Effects of Voice Therapy: A Systematic Review

R. Speyer, *Maastricht, The Netherlands*

Summary. Medical as well as paramedical treatments should be evaluated by scientific methods. This systematic review focuses on the effects of voice therapy, excluding pharmacological or surgical treatments. In general, statistically significant positive but modest and varying therapy effects are found. Many of these effect studies cope with diverse methodological problems. Furthermore, the conclusions of most studies cannot be generalized easily or compared to one another. As a consequence, many issues in the field of effects of voice therapy have yet been unanswered.
Key Words: Systematic review–Dysphonia–Voice disorder–Voice therapy–Therapy effect–Therapy outcome.

INTRODUCTION

As it becomes more and more accepted that medical treatments should be evaluated by scientific methods, paramedical therapies as well need objective evaluation according to current standards of evidence-based medicine. Evaluation of voice therapy fits into this growing interest. The number of studies on the effects of voice therapy are still rather rare. In this article, a review of literature on the effects of voice therapy carried out by speech therapists is presented. Pharmacological or surgical treatments are not included. Some major methodological aspects and choice of evaluation tools of these studies will be discussed.

Methodological aspects

The design of an evaluation study depends on its purpose. The most simple design refers to the study of one specific therapy in patients with the same diagnosis under strictly controlled experimental conditions. It will be very unlikely that the therapy effects found in such a study can be generalized to other groups of patients or therapies. If the request for an evaluation study originates from a health care insurance or an organization responsible for health care budgets, the main focus will be the effectiveness of voice therapy in general. A study of this kind should include all possible phoniatric diagnoses—with an indication for voice therapy—as well as consider the diversity of existing voice therapies.[1–4] Most studies will be neither of these extremes, but will represent a mixed design as a compromise between these two options.

Therapy effects can be determined by applying exactly the same measurements before as well as after finishing therapy. To get objective results, no knowledge about the moment of data collection (before or after therapy), must be given to any judge when rating perceptual or visuoperceptual data (eg, perceptual evaluation of voice or visuoperceptual evaluation of videostroboscopy). Furthermore, results have to be compared using statistical analyses. Another issue is the inclusion of a group of patients that do not receive any treatment (placebo group). The results of the placebo group should be compared with the results of those who did have therapy. Sometimes, for ethical or practical reasons, no placebo group is included. In this case, another existing therapy could be used as control group. Which methodology should be used depends on the specific aim of the study. Besides group effects, the individual performances per patient can be of particular interest. Especially, when the patient population is inhomogeneous, therapy effects may be statistically scarcely significant for a whole group of patients, whereas the result can be quite diverse for subgroups of patients.

Multidimensional assessment tools

In literature, the success or lack of success of a voice therapy is assessed using different aspects of voice production. One of the main voice aspects described in literature is voice quality. Voice quality is described with terms such as breathiness, roughness, and harshness. Multiple systems of perceptual classification have been suggested by different authors: for example, the Buffalo Voice Profile,[5] the Vocal Profile,[6] the Grade, Roughness, Breathiness, Astenicity, and Strain (GRBAS),[7] the multidimensional model for voice production by Perkins,[8] the classification of voice qualities by Wendler,[9] and the SVEC.[10] However, perceptual evaluation involves problems such as the unstable internal standards for comparing speech stimuli[11] and the lack of universally accepted definitions for perceptual concepts.[12] 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 voice range profile, or phonetogram, describes the laryngeal possibilities with respect to the fundamental frequency and the sound intensity.[13,14] 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.

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 using 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 flashes at frequencies that differ slightly from the vibration of the vocal folds, the vibration of the vocal folds

Accepted for publication October 11, 2006.
From the Department of O.R.L. and Head and Neck Surgery, University Hospital Maastricht, Maastricht, The Netherlands.
Address correspondence and reprint requests to R. Speyer, Department of O.R.L. and Head and Neck Surgery, University Hospital Maastricht, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands. E-mail: rspe@skno.azm.nl
Journal of Voice, Vol. 22, No. 5, pp. 565-580
0892-1997/$34.00

doi:10.1016/j.jvoice.2006.10.005

is seen in slow motion. By means of visuoperceptual evaluation, the morphological and functional abnormality of the vocal folds and the glottal waveform can be described. Several protocols have been developed.[15–17] Recent advances in the technique of digital processing of laryngeal images have led toward the development of methods for deriving objective measures from such endoscopic examinations.[1]

Aerodynamic parameters such as maximum phonation time and the phonation quotient (the ratio of vital capacity and maximum phonation time) are widely used clinical measures. These measurements are inexpensive and simple methods for measuring the efficiency of the vocal fold vibration. 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. The demonstrated therapy effects, using the above-mentioned objective evaluation tools, must be compared with the beneficial or negative changes 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.

Many more instruments are available as alternatives for evaluating the vocal fold vibration in an objective way, such as electroglottography, photoglottography, or kymography. Aerodynamic measurements can be completed with averaged airflow measurement during phonation, and diverse measurements can be combined into indexes such as the Dysphonia Severity Index.[18] Usually, the perceptual evaluation of voice quality is considered to be the gold standard for voice assessment. However, it can be expected that patients will not show an abnormality in all aspects of voice, nor an improvement on all these aspects.[4] Voice must be regarded as a multidimensional phenomenon[19,20] and, therefore, the main aspects of voice must be considered when evaluating therapy effects. The Committee on Phoniatrics of the European Laryngological Society 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.[21] Of course, in such a study, the problem of increasing probability of significance has to be addressed.

In this article, a systematic review of the literature on the effects of voice therapy as applied by speech therapists will be undertaken.

METHODS

A literature search was carried out using the electronic databases Pubmed and Embase. All available inclusion dates up to February 2006 were used. The search was limited to English, German, French, Spanish, and Dutch language publications. In Pubmed, the Mesh terms *voice disorders*, *hoarseness*, and *aphonia* were combined with *therapy. Voice training* and the combination *voice* and *treatment outcome* were added. In Embase, the Mesh terms *treatment outcome* was completed with *dysphonia* and *larynx disorders* linked to *therapy.* To identify the most recent publications, the search was supplemented by using free text words (for the period after January 2005): *voice therapy* (Pubmed) and the combinations *dysphonia* or *voice* with *therapy* or *treatment* and *outcome* or *effect* (Embase). A total of 310 articles were found in Pubmed and 197 in Embase. Some articles were obtained from both databases.

Only articles on the effects of voice therapy in case of dysphonia carried out by speech therapists were included, thus, excluding pharmacological or surgical treatments as well as voice training in professional voice users (eg, Timmermans[22]). The search was restricted to therapy of dysphonia on a functional and/or organic base without any neurological origin such as Parkinson's disease. Review articles, case reports, and articles limited to populations smaller than five subjects were excluded. Studies that described only the posttherapy situation without comparable information on the voice status before the onset of therapy were considered of lesser importance and, therefore, excluded as well. The references listed in the selected papers were searched for additional literature. After a first selection based on abstracts, a definitive inclusion was made using the original articles. Finally, 47 studies were included.

RESULTS: THERAPY EFFECTS IN LITERATURE

Studies on therapy effects have become more frequent, especially during the last two decades. Tables 1A–C represent a summary of relevant scientific studies that describe aspects of the effects of voice therapy in dysphonic patients. Only studies that meet the above-mentioned inclusion criteria are listed (see Methods). The studies are classified into three main categories based on phoniatric diagnoses: functional dysphonia, organic dysphonia, and functional plus organic dysphonia (respectively, Tables 1A–C). The first column of the table represents the so-called level of evidence. To rate the study quality, the ABC rating scale according to Siwek et al[23] has been used. Level A refers to high-quality randomized controlled trials, whereas level B refers to well-designed, nonrandomized clinical trials. Level C, consensus or expert opinions, is excluded. These categories are subdivided into two groups according to the way data were handled. The first, largest group uses statistical analyses for comparing pre- *versus* posttherapy data. The second group uses descriptive statistics to evaluate the therapy outcome. Authors are listed in alphabetical order. For each study, the following data are summarized: the number of patients, the diagnostic group(s), the evaluation techniques, the kind of therapy used, and the author's key findings. The number of subjects refers to the group of subjects on which the study results are based, thus, excluding dropouts. Some articles contain extra study groups that fall beyond the purpose of this article. These groups are not mentioned in Tables 1A–C. Sometimes the primary purpose of a study is not to objectify the effects of voice therapy. However, if pre- and posttreatment data are present, the study is included. All 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.[24] Their population ($N = 6$) includes only two patients with