Multisensor Classification System for Triggering FES in Order to Support Voluntary Swallowing

Benjamin Riebold, Holger Nahrstaedt, Corinna Schultheiss, Rainer O. Seidl and Thomas Schauer

Abstract—In dysphagia the ability of elevating the larynx and hyoid is usually impaired. Electromyography (EMG) and Bioimpedance (BI) measurements at the neck can be used to trigger functional electrical stimulation (FES) of swallowing related muscles. Nahrstaedt et al. [1] introduced an algorithm to trigger the stimulation in phase with the voluntary swallowing to improve the airway closure and elevation speed of the larynx and hyoid. However, due to non-swallow related movements like speaking, chewingor headturning, stimulations might be unintentionally triggered. So far a switch was used to enable the BI/EMG-triggering of FES when the subject was ready to swallow, which is inconvenient for practical use. In this contribution, a range image camera system is introduced to obtain data of head, mouth, and jaw movements. This data is used to apply a second classification step to reduce the number of false stimulations. In experiments with healthy subjects, the amount of potential false stimulations could be reduced by 47% while 83% of swallowing intentions would have been correctly supported by FES.

I. INTRODUCTION

The EMG and BI signals measured at the larynx show specific patterns during swallowing [2]. The intention to swallow results in an increased activity of the EMG. As soon as the larynx starts to elevate a decent in the bioimpedance can be observed. These patterns can be used to trigger FES in phase to the swallowing intention [1].

FES is applied over two surface stimulation electrodes which are submentally attached. Stimulation pulses with a frequency of 30 Hz, which have a maximum current amplidude of 25 mA and a maximum pulse width of 500 μs, amplify the muscle contraction during swallowing and help the patient to swallow by elevating the larynx and hyoid. The stimulation is applied until the bioimpedance value rises again which indicates the end of the pharyngeal swallowing phase. Unfortunately, swallowing-like patterns in EMG and BI can also be generated by head movements, talking, chewing or food intake and lead to false stimulation triggering. To achieve a more comfortable therapy for dysphagia patients, an improvement of the stimulation triggering is needed. Range image cameras such as the Intel RealSense are capable to detect movements of the head and mouth. The color- and depth image data is used to track parts of the face by calculating landmarks that are placed on designated positions on the face Fig. 1 shows the positions of the landmarks, that are placed by the Intel RealSense face tracking algorithm[1], as well as the measurement electrodes for EMG/BI measurement at the neck.

We use the camera data to extend the classification process by a second classifier. The flow chart in Fig. 2 illustrates
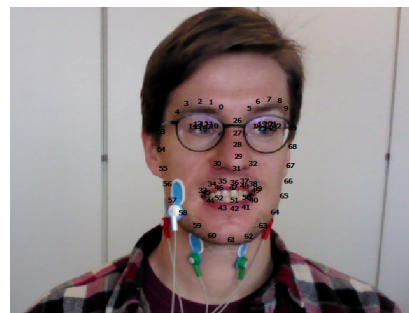
B. Riebold (corresponding author), H. Nahrstaedt and T. Schauer are with the Control Systems Group, Technische Universität Berlin, Germany. C. Schultheiss and R.O. Seidl are with the Unfallkrankenhaus, Ear, Nose and Throat, Berlin, Germany Unit Berlin, Germany (phone: +49-30-314-23680; email: riebold@control.tu-berlin.de).

[1]The Intel RealSense face tracking algorithm is included in the SDK that is available under: https://software.intel.com/de-de/intel-realsense-sdk

Figure 1. Positions of landmarks as given by Intel RealSense.

the complete classification process. The EMG/BI signals,that are recorded with the measurement system described in Nahrsteadt et al. [3], and camera data are synchronized. In a first step, the realtime EMG/BI classifier is used to find all time instances of potential swallow intentions. In a subsequent step we use the camera data to classify in realtime if the EMG/BI pattern was induced by a swallow onset or by a non-swallow related event. Only in the first case, a stimulation is triggered.
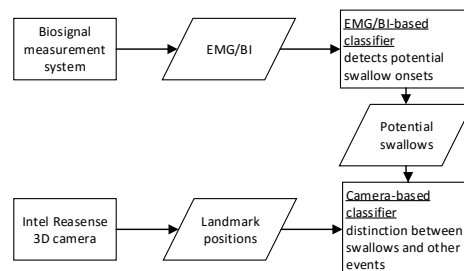

Figure 2. Two-step-classification process.

II. METHODS

A. EMG/BI-based classifier

The BI signal is downsampled from 2 kHz to 40 Hz by taking a window of 50 samples, denoising it with a Wavelet algorithm and taking the mean value as downsampled BI value. Based on the work by Nahrstaedt et al. [1], changes in the downsampled BI signal are detected by a modified sliding window and bottom-up (SWAB) algorithm [4] which is a piecewise linear approximation method (PLA). The length of the approximated lines is defined by the maximum error maxE between a line and the original data. This concept is enhanced by an upper bound maxL of the line length and a maximum difference maxD between the start and end point of a line. The downsampled BI data are continuously fed into the SWAB algorithm, that creates lines. The parameters of the SWAB algorithm are set to $maxE = 1\ \Omega$, $maxD = 0.25\ \Omega$ and $maxL = 1$ s.

In case a new line is found inside the SWAB algorithm, features are calculated for the EMG/BI classifier. The following features are used:

- SNR of the EMG activity during the newest line,

- last four slope values from the line segments that were found by the SWAB algorithm,

- SAX [5] word from the BI and EMG data [6] of the last 0.5 seconds.

The SAX word contains ten symbols and uses an alphabet size of 16 numbers. A random forest classifier is then used to decide which newly created line segment intentionally belongs to the beginning of a swallow.

B. Pre-analysis of camera data

The camera data is used for a second classification of the FES trigger time points generated by the EMG/BI based algorithm. To show how movements result into false classified swallows during the swallowing of saliva and solid boluses, we plotted true and false classified swallowstogether with the movement speed of the lower lip over time in Fig 3 and Fig. 4, respectively. It can be seen that chewing a solid bolus increases the movement speed of the lips. A higher moving speed is correlated with a higher number of false classified swallows. This can be explained as follows: By moving the jaw down the electrodes get shifted down along the larynx which leads to a drop of BI as well. The activation of the yaw muscles is measured as an EMG activity. Since both effects occur simultaneously, the EMG/BI-based classifier alone would trigger a false stimulation. These events need to be suppressed by the additional camera-based classifier. Movements of the head such as looking up and down or left and right, did not show a strong relation to false stimulation triggering in the data we recorded so far.
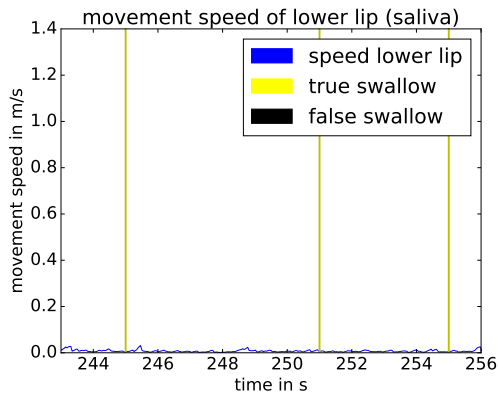

Figure 3. Swallow onsets detected by EMG/BI-basedclassifier while swallowing saliva.

C. Camera-based Classifier

For a proper classification, significant features have to be extracted from the data in a time window before a potential swallow onset. The first group of features models the movement of the mouth and jaw. For that reason, the moving speed of the upper and lower lip, the jaw and the mouth corners were used as features. We use Eq. (1) to get an approximation of the moving speed of a face part. The landmarks $\mathbf{x}_i(n)$ contained in a set L are differentiated and rectified, summed up and divided by K which is the number of elements in L. This is performed for each time instance n. The definition of the landmarks and the position of the
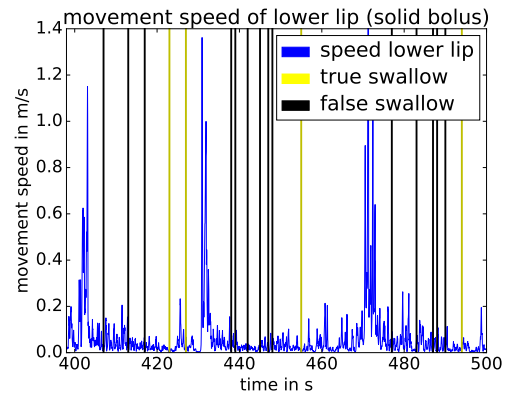

Figure 4. Swallow onsets detected by the EMG/BI-basedclassifier while swallowing bread.

measurement electrodes are shown in Figure 1. We used four different sets for feature calculation referring to the upper and lower lip, the mouth corners and the jaw. The sets of landmarks used for feature calculation are summarized in Eq. (2).

$$speed_L(n) = \frac{1}{K} \sum_L \left| \frac{d\mathbf{x}_l(n)}{dn} \right| \qquad (1)$$

$$
\begin{aligned}
L_{upperLib} &= \{35, 37, 47\} \\
L_{lowerLib} &= \{41, 43, 51\} \\
L_{mouthCorner} &= \{33, 39\} \\
L_{jaw} &= \{61\}
\end{aligned}
\qquad (2)
$$

A second group of features is used to determine if the mouth is open and if the jaw is deflected downwards. Those features are generated by calculating adistances measure between certain landmarks according to Eq. (3) and Eq. (4).

$$mouthOpen(n) =$$
$$\frac{|\mathbf{x}_{43}(n)-\mathbf{x}_{35}(n)|+|\mathbf{x}_{47}(n)-\mathbf{x}_{51}(n)|+|\mathbf{x}_{37}(n)-\mathbf{x}_{41}(n)|}{3} \qquad (3)$$

$$jawDown(n) = |\mathbf{x}_{61}(n) - \mathbf{x}_{31}(n)| \qquad (4)$$

The signals $speed_L$, $mouthOpen$ and $jawDown$ which we obtain from Eq. (1), Eq. (3) and Eq. (4) are used to calculate the features according to Eq. (5). That is done by the weighted summation of the values in an interval prior to the potential stimulation trigger time instance n_t with a sample length I. We used a hanning function with a width of two times I to weight the samples in the interval. Thus the contribution of a sample to the sum decreases with the distance to n_t.

$$f = \sum_{i=0}^{I} w_{2I}(I-i)Signal(n_t - i) \qquad (5)$$

The classification of potential trigger points is done using a vector support machine implementation by Chang et al. [7]. We found that radial basis functions as kernel yield the best results. The data was normalized with a standard scaling procedure implemented in scikit-learn [8]. Thus, each feature was normalized by first subtracting its mean and then dividing by its variance. To find the best suited classifier, a grid search on the parameters γ and C is performed. The

parameter C weights the influence of the penalty term in the cost function that is used for the vector support machine optimization, whereas γ controls the width of the radial basis function used as kernel.

D. Evaluation of both classifiers

The evaluation of the EMG/BI-based and camera-based classifieris performed by using the leave-one- subject-out method by which data from all but one subject are used for training and the remaining data are used for testing. This is repeated for each subject. The safety of the patient is the most important issue. Every swallow intention that is not detected barres the risk of aspiration. For that reason, a suitable classification system needs to have a high sensitivity. The sensitivity S is calculated as shown in Eq. 6. In our case the sensitivity is a measure for the rate of correct classified swallow intentions.

$$S = \frac{true\,positives}{true\,positives + false\,negatives} \qquad (6)$$

III. Results

We conducted experiments with five healthy subjects, that were asked to perform three swallows of four different boluses in five different head postures. This study was approved by the ethic board at the Charité Berlin (EA1/161/09). Swallows in the data set have been automatically segmented by using an offline swallow detection [6]. Additionally, the data sets were manually inspecting to confirm or reject found swallows or to mark non-found swallows. In total, 363 swallows are contained in the data set. The resulting data were used as reference for the evaluation of the described two classifiers.

A. EMG/BI-based classifier

The best training parameters for the random forest classifier are found to be $n\_estimators = 40$, $min\_samples\_leaf = 15$, $min\_samples\_split = 12$ and $max\_depth = 8$. A definition of the parameters can be found in the documentation of scikit-learn [7]. The class weights are adjusted inversely to the class frequency by setting the weights option to balanced. The testing results for all subjects are shown in Table I. As described by Nahrstaedt et al. [6], the pharyngeal swallow phase results in a valley in the BI signal with start, minimum and end point. The hyoid bone and larynx have their maximum displacement around the minimum point.

TABLE I. BI/EMG-BASED CLASSFIER: CONFUSION MATRIX.

Potential swallow onsets	predicted positive	predicted negative
true	335	28
false	273	16038

B. Camera-based classifier

After applying the grid search over the support vector machine parameters γ and C we have chosen the best suited parameter pair for our problem.Among all the classification results we selected the pair with the highest specificity and a sensitivity value above 90% ($\gamma = 0,82 \cdot 10^{-6}$ and $C = 689.96$). The confusion matrix is shown in Table II. The corresponding metrics is displayed in Table III. In average, swallow onsets were detected 190 ms after the true begin of a swallow and 369 ms before the maximal larynx elevation was observed.

TABLE II. CAMERA-BASED CLASSIFIER: CONFUSION MATRIX.

FES trigger points	predicted positive	predicted negative
true	301	34
false	144	129

TABLE III. METRICS OF CAMERA CLASSIFICATION RESULT.

precision	64%
sensitivity	90%
specificity	42%
accuracy	67%

IV. Discussion

The camera-based classifier reduced the number of resulting false stimulations by 47%. However, the number of resulting correct stimulations potential was also sligthy reduced by 11%. The observed timing in the swallow onset detection is sufficient for supporting an elevation of the larynx and hyoid by FES. These preliminary resultsare a starting point for further investigations. For practical use a higher specificity would be desirable to enable a comfortable therapy. Besides the need of a bigger data set to obtain more generalized trainings data and a more robust evaluation. The swallowing performance of dysphagia patients can be improved by adopting special head postures as shown in different publications [9]. The postures can be defined by pitch and yaw angles of the head. The desired head posture could be used to switch the stimulation system on and off. This would suppress most of the disturbances and create a more balanced class ratio for the classification system. Those angles can be measured with the Intel RealSense camera and used to give the patients a graphical biofeedback. Combining these two techniques could dramatically increase the stimulation precision.

References

[1] H. Nahrstaedt, C. Schultheiss, T. Schauer, and R. O. Seidl, "Bioimpedance- and EMG-Triggered FES for Improved Protection of the Airway During Swallowing," *Biomedical Engineering / Biomedizinische Technik*, 2013.

[2] C. Schultheiss, T. Schauer, H. Nahrstaedt, and R. O. Seidl, "Automated Detection and Evaluation of Swallowing Using a Combined EMG/Bioimpedance Measurement System," *The Scientific World Journal*, vol. 2014, p. 405471, 2014.

[3] H. Nahrstaedt, T. Schauer, and R. O. Seidl, "Bioimpedance based measurement system for a controlled swallowing neuro-prosthesis," in *Proc. of 15th Annual International FES Society Conference and 10th Vienna Int. Workshop on FES*, 2010, pp. 49–51.

[4] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *Proceedings 2001 IEEE International Conference on Data Mining*. IEEE, 2001, pp. 289–296.

[5] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. ACM, 2003, p. 2.

[6] H. Nahrstaedt, C. Schultheiss, R. O. Seidl, and T. Schauer, "Swallow Detection Algorithm Based on Bioimpedance and EMG Measurements," in *8th IFAC Symposium on Biological and Medical Systems*, Budapest, Hungary, 2012, pp. 91–96.

[7] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[9] P. Macrae, C. Anderson, and I. Humbert, "Mechanisms of Airway Protection During Chin-Down Swallowing," *Journal of Speech Language and Hearing Research*, vol. 57, no. 4, p. 1251, 2014.