Modelling Signal Dynamics in Voiced Speech Coding

Domingo Docampo¹
Juan Francisco Fariña²

ABSTRACT

We present a non-linear and non-parametric prediction algorithm for modelling the internal dynamics of the speech signal through a state-space geometric description. The proposed model extracts the intrinsic determinism embedded in the voiced speech signal from comparisons among the observed trajectories in state space. The combination of a metric criterion and frequency similarity results in a model which evolves throughout a quasi-periodic trajectory with a pitch period similar to the speech signal under analysis. Results for voiced frames of this model and comparisons with LPC-based methods are promising at medium bit rates.

31.1 Introduction

Time-domain coders differ in the way the receptor generates the excitation to the LPC filter. The quality of the synthetic signal in this type of coders relies on the number of bits used to encode the excitation signal [1]. There are currently some techniques which render acceptable quality within the range from 4.5 to 6.5 Kbps. Reducing the bit-rate without affecting the quality is not an easy task. Kleijn [2] suggests an interpolative algorithm (PWI), based on the quasi-periodic nature of the excitation signal during voiced frames, which allows the reconstruction of the signal from two different sequences delayed in a certain amount of time – typically 20-30 ms.

We are interested in studying whether this quasi-periodic nature of the voiced frames is really caused by an internal determinism of speech dynamics. If that were the case, it would be possible to model the production of voiced speech so that the excitation signal is generated by a dynamic process driven by a trigger sequence. Such an approach was taken in [6], where the speech signal is modeled as the output of a low-dimension determinism generator.

In this paper we propose a non-linear and non-parametric algorithm to model the internal dynamics of the signal through a geometric description of its state space, under the hypothesis that the mechanism for the production of voiced speech is

¹Departamento de Tecnologías de las Comunicaciones, E.T.S.I. Telecomunicación, Campus Universitario, 36200 Vigo, Spain, E-mail: ddocampo@tsc.uvigo.es
²Departamento de Tecnologías de las Comunicaciones, E.P.S.I. Telecomunicación, Univ. Carlos III de Madrid, 28911 Leganés, Madrid, Spain, E-mail: jfarina@tsc.uc3m.es
quasi-deterministic and of low dimension[3]. Our main contribution is the development of a metric which takes into account the non-stationary nature of the speech signal to enable an appropriate neighbor-state selection. It results in a more accurate prediction of speech dynamics which can be exploited to achieve a better compression rate for voiced speech frames.

Section 31.2 is dedicated to introducing qualitative aspects of the Dynamic System Theory. The voiced speech production mechanism is analyzed in this framework in Section 31.3. Next, the coding algorithm which results from this analysis is explained in Section 31.4. Section 31.5 presents the results of applying the algorithm to synthesize actual voiced speech. Finally, Section 31.6 gives the summary and conclusions.

31.2 Dynamic System Model

Modelling and predicting a deterministic dynamic system relies on finding a function $F$ that can mimic, with the desired degree of accuracy, the evolution of the system in state space [5]. We take the state of a deterministic dynamic system $s_t$ as the necessary information to determine its future evolution: the sequence of states the system goes through in time.

Then our objective is to track the deterministic internal dynamics defined as

$$s_{t+N} = F_N(s_t) \tag{31.1}$$

The problem we deal with can be split in two, well-defined stages. First, we should identify a mapping which transforms the actual state of the system to a vectorial representation (the problem of state-space reconstruction). Then, we need to address the modelling of the dynamic evolution of the system efficiently (the problem of the dynamic system approach).

Reconstruction of the attractor is carried out starting from observations of the system $x_n$ — usually the only information we have. We want to find a one-to-one correspondence (embedding) between the attractor states in phase space and the image states built from the observations. The one-to-one requirement is useful since then the future evolution of the system can be completely specified from a point in the reconstructed space. Otherwise, there would be crossed trajectories in some points of the space, in which the prediction would thus be impossible. Besides, a differentiable embedding is desirable, so that the topological structure of the attractor is also preserved. A particular solution is the delay coordinate embedding which associates each state $s$ with a vector in the Euclidean space $\mathcal{R}^m$: the state of the system is represented by a vector (delay coordinate vector) containing the last $m$ samples taken from the observations:

$$d_t = [x_t, \ldots, x_{t-(m-1)}]^T \tag{31.2}$$

Let $S$ denote a compact finite-dimensional set of states of a system and $D$ a mapping from $S$ to $\mathcal{R}^m$. A well-known result by Takens [4] guarantees, under certain conditions, the existence of an embedding. Provided that $m$ is greater than twice the dimension of $S$ and $S$ is a smooth manifold, then $D(S)$ is also a smooth manifold, and $D$ is a diffeomorphism.