Upper Esophageal Sphincter Opening Segmentation with Convolutional Recurrent Neural Networks in High Resolution Cervical Auscultation

Yassin Khalifa, Cara Donohue, James L. Coyle, and Ervin Sejdić, *Senior, IEEE*

***Abstract*—Upper esophageal sphincter is an important anatomical landmark of the swallowing process commonly observed through the kinematic analysis of radiographic examinations that are vulnerable to subjectivity and clinical feasibility issues. Acting as the doorway of esophagus, upper esophageal sphincter allows the transition of ingested materials from pharyngeal into esophageal stages of swallowing and a reduced duration of opening can lead to penetration/aspiration and/or pharyngeal residue. Therefore, in this study we consider a non-invasive high resolution cervical auscultation-based screening tool to approximate the human ratings of upper esophageal sphincter opening and closure. Swallows were collected from 116 patients and a deep neural network was trained to produce a mask that demarcates the duration of upper esophageal sphincter opening. The proposed method achieved more than 90% accuracy and similar values of sensitivity and specificity when compared to human ratings even when tested over swallows from an independent clinical experiment. Moreover, the predicted opening and closure moments surprisingly fell within an inter-human comparable error of their human rated counterparts which demonstrates the clinical significance of high resolution cervical auscultation in replacing ionizing radiation-based evaluation of swallowing kinematics.**

***Index Terms*—Swallowing Accelerometry, Swallowing Vibrations, Cervical Auscultations, Dysphagia, Upper Esophageal Sphincter, Signal Processing, Deep Learning, Supervised Learning, Convolutional Recurrent Neural Networks, GRU.**

I. Introduction

SWALLOWING is a complex process that involves the coordination of various anatomical structures, muscles, and the biomechanical events they perform, in a somewhat sequential order to safely and efficiently transport food and liquids from the oral cavity to the stomach [1], [2]. Because swallowing requires the coordination of multiple subsystems of the body, a variety of medical or surgically related conditions can cause swallowing impairments, also known as dysphagia [2]–[4]. Dysphagia is prevalent with approximately 16%-22% of people over the age of 50, 12%-13% of short-term care patients, and up to 60% of nursing home residents experiencing swallowing difficulties [5]–[7]. Dysphagia can result in aspiration, or the entry of food and/or liquid into the airway below the level of the vocal folds. Aspiration of food and liquids is concerning, especially silent aspiration (Asymptomatic), because it can lead to adverse outcomes including pneumonia, malnutrition, and dehydration [7]–[9], as well as reduced quality of life [10]–[17].

Among the most important physiologic correlates of healthy swallowing function is the duration of upper esophageal sphincter (UES) opening. UES opening enables food and liquid to enter the esophagus [18]–[21]. Reduced UES opening diameter, delayed onset of opening, or premature closure attenuate UES opening duration and can result in pharyngeal residue that in turn can enter the upper (laryngeal penetration) or lower (aspiration) airway, which are known risk factors for pneumonia and airway obstruction [22]. UES opening is the product of hyolaryngeal excursion, bolus propulsion, and neural inhibitory relaxation of the UES itself [21], [22]. UES dysfunction may occur due to neurological diseases that alter the timing of UES relaxation and the delivery of muscular traction forces that act to distend the relaxed UES during swallowing, or due to impaired propulsive forces applied by the oropharyngeal pump [19], [22].

Table I summarizes the different diagnostic modalities that can generate images and signals for the assessment of UES function [19], [24], [25]. The modalities include videofluoroscopic swallow studies (VFSSs), fast pharyngeal CT/MRI, fiberoptic endoscopic evaluation of swallowing (FEES), and non-imaging instrumental tests such as pharyngeal manometry and Electromyography (EMG). Most of these modalities require expertise to perform and highly trained clinicians to interpret. VFSSs are most frequently and actually the best modality to clinically assess swallow kinematic events such as UES opening, because of the ability to dynamically visualize the UES during all phases of the swallow and give exact estimates of the moments when UES opens and closes [18], [19]. However, VFSSs, which use ionizing radiation to produce radiographic images with full temporal resolution, are unavailable or undesirable to many patients, are relatively expensive, and require specialized instrumentation and trained clinicians to perform and interpret, leaving many patients undiagnosed or inaccurately diagnosed, and exposed to ongoing risk of dysphagia-related complications [18].

Y. Khalifa is with Department of Electrical and Computer Engineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA, USA.

C. Donohue and J. L. Coyle are with Department of Communication Science and Disorders, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, USA.

E. Sejdić is with Department of Electrical and Computer Engineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA, USA,
Department of Bioengineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA, USA,
Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA,
Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, USA.
E-mail: esejdic@ieee.org

TABLE I
SUMMARY OF TOOLS USED FOR DIAGNOSTIC ASSESSMENT OF UES.

Modality	Strengths	Weaknesses
VFSS [19]	- Dynamically visualize UES during all phases of swallowing - Provides the exact moments when UES opens and closes	- Subjective interpretation - Radiation exposure
FEES [23]	- Direct visualization of swallowing pharyngeal stage	- Limited in describing UES activity (either probe is covered with bolus or already through UES)
CT/MRI [19]	- Panoramic and full-thickness visualization of oropharyngeal structures	- Hard to conduct - Radiation exposure (CT) - Require synchronization with patient behavior (MRI) - Limited availability
Manometry [19]	- Monitor UES pressure during swallowing - Detect UES impaired relaxation/distension	- Invasive - Subjective interpretation - Limited availability
EMG [19]	- Monitor muscle activations during swallowing - Detect UES impaired relaxation/distension	- Can't tell the exact moments when UES opens/closes - Subjective interpretation

The holy grail of dysphagia clinical evaluation methods has long been a noninvasive and clinically feasible method of accurately identifying the biomechanical events of swallowing that contribute to airway protection such as UES opening. The availability of such methods would enable the development of a screening tool that can differentiate between impaired and healthy swallowing with a high degree of sensitivity and specificity without the uncertainty of clinical examinations or the lack of availability of imaging studies [21], [26]–[29]. To address the obstacle of insufficient access to instrumental testing of swallowing function universally, high resolution cervical auscultation (HRCA) is currently being investigated as an affordable, feasible, non-invasive bedside assessment tool for dysphagia. HRCA combines the use of vibratory signals from an accelerometer with acoustic signals from a microphone attached to the anterior neck region during swallowing. Following collection of signals, advanced machine learning techniques are used to examine the association between HRCA signals and physiological events that occur during swallowing [30], [31].

HRCA has shown strong associations with multiple factors that affect the UES opening process. For instance, HRCA has been used in multiple studies to monitor the pharyngeal bolus propulsion during swallowing from the moment the bolus passes the mandible till the UES closes [32]–[35]. Furthermore, hyolaryngeal excursion has been investigated to be the origin of HRCA signals in many occasions [36]–[38], and later they were successfully used to actually track the location of the hyoid bone during swallowing [31]. The formerly mentioned events are all parts of the UES opening mechanism which proves the potential of HRCA signals in detection of UES opening. While previous studies have monitored changes in HRCA signal features at the moments of UES opening and closure [39].], no studies have used HRCA signals to measure the time of UES opening and closure within a swallow.

As mentioned previously, UES opening is the result of a mechanism that is controlled by multiple events occurring during swallowing, which necessitates the temporal modeling of the whole swallow for the purpose of UES opening detection. Recurrent neural networks (RNNs) have been extensively employed for the time series modeling in the recent years, due to their capability of carrying information from arbitrarily long contexts, selective information transfer across time steps, and affordable scalability [40], [41]. RNNs are seemingly efficient in modeling temporal contexts in time series data and have been used in event detection for many biomedical signals like ECG and EEG [42], [43], but nevertheless using RNNs on raw signals is extremely hard to optimize because of the propagating error signals through huge number of time steps [44], [45]. To overcome this, convolutional neural networks (CNNs) have been utilized for the perception of short contexts and more abstraction before feeding into RNNs for the perception of longer temporal contexts [44]. Known as representation learning, such hybrid architectures allow feeding the machines with raw data to automatically discover representations necessary for the detection problem [45]. These models were first conceived for computer vision applications [44], [46]; however, similar designs are being adopted recently for event detection in biomedical signals [47], [48] in addition to numerous applications in audio and speech signal processing [49].

In this study, we propose an implementation that uses HRCA acceleration signals to estimate the moments at which the UES opens and re-closes during swallowing and compare the estimates to gold-standard judgments of UES opening duration in videofluoroscopic images. The proposed method relies on convolutional recurrent neural networks to extract the dynamics of the swallowing vibrations from HRCA signals and use them to infer the moments when the UES first opens and re-closes during swallowing. Verifying the ability of HRCA signals to demarcate the UES opening among other swallowing physiological events, will promote a new non-invasive sensor-based swallowing assessment technology that is widely available and doesn't add financial or relocation burdens to patients. Moreover, it will help patients get a consistent feedback about their swallowing, while they are swallowing; a feature that will not only help improve the clinic-based swallowing evaluation, but will also be a of a great benefit for the patients towards feeling the progress of the rehabilitation process and maintaining safe swallowing.

II. Methodology

A. Materials and Methods

Permission for this study was granted by the institutional review board of the University of Pittsburgh and all participating patients provided informed consents including consent to publish before enrollment. A total of one hundred and sixteen patients (72 males, 44 females, age: 62.7 ± 15.5) with suspected dysphagia resulting from a variety of diagnoses, underwent an oropharyngeal swallowing function evaluation by a speech language pathologist using VFSS at the University of Pittsburgh Medical Center Presbyterian Hospital (Pittsburgh, PA). Of the sample, 15 patients were diagnosed with stroke while the remaining 101 patients were diagnosed with different medical conditions unrelated to stroke.

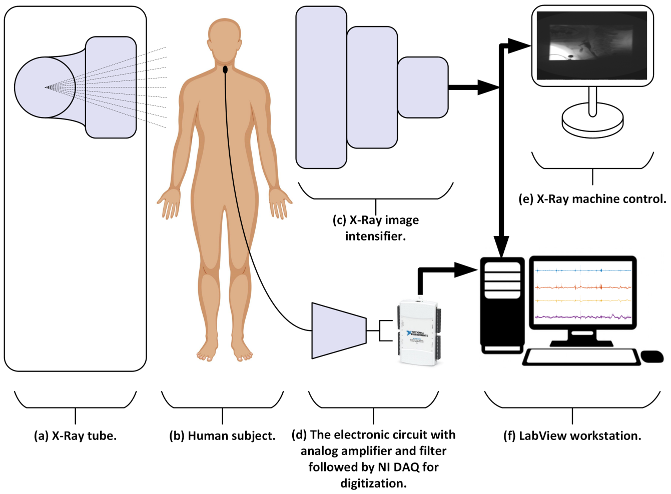

(c) X-Ray image intensifier.
(e) X-Ray machine control.
(a) X-Ray tube.
(b) Human subject.
(d) The electronic circuit with analog amplifier and filter followed by NI DAQ for digitization.
(f) LabView workstation.

Fig. 1. The experimental setup of the study. (a) An X-Ray tube that resides in a table is adjusted in a vertical position to be parallel to the swallowing path. (b) The human subject is standing or comfortably seated between the x-ray tube and the image intensifier with the HRCA sensors attached to the anterior neck. (c) The image intensifier is positioned and adjusted according to the subject height, so that the produced frames capture all of the important anatomical landmarks of the oropharyngeal swallow (jaws, pharynx, and esophagus). (d) The sensors are connected to the electronic circuit that supplies power and performs analog amplification and filtration and then to the NI DAQ for sampling. (e) The video feed is taken directly from the image intensifier to the X-Ray control workstation where clinicians and radiologists create, save , and view the exams. (f) The video feed from the image intensifier is cloned into the video capture card installed on the research workstation which is also connected to the NI DAQ and runs LabView for means of data collection and synchronization.

Swallows for this study, were collected as a part of standard clinical care rather than for research purposes alone. As a result, speech language pathologists who conducted the VFSSs, had the ability to alter the evaluation protocol based on the patient's clinical manifestation of dysphagia. This included how the boluses were administered to patients (i.e. spoon, cup), the volume and viscosity/texture of each bolus of food and liquids, the number of trials, and head position during swallowing (i.e. head/neck flexion, head rotation, head neutral). The following consistencies were used during VFSSs: thin liquid (Varibar thin, Bracco Diagnostics, Inc., < 5 cPs viscosity), mildly thick liquid (Varibar nectar, 300 cPs viscosity), puree (Varibar pudding, 5000 cPs viscosity), and Keebler Sandies Mini Simply Shortbread Cookies (Kellogg Sales Company). Boluses were either self-administered by patients via a cup or a straw or administered by the clinician through the use of a spoon ($3-5$ mL).

This study yielded 710 swallows (132 from patients diagnosed with stroke and 578 from patients with other diagnoses) with an average duration of pharyngeal bolus transit of 869.5±221 msec and an average DUESO of 604.9±150 msec. The collected swallows were classified into three categories: single (single bolus swallowed with one swallow), multiple (single bolus swallowed using more than one swallow), or sequential (multiple boluses swallowed sequentially in a rapid manner). The final data included 224 single, 477 multiple, and 9 sequential swallows.

B. Data Acquisition

The general experimental setup is illustrated in Fig. 1. During all recording sessions, VF equipment was controlled by a radiologist and the patients were comfortably seated with the swallowing sensors attached to the anterior neck region using double sided tape. VF was conducted in the lateral plane using a Precision 500D system (GE Healthcare, LLC, Waukesha, WI) at a pulse rate of 30 pulses per second (PPS) and with the images acquired a frame rate of 30 frames per second (FPS) [50]. The video stream was captured and digitized using an AccuStream Express HD video card (Foresight Imaging, Chelmsford, MA) into movie clips with a resolution of 720 × 1080 at 60 FPS.

A tri-axial accelerometer (ADXL 327, Analog Devices, Norwood, Massachusetts) and a contact microphone (model C 411L, AKG, Vienna, Austria) were used to collect swallowing vibratory and acoustic signals. The accelerometer was mounted into a small plastic case with a concave surface that fits on neck curvature and the case was attached to the skin overlying the cricoid cartilage using a tape. The accelerometer was attached such that its main axes are aligned parallel to the cervical spine, perpendicular to the coronal plane, and parallel to the axial/transverse plane. These axes are referred to as superior-inferior (S-I), anterior-posterior (A-P), and medial-lateral (M-L) respectively. The microphone was mounted towards the right lateral side of the larynx to avoid contact noise with the accelerometer and guarantee a clear radiographic view of the upper airway. Attaching the sensors around the area of cricoid cartilage is logical given that most of the pharyngeal swallowing activity is produced by the anatomical structures present at this level and it has been reported to yield the best signal-to-noise ratio for the acquisition of swallowing signals [34], [35], [51], [52].

The accelerometer has a bandwidth of 1600 Hz after which the response falls to -3dB of the response to low frequency acceleration. In other words, the accelerometer has a low pass filter with a cut-off frequency at 1600 Hz. The contact microphone was chosen as well so that it produces a flat frequency response over the entire range of audible sounds which was proved to pass most of the frequencies encountered during swallowing [52]–[54]. The signals from both the accelerometer and microphone were hardware band-limited to 0.1-3000 Hz with an amplification gain of 10. The cut-off frequencies for the band-limiting filter were chosen so that most of body sway components below 0.2 Hz are suppressed and the signal components with the vast majority of energy are passed [34], [54]–[56]. The signals were sampled using a National Instruments 6210 DAQ at a sampling rate of 20 kHz. Both signals and video were acquired simultaneously using LabView's Signal Express (National Instruments, Austin, Texas) with a complete end-to-end synchronization.

C. VF Image Analysis

Video clips were segmented based on individual swallow events by tracking the bolus in a frame by frame manner. The onset of the pharyngeal swallow event was defined as the frame in which the head of the bolus passes the shadow of the posterior border of the ramus of the mandible and the offset as the frame in which the bolus tail passes through the UES [57], in order to capture the entire duration of pharyngeal bolus flow. Three expert judges trained in swallow kinematic judgments, identified the video frame of first UES opening and the video frame of first UES closure in the segmented videos. All raters who segmented swallowing videos and analyzed UES opening and closure established a priori intra- and inter-rater reliability with ICC's over 0.99. All raters maintained intra- and inter-rater reliability throughout measurements on 10% of swallows with ICC's over 0.xx and were blinded to participant demographics and diagnosis and any bolus condition information.

D. Signals Preprocessing

Numerous physiologic and kinematic events such as coughing and breathing occur in close temporal proximity to the pharyngeal swallow event. These events can contribute to the collected vibratory and acoustic signals [33]. As a first step to overcome confounding noise in the signals due to multi-source environmental data collection and other measurement errors, the signals accrued at a sampling rate of 20 kHz were down-sampled to 4 kHz. A more intense down-sampling could have been adopted as previous studies reported that the frequency with the maximum energy for swallowing accelerometry signals occurs below 100 Hz and the central frequency almost below 300 Hz [34], [58]–[60]. However, we chose down-sampling to 4 kHz so that we match twice the max frequency component present in the acceleration signals (1600 Hz). Down-sampling was performed through applying an anti-aliasing low pass filter then picking up individual samples to match the new rate.

The baseline outputs of accelerometer and microphone (produced by zero-physical input) were recorded earlier before the main data collection procedure and device noise was characterized through modified covariance auto-regressive modeling [58], [61]. The order of the auto-regressive model was 10 and it was determined using the Bayesian information criterion [58]. The coefficients of the auto-regressive model were then used to create a finite impulse response filter (FIR) to remove the device noise from the recorded swallowing signals [58]. Afterwards, the low-frequency noise components and motion artifacts were eliminated from accelerometer signals using fourth-order least-square splines [62], [63]. Particularly, we used fourth-order splines with a number of knots equivalent to $\frac{N \times f_l}{f_s}$, where N is the data length and f_s is the sampling frequency. f_l is called the lower sampling frequency and it is proportional to the frequency associated with motion artifacts. The values for f_l were calculated and optimized in previous studies [62]. Finally, the effect of broadband noise on signals was reduced through wavelet denoising [64]. Specifically, we used tenth-order Meyer wavelets and soft thresholding. The threshold was calculated using $\sigma\sqrt{2 \log N}$, where N is the number of samples and σ is the estimated standard deviation of the noise (calculated through down-sampling the wavelet coefficients) [64], [65].

E. System Design

Due to the fact that there is no specific rule of thumb to calculate the number of layers and layer sizes for a certain problem, the used architecture was fine-tuned based on an experimental approach and by following the best network configurations that achieved good results in similar problems [47], [49], [66]. Particularly, we tested multiple architecture depths that included more layers of CNN (3, 4, and 5 layers) with up to 32 filters per channel and more RNN unit sizes up to 128. The chosen architecture was found to be the most stable among the tested configurations. In other words, it included the smallest number of parameters to be optimized while achieving a detection accuracy that doesn't sharply change when adding more layers or increasing the layer sizes. The used architecture employed also dropout between layers as well as early stopping techniques to control the network from over-fitting to the training data [67].

The longest swallow event duration in the collected dataset was around 1500 msec (90 frames of VF). The signals were divided into chunks 16.67 msec in length (equivalent to one frame in VF or 66 samples in signals). Each signal chunk is composed of 3 axes of acceleration which makes the dimensions 66 samples × 3 channels. The chunks were fed into a 1D convolutional neural network that included two convolutional layers with a max pooling layer in between as in Fig. 2. Both convolutional layers were followed by a rectified linear unit (ReLU). The first convolutional layer applied 16 ”1 × 5” filters per channel which produced 3 ”62 features × 16 channels”. The max pooling layer applied a window of size 2 with 2 strides and reduced the features into ”31 features × 48 channels”. The last convolutional layer was identical to the first one except that it used only one filter per channel which produced ”27 features × 48 channels”.

The complete sequence of features $x_{1:T}$ (for a full swallow) coming out of the convolutional layer was then fed into a 3-layers dynamic RNN with gated recurrent units (GRUs) as building blocks each of 64 units and a sequence of 90 time steps. The RNN computed an output sequence $\hat{y}_{1:T}$ using the following nonlinear model:

$$
\begin{aligned}
r_t^{(k)} &= \begin{cases} \sigma(W_r^{(1)} \left[h_{t-1}^{(1)}, x_t\right] + b_r^{(1)}), & \text{k=1,} \\[2ex] \sigma(W_r^{(k)} \left[h_{t-1}^{(k)}, h_t^{(k-1)}\right] + b_r^{(k)}), & \text{k=2, 3} \end{cases} \\
z_t^{(k)} &= \begin{cases} \sigma(W_z^{(1)} \left[h_{t-1}^{(1)}, x_t\right] + b_z^{(1)}), & \text{k=1,} \\[2ex] \sigma(W_z^{(k)} \left[h_{t-1}^{(k)}, h_t^{(k-1)}\right] + b_z^{(k)}), & \text{k=2, 3} \end{cases} \\
\hat{h}_t^{(k)} &= \begin{cases} tanh(W^{(1)} \left[r_t^{(1)} h_{t-1}^{(1)}, x_t\right] + b^{(1)}), & \text{k=1,} \\[2ex] tanh(W^{(k)} \left[r_t^{(k)} h_{t-1}^{(k)}, h_t^{(k-1)}\right] + b^{(k)}), & \text{k=2, 3} \end{cases} \\
h_t^{(k)} &= \quad z_t^{(k)} \hat{h}_t^{(k)} - z_t^{(k)} h_{t-1}^{(k)}, \qquad \text{k=1, 2, 3} \\
\hat{y}_t &= \quad U h_t^{(3)} + c
\end{aligned}
$$

The output sequence $\hat{y}_{1:T}$ coming out of the RNN was masked (ones/zeros mask) before being fed in to the following stages to balance for the shorter swallows (less than 90 frames). Furthermore, the length of each swallow was

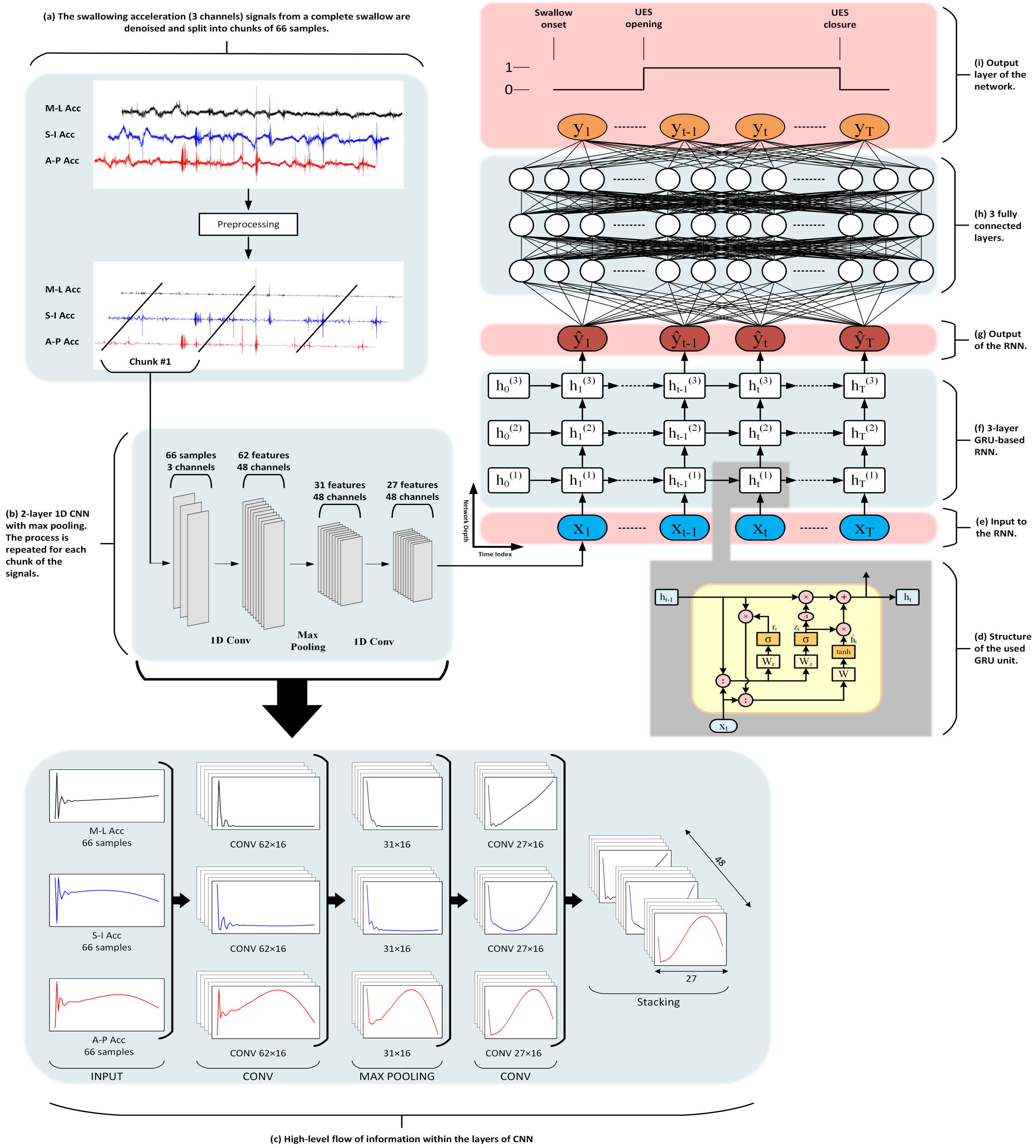


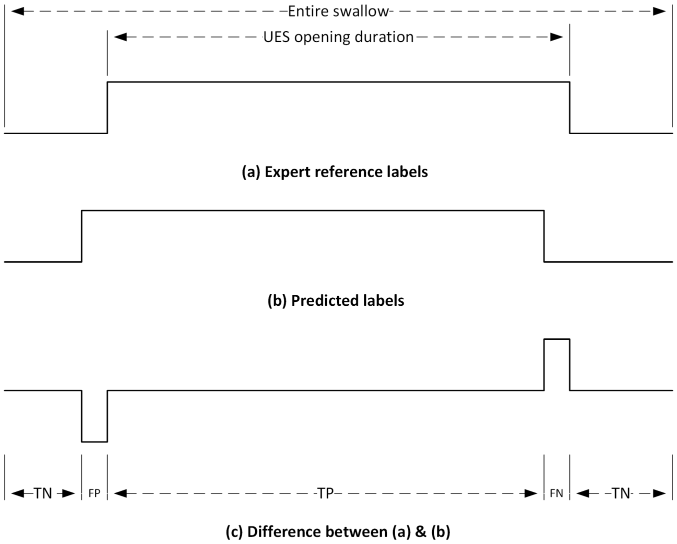
Fig. 2. The architecture and data flow in the UES opening detection system. (a) This part is where the 3-channel acceleration signals from each swallow are denoised and split into equal chunks each of 66 samples (equivalent to 1 VF frame). (b) This part shows the operation of the CNN network part per data chunk. The architecture of the used 1D CNN which is comprised of two layers, the first applies 16 filters on each channel and produces 48 channels. The first CNN layer is followed by a max pooling layer and another CNN layer identical to the fist except that it applies 1 filter per channel then a max pooling layer reduces the size of the features. (c) This is an illustration for the operation of the CNN after training that shows a chunk of 3-channel acceleration pushed throw the first layer of CNN to produce 16 feature-channels per original channel. The length of chunks is shorter after this layer due to convolution on the edges of the chunks (no padding is used). (d) This is an illustration that shows the architecture of the GRU unit with the reset and update parts that help propagate states across time steps. (e) $(x_{1:T})$ is the output train from the CNN for chunks $(1:T)$ which is fed into the RNN units. (f) The architecture of the 3-layer RNN used for time sequence modeling. (g) The output sequence from the last layer of the RNN $(\hat{y}_{1:T})$ is flattened and fed into the first fully connected layer. (h) A diagram of the 3 fully connected layers (each of 128 units) used to combine the features coming out of the RNN. (i) The output layer of the network which composed of 90 units $(y_{1:T})$ that resemble the UES opening mask.

considered in the architecture of the RNN and the same mask was used in the calculation of the cost function for the whole problem. The sequence was then fed in to 4 fully connected layers in order to fuse the temporal features from RNN into a meaningful UES opening segmentation mask. This part of the network featured 3-ReLU activated layers with 128 units and an output layer that assembled 90 units, one for each time step in the swallow as shown in Fig. 2 plus Sigmoid activation for a zeros and ones segmentation mask. Each two fully connected layers were separated by a dropout layer with a drop rate of 20%.

The final cost function was defined as the mean squared error between the zero-padded ground truth $\bar{y}_{1:T}$ labeled by the expert judges and the masked output coming from the final connected layer $\hat{y}_{1:T}$ as follows:

$$MSE = \frac{1}{T}\sum_{i=1}^{T}\left[(\bar{y}_i - \hat{y}_i) \times mask_i\right]^2 \qquad (1)$$

where $mask_i$ is the mask used to compensate for short swallows. We used the Adam optimizer to train the network due to its superiority in convergence without fine tuning for hyper-parameters [68].



(a) Expert reference labels

(b) Predicted labels

(c) Difference between (a) & (b)

Fig. 3. The evaluation procedure for each swallow. (a) The UES opening mask created from the expert manual segmentation in VF images. (b) The UES opening mask as predicted by the proposed algorithm. (c) Comparison is performed between the masks from (a) and (b) to create a confusion matrix. The confusion matrix is created in this way for each swallow included in testing. The values of accuracy, sensitivity, and specificity are calculated through this confusion matrix.

F. Evaluation

The dataset was randomly divided into 10 equal subsets in terms of the number of swallows. A holdout method was repeated 10 times by training with 9 subsets and testing with the remaining one (10-fold cross validation). The results of the proposed system are in the form of a segmentation mask that tells when the UES opens and closes with respect to the start (onset) of the swallow segment as shown in Fig. 3 (b). This mask is calculated for approximately each swallow in the dataset when passed as a test sample through the trained system. In order to acquire a solid evidence about the detection quality of the system, a confusion matrix is constructed for each swallow based on the predicted segmentation mask and the reference mask as labeled by judges. The confusion matrix is then used to calculate accuracy, sensitivity, and specificity as follows:

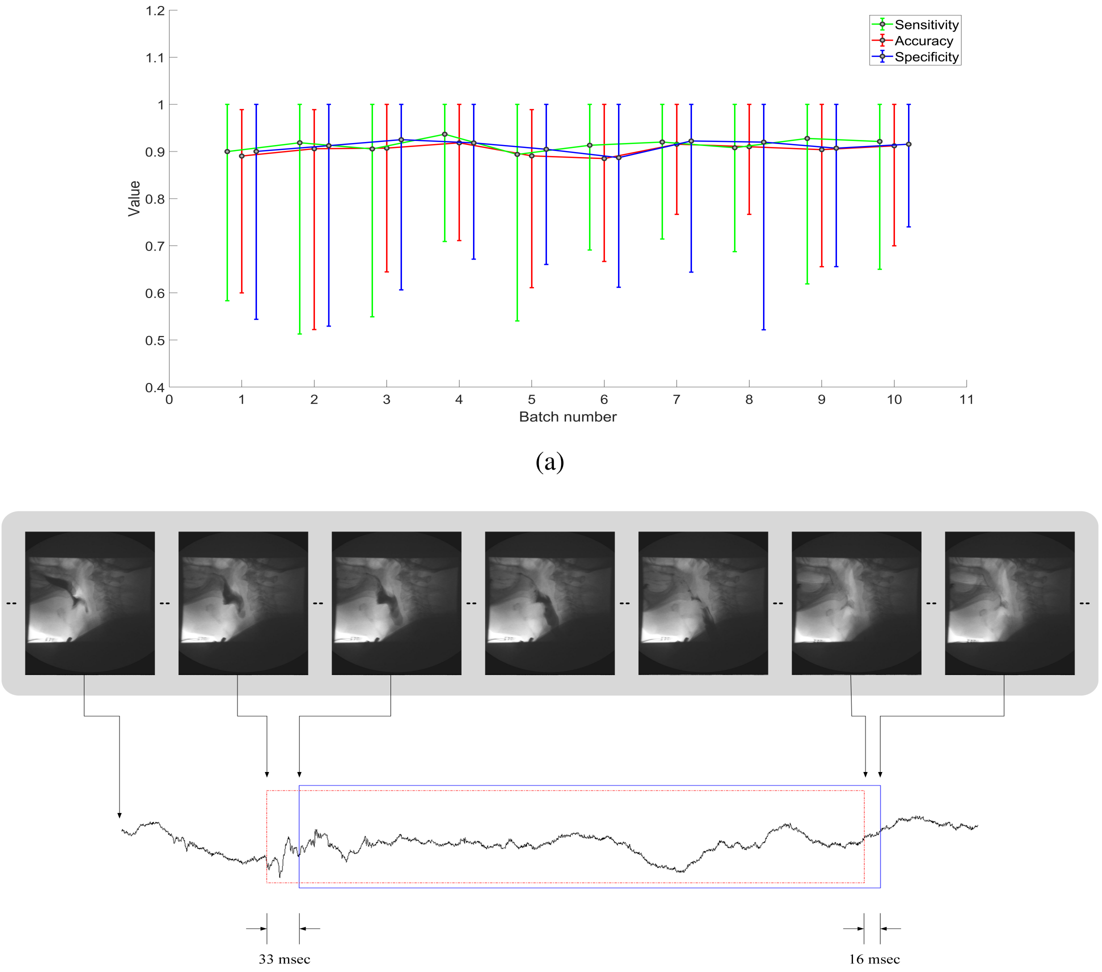
$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+FP+TN+FN} \\ Sensitivity &= \frac{TP}{TP+FN} \\ Specificity &= \frac{TN}{FP+TN} \end{aligned}$$

where TP stands for True Positive, TN stands for True Negative, FP stands for False Positive, and FN stands for False Negative. Furthermore, the difference between the actual and predicted UES opening and UES closure was measured, so that we could compare it to the human judges' tolerance reported in the literature.

G. Clinical Validation

In order to evaluate the proposed system in a clinical environment, it was tested during the workflow of an ongoing clinical experiment performed on 15 (8 males, 7 females, age: 63.7 ± 6.2), community dwelling healthy adults who provided informed consent, and who had no reported current or prior swallowing difficulties. Participants in this validation sample also had no history of neurological disorder, surgery to the head or neck region, or chance of being pregnant based on participant report. The experimental setup of this clinical experiment relied on the same equipment and hardware used for the collection of the main dataset as shown in Fig. 1. This included recording VF in the lateral plane using a Precision 500D system (GE Healthcare, LLC, Waukesha, WI) at a pulse rate of 30 pulses per second (PPS) and with the images acquired a frame rate of 30 frames per second (FPS). The video stream was captured and digitized using an AccuStream Express HD video card (Foresight Imaging, Chelmsford, MA) at 60 FPS. Swallowing vibratory and acoustic signals were acquired concurrently with VF using the same tri-axial accelerometer and microphone (ADXL 327, Analog Devices, Norwood, Massachusetts andmodel C 411L, AKG, Vienna, Austria). The sensors were attached to the same location on the anterior neck to the skin overlying the cricoid cartilage. The signals from both sensors were also band-limited between 0.1-3000 Hz and amplified with a gain of 10 then sampled at a rate of 20 kHz via a National Instruments 6120 DAQ through LabView's Signal Express (National Instruments, Austin, Texas).

The participants in this clinical experiment were community dwelling adults without report of current or prior swallowing difficulties. Therefore, only ten thin liquid boluses (5 at 3mL by spoon, 5 unmeasured self-selected volume cup sips) administered in a randomized order in order to limit x-ray radiation exposure. For all spoon presentations, participants were instructed by the researcher to ”Hold the liquid in your mouth and wait until I tell you to swallow it.” Liquid bolus presentations by cup varied in volume by participant, because participants were instructed by the researcher to ”Take a comfortable sip of liquid and swallow it whenever



(a)

(b)

Fig. 4. Distribution of per swallow based performance measurements in each testing batch of the 10-fold cross validation process and a sample visual of the detection in one of the swallows. A sample of figures showing the timing difference between the automatically detected DUESO by our algorithm and the actual DUESO observed from VF (in frames) for both opening and closure. (a) Distribution for accuracy, sensitivity, and specificity in each batch (min, average, and max). (b) shows a sample full swallow with both the predicted (in red) and the actual DUESO (in blue) marked on the A-P acceleration component and video frames.

you're ready." Fifty swallows, selected randomly from this independent clinical experiment, were used to test the system for UES opening detection after being trained over the full 710 swallows dataset.

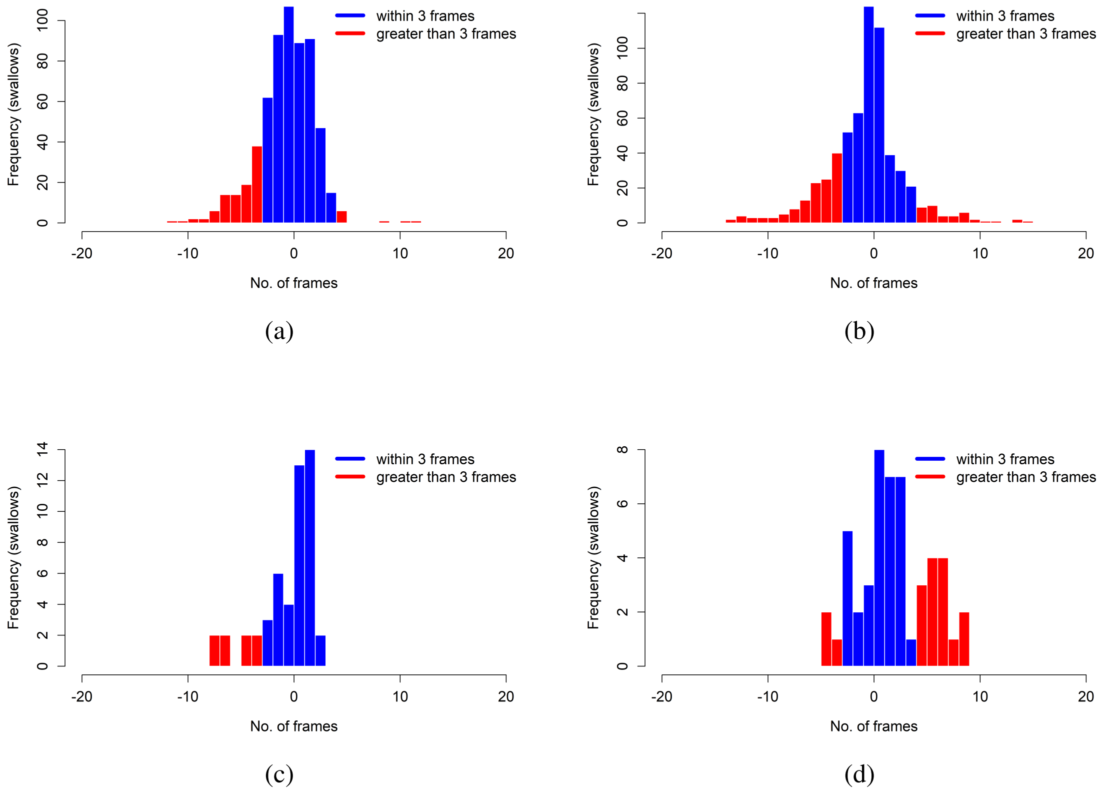
III. Results

A chunk of 3D acceleration (3×133) was first preprocessed to achieve denoising and artifact removal as shown in Fig. 2. After preprocessing, the filtered acceleration segments were fed into the convolutional network (CNN) part of the system as in the snapshot shown in the lower part of Fig. 2. The snapshot represents a sample feature map across the CNN that shows the evolution of inputs (low-level features) into high level features at the final layer of the CNN. The later helps identify more complex features in the input signals and promote distinctive traits while the insignificant features disappear.

TABLE II
Summary of the performance measurements that the proposed system achieved for both the main patient and the independent clinical datasets.

	Main dataset	Independent dataset
Average Accuracy	0.9093	0.8880
Average sensitivity	0.9145	0.8559
Average specificity	0.9119	0.9356
% of swallows with UES opening error $<$ 3 VF frames	82.6	84
% of swallows with UES opening error $<$ 4 VF frames	90	88
% of swallows with UES closure error $<$ 3 VF frames	72.3	66
% of swallows with UES closure error $<$ 4 VF frames	80	74

Fig. 4 (a) shows the performance of the proposed system across the 10-folds of the whole set of swallows. The values presented, represent the distribution of sensitivity, accuracy, and specificity in each fold. Each vertical line has 3 main points that represent the min average and maximum respectively from bottom up. The average accuracy of all folds across the whole dataset was 0.9039 with 0.9145 sensitivity and



(a) (b)

(c) (d)

Fig. 5. The timing difference between the automatically detected DUESO by the proposed system and the actual DUESO observed from VF (in frames) for both opening and closure in the whole dataset and the clinically independent data. The differences between the detected opening frame and the opening frame marked by the judges are highlighted in (a) for the 10 folds within the original dataset and in (c) for the clinically independent data. The differences between the detected closure frame and the closure frame marked by the judges are highlighted in (b) for the 10 folds within the original dataset and in (d) for the clinically independent data. The Positive values indicate that the actual UES opening and closure preceded the predicted UES opening and closure.

0.9119 specificity. Fig. 4 (b) depicts a comparison between DUESO detection from the proposed system against the manual labeling by experts through the use of VF. On average, the network detected UES opening 33 msec earlier and closure 16 msec earlier than true opening and closure as measured by swallow kinematic analysis. The outcome of the algorithm for the whole set of swallows, was calculated and compared to the VF based labels and the differences are shown through the histograms in Fig. 5 (a-b) and Table II. The comparison shows that for 82.6% of the swallows, the opening of UES was detected within a 100 msec ($\approx$ 3 frames at 30 FPS) of the human ratings, and within a 133 msec ($\approx$ 4 frames at 30 FPS) for 90% of the swallows (Fig. 5 (a)). Likewise, the network accurately detected UES closure within a 100 msec ($\approx$ 3 frames at 30 FPS) for 72.3% of the swallows and within a 133 msec ($\approx$ 4 frames at 30 FPS) for more than 80% of the swallows (Fig. 5 (b)). The accepted tolerance for human frame selection $\approx \pm 2.48$ frames at 30 FPS [57].

The system also presented similar results when tested using the swallows from the independent clinical experiment as in Table II. for the 50 swallows, the system achieved an average per swallow accuracy of 0.8880, an average per swallow sensitivity of 0.8559, and an average per swallow specificity of 0.9356. Fig. 5 (c-d) show histograms for the difference between the automatic detection and the reference manual labeling of the DUESO in terms of opening and closure frames. The results showed that UES opening and closure were detected within a 100 msec tolerance in around 84% and 66% of the swallows in the independent test set respectively.

IV. Discussion

The main purpose of this study was to test the feasibility of HRCA in detecting the exact timing of UES opening and closure during swallowing using non-invasive neck-attached sensors independent of VFSS images and to compare the accuracy to human ratings of the DUESO. We have established the fact that UES opening can be best visualized using VF which is clinically impractical due to the delivered radiation doses and unavailability outside clinical care settings. We have also demonstrated the critical rule that UES plays during swallowing and how monitoring its opening and closure will help identify the risks leading to unsafe swallowing. As a necessary part of the optimal goal to create a non-invasive swallowing monitoring system, UES opening/closure detection should help patients with brainstem parts, responsible for swallowing regulation, damaged and/or surgically removed to rehabilitate and relearn how to swallow. These patients will have a consistent feedback to tell if they are correctly performing swallowing compensation maneuvers in which they are taught to improve the hyolaryngeal excursion which would in turn reflect on UES duration/diameter and airway protection in order to maintain a safe function.

Prior studies have only addressed indicators and changes in HRCA signal features at the UES opening and closure moments or during the passage of the bolus through UES, but non of them offered a direct way to detect the DUESO during swallowing. Some of these studies reported the presence of localized maxima of some HRCA signal features at UES opening and closure times [39], [69]. One study also observed changes in the acoustic component of HRCA signals while the bolus passed through the UES [70]. Although these studies were essential for establishing the association between UES opening and HRCA signals, they were just descriptive analyses about the patterns in signal features at certain points of time when physiological events occurred. Therefore, in this study we aimed to explore a more advanced predictive profile to detect the DUESO from HRCA signal through considering the time dependency along the swallowing segment. As such we have demonstrated the system's feasibility on detecting DUESO without VFSS image verification.

One major disadvantage of human ratings is the subjectivity which creates an inter-rater tolerance of 82 msec (≈ ± 2.48 frames at 30 FPS) as reported for measuring swallowing kinematic events [57]. Human ratings of swallow kinematic events can also drift over time and necessitates that raters maintain ongoing intra and inter-reliability over time to maintain an appropriate error tolerance. Having an automated system that is capable of rating the swallowing kinematic events with a comparable human rater accuracy and impregnable to changes over time, is advantageous for swallowing analysis when imaging technology is unavailable, not feasible, or otherwise impractical for evaluating swallowing physiology. Based on the results, we can clearly see that the proposed system accurately detected up to 93.6% of the actual DUESO with low rates of false positives and negatives occurring only at the borders of DUESO as shown in Fig. 4 (b). These results were also achieved regardless of gender, age, or diagnosis of the subjects which assures the wide applicability of the system.

The system also showed robust performance when applied to a completely independent set of swallows that were collected from a different group of participants with different conditions and never seen in the training dataset. In terms of global measurements, the system achieved a close testing accuracy compared to the validation done through the folds of the original dataset (0.888 vs. 0.9035) and the same for sensitivity and specificity. It didn't come short either on the side of temporal properties of the DUESO, where it captured the UES opening and closure within a 100 msec tolerance in most of the swallows in the independent test set. This confirms that the high quality of DUESO detection can be carried over to completely unseen data and assures a high degree of generalization in the proposed system.

It is important to bear in mind that the accuracy of any physiological event detector cannot be judged only through comparison with human ratings which are subject to error too. The sub-events occurring during or after the detected event and their importance to the whole physiological process, control the limits to which the system can be considered accurate because one doesn't want to detect an event with 50 msec accuracy to look for another sub-event that happens within 10 msec of the original event. Previous studies have shown that the important UES events happen slightly after the initial UES opening [21]. For example, in general, entry of the bolus head into the sphincter defines UES opening; however, in 20% of swallows, air precedes entry of the bolus by 30-60 msec [21]. Maximal values of A-P UES diameter were found also to be reached after 70-170 msec of UES opening, depending on the bolus size and other factors [21]. So, it could be argued that a delayed detection of UES opening is not completely inaccurate if it happens within 100 msec (≈ 3 frames at 30 FPS) after the actual opening. Conversely, anatomic abnormalities leading to reduced DUESO (e.g. cricopharyngeal bar, Zenker diverticulum, hypopharyngeal lesions) would be completely undetectable without imaging leading to the need for further research to determine if HRCA can classify patterns of DUESO that indicate the need for imaging to rule out an anatomic diagnosis reducing DUESO.

In Summary, this study along with others, demonstrates advancements in HRCA signal processing and provides substantial evidence that HRCA signals predominantly reflect the patterns in DUESO and combined with our overall growing research portfolio, swallowing physiological activity. These advancements show the capability of HRCA to provide insight into diagnostic physiological aspects of swallow function and push towards the development of more accessible tools for dysphagia screening within clinical settings. Future research directions for this study include enhancing the detection quality of DUESO while reducing the error between the predicted and actual DUESO and investigating whether characteristic differences in HRCA signal signatures may reflect underlying anatomic or other etiologic explanations warranting investigation with imaging. This point is crucial in that some causes of dysphagia are indeed anatomically based, however in situations in which such diagnoses are suspected and imaging is not available immediately, HRCA certainly shows promise toward providing interim information that can guide management. Further, the scope of the study will be expanded to include the detection of maximal A-P UES diameter and its time of occurrence solely from HRCA signals.

V. Conclusion

In this paper, we proposed an ambitious deep architecture for the temporal identification of the DUESO during swallows by using HRCA signals. Swallows from 116 patients were collected under a standard clinical procedure for different swallowing tasks and materials. 3D acceleration signals of full length swallows, were denoised and fed into a network composed of a two-layer CNN, a 3-layer GRU-based RNN, and 3 fully connected layers to generate the temporal mask marking the time of UES opening and closure during swallows. The proposed system yielded an average accuracy of more than 90% of the swallow width and more than 91% of the DUESO width (sensitivity) with a low false positive rate. Moreover, the system showed nearly identical performance when used on an independent testing set from an ongoing clinical trial. Our results have provided substantial evidence that HRCA signals combined with a deep network architecture can be used to demarcate important physiological events that occur during swallowing.

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, 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.

References

[1] A. J. Miller, "The neurobiology of swallowing and dysphagia," *Developmental Disabilities Research Reviews*, vol. 14, no. 2, pp. 77–86, 2008.

[2] N. Bhattacharyya, "The prevalence of dysphagia among adults in the united states," *Otolaryngology–Head and Neck Surgery*, vol. 151, no. 5, pp. 765–769, 2014.

[3] J. Murray, *Manual of Dysphagia Assessment in Adults*. Cengage Learning, Inc, Oct 1998.

[4] C. Lazarus and J. A Logemann, "Swallowing disorders in closed head trauma patients," *Archives of physical medicine and rehabilitation*, vol. 68, pp. 79–84, 03 1987.

[5] I. J. Cook and P. J. Kahrilas, "Aga technical review on management of oropharyngeal dysphagia," *Gastroenterology*, vol. 116, no. 2, pp. 455 – 478, 1999.

[6] S. Lindgren and L. Janzon, "Prevalence of swallowing complaints and clinical findings among 50-79-year-old men and women in an urban population," *Dysphagia*, vol. 6, pp. 187–92, 02 1991.

[7] H. Siebens, E. Trupe, A. Siebens, F. Cook, S. Anshen, R. Hanauer, and G. Oster, "Correlates and consequences of eating dependency in institutionalized elderly," *Journal of the American Geriatrics Society*, vol. 34, no. 3, pp. 192–198, 1986.

[8] B. Martin-Harris and B. Jones, "The videofluorographic swallowing study," *Physical Medicine and Rehabilitation Clinics of North America*, vol. 19, no. 4, pp. 769 – 785, 2008, dysphagia.

[9] R. Ishida, J. B. Palmer, and K. M. Hiiemae, "Hyoid motion during swallowing: Factors affecting forward and upward displacement," *Dysphagia*, vol. 17, no. 4, pp. 262–272, Dec 2002.

[10] E. K. Plowman-Prine, C. M. Sapienza, M. S. Okun, S. L. Pollock, C. Jacobson, S. S. Wu, and J. C. Rosenbek, "The relationship between quality of life and swallowing in parkinson's disease," *Movement Disorders*, vol. 24, no. 9, pp. 1352–1358, 2009.

[11] N. Miller, E. Noble, D. Jones, and D. Burn, "Hard to swallow: dysphagia in Parkinson's disease," *Age and Ageing*, vol. 35, no. 6, pp. 614–618, 11 2006.

[12] E. S. Lun Chow, B. M. Hei Kong, M. T. Po Wong, B. Draper, K. L. Lin, S. K. Sabrina Ho, and C. Por Wong, "The prevalence of depressive symptoms among elderly chinese private nursing home residents in hong kong," *International Journal of Geriatric Psychiatry*, vol. 19, no. 8, pp. 734–740, 2004.

[13] A. Farri, A. Accornero, and C. Burdese, "Social importance of dysphagia: its impact on diagnosis and therapy," *ACTA Otorhinolaryngologica Italica*, vol. 27, no. 2, pp. 83–86, Apr 2007.

[14] B. Gustafsson and L. Tibbling, "Dysphagia, an unrecognized handicap," *Dysphagia*, vol. 6, no. 4, pp. 193–199, 1991.

[15] O. Ekberg, S. Hamdy, V. Woisard, A. Wuttge-Hannig, and P. Ortega, "Social and psychological burden of dysphagia: Its impact on diagnosis and treatment," *Dysphagia*, vol. 17, no. 2, pp. 139–146, Apr 2002.

[16] L. Perry, "Dysphagia: the management and detection of a disabling problem," *British Journal of Nursing*, vol. 10, no. 13, pp. 837–844, 2001.

[17] N. P. Nguyen, C. Frank, C. C. Moltz, P. Vos, H. J. Smith, U. Karlsson, S. Dutta, A. Midyett, J. Barloon, and S. Sallah, "Impact of dysphagia on quality of life after treatment of head-and-neck cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 61, no. 3, pp. 772 – 778, 2005.

[18] S. Singh and S. Hamdy, "The upper oesophageal sphincter," *Neurogastroenterology & Motility*, vol. 17, no. Suppl. 1, pp. 3–12, 2005.

[19] N. K. Ahuja and W. W. Chan, "Assessing upper esophageal sphincter function in clinical practice: a primer," *Current Gastroenterology Reports*, vol. 18, no. 2, p. 7, Jan 2016.

[20] P. Kahrilas, W. Dodds, J. Dent, J. Logemann, and R. Shaker, "Upper esophageal sphincter function during deglutition," *Gastroenterology*, vol. 95, no. 1, pp. 52 – 62, 1988.

[21] I. J. Cook, W. J. Dodds, R. O. Dantas, B. Massey, M. K. Kern, I. M. Lang, J. G. Brasseur, and W. J. Hogan, "Opening mechanisms of the human upper esophageal sphincter," *American Journal of Physiology-Gastrointestinal and Liver Physiology*, vol. 257, no. 5, pp. G748–G759, 1989.

[22] Y. Kim, T. Park, E. Oommen, and G. McCullough, "Upper esophageal sphincter opening during swallow in stroke survivors," *Am J Phys Med Rehabil*, vol. 94, no. 9, pp. 734–739, Sep 2015.

[23] A. L. Merati, "In-office evaluation of swallowing: FEES, pharyngeal squeeze maneuver, and FEESST," *Otolaryngol. Clin. North Am.*, vol. 46, no. 1, pp. 31–39, Feb 2013.

[24] S. G. Butler, L. Markley, B. Sanders, and A. Stuart, "Reliability of the penetration aspiration scale with flexible endoscopic evaluation of swallowing," *Ann. Otol. Rhinol. Laryngol.*, vol. 124, no. 6, pp. 480–483, Jun 2015.

[25] A. M. Kelly, M. J. Drinnan, and P. Leslie, "Assessing penetration and aspiration: how do videofluoroscopy and fiberoptic endoscopic evaluation of swallowing compare?" *Laryngoscope*, vol. 117, no. 10, pp. 1723–1727, Oct 2007.

[26] I. J. Cook, W. J. Dodds, R. O. Dantas, M. K. Kern, B. T. Massey, R. Shaker, and W. J. Hogan, "Timing of videofluoroscopic, manometric events, and bolus transit during the oral and pharyngeal phases of swallowing," *Dysphagia*, vol. 4, no. 1, pp. 8–15, 1989.

[27] A. Daggett, J. Logemann, A. Rademaker, and B. Pauloski, "Laryngeal penetration during deglutition in normal subjects of various ages," *Dysphagia*, vol. 21, no. 4, pp. 270–274, Oct 2006.

[28] Y. Kim, G. H. McCullough, and C. W. Asp, "Temporal measurements of pharyngeal swallowing in normal populations," *Dysphagia*, vol. 20, no. 4, pp. 290–296, 2005.

[29] R. D. Gross, C. W. Atwood, S. B. Ross, K. A. Eichhorn, J. W. Olszewski, and P. J. Doyle, "The coordination of breathing and swallowing in Parkinson's disease," *Dysphagia*, vol. 23, no. 2, pp. 136–145, Jun 2008.

[30] E. Sejdić, G. A. Malandraki, and J. L. Coyle, "Computational deglutition: Using signal- and image-processing methods to understand swallowing and associated disorders [life sciences]," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 138–146, Jan 2019.

[31] S. Mao, Z. Zhang, Y. Khalifa, C. Donohue, J. L. Coyle, and E. Sejdic, "Neck sensor-supported hyoid bone movement tracking during swallowing," *Royal Society Open Science*, vol. 6, no. 7, p. 181982, 2019.

[32] E. Sejdić, C. M. Steele, and T. Chau, "Segmentation of dual-axis swallowing accelerometry signals in healthy subjects with analysis of anthropometric effects on duration of swallowing activities," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1090–1097, 2009.

[33] S. Damouras, E. Sejdić, C. M. Steele, and T. Chau, "An online swallow detection algorithm based on the quadratic variation of dual-axis accelerometry," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3352–3359, June 2010.

[34] J. Lee, C. M. Steele, and T. Chau, "Time and time–frequency characterization of dual-axis swallowing accelerometry signals," *Physiological Measurement*, vol. 29, no. 9, p. 1105, 2008.

[35] Y. Khalifa, J. L. Coyle, and E. Sejdić, "Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings," 2019, under review.

[36] D. C. B. Zoratto, T. Chau, and C. M. Steele, "Hyolaryngeal excursion as the physiological source of swallowing accelerometry signals," *Physiological Measurement*, vol. 31, no. 6, p. 843, 2010.

[37] C. Rebrion, Z. Zhang, Y. Khalifa, M. Ramadan, A. Kurosu, J. L. Coyle, S. Perera, and E. Sejdić, "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*, vol. 7, pp. 1–9, 2019.

[38] Q. He, S. Perera, Y. Khalifa, Z. Zhang, A. S. Mahoney, A. Sabry, C. Donohue, J. L. Coyle, and E. Sejdić, "The association of high resolution cervical auscultation signal features with hyoid bone displacement during swallowing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 9, pp. 1810–1816, Sep. 2019.

[39] A. Kurosu, J. L. Coyle, J. M. Dudik, and E. Sejdić, "Detection of swallow kinematic events from acoustic high-resolution cervical auscultation signals in patients with stroke," *Archives of Physical Medicine and Rehabilitation*, 2018.

[40] Z. C. Lipton, J. Berkowitz, and C. Elkan, “A critical review of recurrent neural networks for sequence learning,” *arXiv preprint arXiv:1506.00019*, May 2015.

[41] J. L. Elman, “Finding structure in time,” *Cognitive Science*, vol. 14, no. 2, pp. 179 – 211, Apr. 1990.

[42] S. Chauhan and L. Vig, “Anomaly detection in ECG time signals via deep long short-term memory networks,” in *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics*, Oct. 2015, pp. 1–7.

[43] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, and M. Shah, “Deep learning human mind for automated visual classification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Jul. 2017, pp. 6809–6817.

[44] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, “Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation,” *arXiv preprint arXiv:1506.07452*, Jun. 2015.

[45] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[47] J. H. Tan, Y. Hagiwara, W. Pang, I. Lim, S. L. Oh, M. Adam, R. S. Tan, M. Chen, and U. R. Acharya, “Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals,” *Computers in Biology and Medicine*, vol. 94, pp. 19–26, Mar. 2018.

[48] Z. Xiong, M. P. Nash, E. Cheng, V. V. Fedorov, M. K. Stiles, and J. Zhao, “ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network,” *Physiological Measurement*, vol. 39, no. 9, p. 094006, Sep. 2018.

[49] S. Adavanne, A. Politis, J. Nikunen, and T. Virtanen, “Sound event localization and detection of overlapping sources using convolutional recurrent neural networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 34–48, Mar. 2019.

[50] H. S. Bonilha, J. Blair, B. Carnes, W. Huda, K. Humphries, K. McGrattan, Y. Michel, and B. Martin-Harris, “Preliminary investigation of the effect of pulse rate on judgments of swallowing impairment and treatment recommendations,” *Dysphagia*, vol. 28, no. 4, pp. 528–538, Dec 2013.

[51] J. A. Cichero and B. E. Murdoch, “The physiologic cause of swallowing sounds: answers from heart sounds and vocal tract acoustics,” *Dysphagia*, vol. 13, no. 1, pp. 39–52, 1998.

[52] J. M. Dudik, J. L. Coyle, and E. Sejdić, “Dysphagia Screening: Contributions of Cervical Auscultation Signals and Modern Signal-Processing Techniques,” *IEEE Trans Hum Mach Syst*, vol. 45, no. 4, pp. 465–477, Aug 2015.

[53] K. Takahashi, M. E. Groher, and K. Michi, “Methodology for detecting swallowing sounds,” *Dysphagia*, vol. 9, no. 1, pp. 54–62, 1994.

[54] J. A. Cichero and B. E. Murdoch, “Detection of swallowing sounds: methodology revisited,” *Dysphagia*, vol. 17, no. 1, pp. 40–49, 2002.

[55] J. Lee, E. Sejdić, C. M. Steele, and T. Chau, “Effects of liquid stimuli on dual-axis swallowing accelerometry signals in a healthy population,” *BioMedical Engineering OnLine*, vol. 9, no. 1, p. 7, Feb 2010.

[56] A. El-Jaroudi, M. S. Redfern, L. F. Chaparro, and J. M. Furman, “The application of time-frequency methods to the analysis of postural sway,” *Proceedings of the IEEE*, vol. 84, no. 9, pp. 1312–1318, 1996.

[57] G. L. Lof and J. Robbins, “Test-retest variability in normal swallowing,” *Dysphagia*, vol. 4, no. 4, pp. 236–242, Dec 1990.

[58] E. Sejdić, V. Komisar, C. M. Steele, and T. Chau, “Baseline characteristics of dual-axis cervical accelerometry signals,” *Annals of Biomedical Engineering*, vol. 38, no. 3, pp. 1048–1059, Mar 2010.

[59] J. M. Dudik, I. Jestrović, B. Luan, J. L. Coyle, and E. Sejdić, “A comparative analysis of swallowing accelerometry and sounds during saliva swallows,” *Biomedical engineering online*, vol. 14, no. 1, p. 3, 2015.

[60] J. M. Dudik, I. Jestrović, B. Luan, J. L. Coyle, and E. Sejdić, “Characteristics of dry chin-tuck swallowing vibrations and sounds,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 10, pp. 2456–2464, Oct 2015.

[61] L. Marple, “A new autoregressive spectrum analysis algorithm,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 441–454, 1980.

[62] E. Sejdić, C. M. Steele, and T. Chau, “A method for removal of low frequency components associated with head movements from dual-axis swallowing accelerometry signals,” *PLOS ONE*, vol. 7, no. 3, pp. 1–8, 03 2012.

[63] ——, “The effects of head movement on dual-axis cervical accelerometry signals,” *BMC Res Notes*, vol. 3, p. 269, Oct 2010.

[64] ——, “A procedure for denoising dual-axis swallowing accelerometry signals,” *Physiol Meas*, vol. 31, no. 1, pp. 1–9, Jan 2010.

[65] J. M. Dudik, A. Kurosu, J. L. Coyle, and E. Sejdić, “A statistical analysis of cervical auscultation signals from adults with unsafe airway protection,” *Journal of NeuroEngineering and Rehabilitation*, vol. 13, no. 1, p. 7, Jan 2016.

[66] S. P. Shashikumar, A. J. Shah, G. D. Clifford, and S. Nemati, “Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD ’18. London, United Kingdom: ACM, Jul. 2018, pp. 715–723.

[67] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[68] Y. Bengio, *Practical Recommendations for Gradient-Based Training of Deep Architectures*, ser. Lecture Notes in Computer Science. Springer, 2012, vol. 7700, ch. 26, pp. 437–478.

[69] A. L. Perlman, X. He, J. Barkmeier, and E. V. Leer, “Bolus location associated with videofluoroscopic and respirodeglutometric events,” *Journal of Speech, Language, and Hearing Research*, vol. 48, no. 1, pp. 21–33, 2005.

[70] S. Morinière, M. Boiron, D. Alison, P. Makris, and P. Beutter, “Origin of the sound components during pharyngeal swallowing in normal subjects,” *Dysphagia*, vol. 23, no. 3, pp. 267–273, Sep 2008.