The Laryngoscope

Impact of Audio-Visual Complexity on Symptomatology of Laryngeal Dystonia: A Virtual Reality Study

Jimmy Petit, PhD; Stefan K. Ehrlich, PhD; Garrett Tougas, BS; Jacob M. Bernstein, BA; Nicole E. Buie, BS; Kristina Simonyan, MD, PhD

Background: Laryngeal dystonia (LD) is an isolated focal dystonia characterized by involuntary spasms in laryngeal muscles selectively impairing speech production. Anecdotal observations reported the worsening of LD symptoms in stressful or vocally demanding situations.

Objectives: To examine the impact of surrounding audio-visual complexity on LD symptomatology for a better understanding of disorder phenomenology.

Methods: We developed well-controlled virtual reality (VR) environments of real-life interpersonal communications to investigate how different levels of audio-visual complexity may impact LD symptoms. The VR experiments were conducted over five consecutive days, during which each patient experienced 10 h of 4100 experimental trials in VR with gradually increasing audio-visual complexity. Daily reports were collected about patients' voice changes, as well as their comfort, engagement, concentration, and drowsiness from using VR technology.

Results: After a weekly VR exposure, 82% of patients reported changes in their voice symptoms related to changes in background audio-visual complexity. Significant differences in voice symptoms were found between the first two levels of the audio-visual challenge complexity independent of study sessions or VR environments.

Conclusion: This study demonstrated that LD symptoms are impacted by audio-visual background across various virtual realistic settings. These findings should be taken into consideration when planning behavioral experiments or evaluating the outcomes of clinical trials in these patients. Moreover, these data show that VR presents a reliable and useful technology for providing real-life assessments of the impact of various experimental settings, such as during the testing of novel therapeutic interventions in these patients.

Key Words: background noise, isolated dystonia, virtual reality.

Level of Evidence: 3

***Laryngoscope*, 135:787–793, 2025**

INTRODUCTION

Laryngeal dystonia (LD) is an isolated focal dystonia characterized by involuntary spasms in laryngeal muscles that selectively impair the production of speech but not other vocal behaviors, such as whispering, laughing, or crying.[1] LD symptoms detrimentally affect patients' quality of life, chronically restricting their ability to communicate during daily and professional activities. LD symptoms are known to wax and wane, often worsening in stressful situations, during telephone conversations, or due to temporary voice demands, such as public speaking or communication in crowded and noisy surroundings.[2,3] However, systematic investigations of this clinical phenomenon and the impact of various situational environments on LD symptoms remain scarce due, in part, to the unfeasibility of continuous patient evaluations during their daily social activities or public engagements. This, in turn, hinders the complete understanding of the disorder phenomenology and the factors influencing LD symptom fluctuations, which further lowers the translational impact of behavioral and therapeutic interventions.

In this study, we used a novel approach of virtual reality (VR) to examine how realistic situational environments with different audio-visual backgrounds, such as being in a restaurant, shopping mall, office, or an outdoor urban setting, may impact LD symptomatology. Over the past decade, VR has steadily increased its applicability in medical settings, especially in the training of healthcare professionals and the rehabilitation of stroke patients to restore balance, gait, and upper limb motor function.[4–6] In other forms of dystonia, a recent study successfully implemented VR to control the administration of stimuli during the assessment of vestibular heading discrimination in patients with cervical dystonia.[7] Although not

From the Department of Otolaryngology-Head & Neck Surgery (J.P., S.K.E., G.T., J.M.B., N.E.B., K.S.), Harvard Medical School and Massachusetts Eye and Ear, Boston, Massachusetts, U.S.A.; Department of Neurology (K. S.), Massachusetts General Hospital, Boston, Massachusetts, U.S.A.

Additional supporting information may be found in the online version of this article.

Editor's Note: This Manuscript was accepted for publication on September 09, 2024.

Jimmy Petit, Stefan K. Ehrlich, Garrett Tougas, Jacob Bernstein, and Nicole E. Buie have nothing to disclose. Kristina Simonyan receives funding from the National Institutes of Health (R01NS088160, R01NS124228, R01DC011805, R01DC012545, R01DC019353, P50DC01990) and serves on the Scientific Advisory Board of the Tourette Association of America and the Voice Foundation.

The authors have no funding, financial relationships, or conflicts of interest to disclose.

Send correspondence to Kristina Simonyan, Department of Otolaryngology-Head and Neck Surgery, Harvard Medical School, 243 Charles Street, Suite 421, Boston, MA; Email: kristina_simonyan@meei.harvard.edu

DOI: 10.1002/lary.31800

used to date in LD patients, VR allows for controlled, reliable, and reproducible simulation of realistic environments that are otherwise unfeasible in a typical clinical or research setting. For example, without VR, comparable data in LD patients for this study could have been collected only in field studies at restaurants, shopping malls, work offices, and outdoors. As such an approach is impractical, we developed experimental VR environments of real-life situations with realistically looking avatars and audio-visual backgrounds that simulated common social, work, and outdoor settings to investigate their impact on LD symptomatology. Based on the previous anecdotal clinical observations, we hypothesized that LD symptoms would worsen in environments that require increased vocal demands during speaking.

MATERIAL AND METHODS

Study Participants

Fourteen patients with isolated focal LD participated in the study (see detailed demographics in Table I). All patients were native English speakers and right-handed or ambidextrous based on the Edinburgh Handedness Inventory. No patient had any past or present history of neurological (except for LD and co-occurring dystonic tremor of voice), psychiatric, or laryngeal problems as determined by the review of a case history, laryngeal/neurological evaluations, and perceptual analysis of voice and speech recordings. All patients had normal cognitive function as assessed with the Montreal Cognitive Assessment (MoCA). Eleven out of 14 patients received at least one botulinum toxin (BoNT) injection. Among these, 8 patients reported no benefits and discontinued the treatment. The remaining 3 patients were on active BoNT treatment to manage their symptoms. All patients who received BoNT injections participated in this study at least 3 months after their last injection and were fully symptomatic at the time of study participation. None of the patients were on any centrally acting medications for at least 2 weeks prior to and during study participation.

TABLE I.
Patient Demographics.

Number of participants	14
Age (years; mean ± st.dev.)	54.0 ± 12.9
Sex (female/male)	8/6
Handedness Edinburgh	13 Right; 1 Ambidextrous
Language	Native English
Cognitive status (MoCA; mean ± st.dev.)	28.0 ± 1.7
Last botulinum toxin treatment (years; mean ± st.dev.)	11 patients: 3.4 ± 4.6
LD phenotype	8 ADLD
	1 ADLD/DTv
	4 ABLD
	1 ABLD/DTv
LD age of onset (years; mean ± st.dev.)	34.0 ± 19.4
LD duration (years; mean ± st.dev.)	14.2 ± 10.1
LD severity (mean ± st.dev.)	
Self-reported voice effort to talk (no effort 1 – constant struggle 9)	7.4 ± 1.0

LD = laryngeal dystonia; MoCA = Montreal Cognitive Assessment; st.dev. = standard deviation.

Ethical Compliance

All patients gave written informed consent before study participation, approved by the Institutional Review Board of Mass General Brigham. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines.

Study Procedures

The study was conducted over five consecutive days (Fig. 1A). All patients completed two experimental sessions per day, each approximately 60 min, with a 1-h break between sessions. During each session, the patient was seated in a comfortable armchair in front of a desk, wearing a head-mounted VR display (HTC VIVE Pro) to deliver the experimental paradigm.

All VR scenes were developed using Unity Pro software, version 2021.3.8f1. Each VR scene included an avatar embedded in a realistic environment (Fig. 1B). Six avatars (three females and three males) and four VR environments (a restaurant, a shopping mall, an office, and an outdoor urban setting), each with four different levels of audio-visual complexity (Levels 1 to 4), were used in the experiment (Fig. 1B). The avatar articulated grammatically correct English sentences, which were prerecorded by a native English female or male speaker. A total of 40 sentences loaded with vowels and voiceless consonants to elicit LD symptoms[8] were used as experimental stimuli in each session.

Level 1 of each VR environment included a quiet background setting, during which the patient and the avatar had one-to-one interactions without additional audio-visual background features (Fig. 1B). The audio-visual virtual features were gradually increased in their complexity in the background from Level 2 to Level 4 (Fig. 1C). Specifically, in Level 2, additional virtual characters in motion were added to the background, away from the avatar and patient. In Level 3, these virtual characters were positioned closer to the avatar and the patient. In Level 4, additional virtual characters were added, such as avatars dancing to loud music in the restaurant scene, children avatars running in the shopping mall scene, a pizza party in the office scene, or avatars running away from a burning building (Fig. 1B, Table II). The loudness of each scene level, measured by the mean decibel full scale (dB FS) of audio recordings, was −49.76 at Level 1, −42.57 at Level 2 (+15.6% from level 1), −38.31 at Level 3 (+26.0% from level 1), and − 36.67 at Level 4 (+30.3% from level 1) (Fig. 1C).

For each patient, the avatar and the initial VR environment were pseudo-randomized on the first day of the study. Each patient was presented with two different VR scenes of the same level per session; each scene consisted

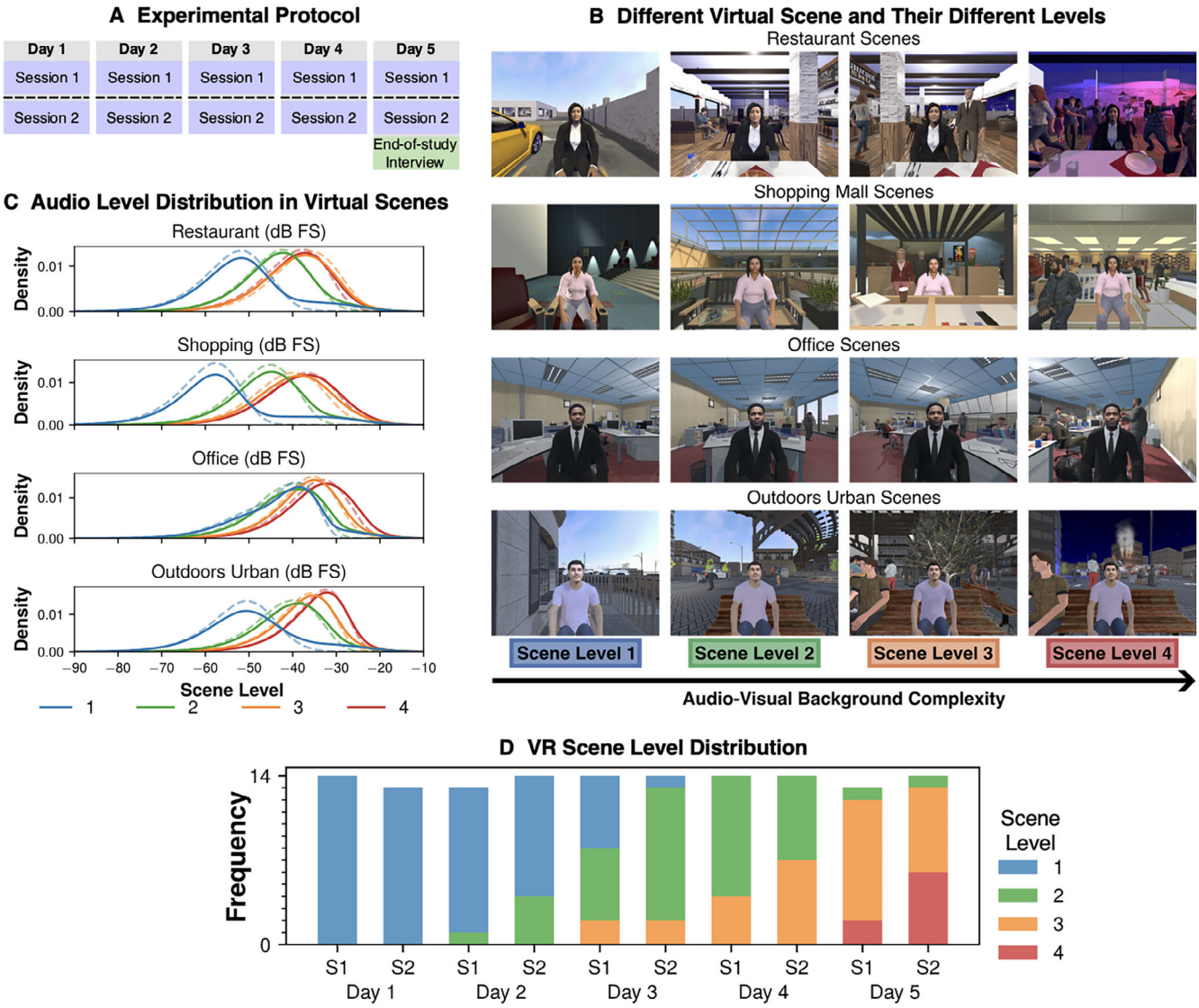

Fig. 1. (A) The experimental protocol included five consecutive days. Each patient underwent two sessions each day. Patients completed the user experience questionnaire after sessions of each day and the end-of-the-study interview at the end of the last session of the last day. Dashed lines between sessions indicate a between-session break. (B) Different VR scenes and their different levels. Rows represent various environments, including a restaurant, a shopping mall, an office, and an outdoor urban environment. Columns show the respective scenes from Levels 1 to Level 4. (C) Audio level distribution in virtual scenes. Decibel Full Scale (dB FS) distributions of audio recordings of each virtual scene, with two main components: the pre-recorded native English voices articulating the experimental stimulus combined with the environment-specific ambient sounds in solid lines, and only environment-specific ambient sounds in dash-colored lines. (D) VR scene level distribution. Patient participation in each level and session throughout the experiment. [Color figure can be viewed in the online issue, which is available at www.laryngoscope.com.]

of 10 blocks. Patients were allowed to take a short break between blocks if needed. The first session of the first day started with the Level 1 VR environment and progressively increased between experimental sessions but not within a session. The leveling up was performed by considering (1) the patient's feedback about the scene difficulty for communication and (2) a minimum proportion of all patients at each scene level for comparability of different scene levels. The goal was to avoid patients' boredom, maintain their engagement in the study, increase the challenge of the task, and have the patients experience different real-life scenarios. Specifically, all patients (100%) were set to participate in Level 1; at least 80% of patients were set to participate in Level 2; at least 60% of patients were considered to participate in Level 3; and at least 40% of patients were set to participate in Level 4 (Fig. 1D, Supplemental Fig. 1A).

During each session, the patient repeated sentences that were pseudo-randomly spoken by the avatar, followed by an open-ended question from the avatar to engage the patient in free speech for 30 s. Each session lasted 60 min; thus, each patient experienced 10 h of the VR environment, completing a total of 4000 sentence trials and 100 free-speech trials. Following each session, the

TABLE II.
Description of Auditory Components Per Scene Level.

	Level 1	Level 2	Level 3	Level 4
Restaurant	Continuous far unintelligible conversation. Occasional drilling sound and car engine starting.	Continuous unintelligible conversation. Occasional drilling sound and cutlery clinging sound.	Continuous unintelligible multiple conversations. Occasional service ring bell, cash register, and cutlery clinging sound.	Continuous techno music. Occasional service ring bell. Many people dancing.
Shopping mall	Continuous far unintelligible conversation.	Continuous soft pedestrian sound. Occasional soft baby babble.	Continuous soft pedestrian sound. Occasional low baby babble, child screaming, mall announcement ringtone, and people coughing.	Continuous pedestrian sound. Occasional baby babble, child screaming, mall announcement ringtone, people coughing, and shopping alarm.
Office	Continuous soft computer running sound.	Continuous soft computer running sound. Frequent keyboard typing sound. Occasional printing sound.	Continuous unintelligible conversation and computer running sound. Frequent keyboard typing, mouse clicking, and phone ringing.	Continuous far sirens and office party multiple conversations. Frequent keyboard typing, mouse clicking, and phone ringing.
Urban outdoors	Frequent soft sound of passing cars. Occasional soft chirping bird.	Continuous unintelligible conversation. Frequent sound of passing cars.	Continuous sound and rain sound and various traffic sounds. Occasional distant honks and soft baby babbles.	Continuous sound of rain, traffic, sirens, and a helicopter sound. Frequent honks. Apartment on fire in the background. Several adults running.

patients completed a structured questionnaire about their experience in the VR environment (User Experience Questionnaire), which included an 11-item Likert scale (range: −5 to 5 or 0 to 10, as applicable) for the self-assessment of voice symptoms, comfort, engagement, concentration, and drowsiness from being in the VR environment (Table III). At the end of the study, a semi-structured interview (End-of-Study Interview) was conducted with each patient to collect information about their perceived experience and perspectives in the VR environment (Table IV).

Data Processing and Statistical Analysis

Data from each patient's User Experience Questionnaire and the exposure time to each scene level were extracted and used as variables of interest in the analysis (Supplementary Figs. 1 and 2).

As the first step, the outlier analysis identified two patients (both with adductor LD) whose variables of interest had a median of more than two standard deviations compared with the median of all patients. One of these patients had the outlier scores on the questions related to concentration at Level 1 and engagement at Levels 1 and 2. The other patient had the outlier scores on the questions related to the self-reported changes in voice symptoms at Level 1, concentration and drowsiness at Level 2, drowsiness at Level 3, and engagement at Level 4. Additionally, one patient experienced discomfort with the VR headset and, therefore, did not progress further than Level 2. All three excluded outlier patients were fully symptomatic, with their self-reported voice symptoms scored at 7.0, which was within the range of voice symptoms of the entire patient cohort (Table I). It is, therefore, unlikely that these excluded patients were phenomenologically different from the rest of the cohort or that their exclusion resulted in a non-representative patient sample for statistical analysis.

The final cohort included in statistical analysis comprised 11 LD patients (6 females/5 males, mean age of 54.5 ± 10.7 years). The primary outcome was defined as the patient's self-assessments of their voice symptoms in VR environments with various audio-visual background complexity. We specifically refrained from the conduct of objective perceptual acoustic evaluation of LD symptoms in this study because the goal was to capture and evaluate patients' individual experiences related to their LD symptomatology during situational communication.

TABLE III.
User Experience Questionnaire.

Q1	Did you experience any changes in your voice symptoms? (−5, Worsened to +5, Improved)
Q2	Rate your level of comfort throughout the session (−5, Uncomfortable to +5, Comfortable)
Q3	Rate your level of engagement in the task (0, None to 10, High)
Q4	Rate your level of concentration throughout the session (0, None to 10, High)
Q5	Rate the level of drowsiness you experienced during the experiment (0, None to 10, High)

TABLE IV.
End-of-Study Interview Questionnaire.

Q1	How did your voice quality change throughout the week?
Q2	How do you describe the whole week?
Q3	Describe your level of engagement/fun/boredom throughout the week.
Q4	What is your general assessment of the duration of the experiment (1 week), the number of sessions per day (2), and the duration of each session (1.5–2 h)? Do you think there were enough breaks?
Q5	What is your opinion about the scenes and levels?
Q6	How do you describe the leveling up in terms of increased challenges?
Q7	What is your opinion about VR vs. non-VR experimental settings?

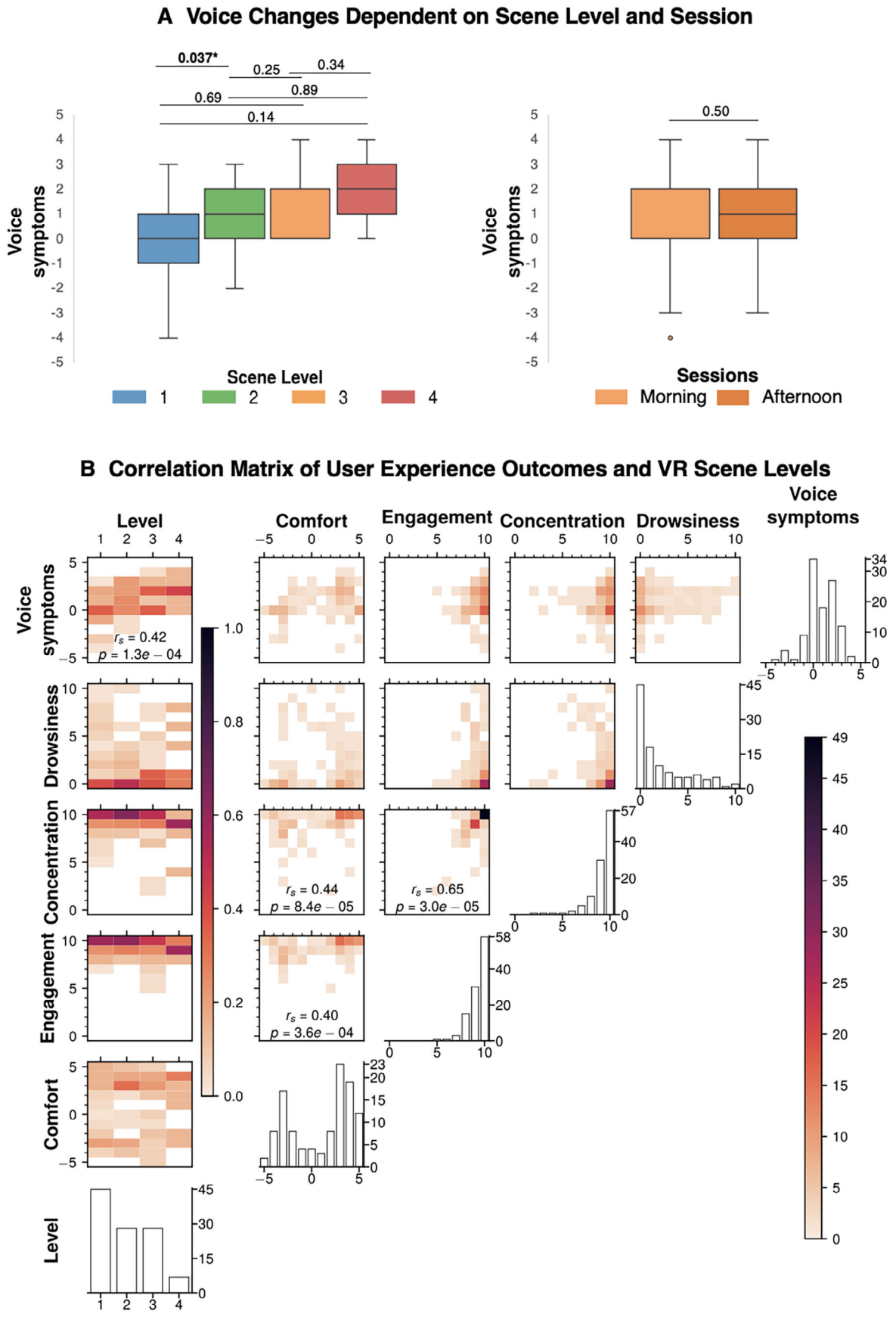

Fig. 2. Legend on next page.

Statistical analysis of between-group and within-group self-assessed voice symptom effects was performed using a linear mixed-effects (LME) model with three fixed factors, including session (2 factors encoding for morning or afternoon session), level (4 factors encoding for scene level difficulty), scene (12 factors encoding every couple of VR scene type used in one session, e.g., restaurant/office); a random intercept was used for every patient as a random factor at the overall statistical significance of $p \leq 0.05$. Pairwise post hoc comparisons within significant factor levels were performed using the Scheffé test to account for multiple comparisons. Next, Spearman rank correlation analyses were performed between the variables of interest to examine the global effect of the scenes on patients. The significance of correlations was assessed using a permutation test at $r_s \geq 0.3$ and Holm-Bonferroni-adjusted $p \leq 0.01$.

Finally, the qualitative analysis of the End-of-Study Interview was conducted to identify the lexical fields used in the answers, group them into clusters of similar positive, negative, or neutral answers, and extract their frequency.

Data Availability

All data relevant to research information of the datasets used in this study are included in the manuscript. The dataset may be requested from the corresponding author (KS), subject to the Data User Agreement, and approval by the Mass General Brigham Data and Tissue Sharing Committee.

RESULTS

The LME analysis found overall significant effects of level ($F_{3,47} = 6.60$, $p = 0.001$) and session ($F_{1,47} = 4.47$, $p = 0.04$) on LD symptoms but not scene or any interactions between these factors (all $F \leq 1.69$, $p \geq 0.18$). Pairwise post hoc tests found significant changes in self-reported voice symptoms between scenes of Level 1 and Level 2 (difference $= -2.59$; 95% confidence interval [CI] $= -5.08$ to -0.12, $p = 0.037$) but not between other levels or between morning and afternoon sessions ($p = 0.50$) (Fig. 2A). Self-reported changes in voice symptoms were significantly and positively correlated with scene level ($r_s = 0.42$, $p = 1.3\text{e-}04$). Additionally, there were significant positive correlations between self-reported engagement and concentration ($r_s = 0.65$, $p = 3.0\text{e-}05$), engagement and comfort ($r_s = 0.40$, $p = 3.6\text{e-}04$), and comfort and concentration ($r_s = 0.44$, $p = 8.4\text{e-}05$) during VR sessions (Fig. 2B).

The End-of-Study interview showed that 9 (82%) out of 11 patients perceived a change in their voice symptoms, either worsening or improving throughout the study. Five (45%) patients found the experiment to be challenging due to the increased complexity of audio-visual background noise during speaking. However, two of them indicated that they nevertheless enjoyed the experiment. When asked about their level of engagement throughout the 5 days of study, 6 (55%) patients reported experiencing steady engagement, whereas 3 (27%) patients stated a varying level of engagement, particularly between morning and afternoon sessions. Eight (73%) patients found the study duration to be acceptable and manageable in terms of session frequency and duration; 3 (27%) patients deemed the study too long.

When questioned about the virtual scenes, level-to-level difficulty, and the use of the VR headset, 7 (64%) patients reported that being in the VR environment helped increase their engagement, immersion, or focus during the study. Eight (73%) patients reported increased difficulty throughout the scene levels, whereas 4 (36%) patients found no relationship between difficulty/challenge and scene level or reported that the task became easier as the level increased. One patient fell into both categories. This patient acknowledged an increased distraction as the scene level increased, but also stated that it helped maintain their engagement, and as a result, they found the task easier as the difficulty increased.

DISCUSSION

In this study, we used a novel VR experimental paradigm of simulated realistic daily situations to demonstrate that the majority (82%) of LD patients report changes in their voice symptoms associated with increased audio-visual background noise complexity. Interestingly, we found that a statistically significant difference is observed between Levels 1 and 2 but not other levels that have higher audio-visual complexity. That is, the largest change in voice symptoms was perceived to be at the initial injection of the background noise, while the subsequent increases had only incremental effects on LD symptoms. This finding suggests that background noise, but not its increased complexity, affects LD symptomatology. Moreover, there were no differences in self-reported voice symptoms between different VR scenes, which indicates that various noisy environments may equally affect voice symptoms independent of their situational configuration. Lastly, contrary to the previous reports of voice symptom fluctuations throughout the day,[9] no differences were found in self-reported assessments of LD symptoms between morning and afternoon sessions.

There is a possibility that habituation to the experimental stimulus or virtual scenes might have somewhat

Fig. 2. (A) Voice changes dependent on scene level (left) and session (right). The numbers indicate the corrected *p*-values computed using the Scheffé test. The asterisk indicated the statistical significance. (B) The correlation matrix of user experience outcomes and VR scene levels. Pairwise correlations between scene levels and each question of the user experience questionnaire. The color bar shows the answer frequency in 2D histograms. The histograms involving the "Level" variable had bins normalized between 0 and 1 for each level, thus enhancing comparability between levels. Annotations show the Spearman correlation coefficient. Only significant correlations using a permutation test at $r_s \geq 0.3$ and a Holm-Bonferroni-adjusted $p \leq 0.01$ are annotated. The numbers indicate adjusted *p*-values computed using a permutation test with the Holm-Bonferroni test. [Color figure can be viewed in the online issue, which is available at www.laryngoscope.com.]

diminished the impact of different levels. Additionally, the use of the Likert scale for self-assessment of symptoms might have introduced the agreement bias, inclining patients to agree with a repeatedly presented statement regardless of how they felt during the experiment.[10,11] However, it is important to note that the level of comfort, engagement, concentration, and drowsiness from being in the VR environment was not correlated with either voice symptoms or scene level. On the contrary, the majority of patients reported that the use of the VR headset helped them increase their engagement, immersion, and focus and minimized potential surrounding distractions during the study. This statement is supported by data showing that the level of engagement was strongly and positively associated with the level of concentration, indicating that patients stayed focused and motivated throughout the study, which likely helped with higher concentration during the experiment. Taken together, this study demonstrated that the VR-based realistic experimental conditions provide an adequately controlled challenge to LD patients without provoking discomfort or causing adverse events. VR may be reliable and useful for future integration with various experimental paradigms for testing novel therapeutic interventions in LD patients, such as non-invasive neuromodulation or neurofeedback rehabilitation.[12]

A few limitations of this study should be acknowledged. One is related to a relatively small sample size in this first VR study in LD patients. Another limitation is that our experimental protocol tested different scene levels in the incremental order of complexity, which helped keep the patients engaged and minimize their stress and anxiety in unexpectedly changing VR environments. Future studies should consider larger study cohorts and the randomization of scene levels to further increase the complexity of virtual settings and to avoid potential habituation effects.

CONCLUSION

This study showed that the severity of voice symptoms in LD patients is impacted by the initial but not subsequent increases in audio-visual background noise complexity in various realistic environments. Future studies in LD patients may have a higher translational potential if the outcomes of interventions are evaluated in simulated settings with increased audio-visual background noise rather than noise-free labs or clinical settings.

ACKNOWLEDGMENT

This study was funded by the grant R01DC019353 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, to KS.

BIBLIOGRAPHY

1. Guiry S, Worthley A, Simonyan K. A separation of innate and learned vocal behaviors defines the symptomatology of spasmodic dysphonia. *Laryngoscope*. 2019;129:1627-1633.
2. Blitzer A, Kohli N. Laryngeal dystonia: phenomenology, genetics, and management. *Toxicon*. 2023;233:107258.
3. Brin MF, Blitzer A, Velickovic M. Anatomy of the larynx. *Movement Disorders of the Larynx Neurologic Disorders of the Larynx*. New York: Thieme; 2009.
4. Kang HK, Kim Y, Chung Y, Hwang S. Effects of treadmill training with optic flow on balance and gait in individuals following stroke: randomized controlled trials. *Clin Rehabil*. 2012;26:246-255.
5. Subramanian SK, Lourenço CB, Chilingaryan G, Sveistrup H, Levin MF. Arm motor recovery using a virtual reality intervention in chronic stroke: randomized control trial. *Neurorehabil Neural Repair*. 2013;27:13-23.
6. Pottle J. Virtual reality and the transformation of medical education. *Future Healthc J*. 2019;6:181-185.
7. Agharazi H, Wang A, Guha A, Gupta P, Shaikh AG. Unraveling the twist: spatial navigational challenges in cervical dystonia. *Mov Disord*. 2023;38:2116-2121.
8. Simonyan K, Barkmeier-Kraemer J, Blitzer A, et al. Laryngeal dystonia: multidisciplinary update on terminology, pathophysiology, and research priorities. *Neurology*. 2021;96:989-1001.
9. Ludlow CL, Domangue R, Sharma D, et al. Consensus-based attributes for identifying patients with spasmodic dysphonia and other voice disorders. *JAMA Otolaryngol Head Neck Surg*. 2018;144:657-665.
10. Schuman H, Ludwig J. The norm of even-handedness in surveys as in life. *Am Sociol Rev*. 1983;48:112-120.
11. Pasek J, Krosnick JA. Optimizing survey questionnaire design in political science: insights from psychology. In: Leighley JE, ed. *The Oxford Handbook of American Elections and Political Behavior*. Oxford University Press; 2010 0.
12. Simonyan K, Ehrlich SK, Andersen R, et al. Brain–computer interfaces for treatment of focal dystonia. *Mov Disord*. 2022;37:1798-1802.