Critical Review: The Effectiveness of Voice Training in Preventing Vocal Pathology in Actors

Julie Cohn
M.Cl.Sc. (SLP) Candidate
University of Western Ontario: School of Communication Sciences and Disorders

This critical review explores the impact of training on the quality of actors' voices. Despite the wealth of information on the benefits of therapy for occupational voice users, little research has focused on the actor. Studies thus far have been inconclusive, or suggestive, in nature, using small sample sizes and non-uniform methods, lacking control groups and statistical power, or following treatment trends without defining specific research goals. Studies using better blinding, more consistent scheduling, randomization, and relevant control groups will help researchers discover the true impact of voice training on actors' vocal health.

Introduction

Professional actors are occupational voice users, yet their vocal needs differ drastically from other professionals; stage actors must often engage in vocal violence and give emotionally charged performances, often using a style of resonance, accent, prosody, and pitch focus different from their own. Actors frequently give unamplified performances to large audiences in acoustically impoverished venues, and must often work late into the night, on an irregular schedule.

The most common causes of vocal pathology in actors include: hyperfunction, glottal attack, extension beyond the voice's natural range, allergies, asthma, and the stress of volatile careers leading to tension (Brodnitz, 1954). These factors commonly lead to acute laryngitis, nodules, and contact granuloma (Brodnitz, 1954). In 2004, Laukkanen, Syrja, Laitala, and Leino defined vocal training as "an activity aiming through certain maneuvers – usually muscular functions – at changing the existing voice production related muscular functions in such a way that the result is an optimal voice and an optimal phonation" (p. 29). Stemple, Glaze, and Klaben (2000) purport that training should include teaching alignment, relaxation, articulatory isolation, forward resonation placement, abdominal and diaphragmatic breath support, easy onsets, encouraging an open upper vocal tract, and adequate hydration, as well as prescribed programs such as Vocal Function Exercises or Resonant Voice Therapy. In 2000, Roy, Ryker, and Bless noted that "although most voice teachers maintain that healthy vocal technique can preserve voice and prevent damage, to date there are no objective data to support this claim" (p. 216).

Objectives

The primary objective of this paper is to critically evaluate the existing literature regarding the vocal health of actors with voice training versus those without.

Methods

Search Strategy

Computerized databases were searched, including the following: Elsevier, Google Book Search, Google Scholar, and PubMed. The following key terms were used: (actor voice) AND (performer voice). The search was limited to articles written in English, but no limits were set regarding year of publication.

Selection Criteria

Studies selected for inclusion in this critical review paper were required to investigate the implementation of a specific voice training program on actors. Because of the small pool of evidence in this area, search strategies were kept broad in order to yield results, and no limits were set on the demographics of research participants or outcome measures.

Data Collection

Search results yielded three articles of the following types in keeping with the aforementioned selection criteria: non-randomized clinical trial (1), single group pre-posttest study (1), and randomized clinical trial (1).

Results

Timmermans, De Bodt, Wuyts, and Van de Heyning (2004) conducted a non-randomized clinical trial with a between-groups design in order to determine

if there was a difference between the qualities of actors' voices with versus without training. One group of 23 subjects received no voice training; the other group of 23 received training in vocal hygiene for nine months and voice training for 18 months. The voice training consisted of 90 hours of technical workshops and vocal coaching lead by teachers and an SLP, and lectures on breathing, articulation, voicing, and vocal hygiene. Each group of students was assessed before and after the training period. The researchers found that voice quality changed due to time and training, with training creating a more significant effect. Self-assessment did not seem to be affected by training. This study verifies the need for a well-organized voice training program for actors; however, the researchers do acknowledge the limited usefulness of the vocal hygiene lectures employed in this study (even afterward, actors were reluctant to make the recommended lifestyle changes). The researchers suggest that future studies should include measures of better vocal technique (breath support, laryngeal physiology, and speaking rate).

The researchers were very thorough with regard to the number and range of the scales they used to measure the students (good content validity). They were tested on a multidimensional test battery, including: the GRBAS scale (G = overall hoarseness, R = roughness, B = breathiness, A = asthenia, and S = strain), videolaryngostroboscopy, maximum phonation time, jitter, lowest intensity, highest frequency, Dysphonia Severity Index (DSI), Sound Pressure Level in dB(A), a vocal hygiene survey, and the Voice Handicap Index (VHI). Before the training, the students were also interviewed on their medical history, medications, and vocal hygiene; after the training, only vocal hygiene was reassessed. Each subject was assessed on the following measures at each assessment: investigator's perceptual evaluation, acoustic analysis, aerodynamic measurement, and self-evaluation.

Differences for various variables were found on the DSI and VHI between trained and untrained groups using a paired-samples t-test. Using a paired-samples t test was appropriate, given the two-group set-up of the study and the need to compare these groups on multiple measures. The researchers did not, however, make note of any corrections they may have made for multiple comparisons; in order to ensure the alpha levels did not add up, and to minimize the chance of a Type 1 error, some such correction would have been prudent. Because t tests are affected by sample size, however, results may have been skewed due to the relatively small sample size of the study; the researchers did not include this consideration in their analysis. The following variables were investigated: the potential interaction effect (i.e. whether or not the evolution in time was influenced by the training) (investigated using a two-way ANOVA repeated measures test, which was appropriate for examining potential interactions between time and training), the time effect (i.e. accounting for the time involved in the study and ruling that out as a factor) (investigated comparing subjects from one assessment to another, which was a valid way to compare, although researchers do not give details on which analysis tool was used), and daily habits (assessed using the Wilcoxon signed ranks test, which was used appropriately in that it looks for significant differences over time within each group).

Because the vocal hygiene information was taught and then abandoned for nine months before the trained group was re-assessed, potential gains could have been made and then subsided as time passed, so the construct validity is in question. However, this type of delayed testing more accurately reflects students' ability to retain the information over a longer time period, and is therefore more externally valid. Another major weakness was that the subjects were selected and broken into groups based on their school major (actors in one group, film directors in another). The results of this study are therefore quite weak with regard to validity, reliability (especially external), and clinical relevance, as directors do not have the same voice demands as actors, and thus cannot serve as a reliable control group. The groups were also likely very different in their vocal styles at the beginning of the trial. Also, neither subjects nor raters were blinded regarding these groups, making the results less valid. These factors render the study's results suggestive at best. Other weaknesses include the following: students were only reassessed using videolaryngostroboscopy if they were found to have structural anomalies during the first assessment (while it is an unlikely result, it would be important to know if any of the students had developed new problems after the training); results of the daily habits questionnaire may have been skewed (because the actors would have had an increase in vocal demands during a full-time acting school program, while the directors would not have had this increase); finally, the researchers do not delineate which subjects whose nodules resolved after training were in which groups, making analysis of training effects impossible.

Given the type of study conducted, and the significant changes due to training found in each testing area, this article is at evidence level 2. Due to the study's aforementioned weaknesses, however, the results are more suggestive than compelling.

In a 2000 single group pre-posttest (within-groups repeated measures) study of vocal violence in actors, Roy, Ryker, and Bless investigated whether or not there was a difference between the qualities of actors' voices with versus without Hygienic Laryngeal Release Training (HLRT). The goal of the study was to

examine the following: 1) pre- versus post-vocal violence (the "vocal violence effect"), 2) pre- versus post-training (the "training effect"), and 3) the interaction of 1) and 2) (the "interaction effect"). The procedure followed the following format: vocal measurements, vocal violence, vocal measurements repeated, training in HLRT, vocal measurements, vocal violence, vocal measurements repeated. The subjects' Modal, Minimum, and Maximum F0 were measured using: flexible laryngovideostroboscopy and Electroglottography (EGG). For the vocal violence portion of the study, the subjects were instructed to produce four violent vocal behaviours (grunting, groaning, sobbing, and shouting). The researchers found no significant vocal violence effect (comparing the before- and after- vocal violence portions of the study at each session), suggesting no damage will come from short-term vocal violence at a comfortable pitch. The researchers also found a significant interaction effect (vocal violence had a smaller effect on the acoustic signal after training), suggesting training can protect the voice from damage.

The recordings used in this study were analyzed on 18 different parameters. The results were then examined using a repeated measures analysis of variance (ANOVA), and then the SAS system's MIXED procedure with an unstructured covariance matrix. The former analysis was appropriate in that it allowed researchers to study the interaction effect of the two measures (vocal violence and training). The latter analysis allows for the inclusion of subjects with some missing values, so it was appropriate for the study if some values were, in fact, missing (the researchers do not give any further information on the issue). The researchers' vagueness with regard to missing values leaves the number of missing values unknown; if many were missing, the results may have been skewed, especially given the small sample size (27 subjects).

With regard to weaknesses of the study, no control or placebo group was used. Because there was no comparison group, results may have been due to a temporal effect alone, and are therefore merely suggestive; the authors do, however, acknowledge this limitation. In a study in which actors are actively working on the task to be measured it is difficult to blind the subjects, however, the report gives no indication regarding whether or not the testers and/ or trainers were blinded, leaving this measure of validity inconclusive. As vocal violence only leads to major damage over time, the fact that the researchers found no significant vocal violence effect only after a short time is suggestive but not compelling. Furthermore, as the researchers found no differences in vocal violence effect when comparing the before- and after- violence conditions, it is impossible to claim any positive influence of vocal training, or indeed that HLRT does not prevent immediate harm from vocal violence as the researchers did (since no damage was being incurred by the vocal violence anyway). This oversight calls into question the study's construct validity, and renders the results equivocal rather than suggestive or compelling. On many of the parameters, no significant differences were found, again leading to the results of this study being merely equivocal.

This study's main strength was its thoroughness. Three different pitch levels were studied, casting a wide net over the potential types of damage which may have occurred (good content validity), and yielding significant results which may have been overlooked in a more narrowly-focused study. Given these results and the type of study, this article is at evidence level 2. Although there were several weaknesses in the execution and analysis of this study, HLRT did yield significant improvements, and is therefore a useful and clinically relevant technique to consider when working with actors.

Laukkanen, Syrja, Laitala, and Leino (2004) conducted a randomized clinical trial with a between groups design investigating the effects of vocal exercising with and without spectral biofeedback (SB) on actors' voices. The goal was to determine if there was a difference between the quality of actors' voices with traditional training (TT) (which uses sensory feedback and imitation) versus visual biofeedback. Subjects were selected from an acting class and were randomly assigned to two groups. Subjects were recorded while reading a text sample. Both the SB and TT groups were then given the same instructions for vocalizing, but those in the SB group were also told to produce strong spectral components between 3000 and 5000 Hz without using excessive effort. The SB group watched a spectrum of their voices while completing the exercises. After the training, both groups were recorded again. The researchers then analyzed pitch (F0 and Long-Term-Average Spectrum), loudness (SPL), and voice quality (rated twice each by two novel professional voice trainers). For both groups, after training, SPL was higher, the relative loudness of F0 decreased, the spectral slope became less steep (indicating tighter and faster vocal adduction and an increase in sound energy), and voice quality was rated as higher for 10 of the 12 subjects. All of these results indicate that training was beneficial in vocal production. Additionally, the researchers found significantly tighter adduction in the SB group compared to the control group after training.

The researchers conducted an analysis of variance using a linear mixed effects model, which was appropriate given their desire to rule out the potential effects of SPL. The researchers were thus able to prove that variations in SPL did not single-handedly account

for changes in spectral parameters; regardless of SPL, the training had a significant, independent, positive effect on F0, frequencies of 3-4 kHz, and general voice quality (formant changes were not simply brought about by increasing loudness, therefore it was the training that improved voice quality). This analysis was critical in ruling out potentially confounding variables and ensuring good construct validity. ANOVA was an appropriate measure because it allowed researchers to study the interaction between training type and voice improvements based on several different parameters (F0, SPL, and voice quality).

There were several disadvantages of this study. The authors admit that the small number of subjects made conclusions tenuous. The pre-test results were not analyzed between groups with regard to how the subjects were performing (thus it is possible that there was a difference between the two groups before the training, even though they were randomized). No information regarding randomization methods was given. Due to the nature of this study, blinding the subjects would be extremely difficult (subjects had to be instructed regarding the vocal goals being studied in order to use the feedback they were receiving).

This paper has several advantages: the researchers were thorough in ensuring inter- and intra-rater reliability and increasing the study's validity by using two blind trainers, presenting the vocal clips in randomized pairs, and having each trainer rate the results twice. Subjects were also matched on experience level before the experiment; thus, it was ensured that all subjects had no prior training. While information was not given regarding the specific recruitment process for subjects, which allows for a potential selection bias, subjects were randomized into the two groups. Making these provisions lead to good construct, content, and statistical conclusion validity; the researchers ensured they were measuring what they had intended to, did so thoroughly, and used statistical techniques thoroughly to rule out other possible explanations for their statistically significant findings. Due to the type of study and the thoroughness regarding data recruitment and analysis that the authors exhibited, this article is at evidence level 1, and its findings are the most compelling of the three.

Discussion

Two of the three studies (Timmermans et. al., 2004; Roy et. al., 2000) were less than compelling because of their design. In order to be more compelling, future studies should use random assignment to implement a between-groups design using a control or placebo group in order for a valid comparison to be made. This type of study would more likely ensure that any changes were due to the specific training being studied, and not because of temporal or other external effects. Greater power will be gleaned from future studies with larger subject pools.

In future studies, it will be crucial to first identify the incidence of vocal pathology in actors, in order to prove the study is able to reduce incidences of vocal damage. It would also be useful to find situations in which vocal damage has been incurred, and then treat half of the group in which it was found in order to see if vocal damage could be reversed with adequate training. Without these comparison groups, it will be impossible to claim any positive influences of vocal training. For example, in the Roy et. al. (2000) study, no subjects incurred changes due to the vocal violence committed, so it was impossible to claim any positive influence of the training.

Each of the three studies used relatively short warm-up, training, testing, and follow-up sessions, which may not have been long enough for changes to emerge. In order to attain better validity, future studies should include longer sessions of each type and more testing sessions at longer intervals after the training. Following these measures will ensure that any potential changes due to voice training, vocal hygiene information sessions, HLRT, and/or Spectral Biofeedback intervention have been discovered, leading to more compelling evidence.

In order to ensure good content and construct validity, further studies would benefit from analyzing each aspect of training separately or by controlling for confounding variables in order to discern which processes were specifically linked to the changes found. This point pertains most to the Roy et. al. (2000) study, which assessed subjects on a wide variety of parameters, but conducted only general analysis, without enough data in each individual area to make specific conclusions. In the Laukkanen et. al. (2004) study, researchers took these efforts to control for SPL.

In order to be compelling and glean more clinical significance, further studies should also more closely replicate real-life rehearsal and performance conditions. For example, the Roy et. al. (2000) study conducted their analysis using sustained vowels only, which have different dynamic realities than the voicing and devoicing found in running speech. The study also assessed subjects while they stood against a leaning board, which provided resistance as the actors performed the vocally violent behaviours. As these boards are not used in a non-clinical environment in rehearsal or performance, they likely negatively affected the external and construct validity of this study making its findings less clinically relevant than they otherwise may have been. Further studies in this area should therefore include running speech and realistic physical conditions in order to be more externally valid and thus more clinically significant.

Clinical Implications

While there are many improvements to be made in this field of research and myriad unanswered questions, new and clinically relevant information has been uncovered in the three articles examined here.

The Timmermans et. al. (2004) article has shown that training is beneficial to voice quality; however, it has not yet been proven that training can affect self-assessment skills. Clinicians should therefore continue with their existing regimens for assisting clients to improve their self-assessment skills. Giving information regarding vocal hygiene has shown to yield limited success. This client reticence highlights the need for clinicians to be diligent in reinforcing the importance of vocal hygiene in order for actor clients to stay motivated when making these changes.

Roy et. al.'s (2000) conclusion that vocal training is significantly successful in protecting the vocal tract from damage at the ends of an actor's pitch range is an important and clinically relevant finding. Actors use these pitch extremes frequently, so even small changes at the upper ranges of pitch could lead to major consequences over time. Clinicians should therefore be vigilant regarding voice quality at high pitches in training in order to most efficiently prevent vocal pathology in actors.

Evidence from the Laukkanen et. al. (2004) study showed increased vocal effort levels in the Spectral Biofeedback group, and subjects reported that SB was interesting; these results indicate that clients may find this type of feedback motivating in clinical applications. Some participants in the study, however, found SB distracted them from auditory feedback; it may be that certain people do not benefit from SB. Clinicians could therefore use this tool as a trial method before committing the actor to using SB throughout therapy in order to ensure it is a helpful technique for that client. Higher levels of motivation could also lead to vocal hyperfunction, in turn impeding positive change. Using visual feedback could also lead to task/ goal confusion, as actors focus on external rather than internal cues. Overreliance on external cues could in turn lead to poor self-monitoring skills, poor transfer and carry-over, and even diminishing voice quality without external feedback, as actors will have to rely on themselves for feedback while on stage and in most rehearsals. Clinically, any improvements created by SB would therefore only be relevant, important, and useful if actors could maintain their internal feedback mechanisms. In this fashion, when the external reinforcers were removed, actors' voice quality and self-monitoring skills would not suffer, in keeping with the risk reduction/ Evidence-Based Practice model. Using other sensory feedback and careful monitoring concurrently with SB mitigates the problems of hyperfunction and overreliance on external cues, and will likely lead to better performance when clients need to transfer their skills to non-clinic environments. Given these points, SB is a valid tool for clinicians working in educational facilities and for actors developing their skills, but because rehearsal and performance situations rarely afford the time or resources for SB training, it is not likely feasible for clinicians working with professional organizations or actors already working in the field.

As studying actors and their voices is such a new area of the field, any new information or significant results involving analytically sound data will have clinical implications, if only to point clinicians in the right direction and motivate further research. Conducting further and more specific research on the topic of actors' voices is especially important because there are currently no established standardized protocols in the literature for avoiding vocal deterioration in this population. In keeping with the risk reduction/ Evidence-Based Practice model, garnering any new information regarding potential risk reduction for actors is crucial in helping them attain successful careers and lead healthy lives.

References

Brodnitz, F.S. (1954). Voice problems of the actor and singer. *Journal of Speech and Hearing Disorders*, 19 (3), 322-326.

Laukkanen, A.M., Syrja, T., Laitala, M., and Leino, T. (2004). Effects of two-month vocal exercising with and without spectral biofeedback on student actors' speaking voices. *Logopedics, Phoniatrics, Vocology*, 29 (2), 66-76.

Roy, N., Ryker, K., and Bless, D. (2003). Vocal violence in actors: An investigation into its consequences and the effects of Hygenic Laryngeal Release Training. *Journal of Voice*, 14 (2), 215-230.

Stemple, J.C., Glaze, L.E., and Klaben, B.G. (2000). *Clinical Voice Pathology: Theory and Management* (3rd Ed.). San Diego: Singular Publishing Group.

Timmermans, B., De Bodt., M.S., Wuyts F.L., and Van de Heyning, P.H. (2004). Training outcome in future professional voice users after 18 months of voice training. *International Journal of Phoniatrics, Speech Therapy and Communication Pathology*, 56 (2), 120-129.