RESPIRATORY PROPERTIES OF BACKCHANNELS IN SPONTANEOUS MULTIPARTY CONVERSATION

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

ABSTRACT

In this paper we report on first results of a newly started project focussing on interactional functions of breathing in spontaneous multiparty conversation. Specifically, we investigate respiratory patterns associated with backchannels (short feedback expressions), and compare them with breathing cycles observed during longer stretches of speech or while listening to interlocutor's speech. Overall, inhalations preceding backchannels were found to resemble those in quiet breathing to a large degree. The results are discussed in terms of temporal organisation and respiratory planning in these utterances.

Keywords: Breathing, multiparty conversation, turn-taking, verbal feedback

1. INTRODUCTION

Right from the moment humans draw their first breath and cry loudly announcing their arrival into the world, respiration is inextricably linked to sound production. With age, increasing motoric coordination within the individual leads to a shift from inarticulate cries to what we recognise as speech. The role of breathing in these coordinative processes has been studied before and is relatively well known [12, 13].

Similar coordination, this time *between* individuals, is necessary for emergence of interactional events recognisable as dialogues and conversations. This organisation is perhaps most readily visible in the patterns of alternately speaking and staying silent. In this way, respiratory mechanisms too become incorporated into the dense network of interpersonal dependencies underlying human communication. However, so far only a handful of studies addressed interactive facets of breathing.

An aspect which has received particular attention is sensitivity of respiration to alternation between speaking and listening. Namely, speech breathing is characterised by short inspirations followed by prolonged expirations and is a dramatic departure from the more symmetric tidal breathing cycle [16, 19]. In addition, some studies (e.g. [16]) reported that in subjects listening to their interlocutor's speech the breathing cycle deviates from the tidal pattern towards the asymmetric cycle characteristic of speech breathing. The finding was interpreted in the light of increasing motoric activation ([5], cf. [2]).

Finally, certain adaptations of the breathing cycle to specific turn configurations have been noted in literature [19, 14]. For instance, inhalations coinciding with turn-holds were found to be shorter than those before turn-changes, and unsuccessful attempts at interrupting the interlocutor were initiated later in the exhalatory phase than successfull interruptions or non-interruptive speech.

By contrast, little is known about respiratory characteristics of backchannels, which, due to their limited lexical repertoire and their brevity might require less preparation and smaller expiratory capacity than full dialogue turns. Even though in most earlier studies (e.g. [16, 17, 19]) backchannels were subsumed under the "quiet breathing" rubric, the decision was based on turn-taking rather than respiratory grounds.

One piece of evidence suggesting that backchannels do indeed have different respiratory demands than longer stretches of speech comes from a pilot study of dyadic conversations in Estonian. In those dialogues backchannels were distributed more uniformly within the respiratory cycle than longer non-backchannel turns [1], indicating they might require less respiratory planning than evidenced in (primarily read) speech [7]. At the same time, Rahman et al. [17] were able to discriminate between silent breathing and periods of listening based on increased cycle amplitude, an outcome attributed largely to presence of backchannels. However, it is not clear whether the increased volume was, as the authors suggest, due to deeper inhalations (indicating anticipatory activity) or simply due to the greater amount of air being exhaled during the vocalisation. Finally, [20] found that short answers are more likely to be preceded by an inhalation than longer ones.

In this paper we aim to fill the existing gap by reporting first results on respiratory profiles of backchannels, which consist in short (often monosyllabic) feedback expressions serving as a basic grounding mechanism in dialogue [21]. We draw our data from a corpus of spontaneous multiparty conversa-

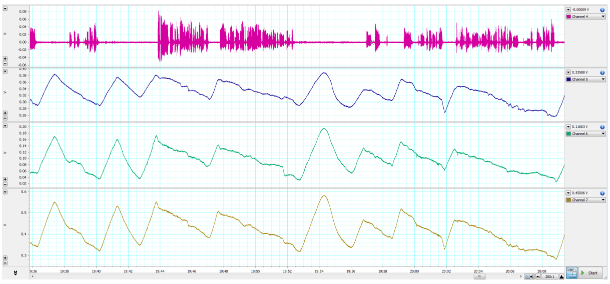

Figure 1: Speech recording (channel 1) and respiratory measurements from rib-cage and abdomen belts (channels 2-3) for one speaker. The bottom channel shows the weighted sum of the two belts.

tions in Swedish, which we are currently collecting as part of a freshly started project investigating communicative functions of breathing.

2. METHOD

Two recordings of three-party conversations in Swedish (21:55 and 27:18 minutes long) were used in the present study. All participants were native speakers of Swedish. The topic and the course of interaction were not restricted in any way.

Each participant's breathing was recorded using Respiratory Inductance Plethysmography, which measures changes in cross-sectional area of the rib cage and the abdomen by means of two elastic belts worn at the level of the armpits and the navel. Before the recording individual contributions of each belt to total lung volume change were assessed using the iso-volume manoeuvre [15]. Vital capacity and resting respiratory level were also estimated. Participants were recorded standing at a high table (95 cm), and were asked to avoid large torso movements, which would otherwise distort the respiratory trace.

The signal from the belts was sampled by RespTrack processors, designed and built at Stockholm University, and captured by PowerLab (ADInstruments). The summed signal from the two belts corresponding to the total lung volume change was captured as well. A sample signal is shown in Figure 1.

Cycles in the summed respiratory signal were identified automatically by replacing each sample value with a z-score calculated within a moving 10-second window, and locating signal maxima and minima which differ by at least 1 standard deviation in amplitude. The result was subsequently compared with manually corrected segmentations. As the differences were small (formal evaluation is forthcoming), the automatic annotations were used in subsequent analyses. Annotation errors (inhalations coinciding with speech), most likely due to large body movements were excluded from the analysis.

The respiratory signal was downsampled to 100 Hz and the following features, used in earlier studies of respiration [4, 17], were extracted for each cycle: amplitude, inhalation duration and slope, exhalation duration and slope, as well as inhalation-to-exhalation duration ratio. Amplitude was converted to z-scrores to normalise for speaker differences. Consequently, slope values are expressed in terms of change in standard deviations per second. Due to the exploratory character of the present study, we restrict ourselves to a descriptive account, leaving statistical modelling for a later study using a bigger data set.

Speech was collected with close-talking condenser microphones (Sennheiser HSP 4) and routed to PowerLab to allow synchronisation with the respiratory signal. Data collection took place in a sound-treated studio in Phonetics Laboratory, Stockholm University. The setup is described in greater detail in [6].

Voice activity detection was performed semi-automatically by manual correction of intensity-based segmentations done in Praat [3]. Talkspurts shorter than 1 second were classified as *very short utterances* (VSUs). This class of utterances was previously shown to capture a large proportion of backchannels [10].

Subsequently, every cycle was assigned to one of three classes depending whether it coincided with no speech activity, a VSU or a non-VSU speech segment. Overall, the analysed material included 703 silent cycles, 384 speech (non-VSU) cycles and 389 VSU cycles.

3. RESULTS

Distributions of (z-score normalised) amplitude in respiratory cycles coinciding with stretches of silence, speech and VSUs are plotted as kernel density estimates in Figure 2. It is apparent both from the figure and from mean amplitude values (-0.27, -0.13 and 0.29 for silent, speech and VSU cycles, respectively) that VSU cycles are indeed more similar to silent cycles than to speech breathing, which is characterised by substantially greater amplitude. More importantly, however, in our data the three distributions overlap to a large degree. The result is somewhat surprising and contrary to the intuitive assumption that speech tends to be preceded by deeper inhalations than silent breathing [17]. It should be borne in mind, however, that the analysed material consisted in friendly, non-competitive conversations recorded in a quiet laboratory environment with participants standing in close proximity to one another. It is likely that these conditions do not require lung volumes exceeding the tidal volume.

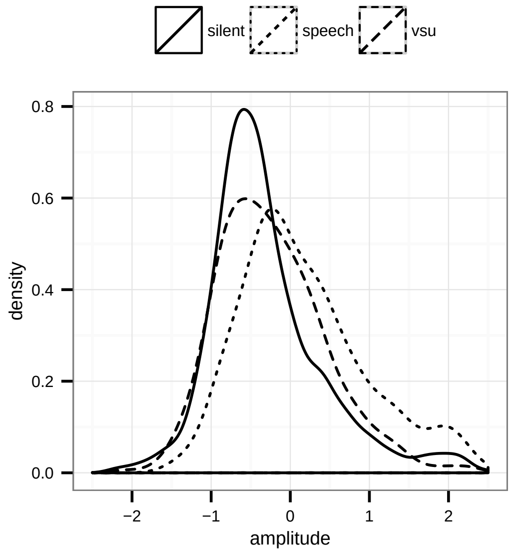

Figure 2: Kernel density estimation of amplitude (z-score normalised) in respiratory cycles coinciding with silence, speech and VSUs

In the left panel of Figure 3 we plot density distributions of inhalation durations in the three cycles types. Predictably, inhalations preceding speech are on the whole shorter than during quiet breathing (on average 1.10 s and 1.24 s, respectively). There is, however, also a sizeable number of long pre-speech inhalations. As before, backchannels (with a mean cycle duration of 1.24 s) cluster together with quiet breathing.

Much larger deviations of backchannels from the silent breathing pattern are observed when their exhalation durations are compared (Figure 3, middle panel). The corresponding mean values for silent, speech and VSU cycles equal 1.84, 3.97 and 2.52 s, respectively. This is understandable—due to increased resistance of the articulators, in backchannels the exhalatory part of the cycle is extended by the duration of the vocalisation itself. Finally, exhalations preceding speech stand out sharply against the other two groups, the difference being much larger than for inhalations. This is also expected since exhalations can be extended much more freely than inhalations can be compressed.

A particularly clean grouping in the data is revealed when distributions of inhalation-to-exhalation duration ratios are plotted, as in the right panel of Figure 3. Here, silent and speech breathing form two extremes, with VSUs placed half-way between them (mean values of this feature for silent, speech and VSU cycles are 0.31, 0.59 and 0.77, respectively). Two points should be made, however. First, given very minor differences in inhalation durations between VSU and quiet breathing, the observed pattern needs to be attributed largely to variation in exhalation durations. Second, even though silent cycles are comparatively least skewed, they are nowhere near perfectly symmetrical. In fact, in 20% of all silent cycles the exhalation is at least twice as long as the inhalation.

Due to space limitations, we do not include density distributions of inhalation and exhalation slopes. However, the results follow the expected pattern:[1] inhalations in silent breathing cycles are on average less steep than in speech cycles (-0.23 and -0.06, respectively), and backchannels resemble closely the former (with a mean of -0.21). The opposite is true for exhalation slopes with listening cycles being faster (-0.27) than speech cycles (0.04) and with backchannels falling almost exactly half-way between these values (-0.16).

4. DISCUSSION

The results summarised in the previous section point towards a very close affinity between breathing patterns coinciding with VSUs and during periods of listening. In particular, similar to listening cycles, VSUs are preceded by longer and slower inhalations than those found in speech. While these results will need to be supported by suitable statistical analysis, given the small differences in both cycle amplitude, inhalation duration and inhalation slope, we find limited evidence for respiratory planning in these utterances. It is likely that, given their short durations and their perceptual and interactional unobtrusiveness [9], VSUs have modest respiratory demands and can be easily produced also at lower lung volumes, thus requiring little respiratory readjustments. Consequently, even though we have not addressed the issue directly, the results are in agreement with Aare et al. [1], who found backchannels and short backchannel-like utterances (discourse markers, short answers) to be distributed more uniformly within the respiratory cycle than longer stretches of speech. The findings thus provide support for those descriptions of backchannels which stress their freedom in terms of temporal organisation [9, 11].

Finally, our results have certain implications for studies which subsume cycles coinciding with VSUs

---

[1] The reader is reminded that slope is expressed in z-scored amplitude change per second. Thus, for inhalations higher values correspond to steeper (positive) slopes. Conversely, for exhalations lower values correspond to steeper (negative) slopes. However, due to the normalisation procedure, the sign no longer reflects slope direction.

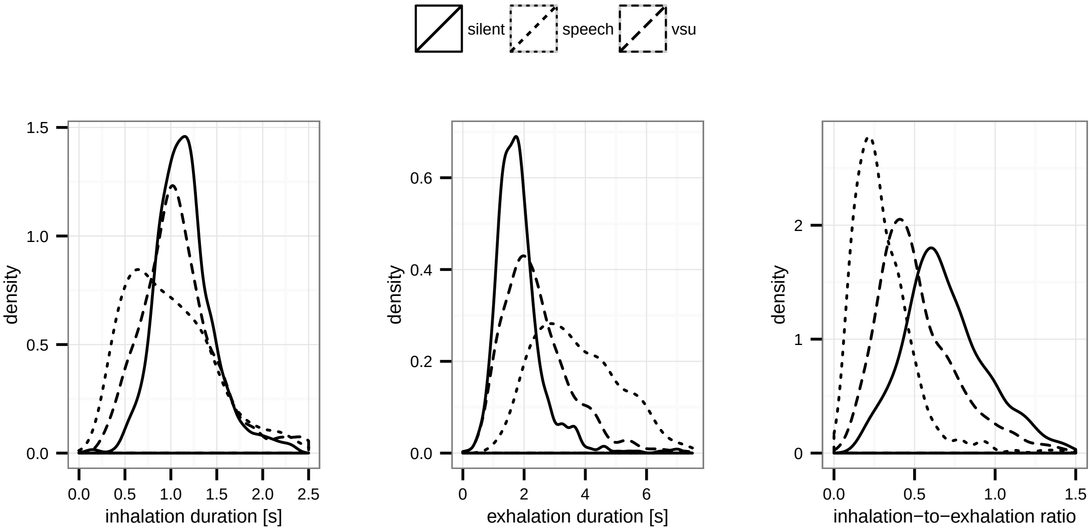

Figure 3: Kernel density estimation of inhalation duration (left), exhalation duration (middle) and inhalation-to-exhalation duration ratio (right) in respiratory cycles coinciding with silence, speech and VSUs.

under the silent cycle category. While it is true that as far as the inhalation is concerned, there is very little difference between the two, the fact that in VSU cycles the exhalatory part is extended by the duration of the vocalisation might have a marked impact on obtained results. For instance, in our data it shifts mean inhalation-to-exhalation ratio by 8%, from 0.77 for silent cycles only to 0.71 for the mixed set.

Beyond short feedback utterances, we have found what might seem like surprisingly small variation in amplitude across speech and silent cycles. We have hypothesised that this is largely due to the fact that our recordings took place in a quiet recording studio, which did not require air volumes much larger than those found in tidal breathing. In that respect our material is likely to be different from data collected in the field [17]. We have also found that the three cycle types considered here are best differentiated by the ratio of inhalation to exhalation duration, and we have noted the antisymmetry of the listening cycle reported previously for dyadic conversations [19, 16].

5. CONCLUSIONS AND FUTURE WORK

The contributions of the present paper are two-fold. First, we have described a fully automatic setup allowing zero-manual effort analysis of respiratory cycles from a recording of breathing in a naturalistic non-restrained multiparty conversation. While speech segmentation was done in a semi-automatic manner, we are currently working on applying more sophisticated voice activity detection algorithms, used in our previous work [8], to the data. Finally, following [10] backchannels have been operationalised as very short utterances, removing much of the definitional confusion surrounding the term and requiring no manual work.

Second, using this setup we have provided a preliminary account of respiratory properties of short feedback utterances, one of the most pervasive phenomena in spontaneous interactions, which have been shown to combine characteristics of both silent and speech breathing. We have also discussed implications of the fidings for accounts of respiratory planning in speech as well as possible effects of including VSU cycles in the category of silent breathing on obtained results.

In addition to reproducing these findings on a data set comprising more speakers, in the future we plan to expand on the results presented above by investigating respiratory properties of specific turn configurations (silences and overlaps followed by speaker change or more speech from the previous speaker). We also intend to replace z-score normalisation with lung volumes expressed as percentages of speaker's vital capacity with a view to obtaining more easily interpretable results. Last but not least, we are currently looking into more dynamic descriptions of the breathing cycle [18] in addition to methods based on extraction of specific features.

6. ACKNOWLEDGEMENTS

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

7. REFERENCES

[1] Aare, K., Włodarczak, M., Heldner, M. 2014. Backchannels and breathing. *Proceedings of FONETIK 2014* Stockholm, Sweden. 47–52.

[2] Bailly, G., Rochet-Capellan, A., Vilain, C. 2013. Adaptation of respiratory patterns in collaborative reading. *Proceedings of Interspeech 2013* Lyon, France. 1653–1657.

[3] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer. Computer program. http://www.praat.org/.

[4] Boiten, F. 1993. Component analysis of task-related respiratory patterns. *International Journal of Psychophysiology* 15(2), 91–104.

[5] Conrad, B., Schönle, P. 1979. Speech and respiration. *Archiv für Psychiatrie und Nervenkrankheiten* 226(4), 251–268.

[6] Edlund, J., Heldner, M., Włodarczak, M. 2014. Catching wind of multiparty conversation. *Proceedings of Multimodal Corpora 2014* Reykjavík, Iceland.

[7] Fuchs, S., Petrone, C., Krivokapić, J., Hoole, P. 2013. Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics* 41(1), 29–47.

[8] Heldner, M., Edlund, J. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38(4), 555–568.

[9] Heldner, M., Edlund, J., Hirschberg, J. 2010. Pitch similarity in the vicinity of backchannels. *Proceedings of Interspeech 2010* Makuhari, Japan. 3054–3057.

[10] Heldner, M., Edlund, J., Hjalmarsson, A., Laskowski, K. 2011. Very short utterances and timing in turn-taking. *Proceedings of Interspeech 2011* 2837–2840.

[11] Heldner, M., Hjalmarsson, A., Edlund, J. 2013. Backchannel relevance spaces. Asu, E. L., Lippus, P., (eds), *Nordic Prosody: Proceedings of the XIth Conference, Tartu 2012* Frankfurt am Main. Peter Lang 137–146.

[12] Hixon, T. J., Goldman, M. D., Mead, J. 1973. Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung. *Journal of Speech, Language and Hearing Research* 16(1), 78–115.

[13] Hixon, T. J., Mead, J., Goldman, M. D. 1976. Dynamics of the chest wall during speech production: Function of the thorax, rib cage, diaphragm, and abdomen. *Journal of Speech, Language and Hearing Research* 19(2), 297–356.

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

[15] Konno, K., Mead, J. 1967. Measurement of the separate volume changes of rib cage and abdomen during breathing. *Journal of Applied Physiology* 22(3), 407–422.

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

[17] Rahman, M. M., Ali, A. A., Plarre, K., al'Absi, M., Ertin, E., Kumar, S. 2011. mConverse: Inferring conversation episodes from respiratory measurements collected in the field. *Proceedings of the 2nd Conference on Wireless Health* San Diego, CA. 1–10.

[18] Ramsay, J. O., Silverman, B. W. 2005. *Functional data analysis*. Springer Series in Statistics. Berlin: Springer.

[19] Rochet-Capellan, A., Fuchs, S. 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.

[20] Torreira, F., Bögels, S., Levinson, S. C. 2015. Breathing for answering: the time course of response planning in conversation. *Frontiers in Psychology* 6, 1–11.

[21] Yngve, V. 1970. On getting a word in edgewise. *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society* Chicago. 567–577.