Stylohyoid and posterior digastric potential evaluation for a swallowing detection, with intramuscular EMG

Adrien Mialland, Ihab Atallah, Agnès Bonvilain

***Abstract*—Total laryngectomy consists of the removal of the larynx and requires a tracheostomy to allow breathing, where the trachea is sewn on the anterior throat. This permanently separates the airway from the esophagus and any attempt to set it back in place would require to emulate the protective mechanism of the larynx, to avoid any aspiration during swallowing. So, with the aim to allow for the feasibility of an *implantable active artificial larynx*, and based on our previous investigations [1]–[4], this paper focuses on the stylohyoid and the posterior digastric muscles to evaluate their potential for a real-time and implantable swallowing detection algorithm. Indeed, we have shown that they activate at the beginning of swallowing, predominantly for swallowing, and that they are easily accessible, with no further impairment required during any sensor implantation. So, we measured them with intramuscular electromyography (EMG), along with the surface EMG of the submental muscles, to provide a basis for comparison. The swallowing sound was also measured with an accelerometer, to measure the vibration it generates through the skin, and define a temporal limit for the detection. 4 swallowing tasks and 13 non-swallowing tasks were investigated, to assess the muscles in terms of detection performances and abilities to provide early detection. Sub-groups of participants with similar swallowing, based on the Euclidean distance, were also explored. Thus, we found that the stylohyoid and the posterior digastric outperform the submental muscles and previously reported results in the literature, with increased performances when combined. Besides, the tasks that are the most related to the oral preparatory stage of swallowing tend to provide similar activity than the swallowing tasks, and are responsible for the most drastic drop in performances. Finally, even though we found promising results in terms of earliness, a real-time swallowing detection would benefit both from a stratified approach and from the development of an algorithm that is inherently built around a temporal constraint.**

***Index Terms*—swallowing detection, stylohyoid, digastric, electromyography, total laryngectomy, deglutition.**

I. Introduction

When it comes to event detection, the past decades have provided a large corpus of algorithms able to automatically abstract complex representations of each type of events, and maximize their differences [5]. These machine-learning approaches have, therefore, become ubiquitous and it has been used in medical engineering to identify specific events in physiological signals. Specifically, the complexity of the swallowing process has given rise to multiple measurement methods to better access its subtleties [6], and the detection algorithms make it possible to consider advanced clinical methods to treat and follow patients, even in free-living conditions [7], [8]. However, these approaches focus on non-invasive measurements, to allow bedside assessments or daily monitoring, and have shown limited performances in complex every day environments [3]. These shortcomings emphasize the difficulty to provide these algorithms with high-quality data, that best represents the full range of possibilities. Consequently, the algorithm cannot abstract a suitable representation of each type of task to recognize. In other words, whatever the algorithm, its performances inherently depend on the information the signals at hand can provide, and non-invasive measurement methods are likely to provide limited insight into swallowing. Indeed, swallowing requires a complex sequencing of multiple anatomical structures in the neck [9], and invasive measurement may allow to access more suitable signals.

Yet, we aim at the feasibility of an *implantable active artificial larynx* as an airway rehabilitation method following total laryngectomy, which consists of the full resection of the larynx and therefore the loss of its functions. Especially, it is involved in the protection of the airway during swallowing, to avoid any aspiration. Therefore, the surgery permanently separates the air passage from the bolus passage with the creation of a tracheostomy, where the trachea is sewn on the anterior neck to allow breathing. As a consequence, no air passes through the nose and the mouth anymore. So, the restoration of the natural airway requires to emulate the switching mechanism of the larynx, and a real-time swallowing detection would be beneficial to temporarily force the trachea to be closed with an *active closure mechanism*. However, such a system should ideally forbid any detection failure to effectively protect the airway, and allow early swallowing detection to close the trachea as soon as possible. But the current approaches have failed to meet these criteria [3].

These requirements translate into the need for measurements that provide signals with early, dedicated, and stable data about swallowing, and skeletal muscles are particularly suited. The measurements of their activity inform about the state of a movement, and the swallowing involves multiple contractions throughout the process to move the larynx, protect the airway, and force the bolus down the throat. Also, many swallowing muscles are still in place after total laryngectomy, and muscles that significantly activate at the beginning would provide valuable data. Specifically, we focused on the stylohyoid and the posterior digastric muscles that we measured with intramus-

A. Mialland is with Univ. Grenoble Alpes, CNRS, Grenoble INP, Gipsa-lab, 38000 Grenoble, France

A. Atallah is with Univ. Grenoble Alpes, Otorhinolaryngology, CHU Grenoble Alpes, 38700, La Tronche, France.

A. Bonvilain is with Univ. Grenoble Alpes, CNRS, Grenoble INP, Gipsa-lab, 38000 Grenoble, France

cular EMG, and this choice follows an extensive investigation that we conducted to characterize their timings and recruitment pattern. First off, the literature suggests their activation at the beginning of swallowing [3], but only via indirect imaging methods on humans, because of the difficulty to measure them. We, therefore, developed the first standardized procedure that allows their simultaneous and independent measurements with intramuscular EMG [4]. This allowed us to show that both muscles activate at the same time as the mylohyoid muscle [2], which is known to activate at the beginning. Also, we further evaluated their recruitment patterns and compared swallowing tasks with a large set of non-swallowing tasks. We found a clear predisposition of the stylohyoid muscle for swallowing with great stability, and the posterior digastric muscle might provide relevant data and had a characteristic burst of activity during swallowing [1]. Besides, these muscles are easily accessible during the surgery, and would not require further impairment of the neck during any sensor placement.

These results suggest that the stylohyoid and the posterior digastric muscles have a significant potential for an implanted and real-time detection of swallowing. But we only provided a characterization of their activity, and their potential must also be evaluated against detection algorithms. Besides, the requirement for an early detection stems from the fact that the airway is rapidly at risks of aspiration during swallowing. But the objective definition of a time limit, at which the airway must absolutely be closed, would allow to better evaluate the performances of a real-time detection. We defined that limit to be the moment the bolus starts to flow through the upper esophageal sphincter (UES). Indeed, this event is associated with a drastic increase in pressure and deformation in the pharynx, under the action of the pharyngeal constrictor, and is suggested to be linked to the laryngeal vestibule closure, which closes the airway during swallowing [10]. It, therefore, changes the previously relaxed conditions in the pharynx, where the bolus is essentially moved by the tongue [11], [12]. So, the moment the bolus starts to flow through the UES has been shown to be accessible through the sound produced by the swallowing and the vibration it generates, which we measured with an accelerometer [13]. This measurement gives access to various burst of sound separated in time that are linked to swallowing events, and where the passage of the bolus though the UES is the only one the be present 100% of the time [4], [13].

So, this paper intends to evaluate the potential of the stylohyoid and the posterior digastric muscles, measured with intramuscular EMG, for a real-time and implanted swallowing detection system. The submental muscles were also measured with surface electrodes to provide a basis for comparison, as they constitute a well-studied swallowing muscle group and are often included in swallowing detection strategies [3]. Finally, the moment the bolus starts to flow through the UES has been measured with an accelerometer, to define a precise time limit for the airway to be closed.

II. Materials and Methods

The data are extracted from signals that we acquired in a larger study, that we conducted to develop the first standardized procedure that allows direct functional measurement and evaluation of the stylohyoid and the posterior digastric muscles independently, through intramuscular EMG [4]. We also measured the surface EMG of the submental muscles, which we included in the current evaluation to provide a basis for comparison. Besides, the swallowing sound was measured with an accelerometer, to measure the vibration it generates through the skin, and access the moment the bolus starts to flow through the UES. This later event will be called *UES bolus flow*, for conciseness, and defines a time limit for the airway to be protected. The acquisition method is briefly detailed hereafter, but the full procedure is available in the related paper [4].

A. Subjects

Seventeen healthy adults (8 males/ 9 females) with no history of dysphagia, neck surgery, immune deficiency, or any neurological impairment participated in this study. The mean age was 36.1 ± 13.6 and all enrolled participants met the following inclusion criteria: an age of 18 years or higher and a body mass index (BMI) of 25 or lower. Participants with BMI higher than 25 were intentionally excluded from this study to avoid a potentially excessive amount of fat in the region of the neck that would make the proper placement of the sensors more complex. This study was approved by the research ethics committee of Sud-Méditerranée III of Nîmes in France (Protocol ID: 38RC22.0096).

B. Sensors placement

The participants were asked to comfortably sit on a chair and each sensor was placed by an otolaryngologist in the following order. A swallowing could be performed to confirm the absence of pain or discomfort:

Swallowing Sound: The accelerometer (TN1012/ST, 1600Hz, ADInstrument) was fixed with a hypoallergenic paper tape on top of the cricoid cartilage, as it was suggested to provide the greatest signal-to-noise ratio [14]. It is positioned to record the antero-posterior vibrations.

Surface EMG: 2 differential electrodes were placed under the left part of the submental area, with their center approximately $2cm$ apart, and a ground electrode was placed over the right clavicle. Both areas were first cleaned with a dedicated abrasive and conductive paste to reduce the electrode-skin impedance.

Intramuscular EMG: To isolate the stylohyoid and the posterior digastric activity, they are targeted close to their origin and insertion points: the stylohyoid muscle was targeted next to its insertion point, at the level of the junction between the body and the greater horn of the hyoid bone. In that region, the posterior digastric is composed of its intermediate tendon, which cannot produce EMG signals and, therefore, cannot be mistaken for the stylohyoid muscle activity. As for the posterior digastric muscle, it was targeted in its posterior portion, close to its origin at the level of the mastoid notch. In that region, it separates from the stylohyoid, which originates from the styloid process upper in the neck. Besides, both muscles are directly accessible behind the skin. While the otolaryngologist was slowly inserting the needles, a second trained operator monitored the EMG activity on the computer to look for muscle activities. It could be requested to the participant to swallow to elicit an event. Once an event was visible on the signal, the needles were no further inserted. They were then secured with a Steri-Strip so that they cannot come out of the muscles.

C. Tasks performed

4 swallowing and 13 non-swallowing tasks were performed by the participants. The non-swallowing tasks are chosen to cover the largest set of possible combinations of muscle contractions. We fused most of the tasks that have previously been considered in independent works [3] on the stylohyoid and posterior digastric muscles, with additional ones to provide a comprehensive comparison. So, each task is first prepared, then 2 seconds with no motion are acquired, the task is performed upon vocal cue, and 2 more seconds with no motion are acquired. This ensures to record a clear event. Also, a button is pressed when the vocal cue is emitted and when the task is finished, to roughly place temporal markers that delineate the tasks.

SWALLOWING TASKS: the participants performed 5 swallowing of Saliva, water ($10ml$), thick liquid (compote), and solid bolus (madeleine).

NON-SWALLOWING TASKS: the participant performed (1) 3 times mouth opening, lips purseing, teeth clenching, smiling, whistling, coughing, blowing through a straw, speaking: counting from 1 to 10, vocalizing: saying *"iii"* in ascending and descending order. (2) 3 times movement tasks being lateral jaw movements, lateral head movements, and vertical (flexion-extension) head movements. (3) 5 times chewing, that are actually recorded at the moment the solid bolus swallowing tasks is performed, with a pause in between, to separate the two. All tasks are performed to a comfortable full extent and at a natural pace.

D. Signals Acquisition

All four signals were acquired with a Bio-Amp (FE234, 4 channels) pre-amplifier and a PowerLab (35 series, 4 channels) acquisition system from ADInstrument. All signals were first processed with analog filters. A $10-1000Hz$ band-pass filter was applied on both intramuscular EMG and a $10-500Hz$ band-pass filter was applied on the surface EMG: the wider frequency band applied to the intramuscular EMG stems from its frequency content that contains higher frequencies [15]. Each channel was then sampled with a 16 bits analog-digital-converter, at a rate of $4000Hz$. Once digitized, a 2^{nd} order high-pass Butterworth digital filter with a cut-off frequency of $20Hz$ is applied on all signals, followed by a 4^{th} order Butterworth notch filter to eliminate the $50Hz$ line noise.

1) Event Onset, offset: Onset and offset of all muscles and for all activities were first localized. All raw EMG signals are transformed with the Teager-Kaiser energy operator (TKEO) $\Psi[x(n)] = x(n)^2 - x(n-1)x(n+1)$, as it was shown to improve the signal-to-noise ratio [16]. Also, EMG is prone to background noise and especially spurious spikes from intramuscular EMG, because of a few surrounding muscle fibers that react to the presence of the needles. To account for it, we used the generalized likelihood ratio (GLR) method [4], [17], computed on a $100ms$ sliding window, with an exponential probability density function to represent the TKEO-transformed raw EMG [18]. This allows to model the signals to effectively abstract from the noise, and an adaptive threshold of 10% of the maximum of the current GLR signal is placed as a guideline. In addition, the root-mean-square (RMS) amplitude is computed on all raw EMG with a $100ms$ sliding window, and a 3 standard deviation threshold is also placed as a guideline. Then, the resultant GLR and RMS signals are compared, and the onset and offset of each event are manually placed accordingly, to avoid any false positive. Also, in case of transitory or absence of activity that would make difficult to effectively locate the onset and offset, the previously described rough temporal markers, placed when the tasks is performed, were used as the default onsets and offsets.

2) UES bolus flow: To localize the UES bolus flow, the accelerometer signals are first transformed with TKEO. Indeed, the burst of sound that corresponds to the UES bolus flow is the most prominent [4], and the TKEO-transform signals allowed to further bring that event out. It could therefore easily be differentiated from other event, and its beginning could be placed manually with no difficulty [4].

E. Event Segmentation

The previously localized onsets and offsets are used to extract each event from the raw signals, which are then segmented into multiple overlapping windows of fixed length. So, each window represents an instance to be classified and defines the temporal resolution of a real-time detection strategy [19], [20]. However, the smaller the window the harder it is to obtain robust data from it that are representative of the overall ongoing event. So, we used a common value of $200ms$, with $150ms$ overlap between windows, to mitigate the effect and still increase the temporal resolution [20].

F. Feature Extraction

The higher dimensional input space of the raw windows is reduced into a lower dimensional space with the extraction of a set of selected time domain features [19]. Indeed, a large corpus of features has been proposed in the field of EMG event detection, to extract highly informative content. Also, it has been proved that in certain cases, a subset of features is sufficient to describe the overall category they belong to [21]. We therefore choose to stick with the representative ones to provide a general comparison of the signals: root mean square (RMS), moving average (MAV), wavelength (WL), second order moment (M2), zero crossing (ZC), slope sign change (SSC), simple square integral (SSI), myopulse percentage rate (MYOP), Willison amplitude (WAMP), difference variance value (DVARV), difference absolute mean value (DAMV), and difference absolute standard deviation value (DASDV). The same set of features is extracted for each window and for each signal (Table I).

Each feature is then standardized as follows:

TABLE I: Features extracted from each window and for each signal.

Feature	Equation
RMS	$\sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2}$
MAV	$\frac{1}{N}\sum_{n=1}^{N} \lvert x_n \rvert$
WL	$\sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$
M2	$\sum_{n=1}^{N-1} (x_{n+1} - x_n)^2$
ZC	$\sum_{n=1}^{N-1} [\phi(x_n \times x_{n+1}) \cap \lvert x_n - x_{n+1} \rvert \geqslant thr]$
SSC	$\sum_{n=2}^{N} \phi[(x_n - x_{n-1}) \times (x_n - x_{n+1})]$
SSI	$\sum_{n=1}^{N-1} x_n^2$
MYOP	$\frac{1}{N}\sum_{n=1}^{N} \phi(x_n)$
WAMP	$\frac{1}{N}\sum_{n=1}^{N-1} \phi(\lvert x_n - x_{n+1} \rvert)$
DVARV	$\frac{1}{N-2}\sum_{n=1}^{N-1} (x_{n+1} - x_n)^2$
DAMV	$\frac{1}{N-1}\sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$
DASDV	$\sqrt{\frac{1}{N-1}\sum_{n=1}^{N-1} (x_{n+1} - x_n)^2}$

$$\hat{x}_j^{(i)} = \frac{x_j^{(i)} - \mu_{x_j}}{\sigma_{x_j}} \tag{1}$$

Where $x_j^{(i)}$ is the $j\text{-}th$ feature from the $i\text{-}th$ feature vector, and μ_{x_j} and σ_{x_j} are its mean and standard deviation, respectively.

$$\phi(a) = \begin{cases} 1, & \text{if } a \geqslant thr \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

G. Feature Selection

Before any training of a classifier and for any signals or task combination, a feature selection process is conducted to only retain the features that add relevant information. This procedure is performed with the Boruta algorithm, which iteratively removes the features that are proved by a statistical test to be less relevant than random probes [22], [23]. More precisely, for each feature independently, a 'shadow feature' is created, which consist in the random reordering of the original feature values. The dimensionality of the data is therefore doubled, where half of the dimensions contain random fluctuations. Classifications are then performed with a random forest [24] and each feature importance is computed. Then, within the original features, only the ones that exhibit better performances than the best shadow feature are considered relevant. Finally, the procedure is repeated until the relevant features have been considered for a statistically significant number of times.

H. Classification

The extracted vectors of features constitute the dataset used to train and evaluate a classifier and is split into a training set (85%) and test set (15%), with a stratified sampling. Each vector is associated a label of 1 if it comes from a swallowing task, and 0 if it comes from a non-swallowing task. Tow classifiers are considered: LDA [25] and SVM [26]. While the LDA is a linear classifier, the SVM is more versatile and can build a wide range of decision functions using kernel functions. In our approach, we employed the SVM with a radial basis function (RBF) kernel to evaluate the benefit of a non-linear classifier.

To deal with the unbalanced dataset, precision, recall, and F1-scores were used to evaluate the performances of each classifier [27]. Also, recall is calculated for swallowing again, but at the task level, which we call *S-Recall*. It is defined as the number of correctly classified swallowing over the total number of swallowing. A swallowing is considered correctly classified if at least one associated feature vector is correctly classified.

Finally, for any training of a classifier and any signals or task combination, a 5-fold cross-validation with 10 repetitions and a stratified sampling is performed on the training set to look for the hyper-parameters that maximize the F1-scores.

I. EMG Signals evaluation

To evaluate the performances of the stylohyoid and the posterior digastric muscles, they are compared with the submental muscles in three different scenarios. For each scenario, all 7 combinations of the 3 signals are compared:

1) Benchmark Scenario: The literature provides very few studies to compare with, and none of them included the stylohyoid and the posterior digastric muscles. Therefore, to provide meaningful comparisons and evaluate the potential benefit of adding these two signals, the most relevant studies are used as a benchmark, from which we reproduced subgroups of comparable tasks. First off, [8] used the submental sEMG signal to record regular, effortful, and Mendelsohn maneuver swallowings, along with lip press, tongue movements, and head movements, in an attempt to develop a swallowing rehabilitation tool. To compare with their setup, we used all 4 swallowing tasks that we recorded along with head lateral, head vertical, and lips purseing tasks. Besides, [28] used submental, intercostal, and diaphragm surface EMG, to compared healthy volunteers and total laryngectomees in a swallowing detection strategy. It involved saliva, water, and banana swallowing, along with coughing, speaking, standing, reaching overhead, twisting, walking, and sitting. We compare their approach against our 4 swallowing tasks, along with coughing, speaking, vocalizing, head vertical, and head lateral tasks.

2) Tasks Combinations Scenario: to better evaluate the signals in relation to the various tasks, a sequential forward selection strategy is applied and adapted from its standard usage for feature selection [29]. It basically selects the best scoring tasks, based on F1-score, starting with all 4 swallowing tasks as a basis plus the selected task. Then, it iteratively adds one task, selected as the best among the remaining tasks, until all 17 tasks have been selected. Therefore, it allows to track the evolution of the signals performances in swallowing detection as more tasks are added.

3) Subject Clustering Scenario: to better explore the variability among participants, we used a clustering strategy to gather similar individuals into 2 subgroups, which could potentially improve discrimination ability while still retaining inter-subject variability. To do so, we used an agglomerative hierarchical clustering which iteratively gather subgroups into pairs [30]. More precisely, initial clusters are irreversibly agglomerated into similar pairs until the required number of groups is achieved, and where initial clusters are defined by each single data point being part of its own cluster. However, to gather similar participants, we used a slightly different version where each initial cluster is defined by the data points from each participant. Also, only the data points from the swallowing tasks are used to gather participants by their swallowing similarities while non-swallowing tasks can still express substantial variability. To assess the similarity between tow clusters, we used the average euclidean distance between each point of each cluster. Finally, a Wilcoxon signed ranks test [31] is used to compare the classification performances of each of the 3 groups based on F1-score.

III. Results

All 17 participants completed the full acquisition procedure. Few tasks could not be processed for feature extraction, because of bad signal quality, and were therefore excluded. Once the feature vectors were extracted, and depending on signal combinations, it yielded a maximum dataset size of 5441 swallowing vectors and 44314 non-swallowing vectors, and a minimum dataset size of 4243 swallowing vectors and 32422 non-swallowing vectors.

A. Benchmark Scenario

Among all tasks, specific sub-groups were chosen to allow comparisons with previously reported works:

First off, Table II shows our classification outcomes that we compare to the results of [8]. They used submental surface EMG to detect swallowing as a whole event and got a precision of 83.9% (12.9), a recall of 92.3% (10.4), and an F1-score of 87.9%. As for our results, while the SVM consistently outperformed the LDA classifier, we were able to recover comparable performances with the submental surface EMG as well. Only the posterior digastric performed worse than [8], and every other combination yielded better results, with increasing performances as more signals are considered. The best performances are obtained with all 3 channels. Also, the stylohyoid was consistently part of the combinations that yield the best and outperformed the results of [8], even when considered alone.

Then, Table III shows our classification outcomes that we compare to the results of [28]. They used submental, intercostal, and diaphragm surface EMG to detect segmented swallowing events and got a precision of 77% (8), a recall of 57% (3), and a F1-score of 66%. As for our results, while the SVM consistently outperformed the LDA classifier, we were able to recover comparable performances with the posterior digastric muscle alone, even though the setup is not strictly the same. However, the posterior digastric yielded the worst performances and every other combination yielded better results than [28], with increasing performances as more signals are considered. The best performances are obtained with all 3 channels. Also, the stylohyoid was consistently part of the combinations that yielded the best, even when considered alone.

Finally, for both comparison, the posterior digastric was still able to provide among the best results when combined with the stylohyoid, despite yielding the worst when considered alone.

B. Tasks Combinations Scenario

A sequential forward selection strategy was used and adapted to task selection to evaluate the effect of an increasing number of tasks on the classification performances, based on F1-score. It started with the 4 swallowing tasks and progressively added, one by one, all 13 non-swallowing tasks.

First off, Figure 1 shows the results for each of the 7 signal combinations, both for LDA and SVM classifiers, which makes a total of 14 F1-score curves dependent on tasks number. So, the SVM consistently outperformed the LDA classifier, except for the submental muscles, where both classifiers provided comparable results. However, The posterior digastric yielded the worst anyway, followed by the submental muscle and the stylohyoid. The F1-scores then kept increasing as more signals combination were considered, with the best performances obtained with all 3 channels. More specifically, when all 13 non-swallowing tasks are considered, the 3 best F1-scores were obtained for the SH-PD-SM with 84.98%, the SH-PD with 72.28%, and the SH-SM with 65.84% (detailed in the last scenario, Table IV). But more interestingly and for every case, when 12 out of 13 non-swallowing tasks are considered, a general trend with a drastic improvement can be seen, where the 3 best F1-score were obtained for the SH-PD-SM with 95.75%, the SH-PD with 90.94%, and the SH-SM with 90.66%. Finally, the stylohyoid was consistently part of the combinations that yielded the best and the posterior digastric was still able to provide among the best results when combined with the stylohyoid, despite yielding the worst when considered alone.

However, this only gives an overview of the performances based on the number of tasks considered for classification. Therefore, in order to look for any trend in their individual performances, the Figure 2 shows the ranks of each task on a radar plot, obtained from all 14 F1-score curves, along with their mean rank. So, it reveals that most of the tasks are likely to fall in a limited range of rank, with a minimum average rank of 2.5 for the lateral head task, and a maximum average rank of 13 for the mastication task. Also, the latter implies that mastication is always found in the 13th rank and is, therefore, the task responsible for the drastic drop in F1-score when all tasks are considered (Figure 1). Moreover, mastication is followed by the lateral jaw and mouth opening tasks, which are respectively ranked 10.8 and 10.3 on average. Yet, these three tasks are the most related to swallowing and more precisely to the oral preparatory stage of swallowing [3].

C. Subject Clustering Scenario

The variability among participants was explored with a clustering method, to identify potential sub-groups of similar participants. No more than 2 clusters were extracted

TABLE II: Classification results of our sub-group of tasks (saliva, water, solid, thick, head lateral, head vertical, lips purseing) chosen to be compared with the results of [8]. They used submental surface EMG to detect swallowing as a whole event. We got comparable results with the submental muscles as well and got improved performances with the stylohyoid and the posterior digastric.

	LDA				SVM			
	Precision	Recall	F1-score	S-Recall	Precision	Recall	F1-score	S-Recall
SH-PD-SM	95.31 (0.41)	89.72 (1.25)	92.43 (0.62)	98.27 (0.54)	98.87 (0.37)	99.31 (0.29)	99.09 (0.15)	99.94 (0.16)
SH-SM	94.26 (0.66)	90.93 (1.13)	92.57 (0.68)	98.11 (0.87)	98.54 (0.45)	97.83 (0.64)	98.18 (0.38)	99.85 (0.23)
PD-SM	84.03 (1.32)	80.89 (1.43)	82.43 (1.09)	92.88 (1.12)	95.74 (0.61)	94.43 (0.86)	95.08 (0.52)	99.03 (0.63)
SH-PD	95.07 (0.63)	89.97 (0.95)	92.45 (0.59)	97.66 (0.69)	98.42 (0.43)	97.97 (0.47)	98.19 (0.31)	99.74 (0.35)
SH	93.87 (0.98)	92.17 (0.87)	93.01 (0.49)	97.91 (0.71)	96.93 (0.52)	95.58 (0.66)	96.25 (0.37)	99.24 (0.59)
PD	81.15 (2.12)	47.07 (2.59)	59.58 (2.28)	61.34 (2.61)	81.31 (1.61)	70.06 (2.41)	75.27 (1.88)	88.06 (2.28)
SM	83.06 (1.21)	80.65 (1.48)	81.83 (0.86)	91.66 (1.24)	88.44 (1.13)	83.54 (1.38)	85.92 (0.78)	94.93 (1.29)

Closest model to the results of [8]; SH: stylohyoid, PD: posterior digastric, SM: submental; Values are mean (SD).

TABLE III: Classification results of our sub-group of tasks (saliva, water, solid, thick, coughing, speaking, vocalizing, head vertical, head lateral) chosen to be compared with the results of [28]. They used submental, intercostal, and diaphragm surface EMG to detect segmented swallowing events. We got comparable results with the posterior digastric alone and got improved performances with the stylohyoid or its combination with the posterior digastric.

	LDA				SVM			
	Precision	Recall	F1-score	S-Recall	Precision	Recall	F1-score	S-Recall
SH-PD-SM	92.51 (0.52)	81.58 (1.38)	86.71 (0.91)	94.59 (1.13)	97.79 (0.45)	98.02 (0.52)	97.91 (0.32)	99.78 (0.32)
SH-SM	92.56 (0.72)	83.01 (1.31)	87.53 (0.81)	94.74 (0.84)	95.86 (0.76)	94.19 (0.81)	95.02 (0.55)	99.26 (0.58)
PD-SM	86.67 (1.24)	74.76 (1.56)	80.28 (1.09)	89.98 (1.29)	93.42 (1.03)	90.99 (1.06)	92.19 (0.74)	98.37 (0.82)
SH-PD	90.82 (0.96)	82.38 (1.27)	86.39 (0.84)	94.91 (1.07)	96.01 (0.75)	94.64 (0.69)	95.32 (0.56)	99.32 (0.46)
SH	90.92 (1.01)	84.87 (1.16)	87.79 (0.81)	94.74 (1.09)	93.31 (0.96)	88.55 (1.13)	90.87 (0.81)	96.78 (0.98)
PD	75.03 (2.58)	36.54 (1.78)	49.15 (1.95)	50.79 (2.51)	75.65 (2.15)	58.88 (2.35)	66.22 (1.88)	78.95 (2.82)
SM	85.07 (1.04)	73.87 (1.48)	79.08 (1.02)	88.21 (1.42)	90.44 (0.97)	74.94 (1.64)	81.96 (1.08)	89.19 (1.54)

Closest model to the results of [28]; SH: stylohyoid, PD: posterior digastric, SM: submental; Values are mean (SD).

to preserve relevant diversity and datasets sizes, where the cluster 1 contained 8 participants, and the cluster 2 contained 9 participants. The performances of each cluster were then compared to all participants and the results are reported in Table IV. Only the SVM classifier is reported for clarity, and because of the poor results the LDA classifier provides (Figure 1).

Both clusters yielded significant F1-score improvement but only cluster 1 was consistent across signal combination, while cluster 2 had no improvement for the posterior digastric and a slight deterioration for the stylohyoid, when they are considered alone. The largest increase in F1-score was 12.34% from cluster 1, when the posterior digastric and the submental muscles are combined. In all cases, when considered alone, the posterior digastric yielded the worst, followed by the submental muscles, while the stylohyoid showed a drastic improvement. The performances then kept increasing as the signals were considered in combination and the stylohyoid was consistently part of the combinations that yielded the best.

However, this only gives an overview of the performances without a temporal limit for a detection. Therefore, in order to assess the potential of each signal combination to provide an early detection, the Figure 3 shows the cumulative frequencies of their S-Recall scores, based on the earliest moment a swallowing is detected. Thus, the maximum S-Recall values are those available in Table IV, when the full swallowing event is finished. Besides, the time was normalized to the UES bolus flow time, used as the 0 *second* reference and that we defined as the temporal limit for a detection. Finally, the curves are normalized to their maximum S-Recall value (Table IV) to focus on earliness. So, the earliest detection was systematically obtained with the posterior digastric muscle, where the group of all participants reached 79.39% of its maximum S-Recall value before the UES bolus flow, 64.55% for cluster 1, and 83.3% for cluster 2. Also, the latter gave a statistically significantly superior value but, overall, no clear tendency emerged and the capabilities of the signals to provide an early detection either remained the same or worsened. Finally, the posterior digastric was consistently part of the signal combinations that yielded the best, while the submental muscle, the stylohyoid, and their combination yielded the worst, with a lowest value of 31.79% for the submental muscle from cluster 2.

IV. Discussion

A real-time and implantable swallowing detection would be beneficial for the conception of an implantable active artificial larynx, following the full resection of the larynx. The surgery, called total laryngectomy, impairs the natural airways, and causes the loss of the protective mechanism of the trachea during swallowing. Consequently, the trachea is sewn on the anterior neck (tracheostomy), and any attempt to set it back in place requires to emulate the protective mechanism of the larynx, to close the airway as early as possible during swallowing. Therefore, on the basis of our previous investigations [1]–[4], we focused on an active mechanism and evaluated the potential of the stylohyoid and the posterior digastric muscles for a real-time and implantable swallowing detection. They were measured with intramuscular EMG, along with the submental muscle with surface EMG, to provide a basis for comparison,

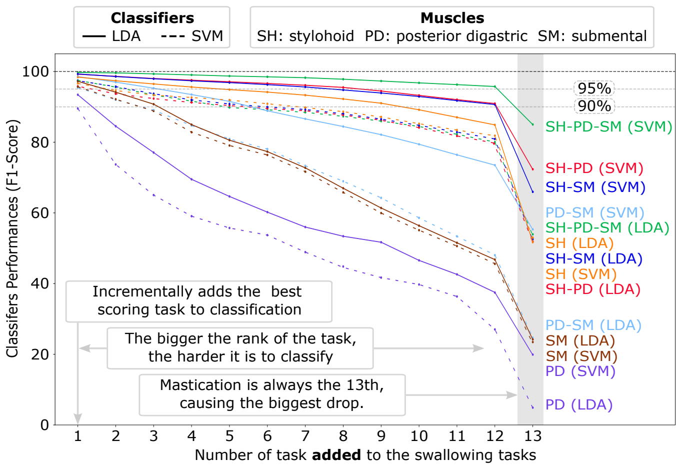

Fig. 1: SH: stylohyoid, PD: posterior digastric, SM: submental; Number of tasks added to the swallowing tasks that shows the effect of an increasing number of tasks on F1-score. Only the task number is expressed as each curve yielded a different arrangement. The SVM outperformed the LDA, SH is included in the best combinations, and a drastic drop is systematically observed when a 13th task is added.

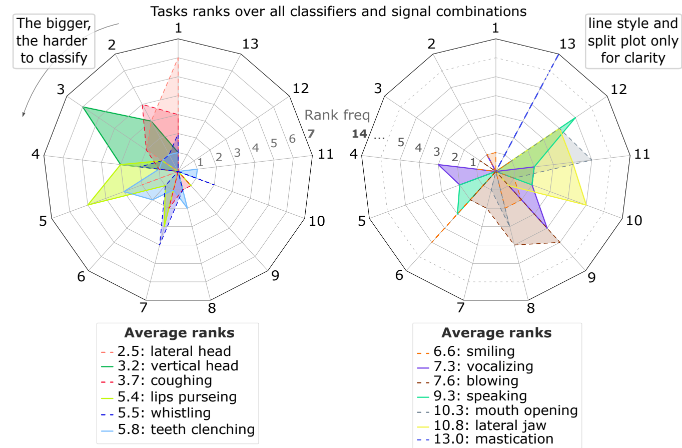

Fig. 2: Radar plot of the ranks of each task obtained from all 14 F1-scores curves from Figure 1. The legend expresses the tasks with their average rank, in an ascending order. Most of the tasks are likely to fall within a limited range of ranks, the mastication task is always at the 13th rank, and the 3 last tasks are the most related to the oral preparatory stage of swallowing. Note: line style is only for clarity.

and the UES bolus flow, to define a temporal limit for the detection.

The stylohyoid and the posterior digastric were evaluated in 3 scenarios. First, they were compared with previous representative work. Second, we explored the effect of an increasing number of tasks on their detection performances. Third, we assessed the effect of similar subjects clustering on their detection performances and their potential to provide an early detection. So, as a general trend, the stylohyoid was consistently part of the signal combinations that yielded the best, which is supported by our previous investigations on its recruitment pattern. Indeed, it exhibited marked, high and stable amplitude during swallowing, compared to most of the reaming tasks [1]. Besides, the posterior digastric alone gave the worst results but was still able to yield among the best when combined with the stylohyoid. These results are reinforced by previous works that studied the nerve conduction of both muscles [32]. The authors showed that they are the only swallowing muscles innervated by the facial nerve, which is primarily involved in mimicry, while activating predominantly for swallowing and mouth opening. The authors highlighted that their intermediary position may facilitate electrophysiological identification of swallowing. Therefore, along with our findings, this could underscore the ability of the stylohyoid and the posterior digastric muscles to work in synergy.

Besides, our comparison with tow previously reported work show the substantial improvement that our approach gives, and further emphasis the disadvantage that the submental muscles have in comparison to the stylohyoid and the posterior digastric muscles. Also, the approach of [28] that used intercostal and diaphragm EMG was easily outperformed by our approach. This may be due to the inherent link of the intercostal and diaphragm muscles to the respiratory system, which is not only involved in swallowing but also in various activity. In comparison, the stylohyoid and the posterior digastric muscles tends to predominantly activate for swallowing [1] and are, ultimately, directly linked to the larynx.

With regard to the classification of a various number of tasks (Figure 1), and the ranking of their performances (Figure 2), we showed that the most related task to swallowing (mastication, lateral jaw, mouth opening) are the ones responsible for the biggest drop in F1-score. In addition, mastication was always found at the latest rank, which emphasizes the necessity to engage in a broader strategy, for an effective detection. For example, while the swallowing bolus transportation (pharyngeal) stage should obviously remain at the center of the strategy, including muscles that are representative of the preceding oral preparatory stage could substantially increase the performances. Indeed, even when removing the mastication tasks only, the F1-score could reach 95.75% (+10.77%) for the SH-PD-SM, 90.94% (+18.66%) for the SH-PD, and 90.66% (+24.82%) for the SH-SM signal combinations. However, the fusion of data from multiple sensors and neck area, to maximize the variety of informative content, is not new and some work has shown remarkable results in the past [33]–[35]. However, these often require the larynx to be still in place and do not propose particular tasks comparison. Yet, the stylohyoid and the posterior digastric muscles are still in place after a total laryngectomy, with little to no impairments, and are easily accessible [3]. So, the simple yet informative task ranking we propose (Figure 2) offers a guideline to focus on tasks or families of tasks that could improve the detection performances.

As for the clustering strategy, the 2 clusters of participants yielded statistically significant improvements, with particular consistency with cluster 1, over the group of all participants

TABLE IV: SVM performances following the clustering of similar participants into 2 groups. Cluster 1: 8, cluster 2: 9. Both clusters yielded increased performances, but with more consistency in cluster 1.

	Precision	Recall	F1-score	S-Recall
All Participants				
SH-PD-SM	90.36 (1.03)	80.27 (1.31)	84.98 (0.76)	96.31 (1.49)
SH-SM	82.51 (1.52)	54.77 (1.56)	65.84 (1.33)	84.83 (2.31)
PD-SM	80.79 (2.01)	41.98 (1.49)	55.25 (1.51)	70.65 (2.21)
SH-PD	76.72 (1.01)	68.39 (1.51)	72.28 (1.44)	90.11 (1.44)
SH	73.86 (1.86)	39.71 (1.74)	51.65 (1.68)	68.46 (2.58)
PD	61.51 (6.01)	11.81 (0.93)	19.81 (1.92)	19.95 (1.81)
SM	47.71 (3.11)	15.91 (1.29)	23.85 (1.74)	33.51 (1.57)
Cluster 1				
SH-PD-SM	92.65 (1.29)	86.47 (1.67)	89.45 (1.13)*	97.71 (1.33)
SH-SM	86.81 (1.64)	63.82 (2.41)	73.56 (1.83)*	90.76 (2.26)
PD-SM	86.17 (2.23)	55.62 (2.53)	67.59 (2.11)*	81.93 (3.04)
SH-PD	86.37 (2.01)	69.58 (2.37)	77.07 (1.74)*	90.56 (2.33)
SH	77.22 (2.37)	49.23 (2.43)	60.13 (2.15)*	77.17 (3.69)
PD	75.45 (9.94)	15.01 (1.21)	25.04 (2.91)*	26.77 (1.76)
SM	53.39 (4.07)	24.84 (2.15)	33.87 (2.57)*	36.38 (3.82)
Cluster 2				
SH-PD-SM	91.77 (1.45)	85.06 (1.58)	89.29 (1.11)*	96.81 (1.42)
SH-SM	83.35 (2.28)	62.61 (2.06)	71.64 (1.74)*	89.56 (2.65)
PD-SM	81.55 (2.44)	46.78 (2.31)	59.45 (2.05)*	71.46 (3.21)
SH-PD	80.67 (1.66)	73.91 (1.86)	77.14 (1.52)*	93.22 (2.11)
SH	72.52 (2.91)	34.98 (2.34)	47.19 (2.34)	58.06 (3.34)
PD	75.92 (8.98)	10.96 (1.71)	19.15 (3.19)	20.05 (2.81)
SM	47.52 (3.52)	21.94 (1.13)	29.98 (2.01)*	36.25 (1.05)

■ biggest improvement of the clusters over all participants, per signal combination; SH: stylohyoid, PD: posterior digastric, SM: submental, * $p < 0.001$: F1-scores comparison between all participants and the clusters with a Wilcoxon signed ranks test.

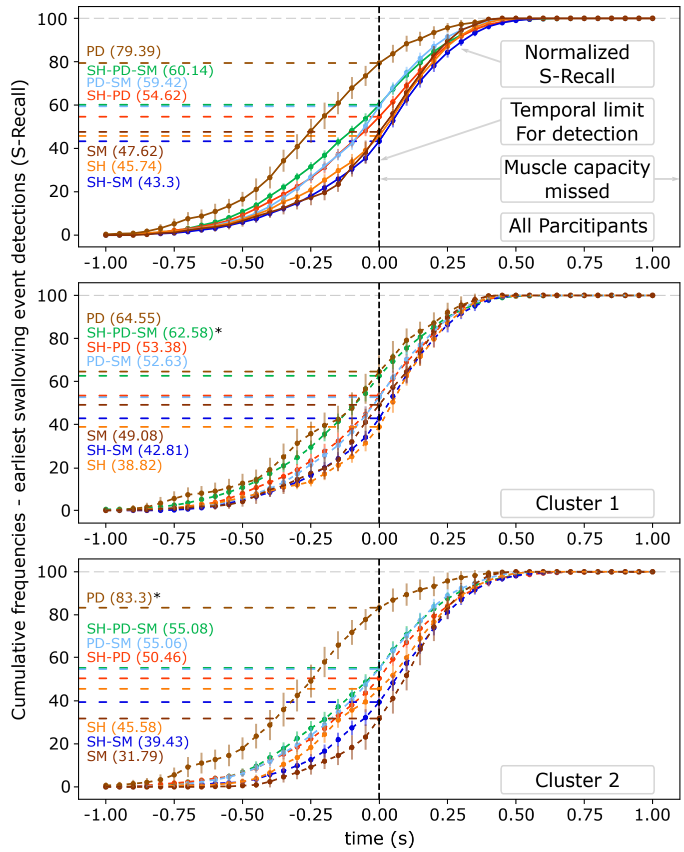

Fig. 3: SH: stylohyoid, PD: posterior digastric, SM: submental, * $p < 0.001$: F1-scores comparison, at the temporal limit for the detection, between all participants and the clusters, with a Wilcoxon signed ranks test; Cumulative frequencies of the S-Recalls, based on the earliest moment a swallowing is detected, and normalized to their maximum values (Table IV). The posterior digastric provided the earliest detection, with a net improvement in cluster 2. But overall, the clusters tend to worsen the performances.

(Table IV). This highlights the potential of employing a more targeted detection strategy, wherein sub-groups of representative individuals can introduce valuable diversity while preserving specific characteristics. In that sense, [8] built their algorithm with all participants, but evaluated the performances at the individual level, both from healthy and head and neck cancer (HNC) survivors. They emphasized the necessity of a more stratified approach, especially in the HNC group to mitigate the effect of the heterogeneity of the profiles.

Besides, the ability of each signal combination to provide an early detection (Figure 3) shows a clear advantage of the posterior digastric alone, but decreased performances for the others. Yet, our previous investigation on their timings [2] showed that the posterior digastric peak activity occurs first, with the majority of its activity that precedes the UES bolus flow. On the other hand, the peak activity of the stylohyoid and the submental muscles occurs later, respectively slightly around and before the UES bolus flow, with their activity that lasts throughout the whole swallowing. So, even though all muscles start at the beginning of swallowing [2], this probably explains why they tend to yield the worst in terms of earliness of detection. Besides, the swallowing is a transient event that essentially provide dynamic EMG signals, and the field of prosthetic limb myoelectric control has previously emphasized the increased difficulty to classify transient EMG in real-time and with short time windows [36]. The increase in signal stationarity improves the performances [37], and prosthetic limbs actually take advantage of the possibility to generate static signals through voluntary contraction of residual limb muscles. Finally, the clustering strategy added little improvement and actually tended to worsen the performances overall. Thus, all these aspects may suggest the requirement of an algorithm that is inherently centered around a timely constrained detection, which is not the case for the standard classifiers we used.

So, from the perspective of an implantable active artificial larynx, an effective detection requires measurements that provide signals with early, dedicated, and stable data about swallowing. So, the stylohyoid clearly outperformed the submental muscles in terms of F1-score (Figure 1), and the posterior digastric showed the best results in terms of earliness (Figure 3). However, these two muscles did not outperform both F1-score and earliness at the same time, but their combination still drastically improves the detection performances (Figure 1). Thus, they would likely benefit from algorithms that effectively capture the temporal dynamics of the signals, and the field of early time series detection has provided various examples that introduce temporal constraints [38]. Also, dealing with specific tasks could improve the performances (Figure 2) and, for these reasons, we confirm the potential of the stylohyoid and the posterior digastric muscles for a real-time and implantable detection of swallowing, that should be included in further development.

Finally, it should be emphasized that the present stylohyoid and posterior digastric EMG signals contain inherent imperfections because of the measurement method. First off, the concentric needles used for intramuscular EMG only reflect the activity of a small region of the muscle, because of the little recording area they have [39]. Also, concentric needles tend to move within the muscle, which may add undesired instability in the signals, in case of any regional firing. On the other hand, surface EMG electrodes provide a more global assessment of the muscle, by averaging the activity of fibers from a broader area. The resultant signals are, therefore, often more stable, with a higher signal-to-noise ratio. As a consequence, the extracted feature vectors from the stylohyoid and the posterior digastric muscles may contain additional variability, and any implantable EMG sensor would reduce that effect, as they provide a recording area comparable to the surface electrodes [40]. More specifically, epimysial electrodes are essentially the implantable counterpart of the surface electrodes, and is meant to be sewn on top of the muscle. In addition, this type of electrode has recently shown its long term implantable, along with the possibility to effectively manipulate a robotic prosthetic arm in real-time [41], [42]. Besides, the feasibility of an implantable active artificial larynx depends on the robustness of the detection algorithm and its ability to deal with the various contexts encountered on a daily basis. So, while the development of such algorithm should be the focus of future work, its flexibility heavily depends on the type of information it has at hand. Therefore, the stylohyoid and the posterior digastric muscles are essential, but future research should probably look for additional signals of different nature that would diversify the view of the neck activity. For instance, complementary signals about a family of non-swallowing task could facilitate their discrimination. Also, the measurement of sensory inputs could provide essential data, and recent works have provided innovative electrodes, which have shown long term implantability [43], [44]. However, minimizing the number implanted sensors is essential and should be considered in future research. Finally, this paper focus on the detection aspects of an implantable active artificial larynx but, while the field of robotic prosthetic limb has already shown the possibility of a real-time detection of muscle events [45], future research should initiate the development of a protective mechanism, to better evaluate the feasibility of such a system as a whole.

V. Conclusion

The development of a real-time swallowing detection for an implantable active artificial larynx requires to emulate the airway protective mechanism of the larynx. Therefore, this paper evaluated the potential of the stylohyoid and the posterior digastric muscles measured with intramuscular EMG. The submental muscles were also measured with surface EMG to provide a basis for comparison, and the swallowing sound was measured with an accelerometer to define a temporal limit. The latter allowed to assess the ability of each signal and their combinations to provide an early detection. So, we showed that the stylohyoid outperformed the submental muscles in terms of F1-score and were consistently part of the combinations that yielded the best. As for the posterior digastric, it was particularly effective in providing an early detection. However, these two muscles alone did not outperformed both F1-score and earliness at the same time, but their combinations still provide among the best results. Also, we found that the tasks related to the oral preparatory stage of swallowing are responsible for the biggest drop in performances, and a real-time detection of swallowing would benefit from a more stratified and timely constrained algorithm. Thus, we confirm the interest of the stylohyoid and the posterior digastric muscle for future investigations.

Acknowledgments

This research has been carried out with funding from the Région Auvergne Rhône Alpes.

References

[1] A. Mialland, I. Atallah, and A. Bonvilain, "Stylohyoid and posterior digastric recruitment pattern evaluation in swallowing and non-swallowing tasks," *Innovation and Research in BioMedical engineering*, 2023, submitted.

[2] ——, "Stylohyoid and posterior digastric timing evaluation," *Engineering in Medicine and Biology Conference*, 2023, submitted.

[3] ——, "Toward a robust swallowing detection for an implantable active artificial larynx: a survey," *Medical & Biological Engineering & Computing*, 2023.

[4] ——, "Stylohyoid and posterior digastric measurement with intramuscular emg, submental emg and swallowing sound." *Biomedical Engineering Systems and Technologies*, 2022.

[5] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, pp. 381–386, 2020.

[6] C. M. Steele, "The Blind Scientists and the Elephant of Swallowing: A Review of Instrumental Perspectives on Swallowing Physiology," *Journal of Texture Studies*, vol. 46, pp. 122–137, 2015.

[7] Y. Khalifa, J. L. Coyle, and E. Sejdić, "Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings," *Scientific Reports*, vol. 10, p. 8704, 2020.

[8] G. Constantinescu, K. Kuffel, D. Aalto, W. Hodgetts, and J. Rieger, "Evaluation of an Automated Swallow-Detection Algorithm Using Visual Biofeedback in Healthy Adults and Head and Neck Cancer Survivors," *Dysphagia*, vol. 33, pp. 345–357, 2018.

[9] S. M. Shaw and R. Martino, "The Normal Swallow: Muscular and Neurophysiological Control," *Otolaryngologic Clinics of North America*, vol. 46, pp. 937–956, 2013.

[10] A. Kurosu, J. L. Coyle, J. M. Dudik, and E. Sejdic, "Detection of Swallow Kinematic Events From Acoustic High-Resolution Cervical Auscultation Signals in Patients With Stroke," *Archives of Physical Medicine and Rehabilitation*, vol. 100, pp. 501–508, 2019, number: 3.

[11] M. Hasegawa, M. Kurose, K. Okamoto, Y. Yamada, T. Tsujimura, M. Inoue, T. Sato, T. Narumi, N. Fujii, and K. Yamamura, "Differential Response Pattern of Oropharyngeal Pressure by Bolus and Dry Swallows," *Dysphagia*, vol. 33, pp. 83–90, 2018.

[12] A. Pal, R. B. Williams, I. J. Cook, and J. G. Brasseur, "Intrabolus pressure gradient identifies pathological constriction in the upper esophageal sphincter during flow," *American Journal of Physiology-Gastrointestinal and Liver Physiology*, vol. 285, pp. G1037–G1048, 2003.

[13] S. Morinière, M. Boiron, L. Brunereau, P. Beutter, and F. Patat, "Pharyngeal Swallowing Sound Profile Assessed after Partial and Total Laryngectomy," *Dysphagia*, vol. 26, pp. 366–373, 2011.

[14] J. M. Dudik, J. L. Coyle, and E. Sejdić, "Dysphagia Screening: Contributions of Cervical Auscultation Signals and Modern Signal-Processing Techniques," *IEEE Transactions on Human-Machine Systems*, vol. 45, pp. 465–477, 2015.

[15] R. Merletti and P. Di Torino, "Standards for reporting emg data," *J Electromyogr Kinesiol*, vol. 9, no. 1, pp. 3–4, 1999.

[16] S. Wang, S. Zhu, and Z. Shang, "Comparison of different algorithms based on TKEO for EMG change point detection," *Physiological Measurement*, vol. 43, p. 075001, 2022.

[17] Q. Xu, Y. Quan, L. Yang, and J. He, "An Adaptive Algorithm for the Determination of the Onset and Offset of Muscle Contraction by EMG Signal Processing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, pp. 65–73, 2013.

[18] S. E. Selvan, D. Allexandre, U. Amato, and G. H. Yue, "Unsupervised Stochastic Strategies for Robust Detection of Muscle Activation Onsets in Surface Electromyogram," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, pp. 1279–1291, 2018.

[19] A. Phinyomark, E. Campbell, and E. Scheme, "Surface electromyography (emg) signal processing, classification, and practical considerations," *Biomedical Signal Processing: Advances in Theory, Algorithms and Applications*, pp. 3–29, 2020.

[20] A. Jaramillo-Yánez, M. E. Benalcázar, and E. Mena-Maldonado, "Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review," *Sensors*, vol. 20, p. 2467, 2020.

[21] A. Phinyomark, R. N. Khushaba, E. Ibáñez-Marcelo, A. Patania, E. Scheme, and G. Petri, "Navigating features: a topologically informed chart of electromyographic features space," *Journal of The Royal Society Interface*, vol. 14, p. 20170734, 2017.

[22] M. B. Kursa, A. Jankowski, and W. R. Rudnicki, "Boruta–a system for feature selection," *Fundamenta Informaticae*, vol. 101, no. 4, pp. 271–285, 2010.

[23] M. B. Kursa and W. R. Rudnicki, "Feature selection with the boruta package," *Journal of statistical software*, vol. 36, pp. 1–13, 2010.

[24] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[25] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI communications*, vol. 30, no. 2, pp. 169–190, 2017.

[26] C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.

[27] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.

[28] J. McNulty, K. de Jager, H. T. Lancashire, J. Graveston, M. Birchall, and A. Vanhoestenberghe, "Prediction of Larynx Function Using Multichannel Surface EMG Classification," *IEEE Transactions on Medical Robotics and Bionics*, pp. 1032–1039, 2021.

[29] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[30] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, pp. 86–97, 2012.

[31] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine learning research*, vol. 7, pp. 1–30, 2006.

[32] T. Kurt, N. Gürgör, Y. Seçil, N. Yıldız, and C. Ertekin, "Electrophysiologic identification and evaluation of stylohyoid and posterior digastricus muscle complex," *Journal of Electromyography and Kinesiology*, vol. 16, pp. 58–65, 2006.

[33] S. Roldan-Vasco, S. Restrepo-Agudelo, Y. Valencia-Martinez, and A. Orozco-Duque, "Automatic detection of oral and pharyngeal phases in swallowing using classification algorithms and multichannel EMG," *Journal of Electromyography and Kinesiology*, vol. 43, pp. 193–200, Dec. 2018.

[34] M. A. Crary, L. Sura, and G. Carnaby, "Validation and Demonstration of an Isolated Acoustic Recording Technique to Estimate Spontaneous Swallow Frequency," *Dysphagia*, vol. 28, pp. 86–94, 2013.

[35] H. Nahrstaedt, C. Schultheiss, R. Seidl, and T. Schauer, "Swallow Detection Algorithm Based on Bioimpedance and EMG Measurements," *IFAC Proceedings Volumes*, vol. 45, pp. 91–96, 2012.

[36] T. Lorrain, N. Jiang, and D. Farina, "Influence of the training set on the accuracy of surface EMG classification in dynamic contractions for the control of multifunction prostheses," *Journal of NeuroEngineering and Rehabilitation*, vol. 8, p. 25, 2011.

[37] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, "Feature extraction of the first difference of EMG time series for EMG pattern recognition," *Computer Methods and Programs in Biomedicine*, vol. 117, pp. 247–256, 2014.

[38] A. Gupta, H. P. Gupta, B. Biswas, and T. Dutta, "Approaches and applications of early classification of time series: A review," *IEEE Transactions on Artificial Intelligence*, vol. 1, pp. 47–61, 2020.

[39] R. Merletti and D. Farina, "Analysis of intramuscular electromyogram signals," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, pp. 357–368, 2009.

[40] K. A. Yildiz, A. Y. Shin, and K. R. Kaufman, "Interfaces with the peripheral nervous system for the control of a neuroprosthetic limb: a review," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, p. 43, 2020.

[41] M. Ortiz-Catalan, E. Mastinu, P. Sassu, O. Aszmann, and R. Brånemark, "Self-Contained Neuromusculoskeletal Arm Prostheses," *New England Journal of Medicine*, vol. 382, pp. 1732–1738, 2020.

[42] M. Ortiz-Catalan, B. Håkansson, and R. Brånemark, "An osseointegrated human-machine gateway for long-term sensory feedback and motor control of artificial limbs," *Science Translational Medicine*, vol. 6, 2014.

[43] P. Sabetian and P. Yoo, "Feasibility of differentially measuring afferent and efferent neural activity with a single nerve cuff electrode," *Journal of Neural Engineering*, 2020.

[44] P. Sabetian and P. B. Yoo, "Optimizing a novel nerve cuff electrode to record bidirectional neural activity," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2019.

[45] N. Parajuli, N. Sreenivasan, P. Bifulco, M. Cesarelli, S. Savino, V. Niola, D. Esposito, T. J. Hamilton, G. R. Naik, U. Gunawardana *et al.*, "Real-time emg based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation," *Sensors*, vol. 19, 2019.