From black boxes to gray boxes: A look behind the curtain of nonlinear speech models.

Gernot Kubin,
Signal Processing and Speech Communication Laboratory
Graz University of Technology, Graz, Austria

Abstract - When we started our work on nonlinear dynamical modeling of speech signals some 15 years ago we did this based on a strong mathematical existence theorem due to F. Takens that would guarantee us to build signal models able to re-generate perfectly equivalent observables without even attempting to identify the structure of the underlying physical signal generation mechanism. Combining this existence theorem with automatic machine learning methodologies has naturally led to a framework where observed speech signals (from sustained, stationary sounds) can more or less automatically be converted to a nonlinear black-box oscillator with a parameterization adapted to the data, but with no insight into the relationship between these parameters and their physical meaning.

Many subsequent years have been devoted to incremental improvements of the structures used for representation of the black-box oscillators and to the tuning of the associated machine learning algorithms. However, it was only during the course of the past five years, where our work was embedded in the COST277 environment, that the constant challenge from our colleagues in speech syntesis made us relate our model back to the human speech signal generation apparatus. By now, we have converged to a gray-box model where some light is shed on the relationship of the model structure (in terms of a decomposition into a small number of interacting modules) to the underlying physics. The model is still no glass-box as we have maintained an emphasis on the ability to automatically identify all the parameters from data without other physical measurements than the speech waveform itself. This model has been referred to as the oscillator-plus-noise model and condenses what we have learnt in the COST action. At the eve of its completion, we are about to lift the curtain even further by working towards glass-box models that will use limited prior assumptions on model complexity to uncover more details of the internal system structure of the physical modules).