Effect of voice therapy: measurement and evaluation

R. Speyer

Effect of voice therapy: measurement and evaluation

R. Speyer

The work presented in this thesis was performed at the Department of ENT/Phoniatrics of the University Medical Center Utrecht and supported by the Dutch Health Care Insurance Council.

ISBN: 9064642788

Effect of voice therapy: measurement and evaluation

Effecten van stemtherapie: meten en evalueren

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. dr. W.H. Gispen,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op
dinsdag 14 oktober 2003 des middags te 16:15 uur

door

Renée Speyer

Geboren op 9 augustus 1963 te Nijmegen

Promotor: Prof. Dr. P.H. Dejonckere

Co-promotor: Dr. Ir. G.H. Wieneke

Faculteit: Department of E.N.T./Phoniatrics University Medical Center Utrecht

Contents

Promotor:	Prof. Dr. P.H. Dejonckere
Co-promotor:	Dr. Ir. G.H. Wieneke
Faculteit:	Department of E.N.T./Phoniatrics, University Medical Center, Utrecht

Chapter 1

General Introduction

General Introduction

Introduction

As it becomes more and more accepted that a medical treatment must be evaluated by scientific methods, paramedical therapies as well need objective evaluation according to current standards of evidence based medicine. Evaluation of logopedic voice therapy[1] fits into this growing interest. If this interest comes from organizations responsible for health care budgets or from health care insurances, the main point will be the effectiveness of voice therapy in general. This is directly related to the actual clinical practice of speech therapists and ENT-specialists or phoniatricians. A study of this kind of effectiveness should therefore include all kind of phoniatric diagnoses - as far as there is an indication for voice therapy -, as well as the diversity of existing voice therapies. Such a study answers another question than a study of a restricted group of patients undergoing a specific therapy under well-controlled experimental conditions. It will be very unlikely that the therapy effects found in such a study, can be generalized to other groups of subjects or therapies. Thus, for the effectiveness of voice therapy in general, it will be necessary to use a large group of patients suffering from diverse functional dysphonias or organic pathologies of the vocal folds with a functional component, without restrictions on the type of voice therapy.

The success or lack of success of a voice therapy can be assessed using different aspects of the voice production: for example, the changes in voice quality, in the biomechanical properties of the vocal folds, or in the self-evaluation or satisfaction by the patient. If voice is regarded as a multidimensional phenomenon (Hirano, 1989) all aspects considered as it is expected that patients will not show an abnormality in all aspects of voice, nor an improvement on all of these aspects. The Committee on Phoniatrics of the European Laryngological Society (Dejonckere et al., 2001) made the following recommendations for a minimal set of multidimensional measurements for functional assessment of voice pathology: perceptual rating, videostroboscopy, acoustic analysis, aerodynamic measures, and subjective rating by the patient. The only way to gain insight into the complete effects of voice therapy, is to design a study in which changes of different aspects on the voice are examined by a diversity of evaluation instruments. The individual effects per patient on the evaluation tools can be of particular interest. The measurement of group effects reveals no detailed information on individual success with respect to the variables that have been studied. In particular, when the patient population is inhomogeneous, therapy effects may be statistically significant for a whole group of patients, whereas the result can be quite different for subgroups of patients.

The literature on the effects of voice therapy in dysphonic subjects can be divided according to their choice of using either specific or broad patient populations, or using

[1] In the present study 'voice therapy' will be used as equivalent to 'logopedic voice therapy'. The specification 'logopedic' is mainly used in European countries, but unknown in other countries, like in the USA. A logopedic voice therapy does not include pharmacological or surgical treatments..

one specific therapy or several types of voice therapies. In the latter case, a diversity of therapies adjusted to the phoniatric diagnoses and patient's needs are included in the study. Furthermore, a classification of the studies in literature can be made, based on the different type of assessment tools used to evaluate therapy effects or based on methodological aspects of the study design such as the use of a placebo group or the applied statistical analyses. For ethical and practical reasons, very often no placebo group is included. Such a group consists of control subjects receiving no voice therapy. A multiple baseline design, providing the possibility of repeated measurements during a prolonged baseline period, could partially be an alternative. Furthermore, not all studies describe pre- and post-therapy measurements. Sometimes, only after accomplishing therapy, the voice is thoroughly assessed. But even if studies do have pre- and post-therapy measurements, the voice assessment is frequently performed knowing of which voice sample originates from which moment of measurement, instead of using a more objective assessment in which no information is available on whether a voice sample has been recorded pre- or post-therapy. Finally, many studies lack statistical analyses to verify the significance of the differences between the measurements, and provide only a global description of the data found.

Multidimensional assessment tools

Usually, the perceptual evaluation of voice quality is considered to be the gold standard for voice assessment. In the practice of a speech therapist it will be very often the only evaluation tool available. Voice quality is then described in terms like breathiness, roughness, harshness, etc. Multiple systems of perceptual classification have been suggested by different authors: for example, the Buffalo Voice Profile (Wilson, 1979), the Vocal Profile (Laver,1980), the GRBAS (Hirano, 1981), the multidimensional model for voice production by Perkins (1983), the classification of voice qualities by Wendler (1986) and the SVEC (Hammerberg, 1998). However, perceptual evaluation involves problems like the unstable internal standards for comparing speech stimuli (Gerratt et al., 1993) and the lack of universally accepted definitions for perceptual concepts (Kreiman et al., 1994). Another way of evaluating voice quality in a more objective manner, is acoustic analysis. Algorithms describe per analyzed sample, for example, the variability in pitch period and in peak-to-peak amplitude (jitter and shimmer) or the ratio of energy of inharmonic to harmonic components (noise). This method shows imperfections as well, for example, the possibility of errors in pitch tracking, the inadequacy of acoustic analysis in very aperiodic vocal vibrations, and the use of unnatural speech samples such as sustained vowels. The acoustic analysis is considered to be more directly related to the vocal folds vibration, as is laryngostroboscopy.

The technique of laryngostroboscopy provides direct information on the source of sound production: the vocal folds. Video recordings are made of the laryngeal structures and the vocal fold vibration employing rigid or flexible scopes. Two sources of light are used: normal light and stroboscopic light. The use of stroboscopic light during the vibration of the vocal folds, can provide the optical illusion of a static image, when the frequencies of the light flashes and the vocal fold vibration are equal. When the light is

flashed at frequencies that differ slightly from the vibration of the vocal folds, the vibration of the vocal folds is seen in slow motion. By means of visuo-perceptual evaluation the morphological and functional abnormality of the vocal folds and the glottal waveform can be described. Several protocols have been developed (e.g. Hirano et al, 1993, Dejonckere et al, 1998, Poburka, 1999). Recent advances in the technique of digital processing of laryngeal images have lead towards the development of methods for deriving objective measures from such endoscopic examinations.
The voice range profile, or phonetogram, describes the laryngeal possibilities with respect to the fundamental frequency and the sound intensity (Damsté, 1970, Schutte & Seidner, 1983). The maximal and minimal intensity that the patient can produce is plotted against the fundamental frequency. The voice range profile is considered to be a useful tool in the evaluation of therapy effects, because it represents the maximal vocal capacities. Aerodynamic parameters like maximum phonation time and the phonation quotient (the ratio of vital capacity and maximum phonation time) are widely used clinical measures. Only more recently, the quality-of-life measurements have become part of the voice assessment procedures. When the effects of therapy are evaluated, the patient's well-being cannot be neglected. Shown therapy effects demonstrated, using the above-mentioned objective evaluation tools, must be compared with the beneficial or negative changes as experienced by the patient him- or herself. In literature, a growing interest is found in the self-evaluation of patient's handicap as a result of the voice disorder.

Therapy effects in literature

Studies on therapy effects have become more frequent, especially during the last two decades. Table I represents a review of relevant scientific studies that describe aspects of the effects of voice therapy in dysphonic patients. Studies that consider merely outcome measurement after laryngeal surgery or pharmacological treatment, are not quoted. The list is restricted to subjects with dysphonia on a functional or organic base without any neurological origin such as Parkinson's disease. The studies are classified into three main categories based on phoniatric diagnoses: functional dysphonia (or muscle tension dysphonia), organic dysphonia, and functional plus organic dysphonia (respectively, Table IA, IB, and IC). These categories are subdivided into three groups according to the way data were handled. One group consisted of studies that described the post-therapy situation without comparable information on the voice status before the onset of therapy. Therefore, this first group, including studies by Koufman and Blalock (1982, 1991), Lancer et al. (1988) and Raabe and Pascher (1999), is considered of less importance. Another group of studies showed descriptive statistics to evaluate the therapy outcome. The majority of studies, the third group, used statistical analyses for comparing pre- versus post-therapy data. For each study, the following data are summarized: the number of subjects, the diagnostic group(s), the evaluation techniques, the kind of therapy used, and the effects of therapy or author's main key findings. The last two groups of studies will be described briefly.

Functional dysphonia

One of the earliest studies on voice therapy effects in patients with functional dysphonia was done by Wedin and Ögren (1982). Their population (N=6) included only two subjects with phonasthenic symptoms, two professional singers, and two subjects with normal untrained voices. After a voice training program, the authors concluded that their training seemed to be effective in bringing the pitch to its optimal range. However, the change was greater for the professional and normal subgroup than for the phonasthenic subjects. No exact data were available. Furthermore, the group of dysphonic patients was too small to justify generalizations for other patients with dysphonia. Another study by D'Antonio (1987) was based on the results of a single subject receiving therapy. This study intended to illustrate the usefulness of a 'perceptual-physiologic' approach to the evaluation and therapy of dysphonia rather then to demonstrate the effects of voice therapy. In this approach, a detailed case history is combined with auditory-perceptual, aerodynamic, and videonasoendoscopic evaluations. According to the author, this approach emphasizes the importance of quantitative analysis of voice production and the determination of the physiologic consequences of the abnormal voicing. A more recent study done by Prathanee (1996) described the positive therapy effects in seven subjects with mutational falsetto voices after ear training practice. The precise data were not mentioned and, moreover, the number of subjects was very small. Enderby and John (1999) studied a group of 99 subjects with non-organic dysphonia. The patients described the change after speech and language therapy using 11-point self-evaluation scales that were related to the domains of impairment, disability, handicap and well-being. Outcome scores, representing the percentage of change after therapy, are available for five different speech and language services. However, an unknown number of patients had not yet finished their therapy within the 9-month trial. It is concluded that different speech and language services have different impacts on the number and type of domains involved and that patients are being discharged at different phases of their recovery.

Another seven studies on functional dysphonia provided statistical analyses of the pre- versus post-therapy data to support their conclusions. Hammerberg (1987) included twelve male subjects with functional mutational disorders receiving voice therapy (larynx depressing exercises) and psychological counseling. The evaluation tools used were perceptual evaluation and analysis of the distribution of the fundamental frequency as derived from the acoustic signal. After therapy a statistically significant decrease of the deviant voice qualities instability, breathiness, hypofunction (laxness), and diplophonia was demonstrated using a Wilcoxon's matched-pairs signed ranks test. The most deviant voice qualities, such as instability, breathiness and diplophonia, had diminished, while the pitch and register had stabilized, showing a reduced variability. This resulted in an improved perceptual impression of the voice. In contrast to the rather small group of subjects used in the study by Hammarberg, Kitzing and Åkerlund (1993) used a large group of 174 subjects with non-organic voice disorders. Tape recordings before and after therapy were analyzed by long-time averaged voice spectrograms (LTAS) and compared with the results of a global perceptual rating of the voice qualities on a three-point scale (t-test on paired observations). There was no significant change of the LTAS in voices with negligible perceptual amelioration after therapy. In voices with considerable perceptual changes after therapy, the LTAS showed only an increase in intensity, but the general configuration of the spectral envelope remained unchanged. There was only a weak positive correlation between the quality ratings and parameters of the spectra. In two studies by Roy and Leeper (1993) and Roy et al. (1997) the effects of the manual laryngeal musculoskeletal tension reduction technique were evaluated by means of a perceptual severity rating and acoustic analysis. The latter study which can be considered as an extension of the first, included a population of 25 women with functional dysphonia and focused on short-term as well as on long-term therapy outcomes. By means of repeated-measures analysis of variance, time trends within the data were evaluated. Subjects demonstrated consistent and significant improvement across perceptual and acoustic indices of vocal function immediately after therapy and during the follow-up period. The authors concluded on the basis of patient reports that the short-term results were impressive, but that the long-term results were less robust. Finally, three studies done by Carding and Horsley (1992) and Carding et al. (1998, 1999) are being mentioned. The patients in all three studies ($30<N<45$) had non-organic dysphonia. The studies used a similar study design in which indirect voice therapy was compared with a therapy in which direct and indirect therapy were combined. Indirect therapy techniques focused on managing the aspects which contribute to the voice problem (such as vocal abuse patterns or poor vocal hygiene). Direct therapy techniques focused on modifying certain aspects of improper voice production in order to promote appropriate and efficient voice production. A control group receiving no therapy was included. The main findings of the most recent study (1999) applying non-parametric tests showed statistically significant differences between all three groups in the amount of change for voice severity ratings ('blinded' assessment), electrolaryngographic data (visual interpretation of the Lx waveform), shimmer measurements and on ratings provided by a patient questionnaire. Other parameters such as fundamental frequency, signal-to-noise ratio, and jitter failed to show significant differences between the three groups. Most of the patients of the control group (86%) showed no significant change on any of the parameters, whereas respectively 46% and 93% of the indirect and combined therapy group showed changes in voice quality on all parameters.

Organic dysphonia

As early as in 1981, Gould et al. described the effects of voice therapy according to each patient's need on contact granuloma of the vocal fold (N=17). The evaluation of the success of the therapy was based on laryngoscopic findings, a perceptual rating by a speech pathologist and patient's subjective evaluation. The authors concluded that voice therapy was an effective mode of therapy in many cases of contact granuloma. However, some patients did not improve at all. Heuer et al. (1997) focused on patients with unilateral recurrent nerve lesions (N=41) and formed four groups of patients according to gender and type of therapy (voice therapy or combination of laryngeal surgery and voice therapy). Acoustic parameters, aerodynamic measurements and a measurement of glottal function (the quasi-open quotient using laryngeal EMG) were used in the evaluation of therapy effects. Only pre- and post-therapy data of so-called representative subjects per group were displayed. The authors found their data promising in differentiating between groups of patients who could be treated by voice therapy alone and those who require surgery. However, the conclusions drawn from these last mentioned studies are restricted to the specific groups of voice patients included in the experiment. Gordon et al. (1997) included a larger and more diverse group of subjects suffering from dysphonia resulting from vocal misuse or abuse, with a variety of secondary pathologies including soft nodules, polyps, Reinke's oedema and fold thickening (N=200). Patients with significant disordered air usage were assigned to a therapy program in order to normalize the aerodynamic parameters. The other patients were put on a monitoring program including voice hygiene advice and relaxation exercises. The outcome of the therapy was assessed by airflow test scores and videolaryngoscopy and was translated freely in terms of: problem resolved, prolongation of therapy or monitoring necessary, or therapy discontinued due to non-compliance. This resulted in the successful outcome for 41.5% of the referrals in the advice and monitoring group, and for 35% of the referrals in the voice therapy group. The program was discontinued without result for 11% of the referrals. Some 12.5% of the subjects had received sufficient reassurance and advice from initial attendance at the ENT clinic and rejected the offer of further voice assessment.

The studies mentioned above, used no statistical analyses to test for the significance of changes between pre- and post-therapy data. The following three studies on therapy effects did use statistical analyses to test their data. All patients involved were diagnosed as having vocal fold nodules. Murry and Woodson (1992) used a global four-point scale of perceptual improvement in order to compare pre- and post-therapy recordings. Two judges rated pairs of pre- and post-therapy recordings, without any knowledge of the purpose of the study. They did not know if these patients underwent any therapeutic or surgical procedures. The subjects were divided in three groups according to the type of therapy: voice therapy (N=28), voice therapy following surgery (N=20), or combined treatment by an otolaryngologist and a speech pathologist (N=11). The pre- and post- therapy judgements were subjected to a complex χ^2 analysis (Chi square) and mean differences between pairs of groups were analyzed using a Mann-Whitney U-test. The authors concluded that a satisfactory improvement in the voice could be obtained using any of the three approaches. These findings are in line with the results of Verdolini et al. (1995) and Holmberg et al. (2001). Verdolini et al. assessed the effects of two types of voice therapy (confidential and resonant therapy) using measurements of phonatory effort, auditory-perceptual status of voice, and laryngeal appearance. On all measurements, a greater proportion of subjects receiving therapy improved, as compared with a group of control subjects who received no therapy. Statistical analyses (z-scores and corresponding level of significance) indicated that these results exceeded chance levels for the combined therapy groups, but not for the control group. Association tests showed that the likelihood of benefitting from therapy directly covaried with estimates of ongoing compliance (continued utilization of therapy techniques following therapy discontinuation), but not with therapy type. The results, however, were based on a total group of 13 subjects. Holmberg et al. (2001) used a population of patients with vocal nodules, that was even smaller (N=11). Analyses of variances tested the effects of a behaviorally based voice therapy protocol. The perceptual and physiological progressive changes suggested that voice therapy had a positive effect for the majority of the patients.

Functional and organic dysphonia
There are a number of studies in which mixed groups of subjects are included; that is a population in which both functional and organic dysphonia are admitted. Xu et al. (1989) assessed the effects of a diaphragm support breath pattern (yawning breath pattern) in voice therapy in patients with incomplete glottal closure (N=30), vocal nodules (N=41), and recurrent laryngeal nerve paralysis (N=20). Based on laryngoscopic examination, vocal function tests, perceptual evaluation, and patient's subjective evaluation, the group changes after therapy were globally described. The authors concluded that about one third of the patients mastered the yawning breath pattern perfectly, resulting in satisfactory improvement of the symptoms. The performances of the other patients varied between fair and slight or unsatisfactory improvement. Murry and Rosen (2000) as well as Casper (2001) used evaluation instruments that were restricted to quality-of-life measurements. Murry and Rosen studied the pre- and post-therapy data on the Voice Handicap Index, a patient self-assessment questionnaire, in a group of 37 subjects suffering from muscle tension dysphonia, benign vocal fold lesions or unilateral laryngeal nerve paralysis. Subjects received surgery and/or voice therapy. In general, a 50% or greater improvement in the mean VHI was found. Overall, 81% of the patients reported a reduced perception of voice handicap. A histogram of the pre- and post-therapy group data per diagnostic category, visualizes positive therapy effects. Using norm values from literature, the post-therapy histogram was significantly shifted in relation to the pre-therapy histogram. However, in the description of therapy results, the authors did not distinguish between voice therapy and surgery. Neither did Casper (2001) while evaluating the therapy effects on two patient self-evaluation scales in a group of 184 adults with dysphonia. Four diagnostic categories were included: benign lesions, unilateral vocal fold paralysis, postoperative dysphonia and functional problems (muscle tension dysphonia). The overall conclusion that patients perceived great benefits of voice therapy, whether surgical or behavioral, was not further specified per therapy group. After completion of the therapy seventy-nine percent of the patients rated their voices better than before therapy, and 8% believed that they were back to normal. However, no exact numbers on therapy groups were given, nor any results of a statistical analysis.

Three studies on mixed subject groups did provide pre- and post-therapy data that were statistically tested for significant differences: Kotby et al. (1991), Fex et al. (1994) and Bassiouny et al. (1998). All three studies described the effects of the accent method of voice therapy, a holistic approach for behavior modification of the voice. The evaluation instruments used by Kotby et al. in a group of 28 subjects consisted of a patient's grading of the voice function, auditory perceptual assessment, indirect microlaryngovideostroboscopy and some aerodynamic values. Positive significant therapy effects (paired t tests) supported the notion that voice therapy was indicated mainly in cases of habitual functional voice disorders and in selected pathological lesions (nodules), as well as in some organic laryngeal ailments (vocal fold paralysis). Fex et al. (1994) used acoustic analysis in a small group of ten subjects with functional voice disorders amongst whom three had developed bilateral vocal nodules. The Wilcoxon signed-ranks test was used for statistical analysis. Significant improvement on pitch and amplitude perturbation quotients, normalized noise energy and fundamental frequency were found. A much more detailed study was done by

Bassiouny et al. (1998). These authors compared the results of 42 subjects with a variety of vocal pathology divided into two therapy groups: one group receiving only voice hygiene advice and another group receiving accent exercises a well. A diversity of evaluation instruments was used: a patient's own grading of severity of voice dysfunction, auditory perceptual assessment, visuo-perceptual evaluation of videolaryngostroboscopy, aerodynamic measures, acoustic analysis, and inverse filtering measures. The difference in improvement in both therapy groups at the end of the observation period was generally significant (paired t test) in favor of the group receiving accent exercises as well.

Summarizing literature

Summarizing the literature on the effects of voice therapy in dysphonic subjects, the overall impression is that statistically significant positive but modest and varying therapy effects are found. Many of these studies in literature, however, cope with diverse methodological problems. For example, in case of perceptual evaluation, it is very often unclear whether the data have been offered to the listeners in randomized order and without any information on pre- or post-therapy status of the voice samples. Some studies use very subjective instruments to evaluate therapy effects without any statistical foundation. Furthermore, the lack of a good alternative for a control group receiving no therapy, or a prolonged baseline period in which repeated measurements are performed, invalidates many study designs, and, of course, those therapy outcome studies that have no pre-therapy data at all, are of very little value.

However, besides such methodological difficulties, other problems appear as well when looking at the possibilities of generalizing the therapy effects found in literature. The results found in literature are insufficient to be generalized to the more common situation in which patients visit an ENT-specialist or phoniatrician, and then, after being diagnosed, get referred to a voice therapist if necessary. Usually, the results of the effect studies are based on small or restricted groups of subjects who belong to specific phoniatric diagnostic categories. Generally, the types of voice therapy included have been limited, as well as the number of speech therapists involved. Furthermore, very often only restricted sets of assessment instruments have been used in the experiments. Therefore, the conclusions of most of the above-mentioned studies cannot be easily generalized or compared to one another. If one is interested in the overall effects of voice therapy in such a clinical setting, several kinds of voice therapies, a large group of voice therapists and diverse pathologies should be included in the study.

Scope of the thesis

This study aimed at an evaluation of the standard situation in the Netherlands, by looking at the real practice of referrals of a Phoniatric department. As few restrictions as possible were made concerning phoniatric diagnoses or voice therapies. However, for methodological reasons, only patients with a chronic dysphonia were admitted to the study. In those patients, spontaneous recovery over a period as long as the therapy lasted, was not expected. Furthermore, acute dysphonia is - as a rule - not an indication for voice therapy. All patients suffered from a functional dysphonia or some benign vocal fold lesion. Patients with psychogenic dysphonia or aphonia were excluded. In this study, the overall group effects as well as the effects for the individual patients were investigated.

With financial support by the Health Care Insurance Council, a national research project on the effects of voice therapy was started in 1997 at the Institute of Phoniatrics of the University Hospital Utrecht in the Netherlands. In order to objectify the effects of voice therapy, following issues are taken into account in this study: the population of subjects is large enough to allow for later generalization of the results to comparable experimental designs or situations. Measurements are available from before therapy onset as well as after accomplishment of therapy. Statistical analyses-are used to test for differences between the pre- and post-therapy data. As for ethical reasons, no placebo group or group consisting of control subjects receiving no voice therapy could be included, only chronically dysphonic adults were admitted to the research assuming no spontaneous recovery. By means of repeated measurements during a prolonged baseline period, a so-called multiple baseline design, the chronic character of the dysphonia was checked. A final subject refers to the assessment tools to evaluate therapy effects. Voice is considered to be a multidimensional phenomenon, that means that patients are not expected to show abnormal baseline data or positive therapy effect on all aspects of the voice. Therefore, the assessment procedures should cover all these aspects or dimensions as far as possible. In this study the choice of multidimensional evaluation tools is mainly conform the recommendations by the Committee on Phoniatrics of the European Laryngological Society (Dejonckere et al., 2001): perceptual evaluation, acoustic analysis, videolaryngostroboscopy, voice range profile, aerodynamic assessment, and quality-of-life measurement.

The first study (chapter two) of this thesis describes the amount of intra subject variability to be expected when using an objective acoustic analysis for evaluation of voice quality. It also considers how such variability affects the assessment of therapy outcome. As the variability of perturbation parameters increases with the magnitude of the parameter, the inaccuracy was characterized by a relative error. Using these error measurements, the relation between ratios of the post-therapy to the pre-therapy data and the corresponding level of significance of improvements could be computed.
The results of this study are used in the subsequent evaluation of voice therapy effects for individual patients.

A second study (chapter three) was conducted to examine whether objective measurements derived from digitized laryngeal stroboscopic images could be used to detect changes in vocal fold vibration and in the size of benign lesions after three months of voice therapy.

Chapter four describes on the one hand, the effect of voice therapy on the voice quality as determined by perceptual rating and acoustic analysis. On the other hand, the effects on the vibrational pattern of the vocal folds were determined using visuo-perceptual evaluation of videolaryngostroboscopy. This study relates the effects of therapy as found on the different instruments as well as possible relations in therapy success and phoniatric diagnoses.

Another instrument to evaluate therapy effects, the voice range profile, is used in a study presented in chapter five. This study had two objectives: to investigate which parameters in the voice range profile of dysphonic patients showed significant changes after voice therapy and, furthermore, to determine the size of the demonstrated effects and any possible relationship to phoniatric diagnoses.

However, besides improvements of the vocal fold vibration and the voice quality, patients are supposed to experience less problems after voice therapy in their working or social environment. Two quality-of-life measurements are used to investigate this (chapter six). Therapy outcome as well as the differences and similarities between the two self-assessment instruments are discussed.

A general discussion in which the findings of these studies are integrated, is presented in chapter seven.

Literature

Bassiouny S. Efficacy of the accent method of voice therapy. *Folia Phoniatr.Logop.* 1998;50:146-164.

Carding PN, Horsley IA. An evaluation study of voice therapy in non-organic dysphonia. *Eur.J.Disord.Commun.* 1992;27:137-158.

Carding PN, Horsley IA, Docherty GJ. The effectiveness of voice therapy for patients with nonorganic dysphonia. *Clin.Otolaryngol.* 1998;23:310-318.

Carding PN, Horsley IA,, Docherty GJ. A study of the effectiveness of voice therapy in the treatment of 45 patients with nonorganic dysphonia. *J.Voice* 1999;13:72-104.

Casper JK. Treatment outcomes in occupational voice disorders. In Dejonckere PH, ed. *Occupational voice: care and cure.* The Hague: Kugler Publications; 2001:187-199.

D'Antonio L, Lotz W, Chait D, Netsell R. Perceptual-physiologic approach to evaluation and treatment of dysphonia. *Ann.Otol.Rhinol.Laryngol.* 1987;96:187-190.

Damsté PH. The phonetogram. *Pract.Otorhinolaryngol.(Basel)* 1970;32:185-187.

Dejonckere PH, Crevier L, Elbaz E, Marraco M, Millet B, Remacle M, Woisard V. Quantitative rating of video-laryngostroboscopy: a reliability study. *Rev.Laryngol.Otol.Rhinol.* 1998;119:259-260.

Dejonckere, PH, Bradley, P, Clement, P, Cornut G, Crevier-Buchman L, Friedrich G,van de Heyning P, Remacle M, Woisard V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur.Arch.Otorhinolaryngol.* 2002;258:77-82.

Enderby PM, John A. Therapy outcome measure in speech and language therapy: comparing performance between different providers. *Int. J. Language & Communication Disorders* 1999;34:17-429.

Fex B, Fex S, Shiromoto O, Hirano M. Acoustic analysis of functional dysphonia: before and after voice therapy (Accent method). *J.Voice* 1994;8:163-167.

Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research.* 1993;36:14-20.

Gordon MT, Pearson L, Paton F, Montgomery R. Predictive assessment of vocal efficiency (PAVE). A method for voice therapy outcome measurement. *J.Laryngol.Otol.* 1997;111:129-133.

Gould WJ, Bloch CS, Hirano M. Effect of voice therapy on contact granuloma of the vocal fold. *Ann.Otol.Rhinol.Laryngol.* 1981;90:48-52.

Hammarberg B. Pitch and quality characteristics of mutational voice disorders before and after therapy. *Folia Phoniatr.* 1987;39:204-216.

Hammarberg B. Perception and acoustics of voice disorders - a combined approach. *Proceedings of Voicedata98: Symposium on databases in voice quality research and education.* University of Utrecht; 1998;1-6.

Heuer RJ, Sataloff RT, Emerich K, Rulnick R, Baroody M, Spiegel JR, Durson G, Butler J. Unilateral recurrent laryngeal nerve paralysis: the importance of "preoperative" voice therapy. *J.Voice* 1997;11:88-94.

Hirano M. *Clinical examination of voice.* Wien: Springer-Verlag, 1981.

Hirano M. Objective evaluation of the human voice: clinical aspects. *Folia Phoniatr.* 1989;41:89-144.

Hirano M, Bless DM. *Videostroboscopic examination of the larynx.* London: Whurr Publishers, 1993.

Holmberg EB, Hillman RE, Hammarberg B, Södersten M, Doyle P. Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *J.Voice* 2001;15:395-412.

Kitzing P, Akerlund L. Long-time average spectrograms of dysphonic voices before and after therapy. *Folia Phoniatr.Basel.* 1993;45:53-61.

Kotby MN, El-Sady SR, Basiouny SE, Abou-Rass YA, Hegazi MA. Efficacy of the accent method of voice therapy. *J.Voice* 1991;5:316-320.

Koufman JA, Blalock PD. Classification and approach to patients with functional voice disorders. *Ann.Otol.Rhinol.Laryngol.* 1982;91:372-377.

Koufman JA, Blalock PD. Functional voice disorders. *Otolaryngol.Clin.North Am.* 1991;24:1059-1073.

Kreiman J, Gerratt BR, Berke GS. The multidimensional nature of pathologic vocalquality. *J.Acoust.Soc.Am.* 1994;96:1291-1302.

Lancer JM, Syder D, Jones AS, Le BA. The outcome of different management patterns for vocal cord nodules. *J.Laryngol.Otol.* 1988;102:423-427.

Laver J. *The phonetic description of voice quality.* Cambridge: Cambridge University Press, 1980

Murry T, Woodson GE. A comparison of three methods for the management of vocal fold nodules. *J.Voice* 1992;6:271-276.

Murry T, Rosen CA. Outcome measurements and quality of life in voice disorders. *Otolaryngol.Clin.North Am.* 2000;33:905-916.

Perkins WH. Quantification of vocal behavior: a foundation for clinical management of voice. In: Bless DM, Abbs JH., eds. *Vocal fold physiology*, San Diego; College-Hill Press, 1983:426-431.

Poburka BJ. A new stroboscopy rating form. *J.Voice* 1999;13:403-413.

Prathanee B. Mutational falsetto voices: voice therapy. *J.Med.Assoc.Thai.* 1996;79:388-394.

Raabe J, Pascher W. [Reinke's edema: an investigation of questions related to etiology, prognosis and the effectiveness of therapeutic methods] Das Reinke-Odem: Eine Untersuchung zu Fragen der Atiologie, der Prognose und der Wirksamkeit therapeutischer Interventionen. *Laryngorhinootologie* 1999;78:97-102.

Roy N, Leeper HA. Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: perceptual and acoustic measures. *J.Voice* 1993;7:242-249.

Roy N, Bless DM, Heisey D, Ford CN. Manual circumlaryngeal therapy for functional dysphonia: an evaluation of short- and long-term treatment outcomes. *J.Voice* 1997;11:321-331.

Schutte HK, Seidner W. Recommendation by the Union of European Phoniatricians (UEP): standardizing voice area measurement/phonetography. *Folia Phoniatr.* 1983;35:286-288.

Verdolini MK, Burke MK, Lessac A, Glaze L, Caldwell E. Preliminary study of two methods of treatment for laryngeal nodules. *J.Voice* 1995;9:74-85.

Wedin S, Ogren JE. Analysis of the fundamental frequency of the human voice and itsfrequency distribution before and after a voice training program. *Folia Phoniatr.Basel.* 1982;34:143-149.

Wendler J, Rauhut A, Krüger H. Classification of voice qualities. *Journal of Phonetics* 1986;14:483-488.

Wilson DK. *Voice problems of children.* Baltimore: Wiliams & Wilkins, 1979[2].

Xu JH, Ikeda Y, Komiyama S. Bio-feedback and the yawning breath pattern in voice therapy: a clinical trial. *Auris.Nasus.Larynx.* 1991;18:67-77.

Table IA Functional dysphonia

Diagnose	Data analysis	Reference	Subjects[A]	Evaluation technique[B]	Treatment(s) / Groups (G)	Author(s)'s conclusions / key findings
Functional dysphonia	Description post-therapy	Koufman & Blalock '82	52	6 (overall judgement on voice based on 1,2,4)	G1 voice therapy (N=35) G2 voice therapy & surgery (N=9) G3 voice therapy & psychiatric therapy (N=3) G4 no therapy (N=5)	Voice therapy is essential for patients with functional voice disorders and particularly preoperatively for vocal abusers with surgical lesions.
		Koufman & Blalock '91	322	6 (overall judgment of success versus failure based on patient/medical history, clinical examination voice laboratory data, ...)	voice therapy	Overall response to treatment reveals a successful outcome in 69% of cases, with patient motivation and compliance to voice therapy being the principal predictors of success. Continued cigarette smoking seems to be a factor correlated with a relatively poor outcome in functional voice disorders.
	Descriptive statistics	D'Antonio '87	1 ventricular fold phonation	245	'perceptual-physiologic approach'	The 'perceptual-physiologic approach' is important to both the evaluation and the treatment of voice disorders.
		Enderby & John '99	99 non-organic dysphonia	1	speech and language therapy	Evidence is given on the differences between the types of patients being referred to different providers of speech and language therapy. Different services have different impacts on the number and type of domains and services discharge their patients at different stages of their recovery.

Table IA Functional dysphonia

		Prathanee '96	7 mutational falsetto voices (aged 11-26 years)	2	ear training	Voices of all patients improved: 4/7 became "normal" within 5 sessions.
		Wedin & Ögren '82	2 subjects with phonasthenic symptoms, 2 professional singers, 2 untrained voices	6 (F_0, LTAS)	intensive 5-day voice training program	Generally, training seems to be effective in bringing the pitch to its optimal range.
	Statistical analysis	Carding & Horsley '92	30	1,2,6 (electrolaryngography, F_0-analysis)	G1 indirect therapy (N=10) G2 direct & indirect therapy (N=10) G3 no therapy / controls (N=10)	90% of G2 showed successful return to normal voice functioning as compared to 60% of G1 and 10% of G3. Statistically significant differences between the 3 groups after therapy on perceptual ratings and patient's questionnaires were found. Similar trends were found on electrolaryngography and F_0-analysis.
		Carding et al. ' 98	45	123	G1 indirect therapy (N=15) G2 direct & indirect therapy (N=15) G3 no therapy controls (N=15)	Significant differences exist between the three groups on the self-report questionnaire, voice quality ratings and pitch perturbation measurements. (13/15 patients of G3 showed no significant change on any of the measures; 7/15 of G1 and 14/15 of G2 showed significant improvement in voice quality.)

Table IA Functional dysphonia

		Carding et. al. '99	45	1,2,3,4,6 (electrolaryngography, F_0-analysis)	G1 indirect therapy (N=15) G2 direct & indirect therapy (N=15) G3 no therapy / controls (N=15)	Significant differences exist between the 3 treatment groups in the magnitude of change for voice severity, electrolaryngograph, and shimmer measurements and on ratings from a patient questionnaire. (Other measurements failed to show significant differences between the 3 groups.) 86% of the patients of G3 showed no significant change on any of the measurements; 46% of the patients of G1 en 93% of G2 showed significant changes in voice quality.
		Hammer-berg '87	12 functional mutational disorders (male adolescents / adults)	2,6 (F_0-distribution analysis)	voice therapy: larynx depressing exercises, psychological counseling.	After therapy the most deviant voice qualities had improved, together with a stabilization of the pitch and the register. These changes were statistically significant. (Observations show an improvement of the mutational voice disorder, as can be seen from the stabilization of the prepubertal pitch and normalization of the post-therapy voice.)
		Kitzing & Åkerlund '93	174 non-organic voice disorders (functional dysphonia)	2,6 (LTAS)	voice therapy	Nonsignificant change of the LTAS in voices with negligible improvement after therapy. In the voices with (perceptually) considerable change after therapy, the LTAS showed only an increase in intensity, but the general configuration of the spectral envelope remained unchanged.

Table IA Functional dysphonia

		Roy & Leeper '93	17	23	'manual laryngeal musculoskeletal tension reduction'	Results indicated a significant change to normal vocal functioning of the majority of patients within one treatment session. (82% (14/17) were considered as either normal or exhibiting only mildest dysphonic symptoms at perceptual evaluation of connected speech samples.)
		Roy et al. '97	25 (women)	2,3,6 (F_0)	'manual laryngeal musculoskeletal tension reduction'	Subjects demonstrated consistent improvement on perceptual and acoustic indices of vocal function immediately after treatment and during the follow-up period. (After therapy, 64% (16/25) were considered as either normal or exhibiting only mildest dysphonic symptoms at perceptual evaluation of connected speech samples. Furthermore, significant reductions in jitter, shimmer, and SNR were found, but no significant changes in F_0.) Short-term results are impressive, but on the basis of patient reports, long-term results are less robust.

A adult men and women, unless mentioned otherwise

B 1. Quality of life measures
2. Perceptual evaluation
3. Acoustical analysis
4. Videolaryngo(stro)scopy
5. Aerodynamic measure
6. Other

Table 1B Organic dysphonia

Diagnose	Data analysis	Reference	Subjects[A]	Evaluation techniques[B]	Treatment(s) / Groups (G)	Author(s)'s conclusions / key findings
Organic dysphonia	Descrip-tion post-therapy	Lancer et al.'88	20 unilateral vocal cord nodules	1 (long-term outcome)	G1 surgery (N=8) G2 surgery & voice therapy (N=6) G3 voice therapy (N=6)	Significant relationship found between recurrence rates and presence or absence of speech therapy treatment (with or without surgery) which supports the suggestion that speech therapy reduces the incidence of recurrence after treatment in vocal nodule patients. A significant relationship between smoking and recurrence is found.
		Raabe & Pascher '99	48 Reinke-oedema	2,4,6 (severity of breathing disorder)	G1 'decortication' & voice therapy (N=16) G2 'Mucosuction' & voice therapy (N=10) G3 voice therapy (N=12) G4 no therapy / controls (N=10)	'Decortication' is the most invasive method regarding voice function, but does not prevent recurrences. Mucosuction leads to faster voice rehabilitation, and there were fewer scars. Voice therapy combined with stopping smoking is necessary in every case. A considerable breathing disorder remained in all therapy groups. To obtain a good voice function, not every Reinke-edema has to be operated on. But if an operation is performed, a consistent voice therapy must follow.

Table 1B Organic dysphonia

	Descriptive statistics	Gordon et al. '97	143 dysphonia resulting from vocal misuse or abuse with a variety of secondary pathologies	6 (% based on 4 and 5: problem resolved, prolongation therapy or monitoring, therapy stop due to non-compliance)	G1 therapy program (N=74): normalization of aerodynamic parameters (subjects with sign. disorder of air usage) G2 monitoring program (N=69): voice hygiene advice & relaxation exercises (subjects without disorder of air usage)	Successful resolution of the problem: 41.5% of G1 and 35 % of G2.
		Gould et al. '81	17 contact granuloma	124	voice therapy (methods varied according patient's needs)	As a result of voice therapy: 9/14 granuloma disappeared, 4/14 reduced in size, 1 did not change; 4/17 phonation returned to normal, 7/17 improved, 6/17 did not change; 4/17 patients felt complete recovery, 10/17 showed some improvement, 3/17 showed no change.
		Heuer et al. '97	41 unilateral recurrent nerve lesions	3,5,6 (quasi-open quotient): exact data on 1 representative patient per group	G1 voice therapy (13 women) G2 voice therapy & surgery (6 women) G3 voice therapy (14 men) G4 voice therapy & surgery (8 men)	Findings indicate 1. some elements of objective voice assessment may provide useful prognostic information, 2. pre- and post-therapy objective measures are helpful in confirming subjective estimates of improvement, 3. treatment of a substantial percentage of patients using nonsurgical therapy alone may be satisfactory.

Table 1B Organic dysphonia

	Statistical analysis	Holmberg et al. '01	11 bilateral vocal nodules (women)	2,4,6 (mean SPL, F_0)	behaviorally based voice therapy (5 phases): vocal hygiene, direct facilitation, respiration, relaxation, and carryover	The results of decreased nodules and improved voice quality suggest that the voice therapy had a positive effect on the majority of the patients. (Significant effects of therapy were found for overall dysphonia, press, instability, gratings, roughness, vocal fry, and "scrape"; nonsignificant group effects were found for breathiness, aphonic instances, and lack of sonority.)
		Murry & Woodson '92	59 vocal fold nodules (including 1 adolescent)	2	G1 voice therapy (N=28) G2 surgery & therapy (N=20) G3 integrated management procedure (by an otolaryngologist-speech pathologist) (N= 11)	Satisfactory improvement of the voice can be obtained using any of the three approaches (therapy groups). 22/59 achieved the maximum rating of improvement (no perceptual indication of a voice disorder), while 4/59 showed no improvement; of those who obtained a maximal rating, 55% were in G3, 39.3% in G1, and 25% in G2.
		Verdolini-Marston '95	13 laryngeal nodules (women)	2,4,6 (phonatory effort)	G1 confidential voice therapy (N=5) G2 resonant voice therapy (N=3) G3 no therapy / controls (N=5)	On all measures a greater proportion of therapy subjects improved over the initial 2-week period, as compared with control subjects. The likelihood of benefitting from therapy directly covaried with compliance scores but not with therapy type. (Statistically supported.)

A adult men and women, unless mentioned otherwise

B 1. Quality of life measures
2. Perceptual evaluation
3. Acoustical analysis
4. Videolaryngo(stro)scopy
5. Aerodynamic measure
6. Other

Table 1C Functional plus organic dysphonia

Diagnose	Data analysis	Reference	Subjects[A]	Evaluation technique[B]	Treatment(s) / Groups (G)	Author(s)'s conclusions / key findings
Functional & organic dysphonia	Descriptive statistics	Casper '01	184 diverse (benign lesions, unilateral vocal fold paralysis, postoperative dysphonia, functional problems)	1	G1 surgery G2 voice therapy	Patient-perceived benefits of voice treatment, whether surgical or behavioral, are great. (On the completion of therapy, 79% of the patients rated their voices as better than before treatment and 8% as back to normal.)
		Murry & Rosen '00	37 diverse (muscle tension dysphonia, benign vocal fold lesions, unilateral vocal fold paralysis)	1	surgery and/or voice therapy	81% of the patients demonstrated a reduced perception of voice handicap, whether they were treated with surgery, voice therapy, or a combination. (50% or greater improvement of the mean Voice Handicap Index)
		Xu et al. '89	91 (41 vocal nodules, 20 recurrent laryngeal nerve paralysis, 30 incomplete glottal closure)	2,4,5,6 (voice range, SPL)	yawning breath pattern	About 1/3 of the patients mastered the yawning breath pattern perfectly and their symptoms improved satisfactorily.
	Statistical analysis	Bassiouny '98	42 diverse (nonorganic dysphonia, minimal associated pathological lesion, vocal fold immobility)	1,2,3,4,5,6 (inverse filtering)	G1 voice hygiene advice & Accent method (N=21) G2 voice hygiene advice (N=21)	The difference in improvement for most of the parameters in G1 and G2 after therapy was generally significant in favor of G1. The improvement from pre-test to mid-test to post-test values followed a linear tendency.

	Fex '94	10 (7 normal vocal cords, 3 bilateral vocal nodules)	3	Accent method	Significant improvement on pitch and amplitude perturbation quotient, NNE for 1-4 kHz and F_0.
	Kotby et al. '91	28 diverse (functional dysphonia, minimal associated pathological lesion, vocal fold immobility)	1245	Accent method	Patient's complaint: 89% showed a positive grade shift; Auditory perceptual assessment: 68% showed a significant degree shift in the overall grade of dysphonia; Indirect laryngovideostroboscopy showed a reduction in nodule size (6/6) and in the maximal phonatory gap (4/6); Aerodynamic parameters varied in the degree of significance.

A adult men and women, unless mentioned otherwise

B 1. Quality of life measures
2. Perceptual evaluation
3. Acoustical analysis
4. Videolaryngo(stro)scopy
5. Aerodynamic measure
6. Other

Chapter 2

The use of acoustic parameters for the evaluation of voice therapy for dysphonic patients

R. Speyer, G.H. Wieneke, P.H. Dejonckere
Submitted (Acustica united with Acta Acustica)

The use of acoustic parameters for the evaluation of voice therapy for dysphonic patients

Abstract

This study describes the amount of intra-subject variability to be expected when using an objective, acoustic analysis to evaluate voice. It also considers how such variability affects the assessment of a therapy outcome. In total, 82 patients suffering from chronic dysphonia were asked to produce three sustained vowels /a:/ at a comfortable pitch and loudness level. Three recordings were made over a six-week period. For each sample, acoustic variables characterizing jitter, shimmer and harmonic-to-noise ratio were computed. Statistical analysis was then used to estimate the expected intra-subject variability per acoustic parameter.

The variability of perturbation parameters increases with the magnitude of the parameter. Therefore, the inaccuracy was characterized by a relative error (coefficient of variation). For the acoustic parameters, distinctive inaccuracies in series of consecutive sustained vowels were found: the coefficients of variation ranged from 14% to 33%. When an individual therapy is evaluated by acoustic perturbation measures, the ratio of the post-therapy to the pre-therapy value generally had to be below 0.5 to 0.4 to consider the improvement significant.

Introduction

The diagnostic assessment of voice disorders and the evaluation of the effectiveness of voice therapy require accurate and reliable voice quality measurements. The final judgement on voice quality is generally made on the grounds of perceptual assessment. However, listeners may use different perceptive phenomenon when rating [1] and may attribute different severity levels to the points of the rating scale. In that light, parameters derived from the acoustic signal might offer an alternative to perceptual rating. Acoustic analysis has an advantage over perceptual evaluation of voice: it avoids the problems of variability, which are caused by differences among the listeners and variation of the internal standards within listeners. Whether acoustic parameters have clinical relevance is still under discussion. Some clinicians do not attribute any value to these measures. Others see the acoustic parameters as a source of additional information about the quality of the vocal fold vibrations of a patient. At the very least, acoustic parameters can provide information about the stability of the vibrations of the vocal folds. As irregular vibrations of the vocal folds are an indication of pathology in one or both vocal folds, a phoniatric examination will generally include a measure of regularity of the vocal fold vibrations. The parameters that are generally used characterize the perturbations of the fundamental frequency or the amplitude of the voice signal and, additionally, the relative amount of non-harmonic power.

When evaluating the effectiveness of voice therapy, it is preferable to use parameters with a low intra-subject variability. As a therapy does not always result in a normal voice, the pre- and post-therapy differences will be tested for significance in a group of patients. The less intra-subject variability, the more sensitive the test will be. If the difference between the pre- and post-therapy data of a group of patients proves to be significant, a next step may be to investigate the distribution of the successes among individual patients. It may be important to know, if only some patients were very successful or if all patients showed slight improvement. This information may also help identify factors that predict success. When evaluating an individual therapy that did not result in a normal voice, one wants to know whether the value of a parameter after therapy is really (that is, statistically significant) different from that before the therapy. The value of that change will depend on the measuring error of the parameter.

The error in acoustic perturbation measures will be based on several sources of variability. When computing perturbations characterized by the period-to-period differences, the imperfect detection of corresponding points in successive periods will introduce errors in the value of the acoustic parameters. Small deviations, such as those due to additional noise, may have effects on the results, unrelated to an abnormality in the vibration of the vocal folds. Therefore different analysis programs may yield different results for the same sample [2]. Another source of variability is inherent to the estimation of parameters of a random process from a sample with limited duration. Although the underlying parameters of the process are constant, the results for a sample of limited duration will fluctuate [3, 4]. However, the variation in the vibratory pattern within a subject from moment to moment may be the most important factor. It will yield different results for subsequent utterances and for utterances on different days. Many factors - such as intensity, fundamental frequency, stress, time of day, and mood - appear to influence the perturbation of the vibration of the vocal folds in subjects with a normal voice [5, 6, 7, 8]. In dysphonic patients, additional factors related to the pathology may contribute to the variation in the vibrational pattern.

Whereas the reliability of perceptual ratings has been amply studied, less is known about the amount of variability of acoustic parameters related to the stability of voice production. Several studies give the distribution of the values of acoustic parameters for a group of subjects without voice problems. The aim of these studies is to discriminate between patients and normal subjects [9, 10]. The discriminating power of acoustic parameters was tested in studies with dysarthric patients due to neurological problems [11, 12] and with dyphonic patients [13]. These studies provide no information about the intra-subject variability of the parameters when measured on different occasions. Some studies [14, 15, 16, 17, 7] do mention the intra-subject variability of acoustic parameters when measured on different occasions in a group of normal (generally young) subjects. Whether these results may be generalized to patients with abnormal voices is questionable, though. Variability may be expected to increase along with the degree of abnormality. This effect can be seen in the figures presented in the study by Linville and Korabic [18], who studied the jitter in a group of older women. Moreover, in the studies with normal subjects, the recordings were generally made under controlled experimental conditions. The data may therefore be too optimistic to apply to everyday clinical practice. For a patient, it may be difficult to produce a sustained vowel for longer than three seconds and do so at the same fundamental frequency and intensity in several trials. In the study by Kent et al. [19] the variability of acoustic parameters within a group of dysarthric patients was determined for successive sustained vowels at the beginning and end of a session. Variation on different days is not included in these variabilities. The results for patients with a neurological problem may differ from those for patients with a chronic voice disorder. Theoretically, several of the sources of variability can be controlled even in a clinical situation. However, in actual practice, many of these sources will be present. In a study on the effectiveness of voice therapy in actual clinical practice these sources should be included.

In an evaluation project of voice therapy for chronic dysphonic patients, multiple voice recordings were made on several days during a baseline period. These measurements were used to determine the variability of acoustic parameters in this patient group.

The first aim of this study was to estimate the variability of the acoustic parameters in the real world of clinical assessment. All kinds of variability that can occur in clinical practice are included in the estimated variability measures. The parameters were derived from sustained vowel productions of patients with a chronic voice disorder produced in succession and on several days. The second aim was to determine the degree of change in the value of the acoustic parameters that would be necessary in a clinical situation to conclude that there was a significant change after therapy.

Method

Subjects

This study used pre-treatment data on 82 patients participating in a research project on the efficacy of logopedic voice therapy. Only nearly-periodic voice signals were used. Two patients were excluded from a group of 84 patients with complete data, because of an insufficient number of voice samples with a nearly-periodic signal. All patients exhibited some kind of chronic dysphonia, either functional or organic resulting from a benign laryngeal pathology. Table 1 shows the frequency of the etiologic categories represented among the group as diagnosed by an ENT-specialist. The group of patients consisted of 34 men and 48 women ranging in age from 17 to 76 years. The average age for the female participants was 40 and for the male subjects 48 years of age.

Table 1. Distribution of patients by diagnostic categories.

Phoniatric diagnosis	N
Muscle tension dysphonia	14
Submucosal swelling	20
Vocal fold nodules	9
Vocal fold polyps	5
Unilateral vocal fold paralysis	10
Other: slight vocal fold abnormalities	15
Other: severe vocal fold abnormalities	9

Procedure
Over a period of six weeks, three recordings were made of a patient's voice. The first recording was made at the Department of Phoniatrics at the University Medical Center in Utrecht, as part of the voice assessment of the patient. The next two recordings were made at the office of a speech therapist before the actual therapy started. The time of day at which voices were recorded, was not standardized. Subjects were asked to produce three sustained vowels /a:/ of several seconds long. The patients were instructed to choose a comfortable pitch and loudness level.

Instrumentation
Voice samples were recorded on digital audio tape (recorder Sony DA-7 or Aiwa HD-S200) combined with a condenser microphone (Sennheisser K3 or Sony ECM-221) held at a mouth-to-microphone distance of approximately five centimeters and an angle of about 45^{o} by means of a harmonica holder. The speech samples were digitized at a sample frequency of 50.0 kHz using the Computerized Speech Lab Model 4300 (Kay Elemetrics Corporation). The first and last 200 ms of the sustained phonation were removed. The remaining signal was provided with a gradual start and stop (quarter of a sinus) over 50 ms. The duration of the sample varied according to the patient's possibilities, but contained at least 100 cycles.

The Multi-Dimensional Voice Program (Kay Elemetrics Corporation) was used to compute nine acoustic parameters for each sample. In order to ensure that only nearly-periodic signals were analyzed, for each sample the instantaneous frequency was displayed on the computer screen. If periods with a double (or more) duration or half (or less) the mean duration were seen, the sample was not used. In some cases the sample duration was reduced to exclude the abnormal period durations.

The perturbation of the pitch period (jitter) was evaluated by four measures derived from the period-to-period differences of the period duration. The percentage jitter (Jitt) is the mean of the absolute value of this difference relative to the mean period duration [20]. The other measures - Relative Averaged Perturbation (RAP), Period Perturbation Quotient (PPQ) and Smoothed Period Perturbation Quotient (SPPQ) - use the difference of a successive period durations to mean the duration of the surrounding cycles. These parameters were highly correlated with the parameter Jitt (R>0.94). Consequently, when analyzed in the same way, the results are very similar for all four jitter measures. Only the results for the parameters Jitt will be presented. The period-to-period variability of the peak-to-peak amplitude (shimmer) was evaluated in terms of three parameters. The percentage shimmer (Shim) is the mean of the absolute period-to-period difference in peak-to-peak amplitude divided by the mean amplitude. The two other parameters - Amplitude Perturbation Quotient and Smoothed Amplitude Perturbation Quotient, which are defined similar to the PPQ and SPPQ - were highly correlated with the parameter Shim (R>0.84). As the results for these two parameters did not substantially differ from the results for the parameter Shim, only the results for the latter will be reported here. For a general evaluation of the noise presence in the analyzed signal three parameters were studied. Three parameters were selected because they are not highly correlated among themselves. Noise-to-Harmonic Ratio (NHR) was determined as the average ratio of energy of the inharmonic components in the range 1500-4500 Hz to the harmonic components energy in the range 70-4500

Hz. The Voice Turbulence Index (VTI) is an average ratio of the spectral inharmonic high-frequency energy in the range 2800-5800 Hz to the spectral harmonic energy in the range 70-4500 Hz. VTI measures the relative energy level of high-frequency noise. The Normal Noise Energy (NNE) [21] represents the ratio of mean spectral level between harmonic peaks in short-period spectra to the total spectral level. The Voice Assessment program of Dr. Speech (Tiger Electronics Inc.) provided the NNE.

Statistical analysis

For perturbation parameters that estimate the underlying parameters of a random process (such as jitter) the variability is expected to be dependent upon the values itself. In those cases, a multiplicative model for the measuring error representing the variability is more adequate than a linear model. That means that, if μ_i is the true value for subject i and p_{ij} is a normally distributed error term representing the variability (mean = 1 and standard deviation σ_ε), a measurement X_{ij} is:

$$X_{ij} = p_{ij} \,.\, \mu_i \qquad (1)$$

The variance of X_{ij} within one subject will be approximately:

$$\mathrm{Var}[X_{ij}] = \mathrm{Var}[p_{ij}\,].\,\mu_i^2 \qquad (2)$$

When several measurements of the same subject are available, the value μ_i can be approximated by the mean (X_i) of these values. According to Eq. (1), the standard deviation (SD_i) computed within subjects will show a positive relation with the subject's mean value (X_i), and the error in a measurement is approximately proportional to the measured value. Therefore, the coefficient of variation (CV) is an adequate variable to describe the error. Using Eq. (2), the CV is found as follows:

$$CV = (\text{standard deviation / mean}) = \text{standard deviation of } p_{ij} = \sigma_\varepsilon$$

To estimate the value of σ_ε from a sample of data X_{ij}, a logarithmic transformation of Eq. (1) could be used. The result is a linear relation and with an analysis of variance for repeated measurements of subjects, the variance of $\ln(p_{ij})$ was found. When the errors are small, mathematically, this variance is roughly equal to the variance of p_{ij}, because the mean of p_{ij} is one. However, when sufficient values X_{ij} of each subject i are available, another procedure is possible. The values of p_{ij} for each individual measurement are approximated by X_{ij} / X_i. The variance of p_{ij} can then be estimated by the within-cell variance in an analysis of variance of (X_{ij} / X_i) with subjects as the single factor. Simulations showed that in this study the second procedure yielded more accurate results than a logarithmic transformation. This was due to the fact that the errors were relatively large and that nine values were available for the computation of the mean in each subject. In this study, the factor p_{ij} was split up into a factor representing the variability among successive trials (t_{ijk}) and a factor related to different days (d_{ik}). When the factor p_{ij} is equal to the product of d_{ik} and t_{ijk}, the variance of p_{ij} is approximately the sum of the variances of the factors d_{ik} and t_{ijk}, assuming that these factors are independent and their mean is one.

To test whether a multiplicative model was appropriate for a parameter, the dependency of the standard deviation on the mean value was determined for each acoustic parameter. Scatter plots were made of the standard deviation calculated over the three trials for each patient per day as a function of the total mean over nine trials for each patient. The plots were visually inspected for the dependency of the standard deviation on the mean and a regression analysis was performed. When a significant positive slope was found together with a non-significant intercept, the error was expressed as the coefficient of variation (CV). For the computation of the coefficient of variation the data were replaced by the values divided by the mean of the data for all nine samples (three trials on three days) of the subject. This mean value was considered to be the best approximation of the true value of the acoustic parameter for a particular voice. If no dependence was found between the standard deviation and the mean value, an absolute error was computed. The original data were used in the same procedure as the relative data.

The data on the repetitions within each patient were assumed to be normally distributed. Standard deviations among three repetitions on one day or among the mean values on three days were used to characterize these variations. Among the patients, extreme values for the standard deviation were found. Inclusion of these extreme values make the variance among data rise to an unrealistic level. Therefore, the pooled error variance was not computed directly by an analysis of variance. First, the standard deviations among the subjects were inspected for the presence of extreme values that were more than 1.5 times the interquartile range above the 75th percentile. Then, these extreme values were reduced to the maximum of the not extreme values. After that the pooled variance was calculated. The same procedure was used for the standard deviations computed for the mean values on the three days. Before applying this procedure, this standard deviation was corrected for the variation on one day that is included in the variation of the mean. The correction was performed in each subject by subtracting the pooled variance for one day divided by three (i.e. the number of repetitions on each day) from the variance among three days.

The Friedman test for repeated measures was used to test systematic differences between trial 1, 2 and 3. In this test, data from all three days was used. As the test was used to demonstrate no difference between the trials, the increase of the degrees of freedom was acceptable. A second Friedman test was performed for the mean values on day 1, 2 and 3. The Statistical Package for the Social Sciences (version 8.0.0) was used for the calculations (SPSS Inc.).

Results

Table 2 shows the distribution of parameter values among the patients in this study. Generally, the distributions were skewed toward abnormal values, though not all patients had abnormal values for all the parameters.

The duration of the analyzed period of the sustained vowels produced by the patients varied from 0.9 to 14.9 seconds. The median, the 25th and 75th percentile value were 3.5, 2.7, and 4.7 seconds, respectively. In about 25% of the patients, all nine vowel durations were shorter than 3.9 seconds.

Table 2. Median, 25th and 75th percentile of the distribution of the voice quality parameters of the patients averaged over all nine sustained vowels. The last column shows the normal threshold as used in the MDVP program. The value for the NNE comes from Kasuya et al. [21].

	Median	25th perc.	75th perc.	Norm
Jitt [%]	1.85	1.09	3.07	1.04
Shim [%]	5.72	4.67	8.07	3.81
NHR [-]	0.15	0.13	0.18	0.19
VTI [-]	0.07	0.048	0.09	0.06
NNE [dB]	-7.8	-11.9	-4.4	-11

Systematic differences

None of the parameters showed a clear significant difference between the values determined for the repetitions on one day (Friedman, $p>0.16$, except for Shim: $p=0.025$). For recordings on different days, for the parameters Shim and VTI a significant difference was found (Friedman, $p<0.001$). During the recordings on the first day, higher shimmer and VTI values were found than on the second and third day (medians for shimmer: 6.1%, 5.4% and 5.3%; for VTI: 0.071, 0.059 and 0.052). In the other parameters no trend or learning effect was found.

Type of error

The question of which kind of values - relative or absolute - best describe the errors was investigated by performing a linear regression of the standard deviations on the subjects' mean values. Scatter plots of these data showed a characteristic triangular spread of data points (example in Fig. 1A) for all parameters, except the NNE. The characteristic spread corresponds to the results of a model in which the values are the product of a random error term and a true patient value (Fig. 1B). The results of the linear regression are shown in table 3. Clearly, relative errors should be used for the parameters Jitt, Shim, and VTI. For those parameters, the slope is significant while the intercept is small and statistically not significant. For NHR and NNE the results were not directly clear. However, the computation of NNE values for cases with noise levels of the same order as the harmonic spectral power (values near zero dB), will be not reliable. An analysis that exclude values of -3 dB or higher yielded a non-significant slope (table 3). Therefore, absolute errors were computed for the NNE. In the scatter plot for the NHR a clear positive relation between the standard deviation and the mean value was present. However, the regression line crossed the x-axis at about 0.07 leading to a small but significant negative intercept. As the intercept was much smaller than the total variation of the standard deviation, relative errors were computed for the NHR as well.

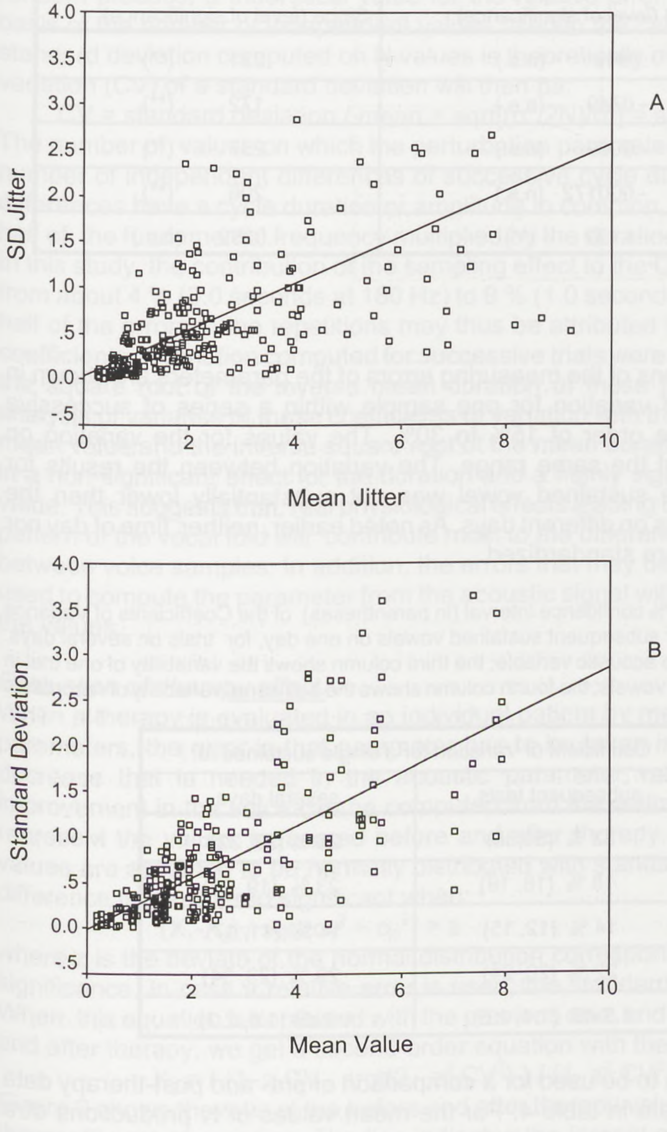

Figure 1.
A: Scatter plot of the standard deviation of the jitter (SD jitter) for three consecutive sustained vowels against their mean value (Mean jitter). The MDVP variable jitter percent (Jitt) is used.
B: Results of a simulation with the model $X_{ij} = p_{ij} \cdot \mu_i$. Scatter diagram of standard deviation computed three times for three values in 82 subjects plotted against the mean value of these nine values. p_{ij} was normally distributed with mean=1 and SD= 0.33. μ_i was distributed as exp(N[0.67, 0.70]) approximating the distribution of the jitter found in this study. A regression line is drawn through the data.

Table 3.. Results of the regression analysis of standard deviation for three subsequent repetitions of a sustained vowel as function of the mean value of the parameter.

Acoustic variable	Intercept (level of significance)	Slope (level of significance)
Jitt	.0949 (n.s.)	.231 (**)
Shim	-.0749 (n.s.)	.172 (**)
NHR #)	-.0009 (n.s.)	.237 (**)
VTI	-.000172 (n.s.)	.206 (**)
NNE	1.132 (**)	-.0252 (n.s.)

(n.s.) not significant, (**) $p<.001$
#) (NHR-0.07)

Variability.
The mean standard deviations of the measuring errors of the parameters are shown in table 4. The coefficients of variation for one sample within a series of successive sustained vowels are in the order of 15% to 30%. The values for the variation on different days vary in about the same range. The variation between the results for successive repetitions of a sustained vowel was not substantially lower than the variation found for recordings on different days. As noted earlier, neither time of day nor the recording conditions were standardized.

Table 4. Mean values and the 95% confidence interval (in parentheses) of the Coefficients of Variance (%) found among the patients for subsequent sustained vowels on one day, for trials on several days. The first two columns indicate the acoustic variable; the third column shows the variability of one trial in a series of subsequent sustained vowels; the fourth column shows the additional variability of recordings on several days.

	Acoustic variable	Coefficient of Variation for a single sustained /a: /	
		subsequent trials	several days
Jitter	Jitt	33 % (30,36)	28 % (23, 32)
Shimmer	Shim	18 % (16, 19)	23 % (19, 26)
Noise	NHR	14 % (12, 15)	14 % (11, 17)
	VTI	22 % (20, 24)	29 % (23, 34)
	NNE	1.5 dB (1.4, 1.6)	1.9 dB (1.6, 2.2)

The coefficients of variation to be used for a comparison of pre- and post-therapy data can be derived from the data in table 4. For the mean values of N productions of a sustained /a:/ produced both before and after the therapy, the coefficient of variation (CV_N) can be estimated as follows: $CV^2_N = CV^2_{day} + CV^2_{trials}/N$. This equation expresses the reduction of the CV by averaging over the results of N voice samples. As the CV for trials and days are of the same magnitude, averaging over subsequent trials is significant up to N of about 5. At that level, the term CV^2_{trials}/N contributes only 10% to the value of CV_N.

Part of the variability may be explained by the fact that each sustained vowel is the realization of a random process. If we consider jitter and shimmer as results of a random process, a theoretical value for the relative error can be determined on the basis of the number of independent values used in the calculation. The variance of a standard deviation computed on N values is theoretically $\sigma^2/2N$ [22]. The coefficient of variation (CV) of a standard deviation will then be:

$$CV = \text{standard deviation} / \text{mean} = \text{sqrt}[(\sigma^2/2N)/\sigma^2] = \text{sqrt}(1/2N)$$

The number of values on which the perturbation parameters are based is equal to the number of independent differences of successive cycle durations. As pairs of these differences have a cycle duration or amplitude in common, this number is equal to the half of the fundamental frequency multiplied by the duration of the sample in seconds. In this study, the contribution of the sampling effect to the CV can be estimated to vary from about 4 % (3.0 seconds at 180 Hz) to 9 % (1.0 seconds at 120 Hz). At most about half of the errors in the repetitions may thus be attributed to sampling. However, the coefficients of variation computed for successive trials were only weakly correlated with the square root of the inverse mean duration of these trials ($R=0.08$ to 0.21). An analysis of variance of these coefficients of variation with the patients as factor and the mean value and the inverse square root of the mean duration as co-variates, resulted in a non-significant effect for the duration and a highly significant effect for the mean value. This suggests that, real physiological effects leading to variation in the vibrational pattern of the vocal fold will contribute most to the differences in acoustic parameters between voice samples. In addition, the errors that may be attributed to the algorithm used to compute the parameter from the acoustic signal will contribute to the spread of the values.

Evaluation of therapy effect.

When a therapy is evaluated in an individual patient by means of one of the acoustic parameters, the error in that parameter has to be taken into account. The minimum decrease that is needed in the acoustic parameter value to make a significant improvement in the voice can be computed from the estimated errors. Let X_1 and X_2 represent the values measured before and after therapy. When the errors in these values are assumed to be normally distributed with standard deviations σ_1 and σ_2, a difference is considered significant when:

$$(X_1-X_2) / \text{sqrt}(\sigma_1^2 + \sigma_2^2) > z$$

where z is the deviate of the normal distribution corresponding with a chosen level of significance. In case a relative error is used, the standard deviation is $\sigma_x = CV.X$. When this equation is combined with the previous one, and assuming equal CV before and after therapy, we get a second-order equation with the following solution:

$$X_2 = (\,(1- z.CV\,.\,\text{srqt}(2 - z^2.CV^2)\,) / (1- z^2.CV^2)\,)\,.\,X_1\,.$$

Figure 2 shows the ratio of the before and after therapy values (X_2 / X_1) plotted against the coefficient of variation. The line indicates the largest ratio that corresponds to a significant ($p<0.025$, one-sided) improvement in vocal fold vibration. With an increasing value of the coefficient of variation, the ratio decreases. This indicates that, with an increasing value of the CV, the post-therapy value of a parameter had to decrease to a greater extent for a significant therapy effect. As stated above, the coefficient of variation to be used in a comparison between parameter values before and after therapy can be reduced by using the mean result of more than one sustained vowel.

Research is generally concerned with the comparison of groups. The variance computed within the groups will include the measurement error. Therefore, statistical tests comparing group means, takes this measurement error into account. If the variance among the subjects is mainly due to a measurement error, averaging over several samples of the same subject may lower the group's standard deviation. Consequently, tests on differences between groups will be more sensitive.

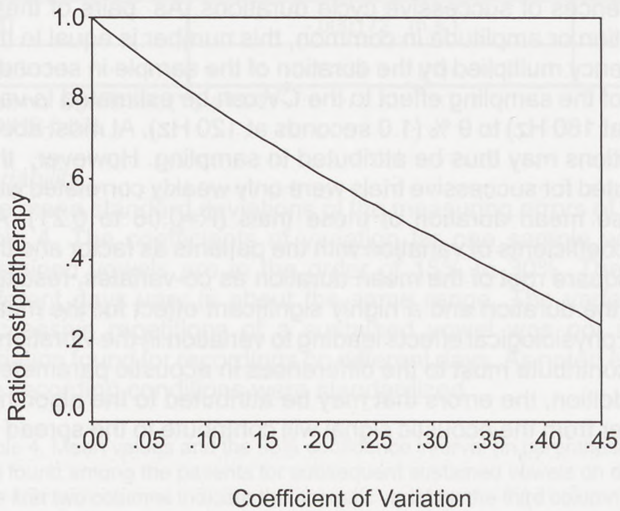

Figure 2. Theoretical relation between the decrease after therapy (ratio post/pre) at which a significant improvement at a level of p=0.025 (one-sided) is achieved and the relative error (Coefficient of Variation) in the acoustic parameter.

Discussion

This study determined the variability in acoustic parameters of sustained vowels produced at a comfortable loudness and pitch by dysphonic patients in a clinical setting. The acoustic parameters appeared to vary more as the magnitude of the parameter increased. The experimental results fit well with a multiplicative model of variations. The only exception was the NNE, most likely because this parameter has a logarithmic scale. The inaccuracy of the other acoustic parameters should therefore be described by a relative error, such as the coefficient of variation. This is in contrast to the confidence intervals found for perceptual ratings. Those confidence intervals are only weakly related to the rating itself [23, 24].

Distinctive inaccuracies in series of consecutive sustained vowels were found for the acoustic parameters. The coefficients of variation varied from 14% to 33% for a single sustained vowel /a:/ in a series of successive trials. The coefficients of variation for the results of different days (14% to 29%) were not substantially higher than those for successive trials. Higher values might be expected, because the conditions of the vocal folds might be less similar on different days than during successive trials. The sources of the variability will be diverse. Differences due to sampling, computational errors in determining cycle values, and physiological variations will contribute to the variation. The sampling factor due to the limited length of the samples can theoretically explain at most half of the value of the coefficient of variation. The errors for a stationary voice may be estimated from repeated analyses using different analysis periods from the same sustained vowel. Kent et al. [19] reported the standard deviations among two edited signals of about 3 seconds from the same phonation. For the parameters Jitt, Shim, and NHR they found values of 0.18, 0.37, and 0.01 . Assuming a patient population comparable to the patients in this study, these values would result in an error in the order of 6 to10% which include the effect of sampling and period detection. This error corresponds with about one-third to one-half of the coefficients of variation in this study. As no clear correlation with the sample duration was found, it is suggested that slight differences in the vibration mode of the vocal folds in successive phonations may be an important source for variation. The physical system of two vocal folds with abnormalities on or within these folds may yield a mechanically complex system with many vibration modes. Subtle differences at the start of phonation may result in different vibratory patterns during the rest of the phonation. Therefore, the coefficients of variation must be regarded as estimates of average perturbation parameters more than estimates of the parameters of a stationary process.

The values for the coefficients of variation are estimated under the assumption that the variation among the patients is random. However, the difference in variation of the perturbation measures among the patients may be related to the type of abnormality and to patient characteristics such as gender. Real differences in the variability across the patients cannot be excluded. Because the number of patients was small in relation to the number of diagnoses, the dependency of the variability on the diagnosis could not be analyzed. Furthermore, the possible associations of other factors with the variability was not investigated, because these factors can be related to the type of abnormality. For example, gender was significantly related to the diagnosis. In conclusion, the results shown in table 4 should be regarded as an estimate of the average variability of the acoustic parameters computed for the average sustained vowel among dysphonic patients. It should also be noted that the values for the CV's may be restricted to the computation of the parameters with MDVP.

When a parameter measured prior to therapy had to be compared with that measured after therapy, the precision of the comparison can be increased by averaging. The coefficients of variation for successive trials and different days are of about the same magnitude. Therefore, it makes little difference whether the results of sustained vowels on one day or the results of recordings on several days are averaged. In clinical practice, averaging over successive trials will be preferred. The results of this study suggest that at least five trials should be used. Using data from much more successive trials is less effective, because the variability between different days is not reduced. The effect of averaging was verified experimentally by Scherer et al. [25], who determined how many trials were needed to establish a stable mean value for jitter, shimmer, or NHR. The numbers Scherer at al. suggested are higher than the above-mentioned recommendations. The reason is that in the study by Scherer the reduction in variability was not limited by the day-to-day variation.

In the introduction it was stated that the variability of acoustic measures could be greater in dysphonic patients than in subjects with a normal voice. In the literature, several values can be found for the variability of acoustic perturbation parameters for young subjects with a normal voice. However, the procedures applied and the programs selected to compute the parameters are not the same throughout the literature. This makes it difficult to compare the results from the literature with those from this study. If the methodological differences are ignored, the coefficient of variation computed from results in the literature vary for jitter on different days from 26% to 62% for normal young women and from 15% to 46% for young normal man. Although the highest values in the literature may be attributed to the computational method [15], these values are not substantially lower than those found in this study for dysphonic patients. One reason for the high values of the coefficients of variation for normal voices may be that the numerator of the coefficient is small for normal voices. It should be realized that , when coefficients of variation are similar for normal voice and pathologic voices, the error itself will be much larger in patients as a result of the abnormal high mean values. Therefore, the use of a measure of relative error is essential.

The implications for the evaluation of voice therapy in an individual patient can be derived from Fig. 2. With the outcomes shown in table 4 and the number of sustained vowel productions over which is averaged, the value of the coefficient of variation can be computed. With this value, the ratio of post- to pre-therapy parameter value can be determined necessary for a significant improvement in the acoustic parameter. An improvement of the vibrational pattern of the vocal folds as far as detected by the acoustic parameter may then be assumed. For acoustic perturbation measures such as jitter, and shimmer, a decrease of at least 30 to 50 percent will generally be required for the therapy to have a real effect.

Conclusions

- The measurement errors in acoustic perturbation parameter should be expressed relative to the value of the parameter.
- For chronic dysphonic patients, the coefficients of variation of jitter, and shimmer for a sustained vowel as determined by MDVP are in the order of 20 % to 30 % for successive single trials as well as trials on different days.
- For reducing the error, averaging over successive trials can be restricted to 5 trials.
- For a significant result of an individual therapy, the jitter, and shimmer values had to decrease to less then about 50 or 70 percent of the pre-therapy value depending on the number of sustained vowels (1 or 5) that are used.

References

1. Kreiman, J., Gerratt, B.R. The perceptual structure of pathologic voice quality. *J. Acoust. Soc. Am.* 1996;100:1787-1795.

2. Rabinov, C.R., Kreiman, J., Gerratt, B.R., Bielamowicz, S. Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *J. Speech Hear. Res.* 1995;38:26-32
3. Linville, S.E., Korabic, E.W., Rosera, M. Intraproduction variability in jitter measures from elderly speakers. *J.Voice* 1990;4:45-51.
4. Jafari, M., Till, J.A., Truesdell, L.F., Law-Till, C.B. Time-shift , trial, and gender effects on vocal perturbation measures. *J.Voice* 1993;7:326-336.
5. Gelfer, M.P. Fundamental frequency, intensity, and vowel selection: Effects on measures of phonatory stability. *J. Speech Hear. Res.* 1995;38:1189-1198.
6. Huang, D.Z., Minifie, F.D., Kasuya, H., Lin, S.X. Measures of vocal function during changes in vocal effort level. *J. Voice* 1995;9:429-438.
7. Bough, I.D. J., Heuer, R.J., Sataloff, R.T., Hills, J.R., Cater, J.R. Intrasubject variability of objective voice measures. *J.Voice* 1996;10:166-174.
8. Ferrand, C.T. The effects of time constraints on phonatory stability in normally speaking adult women. *J. Voice* 1998;12:175-181.
9. Ferrand, C.T. Effects of practice with and without knowledge of results on jitter and shimmer levels in normally speaking women. *J. Voice* 1995;9:419-423.
10. Fernández Liesa, R., Damborenea Tajada, D., Rueda Gormedino, P., García y García, E., Leache Pueyo, J., Campos del Álamo, M.A., Llorente Arenas, E., Naya Gálvez, M.J. Análisis acústico de la voz normal en adultos no fumadores. (Acoustic analysis of the normal voice in non-smoker adults) *Acta Otorrinolaring. Esp.* 1999;50:134-141.
11. Zwirner, P., Murry, T., Woodson, G.E. Phonatory function of neurologically impaired patients. *J.Commun.Disord.* 1991;24:287-300.
12. Robert, D.,.Pouget J., Giovanni, A., Azulay, J., Triglia, J. Quantitative voice analysis in the assessment of bulbar involvement in Amyotrophic Lateral Sclerosis. *Acta Otolaryngol. (Stockh)* 1999;119:724-731.
13. Hirano, M. Objective evaluation of the human voice: Clinical aspects. *Folia Phoniatr.* 1989;41:89-114.
14. Linville, S.E. Intraspeaker variability in fundamental frequency stability: An age-related phenomenon? *J.Acoust.Soc.Am.* 1988;83:741-745.
15. Higgins, M.B., Saxman, J.H. A comparison of intrasubject variation across sessions of three vocal frequency perturbation indices. *J.Acoust.Soc.Am.* 1989;86:911-916.
16. Stone, R.E., Rainey, C.L. Intra- and intersubject variability in acoustic measures of normal voice. *J. Voice* 1991;5:189-196.
17. Dwire, A., McCauley, R. Repeated measures of vocal fundamental frequency perturbation obtained using the Visi-Pitch. *J.Voice* 1995;9:156-162.
18. Linville, S.E., Korabic, E.W. Fundamental frequency stability characteristics of elderly women's voices. *J.Acoust.Soc.Am.* 1987; 81: 1196-1199.
19. Kent, R.D.,Vorperian H.K., Duffy, J.R. Reliability of the Multi-Dimensional Voice Program for the analysis of voice samples of subjects with dysarthria. *Am.J.Speech-Language Path.* 1999;8:129-136.
20. Titze, I.R. *Workshop on Acoustic Voice Analysis.* National Center for Voice and Speech, 1995:16.

21. Kasuya, H., Ogawa, S., Mashima, K., Ebihara, S. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *J. Acoust. Soc. Am.* 1986;80:1329-1334.
22. Kendall, M.G., Stuart, A. *The advanced theory of statistics.* London & High Wycombe: C. Griffin & Comp, 1969:234.
23. Krom de, G. Consistency and reliability of voice quality ratings for different types of speech fragments. *J. Speech Hear. Res.* 1994;37:985-1000.
24. Kreiman, J., Gerratt, B.R., Kempster, G.B., Erman, A., Berke, G.S. (1993) Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J. Speech Hear. Res.* 1993;36:21-40.
25. Scherer, R.C., Vail, V.J., Guo, C.G. (1995) Required number of tokens to determine representative voice perturbation values. *J. Speech Hear. Res.* 1995;38:1260-1269.

Chapter 3

Effects of voice therapy as objectively evaluated by digitized laryngeal stroboscopic imaging

R. Speyer, G. Wieneke, E. Ghazi Hosseini,
P. A. Kempen, W. Kersing, P.H. Dejonckere
Annals of Otology, Rhinology & Laryngology
2002; 111: 902-909

Effects of voice therapy as objectively evaluated by digitized laryngeal stroboscopic imaging

Abstract

Objective measurements derived from digitized laryngeal stroboscopic images were used to demonstrate changes in vocal fold vibration and in the size of benign lesions after three months of voice therapy. Forty chronically dysphonic patients were studied. Using a rigid stroboscope, pre treatment and post treatment recordings were made of the vocal folds at rest and under stroboscopic light during phonation. From each recording, images of the positions at rest, during vibration at maximal opening and at maximal closure were digitized. The surface area of any lesions and of the glottal gap were independently measured in the digitized images by two experienced laryngologists. Referential distances were determined in order to compensate for discrepancies in magnification in the various recordings. After three months of voice therapy, significant improvement in lesion size and degree of maximal closure during vibration could be demonstrated in about 50% of the patients. The degree of maximal opening did not prove to be a significant parameter.

Introduction

The expectation that the therapeutic effects of a treatment should be underpinned by scientific research is increasingly prevalent. A precondition for such a study is the availability of relevant outcome variables. In the case of voice therapy, different sets of treatment outcome parameters provide us with diverse kinds of information concerning treatment effects. Common approaches in research on voice dysfunction are the perceptual evaluation of voice quality and acoustic analysis. Although perceptual measurements of breathiness and roughness have gained widespread acceptance as standard parameters in voice research, they remain essentially subjective in nature.

Laryngostroboscopy, however, enables the laryngologist to focus directly on the source of sound production: the vocal folds, and thus facilitate objective measurement. Videostroboscopy can be used as a tool for visual-perceptual evaluation of the glottal waveform and its abnormalities. For instance, a reduction in the dimensions of the glottal gap when the vocal folds are maximally closed may be observed as a result of voice therapy as well as of laryngoplastic phonosurgery (see, among others: D'Antonio et al[1]). Several protocols for evaluating data obtained using videostroboscopy have been developed (e.g. Hirano et al[2], Dejonckere et al[3], Poburka[4]). Recent advances in the technique of digital processing of laryngeal images have enabled the development of methods for deriving objective measurements from such endoscopic examinations (e.g. Colton et al[5], Sercarz et al[6], Johnson et al[7], Omori et al[89], Gonçalves & Leonard[10], Hanson et al[11], Jeannon et al[12], Björck & Hertegård[13]). Quantitative measurements derived from videolaryngostroboscopic images of vocal fold vibration in normal speakers of both sexes were published by Woo[14], and his reference data can be used to identify dysphonic patients as opposed to normal subjects. Hitherto, a handful of studies have addressed the evaluation of the effects of voice therapy using videolaryngostroboscopy (Kotby et al[15], Verdolini et al[16], Omori et al[8], Bassiouny[17]). Omori et al[8] have adopted new methods to measure dimensions in digitized laryngeal stroboscopic images.

The purpose of this study is to examine whether objective measurements derived from digitized laryngeal stroboscopic images can be used to detect significant changes in closure during vocal fold vibration and/or changes in the dimensions of benign lesions which may have occurred after three months of voice therapy. A majority of these benign vocal fold lesions is supposed to be related to voice abuse and/or misuse. The present study is part of a research project on the overall effects of logopedic voice therapy for patients as found in the daily clinical practice. As a consequence, this study includes patients with a variety of indications for voice therapy and various approaches in voice therapy. This is in contrast to most studies in the literature that concentrate on a specific patient population or therapy. As interaction between therapy and diagnosis is to be expected, the voice therapies were adapted to the diagnosis. The main goal was to investigate the efficacy of voice therapy in general. For ethical reasons, it was not possible in our experimental design to use a control group who did not receive any voice therapy. Therefore, only chronically dysphonic patients were selected to participate in our study. Because of the chronic nature of the dysphonic disorders, any change found in digitized laryngeal imaging could be considered to be the result of voice therapy.

Methods

Subjects

The study reported in this paper is part of a larger study on the effects of voice therapy which employs a multidimensional assessment approach. Patients participating in this study were selected according to the following criteria: all patients had to suffer from chronic dysphonia, the onset of the dysphonic problems had to have occurred at least four months before visiting an O.R.L.-specialist/phoniatrician at the Phoniatric Department of the University Hospital Utrecht, and the exact medical diagnosis had to have been demonstrated by comprehensive phoniatric investigation. Those under eighteen years of age were excluded from participating in this study because of possible problems of voice maturation or mutation. Furthermore, coexisting speech or language disorders, as well as malignant or pre-malignant lesions, hormonal voice disorders, laryngeal papillomas, gastroesophageal reflux, substitution of voice after laryngectomy, spasmodic dysphonia or psychogenic dysphonia, were also exclusion criteria. The aim was to avoid, as far as possible, parameters other than voice therapy affecting the outcome. If, according to the phoniatrician, voice therapy was indicated, patients were referred to speech therapists in their residential area. Indications included trial therapy and pre-surgical voice therapy, whereby patients received voice therapy adjusted to their specific phoniatric diagnosis for at least three months, with a frequency of thirty minutes twice a week or sixty minutes once a week. Seventy-one subjects who satisfied the inclusion criteria without triggering the exclusion criteria completed three months of voice therapy.

A videostroboscopic recording was made before and after these three months, and out of the 71 patients, both pre therapy and post therapy stroboscopic recordings were suitable for further computerized analysis in 40 cases. Of this group 36 subjects presented with pathological vocal fold closure and 25 subjects with a benign vocal fold lesion. With three exceptions, all of the patients with lesions were categorized by a phoniatrician as also presenting with glottal insufficiency. The main reason for not including a patient's data in the analysis was the inadequate quality of the video images concerned. The poor quality was either caused by low light levels, involved images that were out of focus, or resulted from incomplete glottic images due to an overhanging epiglottis, prominent false cords or collapsed arytenoids (Jeannon et al[12]). Furthermore, because of the variable distance between the laryngoscope and the vocal folds, surface area measurements (in pixels) taken from vocal fold images require calibration to allow for variations in the magnitude of magnification. In order to place stroboscopic pre treatment and post treatment recordings on the same scale, a referential distance was identified and measured. (Figures 1, 2 and 3.)

The group of 40 patients consisted of 27 women and 13 men, with an age range of 18 to 76 years. The average age of the female subjects was 38 and of the men 52 years. The subjects represent a sample reflecting the diversity of laryngeal pathology in chronic dysphonia as found in the clinical setting (Table 1) and are used to evaluate overall effects of voice therapy in such clinical practice. The chronic character of the dysphonia was confirmed by acoustic measurements on speech samples collected before therapy onset (baseline period) on three different days within six weeks. No significant systematic differences between these baseline measurements could be demonstrated.

Figure1

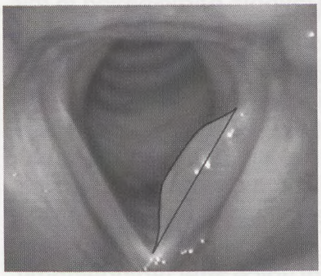

Figure 2

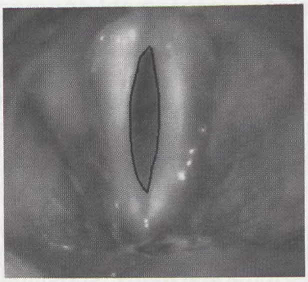

Figure 3

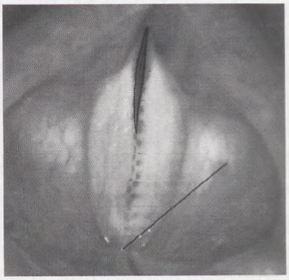

Figure 1: Measurement of the surface area of the lesion.
Figure 2: Measurement of the surface area of the maximal opening of the vocal cords during phonation.
Figure 3: Measurement of the surface area of the maximal closing of the vocal cords during phonation and corresponding referential distance (distance between telangiectasia and ventral commissure)

Table 1. Distribution of patients by diagnostic categories.

Phoniatric diagnosis	N
Muscle tension dysphonia	5
Submucosal swelling	7
Vocal fold edema (Reinke edema)	2
Vocal fold nodules	7
Vocal fold polyp	3
Unilateral vocal fold paralysis	4
Other: slight vocal fold abnormalities*	9
Other: severe vocal fold abnormalities**	3

* including chronic laryngitis, slight sulcus, contact ulcus, small (pseudo)cyst, etc.
** including radiotherapy, scar, etc.

Procedure

A rigid endoscope was employed to achieve an image quality superior to that realized by flexible techniques of endoscopy. If necessary a topical anesthetic (Lidocaine 4%) was applied. Recordings on video tape (U-matic videocassette recorder Sony VO-5630) were made of vocal fold vibration during repeated stable phonation of a sustained vowel /a:/ or /i:/ at comfortable pitch and loudness using normal as well as stroboscopic light (Stroboscope Kay RLS 9100). In addition, images of the vocal folds in rest position were recorded.

A pentium II IBM MS-DOS compatible computer (internal memory 128 Mb) combined with a PCI video-adapter (ATI, RAM memory 8 Mb) served as workstation. Selections of the video recordings were digitized and stored using a frame grabber (FastCap version 2.2.0 by FastMultimedia, Inc). By means of the software program VideoCapture (version 5.02 by Ulead Systems, Inc.) video frames were selected for further analysis. From each pre treatment and post treatment recording three images were captured. One image had been made during the rest position of the vocal folds, providing a clear image of any lesion present, while the other two images were taken during vibration of the folds under stroboscopic light. The second image was chosen at the moment of the maximal opening of the vocal cords, and the third at the moment of maximal closure. The moment of maximal opening during vibration was considered to be a global measure for the elasticity of the vocal folds, while the moment of maximal closure could reveal glottal insufficiency, which is presumed to be correlated to breathiness. Since quality was the criterion in selecting the images of maximal opening or closure from the whole recording, they need not have belonged to a single vibratory cycle. Three images were obtained from the pre treatment, and three from the post treatment recordings, so that six images per patient were available for quantitative measurement.

With the aid of software program Scion Image, distances and surface areas were measured in the digitized images. Scion Image is an IBM compatible version of a software program developed for Macintosh systems: NHI-Image by W. Rasband (National Institutes of Health, Bethesda). Measurements were made with a screen resolution of 1024 horizontal by 768 vertical pixels. Scion Image provides the ability to select the red, blue or green component of an image separately and show them in shades of grey. In practice, green images shown in shades of grey appeared to produce the sharpest pictures of the vocal folds. In some cases enhancement techniques were used to optimize the contrast and brightness of the images.

Contours of surfaces and distances of anatomical reference features were traced manually using a mouse. Areas within the traced contours were computed by the program. During the procedure, each patient's pre treatment and corresponding post treatment images were shown on a monitor simultaneously without any indication as to which was which, and no soundtrack was provided. The area of the lesion was measured in the first image, the area of the maximal visible glottal gap during phonation was measured in the second image and the area of the smallest gap in the third image, all measurements being carried out independently by two experienced phoniatricians. A referential distance that was visible in the pre treatment as well as in the post treatment image was chosen and measured. Different kinds of referential distances were used such as the distance between two blood vessels, the vocal fold length, the width of an arytenoid or of the epiglottis. The error of intraobserver measurement was determined by performing independently the complete measurements twice, with an interval of at least six weeks. The intraobserver variance for the difference between pre and (corrected) post-treatment measurements was computed by means of an analysis of variance (Shrout & Fleiss[18]).

A correction was applied to the post treatment values of the different parameters, to allow for discrepancies in magnification in pre and post treatment video recordings (Appendix). Surface areas calculated from the post treatment recordings were multiplied by a correction factor to correspond to the pre treatment recording data. This correction factor was the squared ratio of the referential distance as measured in the pre treatment recording to the same distance as measured in the post treatment recording. Squared ratios were used because the videostroboscopic variables in this study are expressed as surface areas rather than linear distances. In order to determine the effect of voice therapy, pre treatment values were subtracted from adjusted surface areas after therapy, resulting in quantified treatment effect data.

Results

A Wilcoxon matched-pairs signed-ranks test was used as an overall test for evaluating differences between pre and post treatment data. The results revealed that the size of the laryngeal lesion as well as the glottal insufficiency or area of the glottal gap in maximal closure position of the vocal folds during vibration, did show significant improvements ($p = .01$ and $p=.001$) . The 25th, the median and the 75th percentiles for change in lesion are -52%, -36% and -3%, and for change in maximal closure these values are -88%, -46% and -17%. The change in maximal opening position of the folds was not significant.

The treatment effects after three months of voice therapy in relation to the pre treatment data are specified for each patient in figures 4 to 6. If no lesion or no air leakage in the closed position of the vibratory cycle was detected in the pre treatment examination, then these data are left out of the corresponding figures and analyses. The horizontal dotted lines in these figures represent the 95% confidence interval of the intraobserver variability of differences as assessed by means of an analysis of variance of the data of the repeated assessment. If no relevant differences are apparent between pre treatment and post treatment videostroboscopic recordings, the corresponding data point will be plotted between these lines. If, after three months of therapy, a lesion had completely disappeared, the value data point would be plotted on the sloping solid line in figure 4. In case of the maximal closure, this line represents a situation in which there is no air leakage occurring during closure of the folds after therapy. Therefore, in the case of the maximal possible improvement being achieved after therapy, the value concerned will be plotted near the solid lines.

The figures show that, in the case of the larger lesions, a decrease in the dimensions of the abnormality was generally achieved. Only a small number of patients had completely recovered after three months of therapy, and in some subjects the lesion did not change at all significantly as result of therapy. Where small lesions were concerned, it was especially difficult to verify improvement in view of the margins of measurement error. In some 50% of the patients a significant improvement was found. A regression analysis for the data below the lower limit of the 95% confidence interval of figure 4 showed that the size of these lesions had been reduced on the average by about 46%. A complete recovery, that is the state indicated by the post treatment value within the confidence interval around the absence of lesion, was found in 11% of the patients who had presented with a lesion before therapy. Among the diagnostic categories, no clear differences in the pattern of recovery could be demonstrated in this study. A slight enlargement of the lesion was observed in two patients with vocal fold nodules and two patients with submucosal swelling.

The data on maximal opening of the vocal folds (figure 5) showed that an important group of subjects underwent an enlargement of this parameter but a similar number of patients showed a reduction. This parameter did not systematically change as result of three months of voice therapy.

Figure 6 shows that in the case of larger deviations that imply greater insufficiency of glottal closure, a decrease in the size of the abnormality was generally attained. As in those subjects who had presented with lesions, in some 50% of the patients a significant improvement was demonstrated. Regression analysis for the data below the lower limit of the 95% confidence interval obtaining in figure 6 showed that the minimal glottal gap during vibration reduced on the average by about 62%. A complete closure within the confidence interval was found in 20% of the patients with insufficient glottal closure before therapy. The diverse diagnostic categories showed similar patterns of recovery except in the cases of four patients presenting with vocal fold paralysis for whom the results of treatment were inconsistent. Deterioration of the maximal closing position of the folds appeared unambiguously in two subjects diagnosed as suffering from unilateral vocal fold paralysis.

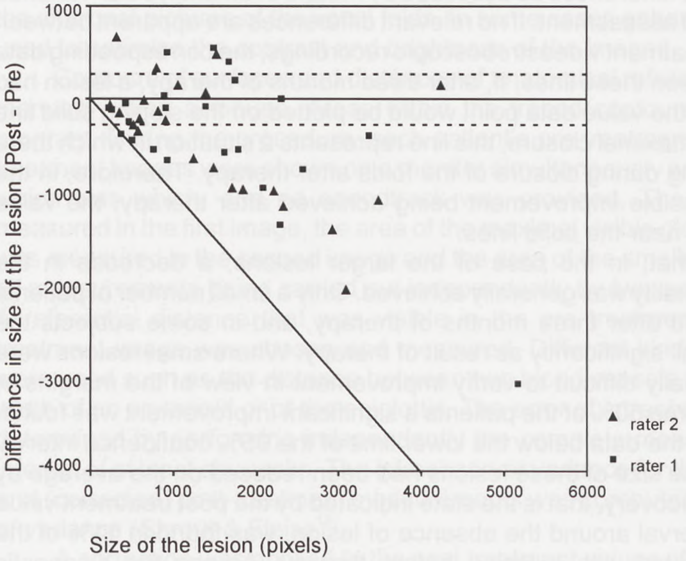

Figure 4:
Change in surface area of lesion after three months of voice therapy (N = 25) plotted against the pre treatment surface area as assessed by two phoniatricians. The horizontal dotted lines define the 95% confidence interval of the intraobserver variability (plus or minus 270 pixels). The solid line represents the position of values for patients in whom the lesion had completely disappeared. The sloping dotted line is the regression line through the data below the lower limit of the confidence interval.

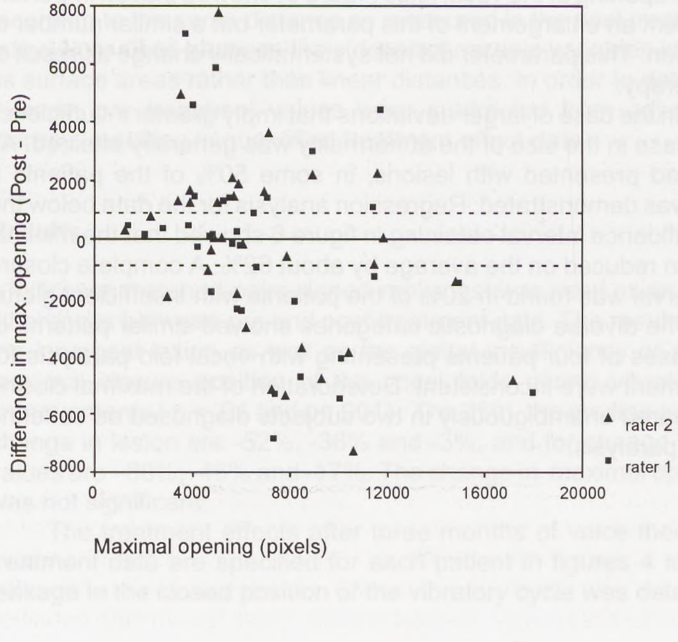

Figure 5:
Change in maximal opening of the vocal cords during phonation after three months of voice therapy (N = 35) plotted against the pre treatment maximal opening as assessed by two phoniatricians. The horizontal dotted lines define the 95% confidence interval of the intraobserver variability (plus or minus 888 pixels).

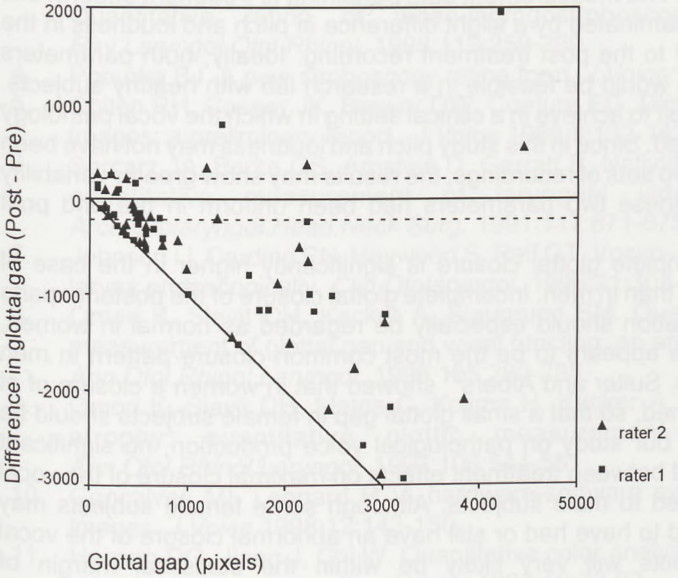

Figure 6:
Change in maximal closure of the vocal cords during phonation after three months of voice therapy (N = 36) plotted against the pre treatment glottal gap as assessed by two phoniatricians. The horizontal dotted lines define the 95% confidence interval of the intraobserver variability (plus or minus 212 pixels). The solid line represents the position of values for patients in whom the glottal insufficiency had completely disappeared. The sloping dotted line is the regression line through the data below the lower limit of the confidence interval.

Discussion

The following conclusions can be drawn from the results of this study. In the first place, objective measurements derived from digitized laryngeal stroboscopic images can be used to investigate the effects of voice therapy. It is evidently possible to use such measurements to evaluate whether voice therapy has positive effects on the physiology of the vocal folds or does not. Measurements taken from laryngeal stroboscopic recordings, before and after three months of voice therapy, can provide valuable information concerning therapeutic effects and the usefulness of possible continuation of voice therapy.

Secondly, valuable parameters in the stroboscopic imaging appeared to be the size of a laryngeal lesion as well as the glottal insufficiency or area of the glottal gap during maximal closure of the vocal folds during vibration. No useful information was yielded by measuring the maximal opening position of the folds. Thirdly, it was demonstrated that in this group of chronically dysphonic patients with diverse pathologies, a significant improvement after three months of voice therapy was realized in about 50% of the patients.

In this study, all videostroboscopic recordings made of vocal fold vibration during stable phonation were at comfortable pitch and loudness. Fundamental frequency as well as loudness are factors which influence vocal fold vibration. In normal phonation, an increase in sound intensity is positively correlated with glottal closure, whereas in the case of female subjects an increase in fundamental frequency, whilst avoiding transition among registers, can be related to a decrease in closure (Sulter et al[19]). However, other authors did not find the degree of incomplete closure to be affected by changes in pitch (Södersten & Lindestad[20]). The measurement data pertaining to treatment effect in this study may have been contaminated by a slight difference in pitch and loudness in the pre treatment as opposed to the post treatment recording. Ideally, both parameters should be controlled. This would be feasible in a research lab with healthy subjects. However, it would be difficult to achieve in a clinical setting in which the vocal pathology of patients is being assessed. Since in this study pitch and loudness may not have been entirely consistent in the two sets of recordings, the results may show greater variability than they would have if these two parameters had been uniform in pre and post treatment recordings.

The degree of incomplete glottal closure is significantly higher in the case of normally speaking women than in men. Incomplete glottal closure of the posterior parts of the glottis during phonation should especially be regarded as normal in women, whereas complete closure appears to be the most common closure pattern in men (Södersten & Lindestad[20]). Sulter and Albers[21] showed that in women a closure of at least 90% should be attained, so that a small glottal gap in female subjects should be considered as normal. In our study on pathological voice production, no significant differences could be found between treatment effects on maximal closure of the vocal folds of female as opposed to male subjects. Although some female subjects may erroneously be considered to have had or still have an abnormal closure of the vocal folds, the observed effects will very likely be within the statistical margin of measurement error.

Conclusion

The results of this study show that objective measurements in digitized laryngeal stroboscopic images can be used to investigate the effects of voice therapy on the physiology of the vocal folds. In about 50% of the dysphonic patients in this study, a significant improvement after three months of voice therapy could be demonstrated. The size of the laryngeal lesion as well as the glottal insufficiency or area of the glottal gap in maximal closure position of the vocal folds during vibration, proved to be valuable parameters. In contrast, measuring the maximal opening position of the folds failed to yield useful data.

Acknowledgments
We are indebted to B. Altorffer for technical assistance in the digital processing of laryngeal videostroboscopy.

References

1. D'Antonio LL, Wigley TL, Zimmerman GJ. Quantitative measures of laryngeal function following Teflon injection or thyroplasty type I. *Laryngoscope* 1995;105:256-262.

2. Hirano M, Bless DM. *Videostroboscopic examination of the larynx.* London: Whurr Publishers, 1993.
3. Dejonckere PH, Crevier L, Elbaz E, Marraco M, Millet B, Remacle M, Woisard V. Quantitative rating of video-laryngostroboscopy: a reliability study. *Rev.Laryngol.Otol.Rhinol.* 1998;119:259-260.
4. Poburka BJ. A new stroboscopy rating form. *J.Voice* 1999;13:403-413.
5. Colton RH, Casper JK, Brewer DW, Conture EG. Digital processing of laryngeal images: a preliminary report. *J.Voice* 1989;3:132-142.
6. Sercarz JA, Berke GS, Arnstein D, Gerratt B, Natividad M. A new technique for quantitative measurement of laryngeal videostroboscopic images. *Arch.Otolaryngol.Head.Neck Surg.* 1991;117:871-875.
7. Johnson IJ, Carding PN, Heywood S, Relf GT. Vocim--a system for measuring the larynx endoscopically. *Clin.Otolaryngol.* 1996;21:308-311.
8. Omori K, Slavit DH, Kacker A, Blaugrund SM. Quantitative videostroboscopic measurement of glottal gap and vocal function: an analysis of thyroplasty type I. *Ann.Otol.Rhinol.Laryngol.* 1996;105:280-285.
9. Omori K, Slavit DH, Matos C, Kojima H, Kacker A, Blaugrund SM. Vocal fold atrophy: quantitative glottic measurement and vocal function. *Ann.Otol.Rhinol.Laryngol.* 1997;106:544-551.
10. Goncalves MI, Leonard R. A hardware-software system for analysis of video images. *J.Voice* 1998;12:143-150.
11. Hanson DG, Jiang J, Chi W. Quantitative color analysis of laryngeal erythema in chronic posterior laryngitis. *J.Voice* 1998;12:78-83.
12. Jeannon J-P, Carding PN, Wilson JA. Vocim analysis of laryngeal images: is breathiness related to the glottic area? *Clin.Otolaryngol.* 1998;23:351-353.
13. Björck G, Hertegård S. Reliability of computerized measurements of glottal insufficiency. *Log.Phon.Vocol.* 1999;24:127-131.
14. Woo P. Quantification of videostrobolaryngoscopic findings--measurements of the normal glottal cycle. *Laryngoscope* 1996;106:1-27.
15. Kotby MN, El-Sady SR, Basiouny SE, Abou-Rass YA, Hegazi MA. Efficacy of the accent method of voice therapy. *J.Voice* 1991;5:316-320.
16. Verdolini MK, Burke MK, Lessac A, Glaze L, Caldwell E. Preliminary study of two methods of treatment for laryngeal nodules. *J.Voice* 1995;9:74-85.
17. Bassiouny S. Efficacy of the accent method of voice therapy. *Folia Phoniatr.Logop.* 1998;50:146-164.
18. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979;86:420-428.
19. Sulter AM, Schutte HK, Miller DG. Standardized laryngeal videostroboscopic rating: differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency, and age. *J.Voice* 1996;10:175-189.
20. Södersten M, Lindestad PA. Glottal closure and perceived breathiness during phonation in normally speaking subjects. *J.Speech.Hear.Res.* 1990;33:601-611.
21. Sulter AM, Albers FW. The effects of frequency and intensity level on glottal closure in normal subjects. *Clin.Otolaryngol.* 1996;21:324-327.

Appendix:
Image calibration for magnification differences of pre and post treatment video recordings

In this study the following correction procedure is used.
If r_1 = referential distance in pre treatment image
r_2 = referential distance in post treatment image
then the correction factor (c) for discrepancies in magnification in pre and post treatment video recording is defined as:

$$c = (r_1 / r_2)^2$$

The post treatment surface area is multiplied by the correction factor to correspond to pre treatment data.
If a_1 = surface area in pre treatment image
a_2 = surface area in post treatment image
then the adjusted surface area (a_2') is

$$a_2' = c \cdot a_2$$

The pre treatment data are subtracted from the adjusted surface area in order to determine the effect of voice therapy. The quantified treatment effect value (e) is

$$e = a_2' - a_1$$

Chapter 4

Effects of voice therapy on the voice range profiles of dysphonic patients

R. Speyer, G.H. Wieneke, I. Wijck-Warnaar P.H. Dejonckere
In press (Journal of Voice)

Effects of voice therapy on the voice range profiles of dysphonic patients

Abstract

In a group of chronically dysphonic patients, a voice range profile, or phonetogram, was recorded before and after receiving voice therapy and again three months later. The voice range profiles took a wide variety of shapes. Therefore, only measures that did not depend on a smooth contour could be used to describe changes before and after therapy. The main effect of voice therapy was an enlargement on the side of low frequency and low intensity.

Introduction

The voice range profile (VRP), or phonetogram, describes the laryngeal possibilities with respect to fundamental frequency and sound intensity[1,2]. The maximal intensity range of sound production as a function of the fundamental frequency in the range at which the vocal folds vibrate is visualized on a two-dimensional surface. Generally, the VRP is determined by requesting the subject to generate sound by producing a sustained vowel at both maximal and minimal intensity for all possible fundamental frequencies. A closed contour can be constructed by connecting the points of maximal and minimal intensity. The use of a VRP is usually associated with the assessment of the singing voice, but it might be useful to evaluate the effects of voice therapy as well. For example, a larger surface area of the voice range profile area may indicate improvement in voice capacities and therefore be considered a positive effect of therapy.

In the literature, significant differences have been found between healthy and dysphonic subjects. Some authors provide averaged data on a variety of phoniatric disease groups[3,4] pathologies such as vocal nodules[5], superior laryngeal nerve paresis or paralysis[6], or non-organic dysphonia[7]. Airainer & Klingholz[8] showed how the computer-aided evaluation of voice range profiles can help distinguish types of functional dysphonia. These data suggest that the VRP can indeed be useful to detect improvement due to voice therapy.

In order to compare voice range profiles before and after voice therapy, the recording must be reproducible. To that end, several studies have documented the variability or stability of voice range profiles in healthy subjects[9,10,11,12,13]. Sihvo et al.[13] found that the standard deviations of intrasubject sound level variations within a series of consecutive phonations were 3 dBA in the loudest phonations and between 3 and 5 dBA in the softest ones, depending on pitch. When phonating at the same pitch ten times in succession prior to the next target tone, the intrasubject sound level variation was 2 dBA in soft and 1 dBA in loud phonation. These last findings are within the 2-3 dB difference that can be expected in successive trials, according to Coleman[14]. The total fundamental frequency range may vary as much as plus or minus 1- 2 semitones in retesting over a period of maximally one year[9,10,11]. Other studies have investigated sources of variation such as methodological variables[15,16] or spectrum factors[17]. Very few studies have considered the variability in the voice range profiles of patients with voice problems, however. One of these, by Kotby et al. [18], concerned fundamental frequency stability in patients with functional dysphonia. Another, by Gramming et al.[19], included a subgroup of patients with non-organic dysphonia in an investigation of short-term voice SPL variations. Furthermore, data on test-retest mean differences in the frequency range and the upper and lower contours of the VRP, as reported by Behrman et al.[20] for a small group of dysphonic patients undergoing laryngeal surgery, are in conformity with data on variability in healthy subjects. From these data on variability the magnitude of the changes that are necessary in order to detect improvement in a VRP can be derived.

The VRP was used to describe the effect of diverse kinds of treatments. For example, the results of laryngeal surgery were evaluated by Behrman et al.[20], Ikeda et al.[21], and Uloza & Šiupšinskiene[22], the results of radiotherapy on early glottic cancer were evaluated by Verdonck-de Leeuw et al.[23], the outcome of treatment with medication was evaluated by Pedersen[24], and the effects of training on the singing voice were studied by Klingholz[25] and Sulter et al.[26] Only in some cases did these papers describe improvements in the VRP, and the data were generally not statistically tested.

In contrast to the therapeutic procedures covered in the above-mentioned studies, phonetography has rarely been used to evaluate the effects of logopedic voice therapy in dysphonic subjects. Schulz-Coulon[27] gave some clinical examples of voice range profiles measured before and after voice therapy. Other reports on single cases - describing just one subject - are more common (see for additional references: Heylen[5]). Few papers present thorough research on the effects of voice therapy in a group of chronic dysphonic patients. One of these studies, by Hirano[3], was concerned with changes after therapy in adult voice patients. In that study, phonetography formed part of a larger multidimensional assessment protocol covering diverse therapies. Although changes in the voice range profile were measured on several parameters, a subgroup of patients with vocal fold nodules showed no significant changes following voice therapy[3,28]. In another study, Åkerlund [29] looked for changes in non-organic dysphonic patients; significant changes were evident in averaged phonetogram contours per gender, both before and after voice treatment. Female subjects achieved higher sound levels after therapy, as reflected by the elevation of the upper phonetogram contour. Male subjects displayed a lowering of the minimal intensity contour. This study of the clinical usefulness of the VRP was extended to include the assessment of voice therapy in patients with a more diverse pathology as compared to the earlier studies described in the literature.

An intriguing issue is how to compare voice range profiles with each other. Voice range profiles are rather difficult to use when comparing pre- and post-treatment data or different groups of subjects (normal versus pathological voicing). The difficulty lies in the two-dimensional data structure: frequency by intensity. In the absence of parameters describing the voice range profile as a whole, it is difficult to compare one voice range profile with another as well as to determine standard values for voice range profiles[30]. To overcome this obstacle, various authors have proposed methods for recording and analyzing voice range profiles[31,32,30]. A more extensive survey of the literature on voice range profiles is found in Heylen[5].

Parameters derived from the voice range profiles can be robust and simple: for example, the maximum and minimum of the frequency and intensity range, and the area. These parameters can be computed for any VRP irrespective of its shape. Several studies have applied these measures (see for example: Ikeda et al.[21], Heylen[5], Sulter et al.[30]). Some authors calculated specific data points such as the so-called center of gravity; this refers to the coordinate of the median semitone note and the median intensity[12]. Others performed a regression analysis in order to characterize the overall slope of the voice range profile[30,32]. Slopes fitted to parts of the highest and lowest intensity contour were proposed by Orr et al.[33]. These measures require the VRP to be more or less smooth in shape. Otherwise, the result of the computation will be meaningless. The VRP had to meet even higher standards of shape and quality for the computations introduced by Klingholz and Martin[32]. They fitted two or three ellipses to the contour of the VRP in the frequency range of the chest and falsetto register or to the chest, middle, and falsetto register. Measures that require smooth, well-shaped VRP's, stand in contrast to measures that characterize the irregularity of the contour. Sulter et al.[30] introduced such measures as contour regularity and a description of the irregularity by means of Fourier descriptors. Eichel [34] suggested a combination of distances from the comfortable F_0 and intensity point to the contour of the VRP in the direction of lower or higher frequency and intensity. Measures that require a smooth VRP contour may not be suitable for the description of pathological VRP's. The shape of the VRP in patients with severe pathology of the vocal folds may deviate sharply from the shape of the VRP of a normal voice. Therefore, to describe the effects of voice therapy in patients with a dysphonic voice, it is preferable to use robust and simple VRP parameters that are not very sensitive to the irregularity of the VRP profiles.

This study has two objectives. The first is to investigate which parameters in the voice range profile of dysphonic patients show significant changes after voice therapy, and therefore may in general be useful in an assessment of voice therapies. In order to get a clinical useful result, patients with diverse pathology were included in the study. The second is to determine the size of the demonstrated effects and how these might relate to phoniatric diagnoses.

Methods

Subjects

Patients participating in this study were selected according to the following criteria. All patients had to suffer from chronic dysphonia. The chronicity of the dysphonia was assumed by an early onset of the dysphonic problems of at least four months before visiting an O.R.L. specialist/phoniatrician. The exact medical diagnosis had to be demonstrated by a comprehensive phoniatric investigation. Patients under 18 years of age were excluded from participating in this study because of possible problems of voice maturation or mutation. Other exclusion criteria were coexisting speech or language disorders but also malignant or pre-malignant lesions, hormonal voice disorders, laryngeal papillomas, substitution of voice after laryngectomy, and spasmodic dysphonia or psychogenic dysphonia.

Pre-therapy and post-therapy data on 17 subjects were incomplete as a result of technical problems (N = 8), extreme deviant voice quality (N =5), or problems in scheduling appointments (N = 4). Of the original 79 subjects, 62 thus remained in the study, having both pre-therapy and post-therapy VRP's available. Out of those remaining, 47 subjects were willing to return to the hospital for follow-up measurements. Table I shows the frequency of the etiologic categories as diagnosed by a phoniatrician. The group of patients enrolled in the study comprised 28 men and 34 women, with an age range from 18 to 76 years. The average age for the female participants was 40 and for the male subjects 48 years.

Together, the subjects reflect the diversity of laryngeal pathology in chronic dysphonia as found in a clinical setting. On the basis of this representative sample, the present study evaluates the overall effects of voice therapy in clinical practice. Before therapy onset (baseline period), speech samples were collected on three different days within six weeks. No significant systematic differences in jitter, shimmer, and noise-to-harmonic ratio between these baseline measurements could be demonstrated. As no significant changes were found, the assumed chronicity of the voice problems was confirmed.

Table I. Distribution of patients by diagnostic categories.

Phoniatric diagnosis	N
Muscle tension dysphonia	10
Vocal fold edema (Reinke edema)	6
Vocal fold nodules	7
Vocal fold polyp	5
Unilateral vocal fold paralysis	4
Other: slight vocal fold abnormalities*	23
Other: severe vocal fold abnormalities**	7

* including chronic laryngitis, slight sulcus, contact ulcus, small (pseudo)cyst, etc.
** including radiotherapy, scar, etc.

Procedure

If the phoniatrician considered voice therapy to be indicated, the patients were referred to speech therapists in their residential area. The indications for referral included trial therapy and pre-surgical voice therapy. The general elements of logopedic voice therapy included voice hygiene advice, exercise training and integration of the newly obtained vocal behavior in spontaneous voicing and speaking. The patients received voice therapy that was adapted to their individual needs and voicing possibilities. No special instructions were given on type of voice therapy. In order to avoid a therapist-related effect, no more than two patients were referred to the same therapist. A total number of 68 speech therapist participated in this study. The therapy lasted at least three months and was given for 30 minutes twice a week or 60 minutes once a week. If necessary according to the ORL-specialist/phoniatrician, the patients were treated during a second period of three months for either 30 or 60 minutes per week.

A voice range profile was made prior to initiating and after finishing voice therapy as well as after a period of three months of no therapy, the so-called follow-up measurement. An automated procedure for obtaining a voice range profile was used (Pabon phonetograph[35], Laryngograph Ltd, London). During phonation, each intensity-frequency combination produced by the patient is recorded with a resolution of one decibel by one semitone. The result is directly displayed on a monitor in a VRP chart. During the recording session, the patients could observe their performance on the monitor. Although, the Pabon phonetograph can produce a so-called filled VRP, only the minimum and maximum contour of the voice range profile was used. Patients were instructed to phonate using a sustained vowel /a:/ as loud and as soft as possible over the maximum frequency range that could be produced by the patient. This way, all frequencies could be assessed. In order to motivate patients to perform to their maximum capacity, the investigator provided verbal support and auditory examples if necessary.

The phonetograph used in this research allows the clinician to adjust the acceptation threshold based on the level of jitter[36]. The equipment will not accept voices with jitter at a higher level than the chosen threshold value; therefore, those utterances will not be registered in the voice range profile. All recordings of the VRP are made using the same threshold level corresponding to normal voicing. This procedure guarantees that only quasi-periodic signals are used.

Analysis

Four groups of parameters were used in determining the effects of voice therapy: minimum and maximum contours, surface areas (integration of the intensity interval per semitone), slopes determined for several parts of these contours, and overall measures (see Table II for the exact definitions). The contours of the voice range profile were measured in terms of the minimum frequency, the maximum frequency (both expressed in semitones), the softest intensity, and the loudest intensity. Within the voice range profile, various surface areas were determined: the total area, the area around the speaking frequency, and the areas below and above the speaking frequency area. These last surface areas were determined over the full as well as over a restricted frequency range next to the speaking frequency area. The speaking frequency was determined by means of acoustic analysis (Multi Dimensional Voice Program by Kay Elemetrics Corporation). That analysis used the mid-segment of a reading text that had been recorded during the same sessions at which the voice range profiles had been made. The mean value of the speaking frequency of both sessions was used for the comparison of two VRP's. Four slopes were calculated: the overall slope of the average minimum and maximum contour for the full frequency range and three slopes fitted to the contours of the VRP. These slopes were: the slope of the maximum intensity contour below the speaking frequency, the slope of the maximum intensity contour above the speaking frequency, and the slope of the minimum intensity contour. If there was a jump in the contour of more than 10 dB, the semitone preceding this change will was considered as the final frequency when calculating either one of the last two slopes.

Table II. Parameters used in the comparison of the voice range profiles. SF=mean speaking frequency determined for each patient individually from text reading.

Group	Parameter	Definition / Notes
Minimum and maximum of contours	minimum frequency (ST)	lowest semitone
	maximum frequency (ST)	highest semitone
	softest intensity (dB)	minimum SPL
	loudest intensity (dB)	maximum SPL
Surface areas (dB*semitone)	total surface area	surface area between the maximum and minimum intensity contour
	surface area around the SF (speaking frequency area)	surface area between the maximum and minimum intensity contour for a limited frequency range: two semitones above and two semitones below the semitone in which the SF, resulting in an interval of five semitones
	surface area below the SF area	surface area between the maximum and minimum intensity contour for a limited frequency range: starting at three semitones below the SF up to the lowest semitone
	surface area above the SF area	surface area between the maximum and minimum intensity contour for a limited frequency range: starting at three semitones above the SF up to the highest semitone
	restricted area below the SF area	surface area between the maximum and minimum intensity contour for a limited frequency range: starting at three semitones below the SF up to six semitones below the SF
	restricted area above the SF area	surface area between the maximum and minimum intensity contour for a limited frequency range: starting at three semitones above the SF up to eight semitones above the SF
Slopes (dB/semitone)	overall slope	slope of the average minimum and maximum contour for the full frequency range
	slope of maximum intensity contour below the SF	frequency range: minimum frequency plus one semitone up to and including the SF
	slope of maximum intensity contour above the SF	frequency range: SF up to and including the semitone with max. intensity. If the maximum contour drops 10 dB or more, this semitone is the last point of the frequency range.
	slope of minimum intensity contour	frequency range: one semitone above lowest frequency up to two semitones below maximum frequency. If an increase of more than 10 dB occurs in the contour, the range stops before the increase.
Overall measures	mean dB range (dB)	mean of the intensity ranges for all semitones in the VRP
	median frequency (Hz)	median frequency for the full frequency range
	median intensity (dB)	median intensity for the full intensity range
	contour regularity	total surface area divided by the squared perimeter and standardized to the result for a circle

Furthermore, three overall measures were calculated: the mean intensity range for the full frequency range, the median frequency and median intensity for the full voice range profile, and the contour regularity as defined by Sulter[30]

Wilcoxon signed-rank tests were used to test for systematic differences between pre- and post-treatment data as well as between post-treatment and follow-up data. Apart from the tests using all subjects, tests were performed for the two genders and for each diagnosis group. The computations were made with SPSS 8.0.0 (SPSS Inc., USA).

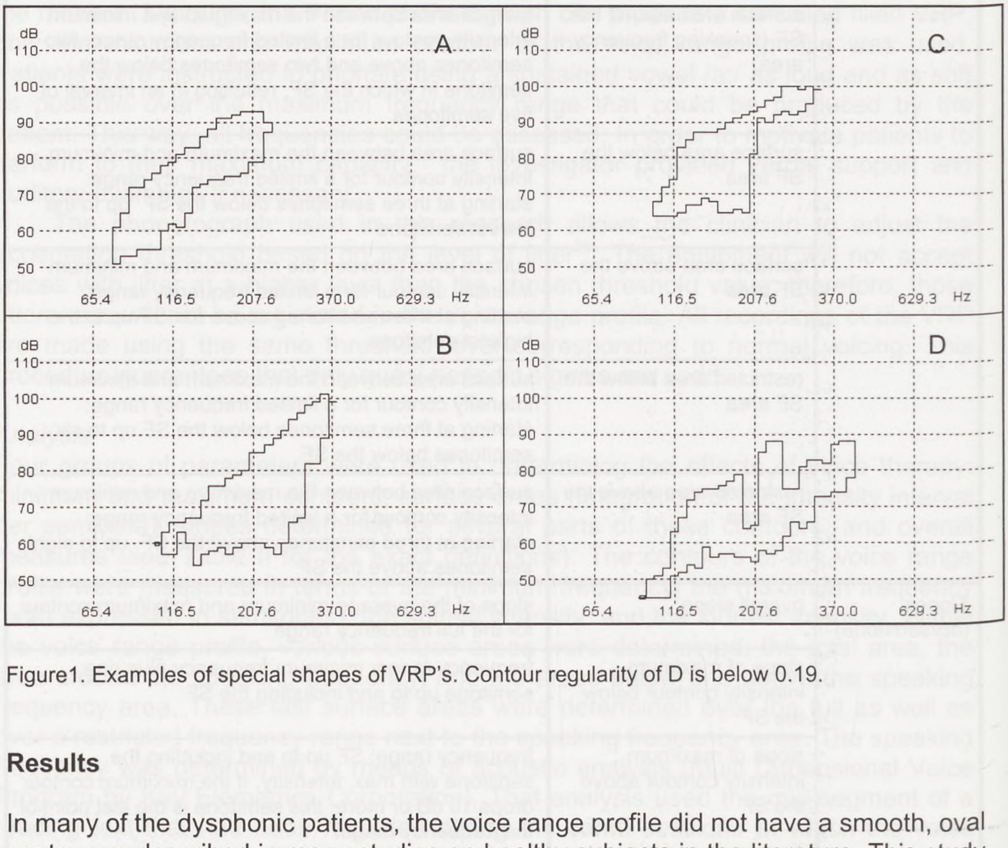

Figure1. Examples of special shapes of VRP's. Contour regularity of D is below 0.19.

Results

In many of the dysphonic patients, the voice range profile did not have a smooth, oval contour as described in many studies on healthy subjects in the literature. This study on dysphonic patients found diverse shapes of VRP's including normal shapes but also long-drawn-out, triangular, or irregular forms (see Figure 1). VRP's with a smooth contour resembling the normal shape were found in about 40% of the patients. In another 40%, the VRP showed a long, narrow extension in the higher frequency range (Figure 1C), which presumably corresponds to the restricted dynamic voice possibilities in the falsetto register. In five cases, the high frequency part was not connected to a main body of the VRP. In some cases, a sudden change in intensity was found at a frequency where no change in register was to be expected. In certain patients, a drop in the minimum intensity contour occurred at one specific frequency. The last three possibilities were observed among roughly 20% of the patients. The VRP's of those 20% differed from the VRP's of the rest in having a low contour regularity (see Table II for a definition). The optimal threshold was 0.19, differentiating VRP's with more or less smooth contours (Figure 1A,B,C) from VRP's with irregular contours (for example Figure 1D) or VRP's consisting of more than one closed contour.

Pre- and post-therapy data

Table III shows the median as well as the 25th and 75th percentiles for the pre-therapy data and the corresponding figures for the post-therapy data. For all patients (N=62),the pre-therapy data were compared to the post-therapy data with a Wilcoxon signed-rank test.). After therapy, the following significant differences at a *p* level of less than 0.01 were found: a decrease in minimum frequency and softest intensity, and an enlargement of the total surface area of the VRP and the speaking frequency area as well as both surface areas below the speaking frequency area. Furthermore, the slope of the maximal intensity contour above the speaking frequency proved to become less steep. Although many variables were tested, the size and *p* levels of the significant differences indicate that there is a real overall significant effect. Even a simple Bonferroni correction ($p<0.05/18$) leads to an overall significance level of $p=0.05$. In general, the low frequency part and the speech range of the VRP are larger after therapy. The significant decrease of the slope on the maximum intensity contour above the speech frequency has no direct clinical implications.
all patients.

Bivariate Spearman correlation coefficients for the parameter differences between post- and pre-treatment data were low (|R|<0.6), except for pairs of parameters that were obviously related: the full and corresponding restricted surface areas (R>0.72), the maximal frequency and median frequency (R=0.81), and the mean intensity range and total surface area (R=0.67). The total surface area was only highly correlated with the area above the speaking frequency area (R=0.91). The mutual correlations between the extreme points of the VRP were all below 0.30. The correlations of these points with the surface areas were generally lower than 0.61. The exceptions were the correlation of the minimal frequency and the surface area below the speaking frequency area (R=-0.77). The generally low correlations indicate that many different changes in the VRP are possible. In other words, the effect of voice therapy on the VRP is multidimensional. The multidimensional character was confirmed by factor analysis based on the Spearman correlations, which yielded six factors explaining 82% of the total variance when all parameters were used. When using only the parameters that were significantly different for the pre- and post-therapy data, two factors were found to explain 66% of the variance. After rotation, the change in the lowest frequency and the areas below the speaking frequency range were found in one of the factors. The change in the minimal intensity and the speaking frequency area loaded highly on the second factor.

Table III. Median and 25th and 75th percentiles for the pre- treatment data, and the post- minus pre-treatment data (** 0.001<p<0.01 ; *** p)

Parameter	Pre-treatment data (N=62)		Post- minus pre-treatment data (N=62)	
	Median	25',75' perc.	Median	25',75' perc.
minimum frequency (ST)	28.5	25.0; 33.0	-1.0**	-2.2; 0.0
maximum frequency (ST)	54.5	51.0; 60.2	1	-2.2; 2.2
softest intensity (dB)	52	46.0; 55.0	-2.0**	-6.0; 1.0
loudest intensity (dB)	99.5	94.8; 104.8	-1	-4.0; 3.0
total surface area (dB * ST)	614	392; 827	64**	-38; 180
surface area around the SF (dB * ST)	128	98; 169	14***	-6; 39
surface area below the SF (dB * ST)	90	38; 150	18**	-12; 52
surface area above the SF (dB * ST)	357	153; 536	17	-64; 87
restricted area below the SF (dB * ST)	62	36; 79	12**	-9; 24
restricted area above the SF (dB * ST)	168	122; 229	2	-14; 39
overall slope	1.2	0.73; 1.58	-0.04	-0.27; 0.28
slope of max. intensity contour below the SF	2.31	1.70; 3.50	0.14	-0.52; 0.75
slope of max. intensity contour above the SF	1.34	0.86; 1.81	-0.27***	-0.66; 0.19
slope of minimum intensity contour	0.78	0.41; 1.13	0.05	-0.33; 0.47
mean dB range (dB)	21.6	16.6; 27.9	1.1	-2.3; 4.6
median frequency (Hz)	42.2	38.9; 44.5	0	-2.0; 1.0
median intensity (dB)	75	69.0; 79.0	-0.5	-5.6; 5.5
contour regularity	27	20.7; 35.5	-1	-6.3; 6.1

A second analysis was performed to check for a possible influence of voice range profiles with irregular contours on the overall results. All subjects showing a contour regularity of less than 0.19 in either one of the pre- or post- treatment voice range profiles were excluded, leaving 43 subjects. The tests with these patients revealed significant differences in the same parameters as found in the whole group. Despite the lower number of subjects, higher significance levels were reached (all $p<0.003$).

As therapy effects may depend on gender, separate analyses were performed for men and women. Men (N=28) showed a significant decrease in minimum frequency, softest intensity, and median frequency (all $p<0.01$). The total surface area was enlarged ($p<0.02$), as were the full ($p<0.01$) and restricted ($p<0.02$) areas below the speaking frequency. For women (N=34), the only significant enlargement of the surface area was around the speaking frequency ($p<0.03$). The slope of the maximum intensity contour above this frequency was less steep for women ($p=0.000$). The therapy effects were similar for both sexes, though they were clearly more pronounced in men. Only the therapy effect in the slope on the maximum intensity contour above the speaking frequency was significantly different in men and women. This effect is only found for women.

Men and women were distributed more or less evenly over the diagnostic categories, with the exception of vocal fold nodules (N=7) and polyps (N=5), which was comprised of women. Therefore, the statistical procedure for the women was repeated excluding these two categories. The demonstrated effects of therapy were similar to the results of the analysis that included all diagnostic categories.

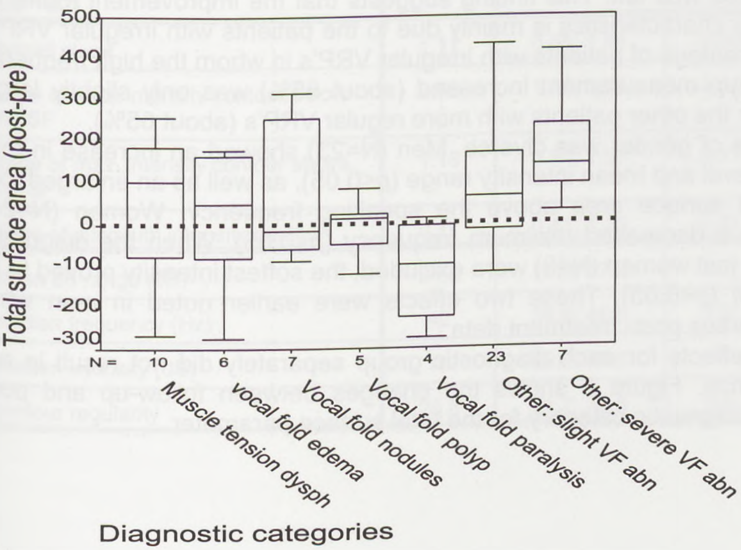

Figure 2. Therapy effects (post- minus pre-treatment data) for the total surface area parameter per diagnostic category. (Two outliers are out of sight.)

No obvious differences in therapy effects were observed for the diagnostic groups when analyzed separately. Figure 2 shows the differences between post- minus pre-treatment data per diagnostic category for the total surface parameter. Although several diagnostic groups comprised a small number of patients, tests were performed for each of the diagnostic groups separately. Only in the groups containing patients with slight and severe vocal fold abnormalities (N=23 and 11) were significant differences found for most of the parameters that were significant when the analysis included all patients. Patients with paralysis showed no beneficial changes. In the other groups, about 70% of the patients showed improvement in the parameters that had changed significantly when tested for all patients.

Post-therapy and follow-up data

The number of patients who came back for a follow-up measurement (47) was lower than the number who were enrolled in the study (62). Differences between the post-therapy data and the follow-up data were statistically analyzed (see Table IV). Including all 62 subjects, a significant change in maximum frequency ($p<0.05$) and loudest intensity ($p<0.01$) could be demonstrated. The total surface area as well as the surface area above the speaking frequency showed an enlargement (respectively, $p<0.05$ and $p<0.01$). The mean intensity range was broadened and the median frequency heightened. The significant changes between pre- and post-treatment data mainly occurred at the softer, lower region of the voice range profiles. The overall differences between post- and follow-up data appeared in the maximum intensity contour at higher frequencies. If patients with a severe irregularity in either one of the voice range profiles (contour regularity <0.19; N=15) were excluded, no significant difference was left. This finding suggests that the improvement found for the high frequency characteristics is mainly due to the patients with irregular VRP's. However, the percentage of patients with irregular VRP's in whom the high frequency part in the follow-up measurement increased (about 85%) was only slightly larger than that found for the other patients with more regular VRP's (about 65%).

The influence of gender was diverse. Men (N=23) showed an increase in their loudest intensity level and mean intensity range ($p<0.05$), as well as an enlarged total surface area and surface area above the speaking frequency. Women (N=24), however, showed a decreased minimum frequency ($p<0.05$). When the diagnostic groups containing just women (N=8) were excluded, the softest intensity proved to be significant as well ($p<0.05$). These two effects were earlier noted in men when comparing pre- versus post- treatment data.

Testing the effects for each diagnostic group separately did not result in any significant difference. Figure 3 shows the changes between follow-up and post-therapy data per diagnostic category for the total surface parameter.

Table IV. Median and 25th and 75th percentiles for the pre-treatment data, and the follow-up minus post-treatment data (* 0.05<p<0.01; ** p<0.01)

Parameter	Pre-treatment data (N=47)		Follow-up minus post-treatment data (N=47)	
	Median	25',75' perc.	Median	25',75' perc.
minimum frequency (ST)	29	24.0; 33.0	0	-1.0; 1.0
maximum frequency (ST)	54	52.0; 59.0	1.0*	-1.0; 3.0
softest intensity (dB)	52	46.0; 55.0	-1	-4.0; 2.0
loudest intensity (dB)	98	94.0; 104	2.0**	0.0; 5.0
total surface area (dB * ST)	614	394; 823	27*	-26; 94
surface area around the SF (dB * ST)	128	97; 168	3	-11; 21
surface area below the SF (dB * ST)	90	43; 149	-3	-30; 19
surface area above the SF (dB * ST)	363	154; 534	30**	-25; 92
restricted area below the SF (dB * ST)	62	37; 85	-2	-12; 7
restricted area above the SF (dB * ST)	169	123; 232	3	-12; 17
overall slope	1.11	0.65; 1.44	0.08	-0.24; 0.29
slope of max. intensity contour below the SF	2.58	1.76; 3.55	-0.16	-0.60; 0.56
slope of max. intensity contour above the SF	1.15	0.81; 1.57	0.11	-0.26; 0.54
slope of minimum intensity contour	0.77	0.41; 1.16	0.07	-0.52; 0.39
mean dB range (dB)	21.9	17.9; 27.2	1.7*	-2.0; 4.0
median frequency (Hz)	42	38.5; 44.5	0.5*	-0.5; 1.5
median intensity (dB)	75	69.0; 79.0	-1	-4.5; 3.0
contour regularity	26.8	21.2; 35.5	0.6	-8.9; 4.8

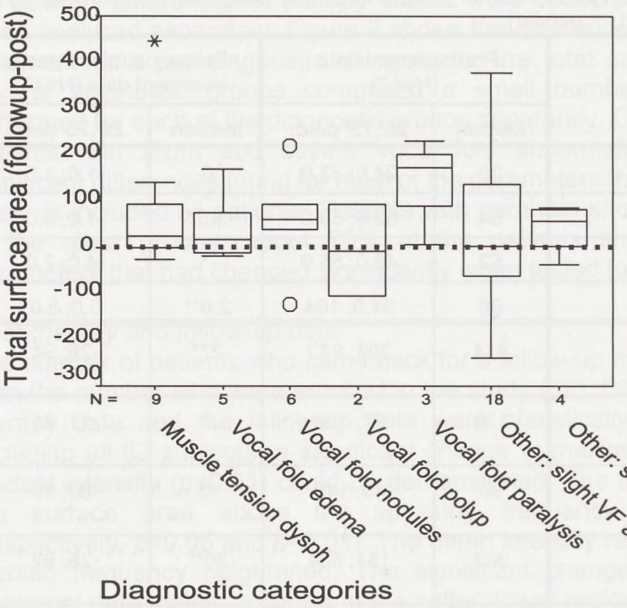

Figure 3. Follow-up data minus post-treatment data for the total surface area parameter per diagnostic category.

Discussion and conclusion

One of the main drawbacks of using voice range profiles of dysphonic patients is the diversity and irregular shape of the VRP contours. Although the majority of the voice range profiles in this study showed regular, more or less smooth contours, there were also many exceptions. Some subjects had separated voice areas in the VRP, jagged boundaries of the lower or upper contour, or restricted dynamic falsetto phonations. Because of this diversity, parameters that require more or less smooth VRP contours are not suitable for studying dysphonic patients. Therefore, in research on dysphonic patients, robust and simple VRP measures should be used to describe the effects of voice therapy.

From a physical point of view, the recurrence of irregular contours in the VRP of dysphonic patients is not remarkable. Many of the abnormalities in those patients are characterized by local changes in the mass or elasticity of the vocal folds. The vibration pattern of the folds will be influenced by these irregularities and may result in complex vibration modes. This may explain that in some VRP's in a small frequency band a much larger intensity range was found than for the frequencies outside this band.

Another question concerns the method of recording voice range profiles, specifically, the use of automated or clinician-elicited voice range profiles[37]. The main problem is when to accept a certain phonation as an adequate data point in the VRP.

Titze described a so-called acceptability criterion in terms of stability as a required minimum number of repetitions with a certain maximum amount of variation in intensity and frequency between these repetitions. When the clinician records the VRP manually, the criterion for accepting a maximum or minimum intensity point will be subjective. In the Pabon phonetograph a fixed jitter threshold for accepting data can be set. As a measurement is rejected when the jitter of the voice exceeds this threshold, the criterion for accepting a point in a VRP will be more objective.

Jitter may be relatively high at the extremes of the possibilities of a voice, and, therefore, the contours of the VRP may depend on the jitter threshold used. With a lower threshold, a larger VRP may be found because more voicings will be accepted. Another effect of the jitter threshold may be a slight arbitrariness about whether a phonation is accepted or not. When the jitter values at the maximum or minimum intensities are near the jitter threshold, small differences in jitter will determine whether a point in the VRP is accepted or not. When such small differences in jitter occur for neighbouring frequencies, an irregular shape of the VRP contour may result. However, when a subject reaches a certain intensity/frequency point more than once, this effect will be reduced. The use of the jitter threshold may have influenced the therapy effects by a change of the amount of jitter at the borders of the VRP before and after therapy. However, it is questionable whether the jitter at a new margin of the possibilities of the voice will be very different.

The main finding after voice therapy was an enlargement of the VRP in the low frequency range. This effect was revealed by a lowering of the minimal frequency and intensity, an increase in the surface areas below the speaking frequency, and consequently an increase in the total surface area. The restricted areas on both sides of the speaking frequency area did not provide any new information on therapy effects when compared to the information already given by the full areas above and below the speaking frequency. Three months after the end of voice therapy, significant changes in the voice range profiles still occurred. In contrast to changes after therapy, which were characterized by increased possibilities at the lower frequencies and intensities, the follow-up data showed improvement at higher frequencies and intensities.

The enlargement on the low frequency side of the VRP directly after voice therapy may be the result of the emphasis put on the speaking voice as well as phonation in the chest register during voice therapy. Furthermore, voice relaxation exercises may primarily stimulate the voice production at low frequency. An increase of the lower frequency formants caused by an acquired change of the shape of the vocal tract cannot be ruled out in this respect as well. The enlargement of the VRP three months after therapy may be the result of a patient's awareness during therapy of vocal hygiene instructions as given by the speech therapist and then, after finishing therapy, a weakening of compliance with these instructions.

The therapy effects appeared to be slightly dependent on gender. Generally, the beneficial changes were found to occur more in men than in women. An inexplicable difference between the therapy effects between men and women is the significant reduction in the slope of the regression line at the maximal intensity contour above the speaking frequency in women but not in men.

Differences between the diagnostic categories were not clear. Less pronounced differences might go unobserved, due to the diversity in phoniatric diagnoses and the relatively small number of patients in each category. The only clear observation was that the VRP's of patients with paralysis did not change much.
Even though no significant changes in some of the voice range profiles could be demonstrated, this does not imply that voice therapy had no beneficial effects at all on these subjects. It merely indicates that the effects could not be found by means of phonetography. As the voice is considered to be a multidimensional phenomenon, the evaluation of voice therapy will require a multidimensional assessment instrument, including instruments such as perceptual evaluation, laryngostroboscopy and self-evaluation by the patient[38]. This aspect of multidimensionality can even be seen in the association between the parameters of the voice range profiles; no high correlations between the parameters were found. Looking at the parameters in which significant therapy effects were found, and considering the correlations found between the effects, the following parameters are recommended when evaluating the effects of voice therapy on the voice range profiles of chronically dysphonic patients. First, the investigator can choose among the surface area below the speaking frequency area, the minimum frequency, and the softest intensity. Second, the investigator can use the surface area above the speaking frequency area, the maximum frequency, or the loudest intensity level. Third, the surface area around the speaking frequency may be used; the total surface area may serve as a kind of integrated parameter.

Reference List

1. Damsté PH. The phonetogram. *Pract.Otorhinolaryngol.(Basel)* 1970;32:185-187.
2. Schutte HK, Seidner W. Recommendation by the Union of European Phoniatricians (UEP): standardizing voice area measurement/phonetography. *Folia Phoniatr.* 1983;35:286-288.
3. Hirano M. Objective evaluation of the human voice: clinical aspects. *Folia Phoniatr.* 1989;41:9-144.
4. Heylen L, Wuyts FL, Mertens F, De Bodt M., Pattyn J, Heyning PH van de. Comparison of the results of the frequency & the intensity data of the BSGVD with phonetogram characteristics. *Acta Otorhinolaryngol.Belg.* 1996;50:353-360.
5. Heylen L. *De klinische relevantie van het fonetogram: een onderzoek bij kinderen en leerkrachten.* Doctoral dissertation. University of Antwerpen: 1997.
6. Eckley CA, Sataloff RT, Hawkshaw M, Spiegel JR, Mandel S. Voice range in superior laryngeal nerve paresis and paralysis. *J.Voice* 1998;12:340-348.
7. Gramming P, Åkerlund L. Non-organic dysphonia. II. Phonetograms for normal and pathological voices. *Acta Otolaryngol.Stockh.* 1988;106:468-476.
8. Airainer R, Klingholz F. Quantitative evaluation of phonetograms in the case of functional dysphonia. *J.Voice* 1993;7:136-141.
9. Gelfer MP. Stability in phonational frequency range. *J.Commun.Disord.* 1989;22:181-192.

10. Hollien H, Dew D, Philips P. Phonational frequency ranges of adults. *J.Speech Hear.Res.* 1971; 14:755-760.
11. Kotby MN, Abul-Ela MY, Orabi AA. Voice range profile as a quantitative measure of vocal function: normative data. Kotby MN. *Proceedings of the XXIVth World congress of the International Association of Logopedics & Phoniatrics, Cairo.* 1995;46-49
12. Mersbergen M van, Verdolini K, Titze IR. Time-of-day effects on voice range profile performance in young, vocally untrained adult females. *J.Voice* 1999;13:518-528.
13. Sihvo M, Laippala P, Sala E. A study of repeated measures of softest and loudest phonations. *J.Voice* 2000;14:161-169.
14. Coleman RF. Sources of variation in phonetograms. *J.Voice* 1993;7:1-14.
15. Reich AR, Frederickson RR, Mason JA, Schlauch RS. Methodological variables affecting phonational frequency range in adults. *J.Speech Hear.Disord.* 1990;55:124-131.
16. Zraick RI, Nelson JL, Montague JC, Monoson PK. The effect of task on determination of maximum phonational frequency range. *J.Voice* 2000;14:154-160.
17. Gramming P, Sundberg J. Spectrum factors relevant to phonetogram measurement. *J.Acoust.Soc.Am.* 1988;83:2352-2360.
18. Kotby MN, Titze IR, Saleh MM, Berry DA. Fundamental frequency stability in functional dysphonia. *Acta Otolaryngol. (Stockh.)* 1993;113:439-444.
19. Gramming P, Sundberg J, Åkerlund L. Variability of phonetograms. *Folia Phoniatr.* 1991;43: 79-92.
20. Behrman A, Agresti CJ, Blumstein E, Sharma G. Meaningful features of voice range profiles from patients with organic vocal fold pathology: a preliminary study. *J.Voice* 1996;10:269-283.
21. Ikeda Y, Masuda T, Manako H, Yamashita H, Yamamoto T, Komiyama S. Quantitative evaluation of the voice range profile in patients with voice disorder. *Eur.Arch.Otorhinolaryngol.* 1999;256 Suppl 1:S51-S55
22. Uloza V, Šiupšinskiene N. Voice Range Profile as a tool for clinical voice assessment and monitoring. Kotby MN. *Proceedings of the XXIVth World congress of the International Association of Logopedics & Phoniatrics, Cairo.* 1995:37-40.
23. Verdonck-de Leeuw I , Hilgers FJ, Keus RB et al. Multidimensional assessment of voice characteristics after radiotherapy for early glottic cancer. *Laryngoscope* 1999;109:241-248.
24. Pedersen MF. Computed phonetograms in adult patients with benign voice disorders before and after treatment with a nonsedating antihistamine (loratadine). *Folia Phoniatr.* 1991;43:60-67.
25. Klingholtz F. Die Stimme des Sangers im Phonetogramm. *Laryngol.Rhinol.Otol.(Stuttg.)* 1989;68; 1:62-66.
26. Sulter AM, Schutte HK, Miller DG. Differences in phonetogram features between male and female subjects with and without vocal training. *J. Voice* 1995;4:363-377.
27. Schultz-Coulon HJ. *Stimmfeldmessung*. Berlin: Springer-Verlag; 1990.

28. Hirano M, Tanaka S, Fujita M, Terasawa R. Fundamental frequency and sound pressure level of phonation in pathological states. *J.Voice* 1991;5:120-127.
29. Åkerlund L. Averages of sound pressure levels and mean fundamental frequencies of speech in relation to phonetograms: comparison of nonorganic dysphonia patients before and after therapy. *Acta Otolaryngol.Stockh.* 1993;113:102-108.
30. Sulter AM, Wit HP, Schutte HK, Miller DG. A structured approach to voice range profile (phonetogram) analysis. *J.Speech Hear.Res.* 1994;37:1076-1085.
31. Bloothooft G, Binnenpoorte D. Towards a searchable resource of phonetograms. In:ed. *Voicedata98: Symposium on databases in voice quality research and education.* University of Utrecht: G. de Krom: 1998.
32. Klingholtz F, Martin F. Die quantitative Auswertung der Stimmfeldmessung. *Sprache-Stimme-Gehör* 1983;7:106-110.
33. Orr R, Jong F de, Cranen B. Het fonetogram als discriminerend instrument. *Stem-, Spraak- en Taalpathologie* 2000;9:22-37.
34. Eichel HW. Der Stimmfeldindex, ein Vorschlag zur quantitativen Auswertung des Stimmfeldes für die Praxis. *Sprache-Stimme-Gehör* 1988;12:63-64.
35. Pabon JPH. Objective acoustic voice-quality parameters in the computer phonetogram. *J.Voice* 1991;3:203-216.
36. Pabon JPH, Plomp R. Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *J.Speech Hear.Res.* 1988;31:710-722.
37. Titze IR, Wong D, Milder MA, Hensley SR, Ramig LO. Comparison between clinician-assisted and fully automated procedures for obtaining a voice range profile. *J.Speech Hear.Res.* 1995;38:526-535.
38. Dejonckere PH, Bradley P, Clement P et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur.Arch.Otorhinolaryngol.* 2001;258:77-82.

Chapter 5

Documentation of progress in voice therapy: perceptual, acoustic and laryngostroboscopic findings pre- and post-therapy

R. Speyer, G.H. Wieneke, P.H. Dejonckere
Submitted (Journal of Voice)

Documentation of progress in voice therapy: perceptual, acoustic and laryngostroboscopic findings pre- and post-therapy

Summary

The effect of voice therapy in a group of chronically dysphonic patients with diverse diagnoses was studied according the normal clinical procedure. The results were evaluated by perceptual rating, acoustic analysis, and the assessment of laryngostroboscopic recordings. Although the group effects for the differences between post- and pre-therapy data were clearly significant, the effects of voice therapy for the individual patients were divergent. For each of the three evaluation methods, a significant improvement was found for about 40% to 50% of the patients. The diversity of the therapy outcome among the patients could not be explained by the pre-therapy status nor to age, gender, or diagnosis groups. In general, the perceptual ratings and the acoustic parameters from the baseline data were clearly correlated. However, these characterizations of the voice were only moderately correlated with the visual evaluation of the vocal fold vibrations. Relations between the three evaluation tools for the changes due to voice therapy were very weak. The low correlation among the three methods suggests that a multidimensional evaluation of the voice is necessary to give a complete picture of the therapy outcome.

Introduction

The growing importance of evidence-based medicine is stimulating the demand for scientific research on the effects of (para) medical therapy. As society has a greater stake in such research, funding is becoming more readily available. Thus far, few studies have been done on the effects of voice therapy. In some of them, no comparison was possible between pre- and post-therapy data, because no measurements had been made prior to therapy (for example Koufman and Blalock[1], and Raabe and Pascher[2]) Furthermore, the number of subjects studied is often too small to permit generalization of the results (for instance, Fex et al.[3], Holmberg et al.[4], and Verdolini et al.[5]). Almost all other studies published over the last two decades concern either specific groups of subjects (for example Kitzing and Åkerlund[6], Murry and Woodson[7], and Heuer et al.[8]) or voice therapies in which a strict protocol was used (for example Roy et al.[9], Bassiouny[10], and Kotby et al.[11]). Conclusions drawn from these experiments can only apply to those groups of patients (e.g., persons with vocal nodules) or therapies involved in the studies. Thus, the findings may not be useful for voice therapy in everyday clinical practice. If the aim is to provide guidance on how to deal with referrals, the study would have to include many different diagnoses and diverse therapies. For the reader interested in this aspect, Carding[12] is a recommended reading on the effects of therapy.

A major obstacle to the investigation of therapy effects is an ethical issue: is it acceptable to include a placebo group in the study, whether or not therapy is given later?. Most studies on the effect of voice therapy do not use a placebo group and therefore do not deal with this issue. One way to circumvent it is by taking repeated measurements over a prolonged period before starting therapy. When researchers use a multiple baseline design, the pre-therapy measurements can be used to test for spontaneous recovery. Another question is which instruments should be used to evaluate the effects of voice therapy. The perceptual evaluation of voice quality is generally taken as the gold standard for voice assessment. In fact, in the clinical practice of a speech therapist, other evaluation tools are rarely available. Therefore, the main goal of voice therapy is usually to improve the patient's voice quality, as perceived by the therapist. However, the literature describes some of the problems that may arise: e.g., unstable internal standards that may vary over time for comparing speech samples[13] and the absence of universally accepted definitions for perceptual concepts such as breathiness or roughness[14]. Besides the subjective evaluation tools, there are also some objective measures such as acoustic analysis or laryngostroboscopy. These methods pertain more directly to vocal fold vibration. Therefore, the judgments based on them may reflect different dimensions of the voice than those emanating from perceptual evaluation. In acoustic analysis, several issues have to be taken into account: the possibility of errors in period tracking thus finding the correct period duration; the inadequacy of acoustic analysis in very irregular voicing; and unnatural aspects of speech samples such as sustained vowels. Direct insight into the changes in vocal fold vibration is provided by another technique, namely laryngostroboscopy. Stroboscopic images shed light on the regularity of the vibration as well as on changes in morphology and their effect on vibration. A visuo-perceptual evaluation of video recordings of the stroboscopic recordings permits the investigator to detect abnormalities in the source process of the voice. The purpose of visual evaluation is to describe the physiology of the vibration pattern of the vocal folds. But while laryngostroboscopy has certain advantages, the investigator still has to deal with the problem of intra- and intervariability in the data. These variabilities can be partly avoided by digital processing of the images, but such a procedure is restricted to high quality recordings[15].

This study considers the effects of voice therapy in a group of chronically dysphonic patients. A multidimensional assessment protocol is used in order to be informed about the different effects on voice and vocal fold vibration. If a restricted instrumentarium is used, just part of the effects of therapy may be detected. In this study the effect on the voice was determined by perceptual rating and acoustic analysis; and the effect on the vibrational pattern of the vocal folds, or the vocal fold biomechanics, was evaluated by laryngostroboscopy. For ethical reasons, no placebo group was included. Instead, a prolonged baseline period was used, allowing for repeated measurements during the baseline period. These measurements were used to confirm that no spontaneous recovery of the chronic dysphonia could be presumed. The aim was to evaluate the efficacy of voice therapy as used in actual practice. Therefore, no restrictions were placed on the type of voice therapy administered, and patients with a wide range of laryngeal diagnoses were admitted to the study.

Methods

Subjects

Patients were eligible to take part in this study if they met two criteria. Their dysphonia had to be chronic. Chronicity was assumed when the dysphonic problems had started at least four months before the patients visited an otolaryngologist at the Phoniatric Department of the University Hospital Utrecht. And the diagnosis must have been based on a complete ENT examination including videostroboscopy. Patients were excluded from the study if they were under 18 years of age, because voice maturation or mutation might confound the results. The other exclusion criteria were coexisting speech or language disorders, malignant or pre-malignant lesions, hormonal voice disorders, laryngeal papillomas, substitution voices after oncological surgery, spasmodic dysphonia, or psychogenic dysphonia.

In total, 78 subjects participated in the study. Table I shows the frequency of the etiologic categories as diagnosed by an otolaryngologist. The group of patients comprised 35 men and 43 women, ranging in age from 18 to 76 years. The female participants were on average 40 years old, the male subjects 46.

Table I. Distribution of patients by diagnostic categories.

Laryngeal diagnosis	N =78
Muscle tension dysphonia	12
Submucosal swelling	8
Vocal fold nodules	10
Vocal fold polyps	6
Unilateral vocal fold paralysis	7
Other: slight vocal fold abnormalities	24
Other: severe vocal fold abnormalities	11

Procedure

If the otolaryngologist found voice therapy to be indicated, the patient was referred to a speech therapist near home. The therapist received a copy of the patient's medical file and the voice therapy was adjusted to the patient's needs. No restrictions were placed on the therapist's choice of voice therapy applied. Therapy lasted at least three months and was given for 30 minutes twice a week or 60 minutes once a week. If necessary, a patient was given a second three-month period of therapy for either 30 or 60 minutes per week.

Within six weeks prior to therapy, three audio recordings were made of the patient's voice. These baseline measurements served as an individual point of reference for evaluating the effects of therapy. The time of day for recording was not standardized. The only videolaryngostroboscopic recording session during the six-week pre-therapy period was on the first visit. Both the audio and the videolaryngostroboscopic recordings were repeated after the patients finished voice therapy.

Audio recordings

Audio samples were recorded on digital audio tape (recorder: Sony DA-7 or Aiwa HD-S200). Attached to the recorder was a condenser microphone (Sennheisser K3 or Sony ECM-221) placed in a harmonica holder at a mouth-to-microphone distance of approximately five centimetres and an angle of approximately 45 degrees. The subjects were asked to read aloud a text and produce three sustained vowels /a:/ of several seconds duration. They were instructed to choose a comfortable pitch and loudness level. The speech samples were digitized at a sample frequency of 50.0 kHz using the Computerized Speech Lab Model 4300 (CSL from Kay Elemetrics Corporation, Pine Brook, New Jersey).

Perceptual evaluation

The mid-section of the recorded text - a segment of about ten seconds - was used for blind perceptual evaluation by a panel of five expert listeners. This panel consisted of two otolaryngologists and three experienced speech therapists. Prior to the experiments, the members of the panel took part in four group training sessions to learn the listening and scoring procedure. Reading fragments of dysphonic voices were judged and discussed during sessions of about 45 minutes. The three therapists in the panel did not give treatment to any of the patients involved in this experiment, and are working in a clinical setting. The speech therapists that give the voice therapy to the patients work in their own practices.

The instrument used for perceptual evaluation consisted of three visual analogue scales representing grade (the overall impression of deviance in voice quality), breathiness, and roughness as proposed by Hirano[16]. Their scores ranged from 0 to 100 millimetres, representing a normal voice at one end and an extremely severe disorder or deviant voice quality at the other end of the scale.

For each patient, there were three baseline measurements and one post-therapy measurement. These four voice samples were combined to make six pairs, consisting of two baseline samples (three pairs) or one of the baseline samples and the post-therapy sample (three pairs). The pairs of voice samples for each patient were presented to the listeners at random. They rated each sample of a pair on a visual analogue scale. The mean rating of the five listeners for each recorded sample was computed for each patient. Additionally, differences between voice samples within a pair were computed. The visual analogue scales were displayed on a computer monitor, and the score was entered by means of a mouse click on the correct position. Listeners were allowed to listen to the stimuli as often as convenient. Rating took place in eight individual listening sessions of about one and a half hours.

Acoustic analysis

The relatively stable mid-vowel sections of the sustained vowels were analyzed by the Multi-Dimensional Voice Program (MDVP from Kay Elemetric Corporation). The first and last 200 ms of the sustained phonation were removed. The quality of the signals was visually inspected, and only quasi-periodic signals were analyzed [17] to assure a correct acoustic analysis. Three acoustic parameters were computed for each sample. The percentage of jitter (Jitt %) gives an indication of the variability of the pitch period within the analyzed voice sample. It represents the relative period-to-period (very short term) variability. The percentage of shimmer (Shim %) gives an indication of the period-to-period variability of the peak-to-peak amplitude. The evaluation of the noise present in the signal was expressed as noise-to-harmonic ratio (NHR): the average ratio of energy of the inharmonic components in the 1500-4500 Hz range to the harmonic components energy in the 70-4500 Hz range.

Laryngostroboscopy
If possible, a rigid endoscope was employed during videolaryngostroboscopy. Some subjects could not bear rigid videoendoscopy. In those cases, a flexible endoscope had to be used in pre- and post-therapy assessments. If necessary, a topical anesthetic (lidocaine hydrochloride 4%) was applied. Recordings on videotape (U-matic videocassette recorder VO-5630, Sony) were made of the vocal fold vibration during repeated stable phonation of a sustained vowel /a:/ or /i:/ at a comfortable pitch and loudness using normal as well as stroboscopic light (Stroboscope RLS 9100, Kay Elemetrics Corp, Pine Brook, New Jersey). In addition, images of the vocal folds in a resting position were recorded.

The recordings were made by several otolaryngologists. At least one year after the recordings were made, the actual visuo-perceptual evaluation of the recordings was done by the senior member of these otolaryngologists, including some repeated measurements. The time between the repeated evaluations was a few months. This otolaryngologist judged both pre- and post-therapy recordings by means of visuo-perceptual evaluation. The recordings were randomized per subject. The judge did not know whether the recordings dated from before or after therapy. Visual analogue scales were used to evaluate the recordings on six parameters: the degree of insufficient vocal fold closure; the maximum amplitude of the vocal fold vibration; the quality of the mucosal wave; the regularity of the vocal fold vibratory pattern as evaluated in stroboscopic slow motion; and, in the presence of a lesion, its size and its impact on vocal fold vibration. For all scales, a minimum score of zero represented normality, while a maximum score of 100 represented extreme abnormality.

Statistical analysis
The group differences between the post-therapy and baseline data were tested for significance by analyses of variance for each parameter of the perceptual ratings and of the logarithmic transformed acoustic parameters. These data were sufficiently normally distributed to allow analyses of variance. The group differences of the visuo-perceptual parameters of the laryngostroboscopic recordings were tested for significance by means of Wilcoxon matched pairs signed-ranks tests. For the analysis of the perceptual data, the rater's mean values before and after therapy were used. Additionally, the differences between baseline pairs and pairs with one baseline sample and the post-therapy sample were evaluated. All computations were performed using SPSS 8.0.0 (SPSS Inc., Chicago, Illinois).

Apart from the tests of group effects, individual therapy effects were investigated. For each patient, the difference between the mean baseline and post-therapy data was classified according to the following scheme. If the pre- <u>and</u> post-therapy scores on a parameter were both below the value that is generally used as a upper limit for a normal voice (norm value), the result was classified in group 1. Next, the patient's post- versus pre-therapy difference was tested for significance using the

90% confidence interval of this difference. Patients for whom no significant difference was found were classified in group 2. If a score decreased significantly (denoting improvement), the patient was classified in group 3. When a significant increase (denoting deterioration) was found, the patient was classified in group 0. For the perceptual data, an analysis of variance per patient was applied on the baseline and post-therapy mean ratings to test the differences in ratings between these moments of measurement for significance higher than 5% one sided. For the acoustic parameter values, the standard deviations for the 90% confidence intervals were estimated using the pooled variance among the patients for the nine measurements (three /a/'s on three different days) made during the baseline period. The acoustic parameters were first divided by the patient's mean value of the pre-therapy data in order to obtain a relative error (coefficient of variation). The standard deviation of the evaluation of the laryngeal images was derived from repeated assessments by the senior otolaryngologist . The norm values for the perceptual data were computed by doubling the standard error of the pre-therapy data for the rating averaged over five listeners. The acoustic norm values were given by the upper limit of normality as described in the manual of MDVP increased by twice the standard deviation computed for the given norm values.

The relationship between the different evaluation scales was studied by factor analyses based on nonparametric Spearman's rank-order correlation coefficients. Relations among the baseline data as well as among effect data (post-therapy minus pre-therapy data) were determined. The associations between the baseline data and the therapy effects on all three evaluation tools were determined by nonparametric Spearman's correlation coefficients.

Results

Baseline measurements

In order to make sure that the effects demonstrated after therapy were the result of voice therapy, pre-therapy measurements were tested for significant differences or trends towards normality. A significant improvement over the successive baseline measurements could denote spontaneous recovery. An analysis of variance was used to test for significant differences between the three baseline measurements. The distributions of perceptual data were sufficiently normal, but the acoustic data needed logarithmic transformation. No significant differences were found between the perceptual baseline data nor between the acoustic baseline data (all $p>0.12$). Additionally, a test for a trend in grade and jitter over the successive baseline data was performed by fitting a linear regression line per patient over the grade or jitter data. One-sample t tests were applied to the patient's regression coefficients of the regression lines using zero as a test value. Again, no significant trends over the baseline measurements were found ($p>0.73$). These findings confirm the chronicity of the dysphonia. Thus, all changes detected after therapy will be regarded as the result of voice therapy.

Effects of therapy

- Perceptual evaluation

Table II contains descriptive statistics of the pre-therapy data derived from averages of responses obtained from five expert listeners. All patients were judged to demonstrate some degree of overall deviance in voice quality before the initiation of therapy. Two analyses of variance were performed on the data. First, we tested for differences between pre- versus post-therapy raw scale data. Then we tested for differences within the two types of pairs of voice samples. Both analyses showed significant improvement on the parameters of grade (overall deviance in voice quality) and roughness. Differences between the raw data were significant at $p<0.007$ and between the pairs of samples at $p<0.03$. Either of the analyses demonstrated any significant changes on the parameter of breathiness. The interaction terms (patient x moment of measurement) for all three parameters were highly significant ($p<0.001$). This indicates that the success of the therapy differed greatly among the individual patients.

Although statistically significant on grade and roughness, the panel of listeners estimated the perceptual changes on all three parameters to be very small on average. The mean decrease on grade, roughness, and breathiness was -2.4, -2.3, and -1.8 mm on a scale of 100 mm. The averaged differences between pairs of voice samples consisting of two baseline samples were indeed almost equal to zero (0.04). However, the differences between pairs of voice samples consisting of one baseline and one post-therapy sample were not very large either (-2.2). The correlation coefficients between the three perceptual parameters varied for baseline and effect data between +0.60 to +0.87 and +0.52 to +0.90. A factor analysis applied on the effects in perceptual parameters confirmed the relationship between the changes in the parameters attributed to therapy. In this analysis, only one factor was extracted explaining 74% of the total variance.

Table II also displays the magnitude of the effects of therapy. Data on subjects showing no deviant baseline data and no significant change after therapy (effect group 1) were excluded from the computations. Figure 1 gives an overview of individual therapy effects on overall deviance in voice quality as judged by the panel. The effect of therapy is plotted as a function of the pre-therapy data, while markers indicate the effect groups. The individual effects diverge widely and show no direct relation to the pre-therapy value.

The speech therapists who performed the therapies were asked to evaluate the effects of their therapy with the same perceptual instrument that the panel of expert listeners had used (Table III). This group of therapists was not specially trained in rating this instrument. In the judgment of the therapists, the voice abnormalities in the pre-therapy period were more severe than the panel had judged them to be. On the other hand, the therapists discerned greater differences between the pre- and post-therapy data. The therapists' median baseline and effect data were respectively twice and ten times as high as the data of the panel. Nonparametric correlations between the baseline scores awarded by the panel and the speech therapists were high for the parameters grade and breathiness ($R>+0.68$) but low for roughness ($R=+0.27$). The correlations between the effect scores were low (all $R<+0.17$).

Table II. Perceptual evaluation: baseline versus post-therapy data.
Descriptive statistics of baseline data, level of significance (*p*) for the comparison of differences between types of pairs of voice samples/rough baseline versus post-therapy data (effect data), numbers of patients in the effect groups, and descriptive statistics of effect data (effect group 1 excluded). The maximum score of the scales is 100. Effect groups: 0=significantly increased score; 1=normal at start and after therapy; 2=no significant difference; 3= significantly decreased score.

Dependant variable	Baseline data (N=73)		*p*<	Numbers in Effect group (N=73)				Effect data (excl. effect group 1)		
Perceptual evaluation	Median	25', 75' perc.		0	1	2	3	Median	25', 75' perc.	N
Grade	33	22 , 43	.03 / .007	9	0	43	21	-1.8	-4.5, 1.3	73
Roughness	24	13 , 35	.02 / .001	7	6	37	23	-2	-4.8, 0.1	67
Breathiness	28	16 , 38	ns / ns	7	3	45	18	-0.4	-4.8, 1.6	70

Table III Perceptual evaluation: speech therapists versus panel of expert listeners.
Descriptive statistics of baseline and effect data (for speech therapists and panel separately), and correlations coefficients (Spearman's rho) for the association between data of the speech therapists and the panel.

Dependant variable	Baseline data (N=60)					Effect data (N=60)				
Perceptual evaluation	Speech therapists		Panel		R	Speech therapists		Panel		R
	Median	25',75' perc.	Median	25',75' perc.		Median	25',75' perc.	Median	25',75' perc.	
Grade	63	28, 74	34	23, 49	.76	-16	-33, -3	-1	-4, 2	.13
Roughness	50	22, 77	25	14, 37	.27	-17	-38, -1	-2	-5, 1	.17
Breathiness	56	19, 76	27	16, 39	.68	-15	-42, -5	1	-2, 2	.00

Table IV. Acoustic analysis: baseline versus post-therapy data.
Descriptive statistics of baseline data, level of significance (*p*) of the difference between post minus pre-therapy data (effect data), numbers of patients in the effect groups, and descriptive statistics of effect data (effect group 1 excluded). Effect groups: 0=significantly increased score; 1=normal at start and after therapy; 2=no significant difference; 3= significantly decreased score.

Dependant variable	Baseline data (N=77)		*p*<	Numbers in Effect group (N=77)				Effect data (excl. effect group 1)		
Acoustic analysis	Median	25', 75' perc.		0	1	2	3	Median	25', 75' perc.	N
Jitt	1.7	1.1 , 3.1	.002	6	25	17	29	-0.8	-1.7, 0.2	52
Shim	5.4	4.6 ,7.6	.002	7	18	35	17	-1.0	-2.2, 0.5	59
NHR	0.15	0.13 , 0.18	.003	0	67	4	6	-0.07	-0.10, 0.048	10

- Acoustic analysis

Table IV presents the median and percentiles of the acoustic baseline data. After logarithmic transformation, an analysis of variance was performed to test for group differences between pre- and post-therapy data. For all three acoustic parameters, the results showed positive significant therapy effects (all $p<0.003$). A highly significant patient factor ($p<0.001$) confirmed the heterogeneity of the patient population. The interaction term (patient x moment of measurement) appeared to be highly significant ($p<0.001$) as well, indicating that the therapy effect differed among the individuals. In other words, not all patients showed the same degree of improvement.

When we exclude effect group 1, the median changes on Jitt, Shim, and NHR are -0.8 %, -1.0 %, and -0.065 (Table IV). The correlations among the three acoustic parameters for baseline data ranged from +0.71 to +0.90 and for effect data from +0.61 to +0.68. After performing factor analysis on the effect data, only one factor was extracted explaining 73% of the total variance, while communalities ranged from 0.68 to 0.82.

The number of subjects per effect group (Table IV) describes the diversity in therapy results at the individual level. Many subjects, especially in the case of the parameter NHR, showed pre-therapy data that were already lower than the norm value. For these subjects, no further improvement would be possible on the parameters concerned. Therefore, NHR seems to be a less suitable parameter for demonstrating the effects of therapy. Figure 2 shows the therapy effect for jitter as a function of the pre-therapy values. As correlations between the acoustic parameters are rather high, this picture applies to the results on shimmer as well. If data are abnormal before therapy, this abnormality was reduced after therapy to nearly normal values, except for some cases. Thus, subjects with more deviant acoustic baseline data improved generally more than those with data closer to the norm value at the beginning of the therapy. This effect differs from the therapy outcome of the perceptual evaluation, where no clear association between therapy effect and pre-therapy data could be demonstrated.

- Laryngostroboscopy

Table V contains descriptive statistics for the pre-therapy data for the visuo-perceptual evaluation. Not all aspects were abnormal in all patients. Depending on the parameter, between 7% and 32% of the patients had normal parameter values before and after therapy. A Wilcoxon matched pairs signed-ranks test was used to test for significant group differences between baseline and post-therapy data. All but one of the differences were significant at the $p<0.001$ level. The only exception concerned the parameter describing the regularity of the vibration of the mucous membranes, for which a significance level of $p<0.004$ was found. The median changes in ratings after therapy plus the 25^{th} and 75^{th} percentiles were determined over all patients except those who were rated as normal before and after therapy (effect group 1; see Table V).

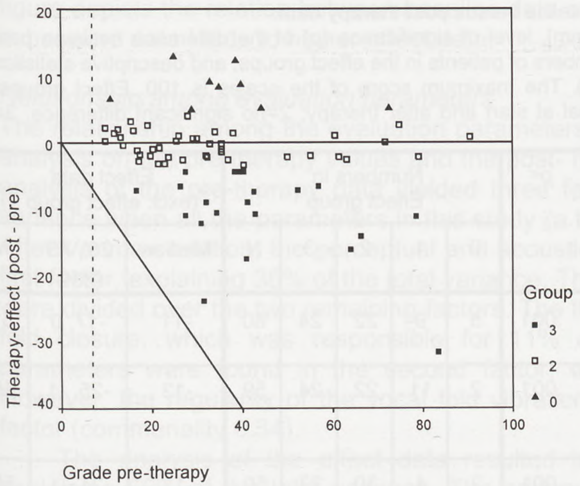

Figure 1. Individual therapy effects for Grade (Perceptual evaluation: the overall deviance in voice quality). The post- minus pre-therapy values are plotted against the mean baseline value for each patient. The oblique line represents the position of data for patients that show maximal improvement. Group numbers: see text.

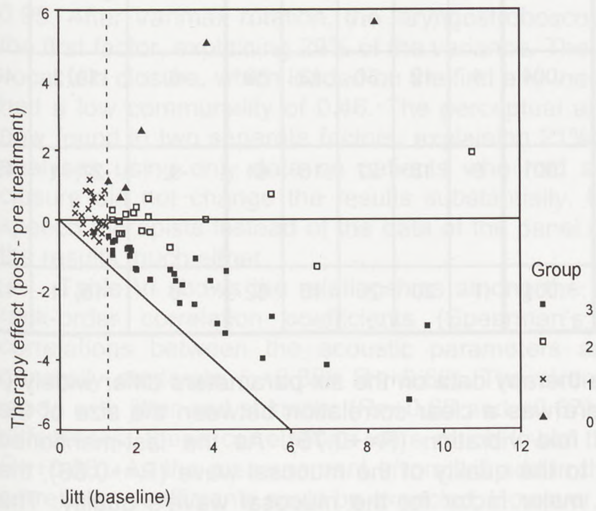

Figure 2. Individual therapy effects for Jitt (Acoustic analysis by MDVP: relative jitter). The post- minus pre-therapy values are plotted against the mean baseline value for each patient. The oblique line represents the position of data for patients that show maximal improvement. Group numbers: see text.

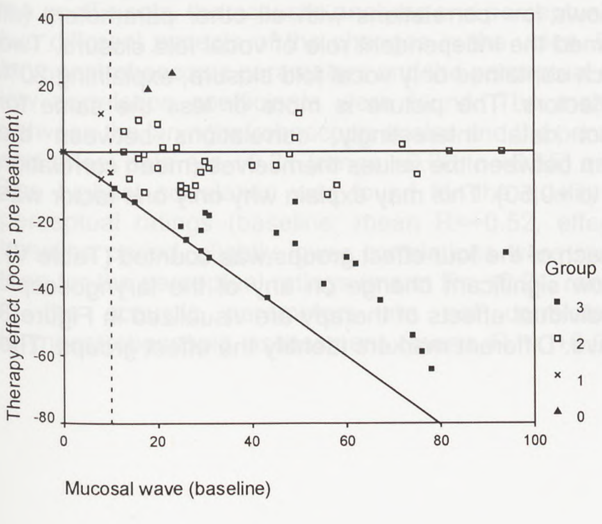

Figure 3. Individual therapy effects for the quality of the mucosal wave (Visual assessment of laryngostroboscopic recordings). The post- minus pre-therapy values are plotted against the mean baseline value for each patient. The oblique line represents the position of data for patients that show maximal improvement. Group numbers: see text.

Table V. Visuo-perceptual evaluation: baseline versus post-therapy data.
Descriptive statistics of baseline data [mm], level of significance (*p*) of the difference between post minus pre-therapy data (effect data), numbers of patients in the effect groups, and descriptive statistics of effect data (effect group 1 excluded). The maximum score of the scales is 100. Effect groups: 0=significantly increased score; 1=normal at start and after therapy; 2=no significant difference; 3= significantly decreased score.

Dependant variable Visuo-perceptual evaluation	Baseline data			*p*<	Numbers in Effect group					Effect data (excl. effect group 1)		
	Median	25', 75' perc.	N		0	1	2	3	N	Median	25', 75' perc.	N
vocal fold closure	20	13, 34	61	.001	5	9	22	24	60	-11	-17, 0	51
amplitude vocal fold vibration	21.5	12, 40	60	.001	2	11	22	24	59	-13	-25, -1	48
quality mucosal wave	29	17, 50	60	.001	2	4	30	23	59	-11	-24, -1	55
regularity vocal fold vibration	25	14, 43	61	.004	4	12	30	13	59	-6	-19, 2	47
impact lesion on vocal fold vibration	27	0, 42	61	.001	0	18	27	16	61	-9	-27, -1	43
size of lesion	20	0, 37	63	.001	1	20	26	15	62	-6	-15, -1	42

The correlations between the pre-therapy data on the six parameters differ widely ($-0.06<R<+0.75$). As expected, there was a clear correlation between the size of the lesion and its impact on vocal fold vibration ($R=+0.75$). As the last-mentioned parameter is also clearly related to the quality of the mucosal wave ($R=+0.66$), the existence of a lesion may be a major factor for the mucosal wave's quality. The degree of vocal fold closure shows low correlations with all other parameters (all $R<+0.27$). Factor analysis confirmed the independent role of vocal fold closure. Two factors are extracted, one of which contained only vocal fold closure, explaining 20% of the 68% explained by both factors. The picture is more or less the same for correlations between the effect data. Interestingly, correlations between the differences are slightly higher than between the values themselves (mean correlation coefficient increased from +0.35 to +0.50). This may explain why only one factor was found for the therapy effects.

The number of patients in each of the four effect groups was counted (Table V). Twenty-one patients did not show significant change on any of the laryngoscopic parameters after therapy. The individual effects of therapy are visualized in Figure 3 for the quality of the mucosal wave. Different markers identify the effect groups. The figure depicts the relation between baseline data and effect data: greater abnormality in baseline data leads to higher effect data.

Relationship among evaluation parameters
The relationship among the evaluation parameters was studied by performing factor analysis on the pre-therapy values and the post- minus pre-therapy differences. The analysis of the pre-therapy data yielded three factors describing 72% of the total variance when all the parameters in this study (a total of 12 parameters) were used. After Varimax rotation, the perceptual and acoustic parameters loaded highly on the first factor, explaining 36% of the total variance. The laryngostroboscopic parameters were divided over the two remaining factors. The third factor was the degree of vocal fold closure, which was responsible for 11% of the total variance. The other parameters were found in the second factor, explaining 25 % of the variance. However, the regularity of the vocal fold vibrations was scarcely described by this factor (communality 0.34).

The analysis of the effect data resulted in the extraction of three factors explaining 66% of the total variance. The communalities varied between 0.33 and 0.95. After Varimax rotation, the laryngostroboscopic parameters loaded mainly on the first factor, explaining 29% of the variance. The only exception was the degree of vocal fold closure, which loaded on the first and the second factor. But this parameter had a low communality of 0.46. The perceptual and the acoustic parameters were now found in two separate factors, explaining 21% and 19% of the variance. Factor analyses using only data on patients who had a lesion or insufficient vocal fold closure did not change the results substantially. Using the perceptual data of the speech therapists instead of the data of the panel of expert listeners did not change the results much either.

Table VI shows the relationships among the parameters in more detail, giving rank-order correlation coefficients (Spearman's rho) for all parameters. The correlations between the acoustic parameters and the perceptual ratings were generally moderate (+0.38< R<+0.68). The strongest correlations were found for grade with jitter and shimmer (R=+0.68 and +0.67). For the post- minus pre-therapy differences, these correlations were much lower than for the pre-therapy data (all R<+0.38). As the measurement errors will add in the differences, a reduction of the correlation coefficients could be expected. However, in view of the strong decrease of the coefficients, the acoustic analysis and perceptual evaluation apparently represent two different aspects of the changes in the voice. For the correlations between the laryngostroboscopic parameters and the perceptual evaluation and acoustic analysis, low correlation coefficients were found. The mean coefficients for the relations between the laryngostroboscopic scales and the perceptual evaluation for pre-therapy and effect data are +0.31 (range +0.06 to +0.55) and +0.18 (range +0.07 to +0.33). The highest correlation was found for the quality of the mucosal wave and the perceptual ratings (baseline: mean R=+0.52, effect: mean R=+0.30). During the baseline period, slightly lower correlations were found for the acoustic parameters than for the perceptual ratings (mean R= +0.24, range -0.09 to +0.47). The changes in the acoustic parameters are not correlated with the changes in the laryngostroboscopic assessment (mean R = +0.01, range -0.18 to +0.23). The correlations indicate that the visual evaluation of the vibration of the vocal folds is more strongly related to the perceptual ratings than to the acoustic parameters.

Table VI. Correlation coefficients (Spearman's rho) for the perceptual evaluation, the acoustic analysis and the videolaryngostroboscopy. The figures on the white background refer to correlations between baseline data. The figures on the gray background refer to correlations between effect scores (post- minus pre-therapy data). The numbers of patients involved are in parentheses.

Dependant variable	Perceptual evaluation			Acoustic analysis			Videolaryngostroboscopy					
	Grade (G)	Roughness (R)	Breathiness (B)	Jitt	Shim	NHR	vocal fold closure	amplitude vocal fold vibration	quality mucosal wave	regularity vocal fold vibration	impact of lesion on vocal fold vibration	size of lesion
G		0.70** (73)	0.90** (73)	0.29* (72)	0.21 (72)	0.16 (72)	0.33* (56)	0.23 (55)	0.29* (55)	0.20 (55)	0.18 (57)	0.16 (58)
R	0.85** (73)		0.52** (73)	0.26* (72)	0.14 (72)	0.19 (72)	0.26 (56)	0.12 (55)	0.07 (55)	0.20 (55)	0.11 (57)	0.11 (58)
B	0.87** (73)	0.60** (73)		0.38** (72)	0.27* (72)	0.23 (72)	0.30* (56)	0.14 (55)	0.25 (55)	0.13 (55)	0.11 (57)	0.08 (58)
Jitt	0.68** (72)	0.56** (72)	0.67** (72)		0.61** (77)	0.68** (77)	-0.04 (60)	0.00 (59)	0.01 (59)	0.03 (59)	-0.16 (61)	-0.17 (62)
Shim	0.67** (72)	0.57** (72)	0.58** (72)	0.79** (77)		0.63** (77)	0.11 (60)	0.14 (59)	0.23 (59)	0.18 (59)	0.00 (61)	-0.07 (62)
NHR	0.50** (72)	0.46** (72)	0.38** (72)	0.71** (77)	0.90** (77)		-0.18 (60)	0.05 (59)	0.10 (59)	0.08 (59)	-0.03 (61)	-0.15 (62)
VF closure	0.17 (57)	0.12 (57)	0.33* (57)	0.21 (61)	-0.02 (61)	-0.09 (61)		0.33* (58)	0.47** (58)	0.43** (58)	0.36** (59)	0.34** (60)
amplitude VF vibrat.	0.36** (56)	0.23 (56)	0.36** (56)	0.26* (60)	0.23 (60)	0.26* (60)	-0.06 (60)		0.72** (59)	0.55** (59)	0.58** (58)	0.44** (59)
mucosal wave	0.55** (56)	0.44** (56)	0.58** (56)	0.47** (60)	0.41** (60)	0.44** (60)	0.14 (60)	0.70** (60)		0.56** (59)	0.62** (58)	0.55** (59)
regularity VF vibrat.	0.27* (57)	0.41** (57)	0.20 (57)	0.25 (61)	0.22 (61)	0.22 (61)	0.06 (60)	0.30* (60)	0.49** (60)		0.49** (58)	0.40** (59)
impact lesion on VF vibrat.	0.41** (57)	0.31* (57)	0.54** (57)	0.36** (61)	0.34** (61)	0.31* (61)	0.27* (60)	0.42** (59)	0.66** (59)	0.40** (60)		0.64** (60)
size of lesion	0.11 (59)	0.06 (59)	0.24 (59)	0.16 (63)	0.12 (63)	0.15 (63)	0.16 (61)	0.30* (60)	0.49** (60)	0.22 (60)	0.75** (60)	

** Correlation is significant at the 0.01level (2-tailed);
* Correlation is significant at the 0.05 level (2-tailed).

As mentioned above, factor analysis of therapy effects yielded three factors. That means that three kinds of changes could be found independent of each other. These changes appeared to be represented by the three evaluation methods (perceptual evaluation, acoustic analysis, laryngostroboscopy). Therefore, a patient will not usually show improvement on all of the evaluation parameters. Only four patients showed significant improvement on more than eight parameters out of a total of 12. On the other hand, the majority of the patients (64) showed significant improvement on at least two parameters. For each evaluation method separately, we determined how many patients showed significant improvement (effect group=3) on at least one of the parameters of that specific instrument (Table VII data on the diagonal). The off-diagonal data in table VII show the number of patients with a significant improvement in two instruments. That means that for at least one of the parameters of the first evaluation instrument the patient was classified in effect group 3 and that the same was true for the second evaluation instrument. The number of patients with such a significant improvement found in two evaluation methods is about half the number as when counting the number of patients in effect group 3 of one method. This number could be expected approximately when the classifications of patients in effect group 3 for the three evaluation methods are not related. This result confirms the independence of the aspects represented by the three evaluation methods at the level of therapy success of the individual patient.

Table VII. Number of patients with a significant improvement (e-group=3) in at least one parameter of one evaluation method (diagonal) or in at least one parameter of two evaluation methods (off-diagonal). Percentages are shown for the patients within one method.

	Perceptive	Acoustic	Stroboscopic
Perceptive	31 (42%)	15	16
Acoustic		33 (42%)	17
Stroboscopic			40 (51%)

Patient-related parameters

Subjects were grouped on the basis of the patient-related factors: age, gender, or laryngeal diagnosis. Using nonparametric correlations or Mann-Whitney *U* tests, we tested the data for significant differences. Spearman's correlation coefficients showed that the age of the patient was not related to either the mean pre-therapy data or the magnitude of the therapy effects for all parameters (both mean $R=+0.10$; $-0.22<R<+0.22$, and $+0.09<R<+0.20$), except for the acoustic parameter NHR ($R=+0.36$). In case of NHR, the older subjects showed slightly higher baseline values. As for the factor of gender (Mann-Whitney *U* test), male subjects showed higher pre-therapy values for Shim and NHR ($p=0.010$ and $p=0.003$), while female subjects had more pre-therapy lesions ($p=0.005$). However, no clear significant differences in therapy effects were found (Mann-Whitney *U* test, all $p>0.042$).

The results of statistical tests within the diagnostic categories are difficult to compare due to the large variation in the number of patients assigned to each category. However, the distribution of post- minus pre-therapy changes in the categories showed some clear differences. For the group of patients with slight vocal fold abnormalities, significant improvement was found for nearly all parameters. The group of patients with vocal fold nodules showed significant success on the perceptual and laryngostroboscopic parameters. All other diagnostic groups showed less or no improvement after therapy. For the perceptual ratings, there was a trend for the group of patients with unilateral vocal fold paralysis and submucosal swelling to deteriorate slightly after therapy. The therapy effects on the laryngostroboscopic parameters show, on average, no change in patients with unilateral vocal fold paralysis (none of them showed recovery from paralysis) and slight deterioration in patients with muscle tension dysphonia. Positive effects were found for the other diagnoses. Patients with muscle tension dysphonia and severe vocal fold abnormalities were found to have a particularly high variation in therapy effects.

Discussion

Effects of voice therapy

This study of changes on parameters of voice quality and laryngostroboscopy reveals rather modest effects of voice therapy in chronically dysphonic patients. Nevertheless, the statistical analysis for the group effects revealed that overall improvement was highly significant. However, the effects could be very diverse at the individual level. A minimum of at least one significant improvement on one of the three perceptual or one of the three acoustic parameters was achieved by 42% of the patients. The rate of success on at least one of the six stroboscopic parameters was 51%. This outcome is slightly higher than the effects found for the perceptual and acoustical analysis, but the number of parameters involved was higher as well. There was no significant relation between laryngeal diagnosis and therapy effect, although some tendencies in therapy effects could be determined. For instance, patients with unilateral vocal fold paralysis and muscle tension dysphonia showed poor improvement. The modest results may be ascribed to the restriction to include only patients with chronic symptoms. It may be expected that chronic abnormalities recover with more difficulty than dysphonic problems that have not been present over such a long period of time. On the other hand, factors that were not taken into account, such as motivation or smoking habits, may have greater bearing on the outcome of therapy than those that were actually included in the study.

Differences among evaluation methods

We could only demonstrate marginal improvement in the aspect of perceptual voice quality, whereas we found more distinct changes in the laryngostroboscopic and acoustic parameters. Although the number of patients with a significant improvement after therapy (effect group 3) was quite similar for the three evaluation instruments, the median improvement was different. The median decrease on the parameter Grade was less than 1 on a scale from 0 to 100, whereas this decrease for Jitt was -0.8 on a scale from 0 to approximately 10 and the mean decrease on the laryngostroboscopic parameters was about 9 on a scale from 0 to 100.

Often, only one or two of the three evaluation instruments are used to assess the success of a therapy. However, the results of these methods for individual patients are not identical and therefore the degree of therapy success depends on the instrument used. The differences between the methods might be related to the different types of voicing on which the results are based. The mid-sections of the sustained vowels were used for the acoustic analysis, and the laryngostroboscopic recordings were made during sustained phonation. The perceptual evaluation, however, was made on a fragment of a reading text. The use of the relatively stable fragment, excluding vowel onset and offset, ignores irregularities that may occur more frequently in running speech. This may explain why prior to therapy, many subjects showed no abnormality or a very low amount on the acoustic parameters. However, according to de Bodt[18], pathological conditions of the voice will be considered more severe when the perceptual evaluation is based on sustained vowels rather than on running speech. On the other hand, Revis et al.[19] stated that judgments on stabilized sustained vowels were confirmed as less severe than judgments on connected speech. In literature, various statements can be found on the use of diverse types of voice samples in voice assessment. However, as the evaluation instruments require different voice samples, this problem cannot be avoided. A reading text, for example, may be considered to approach natural speech and, therefore, be used for perceptual evaluation. During laryngostroboscopic recording, however, a stable phonation is required.

In our study, voice therapy has more apparent effects on the laryngostroboscopic parameters than on perceptual alterations, which is also the case for acoustic changes. Both findings show that improvements in the vibratory mechanism of the vocal folds do not always go hand in hand with perceptual changes of voice quality in running speech. Correlations between the perceptual evaluation instrument and the other two methods are rather low. Subjects having normal values on acoustic or videostroboscopic parameters may demonstrate deviant data on the perceptual evaluation scales. The acoustic and laryngostroboscopic parameters describe improvements in the stationary vibration of the vocal folds, changes that are apparently not always evident to the listener. The hearing of voice in terms of voice quality perception is the outcome of a much more complex process, involving both speaker and listener.

Voice therapy evaluation methods

When the aim of voice therapy is restricted to the improvement of communication, perceptual rating may be considered as the only meaningful outcome measure. Indeed, the perceptively abnormal voice is the patient's main complaint and his or her major handicap in communication. Therefore, improvement of voice quality will be the main objective of voice therapy. However, in several individual patients, clear improvements were found in laryngostroposcopic or acoustic parameters and not in the perceptual ratings. These findings indicate that in those patients an abnormality in the vocal fold vibration reduced, without a better perceptual rating. When therapy outcome is based on perceptual improvement only, this study suggests that voice therapy for these chronically dysphonic patients may not be indicated. However, especially in a medical setting, physiological improvements will be considered important as well. Without offering therapy to these patients, the beneficial effects on the physiological aspects of the vocal fold and on the vocal fold biomechanics would be withheld from them.

Judgments made by speech therapists versus panel of expert listeners
In this study, speech therapists gave a noticeably more positive evaluation of therapy outcome than an independent panel of expert listeners did. The fact that the therapists judged the situation before therapy to be more severe than the panel, indicates that the therapists assigned a lower severity level to the highest point of the scale than the panel did. A change towards a normal voice will therefore be larger for the therapist than for the panel. However, the panel's assessment, being based on a fragment of a reading text, may not be comparable to that of a therapist, who has assessed the patient's voice under many different circumstances and possibly in more natural conversation. Furthermore, the panel was offered voice samples at random per patient, whereas the speech therapist was fully aware of the moment of measurement. In view of these differences, the results emphasize that evaluation by the perceptual assessment of the voice by therapists themselves should be interpreted with caution. An essential methodological aspect for this effectiveness-study is that the evaluation occurs without the rater's awareness or even suspicion whether the rated material is pre- or post-treatment. This requires a time-consuming preparation of the auditory and visual material that will be rated, but guarantees the validity of the conclusions.

Multidimensionality of voice
Not all patients showed abnormal pre-therapy data on all the evaluation methods, nor did all patients show significant improvement by all the methods. The overall positive effects were significant, but the individual benefits of therapy were highly diverse. The majority of the patients showed significant improvement as measured by one or two of the methods but not on all parameters. Due to the independence of the therapy successes according to the three methods, a patient can show, for example, improvement upon laryngostroboscopic examination without a better perceptual impression of the voice. Therefore, in order to evaluate the overall effect of voice therapy, all dimensions must be considered[20]. This study covered only three evaluation methods: perceptual evaluation, acoustic analysis, and laryngostroboscopy. It might be useful to incorporate other instruments as well. As the individual effects tend to vary widely, subjects showing no benefit from therapy on perceptual, acoustic, or laryngostroboscopic parameters may improve in other dimensions of voice. The Committee on Phoniatrics of the European Laryngological Society[21] recommends a multidimensional minimal set of basic measurements for the assessment of voice pathology, using perception, videostroboscopy, acoustics, aerodynamics, and subjective rating by the patient.

Conclusions

- Perceptual evaluation of pre- and post-therapy voice quality of chronically dysphonic patients showed as a group a significant improvement for the parameters Grade and Roughness. However, the average changes were rather small. The effects of the therapy considerably differed among the individual patients. No relation was found between the magnitude of the observed change, and the pre-therapy degree of deviance. A randomized rating of the material without reference to the moment of recording appeared to be essential for a valid evaluation, as the therapists clearly tended to overestimate the favorable change in voice quality. The therapist's opinion is probably influenced by other factors then only voice quality.

- For all three acoustic parameters (Jitter, Shimmer and Noise-to-Harmonics Ratio), the results showed significant positive therapy effects for the whole group of patients. The therapy effects considerably differed from one patient to another. Subjects with more deviant acoustic baseline data improved more than those with data closer to the norm value at the beginning of the therapy.
- Ratings of laryngostroboscopic recordings without the rater's awareness whether the recordings are pre- or post-treatment showed clear significant favorable group differences between baseline and post-therapy data. Just as found for the acoustic measures, greater abnormality in baseline data corresponded with a greater therapy effect.
- The most distinct therapy results were found in patients with vocal fold nodules and slight vocal fold abnormalities. Patients with muscle tension dysphonia and severe vocal fold abnormalities had a particularly high variation in therapy effects.
- As not all patients show abnormal pre-therapy data on all the evaluation methods, and not all patients show significant improvement by all the methods, multidimensional assessment is to be recommended for evaluation of voice therapy effects.

References

1. Koufman JA. & Blalock PD. Functional voice disorders. *Otolaryngol.Clin.North Am.* 1991;24:1059-1073.
2. Raabe J. & Pascher W. Das Reinke-Ödem: eine Untersuchung zu Fragen der Ätiologie, der Prognose und der Wirksamkeit therapeutischer Interventionen. *Laryngorhinootologie.* 1999;78: 97-102.
3. Fex B, Fex S, Shiromoto O & Hirano M. Acoustic analysis of functional dysphonia: before and after voice therapy (Accent method). *J.Voice* 1994;8:163-167.
4. Holmberg EB, Hillman RE, Hammarberg B, Södersten M & Doyle P. Efficacy of a behaviourally based voice therapy protocol for vocal nodules. *J.Voice* 2002;15:395-412.
5. Verdolini MK, Burke MK, Lessac A, Glaze L & Caldwell E. Preliminary study of two methods of treatment for laryngeal nodules. *J.Voice* 1995;9:74-85.
6. Kitzing P & Åkerlund L. Long-time average spectrograms of dysphonic voices before and after therapy. *Folia Phoniatr.Basel.* 1993;45:53-61.
7. Murry T & Woodson GE. A comparison of three methods for the management of vocal fold nodules. *J.Voice* 1992;6:271-276.
8. Heuer RJ, Sataloff RT, Emerich K, Rulnick R, Baroody M, Spiegel JR, Durson G & Butler J. Unilateral recurrent laryngeal nerve paralysis: the importance of "preoperative" voice therapy. *J.Voice* 1997;11:88-94.

9. Roy N, Bless DM, Heisey D & Ford CN. Manual circumlaryngeal therapy for functional dysphonia: an evaluation of short- and long-term treatment outcomes. *J.Voice* 1997;11:321-331.
10. Bassiouny S. Efficacy of the accent method of voice therapy. *Folia Phoniatr.Logop*. 1998;50:146-164.
11. Kotby MN, El-Sady SR, Basiouny SE, Abou-Rass YA & Hegazi MA. Efficacy of the accent method of voice therapy. *J.Voice* 1991;5:316-320.
12. Carding P. *Evaluating voice therapy: measuring the effectiveness of treatment.* London Philadelphia: Whurr Publishers, 2000.
13. Gerratt BR, Kreiman J, Antonanzas-Barroso N & Berke GS. Comparing internal and external standards in voice quality judgments. *J.Speech Hear.Res.* 1003;36:14-20.
14. Kreiman J, Gerratt BR & Berke GS. The multidimensional nature of pathologic vocal quality. *J.Acoust.Soc.Am*. 1994;96:1291-1302.
15. Speyer R, Wieneke G H, Ghazi Hosseini E, Kempen PA, Kersing W & Dejonckere PH. Effects of voice therapy as objectively evaluated by digitized laryngeal stroboscopic imaging. *Annals of Otology, Rhinology & Laryngology.* 2002;111:902-909.
16. Hirano M. *Clinical examination of voice.* Wien: Springer-Verlag, 1981.
17. Titze I R. Workshop on acoustic voice analysis: summary statement. *Denver. Conference Proceeding* 1995.
18. Bodt de M. *Een onderzoeksmodel voor stemevaluatie: de relatie tussen subjectieve en objectieve parameters in de beoordeling van de normale en pathologische stemfunctie.* Doctoral dissertation. University of Antwerpen:1997.
19. Revis J, Giovanni A, Wuyts F & Triglia J. Comparison of different voice samples for perceptual analysis. *Folia Phoniatr.Logop* 1999;51:108-116.
20. Hirano M. Objective evaluation of the human voice: clinical aspects. *Folia Phoniatr.* 1989;41: 89-144.
21. Dejonckere PH, Bradley P, Clement P, Cornut G, Crevier-Buchman L, Friedrich G, Heyning van de P, Remacle M &Woisard V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur.Arch.Otorhinolaryngol.* 2001;258:77-82.

Chapter 6

Self-assessment of voice therapy for chronic dysphonia

R. Speyer, G.H. Wieneke, P.H. Dejonckere
Accepted (Clinical Otolaryngology)

Self-assessment of voice therapy for chronic dysphonia

Abstract

The effects of voice therapy in a group of chronically dysphonic patients are determined using two quality-of-life instruments: the Voice Handicap Index or VHI (Jacobson et al.[1]), and a simple three-item outcome scale (three visual analog scales). Both instruments measure changes in the quality of the voice itself and in the extent of impairment resulting from the dysphonia as experienced by the patient in social and occupational settings. Statistical tests conducted on pre- and post-treatment data indicated significant improvements on both instruments for the group as a whole. At the individual level, however, the effects were diverse. For roughly 50% of the subjects a significant improvement could be established.

The positive changes as measured with the three-item scale were greater than those measured with the VHI. The results suggest that the two quality-of-life instruments measure slightly different aspects of the subjective perception of the therapy effects. In order to obtain a general evaluation of the patient's handicap, it may suffice to ask some simple questions.

Introduction

The outcome of therapy is increasingly evaluated in terms of the patient's quality of life. Besides seeking improvement in the patient's vocal fold vibration and voice quality, the clinician expects voice therapy to help the patient function better at work and in social settings.

One way to evaluate improvements in quality of life (QOL) is simply to ask patients to rate their sense of well-being before and after therapy and to quantify the change in their impairment. Of course, their answers may be strongly influenced by certain aspects of daily living, thereby reflecting the respondent's emotional perception of the situation. A questionnaire about specific aspects of daily living would give a more objective picture of the changes in a patient's situation. In that vein, several quality-of-life instruments have been developed to quantify how dysphonia impacts a patient's psychosocial functioning. Notable examples are the following: the Voice Handicap Index developed by Jacobson et al.[1], the Voice Outcome Survey by Gliklich et al.[2], the Voice-Related Quality-of-Life instrument by Hogikyan and Sethuraman[3], and the Outcome Scale by Casper[4].

The clinician might suspect that asking a few simple questions would produce a different picture than administering a detailed questionnaire. If so, it would be inappropriate to use the results of one of these methods to validate those of the other. Instead, the two methods should be construed as complementary, capable of shedding light on different aspects of the patient's quality of life. Following that train of thought, an investigation of the effects of voice therapy on chronically dysphonic patients was carried out at the Institute of Phoniatrics (University Medical Center Utrecht). The study used a multiparametric approach, including two self-assessment instruments: the Voice Handicap Index, and a simple three-item outcome scale. The latter quantifies the severity of the disorder and the extent of impairment resulting from the dysphonia as perceived by the patients. Their answers give insight into the relationship between these two methods of measuring quality of life. The ratings of self-perception of the disorder can deviate from the perceptual impression of the disorder by other people, and from objective parameter for voice quality or normality of vocal fold vibration. Therefore, the relation of the self-assessment scales and perceptual, acoustic and laryngostroboscopic parameters measured at the same occasions was studied as well.

The Voice Handicap Index (VHI) has recently been applied to patient groups comparable to those in our study. For instance, it was used to compare data gathered before and after medialization laryngoplasty in patients with unilateral laryngeal paralysis (Billante et al.[5]; Spector et al.[6]). It was also used to evaluate two approaches to treatment for teachers with voice disorders (Roy et al.[7]), and to evaluate the results of vocal fold surgery, voice therapy, and/or medical therapy (Rosen et al.[8]). The last-mentioned study included patients suffering from unilateral vocal fold paralysis, vocal cyst or polyp, or muscle tension dysphonia.

The present study has two objectives. First, we seek to determine the effects of voice therapy practiced by speech therapists for patients with a chronic voice disorder. To that end, we apply the two quality-of-life instruments to a group of patients suffering from chronic dysphonia, which they developed as a result of diverse voice pathologies. Secondly, we compare the two self-assessment instruments in order to determine their specific utility.

Methods

Subjects

In order to take part in this study, patients had to meet certain criteria. Their dysphonia had to be chronic. It must have started at least four months before visiting an ORL specialist/phoniatrician at the Phoniatric Department of the University Hospital Utrecht, and the diagnosis must have been confirmed by phoniatric examination. Persons under 18 years of age were excluded because of the possibility that voice maturation or mutation might be involved. Other exclusion criteria were coexisting speech or language disorders, malignant or pre-malignant lesions, hormonal voice disorders, laryngeal papillomas, substitution of voice after laryngectomy, spasmodic dysphonia, and psychogenic dysphonia.

In total, 77 subjects were included in the study. Table 1 shows the distribution of etiologic categories as diagnosed by the phoniatrician. There were 34 men and 43 women in the study, ranging in age from 18 to 76 years. The average age was 40 years for the female subjects and 47 for the males. In a few cases, it was a trial therapy or a preliminary intervention prior to phonosurgery.

The range of etiologies for these subjects reflects the diversity of laryngeal pathology in chronic dysphonia as found in the clinical setting. Thus, this sample may be used here to evaluate the overall effects of voice therapy in clinical practice. The chronic nature of the dysphonia was confirmed by acoustic and perceptual analyses performed on speech samples that were recorded before starting therapy (i.e., in the baseline period). The recordings were made on three separate days within a six-week period. We did not find any significant differences between these baseline measurements. That is, experienced raters did not find any systematic disparities when conducting blinded perceptual ratings of the severity of the voice disorder. Nor systematic differences in jitter on recorded voice samples of sustained phonation of the vowel /a/ were found.

Table 1. Distribution of patients by diagnostic category.

Phoniatric diagnosis	N =77
Muscle tension dysphonia	12
Submucosal swelling	7
Vocal fold nodules	10
Vocal fold polyps	6
Unilateral vocal fold paralysis	7
Other: slight vocal fold abnormalities	24
Other: severe vocal fold abnormalities	11

Quality-of-Life instruments

This study applies and compares two self-assessment instruments: the Voice Handicap Index (Jacobson et al.[1]), and a simple three-item visual analog outcome scale.

The Voice Handicap Index is a 30-item questionnaire on the perception of impairment/handicap. It consists of a functional, a physical, and an emotional subscale, each with ten items (translated into Dutch by C.J. van As). The statements reflect the variety of experiences that a patient with a voice disorder may encounter. The patient responds by rating each statement on a scale of 0 to 4: never, almost never, sometimes, almost always, and always. When the points thus awarded are summed up, the total score can lie anywhere between zero and 120.

The three-item outcome scale (see Appendix I) quantifies the severity of the voice disorder and the extent of impairment experienced by the patient. The first visual analog scale (Item 1) refers to the overall severity of the disorder in terms of its effect on voice production. The second scale (Item 2) refers to the psychosocial impact of the disorder on one's occupational activities (if relevant). The third scale (Item 3) refers to the impact on daily living. Retirees, students, and persons not engaged in regular occupational activities were allowed to skip the item on work-related impairment. A score of zero indicates normal voicing conditions, while a score of 100 (the maximum) indicates extreme voice impairment or a severe handicap.

Perceptive evaluation, acoustic analysis and laryngostroboscopic examination

Audio recordings were made of each patient's voice. The subjects were asked to produce three sustained phonations of the vowel /a/, each lasting several seconds, at a comfortable pitch and loudness. Then the subject's voice was recorded while reading a text aloud.

The midsection of the reading text (about 10 seconds) was used for perceptual evaluation. The overall severity of the voice disorder (Grade) was rated on a visual analog scale. The scores ranged from 0 for a normal voice to 100 millimeters for an extremely severe disorder.

Acoustic analysis was performed on the sustained phonation using the Multi-Dimensional Voice Program (MDVP, Kay Elemetrics Corporation). The first and last segments of the phonation, each 200 ms, were removed. For each voice sample, we calculated the percentage of jitter (Jitt%). This gives an indication of the variability of the pitch period within the sample. It represents the relative period-to-period variability.

Laryngostroboscopy was performed using a rigid or flexible endoscope. Recordings on videotape (U-matic videocassette recorder VO-5630, Sony) were made of the vocal fold vibration during repeated stable phonation of a sustained vowel /a:/ or /i:/ at a comfortable pitch and loudness level using normal as well as stroboscopic light (Stroboscope RLS 9100, Kay Elemetrics Corp, Pine Brook, New Jersey). In addition, images of the vocal folds in a resting position were recorded. Visual analogue scales were used to evaluate the recordings on six parameters: the degree of insufficient vocal fold closure; the maximum amplitude of the vocal fold vibration; the quality of the mucosal wave; the regularity of the vocal fold vibratory pattern as evaluated in stroboscopic slow motion; and, in the presence of a lesion its impact on vocal fold vibration.

Procedure

If the phoniatrician found voice therapy to be indicated, the patients were referred to a speech therapist near home. The referrals included trial therapy and pre-surgical voice therapy. Treatment appropriate to a patient's specific needs and voicing possibilities was given over a period of at least three months for 30 minutes twice a week or 60 minutes once a week. If necessary, the patients were given a second series of sessions over a three-month period for either 30 or 60 minutes per week.

The subjects completed the form for the Voice Handicap Index before starting therapy and then after it was all over. They filled in the three-item outcome scale on three occasions over a period of six weeks before starting voice therapy. Furthermore, a retrospective scaling was made during the first visit to the clinic; at that point, the patient evaluated the situation of a month earlier. Together, these four scalings constitute the baseline measurements. They provide a point of reference for evaluating the effects of therapy on an individual. After finishing voice therapy, the patient filled in the three-item scale once more. At the time the three baseline measurements were made, audio recordings were made as well; these were to be subjected to acoustic and perceptual evaluation.

Statistical analysis

The distribution of the data from the three-item self-evaluation forms was sufficiently normal to allow us to perform analyses of variance to test for differences. An analysis of variance was used to test for significant differences between the baseline observations on the three-item scale as well as the differences between the baseline and the post-therapy observations. The therapy outcomes compiled in the Voice Handicap Index had to be tested nonparametrically by means of a Wilcoxon matched-pairs signed rank test.

Besides testing for group effects, we also investigated the effects of therapy on individual patients. For each parameter, we classified the patients into four groups according to the degree to which the treatment was considered successful. Concretely, the classification was based on the difference between the pre- and post-therapy scores. If both observations were within normal limits, the patient was placed in group 1. Otherwise, we carried out a statistical test to determine whether the difference deviated significantly from zero. If not, we concluded that the therapy had not affected the voice problem. In that case, the patient was placed in group 2. If a parameter value had decreased significantly, we concluded that the patient's situation had improved (group 3). If a parameter value had increased significantly, we concluded that the patient's situation had worsened (group 0).

For the three-item outcome scale, the ceiling for having no complaints (the normal limit) was set at 5 mm. A difference on this scale was assumed to be significant when the post- minus the pre-therapy values, divided by the subject's standard deviation on the three baseline scores (Z value), was above 1.96 (significance level 5% two-tailed).

For the VHI, the normal limit and the threshold for significant change were based on the values reported by Jacobson et al.[1]. This author found 95% confidence intervals of +/- 8 and +/- 18 points for the subscales and the total score. This means that when we find differences between post- and pre-treatment data of at least 8 points on the subscales and 18 points on the total score, we may consider the changes after therapy to be statistically significant. The same values were used to interpret the difference from a score of zero. That means that the normal limits were set at 8 and 18 for the subscales and the total score.

For the perceptual rating, the acoustic parameter and the laryngostroboscopic ratings, a similar classification was performed. The specific threshold values for these classifications can be found in Speyer[9] (2003)

The association between the two quality-of-life instruments as well as between the subscales per instrument is determined by nonparametric Spearman's correlation coefficients. Furthermore, factor analyses based on Spearman's correlations were applied to the baseline observations as well as to the effect data (post- minus pre-treatment data).

Results

Voice Handicap Index

Table 2 presents, among other things, descriptive statistics for the baseline data. The maximum score that can be attained on the subscales is 40, while the maximum total score is 120. The scores for some patients were rather high, especially on the physical scale.

Table 2 also gives the median and percentiles of post- minus pre-treatment differences. A Wilcoxon matched-pairs signed rank test was performed in order to test for significant differences between these observations. The overall tests showed significant improvement at all scales. The differences were significant at the level of $p<0.001$, except for the emotional subscale, which showed significant results at a level of $p<0.05$. The correlations between the effects on the different subscales were above 0.55. Although the overall level of significance is high, the improvement for the individual patient as expressed in the number of subjects per effect group is rather poor (Table 2). This is partly due to the high value set for the normal limit and the threshold for significant improvement, both values being derived from the literature (Jacobson et al.[1]). When the confidence interval of the total score was halved to +/- 9, the number of patients with a significantly lower score rose from 11 (22%) to 18 (31%).

Table 2.Voice Handicap Index: baseline versus post-therapy data.
Descriptive statistics of baseline data and of the differences in post- minus pre-treatment data and level of significance (*p*), numbers of patients in the effect groups. The maximum of the subscore scales is 40, that of the total score is 120. Effect groups: 0=significantly increased score; 1=normal at start and after therapy; 2=no difference; 3= significantly decreased score.

Dependent variable	Baseline data (N=69)		Post-minus pre-treatment data (N=52)		p<	Numbers in effect groups (N=52)			
	Median	25', 75' perc.	Median	25', 75' perc.		0	1	2	3
Physical subscore	17	12, 22	-4	-6, 1	0.001	0	7	36	9
Functional subscore	9	5, 15	-2	-5, 0	0.001	0	22	26	4
Emotional subscore	6	2, 13	-1	-5, 1	0.012	1	27	16	9
Total score	33	22, 48	-6	-14, 0	0.001	0	6	35	11

The mean improvement on the VHI scales is limited. Compared to the scale maximum, the decrease in rating lay between 3 and 9%. Considering only those subjects showing significant improvement, the median changes on the physical, functional, emotional, and total score were -14, -16, -12, and -31 points (-26 to -31% of the scale maximum).

One reason for the slight decrease in the VHI scores may be that many of the items within a (sub)score were not abnormal prior to therapy. Because the scores for these items therefore cannot decrease, the sum of the changes within a (sub)score will be lower. In the baseline period, on the average, each subject gave about half of the 30 items a score below or equal to one (mean number of items = 16.2, standard deviation 7.1). The median improvement among the patients using only the items with a higher rating (more than one) in the baseline was –11% to -17% of the scale maximum. This is higher than the –3% to –9% as found for all items. The significance of the pre- versus post-therapy values increased as well. When the mean of the items with a baseline value above one is used for each patient, the change in the total score is negative for all of them. With the original scores, a negative change was found for only 38 of the 52 subjects. The number of patients with a significant change in the total score (decrease more than 18 points) increased from 11 to 28. The VHI total score on the items with ratings of zero or one rose by 3.6 points (+3%). This change in a positive direction ($p<0.02$) may indicate a slight worsening of the impairment on these items.

The individual therapy effects are weakly related to the pre-therapy values. This is shown for the total score in Figure 1. The high ratings on the impairment at the start of therapy appeared to be only slightly related to a better outcome.

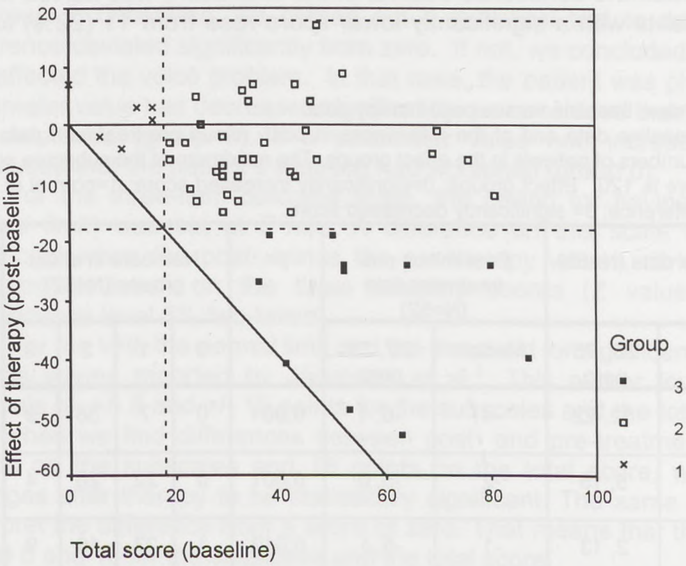

Figure 1. Individual therapy effects for the total score (VHI). The post- minus pre-therapy values are plotted against the baseline value for each patient with a complete set of VHI data. The oblique line represents the position of data for patients that show maximal improvement. Group numbers: see text.

Three-item outcome scale

Among other things, Table 3 presents descriptive statistics for the three-item outcome scale. An analysis of variance was used to test for significant differences among the four baseline values. The differences in severity were not significant, nor were those for impairment in daily living. However, the differences for the effect of the impairment on one's occupation did prove to be significant ($p=0.015$). There was a slight trend toward improvement; the successive means of the baseline data were 59, 56, 54, and 51. Yet taking the average standard deviation of about 25 into account, the change is small.

An analysis of variance was used to test the group differences between the post- and pre-therapy scores. The two factors were the patients and the moment of observations (pre- or post-therapy scores). For all three items, both factors (the effect of therapy and the patient) were significant at $p=0.000$. The average decrease in rating on the severity scale was 22 mm, on the scale for occupational impairment, it was 24 mm, and on the scale for impairment in daily living, it was 23 mm. The overall average standard deviation was 24.

Table 3 shows the number of patients in each effect group. In total, 19 patients did not show any significant change on any of the three parameters. This is partially due to their highly variable baseline scores. Regarding subjects with a significant positive therapy effect (effect group 3), the average improvement in the voice disorder severity, in impairment at work and in daily living was respectively 38, 40, and 43 mm on the 100mm scale.

Table 3. Three-item outcome scale. Descriptive statistics of the baseline data and the post- minus pre-treatment differences. Median and percentiles are given for comparison with the VHI data. The maximum of the scales is 100. Level of significance of these differences (*p*). Numbers of patients in the effect groups: 0=significantly increased score; 1=normal at start and after therapy; 2=no difference; 3= significantly decreased score.

Dependent variable	Baseline data			Post- minus pre-treatment data			p<	Numbers in effect group			
	Median	25',75 perc.	N	Median	25',75 perc.	N		0	1	2	3
Voice disorder severity	54	38, 67	77	-23	-35, -3	76	0.001	2	0	35	39
Impairment in occupation	61	42, 69	52	-19	-38, -8	51	0.001	1	0	26	24
Impairment in daily living	59	38, 75	77	-19	-42, -2	76	0.001	1	4	36	35

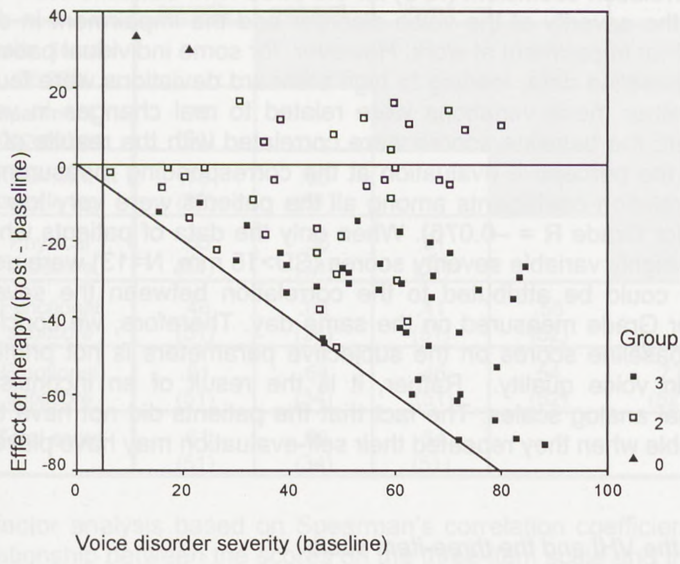

Figure 2. Individual therapy effects for the voice disorder severity scale (three-item outcome scale). The post- minus pre-therapy values are plotted against the mean baseline value for each patient. The oblique line represents the position of data for patients that show maximal improvement. Group numbers: see text.

The individual effects of therapy are visualized in Figure 2 for the overall severity rating. Post-therapy minus averaged pre-therapy data are plotted against the baseline scores. The distinguished groups of therapy effects are marked on the graph.

The therapy effects are more distinct for the three-item outcome scale than for the VHI scales. On the three-item outcome scales for the patients for whom the VHI pre- and post-therapy scores were available (N=52), the differences were -26 to -29%. These percentages are much greater than those for the raw VHI data (-3 to -9%) as well as for the values found when only the high scores are summed (-11 to -17%).

Patient-related factors

The age of the patient was not related to the magnitude of the therapy effect. The absolute values of Spearman's correlation coefficients for all seven scales were below 0.20. In the baseline data, there was a tendency for the functional VHI subscore to be higher for older patients (R= 0.40; $p<0.001$). We found no significant differences in therapy effect for gender or for functional versus non-functional disorders (Mann-Whitney *U* test, all $p>0.05$).

Reliability

According to Benninger et al.[10], the VHI is a robust instrument with a Cronbach's alpha coefficient (an indicator of internal consistency and reliability) of 0.95. The test-retest reliability of the VHI is reported to be 0.92.

The intraclass correlation coefficient (ICC) for the three-item outcome scale was 0.89 for the items on the severity of the voice disorder and the impairment in daily living, while it was 0.92 for impairment at work. However, for some individual patients, wide variations in the baseline data, leading to high standard deviations were found. In order to check whether these variations were related to real changes in voice quality, for each patient, the baseline scores were correlated with the results of the acoustic analysis and the perceptive evaluation at the corresponding measurement points. The mean correlation coefficients among all the patients were very low (for jitter R = –0.004 and for Grade R = –0.076). When only the data of patients whose baseline data showed highly variable severity scores (SD>15 mm, N=13) were used, again no significance could be attributed to the correlation between the severity scores and the jitter or Grade measured on the same day. Therefore, we conclude that the variability in baseline scores on the subjective parameters is not primarily caused by changes in voice quality. Rather, it is the result of an inconsistent application of the visual analog scales. The fact that the patients did not have their previous scores available when they repeated their self-evaluation may have played a role in this respect.

Relationship between the VHI and the three-item scale

Nonparametric correlations were computed between the baseline data as well as between the effect scores (post- minus pre-treatment data) for both instruments of self-assessment (Table 4). The average correlation coefficients for the items on the three-item outcome scale for baseline and effect data were, respectively, 0.68 (range

0.52 to 0.78) and 0.72 (range 0.65 to 0.79). The correlation coefficients for pairs of subscores of the VHI varied around 0.6. The range was 0.54 to 0.60 for the baseline data and 0.54 to 0.73 for the effect data. The correlation between the two instruments varied from 0.39 to 0.64 with respect to the baseline data and from 0.41 to 0.64 for the effect scores. The mean of these coefficients was about 0.5 for both methods. This is not far below the values found for the correlation coefficients within the two instruments.

As the scores are subject to random measurement errors, the correlation coefficients cannot attain a value of one, even when the scores are perfectly correlated. Using the estimates of the errors, a theoretical maximum for the correlation coefficient was computed for each pair of scores. The maximal correlation between the three questions comprising the three-item outcome scale was estimated to be 0.71. The maximal correlation between the subscales of the VHI was about 0.65; between the subscales of the two instruments, it was about 0.68. The correlations found between subscales per instrument are only slightly lower than these theoretical values. Apparently, in the patient's opinion, the influence of the voice disorder on the aspects represented by the subscores is roughly similar.

Table 4. Spearman's correlation coefficients for the Voice Handicap Index and the three-item outcome scale. The figures on a gray background refer to correlations between baseline data. The other figures refer to correlations between effect scores (post- minus pre-treatment data). The numbers of patients involved are in parentheses.

Dependent variable	Voice disorder severity	Impairm. in occupat.	Impairm. in daily living	Physical subscore	Funct. subscore	Emot. subscore	Total score
Voice disorder severity		.73 (52)	.78 (77)	.60 (69)	.53 (69)	.41 (69)	.58 (69)
Impairment in occupation	.73 (51)		.52 (77)	.64 (46)	.38 (46)	.42 (46)	.54 (46)
Impairment in daily living	.79 (76)	.65 (51)		.54 (69)	.55 (69)	.39 (69)	.56 (69)
Physical subscore	.60 (51)	.58 (34)	.51 (51)		.60 (69)	.52 (69)	
Functional subscore	.46 (51)	.62 (34)	.41 (51)	.55 (52)		.72 (69)	
Emotional subscore	.51 (51)	.64 (34)	.46 (51)	.54 (52)	.73 (52)		
Total score	.61 (51)	.68 (34)	.54 (51)				

A factor analysis based on Spearman's correlation coefficients confirmed the clear relationship between the scores on the three-item scale and the VHI. Using the three subscores of the VHI, the severity scale, and the scale on impairment in daily living, we were able to extract one factor that explains 66% of the total variance. The communalities ranged from 0.60 to 0.75. Even when the occupational impairment was included, we were still able to extract one factor explaining 65% of the total variance. In this analysis, we excluded all patients who did not have a job, leaving 46 subjects. A factor analysis performed on the effect scores (post- minus pre-treatment data) also resulted in one factor describing 67% of the total variance and communalities ranging from 0.61 to 0.73. We conclude that little differentiation is possible between the parameters of the two self-assessment instruments.

In some cases, self-evaluation of the severity of the voice disorder may be strongly influenced by a specific aspect in the patient's live, which corresponds with one of the questions of the VHI. In that event, the correlation between a single item and the three-item outcome scores may be higher for a single item than for the subscore sum. We considered this possibility, calculating the correlation (Spearman's correlation coefficients) for all of the individual questions in the VHI with the severity score. All of the resulting coefficients were lower than the coefficients found for the corresponding subscores. As the specific item may vary among the patients, the same computations were performed with the most abnormal scores (rating > 1) within each VHI subscore, as determined for each patient separately. Again, the correlations were of the same magnitude or lower than the coefficients found for the subscores. These results suggest that, in general, the rating on the three-item outcome scale is not determined by only one or just a few of the items in the VHI.

Relationship between the self-assessment and other clinical measures.

Nonparametric correlation coefficients were calculated for baseline data on quality-of-life measures, on the one hand, and for baseline data on perceptively rated overall severity of the voice disorder (Grade), jitter (Jitt%) and laryngostroboscopic parameters, on the other hand. The last three mentioned parameters are evaluation tools used in a clinical setting and regarded as objective measures for therapy evaluation. All the correlations were low. The correlation coefficients of the 3-item scale and the VHI scores with Grade varied between 0.29 and 0.37 for mean baseline data. The calculations for Jitt% yielded only for the four VHI scores significant correlation coefficients (R=0.32 to 0.34). For the other clinical measures significant correlations were found for the VHI scores with the maximum amplitude of the vocal fold vibration, the quality of the mucosal wave, and the impact of a lesion (range R=0.32 to 0.56).

The Spearman correlation coefficients for the post-pre therapy differences of the patients' rating of severity of their voice disorder and the differences in parameters for voice quality and the laryngostroboscopic parameters, were of the same magnitude as those for the baseline data. However, the highest coefficient for the severity rating was found for the correlation with the improvement of the vocal fold closure.(R= 0.40, p=0.01). The correlations with the maximum amplitude of the vocal fold vibration, the quality of the mucosal wave, and the regularity of the vocal fold vibratory pattern were significant as well. The impact on vocal fold vibration of a changed lesion had no effect on the subjectively perceived severity of the voice disorder (R<0.13). The changes in the parameters Grade and Jitt% did not significantly correlate with the patient's severity ratings. For the VHI scores, only the change in vocal fold closure correlated significantly with the change in the total score (R=0.43) and with the change in the functional and emotional subscore (R=0.41 and 0.33). Low correlations between the pre- to post-therapy changes of the self-assessment ratings and the more objective clinical measures may be caused by several effects: The internal scales for severity may widely vary among the patients; not all the patients will associate the severity with the same aspects of voice abnormality; and the changes due to the therapy were not found in the same parameters in all patients. These problems are reduced when only the classification of the individual therapy effect is used and the pre- to post-therapy changes of all parameters are taken into account. To that end, the numbers of patients whose self-assessment indicated that their voice problem was significantly reduced or was not changed, were counted in the groups with or without a significant improvement in at least one of the clinical measures. The results for the patient's rating of severity and the total VHI are shown in Table 5. In 40% of the patients, a significant improvement on the 3-item severity scale corresponded with a significant improvement in at least one of the clinical measures. Only 9% of the patients experienced an improvement of their voice disorder without any clear objective improvement. On the other hand, 36% of the patient did not find their therapy significantly successful, whereas clinically beneficial effects were found. No conclusions can be drawn from the table for the VHI, because of the high threshold for a significant change of the total score.

Table 5. Crosstabulation of the number of patients and percentage of the total number of patients classified according to their significant therapy effect on self-assessment scales and the best of the clinical measures. The clinical measures were perceptive rating (Grade), acoustic parameter (Jitt) and assessment of laryngostroboscopic recordings. The highest classification group was used for crosstabulation. Classification group= 0: significant deterioration after therapy; group = 1: pre- and post therapy not abnormal; group = 2: no significant therapy effect; group = 3: significant improvement after therapy. One patient who was in group 0 for the clinical measures is not in the table. (As not all patients completed both VHI questionnaires, the number of patients differs for the severity scale and the total VHI score.)

Clinical measures	Severity scale Classification (N=75)			Total score VHI Classification (N=51)		
	Group 0	Group 2	Group 3	Group 1	Group 2	Group 3
Group 2	1 (1%)	8 (11%)	7 (9%)	1 (2%)	8 (16%)	3 (6%)
Group 3	1 (1%)	27 (36%)	30 (40%)	5 (10%)	26 (51%)	8 (16%)

Discussion and conclusion

Opinions proved to differ widely among the chronically dysphonic patients about the success of their voice therapy. This divergence showed up in the VHI and on the three-item outcome scale. The statistical tests indicated that, on average, the improvement in the patients' situation was highly significant. Yet at the individual level, improvement was significant for only about half of the subjects. About 8% of the patients considered their situation after therapy to be normal.

The correlations between the self-assessment ratings and the perceptual rating of voice quality by expert listeners appeared to be low. One of the reasons mentioned before may be a marked variation among the patients about the severity level assigned to the endpoint of the scale. On the other hand, patients may base their judgement of the severity of their voice problem on more factors than strictly voice quality. In the baseline data the mean correlation coefficient between the self-assessment ratings and Grade was 0.35, but the coefficients between the VHI total score and the maximum amplitude of the vocal fold vibration or the quality of the mucosal wave were higher (R=0.50 and 0.51). Low correlation coefficients were found between the self-assessment ratings and jitter as well. This result is in line with the poor correlations found by Hsiung et al.[11] between the Voice Handicap Index and acoustic parameters in dysphonic patients.

The opinion of the patient about the improvement after therapy is not related with the same clinical measures as their judgement of the severity of the voice disorder before therapy. The main factor determining the patient's assessment of the change in the voice disorder appeared to be the change in vocal fold closure. Only the regularity of the vocal fold vibration as assessed in the laryngostroboscopic recordings is clearly related to the severity rating as well. The improvement in voice quality, as it is judged by experts, seems to play a minor role in the subjective perception of the patient. If only the success of the therapy as assessed by the clinical measures and the patient are compared, an agreement on a significant improvement or not, exists for 51% of the patients. A distinct 'placebo' effect, that is a significant improvement in the self-assessment without any clinical improvement, was not found. For 36% of the patients no significant positive therapy effect could be derived from their self-assessment of the severity of the voice disorder, whereas significant clinical improvements in these patients were found. These improvements were apparently too subtle for the patient to be felt as a beneficial effect.

The subjective improvement of the voice disorder was greater when measured on the three-item scale than on the VHI On a scale of 0 to 100 mm, the mean improvement in the severity score on the three-item scale was -22 mm. On a scale of 0 to120, the mean change in the total score of the VHI was -6. The percentage of patients with a significantly lower score after therapy was 51% for the severity score of the three-item scale and 22% for the VHI total score. Even when the confidence interval of the VHI score was halved (to +/- 9), the latter percentage only went up to 31%. However, the fact that about half of the items in the sum of the VHI had more or less normal scores before therapy partly explains the narrower gap between the scores before and after treatment. Indeed, a score that is normal before therapy cannot contribute to a decrease in the sum for a subscore or total score. If only the VHI items that are abnormal (with a score above one) are used in the comparison of pre- and post-therapy scores, the difference between the VHI results and those for the three-item outcome scale turn out to be much smaller. Nevertheless, the improvement measured with the instrument consisting of three simple questions is more pronounced than the improvement measured with the questionnaire containing many questions.

The correlations between the scales of the VHI and the three-item outcome scales were only moderate. This might imply that the two instruments measure different aspects of the patients' well-being. However, the correlation coefficients are only slightly below the theoretical maximum for the correlation coefficients as estimated with the random measurement errors. A factor analysis revealed only one factor. Still, differences between the two scales may be explained by the fact that the VHI does not cover certain aspects that the patient considers important. Another reason may be that the rating on a general question might be more susceptible to the patient's emotions than the specific questions in the VHI would be. However, we did not find a particularly high correlation between the scores on the three-item scale and the emotional subscale of the VHI In conclusion, it seems that the two quality-of-life instruments refer to slightly different aspects of the patient's subjective evaluation of therapy. Therefore, some caution should be exercised when applying a simple outcome scale to validate new questionnaires on quality-of-life instruments with respect to dysphonia.

A score based on many different aspects may yield a more objective indication of a patient's impairment. However, when the patient considers many of the aspects in the questionnaire irrelevant, that instrument will underestimate the effect of treatment compared to the result produced by simple scales posing a single, more general question. The items perceived as irrelevant would contaminate the sum scores of the questionnaire and may enhance the variance of the sum scores. On the other hand, averaging over many different questions may reduce the variability of the result. This effect could make a battery of questions a more valuable instrument than one containing a single question. In this study, the test-retest reliability was not determined for both of the quality-of-life instruments. Nonetheless, the results do not suggest that the questionnaire is more reliable than the simple questions. The reliability of the VHI as reported in the literature is not substantially higher than that computed for the single questions on the three-item scale. Aside from that, one advantage of using a questionnaire such as the VHI is that it yields information about the type of impairments that are most important to the patient. For a general evaluation of the patient's handicaps before and after voice therapy, some simple questions rated on a visual analog scale may suffice. Indeed, this simple method may even have some advantages over a more detailed questionnaire. It may save time, there is no need for extensive explanation or supervision by the therapist or specialist, and it would probably persuade even unmotivated patients to participate.

Acknowledgments

We are grateful to Dr. C.J. van As for the possibility to use the Dutch translation of the Voice Handicap Index. This translation was a collaborative effort of the Institute of Phonetic Sciences of the University of Amsterdam, the Netherlands Cancer Institute / Antoni van Leeuwenhoek Hospital in Amsterdam, and the Department of Language and Speech of the University of Nijmegen.

Literature

1. Jacobson B.H., Johnson A., Grywalski C. et al. The Voice Handicap Index (VHI): development and validation. *American Journal of Speech-Language Pathology* 1997;6:66-70.

2. Glicklich R.E., Glovsky R.M., Montgomery W.W. Validation of a voice outcome survey for unilateral vocal cord paralysis. *Otolaryngol.Head.Neck Surg.* 1999;120:153-158.
3. Hogikyan N.N., Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J.Voice* 1999;13:557-569.
4. Casper J.K. Treatment outcomes in occupational voice disorders. In Dejonckere P.H. ed., *Occupational voice: care and cure.* The Hague: Kugler Publications, 2001:187-199.
5. Billante C.R., Spector B., Hudson M. et al. Voice outcome following thyroplasty in patients with cancer-related vocal fold paralysis. *Auris.Nasus.Larynx.* 2001;28:315-321.
6. Spector B.C., Netterville J.L., Billante C. et al. Quality-of-life assessment in patients with unilateral vocal cord paralysis. *Otolaryngol.Head.Neck Surg.* 2001;125:176-182.
7. Roy N., Gray S.D., Simon M. et al. An evaluation of the effects of two treatment approaches for teachers with voice disorders: a prospective randomized clinical trial. *J.Speech Lang.Hear.Res.* 2001;44:286-296.
8. Rosen C.A., Murry T., Zinn A. et al. Voice Handicap Index change following treatment of voice disorders. *J.Voice* 2000;14:619-623.
9. Speyer R., Wieneke G.H., Dejonckere, P.H. Documentation of progress in voice therapy: perceptual, acoustic and laryngostroboscopic findings pre- and post-therapy. *J.Voice* 2003 (submitted).
10. Benninger M.S., Gardner G.M., Jacobson B.H. et al. New dimensions in measuring voice treatment outcomes. In Sataloff R.T. ed., *Professional voice: the science and art of clinical care*. San Diego: Singular Publishing Group, 1997:789-794.
11. Hsiung M.-W., Pai L., Wang H.-W. Correlation between voice handicap index and voice laboratory measurements in dysphonic patients. *Eur.Arch.Otorhinolaryngol.* 2002;259:97-99.

Appendix 1. Three-item outcome scale

Item 1 How would you rate the severity of your voice disorder?

Item 2 To what extent does your voice disorder bother, inconvenience, or restrict you in your work? (For instance, how much does it hinder you when speaking with clients, teaching, talking on the phone, attending meetings, and so forth?)

Item 3 To what extent does your voice disorder bother, inconvenience, or restrict you in your social activities and in your daily living? (For instance, how much does it hinder you when talking to friends and relatives, shopping, singing, and so forth?)

Chapter 7

General discussion and conclusions

General discussion and conclusions

Introduction

The growing importance of evidence-based medicine has stimulated the demand for scientific research on the effects of (para-)medical therapy. As a consequence, funds became more readily available for studies on the effectiveness of therapies. It was within this scope, that this study was financially supported by the Health Care Insurance Council. If positive effects on therapies cannot be proven scientifically, budgets for this kind of therapies might be cut for economic reasons. Important issues, like whether or not voice therapy is successful, or how to define such success in objective terms, have led to this attempt to investigate the effects of logopedic voice therapy. In contrast to some medical and surgical treatments, of which the effects are obvious and more easily assessed, the correct evaluation and measurement of voice therapy is under discussion. The perceptual evaluation of voice disorders seems to be the golden standard for most speech therapists. However, this study strongly suggests, that other evaluation instruments cannot be neglected. In general, therapy effects found for parameters of acoustic analysis or laryngostroboscopy for example, are more distinct than the rather small changes in the perception of the dysphonia.

This study on voice therapy used a prolonged baseline design in order to perform repeated measurements in a situation without therapy. During this baseline period, changes on parameters of the perceptual evaluation, the acoustic analysis and the patient's self-evaluation were tested for significance. No significant changes or trends towards normal values were found. These findings confirmed the assumed chronic character of the dysphonic problems in the subject population. However, this is no conclusive prove that, within the patient population, no spontaneous recovery occurred. A baseline period of about six weeks is still a short period of time compared with the 12 or 24 weeks duration of the therapy. Furthermore, only a limited number of parameters has been measured repeatedly. For example, the videostroboscopic recordings could not be carried out three times within six weeks because of ethical as well as practical reasons. As voice is considered to be a multidimensional phenomenon, in fact changes on all the parameters used in this study should have been tested for statistical significance. A conclusive proof that the changes found in a study are due to the therapy, is only possible if a placebo group is included in the study design. However, a procedure that guarantees a real placebo condition, will not be easily realized in a normal clinical situation. A patient, fully informed about a medical abnormality found, will generally not accept that he or she receives no therapy during a long period. The restriction to patients with chronic disorders and the use of measurements in a prolonged baseline are considered to be a second best solution.

Effects of logopedic voice therapy

The findings of this study allow us to conclude that the positive effects of logopedic voice therapy for chronically dysphonic patients as a group are statistically significant when studied in terms of the changes on parameters of laryngostroboscopy, perceptual evaluation, acoustic analysis, phonetography, and the patient's self-evaluation. No significant therapy effects were found for the aerodynamic parameters maximal phonation time and phonation quotient. The therapy effects are in line with the general positive effect of specific therapies, found in literature (See Chapter 1 General introduction). However, as mentioned before, the individual effects of therapy were highly diverse. For some patients the therapy was highly successful, but for other patients almost no improvement was achieved. The majority of the patients (64/78) showed a significant improvement for at least one of the objective evaluation tools, but, generally, not all parameters that were abnormal before therapy, normalized in a patient. The fact that the patients could not be divided in a group with success on all or nearly all abnormal parameters and a group without significant improvements, illustrates the multidimensionality of the effect of voice therapy for patients with a chronic dysphonia. It also explains, at least partly, the modest, but significant results for group effects of the individual parameters. A patient with a decrease of abnormality in one parameter may not show any improvement in another parameter.

The therapy effects could not be explained by a relationship between a patient related variable and the effects. Relations between medical-etiological diagnoses and therapy effects proved not statistically significant. Clear relationships with pre-therapy data or other patient related data were not found either. Apparently, other factors, which were not included in this study, are more important. These factors can be, for example, smoking habits, or the motivation and compliance of the patient with the functional / behavioral approach. The patient is generally supposed to change his/her voice related behavior, to do voice exercises at home and to integrate, what has been learnt, in spontaneous communication. The patient's acceptance to play such an active role in the therapy, instead of undergoing some kind of medical treatment in which patient's activities are restricted to, for example, taking medication, will be important for a successful therapy. On the other hand, apart from these patient related factors, the therapy strategy may have been not adequate for all patients. The therapists were asked to give a description of their therapy and estimate the weight of the different therapy elements in the therapy they had employed. As the response to this question from the therapists was low, the data were insufficient for a good analysis of a relation with therapy success. As long as no prognostic parameters are available that predict which patients will benefit from therapy, one can consider voice therapy necessary for the total patient population with chronic dysphonia.

It is very likely, that a patient population as found in the Phoniatric Department of the University Hospital Utrecht, differs from the population in a speech therapist's practice. The patient population of the Phoniatric Department has already been seen by other medical specialists or family doctors before being referred to the Department. This group of patients will show more severe or persistent voice disorders than the common patient seen by speech therapists in their practices.

Therefore, some caution must be taken in generalizing the results from this study to other groups of dysphonic patients receiving voice therapy. The outcome of voice therapy for these other groups, with less severe or less persistent disorders, remains unclear and requires further investigation in future research projects.

Judgments made by speech therapists ,versus a panel of expert listeners

Usually, the speech therapists themselves decide on whether the outcome of voice therapy can be considered successful or not. In this study it appeared that speech therapists gave a noticeably more positive subjective/perceptual evaluation of therapy outcome than the independent panel of expert listeners. The fact that the therapists judged the situation before therapy to be more severe than the panel, indicates that the therapists assigned a lower severity level to the highest point of the scale than the panel did. The expert listeners working in a clinical setting may be more accustomed to severely impaired subjects than the majority of the extramural speech therapists . Consequently, the internal standards for maximum severity will be different for the panel and the therapists. Another consequence of the more severe baseline judgements of the extramural speech therapists is that a patient's performance as judged by a speech therapist can show more improvement after therapy, than when judged by the panel. This may partly explain the larger post- minus pre-therapy differences. However, a more important reason for a discrepancy between the two judgements may be that the voice samples were offered at random per patient to the panel, whereas the speech therapist was fully aware of the moment of measurement. Moreover, the panel compared two samples of which one was recorded before therapy, whereas the speech therapists had the earlier rating not at their disposal as a point of reference. The effects at the end of therapy may have been exaggerated by the speech therapists, indicating a tendency rather than an absolute scaling. However, the correlations between the ratings of the panel and the therapists were low, which does not fit with a mere scale factor between the rating of the extramural speech therapists and the panel. As the correlations between a therapist's perceptual evaluation and the laryngostroboscopic and acoustic parameters were not higher than those for the panel, there is no reason to believe that the therapist's ratings would reflect the real changes more accurately than the panel's ratings. Therefore, evaluation by means of a perceptual assessment by the voice therapists themselves should be interpreted with caution. Consequently, an essential methodological aspect of this kind of effectiveness-studies is, that the evaluation occurs without the rater's awareness, or even suspicion, whether the data to be rated, are pre- or post-treatment. This requires a time-consuming preparation of the auditory and visual material that will be rated, but guarantees the validity of the conclusions.

One may argue, that the panel's assessment, being based on a fragment of a reading text, is not comparable to that of a therapist, who had assessed the patient's voice under many different circumstances and possibly in more natural conversation. Actually, the judgement of the therapist may be based on the patient's best moments and would therefore not reflect the overall status of the voice. Furthermore, the patient has been trained to apply certain therapeutic techniques during sessions of voice therapy. The patient may not (yet) be ready to use these new techniques in other settings, like the clinical one in our department.

Dropouts

The number of subjects that did not accomplish the full period of therapy was unexpectedly high. About 25% of the patients that were included at the beginning of this study broke off their therapy course. The reasons for leaving the therapy course were diverse: private matters (such as finance, family affairs, illness) or lack of motivation. Generally, in literature, such high numbers of subjects dropping out from therapy, are not reported. In many papers no information on drop-outs can be found at all. It could be that in these studies all patients did accomplish the full therapy period. This seems only possible if all patients in these studies were very motivated. Such a highly motivated group of patients can hardly be considered to be the average, ordinary subject population in actual clinical practice.

Beforehand, it might be expected that not all subjects included in a study, will complete their therapy course without anyone stopping early. An essential aspect of logopedic voice therapy is, that patients are supposed to change their vocal behavior over a longer period of time, as well as to invest quite an amount of energy and time in practicing voice exercises and integrating improved voice hygiene and better voice technique in spontaneous communication. At the beginning of the therapy, at least part of the patients may not have realized, or was not fully aware, of the considerable effort they were supposed to make. Moreover, patients may be satisfied with the achieved results as soon as the communication isn't too problematic anymore. Although they may still have a poor voice, according to the professional view of their speech therapist, they stop with the therapy. The patient may find it not worth trying to achieve a perceptually better voice.

Multidimensionality of voice

Usually, a patient's main complaint will refer to the perceptually disordered voice quality that often goes together with a sore throat or voice fatigue. The perceptual evaluation will be the main, if not the only evaluation instrument that is used by a speech therapist to assess the success of a therapy. However, it is shown that the perceptually disordered voice quality covers just one aspect of voice abnormality. Other parameters of the different assessment instruments of voice, including videostroboscopy, acoustic analysis, voice range profiles or patient's self-evaluation, show significant deviations from normal, which are independent from the perceptual abnormalities. Apparently, in order to make a complete evaluation of the effect of voice therapy, a multidimensional assessment must be used, including all different aspects of the vocal function and not only the perceptual rated voice quality. The Committee on Phoniatrics of the European Laryngological Society (Dejonckere et al.,

2001) made the following recommendations for a minimal set of multidimensional measurements for functional assessment of voice pathology: perceptual rating, videostroboscopy, acoustic analysis, aerodynamic measures, and subjective rating by the patient.

These conclusions have consequences for the evaluation of speech therapy, also when practiced outside a clinical setting. Although the group effects on perceptual evaluation were significant, the mean perceptual changes were very small. Therefore, when a therapy outcome is based on perceptual improvement only, firm conclusions on results cannot be made. And consequently, it would suggest that voice therapy for the chronically dysphonic patients would not seem to be appropriate. However, especially in a medical setting, physiological improvements are being considered important as well. Moreover, changes in the feeling of well-being can still make an important difference for an individual patient. Without offering therapy to these patients, the beneficial effects on the physiology of the vocal folds, on the vocal fold biomechanics and the patient's well-being would be withheld from them.

Factor analyses applied on the baseline data showed that the various evaluation instruments represented separate factors. Parameters of the different instruments that could be expected to be related according to underlying (hypothetical) theories did not share the same factor. For example, an insufficient closing of the vocal folds during vibration will cause a turbulent airflow during the closing phase. This turbulence is thought to have a relation with the perceptual impression of breathiness and the acoustic noise-to-harmonic ratio. However, the rating of the degree of vocal fold closure and breathiness or NHR loaded on two different factors. Thus, other factors will be more important for the generation of acoustic noise. It could be expected that an irregular vocal fold vibration is related to increased jitter. Irregular vibration is also thought to correlate with the perceptive impression of roughness. Again, the parameters were in separate factors. Simple theoretical considerations appear to be inadequate oversimplifications of the actual events. Another conclusion is that the multidimensional character of the voice leads to the independence of the results of different instruments. No relation between the dimensions and specific abnormalities of the vocal fold vibration could be established.

Suggestions for further research

This study is only a first step in the field of outcome measurement in logopedic voice therapy. It provided some answers on the overall effects of voice therapy, but many questions have to be investigated in future research, for example:

- Further research will be essential for a better understanding of the interactions of medical-etiologic diagnoses and the effects of therapy. Larger groups of subjects per diagnostic category will be needed;

- Patients outside the academic hospital setting must be included in new studies on therapy effects as well. This population will be an important part of the patients in the usual logopedic practice.
- Special attention must be paid to the possible influence of patient's motivation and faithfulness to the therapy, the amount of voice exercises at home, smoking habits, and vocal abuse;
- The overall efficacy of speech therapy over time and its therapy frequency must be subject of study. Special attention must be given to the required number of therapy sessions before reaching a patient's maximal improvement. What are the criteria to end voice therapy?;
- Other evaluation instruments such as aerodynamic measures (apart from maximum phonation time and phonation quotient, for example, a flow-glottogram), can be added to the multidimensional assessment of voice characteristics as described in this study;
- The therapy outcome of the diverse types of voice therapy must be compared to one another. Is giving advices concerning vocal hygiene and avoiding vocal abuse sufficient, or is it absolutely necessary to exercise the new, adjusted use of the voice outside the therapy situation in making the therapy outcome successful?;
- How many voice samples are required for a good judgement of the voice quality of a patient, before and after therapy? Is a personal contact between patient and rater an advantage or a disadvantage in this respect?

Conclusions

Based on this thesis following conclusions can be drawn:

- The effects of logopedic voice therapy for chronically dysphonic patients as a group are significant, but quantitatively rather modest. Furthermore, the individual effects of therapy are highly diverse.
- A multidimensional assessment instrumentarium must be used when evaluating the overall effects of voice therapy. The minimal assessment protocol should include laryngostroboscopy, perceptual evaluation, acoustic analysis, self-evaluation by the patient, and a maximal performance task such as a voice range profile.

Summary

Summary

The demand for scientific research on the effects of (para) medical therapy has been stimulated by the growing importance of evidence-based medicine. This thesis describes a first approach for measuring the result of logopedic voice therapy. The subject population consists of patients with chronic dysphonia.

Chapter 1, the introduction, describes the multidimensional instrumentarium used for functional assessment of voice pathology. In addition, a review of relevant scientific studies on the effects of logopedic voice therapy in dyphonic patients is presented.

Chapter 2 deals with the amount of intra-subject variability to be expected when using an objective, acoustic voice evaluation analysis. It also describes in what way this variability affects the assessment of a therapy outcome. In total, 82 patients suffering from chronic dysphonia were asked to produce three sustained vowels /a:/ at a comfortable pitch and loudness level. Three recordings were made over a six-week period. For each sample, acoustic variables characterizing jitter, shimmer and harmonic-to-noise ratio were computed. Statistical analysis was then used to estimate the expected intra-subject variability per acoustic parameter.

The variability of perturbation parameters increases with the magnitude of the parameter. Therefore, the inaccuracy was characterized by a relative error (coefficient of variation). For the acoustic parameters, distinctive inaccuracies in series of consecutive sustained vowels were found: the coefficients of variation ranged from 14% to 33%. When an individual therapy was evaluated by acoustic perturbation measurements, the ratio of the post-therapy to the pre-therapy value generally had to be below 0.5 to 0.4, for ensuring the improvement to be significant.

In **Chapter 3** therapy effects are described, while using objective measurements, derived from digitized laryngeal stroboscopic images. These measurements were used to demonstrate changes in vocal fold vibration and in the size of benign lesions after three months of voice therapy. Forty chronically dysphonic patients were studied. Using a rigid stroboscope, pre-treatment and post-treatment recordings were made of the vocal folds at rest and under stroboscopic light during phonation. From each recording, images of the positions at rest, during vibration at maximal opening and at maximal closure were digitized. The surface area of any lesions and of the glottal gap were independently measured in the digitized images by two experienced laryngologists. Referential distances were determined in order to compensate for discrepancies in magnification in the various recordings. After three months of voice therapy, significant improvement in lesion size and degree of maximal closure during vibration could be demonstrated in about 50% of the patients. The degree of maximal opening did not prove to be a significant parameter.

Chapter 4 describes therapy effects using voice range profiles or phonetograms. This study had two objectives. The first was to investigate which parameters in the voice range profile of dysphonic patients showed significant changes after voice therapy, and therefore may be useful in general for the assessment of voice therapies. In order to get a clinically useful result, patients with diverse pathologies were included in the study. The second objective was to determine the size of the demonstrated effects and to investigate how these might relate to phoniatric diagnoses.

In a group of chronically dysphonic patients, a voice range profile was recorded before and after receiving voice therapy and again three months later. The voice range profiles took a wide variety of shapes. Therefore, only measurements that did not depend on a smooth contour could be used to describe changes before and after therapy. The main effect of voice therapy was an enlargement on the side of low frequency and low intensity.

Chapter 5 examines therapy effects evaluated by perceptual rating, acoustic analysis, and the assessment of laryngostroboscopic recordings. Although the group effects for the differences between post- and pre-therapy data were clearly significant, the effects of voice therapy for the individual patients were divergent. For each of the three evaluation methods, a significant improvement was found for about 40% to 50% of the patients. The diversity of the therapy outcome among the patients could not be explained by the pre-therapy status nor to age, gender, or diagnosis groups. In general, the perceptual ratings and the acoustic parameters from the baseline data were clearly correlated. However, these characterizations of the voice were only moderately correlated with the visual evaluation of the vocal fold vibrations. Relations between the three evaluation tools for the changes due to voice therapy were very weak. The low correlation among the three methods suggests that a multidimensional evaluation of the voice is necessary to give a complete picture of the therapy outcome.

Chapter 6 focuses on therapy effects using two quality-of-life instruments: the Voice Handicap Index or VHI (Jacobson et al.[1]), and a simple three-item outcome scale (three visual analog scales). Both instruments measure changes in the quality of the voice itself and in the extent of impairment resulting from the dysphonia as experienced by the patient in a social and an occupational environment. Statistical tests conducted on pre- and post-treatment data indicated highly significant improvements on both instruments for the group as a whole. At the individual level, however, the effects were diverse. For roughly 50% of the subjects a significant improvement could be established. The positive changes as measured with the three-item scale were greater than those measured with the VHI. In order to obtain a rough indication of the patient's handicap, it may suffice to ask some simple questions.

Chapter 7 gives a general discussion and the conclusions on the measurement and evaluation of logopedic voice therapy in chronic dysphonic patients. The findings of this thesis confirm the positive significant group effects of therapy, when studied in terms of changes of parameters of laryngostroboscopy, perceptual evaluation, acoustic analysis, phonetography and self-evaluation by patients. However, the individual effects of therapy can be highly diverse. Remarks are made on the noticeably more positive evaluation of therapy outcome by speech therapists compared to the assessment by an independent panel of expert listeners. Evaluation by means of perceptual assessment by the voice therapists themselves, should be interpreted with caution, as it is essential for an effectiveness-study that the evaluation occurs without the rater knowing, or even suspecting, whether the rated material is pre- or post-treatment. In order to evaluate the overall effect of voice therapy, a multidimensional assessment instrumentarium has to be applied. Factor analyses, applied on our data, showed, that the diverse instruments represent mutually independent factors. Parameters, which, based on commonly used underlying hypothetical theories, were thought to be related, did not share the same factor. The relations between the assessment instruments and their parameters are rather complex.

1. Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI): development and validation *American Journal of Speech-Language Pathology* 1997;6:66-70.

Samenvatting

Samenvatting

Naarmate het belang van 'evidence based medicine' groeit, neemt ook de vraag naar wetenschappelijk onderzoek op het gebied van (para-)medisch handelen toe. Deze thesis beschrijft een eerste poging om de effecten van logopedische stemtherapie te meten. De subjectpopulatie bestaat uit patiënten met chronische dysfonie.

Hoofdstuk 1, de introductie, beschrijft het multidimensionele instrumentarium dat wordt gebruikt voor functioneel onderzoek van stempathologie. Daarnaast wordt een overzicht gegeven van relevante wetenschappelijke studies aangaande effectmeting van logopedische stemtherapie bij dysfone patiënten.

Hoofdstuk 2 behandelt de intrasubjectvariabiliteit die kan worden verwacht als gebruik wordt gemaakt van een objectieve, akoestische stemanalyse. Ook wordt beschreven op welke wijze deze variabiliteit de beoordeling van therapie-effecten beïnvloedt. In totaal is aan 82 patiënten met chronische dysfonie gevraagd om drie aangehouden klinkers /a:/ te produceren op een comfortabele toonhoogte en luidheid. Binnen een periode van zes weken werden drie opnames gemaakt. Voor ieder spraaksample werden akoestische variabelen voor jitter, shimmer en harmonic-to-noise ratio berekend. Met behulp van statistische analysen werd de te verwachten intrasubjectvariabiliteit per akoestische parameter geschat.

De variabiliteit van pertubatieparameters neemt toe met de grootte van de parameter. Daarom werd de onnauwkeurigheid gekarakteriseerd door een relatieve fout (variatiecoëfficiënt). Voor de akoestische parameters werden verschillende onnauwkeurigheden gevonden in een reeks van aangehouden klinkers: de variatiecoëfficiënten varieerden van 14% tot 33%. De dag tot dag variatie bleek van dezelfde orde als de variatie in een opeenvolgende reeks op één dag. Als een individuele therapie beoordeeld werd aan de hand van akoestische perturbatiemetingen, dan moest de ratio van posttherapie- gedeeld door pretherapiewaarde in het algemeen beneden de 0.5 tot 0.4 zijn, opdat de verbetering significant zou zijn.

In **hoofdstuk 3** worden de therapie-effecten beschreven die worden gemeten met behulp van objectieve metingen afkomstig uit gedigitaliseerde laryngale stroboscopische foto's. Deze metingen werden gebruikt om veranderingen aan te tonen in de stemplooitrilling en in de grootte van benigne laesies na drie maanden stemtherapie. Veertig patiënten met chronische dysfonie werden gevolgd. Met behulp van een rigide stroboscoop, werden voor en na behandeling opnames gemaakt van de stemplooien in rust en met stroboscopisch licht gedurende fonatie. Van iedere opname werden beelden gedigitaliseerd van de stemplooien in rust en gedurende trilling tijdens maximale opening en maximale sluiting. De oppervlakte van de laesies en van de glottisopening werden onafhankelijk door twee foniaters/KNO-artsen gemeten in de gedigitaliseerde beelden. Er werden referentie-afstanden bepaald zodat verschillen in vergroting tussen de opnamen konden worden gecompenseerd. Na drie maanden therapie konden significante verbeteringen in laesiegrootte en de mate van maximale sluiting gedurende trilling in ongeveer 50% van de patiënten worden aangetoond. De mate van maximale opening bleek geen significante parameter.

Hoofdstuk 4 beschrijft therapie-effecten die zijn gemeten met behulp van fonetografie. Een fonetogram legt de omvang van de stem vast in een vlak met de luidheid en grondfrequentie als assen. Deze studie had twee doelstellingen. Ten eerste werd onderzocht welke parameters in het fonetogram van dysfone patiënten significante veranderingen na therapie lieten zien en derhalve zinvol zouden kunnen zijn bij het evalueren van stemtherapieën in het algemeen. Teneinde een klinisch zinvolle uitkomst te verkrijgen, werden in deze studie patiënten met uiteenlopende pathologieën geïncludeerd. Het tweede doel was om de grootte van de aangetoonde effecten te bepalen en de eventuele relaties met de foniatrische diagnosen te onderzoeken.

In een groep chronisch dysfone patiënten, werd een fonetogram opgenomen voor en na stemtherapie en opnieuw drie maanden later. De fonetogrammen waren zeer verschillend van vorm. Om de veranderingen voor en na therapie te beschrijven, konden derhalve alleen maten worden gebruikt die niet afhankelijk waren van een regelmatig contour. Het belangrijkste effect van stemtherapie bleek een uitbreiding aan de kant van de lagere frequenties en zachtere intensiteiten.

Hoofdstuk 5 onderzoekt de therapie-effecten zoals geëvalueerd met behulp van perceptieve evaluatie, akoestische analyse en laryngostroboscopisch onderzoek. Alhoewel de groepseffecten voor de verschillen tussen post- en pretherapiedata duidelijk significant waren, bleken de stemtherapie-effecten voor de individuele patiënt zeer uiteen te lopen. Voor elk van de drie evaluatiemethoden werd een significante verbetering gevonden voor ongeveer 40% tot 50% van de patiënten. De verschillen in therapieresultaten tussen de patiënten onderling konden niet worden verklaard door de pretherapiestatus, noch door leeftijd, geslacht of diagnosegroep. Over het algemeen waren de perceptieve beoordelingen en de akoestische parameters van de baselinedata duidelijk gecorreleerd. Deze stemkenmerken bleken echter slechts matig te correleren met de visuele evaluatie van de stemplooitrillingen. Relaties tussen de veranderingen ten gevolge van therapie zoals gemeten met deze drie evaluatie-instrumenten, waren erg zwak. De lage correlaties tussen de drie methoden suggereert dat een multidimensionele evaluatie van stem noodzakelijk is om een volledig beeld te kunnen geven van de therapie-effecten.

Hoofdstuk 6 richt zich op therapie-effecten gemeten met behulp van twee quality-of-life instrumenten: de Voice Handicap Index (Jacobson et al.[1]) en een eenvoudige drie-item effectschaal (drie visueel analoge schalen). Beide instrumenten meten veranderingen in de stemkwaliteit zelf en in de mate van beperking als gevolg van de dysfonie zoals ervaren door de patiënt binnen zijn of haar sociale omgeving alsook binnen de werksituatie. Statistische testen uitgevoerd over pre- en posttherapiedata toonden op beide instrumenten hoog significante groepsverbeteringen aan. Op het individuele niveau echter, liepen de effecten uiteen. Voor ruwweg 50% van de subjecten kon een significante verbetering worden aangetoond. De positieve veranderingen gemeten met de drie-item schaal waren groter dan welke gemeten met de VHI. Teneinde een globale indicatie te verkrijgen van de door de patiënt ervaren handicap, zou het stellen van enkele eenvoudige vragen mogelijk al voldoende zijn.

Hoofdstuk 7 biedt een algemene discussie en de conclusies aangaande het meten en evalueren van logopedische stemtherapie in chronisch dysfone patiënten. De bevindingen in deze thesis bevestigen de positief significante groepseffecten van therapie indien deze worden bestudeerd in termen van veranderingen in parameters van laryngostroboscopie, perceptieve evaluatie, akoestische analyse, fonetografie en zelfevaluatie door de patiënt. De individuele effecten zijn echter zeer divers. De beduidend positievere evaluatie van therapie-effecten door logopedisten vergeleken met de beoordeling door een onafhankelijk ervaren luisterpanel wordt besproken. Evaluatie door middel van perceptieve beoordeling door de behandelende stemtherapeuten zelf moet met enige terughoudendheid worden geïnterpreteerd aangezien het essentieel is voor effectonderzoek dat de evaluatie plaats heeft zonder dat de beoordelaar weet of zelfs een vermoeden heeft of het te beoordelen materiaal van voor of na de behandeling dateert. Opdat de overall effecten van stemtherapie in kaart kunnen worden gebracht, blijkt een multidimensioneel onderzoeksinstrumentarium noodzakelijk. Factor analyses toegepast op onze data toonden aan dat de verschillende instrumenten wederzijds onafhankelijke factoren vertegenwoordigen. Parameters die op basis van onderliggende hypothesen verondersteld werden samenhang te vertonen, hadden niet dezelfde factor gemeenschappelijk. De relaties tussen onderzoeksinstrumenten en bijbehorende parameters zijn complex.

1. Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI): development and validation *American Journal of Speech-Language Pathology* 1997;6:66-70.

Dankwoord

Dankwoord

Met de verdediging van dit proefschrift wordt een periode van zes jaar afgesloten. Zonder de hulp van velen was mij dit nooit, nooit gelukt.

Prof. Dr. P.H. Dejonckere, beste Philippe, als promotor van het geheel ben jij de initiator geweest van het hier beschreven onderzoek. Dankzij jou, werd mij de gelegenheid geboden tot promoveren.

Dr. ir. G.H. Wieneke, beste George, zonder jou was ik na drie maanden allang in Nijmegen ondergedoken. Je weet dat ik de samenwerking gedurende de afgelopen zes jaren enorm heb gewaardeerd en hoop je ook de komende tijd nog regelmatig tegen te komen. Met engelengeduld heb je geprobeerd mij de beginselen van wetenschappelijk onderzoek bij te brengen. Jouw persoonlijkheid, jouw relativerende vermogen (als ik weer eens verhit kwam binnenstormen) alsook jouw altijd gewillige oor (tot en met mijn konijnenverhalen aan toe) zijn de voornaamste impuls geweest de eindstreep te halen.

Mw. v.d. Meulen, beste Bonny, je bent een hartstikke gezellig mens! Dat eerste jaar met z'n allen, kakelend op een kluitje in het secretariaat als gevolg van chronisch plaatsgebrek, zal mij bijblijven als het beste AZU-jaar. Bedankt voor alle babbels met streepjesbonbons.

Ing. R. de Winter, beste Rinus, laat je niet kisten. Het werk liep je over, maar altijd was je bereid mij bij te staan in mijn computergestuntel. Zo mogelijk, onder het genot van een bakkie leut.

Beste Evelien, Vivienne, Anne, Merlijn, Lonneke, ... : geen overleven op het werk zonder sociale ruggensteun!

Dr. W. Kersing, beste Willem, bedankt voor jouw pianotalenten in de nachtelijke AZU-uren en het *even* willen participeren in het onderzoek. Na vele, vele plaatjes en /a:/-tjes, wie weet, misschien volgend jaar weer?

Dr. R.J.B. Hemler, beste Rafaël en lotgenoot, beiden toch een boekje verder mede dankzij herhaald 'werkoverleg'.

Stagiaires van velerlei aard, ook jullie wil ik bedanken voor jullie inzet. Succes met al die toekomstplannen.

Nieuwe collegae te Maastricht, bedankt voor de mogelijkheid de laatste loodjes vanuit het verre zuiden te kunnen volbrengen.

Vrienden en vriendinnen, om de hoek en verder weg: niets is belangrijker dan even kunnen bijkomen op de koffie, bij jullie of bij mij.

Aan het thuisfront. Het eind is gehaald, immers '**Alles is tijdelijk**'.
En nu maar weer eens (meer) gezellige dingen doen.

En alle anderen, niet expliciet genoemd maar toch aanwezig: ook U, bedankt!

Met de verdediging van dit proefschrift wordt een periode van zes jaar afgesloten. Zonder de hulp van velen was mij dit nooit, nooit gelukt.

Prof. Dr. P.H. Dejonckere, beste [illegible] initiator geweest van het hier beschreven onderzoek. Dankzij jou, werd mij de gelegenheid geboden te promoveren.

Dr. Ir. G.H. Wieneke, beste George, zonder jou was ik na drie maanden allang in Nijmegen ondergedoken. Je weet dat ik de samenwerking gedurende de afgelopen zes jaren enorm heb gewaardeerd en hoop je ook de komende tijd nog regelmatig tegen te komen. Met engelengeduld heb je geprobeerd mij de beginselen van wetenschappelijk onderzoek bij te brengen. Jouw persoonlijkheid, jouw relativerende vermogen (als ik weer eens verhit kwam binnenstormen) alsook jouw altijd gewillige oor (tot en met mijn konijnenverhalen aan toe) zijn de voornaamste impuls geweest de eindstreep te halen.

Mw. v.d. Meulen, beste Bonny, je bent een hartstikke gezellig mens! Dat eerste jaar met z'n allen, kakelend op een kluitje in het secretariaat als gevolg van chronisch plaatsgebrek, zal mij bijblijven als het beste AZU jaar. Bedankt voor alle babbels met streepjesbonbons.

Ing. R. de Winter, beste Rinus, laat je niet kisten. Het werk loopt je over, maar altijd was je bereid mij bij te staan in mijn computergestuntel. Zo mogelijk, onder het genot van een bakkie leut.

Beste Evelien, Vivienne, Anne, Merlijn, Lonneke, ... geen overleven op het werk zonder sociale ruggensteun!

Dr. W. Kersing, beste Willem, bedankt voor jouw pianotalenten in de nachtelijke AZU-uren en het even willen participeren in het onderzoek. Na vele, vele plaatjes en [illegible] wie weet, misschien volgend jaar weer?

Dr. R.J.B. Hamler, beste Rafaël en lotgenoot, beiden toch een boekje verder mede dankzij herhaald 'wandoverleg'.

Stagiaires van allerlei aard, ook jullie wil ik bedanken voor jullie inzet. Succes met al de toekomstplannen.

Curriculum Vitae

Curriculum Vitae

Renée Speyer was born in Nijmegen on August 9th 1963. She attended the Canisius College Mater Dei and obtained her diploma (VWO-B) in 1981.

In 1986 she graduated as a speech therapist ('Hogeschool voor Logopedie en Akoepedie' in Nijmegen) and exercised her profession in a private practice (until 2003). She combined these activities with a job at a rehabilitation center (1989 to 1992) and at the neurology department of a regional hospital (1995 to 1997).

In 1986 she started a study in educational pedagogics (Propaedeutic examination) followed by a study of Speech and Language Pathology at the University of Nijmegen (Main phase). She received her Master's degree in 1990. In 1997 she became affiliated to the Department of Phoniatrics of the University Medical Center Utrecht, where she carried out the research project which is reported on in this thesis. Currently she is working as a researcher at the Department of Otorhinolaryngology and Head & Neck Surgery at the University Hospital Maastricht.

Curriculum Vitae

Renée Speyer was born in Nijmegen on August 9th 1963. She attended the Canisius College Mater Dei and obtained her diploma (VWO-B) in 1981.

In 1986 she graduated as a speech therapist ('Hogeschool voor Logopedie en Akoepedie' in Nijmegen) and exercised her profession in a private practice (until 2003). She combined these activities with a job at a rehabilitation center (1989 to 1992) and at the neurology department of a regional hospital (1995 to 1997).

In 1986 she started a study in educational pedagogics (Propaedeutic examination) followed by a study of Speech and Language Pathology at the University of Nijmegen (Main phase). She received her Master's degree in 1990. In 1997 she became affiliated to the Department of Phoniatrics of the University Medical Center Utrecht, where she carried out the research project which is reported on in this thesis. Currently she is working as a researcher at the Department of Otorhinolaryngology and Head & Neck Surgery at the University Hospital Maastricht.

A 2040785