Prediction of larynx function using multichannel surface EMG classification

Johnny McNulty*, Kylie de Jager, Henry T. Lancashire, James Graveston, Martin Birchall, Anne Vanhoestenberghe

Abstract—Total laryngectomy (TL) affects critical functions such as swallowing, coughing and speaking. An artificial, bioengineered larynx (ABL), operated via myoelectric signals, may improve quality of life for TL patients. To evaluate the efficacy of using surface electromyography (sEMG) as a control signal to predict instances of swallowing, coughing and speaking, sEMG was recorded from submental, intercostal and diaphragm muscles. The cohort included TL and control participants. Swallowing, coughing, speaking and movement actions were recorded, and a range of classifiers were investigated for prediction of these actions. Our algorithm achieved F1-scores of 76.0 ± 4.4 % (swallows), 93.8 ± 2.8 % (coughs) and 70.5 ± 5.4 % (speech) for controls, and 67.7 ± 4.4 % (swallows), 71.0 ± 9.1 % (coughs) and 78.0 ± 3.8 % (speech) for TLs, using a random forest (RF) classifier. 75.1 ± 6.9 % of swallows were detected within 500 ms of onset in the controls, and 63.1 ± 6.1 % in TLs. sEMG can be used to predict critical larynx movements, although a viable ABL requires improvements. Results are particularly encouraging as they encompass a TL cohort. An ABL could alleviate many challenges faced by laryngectomies. This study represents a promising step toward realising such a device.

Index Terms—Artificial larynx, coughing, total laryngectomy, pattern recognition, speech, surface electromyography (sEMG), swallowing.

I. INTRODUCTION

An estimated 177,422 individuals worldwide were diagnosed with laryngeal cancer in 2018 [1]. In moderate to severe cases, total laryngectomy (TL) is often necessary during treatment. TL involves complete removal of the larynx, including vocal cords. The trachea is re-connected to a stoma created at the base of the neck, through which individuals respire. The mouth and nose remain connected to the oesophagus. 299 patients underwent TL in the UK in 2014 [2], and 3,414 in the USA in 2008 [3]. Laryngectomies face many critical physiological and social challenges, with impaired swallowing, coughing and speaking. Dysphagia incidence as high as 71.8 % has been reported [4], comprising problems such as increased swallow durations, avoidance of certain food types and poor bolus clearance. Coughing remains as one mechanism to clear mucus from the airway. However, the stoma cover must be removed each time a TL coughs. As the urge to cough is often sudden, the stoma cover may become clogged if not removed in time.

Speech, impaired due to removal of vocal cords, may be enabled through oesophageal speech, tracheo-oesophageal puncture (TEP) with speech prosthesis or use of an electrolarynx (EL) [5]. These methods however assist with speech only, no other functions such as swallowing and coughing. Oesophageal speech can be difficult to learn, an EL is not hands-free, and TEP requires regular maintenance and is often not hands-free either. Additional drawbacks exist from a social aspect, with reported worsening of self-image and social isolation [6]. An artificial bioengineered larynx (ABL), consisting of a tracheal prosthesis with a valve enabling respiration, has recently been developed [7]. One patient reported improved breathing, swallowing and smell, and was also able to speak quietly while the tracheostomy was closed.

An ABL would improve quality of life for laryngectomies, by enabling normal larynx functions such as swallowing and coughing and supporting more natural speech. An ABL must be safe and intuitive for the user, opening and closing correctly to generate coughing pressures and expiration flow rates following aspiration (entry of a foreign object, such as food, into the airway). An ABL must close quickly to enable swallow functionality, mitigating aspiration as a regular larynx does, as well as enable speech by vibrating as the vocal cords do when air passes through. To provide intuitive control of an ABL, this paper investigates using surface electromyography (sEMG) as a native signal to predict larynx movements.

Prediction of user intention by sEMG pattern recognition has been extensively investigated for applications including prosthetic limbs and exoskeletons [8, 9, 10], and functions associated with the larynx [11, 12]. Commonly targeted muscle groups in such studies are the suprahyoid (SH) muscles, located in the submental space underneath the chin, and infrahyoid (IH) muscles, located in the anterior neck.

Anfšt and Troster [13] used sEMG of the IH muscles to
distinguish between swallow and non-swallow events based on a signal segmentation and similarity search method. A true positive rate (TPR) of 82% was achieved; however, positive predictive value (PPV) was just 17%; swallowing was predicted almost 5 times as frequently as it occurred. Roldan-Vasco et al. [11] classified swallow phases, using sEMG from the masseter, orbicularis oris, submental, and IH muscles. Support vector machine (SVM) and artificial neural network (ANN) classifiers were compared, with the SVM resulting in superior classification of oral phase (TPR: 90 ± 6%, PPV: 90 ± 5%) and pharyngeal phase (TPR: 92 ± 4%, PPV: 94 ± 4%). Swallow detection has also been investigated with alternative methods to sEMG. Using tongue pressure on the hard palate, and a time-delayed ANN, to classify swallow events, Hadley et al. [14] achieved a TPR of 90%, and a PPV of 80%. These studies demonstrate the feasibility of detecting swallow events. Auscultation and accelerometry are further means by which swallows are commonly analysed [15], [16]. However, popular transducer locations in these studies, such as the cricoid cartilage [17], are affected by TL. We expect sEMG to be more suitable for our aims, as it permits recording of relevant signals from locations distant to the larynx.

We investigated voluntary cough detection in a prospective study [18]. Using signal amplitude thresholding of sEMG of the intercostal and diaphragm muscles, we detected 79% of coughs 100 ms in advance of exhalation. For control, thresholding is unsatisfactory, as other movements involving the intercostal and diaphragm muscles will cause a high false positive rate. A pattern recognition approach may provide more robust detection. IH muscle sEMG pattern recognition has been applied to speech related tasks such as pitch estimation [12], [19]. A silent speech recognition system for laryngectomees has also been developed, using sEMG from a cohort of muscle groups [20]. Although promising, many of these results are not applicable for laryngectomees, as the IH muscles are often removed or significantly reduced during TL.

For an ABL, a delay in swallow detection presents a risk of aspiration. An acceptable delay may be inferred from the duration of the late oral and pharyngeal phases, when the bolus is propelled toward the oropharynx as the swallow reflex is initiated, and contraction of the SH and thyrohyoid muscles results in larynx ascension and contributes to closure of the airway. Transition from pharyngeal to oesophageal phase occurs after the upper oesophageal sphincter opens, allowing the bolus to pass through [21]. Thus, if a control signal is not delivered prior to the latter stage of the pharyngeal phase, aspiration may occur. The late oral phase ranges from 308 ± 32 ms to 900 ± 300 ms, and the pharyngeal phase from 648 ± 194 ms to 1500 ± 450 ms [11], [22], [23]. Thus, a maximum delay of 500 ms may be reliable, conservative, and safe [14].

We investigated sEMG pattern recognition of swallow, cough and speech actions, within a cohort inclusive of TLs. Previous work typically focused on binary classification, aspects of a specific function, or on an alternative goal such as pitch estimation. We extend this to investigate a range of larynx functions, recording from muscle groups available in post-laryngectomy users. Accurate, timely classification of user action is necessary to inform and actuate state changes of our envisioned ABL. Validation of using sEMG to distinguish between critical larynx functions in laryngectomees is an essential precursor to an implantable EMG control system, and therefore constitutes a significant step toward ABL control.

II. METHODS

The study was approved by UCL ethics committee, project 5697/006. Fig. 1 illustrates our classification algorithm, from data collection to determination of results.

A. Data collection

1) Cohort: recruitment target of 10 participants (5 TL, 5 control). TL participants were recruited via the UK National Association of Laryngectomee Clubs, control participants were recruited internally. Eligibility criteria: over 18 years old; no neuromuscular disorder nor disease affecting voice (other than TL). All participants gave informed consent.

2) Experimental protocol: Each participant attended 3 sessions, on separate days. In each session, they carried out actions in a pseudorandomised order, during which sEMG was recorded. Recordings in each session were as follows:

- Swallowing: 15 recordings, 1 swallow per recording: 5 dry (saliva only); 5 liquid (water); 5 solid (banana).
- Coughing: 3 recordings, 5 voluntary coughs [24] per recording for a total of 15 coughs per session.
- Speaking: 3 recordings, each with 10 randomly selected phrases from the Harvard Sentences [25], read aloud.
- Everyday movements: 6 recordings of: standing, reaching overhead, twisting, walking, and sitting.

Across 30 sessions (10 participants, 3 sessions each), this protocol amounted to a total of 810 specified recordings.

3) Hardware setup (Fig. S1): The skin was cleaned with an alcohol wipe. Two submental electrodes (EL513, 10 mm diameter, BIOPAC Systems UK) were placed on the midline, posterior to the mental protuberance, with 20 mm interelectrode distance. Three electrodes (EL503, 11 mm diameter, BIOPAC) were placed on the right 9th/10th intercostal space close to the anterior axillary line, with 35 mm interelectrode distance. The posterior two electrodes formed the intercostal recording dipole. The anterior electrode and a single electrode placed on the left 9th/10th intercostal space formed the diaphragm recording dipole. Two reference electrodes (EL503) were placed on the midline over the sternum. Two wireless EMG recorders (BIOPAC BN-EMG2 BioNomadix, 2 kHz sampling rate, 2,000× gain, 5 to 500 Hz bandpass filter) were placed at the waist and on the head to minimise relative cable length and motion artefacts.

In addition to sEMG, reference measures were recorded simultaneously, to assist in identifying onset and offset of each action. High-speed video of the neck and submental region

---

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMRB.2021.3122966, IEEE Transactions on Medical Robotics and Bionics

---
through amplitude thresholding and visual assessment, and replaced with the mean of the signal after artefact removal.

Additionally, a 50 Hz notch filter (order 2) was applied to all recordings to attenuate powerline interference.

C. Data segmentation

Onset and offset times of each action were determined using a combination of amplitude thresholding of the sEMG signal envelope [26], [27], and the reference measure for each activity (swallow – video, speech – audio, cough – pneumotachometry). Different amplitude thresholds and muscles were used for each action (Table I), selected through empirical observation, and trial and error (see supplemental material for additional information).

The reference measures provided an approximation of action timings. For swallows, the high-speed videos were used. For the control group, the initial ascension of the larynx and subsequent return to base position provided the reference timings. In the TL group, movement of the bolus was identified, and the onset/offset of movement in the anterior neck in surrounding video frames used as timing reference.

Reference timings for coughs were based on peaks of the derivative air flow rate (points of greatest change) and zero-crossings in the pneumotachometry signal. Given n coughs per recording (n was not always 5, due to participants mistiming or stopping), the 2n largest peaks were located to give an initial approximation of reference onset and offset. Onsets were adjusted to the earliest zero-crossing occurring within 1 second prior, to approximate the transition from inhalation to larynx closure. Reference timings were manually marked where this method was unsuitable, due to one of the following: unusable pneumotachometry data, in which case an estimate was made using sEMG; peaks in the derivative airflow not associated with a cough (e.g from a sharp breath); multiple peaks detected for a single onset or offset, due to a particularly strong cough, or weak cough elsewhere in the signal.

Reference timings for speech were determined by applying the MATLAB Voice Activity Detector to the sound recordings to identify speech intervals.

Once reference timings were set, for each recording the final onset/offset of each action were determined as follows:

- Signal envelope extracted via moving average filter, applied over 40 ms to full wave rectified sEMG signal.
- Amplitude threshold calculated from signal envelope, according to Table I.

### TABLE I: sEMG CHANNEL AND THRESHOLD FOR EACH ACTION.

<table>
<thead>
<tr>
<th>Action</th>
<th>sEMG Channel</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swallow</td>
<td>Submental</td>
<td>$\mu + 0.5\sigma$</td>
</tr>
<tr>
<td>Cough</td>
<td>Intercostal¹, diaphragm¹</td>
<td>$\mu + \sigma$</td>
</tr>
<tr>
<td>Speech</td>
<td>Submental</td>
<td>$\mu + 0.25\sigma$</td>
</tr>
</tbody>
</table>

$\mu$ is the mean of the rectified signal envelope, and $\sigma$ its standard deviation (derived from entire envelope, not noise-level activity only).¹Digitally high-pass filtered with cut-off at 50 Hz, using a 20th-order Chebyshev type-1 filter, to attenuate ECG interference. (Note: the filter was only for class labelling. During the feature extraction phase this high-pass filter was not applied).
• sEMG segments crossing the threshold were marked, corresponding to the “Active segments” in Fig. 2.
• These segments were compared with the corresponding reference timings for the recording.
• If a segment coincided with the reference, the first/last threshold crossings were taken as action onset/offset.
• If multiple segments coincided with the same reference (due to fluctuations around the threshold) the first crossing of the first segment and last crossing of the last segment were used as onset and offset.
• If a segment of sEMG activity did not coincide with the marked reference timings, it was not considered as belonging to one of the specified actions (swallow, cough, speech) and therefore, disregarded.
• For coughs, onset/offset was taken as the earliest/latest of the threshold crossings between the intercostal and diaphragm channels.

Fig. 2 displays an example of this procedure for a swallow recording (Fig. S2 – cough, Fig. S3 – speech).

Processed sEMG recordings were divided into 128 ms frames, with 64 ms overlap. Each frame between action onset and offset was assigned that action’s class label (swallow, cough, or speech). Transition frames (those involving a change in action) were assigned the class label of the majority of the data in the frame. Null frames encapsulated all activity (primarily baseline sEMG between actions, and movement activity) outside swallowing, coughing or speech, resulting in a total of 4 classes: swallow; cough; speech; and null.

D. Feature extraction

15 features, listed with equations in supplemental Table S1, were extracted from the time domain (TD), frequency domain (FD), and time-frequency domain (TFD) for each of the 3 channels, resulting in a 45-fold feature vector characterising each frame. TD features were: mean absolute value (MAV); Teager-Kaiser operator (TKO) [28]; zero-crossing rate (ZCR); slope-sign change (SSC); Willison amplitude (WAMP); and waveform length (WL). FD features were: mean frequency (MNF); median frequency (MDF); modified mean frequency (MMNF); and modified median frequency (MMDF). The multiscale variance of wavelet coefficients (WVAR) was extracted as a TFD feature: the signal was decomposed to the 4th level using the “db4” wavelet filter with the maximal overlap discrete wavelet transform algorithm [29]. The variance of the resulting 4 sets of detail coefficients and 1 set of approximation coefficients returned a 5-fold feature vector.

These features were selected due to either their high efficacy in sEMG prediction studies [12], [30] or their ability to distinguish between the target class and ECG artefacts (present in the intercostal and diaphragm channels) [31].

The ZCR, WAMP, and SSC features use a threshold to reduce the impact of noise-level fluctuations in the signal, often set between 10 - 50 mV [30]. However, Kamavuako et al. [32] reported that low thresholds may be advantageous when ZCR and SSC are part of an ensemble of features. A threshold of zero offered a reasonable trade-off between performance and generalisation across sessions and subjects. Thus, we chose a low, but non-zero, threshold of 5 µV.

E. Model development

The dataset was first divided into a TL group and a control group, with separate models developed for each group. A randomised, 5-fold cross-validation (CV) split was applied to each group for model training and testing, in a stratified (maintaining approximately equal class proportions) manner. Partitioning was recording-based rather than frame-based, to preserve the sequential nature of the data for post-processing and further analysis. A leave-one-participant-out model was also developed for each group (TL/control), the results of which are included as supplemental material.

During each iteration of CV, features in the training set were scaled to the range [-1, 1]. The parameters used for this transformation were applied to the test set.

Hyperparameter selection used nested, grid-search CV [33]. During training, a nested CV loop was established by dividing the training set into 5 further subsets. Within each subset, the null class was undersampled to the size of the speech class, to reduce processing time. Classifiers were trained with each combination of their hyperparameters. The combination with highest average F1-score across the 5 subsets of the nested loop was selected for classifier training in the outer loop.

The classifiers investigated in this study were an ANN, random forest (RF), SVM with radial basis function (RBF) kernel, and linear discriminant analysis (LDA). Table II shows the hyperparameter ranges included in the grid-search. The ANN was trained using the scaled conjugate gradient backpropagation method [34], and contained one hidden layer. The SVM was implemented using the LIBSVM package [35].

F. Post-processing

To reduce spurious predictions arising from isolated misclassifications, a sequential smoothing method was applied to the output of the classifier. A class change required two successive frames of the same class to be predicted.

G. Measurement of event-based performance

Predictive performance was measured in relation to events and the individual frames comprising them. Event-based performance is presented in terms of TPR, PPV and F1-score.
**TABLE II**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Hyperparameter</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>nNeurons</td>
<td>10 log-spaced values from 10 to 10^3</td>
</tr>
<tr>
<td></td>
<td>Transfer func.</td>
<td>Tanh, sigmoid, ReLU</td>
</tr>
<tr>
<td>RF</td>
<td>nTrees</td>
<td>50, 100, 200, 300, 500</td>
</tr>
<tr>
<td></td>
<td>mTry</td>
<td>( \sqrt{\text{NF}} ), 10:10:70% * NF</td>
</tr>
<tr>
<td>SVM</td>
<td>c</td>
<td>10 log-spaced values from 10^3 to 10^2</td>
</tr>
<tr>
<td></td>
<td>g</td>
<td>10 log-spaced values from 10^3 to 10</td>
</tr>
<tr>
<td>LDA</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

\( nNeurons \) is the number of hidden layer neurons, \( nTrees \) the number of decision trees comprising the RF, \( mTry \) the number of features evaluated at each node, \( NF \) the number of features, \( c \) is the penalty parameter and \( g \) is the gamma parameter in the RBF kernel function.

F1-score is a measure of the harmonic mean of TPR and PPV:

\[
F1 = 2 \times \frac{\text{PPV} \times \text{TPR}}{\text{PPV} + \text{TPR}}
\]

An event is comprised of \( N \) successive frames of a particular class. True positives (TP), false positives (FP) and false negatives (FN) for an event were defined as follows:

- **TP**: At least 1 of \( N \) frames comprising an event correctly classified (at least 2 when the post-processing strategy is implemented).
- **FP**: Incorrectly predicted frame(s). Each sequence of FP frames of the same class is counted as 1 FP.
- **FN**: 0 of \( N \) frames comprising an event correctly classified.

Event-based results are calculated in this way to assess class performance individually. It is feasible for one event comprising multiple frames to be marked as both a TP with respect to one class, and a FP with respect to another.

\section*{H. Measurement of frame-based performance}

Frame-based performance is presented in the form of a confusion matrix, with TPR and PPV also given.

\subsection*{I. Swallow detection delay}

In addition to predictive performance, swallow detection delay was calculated for positively identified swallows. This was measured as the difference between true onset (Section II.C) and predicted onset following post-processing (i.e. after a second successive swallow frame has been detected).

\section*{III. RESULTS}

\subsection*{A. Data collection}

Demographics for the 10 participants are provided in Table III. All TLs had a TEP prosthesis. Fig. 3 shows sEMG from a TL participant (ECG has been filtered for illustration purposes - see Fig. S4 for unfiltered version). 818 sEMG recordings were captured, more than the amount specified by the protocol (810 total), as in some cases a recording had to be paused and partially repeated. Table S2 displays the amount of data contributed, in number of frames, per participant.

\subsection*{B. sEMG pre-processing}

142/818 sEMG recordings across 13 sessions were contaminated with 50 Hz noise and harmonics and digitally filtered as outlined in Section II.B.

\subsection*{C. Data segmentation}

32/95 cough recordings were marked manually: 8/32 due to unusable pneumotachometry data, and 24/32 due to detected peaks in the derivative airflow requiring manual correction.

\subsection*{D. Event-based performance}

Fig. 4 summarises the results of the event-based analysis with post-processing strategy for the control and TL group (see Table S3 for tabular version of results). The RF classifier achieved greatest event-based performance, as measured by F1-score. Results were also calculated using a leave-one-participant-out CV method (Table S4).

\subsection*{E. Frame-based performance}

Given the RF classifier achieved greatest event-based performance, (Section III.D), results in this section and III.F are based on this strategy. Average frame-based results for each class, across the 5 folds, are outlined in Table IV for the control group and Table V for the TL group.

\subsection*{F. Swallow detection delay}

Detection delays for the swallow events are shown in Fig. 5. Delays are plotted cumulatively over intervals of 64 ms, the frame rate used in this study, and are presented as a proportion of all swallows (not as a proportion of positively-identified swallows only). Missed swallows consist of both undetected swallows and those with a detection delay exceeding 500 ms (this analysis does not take processing time into account). In the control group 75.1 ± 6.9 % of swallows were detected within 500 ms, 63.1 ± 6.1 % in the TL group.

\section*{IV. DISCUSSION}

Our study is the first investigation of sEMG for predicting larynx movements to include laryngectomees, specifically accounting for key anatomical differences, in a multiclass
classification problem of this nature. The discussion relates to the RF classifier (best performance in control and TL groups).

Speech F1-scores were 70.5 ± 5.4 % (control) vs 78.0 ± 3.8 % (TL). TL participants had high sEMG across all channels during speech. Activity was concentrated in the submental channel in the controls, with only some using intercostal and diaphragm muscles. This inconsistency likely contributed to inferior speech performance as the model is trained on mixed participant data. The heavier breathing (increased intercostal and diaphragm activity) during speech in the TL group is to maintain pressure through a TEP speech prosthesis [36], [37].

For swallowing, aspiration rate is a key consideration due to the risks of respiratory conditions that may develop following incomplete airway clearance. A low, but non-zero, rate is, however, acceptable, with 3 % reported in older (69-85 years), non-laryngectomy populations [38]. Although not directly analogous to swallow detection delay performance, aspiration rate provides a benchmark for an ABL. The rate of undetected swallows is 10.5 ± 4.0 % (control) and 28.3 ± 7.4 % (TL). Using 500 ms as the longest acceptable delay (Fig. 5), missed swallow rate is 24.9 ± 6.9 % (control) and 36.9 ± 6.1 % (TL). A model tailored to a specific participant may improve performance, reducing missed swallow rate [14], [39].

During a cough, sEMG increases after larynx closure (end of inhalation), making advanced prediction of a natural cough challenging. An ABL should include high-level user control to close the ABL to build-up pressure in the lungs, in preparation for a cough. The detection demonstrated in this study would trigger the ABL opening. Alternatively, the trigger could be a preset, user-specific, pressure [24]. Accurate cough detection is also important to reduce false positive swallows and speech.

As we present a novel sEMG classification problem, comparisons with other studies are limited. TPR and PPV for swallows in the control group compare favourably to those of Amft and Troster [13]. TPR is improved from 82 % in Amft and Troster to 89.5 ± 4.0 %, and PPV from 17 % to 66.6 ± 7.9 %.

For swallowing, aspiration rate is a key consideration due to the risks of respiratory conditions that may develop following incomplete airway clearance. A low, but non-zero, rate is, however, acceptable, with 3 % reported in older (69-85 years), non-laryngectomy populations [38]. Although not directly analogous to swallow detection delay performance, aspiration rate provides a benchmark for an ABL. The rate of undetected swallows is 10.5 ± 4.0 % (control) and 28.3 ± 7.4 % (TL). Using 500 ms as the longest acceptable delay (Fig. 5), missed swallow rate is 24.9 ± 6.9 % (control) and 36.9 ± 6.1 % (TL). A model tailored to a specific participant may improve performance, reducing missed swallow rate [14], [39].

During a cough, sEMG increases after larynx closure (end of inhalation), making advanced prediction of a natural cough challenging. An ABL should include high-level user control to close the ABL to build-up pressure in the lungs, in preparation for a cough. The detection demonstrated in this study would trigger the ABL opening. Alternatively, the trigger could be a preset, user-specific, pressure [24]. Accurate cough detection is also important to reduce false positive swallows and speech.

As we present a novel sEMG classification problem, comparisons with other studies are limited. TPR and PPV for swallows in the control group compare favourably to those of Amft and Troster [13]. TPR is improved from 82 % in Amft and Troster to 89.5 ± 4.0 %, and PPV from 17 % to 66.6 ± 7.9 %.

Our TPR is equal to Hadley et al. (90 %), but their PPV is larger (80 %) [14]. Speech was a common cause of false swallows in Hadley et al., as in our study (Tables IV and V).

Roldan-Vasco et al. [11] achieved TPRs of 90 ± 6 % and 92 ± 4 % for oral and pharyngeal phases of swallowing, with PPV of 90 ± 5 %, and 94 ± 4 % respectively. This is superior to our control group frame-based results for swallows: TPR (57 ± 3 %); PPV (77 ± 8 %). This may be due to various reasons beyond the selection of features or classifiers themselves. A lesser number of sEMG channels were used in our study (3 vs 8). The range of actions included in our study confounded swallow prediction, as evidenced by 45/593 swallow frames misclassified as speech, and 208/593 as null (Table IV). Additionally, Roldan-Vasco et al. discarded transition frames from both training and testing, whereas we assigned them to the class of the majority of the frame’s data.

Using a participant-dependent 10-fold CV approach with

### TABLE IV

<table>
<thead>
<tr>
<th>Predicted Class</th>
<th>Swallow</th>
<th>Cough</th>
<th>Speech</th>
<th>TPR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Null</strong> (n=18152)</td>
<td>17631 (438)</td>
<td>208 (46)</td>
<td>41 (14)</td>
<td>392 (172)</td>
</tr>
<tr>
<td><strong>Swallow</strong> (n=593)</td>
<td>208 (33)</td>
<td>340 (23)</td>
<td>0</td>
<td>45 (18)</td>
</tr>
<tr>
<td><strong>Cough</strong> (n=365)</td>
<td>125 (30)</td>
<td>0</td>
<td>239 (27)</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Speech</strong> (n=2492)</td>
<td>1457 (104)</td>
<td>20 (7)</td>
<td>0</td>
<td>1015 (205)</td>
</tr>
</tbody>
</table>

**TABLE V**

<table>
<thead>
<tr>
<th>Predicted Class</th>
<th>Swallow</th>
<th>Cough</th>
<th>Speech</th>
<th>TPR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Null</strong> (n=20278)</td>
<td>19634 (757)</td>
<td>106 (34)</td>
<td>33 (13)</td>
<td>505 (70)</td>
</tr>
<tr>
<td><strong>Swallow</strong> (n=754)</td>
<td>390 (24)</td>
<td>336 (56)</td>
<td>0</td>
<td>28 (9)</td>
</tr>
<tr>
<td><strong>Cough</strong> (n=444)</td>
<td>238 (69)</td>
<td>0</td>
<td>175 (44)</td>
<td>31 (20)</td>
</tr>
<tr>
<td><strong>Speech</strong> (n=3186)</td>
<td>894 (155)</td>
<td>4 (4)</td>
<td>10 (11)</td>
<td>2278 (311)</td>
</tr>
</tbody>
</table>

**PPV (%)**

| Null | 91 (0.5) | 77 (8) | 86 (4) | 70 (7) |
| Cough | 93 (1) | 75 (8) | 80 (9) | 80 (2) |

Results presented are the mean frame results across 5 folds, with standard deviation in parentheses.
accelerometer data, Li et al. [40] reported F1-scores of 94 \% for coughing and 93 \% for speech. Our results are similar for cough, 93.8 \pm 2.8 \% (control, event-based) and lower for speech, 70.5 \pm 5.4 \%. However, the segment size in Li et al., 2.56 s, would be too long for ABL control as the actions in our study are typically shorter than this segment size.

There is a discrepancy between frame-based and event-based swallow, cough and speech results in our study. PPV is greater than TPR for the frame-based results (Tables IV and V), but smaller for event-based results (Fig. 4). This is in part due to our event scoring criteria. Positive event detection (true and false) required only two successive frames (two due to post-processing method) of the same class to be detected, incurring higher TPR and lower PPV. In a functional ABL, operated using sEMG, frames comprising an event may be unequal in importance given the need for early detection of swallows for example. Additionally, the risk associated with missing a swallow event is greater than that imposed by over-prediction. A system which tends towards positive event detection, further informed by latency analysis, is desirable. Hence the design of our event scoring criteria.

The discrepancy between frame-based and event-based results may also be a consequence of the semi-automatic method of determining event onset/offset. Manual labelling, as in Roldan-Vasco et al. [11], is cumbersome for large datasets and automation reduces researcher bias. Although our reference-threshold method mitigated imprecision, there were still observed instances of imprecise event onset/offset. Furthermore, merging multiple segments crossing the amplitude threshold, while overlapping with the same reference segment, resulted in classing some near-baseline sEMG frames as swallow, cough or speech (Fig. S3 illustrates this issue). Such frames were likely detrimental in training and testing the algorithm, with frames frequently misclassified as the null class, although this is also attributable to class-imbalance. Our main goal was to detect the event overall; thus, classing these frames as null was undesirable, as it would split one event into multiple, artificially raising the TPR for events. A solution may be to use two separate sets of timings, one for sEMG frames with elevated activity, and one for event onset/offset. A model could be trained and evaluated in terms of frames on the former, and event detection on the latter.

We expect an EMG-controlled ABL would require target user data to be present in the training set. This training set may either consist of data from multiple TLs, inclusive of the target user, or of data exclusive to the target user. The substantial performance gap between Table S3 and Table S4 indicates a leave-one-participant-out approach is not viable (at least with respect to our model design and number of participants).

Our study has limitations. There is a small sample size (5 controls, 5 TLs). This has particular ramifications for the leave-one-participant-out models (Table S4). A larger sample size may return a model less sensitive to differences between individuals. There are more actions involving the larynx and selected muscles than those we examined, such as yawning or sneezing. Involuntary coughs may also yield patterns distinct from the voluntary coughs in our study. We consider breathing to be covered by the null set, and did not explicitly aim to identify this action. However, as breaths range in intensity, a deep breath could exhibit high activity in the intercostal and diaphragm channels which may confound predictions. Control group participants were familiar with sEMG practices which may have contributed to cleaner sEMG recordings. Analysis of swallow onset delays did not account for processing time as this work was carried out in MATLAB which would not be used for real-world implementation of an ABL. The maximum acceptable delay will be reduced by the duration required to read-in EMG data and produce the output class. Additionally, the time taken to actuate the mechanical aspect of the ABL would need to be factored in for real-world implementation. Reference onset/offset times for swallowing were determined by a single assessor. In an experiment with the control group, we shifted video times randomly, in an approximately uniform manner, by either -200, -100, 0, 100 or 200 ms. Frame-based and event-based TPR and PPV differed by < 1 \% from the original results. Thus, unless disagreement between assessors over video timings is large, the difference in overall outcome is negligible. Further limitations are that our model consists of same-session data present in training and testing sets, we have not accounted for variation in sEMG across days, and that we did not normalise data across sessions or participants, due to the lack of a specific baseline signal to normalise to.

Future work will address some of these limitations and focus on a real-time system and participant-specific model, including analysis of performance on unseen days. Deep learning methods and more sophisticated neural networks accounting for the time-series nature of sEMG signals may improve predictions of larynx function. Although the current algorithm was designed for an ABL control system, it may be beneficial to other areas such as dysphagia monitoring.
V. CONCLUSION

eEMG was used to predict the larynx functions of swallow, cough and speech, through a pattern recognition approach. This was achieved for a control group and a laryngectomy group, specifically accounting for muscles excised during total laryngectomy. Improvements are needed however, particularly in relation to how quickly a swallow can be detected. Additionally, it would be useful to investigate a wider range of larynx functions as some movements, outside the scope of this study, may share similar muscle activity to swallows, coughs, or speech. This work presents the possibility of improved treatment for those requiring total laryngectomy, using eEMG as a control signal for an artificial, bioengineered larynx.

ACKNOWLEDGMENT

We thank: the participants for assisting in this study and the National Association of Laryngectomee Clubs; the Wellcome Trust for funding this work [Grant 106574/Z/14/Z]; the Restoration of Appearances and Function Trust (RAFT, UK) and the Therapeutic Acceleration Support (TAS) Fund for additional funding of K. de Jager; and K. Yue for assistance with the pneumotachometry kit used in data collection.

REFERENCES