Temporal aspects of breathing and turn-taking in Swedish multiparty conversations

Jonna Hammarsten[1], Roxanne Harris[1], Nilla Henriksson[1], Isabelle Pano[1], Mattias Heldner[2] and Marcin Włodarczak[2]
[1] CLINTEC, Division of Speech and Language Pathology, Karolinska Institutet, Stockholm
[2] Department of Linguistics, Stockholm University, Stockholm

Abstract

Interlocutors use various signals to make conversations flow smoothly. Recent research has shown that respiration is one of the signals used to indicate the intention to start speaking. In this study, we investigate whether inhalation duration and speech onset delay within one's own turn differ from when a new turn is initiated. Respiratory activity was recorded in two three-party conversations using Respiratory Inductance Plethysmography. Inhalations were categorised depending on whether they coincided with within-speaker silences or with between-speaker silences. Results showed that within-turn inhalation durations were shorter than inhalations preceding new turns. Similarly, speech onset delays were shorter within turns than before new turns. Both these results suggest that speakers 'speed up' preparation for speech inside turns, probably to indicate that they intend to continue.

Introduction

In a conversation people exchange roles from speaker to listener on a regular basis. In order for turn-taking to proceed smoothly people use various cues – both conscious and unconscious ones (Rochet-Capellan et al., 2014). By means of different verbal and nonverbal cues, conversation partners can show the intention to take, hold or release the turn. These cues include, among other things, syntax, prosody and communicative silences (Local & Kelly, 1986) as well as head and body movements (Hadar et al., 1983, 1984).

From a physiological point of view, the foundation of speech production lies within the respiratory patterns of inhalations and exhalations. A great majority of speech sounds are formed when air is forced from the lungs via the glottis, through the oral and nasal cavities, where they become audible (Ohala, 1990). A typical breathing pattern during speaking consists of short and fast inhalations, followed by extended exhalations (Hixon et al., 1973).

Even though breathing has been mentioned as a potentially important turn-taking cue, few empirical studies have explored this area. However, according to Ishii et al. (2014), speakers' inhalations can be effective signals for predicting whether the turn will be released or kept. Ishii et al. (2014) also found that listeners' inhalations can predict attempts to take the turn. Similarly, results in Rochet-Capellan et al. (2014) suggest that most successful turns are initiated right after a new inhalation, and that breathing cycles at the start of a turn are more symmetric than inside a turn. Furthermore, Rochet-Capellan et al. (2014) found that speech onset delay (the interval between exhalation onset and speech onset) was kept to a minimum in successfully taken turns.

In line with the findings of Ishii et al. (2014) and Rochet-Capellan et al. (2014), the current study explores whether speakers show the intention to hold the turn using inhalation duration and speech onset delay as cues. We hypothesise that both inhalation durations and speech onset delays will be shorter within turns than when new turns are initiated.

Method

Participants

Participants were recruited by email sent to students at Stockholm University and KTH, Royal Institute of Technology. In this study two conversations were chosen for analysis. In one of the conversations, the speakers were a brother, aged 24, a sister, aged 28 and their father, aged 67. The other conversation included three students who did not know each other; a

female aged 23, a female aged 24 and a male aged 27. All participants were native Swedish speakers. The participants were not fully aware of the purpose of the study or the details of interest.

Procedure

The participants were instructed to converse freely in Swedish for about 20 minutes. The recordings were made in a sound treated room in the Phonetics Laboratory at Stockholm University. They were standing in an upright position around a table of 95 cm height. This position was preferred to minimize disruption of the breathing signal. The sound was recorded with close-talking directional microphones, (Sennheiser HSP 4). Respiratory Inductance Plethysmography (Watson, 1980) was used to measure respiratory activity. Each participant wore two elastic belts – one around the ribcage at armpit level and one around the abdomen, at the navel (Figure 1). Coils embedded in each belt are sensitive to changes in cross-sectional area due to breathing. The belts were connected to a RespTrack, a respiratory belt processor designed and built at Stockholm University (Edlund et al., 2014).

The respiratory signal from each belt was weighted and summed by means of a potentiometer, also part of RespTrack. The resulting signal (Figure 2) was captured by an integrated physiological data acquisition system (PowerLab by ADInstruments).

Figure 1. Students in the Phonetics Laboratory wearing transducer belts, each connected to a RespTrack, during a recording session.

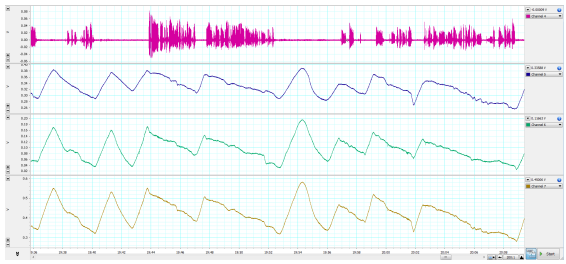

Figure 2. Respiratory signals shown as one wave for each transducer belt together with their summed wave below. The speech signal at the top.

The following parameters were estimated before the recording for each speaker: vital capacity (maximum exhalation produced after maximum inhalation), isovolume manoeuver (net lung volume change) (Konno & Mead, 1967), and resting expiratory level (measured after a relaxed sigh).

Annotation and feature calculation

Stretches of speech and silence were detected automatically in ELAN (Wittenburg et al., 2006) based on the intensity threshold for a silent portion of the signal, and then corrected manually. Next, inhalations and exhalations in the respiratory cycles were labelled automatically using an algorithm described in Włodarczak & Heldner (in press). This was also followed by manual correction in Praat (Boersma & Weenink, 2015).

Silent portions were classified as within-speaker silences (WSSs) or between-speaker silences depending on whether the same speaker or a new speaker continued the conversation after the silence (Jaffe & Feldstein, 1970).

Inhalation duration was calculated from the manually corrected respiratory annotations. Speech onset delay was calculated from the offset of the inhalation to the onset of speech. Cases were speech onset preceded inhalation offset were excluded. Overlaps as well as stretches of speech shorter than 1 second, which correspond primarily to short feedback expressions (Heldner et al., 2011) were not included in the analysis.

The final analysed sample contained 68 WSSs and 138 BSSs. Data analysis was done in SPSS, version 22.0 (IBM Corp, 2012).

Results

Comparison of mean parameter values in the two interval types shows that inhalation in

WSSs are shorter by 0.345 s (cf. Table 1), indicating that speakers inhale more quickly when they intend to hold the turn. A mean difference was also observed with respect to speech onset delay, whose duration was shorter by 0.356 s in WSSs than in BSSs (cf. Table 2). Thus, speakers tend to start speaking more quickly after the inhalation offset when holding the turn.

Histograms of inhalation duration and speech delay showed the distributions in the sample were positively skewed (cf. left panels in Figures 3 & 4). In order to remove the skew, the values were transformed using logarithms with a base of 10 (cf. right panels of Figures 3 & 4). The variations in log-transformed inhalation duration and speech onset delay of WSSs and BSSs are summarized in boxplots in Figure 5.

The log-transformed distributions satisfied the preconditions of an independent parametric t-test. With degrees of freedom corrected due to unequal variances between conditions (see Tables 1 and 2), significant differences between WSS and BSS categories were obtained both for inhalation duration ($t_{117.003} = 6.432$, $p < 0.0001$) and speech onset delay ($t_{202.883} = 3.428$, $p < 0.001$). The t-test thus revealed significant differences between WSSs and BSSs for both conditions in the current sample.

Table 1. Mean values and standard deviation of inhalation duration (in seconds) for Within Speaker Silences (WSSs) and Between Speaker Silences (BSSs).

Interval type	Mean	Std. Dev.
WSS	0.540	0.272
BSS	0.885	0.445

Table 2. Mean values and standard deviation of speech onset delay (in seconds) for Within Speaker Silences (WSSs) and Between Speaker Silences (BSSs).

Interval type	Mean	Std. Dev.
WSS	0.116	0.088
BSS	0.472	0.976

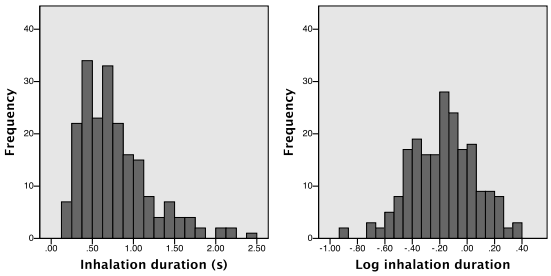

Figure 3. Histograms of inhalation durations (s) in raw values (left panel) and after log-transformation (right panel).

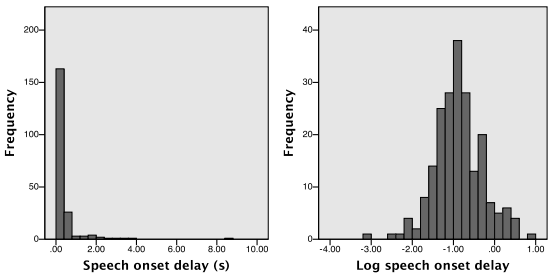

Figure 4. Histograms of speech onset delays (s) in raw values (left panel) and after log-transformation (right panel).

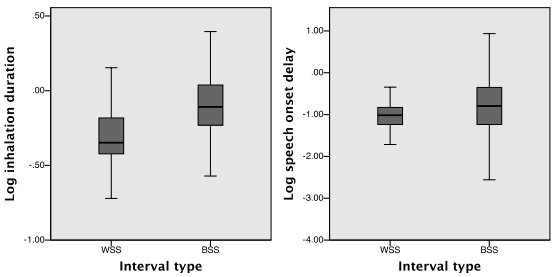

Figure 5. Boxplots of log-transformed inhalation duration (left panel) and speech onset delay (right panel) for the two interval types: Within Speaker Silences (WSSs) and Between Speaker Silence (BSSs).

Discussion

The aim of the study was to investigate whether inhalation duration and speech onset delay differs between turn-holding and turn-taking. In line with our hypothesis and earlier findings by Rochet-Capellan et al. (2014) and Ishii et al. (2014), the results showed significant differences between the conditions with lower values for both inhalation and speech onset delay within a turn compared to when a new turn was initiated.

The differences observed in inhalation duration were expected since the speaker holding the turn might want to keep his/her

inhalation to a minimum in order not to lose the turn. In a similar vein, Rochet-Capellan et al. (2014) suggested that speakers will adapt their breathing in order to hold the turn. Specifically, more rapid inhalations within a turn will reduce pauses and increase speaker's chances to continue without interruption.

Differences observed in speech onset delay were also expected, and for a similar reason. Specifically, the speaker might want to start speaking faster upon exhalation onset while holding the turn to reduce pause duration and counteract a possible interruption attempt from a dialogue partner.

Since the results are based on a small sample, containing in total 6 participants and 40 minutes worth of conversation, they need to be treated as preliminary. The environment of the phonetics laboratory in which the study took place could have affected conversational flow. Nevertheless, the material used in the present study was significantly more natural and spontaneous than the dialogues used by Rochet-Capellan et al. (2014) and Ishii et al. (2014). In spite of less tight experimental control similar results were obtained, indicating that the current setup allows for studying respiration in spontaneous multiparty interactions.

Despite the significance of the presented results, they can only be applied to the current sample. Future studies within the field of breathing and turn-taking are needed. They should include a larger amount of speakers as well as a larger number of conversations analysed, to ensure generalizability of the results and to the lay foundation for technological advances, such as dialogue systems and synthetic speech.

Acknowledgements

This work was funded in part by the Swedish Research Council project 2014-1072 *Andning i samtal* (*Breathing in conversation*).

References

Boersma P and Weenink D (2015). Praat: doing phonetics by computer [Computer program] (Version 5.3.84). Retrieved from http://www.praat.org/

Edlund J, Heldner M and Włodarczak M (2014). Catching wind of multiparty conversation. In: J Edlund, D Heylen & P Paggio, eds, *Proceedings of Multimodal Corpora: Combining applied and basic research targets (MMC 2014)*. Reykjavik, Iceland.

Hadar U, Steiner T, Grant E C and Rose F C (1983). Head movement correlates of juncture and stress at scentence level. *Language and Speech* 26, 117-129.

Hadar U, Steiner T, Grant E C and Rose F C (1984). The timing of shifts of head postures during conversation. *Human Movement Science* 3, 237-245.

Heldner M, Edlund J, Hjalmarsson A and Laskowski K (2011). Very short utterances and timing in turn-taking. In *Proceedings Interspeech 2011*. Florence, Italy, 2837-2840.

Hixon T J, Goldman M D and Mead J (1973). Kinematics of the chest wall during speech production: Volume displacement of the rib cage, abdomen, and lung. *Journal of Speech, Language and Hearing Research* 16, 78-115.

IBM Corp. (2012). SPSS Statistics for Macintosh [Computer program] (Version 22.0). Armonk, NY: IBM Corp.

Ishii R, Otsuka K, Kumano S and Yamato J (2014). Analysis of respiration for prediction of "who will be next speaker and when?" in multi-party meetings. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, 18-25.

Jaffe J and Feldstein S (1970). *Rhythms of dialogue*. New York, NY, USA: Academic Press.

Konno K and Mead J (1967). Measurement of the separate volume changes in the rib cage and abdomen during breathing. *Journal of Applied Physiology* 22, 407-422.

Local J K and Kelly J (1986). Projection and 'silences': Notes on phonetic and conversational structure. *Human Studies* 9, 185-204.

Ohala J J (1990). Respiratory activity in speech. In: W J Hardcastle & A Marchal, eds, *Speech Production and Speech Modelling*: Springer, 23-53.

Rochet-Capellan A, Bailly G and Fuchs S (2014). Is breathing sensitive to the communication partner? In: N Campbell, D Gibbon & D Hirst, eds, *Proceedings of Speech Prosody 2015*. Dublin, Ireland: Trinity College, 613-617.

Watson H (1980). The technology of respiratory inductive plethysmography. In: F D Stott, E B Raftery & L Goulding, eds, *Proceeding of the Second International Syposium on Ambulatory Monitoring (ISAM 1979)*. London: Academic Press.

Wittenburg P, Brugman H, Russel A, Klassmann A and Sloetjes H (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* 1556-1559.

Włodarczak M and Heldner M (in press). Respiratory properties of backchannels in spontaneous multiparty conversation. In *Proceedings ICPhS 2015*. Glasgow, UK.