Voice analysis in adductor spasmodic dysphonia: Objective diagnosis and response to botulinum toxin

Antonio Suppa\textsuperscript{a,b,1}, Francesco Asci\textsuperscript{a,1}, Giovanni Saggio\textsuperscript{c}, Luca Marsili\textsuperscript{d}, Daniele Casali\textsuperscript{e}, Zakarya Zarezadeh\textsuperscript{e}, Giovanni Ruoppolo\textsuperscript{f}, Alfredo Berardelli\textsuperscript{a,b,\ast}, Giovanni Costantini\textsuperscript{c}

\textsuperscript{a} Department of Human Neurosciences, Sapienza University of Rome, Italy
\textsuperscript{b} IRCCS Neuromed Institute, Pozzilli, IS, Italy
\textsuperscript{c} Department of Electronic Engineering, University of Rome Tor Vergata, Italy
\textsuperscript{d} Gardner Family Center for Parkinson’s Disease and Movement Disorders, Department of Neurology, University of Cincinnati, Cincinnati, OH, USA
\textsuperscript{e} Department of History, Humanities and Society, University of Rome Tor Vergata, Italy
\textsuperscript{f} Department of Sense Organs, Otorhinolaryngology Section, Sapienza University of Rome, Italy

\section*{ARTICLE INFO}

Keywords:
Adductor-type spasmodic dysphonia
Cepstral analysis
Machine-learning
Botulinum toxin

\section*{ABSTRACT}

\textbf{Introduction:} Adductor-type spasmodic dysphonia is a task-specific focal dystonia characterized by involuntary laryngeal muscle spasms. Due to the lack of quantitative instrumental tools, voice assessment in patients with adductor-type spasmodic dysphonia is mainly based on qualitative neurologic examination. We evaluated patients with cepstral analysis and specific machine-learning algorithms and compared the results with those collected in healthy subjects. In patients, we also used cepstral analysis and machine-learning algorithms to investigate the effect of botulinum neurotoxin type A.

\textbf{Methods:} We investigated 60 patients affected by adductor-type spasmodic dysphonia before botulinum neurotoxin type A therapy and 60 age and gender-matched healthy subjects. A subgroup of 35 patients was also evaluated after botulinum neurotoxin type A therapy. We recorded the sustained emission of a vowel and a sentence by means of a high-definition audio recorder. Voice samples underwent cepstral analysis as well as machine-learning algorithm classification techniques.

\textbf{Results:} Cepstral analysis was able to differentiate between healthy subjects and patients, but receiver operating characteristic curve analysis demonstrated that machine-learning algorithms achieved better results than cepstral analysis in differentiating healthy subjects and patients affected by adductor-type spasmodic dysphonia. Similar results were obtained when differentiating patients before and after botulinum neurotoxin type A therapy. Cepstral analysis and machine-learning measures correlated with the severity of voice impairment in patients before and after botulinum neurotoxin type A therapy.

\textbf{Conclusions:} Cepstral analysis and machine-learning algorithms are new tools that offer meaningful support to clinicians in the diagnosis and treatment of adductor-type spasmodic dysphonia.

\section*{1. Introduction}

Adductor-type spasmodic dysphonia (ASD) is an adult-onset focal, task-specific laryngeal dystonia characterized by intermittent strained and strangled voice due to involuntary thyroarytenoid muscle contraction [1–7]. Currently, the diagnosis of ASD is based on neurological and phoniatric clinical examination [3,5,8,9]. ASD symptoms can be improved by botulinum neurotoxin type A (BoNT-A) injections into the vocal folds [1,4–7].

Several authors are now using voice cepstral analysis to achieve an objective examination of voice disorders including ASD [10,11]. The most commonly used outcome measure of cepstral analysis is the cepstral prominence peak (CPP) [10,12,13]. Reduced CPP values have been demonstrated in patients with several voice disorders [11,14–16].

Novel machine-learning algorithms have recently improved the classification accuracy of selected features in target variables when compared to more conventional procedures [17–19]. Machine-learning algorithms may therefore increase the accuracy of cepstral analysis.

In this study, we examined voice samples using conventional cepstral analysis by measuring CPP values in 60 ASD patients and we also
compared CPP from voice samples recorded in patients before and after BoNT-A therapy to objectively quantify the symptomatic improvement of ASD following pharmacological treatment.

Our study is the first to use advanced machine-learning algorithms to further examine and classify voice samples, with the aim of determining whether the use of these algorithms could increase the accuracy of objective ASD diagnosis and evaluation of BoNT-A response. We here used receiver operating characteristic (ROC) analysis to compare voice analysis by conventional cepstral analysis with that obtained by more recent procedures using machine-learning algorithms in 60 ASD patients.

2. Methods

2.1. Subjects

Sixty patients with ASD (9 men; age ± SD 64.1 ± 13y, [37-91y]) and 60 age and gender-matched HS (15 men; age ± SD 58.5 ± 12.1y, [33-85y]) were recruited from the Movement Disorders Clinic at the Department of Human Neurosciences, Sapienza University of Rome, Italy. All participants were native Italian-speakers. The clinical diagnosis of ASD was made according to current standardized criteria [2]. Patients underwent phoniatric evaluation and laryngoscopy. We excluded smokers and subjects with bilateral/unilateral hearing loss, respiratory disorders, and other conditions affecting the vocal cords. None of the patients enrolled in the study had any improvement of symptoms after taking oral medications or alcohol and, at the time of evaluation, none was taking any drugs acting on the central nervous system. Participant demographic and clinical features are summarized in Table 1. ASD symptoms were scored using the Italian version of the Voice Handicap Index (VHI) [8,9] and the Dysphonia Clinical Scale (DCS) [16,20]. Cognitive function and mood were evaluated using the Mini-Mental State Evaluation (MMSE) and the Hamilton Depression Scale (HAM-D).

ASD patients were chronically treated with BoNT-A injections and examined at least 6 months after the last injection. A subgroup of 35 patients (6 men; age ± SD 65 ± 11.1y, [37-86y]) was also evaluated one month after BoNT-A therapy (Table 1; Supp.1–2).

Participants gave written informed consent to the study, which was approved by the institutional review board in accordance with the Declaration of Helsinki.

2.2. Voice recordings

Participants performed two separate speech tasks: 1) produce a sustained close-mid front unrounded vowel /e/for 5 s, and 2) read the Italian sentence: “Nella casa in riva al mare Maria vide tre cani bianchi e neri”. Participants were asked to speak with their usual voice intensity, pitch, and quality [10,13]. All the recordings took place early in the morning (8:30 am to 10:00 am).

Voices were acquired using a Shure WH20 Dynamic Headset Microphone (Shure Incorporated, USA) placed at 5 cm from the mouth and recorded in “.wav” format with a high-definition audio-recorder H4n Zoom (Zoom Corporation, Tokyo, Japan) at 44.1 kHz sampling frequency and 16 bit of resolution.

2.3. Cepstral analysis

SpeechTool was used to perform cepstral analysis on voice samples [11,12]. First, the cepstrum, which represents the Fourier transform of the logarithm of the power spectrum of the signal, in the quefrency domain, was calculated. It consists of a series of peaks, called rahmonics: the first cepstral peaks are related to the spectrum envelope, while the higher peak (CPP) is related to the power and represents the difference in dB between the amplitude of the cepstral peak and the corresponding amplitude value on the linear regression line. Finally, we calculated the smoothed CPP (CPPS) averaging CPP values across shorter time windows (2 ms) [11,12].

2.4. Machine-learning analysis

The digitized voice samples were pre-processed through feature extraction using modified INTERSPEECH2016 Computational-ParalinguisticsChallenge (ComParE) included in OpenSMILE (audEERING GmbH, Germany) [21]. A predefined “configuration file” was used to extract a dataset of 6139 features, mostly belonging to energy, spectrum, zero crossing, pitch relations, and low-level descriptors [22]. To identify relevant features for acoustic analysis of speech signals [22], extracted features were pre-processed through feature selection using Weka [23]. Weka contains a collection of algorithms for data analysis and predictive modeling, including the sequential minimal optimization (SMO) method, which is a support vector machine algorithm designed to perform the classification.

Hence the SMO was trained by analyzing CPPS plus a number of features selected by the supervised attribute filter of Weka suite. We applied selected features by means of gain ratio attribute evaluation, a computational algorithm included in Weka software that evaluates the worth of an attribute by measuring the gain ratio with respect to the class [23]. We selected 31 features (CPPS + 30) according to previous studies [24]. To correlate the SMO output with patients’ clinical features, we used a feedforward Artificial Neural Network (ANN), trained using all the selected features (CPPS + 30). ANN was made of 31-input layer, a ten-neurons middle hidden layer and a one-neuron output layer. The ANN model allowed us to obtain as output, the likelihood ratio (LR), a continuous numerical value for each patient, providing a measure of the severity of voice impairment. Normalized LRs are shown in Supp.2.

2.5. Statistical analysis

The normality of the demographic and anthropometric parameters (age, gender, height, and weight) was assessed using the Kolmogorov-Smirnov test. Mann-Whitney U test was used to compare demographic and clinical scores in HS and patients. The Wilcoxon signed rank test was used to compare VHI and DCS scores in patients before and after BoNT-A therapy. Unpaired Student t-test was used to compare the CPPS values in patients before and after BoNT-A therapy and HS, whereas paired Student t-test was used to compare the same measures in patients before and after BoNT-A therapy.

ROC analyses were performed to identify the optimal diagnostic cut-off values of CPPS alone and SMO (CPPS + selected features), calculated during the emission of the vowel and the sentence, for discriminating

| Table 1 | Demographic and clinical features of ASD patients before and after BoNT-A and HS. DD: disease duration; VHI: voice handicap index; DCS: dysphonia clinical scale; MMSE: Mini-Mental State Evaluation; HAM-D: Hamilton Depression Rating Scale. CPPS scores were calculated from the emission of the vowel. Results are expressed as average ± standard deviation. Please note: VHI and DCS scores after and CPPS scores after refer to a subgroup of 35 ASD patients after BoNT-A therapy. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age (years) | Weight (Kg) | Height (cm) | DD (years) | VHI before | VHI after | DCS before | DCS after | MMSE | HAMD | CPPS before | CPPS after |
| ASD | 64.1 ± 13 | 65.3 ± 11 | 161.1 ± 16.1 | 7.7 ± 7.1 | 56.4 ± 23.2 | 41 ± 21.8 | 2.2 ± 0.7 | 1.2 ± 0.8 | 29.7 ± 0.7 | 6.8 ± 4.9 | 4.2 ± 2.6 | 5.1 ± 2.6 |
| HS | 59 ± 12 | 70.1 ± 11.3 | 164.2 ± 9.2 | - | - | - | - | - | - | 29.9 ± 0.4 | 3.4 ± 1.4 | 7.7 ± 1.4 | - |
between: 1) patients before BoNT-A and HS; 2) patients before and after BoNT-A; and 3) patients after BoNT-A and HS. Cut-off values were calculated as the point of the curves with the highest Youden Index (sensitivity + specificity − 1) to maximize the sensitivity and specificity of the diagnostic tests. The positive and negative predictive values were also calculated. According to standardized procedures [25], we compared the area under the curves (AUCs) in the ROC curves calculated from the CPPS values alone and from SMO (CPPS + selected features) to verify the optimal test for discriminating within the three subgroups. All ROC analyses were performed using WEKA software.

The Spearman rank correlation coefficient was used to assess correlations between all patients’ clinical features and output measures of cepstral analysis (CPPS values) and machine learning (LR values). Values in patients than HS when comparing CPPS and that calculated with SMO ([CPPS + selected features] during emission of the vowel and the sentence were not statistically significant (difference between AUCs = 0.04, z = -0.96, p = 0.34), suggesting comparable performances in the two speech tasks (Fig. 2A).

3.1.1. Cepstral analysis

Cepstral analysis allowed us to accurately distinguish voice samples collected in HS from those recorded in the 60 patients before BoNT-A. When comparing CPPS during emission of the vowel in HS and patients, unpaired t-test showed lower values in patients than HS (t = 9.14; p < 0.01) (Fig. 1A). Unpaired t-test also showed lower values in patients than HS when comparing CPPS during emission of the sentence in the two groups (t = 10.99; p < 0.01) (Fig. 1B). Results of ROC analyses referring to CPPS during emission of the vowel and sentence are shown in Supp.3; Fig. 2A. Differences between the two ROC curves obtained from the CPPS during emission of the vowel and the sentence were not statistically significant (difference between AUCs = 0.04, z = -0.96, p = 0.34), suggesting comparable performances in the two speech tasks (Fig. 2A).

3.1.2. Machine-learning

When discriminating between HS and patients, the artificial classifier based on SMO using CPPS + selected features allowed us to significantly improve the diagnostic performance of our test compared to cepstral analysis based on CPPS alone.

Results of ROC analyses referring to the comparison of CPPS + selected features during emission of the vowel and the sentence in HS and patients are shown in Supp.3; Fig. 2B. The two AUCs showed no significant differences (z-statistic = -0.14, p = 0.89), suggesting comparable performances in the two speech tasks (Fig. 2B).

In comparing AUCs of the ROC curves obtained from CPPS alone and that calculated with SMO using CPPS + selected features during emission of the vowel, we found that the two AUCs showed significant differences, with the AUC from SMO significantly greater than that obtained using CPPS alone (z-statistic = 2.0; p < 0.05) (Fig. 3C). When applying the same procedure to the sentence, the AUC from SMO was again greater than that obtained using CPPS alone (z-statistic = -2.95; p < 0.05) (Fig. 3D; Supp.3).

3.2. ASD before and after BoNT-A

Our analysis demonstrated a significant clinical effect of BoNT-A on ASD symptoms. Wilcoxon signed rank test demonstrated that after BoNT-A treatment, patients had a significant reduction in VHI (before BoNT-A: 59.7 ± 23.6; after BoNT-A: 41 ± 21.8; z = -4.5, W = 40.5, p < 0.05) and DCS scores (before BoNT-A: 2.2 ± 0.6; after BoNT-A: 1.4 ± 0.8; z = -4.7, W = 18.5, p < 0.05), suggesting a symptomatic improvement in ASD (Table 1; Supp.1–2).

3.2.1. Cepstral analysis

Cepstral analysis accurately distinguished between voice samples collected in the same subgroup of 35 patients before and after BoNT-A.

When comparing CPPS during emission of the vowel before and after BoNT-A, paired t-test showed significantly higher values for patients after than before BoNT-A (t = 5.39; p < 0.01) (Fig. 1C). Paired t-test showed significantly higher values for patients after than before BoNT-A also when comparing CPPS during emission of the sentence (t = 5.17; p < 0.01) (Fig. 1D).

Results of ROC analyses referring to CPPS during emission of the vowel and the sentence are shown in Supp.3; Fig. 2C. Differences between the two ROC curves obtained from CPPS during emission of the vowel and sentence were not statistically significant (difference between AUCs = 0.03, z = -0.96, p = 0.34), suggesting comparable performances in the two speech tasks (Fig. 2C).

3.2.2. Machine-learning

When discriminating between patients before and after BoNT-A, the SMO using CPPS + selected features resulted in a significant improvement in the diagnostic performance of our test compared to cepstral analysis based on CPPS alone.

Results of ROC analyses referring to the comparison of CPPS + selected features extracted from the sustained emission of the vowel and the sentence in patients before and after BoNT-A are shown in Supp.3; Fig. 2D. The two AUCs showed no significant differences (z-statistic = -0.47, p = 0.64) (Fig. 2D).

When comparing AUCs of the ROC curves obtained from CPPS alone and that calculated with SMO using CPPS + selected features during emission of the vowel, the two AUCs showed significant differences, with the AUC from SMO significantly greater than that obtained using CPPS alone (z-statistic = 2.0; p < 0.05) (Fig. 3C). When applying the same procedure to the sentence, the AUC from SMO was again greater than that obtained using CPPS alone (z-statistic = -2.95; p < 0.05) (Fig. 3D; Supp.3).

3.3. ASD after BoNT-A and HS

3.3.1. Cepstral analysis

We accurately distinguished voice samples collected in HS from those recorded in the 35 patients after BoNT-A.

When comparing CPPS values during emission of the vowel in HS and in patients after BoNT-A, unpaired t-test showed lower values in patients than HS (t = 5.71; p < 0.01) (Fig. 1E). Unpaired t-test showed lower values in patients than HS also when comparing CPPS during emission of the sentence (t = 6.46; p < 0.05) (Fig. 1F).

Results of ROC analyses referring to CPPS during emission of the vowel and the sentence are shown in Supp.3; Fig. 2E. Differences between the two ROC curves obtained from CPPS during emission of the vowel and the sentence were not significant (difference between AUCs = 0.02, z = -0.26, p = 0.79), suggesting comparable performances in the two speech tasks (Fig. 2E).

3.3.2. Machine-learning

When discriminating between patients after BoNT-A and HS, SMO with CPPS + selected features resulted in a significant improvement in the diagnostic performance of our test compared to cepstral analysis based on CPPS alone.

Results of ROC analyses referring to the comparison of CPPS + selected features during emission of the vowel and the sentence are shown in Supp.3; Fig. 2F. The two AUCs showed no significant difference (z-statistic = -0.09, p = 0.93), suggesting comparable performances in the two speech tasks (Fig. 2F).

In comparing AUCs of the ROC curves obtained from CPPS alone
and that calculated with SMO using CPPS + selected features during emission of the vowel, we found that the two AUCs showed significant differences, with the AUC from SMO significantly greater than that obtained using CPPS alone (z-statistic = -2.73; p < 0.01) (Fig. 3E). When applying the same procedure to the sentence, the AUC from SMO was again significantly greater than that obtained using CPPS alone (z-statistic = 2.9; p < 0.05) (Fig. 3F; Supp.3).

### 3.3.3. Correlation analysis

Spearman’s test found that DCS scores negatively correlated with CPPS values calculated during the emission of a vowel but not of a sentence, in patients before (r = -0.5, p < 0.01) and after BoNT-A (r = -0.41, p = 0.02). The correlation with DCS scores was also present when considering LR values instead of CPPS, calculated during the emission of a vowel but not of a sentence, before (r = -0.52, p < 0.01) and after BoNT-A (r = 0.48, p < 0.01). CPPS and LR values correlated in patients before (r = 0.6, p < 0.01) and after BoNT-A (r = -0.6, 

---

**Fig. 1.** CPPS in patients before BoNT-A and HS (panels A, B), in patients after BoNT-A and HS (C, D), and in patients before and after BoNT-A (E, F) during the emission of the vowel (left column) and the sentence (right column).
Fig. 2. Comparison of ROC curves calculated with CPPS alone (left column) and CPPS + selected features (right column), each used to differentiate patients before BoNT-A and HS (panels A, B), patients before and after BoNT-A (C, D), and patients after BoNT-A and HS (E, F) during the emission of the vowel (grey line) and the sentence (black line).
Fig. 3. Comparison of ROC curves calculated with CPPS values alone (black line) and CPPS + selected features (grey line), each used to differentiate patients before BoNT-A and HS (panels A, B), patients before and after BoNT-A (C, D), and patients after BoNT-A and HS (E, F) during the emission of the vowel (left column) and the sentence (right column).
p < 0.01). Lastly, in ASD there was no correlation between CPPS/LR values and the VHI scores or the remaining patients’ clinical features.

4. Discussion

In our study, cepstral analysis showed reduced CPPS in ASD compared with HS. CPPS increased after BoNT-A, without fully normalizing. Voice analysis with advanced machine-learning algorithms better distinguished between patients and HS compared to conventional cepstral analysis. Furthermore, we found that machine-learning algorithms distinguished between patients before and after BoNT-A with higher accuracy than cepstral analysis, thus better quantifying the symptomatic effect of BoNT-A. Finally, cepstral and machine-learning analysis performed during the sustained emission of the vowel and of the sentence gave similar results in HS and patients. However, we found a significant correlation between the severity of voice impairment (i.e. DCS scores) and objective voice parameters (i.e. CPPS or LR values) only during the emission of the vowel and not of the sentence.

All participants were native Italian speakers, thus excluding confounding due to linguistic differences. Moreover, they were all non-smokers and had comparable anthropometric features. ASD was diagnosed according to standardized procedures [2,3]. Concerning the specific timing of voice recordings, we examined patients before and after BoNT-A according to botulinum pharmacokinetics [26]. Since some patients with ASD complained of voice worsening over the day, we have collected voice samples in all participants early in the morning in order to reduce variability in voice recordings [1–6,26].

4.1. Voice analysis in ASD

Cepstral analysis demonstrated lower CPPS in patients than in HS, confirming previous observations [10–16,24,27–29]. Also, in line with previous reports [10–16,24,27–29], we found a negative correlation between CPPS and DCS scores (i.e. the lower CPPS, the greater voice impairment), supporting the hypothesis that CPPS scores are helpful to quantify overall disability in ASD. In a previous study using a large sample of patients with heterogeneous voice disorders, Henam-Ackah et al. demonstrated that abnormally low CPPS helped to objectively identify patients with any kind of voice impairment [11,12]. Using a cut-off of 4 for CPPS obtained during the emission of a standardized sentence, Henam-Ackah et al. demonstrated a ROC analysis with high sensitivity and specificity [11,12]. Our ROC analyses obtained during the emission of a sentence identified a similar diagnostic cut-off value of 4.58 for CPPS and comparable values of sensitivity and specificity.

Our study is the first to demonstrate that the machine-learning algorithm is able to better discriminate between HS and ASD patients compared to conventional cepstral analysis using CPPS alone, as demonstrated by previously unreported high values of sensitivity (98.3%) and specificity (98.3%). Moreover, LR values also correlated with DCS scores. Several methodological and pathophysiological factors may explain the superior performance of the machine-learning algorithm compared to conventional cepstral analysis [30]. The human voice represents a complex phenomenon characterized by high-dimensional data based on an exponential number of features [24,27–30]. Furthermore, in ASD patients, involuntary abnormal vocal cord contractions provide further complexity. The machine-learning algorithm dynamically combines selected features extracted from a large dataset, and offers the opportunity to automatically learn and improve from experience, without being explicitly programmed [17–19,21–23]. Accordingly, the machine-learning algorithm is superior to recently reported efficient methods based on multivariate analysis including those used to calculate the acoustic voice quality index and the cepstral spectral index of dysphonia [10,13–15,24,27,28]. Machine-learning algorithm thus offers new advanced tools for the analysis of high-dimensional and multimodal biomedical data including the human voice. Image-based machine-learning analysis of a heterogeneous sample of voice spectrograms using a deep neural network has recently been proposed as an indirect tool to discriminate different voice disorders [30]. However, our machine learning algorithm directly analyses voice recordings and appears better able to identify features of voice analysis.

4.2. Voice analysis and the objective evaluation of BoNT-A therapy

Cepstral analysis demonstrated higher CPPS in patients after than before BoNT-A, which is consistent with our findings in a previous study [16]. ROC curve analysis demonstrated moderate sensitivity and high specificity, suggesting accurate detection of BoNT-A effects in ASD. The well-known pharmacological action of BoNT-A consists of chemodenervation of injected muscles, leading to a reduction in muscular activation [1–7]. In ASD, BoNT-A reduces vocal cord tension, thus causing decreased voice aperiodicity and increased speech quality. Accordingly, the improvement in CPPS in patients after BoNT-A reflects the ability of this parameter to detect irregular vibration of the vocal cords and voice periodicity [7,27].

When evaluating the effect of BoNT-A in ASD, the machine-learning algorithm again provided higher sensitivity and specificity values than conventional cepstral analysis using CPPS alone and other methods based on multivariate analysis [27]. This suggests the superiority of the machine-learning algorithm compared to cepstral analysis when analyzing complex high-dimensional data. We speculate that by reducing vocal cord tension, BoNT-A changes specific voice features that are more easily detected by the machine-learning algorithm.

Although BoNT-A clinically improved ASD symptoms, it did not restore voices within normal ranges. Accordingly, our conventional cepstral analysis demonstrated lower CPPS in patients after BoNT-A than in HS, reflecting a less harmonic and periodic voice in the first group. The observation that CPPS and LR values both correlated with DCS scores in patients after BoNT-A therapy, supports the clinical utility of these measures to quantify voice impairment in ASD. However, although BoNT-A injection increased CPPS in ASD, it was not able to restore them within normal ranges. Results were similar when using the machine-learning algorithm, which discriminated between patients after BoNT-A and HS with an even greater performance than cepstral analysis. Furthermore, when evaluating the effect of BoNT-A, the machine-learning algorithm operated with a higher sensitivity and specificity compared to conventional cepstral analysis.

We achieved comparable results using voice samples collected during the emission of the vowel and the sentence in HS and ASD before and after BoNT-A [29]. However, we found significant correlations between the severity of voice impairment, as tested by the DCS scores, and CPPS or LR values, only when considering the emission of the vowel. This finding would probably reflect the ability of a trained human evaluator to recognize subtle impairment of voice parameters more precisely during the emission of a vowel than a sentence. Accordingly, the DCS score would be invariably more influenced by changes in the vowel than in the sentence. Hence, we suggest that from a clinical point of view, it would be more useful to implement a speech task based on the vowel, being more reliable among different languages.

The observation that, differently from the DCS, the VHI scores did not correlate with CPPS or LR values in patients with ASD, during the emission of the vowel and the sentence, before and after BoNT-A. This would likely reflect the specific features of the VHI questionnaire that, differently from the DCS scale, consists of patients’ self-administered scale based on qualitative/perceptive evaluation of the voice.

Given that ASD is a rare disease and that machine-learning analysis requires large amounts of data, we recognize that the low number of ASD patients represents the main limitation of our study. Future studies are needed to improve the application of machine-learning analysis in voice disorders. Since in the present study we did not record voice samples serially, we cannot exclude variability in voice recordings due to daily fluctuations in voice parameters.
5. Conclusion

Cepstral analysis discriminates between ASD and HS and assess the effect of BoNT-A. Furthermore, machine-learning algorithms are superior to conventional cepstral analysis in supporting clinicians in the diagnosis and therapeutic follow-up of ASD. We suggest that an objective clinical diagnosis of ASD should involve a two-step algorithm based on the correct classification of voices through machine-learning, followed by calculation of CPPS or LR values to quantify the severity of voice impairment. Similarly, the same two-step algorithm could assist clinicians in monitoring the effects of BoNT-A in ASD.

Authors’ roles

1) Research project: A. Conception (AS, GR, GS); B. Organization (FA, GC, LM, GR, GS); C. Execution (FA, GC, LM, DC, ZZ); 2) Statistical Analysis: A. Design (AS, FA, GC); B. Execution (FA, GC, DC); C. Review and Critique (AS, GS, AB); 3) Manuscript: A. Writing of the first draft (AS, FA, LM); B. Review and Critique (GS, GC, AB).

Funding sources to declare

None.

Data availability

The anonymized database used and analyzed during the current study is available from the corresponding author on reasonable request.

Declaration of competing interest

The authors report no disclosures.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.parkreldis.2020.03.012.

References