Mealcoach: Contact Microphone-Based Meal Supervision For Post-Stroke Dysphagia Patients

Yanbin Gong [1], *Jia Qiao* [2], *Yandao Huang* [1], *Qian Zhang* [1], *Zulin Dou* [⋆2]

[1] The Hong Kong University of Science and Technology, Hong Kong, China
[2] Dept of Rehabilitation Medicine, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

ABSTRACT

Meal supervision for post-stroke dysphagia patients significantly improves prognosis during rehabilitation. Aspiration often occurs during meals, which may further incur aspiration pneumonia. Therefore, it's necessary to know the patient's swallowing ability as well as the occurrence of cough. Recently, some researchers have detected swallowing or coughing with audio signals and have made remarkable achievements. However, the users need to stay in quiet environments or wear uncomfortable cervical auscultation devices because the signals generated by swallowing are weak. In this work, we present *MealCoach*, a system that utilizes a contact microphone to collect high-quality signals to identify the events during meals. We take advantage of the insensitivity of contact microphones to ambient noise for free-living environment supervision. After balancing the wearing experience and identification accuracy, we elaborately select the optimal site to leverage the unique characteristics of cricoid cartilage movement during meals to accurately identify swallowing, coughing, speaking, and other events during meals. We collected data from thirty PSD patients in the hospital and evaluated our system, and the results demonstrate that *MealCoach* achieves a mean classification accuracy of 95.4%.

Index Terms— Post-Stroke Dysphagia, Meal Supervision, Contact Microphone, Cricoid Cartilage, Deep Learning

1. INTRODUCTION

Stroke is recognized as the third leading cause of severe long-term disability globally. One in four people may have a stroke in their lifetime. Post-stroke dysphagia (PSD) is a common sequela that increases mortality and morbidity due to aspiration and aspiration pneumonia [1]. Swallowing requires a series of muscles to work together. Due to partial brain damage in stroke patients, they are prone to swallowing incorrectly, which can lead to aspiration. Aspiration often occurs during meals, which may further incur aspiration pneumonia that causes the highest attributable mortality following stroke [2]. Therefore, it is essential to monitor the patient's swallowing status at mealtime and to identify the onset of coughing, which allows for timely intervention by the medical team. When a cough occurs after a patient swallows, there is a high probability that aspiration has occurred. Accurate identification of each swallow and cough can effectively enhance the patient's recovery.

Recently, researchers have managed to detect swallowing or coughing for dysphagia monitoring. Olubanjo *et al.* proposed a real-time swallowing detection system with a throat microphone [3]. Statistical features were utilized to discriminate swallowing from other events. The result is unsatisfactory due to the low signal-to-noise ratio (SNR) of swallowing generated sound. To increase SNR, Subramani *et al.* placed cervical auscultation (CA) device on the patient's neck over the lateral border of the trachea to capture the acoustic signatures [4, 5]. However, wearing CA devices may cause discomfort to PSD patients. Some researchers manage to complement the audio with an inertial measurement unit (IMU) to mitigate interference from ambient noise. Coughtrigger [6] first detects the candidate cough by IMU, then triggers the audio cough detection. Coyle *et al.* integrated a microphone and an IMU as high-resolution cervical auscultation (HRCA) devices to detect and analyze the swallowing events [7, 8]. However, patients were still required to stay in a quiet environment with a neutral head position because the IMU signals are sensitive to body motion, which is not applicable in the daily scenario. Therefore, HRCA is more useful for bedside screening but not for daily dynamic monitoring.

Contact microphone has been applied for dysphagia in recent research [9]. It has the advantages of being insensitive to environmental noise, low cost, and easy to wear. However, because of the low intensity of the swallowing signals, it is easy to overlook them among the other vibration signals produced by vocal cord vibration. In this work, we combine clinical experience with signal-processing techniques to deal with the issue. We propose *MealCoach*, a contact microphone-based four-category classification system to identify the *swallow*, *cough*, *speaking*, and *other* body movements during the meals in free-living environments. Detecting these events has significant clinical value because a cough immediately after a swallow indicates aspiration. We observe that the hyoid bone and larynx movement can be unique characteristic of swallowing, which can help us detect swallowing. We analyze the signal quality picked up by the contact microphone at various locations around the neck and manage to capture the signals with unique characteristics. We also consider the wearing convenience and comfort of designing the wearing method based on clinical experience and patient surveys. We choose the optimal site for the contact microphone to be deployed, satisfying the signal quality and the patient's wearing willingness. A signal processing and deep learning-based pipeline are designed for the classification task. We first filter out the buzzing leaked by the improperly shielded capsule, then augment the data by shifting and masking. Because Convolutional Neural Networks (CNNs) have proven very effective for audio classification tasks [10], we apply CNN to identify the non-continuous meal events. We first convert the vibration signals to Mel Spectrogram and then feed them into the robust ResNet-34 [11] for classification.

The main contributions of this work are: **(1)** We propose *MealCoach*, a contact microphone-based meal supervision system, which is useful for early diagnosis of aspiration. **(2)** We pick the sens-

⋆ Corresponding author.
This research is supported in part by Guangzhou Science and Technology Plan Project 2023B03J1234, RGC under Contract CERG 16203719, 16204820, and Contract R8015, FS112.

ing modality under the clinical property and needs. We carefully select the deployment site of the contact microphone considering the patient's wearing comfort as well as the actual usage scenario. **(3)** We evaluate *MealCoach* in PSD patients. We set suitable inclusion/exclusion criteria and prepared specific bolus to ensure the reasonable validity of the experiment. The experiments under the supervision of medical professionals demonstrate that our system achieves high accuracy and robustness, which meets clinical needs.

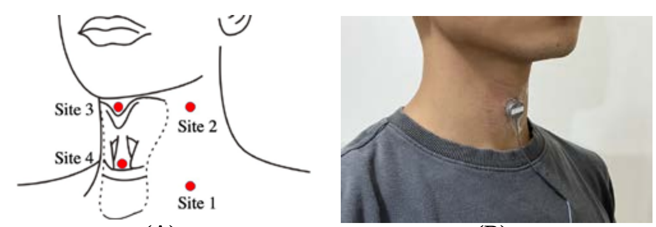
2. METHODOLOGY

2.1. Optimal Site Selection

Selecting a suitable site to attach the contact microphone is of great significance as it determines the signal quality as well as the wearing experience. We expect that wearing the sensor will not cause discomfort to the patient, while the sensor can collect sufficiently obvious signals. As detecting swallowing is the toughest task, we conduct a preliminary study for site selection. After analyzing the solution of current state-of-art works and conducting clinical studies, we selected four alternative locations, which are shown in Fig. 1(A). We asked patients about their wearing experience, and conducted a preliminary study of signal quality on healthy people for ethical reasons. We deploy the contact microphone on each site to pick up signals under various behaviors. We further evaluate them theoretically and experimentally and select the optimal site. We analyze the SNR and distribution on the statistical features of the signals, including windowed energy (WE), peak frequency (PF), and Shannon entropy (SE) [12]. In order to distinguish swallowing signals from signals induced by body movements, we calculate the distance between events:

$$\begin{aligned} DIS(s, o) = l^2\text{-}norm(&MMS(WE(s) - WE(o)), \\ &MMS(PF(s) - PF(o)), MMS(SE(s) - SE(o))) \end{aligned} \tag{1}$$

where DIS indicates the distance between two samples, s represents swallowing signals and o represents signals of other events. MMS is the MinMaxScaler to standardize the feature distance between 0 and 1. And the distance is measured by l^2-norm.



(A) (B)

Fig. 1. (A) Four alternative locations. (B) Contact microphone attached superior to the cricoid cartilage.

Site 1: Over the lateral border of the trachea inferior to the cricoid cartilage. Site 1 was considered the optimal site for throat microphones as this site showed the greatest averaged magnitude of the SNR with the smallest variance[13], and is used by some researches[3, 12]. But we find that the swallowing signals detected here are very weak and it is hard for medical professionals to distinguish the swallow from other events. In addition, the carotid signals here may also influence judgment. The distribution analysis of statistical features shows that the swallowing signals are individual-specific, making it difficult to distinguish swallowing and other events on different subjects.

	Site 1	Site 2	Site 3	Site 4
SNR(dB)	1.37	1.75	6.18	4.15
Feature Distance	0.39	0.35	0.51	0.55

Table 1. Signal qualities of different sites.

Site 2: Two inches on either side of the hyoid bone. Site 2 is specially applied for swallow detection, and experiments on multiple participants show that the swallowing signal strength here is the strongest [14]. However, the absolute signal strength is still relatively weak compared to the sound of a hand sliding across the neck, and it is still difficult to distinguish it from other signals induced by neck movement. For this site, we find the same problem as Site 1 that it is hard to cluster the signals by events rather than by subjects.
Site 3: Superior to the thyroid cartilage. We observe that the hyoid bone and larynx movement can be a unique characteristic of swallowing, and thus detecting this motion can help us detect swallowing. Elevation of the thyroid cartilage produces strong and distinctive collision signals. However, some compensatory actions may also lead to slight movement of the thyroid cartilage that may generate strong signals, which may be misclassified as a swallow, while a swallow is not initialized successfully. What's more, applying pressure over the thyroid cartilage can stimulate swallowing and cause discomfort to the patients, thus the position is not suitable for patients with swallowing disorders.
Site 4: Inferior to the thyroid cartilage and superior to the cricoid cartilage. Similar to Site 3, the upward and downward movement of the cricoid cartilage during swallowing will strike the central rubber pad of the contact microphone twice, resulting in two very distinct signals. The signals here are not as strong as from Site 3, However, the unique signal characteristics generated by the collision can still be collected. The vibration of the muscles when swallowing initiates produces signals, but its strength is significantly weaker than the two signals produced by the impact of the cricoid cartilage against the rubber pad. Thus, the signals from swallowing are clearly distinguishable from the weak signals of muscle movement and the strong signals from the vibration of the vocal cords when coughing or speaking. In addition, placing the microphone here will not cause any discomfort to the patient.

To illustrate the signal qualities among the sites, we shifted the swallowing signals collected from the same participant on different sites and plot on them on the same figure. In other participants, the relationship between these signals is consistent. As shown in Fig. 2, the signals near the peak are induced by swallowing. Site 1 and Site 2 have similar signals, while Site 2 has a higher SNR and stronger peak signal. For Site 4, the signals are much stronger, especially the characteristic signals induced by cricoid cartilage movement. The signals from Site 3 is the enhanced edition of Site 4, but considering the comfort of wearing, we give up on it and do not draw it in Fig. 2 in order not to block other signals . We also compare the SNR and feature distance of the signals picked up from different sites, as shown in Table 1, and we can find the signals from Site 4 are more suitable. Therefore, we choose **Site 4**, superior to the cricoid cartilage, to attach the contact microphone.

2.2. Meal Event Classifier

To identify the significant events during meals, we build a four-category classification model to identify *swallow*, *cough*, *speaking*,

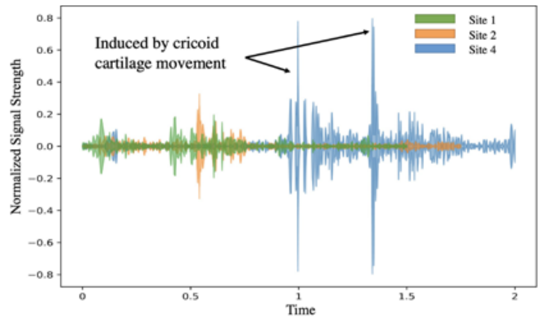


Fig. 2. Swallow signals collected from the same participant.

and *others* that represents possible body movement for meal supervision. Instead of inputting the vibration signals directly, we preprocess the signals and convert them to Mel Spectrogram [15], then deal with the task with the image classification model.

Fixed-size sliding windows are created to detect and identify events. Based on the clinical observation that the duration of most swallowing events of target patients is within two seconds, we segment the data into two seconds. As the sample rate is set to 48kHz, the shape of the data is (1,96000). After getting the data, we filter out the ripple noise with a notch filter. To reduce power consumption for long-time monitoring, we identify meal events based on the energy of the vibration signals. The threshold is set to half the strength of the average PSD patient's swallowing signals. When the signal strength is lower than this threshold, we classify the event as *others*.

Once the events are detected, we convert the signals to Mel Spectrograms by Short-time Fourier Transform (STFT). We set the size of the Fast Fourier Transform (FFT) to 1024, the hop length to 512, and the number of frequency bands to 64. Thus, we generate the Mel Spectrograms shape of (1,64,188). We augment the data at the training stage to make the model more robust. We randomly shift the window before the data is converted to a spectrogram to simulate the inefficiency of the event detection algorithm. Then we mask the spectrogram in the time and frequency domains for further augmentation.

Finally, we train a neural network for event classification. We choose ResNet [11] because it is mature, robust, and has an excellent performance in multi-scenario tasks. Other neural network architectures can also be applied with the input Mel Spectrogram [10], but we do not find a clear advantage from these networks. Due to the limited data size, we fine-tune a pre-trained ResNet, which is optimized by Adam optimization. We empirically select ResNet-34 because its experimental performance outperforms ResNet-18, ResNet-50 and ResNet-101. The performance of the model is determined by the model's capacity, the complexity of the task, and the amount of data. The workflow is described in Algo 1.

Algorithm 1 Meal Event Classifier

Input: $data$: a window of two seconds signals
Output: $Event$: an identified event, *swallow*, *cough*, *speaking*, or *others*

```
data ← Notch_Filter(data, 50Hz, 100Hz)
if Energy(data) < Threshold then
    Return Event ← others
end if
data ← Mel_Spec(data, fft=1024, hop=512, mels=64)
Return Event ← ReNet-34(data)
```

3. EVALUATION

3.1. Experiment Setup

Hardware Setup. We use CM-01B contact microphone to pick up the audio signals. Compared with other contact microphones, CM-01B is sized for deployment over the cricoid cartilage and is lightweight. The contact microphone collects vibration signals through a rubber head, which is attached to the neck right inferior to the hyoid bone, and superior to the cricoid cartilage by a sterile wound dressing, as shown in Fig. 1(B). The microphone records the vibration signals with the sample rate set to 48 kHz.

Participants and Experimental Protocol. We recruited PSD patients to validate the effectiveness of *MealCoach*.[1] The inclusion criteria are that PSD patients with a PAS score greater than or equal to 4 with a clear diagnosis of overt aspiration/no aspiration by routine standardized swallowing angiography. At the same time, patients with disturbance of consciousness, tracheotomy, or continuous blood oxygen saturation below 95%, or using artificial ventilators to assist ventilation, or combined with cognitive impairment, or combined with head and neck lesions were excluded.

After screening, thirty PSD patients were selected to participate in the experiment (aged 16 to 69, 15 male and 15 female). Evaluated by doctors and therapists, all participating patients were in a safe recovery phase and were assessed to be able to have eating training. To ensure safety, the experiment was conducted in the doctor's office of the inpatient department. The patients were instructed to swallow, cough, speak, or have other acts under the supervision of the medical team. The experiment would be stopped early when the patient felt tired, unwell, or had significant difficulty initiating swallowing. The patients were asked to swallow 3 ml water or 3 ml prepared boluses whose viscosity was 187 cP measured with a rheometer at 25 °C and a speed of 50 s-1. Each patient would swallow between 3 to 18 times, with a total of 294 swallows collected. For other events, an average of twenty actions per patient were performed, with a total of approximately 600 data collected for each category.

We also collected data from twenty healthy volunteers for model training (aged 18 to 56). The experiments were conducted in free-living conditions full of noise, which simulated the meal environment. Participants were asked to swallow, cough, talk, and move necks or other body parts. The experiments were recorded and labeled simultaneously by researchers. Fifty samples were collected for each individual for each type of event, with a total of 1000 samples collected for each type.

3.2. Model Validation

Our model is evaluated by k-fold cross-validation. As each swallowing or coughing data by the same person may be similar, we consider Leave-One-Subject-Out Cross Validation (LOSOCV) to evaluate the generalization ability of the participant-independent model [16], where only the data of the participant being tested will be left and other subjects' data are used for model training.

We evaluate the overall performance of *MealCoach* by its classification accuracy. In addition, as this work focuses much on swallowing identification, we further evaluate it by recall and precision, which measures swallowing detection ability and the chance of getting false positives. These metrics show how useful the model is in reality. To better express the model's accuracy, we balance the size

[1] Experiments were conducted following the ethical policies of our institutions. (Issued number: HKUST: HREP-2022-0120, and SYSU: [2022]02-192-01.)

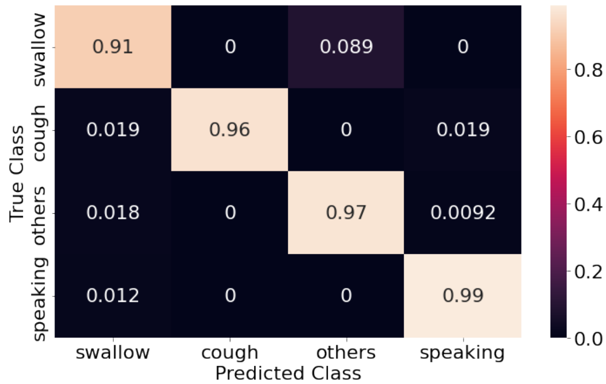


Fig. 3. Confusion matrix of *MealCoach* on PSD patients.

of each class on the test set. In actual scenarios, swallowing occurs most frequently, and coughing events occur much less frequently than other events.

3.3. Evaluation Result

Performance on PSD Patients. We first validated the performance of *MealCoach* in PSD patients. The confusion matrix of classification result is plotted in Fig.3, demonstrating that *MealCoach* achieves an average classification accuracy of 95.4%. This result is facilitated by the large differences in the signal characteristics of the different behaviors during mealtime. During swallowing, laryngeal bone uplift and repositioning against the rubber probe produced signals in two segments, which is unique enough to get identified. The vibration signals induced by coughing are short and strong. As for speaking, the produced signals are continuous and strong, which differs from coughing signals. In addition, when performing other body movements or resting, the contact microphone collected no obvious signal but just some noises. Therefore, *MealCoach* can classify the events well in the time-frequency space.

We analyze the misclassified samples and find that motion artifacts caused by compensatory behavior (e.g., swinging the head or using hand to support) are evident for some patients, leading to the misclassification of *swallowing* as *others* or *speaking*. The reason may be that PSD patients are hard to initiate swallowing due to weak muscles after several swallows, and they tend to perform some compensation actions. Therefore, the swallowing movement is particularly weak, or the swallowing is not even initiated successfully, resulting the misclassification.

Swallowing Detection. Swallowing detection is one of our major contributions. In this section, we compared *MealCoach* with other related works. The results in Table 2 show that our work achieves 91% recall and 95% precision for swallowing detection, which is significantly superior compared to other works. Since the contact microphone does not pick up the sound of air pressure reverberating through the airway during swallowing but collects the sound of cricoid cartilage hitting the contact microphone directly, the signals generated during swallowing have a high SNR. Therefore, *MealCoach* is more robust in noisy environments. Constantinescu *et al.* [17] also achieved good performance. However, their system is based on sEMG with a more complicated device attached to the jaw, and may cause discomfort.

	Recall	Precision	Noisy Env.
Olubanjo *et al.* [3]	0.80	0.68	No
Constantinescu *et al.* [17]	0.92	0.84	Yes
Khalifa *et al.* [18]	0.85	0.84	No
Nakamura *et al.* [19]	0.86	0.74	Yes
MealCoach	**0.91**	**0.95**	**Yes**

Table 2. Comparison with related works.

4. DISCUSSION

In addition to existing functionalities, in this section, we discuss some potential future works for facility improvement.

4.1. Respiration Detection

Respiratory-swallowing coordination is important for the assessment and intervention of swallowing disorders [20]. During our experiments, we find that when the contact microphone is deployed over the cricoid cartilage, a distinct respiratory sound could be heard. In future studies, we can combine respiratory-swallowing coordination analysis with respiratory-phase recognition algorithms [21] to design new clinical interventions.

4.2. Chewing Detection

Chewing and swallowing information, when combined, are considered important features in assessing dietary health. There are much research works about chewing detection through various sensors [19, 22]. Since our sensor deployment site is designed to identify swallowing, we could not extract significant chewing information from the contact microphone. Therefore, in future work, we may consider combining other sensors to extract swallowing and chewing signals.

5. CONCLUSION

In this paper, we present *MealCoach*, a contact microphone-based meal supervision system designated for PSD patients. The user only needs to attach the contact microphone to the superior of the cricoid cartilage; then the system can automatically identify each swallow, cough, and speech when they are eating in a free-living environment. Based on the information collected, the medical team can assess the user's status, determine if aspiration has occurred, and update the rehabilitation program. The system is widely deployable due to a minimal set-up requirement – a contact microphone. Based on clinical observations and experiments, we have selected the optimal site that allows us to obtain uniquely characterized swallowing signals with strong strength. We use signal processing and deep learning techniques to make the system robust and can extract features from different levels. By adapting pre-trained ResNet-34, we can accurately identify each swallow and cough in various noisy environments, thus contributing a lot to the early identification of aspiration. The experiments on thirty PSD patients in different courses show that *MealCoach* is robust and delivers a high performance of 95.4% classification accuracy. Our experimental analysis also reveals why some events are misidentified. We find that in some fatigue-prone PSD patients, involuntary compensatory swallowing activity is likely, which is likely to increase the risk of aspiration. This finding has important clinical value, which can help to continuously assess the patient's recovery status and help design corresponding rehabilitation methods.

6. REFERENCES

[1] Marc Randall, Katie Neurology Robson, and Lisa Woodhouse, "Post stroke dysphagia: a review and design considerations for future trials," .

[2] AR Gujjar, E Deibert, EM Manno, S Duff, and Michael N Diringer, "Mechanical ventilation for ischemic stroke and intracerebral hemorrhage: indications, timing, and outcome," *Neurology*, vol. 51, no. 2, pp. 447–451, 1998.

[3] Temiloluwa Olubanjo and Maysam Ghovanloo, "Real-time swallowing detection based on tracheal acoustics," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014, pp. 4384–4388.

[4] Siddharth Subramani, MV Achuth Rao, Divya Giridhar, Prasanna Suresh Hegde, and Prasanta Kumar Ghosh, "Automatic classification of volumes of water using swallow sounds from cervical auscultation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1185–1189.

[5] Siddharth Subramani, Achuth Rao, Anwesha Roy, Prasanna Suresh Hegde, and Prasanta Kumar Ghosh, "Segnet-based deep representation learning for dysphagia classification," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 1141–1145.

[6] Shibo Zhang, Ebrahim Nemati, Minh Dinh, Nathan Folkman, Tousif Ahmed, Mahbubur Rahman, Jilong Kuang, Nabil Alshurafa, and Alex Gao, "Coughtrigger: Earbuds imu based cough detection activator using an energy-efficient sensitivity-prioritized time series classifier," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 1–5.

[7] Faezeh Movahedi, Atsuko Kurosu, James L Coyle, Subashan Perera, and Ervin Sejdić, "A comparison between swallowing sounds and vibrations in patients with dysphagia," *Computer methods and programs in biomedicine*, vol. 144, pp. 179–187, 2017.

[8] Cara Donohue, Yassin Khalifa, Subashan Perera, Ervin Sejdić, and James L Coyle, "A preliminary investigation of whether hrca signals can differentiate between swallows from healthy people and swallows from people with neurodegenerative diseases," *Dysphagia*, vol. 36, no. 4, pp. 635–643, 2021.

[9] Joshua M Dudik, Atsuko Kurosu, James L Coyle, and Ervin Sejdić, "Dysphagia and its effects on swallowing sounds and vibrations in adults," *Biomedical engineering online*, vol. 17, no. 1, pp. 1–18, 2018.

[10] Shawn Hershey, Sourish Chaudhuri, Daniel PW Ellis, Jort F Gemmeke, Aren Jansen, R Channing Moore, Manoj Plakal, Devin Platt, Rif A Saurous, Bryan Seybold, et al., "Cnn architectures for large-scale audio classification," in *2017 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE, 2017, pp. 131–135.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[12] Temiloluwa Olubanjo and Maysam Ghovanloo, "Tracheal activity recognition based on acoustic signals," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 1436–1439.

[13] Koji Takahashi, Michael E Groher, and Ken-ichi Michi, "Methodology for detecting swallowing sounds," *Dysphagia*, vol. 9, no. 1, pp. 54–62, 1994.

[14] Q Pan, Naoto Maeda, Yousuke Manda, Naoki Kodama, and Shougo Minagi, "Validation of the optimal site in the neck region for detecting swallowing sounds," *Journal of oral rehabilitation*, vol. 43, no. 11, pp. 840–846, 2016.

[15] Steven Davis and Paul Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE transactions on acoustics, speech, and signal processing*, vol. 28, no. 4, pp. 357–366, 1980.

[16] Atesh Koul, Cristina Becchio, and Andrea Cavallo, "Cross-validation approaches for replicability in psychology," *Frontiers in Psychology*, vol. 9, pp. 1117, 2018.

[17] Gabriela Constantinescu, Kristina Kuffel, Daniel Aalto, William Hodgetts, and Jana Rieger, "Evaluation of an automated swallow-detection algorithm using visual biofeedback in healthy adults and head and neck cancer survivors," *Dysphagia*, vol. 33, no. 3, pp. 345–357, 2018.

[18] Yassin Khalifa, James L Coyle, and Ervin Sejdić, "Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings," *Scientific Reports*, vol. 10, no. 1, pp. 1–13, 2020.

[19] Akihiro Nakamura, Takato Saito, Daizo Ikeda, Ken Ohta, Hiroshi Mineno, and Masafumi Nishimura, "Automatic detection of chewing and swallowing," *Sensors*, vol. 21, no. 10, pp. 3378, 2021.

[20] Bonnie Martin-Harris, "Clinical implications of respiratory–swallowing interactions," *Current opinion in otolaryngology & head and neck surgery*, vol. 16, no. 3, pp. 194, 2008.

[21] Chen-Hsuan Shih, Naofumi Tomita, Yanick X Lukic, Álvaro Hernández Reguera, Elgar Fleisch, and Tobias Kowatsch, "Breeze: Smartphone-based acoustic real-time detection of breathing phases for a gamified biofeedback breathing training," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 3, no. 4, pp. 1–30, 2019.

[22] Yanxin Wei, Khairun Nisa' Minhad, Nur Asmiza Selamat, Sawal Hamid Md Ali, Mohammad Arif Sobhan Bhuiyan, Kelvin Jian Aun Ooi, and Siti Balqis Samdin, "A review of chewing detection for automated dietary monitoring," *Journal of the Chinese Institute of Engineers*, vol. 45, no. 4, pp. 331–341, 2022.