcripts prepared in Praat were converted to Annotation Pro native format and processed further using two Annotation Pro plug-ins specifically designed for the abovementioned calculations of *NSR* and *nPVI* (available from http://annotationpro.org/plugins/).

The results were exported as CSV files and analysed using SPSS statistical package (IBM Corp., 2012). Analyses included descriptive statistics calculated for all the speakers and dialogues, and regression analyses as well as some others, inspired by the data themselves.

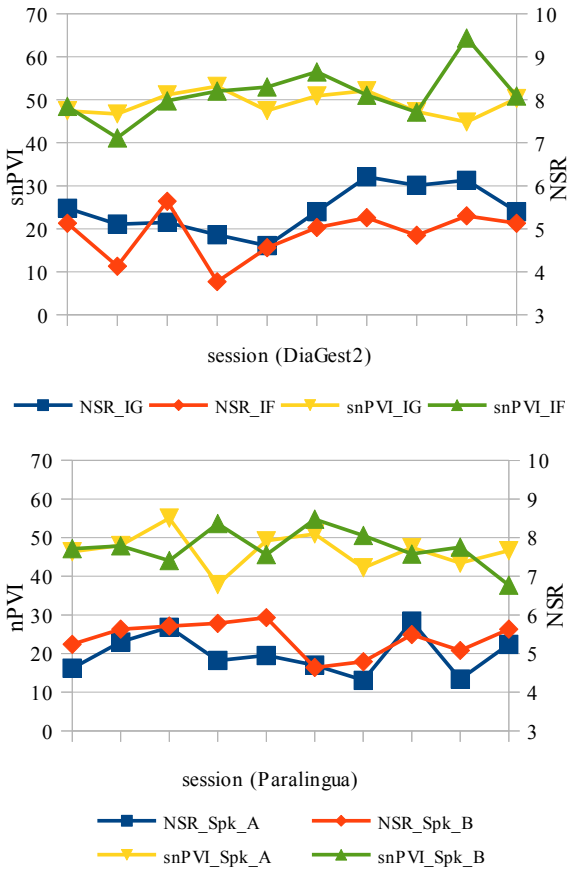

Figure 2. Mean normalised speech rate and mean snPVI for DiaGest2 (upper panel) and Paralingua (lower panel) dialogues

3.2.1. Speech rate analysis

The mean values of *NSR* calculated from 30-second time windows 5.23 (std.dev.=0.93, range=1.2–8.4) for DiaGest2 and 5.13 (std.dev.=1.01, range=3.2–7.4) for Paralingua. The difference between the mean values of the NSR for the two data sets turned out to be non-significant at p=0.05. DiaGest2 data were analysed additionally to test the difference between IG and IF speech rate (using Student's t-test). For the

difference was significant for both the 30-second time windows (IG mean=5.43, std.dev.=0.99; IF mean=4.88, std.dev.=1.19; p>0.0001) and the 60-second time windows data (IG mean=5.51, std.dev.=0.72; IF mean=4.92, std.dev.=0.86; p>0.0001).

Linear regression analysis was carried out for the data based on 10, 30 and 60 second time windows. While for the smallest window size no significant results have been found in the two corpora, with a 30 second window, four significant results of regression analysis were found in the DiaGest2 data and one in Paralingua. For the 60 second window, one significant result of regression was found in each corpus.

Significant results for the 30 second time window were found in dialogues 2 (R=0.314, p=0.024), 6 (R=0.629, p<0.001), 8 (R=0.415, p=0.007) and 9 (R=0.39, p=0.013) from the DiaGest2 data and dialogue 12 (R=0.432, p=0.006) from the Paralingua corpus. The significant results for the 60-second frame were found in different dialogues: dialogue 3 (R=0.617, p=0.021) from the DiaGest2 corpus and dialogue 7 (R=0.983, p=0.008) from the Paralingua corpus.

The results may suggest that alignment, understood as linear correlation between the speech rates of interlocutors, occurres rarely in this kind of task dialogues. On the other hand, when the phenomenon takes place, the correlation is very strong (98% in dialogue 7 from the Paralingua data) or medium (40%–60% in the other dialogues). Moreover, local alignment (30-second time window) occurred more often, than the global one. In this data alignment occurred more often in dialogues which participants had different roles (IG and IF). Although IF's speech rate showed a global tendency to be lower, the difference did not prevent the occurrence of alignment, but might even encourage it.

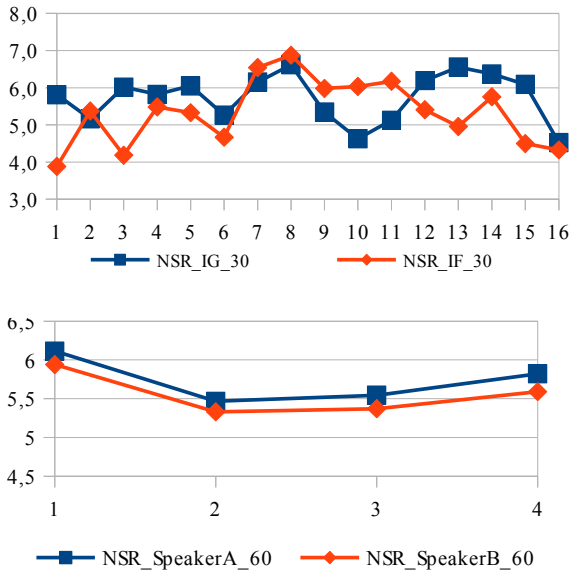

Figure 3. *NSR* changes in dialogue 6 from DiaGest2 corpus (30 second window; upper panel) and in dialogue 7 from Paralingua (60 second window; lower panel)

An additional Chi-squared test was used to test the hypothesis that alignment would be more frequent when interlocutors are of different sex. Significant and strong results were found for the DiaGest2 data alone (Pearson's chi-squared=23.333, p<0.001) in contrast with non-significant results for the Paralingua data (Pearson's chi-squared=1.667, p=0,4). The results were not significant for the 30-second time window data including both DiaGest2 and Paralingua results (Pearson's chi-squared=3.3, p=0.069). In DiaGest2 data three occurrences of alignment were found between participants of different sex (out of five mixed pairs), and only one (dialogue 2) in a dialogue between two women. This may suggest that there is a higher probability of the occurrence of the local alignment between interlocutors of different sex. When it comes to the 60-second time window, the two cases of alignment found occurred in dialogues between two women (due to the small number of alignment occurrence no Chi-squared test was performed). This result may suggest further differences between local and global alignment.

3.2.2. The analysis of snPVI

The mean value of *snPVI* (30-second time window) equalled to 50.6 (std.dev.=8.8, range=20.4–80.30) for DiaGest2 and 48.2 (std.dev,=6.1, range=35.2–70.1) for Paralingua. The difference between the mean values of *snPVI* between the DiaGest2 and Paralingua data turned out to be statistically significant at p<0.01. DiaGest2 data were analysed additionally to test the difference between IG and IF speech rate. For the difference was significant for the 30-second time windows (IG mean=49.22, std.dev.=6.58; IF mean=51.41, std.dev.=11.84; p=0.041) but non-significant for the 60-second time windows data (IG mean=48.27, std.dev.=9.15; IF mean=49.73, std.dev.=12.69; p=0.39).

For the values of the *snPVI* linear regression analysis was carried out for each pair of speakers using 30 and 60 second time windows. Results were found significant (p<0.05) for two pairs for each size of the time window in the DiaGest2 data and for three pairs at 30 second time window in the Paralingua data set.

The significant results for 30-second window were found in dialogues 6 (R=0.306, p=0.026) and 10 (R=0.338, p=0.014) from the DiaGest2 data and dialogues 2 (R=0.502, p=0.022), 14 (R=0.896, p=0.004) and 15 (R=0.816, p=0.005) from the Paralingua data. The significant results for the larger time window were found in dialogues 6 (R=0.86, p<0,001) and 7 (R=0.631, p=0.011) from the DiaGest2 data.

Again alignment occurred rarely, but the correlations were often medium (dialogue 2 from DiaGest2) or strong. Only two weak correlations were found within the local time window data (dialogues 6 and 10 from DiaGest2). Like previously, local alignment occurred more frequently than global and there was more alignment in the DiaGest2 corpus. The difference in the mean *snPVI* between IG and IF participants has disappeared for the 60-seconds time window data but prevailed for the local data.

Significant results (the occurrence of alignment) were found in dialogue 6 from the DiaGest data for both measures for the local time window, and in both windows for *snPVI*. This dialogue seems to be exceptional, since all the other dialogues show alignment only once across all the analyses.

Chi-squared tests were also performed for the *snPVI* alignment results. Significant results were found for the DiaGest2 data (Pearson's chi-squared=25.0, p<0.001 for 30-second frame, Pearson's chi-squared=25.0, p<0.001 for 60-

second frame). The results are the same, since alignment occurred in both cases within two mixed pair dialogues.

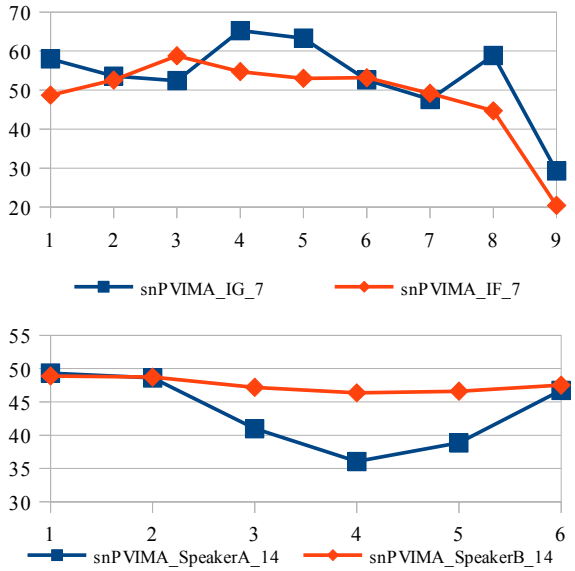

Figure 4. *snPVI* changes in dialogue 7 from DiaGest2 corpus at 60-second time window (upper panel) and in dialogue 14 from Paralingua corpus at 30-second time window (lower panel)

The results were not significant for the data from both corpora analysed together (Pearson's chi-squared=1.818, p=0.479 for 30-second frame, Pearson's chi-squared=2.716, p=0.127 for 60-second frame) or the Paralingua data separately (Pearson's chi-squared=1.27, p=1.0 for 30-second window). Results obtained from the DiaGest2 data may suggest a correlation between interlocutors sex and the probability of alignment. This suggestion may be strengthened by the fact, that two, out of three cases of alignment from the Paralingua data occurred in mixed pairs. However, the test results were non-significant for the whole data.

4. Conclusions and further work

In the present study, an attempt was made to explore and detect inter-speaker convergence in the dimensions of speech rate *(NSR)* and syllable timing pairwise variability *(snPVI)*. According to Kousidis [34], alignment in task-oriented dialogues may be much more difficult to detect. Nevertheless, such attempts have been made, e.g. [28], although results are not always clear-cut.

The overall speech rate means observed for both of the analysed corpora are similar (5.23 vs. 5.13 syll/sec) although the min-max range appeared to be wider in the DiaGest corpus (the overall rates are also in line with the rates for Polish normal reading rate [39]). Malisz [40] reported a higher overall mean around 6.9 syll/sec for Polish dialogues but using only fluent and coherent utterances with no unintelligible parts, false starts or hesitation markers while in the present work all utterances were used with the only exception of the segments labelled as filled pauses that were arbitrarily excluded from the analyses.

The results may suggest that the SR co-variability may be prevalently a local phenomenon that tends to occur most clearly within homogeneous, fluent stretches of conversation and may work very quickly and precisely as in the case of syllable boundary alignment [7]. On the other hand, it does not exclude the possibility of long-term, more permanent speech rate and syllable length variability accommodation even in the participants of task-oriented dialogues [23,34]. Verification of this claim, however, requires studies on larger corpora so that less prominent tendencies can be detected.

The influence of participants' gender on the level of *NSR* and *snPVI* convergence was also investigated. The results for the DiaGest data are coherent with Street's observations [3] and show less convergence in female-female pairs. As no influence was observed in the Paralingua data, one may hypothesize that the results may be somehow related to the mutual visibility condition. However, the factor of mutual visibility may somehow account for worse results in the timing-related convergence independently of gender matching.

In the DiaGest2, more manual activity (re-contructing a figure made of paper) was necessary while in Paralingua task, the participants had to scrutinize images (in order to find differences). It is difficult to judge which of the tasks tended to consume more attention and re-direct it from the partner to an object.

The differences in the overall ranges of *NSR* and *snPVI* might reflect the discrepancies in timing strategies applied depending on the presence or absence of "symmetry" in the task-solving roles. These observations are aimed to be further investigated also using other methods of syllable duration analysis, e.g. those accounting for deceleration/acceleration patterns [41,42] as well as more detailed analysis of pausing schemes. Note that although filled pauses were excluded from the present computations, it was observed that in DiaGest the number of pauses was considerably higher for IGs than for IFs while in Paralingua the number of pauses produced by the interlocutors was similar even though the total number of pauses was similar in both corpora. This, however, also requires further analyses as the proportion of the total speaking time between the dialogue participants should be taken into account.

Further research will include the analysis of alignment of timing-related phenomena in shorter stretches of speech between major changes of the conversation stages or topic shifts, fluency breakdowns and other phenomena that may significantly influence (hinder) the process of alignment between speakers.

Acknowledgements

Recording of the Paralingua database was supported from the financial resources for science in the years 2010–2012 as a development project (O R00 0170 12). DiaGest2 corpus comes from DiaGest2 project (N N104 010337, 2009-2010).

5. References

[1] Giles, H., Taylor, D. M. and Bourhis, R. "Towards a theory of interpersonal accommodation through language: some canadian data", Language in Society, pp. 177–192, 2010.

[2] Garrod, S., & Pickering, M., Why is conversation so easy? Trends in Cognitive Sciences, 8, pp. 8 – 11, 2004.

[3] Street, R. L. “Speech convergence and evaluation in fact-finding interviews. Human Communication Research”, Vol. 11, No. 2, pp. 139–169, 1984.
[4] Pardo, J. “On phonetic convergence during conversational interaction”, Journal of the Acoustical Society of America, 119, (2382-2393), 2006.
[5] Kousidis, S., “A Study of Accomodation of Prosodic and Temporal Features in Spoken Dialogues in View of Speech Technology Applications”, Doctoral Thesis. Dublin Institute of Technology, 2010.
[6] Kim, M., Horton, W. S. & Bradlow, A. R., “Phonetic convergence in spontaneous conversations as a function of interlocutor language distance”, Laboratory Phonology, 2, 125–156, 2011.
[7] Włodarczak, M., Simko, J., Wagner, P., “Syllable boundary effect: Temporal entrainment in overlapped speech”, Proceedings of Speech Prosody 2012.
[8] Branigan, H., Pickering, M., & Cleland, A., Syntactic co-ordination in dialogue. Cognition, 75, pp.13 – 25, 2000.
[9] Parrill, F., & Kimbara, I., “Seeing and hearing double: The influence of mimicry in speech and gesture on observers”, Journal of Nonverbal Behavior, 30, pp. 157–166, 2006.
[10] Mol, L., Krahmer, E., Maes, A., & Swerts, M. “Adaptation in gesture: Converging hands or converging minds?”, Journal of Memory and Language, 66, pp. 249–264, 2012.
[11] Bergmann K, Kopp, S., “Gestural Alignment in Natural Dialogue”, Proceedings of the 34th Annual Conference of the Cognitive Science Society (CogSci 2012), Cooper RP, Peebles D, Miyake N (Eds); Austin, TX: Cognitive Science Society: pp. 1326–1331, 2012.
[12] Zwaan, R. A. & Radvansky, G. A., “Situation models in language comprehension and memory”, Psychological Bulletin, 123, 162–185, 1998.
[13] Pickering, M. J. & Garrod, S., “Toward a mechanistic psychology of dialogue”, Behavioral and Brain Sciences, vol. 27, pp. 169–226, 2004.
[14] Krauss, R. M., Pardo, J. S., “Is alignment always the result of automatic priming?”, Behavioral and Brain Sciences, 27(02), pp. 203–204, 2004.
[15] Porzel, R., Scheffler, A. & Malaka, R. “How entrainment increases dialogical efficiency”, Proceedings of Workshop on Effective Multimodal Dialogue Interfaces, Sydney, 2006.
[16] Pickering, M. J. & Garrod, S., “Alignment as the Basis for Successful Communication”, Research on Language and Computation, Volume 4, Issue 2–3, pp. 203 – 228, 2006.
[17] Reitter, D. & Moore, J. D., “Predicting success in dialogue”, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp. 808–815, 2007.
[18] Karpiński, M., “New Challenges in Psycholinguistics: Interactivity and Alignment in Interpersonal Communication”, Festschrift for Prof. Piotra Łobacz (in print).
[19] Sacks, H., Schegloff, E., Jefferson, G., A simplest systematics for the organization of turn-taking for conversation, Language, vol. 50, 4, 1974.
[20] Wilson, M. & Wilson, T. P. “An oscillator model of the timing of turn taking”, Psychonomic Bulletin and Review, vol. 12, no. 6, pp. 957–968, 2005.
[21] O’Dell, M., Nieminen,T., “Coupled Oscillator Model for Speech Timing Overview and Examples”, Nordic Prosody. Proceedings of the Xth Conference, 2009.
[22] O’Dell, M., Lennes, M., Werner, S., Nieminen, T., “Looking for Rhythms in Conversational Speech”, Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, pp. 1201—1204, 2007.
[23] Buder, E. H. & Eriksson, A., “Time-series analysis of conversational prosody for the identification of rhythmic units”, Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1–7. August, 1999.
[24] Edlund, J., Heldner, M. & J. Hirschberg, “Pause and gap length in face-to-face interaction”, Proceedings of Interspeech, pp. 2779–2782, 2009.
[25] Vaughan, B., “Prosodic synchrony in co-operative task-based dialogues: A measure of agreement and disagreement”, Proceedings of Interspeech, pp. 1865–1867, 2011.
[26] De Looze, C. and Rauzy, S., “Measuring speakers’ similarity in speech by means of prosodic cues: methods and potential”, Proceedings of Interspeech, pp. 1393–1396, 2011.
[27] Levitan, R., Hirschberg, J., “Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions”, In: Proceedings of Interspeech, Florence, Italy, pp. 3081–3084, 2011.
[28] Truong, K. P. & Heylen, D., “Measuring prosodic alignment in cooperative task-based conversations”, Proceedings of Interspeech, 9-13 September 2012.
[29] Manson, J.H., Bryant, G.A., Gervais, M.M., Kline, M. A., Convergence of speech rate in conversation predicts cooperation. Evolution and Human Behavior, 34, 6, pp. 419–426, 2013.
[30] Karpiński, M. & Jarmołowicz-Nowikow, E. “Prosodic and Gestural Features of Phrase-internal Disfluencies in Polish Spontaneous Utterances”, Proc. Speech Prosody, Chicago, 2010.
[31] Klessa, K., Wagner, A., Oleśkowicz-Popiel, M., Karpiński, M., “Paralingua – a new speech corpus for the studies of paralinguistic features”, Procedia – Social and Behavioral Science, vol. 95, pp. 48–58, 2013.
[32] Boersma, P. & Weenink, D., Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Available from http://www.praat.org/, accessed in November 2013.
[33] Klessa, K., Karpiński, M., Wagner, A., Annotation Pro – a new software tool for annotation of linguistic and paralinguistic features. Proceeding of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop, Aix en Provence, 2013.
[34] Kousidis, S., Dorran, D., Wang, Y., B. Vaughan, Cullen, C., Campbell, D., McDonnell, C. and E. Coyle, “Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues”, Proceedings of Interspeech, pp. 1692–1695, 2008.
[35] Grabe, E. & Low, E. L. “Durational variability in speech and the rhythm class hypothesis”, In C. Gussenhoven & N. Warner (Eds.), Laboratory Phonology 7, pp. 515–546, 2002.
[36] Asu, E.L., Nolan, F. "Estonian and English rhythm: a two-dimensional quantification based on syllables and feet." Proc. Speech Prosody, Dresden, Germany, 2006.
[37] Barry, W.J.; Andreeva, B.; Russo, M.; Dimitrova, S.; Kostadinova, T., "Do rhythm measures tell us anything about language type?" Proc. of the 15th ICPhS. Barcelona, 2693-2696, 2003.
[38] Deterding, D. The measurement of rhythm: a comparison of singapore and british english. Journal of Phonetics, 29:217–230, 2001.
[39] Yu, J., Gibbon, D., Klessa, K. “Computational annotation-mining of syllable durations in speech varieties”, to appear in Proc. Speech Prosody, Dublin, Ireland, 20-23 May 2014.
[40] Malisz, Z., “Speech rhythm variability in Polish and English: A study of interaction between rhythmic levels”, PhD Thesis, Faculty of English, Adam Mickiewicz University, 2013.
[41] Gibbon, D., “TGA: a web tool for Time Group Analysis”. Proc.Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop, Aix en Provence, 2013.
[42] Klessa, K. and Gibbon, D. “Annotation Pro + TGA: automation of speech timing analysis”. To appear in Proceedings of Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26-31 May 2014.