Catching wind of multiparty conversation

Jens Edlund[*]**, Mattias Heldner**[†]**, Marcin Włodarczak**[†]

[*]KTH Speech, Music and Hearing, [†]Department of Linguistics, Stockholm University, Sweden
edlund@speech.kth.se, {heldner, wlodarczak}@ling.su.se

Abstract

The paper describes the design of a novel corpus of respiratory activity in spontaneous multiparty face-to-face conversations in Swedish. The corpus is collected with the primary goal of investigating the role of breathing for interactive control of interaction. Physiological correlates of breathing are captured by means of respiratory belts, which measure changes in cross sectional area of the rib cage and the abdomen. Additionally, auditory and visual cues of breathing are recorded in parallel to the actual conversations. The corpus allows studying respiratory mechanisms underlying organisation of spontaneous communication, especially in connection with turn management. As such, it is a valuable resource both for fundamental research and speech techonology applications.

Keywords: breathing, multiparty conversation, turn-taking, respiratory inductance plethysmography, physiological measurements

1. Introduction

Even though we may not be aware of it, much breathing in dialogue is both clearly audible and visible. Consequently, it has been suggested that respiration is used in coordination of dialogue flow (Schegloff, 1996; Local and Kelly, 1986), e.g. by indicating intention to take or release a speaking turn. As a result, breathing is likely to provide a more direct access to speaker's communicative intentions than is otherwise available. However, few studies addressed interactional aspects of breathing. While notable exceptions exist, for instance (McFarland, 2001; Winkworth et al., 1995), even those studies were based on interactions which were not entirely spontaneous. In addition, no account exists of breathing in dialogue between more than two speakers, which is likely to show a greater range of respiratory patterns due to increased turn management complexity.

These omissions are particularly glaring given the potential relevance of breathing to speech technology application. As dialogue turns are normally preceded by deep and easily perceivable inhalations and followed by marked exhalations, presence of breathing noises could be used to improve turn management strategies implemented in the state-of-the-art dialogue systems. For instance, loud inhalations during system output could be used to detect user interruptions *prior to* the actual speech onset. Likewise, identification of post-completion exhalations should reduce the number of pause interruptions, which are a major problem in current speech technology applications.

Consequently, studying breathing in conversation is highly relevant from the point of view of both fundamental and applied research. On the one hand, it contributes significantly to the understanding of physiological constraints driving speech production and organisation of human interaction. On the other hand, it informs computational models of human interaction and paves the way towards more human-like embodied conversational agents capable of using previously unavailable cues.

Motivated by these goals, we have begun collection of a multimodal corpus of spontaneous multiparty conversations which includes physiological measurements relevant to breathing. Below we outline the recording setup and briefly discuss possible applications of the corpus.

2. Data acquisition setup

The recordings take place at the Phonetics Laboratory, Stockholm University in a quiet, sound-treated room. As it was observed that a standing position minimises noise in the respiratory signal due to body movement, subjects are recorded standing at a table 95 cm in height. No task is used; instead participants are asked to talk spontaneously on a topic of their choice. The recording setup is shown in Figure 1.

Respiratory activity is measured using using respiratory inductance plethysmography (Watson, 1980), which quantifies changes in rib cage and abdominal cross sectional area by means of two elastic transducer belts (Ambu RIPmate) placed at the level of the armpits and the navel, respectively. Contributions of individual belts to the net lung volume change are estimated using isovolume manoeuvres (Konno and Mead, 1967).

Figure 1: Recording setup. The white boxes are earlier prototypes of our respiratory belt processors.

Figure 2: The second prototype of the RespTrack processor.

The belts are connected to a dedicated respiratory belt processor (RespTrack, Figure 2) designed and built in the Phonetics laboratory at Stockholm University. The RespTrack processor was 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 for example 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 3).

The signal is collected by an integrated physiological data acquisition system (PowerLab by ADInstruments), which also allows connecting other measuring instruments, such as air-flow masks or electroglottographs. A sample signal is presented in Figure 3.

High-quality audio is recorded with close-talking directional microphones (Sennheiser HSP 4), and video is captured by GoPro Hero3+ cameras.

We plan to expand the setup by including contact microphones attached to speakers' necks (throat microphones) with a view to obtaining clearer recordings of inhalation and exhalation noises. Additionally, we will use thermal probes placed in the nostril to be able to distinguish nasal and mouth breathing. All these extensions are fully compatible with our current recording setup and will be presented during the workshop.
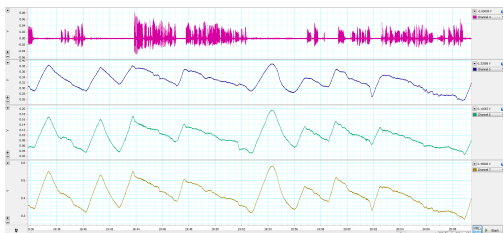

Figure 3: Speech recording (channel 1) and respiratory measurements from rib-cage and abdomen belts (channels 2-3) for one speaker. The bottom channel shows the weighted sum of the two belts.

Minimally, the corpus will be annotated with interactional events derived from voice activity detection, as well as (semi-)automatically detected inhalation and exhalation events in the respiratory data.

3.   Possible applications of the corpus

Our primary rationale for the corpus collection is studying the role of breathing in turn taking. Above all, it will allow a stringent quantitative investigation of previously untested claims made in literature, for instance about the role of inhalations as an interactionally salient cue to speech initiation, exhalations as a turn yielding device and breath holding as a marker of turn incompleteness. Furthermore, detection of pre-speech inhalations should allow to infer speaker's *intention* to initiate a new turn, whether or not this intention is realised or abandoned. Thus, respiratory data will also shed light on "hidden" events in dialogue, which are otherwise unavailable for analysis.

Furthermore, the corpus could serve as a test bed for computational models of turn-taking. In particular, the combination of physiological measurements with audio recordings of respiratory noises will provide valuable training data for automatic detection and classification of interactionally salient breathing.

Last but not least, given scarcity of corpora of spontaneous multiparty interactions, it is expected that the corpus will be a valuable resource for many other dialogue studies not necessarily related to studying respiration. We plan to make the corpus available for research use.

Acknowledgements

4.   References

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

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

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)*, London. Academic Press.

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.