Automatic estimation of laryngeal vestibular closure duration using high resolution cervical auscultation (HRCA) signals

Aliaa Sabry, MD, PhD[1,2], Amanda S. Mahoney, MA, SLP[1], Shitong Mao, MS[3], Yassin Khalifa, MS[3], Ervin Sejdic, PhD[3,4], James L. Coyle, PhD[1]

[1] Department of Communication Science and Disorders, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA 15260 USA

[2] Krembil Research Institute, University Health Network (UHN), Toronto, ON M5T 0S8, Canada (current affiliation)

[3] Department of Electrical and Computer Engineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA 15260 USA

[4] Department of Bioengineering, Swanson School of Engineering Department of Biomedical Informatics, School of Medicine Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, PA 15260 USA.

Corresponding author: Aliaa Sabry

Address: 160- 500 University Avenue, Toronto, ON M5G 1V7

Tel.: 647-836-7430

E-mail: Aliaa.Elbahnasy@uhnresearch.ca

Key words: dysphagia, laryngeal vestibular closure duration, cervical auscultation, machine learning.

Conflict of Interest: no relevant conflicts of interest.

Funding: Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institute of Health under Award no. R01HD092239 (E. Sejdic, PI), while the data were collected under Award no. R01HD074819 (E. Sejdic, J. Coyle, co-PI's). The work was also supported by the NSF Career Award #1652203 (E. Sejdic, PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health or the National Science Foundation.

Abstract

Purpose: Safe swallowing requires adequate protection of the airway to prevent swallowed materials from entering the trachea or lungs (i.e., aspiration). Laryngeal vestibular closure (LVC) is the first line of defense against swallowed material entering the airway. Absent LVC or mistimed/shortened closure duration can lead to aspiration, adverse medical consequences, and even death. Laryngeal vestibular closure mechanisms can be judged commonly through the videofluoroscopic swallowing (VFS) study, however, this type of instrumentation exposes patients to radiation and is not available or acceptable to all patients. There is growing interest in noninvasive methods to assess/monitor swallow physiology. In this study, we hypothesized that our non-invasive sensor-based system, which has been shown to accurately track hyoid displacement and upper esophageal sphincter opening duration during swallowing, could predict laryngeal vestibular status, including the onset of LVC and laryngeal vestibular re-opening (LVO) in real time and estimate the closure duration with a comparable degree of accuracy as trained human raters. **Methods:** The sensor-based system used in this study is high-resolution cervical auscultation (HRCA). Advanced machine learning techniques enable HRCA signal analysis through feature extraction and complex algorithms. A deep learning model was developed with a dataset of 588 swallows from 120 patients with suspected dysphagia and further tested on 45 swallows from 16 healthy participants. **Results:**

The new technique achieved an overall mean accuracy of 74.90% and 75.48%, for the two data sets respectively, in distinguishing LVC status. Closure duration ratios between automated and gold-standard human judgment of LVC duration were 1.13 for the patient data set and 0.93 for the healthy participant data set. **Conclusion:** This study found that HRCA signal analysis using advanced machine learning techniques can effectively predict LV status (closure or opening) and further estimate LVC duration. HRCA is potentially a non-invasive tool to estimate LVC duration for diagnostic and biofeedback purposes without x-ray imaging.

Introduction

Swallowing is a complex neuromuscular process involving the integration of two distinct but related functions: airway protection and bolus transport. This complex process involves volitional and reflexive neural activities paired with coordinated contraction of many paired muscle groups. The result of this process is specific biomechanical events, which are executed in a sequential temporal order to ensure safe and efficient swallowing. Although there is variability within and among humans, any disturbance of these biomechanical events caused by disease can lead to swallowing disorders, known as dysphagia.

Entrance of food or liquid into the airway during the pharyngeal stage of swallowing is known as aspiration. Aspiration is generally considered the most concerning component of swallowing dysfunction and may lead to possibly fatal pulmonary consequences, especially for individuals with neurologic and neurodegenerative diseases (Cabib et al., 2016) or already-compromised respiratory systems. Laryngeal vestibular closure (LVC) is usually considered the primary and most critical aspect of laryngeal function during swallowing, providing protection for the airway against the entrance of swallowed materials. LVC is defined as the collapse of the laryngeal inlet via arytenoid adduction, and arytenoid approximation to the epiglottis during epiglottic inversion (Logemann et al., 1992). The closure of the laryngeal airway occurs in a peristaltic-like motion, by a caudal to rostral compression while the larynx shortens facilitating approximation of the epiglottis to the laryngeal inlet. This pattern of closure, which is observable through videofluoroscopic studies (VFS) of swallowing function, prevents airway invasion by closing off the airway while squeezing aberrant swallowed material out of the laryngeal vestibule (LV) (Ekberg, 1982; Ekberg & Nylander, 1982).

Timely and complete LVC is vital to safe and successful swallowing. Incomplete closure, or shortened LVC duration may cause laryngeal penetration, in which swallowed material enters the LV remains above the level of the vocal folds, and/or tracheal aspiration of swallowed materials (Mann et al., 1999; Robbins et al., 1993,). Shortened LVC duration is significantly associated with an increased incidence of aspiration (Cabib et al., 2016). In fact, shortened LVC duration is the primary impairment for predicting aspiration in patients following stroke (Power et al., 2007).

The published literature reports a wide range of LVC durations, with mean values from 0.31 to 1.07, depending on the presence or absence of certain factors (Humbert et al., 2018; Logemann et al., 1992; Logemann et al., 2000; Logemann et al., 2002; Molfenter & Steele, 2012; Ohmae et al., 1995; Ohmae et al., 1996; Park et al., 2010). Prolonged LVC duration has been observed with increasing bolus volumes, longer pharyngeal transit durations (Kang et al., 2010; Kendall et al., 2003; Kim et al., 2005; Kim et al., 2010; Martin-Harris et al., 2003; Rofes et al., 2010; Rosenbek et al., 1996), and during the performance of swallow maneuvers such as the effortful swallow and the chin down posture (Hind et al., 2001; Macrae et al., 2014; Young et al., 2015). Intentionally increasing LVC duration during swallowing in patients with shortened LVC duration has been investigated as a method of improving airway protection for decades. The supraglottic swallow maneuver, described in 1993, was designed to volitionally close the upper airway before swallowing in patients with a supraglottic laryngectomy whose epiglottis had been resected (Mendelsohn & Martin, 1993). This maneuver, and its sibling the super-supraglottic swallow, which exaggerates contact between the arytenoids and epiglottic base in non-resected patients, has been adapted for use in patients with dysphagia whose laryngeal anatomy remains intact, and are mainstays of dysphagia compensatory management for many patients (Lazarus et al., 1993). Many literatures demonstrated that healthy individuals and individuals with dysphagia due to stroke could volitionally prolong LVC after training (Azola et al., 2015; Lazarus et al., 1993; Macrae et al., 2014; Mendelsohn & Martin, 1993; Young et al., 2015). Direct volitional control of the timing and duration of LVC has enormous rehabilitation potential for individuals with dysphagia.

VFS, a real-time dynamic x-ray technique, is the only standard instrumental assessment to visualize LVC and to determine LVC duration during swallowing (Martin-Harris & Jones, 2008). The duration of LVC is the measure of how long the LV remains completely closed. In VFS images, complete LVC is defined as no visible air space or barium contrast in the LV given complete contact of the arytenoids to the base of the epiglottis and full epiglottic inversion over the base of the arytenoids (Logemann et al., 1992). VFS can be used to train volitional prolongation of LVC by providing patients with kinematic visual biofeedback. However, VFS has inherent challenges such as patients’ exposure to radiation. Radiation safety standards limit exposure time during VFS, thus data collection opportunities are time sensitive and despite its superior visualization of the entire aerodigestive mechanism during swallowing, the use of VFS for visual biofeedback during treatment to acquire compensatory volitional augmentation of

LVC is impossible. VFS may not be feasible in facilities without x-ray departments and facilities may not have qualified clinicians to perform and interpret the VFS images. Additionally, some patients may refuse x-ray testing or have other conditions limiting its accessibility or feasibility (Bonilha et al., 2013; Nierengarten, 2009; Steele et al., 2007; Zammit-Maempel et al., 2007).

Although acquiring temporal measurements of LVC duration would be invaluable when managing many patients with dysphagia, it is rarely quantified during imaging studies of swallowing function. During VFS studies, LVC is typically judged as present, absent, or incomplete but temporal measurements are not assessed.

There are limitations in a typical clinical setting that prevent frequent temporal measurement of LVC, which result in these broad categorical judgements. Swallow kinematic analysis using frame-by-frame review of VFS images is not typically performed by clinicians because very few have the required training or confirmation of their judgment reliability. Some clinicians may not have the ability to record VFS images for secondary review due to lack of equipment or limited access to archived materials. Additionally, a minimum temporal resolution of 30 frames per second is required to properly assess LVC duration. Recording at reduced frame rates (i.e., 7.5 or 15 frames per second), a common practice, is inadequate for accurately capturing LVC timing due to its short duration (Bonilha et al., 2013).

Adding temporal measures to the evaluation of LVC could provide clinicians with objective swallowing kinematic data, which could be compared to published, normative data, and provide clinical evidence of increased risk of airway compromise (Humbert et al., 2018; Molfenter & Steele, 2012). Successfully achieving this goal would help initiate appropriate compensatory interventions to reduce dysphagia complications through timely diagnosis. The benefits of having objective LVC data and the limitations of using VFS indicates that clinicians would benefit from a non-invasive, alternative method to estimate LVC duration. Naturally the ability to obtain LVC information noninvasively would revolutionize efforts to stabilize or improve LVC timing and duration in people with dysphagia.

One potential non-invasive alternative for quantifying LV temporal measures is high-resolution cervical auscultation (HRCA). Traditional cervical auscultation (CA) is a method by which a clinician uses a stethoscope on a patient's throat to assess swallowing and airway sounds. The cardiac analogy hypothesis suggests that cervical auscultation acoustic signals are generated via vibrations caused by valve and pump systems within the upper aerodigestive tract. As with heart valves that open and close during the cardiac cycle, valves in the upper aerodigestive tract produce characteristic acoustic signals during different stages of swallowing (Cichero & Murdoch, 1998). However, the transmission of swallow information may be incomplete due to the limited receiving bandwidth of a stethoscope, and the interpretation of these sounds by judges listening through a stethoscope can be bounded by the limits of the hearing frequency range of humans. Likewise, numerous well-designed studies have confirmed the very low inter-judge agreement for CA sounds rendering it a relatively weak diagnostic method (Leslie et al., 2004). Therefore, CA cannot be considered a valid and reliable screening or assessment tool for swallowing function due to imprecise and incomplete interpretation of these signals (Sejdic et al., 2018).

HRCA exhibits unbiased and reliable interpretations as compared to conventional CA assessment. HRCA uses high resolution accelerometers and microphones, attached to patients' necks, to record vibratory and acoustic signals during swallowing (Dudik et al., 2015; Movahedi et al., 2016). In line with the cardiac analogy hypothesis, the striking of the epiglottis and arytenoids may be the valve activity that generates swallowing sounds and vibrations during LVC, which can be recorded with HRCA.

HRCA is an easily mobile, non-invasive tool, which is suitable for daily monitoring of swallow function. Advanced technology using artificial intelligence through machine learning techniques enables HRCA signal analysis by using feature extraction and complex algorithms. HRCA has recently shown promise in the autonomous detection of many swallow kinematic events. HRCA signals have been found to be associated with hyoid bone displacement (He et al., 2019), LVC, and the contact of the base of the tongue with the posterior pharyngeal wall (Kurosu et al., 2019). Furthermore, HRCA successfully detected vertical and horizontal displacements of the hyoid bone (Rebrion et al., 2018) and the diameter of upper esophageal sphincter maximal opening (Shu, 2019). Given recent advances in signal processing algorithms, HRCA could provide a fundamental contribution to dysphagia management.

In this study we investigated the ability of advanced machine learning techniques to predict LVC and LVO through HRCA signal analysis, thus allowing a predicted estimation of LVC duration. We hypothesized that by analyzing HRCA signals using machine learning techniques, we could predict LVC and LVO status in real time and estimate the duration of LVC with a comparable degree of accuracy as trained human raters. Successfully achieving this aim would significantly improve LVC duration estimation by making it more automatic and objective.

Methods

Data collection and equipment

Two sets of data were collected; the first dataset was composed of 588 swallows from 120 enrolled patients with various diagnoses and etiologies of dysphagia, the second was composed of 45 swallows from 16 healthy community dwellers. Patient and healthy participant characteristics can be found in Table 1.

All patients and healthy participants underwent VFS at University of Pittsburgh Medical Center Presbyterian Hospital. Since the aim of this study was to investigate the feasibility of our system's ability to predict LVC regardless of other variables, we intentionally did not control for patient variables including the patient's diagnosis or characteristics of swallowed materials. Data for patients was collected during routine clinical VFS studies, which resulted in various volumes and consistencies of swallowed material. Healthy participants swallowed only thin liquids of various volumes. All patient and healthy participants in this study signed informed consents and the data collection protocol was approved by the Institutional Review Board of the University of Pittsburgh.

Please insert Table 1 here.

VFSs for patients were conducted in the lateral plane using an x-ray machine (Ultimax system, Toshiba, Tustin, CA) with a pulse rate of 30 fps. Healthy participant data was collected in the lateral plane with a Precision 500D x-ray system (GE Healthcare, LLC, Waukesha, WI) with a pulse rate of 30 fps. To ensure that different resolutions did not affect judgment of kinematic events, we resampled a subset of the original VFS data to match the sample rate of the new machine. Five judges labeled nine swallowing kinematic events, including LVC and LVO, using native and resampled resolutions. The level of agreement between human labels at the different resolutions was excellent for all measures, with inter-judge ICCs at or above .99. VFS videos were captured on an AccuStream Express HD video card (Foresight Imaging, Chelmsford, MA) and digitized with a sampling rate of 60 frames per second then saved to a hard disk using LabView's Signal Express (National Instruments, Austin, Texas).

The sensor signals were collected concurrent to VFS examinations using a tri-axial accelerometer neck sensor and contact microphone. The accelerometer (ADXL 327, Analog Devices, Norwood, Massachusetts) was attached to the midline of participant's anterior neck at the level of the cricoid cartilage with surgical tape to obtain the best contact (Takahashi et al., 1994). The sensors' axes were aligned to the anatomical directions of anterior-posterior [AP], superior-inferior [SI], and medial-lateral [ML] respectively. The sensor was powered by a power supply (model 1504, BK Precision, Yorba Linda, California) with a 3V output, and the resulting signals were bandpass filtered from 0.1 to 3000 Hz and amplified ten folds (model P55, Grass Technologies, Warwick, Rhode Island). The microphone (model C411L, AKG, Vienna, Austria), which was powered by a power supply (model B29l, AKG, Vienna, Austria), was placed below the accelerometer and slightly towards the right lateral side of the trachea. This location has previously been described to be appropriate for collecting swallowing sound signals without interfering with visualization of the proximal trachea or larynx (Cichero & Murdoch, 2002; Takahashi et al., 1994). All signals acquired by the accelerometer and microphone were fed into a National Instruments 6210 DAQ and recorded at 20 kHz by the LabView program (Signal Express, National Instruments, Austin, Texas). This setup has been shown to be effective at detecting swallowing activity in previous studies (Dudik et al., 2016; Lee et al., 2010).

Data labeling

All videos were segmented into individual swallows. Swallow durations were defined as the frame in which the head of the bolus reached the ramus of the mandible (onset) to the frame in which the hyoid returned to its lowest position following clearance of the bolus from the pharynx (offset). The corresponding HRCA signals were also segmented according to the frames of onset and offset. Reliability of segmentation was established on 10% of the videos with ICCs of over .99 and intra-rater reliability and was maintained throughout testing to avoid judgment drift.

Two trained raters labeled the first closure and first re-opening of the LV from VFS x-ray videos for each swallow sample (Fig.3). Reliability was established on 10% of the videos with ICCs of over .99 and intra-rater reliability was maintained throughout testing to avoid judgment drift. The criteria in judging the LV status are listed in Table 2.

Please insert Table 2 here.

Once the onset values for LVC and LVO were recorded by judges, the data was entered into machine learning routines to enable training and testing of the accuracy of the algorithms.

Deep neural network architecture, training, and testing

An advanced hybrid deep neural network combining a Convolutional Neural Network and Recurrent Neural Network, called a Convolutional Recurrent Neural Network (CRNN), was used to build the relationship between the HRCA signals and the LVC duration by predicting the LVC and LVO statuses. Artificial Neural Networks are loosely based on the neuronal networks in humans. They are typically organized in "layers" and contain "learning rules", which allow the network to recognize underlying patterns between input and output. The network is repeatedly trained based on observed datasets until it recognizes the patterns, and then the model is tested on a novel or "unseen" dataset to evaluate the model fit, or how well the network has "learned".

In this study, the two LV statuses (opened and closed) were coded as '0' and '1' respectively. The human-labeled LV statuses were translated to the computer program through this binary sequence (Fig. 1). The CRNN model was given the binary sequence for each swallow frame series (i.e. the first frame through the last frame of the swallow), with the corresponding HRCA signal segments. The CRNN was trained to mathematically model the relationship between the HRCA signals and the LV statuses.

Please insert Figure 1 here.

A 10-fold cross validation technique was used to develop the CRNN model. In 10-fold cross validation, all samples are divided into 10 non-overlapping training groups. During training, nine of the ten groups are used to “train” the model by providing feedback to help the model predict the human labels using signals only. The remaining sample is used as a validation set to evaluate, or essentially help the model find parameters (i.e. other factors), which may not have been identified during training with the initial nine groups. This process is repeated a total of 10 times with each sample used as a validation set once.

For this study, the 588 patient swallowing samples were randomly divided into 10 patient-specific training groups. In other words, an individual patient’s swallows were contained within one group and not spread across any of the remaining nine groups. The groups were used for training and validating the CRNN to predict LVC and LVO based on HRCA signals alone. Once the 10-fold validation was completed, the “unseen” dataset of 45 healthy participant swallows was used as a testing set to evaluate the final model fit (i.e. to determine how well the model could predict LVC and LVO using HRCA signals without having ever “seen” the data) to evaluate how well the model generalized to new information.

Results

The following results reveal the accuracy of the CRNN model. We use the term "accuracy" to characterize the percentage of the frames that were correctly predicted, as compared to the human labels. First, the accuracy of the model to predict the frame number of the onset of the LVC (within +/- 3 frames of the human label (Lof & Robbins, 1990)) for the patient dataset was 62.07% (mean error value = 0.19 +/- 4.5 frames) and the frame number for the onset of LVO was 60.03% (mean error value = 0.08 +/- 4.9 frames). For the healthy participant dataset, whose data were not included in the training process, the accuracy of model prediction for the frame number for the onset of LVC (within +/- 3 frames of the human label) was 66.22% (mean error value = 0.73 +/- 5.2 frames) and the frame number for the onset of LVO was 64.44% (mean error value = 0.73 +/- 5.2 frames). Figure 2 illustrates the frame error distribution for the validation sets and the testing set.

Please insert Figure 2 here.

Mean overall accuracy is the ratio of the number of frames that were correctly predicted by the algorithm (whether the LV was opened or closed) over the total number of frames for all swallows. The model's mean overall accuracy for predicting LV status (whether the LV was opened or closed) across the 10 groups from the training set of patient swallows was 74.90%. The accuracy levels of the 10 validation groups for the LV status prediction are shown in Fig. 3. The mean overall accuracy for distinguishing LV status (opening and closure) from the testing dataset of 45 healthy participant swallows was 75.48%.

Finally, to evaluate the model's predictive ability for LVC duration, we used a duration ratio. The duration ratio was calculated as the predicted number of frames for which the LV is closed over the human labeled LVC frames for which the LV is closed. The closer the ratio is to 1, the closer the model's prediction was to the human calculated duration. The duration ratio for the 10 patient validation groups is listed in Table 3. The overall mean value for the duration ratio from the patient dataset was 1.13, indicating that the model slightly overestimated the number of frames in which the LV was closed. The overall mean value for the duration ratio from the healthy participant dataset was 0.93, indicating that the model slightly underestimated the number of frames in which the LV was closed.

Please insert Figure 3 here.
Please insert Table 3 here.

Discussion

The primary aim of this study was to determine the feasibility of HRCA signals to predict LV status (open, closed) during swallowing with an advanced computer-aided approach, and thus non-invasively estimate the duration of LVC. We demonstrated that a highly complex and non-linear relationship between the LV status and HRCA signals can be established via advanced deep learning algorithms, such as the proposed hybrid neural network in this study.

The CRNN model autonomously predicted LV status based on HRCA signal input alone, independent from the manual analysis of the VFS videos by human judges, which were used to assess the model's performance. Our experimental results revealed that the overall accuracy of the model to distinguish the LV status (open, closed) was around 75% for both validation and testing datasets, suggesting that the CRNN algorithm is capable of distinguishing LV status (open, closed) based only on HRCA signals and, therefore, LVC duration.

The mean accuracies for machine predicted LVC and LVO frames for the testing group of healthy participants' "unseen data" were higher than the accuracies for the training and validation sets of patient "seen data", which underscores the robustness of the CRNN model. It is unclear why the participant testing data had larger mean error values than the patient data, but a possible explanation could be differences between patient vs. healthy swallow kinematics. The algorithm was trained and validated only on disordered swallows but was tested on healthy swallows. Regardless, the higher accuracies seen in the tested set support the utility of the algorithm; however, the system is not yet ready for clinical implementation. This study established feasibility and illustrated the model's relatively impressive performance in accurately identifying very short-duration events. These events were detected from among all events occurring during a swallow sequence. We intend to hone the system's precision in future investigations.

HRCA also has the potential to be used as a non-invasive biofeedback tool during swallowing rehabilitation. Dysphagia management is designed to target the underlining biomechanical impairment during swallowing, which can be achieved through behavioral modifications such as swallowing maneuvers. However, when training swallowing maneuvers, patients are expected to exert volitional control over laryngeal structures. This presents treatment challenges when imaging-based visual biofeedback is unavailable because individuals with dysphagia may not be familiar with laryngeal function. Providing the patients with extrinsic feedback could improve patient compliance, accurate performance, and overall outcomes, as has been demonstrated with other signal-based biofeedback methods (Martin-Harris et al., 2017; Steele et al., 2012).

In clinical settings, the combination of clinician's verbal feedback with visual biofeedback (i.e. kinematic feedback such as videofluoroscopy or FEES, or non-kinematic such as signal waveforms, numerical data, or graphs) corresponding to the patient's target movement can intensify the impact of extrinsic feedback (Crary & Groher, 2000; Humbert & Joel, 2012). Unlike limbs, the volitional control of the larynx is a relatively obscure act without externally observable activity upon which to base motor learning. The amplified effect of combined extrinsic feedback may augment the patient's intrinsic feedback system, which monitors the movement of muscles, joints, and general body position, thus allowing the patient to make more accurate approximations of targeted gross and fine movements (Abbruzzese et al., 2014; Gandevia et al., 2002) and, ultimately, support learning the target task (Dayan & Cohen, 2011; Taubert et al., 2011).

HRCA can provide biofeedback by estimating LVC and LVO, thereby providing LVC duration to patients. Using HRCA in this way would limit radiation exposure and could improve patient accuracy for targets related to LVC and LVO onset and volitional LC prolongation, thus promoting better airway protection.

Methods of improving skill acquisition, along with schedules for dosage and intensity, and reinforcement and feedback, are important components of rehabilitation treatment taxonomies (Hart et al., 2019). Imagine, for example, there is an HRCA visual biofeedback device, which provides the patient with a simple visual representation of laryngeal closure and opening (e.g., red (open) or green (closed) lights) as biofeedback. This type of system could provide the clinician and patient with LVC duration information as well as provide the patient with visual feedback during skill acquisition to help support them achieve their therapy goal.

HRCA provides an objective tool to noninvasively analyze laryngeal behavior during swallowing, which can provide trackable outcome measures and help demonstrate and document the efficacy of interventions to reduce aspiration risk.

The newly proposed machine learning technique using a CRNN model enabled us to analyze HRCA signals associated with specific swallowing kinematic events (LVC, LVO), and aligns with other research in our lab demonstrating the association between HRCA signals and hyoid bone displacement (He et al., 2019), LVC, the contact of the base of the tongue with the posterior pharyngeal wall (Kurosu et al., 2019), and the diameter of upper esophageal sphincter maximal opening (Shu, 2019).

This new technique has potential for further non-invasive swallowing function examination for other kinematic events such as tongue base retraction or epiglottic inversion, which could not be completely perceived or precisely analysed previously.

The aim of this study was to determine the ability of the sensors and the CRNN to independently predict the LV status regardless of age, gender, or diagnosis; however, these considerations provide interesting directions for future research. Researchers could investigate systematic changes in model predictions of LVC and LVO. Considerations for changes include varying bolus volumes and consistencies, various patient characteristics (e.g. age, gender, diagnosis), and disease characteristics (e.g. disease/dysphagia severity, infarct location from stroke, and degenerative disease progression).

Further considerations for future research include exploring factors for machine learning, such as model structure, learning algorithms, and hyperparameter tuning. These factors may improve the accuracy of the CRNN model, thus ensuring the identification of “safe” swallows and avoiding the over or under estimation of LV closure. Ideally, clinical trials should investigate the efficacy of HRCA as a non-invasive biofeedback tool to augment training in volitional laryngeal closure and to establish its use as a swallowing intervention to reduce aspiration.

Limitations

One limitation of the current study is that the model was trained on patient swallows and did not incorporate healthy swallows, which may have improved its performance. These machine learning algorithms perform more robustly when they are trained on heterogeneous exemplars (i.e., swallows) from the population under investigation. We also conducted training and testing of the model with relatively small sample sizes. Generally, larger training sample sizes are preferred in the machine learning process. A larger sample of swallows would have increased the opportunity for the model to characterize less common perturbations in swallow physiology; the accuracy in modelling the novel test data subset would most likely be improved. Our results are considered preliminary and will likely improve as we increase the sample size and train the model with healthy swallows; however, this study demonstrates the feasibility of using HRCA to predict LV status and LVC duration.

Conclusion

This study found that HRCA signal analysis using an advanced machine learning technique can effectively predict LV status (opening or closure) and accurately estimate LVC duration. This provides a potential non-invasive tool to estimate LVC duration for diagnostic and biofeedback purposes in managing patients with dysphagia as an adjunct to x ray imaging.

Acknowledgment

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R01HD092239 and National Science Foundation CAREER Award Number 1652203, while the data was collected under Award Number R01HD074819. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or National Science Foundation.

References

Abbruzzese, G., Trompetto, C., Mori, L., & Pelosin, E. (2014). Proprioceptive rehabilitation of upper limb dysfunction in movement disorders: A clinical perspective. *Frontiers in Human Neuroscience*, *8*, 1-8.

Azola, A. M., Greene, L. R., Taylor-Kamara, I., Macrae, P., Anderson, C., & Humbert, I. A. (2015). The relationship between submental surface electromyography and hyo-laryngeal kinematic measures of Mendelsohn Maneuver duration. *Journal of Speech, Language, and Hearing Research, 58*(6), 1627-1636.

Bonilha, H. S., Blair, J., Carnes, B., Huda, W., Humphries, K., McGrattan, K., Michel, Y. & Martin-Harris, B. (2013). Preliminary investigation of the effect of pulse rate on judgments of swallowing impairment and treatment recommendations. *Dysphagia, 28*(4), 528-538.

Bonilha, H. S., Humphries, K., Blair, J., Hill, E. G., McGrattan, K., Carnes, B., Huda, W. & Martin-Harris, B. (2013). Radiation exposure time during MBSS: Influence of swallowing impairment severity, medical diagnosis, clinician experience, and standardized protocol use. *Dysphagia*, *28*(1), 77-85.

Cabib, C., Ortega, O., Kumru, H., Palomeras, E., Vilardell, N., Alvarez-Berdugo, D., Muriana, D., Rofes, L., Terre, R., Mearin, F. & Clave, P. (2016). Neurorehabilitation strategies for poststroke oropharyngeal dysphagia: From compensation to the recovery of swallowing function. *Annals of the New York Academy of Sciences*, *1380*(1), 121-138.

Cichero, J. A., & Murdoch, B. E. (1998). The physiologic cause of swallowing sounds: Answers from heart sounds and vocal tract acoustics. *Dysphagia*, *13*(1), 39-52.

Cichero, J. A., & Murdoch, B. E. (2002). Detection of swallowing sounds: Methodology revisited. *Dysphagia*, *17*(1), 40-49.

Crary, M. A., & Groher, M. E. (2000). Basic concepts of surface electromyographic biofeedback in the treatment of dysphagia: A tutorial. *American Journal of Speech-Language Pathology*, *9*(2), 116-125.

Dayan, E., & Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron*, *72*(3), 443-454.

Dudik, J. M., Jestrovic, I., Luan, B., Coyle, J. L., & Sejdic, E. (2015). Characteristics of dry chin-tuck swallowing vibrations and sounds. *IEEE Transactions on Biomedical Engineering*, *62*(10), 2456-2464.

Dudik, J. M., Kurosu, A., Coyle, J. L., & Sejdic, E. (2016). A statistical analysis of cervical auscultation signals from adults with unsafe airway protection. *Journal of Neuroengineering and Rehabilitation*, *13*(1), 7.

Ekberg, O. (1982). Closure of the laryngeal vestibule during deglutition. *Acta oto-laryngologica*, *93*(1-6), 123-129.

Ekberg, O., & Nylander, G. (1982). Cineradiography of the pharyngeal stage of deglutition in 150 individuals without dysphagia. *The British Journal of Radiology*, *55*(652), 253-257.

Gandevia, S. C., Refshauge, K. M., & Collins, D. F. (2002). Proprioception: Peripheral inputs and perceptual interactions. In *Sensorimotor control of movement and posture*. Boston, MA: Springer.

Hart, T., Dijkers, M. P., Whyte, J., Turkstra, L. S., Zanca, J. M., Packel, A., Van Stan, J.H., Ferraro, M., & Chen, C. (2019). A theory-driven system for the specification of rehabilitation treatments. *Archives of Physical Medicine and Rehabilitation*, *100*(1), 172-180.

He, Q., Perera, S., Khalifa, Y., Zhang, Z., Mahoney, A. S., Sabry, A., Donohue, C., Coyle, J. & Sejdic, E. (2019). The association of high resolution cervical auscultation signal features with hyoid bone displacement during swallowing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(9), 1810-1816.

Hind, J. A., Nicosia, M. A., Roecker, E. B., Carnes, M. L., & Robbins, J. (2001). Comparison of effortful and noneffortful swallows in healthy middle-aged and older adults. *Archives of Physical Medicine and Rehabilitation*, *82*(12), 1661-1665.

Humbert, I. A., & Joel, S. (2012). Tactile, gustatory, and visual biofeedback stimuli modulate neural substrates of deglutition. *Neuroimage*, *59*(2), 1485-1490.

Humbert, I. A., Sunday, K. L., Karagiorgos, E., Vose, A. K., Gould, F., Greene, L., Azola, A., Tolar, A & Rivet, A. (2018). Swallowing kinematic differences across frozen, mixed, and ultrathin liquid boluses in healthy adults: Age, sex, and normal variability. *Journal of Speech, Language, and Hearing Research*, *61*(7), 1544-1559.

Kang, B. S., Oh, B. M., Kim, I. S., Chung, S. G., Kim, S. J., & Han, T. R. (2010). Influence of aging on movement of the hyoid bone and epiglottis during normal swallowing: A motion analysis. *Gerontology*, *56*(5), 474-482.

Kendall, K. A., Leonard, R. J., & McKenzie, S. W. (2003). Sequence variability during hypopharyngeal bolus transit. *Dysphagia*, *18*(2), 85-91.

Kim, S. J., Han, T. R., & Kwon, T. K. (2010). Kinematic analysis of hyolaryngeal complex movement in patients with dysphagia development after pneumonectomy. *The Thoracic and Cardiovascular Surgeon*, *58*(02), 108-112.

Kim, Y., McCullough, G. H., & Asp, C. W. (2005). Temporal measurements of pharyngeal swallowing in normal populations. *Dysphagia*, *20*(4), 290-296.

Kurosu, A., Coyle, J. L., Dudik, J. M., & Sejdic, E. (2019). Detection of swallow kinematic events from acoustic high-resolution cervical auscultation signals in patients with stroke. *Archives of Physical Medicine and Rehabilitation*, *100*(3), 501-508.

Lazarus, C., Logemann, J. A., & Gibbons, P. (1993). Effects of maneuvers on swallowing function in a dysphagic oral cancer patient. *Head & Neck*, *15*(5), 419-424.

Lee, J., Sejdic, E., Steele, C. M., & Chau, T. (2010). Effects of liquid stimuli on dual-axis swallowing accelerometry signals in a healthy population. *Biomedical Engineering Online*, *9*(1), 7.

Leslie, P., Drinnan, M. J., Finn, P., Ford, G. A., & Wilson, J. A. (2004). Reliability and validity of cervical auscultation: a controlled comparison using videofluoroscopy. *Dysphagia*, *19*(4), 231-240.

Lof, G. L., & Robbins, J. (1990). Test-retest variability in normal swallowing. *Dysphagia*, *4*(4), 236-242.

Logemann, J. A., Kahrilas, P. J., Cheng, J. O. A. N., Pauloski, B. R., Gibbons, P. J., Rademaker, A. W., & Lin, S. (1992). Closure mechanisms of laryngeal vestibule during swallow. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *262*(2), G338-G344.

Logemann, J. A., Pauloski, B. R., Rademaker, A. W., Colangelo, L. A., Kahrilas, P. J., & Smith, C. H. (2000). Temporal and biomechanical characteristics of oropharyngeal swallow in younger and older men. *Journal of Speech, Language, and Hearing Research*, *43*(5), 1264-1274.

Logemann, J. A., Pauloski, B. R., Rademaker, A. W., & Kahrilas, P. J. (2002). Oropharyngeal swallow in younger and older women. *Journal of Speech, Language, and Hearing Research*, *45*(3), 434-445.

Macrae, P., Anderson, C., Taylor-Kamara, I., & Humbert, I. (2014). The effects of feedback on volitional manipulation of airway protection during swallowing. *Journal of Motor Behavior*, *46*(2), 133-139.

Mann, G., Hankey, G. J., & Cameron, D. (1999). Swallowing function after stroke: Prognosis and prognostic factors at 6 months. *Stroke*, *30*(4), 744-748.

Martin-Harris, B., Brodsky, M. B., Price, C. C., Michel, Y., & Walters, B. (2003). Temporal coordination of pharyngeal and laryngeal dynamics with breathing during swallowing: Single liquid swallows. *Journal of Applied Physiology*, *94*(5), 1735-1743.

Martin-Harris, B., Garand, K. L., & McFarland, D. (2017). Optimizing respiratory-swallowing coordination in patients with oropharyngeal head and neck cancer. *Perspectives of the ASHA Special Interest Groups*, *2*(13), 103-110.

Martin-Harris, B., & Jones, B. (2008). The videofluorographic swallowing study. *Physical Medicine and Rehabilitation Clinics of North America*, *19*(4), 769-785.

Mendelsohn, M. S., & Martin, R. E. (1993). Airway protection during breath-holding. *Annals of Otology, Rhinology & Laryngology*, *102*(12), 941-944.

Molfenter, S. M., & Steele, C. M. (2012). Temporal variability in the deglutition literature. *Dysphagia*, *27*(2), 162-177.

Movahedi, F., Kurosu, A., Coyle, J. L., Perera, S., & Sejdic, E. (2016). Anatomical directional dissimilarities in tri-axial swallowing accelerometry signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(5), 447-458.

Nierengarten, M. B. (2009). Evaluating dysphagia: Current approaches. *Oncology Times*, *31*(14), 29-30.

Ohmae, Y., Logemann, J. A., Hanson, D. G., Kaiser, P., & Kahrilas, P. J. (1996). Effects of two breath-holding maneuvers on oropharyngeal swallow. *Annals of Otology, Rhinology & Laryngology*, *105*(2), 123-131.

Ohmae, Y., Logemann, J. A., Kaiser, P., Hanson, D. G., & Kahrilas, P. J. (1995). Timing of glottic closure during normal swallow. *Head & neck*, *17*(5), 394-402.

Park, T., Kim, Y., Ko, D. H., & McCullough, G. (2010). Initiation and duration of laryngeal closure during the pharyngeal swallow in post-stroke patients. *Dysphagia*, *25*(3), 177-182.

Power, M. L., Hamdy, S., Singh, S., Tyrrell, P. J., Turnbull, I., & Thompson, D. G. (2007). Deglutitive laryngeal closure in stroke patients. *Journal of Neurology, Neurosurgery & Psychiatry*, *78*(2), 141-146.

Rebrion, C., Zhang, Z., Khalifa, Y., Ramadan, M., Kurosu, A., Coyle, J. L., Perera, S. & Sejdic, E. (2018). High-resolution cervical auscultation signal features reflect vertical and horizontal displacements of the hyoid bone during swallowing. *IEEE Journal of Translational Engineering in Health and Medicine*, *7*, 1-9.

Robbins, J., Levine, R. L., Maser, A., Rosenbek, J. C., & Kempster, G. B. (1993). Swallowing after unilateral stroke of the cerebral cortex. *Archives of Physical Medicine and Rehabilitation*, *74*(12), 1295-1300.

Rofes, L., Arreola, V., Romea, M., Palomera, E., Almirall, J., Cabre, M., Serra-Prat, M., & Clave, P. (2010). Pathophysiology of oropharyngeal dysphagia in the frail elderly. *Neurogastroenterology & Motility*, *22*(8), 851-e230.

Rosenbek, J. C., Roecker, E. B., Wood, J. L., & Robbins, J. (1996). Thermal application reduces the duration of stage transition in dysphagia after stroke. *Dysphagia*, *11*(4), 225-233.

Sejdic, E., Malandraki, G. A., & Coyle, J. L. (2018). Computational deglutition: Using signal-and image-processing methods to understand swallowing and associated disorders [life sciences]. *IEEE Signal Processing Magazine, 36*(1), 138-146.

Shu, K. (2019). *Association Between Diameter of Upper Esophageal Sphincter Maximal Opening and High-Resolution Cervical Auscultation Signal Features* (Doctoral dissertation, University of Pittsburgh).

Steele, C., Allen, C., Barker, J., Buen, P., French, R., Fedorak, A., Irvine Day, S., Lapointe, J., Lewis, L., MacKnight, C., McNeil, S., Valentine, J., & Walsh, L. (2007). Dysphagia service delivery by speech-language pathologists in Canada: Results of a national survey. *Canadian Journal of Speech-Language Pathology and Audiology*, *31*(4), 166-177.

Steele, C. M., Bennett, J. W., Chapman-Jay, S., Cliffe Polacco, R., Molfenter, S. M., & Oshalla, M. (2012). Electromyography as a biofeedback tool for rehabilitating swallowing muscle function. *Applications of EMG in Clinical and Sports Medicine*, 311-328.

Takahashi, K., Groher, M. E., & Michi, K. I. (1994). Methodology for detecting swallowing sounds. *Dysphagia*, *9*(1), 54-62.

Taubert, M., Lohmann, G., Margulies, D. S., Villringer, A., & Ragert, P. (2011). Long-term effects of motor training on resting-state networks and underlying brain structure. *Neuroimage*, *57*(4), 1492-1498.

Young, J. L., Macrae, P., Anderson, C., Taylor-Kamara, I., & Humbert, I. A. (2015). The sequence of swallowing events during the chin-down posture. *American Journal of Speech-Language Pathology*, *24*(4), 659-670.

Zammit-Maempel, I., Chapple, C. L., & Leslie, P. (2007). Radiation dose in videofluoroscopic swallow studies. *Dysphagia*, *22*(1), 13-15.

Figure captions

Figure 1. Illustrates using the temporal binary classification method to train the CRNN architecture. The events of LVC and LVO were labeled by an experienced rater in kinematic analysis of VFS videos. The numbers ’0’ and ’1’ represent the opening and closure of LV respectively.

Figure 2. The frame error distribution for the validation results. The red bars represent an error no larger than 3 frames. (a) & (b) show the distribution of onset of LVC and onset of LVO respectively for the 10-fold validation dataset, which contained 588 swallowing samples. (c) & (d) show the distribution of onset of LVC and onset of LVO respectively for the testing dataset, which contained 45 unseen swallowing samples.

Figure 3. The accuracy levels for the LV status prediction across the 10 validation groups.