David Zhang · Kebin Wu

Pathological Voice Analysis

Springer

Pathological Voice Analysis

David Zhang • Kebin Wu

Pathological Voice Analysis

Springer

David Zhang
The Chinese University
of Hong Kong (Shenzhen)
Shenzhen Institute of Artificial
Intelligence and Robotics for Society
Guangdong, China

Kebin Wu
Huawei Technologies
Beijing, China

ISBN 978-981-32-9195-9 ISBN 978-981-32-9196-6 (eBook)
https://doi.org/10.1007/978-981-32-9196-6

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The human voice contains rich information about the speaker, such as identity, gender, health, situation, and emotion. However, the biomedical value of voice is less addressed compared with its biometrics applications such as speech recognition and speaker identification. In this book, we present the importance and value of voice in aiding diagnosis through pathological voice analysis.

We systematically introduce our research works on pathological voice analysis from the following three aspects: (1) a review on pathological voice analysis and a guideline on voice acquisition for clinical application; (2) design appropriate signal processing algorithms for pathological voice; and (3) extract biomedical information in voice to improve disease detection, such as dictionary-based feature learning and multi-audio fusion. Experimental results have shown the superiority of these techniques. These proposed methods can be used in applications of voice quality assessment, disease detection, disease severity prediction, and even the analysis of other similar signals. This book will be useful to researchers, professionals, and postgraduate students working in the field of speech signal processing, pattern recognition, biomedical engineering, etc. This book also will be very meaningful for interdisciplinary research.

The book is organized as follows: In Chap. 1, the development of pathological voice analysis is systematically reviewed. Then important factors in the acquisition of pathological voice, sampling rate in particular, are discussed in Chap. 2. After that, two of the widely used signal processing steps in pathological voice analysis, which are pitch estimation and glottal closure instant (GCI) detection, are studied in Chaps. 3 and 4, respectively. In Chap. 5, feature learning based on spherical K-means is proposed, in contrast to the traditional handcrafted features. In Chaps. 6 and 7, we investigate multi-audio fusion in pathological voice analysis so as to make full use of the multiple audios collected for each subject. Finally, we summarize the book and give a short introduction to future works that may contribute to the further development of pathological voice analysis in Chap. 8.

Our team has been working on pathological voice analysis for more than 5 years. We appreciate the related grant supports from the GRF fund of the HKSAR

Government, Research projects from Shenzhen Institutes of both Big Data and Artificial Intelligence & Robotics for Society, and the National Natural Science Foundation of China (NSFC) (61020106004, 61332011, 61272292, and 61271344). Besides, we thank Prof. Guangming Lu and Prof. Zhenhua Guo for their valuable suggestions about our research on pathological voice analysis.

Guangdong, China David Zhang
Beijing, China Kebin Wu
April 2020

Contents

Chapter 1
Introduction

Abstract Recently, computer-based pathological voice analysis has become increasingly popular. This chapter first discusses the necessity for the study of pathological voice analysis, and then a systematic review is presented, including the diseases which may lead to pathological voice, current development of pathological voice analysis in the aspects of voice recording, feature extraction, and classification (regression). Finally, the key problems and challenges in pathological voice analysis are discussed. After reading this chapter, people will have some shallow ideas on pathological voice analysis.

Keywords Pathological voice analysis · Disease detection · Disease monitoring

Recently, computer-based pathological voice analysis has become increasingly popular. This chapter first discusses the necessity for the study of pathological voice analysis, and then a systematic review is presented, including the diseases which may lead to pathological voice, current development of pathological voice analysis in the aspects of voice recording, feature extraction, and classification (regression). Finally, the key problems and challenges in pathological voice analysis are discussed. After reading this chapter, people will have some shallow ideas on pathological voice analysis.

1.1 Pathological Voice Analysis

In this book, voice refers to all possible sounds produced by the human vocal system, such as vowels, continuous speech, and coughing sounds. While studying voice has always been a great interest for researchers from various fields, we mainly focus on the voice analysis in the fields of machine learning and signal processing in this book.

In recent years, speaker and speech recognition are two main applications that adopt machine learning technique on voice. In speaker recognition, voice

D. Zhang, K. Wu, *Pathological Voice Analysis*,
https://doi.org/10.1007/978-981-32-9196-6_1

pronounced by a speaker is termed as a biometric signal including identity related characteristic so that it can be used to recognize the identity of the speaker (Kinnunen and Li 2010). Speech recognition, however, is to recognize and translate spoken language into text (Rabiner 1989). In both cases, signal processing (such as de-noising (Milone et al. 2010) and pitch estimation (Atal 1972)) and machine learning technique (Miro et al. 2012; Graves et al. 2013) play significant roles.

Unlike speaker and speech recognition which has been widely studied and even adopted in industry extensively, the biomedical value of voice is less emphasized. The production of voice requires the cooperation of multiple organs: (1) the nervous system that coordinates the operation of various tissues and organs in the process of voice production; (2) the respiratory system that provides energy, including lungs and tracheas; and (3) the vocal cords and vocal tracts that function as vibrators and resonators, respectively. When a certain disease affects either of the three abovementioned system directly or indirectly, pathological voice may be generated. For example, studies have shown that voice quantitative analysis can be used to detect neurological diseases such as Alzheimer's disease (AD) (López-de-Ipiña et al. 2013), Parkinson's disease (PD) (Tsanas 2012), and stroke (Le et al. 2014). When suffered from respiratory diseases such as lung cancer and chronic obstructive pneumonia, patients may catch on symptoms like dysphonia and cough (Lee et al. 2008). In addition, organic lesions occurring in vocal cord and tract can lead to hoarse voice. For example, Saudi et al. classified healthy voice and voice collected from patients with vocal cord nodules, abscesses, polyps, paralysis, edema, and vocal cord cancer with an accuracy of 92.86% (Saudi et al. 2012). Therefore, pathological voice analysis can be used to help diagnose (even monitor) the healthy status of vocal system.

In clinical applications, however, there exist many alternative methods to diagnose diseases that affect voice. Hence, the necessity of pathological voice analysis is furthermore clarified. Firstly, comparing with common medical signals such as electroencephalogram (EEG) and electrocardiogram (ECG), voice signal collection does not need special acquisition devices. For instance, Little et al. of Oxford University launched a PVI (Parkinson's voice initiative) to collect voices of 10,000 Parkinson's patients worldwide in order to realize Parkinson's disease detection and monitoring based on voice analysis. In this initiative, the signal acquisition device is by telephone which is quite ubiquitous. Secondly, pathological voice analysis is user-friendly, painless, and non-invasive. Traditional laryngoscopy in diagnosis inserts a tubular endoscope into the larynx through the mouth to observe the interior of the larynx, which may cause nausea and vomiting in patients and lead to lesion due to the friction between instruments and skin, mucosa, tissues, and organs. Besides, unqualified sterilization and anesthesia are also high-risk factors. In contrast, collecting voice is a non-invasive and painless process so that patients may be more actively involved in voice acquisition. Thirdly, voice analysis may be the key method for the early detection of some certain diseases. Rusza et al. found that most patients with early Parkinson's disease had some degree of speech disorder and it often appeared earlier than other symptoms like dyskinesia (Rusz et al. 2011). At present, the cause of Parkinson's disease (PD) is not completely understood by researchers and there is no radical cure for the disease. If patients can get treatment at the early stage of PD onset, the development of the disease can be delayed (King et al. 1994). Therefore, if people with high-risk of PD can be screened regularly based on voice analysis, it is possible to detect the disease as soon as possible so as to obtain early treatment. Fourthly, disease diagnosis and monitoring by voice analysis is automatic and objective. In recent years, the traditional diagnosis and monitoring of Parkinson's disease is mainly decided by a doctor subjectively based on the patient's medical history, self-description of symptoms, and the reaction to certain drugs (such as levodopa). Besides, brain computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) are also employed to exclude some similar diseases. Obviously, the widely used diagnostic method relies too much on doctors, and thus the diagnostic results are possible to be affected by doctors' medical skills, emotions, and physical state. In contrast, objective voice analysis can improve the reliability of initial diagnosis and follow-up: in diagnosis, objective measurement and analysis of patient's voice offer more accurate information so as to reduce the rate of missed diagnosis and misdiagnosis; in follow-up, objective voice analysis method provides the basis for doctors to adjust the patients' dosage reasonably. For example, Tsanas et al. established a Parkinson's disease detection model based on voice analysis and obtained an accuracy of 98.5 ($\pm$2.3%) (Tsanas 2012). At the same time, the author extended the application of voice analysis to the severity prediction of PD (Tsanas 2012). In addition, Wang Jianqun's research showed that the objective analysis of pathological voice can also be used to effectively evaluate the effect of surgery (Wang et al. 2004). Fifth, voice analysis can promote the development of telemedicine. Traditional diagnostic methods require face-to-face contact between patients and doctors and patients need to visit hospitals frequently for follow-up, which will bring great inconvenience to the elderly, especially those with disabilities such as PD and stroke. On the contrary, the objective voice measurement does not require the direct participation of doctors so that patients are expected to get diagnosis and severity prediction at home, eliminating the inconvenience of visiting hospitals. Moreover, the convenience of objective voice analysis enables patients to make quantitative measurements at home many times a day. The timely and frequent measurements enable doctors to give more appropriate and reasonable treatment plans accordingly. Sixth, voice analysis is highly cost-effective. The low cost of voice acquisition and analysis makes it possible for more patients to obtain early detection, early treatment, and regular follow-up. Finally, voice analysis may improve the life quality for patients. It is well known that patients with pathological voice often present certain obstacles in communication, which affect their quality of life. Song et al. analyzed the spectral characteristics of pathological voice and proposed a voice enhancement algorithm to reduce the wheezing in patients' voice without losing biometric characteristics (Song et al. 2013).

1.2 Computerized Voice Analysis

The biomedical significance of voice is less emphasized compared with its other applications, such as speaker recognition and speech recognition. This section presents a review of the computerized voice analysis in biomedical field so that the biomedical value of voice can be reemphasized. Firstly, we show by a literature review that many diseases result in pathological voice and these diseases are categorized into three classes based on the voice production mechanism: nerve system diseases, respiratory system diseases, and diseases in vocal folds and vocal tract. Secondly, the three steps in computerized voice analysis, which are voice recording, feature extraction, and classification and regression, are reviewed separately to show the current development of voice analysis in the biomedical field. Finally, the potential challenges to be addressed in computerized voice analysis are discussed in terms of data level and algorithm level. Some suggestions concerning future research direction are also presented. Computerized voice analysis is a promising complementary tool for disease diagnosis and monitoring; however, there are several inevitable challenges to be handled before it is put into practical use.

1.2.1 Introduction

Health related studies are bringing along a growth of attention from governmental agencies, charities, and companies since diseases pose increasing social and economic burdens on society (The US Burden of Disease Collaborators 2018; Milken Institute 2018). Technologies from different fields are often adopted to assist the research. One of them is to utilize the modern computerized technique since it possesses two advantages: (1) automation that helps to reduce manual labor and (2) objectiveness which enables the detection and monitoring of disease less dependent on the doctors' expertise. This section presents a literature review of biomedical analysis of voice that is based on computerized technology. Even though voice is one kind of biomedical signal, its biomedical value is far less emphasized than its other applications, such as speaker recognition and speech recognition. Hence, this review aims to reemphasize the significance of voice in the biomedical field and discuss the accompanied challenges.

As shown in Titze (1994) and (Mekyska et al. 2015), the term voice can be defined in both a broad and a narrow sense. While it refers to the sound in which the vocal folds vibrate in the narrow sense, voice may be taken as synonymous of speech in the broad sense. In this section, we take the broad definition and sounds pronounced by a human being, such as simple vowels, continuous speech, and cough, are all included. Voice, as a biomedical signal, is generated under the cooperation of multiple systems and diseases in these systems often show abnormality in voice. The commercial database MEEI (Massachusetts Eye and Ear Infirmary) (Elemetrics 1994) developed by Massachusetts Eye and Ear Infirmary Voice & Speech Lab is frequently exploited to analyze pathological voices. In this dataset, voices of laryngeal diseases are included, such as vocal fold polyp, adductor spasmodic dysphonia, keratosis leukoplakia, vocal nodules, and vocal fold paralysis. In recent decades, studies suggest that voice analysis can also be used to detect and monitor some neurological disorders, such as Alzheimer's disease (AD) (Lopez-de Ipina et al. 2013) and Parkinson's disease (PD) (Tsanas et al. 2010, 2012; Little et al. 2009; Rusz et al. 2011; Mandal and Sairam 2013; Tsanas 2012; Little 2007). Diseases related to respiratory system can also result in pathological voice. In Lee et al. (2008), it was pointed out that 90% of lung cancer patients were perceived as dysphonic. Finally, pathological voice can be generated indirectly. Renal failure, for example, affects voice by generating disturbances in the respiratory system (Kumar and Bhat 2010; Hamdan et al. 2005; Jung et al. 2014).

In spite of these studies, there are few summaries to clarify the biomedical value of voice. In this work, we aim to reemphasize the biomedical value of voice. Firstly, a review of the existing literature is provided to demonstrate that lots of diseases lead to abnormal voices. Besides, these diseases are classified into three categories according to the mechanism of voice production. Secondly, three steps in computerized voice analysis in biomedical applications are reviewed separately, offering a basic summary of its present situation. Thirdly, we discuss the potential challenges to be handled in computerized voice analysis from both data level and algorithm level. Additionally, we also present some suggestions concerning future research direction on computerized voice analysis.

The remainder of this section is organized as follows. In Sect. 1.2.2, a literature review is presented showing that many diseases affect voice and these diseases can be categorized into three classes, according to the mechanism of voice production. In Sect. 1.2.3, we review the present situations for the three steps in computerized voice analysis separately. Section 1.2.4 presents the discussion of potential challenges in the computerized voice analysis from both data and algorithm level. Suggestions concerning future research directions are also included. Conclusions are given in Sect. 1.2.5.

1.2.2 Biomedical Value of Voice

Recently, there have been more and more studies analyzing the pathological voice, as can be seen in Fig. 1.1. Various diseases can lead to pathological voice. In this section, the biomedical value of voice is reemphasized by literature review and the diseases resulting in pathological voice are categorized. The reviewed literature use voice as a biomedical signal and adopt computer-aided technology to implement analysis.

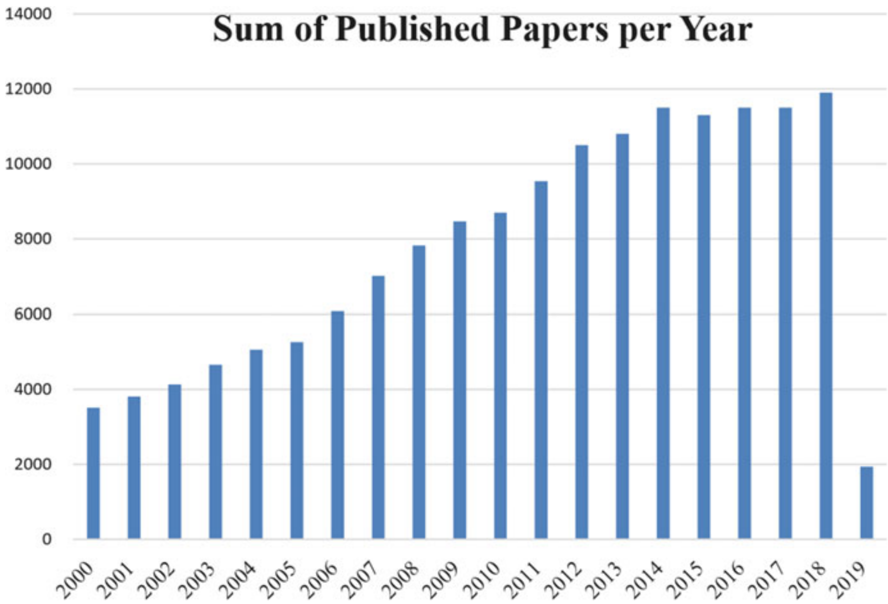

Fig. 1.1 The number of published papers on the topic of pathological voice analysis. The numbers were obtained by searching "pathological voice analysis" in Google Scholar on March 22, 2019

1.2.2.1 Mechanism of Voice Production

Before the review, we present a brief description of how voice is produced, aiming to show the organs involved in phonation.

Air flowing is an essential part of voice production. The airflow first originates from lungs, goes up along trachea to larynx, and then exits from mouth and nose. Systems involved in this process can be classified into three categories depending on their functions: (1) The nerve system coordinates the interactions of involved systems in voice production: (2) The respiratory system, including lungs and trachea, is regarded as the energy source (Tsanas 2012; David 2010); (3) vocal cords and vocal tract are called as vibrator and resonator, respectively. When producing a voiced sound, vocal cords are driven by airflow from respiratory system to open and close repeatedly, forming vibrations. Vocal tract, which is assumed as a tube formed by several anatomical segments: pharynx, oral cavity, nasal cavity, and soft palate changes its shape flexibly with different syllables to form resonance (Titze 1994).

With the coordination of the three abovementioned systems, voice can be produced. When a disease occurs in these systems, pathological voice may be generated.

1.2.2.2 Voice and Diseases

Diseases that affect voice can be grouped into three classes, according to the categorization of involved systems in voice production in Sect. 1.2.2.1.

Voice and Diseases in Nerve System

In this category, three typical diseases affect voice, which are Alzheimer's disease (AD), Parkinson's disease (PD), and stroke.

AD patients are often featured with progressive cognitive deterioration. Most of them have difficulty in communicating and their emotional responses are impaired. These characteristics show signs in voice. Therefore, it was proposed to extract quantitative features from voice for early diagnosis of AD. For instance, the Automatic Spontaneous Speech Analysis (ASSA) and Emotional Response Analysis (ERA) on speech were employed to diagnose Alzheimer's disease (AD) as well as to predict its severity degree in Lopez-de Ipina et al. (2013).

PD is another neurodegenerative disorder resulting in voice impairment and there have been varieties of literature showing that voice analysis is one effective non-invasive way for PD diagnosis (Tsanas et al. 2010, 2012; Little et al. 2009; Rusz et al. 2011; Mandal and Sairam 2013; Tsanas 2012; Little 2007) and monitoring (Tsanas et al. 2011; Wang et al. 2016; Arora et al. 2015). As pointed out in Tsanas et al. (2012), PD leads to the malfunction of nerve systems, making the cooperation of organs in voice production less appropriate and delicate. The produced voice is hoarse, breathy, less articulated and the prosody aspect of voice is shown as monoloudness, reduction of stress, and monopitch (Schulz and Grant 2000). Based on this observation, Tsanas (2012) conducted voice analysis to detect PD and the accuracy rate reported was $98.5 \pm 2.3\%$. It should be highly stressed that abnormality in the voice is often shown earlier than other symptoms (Harel et al. 2004), making voice analysis more important for the early diagnosis of PD. In addition, Tsanas also made progresses in predicting the symptom severity of PD, which is often described by the Unified Parkinson's Disease Rating Scale (UPDRS). The reported mean absolute error in predicting UPDRS by regression based on voice features was 1.59 ± 0.17 (UPDRS ranges from 0 to 176) (Tsanas 2012). In Garcia et al. (2017), the i-vector framework was utilized to model the speech of PD patients and thus to predict their neurological state. These experimental results show that it is promising to use computerized voice analysis for PD detection and monitoring. Actually, there is a project, named as Parkinson's Voice Initiative (Tsanas and Little 2012), aiming to collect large numbers of voice samples from healthy volunteers and PD patients. With this large database, the model of voice analysis can be further optimized for PD detection and monitoring. Readers may refer to the work in Benba (2016) for another review on voice based PD assessment.

Stroke disease often leads to aphasia. Patients with aphasia have difficulty in communication, i.e., speaking rate is slow and the speech length tends to be short.

Thus, speech signals have been employed for stroke monitoring. In Le et al. (2014), a system was devised to assess the speech quality of patients with aphasia objectively, using acoustical measures such as voiced duration and the rate of clear speech over all speech. Similarly, another automatic system for detecting voice disorders, particularly for stroke, was presented in Brauers et al. (2006). This system was designed for home use, making it especially useful for patients during stroke rehabilitation and those with suspicious symptoms of stroke.

A patient with any of these diseases often suffers from a shorter life span and impacted life quality. When literature evidence showing that automatic voice analysis is promising for the detection and monitoring of these diseases, voice analysis needs more attention.

Voice and Diseases in Respiratory System

A malfunctioning respiratory system shows signs in voice. For instance, cough is a common symptom and its frequency, strength, and other characteristics of cough can be used to screen respiratory diseases (Shrivastav et al. 2014). In Lee et al. (2008), it was pointed out that 90% of the lung cancer patients are perceived as dysphonic. An acoustical analysis for the chronic obstructive pulmonary disease (COPD) was carried out in Shastry et al. (2014) and the experimental results showed that the values of acoustical measures extracted from patients with COPD and healthy subjects, respectively, were significantly different.

In addition to diseases affecting respiration systems directly, illness originating from other organs can influence the produced voice indirectly. One main concern is the renal failure disease, which leads to disturbances in the pulmonary system and the voices of patients often have increased pitch (Kumar and Bhat 2010). Experiments demonstrated that the acoustical differences between patients with renal failure and healthy controls were significant and the underlying pathophysiology for the differences was explained in Kumar and Bhat (2010). In Hamdan et al. (2005) and Jung et al. (2014), the effects of hemodialysis, which is a typical treatment for renal failure, on voice were investigated. Cystic fibrosis (CF) disease is another disease affecting voice indirectly and it generates drastic respiratory symptoms such as short breathiness and chronic coughing. In Louren et al. (2014), experimental results demonstrated that acoustical features measuring deviations of vocal parameters display significant differences between CF patients and the control group.

Since dysfunction of the respiration system often causes chronic cough, shorter breath, and even death in severe case, an effective and non-invasive monitoring tool will bring convenience. The studies indicate that acoustical analysis may be one promising solution.

Voice and Diseases in Vocal Folds and Vocal Tract

Diseases in vocal folds and tract have direct impacts on voice. In fact, studies that use voice for biomedical purposes are primarily focused on diseases in this category. Besides, computerized voice analysis system has been put into practical use in the ENT (ear, nose, and throat) department of many hospitals.

Among all systems involved during voice production, vocal folds are the most sensitive tissue (Arjmandi and Pooyan 2012). It was pointed out in Akbari and Arjmandi (2014) that vocal folds, along with vocal tract, are the most important components so that voice can be highly influenced by diseases in these systems. In Campisi et al. (2000), the voice pattern for patients with vocal cord nodules was analyzed. It was demonstrated in Oguz et al. (2007) that unilateral vocal cord paralysis, another disease in vocal folds, had effects on acoustic measures. Besides, many literature treat voices of multiple diseases as a single class and attempt to distinguish them from normal voices (Dibazar et al. 2002; Godino-Llorente et al. 2006; Henrıquez et al. 2009; Markaki and Stylianou 2009, 2011; Parsa and Jamieson 2000). Study in Saudi et al. (2012) showed that healthy subjects could be discriminated from patients with six vocal fold diseases, including cyst, polyps, nodules, paralysis, edemas, and carcinoma by acoustical analysis, with an accuracy of 92.86%. In Akbari and Arjmandi (2014), three other diseases affecting the vibration function of vocal folds were analyzed. More types of vocal folds diseases resulting in pathological voices can be found in the widely used database (MEEI Elemetrics 1994). Literature concerning voice analysis for diagnosis of vocal folds diseases fall into two classes. On one hand, some aim to screen pathological voices by treating voices of several vocal folds diseases as one group and comparing extracted acoustical measurements with that of normal voices (Dibazar et al. 2002; Godino-Llorente et al. 2006; Henrıquez et al. 2009; Hadjitodorov and Mitev 2002; Vikram and Umarani 2013; Boyanov and Hadjitodorov 1997; Saeedi et al. 2011; Arjmandi et al. 2011). On the other hand, there are some research attempting to use voice to identify the type of vocal folds diseases (Markaki and Stylianou 2009; Arjmandi and Pooyan 2012; Jothilakshmi 2014; Alsulaiman 2014; Cavalcanti et al. 2010). Arjmandi and Pooyan presented an algorithm to classify six different voice disorders: paralysis, nodules, polyp, edema, spasmodic dysphonia, and keratosis (disease in the vocal tract) (Arjmandi and Pooyan 2012). Likewise, the automatic system developed in Jothilakshmi (2014) was designed to classify ten types of diseases by voice analysis.

Three main elements determining the shape of vocal tract are tongue, nose cavity, and oral cavity. Disease in these three places can lead to unnatural shape change of vocal tract and thus abnormal voice is generated. Oral, head, and neck cancer (OHNC), which includes any cancer starting in the upper digestive tract (Zhou et al. 2012), often affects vocal tract severely and consequently leads to degraded speech intelligibility. Experiments in Maier et al. (2009) proved that the speech intelligibility of OHNC patients was significantly worse than that of the control group by analyzing with an automatic speech recognition system (ASR). Besides, there are investigations implemented to compare the acoustical parameter changes