

Expert System-Type Approach to Voice Disorders: Scheduling Botulinum Toxin Treatment for Adductor Spasmodic Dysphonia

Anthony P. Salvatore,
Amitava Biswas,
Vladik Kreinovich,
Bertha Manriquez
University of Texas at El Paso
El Paso, TX 79968, USA
Emails: {asalvatore,vladik}@utep.edu

Michael P. Cannito
Speech and Hearing Center
School of Audiology and
Speech-Language Pathology
University of Memphis
Memphis, TN 38105, USA
Email: mcannito@memphis.edu

Robert J. Sinard
Department of Otolaryngology
University of Texas
Southwestern Medical Center at Dallas
5323 Harry Hines Boulevard
Dallas, TX 75390, USA

Abstract—One of the most debilitating disorders is adductor spasmodic dysphonia (ADSD), a voice disorder caused by involuntary movements of the muscles of the larynx (voice box). For treating ADSD, botulinum toxin (BT) injections turned out to be very useful. However, the effects of BT are highly variable, so at present, there is no objective criterion of when such a BT treatment is necessary. It is therefore desirable to develop such a criterion.

In this paper, we show that traditional statistical techniques are unable to generate such a criterion, while a natural expert system approach seems to be capable of generating reasonably simple rules that determine when a BT treatment is necessary.

I. INTRODUCTION

A. Clinical Condition

In this paper, we analyze a voice disorder called *adductor spasmodic dysphonia* (ADSD). Spasmodic dysphonia (SD) is a voice disorder caused by involuntary movements of the muscles of the larynx (voice box). Patients with SD have occasional difficulty saying words and/or experience sufficient difficulty that interfere with communication. Spasmodic dysphonia causes the voice to break or to have a tight, strained, or strangled quality.

ADSD is one of the three main types of SD, when sudden involuntary muscle movements or spasms cause the vocal folds (or vocal cords) to slam together and stiffen. These spasms make it difficult for the vocal folds to vibrate and produce voice. Words are often cut off or difficult to start because of the muscle spasms. As a result, the voice of an individual with adductor spasmodic dysphonia is commonly described as strained or strangled and full of effort.

B. Treatment

It is known that botulinum toxin (BT) injections of the vocal folds improve the speech of ADSD patients; see, e.g., [8].

Not only do patients themselves feel the improvement, this improvement can also be detected – in a blind-control situation – both by measurements and by expert listeners [3], [7].

Specifically, expert listeners “perceived significant improvement in ADSD connected speech at one month following botulinum toxin injection relative to pre-injection levels” [3].

C. Problem

The problem with BT treatment is that while there have been studies evaluating pre- and post-BT treatment changes in the voice and speech fluency of ADSD individuals (see, e.g., [3], [7], [8]), the nature of speech stability of individuals with spasmodic dysphonia (SD) treated with botulinum toxin (BT) is still poorly understood.

The situation is made even more complex by the fact that the severity of SD varies from time to time and from place to place; see, e.g., [4].

As a result, a patient him/herself, in collaboration with a clinician, decided when it is time to get a BT injection. This is typically based on subjective impressions of vocal quality and the amount of speaking effort. For example, [15] concludes that “objective parameters used to measure vocal function may not adequately reflect the handicapped experienced by the patient.”

There is no known method for determining when such an injection is necessary based on the measurable characteristics of the patient’s speech. The objective of this research is to describe a methodology which is, in our opinion, capable of providing such a method.

II. WHICH SPEECH CHARACTERISTICS SHOULD WE USE

Not only it is difficult, based on the objective measurements of speech, to determine when a patient needs a BT treatment, it is sometimes difficult even to determine, based on the measured speech characteristics, whether a patient has a speech disorder at all and whether this speech disorder is SD or some other type of speech disorder.

For example, in [6], it was proposed that the speaking rate may be one of the measures that provide us with an

information on whether an individual has a SD (and, correspondingly, whether a BT treatment is needed). However, by comparing the speaking rate of normal and SD speakers, the authors of [2] found out that speech disfluency (\approx stuttering) is “not a defining feature of SD”, although “it does contribute significantly to the overall clinical impression of severity of the disorder.”

A similar conclusion was reached in [13] with respect to other speech characteristics such as vowel prolongation.

While each individual characteristic is not sufficient to distinguish between normal and SD individuals, a *combination* of these characteristics helps. Specifically, in [11], it was shown that it is possible to differentiate between normal speakers and SD individuals using the following characteristics:

- total words per minute,
- disfluencies per 100 words,
- total inter-word interval,
- total articulation time,
- total reading errors, and
- total speaking time.

(This selection is also justified by the fact that while, as we have mentioned, the severity of SD varies from time to time and from place to place, the above characteristics remain reasonably stable [4].)

In a follow-up study [12], it was shown that the same six fluency measures can not only differentiate between normal and SD individuals, they also enable us to differentiate between normal, SD, and stuttering individuals.

This differentiation success supports the position that the above six fluency measures reflect the status of the speech behaviors that characterized a variety of disorders such as SD.

Therefore, in this research, we use the same six fluency measures to determine when a BT treatment should be applied.

III. EXPERIMENT: DESCRIPTION

A. Participant

The participant was a 54 year-old male when the study began. He was diagnosed as adductor spasmodic dysphonic by a speech language pathologist and an otolaryngologist approximately one year before this study was initiated. He had three BT injections during the year preceding this study. He did not exhibit nor was he diagnosed with any neurological disorders; he did not have a history of any psychiatric disorder. During the duration of this study, he was reporting that he was in good health.

He is employed as an insurance agent, is successful and carries on a very active life style. The participant did not receive any directed behavioral treatment for his vocal disorder during this study other than some periodic counseling following his reading of “The Rainbow Passage” [5] suggesting that he should project his voice when he felt laryngeal tightness. This attempt to project his voice is a vocal behavior that does not affect the corresponding speech disorder.

He had to travel to another city for his BT injections. He resided in the same city where the speech recordings were

made. He signed a consent form approved by the University of Texas at El Paso Institutional Review Board.

B. Recording Procedure

Recordings of the participant’s speech were made approximately one week before a scheduled injection, one week after an injection, and monthly after the injection. There was some variability in this schedule due to job demands, vacations, and holidays.

The participant was given a large-print version of the first paragraph (98 words) of “The Rainbow Passage” [5]. The participant was instructed to read the passage aloud using his normal speaking pattern. The audio recording of his reading was carried out in an IAC booth. A Marantz PMD430 audio recorder and a Realistic dynamic omni-directional microphone placed approximately 15 cm from the participant’s mouth were used to record the participant’s readings.

C. Temporal Acoustic Analysis

The primary software/hardware system for signal analysis was the Computerized Speech Lab (CSL) (Kay Elemetrics). Each audio recording of each passage reading was digitized at a sampling frequency of 20 KHz using commercial software and 16-bit hardware with anti-aliasing at 10 KHz. A constant input intensity level was maintained across samples during digitization to preserve inter-sample intensity relationships. The passages were then analyzed using the CSL.

A graduate student in speech-language pathology who had experience in acoustical analysis evaluated the temporal characteristics of the 24 readings. The following fluency measures were analyzed and calculated:

- total pause time,
- total articulation time,
- words per minute,
- frequency of speech disfluencies, and
- total speaking time.

Specifically:

- Total speaking time (TST) was determined by positioning the left vertical time cursor at the onset of acoustic energy for the first word of the reading and the right vertical time cursor at the offset of acoustic energy for the last word in the passage.
- The syntactic pause time (SP) was determined by measuring the duration of all syntactic inter-word intervals at or exceeding 20 ms in duration. The measured results were then tabulated and summed.
- The non-syntactic pause time (NSP) was determined by measuring the duration of all non-syntactic inter-word intervals at or exceeding 20 ms in duration. The measured results were then tabulated and summed.

These measurements were made with a millisecond accuracy.

D. Counting and the Resulting Analysis

We also counted:

- the number NN of non-syntactic pauses,
- the number NS of syntactic pauses,

- the number NE of reading errors, and
- the number of disfluencies ND.

Since the text consisted of 98 words, these numbers almost coincide with the corresponding frequencies – e.g., ND almost coincides with the disfluency frequency that is defined as the number of disfluencies per 100 words.

Based on these measured and counted values, several other characteristics were also determined:

- The average NSP time (AN) was computed by dividing the total NSP time by the number NN of non-syntactic pauses.
- The average SP time (AS) was computed by dividing the total SP time by the number NS of syntactic pauses.
- The total inter-word interval (pause) time (TPT) was determined as the sum of SP and NSP pause times.
- The total articulation time (TAT) was determined by subtracting total inter-word interval time/pause time TPT from total speaking time TST.
- The number of words per-minute WPM was calculated by dividing the total number of words read (98) by the total speaking time (TST) in minutes.

IV. RESULTS

As a result, we got the following 10 measurement records, 6 immediately pre-botox (Y1–Y6), when the patient felt that a BT treatment is necessary, and 4 post-botox (N1–N4):

No.	date	NSP	NN	AS	SP	NS	AS
Y1	03/26/01	1.57	16	0.10	6.36	10	0.64
Y2	08/20/01	0.98	10	0.10	4.61	9	0.51
Y3	04/22/02	0.07	11	0.01	5.88	9	0.65
Y4	08/27/02	0.43	6	0.07	5.60	9	0.62
Y5	02/11/03	0.33	8	0.04	4.77	9	0.53
Y6	05/12/03	0.50	7	0.07	5.70	7	0.81
N1	05/04/01	1.66	11	0.15	6.66	9	0.74
N2	09/24/01	1.03	11	0.09	8.50	10	0.85
N3	10/02/02	1.93	12	0.16	6.39	10	0.64
N4	03/11/03	0.04	9	0.00	5.81	7	0.83

No.	TST	TPT	TAT	WPM	NE	ND
Y1	41.6	7.93	33.7	141	2	0
Y2	39.4	5.59	33.8	149	1	1
Y3	36.5	5.95	30.5	161	1	0
Y4	33.6	6.03	27.5	175	0	0
Y5	32.4	5.11	27.3	182	1	0
Y6	36.7	6.19	30.6	160	1	0
N1	36.2	8.32	27.8	163	1	0
N2	36.1	9.53	26.5	163	1	0
N3	34.7	8.31	26.4	170	2	0
N4	34.7	5.85	28.8	169	0	0

V. TRADITIONAL STATISTICAL APPROACH DID NOT WORK WELL

A. General Idea

Our objective is to differentiate between pre- and post-botox situations. The traditional statistical approach to such a differentiation would be to find a linear or non-linear discrimination function based on these measured results.

B. Linear Approach

First, we applied, to the above data, statistical software that provides linear discrimination, i.e., that looks for a linear expression $c(x) = c_0 + c_1 \cdot x_1 + \dots + c_n \cdot x_n$ of the measured values that is positive for all pre-botox values and negative for all post-botox points.

Alas, the best linear separation correctly classified only 8 out of 10 points: readings Y4 and Y5 were erroneously classified as post-botox.

C. Quadratic Approach

In principle, we could extend this procedure to use quadratic discrimination techniques, i.e., techniques in which we look for a quadratic discrimination expression

$$c(x) = c_0 + \sum_{i=1}^n c_i \cdot x_i + \sum_{i=1}^n \sum_{j=1}^n c_{ij} \cdot x_i \cdot x_j.$$

The problem with this approach is that even for $n = 6$ directly measured and counted variables, we need $1 + 6 + 6 \cdot (6 + 1)/2 = 28$ parameters to find a general quadratic discrimination function, and we only have 10 measurement results – i.e., 10 equations to determine these parameters.

Of course, we can always fit 28 parameters so that they satisfy 10 equations – but there are many such fits, and each of them will be a purely mathematical fit which says nothing about the actual separation between pre- and post-botox cases.

VI. NEW IDEA: EXPERT SYSTEM APPROACH

A. Expert System Approach: General Idea

Since the traditional statistical approach does not work well, we decided to apply a new approach motivated by expert systems. In expert systems, relations between quantities are usually described not in terms of numerical relations, but rather in terms of rules, like “if the robot is too close to the obstacle, it should slow down”. The simplest interpretation of this rule is that we select some threshold for distance, so that:

- if the distance exceeds the selected threshold, the robot should continue its movement, while
- if the distance is smaller than the selected threshold, the robot should slow down.

Usually, there are several rules. For example, for the robot control, if there is an obstacle nearby, then we should slow down. However, if the robot is far from all the obstacles, it does not necessarily mean that we should continue moving: there may be other reasons to slow down – e.g., if the target destination is close.

So, in general, we must consider a hierarchy of rules. Let us discuss how we can generate these rules.

B. Formulation of the Problem

Suppose that we have the table consisting of several records (corresponding to several objects).

There are two classes of objects. We know, for each record, whether this record describes an object of Class 1 or an object of Class 2. Each record contains n numbers x_1, \dots, x_n that characterize the corresponding object. Our objective is to find the rules that classify the objects into classes based on the corresponding values x_i .

C. New Idea: Motivation

The ideal case is when a single rule will suffice, i.e., when it is sufficient to select a single variable x_i , a threshold t_i , and then classify the object to Class 1 or Class 2 depending on whether the value of x_i is larger or smaller than the threshold.

This is a rare situation, so what we can do instead is find the variable x_i and a threshold t_i for which the corresponding threshold-based rule covers as many objects as possible. This will be our first classification rule.

Then, we consider two new situations:

- a situation in which $x_i < t_i$, and
- a situation in which $x_i > t_i$.

For each of these situations, we select a subtable consisting of only the objects for which $x_i < t_i$ (correspondingly, only the objects for which $x_i > t_i$). For each of these subtables, we select a new variable x_j and a new threshold t_j – in general, different variables x_j and different threshold t_j for different subtables. This will be our second rule.

For each subtable, the second rule divides the subtable into two new sub-subtables, etc.

At the end, we get a hierarchical tree of rules that provide the desired classification.

D. Applying This Idea to the Patient Data Table

By testing all the measured, counted, and computed values x_i described in the above tables, we can see that the best classification happens when we select TST as the classifying parameter x_i and select the value 36.3 as the corresponding threshold. Specifically:

- when the total speaking time TST exceeds 36.3, then we are guaranteed to be in a pre-botox situation;
- when TST is smaller than 36.3, then we probably are in a post-botox situation, but there are two cases when it is not so: cases Y4 and Y5 (by the way, the same cases for which linear discrimination did not work).

So, our first rule is: if TST is larger than 36.3, apply botox, otherwise, we need a second rule.

To derive the second rule, we need to consider all the records for which TST is smaller than 36.3, i.e., we need to consider a subtable consisting of all the post-botox (N) records plus the records Y4 and Y5. For this subtable, we again consider all possible variables x_j and all possible thresholds t_j . It turns out that there exists a variable x_j and a threshold t_j that enables us to completely classify this subtable: namely, as x_j , we can take the average syntactic pause time AS, and as the threshold, we can take $t_j = 0.63$. Indeed:

- the value of AS exceeds 0.63 for all post-botox records, and
- the value of AS is smaller than 0.63 for both pre-botox records Y4 and Y5 from this subtable.

As a result, we arrive at the following classification rules:

E. Resulting Classification Rules

To check whether a patient needs a botox injection, we must use the total speaking time TST and the average syntactic pause time AS:

- if TST exceeds 36.3, the patient needs an injection;
- otherwise, the patient needs an injection if and only if AS is smaller than 0.63.

These two simple rules use only two variables and perfectly describe all the measurement results.

VII. CONCLUSION AND FUTURE WORK

Botulinum toxin (BT) injections are a useful tool for taking care of the adductor spasmodic dysphonia – a rather severe voice disorder. Unfortunately, little is known about the exact effect of BT, so at present, there is no objective criterion of when such a BT treatment is necessary. In this paper, we have shown that it is possible to produce simple rules that describe when to use and when not to use the BT treatment for a given patient. To derive these rules, we used an expert system-type approach.

Of course, the exact thresholds are currently based on a single patient. More experimental data is needed to describe how these thresholds will vary across patients (and what other rules are necessary for other patients).

VIII. ACKNOWLEDGMENTS

This work was supported in part by NASA under cooperative agreement NCC5-209, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, by the Army Research Laboratories grant DATM-05-02-C-0046, and by the NIH grant 3T34GM008048-20S1.

The authors are thankful to the anonymous referees for valuable suggestions.

REFERENCES

- [1] Aronson, A.E., McCaffrey, T.V., Litchy, W.J., and Lipton, R.J. (1993). *Laryngoscope*, Vol. 103, pp. 683–692.
- [2] Cannito, M.P., Burch, A.R., Wats, C., Rappold, P.W., Hood, S.B., and Sherrard, K. (1997). Disfluency in Spasmodic Dysphonia: A Multivariate Analysis, *Journal of Speech, Language and Hearing Research*, Vol. 40, pp. 627–641.
- [3] Cannito, M.P., Woodson, G.E., and Murry, T. (1999) Perceptual Scaling of Spasmodic Dysphonia Before and After Botulinum Toxin Treatment, In: Schutte, H.K., Dejonckere, P., Leezenberg, Mondelaers, B., and Peters, H.F.M. (eds.), *Communication and Its Disorders: A Science in Progress*, Nijmegen University Press, pp. 161–163.
- [4] Cimino-Knight, A.M., and Sapienza, C.M. (2001). Consistency of Voice Produced by Patients With Adductor Spasmodic Dysphonia: A Preliminary Investigation. *Journal of Speech, Language, and Hearing Research*, Vol. 44, pp. 793–802.

- [5] Fairbanks, G. (1960) *Voice and Articulation Drillbook*, New York: Harper & Row.
- [6] Ford, C.N., Bless, D.M., and Patel, N.Y. (1992). Botulinum Toxin Treatment of Spasmodic Dysphonia: Techniques, Indications, Efficacy, *Journal of Voice*, Vol. 6, No. 4, pp. 370–376.
- [7] Langeveld, T.P.M., Houtman, E. H., Briaire, J.J., van Rossum, M., Zwinderman, A.H., and Baatenburg De Jong, R.J. (2001). Evaluation of Voice Quality in Adductor Spasmodic Dysphonia Before and After Botulinum Toxin Treatment. *Ann. Otolaryngol.*, Vol. 110, pp. 627–634.
- [8] Ludlow, C.L., Naunton, R.F., Terada, S., and Anderson, B.J. (1991). Successful treatment of selected cases of abductor spasmodic dysphonia using botulinum toxin injection. *Otolaryngology-Head and Neck Surgery*, Vol. 104, No. 6, pp. 849–855.
- [9] Lundy, D.S., Ling Lu, F., Casiano, R.R., and Xue, J.W. (1998). The Effect of Patient Factors on Response Outcomes to Botox Treatment of Spasmodic Dysphonia, *Journal of Voice*, Vol. 12, No. 4, pp. 460–466.
- [10] Rosenthal, R., and Rosnow, R.L. (1991). *Essentials of Behavioral Research: Methods and Data Analysis*, New York City: McGraw Hill, Inc.
- [11] Salvatore, A.P., Cannito, M.P., and Gutierrez, G.S. (1999). Spasmodic Dysphonia: A Neural Net Analysis, *Journal of Medical Speech-Language Pathology*, Vol. 7, No. 2, pp. 169–174.
- [12] Salvatore, A.P., Cannito, M.P., Biswas, A., Ingham, R., Bender, B.K., and Manriquez, B. (2002) Artificial Neural Network Categorization of Speech Fluency Measures Across Speech Disorders, *Abstracts of the Motor Speech Conference*, Williamsburg, VA.
- [13] Sapienza, C.M., Walton, S., and Murry, T. (1999). Acoustic Variations in Adductor Spasmodic Dysphonia as a Function of Speech Task. *Journal of Speech, Language, and Hearing Research*, Vol. 42, pp. 127–140.
- [14] Truong, D.D., Rontal, M., Rolnick, M., Aronson, A.E., and Mistua, K. (1991). Double-Blind Controlled Study of Botulinum Toxin in Adductor Spasmodic Dysphonia, *Laryngoscope*, Vol. 101, No. 1, pp. 630–634.
- [15] Woodson, G.E., Zwirner, P., Murry, T., and Swenson, M.R. (1992). Functional Assessment of Patients With Spasmodic Dysphonia. *Journal of Voice*, Vol. 6, No. 4, pp. 338–343.
- [16] Zwirner, P., Murry, T., Swenson, M., and Woodson, G.E. (1992). Effects of Botulinum Toxin Therapy in Patients With Adductor Spasmodic Dysphonia: Acoustic, Aerodynamic, and Videendoscopic Findings. *Laryngoscope*, Vol. 102, pp. 400–406.