
Megan K. O’Brien, a, b Olivia K. Botonis, a Elissa Larkin, c Julia Carpenter, c Bonnie Martin-Harris, d Rachel Maronati, a KunHyuck Lee, e Leora R. Cherney, b, c, d Brianna Hutchison, f Shuai Xu, g John A. Rogers, g Arun Jayaraman, a, b

aMax Nader Lab for Rehabilitation Technologies and Outcomes Research, Shirley Ryan AbilityLab, Chicago, IL, USA; bDepartment of Physical Medicine and Rehabilitation, Northwestern University, Chicago, IL, USA; cThink and Speak Lab, Shirley Ryan AbilityLab, Chicago, IL, USA; dDepartment of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA; eImpriMed, Palo Alto, CA, USA; fDepartment of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA; gDepartments of Materials Science and Engineering, Center for Bio-Integrated Electronics, Biomedical Engineering, Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA

Keywords
Clinical decision support · Machine learning · Monitoring · Precision medicine · Wearable sensors

Abstract
Introduction: Difficulty swallowing (dysphagia) occurs frequently in patients with neurological disorders and can lead to aspiration, choking, and malnutrition. Dysphagia is typically diagnosed using costly, invasive imaging procedures or subjective, qualitative bedside examinations. Wearable sensors are a promising alternative to noninvasively and objectively measure physiological signals relevant to swallowing. An ongoing challenge with this approach is consolidating these complex signals into sensitive, clinically meaningful metrics of swallowing performance. To address this gap, we propose 2 novel, digital monitoring tools to evaluate swallows using wearable sensor data and machine learning.

Methods: Biometric swallowing and respiration signals from wearable, mechano-acoustic sensors were compared between patients with poststroke dysphagia and nondysphagic controls while swallowing foods and liquids of different consistencies, in accordance with the Mann Assessment of Swallowing Ability (MASA). Two machine learning approaches were developed to (1) classify the severity of impairment for each swallow, with model confidence ratings for transparent clinical decision support, and (2) compute a similarity measure of each swallow to nondysphagic performance. Task-specific models were trained using swallow kinematics and respiratory features from 505 swallows (321 from patients and 184 from controls). Results: These models provide sensitive metrics to gauge impairment on a per-
Swallowing is a complex process, requiring intricate coordination of nerves and muscles to move a substance (bolus) from the mouth to the stomach. Dysphagia, or difficulty swallowing, occurs when this process is compromised, such as from muscle weakness or damage to the nervous system (e.g., from a stroke), and affects approximately 1 in 25 adults annually in the USA [1]. The consequences of impaired swallowing can be dire, with significant risk of health complications and even death [2]. Patients typically experience overt symptoms of impaired swallowing such as coughing, choking, and regurgitation. Reduced ability or desire to swallow may lead to malnourishment, dehydration, or inability to take oral medication. Symptoms of impaired swallowing are not always evident; for example, silent aspiration occurs when a bolus enters the airway without triggering observable symptoms and may affect 2–25% of acute stroke patients [3]. Coordination with respiratory processes is essential to protect the airway while swallowing [4, 5], and these dynamics have been known to change with age and disease [6–8].

Given the prevalence and profound impact of impaired swallowing, early detection of dysphagia is critical to design appropriate care plans and improve patient outcomes. Current techniques include imaging procedures, which are invasive and costly, and bedside examinations, which rely on subjective clinician observations. Although generally effective for identifying impaired swallowing, these techniques are only performed intermittently in the clinic, are not administered for every patient, and may not be sensitive to detect milder symptoms. Furthermore, many procedures lack sensitivity because they assign a single score or rating to a patient’s overall swallowing ability; however, patients might have some swallows that are functional and safe and others that are not because swallowing is inherently variable [9, 10] and affected by bolus characteristics [11], attention [12], and posture [13]. Alternative tools are needed to capture this intrapersonal swallowing variability, improve risk assessment by gauging each swallow independently, and characterize subtle functional changes over time.

Wearable sensors have been proposed as a noninvasive means of obtaining continuous, precise, and objective measures of swallowing. Early results are promising to discriminate underlying biomarkers of impairment [14–17]. However, advocates consistently identify the need to translate these methods into clinical practice without providing user-friendly tools to do so. Additionally, few studies have simultaneously captured respiration during swallowing using wireless sensing platforms. To address these gaps, we introduce 2 machine learning approaches to translate complex sensor data – namely, biometric signals of swallowing and respiratory dynamics – into novel, multidimensional metrics that evaluate individual swallows and quantify deviations from healthy behaviors. This proof-of-concept study aims to demonstrate the methodological feasibility of developing sensitive metrics and interpretable visual tools to quantify swallowing behaviors from wearable sensors using machine learning techniques. Subsequently, this framework can be leveraged with additional training data to further develop optimized, validated models for the detection and monitoring of dysphagia. Recognizing the frequency and severity of poor swallows outside of a standard clinical evaluation would improve our understanding of real-world swallowing behaviors, which in turn may facilitate intervention and data-driven treatment to optimize health outcomes.

Methods

Participants
Individuals with poststroke dysphagia were recruited from the acute inpatient rehabilitation unit at the Shirley Ryan AbilityLab (Chicago, IL, USA). All individuals provided written informed consent prior to participation. Inclusion criteria were a stroke resulting in dysphagia, at least 18 years of age, and able and willing to give consent and follow study procedures. Exclusion criteria were diagnosis of neurodegenerative pathology as a comorbidity; pregnant or nursing; or presence of skin allergies, irritation, or open wounds. Medical clearance was obtained from each patient’s primary physician prior to participation.
Twelve patients with dysphagia were recruited in this proof-of-concept study; 3 patients were excluded from analysis because of limited data availability (due to software error) or a change in medical status (which prevented continued participation). All patients were diagnosed as dysphagic in their medical records following standard hospital intake evaluation. Demographics and medical characteristics of the final patient cohort (N = 9; 5M/4F; 59.4 ± 11.3 years) are given in online suppl. Table 1 (for all online suppl. material, see www.karger.com/doi/10.1159/000517144). Individuals with no known health problems or history of stroke were recruited as controls from a sample of convenience (N = 10; 3M/7F; 28.5 ± 6.1 years).

Devices
Two flexible, wireless, research-grade mechano-acoustic sensors [18] were used to record kinematics of swallowing and respiration (Fig. 1), sampling triaxial acceleration (±2 g) at 1,600 Hz from the z-axis (anteroposterior plane) and 200 Hz from the x- and y-axes. One sensor was placed on the throat (suprasternal notch) to capture the laryngeal motion accompanying swallowing [18, 19]. The second sensor was placed on the ribcage (midaxillary at the level of the xiphoid process; on the unaffected side for patients with stroke or on the dominant side for controls) to record laryngeal swallow response and respiratory movement patterns, respectively. Orientation of x, y, and z sensor axes is shown relative to anatomical planes (S-I, superior-inferior; M-L, medial-lateral; A-P, anterior-posterior). The sensor is a wireless, flexible device that communicates with a custom app via Bluetooth. Raw triaxial accelerometer recording example from the throat sensor during a liquid bolus swallow (left) and from the ribcage sensor during quiet sitting (right).

Protocol
Swallowing assessments were administered and scored by a trained speech-language pathologist. While wearing the sensors, participants first sat quietly for 30 s to capture baseline respiration. They then performed a series of orofacial movements and swallowing tasks, including natural or effortful swallows of saliva, as well as liquids and foods of different consistencies and presentations (Table 1). Only tasks that adhered to patients’ prescribed dysphagia diets were attempted. Each task was performed at least twice to capture within-subject swallow variability.

Patients participated in 2 sessions: within 1 week of admission (Adm) and within 1 week before discharge (Dis) from the inpatient rehabilitation program. At each session, dysphagia severity was
rated using the Mann Assessment of Swallowing Ability (MASA) [20], a widely acknowledged and validated bedside tool [21] (on-line suppl. Table 1). Healthy individuals performed the same protocol in a single session, serving as a nondysphagic control group.

**Data Processing and Feature Extraction**

The data pipeline is illustrated in online suppl. Fig. 1. Sensor data were cleaned, filtered, and clipped around each identified swallow. Trials containing competing peaks near the swallow event or excessively noisy signals were discarded to ensure models were trained using sensor data confidently attributable to swallowing. To mitigate inter- and intrasubject variability due to variations in sensor placement [22], each subject’s sensor data were normalized by a within-session task.

Thirty-six features were extracted from the normalized throat and ribcage sensor signals during each swallow, including descriptive statistics, time and frequency domain measures, and respiratory-swallow coordination parameters (e.g., apnea duration and swallow timing during inhale-exhale cycles). Correlation-based feature selection was implemented to reduce model complexity and remove redundant variables (Pearson correlation coefficient $p \geq 0.6$). From this set, 17 features were selected for supervised learning (online suppl. Tables 2, 3).

**Model Development**

We used dysphagia severity classifications, operationally defined by the MASA, as an example ground truth for model training. Models were implemented using Python scikit-learn 0.23.1 [23].

The first model, named the Severity Probability Model, was designed to provide transparent, multifaceted evaluation of swallowing behavior. This approach estimates the presence and severity of impairment for each swallow and computes the probability that the swallow belongs to each possible severity class. A Random Forest classifier was trained on sensor data using a leave-1-subject-out nested cross-validation to tune hyperparameters (number of trees, maximum depth, and maximum number of leaf nodes). Separate models were trained for different bolus types, since bolus characteristics can differentially affect swallowing [10, 11]. Online suppl. Table 4 shows representative hyperparameter values for each task, tuned on all patient data to provide a sense of the optimization output. To evaluate model predictions as a function of ground truth severity, we used a general linear model to compare the output probabilities across MASA groups for each predicted severity class.

The second model, named the Distance Model, evaluates individual swallow impairment by representing sensor data as a measurable “distance” from healthy swallowing behavior. This approach simplifies multidimensional sensor features into a single metric to track swallowing performance over time. Patients who completed both the Admission and Discharge sessions ($N = 7$) were tested by iteratively holding out each patient’s data and training on the controls and remaining patient data. We applied linear discriminant analysis to represent each swallow in a standardized, 2-dimensional subspace, clustered by dysphagia severity. A single-class support vector machine with a nonlinear decision boundary was fit to the control data. Individual swallow quality was quantified via a distance $d$ from the boundary, with increasing distance values representing greater deviation from healthy swallowing. We examined group-level statistics of the resulting distance values using a generalized linear mixed effects model with repeated measures. Distance values for each patient were averaged across swallows for a given task and session, since some patients had more swallows than others. We included task, session, and their interaction as fixed effects and patients as a random effect.

All statistical analyses were conducted in SAS Studio 3.8 (SAS Institute; Cary, NC, USA), with $p$ values <0.05 considered significant. Post hoc tests were adjusted for multiple comparisons using a simulated distribution. For additional details about model development, please refer to online suppl. Methods.

**Results**

After processing, 505 total swallow instances (184 controls and 321 poststroke patients) were available for model development (Table 1). Swallow trials from representative patients and tasks are shown in online suppl. Fig. 2.

### Table 1. Number of swallows available for model training in each task and dysphagia severity level

<table>
<thead>
<tr>
<th>Task (bolus and oral presentation)</th>
<th>Dysphagia severity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None (controls)</td>
</tr>
<tr>
<td>Saliva: natural dry swallow</td>
<td>12</td>
</tr>
<tr>
<td>Saliva: effortful dry swallow</td>
<td>18</td>
</tr>
<tr>
<td>Liquid: 2.5 mL</td>
<td>18</td>
</tr>
<tr>
<td>Liquid: 5 mL</td>
<td>16</td>
</tr>
<tr>
<td>Liquid: natural sip</td>
<td>27</td>
</tr>
<tr>
<td>Liquid: straw sip</td>
<td>20</td>
</tr>
<tr>
<td>Puree: spoonful</td>
<td>13</td>
</tr>
<tr>
<td>Puree: spoonful with effortful</td>
<td>20</td>
</tr>
<tr>
<td>Soft solid: spoonful</td>
<td>22</td>
</tr>
<tr>
<td>Hard solid: natural bite</td>
<td>18</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>184</strong></td>
</tr>
</tbody>
</table>
Fig. 2. Severity Probability Model to classify impairment in each swallow with transparent, quantifiable decision support. a Each swallow is assigned a severity level (e.g., none, mild, moderate, and severe) using a Random Forest classifier trained on MASA dysphagia ratings. b Additional model predictions and sensor-based metrics provide comprehensive information about a swallow based on quantifiable metrics. Left: model predictions compute the probability of belonging to each class, thereby moderating model confidence and identifying swallows that might border multiple classes. Middle: task-specific swallowing acceleration signals recorded from the throat sensor are compared to average signals from non-dysphagic controls. Data are aligned by the time of the peak swallow acceleration (vertical dashed gray line), taken from the throat sensor signal. Right: swallow and apnea timing in the respiratory cycle reveals characteristics about respiratory-swallow coordination to further assess aspiration risk. Inhalation-exhalation patterns (red lines; In, inhale; Ex, exhale) show the respiratory phases occurring immediately before and after the identified swallow apnea (gray box), relative to the timing of the peak swallow acceleration (vertical dashed gray line). Acceleration signals are min-max normalized using a standardized in-session task.

Severity Probability Model Delivers Transparent Severity Classification for In-Depth Swallowing Assessment

Figure 2 illustrates the model output for a patient with moderate dysphagia swallowing various boluses in a 1-h window. The predicted severity of each swallow is supplemented with (1) severity class probabilities, (2) a visual comparison of the patient’s throat sensor signal with the average control signal, and (3) measures of respiratory-swallow coordination. In this proof-of-concept, dysphagia severity classification is confined to none, mild, or moderate based on the MASA scores represented in our patient cohort. Out of 20 labeled swallows, 3 were classified as moderate, 7 as mild, and 10 as nondysphagic. Classification probabilities varied across each swallow, indicating that some swallows were more confidently attributable to a severity class than others. Respiratory-swallow coordination features revealed variation in apnea timing and inhalation-exhalation patterns, depending on the bolus and swallow trial. Previous studies of respiratory-swallow coordination have found that inhalation-swallow-inhalation (In-In) patterns [24, 25] and short-
ened or variable apnea duration [26, 27] are associated with aspiration risk.

For each model, feature importance varied across the different bolus types and oral presentation methods. Total power of throat acceleration and relative duration of the previous inhale to swallow (old phase) were the highest weighted features on average (online suppl. Table 5). There was a significant effect of MASA severity on the probability of swallows being classified as none (F[2,502] = 11.5, p < 0.001) or mild (F[2,502] = 18.4, p < 0.001) (Fig. 3). Post hoc tests revealed that the model generally predicted lower probabilities for the none class and greater probabilities of classifying swallows as mild. Asterisks (*) indicate significant differences between true dysphagia categories (p < 0.05).

**Discussion**

In summary, this proof-of-concept demonstrates the feasibility of collecting simultaneous swallow-respiratory wearable sensor data in a clinical setting to create sensitive, multidimensional metrics on swallowing behavior. We demonstrated that mechano-acoustic sensor data can capture subtle differences in individual swallows across bolus types and dysphagia severities. Both models can quantify intrasubject swallow variability, as well as patient-specific changes in performance following inpatient rehabilitation (even when those changes were not indicated by a clinical screening tool).

Noninvasive detection of swallowing impairment has garnered increasing interest in recent years, likely bolstered by advances in sensor technology and machine learning techniques. A primary focus has been accelerometry [15, 17, 28, 29] or audio signals (i.e., microphone) to classify impairment via swallowing, coughing, and other behaviors [19, 30–32]. Results are promising; in a recent prospective study of 344 individuals at risk for oropharyn-
geal dysphagia, Steele et al. [17] trained a regularized linear discriminant analysis on dual-axis accelerometer signals and videofluoroscopy, achieving \( \sim 90\% \) sensitivity and \( \sim 60\% \) specificity for detecting impaired swallow safety (when material entered the airway; “penetration-aspiration”) and \( \sim 80\% \) sensitivity and \( \sim 60\% \) specificity for detecting impaired swallow efficiency (when material remained in the pharynx). In the current study, we utilized accelerometry features from the time and frequency domains to capture mechano-acoustic properties of pharyngeal swallowing. Classification accuracy may be further improved by incorporating sensor features from other physiologically relevant modalities (i.e., audio data) [19].

We have also introduced respiratory features, computed from similar accelerometry approaches, to characterize respiratory-swallow coordination.

Research and development of automated diagnostic tools is a major focus of the digital health revolution; however, such tools are rarely implemented in clinical settings. There may be concerns about overreliance on imperfect automated tools, especially for tasks requiring a high cognitive load [33], or failed uptake stemming from a fundamental mistrust of any imperfections. Sensor signals may be discriminative for detecting impaired swallowing, but the data itself is difficult to interpret and transform into actionable insights. Thus, the clinical utility of these models depends on their ability to deliver interpretable information that neither overwhelms nor oversimplifies. To this end, we presented the Severity Probability Model, which classifies swallowing impairment while providing transparent estimates of model confidence and visual confirmation of clinically relevant sensor measures. Such a model could verify or supplement subjective judgment of per-swallow impairments and empower clinicians to resolve potential ambiguities from the automated model predictions. With more train-
ing data and per-swallow ground truth labels, these approaches could be extended to additional severity levels or adapted for alternative classification schemes, such as levels of swallowing safety or efficiency.

In contrast to the traditional classification-based approach, we also presented the Distance Model, which transforms a broad set of sensor features into a novel metric of swallowing performance relative to nondysphagic controls. This approach quantifies relative differences in individual swallows, is sensitive to task-specific changes over time, and lends itself to an intuitive visual tool for continuous monitoring. Significant or concerning changes toward or away from target performance may trigger additional in-depth clinical evaluation to pinpoint the physiological and anatomical mechanisms underlying such changes. Group-level statistics showed a significant effect of task on distance values, with poststroke puree boluses farther from the control boundary relative to natural sip and saliva tasks. Although thicker consistency boluses (soft/hard solids) may be expected to show significant differences from healthy swallowing, fewer patient swallows were available for analysis in these tasks compared to the puree (Table 1) due to physician-prescribed diet limitations. There was no effect of session on distance values; this is not altogether surprising given the potential confounds for this small cohort, which may include true functional declines, day-to-day swallow variability, and promotion to more challenging diets.

The primary limitation of this study is the small sample size available for model training. Expanding the dataset to include more participants, swallows per bolus, and swallows per severity level is critical to reliably report the model performance and generalizability. Second, although the MASA is a frequently used clinical tool for screening and monitoring poststroke dysphagia, its limited sensitivity and specificity for identifying swallowing abnormality (73% and 89%, respectively [20]) and single score format may not accurately reflect impairment with age. Group-level statistics showed a significant effect of task on distance values, with poststroke puree boluses farther from the control boundary relative to natural sip and saliva tasks. Although thicker consistency boluses (soft/hard solids) may be expected to show significant differences from healthy swallowing, fewer patient swallows were available for analysis in these tasks compared to the puree (Table 1) due to physician-prescribed diet limitations. There was no effect of session on distance values; this is not altogether surprising given the potential confounds for this small cohort, which may include true functional declines, day-to-day swallow variability, and promotion to more challenging diets.

The primary limitation of this study is the small sample size available for model training. Expanding the dataset to include more participants, swallows per bolus, and swallows per severity level is critical to reliably report the model performance and generalizability. Second, although the MASA is a frequently used clinical tool for screening and monitoring poststroke dysphagia, its limited sensitivity and specificity for identifying swallowing abnormality (73% and 89%, respectively [20]) and single score format may not accurately reflect impairment across multiple swallows. Swallowing is inherently variable, and previous work has found that individuals with dysphagia do not exhibit unsafe swallowing consistently [17, 34]. As such, training models on a single MASA score limits our ability to effectively validate individual swallow classifications using traditional model performance metrics at this stage. Future studies should pair sensor data with validated per-swallow impairment measures (i.e., videofluoroscopy [17, 35]) to better evaluate model performance. Finally, this dataset was not age-balanced for the explored cohorts and therefore does not account for age-related swallowing changes (i.e., increased apnea duration or altered inhalation-exhalation patterns) [6]. It is unclear how skin laxity and adipose tissue at the suprasternal notch, which often accompany age, specifically affect mechano-acoustic signal properties during swallowing. While this study intended to demonstrate methodological feasibility rather than to optimize model performance, it is imperative that future work incorporates training data from age-matched healthy controls to differentiate the effect of oropharyngeal dysphagia on sensor signals from typical swallowing changes that occur with age.

Development of interpretable, user-friendly tools is critical to optimize the clinical utility of novel, sensor-based measurement techniques. Additional, large-scale training data and real-world validation is required to refine the algorithms and maximize their accuracy and performance. Future work will expand the training dataset with ground truth impairment for each swallow, validate model performance using clinic and community sensor data, and assess the clinical utility of these models in dysphagia detection and treatment.

Acknowledgment

The authors would like to thank Nsude Okeke-Ewo for assistance with data collection.

Statement of Ethics

All individuals provided written informed consent prior to participation. The study was approved by the Institutional Review Board of Northwestern University (Chicago, IL, USA; STU00205532) in accordance with federal regulations, university policies, and ethical standards regarding research on human subjects.

Conflict of Interest Statement

S.X. and J.A.R. hold equity in the company Sonica Health, which makes wearable sensors for medical applications. A patent on the wireless sensor and methods for medical use has been filed by Northwestern University with J.A.R., S.X., B.M.-H., L.R.C., A.J., and M.K.O. as inventors. The remaining authors have no conflicts of interest to declare.

Funding Sources

This work was supported by the Shirley Ryan AbilityLab, with partial funding from the NIH under an institutional training grant at Northwestern University (T32HD007418, awarded to M.K.O.). S.X. and J.A.R. recognize support from contract 75A50119C00043 award-
ed by the Biomedical Advanced Research and Development Authority. S.X., B.M.-H., and J.A.R. recognize support from R41AG062023 by the National Institutes of Health and grant ID 17777 from the Michael J. Fox Foundation. S.X. and J.A.R. recognize support from R43AG060812 and R41AG062023-02SI by the National Institutes of Health. The funding sources had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or decision to submit the manuscript for publication.

Author Contributions


References

4 Martin-Harris B. Coordination of respiration and swallowing. GI Motil online. 2006 May.