Breath Holds in Chat and Chunk Phases of Multiparty Casual Conversation

Kätlin Aare[1,3]*, Emer Gilmartin*[2]*, Marcin Włodarczak*[1]*, Pärtel Lippus*[3]*, Mattias Heldner*[1]

[1]Stockholm University, Sweden
[2]ADAPT Centre, Trinity College Dublin, Ireland
[3]University of Tartu, Estonia

{katlin.aare, marcin.wlodarczak, mattias.heldner}@ling.su.se, gilmare@tcd.ie,
partel.lippus@ut.ee

Abstract

Breathing has been associated with all the basic turn organisation categories, with breath holds, where a participant holds their breath for a short period of time, specifically claimed as markers of turn incompleteness. Casual conversation comprises stretches of interactive chat interspersed with more monologic chunks where one speaker dominates the conversation, often telling a story or giving an extended opinion. We investigate the interplay of breath holds with chat and chunk phases in 16 three-party casual conversations in Estonian and Swedish, using data annotated for respiratory and speech activity. Our results demonstrate that breath holds occur predominantly before transitions between chat and chunk, and that within chunk breath holds are largely produced by listeners, implying that the main speaker does not need to resort to the use of breath holds to retain their turn.

Index Terms: turn taking, speech breathing, breath holds, spontaneous speech, multiparty conversations

1. Introduction

Casual conversation has been described as an emergent property of congregating people. Breath has been linked with turn organisation by participants in conversation, and breath holding in particular is associated with a speaker's intention to retain the conversational floor. In this paper we investigate breath holds in multiparty casual conversation. We begin with a brief overview of breath holds and their connection to turn dynamics in conversation. We then review current understanding of casual conversation with particular focus on its chat and chunk structure. We describe our investigation of breath holds in chat and chunk phases of a collection of 8 Estonian and 8 Swedish 3-party conversations, and discuss the results.

1.1. Breath

Breathing has been associated with all the basic turn organisation categories: it has been suggested that an audible inhalation is a pre-beginning element in turn-taking and projects the onset of talk [1], while it has also been claimed that breath holding can function as a marker of turn incompleteness [2], and exhaling can be a turn-yielding device [2, 3, 4].

Breath holds are voluntary parts of respiratory activity when air is neither inhaled nor exhaled. While breath holds often occur to protect the airway with a glottal closure, such as during swallowing, coughing, or sneezing, they occur frequently in conversations and have been claimed to function as a marker of turn incompleteness [2, 3, 4]. An overview of breath hold characteristics and variance in lung volume and timing within exhalations in spontaneous Estonian is provided in

[5]. Research on conversational silences offers some insight into how breath holds may affect turn organisation in conversations. In particular, [3] discuss "holding silences" – intra-turn pauses following conjunctionals, ending in glottal closures maintained through silence and released at the beginning of the following word produced by the same speaker. They differentiate these from "trail-off silences" which end with an audible exhalation not terminated in glottal closure, thus not considered a "closure piece". Differences between the two are established by other phonetic features as well, such as no noticeable slowing of speech tempo or decrease in loudness in the "trail-off silences", while these features were found to characterise the "closure pieces". Jefferson adds that if the current speaker produces a "holding silence" even if overlapping talk occurs, they still maintain their turn [6, p195].

More specific information about breathing-related turn-taking mechanisms include results from a study on German by [7], indicating the adaptation of breathing to dialogue turns and evidence for inter-personal coordination of breathing in turn-taking at a global level. They explain that breathing profiles are different depending on whether speakers are trying to claim the turn or if they are holding the turn: in general, respiratory cycles in turn-taking were longer than in turn-holding. Further, their data on spontaneous German speech demonstrated that turn-taking was more successful after a new inhalation, indicating that speakers coordinate their breathing with turn-taking.

1.2. Casual conversation

Casual conversation, is ubiquitous in human life, and has been viewed as the most basic use of speech [8]. Such 'talk for the sake of talking' is often multiparty, follows no pre-arranged agenda, and affords equal rights to all speakers [9, 10, 11]. Its purpose is not the pursuit of a short-term practical goal such as a business transaction, but rather to avoid threatening or unfriendly silence and promote social bonding [12, 13]. Casual conversation is described as 'interpersonal' rather than 'instrumental' [14], and can contain smalltalk or interactive chat in addition to stretches of gossip, opinion or narrative [9, 15]. In early work, Laver focussed on smalltalk at the 'psychologically crucial margins of interaction' at the beginning and end of speech exchanges. He suggested that light smalltalk eased transition from initial greetings to the main business of the interaction and back to closing sequences and leave taking. He also proposed that such talk suspended power or social differentials between interlocutors, and tended to treat topics which were uncontroversial. [16]. Ventola described entire casual conversations in terms of distinct phases [17]; often beginning with ritualised opening greetings, followed by approach segments of light uncontroversial small talk, and in longer con-
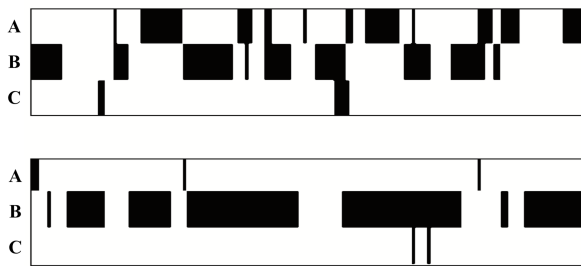
Figure 1: *Two 20-second extracts from a 3-party Swedish conversation in the dataset. The upper panel shows a stretch of chat, while the lower panel shows a stretch of chunk where the middle speaker is the chunk owner. A, B, C indicate different speakers' voice activity.*

versations leading to more informative centre phases consisting of sequential but overlapping topics, and then back to ritualised leavetakings. Slade and Eggins divided casual conversation into distinct phases of 'chat' and 'chunk' [9]. Chunks are segments where 'one speaker takes the floor and is allowed to dominate the conversation for an extended period', and the chunk appears to move through predictable stages – that is, it is generic and amenable to genre analysis. 'Chat' segments, on the other hand, are described as highly interactive, appearing to be managed locally, unfolding move by move or turn by turn. Slade found that around fifty percent of casual conversation recorded during tea breaks could be classified as chat, while the rest comprised longer form chunks of storytelling, observation/comment, opinion, gossip, and joke-telling, with narrative monologue appearing as the most frequent chunk genre ([18]. Figure 1 shows examples of chat and chunk phases taken from the dataset described in this paper.

1.3. Research goals

We investigate breath hold occurrence in two languages and in chat and chunk phases. We focus on the timing, rate, and production of breath holds by listeners and speakers.

Firstly, we are interested whether the occurrence of breath holds is affected by the conversation language, or by the conversational phase in which breath holds occur. Specifically, are breath holds more common in the chat phases of a conversation where speaker transitions are frequent, or in the chunk phases where one speaker dominates? We expect to see more breath holds in chats as that is the more dynamic part of the conversation.

Secondly, we look at the relative timing of breath hold offsets within the conversational phase they coincide with: we expect breath holds to occur in both phases but demonstrate differences in chunks when it comes to the producer of breath holds.

Thirdly, as breath holds have been claimed to mark turn incompleteness, we check if breath holds are produced by listeners or speakers, and if the chunk owners/speakers produce breath holds between their utterances in a chunk in casual conversations. As chunks are largely narrative, and therefore of predictable structure, we would like to see if breath hold frequency in chunk owner speech reflects Local and Kelly's observations on breath holds in turn incompleteness.

Finally, we investigate whether listener-produced breath holds co-occur with other listener activities in conversation, such as laughter and backchannels. If breath holds are directly

followed by those other activities, breath holds could help produce them in a timely way by allowing speakers to prepare (and stay ready) for speech production ahead of time.

2. Materials & methods

The data used in this work are three-party spontaneous conversations in Estonian or Swedish. In both subsets, the participants in each conversation knew each other well and were native speakers of the language. The conversations lasted approximately 25 minutes. The data in the Estonian-speaking subset were collected from 24 unique speakers (13F and 11M; mean age = 24.5 years, SD = 2.75) with an average Body Mass Index of 22.5 (SD = 2.3), participating in 8 different conversations. The data in the Swedish-speaking subset were collected from 24 unique speakers (12F and 12M; mean age = 25.5 years, SD = 10) with an average Body Mass Index of 22.7 (SD = 2.5), participating in 8 different conversations.

2.1. Method

The conversations were recorded in a sound-treated room in the Phonetics Laboratory at Stockholm University. Respiratory signals, audio and video were recorded separately for each participant. The speakers were asked to wear tight-fitting shirts for the recording procedure to maximize the accuracy of the signals from the transducer belts. During the experiment, the speakers were standing around a 1-meter-high round table for the calibration and entire conversation duration, and were instructed to avoid large movements.

Respiratory activity was measured with Respiratory Inductance Plethysmography [19], which quantifies changes in the rib cage and abdominal cross-sectional areas by means of two elastic transducer belts (Ambu RIP-mate) placed at the level of the armpits and the navel. The calibration of the belts for the estimated volume change between the two chest walls was achieved by performing the isovolume maneouvre [20]. A more detailed setup description is provided in [4].

The respiratory signal was recorded using PowerLab (ADInstruments). Audio was captured using head-worn microphones with a cardioid polar pattern (Sennheiser HSP 4). Each participant was facing a GoPRO Hero 3+ camera recording the movements of the upper part of the torso and the head.

The audio signal was annotated by identifying sections of silence in the speech signal automatically with ELAN [21]. In the Estonian-speaking subset, this was followed by a manual adjustment of the consequent interval borders as well as changing interval labels according to their function (i.e. separating backchannels from the rest of speech as they are considered feedback units which do not claim the conversational floor [22], marking laughter and coughing, etc.) in Praat [23]. Annotation of the respiratory data in both subsets was carried out automatically using Praat and Python scripts [24]. Automatic annotation of breath holds was based on a method by proposed for airflow signals by Noto et al. [25], adapted for RIP recordings and implemented in the RespInPeace toolkit for processing and analysis of respiratory data [26][1].

The conversations were manually segmented into phases by identifying the 'chunks' using the first, structural part of Slade and Eggins' definition - 'a segment where one speaker takes the floor and is allowed to dominate the conversation for an extended period' [9]. Chunks were annotated with the identity of the chunk owner. All other interaction was considered chat.
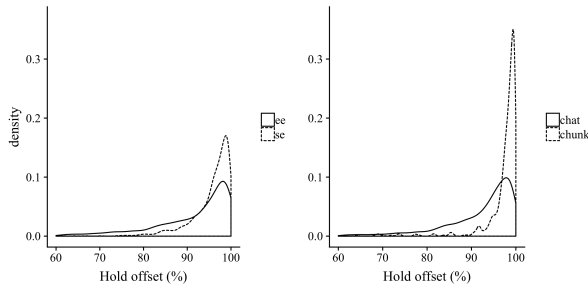
---

[1] https://gitlab.com/mwlodarczak/RespInPeace

Figure 2: *Relative breath hold offset. Left: differences in languages. Right: differences in conversational phases*
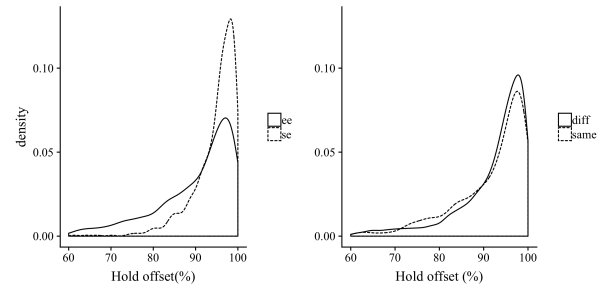

Figure 3: *Relative breath hold offset in chat. Left: differences between languages. Right: differences depending on whether the breath holder is the same person who owns the next chunk or not.*
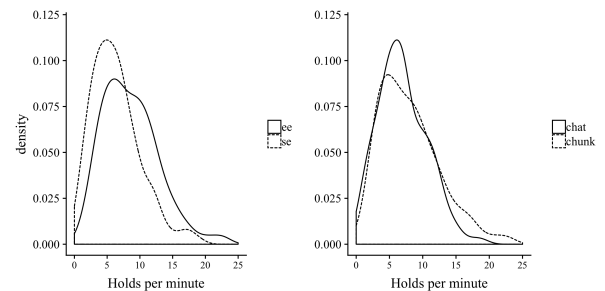

Figure 4: *Breath hold rate. Left: differences in languages. Right: differences in conversational phases*

3. Results

The data consisted of 2169 chat phases and 2088 chunk phases. On average, chunk phases took up 26% ($SD = 10\%$) of the conversation. The median values for chat and chunk durations were 159 and 26 seconds, respectively.

A total of 2214 (1337 in Estonian, 877 in Swedish) breath holds in chat and 637 (345 in Estonian, 292 in Swedish) breath holds in chunk phases were identified across the 16 conversations. The median duration of a breath hold was 0.47 seconds ($M = 0.63$, $SD = 0.5$).

More than half of all phases (54%) contain at least one breath hold. Joint probability was calculated for the occurrence of two events at the same time: breath holds and conversational phases. It is likelier for chat phases to co-occur with breath holds rather than not (30% and 20%, respectively), and it is likelier for chunk phases to co-occur with no breath holds rather than with (26% and 24%, respectively). For this calculation, chunk ownership was not taken into account.

3.1. Location of breath holds

Our results indicate that breath holds cluster before conversational phase changes - near chat or chunk phase ends - approximately fitting into the last 10% of phase duration.

In chunks, the vast bulk of breath holds (approximately 95%) are produced by listeners (non-chunk owners), and occur close to chunk ends, around the last 2% (median absolute value = 12 sec) of chunk duration approximately. In chats, breath holds are mostly located in the last 6% (median absolute value = 60 sec) of the phase. However, for chats phase ownership is undetermined and thus is not analysed here. Figure 2 illustrates the results in more detail.

The effects of conversational phase (chat or chunk) and language on breath hold location (measured at the offset of the breath hold) were tested with a two-way ANOVA. Both language and phase, as well as their interaction had a significant effect ($p < .001$) on the outcome. Considering a base value of chat phase in Estonian, breath holds would occur later in chats compared to chunks, and later in Swedish compared to Estonian. The interaction of the two variables, which realizes as chats in Swedish, results in slightly earlier occurrence of breath holds.

Further, we looked at whether the participant who produced a breath hold in the end of a chat phase would also be the owner of the following chunk phase. The results show that in 66% (68% in Estonian, 61% in Swedish) of cases, the breath holder does not own the following chunk. We used Kolmogorov-Smirnov tests to see whether the distribution of the breath hold location in the phase differed when the following chunk owner was the same as or differed from the breath holding person, and if language had an effect on breath hold location within the chat phases. The results demonstrate that there are no differences in the distribution of relative breath hold location depending on whether the following chunk owner is the same person as the breath holder or not ($p = .1$), while language does have an effect as the distributions of relative breath hold location are significantly different between Estonian and Swedish ($p < .001$). The results are illustrated on Figure 3.

3.2. Rate of breath holds

Rate of breath holds per minute was calculated for both languages and conversational phases. The median values for breath hold rate were 6.8 and 7.3 breath holds per minute for chats and chunks, respectively. In Estonian, median rates were 7.18 breath holds per minute in chat phases and 8.7 breath holds per minute in chunk phases. In Swedish, median rates were 5.5 breath holds per minute in chat phases and 6 breath holds per minute in chunk phases. The results are illustrated on Figure 4.

We fitted a linear model for log(rate) and tested the effects of language and conversational phase with a two-way ANOVA. Compared to the baseline of a chat phase in Estonian, a chunk phase would have a higher breath hold rate ($p < .05$), and in Swedish the breath hold rate in chats would be lower compared to the rate in Estonian chats ($p < .001$).

3.3. Breath holds and listener activity

As breath holds in chunks were predominantly produced by participants in a listener role and occurred very late in the phase, we were interested in listener activity following breath holds, such
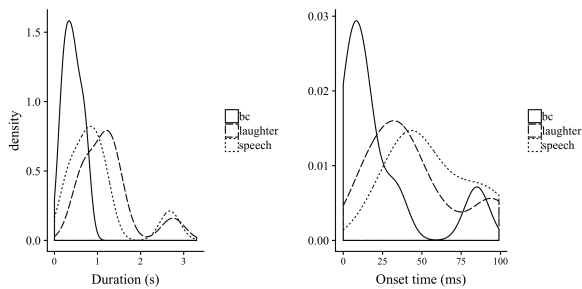
781

Figure 5: *Post-breath-hold production properties. Left: durations of the backchannels, laughter, and speech that begin within a 1-second window after the hold offset. Right: onset of backchannels, laughter, and speech after the offset of the breath hold within a 1-second time window. The y-scale differs on the panels for presentation purposes.*

as laughter and backchannels. We analysed all listener activity in a 1-second post-breath-hold window in Estonian chunks (n = 345), as information about backchannels and laughter was unavailable for Swedish data. The main results are illustrated on Figure 5.

The data showed that within 1 second from breath hold offset, listener activity occurred in 8% of the cases. After removing rarer events such as lip smacking and coughing, listeners produced mostly backchannels, laughter or speech. Of the remaining data, 27% were backchannels, 42% laughter, and 31% speech.

We also looked at how close the onset of the following production was to breath hold offset. In a 1-second window, backchannel onsets were closest to breath hold offsets (often produced without a pause between them), while speech onset was the farthest.

Additionally, the duration of the post-breath-hold speech starting within 1 second from breath hold offset was measured. The median value of the speech spurts was 815 ms.

4. Discussion

Casual conversation has been reported to comprise phases of interactive chat interspersed with more monologic chunk phases. This pattern is evident in the data studied here. The mean duration for chunks of 30s is very close to that found for English casual conversation in [27], perhaps reinforcing a common preferred chunk duration in these three languages.

In chunks, the vast bulk of breath holds (approximately 95%) are produced by listeners (non-chunk owners), and occur close to chunk ends, within the last 2% of chunk duration approximately. In chats, breath holds are mostly located in the last 10% of the phase. This tendency for breath holding near the very end of chunks could reflect preparation by listeners to act. Surprisingly, only 8% of such breath holds were in fact followed by laughter, short utterances, or backchannels by the breath holder. Additionally, the results showed that in 66% of breath holds, the breath holder was not the next chunk holder. This could indicate that for starting a new chunk, breath holds are not commonly used as a preparatory mechanism.

The results on breath hold rate did not show any significant differences between chat and chunk. However, the breath hold rates are relatively high - one explanation for this is that the breath holds analysed include the breath hold-like plateaus

which often occur around resting expiratory level (REL) in silent breathing [5]. That would also explain why only 8% of breath holds are followed by some listener activity. The same thing might also inflate the number of breath holds in chunks, resulting in comparable hold rates in the two phases. However, that does not explain why most of the breath holds occur towards the end of these phases.

While the vast bulk of pauses in chunks are within speaker, the finding that 95% of breath holds in chunks in the data were produced by listeners, and not by the chunk holder, indicates that breath holding is not used by chunk owner speakers to maintain their turn in the extended near-monologues characteristic of chunks. This may indicate that listeners generally use their knowledge of speech genre to cede the floor to the chunk owner for the duration of the chunk, in line with Slade and Eggins' theory of chunk composition, while the chunk owner expects to retain the floor for the duration of their contribution.

The differences observed between the two languages are intriguing and might indicate an interesting research topic to see if they are rooted in language-specific communication styles or are simply minor statistical differences.

5. Conclusions

These preliminary investigations have shown that the phase of conversation under consideration is important in the study of breath holds and their occurrence in spoken interaction. The positioning of breath holds towards the end of phases is interesting, and will require further investigation, while the absence of breath holding in chunk owner speech reinforces Sacks, Schegloff, and Jefferson's assertion that there are many speech exchange systems available in human spoken interaction and these may have differing turn management paradigms. In the work described here, chunk phases of conversation do not seem to conform to earlier observations of the turn retention function of breath holding in conversation, further demonstrating that casual conversation should be analysed in terms of its constituent phases. The results of this study could be used to develop speech technology applications, such as conversational agents, as well as to broaden the scope of established conversation analytical methods by including the breathing signal in future research.

6. Acknowledgements

7. References

[1] E. A. Schegloff, *Turn organization: One intersection of grammar and interaction*. Cambridge University Press, 1996, p. 52–133.

[2] P. French and J. Local, "Turn-competitive incomings," *Journal of Pragmatics*, vol. 7, no. 1, p. 17–38, 1983.

[3] J. Local and J. Kelly, "Projection and 'silences': Notes on phonetic and conversational structure," *Human Studies*, vol. 9, no. 2–3, p. 185–204, 1986.

[4] J. Edlund, M. Heldner, and M. Włodarczak, *Catching wind of multiparty conversation*, 2014.

[5] K. Aare, M. Włodarczak, and M. Heldner, "Breath holds in spontaneous speech," *Journal of Estonian and Finno-Ugric Linguistics Special issue: Aspects of speech studies II*, no. 10 (1), p. 13–34, 2019.

[6] G. Jefferson, "On a failed hyothesis:'conjunctional's as overlap-vulnerable'," *Tilburg papers in language and literature*, 1983.

[7] A. Rochet-Capellan and S. Fuchs, "Take a breath and take the turn: how breathing meets turns in spontaneous dialogue," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1658, p. 20130399, 2014.

[8] B. Malinowski, "The Problem of Meaning in Primitive Languages," in *The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism*, 4th ed. London: Kegan Paul, Trench, Trübner, 1936, pp. 296–336.

[9] S. Eggins and D. Slade, *Analysing casual conversation*. Equinox Publishing Ltd., 2004.

[10] K. P. Schneider, *Small talk: Analysing phatic discourse*. Hitzeroth Marburg, 1988, vol. 1.

[11] C. Cheepen, *The predictability of informal conversation*. Pinter London, 1988.

[12] R. Dunbar, *Grooming, gossip, and the evolution of language*. Harvard Univ Press, 1998.

[13] D. Abercrombie, *Problems and principles: Studies in the Teaching of English as a Second Language*. Longmans, Green, 1956.

[14] G. Brown and G. Yule, *Teaching the Spoken Language*. Cambridge University Press, 1983, vol. 2.

[15] S. Thornbury and D. Slade, *Conversation: From description to pedagogy*. Cambridge University Press, 2006.

[16] J. Laver, "Communicative Functions of Phatic Communion," in *Organization of behavior in face-to-face interaction*, A. Kendon, R. M. Harris, and M. R. Key, Eds. Oxford, England: Mouton, 1975, pp. 215–238.

[17] E. Ventola, "The structure of casual conversation in English," *Journal of Pragmatics*, vol. 3, no. 3, pp. 267–298, 1979.

[18] D. Slade, *The texture of casual conversation: A multidimensional interpretation*. Equinox, 2007.

[19] H. Watson, "The technology of respiratory inductive plethysmography," in *Proceeding of the Second International Symposium on Ambulatory Monitoring (ISAM 1979)*. Academic Press London, 1980, p. 537–563.

[20] K. Konno and J. Mead, "Measurement of the separate volume changes of rib cage and abdomen during breathing," *Journal of Applied Physiology*, vol. 22, no. 3, p. 407–422, 1967.

[21] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "Elan: a professional framework for multimodality research," in *Proceedings of LREC*, vol. 2006, 2006.

[22] V. H. Yngve, *On getting a word in edgewise*. Chicago Linguistic Society, 1970, p. 567–578.

[23] P. Boersma and D. Weenink, *Praat: doing phonetics by computer [Computer program], Version 5.1. 44*, 2010.

[24] H. Buschmeier and M. Wlodarczak, "Textgridtools: A textgrid processing and analysis toolkit for python," ser. Tagungsband der 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV 2013), 2013, p. 152–157.

[25] T. Noto, G. Zhou, S. Schuele, J. Templer, and C. Zelano, "Automated analysis of breathing waveforms using BreathMetrics: A respiratory signal processing toolbox," *Chemical Senses*, vol. 43, pp. 583–597, 2018.

[26] M. Włodarczak, "RespInPeace: Toolkit for processing respiratory belt data," in *Proceedings of Fonetik 2019*, M. Heldner, Ed., Stockholm, Sweden, 2019, pp. 115–118.

[27] E. Gilmartin and N. Campbell, "Capturing Chat: Annotation and Tools for Multiparty Casual Conversation." in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, N. C. C. Chair), K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds. Paris, France: European Language Resources Association (ELRA), 2016.