CHARACTERIZING THE COORDINATION OF SPEECH PRODUCTION AND BREATHING

Jeffrey E. Kallay[1], Ulrich Mayr[2], Melissa A. Redford[1]

[1]University of Oregon, Linguistics, [2]University of Oregon, Psychology
jkallay@uoregon.edu, mayr@uoregon.edu, redford@uoregon.edu

ABSTRACT

Previous studies have concluded that breath intake patterns during speech emerge as a function of planning processes. Little work has tested for similar effects of respiratory recovery on these patterns. Moreover, previous work has relied on one-by-one elicitation of read sentences which incorporates a direct cue to upcoming length, allowing for anticipatory effects to emerge but prohibiting a test of preceding material on intakes. The current study investigated the relative influences of recovery and anticipatory factors on breath intakes in a connected speech task that better approximates spontaneous production. Participants (N = 6) were asked to recite a passage of 20 unrelated sentences from memory. Results revealed a significant effect of preceding utterance length on presence of breath intakes during pauses, but not of following utterance length. It is concluded that respiratory recovery drives breath intakes in connected speech.

Keywords: speech production, breathing, speech planning, respiratory recovery.

1. INTRODUCTION

Speech and language researchers have long assumed that speech breathing patterns are governed by spoken language planning processes. As early as the 1960s, Henderson and colleagues [5] found that speakers took breaths overwhelmingly at grammatical junctures in both read and spontaneous speech. This notwithstanding that their more general pausing patterns differed greatly between the 2 speaking styles as a function of the relative cognitive demands of each. These early findings suggested that breath intakes during speech production are planned largely around syntactic structure, regardless of speech fluency. Many subsequent studies have demonstrated similar relations between syntactic structure and intake patterns, e.g. [3,6,9,10]. More recent work, however, has suggested instead that speech breathing patterns reflect speech motor planning. For example, Whalen and Kinsella-Shaw [8] asked speakers to read sentences of variable length and syntactic complexity. They found that breath intakes were longer in duration prior to longer sentences, while there was no effect of the syntactic structure. More recently, Fuchs and colleagues [2] asked subjects to read target sentences of variable length and complexity that were preceded by a carrier sentence designed to induce intermediate intake. Both the depths and durations of intakes were greater before longer sentences; again, there were no effects of syntactic complexity. These results suggest that it is the amount of speech to be produced rather than linguistic content which drives breath intake patterns while speaking.

Of course, there are reasons to question the evidence for planning effects on breathing. For one, the studies typically present written sentences one at a time to speakers thereby providing them with explicit, speech-external cues to sentence length. Moreover, a one-by-one elicitation method precludes the investigation of physiological effects on intake patterns. This is problematic because breath pauses are "physiological necessities" [7:54], which could suggest that breath intakes are governed by respiratory recovery rather than by planning. A recovery hypothesis predicts that the probability of an intake increases with the length of the preceding utterance rather than with the length of a subsequent utterance.

The current study was designed to provide a more stringent test of the planning hypothesis and to test the recovery hypothesis. Specifically, we investigated the effects of utterance length and duration on inter-utterance breath intakes in memorized passages of connected speech. The passages consisted of long and short sentences that were ordered to obtain a zero (lag-1) autocorrelation sequence. In this way, the elicitation method and materials removed external cues to sentence length. The resulting speech was also closer to spontaneously produced speech than that which is typically elicited in studies on speech breathing.

2. METHODS

2.1. Participants

Samples of memorized connected speech were collected from 6 healthy college-aged American English speakers (1M;5F). Participants were recruited through the Human Subjects Pool administered by the Psychology department at University of Oregon. All had a self-reported history of typical speech and language development.

2.2. Stimuli

Speech samples were elicited with a 20-sentence passage of meaningful but unrelated sentences. Half of the sentences were short (6 syllables) and the other half were long (12 syllables), with syntactic structure controlled. Words were limited to the top 1000 frequency in the Corpus of Contemporary American English (COCA) [1] to avoid frequency effects, and no content words were repeated. Sentence order was randomized to create a zero (lag-1) autocorrelation sequence.

To facilitate memorization, sentences were blocked into chunks with vertical lines on the page, and participants were invited to memorize the passage in those chunks. To mitigate possible effects of the chunking on intake patterns, half of the participants were given the passage in chunks of 4 sentences, and the other half in chunks of 5 sentences.

2.3. Procedures

Participants were told that their goal was to recite the passage as fluently as possible by the end of the study session. They were then given a sheet of paper containing the passage in paragraph form and left alone for 20 minutes to memorize it through rote learning. Subsequent to this, participants were asked to practice reciting the passage from memory for 15 minutes. Following each recitation, the experimenter provided feedback on any mistakes that were made. Following this, the speech samples analyzed in the current study were elicited. Participants were asked to recite the entire passage from memory 3 times in a row as naturally as possible. This was repeated 4 times, yielding 12 recitations per speaker. To facilitate fluency, prompts were provided on a computer screen as the first 2 words of each sentence. The presentation of prompts was self-paced by the participant.

2.4. Recordings and Segmentation

The speech samples were audio recorded with a Shure SM81-LC cardioid condenser microphone connected to a Marantz PMD660 digital audio recorder. The microphone was boom-mounted and placed directly in front of the participants' mouths.

The recordings were segmented into pause-delimited utterances by a trained researcher. Pauses were identified as clear visual (in a spectrogram) and audible breaks in the speech stream of no less than 100 milliseconds. An additional 50 milliseconds was added to this criterion for initial or unreleased final plosives at a pause boundary. Utterances were orthographically transcribed and coded for number of syllables spoken. Pauses were coded for visual and audible presence of breath intake, and intake duration was measured.

2.5. Analyses

Hierarchical logistic regression analyses were conducted on the effects of preceding and following utterance lengths on the presence of inter-utterance breath intakes for utterances that were 6 or 12 syllables long (i.e., complete sentences from the passage). Utterance length was entered as a predictor in the base models. Repetition number and block were then added in succession and the new models compared to the previous ones with ANOVA analysis. Hierarchical linear regression was used to assess the effects of the preceding and following sentence lengths on the durations of breath intakes, with repetition and block again added to the base models.

Hierarchical logistic regression was also used to assess the extent to which the presence or absence of inter-utterance breath intake was predictable from the lengths (in syllables) and durations (in msec.) of all preceding and following utterances. Length, duration, and their interaction (i.e., articulation rate) were entered as predictors in separate base models. Effects of repetition number and block were again added to the models in succession, and the new models compared to the preceding ones. Hierarchical linear regression was then used to investigate the effects of the preceding and following utterance lengths and durations on breath intake durations. Utterance length, duration, and their interaction were again entered as predictors in separate base models, along with repetition and block.

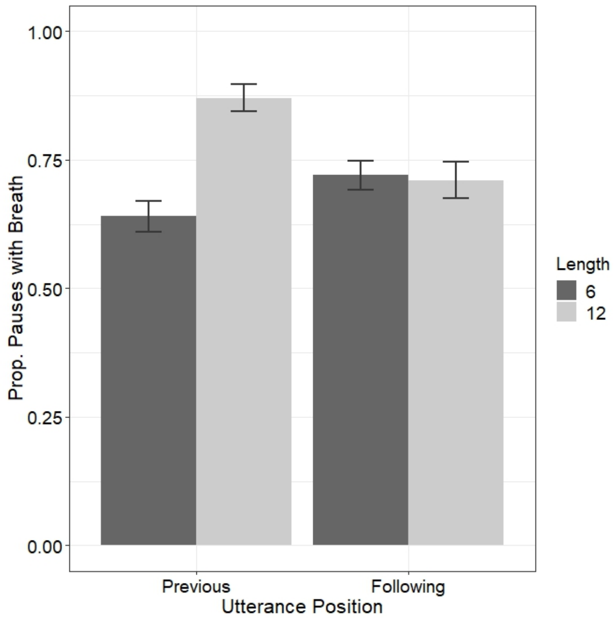
Each of the preceding analyses were also run as mixed effects models with random slopes and intercepts specified for speaker. None of the results differed in a meaningful way, suggesting that the effects reported here are not being driven by individual patterns of a subset of speakers. No significant effects of passage repetition or block were found in any of the analyses.

3. RESULTS

3.1. Breath intake by experimentally-defined sentence length

When looking at just the pause-delimited utterances which represented a single, complete sentence from the passage, there was a significant effect of preceding utterance length on the presence of breath intake during a pause ($z = 5.05$, $p < .001$). When the preceding utterance was 12 syllables long, speakers took a breath 87% of the time, compared to only 64% when it was 6 syllables long. There was not a similar effect of following utterance length, with breaths before 72% of 6-syllable and 71% of 12-syllable utterances. Figure 1 shows the proportions of pauses that contained a breath intake within each condition.

Figure 1: Proportion of pauses with breath intake by utterance position and length. Standard Errors of the proportions are represented by the error bars.



There was also a significant effect of preceding sentence length on the durations of breath intakes ($t = 2.46$, $p = .01$) but not of following sentence length. The mean intake duration was 351 msec. ($SD = 126$) when following 12-syllable sentences and only 315 msec. ($SD = 129$) when following 6-syllable sentences. The mean intake duration was also longer before 12-syllable sentences ($M = 333$ msec., $SD = 125$) than before 6-syllable sentences ($M = 301$ msec., $SD = 102$), but the difference between these was not as pronounced as between the preceding sentences of differing lengths.

3.2. Breath intake by speaker-defined utterance lengths and durations

Preceding utterance length also had a significant effect on the presence of breath intake during a pause ($z = 7.71$, $p < .001$) when breath intake patterns were analyzed with reference to speaker-imposed junctures on the memorized passage. As shown in Figure 2, all 6 speakers were more likely to take a breath following a longer utterance than following a shorter one. The mean utterance length before a breath intake was 9.25 syllables ($SD = 6.56$), and only 5.56 syllables ($SD = 4.52$) before a non-breath pause. There was no effect of preceding utterance duration on the likelihood of inter-utterance intake.

Figure 2: Absence or presence of breath intake during a pause by speaker and preceding utterance length.

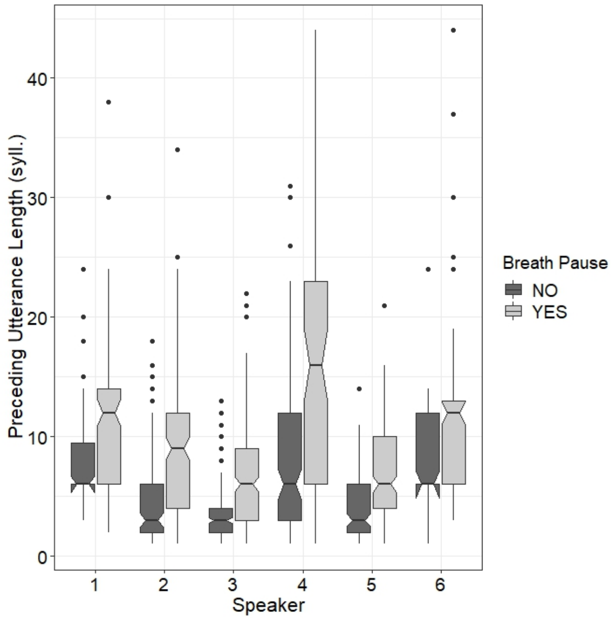
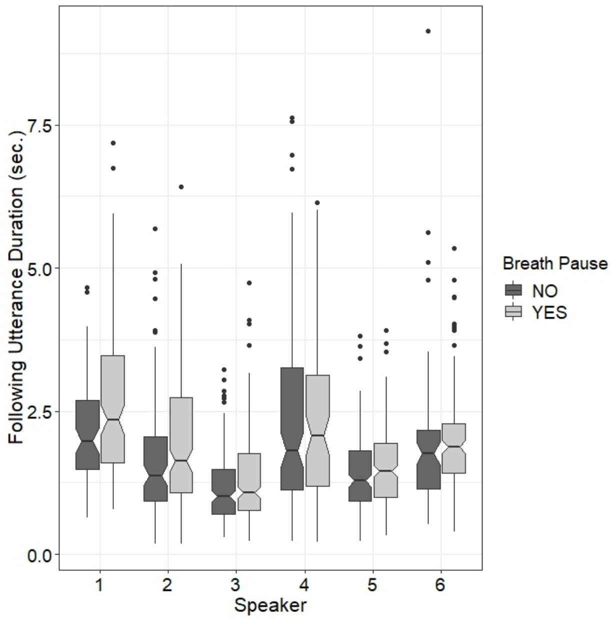


Figure 3: Absence or presence of breath intake during a pause by speaker and following utterance duration.



Following utterance duration also had a significant effect on the presence of breath intake during a pause ($z = 4.36$, $p < .001$). Figure 3 shows that all 6 speakers were more likely to take a breath before a longer utterance than a shorter one. The mean utterance duration following a breath intake was 1819 msec. (*SD* = 1077), and only 1676 msec. (*SD* = 1119) following a non-breath pause. There was no effect of following utterance length on the likelihood of inter-utterance breath intakes.

The interaction between utterance length and utterance duration (i.e. articulation rate) was also a significant predictor of breath intake for both preceding (z = -3.84, p < .001) and following utterances (z = -3.33, p < .001). All 6 speakers were more likely to take a breath following an utterance with a faster articulation rate than following a slower one. The mean articulation before a breath intake was 4.41 (*SD* = 1.22) syllables per second, and only 3.67 (*SD* = 1.32) syllables per second before a non-breath pause. The patterns for the articulation rate of the following utterance were less systematic across speakers.

There were no effects of preceding utterance length or duration on the durations of breath intakes, but there were significant effects of following utterance length (t = -4.32, p < .001) and duration (t = 3.80, p < .001). However, there was a large amount of variance overall among the intake durations (*range* = 64 - 1194 msec., *SD* = 157 msec.), and the patterns of the relationships across individual speakers were unsystematic.

4. DISCUSSION

At first glance, our findings suggest that both recovery and planning processes affect speech breath intake patterns. Both preceding utterance length and following utterance duration predicted intakes. However, closer inspection of the data points to respiratory recovery as the leading determinant of these patterns. Recall that previous work has found speakers to take breaths primarily at grammatical junctures, even in spontaneous connected speech. The effect of preceding utterance length but not of preceding utterance duration implies that speakers are taking advantage of pauses at these junctures as opportunities to breathe following production of a long string of syllables. This also fits the assertion that “breathing in speech is subservient to pausing, and not the other way around” [4:63].

Conversely, speakers were not planning intakes according to the number syllables to be produced, but rather were afforded a certain amount of time with which to produce more speech given the magnitude of the preceding intake. Thus the effect of following utterance duration but not of following utterance length. This conclusion is supported by the analyses based on 6- and 12-syllable utterances. While speakers were more likely to breathe following longer sentences than shorter ones, the following sentence length had no effect because intakes were not planned in anticipation of production units.

The clear effect of preceding utterance articulation rate on intakes is also unsurprising given the strong positive relationship between articulation rate and utterance length (t = 30.48, p < .001). And as previously concluded, “at fast rates… the physiological need to breathe is the sole determinant of pausing” [3:98].

Finally, strong conclusions on the relationship between utterance length and intake duration were elusive due to considerable variability in the latter. This may be because intake depth is not perfectly correlated with intake duration, and intake depth may be the more physiologically-relevant measure of speech breathing than intake duration. After all, breath intakes function to support the constant airflow that speech production requires, but long shallow breaths may not meet respiratory needs as well as long deep breaths. Unfortunately, the lack of a fixed mouth-to-mic distance in the present study precluded us from analyzing acoustic intensity, which could have provided some insight into intake depth.

In on-going work, we are coupling kinematic and acoustic measures to validate the present results and obtain a robust measure of inhalation depth and duration. We are also modifying the elicitation method to reduce the difficulty of the memorization task for participants. Instead of semantically-unrelated sentences, new passages have been constructed to create several chunks of text within which the sentences are thematically-related even while all other factors (e.g., frequency, syntax) are controlled. It is hoped that these new passages will allow participants to produce the passage without breaks that are clearly memory-related. The introduction of chunks will also allow us to investigate whether or not higher-level language structures influence breath intake patterns.

5. ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (NIH), award #R01HD087452.

6. REFERENCES

[1] Davies, M. 2008-present. *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present.* https://corpus.byu.edu/coca/

[2] Fuchs, S., Petrone, C., Krivokapic, J., Hoole, P. 2013. Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics* 41, 29-47.

[3] Grosjean, F., Collins, M. 1979. Breathing, pausing and reading. *Phonetica* 36(2), 98-114.

[4] Grosjean, F., Grosjean, L., Lane, H. 1979. The patterns of silence: Performance structures in sentence production. *Cognitive Psychology* 11, 58-81.
[5] Henderson, A., Goldman-Eisler, F., Skarbek, A. 1965. Temporal patterns of cognitive activity and breath control in speech. *Language and Speech* 8, 236-242.
[6] Huber, J.E., Darling, M., Francis, E.J., Zhang, D. 2012. Impact of typical aging and Parkinson’s disease on the relationship among breath pausing, syntax, and punctuation. *American Journal of Speech-Language Pathology* 21, 368-379.
[7] Vaissiere, J. 1983. Language-independent prosodic features. In: Cutler, A., Ladd, DR. (eds.), *Prosody: Models and measurements*. Heidelberg: Springer-Verlag, 53-66.
[8] Whalen, D.H., Kinsella-Shaw, J.M. 1997. Exploring the relationship of inspiration duration to utterance duration. *Phonetica* 54, 138-152.
[9] Winkworth, A.L., Davis, P.J., Adams, R.D., Ellis, E. 1995. Breathing patterns during spontaneous speech. *Journal of Speech, Language, and Hearing Research* 38, 124-144.
[10] Winkworth, A.L., Davis, P.J., Ellis, E., Adams, R.D. 1994. Variability and consistency in speech breathing during reading. *Journal of Speech, Language, and Hearing Research* 37, 535-556.