Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Automatic Acoustics Measurement of Audible Inspirations in Pathological Voices

Eduardo Castilllo-Guerra and Williams Lee

University of New Brunswick, P.O. Box 4400, 15 Dineen Dr., D36 Head Hall, Fredericton,
NB, Canada E3B 5A3
ecastill@unb.ca

Audible inspiration is a type of speech perturbation used in conjunction with other acoustic observations to assess different types of pathologic conditions of speech associated with neurological or vocal cord disorders. The perception of this voice perturbation is very subjective and difficult to appraise in a consistent form across multiple utterances, subjects and disorders. This work reports an algorithm to model the perception of audible inspirations. It automatically segments the inspirations in continuous speech based on time-frequency characteristics and estimates the magnitude of the perturbation through a linear combination of the number, duration and the intensity of the inspirations. The algorithm was evaluated with the Massachusetts Eye and Ear Infirmary Voice database and two other databases containing recording from motor speech disorders. Results: a new method to automatically segment inspiratory phonation was developed. It provided an average segmentation accuracy of 84.4% and enabled accurate objective judgments of the perturbations associated with audible inspirations.

1 Introduction

The use of instruments to assist the diagnosis and rehabilitation of speech disorders is common today in clinical practice. It has been enabled by an accelerated development of technology and an increasing interest of practitioners to provide accurate assessments of the acoustic perturbations. Acoustic analysis of speech is an interesting alternative of the instrumental assessment that provides objective measurements of different types of perturbations observed in disordered speech. This approach has a special relevance for pathologic conditions associated with neurological disorders that manifest speech perturbations since early stages of the disorders. In such cases, each acoustic perturbation is often assessed individually and later combined in a multidimensional analysis to achieve the general assessment. Some types of dysarthria are often studied with this approach [1] to assess the magnitude of the neurological lesion(s) and determine more effective rehabilitation strategies. Audible inspiration (AI) is often part of the multidimensional profile used to study these disorders.

AI is the result of excessive constriction of the airway due to a variety of physiological or neuromuscular problems. It has been associated with utterances resulting from turbulent airflow passing through adducted vocal folds [2, 3, 4, 5], edema, paralysis, or due to muscular "slowness" caused by reduced control over speech musculature [1]. Other researchers have associated AI with prolapse of the vocal tract (stridor [6]) and with subglottal obstructions (wheezing [7]).

The judgment of acoustic perturbations caused by AIs is commonly performed perceptually. It is often made based on a combination of different factors such as: intensity and frequency of occurrence. Normal voices can have occasional low-intensity AIs but they become pathologic if they affect normal prosodic patterns. Duration and intensity also provide relevant information about the originating causes. A more constricted airway imposes longer inhalation periods and louder inspirations. A stronger inhalation produces inspirations with shorter duration but more intense. This is because the pressure in the constricted area drops causing vibration of both the glottis and surrounded tissue. The perceptual method, however, is very subjective and difficult to appraise consistently across multiple trials, subjects and disorders [8].

Orlikoff et al. [4] previously investigated inspiratory phonation. They reported acoustic and physiologic characteristics of inspiratory phonations associating the AI perturbations with the number, intensity, duration and fundamental frequency of the inspirations. However, the way these characteristics interact when modeling the perception of this type of perturbation requires further research. A better understanding of the interaction between these features can lead to automatic acoustic measurements with great benefit to research, assessment and rehabilitation. Such measurements require automatic segmentation techniques that are challenged by the complexity of the disordered speech.

This research is focused on developing fully automated segmentation techniques and multivariable models of the perception of AIs aiming to provide a source of reference to clinicians.

2 Materials and methods

This research consisted in two main sections: the detection of AIs in continuous speech with independence of the text spoken and the modeling of the acoustic perception of AI perturbations. The work relied on the analysis of pathological voices from three databases. The Massachusetts Eye and Ear Infirmary Voice database (MEEI) [9] and two motor speech disorders databases reported in [10] and [11].

The first 350 pathological voices from the MEEI database were used to develop the segmentation algorithm. This database provided utterances from a wide variety of organic, neurological, traumatic and psychogenic voice disorders. All subjects studied recorded a fragment of the "Rainbow" passage [13] containing 12 seconds of continuous speech.

The other two databases were used to automatically assess the perception of AI perturbations manifested in utterances from subjects with different types of dysarthria. Both databases contained utterances with fragments of the passage “The Grand Father” [13]. The recordings were made by subjects undergoing several types of dysarthrias among which are: Parkinson disease (PD), amyotrophic lateral sclerosis (ALS), organic voice tremor (OVT), chorea (HC), dystonia (HD), flaccid dysarthria, spastic dysarthria (SD) and ataxic dysarthria (AD). A total of 108 disordered utterances and 19 normal speech recording were processed from these databases.

The method followed to develop the segmentation algorithm was based on the derivation of time-frequency characteristics of the inspiratory phonation from a detailed analysis of the data. The performance of the automatic algorithm was compared to a manual segmentation conducted by the researchers. All selected utterances from the MEEI database were listened and transcribed denoting the number, position and intensity of the inspirations. All inspirations were classified in normal or pathologic according to an intensity-based threshold derived heuristically from normal AI levels. The performance was evaluated considering the correct detection, false acceptance and false rejection indexes.

The modeling of the acoustic perception of AI perturbations was implemented on the motor speech disorder databases. All utterances were processed manually and automatically as performed in the first part. Four acoustic measurements were derived from the segmented inspirations and correlated with the judgments provided by three speech language pathologists. Different linear combinations of the acoustic measurement were studied to model the perception of the AI perturbations.

All participating judges had five or more years of experience working with disordered speech [8]. The perceptual assessment was performed in a scale equally spaced with 0 corresponding to normal inspirations and 6 to excessive amount of perturbation. The average perception of the judges was selected to evaluate the performance of the automated algorithm.

3 Characterization of AIs

The characteristics of AIs were studied in the time and frequency domains to produce observations that identify the inspirations in continuous speech.

3.1 Time-domain characterization

The time-domain characteristics of AIs exhibit three main phases (see Fig. 1a):

Phase 1- *The time interval between the inhalation of air starts and the constricted area begins to vibrate*. This is typically the only phase that occurs in normal inspirations. In this phase the air is inhaled in a controlled way with normal pressure in the constricted area. The acoustic waveform is semi-periodic with high noise level.

Phase 2- *The time interval when the constricted area is vibrating*. In this interval, the pressure in the constricted area drops causing vibration. The waveform of this phase resembles a periodic signal with low fundamental frequency and a reduced amount of noise with respect to previous phase.

Phase 3- *The time interval between the constricted area stop to vibrate and the inhalation of air stops*. This phase is similar to phase 1. It occurs when the vibration stops due a reduction of the air stream.

AIs are segments of speech that commonly occur at the end of words or phrases, when the volume of air in the lungs is insufficient. The intensity of AIs depends on the amount of air inhaled, the physical constriction of the airway and the air stream force. Softer inhalations require more time to replenish the lung capacity than more intense ones. This results in longer phase 1 and 3 but a shorter phase 2 with lower intensity. Stronger inhalations have short phase 1 and 3 with longer phase 2 characterized by high intensity and low noise levels.

Fig. 1b shows the contour of the RMS intensity signal of a typical AI. It reflects an initial low intensity corresponding to phase 1, followed by an average intensity increment associated to phase 2 succeeded by a decrement on intensity again due to phase 3.

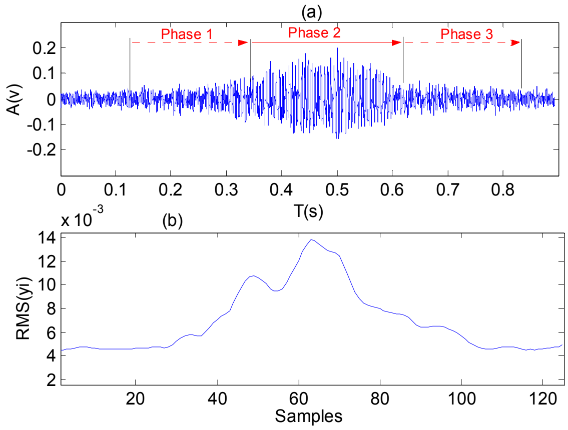

Fig.1 a) Typical AI waveform. b) RMS oscillogram obtained with 30ms frames and 66.6% overlap.

Previous time-domain characteristics were used for automatic detection of AIs in continuous speech. A sonority analysis was used to identify the segments between words with higher chances to host AIs. This analysis was performed based on correlation, energy and zero crossing criteria [8] relying on 30ms frames overlapped 20ms.

Although time-domain analysis is effective to detect AIs, there are other types of sound uttered in continuous speech that have similar characteristics. This is the case, for instance, of some types of monosyllables and syllables with fricative consonants. However, these types of sound can be effectively differentiated in the frequency domain.

3.2 Frequency domain characterization

The following observations were consistently present in the spectrum of typical AIs:

- There is an absence of harmonic structure when the intensity of the inspiration is weak. A main frequency and a couple of harmonics can occur in strong inspirations but it doesn't have a well-defined harmonic structure.
- In the inverse LPC spectrum: (a) there is a resonance in the 1200-2700 Hz band, (b) one or two other resonances can be present in the 3-4.4 kHz band, (c) no resonance can occur in the 500-1200 Hz band, and (d) a resonance in the band below 500 Hz can occur in very loud AIs (see Fig. 2).

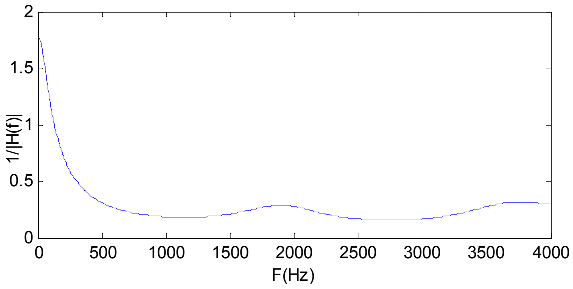

Fig.2 Spectral magnitude response of the inverse LPC filter modeling the vocal tract during AIs.

- There is a high ratio between the first spectral resonance and the second measured with respect to the middle trough (more than 13 times).
- The spectral magnitude of the band above 5 kHz is smaller than the band below (fricative sounds have opposite spectral magnitudes in these bands).

4 AI segmentation algorithm

The AI segmentation is based on a combination of time and frequency characteristics. The algorithm is implemented as follows:

a. Locate the segments that are candidates to have AIs using sonority analysis (marked with lighter color in Fig. 3). These segments are located in long unvoiced intervals between words.

b. Split such segments into 30ms frames overlapped 20ms to obtain a vector of RMS values. The RMS vector is smoothed with an averaging filter of 7 samples.

c. Obtain the 40 percentile of the RMS vector to estimate the lower threshold required to detect the starting trough of the AIs. Troughs below this threshold are candidates to be the starting point of the AI segment.

d. Obtain the 80 percentile of the RMS vector containing only voiced segments (darker color in Fig. 1). This is necessary to estimate the upper threshold required for discarding voiced segments incorrectly detected as unvoiced (i.e. first segment marked with lighter color in Fig. 3).

e. Detect peaks in the RMS signal between the upper and lower thresholds having adjacent troughs below the minimum threshold. The duration of the segment must be longer than a minimum inspiration length of 250ms (heuristically determined from data).

f. Find the ratio between the central peak and the first trough. AIs exhibited a range between 1.5 and 17. These values were obtained from a detailed analysis of the acoustic data studied.

g. Obtain the average RMS value of the segment calculated between each adjacent trough and compare it to a global low-magnitude threshold. This threshold discards AI segments with low intensity (non-audible). This threshold was set at 0.0020 based on data analysis.

The segments meeting the previous criteria can still include sounds with time-domain characteristics similar to AIs. The frequency domain analysis is used to extract the AIs only. The spectral analysis is performed through the following steps:

h. Apply a pre-emphasis filter and estimate the inverse linear prediction spectrum.

i. Verify that: there isn't any resonance in the 500-1200 Hz band of the spectrum, there is a resonance in the 1200-2400 Hz band and 1 or 2 resonances in the 2400-4000 Hz band.

j. Determine that the ratio between the first and second resonances is greater than 13 when estimated with respect to the middle though.

k. Estimate the spectrum of the utterance and determine that the harmonic structure only contains the main resonant frequency and one or two harmonics.

l. Verify that the ratio of the band below and over 5 kHz is greater than unity.

Fig. 3 shows a processed utterance highlighting the detected AIs and the segments where the inspirations were searched using the previous algorithm.

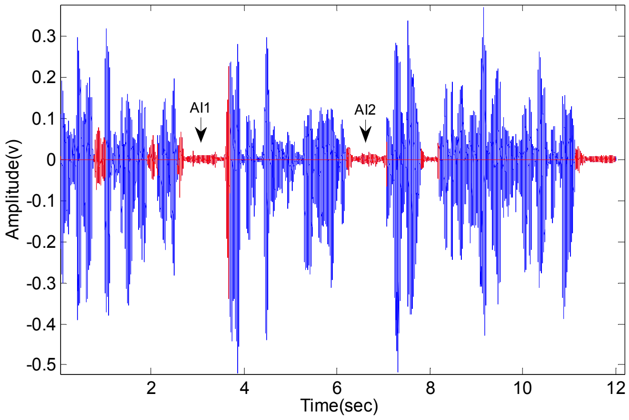

Fig. 3 Results of the AI segmentation algorithm. Unvoiced segments are highlighted and the detected AIs are labeled.

5 AI perception index

Four measurements were selected to model the perception of AI perturbations: the number of audible inspirations (NAI), the ratio between the maximum peak and the left trough from the RMS signal (as calculated in the detection algorithm, P/D), the duration of the inspiration (DAI) and the average RMS magnitude of the AI segment (AI_{rms}). The measurements and all possible linear combinations of them were submitted to statistical analyses to model the average perceptual judgments.

6 Analysis and results

6.1 MEEI database

The utterances of the MEEI database were analyzed manually and automatically. Table 1 shows a summary of the segmented inspirations with both methods.

Method	Number of detected segments	False Acceptance (%)	False Rejection (%)	Correctly Detected (%)
Manual	672	0 (0)	0 (0)	672 (100)
Automatic	624	77 (11.46)	32 (4.8)	547 (81.39)

Table 1 Summary of AI segmentation methods.

The automatic segmentation algorithm detected 624 AI segments of a total of 672 annotated manually. A fraction of the detected segments (29) corresponded with actual inspirations considered with normal intensity level in the manual assessment. A total of 48 segments did not correspond to segments containing inspirations and 32 inspirations were not detected. The total accuracy of the automatic segmentation was 81.39%.

A further analysis on the missing segments denoted that the decision of the segmenter was sometimes affected by low signal-to-noise ratio and the occurrence of non-AI-related acoustic perturbations. In many of those cases, the main source of error was caused by the voicing algorithm. Some of the extremely severe AIs were also occasionally missed because they sometimes presented longer than normal voiced segments or intensity levels beyond the thresholds estimated. The results, however, suggested that the algorithm is effective detecting AIs in continuous speech.

6.2 Dysarthria database

The performance of the objective algorithms was evaluated with the acoustic recording available in the dysarthria databases. The segmentation algorithm detected 222 AI segments in 127 utterances studied. The manual evaluation implemented by the researchers highlighted 254 AI segments. The segmentation accuracy achieved with this database was 87.41%.

The acoustic measurements described in section five were extracted from the segments located automatically. The number and position of the inspirations that influenced the judges' perception were not available for these databases. Therefore, the performance criteria when modeling the perception of AI perturbations was based on the correlation coefficient between the objective measurements and the average perceptual judgments (average correlation among judges was 86.7%). Table 2 shows the performance of each objective measurement.

Indicators	NAI	P/D	DAI	AI_{rms}
R	0.796	0.474	0.681	0.625
P values	<0.001	<0.005	<0.001	<0.001

Table 2 Correlation analysis between measurements and perceptual judgments.

NAI accounted for the largest percent of the data variability (79.6%) followed by the mean duration and intensity measurements ranging between 62% and 69%. The P/D measurement, however, did not account for a significant percent of the targeted data.

The results of the multivariable analysis are shown in Table 3. Only linear combinations of the variables were considered to prevent overfitting the datasets. Non-linear methods can provide better results but often are less generalisable to new acoustic data. The analysis implemented relied on "Best Subsets Regression Analysis" [12] which generates models formed with linear combinations of all variables. The maximum correlation criterion was used to select the models providing the best performance. The models were evaluated with four types of statistics:

(1) Correlation coefficient (*R*): described the proportion of the data variation explained by the measurements integrating the model (a magnitude closer to one indicates higher correlation between the model output and the target data).

(2) Adjusted correlation coefficient (R_{adj}): a modified version of R that has been adjusted for the number of variables composing the model [14].

(3) Mallows' statistic (C-p): the addition of the fitted response values divided by the variance of the model's error term (a magnitude closer to the number of variables in the model indicates higher performance).

(4) Standard deviation (S): reflects the estimated standard deviation of the model's error term (a smaller magnitude indicates higher performance).

The statistics R is useful to compare models of the same order and R_{adj}, C-p and S statistics are useful to compare models of different orders [14].

ID	Model Order	Performance Indices				Acoustic Measurements			
		R	R_{adj}	C-p	S	NAI	P/D	DAI	AI_{rms}
1	1	0.796	0.794	9.9	0.941	X			
2	1	0.681	0.676	71.7	1.139			X	
3	1	0.625	0.621	98.2	1.215				X
4	1	0.474	0.470	158.6	1.370		X		
5	2	**0.812**	**0.808**	**3.1**	**0.914**	**X**		**X**	
6	2	**0.810**	**0.807**	**3.9**	**0.916**	**X**			**X**
7	2	0.800	0.796	9.7	0.937	X	X		
8	2	0.691	0.685	68.6	1.120			X	X
9	3	**0.814**	**0.809**	**3.8**	**0.912**	**X**		**X**	**X**
10	3	0.812	0.807	4.7	0.915	X	X	X	
11	3	0.810	0.805	5.9	0.919	X	X		X
12	3	0.699	0.689	67.0	1.122		X	X	X
13	4	0.815	0.808	5.0	0.913	X	X	X	X

Table 3 Best subsets regression analysis of AI data.

The second-order models conformed by NAI-DAI and NAI-AI_{rms} achieved similar correlation indexes. This indicated that the duration and intensity of the inspirations could be contributing redundant information. This was verified and later explained by the high correlation found between the two measurements (0.818).

Model 7 composed by NAI-DAI-AI_{rms} had the highest R_{adj}, the C-p index closer to the number of variables, and smaller standard deviation. However, this model didn't provide a significant improvement over other multivariable models.

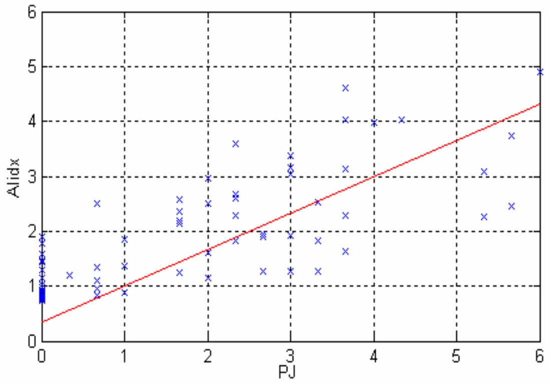

Fig. 4 Plot of PJs versus AI_{idx} index.

Fig 4 shows the plot of the judgments made with this model versus the PJs (AI_{idx} = -0.087+0.414 NAI+0.00971 DAI+ 42.3 AI_{rms}). The continuous line indicates a linear trend between both methods. A linear relationship appears between both methods but certain disagreement is observed in the lower scale values. This reflects mismatches in the low-level threshold calculated with the algorithm and the natural threshold used by the judges.

Fig. 5 shows the judgments obtained with the algorithm AI_{idx} for each type of disorder included in the databases. It is observed that FD and ALS groups have higher mean scores (similar results were observed on the perceptual analysis). Other groups such as: SD, HD, HC and HO also present cases with high scores, which are expected since AIs have also been reported in these groups [3]. This is consistent with the analysis reported in [1].

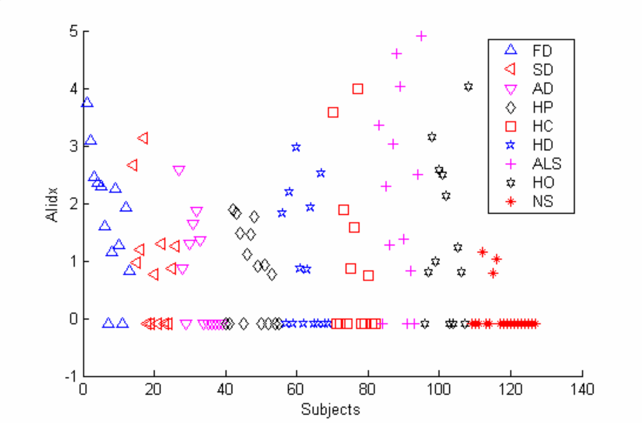

Fig. 5 Performance of AI_{idx} measurement.

7 Conclusion

This investigation studied the modeling of acoustic perception of AIs in disordered speech. A text-independent algorithm was developed to detect AIs automatically. The average segmentation accuracy of algorithm reported is 84.4%. The segmenter relied on sonority analyses and time-frequency characteristics to automatically extract the segments containing the inspirations.

Different acoustic measurements were extracted from the segmented inspirations to model the perception of the speech perturbations caused by AIs. It was noted that the number, the duration and the intensity of the inspirations were relevant to the perceptual judgments performed by the three judges. NAI was the single-variable model that produced the highest correlation with respect to perceptual judgments followed by measurements of intensity and duration.

Multivariable models evidenced higher descriptive capacity than single measurements, denoting that perceptual judgments are better modeled by the combination of multiple variables. This corresponded with the opinion of judges that based their evaluations on a combination of acoustics events perturbing normal prosodic patterns. The high-order models (orders 3 & 4) provided only modest performance improvement indicating that the second-order models are sufficient to represent the perception of AI perturbations. However, because the measurements of intensity and duration were needed for the segmentation stage, the model combining the three measurements (NAI-AI_{RMS}-DAI) was obtained at practically no extra cost.

The multivariable model, AI_{idx}, showed 80.9% correlation with respect to perceptual judgments, which is in the order of the correlation achieved among clinicians. Therefore, the model is an alternative reference judgment for clinicians when analyzing perturbation produced by AIs in pathological voices.

Acknowledgments

The authors appreciate the collaboration of the clinicians: Beth McCann from Stan Cassidy Center for Rehabilitation of New Brunswick, Rodolfo Gutierrez and Rafael Espinosa from the Department of Neurology of AMC Hospital in Santa Clara. Thanks also to Dr. Norma Mendez for her contribution to the paper. Authors appreciate NSERC and UNB for the financial contributions to this work.

References

[1] F.L. Darley, A.E. Aronson, J.R. Brown, “Differential diagnostic patterns of dysarthria” *J. Speech Hear. Res.* 12, 246-249 (1969a)

[2] S.M. Grau, M.P. Robb, A.T. Cacace, “Acoustic correlates of inspiratory phonation during infant cry”, *J. Speech Hear. Res.* 38, 373-381 (1995)

[3] C. Kelly, K. Fisher, “Stroboscopic and acoustic measures of inspiratory phonation”, *J. of Voice* 13, 389-402 (1999)

[4] R. Orlikoff, R. Baken, D. Kraus, “Acoustic and physiologic characteristics of inspiratory phonation”, *J. Acoust. Soc. Am.* 102, 1838-1845 (1997)

[5] R. Williams, I. Farquharson, J. Anthony, “Fiberoptic laryngoscopy in the assessment of laryngeal disorder”, *J. of Laryngology and Otology* 89, 299-306 (1975)

[6] A. Aronson, “Clinical voice disorders”, *New York: Thieme Inc.* 3rd edition (1990)

[7] N. Gavriely, Y. Palti, G. Alroy, J. Grotberg, “Measurement and theory of wheezing breath sounds”, *J. of Applied Physiology* 57, 481-492 (1984)

[8] E. Castillo-Guerra, “A modern Approach to Dysarthria Classification”, *Ph.D. Thesis,* University of New Brunswick (2004)

[9] Kay Elemetrics Corp., “Disordered speech database”, Massachusetts Eye and Ear Infirmary, *Voice and Speech Lab*, Boston M.A., ver. 1.03 (1994)

[10] F.L. Darley, A.E. Aronson, J.R. Brown, “Motor speech disorders”, *W.B. Saunders* Philadelphia, PA. (1975)

[11] E. Castillo-Guerra, N.L. Mendez, “Methodology to obtain a Spanish database of dysarthric speech”, *Proceedings of VII International Symposia of Social Communication*, Santiago de Cuba, (2001)

[12] Minitab, Inc., “MinitabTM statistical software”, ver. 1.1, www.minitab.com (2006)

[13] R. J. Baken, “Clinical measures of speech and voice”, *Singular Publishing Group*, Inc. (1996)

[14] R. R. Hocking, “A Biometrics Invited Paper: The Analysis and Selection of Variables in Linear Regression”, *Biometrics* 32, 1-49 (1976)