Transitions, pauses and overlaps: Temporal characteristics of turn-taking in Czech

Lenka Weingartová, Eliška Churaňová, Pavel Šturm

Institute of Phonetics, Charles University in Prague, Czech Republic

lenka.weingartova@ff.cuni.cz

Abstract

This study aims to describe temporal characteristics of pausing and turn-taking phenomena in conversation. The material comes from the VASST corpus of contemporary Czech and uses four spontaneous dialogues in the form of an informal interview. We describe both general and idiosyncratic effects found in our data and compare them with results from other languages. In our material, transitions with a silent gap, overlaps and back-channels all display notably similar temporal distributions with the median around 360 ms and a marked skewing. The four dialogues did not differ in the proportion of turns belonging to the interviewer (58 %) vs. interviewee (42 %), which is hypothesized to characterize the experimental task. Despite a number of general tendencies, individual differences in pausing and turn-taking behaviour of the speakers were found as well. For instance, the ratio of pauses and gap transitions proved to be highly dialogue-specific. We also gathered evidence for a substantial change in the speech behaviour of the interviewer resulting from a change of her communication partner.

Index Terms: conversation, turn-taking, transition, overlap, back-channelling, pause

1. Introduction

The structure of conversation and turn-taking has been systematically investigated since the 1970s. A pioneering study in this regard is that of Harvey Sacks and his colleagues [1], who introduced an early model attempting to describe and explain the organization of turn-taking in natural conversation. A key assumption is that the structure of syntactic and semantic units should allow the listener to anticipate the end of the speaker's turn and to take over at a *transition-relevance place* (TRP). In addition to these syntactic cues, however, prosodic features (intonation contour, final lengthening, loudness) also play an important role [2], [3], [4].

The model in [1] takes into account several further assumptions that are supposed to describe any modifications in the organization of conversation. One of these is the temporal principle of "minimizing gap and overlap", i.e. of making the transition of speakers as smooth as possible. Several studies (e.g. [5], [6]) challenged this principle, providing a new set of data that yielded the most frequent pause duration at speaker transitions approximating 200 milliseconds, while gaps or overlaps of less than 10 ms were scarce. Thus the close succession of turns remains to be further investigated [6].

The distribution of *gaps* (pauses at turn transitions) and *overlaps* (intervals with overlapping speech at transitions) has been widely investigated in English and a few other languages. In the material of [6] overlaps appeared in approximately 40 % of all turn transitions. The importance of overlaps was also emphasized by ten Bosch et al. [7] and Shriberg et al. [8]. The latter, investigating multi-participant dialogues, found overlaps in 17 % of all words, and in 54 % of intonation units without a pause. Similarly, in an analysis of 26 meetings featuring several participants, overlaps took up 12 % of all speaking time [9]. These findings imply that overlaps represent a relevant feature in the organization of conversation, and it is thus necessary to include them in any model of turn-taking.

Naturally, pauses (not only) at turn transitions have been the focus of much research. Discontinuities in speech may be classified in several ways based, for instance, on their function in conversation (pause, gap, lapse; [1], [6]), their form (silent vs. filled pauses, different forms of hesitations; [10], [11], [12]), their relation to syntax (boundary vs. hesitation pauses; [13]), their duration [14], [15] or on the speaker's intention to continue or pass the word [11], [16]. This also influences the used terminology which is not unified. The issue of what constitutes a pause is also debated. Usually, the authors determine a minimum cut-off boundary in the range of 100-150 milliseconds [4], [17], [18], [19], [20], but other decisions are possible, e.g. [13], [14], [21]. The detection threshold for a pause in turn transitions was determined to be 120 milliseconds [22]. Nevertheless, it is important to keep in mind that decisions on both pause type and minimum pause duration always depend on the purpose of the analysis.

One of the central concerns of this study is the relation between turn transitions and the duration of pauses. This was investigated for instance by Wennerstrom and Siegel [4]. Their results suggest that the probability of a turn transition decreases within the first 500 milliseconds of the pause, and then increases for longer pause durations (approximately 1500 milliseconds). The probability of giving the floor to another speaker is thus highest for short and long pauses, while lowest for pauses of middle duration – which, in other words, tend to occur within a single turn.

Both duration and occurrence of pauses demonstrate a great amount of variability in different languages, speech styles and speech tempos. Studies also dealt with a strong influence of individual habits on the duration of pauses [23], [24]. However, it was ascertained that pauses are connected to major syntactic breaks and coincide with prosodic boundaries of intonational phrases [18], [25]. The duration of the pause has been shown to correlate with the strength of prosodic boundaries [26] and the length of the phrase (e.g. [27]).

We may hypothesize that the organization of conversation is to a certain degree universal and independent of the particular language. To our knowledge, none of the above mentioned experiments have been replicated in Czech. The present paper therefore aims to compare the findings from well-investigated languages with our data on Czech, but also to enlarge the scope of interest and explore idiosyncratic patterns of speakers and accommodation towards the dialogue partner.

10.21437/SpeechProsody.2014-87

2. Method

2.1. VASST corpus

The speech material was taken from the VASST corpus (the acronym meaning *group and style variation* in Czech, see [28]), which is currently being built at the Institute of Phonetics in Prague. The aim of the corpus is to capture the variability of contemporary spoken Czech in different styles and sociolinguistic groups. To this day, the corpus comprises 168 speakers from eight different regions and three specific social groups. Each subject provided five different speaking styles (ordered by formality): a read list of sentences, read continuous text, picture description, controlled interview and spontaneous dialogue.

For this study, the last part of the corpus – spontaneous dialogue – was used. It should be noted that the level of spontaneity of the dialogues is subject to discussion and depends on several factors (familiarity with the interviewer, character of the subject, position within the dialogue, etc.). The fact of being recorded and interviewed presented a rather unnatural situation for the subjects, however it was done in a familiar environment (at the participants' homes) and the experimenters were instructed to make the subjects as comfortable as possible. It could be stated that the recordings we selected from the corpus show a high degree of spontaneity.

2.2. Speech material

Four dialogues were analyzed with a total duration of 85 minutes (on average 21 minutes per speaker pair). The speakers were female, aged 75 to 85, who all came from the same region in Northern Bohemia and spoke colloquial Czech. The experimenter (i.e. their dialogue partner) was in all cases the same, a female student of Phonetics.

The utterances were recorded on a portable professional device Edirol HR-09, with a sampling frequency of 48 kHz and a 16-bit quantization. Afterwards, the recordings were downsampled to 32 kHz and manually post-processed and labelled in Praat [29] by experienced phoneticians (including two of the authors). The boundaries of breath-groups were marked for both dialogue partners, as well as pauses and turn transitions. A breath-group was defined as a stretch of speech of one speaker between his two breath intakes. A turn is the stretch of speech (consisting of one or more breath-groups) of one speaker uninterrupted by the second (with the exclusion of back-channels, see below). Another solution was adopted by [30] and [31] whose main analysis unit was the interpausal stretch. However, this information is obtainable from our annotation as well.

Pauses within speakers' turns were categorized as follows:

- *silent pauses* (Ps): unfilled pause
- *hesitations* (Ph): pause containing a hesitation sound
- *breath pauses* (Pb): pause containing breath intake

The minimum duration of a pause was set to 120 milliseconds; shorter pauses were treated as part of segmental articulation. Pauses that partake in speaker change (i.e. gaps) were not included in this category, see below.

For labelling the turn-taking phenomena, a classification based on [5] was used:

- *gap transitions* (Tg): speakers switch their turns following a gap
- *overlaps* (To): speakers switch by overlapping each other's speech
- *back-channels* (Tbc): intervals of short overlap not resulting in speaker change; often consisting of a single sound or word

It was shown that this classification can cover as much as 96 % of all turn transition phenomena [5].

In the next step, the duration of turns, pauses and turn transitions was measured. Since the duration of transitions or pauses cannot be expected to be normally distributed (the values are only positive and the number of short and long durations is bound to be extremely different), medians instead of arithmetic means are reported. To assess the significance of the results, non-parametrical statistic tests, i.e. Mann-Whitney U test and Kruskal-Wallis one-way ANOVA, were used.

3. Results

3.1. Transitions and overlaps

Gap transitions (Tg) represent the most frequent type of turn-taking in our material. They were over two times more

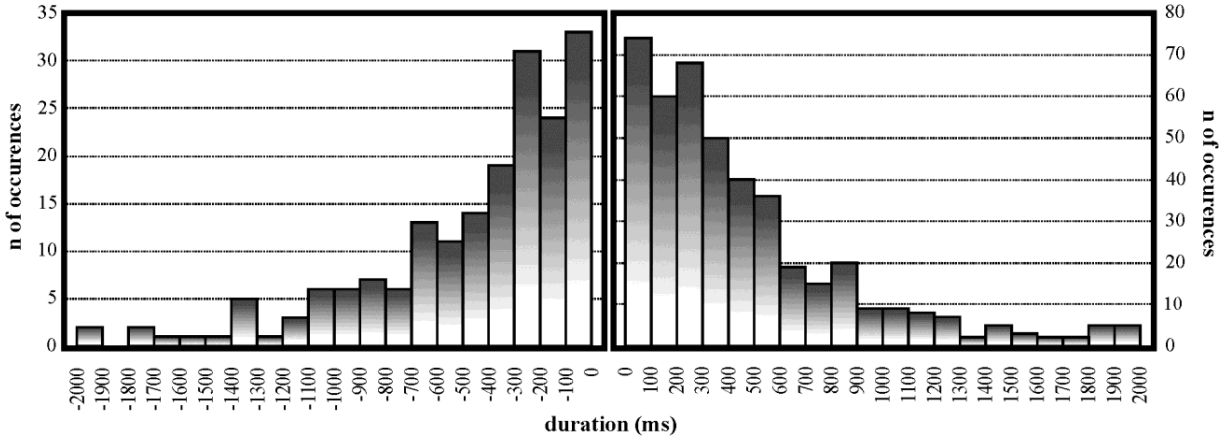

Figure 1: *A histogram of the durations of* OVERLAPS *(on the left) and* GAP TRANSITIONS *(on the right).*

frequent than overlaps (To), and this tendency holds true for individual dialogues as well, although in one of them the ratio amounted almost to 3:1. However, we found a discrepancy between the subjects and the experimenter: the Tg/To ratio was only 1.4 for the subjects, while it was as high as 4.6 for the experimenter (who was instructed not to interrupt the subjects if possible). In terms of percentage, gap transitions constitute 52 % of all turn-taking phenomena, whereas overlaps comprise 22 %.

Figure 1 shows the durational distribution of gap transitions and overlaps. Although both phenomena are inherently different in nature, they display strikingly similar temporal characteristics. The medians for Tg and To are 333 ms and 353 ms, respectively, and both distributions are massively skewed. 57 % of Tg's and 57.5 % of To's are shorter than 400 ms, while 89 % of Tg's and 88 % of To's are shorter than 1 s.

If we pool together gap transitions and overlaps from all speakers, the distributions of Tg × To show no significant difference (Mann-Whitney U test: $p = 0.83$). However, if we divide the transitions between the speakers and assign each to the speaker that follows, we discover that the experimenter and one of the subjects are differentiated in their realization of Tg's and To's – the subject in Dialogue 3 had significantly longer overlaps than gap transitions (Mann-Whitney: $p < 0.05$), while the experimenter had longer Tg's than To's (Mann-Whitney: $p < 0.01$). More importantly, if we compare the duration of overlaps of the subjects with those of the experimenter, the latter are significantly shorter (Mann-Whitney: $p < 0.05$).

Turn transitions in close succession, i.e. those realized with a gap/overlap not exceeding 10 ms (investigated for instance by [6]), constitute 1.7 % of all transitions in our material. According to [22], this boundary can be raised to 120 ms, which is the detection threshold for a no-gap, no-overlap transition in a dialogue. In this respect, smooth or close transitions were realized in 18.5 % of all cases.

3.2. Back-channelling

As expected given the nature of the interview, the back channels (Tbc) were more frequent (1.8 times) with the experimenter than the subjects. Back-channelling can be used as a supportive affirmation and encouragement on part of the experimenter in order to induce subjects to continue speaking. However, there were intra-speaker differences in the experimenter's back-channelling behaviour (see below, section 3.4). The number of back-channels on the part of subjects was too low to permit quantitative analysis.

Out of all turn-taking phenomena, back-channels constitute 26 %, so they occur more often than overlaps (22 %). Interestingly enough, the durational properties of back-channelling are not significantly different from gap transitions or overlaps (Kruskal-Wallis ANOVA: $p > 0.05$). The median duration of back-channels was 384 ms.

3.3. Pauses

The overall duration of within-turn pauses constitutes 16 % of the total duration of the speech material. Since the number of experimenter's pauses was low (due to briefness of her turns), in the following text only pauses of the subjects are reported.

In individual dialogues, breath pauses (Pb) constitute 9-14 % of the duration, while silent pauses (Ps) up to 5 % and hesitation pauses (Ph) only up to 3 % of the subjects' speaking time. Since hesitation pauses were infrequent, differed greatly in manifestation and were problematic in terms of identifying and labelling, we decided to exclude them from further analyses.

Breath pauses were significantly longer than silent pauses (Mann-Whitney: $p < 0.001$), which was the case for all subjects. For three of the subjects the median duration of a breath pause ranged between 425 and 470 ms, while one (D2) was significantly different with 572 ms (Kruskal-Wallis ANOVA: $H(3, n = 894) = 23.2$; $p < 0.001$). Interestingly, the same speaker had the lowest median duration of silent pauses (274 ms), while the other three speakers clustered between 370 and 401 ms.

Working under the paradigm of [1], pauses can be considered *transition relevant places* (TRPs) where the change of speakers does not occur (similarly to [30]). It would therefore be interesting to compare the duration of breath and silent pauses to gap transitions, where the change of speakers takes place.

We discovered a significant difference between the duration of breath pauses and gap transitions – breath pauses were considerably longer (Mann-Whitney: $p < 0.001$). This holds true for all dialogues but one (D1). The case of silent pauses and gap transitions was less clear, as in two dialogues (D1, D2) Tg's were longer than Ps's, whereas in D3 and D4 it was the other way round.

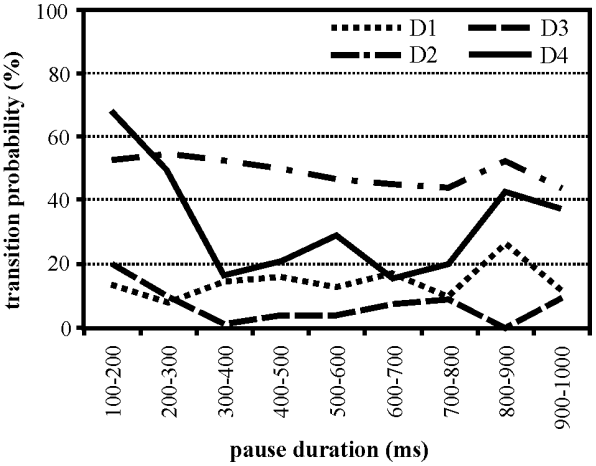

Figure 2: *The probability of turn transition as a function of pause duration. D1-D4 represent individual dialogues.*

We also computed transition probability by determining the percentage of gap transitions in all "silences" (Tg, Ps and Pb). Unlike [4], we did not find any drop in the percentage around 500 ms, nor an increase over 1500 ms. Instead, the percentage curve drops to about 20 % around 300 ms and stays at that level. Figure 2 shows the percentage in steps of 100 ms according to individual dialogues.

It is remarkable that between 300 and 1000 ms (over 1000 ms the number of cases starts to be too low) the percentage of gap transitions stays quite similar and highly dialogue-specific. For D1, gap transitions comprise around 15 % of all silent intervals, D2 has the highest percentage (48 % on average), D3 with only 5 % of Tg's represents the other extreme, and D4 fluctuates around 26 %.

3.4. The experimenter

Since the experimenter was the same in all dialogues, it presents an interesting opportunity to examine changes in her speech behaviour associated with change of dialogue partner.

The frequency of her turns paralleled the speaking behaviour of the subject, resulting in a substantial change in the average values from 1.7 turns per minute (D3) up to 10.5 turns per minute (D2). The duration of turns of the subject and the experimenter also seemed to be in a direct relationship, but the low number of dialogues prevented statistical verification.

Regardless of turn duration or frequency, it is remarkable to see that the experimenter uttered in all cases 58 % of the turns (the exact range was 58.1 to 58.8), leaving the remaining 42 % to the individual subjects. This could be the consequence of the task and of her role as interviewer.

Concerning the experimenter's transition behaviour, it can be seen from the data that her overlap strategy varied. Although in all four dialogues she produces approximately 1/3 of the overlaps, their duration differs significantly (Kruskal-Wallis ANOVA: $H(3, n = 57) = 8.5$; $p < 0.05$). Moreover, the duration and frequency of her back-channels differed as well. While in D3 the back-channels appear only 0.7 times per minute, in D2 it increases to more than three times per minute. This is in inverse relation to the length of the Tbc's – in D3 the back-channels were longest (median 593 ms), in D2 they were shortest (median 322 ms).

4. Discussion

Identical instructions to all participants resulted in a number of similar tendencies in the conversational structure of the dialogues, but several individual differences were also found.

Despite the fact that gap transitions, overlaps and back-channels are phenomena of a different nature and they manifest differently in the flow of speech, their temporal characteristics show remarkable similarities in our speech material.

The most frequent strategy for changing turns seems to be a transition with a silent gap between individual turns. Overlaps occured in our dialogues less often than reported in [6], where they constituted around 40 % of all transitions. We detected 30 % of overlaps; 22 % if back-channelling is also included as a transition phenomenon.

Concerning the findings of [5] and [6], who point out that the principle of minimizing gap and overlap postulated first by [1] may not be the main cause governing transition duration, our results show that although transitions without perceptible gap (that is ±120 ms) do not constitute the majority of transitions (only 18.5 %), the frequency of both gap transitions and overlaps generally does increase as their duration approaches zero.

If we compare the frequency of short transitions (without a perceptible gap), longer transitions and longer overlaps with the findings of [22: 511], our data are well in range of what the author found for 12 different languages. In our data, no-gap, no-overlap transitions constituted 18.5 % of all transitions, longer gap transitions 57.1 %, and longer overlaps 24.3 %. The author also noticed remarkable similarities in the ratio of these categories within language families – but in his material, Slavonic languages were not represented, so more research would be needed to see whether this hypothesis is valid also for them.

Durational characteristics of pauses and their relation to the turn-taking phenomena were also investigated. Gap transitions were significantly shorter than breath pauses. This may suggest that often the communication partner takes the turn before the other finishes breathing in. Despite breath pauses being physiologically conditioned, there also seem to be some idiosyncrasy in their duration. One of our speakers had a significantly longer duration of breath pauses and, interestingly enough, a significantly shorter duration of silent pauses at the same time. It is possible that this represents some kind of unconscious compensation on the part of the speaker. That the duration of pauses may be speaker-dependent and a useful tool for speaker identification has already been pointed out not only by forensic phoneticians (e.g. [11], [32]).

On the other hand, we could not replicate the findings of [4], who observed a marked drop in the probability of transition in pauses around 500 ms of length. We detected no such change in transition probability; however, this probability displayed remarkable and very stable inter-dialogue differences (shown in Figure 2).

In our experimental design, the experimenter received instructions concerning her conversational behaviour. The data suggest that she indeed performed her role as an interviewer – she uttered more turns and of shorter duration than the subjects (around 58 % in each dialogue). Furthermore, she took turns significantly less often by overlapping her communication partner; and, if she did, the overlaps were significantly shorter.

The conspicuous similarity of experimenter's turn percentage in all dialogues could be attributed to the experimental task. In future research we therefore plan to compare it with the controlled interview, which is the other conversational task from the VASST corpus, to see whether the percentage of turns changes. It is also possible to be a characteristics of the experimenter – again, we will examine the same dialogue task performed by other experimenters to verify this hypothesis.

Despite her consistencies mentioned above, the experimenter shows notable changes in her behaviour depending on the dialogue partner. This concerns mainly the duration of overlaps and the duration and frequency of back-channels. The notion of accommodation or entrainment may be evoked in this regard (see e.g. [7]), and these relationships should be investigated further.

The VASST corpus of contemporary Czech with its different speaking styles offers an excellent opportunity to research conversation behaviour in contrast to other speaking styles. In future studies the results will be verified on a larger amount of speech data, and the scope of the research should be enlarged to cover other prosodic phenomena in turn-taking, as well as to discover further individual patterns of speakers' behaviour.

5. Acknowledgements

The support of the Programme of Scientific Areas Development at Charles University in Prague (PRVOUK), subsection 10 – Linguistics: Social Group Variation is acknowledged. The first author was funded by a grant of the Czech Science Foundation (GACR 406/12/0298). The authors would like to thank all the students who participated in the recording and post-processing of the material. Special thanks also to Jan Volín for inspiring the paper and for his helpful insights and comments.

6. References

[1] Sacks, H., Schegloff, E. and Jefferson, G., "A simplest systematics for the organization of turn-taking for conversation", Language, 50(4): 696–735, 1974.

[2] Duncan, S., "Some signals and rules for taking speaking turns in conversations", Journal of Personality and Social Psychology, 23(1): 283–292, 1972.

[3] Ford, C. and Thompson, S., "Interactional units in conversation: Syntactic, intonational and pragmatic resources for the management of turns", in E. Ochs, E. Schegloff and S. Thompson [Eds.], Interaction and grammar, 134–184, Cambridge: Cambridge University Press: 1996.

[4] Wennerstrom, A. and Siegel, A. F., "Keeping the floor in multiparty conversations: Intonation, syntax and pause", Discourse Processes, 36: 77–107, 2003.

[5] Weilhammer, K. and Rabold, S., "Durational aspects in turn taking", Proceedings of the International Conference of Phonetic Sciences 2003, Barcelona, Spain, 2003.

[6] Heldner, M. and Edlund, J., "Pauses, gaps and overlaps in conversation", Journal of Phonetics 38: 555–568, 2010.

[7] ten Bosch, L., Oostdijk, N. and Boves, L., "On temporal aspects of turn taking in conversational dialogues", Speech Communication, 47: 80–86, 2005.

[8] Shriberg, E., Stolcke, A. and Baron, D., "Observations on overlap: Findings and implications for automatic processing of multi-party conversation", Proceedings of Eurospeech, vol. 2, Aalborg, Denmark: 1359–1362, 2001.

[9] Çetin, Ö. and Shriberg, E., "Analysis of overlaps in meetings by dialog factors, hot spots, speakers and collection site: Insights for automatic speech recognition", Proc. ICSLP, Pittsburgh: 293–296, 2006.

[10] Maclay, H. and Osgood, C. E., "Hesitation phenomena in spontaneous English speech", Word, 15: 1944. 1959.

[11] van Donzel, M. E. and Koopmans-van Beinum, F. J., "Pausing strategies in discourse in Dutch", Proceedings ICSLP '96, Philadelphia, USA, vol. 2: 1029–1032, 1996.

[12] Rose, R. L., "Crosslinguistic Corpus of Hesitation Phenomena: A corpus for investigating first and second language speech performance", Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013), Lyon, France: 992–996, 2013.

[13] Boomer, D. S. and Dittmann, A. T., "Hesitation pauses and juncture pauses in speech. Language and Speech", 5: 215–220, 1962.

[14] Goldman–Eisler, F., "Pauses, clauses, sentences", Language and Speech, 15: 103–113, 1972.

[15] Campione, I. and Véronis, J., "A large-scale multilingual study of silent pause duration", Proceedings of Eurospeech 2002: 199–202, 2002.

[16] Local, J. and Kelly, J., "Projection and "silences": Notes on phonetic and conversational structure", Human Studies, 9: 185–204, 1986.

[17] Dankovičová, J., "The minimum pause duration in spontaneous speech", PROPH – Progress Reports from Oxford Phonetics 5: 17–24, 1992.

[18] Butcher, A., "Aspects of the speech pause: Phonetic correlates and communicative functions", Aipuk (Arbeitsberichte Institut für Phonetik Kiel): 15, 1981.

[19] Hieke, A., Kowal, S. and O'Connell, M., "The trouble with "articulatory" pauses", Language and Speech, 26(3): 203–214, 1983.

[20] Hansson, P., "Prosodic phrasing and articulation rate variation", Proceedings of Fonetik, TMH-QPSR, 44: 173–176, 2002.

[21] Goldman-Eisler, F., "Psycholinguistics: Experiments in spontaneous speech", New York: Academic Press, 1968.

[22] Heldner, M., "Detection thresholds for gaps, overlaps and no-gap-no-overlaps", Journal of the Acoustical Society of America 130(1): 508–513, 2011.

[23] Goldman-Eisler, F., "The distribution of pause durations in speech", Language and Speech, 4: 232–237, 1961.

[24] Ruder, K. F. and Jensen, P. J., "Fluent and hesitation pauses as a function of syntactic complexity", Journal of Speech and Hearing Research, 15: 49–58, 1972.

[25] Ferreira, F., "Effects of length and syntactic complexity on initiation times for prepared utterances", Journal of Memory and Language, 30: 210–233, 1991.

[26] Zellner, B., "Pauses and the temporal structure of speech", in E. Keller [Ed.], Fundamentals of speech synthesis and speech recognition, 41–62, Chichester: John Wiley, 1994.

[27] Zvonik, E. and Cummins, F., "The effect of surrounding phrase lengths on pause duration", Proceedings of Eurospeech 2003, Geneva, Switzerland: 777–780, 2003.

[28] Volín, J. and Weingartová, L., "Současný stav zkoumání zvukové stránky mluvních stylů", in preparation.

[29] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer" [Computer program], version 5.3.41, retrieved from http://www.praat.org/, 2013.

[30] Caspers J., "Local speech melody as a limitating factor in the turn-taking system in Dutch", Journal of Phonetics, 31: 251–276, 2003.

[31] Gravano A. and Hirschberg, J., "Turn-taking in task-oriented dialogues", Computer Speech and Language, 25: 601–634, 2011.

[32] Foulkes, P. and French, J. P., "Forensic speaker comparison: a linguistic-acoustic perspective", in P. Tiersma and L. Solan [Eds.], Oxford Handbook of Language and Law, 557–572, Oxford: Oxford University Press, 2012.