



HAL
open science

Swallowing sound recognition at home using GMM

Hajer Khlaifi, Dan Istrate, Jacques Demongeot, Jérôme Boudy, Dhafer Malouche

► **To cite this version:**

Hajer Khlaifi, Dan Istrate, Jacques Demongeot, Jérôme Boudy, Dhafer Malouche. Swallowing sound recognition at home using GMM. Journées d'Etude sur la TéléSANTé, 6ème édition, Pôle Capteurs, Université d'Orléans, May 2017, Bourges, France. hal-01692422

HAL Id: hal-01692422

<https://hal.science/hal-01692422v1>

Submitted on 25 Jan 2018

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

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

Swallowing Sound Recognition at Home Using GMM

Hajer Khlaifi^{*†}, Dan Istrate[†], Jacques Demongeot[‡], Jérôme Boudy[§], Dhafer Malouche^{*}

[†]Sorbonne University, University of Technology of Compiègne, CNRS,
UMR 7338 Biomechanics and Bioengineering, 60200 Compiègne, France

[‡]Université Grenoble Alpes (UGA), AGEIS, Faculté de Médecine, 38700 La Tronche

[§]Télécom SudParis, 9, Rue Charles Fourier, Evry, France

^{*}National School of Engineers of Tunis, Signals and Systems Research Unit,
University of Tunis El Manar, Tunisia,

hajer.khlaifi@utc.fr

Abstract

Aiming for autonomous living for the people after a stroke is the challenge these days especially for swallowing disorders or dysphagia. It is in this context that the e-swallowhome project proposes to develop tools, from hospital care until the patient returns home, which are able to monitor in real time the process of swallowing. This paper proposes a non-invasive acoustic based method to differentiate between swallowing sounds and other sounds in normal ambient environment during food intake. Gaussian Mixture Models (GMM) adapted through the Expectation Maximization (EM) algorithm was employed for classification and test combination using the leave-one-out approach according to the small amount of data in our database. Results provide a good recognition rate of 100 %.

Keywords: Sound recognition, GMM, acoustical features

I. Introduction

Increased life expectancy increases the risk of stroke. Its incidence in the elderly is important ; about 130 000 French people are affected each year by a stroke. In France, according to the French National Authority for Health (HAS) [4], stroke is the leading cause of acquired disability in adults, the second cause of dementia and the third cause of death. The increase in survival (e.g., 79.3 years for French men and 85.5 for French women in 2014) leads to an increase in care needs and social charges [7] because of the high frequency of medical complications estimated between 40% and 96% [16]. Among them, swallowing disorders or dysphagia are frequent, 42 to 76% of patients have these problems after

acute stroke [30]. Its prevalence increases with age, from 9% for subjects aged 65-74 living at home to 28% after 85 years [23]. Numerous studies have attempted to establish the incidence of dysphagia after a stroke with figures ranging from 23% to 50% [26]. In institutions, it reaches 30 to 60% [3]. As for patients living at home, a study by questionnaire found that 10 to 16% of them had symptoms of dysphagia [23]. These disorders are responsible for severe complications such as aspiration, pneumonia and undernutrition. They are responsible for about 4000 deaths per year in France [23]. The incidence of pneumonia is estimated to be around one third of patients [10]. Pneumonia is associated with a quarter of deaths in the first month after stroke [30]. Some studies show that approximately 4% of elderly people living at home are malnourished [3]. They are also responsible for the degradation of the patient's quality of life by causing them anxiety associated with meals, loss of eating pleasure and then depression and social isolation.

These disorders require multidisciplinary care, requiring a good organization and articulation of the health and social system. In this order, the e-swallowhome project participates in the home monitoring of the swallowing process by proposing automatic methods for monitoring and evaluating functional rehabilitation. The aim of our work is to start with different signals and ambulatory way to automatically detect swallow event firstly and in a second time identifying swallow event with and without disorders. In order to recognize swallow event, the aim of this paper is to differentiate between swallow event sound and other event sound which can be present during food intake using GMM model.

II. State of the art

Swallowing disorders are frequent after the onset of neurological disorders such as stroke, which causes lesions in certain parts of the brain that may be the cause of this dys-

phagia and may thus have disrupted the mechanisms involved in the function of swallowing. The clinical evaluation of swallowing disorders includes a number of tools. The Videofluoroscopic Swallowing Study (VFSS) is considered as the preferred instrument by most practicing swallowing clinicians thanks to its ability to allow to visualize bolus flow versus structural motion in the upper aerodigestive tract in real time. The VFSS helps to identify the physiologic causes of the aspiration and clearly identify the symptoms of the patients swallowing disorder [5] [19] and estimate the degree of swallowing impairment allowing clinicians to make an accurate judgment about the sensory or motor impairment [20]. Also, ultrasound imaging is an excellent non invasive opportunity to study the dynamics of the oropharyngeal system and the muscles and other soft tissues of the oropharynx during swallowing [15]. It was used to examine movements of the lateral pharyngeal wall [21], to visualize and examine tongue movement during swallowing [25] [28]. It was also used to evaluate normal and abnormal swallowing by tracking the motion of the hyoid bone [27]. Ultrasound imaging was also used to assess laryngeal phonation function by identifying the morphology of vocal folds and quantifying the tissue horizontal displacement velocity in the vibrating portion of vocal folds [11]. Position of the head during the meal is advised to be leaned forward. Considering this assumption, U. Imtiaz et al. [14] presented a system comprised of an Inertial Measurement Unit (IMU) and an Electromyographic sensor (EMG), able to measure the head-neck posture and submental muscle activity during swallowing. An other study [12] was established to evaluate the influence of two swallowing maneuvers (effortful swallow) on anterior suprahyoid surface through electromyographic measurement and pharyngeal manometric pressure. They found that significant change in both suprahyoid surface electromyographic and pharyngeal pressures during effortful swallow compared with normal swallow. Same researchers [13] have proved that Tongue-to-palate emphasis during execution of effortful swallowing increases amplitudes of submental surface electromyography, orolingual pressure, and upper pharyngeal pressure to a greater degree than a strategy of inhibiting tongue-to-palate emphasis. Swallowing assessment is also supported by acoustic point of view. In 1990, Vice et al. [29] described the detailed pattern of cervical sounds in newborn infants during suckle feeding. As normal swallowing process can be divided into three stages (oral, pharyngeal and esophageal phases), swallowing sound can be divided into three different segments [29] : Initial Discrete Sound (IDS), Bolus Transit Sound (BTS) and Final Discrete Sound (FDS). The opening of the crico-pharynx in pharyngeal phase contributes to the IDS and bolus transition into the esophagus in pharyngeal phase contributes to BTS. Esophageal phase contributes to the FDS. Zahra Moussavi

[22] proved that the swallowing sounds have three stages by applying a Hidden Markov Model (HMM). An Automated method [1] has been set up to separate swallowing sounds from breathing sounds using multilayer feedforward neural networks. The algorithm was able to detect 91.7% of swallows correctly. Lazareck et al. [18] [17] proposed a non-invasive, acoustic-based method to differentiate between individuals with and without dysphagia. Wearable monitoring swallow system was developed [6] in order to assess an overall food and drink intake habits as well as other swallow related disorders. For example, piezo-respiratory belt able to be used for identifying non-intake swallows, solid intake swallows, and drinking swallows is in development.

II.1. Swallowing phases

Swallowing transports food from the mouth to the stomach while protecting the airways. Classically, it is described in three phases: oral, pharyngeal and esophageal. Swallowing is initiated by the introduction of the food into the oral cavity and its preparation by the process of mastication and insalivation. Once the bolus is formed and is ready to be expelled to the pharynx, the tip of the tongue is raised up and applied against the alveolar ridge of the upper incisors and the tongue takes the form of a spoon where the bolus slips and forms a single mass. The latter moves backwards by the moving tongue which gradually applies to the palate from front to back. At this moment, the soft palate realizes the closure of the oropharynx and prevents the penetration of the bolus into the pharynx, however, the larynx is still open. By the time the bolus penetrates the throat isthmus, the oral time (voluntary time) is over. The back of the tongue moves forward and forms an inclination allowing the bolus to move towards the oro-pharyngeal cavity. Thus, the pharyngeal phase is triggered by the contact of the bolus with the sensory receptors of the throat isthmus and of the oro-pharynx.

The pharyngeal process is a continuous phenomenon in time, considered as a reflex accompanied simultaneously by the velo-pharyngeal closure by the velum, by the laryngeal occlusion assured by the elevation of the larynx, and by the retreat of the tongue's base, the movement at the bottom of the epiglottis, the pharyngeal peristalsis and finally the opening of the upper sphincter of the esophagus allow the passage of the food bolus into the esophagus. This phase lasts less than one second. The opening of the upper sphincter of the esophagus is initiated by the arrival of the pharyngeal peristalsis and passage through the esophagus is ensured by the continuity between the pharyngeal and the esophageal peristalsis.

The pharyngeal stage triggered by the end of the oral stage becomes involuntary and is called the swallowing reflex. Velum is raised to close the nasal cavities and avoids any passage of food into the nose and facilitates the passage

of the bolus downwards towards the esophagus through the pharynx. Precisely at this moment passage of food into the trachea is avoided. The larynx open during chewing, to allow breathing, is closed as soon as the bolus arrives on the base of the tongue. At the same time, the vocal cords are embracing ensuring airway closure, the moving cartilages of the larynx (arytenoids) swing forward in the laryngeal vestibule, covered by the rocking of the epiglottis. The larynx is pulled up and down by the hyoid muscles, which places it under the protection of the base of the tongue.

As a result, breathing is interrupted and at the same time the last stage of swallowing begins with the introduction of the bolus into the esophagus and progresses through the esophageal peristaltic waves (muscle contractions) to the stomach. The esophagus is indeed a muscular tube, the two extremities of which are sphincteric zones: the upper esophageal sphincter and the lower esophageal sphincter acting as a valve preventing reflux of food or acidity from the stomach.

The pharyngeal and esophageal phases constituting the reflex of swallowing, without voluntary control [8] [9] [15]. The swallowing reflex is the result of the laryngo-pharyngeal activity, which is triggered by intrapharyngeal and sometimes laryngeal stimuli and is introduced in humans during the fourth month of intrauterine life and observed by ultrasound from the 6th month [9] in order to protect airways from any passage of food.

Given the large number of dysphagic patients in the stroke population of up to 76% [30], this thesis topic from the e-sWallhome project (ANR project) participates in the monitoring of patients who have suffered a stroke proposing automatic methods of monitoring and evaluation of the functional rehabilitation of swallowing disorders. Using different sensors which are detailed below in section III, we must be able to both detect in real time swallowing events and cases of distress of the person during food intake such as aspiration, asphyxiation accompanied or not with falls. In this paper, we propose a GMM model adapted through EM algorithm for classification of sound events that may be present during food intake.

III. Materials

In this project, non-invasive equipment is used to simultaneously record sounds associated with the swallow and also environmental sound, respiratory habits and electrical activity of neck muscles during food intake. Sounds were recorded using discrete module of miniature omni-directional microphone capsule (Sennheiser-ME 102) with IMG Stageline MPR1 Microphone Pre-Amplifier over supra-sternal notch able to record simultaneously swallowing sound and ambient environmental sound. Sounds were recorded at 44.1 KHz sampling rate. Respiratory signals were recorded using a respiratory plethysmographic induc-

tive (RPI) Visuresp system (RBI, France) for measuring respiratory volume in both spontaneously ventilated subjects and patients receiving ventilatory assistance (at home or in a hospital setting). Visuresp is in the form of a vest connected to an oscillator. The both are connected to a signal processing box. This housing is itself connected to a PC computer through a CAN card (Digital Analog Conversion). Respiratory signals were recorded with a sampling rate of 40 Hz. Electrical activity of neck muscles were recorded using a grid of 64 monopolar surface EMG electrodes which was evenly spaced and glued over the front of neck region covering infrahyoid muscles. EMG Electrode Conductive Gel was putted on the electrodes before it was glued to the neck. EMG signals were recorded at 2000 Hz sampling rate. Three signals were recorded simultaneously. In this paper, only recorded sound are used.

Subjects and method

In this study, signals were recorded from healthy subjects with no history of swallowing dysfunction. They were recorded from 10 persons (5 females and 5 males) aged between 23 and 34 years. Participants were seated comfortably on a chair. They were told that the equipment would record swallowing sound and respiration signals. During the test, all of them were asked to drink or eat freely to promote an ambient environment. They were fed with thick liquid (compote-100 grams) using teaspoon and a glass of water (200 ml).

IV. Proposed system

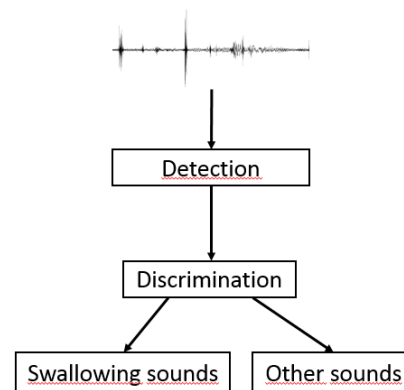


Figure 1. Proposed System

The proposed system allows us to differentiate between swallow sounds and other sounds. The swallowing database acquired in the laboratory and a part of the database used by Mr. Sehili [24] during his thesis were used in this paper. The database is described in the Table 1. It have to be noted that the automatic Detection step is not presented and swal-

Sounds	Number of Samples	Total Duration (s)
Swallowing	188	99,479
Breathing	9	20,487
Cough	8	24,778
Dishes	18	17,336
Speech	16	16,576
Sneeze	9	19,522
Yawn	6	18,541

Table 1. Database

lowing sound segments are manually extracted from each recorded signal. Each sample is then sliced in 16 ms frames with a 50% overlap. Then, for each frame, parameters are calculated and putted into GMM model for recognition.

IV.1. Acoustical Features

Swallowing sound signals (188 signals) were manually extracted from each recorded signal by listening and monitoring the signals in both time and frequency domains. A re-sampling to 16 Khz was not used directly into GMM model. An intermediate step consists in acoustical feature extraction. Features are calculated on a 16 ms window with an overlap of 50%. We use in this study Linear Frequency Cepstral Coefficients (LFCC) and Mel Frequency Cepstral Coefficients (MFCC). MFCCs are very often used in automatic speech recognition. In addition, it was proved [2] that the MFCCs have the most robust characteristics in this domain. LFCCs are mostly used in speaker recognition. We use in addition delta and delta-delta to add temporal aspect. We use also additional parameters which are zero crossing rate (ZCR), roll-off (RF) and centroid using their delta and delta-delta too.

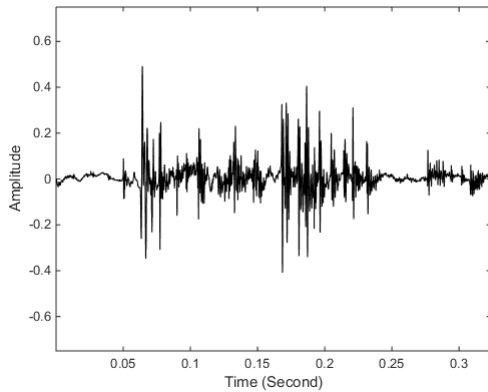


Figure 2. Example of a swallow sound event

IV.2. Principle of the GMM algorithm

Classification has now become an important part of exploratory and decision-making analysis, both in different domains. The exploratory objective is typically to made unsupervised classification or automatic classification. The decision-making purpose of this refers to the supervised classification which aims to discover an hypothetical partition in a set of observations referring to a supervised classification. Accordingly, it is necessary to have methods that adapt to the diversity of the data to be analyzed, which are in our case sounds that can be present during the food intake. Thanks to their flexibility, the finite mixture probability distribution model (Gaussian Mixture Model) correspond to these requirements.

A Gaussian mixture model is a parametric probability density function represented as a weighted sum of M component Gaussian densities as given by the equation below :

$$p(x|\Theta) = \sum_{i=1}^M \pi_i f(x|\mu_i, \Sigma_i) \quad (1)$$

where x is a d -dimensional continuous-valued data vector; in our case it represents feature acoustic vector, $\pi_i, i = 1, \dots, M$, are the mixture weights, and $f(x|\mu_i, \Sigma_i), i = 1, \dots, M$, are the component Gaussian densities. The mixture weights satisfy the constraint that $\sum_{i=1}^M \pi_i = 1$. Each component density is a d -variate Gaussian function which takes the form :

$$f(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu_i)' \Sigma_i^{-1} (x - \mu_i)\right\} \quad (2)$$

which μ_i is the mean vector and Σ_i is the covariance matrix. Gaussian Mixture Model is parametrized by the mean vector, covariance matrix and mixture weights denoted :

$$\Theta = (p_i, \mu_i, \Sigma_i) \quad (3)$$

IV.3. Expectation Maximisation EM-algorithm

Given training acoustic vectors and an initial GMM configuration(θ), we aim to estimate the parameters of the optimal GMM model which corresponds to the distribution of the training features vectors. There are several techniques to estimate model parameters The well established method is Expectation Maximization (EM) an iterative algorithm which its basic idea is to start with the initial model to estimate a new model $\bar{\Theta}$ with maximum likelihood ($p(x|\bar{\Theta}) \geq p(x|\Theta)$). At the iteration, the new model is becoming the initial model. This step is repeated until a convergence threshold ϵ is reached which is expressed by the equation below :

$$L(\bar{\Theta}; x) - L(\theta; x) < \epsilon \quad (4)$$

where L is the likelihood of the model. On each EM iteration, the re-estimation of parameters is done in the following way :

Gaussian Component Weights

$$\bar{\pi}_i = \frac{1}{T} \sum_{i=1}^T P(z|x_i, \Theta) \quad (5)$$

Mean

$$\bar{\mu}_i = \frac{\sum_{i=1}^T P(z|x_i, \Theta)x_i}{\sum_{i=1}^T P(z|x_i, \Theta)} \quad (6)$$

Variance

$$\bar{\Sigma}_i = \frac{\sum_{i=1}^T P(z|x_i, \Theta)(x_i - \bar{\mu}_i)(x_i - \bar{\mu}_i)'}{\sum_{i=1}^T P(z|x_i, \Theta)} \quad (7)$$

where μ_i , x_i and Σ_i refer to the i^{th} elements of μ , X and Σ and z refers to the z^{th} component.

V. Results

GMM model adapted through EM algorithm, described in section IV, was trained using MFCCs and LFCCs coefficients, delta and delta-delta to include the signal temporal variations. Data in our test were divided into two clusters ; the first cluster contains only swallowing samples and the second cluster contains all other samples as described in Table 1. Depending on the small amount of data in this database, we choose leave-one-out algorithm in order to obtain statistically significant results. Firstly, we have trained the model with different features and different GMM orders to find the optimal GMM order. Then, optimal GMM order was set to 6. The best results are obtained with MFCCs features with a recognition rate of 100 %. This high rate can be explained by the small size of the database as well as the absence of noise in the records we have. Our model need to be tested in real environmental conditions.

VI. CONCLUSIONS AND PERSPECTIVES

This paper proposes an automated method to identify swallowing sounds from other sounds in healthy subjects. GMM adapted through EM algorithm was trained with several acoustical features (MFCC, LFCC and their delta and delta-delta). This method achieves a recognition rate of 100 % for swallowing sound.

The outcomes of this study may be utilized for the differentiation of the different phases of swallowing.

ACKNOWLEDGMENTS

This work was funded within the framework of EBIOMED IUIS Chair (Institut Universitaire d'Ingénierie de Santé).

References

- [1] M. Aboofazeli and Z. Moussavi. Automated classification of swallowing and breath sounds. *IEEE*, 3:7803–8439, 2004.
- [2] S. B. Davis and P. Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE*, 28:357–366, 1980.
- [3] H. A. de Santé. *Stratégie de prise en charge en cas de dénutrition protéino-énergétique chez la personne âgée*. 2007.
- [4] H. A. de Santé. *Accident vasculaire cérébral : méthodes de rééducation de la fonction motrice chez l'adulte Méthode Recommandations pour la pratique clinique* . 2012.
- [5] W. J. Dodds, J. A. Logemann, and E. T. Stewart. Radiologic assessment of abnormal oral and pharyngeal phases of swallowing. *American Roentgen Ray Society*, 154:965–974, 1990.
- [6] B. Dong and S. Biswas. Swallow monitoring through apnea detection in breathing signal. *IEEE*.
- [7] G. A. Donnan, M. Fisher, M. Macleod, and S. M. Davis. Stroke. *Lancet*, 371:16121623, 2008.
- [8] V. W. B. et al. *La rhabilitation de la dglutition chez ladulte Le point sur la prise en charge fonctionnelle*. Solal, Collec-tion Le Monde du verbe, 2e dition revue et augmente, 2011.
- [9] M. Guatterie and V. Lozano. dglutition-respiration : couple fondamental et paradoxal , unit de rducation de la dglutition. *kinra*, 42:1–9, 2009.
- [10] R. Hilker, C. Poetter, N. Findeisen, J.Sobesky, A. Jacobs, M. Neveling, and W. D. Heiss. Nosocomial pneumonia after acute stroke : implications for neurological intensive care medicine. *Stroke*, 34:975–981, 2003.
- [11] T.-Y. Hsiao, C.-L. Wang, C.-N. Chen, F.-J. Hsieh, and Y.-W. Shau. Noninvasive assessment of laryngeal phonation function using color doppler ultrasound imaging. *Ultrasound in Med. & Biol.*, 27:1035–1040, 2001.
- [12] M.-L. Huckabee, S. G. Butler, M. Barclay, and S. Jit. Sub-mental surface electromyographic measurement and pharyn-geal pressures during normal and effortful swallowing. *Arch Phys Med Rehabil*, 86:2144–2149, 2005.
- [13] M.-L. Huckabee and C. M. Steele. Electromyographic mea-sures and pharyngeal pressure during effortful swallow. *Arch Phys Med Rehabil*, 87:1067–1072, 2006.
- [14] U. Imtiaz, K. Yamamura, W. Kong, S. Sessa, Z. Lin, L. Bar-tolomeo, H. Ishii, , M. Zecca, Y. Yamada, and A. Takanishi. Application of wireless inertial measurement units and emg sensors for studying deglutition preliminary results. *IEEE*, 978:4244–7929, 2014.
- [15] B. Jones. *Normal and abnormal swallowing : Imaging In Diagnosis And Therapy*. Springer, Second Edition, 2012.
- [16] P. Langhorne, D. Stott, L. Robertson, J. MacDonald, L. Jones, C. McAlpine, F. Dick, G. Taylor, and G. Mur-ray. Medical complications after stroke a multicenter study. *Stroke*, 31:1223–1229, 2000.
- [17] L. J. Lazareck and Z. Moussavi. Classification of normal and dysphagic swallows by acoustical means. *IEEE Transactions on Biomedical Engineering*, 51:0018–9294, 2004.

- [18] L. J. Lazareck and Z. Moussavi. Swallowing sound characteristics in healthy and dysphagic individuals. *IEEE*, 3:7803–8439, 2004.
- [19] J. A. Logemann. Behavioral management for oropharyngeal dysphagia. *Folia Phoniatr Logop*, 51:199–212, 1999.
- [20] B. Martin-Harris and B. Jones. The videofluorographic swallowing study. *Phys Med Rehabil Clin N Am*, 19:769–785, 2008.
- [21] J. L. Miller and K. L. Watkin. Lateral pharyngeal wall motion during swallowing using real time ultrasound. *Dysphagia*, 12:125–132, 1997.
- [22] Z. Moussavi. Assessment of swallowing sounds stages with hidden markov model. *IEEE*, 6:7803–8740, 2005.
- [23] A. G. Rouah. *La prise en charge des troubles de la déglutition en EHPAD, Etude descriptive des pratiques professionnelles de Médecins Coordonnateurs dans 27 EHPAD d'un groupe privé associatif*. 2013.
- [24] M. E. A. Sehili. *Reconnaissance des sons de l'environnement dans un contexte domotique*. PhD thesis, Institut National des Telecommunications, 2013.
- [25] T. H. Shawker, B. Sonies, M. Stone, and B. J. Baum. Real-time ultrasound visualization of tongue movement during swallowing. *J Clin Ultrasound*, 11:485–490, 1983.
- [26] S. Singh and S. Hamdy. review dysphagia in stroke patients. *Postgrad Med J*, 82:383391, 2006.
- [27] B. C. Sonies, C. Wang, and D. J. Sapper. Evaluation of normal and abnormal hyoid bone movement during swallowing by use of ultrasound duplex-doppler imaging. *Ultrasound in Med. & Bml.*, 22:1169–1175, 1996.
- [28] M. Stone and T. H. Shawker. An ultrasound examination of tongue movement during swallowing. *Dysphagia*, 1:78–83, 1986.
- [29] F. L. Vice, J. M. Heinz, G. Giuriati, M. Hood, and J. F. Bosma. Cervical auscultation of suckle feeding in newborn infants. *Developmental Medicine and Child Neurology*, 32:760–768, 1990.
- [30] Z. Zhou. *Accidents Vasculaires Cérébraux (AVC) : Conséquences Fonctionnelles et Dysphagie Associée*. PhD thesis, Université de Limoges, 2009.