# Mitigating Dysphonia, Pain, and Vocal Handicap after Violent Video Game Voice Overs: A Pilot Investigation into Vocal Combat Technique Training

*Katelyn Reid, *,†,‡Victoria S. McKenna, and §D'Arcy Smith, *†‡§Cincinnati, Ohio

**Summary: Objective.** to investigate how the Vocal Combat Technique (VCT) may mitigate vocal symptoms in voice over actors using vocal violence.

**Methods.** Five voice over actors (3 male, 2 females, Mean = 29.6 years) completed two study sessions of 45 minutes-to-1 hour of vocally violent voice over work held approximately 4 weeks apart. During session one, participants completed voice over work as they typically would, whereas they received 3-hours of VCT training to improve/assist in healthy vocal techniques prior to session two. Pre- and post-session measures for both sessions included self-perceptual ratings of vocal symptoms, auditory-perceptual evaluation, and traditional acoustic measures of frequency and perturbation.

**Results.** Participants showed substantial mitigating effects of VCT training on acoustic perturbation measures (jitter, shimmer, harmonics-to-noise ratio), and self-ratings of vocal symptoms (Vocal-Handicap Index-10, McGill Pain Scale for vocal discomfort, and Evaluation of the Ability to Sing Easily) with calculated medium to large effect sizes ($d = 0.61-1.95$). There were no changes in auditory perceptual ratings across sessions.

**Conclusion.** Our pilot investigation yielded positive improvements in vocal symptomology in five voice over actors who were trained in VCT. Next steps should include a larger enrollment of voice actors to determine optimal preventative and recovery techniques.

**Key Words:** Professional voice−Perception−Voice over−Video game−VCT.

## INTRODUCTION

For decades, teachers and vocal arts performers have been identified as at-risk of vocal injury due to the duration and intensity of their voice use.[1−3] More recently, a new occupational category has become a growing concern for the voice care community: actors who voice the soundtracks for "Action," "Shooter" and "Fighting" genres of video games. In many of these games, characters regularly encounter or perpetrate physical violence, requiring the actors who portray them to repeatedly vocalize intense physical exertion, danger, or pain.[4] The impact of these vocalizations on vocal health and wellness is not well established, despite the fact that the video game industry is growing at an exponential rate.[5,6]

For the typical vocal performer, phonotraumatic injury is the number one cause of dysphonia.[7] It follows then that the video game performer, who consistently engages in higher levels of vocally intensive behaviors than their peers, may be even more susceptible to vocal damage. Actors are required to work 2−4 hour long sessions, while shouting commands, screaming, growling, grunting and more. In part, due to these strenuous demands, video game performers initiated a strike in 2016. One performer interviewed for a radio broadcast stated her voice "is dead" after her recording sessions, causing loss of voice for up to one-week. She also reported that a co-worker needed "vocal surgery and will not be able to work for months."[8] Another news article discussing the strike stated "…actors said they've tasted blood in their throats during prolonged sessions. One actor fainted during an audition after screaming for too long."[9] Unfortunately, news articles, interviews, and conference presentations are the current evidence describing the problems in this industry, leaving a major gap in quantifiable, peer-reviewed research data on the full vocal implications of this violent voice over work.

To our knowledge, few have systematically studied the impact of violent video game recording sessions on voice over actors. A conference paper by Ogrodnik et al[10] examined the degree of vocal fatigue experienced by video game voice over artists using acoustic measures derived from neck-surface accelerometer signals. Their study revealed a decrease in Cepstral Peak Prominence (CPP) across the video game recording session, which is significant as an indicator for vocal fatigue as shown in prior research.[11,12] They also found, via analysis with the Acoustic Voice Quality Index (AVQI), that scores were pathologic up to 48-hours following the recording session in some participants. Although this study provided important quantitative data, it was limited because it did not investigate the subjective experiences of the performers, nor any methods to mitigate traumatic vocal effects. It then becomes critical to identify measures that quantify problems associated with this type

Accepted for publication June 10, 2022.

Presentations: This work was presented as an oral presentation at Fall Voice, October 18, 2019 in Dallas, Texas.

COIs: D'Arcy Smith is the developer of the Vocal Combat Technique. Katelyn Reid has been paid to provide lectures on the vocal mechanism and vocal health at Vocal Combat Technique training workshops.

From the *Department of Otolaryngology-Head & Neck Surgery, University of Cincinnati, Cincinnati, Ohio; †Department of Communication Sciences and Disorders, University of Cincinnati, Cincinnati, Ohio; ‡Department of Biomedical Engineering, University of Cincinnati, Cincinnati, Ohio; and the §Undergraduate Department of Acting, College-Conservatory of Music, University of Cincinnati, Cincinnati, Ohio.

Address correspondence and reprint requests to Katelyn Reid, University of Cincinnati, 3113 Bellevue Ave, Cincinnati, OH 45219. E-mail: grohkn@ucmail.uc.edu

Journal of Voice, Vol. ■■, No. ■■, pp. ■■−■■

0892-1997



https://doi.org/10.1016/j.jvoice.2022.06.015

of voice over work and develop preventative and restorative techniques for this group.

Training programs are one avenue sought by performers to hone their craft and reduce the likelihood of vocal trauma. Roy et al[13] investigated the impact of *Hygienic Laryngeal Release* on voice acoustics for stage performers required to perform vocally violent sounds. Their training program comprised typical techniques, referred to as "open track techniques," used by voice trainers to limit the potential damage incurred by stage performers. Open track techniques include warm-ups, warm-downs, muscular release, breath support, use of resonant voice and placement of pitch, and diversion of vocal effort away from the larynx. Their results showed modest evidence that vocal training helped to maintain and, at times, improve vocal stability following vocally violent sound productions; trainees showed reduced perturbation effects across their pitch range.

Although techniques described by Roy et al[13] seemed to help to mitigate negative effects on the voice for stage performers, these techniques do not necessarily translate to the demands of the gaming community. Open track techniques are known to place more value on safety rather than realism.[14,15] In the absence of a viable training protocol that both protects the actor and meets the industry's performance standard, individuals tend to approach these violent sounds for video games instinctively. This may be one reason for the rise in self-reports that these performances are to the detriment of their overall vocal health.[8,9] As such, there is a quantifiable need not only for more direct, occupation-based research on the impact of these violent video game voice over sessions, but also for the creation of customized training programs that will help address this issue and safeguard the long-term viability of this type of vocal professional. Our research begins to answer the call to action by Cazden,[4] who stated, "There is a further need for direct, occupation-based research on these vocally stressful voice over sessions. We who are committed to the care of the professional voice should begin to include these interactive media professionals in our scope of concern."

Therefore, we investigated a training program specifically aimed at safely producing the vocally violent acts required in video games to determine if specialized training would help to mitigate dysphonia and vocal symptoms. Referred to as Vocal Combat Technique (VCT), this formalized 3-hour vocal training program was specifically adapted for voice over actors. It instructs performers in "open tract" techniques appropriate for stage work, but also incorporates specialized training for the more aggressive and "realistic" sound productions required by the video game industry. In part, it utilizes the surrounding structures of the vocal tract (eg, supraglottic structures) to increase the ability to produce desired grunts, growls, and screams, while limiting phonotraumatic behaviors. The method of using supraglottic structures to add turbulence/roughness and volume to a sound has been researched in rock singers and theater actors, with preliminary evidence that that use of these structures does not result in laryngeal pathology and as such can support the production of more agressive sounds safely.[16,17]

The present study describes a pilot investigation into the potential impacts of VCT on self-perceived vocal symptoms, voice acoustics, and auditory-perceptual ratings of vocal quality. The study design included two separate video game recording sessions: the first session was completed without any prior vocal training, and the second session was completed following VCT. Pre and post testing were completed immediately before and after each session as well as 48-hours afterwards in order to capture any lingering vocal symptoms and determine the performer's readiness and eagerness to return to work. With this design, each participant acted as their own control and also provided foundational evidence of the impact of voice over work on clinical measures. We hypothesized that VCT training would decrease the degree of symptoms of vocal fatigue and pain experienced by participants, reduce acoustic and auditory-perceptual indicators of dysphonia, and increase the actor's ability and eagerness to return to work when compared to no VCT training.

## METHODS

### Participants

The institutional review board at the University of Cincinnati approved this study and all participants provided informed, written consent prior to participation. Although 7 voice over actors were initially enrolled in the study, only five completed the entire protocol (two female, three male, M = 29.6 years, SD = 12.4 years, range = 21−54 years). Therefore, the analysis refers to only these five participants here forward. Participants were required to be fluent in English, have no history of prior or present vocal injury, and no prior instruction in VCT. All participants self-reported working as professional actors an average of 2.9 years (range = 1−5 years). None of the participants worked in the video game industry prior to this study, which was not unexpected due to lack of gaming industry presence in the region of their recruitment.

### Protocol

Subjects underwent two separate, 45 minutes to 1-hour long, Simulated Video game Recording Sessions (SVRS) conducted in a sound treated room at the University of Cincinnati College-Conservatory of Music. These sessions were virtually guided by an industry professional with over 25 years of experience. Typical vocal tasks included grunting, growling, shouting, coughing, crying, screaming, and panting, for a total of 80 different target utterances, each repeated multiple times. The professional leading the session recorded the number of attempts at each target and determined when the task met the appropriate industry standard.

The first session was recorded prior to undergoing VCT training (referred here forward as session one). Three weeks after the initial session, the subjects underwent 3-hours of instruction in VCT, led by author D.S. This method trains voice over artists to vocalize aggressive/intense sounds with upper thoracic expansion to counteract vocal tension, compression, and strain. It also utilizes positioning and vibration of supraglottic structures for the addition of roughness or turbulence to the sound. This training was followed by another week-long rest period before the second session of SVRS was recorded (referred to as session two). Therefore, the time between sessions one and two of SVRS was 4 weeks.

### *Measures*

Prior to each SVRS, baseline voice recordings and self-ratings were obtained. Participants completed sustained vowels of /i/ and /ɑ/ at a comfortable pitch and volume, and sentences from the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V).[18] Acoustic recordings were acquired with a RøDE NTK condenser microphone at a sampling rate of 44.1 kHz. All acoustic recordings were made at a microphone-to-participant distance of 15 cm.

Participants completed the ten question Vocal Handicap Index (VHI-10).[19] The VHI-10 has the user rate perceived vocal handicap on daily activities on a 5-point Likert scale (0−4). A rating of “0” indicates that the prompt is “never” a problem, whereas a rating of “4” indicates that it is “always” a problem. The responses of the ten questions are summed to a total score in which a higher total score indicates a greater vocal handicap. Participants also completed the Evaluation of the Ability to Sing Easily (EASE) questionnaire.[20] The EASE includes 22 questions targeting the subtleties of a vocal performer’s voice and the changes to the mechanism as a function of vocal load. Ratings are made on a Likert rating scale in which a score of “one” indicates they experience the problem “not at all” and a score of “four” indicates that it is an “extreme” problem. For the purposes of this study, some questions were slightly modified for the task at hand. For example, the statement “I am having difficulty singing softly” was modified to “I am having difficulty speaking softly.” The EASE consists of three subscales, including 10-questions on vocal fatigue (VF), 10-questions on pathologic risk indicators (PRI), and two questions on vocal concern (VC). The 20 questions across EASE-VF and EASE-PRI can be summed for a total EASE score (EASE-total), whereas information on EASE-VC is interpreted separately.

Immediately following each SVRS the same acoustic recordings were captured and participants were once again asked to complete the EASE. Additionally, they completed the Short-form McGill Pain Questionnaire (SF-MPQ).[21] The SF-MPQ rates pain symptoms such as “throbbing,” “stabbing,” and “aching” on a scale of 0 = none, 1 = mild, 2 = moderate, and 3 = severe. Participants were instructed to rate their amount of discomfort related to their voice use for all 15 questions, which were them summed to a total score. Finally, a separate Procedure Pain Index question has them rate their degree of pain with 0 = no pain, 1 = mild, 2 = discomforting, 3 = distressing, 4 = horrible, and 5 = excruciating.

A brief follow-up questionnaire was sent out 48-hours after each SVRS. This consisted of the VHI-10 again as well as specific questions related to their recovery and perceived interest and readiness to complete voice over work again. The three questions were as follows: Q1- “How much difficulty are you having returning to your daily vocal activities?”; Q2- “How likely would you be to sign-up to record another voice recording session today?”; and Q3- “I am confident in my ability to safely record a video game voice over recording session.” Questions were rated on a 1−4 Likert range in which a low score was described as “none or not at all” and a higher score was anchored as “an extreme amount or extremely.”

## Data processing

### *Self-reported measures*

Scores from the EASE, VHI-10, SF-MPQ, and questions related to persisting levels of vocal impairment and readiness were transferred to an excel spreadsheet.

### *Auditory-perceptual measures*

CAPE-V recordings were de-identified and blinded for session (1 or 2) and timing (pre- or post-session). A total of 20 recordings (5 participants × 2 sessions × 2 recordings) were presented in randomized order to certified speech-language pathologist (SLP) with more than 10 years of clinical and research experience in voice. Acoustic recordings were presented in a quiet room. Loudness level was set by the SLP and she was allowed to listen to each recording as many times as she wished. CAPE-V ratings of all vocal percepts (eg, overall severity, roughness, breathiness) were made by placing a single line on a horizontal visual-analog scale (VAS). These were measured by the rater to the nearest millimeter (mm) and placed into an excel.

### *Acoustic measures*

Traditional voice acoustic measures were extracted using Praat (ver. 6.1.25).[22] Two trained technicians completed reliability training prior to data extraction in which they met a benchmark of >90% inter-rater reliability (intra-class correlation coefficient [ICC]) for each acoustic measure. Next, each technician extracted data from two participants (2 technicians × 2 participants = total of four participants). Then both technicians extracted acoustic information from the last participant, blinded to each other’s extractions. This last extraction was used to calculate inter-rater reliability for our sample. Reliability was calculated separately for each acoustic measure, resulting in an average ICC = 0.92 (range of ICC = 0.83−0.99), indicating good-to-excellent reliability across the acoustic measures.[23]

Acoustic extraction was completed using the “Voice Report” function in Praat. The targeted measures were mean fundamental frequency ($f_o$), jitter (%), shimmer (%), sound pressure level (SPL; dB), and harmonics-to-noise ratio (HNR; dB). Measures were extracted from steady-state portions of vowel productions from both the sustained vowels as well as vowels produced during continuous speech in the CAPE-V sentences. For a complete list of the words targeted for analysis, please see Table 1. The steady-state portion of the vowel was determined by viewing both the acoustic waveform as well as a spectrogram with formant tracking enabled. This view helped to determine when transitions to other phonemes occurred and reduce variability of coarticulation from extracted vowels.

### Statistics

A formalized statistical analysis was not completed due to the small sample size of our pilot study. Instead, descriptive summaries are reported in tables and graphical form. Cohen's *d*, calculated as ($M_1$-$M_2$)/pooled SD,[24] was determined to provide evidence of the size of the changes between sessions one and two, as needed. Finally, a sample size calculation was computed to provide guidance for the number of participants that are needed to detect significance in follow-up investigations.

## RESULTS

### Self-reported measures

Participants completed the EASE questionnaire immediately before and after each voice over session. The EASE-total as well as its two subscales (EASE-VF, EASE-PRI) showed large effect size differences ($d$ = 0.97−1.07) in the amount of change observed across each session (see Table 2 for values and effect sizes). The EASE-VC, a subscale with only two questions regarding vocal concern, showed limited change with only a small effect size.

The VHI-10 was captured prior to each session and then 48-hours later. One participant did not have a post VHI-10 score and was excluded from this analysis. The baseline VHI-10 scores for each session were consistent at an average of 4.75 (session 1) and 4.80 (session two), showing similar starting points at each session. Following session one, VHI-10 increased to an average of 11.00, whereas following session two, it maintained at a score of 4.40. When comparisons were made of the change across sessions, a large effect size ($d$ = 1.95) was observed.

Following each session, participants completed the SF-MPQ for vocal discomfort. Figure 1 provides a boxplot distribution of responses. The change from session 1 (M = 12.2) to session 2 (M = 8.2), yielded a medium-to-large effect size ($d$ = 0.72). The Procedure Pain Index, a single question asking about the degree of pain on a 0−5 Likert scale, noted an improvement following session two (M = 1.4, mild pain) compared to session one (M = 2.8, discomforting-to-distressing pain).

Finally, participants completed three questions on persisting symptoms and readiness for future work. One participant did not complete a post-assessment for session 1,

**TABLE 1.**
**Targeted Vowels During Sustained Productions and Continuous Speech Used to Extract Voice Acoustic Measures**

| Vowel | Stimuli |
|---|---|
| **/ ɑ /** | • Sustained /ɑ/ × 5 sec. Repeated two times. |
| | • Vowel in "spot" from the sentence "The blue spot is on the key again." |
| | • First vowel in the word "mama" from the sentence "My mama makes lemon muffins." |
| /i/ | • Sustained /i/ × 5 seconds. Repeated two times. |
| | • Vowel in "key" from the sentence "The blue spot is on the key again." |
| | • Vowel in "he" from the sentence "How hard did he hit him?" |
| | • Vowel in "eat" from the sentence "We eat eggs every Easter." |
| | • Vowel in "keep" from the sentence "Peter will keep at the peak." |

**TABLE 2.**
**Mean (SD) of Self-ratings for the EASE and VHI-10**

| Session | Timepoint | EASE (Total) | EASE-VF | EASE-PRI | EASE-VC | VHI-10 |
|---|---|---|---|---|---|---|
| 1 | Pre | 30.40 (6.89) | 16.80 (3.87) | 13.60 (3.38) | 2.60 (0.80) | 4.75 (7.16) |
| | Post | 45.00 (4.52) | 25.40 (2.42) | 19.60 (2.33) | 4.00 (1.79) | 11.00 (5.34) |
| | Change | 14.60 (7.79) | 8.60 (4.45) | 6.00 (3.90) | 1.40 (1.74) | 4.50 (2.60) |
| 2 | Pre | 29.20 (4.12) | 16.00 (2.45) | 13.20 (2.64) | 2.40 (0.80) | 4.80 (6.49) |
| | Post | 35.20 (6.05) | 20.60 (3.50) | 14.60 (2.58) | 3.20 (1.60) | 4.40 (4.50) |
| | Change | 6.00 (8.24) | 4.60 (3.72) | 1.40 (4.84) | 0.80 (1.60) | -0.40 (2.42) |
| Effect Size | Cohen's *d* | 1.07 | 0.97 | 1.05 | 0.36 | 1.95 |
| | | Large | Large | Large | Small | Large |

The EASE has three sub-tests including the EASE-VF (Vocal Fatigue), EASE-PRI (Pathologic Risk Indicators), and EASE-VC (Vocal Concern). The EASE-total is a sum of both the EASE-VF and EASE-PRI. Effect size is calculated from the change observed for each session

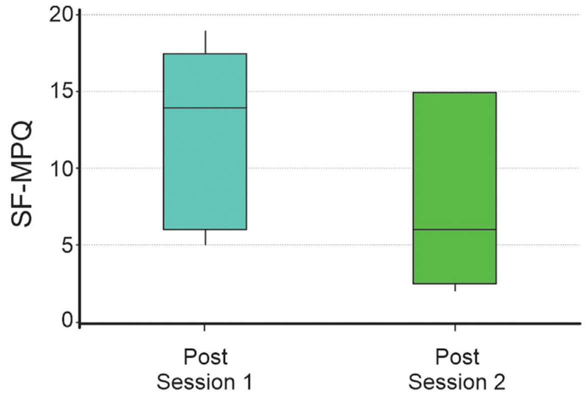


**FIGURE 1.** Boxplot of answers from the short form McGill Pain Questionnaire (SF-MPQ) taken immediately after the voice over session. Lower scores represent lower degrees of vocal pain, whereas higher scores are indicative of greater degrees of reported pain. The effect size difference between session 1 and session 2 was $d = 0.72$.

leading to only 4 responses for this session. A summary of questions and responses can be found in Figure 2. Overall, the data seemed to trend toward improvement in vocal symptoms and increased willingness to participate in voice over work, with concurrent reduction in variability of the responses. Effect sizes were not calculated for individual questions.

### Auditory-perceptual ratings

CAPE-V ratings, in general, were higher prior to initiating session 1 compared to when beginning session 2. For example, the overall severity rating for pre-session one was an average of 14.6 mm, whereas it was 8.6 mm for pre session two (see Table 3). Nevertheless, all average CAPE-V averaged values were low across all percepts, indicating normal auditory-perceptual vocal quality across all timepoints. Further, there was limited change across any percept; at most, change occurred up to an average of 4 mm. Change was not consistent in either direction for either session, meaning that some scores increased (became worse) in the post session analysis, whereas some decreased (became better). Effect sizes (Cohen's $d$) were not calculated for CAPE-V ratings, as visual inspection of changes pre/post session were deemed negligible and representative of normal vocal quality.

### Acoustic measures

During session 1, jitter and shimmer both increased to a greater degree than in session two. The degree of change between sessions for jitter and shimmer showed large ($d = 1.02$) and medium ($d = 0.61$) effect sizes, respectively. Conversely, HNR decreased during session 1, indicating a reduction in harmonic content in the vocal signal, whereas it increased in session two, possibly due to improvement in voice quality or concurrent increase in intensity (ie, SPL). The degree of change across sessions showed a large effect size ($d = 0.80$) for HNR. All averaged session values, change within session, and effect sizes can be found in Table 4.

There appeared to be no change in mean $f_o$ during session one, but an increase in mean $f_o$ across session two of 6.83 Hz, yielding a medium effect size; however, this increase in mean $f_o$ is likely not clinically meaningful. SPL increased by 1.34 dB and 1.75 dB for sessions one and two, respectively. The degree of difference between these was small with $d = 0.27$.

### Voice over director ratings

The director of the SVRS had each participant repeat 80 separate targets utterances (eg, "Grenade!", "Where is he?!") for a set number of repetitions (2−4 depending on the utterance), totaling 284 trials. All participants completed all trials and the director rated each target's acceptability. The total number of acceptable targets was divided by the total number of possible targets to provide a percentage of acceptability. On average, participants had 47% of utterances deemed acceptable in session 1 (range 27−70%), and an

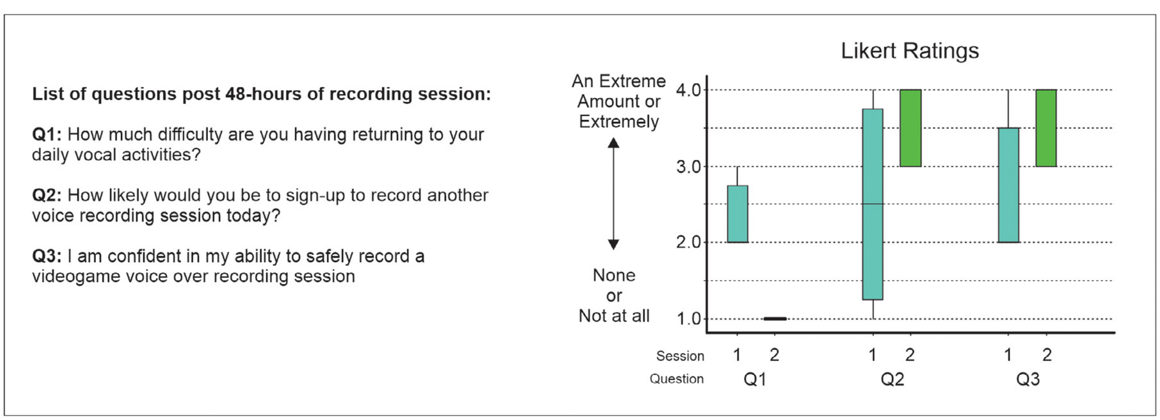


**FIGURE 2.** Boxplot of Likert-ratings taken 48-hours after each voice over session.

**TABLE 3.**
**Mean (SD) of CAPE-V Ratings (mm) Pre- and Post-sessions 1 and 2**

| Session | Timepoint | Overall Severity | Roughness | Breathiness | Strain | Pitch | Loudness |
|---|---|---|---|---|---|---|---|
| 1 | Pre | 14.6 (5.0) | 9.0 (6.8) | 10.7 (6.2) | 6.5 (3.7) | 0 (0) | 0 (0) |
| | Post | 16.3 (7.3) | 8.8 (7.9) | 12.3 (5.8) | 10.7 (7.9) | 1.7 (3.4) | 0 (0) |
| 2 | Pre | 8.6 (3.3) | 4.1 (4.4) | 5.2 (3.7) | 5.2 (4.1) | 1.3 (2.6) | 0 (0) |
| | Post | 10.8 (7.9) | 3.0 (1.6) | 5.8 (3.5) | 7.7 (9.4) | 1.1 (2.2) | 0 (0) |

average acceptability of utterances of 73% in session 2 (range 64%−90%).

## Sample size calculation

A sample size calculation was performed in G*Power, ver. 3.1.[25] The effect sizes of the perturbation measures (jitter, shimmer, HNR) and perceptual measures (EASE, SF-MPQ, VHI-10) showed a range from medium to large (ie, $d = 0.61-1.95$). With this information, we took a conservative approach of a medium effect size ($d$ = 0.61) for the sample size calculation. With the statistical plan of a paired, one-tailed $t$ test, power of .80, significance of $P < 0.05$, a sample size of 19 subjects is needed in a follow-up study to detect significance. Should a control group of subjects be used as a comparison, a two-sample $t$ test with the same input parameters would result in the need for 34 subjects in each group, or 68 subjects in total.

## DISCUSSION

Voicing for violent video games is one of the most vocally demanding jobs a performer can undertake. Without effective training for healthy production of these aggressive sounds, performers are frequently putting themselves at risk of vocal injury in the name of authenticity. In this study we sought to determine the performer's perception of the experience, listener perception of vocal quality, and objective changes to voice acoustics. We also investigated whether VCT could mitigate some of the negative effects of these sessions and even increase a performer's work readiness and confidence towards safely completing subsequent recording sessions.

First, we hypothesized that VCT would help to mitigate dysphonia, which we quantified through changes in voice acoustics and auditory-perceptual ratings. Our hypothesis was partially supported when the acoustic perturbation measures showed medium-to-large effect size differences when comparing the change within each of the sessions. Specifically, there was a smaller amount of change for jitter in session two (Pre = 0.80%, Post = 0.84%, Change = 0.04%) following VCT training, compared to the session one (Pre = 0.69%, Post = 0.98%, Change = 0.29%), as well as smaller amount of change for shimmer in session two (0.18%) compared to session one (0.65%). Jitter and shimmer are measures of periodicity and amplitude variation, respectively, with greater values indicating higher rates of instability. Previous research has shown that jitter and shimmer can be indicative of overall severity of perceived dysphonia.[26−28] Therefore, the smaller change in session two indicates that actors were able to maintain their typical speaking voices and experienced less dysphonia over the course of the session. Roy et al[13] reported reduced shimmer, amplitude perturbation quotient, and voice turbulence index following vocal training during sustained modal $f_o$, similar to our findings here.

We also found that HNR decreased following session one (pre = 17.08 dB, post = 15.91 dB), indicating a worsening of the voice acoustic signal. Interestingly, HNR in fact slightly *increased* following session two (pre = 15.94 dB, post = 16.03 dB), showing a stability and slight strengthening of harmonic content. The difference between these

**TABLE 4.**
**Mean (SD) of Acoustic Values Before and After Sessions 1 and 2**

| Session | Timepoint | Mean $f_o$ (Hz) | SPL (dB) | Jitter (%) | Shimmer (%) | HNR (dB) |
|---|---|---|---|---|---|---|
| 1 | Pre | 164.29 (60.84) | 73.46 (1.67) | 0.69 (0.23) | 5.34 (1.77) | 17.08 (2.60) |
| | Post | 164.50 (63.91) | 74.81 (1.47) | 0.98 (0.47) | 5.99 (2.28) | 15.91 (2.49) |
| | Change | 0.21 (8.94) | 1.34 (1.69) | 0.29 (0.33) | 0.65 (0.90) | -1.17 (1.34) |
| 2 | Pre | 166.44 (62.87) | 73.02 (1.30) | 0.80 (0.14) | 5.11 (0.83) | 15.94 (0.83) |
| | Post | 173.26 (72.87) | 74.77 (1.27) | 0.84 (0.23) | 5.30 (1.25) | 16.03 (1.64) |
| | Change | 6.83 (12.66) | 1.75 (1.29) | 0.04 (0.11) | 0.18 (0.57) | 0.09 (1.80) |
| Effect Size | Cohen's $d$ | 0.60 | 0.27 | 1.02 | 0.61 | 0.80 |
| | | Medium | Small | Large | Medium | Large |

The average change and its SD were also calculated within session. Effect size (Cohen's $d$) was determined from the average change between sessions 1 and 2 resulted in a large effect size. Still, HNR values must be interpreted in the context of other vocal behaviors. Specifically, HNR is closely linked to vocal intensity, in which an increase in intensity may act to artificially inflate HNR.[29] Our data showed an increase of 1.75 dB SPL following session two, which could have also led to this slight increase noted here. However, it should also be noted that there was a similar increase in vocal intensity following session one, but HNR values fell, even with the artificial bump that increased vocal intensity may have had. Knowing this, we suspect that VCT helped the actors to maintain their harmonicity (as also evidenced by their small changes in jitter and shimmer) and that coupled with an increased intensity, helped to maintain HNR values over session 2.

Our acoustic findings are consistent with previous studies.[10,13] Ogrodnik et al[10] found that CPP, a measure of harmonic strength and periodicity,[30] decreased over the course of a voice acting session. The measures used to assess perturbation in our study (jitter, shimmer, and HNR) also indicated a decrease in periodicity of the signal following the untrained session (session one). Like Roy et al[13] who reported no modal $f_o$ changes in stage actors following vocal violence, we also found no meaningful change in our measure of mean $f_o$ across sessions. However, Roy et al[13] did report an increase in $f_o$ range due to an increase in max $f_o$ post vocal violence (increase of approximately 20−50 Hz). The authors hypothesized an increase in $f_o$ range may have been caused by a physiological "warm up" factor from vocally violent sounds or a possible reflection of pretest anxiety and increased laryngeal tension that resolved once the task was completed. These acoustic measures may be an area of inquiry in vocal training studies moving forward.

Unlike our acoustic findings, our hypothesis regarding auditory-perceptual ratings of dysphonia was not supported by our analysis. That is, all CAPE-V values across all percepts were deemed "normal" and did not change in any consistent direction (improve or worsen) across either session. This result though, is not surprising. It is notoriously difficult for a listener to perceive additional somatosensations experienced by the speaker. For example, experiences of vocal fatigue and laryngeal pain, often quantified as percepts of vocal effort and strain, result in some of the most unreliable listener estimations of severity[31] and inconsistent correlations between listener and speaker ratings.[32−35] There are many symptoms someone may be experiencing that do not manifest as dysphonia to a trained clinical ear, which is a consistent problem noted across clinical literature.

Conversely, there were substantial mitigating effects of VCT training on self-reported vocal symptoms. Nearly all participants had increased VHI-10, EASE and SF-MPQ scores when they performed the SVRS without prior training. When VHI scores were compared across sessions, a large effect size was demonstrated. Pre training participants went from a baseline average of 4.75−11.00, with 11 being the cutoff point for within normal limits on the VHI-10.[36] Following VCT, VHI-10 scores decreased from 4.80 pre session to 4.40 post-session. This indicates that a significant increase in perception of vocal handicap occurred without training, whereas participants were able to maintain if not decrease their handicap perception post training.

The EASE questionnaire, created to assess the effects of vocal load on vocal performers, also demonstrated a large effect size for change across sessions on two of their subscales (EASE-VF and EASE-PRI) and total EASE score which is the sum of these two subscales. The EASE-VF subscale questions correspond to physical symptoms of vocal fatigue (eg, My throat muscles are feeling overworked; My voice is tired). With similar baseline average scores prior to sessions one and two (16.80 and 16.00, respectively) for this subscale, significant change across sessions ($d$ = 0.97) was noted with decreased average scores observed post training (20.60) as compared to pre training (25.40), indicating that VCT has the potential to mitigate the vocal fatigue experienced by voice over actors engaging in violent sound production. Similarly, EASE-PRI questions, which correspond to the presence of edema or vocal pathology (eg, My voice cracks and breaks; I am having difficult speaking softly,) also demonstrated an even larger effect size for change across sessions ($d$ = 1.05). From once again similar average baseline measures (13.60 pre, 13.20 post training), scores increased by an average of 6 on this subscale without training, compared to an average increase of 1.4 with training. This indicates that performers were experiencing more symptoms of vocal edema and possible injury when they performed the session without training. This result is of particular significance due to the anecdotal evidence that voice over artists have injured themselves undertaking these vocally violent recording sessions. Preliminarily, this indicates that VCT has the chance to decrease the amount of inflammation and potential injury to the larynx, which is in support of our hypothesis. The two EASE-VC questions corresponding to a performers concern for their voice also supported our hypothesis, but with only a small effect size ($d$ = .36) for change across sessions.

Due to the intensity of the violent sounds being produced by this population and the anecdotal reports of strain and injury, we also administered the SF-MPQ to assess the level of pain experienced during these recording sessions. Once more, participants' ratings improved with training showing a medium to large effect size across sessions ($d$ = 0.72). The Procedure Pain Index aspect of the SF-MPQ which asks for a rating of the total pain experienced during the session revealed participants felt "discomforting-to-distressing" levels of pain when performing the session without training. Post training in VCT, performers reported pain as only "mild." It is of note that though VCT did decrease the overall pain experienced, the training did not completely eliminate it. Given the nature of the work involved, some degree of discomfort may be unavoidable for voice-over actors in the video game industry. However, the efficacy of even a single session provides hope that vocal discomfort may be significantly reduced through further practice and increased proficiency with VCT.

To examine how these types of intense sessions may affect the stability of a voice over actor's career and life outside of their work, we developed three questions targeting ease of returning to normal daily vocal function, readiness to return to work to record again, and comfort with performing the required recording content safely. Our results revealed that they had some difficulty returning to normal voice use following session one. After training in VCT, all participants unanimously stated they had no difficulty returning to normal voice use afterwards, and most of them stated they felt more ready to perform a subsequent recording and were more confident in their ability to do so safely.

The results of the formalized questionnaires and follow-up questions are especially compelling and provide a foundation of research-based data describing the experiences of voice over actors undergoing voice changes. Additionally, our study only required a 1-hour SVRS, a quarter to a half of the industry standard duration for a session, and performers had increased acoustic measures of dysphonia, experienced symptoms of vocal fatigue and possible injury up to 48-hours afterwards and felt discomforting to distressing levels of pain post performance. This further highlights the need for the development and implementation of training programs and preventative techniques for this population to have viable and sustainable careers in this rapidly growing industry.

In order to fully assess the viability of VCT for this population, a Voice Director active in the video game industry guided the sessions and determined whether an actor's production met the industry standard. Our analysis of the Voice Director's feedback showed that not only does VCT mitigate perceived vocal fatigue and potential damage, but it also improved the quality of the target sounds performed during the SVRS. Prior to training, the participants acceptable productions averaged to 47%, whereas after training, acceptability increased to 73%. This finding shows that beyond mitigating dysphonia and discomfort, VCT may in fact aid in a more successful production of the target sound.

## Limitations and future directions

Our results should be interpreted in the context of the limitations of our study. First and foremost, this study was small with a total of five participants completing the entire protocol. We have calculated and proposed the number of subjects needed for follow-up endeavors. These calculations include a potential of 19 subjects if the design of the study were to remain the same, in which each subject acted as their own control (ie, each participant completed session one without any previous training). However, with this design, it cannot be stated for certain the VCT alone increased the participants' accuracy with the target sounds, as repeat exposure to the task may also have been responsible. Therefore, a control group of those who have not received VCT should be added to mitigate repeated sampling and learning effects. Additionally, blinding could be added to the director so that they do not know if the performer has received VCT or not.

With a larger group of subjects, it may also be beneficial to vary the experience of the vocal performers. Our participants had an average of only 2.9 years of voice over performance and no prior video game recording experiences. It is possible that their lack of experience and familiarity with what the SVRS included may have contributed to the drop out of two performers after completing only the first SVRS. A longitudinal study to determine how preventative techniques could improve attrition in this field may also be beneficial as it is possible voice over actors are not remaining consistent with their work due to the discomfort and vocal effects they are experiencing when performing them. There is also the potential for the application of VCT for other non-performance related occupations that involve heavy use of intense or aggressive vocalizations and elevated rates of dysphonia, like fitness instructors[37−39] and military drill instructors.[40]

Other limitations include that the EASE questionnaire is validated for singers and six of its questions specifically target that type of vocal production, which may or may not have felt relevant to any participants who were not vocalists. Nevertheless, we chose to use this questionnaire because we did not feel there was another significantly sensitive, validated questionnaire for rating the subtle aspects of dysphonia relating to vocal load that may be present in a vocal performer's voice. Future research on VCT could further explore the effects of vocal load by utilizing the Borg Category Ratio 10 (CR-10) to measure vocal effort.[41] This scale has been shown to elucidate aspects of vocal effort not targeted by the P5 question on the VHI-10, "I feel as though I have to strain to produce voice,"[19] which targets the frequency of the straining and not the severity.[42] If this questionnaire were to be administered following some of the more aggressive sound productions (eg, video game tasks vary from running sounds to being set on fire), a greater understanding of the treatment's specific effect on vocal effort could be obtained.

Finally, we limited our acoustic analysis to more traditional acoustic measures. However, Ogrodnik et al[10] showed a change in CPP across sessions, perhaps indicating a sensitivity with this particular measure. CPP has also been shown to be sensitive to vocal loading and fatigue in controlled experimental paradigms.[11,12] Likewise, Roy et al[13] examined additional measures of $f_o$ (range, max) that could provide insights into vocal edema and tension. Therefore, a larger set of acoustical measures could be examined in follow-up works.

## Conclusion

Our study is one of the first of its kind to quantify vocal symptoms in voice over actors during violent video game performances and provide evidence of the impact specialized training has to help offset any fatiguing and damaging effects. Our hypotheses were supported when we saw substantial improvements in the form of medium-to-large effect size changes, in both objective acoustic perturbation and subjective patient reports once participants were trained with VCT. Further research into this technique would not only benefit from a larger sample size, but also from an attempt to recruit a participant population more actively engaged in the video game industry. Our study highlights the need not only for training programs to help performers healthy produce these violent sounds, but also for investigation into preventative measures and optimal post-session recovery techniques.

## FUNDING

This work was supported by the University of Cincinnati College-Conservatory of Music Faculty Development Fund awarded to author D.S.

## Acknowledgments

The authors thank Kelly Carraro and Haley Doerr for their assistance with acoustic data processing. Thank you to John Paul Giliberto, M.D. for his contributions to the protocol design.

## REFERENCES

1. Hunter EJ, Cantor-Cutiva LC, van Leer E, et al. Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *J Speech Lang Hear Res*. 2020;63:509–532.
2. Pestana PM, Vaz-Freitas S, Manso MC. Prevalence of voice disorders in singers: systematic review and meta-analysis. *J Voice Off J Voice Found*. 2017;31:722–727. https://doi.org/10.1016/j.jvoice.2017.02.010.
3. Roy N, Merrill RM, Gray SD, et al. Voice disorders in the general population: prevalence, risk factors, and occupational impact. *The Laryngoscope*. 2005;115:1988–1995.
4. Cazden J. Screaming for attention: the vocal demands of actors in violent interactive games. *J Voice Off J Voice Found*. 2017;31:1–2. https://doi.org/10.1016/j.jvoice.2016.02.001.
5. Palandrani P. *Video Games & Esports: Building on 2020's Rapid Growth*. Global X; 2021.. Published March 9Available at; https://www.globalxetfs.com/video-games-esports-building-on-2020s-rapid-growth/. Accessed March 18, 2022.
6. Witkowski W. *Videogames Are a Bigger Industry Than Movies And North American Sports Combined, Thanks to the Pandemic*. MarketWatch; 2021. Published January 2Available at; https://www.marketwatch.com/story/videogames-are-a-bigger-industry-than-sports-and-movies-combined-thanks-to-the-pandemic-11608654990 . Accessed March 19, 2022.
7. Guss J, Sadoughi B, Benson B, et al. Dysphonia in performers: toward a clinical definition of laryngology of the performing voice. *J Voice Off J Voice Found*. 2014;28:349–355. https://doi.org/10.1016/j.jvoice.2013.10.004.
8. Canadian Broadcasting Corporation. "One Time, I Actually Threw up": Why Video Game Voice Actors Are on Strike.; 2016. https://link.gale.com/apps/doc/A470619521/CPI?u=ucinc_main&sid=summon&xid=a1ba5b02.
9. Frankel T. *In $24* Billion Video Game Industry, Voice Actors Struggle to Make a Living. Chicago Tribune; 2017.. Published October 29Available at; https://www.chicagotribune.com/business/blue-sky/ct-wp-bsi-video-game-voice-actors-20171029-story.html. Accessed March 27, 2022.
10. Ogrodnik N, Li-Jessen NYK, Mongeau L. Quantification of occupational vocal fatigue in voice actors. In: 2018.
11. Fujiki RB, Chapleau A, Sundarrajan A, et al. The interaction of surface hydration and vocal loading on voice measures. *J Voice*. 2017;31:211–217.
12. Sundarrajan A, Fujiki RB, Loerch SE, et al. Vocal loading and environmental humidity effects in older adults. *J Voice*. 2017;31:707–713.
13. Roy N, Ryker KS, Bless DM. Vocal violence in actors: an investigation into its acoustic consequences and the effects of hygienic laryngeal release training. *J Voice*. 2000;14:215–230. https://doi.org/10.1016/S0892-1997(00)80029-6.
14. Lishner DA, Groves CL, Chrobak QM. Are violent video game-aggression researchers biased? *Aggress Violent Behav*. 2015;25:75–78. https://doi.org/10.1016/j.avb.2015.07.010.
15. Fall S T. Retract then attack. *Fight Master*. 2012;34:18–19.
16. Aaen M, McGlashan J, Christoph N, et al. Extreme vocal effects distortion, growl, grunt, rattle, and creaking as measured by electroglottography and acoustics in 32 healthy professional singers. *J Voice Off J Voice Found*. 2021. https://doi.org/10.1016/j.jvoice.2021.11.010. Published online December 28S0892-1997(21)00391-X.
17. Guzman M, Ortega A, Olavarria C, et al. Comparison of supraglottic activity and spectral slope between theater actors and vocally untrained subjects. *J Voice*. 2016;30:767.e1–767.e8. https://doi.org/10.1016/j.jvoice.2015.10.017.
18. Kempster GB, Gerratt BR, Verdolini AK, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132. https://doi.org/10.1044/1058-0360(2008/08-0017.
19. Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI): development and validation. *Am J Speech Lang Pathol*. 1997;6:66–70. https://doi.org/10.1044/1058-0360.0603.66.
20. Phyland DJ, Pallant JF, Thibeault SL, et al. Measuring vocal function in professional music theater singers: construct validation of the Evaluation of the Ability to Sing Easily (EASE). *Folia Phoniatr Logop Off Organ Int Assoc Logop Phoniatr IALP*. 2014;66:100–108. https://doi.org/10.1159/000366202.
21. Melzack R. The short-form McGill pain questionnaire. *Pain*. 1987;30:191–197. https://doi.org/10.1016/0304-3959(87)91074-8.
22. Boersma P. Praat, a system for doing phonetics by computer. *Glot Int*. 2001;5:341–345.
23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.
24. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Second Lawrence Erlbaum Associates; 1988.
25. Faul F, Erdfelder E, Buchner A, et al. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*. 2009;41:1149–1160. https://doi.org/10.3758/BRM.41.4.1149.
26. Fujiki RB, Thibeault SL. Examining relationships between GRBAS ratings and acoustic, aerodynamic and patient-reported voice measures in adults with voice disorders. *J Voice*. 2021 S0892199721000746. https://doi.org/10.1016/j.jvoice.2021.02.007. Published online March.
27. Ma EPM, Yiu EML. Multiparametric evaluation of dysphonic severity. *J Voice*. 2006;20:380–390. https://doi.org/10.1016/j.jvoice.2005.04.007.
28. Martin D, Fitch J, Wolfe V. Pathologic voice type and the acoustic prediction of severity. *J Speech Lang Hear Res*. 1995;38:765–771. https://doi.org/10.1044/jshr.3804.765.
29. Brockmann-Bauser M, Bohlender JE, Mehta DD. Acoustic pertubation measures improve with increasing vocal intensity in individuals with and without voice disorders. *J Voice*. 2018;32:162–168. https://doi.org/10.1016/j.jvoice.2017.04.008.
30. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Lang Hear Res*. 1994;37:769–778. https://doi.org/10.1044/jshr.3704.769.
31. Kelchner LN, Brehm SB, Weinrich B, et al. Perceptual evaluation of severe pediatric voice disorders: rater reliability using the consensus auditory perceptual evaluation of voice. *J Voice*. 2010;24:441–449. https://doi.org/10.1016/j.jvoice.2008.09.004.
32. Eadie TL, Nicolici C, Baylor C, et al. Effect of experience on judgments of adductor spasmodic dysphonia. *Ann Otol Rhinol Laryngol*. 2007;116:695–701. https://doi.org/10.1177/000348940711600912.

33. Eadie TL, Kapsner M, Rosenzweig J, et al. The role of experience on judgments of dysphonia. *J Voice*. 2010;24:564–573. https://doi.org/10.1016/j.jvoice.2008.12.005.
34. Johnson J. *A Comparison between Self-Rated and Listener-Rated Outcomes in Tracheoesophageal Speech*. University of Washington; 2013.
35. Lee M, Drinnan M, Carding P. The reliability and validity of patient self-rating of their own voice quality. *Clin Otolaryngol*. 2005;30:357–361. https://doi.org/10.1111/j.1365-2273.2005.01022.x.
36. Arffa RE, Krishna P, Gartner-Schmidt J, et al. Normative values for the voice handicap index-10. *J Voice*. 2012;26:462–465. https://doi.org/10.1016/j.jvoice.2011.04.006.
37. Long J, Williford HN, Olson MS, et al. Voice problems and risk factors among aerobics instructors. *J Voice*. 1998;12:197–207. https://doi.org/10.1016/S0892-1997(98)80039-8.
38. Rumbach A, Khan A, Brown M, et al. Voice problems in the fitness industry: factors associated with chronic hoarseness. *Int J Speech Lang Pathol*. 2015;17:441–450. https://doi.org/10.3109/17549507.2014.987820.
39. Rumbach AF. Vocal problems of group fitness instructors: prevalence of self-reported sensory and auditory-perceptual voice symptoms and the need for preventative education and training. *J Voice*. 2013;27:524.e11–524.e21. https://doi.org/10.1016/j.jvoice.2013.01.016.
40. Spellman J, Coulter M, Roth C, et al. Prevalence, characteristics and impact of dysphonia in US marine corps drill instructors. *J Voice*. 2020;34:694–701. https://doi.org/10.1016/j.jvoice.2019.02.015.
41. Borg G. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc*. 1982;14:377–381.
42. van Leer E, van Mersbergen M. Using the borg CR10 physical exertion scale to measure patient-perceived vocal effort pre and post treatment. *J Voice*. 2017;31:389.e19–389.e25. https://doi.org/10.1016/j.jvoice.2016.09.023.