

Long-term Follow-up of Patients with Spasmodic Dysphonia Repeatedly Treated with Botulinum Toxin Injections

¹PH Dejonckere, ²C Manfredi

¹Department of Otolaryngology, University Medical Center Utrecht, Utrecht, Netherlands, Neurosciences Catholic University of Leuven Belgium, Federal Institute of Occupational Diseases, Brussel

²Department of Electronics and Telecommunications, Università degli Studi di Firenze, Italy

Correspondence: PH Dejonckere, Department of Otolaryngology (Phoniatics), University Medical Center Utrecht, AZU F.02.504 Po Box 85500, NL 3508GA, Utrecht, Netherlands, Phone: 0031887557729, e-mail: ph.dejonckere@umcutrecht.nl

ABSTRACT

Adductor spasmodic dysphonia (SD) is a focal laryngeal dystonia mainly resulting in a strained voice quality with spastic voice breaks and frequency shifts, perturbing fluency and intelligibility. SD-patients report unusually high impairment of their quality of life. The standard treatment is botulinum toxin injection in the thyroarytenoid muscles, in order to interfere with the perturbed sensory feedback loop of kinetic muscle tension regulation. The globally favorable effects are temporary, but the botulinum injections can be repeated. There is a lack of information about long-term effects. This is the first study investigating effects over several years, and comparing self-evaluation of patients with objective multimodal acoustic analysis. Results show that 72% of the individual injections are successful. The effects of botulinum are not reduced after repeated injections. In contrary, the self-perceived improvement increases in average over time. When self-evaluations preinjection are considered, patients tend to evaluate their voice and their handicap as worsening over time. This contrasts with the results of multimodal acoustic analysis. Objective data reveal a relative stability over time for as well pre- as postinjection. This seems to indicate that there is no shift over time in the objective severity of deviance in voice quality.

Keywords: Spasmodic dysphonia, Long-term assessment, Self-evaluation, Acoustic analysis.

INTRODUCTION

Adductor spasmodic dysphonia (SD) is a focal laryngeal dystonia mainly resulting in a strained voice quality with spastic voice breaks and frequency shifts, perturbing fluency and intelligibility.⁸ The current standard recommended treatment is Botulinum Toxin (BT) injection in the thyroarytenoid muscles, in order to interfere with the perturbed sensory feedback loop of kinetic muscle tension regulation.¹⁰ The mode of action of this toxin is at cholinergic nerve terminals where it inhibits the release of acetylcholine. However, the effects are only temporary in part because of the formation of remodeled neuromuscular junctions after a few months, but the botulinum injections can be repeated. Assessment of treatment efficacy in SD is difficult as well objectively as subjectively. Watts et al (2006) found that the evidence from randomized controlled trials supporting the effectiveness of botulinum toxin for management of spasmodic dysphonia was still deficient.¹³ Three approaches are available: Perceptual evaluation by expert listeners, acoustic measurements and self-evaluation by the patient. Cannito et al (2004) compared perceptual evaluation pre- and postbotulinum injection.² Voice quality and fluency improved for most patients following treatment. However, Braden et al (2010) found only a very weak correlation between the patient's assessment of voice impairment and the clinician's perceptual judgment of voice impairment.¹ It is well known that SD-patients report

unusually high scores on the Voice Handicap Index⁴—a widespread instrument for measuring the psychosocial handicapping effects of a voice disorder—as they experience their disease as seriously impairing their quality of life.⁷ Achieving advanced acoustic analysis, Sapienza et al (2002) showed that the number of atypical acoustic events decreased following botulinum injection. However, deviant acoustic characteristics as strong aperiodicity, phonatory breaks and frequency shifts, considered as typical for SD, cannot be adequately analyzed by traditional software programs, particularly in running speech.³ Using a specific acoustic analysis program, Siemons-Lühning et al (2009) demonstrated a quite satisfactory correlation between perceptual ratings and acoustic measurements.¹¹ However, there is still a lack of information about (very) long-term effects in patients receiving repeated botulinum injections. This is the first study investigating effects over several years in repeatedly treated patients, and comparing self-evaluation of patients with objective multimodal acoustic analysis.

MATERIALS AND METHODS

In the current study, covering the period 1992 to 2009, long-term evolution is analysed in 19 patients (11 females and 9 males) having been injected with botulinum toxin (Botox Allergan: 5 units/vocal fold) between 3 and 22 times over

periods of 2 to 16 years. All these patients asked for repeated injections. Our approach is based on:

1. A patient's self-evaluation on two visual analog scales: Voice quality itself and social handicap. These two scores in this study are averaged. Each patient was asked to give his/her evaluation just before the injection, and between 1 and 4 months after the injection.
2. An acoustic analysis is achieved with a computerized program for signal analysis that is suited for irregular voices, and that mainly deals with voicing and 'cycle-to-cycle aperiodicity' (Jitter) criteria.^{5,6} Material is a standard, neutral, phonetically selected constantly voiced sentence. As the sentence is phonetically selected for being constantly voiced, some face validity for the relevance of the parameters may be accepted: The more voicing, the better the voice quality.

The analysis program "Ampex" (Auditory Model Based Pitch Extractor) was created (1992) and further developed by Van Immerseel and Martens was used for the acoustic measurements.¹² It has proven to be able to extract in a valid way the period in irregular signals with background noise. It also detects low frequency components (< 0.1 KHz), is suited for running speech and has been efficiently used for substitution voices and for spasmodic dysphonia.⁵ The major advantage of this program is that it includes the three deviant acoustic events that were found relevant for characterizing SD: Aperiodicity, phonatory breaks and frequency shifts⁹ without requesting subjective intervention of an experimenter for placing cursors and identifying deviant events, as in the experiment of Sapienza et al (2002).

The acoustic analysis is performed in three stages. In the first stage, short-term acoustic features are extracted every 10 ms by the auditory model described in Van Immerseel and Martens (1992).¹² Then these features are employed to distinguish speech frames from background (silence) frames. Finally, a global analysis of the short-term acoustic feature patterns over the entire recording is performed to produce a limited set of features that are expected to characterize the voice of the recorded speaker.

Every 10 ms, the auditory model produces a set of more than 30 features, but for the present study, only four of them are relevant, namely, the energy (E), the voicing evidence (VE), the voiced/ unvoiced nature (VU) and the pitch frequency (F_0) (in case of voicing) of the frame.

The speech/background classification of the frames is based on an analysis of the smoothed energy pattern. The smoothed energy of frame i is computed as the mean of the energies in frames $i - 2$ to $i + 2$. In the first step, a background threshold is determined as 1.1 times the minimal energy and 0.05 times the maximum energy found in the recording. All frames exceeding this threshold are initially labeled as speech and the others as background. However, to avoid that too many weak parts of speech (e.g. closures of plosives, weak consonants) are classified as background, any interval shorter than 100 ms that was labeled as background is converted to speech again.

The first feature emerging from the global analysis stage characterizes the ability of the speaker to produce voicing. It comes in two flavors: The proportion of voiced frames (PVF) in the entire recording and the proportion of voiced speech frames (PVS). Because pauses and weak speech sounds are typically unvoiced, PVS is expected to be larger than PVF.

The second feature is the average voicing evidence (AVE) in the voiced frames. It characterizes the degree of regularity/periodicity in the voiced frames. Since, the real background frames are normally unvoiced the analysis is performed on all frames, and not just on the speech frames, in the hope to be more robust against possible errors of the speech/background classification which is after all purely energy based whereas the voicing evidence is derived from an analysis of all the subband signals created by the auditory model.

The third feature being assessed is the traditional 'Jitter'. Jit represents the F_0 -jitter in all voiced frame pairs (= 2 consecutive frames). The corrected jitter goes even one step further and reports the average jitter only in frames with a "reliable" F_0 estimate. The vocal frequency estimate F_0 is designated reliable if it deviates less than 25 % from the average over all voiced frames. The formula which is used to compute the jitter is:

$$\text{Jitter} = \frac{\text{sum of VE}(i) * |T_0(i) - T_0(i-1)|}{\text{sum of VE}(i) * T_0(i-1), T_0 = 1/F_0}$$

A fourth and last feature is the 90th percentile (VL 90) of the voicing length distribution. It is considered to be a robust estimate of the maximum voicing duration. The voicing length is defined as the number of consecutive voiced frames in the data.

Summarizing, the following features have been estimated:

- *PVF/PVS*: PVF is the proportion of voiced frames and depends on the pauses appearing in speech. Also the PVS, the proportion of voiced speech frames is computed, thus considering only frames that are classified as speech in the first step of the analysis. Since pauses and weak sounds are typically unvoiced, PVS will typically be larger than PVF. For vowels it should be expected that PVS = 100%: The better the voice, the highest the percentages
- *AVE*: The average voicing evidence. The more regular (periodic) the voice frames are, the higher the AVE will be
- *VL 90 parameter*: The 90th percentile of the voicing length distribution as a robust estimate of the maximum voicing duration. The voicing length is defined as the number of consecutive voiced frames found in the data. Phonatory breaks reduce this parameter
- *JIT and JITc*: The cycle-to-cycle period perturbation and the corrected cycle-to-cycle period perturbation. Better voices show limited jitter
- *PVFU*: The percentage of frames with "unreliable" F_0 is considered as a second F_0 -instability factor. Frequency shifts make F_0 unreliable
- *DUR (duration)*: Total time used for reading the sentence. Improvement in fluency shortens duration.

As acoustic parameters have different units, a z-transformation was first achieved (and, when relevant, a sign inversion) before creating a global acoustic score combining all 9 parameters without weighting. To illustrate this, Figure 1 shows the globally averaged percentage of pre-post variation (y-axis) for each acoustic parameter.

In average, the postinjection voice demonstrates on the one hand an increase in voicing length, a higher proportion of voiced frames and voiced speech frames and a higher voicing evidence. On the other hand, duration needed for pronouncing the sentence, jitter and percentage of frames with unreliable F_0 decrease.

RESULTS AND DISCUSSION

1. Moments of treatment clearly determine—as a general rule—a saw teeth is as well objective as subjective parameters. There is no statistically significant reduction of time delay between the injections over time, but all patients were informed about the limited duration of the effect.
2. Each of the 19 patients shows an average—overall of his/her treatments—improvement as well acoustically as self-rated. However, not every injection was successful, the global percentage of successful injections is 72% (subjective + objective improvement). However, the current study can obviously not be considered as a treatment-efficacy study, as all included patients were asking themselves for repeated botulinum injections, suggesting any perceived benefit.
3. Globally (i.e. considering as well the pre- as the post-score), over time, the combined acoustic score tends to less deviance, while self-assessment score tends to more deviance. The two slopes for all data of the 19 patients significantly differ ($p < 0.05$).
4. The slope of the preinjection self-evaluations over time is an average significantly steeper than that one for the self-evaluations postinjection (Fig. 2). This means that in general,

the subjective effect of a new injection is rated as stronger over time, and this effect is due to a worsening of the self-evaluation preinjection. The self-evaluation scores after injection remain remarkably stable.

5. The slope of the preinjection combined acoustic scores over time does in average not differ from the slope of the combined acoustic scores postinjection (Fig. 3). This indicates that there is no shift over time in the objective severity of deviance in voice quality. The pathology does not show progression in severity, but there is also no cure.

CONCLUSION

Long-term evolution is analysed in 19 patients (11 females and 9 males) having been injected with botulinum toxin (Botox Allergan: 5 units/vocal fold) between 3 and 22 times over periods of 2 to 16 years. All these patients asked for repeated injections.

72 % of all individual injections appear to be successful as well objectively (acoustic analysis) as subjectively (patient's self rating).

The effects of botulinum are not reduced after repeated injections. In contrary, the self-perceived improvement after an injection increases in average over time. This effect is due to a slight worsening of the self-evaluation preinjection. When only self-evaluations preinjection are considered, patients tend to evaluate their voice and their handicap as worsening over time. The self-evaluation scores after injection remain remarkably stable.

This contrasts with the results of acoustic analysis, including measurements of voicing, F_0 -irregularity and fluency. Objective data reveal a relative stability of voice quality over time in as well pre- as postinjection, within most cases an improvement after each individual injection of botulinum toxin. This seems to indicate that in repeatedly injected patients, there is no shift over time in the objective severity of deviance in voice quality. The pathology does not show progression in severity, but there is also no cure.

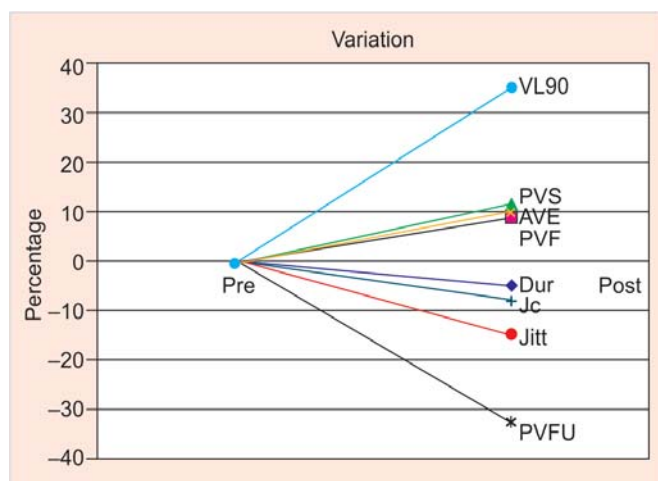


Fig. 1: Globally averaged percentage of pre-post variation (y-axis) for each acoustic parameter

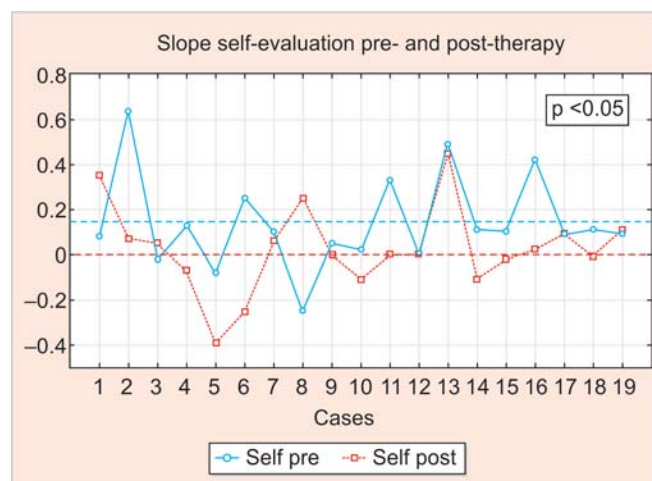


Fig. 2: Slopes of the pre- and postinjection self-evaluations over time for all 19 patients

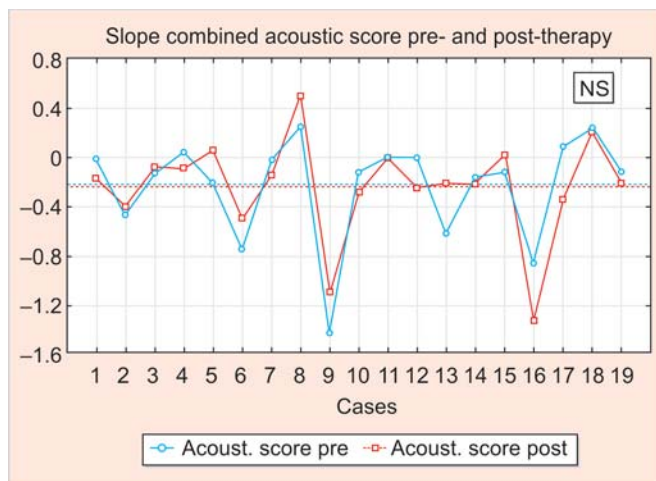


Fig. 3: Slopes of the pre- and postinjection combined acoustic scores over time for all 19 patients

REFERENCES

1. Braden MN, Johns MM, Klein AM, Delgaudio JM, Gilman M, Hapner ER. Assessing the effectiveness of botulinum toxin injections for adductor spasmodic dysphonia: Clinician and patient perception. *J Voice* 2010;24:242-49.
2. Cannito MP, Woodson GE, Murry T, Bender B. Perceptual analyses of spasmodic dysphonia before and after treatment. *Arch Otolaryngol Head Neck Surg* 2004;130:1393-99.
3. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessments techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol* 2001;258:77-82.
4. Jacobson BH, Johnson A, Grywalsky C, Silbergleit A, Jacobson G, Benninger MS, Newman CW. The voice handicap index

(VHI): Development and validation. *American J of Speech and Langu Pathol* 1997;6:66-70.

5. Moerman MBJ, Pieters G, Martens JP, Van der Borgt MJ, Dejonckere PH. Objective evaluation of quality of substitution voices. *Eur Arch Otorhinolaryngol* 2004;261(10):541-47.
6. Moerman MBJ, Martens JP, Van der Borgt MJ, Peleman M, Gillis M, Dejonckere PH. Perceptual evaluation of substitution voices: Development and evaluation of the (I). INFVo rating scale. *Eur Arch Otorhinolaryngol* 2006;263(2):183-87.
7. Moerman MBJ, Dejonckere PH, Liefink AF. Comparing voice pathologies with the voice handicap inventory: Is a weighting factor required? www.siicsalud.com/des/expertos.php/89282, 2008.
8. Rees JC, Blalock PD, Kemp SE, Halum SL, Koufman JA. Differentiation of adductor-type spasmodic dysphonia from muscle tension dysphonia by spectral analysis. *Otolaryngology—Head and Neck Surgery* 2007;137:576-81.
9. Sapienza CM, Cannito MP, Murry T, Branski R, Woodson G. Acoustic variations in reading produced by speakers with spasmodic dysphonia pre-Botox injection and within early stages of post-Botox injection. *J Speech Language Hearing Res* 2002;45:830-43.
10. Schwartz SR, Cohen SM, Dailey SH, Rosenfeld RM, Deutsch ES, Gillespie MB, et al. Clinical practice guideline: Hoarseness (dysphonia). *Otolaryngol Head Neck Surg* 2009;141(3 Suppl 2):S1-31.
11. Siemons-Lühring DI, Moerman M, Martens JP, Deuster D, Müller F, Dejonckere PH. Spasmodic dysphonia, perceptual and acoustic analysis: Presenting new diagnostic tools. *Eur Arch Otorhinolaryngol* 2009;266:1915-22.
12. Van Immerseel LM, Martens JP. Pitch and voiced/unvoiced determination with an auditory model. *J Acoust Soc Am* 1992;91:3511-26.
13. Watts C, Nye C, Whurr R. Botulinum toxin for treating spasmodic dysphonia (laryngeal dystonia): A systematic Cochrane review. *Clinical Rehabilitation* 2006;20(2):112-22.