



**HAL**  
open science

# Detection of Pharyngolaryngeal Activities in Real-World Settings Using Wearable Sensors

Lila Gravelier, Maxime Le Coz, Jérôme Farinas, Julien Pinquier

## ► To cite this version:

Lila Gravelier, Maxime Le Coz, Jérôme Farinas, Julien Pinquier. Detection of Pharyngolaryngeal Activities in Real-World Settings Using Wearable Sensors. 46th annual international conference of the IEEE engineering in medicine and biology society (EMBC 2024), EMBC: Engineering in Medicine and Biology Society, Jul 2024, Orlando (Florida), United States. à paraître. hal-04561003

**HAL Id: hal-04561003**

**<https://hal.science/hal-04561003>**

Submitted on 26 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Detection of Pharyngolaryngeal Activities in Real-World Settings Using Wearable Sensors

Lila Gravellier  
IRIT, Université de Toulouse  
CNRS, Toulouse INP, UT3  
Swallis Medical  
Toulouse, France  
lila.gravellier@irit.fr

Maxime Le Coz  
Swallis Medical  
Toulouse, France  
m.lecoz@swallismedical.com

Jérôme Farinas  
IRIT, Université de Toulouse  
CNRS, Toulouse INP, UT3  
Toulouse, France  
jerome.farinas@irit.fr

Julien Pinquier  
IRIT, Université de Toulouse  
CNRS, Toulouse INP, UT3  
Toulouse, France  
julien.pinquier@irit.fr

**Abstract**—Automatic detection and identification of pharyngolaryngeal activities (PLA) such as swallowing, speaking, and coughing, are crucial steps in developing a system for diagnosing dysphagia (swallowing disorders) using high-resolution cervical auscultation (HRCA) signals. Using sensors placed on the throat, HRCA provides clinicians with a non-invasive means of assessing swallowing. This study introduces a new approach based on depthwise CNN and LSTM networks to automatically detect and identify PLA in synchronized accelerometer and microphone signals. This method achieves an  $F_{\text{score}}$  of 88% for swallow detection across seven hours of HRCA signals recorded from 42 healthy subjects under both supervised and ecological conditions. It outperforms state-of-the-art systems in real-world settings. On other activities, the system reaches an  $F_{\text{score}}$  of 93% for phonation and 86% for airway defense mechanisms, making it a promising tool to assist clinicians in the analysis of HRCA signals and a reliable base towards non-invasive assessment of dysphagia.

**Index Terms**—Event Detection, Deep Learning, Swallowing, Microphone, Accelerometer, High-Resolution Cervical Auscultation, Pharyngolaryngeal Activities

## I. INTRODUCTION

Swallowing disorders (dysphagia) can become life-threatening without early diagnosis and treatment [1]. However, current reference examinations (X-rays, fibroscopy) are not adapted to screen patients on a massive scale, due to their invasiveness, inaccessibility or cost [2]. High-resolution cervical auscultation (HRCA) is a promising alternative that involves placing non-invasive sensors on the patient’s throat and interpreting the captured signals to assess dysphagia. Microphones and accelerometers, combined or separately, were successfully used to identify abnormal swallowing events [3], [4] and describe displacements of anatomical structures relevant to swallowing assessment [5], [6].

These analyses and interpretations depend on identifying the locations of swallows within the signals, but this task remains time-consuming when carried out manually. Several researchers have explored methods for automated event detection, but the best results were conducted exclusively on patients, and mostly in highly controlled conditions compatible with concurrent videofluoroscopy (X-rays) [7], [8]. Achieving

This work has been carried out under the funding of an industrial doctorates fellowship from the National Association for Research and Technology (ANRT), France.

performant detection of swallowing events, both in healthy individuals and patients, is essential for screening swallowing disorders. In their scoping review, So et al [9] point out that available swallow detection studies within the healthy population suffer from limitations either due to lack of performance, data, precision in the protocol or in the description of evaluation metrics. Unlike X-rays and fibroscopy, HRCA can also provide insights for analysis under real-life conditions, giving healthcare professionals more information on the natural behaviour of the patients. Swallowing detection methods will need to be adapted and tested to deal with those often challenging circumstances. Sazonov et al. [10] are among the few who tested a detection method in ecological settings. They tested their method on a group of twenty individuals while other ecological studies typically involved fewer participants, limiting their generalizability. They demonstrated a decrease in performance when transitioning towards a complete ecological setting, by introducing elements like conversation or background noise.

Additionally, the presence or absence of other pharyngolaryngeal activities (PLA) like coughing, throat clearing, and vocal characteristics can be crucial diagnostic indicators of swallowing inefficiencies [11], [12]. Automating the analysis of ecological recordings by identifying and locating PLA would significantly enhance patient recording analysis. A recent work has led to the development of the first algorithm capable of detecting swallowing, coughing, and speaking in controlled scenarios, specifically in thirty post-stroke dysphagic patients [13]. They achieved high precision and recall rates of 95% and 91%, respectively, in swallow detection as well as high scores for the speaking and coughing class.

In our present work, we describe a novel method, called *PLA-Tracker*, for detecting three types of PLA in HRCA signals from the healthy population. Our goal is to provide a comprehensive solution that: (i) achieves state-of-the-art performance in controlled swallowing detection scenarios; (ii) outperform precedent studies under ecological conditions; and (iii) simultaneously detects phonation and airway defense mechanisms.

## II. MATERIALS

All vibroacoustics signals used in this study were recorded with the Swallis DSA (@Swallis Medical, France). This device includes a camera and a necklace equipped with a tri-axis accelerometer and a microphone (20 kHz sampling frequency) which are positioned on the neck respectively at the anterior part of the thyroid cartilage and the cricoid cartilage. These positions are among the optimal locations for swallowing detection according to Takahashi [14] and Mamun [15]. Figure 1 shows the device and the sensors position and figure 2 is an example of signals captured by the device when a subject speaks and then swallows.

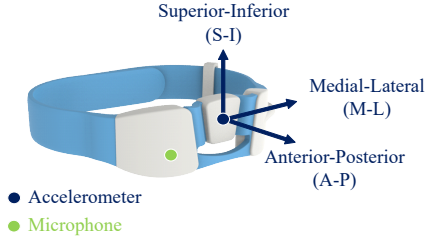


Fig. 1. Swallis DSA<sup>TM</sup> necklace and location of the sensors.

Forty-two healthy subjects (23 females and 19 males) aged from 22 to 57 years (mean  $34 \pm 10$  y.o.) were recruited for this study. Prior to participation in this study, all subjects provided written informed consent in accordance with ethical guidelines. The data collected from these subjects were treated with the utmost confidentiality and in compliance with the regulations of the French data protection authority (CNIL 2225691). To be included, they should neither have undergone neck surgery nor have been diagnosed with dysphagia. Each subject first completed a supervised feeding scenario and then an ecological one. During the first scenario, the operator asked the subject to do several PLA summarized in table I. The international dysphagia diet standardisation initiative (IDDSI) [16] framework was used to characterize the different textures swallowed, except for saliva.

TABLE I  
TASKS INCLUDED IN THE SUPERVISED SCENARIO.

PLA	Bolus	Volume	Repetition
Swallowings	Saliva	Variable	3
	Water (0)*	10 mL	3
	Gelled water (3)	2.5 mL	3
		5 mL	3
	Gelled water (4)	2.5 mL	3
		5 mL	3
	Biscuits (7)	3 g	3
	Water (0)	10 cL	1
Throat clearing			1
Cough	-	-	1
Phonation			1 sentence

\*In brackets the IDDSI texture level.

During the second scenario, they were asked to consume a snack consisting of three food items (12 g of IDDSI 7 biscuits,

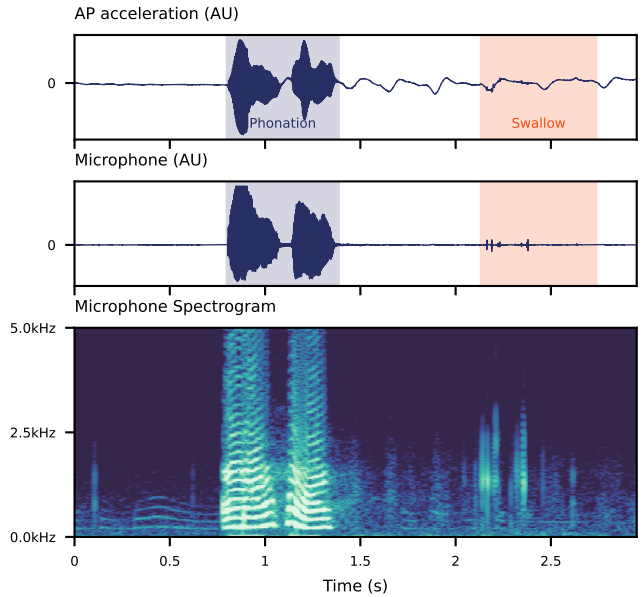


Fig. 2. Illustration of phonation and swallowing through the sensors.

10 cL glass of IDDSI 0 water and 100 g of IDDSI 4 gelled water) while reproducing their everyday, natural eating habits. To obtain even more natural recordings, participants were not limited in time or position and could talk with the operators. For each scenario, a continuous recording per subject was acquired without interruption.

HRCA experts annotated 84 recordings (7 h 46 min) using HRCA signals and synchronised video. Table II lists the cumulative duration of annotated PLA.

TABLE II  
COUNT OF ANNOTATED PLA SEGMENTS AND CUMULATIVE DURATION BY SCENARIO.

PLA	Supervised (Sup)	Ecological (Eco)	ALL
Swallowing	1722* (24 min)	1005 (11 min)	2727
Cough/Throat clearing	129 (2 min)	24 (14 s)	133
Phonation	1493 (26 min)	983 (17 min)	2476

\*Including 880 requested swallows.

## III. AUTOMATIC DETECTION OF PHARYNGOLARYNGEAL ACTIVITIES

Our goal is to automatically detect swallows (S) of different volumes, textures, either spontaneous or supervised, as well as phonation (P) and airway defense mechanisms (ADM) in both supervised and ecological scenarios. *PLA-Tracker* combines two tasks: an activity detector to select segments in the recording and a classifier based on a neural network to predict the class of each segment.

### A. Activity detection

In this corpus, the annotated PLA segments represent less than one-eighth of the total recordings, leading to the majority of the signal consisting of silence or non-PLA events (classified as "Others"). The use of sliding windows as input data results in suboptimal learning for the systems presented

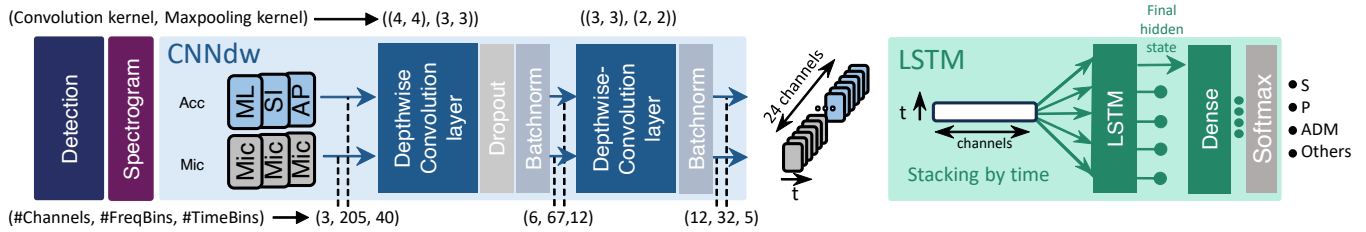


Fig. 3. Detailed configuration of the *PLA-Tracker* pipeline with depthwise convolution (CNNdw+LSTM). Convolution layers are composed of a convolution block, an activation layer (ReLU) and a max pooling layer.

in this work. Our proposed solution is the creation of an activity detector designed to identify pertinent candidates for the classification process. As Morinière et al. have shown that swallows necessarily generates at least one high energy peak in the acoustic signal [17], our detector employs a peak detection method applied to the smoothed energy of the acoustic signal. The process is the following:

- Calculate energy with window size 250 ms (step 25 ms);
- Smooth energy with 8-samples average sliding window;
- Perform peak detection.

Two advantages have been identified when employing this approach: 1) reduction in the volume of analyzed data, coupled with a reduction of the dataset imbalance; 2) improved stability by providing the classifier with data exclusively centered around peaks. These aspects result in more robust and efficient learning by neural networks. Moreover, this detection does not require prior knowledge of the content of the recordings.

### B. Input segments

For each high-energy peak detected, a spectrogram is generated for each source channel over a 500 ms segment centered on the peak. Although swallowing sounds can last from 250 to 800 ms [18], using a 500 ms window enhances our ability to detect short events while avoiding overlap. The four spectrograms are computed using an NFFT of 1024 samples and a step size of 256 samples. Subsequently, they are cropped to retain only frequencies between 0 and 4 kHz as main frequencies associated with swallowing remain below 4 kHz [19].

### C. Classification

To classify the segments, we developed a convolutional neural network with depthwise separable convolutions (CNNdw) followed by a long short-term memory (LSTM) network. Its architecture is detailed in figure 3. As demonstrated in [8], CNN is a robust feature extractor when applied to HRCA signal spectrograms. However, in preliminary experiments, the global convolutions applied to the four input sources tended to disproportionately highlight the audio source in the CNN’s output. While the microphone is highly valuable for HRCA signal analysis and is often used alone, the accelerometer contributes supplementary data that enhances the ability to distinguish key events. CNNdw preserves sensor data separation in CNN layers, ensuring an accurate HRCA signal representation at the CNN’s output. With three signals from the accelerometer and only one from the microphone, various combinations were

possible to create two equal-sized sensor groups. So far, the most effective one is to place all three accelerometer axes in one group and replicate the microphone signal three times in the other group. Then, at the output of the CNNdw network, the resulting channels are concatenated over time, and each time step is subsequently fed into an LSTM layer. The final hidden state is then passed through a dense neural network. Figure 3 details the structure of each block from the proposed method. The final output determines the class: Swallow (S), Phonation (P), Airway Defense Mechanism (ADM) or ”other”. This final class comprises all sounds that are sufficiently loud to be selected by the activity detector but do not belong to the primary classes (S/P/ADM). It includes, for example, noises like utensil clatter, necklace impacts or friction, burps, and sneezes.

### D. Training and testing

To get consistent results and assess the model’s robustness, a 14-fold cross validation was performed. For each split, the recordings from 3 subjects were put in the testing set, 3 others in the validation set, and the 36 left in the training set. This method ensures that a subject cannot appear in both train and test part (speaker independent system). It is a critical condition to validate the functioning of this model on a new unseen subject. An adaptive learning rate was used, starting at  $10^{-3}$  and halved when the loss function reaches a plateau. The network underwent training for 300 epochs with batch sizes of 64. We used cross-entropy as the loss function and weighted Adam optimizer (AdamW) with a weight decay of 0.1. The validation set was used to set the optimal number of epochs for the final test.

### E. Evaluation method

To evaluate the detection performance of the system, an activity was considered as hit (true positive) if at least one predicted segment with the correct label (S/P/ADM) overlapped with it. Otherwise the activity was considered as missed (false negative). Any predicted PLA segment that did not overlap with an annotated segment of the corresponding label was recorded as a false alarm (false positive). Precision, recall, and  $F_{score}$  were used as evaluation metrics.

## IV. RESULTS AND DISCUSSION

### A. Results

Table IV summarizes the method’s performance for the three distinct classes on the supervised and ecological scenarios.

TABLE III  
*PLA-Tracker* COMPARED WITH STATE-OF-THE-ART SWALLOWING DETECTION SYSTEMS

Related work control level	Method	Subjects	Other PLA in recordings	Continuous recordings	Swallows	Protocol	Precision	Recall	Weighted Accuracy
High control	Khalifa et al. (2020) [7]	<b>248</b>	No	No	<b>3144</b>	Swallowing tasks 4 textures	85.3	83.8	-
	<i>PLA-Tracker</i>	42	<b>Yes</b>	<b>Yes</b>	1722	PLA tasks 5 textures (Sup)	<b>86.7</b>	<b>89.8</b>	-
Control	Gong et al. (2023) [13]	30	Yes	Yes	294	PLA tasks 2 textures	<b>95</b>	<b>91</b>	-
	<i>PLA-Tracker</i>	<b>42</b>	Yes	Yes	<b>1722</b>	PLA tasks 5 textures (Sup)	86.7	89.8	-
Ecological	Sazonov (2010) [10]	20	Yes	Yes	<b>9966</b>	Meal	-	-	84.7
	<i>PLA-Tracker</i>	<b>42</b>	Yes	Yes	1005	Meal (Eco)	-	-	<b>96.4</b>

Results from the CNN+LSTM model, without the depth-wise aspect, were included to demonstrate its impact. In the CNN+LSTM version, the four sources (three axes of the accelerometer and the microphone) were directly inputted into the convolutional layers as four channels.

TABLE IV  
SYSTEM COMPARISON ON THE SUPERVISED (SUP) SCENARIO AND SCORES ON SUP, ECOLOGICAL (ECO) AND BOTH SCENARIOS.

Model	Train*	Test*	F <sub>score</sub>		
			S	P	ADM
CNN+LSTM	Sup	Sup	83.2	91.7	82.7
CNNdw+LSTM ( <i>PLA-Tracker</i> )			<b>87.7</b>	<b>92.0</b>	<b>91.5</b>
<i>PLA-Tracker</i>	ALL	Sup	88.4	93.0	90.2
		Eco	87.9	92.5	65.5
		ALL	<b>88.2</b>	<b>92.7</b>	<b>85.7</b>

\*Each experiment uses the 14-fold cross validation with train, validation, and test sets and ensures a speaker independent system.

The CNNdw+LSTM model surpasses the CNN+LSTM version across all classes in the supervised scenario. This suggests that isolating sensors for convolution operations might play a crucial role in preserving or generating information, in contrast to the global convolution approach. Further investigation is necessary to evaluate the effects of both structural approaches within a multimodal context.

Although the system effectively detects most events, it tends to over-predict. For instance, it achieves a recall of 89.8% and a precision of 86.7% for the swallowing class on the supervised scenario (see Table III). Swallowing false alarms primarily occur during chewing or when subjects place food in their mouths. In both cases, impulsive activities manifest simultaneously in both sensor signals, contributing to event confusion. Addressing this issue may involve augmenting these specific noisy events in the training set or implementing signal filtering techniques. Missed swallows, on the other hand, are predominantly associated with very low-energy swallows, such as spontaneous saliva swallows. Work on spectrogram image enhancement or augmentations to train the

system in artificially faded events could lead to recall increase.

Overall, the system achieves an F<sub>score</sub> above 85% for all classes except the airway defense mechanisms in the ecological scenario. This is mostly due to the scarcity of these events (only 24 instances) and 13 false alarms, which significantly impact precision despite an average false alarm rate of approximately 1 every 10 minutes. Most airway defense mechanisms false alarms stem from confusion with voiced laughter, generating impulsive and energetic signals. In the larger global corpus, the score increases to 85.7%. Other classes are adequately represented in both scenarios (see table II), indicating that the obtained scores on these classes should remain consistent when applied to new data.

### B. Comparison with related work

The comparison was categorized based on the level of control within the recording protocol, as illustrated in table III, with the most pertinent studies selected for each category. Most studies on swallowing detection in HRCA signals are conducted exclusively with patients. However, it is also essential to demonstrate the effective functionality of this task in healthy subjects [9]. Khalifa et al. [7] and Gong et al. [13] achieved the best results in swallowing detection for the patient population. Khalifa’s study analyzed short recording intervals containing only requested swallows. Their recordings were made simultaneously with videofluoroscopy, which must be activated for the shortest possible time around the swallowing event to minimize radiation exposure to the patient. Thus, their conditions were highly controlled. *PLA-Tracker* outperforms their results under more adverse conditions. Gong et al.’s study focused on continuous recordings, similar to the supervised scenario presented in this work. In their approach, patients were administered specific food boluses sequentially, with an operator simultaneously marking the timestamps of swallows. It appears that only the requested swallowing activities were annotated. A direct comparison of their findings with ours is challenging due to the prevalence of spontaneous events in our supervised scenario (see table II), where spontaneous swallows occur nearly as frequently as controlled ones. Nev-

ertheless, Gong et al. reported excellent results, achieving a precision of 95% and a recall of 91% under these conditions. In our supervised scenario, the recall rate stood at 89.8%, with controlled swallows detected at a rate of 94.7% and spontaneous swallows at 88.3%. Therefore, our results are comparably close to those achieved by Gong et al. under analogous conditions. Finally, the study by Sazonov et al. [10] is the only one that enables comparison in complete ecological conditions for this swallowing detection task. The weighted accuracy was defined in their work as:

$$\text{Weighted Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

with TP, FP as true and false positive and TN, FN as true and false negative.

Calculations of the weighted accuracy for the *PLA-Tracker* in the ecological scenario surpassed the previously reported results (see table III). No significant difference in performance was observed between the supervised and ecological scenarios, contrary to their findings. This discrepancy could be attributed to the inclusion of a substantial number of spontaneous events in the supervised scenario, making it closely resemble an ecological setting. Consequently, the algorithm demonstrates stable detection of swallowing across various settings, marking a significant contribution to the field.

Therefore, when applied to continuous recording conditions, the *PLA-Tracker* outperforms previous studies for swallow detection but also integrates additional pertinent information into the analysis. In the context of dysphagia diagnosis, the relationship between airway defensive mechanisms and swallows, as well as the post-swallow voice characteristics, are highly relevant indicators. Obtaining an overview of activities in a long recording will allow an initial dysphagia assessment even without further characterization.

## V. CONCLUSION

This work introduces a new method to structure HRCA signals. Within ecological conditions, this method outperforms the swallow detection state-of-the-art on the healthy population reaching an  $F_{\text{score}}$  of 88%. Additionally, it extends its utility by successfully identifying two other relevant activities for dysphagia assessment, yielding  $F_{\text{scores}}$  of 93% for phonation and 86% for airway defense mechanisms when evaluated on the seven-hour corpus. While the precision of the system could be enhanced through the use of augmentation techniques, the *PLA-Tracker* currently extracts relevant information from the HRCA signals to assist clinicians in their analysis.

## REFERENCES

- [1] Jeri A. Logemann, "Dysphagia: Evaluation and Treatment," *FPL*, vol. 47, no. 3, pp. 140–164, 1995, Publisher: Karger Publishers.
- [2] Olga Birchall, Michelle Bennett, Nadine Lawson, Susan M. Cotton, and Adam P. Vogel, "Instrumental Swallowing Assessment in Adults in Residential Aged Care Homes: A Scoping Review," *Journal of the American Medical Directors Association*, vol. 22, no. 2, pp. 372–379.e6, Feb. 2021.
- [3] E. Sejdic, C. M. Steele, and T. Chau, "Classification of Penetration–Aspiration Versus Healthy Swallows Using Dual-Axis Swallowing Accelerometer Signals in Dysphagic Subjects," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1859–1866, July 2013.
- [4] Joshua M. Dudik, James L. Coyle, Amro El-Jaroudi, Zhi-Hong Mao, Mingui Sun, and Ervin Sejdic, "Deep learning for classification of normal swallows in adults," *Neurocomputing*, vol. 285, pp. 1–9, Apr. 2018.
- [5] Cara Donohue, Shitong Mao, Ervin Sejdic, and James L. Coyle, "Tracking Hyoid Bone Displacement During Swallowing Without Videofluoroscopy Using Machine Learning of Vibratory Signals," *Dysphagia*, vol. 36, no. 2, pp. 259–269, Apr. 2021.
- [6] Yassin Khalifa, Cara Donohue, James L. Coyle, and Ervin Sejdic, "Upper Esophageal Sphincter Opening Segmentation With Convolutional Recurrent Neural Networks in High Resolution Cervical Auscultation," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 2, pp. 493–503, Feb. 2021.
- [7] Yassin Khalifa, James L. Coyle, and Ervin Sejdic, "Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings," *Sci Rep*, vol. 10, no. 1, pp. 8704, Dec. 2020.
- [8] Naomi Kuramoto, Kazuhiro Ichimura, Dushyantha Jayatilake, Tomoya Shimokakimoto, Kikue Hidaka, and Kenji Suzuki, "Deep Learning-Based Swallowing Monitor for Realtime Detection of Swallow Duration," in *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, July 2020, pp. 4365–4368, IEEE.
- [9] Bryan Pak-Hei So, Tim Tin-Chun Chan, Liangchao Liu, Calvin Chi-Kong Yip, Hyo-Jung Lim, Wing-Kai Lam, Duo Wai-Chi Wong, Daphne Sze Ki Cheung, and James Chung-Wai Cheung, "Swallow Detection with Acoustics and Accelerometric-Based Wearable Technology: A Scoping Review," *IJERPH*, vol. 20, no. 1, pp. 170, Dec. 2022.
- [10] Edward S. Sazonov, Oleksandr Makeyev, Stephanie Schuckers, Paulo Lopez-Meyer, Edward L. Melanson, and Michael R. Neuman, "Automatic Detection of Swallowing Events by Acoustical Means for Applications of Monitoring of Ingestive Behavior," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 626–633, Mar. 2010.
- [11] James A. Curtis and Michelle S. Troche, "Handheld Cough Testing: A Novel Tool for Cough Assessment and Dysphagia Screening," *Dysphagia*, vol. 35, no. 6, pp. 993–1000, Dec. 2020.
- [12] Shanmugam Murugappan, Suzanne Boyce, Sid Khosla, Lisa Kelchner, and Ephraim Gutmark, "Acoustic characteristics of phonation in "wet voice" conditions," *The Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2578–2589, Apr. 2010.
- [13] Yanbin Gong, Jia Qiao, Yandao Huang, Qian Zhang, and Zulin Dou, "Mealcoach: Contact Microphone-Based Meal Supervision For Post-Stroke Dysphagia Patients," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Sydney, Australia, July 2023, pp. 1–5, IEEE.
- [14] Koji Takahashi, Michael E. Groher, and Ken-ichi Michi, "Methodology for detecting swallowing sounds," *Dysphagia*, vol. 9, no. 1, 1994.
- [15] Khondaker A. Mamun, Catriona M. Steele, and Tom Chau, "Swallowing accelerometry signal feature variations with sensor displacement," *Medical Engineering & Physics*, vol. 37, no. 7, pp. 665–673, July 2015.
- [16] Julie A. Y. Cichero, Peter Lam, Catriona M. Steele, Ben Hanson, Jianshe Chen, Roberto O. Dantas, Janice Duivestain, Jun Kayashita, Caroline Lecko, Joseph Murray, Mershen Pillay, Luis Riquelme, and Soenke Stanschus, "Development of International Terminology and Definitions for Texture-Modified Foods and Thickened Fluids Used in Dysphagia Management: The IDDSI Framework," *Dysphagia*, vol. 32, no. 2, pp. 293–314, Apr. 2017.
- [17] Sylvain Morinière, Michèle Boiron, Daniel Alison, Pascal Makris, and Patrice Beutter, "Origin of the Sound Components During Pharyngeal Swallowing in Normal Subjects," *Dysphagia*, vol. 23, no. 3, pp. 267–273, Sept. 2008.
- [18] Julie A. Y. Cichero and Bruce E. Murdoch, "Acoustic Signature of the Normal Swallow: Characterization by Age, Gender, and Bolus Volume," *Ann Otol Rhinol Laryngol*, vol. 111, no. 7, pp. 623–632, July 2002.
- [19] Joshua M. Dudik, Iva Jestrović, Bo Luan, James L. Coyle, and Ervin Sejdic, "A comparative analysis of swallowing accelerometry and sounds during saliva swallows," *BioMedical Engineering OnLine*, vol. 14, no. 1, pp. 3, Jan. 2015.