Received: 24 August 2023 | Revised: 17 November 2023 | Accepted: 13 December 2023

DOI: 10.1002/lio2.1204

ORIGINAL RESEARCH

Laryngoscope Investigative Otolaryngology

Development and evaluation of a new intraoral voice assist device called the voice retriever

Taishi Yamada DDS[1] | Kohei Yamaguchi DDS, PhD[1] | Ayane Horike DDS[1] | Kohei Takahashi Dyploma of Speech Therapist[2] | Sirinthip Amornsuradech DDS, MPH[1,3] | Kazuharu Nakagawa DDS, PhD[1] | Kanako Yoshimi DDS, PhD[1] | Haruka Tohara DDS, PhD[1]

[1]Department of Dysphagia Rehabilitation, Division of Gerontology and Gerodontology, Tokyo Medical and Dental University, Tokyo, Japan

[2]Department of Rehabilitation, Japanese Red Cross Osaka Hospital, Osaka, Japan

[3]Department of Community Dentistry, Mahidol University, Bangkok, Thailand

Correspondence
Haruka Tohara Tokyo Medical and Dental University, 1-5-45, Yushima, Bunkyo, Tokyo, 113-8549, Japan.
Email: harukatohara@hotmail.com

Abstract

Objective: Patients lose their voice after laryngectomy for laryngeal cancer or aspiration prevention surgery for severe dysphagia. To assist such patients, we developed and verified the utility of a novel vocalization method using a device termed the voice retriever (VR), in which the sound source is placed in the mouth.

Methods: We investigated the effectiveness of the VR in patients. The VR consists of a mouthpiece with a built-in speaker and a dedicated application that serves as the sound source. We compared the speech intelligibility and naturalness in normal participants using VR and an electrolarynx (EL) for the first time as well as the voice-related quality of life (V-RQOL) in patients with dysphonia before and after using the VR.

Results: The VR produced significantly higher 100-syllable test scores as well as fluency, amount of additional noise, intonation, intelligibility and overall long reading test ratings in the first-time VR and EL users. Furthermore, the VR use significantly improved the V-RQOL of participants with dysphonia.

Conclusion: Compared to EL, VR allows more effective speech improvement in participants without experience using an alternative vocalization method and improves the V-RQOL in patients with dysphonia.

Level of Evidence: Step 4.

KEYWORDS

dysphonia, electrolarynx, speech intelligibility, vocalization

1 | INTRODUCTION

Dysphonia may have several causes, including laryngectomy for laryngeal cancer, laryngeal-tracheal separation for severe dysphagia, and ventilator management due to decreased respiration caused by neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS). Among patients diagnosed with laryngeal cancer, 54% lose their voice after total laryngectomy.[1] In Japan, 20% of patients with ALS are on tracheostomy positive-pressure ventilation,[2] which has a direct and major effect on their quality of life.[3] Currently, an electrolarynx

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 The Authors. *Laryngoscope Investigative Otolaryngology* published by Wiley Periodicals LLC on behalf of The Triological Society.

(EL) is the most widely used alternative vocalization method for patients who have lost their voice.[4]

An EL is a hand-held machine with a vibrating tip, which when pressed against the neck vibrates the air in the pharynx through the skin to enable articulation.[5] Despite their widespread use, EL devices are associated with numerous problems, including the location of the sound source outside the body. As a result, the vibrating sound made by the EL is directly heard by the listener. Additionally, training is required to correctly position the EL and an EL cannot be used if a patient has trouble using both hands.[6] Furthermore, if the skin of the neck is hardened by radiotherapy, vibrations are not adequately transmitted to the pharynx, making the device unusable.[5]

Esophageal speech involves training a patient to first collect air in the stomach and lower esophagus. The air is then propelled to the upper part of the esophagus and pharynx, where it causes the walls to vibrate, thereby producing sound. This sound is articulated in the mouth to generate intelligible speech. The challenges associated with this technique of voice production are as follows: (1) long periods of training for patients to master; (2) difficulty in learning the method; and (3) the maximum phonation time is relatively short, and it is interrupted by the process of gathering more air for the upcoming sound. However, this method does not require the use of hands or battery-powered devices, and with adequate training, patients can produce short bursts of intelligible speech.[7]

To solve the aforementioned problems, it is useful to place the sound source inside the pharynx rather than on the neck to improve the ease of device operation. Talking modulator is an effector used in the song "Sweet Emotion" by the rock band "Aerosmith." The device changes the sounds made by an electric guitar by guiding them through a tube to the mouth. Inspired by the talking modulator, we developed a novel intraoral voice assistance device called the voice retriever (VR), in which the sound source is placed on the palate.

We hypothesized that the performance of VR would be similar to that of EL and that the VR would considerably improve the voice-related quality of life (V-RQOL)[8] of patients with dysphonia. Herein, we describe our newly developed VR. The aim of this study was to verify the usefulness of VR by (1) comparing speech intelligibility and naturalness of normal participants using a VR and EL for the first time and (2) comparing the V-RQOL of patients with dysphonia before and after VR use.

2 | MATERIALS AND METHODS

2.1 | Experimental design

This study consisted of two trials: a comparative analysis of the functionality of the EL and VR in healthy adults (Trial 1) and a comparative analysis of the V-RQOL before and after VR use among patients with dysphonia (Trial 2). Written informed consent was obtained from all participants. This study was approved by the Ethics Review Committee of the School of Dentistry at Tokyo Medical and Dental University (Ethics approval number: D2020-071; UMIN study ID: UMIN000045151).

FIGURE 1 Structure of the voice retriever (VR).

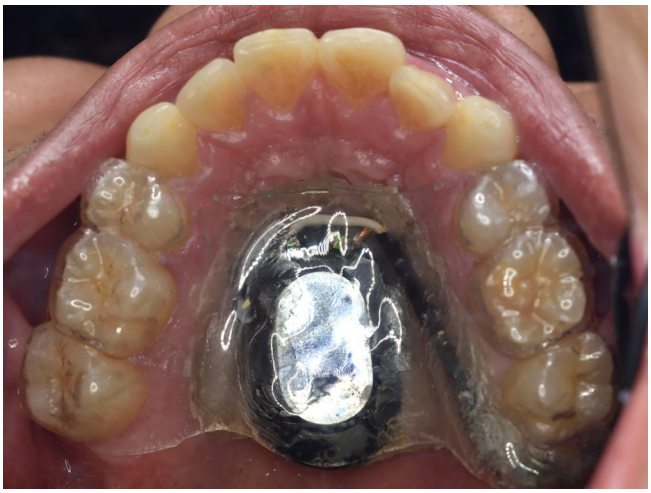

FIGURE 2 Wiring runs through the oral vestibule from the most distal posterior molar to the corner of the mouth.

2.2 | Voice retriever

The VR device consists of a mouthpiece with built-in speakers and dedicated application that serves as a sound source. Before obtaining an impression of the maxilla, an initial intraoral examination of the participants was conducted by a dentist to check for the presence of upset teeth and a strong strangulation reflex.

To create the VR, a speaker ($23 \times 16 \times 4.6$ mm^3) was placed at the center of the palate on a plaster model (Figure 1), with wiring embedded in Ortho-fast (GC, Tokyo, Japan), which is a cold-curing resin, to facilitate its passage through the oral vestibule from the most distal posterior molar to the corner of mouth (Figure 2). A hard capture sheet (Shofu, Tokyo, Japan) of diameter 1 mm was produced by suction shaping using an Erkoform 3D Plus (ERKODENT, Pfalzgrafenweiler, Germany) device.

The original sound output from the mouthpiece speaker was generated using a software synthesizer called "Laryngeal Original Sound

Generation Module,"[9] which was developed for a dedicated VR application (Crimson Technology Inc., Tokyo, Japan). The software processes signals to generate a waveform simulating original laryngeal sounds derived from vocal fold vibrations. The software synthesizer processes the signal to generate a waveform that modulates original laryngeal sounds from vocal cord vibrations. Finally, the participants tap an smart phone screen, which causes sounds to emanate from the speaker on the mouthpiece, enabling vocalization.

2.3 | Trial 1

2.3.1 | Participants

The trial participants included healthy volunteers of both sexes, aged ≥20 years and lacked experience with EL use. The participants were recruited from June 2022 to January 2023. The exclusion criteria were as follows: difficulty obtaining mouthpiece impressions because of strong tooth movement, mouth opening disorder or strangulation reflex, and previous use of VR or EL.

Regarding sample size, in a previous study,[10] the mean ± standard deviation (SD) of intelligibility of tracheoesophageal speech (with tracheoesophageal puncture incorporating an inserted silicone valved prosthesis)[11] and esophageal speech was 92.27 ± 6.12 and 96.58 ± 0.05, respectively. Assuming that the effects of the intervention in this study would be similar to those previously reported, the sample size was calculated as follows: $28 \times 2 = 56$ participants, with an effect size of 0.72 for between-group comparisons via a paired *t*-test using G power (Version 3.1.9.6; Kiel University, Kiel, Germany). As a result, 33 participants were recruited after adjusting for a dropout rate of 20%.

2.3.2 | Experimental methodology

The participants were instructed on how to use the EL (YOURTONE; DENSEI COMTEC Inc., Hokkaido, Japan) and VR. They were asked to hold their breath while speaking to (1) prevent phonation and air flow unachievable by laryngectomees and (2) approximate the acoustic effects caused by anatomical changes after laryngectomy.[12] Subsequently, they performed a 100-syllable test[13] in Japanese and long reading test of "The North Wind and the Sun"[14] using each device for subjective evaluation. Measurements were taken in a soundproof room. Before these tests, the participants were asked to read "The North Wind and the Sun" written in Japanese once using each device for practice. The first 15 participants used the EL first, whereas the other 15 used the VR first. The readings of "The North Wind and Sun" and the 100-syllable test were recorded with a Zoom H1n voice recorder (ZOOM CORPORATION, Tokyo, Japan). The recorder was placed at face level. The recordings were evaluated by five raters who were uninformed to the device used during the specific reading. The evaluators were native Japanese-speaking undergraduate students who were not experts in listening comprehension tests.

When conducting the 100-syllable Japanese speech intelligibility test, the participants were asked to randomly read 100 Japanese monosyllables while being recorded. Five raters listened to the recordings and evaluated speech intelligibility as the percentage of syllables correctly heard out of the 100 monosyllables. The average score of the five raters was analyzed.

Subjective evaluation of a long passage called "The North Wind and the Sun," was also performed. Five raters listened to the recordings and rated them according to the visual analog scale[15] based on tonicity, fluency, voice onset, amount of additional noise intonation tempo intelligibility and overall impression.[16]

2.4 | Trial 2

2.4.1 | Participants

The participants were recruited from July 2021 to November 2022. The participants included those who were unable to speak, including those who had undergone total laryngectomy and those without clear impairment of oral organs such as the tongue and lips. The exclusion criteria were as follows: difficulty following instructions due to severe dementia, significantly agitated teeth, difficulty taking oral impressions because of a mouth opening disorder, a strong strangulation reflex, and difficulty wearing a mouthpiece.

Regarding the sample size, a previous study[3] found that the mean (SD) V-RQOL values obtained using two different vocalization methods were 76.5 ± 15.9 and 53.5 ± 24.3, respectively. Assuming similar pre- and post-intervention effects in this study and a correlation coefficient of 0.6 between pre- and post-intervention V-RQOL, the sample size was calculated as $n = 8$ based on a corresponding *t*-test.

2.4.2 | Experimental methodology

The participants were fitted with a VR by a dentist at an outpatient clinic of the Department of Dysphagia Rehabilitation, Tokyo Medical and Dental University Hospital or at the participant's home. Specific procedures included maxilla impressions as well as V-RQOL evaluation before and 1–3 months after VR fitting.

2.4.3 | Measures

V-RQOL is used to assess the effect of dysphonia on a participant's QOL. A 5-point Likert scale was used to answer 10 items classified into two domains as follows: social–emotional (items 4, 5, 8, and 10) and physical functioning (items 1, 2, 3, 6, 7, and 9). The total V-RQOL score ranges from 0 to 100, with a higher score indicating a good V-RQOL.

2.5 | Statistics

In Trial 1, a paired *t*-test was used to analyze the measured numerical values related to EL and VR use. In Trial 2, the V-RQOL values before and after VR use were compared using a paired Wilcoxon test. Statistical significance was set at $p < 0.05$. Given the small number of participants in this study and difficulty in confirming the normality of the distribution, a sensitivity analysis was performed using Wilcoxon test. The statistical analyses were performed using IBM SPSS for Windows, Version 28.0 (IBM Japan, Tokyo, Japan).

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and the tenets of the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Written informed consent was obtained from all participants or their legal representatives.

3 | RESULTS

3.1 | Trial 1

We included 33 participants (12 male and 21 female participants); among them, three dropped out because of a strangulation reflex. The mean age of the included participants was 31.56 ± 6.84 (age range: 23–49) years. Between-group comparisons of the 100-syllable and subjective evaluation of long reading tests differed based on speech, comprehension, noise level, and overall impression of VR use (Table 1).

3.2 | Trial 2

Trial 2 included eight participants (7 male participants and 1 female participant), with a mean age of 77.1 ± 5.15 years (Table 2). All participants used VR until the second trial, and no one dropped out. Among the eight participants, six independently performed ADL, whereas two were on ventilators and had difficulty walking independently; among them, one required total assistance (Table 2). The study findings indicated that V-RQOL after more than 1 month of VR use was significantly better than that before VR use ($p = .028$).

TABLE 1 Comparison of voice retriever (VR) and electrolarynx (EL) use in healthy adults.

	VR	EL	*p*-value
100-syllables test	14.6(3.28)	12.8(2.93)	<.001*
Subjective evaluation			
Tonicity	2.58(0.97)	2.43(1.02)	.427
Fluency	5.26(0.98	4.60(1.34)	.03*
Voice onset	4.23(1.23)	3.17(1.45)	0.003*
Amount of additional noise	4.34(1.22)	2.96(1.18)	<0.001*
Intonation	3.23(1.11)	2.26(1.04)	<0.001*
Tempo	6.42(1.11)	5.54(1.94)	0.055
Intelligibility	5.01(1.19)	4.04(1.68)	0.006*
Overall Impression	5.20(1.00)	4.16(1.45)	0.001*

Note: Mean (SD).
Abbreviations: EL, electrolarynx; VR, voice retriever.

4 | DISCUSSION

This study demonstrated that the VR allows better speech production than an EL in healthy participants who were using an alternative vocalization method for the first time. Moreover, it showed that VR use significantly improves the V-RQOL of patients with dysphonia.

Several anatomical requirements are required to achieve natural articulation while placing the sound source in the oral cavity. First, the sound has to be radiated toward the posterior part of the mouth to produce a voiceless velar stop. Additionally, as the anterior teeth and palatal folds are important sites for the articulation of the "t" and "s" sounds, the mouthpiece was designed to be clipped on, covering only the molars. Accordingly, because the retention of the device is dependent on teeth, the oral condition affects adaptation conditions. Our participants had sufficient occlusal support for the molars. For individuals with many missing teeth, it will be necessary to fabricate dentures before VR fitting. Therefore, it is desirable for dentists to make and adjust this device.

Here, among healthy individuals who had never used a VR or EL, factors that led to significant between-device differences in the 100-syllable test included the simplicity of VR operation and the fact that the sound source of the VR was located inside the body. While an EL vibrates air in the pharynx from outside the body, the VR vibrates the air directly in the oral cavity, resulting in less noise. Furthermore, subjective evaluation showed significantly higher values for the VR except with regard to tonicity and tempo, suggesting that the low noise involved in VR contributed to the results.

Our findings demonstrated that V-RQOL was significantly improved by the use of the VR. The dysphonia in the eight patients in Trial 2 was caused by total laryngectomy due to laryngeal cancer (six patients) and tracheostomy and aspiration prevention surgery due to neurodegenerative diseases such as ALS (two patients). None of the patients had established effective substitute speech before VR use. Notably, the patient with ALS had difficulty pressing the EL against their neck because of upper extremity muscle weakness. Patients with postoperative laryngeal cancer had not acquired alternative vocalization because of difficulty in using EL due to radiotherapy[4] or discomfort with the high noise level of the EL.[17]

An advantage of VR over EL is its ease of learning; VR does not require the movement of holding the EL in the hand and pressing it against the neck (Figure 3), which is a hinderance in learning EL, and the sound can be produced in the oral cavity placing the mouthpiece. Therefore, VR use is easier to learn than EL use, especially for beginners, and is considered more stable for vocalization.

TABLE 2 Participant characteristics and V-RQOL.

						V-RQOL	
ID	Age	Sex		Note	Acquisition status of alternative vocalization	Before	After
1	73	M	Hypopharyngeal cancer		Because of dissatisfaction with the sound quality of EL	0	47.5
2	84	F	Hypopharyngeal cancer		Due to failure to use EL well.	7.5	17.5
3	82	M	Hypopharyngeal cancer		Hadn't tried anything yet.	12.5	97.5
4	82	M	Hypopharyngeal cancer		Due to failure to use EL well.	15	75
5	68	M	Invasive ventilation due to ALS	Full assistance	Unable to use EL due to muscle weakness	75	72.5
6	78	M	Hypopharyngeal cancer		Due to failure to use EL well.	35	72.5
7	73	M	Hypopharyngeal cancer		Due to failure to use EL well.	32.5	32.5
8	77	M	Invasive ventilation due to ALS		Unable to use EL due to muscle weakness	30	70

Note: *Mean (SD).
Abbreviations: ALS, amyotrophic lateral sclerosis; EL, electrolarynx; V-RQOL, voice related QOL.

FIGURE 3 How to use EL for healthy people.

A disadvantage of VR is that the thickness of the mouthpiece inhibits tongue movement; some people might drop out because of the strangulation reflex. Moreover, the saliva drips from the corners of the mouth because the mouthpiece is currently wired.

This study has several limitations. First, the outcomes assessed only included the V-RQOL, which is a subjective assessment completed by the patients, with objective items assessed. Furthermore, the participants with dysphonia do not have a single primary disease, with laryngeal cancer and ALS being the primary causes. Therefore, larger, disease-specific intervention studies are required to validate the utility of VR use. Despite these limitations, this study provides useful findings for both clinicians and patients with dysphonia. For individuals without experience in using any alternative vocalization methods, the VR can facilitate the production of better speech than EL. Additionally, as it is simpler than alternative articulation methods, the VR can be effectively used, even by patients with poor manual dexterity. Therefore, the VR may be a new option for patients requiring an alternative to existing vocalization methods such as the EL.

5 | CONCLUSIONS

We developed a novel intraoral speech aid that facilitated better speech production than EL in healthy participants without experience. Furthermore, the VR significantly improved the V-RQOL of patients with dysphonia. In the future, long-term intervention studies should be conducted among patients with the same disease to further investigate whether the VR differs from alternative vocalization methods among patients with specific diseases and to identify patient groups in which VR is highly effective.

ACKNOWLEDGMENTS

The authors wish to acknowledge Tsuda Akira for developing the application that facilitates the use of the voice retriever. The authors also thank all staff members and individuals involved in the study.

CONFLICT OF INTEREST STATEMENT

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

ORCID

Taishi Yamada https://orcid.org/0009-0001-4277-5053
Kanako Yoshimi https://orcid.org/0000-0002-5005-036X

REFERENCES

1. Abdel Tawab HM, AbdElMessih MW, AlNaggar NAA, El Sharkawy LS. Study of the epidemiology and management of laryngeal cancer in Kasr Al-Aini Hospital. *Egypt J Otolaryngol*. 2014;30(3):208-214.
2. Ogino M. End of life care for patients with ALS in Japan. *Rinsho Shinkeigaku*. 2008;48:973-975.
3. Moukarbel RV, Doyle PC, Yoo JH, Franklin JH, Day AM, Fung K. Voice-related quality of life (V-RQOL) outcomes in laryngectomees. *Head Neck*. 2011;33(1):31-36.
4. Liu H, Ng ML. Electrolarynx in voice rehabilitation. *Auris Nasus Larynx*. 2007;34(3):327-332.

5. Kaye R, Tang CG, Sinclair CF. The electrolarynx: voice restoration after total laryngectomy. *Med Devices (Auckl)*. 2017;10:133-140.
6. Gates GA, Hearne EM III. Predicting esophageal speech. *Ann Otol Rhinol Laryngol*. 1982;91(4 Pt 1):454-457.
7. Chen HC, Kim Evans KF, Salgado CJ, Mardini S. Methods of voice reconstruction. *Semin Plast Surg*. 2010;24(2):227-232.
8. Aaby C, Heimdal JH. The voice-related quality of life (V-RQOL) measure—a study on validity and reliability of the Norwegian version. *J Voice*. 2013;27(2): 258.e29-258.e33.
9. Morise M, Yokomori F, Ozawa K. WORLD: a vocoder-based high-quality speech synthesis system for real-time applications. *IEICE Trans Inf Syst*. 2016;E99.D(7):1877-1884.
10. Eadie TL, Otero D, Cox S, et al. The relationship between communicative participation and postlaryngectomy speech outcomes. *Head Neck*. 2016;38(suppl 1):E1955-E1961.
11. Blom ED. Tracheoesophageal speech. *Tracheoesophageal Speech (Semin. Speech Lang.)*. Thieme Medical Publishers, Inc; 1995: 191-204.
12. Guo L, Nagle K, Heaton JT. Generating tonal distinctions in mandarin Chinese using an electrolarynx with preprogrammed tone patterns. *Speech Commun*. 2016;78:34-41.
13. Watarai A, Yasunaga Y, Nakao J, et al. Groin and anterolateral thigh flaps for hemiglossectomy reconstruction: a comparison based on Japanese speech intelligibility. *Auris Nasus Larynx*. 2023;50(1): 110-118.
14. Kawanishi M, Kajikawa H, Yamamura K, et al. Aphasia following left putaminal hemorrhage. Statistical analysis of factors affecting prognosis. *Neurol Res*. 2002;24(8):817-821.
15. Sung YT, Wu JS. The visual analogue scale for rating, ranking and paired-comparison (VAS-RRP): a new technique for psychological measurement. *Behav Res Methods*. 2018;50:1694-1715.
16. Moerman M, Pieters G, Martens JP, van der Borgt MJ, Dejonckere P. Objective evaluation of the quality of substitution voices. *Eur Arch Otorhinolaryngol*. 2004;261(10):541-547.
17. Cox SR, Doyle PC. The influence of electrolarynx use on postlaryngectomy voice-related quality of life. *Otolaryngol Head Neck Surg*. 2014;150(6):1005-1009.

How to cite this article: Yamada T, Yamaguchi K, Horike A, et al. Development and evaluation of a new intraoral voice assist device called the voice retriever. *Laryngoscope Investigative Otolaryngology*. 2024;9(1):e1204. doi:10.1002/lio2.1204