



Using an Automated Speech Recognition Approach to Differentiate Between Normal and Aspirating Swallowing Sounds Recorded from Digital Cervical Auscultation in Children

Thuy T. Frakking^{1,2,3} · Anne B. Chang^{4,5,6} · Christopher Carty^{1,7} · Jade Newing⁸ · Kelly A. Weir^{9,10} · Belinda Schwerin⁸ · Stephen So⁸

Received: 22 July 2021 / Accepted: 19 January 2022 / Published online: 29 January 2022
© Crown 2022

Abstract

Use of machine learning to accurately detect aspirating swallowing sounds in children is an evolving field. Previously reported classifiers for the detection of aspirating swallowing sounds in children have reported sensitivities between 79 and 89%. This study aimed to investigate the accuracy of using an automatic speaker recognition approach to differentiate between normal and aspirating swallowing sounds recorded from digital cervical auscultation in children. We analysed 106 normal swallows from 23 healthy children (median 13 months; 52.1% male) and 18 aspirating swallows from 18 children (median 10.5 months; 61.1% male) who underwent concurrent videofluoroscopic swallow studies with digital cervical auscultation. All swallowing sounds were on thin fluids. A support vector machine classifier with a polynomial kernel was trained on feature vectors that comprised the mean and standard deviation of spectral subband centroids extracted from each swallowing sound in the training set. The trained support vector machine was then used to classify swallowing sounds in the test set. We found high accuracy in the differentiation of aspirating and normal swallowing sounds with 98% overall accuracy. Sensitivity for the detection of aspiration and normal swallowing sounds were 89% and 100%, respectively. There were consistent differences in time, power spectral density and spectral subband centroid features between aspirating and normal swallowing sounds in children. This study provides preliminary research evidence that aspirating and normal swallowing sounds in children can be differentiated accurately using machine learning techniques.

Keywords Cervical auscultation · Deglutition · Swallowing sounds · Classifier · Machine learning · Aspiration

✉ Thuy T. Frakking
thuy.frakking@health.qld.gov.au

¹ Research Development Unit, Caboolture Hospital, Metro North Hospital & Health Service, McKean St, Caboolture, QLD 4510, Australia

² Centre for Clinical Research, School of Medicine, The University of Queensland, Herston, QLD 4029, Australia

³ Speech Pathology Department, Gold Coast University Hospital, Gold Coast Hospital & Health Service, 1 Hospital Boulevard, Southport, QLD 4215, Australia

⁴ Department of Respiratory Medicine, Queensland Children's Hospital, 501 Stanley St, South Brisbane, QLD 4101, Australia

⁵ Child Health Division, Menzies School of Health Research, Charles Darwin University, PO Box 41096, Casuarina, NT 0811, Australia

⁶ Australian Centre for Health Services Innovation, Queensland University of Technology, Level 7, 62 Graham St, South Brisbane, QLD 4101, Australia

⁷ Griffith Centre of Biomedical and Rehabilitation Engineering, Menzies Health Institute Queensland, Griffith University, Gold Coast 4222, Australia

⁸ School of Engineering and Built Environment, Griffith University, Parklands Dr, Southport, QLD 4215, Australia

⁹ Menzies Health Institute QLD & School of Health Sciences & Social Work, Griffith University, Gold Coast Campus, 1 Parklands Avenue, Southport, QLD 4222, Australia

¹⁰ Allied Health Research, Gold Coast University Hospital, Gold Coast Hospital & Health Service, 1 Hospital Boulevard, Southport, QLD 4215, Australia

Introduction

Up to one third of children with paediatric feeding disorders have identified oropharyngeal aspiration (abbreviated to aspiration), where fluids, foods and/or saliva enter the trachea below the level of the true vocal cords pre-, during and/or post-swallowing [1]. Clinical signs and symptoms of aspiration can be overt (e.g. reflexive cough generated to expel aspirated material) or considered silent when there is an absence of cough within 20 s of the aspiration event [2]. The prevalence of silent aspiration in infants and children is between 80 and 89% in children with paediatric feeding disorders [3, 4]. The accurate detection of aspiration, including silent aspiration, is important in infants and children because missed detection of aspiration can lead to acute and chronic lung sequelae [5–7], reduced nutritional growth and development [8] and decreased caregiver health-related quality of life [9].

In current clinical practice, a videofluoroscopic swallow study (VFSS) is the preferred instrumental assessment for aspiration in infants and children due to the ability to visualise all phases of the swallow, including direct visualisation of aspiration events [10]. However, VFSS is not readily available, involves exposure to ionising radiation (although at safe levels) [11–13] and may not replicate typical mealtimes due to time constraints associated with minimisation of screening time [14], observations of eating and drinking in an unfamiliar environment and difficulties replicating typical infant formula or breastmilk with barium impregnated fluids [15, 16]. Paediatric VFSS is also limited by accessibility due to the high costs associated with medical imaging equipment and the requirement of multiple health professionals specifically trained in conducting and interpreting paediatric VFSS [14]. As such, cervical auscultation has the potential to complement instrumental assessment and facilitate assessments which are more representative of typical mealtimes for infants and children.

Cervical auscultation (CA) is the most commonly used adjuvant to the clinical feeding evaluation across the United Kingdom, Ireland and Australia [17, 18]. CA is a repeatable, non-invasive technique that uses a stethoscope, digital accelerometer or digital microphone to capture swallow and breath sounds generated during the oral preparatory, oral and pharyngeal phases of swallowing [19, 20]. The use of CA has demonstrated high sensitivity 0.85 (95% CI 0.62–0.97) and negative predictor value 0.92 (95% CI 0.78–0.98), as well as good to very good reliability (inter-rater kappa = 0.81; 95% CI 0.79–0.84, intra-rater kappa range 0.72–0.98) in detecting aspirating swallows [21, 22]. Recent studies have also shown that specific sound features of swallow [23] and post-swallow

breath [19] sounds can accurately differentiate between aspirating and non-aspirating swallows in children based on digital cervical auscultation. In recent years, machine learning has gained popularity in health care for diagnostics and outcome prediction. Machine learning involves using statistical algorithms to model underlying relationships between variables (or features), to make predictions. For diagnostic applications, supervised machine learning approaches are most common because the algorithms are trained to classify the presences or absence of disease or dysfunction. Support Vector Machine (SVM) algorithms have gained popularity in the machine learning community as they can provide accurate predictions when the relationship between the features and the outcome are non-linear. SVM classifiers are considered the most appropriate for use in sound classification, given its non-linear ability to differentiate binary classes (i.e. aspiration versus non-aspiration) using feature selection techniques on smaller datasets [24–26]. To date, radial basis classifiers have been shown to accurately detect aspirating swallows in children with paediatric feeding disorders related to neurological conditions (sensitivity of 79.4% [27] and 92.2% [28]). These studies [27, 28] demonstrated that using machine learning can accurately classify aspiration in children but 100% diagnostic test accuracy remains elusive.

Previous studies using machine learning are limited by several factors including their use of pre-processed sounds and/or pre-selected input features to classify aspirating swallows. Pre-processed sounds refer to the process of removing contaminating signal components which are not related to swallowing sounds. This may include signal (e.g. cough, breath sounds) or disturbance noises (e.g. vasomotion of major arteries, head movement) [29]. Using pre-selected mathematical or physiological features to classify aspirating swallow sounds in adults likely resulted in unintentional bias and reduced accuracy of the classifier of previous CA studies [30]. Also, previous classifier accuracy results were based on swallowing sounds collected from children with large age ranges (2–11.6 years [27] and 2–9.9 years [28]) and on a variety of fluid viscosities. These are limitations as age and viscosity influences swallowing kinematics and acoustic swallowing parameters of amplitude, frequency and duration for children [31–33]. To minimise confounding factors of age and viscosity on swallowing sounds, studying swallowing sounds in children of a narrow age range and on single viscosities are necessary. Additionally, symptoms of paediatric swallowing disorders can be heterogeneous and nonspecific to underlying medical aetiology; therefore, an approach which can accommodate such heterogeneity whilst detecting key attributes of aspiration swallowing sound features has the potential to increase the accuracy in detecting aspiration amongst infants and children.

Automatic speaker recognition (ASR), an approach which aims to identify a human speaker from a digital recording of their voice, may provide increased accuracies to the detection of aspirating swallow sounds in children. In conventional ASR, the speech signal is represented as temporal changes in the shape of the human vocal tract, which contains speaker-specific information that can be used to identify a person. Engineering models of the vocal tract shape mathematically represent the resonant frequencies as the spectral envelope of the power spectral density of the speech signal [34]. The frequency locations of strong spectral peaks or formants, which represent resonances in the vocal tract, convey information on the speech content as well as the speaker characteristics [35]. Based on a similar premise, aspirating swallow sounds are the result of fluid flow into the trachea, below the level of the vocal cords. It is therefore reasonable to hypothesise that formant-based speech features, such as LPCCs (or linear prediction cepstral coefficients) [36], MFCCs (or Mel-frequency warped cepstral coefficients) [37] and PLPs (or perceptual linear prediction) [38], could convey useful information for discriminating aspirating from non-aspirating swallowing sounds using machine learning techniques. To date, CA has differentiated between safe and unsafe swallows [39], swallow function [40] and correlated specific swallow kinematic events such as hyoid bone displacement [41–43] and laryngeal closure [44] in adults. Given that ASR collates signals based on the temporal changes in the shape of the human vocal tract and the promising correlates of CA to specific swallow kinematics in adults, the differentiation of aspirating from non-aspirating swallowing sounds may be possible. This study aimed to investigate the accuracy of using an ASR approach to differentiate between normal and aspirating swallowing sounds recorded from digital cervical auscultation in children.

Methods

This study was approved by a Human Research Ethics Committee. Data analyses were completed on two groups of children where participants were prospectively recruited and informed consent obtained from the caregivers and assent from older children.

Participants

We had two groups of children: (i) typically developing and (ii) feeding disorders.

Group (i) was recruited from the generally community, median age 13 months (range 4–33 months, 52.1% males). Their inclusion criteria were as follows: aged 4–36 months and confirmation of normal oral feeding development via a Pre-Feeding Checklist [45] for infants aged 4–7 months

or confirmation of normal oromotor functioning via the Schedule for Oral Motor Assessment (SOMA) [46] for children aged 8–36 months. Children were excluded if they had a medical history of any of the following: developmental delay, visual/hearing impairment, neurological impairment, aerodigestive tract structural abnormalities, genetic syndromes, paediatric feeding disorder, neurodevelopmental disorder and/or prematurity (< 37 weeks of gestation) [31]. Children included in this study were chosen based on closest match in age (months) to Group (ii).

Children from group (ii) were aged 2–71 months (median age 10.5 months, 61.1% males). Diagnoses included congenital syndromes (e.g. Beckwith-Wiedemann, Cru de Chat, Pierre Robin Sequence), neurological (e.g. cerebral palsy), respiratory (e.g. bronchiectasis, chronic cough) anatomical anomalies (e.g. oesophageal atresia, congenital myopathy, tracheoesophageal fistula) and other (e.g. developmental delays, failure to thrive).

Procedure

Normal Swallows from Typically Developing Children Group

This dataset consisted of 106 sound clips of thin fluid initial and subsequent swallows from 23 healthy infants/children. Depending on the children's preferences and gross motor development, children were seated upright at a small chair and table, in a high chair or positioned on their caregiver's lap. Children were fed by their caregiver, the researcher or allowed to independently feed themselves depending on their expressed preferences and/or fine motor skills development. Children were first offered three single bites or single sips of age-appropriate food/fluid consistencies in a standardised order of puree, lumpy mash, chewable solids and thin fluids. Bolus sizes were not standardised to allow children to eat and drink bolus volumes which replicated their typical mealtime experiences. After this, children were allowed to free eat and drink the remainder of food/fluids for a period of 30 min.

Aspirating Swallows from Children with Paediatric Feeding Disorders Group

The second dataset consisted of 18 thin fluid swallows from 18 patients with confirmed aspiration via VFSS with various underlying medical aetiologies. Depending on the children's age and level of gross motor development, children were seated either semi-reclined or upright on a tumbleform chair, which was positioned on top of a Videofluoroscopic Imaging Chair. Children were viewed in the lateral position and offered a standard VFSS protocol of two presentations of puree, lumpy mash, chewable solid, extremely thick, moderately thick, mildly thick, slightly thick and/or thin fluids by

either their caregiver or a speech pathologist. The order of the protocol was consistent for all children; however, adaptations to the type of foods presented were made based on individual medical status and feeding development of each participant. Conferral on the presence/absence of aspiration was jointly completed by a paediatric speech pathologist and radiologist; and objectively rated by a speech pathologist post-VFSS, using the Penetration–Aspiration Scale [47]. Scores of 6, 7, or 8 were considered aspirating swallows due to the entry of material below the level of the vocal cords. Only one swallow on thin fluids per patient was used in this study. Aspirating swallows used were chosen based on the absence of extraneous noises as a result of a loss of flat contact of the microphone resulting in capture of ambient room noise.

Equipment

An omnidirectional condenser microphone (C417, AKG Acoustics, Vienna, Austria) (sensitivity at 1 kHz of 10 mV/Pa, impedance 200, frequency range 20 to 20,000 Hz) was placed on the skin surface lateral (< 1 cm) to the cricoid cartilage aligned at the level of the 6th cervical posterior vertebrae (Fig. 1). Palpation of the cricoid cartilage was performed prior to microphone attachment to help guide accurate placement for all participants. To ensure the microphone was consistently placed, all were placed by the principal researcher only using a fitted circular O-ring and secured with microfoam tape. The use of a circular O-ring and microfoam tape maximises flat surface contact, reduces pick up of ambient room noise and allows the infants and children to move as they typically would during mealtimes without interfering on sound quality. All swallowing sounds were digitally recorded (Digital H4n Handy Recorder, Zoom Corporation, Tokyo, Japan) and were continuously monitored prior to and throughout the recordings using headphones

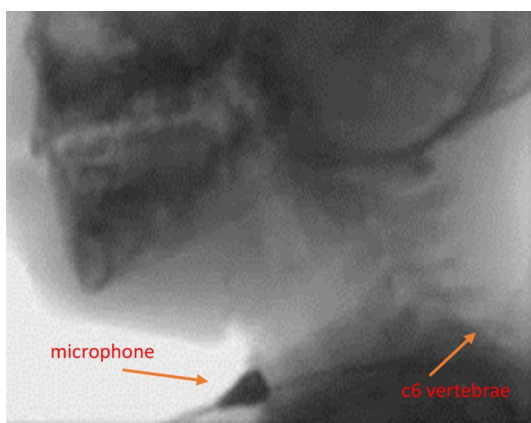


Fig. 1 Placement of microphone during a videofluoroscopic swallow study (VFSS)

(Model ATH-M50, Audio-Technica, Taiwan). For quality control, adjustments to the microphone placement and taping were made for the detection of extraneous noises (e.g. loss of flat contact resulting in capture of ambient room noise) and/or loss of accurate microphone placement (e.g. children pulling off the microphone).

Visual images of children drinking thin fluids were captured via a digital video-recorder (Model DCR-DVD605E, Sony Corporation, Tokyo, Japan) for the normal group, whilst VFSS images were captured at 15 frames per second via a digital fluoroscopy unit (Toshiba KXO-80G, North Ryde, NSW, Australia) for the aspiration group. Visual images for both groups were all simultaneously recorded on the Digital Swallowing Workstation (KayPentax, Pentax, New Jersey, USA). Where possible, nasal airflow direction was also simultaneously recorded via placement of an infant or paediatric sized nasal cannula, which was secured firmly behind the ears. A vocal signal was used at the beginning of all assessments to enable accurate synchronisation of acoustic and visual data. Visual and audio recordings of children drinking thin fluids were downloaded from the Digital Swallowing Workstation (KayPentax, Pentax, New Jersey, USA) onto an external hard drive and synchronised on video-editing software (Sony Vegas Movie Studio 9, Madison, WI). Manual segmentation for time points of normal swallows were completed by a research assistant and clinician researcher with speech pathology experience trained to identify the start and end points of the swallows. The start point of a swallow sound was defined as the commencement of a fluid flushing sound (audio data) and commencement of laryngeal motion associated with pharyngeal activity (visual data). The end point of a swallow was defined as the cessation of a fluid flushing sound (audio data), combined with no laryngeal motion (visual data). Inter-rater reliability for swallow segmentation was performed on a random selection of 25% of swallows and intra-class coefficients of > 0.99 were obtained for both raters [31, 48].

Preprocessing and Feature Extraction from Swallowing Sounds

An overview of the processing that was performed is summarised in Fig. 2. No downsampling was applied to the swallowing sounds to record the entire dynamic range. In the feature extraction stage, the swallow sounds were passed through the following pre-emphasis high-pass filter.

$$H(z) = 1 - 0.97z^{-1}$$

The purpose of the pre-emphasis filter was to compensate for any spectral tilt by flattening the power spectrum so that higher frequency formants were emphasised [36]. The filtered sounds were then segmented into 20 ms overlapping

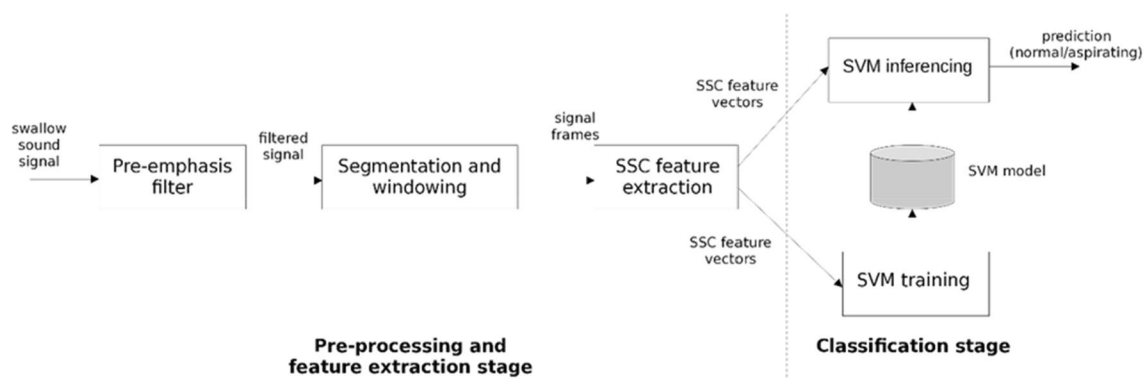


Fig. 2 Flowchart of the pre-processing, feature extraction, and classification stages

frames with a 10 ms update and a Hamming window was applied to each frame.

Feature vectors were then computed for each frame. There are several speech features that are commonly used in automatic speaker recognition, such as LPCCs, MFCCs and PLPs. The purpose of the feature extraction process is to derive useful parameters of the spectral envelope shape of the power spectrum, specifically, the frequency locations of the formants [35]. Whilst MFCCs are very popular features used in conventional ASR, they are also well known to be sensitive to additive noise. Early experiments as part of this study were performed using MFCCs to identify aspirating sounds; however, results showed relatively poor classification performance. Rather than using cepstral-based features, which utilise both formant and non-formant regions of the power spectrum, it was discovered that features based on formant locations, specifically, spectral subband centroids (SSCs) [49], were better suited for this task.

To compute SSCs, the power spectrum is filtered by a Mel-frequency warped triangular filterbank (similar to that used in MFCCs), which results in M subbands, where M is the number of triangular filters used. In this study, M was set to 26. The spectral centroid frequency of each subband is computed and these centroid frequencies together form a M -dimensional SSC feature vector. Spectral subband centroids tend to be more robust because they are more heavily influenced by the location of the formants, which can be less sensitive to broadband noise [49].

Based on features from a previous study on SVM classification of acoustic events [50], a feature vector for each patient is then formed by computing the mean and standard deviation of the SSC vectors and concatenating them together, giving the final feature dimension of 52.

SVM Classifier Design

There was a total of 124 swallows, of which 18 were aspirating and 106 normal swallows. Training and test sets

were formed using a 50/50% stratified random split, which ensured the same distribution of classes in both sets. The training and test set each comprised 53 normal and 9 aspirating swallows. Since SSCs are frequency locations that are required to be in ascending order, feature standardisation was not applied. The class labels for normal and aspirating swallows were set to 0 and 1, respectively.

The classifier used in this study was the support vector machine (or SVM) [51], which is well suited to situations where training data are limited. The SVM is a large margin classifier that has good generalisation properties and is relatively robust to outliers, since the computed hyperplane that separates the two classes is determined only by a smaller subset of training vectors known as the support vectors. By using different SVM kernel functions, which effectively transform the feature vectors into a higher dimensional vector space in order to achieve better class separation, a non-linear decision hyperplane can be efficiently computed [52].

The SVM classifier in the Python library scikit-learn [53] was utilised in the experiments for training and testing. In order to determine the optimal SVM parameters, such as the type of kernel (sigmoid, radial basis function or polynomial), kernel parameters (gamma and polynomial order) and regularisation parameter C , a grid search was performed using a five-fold cross-validation on the training set. In this study, the best parameters found were $C = 1$ using a second order polynomial kernel. To handle the class imbalance problem, the regularisation parameter C for each class was weighted by a factor that was inversely proportional to the class frequencies, as is implemented in the “balanced mode” of the SVC in scikit-learn.

Results

Confusion matrices summarises prediction results on a classification problem and often used in the field of machine learning. The confusion matrix of the normal (0) and

aspirating (1) classes for the SVM classifier are shown in Fig. 3. This matrix shows that of the true normal swallows, the classifier correctly identified 53 as normal swallows, and 0 as aspirating. Of the true aspirating swallows ($n=9$), the classifier labelled 1 as a normal swallow and 8 as aspirating swallows. The matrix demonstrates that the proposed method accurately identified all normal swallows with only one false negative, i.e. the classifier identified an aspirating swallow as normal. Table 1 lists the overall performance of the SVM classifier, where the total accuracy was found to be 98%.

Aspirating swallow sounds had different time domains and power spectral density (PSD) features compared to

True Label	Normal	53	0
	Aspirating	1	8
		Normal	Aspirating
		Predicted Label	

Fig. 3 Confusion matrix between normal and aspirating classes

Fig. 4 Time domain representations of and regions of interest (A, B, C) in **a** an aspirating swallow; and **b** a normal swallow

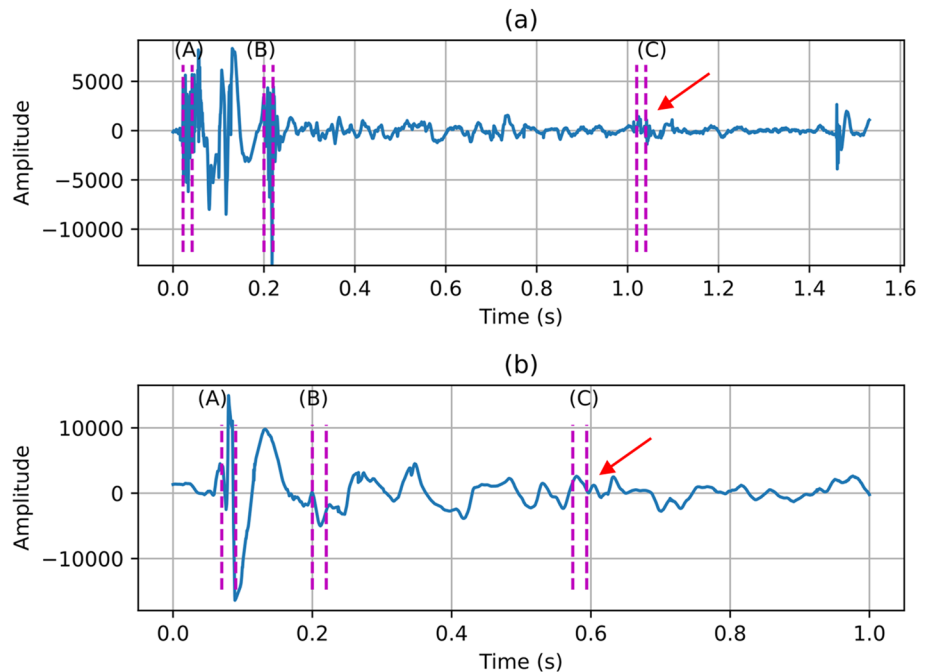


Table 1 SVM classifier performance between patients with aspirating and normal swallows (total accuracy = 0.98)

	PPV or precision	NPV	Specificity	Sensitivity or recall	F ₁ score
Normal swallow (0)	0.98	1.00	0.89	1.00	0.99
Aspirating swallow (1)	1.00	0.98	1.00	0.89	0.94

The F₁ score, which is defined as the harmonic mean of the precision and recall, is another measure of accuracy that is more conservative, balances the contribution of false negatives and false positives to the final metric, and is better suited to cases where classes are unbalanced. PPV positive predictive value, NPV negative predictive value

normal swallows. To demonstrate this difference, Fig. 4 shows randomly selected time domain segments (each with labels for three points of interest A, B and C) of aspirating swallows (subplot 2(a)) and normal swallows (subplot 2(b)). The corresponding PSD for both aspirating and normal segments and each point of interest A, B and C are shown in Figs. 5, 6 and 7. In Figs. 5, 6 and 7, the dashed red vertical lines indicate the location of the individual SSCs. Since the SSCs are computed as the centroid of different frequency subbands, they will tend to be more concentrated around strong spectral peaks. This higher concentration of SSCs can be seen in the low frequencies where the spectral components are strong. Therefore, the SSCs, which are highlighted in Figs. 5, 6 and 7, are able to capture the high-frequency spectral peak that is near the 20 kHz, which appears to be a

Fig. 5 Power spectral density after pre-emphasis (blue line) and SSC frequencies (dashed red lines) of region (A) from Fig. 4: **a** an aspirating swallow; and **b** a normal swallow

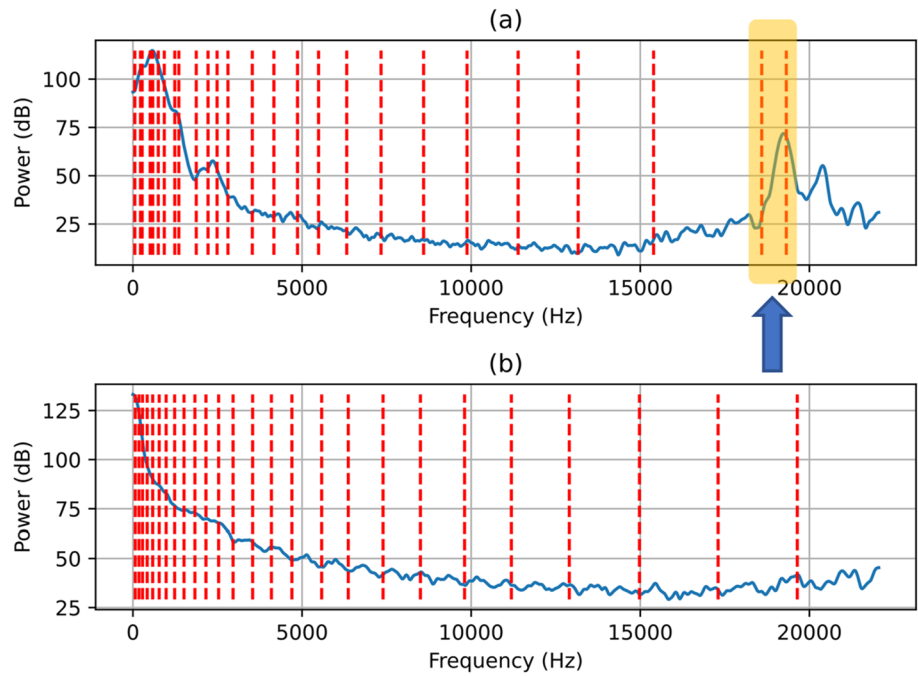
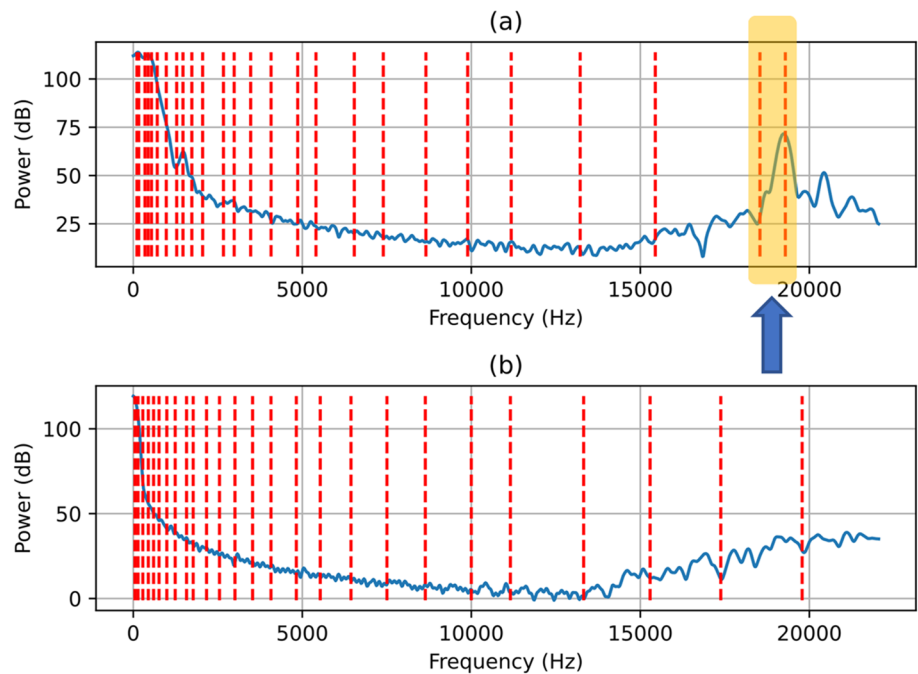


Fig. 6 Power spectral density after pre-emphasis (blue line) and SSC frequencies (dashed red lines) of region (B) from Fig. 4 in: **a** an aspirating swallow and **b** a normal swallow



characteristic of the power spectral density of the aspirating swallow sound.

In order to determine whether these characteristic SSCs are a common feature that can be used for the discrimination between normal and aspirating swallow sounds, Fig. 8a and b shows all the SSCs from the

entire aspirating and normal swallow sounds dataset. As shown in the highlighted region in Fig. 8a, there are high-frequency SSCs that only appear in aspirating swallow sounds.

Fig. 7 Power spectral density after pre-emphasis (blue line) and SSC frequencies (dashed red lines) of region (C) from Fig. 4 in: **a** an aspirating swallow and **b** a normal swallow

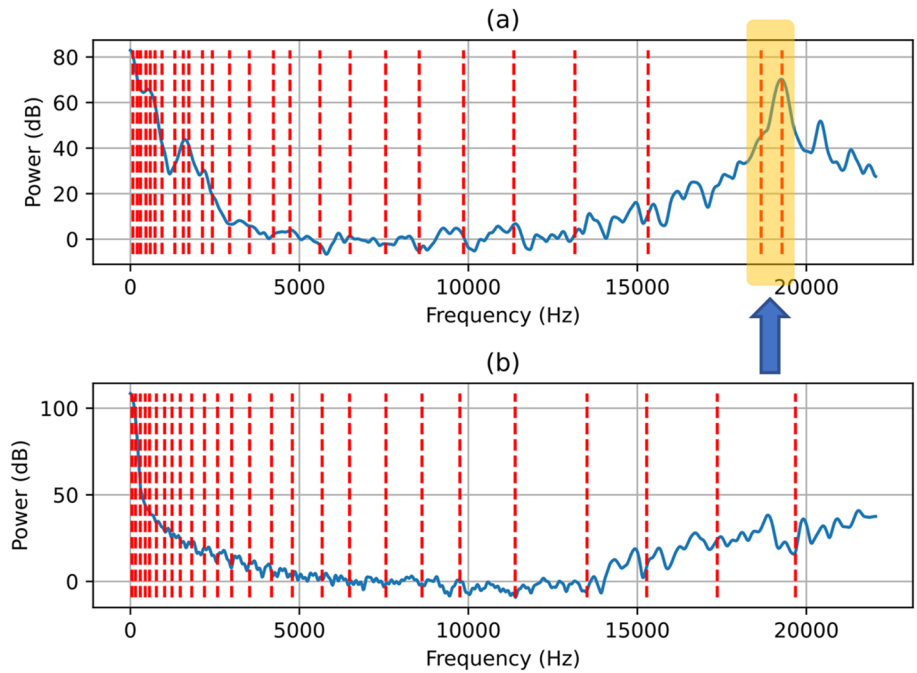
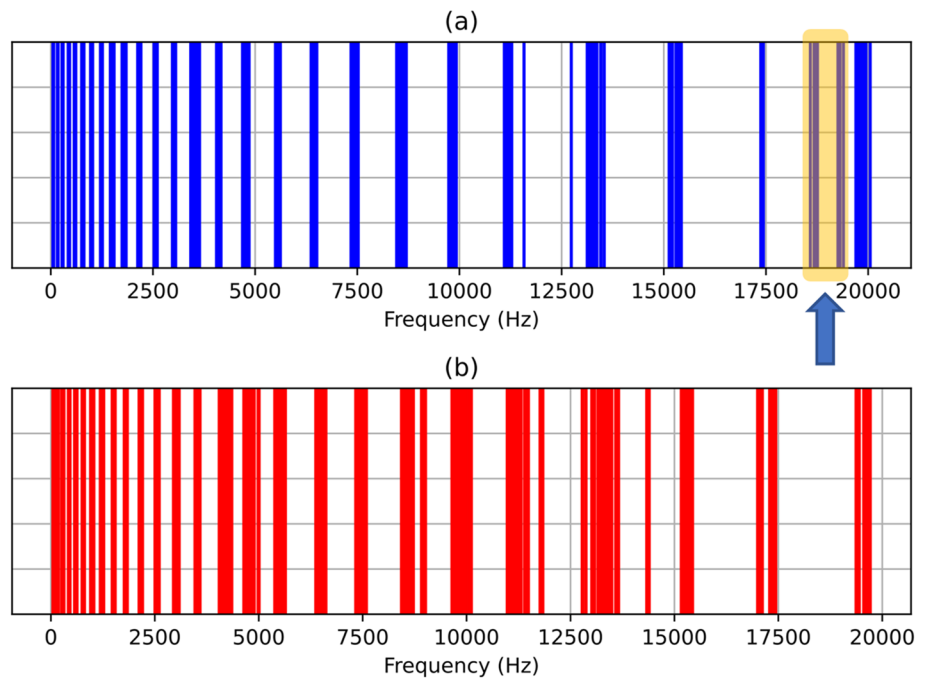


Fig. 8 Plot of mean SSCs for all **a** aspirating swallows (18) and **b** normal swallows (106). There are 26 “bands” that represent the 26 SSC features. Within each band, a vertical line represents the mean SSC for an individual swallow sound. The highlighted region in **a** shows SSCs that only appear in aspirating swallow sounds



Discussion

To our knowledge, this is the first study to use an ASR approach to determine a classifier for the differentiation between normal swallows in healthy children and aspirating swallows in a cohort of children with paediatric feeding disorders using digital CA. High accuracy for the detection of aspirating and normal swallows were found,

with sensitivities of 89% and 100%, and NPVs of 1.00 and 0.98, respectively. We demonstrated differences in time, power spectral density and SSC features between aspirating and normal swallows in children.

Previous studies investigating the use of a classifier to detect aspirating swallow sounds in children have documented lower overall accuracies of 79.8% [27] and 89.6% [28] when compared to the overall accuracy of 98% found in our study. It is possible that the lower accuracies previously

reported were at increased risk of type I errors due to multiple swallow sound data points collected and used from each participant with a paediatric feeding disorder. The same studies also used swallow sounds recorded via an accelerometer on a combined range of fluid viscosities and solid consistencies [27, 28]. Exclusively focussing on thin fluids is essential in the development of a swallowing sounds classifier because there are known differences in the acoustic properties (e.g. intensity, duration, frequency) between thin and puree bolus' in children [32] and adults [54, 55]. Thus, our higher reported accuracy may have been facilitated by exclusively using thin fluid swallowing sounds across two independent groups of healthy children and children with paediatric feeding disorders, thereby minimising type I error and the heterogeneity of swallowing sounds.

Our reported sensitivity of 89% for the detection of aspiration is higher than previously documented sensitivities of between 33 and 92% for routinely performed clinical feeding evaluations in children when compared with gold standard tests, such as the VFSS or fiberoptic endoscopic evaluation (FEES) [10, 56–63]. In comparison with cervical auscultation-specific diagnostic test accuracy data in children, our overall sensitivity for the differentiation between aspirating and normal swallows is superior to subjective clinician judgement when cervical auscultation was used in conjunction with the clinical feeding evaluation 85% (95% CI 0.62–0.97) [21] or in isolation 93.9% (95% CI 91.8–95.6) [22]. The improvements in accuracy for the differentiation between aspirating and normal swallows found in our study are likely attributed to the ASR approach used. Given that the ASR approach is based on temporal changes to the shape of the vocal tract, it is plausible that the entry of fluids into the trachea and below the level of the vocal cords caused changes to features of the swallowing sound which would not be audible on normal swallows.

In our study, consistent differences in time, power spectral density and SSC features were found between aspirating and normal swallows in children. For aspirating swallow sounds, the power spectral density at specific swallowing time points contained strong peaks in the high frequencies when compared to normal swallowing sounds. In addition, only a specific SSC feature appears in aspirating swallowing sounds and is absent from all normal swallowing sounds. We hypothesise that the SSC and power spectral density features on aspirating swallow sounds are one out of 21 physiologic components in the oropharyngeal mechanics of feeding/swallowing in infants [64, 65]. Martin-Harris and colleagues [65] completed the largest known prospective study on 300 infants which investigated the quantification of swallowing function. Based on the gold standard VFSS and validity testing, Martin-Harris and colleagues [65] demonstrated that aspiration formed one of five domains

of swallowing in infants. As such, the contributions of our preliminary work of SSC and power density features lay the foundation for future approaches to develop an aspiration classifier for swallowing sounds which correlate with quantifiable physiological components of the paediatric feeding/swallowing mechanism.

Whilst our study uniquely described the use of machine learning to accurately classify aspirating swallow sounds in children, there are limitations to our study. Firstly, VFSS was not used to objectively classify swallowing sounds obtained in the normal group of children. Our team acknowledge that obtaining ethical approval and caregiver consent for radiation exposure with VFSS for healthy infants and children are unlikely. Instead, we used the results of the SOMA [46] and/or pre-feeding checklists [45] to objectively categorise normal feeding/swallowing skills. Secondly, our small sample of aspirating swallows ($n=18$) was a limitation and there is the potential for overfitting based on this small specific dataset. To overcome any bias towards overfitting, we used stratification and class weightings in the SVM to preserve the class balance between aspirating and normal swallows. This ensured consistent proportions of the two swallow types used for machine learning. Future studies with larger sample sizes of aspirating swallows in children with paediatric feeding disorders are required to further validate use of the ASR approach for the classification of aspiration. In addition, comparative studies of different classifiers are required to provide reliable interpretation of reported diagnostic accuracies which are performed on the same dataset to reduce known bias' in machine learning within healthcare [66].

The potential benefits of an accurate classifier for the detection of aspiration based on swallowing sounds provide an important step towards cervical auscultation being a complementary diagnostic instrumental assessment to VFSS. Cervical auscultation has the potential capability of helping health professionals worldwide working in the field of paediatric feeding disorders to personalise feeding management plans which are patient and family centred [67] due to its non-invasive nature, low cost and repeatability in multiple environments. However, further studies are required which uses machine learning principles to systematically understand swallowing sound properties in a variety of clinical populations in infants and children. Speech pathologists continue to have an important role in the assessment of aspiration risk and the complex richness of information (e.g. breath sounds, respiratory status, positioning) that is obtained in a clinical feeding evaluation whilst further work on refining current swallowing classifiers continues.

Conclusion

This study provides preliminary evidence that the use of machine learning techniques could accurately classify aspirating swallowing sounds collected from digital cervical auscultation in children with a high degree of accuracy (98%), sensitivity (89%) and PPV (100%).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. TTF is supported by a Metro North Clinician Researcher Fellowship. ABC is supported by NHMRC Practitioner Fellowship (no. 1058213).

Declarations

Competing interests None declared.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Weir K, McMahon S, Barry L, Ware R, Masters IB, Chang AB. Oropharyngeal aspiration and pneumonia in children. *Pediatr Pulmonol.* 2007;42(11):1024–31.
- Weir K, McMahon S, Barry L, Masters IB, Chang AB. Clinical signs and symptoms of oropharyngeal aspiration and dysphagia in children. *Eur Respir J.* 2009;33(3):604–11.
- Weir K, McMahon S, Taylor S, Chang AB. Oropharyngeal aspiration and silent aspiration in children. *Chest.* 2011;140(3):589–97.
- Velayutham P, Irace AL, Kawai K, et al. Silent aspiration: Who is at risk? *Laryngoscope.* 2017.
- Boesch RP, Daines C, Willging JP, et al. Advances in the diagnosis and management of chronic pulmonary aspiration in children. *Eur Respir J.* 2006;28(4):847–61.
- Dodrill P, Gosa M. Pediatric dysphagia: physiology, assessment, and management. *Ann Nutr Metab.* 2015;66:24–31.
- Tutor JD, Srinivasan S, Gosa MM, Spentzas T, Stokes DC. Pulmonary function in infants with swallowing dysfunction. *PLoS ONE.* 2015;10(5):e0123125.
- Bell KL, Benfer KA, Ware RS, et al. Development and validation of a screening tool for feeding/swallowing difficulties and undernutrition in children with cerebral palsy. *Dev Med Child Neurol.* 2019;61(10):1175–81.
- Lefton-Greif MA, Okelo SO, Wright JM, Collaco JM, McGrath-Morrow SA, Eakin MN. Impact of children's feeding/swallowing problems: validation of a new caregiver instrument. *Dysphagia.* 2014;29(6):671–7.
- Arvedson JC. Feeding children with cerebral palsy and swallowing difficulties. *Eur J Clin Nutr.* 2013;67(Suppl 2):S9–12.
- Hersh C, Wentland C, Sally S, et al. Radiation exposure from videofluoroscopic swallow studies in children with type 1 laryngeal cleft and pharyngeal dysphagia: a retrospective review. *Int J Pediatr Otorhinolaryngol.* 2016;89:92–6.
- Im HW, Kim SY, Oh B-M, Han TR, Seo HG. Radiation dose during videofluoroscopic swallowing studies and associated factors in pediatric patients. *Dysphagia.* 2019.
- Layly J, Marmouset F, Chassagnon G, et al. Can we reduce frame rate to 15 images per second in pediatric videofluoroscopic swallow studies? *Dysphagia.* 2019:1–5.
- Batchelor G, McNaughten B, Bourke T, Dick J, Leonard C, Thompson A. How to use the videofluoroscopy swallow study in paediatric practice. *Arch Dis Child Educ Pract Ed.* 2019;104(6):313.
- Cichero JAY, Nicholson T, Dodrill P. Liquid barium is not representative of infant formula: characterisation of rheological and material properties. *Dysphagia.* 2011;26:264–71.
- Frazier J, Chestnut AH, Jackson A, Barbon CEA, Steele CM, Pickler L. Understanding the viscosity of liquids used in infant dysphagia management. *Dysphagia.* 2016;31(5):672–9.
- Bateman C, Leslie P, Drinnan MJ. Adult dysphagia assessment in the UK and Ireland: are SLTs assessing the same factors? *Dysphagia.* 2007;22(3):174–86.
- Rumbach A, Coombes C, Doeltgen S. A survey of Australian dysphagia practice patterns. *Dysphagia.* 2018;33(2):216–26.
- Almeida STD, Ferlin EL, Maciel AC, et al. Acoustic signal of silent tracheal aspiration in children with oropharyngeal dysphagia. *Logop Phoniatr Vocol.* 2018;43(4):169–74.
- Frakking TT, Chang AB, O'Grady KA, Walker-Smith K, Weir KA. Cervical auscultation in the diagnosis of oropharyngeal aspiration in children: a study protocol for a randomised controlled trial. *Trials.* 2013;14(377):377.
- Frakking TT, Chang AB, O'Grady KF, David M, Walker-Smith K, Weir KA. The use of cervical auscultation to predict oropharyngeal aspiration in children: a randomized controlled trial. *Dysphagia.* 2016;31(6):738–48.
- Frakking TT, Chang AB, O'Grady KF, David M, Weir KA. Reliability for detecting oropharyngeal aspiration in children using cervical auscultation. *Int J Speech Lang Pathol.* 2017;19(6):569–77.
- Frakking T, Chang A, O'Grady K, David M, Weir K. Aspirating and nonaspirating swallow sounds in children: a pilot study. *Ann Otol Rhinol Laryngol.* 2016;125(12):1001–9.
- Vieira AMS. *Machine learning.* Cambridge: Academic Press; 2019.
- Miyagi S, Sugiyama S, Kozawa K, Moritani S, Sakamoto SI, Sakai O. Classifying dysphagic swallowing sounds with support vector machines. *Healthcare.* 2020;8(2):103.
- Sidey-Gibbons JAM, Sidey-Gibbons CJ. *Machine learning in medicine: a practical introduction.* BMC Med Res Methodol. 2019;19(1):64.
- Lee J, Blain S, Casas M, Kenny D, Berall G, Chau T. A radial basis classifier for the automatic detection of aspiration in children with dysphagia. *J Neuroeng Rehabil.* 2006;3:14.
- Merey C, Kushki A, Sejdić E, Berall G, Chau T. Quantitative classification of pediatric swallowing through accelerometry. *J Neuroeng Rehabil.* 2012;9(1):34–34.
- Coyle JL, Sejdić E. High-resolution cervical auscultation and data science: new tools to address an old problem. *Am J Speech Lang Pathol.* 2020;29(2S):992–1000.
- Dudik JM, Coyle JL, El-Jaroudi A, Mao Z-H, Sun M, Sejdić E. Deep learning for classification of normal swallows in adults. *Neurocomputing.* 2018;285:1–9.

31. Frakking TT, Chang AB, O'Grady KF, Yang J, David M, Weir KA. Acoustic and perceptual profiles of swallowing sounds in children: normative data for 4–36 months from a cross-sectional study cohort. *Dysphagia*. 2017;32(2):261–70.
32. Hennessey NW, Fisher G, Ciccone N. Developmental changes in pharyngeal swallowing acoustics: a comparison of adults and children. *Logoped Phoniatr Vocol*. 2018;43(2):63–72.
33. Steele CM, Alsanei WA, Ayanikalath S, et al. The influence of food texture and liquid consistency modification on swallowing physiology and function: a systematic review. *Dysphagia*. 2015;30(2):272–3. <https://doi.org/10.1007/s00455-014-9578-x>.
34. Atal BS, Hanauer SL. Speech analysis and synthesis by linear prediction of speech wave. *J Acoust Soc Am*. 1971;50(2):637–55.
35. Rabiner LR, Schafer RW. Theory and applications of digital speech processing. 1st ed. Upper Saddle River: Pearson; 2011.
36. Rabiner LR, Juang BH. Fundamentals of speech recognition. Englewood Cliffs, NJ: PTR Prentice Hall; 1993.
37. Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans Acoust Speech Signal Process*. 1980;28(4):357–66.
38. Hermansky H. Perceptual linear predictive (PLP) analysis of speech. *J Acoust Soc Am*. 1990;87(4):1738–52.
39. Dudik JM, Kurosu A, Coyle JL, Sejdić E. A statistical analysis of cervical auscultation signals from adults with unsafe airway protection. *J Neuroeng Rehabil*. 2016;13(1):7–7.
40. Donohue C, Khalifa Y, Perera S, Sejdic E, Coyle JL. A Preliminary investigation of whether HRCA signals can differentiate between swallows from healthy people and swallows from people with neurodegenerative diseases. *Dysphagia*. 2020.
41. Donohue C, Mao S, Sejdic E, Coyle JL. Tracking hyoid bone displacement during swallowing without videofluoroscopy using machine learning of vibratory signals. *Dysphagia*. 2020.
42. He Q, Perera S, Khalifa Y, et al. The association of high resolution cervical auscultation signal features with hyoid bone displacement during swallowing. *IEEE Trans Neural Syst Rehabil Eng*. 2019;27(9):1810–6.
43. Rebrion C, Zhang Z, Khalifa Y, et al. High-resolution cervical auscultation signal features reflect vertical and horizontal displacements of the hyoid bone during swallowing. *IEEE J Transl Eng Health Med*. 2019;7:1–9.
44. Mao S, Sabry A, Khalifa Y, Coyle JL, Sejdic E. Estimation of laryngeal closure duration during swallowing without invasive X-rays. *Future Gener Comput Syst*. 2021;115:610–8.
45. Morris SE, Klein MD. Pre-feeding skills: a comprehensive resource for mealtime management. Pro-Ed; 2000.
46. Skuse D, Stevenson J, Reilly S, Mathisen B. Schedule for oral-motor assessment (SOMA): methods of validation. *Dysphagia*. 1995;10:192–202.
47. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. *Dysphagia*. 1996;11(2):93–8.
48. Rousson V. Assessing inter-rater reliability when the raters are fixed: two concepts and two estimates. *Biom J*. 2011;53(3):477–90.
49. Paliwal KK. Proceedings of the 1998 IEEE international conference on acoustics, speech and signal processing, ICASSP '98 (Cat. No.98CH36181), vol. 612. Vol 2: IEEE; 1998:617–20.
50. Temko A, Nadeu C. Classification of acoustic events using SVM-based clustering schemes. *Pattern Recogn*. 2006;39(4):682–94.
51. Vapnik VN. The nature of statistical learning theory. New York: Springer; 1995.
52. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
53. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–30.
54. Cichero JAY, Murdoch BE. Acoustic signature of the normal swallow: characterization by age, gender, and bolus volume. *Ann Otol Rhinol Laryngol*. 2002;111(7 Pt 1):623–32.
55. Youmans SR, Stierwalt JA. Normal swallowing acoustics across age, gender, bolus viscosity, and bolus volume. *Dysphagia*. 2011;26:374–84.
56. Arvedson JC. Assessment of pediatric dysphagia and feeding disorders: clinical and instrumental approaches. *Dev Disabil Res Rev*. 2008;14(2):118–27.
57. Calvo I, Conway A, Henriques F, Walshe M. Diagnostic accuracy of the clinical feeding evaluation in detecting aspiration in children: a systematic review. *Dev Med Child Neurol*. 2016;58(6):541–53.
58. DeMatteo C, Matovich D, Hjartarson A. Comparison of clinical and videofluoroscopic evaluation of children with feeding and swallowing difficulties. *Dev Med Child Neurol*. 2005;47:149–57.
59. Hartnick CJ, Hartley BEJ, Miller C, Willging JP. Pediatric fiberoptic endoscopic evaluation of swallowing. *Ann Otol Rhinol Laryngol*. 2000;109(11):996–9.
60. Leder SB, Karas DE. Fiberoptic endoscopic evaluation of swallowing in the pediatric population. *Laryngoscope*. 2000;110(7):1132–6.
61. Link DT, Willging JP, Miller CK, Cotton RT, Rudolph CD. Pediatric laryngopharyngeal sensory testing during flexible endoscopic evaluation of swallowing: feasible and correlative. *Ann Otol Rhinol Laryngol*. 2000;109:899–905.
62. Sittin M, Arvedson J, Visotcky A, et al. Fiberoptic endoscopic evaluation of swallowing in children: feeding outcomes related to diagnostic groups and endoscopic findings. *Int J Pediatr Otorhinolaryngol*. 2011;75(8):1024–31.
63. Willging JP, Thompson DM. Pediatric FEESST: fiberoptic endoscopic evaluation of swallowing with sensory testing. *Curr Gastroenterol Rep*. 2005;7:240–3.
64. Lefton-Greif MA, Lefton-Greif MA, McGrattan KE, et al. First steps towards development of an instrument for the reproducible quantification of oropharyngeal swallow physiology in bottle-fed children. *Dysphagia*. 2018;33(1):76–82.
65. Martin-Harris B, Carson KA, Pinto JM, Lefton-Greif MA. BaBy-VFSSImP (c) a novel measurement tool for videofluoroscopic assessment of swallowing impairment in bottle-fed babies: establishing a standard. *Dysphagia*. 2020;35(1):90–8.
66. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical machine learning in health care. *Annu Rev*. 2020;4:123–44.
67. McAllister S, Kruger S, Doeltgen S, Tyler-Boltrek E. Implications of variability in clinical bedside swallowing assessment practices by speech language pathologists. *Dysphagia*. 2016;31(5):650–62.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Thuy T. Frakking PhD

Anne B. Chang PhD

Christopher Carty PhD

Jade Newing BEng(Hons)

Kelly A. Weir PhD

Belinda Schwerin PhD

Stephen So PhD