12-2012

Accuracy of perceptual and acoustic methods for the detection of inspiratory loci in spontaneous speech

Yu-Tsai Wang
National Yang-Ming University, Taipei, Taiwan, yutsaiwang@ym.edu.tw

Ignatius S. B. Nip
San Diego State University, inip@mail.sdsu.edu

Jordan R. Green
University of Nebraska-Lincoln, jgreen4@unl.edu

Ray D. Kent
University of Wisconsin–Madison, kent@waisman.wisc.edu

Jane Finley Kent
University of Wisconsin–Madison

See next page for additional authors

Published in *Behavior Research Methods* 44:4 (December 2012), pp. 1121–1128;
doi: 10.3758/s13428-012-0194-0 ; PMCID: PMC3394888

Published online February 24, 2012.

Accuracy of perceptual and acoustic methods for the detection of inspiratory loci in spontaneous speech

Yu-Tsai Wang,[1] Ignatius S. B. Nip,[2] Jordan R. Green,[3]
Ray D. Kent,[4] Jane Finley Kent,[4] and Cara Ullman[3]

1. School of Dentistry, National Yang-Ming University, Rm. 411, No. 155, Sec. 2, Li-Nong St. Beitou District, Taipei, 11221, Taiwan
2. School of Speech, Language, and Hearing Sciences, San Diego State University, San Diego, CA, USA
3. Department of Special Education and Communication Disorders, University of Nebraska–Lincoln, Lincoln, NE, USA
4. Waisman Center, University of Wisconsin–Madison, Madison, WI, USA

Corresponding author – Yu-Tsai Wang, email yutsaiwang@ym.edu.tw

Abstract
The present study investigates the accuracy of perceptually and acoustically determined inspiratory loci in spontaneous speech for the purpose of identifying breath groups. Sixteen participants were asked to talk about simple topics in daily life at a comfortable speaking rate and loudness while connected to a pneumotach and audio microphone. The locations of inspiratory loci were determined on the basis of the aerodynamic signal, which served as a reference for loci identified perceptually and acoustically. Signal detection theory was used to evaluate the accuracy of the methods. The results showed that the greatest accuracy in pause detection was achieved (1) perceptually, on the basis of agreement between at least two of three judges, and (2) acoustically, using a pause duration threshold of 300 ms. In general, the perceptually based method was more accurate than was the acoustically based method. Inconsistencies among perceptually determined, acoustically determined, and aerodynamically determined inspiratory loci for spontaneous speech should be weighed in selecting a method of breath group determination.

Keywords: accuracy, breath group, spontaneous speech, acoustics

During speech, breathing patterns are constantly changing to balance the varying demands of an utterance with those of underlying homeostatic respiration. Among primates, humans appear unique in this refined and flexible capability for sound production (MacLarnon & Hewitt, 1999). During speech, the duration of inspiration typically represents only 9%–19% of the full breath cycle (inspiration + expiration; Loudon, Lee, & Holcomb, 1988). The characteristic respiratory pattern for speech (quick inspiration and a gradual and controlled expiration) inevitably imposes a breath-related structure on vocal output. This structure is commonly known as the breath group, a sequence of syllables or words produced on a single breath. Management of breath groups is one aspect of efficient and effective communication, for optimum vocal performance in both healthy and disordered speakers. The location and degree of inspiration must be planned prior to the production of an utterance to ensure that there is adequate aerodynamic power for conveying the linguistic properties of an utterance (Winkworth, Davis, Adams, & Ellis, 1995).

Identification of breath groups is often a fundamental step in the analysis of recorded speech samples, especially for reading passages, dialogs, and orations. Inspirations mark intervals of speech that can be subsequently examined for prosody and related variables. The usefulness of breath groups has been demonstrated in studies of (1) normal speech breathing (Hoit & Hixon, 1987; Mitchell, Hoit, & Watson, 1996), (2) the development of speech in infants (Nathani & Oller 2001), (3) design of speech technologies such as automatic speech recognition and text-to-speech synthesis (Ainsworth,

1973; Rieger, 2003), and (4) the assessment and treatment of speech disorders (Che, Wang, Lu, & Green, 2011; Huber & Darling, 2011; Yorkston, 1996). Common to these various applications is the need to identify groupings of syllables or words produced on a single breath, which is the inevitable respiratory imprint on spoken communication.

Breath groups have been determined in three ways: perceptually (by listening to the speech output), acoustically (usually by detecting pauses or silences that exceed a criterion threshold), and physiologically (typically by recording chest wall movements or the direction of airflow during speech). The physiologic method may be considered the gold standard; however, it is not always easily incorporated into studies of speech and cannot be used to analyze previously recorded samples (such as archival recordings) that did not employ physiological measures. Although many studies identify and evaluate breath groups perceptually and acoustically, the basic question about breath group studies using perceptual and acoustic methods is how well they correlate with physiologic analysis.

Perceptual determination is an indirect detection based on auditory judgments of speech features associated with the respiratory cycle (Bunton, Kent, & Rosenbek, 2000; Oller & Smith, 1977; Schlenck, Bettrich, & Willmes, 1993; Wang, Kent, Duffy, & Thomas, 2005; Wozniak, Coelho, Duffy, & Liles, 1999). Both the perceptual and acoustic methods can be applied to previously recorded speech samples and can be accomplished with only modest investment in hardware or software. Although most studies have investigated breath groups using either of these indirect methods, the accuracy of these approaches has not been tested; the perceptual method is entirely subjective, based on listeners' impressions; the acoustic method requires the user to specify a minimum duration for an acceptable pause. Therefore, silent portions in the speech signal that may be pauses but do not exceed this criterion are not investigated (Campbell & Dollaghan, 1995; Green, Beukelman, & Ball, 2004; Walker, Archibald, Cherniak, & Fish, 1992; Yunusova, Weismer, Kent, & Rusche, 2005).

In contrast to the indirect methods, the physiologic determination directly detects inspiratory and expiratory events through either airflow (Wang, Green, Nip, Kent, Kent, & Ullman, 2010) or chest-wall movements (Bunton, 2005; Forner & Hixon, 1977; Hammen & Yorkston, 1994; Hixon, Goldman, & Mead, 1973; Hixon, Mead, & Goldman, 1976; Hoit & Hixon, 1987; Hoit, Hixon, Watson, & Morgan, 1990; McFarland, 2001; Mitchell et al., 1996; Winkworth et al., 1995; Winkworth, Davis, Ellis, & Adams, 1994). Physiologic detection requires adequate instrumentation and may impose at least slight encumbrances on participants, such as the need to wear a face mask for oral airflow measures.

The present study is a follow-up to an earlier investigation of breath group detection in a task of passage reading. Because accuracy of breath group detection may be affected by the speaking task, it is necessary to examine the performance of different methods of detection in at least spontaneous speech and passage reading, which have been primary tasks in the study of speech production. Studies have shown that these two speaking tasks are associated with somewhat different patterns in breath group structure (Wang, Green, Nip, Kent, & Kent, 2010).

Method

Participants and stimuli

Sixteen healthy adults (6 males, 10 females), ranging in age from 20 to 64 years (M = 40, SD = 15), participated in the study. All participants were native speakers of North American English, with no self-reported history of speech, language, or neurological disorders. Participants had normal or corrected hearing and vision. Participants were screened to ensure that they had adequate speech, language, and cognitive skills required to discuss simple topics regarding daily life. In addition to the 16 speakers, three individuals from the University of Wisconsin–Madison judged where inspiratory loci fell in each speaking sample on the basis of auditory-perceptual cues in the audio recording.

Experimental protocol

Participants were seated and were instructed to hold a circumferentially vented mask (Glottal Enterprises MA-1 L) tightly against their faces. Expiratory and inspiratory airflows during the speaking tasks were recorded using a pneumotachograph (airflow) transducer (Biopac SS11lA) that was coupled to the facemask. Previous research has demonstrated that facemasks do not significantly alter breathing patterns (Collyer & Davis, 2006). Although respiratory activity may be affected by the participants' use of facemasks in combination with the hand and arm muscle forces needed to hold the mask tightly against the face, participants in the present study were talking comfortably. Audio signals were recorded digitally at 48 kHz (16-bit quantization) using a professional microphone (Sennheiser), which was placed approximately 2–4 cm away from the vented mask. Participants were also video-recorded using a Canon XL-1 s digital video recorder; however, only the audio signals were used for the analysis of breath group determination.

Participants were asked to talk about the following topics with a comfortable speaking rate and loudness 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. The topics were presented on a large screen using an LCD projector. Participants were given time to formulate their responses to the topics before the recording was initiated to obtain reasonably organized and fluent spontaneous speech samples. Each response was required to be composed of at least six breath groups (as monitored by an airflow transducer).

Breath group determination

Aerodynamics — Data from the pneumotachometer and the simultaneous digital audio signal were recorded using Biopac Student Lab 3.6.7. The airflow signal was sampled at 1000 Hz and subsequently low-pass filtered (F_{LP} = 500 Hz). The resultant airflow signal was later used to visually identify actual inspiratory loci, represented by the upward peak in the airflow trace indicating inspiration (Figure 1), whereas a downward trend in the signal indicated expiration. On the rare occasions where there was uncertainty about the location of the inspiratory location, the first and the second authors examined the airflow traces in order to reach a consensus agreement on the inspiratory location.

Perception — Breath groups for the speech samples were determined perceptually by three judges at the University of Wisconsin–Madison. The judges were native English speakers trained on how to identify breath groups using known perceptual cues that signal the production of inspiratory pauses. The judges for the determination of breath group were trained to learn how to determine the location of breath groups on the basis of possible cues before performing their tasks. They were asked to listen to other conversation speech samples and to mark the points on their transcription sheets at which inspiration occurred. When the inspiration was not audible, the judges estimated the inhalation point on the basis of auditory-perceptual impression and various acoustic cues, such as longer pause duration, f0 declination, and longer phrase-final duration, which are fairly reliable indicators of pauses in normal speech and infant vocalization (Nathani & Oller, 2001; Oller & Lynch, 1992). The judges were also provided with a standard set of instructions explaining the task (see the Appendix). In addition, the judges were allowed to listen to the speech samples repeatedly to ensure that they were confident in their determination on the breath group location. The procedures of breath group determination were as follows:

1. The speech samples were orthographically transcribed by a trained transcriptionist who did not serve as a judge in the determination of inspiratory loci.
2. Punctuations and upper- and lower-case distinctions (except for the pronoun I and proper names) were removed from the orthographic transcripts to prevent the judges from analyzing breath groups on the basis of punctuation and related visual cues in the transcript. Three spaces separated each word to prevent the judges from using word order to separate breath groups.
3. The speech samples prepared for the judges for the task of breath group determination were randomized for order of speaker, using a table of random numbers.
4. The judges listened to the speech samples at normal loudness and marked perceived inspiratory loci on the transcripts. The judges were asked to make a best

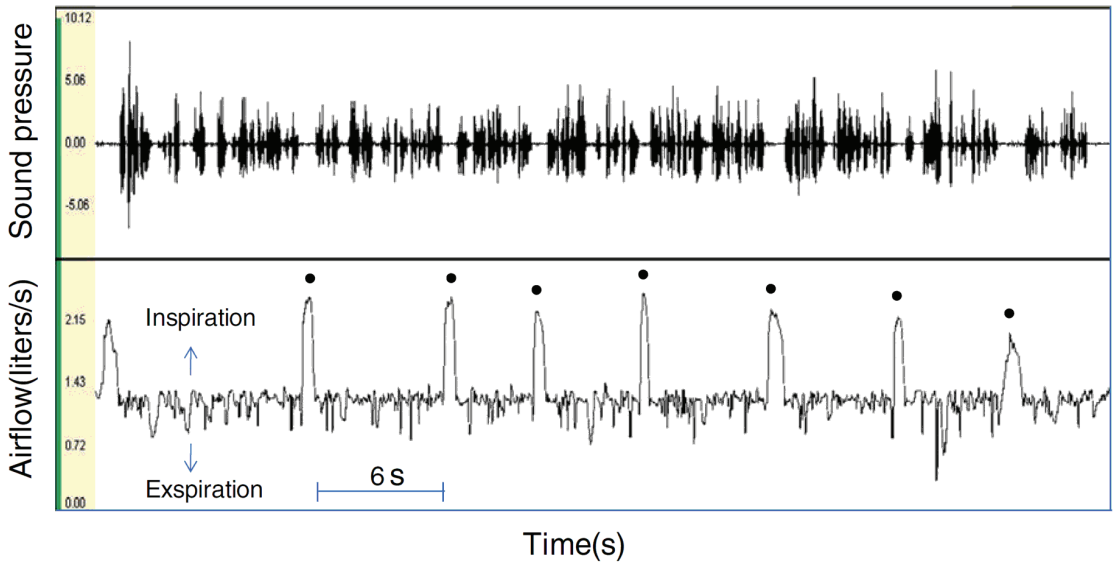

Figure 1. A demonstration of the locations of inspiration indicated by the dots for the recorded speech sample based on the aerodynamic signal (the lower panel). The upper panel is the corresponding sound pressure of the acoustic signal. The arrows indicate the direction of airflow.

guess of the inhalation location on the basis of their auditory-perceptual impressions when inspirations were not obvious. Therefore, these judgments could be based on multiple cues available to listeners, such as longer pause duration, f_0 declination, and longer phrase-final word or syllable duration. The judges were allowed to listen to the digitized speech samples repeatedly until they were satisfied with their determination of the breath group location.

5. The perceptual judgments of inspiratory loci were compared across each possible pairing of the three judges and across all the three judges to gauge the interjudge reliability. Measurement reliability was defined as the number of points that the judges agreed upon an inspiratory location divided by the total number of perceptually determined inspiratory loci by the three judges.

Acoustics — A custom MATLAB algorithm called speech pause analysis, or SPA (Green et al., 2004), determined the acoustically identified locations of the breath groups for the speech samples. The software required that a section of pausing be identified manually to specify the minimum amplitude threshold for speech. The software also required specification of durational threshold values for the minimum pause and speech segment durations. For the present study, five pause duration thresholds were tested: 150, 200, 250, 300, and 350 ms. These were selected to cover the range of pause duration thresholds typically used in previous studies; for example, inspiratory loci have been defined as pauses greater than 150 ms (Yunusova et al., 2005), 250 ms (Walker et al., 1992), or 300 ms (Campbell & Dollaghan, 1995). The minimum threshold for speech segment duration was held constant at 25 ms. Once these parameters were set, the acoustic waveform was rectified, and then signal boundaries were identified on the basis of the portions of the recording that fell below the signal amplitude threshold and above the specified minimum pause duration (e.g., 250 ms). Portions that exceeded the minimum amplitude threshold were identified as speech. Adjacent speech regions were considered to be a single region if a pause region was less than the minimum pause duration. Finally, all the speech and pause regions in the speech samples were calculated by the algorithm.

Accuracy

The loci of inspiration determined by the airflow signal for all speakers were marked first. Inspiratory loci in the aerodynamic signal were taken as the true inspiratory events because they reflected the physiologic events. The aerodynamically determined inspiratory loci were set to determine the accuracy of the perceptually determined loci and acoustically determined loci. Once inspiratory loci were determined using each of the three methods, the loci between conditions were compared. First, the number of perceptually or acoustically judged inspiratory loci was totaled. These loci were then compared with those identified using the aerodynamic signal. Loci identified perceptually and acoustically were then coded as a *true positive* when loci identified by the judges matched an inspiration identified by the aerodynamic method. Loci for which judges perceived an inspiration but that were not indicated in the aerodynamic signal were coded as a *false positive*. Aerodynamically determined loci that were not identified by the judges were coded as a *miss*.

Statistical analysis

Signal detection analysis (MacMillan & Creelman, 1991) was used to evaluate which perceptually based method and which pause threshold used in acoustic analysis yielded the most accurate results. Specifically, sensitivity as indicated by the true positive rate (TPR), the false positive rate (FPR; 1 − specificity), accuracy, and d' values were determined for each perceptual judgment and for each pause threshold.

Results

Accuracy

The total number of inspirations determined from the airflow signal for all speakers was 1,106. The number of pauses greater than 150 ms detected by the SPA algorithm was 2,281, which was considered the total number of potential inspirations for judges to make their decisions.

Perception — The total number of inspiratory loci determined perceptually by the three judges was 1,177. The number of inspiratory locations determined individually by judge 1 (J1), judge 2 (J2), and judge 3 (J3) was 1,088, 1,094, and 1,054, respectively. The number of consistent judgments between at least two of the three judges (i.e., J1J2, J1J3, J2J3, or J1J2J3), was 1,080. The number of consistent judgments across all three judges was 979. The highest interjudge reliability between two of the three judges was .92 (1,080/1,177). The interjudge reliability across all the three judges was .83 (979/1,177).

Referenced to the 1,106 actual inspiratory loci, J1 correctly identified 1,066, missed 42, and added 22 (false alarm). J2 correctly identified 1,065, missed 43, and added 29. J3 correctly identified 1,010, missed 98, and added 44. The loci that were consistent between at least two of the three judges were 1,068 correctly identified, 40 missed, and 12 added. The loci that were consistent across all three judges were 976 correctly identified, 132 missed, and 3 added.

Table 1 shows the sensitivity, specificity, accuracy, and *d'* data for the perceptual judgments. Inspiratory locations were perceived correctly (TPR) about 95% of the time on average, and the false alarm rate (FPR) varied among the three judges. J1 had the highest sensitivity, specificity, accuracy, and *d'*. When the decision was based on the agreement across all three judges, the specificity was increased substantially, but the sensitivity and accuracy were decreased to 88%. However, when the decision was based on agreement between at least two of the three judges, the specificity was near 99%, and the sensitivity, accuracy, and *d'* were all at their highest. Overall, the best discrimination of the perceptual judgment of inspiratory loci in spontaneous speech was based on the consistency between at least two of the three judges. However, as is shown in the receiver operating characteristic (ROC) curve of Figure 2, the separate results for the three judges are clustered rather tightly.

Acoustics — The number of pauses acoustically determined by the SPA algorithm is given in parentheses in the following summary for the five different pause thresholds: 150 ms (2,281), 200 ms (1,864), 250 ms (1,657), 300 ms (1,513), and 350 ms (1,406). Table 2 shows the sensitivity, specificity, accuracy, and *d'* data for the SPA algorithm results. Figure 2 shows the ROC for the combined perceptual and acoustic results. The TPR (sensitivity) values of the five different pause thresholds were all above 98%, but the FPR differed greatly among different threshold values, with smaller thresholds resulting in greater FPRs. The smaller thresholds had near perfect sensitivity but very poor specificity and, consequently, lower accuracy. Thus, in terms of the *d'* value, the SPA acoustically determined inspiratory loci of 300-ms threshold had the best performance.

As compared with the actual inspiratory locations determined by the aerodynamic signal, the perceptually determined method with the best performance had smaller TPR and FPR but larger accuracy and *d'* than did the acoustically determined method for this spontaneous speech task (Table 2). Moreover, the sensitivity values of the five different pause thresholds were all higher than those of perceptual judgments, but the specificity values were much larger and varied widely (Table 2). Consequently, on the basis of accuracy and *d'* analysis, the performance of the perceptually based breath determination of breath groups is judged to be better than that of the acoustic method of pause detection.

Discussion

The present study indicates that (1) the greatest accuracy in the perceptual detection of inspiratory loci was achieved with agreement between two of the three judges; (2) the most accurate pause duration threshold used for the acoustic detection of inspiratory loci was 300 ms; and (3) the perceptual method of breath group determination was more accurate than the acoustically based determination of pause duration.

For the perceptual approach, the criterion of agreement between two of the three judges yielded the highest TPR, accuracy (.977), and *d'* (4.116). This approach had approximately 1.75% (40/2,281) false negatives and 0.53% (12/2,281) false positives. Apparently, the more stringent criterion of consistency across all three judges led to an increase of false negatives that was much larger than the decrease of false positives, thereby reducing both accuracy and *d'*. In contrast, the most accurate approach for detecting inspiratory loci on the basis of listening in a reading task (Wang, Green, Nip, Kent, Kent, & Ullman, 2010) was agreement across all three judges, which achieved an accuracy of .902, a *d'* of 4.140, and a small number of both false negatives (approximately 10%) and false positives (0%). The accuracy of the perceptual approach was better for spontaneous speech in the present study than it was for passage reading in the study by Wang, Green, Nip, Kent, Kent, and Ullman (2010). The differences between spontane-

Table 1. The sensitivity, specificity, accuracy, and *d*-prime data of perceptual judgments determined by judge (J1, J2, J3), by the consistency of at least 2 of the 3 judges, and by the consistency of all the 3 judges. True positive rate (TPR) refers to sensitivity, whereas false positive (FPR) refers to 1 – specificity.

Judge(s)		Inspiratory location		TPR	FPR	Accuracy	*d*-prime	beta (ratio)
		Yes	No					
J1	Yes	1066	22	0.9621	0.0188	0.972	3.856	1.799
	No	42	1151					
J2	Yes	1065	29	0.9612	0.0247	0.968	3.729	1.452
	No	43	1144					
J3	Yes	1010	44	0.9116	0.0375	0.938	3.131	1.960
	No	98	1129					
2	Yes	1068	12	0.9639	0.0102	0.977	4.116	2.915
	No	40	1161					
3	Yes	976	3	0.8809	0.0026	0.941	3.979	25.122
	No	132	1170					

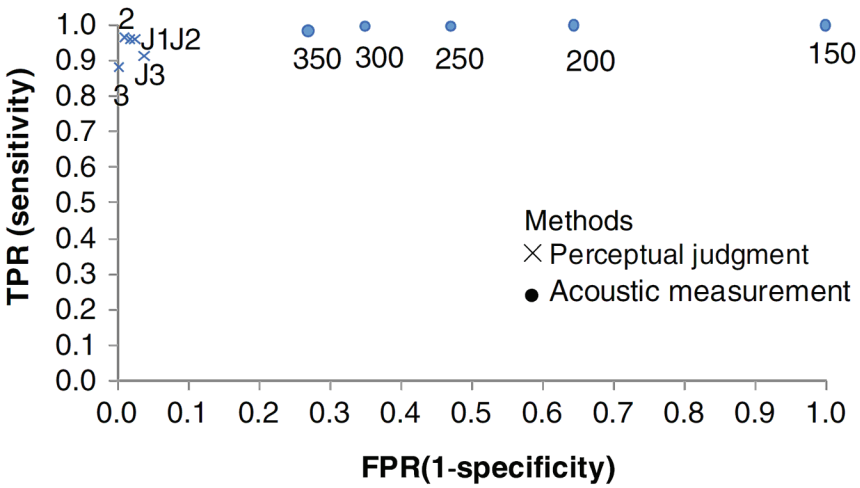

Figure 2. Receiving operator characteristic curve for the perceptual and acoustic methods of breath group determination. Perceptual results are shown for each judge and agreements between two judges (2) and three judges (3). Acoustic results are shown for various thresholds of pause duration.

ous speech and reading are likely explained by differences in breath group structure, as discussed in Wang, Green, Nip, Kent, and Kent (2010). Breath groups had longer durations for spontaneous speech, as compared with reading. In addition, inspiratory pauses for spontaneous speech are more likely to fall in grammatically inappropriate locations, potentially making the inspirations to be more perceptually salient to the judges.

Using acoustic algorithms to identify inspiratory loci, the optimal threshold of pause detection in the present study was 300 ms, which achieved an accuracy of .817 and a *d'* value of 2.994. With this threshold, the false negative rate is 0.2% (5/2,281), but the false positive rate is much higher, approximately 18% (412/2,281). Wang, Green, Nip, Kent, Kent, and Ullman (2010) reported that the most accurate pause duration threshold for detecting inspiratory loci in the reading task was 250, which achieved an accuracy of .895, a *d'* of 3.561, a zero rate of false negatives, and an approximately 10% rate of false positives. Task effects between reading and spontaneous speech occurred for the acoustic method, much as they did for the perceptual method. The accuracy and *d'* values in spontaneous speech were lower than those in reading. Furthermore, the false negative rate and false positive rate in spontaneous speech were both raised when compared with reading. Consequently, the acoustically determined method in spontaneous speech performed more poorly than for reading, which is likely related to the task differences in the breath group structure and perhaps in cognitive-linguistic load.

Because the minimum inter-breath-group pause in reading for healthy speakers is 250 ms (Wang, Green, Nip, Kent, & Kent, 2010), the 150- and 200-ms thresholds produced no false negatives but many false positives, which lowered their accuracy. In contrast, with thresholds above 200 ms, the decrease in the number of false positives was substantially more than the increase of the number of false negatives, which increased the accuracy. Generally speaking, the false positive rate differed among different pause thresholds, indicating that the selection of the pause threshold is very sensitive to the detection of false positives in spontaneous speech. Because the spontaneous speech samples in the present study were produced fluently by healthy adults who were familiar with the topics to be addressed, there was negligible occurrence of prolonged cognitive hesitations or articulatory or speech errors. Therefore, the present findings may not apply to speech produced by talkers with neurological or other impairments, whose speech might be characterized by either a faster or a slower

Table 2. The sensitivity, specificity, accuracy, and *d*-prime data of acoustically determined by SPA algorithm. True positive rate (TPR) refers to sensitivity, whereas false positive rate (FPR) refers to 1 – specificity.

Threshold (ms)		Inspiratory location 1		TPR	FPR	Accuracy	*d*-prime	beta (ratio)
		Yes	No					
150	Yes	1106	1175	0.9995	0.9996	0.485	–0.017	1.056
	No	0	0					
200	Yes	1106	758	0.9995	0.6451	0.668	2.947	0.004
	No	0	417					
250	Yes	1104	553	0.9982	0.4706	0.757	2.983	0.015
	No	2	622					
300	Yes	1089	317	0.9846	0.2698	0.854	2.774	0.117
	No	17	858					

speaking rate and with more pauses of long durations unrelated to inspiration. A threshold of 300 ms might potentially be either too short for individuals who speak significantly slower or too long for speakers with faster than typical speaking rates.

Taking together the present results and those of Wang, Green, Nip, Kent, Kent, and Ullman (2010), it can be concluded that for both spontaneous speech and passage reading, the perceptual method of breath group determination is more accurate than the acoustic method based on pause duration. The ability of listeners to identify breath groups is no doubt aided by their knowledge that speech is typically produced on a prolonged expiratory phase. Simple acoustic measurements of pauses are naive to this expectation, which is one reason perceptual assessment can be more accurate than acoustic pause detection. The larger *d'* obtained for the perceptual approach may indicate that listeners are sensitive to many cues beyond pause duration. Factors related to physiologic needs, cognitive demands, and linguistic accommodations that affect the locations of inspirations and the durations of inter-breath-group pauses are possibly perceptible by human ears. Perceptual cues for inspiration include the occurrence of pauses at a major constituent boundary, anacrusis, final syllable lengthening, and final syllable pitch movement (Wozniak et al., 1999). Some of these factors could be included in an elaborated acoustic method that relies on more than just pause duration.

The choice of method for breath group determination should be based on a consideration of the risk–benefit ratio. If errors cannot be tolerated, physiologic methods are preferred, if not mandatory. But if this is not possible (as in the analysis of archived audio signals), the choice between perceptual and acoustic methods should weigh the risk of greater errors (likely to occur with the acoustic method) against the relative costs (in terms of both analysis time and technology). As is shown in Figure 2, the results for any one judge in the perceptual method were more accurate than those for any of the pause duration thresholds used in the acoustic study. Perceptual determination appears to be a better choice, on the basis of accuracy alone. Of course, these findings pertain to studies interested in identifying only inspiratory pauses, and not those located at phrase and word boundaries; the high false positive rates obtained for the acoustic method suggest that this approach may be well suited for this purpose, although additional research is needed. If it is desired to examine the relationship between breath groups and linguistic structures, preparation of a transcript is necessary for any method of breath group determination. Finally, it should be recognized that the present results and those of Wang, Green, Nip, Kent, Kent, and Ullman (2010) pertain to healthy adult speakers. Generalization of the results to younger or older speakers or to speakers with disorders should be done with caution.

Acknowledgments — This work was supported in part by Research Grant number 5 R01 DC00319, R01 DC009890, and R01 DC006463 from the National Institute on Deafness and Other Communication Disorders (NIDCD-NIH) and NSC 100-2410-H-010-005-MY2 from the National Science Council, Taiwan. Additional support was provided by the Barkley Trust, University of Nebraska–Lincoln, Department of Special Education and Communication Disorders. Some of the data were presented in a poster session at the 5th International Conference on Speech Motor Control, Nijmegen, 2006. We would like to acknowledge Hsiu-Jung Lu and Yi-Chin Lu for data processing.

Appendix

The instruction of breath group determination for conversational speech samples — You will be provided with a transcription of the conversational speech samples without punctuations for each speaker in the present study. The task is to mark the points at which speakers stop for a breath. When you identify this point, place a mark on the corresponding location on the transcript. Make your best guess as to where the speaker stops to take a breath. Sometimes you can hear an expiration and/or inspiration, but in other cases you may have to make the judgment based on other cues, such as longer pause duration, f_0 declination, and longer phrase-final duration. In this task, you can listen to the sound files repeatedly before you are confident in your determination on the breath group location. Do you have any questions?

References

Ainsworth, W. (1973). A system for converting English text into speech. IEEE Transactions on Audio and Electroacoustics, 21, 288–290.

Bunton, K. (2005). Patterns of lung volume use during an extemporaneous speech task in persons with Parkinson disease. Journal of Communication Disorders, 38, 331–348.

Bunton, K., Kent, R. D., & Rosenbek, J. C. (2000). Perceptuoacoustic assessment of prosodic impairment in dysarthria. Clinical Linguistics and Phonetics, 14, 13–24.

Campbell, T. F., & Dollaghan, C. A. (1995). Speaking rate, articulatory speed, and linguistic processing in children and adolescents with severe traumatic brain injury. Journal of Speech and Hearing Research, 38, 864–875.

Che, W. C., Wang, Y. T., Lu, H. J., & Green, J. R. (2011). Respiratory changes during reading in Mandarin-speaking adolescents with prelingual hearing impairment. Folia Phoniatrica et Logopaedica, 63, 275–280.

Collyer, S., & Davis, P. J. (2006). Effect of facemask use on respiratory patterns of women in speech and singing. Journal of Speech Language and Hearing Research, 49, 412–423.

Forner, L. L., & Hixon, T. J. (1977). Respiratory kinematics in profoundly hearing-impaired speakers. Journal of Speech and Hearing Research, 20, 373–408.

Green, J. R., Beukelman, D. R., & Ball, L. J. (2004). Algorithmic estimation of pauses in extended speech samples of dysarthric and typical speech. Journal of Medical Speech-Language Pathology, 12, 149–154.

Hammen, V. L., & Yorkston, K. M. (1994). Respiratory patterning and variability in dysarthric speech. Journal of Medical Speech-Language Pathology, 2, 253–261.

Hixon, T. J., Goldman, M. D., & Mead, J. (1973). Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung. Journal of Speech and Hearing Research, 16, 78–115.

Hixon, T. J., Mead, J., & Goldman, M. D. (1976). Dynamics of the chest wall during speech production: Function of the thorax, rib cage, diaphragm, and abdomen. Journal of Speech and Hearing Research, 19, 297–356.

Hoit, J. D., & Hixon, T. J. (1987). Age and speech breathing. Journal of Speech and Hearing Research, 30, 351–366.

Hoit, J. D., Hixon, T. J., Watson, P. J., & Morgan, W. J. (1990). Speech breathing in children and adolescents. Journal of Speech and Hearing Research, 33, 51–69.

Huber, J. E., & Darling, M. (2011). Effect of Parkinson's disease on the production of structured and unstructured speaking tasks: Respiratory physiologic and linguistic considerations. Journal of Speech, Language, and Hearing Research, 54, 33–46.

Loudon, R. G., Lee, L., & Holcomb, B. J. (1988). Volumes and breathing patterns during speech in healthy and asthmatic subjects. Journal of Speech and Hearing Research, 31, 219–227.

MacLarnon, A. M., & Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. American Journal of Physical Anthropology, 109, 341–363.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York: Cambridge University Press.

McFarland, D. H. (2001). Respiratory markers of conversational interaction. Journal of Speech Language and Hearing Research, 44, 128–143.

Mitchell, H. L., Hoit, J. D., & Watson, P. J. (1996). Cognitive-linguistic demands and speech breathing. Journal of Speech and Hearing Research, 39, 93–104.

Nathani, S., & Oller, D. K. (2001). Beyond ba-ba and gu-gu: Challenges and strategies in coding infant vocalizations. Behavior Research Methods, Instruments,& Computers, 33, 321–330.

Oller, D. K., & Lynch, M. P. (1992). Infant vocalizations and innovations in infraphonology: Toward a broader theory of development and disorders. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), Phonological development: Models, research, implications (pp. 509–536). Parkton, MD: York Press.

Oller, D. K., & Smith, B. L. (1977). Effect of final-syllable position on vowel duration in infant babbling. Journal of the Acoustical Society of America, 62, 994–997.

Rieger, J. M. (2003). The effect of automatic speech recognition systems on speaking workload and task efficiency. Disability and Rehabilitation, 25, 224–235.

Schlenck, K. J., Bettrich, R., & Willmes, K. (1993). Aspects of disturbed prosody in dysarthria. Clinical Linguistics & Phonetics, 7, 119–128.

Walker, J. F., Archibald, L. M., Cherniak, S. R., & Fish, V. G. (1992). Articulation rate in 3- and 5-year-old children. Journal of Speech & Hearing Research, 35, 4–13.

Wang, Y.-T., Green, J. R., Nip, I. S. B., Kent, R. D., & Kent, J. F. (2010). Breath group analysis for reading and spontaneous speech in healthy adults. Folia Phoniatrica et Logopaedica, 62, 297–302.

Wang, Y.-T., Green, J. R., Nip, I. S. B., Kent, R. D., Kent, J. F., & Ullman, C. (2010). Accuracy of perceptually based and acoustically based inspiratory loci in reading. Behavior Research Methods, 42, 791–797.

Wang, Y.-T., Kent, R. D., Duffy, J. R., & Thomas, J. E. (2005). Dysarthria in traumatic brain injury: A breath group and intonational analysis. Folia Phoniatrica et Logopedica, 57, 59–89.

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

Winkworth, A. L., Davis, P. J., Ellis, E., & Adams, R. D. (1994). Variability and consistency in speech breathing during reading: Lung volumes, speech intensity, and linguistic factors. Journal of Speech and Hearing Research, 37, 535–556.

Wozniak, R. J., Coelho, C. A., Duffy, R. J., & Liles, B. Z. (1999). Intonation unit analysis of conversational discourse in closed head injury. Brain Injury, 13, 191–203.

Yorkston, K. (1996). Treatment efficacy: Dysarthria. Journal of Speech and Hearing Research, 39, 546–557.

Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M. (2005). Breath-group intelligibility in dysarthria: Characteristics and underlying correlates. Journal of Speech, Language, and Hearing Research, 48, 1294–1310.