# Semi-Automatic Parameter Extraction for Formant Speech Synthesizers: The Genetic Algorithms Case

S. N. RAPTIS[†‡], S. G. TZAFESTAS[†], and G. CARAYANNIS[‡]

[†] *Intelligent Robotics and Automation Laboratory*
*National Technical University of Athens*
*Zographou GR-157 73, Athens, GREECE*

[‡] *Institute for Language and Speech Processing*
*Artemidos 6 & Epidavrou, GR-151 25, Athens, GREECE*

## Abstract

Designing a high quality formant TtS system, can be roughly viewed as a "synthesis based on analysis" task.

An appropriately selected corpus of prerecorded natural speech is used as the base for an analysis procedure which will yield a set of contours describing the required time evolution for each parameter of the synthesizer model. The adequacy of these contours can be measured via re-synthesis and comparison of the spectral characteristics of the resulting waveform to the original corpus.

Having performed this speech-to-contours conversion, an additional decision should be made concerning the definition and selection of the basic units of the system: the segments. Then, the contours will need to be broken down to (a) a set of target values for the segments and (b) a set of segmental/prosodic rules applied on them. These two components are highly interrelated presenting an (in effect) infinite number of possible "choices".

From the above steps, the main challenge is presented by the speech-to-contours conversion procedure. While good estimates can be found for some of the synthesizer parameters (e.g. for the formants and the bandwidths of voiced speech fragments), other are quite hard to determine (e.g. the voice source parameters, the parallel amplitudes, etc).

Voice source parameter estimation traditionally relies on either direct estimation techniques or on glottal inverse filtering techniques followed by an appropriate parametrization phase. This parametrization is carried out by fitting an appropriate model, e.g. the LF model. However, a better fitting of an estimate of the glottal source derivative does not necessarily guarantee an overall synthetic signal "closer" to the target natural one.

As an alternative, genetic algorithms (GAs) can be employed to perform this very task. They are able to locate "good" values for the parameters where the "goodness" is measured in the context of the overall system output rather than solely in the context of fitting the glottal source estimate.

Making no explicit assumptions on the way the parameters to be optimized are related to each other and to the overall system output, GAs performs a stochastic goal-driven search of the parameter space aiming at the minimization of a cost criterion; possibly the minimization of the spectral distance of the produced synthetic speech from the corpus.

GAs can be smoothly integrated with any other analysis technique to provide a multiple-experts system for efficiently performing the estimation of any parameter of a formant synthesizer and the formulation of high-quality contours. Collecting a large enough set of such contours, it is then possible to (fully or semi-automatically) formulate segmental rules for the synthesizer.

The results of some experiments that have been carried out are presented to clarify the discussed issues and demonstrate the efficiency of a GA-based estimation system.