TOWARDS A BASIC PROTOCOL FOR FUNCTIONAL ASSESSMENT OF SUBSTITUTION VOICES: PRELIMINARY RESULTS OF AN INTERNATIONAL TRIAL

M.B.J.Moerman 1,2, J.P.Martens 3, D.Chevalier 4, G.Friedrich 5, M.Hess 6, G.Lawson 7, A.K.Licht 6, F.Ogut 8, E.Reckenzaun 5, M.Remacle 7, V.Woisard 9, P.H.Dejonckere 2

1Department of ENT/head and Neck Surgery, AZ Jan Palfijn-Maria Middelares Gent, Ghent, Belgium; 2Institute of Phoniatrics, University Medical Centre Utrecht, Utrecht, The Netherlands; 3Electronics and Information Systems Department, Ghent University, Ghent, Belgium; 4Service ORL Hospital Huriez CHU, Lille, France; 5Dept of Phoniatrics, Speech and Swallowing, Graz, Austria; 6University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 7Service ORL et chirurgie cervicofaciale, Cliniques Universitaires UCL de Mont Godinne, Yvoir, Belgium; 8Ege Üniv. KBB Anabilindah Bornova Izmir; Turkey; 9ORL-Chirurgie cervico-faciale, CHU-hôpital Larrey, Toulouse, France

Abstract: We performed an international trial with a newly developed multidimensional assessment protocol for substitution voices based on the European Laryngological Society protocol for ‘common dysphonia’. However, as sound production in SV is largely irregular, we needed some adaptations, in particular of the dimensions perception, acoustic analysis and visual evaluation. The protocol consisted of clinical information, the IINFVo perceptual rating scale, visual examination (level and quality of vibration), acoustic registration of vowels, cvc, cvcvv and text (which was later analysed with the Auditory Model Based Pitch Extractor (AMPEX)), aerodynamic measurements (Vital Capacity, Maximum Phonation Time, and Maximal Intensity), and self-evaluation (of i) voice quality and ii) degree of invalidity).

Six centers participated. We retained 96 suitable files (out of 102). Variance analysis demonstrates significance for i) all perceptual parameters (except for Voicing), MPT and Maximal Intensity and ii) type of surgery and/or main anatomical vibration source. There is no correlation at all between the patient’s perception of his/hers disability and perceptual parameters, MPT, quality of vibration. Correlation between the acoustical analysis and the subjective rating was only moderate (Pearson < 0.62; standard deviation >1.8).

Keywords : substitution voices, multidimensional protocol, acoustics, perception, dysphonia

I. INTRODUCTION

In 2001 the European Laryngological Society advocated a protocol for a multidimensional voice assessment for laryngeal dysphonia [1]. This assessment protocol consists of 5 dimensions: perceptual analysis, acoustic measurements, visual evaluation (videostroboscopy), aerodynamic measurements and self-assessment. However, the ELS assessment protocol seems not applicable to substitution voicing (SV).

Substitution voicing is defined as voicing without two true vocal folds [2] and occurs after total laryngectomy (esophageal and tracheo-esophageal speech), partial laryngectomy (except for horizontal supraglottic laryngectomy), cordectomy from type III on (in which a large part of the vocalis muscle has been removed), severe laryngeal trauma etc. Most of these voices are rated as a G3 on the GRBAS scale, whereas there exists a large quality variety within SV [3, 4, 5, 6, 7]. Furthermore, the acoustic signal is largely irregular and can not reliably be analyzed by traditional acoustic programs (e.g. Kay electromics, EVA). Therefore, we tried to design a clinical assessment protocol for this specific type of severe dysphonia, through i) substituting the perceptual evaluation standard (GRBAS) and the acoustic assessment method, and ii) adapting visual evaluation, aerodynamic measurements and self-assessment. This manuscript describes the preliminary results of an international trial which is still going on.

II. METHODS

Perceptual evaluation scale: A new perceptual evaluation scale, called IINFVo, was proposed and studied for its reliability [8]. In this scale 5 parameters are defined: overall Impression (I), impression of Intelligibility (I), unintended additive noise (N), Fluency (F), Voicing (Vo). Reliability of the scores of both professional and semi-professional jury members was studied on speech samples derived from native Dutch
Multicenter trial with the ELS protocol adapted for SV (E).

Acoustic analysis: The wav files of the same speech samples were analysed by the auditory model of Van Immerseel and Martens [9]. This auditory model has a built-in pitch extractor, called AMPEX (Auditory Model–Based Pitch Extractor), which has been proven to outperform most other pitch extractors in circumstances with background noise. The auditory model internally works with signal windows that are considerably larger than 10 ms and facilitate the extraction of evidence for pitch values lower than 100 Hz. Every 10 ms, it produces a 27-dimensional feature vector consisting of 23 spectral parameters, a voiced/unvoiced flag (VU = 0 or 1), a fundamental frequency (Fo) or pitch (zero if unvoiced), a voicing evidence (VE) and a frame energy (E).

Multicenter trial with the ELS protocol adapted for SV: the ELS protocol for laryngeal dysphonia, modified for SV (consisting of the same 5 dimensions, but with i) substitution of the GRBAS and the traditional acoustic analysis, and ii) adaptation of the other three dimensions) was lately tested by several international centres. Six centres participated in the dimensions perceptual evaluation, visual evaluation, aerodynamic measurements and self-assessment rating; four centres also participated in the voice recordings. Until now, statistics on the international data comprised correlation and variance analysis.

III. RESULTS

A. perceptual evaluation scale
Inter judge agreement, as measured on the ratings of the 102 voices recorded in Ghent, is good for semi-professionals and excellent for professionals [8].

B. acoustic analysis
Properly defined acoustic parameters derived from the auditory analysis seem to demonstrate the following (average) ordering of voices according to their over-all quality: (i) normal voicing followed by (ii) voicing with one vocal fold, (iii) TE voicing and (iv) E voicing [2]. However, the demonstrated differences between TE voices and E voices are rather small.

C. multicenter trial with the ELS protocol adapted for SV
We collected 102 files (16 female, 85 male, 1 unidentified) from which 2 were not further specified and 4 did not concur with the definition of SV. The distribution of the 96 remaining samples categorized according to 5 main surgery types was: 11 fronto-lateral laryngectomy/tucker; 11 total laryngectomy with myotomy, 15 total laryngectomy without myotomy and/or with or without pharyngectomy or reconstruction; 22 cricohyoido(epiglott)pexy; 37 cordectomy (from type III on). This population is largely different from the population recorded in Ghent (the people recorded in Ghent were mainly TL).

Perception: An analysis of the correlations between perceptual parameters showed that the highest values are found between ‘General Impression’ and ‘Voicing’ (r=0.83) and between ‘Impression of Intelligibility’ and ‘Fluency’ (r=0.86).

Perceptual parameters and type of surgery/anatomical structure: Variance analysis is significant for all perceptual parameters (IINF) except for Vo and the type of surgery. This significance is mainly due to the lower scores of the TL group (with or without myotomy), except for the parameter Noise where CHEP scores worst. Regarding the perceptual parameters and the ‘main vibrating anatomical structure’, vibration at the esophageal segment scores worst. Fig.1 gives an example of the perceptual parameter ‘impression of intelligibility’ and the main anatomical vibratory source. This is in agreement with the former acoustic analyses based on the AMPEX model.

Aerodynamic measurements:
Variance analysis demonstrates a significance level for MPT and i) Type of Surgery (p=0.03) and ii) main anatomical structure producing vibration (p=0.0003). The level of significance for Maximal Intensity is 0.0004 for Type of Surgery and 0.0034 for Main anatomical structure producing vibration. Further analysis demonstrates a significant difference between 1) cordectomy and i) TL (p=0.046) and ii) CHEP (p=0.0002), 2) TL with myotomy and CHEP (p=0.009), 3) CHEP and Tucker/FL (p=0.008).

Self-assessment: There is no correlation at all between the patient’s perception of his /her disability and perceptual parameters, MPT, quality of vibration.

Acoustics: As the number of TL was not in proportion to the amount of cordectomies and as the AMPEX model was initially trained on the Ghent database (which mainly consisted out of TL-files), we added 19 additional files from the former Ghent database to the international database before performing calculations. The same 8 acoustic parameters as in the Ghent study were extracted from the text passages. Through linear combination of the acoustic features we designed a regression model and applied it on 4 of the 5 subsets (IINFVo), predicting the 5th subset. This was then compared to the subjective ratings of the clinician. Pearson correlation and standard deviation (between the
prediction and the clinician’s score) were only moderate (Pearson < 0.62; standard deviation >1.8). Fig. 2 demonstrates the values for the predictive scores and subjective scores for ‘impression of intelligibility’.

IV. DISCUSSION

The IINFVo rating scale seems to constitute a reliable tool for the perceptual assessment of substitution voices and could form a viable alternative to the GRBAS scale. In contradiction to the original reliability score conducted on the Ghent data, the analysis of the international data demonstrates that correlation between the two ‘I’s is sufficiently low (r=0.7) not to discard anyone of them. As the first I includes all features and reflects a general appreciation of the voice which is similar to the definition of Filter and Hyman, this is in agreement with their statement that ‘Intelligibility’ and ‘Acceptability’ only share 45 % common variance and thus advocate including both in a research design [10]. The highest correlations are now found between ‘Impression of Intelligibility’ and ‘Fluency’ (0.86) and between ‘General Impression’ and ‘Voicing’ (0.83). The latter supports the theory that SV are perceived as qualitatively better (General Impression) when speech is voiced and unvoiced where it is supposed to be voiced or unvoiced. The high correlation between ‘Impression of Intelligibility’ and ‘Fluency’ may support the theory that Intelligibility is mainly determined by the voicing length and fluent speech production and to a lesser extent by voicing itself. Variance analysis suggests that perception can differentiate surgery type and vibration source. Further analysis reveals that this is mainly due to the worst scores for TL patients on the parameters IIFVo and the worst scores for the CH(E)P patients on the dimension Noise.

There is only a low agreement between the perceptual evaluation by professionals and the self-evaluation of the patient’s voice (the highest is for ‘Voicing’: 0.46). Together with the fact that there is no correlation at all with the perceived disability, our data could, surprisingly, suggest that oncology patients mainly suffer from other co-morbidities (e.g. dysphagia, existence of a stoma) or psychological distress.

The AMPEX acoustic analysis seems capable of differentiating between various SV types [2]. Preliminary results in this trial however, show only a moderate agreement between the predicted scores and the subjective rating. There can be various reasons for the low concordance. First, the AMPEX model was formerly trained on mainly laryngectomy speech. The fact that there are far more cordectomies and partial laryngectomies in the international trial can induce errors. Secondly, the subjective ratings were performed by only one clinician. Although the clinicians had the availability of reference speech samples for each acoustic feature, there were no real training sessions preceding the rating. For this, we will compare the subjective scores of the clinician with our personal rating and additionally with an independent jury. If the concordance with these last ratings and the AMPEX is substantially better, we advocate an intensive training of the IINFVo scale, eventually through developing an audio CD in several languages. This of course implies a large database.
REFERENCES


