Folia Phoniatrica
et Logopaedica

Breath Group Analysis for Reading and Spontaneous Speech in Healthy Adults

Yu-Tsai Wang[a]   Jordan R. Green[b]   Ignatius S.B. Nip[c]   Ray D. Kent[d]
Jane Finley Kent[d]

[a]School of Dentistry, National Yang-Ming University, Taipei, Taiwan; [b]Department of Special Education and Communication Disorders, University of Nebraska-Lincoln, Lincoln, Nebr., [c]School of Speech, Language, and Hearing Sciences, San Diego State University, San Diego, Calif., and [d]Waisman Center, University of Wisconsin-Madison, Madison, Wisc., USA

Abstract

Aims: The breath group can serve as a functional unit to define temporal and fundamental frequency (f_0) features in continuous speech. These features of the breath group are determined by the physiologic, linguistic, and cognitive demands of communication. Reading and spontaneous speech are two speaking tasks that vary in these demands and are commonly used to evaluate speech performance for research and clinical applications. The purpose of this study is to examine differences between reading and spontaneous speech in the temporal and f_0 aspects of their breath groups. *Methods:* Sixteen participants read two passages and answered six questions while wearing a circumferentially vented mask connected to a pneumotach. The aerodynamic signal was used to identify inspiratory locations. The audio signal was used to analyze task differences in breath group structure, including temporal and f_0 components. *Results:* The main findings were that spontaneous speech task exhibited significantly more grammatically inappropriate breath group locations and longer breath group duration than did the passage reading task. *Conclusion:* The task differences in the percentage of grammatically inadequate breath group locations and in breath group duration for healthy adult speakers partly explain the differences in cognitive-linguistic load between the passage reading and spontaneous speech.

Copyright © 2010 S. Karger AG, Basel

Introduction

The respiratory system provides an aerodynamic source of energy and maintains a roughly constant subglottal air pressure during speech production through fairly precise, ongoing control of the respiratory musculature [1, 2]. Speech is structured in terms of breath groups based on the patterns of airflow from the lungs [3]. The features of breath groups are governed not only by respiratory needs, but also by the varying demands of grammatical structure [4]. Because the location and durations of breath groups are determined by physiologic needs, linguistic accommodations, and cognitive demands, these features may differ across speaking tasks such as passage reading and spontaneous speech.

Characteristics of nonspeech and speech breathing for reading and spontaneous speech in healthy speakers in

Yu-Tsai Wang
School of Dentistry, Room 411, National Yang-Ming University
No. 155, Sec. 2, Li-Nong St. Beitou District
Taipei 11221 (Taiwan)
Tel. +886 2 2826 7237, Fax +886 2 2826 4053, E-Mail yutsaiwang@ym.edu.tw

different age-groups and gender have been reported [5–13], but none of these studies reported fundamental frequency (f_0) features within breath groups. The breath group has been proposed as a useful functional unit of prosodic analysis, helping to define temporal and f_0 features for connected speech [14], especially because these features are determined by locations of inspiration. Inspiratory locations usually precede linguistic structural boundaries following grammatical rules; however, inspirations at grammatically inappropriate loci in utterances sometimes occur even for healthy speakers [5, 6, 12–14]. Bunton [5] reported a 19% occurrence of inappropriate breath locations for normal extemporaneous speech for 3 aged men and 3 aged women. Hammen and Yorkston [6] reported a 2.1% occurrence of inappropriate inspiratory locations for reading passages for 22 females and 2 males. Winkworth et al. [12, 13] reported 3.2 and 15.3% occurrences of inappropriate inspiratory locations for reading passages and spontaneous speech, respectively, for 6 healthy young women.

The temporal features of breath groups are mainly described in terms of breath group duration (BGD), inter-breath-group pause (IBP), and inspiratory duration (ID). Statistics on these parameters portray the basic ventilatory pattern of speech. For example, average BGD values range from 3.36 [13] to 3.58 s [11] for reading and from 2.42 [5] to 3.84 s [12] for spontaneous speech. The ID value for reading is 0.59 s [11]. But the full understanding of how these temporal measures vary with speaking task awaits systematic investigation with suitably sensitive methods.

Determining how reading and spontaneous speech tasks differentially affect breath group organization is important because they are often an integral part of the clinical assessment battery used to evaluate dysarthria and other disorders of speech and voice. In addition, the understanding of breath group patterning is important to the improvement of naturalness in speech synthesis [15]. More generally, breath group organization contains a rich source of segmental and prosodic cues used by listeners to perceive and comprehend speech [16]. Proper intonational variations within the breath group provide listeners with cues about linguistic structure [17].

The current investigation extends extant speech breathing studies by examining task differences on temporal and f_0 parameters in a relatively larger number of healthy adult talkers based on aerodynamically determined inspiratory loci.

The purposes of this study were (1) to compare the occurrence of inappropriate inspiratory locations between passage reading and spontaneous speech, and (2) to analyze temporal and f_0 patterns of speech breathing, including BGD, IBP, ID, and mean of the f_0 within the breath group (mean f_0), maximum of the f_0 within the breath group (max f_0), and range of the f_0 within the breath group (range f_0) between passage reading and spontaneous speech in normal adult speech based on actual inspiratory locations determined aerodynamically.

Methods

Participants
Participants were 16 healthy adults (6 males, 10 females), aged 20–64 years (mean: 40.3 years; standard deviation: 14.8 years). Participants were native speakers of North American English with no reported speech and language disorders. Participants had adequate auditory, visual, language and cognitive skills to read passages and answer questions.

Stimuli
Speech samples, including the 'Bamboo' [18] and 'Grandfather' passages [19] and spontaneous speech, were obtained from each participant. The first task involved reading of the 'Bamboo' and 'Grandfather' passages at a comfortable speaking rate and loudness. The 'Bamboo' passage was designed to maximize the number of voiced consonants at word and phrase boundaries so that pauses in speech can easily be identified. To obtain spontaneous speech samples, participants were then asked to talk about the following six topics in as much detail as possible: their family, activities in an average day, their favorite activities, what they do for enjoyment, and their plans for their future. Each answer was at least 1 min in length and consisted of at least 6 breath groups (as monitored by an airflow transducer). Participants were given time to familiarize themselves with the passages and to formulate answers to a question before the recording was initiated.

Experimental Protocol
Participants were seated and were instructed to hold a circumferentially vented mask (Glottal Enterprises MA-1L) tightly against their face. The mask was coupled to an airflow transducer (Biopac SS11lA), which was used to continuously record expiratory and inspiratory flows during the speaking tasks. The face-mask was reported not to affect the breathing patterns [20]. A professional microphone (Sennheiser) was placed approximately 2–4 cm away from the vented mask. The speaking tasks were presented via PowerPoint on a large screen using an LCD projector (ViewSonic PJ501). Participants were video-recorded using a Canon XL-1s digital video recorder. Video was sampled using Microsoft Windows Movie Maker. Audio signals were recorded at 48 kHz, 16-bit signal with the video. For each video recording, Adobe Audition 1.5 was used to separate the audio signal from the video signal, so that the audio signal could be used for the analysis of breath group structure.

The audio signal and the output signals from the airflow transducer were recorded simultaneously using Biopac Student Lab 3.6.7. Airflow was sampled at 1,000 Hz and low-pass-filtered at 500 Hz. This signal was subsequently used for the identification
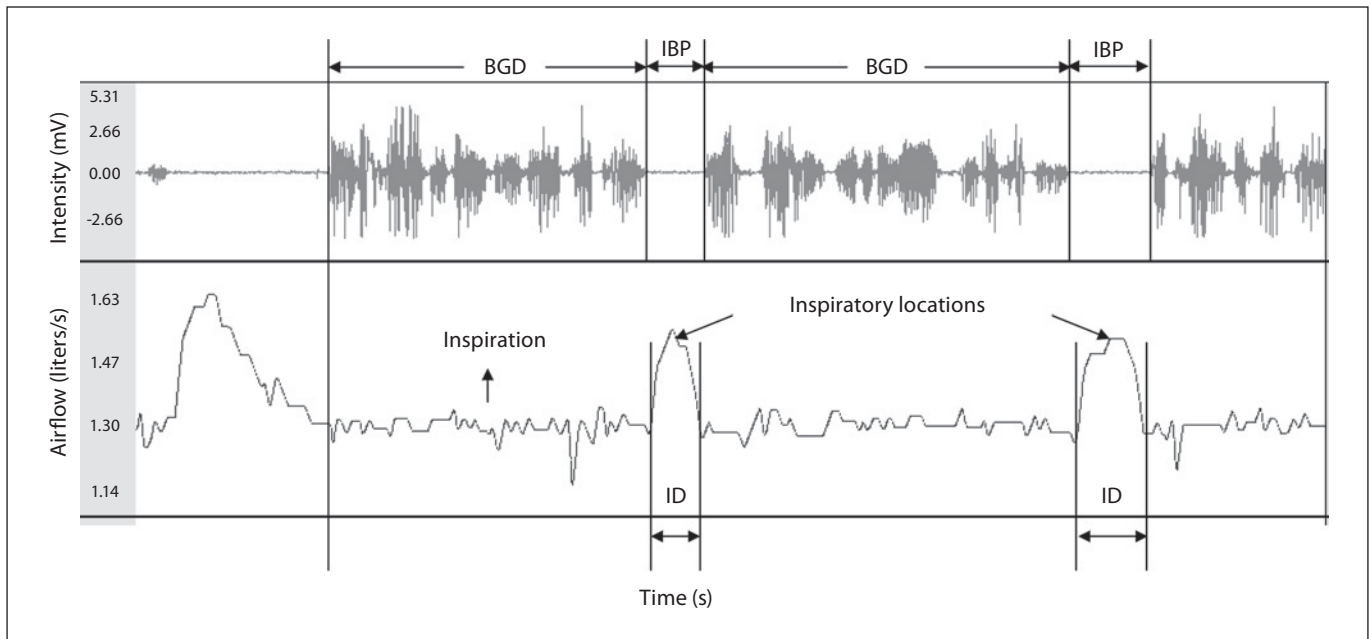
Fig. 1. A demonstration of measures of BGD, IBP and ID based on acoustic and aerodynamic signals. The arrows indicate the locations of inspiration for the Bamboo passage.

of actual inspiratory loci. An experimenter marked all the onsets of a new breath on each airflow signal, as indicated by an easily identified peak in the trace (fig. 1). The total numbers of inspirations determined from the airflow signals were 273 and 1,106 for passage reading and spontaneous speech, respectively.

Appropriateness of Inspiratory Locations

The appropriateness of inspiratory locations for the passage reading and spontaneous speech samples was determined by a judge with training in linguistics based on the rules given by Henderson et al. [21]. Inspirations locating at the end of a sentence or punctuation points such as comma or colon, or before noun, verb, adverbial phrases or other phrases are considered appropriate. Inspirations occurring within phrases or words are considered syntactically inappropriate. The percentage of appropriate breath group loci was calculated to compare the appropriateness of inspiratory locations between the passage reading and the spontaneous speech tasks.

Figure 1 shows inspiratory locations and measures of breath group structure based on acoustic and airflow signals. Top and bottom panels represent waveform and airflow signal from Biopac, respectively. For all tasks, the first BGD was not included in the analysis because the timing patterns associated with the first part of each utterance were expected to be variable and, therefore, nonrepresentative.

Temporal Components

As shown in figure 1, inspiratory locations were used to segment acoustic signals into BGD and IBP. BGD in this study was defined as the duration of groups of speech events produced on a single breath [3], and was measured from the start to the end of the speech signal produced on a breath group based on the acoustic waveform. IBP was measured as the interval between successive BGDs. ID was measured manually between the nearest minima on both sides of each inspiration and indicates actual inspiratory behavior for each IBP.

f_0 Components

After the temporal breath group parameters had been measured, a pitch trace was generated with TF32 [22] for each breath group sample. When the pitch tracking algorithm generated errors, the raw f_0 trace was corrected manually using TF32 software [22], most frequently required to delete erroneous f_0 trace occurring on stop bursts or noise signals and to add a portion of the f_0 trace on which phonation occurred but without f_0 trace, as previously reported [14]. The manually corrected f_0 traces within each breath group sample were used to obtain measures of mean f_0, max f_0, and range f_0 (maximum f_0–minimum f_0).

Measurement Agreement

To estimate intra- and interanalyst measurement agreement, the first author and another individual with experience in acoustic measurement remeasured acoustic data produced by 2 randomly selected participants (12.5% of the entire data corpus). These measurements were taken for both the passage reading and spontaneous speech samples approximately 2 months after completion of the first measures. The Pearson correlation coefficient of BGD between the two measures was 0.99 for intra-analyst and 0.99 for interanalyst. The Pearson correlation coefficient of IBP between the two measures was 0.99 for intra-analyst and 0.99 for

Table 1. Means and standard deviations for BGD, IBP, ID, f_0 mean, f_0 max, and f_0 range in the reading and spontaneous speech samples

Parameter	Passage reading	Spontaneous speech	t(15)	p
BGD, s	3.50 ± 0.62	4.35 ± 0.72	−3.85	0.002
IBP, s	0.65 ± 0.16	0.70 ± 0.12	−1.09	0.295
ID, s	0.55 ± 0.12	0.58 ± 0.08	−1.09	0.295
f_0 mean, Hz				
Male	118 ± 12	112 ± 11	1.93	0.073
Female	186 ± 24	184 ± 243		
f_0 max, Hz				
Male	169 ± 15	166 ± 17	−1.06	0.304
Female	269 ± 37	277 ± 35		
f_0 range, Hz				
Male	99 ± 10	97 ± 14	0.31	0.758
Female	197 ± 33	196 ± 35		

interanalyst. The mean absolute difference between the two measures was 11.9 ms for intra-analyst and 13.2 ms for interanalyst in BGD; 11.6 ms for intra-analyst and 12.7 ms for interanalyst in IBP, respectively.

Statistical Analysis
χ^2 test was used to analyze task differences in the appropriateness of inspiratory locations. Paired t tests were performed for task differences in temporal parameters (including BGD, IBP, and ID) and f_0 parameters (including f_0 mean, f_0 max, and f_0 range) of breath group structure at $\alpha = 0.05$ level.

Results

Appropriateness of Inspiratory Loci
The number of inappropriate breathing locations was 5 out of 273 (1.8%) and 143 out of 1,106 (13%) for the passage reading and the spontaneous speech task, respectively. The number of inappropriate breathing locations was significantly larger for the spontaneous speech task than for the passage reading task [$\chi^2(1) = 24$, p = 0.0001].

Breath Group Structure
Summaries of BGD, IBP, ID, f_0 mean, f_0 max and f_0 range data for the passage reading and the spontaneous speech tasks for each participant are shown in table 1.
Breath Group Duration. For the passage reading task, the mean and standard deviation of the total 273 BGDs were 4.05 and 1.5 s, and the range was 8.43 s, from a minimum of 0.93 s to a maximum of 9.36 s. For the spontaneous speech task, the mean and standard deviation of

the total 1,106 BGDs were 4.88 and 1.93 s, and the range was 13.12 s, from a minimum of 0.9 s to a maximum of 14.02 s. A paired t test was performed based on the mean values of BGD for different tasks for each participant. The spontaneous speech task had a significantly longer BGD than the passage task.
Inter-Breath-Group Pause. For the passage task, the mean and standard deviation of the total 273 IBPs were 0.64 and 0.24 s, respectively, and the range was 1.55 s, from a minimum of 0.25 s to a maximum of 1.8 s. For the spontaneous speech task, the mean and standard deviation of the total 1,106 IBPs were 0.69 and 0.28 s, respectively, and the range was 3.16 s, from a minimum of 0.23 s to a maximum of 3.4 s. There was no significant difference for IBP between passage and spontaneous speech tasks.
Inspiratory Duration. For the passage reading task, the mean and standard deviation of the total 273 IDs were 0.54 and 0.18 s, respectively, and the range was 1.02 s, from a minimum of 0.19 s to a maximum of 1.21 s. For the spontaneous speech task, the mean and standard deviation of the total 1,106 IDs were 0.57 and 0.18 s, respectively, and the range was 1.37 s, from a minimum of 0.19 s to a maximum of 1.56 s. There was no significant difference in IBP between the passage and spontaneous speech tasks.
Mean f_0. The task difference of f_0 mean was not significant.
Max f_0. The task difference of f_0 max was not significant.
Range f_0. There was no significant difference in f_0 range between passage and spontaneous speech tasks.

Discussion

The results of this study confirm and extend earlier reports on respiratory function in speech. The main result of the current study is that the spontaneous speech task exhibited significantly more grammatically inappropriate BG locations and longer BGD than did the passage reading task.

Appropriateness of Inspiratory Loci
The percentages of inappropriate inspiratory locations found in this study were similar to values found in previous studies for both reading [6, 13] and spontaneous speech [12, 14]. Some of the grammatically inappropriate inspiratory loci were due to the insertion of a filler, but none occurred within words. Therefore, the significantly

greater number of inappropriate breathing locations for spontaneous speech than for reading was unlikely due to poor planning of the utterances, but rather due to greater efforts required to coordinate inspiratory locations into a less predictable grammatical structure. Another possible reason is the heavier cognitive load required for spontaneous speech than for oral reading. Increased cognitive-linguistic demands have been reported to lead to a reduced number of syllables per breath group, slower speaking rate, and a greater lung volume expended per syllable [10]. In the current study, inappropriate inspiratory locations probably had little or no impact on speech intelligibility given that (1) none of them occurred within words, and (2) segmental and prosodic features within breath groups were intact within breath groups.

Breath Group Structure

Compared to previous reports, the BGD values observed in the current study were longer in spontaneous speech [5, 12], but comparable in reading [11, 13]; moreover, the ID values were comparable to those in a previous report [11]. Differences among studies are probably due to variations in the methods used to elicit spontaneous speech samples. In this study, the significantly longer BGD in spontaneous speech than in reading for healthy adult speakers is probably due to the differences in cognitive-linguistic loading between these two tasks [23]. There were no significant task differences in IBP or ID, which indicates that the inspiratory control during speech was consistent between these different tasks for healthy adult talkers. The above results suggest that the overall speech breathing cycle (IBP + BGD) in the spontaneous speech task was longer than that in reading.

The noninspiratory pause, defined as IBP minus actual ID, might be an index of the efforts involved with coordinating speech production subsystems and cognitive load in the communicative task. That is, the portion of pause that is not accounted for by actual inspiration may be determined by other factors, including motor control and cognitive effort. Further studies compiling acoustic and aerodynamic measures are needed to test this hypothesis by recruiting participants with speech motor disorders or cognitive deficits. The absence of task differences in f_0 mean, f_0 max, or f_0 range indicates: (1) f_0 control is uniform for these speaking behaviors, which simplifies the programming of laryngeal behavior in connection with respiratory activity, and (2) either task is suitable for assessing f_0 of healthy talkers during connected speech. However, because all the participants in this study had normal vocal function, additional studies are required to explore the possibility of f_0 differences across tasks in participants with impaired vocal control.

Implications for Speech Breathing

Speech respiration differs from resting respiration in having a shorter inspiratory duration with increased velocity of airflow, and a longer expiratory duration with a decrease in velocity. Conrad and Schonle [23] concluded that respiratory patterns for a variety of tasks fall along a continuum from those produced during rest to those produced during speech. They noted that the degree of activation of the respiratory pattern for speech is determined by the degree of internal verbalization and that respiratory patterns for different tasks become more speechlike as they increased in their cognitive-linguistic processing demands. For example, vocalized arithmetic showed a much stronger speech pattern than did reading. Increased internal verbalization (cognitive-linguistic processing) also could explain the longer BGD for spontaneous speaking in the present study. If spontaneous speaking is taken to represent a high cognitive-linguistic load task, then the respiratory pattern for relatively unconstrained speech has the following temporal profile: BGD of about 4–5 s, ID of 0.6 s, and a breath group interval of 0.7 s. The ratio of BGD to ID is about 8:1. These values may be useful for clinical application, including assessment of respiratory function for speech or as guidelines for intervention. The fact that global features of f_0 pattern are highly similar across reading and spontaneous speaking tasks is evidence of a simplifying regularity in the control of laryngeal function vis-à-vis respiratory patterns.

References

1 Hixon TJ, Mead J, Goldman MD: Dynamics of the chest wall during speech production: function of the thorax, rib cage, diaphragm, and abdomen. J Speech Hear Res 1976;19: 297–356.

2 Hixon TJ, Goldman MD, Mead J: Kinematics of the chest wall during speech production: volume displacements of the rib cage, abdomen, and lung. J Speech Hear Res 1973;16: 78–115.

3 Kent RD, Read C: The Acoustic Analysis of Speech, ed 2. San Diego, Singular, 2002.

4 Grosjean F, Collins M: Breathing, pausing and reading. Phonetica 1979;36:98–114.

5 Bunton K: Patterns of lung volume use during an extemporaneous speech task in persons with Parkinson disease. J Commun Disord 2005;38:331–348.

6 Hammen VL, Yorkston KM: Respiratory patterning and variability in dysarthric speech. J Med Speech Lang Pathol 1994;2: 253–261.

7 Hodge MM, Rochet AP: Characteristics of speech breathing in young women. J Speech Hear Res 1989;32:466–480.

8 Hoit JD, Hixon TJ: Age and speech breathing. J Speech Hear Res 1987;30:351–366.

9 Hoit JD, Hixon TJ, Altman ME, Morgan WJ: Speech breathing in women. J Speech Hear Res 1989;32:353–365.

10 Mitchell HL, Hoit JD, Watson PJ: Cognitive-linguistic demands and speech breathing. J Speech Hear Res 1996;39:93–104.

11 Solomon NP, Hixon TJ: Speech breathing in Parkinson's disease. J Speech Hear Res 1993; 36:294–310.

12 Winkworth AL, Davis PJ, Adams RD, Ellis E: Breathing patterns during spontaneous speech. J Speech Hear Res 1995;38:124–144.

13 Winkworth AL, Davis PJ, Ellis E, Adams RD: Variability and consistency in speech breathing during reading: lung volumes, speech intensity, and linguistic factors. J Speech Hear Res 1994;37:535–556.

14 Wang YT, Kent RD, Duffy JR, Thomas JE: Dysarthria in traumatic brain injury: a breath group and intonational analysis. Folia Phoniatr Logop 2005;57:59–89.

15 Keller E, Bailly G, Monaghan A, Terken J, Huckvale M (eds): Improvements in Speech Synthesis: COST 258: The Naturalness of Synthetic Speech. Chichester, Wiley & Sons, 2001.

16 Lieberman P: Intonation, Perception, and Language. Cambridge, MIT Press, 1967.

17 Lieberman P: Some acoustic and physiologic correlates of the breath group. J Acoust Soc Am 1966;39:1218.

18 Green JR, Beukelman DR, Ball LJ: Algorithmic estimation of pauses in extended speech samples of dysarthric and typical speech. J Med Speech Lang Pathol 2004;12:149–154.

19 Darley FL, Aronson AE, Brown JR: Motor Speech Disorders. Philadelphia, Saunders, 1975.

20 Collyer S, Davis PJ: Effect of facemask use on respiratory patterns of women in speech and singing. J Speech Lang Hear Res 2006;49: 412–423.

21 Henderson A, Goldman-Eisler F, Skarbek A: Temporal patterns of cognitive activity and breath control in speech. Lang Speech 1965; 8:236–242.

22 Milenkovic P: Time-Frequency Analysis for 32-Bit Windows. Madison, 2001.

23 Conrad B, Schonle P: Speech and respiration. Arch Psychiatr Nervenkr 1979;226: 251–268.