Backchannels and breathing

Kätlin Aare, Marcin Włodarczak, Mattias Heldner
Department of Linguistics, Stockholm University, Sweden
k2tlin.a@hotmail.com, wlodarczak@ling.su.se, heldner@ling.su.se

Abstract

The present study investigated the timing of backchannel onsets within speaker's own and dialogue partner's breathing cycle in two spontaneous conversations in Estonian. Results indicate that backchannels are mainly produced near the beginning, but also in the second half of the speaker's exhalation phase. A similar tendency was observed in short non-backchannel utterances, indicating that timing of backchannels might be determined by their duration rather than their pragmatic function. By contrast, longer non-backchannel utterances were initiated almost exclusively right at the beginning of the exhalation. As expected, backchannels in the conversation partner's breathing cycle occurred predominantly towards the end of the exhalation or at the beginning of the inhalation.

Introduction

Conversational turn-taking involves coordination between participants exchanging the roles of speakers and listeners, and backchannel communication is part of this system. Backchannels (Yngve, 1970) are short, typically mono- or disyllabic (Gardner, 2001) listener responses in dialogues or conversation. The term backchannel has been coined to refer to the background channel through which the listener can give feedback to the speaker without claiming the conversational floor. Backchannels indicate that the listener is following and understanding the speaker (e.g. Heldner, Hjalmarsson, & Edlund, 2013). In face-to-face dialogues participants make use of visible as well as audible means of communication, backchannels can therefore be both verbal and non-verbal. Verbal backchannels can be more generic like *uh-huh* or *m-hm*, or more specific to signal what the addressee has understood, like *oh* or other markers for surprise, for example. Research has shown that listeners show a great variety of behaviors to contribute specific responses (Bavelas & Gerwing, 2011).

Respiration during speech can be both audible and visible, and breathing patterns in speech have been claimed to be relevant for conversational organization. For instance, an audible inhalation before an utterance has been suggested to be a “pre-beginning” element in turn-taking mechanisms (Schegloff, 1996). The respiratory pattern changes during spontaneous conversations. It has been noted that the quiet breathing cycle is repeated about 12 times per minute, and exhalation is slightly longer than inhalation. The frequency of breathing changes for speech breathing, with the inhalation phase being considerably shorter than the exhalation phase to minimize interruption to the flow of speech (Hixon, 1987). It has also been shown that most speakers take a deeper breath before longer or more complicated sentences (Fuchs et al., 2008; Winkworth, Davis, Adams, & Ellis, 1995). Prephonatory movements of the rib cage and abdomen have been reported to be adaptive to different speech tasks, indicating that there may indeed be preparatory respiratory processes occurring during listening and preparation of turn onset (McFarland, 2001).

To summarize, next speakers prepare turn onset among other things by inhaling, and this is potentially an important turn-taking signal. By contrast, it remains unclear if and how listeners prepare the onset of backchannels.

Backchannels are typically short, brief and quiet, and these characteristics do not require as much exhaled air and effort as longer utterances. Furthermore, backchannels carry relatively little propositional content and they are not supposed to claim the conversational floor. All of this taken into account, it is conceivable that backchannels are not planned the same way longer utterances are, and furthermore that they do not necessarily have to be initiated at the beginning of the (listener's) exhalation phase. In this study, we will explore our intuition that backchannels may occur more freely in the respiratory cycle than longer utterances. We will also explore whether this is related to their non-floor claiming properties, or just to their relative shortness. Finally, we will explore how backchannels are timed relative to the other speaker's breathing cycle.

Method

For the purpose of this exploratory study, we recorded respiratory activity synchronized with audio in two spontaneous two-party dialogues of approximately 20 minutes each. The subjects were two females and two males, aged 18-25, all native speakers of Estonian. The subjects all knew each other. The first dialogue was between two sisters, and the other one between two young men who had known each other for one and a half years. They had no knowledge of the aim of the experiment before the recording. They were free to talk about any topic throughout the recording session. None of the subjects reported any speech or hearing disorders. One speaker had suffered from a breathing disorder caused by low blood pressure, and two were smokers. All subjects were of slim body type and wore tight-fitting clothes.

The recordings took place in a quiet, sound-treated room in the Phonetics Laboratory at Stockholm University. To minimize noise in the respiratory signals caused by body movement, the subjects were recorded standing facing each other at a bar table keeping their hands on the table.

Respiratory activity was measured 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. The belts were connected to dedicated respiratory belt processors (RespTrack) designed and built in the Phonetics Laboratory at Stockholm University. The RespTrack processor was designed for ease of use, and optimized for low noise and low interference 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 the processors supplied with the belt, there is no high-pass filter, thus the amplitude will not decay during periods of 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 a sum signal allowing a direct estimation of lung volume change. The calibration of the belts for the estimated volume change between the two chest walls was achieved by performing the isovolume maneuvre (Konno & Mead, 1967).

Audio was captured using head-worn microphones with a cardioid polar pattern (Sennheiser HSP 4). The audio and belt processor signals were recorded synchronously using an integrated physiological data acquisition system consisting of LabChart software and PowerLab hardware (ADInstruments, 2014), which also allows connecting other measuring instruments, such as air-flow masks or electroglottographs. Figure 1 shows an example of synchronized audio and respiratory measurements from one speaker. The setup is described in greater detail in Edlund, Heldner, & Włodarczak (2014).

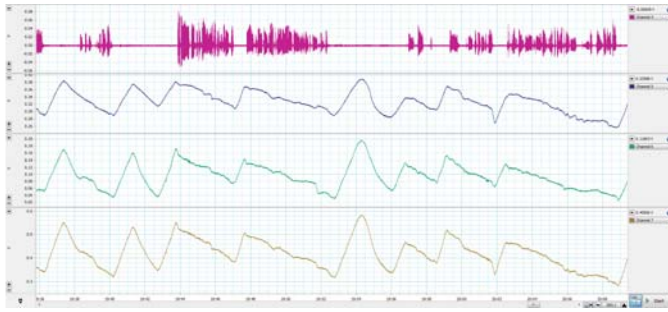

Figure 1. An example of synchronized audio and respiratory measurements from one speaker. The channels (from top to bottom) show the audio signal, the rib-cage signal, the abdomen signal, and the weighted sum of the two belts.

The audio and breathing signals were subsequently manually annotated using Praat (Boersma & Weenink, 2014). The rib cage and abdomen movements were used to segment the breathing signals into periods of inhalations and exhalations. The speech signal was segmented into intervals of pauses, utterances or backchannels, the latter delimited by pauses of at least 500 ms. A Praat script was used to extract timing of speech and breathing events.

Speech onsets were normalized with respect to their relative position within the breathing phase they coincided with: exhalation within speaker's own breathing cycle, inhalation or exhalation within interlocutor's breathing cycle.

Results

Backchannels vs. utterances

A total of 277 backchannels and 732 (non-backchannel) utterances were included in the analyses. A small number of backchannels was excluded from analysis, either because they were produced in the inhalation phase *(N=1)*, or because they erroneously spanned more than one breathing cycle *(N=4)*. The remaining backchannels were mostly short markers of agreement (*m-hm, ahah, jajah* 'yes-yes', *okei*), but also of surprise (*tegelt* 'really'). Figures 2 and 3 show the distribution of normalized onset times for utterances and backchannels, respectively.

As expected, there was a strong tendency for non-backchannel utterances to start early in the exhalation phase. About 44% of all utterances started within the first tenth of the exhalation (i.e. the first two bins). This tendency was considerably weaker in the backchannels, where only about 27% started within the first tenth of the exhalation, and where another mode in the distribution was discernable in the second half of the exhalatory phase. Thus, the backchannels were more evenly distributed across the exhalation phase than the non-backchannel utterances. This is in line with previous findings on German (Fuchs, personal communication), where the tendency was even more marked and backchannels were equally likely throughout the breathing cycle.

Longer vs. shorter utterances

To explore whether the observed difference between backchannels and utterances was related to the relative shortness of the backchannels rather than their non-floor claiming properties, the utterance data was split into two groups based on duration. As 99% of the backchannels were shorter than 0.8 s, this duration was used as the criterion for separating short utterances (<0.8 s) from longer utterances (>0.8 s). Manual inspection of the former revealed that these utterances consisted mainly in short answers, pause-delimited discourse markers, and stretches of disfluent or otherwise incomplete turns.

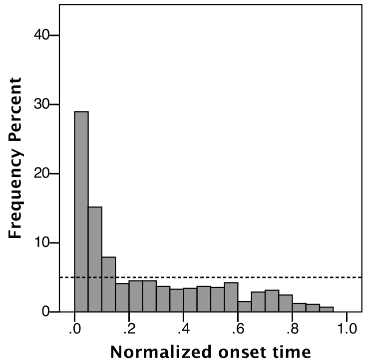

Figure 2. Distribution of normalized onset time for non-backchannel utterances.

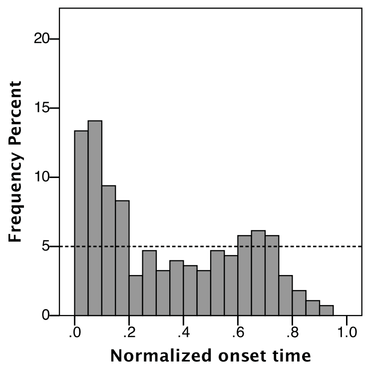

Figure 3. Distribution of normalized onset time for backchannels.

A total of 216 shorter utterances and 516 longer utterances were identified. Figures 4 and 5 show the distribution of normalized onset times for longer and shorter utterances, respectively.

The longer utterances displayed a pattern similar to that observed for all utterances (cf. Figure 2), although the tendency was stronger. About 50% of all longer utterances started in the first tenth of the exhalation. The shorter utterances showed a pattern markedly different from the longer ones. Here, only about 32% of the shorter utterances started in the first tenth of the exhalation and there was a second mode in the distribution around 0.7.

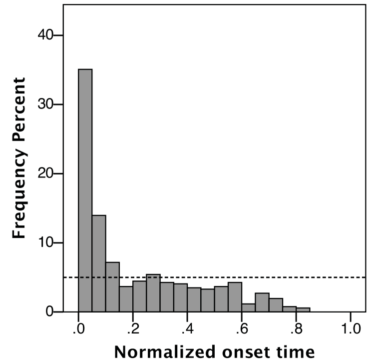

Figure 4. Distribution of normalized onset time for longer utterances (>0.8 s).

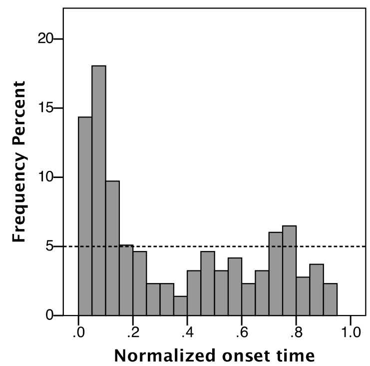

Figure 5. Distribution of normalized onset time for shorter utterances (>0.8 s).

Thus, shorter utterances were more evenly distributed in the exhalation phase, and behaved similarly to backchannels (cf. Figure 3).

Backchannels in the other speaker's breathing pattern

Finally, we wanted to explore if there is a pattern in how backchannels are timed relative to the other speaker's breathing cycle. Therefore, we calculated normalized onset times relative to the other speaker's inhalations and exhalations. All backchannel occurrences *(N=282)* were included in this analysis. Figures 6 and 7 show the distribution of onset time for backchannels normalized relative to exhalations and inhalations in the other speaker's speech, respectively.

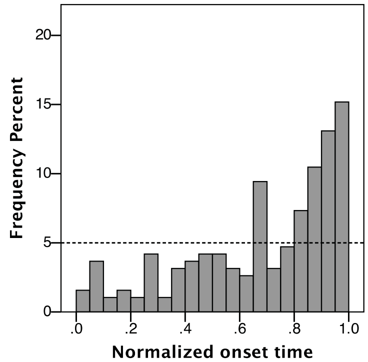

Figure 6. Distribution of normalized onset time for backchannels in the other speaker's exhalations.

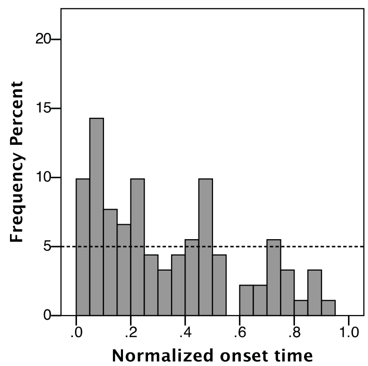

Figure 7. Distribution of normalized onset time for backchannels in the other speaker's inhalations.

The majority of the backchannels (67.5%) were produced during the other speaker's exhalations. The shape of the distribution for exhalations (Figure 6) shows that backchannels were increasingly more frequent towards the end of the other speaker's exhalation.

For the remaining backchannels produced during the other speaker's inhalations, the pattern was the reverse with decreasingly less backchannels towards the end of the other speaker's inhalation (Figure 7).

Discussion

The comparison of backchannels and non-backchannel utterances (Figures 2 and 3) indicates a clear distinction in their temporal organization with respect to speaker's own the respiratory cycle: non-backchannels are initiated predominantly towards the beginning of the exhalation, a tendency which is less pronounced in backchannels where another, somewhat smaller, peak is present towards the end of the exhalatory phase. While this observation suggest a functionally motivated difference, results in Figures 4 and 5, in which non-backchannel utterances where further split depending on their duration, contradict this hypothesis. Specifically, backchannels and comparably short non-backchannels behave very similarly. They are distributed more uniformly than longer utterances with two local maxima: one near the beginning of the exhalation and another between 70 and 80% of its duration. Consequently, it suggests that duration rather than pragmatic function is the decisive factor determining turn initiation patterns. Simply put, if an upcoming turn is short enough, it is produced immediately, without the need for a deep inhalation characteristic of longer stretches of speech. Not surprisingly, backchannel onsets did not always coincide with exhalation in the interlocutor's breathing cycle. Instead, they were most common around the transition between exhalation and inhalation. Insofar as this location corresponds to partner's turn or phrase boundaries, the observed pattern is most likely brought about by the underlying grounding mechanism, whereby feedback acknowledges the new piece of information produced in the previous turn constituent.

Conclusions

The present study revealed that backchannels and non-backchannel utterances of corresponding length are timed in a similar way within the speaker's breathing cycle. They are most likely to be initiated towards the beginning of the exhalation or roughly around 70% of its duration. By contrast, longer non-backchannel utterances are extremely rare anywhere but at the very onset of the exhalatory phase. The observed similarity indicates that timing of speech with respect to the respiratory phase is motivated by turn length, and not its pragmatic function. Consequently, backchannels cannot be distinguished from non-backchannels on the basis of position within the respiratory cycle alone. At the same time, backchannels were found to occur most frequently in the vicinity of interlocutor's exhalation offset, which is likely to reflect processes related to grounding of new information.

Acknowledgements

The work was funded in part by the Swedish Research Council (VR) project *Samtalets rytm* (2009-1766).

References

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

Bavelas, J. B., & Gerwing, J. (2011). The Listener as Addressee in Face-to-Face Dialogue. *International Journal of Listening, 25*(3), 178-198.

Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program] (Version 5.3.75). Retrieved from http://www.praat.org/

Edlund, J., Heldner, M., & Włodarczak, M. (2014). Catching wind of multiparty conversation. In J. Edlund, D. Heylen & P. Paggio (Eds.), *Proceedings of MMC 2014*. Reykjavik, Iceland.

Fuchs, S., Hoole, P., Vornwald, D., Gwinner, A., Velkov, H., & Krivokapić, J. (2008). The Control of Speech Breathing in Relation to the Upcoming Sentence. In *Proceedings of the 8th International Seminar on Speech Production (ISSP 2008)* (pp. 77-80). Strasbourg, France.

Gardner, R. (2001). *When Listeners Talk: Response Tokens and Listener Stance*. Amsterdam: J. Benjamins Publishing.

Heldner, M., Hjalmarsson, A., & Edlund, J. (2013). Backchannel Relevance Spaces. In E. L. Asu & P. Lippus (Eds.), *Nordic Prosody: Proceedings of the XIth Conference, Tartu 2012* (pp. 137-146). Frankfurt am Main, Germany: Peter Lang.

Hixon, T. J. (1987). Respiratory Function in Speech. In T. J. Hixon (Ed.), *Respiratory Function in Speech and Song* (pp. 1-54). Boston, MA, USA: Little Brown.

Konno, K., & Mead, J. (1967). Measurement of the Separate Volume Changes in the 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*(1), 128–143.

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

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.

Winkworth, A. L., Davis, P. J., Adams, R. D., & Ellis, E. (1995). Breathing patterns during spontaneous speech. *Journal of Speech and Hearing Research, 38*(1), 124-144.

Yngve, V. H. (1970). On getting a word in edgewise. In *Papers from the sixth regional meeting of the Chicago Linguistic Society* (pp. 567-578). Chicago, IL, USA: Chicago Linguistic Society.