Restoration of Natural Prosody in Pathological Voices

F. N. Reale, R. Vetter, J. M. Solà I Caros, P. Renevey

The intelligibility and quality of pathological voices is often hindered by insufficient vocal power and insufficient stability of too low a pitch. In order to overcome these deficiencies, a voice restoration device based on a multi-resolution approach is proposed. Preliminary tests have shown promising results in terms of listener perceived speech quality and intelligibility.

The degraded fundamental characteristics of pathological voices often engender a decrease in a patient’s speech intelligibility and thereby a limitation in his social oral interaction. To overcome these deficiencies, a voice restoration device which aims at improving the speech quality and intelligibility through advanced signal restoration techniques was recently proposed. Herein, focus is on the restoration of natural prosody. Prosody is the set of patterns of stress and intonation in speech. From the acoustical point of view, prosody relates to variations in pitch (tonal height), loudness and syllable length. Prosody is highly important to emphasize words or sentences and to express emotions like surprise and anger. Thus, prosody appears as the principal cue which distinguishes human and robot-like voicing. In pathological voices, degradation in prosody may be observed in terms of too low a pitch and a leak in pitch stability. Often it would be even more appropriate to speak of a pitch breakdown, since a periodical voice generation is sustained only intermittently.

Laryngectomy, for instance, is a surgical intervention that consists in partial or total removal of the larynx to treat laryngeal cancer. After laryngectomy, rehabilitation techniques allow the patients to re-acquire the ability to speak. However, prosody is highly degraded due to the lack of vocal cords, which constitutes in healthy subjects a high quality acoustical oscillator.

The project aim is to develop a portable device with integrated microphones, signal processing facilities, amplifier and loudspeakers. Its task is to process and improve pathological voices in real time. To this end, the glottal excitation and articulation information of the pathological voice are separated in a primer signal analysis. The pathological excitation is then replaced by concatenation of randomly chosen glottal waves from a reference database extracted from healthy voices. In order to restore natural prosody and avoid a robotic monotonic voice, the interval between subsequent glottal waves is obtained trough a multi-resolution approach.

Some classes of pathological rehabilitated laryngectomisees’ voices constitute therefore a primary application field of a natural prosody voice restoration device.

This device is under development in the context of a joint project between the Swiss Center for Electronics and Microtechnology (CSEM), the University of Applied Science of West Switzerland (HEIG-VD) and the ORL laboratory of the University Hospital of Lausanne (CHUV). It is financially supported by the Swiss League against Cancer and by the Gebert Rüf Stiftung.

Figure 1: The signal is separated in articulation information and excitation. The excitation of the voiced frames is then restored using the prosody multi-resolution approach

Figure 2: Correlation between pitch and signal variance

Short term pitch variability is reproduced through a statistical variation model. The model is estimated from short term pitch variations of healthy subjects. For middle term pitch variability the correlation at the middle term time scale between pitch and signal variance highlighted in recent studies is exploited. Therefore, middle term variations are estimated from a bi-dimensional frequency-energy approach. Eventually, long term variability is obtained through statistical clustering of instantaneous frequency estimations such as to discard influences of large pitch variations or pitch breakdowns (Figure 1).

The results obtained during a preliminary study (Figure 2) show a significant improvement of the quality of restored voices according to the opinions of medical and speech processing experts and clearly demonstrate the feasibility of the proposed approach.

