EVALUATION OF A NEW MODEL FOR VOWELS SYNTHESIS WITH PERTURBATIONS IN ACOUSTIC PARAMETERS

G.A. ALZAMENDI†, G. SCHLOTTHAUER†, H.L. RUFINER‡ and M.E. TORRES†‡

† *Laboratorio de Señales y Dinámicas no Lineales (FI-UNER) -CONICET*
‡ *Centro de I+D en Señales, Sistemas e Inteligencia Computacional (FICH-UNL) -CONICET*
metorres@santafe-conicet.gov.ar

***Abstract*— Voice signal contains intrinsic irregularities which become more evident in the presence of pathologies. The acoustic parameters are very useful for the clinical assessment of voice and pathologies detection. Most existing voice models handle irregularities as additive noise and not as information carriers. In this work, a new model is proposed allowing to generate synthesized voices with previously selected acoustical parameters shimmer and jitter. Artificial voices are generated from a glottal source signal, obtained by conveniently disturbing amplitudes and periods, and then filtered using an autoregressive linear filter. Models were developed for amplitude and period perturbations based on statistical methods. Several signals were generated and the performance of the model was analyzed. The quality of synthesized voices was evaluated using an objective quality measurement. The obtained jitter and shimmer values mostly agreed with the theoretically predicted values. These results suggest that this model is useful for artificial voices generation.**

***Keywords*— irregular-voice, voice synthesis model, acoustic parameters, jitter, shimmer.**

I. INTRODUCTION

Study and modeling of voice generation mechanisms cover diverse scientific fields and demand interdisciplinary points of view, due to the complexity and diversity of the involved elements. The main issues are the analysis of the anatomical structures and the phenomena involved in the speech processes, considering dynamical behaviors and structural relationships.

Applications of speech models include methods for speaker recognition, techniques to improve the quality of artificial voices, strategies applied to man–machine interfaces and a variety of techniques for modeling, conditioning, synthesis, compression and transmission of speech signals. The different models developed to analyze and imitate the process of voice generation differ on the employed strategies and methods, and depend on the considered application. Recently, voice models have been applied to the study and synthesis of pathological voices. This made possible to develop a knowledge and understanding of the etiologies and alterations that can be found in different voice disorders (Schlotthauer, 2010; Torres *et al.*, 2009). Moreover, it has been demonstrated that even normal voices present intrinsic irregularities and that they are responsible of the degree of perceived naturalness (Baken and Orlikoff, 2000; Schlotthauer, 2010).

Acoustical parameters are usually employed in the practice of clinical medicine. Added to perceptual analysis and specific tests, they allow specialists to characterize the voice of an individual and to determine the presence of pathologies (Schlotthauer, 2010). *Shimmer* and *jitter* are the parameters most frequently used for quantifying instantaneous alterations in amplitude and frequency, respectively. It has been proved that these parameters are useful to characterize different types of voice and are sensitive to voice disorders (Baken and Orlikoff, 2000; Brockmann *et al.*, 2011; Velasco García *et al.*, 2011).

The purpose of this study is to propose and develop a simple voice synthesis model based on acoustical parameters of interest in the voice clinical practice. Particularly, here the focus is centered in the quantities *shimmer* and *jitter*, considering both healthy and pathological voices. For the validation of the proposed model two approaches are followed. First, the quantities estimated from the synthesized signals will be contrasted with the theoretical values. Also, the synthesized speech quality will be evaluated by means of an objective quality measure, with high correlation with psycho–acoustic perception.

This article is organized as follows: in Section II the proposed model is introduced, the methodology is presented and the used materials are detailed. In Section III the experimental results are shown and discussed. In Section IV the conclusions and future works are presented.

II. MATERIALS AND METHODS

In this work, we propose a voice synthesis method based on the *source-filter* speech production model. This approach possesses a simple theoretical framework and has proved to be useful in a variety of applications (Proakis *et al.*, 1993). The model is inspired on the physiology of the voice system and the phonation process, where the air flow coming from the lungs is modified by the actions of the vocal folds producing regular pulses, called *Glottal Pulses* (GP). Those are acoustically transmitted through the *Vocal Tract* (VT), giving as result the proper voice signal (Rufiner, 2009). Each component of our model is detailed in the following sections.

A. Glottal source

The morphology of the *Glottal Source* (GS), considered in the model, depends on the speech sounds to be ana-