Spasmodic dysphonia, perceptual and acoustic analysis: presenting new diagnostic tools

Denise Irene Siemons-Lühring · Mieke Moerman · Jean-Pierre Martens · Dirk Deuster · Frank Müller · Philippe Dejonckere

Received: 10 January 2009 / Accepted: 5 May 2009 / Published online: 3 October 2009 © The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract In this article, we investigate whether (1) the IINFVo (Impression, Intelligibility, Noise, Fluency and Voicing) perceptual rating scale and (2) the AMPEX (Auditory Model Based Pitch Extractor) acoustical analysis are suitable for evaluating adductor spasmodic dysphonia (AdSD). Voice recordings of 12 patients were analysed. The inter-rater and intra-rater consistency showed highly significant correlations for the IINFVo rating scale, with the exception of the parameter Noise. AMPEX reliably analyses vowels (correlation between PUVF (percentage of frames with unreliable F0/voicing 0.748), running speech (correlation between PVF (percentage of voiced frames)/voicing 0.699) and syllables. Correlations between IINFVo and AMPEX range from 0.608 to 0.818, except for noise. This study indicates that IINFVo and AMPEX could be robust and complementary assessment tools for the evaluation of AdSD. Both the tools provide us with the valuable information about voice quality, stability of F0 (fundamental frequency) and specific dimensions controlling the transitions between voiced and unvoiced segments.

Keywords Voice · Spasmodic dysphonia · Perception · Acoustic analysis · IINFVo · AMPEX

Introduction

Spasmodic dysphonia (SD) is a severe voice disorder characterised by involuntary disruption of phonation [1]. The aetiology of SD was originally thought to be psycho-neurotic, but it has become clear that the cause is neurological. Nowadays, injecting botulinum toxin A is the preferred therapy [1].

There are three general types of SD: adductor, abductor and mixed adductor/abductor. The most common of these is adductor spasmodic dysphonia (AdSD): it is characterised by strained, strangled and effortful phonation with words frequently cut off or difficult to start. Irregularly intermittent increases in adductor muscle activity coincide with phonatory offsets [1].

We define AdSD as “a focal laryngeal dystonia resulting in a strained voice quality with spastic breaks” which according to Rees et al. [2] improves with botulinum toxin A injections.

A reliable methodology is essential for evidence-based (voice) assessment and/or therapy. The European
Dejonckere et al. [5] demonstrated that the parameters across all parameters except for the parameter "strain". (VPA) and GRBAS) only the GRBAS scale was reliable. Buffalo Voice Profile, The Vocal Profile Analysis Scheme [4] found that by comparing three perceptual scales (The developed, but only few rated for reliability. Webb et al. [5] found that by comparing three perceptual scales (The Buffalo Voice Profile, The Vocal Profile Analysis Scheme (VPA) and GRBAS) only the GRBAS scale was reliable across all parameters except for the parameter "strain". Dejonckere et al. [5] demonstrated that the parameters asthenicity, strain and instability provide low intrarater agreement values: 0.69 for breathiness and 0.65 for asthenicity/strain and concluded that only the G, R and B parameters have clinical importance. This confirms the German perceptual rating with the RBH scale (R Rauigkeit, B Behauchtheit, H Heiserkeit) [6].

It seems that in substitution voicing (SV), the GRBAS is not applicable because of the extreme severity of the voice pathology (often SV scores as a G3) and the inability of the GRBAS scale to describe variations in quality or minor differences. GRBAS is also inadequate in scoring fluency problems, which are one of the main characteristics of spasmodic dysphonia [3].

Moerman et al. [3] developed a rating scale, called II-NFVo, which is appropriate for extremely deviant voices such as substitution voices. It consists of five parameters “Overall Impression”, Impression of Intelligibility”, “Noise”, “Fluency” and “Voicing”, which are all scored on a visual analogue scale (VAS) from 0 to 10. The II-NFVo was found to be reliable when scored by professionals [7].

Regarding the acoustic analysis, the computer program traditionally used for this purpose, MDVP (Kay Elemetrics), is geared to the peaks of the signal in determining the F0 and only works reliably if the acoustic signal (1) contains little or no noise and (2) shows a certain amount of periodicity [8]. Although the evaluation of running speech better matches reality, MDVP can only thoroughly analyse sustained vowels. SD cannot reliably be analysed by MDVP because of its difficulties in determining the fundamental frequency (F0), the strong aperiodicity of the signal and the high amount of voice breaks. Voice onset problems and voiced/unvoiced variations are the bigger challenge in acoustic analysis.

Moerman et al. [3] used the Auditory Model based Pitch Extractor (AMPEX), developed by Van Immerseel and Martens [9] to analyse substitution voices. They proved it is robust in differentiating (1) normal speech, (2) one vocal cord speech, (3) tracheo-oesophageal speech and oesophageal speech. AMPEX was used to analyse various speech types (Table 1: “Sustained vowels”, “Syllables p2, b2, p3 and b3, “text (running speech)” and “count (digit strings)”). For its analysis, the auditory model produces a 27-dimensional feature vector every 10 ms. Each vector consists of 23 spectral parameters, an energy value, a V/U (voiced/unvoiced) value, F0 value and F0 evidence. AMPEX is also able to extract the F0 against background noise (Table 2).

Our investigation focuses on whether AMPEX could be a robust assessment tool in analysing AdSD.

Based on the basic protocol by Dejonckere [approved by the guidelines elaborated by the committee on phoniatrics of the European Laryngological Society (ELS)] and the findings of Moerman we evaluate (1) the acoustic analysis of sustained vowels and running speech with AMPEX and

---

### Table 1 Content of the voice recordings of AdSD patients consisting of vowels, syllables (VCV and CVCVCV) and running speech (text)

| Speech type | Description |
|-------------|-------------|
| Vowels | Vowels /a/, /i/, /u/ |
| p2 | The VCV utterances /apa/, /ipi/, /upu/ |
| h2 | The VCV utterances /aba/, /ibi/, /uba/ |
| p3 | The CVCVCV utterances /papa/, /pipi/, /pupu/ |
| h3 | The CVCVCV utterances /babab/, /bibibi/, /babubu/ |
| Text | German phonetic rich text: Einst stritten sich Nordwind und Sonne, wer von ihnen beiden wohl der Stärkere wäre, als ein Wanderer, der in einen warmen Mantel gehüllt war, des Weges kam. Sie wurden einig, dass derjenige für den Stärkeren gelten sollte, der den Wanderer zwingen würde, seinen Mantel abzulegen |

### Table 2 Parameters of the acoustic analysis by AMPEX for the voice quality of AdSD patients

| Parameters | Description |
|------------|-------------|
| PVF | Percentage of voiced frames |
| PVS | Percentage of voiced speech frames |
| AVE | Mean voicing evidence of voiced frames |
| Jit | Evidence weighted F0-variation in voiced frames |
| Jc | Evidence weighted F0-variation in reliable voiced frames |
| PUVF | Percentage of unreliable voiced frames |
| VL90 | 90th percentile of the voicing length distribution |
The perceptual evaluation with IINFVo in AdSD [5, 10]. In addition, we asked patients to self assess them by responding to the question: rate the quality of your voice on a scale from 0% (worst voice quality ever experienced) to 100% (best voice quality ever experienced).

A visual evaluation of the glottis function and the performance of aerodynamic measures, although of clinical interest were omitted because our focus was on alternative measurement methods specifically aimed at acoustical analysis and perceptual evaluation. The aim of this study is to determine whether the alternative tools (IINFVo and AMPEX) are suitable for describing voice quality in AdSD.

Materials and methods

Research design

In this pilot-study, the reliability and validity of a multidimensional research instrument for patients are examined. It takes a cross-sectional approach, which focuses on diagnosis at different stages and severities of spasmodic dysphonia.

Sampling

Patients

Over a 5-month period patients, with a clinical diagnosis of AdSD of varying severities, were recruited at the University Hospital of Hamburg (Germany), where they were routinely treated with botulinum toxin A injections. The voice recordings were made immediately before the injection.

Twelve patients were included in the study: four males and eight females. The average age was 62.25 years for the male group and 63.75 years for the female group. The average age of the whole group was 63.25 years.

All patients were asked if they would agree to take part in the study and they all gave their written informed consent.

Perceptual evaluation

IINFVo

The IINFVo rating scale consists of the following parameters [7]:

1. Impression: overall voice quality.
2. Intelligibility: the impression of intelligibility.
3. Noise: unintended additive noise.
4. Fluency: the perceived smoothness of the sound production.
5. Quality of voicing: is voicing voiced or unvoiced when it is supposed to be voiced or unvoiced.

The perceptual evaluation was carried out by three professional judges: an ear–nose–throat/phoniatrician medical doctor (Belgium), a phoniatrician medical doctor (Germany) and a speech pathologist (Germany), all with longstanding experience in speech and voice pathology. The study was conducted blindly. Neither the patients, nor the severity of the voice pathology were known by the judges. Only the diagnosis of AdSD was known.

The inter-individual agreement was first calculated per jury member pair (J1–J2, J1–J3 and J2–J3) and then converted into a mean Pearson’s correlation coefficient for each parameter.

After 3 months, the intra-rater agreement was calculated for one of the judges and converted into a mean Pearson’s correlation coefficient for each parameter.

Following the same procedure, we also computed Kendall’s tau as an alternative indicator of interrater agreement, an indicator that does not require a normal distribution of the scores.

Objective acoustic analysis

Recordings

Table 1 shows the utterances to be made: vowels, syllables (VCV and CVCVVCV) and running speech (text). The utterances were separated by long pauses of at least 1 s. Each utterance was stored in a separate WAV file.

The recordings were made in a standard (quiet) clinical room, 4 × 4 m. The System SpeechStudio from Laryngograph with adapted headset-microphone (Labtec) was used. The samples were recorded digitally with a sampling frequency of 16 kHz and stored in linear 16 bit Pulse-Code-Modulation (PCM) in a windows WAV format. Subsequently, they were converted with Cool-Edit to 22,050 Hz. The recording level was adjusted as to avoid peak clipping.

Acoustic analysis

AMPEX Each recording session was first manually (with the help of PRAAT) divided into its constituent parts “vowels”, “syllables” (p2, b2, p3, b3), “text” and “count” (Table 1).

The timing information is stored in short text grid files; AMPEX was able to analyse all voice recordings: it used the timing information to select the part it was supposed to analyse.
The acoustic analysis produces the following parameters [3].

1. Percentage of voiced frames (PVF) and percentage of voiced speech frames (PVS)
2. Average voicing evidence (AVE)
3. Jitter (Jit), jitter corrected (Jc) and percentage of frames with unreliable F0 (PUVF)
4. Voicing length (VL 90).

Because the perceptual parameters of the IINFVo are ordinal, we determined the Spearman’s ranking correlation coefficient for comparing the perceptual evaluation with the acoustic analysis.

Statistical analysis

The statistical analysis was done using SPSS 15.0. The Kolmogorov–Smirnov statistic and a Q–Q Plot were conducted in order to check that the data were distributed normally. A non-significant result in the Kolmogorov–Smirnov statistic (Sig value > 0.05) indicates normality.

A multiple regression analysis for the parameters of AMPEX and the IINFVo could not be performed because of the small number of patients in the study (N = 12). (The sample size should be at least 15 subjects per predictor, Stevens (1996), p.72.)

Perceptual evaluation

Normal distributions were found for the means of the professional judges in the elements Impression (Sig.. value = 0.200), Noise (Sig. value = 0.130), Fluency (0.077) and Voicing (0.200). Only for the parameter “Intelligibility” was the Sig. value significant (0.031), indicating that this parameter showed an abnormal distribution.

Results

Consistency measures

Perceptual evaluation

Inter-rater consistency The inter-rater consistency was measured for the professional judges by determining the Pearson’s correlation coefficient, and Kendall’s tau as an alternative, for each member and the means of the others per parameter of the IINFVo. Table 3 shows the results (Pearson’s Correlation and Kendall’s tau) for each jury member when he is compared with others. It also shows the mean, the minimum and the maximum values of the results.

| Parameter | Impression | Intelligibility | Noise | Fluency | Voicing |
|-----------|------------|----------------|-------|---------|---------|
|           | PCC        | K-tau          | PCC   | K-tau   | PCC     | K-tau   | PCC   | K-tau   | PCC     | K-tau   |
| P1–23     | 0.976**    | 0.967**        | 0.924** | 0.772** | 0.015   | 0.052   | 0.928** | 0.775** | 0.930** | 0.833** |
| P2–13     | 0.864**    | 0.718**        | 0.978** | 0.869** | 0.174   | 0.297   | 0.850** | 0.688** | 0.835** | 0.645** |
| P3–12     | 0.847**    | 0.784**        | 0.947** | 0.785** | 0.343   | 0.349   | 0.703*  | 0.739** | 0.833** | 0.700** |
| Means     | 0.896      | 0.823          | 0.950  | 0.809   | 0.177   | 0.233   | 0.827   | 0.734   | 0.866   | 0.726   |
| Minimum   | 0.847      | 0.718          | 0.947  | 0.772   | 0.015   | 0.052   | 0.703   | 0.688   | 0.833   | 0.645   |
| Maximum   | 0.976      | 0.967          | 0.978  | 0.869   | 0.343   | 0.349   | 0.928   | 0.775   | 0.930   | 0.833   |

The comparisons are made for all the parameters of the IINFVo

P Professional judge, PCC Pearson’s correlation coefficient, K-tau Kendall’s tau

*Significant level 0.05 (two-tailed)

**Significant level 0.01 (two-tailed)
Relation between perceptual and acoustic parameters

The mean values for the three professional judges were computed and used for the Spearman’s ranking order correlation analysis with the parameters of AMPEX.

Table 8 gives the correlations found for the parameters VL90, PUVF, PVF, Jc and jitter. Significant correlations were found for the following combinations:

| VL90 | Syllables b3 | Intelligibility |
|------|--------------|----------------|
| PUVF | Vowels       | Impression, Intelligibility, Voicing |
| PVF  | Vowels Text  | Impression, Intelligibility, Voicing, Impression |
| Jc   | Vowels       | Impression, Intelligibility, Voicing |
| Jitter | Vowels     | Impression, Intelligibility, Voicing |
| PVS  | Vowels       | Impression, Intelligibility, Fluency |

For Syllables p3 (AMPEX), and for Noise (IINFVo) no significant correlations were found.

From Moerman et al. [3], the values of the AMPEX parameters for four different voice types [normal, tracheoesophageal (TE), oesophageal (E) and a voice with one vocal cord (CVC)] are known. They can be compared with the corresponding values for “Vowels” and “Text” that we obtained for our spasmodic dysphonia voices in Tables 9 (Vowels) and 10 (Text).

Table 11 shows the values for normal speech and AdSD for “Syllables p3, b3” and for the parameters PVF, PUVF and VL90.

Discussion

The aim of this study was to determine if two alternative tools (IINFVo and AMPEX) are suitable for analysing the severe voice pathology of spasmodic dysphonia. The relative infrequency cases of SD should be taken into consideration when interpreting the statistics.

IINFVo

The inter-rater consistency for professionals shows high correlations (the mean scores of the three judges range from 0.827 to 0.950), except for “Noise” (0.177). Also Kendall’s tau shows significant values: ranging from 0.726 to 0.823, except for noise (0.233).

The intra-rater consistency, measured for one professional judge 3 months later, also showed significant values for all parameters in the IINFVo scale, ranging from 0.863 to 0.943, except for the parameter “Noise” (0.224).
Table 5 shows that the mean of the MAD values of the professional judges for the parameters “Impression”, “Intelligibility”, “Fluency” and “Voicing” are between 0.56 and 1.42 on a 0–10 scale. As such it suggests that these parameters are valuable in evaluating SD voicing for professionals.

In further investigation, the parameter “Noise” may probably be redefined for patients with AdSD and the training of the rater should focus on this specific problem.

When comparing the data of AdSD with the data of normal voices and of patients with substitution voices (tracheoesophageal and oesophageal voices, voices with one vocal cord) as shown in Tables 9 and 10, one parameter was seen to be very deviant: AVE, the average voicing evidence. The AVE for AdSD was only 0.8 for “Vowels” and 0.7 for “Text”. The corresponding values for normal voices were 6.8 (Vowels) and 5.1 (Text). Even the severe voice pathologies (tracheoesophageal and oesophageal voices, voices with one vocal cord) showed significantly higher values than AdSD voices: 2.8–5.1 for “Vowels” and 2.5–3.8 for “Text”. This could indicate the specific difficulty experienced by AdSD patients in producing voiced/unvoiced sounds where they are required.

Comparing the percentage of voiced frames (PVF) for the syllables p3 and b3 the people with normal speech only show a mean value of 3.0, whereas the AdSD patients show nearly the same mean value as for syllables p3 (12.9). This shows the instability of the F0 for syllables loaded with voiced/consonants.

The mean VL90 of AdSD patients in syllables p3 is higher (55.2) than that of normal speakers (21.0) whereas in syllables b3 the two are very comparable (AdSD 60.5, normal speakers 57). This would suggest that voices with AdSD may not be able to accurately control the changes between voiced and unvoiced segments.

Comparing the percentage of voiced frames (PVF) for the syllables loaded with unvoiced consonants (p3) of normal speakers and AdSD patients it follows that the mean value for normal speakers (40.2) is about twice as high as the mean value for AdSD patients (21.0), whereas the mean values for syllables loaded with voiced consonants (b3) are nearly similar (65.2 and 65.6). This would indicate that the syllables with voiceless consonants present more problems for the voicing part.
This would confirm the suggestions of Roy [12] et al. for certain task specificity. Roy poses the question, based on the clinical observations, whether sentences loaded with voiced consonants are more difficult, e.g. that sentences loaded with voiced consonants would provoke more frequent and severe spasms/voice breaks [12]. Our findings would suggest that the alternation between voiced and unvoiced segments in syllables with unvoiced consonants is the greater challenge in AdSD. Hence, task specificity is assumable, but is more complex and needs further investigation.

IINFVo and AMPEX

The results in Table 8 show significant correlations, ranging from 0.608 to 0.818, between all the parameters of the IINFVo and AMPEX except for “Noise”. It emphasises that both tools are able to measure specific voice dimensions in spasmodic voicing. Both the assessment methods are complementary and seem to provide us with the valuable information without being mutually redundant.

Conclusion

This study suggests that multidimensional voice assessment consisting of objective acoustic analysis using AMPEX and perceptual evaluation using IINFVo could be a robust tool for assessing severe voice pathology such as spasmodic dysphonia. In follow-up studies, these tools may be used for determining markers for segregating and diagnosing AdSD.

Clinical observations by Roy et al. suggest that patients with AdSD would have more difficulties in speaking sentences loaded with voiced consonants [12]. Our findings indicate that voices with AdSD may not be able to control the changes between voiced and unvoiced segments. Our findings confirm the possible presence of task specificity, but also illustrate complexity and the need for further investigation.

Acknowledgments We thank the University Hospital Hamburg, Germany, for recording patients with spasmodic dysphonia and the Department of Phoniatrics and Paediatric Audiology, University Hospital Münster, Germany, for scoring the samples with IINFVo.

Conflict of interest statement The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix I

Spectrogram of vowel/a:/showing high irregularity and voice breaks (PRAAT analysis)

References

1. Zwirner P, Murry T, Woodson GE (1993) A comparison of bilateral and unilateral botulinum toxin treatments for spasmodic dysphonia. Eur Arch Otorhinolaryngol 250:271–276
2. Rees CJ, Blalock PD, ShE Kemp, StL Halum, Koufman JA (2007) Differentiation of adductor-type spasmodic dysphonia from muscle tension dysphonia by spectral analysis. Otolaryngol Head Neck Surg 137:576–581
3. Moerman MBJ, Martens JP, Van der Borgt MJ, Peleman M, Gillis M, Dejonckere PH (2006) Perceptual evaluation of substitution voices: development and evaluation of the (I)INFVo rating scale. Eur Arch Otorhinolaryngol 263:183–187
4. Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen N, Wilson JA (2004) The reliability of three perceptual evaluation scales for dysphonia. Eur Arch Otorhinolaryngol 261:429–434
5. Dejonckere PH, Remacle M, Fresnel-Ebaz E, Woisard V, Crevier-Buchman L, Millet B (1996) Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. Rev Laryngol Otol Rhinol (Bord) 117(3):219–224
6. Ptok M, Schwemmle C, Iven C, Jessen M, Nawka T (2005) Zur auditiven Bewertung der Stimmeigenschaften HNO 54:793–802
7. Moerman M, Martens JP, Crevier-Buchman L, de Haan E, St Grand, Chr Tessier, Woisard V, Dejonckere PH (2006) The IINFVo perceptual rating scale for substitution voicing: development and reliability. Eur Arch Otorhinolaryngol 263:435–439
8. Wempe T, Boersma P (2004) The interactive design of an F0-related spectral analyser. www.praat.org
9. Van Immerseel LM, Martens JP (1992) Pitch and voice/unvoiced determination within an auditory model. J Acoust Soc Am June:3511–3526
10. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchmann L, Friedrich G, Van De Heyning P (2001) A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques (Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS)). Eur Arch Otorhinolaryngol 258:77–82

11. Stevens J (1996) Applied multivariate statistics for the social sciences, 3rd edn. Lawrence Erlbaum, Mahway

12. Roy N, Mauszycke SC, Merrill RM, Gouse M, Smith ME (2007) Toward improved differential diagnosis of adductor spasmodic dysphonia and muscle tension dysphonia. Folia Phoniatrica et Logopaedica 59:83–90