1062

Original Paper

Phonetica 1997;54:138–152

Received: May 28, 1996
Accepted: April 1, 1997

Exploring the Relationship of Inspiration Duration to Utterance Duration

D.H. Whalen, J.M. Kinsella-Shaw

Haskins Laboratories, New Haven, Conn., USA

Abstract

Previous work has indicated that there may be a positive relationship between the duration and extent of inspiration and the length of an upcoming utterance. However, none of that work has uniquely implied a role of planning. We attempted to avoid some of the alternative explanations by forcing subjects to utter single sentences ranging in length from 5 to 82 syllables (mean of 27), after inspiring fully and then expiring down to a set level before uttering the sentence. For all 3 subjects, there was a significant positive relationship between utterance length and inspiration duration, regardless of whether inspiration was measured physiologically or acoustically. The 2 subjects with the higher correlations in the articulatory measures also expended air more quickly during the shorter sentences than longer ones, while the other subject had no correlation with exhalation rate. Complexity of the sentence, calculated as the number of clauses in the sentence, did not affect inspiration duration. The individual differences need further investigation, but there is a positive correlation between the duration of the sentence to be said and the inspiration before it when the speaker is required to read sentences while using only one breath.

When we prepare to speak, we often need to draw a breath. Since most speech segments depend on a continued airflow for successful production, we can expect that utterances with more syllables will require more air to sustain them. A natural expectation, then, is that longer utterances should be preceded by longer inspirations, as has occasionally been asserted [e.g., Froeschels, 1931, cited in Weiss, 1967]. Certainly the amount of air available at the onset of speech should optimally be sufficient for the sentence(s) that a speaker would like to utter. Even that is not a prerequisite, since speakers inspire not only at sentence boundaries, but often at other syntactic boundaries [Grosjean and Collins, 1979; Sugito et al., 1990; Winkworth et al., 1994]. However, it is clear that we do, in the normal case, plan our speech, and having enough air to produce the speech would seem the most likely strategy. The sound of inspiration, then, would give some indication of speech planning, which might help the listener. There is some evidence that listeners use the sound created by the inspiration in the perception of speech, at least of synthetic speech [Whalen et al., 1995; Whalen and

KARGER

E-Mail karger@karger.ch
Fax +41 61 306 12 34
http://www.karger.ch

0031–8388/97/0544–0138$12.00/0

This article is accessible online at:
http://BioMedNet.com/karger

D.H. Whalen
Haskins Laboratories, 270 Crown Street, New Haven, CT 06511 (USA)
Tel. 203-865-6163, Fax 203-865-8963
E-Mail whalen@haskins.yale.edu

Sheffert, 1996, 1997]. But whether or not there are perceptual effects of the inspiration sound, duration may be an indication of a component of speech planning.

There are several reasons to think that there might not be a relationship between inspiration duration and utterance duration, however. It might be, for example, that speakers prefer to attain a particular level of lung volume before they begin to speak. In that case, the inspiration would only be related to the lung volume at the initiation of the inspiration, and the volume at the end of an utterance might show a correlation with utterance length. Even that correlation might fail to appear if speakers used higher air flows for shorter utterances, so that both the starting and ending volumes were near a single preferred level. In addition, virtually every long sentence has natural syntactic break points at which an additional inspiration would not be disruptive, so further air intake can occur in the middle of utterances. This might further weaken any correlation between utterance duration and inspiration duration.

It is also true that a correlation in running speech might not indicate planning on the part of the speaker. Since sentences can be uttered together, without being separated by an inspiration, it could be that a speaker would take a breath which would be more than enough for one sentence and then proceed to the next. A larger inspiration, then, might be followed by a longer utterance, but this might not indicate that the longer utterance was the one planned for. It might be that speakers would prefer to continue with (unplanned) excess capacity than to recharge and then continue.

Previous studies have sometimes found a correlation between inspiration duration and utterance length, but they do not unambiguously point to a planning component. Atkinson [1973] found a relationship between utterance duration and lung volume increment for several sentences. No formal statistics were performed though it was clear that a relationship existed. However, the range of durations was quite small, and there was no control over the initial lung volume. Some of the durational differences were due to differences in contrastive stress rather than differences in number of syllables (or segments) to be uttered. It is not clear whether this difference should be equated with a difference based on more typical durational changes.

Gelfer et al. [1983] also found a correlation between depth of inspiration and following utterance duration. However, as they point out, their long and short utterances were produced together in blocks, so the added depth could just as easily have been due to the preceding utterance as the following. There is some evidence (most obviously in their figure 9–3) that the duration of the inspiration is longer for the longest utterance compared with the two shorter ones. Winkworth et al. [1994] similarly found a correlation between utterance duration and inspiration for readings of the 'Rainbow Passage', but it was not possible to determine whether the preceding or the upcoming utterance was more relevant. In addition, since sentence boundaries were often not accompanied by an inspiration, it was not possible to determine whether the planning included all the sentences that were produced in one breath or if the sentences were opportunistically produced when enough lung capacity remained for them. Their spontaneous speech experiment [Winkworth et al., 1995] also showed a correlation of inspiration volume and utterance duration but had a similar ambiguity, since subjects often did not take a breath where one would have been syntactically allowed.

On the other hand, Wilder [1983] found a correlation between expiratory volume and sentence lenght (with durations of 10, 20 or 30 syllables), but this relationship was obtained not by increasing the initial volume (which might lead to longer inspirations) but by ending lower in the expiratory reserve volume. Recovery from a lower level would, in all likelihood, lead to longer inspirations, but Wilder [1983] does not report those numbers. However, the speech was continuous and the sentences blocked by duration. So it would not be possible to tell whether those numbers were influenced by the preceding or by the following utterance.

Sugito et al. [1990] also found a correlation between utterance duration and inspiration duration for a single speaker of Japanese. However, they report only the context preceding the inspiration, not that following. Grosjean and Collins [1979] found longer pauses at the end of longer constituents in the sentence (though it is not possible to tell from their report how long the actual utterance preceding the inspiration is). These studies would lead us to think that speech might be initiated at a set volume, with later inspirations occurring when a major syntactic break occurs and more lung volume is needed. There may be a correlation at both locations, of course.

The present study was designed to determine whether planning for the duration of the upcoming utterance is apparent in the duration of inspiration or not. To this end, we monitored lung volume for the reading of isolated sentences of widely varying duration. The inspiration demands were kept relatively constant by requiring the subjects to exhale to a particular level before beginning an utterance. Another aim of this study was to make a first pass at measuring whether the syntactic complexity as well as the duration of the sentence affected the inspiration duration. This was done by the simple metric of varying the number of clauses as well as the number of syllables. Although the task was far from a natural one, it is the only way of imposing the limitations needed to address just the question of whether initial inspiration duration is related to utterance duration.

The experiment reported here represents an attempt to avoid some of the alternative explanations that could have applied to the earlier studies. It thus avoided the use of connected speech, since that does not allow us to determine whether the breath is being planned for the upcoming speech or is a result of the preceding speech. We also attempted to start the subject's inspiration from a constant level, since the duration of the inspiration should be dependent on how much volume needs to be added.

Method

Stimuli

The sentences to be used needed to vary widely in length, but they also needed to be equally plausible. It would not be surprising if implausible sentences were preceded by longer inspirations, as speakers attempted to interpret the sentence well enough to give it a plausible reading. Indeed, Mitchell et al. [1996] find longer inspirations for discussions of more difficult topics compared to easier topics. (Mitchell et al. [1996] claim that there is no relation between difficulty and inspiration duration, since they adopted an alpha level of 0.01 for all effects in their study. The duration effect has a p value of 0.016, which we, in contrast, take to indicate a significant result.) Even sentences that are relatively plausible in isolation may be less so in the course of an experiment, in which they will come in succession and thus potentially be linked together as a kind of text. In addition, sample sentences generated by different authors will have differences in lexical selection, syntax, etc., that are difficult to control for. We felt that the best way of avoiding such problems was to select the sentences from an actual text written by one individual. This text could then be read beforehand so that the context of each sentence would be known, thus enhancing the plausibility of the individual sentences.

The sentences were selected from the first five chapters of *Hard Times,* by Charles Dickens. Limiting the sentences to the beginning of the book made it easier to sign up the subjects, since they did not have to read the entire book before participating. More importantly, the memory for the context that the sentences occurred in should be better for a shorter segment of the book than for the whole book.

Dickens uses a wide range of sentence durations, so it was easy to select the samples. We chose 48 sentences, ranging in length from 5 to 82 syllables (with a mean of 27; see 'Appendix'). Seven sentences were shortened by the removal of entire clauses. These sentences are indicated in the 'Appendix' by an apostrophe after the page number. The number of clauses varied from 1 to 7 with a mean of 1.8. It was not possible to fully cross the number of syllables and the number of clauses, of course, since it is impossible to get, say, 7 clauses in a 5-syllable sentence. But there was a fair amount of overlap, especially for one- and two-clause sentences (see 'Appendix'). For example, among the sentences with 13–23 syllables, there were 7 two-clause and 7 one-clause sentences. Since there were too many sentences for the subjects to memorize them all, the stimuli were presented in a readable format. Stimuli were printed in 48-point type, one to a page. These could then be set on a reading stand at a convenient distance from the subject. Since the wearing of glasses did not interfere with the apparatus used, our one subject who wore glasses was able to see the stimuli as easily as the other two.

Subjects

Three young native speakers of American English participated in the experiment. They were colleagues at Haskins Laboratories, and so were aware of the fact that we were researching speech. Once the task and apparatus were described, it was clear that speech breathing was the object. None of them, however, knew of our specific interest in inspiration duration. To avoid distortions in the measurements due to tissue compliance, we selected only slender males. The subjects had no reported speech problems. S1 had been treated for bronchial asthma until age 6. Otherwise, they reported no long-term breathing abnormalities. Additionally, none had ever been heavy smokers.

Apparatus

Lung volume was measured via the Respitrace® inductance plethysmograph (Ambulatory Monitoring Inc., Ardsley, N.Y., USA). This consists of two expandable bands containing transducers for measuring rib cage and abdomen cross-sectional area. The measurements from these transducers can then be used to assess overall lung volume. We calibrated the Respitrace via a Collins Survey III 8-liter spirometer. The spirometer itself was calibrated with a 3-liter syringe. The measurements from the spirometer were output simultaneously with those of the Respitrace for later comparison. The data channels and their sampling rates were as follows: speech audio (20 kHz), Respitrace abdomen, rib cage and sum (all 625 Hz) and spirometer (625 Hz). Each signal went through an antialias filter at the Nyquist rate and was digitized with 12 bits of resolution onto a VAX computer [Whalen et al., 1990].

Procedure

In the week preceding the experiment, each subject read the first five chapters of *Hard Times*. In this way, the context of the sentences was familiar to him. In each experimental run, the subject was first fitted into Respitrace apparatus. One band was placed across the chest, just below the armpits, and the other across the abdomen at the level of the umbilicus. Surgical tape was then applied at three locations for each band to prevent slippage. The subject was then seated in a chair with adjustable height and headrest. The height was adjusted so that the subject's feet rested on the ground. The headrest was put in a position where the subject could read the stimuli without moving his head. Once a comfortable position was attained, the subject was instructed not to move. After the 5-min period in which the Respitrace reached its internal calibration level, we performed a spirometer calibration. The subject breathed into the spirometer for at least three respiratory cycles, taking somewhat larger breaths than he would during normal conversation.

In order to be sure that the initial conditions were as similar as possible for each sentence, we controlled the breathing cycle immediately prior to the reading of the trial. We needed to have each trial preceded by a full inspiration (so that a good supply of oxygen would be available, and thus not requiring an inspiration for physiological reasons) followed by an expiration to a set level, so that the initial volume before the inspiration for speech was the same for each sentence. Therefore we had the subject perform several cycles of taking a full breath and expelling it. During those cycles, the experimenter was monitoring the volume measurement on an oscilloscope. At approximately 80% of the volume drop in that cycle, we selected a conveniently marked portion of the oscilloscope screen as the point at which the test trials would begin. This is the level that the subject would expel down to before taking a breath to utter the sentence. We chose the 80% point as one which would almost certainly require some inspiration before the sentence could be read, and yet which would not require extraordinary means to be reached.

Each trial had the following form: First the subject would inspire to a fairly full volume. Then an experimenter would place the next sentence on the reading stand. At that point, the subject would begin to exhale while preparing to utter the sentence printed. The second experimenter, meanwhile, was monitoring the lung volume to see when the criterial level was reached. Once it was reached, the second experimenter said 'go' loudly enough both to be heard by the subject and to be recorded on the audio channel of the input. Then the subject took as deep an inspiration as he thought necessary to say the sentence in a single breath. Misreadings were redone immediately, beginning with the full inspiration and expiration to the set level. Trials in which a second (or third) inspiration was taken were not redone.

The sentences were read once each, and in the order listed in the 'Appendix'. This is also the order they appear in the text itself. The length of the sentences was automatically randomized fairly well by this choice, but, more importantly, the context for each sentence followed naturally from the one before. None of the sentences had occurred more than 5 pages apart, and most were within half a page of the previous one. At the end of the session, the spirometer calibration was again performed.

Analysis

Since the subjects were instructed to produce the sentences in a single breath, those utterances with two inspirations can be considered to be errors. For the most part, we will focus on those utterances that were successfully produced, although we will also have some comparisons to an analysis with all utterances, to show how similar the two are.

Inspiration duration was measured both by analysis of the Respitrace signals and by acoustic/auditory determination. For both analyses, three points were based on the audio signal: the beginning of the 'go' signal, sentence onset and sentence offset. These we set by visual and auditory inspection. An example of these points, along with the physiological measurement points, can be seen in figure 1. Lung volume measurements were taken from the Respitrace 'sum' signal, which combines the rib cage and abdomen signals. The onset of the inspiration and its endpoint were selected via the peak/valley/plateau routine in the Haskins Laboratories program HADES [Rubin, 1995]. These routines measure inflection points in the signal that are above certain (experimenter-selected) noise levels. (Prior to analysis, the spirometer and Respitrace sum signals were smoothed by a 21-point triangular window.) When the signal is changing gradually (as with the initial exhalation and with the exhalation during the speech), the algorithm considers the signal to constitute a 'plateau'. The onset of the inspiration was taken as the right edge of the plateau formed by the exhalation before the utterance. The end of the inspiration was the first peak (a local maximum, taking the noise level into account) in lung volume after the inspiration began. If there were secondary inspirations during the utterance, they were marked in a similar fashion. Figure 1 also shows the onset of the inspiration after the utterance, although this measure was not used in the present study. Figure 2 shows an example of the acoustic measurements. This analysis was performed by the first author, using visual and auditory criteria. While it is possible to inhale quietly, there was, in fact, only one token (for S1) in which it was not possible to hear the inspiration. Measurements were typically easy to make; subjectively, onsets were 'hard' to locate in 36.1, 4.2 and 31.3% of the cases for subjects 1–3, respectively. Similarly, offsets were 'hard' in 48.9, 0, and 52.1% of the cases. Uncertainty about the location of the onset or the offset was typically on the order of 5 ms; the measurements reported are for a single location within that range.

Inspiration duration is the difference between the onset and offset of the inspiration (C-B in fig. 1 and 2). Utterance duration is the difference between the onset and offset of the sentence's acoustic realization (E-D in fig. 1). We also analyzed the slope of the expiration function, an indicator of the rate of expiration. This was calculated as the difference in lung volume at sentence onset (D) and offset (E) divided by the duration. The sentence reaction time is the difference between the 'go' signal and the onset of inspiration (D-A for both figures). The inspiration reaction time is the difference between the 'go' signal and the onset of inspiration (B-A for both).

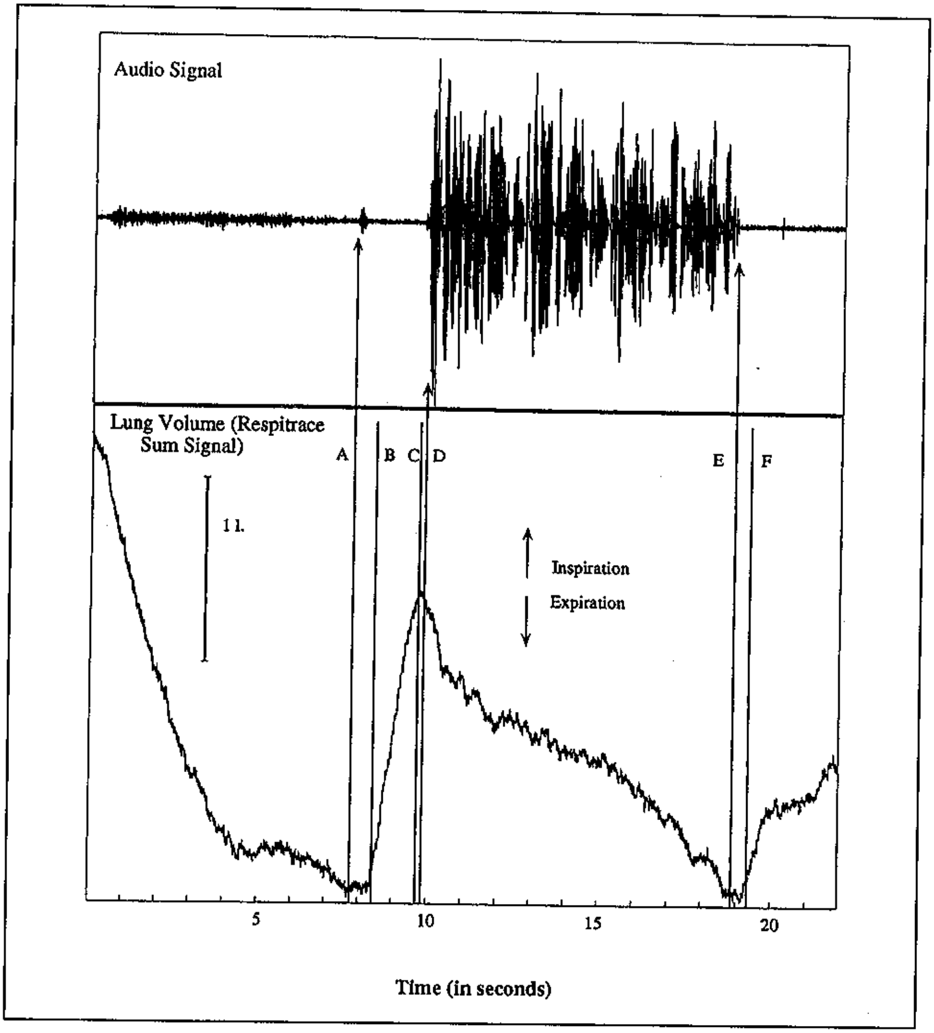

Fig. 1. Example of the speech audio signal (top panel) and the physiological measurements (bottom panel). The leftmost edge is the point at which the subject began exhaling down to the criterial level. The 'go' signal (A) is acoustically defined from the speech channel. Inspiration onset (B) is the edge of the 'plateau' formed by the exhalation. Peak volume (C) is defined as the local maximum in the volume signal. Sentence onset (D) is based on the speech signal, as is sentence offset (E). Inspiration after the sentence (F) was not analyzed in the present results.

Because there is the possibility that the volume increase is not completely correlated with duration, we also estimated the area under the curve of the inspiration. This was done by taking the sum of the difference between the magnitude of each of the samples in the inspiration and magnitude of the beginning sample. Since these values were always positive, it was not necessary to do any further processing to obtain a valid area measure. If subjects took different amounts of time for inspiring the same volume, we would want to interpret the duration results differently than if there was a constant rate of inspiration.

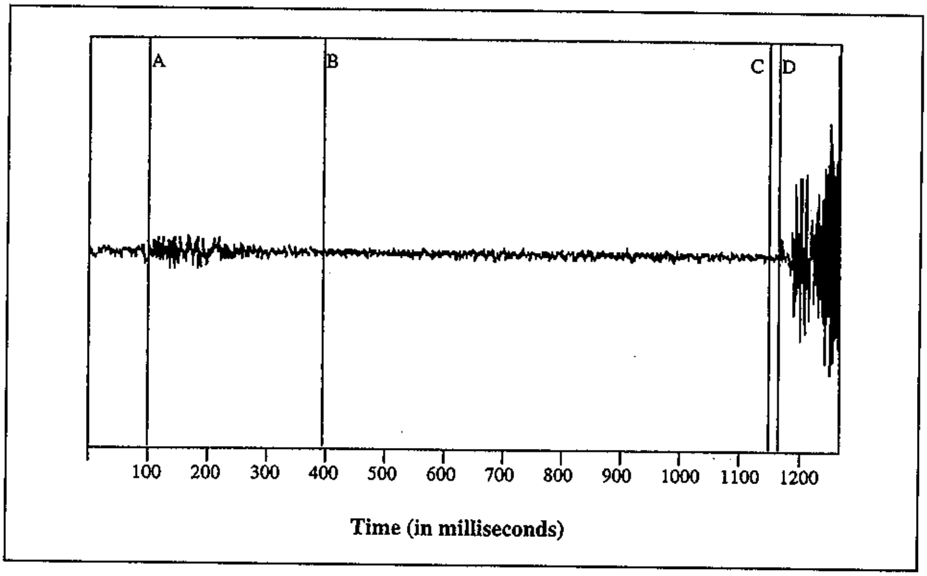

Fig. 2. Example of the acoustic measurement of the inspiration. The 'go' signal (A) is the same as for the articulatory measurement. Inspiration onset (B) and offset (C) were located both from the amplitude contour and by listening. Speech onset (D) is the same as for the articulatory measurement.

Results

The correlations between the spirometer signal and the Respitrace sum signal in the calibration conditions were quite high. For the pretest calibration, they were 0.98, 0.90 and 0.94 for S1, S2 and S3, respectively. For the post test calibration, they were 0.98, 0.93 and 0.92. Since the three breath cycles measured were slightly larger than those used in normal speech but not so exaggerated as a vital capacity measurement would have been, we can thus be fairly confident that the apparatus was recording accurately throughout the session.

Although it is quite common for speakers to inspire in the course of a sentence, we were interested in utterances produced with a single breath. If a second inspiration is made, it would not be clear whether the initial inspiration should be correlated with the duration of the utterance as a whole or only that portion up to the second inspiration. Except where noted, then, the following correlations will exclude the utterances with two or more breaths for all subjects. There were 7 such sentences for S1, 5 for S2 and 1 for S3.

For the subjects in this paradigm, duration and area under the curve were highly correlated (0.92, 0.98 and 0.90 for S1, S2, and S3, respectively). If we perform the analyses of interest using area under the curve rather than duration, slight differences in the magnitude of the correlations occur, but none made any material difference to the arguments presented here. Therefore, to simplify the presentation, the remaining correlations will use duration.

Table 1. Mean values of physiologically determined measurements

Subjects	Sentences n	Inspiration duration	SD	Range	Sentence duration	SD	Range	Duration per syllable
S1	48	707.0	258.6	371–2,030	2,879.4	2,049.5	543–8,729	106.7
	41	670.7	181.1	371–1,393	2,367.9	1,607.2	543–7,064	106.8
S2	48	541.0	206.1	249–1,084	2,888.3	2,022.3	481–8,410	107.1
	43	507.1	185.5	249–1,084	2,400.0	1,450.5	481–5,665	106.6
S3	48	1,071.5	245.4	507–1,446	3,041.8	2,075.9	662–9,227	112.8
	47	1,069.2	247.6	507–1,446	2,969.5	2,036.3	662–9,227	112.4

The first line for each subject includes all sentences, even those with two breaths. Since duration per syllable (last column) was calculated by dividing the duration of the sentence by the number of syllables, those values include the duration of those secondary breaths. All duration values are in milliseconds.

Table 2. Mean values of acoustically determined measurements

Subjects	Sentences n	Inspiration duration	SD	Range
S1	40	583.8	117.7	410–925
S2	43	572.2	183.5	350–1,104
S3	47	1,006.9	230.3	455–1,435

All duration values are in milliseconds.

Utterance duration can also be defined in two ways, either in terms of the number of syllables or the actual duration of the expressed sentence. For the present experiment, utterance duration and number of syllables were extremely well correlated (0.99, 0.98 and 0.99 for S1, S2, and S3, respectively). Although we can still assume that there were large differences among particular syllables, these differences seem to have been well balanced across the sentences. Only immaterial differences in the magnitude of the correlations for the factors of interest were obtained by using syllables rather than utterance duration, so all of the reported results will be in terms of actual spoken duration.

Table 1 presents the mean values for the pysiologically measured duration of the inspirations, along with the acoustically determined duration of the utterances and the average duration per syllable. The ranges are rather large, as expected, since we had a large range of sentence durations. Table 2 presents the duration of the inspirations as measured acoustically. We used correlations to determine whether there was a link between various of our measurements, as will be reported shortly. There was, however, a discrepancy between the physiological measurements and the acoustic ones that we will address first.

The temporal resolution on the physiological measurements appears to be lower than that of the acoustic measurements. Although the beginnings of the inspirations as marked by the HADES algorithm were clearly the earliest point at which it was justifiable to indicate an increase in volume, this point was often later than the audible onset of the inspiration. Acoustic onset averaged 53 ms before articulatory onset for S1, and 84 ms for S2; acoustic onset averaged 18 ms *after* articulatory onset for S3. The physiologically determined ends of the inspiration were also some tens of milliseconds beyond what was reasonable, since the peak often occurred after the onset of speech. For the 3 subjects, this happened in 36, 12, and 5 cases out of the 48 stimuli; the mean value for the peak was after the onset of the sentence for S1 (91.5 ms), while it preceded the onset for S2 (27.6 ms) and S3 (51.1 ms). Although it is possible to speak during inspiration, the acoustic output is quite different from normal, egressive phonation. The present stimuli were clearly normally produced, so that the volume measure must be somewhat inaccurate, either in the volume measure itself, or in the temporal alignment of that measurement. There is some temporal delay in the filtering process, but that is on the order of samples, not tens of milliseconds. It is possible that there is some inertial delay in the lungs themselves, or that auxiliary mechanisms effect changes in pressure at the extremes of the volume and/or at the change from inspiration to expiration. The bands of the Respitrace measure circumference at two points, and treat the lungs as a single cylinder. Thus the total volume will be accurate if there are no changes in the shape of the top and bottom of that cylinder. Such changes may in fact occur, leading to small changes in the actual volume that will not be reflected in the Respitrace signal. If both onset and offset have similar adjustments that result in similar delays, of course, the duration measure will be accurate. The high correlation between the acoustically and articulatorily determined durations suggests that such compensatory action may have occurred. The existing literature is of little help in resolving this dilemma: We can find no study which looked at time scales as narrow as this, and so it is unclear how general this phenomenon is. We report the articulatory values as the automatic algorithm generated them, even though they must contain some yet to be determined error in either or both of the temporal and volume domains.

Inspiration duration, measured physiologically, was positively correlated with duration of utterance. The values of the correlations across the 48 sentences, including those with more than one breath, were: S1, 0.56 ($p<0.001$), S2, 0.65 ($p<0.001$), S3, 0.29 ($p<0.05$). If we remove the multiple breath utterances, the values change to: S1, 0.53 ($p<0.001$), S2, 0.53 ($p<0.001$), S3, 0.29 ($p=0.052$). Thus after we control for initial lung volume and limit the utterance to a single sentence, we do find that subjects take a longer inspiration before longer utterances. The amount of variance accounted for (8–28%) is rather small, however.

The differences in the size of the correlation among the subjects may indicate a difference in strategy or respiratory organization. While the sample size of 3 is far too small to explore the generality of the differences, there are two other aspects of the results that support the existence of a true difference between S1 and S2 (the 'more highly correlated' subjects) on the one hand and S3 (the 'weakly correlated' subject) on the other. First, the same subsets of subjects patterned together in whether they showed different rates of exhalation (i.e., different slopes of the exhalation line) for different lengths of utterance. The two 'more highly correlated' subjects showed significantly shallower slopes for the longer sentences ($r=0.42$ and 0.44 respectively, $p<0.01$ for both). The 'weakly correlated' subject, however, showed no such correlation ($r=0.001$). Second, the 'more highly correlated' subjects inspired to higher levels for longer utterances, while the 'weakly correlated' one did not. The correlations were 0.45 ($p<0.001$) for S1, 0.53 ($p<0.001$) for S2, and 0.26 (n.s.) for S3. Inspection of the

Table 3. Mean reaction time (RT) measurements

Subjects	Sentences n	Sentence RT	SD	Phys. RT	SD	Acoust. RT	SD
S1	40	1,036.4	162.4	460.3	83.6	407.3	75.1
S2	43	1,134.8	263.0	600.1	149.5	516.5	148.5
S3	47	1,816.8	240.0	696.5	185.7	714.9	222.0

Sentence RT is the time from the 'go' signal to the start of the utterance. Phys. RT is the time from the 'go' signal to the onset of the breath, measured physiologically. Acoust. RT is the time from the 'go' signal to the onset of the breath, measured acoustically. All duration values are in milliseconds.

individual plots for the sentences produced by S3 suggested that he attained a lung volume from within a rather broadly defined range that would allow the sentence to be completed, and then he produced each sentence at the same rate of exhalation. There appears to have been a constraint that the longer sentences should not go too low in volume at the end, thus accounting for the correlation of inspiration duration and sentence duration. But this seems to have been a rather loose constraint, accounting for the small magnitude of the correlation.

All of the subjects tended to end at lower volumes for longer utterances ($r=0.77$, 0.82, 0.66 for S1, S2, and S3, respectively, $p<0.001$). Thus the larger volume inspired at the beginning of longer sentences was not sufficient to fully account for the difference in the requirements for longer versus shorter sentences.

The acoustic measures again show a correlation between utterance duration and inspiration duration, and with less indication of individual differences. The correlations of the two durations (excluding sentences with two or more inspirations and the one inaudible inspiration for S1) were 0.60, 0.58 and 0.50 ($p<0.001$) for S1, S2 and S3, respectively. Although the magnitude of the correlations was quite similar for the 3 subjcets with this measurement, S3 still had a clearly different pattern visually. For utterances less than 7.5 s in duration, the inspiration durations covered the full range, from longest to shortest. For longer sentences, only the upper half of the range was used.

Although this task was not strictly a reaction time paradigm, it is of some interest to look at the time taken between the 'go' signal and both the sentence and the inspiration (table 3). Time between the 'go' signal and the sentence was correlated with utterance duration (0.60, 0.71, and 0.59, $p<0.001$), as it must be if the inspiration duration itself is so correlated. There was also a longer delay between the signal and the onset of inspiration, at least for 1 subject, both in the articulatory (0.30, $p<0.10$, 0.60, $p<0.001$, and 0.24, $p=0.10$ for S1, S2 and S3, respectively) and the acoustic measurements (0.30, $p<0.10$, 0.52, $p<0.001$, and 0.18, n.s.). Since the speed of response was not emphasized in the instructions, these results must be treated with caution. But they suggest that there may even be a component of planning for the onset of inspiration that mirrors the planning aspects of the inspiration itself.

The possible influence of the number of clauses on the inspiration was studied via two ANOVAs, based on the physiological measurements. In the first, data from all 3 subjects were analyzed togheter, without regard to speaker identity. There was a single grouping factor, Clause, which was used for the single dependent variable, inspiration duration. Since the number of clauses was well correlated with the number of syllables ($r = 0.60$, $p < 0.001$ for the 48 sentences), Number of Syllables was used as a covariate. All utterances were used, whether or not they had a secondary breath. In this analysis, Clause was not significant [$F(1,137) < 1$, n.s.]. To study the effect without recourse to a covariate, we examined just those sentences with 13–23 syllables. As it turned out, there were 7 sentences with one clause and 7 with two clauses in that range. There again, the grouping factor Clause was used for the dependent variable inspiration duration, for all utterances by the 3 subjects. Clause was, again, not a significant factor [$F(1,40) < 1$, n.s.]. The number of clauses does not seem to influence the duration of the inspiration independently of the number of syllables in the sentence.

Discussion

Speakers take longer breaths before longer sentences. This relationship held for sentences uttered in isolation and beginning at a fixed volume of air. Although the correlation was present for all 3 subjects in this experiment, there appeared to be two strategies. One strategy, used by 2 of the speakers, not only had larger correlations between the size of the inspiration and the length of utterance but also two other features: slower rates of exhalation for longer sentences compared with shorter, and higher beginning volumes for longer sentences. The 3rd speaker, on the other hand, used rates of exhalation that were not correlated with the utterance length, and beginning volume was also independent of sentence duration. While the number of subjects is not large enough to draw firm conclusions, the co-occurrence of these three independent factors is indicative of a difference in strategy. All 3 speakers showed a negative correlation between length of utterance and final lung volume, indicating a willingness to use more of the vital capacity with longer sentences. There was no effect of the number of clauses in the sentences on inspiration duration independent of the number of syllables.

The present design avoided some of the alternative explanations that might be proposed for such a relationship. For example, if we look at connected speech, we cannot tell whether the inspiration is associated more with the preceding utterance or the following [Gelfer et al., 1983; Winkworth et al., 1994]. Similarly, if connected speech is studied and there is not an inspiration before each sentence (which is quite common), it is not possible to know whether the speaker intended to produce two sentences in one breath or simply continued from the first (planned) sentence to the second (unplanned) one when there was sufficient lung capacity left for its production. Long inspirations can be associated with the recovery from a preceding long sentence as well. Perhaps a paradigm in which subjects produce spontaneous speech but give some indication of when they have changed their mind about what they will say would help illuminate this issue for less constrained circumstances. Our paradigm was not particularly natural, since it is seldom the case that speakers will spontaneously decide to exhale to a low volume and then utter a single sentence. However, this constraint did allow us to specify how many syllables the speakers planned to say.

The predictive power of the duration of inspiration for upcoming utterance duration is not great. For 2 of the subjects, about 29% of the variance in utterance duration was accounted for by the correlation with inspiration as measured physiologically, while the 3rd subject's correlation accounted for only 8%. For the acoustic measurements, which might be more relevant to perception, the range was 25–36%. So, in more naturalistic speech, when it is not usually possible to determine what lung volume the speaker is starting with, the duration of the inspiration would not typically give a very solid indication of how long the speaker was planning to talk. Extreme cases can be found, at least for long inspirations followed by short utterances, in which the utterance length is completely unexpected. The total lack of inspirations in speech synthesis also seems to have some effect on its perception [Whalen et al., 1995].

The syntactic complexity of an utterance, as measured here by a simple count of the number of clauses, did not influence the duration of the inspiration. Like all negative results, this one must be interpreted lightly. This is especially true given the conditions of the experiment. The subject was reading the sentences, and so, in some senses, was not planning them. If, as seems likely, inspiration is indicative of a planning stage for the utterance's content as well as its duration, then we might expect to see an effect for spontaneous sentences that begin from a particular lung volume. With a large enough sample, for example, we could compare only those utterances that began at one volume or another. Such an approach would also need to be restricted, though, to utterances comprising a single sentence, since second sentences may not have been planned at the time of the inspiration. This would require an extremely large amount of data. In the present experiment, it was also the case that the context of the sentences was well known, by design, and we might expect that the degree to which a sentence is supported by its context would affect the size of the inspiration. Less supported sentences would probably take longer to produce [Hunnicutt, 1985], and so might require longer inspirations. On the linguistic side, clause count is not necessarily the best indicator of complexity. Other factors such as degree of syntactic branching, number and type of dependencies, etc., might equally well play a role. And, finally, it may truly be the case that syntactic complexity does not affect inspiration duration. Other studies will have to be designed to further our understanding of that relationship.

There are presumably many factors that affect the duration of the inspiration. The number of syllables has been shown to be an important one, intuitively enough. Another is the loudness at which the speaker intends to speak [Draper et al., 1959; Hixon et al., 1973]. There might also be differences depending on differences in airflow for the segments in the upcoming utterance, such as the higher flow for low vowels compared with high vowels [Atkinson, 1973] or (for some conditions) voiceless fricatives compared with voiced stops [Gelfer, 1987; Russell and Stathopoulos, 1988]. Inspiration may be used on occasion as a temporal place-keeper while the details of the utterance are worked out. Different speakers may organize their intent to speak at different parts of their breath cycle, leading to differences in how much air must be taken in [Schönle et al., 1988]. Whatever else may go into the determination of inspiration duration, the present results clearly show it to depend, when starting from a comparable lung volume, on the duration of the upcoming utterance.

Appendix

Sentences used in the experiment. The order in which the sentence occurred, in both the text and the experiment, is indicated in the number at the left. The next number is the number of syllables in the sentence, followed by the number of clauses. Page numbers, in brackets at the end of the sentence, refer to the Signet Classic edition of *Hard Times*, by Charles Dickens. An apostrophe with the page number indicates that the sentence has been shortened, usually by removing one clause.

1.	15	1	The scene was a plain, bare, monotonous vault of a school-room [p. 11].
2.	24	2	The emphasis was helped by the speaker's voice, which was inflexible, dry and dictatorial [p. 11].
3.	22	1	The speaker's obstinate carriage, square coat, square legs and square shoulders all helped the emphasis [p. 11'].
4.	15	1	He seemed a kind of cannon loaded to the muzzle with facts [p. 13'].
5.	15	2	It was his school, and he intended it to be a model [p. 18].
6.	19	2	Almost as soon as they could run, they had been made to run to the lecture-room [p. 18].
7.	45	5	No little Gradgrind had ever associated a cow in a field with that famous cow with the crumpled horn who tossed the dog who worried the cat who killed the rat who ate the malt [p. 19'].
8.	29	1	It had only been introduced to a cow as a graminivorous ruminating quadruped with several stomachs [p. 19].
9.	38	1	They had a little conchological cabinet, and a little metallurgical cabinet, and a little mineralogical cabinet [p. 19].
10.	36	2	He was an affectionate father, after his manner, but he would probably have described himself as 'an eminently practical' father [p. 20'].
11.	36	2	The clashing and banging band attached to the horse-riding establishment which had there set up its rest in a wooden pavilion was in full bray [p. 20].
12.	51	2	Thomas Gradgrind took no heed of these trivialities of course, but passed on as a practical man ought to pass on, either brushing the noisy insects from his thought, or consigning them to the House of Correction [p. 21].
13.	6	1	This brought him to a stop [p. 21].
14.	8	1	Both rose, red and disconcerted [p. 21].
15.	22	1	There was an air of jaded sullenness in them both, and particularly in the girl [p. 22].
16.	53	2	Yet, struggling through the dissatisfaction of her face, there was a light with nothing to rest upon, a fire with nothing to burn, a starved imagination keeping life in itself somehow, which brightened its expression [p. 22].
17.	17	1	He was a rich man, banker, merchant, manufacturer, and what not [p. 23].
18.	54	1	In the formal drawing-room of Stone Lodge, standing on the hearth-rug, warming himself before the fire, Mr. Bounderby delivered some observations to Mrs. Gradgrind on the circumstance of its being his birthday [p. 24].
19.	8	1	I hadn't a shoe to my foot [p. 24].
20.	33	2	Mrs. Gradgrind, a little, thin, white, pink-eyed bundle of shawls, of surpassing feebleness, mental and bodily, hope it was a dry ditch [p. 24'].
21.	10	1	Mrs. Gradgrind faintly looked at the tongs [p. 24'].
22.	14	2	Mrs. Gradgrind, stunned as usual, collapsed and gave it up.
23.	10	1	My mother left me to my grandmother [p. 25].
24.	55	2	Mrs. Gradgrind, weakly smiling, and giving no other sign of vitality, looked (as she always did) like an indifferently executed transparency of a small female figure, without enough light behind it [p. 25].
25.	55	1	His pride in having at any time of his life achieved such a great social distinction as to be a nuisance, an encumbrance, and a pest, was only to be satisfied by three sonorous repetitions of the boast [p. 25].
26.	8	1	We were peeping at the circus [p. 26].
27.	21	2	I declare you're enough to make one regret ever having had a family at all [p. 26].

28.	18	1	Mr. Gradgrind did not seem favourably impressed by these cogent remarks [p. 26].
29.	6	1	He frowned impatiently [p. 26].
30.	41	2	Mrs. Gradgrind was not a scientific character, and usually dismissed her children to their studies with this general injunction top choose their pursuit [p. 27].
31.	52	1	The simple circumstance of being left alone with her husband and Mr. Bounderby was sufficient to stun this admirable lady again without collision between herself and any other fact [p. 27].
32.	14	2	So she once more died away, and nobody minded her [p. 27].
33.	66	7	The girl wanted to come to the school, and Mr. Gradgrind wanted girls to come to the school, and Louisa and Thomas both said that the girl wanted to come, and that Mr. Gradgrind wanted girls to come, and how was it possible to contradict them when such was the fact [p. 29].
34.	8	1	I am much of your opinion [p. 29].
35.	29	4	He went his way, but she stood on the same spot, rubbing the cheek he had kissed with her handkerchief until it was burning red [p. 30].
36.	12	1	She was still doing this five minutes afterward.
37.	25	2	It was a town of red brick, or of brick that would have been red if the smoke and ashes had allowed it [p. 30].
38.	40	3	It was a town of machinery and tall chimneys, out of which interminable serpents of smoke trailed themselves forever and ever, and never got uncoiled [p. 30].
39.	82	4	It had a black canal in it, and a river that ran purple with ill-smelling dye, and vast piles of building full of windows where there was a rattling and a trembling all day long, and where the piston of the steam-engine worked monotonously up and down like the head of an elephant in a state of melancholy madness [p. 31].
40.	15	2	You saw nothing in Coketown but what was severely workful [p. 31].
41.	13	2	This man lives at Pod's End, and I dont't quite know Pod's End [p. 33].
42.	11	1	So they stopped for a moment, looking about [p. 33].
43.	72	4	The question was unexpectedly and suddenly answered for her by the colourless boy Bitzer, who came round the corner with such blind speed and so little anticipating a stoppage on the pavement that he brought himself up against Mr. Gradgrind's waistcoat and rebounded into the road [p. 34].
44.	32	4	Bitzer picked up his cap, which the concussion had knocked off, and backing, and knuckling his forehead, pleated that it was an accident [p. 34].
45.	5	1	They're famous for it [p. 34].
46.	11	1	He frightened me so with his cruel faces [p. 34'].
47.	9	1	They bruise themselves very bad sometimes [p. 35].
48.	15	1	She glanced up at his face, with mingled astonishment and dread [p. 35].

Acknowledgments

This research was supported by NIH grant HD-01994 to Haskins Laboratories. We are grateful to E.R. Wiley and Anders Löfqvist for technical assistance. This research was also presented at the 128th Meeting of the Acoustical Society of America, December, 1994, Austin, Tex. We thank Elaine Stathopoulos, Vincent Gracco, Richard McGowan, and Katherine S. Harris for helpful comments.

References

Atkinson, J.: Aspects of intonation in speech: implications from an experimental study of fundamental frequency; PhD thesis, University of Connecticut, Storrs (unpublished, 1973).

Draper, M.H.; Ladefoged, P.; Whitteridge, D.: Respiratory muscles in speech. J. Speech Hear. Res. *2:* 16–27 (1959).

Froeschels, E.: Lehrbuch der Sprach- und Stimmstörungen (Deuticke, Wien 1931).

Gelfer, C.E.: A simultaneous physiological and acoustic study of fundamental frequency declination; PhD thesis, City University of New York (unpublished, 1987).

Gelfer, C.E.; Harris, K.S.; Collier, R.; Baer, T.: Is declination actively controlled?; in Titze, Scherer, Vocal fold physiology: biomechanics, acoustics and phonatory control, pp. 113–126 (Denver Center for the Performing Arts, Denver 1983).

Grosjean, F.; Collins, M.: Brathing, pausing and reading. Phonetica *36:* 98–114 (1979).

Hixon, T.J.; Goldman, M.D.; Mead, J.: Kinematics of the chest wall during speech productions: volume displacements of the rib cage, abdomen, and lung. J. Speech Hear. Res. *16:* 78–115 (1973).

Hunnicutt, S.: Intelligibility versus redundancy – conditions of dependency. Lang. Speech *28:* 47–56 (1985).

Mitchell, H.L.; Hoit, J.D.; Watson, P.J.: Cognitive-linguistic demands and speech breathing. J. Speech Hear. Res. *39:* 93–104 (1996).

Rubin, P.E.: HADES: A case study of the development of a signal analysis system; in Syrdal, Bennett, Greenspan, Applied speech technology, pp. 501–520 (CRC, Boca Raton 1995).

Russell, N.K.; Stathopoulos, E.: Lung volume changes in children and adults during speech production. J. Speech Hear. Res. *31:* 146–155 (1988).

Schönle, P.W.; Machetanz, J.; Conrad, B.: Der Einfluss der Atmung auf die sprechmotorische Reaktionszeit. Folia phoniat. *40:* 123–127 (1988).

Sugito, M.; Ohyama, G.; Hirose, H.: A preliminary study on pauses and breaths in reading speech materials. Annu. Bull. RILP *24:* 121–130 (1990).

Weiss, D.A.: Cluttering. Folia phoniat. *19:* 233–263 (1967).

Whalen, D.H.; Hoequist, C.E.; Sheffert, S.M.: The effects of breath sounds on the perception of synthetic speech. J. acoust. Soc. Am. *97:* 3147–3153 (1995).

Whalen, D.H.; Sheffert, S.: Perceptual use of vowel and speaker information in breath sounds; in Bunnell, Idsardi, Proceedings ICSLP 96, pp. 2494–2497 (Citation Delaware, New Castle 1996).

Whalen, D.H.; Sheffert, S.: Normalization of vowels by breath sounds; in Johnson, Mullennix, Talker variability in speech processing, pp. 133–144 (Academic Press, San Diego 1997).

Whalen, D.H.; Wiley, E.R.; Rubin, P.E.; Cooper, F.S.: The Haskins Laboratories' pulse code modulation (PCM) system. Behav. Res. Meth. Instrum. Comput. *22:* 550–559 (1990).

Wilder, C.N.; Chest wall preparation for phonation in female speakers; in Bless, Abbs, Vocal fold physiology: contemporary research and clinical issues, pp. 109–123 (College-Hill Press, San Diego 1983).

Winkworth, A.L.; Davis, P.J.; Adams, R.D.; Ellis, E.: Breathing patterns during spontaneous speech. J. Speech Hear. Res. *38:* 124–144 (1995).

Winkworth, A.L.; Davis, P.J.; Ellis, E.; Adams, R.D.: Variability and consistency in speech breathing during reading: lung volumes, speech intensity, and linguistic factors. J. Speech Hear. Res. *37:* 535–556 (1994).