SwallowNet: Recurrent Neural Network Detects and Characterizes Eating Patterns

Dzung Tri Nguyen*, Eli Cohen*, Mohammad Pourhomayoun^, Nabil Alshurafa*[†]

*Electrical Eng. Comp. Science Dept.
Northwestern University
Evanston, IL, USA

^Computer Science Dept.
CSULA
Los Angeles, CA, USA

[†]Dept. Preventive Medicine
Northwestern University
Chicago, IL, USA

***Abstract*—Passively detecting and counting the number of swallows in food intake enables accurate detection of eating episodes in free-living participants, and aids in characterizing eating episodes. On average, the more food consumed, the greater the number of swallows; and swallows have been shown to positively correlate with caloric intake. While passive sensing measures have shown promise in recent years, they are yet to be used reliably to detect eating, impeding the development of timely intervention delivery that change poor eating behavior. This paper presents a novel integrated wearable necklace that comprises two piezoelectric sensors vertically positioned around the neck, an inertial motion unit, and long short-term memory (LSTM) neural networks to detect and count swallows. A unique correlation of derivative features creates candidate swallows. To reduce the FPR features are extracted using symmetric and asymmetric windows surrounding each candidate swallow to feed into a Random Forest classifier. Independently, a LSTM network is trained from raw data using automated feature learning methods. In an in-lab study comprising confounding activities of 10 participants, results show a 3.34 RMSE of swallow count using LSTM, and a 76.07% average F-measure of swallows, outperforming the Random Forest classifier. This system thus shows promise in accurately detecting and characterizing eating patterns, enabling passive detection of swallow count, and paving the way for timely interventions to prevent problematic eating.**

***Index Terms*—Eating Detection; Wearable; Piezoelectric; Inertial Motion Unit; Deep Learning; Recurrent Neural Network.**

I. Introduction and related work

Employing passive sensors in wearable devices to detect and characterize episodes of eating has been an important research challenge to reduce the burden of participant self-report. There are many aspects of the eating process that can be characterized, such as hand-to-mouth gestures, bites, chews, and swallows. Based on these building blocks, higher level semantic information can be inferred such as the mass of ingested food, caloric intake and ultimately eating behavior. This paper focuses on the detection and characterization of swallows, which has been shown to positively correlate with caloric intake [1].

Miniature, low-power sensors are a key component that makes swallow detection feasible. Based on the type of sensors, swallow detection systems can be categorized as acoustical or electrical. The acoustical approaches use a microphone to capture swallow sounds. The microphone can be placed directly at the throat region [2], [3], or in the ear canal capturing bone conduction [4]. It is understood that audio sensors are affected by environmental noises, especially when noises are in the same frequency range as the desired signal, a challenge commonly known as ”the cocktail party problem.” Rahman et al. [3] design a special piezoelectric microphone system that reduces the interference of environmental noises. Päßler et al. [5] propose an additional microphone to capture environmental noise only, as a reference to denoise the signal recorded from the in-ear microphone.

The other approach uses mechanoelectrical sensors placed directly around the neck to track skin motion and muscle activation during ingestion. Amft and Troster [2] measure electromyography signals of muscle activation at infrahyoid and submental positions. Kalatarian et al. [6], [7] place piezoelectric film sensor at the lower region of the neck to detect and classify swallows into different food categories. Piezoelectric sensors are less intrusive than acoustic ones and can be effective even in noisy environments.

Recognition algorithms play an equally important role in detecting swallows. A tutorial by Bulling et al. [8] sketches out a sample data processing pipeline for human activity recognition. While the domain of the tutorial is different from our application, it presents a summary of the standard data processing techniques used to detect activities from time-series based signals. However, every system adopts its own unique approaches to preprocessing, feature extraction and classification, depending on their intended outcome. In particular, authors in [9] propose the usage of statistical features from a spectrogram to perform feature extraction of piezoelectric signals.

This paper focuses on a new design using piezoelectric sensors to detect and count swallows. The main challenges presented by piezoelectric sensors is that the signal can easily be affected by head movement, talking, and even chewing. This paper provides novel methods that help address those problems both in hardware through a combination of multiple sensors, and in software using advanced statistical machine learning and deep learning. The main contributions of this paper comprise:

- The design of a wearable necklace using two piezoelectric sensors and an inertial motion unit to capture swallows during eating episodes.
- Segmenting candidate swallows using a unique out-of-phase feature calculated from piezoelectric signals. Features are extracted in multiple symmetric and asymmetric

windows surrounding each candidate to capture different parts of a swallow.
- SwallowNet: a recurrent neural network framework that detect swallows on a continuous data stream after being trained purely from raw data using automated feature learning methods.

II. The necklace

A. *Multiple piezoelectric sensor configuration*

Prior literature of swallow detection using piezoelectric sensors have focused on using a single sensor [10] [11] [7]. Authors [7] further investigate placement of a single piezoelectric sensor at different positions on the neck, and conclude that measuring signal at lower regions of the neck provides the highest accuracy.

The process of swallowing is a complex neuromuscular activity comprising oral, pharyngeal and esophageal phase. During the pharyngeal phase, the laryngeal closure occurs to prevent aspiration during swallows, followed by hyoid elevation, and then the food moves to the esophageal phase. During this transition a differential in pressure is sensed between the upper and lower part of the neck. As a result, this effort presents a neckworn sensor combining multiple signals from different positions on the neck to capture this differential and reliably detect swallows. This paper refers to the top signal as the laryngeal signal, and the bottom as the esophageal signal. At the beginning of a swallow, the trachea presses on the bottom piezoelectric sensor and creates a peak in the bottom signal. This bottom peak is followed by another peak in the top signal approximately one second later when the larynx presses on the top piezoelectric sensor. Thus during a swallow, there are several alternating peaks from both the larynx and esophageal signal, spaced by a constant time. This is illustrated in Figure 1, where vertical lines are placed at alternating top and bottom peaks.

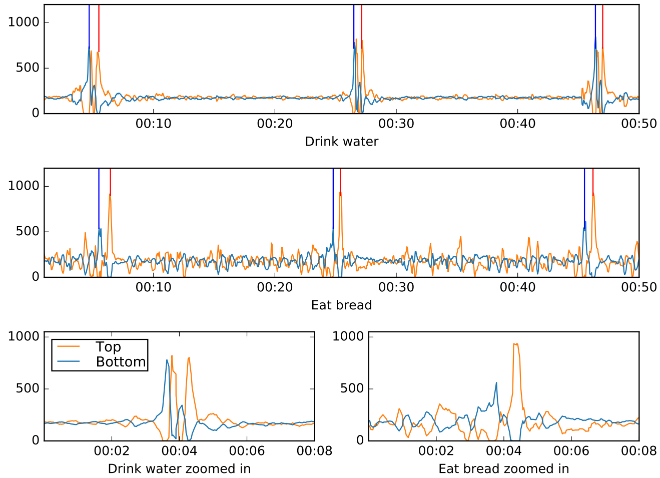

Fig. 1: Piezoelectric signals during three consecutive swallows of drinking water and eating bread (top and middle), and the close up version of individual swallow (bottom).

Only the swallow produces this distinct pattern. Other actions such as chewing, head movement, or speaking do alter individual signals, however they do not generate alternating top and bottom peaks. For example, when a subjects speaks, the top and bottom sensors are usually activated at the same time.

B. *Inertial motion sensors*

Beside the piezoelectric sensors, an inertial motion unit (IMU) is used to track movement of the neck. The IMU combines an accelerometer, gyroscope, and magnetometer to monitor position and orientation of the neck. Theoretically, the absolute position and orientation can be recovered through double integration of acceleration and single integration of angular velocity. However, the addition of noise from motion and the sensor make this problem challenging and require fusion algorithms [12] as integration of noisy signals drift over time. This paper uses the BNO055 IMU from BOSCH, which combines motion sensors and a processor running a fusion algorithm in one single chip. Absolute orientation in the form of quaternion and linear acceleration (without gravity effects) are obtained from the BNO055.

C. *The integrated necklace*

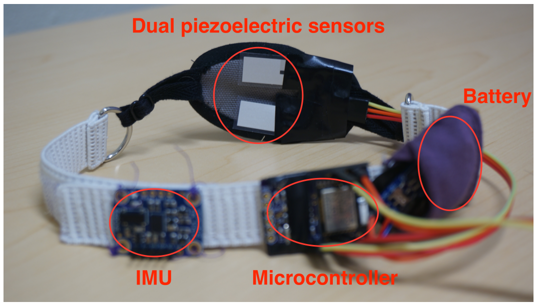

Fig. 2: The necklace: piezoelectric sensors in the middle, IMU in the back, microcontroller and battery on the side.

Piezoelectric sensors and the IMU are mounted on a wearable necklace, as in Figure 2. The necklace is worn low on the neck as opposed to around the larynx, which would impact aesthetic perception and user comfort.

Data is acquired at a sampling rate of 100Hz by a microcontroller running a 16MHz ARM Cortex-M0 processor. Piezoelectric sensors are sampled through an internal analog-to-digital converter, while the IMU transfers data directly through I^2C interface. The microcontroller then converts data into a custom defined format and transmit data to a nearby client using Bluetooth Low Energy (BLE). Since the maximum size of a BLE packet is 20 bytes and our data exceeds this limit (32 bytes), we perform a lossy compression through the removal of least significant bits.

The following sections describe two approaches to analyzing the data from the necklace. The Statistical Machine Learning approach (SML) applies preprocessing, segmentation, feature extraction, feature selection, classification, and fusion algorithms to train and test a predictive model. The Deep Learning approach (SwallowNet), however, trains completely from raw data, empowering the neural network to build its own internal representation of the features and classification boundaries of a swallow.

III. Statistical Machine Learning

The diagram in Figure 3 summarizes the main stages of the data processing pipeline for the SML approach. Firstly, signal processing algorithms are applied to the raw data to denoise and normalize the signals. Secondly preprocessed data from two piezoelectric sensors are fed through a segmentation algorithm to find candidate swallows. After that, a machine learning algorithm is used to build a model that classifies swallow candidates into true (positive) parts of swallows and false (negative) swallows which correspond to other activities such as head motion or talking). Finally, a postprocessing fusion step merges the nearby positive candidates to produce a single prediction for each swallow.

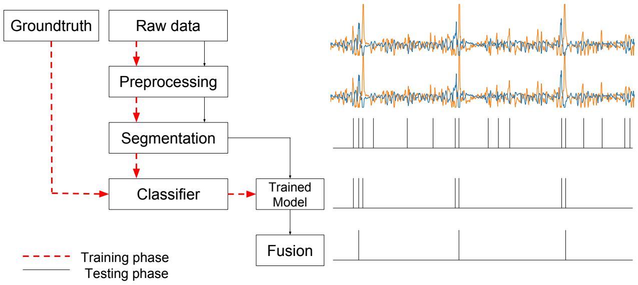

Fig. 3: The data processing pipeline

A. Preprocessing

Data from two piezoelectric sensors and the IMU are smoothed using a Savitzky-Golay filter, which has been shown to be effective in retaining representative peaks of swallows while increasing the signal-to-noise ratio in data from piezoelectric sensors [13]. Empirical findings show best performance using a window size of 9 (.09s) and a low degree polynomial order of 3.

B. Segmentation

Given a continuous data stream, the segmentation algorithm finds the candidate's moments where swallows most likely occur. The algorithm might return multiple candidates for one swallow, or negative candidates created by other activities, but it should identify a large part of the signal that corresponds with a swallow. Thus the most important evaluation for the segmentation algorithm is the recall rate, which is the ratio of number of true swallows (that has at least one corresponding candidate swallow identified by the segmentation algorithm) to the total number of swallows.

While a sliding window approach is often used for segmentation, prior to feature extraction, this approach generates an imbalanced dataset since the proportion of swallows to non-swallow events is small. To minimize the number of negative candidates, another preprocessing step is performed to further amplify the distinction of a swallow from a non-swallow.

As discussed above, a swallow happens when the top and bottom piezoelectric signals exhibit alternating peaks. As shown in Figure 1, when the top signal increases, the bottom one decreases, and vice versa, creating an out-of-phase effect. This phenomenon can be captured quantitatively by calculating the correlation of derivatives between the two signals as in the following equation:

$$corr_t = -\frac{dTop}{dt}\frac{dBottom}{dt}$$
$$= -\frac{(Top_t - Top_{t-1})}{dt}\frac{(Bottom_t - Bottom_{t-1})}{dt}$$

A local maxima finding algorithm with the look up window of 16 samples (0.16 seconds) is then applied to the correlation of derivatives. These local maximas represent the center of the swallow candidates. Figure 4 visualizes the segmentation for drinking water and eating bread. While there are some negative candidates due to chewing or head movement (false positives), all of the true swallows are captured. In the next step, the system relies on the classifier to reduce the false positive rate.

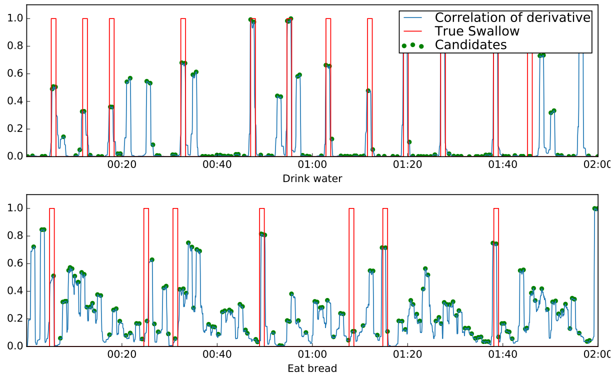

Fig. 4: Correlation of derivatives. Local maximas (green points) are swallowing candidates. Note that eating bread has many false positive candidates due to chewing, which is filtered through feature extraction and classification in the following stage.

C. Feature Extraction

In this step, the SML model is built to differentiate between positive and negative swallow candidates.

1) Multiple windows for feature extraction: Feature extraction is performed in multiple symmetric and asymmetric windows of different sizes around each candidate to capture the different temporal stages of a swallow. For example, a window of [-3s, 0s] before the swallow captures a drop in the lean forward angle which often precedes a swallow.

2) Lean forward angle from IMU: The IMU returns absolute orientation in the form of quaternion. The quaternion is a four dimensional vector $\boldsymbol{q}$ representing the rotation axis and the angle of rotation around that axis. The quaternion can be projected into different planes to gain physical angles and infer activities such as leaning forward and to the side, and to determine the orientation that the subject is facing. Not all of these angles are related to the eating process, however. The most informative one is the Lean Forward Angle (LFA), the angle between the IMU and Earth's surface for example, when the subject sits straight, the LFA is close to 90 degrees. LFA is calculated by applying the dot product of the normal vectors of two planes:

$$LFA = acos < \boldsymbol{n_1}, \boldsymbol{n_2} >$$

where the normal vector of Earth's surface is simply the z-axis, and the normal vector of the IMU is obtained through the quaternion transformation:

$$\boldsymbol{n_1} = [0, 0, 1] \qquad\qquad \boldsymbol{n_2} = \boldsymbol{q n_1 q^{-1}}$$

where q is a unit quaternion that rotates n_1 to obtain the normal vector of the IMU.

3) Features: Table I summarizes the statistical features extracted from the signals, which have been shown to be useful in representing time-series signals [8] [9]. We extract these features for each of the laryngeal, esophageal, and LFA signals and acceleration energy within each window as described in Section C1. A correlation-based feature subset selection (CFS-Subset) algorithm is applied to obtain nineteen optimal features for swallow detection.

Statistical features	Time series features
Mean, variation	Count above mean
Median	Count below mean
Max, min	First location of maximum
Skew	First location of minimum
RMS	Longest strike above mean
Kurtosis	Longest strike below mean
1st, 3rd quantile	Number CWT peaks
Inner quantile range	Number of peaks
	Symmetry looking
	Polynomial fitting features

TABLE I: List of features

D. Classification and Fusion

A Random Forest classifier is trained on the generated features. A Random Forest model comprises multiple decision tree classifiers which are trained on different subset of the features, and corrects for decision trees habit of overfitting the training set.

There might be multiple candidate swallows corresponding to a single swallow since there are multiple peaks resulting from the top-bottom out-of-phase shift. The fusion could have been done prior to the classification stage, however, the presence of negative candidates makes merging more challenging at that point, so the fusion is done post-classification. A mean-shift clustering algorithm [14] is used since it does not require prior knowledge of the number of clusters and does not constrain the shape of the clusters.

IV. SwallowNet

The data from the necklace belongs to the sequential data category. It has variable length, and contains temporal dependencies across different data channels. In the previous section, the temporal dynamics are represented through the usage of multiple symmetric and asymmetric feature extraction windows. However, these windows are fixed and thus sometimes might miss a long range interaction, and at other times might be redundant.

A recurrent neural network model [15] is designed specifically for sequential data. RNN models can be trained on one set of input sequence, and then generalized to a different length test sequence. RNN achieves this property through the inclusion of cycles in its computation graph, and also sharing of parameters across time.

The following particular implementation of RNN is utilized to capture swallows from the continuous data stream of the necklace. First, supervised sequence labeling models are used [16], where the output sequence has the same length as the input sequence. Second, long short-term memory (LSTM) is used as the recurrent layer to avoid the vanishing gradient problem common when applying RNN [17]. LSTM has been employed successfully in many applications related to sequential data such as speech recognition [18], video action recognition [19] and wearable action recognition [20].

It is enticing to build a complicated model with multiple recurrent and constitutional layers. The necklace data, however, is in fact simple and does not have many states. When the subject rests, the piezoelectric signals always go back to the original value of 0.5V. When the subject starts eating, the signal alternates between the state of chewing, head movement or swallows. The swallowing pattern does not exhibit large variability in time since it is difficult to swallow faster or slower involuntarily. The temporal dynamics of piezoelectric signals and the lean forward angle are also much shorter in range compared to speech or wearable gesture signals. Thus the neural network model SwallowNet is designed to have a single recurrent layer combined with one nonlinear transformation layer for feature extraction.

A. Network architecture

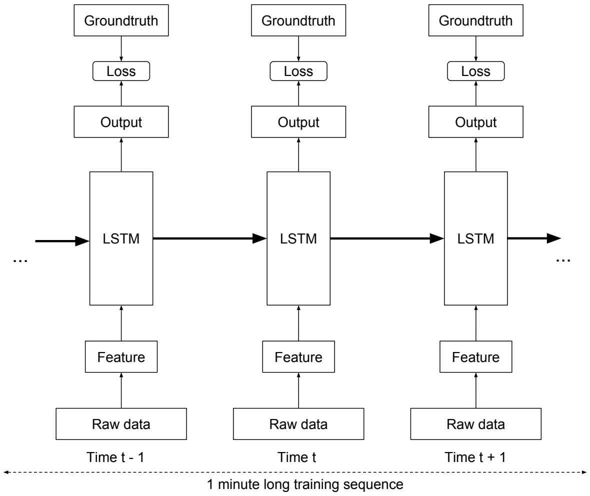

Fig. 5: SwallowNet architecture

Figure 5 shows the architecture of SwallowNet. The data stream is split into chunks of 20 samples (0.2 second), which spans approximately one fifth of one single swallow. These chunks are then transformed through a nonlinear embedding layer which resembles feature extraction:

$$e_t = relu(W_f * x_t + b_f)$$

where $relu$ is the rectified linear unit activation function, W_f, b_f are the parameters of the embedding layer, and x_t, e_t are the original data and output of the embedding layer respectively. Feature extraction are no longer fixed functions, instead the weights W_f, b_f are updated during the training process. The network learns the optimal representation to differentiate between a swallow and a non-swallow chunk.

Feature f_t is then fed into the LSTM layer to learn the temporal dynamics of the signals. A LSTM layer has internal state C_t and output h_t, which are updated recurrently throughout the sequence. LSTM utilizes forget gate f_t, input gate i_t and output gate o_t to implement this update. The following equations describe the update rules for the gates, where σ is the sigmoid activation function:

$$f_t = \sigma(W_f[e_t, h_{t-1}] + b_f)$$
$$i_t = \sigma(W_i[e_t, h_{t-1}] + b_i)$$
$$o_t = \sigma(W_o[e_t, h_{t-1}] + b_o)$$

From these gates, the internal states and output can be obtained as the following equation:

$$C_t = f_t \circ C_{t-1} + i_t \circ tanh(W_c[e_t, h_{t-1}] + b_c)$$
$$h_t = o_t \circ tanh(C_t)$$

where $\circ$ represents the element wise vector multiplication. Outputs from LSTM layers are transformed through another linear transformation layer to obtain two dimensional outputs for each chunk. The loss L of the network is then calculated as the cross entropy loss between ground truth y_t and the soft-max activation of the output layer out_t, summed over the whole sequence:

$$L = \sum_t [y_t \log out_t + (1 - y_t) \log(1 - out_t)]$$

B. Training

SwallowNet is trained on piezoelectric signals, LFA, and acceleration energy. At each iteration, 32 data sequences (a sequence is one minute of data) and their corresponding labels are fed into the optimization. To increase the training set, data augmentation is used by scaling the sequences by a random number between 0.8 and 1.2. This range is selected empirically to introduce realistic noise into the data, while not drastically distorting signal shape.

The dimension of the embedding layer e_t is selected to be 32 to compress the original data (20 samples * 4 channels = 80). The dimension of LSTM layer is set at 32 (SwallowNet32) and 64 (SwallowNet64). The network is trained using the Adam optimization algorithm [21] with a learning rate of 1e-3. The number of training iterations is fixed throughout the whole experiment. The backpropagation through time algorithm updates both feature representation and LSTM weights at each optimization iteration, instead of training each layer separately.

V. Data collection

Data is collected from 10 subjects (7/3 male/female, mean age=26, mean BMI=22). Each subject wears the necklace, along with a chest-mounted GoPro camera (to closely record the neck region) and in-ear microphone (to record eating sounds) to aid reviewers in labeling swallows. The video from the GoPro camera and the in-ear audio are merged later and annotations of swallows are provided by independent reviewers. The combination of audio and video further enhance the quality of the ground truth data generated by the reviewers.

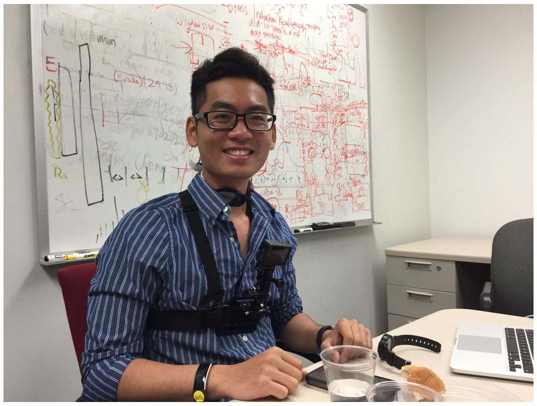

Fig. 6: Experiment setup

The experiment setup allows subjects to eat as freely as possible, without disrupting the eating process by having to report swallows. The subjects perform each of the following activities for two minutes, followed by a 30 second rest period: eat soup, eat bread, drink water from cup, eat salad, drink water from straw, eat chicken, make a phone call, eat chips, brush hair. There are two non-eating actions in the experiment: make a phone call represents talking, and brushing hair represents head movement.

VI. Results

This section reports the optimal features selected for prediction using a Random Forest. Results using Leave One Subject Out Cross Validation (LOSOCV) are provided, along with the root mean square error (RMSE) of predicted swallows for each participant.

A. Statistical Machine Learning

Applying the correlation of derivative feature to generate candidate swallows achieves a 99.4% recall rate. Compared to a sliding window segmentation of 1 second, correlation of derivatives segmentation reduce the number of false positives by 39.5%.

Table II shows the resulting predictive features identified by CFS-Subset feature selection, using different thresholds to generate the ranking. The most predictive features ranked first include all of the top, bottom, LFA and acceleration energy signals.

B. SwallowNet

Detection of swallows is evaluated using the event-based evaluation method [8] using a threshold of 1.5s. Concretely, a prediction within 1.5s of the true swallow is considered a true positive.

Table III shows both SwallowNet32 and SwallowNet64 outperform Random Forest in F-score. F-scores averaged over male and female subjects are similar, although slightly better for females. SwallowNet32 and SwallowNet64 also have similar F-scores, even though they are both trained from scratch using random weights.

SwallowNet32 and SwallowNet64 also outperforms Random Forest in predicting the number of swallows for each participant using LOSOCV (see Table IV). To count swallows,

Rank	Signal	Feature	Window	
1	Bottom	max	-2s	2s
	Top	quart1	0s	3s
	Top	3rd coef	0s	3s
	Lean Forward	first location of maximum	0s	3s
	Top	skew	-4s	1s
	Top	mean	-1s	4s
	Accel energy	3rd coef	-1s	4s
	Bottom	mean	-6s	0s
2	Top	mean	-2s	2s
	Bottom	kurtosis	-2s	2s
	Top	mean	0s	3s
3	Top	kurtosis	-3s	0s
	Top	RMS	0s	3s
	Bottom	max	-4s	1s
	Bottom	skew	-4s	1s
	Top	number cwt peaks	-1s	4s
	Bottom	number cwt peaks	-1s	4s
	Bottom	median	-6s	0s
	Accel energy	IRQ	-6s	0s

TABLE II: Top 19 features

Subject	Random Forest	SwallowNet64	SwallowNet32
1	70.73	83.86	83.69
2	74.78	72.84	78.05
3	39.54	54.42	52.51
4	60.24	47.17	56.45
5	77.76	74.06	69.95
6	87.20	89.31	92.24
7	71.45	77.45	77.03
8	75.90	74.75	80.53
9	57.39	73.46	79.17
10	50.77	91.63	91.08
Average Males	65.86	73.32	74.85
Average Females	68.24	75.22	78.91
Average	66.58	74.16	76.07

TABLE III: F-score of swallows using event based evaluation SML use the number of merged true positive candidates, while SwallowNet use the number of positive prediction interval.

Subject	Number of true swallows	Random Forest	SwallowNet64	SwallowNet32
1	87	63	80	78
2	63	73	74	60
3	65	40	60	57
4	57	57	37	46
5	98	105	76	78
6	87	85	91	94
7	90	78	86	84
8	101	102	97	94
9	119	80	90	106
10	80	31	60	68
RMSE		5.88	4.87	3.34

TABLE IV: Counting swallows

VII. Limitations and Conclusions

In summary, this paper presents an integrated wearable necklace that combines two piezoelectric sensors with an inertial motion unit. The placement of piezoelectric sensors at the larynx and trachea is critical. Placing them elsewhere, such as above and below the larynx, generates correlated signals similar to each other, and thus limited information is gained using multiple sensors.

The paper also shows that given enough labeled data, a deep neural network model outperforms statistical machine learning model in detecting swallows, resulting in a 76.07% F-score compared to 66.6% F-score using LOSOCV, and a RMSE of 3.34 in swallow count. SwallowNet does not require segmentation, and generalizes well to variable length test sequences. Thus SwallowNet proves to be a suitable model for detecting and characterizing eating. Future work will combine audio with piezoelectric signals to improve detection of swallows.

References

[1] Shibo Zhang, Rawan Alharbi, William Stogin, Kevin Moran, Angela F. Pfammatter, Bonnie Spring, and Nabil Alshurafa. Machine learning algorithms applied to detect feeding gestures. *The Obesity Society*, 2016.

[2] Oliver Amft and Gerhard Troster. Methods for detection and classification of normal swallowing from muscle activation and sound. In *2006 Pervasive Health Conference and Workshops*, pages 1–10. IEEE, 2006.

[3] Tauhidur Rahman, Alexander Travis Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury. Bodybeat: a mobile system for sensing non-speech body sounds. In *MobiSys*, volume 14, pages 2–13, 2014.

[4] Jun Nishimura and Tadahiro Kuroda. Eating habits monitoring using wireless wearable in-ear microphone. In *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*, pages 130–132. IEEE, 2008.

[5] Sebastian Päßler, Matthias Wolff, and Wolf-Joachim Fischer. Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food. *Physiological measurement*, 33(6):1073, 2012.

[6] Haik Kalantarian, Nabil Alshurafa, and Majid Sarrafzadeh. A wearable nutrition monitoring system. In *2014 11th International Conference on Wearable and Implantable Body Sensor Networks*, pages 75–80. IEEE, 2014.

[7] Haik Kalantarian, Nabil Alshurafa, Tuan Le, and Majid Sarrafzadeh. Monitoring eating habits using a piezoelectric sensor-based necklace. *Computers in biology and medicine*, 58:46–55, 2015.

[8] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):33, 2014.

[9] Nabil Alshurafa, Haik Kalantarian, Mohammad Pourhomayoun, Shruti Sarin, Jason J Liu, and Majid Sarrafzadeh. Non-invasive monitoring of eating behavior using spectrogram analysis in a wearable necklace. In *Healthcare Innovation Conference (HIC), 2014 IEEE*, pages 71–74. IEEE, 2014.

[10] Akira Toyosato, Shuichi Nomura, Atsuko Igarashi, Naoko Ii, and Akiko Nomura. A relation between the piezoelectric pulse transducer waveforms and food bolus passage during pharyngeal phase of swallow. *Prosthodontic research & practice*, 6(4):272–275, 2007.

[11] Qiang Li, Kazuhiro Hori, Yoshitomo Minagi, Takahiro Ono, Yong-jin Chen, Jyugo Kondo, Shigehiro Fujiwara, Kenichi Tamine, Hirokazu Hayashi, Makoto Inoue, et al. Development of a system to monitor laryngeal movement during swallowing using a bend sensor. *PloS one*, 8(8):e70850, 2013.

[12] Sebastian OH Madgwick. Automated calibration of an accelerometers, magnetometers and gyroscopes-a feasibility study. Technical report, Technical report, 2010.

[13] Nabil Alshurafa, Haik Kalantarian, Mohammad Pourhomayoun, Shruti Sarin, Jason J Liu, and Majid Sarrafzadeh. Non-invasive monitoring of eating behavior using spectrogram analysis in a wearable necklace. In *Healthcare Innovation Conference (HIC), 2014 IEEE*, pages 71–74. IEEE, 2014.

[14] Keinosuke Fukunaga and Larry Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1):32–40, 1975.

[15] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.

[16] Graves Alex. *Supervised Sequence Labelling with Recurrent Neural Networks*. PhD thesis, Technische Universität München, 2008.

[17] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, and Jürgen Schmidhuber. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.

[18] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE, 2013.

[19] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2625–2634, 2015.

[20] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[21] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.