RESEARCH REPORT

Automated assessment of hyoid movement during normal swallow using ultrasound

Joan K.-Y. Ma[1] | **Alan A. Wrench[1,2]**

[1]Clinical Audiology, Speech and Language Research Centre, Queen Margaret University, Edinburgh, UK

[2]Articulate Instruments Ltd, Edinburgh, EH21 6UU, UK

Correspondence
Joan K.-Y. Ma, Queen Margaret University, Queen Margaret University Drive, Musselburgh EH21 6UU, UK.
Email: jma@qmu.ac.uk

Abstract

Background: The potential for using ultrasound by speech and language therapists (SLTs) as an adjunct clinical tool to assess swallowing function has received increased attention during the COVID-19 pandemic, with a recent review highlighting the need for further research on normative data, objective measurement, elicitation protocol and training. The dynamic movement of the hyoid, visible in ultrasound, is crucial in facilitating bolus transition and protection of the airway during a swallow and has shown promise as a biomarker of swallowing function.

Aims: To examine the kinematics of the hyoid during a swallow using ultrasound imaging and to relate the patterns to the different stages of a normal swallow. To evaluate the accuracy and robustness of two different automatic hyoid tracking methods relative to manual hyoid position estimation.

Methods & Procedures: Ultrasound data recorded from 15 healthy participants swallowing a 10 ml water bolus delivered by cup or spoon were analysed. The movement of the hyoid was tracked using manually marked frame-to-frame positions, automated hyoid shadow tracking and deep neural net (DNN) tracking. Hyoid displacement along the horizontal image axis (HxD) was charted throughout a swallow, and the maximum horizontal displacement (HxD max) and maximum hyoid velocity (HxV max) along the same axis were automatically calculated.

Outcomes & Results: The HxD and HxV of 10 ml swallows are similar to values reported in the literature. The trajectory of the hyoid movement and its location at significant swallow event time points showed increased hyoid displacement towards the peak of the swallow. Using an interclass correlation coefficient, HxD max and HxV max values derived from the DNN tracker and shadow tracker are shown to be in high agreement and moderate agreement, respectively, when compared with values derived from manual tracking.

Conclusions & Implications: The similarity of the hyoid tracking results using ultrasound to previous reports based on different instrumental tools supports the possibility of using hyoid movement as a measure of swallowing function in ultrasound. The use of machine learning to automatically track the hyoid movement potentially provides a reliable and efficient way to quantify swallowing function. These findings contribute towards improving the clinical utility of ultrasound as a swallowing assessment tool. Further research on both normative and clinical populations is needed to validate hyoid movement metrics as a means of differentiating normal and abnormal swallows and to verify the reliability of automatic tracking.

WHAT THIS PAPER ADDS

What is already known on this subject

- There is growing interest in the use of ultrasound as an adjunct tool for assessing swallowing function. However, there is currently insufficient knowledge about the patterning and timing of lingual and hyoid movement in a typical swallow. We know that movement of the hyoid plays an essential role in bolus transition and airway protection. However, manual tracking of hyoid movement is time-consuming and restricts the extent of large-scale normative studies.

What this study adds

- We show that hyoid movement can be tracked automatically, providing measurable continuous positional data. Measurements derived from this objective data are comparable with similar measures previously reported using videofluoroscopy and of the two automatic trackers assessed, the DNN approach demonstrates better robustness and higher agreement with manually derived measures. Using this kinematic data, hyoid movement can be related to different stages of swallowing.

Clinical implications of this study

- This study contributes towards our understanding of the kinematics of a typical swallow by evaluating an automated hyoid tracking method, paving the way for future studies of typical and disordered swallow. The challenges of image acquisition highlight issues to be considered when establishing clinical protocols. The application of machine learning enhances the utility of ultrasound swallowing assessment by reducing the labour required and permitting a wider range of hyoid measurements. Further research in normative and clinical populations is facilitated by automatic data extraction allowing the validity of prospective hyoid measures in differentiating different types of swallows to be rigorously assessed.

INTRODUCTION

Ultrasound imaging has been developed extensively over the last decade to analyse lingual articulation and for clinical intervention and assessment. The opportunity arises to pivot this expertise and equipment towards the assessment of swallowing function. A recent rapid review (Allen et al., 2021) highlighted the potential value of using ultrasound imaging as an adjunct to the clinical assessment of swallowing, with the need to further establish normative baselines for ultrasound-derived objective measures of swallowing function.

Using ultrasound to assess swallowing function presents unique challenges compared with speech. For speech, the primary interest is in the shape of the tongue relative to the palate, as the shape of the resulting air cavity has a close relation to the resonant structure of the speech acoustics. Since ultrasound reflects strongly from a tissue-to-air boundary, the tongue contour is well defined for vowels and still largely for consonants. Swallowing assessment differs in that a bolus is present in the oral cavity, allowing ultrasound to penetrate beyond the tongue surface and reflect off the bolus-to-air boundary. As a result, the task of determining the tongue surface in swallowing is computationally more complex with multiple reflections from the tongue surface, the bolus surface and the palate all in the same image, and therefore unable to draw parallel from the automated methods in the speech literature to find a single best tongue surface contour. Another significant difference between speech and swallowing is the dynamics of the tongue and hyoid. During speech, the tongue is the active articulator, and there is low to moderate movement of the hyoid. However, in swallowing, while the movement of the tongue provides information about bolus manipulation and transition, the superior and anterior movement of the hyoid during the pharyngeal phase plays an important role in facilitating the transition of the bolus and in airway protection (Vandaele et al., 1995). The latter is known to correlate with penetration and aspiration (Steele et al., 2011), a negative consequence of dysphagia that a swallowing intervention often tries to rectify (Vandaele et al., 1995). The hyoid is relatively easily identified in ultrasound images due to the shadow produced by the absorption of ultrasound energy by the hyoid bone. Hyoid displacement has previously been investigated as an indicator of the functioning of pharyngeal swallow in normal and disordered populations using VFSS (Kang et al., 2010; Paik et al., 2008). A recent study comparing hyoid excursion in ultrasound and videofluoroscopic study of swallowing reported a strong correlation between the two instrumental measures during a dry and a liquid swallow, but not for a puree swallow (Winiker et al., 2021).

Due to its salience in ultrasound image sequences and its relevance for swallowing function, the extent of anterior hyoid movement has been investigated by several research groups as a clinical marker for swallowing assessment (Chen et al., 2017; Hsiao et al., 2012; Lee et al., 2016). The recent rapid review (Allen et al., 2021) on ultrasound for swallowing assessment highlighted that while the hyoid movement is the most explored feature, there is still no consensus on the best methodology for measuring the hyoid excursion. The authors call for further development of measurement metrics for the ultrasound assessment of swallowing, including representative measures of swallowing function with clinical utility, the establishment of normative data sets and exploration of the reliability of different measures.

In a typical swallow, the superior and anterior movement of the hyoid contributes to the elevation of the larynx, deflection of the epiglottis leading to the closure of the laryngeal vestibule, and the opening of the upper oesophageal sphincter (UES) (Ekberg, 1986). The dynamic movement of the hyoid is commonly measured by a single hyoid displacement value, which is defined as the distance between the resting position of the hyoid and the position of its maximal advancement. Reduction in hyoid displacement is reported to be associated with poor swallowing function (Hsiao et al., 2012), increased residue and aspiration (Lee et al., 2016). However, not all studies show decreased hyoid displacement correlating with disordered swallow. Kendall and Leonard (2001) found that older participants with dysphagia advanced the hyoid farther than younger controls using a small 1 ml water bolus, though the advancement was slower and the hyoid remained maximally advanced for a shorter duration. This strategy was not observed for a larger bolus volume. A study of sarcopenic elderly also showed slightly increased hyoid displacement for small bolus swallows in individuals with reduced swallowing function compared with age-matched control (Chen et al., 2020).

A high degree of variability in hyoid displacement has been reported within the healthy population (Molfenter & Steele, 2011). Hyoid displacement can be influenced by factors such as body size (Molfenter & Steele, 2014), age (Kang et al., 2010; Logemann et al., 2002), and bolus volume (Ishida et al., 2002; Kim & McCullough, 2008; Nagy et al., 2015). This variability in healthy swallows leads to the question of the representativeness of hyoid displacement in quantifying the hyoid excursion during a swallow, especially in distinguishing normal and abnormal swallows (Ekberg, 1986).

Peak hyoid velocity, measuring the rate of change of the distance between the hyoid and mandible during the pharyngeal phase of the swallow, has been used to measure hyoid function. Barikroo et al. (2015) and Nagy et al. (2014) investigated the impact of bolus volume on peak hyoid velocity. Both studies reported an increase in peak velocity for 20 ml bolus compared with the 5 ml bolus, suggesting the speed of hyoid movement could also be an important feature of hyoid dynamics. Chen et al. (2020) reported increased hyoid velocity in individuals with reduced hyoid displacement in individuals with sarcopenia. They hypothesized that increased hyoid velocity is due to adaption and compensation for the reduced displacement to maintain normal swallowing function. Since both displacement and velocity measures have been found to be informative and because they have been reported in several studies, they will be used in this study as the basis for comparisons between tracking methods.

If ultrasound is to be adopted as a clinical tool for evaluating swallowing, it is imperative for the measurement protocol to be of high clinical utility. Donohue et al. (2021) discussed both the challenge of training clinicians and the time required to complete the frame-by-frame tracking of the hyoid movement during a swallow in the clinical setting. One potential solution to address this would be to develop a robust and accurate automated method of estimating hyoid position in every frame of the ultrasound image sequence and automatically identifying the maximum hyoid excursion and peak velocity.

Automated tracking of hyoid movement during a swallow has previously been applied to VFSS (30 frames/s) (Kaneyama et al., 2016; Lee et al., 2017; Spadotto et al., 2008). Most recently, Lee et al. (2021) used their MATLAB-based spatio-temporal analyser for motion and physiologic study (STAMPS) to evaluate the kinematics of hyoid movement in VFSS for post-stroke patients and identify features that predict successful recovery of swallowing function. Using 2 ml bolus volume, they showed significantly reduced normalized maximum horizontal displacement and reduced velocity of the anterior hyoid movement in individuals with poor prognosis for swallowing functions compared with those with a good prognosis. Such findings highlight the potential contribution of an automatic tracker and the quantitative measurement of hyoid kinematics in our understanding of swallowing physiology and enhancing the clinical utility of ultrasound in the clinical assessment of swallowing. This study will describe and assess two automated methods of tracking the hyoid position in an ultrasound image sequence.

Aims

In order to establish the practicality of ultrasound as a tool for swallowing assessment, we will discuss our recording protocol and the application of automated hyoid position estimation in exploring the hyoid kinematics throughout the execution of a swallow. Specifically, this study aims to do the following:

- To analyse the midsagittal hyoid dynamics and relate the patterns to the different stages of a normal swallow.
- To evaluate two methods of tracking the hyoid position from ultrasound images with the potential for automatic estimation of hyoid displacement and peak velocity measures.

METHOD

Participants

Data from 15 participants from two data sets were used for the present study. These included six participants (three males and three females aged between 22 and 37 years) from cohort G and nine participants (three males and six females aged between 27 and 72 years) from cohort R. All participants had self-reported normal swallowing function and no history of medical conditions or medication usage that might affect oral sensorimotor function or swallowing.

Stimuli

The data used for this study are part of two larger scale experiments in which the effect of taste or bolus consistency on swallowing was examined. In both experiments, water was used as control. The 10 ml water bolus trials from all participants in these cohorts were selected for the current analysis. The water bolus was delivered using a 60 ml flexi-cut cup in cohort G and a tablespoon in cohort R.

Ultrasound equipment

A pocket-sized ultrasound system (Micro, Articulate Instruments Ltd, Edinburgh, UK) operating in standard B-mode was used for recording. This portable system is already used by speech and language therapists (SLTs) clinically for the treatment of speech sound disorders and has some advantageous characteristics. The system records images at a fast rate of between 80 and 119 frames/s, which is important for capturing the fast pharyngeal phase of

bolus transit. Image processing is carried out within programmable software in real-time, meaning that hyoid position estimations have the potential to be overlaid on live images as an augmentation in the future. The acquisition of synchronized audio channel and camera-derived video with the ultrasound images allows recording the swallowing sound, approach and delivery of the bolus into the mouth and movement of the lips and jaw during the oral preparatory phase.

Each frame of the ultrasound image displayed on the screen was generated by the raw scanline data recorded and stored by the Micro system. Each frame consists of a matrix of pixel values, defined by the number of scanlines and pixels per scanline. Here, each frame is a 64 × 842 matrix. The number of scanlines depends on the percentage of the field of view (FOV) with a maximum of 64 scanlines at 100% FOV for the 20 mm radius 2–4 MHz probe used in the current experiment. In our set-up, there are typically 842 pixels along a scanline representing distances in the range 0–80 mm (as the depth is set to 80 mm).

Different probes were trialled in the experimental set-up. A 2–4 MHz 60 mm radius convex probe was initially considered. However, for participants with an acute angle between neck and jaw, image blackouts (resulting from the partial loss of probe contact) obscured the view of the hyoid movement. The addition of a 20 mm deep gel acoustic standoff (SONOKIT soft 200 × 100 × 20 mm; Sonogel Vertriebs GmbH, Bad Camberg, Germany) between the probe and the submental surface reduced these blackouts but did not remediate the problem in all cases. Although a specially contoured layer of standoff material or water-filled balloon (Chen et al., 2020; Kwong et al., 2020) might provide a better solution, this was not available at the time of recording. As a result, the 60 mm radius probe was set aside in favour of a smaller 20 mm radius 2–4 MHz probe. For most participants, this probe with a maximum 92° FOV was capable of imaging the hyoid throughout its full excursion and the anterior oral cavity. The probe made light contact with the submental tissue and did not interfere with laryngeal elevation. The probe was set to 3 MHz, and this low frequency provided good penetration and a bright tongue contour throughout the swallow. A depth of 80 mm was selected, and the resulting frame rate was 81 ultrasound frames/s.

A probe stabilization headset (UltraFit, Articulate Instruments) (Pucher et al., 2020) was fitted to each participant to maintain the probe in the midsagittal plane and reduce movement relative to the head (Figure 1). This minimizes the potential inconsistency in hyoid measurements between different frames resulting from probe or head movement rather than the actual movement of the structure of interest. The headset was fitted with a lightweight NTSC micro-camera mounted on the right side, offering

FIGURE 1 UltraFit headset holds the convex 20 mm radius ultrasound transducer in a midsagittal position while a side-mounted camera monitors the bolus approaching the mouth [Colour figure can be viewed at wileyonlinelibrary.com]

a profile image of the lips and chin. The recorded images were deinterlaced to provide 59.94 frames/s and an image size of 640 × 240 pixels.

AAA recording and analysis software

The ultrasound system was connected by USB to a computer laptop running Windows 10. The micro-camera was connected to the same laptop via a USB video capture card. The computer acted as the display for the ultrasound and AAA software application (Articulate Instruments Ltd) provided imaging, recording, synchronization of ultrasound with audio and video, and analysis facilities. Measurements of ultrasound image features and distances were automatically scaled in millimetre units.

Data collection

All the experiments were conducted in the Speech Lab at Queen Margaret University. Ultrasound recording was performed by two undergraduate students with no previous ultrasound experience, supervised by a senior experimental officer (SEO) with knowledge of the equipment. The Ultrafit headset was fitted on the participant by the SEO, with the ultrasound probe held in place in a midsagittal

submental position. Ultrasound gel was then applied, and the probe orientation was adjusted so that the tongue tip and the hyoid shadow were visible. Where this was not possible, the hyoid shadow was prioritized. Each trial began with a 'beep' sound when the record button on the AAA software was clicked, which cued the participant raised the spoon or cup to the mouth. This ensured that the entire natural swallow (from bolus approaching the mouth to the end of the pharyngeal swallow) was recorded.

Labelling swallowing events and region

Five swallow events were labelled in each swallow trial by the first author, who is a SLT with experience in dysphagia and instrumental evaluation of swallowing. The five swallow events were defined as when the beginning of bolus enters the oral cavity (S1), completion of bolus entering the oral cavity (S2), tongue tip makes initial contact with the anterior palate (S3), head of the bolus passes the hyoid shadow (S4) and completion of bolus passing the hyoid shadow (S5). These time points were identified by subjective estimation of the movement of bolus in the ultrasound image. The identification of S1 was also assisted by synchronized video of the lips. Two of the participants (RP7 and RP9) were excluded in the analysis of the swallow event due to incorrect depth setting at the time of the recording, which resulted in the tongue surface and palate not being fully visible in the entire swallow, and swallowing events could not be consistently identified. The second author relabelled the swallow events in 20% of the data, and the variance across the time labels for the swallow events was used to calculate the interrater reliability.

A temporal swallow region was marked for each trial, from 1 s before S1 to 1 s after S5. The movement of the hyoid was then tracked spatially within this swallow region.

Manual hyoid tracking

In adults, the hyoid has a significant midsagittal cross-section that absorbs ultrasound energy and generates a shadow beyond the hyoid from the perspective of the probe. A bright reflection is also often visible where the ultrasound beam encounters the hyoid, highlighting the anterior–inferior surface of the hyoid. The position of the hyoid was manually annotated by a trained SLT with previous ultrasound experience for each trial at the base and in the angular midpoint of the hyoid shadow on the bright reflection.

The position of the hyoid was marked on a keyframe at the start of the swallow. Additional keyframes were anno-

tated when a change in hyoid position was observed. The software then interpolated the hyoid position linearly in the intervening frames. The annotated points were superimposed on the ultrasound image, and the sequence of movements could be played back at the original speed or in slow motion. If the interpolated position deviated from the underlying image, a new keyframe was annotated to resolve the difference. On average, 15–30 manually annotated keyframes were needed to accurately track the hyoid movement over the duration of the swallow (approximately 5 s).

The interpolated output of the manual tracking can be charted as time series of x and/or y Euclidean coordinates using an arbitrary coordinate system. For this study, the measurement axes were set orthogonal to the image axes. Although this means the axes are dependent on the probe angle, the probe angle is set to include the hyoid at rest on the far left of the image, so there is some consistency between probe orientation, axis values, and anatomical features.

A second trained annotator, who is also a SLT with previous ultrasound experience, manually tracked the movement of the hyoid in 20% of the data. The interrater reliabilities across the two annotators for HxD max and HxV max were calculated.

Hyoid shadow tracking algorithm

Taking advantage of the hyoid absorption shadow, an algorithm (described in the additional supporting information) was developed within AAA to detect the position of the base of the hyoid shadow. For the shadow tracking, the hyoid shadow was searched for within an angular region of interest (ROI), typically within the range of scanlines 5–40 (8–62% of the FOV). A frame-to-frame continuity constraint was set (typically 9% of the FOV) to avoid the tracker jumping from the tracking of the hyoid shadow to the thyroid shadow or a blackout.

The hyoid position estimate tended to jitter between scanlines if the single best pixel value was selected. Smoothing was applied by searching for sets of five consecutive scanlines that had the highest average probability of being the hyoid position. The middle scanline of the best average set was then chosen as the estimate. The algorithm thus detected the centre of the hyoid shadow rather than the leading or trailing edge.

The algorithm is fast and works well if the following conditions are met.

- The hyoid/mandible shadow is clearly visible.
- The base of the hyoid/mandible shadow remains within the ROI.

- The thyroid shadow does not appear within the ROI for the hyoid.
- There are no blackouts due to loss of probe contact.

Deep neural net (DNN) approach

The DeepLabCut neural network training and analysis software package (Mathis et al., 2018; Nath et al., 2019) was used to estimate the hyoid position. DeepLabCut applies transfer learning so that a few examples of annotated images from a new domain can be used to retrain an initial network, Deepercut (Insafutdinov et al., 2016), that has been trained on thousands of images of human whole-body poses (arms, legs, torso, head). It is perhaps surprising that it should successfully estimate the hyoid position from ultrasound images which are of a significantly different type. Nevertheless, it is designed to identify poses of human articulation, and in that aspect, the movement of the hyoid is similar to the movement of, for example, an elbow joint.

The DeepLabCut MobileNetV2_1.0 network was trained on 40 manually labelled frames from a single repetition (rep 2) of each of the following participants GP1, GP3, GP4, GP6, RP1, RP2, RP4, RP7 and RP 8. This labelling was performed by the second author who has an engineering background and extensive experience in ultrasound tongue imaging. Other data from these participants and all recordings of GP2, GP5, RP3, RP5, RP6 and RP9 were unseen by the network during training. Ultrasound data were generated as a video of width 320 height 240 pixels (providing a resolution of 2.4 pixels/mm) at 81 Hz. Unlike the shadow tracker, no regions of interest needed to be specified. In general, the training time depends on the size of the images. Using a NVIDIA GTX 1060 graphics processor, 320×240 video training data took approximately 8 h for 900,000 iterations to be completed. Analysis of all 75 videos using the trained network operates in real-time (approximately 200 frames/s). The DeepLabCut analysis generated *.csv text files with the coordinates of the estimated hyoid position measured in pixels. The 75 *.csv files were then imported into the AAA analysis software (5 min) which automatically rescaled the coordinates in millimetres, ready for analysis

Post-processing

Although the hyoid movement was tracked in both x and y dimensions in the analyses above, the dynamic movement on the x-axis was selected to represent the hyoid movement, as it is the most commonly reported hyoid displacement measure in previous studies (e.g., Chen et al.,

2017; Lee et al., 2016). Hyoid displacement at the labelled swallowing events (S1–S5), maximum hyoid displacement parallel to the x-axis (HxD max) and the associated maximum hyoid velocity (HxV max) were extracted from both the manual tracking and the automatic tracking to characterize the hyoid kinematics during each swallow. From the x-axis trajectory of the hyoid movement within a swallow, the HxD max was extracted by automatically finding and measuring the difference between the maximum and minimum smoothed HxD values within the tracked swallow region. The HxV max was automatically found within a constrained temporal region ± 500 ms relative to HxD max. This restriction was applied in order to ensure that the velocity peak was associated with the maximum excursion, which is expected to be associated with the pharyngeal phase. Without this search region for the HxV max, an early hyoid raising was detected and incorrectly identified as HxV max in 12% of the trials before the bolus even entered the mouth (i.e., before S1).

The outputs of the two automated trackers exhibit frame-to-frame estimation error, which appears as high-frequency noise on the underlying hyoid movement chart. Based on the assumption that the true movement of the hyoid is smooth over time, a fourth-order Savitsky–Golay filter was applied to the x and y coordinate time series separately. A low-pass cut-off frequency of 13 Hz was used to smooth the time series. Similar smoothing is implemented in the videofluoroscopic hyoid tracker (STAMPS) developed by (Lee et al., 2017). A further low-pass filter with an 8 Hz cut-off was applied to the derivative of this smoothed displacement time series when calculating the velocity (HxV). For consistency, the exact same smoothing filters were applied to the manually tracked data.

Statistical analysis

IBM SPSS Statistics for Windows, version 23 (IBM Corp., Armonk, NY, USA) was used to calculate the interclass correlation coefficient (ICC) to evaluate the interrater reliability of the manual measurement, and to compare the performance of the three different tracking methods (manual, shadow and DNN) on both HxD max and HxV max. In the case of the interrater reliability of the manual tracking, the HxD max and HxV max measurements were repeated in 20% of the data, and the level of agreement was compared by ICC using a random two-way absolute agreement model. Similarly, the performance of the two automatic trackers was compared with that of the manual tracker by comparing the HxD max and HxV max using a random two-way absolute agreement model of ICC. The results of the ICC were interpreted with reference to previous literature (Koo & Li, 2016), with ICC < 0.50 indicative
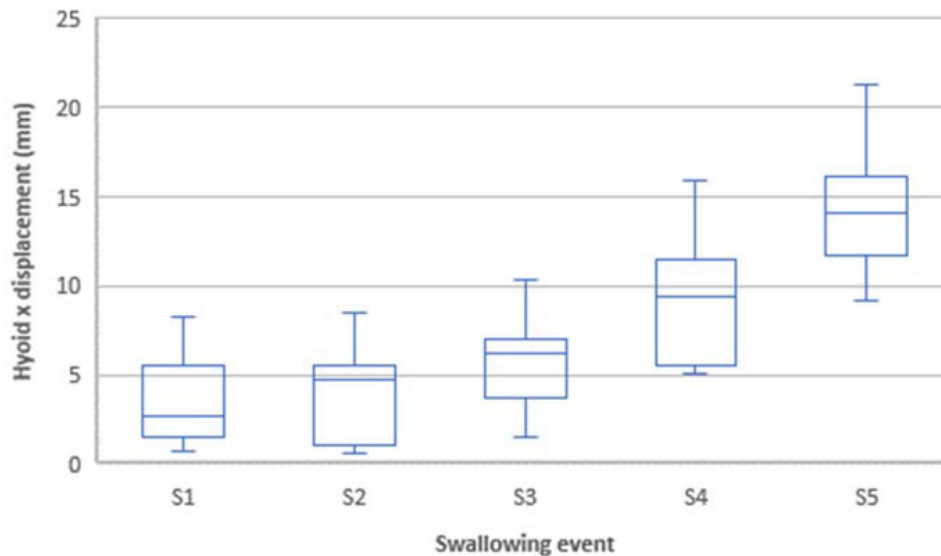
FIGURE 2   Average hyoid *x*-axis displacement value at the five labelled time points across five swallows of 13 participants [Colour figure can be viewed at wileyonlinelibrary.com]

Note: S1, Beginning of the bolus enters the cavity; S2, completion of the bolus entering the oral cavity; S3, the tongue tip makes initial contact with the anterior palate; S4, head of the bolus passes the hyoid shadow; and S5, completion of the bolus passing the hyoid shadow.

of poor reliability, 0.5–0.75 moderate reliability, 0.75–0.90 good reliability and > 0.90 representative of excellent reliability.

RESULTS

Dynamic movement of the hyoid during swallow

The movement of the hyoid parallel to the *x*-axis was manually tracked throughout the duration of the swallow. In terms of the maxima, we found an average of 16.0 mm HxD max (SD = 4.4 mm) and an average HxV max of 49.4 mm/s (SD = 18.7 mm/s) across all participants. Figure 2 shows the mean and standard deviation (SD) of the HxD values at S1, S2, S3, S4 and S5 averaged across five repetitions of 13 participants. Overall, the HxD value remained relatively low during the stage of bolus entering the oral cavity (S1 and S2), with the increase in HxD noted when the tongue tip moves forward to make contact with the anterior hard palate. The HxD value continued to increase and peaks at around S5 where the entire bolus has passed the hyoid shadow.

    While most previous studies have focused on the *x*-axis displacement of the hyoid or the distance between hyoid and mandible, a fuller understanding of the hyoid movement can be gained by observing displacement in two dimensions. Figure 3 shows a plot of the hyoid movement of the swallow region for rep1 recorded and manu-

ally tracked for each participant. As a means of displaying time, each trace starts faintly and becomes bolder through the swallow. Variations in the movement pattern can be observed across participants, largely due to the extent and direction of a raising component of the movement parallel to the *y*-axis. A general movement pattern of the hyoid was observed in all participants (Figure 3) to include a posterior extremum (point A), an anterior extremum (point C), and an intermediate superior extremum between points A and C (point B). The movement pattern is schematized in Figure 4. It began with a marked retraction of the hyoid (towards A), followed by upward and forward movements (towards B). The hyoid continues to move forward relative to the *x*-axis (towards C) and finally returns to the resting position towards the end of the swallow.

Interrater reliability

Interrater reliability in the time labels of the five swallow events was calculated by comparing the variance in 20% of the data marked by the first and second raters. One repetition from each participant was included in this second manual annotation to avoid bias, as tracking was more challenging for some participants than others. A two-way analysis of variance (ANOVA) was used, and the results showed no significant difference in the time labels between the first and the second raters across the five time points $F_{(4, 60)} = 0.79$, $p > 0.05$.
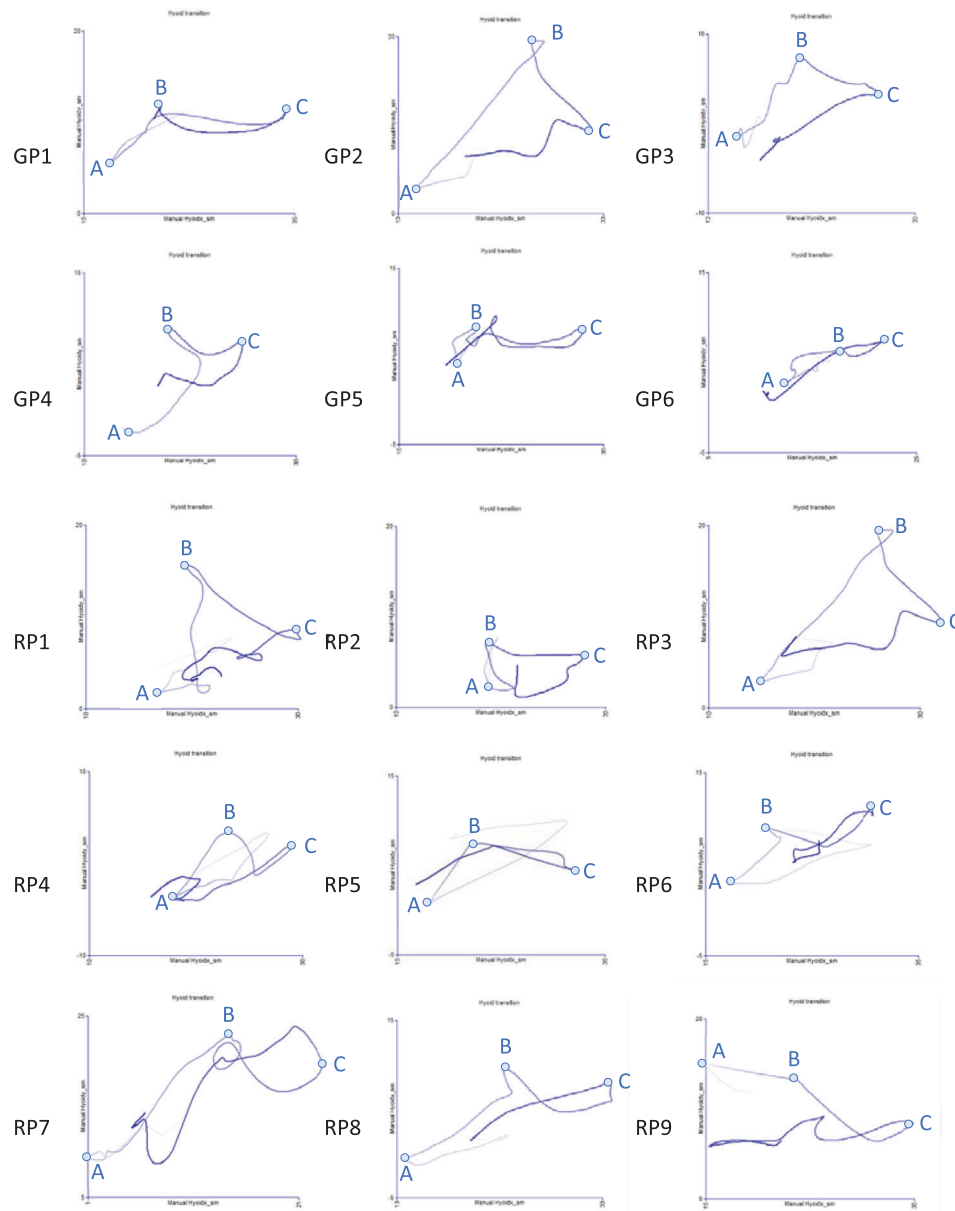
FIGURE 3 Movement of the hyoid in two dimensions for 15 recordings, one from each participant [Colour figure can be viewed at wileyonlinelibrary.com]

Note: The trace starts faintly, becoming bolder as time progresses through the annotated region. All plots were presented with the same aspect ratio within 20 × 20 mm *x*- and *y*-axes extents. The individual hyoid trajectory is marked by three extrema: point A marks the posterior extremum of the hyoid at the beginning of a swallow; point C marks the anterior extremum of the hyoid; and point B refers to an intermediate superior extremum between points A and C.

For the manual hyoid tracking, the interrater reliability was measured by having a second annotator repeat the tracking of 20% of the data. Similarly, one repetition from each participant was included. ICC was used to analyse the interrater reliability of the HxD max and HxV max values derived from tracks generated by annotators 1 and 2, based on a two-way mixed, absolute agreement and average measures model. The results are summarized in Table 1. The ICC showed good interrater reliability for HxD max and HxV max, with a standard error of measurement (95% confidence interval) computed as 19.49% and 33.38%, respectively.

Reliability between manual and automatic hyoid trackers

Manual tracking was compared with two different methods of automatic hyoid tracking (shadow versus DNN).
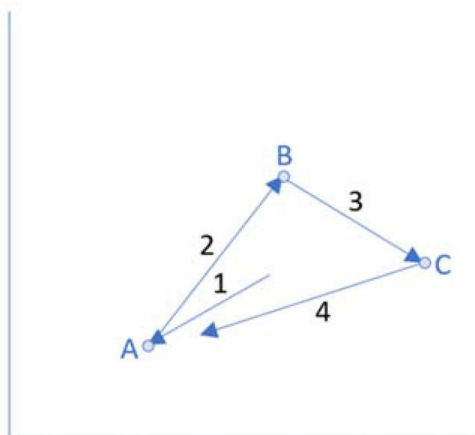
FIGURE 4 General pattern of hyoid movement [Colour figure can be viewed at wileyonlinelibrary.com]
Note: 1, Retraction towards point A to make space for the bolus; 2, raising to point B to aid tongue dorsum raising; 3, advancing to point C leading to laryngeal elevation; and 4, relaxation and returning to the resting position.

FIGURE 5 Ultrasound image from RP8 rep1 showing measurement along the *x*-axis (long blue horizontal line) [Colour figure can be viewed at wileyonlinelibrary.com]
Note: Shadow tracker (H); deep neural net (DNN) tracker (left end of the angled red line); two manual trackers (left end of the short horizontal pink and green lines).

TABLE 1 Interrater reliability measure between two annotators on manual tracking

	ICC	95% Confidence interval
HxD max	0.86	0.65–0.95
HxV max	0.76	0.44–0.91

Figure 5 shows an example of comparing the four trackers (two manual and two automatic measures) on a single frame of the ultrasound image. Figures 6 and 7 each show an example of the tracked HxD and HxV values throughout a swallow measured by the four different trackers. The HxD max and HxV max were automatically extracted from the hyoid traces for each tracker and compared across the manual and the automatic trackers.

Figure 8 summarizes the comparison of the HxD max between manual tracking and the two automatic hyoid trackers by displaying the mean and SD for each participant. The mean HxD max values were 16.0 mm (SD = 4.4 mm) for the manual tracker, 17.0 mm (SD = 4.40 mm) for the shadow tracker and 15.7 mm (SD = 4.0 mm) for the DNN tracker. The HxD max agreement between the three tracking methods was assessed using ICC, and the results are displayed in Table 2. Moderate agreement was observed between the manual tracking and the shadow tracker, and good agreement between the manual tracking and the DNN tracker. This shows that the HxD max calculated by DNN is closer to that of the manual tracker than the estimation of the shadow tracker.

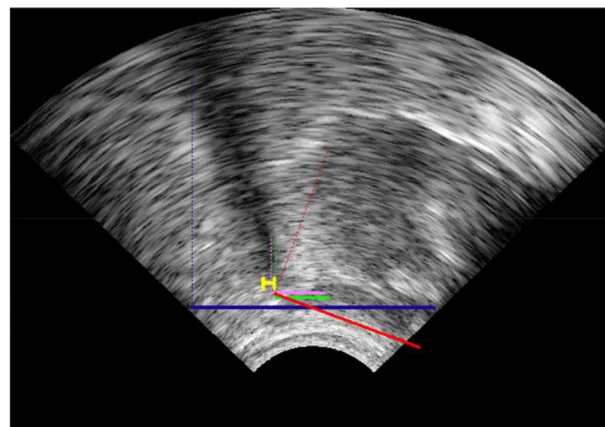In a further investigative step, the HxD max values were normalized by dividing them by the maximum hyoid–mandible distance (as determined from the DNN tracker, which output estimates of mandible position as well as hyoid) and replotted (Figure 9). Normalization made little overall difference but did scale the mean values for RP7 and RP8, making them more in accordance with values for other participants.

The mean HxV max value for each participant by the three tracking methods and its SD are displayed in Figure 10. The mean HxV max values were 49.4 mm/s (SD = 18.7 mm/s) for the manual tracker, 63.8 mm/s (SD = 30.4 mm/s) for the shadow tracker and 53.4 mm/s (SD = 18.7 mm/s) for the DNN tracker. The agreement between the manual tracking and the two automatic trackers were compared using ICC, and the results are summarized in Table 3. HxV max values obtained with both the shadow and DNN trackers showed a moderate correlation to manual tracking, although a higher ICC and 95% confidence interval were noted for the DNN tracker.

DISCUSSION

The first aim of this study was to investigate the midsagittal hyoid dynamics in relation to different stages of a normal swallow. Hyoid dynamics were examined by tracing the trajectory of hyoid movement throughout a swallow and quantitative measures of hyoid movement in a horizontal dimension (HxD max and HxV max). The trajectory of the hyoid movement observed was consistent with previous reports in the literature (Molfenter & Steele, 2011; Paik et al., 2008). The beginning of a swallow was characterized by retraction of the hyoid, which facilitates tongue
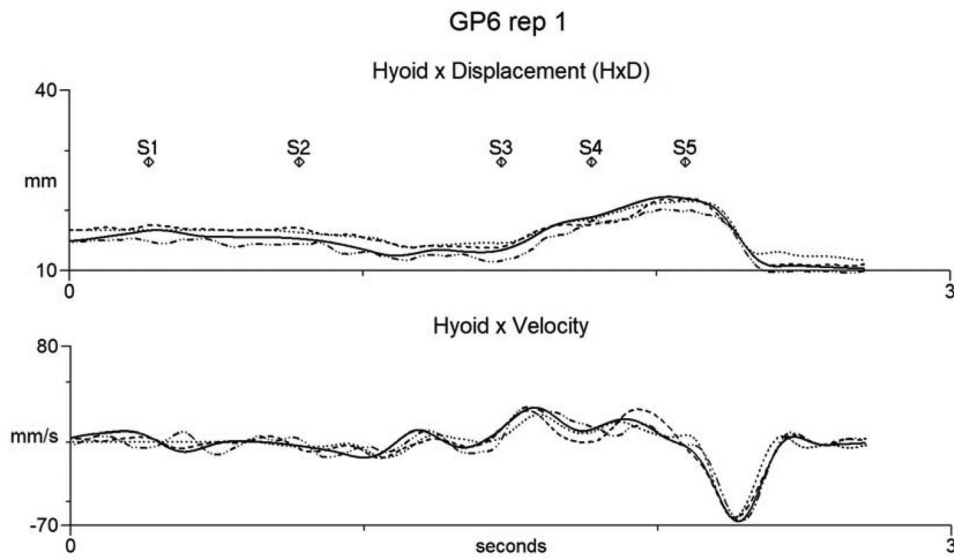
GP6 rep 1

Hyoid x Displacement (HxD)

Hyoid x Velocity

FIGURE 6 Tracked displacement and velocity along the *x*-axis of the ultrasound image sequence for participant GP6 rep1
Note: Solid = manual tracker 1; dot = manual tracker 2; dot–dot–dash = shadow tracker; and dash = deep neural net (DNN) tracker.
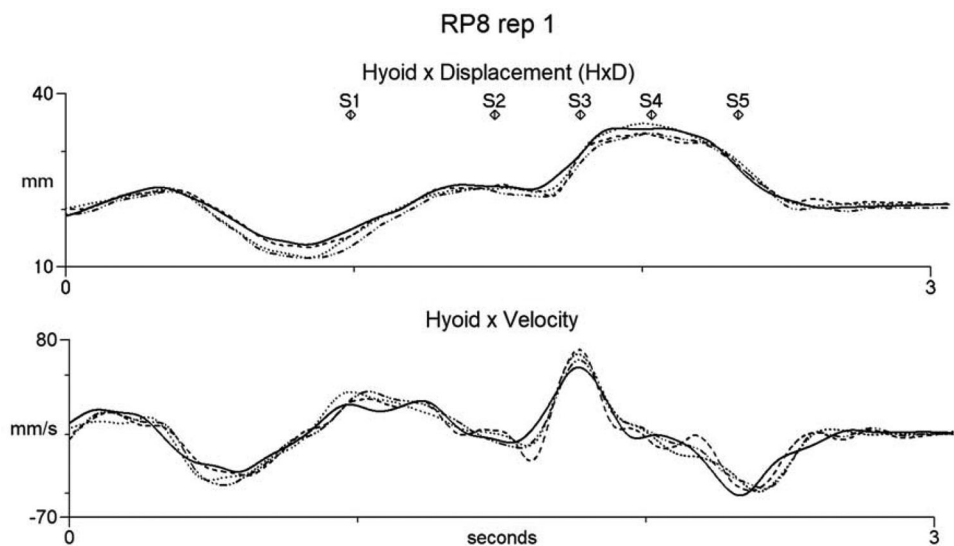
RP8 rep 1

Hyoid x Displacement (HxD)

Hyoid x Velocity

FIGURE 7 Tracked displacement and velocity along the *x*-axis of the ultrasound image sequence for participant RP8 rep1.
Note: Solid = manual tracker 1; dot = manual tracker 2; dot–dot–dash = shadow tracker; and dash = deep neural net (DNN) tracker.

retraction, increasing the volume of the anterior oral cavity to make space for the bolus to enter the mouth at the beginning of the oral preparatory phase. The hyoid then often moves superiorly to support tongue body raising in order to hold the bolus in the oral cavity. The hyoid then moves anteriorly, resulting in laryngeal elevation, airway protection and UES opening at the peak of the swallow. The hyoid remains in the advanced position for a short duration of time, during which the bolus passes from the oral cavity into the pharyngeal space. The hyoid then moves posteriorly and inferiorly, returning to the resting position.

The HxD max value reported in this study was comparable with that of 10 ml water swallow studies previously reported in the literature (Macrae et al., 2012; Molfenter & Steele, 2011). In the current study, HxD max was calculated by the difference between the maximum and minimum Hx values extracted from the swallow region defined by when the bolus starts to approach the mouth until the completion of the swallow. While the position of the hyoid before the beginning of a swallow has been used in most previous studies as the resting position (e.g., Hsiao et al., 2012; Lee et al., 2016), Molfenter and Steele (2011) reported
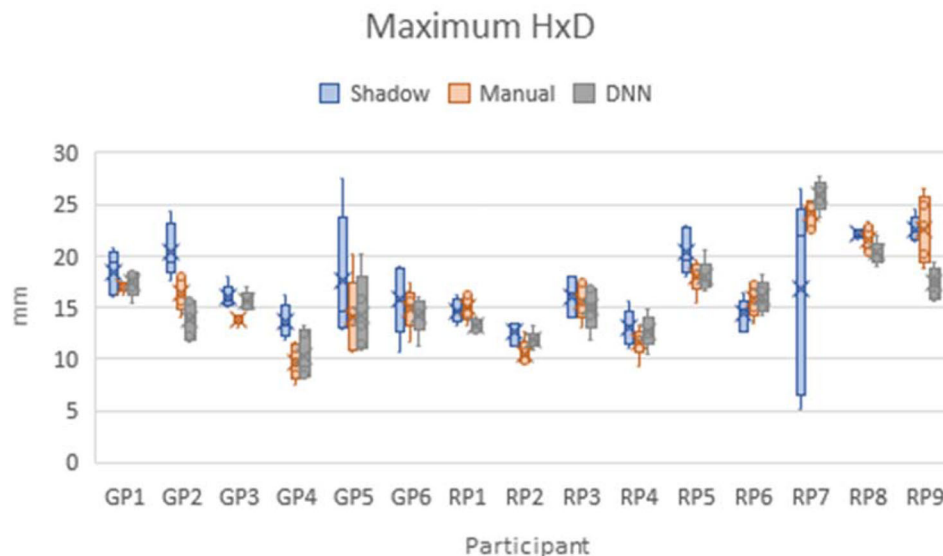
FIGURE 8 Box plot for HxD max for each participant using manual, shadow and deep neural net (DNN) tracking [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Agreement of HxD max between the manual tacker and two automatic trackers (shadow and deep neural net—DNN)

	ICC	95% Confidence interval
Manual versus shadow	0.62	0.46–0.74
Manual versus DNN	0.85	0.77–0.90

variability in how the rest frame is defined, including hyoid position before swallow (Vandaele et al., 1995), after swallow (Winiker et al., 2021) or the moment before bolus being propelled into the pharynx (Kim & McCullough, 2008). From observing the recorded data, it was noted that the lowest hyoid position was not always found before the swallow, especially when there is an anticipatory pre-swallow raising of the hyoid position. In these cases, a rest position may not necessarily be captured within a recording, especially if the cineloop buffer on the ultrasound system only allows a few seconds of recording time. Some instruction to the participant may be required in future studies to ensure that the hyoid is not already in a primed position at the start of a recording. Ensuring the hyoid is truly at rest in this way might improve the consistency of the absolute and normalized displacement in a swallow.

The second aim of the current study was to evaluate two different methods of automatic tracking the hyoid position from ultrasound images in estimating the hyoid displacement and peak velocity measures. The current results showed that the DNN tracker had a higher level of agreement to that of the manual tracker than the shadow tracker. However, there are advantages and disadvantages to both automatic trackers. The shadow tracker required

the performance for each trial to be reviewed and regions of interest to be modified, and the tracker rerun where it was deemed that hyoid lay outside the default region specification or where blackout regions need to be excluded to prevent the tracker from identifying them as a hyoid shadow. This, in effect, makes the shadow tracker only semi-automatic. However, the shadow tracker is currently implemented within the AAA acquisition and analysis software which runs on a standard business laptop. It is capable of running in real-time, displaying the hyoid position on top of the live display, which could potentially support the identification of the hyoid movement without additional signal processing in clinical practice.

The DNN tracker significantly outperformed the shadow tracker. The results showed that the DNN tracker had good agreement with HxD max and HxV max values assigned manually. It recovered well from blackouts even if these obscure the hyoid for parts of the recording. Although only a small number of frames were labelled ($n = 360$) from a small number of participants ($n = 9$), the DNN algorithm achieved hyoid tracking performance close to manual annotation. Improved performance could certainly be obtained by exposing the training to a more significant number of participants. Of the five participants unseen in the DNN training, only RP9 showed a considerable discrepancy in the HxD estimate (Figure 8). A review of the tracking for this participant indicated that it failed to estimate the extrema of hyoid positions, perhaps because there were very few examples in the training set of the hyoid in these extreme positions. Consideration should therefore be given to increasing representation in the training frames of these extrema. The DNN tracker
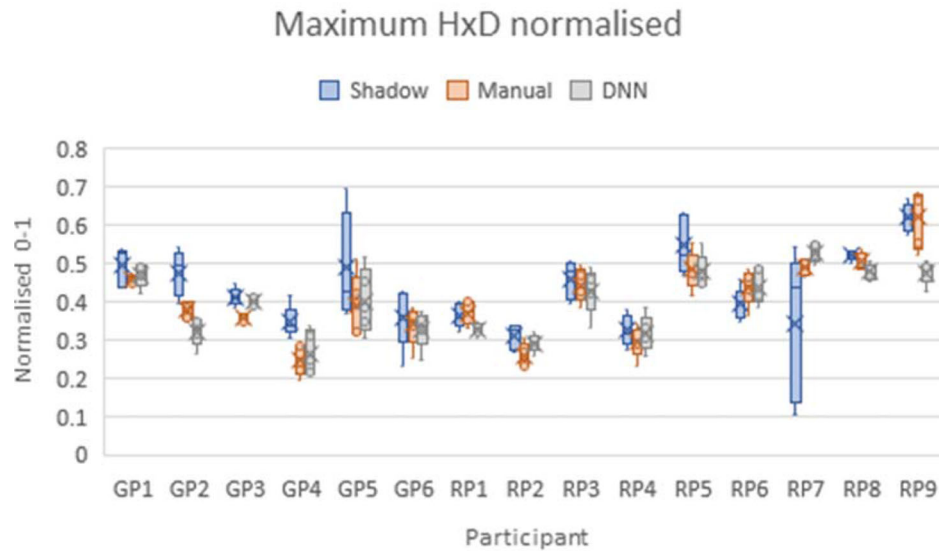
Maximum HxD normalised

FIGURE 9 Values from Figure 8 divided by the maximum mandible to hyoid distance for each recording to give a normalized HxD for each participant [Colour figure can be viewed at wileyonlinelibrary.com]
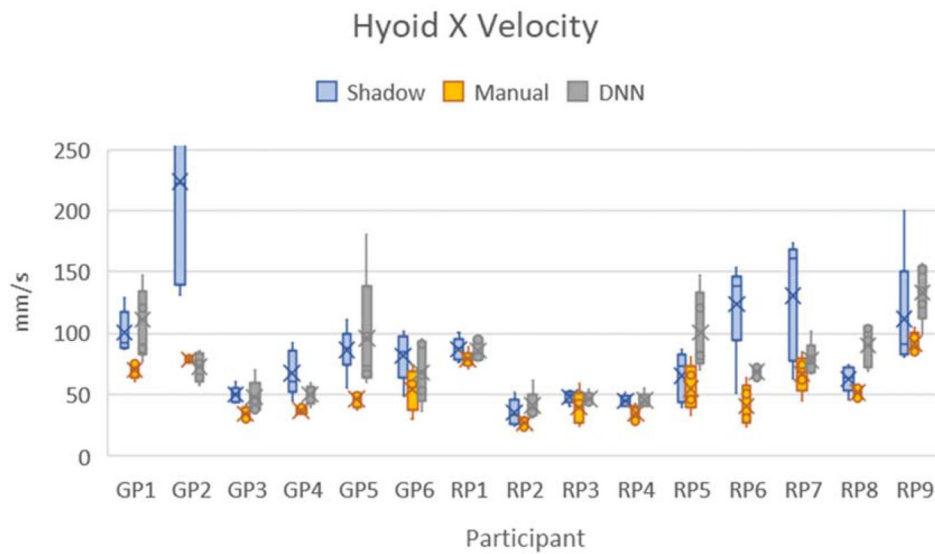
Hyoid X Velocity

FIGURE 10 Box plot for HxV max for each participant using manual, shadow and deep neural net (DNN) tracking [Colour figure can be viewed at wileyonlinelibrary.com]

provides a rating of its confidence in the position of the hyoid. In the case of a blackout where the position of the hyoid is obscured, confidence is rated low, allowing a warning to be issued. The network used in the current study was trained on data from a single probe from a single ultrasound system with a standardized FOV and depth. The DNN may need to be retrained for a different machine and/or probe. The DNN tracker is not currently integrated into the AAA acquisition and analysis software. Running in real-time requires an NVIDIA graphics card with 1000

TABLE 3 Agreement of HxV max between the manual tacker and the two automatic trackers (shadow and deep neural net—DNN)

	ICC	95% Confidence interval
Manual versus shadow	0.53	0.21–0.72
Manual versus DNN	0.73	0.60–0.83

or more Cuda cores and the DeepLabCut package of compilers and drivers to be loaded on the client computer. A preconfigured high-end computer would be required

to implement this tracker in a clinical setting, increasing costs but with the potential to make hyoid displacement assessment fully automatic.

CLINICAL IMPLICATIONS, LIMITATIONS AND FUTURE WORK

The results of this study contribute to the development of ultrasound as an emergent clinical tool for the evaluation of swallowing. The reported ultrasound analysis of a normal swallow addresses some of the challenges in swallowing data acquisition using ultrasound, such as in the selection of probe, headset stabilization and minimizing image blackout. Although based on a small sample size, the measures derived from continuous tracking of the hyoid show consistency with previous data reported in the literature. Automating the measurement process overcomes the practical obstacle of time-consuming manual measurements in a clinical setting and enables larger scale normative studies.

The current data set included swallowing data with either cup or spoon as the mode for bolus delivery. The different methods of bolus delivery could potentially impact the hyoid position in the oral preparatory stage of the swallow. However, the differences in the movement patterns were not compared statistically in the current study due to the small sample size. A larger scale study would be required to study the effect of bolus delivery on the oral preparatory stage of normal swallow.

HxD max and HxV max used in the current study are commonly applied measures in the literature in part because they can be derived by subjectively identifying from within a swallowing sequence only two frames of minimum and maximum hyoid displacement in a single movement dimension. Given the variability in the hyoid movement reported in the literature (Molfenter & Steele, 2011), future work exploring a more comprehensive set of metrics derived from the ultrasound measurements that includes different dynamic and timing characteristics of the hyoid movement and their objective measurement are warranted. The automatic hyoid tracking proposed in this paper provides a necessary and practical first step in developing such metrics.

Two automatic trackers were developed in the current study to potentially improve the clinical utility of ultrasound in swallowing assessment. Of the two, the DNN tracker was more robust and accurate than the shadow tracker. Measures derived from the DNN tracker showed high agreement with those derived from manual tracking. However, it should be noted that the training data for the DNN tracker was also included in the calculation of the agreement due to the small sample size in this study. Nonetheless, the DNN tracker still offers a promising approach to hyoid tracking and the performance of the DNN tracker is likely to improve with additional training data. Translating the implementation of the DNN tracker from the research setting to a clinical setting is, however, challenging and further development is needed to both improve its accuracy and integrate its use into clinical practice.

CONFLICT OF INTEREST
Joan K.-Y. Ma declares no competing interests. Alan Wrench is the director of Articulate Instrument Ltd.

DATA AVAILABILITY STATEMENT
Data are available on request from the authors.

REFERENCES
Ahmed, N., Sonja, M.M., Péladeau-Pigeon, S.S. & Catriona, M.S. (2014) The Effect of Bolus Volume on Hyoid Kinematics in Healthy Swallowing. *BioMed Research International*, 2014. https://doi.org/10.1155/2014/738971

Allen, J.E., Clunie, G.M., Slinger, C., Haines, J., Mossey-Gaston, C., Zaga, C.J. et al. (2021) Utility of ultrasound in the assessment of swallowing and laryngeal function: a rapid review and critical appraisal of the literature. *International Journal of Language & Communication Disorders*, 56(1), 174–204.

Barikroo, A., Carnaby, G. & Crary, M. (2015) Effects of age and bolus volume on velocity of hyolaryngeal excursion in healthy adults. *Dysphagia*, 30(5), 558–564

Chen, Y.-C., Chen, P.-Y., Wang, Y.-C., Wang, T.-G. & Han, D.-S. (2020) Decreased swallowing function in the sarcopenic elderly without clinical dysphagia: a cross-sectional study. *BMC Geriatrics*, 20(1), 419. https://doi.org/10.1186/s12877-020-01832-0

Chen, Y.-C., Hsiao, M.-Y., Wang, Y.-C., Fu, C.-P. & Wang, T.-G. (2017) Reliability of ultrasonography in evaluating hyoid bone movement. *Journal of Medical Ultrasound*, 25(2), 90–95.

Donohue, C., Mao, S., Sejdić, E. & Coyle, J.L. (2021) Tracking hyoid bone displacement during swallowing without videofluoroscopy using machine learning of vibratory signals. *Dysphagia*, 36(2), 259–269.

Ekberg, O. (1986) The normal movements of the hyoid bone during swallow. *Investigative Radiology*, 21(5), 408–410.

Hsiao, M.Y., Chang, Y.C., Chen, W.S., Chang, H.Y. & Wang, T.G. (2012) Application of ultrasonography in assessing oropharyngeal dysphagia in stroke patients. *Ultrasound in Medicine and Biology*, 38(9), 1522–1528.

Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M. & Schiele, B. (2016) *DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model*. In Computer Vision-ECCV

2016; Springer International Publishing: Cham, Switzerland, 2016 pp. 34–50

Ishida, R., Palmer, J.B. & Hiiemae, K.M. (2002) Hyoid motion during swallowing: factors affecting forward and upward displacement. *Dysphagia*, 17(4), 262–272.

Kaneyama, S., Sumi, M., Takabatake, M., Kasahara, K., Kanemura, A., Koh, A. et al. (2016) Preliminary evaluation of the pathomechanisms of dysphagia after occipitospinal fusion. *Spine*, 41(23), 1777–1784.

Kang, B.-S., Oh, B.-M., Kim, I.S., Chung, S.G., Kim, S.J. & Han, T.R. (2010) Influence of aging on movement of the hyoid bone and epiglottis during normal swallowing: a motion analysis. *Gerontology*, 56(5), 474–482.

Kendall, K.A. & Leonard, R.J. (2001) Hyoid movement during swallowing in older patients with dysphagia. *Archives of Otolaryngology–Head & Neck Surgery*, 127(10), 1224–1229.

Kim, Y. & Mccullough, G.H. (2008) Maximum hyoid displacement in normal swallowing. *Dysphagia*, 23(3), 274–279.

Koo, T.K. & Li, M.Y. (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163.

Kwong, E., Ng, K.-W.K., Leung, M.-T. & Zheng, Y.-P. (2020) Application of ultrasound biofeedback to the learning of the Mendelsohn maneuver in non-dysphagic adults: a pilot study. *Dysphagia*, 36(4), 650–658.

Lee, W.H., Chun, C., Seo, H.G. et al. STAMPS: development and verification of swallowing kinematic analysis software. *BioMedical Engineering OnLine*, 16, 120 (2017). https://doi.org/10.1186/s12938-017-0412-1

Lee, W.H., Lim, M.H., Seo, H.G. et al. Hyoid kinematic features for poor swallowing prognosis in patients with post-stroke dysphagia. *Scientific Reports*, 11, 1471 (2021). https://doi.org/10.1038/s41598-020-80871-4

Lee, Y.S., Lee, K.E., Kang, Y., Yi, T.I. & Kim, J.S. (2016) Usefulness of submental ultrasonographic evaluation for dysphagia patients. *Annals of Rehabilitation Medicine*, 40(2), 197–205.

Logemann, J.A., Pauloski, B.R., Rademaker, A.W. & Kahrilas, P.J. (2002) Oropharyngeal swallow in younger and older women. *Journal of Speech, Language, and Hearing Research*, 45(3), 434–445.

Macrae, P.R., Doeltgen, S.H., Jones, R.D. & Huckabee, M.-L. (2012) Intra- and inter-rater reliability for analysis of hyoid displacement measured with sonography. *Journal of Clinical Ultrasound*, 40(2), 74–78.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W. et al. (2018) DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9), 1281–1289.

Molfenter, S.M. & Steele, C.M. (2011) Physiological variability in the deglutition literature: hyoid and laryngeal kinematics. *Dysphagia*, 26(1), 67–74.

Molfenter, S.M. & Steele, C.M. (2014) Kinematic and temporal factors associated with penetration–aspiration in swallowing liquids. *Dysphagia*, 29(2), 269–276.

Nagy, A., Molfenter, S.M., Péladeau-Pigeon, M., Stokely, S. & Steele, C.M. (2015) The effect of bolus consistency on hyoid velocity in healthy swallowing. *Dysphagia*, 30 (4), 445–451.

Nath, T., Mathis, A., Chen, A.C., Patel, A., Bethge, M. & Mathis, M.W. (2019) Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, 14(7), 2152–2176.

Paik, N.-J., Kim, S.J., Lee, H.J., Jeon, J.Y., Lim, J.-Y. & Han, T.R. (2008) Movement of the hyoid bone and the epiglottis during swallowing in patients with dysphagia from different etiologies. *Journal of Electromyography and Kinesiology*, 18(2), 329–335.

Pucher, M., Klingler, N., Luttenberger, J. & Spreafico, L. (2020) Accuracy, recording interference, and articulatory quality of headsets for ultrasound recordings. *Speech Communication*, 123.83–97.

Spadotto, A.A., Gatto, A.R., Cola, P.C., Montagnoli, A.N., Schelp, A.O., Silva, R.G.D.A. et al. (2008) Software para análise quantitativa da deglutição. *Radiologia Brasileira*, 41(1), 25–28.

Steele, C.M., Bailey, G.L., Chau, T., Molfenter, S.M., Oshalla, M., Waito, A.A. et al. (2011) The relationship between hyoid and laryngeal displacement and swallowing impairment. *Clinical Otolaryngology*, 36(1), 30–36.

Vandaele, D.J., Perlman, A.L. & Cassell, M.D., (1995) Intrinsic fibre architecture and attachments of the human epiglottis and their contributions to the mechanism of deglutition. *Journal of Anatomy*, 186, 1–15.

Winiker, K., Burnip, E., Gozdzikowska, K., Guiu Hernandez, E., Hammond, R., Macrae, P. et al. (2021) Ultrasound: validity of a pocket-sized system in the assessment of swallowing. *Dysphagia*, 36(6), 1010–1018.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.