Swallowing Sound Waveform and Its Clinical Significance:
Evaluation Using Ultrasonography

Tsuyoshi HONDA¹, Takuro BABA¹, Keiko FUJIMOTO¹, Kan NAGAO¹, Akira TAKAHASHI², Tetsuo ICHIKAWA¹

キーワード：Dysphagia, Swallowing sound, Ultrasonography

Abstract: Background: Cervical auscultation is a technique frequently used for the screening of dysphagia. However, this method is difficult to evaluate objectively and it is unclear how sound is generated during the swallowing process. The aim of this study was to analyze the waveform of swallowing sound and clarify the sound production process using recordings of swallowing sounds and ultrasound images (USI), performed simultaneously.

Materials and Methods: Commercial natural spring water and natural carbonated water were used in experiments 1 and 2, respectively. In experiment 1, a microphone was attached to the skin of the neck of 20 young participants and swallowing sounds were recorded and analyzed. In experiment 2, swallowing processes in three participants were recorded by a medical ultrasonography apparatus. The ultrasonic probe was placed on the skin over one of the thyroid cartilages or the thyroid gland.

Results: The swallowing sound wave (SSW) was divided into three sectional periods. The mean duration of the first, second, and third SSW was 210 ± 147 ms, 458 ± 113 ms, and 91 ± 61 ms, respectively. The mean intensity ratio of the first, second, and third SSW was 7.8 ± 5.2, 29.2 ± 16.5, and 5.8 ± 5.1, respectively. When the ultrasonic probe was placed on the skin over one of the thyroid cartilages, in the phase between the production of the second SSW and the silent period, the USI revealed an accumulation of swallowed material around the valleculae and oropharynx. In the silent period of the second SSW, the swallowed material accumulated around the hypopharynx. When the ultrasonic probe was placed on the skin over the thyroid gland, in the silent period of the second SSW, the USI revealed that the swallowed material had passed through esophagus.

Conclusion: Waveform and USI findings from this study suggest that swallowing sound can be divided into three sectional periods: an oral phase, a pharyngeal phase, and a repositioning phase.

I. Introduction

Recently, the aging rate of the global population has been increasing. Japan, in particular, is becoming a super-aged society; in 2010¹ the elderly (≥ 65 years old) comprised over 23% of the population. Dysphagia is one of the major disabilities and diseases associated with aging. It reduces the patient's quality of life through malnutrition and aspiration pneumonia. The serious and critical consequences of dysphagia also reduce activities of daily living. Therefore, it is important to evaluate the risk of dysphagia early. A variety of clinical examinations have been proposed as screening methods for dysphagia; among these, the repetitive saliva

¹ Department of Oral and Maxillofacial Prosthodontics, Institute of Health Biosciences, The University of Tokushima Graduate School
² Department of Oral Surgery, Institute of Health Biosciences, The University of Tokushima Graduate School
swallowing test\(^7,8\), the modified water swallow test\(^9\), and the food test\(^{10}\) are commonly used. In addition, cervical auscultation, a technique for listening to and assessing swallowing sounds using a stethoscope, is a very simple and effective method also commonly utilized\(^{11-18}\). However, this technique is difficult to evaluate objectively, and it is unclear how sound is generated during the swallowing process. The relationship between the swallowing sound and process has to be clarified to ensure an adequate examination of dysphagia with the cervical auscultation.

Imaging examinations, in which the movements of swallowing and bolus from the oral cavity to the esophagus can be directly detected, are also very useful for the examination of dysphagia. In particular, video-fluoroscopic examination is considered the gold-standard for evaluating dysphagia\(^{19}\); however, a limitation of this method is the associated radiation risk\(^{20}\). Ultrasonography, an ultrasound-based imaging technique for visualizing internal body structures, is a safe alternative method because it does not produce ionizing radiation\(^{21,22}\). There has been no study on the measurements of the swallowing process and sound using ultrasonography.

The aim of this study was to determine the swallowing sound production process by analyzing the waveform of swallowing sound recordings and ultrasound images (USI) taken simultaneously and to validate the cervical auscultation for examinations of dysphagia.

II. Materials and Methods

1. Participants

Participants were randomly recruited from undergraduates and graduate students in the University of Tokushima and were selected from a group of volunteers based on the following inclusion criteria: i) normal dentition, ii) no history of stomatognathic disorder and dysphagia, iii) more than 3 of repetitive saliva swallowing test value (RSST value), iv) informed consent obtained before inclusion in this study. Twenty-three young and healthy volunteers were enrolled as participants in this study (experiment 1: 10 men and 10 women, mean age = 25.8 years; experiment: 2, men and one woman, mean age = 28.3 years). This research was conducted with the approval by the clinical research ethical review board of Tokushima University Hospital (No.1406).

2. Swallowing materials

A Commercial, natural spring water (CRYSTAL GEYSER\(^7\), Otsuka Foods, Japan) and natural carbonated water (Minami Alps Tennensui Sparkling\(^6\), Suntory Beverage & Food, Tokyo, Japan) at approximately 25°C were used in experiments 1 and 2, respectively.

3. Recording of swallowing sounds

The swallowing sound recording system is outlined in Figure 1. An ultra-small condenser microphone (AT9903, Audio-Technica, Tokyo, Japan; frequency characteristic: 30-18000 Hz) was used to detect swallowing sounds. The tip of the sensor was enlarged to 10 mm in diameter using an auto-curing resin (UNIFAST II Clear, GC, Tokyo, Japan) for ease of attachment to the skin. The microphone was attached to the skin of the neck using double-sided tape (1517, 3M Japan, Tokyo, Japan). The output signal was amplified through an amplifier (AT-MA2, Audio-Technica) and a low-pass filter (NF 3611, NF CORPORATION, Yokohama, Japan; fc: 10 kHz, -24 dB ± 2 dB/ oct). The signals were digitized as 16 bit at a 20 kHz sampling rate using an analog-digital card (BPC-0600, Interface Corporation, Hiroshima, Japan), and analyzed using a waveform analysis software (DADiSP, CAE Solutions Corporation, Tokyo, Japan) on a personal computer.

4. Ultrasound imaging

The swallowing processes were recorded by a medical ultrasonography apparatus (LOGIQ Book XP Enhanced, GE Health Care Japan, Tokyo, Japan). The scanning probe (GE 8L-RS Linear Array Ultrasound Probe, Frequency: 6 MHz, Dynamic range: 90 dB, Gain: 58 dB) was placed on the skin over one of the thyroid cartilages or the thyroid gland, so as to obtain axial transverse images.

Before the recordings, anatomical structures and swallowed material in the USI were identified through video-endoscopy (FNL-10RP3, HOYA, Tokyo, Japan). Figure 2 presents the USI of three swallowing materials; carbonated water was observed as a white, bubble-like feature. The chewable, green-colored, moderately thickened water (100 ml pure water containing 2 g of a thickening agent [Tsururinko Quickly\(^\circledR\), CLINICO, Tokyo, Japan]) which accumulated around the valleculae during chewing, was more clearly visible as a white, bubble-like feature.

Figures 3 and 4 present the USI of the thyroid cartilage and thyroid gland positions, respectively. In the thyroid cartilage position, the accumulation of swallowed material around the valleculae and oropharynx was observed as highly echogenic regions in the USI. These regions passed through the hypopharynx and were therefore found on the equivalent site on the USI. In the thyroid gland position, the esophagus was clearly observed in the USI because there was little acoustic shadow of the thyroid and cricoid cartilages. When
5. Measurement procedures

In experiment 1, participants were seated in a soundproof room, on a straight-backed chair and in an upright position. The microphone was attached to the skin over the lateral laryngeal prominence (Figure 1). Participants were asked to swallow 5 ml of water (administered into the sublingual region using a syringe) in one complete swallow, with the chin in the normal position. The recording was initiated 2 s before the investigators commanded to swallow water and stopped 2 s after completion of the swallow, as determined by the participants. Measurements for each condition were repeated three times, separated by an interval of 3 min.

In experiment 2, the swallowing processes were recorded using a medical ultrasonography apparatus to detect swallowing sound. The scanning probe was placed on the skin, to the left, over one of thyroid cartilages or the thyroid gland. The participants were asked to swallow 5 ml of carbonated water, in a same manner as in experiment 1. The two acoustic and visual signals were transferred to a visual recorder (AQ-VU, TEAC, Tokyo, JAPAN). The visual signals were recorded as 30 frames per second and acoustic signals were recorded as 16 bit at a 2 kHz sampling rate. Images for each swallowing sound event were evaluated.

III. Results

Figure 5 shows a typical swallowing sound waveform
based on the overall data from the 20 participants. The waveform was divided into three sectional periods through the overall inspections of digital signal data. The first swallowing sound wave (first SSW) was defined as the period from the beginning of the swallowing wave (first peak) to the beginning of the following large peaks, the second SSW was the period from the beginning of the large peak to the beginning of the third peak, and the third SSW was the period from the beginning of third peak to the offset of the swallowing wave. Each first and second SSW comprised a swallowing wave production followed by a silent period in the latter part. The segmentations of each SSW were interactively generated by one examiner with a comparison to the noise level of the resting period. Duration and peak intensities of each SSW were measured in experiment 1. The peak intensity was evaluated as the proportion of peak-to-peak value (P-P value) to the noise level of the resting period (Figure 6).

The mean duration and peak intensity ratio of the three SSWs when 5 ml water was swallowed is presented in Figure 7. The mean duration of first, second, and third SSW was 210 ± 147 ms, 458 ± 113 ms, and 91 ± 61 ms, respectively, and the overall duration of swallowing sound was 739 ± 179 ms. The mean intensity ratio of the first, second, and third SSW was 7.8 ± 5.2, 29.2 ± 16.5, and 5.8 ± 5.1, respectively.

Figure 8 displays the typical USI for each swallowing sound waveform event which occurred when the ultrasonic
probe was placed on the skin over one of the thyroid cartilages. In the phase between the production of the second SSW to the silent period, the USI revealed an accumulation of the swallowed material around the valleculae and oropharynx. The swallowed material then passed through the hypopharynx in the silent period of the second SSW. There was no swallowed material was observed on the USI in the first and third SSW.

Similar findings were found in the three participants.

Typical sonograms for each swallowing sound waveform event when the ultrasonic probe was placed on the skin over the thyroid gland are presented in Figure 9. In the silent period of the second SSW, the swallowed material passed through esophagus, while in the first and third SSW, and the production period of the second SSW, there was no swallowed material was observed on the USI. Similar findings were found in the three participants.

**IV. Discussion**

A variety of models on swallowing dynamics have been proposed, and the three-stage theory comprising oral, pharyngeal and esophageal phases is generally accepted. The solid bolus translation for swallowing is also understood in the same way. In addition, a five-stage model for ingestion/swallowing was reported by Leopold et al. as follows: preoral (anticipatory), preparatory, oral (propulsive), pharyngeal, and esophageal. Furthermore, Palmer proposed a process model in which a proportion of the solid bolus is translated around the valleculae during oral processing. In this present study, swallowing processes at the time of the swallow water command followed the five-stage process model.

The swallowing sound was divided into three sectional periods. The first SSW, with a duration of 210 ms, was the first component of the overall swallowing sound. While there have been no previous studies on swallowing sound in the oral stage, Leonard et al. reported that the latency between the first movement of the bolus tip past the posterior nasal spine which leads to a swallow, and the first hyoid elevation was 210 ms. This latency corresponds to the duration of first SSW recorded in this present study. The findings from the sonogram in this period suggest that sound is generated when the swallowed material is held in the oral cavity and then transferred to the pharynx. These results indicate that posterior movements of the tongue and hyoid bone are the swallowing movements corresponding to the first SSW. Although the definition of the first SSW as a sound prior to the pharyngeal reflex may be questionable, the wave profile in addition to the USI findings, indicate that it should be interpreted as the oral stage.

The second SSW, with a duration of 460 ms, was the middle component of the swallowing sound and consisted of continuous and large waveforms in the early part, and a silent period in the latter part. In previous studies, the swallowing sound which occurs in the pharyngeal stage was found to be focused and associated with two major events: the movement of the larynx and epiglottis closure, and the passing of the bolus from the pharynx to the esophagus. In this present study, these two events could not be separated in the wave profile, despite analysis of the relevant USI. Boiron et al. found that the mean total duration of the pharyngeal stage was 620 ms when 5 ml water was swallowed and Hammoudi et al. reported a that mean total duration was 441 ± 150 msec for the swallowing of solid material. These durations would be corresponded to the total duration of second SSW and third SSW in this study. Therefore, the second SSW would be generated when the pharyngeal reflex occur when the swallowing material is transferred from the pharynx to the esophagus. These findings suggest that sound production of the second SSW is related to the larynx movements beginning.
with elevation of the hyoid bone, and followed by closure of the epiglottis and the passage of the swallowed material from the pharynx to the esophagus.

The third SSW, the last component of the swallowing sound, was observed after the material flowed into the esophagus; it had a mean total duration of less than 100 ms, with a weak and short waveform. Moriniere et al.\(^{30}\) reported that the mean duration between the descent and opening of larynx was 72 ms. The sound production of the third SSW could therefore be related to repositioning of the hyoid bone, the larynx, and the epiglottis.

Previous studies have demonstrated that the intensity of swallowing sound is dependent on age, gender, viscosity, and the volume of food bolus; however, no studies have evaluated the relative intensity of each swallowing sound component. According to Klahn et al.\(^{30}\) and Perlman et al.\(^{30}\), peak sound intensity occurs when the upper sphincter muscle opens completely and the bolus passes through esophageal orifice; this period corresponds to the second SSW in this present study.

V. Conclusion

In conclusion, the waveform and USI findings from this study suggest that swallowing sound can be divided into three sectional periods which are associated with each process of a serial swallowing movements: posterior movements of the tongue and hyoid bone as an oral phase, larynx movements with the hyoid bone elevation, the epiglottis closure and the bolus passage from the pharynx to the esophagus as a pharyngeal phase, repositioning of the hyoid bone, and epiglottis as a repositioning phase. This finding could be effective to validate the cervical auscultation for a better examination.

References