Published in final edited form as:
Behav Res Methods. 2010 August ; 42(3): 791–797. doi:10.3758/BRM.42.3.791.

Accuracy of perceptually based and acoustically based inspiratory loci in reading

Yu-Tsai Wang,
National Yang-Ming University, Taipei, Taiwan

Jordan R. Green,
University of Nebraska–Lincoln, Lincoln, Nebraska

Ignatius S. B. Nip,
San Diego State University, San Diego, California

Ray D. Kent,
University of Wisconsin, Madison, Wisconsin

Jane Finley Kent, and
University of Wisconsin, Madison, Wisconsin

Cara Ullman
University of Nebraska–Lincoln, Lincoln, Nebraska

Abstract

Investigations of speech often involve the identification of inspiratory loci in continuous recordings of speech. The present study investigates the accuracy of perceptually determined and acoustically determined inspiratory loci. While wearing a circumferentially vented mask connected to a pneumotach, 16 participants read two passages. The perceptually determined and acoustically determined inspiratory loci were compared with the actual loci of inspiration, which were determined aerodynamically. The results showed that (1) agreement across all three judges was the most accurate of the approaches considered here for detecting inspiratory loci based on listening; (2) the most accurate pause duration threshold for detecting inspiratory loci was 250 msec; and (3) the perceptually based breath-group determination was more accurate than the acoustically based determination of pause duration. Inconsistencies among perceptually determined, acoustically determined, and aerodynamically determined inspiratory loci are not negligible and, therefore, need to be considered when researchers design experiments on breath groups in speech.

Continuous speech is structured in terms of breath groups that are based on the patterns of airflow from the lungs (Hixon, Goldman, & Mead, 1973; Hixon, Mead, & Goldman, 1976; Kent & Read, 2002). The pattern of breath-group structure contributes to speech intelligibility and prosody (Wang, Green, Nip, Kent, & Kent, 2010; Wang, Kent, Duffy, & Thomas, 2005). In speech research, the breath group has commonly served as a functional unit for delineating detailed analyses of the kinematic, acoustic, and perceptual aspects of speech. Thus, the accurate identification of inspiratory loci in continuous speech is a prerequisite of valid breath-group analysis.

Corresponding author — Y.-T. Wang, yutsaiwang@ym.edu.tw.

To identify inspiratory loci in continuous speech, clinicians and researchers have used three methods: (1) direct detection of inspiration by recording either chest-wall movements (Bunton, 2005; Forner & Hixon, 1977; Hammen & Yorkston, 1994; Hixon et al., 1973; Hixon et al., 1976; Hoit & Hixon, 1987; Hoit, Hixon, Watson, & Morgan, 1990; McFarland, 2001; Mitchell, Hoit, & Watson, 1996; Winkworth, Davis, Adams, & Ellis, 1995; Winkworth, Davis, Ellis, & Adams, 1994) or oral airflow (Wang et al., 2010); (2) indirect detection based on a presumed pause duration identified visually in acoustic recordings of speech (Campbell & Dollaghan, 1995; Walker, Archibald, Cherniak, & Fish, 1992; Yunusova, Weismer, Kent, & Rusche, 2005); and (3) indirect detection based on auditory–perceptual judgments of speech samples (Bunton, Kent, Kent, & Rosenbek, 2000; Oller & Smith, 1977; Schlenck, Bettrich, & Willmes, 1993; Wang et al., 2005; Wozniak, Coelho, Duffy, & Liles, 1999). Although many studies have relied on indirect methods to identify inspiratory loci in continuous speech, the accuracy of each technique has not been satisfactorily established.

A very small number of studies have reported on the reliability of auditory–perceptual identification of breath groups during speech, and they indicated a fairly high level of agreement (Wang et al., 2005; Wozniak et al., 1999). For example, the point-to-point intrarater reliability of intonational boundary determination using perceptual criteria in conversational discourse analysis was as high as 90.1% for speakers with closed head injury (Wozniak et al., 1999). The most identifiable intonation boundary in Wozniak et al.'s study was the occurrence of a pause at a major constituent boundary; however, additional perceptual cues included final syllable pitch movement, final syllable lengthening, and anacrusis (one or more unstressed syllables at the beginning of a phrase). Wang et al. (2005) also reported a satisfactory degree of reliability in the listening-based determination of inspiratory loci. This was true both for speakers with traumatic brain injury and for healthy speakers. Specifically, three judges marked inspiratory loci along the prepared texts while listening to the speech samples, and only the breath groups with consistent agreement across two of the three judges were used for further breath-group analysis. In other words, it was assumed that perceptual judgments of inspiratory loci of consistency between two of the three judges were more reliable and accurate than judgments made by a single listener (Wang et al., 2005).

Although the reliability of perceptually based breathgroup determination may be satisfactory, the accuracy of perceptually determined inspiratory loci has not been established. Because inspiratory loci determined by perceptual judgment may not be congruent with the actual inspiratory loci in continuous speech, it is possible that judges reliably identify loci of inspiration where the speaker does not take a breath (false alarms), and/or they miss loci of inspiration when the speaker does take a breath (misses). False alarms and misses may incorrectly segment speech samples into breath groups and, consequently, undermine experimental findings. The present investigation evaluates the sensitivity, specificity, and accuracy of the determination of inspiratory loci in reading by comparing perceptually determined inspiratory loci and actual inspiratory loci detected by oral airflow signals.

A more objective alternative to perceptual judgments of inspiratory loci is the identification of pauses in the acoustic speech signal using computer algorithms. This approach is significantly more efficient than using human listeners to identify inspiratory loci, an approach that constrains the scope and depth of research by limiting the number of participants and utterances that can be studied. Therefore, if it can be established that acoustic pause identification is both reliable and valid as a means of identifying inspiratory loci, then acoustic analysis can be used in place of, or as a complement to, subjective perceptual judgments.

The essential idea of acoustic analysis is that inspiratory loci are determined visually or algorithmically by identifying pauses in continuous recordings of speech waveforms that exceed a predefined minimal pause duration presumed to reflect interruption of speech for the purpose of breathing. For example, inspiratory loci have been defined as pauses greater than 150 msec (Yunusova et al., 2005), 300 msec (Campbell & Dollaghan, 1995), or 250 msec (Walker et al., 1992). Thus, different studies have adopted different pause-duration thresholds to segment continuous speech into breath groups and perform breathgroup analysis. It is likely that different speaking tasks have different optimal pause-duration thresholds, but these optimal thresholds are still unknown.

Another unexplored possibility is that the rate of false alarms and misses may vary as a function of pause-duration threshold. For example, because the pause between breath groups during reading for healthy speakers is approximately 250 msec (Wang et al., 2010), thresholds lower than 250 msec would be expected to produce more false positives than misses. These issues highlight the need to investigate the accuracy of acoustically determined inspiratory loci in reading, particularly since it varies as a function of pause-duration threshold.

The purposes of the present study are to investigate the accuracy of perceptually determined and acoustically determined inspiratory loci in reading, as based on actual inspiratory loci that are determined objectively by oral airflow, for a group of healthy adults. The following issues were to be addressed: (1) whether perceptual judgments of inspiratory loci that are consistent between two judges are more accurate than judgments made by a single judge or that are consistent among three judges, (2) whether there are optimal pause-duration thresholds for acoustically determined inspiratory loci, and (3) whether the accuracy of perceptually determined inspiratory loci is better than that of acoustically determined inspiratory loci.

Method

Participants

Participants who served as speakers were 16 healthy adults (6 males, 10 females), 20 to 64 years of age ($M = 40.3$, $SD = 14.8$). Participants were native speakers of North American English who reported no speech and language disorders and had adequate auditory, visual, language, and cognitive skills to read passages. Three additional individuals served as judges to determine inspiratory loci by listening to speech samples.

Stimuli

Speech samples, including the bamboo passage (Green, Beukelman, & Ball, 2004) and the grandfather passage (Darley, Aronson, & Brown, 1975), were obtained from each participant. The grandfather passage has been used extensively in prior investigations on speech production. The bamboo passage, which has not been used as extensively as the grandfather passage, was selected because it was specifically designed to elicit a large number of voiced consonants at word and phrase boundaries, allowing pauses in speech to easily be identified.

Experimental Protocol

Participants were seated and were instructed to hold a circumferentially vented mask (Glottal Enterprises MA-1L) tightly against their face. Expiratory and inspiratory flows during the speaking tasks were recorded using a pneumotach (airflow) transducer (Biopac SS11lA) that was coupled to the face mask. The use of the face mask was judged acceptable because a prior investigation suggested that face masks do not significantly alter breathing patterns (Collyer & Davis, 2006). However, it must be acknowledged that respiratory activity may have been affected by the use of face masks, as well as by the use of hand and arm muscle force to hold the mask tightly against the face.

Audio signals were recorded digitally at 48 kHz (16-bit) using a professional microphone (Sennheiser) that was placed approximately 2–4 cm from the vented mask. Participants were also videotaped using a Canon XL-1s digital video recorder; however, only the audio signals were used for the analysis of breath-group determination.

The speaking tasks were presented via PowerPoint on a large screen using an LCD projector. The task involved reading of the grandfather and bamboo passages at a comfortable speaking rate and loudness. To avoid the influence of unfamiliarity with the reading material, participants were given time to familiarize themselves with the passages before the recording was initiated.

Breath-Group Determination

Aerodynamics—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 1000 Hz and low-pass filtered at 500 Hz. The processed airflow signal was then used to manually identify actual inspiratory loci, which were indicated by the easily identified peak in the airflow trace (Figure 1). The upward airflow signal represents inspiratory direction, and the downward airflow signal shows expiratory direction. In general, the peaks were prominent and, therefore, easy to identify in the airflow trace. On the few occasions when analysts were uncertain about the location of a peak, the first and third authors rechecked and discussed the peaks to reach an agreement.

Perception—The locations of the breath groups for the speech samples were determined by three trained judges at the University of Wisconsin–Madison, who were native English speakers. The judges for this task were trained on breath-group definitions and the cues that determine the inspiratory loci. They were asked to listen to the speech samples and to mark on the transcription sheets the locations at which inspiration occurred. The judges were asked to make a best guess of the inhalation location on the basis of an auditory–perceptual impression 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. Judges were allowed to listen to the digitized speech samples repeatedly, until they were satisfied with their determination of the breath-group location. The procedures of breath-group determination are described in more detail in the following four steps.

In the first step, the speech samples were orthographically transcribed by an experienced judge.

In the second step, all punctuation and uppercase and lowercase distinctions (except for the pronoun “I” and proper names) of the orthographic transcripts were removed, in order to prevent judges from analyzing breath groups on the basis of the sequence of words in the transcript. The space between words was three times standard spacing in order to minimize the effects of word sequence.

In the third step, the speech samples prepared for the judges for breath-group determination were randomized by the order of participants using a table of random numbers. Three listeners heard the randomly arranged speech samples and then determined the locations of inspiration on the basis of their perceptual judgments.

In the fourth step, the consistency of the auditory–perceptual determinations of inspiratory loci were compared across two of the three judges and across the three judges to gauge interjudge reliability. *Measurement reliability* was defined as the counts of consistency among judges, divided by the total number of perceptually determined inspiratory loci by the three judges.

Acoustics—The acoustically identified locations of the breath groups for the speech samples were determined by an algorithm called *speech pause analysis* (SPA; Green et al., 2004). For this analysis, the user manually identified a section of pausing in order to specify the minimum amplitude threshold in each audio recording of speech and to specify durational threshold values for the minimum pause and speech segment durations. For the present study, the following five pause-duration thresholds were tested: 150, 200, 250, 300, and 350 msec. The minimum speech segment duration threshold was kept constant at 25 msec.

The automated algorithm identified pauses in connected speech by first identifying signal boundaries associated with each possible pause region on the basis of values in a rectified version of the acoustic waveform that were below the signal amplitude threshold and specified minimum pause duration (e.g., 250 msec). Speech regions were identified as values that were above the minimum amplitude threshold. If a pause region was less than the minimum pause duration, flanking speech regions were joined. Finally, all the speech and pause regions in the speech samples were calculated automatically.

Pauses of the reading samples were hand measured by the first author to evaluate the accuracy of the SPA algorithm. These speech samples were displayed using the TF32 computer program (Milenkovic, 2001). The experimenter measured all pause durations longer than 150 msec visually. Different pause-duration thresholds (including 150, 200, 250, 300, and 350 msec) were used to count the numbers of correct loci of inspiration, false alarms, and misses. These results were then compared with the SPA outputs.

Accuracy

The loci of inspiration determined by the airflow signal for all speakers were marked first. Inspiratory loci in the aerodynamic signal were considered to be the true inspiratory events and the perceptually determined loci and acoustically determined loci were compared against them for accuracy. For each comparison, the following measures were counted and calculated: (1) total number of perceptually or acoustically judged inspiratory loci, (2) loci where judges noted an inspiration that did occur (true positive), (3) loci where judges noted an inspiration that did not occur (false positive), and (4) loci where judges missed an inspiration that did occur (miss). Then the true positive rate (TPR) or sensitivity, false positive rate (FPR) or 1 2 specificity, accuracy, and d' values were calculated for each perceptual judgment and each threshold.

Statistical Analysis

Signal detection analysis (sensitivity, specificity, accuracy) (Macmillan & Creelman, 1991) was used to determine the perceptually based method with the best performance and the optimal pause threshold for the acoustically based methods.

Results

Accuracy

The total number of inspirations determined from the airflow signal for all speakers in the present study was 273. The longest pause (1,081 msec) that was related to a hesitation in the reading speech samples was excluded from the following analysis. The number of pauses greater than 150 msec detected by the SPA algorithm was 408, and this was considered to be the total inspirations for judges to use to make their decisions in the present study.

Perception—For the reading speech samples, the total number of inspiratory locations determined perceptually by the three judges was 305. The numbers of inspiratory locations determined individually by Judge 1 (J1), Judge 2 (J2), and Judge 3 (J3) were 260, 278, and 271, respectively. The number of consistent judgments between at least two of the three judges (i.e., J1J2, J1J3, J2J3, and J1J2J3) was 271. The number of consistent judgments across all three judges was 233. The interjudge reliability values between two of the three judges and across all three judges were .89 and .76, respectively.

Referenced to the 273 actual inspiratory loci, J1 identified 259, missed 14, and added 1. J2 identified 247, missed 26, and added 31. J3 identified 257, missed 16, and added 14. The loci consistent between at least two of the three judges were as follow: identified 260, missed 13, and added 11. The loci consistent across all three judges were as follow: identified 233, missed 40, and added 0.

Table 1 shows the sensitivity and specificity data for the results of perceptual judgments. Individual differences between judges were evident. Inspiratory locations were perceived correctly (true positive rate, TPR) 90% of the time, and the false alarm rate (false positive rate, FPR) varied among the three judges (Figure 2). Although J1 had the highest accuracy and d', the bias (β) also increased. When the decision was based on the consistency among the three judges, the specificity was increased, but the sensitivity and accuracy were lowered to 85% and 90%, respectively. However, when the decision was based on the consistency between at least two of the three judges, on average, the sensitivity, specificity, and accuracy were all raised. Overall, the best discrimination of the perceptual judgment of inspiratory loci in reading was based on consistency across the three judges, since hits were just as important as correct rejections.

Acoustics—For the five different pause thresholds determined by the SPA algorithm—150, 200, 250, 300, and 350 msec—the numbers of pauses were 408, 348, 316, 293, and 284, respectively. Table 2 shows the sensitivity and specificity data for the SPA algorithm. The TPR (sensitivity) values of the five different pause thresholds were all above 95%, but the FPR differed greatly among different thresholds, with smaller thresholds resulting in greater FPRs. The smaller thresholds had near-perfect sensitivity but poor specificity and, consequently, lower accuracy. Thus, in terms of accuracy, the SPA acoustically determined inspiratory loci at the 250-msec threshold had the best performance.

In order of increasing pause-duration thresholds, the number of pauses for each hand-measured threshold equalled 363, 341, 319, 297, and 284. Table 3 shows the sensitivity and specificity data for these hand measurements. The trends for sensitivity and specificity values of the five different pause thresholds were identical to those acoustically determined by the SPA algorithm. The primary difference between the two methods was the FPR (1 2 specificity) values. Although the 150-, 200-, and 250-msec thresholds could detect all of the actual inspiratory loci, a high frequency of false alarms occurred at these thresholds. It was evident that the larger threshold had a smaller FPR difference (Figure 2). In terms of accuracy, the 250-msec threshold had the best ability to discriminate an inspiration from a noninspiration.

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 the reading task

(Figure 2). Moreover, the sensitivity values of the five different pause thresholds were all higher than those for perceptual judgments, but the specificity values varied widely (Figure 2). Generally speaking, the performance of the perceptually based breath-group determination was better than that of the acoustically based method in reading.

Discussion

The present study indicates that (1) the most accurate of the approaches considered here for detecting inspiratory loci based on listening was agreement across all three judges; (2) the most accurate pause-duration threshold for detecting inspiratory loci was 250 msec; and (3) the perceptually based breath-group determinations were more accurate than the acoustically based determinations of pause duration.

Although the highest d' was obtained with a criterion of consistency across all three judges, it is noteworthy that Judge 1 achieved a value of d' nearly as high as that for consistency across all three judges. That is, this judge was particularly adept at determining inspiratory loci. Individual differences among the judges appear to be substantial, and this fact limits generalization of the results.

The most accurate of the approaches considered here for detecting inspiratory loci in reading by listening was agreement across all three judges, but the accuracy is not perfect (.902). Apparently, the criterion of consistency among all three judges was so stringent that there were still 40 (~10%) false negatives. The optimal pauseduration threshold for separating reading speech samples into breath groups in the present study was 250 msec; however, as compared with the actual inspiratory locations determined by the aerodynamic signal, its accuracy was only .895. If we adopt the optimal threshold (250 msec), the false negative rate (miss rate) was zero, but the false positive rate (false alarm rate) was approximately 10% (43/408). These findings demonstrate the unavoidable limitations of both perceptually based and acoustically based methods in determining inspiratory loci in reading and partly explain the differences of the false positive rate and miss rate between the perceptually based method and the selected acoustically based method with a 250-msec pause threshold.

Because the minimum inter-breath-group pause in reading for healthy speakers is 250 msec (Wang et al., 2010), the 150-, 200-, and 250-msec thresholds produced no false negatives (misses) but many false positives, which lowered their accuracy. In contrast, with thresholds above 250 msec, the number of false positives decreases, and the number of false negatives increases, which reduces accuracy and d'. 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.

Because the reading passages in the present study were read fluently by healthy adults who were familiar with the two passages, there were negligible occurrences of prolonged cognitive hesitations or of articulatory or speech errors. Therefore, the present findings may not apply to spontaneous speech samples, where longer pauses not related to inspiration may occur more frequently, or to speech produced by talkers with neurologic impairments, whose speech might be characterized by either a faster or a slower speaking rate. A threshold of 250 msec might be either too short for individuals who speak significantly slower or too long for speakers with faster than typical speaking rates. Further studies on potential differences in the duration of the inhalation pauses produced by speakers with dysarthria or apraxia of speech at the appropriate grammatical boundaries versus in the middle of the grammatical clauses are needed.

The perceptually based method may have exhibited a larger d' than acoustically based methods did for the reading task because many more cues—not only pause duration—can be considered by a listener. Factors related to physiologic needs, cognitive demands, and linguistic accommodations that affect the locations of inspirations and the durations of inter-breath-group pauses may possibly be perceptible by human ears. Perceptual cues for inspiration include the occurrence of a pause 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 goes beyond pause duration. In the present study, the impact of the airflow mask on inspiration acoustics is not known; however, it should be noted that the accuracy of perceptual judgments may have been enhanced because airflow through the circumferentially vented mask might have made the inspirations more audible (the “Darth Vader effect”).

It is not entirely surprising that a single pause-duration threshold is less effective than listening, since durations of inspiratory pauses are known to vary, depending on the influence of multiple factors (e.g., physiologic needs, cognitive demands, linguistic accommodations). Furthermore, the impact of these factors on inspiratory pauses may vary across different age groups and genders, across speaking tasks (such as reading and spontaneous speech), and across speakers’ speech motor control (such as dysarthrias, apraxia of speech, and hearing impairment). Additional studies on these factors are needed to advance our understanding of speech breathing behaviors.

The perceptually based method for determining inspiratory loci is more accurate than the acoustically based method focusing on pause duration, but false negatives and false positives still occur. How the erroneous identification of 10% of the inspiratory loci affects the outcomes of research projects will have to be evaluated on a case-bycase basis. The present investigation, however, suggests that the accuracy of both methods is probably satisfactory for most applications, but the rate of inconsistency for each method is not negligible. Therefore, if errors in the detection of inspiratory loci cannot be tolerated, perceptual or acoustic methods might not be adequate. For example, these methods may be adequate for identifying large group differences but inadequate for testing the efficacy of experimental drugs that target improved respiratory function.

Acknowledgments

This work was supported in part by Research Grants R01 DC00319, R01 DC009890, and R01 DC006463, from the National Institute on Deafness and Other Communication Disorders (NIDCD-NIH), and NSC 94-2614-B-010-001 and NSC 95-2314-B-010-095, 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 Fifth International Conference on Speech Motor Control, Nijmegen, 2006. We acknowledge Hsiu-Jung Lu and Yi-Chin Lu for data processing.

References

Bunton K. Patterns of lung volume use during an extemporaneous speech task in persons with Parkinson disease. Journal of Communication Disorders. 2005; 38:331–348. [PubMed: 15963334]

Bunton K, Kent RD, Kent JF, Rosenbek JC. Perceptuo-acoustic assessment of prosodic impairment in dysarthria. Clinical Linguistics & Phonetics. 2000; 14:13–24. [PubMed: 22091695]

Campbell TF, Dollaghan CA. Speaking rate, articulatory speed, and linguistic processing in children and adolescents with severe traumatic brain injury. Journal of Speech & Hearing Research. 1995; 38:864–875. [PubMed: 7474979]

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

Darley, FL.; Aronson, AE.; Brown, JR. Motor speech disorders. Philadelphia: Saunders; 1975.

Forner LL, Hixon TJ. Respiratory kinematics in profoundly hearing-impaired speakers. Journal of Speech & Hearing Research. 1977; 20:373–408. [PubMed: 895106]

Green JR, Beukelman DR, Ball LJ. Algorithmic estimation of pauses in extended speech samples of dysarthric and typical speech. Journal of Medical Speech–Language Pathology. 2004; 12:149–154. [PubMed: 20628555]

Hammen VL, Yorkston KM. Respiratory patterning and variability in dysarthric speech. Journal of Medical Speech–Language Pathology. 1994; 2:253–261.

Hixon TJ, Goldman MD, Mead J. Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung. Journal of Speech & Hearing Research. 1973; 16:78–115. [PubMed: 4267384]

Hixon TJ, Mead J, Goldman MD. Dynamics of the chest wall during speech production: Function of the thorax, rib cage, diaphragm, and abdomen. Journal of Speech & Hearing Research. 1976; 19:297–356. [PubMed: 135885]

Hoit JD, Hixon TJ. Age and speech breathing. Journal of Speech & Hearing Research. 1987; 30:351–366. [PubMed: 3669642]

Hoit JD, Hixon TJ, Watson PJ, Morgan WJ. Speech breathing in children and adolescents. Journal of Speech & Hearing Research. 1990; 33:51–69. [PubMed: 2314085]

Kent, RD.; Read, C. Acoustic analysis of speech. 2. San Diego: Singular; 2002.

Macmillan, NA.; Creelman, CD. Detection theory: A user's guide. New York: Cambridge University Press; 1991.

McFarland DH. Respiratory markers of conversational interaction. Journal of Speech, Language, & Hearing Research. 2001; 44:128–143.

Milenkovic, P. Time–frequency analysis for 32-bit Windows. Madison: University of Wisconsin; 2001.

Mitchell HL, Hoit JD, Watson PJ. Cognitive–linguistic demands and speech breathing. Journal of Speech & Hearing Research. 1996; 39:93–104. [PubMed: 8820701]

Oller DK, Smith BL. Effect of final-syllable position on vowel duration in infant babbling. Journal of the Acoustical Society of America. 1977; 62:994–997. [PubMed: 908792]

Schlenck KJ, Bettrich R, Willmes K. Aspects of disturbed prosody in dysarthria. Clinical Linguistics & Phonetics. 1993; 7:119–128.

Walker JF, Archibald LMD, Cherniak SR, Fish VG. Articulation rate in 3- and 5-year-old children. Journal of Speech & Hearing Research. 1992; 35:4–13. [PubMed: 1735975]

Wang YT, Green JR, Nip ISB, Kent RD, Kent JF. Breath group analysis for reading and spontaneous speech in healthy adults. Folia Phoniatrica et Logopaedica. 2010; 62:301–306.

Wang YT, Kent RD, Duffy JR, Thomas JE. Dysarthria in traumatic brain injury: A breath group and intonational analysis. Folia Phoniatrica et Logopaedica. 2005; 57:59–89. [PubMed: 15914992]

Winkworth AL, Davis PJ, Adams RD, Ellis E. Breathing patterns during spontaneous speech. Journal of Speech & Hearing Research. 1995; 38:124–144. [PubMed: 7731204]

Winkworth AL, Davis PJ, Ellis E, Adams RD. Variability and consistency in speech breathing during reading: Lung volumes, speech intensity, and linguistic factors. Journal of Speech & Hearing Research. 1994; 37:535–556. [PubMed: 8084185]

Wozniak RJ, Coelho CA, Duffy RJ, Liles BZ. Intonation unit analysis of conversational discourse in closed head injury. Brain Injury. 1999; 13:191–203. [PubMed: 10081600]

Yunusova Y, Weismer G, Kent RD, Rusche NM. Breath-group intelligibility in dysarthria: Characteristics and underlying correlates. Journal of Speech, Language, & Hearing Research. 2005; 48:1294–1310.

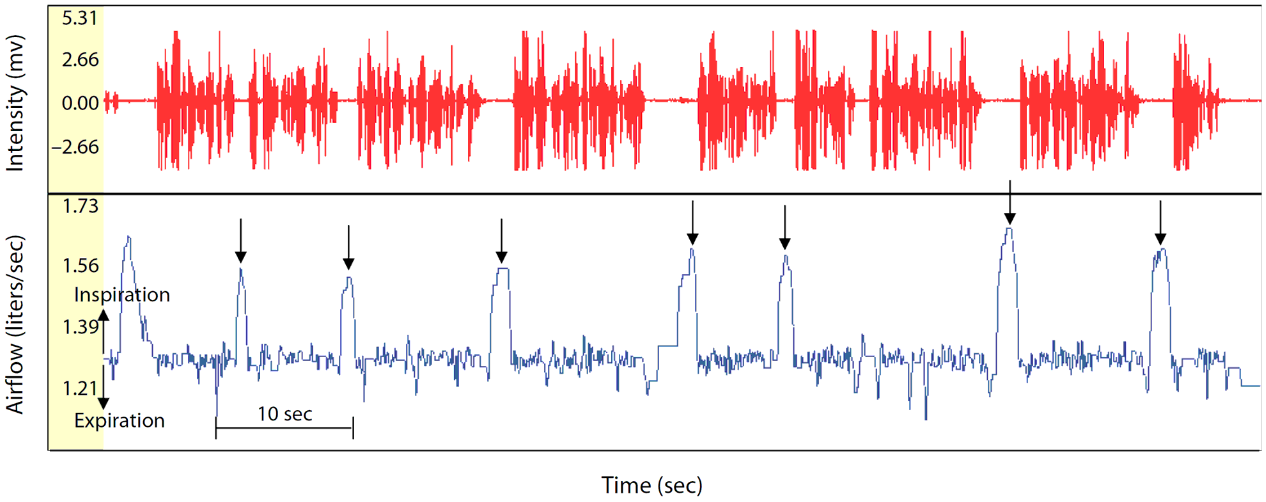

Figure 1.

The lower panel is a demonstration of the locations of inspiration, indicated by the arrows, for the bamboo passage, based on the aerodynamic signal. The upper panel is the corresponding acoustic signal.

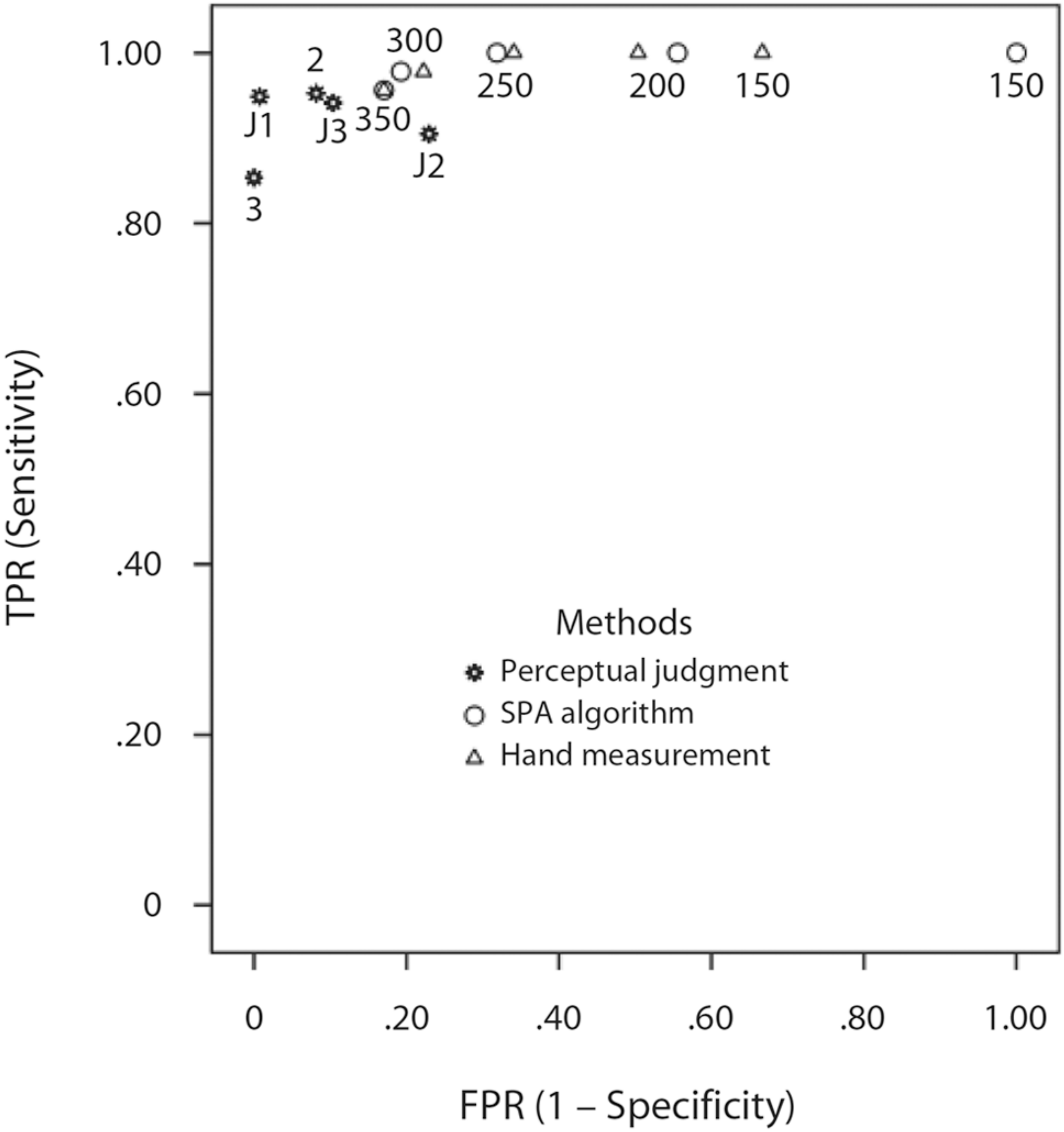

Figure 2.

A receiver operating characteristic curve for perceptual and acoustic methods. For the perceptual judgments, J1, J2, and J3 represent Judges 1, 2, and 3, respectively. The numeral 2 represents the consistency between two of the three judges. The numeral 3 represents the consistency among all three judges. The five different pause-duration thresholds — 150, 200, 250, 300, and 350 msec— are plotted for the speech pause analysis (SPA) and hand-measured acoustic data.

Table 1

Sensitivity and Specificity Data of Perceptual Judgments Determined by Judges 1, 2, and 3 (J1, J2, and J3), by the Consistency of Two of the Three Judges, and by the Consistency of All Three Judges

		Inspiratory Location						
Judge(s)		**Yes**	**No**	**TPR**	**FPR**	**Accuracy**	d'	β
J1	Yes	259	1	.949	.007	.963	4.092	5.377
	No	14	134					
J2	Yes	247	31	.905	.229	.861	2.053	0.558
	No	26	104					
J3	Yes	257	14	.941	.104	.926	2.822	0.651
	No	16	121					
2	Yes	260	11	.952	.081	.941	3.063	0.665
	No	13	124					
3	Yes	233	0	.853	.001	.902	4.140	68.317
	No	40	135					

TPR, true positive rate (sensitivity); FPR, false positive rate (1 – specificity).

Table 2

Sensitivity and Specificity Data of Perceptual Judgments Determined by the Speech Pause Analysis Algorithm

		Inspiratory Location						
Threshold (msec)		**Yes**	**No**	**TPR**	**FPR**	**Accuracy**	***d*´**	***β***
150	Yes	273	135	.999	.999	.669	0.000	1.000
	No	0	0					
200	Yes	273	75	.999	.556	.816	2.949	0.009
	No	0	60					
250	Yes	273	43	.999	.319	.895	3.561	0.009
	No	0	92					
300	Yes	267	26	.978	.193	.922	2.881	0.192
	No	6	109					
350	Yes	261	23	.956	.171	.914	2.656	0.366
	No	12	112					

TPR, true positive rate (sensitivity); FPR, false positive rate (1 – specificity).

Table 3

Sensitivity and Specificity Data of Perceptual Judgments Determined by Acoustic Hand Measurement

		Inspiratory Location						
Threshold (msec)		**Yes**	**No**	**TPR**	**FPR**	**Accuracy**	***d*´**	***β***
150	Yes	273	90	.999	.667	.779	2.659	0.009
	No	0	45					
200	Yes	273	68	.999	.504	.833	3.080	0.008
	No	0	67					
250	Yes	273	46	.999	.341	.887	3.500	0.009
	No	0	89					
300	Yes	267	30	.978	.222	.912	2.780	0.176
	No	6	105					
350	Yes	261	23	.956	.171	.914	2.656	0.366
	No	12	112					

TPR, true positive rate (sensitivity); FPR, false positive rate (1 – specificity).