Inhalation amplitude and turn-taking in spontaneous Estonian conversations

Kätlin Aare, Marcin Włodarczak and Mattias Heldner
Department of Linguistics, Stockholm University, Stockholm

Abstract

This study explores the relationship between inhalation amplitude and turn management in four approximately 20 minute long spontaneous multiparty conversations in Estonian. The main focus of interest is whether inhalation amplitude is greater before turn onset than in the following inhalations within the same speaking turn.

The results show that inhalations directly before turn onset are greater in amplitude than those later in the turn. The difference seems to be realized by ending the inhalation at a greater lung volume value, whereas the initial lung volume before inhalation onset remains roughly the same across a single turn. The findings suggest that the increased inhalation amplitude could function as a cue for claiming the conversational floor.

Introduction

Previous research has shown that speech planning is reflected in respiratory patterns, at least in read speech. For instance, the duration and amplitude of inhalation have been found to correlate positively with the upcoming utterance length in several studies (e.g. Winkworth, Davis, Adams, Ellis, 1995; Fuchs, Petrone, Kirvokapić, Hoole, 2013). Also the location of inhalation is strongly determined by speech planning. Almost all inhalations in read speech occur at major constituent boundaries, such as paragraphs, sentences or phrases (Conrad, Thalacker, Schönle, 1983; Grosjean and Collins, 1979).

By contrast, breathing in spontaneous speech shows a less consistent pattern. It has been claimed that as many as 13% of all inhalations in spontaneous monologues occur at grammatically inappropriate locations (Wang, Green, Nip, Kent, Kent, 2010), possibly due to the additional demands of real-time speech planning. The effect should be even more pronounced in spontaneous conversation, where the communicative demands are different.

A key characteristic of the conversational rhythm is its oscillating pattern – generally, one speaker at a time has the speaking turn and longer stretches of simultaneous speech tend to be avoided. Thus, the exchange of speaker and listener roles needs to be precisely coordinated by means of turn-taking cues indicating the intention to take, hold or release the turn (McFarland, 2001). Breathing patterns have previously been hypothesized to be part of the turn-taking system. Inhalations have been claimed to be an interactionally salient cue to speech initiation (Schegloff, 1996) and to be deeper before turn initiation (Ishii, Otsuka, Kumano, Yamato, 2014). Finally, breath holding and exhalation have been suggested as turn keeping and turn-yielding devices, respectively (French and Local, 1983).

Furthermore, durational properties of respiration have been shown to reflect turn-taking intentions. Speakers tend to minimise pause durations inside the turn by inhaling more quickly, and by reducing the delay between inhalation offset and speech onset (Rochet-Capellan and Fuchs, 2014; Hammarsten, Harris, Henriksson, Pano, Heldner, Włodarczak, this volume). As inhalation duration and depth have been found to correlate, especially in read speech (e.g. Rochet-Capellan and Fuchs, 2013), the amplitude of non-initial inhalations in a speaking turn should also be smaller.

Therefore, in addition to being governed by the demands of phrasing and speech planning, breathing patterns in spontaneous conversation may depend, at least partly, on speaker's communicative goals linked to claiming and keeping the turn. In this study, we investigate whether turns consisting of several breathing cycles show a pattern where the turn-initiating inhalation amplitude is greater than the amplitude of the following inhalations within the turn.

Method

Data acquisition was carried out by recording respiratory activity synchronized with audio and video in spontaneous three-party conversations each lasting approximately 22 minutes. Participants were 20- to 35-year-old (with the mean of 26) healthy native speakers of Estonian with an average Body Mass Index of 21.9. The speakers did not report any history of speech, language, hearing or respiratory disorders, and had never been smokers. They were invited to travel to Stockholm and take part in the recording sessions by personal communication. The speakers in each recording session had known each other for a period from a few weeks to 14 years. With the exception of two couples living together, the speakers described their relationships as friends.

The recordings took place in the Phonetics Laboratory at Stockholm University. The participants had no knowledge of the exact aim of the experiment prior to the recording and they were free to choose the topics of conversations. They were instructed to wear tight-fitting clothes to minimise distortions in the respiratory signals.

Respiratory activity was measured with Respiratory Inductance Plethysmography (Watson, 1980), which quantifies changes in the rib cage and abdomen cross-sectional areas by means of two elastic transducer belts (Ambu RIP-mate) placed at the level of the armpits and the navel. The overall lung volume change was estimated by isovolume manoeuvre (Konno and Mead, 1967). A more detailed setup description is provided in Edlund, Heldner and Włodarczak (2014).

The respiratory signal was recorded using PowerLab (ADInstruments, 2014). Audio was captured using head-worn microphones with a cardioid polar pattern (Sennheiser HSP 4). The speakers were asked to stand around a 1-meter-high table, each facing a GoPRO Hero 3+ camera recording the upper part of the torso, and to avoid large movements.

Annotation of the data was carried out semi-automatically using Praat (Boersma and Weenink, 2015) and Python scripts (Buschmeier and Włodarczak, 2013). The sum of the rib cage and abdomen signals was used to segment the breathing signal into periods of inhalations and exhalations (for details see: Włodarczak and Heldner, in press). A total of 11.9% of the automatically assigned borders were either moved or added manually due to inaccuracy of the automatic annotation. In addition, following Jaffe and Feldstein (1970), silences and overlaps were classified depending on whether they coincided with speaker change or were followed by more speech from the same speaker. Accordingly, the speaking turn has been defined as an uninterrupted series of speech segments from a single speaker. As backchannels, coughing and laughter are commonly considered to be non-interruptive, they were not classified as claiming a turn. This is particularly true of backchannels, which are unplanned and produced by the listener to give short feedback to the speaker (see e.g. Heldner, Hjalmarsson, Edlund, 2013; Aare, Włodarczak, Heldner, 2014). Consequently, only uninterrupted speaking turns that included multiple breath groups were analysed. The amount of data was limited further by excluding some speech stretches that coincided partly with inhalations.

Lung volume levels were normalised with respect to speaker's minimum and maximum lung volumes measured at the calibration stage of the recording. These values correspond to vital capacity (VC), the maximum volume of air exhaled after a maximum inhalation (Hixon, 2006). The final part of analysis was carried out with R (R Core Team, 2015).

Results

Data distribution

Due to the filtration procedure, 50 suitable speaking turns were left for analysis. These turns consisted of 128 breath groups produced by 11 speakers. Table 1 illustrates the distribution of the number of breath groups and inhalations for one speech turn. All multi-breath-group turns included at least two inhalation phases but rarely more. Therefore, to ensure sufficient sample sizes, the results below are based on the data from the first two inhalations in each of the 50 turns.

Table 1. Inhalations per speaking turn.

Number of inhalations	Frequency
2	31
3	10
≥ 4	9
Σ	50

Inhalation start and end levels

The initial and final lung volume levels for the first two inhalations in a turn, normalised to speaker's vital capacity (%VC), are shown in Figure 1.
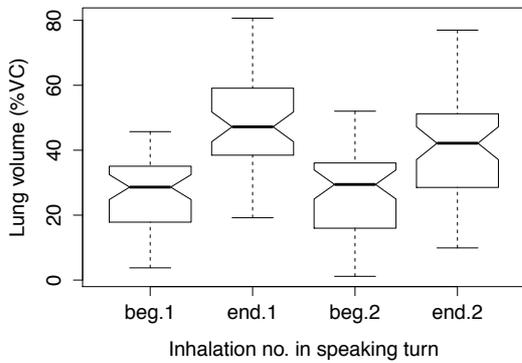

Figure 1. A comparison of the start and end lung volume levels for the first two inhalations in a turn. All data is normalised to speaker's vital capacity. "1" marks the first inhalation of a turn, occurring before turn onset, and "2" marks the second inhalation, located inside the speaking turn; "beg" and "end" indicate the point of measurement – the beginning or end of the corresponding inhalation.

Figure 1 shows that while the initial lung volumes in the first and second inhalation are practically identical (except for a slightly larger range in the latter), the end values differ more considerably. Specifically, the second inhalation's end value is lower in median and both range end points, and is characterised by larger range. Furthermore, the difference between the first and second inhalation's end lung volume means is statistically significant $(t(98) = 2.262, p = .026)$.

A more detailed relationship of the start and end lung volume levels for both inhalations can be seen on Figure 2. The figure shows a marked positive relationship (correlation: $r(48) = .657$, $p = .000$) between the inhalation start and end lung volume levels for both inhalations.

Inhalation amplitude

Inhalation amplitudes (also normalised with respect to speakers' vital capacity) for turn-initial and turn-medial inhalations are presented in Figure 3. The overall inhalation amplitude falls into the range between 26 and 45% of VC.
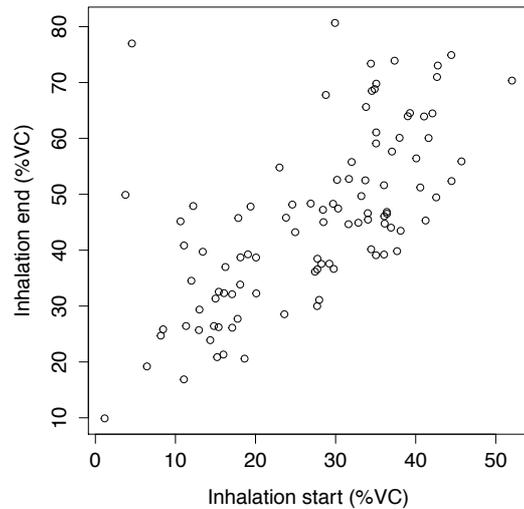

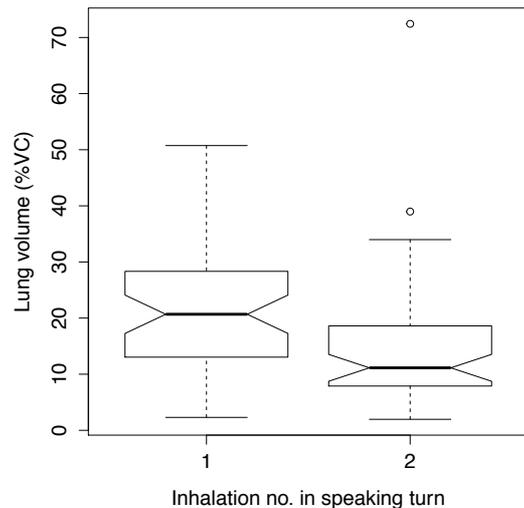*Figure 2. A scatterplot of the normalized inhalation start and end lung volume levels (%VC).*

Figure 3. Inhalation amplitude for the first and second inhalation in a turn.

As can be seen, the first and second inhalation in a turn have different amplitude medians and ranges. The turn-initial inhalation exhibits higher and more symmetrically distributed amplitude values with values ranging to around 50% of vital capacity and a median of 20% VC. The following inhalation has a considerably lower median and a smaller range. The difference between mean amplitude values is also significant $(t(98) = 2.854, p = .005)$.

Discussion

The comparison of turn initial and turn medial inhalations in a single turn shows that the turn-initial inhalation is significantly greater in amplitude. As can be seen in Figure 1, speakers tend to initiate inhalations at approximately the same level in the vital capacity, but inhale more deeply before taking the turn.

This might imply that a deeper inhalation cues the intention to take the turn. As the larger amplitude is accomplished by reaching a higher lung volume, the cue could just as well be the end lung volume level, rather than the amplitude itself. In other words, the actual change in amplitude may be less important than the end level within the vital capacity. We leave this hypothesis for future research.

We may also speculate that a deeper inhalation projects a longer turn, possibly spanning several breath cycles. Simply put, by signalling an intention to produce a longer contribution, speakers might minimise competition for the conversational floor at turn internal pauses. At the same time, more shallow inhalations at turn medial pauses could serve as a phrasing device to indicate how the speaker intended to structure the message, thereby facilitating listener's understanding.

There are a number of additional factors that could provide further insight into the interplay between turn-taking and respiration. For instance, shorter inhalations might require higher airflow, which in turn increases the likelihood of fricative noise. This audible inhalation in itself might also function as a turn-holding or phrasing device. We plan to address this issue in subsequent studies.

Conclusions

This exploratory study focused on the physical constraints governing speech breathing in connection with turn-taking in spontaneous multiparty conversations among native speakers of Estonian. The examination of lung volumes in speech turns spanning several breath cycles indicates that turn initial inhalations are deeper than inhalations later in the turn. This is accomplished by inhaling to a higher lung volume during the first inhalation, with the pre-inhalatory lung volumes remaining relatively stable across consecutive inbreaths.

The use of respiratory cues for the organization of turn-taking is not a new discovery, but this work contributes novel findings regarding lung volumes. Although the limited size of the data set restricts the possible conclusions, the patterns discovered in this study indicate that inhalation depth is sensitive to speakers' intention to start or continue speaking. The results thus provide another evidence in favour of breathing serving as a potentially important turn-taking cue in spontaneous conversation.

Acknowledgements

References

Aare K., Włodarczak M. and Heldner M. (2014). Backchannels and breathing. In: M. Heldner (Ed.), *Proceedings of FONETIK 2014.* Stockholm, Sweden, 47-52.

ADInstruments. (2014). LabChart software and PowerLab hardware (Version 8). New South Wales, Australia: ADInstruments

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

Buschmeier H. and Włodarczak M. (2013). TextGridTools: A TextGrid Processing and Analysis Toolkit for Python. In P. Wagner (Ed.), *Tagungsband der 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV 2013),* 152-157. Dresden: TUDpress.

Conrad B., Thalacker S. and Schönle P. (1983). Speech respiration as an indicator of integrative contextual processing. *Folia Phoniatrica et Logopaedica* 35, 220-225.

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

French P. and Local J. (1983). Turn-competitive incomings. *Journal of Pragmatics* 7, 17-38.

Fuchs S., Petrone C., Krivokapić J. and Hoole P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics* 41, 29-47.

Grosjean F. and Collins M. (1979). Breathing, pausing and reading. *Phonetica* 36(2), 98-114.

Hammarsten J., Harris R., Henriksson N., Pano I., Heldner M. and Włodarczak M. (this volume). Temporal aspects of breathing and turn-taking in Swedish multiparty conversations.

Heldner M., Hjalmarsson A. and Edlund J. (2013). Backchannel relevance spaces. In E.L. Asu and P. Lippus (Eds.), *Nordic Prosody: Proceedings of the*

XIth Conference, Tartu 2012, 137-146. Frankfurt am Main, Germany: Peter Lang.

Hixon T.J. (2006). *Respiratory function in singing: A primer for singers and singing teachers*. Tucson, Arizona: Redington Brown.

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 of rib cage and abdomen during breathing. *Journal of Applied Physiology*, 22(3), 407–422.

McFarland D.H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research* 44, 128-143.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria: http://www.R-project.org/.

Rochet-Capellan A. and Fuchs S. (2013). The interplay of linguistic structure and breathing in German spontaneous speech. In *Proceedings of Interspeech*, 1128–1132.

Rochet-Capellan A. and Fuchs S. (2014). Take a breath and take the turn: how breathing meets turns in spontaneous dialogue. In R. Smith, T. Rathcke, F. Cummins, K. Overy and S. Scott (Eds.), *Philosophical Transactions of the Royal Society B: Biological Sciences,* 369(1658), 1-10.

Schegloff E.A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs, E.A. Schegloff and S.A. Thompson (Eds.), *Interaction and Grammar*, 52-133. Cambridge: Cambridge University Press.

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

Winkworth A.L., Davis P.J., Adams R.D. and Ellis E. (1995). Breathing patterns during spontaneous speech. *Journal of Speech & Hearing Research* 38, 124-144.

Wang Y.T., Green J.R., Nip I.S., Kent R.D. and Kent J.F. (2010). Breath group analysis for reading and spontaneous speech in healthy adults. *Folia Phoniatrica et Logopaedica* 62 (6), 297-302.

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