Characteristics of Dry Chin-Tuck Swallowing Vibrations and Sounds

Joshua M Dudik, Iva Jestrović, Bo Luan, James L. Coyle, Ervin Sejdić, *Member, IEEE*

Abstract—*Objective:* The effects of the chin-tuck maneuver, a technique commonly employed to compensate for dysphagia, on cervical auscultation are not fully understood. Characterizing a technique that is known to affect swallowing function is an important step on the way to developing a new instrumentation-based swallowing screening tool. *Methods:* In this study, we recorded data from 55 adult participants who each completed five saliva swallows in a chin-tuck position. The resulting data was processed using previously designed filtering and segmentation algorithms. We then calculated 9 time, frequency, and time-frequency domain features for each independent signal. *Results:* We found that multiple frequency and time domain features varied significantly between male and female subjects as well as between swallowing sounds and vibrations. However, our analysis showed that participant age did not play a significant role on the values of the extracted features. Finally, we found that various frequency features corresponding to swallowing vibrations did demonstrate statistically significant variation between the neutral and chin-tuck positions but sounds showed no changes between these two positions. *Conclusion:* The chin-tuck maneuver affects many facets of swallowing vibrations and sounds and its effects can be monitored via cervical auscultation. *Significance:* These results suggest that a subject's swallowing technique does need to be accounted for when monitoring their performance with cervical auscultation based instrumentation.

Index Terms—**Accelerometers, Biomedical signal processing, Biomedical transducers, Medical signal detection**

I. INTRODUCTION

DYSPHAGIA is a term used to describe swallowing impairments [1] that commonly develop as a component of neurological conditions [1], [2]. Stroke is the most common cause, though physical trauma or other conditions such as Parkinson's disease can also result in dysphagia [1], [2]. Dysphagia can lead to serious health complications including pneumonia, malnutrition, dehydration, and even death [2], [3]. Assessment of swallowing function typically begins with screening. Screening is a pass-fail level of testing that indicates whether it is likely that the patient has a swallowing disorder, and identifies the need for further examination. A "failed" dysphagia screen typically leads to a non-instrumental clinical or bedside evaluation of oral, facial, and pharyngeal sensorimotor function. This includes the testing of associated communication, cognitive, and feeding behaviors. These

Joshua M Dudik, Iva Jestrović, Bo Luan, and Ervin Sejdić are with the Department of Electrical and Computer Engineering, Swanson School of Enginering, University of Pittsburgh, Pittsburgh, PA, USA. e-mail: jmd151@pitt.edu, ivj2@pitt.edu, bol12@pitt.edu, esejdic@ieee.org. Ervin Sejdić is the corresponding author.

James L. Coyle is with the Department of Communication Science and Disorders, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, USA. email: jcoyle@pitt.edu

clinical observations reveal the signs and symptoms of and potential imminent risk factors for adverse events associated with dysphagia, but cannot specify the physiological nature of the disorder. Screening and clinical assessments are readily available because they involve little to no instrumentation and can be immediately implemented at the patient's bedside.

In some cases, clinical assessments can adequately identify impairments and inform the examiner about how to alleviate them. However, when the possibility of oropharyngeal, laryngeal, or esophageal sensorimotor impairments affecting airway protection exist, instrumental testing is needed since many of these impairments are not observable at the bedside. Clinical assessment is limited in its ability to identify structural abnormalities, evaluate kinematic events, and assess bolus clearance. When such information is necessary to properly assess a patient's condition, instrumentation-based methods of swallowing assessment must be utilized. Currently accepted and widely used techniques such as nasopharyngeal flexible endoscopy and videofluoroscopy allow the examiner to directly view a swallow and provide a great deal of diagnostic information [1], [4]. However, they require skilled expertise to use and are somewhat invasive procedures. Furthermore, these devices are not appropriate for all situations and may not be able to be implemented due to various patient-specific reasons. An accurate, non-invasive method of screening for swallowing disorders that could be used at a patient's bedside alongside existing techniques would be a significant benefit to the field.

Several methods, including a water swallow screen, pulse oximetry, and cervical auscultation, have been proposed over the last two decades as less invasive methods to predict aspiration during swallowing [5], [6], [7]. Water swallow screenings simply involve observing the patient drink a specific volume of water. If suspicious clinical signs of aspiration such as coughing or respiratory distress are observed, the patient is considered to have failed the screen. It has exhibited an acceptable predictive value for aspiration in some studies but is obviously a rather crude exam. Cervical auscultation has traditionally been performed using stethoscopes and microphones as sensors to detect the sounds of swallowing and has not yet demonstrated sufficient ability to characterize swallowing physiology or identify aspiration [8], [9], [10]. Recently however, several studies have reported using dual-axis accelerometers and digital algorithms to automatically detect and analyze throat vibrations during swallowing [11], [12]. Though the raw cervical auscultation signal provides minimal benefit in a clinical environment, it is possible that further mathematical analysis of the signal could improve its usefulness. If cervical auscultation could be refined into a re-

liable, instrumentation-based swallowing screening method it could provide a level of objectivity to a traditionally subjective process.

The source of the swallowing-related signal is likely to be the same for both accelerometer and microphone recordings. However, the type of transducer used has the potential to alter the information obtained. Many types of accelerometers operate on the basis of an externally charged capacitor that can move and produce a varying signal when subject to physical motion [13]. On the other hand, not only do most modern microphones use a pre-polarized film rather than an external power source to produce a capacitive circuit component, but this film only responds to pressure waves rather than the motion of the device [14]. In addition, these two different types of sensors may differ in size, temperature response, sensitivity, and polar patterns as well as having completely different frequency response curves (e.g., [15], [16]). Therefore, the nature of similarities and differences in swallowing signals obtained using accelerometry versus microphones remains an open question.

Numerous compensatory maneuvers that are designed to redirect the flow of swallowed material and mitigate the risk of unwanted behaviors have been investigated. A commonly employed compensatory technique, the "chin-tuck maneuver", is effective for some patients with specific biomechanical swallowing impairments and has been found to reduce aspiration in specific conditions [1], [17], [18]. The patient simply flexes their head and neck so that their chin is brought down towards their chest, otherwise known as sagittal head and neck flexion, and maintains this position when swallowing [18]. As one would expect, this results in a number of changes in the anatomical relationships among structures in the aerodigestive tract during the swallow [19], [20]. For example, the anterior structures of the throat are shifted in the posterior direction when the chin is tilted forward [19], [20]. This reduces the distances between the hyoid bone and the mandible as well as the hyoid bone and the larynx in addition to narrowing the pharynx and airway inlet [19], [20].

There are many potential physiological benefits to using the chin-down maneuver. First, it places the swallowed bolus farther anteriorly in the oral cavity. If a patient has a delayed pharyngeal response, the chin-tuck maneuver gives the patient's muscles more time to react to the oncoming bolus. This maneuver also allows the patient to use gravity to ensure bolus containment. For patients with poor posterior oral containment, this technique can reduce the risk of premature spillage of the bolus. Both of these examples help to minimize the risk of aspiration before or during the swallow. Additionally, the chin-tuck maneuver widens the vallecular space between the tongue base and epiglottis. This enables a larger volume of residue to be retained without passively entering the unprotected airway after a swallow has finished. Shanahan, et al. found a 50% reduction in aspiration among patients with stroke that utilized the chin-tuck maneuver [17]. However all patients who continued to aspirate using this technique aspirated from the pyriform sinuses because the posture directed pharyngeal residue into the unprotected airway. Although it cannot aid all patients equally and is

actually harmful for some patients, the chin-tuck maneuver can increase the protection of the airway during swallowing and has been proven to reduce the risk of aspiration in some dysphagic patients [18], [20], [21]. As a result, it is frequently implemented in the clinical setting [18], [20], [21].

In terms of identifying swallowing disorders, age and sex based differences of the swallowing profile can potentially complicate a mathematical analysis. Our previous study [22] found that males and females produce different swallowing sound and vibration signals, particularly with regards to the frequency attributes of those signals. This means that if one were to use cervical auscultation to differentiate healthy and unhealthy swallows based on the values of these features, then separate algorithms may potentially need to be developed for male and female subjects. On the other hand, we found that participant age had a negligible impact on the values of our signal features [22]. Therefore, any similar classification task would only need to develop a single algorithm that would be applicable to subjects of any age.

Unlike our previous study [22] which focused on swallows made with a neutral head position, this study focuses exclusively on swallows made in the chin-tuck position. Otherwise, the goals of this paper are the same as our previous work. First, we sought to understand the differences between the two simultaneously recorded signal acquisition modalities (swallowing sounds and swallowing accelerometry signals) in the time, frequency, and time-frequency domains. We also sought to determine age and sex effects on the extracted signal features for both signal acquisition modalities. There are some notable differences in the anatomy of the neck and throat between the sexes, particularly in the size of the laryngeal prominence, which could affect either type of recording and should be investigated [23], [24]. Finally, we compare how our chosen features vary between swallows made in either a chin-tuck or neutral head position based on our previous findings [22]. This study should allow us to characterize various attributes of chin-tuck swallowing sounds and vibrations and contribute to the development of a new, non-invasive, instrumental swallowing screening procedure.

II. METHODOLOGY

A. Data collection equipment

Our recording equipment consisted of a dual-axis accelerometer and a contact microphone attached to the participant's neck with double-sided tape. The accelerometer (ADXL 322, Analog Devices, Norwood, Massachusetts) was mounted in a custom plastic case and affixed over the cricoid cartilage in order to provide the highest signal quality [25]. The two accelerometer axes were aligned parallel to the anterior plane of the neck (approximately parallel to the cervical spine) and perpendicular to the same surface (approximately perpendicular to the coronal plane). These axes are referred to as the superior-inferior (S-I) and anterior-posterior (A-P) axes, respectively, in the remainder of the manuscript. The sensor was powered by an external power supply (model 1504, BK Precision, Yorba Linda, California) with a 3V output, and the resulting signals were bandpass filtered from 0.1 to 3000 Hz
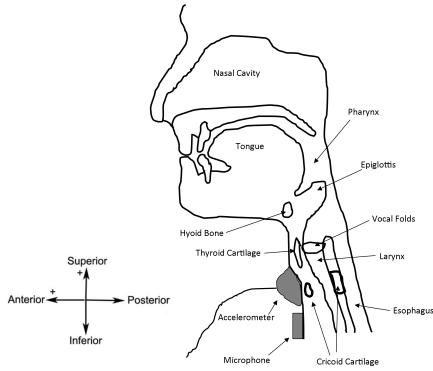
Fig. 1. Location of recording devices during data collection.

with ten times amplification (model P55, Grass Technologies, Warwick, Rhode Island). The voltage signals for each axis of the accelerometer were both recorded by a National Instruments 6210 DAQ and at a sampling rate of 40 kHz by the LabView program Signal Express (National Instruments, Austin, Texas). This set-up has been proven to be effective at detecting swallowing activity in previous studies [11], [26]. The microphone (model C 411L, AKG, Vienna, Austria) was placed below the accelerometer and slightly towards the right lateral side of the trachea, so as to avoid contact between the two sensors but record events from approximately the same location. This location has previously been described to be appropriate for collecting swallowing sound signals and can be seen in Fig. 1 [27], [28]. This microphone was chosen as it had a nearly flat frequency response for the entire range of human-audible sounds (10-20000 Hz [14]). The microphone was powered by a power supply (model B29L, AKG, Vienna, Austria) set to 'line' impedance with a volume of '9' and the resulting voltage signal was sent to the previously mentioned DAQ. Again, the signal was sampled by Signal Express at a rate of 40 kHz.

B. Subjects and protocol

The protocol for the study was approved by the Institutional Review Board at the University of Pittsburgh and participants were recruited from the neighborhoods surrounding the University of Pittsburgh Oakland campus. One participant's data was eliminated from our calculations due to mistakes made during recording, resulting in a total of 55 participants with useful data (28 males, 27 females. average age: 38.9 ± 14.9). All participants confirmed that they had no history of swallowing disorders, head or neck trauma or major surgery, chronic smoking, or other conditions which may affect swallowing performance. These participants were the same as

those used in our previous work on swallows made in a neutral head position [22]. All testing was performed in the iMED laboratory facilities at the University of Pittsburgh.

With their head in the chin-tuck position, each participant was asked to perform five saliva swallows. They were instructed to rest few seconds between each swallow to allow for saliva accumulation. Each unique task was recorded as a separate text file by the Signal Express software and imported into MATLAB (Mathworks, Natick, Massachusetts) for subsequent data processing.

C. Data pre-processing

At an earlier date, the device's baseline output was recorded and modified covariance auto-regressive modelling was used to characterize the noise of the recording system itself [29], [30]. The order of the model was determined by minimizing the Bayesian Information Criterion [29]. These autoregressive coefficients were then used to create a finite impulse response filter that could remove the noise inherit in our devices from the signal [29]. Afterwards, motion artifacts and other low frequency noise were removed from the signal through the use of least-square splines. Specifically, we used fourth-order splines with a number of knots equal to $\dfrac{Nf_l}{f_s}$, where N is the number of data points in the sample, f_s is the 40 kHz sampling frequency of our data, and f_l is equal to either 3.77 or 1.67 Hz for the superior-inferior or anterior-posterior signal, respectively. The values for f_l were calculated and optimized in previous studies [31]. After subtracting this low frequency motion from the signal we denoised the remaining data by using tenth-order Meyer wavelets with soft thresholding [32]. The optimal value of the threshold was determined through previous research to be $\sigma\sqrt{2\log N}$, where N is the number of samples in the data set and σ, the estimated standard deviation of the noise, is defined as the median of the down-sampled wavelet coefficients divided by 0.6745 [32]. Previous research by Wang and Willett demonstrated a useful method for segmenting data sets into two distinct categories based on local variances [33]. For this study, we applied a modified version of their method and used a proven two-class fuzzy c-means segmentation technique to identify the endpoints of each swallowing vibration in our continuously recorded signal [34].

The device noise filtering algorithm was recalculated with respect to the microphone system. Afterwards, a FIR filter was applied to the swallowing sound signal to eliminate the device's noise as done with respect to the accelerometer signals. We also applied the same 10 level wavelet denoising process to further refine the data. No splines or other low-frequency removal techniques were applied to the swallowing sound signals because we had not investigated if such frequencies contained important sound information. We did not develop new segmentation algorithms to extract the five individual swallows from the microphone signal, but instead simply used the time points given by the accelerometer segmentation process.

D. Feature extraction

Our next step involved extracting a number of signal features from our dual-axis swallowing accelerometry and swallowing sound signals. We chose several basic features from the time, frequency, and time-frequency domains in order to broadly characterize our signals' attributes. We used duration, skewness, kurtosis, entropy rate, and Lempel-Ziv complexity as our time domain features in order to provide information about the shape of our signals and the existence or absence of repeating patterns in a cervical auscultation signal. Our frequency domain features, peak frequency, center frequency, and bandwidth, were likewise chosen to characterize the spectral distribution of our signals in a simple and easy to interpret manner. As swallowing signals demonstrate some nonstationary properties and spectrogram representations are often used with cervical auscultation signals, we also calculated the wavelet energy distribution and wavelet entropy of our signals in order to summarize their time-frequency characteristics in an easy to interpret manner.

It should be noted that all of the following features were calculated independently for all three of our signals (anterior-posterior vibrations, superior inferior vibrations, and sounds). To avoid unnecessary repetition, however, the features will be described generically with the understanding that the relevant inputs to each equation are found with regards to only one of our three signals.

In the time domain, the signal skewness and kurtosis were calculated using the standard formulas [35], [36]. Calculating our third time domain feature, the swallow duration, only required converting the Matlab indices given in the segmentation step into proper time units.

To calculate the information-theoretic features we followed the procedures outlined in previous publications [11], [35]. The signals were normalized to zero mean and unit variance then divided into ten equally spaced levels, ranging from zero to nine, that contained all recorded signal values. We then calculated the entropy rate feature of the signals. This is found by subtracting the minimum value of the normalized entropy rate of the signal from 1 to produce a value that ranges from zero, for a completely random signal, to one, for a completely regular signal [11]. The normalized entropy rate is calculated as

$$NER(L) = \frac{SE(L) - SE(L-1) + SE(1) * perc(L)}{SE(1)} \quad (1)$$

where $perc$ is the percent of unique entries in the given sequence L [11]. SE is the Shannon entropy of the sequence and is calculated as

$$SE(L) = - \sum_{j=0}^{10^L - 1} \rho(j) \ln(\rho(j)) \quad (2)$$

where $\rho(j)$ is the probability mass function of the given sequence. Lastly the original signal was quantized again, but this time into 100 discrete levels. This allowed us to calculate the Lempel-Ziv complexity as

$$C = \frac{k \log_{100} n}{n} \quad (3)$$

where k is the number of unique sequences in the decomposed signal and n is the pattern length [37].

Next, in the frequency domain, we determined the bandwidth of the signals along with the center and peak frequencies. The center frequency was simply calculated by taking the Fourier transform of the signal and finding the weighted average of all the positive frequency components. Similarly, the peak frequency was found to be the Fourier frequency component with the greatest spectral energy. We defined the bandwidth of the signal as the standard deviation of its Fourier transform [11].

We also calculated a number of signal features in the time-frequency domain by utilizing a ten-level discrete Meyer wavelet decomposition [38], [39], [40], [41]. The energy in a given decomposition level was defined as

$$E_x = ||x||^2 \quad (4)$$

where x represents a vector of the approximation coefficients or one of the vectors representing the detail coefficients. $|| * ||$ denotes the Euclidean norm [11]. The total energy of the signal is simply the sum of the energy at each decomposition level. From there, we could calculate the wavelet entropy as the Shannon entropy of the wavelet transform. Applying 2 we produce the following expression:

$$WE = -\frac{Er_{a10}}{100} \log_2 \frac{Er_{a10}}{100} - \sum_{k=1}^{10} \frac{Er_{d_k}}{100} \log_2 \frac{Er_{d_k}}{100} \quad (5)$$

where Er is the relative contribution of a given decomposition level to the total energy in the signal and is given as [11]

$$Er_x = \frac{E_x}{E_{total}} * 100\% \quad (6)$$

E. Statistical analysis

Our statistical analysis involved transferring the processed features from Matlab to the SPSS (IBM, Armonk, New York) statistical analysis program. There we ran 16 Wilcoxon signed-rank tests, eight for each accelerometer axis, comparing the value of each relevant accelerometer attribute against the respective swallowing sound data. The swallow duration was left out of these tests since it was already assumed to be identical for each transduction method. A p-value of 0.003 or less was required for significance after applying the Bonferoni correction. We then ran 25 Wilcoxon rank-sum tests, eight for each signal plus one for the duration, to investigate possible sex-based differences in our recordings. In this situation the Bonferoni correction requires a p-value of less than 0.002 for statistical significance. Another 25 tests were run to compare the data gathered in this study to that gathered in [22] in order to characterize any position-dependent variations in our signals. Finally, linear regression curves with respect to participant age were fitted to the 25 signal features in order to characterize any potential age-related influences on our data.

III. RESULTS

Tables I-V and Figs. 2-4 summarize the results of our analysis in terms of each feature's mean and standard deviation.

TABLE I
TIME DOMAIN FEATURES IN THE CHIN-TUCK POSITION FOR MALES

	A-P	S-I	Sounds
Skewness	-0.759 ± 2.549	-0.203 ± 4.149	0.256 ± 5.427
Kurtosis	64.51 ± 238.7	98.15 ± 289.8	249.3 ± 899.3
Entropy Rate	0.989 ± 0.008	0.990 ± 0.007	0.988 ± 0.010
L-Z Complexity	0.057 ± 0.024	0.063 ± 0.029	0.071 ± 0.059
Duration (s)	2.486 ± 1.505		

TABLE II
TIME DOMAIN FEATURES IN THE CHIN-TUCK POSITION FOR FEMALES

	A-P	S-I	Sounds
Skewness	-0.240 ± 1.456	-0.855 ± 5.945	0.270 ± 2.791
Kurtosis	17.97 ± 55.28	81.39 ± 381.7	40.24 ± 301.8
Entropy Rate	0.991 ± 0.005	0.991 ± 0.004	0.988 ± 0.015
L-Z Complexity	0.063 ± 0.025	0.071 ± 0.023	0.074 ± 0.039
Duration (s)	2.395 ± 0.790		

TABLE III
A SUMMARY OF FREQUENCY DOMAIN FEATURES IN THE CHIN-TUCK
POSITION FOR MALES

	A-P	S-I	Sounds
Peak Frequency (Hz)	3.059 ± 3.613	5.735 ± 6.387	47.69 ± 426.8
Center Frequency (Hz)	59.00 ± 139.6	109.4 ± 227.8	704.6 ± 1630
Bandwidth (Hz)	111.7 ± 148.2	152.7 ± 229.3	917.0 ± 1303

TABLE IV
A SUMMARY OF FREQUENCY DOMAIN FEATURES IN THE CHIN-TUCK
POSITION FOR FEMALES

	A-P	S-I	Sounds
Peak Frequency (Hz)	2.270 ± 1.376	6.841 ± 8.697	14.55 ± 14.03
Center Frequency (Hz)	31.58 ± 127.1	45.28 ± 125.5	163.3 ± 468.2
Bandwidth (Hz)	65.78 ± 206.8	65.14 ± 132.0	387.1 ± 598.7

TABLE V
A SUMMARY OF TIME-FREQUENCY DOMAIN FEATURES IN THE
CHIN-TUCK POSITION

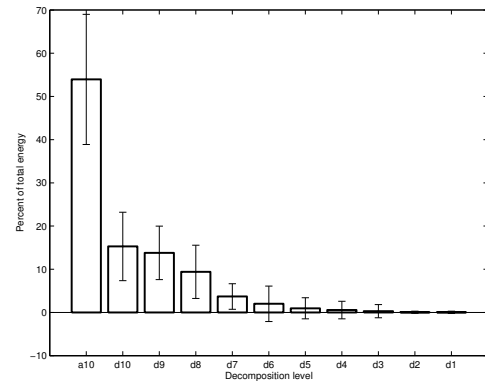|	A-P	S-I	Sounds
Wavelet Entropy Males	1.705 ± 0.701	1.960 ± 0.666	1.504 ± 0.694
Wavelet Entropy Females	1.429 ± 0.519	1.574 ± 0.658	1.163 ± 0.584

Fig. 2. Wavelet energy composition of S-I swallowing accelerometry signals

None of the regression analyses were able to account for more than 4% of the variation in any attribute with respect to participant age. In the majority of cases, the participant's age accounted for less than 1% of the variation of a feature. Due to the negligible amount of impact age had on our chosen signal features, we did not find it worthwhile to segregate our data into separate age categories. All further statistical analyses were carried out without controlling for the age of the participants.

Our statistical analysis demonstrated that only the anterior-posterior kurtosis ($p = 0.162$), the superior-inferior L-Z complexity ($p = 0.868$), and the skewness of both accelerometer signals ($p = 0.009$ and $p = 0.437$ for the A-P and S-I directions respectively) were not significantly different from the respective attributes for swallowing sounds. All of our chosen frequency domain features (peak frequency, center frequency and bandwidth), had significantly greater values for swallowing sounds when compared to the accelerometer signals ($p < 0.001$ for all). In the time domain we found that the kurtosis of swallowing sounds was generally greater than the superior-inferior accelerometer signal ($p < 0.001$) while the L-Z complexity was less when compared to the anterior-posterior accelerometer signal ($p < 0.001$). When compared to either accelerometer signal the entropy rate of swallowing sounds was notably less ($p < 0.001$). We also determined that our only time-frequency attribute, the wavelet entropy, was found to be significantly less for swallowing sounds than either the anterior-posterior ($p < 0.001$) or superior-inferior ($p < 0.001$) accelerometer signals.

This study found a number of attributes which differed between male and female participants. Both the bandwidth and center frequency of all three signals were significantly greater for male subjects ($p < 0.001$ for all). Male participants also produced signals with greater wavelet entropy when compared to females ($p < 0.001$ for all). Lastly, all of our recorded signals had greater kurtosis when recorded from male

participants ($p < 0.001$ for all). None of the remaining signal attributes demonstrated any significant differences between male and female participants.

Comparing the data collected in this study to that collected in [22], no significant differences in swallowing sounds due to the chin-tuck maneuver were detected, though there were a few notable differences for swallowing accelerometry. The anterior-posterior accelerometer signal did show significantly greater kurtosis ($p = 0.001$), peak frequency ($p < 0.001$), and wavelet entropy ($p = 0.001$) in the chin-tuck position when compared to neutral swallows. The superior-inferior accelerometer signal likewise demonstrated a greater kurtosis ($p < 0.001$), center frequency ($p < 0.001$), and bandwidth ($p < 0.001$) in the chin-tuck position.

Figs. 2-4 show the average energy distribution of the wavelet coefficients of all three signals. They all show that the vast majority of swallowing energy is contained in the lowest
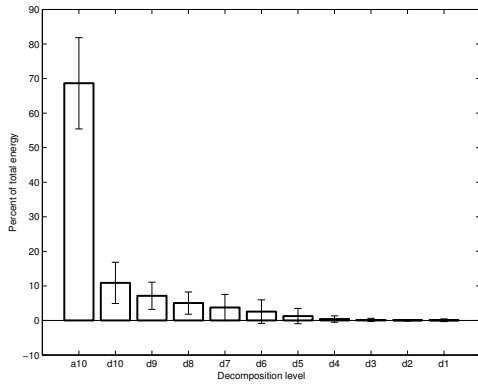
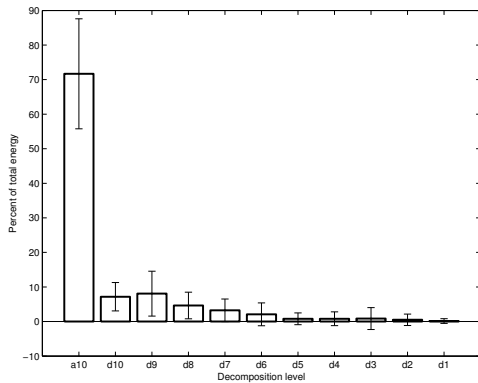Fig. 3. Wavelet energy composition of A-P swallowing accelerometry signals

Fig. 4. Wavelet energy composition of swallowing sounds

frequency components. The a_{10}, d_{10}, and d_9 bands of all three of our signals, corresponding to wavelet frequency components below 80 Hz, contain over 80% of each signal's energy.

IV. DISCUSSION

A. Age and sex effects in the chin-tuck position

Frequency features of all three of our signals demonstrated statistically significant variation between male and female participants. Specifically, the center frequency and bandwidth of all three signals were greater in male participants using the chin-tuck position compared to females. The wavelet entropy for all of our signals was also greater in men, resulting in more chaotic and less predictable data, and is most likely related to the signals' shifts towards higher frequencies. We suspect that these differences may be due to the sex-based variations in the size and position of the laryngeal prominence, since our recording devices were placed just below this structure [23]. The laryngeal prominence tends to protrude further in males, yet undergoes the same motion during a swallow as females [24], [42]. This could produce higher frequency vibrations in male subjects as tissues are displaced faster in the anterior-posterior direction to accommodate the larger moving structure. It is interesting to note that in our previous work, which focused only on swallows made in the neutral position, we found that the center frequency of swallowing vibrations and sounds also depended on the subject's sex [22]. However, in that study, only the superior-inferior accelerometer bandwidth

showed such variation [22]. While this may simply be a recording artifact due to the different position of the head and relative direction of gravity, it may also be a result of differences in the execution of the chin-tuck maneuver in males and females and deserves further investigation.

In the time domain, we also found that the kurtosis of all three signals is greater in men. A higher kurtosis implies that the signal has a greater intensity over a shorter period of time. This conforms to our earlier findings with respect to our frequency domain features, as signals with greater kurtosis tend to include high frequency pulses [36]. This finding adds validity to our previous statistical conclusions.

Lastly, we found that the duration of each individual swallow as determined by our algorithm did not vary with the sex of the participant, nor did it demonstrate any significant dependency on the subject's age. Both of these findings are contrary to past research on the topic [43]. However, these findings are identical to what we found in our previous work on swallows in the neutral position and it is reasonable to assume that the discrepancies are due to the same variations between this and past studies [22]. It is also possible that this is simply the result of using different segmentation schemes. Our study used a clustering algorithm to identify the endpoints of each swallow while other studies may use concurrent videofluoroscopic imaging or another technique to perform the same task. This could easily result in slightly different start and stop times for each swallow. In summary, we believe that our loss of sex dependency on the swallow duration is a result of using different processing and segmentation techniques than these past studies while the loss of age dependence is due to our small sample size and comparatively low statistical power.

In terms of identifying swallowing disorders, these age and sex based differences serve to complicate any future classification work. Our study found that males and females produce different sound and vibration signals, particularly with regards to the frequency attributes of those signals. This means that if one were to use cervical auscultation to differentiate healthy and unhealthy swallows based on the values of these features, then separate algorithms to do so would need to be developed for male and female subjects. A classification algorithm trained on only male or only female subjects would produce incorrect results when used on a member of the opposite group, as the 'normal' range of feature values are not identical for both groups. Meanwhile, training it on a mixed population would make the algorithm too conservative to be an effective screening tool. Such a procedure would not need to be followed for subjects of varying ages, however, as we found that participant age had a negligible impact on the values of our signal features. Therefore, any similar classification task would only need to develop a single algorithm that would be applicable to subjects of any age.

B. Comparing swallowing sounds and swallowing accelerometry signals in the chin-tuck position

The statistical differences found between our vibration and sound signals in the time domain were similar to those found in our previous work on swallows made in a neutral head position

[22]. Our analysis showed that only the superior-inferior accelerometer signal had a lower kurtosis than the microphone signal, indicating that the energy of this accelerometer axis is statistically less temporally focused [36]. Meanwhile, only the anterior-posterior accelerometer signal had a lower Lempel-Ziv complexity when compared to the same data, which indicates that this signal can be compressed further without losing information about the signal [44]. We also found that both accelerometer signals had slightly greater entropy rate features than the microphone signal. However, just like in our previous study, these differences are very small and very close to a value of 1, and so indicate that both swallowing vibrations and sounds follow a predictable pattern when discretized to ten levels [22].

Again, our frequency feature contrasts are similar to those found in our previous work with swallows in the neutral position [22]. We found that swallowing sounds contain higher frequency components than vibrations along either accelerometer axis, indicating that there are higher frequency components of the swallowing event that only one transduction method detected. However, our peak and center frequencies are still significantly lower than what was reported in past studies on swallowing activity [10], [25], [26]. In summary, we concluded that this discrepancy is a result of using recording and segmentation methods that more effectively utilize the lower end of the frequency spectrum [22].

The wavelet energy plots (Figs. 2-4) are distributed as one would expect. The energy follows an exponentially decaying pattern as the frequency band increases, just as seen for data in the neutral position [22]. The overwhelming majority of the energy for all three signals is concentrated in the lowest frequency components, which conforms with the time-scale of swallowing and the temporal dynamics of the corresponding physiology [45].

Ultimately, these results demonstrate that swallowing vibrations and swallowing sounds are distinct signals and cannot be arbitrarily compared. Our study, which utilized accepted transduction and filtering techniques, found that these signals varied significantly across multiple attributes and domains. While they share certain physical similarities and are likely produced by the same events in the body they cannot be considered identical for the purposes of swallowing assessment. These conclusions match our findings from our previous study with swallows made in a neutral head position [22].

C. Head position effects on swallowing sounds and accelerometry

We can draw several interesting conclusions by comparing our data from chin-tuck swallows with data corresponding to swallows made in a neutral position. The first, is that the chin-tuck position does not change the acoustic profile of a swallow. Although we know that this technique does change the subject's physiology and provides greater protection of the airway in some patients, we were unable to detect any changes with our microphone set-up. Our statistical analysis of swallowing sounds demonstrated that none of our chosen features varied significantly due to the subject's head position. This indicates that either the mechanism that produces swallowing sounds is unaffected by the physiological changes that occur in the chin-tuck position or that our microphone's position was unable to detect any changes to the swallowing sounds. Given that the larynx, which our microphone was in contact with and recording from, shows little if any deformation during this maneuver we can assume that our recording method is adequate. This fact has been demonstrated repeatedly for normal swallows. Furthermore, our earlier analysis showed that the vast majority of swallowing energy is in the lowest frequencies. Therefore, swallowing sounds have an average wavelength of several meters. The larynx and pharynx would need to undergo much more drastic conformational changes in shape while using the chin-tuck position than those reported in order to significantly modify such large waveforms through reflection or diffraction [46], [47]. In summary, we have found no differences between swallowing sounds produced in the neutral position and swallowing sounds made in the chin-tuck position.

Based on our statistical analysis we do not see any evidence of the relative angle of gravitational pull affecting our recordings. If the angle of our accelerometer relative to gravity was significant then we would expect to see complementary changes in one or more of our chosen features, where the feature for the signal from one accelerometer axis would increase and the other decrease, when comparing neutral and chin-tuck data. This would be the result of one axis more closely aligning with the direction of the gravitational pull while the other axis is brought out of alignment. However, we found no complementary changes in any of our accelerometer features and so conclude that the significant features in our comparison are not simply an artifact of the angle of the device relative to gravity.

We did find several features that varied significantly between the neutral and chin-tuck conditions. In particular, the kurtosis of both the anterior-posterior and superior-inferior accelerometer signals was significantly greater in the chin-tuck position. Much like when comparing sex-dependant effects, this demonstrates that the energy of swallowing vibrations are more temporally focused for both accelerometer axes. We also found that the peak frequency of the anterior-posterior accelerometer signal as well as the center frequency and bandwidth of the superior-inferior accelerometer signal are greater in the chin-tuck position. From these findings, we can see that swallows made in the chin-tuck position produce higher frequency vibrations than those in the neutral position. This shift is most likely a result of the physiological changes made when entering the chin-tuck position, though the exact source of the different vibrations cannot be determined through this study.

V. CONCLUSION

In this study, we recorded data from healthy adult subjects performing saliva swallows in the chin-tuck position using both a dual-axis accelerometer and a contact microphone. None of our chosen time or frequency domain features showed any significant dependence on the subject's age, but several of

these features did vary between male and female participants. We also found that the majority of these features were significantly different between swallowing sounds and swallowing vibrations. These findings complement our previous work on swallows made in a neutral position and further support the idea that, despite their similar transduction methods, swallowing vibrations and sounds contain unique information about the underlying physiology. Finally we found that, when compared to neutral swallows, swallowing sounds do not significantly change when recorded in the chin-tuck position while a number of time and frequency features did significantly vary for swallowing vibrations. These findings provide important insights for designing a method to characterize swallowing performance based solely on cervical auscultation.

ACKNOWLEDGEMENT

REFERENCES

[1] M. Spieker, "Evaluating dysphagia," *American Family Physician*, vol. 61, no. 12, pp. 3639–3648, June 2000.

[2] D. Smithard *et al.*, "Complications and outcome after acute stroke. does dysphagia matter?" *Stroke*, vol. 27, no. 7, pp. 1200–1204, July 1996.

[3] P. Leslie *et al.*, "Cervical auscultation synchronized with images from endoscopy swallow evaluations," *Dysphagia*, vol. 22, no. 4, pp. 290–298, October 2007.

[4] J. Coyle *et al.*, "Oropharyngeal dysphagia assessment and treatment efficacy: Setting the record straight (response to campbell-taylor)," *Journal of the American Medical Directors Association*, vol. 10, no. 1, pp. 62–66, January 2009.

[5] B. Sherman *et al.*, "Assessment of dysphagia with the use of pulse oximetry," *Dysphagia*, vol. 14, no. 3, pp. 152–156, June 1999.

[6] C. Ertekin *et al.*, "Electrodiagnostic methods for neurogenic dysphagia," *Electroencephalography amd Clinical Neurophysiology/Electromyography and Motor Control*, vol. 109, no. 4, pp. 331–340, August 1998.

[7] P. Leslie *et al.*, "Reliability and validity of cervical auscultation: A controlled comparison using videofluoroscopy," *Dysphagia*, vol. 19, no. 4, pp. 231–240, 2004.

[8] J. M. Fontana *et al.*, "Swallowing detection by sonic and subsonic frequencies: A comparison," in *Proc. of 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC, 2011)*, Boston, MA, USA, Aug./Sep. 30–3, 2011, pp. 6890–6893.

[9] E. Sazonov *et al.*, "Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior," *Physiological Measurement*, vol. 29, no. 5, pp. 525–541, May 2008.

[10] A. Santamato *et al.*, "Acoustic analysis of swallowing sounds: A new technique for assessing dysphagia," *Journal of Rehabilitation Medicine*, vol. 41, no. 8, pp. 639–645, March 2009.

[11] J. Lee *et al.*, "Effects of stimuli on dual-axis swallowing accelerometry signals in a healthy population," *Biomedical Engineering Online*, vol. 9, no. 7, pp. 1–14, February 2010.

[12] V. Gupta *et al.*, "Acceleration and EMG for sensing pharyngeal swallow," in *Proc. of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 1993)*, San Diego, CA, USA, Aug. 28–31, 1993, pp. 1221–1222.

[13] *ADXL322: Small and Thin 2g Accelerometer Data Sheet*, 6th ed., Analog Devices, 2005.

[14] *MicroMic, The Original: C411*, 1st ed., AKG Acoustics, 1991.

[15] T. K. Kim *et al.*, "Comparison of an accelerometer and a condenser microphone for mechanomyographic signals during measurement of agonist amd antagonist muscles in sustained isometric muscle contractions: The influence of the force tremor," *Journal of Physiological Anthropology*, vol. 27, no. 3, pp. 121–131, May 2008.

[16] N. Audibert and A. Amelot, "Comparison of nasalance measurements from accelerometers and microphones and preliminary development of novel features," in *Proc. of 12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011)*, Florence, Italy, Aug. 27–31, 2011, pp. 1–4.

[17] T. K. Shanahan *et al.*, "Chin-down posture effect on aspiration in dysphagic patients," *Archives of Physical Medicine and Rehabilitation*, vol. 74, no. 4, pp. 736–739, July 1993.

[18] J. Lewin *et al.*, "Experience with the chin tuck maneuver in postesophagectomy aspirators," *Dysphagia*, vol. 16, no. 3, pp. 216–219, June 2001.

[19] M. Bulow *et al.*, "Videomanometric analysis of supraglottic swallow, effortful swallow and chin tuck in healthy volunteers," *Dysphagia*, vol. 14, no. 2, pp. 67–72, March 1999.

[20] M. Welch *et al.*, "Changes in pharyngeal dimensions effected by chin tuck," *Archives of Physical Medicine and Rehabilitation*, vol. 74, no. 2, pp. 178–181, February 1993.

[21] J. Logemann, "Noninvasive approaches to deglutitive aspiration," *Dysphagia*, vol. 8, no. 4, pp. 331–333, September 1993.

[22] J. M. Dudik *et al.*, "A comparative analysis of swallowing accelerometry and sounds during saliva swallows," *Biomedical Engineering Online*, 2014, accepted for publication.

[23] H. Gray, *Anatomy of the Human Body*, 20th ed., W. Harmon, Ed. Lea & Febiger, 1918.

[24] V. Kuhl *et al.*, "Sonographic analysis of laryngeal elevation during swallowing," *Journal of Neurology*, vol. 250, no. 3, pp. 333–337, March 2003.

[25] S. Youmans and J. Stierwalt, "An acoustic profile of normal swallowing," *Dysphagia*, vol. 20, no. 3, pp. 195–209, July 2005.

[26] S. Hamlet *et al.*, "Stethoscope acoustics and cervical auscultation of swallowing," *Dysphagia*, vol. 9, no. 1, pp. 63–68, January 1994.

[27] K. Takahashi *et al.*, "Methodology for detecting swallowing sounds," *Dysphagia*, vol. 9, no. 1, pp. 54–62, January 1994.

[28] J. Cichero and B. Murdoch, "Detection of swallowing sounds: Methodology revisited," *Dysphagia*, vol. 17, no. 1, pp. 40–49, January 2002.

[29] E. Sejdić *et al.*, "Baseline characteristics of dual-axis cervical accelerometry signals," *Annals of Biomedical Engineering*, vol. 38, no. 3, pp. 1048–1059, March 2010.

[30] L. Marple, "A new autoregressive spectrum analysis algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 4, pp. 441–454, August 1980.

[31] E. Sejdić *et al.*, "A method for removal of low frequency components associated with head movements from dual-axis swallowing accelerometry signals," *PLoS ONE*, vol. 7, no. 3, pp. 1–8, March 2012.

[32] ——, "A procedure for denoising of dual-axis swallowing accelerometry signals," *Physiological Measurements*, vol. 31, no. 1, pp. N1–N9, January 2010.

[33] Z. Wang and P. Willett, "Two algorithms to segment white gaussian data with piecewise constant variances," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 373–385, February 2003.

[34] E. Sejdić *et al.*, "Segmentation of dual-axis swallowing accelerometry signals in healthy subjects with analysis of anthropometric effects on duration of swallowing activities," *IEEE Transactions of Biomedical Engineering*, vol. 56, no. 4, pp. 1090–1097, April 2009.

[35] J. Lee *et al.*, "Time and time-frequency characterization of dual-axis swallowing accelerometry signals," *Physiological Measurement*, vol. 29, no. 9, pp. 1105–1120, August 2008.

[36] B. Everitt and A. Skrondal, *The Cambridge Dictionary of Statistics*, 4th ed. Cambridge University Press, October 2010.

[37] M. Aboy *et al.*, "Interpretation of the lempel-ziv complexity measure in the context of biomedical signal analysis," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2282–2288, November 2006.

[38] E. Sejdić *et al.*, "Time-frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Processing*, vol. 19, no. 1, pp. 153–183, Jan. 2009.

[39] S. Stanković *et al.*, *Multimedia Signals and Systems*. New York, NY: Springer US, 2012.

[40] A. Cohen and J. Kovačević, "Wavelets: The mathematical background," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 514–522, Apr. 1996.

[41] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[42] T. Burnett *et al.*, "Laryngeal elevation achieved by neuromuscular stimulation at rest," *Journal of Applied Physiology*, vol. 94, no. 1, pp. 128–134, September 2003.

[43] J. Cichero and B. Murdoch, "Acoustic signature of the normal swallow: Characterization by age, gender, and bolus volume," *Annals of Otology, Rhinology, and Laryngology*, vol. 11, no. 7, pp. 623–632, July 2002.

[44] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 75–81, January 1976.

[45] J. Tracy *et al.*, "Preliminary observations on the effects of age on oropharyngeal deglutition," *Dysphagia*, vol. 4, no. 2, pp. 90–94, April 1989.

[46] B. Crowell, *Light and Matter*, 1st ed. Benjamin Crowell, Creative Commons, June 1998, vol. 1, no. 1.

[47] D. Brewster, *A Treatise on Optics*, 1st ed. London: Printed by A&R Spottiswoode, New-Street-Square, February 1831, vol. 1, no. 1.