The Impact of a Vocal Loading Task on Voice Characteristics of Female Speakers With Benign Vocal Fold Lesions

*Nicole Free, †Joseph C. Stemple, *Julian A. Smith, and *Debra J. Phyland, *Melbourne, Australia, and †Lexington, Kentucky

Summary: Objectives. To examine the effect of a vocal loading task on measures of vocal structure and function in females with benign vocal fold lesions (BVFLs) and determine if change is observed in voice and lesion characteristics.

Study Design. Prospective cohort study.

Methods. Twenty-eight (n = 28) female subjects with phonotraumatic BVFLs completed a vocal loading task of 30 minutes of reading aloud at 75-85 dBA. Multidimensional voice evaluation was completed pre- and post-load, including audio and videostroboscopy recordings and images for expert perceptual ratings and acoustic and aerodynamic evaluation. Subjects also scored themselves using a 10 cm visual analogue scale for Perceived Phonatory Effort, and completed the Evaluation of Ability to Voice Easily, a 12 item self-report scale of current perceived speaking voice function. An exploratory rather than confirmatory approach to data analysis was adopted. The direction and magnitude of the change scores (pre- to post-load) for each individual, across a wide variety of instrumental and self-report measures, were assessed against a Minimal Clinically Important Difference criteria.

Results. Observations of change and the direction of change in vocal response of individuals with BVFLs to 30 minutes of loud vocal load was variable. Minimal to no change was noted for participants pre- to post-load as rated perceptually, for auditory and videostroboscopy samples. For most instrumental measures, change was shown for many participants including an overall improvement in aerodynamic and acoustic measures of function and efficiency post-load for 20 participants (77%) and decline in function for 4 participants (15%). Self-reported effort and vocal function post-load was multidirectional with similar numbers of participants reporting no change, improved function or a decline.

Conclusion. Subjects with BVFLs demonstrate change in vocal function following 30 minutes of vocal load. While this change can be variable and multidirectional, overall improvement was observed in instrumental measures of function and efficiency for most participants. Some participants perceived this change to be an increase in effort, some a reduction in effort and some perceived no change. Improved vocal function despite relative lesion stability can seemingly occur after loading in some pathological voices.

Key Words: Benign vocal fold lesion−Phonotrauma−Vocal load−Vocal loading task−Voice disorders.

INTRODUCTION

Vocal load is a term used to quantify voice use and is derived from characteristics such as duration of phonation time, loudness, pitch, quality and efficiency of voice production,[1] and is also referred to as vocal demand.[2] Vocal loading tasks (VLTs) have been designed and implemented to study how the larynx responds to a controlled high demand task, with similar loading tasks used in sports medicine[3] and other medical fields.[4] These tasks vary in nature, provide insight regarding range of function and can have diagnostic value. VLTs are used to study how the larynx responds to stress, to challenge the laryngeal mechanism, and in many studies are considered a negative stressor aimed to compromise laryngeal function and to result in observable fatigue.[5]

Manipulation of extrinsic and intrinsic factors in an effort to load the larynx and observe evidence of change in vocal function has been explored in previous research.[5] Extrinsic factors include background noise, room acoustics and caffeine consumption. Examples of intrinsic manipulation include elevated intensity, altering vocal quality, or prolonged periods of reading or sustained vowels.[6-9] Many studies track potential vocal demand response using multidimensional assessment, commonly including videostroboscopic, aerodynamic, acoustic, and auditory-perceptual evaluation in addition to a patient reported phonatory effort scale or subjective self-rating on specific parameters.[6,10-12]

Benign vocal fold lesions (BVFLs) are laryngeal pathologies that result in impaired voice production due to incomplete glottic closure, disturbed vibratory characteristics and increased vocal fold mass.[13] They have commonly been attributed to acute or repeated episodes of phonotrauma in individuals with laryngeal structure potentially predisposed to cumulative damage.[14-16] While the genesis of these lesions is not yet fully established, phonotraumatic BVFLs generally include nodules, polyps, pseudocysts and reactive lesions.[16] Typical symptoms include impaired vocal quality, reduced pitch range, reduced vocal stamina, increased vocal effort and throat discomfort.[17] It is unknown whether

Accepted for publication November 9, 2021.

From the *Department of Surgery, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia; and the †Department of Communication Sciences and Disorders, and Rehabilitation Sciences PhD Program, University of Kentucky, Lexington, Kentucky.

Address correspondence and reprint requests to Nicole Free, Department of Surgery, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Clayton, Victoria, Australia. E-mail: nicole.free@monash.edu

Journal of Voice, Vol. 38, No. 4, pp. 964.e1−964.e16
0892-1997

https://doi.org/10.1016/j.jvoice.2021.11.009

patients with BVFLs present with stability in vocal function, or if variability in the nature and degree of these associated vocal symptoms across time and voicing contexts is experienced. Such fluctuations would support a hypothesis that the lesions themselves may vary in size and/or impact on voice production. Vocal load has been suggested to be a major influence on such potential fluctuations whereby an increased vocal load can lead to increased phonotrauma and lesion expression.[14] It therefore follows that a reduction in vocal load may reduce phonotrauma and minimise the impact of a BVFL on vocal impairment symptoms. Indeed, some voice therapy protocols call for varying degrees of vocal conservation from complete voice rest to conservative voice use.[18,19]

Two previous studies have investigated the impact of a VLT on voice characteristics in small samples of participants with one type of BVFL, vocal fold nodules. These studies included aerodynamic, acoustic and self-report measures however, to date, laryngeal examination has not been implemented to assess the impact of VLTs on lesion characteristics. While a small number differences were observed in higher subglottic pressures and ratings of subjective effort, in general, the majority of observed changes in voice characteristics, such as increased sound pressure level and fundamental frequency, and deterioration in vocal quality were comparable between participants with nodules and controls.[11,20] Remacle, Morsomme, Berrue and Finck[11] implemented a 2-hour VLT at 70-75 dB with 16 participants, and their findings of a similar effect on both speakers with and without BVFLs highlight that additional intrinsic or extrinsic factors should be considered in the development of loading tasks with this population. In Niebudek-Bogusz, Kotylo and Sliwinska-Kowalska's[20] study of dysphonic female speakers, including 16 participants with nodules, participants spoke over 80 dB of background noise for a 30 minute VLT. Their findings are difficult to translate to the BVFL population given the heterogeneity of their cohort, however a range of responses were observed in their group - with deterioration, no change and improvement observations post load. The impact of VLTs for other groups of dysphonic speakers have also been investigated in a small number of studies, such as patients with paralysis and functional dysphonia.[21,22] The nature of vocal demand response in participants with BVFLs and whether there are differences in response between those with or without BVFLs has not yet been ascertained. This knowledge would seem critical to voice care management especially for guidelines around thresholds of voice use for optimal vocal function – how much, how loud, how high. This understanding has the potential to reduce the risk of further phonotraumatic injury leading to lesion progression and worsening of vocal impairment symptoms.

Objective

To examine the effect of a VLT on measures of vocal structure and function in participants with BVFLs.

Research questions

1. Is change observed in measures of vocal function and lesion characteristics in female speaks with BVFLs after 30-minutes of a VLT at 75-85dBA?
2. Are observations consistent across the cohort, or is there variability within the group?
3. Is the direction of change consistent across measures of vocal function and lesion characteristics?

METHOD

Approval for the conduct of this study and subject recruitment was obtained from the Monash University Human Ethics Committee (project number 9681).

Subjects

In this prospective cohort study, 28 females with a diagnosis of a BVFL were recruited via a private multi-disciplinary voice clinic in Melbourne by convenience sample between October 2017 and May 2019. Patients with a diagnosis of BVFL were provided a written invitation to participate in the study by their treating ENT Surgeon or Speech Pathologist and participants then contacted the researcher to express interest and consent to participation. Demographic information for participants is presented in Table 1. Participants who identified themselves as singers were varied in their main singing genre, and amateur and professional status. Singers were considered trained if they had undertaken tertiary education or had received more than twenty 1:1 singing lessons. Baseline self-report scores for the voice related quality of life tool, the Voice Symptom Scale (VoiSS)[23] are also presented. The median score of 46 (38.5, 57.5), indicating moderate impact, is consistent with previous research of participants with BVFLs.[24] The range of scores from 13 to 79 indicates the self-perceived severity and impact of dysphonia was broad within the cohort. VoiSS score was evaluated to determine if a difference in self-perception of voice impairment was evident for trained singers and the remaining cohort. A Mann-Whitney U test revealed no significant difference in total scores for trained singers and the rest of the sample group ($U = 127$, $z = 1.44$, $P = 0.159$, $r = 0.27$).

Inclusion criteria

Diagnosis of BVFL via laryngeal stroboscopy by an Otolaryngologist/Laryngologist was required for inclusion. BVFLs were classified according to their clinical diagnosis and morphology, and included pre nodular oedema, nodules, polyps and pseudocysts. Participants were all non-smokers, perceived themselves to be well hydrated and had no recognised risk factors such as excessive alcohol or caffeine consumption.

Exclusion criteria

Patients with a diagnosis of vocal fold cyst on stroboscopy were excluded from this study as the distinct histopathologic classification of epithelial lining surrounding the lesion distinguishes them from other types of phonotraumatic lesions

TABLE 1.
Participant Characteristics

Age (years)	Median (IQR)	29 (25-35.5)		
	Range	21-53		
Age group	18-29 yrs	16	57%	
	30-39 yrs	8	29%	
	40-49 yrs	2	7%	
	50+	2	7%	
Primary Lesion/s	Nodules	13	46%	
	Polyp/s	8	29%	
	Pseudocyst	6	21%	
	Pre nodular oedema	1	4%	
Onset	Sudden	10	36%	
	Progressive	18	64%	
Time since onset	1-3 mo	4	14%	
	3-6 mo	2	7%	
	>6 mo	22	79%	
Trained singer	Yes	12	43%	
	No	16	57%	
Voice Symptom Scale (VoiSS)		Median (IQR)	Min-Max	Possible range
	Total score	46 (38.5-57.5)	13-79	0-120
	Impairment	29 (22.5-35.3)	11-40	0-60
	Emotional	7 (3.5-12.8)	0-23	0-32
	Physical	10 (7-13.5)	0-20	0-28
		Median	*Min-Max*	
VoiSS total score by trained singer status	*Yes*	*42*	*31-79*	
	No	*52*	*13-69*	

and likely results in greater lesion stability. Three patients presenting with evidence of vocal fold haemorrhage, varix or marked hypervascularity were also excluded from this study, as the researchers hypothesised the VLT had the potential to increase the risk of further vascular trauma. Patients with any concomitant diagnosis, such as paresis or uncontrolled reflux, were excluded from this study.

Loading task protocol

The participant read aloud for a period of 30 minutes maintaining an intensity of 75-85 dBA. This intensity level was selected based on previous VLT literature, and with reference to an intensity level above the normal range of mean speaking SPL for Australian female speakers (mean 69.75 +/-2.78 dB) to present a vocal challenge to participants.[5,25] No training was provided to the participant. The Vocal Loading Test module of lingWAVES voice analysis program was implemented (WEVOSYS Medical Technology, Baunach, Germany). The lingWAVES system consists of software and standardized recording hardware, USB connector and a pre calibrated and certified sound level meter. The hardware provides sound level data in real time to the software program and allows the participant and researcher to track dBA visually. The module was set to 30 minutes with minimum and maximum thresholds of 75 dBA to 85 dBA. A tone sounded each time the participant dropped below the minimum level, or exceeded the maximum threshold. The participant completed the task standing, with a distance of 30 cm from mouth to microphone. While participants were instructed to cease the task if they experienced discomfort, all participants completed the 30 minutes VLT. Participants were permitted to sip water during the VLT as desired, though the majority of participants chose not to do so.

When elevated intensity is employed in a VLT, there is evidence a duration greater than 1 hour is required for observable changes in function with normophonic speakers.[5] Given this was a novel investigation of a treatment seeking participant group with a range of BVFLs and the first study to use stroboscopy to examine the effect of a VLT on lesion characteristics and laryngeal function, a more conservative duration of 30 minutes was selected.

Evaluation protocol

Participants completed a multidimensional voice evaluation before and immediately after their vocal loading tasks.

Perceptual evaluation

Voice recording was performed using the Voice Protocol module of lingWAVES voice analysis program. Voice samples were recorded using a WEVOSYS SPL meter microphone positioned 30 cm from the participant's mouth in a quiet room with ambient noise of 30 dBA. Participants read the first paragraph of "The Rainbow Passage" (Fairbank, 1960). Due to software error, one post load voice recording was not able to be included. A total of 27 pairs of pre- and post-loading reading samples were collated in random pre-post/post-pre order, with 10% (3 pairs) repeated for inter- and intra-rater reliability assessment. The reading samples were rated by three Speech Pathologists with a minimum of 16 years clinical experience in voice. The raters completed a calibration session with five voice samples not from this study and discussed the specific parameters for evaluation. Samples were then individually rated using the 6 point Perceptual Voice Profile (PVP) scale[26] ranging from normal to severe for three selected parameters: breathy, rough and strain. A global rating was also given for each pair of samples; raters scored the second sample of each pair as same, improved or worse compared to the first sample.

Videostroboscopy

Laryngeal videostrobscopy was performed and recorded using Xion EndostrobE with 70 degree rigid scope and Digital Video Archive Software (DiVAS) version 2.5 (XION GmbH, Berlin, Germany). Participants made their best effort to complete /i/ sustained for 5-7 seconds at comfortable pitch and loudness, high pitch, low pitch, loud and soft. Pre- and post-recordings were made for a total of 27 participants as one participant was excluded from this task as she was unable to tolerate rigid stroboscopy.

Recordings were reviewed by the primary investigator and collated for evaluation. A still image of maximal abduction and a minimum of 5 seconds of vibration at comfortable pitch and loudness were selected, with additional best segments of high, low, loud and soft voicing at similar frequency (Hz) and intensity (dB) levels pre- and post-load. Stroboscopy videos for one participant were excluded as the minimum 5 seconds of continuous voicing at comfortable pitch and loudness was not achieved. Still images of maximal vocal fold abduction for evaluation of lesion characteristics were not obtained for another participant. Therefore a total of 26 pairs of pre- and post-loading recordings and still images were collated in random pre-post/post-pre order and 10% were repeated for inter- and intra-rater reliability assessment. Recordings and images were evaluated by four blinded, expert raters: two Laryngologists and two Speech Pathologists. Raters scored the second sample of each set as no change, worse or improved compared to the first sample on three parameters; vocal fold closure, vibratory characteristics and lesion characteristics. Raters were provided with two pairs of videos and still images for calibration and rated the examples along with a guide rating form.

Aerodynamic evaluation

The KayPENTAX Phonatory Aerodynamic System (PAS) Model 6600 (Kay PENTAX Corp, Lincoln Park, NJ) was used for aerodynamic recordings. The Voicing Efficiency module was completed with three sets of 5-7 serial repetitions of /pa/ at comfortable pitch and loudness for evaluation of glottal airflow during voicing (Flow) (L/s) and mean peak air pressure/estimated mean subglottic pressure (Psub) (cm H20), with an average value calculated. Phonation threshold pressure (PTP) (cm H20) was evaluated in the same module but with 5-7 repetitions of /pa/ at softest voicing without whispering. An average value from the 3-5 smallest consecutive and regular peaks was calculated for PTP ensuring voicing was achieved for each repetition. The Maximum Sustained Phonation module was completed for evaluation of phonation time (MPT) (s) and expiratory volume (EV) (L), and a consistent instruction was provided to participants to sustain /a/ vowel for as long as possible with encouragement during the task to continue to maximum duration. Due to software error, one post-load recording for Flow, Psub and MPT, and two recordings of EV were unable to be included.

Acoustic evaluation

Voice recordings for acoustic analysis were performed using the Voice Protocol module of lingWAVES voice analysis program pre- and post-loading task. Participants provided three repetitions of 5 second sustained /a/ vowel for glottal to noise excitation ratio (GNE) calculation. Glissandos from modal to highest pitch, and modal to lowest pitch on vowels /a/, /i/, and /u/ were provided to evaluate minimum f0 (min f0), maximum f0 (max f0) and semitone range (SR). Finally, mean fundamental frequency (mean f0) and mean intensity (SPL) were obtained from the reading of the first paragraph of the Rainbow Passage (Fairbank, 1960). A total of 27 pairs of pre- and post-loading voice recordings were collated as one post-load voice recording was not able to be included due to software error.

Subjective self-rating

Participants scored themselves using a 10 cm visual analogue scale (VAS) for Perceived Phonatory Effort (PPE) – with "very easy" and "very effortful" marking each end of the scale. In addition all participants completed the Evaluation of Ability to Voice Easily (EAVE), a newly developed and validated 12 item self-report scale of current perceived speaking voice function.[27]

Statistics

As the aim of this study was to explore the impact of vocal loading on a variety of different parameters, no directional hypotheses were specified. In this context it was decided that the use of inferential Null Hypothesis Statistical Testing (NHST) techniques would not be appropriate, and an exploratory, rather than confirmatory, approach was adopted. A review by Balluerka, Gómez and Hidalgo[28] recommended that graphical techniques be utilised by all researchers as they provide "essential information to get a better understanding of the set of data in hand" (p. 66). Consistent with this advice, a series of spaghetti plots (plotting each person's pre- and post-loading scores as separate lines) were generated to explore the direction and magnitude of change. On a substantial number of parameters assessed, these plots showed some subjects recording an increase in scores from pre- to post-load and others a decrease. This differential response to load identified by the plots suggested that group-based NHST statistical techniques were not appropriate. However, inspection of change scores at the individual, rather than group level, allowed detection of both the direction of the change and the magnitude.

In addition to the output data for instrumental measures, minimal clinically important difference (MCID) values were set for each parameter to assess the change scores. The selection of MCIDs was primarily determined by a distribution-based approach using published normative data, applying a relatively conservative one standard deviation of the normal mean to represent a MCID, with further consideration of the clinical application of the parameter by the researchers to establish a value.[25,29-31] Where normative data was not available, previous use of the parameter in the VLT literature informed this selection, such as an MCID for PTP which was based on previous reports of a statistically significant increase for female speakers at comfortable voicing at approximately 0.75cm/H20.[32,33] To date there have been no published MCID values for aerodynamic or acoustic voice analysis parameters. MCIDs allow investigation of minimal values at which a patient may note change or benefit in function allowing meaningful clinical interpretation and were considered more appropriate with a small sample size.[34,35] Change scores were calculated by subtracting scores for each parameter collected post-loading from those obtained prior to the loading task. Data was evaluated using SPSS Version 27 and group descriptive statistics are stated as median and interquartile range as data was non parametric.

RESULTS

Reliability

Perceptual ratings

Intra-rater reliability. An intraclass correlation coefficient (ICC) model 3 was used to evaluate intra-rater reliability for the perceptual ratings using the PVP parameters breathy, strain and rough with raters representing a fixed effect and samples a random effect.[36] Raters presented with good to excellent intra-rater reliability (rater 1 ICC 0.95, $P < 0.001$; rater 2 ICC 0.81, $P < 0.001$; rater 3 ICC 0.89, $P < 0.001$). Percentage agreement was calculated to ensure statistical analysis of ICC was reflective of clinical implementation of the PVP and agreement was determined to be a rating within +/- 1 on the 6-point scale. Percentage agreement was reflective of the ICC values obtained (rater 1 100%, rater 2 94%, rater 3 94%). For global perceptual rating of "no change," "better" or "worse" post loading, a percentage agreement was calculated for the three repeated pairs. Rater 2 presented with 100% agreement with ratings, while rater 1 and 3 had one consistent rating among the three repeated pairs and therefore excluded from further analysis of this data.

Inter-rater reliability. An ICC model 3 was used to evaluate the inter-rater reliability of the PVP parameters.[36] Strain and rough were found to have good reliability between raters (strain ICC 0.80 $P < 0.001$; rough ICC 0.86 $P < 0.001$), while breathy was found to have poor agreement (breathy ICC 0.530 $P < 0.001$). Percentage agreement was calculated and was reflective of the ICC values obtained (breathy 54%, strain 72%, rough 81%).

Stroboscopy ratings

Intra-rater reliability. A percentage agreement was calculated to evaluate intra-rater reliability for stroboscopy ratings. Rater 4 reached 100% agreement on repeated samples of video ratings of closure pattern and vibratory characteristics, while raters 1, 2 and 3 each reached 50% agreement. Rater 4 was therefore the only rater included in further analysis. For ratings of lesion characteristics from still images both rater 3 and 4 had 100% agreement on repeated samples. Rater 1 and 2 were excluded from further analysis of this data, with 66% agreement on repeated samples.

Inter-rater reliability. An ICC model 3 was used to evaluate inter-rater reliability of lesion characteristic ratings from rater 3 and 4. The ICC value (ICC 0.51, $P = 0.026$) and subsequent percentage agreement (58%) were both poor therefore this data was not included in further analysis.

Results

Perceptual

A MCID of +/- 1 scale degree was applied to this data to evaluate whether change was observed for individuals post load. Six (22%) of the strain ratings exceeded MCID post load, with 3 (11%) rated with less strain, and 3 (11%) with more strain. Two (7%) post load rough ratings exceeded MCID with one sample rated as less rough, and one rated as more rough. Seventy-seven percent of voice samples were rated with no change in strain post load and 93% with no change in roughness post load. Finally, using the global perceptual rating, Rater 2 rated 12 (44%) of post load samples as the "same" as post load, 12 (44%) were rated as better post load and 3 (11%) as worse.

Stroboscopy

Videos. The majority of post load videos were rated with no change on vibratory characteristics or glottal closure. Seven (27%) participants were rated with a positive change in vibration post load and none were rated as worse. Six (23%) participants were rated with improved glottic closure post load and 1 (4%) with worse closure. Seventy-three percent of stroboscopy videos were rated with no change post load for vibration and glottic closure.

Still images. Due to poor inter-rater reliability on this rating task, no analysis of this data is presented. It should be noted that there was no evidence of acute bleeding or overt

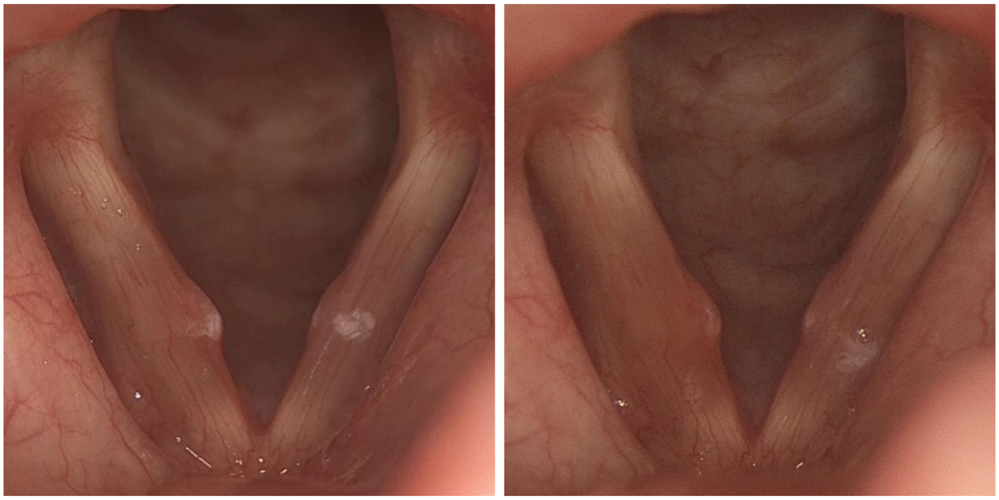

IMAGE 1. Participant 15 pre- and post-loading with increased redness surrounding the right vocal fold lesion.

vascular trauma post loading. At the time of stroboscopy recording, the principal researcher noted participant 15 presented with a mild increase in redness or inflammation in the vocal fold tissue surrounding her right vocal fold lesion post loading (Image 1). This was not noted in any other participant. All other lesion presentations were deemed by the researcher to be similar pre- and post-loading and two further sets of stroboscopy images are provided as typical examples of pre- and post-loading images (Image 2 and 3).

Aerodynamic measures

Descriptive data for aerodynamic measures are presented in Table 2. The difference between pre- and post-load aerodynamic measures was calculated and each individual score checked to see if a MCID was exceeded to represent change versus no change. Less than 50% (n = 12) of participants had post load difference in scores that exceeded the MCID for the parameter Psub, with near even numbers for participants with a decrease (n = 7) and increase (n = 5) in post load score demonstrating that change post load was multidirectional. EV demonstrated little change post load, with 23% (n = 6) of post load scores exceeding MCID. Three participants presented with a decrease in score and three with an increase. For measures Flow, PTP and MPT greater than 50% of participants exceeded the MCID indicating the majority of participants presented with change post load. A trend in direction of change was observed for these three parameters with a majority of participants demonstrating decrease in Flow (n = 11) and PTP (n = 13) and an increase in MPT (n = 13). Exploration of the difference in individual and group scores using spaghetti plots for the parameters Psub and PTP are provided in Figures 1 and 2, further demonstrating the multi directional nature of post load change. The magnitude of change was explored by plotting the difference in scores from pre- to post-load, with the greatest magnitude observed for parameters PTP and MPT, presented in Figures 3 and 4. The range of post load difference in score is presented in Table 2; <0 indicates a lower score and >0 a higher score post load.

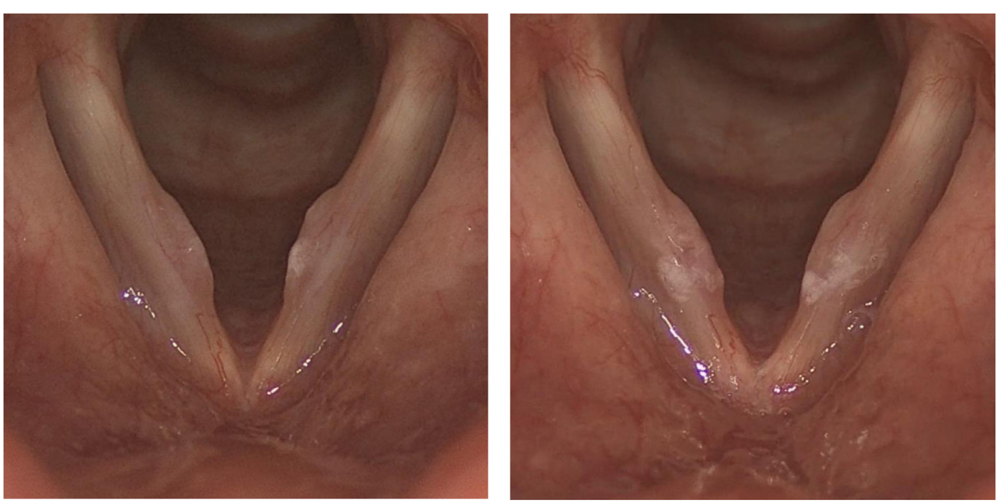

IMAGE 2. Participant 16 pre- and post-loading.

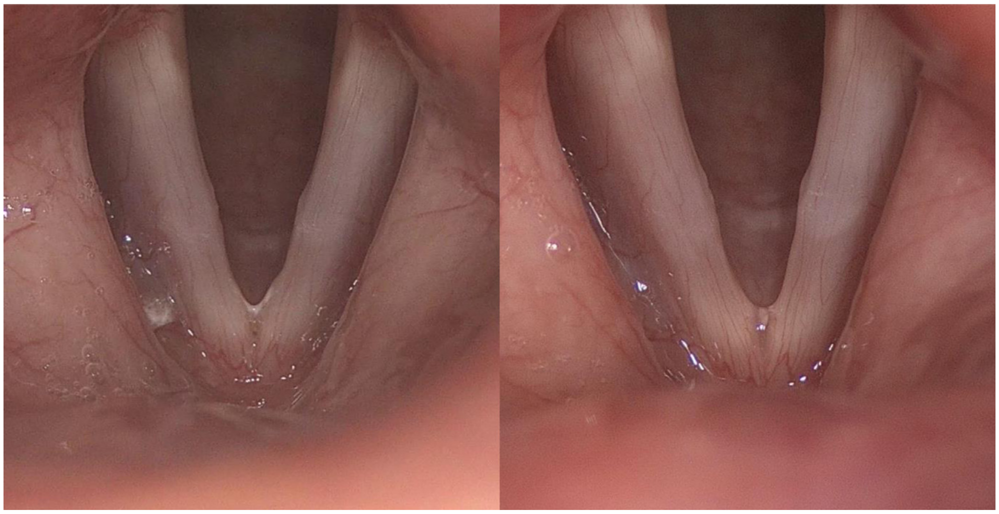

IMAGE 3. Participant 13 pre- and post-loading.

Acoustic measures

Descriptive data for acoustic measures are presented in Table 3. Individual scores across acoustic measures were evaluated and the difference between pre- to post-load scores checked to see if a MCID was exceeded. Min f0 demonstrated minimal change post load, with 19% (n = 5) of participants' post load difference in scores exceeding the MCID. A decrease in min f0 was observed for two participants and an increase for three. While only 37% (n = 10) of participants had a difference in post load mean f0 that exceeded the MCID, all 10 demonstrated an increase on this parameter and the pattern of individual pre- to post-load change on this parameters are presented in Figure 5. The majority of participants presented with change post load on measures max f0 and SR range, with 78% (n = 21) and 56% (n = 15) of participants exceeding the MCID on these parameters respectively. While this change was somewhat multidirectional, a trend was observed with the majority of participants demonstrating an increase in max f0 (n = 14, 52%) and SR post load (n = 11, 41%). A trend in the direction of change for GNE and SPL was also observed. Thirty-seven percent (n = 10) of participants presented with a decrease in GNE post load, and 7% (n = 2) an increase. Finally, the 48% of participants that exceeded the MCID for SPL all presented with an increase post load (n = 13). Spaghetti plots depicting individual scores for the parameters mean f0 and GNE are provided in Figures 5 and 6, and the difference in pre- to post-load SR score is presented in Figure 7. The range of post load difference in score for acoustic parameters is presented in Table 2, and SR was observed to have the greatest magnitude of change (Figure 7).

Participant self-reported measures

Participants completed PPE 10 cm VAS and the EAVE pre- and post-load to evaluate self-perception of phonatory effort and vocal function. The difference in individual pre- to post-load scores was calculated to determine if the MCID value was exceeded. Descriptive, MCID and range of difference data are presented in Table 4. Sixty-four percent (n = 18) of participants had post load scores that

TABLE 2.
Median, Interquartile Range, MCID and Difference Data for Pre- and Post-Loading Aerodynamic Measures

Variable			Median	IQR	MCID	n	No Change	Change	*Decrease Increase*	*Range of Difference*
Flow	ml/s	Pre	0.25	0.19, 0.34	+/- 0.05	27	12 (44%)	15 (56%)	*11 (41%)*	*-0.05-0.19*
		Post	0.23	0.19, 0.28					*4 (15%)*	*0.05-0.1*
Psub	cm H20	Pre	10.02	9.09, 13.08	+/- 1	27	15 (56%)	12 (44%)	*7 (26%)*	*-1.14 to -5.52*
		Post	10.25	7.75, 11.88					*5 (19%)*	*1.23-2.86*
PTP	cm H20	Pre	4.02	2.79, 5.37	+/- 0.75	28	12 (43%)	16 (57%)	*13 (46%)*	*-0.81 to -3.34*
		Post	3.82	1.97, 5.26					*3 (11%)*	*1.05-1.45*
MPT	sec	Pre	12.31	10.27, 16.43	+/- 2	27	11 (41%)	16 (59%)	*3 (11%)*	*-2.45 to -3.96*
		Post	14.55	11.61, 20.89					*13 (48%)*	*2.38-13.34*
EV	L	Pre	2.76	2.25, 3.40	+/- 0.5	26	20 (77%)	6 (23%)	*3 (12%)*	*-0.61 to -1*
		Post	2.87	2.46, 3.37					*3 (12%)*	*0.62-2.23*

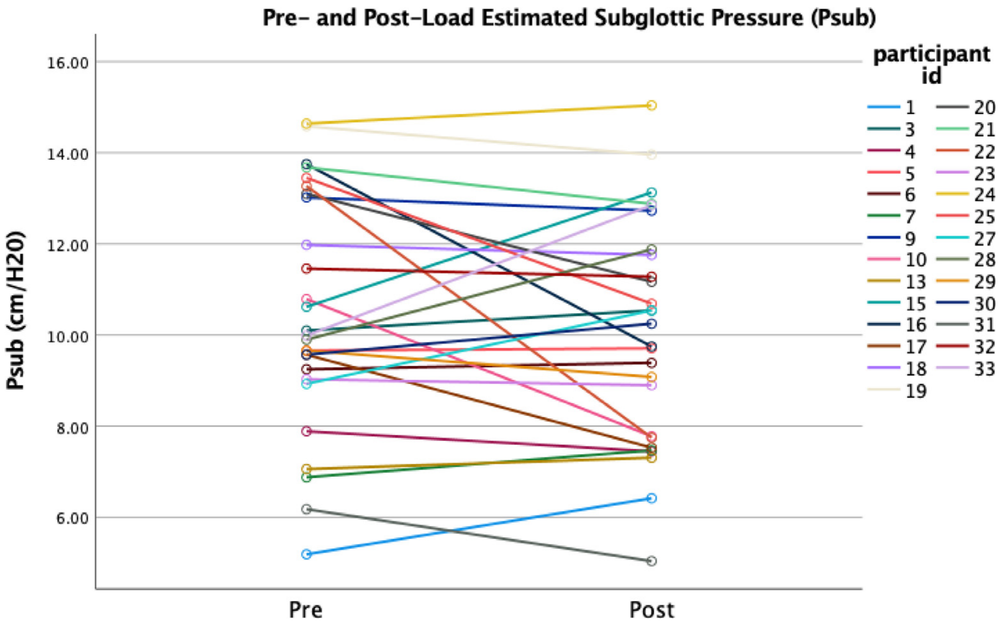

FIGURE 1. Individual participant pre- and post-load subglottic pressure (Psub).

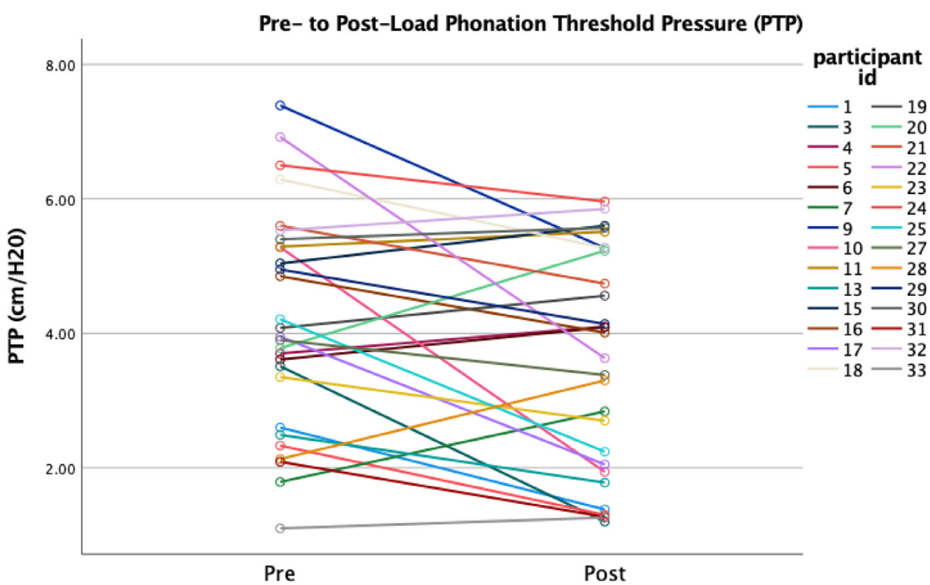

FIGURE 2. Individual participant pre- and post-load phonation threshold pressure (PTP).

exceeded the MCID for both PPE and EAVE. This difference in score was multidirectional, with similar numbers of participants showing an increase, decrease, or no change in score post load.

Patterns of MCID change for participants across instrumental parameters

As previously described, the calculation of MCID to explore meaningful change versus no change was undertaken for individuals across all 13 instrumental and self-report parameters. A range of outcomes were observed.

The direction of potential change (according to MCID) from pre- to post-load for individual participants was then compared for six key variables to assess consistency. Variables were excluded where the direction of change is not a clear improvement or decline in function; including EV, f0 measures and SPL. This data is presented in Table 5. The direction of change for individual participants across key aerodynamic variables of Flow, Psub, PTP and MPT, was generally consistent with 70% of participants (n = 19) demonstrating overall improved function, 11% (n = 3) no to minimal change and 19% (n = 5) negative change. Similarly, on the key acoustic variables SR and GNE, 50% (n = 13) of participants demonstrated overall improved function, 44% (n = 12) no to minimal change, and 7% (n = 2) negative change. The relationship between these two sets of instrumental measures for individuals was reasonable with 77%

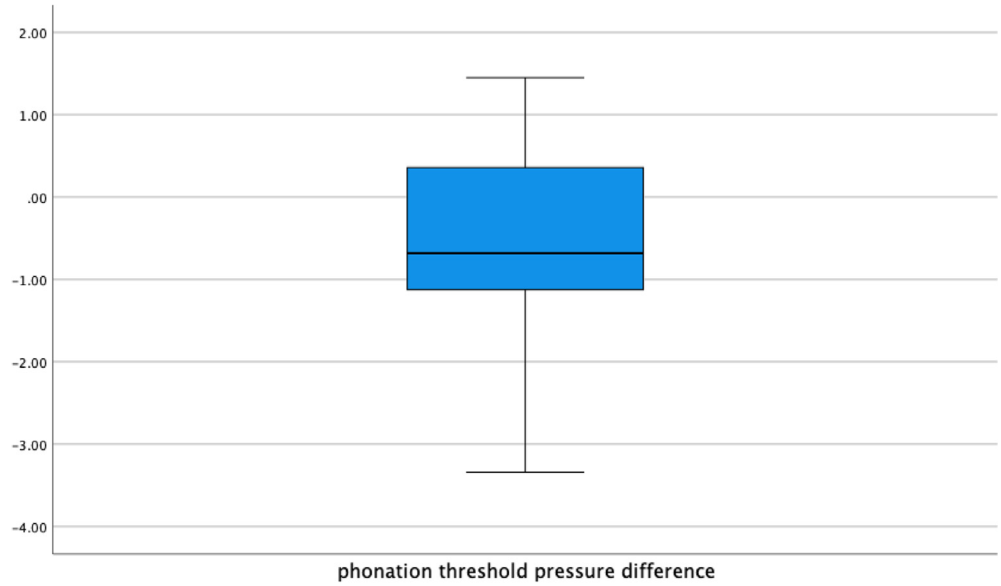

FIGURE 3. Range of difference in pre- to post-load participant scores for phonation threshold pressure (PTP), 0 indicates a lower score and >0 a higher score post load.

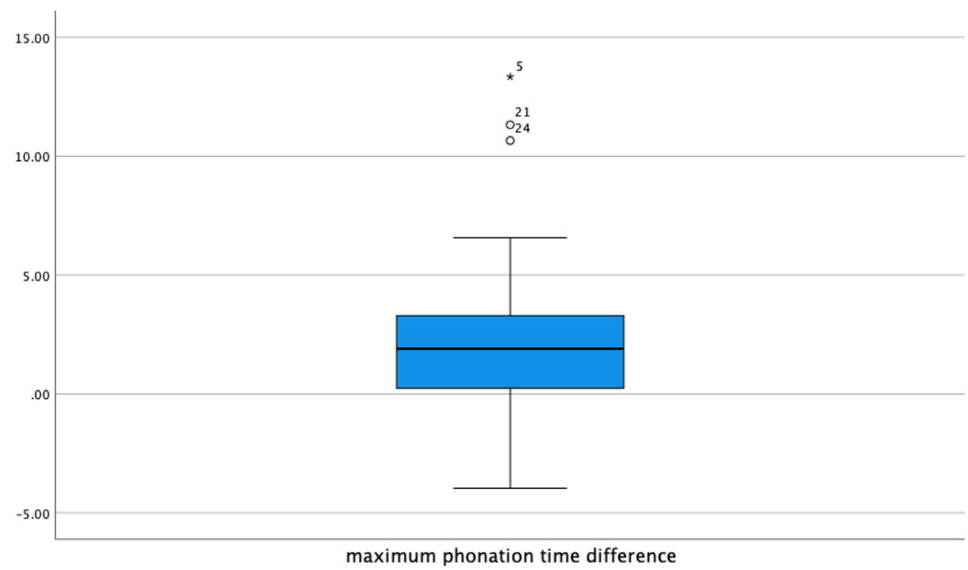

FIGURE 4. Range of difference in pre- to post-load participant scores for maximum phonation time (MPT), 0 indicates a lower score and >0 a higher score post load.

(n = 20) participants showing overall improved function, 8% (n = 2) no to minimal change, and 15% (n = 4) a negative change post load. MCID findings for self-report measures PPE and EAVE are included for review in Table 5, with less agreement noted between the self-report and instrumental measures regarding the direction of change for participants.

DISCUSSION

The results of this study demonstrate that observed change and direction of change in the vocal response of individuals with BVFLs to 30 minutes of loud vocal load can be variable. Minimal to no change was noted for participants pre- to post-load as rated perceptually for auditory and videostroboscopy samples. This would suggest that the impact of this loading task on vocal quality, vocal fold vibration, glottic closure and on the lesions was minimal or that potential change on these parameters is subtle and challenging to interpret consistently by expert raters. In contrast, change in most instrumental measures of vocal function was found and, although some worsened, for the majority of participants these changes were in a positive direction. The overall improvement observed for participants in many aerodynamic and acoustic measures of function and efficiency is novel and merits further consideration.

Previous studies regarding the effects of load on vocal function have differed to our study in their findings on aerodynamic and acoustic outcomes. Specifically, for

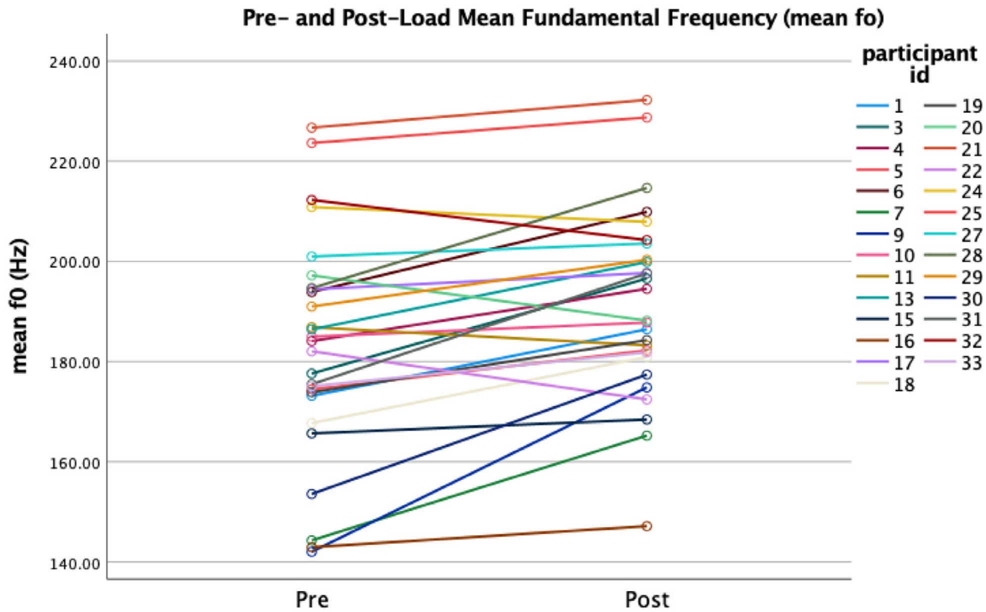

FIGURE 5. Individual participant pre- and post-load mean fundamental frequency (mean f0).

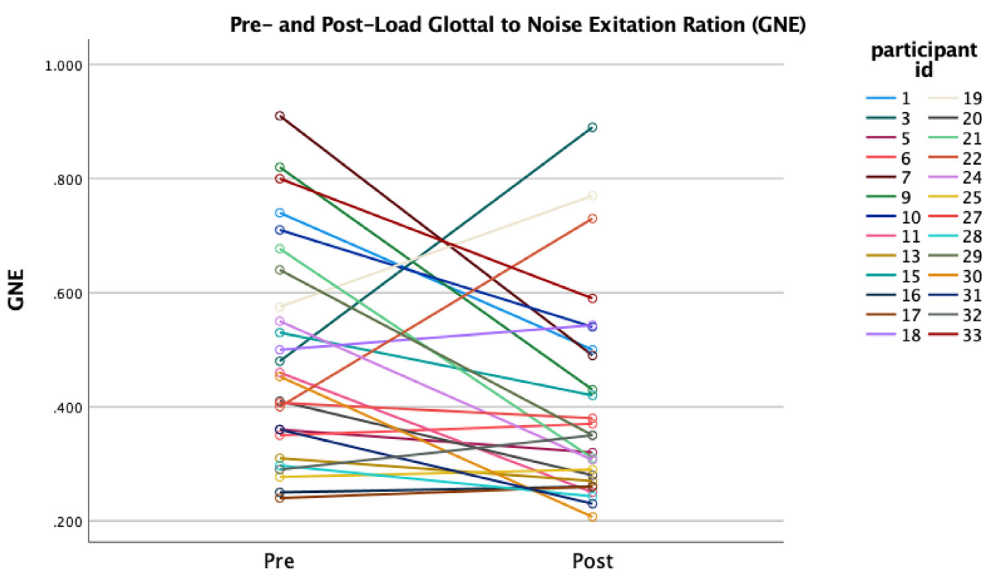

FIGURE 6. Individual participant pre- and post-load glottic to noise excitation ratio (GNE).

example, rather than the improvement indicated by a decrease found for the majority of our participants, an increase in PTP in normophonic subjects has been commonly observed once loading tasks are sustained for 30 minutes or more, or in shorter tasks with additional intrinsic loading factors.[32,37,38] Also, Aronsson, Bohman, Ternstrom and Sodersten[7] found significantly higher subglottic pressure (Psub) in their study of 10 females with vocal nodules compared to a control group after 90 seconds of reading in four levels of increasing background noise whereas increased SPL, f0 and subjective rating of strain were found for both groups. Remacle, Morsomme, Berrue and Finck[11] compared the effect of a 2-hour reading task at 70-75 dB on 16 female teachers with vocal nodules to normophonic teachers. Their results demonstrated similar findings for both groups with an increase in self perceived effort, fatigue and discomfort, progressive increase in f0, frequency range and SPL, and increase in Psub. Teachers with nodules had higher Psub and lower frequency range but otherwise responded comparably to the normophonic group to the VLT. Of interest, aerodynamic measures appeared to improve and stabilise only after 1 hour of loading for both groups which was postulated to be the result of adaptation to task or improved efficiency. These previous findings are also different to the current study whereby post load change in Psub was variable with small numbers of participants presenting with increased (N = 5) and decreased (N = 7) Psub but the majority of participants (N = 15) not changing.

Although the results of previous studies are difficult to compare to the current study, due to differences in the design of the VLT in terms of the task requirements and

TABLE 3.
Median, Interquartile Range and MCID Data for Pre- and Post-Loading Acoustic Measures

Variable			Median	IQR	MCID	n	No Change	Change	*DecreaseIncrease*	*Range of Difference*
Mean f0	Hz	Pre	184.12	173.19, 194.69	+/- 1 semitone	27	17 (63%)	10 (37%)	*0*	
		Post	188.01	180.91, 203.58					*10 (37%)*	*1 semitone*
Min f0	Hz	Pre	143.77	127.83, 147.50	+/- 1 semitone	27	22 (81%)	5 (19%)	*2 (7%)*	*NA*
		Post	143.41	131.85, 150.50					*3 (11%)*	
Max f0	Hz	Pre	630.51	474.35, 848.08	+/- 1 semitone	27	6 (22%)	21 (78%)	*7 (26%)*	*NA*
		Post	787.50	540.48, 905.50					*14 (52%)*	
SR		Pre	27	20, 33	+/- 2 semitone	27	12 (44%)	15 (56%)	*4 (15%)*	*-2 to -8*
		Post	29	23, 33					*11 (41%)*	*2-18*
GNE		Pre	0.46	0.34, 0.65	+/- 0.2	27	15 (56%)	12 (44%)	*10 (37%)*	*-0.21 to -0.78*
		Post	0.35	0.27, 0.52					*2 (7%)*	*0.33-0.41*
SPL	dBA	Pre	67.91	64.93, 70.47	+/- 3	27	14 (52%)	13 (48%)	*0*	
		Post	69.60	68.83, 72.15					*13 (48%)*	*3.4-6.05*

duration, they do seem to highlight that vocal function as measured by aerodynamic and acoustic parameters can change after loading. Given that the direction of this change can be positive, these results may provide further evidence for an overall adaptation to task, training effects, or improved efficiency post load in some participants with BVFLs. Another possible post load outcome may be a subtle increase in vocal fold oedema, leading to improved vocal fold closure, and in turn, improvement in aerodynamic measures with reduced glottal flow. This subtle or subclinical increase in oedema may not be perceptible on stroboscopy even by expert raters. Interestingly, unlike Sundarrajan, Huber and Sivasankar[39] who reported an increase in lung volumes post loading in younger adult participants (18-23 years), an increase in MPT without an increase in EV was observed in this study. This finding further supports improved glottic closure post load and is in keeping with the report by Remacle, Morsomme, Berrue and Finck.[11]

An increase in mean f0 and SPL has been consistently observed in response to VLTs in normophonic speakers and in participants with BVFLs.[7,10,40] Remacle, Morsomme, Berrue and Finck[11] hypothesised that an increase in mean f0 is likely due to increased muscle tension in the larynx or peri-larynx and results from stiffness of muscle or vocal fold cover rather than tissue changes or increased oedema where one might expect a decrease in mean f0. Remacle, Finck, Roche and Morsomme[41] observed an increase in mean f0, max f0 and f0 range as load increased, with these parameters measuring significantly higher in a 70- to 75-dB session compared with the 60- to 65-dB session. The findings of this study of a higher mean f0, max f0, SR and SPL for many participants are consistent with the literature. In addition, the principal researcher noted an improved glissando task performance post load with reduced phonation breaks.

In summary, the aerodynamic and acoustic data suggest that consistent improvement was observed for most participants. Perhaps surprisingly, despite the presence of BVFLs, loading seemed to have a positive effect on vocal function for many participants according to these instrumental measures and, for these speakers, may indicate a laryngeal or phonatory adaptation to this loading task or improved efficiency of voice production post loading. Also, on investigation of individual participants' performance across parameters, the direction of change was generally consistent across tasks.

Four individuals presented with an observed decline in function post load across instrumental measures (P3, P15, P27, P28). On review of specific characteristics of these participants, it was noted three of four were not trained singers and all had bilateral lesion presentations (P15, P27 nodules; P3, P28 left pseudocysts and right reactive lesions). Two of these participants (P27, P28) rated themselves as worse post load on self-report measures. Videostroboscopic still images of vocal fold abduction for P15 are presented in Image 1, and show increased redness/inflammation post load. This negative change in lesion characteristics is therefore

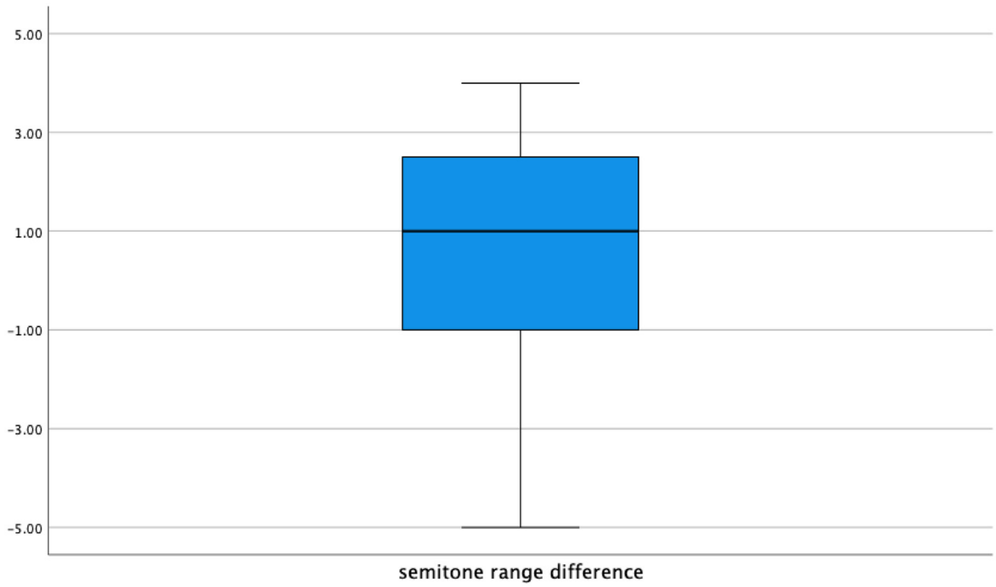

FIGURE 7. Range of difference in pre- to post-load participants scores for semitone range (SR), <0 indicates a lower score and >0 a higher score post load.

consistent with the findings of decline on instrumental measures for this participant. Baseline VOISS and EAVE scores for P15 were also both above the 75th percentile for the group, indicative of a higher self-perceived severity of vocal effort and impairment, though interestingly she did not perceive an increase in effort or decline in function post load.

Previous findings regarding the direction of change in PPE post load have not been consistent. Laukkanen, Jarvinen, Artkoski, Waaramaa-Maki-Kulmala, Kankare, Sippola, Syrja and Salo[10] found participants with previous voice training self-reported improved throat symptoms after 30 minutes of a loading task. Aronsson, Bohman, Ternstrom and Sodersten[7] found increased self-rating of strain for both controls and participants with nodules after short 90 second loading tasks at four increasing levels of background noise, and similarly Remacle, Morsomme, Berrue and Finck[11] found an increase in self perceived effort, fatigue and discomfort for teachers who were normophonic and for those with vocal nodules. The results of PPE in this study were multidirectional, with 18 (64%) participants showing a MCID of +/- 1 cm post load in both positive and negative directions. Similarly, 18 (64%) participants perceived either a positive or negative change in the physical aspects of their speaking voice according to MCID score post load on the EAVE. Therefore the consistency of findings for both PPE and EAVE suggest that many participants with BVFLs did perceive change after loading, for some an improvement and others a decline.

The multidirectional outcomes on self-report of effort and vocal function post load is in contrast to the generally positive changes shown in instrumental measures of vocal function. While the instrumental data suggests improved efficiency after loading, some participants perceived this change to be an increase in effort and decline in vocal function. Certainly, while instrumental measures provide insight into vocal function, they do not necessarily inform us how hard a speaker is working to achieve best voice for the task. Discrepancy between self-report and objective parameters of vocal function has previously been described[5] and highlights the multidimensional nature and complexity of voice measurement.

A strength of this study is in the exploratory analysis approach taken which yielded unexpected results and could have important implications for future clinical research in this area. Potentially, there may be previous pre/post loading studies that have concluded "no statistically significant difference" but they might have included cases which showed both positive and negative change scores, which cancelled out at

TABLE 4.
Median, Interquartile Range and MCID Values for Pre- and Post-Loading Participant Self-Reported Measures

Variable			Median	IQR	MCID	n	No Change	Change	*DecreaseIncrease*	*Range of Difference*
PPE	cm	Pre	3.30	1.73, 6.83	+/- 1 cm	28	10 (36%)	18 (64%)	*11 (39%)*	*-1.3 to -5.1*
		Post	4.50	1.43, 7.55					*7 (25%)*	*1.4-3.5*
EAVE	Score	Pre	25.5	24, 33	+/- 3	28	10 (36%)	18 (64%)	*8 (29%)*	*-3 to -11*
	/48	Post	26.0	22.25, 32.75					*10 (36%)*	*3-7*

TABLE 5.
MCID Representation for Aerodynamic, Acoustic and Self-Reported Measures: +/- Indicates Improved or Declined Function, Blank no Change and NA Missing Data

ID	Aerodynamic Measures				Acoustic Measures		Aerodynamic Overall	Acoustic Overall	Instrumental Overall	Self-Report Measures	
	Flow	Psub	PTP	MPT	SR	GNE				PPE	EAVE
1		-	+	+	+	+	+	+	+		+
3			+	-		-		-	-	+	+
4	-				+	+	-	+	+	+	+
5			+	+	+		+	+	+	-	-
6	+			+	+		+	+	+		
7	+		-	+	+	+	+	+	+		
9			+	+	-	+	+		+	-	-
10	-	+	+				+		+	+	+
11	NA	NA				+	NA	+	NA		
13				+			+		+		
15	-	-		-			-		-		
16	+	+	+				+		+	+	+
17		+	+	NA			+		+		
18	+		+		+		+	+	+	-	+
19	+			+	+		+	+	+	-	-
20	+	+	-		+		+	+	+	-	-
21			+		-	+	+		+	-	-
22	+	+	+	+	+	-	+		+	-	+
23					NA	NA		NA	NA	-	
24	-			+	-	+				+	+
25	+	+	+	+			+		+		+
27	+	-		-	-		-	-	-	-	-
28		-	-				-		-	-	-
29	+		+	+	+	+	+	+	+	+	-
30				+		+	+	+	+		
31		+	+		+		+	+	+	+	+
32	+			+			+		+		
33		-				+	-	+		-	
Improved sum							19 (70%)	13 (50%)	20 (77%)	7 (25%)	10 (38%)
Worse sum							5 (19%)	2 (7%)	4 (15%)	11 (39%)	8 (29%)
No change							3 (11%)	12 (44%)	2 (8%)	10 (38%)	10 (38%)

the group level. Although the majority of researchers in this topic area have utilised a traditional null hypothesis statistical testing (NHST) approach to the analysis of data, there is a growing literature raising serious concerns about the appropriateness and usefulness of these techniques. Some reviews go so far as to suggest that NHST generated statistical techniques have been "applied ritualistically and mindlessly as the dominant doctrine",[42] and that this dependence may have "retarded the growth of cumulative knowledge".[43] Lakens[44] concluded his review of this topic by suggesting it is important for researchers to carefully consider "what it is we actually want to know" (p. 645) and to ensure techniques and interpretations of results are appropriate.

Limitations and future directions

Inter-rater reliability for the parameter "breathy" was found to be poor in this study. While this finding is not widely reported in the literature, similar challenges utilising this parameter to evaluate change post load have been described.[45] In the current study, no calibration of voice recordings for loudness pre- and post-loading was made. This was a deliberate decision as an increase in loudness was anticipated post load and masking this change through calibration would impact on expert raters' perception of change. It is possible that the significant increase in intensity impacted expert perception of breathiness. For the global perceptual rating we found poor intra-rater reliability. In retrospect, this task requiring perceptual judgement of potentially subtle quality differences was novel and it may have warranted group discussion for consensus prior to rating, in order to establish perceptual anchors.[46] Similarly intra-rater and inter-rater reliability for stroboscopy videos and still images were generally poor. Again, this may highlight inherent difficulty rating subtle or minimal, if any, differences in form and function of pathology across two samples. For future studies, the implementation of structured rating tools or discussion between raters to reach a consensus may be a more successful way to approach this data. In addition the possibility of response bias should be considered when implementing perceptual analysis of pre- and post-stimulus observations. This study attempted to control for response bias by randomising the order or pre- and post-load voice and stroboscopy recordings, including the order of repeated samples for intra-rater reliability. Given that so few samples were rated as changed post load, order effect was not analysed further but is an important consideration in future research.

It is possible that the findings of change in instrumental measures of vocal function were due to inter-test measurement differences or training effects associated with repeating the same vocal tasks. The use of the MCIDs would however negate the likelihood of this as instructions did not differ and all participants had completed the tasks prior to involvement in this study. Establishing robust MCID values in this study was attempted using a combination of a distribution-based method using normative data, clinical application of parameters and previous use of parameters in the loading literature. The selected values may have been too low, resulting in parameters presenting as inaccurately sensitive to change. This should be considered in interpretation of results, and certainly future work on establishing reliable MCID values for voice research is needed.

Previous research has not shown consistent changes in stroboscopic findings in normophonic subjects post load. After vocal load, Linville[47] found inconsistent changes in glottic configuration for participants, Solomon and Di Mattia[33] showed spindle-shaped glottic closure post load in three of four participants whereas Stemple, Stanley and Lee[12] found an anterior glottic chink in 6 of 10 subjects who did not present with this glottic gap pre load. Given the participants in our study generally presented with hourglass glottic closure patterns secondary to BVFLs, previous stroboscopy findings on normophonic subjects are likely not applicable. There have been no known previous studies using videostroboscopy to evaluate post load change with the BVFLs population.

A relatively high number of professional voice users were recruited to this study including 12 participants who were trained singers. While aerodynamic and acoustic evaluation presents data that may be considered objective, these analyses are taken from functional voicing activities. It is hypothesised that professional voice users may have skill at masking the impact of BVFLs on vocal function assessed by these activities. Additionally professional voice users likely present with higher thresholds for vocal load.[9] Given the cohort of this study included more singers than in previous studies, it may in part explain why a positive response to loading was observed in instrumental measures. Of note, no pattern of difference was observed within the singer group, according to trained status or whether a professional singer, or between singers and the rest of the cohort across all parameters, with similar variability in response to loading for both groups. A similar observation was made in a study Laukkanen, Jarvinen, Artkoski, Waaramaa-Maki-Kulmala, Kankare, Sippola, Syrja and Salo[10] where previous training did not correlate with acoustic or self-report outcome measures following a 45-minute VLT. While no differences were found in this study, this may be evident with a larger sample size and warrants future study.

The VLT utilised was of a relatively conservative duration, given this was a study with a novel participant group. If the design of a VLT increased in duration or if additional intrinsic or extrinsic factors were implemented, evidence of vocal fatigue may well be observed. Further investigation of potential thresholds of vocal load is needed for participants with BVFLs and for normophonic speakers.

Designing a VLT that results in a consistent vocal response across speakers can be challenging in both normophonic and pathological sample groups and this study has proven no exception. A speaker's individual response to BVFL, pattern of vocal stamina and resilience are likely to be impacted by many factors and therefore variability in immediate vocal demand response to a VLT is anticipated. Some individuals may increase loudness in an efficient manner resulting in little to no change or improvement in observable measures of vocal function. Others may perform with less efficiency and present with lower thresholds before fatigue is observed. This variability of response was observed in the current study, certainly in some measures more than others, highlighting both the need to explore individual responses and to use multiple measures to further our understanding of the effect of loading on vocal function.

CONCLUSION

Participants with BVFLs demonstrate change in vocal function following 30 minutes of vocal load. Some participants perceived this change to be an increase in effort, some a reduction in effort and some perceived no change. While this change can be variable and multidirectional, overall improvement in instrumental measures of function and efficiency was observed for many participants. Lesion characteristics and vocal fold closure and vibration generally were not observed to change post load, or if change occurred this was likely to be subtle.

The clinical implications of this study's findings for treatment-seeking patients with BVFLs are somewhat surprising and suggest that recommendations for conservative voice use and relative vocal rest may not be necessary for certain individuals. Our results support the notion of evaluating individual response to increased vocal load or vocal demand when required, and where positive response is observed there may be value in further training for these speakers to use their voice well in these contexts. Although the cumulative effects of repeated loading for people with BVFLs have not been investigated in this study, further exploration of the potential to achieve sustained improvement across time in vocal function and in lesion characteristics (or even possible resolution) is warranted.

This study highlights the complexity of investigating vocal load and function and that immediate responses to vocal loading differ amongst individuals with BVFLs. The improvement in instrumental measures post load in this study suggest that controlled vocal load may have an immediate positive impact on vocal function for some patients with BVFLs. Further research is needed to determine the thresholds beyond which vocal load or vocal demand may result in decline in vocal function and physiology. At the very least, it would seem that clinical recommendations to reduce continuous vocal load for all people with established BVFL may not be universally well-founded without further exploration of an individual's threshold for change and the nature of their response. Indeed, to the contrary, improved vocal function and relative lesion stability can seemingly occur after loading in some pathological voices.

DECLARATIONS OF INTEREST

none

Acknowledgments

J. Pallant for her statistical support, N. Curham, A. Giannini, M. Baxter and P. Paddle for their expert ratings, and our generous subjects.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at doi:10.1016/j.jvoice.2021.11.009.

REFERENCES

1. Phyland DJ. The measurement and effects of vocal load in singing performance. How much singing can a singer sing if a singer can sing songs? *Perspect. ASHA special interest groups SIG 3*. 2017;2:79–88.
2. Hunter EJ, Cantor-Cutiva LC, van Leer E, et al. Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *JSLHR*. 2020;63:509–532.
3. Lollgen H, Leyk D. Exercise testing in sports medicine. *Dtsch Arztebl Int*. 2018;115:409–416.
4. Guazzi M, Bandera F, Ozemek C, et al. Cardiopulmonary exercise testing: What is its value? *J Am Coll Cardiol*. 2017;70:1618–1636.
5. Fujiki RB, Sivasankar MP. A review of vocal loading tasks in the voice literature. *J Voice*. 2017;31:388.e333–388.e339.
6. Bottalico P, Graetzer S, Hunter EJ. Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *J Acoust Soc Am*. 2016;139:2870.
7. Aronsson C, Bohman M, Ternstrom S, et al. Loud voice during environmental noise exposure in patients with vocal nodules. *Logoped Phoniatr Vocol*. 2007;32:60–70.
8. Erickson-Levendoski E, Sivasankar M. Investigating the effects of caffeine on phonation. *J Voice*. 2011;25:e215–e219.
9. Enflo L, Sundberg J, McAllister A. Collision and phonation threshold pressures before and after loud, prolonged vocalization in trained and untrained voices. *J Voice*. 2013;27:527–530.
10. Laukkanen AM, Jarvinen K, Artkoski M, et al. Changes in voice and subjective sensations during a 45-min vocal loading test in female subjects with vocal training. *Folia Phoniatr Logop*. 2004;56:335–346.
11. Remacle A, Morsomme D, Berrue E, et al. Vocal impact of a prolonged reading task in dysphonic versus normophonic female teachers. *J Voice*. 2012;26. 820.e821-820.e813.
12. Stemple J, Stanley J, Lee L. Objective measures of voice production in normal subjects following prolonged voice use. *J Voice*. 1995;9:127–133.
13. White A. Management of benign vocal fold lesions: current perspectives on the role for voice therapy. *Curr Opin Otolaryngol Head Neck Surg*. 2019;27:185–190.
14. Branski RC, Verdolini K, Sandulache V, et al. Vocal fold wound healing: a review for clinicians. *J Voice*. 2006;20:432–442.
15. Haben CM. Voice rest and phonotrauma in singers. *Med Probl Perform Art*. 2012;27:165–168.
16. Lee M, Mau T, Sulica L. Patterns of recurrence of phonotraumatic vocal fold lesions suggest distinct mechanisms of injury. *Laryngoscope*. 2012;131:2523–2529.
17. Vasconcelos D, Gomes AOC, Araujo CMT. Vocal fold polyps: Literature review. *Int Arch Otorhinolaryngol*. 2019;23:116–124.
18. van der Merwe A. The voice use reduction program. *Am J Speech Lang Pathol*. 2004;13:208–218.
19. Holmberg EB, Hillman RE, Hammarberg B, et al. Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *J Voice*. 2001;15:395–412.
20. Niebudek-Bogusz E, Kotylo P, Sliwinska-Kowalska M. Evaluation of voice acoustic parameters related to the vocal-loading test in professionally active teachers with dysphonia. *Int J Occup Med Environ Health*. 2007;20:25–30.
21. Kelchner LN, Lee L, Stemple JC. Laryngeal function and vocal fatigue after prolonged reading in individuals with unilateral vocal fold paralysis. *J Voice*. 2003;17:513–528.

22. Jilek C, Marienhagen J, Hacki T. Vocal stability in functional dysphonic versus healthy voices at different times of voice loading. *J Voice*. 2004;18:443–453.
23. Deary IJ, Wilson JA, Carding PN, et al. VoiSS. *J Psychosom Res*. 2003;54:483–489.
24. Wilson JA, Webb A, Carding PN, et al. The Voice Symptom Scale (VoiSS) and the Vocal Handicap Index (VHI): a comparison of structure and content. *Clin Otolaryngol*. 2004;29:169–174.
25. Sanchez K, Oates J, Dacakis G, et al. Speech and voice range profiles of adults with untrained normal voices: methodological implications. *Logoped Phoniatr Vocol*. 2014;39:62–71.
26. Oates J, Russel A. Learning voice analysis using an interactive multimedia package: development and preliminary evaluation. *J Voice*. 1998;12:500–551.
27. Phyland DJ, Pallant J, Free N, et al. On the EAVE: A new scale for self-evaluation of the physical aspects of speaking voice function In: IALP Voice Composium. 2021.
28. Balluerka N, Gómez J, Hidalgo D. The controversy over null hypothesis significance testing revisited. *Methodology*. 2005;1:55–70.
29. Zraick RI, Smith-Olinde L, Shotts LL. Adult normative data for the KayPENTAX Phonatory Aerodynamic System Model 6600. *J Voice*. 2012;26:164–176.
30. Leung Y, Oates J, Papp V, et al. Speaking fundamental frequencies of adult speakers of Australian English and effects of sex, age, and geographical location. *J Voice*. 2020:1–15. In press.
31. Godino-Llorente JI, Osma-Ruiz V, Saenz-Lechon N, et al. The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders. *J Voice*. 2010;24:47–56.
32. Sivasankar MP, Erickson-Levendoski E. Influence of obligatory mouth breathing, during realistic activities, on voice measures. *J Voice*. 2012;26. 813.e819-813.e813.
33. Solomon NP, Di Mattia MS. Effects of a vocally fatiguing task and systemic hydration on phonation threshold pressure. *J Voice*. 2000;14:341–362.
34. McGlothlin AE, Lewis RJ. Minimal clinically important difference defining what really matters to patients. *JAMA*. 2014;312:1313–1322.
35. Weels G, Beaton D, Shea B, et al. Minimal clinically important differences: Review of methods. *J Rheumatology*. 2001;28:406–412.
36. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice*. 3rd Edition New Jersey: Pearson Prentice Hall; 2009.
37. Chang A, Karnell MP. Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: a study of vocal fatigue. *J Voice*. 2004;18:454–466.
38. Sundarrajan A. Hydration and vocal loading on voice measures. *Open Access Dissertations*. 855. Perdue University; 2016.
39. Sundarrajan A, Huber JE, Sivasankar MP. Respiratory and laryngeal changes with vocal loading in younger and older individuals. *J Speech Lang Hear Res*. 2017;60:2551–2556.
40. Whitling S, Lyberg-Ahlander V, Rydell R. Long-time voice accumulation during work, leisure, and a vocal loading task in groups with different levels of functional voice problems. *J Voice*. 2017;31. 246. e241-246.e210.
41. Remacle A, Finck C, Roche A, et al. Vocal impact of a prolonged reading task at two intensity levels: objective measurements and subjective self-ratings. *J Voice*. 2012;26:e177–e186.
42. Silva-Ayçaguer LC, Suárez-Gil P, Fernández-Somoano A. The null hypothesis significance test in health sciences research (1995-2006): statistical analysis and interpretation. *BMC Med Res Methodol*. 2010;10:1–9.
43. Schmidt FL. Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers. *Psychol Methods*. 1996;1:115–129.
44. Lakens D. The practical alternative to the p value is the correctly used p value. *Perspect Psychol Sci*. 2021;16:639–648.
45. Remacle A, Schoentgen J, Finck C, et al. Impact of vocal load on breathiness: perceptual evaluation. *Logoped Phoniatr Vocol*. 2014;39:139–146.
46. Iwarsson J, Reinholt Petersen N. Effects of consensus training on the reliability of auditory perceptual ratings of voice quality. *J Voice*. 2012;26:304–312.
47. Linville SE. Changes in glottal configuration in women after loud talking. *J Voice*. 1995;9:57–65.