High resolution cervical auscultation and data science: New tools to address an old problem.

James L. Coyle, Ph.D., CCC-SLP, BCS-S
Department of Communication Science and Disorders, School of Health and Rehabilitation Sciences
Department of Otolaryngology, School of Medicine
University of Pittsburgh, Pittsburgh, PA, USA

Ervin Sejdić, Ph.D.
Department of Electrical and Computer Engineering, Department of Bioengineering
Swanson School of Engineering
University of Pittsburgh, Pittsburgh PA, USA

Corresponding Author:

James L. Coyle, PhD
University of Pittsburgh
Department of Communication Science and Disorders
6035 Forbes Tower
Pittsburgh, PA 15260
412.383.6608
jcoyle@pitt.edu

Conflict of Interest Statement:
Financial: The authors disclose that they receive financial support under two grants from the National Institute of Child Health and Human Development (2R01HD074819-04, and 1R01HD092239-01), and the National Science Foundation (NSF Career Award – E. Sejdic; grant #1652203).
Non-financial: the authors disclose that they have no non-financial conflicts of interest.

Funding Statement: This work was supported by grants from the National Institutes of Health – National Institute of Child Health and Human Development: #2R01HD074819-04 (E. Sejdic, J. Coyle, co-PI's), and 1R01HD092239-01 (E. Sejdic, PI), and NSF Career Award (E. Sejdic) #1652203.

High resolution cervical auscultation and data science: New tools to address an old problem.

Abstract

High resolution cervical auscultation (HRCA) is an evolving clinical method for noninvasive screening of dysphagia that relies on data science, machine learning and wearable sensors to investigate of the characteristics of disordered swallowing function in people with dysphagia. HRCA has shown promising results in categorizing normal and disordered swallowing (i.e., screening) and independent of human input, identifying a variety of swallowing physiologic events as accurately as trained human judges. The system has been developed through a collaboration of data scientists, computer-electrical engineers and speech-language pathologists. Its potential to automate dysphagia screening and contribute to evaluation lies in its noninvasive nature (wearable electronic sensors) and its growing ability to accurately replicate human judgments of swallowing data typically formed on the basis of videofluoroscopic imaging data. Potential contributions of HRCA when VFSS may be unavailable, undesired, or not feasible for many patients in various settings, is discussed along with the development and capabilities of HRCA. The use of technological advances and wearable devices can extend the dysphagia clinician's reach and reinforce top-of-license practice for patients with swallowing disorders.

High resolution cervical auscultation and data science: New tools to address an old problem.

Introduction

Why does the use of devices for measuring swallowing function matter? For many years, human judgment of patient function was solely performed by empirical observation of the patient performing a target activity or task. In fact, human judgment has been the gold standard for describing numerous human functions for many decades. But, with the growth of technological advances in computer sciences and sensor technology have come opportunities to meld two areas of science to accomplish two common goals: 1. improving traditional screening, clinical assessment and treatment methods by including technology, and 2. developing individualized treatments designed to address the nuances of a specific patient's impairment patterns. The purpose of this manuscript is to 1) review the current and past use of cervical auscultation in assessing individuals with dysphagia, 2) describe the complex underpinnings of high resolution cervical auscultation (HRCA) and its application to dysphagia assessment, and 3) to describe a current, ongoing project that integrates collaborative HRCA advances in technology and clinical findings.

Limitations of cervical auscultation and rationale for HRCA

Cervical auscultation to observe swallowing function (CA) using ordinary stethoscopes has been a common clinical practice for many years by dysphagia clinicians. Up to one-fourth of dysphagia clinicians use CA in diagnostic and management activities (Bateman et al., 2007; Rumbach et al., 2018; Vogels et al., 2015). The use of CA was implemented following the observation that sounds emanate from the neck during swallowing, and that these sounds may reflect physiologic events occurring during swallowing (Borr et al., 2007). CA is based on the principle that a stethoscope can transmit all available acoustic information from the anterior neck during swallowing, and that a human observer can accurately interpret those sounds into a timeline of physiologic events. This also assumes the ability to form an impression as to the "normalness" of those events. This concept is germane to dysphagia clinical practice, given the longstanding interest in developing inexpensive and noninvasive methods of evaluating swallowing function. Support for CA was first described more than 20 years ago by Cichero & Murdoch (1998), in a theoretical paper in which a cardiac analogy theory was proposed. Briefly, this theory proposes that the upper aerodigestive tract is analogous to the heart. Both consist of several tubes and valves that open and close in a certain pattern, and pumps that squeeze and propel fluids during the cardiac cycle and during swallowing. Furthermore, the theory suggests that ordinary auscultation with a stethoscope, as is used in clinical evaluation of cardiac sounds, should translate to an equivalent interpretation of swallowing function that would be derived from an imaging study. Because of its convenience and low cost, interest in adding stethoscope-based observations has grown in the past 20 years, and many clinicians rely on cervical auscultation in diagnostic assessments, sometimes as a replacement for imaging. Several studies have reported data indicating that specific "sounds" occurring during swallowing represent discrete physiologic and kinematic events, and that these observations may be useful surrogates for videofluoroscopic (VF) imaging studies (Borr et al., 2007; Leslie et al., 2007; Zenner et al., 1995).

Initially, research regarding CA produced results indicating its ability to identify when a swallow occurred, but this quickly spawned research into the nature of those sounds. These studies described and named the sounds, often using a variety of labels (e.g., "lub," "dub," "first and second sound," "pre-click," "click," "swish," Greek alphabet characters, etc.) to reflect what seemed to be associated with swallowing events. These events were observed with concurrent imaging including opening and closing of laryngopharyngeal valves, ventilatory sounds, and bolus flow (Borr et al., 2007). Leslie and colleagues (2007) investigated CA by using an electronic microphone to standardize data acquisition during concurrent imaging studies of swallowing. They described the many inconsistencies in the assumptions underlying CA's utility. The study identified associations between some sounds and observed kinematic events, while also noting an astonishingly broad range of patterns of CA sounds during swallowing in healthy participants. The authors also demonstrated poor inter-judge agreement for CA while underscoring the conflict between the convenience benefits of stethoscope-based CA and its accuracy, cautioning readers that "there is no robust evidence cervical auscultation of swallowing sounds should be adopted in routine clinical practice…" (p. 296). Both studies relied on human interpretation of the sounds produced during the swallows. Regardless of the obvious limitations of the method, CA has persisted in clinical dysphagia work.

CA's limited value as an adjunct to dysphagia assessment lies in the stethoscope's inability to collect and transmit, the entire spectrum of acoustic and vibratory information emanating from the pharynx and larynx during swallowing (Nowak & Nowak, 2018), as well as the human auditory system's limitations in perceiving and interpreting in a standardized manner, the obtained sounds. Stethoscopes are designed for specific purposes and tuned for specific frequency ranges based on those purposes (e.g., heart sounds, ventilatory sounds; adults, children), and likewise, the range of human auditory acuity across independent judges varies widely. To illustrate the challenges presented by auscultation with stethoscopes, Favrat and colleagues (2004) investigated the accuracy of cardiologists, internists, family practitioners, and residents, in identifying cardiac sounds and generation of an accurate diagnosis based on chest auscultation. The expert practitioners were 69% accurate recognizing heart sounds, and correctly diagnosed 62% of the cases, while the residents were 40% and 24% accurate, respectively. This underscores the degree of observation and interpretation imprecision based on auscultation for an actual disorder for which stethoscopes were developed. Since there has been an explosion in the development of electronic data acquisition and analyses over the past 10-15 years, potential alternatives to stethoscope-based CA have received increased attention.

The growth of computerized signal processing capabilities and development of a variety of electronic sensors has delivered an opportunity to investigate the principles underlying CA using techniques that do not rely completely on human judgment, and to capitalize on advanced algorithm-based signal processing, machine learning, and artificial intelligence methods developed by our partners in related Engineering fields. Though other research groups have explored sensor based swallowing observation over the past several years using surface electromyography, piezoelectric sensors, and accelerometers, Sejdić and colleagues described the first steps toward development of a sensor-based HRCA system for use in dysphagia screening (Sejdic, Steele, et al., 2010).

High resolution cervical auscultation (HRCA) was described by Dudik and colleagues (2015) following three years of research that deployed a tri-axial accelerometer and high resolution microphone to accrue the signals. Preliminary studies examining the signal processing of swallowing accelerometry data indicated significant differences in signal features obtained during various bolus conditions and bolus head position during swallowing. In 2013, the authors of this manuscript embarked on a long-term NIH sponsored project that is ongoing, and the results of which have been published or are under analysis, submission, review, or revisions, as well as cited elsewhere in this manuscript. In this study, patients with suspected dysphagia underwent concurrent videofluoroscopy and HRCA signal acquisition. The goals of the study are to 1) develop an autonomous HRCA screening system and test its efficacy in the clinical setting, and 2) compare the accuracy of autonomous and semi-autonomous HRCA prediction of various commonly analyzed swallowing temporal and spatial measurements to gold standard human judgment and raise that accuracy to acceptable levels, in an effort to improve clinical workflow, and to provide a surrogate to VFSS when VFSS is not available, feasible, or desired by the patient. To date, the study methodology has involved the use of three signal sources (VFSS, tri-axial accelerometry, high-resolution microphone) collected simultaneously. Consented participants were comprised of patients referred for a VFSS due to suspected dysphagia. All participants were from an acute, tertiary care teaching hospital. From this cohort, approximately 4,000 imaged swallows were captured and stored. The authors (JC and ES) continue to collect the same type of data, using the same methodology, from a cohort of 200 healthy community dwelling adults. This collaborative clinical and engineering based endeavor permits the 1) development of an automated dysphagia screen while speeding clinical workflow of screening (e.g., nurse dysphagia screens) without compromising accuracy, 2) improvement of objectivity of judgments of swallowing function from imaging data, , and 3) capitalization on the advantages of advanced signal processing techniques within the dysphagia diagnostic process. To develop such a system, traditional human-mediated manual measurement methods of VFSS data measurement serve as the gold standard, and machine learning is deployed to more quickly produce accurate measurements that reflect the same judgments and measurements performed by the human judges.

<u>Current Project: Protocol</u>

To date, we have accrued data from 274 adult patients who were referred for VFSS at the University of Pittsburgh Medical Center campus hospitals, and from 80 healthy community dwelling, age-matched adults recruited from community registries. Patients were referred over the course of routine care due to confirmed or suspected dysphagia, and the examination procedures were controlled by the examining clinicians (i.e., SLP, Radiologist). Data accrual was performed by two SLP’s (VFSS) and two engineers (HRCA) during each examination. All procedures were approved by the Institutional Review Board at the University of Pittsburgh.

After providing informed consent, patients and healthy participants were prepared to undergo a VFSS (GE Ultimax system). Prior to initiation of the VFSS, two sensors were attached to the anterior neck. The tri-axial accelerometer (ADXL 327, Analog devices, Norwood, MA) was positioned at the anterior midline overlying the arch of the cricoid cartilage (based on palpation by the SLP investigators)The microphone (model C111L, AKG, Vienna, Austria) was placed approximately 1cm lateral (right) and inferior to the accelerometer to avoid interfering with the necessary VFSS imaging of the upper airway (Figure 1). For the patient data collection, bolus administration was dictated by the examining clinical speech-language pathologist (SLP) and no effort to modify the VFSS protocol was made by the research team. This ensured that the data set would be consistent with VFSS data obtained during typical conditions that occur during routine clinical VFSS. Patients swallowed varying numbers of boluses of multiple standardized textures and volumes of contrast (Varibar products, Bracco Diagnostics, Monroe Township, N.J.) in a neutral head position, as well as in various postural modifications based on clinician intervention efficacy trial needs. Continuous, written logging by investigators during all data accrual ensured specification of bolus conditions. For the healthy participants (age 18 – 92), a standard research protocol of 10 swallows per participant was followed to minimize x-ray exposure durations (average fluoro time = 0.66 minutes per examination). We also sought to accrue as much data from healthy participants as possible that would align with data accrued from patients to enable a sufficiently robust sample size for the machine learning components of the research. Healthy participants were administered 10 boluses each in the neutral head position. Trials were comprised of the following: 1) five 3mL thin liquid (Varibar Thin, Bracco) boluses, administered by the research SLP from a spoon with a swallow command used to prompt swallows and 2) five unmeasured, self-selected volume boluses of thin liquid, self-administered by participants from a cup without verbal or other prompts to swallow. These bolus size conditions were included in order to capture swallowing under both controlled and natural swallowing conditions, which have been shown to produce different temporal activity during swallowing (Nagy et al., 2013). The rationale for inclusion of a 3ml bolus condition was that this was the most common bolus condition to challenge the patient participants. The order of presentation of the 10 boluses was randomized for each healthy participant.

Fluoroscopy was performed at a pulse rate of 30PPS and images were accrued to a frame grabber card at 60 FPS and later down-sampled to 30FPS to eliminate duplicate frames (Bonilha et al., 2013; Oppenheim & Schafer, 2014). Simultaneously, acoustic and accelerometric signals were accrued directly to a hard drive, time linked to corresponding VFSS imaging data. The sensor placement is illustrated in figure 1, and the details of signal acquisition methods and hardware/software used are described by Dudik and colleagues (2018), as well as in other publications by this research group.

Insert Figure 1 here

<u>Fundamentals of HRCA</u>

The overall aim in developing HRCA is to produce a system that is capable of independently performing some temporal, spatial and kinematic measurements that are traditionally performed by clinicians. After establishing HRCA's accuracy in screening (Dudik, Coyle, et al., 2015), machine learning algorithms are deployed in order to test HRCA's ability to accurately perform some temporal and spatial measurements as accurately as trained human judges. Machine learning is an iterative process by which gold-standard data are first generated (e.g. human temporal and spatial measurements), after which some of that data is used to train computer algorithms to accurately produce acceptably similar judgments as the human judges, and the rest of the data, which is novel to the algorithms, is used to test their accuracy. Training is a computationally expensive but necessary process required to enable algorithms to detect characteristics of signal features that correspond to human-identified temporal or spatial events. As we accrue more data, the training sets grow, resulting in increased precision across an expanding range of conditions and extraneous confounds.

HRCA Data Acquisition

Several commonly used parameters were selected to characterize swallowing impairments. These parameters have been widely reported in the literature over the years. . The general scheme of HRCA data acquisition and analysis is illustrated in Figure 2. All swallow videos were segmented to identify the swallow segments that would be entered into the machine learning processes by trained human judges using image processing software (ImageJ, NIH). Temporal and spatial event measurements were performed based on the methods of others (Lof & Robbins, 1990) to ensure compatibility of measures with historical, published data. Data were recorded manually into spreadsheets and through customized Matlab modules during measurement. All judges underwent standardized training in each measure they were to perform, and their inter- and intra-rater reliability was tested prior to online analysis of study data. All judges returned high inter- and intra-rater reliability (e.g., 80% exact agreement within three frames (0.1 seconds) (Lof & Robbins, 1990) for frame selection during temporal analyses, and excellent intraclass correlation coefficients of 0.90 or greater for pixel-based spatial measures) for each measure. These criteria were also applied during data analyses using to eliminate judgment drift during ongoing measurement/judgment. Events and scores from images that have been coded include categorical measurements (e.g., scores on the penetration aspiration scale (Rosenbek et al., 1996) and measurements of vallecular and pyriform sinus residue using the normalized residue ratio scale (Pearson et al., 2013)). Temporal measurements relying on frame selection include: the video frames indicating first entry of bolus into the pharynx (bolus crosses ramus of mandible) and completion of bolus clearance through the UES (segment duration), onset of hyoid displacement, frame of maximal hyoid displacement, hyoid return to lowest position at the end of the swallow (duration of hyoid displacement), onset and offset of UES opening, and onset and offset of laryngeal closure,. Specific measurement methods for performing temporal measures of VFSS images have been described by Kurosu et al. (2019). Spatial, pixel-based measurements include the position of the hyoid body on each frame (hyoid kinematics), the diameter of the UES at maximal distension, and the position and area of the bolus and its components on each video frame. This latter measurement is being performed in ongoing efforts to develop algorithms to identify and quantify the proportion of boluses that enter the esophagus and that are retained in pharyngeal recesses or that enter the airway. After processing the signals, the VFSS-derived data are entered into the machine learning process to train algorithms.

Insert figure 2 here.

HRCA Data Processing: Pre-processing Deglutition Signals

It is critical to understand the basic data science and engineering definitions used in signal processing. A signal typically represents a quantity recorded via various instruments that represents changes in values. In statistics, signals are typically referred as time series, but in engineering, these recordings as referred as signals as they typically represent a measurable physical quantity. Importantly, signal artifacts must be considered during signal processing. The two artifacts discussed here are related to noise and disturbances.

Signal noise represents physical quantities that contaminate information present in these signals. In many cases, it is assumed that it stems from a random process (e.g., white Gaussian noise), while disturbances also represent signal contaminants that are not stemming from a random process (e.g., coughing, breathing sounds). There is also a major difference between noise and disturbances. Noise typically occupies all frequencies captured by signals, while disturbances are based in specific frequency bands. Sounds and vibrations represents vibration signals that are acquired by microphones and accelerometers, respectively.

Swallowing-related signals such as HRCA signals (i.e., swallowing vibrations or swallowing sounds) or surface electromyography signals are typically contaminated with various disturbances and noise (Dudik, Coyle, et al., 2015). Noise typically originates in electronic equipment used to acquire these signals or elsewhere in the immediate vicinity of data collection, while signal disturbances are caused by physiologic events that occur during the swallowing event (e.g., displacement of structures, bolus flow, breathing, head motions, vasomotion of major arteries). All these additional and simultaneously occurring signal components “contaminate” the targeted swallowing-related signal components and make any subsequent analysis difficult to carry out. This is because it is difficult to understand whether trends observed in the raw data are due to swallowing or due to disturbances and/or noise, or the combination of both. Hence, the first priority is to preprocess these swallowing signals, and remove as much as possible of the contaminating signal components (Sejdic et al., 2019). Steps in the preprocessing and feature extraction of HRCA signals are also illustrated in figure 2.

HRCA Data Processing: Data Reduction

The first task in the signal processing method is to remove any confounding effects of the data acquisition system via a process called whitening (Sejdic, Komisar, et al., 2010). Here, the idea is to develop filters mimicking the frequency behavior of the data acquisition system, and the inverses of these filters are then applied to acquired data to remove any contaminating effects of the data acquisition system. Next, noise needs to be removed from the deglutition signals, and this is typically achieved via a process called denoising (Sejdic, Steele, et al., 2010). Most efficient denoising algorithms are based on wavelets which are state-of-the-art mathematical functions that divide the signal data into components based on their frequency range, to enable each component to be analyzed using a scale that is matched to its resolution (Graps, 1995). Once whitening and denoising steps are completed, one would carry out any normalization steps (e.g., amplitude normalization), and lastly signal segmentation is completed.

Segmentation is the process of identifying the components of the recorded data that represent the event of interest (i.e., a swallow event) and separating the segment from pre- and post-swallow recorded events. For any automated method of segmentation to succeed, a segmentation gold standard must be used to provide the criterion for the onset and offset of any individual swallow, in order to enable comparison of the signal-derived predictions to the actual event duration, ensure face validity of the electronic measurement predictions, and to facilitate machine learning procedures which with multiple iterations of cross validation increase the efficiency and accuracy of the algorithms. Segmentation involves human frame-by-frame viewing and selection of the video frame in which the bolus head enters the pharynx (crosses the plane of the shadow of the mandible), and the frame in which the bolus tail clears through the UES, by trained dysphagia researchers in the swallowing research lab. These results are used to train the algorithms to detect the duration of the swallow.

A number of different algorithms have been proposed over the years to segment swallowing signals into individual swallows (Damouras et al., 2010; Dudik, Kurosu, et al., 2015; Sejdic et al., 2009).The main reason for the variety of algorithms is that this is one of the crucial steps in the analysis of signals, since incorrectly identifying a swallowing segment will obviously skew any subsequent analysis steps.

HRCA Data Processing: Feature Extraction

Once swallowing signals are segmented into individual swallows, signal features are identified and extracted. Most of the current literature considers features in various mathematical domains such as the time domain, frequency domain or the time-frequency domain. Features of segmented swallow signals range in complexity between those that are more common (e.g., standard deviations of these swallowing signals), to more advanced features, such as the entropy rate of these signals, denoting the amount of randomness in these signals. Extracted features can be then used to form various statistical models to examine dependence between independent variables, in this case signal features, and various dependent variables, such as penetration-aspiration scores, hyoid bone displacements in the anterior, posterior, superior and inferior directions (Dudik et al., 2016; Dudik, Kurosu, et al., 2018; Kurosu et al., 2019; Movahedi, Kurosu, Coyle, Perera, & Sejdić, 2017; Rebrion et al., 2019).

On the surface, signal features based on mathematical domains do not appear germane to analysis of clinical data traditionally obtained solely through imaging methods and analyzed by human judges. They are highly relevant from a computational point of view, because extracting features that are directly related to various physiological events that occur during swallowing is of particular relevance to clinicians. However, extracting physiologically identifiable features from swallowing signals requires the use of modern data analytics tools, such as machine learning, which will be described next. Moreover, human judges cannot perceive, nor can their judgment account for, many features of movement-related signals. That is, there are numerous components embedded within signals and images generated during a swallowing VFSS, that a human judge is not capable of identifying and/or discriminating.

<u>HRCA and Machine Learning: Fundamentals</u>

Machine learning is the study of algorithms and various statistical models that can be used to infer about specific patterns in a data set, in a supervised or unsupervised manner. While this scientific discipline has been around for more than 50 years, it has gained much more attention in recent years due to the advances in available computational resources that make the use of these computationally intensive algorithms to solve various problems possible.

Most machine learning algorithms rely on two phases: training and testing phases. During the training phase, one provides data to these algorithms to enable the algorithms to compute and infer about patterns in the dataset, much like the process of inference. The training data from the VFSS images which have been labelled by human judges, (i.e., each data point is labelled as belonging to one of the classes present in the dataset). These classes represent the VFSS measurement parameters described earlier. The training phase typically continues until training conditions, such as the accuracy of the algorithms in identifying human-identified events above a certain a priori percentage criterion, are met. Once the machine learning algorithm achieves desired performance on the training set, the algorithm is then applied to a testing set, (i.e., novel data to which the algorithms have not previously been exposed). The performance metrics such as sensitivity, specificity or recall are then reported. It is important here to clarify that training and testing data need to be separate. In other words, we cannot use the same data points for training and testing phases. In an ideal situation, the training phase is conducted using a dataset that was initially collected specifically for the purpose of training the machine learning algorithm, while the testing phase is conducted on a completely new dataset collected specifically for testing the accuracy of the proposed/used algorithm.

Unfortunately, this is not always possible, especially, in ordinary and often chaotic clinical settings due to a number of different issues such as funding, availability of staff, insufficient numbers of exemplars of the events of interest (e.g., swallows), and other constraints of clinical setting. In these cases, one can use a process called cross-validation in which the available data is randomly split into training and testing data, and the training phase is then completed only using the training data, and the testing phase is completed only using the testing data. This method of developing training and testing data sets from a large mass of clinically-derived data increases the external validity of the resultant algorithms and systems because all factors present in clinical testing environments that are mitigated in controlled studies are present during ordinary data collection and therefore, are components of the data sets.

Clinical Application of Machine Learning

While machine learning algorithms are much more complicated to use and more computationally intensive than other algorithms, they enable us to achieve various tasks that otherwise would be impossible to achieve by humans or other algorithms. For example, machine learning algorithms have been successfully applied in classifying swallowing signals to identify and differentiate swallows exhibiting no aspiration and those with aspiration with a very high accuracy (Celeste et al., 2012; Sejdić et al., 2013). Certainly the ability to noninvasively and continuously monitor and identify adequate from inadequate airway protection during swallowing has clinical applications, but efforts to extend machine learning of HRCA signals to determine the potential diagnostic utility of the system has begun to demonstrate compelling results. For instance, it was recently demonstrated that a combination of machine learning techniques, using non-invasive HRCA acceleration signals, can track the movement of the hyoid bone solely from the HRCA signals with a similar accuracy as trained human judges performing measurements using VFSS images (Mao et al., 2019). This study represents seminal work as it offers an alternative and widely available method for online hyoid bone movement tracking without any radiation risks and provides a pronounced and flexible approach for identifying clinically useful characteristics of dysphagia.

Machine learning has other potential applications that may also increase the speed of interpretation of VFSS imaging data by the clinician. Zhang, Coyle, & Sejdić (2018) recently sought to determine whether machine learning techniques could be used as a surrogate to manual spatial analysis to detect structural features of VFSS data from the video images themselves, demonstrating that unsupervised (i.e., without human input) advanced machine learning algorithms can identify the location of at least half of the body of the hyoid bone, at any point in time of a VFSS sequence. The height of the human hyoid body ranges from 0.6 – 1.2cm (across adult males and females) (Loth et al., 2015). We produced square bounding boxes surrounding the hyoid body on every VFSS frame based on the human judges' frame-by-frame plotting annotations. Through machine learning, a second bounding box denoting the predicted location of the human-determined hyoid body bounding box was generated by the algorithms. The HRCA-generated bounding boxes exhibited >50% overlap with the human-measurement generated bounding boxes 89% of the time continuously throughout the swallow sequences. We acknowledge that routinely 50% does not sound like a very good value; however, given the small dimensions of the hyoid body, accurately locating >50% of a 6-10 mm object is a reasonable preliminary result which we are refining with additional machine learning.

A benefit to this result is a reduction in the time required to analyze this date from15-20 minutes per swallow required by a human judge to annotate the two hyoid body landmarks on each frame of the swallow, to less than 30 seconds per swallow.

Other findings that we have published have demonstrated that HRCA signals combined with signal processing and machine learning techniques can detect a variety of swallow kinematic events with similar accuracy to trained human judges and can differentiate between safe (scores of 1, 2) and unsafe swallows (scores of 3 – 8), as determined by the penetration-aspiration scale, with a high degree of accuracy (Dudik, Coyle, et al., 2018; Dudik, Coyle, et al., 2015; Dudik, Jestrović, et al., 2015; Dudik, Kurosu, et al., 2015; Jestrović et al., 2013; Movahedi, Kurosu, Coyle, Perera, & Sejdic, 2017; Sejdić et al., 2013). We have examined the association between HRCA signals and component scores of various swallow kinematic events from the Modified Barium Swallow Impairment Profile (MBSImP) (Martin-Harris et al., 2008) and found strong associations between HRCA signals and anterior hyoid bone movement (component #9), pharyngoesophageal segment opening (component #14), and pharyngeal residue (component #16) (Donohue et al., 2019; Donohue et al., 2018; Sabry et al., 2019). We have also found a strong association between HRCA signal features and hyoid bone displacement (He et al., 2019; Rebrion et al., 2019; Zhang et al., 2018).

<u>Conclusions and future directions</u>

Incorporation of technology into everyday life is a common practice. Our smart devices, automobiles, and numerous other ordinary and common tools continue to demonstrate that developments in electrical and computer engineering can positively impact ordinary human activities. Likewise, wearable, personalized machine-learning based technologies that provide real-time monitoring of ordinary activities and health conditions (e.g., smart watches, continuous glucose monitoring systems, wearable sweat sensors for endurance athletes) and assist with daily clinical work (e.g., dictation-transcription software) are contributing real-time information that can improve the accuracy and depth of health information needed to provide screening, diagnostic and treatment data to individuals and clinicians in health care settings. Many of these technologies produce similar results as a human judge but significantly more quickly, and many expand clinician capabilities beyond the limits of human judgment.

In the same way that we strive to change the disordered physiology of swallowing in our patients through our observations, developments in advanced signal processing and machine learning in a variety of contexts enrich our observations. These advances show promise in augmenting our ability to not only perform services and procedures more efficiently, but also to perform them with greater depth of inference. But, adoption of new technologies is often met with skepticism. During development of our HRCA system and methods, and after collecting a few hundred samples of acoustic data obtained using

HRCA high resolution microphones, we played these audio files to dysphagia experts with experience in the use of stethoscope-based cervical auscultation. Their response was almost universally “that’s not what swallows sound like.” The sensors had obtained broader spectral and frequency ranges than are possible with a stethoscope. This disbelief is likely rooted in the assumption that the human auditory system has complete receptive and processing capabilities, and that there is no additional information in the acoustic signals because “we can’t hear it.” It will take time for many technological developments to be accepted in mainstream clinical work, and for medicine to embrace the contributions of these new and relatively unfamiliar fields of science are to our own profession and clinical practice, and to fully develop their potential. We are embarking on a clinical trial of our HRCA system to assess its screening effectiveness, in an effort to extend screening beyond the acute care setting. Likewise, we continue testing HRCA’s accuracy in predicting a variety of temporal and spatial measurements in an effort to strengthen clinicians’ impact on patient care. Automated signal-processing based measurements can help shift clinician resources toward actual intervention by reducing some of the tedium of manual measurements that consume so much of the clinical process while increasing their depth.

Numerous devices and systems are under development that capitalize on advances in other areas of science that carry the potential of extending the reach of clinicians. Our own HRCA research is developing results with the hope that such a system that can (in the future) noninvasively analyze some aspects of deglutition on a swallow-by-swallow basis in real time. This could be done either with imaging to expedite measurements and interpretations, or without the use of imaging when it is unavailable, to identify swallowing disorders and impairments, and potentially inform the clinician regarding intervention options when traditional information (e.g., imaging data) is not available. This will broaden the clinician’s capacity to interpret more information more efficiently while extending deployment of the scope of practice to patients who a) have no access to imaging centers for economic or other logistical reasons, b) do not want imaging studies, c) do not have immediate or any access to imaging studies (e.g., underserved regions), , and d) who are physically unable to undergo imaging tests. Moreover, such developments are promising in that they enable clinicians to produce top-of-license practice patterns more efficiently and with comparable accuracy. Collaborations between dysphagia researchers and clinicians, computer and electrical engineers, and many other disciplines, represent the future of development of personalized methods to improve the screening, diagnosis and treatment/management of people with dysphagia.

Acknowledgements

This work is supported by grants from the National Institutes of Health (Eunice Kennedy Shriver Institute for Child Health and Human Development, grant numbers 1R01HD092239-01 and 2R01HD074819-04), and from the National Science Foundation (career award number 1652203). We appreciate our collaborations with the participant registry of the Claude D. Pepper Older Americans Independence Center, and the Pitt+Me Registry. The authors acknowledge and appreciate the contributions of the participants in the described research, the Department of Radiology of the University of Pittsburgh Medical Center, and those of our doctoral, graduate and undergraduate student research associates in the execution of this ongoing work. The authors acknowledge and appreciate the assistance of PhD students Erin Lucatorto and Cara Donohue in the preparation of this manuscript.

References

Bateman, C., Leslie, P., & Drinnan, M. J. (2007). Adult dysphagia assessment in the UK and Ireland: are SLTs assessing the same factors? *Dysphagia, 22*(3), 174-186. https://doi.org/10.1007/s00455-006-9070-3

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. https://doi.org/10.1007/s00455-013-9463-z

Borr, C., Hielscher-Fastabend, M., & Lücking, A. (2007). Reliability and validity of cervical auscultation. *Dysphagia, 22*(3), 225-234. https://doi.org/10.1007/s00455-007-9078-3

Celeste, M., Azadeh, K., Sejdić, E., Berall, G., & Chau, T. (2012). Quantitative classification of pediatric swallowing through accelerometry. *J Neuroeng Rehabil, 9*(1), 34. https://doi.org/10.1186/1743-0003-9-34

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. https://doi.org/10.1007/pl00009548

Damouras, S., Sejdic, E., Steele, C. M., & Chau, T. (2010). An online swallow detection algorithm based on the quadratic variation of dual-axis accelerometry. *IEEE Transactions on Signal Processing, 58*(6), 3352-3359. https://doi.org/10.1109/TSP.2010.2043972

Donohue, C., Khalifa, Y., Sejdic, E., & Coyle, J. L. (2019). How closely do machine ratings of duration of UES opening during videofluoroscopy approximate clinician ratings using kinematic analysis and the MBSImP? Dysphagia Research Society Annual Scientific Meeting, San Diego, CA.

Donohue, C., Zhang, Z., Mahoney, A., Perera, S., Sejdic, E., & Coyle, J. L. (2018). Do machine ratings of hyoid bone displacement during videofluoroscopy match clinician ratings using the MBSImP? American Speech Langage Hearing Association Convention, Orlando, FL.

Dudik, J. M., Coyle, J. L., El-Jaroudi, A., Mao, Z.-H., Sun, M., & Sejdić, E. (2018). Deep learning for classification of normal swallows in adults. *Neurocomputing, 285*, 1-9. https://doi.org/https://doi.org/10.1016/j.neucom.2017.12.059

Dudik, J. M., Coyle, J. L., & Sejdic, E. (2015). Dysphagia screening: Contributions of cervical auscultation signals and modern signal-processing techniques. *IEEE Transactions on Human Machine Systems, 45*(4), 465-477. https://doi.org/10.1109/thms.2015.2408615

Dudik, J. M., Jestrović, I., Luan, B., Coyle, J. L., & Sejdic, E. (2015). A comparative analysis of swallowing accelerometry and sounds during saliva swallows. *Biomed Eng Online, 14*, 3. https://doi.org/10.1186/1475-925x-14-3

Dudik, J. M., Kurosu, A., Coyle, J. L., & Sejdic, E. (2015). A comparative analysis of DBSCAN, K-means, and quadratic variation algorithms for automatic identification of swallows from swallowing accelerometry signals. *Computers in Biology and Medicine, 59*, 10-18. https://doi.org/10.1016/j.compbiomed.2015.01.007

Dudik, J. M., Kurosu, A., Coyle, J. L., & Sejdić, E. (2016). A statistical analysis of cervical auscultation signals from adults with unsafe airway protection. *Journal of NeuroEngineering and Rehabilitation, 13*(1), 7. https://doi.org/10.1186/s12984-015-0110-9

Dudik, J. M., Kurosu, A., Coyle, J. L., & Sejdić, E. (2018). Dysphagia and its effects on swallowing sounds and vibrations in adults. *Biomedical Engineering Online, 17*(1), 69. https://doi.org/10.1186/s12938-018-0501-9

Favrat, B., Pecoud, A., & Jaussi, A. (2004). Teaching cardiac auscultation to trainees in internal medicine and family practice: does it work? *BMC Med Educ, 4*, 5. https://doi.org/10.1186/1472-6920-4-5

Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science and Engineering, 2*(2), 50-61. https://doi.org/10.1109/99.388960

He, Q., Perera, S., Khalifa, Y., Zhang, Z., Mahoney, A. S., Sabry, A., Donohue, C., Coyle, J. L., & Sejdić, 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. https://doi.org/10.1109/TNSRE.2019.2935302

Jestrović, I., Dudik, J. M., Luan, B., Coyle, J. L., & Sejdic, E. (2013). Baseline characteristics of cervical auscultation signals during various head maneuvers. *Comput Biol Med, 43*(12), 2014-2020. https://doi.org/10.1016/j.compbiomed.2013.10.005

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. https://doi.org/10.1016/j.apmr.2018.05.038

Leslie, P., Drinnan, M., Zammit-Maempel, I., Coyle, J., Ford, G., & Wilson, J. A. (2007). Cervical auscultation synchronized with images from endoscopy swallow evaluations. *Dysphagia, 22*(4), 290-298. https://doi.org/10.1007/s00455-007-9084-5

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

Loth, A., Corny, J., Santini, L., Dahan, L., Dessi, P., Adalian, P., & Fakhry, N. (2015). Analysis of hyoid-larynx complex using 3D geometric morphometrics. *Dysphagia, 30*(3), 357-364. https://doi.org/10.1007/s00455-015-9609-2

Mao, S., Zhang, Z., Khalifa, Y., Donohue, C., Coyle, J. L., & Sejdic, E. (2019). Neck sensor-supported hyoid bone movement tracking during swallowing. *Royal Society Open Science, 6*(7), 181982. https://doi.org/10.1098/rsos.181982

Martin-Harris, B., Brodsky, M. B., Michel, Y., Castell, D. O., Schleicher, M., Sandidge, J., Maxwell, R., & Blair, J. (2008). MBS measurement tool for swallow impairment—MBSImp: Establishing a standard. *Dysphagia, 23*(4), 392-405. https://doi.org/10.1007/s00455-008-9185-9

Movahedi, F., Kurosu, A., Coyle, J. L., Perera, S., & Sejdic, E. (2017). A comparison between swallowing sounds and vibrations in patients with dysphagia. *Comput Methods Programs Biomed, 144*, 179-187. https://doi.org/10.1016/j.cmpb.2017.03.009

Movahedi, F., Kurosu, A., Coyle, J. L., Perera, S., & Sejdić, E. (2017). Anatomical directional dissimilarities in tri-axial swallowing accelerometry signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25*(5), 447-458. https://doi.org/10.1109/TNSRE.2016.2577882

Nagy, A., Leigh, C., Hori, S. F., Molfenter, S. M., Shariff, T., & Steele, C. M. (2013). Timing differences between cued and noncued swallows in healthy young adults. *Dysphagia, 28*(3), 428-434. https://doi.org/10.1007/s00455-013-9456-y

Nowak, L. J., & Nowak, K. M. (2018). Sound differences between electronic and acoustic stethoscopes. *Biomedical Engineering Online, 17*(1), 104-104. https://doi.org/10.1186/s12938-018-0540-2

Oppenheim, A. V., & Schafer, R. W. (2014). *Discrete-Time Signal Processing*. Pearson.

Pearson, W. G., Jr., Molfenter, S. M., Smith, Z. M., & Steele, C. M. (2013). Image-based measurement of post-swallow residue: the normalized residue ratio scale. *Dysphagia, 28*(2), 167-177. https://doi.org/10.1007/s00455-012-9426-9

Rebrion, C., Zhang, Z., Khalifa, Y., Ramadan, M., Kurosu, A., Coyle, J. L., Perera, S., & Sejdić, E. (2019). High resolution cervical auscultation signal features reflect vertical and horizontal displacement of the hyoid bone during swallowing. *Journal of Translational Engineering in Health and Medicine, 7*(1), 1-9. https://doi.org/10.1109/JTEHM.2018.2881468

Rosenbek, J. C., Robbins, J. A., Roecker, E. B., Coyle, J. L., & Wood, J. L. (1996). A penetration-aspiration scale. *Dysphagia, 11*(2), 93-98.

Rumbach, A., Coombes, C., & Doeltgen, S. (2018). A survey of Australian dysphagia practice patterns. *Dysphagia, 33*(2), 216-226. https://doi.org/10.1007/s00455-017-9849-4

Sabry, A., Mahoney, A., Perera, S., Sejdic, E., & Coyle, J. L. (2019). Are HRCA signal features associated with clinical ratings of pharyngeal residue using the MBSImP? Dysphagia Research Society Annual Scientific Meeting, San Diego, CA.

Sejdic, E., Komisar, V., Steele, C. M., & Chau, T. (2010). Baseline characteristics of dual-axis cervical accelerometry signals. *Ann Biomed Eng, 38*(3), 1048-1059. https://doi.org/10.1007/s10439-009-9874-z

Sejdic, E., Malandraki, G. A., & Coyle, J. L. (2019). Computational deglutition: Using signal- and image-processing methods to understand swallowing and associated disorders. *IEEE Signal Processing Magazine [Life Sciences], 36*(1), 138-146. https://doi.org/10.1109/MSP.2018.2875863

Sejdic, E., Steele, C. M., & Chau, T. (2009). 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, 56*(4), 1090-1097. https://doi.org/10.1109/TBME.2008.2010504

Sejdic, E., Steele, C. M., & Chau, T. (2010). A procedure for denoising dual-axis swallowing accelerometry signals. *Physiological Measurement, 31*(1), N1-9. https://doi.org/10.1088/0967-3334/31/1/n01

Sejdić, E., Steele, C. M., & Chau, T. (2013). Classification of penetration-aspiration versus healthy swallows using dual-axis swallowing accelerometry signals in dysphagic subjects. *IEEE Transactions on Biomedical Engineering, 60*(7), 1859-1866. https://doi.org/10.1109/TBME.2013.2243730

Vogels, B., Cartwright, J., & Cocks, N. (2015). The bedside assessment practices of speech-language pathologists in adult dysphagia. *International Journal of Speech Language Pathology, 17*(4), 390-400. https://doi.org/10.3109/17549507.2014.979877

Zenner, P. M., Losinski, D. S., & Mills, R. H. (1995). Using cervical auscultation in the clinical dysphagia examination in long-term care. *Dysphagia, 10*(1), 27-31. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=7859529

Zhang, Z., Coyle, J. L., & Sejdić, E. (2018). Automatic hyoid bone detection in fluoroscopic images using deep learning. *Scientific Reports, 8*(1), 12310. https://doi.org/10.1038/s41598-018-30182-6

Figure legends

Figure 1 Legend. The sensors on a videofluoroscopic image. Adapted from 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.

Figure 2 legend. Typical setup of HRCA data acquisition and signal processing (top), and examples of acoustic (left) and vibratory (three axes) signals accrued during a sample swallow. Adapted from Sejdic, E., Malandraki, G. A., & Coyle, J. L. (2019). Computational deglutition: Using signal- and image-processing methods to understand swallowing and associated disorders. *IEEE Signal Processing Magazine [Life Sciences], 36*(1), 138-146. Open access.

Figure 3 legend. Steps in the preprocessing (above) and feature extraction (bottom) of the signals from each axis of the triaxial accelerometer (A-P = anterior-posterior axis, S-I = superior-inferior axis, M-L = medial-lateral axis). Adapted from Movahedi, F., Kurosu, A., Coyle, J. L., Perera, S., & Sejdić, E. (2017). Anatomical directional dissimilarities in tri-axial swallowing accelerometry signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25*(5), 447-458. Open access.