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 laryngectomees. 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 respirate. 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]. Laryngectomees 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, trachea-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 laryngectomees, 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.

Amft and Troster [13] used sEMG of the IH muscles to

This manuscript was originally submitted for review on Nov 23, 2020; revised May 17 2021 and Oct 04 2021; accepted on Oct 20, 2021. This work was supported by The Wellcome Trust [Grant 106574/Z/14/Z].

J. McNulty*, K. de Jager and A. Vanhoestenberghe are with the Division of Surgery and Interventional Science, University College London, London, UK (*corresponding author: email: j.mcnulty@ucl.ac.uk). H.T. Lancashire is with the Department of Medical Physics and Biomedical Engineering, UCL. J. Graveston was with the UCL Ear Institute, UCL, M. Birchall is with the UCL Ear Institute, Division of Brain Sciences and Royal National Nose and Throat and Eastman Dental Hospitals, UCL.

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

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 accelerometery 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 $9^{th}/10^{th}$ 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 $9^{th}/10^{th}$ 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

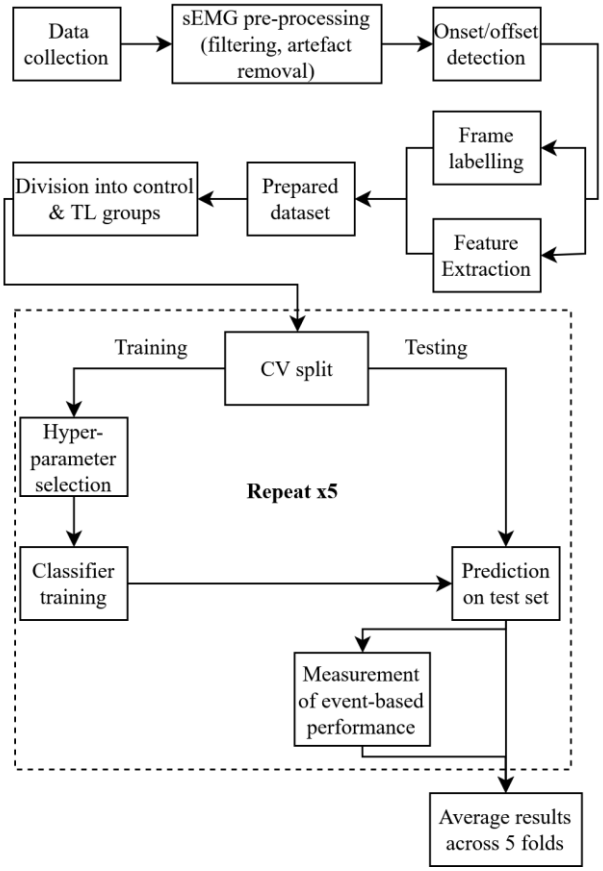

Fig. 1. Development process of algorithm for prediction of larynx function. CV = cross-validation. CV was 5-fold meaning a train/test ratio of 80:20 per CV iteration. The prepared dataset was split into two groups (controls and TLs). A separate model was developed for each group, with both models following the procedure from 'CV Split' onward.

was recorded during swallowing, using a laterally positioned Photron FastCam SA1.1 camera (500 fps, 512×512 pixels). Pneumotachometry was recorded during coughing (2 kHz, BIOPAC DA100C pressure transducer, TSD160A linear pneumotachometer). Sound was recorded during speech (2 kHz, RSPro Unidirectional Electret Condenser Microphone). No reference was recorded for movement actions as the exact timing was not of interest. Movement recordings were considered confounding measurements to impact the efficacy of the classification algorithm as they would in practical use.

B. sEMG pre-processing

All data processing and analysis was carried out using MATLAB 2018a (The MathWorks Inc., USA).

The recordings were visually assessed in time, frequency, and time-frequency domains for 50 Hz powerline interference and harmonics. A digital comb filter (order 40) with notches at the 50 Hz harmonics was applied to contaminated recordings to preserve for use. The recordings were also inspected for artefacts due to the wireless EMG modules temporarily losing connection with the data acquisition unit (an issue related to our purchased wireless sensors). These were characterised by exceptionally large spikes in signal amplitude followed by approximately 1 second of absent data. They were detected 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 miscounting 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.

Action	sEMG Channel	Threshold
Swallow	Submental	$\mu + 0.5\sigma$
Cough	Intercostal†, diaphragm†	$\mu + \sigma$
Speech	Submental	$\mu + 0.25\sigma$

μ is the mean of the rectified signal envelope, and σ its standard deviation (derived from entire envelope, not noise-level activity only).

†Digitally 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).

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

- 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 4^{th} 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.

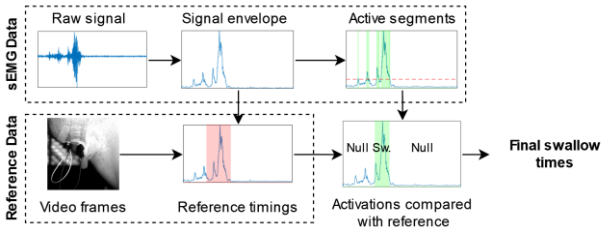

Fig. 2. An example of the data segmentation and labelling process for a swallow (denoted as Sw. in lower right plot). The reference-threshold combination reduced spurious activations, while more precisely identifying the location of action onset and offset than if only one of threshold detection or reference timings was used.

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.

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

TABLE II
LIST OF CLASSIFIERS AND ASSOCIATED HYPERPARAMETERS

Classifier	Hyperparameter	Range
ANN	nNeurons	10 log-spaced values from 10 to 10^3
	Transfer func.	Tanh, sigmoid, ReLU
RF	nTrees	50, 100, 200, 300, 500
	mTry	$\sqrt{NF}$, 10:10:70% * NF
SVM	c	10 log-spaced values from 10^{-3} to 10^2
	g	10 log-spaced values from 10^{-4} to 10
LDA	-	-

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:

$$F1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} \quad (1)$$

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.

H. Measurement of frame-based performance

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

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).

III. Results

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.

TABLE III
DEMOGRAPHICS OF STUDY PARTICIPANTS

	Control (n = 5)	TL (n = 5)
Age (mean ± SD)	32 ± 5.1	69.6 ± 9.8
Female, Male	2, 3	1, 4

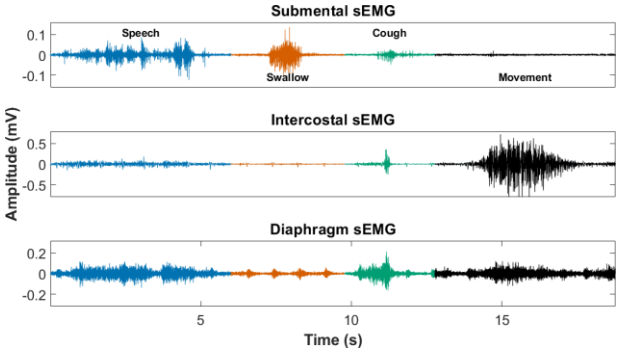

Fig. 3. Example of multichannel sEMG signals. Active sections from recordings for each action have been extracted and concatenated: speech (blue), swallow (orange), cough (green) and movement (black).

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.

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.

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).

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.

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.

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

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

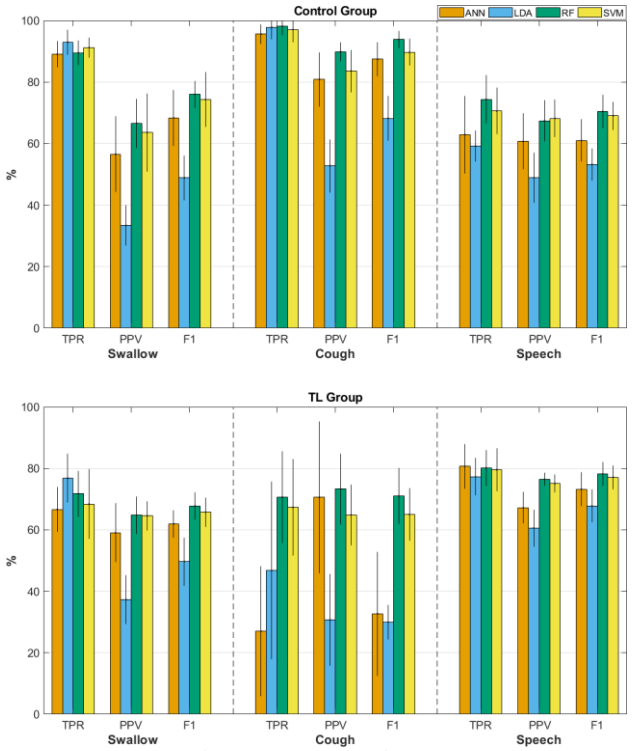

Fig. 4. Overview of event-based results for both groups.

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 %. As swallow detection preceded bolus classification in their work, overestimation was likely preferable to underestimation.

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
FRAME-BASED RESULTS FOR THE CONTROL GROUP

		Predicted Class				
		Null	Swallow	Cough	Speech	TPR (%)
True Class	Null (n=18152)	17631 (438)	208 (46)	41 (14)	392 (172)	97 (1)
	Swallow (n=593)	208 (33)	340 (23)	0	45 (18)	57 (3)
	Cough (n=365)	125 (30)	0	239 (27)	1 (1)	66 (6)
	Speech (n=2492)	1457 (104)	20 (7)	0	1015 (205)	40 (5)
	PPV (%)	91 (0.5)	77 (8)	86 (4)	70 (7)	

Results presented are the mean frame results across 5 folds, with standard deviation in parentheses.

TABLE V
FRAME-BASED RESULTS FOR THE TL GROUP

		Predicted Class				
		Null	Swallow	Cough	Speech	TPR (%)
True Class	Null (n=20278)	19634 (757)	106 (34)	33 (13)	505 (70)	97 (0.4)
	Swallow (n=754)	390 (24)	336 (56)	0	28 (9)	44 (3)
	Cough (n=444)	238 (69)	0	175 (44)	31 (20)	40 (8)
	Speech (n=3186)	894 (155)	4 (4)	10 (11)	2278 (311)	71 (6)
	PPV (%)	93 (1)	75 (8)	80 (9)	80 (2)	

Results presented are the mean frame results across 5 folds, with standard deviation in parentheses.

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

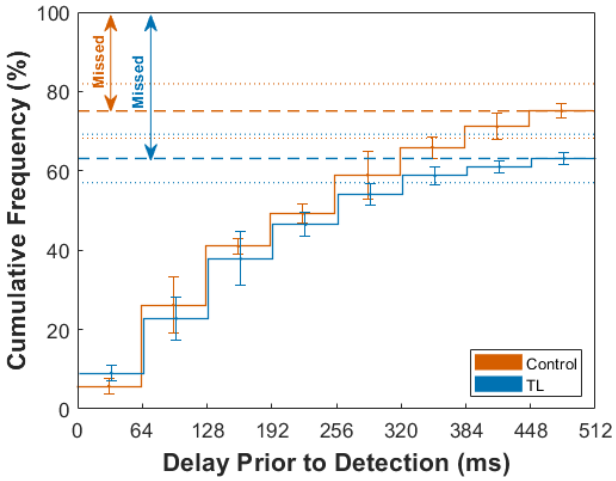

Fig. 5. Cumulative swallow detection delay. Control group (orange): 75.1 ± 6.9 % detected within 500 ms, 63.1 ± 6.1 % in the TL group (blue). The number of missed swallows (undetected swallows plus those detected after 500 ms) across 5 CV folds is indicated by the horizontal lines, dashed = average, dotted = standard deviation.

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 ± 2.8 % (control, event-based) and lower for speech, 70.5 ± 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

sEMG 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 sEMG 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

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018.

[2] NHS, "National Head and Neck Cancer Audit 2014, DAHNO Tenth Annual Report," 2015.

[3] P. T. Maddox and L. Davies, "Trends in total laryngectomy in the era of organ preservation: A population-based study," *Otol. Head Neck Surg*, vol. 147, no. 1, pp. 85–90, 2012.

[4] J. Maclean, S. Cotton, and A. Perry, "Post-Laryngectomy: It's Hard to Swallow," *Dysphagia*, vol. 24, no. 2, pp. 172–179, 2009.

[5] C. G. Tang and C. F. Sinclair, "Voice Restoration After Total Laryngectomy.," *Otol. Clin. North Am.*, vol. 48 (4), pp. 687–702, 2015.

[6] P. Boscolo-Rizzo, F. Maronato, C. Marchiori, A. Gava, and M. C. Da Mosto, "Long-term quality of life after total laryngectomy and postoperative radiotherapy versus concurrent chemoradiotherapy for laryngeal preservation," *Laryngoscope*, vol. 118, no. 2, pp. 300–306, 2008.

[7] C. Debry, A. Dupret-Bories, N. E. Vrana, P. Hemar, P. Lavalle, and P. Schultz, "Laryngeal replacement with an artificial larynx after total laryngectomy: the possibility of restoring larynx functionality in the future.," *Head Neck*, vol. 36, no. 11, pp. 1669–1673, Nov. 2014.

[8] K. Englehart and B. Hudgins, "A Robust, Real-Time Control Scheme for Multifunction Myoelectric Control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, 2003.

[9] P. Shenoy, K. J. Miller, B. Crawford, and R. P. N. Rao, "Online Electromyographic Control of a Robotic Prosthesis," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1128–1135, 2008.

[10] H. He and K. Kiguchi, "A study on EMG-based control of exoskeleton robots for human lower-limb motion assist," *Proc. ITAB*, pp. 292–295, 2007.

[11] 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," *J. Electromyogr. Kinesiol.*, vol. 43, pp. 193–200, 2018.

[12] F. Ahmadi, M. Araújo Ribeiro, and M. Halaki, "Surface electromyography of neck strap muscles for estimating the intended pitch of a bionic voice source," in *2014 IEEE Conf. BioCAS Proc.*, 2014, pp. 37–40.

[13] O. Amft and G. Troster, "Methods for Detection and Classification of Normal Swallowing from Muscle Activation and Sound," in *Pervasive Health Conference and Workshops*, 2006, pp. 1–10.

[14] A. J. Hadley, K. R. Krival, A. L. Ridgel, E. C. Hahn, and D. J. Tyler, "Neural Network Pattern Recognition of Lingual–Palatal Pressure for Automated Detection of Swallow," *Dysphagia*, vol. 30(2), pp. 176–187, 2015.

[15] Y. Khalifa, J. L. Coyle, and E. Sejdić, "Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings," *Sci. Rep.*, vol. 10, no. 1, p. 8704, 2020.

[16] J. M. Dudik, J. L. Coyle, and E. Sejdić, "Dysphagia Screening: Contributions of Cervical Auscultation Signals and Modern Signal-Processing Techniques," *IEEE Trans. Hum-Mach. Syst.*, vol. 45 (4), pp. 465–477, 2015.

[17] K. Takahashi, M. E. Groher, and K. Michi, "Methodology for detecting swallowing sounds.," *Dysphagia*, vol. 9, no. 1, pp. 54–62, 1994.

[18] M. S. Banus, M. A. Birchall, and J. A. Graveston, "Developing control algorithms of a voluntary cough for an artificial bioengineered larynx using surface electromyography of chest muscles: A prospective cohort study," *Clin. Otolaryngol.*, vol. 43, no. 2, pp. 562–566, 2018.

[19] W. De Armas, K. A. Mamun, and T. Chau, "Vocal frequency estimation and voicing state prediction with surface EMG pattern recognition," *Speech Commun.*, vol. 63–64, pp. 15–26, 2014.

[20] G. S. Meltzner, J. T. Heaton, Y. Deng, G. De Luca, S. H. Roy, and J. C. Kline, "Silent Speech Recognition as an Alternative Communication Device for Persons with Laryngectomy," vol. 25, no. 12, pp. 2386–2398, 2017.

[21] K. Matsuo and J. B. Palmer, "Anatomy and physiology of feeding and swallowing: normal and abnormal," *Phys. Med. Rehabil. Clin. N. Am.*, vol. 19, no. 4, pp. 691–697, Nov. 2008.

[22] C. Ertekin, N. Kiylioglu, S. Tarlaci, A. B. Turman, Y. Secil, and I. Aydogdu, "Voluntary and reflex influences on the initiation of swallowing reflex in man," *Dysphagia*, vol. 16, no. 1, pp. 40–47, 2001.

[23] M. Vaiman, E. Eviatar, and S. Segal, "Surface electromyographic studies of swallowing in normal subjects: a review of 440 adults. Report 1. Quantitative data: timing measures.," *Otol. Head Neck Surg.*, vol. 131, no. 4, pp. 548–555, Oct. 2004.

[24] K. Yue *et al.*, "CoughAid - an Assistive Device for Cough Insufficiency," in *BioMedEng19*, 2019.

[25] "IEEE Recommended Practice for Speech Quality Measurements," *IEEE Trans. Audio Electroacoust.*, vol. 17, no. 3, pp. 225–246, 1969.

[26] S. Solnik, P. Rider, K. Steinweg, P. Devita, and T. Hortobágyi, "Teager-Kaiser energy operator signal conditioning improves EMG onset detection," *Eur. J. Appl. Physiol.*, vol. 110, no. 3, pp. 489–498, 2010.

[27] P. Hodges, "A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography," *Electroencephalogr. Clin. Neurophysiol.*, vol. 101, no. 6, pp. 511–519, 2002.

[28] J. F. Kaiser, "Some useful properties of Teager's energy operators," *IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 3, pp. 149–152, 1993.

[29] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press, 2000.

[30] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A Novel Feature Extraction for Robust EMG Pattern Recognition," vol. 1 (1), pp. 71–80, 2009.

[31] H. T. Lancashire, K. de Jager, J. Graveston, and A. Vanhoestenberghe, "ECG reduction in thoracic EMG during coughing with the Teager-Kaiser Operator," in *BioMedEng19*, 2019.

[32] E. N. Kamavuako, E. J. Scheme, and K. B. Englehart, "Determination of optimum threshold values for EMG time domain features; A multi-dataset investigation," *J. Neural Eng.*, vol. 13, no. 4, pp. 1–10, 2016.

[33] T. Fearn, "Double Cross-Validation," *NIR news*, vol. 21, no. 5, pp. 14–15, Aug. 2010.

[34] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993.

[35] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines.," *ACM Trans. Intell. Syst. Tech.*, vol. 2 (3), pp. 27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[36] T. A. Bohnenkamp, K. M. Forrest, B. K. Klaben, and J. M. Stager, "Lung volumes used during speech breathing in tracheoesophageal speakers," *Ann. Otol. Rhinol. Laryngol.*, vol. 120, no. 8, pp. 550–558, 2011.

[37] E. C. Ward, P. Hartwig, J. Scott, M. Trickey, L. Cahill, and K. Hancock, "Speech Breathing Patterns During Tracheoesophageal Speech," *Asia Pacific J. Speech, Lang. Hear.*, vol. 10, no. 1, pp. 33–42, 2007.

[38] S. G. Butler, A. Stuart, L. D. Case, C. Rees, M. Vitolins, and S. B. Kritchevsky, "Effects of liquid type, delivery method, and bolus volume on penetration-aspiration scores in healthy older adults during flexible endoscopic evaluation of swallowing.," *Ann. Otol. Rhinol. Laryngol.*, vol. 120, no. 5, pp. 288–295, May 2011.

[39] C. Castellini, A. E. Fiorilla, and G. Sandini, "Multi-subject/daily-life activity EMG-based control of mechanical hands," *J. Neuroeng. Rehabil.*, vol. 6, no. 1, pp. 1–11, 2009.

[40] Y. L. Cheng, Y. C. Chen, W. J. Chen, P. Huang, and H. H. Chu, "Sensor-embedded teeth for oral activity recognition," *Proc. 2013 ACM Int. Symp. Wearable Comput.*, pp. 41–44, 2013.