

Bionic Voice: Natural Speech Restoration for Voice Impaired Individuals (Pilot Study)

Hamid SHARIFZADEH ¹, Jacqui ALLEN ², Abdolhossein SARRAFZADEH ³, Iman ARDEKANI ⁴

1. Unitec Institute of Technology, New Zealand
2. Auckland Voice and Swallow, New Zealand
3. Unitec Institute of Technology, New Zealand
4. Unitec Institute of Technology, New Zealand

Keywords: Laryngectomy; Voice Reconstruction; Artificial Phonation; Speech Synthesis

INNOVATION

The human voice is the most magnificent instrument for communication, capable of expressing deep emotions, conveying oral history through generations, or of starting a war. However, those who suffer from aphonia (no voice) and dysphonia (voice disorders) are unable to make use of this critical form of communication. They are typically unable to project anything more than hoarse whispers.

Epidemiologic studies of the prevalence of voice disorders in the general adult population are rare. Nevertheless, information from a number of studies suggests that one third of the population have suffered from a temporary vocal impairment at some point in their life and that voice disorders can affect any age group and either sex. In some cases, vocal change is temporary however in those treated for malignant disease or with severe trauma there may be long term disturbance of phonation. This may affect occupation, social function and quality of life.

Within a speech processing framework, we have worked on a novel method to return natural voice to laryngectomised people. This method leverages on recent advances in speech synthesis to deliver a world-first technology. As a pilot study, this project has assessed the acoustic features of laryngectomised speech and has developed required enhancement for natural speech regeneration.

CONTEXT

Whispered speech is useful for quiet and private communications in daily life but becomes the primary communicative mechanism for many people experiencing voice box difficulties. Patients reduced to whispering have generally lost their pitch generation mechanism through physiological blocking of vocal cord vibrations or, in pathological cases, blocking through disease or exclusion by an operation. Typical prostheses for voice impaired patients (oesophageal speech, transoesophageal puncture (TEP), and electrolarynx devices) allow patients to regain limited speaking ability but do not generate natural sounding speech; at best their sound is monotonous or robotised. Additional drawbacks of traditional prostheses are difficulty of use and risk of infection from surgical insertion.

SOLUTION

We have taken the first steps towards the development of a non-surgical, non-invasive alternative method which combines whisper signal analysis with direct pitch insertion and formant enhancement, to reconstruct missing speech and return natural voice to laryngectomised individuals. This pilot study assesses the acoustic features of laryngectomised speech while required enhancement for natural speech regeneration is considered. Our new method implements a speech processing framework which turns the whisper-like voice samples of patients into normal speech.

IMPLEMENTATION

Audio recordings were made of five laryngectomised patients in a sound proof room reading from three randomisations of a balanced word list. Formant contours of speech samples in /hVd/ structure were analysed through combined segmentation-extraction methods. Formant characteristics were used to establish a vowel formant space for laryngectomised speech. Data was translated by spectral enhancement and pitch insertion algorithm into phonated words.

Built upon a parametric speech synthesis algorithm developed by the team, the method converts distorted vowels of laryngectomees into normal speech based on sine wave formant regeneration and artificial pitch modulation. To derive an artificial (but normal sounding) pitch pattern, this method relies upon a harmonic relationship between formant frequencies and pitch period. Then this harmonic is used for modulation of sine wave formants and modifying formant frequencies accordingly.

OUTCOMES

All audio recordings were able to be analysed. Vowel formant information was adequately produced in all samples. Frequency of the first two formants were measured and were comparable to corresponding formant frequencies in phonated speech. Reconstructed words were recognisable to naive listeners.

NEXT STEPS

As a pilot study, this project was focused on a small number of patients with very limited vocabulary size for reconstruction. We have achieved some promising results within early steps thus, the next stages would be: a) migrating to a broader range of speech samples and patient groups, b) implementing the reconstruction algorithm on a digital signal processor unit equipped with a miniature Bluetooth microphone and speaker that can be attached to clothing.

Correspondence:

Hamid Sharifzadeh

Unitec Institute of Technology

hsharifzadeh@unitec.ac.nz