A Prototype Intelligent Assistant to Help Dysphagia Patients Eat Safely At Home

Michael Freed¹, Brian Burns¹, Aaron Heller¹, Daniel Sanchez¹, Sharon Beaumont-Bowman²

¹ SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025 / {firstname.lastname}@sri.com
² Brooklyn College, 2900 Bedford Ave., Brooklyn, NY, 11210 / SharonB@brooklyn.cuny.edu

Abstract
For millions of people with swallowing disorders, preventing potentially deadly aspiration pneumonia requires following prescribed safe eating strategies. But adherence is poor, and caregivers’ ability to encourage adherence is limited by the onerous and socially aversive need to monitoring another’s eating. We have developed an early prototype for an intelligent assistant that monitors adherence and provides feedback to the patient, and tested monitoring precision with healthy subjects for one strategy called a “chin tuck.” Results indicate that adaptations of current generation machine vision and personal assistant technologies could effectively monitor chin tuck adherence, and suggest the feasibility of a more general assistant that encourages adherence to a wide range of safe eating strategies.

Dysphagia Patients Need Help Adhering to Safe Eating Strategies
Dysphagia, or difficulty swallowing, is a widespread and often devastating disorder that affects 10–30% of the elderly population (Barezi and Sullivan, 2000), including 11–16% of individuals living in the community (Bloem et al., 1990; Kawashima et al., 2004) and 55–68% of nursing-home residents (Kayser-Jones and Pengilly, 1999; Steele et al. 1997). It is particularly prevalent among patients with stroke (50–75%), Parkinson’s Disease (up to 95%), and other neurological conditions (Martino et al., 2005; Hunter et al., 1997). Dysphagia creates numerous risks; chief among them is aspiration pneumonia—an infection caused by accidental ingestion of bacteria-laden food into the lungs (Marik and Kaplan, 2003). Mortality ranges from 10–70% (DeLegge, 2002). For those with neurogenic dysphagia, aspiration pneumonia is the most likely cause of death (Marik, 2001).

Clinicians frequently prescribe risk-reducing compensatory strategies such as tucking the chin to the chest before swallowing to protect the airway, and making an “effortful swallow” to clear residual food in the pharynx (Archer et al., 2013). These strategies have been shown to significantly improve patient health and well-being (Low et al., 2001). Patient compliance, however, is poor, as few patients can continuously self-monitor eating behavior (Colodny, 2005; Krisciunas et al., 2012; Shinn et al., 2013). Monitoring by caregivers can be beneficial, but caregivers are often inadequately trained or unavailable (Chadwick et al., 2006; Nund et al., 2014; Warm, Parasuraman, and Matthews, 2008). Such monitoring produces significant relationship strain and caregiver burden (Patterson et al., 2013; Nund et al., 2014). “It is poignant to note that we can assist people with dysphagia by first assisting their caregivers (Cichero and Altman, 2012).”

Our preliminary results indicate the feasibility of shifting the burden of monitoring and encouraging compliance to a device that actively monitors patient-eating behavior and provides appropriate real-time feedback. Dysphagia patients are often advised to eat in front of a mirror to enhance self-awareness. We have constructed a proof-of-concept prototype for a laptop- or tablet-based personal digital assistant application that, when positioned in front of the patient, functions as both a mirror and a source of feedback to optimize safe compensatory strategies. The prototype successfully monitors compliance for a single eating strategy—specifically, whether the patient is compliant with flexing the head downward into a “chin tuck” after spooning food into the mouth.

We hypothesize that currently available machine-vision and machine-learning algorithms can be adapted to monitor adherence with the most commonly prescribed safe eating strategies, and that this monitoring can be done in real time by using widely available consumer electronics hardware. Our ultimate goal is to develop a fully functional, patient-friendly Dysphagia Coach application that provides direct feedback on eating behavior, thereby increasing patients’ eating safety and well-being while decreasing dysphagia-related mortality, morbidity, and healthcare costs (Archer et al., 2013; Krisciunas et al., 2012). Here we describe the initial steps towards this goal including a
breakdown of requirements for monitoring adherence to the chin tucking eating strategy, a description of the technical approach taken in our initial prototype, and initial results in assessing the monitoring accuracy.

**Monitoring Requirements for an Automated Dysphagia Coach**

Current practice for helping patients adhere to safe eating practices relies on human caregivers, with no technology-based support to reduce caregiver burden. However, technology-based solutions exist for the related problem of helping patients perform swallowing rehabilitation exercises. SwallowStrong, for example, uses a custom molded mouthpiece and other specialized hardware to measure key performance indicators for patients’ therapeutic goals as they perform swallowing exercises (Constantinescu et al., 2014). iSwallow, an Apple iOS app marketed as a “Personal Rehabilitation Assistant” (UCDavis-CV&S website), alerts patients when it is time to perform scheduled exercises and provides real-time feedback to the patient using data from microphones taped to the patient’s throat. Such devices are similar to what we would expect of a fully realized Dysphagia Coach in the types of activities monitored and the need to provide real time performance feedback. However, differences in time frame of use (weeks or months vs. years) and objective (recovering lost function vs. maintaining safety) imply a very different patient attitude, with the Dysphagia Coach likely having to meet a higher standard for ease of use for patients to remain motivated to continue using the device. For this reason, we treat the goals of low effort setup and passive monitoring (i.e. without body-attached sensors) as requirements.

To assess whether these requirements could be met using current generation machine vision technology, we developed a prototype Dysphagia Coach to monitor adherence in performing a commonly used safe eating strategy called a chin tuck maneuver. A chin tuck is one of nine typical strategies (Goher and Crary, 2009) prescribed by Speech and Language pathologists (SLPs) after assessing a patient’s swallowing deficits. The maneuver involves lowering the chin to the chest before swallowing. This moves the epiglottis into a protective position over the airway, compensating for a weak or delayed reflex. Examples of other strategies include a head turn, where a patient rotates the head prior to the swallow to compensate for weakness or paralysis on one side of the pharyngeal wall, and making an effortful swallow to clear residual food.

Table 1 summarizes the sequence of actions for a correctly performed chin tuck, errors that might occur at each step, and corresponding system requirements for automatically detecting correct and incorrect actions. We envision the monitoring process starting with a patient sitting at a table in front of an already upright tablet computer and selecting the Dysphagia Coach app. The app would immediately detect the patient’s presence and begin monitoring. If that proves to be more effort than patients are willing or able to make, a real possibility given common comorbidities such as movement disorders and cognitive deficits, an alternative approach could use dedicated always-on hardware.

**Prototype Approach and Initial Results**

As illustrated in Figure 1, our prototype provides a simple chin tuck monitoring and user feedback capability. A monitoring session begins when the system detects a face within an estimated threshold distance from, and orientation, towards the camera (Viola and Jones, 2004), indicating that the patient may be ready to eat (sysreq-1, left image). It recognizes spoons with and without food, and tracks movement to and from the mouth (sysreq-3, middle image). When the system detects that the spoon has been inserted into the mouth, it monitors for a large, downward rotation, or “tuck,” of the head (sysreq-4, right image), followed by an upward rotation into a rest position. If a tuck is detected after spoon insertion and before any subsequent insertion, the system gives the patient positive

<table>
<thead>
<tr>
<th>Action Sequence</th>
<th>Common Errors</th>
<th>System Requirement (sysreq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sit in upright position</td>
<td>Head/trunk misaligned</td>
<td>1. Detect nearby face in view of camera (determine when to initiate monitoring)</td>
</tr>
<tr>
<td>Sip liquid/spoon in food</td>
<td>Swallow before tuck</td>
<td>2. Estimate head/trunk alignment</td>
</tr>
<tr>
<td>Tuck chin to chest</td>
<td>Forget to perform tuck</td>
<td>3. Track cup/utensil movement toward and into mouth.</td>
</tr>
<tr>
<td></td>
<td>Incomplete tuck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Swallow mid-tuck</td>
<td></td>
</tr>
<tr>
<td>Swallow in that position</td>
<td>Raise head during swallow</td>
<td>4. Estimate head/angle to torso in sagittal plane</td>
</tr>
<tr>
<td>Raise the head</td>
<td>Swallow post tuck</td>
<td>5. Detect swallow (timing and omission errors)</td>
</tr>
</tbody>
</table>

*Table 1. Chin Tuck Actions and Errors*
feedback. Otherwise, it indicates that the tuck was omitted. The initial prototype does not incorporate means for estimating head-trunk alignment (sysreq-2) or detecting swallow actions (sysreq-5), although we are exploring approaches to these separately.

Of the requirements explored using our prototype, the need to precisely estimate head angle (sysreq-4) using sensors and processors on commodity hardware raised the most significant question of technical feasibility. We selected an approach incorporating landmark-based machine vision algorithms (Sagonas et al., 2013; Asthana et al., 2014; Cao, Wei, and Sun, 2014; Zhu and Ramanan, 2012) running on a standard, camera-equipped laptop. Landmark algorithms find standard head and neck reference points, such as the tip of the nose, in a video image, and then map them to corresponding points in a reference digital head/neck model. Head pose is then estimated by computing the best 3D rotation fit between points and model (Demonthon and Davis, 1995). Importantly for detecting chin tucks, tilt angle can be treated as approximation of head-torso angle in the sagittal plane.

Discriminating a partial tuck from a complete tuck requires not only detecting downward head rotation, but also determining whether the head-torso angle (the amount of tuck) is large enough given a patient-specific minimum determined by the clinician. Figure 2 illustrates our approach to determining whether landmark-based vision methods can accurately assess head-torso angle. The top row shows three frames from a tuck video sequence, with the green dots indicating automatically identified head and neck landmarks. The second row shows the estimated head angle “pose” of those video frames on a digital head model. Pose is automatically computed by mapping landmarks in the image to corresponding reference points on the model (yellow dots). The third row compares the head inclination angle as estimated by machine-vision software (blue graph line, red data points corresponding to the shown sample frames) to the actual angle measured by hand (green circles). Pilot data (n=5) showed an RMS estimation error of 3.6 degrees. This result was less than the within-subject variability of 5.2 degrees for correctly performed chin tucks, suggesting that estimation error will not be found to be clinically significant upon more comprehensive testing.

**Discussion and Next Steps**

Advances in computing and sensing hardware, machine vision, machine learning, and personal assistant AI are increasing the potential diversity and efficacy of support that might be offered to people with disabilities living at home, improving their safety and quality of life, while reducing dependence on caregivers. As one example poten-
tially affecting millions of people, a Dysphagia Coach would decrease cognitive burden associated with adherence to safe eating strategies (Brehm and Self, 1989), decreasing the effort needed to sustain motivation (Murayen and Baumeister, 2000; Reason, 1995). As an alternative to socially aversive and onerous monitoring by a human caregiver (Bandura, 1997), it would appeal to patient’s needs for self-efficacy (Ryan and Deci, 2000) and thus help reinforce behaviors that reduce morbidity, caregiver burden, and healthcare costs (Archer et al., 2013; Krisciunas et al., 2012).

We designed the described Dysphagia Coach prototype to get feedback from clinicians on the concept, and to gauge the technical feasibility of the approach. Feedback confirmed the potential benefit to patients and provided guidance in prioritizing which safe eating strategies are most important to monitor. Clinicians also indicated strong interest in using the technology to collect patient performance data, both to improve care for individual patients and to advance scientific understanding of dysphagia.

Next steps include creating a corpus of video data with a diverse patient population in diverse settings, using it to test and refine the current non-individualized visual modeling approach, extend it if needed to use person-specific appearance and behavior data, and then generalize to additional eating strategies such as head turns and effortful swallows. In parallel, we will seek to understand user experience considerations affecting adoption and ongoing use. We anticipate that common dysphagia co-morbidities such as hearing loss, visual attention deficits, and cognitive deficits will prevent any single user interaction design from being suitable for all patients, and that multiple user interaction designs will be needed. These efforts will prepare the way for clinical trials to assess patient benefit.

References


Low J, Wyles C, Wilkinson T, Sainsbury R. The effect of compliance on clinical outcomes for patients with dysphagia on vide-


