Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior

Edward Sazonov[1]**, Stephanie Schuckers**[1]**, Paulo Lopez-Meyer**[1]**, Oleksandr Makeyev**[1]**, Nadezhda Sazonova**[2]**, Edward L Melanson**[3] **and Michael Neuman**[4]

[1] Department of Electrical and Computer Engineering, Clarkson University, 8 Clarkson Ave, Potsdam, NY 13699, USA
[2] Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26505, USA
[3] Univesity of Colorado, Denver, School of Medicine, Division of Endocrinology, Metabolism and Diabetes, and the Center for Human Nutrition, Denver, CO 80262, USA
[4] Department of Biomedical Engineering, Michigan Technological University, Houghton, MI 49931, USA

E-mail: esazonov@cias.clarkson.edu

Abstract

A methodology of studying of ingestive behavior by non-invasive monitoring of swallowing (deglutition) and chewing (mastication) has been developed. The target application for the developed methodology is to study the behavioral patterns of food consumption and producing volumetric and weight estimates of energy intake. Monitoring is non-invasive based on detecting swallowing by a sound sensor located over laryngopharynx or by a bone-conduction microphone and detecting chewing through a below-the-ear strain sensor. Proposed sensors may be implemented in a wearable monitoring device, thus enabling monitoring of ingestive behavior in free-living individuals. In this paper, the goals in the development of this methodology are two-fold. First, a system comprising sensors, related hardware and software for multi-modal data capture is designed for data collection in a controlled environment. Second, a protocol is developed for manual scoring of chewing and swallowing for use as a gold standard. The multi-modal data capture was tested by measuring chewing and swallowing in 21 volunteers during periods of food intake and quiet sitting (no food intake). Video footage and sensor signals were manually scored by trained raters. Inter-rater reliability study for three raters conducted on the sample set of five subjects resulted in high average intra-class correlation coefficients of 0.996 for bites, 0.988 for chews and 0.98 for swallows. The collected sensor signals and the resulting manual scores will be used in future research as a gold standard for further assessment of sensor design, development of

automatic pattern recognition routines and study of the relationship between swallowing/chewing and ingestive behavior.

Keywords: swallowing, chewing, mastication, deglutition, energy balance, wearable sensors, energy intake, ingestive behavior, obesity, dysphagia

(Some figures in this article are in colour only in the electronic version)

1. Introduction

A key factor in the maintaining a healthy lifestyle is the balance between energy intake and energy expenditure. Abnormalities in this balance can lead to problems such as obesity, anorexia, bulimia and other eating disorders. Many tools are available which measure energy expenditure, but more research is needed to develop accurate, easy methods to measure energy intake. The most precise method of measuring energy intake is to measure it indirectly through the use of doubly-labeled water (Schoeller 1988). This method provides a measurement of energy intake over a period of several days (typically 7–14), and if weight is stable over the measurement period, it is assumed that daily energy intake equals the average daily energy expenditure. However, the high cost (~$500 per subject) of both the isotope and analytical methods makes this approach impractical as a therapeutic tool.

Dietary self-reporting has been used intensively for the measurement of food intake, but there are numerous shortcomings, particularly in regard to long-term use. Both food frequency questionnaires (Prentice *et al* 1989, Weber *et al* 2001, Champagne *et al* 2002) and self-reported diet diaries (Day *et al* 2001, De Castro 1994) tend to be inaccurate; people tend to miscalculate and underreport their daily intake (Livingstone *et al* 1990, Livingstone and Black 2003, Mertz *et al* 1991).

In this paper, we present the methodology of non-invasive monitoring of chewing (mastication) and swallowing (deglutition) as the first step in developing a wearable non-invasive device for objective quantification of ingestive behavior. Our ultimate goal is to develop a device that will detect each instance of food consumption, differentiate between liquid and solid foods and provide volumetric and weight estimates of energy intake. The device will be able to capture fine nuances of ingestive behavior such as contribution of snacking and night eating to daily energy intake, and impact of rate of ingestion on energy intake, etc. To have the greatest clinical utility in measuring 'free-living humans', the device should be unobtrusive and indistinguishable.

The specific goals for this paper are to develop the sensors, hardware, and software needed to collect reliable data from chewing and swallowing sensors in controlled conditions and to develop a method for manually scoring bites, chews and swallows for the collected data. To evaluate the reliability of manual scoring we use intra-class correlation coefficients. The experimental results suggest high efficiency and reliability of the proposed approach.

In future work, the manual scores of chews and swallows will be used as a gold standard for training of automatic pattern recognition software which will only utilize the information from the wearable sensors (Makeyev *et al* 2007a, 2007b). The manual scores will also be used to generate the statistical model in order to establish a relationship between chewing/swallowing and ingestive behavior.

This paper is organized as follows: in section 2 we present the background on the use of swallowing and chewing for monitoring of ingestive behavior. In section 3, we present

the detailed description of the sensor hardware and software components along with the data collection protocol. In section 4, we present the inter-rater reliability evaluation techniques that were used to evaluate the proposed methodology. Results are presented in section 5. Discussion and conclusions are presented in sections 6 and 7, respectively.

2. Background

Chewing and swallowing have been subjects of several studies addressing various issues of the masticatory sequence (from food ingestion to swallowing) (Spiegel *et al* 1989). Though most of these studies did not look at the issues of energy intake, they provide significant clues that objective measurement of ingestive behavior through monitoring of chewing and swallowing is feasible and has potential.

Lear *et al* (1965) performed a study on the frequency of swallowing. In a study of 20 subjects, they reported a significant difference in the number of swallows per hour recorded during sleep (mean 5.3), inactivity such as reading (mean 23.5) and during food intake (mean 180), which means that the frequency of swallowing is a good predictor for the periods of food intake. This study also reported a significant variance of the swallowing rates among subjects in the range of 80–510 during a meal of a fixed size. This finding illustrates the individuality of the masticatory cycle and emphasizes the need for individual approach to the study of food intake.

One of the strongest cues in favor of studying individual patterns of swallowing and chewing as an indicator of food consumption and a potential indicator of caloric energy intake is reported by Stellar and Shrager (1985). The authors indicate that in a study of ten subjects for a period of one year, the number of self-reported swallows during a day correlated more highly ($r = 0.4$–0.65) with weight gain on the following day than did estimates of caloric intake ($r = 0.1$–0.5).

Other studies of swallowing and chewing concentrated mostly on dysphagia or other issues (Hind *et al* 2001, Agrawal *et al* 1998). However, these studies are of interest because of the various sensors used to detect instances of swallowing and periods of chewing. Presently, videofluoroscopy (Firmin *et al* 1997, Palmer *et al* 1992) and electromyography (EMG) (Cooper and Perlman 1996) are considered the gold standard in studies of swallowing and chewing. The drawbacks of these methods are dependency on bulky and potentially unsafe equipment (videofluoroscopy) and invasiveness (subcutaneous EMG). All of the reported EMG studies have used subcutaneous placement of electrodes in the masseter, suprahyoid and infrahyoid muscles (Cooper and Perlman 1996, Ertekin *et al* 1998, 2002, Palmer *et al* 1992) to avoid interference from the muscles of the neck. Subcutaneous placement of electrodes makes this approach prohibitively invasive.

Other reported sensors include a variety of strain devices. Stellar and Shrager (1985) used an oral sensor shaped like an arch fastened to the back molars. The sensor provided very reliable data on chewing and swallowing, however such setup is not convenient for prolonged wear, requires participation of a dentist at each installation, fitting and removal, and requires a wired interface coming out of the mouth. Pehlivan *et al* (1996) reported an electronic device for measuring the frequency of spontaneous swallowing. The piezoelectric strain sensor was placed on the coniotomy region between the cricoid and thyroid cartilage and was held in place by a band of elastic material. The sensor detected upward and downward motion of the larynx produced by swallowing. Data reported by Pehlivan and in later studies by Ertekin *et al* (1998, 2002) indicate that a laryngeal strain sensor is not appropriate for obese subjects, since under chin fat pad inhibits reliable detection of swallows. McKee and McBride (1998)

studied effects of sex and age on swallowing. The study utilized a four-sensor manometry pressure probe.

Bellisle *et al* (2000) used video recordings and 'edogram' sensors to study chewing and swallowing as indices of the stimulation to eat. Swallowing was detected by an ingenious sensor consisting of a small water-filled balloon maintained on the subject's throat by an adjustable plastic collar. An instance of swallowing induces a change in pressure in the balloon which was detected by a liquid pressure transducer. The same paper described a strain sensor for the detection of chewing during meals. The sensor consisted of a headset terminating on one side by a strain gauge which rested on the cheek just in front of the ear and responded to the movements of the jaw.

The earliest mention of acoustical methods for the detection of swallowing goes back to the original work of Lear *et al* (1965). The authors reported failure of a pneumatic method, which in its essence is very close to the methods for detecting swallowing described by Bellisle *et al* (2000) and Pehlivan *et al* (1996) to detect swallowing in subjects where 'a mass of soft tissue overlay the laryngeal prominence and masked the surface disturbance caused by its movements', which is another consistent indication that sensors detecting larynx and laryngeal prominence movements are not reliable in obese people. They reported very successful experimentation with an acoustical method that detected 'a short sharp noise . . . on the skin lateral to the laryngeal prominence'.

'When detected by instruments of suitable sensitivity, the swallowing sound, regardless of its intensity, can be readily distinguished from other noises heard in the area, such as intrinsic sounds of speech, belching, coughing and snoring, and the extrinsic sounds generated by movements of clothes, sheets, etc against the recording device'. It is important to note that swallowing always occurs when the upper and lower teeth are close together, but not necessarily in contact (Palmer *et al* 1992) and that respiration stops in the pharyngeal phase of swallowing (Lourinia 1996). In other words, intrinsic speech cannot mask the sounds of swallowing.

Another important study on the sonic methods of swallow detection has been conducted by Firmin *et al* (1997) based on the work of Hamlet *et al* (1994) and Vice *et al* (1990, 1995). The procedures generally follow the methodology of cervical auscultation of swallowing commonly used in the dysphagia practice and utilized an ear probe microphone, a neck microphone, a modified electrolaryngograph and a neck-mounted accelerometer. The output of these sensors has been manually scored to detect four different stages of swallowing. The laryngograph has produced the most accurate results while neck and ear microphones have been less accurate. Nevertheless, out of the four listed methods an ear microphone and a neck microphone are the least invasive devices that do not require special fitting of the sensors and can be incorporated into a wearable device.

3. Methodology

Based on the background on studies of swallowing and chewing, we propose the following solution for monitoring of chewing and swallowing. Chewing will be detected by a small piezoelectric strain gauge located below the outer ear and swallowing will be detected by a miniature microphone located over laryngopharynx.

A strain sensor will detect specific motion of the lower jaw by capturing strains created by motion of the posterior border of the mandible's ramus relative to the temporal bone (figure 1). Each bite (the period starting from opening of the mouth for food intake and ending at closing of the mouth) or a period of chewing will be reflected in the strain sensor signal.
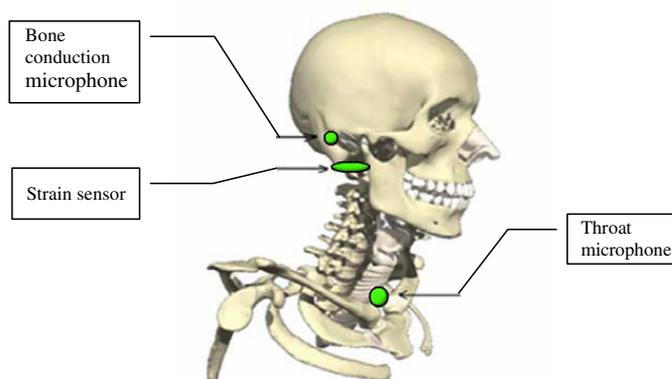
Figure 1. Suggested sensor locations.

A microphone located over the laryngopharynx (figure 1) will detect characteristic sounds of swallowing. Other potential locations may include a bone-conduction microphone placed on the mastoid bone and an in-ear microphone. Laryngopharynx sensor location offers an advantage of a stronger signal, since the microphone is placed closer to the origin of the swallowing sound.

In the following subsections we describe sensors and related hardware, software and protocol used for the collection of multi-modal data from a group of subjects. The multi-modal data are then used in manual scoring of chews and swallows. The obtained manual scores will be used in the future as a gold standard for training of automatic pattern classifier utilizing only sound and strain data for chewing and swallowing detection.

3.1. Sensors

Several variations of microphones and strain sensors were tested in different combinations.

We tested four models of commercially available miniature microphones as the sensing devices. The first model was a piezoelectric bone-conduction microphone EM-L (Temco Inc.). This microphone can be modified to be placed on the mastoid bone behind the ear or used as an ear probe. The second model was a piezoelectric noise-canceling microphone model N4530 (Challenge Electronics). The third model was a modified throat microphone XTM70V (iXradio) usually used for hands-free radio communications. Throat microphones convert vibration signals from the surface of the skin rather than pick up waves of sound pressure, thus reducing the ambient noise. The fourth model was a miniature IASUS NT (IASUS Concepts Ltd) throat microphone. This microphone provides a dynamic range of 46 ± 3 dB with a frequency range of 20–8000 Hz. Youmans (2003) reported the peak frequency of swallowing to be in the range of 1083.02–3286.73 Hz, therefore this microphone is capable of acquiring swallowing sounds.

The microphone tests consisted of recording of several consecutive swallows with subsequent evaluation of sound quality. Sound quality was evaluated both subjectively (i.e. by listening to the recording) and objectively (by visualizing in a sound editor and computing the signal-to-noise ratio). While all four microphones were able to detect swallowing sounds, the throat microphones showed a lower degree of sensitivity to ambient noise. Based on the results of testing the IASUS microphone was selected for the data collection because of its higher sensitivity to swallowing sounds and low noise (figure 2).
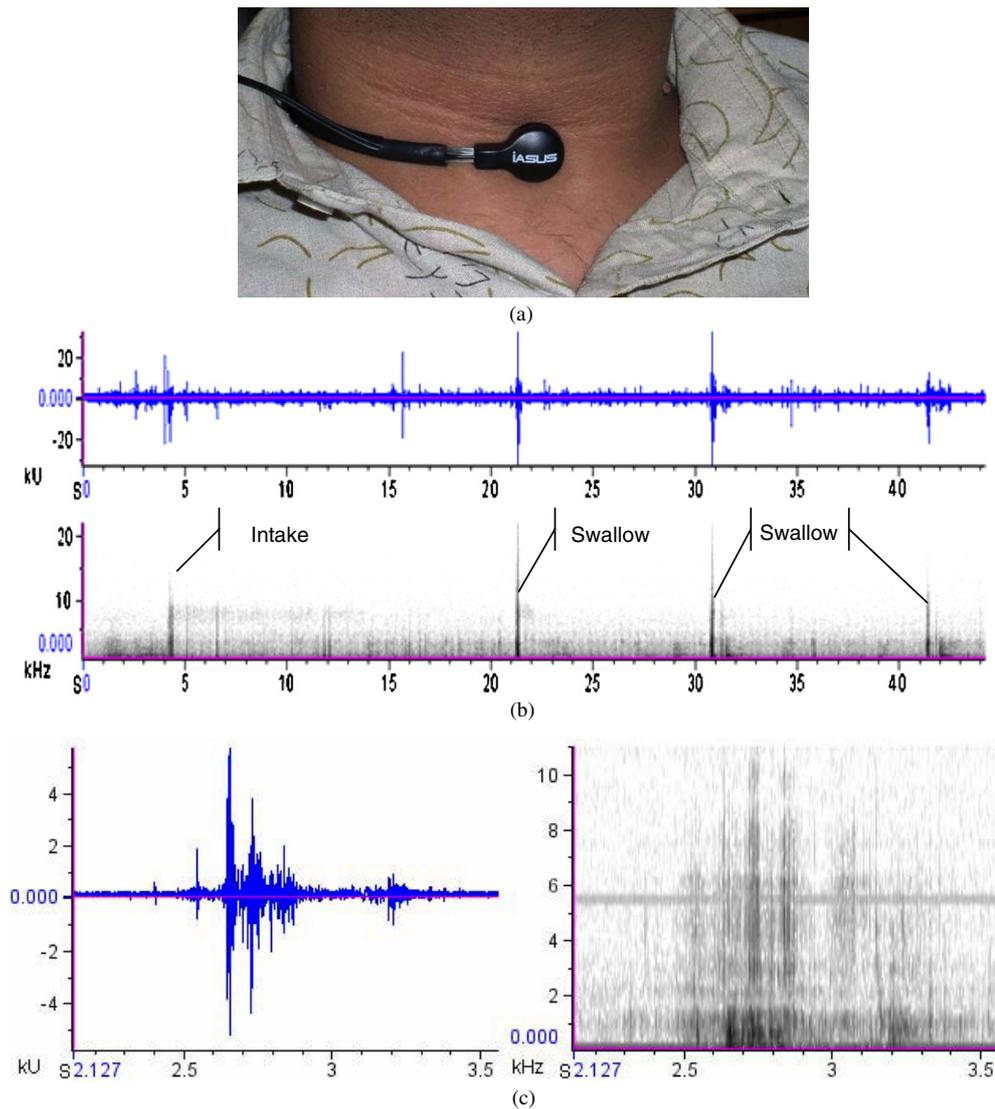
Figure 2. (a) IASUS throat microphone, (b) time series and spectrogram of consuming five peanuts with three swallows and (c) time series and spectrogram of a gulp of water of arbitrary size.

Several configurations of below-the-ear strain sensor for chewing detection were developed and tested. Evaluated types of sensors included foil strain gauges (figure 3) and a piezoelectric sensor (figure 4).

Testing the strain sensors consisted of three distinct activities: orally counting to ten, drinking 50 ml of water and eating a cracker. Different sensors and configurations were evaluated with regard to sensitivity to the characteristic chewing motion and on ability to reject anterior–posterior and medial–lateral head tilts. While all sensor types are able to detect characteristic jaw motion due to chewing, the sensor configurations using foil strain gauges showed a higher degree of sensitivity to the subject's head motion which was especially evident
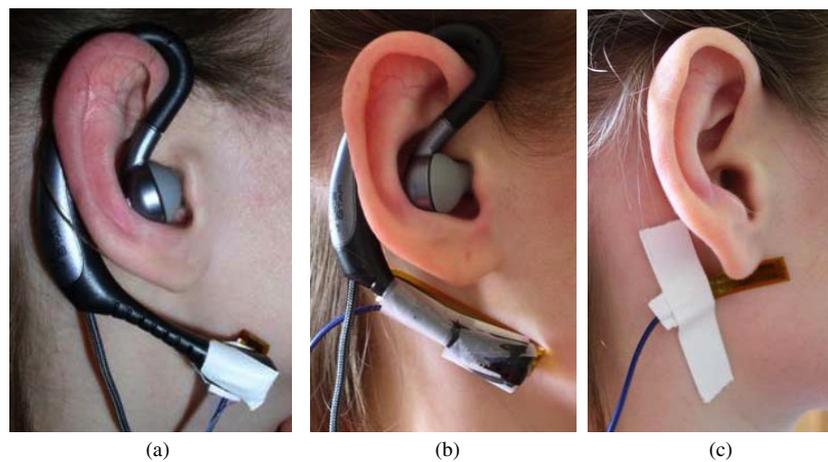
Figure 3. Three configurations of the foil strain gauges: (a) u-shaped strain sensor at the end of the boom, (b) ring strain sensor and (c) skin-attached strain sensor.
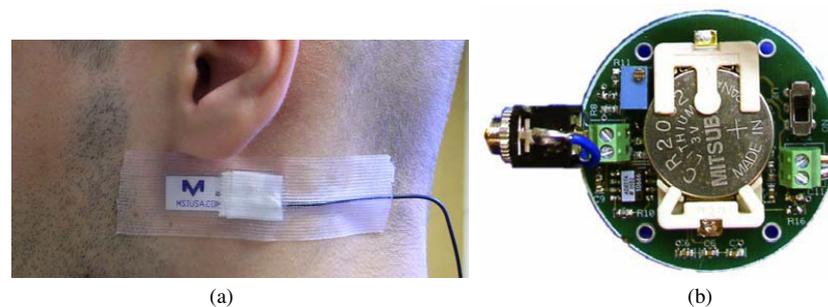

Figure 4. (a) Piezoelectric strain sensor attached by medical tape to the area immediately below the outer ear and (b) strain sensor amplifier.

for head tilting during drinking. Based on the results of the testing, a piezoelectric film sensor (MSI Inc.) was selected to be used for the data collection. Attached by medical tape to the area immediately below the outer ear (figure 4(a)), this sensor is able to detect changes in the skin curvature created by the characteristic motion of the mandible relative to the temporal bone during chewing and bites.

The final set of sensors consisted of (1) an IASUS throat microphone located over laryngopharynx to detect swallowing, (2) a microphone directed outward to detect ambient sounds, (3) a throat microphone located on the mastoid bone to detect swallowing, (4) a piezoelectric strain sensor attached by medical tape immediately below the outer ear to detect chewing and (5) an in-ear microphone XEM98D (iXradio) to detect swallowing.

The block diagram of the system for multi-modal data collection is presented in figure 5.

Microphone signals were amplified by a custom-built pre-amplifier with a variable gain in the range 20–60 dB. The gain of the amplifier was set experimentally for each sound channel to reliably capture the subtle sounds of swallowing without saturating the amplification circuits during normal speech and fixed for the whole data collection process. Amplified signals were recorded through a line-in input of a standard sound card at a sampling rate of 44100 Hz.
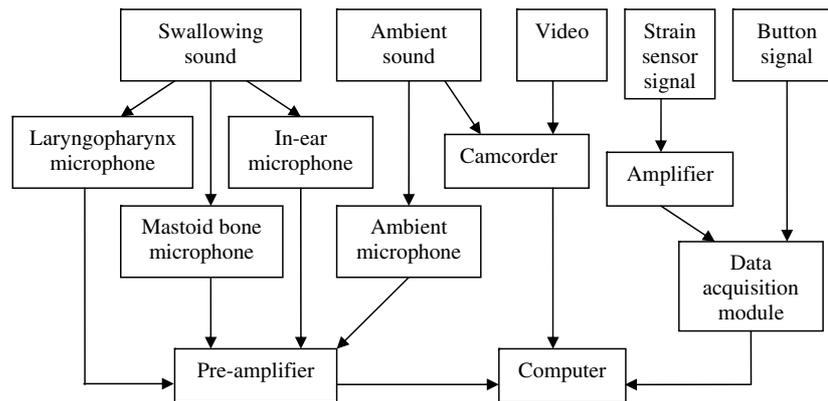
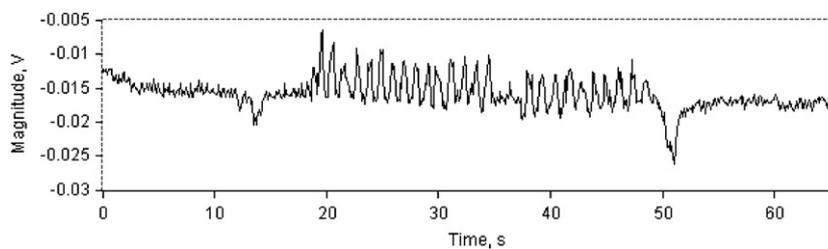**Figure 5.** Block diagram of the multi-modal data collection system.

Figure 6. An example of signal from the piezoelectric strain gauge. A peanut is ingested (∼14 s) and consumed (19–48 s).

The signal from the piezoelectric strain gauge was buffered by a custom-designed amplifier with input impedance of approximately 100 MΩ (figure 4(b)). This buffered signal was acquired by a 16 bit data acquisition module USB-1608FS (www.measurementcomputing.com) at a sampling rate of 100 Hz. An example of the acquired signal is shown in figure 6.

A handheld push-button switch was connected to another input channel of USB-1608FS. Subjects were asked to push the button to indicate swallowing instances which were recorded as a pulse of 5 V.

During each session of the data collection, subjects were videotaped in profile by a camcorder to capture subject activity and ambient sound independent of data acquisition by a computer. Camcorder video was captured at 30 frames per second in an interlaced format. To simplify the scoring process, the video was deinterlaced into a progressive 60 frames per second stream and the sound track was separated from the video.

3.2. Software

Data acquisition software was developed in LabVIEW (National Instruments). The software allows simultaneous capture of four-channel sound (from two sound cards) and up to eight channels of sensor data (such as strain sensor signal and the square wave from the button). All captured data are synchronized in time. Information about the data files and synchronization values was stored in a project file.
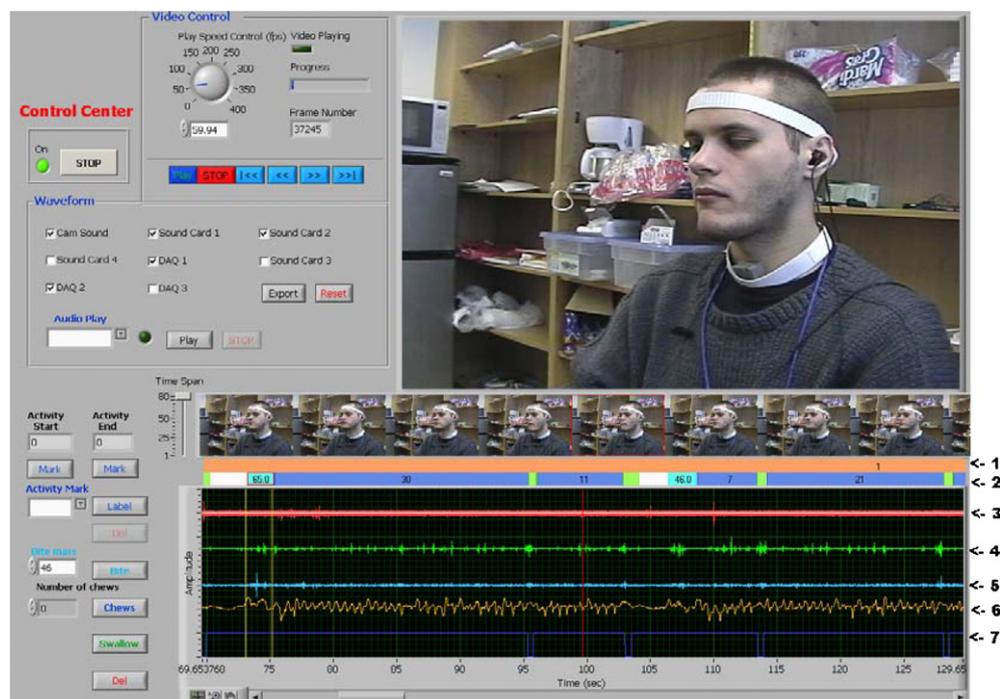
Figure 7. Scoring software graphical user interface: (1) activity mark, (2) bites/chews/swallows track, (3) ambient sound signal, (4) throat microphone signal, (5) bone-conduction microphone signal, (6) strain sensor signal and (7) user button signal.

The scoring software (figure 7), also developed in LabVIEW, allows manual review and playback of the acquired data by a human rater and assignment of event marks to each instance of swallowing, each period of chewing with associated number of chews and bites with associated mass of the bite.

The scoring software also allows the user to zoom in and out in the data window, show all or selected data channels, and browse video frame-by-frame or any specified interval. The same software also allows assignment of labels for long-term activities performed by a subject. For example, periods of consumption of a specific type of food or a specific activity such as silent inactivity or talking can be indicated on the timeline. Manual scoring of the data utilizes all the data channels shown in figure 7, including the video and signals from the sound and strain sensors. A scorer following a predefined protocol identifies target segments of the time series, plays back the sensor data and narrows the boundaries of bites, chews and swallows.

3.3. Data collection protocol

Data collection was performed on a group of 21 generally healthy human subjects, 12 males and 9 females. In addition, since chewing and swallowing detection may be more difficult in obese individuals, 38% of subjects had body mass index (BMI) greater than 30. The mean BMI of the subjects was of 28.98 with a standard deviation of 6.42, subject's minimal BMI of 20.9 and maximum BMI of 42.1. Institutional Review Board approval was obtained for the study. Subjects read and signed the informed consent form. Data collection for each subject

was performed during four visits. Subject's weight, waist and hip circumference (to identify android or gynoid type of obesity if present) were measured at each session. Subject's height was measured once during the first session. Subjects were encouraged to abstain from talking during the study unless they were asked to talk. All subject had no dental problems that would interfere with normal food intake. As much as possible, we attempted to recruit a diverse population in terms of gender, ethnicity, age and body size. However, due to the small sample size, at present, our sample may not be representative of the population. We plan to expand this research to include more subjects.

Each session consisted of three parts: (1) a 20 min inactivity period (10 min of silent inactivity and 10 min of talking where the subject was asked to read aloud), (2) the meal period, consisting of unlimited time to eat the meal of a fixed size plus extra food items at the end of the this period, if desired and (3) a second 20 min inactivity period (10 min of silent inactivity and 10 min of talking). A variety of magazines were provided to entertain the subject during the inactivity periods. Subjects were encouraged to read with a straight neck, holding the magazine in front of the face to avoid obscuring the camcorder's perspective of the subject's neck.

Two fixed sizes of the meal (standard and large) were used with the large size being 50% bigger than the standard. The following food items were included in the meal: a slice of cheese pizza, a can of 1% fat yogurt, an apple and a peanut butter sandwich. The foods were selected to represent different physical properties of the food such as crispiness, softness/hardness and tackiness. The variability in physical properties of food ensured that the proposed methodology was tested on a sample that is representative of the variability in the properties of everyday food. More analysis is needed to determine if sensors are capable of distinguishing between food properties. The provided drink was clear water. All food items were to be consumed unmixed and completely. The weight of the food item was measured after each bite on an electronic scale and recorded by the observer. Water was consumed separately from food.

During the first session, a standard size meal was served and no background noise was allowed during the meal period. During the second session, a standard size meal was served and background noise and talking to the subject were used during the meal period. Noise was introduced to experiment to simulate realistic environments where people may be eating and that can potentially impact results in future sound recognition experiments. To create background noise we used a combined recording of city noise, restaurant noise and segments of music recordings at a fixed volume level. To involve subject in conversation, the operator asked the subject questions not relevant to the purpose of the research. During the third session, a large size meal was served and no background noise was allowed during the meal period. During the fourth session, a large size meal was served and background noise and talking to the subject were used during the meal period.

4. Inter-rater reliability

In this study, the multi-modal data collected on 21 subjects were scored by an experienced rater using the scoring software described in section 3.2. To maintain homogeneity of the scoring process, a specific scoring protocol was developed to specify characteristics of bites, chews and swallows, and provide guidelines on their identification. The 'experienced rater' was trained on several training data collection sessions that were not included in the reported results. After several consecutive scoring sessions and score reviews which produced accurate scores, the rater was considered qualified to score the full study dataset.

The manual score will be used in the future for training of the automatic pattern recognition software and for creating statistical models of volume and mass energy intake, therefore it is

Table 1. Characteristics of human subjects involved in a multi-rater reliability study.

Subject number	Height (cm)	Weight (kg)	Gender	Body mass index	Ethnicity	Age	Waist (cm)	Hip (cm)
1	168	105	Female	37.2	Caucasian	48	119.4	129.5
2	181	83.4	Male	25.5	Caucasian	23	88.9	104.1
3	172	115	Male	38.9	Caucasian	27	119.4	121.9
4	160	55.9	Female	21.8	Asian	26	69.9	91.4
5	155	52	Female	21.6	Caucasian	27	83.8	94
Minimum	155	52		21.6		23	69.9	91.4
Maximum	181	115		38.9		48	119.4	129.5
Mean	167.2	82.3		29		30.2	96.3	108.2
Standard deviation	10.2	28.2		8.4		10.1	22.2	16.9

important to achieve a very high accuracy of scoring. To evaluate the reliability of the scoring process we conducted a multi-rater reliability study in which two new raters were trained using the same scoring protocol. A set of five subjects, two males and three females, was selected randomly from the whole population of subjects. Forty percent of the subjects in the test set have BMI greater than 30, comparable to 38% in the whole population. The mean BMI for the subjects in the reliability study is 29, comparable to 28.98 for the whole population. Detailed characteristics of human subjects involved in a multi-rater reliability study are presented in table 1.

For each of these five subjects only the third and the fourth sessions were used (large size meal with and without noise and talking), resulting in ten sessions in total. The scores obtained on the test set from the two raters were compared to those of the experienced rater. To evaluate the inter-rater reliability of the scores, the scores were split into non-overlapping epochs of fixed duration, and then the number of events (bites, chews, swallows) was calculated for each epoch. In cases when an event occurs on the border between two epochs it was considered belonging to the epoch that contains the midpoint of the event.

4.1. Selection of method for assessment of inter-rater reliability

Two of the popular inter-rater reliability measures were considered on the test set data: multiple raters Kappa (or Fleiss Kappa) (Fleiss 1981) and intra-class correlation coefficient (ICC) (Fleiss and Shrout 1979).

Kappa statistics is a popular measure used in numerous studies (Corwin *et al* 1998, Crowell *et al* 1997, Rybicki *et al* 1998). However, Kappa proved to be a poor choice for the problem at hand. Specifically, Kappa may produce negative values for ordered categorical data such as, in our case, chewing events. Negative Kappa indicates that the observed agreement occurs less often than the chance agreement or that the data are not suitable for the analysis (Juurlink and Detsky 2005). In the case of chews, there is a greater discrepancy between rater's recordings due to the nature of the experiment: chews have high frequency which makes them harder to count. For example, in some particular epoch one rater may detect seven chews and another eight chews, while being close, these values will still be considered as separate categories. From the point of view of Kappa calculation, this case is no different than when one rater detects seven chews and another zero, which indicates much greater disagreement.

Table 2. One-way analysis of variance.

Source of variation	Sum of squares	Degrees of freedom	Mean square
Epochs	$\mathrm{SS_{epochs}} = \frac{1}{k}\sum\limits_{i=1}^{n} x_{i\bullet}^2 - x_{\bullet\bullet}^2/(n \cdot k)$	$n-1$	$\mathrm{BMS} = \frac{\mathrm{SS_{epochs}}}{n-1}$
Error	$\mathrm{SS_{error}} = \mathrm{SS}_T - \mathrm{SS_{epochs}}$	$n \cdot (k-1)$	$\mathrm{EMS} = \frac{\mathrm{SS_{error}}}{(n-1)\cdot(k-1)}$
Total	$\mathrm{SS}_T = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{k} x_{ij}^2 - x_{\bullet\bullet}^2/(n \cdot k)$	$(n \cdot k) - 1$	

Thus, Kappa may not fully capture the agreement for ordered categorical data, where the value of each category matters in the way that some categories are closer to each other.

Due to many problems arising when using Kappa statistics when assessing inter-rater agreement, the use of inter-class correlation (ICC) was suggested instead. ICC is suitable for continuous as well as for ordered categorical data. ICC evaluates rating reliability by finding the portion of total variation across all ratings and all subjects explained by the variation of different ratings of the same subject. Fleiss and Cohen (1973) showed that weighted Kappa statistics and the ICC are essentially equivalent for the ordered categorical data. Thus, the ICC was selected as the most appropriate for the analysis of the data for this particular study.

4.2. Intra-class correlation coefficient

The inter-class correlation was computed for the case where each epoch was rated by each of the same set of raters, who are the only raters of interest (Fleiss and Shrout 1979). In this particular study, all the compared scores are from the same three raters. A one-way analysis of variance designed to calculate the ICC is presented in table 2, where k is the number of raters, n is the number of epochs and x_{ij} represents the event count for epoch i and rater j.

The ICC coefficient was calculated in the following way (Fleiss and Shrout 1979):

$$\mathrm{ICC} = \frac{\mathrm{BMS} - \mathrm{EMS}}{\mathrm{BMS} + \mathrm{EMS}(k-1)}. \tag{1}$$

5. Results

ICC was calculated using equation (1) for the scores produced by the three raters and for the separate periods of each session (first inactivity period, meal, second inactivity period). For the inactivity periods, reliability was assessed only in terms of swallows because no bites and chews occur during the inactivity periods; for the meal periods, reliability was assessed for bites, chews and swallows. ICC coefficients obtained for different epoch durations and averaged for ten experiment sessions are presented in table 3. The highest average intra-class correlation coefficient of 0.996 for bites, 0.988 for chews and 0.98 for swallows (average for the first inactivity period, meal and second inactivity period) were obtained for epoch duration of 120 s (table 3). It should be noted that even epochs of 10 s resulted in high correlation coefficients (>0.92) indicating a high degree of agreement between raters.

6. Discussion

The long-term goal of the proposed methodology is monitoring of ingestive behavior in free-living individuals. We suggest using small, inexpensive and non-intrusive sensors such as

Table 3. Averaged intra-class correlation coefficients for different epoch durations.

Assessed session period and event	Epoch duration (s)					
	10	20	30	45	60	120
First inactivity period (swallows)	0.976	0.975	0.977	0.956	0.978	0.981
Meal (bites)	0.935	0.941	0.962	0.977	0.982	0.996
Meal (chews)	0.928	0.938	0.954	0.973	0.978	0.988
Meal (swallows)	0.943	0.942	0.95	0.967	0.983	0.976
Second inactivity period (swallows)	0.959	0.957	0.953	0.968	0.965	0.983

microphones and strain sensors for capturing chewing and swallowing events. A mastoid bone microphone and a strain sensor can be integrated into a single device worn behind the ear in a manner similar to a wireless phone headset (earpiece). A laryngopharynx microphone can be worn as a medallion on a neck band. Other major advantages of the proposed methodology are no need for special fittings and positioning of the sensors and social acceptance of the sensors since they can be disguised as a wireless headset for a cellular phone and/or a medallion.

Captured chewing and swallowing events can potentially be used to identify periods of food consumption. Such devices may be used in a variety of studies evaluating behavioral aspects of food consumption in free-living individuals. Incorporating the pattern recognition capabilities on the device might allow its use for behavioral modification programs aimed at reducing food intake, similar to the way a pedometer can be used for increasing energy expenditure. The sensor and pattern recognition technology can also be applicable in the study of other diseases, for example anorexia and dysphagia.

The conducted study involving 21 subjects tested the capabilities of proposed sensors. The subject population involved both lean and obese individuals, therefore testing sensor applicability to a wide range of conditions. The collected sensor data were scored by a human rater. The produced scores will later be used to train the automatic sound recognition software that we hope will replace the human rater. The preliminary data that we obtained so far (Makeyev *et al* 2007a, 2007b) allow us to be optimistic about this approach. The manual scoring process relied on additional information (video footage) that will be not available to the automatic scoring software. While this may impact the accuracy of scoring, the highest accuracy of the training dataset is desired for the development of pattern recognition methods, therefore the use of video is justified. Our future goal is to show that a similar reliability of scoring can be achieved by an automatic method operating only on recorded sounds and strain signals.

Reliability of the manual scores was verified by a study conducted on a test set of five subjects involving two additional raters and ICC statistics. The range of the ICC is between 0 and 1 and high value of ICC means that there is little variation between the scores given by different raters. Experimental results obtained in our study (table 3) show that there is a high inter-rater reliability for all the epoch durations and all types of events. Table 3 also demonstrates an increase in ICC value associated with an increase in the epoch duration. There are two potential reasons for this behavior. First, as the epoch duration increases, the number of epochs for a recording decreases and the number of counted events for each epoch increases. This may result in a higher agreement since differences in counts for shorter epochs may compensate each other for longer epochs. Second, an important effect of splitting the scores into epochs is that the same event lying on the border between two epochs may be

marked by different raters in such a way that event midpoints belong to different epochs. In this case the ICC value would decrease, even if the event marks for all raters are similar. Therefore, the fact that the ICC is not equal to 1 does not necessarily mean that there is an actual disagreement between raters. The longer the epochs, the smaller is the possibility that event would lie on a border between two epochs.

Overall, the epoch durations of 30 s to 1 min are characteristic for the time scale of events happening during a meal. For example, it is typical to have 1–2 bites per minute. Therefore, we feel that the epoch sizes of 30 s to 1 min are the most indicative for our study.

The feedback from the raters indicates that some of the subjects had a tendency to move their head and torso a lot during the data collection making the scoring process more difficult. At the same time the raters did not report any situations when it was not possible to score an event unambiguously because of the movements of the subject.

Forty one out of total 252 recordings (16.27%) had partially incomplete data. All cases of incomplete data can be traced to the single reason of operator's error and further subdivided into three categories. These three subcategories are: failure to turn on one of the sensors or camcorder, running out of hard drive space or camcorder tape and failure to provide a synchronization signal. Most of these problems were minor and the experienced rater still was able to score 98.41% of all the sessions using the complete signal channels.

There were no reported problems with sensors or reliability of their positioning during the data collection process.

The biggest problem of most monitoring systems is that the detection of actual events is not always correct: there may be false detections when no event has occurred and there is no guarantee that true events are always detected. These problems are mostly due to the errors of raters and poor readability of the monitoring data. In our case, high inter-rater reliability and high readability of the data suggest the high quality of the rater scores. Moreover, 40% of the subjects in the test set have a BMI of at least 30 which suggests that our monitoring methodology is suitable for obese subjects.

Table 3 shows that consistently high agreement is observed for detecting bites, chews and swallows during meals and during inactivity periods. These results indicate that manually scored recordings of food consumption and inactivity can be used for training of automatic pattern recognition algorithms. Such software will replace human raters for free-living studies of ingestive behavior. The obtained scores of chews and swallows can be used to identify periods of food intake, to differentiate between liquid and solid foods and to produce volume and mass estimates of each meal or snack.

Overall, the proposed methodology presents a first, crucial step into designing a novel approach and novel methodologies for objectively monitoring ingestive behavior of subjects in a free-living environment. The proposed wearable devices may assist in discovering behavioral and seasonal patterns in food intake. Additional analysis of the time patterns of chews and swallows may carry clues to the type of food being consumed. This is a subject of future research. However, even if the future research determines that the food type and energy density of the food cannot be identified from the chewing/swallowing patterns, the device still will be valuable in behavioral monitoring and potentially might be used for behavioral modification programs targeting food intake similarly to the way a pedometer is used in behavioral modification targeting energy expenditure. In addition, the device could boost the accuracy of self-reporting food diaries by asking the subject to document each detected instance of food intake. The same methodology could also be used in the evaluation of dysphagic patients as well as people with other eating disorders.

Technically, implementation of a miniature and concealable wearable device is possible using present-day technology. A microcontroller-based device can capture the sensor data and

store it in a portable flash card that can be read and analyzed on a personal computer. Assuming data rates of 44 kB s^{-1} for the sound signal and 0.1 kB s^{-1} for the strain sensor the storage requirement for 24 h of uncompressed sensor data is less than 4 GB. This kind of capacity can be easily provided by inexpensive SD cards. At the present time the experiments on the automatic recognition of sensor data are performed on a personal computer, but we are confident that the pattern recognition algorithms can eventually be scaled down and implemented using FPGA/ASIC technology. Such miniaturization will allow real-time monitoring and real-time feedback desirable for behavioral modification programs.

7. Conclusion

The study of ingestive behaviors needs a simple methodology for monitoring of the food intake under free-living conditions. The most commonly used method of self-reporting diaries may be insufficiently accurate for this purpose. We propose to utilize counts of bites, chews and swallows as objective indicators of food intake on simple non-invasive sensors which can be implemented as a wearable device. In this method, chewing and bites are detected by a strain sensor positioned below the outer ear while the swallowing is detected by a microphone positioned over laryngopharynx. The hardware/software system described in this paper captures multi-modal sensor data which can be used for manual scoring of chewing and swallowing. The manual scores will be used as a gold standard dataset for further development of the approach. A study of 21 subjects was conducted with the goal of preparing a training dataset for pattern recognition and understanding relation of chews and swallows to food intake. As the first step, acquired data were manually scored by an experienced rater. The reliability of manual scores was assessed by inter-class correlation metrics in a multi-rater study conducted on the sample set of five subjects. The results show very high agreement between the raters which indicates high reliability of the scores. The results of this work will be utilized for training of automatic classifiers for pattern recognition of chews and swallows, and for studying the relationship between chews and swallows and ingestive behavior.

Acknowledgment

References

Agrawal K R, Lucas P W, Bruce I C and Prinz J F 1998 Food properties that influence neuromuscular activity during human mastication *J. Dent. Res.* **77** 1931–8
Bellisle F 2003 Why should we study human food intake behaviour? *Nutr. Metab. Cardiovasc. Dis.* **13** 189–93
Bellisle F, Guy-Grand B and Le Magnen J 2000 Chewing and swallowing as indices of the stimulation to eat during meals in humans: effects revealed by the edogram method and video recordings *Neurosci. Biobehav. Rev.* **24** 223–8
Champagne C M, Bray G A, Kurtz A A, Monteiro J B, Tucker E, Volaufova J and Delany J P 2002 Energy intake and energy expenditure: a controlled study comparing dietitians and nondietitians *J. Am. Diet. Assoc.* **102** 1428–32
Cooper D S and Perlman A L 1996 Electromyography in the functional and diagnostic testing of deglutition *Deglutition and its Disorders: Anatomy, Physiology, Clinical Diagnosis and Management* ed K Lourinia (London: Singular) pp 255–85
Cordain L, Eaton S B, Sebastian A, Mann N, Lindeberg S, Watkins B A, O'Keefe J H and Brand-Miller J 2005 Origins and evolution of the western diet: health implications for the 21st century *Am. J. Clin. Nutr.* **81** 341–54
Corwin M J *et al* 1998 Agreement among raters in assessment of physiologic waveforms recorded by a cardiorespiratory monitor for home use *Pediatr. Res.* **44** 682–90

Crowell D H *et al* 1997 Infant polysomnography: reliability *Sleep* **20** 553–60

Day N E, McKeown N, Wong M Y, Welch A and Bingham S 2001 Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium *Int. J. Epidemiol.* **30** 309–17

De Castro J M 1994 Methodology, correlational analysis, and interpretation of diet diary records of the food and fluid intakes of free-living humans *Appetite* **23** 179–92

Ertekin C, Aydogdu I, Secil Y, Kiylioglu N, Tarlaci S and Ozdemirkiran T 2002 Oropharyngeal swallowing in craniocervical dystonia *J. Neurol. Neurosurg. Psychiatry* **73** 406–11

Ertekin C, Yuceyar N and Aydogdu I 1998 Clinical and electrophysiological evaluation of dysphagia in myasthenia gravis *J. Neurol. Neurosurg. Psychiatry* **65** 848–56

Firmin H, Reilly S and Fourcin A 1997 Non-invasive monitoring of reflexive swallowing *Speech Hear. Lang.* **10** 171–84

Fleiss J L 1981 *The Measurement of Interrater Agreement, Statistical Methods for Rates and Proportions* (New York: Wiley) pp 212–304

Fleiss J L and Cohen J 1973 The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability *Educ. Psychol. Meas.* **33** 613–9

Fleiss J L and Shrout P E 1979 Intraclass correlations: uses in assessing rater reliability *Psychol. Bull.* **86** 420–28

Hamlet S, Penney D G and Formolo J 1994 Stethoscope acoustics and cervical auscultation of swallowing *Dysphagia* **9** 63–8

Hind J A, Nicosia M A, Roecker E B, Carnes M L and Robbins 2001 Links comparison of effortful and noneffortful swallows in healthy middle-aged and older adults *J. Arch. Phys. Med. Rehabil.* **82** 1661–5

Jebb S A and Prentice A M 1997 Assessment of human energy balance *J. Endocrinol.* **155** 183–5

Juurlink D N and Detsky A S 2005 Kappa statistic *CMAJ* **173** 15–6

Kral T V E, Meengs J S, Wall D E, Roe L S and Rolls B J 2003 Effect on food intake of increasing the portion size of all foods over two consecutive days *FASEB J.* **17** A809

Lear C S, Flanagan J B and Moorrees C F 1965 The frequency of deglutition in man *Arch. Oral. Biol.* **10** 83–100

Levine J A, Lanningham-Foster L M, McCrady S K, Krizan A C, Olson L R, Kane P H, Jensen M D and Clark M M 2005 Interindividual variation in posture allocation: possible role in human obesity *Science* **307** 584–6

Livingstone M B E and Black A E 2003 Markers of the validity of reported energy intake *J. Nutr.* **133** 895–920

Livingstone M B E, Prentice A M, Strain J J, Coward W A, Black A E, Barker M E, McKenna P G and Whitehead R G 1990 Accuracy of weighed dietary records in studies of diet and health *Br. Med. J.* **300** 708–12

Lourinia K 1996 *Deglutition and its Disorders: Anatomy, Physiology, Clinical Diagnosis and Management* ed K Lourinia (London: Singular)

Makeyev O, Sazonov E, Schuckers S, Lopez-Meyer P, Melanson E and Neuman M 2007b Limited receptive area neural classifier for recognition of swallowing sounds using continuous wavelet transform *Proc. 29th Annu. Int. Conf. of the IEEE Engineering in Medicine and Biology Society EMBC'2007 (Lyon, France, 23–26 August 2007)* pp 3128–31

Makeyev O, Sazonov E, Schuckers S, Melanson E and Neuman M 2007a Limited receptive area neural classifier for recognition of swallowing sounds using short-time Fourier transform *Proc. Int. Joint Conf. on Neural Networks IJCNN'2007 (Orlando, USA, 12–17 August 2007)* pp 1417.1–6

McKee J and McBride P 1998 Does age or sex affect pharyngeal swallowing? *Clin. Otolaryngol. Allied Sci.* **23** 100–6

Meiselman H L 1992 Methodology and theory in human eating research *Appetite* **19** 49–55

Mertz W, Tsui J C, Judd I T, Reiser S, Hallfrisch J, Morris E R, Steele P D and Lashley E 1991 What are people really eating? The relation between energy intake derived from estimated diet records and intake determined to maintain body weight *Am. J. Clin. Nutr.* **54** 291–95

Palmer J B, Rudin N J, Lara G and Crompton A W 1992 Coordination of mastication and swallowing *Dysphagia* **7** 187–200

Pehlivan M, Yuceyar N, Ertekin C, Celebi G, Ertas M, Kalayci T and Aydogdu I 1996 An electronic device measuring the frequency of spontaneous swallowing: digital phagometer *Dysphagia* **11** 259–64

Prentice A M, Black A E, Murgatroyd P R, Goldberg G R and Coward W A 1989 Metabolism or appetite: questions of energy balance with particular reference to obesity *J. Hum. Nutr. Diet.* **2** 95–104

Prentice A M and Jebb S A 2003 Fast foods, energy density and obesity: a possible mechanistic link *Obes. Rev.* **4** 187–94

Rolls B J, Morris E L and Roe L S 2002 Portion size of food affects energy intake in normal-weight and overweight men and women *Am. J. Clin. Nutr.* **76** 1207–13

Rolls B J, Roe L S, Kral T V E, Meengs J S and Wall D E 2004 Increasing the portion size of a packaged snack increases energy intake in men and women *Appetite* **42** 63–9

Rybicki B A, Peterson E L, Jhonson C C, Kortsha G X, Cleary W M and Gorrel J M 1998 Intra- and inter-rater agreement in the assessment of occupational exposure to metals *Int. J. Epidemiol.* **27** 269–73

Schoeller D A 1988 Measurement of energy expenditure in free-living humans by using doubly labeled water *J. Nutr.* **118** 1278–89

Spiegel T A, Shrager E E and Stellar E 1989 Responses of lean and obese subjects to preloads, deprivation, and palatability *Appetite* **13** 45–69

Stellar E and Shrager E E 1985 Chews and swallows and the microstructure of eating *Am. J. Clin. Nutr.* **42** 973–82

Vice F L, Bamford O, Heinz J M and Bosma J F 1995 Correlation of cervical auscultation with physiological recording during suckle-feeding in newborn infants *Dev. Med. Child Neurol.* **37** 167–79

Vice F L, Heinz J M, Giuriati G, Hood M and Bosma J F 1990 Cervical auscultation of suckle feeding in newborn infants *Dev. Med. Child Neurol.* **32** 760–8

Weber J L, Reid P M, Greaves K A, DeLany J P, Stanford V A, Going S B, Howell W H and Houtkooper L B 2001 Validity of self-reported energy intake in lean and obese young women, using two nutrient databases, compared with total energy expenditure assessed by doubly labeled water *Eur. J. Clin. Nutr.* **55** 940–50

Youmans S R 2003 Increasing the objectivity of the clinical dysphagia evaluation: cervical auscultation and tongue function during swallowing *PhD Dissertation* Florida State University, etd-09232003-010436