An Online Swallow Detection Algorithm Based on the Quadratic Variation of Dual-Axis Accelerometry

Sotirios Damouras, Ervin Sejdić, *Member, IEEE*, Catriona M. Steele, and Tom Chau, *Senior Member, IEEE*

***Abstract*—Swallow accelerometry is an emerging tool for noninvasive dysphagia screening. However, the automatic detection of a swallowing event is challenging due to contaminant vibrations arising from head motion, speech and coughing. In this paper, we consider the acceleration signal as a stochastic diffusion where movement is associated with drift and swallowing with volatility. Using this model, we develop a volatility-based swallow event detector that operates on the raw acceleration signal in an online fashion. With data from healthy participants and patients with dysphagia, the proposed detector achieves performance comparable to previously proposed swallow segmentation algorithms, with the added benefit of online detection and no signal pre-processing. The volatility-based detector may be useful for event identification in other biomechanical applications that rely on accelerometry signals.**

***Index Terms*—Event detection, quadratic variation, semi-martingales, swallow accelerometry.**

I. Introduction

THE biomechnical characterization of human activity has frequently been facilitated through multi axis accelerometry. Example applications include the detection of feeding activities [1] and driver drowsiness [2], the assessment of fall risk in the aged [3], the quantification of muscle contractions [4], the monitoring of cardio-respiratory activity [5], the study of head angular accelerations in impact injuries [6], the tracking of neck and thoracic vibrations for sleep apnea diagnosis [7], and the noninvasive characterization of swallowing activity [8], [9].

In the above applications, there is a common need to automatically extract one or more events of interest from the recorded acceleration signals. Particularly, in swallowing accelerometry, we are interested in automatically identifying signal segments associated with the passage of foodstuffs from the mouth, through the pharynx and into the esophagus [10]. These segments of signals may then be subjected to further analysis, such as feature extraction and classification, to discern between healthy and dysphagic swallowing [11], [12]. Indeed, swallowing accelerometry has been suggested as a potential adjunct to bedside screening for dysphagia [9], [11]–[14].

A number of accelerometric event detection methods have been developed in the literature. For example, Sejdić *et al.* [10] proposed a detection scheme based on the fuzzy clustering of the standard deviation of the acceleration signal into swallow and nonswallow classes. Detected events from two orthogonal axes of vibration were logically combined. This approach was tailored to long accelerometric recordings ($> 10^4$ samples) and yielded high detection accuracies for dry, wet and wet-chin tuck swallows when compared against manual segmentation by human experts. Lee *et al.* [9], used a pseudo-automatic algorithm, whereby a swallow event was detected when the standard deviation of the signal within a 500 ms sliding window exceeded an empirically derived threshold in one of two vibration axes. This method exhibited high detection sensitivity for short swallow sequences but in many cases the detected event times did not agree with results of manual segmentation. These automatic event detection methods have been developed and demonstrated as strictly offline algorithms. Other swallowing accelerometry studies have relied on manual segmentation of vibration signals either by referencing events detected from simultaneously acquired videofluroscopic image sequences [8], [11], or by human expert inspection of acceleration traces [15]. Automatic event detection is necessary for continuous monitoring applications, where data records are too extensive for manual analysis and changes in swallowing function need to be flagged online. Likewise, in screening applications, automatic event detection would reduce the number of manual steps in a swallowing test, thereby mitigating the risk of human error.

In this paper, we introduce a novel approach to the automatic detection of swallowing activity that is particularly conducive to online analyses of dual-axis, continuous, extended duration recordings. We first describe two swallowing accelerometry data sets (Section II) that are used in the detailed formulation of our event detector (Section III). The method is applied to dual axis swallowing accelerometry data and evaluated against respective gold standards (Section IV).

II. Data Sources

We considered two separate dual axis swallowing accelerometry data sets, one derived from healthy participants and the other from patients with dysphagia. In both cases, the data arose from a dual axis accelerometer (ADXL322, Analog Devices) placed midline, below the thyroid notch using double-sided

Manuscript received June 14, 2009; accepted December 23, 2009. Date of publication February 22, 2010; date of current version May 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yufei Huang. This work was supported by the Canada Research Chairs Program, Health Technology Exchange and Ontario Centres of Excellence.

S. Damouras is with the Bloorview Research Institute, Bloorview Kids Rehab, Toronto, ON M4G 1R8, Canada (e-mail: sdamoura@stat.cmu.edu).

E. Sejdić and T. Chau are with the Bloorview Research Institute, Bloorview Kids Rehab and the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON M4G 1R8, Canada (e-mail: esejdic@ieee.org; tom.chau@utoronto.ca).

C. M. Steele is with the Toronto Rehabilitation Institute and the Department of Speech-Language Pathology, University of Toronto, Toronto, ON M4G 1R8, Canada (e-mail: steele.catriona@torontorehab.on.ca).

Digital Object Identifier 10.1109/TSP.2010.2043972

tape. The axes of the accelerometer were aligned to the anterior-posterior (AP) and superior-inferior (SI) anatomical axes as in [9]. Dual axis signals were bandpass filtered in hardware with a pass band of 0.1–3000 Hz, sampled at 10 kHz and passed through an amplifier (P55, Grass Technologies).

The healthy participant data set consisted of dual axis swallowing accelerometry signals corresponding to 295 swallows from 20 individuals without swallowing difficulties. The set included dry saliva swallows, wet swallows (water by cup with the head in the neutral position) and wet chin-tuck swallows (water by cup with the head in the chin-tuck position). The gold standard event times were obtained through manual segmentation by two expert human raters. This healthy participant data set was originally considered in [10], and further details of the collection procedure are contained therein. The patient data set consisted of 266 swallows from 37 individuals with neurogenic dysphagia. Accelerometric data were acquired simultaneous to videofluoroscopy. Participants swallowed two or three 5 mL teaspoons of thin liquid barium (40% weight by volume suspension) with their head in a neutral position. Subsequent to each swallow, the patient was asked to say "ha-ha-ha-ha". At the attending clinician's discretion, the participant may have taken an additional sip by cup. In this case, the gold standard event times were obtained through retrospective review of radiological video by a trained speech language pathologist. This patient data set has not been previously considered.

III. Swallow Detection in Accelerometry Signals

Intuitively, a swallow manifests itself as an irregular vibration, increasing the signal's variance. Thus, our goal is to detect a swallow based on the variability of the accelerometry signal. However, the exact definition and measurement of this variability is challenging. For one, accelerometry signals are highly nonstationary [8], which poses significant challenges to their statistical analysis. Moreover, there are additional sources of variability which are confounded in these signals. The two most prominent ones are measurement error and variability from voluntary or involuntary head movement. Our proposed methodology addresses both types of unwanted variability. To illustrate the effects of head movement in particular, we present in Fig. 1 the accelerometry signal in the SI direction of a healthy subject, while performing five chin-tuck swallows. The obvious excursions in the mean of the signal before and after the swallow are due to head movement. The subject has to lower his chin before the swallow, and then raise it right afterwards to prepare for the next one. This and other types of head movement add to the variability of the signal and induce nonstationarity, both of which significantly complicate swallow detection.

A. Stochastic Modeling of Accelerometry Signal

We propose a model for the accelerometry signal that helps distinguish the variability attributed to swallowing from that due to head movement. The basic premise is to consider the actual acceleration, denoted by $\{X_t\}_{t\geq 0}$, as a continuous stochastic process. We assume X_t follows a stochastic differential equation

$$dX_t = \mu_t dt + \sigma_t dW_t \tag{1}$$

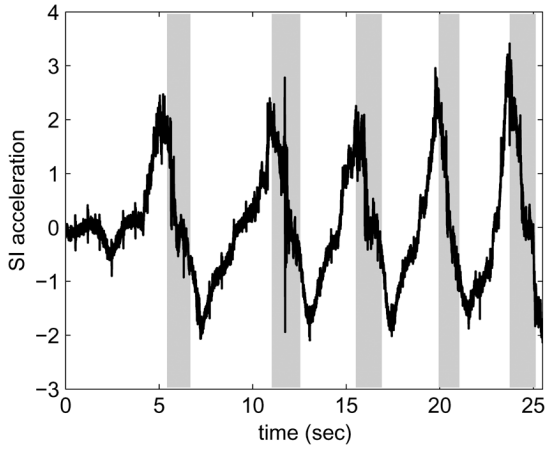

Fig. 1. Accelerometry signal in SI direction during five chin-tuck swallows. Shaded regions delimit swallows, as identified by a speech language pathologist.

where W_t is a Brownian motion process, and μ_t and σ_t are separate processes called *drift* and *volatility*, respectively. Processes that satisfy (1) are called stochastic diffusions, and are described more thoroughly in [16]. They constitute a broad and flexible family of continuous Markov processes, ranging from simple Brownian motion to more elaborate nonstationary, non-Gaussian processes. In simple terms, model (1) states that instantaneous changes in acceleration, dX_t, comprise two terms: a smooth change $\mu_t dt$ and a noisy change $\sigma_t dW_t$. The difference between the two terms is subtle, since they can both be stochastic when the drift is stochastic. The distinguishing characteristic is that $\mu_t dt$ is a gradual change, because μ_t multiplies a time increment dt, whereas $\sigma_t dW_t$ is an erratic change, because σ_t multiplies a Brownian motion increment, i.e., a random quantity. Intuitively, the drift term μ_t represents the gradual rate of change in the mean level of X_t, whereas σ_t is the scaling factor of the noisy fluctuations of X_t.

Based on the previous description of the model, it is natural to associate head movement with drift, since head movement is directional and has a gradual effect on acceleration. A swallow, on the other hand, consists of more abrupt and erratic movements, so we associate it with volatility. We claim that, using our proposed model (1), swallowing activity will be reflected in the accelerometry signal by an increase in its volatility. But volatility is not directly observable, so in practice we need to estimate it. In the next two paragraphs we review the relevant literature and present the approach we will adopt for estimating volatility.

B. Quadratic Variation

We want to estimate the unobserved volatility process σ_t from an equally spaced sample $(X_{t_0}, X_{t_1}, \ldots, X_{t_n})$, where $t_i = i\Delta t$. For now we assume the true acceleration X_t can be recorded without error; we return to the issue of measurement error later on. Even though volatility is a latent process, it can be estimated nonparametrically in a simple way, using the properties of stochastic diffusions. The basic quantity used in the estimation is called the quadratic variation (QV) process, and it acts as the link between volatility and the observed process.

The QV process $[X]_t$ of any stochastic process $\{X_t\}_{t\geq 0}$ is defined as

$$[X]_t = \lim_{\Delta t \to 0} \sum_{i=1}^{\lfloor t/\Delta t \rfloor} \left(X_{i\Delta t} - X_{(i-1)\Delta t}\right)^2. \tag{2}$$

In other words, $[X]_t$ is the limiting sum of the squared increments of the process X up to time t, as the time between increments goes to zero. As such, QV is essentially a measure of the cumulative variability of X. Under fairly general conditions [16], QV is equivalent to

$$[X]_t = \int_0^t \sigma_t^2 dt. \tag{3}$$

Therefore, QV is independent of the drift, and the reason is that the squared $\mu_t dt$ terms are smooth, so they disappear in the limit. In contrast, the squared $\sigma_t dW_t$ terms are sizeable even in the limit, and they converge to $\sigma_t^2 dt$. More importantly, we can associate the volatility in (3) to the observed process in (2). In particular, the QV process $[X]_t$ is strictly increasing, and its rate of increase at any time t is equal to the square of the volatility, σ_t^2. The natural estimator of $[X]_t$, based on a high-frequency, equally spaced sample is

$$\widehat{[X]}_t = \sum_{i=1}^{\lfloor t/\Delta t \rfloor} \left(X_{i\Delta t} - X_{(i-1)\Delta t}\right)^2. \tag{4}$$

The process $\widehat{[X]}_t$ is called *realized QV*, in order to distinguish it from the theoretical QV in (2). It has been used extensively in the statistical literature, especially in financial applications for measuring stock market variability [17]. The properties of $\widehat{[X]}_t$ are described in [16]; in particular it is a consistent estimator of $[X]_t$ as the sampling frequency $(1/\Delta t) \to \infty$.

C. Estimation Under Measurement Error

So far we have assumed the true acceleration X_t is observed without noise. In practice, however, this assumption is unrealistic. The observed signal Y_t is always contaminated with some error, as

$$Y_t = X_t + \epsilon_t \tag{5}$$

where $\{\epsilon_t\}_{t\geq 0}$ is white noise with variance σ_ϵ^2. The physical interpretation of ϵ_t is that of measurement error from the accelerometer, as well as noise from physiological sources such as muscle vibrations. For the rest of the paper, we refer to the noise from both sources as measurement error. This error severely degrades the accuracy of realized QV, as defined in (4). In fact, it can be shown [18] that $\widehat{[Y]}_t \approx \widehat{[X]}_t + 2\lfloor t/\Delta t \rfloor \sigma_\epsilon^2$, where $\lfloor t/\Delta t \rfloor$ is the number of squared differences used in calculating $\widehat{[Y]}_t$. The term $2\lfloor t/\Delta t \rfloor \sigma_\epsilon^2$ represents the estimation bias due to sampling error, and this bias increases with the sampling rate $1/\Delta t$. In particular, $\widehat{[Y]}_t$ converges to infinity instead of $[X]_t$ as $1/\Delta t$ increases, and for large n we can consistently estimate the noise variance as $\hat{\sigma}_\epsilon^2 = \widehat{[Y]}_{t_n}/2n$.

Fortunately, there are ways to overcome the effects of measurement error on realized QV. The simplest is to subsample the process at a lower frequency than the maximum observation frequency [18]. For example, we would calculate the realized QV in (4) using every k^{th} observation in the sample. The benefits of subsampling come from reducing the estimation bias, $2\lfloor t/\Delta t \rfloor \sigma_\epsilon^2$. This subsampling, however, is wasteful because we are using only a fraction $(100/k\%)$ of the available data. We therefore adopt a refinement of this method, due to Zhang *et al.* [19], that uses all of the data. The authors propose creating k subsampled realized QV processes from interleaved samples, and then averaging them to obtain the final QV estimate. These interleaved samples start from 0 to $k-1$, each one increasing by k observations. Formally, the k sparsely sampled realized QV processes are

$$\widehat{[Y]}_t^{(j,k)} = \sum_{i=1}^{\lfloor t/(k\Delta t) \rfloor} \left(Y_{j+ik\Delta t} - Y_{j+(i-1)k\Delta t}\right)^2, \tag{6}$$

for any $j = 0, \ldots, k-1$. To facilitate understanding, Fig. 2 presents a graphical illustration of the sampling scheme for the first two such processes. The proposed estimator of QV is obtained by averaging the k processes in (6) as

$$\widetilde{[Y]}_t = \frac{1}{k} \sum_{j=0}^{k-1} \widehat{[Y]}_t^{(j,k)}. \tag{7}$$

The main question now is what is the best choice of the subsampling parameter k? The parameter k essentially controls the bias-variance trade-off in estimating QV; if it is too small the measurement error will overwhelm the QV estimator, whereas if it is too big the signal's true variability will be lost. The authors in [19] provide an estimate for the optimal k that minimizes the asymptotic mean square error of (7). Specifically, they propose using

$$k^\star \approx \left[\frac{9k_0 \left(\widehat{[Y]}_{n\Delta t}/n\right)^2}{\sum_{i=1}^{\lfloor n/k_0 \rfloor} \left(Y_{ik_0\Delta t} - Y_{(i-1)k_0\Delta t}\right)^4} \right]^{1/3} \tag{8}$$

where k_0 is an initial guess for $k^\star$. A strategy for choosing a good value of k_0 is given in the discussion of the implementation of the algorithm (Section IV-B). For our application to swallowing accelerometry signals we use the realized QV in (7), with

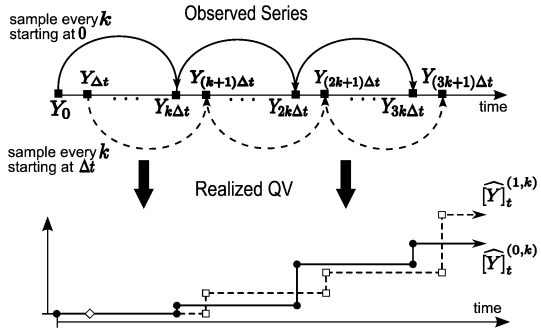

Fig. 2. Illustration of subsampling scheme in (6). The first realized QV process $\widehat{[Y]}_t^{(0,k)}$ uses observations starting from 0 and increasing by $k\Delta t$, whereas the second realized QV process $\widehat{[Y]}_t^{(1,k)}$ uses observations starting from Δt and increasing by $k\Delta t$.

the optimal frequency estimated by (8). More detailed properties and the performance of this volatility estimator on simulated data are presented in [19].

D. Volatility Estimation

Upon obtaining the realized QV, we want to estimate the volatility process σ_t. In other words, we want to measure the first derivative of the realized QV process $[\widetilde{Y}]_t$. For this purpose, we propose a nonparametric kernel smoother. We estimated the volatility process at t by

$$\hat{\sigma_t} = \sqrt{\frac{1}{mh}\sum_{i=0}^{n}[\widetilde{Y}]_{i\Delta t}K\left(\frac{i\Delta t - t}{h}\right)} \tag{9}$$

where $K(u) = (15/4)(u - u^3)I(|u| \leq 1)$ is an optimal kernel for estimating first derivatives [20], $I(\cdot)$ is the indicator function, h is the smoothing parameter, or bandwidth, and $m = \lfloor 2h/dt \rfloor$ is the number of points that fall within the kernel's support. Since the kernel has local support in $[t-h, t+h]$ and observations are equally spaced, the smoothing operation (9) is equivalent to an FIR filter. The choice of the smoothing parameter h is guided by contextual knowledge and is described in the next section.

We give some perspective on our volatility estimate $\hat{\sigma_t}$ by comparing it to alternative approaches for measuring variability. The simplest and most intuitive one would be to use a moving window variance estimator. Moving variance, however, is really appropriate only for first-order stationary signals, because it will incorporate variability due to changes in the signal's mean level. A possible fix for nonstationarity would be to detrend the data within the window before taking the variance, but this is subject to different issues. First of all, knowledge of the form of the trend (e.g., linear or polynomial) is required. Moreover, if the same window is used for both mean and variance, this assumes that they change at the same rate, which is a strict assumption. If, on the other hand, different window sizes are used for the mean and variance, this will increase the number of tuning parameters. Another way of dealing with nonstationarity in the mean would be to take the rolling variance of the first differences of the signal. This approach is closest to ours since it assumes a Markovian structure. By taking first differences, we subtract from the current observation the most recent one, which in a Markov process is the best estimate of the current mean level. Nevertheless, moving variance methods, either with or without taking first differences, do not address measurement error. In comparison, the volatility estimate we propose can be viewed as the average of several moving variance estimators on the subsampled and differenced signal. The main advantage of this framework is the subsampling, which addresses measurement error in an integrated way. We also believe that for dealing with measurement error, subsampling is preferable to preprocessing (e.g., smoothing) because the latter could attenuate the signal's variability.

E. Event Detection Algorithm

We conclude this section by presenting an event detection algorithm based on our volatility estimate. We argue that swallowing activity will be reflected in the accelerometry signal as a surge in its volatility. To capture this effect, we look for peaks in the estimated volatility. We do this by extracting local maxima which are above some threshold H. This thresholding is used to discard local maxima from small statistical fluctuations in the signal. To increase the accuracy of the algorithm, we utilize further information about the swallowing process. We know that the duration of a swallow is fairly short, usually lasting no more than a couple of seconds. Therefore, the peaks in volatility that we want to detect must be localized and, consequently, quite sharp. To quantify the sharpness of the peaks we measure the curvature of the volatility signal. We do this by estimating the second derivative of $\hat{\sigma_t}$ using a kernel smoother, as

$$\hat{\sigma_t}'' = \frac{1}{mh^2}\sum_{i=0}^{n}\hat{\sigma}_{i\Delta t}K\left(\frac{i\Delta t - t}{h}\right) \tag{10}$$

where $K(u) = (105/16)(-5u^4 + 6u^2 - 1)I(|u| \leq 1)$ is optimal for estimating second derivatives [20], and h and m have the same values as those used for estimating $\hat{\sigma_t}$. We use a similar thresholding approach to remove peaks which are not sharp enough. Specifically, we require that the second derivative be negative and below a certain threshold value, denoted by H'', $H'' < 0$. To summarize, our algorithm detects an event at time τ if the following three conditions are satisfied:

1) the estimated volatility is at a local maximum: $(\hat{\sigma}_{\tau-1} \leq \hat{\sigma}_\tau)$ AND $(\hat{\sigma}_\tau > \hat{\sigma}_{\tau+1})$;
2) the estimated volatility is above a certain threshold: $\hat{\sigma}_\tau \geq H$, $H > 0$;
3) the second derivative of the volatility is below a certain threshold: $\hat{\sigma}''_\tau \leq H''$, $H'' < 0$.

The choice of thresholds H and H'' depends on the properties of the event we want to detect, and we discuss this in the next section. We do not use separate threshold values for the signals in each of the two directions, but instead we average the volatilities of the two signals and use thresholding on their average value and its second derivative. We follow this approach because the variability bursts from swallowing are present in both directions, and we can thus easily combine their information. The steps of the algorithm are presented graphically in Fig. 3. Note that the runtime of the algorithm is of order $\mathcal{O}(n)$, where n is the number of observations, and is implemented online. This is easy to verify, since all of the quantities involved in the algorithm (realized QV $[\widetilde{Y}]_t$, estimated volatility $\hat{\sigma_t}$, and its curvature $\hat{\sigma_t}''$) are computed in an online fashion, using one-step updates as new data arrive.

IV. Application to Dual Axis Swallowing Accelerometry Signals

We apply our algorithm to dual axis swallowing accelerometry signals from neurologically intact individuals without swallowing difficulties and patients with neurogenic dysphagia. We begin by presenting an illustrative example of our modeling approach, followed by a discussion of the implementation details.

A. Illustrative Example

In this section we illustrate the workings of our variability estimation based on the signal in Fig. 1. We use the signal in Fig. 1 because it comes from a neurologically intact individual

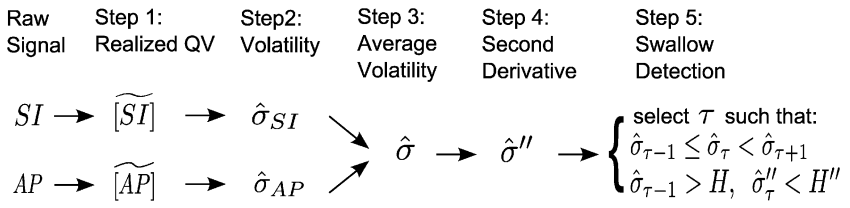

Fig. 3. Steps of swallow detection algorithm.

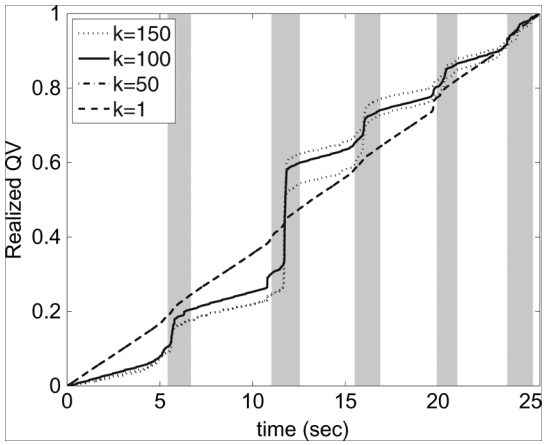

Fig. 4. Realized QV for the signal in Fig. 1 using subsampling with $k = 150$ (dotted line), $k = 100$ (solid line), and $k = 50$ (dash-dot line), and without subsampling ($k = 1$, dashed line). Shaded regions delimit swallows, as identified by a speech language pathologist.

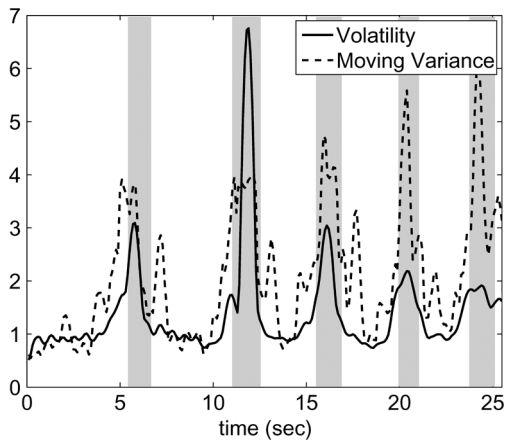

Fig. 5. Estimated volatility (solid line) and linearly detrended moving variance (dashed line) for the signal in Fig. 1. Shaded regions delimit swallows, as identified by a speech language pathologist.

performing chin-tuck swallows, which allows us to test the advantages of our proposed methodology in the presence of head movement. As shown in Fig. 3, the first step is to calculate the realized QV of the signal. Fig. 4 presents the realized QV process using the subsampling scheme in (7) for different values of the parameter k, as well as the realized QV without subsampling as in (4). Swallows are delimited by shaded areas. The plot shows how the use of subsampling can remove the effect of measurement error and reveal the underlying variability bursts from swallowing. Even though the value of the QV process changes with k, the locations of the variability bursts are consistent, which shows that the peak detection algorithm is robust to k.

The next step is to estimate volatility with the kernel smoother in (9). The resulting estimate using a bandwidth of $h = .5$ seconds is shown in Fig. 5 (solid line). For comparison, Fig. 5 also shows the linearly detrended moving variance estimator of the same bandwidth (dashed line). Both volatility and moving variance exhibit peaks during swallowing. However, while the peaks in volatility are well defined, moving variance exhibits additional peaks around the actual swallow as a result of head movement. This demonstrates the ability of our nonparametric modeling framework to deal with nonstationary signals.

B. Implementation

While we have outlined the main steps of our algorithm in Section III-E, we have not yet discussed the choice of the tuning parameters of the algorithm. This set of parameters consists of the subsampling frequency k, the bandwidth h in estimating volatility and its second derivative, and their threshold values H and H'', respectively, used for swallow detection. To maintain consistency and avoid overfitting, we use the same values for these parameters throughout our application, and for both neurologically healthy individuals and patients.

First, we select the subsampling parameter k used in realized QV. For each of our signals we get an estimate of the optimal value $k^\star$ from (7), using an initial guess k_0. For choosing k_0 we suggest plotting the realized QV for different values of k, as in Fig. 4, in order to identify a range of values that gives a clear definition of the variability bursts. We then suggest trying a few initial values in this range, and using the value of k_0 that is closest to its estimated optimum, $k^\star$. For our analysis we tried the initial values $k_0 = 50$, 100, and 150, which gave $k^\star = 111$, 115, and 123, respectively. Thus, the final subsampling parameter in our algorithm is $k^\star = 115$, corresponding to sampling every 11.5 ms. In general, the precise value of $k^\star$ is not critical for the algorithm, so it is not necessary to experiment extensively with k_0. Next, the choice of the bandwidth h of the kernel smoother is guided by contextual information about swallow durations. A typical healthy swallow can range from 1 second to more than 3 seconds [21]. We find that a value of $h = .5$ seconds, leading to a 1-second window, is small enough to detect short swallows and large enough to make estimation robust to statistical error. Finally, we select thresholds H and H'', used for swallow detection. For volatility we use a threshold of $H = 1.5$, which is approximately twice its level at rest. For the second derivative, we use a threshold of $H'' = -2$, which means that volatility at its peak must be at least .5 higher than its value at the edges of the 1-second window.

TABLE I
PERFORMANCE OF PROPOSED EVENT DETECTION ALGORITHM AND ALGORITHM IN [10] ON SWALLOWS FROM 20 HEALTHY INDIVIDUALS. N_S = NUMBER OF SWALLOWS, STANDARD ERRORS IN PARENTHESES

Swallow type	N_S	Proposed algorithm			Algorithm in [10]		
		Recall	Precision	F	Recall	Precision	F
Dry	98	0.8571 (0.0382)	0.9174 (0.0275)	0.8863	1.0000 (0.0105)	0.9159 (0.0268)	0.9561
Wet	102	0.9118 (0.0294)	1.0000 (0.0092)	0.9538	0.9608 (0.0192)	0.9800 (0.0140)	0.9703
Wet chin tuck	95	0.9368 (0.0258)	0.8652 (0.0275)	0.8996	0.8737 (0.0341)	0.9222 (0.0282)	0.8973
Total	295	0.9019 (0.0180)	0.9275 (0.0124)	0.9132	0.9448 (0.0123)	0.9394 (0.0133)	0.9412

C. Performance Measures

Before presenting the results of the event detection algorithm we discuss the performance measures we use. Binary classification measures are not applicable to event detection, because we cannot construct a meaningful contingency table. In particular, true negatives are ill-defined, because they constitute a continuum of times where neither a swallow occurs, nor the algorithm detects one. We can however define the following quantities:

- True Positives: TP $=$ # of correctly detected swallows;
- False Negatives: FN $=$ # of undetected swallows;
- False Positives: FP $=$ # of incorrectly detected swallows.

Based on these quantities we use the following standard measures from information retrieval [22], [23] to assess the performance of our algorithm.

- Recall: $R = TP/(TP + FN)$
- Precision: $P = TP/(TP + FP)$
- $F = 2RP/(R + P)$

Recall is a measure of completeness, and corresponds to sensitivity in classification. Precision is a measure of fidelity, and essentially equals one minus the false alarm rate. Recall and precision are inversely related, so we also use the harmonic average F of the two, to get a combined measure that balances them. For all three measures, values closer to one indicate better performance.

V. RESULTS

A. Healthy Subjects

Table I presents results from each type of swallow, as well as aggregate results. We find that the algorithm achieves on average more than 90% recall and precision, and an average F measure of 91%. In other words, the algorithm detects nearly all swallows while triggering only very few false alarms. We obtain the best results for wet swallows, with an F measure of 95%. Interestingly, we find that even for the wet chin tuck swallows, the most challenging case due to the added head movement, both recall and precision are high, with a F measure of nearly 90%.

It is interesting to compare our results with those of the algorithm proposed in [10]. Unlike our online algorithm, the algorithm in [10] is retrospective, and is based on a less flexible parametric model for the signal, combined with information on the maximum time between swallows. The right part of Table I reports the performance of this algorithm on the same data. The algorithm in [10] yields strong results for dry and wet swallows, with F measures of 95% and 97%, respectively. For wet chin tucks, the F measure drops slightly below 90%.

When directly comparing our algorithm with that proposed in [10], we find that the results are encouraging. Our algorithm achieves nearly equal precision on average, with only a small sacrifice in recall. The difference in the F measure is less than 3%. In particular, we observe that for the challenging case of wet chin tuck swallows, the performance of the two algorithms is identical in terms of their F values.

In summary, we find that our algorithm achieves comparable performance to the one proposed in [10], while having several practical advantages. First, our algorithm is lightweight and is implemented in an online fashion, making it attractive for use in real-time applications. Specifically, when comparing the runtime of the two algorithms for the same data, we find that our algorithm achieves a nearly tenfold improvement (81 s versus 807 s, for processing approximately 30 min of data). Second, it does not make any parametric assumptions in modeling the accelerometry signal. In particular, it does not rely on extra information about the time between swallows. Finally, it proves to be robust against head movement.

B. Dysphagia Patients

We also apply our algorithm to a second data set from subjects with swallowing disorders. The main difficulty in dealing with this data is that the conditions of the study were less controlled, leading to two sources of contamination in the measurements. First, the subjects interacted with the clinician administering the experiment. In particular, they were prompted to talk or cough after each swallow in order to check for aspiration, and they also talked during the experiment to communicate with the clinician. Second, there were swallows that were not identified by videofluoroscopy, because the camera would sometimes fail to capture the cervical region of interest. Fig. 6 shows a sample segment of the estimated volatility for a particular subject. As shown in the graph, most of the peaks in volatility were due to speech. Moreover, we suspect the peak in C to be a missed swallow due to camera misalignment, because it is immediately followed by speech.

In order to remove false positives due to speech, we added an additional step to our swallow detection procedure. We ran an online speech detection algorithm in parallel to our algorithm, in order to identify speech segments. The particular algorithm we used is based on thresholding of the log-energy of the FFT [24]. Whenever the swallow detection algorithm flagged a swallow

TABLE II
PERFORMANCE OF PROPOSED EVENT DETECTION ALGORITHM AND ALGORITHM IN [10] ON SWALLOWS FROM 37 DYSPHAGIC PATIENTS FROM TWO HOSPITALS. N_S = NUMBER OF SWALLOWS, STANDARD ERRORS IN PARENTHESES. VALUES FOR PRECISION AND F REPRESENT LOWER BOUNDS BECAUSE OF THE PRESENCE OF UNIDENTIFIED SWALLOWS

Hospital	N_S	Proposed algorithm			Algorithm in [10]		
		Recall	Precision	F	Recall	Precision	F
TEGH	74	0.8378 (0.0428)	>0.4233 (0.0574)	>0.5624	0.3947 (0.0567)	>0.4839 (0.0645)	>0.4348
TRI	192	0.8802 (0.0234)	>0.3953 (0.0353)	>0.5456	0.3763 (0.0349)	>0.5984 (0.0447)	>0.4620
Total	266	0.8684 (0.0207)	>0.4033 (0.0191)	>0.5508	0.3855 (0.0290)	>0.5412 (0.0243)	>0.4484

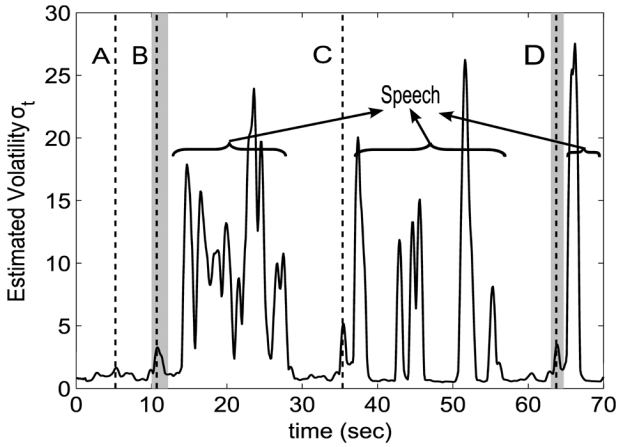

Fig. 6. Estimated volatility from accelerometry signal of patient with dysphagia. Shaded regions delimit swallows, as identified by videofluoroscopy. Dashed lines represent detected swallows from our algorithm: swallow A is false positive, swallows B and D are true positives, and swallow C is classified as false positive, although it probably is a true swallow which was missed by videofluoroscopy. Brackets denote areas of volatility peaks which where identified as speech by the speech detection algorithm and were eventually discarded.

that fell within a speech segment, we discarded it as a false positive. The peaks in the areas below the brackets in Fig. 6 fell within speech segments. Therefore, any swallows that our algorithm would find based on these peaks were discarded.

There were coughs and some true swallows that were not marked by the speech language pathologist during videofluoroscopic review. Unfortunately, our algorithm cannot distinguish between these unmarked events and the identified swallows. As a result, the false positive rate of our algorithm is artificially inflated. Hence, when reporting the results of our algorithm in Table II, the precision and F measure represent lower bounds of the actual values. Clearly from a practical standpoint, a diagnostic device with low precision would generate excessive false alarms. Nonetheless, the reader should be reminded that in this paper we have only discussed detection in isolation. Ultimately, a diagnostic medical device for dysphagia would necessarily include modules for removing nonswallowing events such as speech and coughs. In such case, we would expect the precision of the event detection algorithm to be greatly improved.

We observe that the recall of our algorithm is still very high, with an average value of 86%. This means that the large majority of swallows identified by videofluoroscopy were correctly detected. While we cannot give exact numbers for precision and F measure, for the reasons stated above, the estimated lower bounds are, respectively, 40% and 55%. This means that the false positive rate of our algorithm is less than 60%. Moreover, when comparing with the algorithm in [10], we find that our algorithm delivers drastic improvements in recall (more than a factor of 2 increase), and significant increase in the lower bounds on the F measure.

VI. CONCLUSION

We have modeled the swallow accelerometry signal as a flexible, nonparametric stochastic process. Using volatility estimates based on this model, we have developed an online event detector for isolating swallowing activity. Tests with healthy swallowing signals indicate that our detector yields high rates of recall and precision even in the presence of excessive head motion. Our detector also achieved high recall with patient swallowing data, but our analysis could only provide a lower bound for precision given the presence of swallows not labeled by the gold standard. The proposed event detection method could be applied to other biomechanical studies involving multi axis accelerometry.

REFERENCES

[1] S. Zhang, M. Ang, W. Xiao, and C. Tham, "Detection of activities by wireless sensors for daily life surveillance: Eating and drinking," *Sensors*, vol. 9, no. 3, pp. 1499–1517, 2009.
[2] E. Vurall, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *Proc. Human–Comput. Interact.*, 2007, pp. 6–18.
[3] A. Purwar, R. Myllyla, and W. Chung, "A wireless sensor network compatible triaxial accelerometer: Application for detection of falls in the elderly," *Sens. Lett.*, vol. 6, no. 2, pp. 319–325, 2008.
[4] N. Alves and T. Chau, "Vision-based segmentation of continuous mechanomyographic grasping sequences for training multifunction protheses," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 765–773, 2008.
[5] D. Phan, S. Bonnet, R. Guillemaud, E. Castelli, and N. Thi, "Estimation of respiratory waveform and heart rate using an accelerometer," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Vancouver, BC, Canada, Aug. 20–25, 2008, pp. 4916–4919.
[6] N. Yoganandan, F. Pintar, D. Maiman, M. Philippens, and J. Wismans, "Neck forces and moments and head accelerations in side impact," *Traffic Injury Prevent.*, vol. 10, no. 1, pp. 51–57, 2009.
[7] D. Rendon, J. Ojeda, L. Foix, D. Morillo, and M. Fernandez, "Mapping the human body for vibrations using an accelerometer," in *Proc. 29th Annu. Int. Conf. IEEE Eng. in Med. Biol. Soc.*, Lyon, France, Aug. 22–26, 2007, pp. 1671–1674.
[8] T. Chau, D. Chau, M. Casas, G. Berall, and D. Kenny, "Investigating the stationarity of paediatric aspiration signals," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 13, no. 1, pp. 99–105, 2005.
[9] J. Lee, C. Steele, and T. Chau, "Time and time-frequency characterization of dual axis swallowing accelerometry signals," *Physiol. Meas.*, vol. 29, no. 9, pp. 1105–1120, 2008.

[10] E. Sejdić, C. Steele, and T. Chau, "Segmentation of dual axis swallowing accelerometry signal in healthy subjects with analysis of anthropometric effects on duration of swallowing activities," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1090–1097, 2009.
[11] J. Lee, S. Blain, M. Casas, D. Kenny, G. Berall, and T. Chau, "A radial basis classifier for the automatic detection of aspiration in children with dysphagia," *J. Neuroeng. Rehab.*, vol. 3, no. 1, pp. 14:1–14:17, 2006.
[12] N. Reddy, R. Thomas, E. Canilang, and J. Casterline, "Toward classification of dysphagic patients using biomechanical measurements," *J. Rehab. Res. Develop.*, vol. 31, no. 4, pp. 335–344, 1994.
[13] N. Reddy, A. Katakam, V. Gupta, R. Unnikrishnan, J. Narayanan, and E. Canilang, "Measurements of acceleration during videofluorographic evaluation of dysphagic patients," *Med. Eng. Phys.*, vol. 22, no. 6, pp. 405–412, 2000.
[14] N. Reddy, E. Canilang, J. Casterline, M. Rane, A. Joshi, R. Thomas, and R. Candadai, "Noninvasive acceleration measurements to characterize the pharynegeal phase," *J. Biomed. Eng.*, vol. 13, no. 5, pp. 379–383, 1991.
[15] F. Hanna, S. Molfenter, R. Cliffe, T. Chau, and C. Steele, "Anthropometric and demographic correlates of dual axis swallowing accelerometry signal characteristics: A canonical correlation analysis," *Dysphagia*, in press.
[16] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*. New York: Springer, 1991.
[17] O. Barndorff-Neilsen and N. Shephard, "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *J. Roy. Stat. Soc.—Ser. B*, vol. 64, no. 2, pp. 253–280, 2002.
[18] J. Russell and F. Bandi, "Microstructure noise, realized volatility, and optimal sampling," in *Econometric Soc. 2004 Latin Amer. Meetings—No. 220*, 2004.
[19] L. Zhang, P. Mykland, and Y. Ait-Sahalia, "A tale of two time scales: Determining integrated volatility with noisy high-frequency data," *J. Amer. Stat. Assoc.*, vol. 100, no. 427, pp. 1394–1411, 2005.
[20] T. Gasser, H. Muller, and V. Mammitzsch, "Kernels for nonparametric curve estimation," *J. Roy. Stat. Soc.—Ser. B*, vol. 47, no. 2, pp. 238–252, 1985.
[21] B. Sonies, L. Parent, K. M. K, and B. Baum, "Durational aspects of the oral-pharyngeal phase of swallow in normal adults," *Dysphagia*, vol. 3, no. 1, pp. 1–10, 1988.
[22] C. van Rijsbergen, *Information Retrieval*. London, U.K.: Butterworths, 1979.
[23] Y. Yang, T. Pierce, and J. Carbonell, "A study of retrospective and online event detection," in *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Development in Inf. Retrieval*, Melbourne, Australia, Aug. 24–28, 1998, pp. 28–36.
[24] S. van Gerven and F. Xie, "A comparative study of speech detection methods," in *Proc. ISCA 5th Eur. Conf. Speech Communication and Technology (EUROSPEECH)*, Rhodes, Greece, Sep. 22–25, 1997, vol. III, pp. 1095–1098.

Sotirios Damouras received the B.Sc. degree in statistics from Athens University of Economics and Business, Athens, Greece, in 2002, the M.Sc. degree in financial mathematics from Warwick University, Coventry, U.K., in 2003, and the Ph.D. degree in statistics from Carnegie Mellon University, Pittsburgh, PA, in 2008.

He is currently with the Bloorview Research Institute, Bloorview Kids Rehab, Toronto, ON, Canada. His current research interests include biomedical applications of time-series and stochastic analysis.

Ervin Sejdić (S'00–M'08) received the B.E.Sc. and Ph.D. degrees in electrical engineering from the University of Western Ontario, London, ON, Canada, in 2002 and 2007, respectively.

He is currently with the Bloorview Research Institute and the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada. He was engaged in the field of wireless communications and also focused on signal processing. His current research interests include biomedical signal processing, time-frequency analysis, signal processing for wireless communications, and compressive sensing.

Dr. Sejdić won prestigious research scholarships from the Natural Sciences and Engineering Research Council of Canada in 2003 and 2005.

Catriona M. Steele received the B.A. (Hons.) degree in psychology and German, the M.H.Sc. degree in speech language pathology, and the Ph.D. degree in speech language pathology and neuroscience from the University of Toronto, Toronto, ON, Canada, in 1988, 1991, and 2003, respectively.

She was a practicing Speech Language Pathologist at Baycrest and St. Joseph's Health Centre, Toronto, ON, Canada. She is currently a Senior Scientist at the Toronto Rehabilitation Institute, University of Toronto, where she is also an Associate Professor in the Department of Speech-Language Pathology. Her current research interests include swallowing physiology and rehabilitation.

Dr. Steele received the Canadian Institutes of Health Research New Investigator Award in Aging and an Early Researcher Award from the Ontario Ministry of Research and Innovation.

Tom Chau (S'93–M'98–SM'03) received the B.A.Sc. degree in engineering science and the M.A.Sc. degree in electrical and computer engineering from the University of Toronto, Toronto, ON, Canada, in 1992 and 1994, respectively, and the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1997.

He was with IBM Canada. He is currently a Senior Scientist at the Bloorview Kids Rehab, Toronto, ON, Canada, where he is also an Associate Professor, Graduate Coordinator of the Clinical Engineering Program at the Institute of Biomaterials and Biomedical Engineering, and a Canada Research Chair in Pediatric Rehabilitation Engineering. His current research interests include the development of intelligent technologies and analytical techniques to decode functional intention of individuals with disabilities.