Breathing in Conversation: an Unwritten History

Marcin Włodarczak, Mattias Heldner
Department of Linguistics
Stockholm University
Stockholm, Sweden
{wlodarczak,heldner}@ling.su.se

Jens Edlund
Speech, Music and Hearing
KTH Royal Institute of Techonology
Stockholm, Sweden
edlund@speech.kth.se

Abstract

This paper attempts to draw attention of the multimodal communication research community to what we consider a long overdue topic, namely respiratory activity in conversation. We submit that a turn towards spontaneous interaction is a natural extension of the recent interest in speech breathing, and is likely to offer valuable insights into mechanisms underlying organisation of interaction and collaborative human action in general, as well as to make advancement in existing speech technology applications. Particular focus is placed on the role of breathing as a perceptually and interactionally salient turn-taking cue. We also present the recording setup developed in the Phonetics Laboratory at Stockholm University with the aim of studying communicative functions of physiological and audio-visual breathing correlates in spontaneous multiparty interactions.

1 Introduction

Human face-to-face communication is known to be inherently multimodal. Specifically, multimodal features have been demonstrated to be closely linked to such basic mechanisms of interaction as turn-taking, grounding and interpersonal coordination. In addition, they have also proved useful in developing dialogue systems and computational models of interaction.

At the same time, while some multimodal cues (gaze, manual gestures, head movements, body posture) have received much attention, others remain as yet unexplored, despite their great potential in highlighting important aspects of human-human and human-computer interaction. In this paper we address one such feature. Namely, we argue that studying breathing in conversation is crucial for understanding how speech production is employed in the coordinated and highly context-sensitive domain of conversation, and call for more research in the field. In particular, in the light of perceptual salience of speech breathing suggested by earlier studies (Whalen et al., 1995; Whalen and Sheffert, 1996), we focus on the role of kinematic and audio-visual correlates of respiration in coordination of speaker change in spontaneous conversation.

In the remainder of this paper we briefly discuss earlier research on speech breathing (Section 2) as well as its possible extensions to the domain of spontaneous conversation (Section 3). Subsequently, in Section 4 we describe our newly established respiratory lab at the Department of Linguistics, Stockholm University.

2 Historical look

Breathing is a primary mechanism of voice generation maintaining a suitable level of subglottal pressure required for momentary production needs. As such, it is implicated in many aspects of speech production, such as voice quality (Slifka, 2006), voice onset time (Hoit et al., 1993) and loudness (Huber et al., 2005). Similarly, breathing has been claimed to enter into processes of speech planning and structuring

K. Jokinen and M. Vels. 2015. Proceedings of The 2nd European and the 5th Nordic Symposium on Multimodal Communication.

(Fuchs et al., 2013). However, in line with the methodological stance dominant in traditional phonetics, breathing has been studied almost exclusively in tightly controlled experiments decoupled from communicative context. Consequently, while these and other studies have made important contributions to speech science, they have largely ignored interactive factors at play in conversation, the most common language use.

At the same time, certain findings stirred by the recent wave of interest in speech respiration indicate that breathing plays an important interactional role. For instance, McFarland (2001) observed that speakers synchronise their respiratory cycles prior to speaker change. It was subsequently shown that the synchronisation is brought about by performing a shared task (Bailly et al., 2013) and is therefore similar to other known examples of interspeaker coordination (Shockley et al., 2009). Indeed, there is some evidence than breathing is linked to synchronisation of speech and gesture (Hayashi et al., 2005) and might even be the basis for synchronisation of movement in general (Pellegrini and Ciceri, 2012).

In addition, the listener's breathing cycle was reported to change depending on such properties of perceived speech as tempo or vocal effort (Rochet-Capellan and Fuchs, 2013). While there is considerable controversy as to the exact nature of the underlying alignment mechanism (or mechanisms), it suggests that breathing is implicated in processes of speech perception. Similarly, on the production side, a variety of kinematic adjustments were found depending on where speech was initiated within the respiratory cycle (McFarland and Smith, 1992), thus indicating sensitivity of the respiratory apparatus to the demands of an upcoming vocal task. Clearly, these mechanisms could be also exploited for conversational needs, for instance to coordinate speaker change.

Last but not least, respiratory data have been demonstrated to improve performance of speech and language technology applications. In particular, including breathing noises in synthetic speech enhances its naturalness (Braunschweiler and Chen, 2013) and recall (Whalen et al., 1995). Improvements in performance were also noted for automatic speech recognition (Butzberger et al., 1992) and automatic annotation of prosody (Wightman and Ostendorf, 1994). Finally, respiratory data were successfully used to detect conversational episodes by automatic discrimination between periods of quiet breathing, listening and speaking (Rahman et al., 2011).

3 Conversational perspectives

In spite of the interactional salience of breathing suggested by the work outlined above, studies of breathing in spontaneous conversation are strikingly rare. Conversation analysis has presented some evidence of how audible inspirations and expirations are used as turn-taking and turn-yielding cues, and how breath holds function as a turn-holding device (Schegloff, 1996; Local and Kelly, 1986). However, these findings have so far not been backed up by a comprehensive quantitative analysis of conversational corpora. Moreover, earlier attempts at quantifying breathing in interaction were based on material which was often not entirely spontaneous (McFarland, 2001; Winkworth et al., 1995). Two notable exception is are recent studies by Rochet-Capellan and Fuchs (2014) and Ishii et al. (2014), which measured breathing patterns during pauses coinciding with speaker change or followed by more speech from the previous speaker.

We argue that breathing in dialogue is a potentially fruitful line of research likely to highlight fundamental principles underlying interspeaker coordination and collaborative human action. Respiratory data could be particularly instructive for investigating mechanisms of turn management. Specifically, as turns are normally preceded by easily perceivable inhalations and followed by equally salient exhalations, audio-visual correlates of respiratory events could be an important extension of the set of the more familiar multimodal turn-taking cues. In addition, respiratory data should allow detecting "hidden events" otherwise not easily available for analysis, e.g. abandoned speech initiation attempts (sharp audible inhalations not followed by speech), thus offering more direct access to speakers' intention to initiate or terminate a turn. Similarly, adaptations of the respiratory cycle prior to speaker change, whose preliminary account was presented by McFarland (2001), could shed new light on the long-standing question of mechanisms behind the observed distributions of gaps and overlaps. Importantly, as breathing is by its very nature an embodied activity, it is also likely to provide a valuable insight into interdepenen-

Figure 1: Data acquisition system: PowerLab alongside an audio interface (left) and a RespTrack belt processor (right).

cies between physical and communicative constraints operating in dialogue, for instance the relationship between momentary lung volume and kinematic adaptations prior to speech initiation similar to those found by McFarland and Smith (1992) but set in the fully interactive domain of conversation and subject to temporal constraints of the turn-taking system. Lastly, the links between breathing and other modalities implied by cross-modal synchronisation reported in literature should inform models of sensorimotor coordination both within and between individuals.

In addition to their theoretical significance, studies of respiratory activity in conversation should also help solve some of the key problems in speech and language technology. In particular, loud inhalations might facilitate inferring speaker's intention to initiate a turn and, consequently, provide a shallow, signal-based solution to detecting user barge-ins before their actual onset. Similarly, presence of audible exhalations and breath holds could be used to reason about turn completeness and avoid pause interruptions, which are common in dialogue managers using pause duration as the only turn-yielding cue.

4 Stockholm University Respiratory Lab

In order to answer the questions related to interactional functions of breathing discussed in the previous section, we have developed the following recording setup in the Phonetics Laboratory at Stockholm University. The core of the design is a respiratory inductance plethysmograph (Watson, 1980), which consists of two elastic transducer belts (Ambu RIPmate) measuring changes in cross-sectional area of the rib cage and the abdomen due to breathing. Before each recording, the belts are calibrated using isovolume manoeuvres (Konno and Mead, 1967), which allow estimating contributions of individual belts to the total lung volume change. In addition, vital capacity and resting expiratory levels are also recorded for reference. In order to minimise noise in the signal produced by body movement, participants are recorded standing at a table (about 90 cm high). As the range of respiratory patterns is likely to be sensitive to complexity of turn negotiation and the degree of dialogue competitiveness, we base our studies on multiparty dialogues between three communicative partners.

The belts are connected to dedicated RespTrack processors developed in the Phonetics Lab (see the right panel of Figure 1). The processors were designed for ease of use, and optimised for low noise recordings of respiratory movements in speech and singing. In particular, DC offset can be corrected simultaneously for the rib cage and abdomen belts using a "zero" button. Unlike in the processors supplied with the belts, there is no high-pass filter, thus the amplitude will not decay during breath-holding. A potentiometer allows the signals from the rib cage and abdomen belts to be weighted so that they give the same output for a given volume of air, as well as for the summed signal, enabling direct estimation of lung volume change (see Figure 2).

The signal is recorded by a data acquisition system (PowerLab 16/35 by ADInstruments, left panel

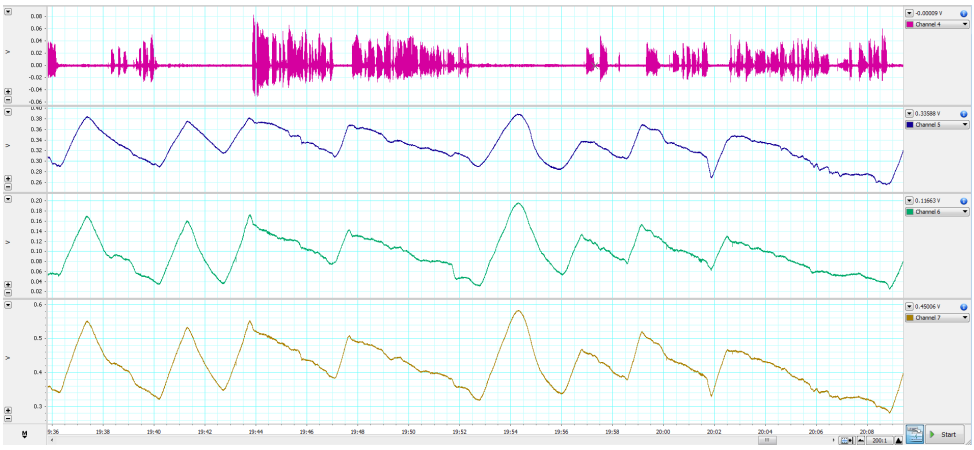

Figure 2: Sample recording for a single speaker: speech (channel 1), respiratory signal from the rib cage and abdomen belts (channels 2 and 3) and the summed respiratory signal (channel 4).

Figure 3: Recording setup. The white boxes are earlier prototypes of the RespTrack processors.

of Figure 1). The system is essentially an analogue-to-digital converter which synchronises the inputs and works with dedicated recording and analysis software (LabChart by ADInstruments). Notably, the system allows connecting other measuring devices, such as airflow masks, which are potentially useful for calibrating the belts. A sample signal is shown in Figure 2.

The setup can be easily adapted to specific recording conditions. For instance, making field recordings is possible by replacing our lab-based data acquisition system with a portable USB-powered unit (DLP-IO8-G Data Acquisition Board by DLP Design). Given the low cost of such devices, they could be also useful for educational purposes, such as student projects.

High quality audio is recorded by close talking microphones (Sennheiser HSP 4) connected to an audio interface (PreSonus AudioBox 1818). The signal is additionally routed to PowerLab to ensure synchronisation with the respiratory trace. As breathing is not only audible but also visible, GoPro Hero3+ cameras are used to record the video.

Our present setup is shown in Figure 3. We are currently conducting a series of pilot studies related to respiratory turn-taking cues as well as temporal patterns of speech initiation within the respiratory cycle. Preliminary results were presented in Aare et al. (2014).

Given that we are particularly interested in communicative functions of audible inhalations and exhalations, we are experimenting with alternative methods of recording clear respiratory noises. Two variants

are being assessed: one in which a dedicated close-talking microphone is placed directly in front of the mouth and one which uses a contact microphone placed on the neck near the larynx (throat microphone). A further extension of the recording setup consists in using thermistor probes placed in speakers' nostrils, which should allow differentiating between breathing through the nose and through the mouth.

The resulting corpus will be segmented into (semi-)automatically derived stretches of speech and silence in the audio signal, and inhalations and exhalations in the respiratory signal. In addition, selected dialogue act categories (interruptions, backchannels, disfluencies) will be annotated. The data set will be made public for research use.

5 Conclusions

This paper has aimed at pointing out potential interest and relevance of respiratory activity to fundamental mechanisms of conversation related to turn management. We have argued that the topic has been long overlooked in breathing research and is ripe for systematic quantitative investigation, especially in the light of the existing evidence of multifaceted interactions between breathing and speech production and perception as well as its possible applications in speech technology. We have also described a recording setup developed at Stockholm University required for such a data collection and analysis effort. We hope to see respiratory activity taking its legitimate place among other better studied multimodal features in the nearest future.

Acknowledgements

The research presented here was funded in part by the Swedish Research Council project 2014-1072 *Andning i samtal* (*Breathing in conversation*).

References

Kätlin Aare, Marcin Włodarczak, and Mattias Heldner. 2014. Backchannels and breathing. In *Proceedings of FONETIK 2014*, pages 47–52, Stockholm, Sweden.

Gérard Bailly, Amélie Rochet-Capellan, and Coriandre Vilain. 2013. Adaptation of respiratory patterns in collaborative reading. In *Proceedings of Interspeech 2013*, pages 1653–1657, Lyon, France.

Norbert Braunschweiler and Langzhou Chen. 2013. Automatic detection of inhalation breath pauses for improved pause modelling in HMM-TTS. In *Proceedings of the 8th ISCA Speech Synthesis Workshop*, pages 1–6, Barcelona, Spain.

John Butzberger, Hy Murveit, Elizabeth Shriberg, and Patti Price. 1992. Spontaneous speech effects in large vocabulary speech recognition applications. In *Proceedings of the workshop on Speech and Natural Language*, pages 339–343. Association for Computational Linguistics.

Susanne Fuchs, Caterina Petrone, Jelena Krivokapić, and Philip Hoole. 2013. Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, 41(1):29–47.

Koji Hayashi, Nobuhiro Furuyama, and Hiroki Takase. 2005. Intra-and inter-personal coordination of speech, gesture and breathing movements. *Transactions of the Japanese Society for Artificial Intelligence*, 20(3):247–258.

Jeannette D. Hoit, Nancy Pearl Solomon, and Thomas J. Hixon. 1993. Effect of lung volume on voice onset time (VOT). *Journal of Speech, Language and Hearing Research*, 36(3):516–521.

Jessica E. Huber, Bharath Chandrasekaran, and John J. Wolstencroft. 2005. Changes to respiratory mechanisms during speech as a result of different cues to increase loudness. *Journal of Applied Physiology*, 98(6):2177–2184.

Ryo Ishii, Kazuhiro Otsuka, Shiro Kumano, and Junji Yamato. 2014. Analysis of respiration for prediction of "who will be next speaker and when?" in multi-party meetings. In *Proceedings of the 16^{t}h ACM International Conference on Multimodal Interaction (ICMI 2014)*, pages 18–25, Istambul, Turkey.

Kimio Konno and Jere Mead. 1967. Measurement of the separate volume changes of rib cage and abdomen during breathing. *Journal of Applied Physiology*, 22(3):407–422.

John Local and John Kelly. 1986. Projection and 'silences': Notes on phonetic and conversational structure. *Human studies*, 9(2):185–204.

David H McFarland and Anne Smith. 1992. Effects of vocal task and respiratory phase on prephonatory chest wall movements. *Journal of Speech and Hearing Research*, 35(5):971–982.

David H. McFarland. 2001. Respiratory markers of conversational interaction. *Journal of Speech, Language and Hearing Research*, 44(1):128–143.

Raffaella Pellegrini and Maria Rita Ciceri. 2012. Listening to and mimicking respiration: Understanding and synchronizing joint actions. *Review of Psychology*, 19(1):17–27.

Md. Mahbubur Rahman, Amin Ahsan Ali, Kurt Plarre, Mustafa al'Absi, Emre Ertin, and Santosh Kumar. 2011. mConverse: Inferring conversation episodes from respiratory measurements collected in the field. In *Proceedings of the 2nd Conference on Wireless Health*, pages 1–10, San Diego, CA.

Amélie Rochet-Capellan and Susanne Fuchs. 2013. Changes in breathing while listening to read speech: the effect of reader and speech mode. *Frontiers in Psychology*, 4(906):1–15.

Amélie Rochet-Capellan and Susanne Fuchs. 2014. Take a breath and take the turn: How breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society B*, 369(1658):1–10.

Emanuel A. Schegloff. 1996. Turn organization: One intersection of grammar and interaction. *Studies in Interactional Sociolinguistics*, 13:52–133.

Kevin Shockley, Daniel C. Richardson, and Rick Dale. 2009. Conversation and coordinative structures. *Topics in Cognitive Science*, 1(2):305–319.

Janet Slifka. 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice*, 20(2):171–186.

H. Watson. 1980. The technology of respiratory inductive plethysmography. In F. D. Stott, E. B. Raftery, and L. Goulding, editors, *Proceeding of the Second International Symposium on Ambulatory Monitoring (ISAM 1979)*, pages 537–563, London. Academic Press.

Doug H. Whalen and Sonya M. Sheffert. 1996. Perceptual use of vowel and speaker information in breath sounds. In H. Timothy Bunnell and William Idsardi, editors, *Proceedings of ICSLP 96*, pages 2494–2497.

Doug H. Whalen, Charles E. Hoequist, and Sonya M. Sheffert. 1995. The effects of breath sounds on the perception of synthetic speech. *The Journal of the Acoustical Society of America*, 97:3147–3153.

Colin W. Wightman and Mari Ostendorf. 1994. Automatic labeling of prosodic patterns. *IEEE Transactions on Speech and Audio Processing*, 2(4):469–481.

Alison L. Winkworth, Pamela J. Davis, Roger D. Adams, and Elizabeth Ellis. 1995. Breathing patterns during spontaneous speech. *Journal of Speech, Language and Hearing Research*, 38(1):124–144.