Restoration of natural prosody in pathological voices

F.N. Reale, Ph. Renevey, J. Sola I Caros and R. Vetter - Swiss Center for Electronics and Microtechnology (CSEM), Neuchâtel, Switzerland

The intelligibility and quality of pathological voices is often hindered by insufficient vocal power and an abnormally low and unstable fundamental frequency. 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 speech quality and intelligibility perceived by the listener.

Motivation
The degraded fundamental characteristics of pathological voices often engender a decrease in a patient’s speech intelligibility and thereby a limitation in their social oral interaction.

Objectives
• Improvement of quality of pathological speech
• Portable system, real time processing
• Robust estimation of fundamental frequency in pathological voices

Context
Prosody relates to variations of the laryngeal excitation signal stimulating the vocal tract.

• Prosody:
  - loudness
  - fundamental frequency
  - sustainability

• Pathological prosody (due to larynx surgical intervention):
  - impaired and unstable fundamental frequency
  - difficulties in its estimation
  - low loudness
  - short sustainability

Method and algorithm
• Articulation information (related to pronunciation) is separated from excitation
• The voiced segments of the excitation are extracted, i.e. only the frames in which vocal cords are active
• The fundamental frequency is estimated through an analysis in the phase space of the signal
• The voiced pathological excitation is then replaced by concatenation of randomly chosen glottal waves from a reference database extracted from healthy voices
• Restored excitation and articulation information

Fundamental frequency estimation
• Embedding of the signal in its phase space
• The periodicities of the signal are studied through the periodicities in phase space
• Vector distance and vector correlation improve pitch estimation when compared to zero crossing technique

Results and conclusions
• Recurrence analysis has been shown to obtain better results than the classical zero crossing approach to estimate the fundamental frequency
• A good fundamental frequency estimation is necessary to recreate a natural prosody
• The future development of the global approach consists of the improvement of the robustness of voiced-unvoiced detection and of the articulation information estimation.

Acknowledgements
This project is financially supported by the Swiss League against Cancer and by the Gebert Rüf Stiftung. It is developed in a collaboration 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).