Chapter 14
Voice Conservation: Towards Creating a Speech-Aid System for Total Laryngectomees

Zdeněk Hanzlíček, Jan Romportl, and Jindřich Matoušek

Abstract. This paper describes the initial experiments on voice conservation of patients with laryngeal cancer in an advanced stage. The final aim is to create a speech-aid device which is able to “speak” with their former voices. Our initial work is focused on applicability of speech data from patients with an impaired vocal tract for the purposes of speech synthesis. Preliminary results indicate that appropriately selected synthesis method can successfully learn a new voice, even from speech data which is of a lower quality.

14.1 Introduction

Speech is a fundamental mean of human communication. However, healthy people usually do not fully appreciate how significant is the ability to speak with own natural voice for the social contact and interaction.

The human vocal tract is a complex and vulnerable system. Its damage can cause various problems with speech production. The damage can be related to an injury or it can be also an inevitable consequence of the treatment of another serious disease, such as mouth or neck cancer.

An example of such a radical treatment is the laryngectomy surgery. This medical intervention is performed on patients with laryngeal cancer when other less invasive types of treatments (e.g. radiation or chemotherapy) fail or are not possible. According to the extent of the carcinoma, various sections of larynx are necessary to
be removed. In the case of so-called total laryngectomy, the removal of the whole larynx is performed. Patients who underwent this surgery are called laryngectomees.

An important part of larynx is the epiglottis – a valve which separates the respiratory and digestive tract during deglutition (swallowing). It protects against breathing in of swallowed food or fluid. This function of epiglottis cannot be reliably substituted. Thus, a permanent surgical separation of both tracts is performed during the laryngectomy surgery. After that, the laryngectomee breathes through a stoma – an opening in the trachea.

Another significant consequence of this surgery is the inability to produce speech in the common manner. It has two main reasons:

- Vocal folds were removed together with the larynx during laryngectomy. The vibrations of vocal folds during the expiration form an excitation component of speech that is then amplified and articulated into speech in the following parts of the vocal tract (primarily the oral cavity).
- Due to the detachment of the respiratory tract, the air from lungs cannot be passed through the mouth.

However, laryngectomees can learn and use several alternative ways of producing speech sounds:

- For producing **tracheo-esophageal speech**, a special voice prosthesis need to be placed between trachea and esophagus that were surgically separated during the laryngectomy. This prosthesis contains a one-way valve that allows the air to flow from lungs into the oral cavity. When the patient wants to talk, the tracheostoma has to be plugged and the exhaled air is pushed through the valve into the mouth where it is articulated into speech. The friction of air passed through the valve produces some vibrations and simulates the function of vocal folds. In principle, this method is similar to the natural way of speech production. The prosthesis needs to be regularly maintained and cleaned.
- **Esophageal speech** is produced solely by the upper part of digestive tract. First, the air is swallowed into the esophagus. Then, it is pushed back into the oral cavity for articulation. The produced speech resembles the belching. This method is very exacting due to the low capacity of esophagus.
- During production of **electrolaryngeal speech**, the function of vocal folds is substituted by an external device (electrolarynx) which is put to the neck where it produces intense mechanical vibrations. These vibrations are transmitted into the oral cavity while the speaker articulates.

All the aforementioned kinds of speech (also called alaryngeal speech) suffers from lack of naturalness and intelligibility. Although, the voice of two laryngectomees can slightly differ according to the various parameters (e.g. proportions) of the vocal tract, the new voice does not bear the former speaker identity. An important speech feature for distinguishing of speaker identity is the fundamental frequency $f_0$, i.e. the frequency of vocal folds oscillation.

Moreover, the changes and dynamics of $f_0$ are the essential means for expressing the prosody of an utterance. Thus, an important consequence of removing vocal