Signal saturation and its effects on time domain parameters

Lorenzo Cioni

Laboratorio di Linguistica, Scuola Normale Superiore, Pisa, Italy

The present paper examines the effects of saturation on the time domain parameters that can be extracted from recordings of speech. The kinds of saturation we are going to discuss in the paper include clipping, zeroing and two's complement whereas the parameters we will take into considerations include root mean square (RMS, as correlated with loudness), zero crossing rate (ZCR), auto correlation (as a tool to extract F0 from speech samples), voice onset time (VOT) and rise time (RT).

INTRODUCTION

Saturation is a physical phenomenon that occurs whenever a quantity exceeds the maximum value that can be managed on the physical system holding that quantity. In case of speech analysis systems saturation can occur either during the recording phase, during the acquisition phase or within the analysis software owing to wrong operations on the acquired signal. In all such cases saturation is a non linearity superimposed on the signal that reflects in a distortion of the signal itself [1, 3]. In many cases there is no way to correct such distortion [2] (owing to the unavailability of either the speakers or the tape recordings) and so what is needed is a way to evaluate the degree of saturation of a signal and the correctness of the (time domain) parameters that can be extracted from such a signal.

THREE MODELS OF SATURATION

Saturation can be modeled with three kinds of non linearity [2, 3]: clipping, zeroing and two's complement (eqs. (1), (2) and (3)).

Such models can be described respectively by the following relations (where $x = x(nT)$ and $y = y(nT)$ are respectively the input and output signal of the non linear system, T is the sampling period and the values x_0, y_0 are the absolute values that define the range of linearity):

(1) $y = kx\,[u(x+x_0) - u(x-x_0)] + y_0\,[u(x-x_0) - u(-x-x_0)]$

with $u(x) = 1$ if $x \geq 0$ and $u(x) = 0$ if $x < 0$ while $k=y_0/x_0$ (usually we have $k=1$);

(2) $y = kx\,[u(x+x_0) - u(x-x_0)]$

and

(3) $y = k\,(x-jx_0)$ if $(j-1)x_0 \leq x \leq (j+1)x_0$

where $j = 2m$, $m\in$ Z. Equation (3) can be written as:

(3') $y = [kx\,[u(x+x_0) - u(x-x_0)]]_{T'}$

where $T' = 2x_0$ is the period of the function.

THE TIME DOMAIN PARAMETERS

A speech signal can be described both in the time domain and in the frequency domain with the use of parameters extracted from the signal itself. As to the time domain parameters in this paper we will limit our analysis to RMS, ZCR, autocorrelation, VOT and RT. Such parameters are correlated with physical quantities and can provide segmental cues for the analysis of speech [4].

RMS [1, 4] is a measure of the energy of a speech signal and when applied to successive windows gives a measure of its changes in amplitude (and so in loudness) over time. It depends on the sum of the squared values of the samples averaged over the length N of a window.

ZCR [4] depends on the number of times a signal crosses the zero line and gives us a measure of the dominant frequency in a signal. It can be used in differentiating between voiced and unvoiced sounds and, together with RMS, can be used to make a simple speech/no speech distinction. Autocorrelation [4] can be used to extract the pitch from a signal by comparing the signal with a delayed version of the same signal. VOT [4] is a measure of the duration of the burst in case of stops followed by a vowel. Last but not least RT [4] can be used as a cue for the distinction between the affricate and fricative sounds and depends on the time interval over which a maximum amplitude is reached.

AN EXPERIMENTAL SETTING

In order to examine the effects of saturation over speech signals (and the derived time domain parameters) an experimental setting has been conceived [2] (Figure 1). According to such a setting speech signals are acquired in digital format so that streams of samples x(nT) are available. Such streams are then passed through the non linear system (NL) characterized by relations (1), (2) and (3') so to produce the new streams y(nT).

The effects of saturation over the signals can be examined by analyzing the signals x(nT), y(nT) and the "error signal" ε(nT) = x(nT) - y(nT) (Diff) and the analysis can be carried out by varying the values x_0 and y_0 so to simulate lower (with higher values) or higher (with lower values) degrees of saturation. The main aim of the analysis is the definition of criteria that, when applied to a generic signal, allow us to understand if the signal has suffered saturation, which kind of saturation and which was the degree of saturation (the famous "how and how much").

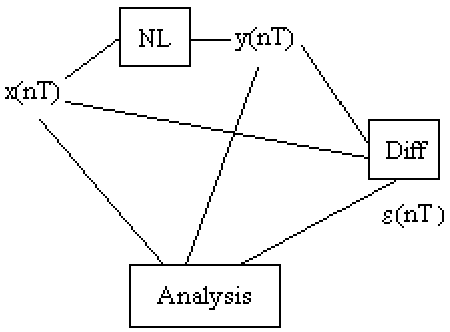

FIGURE 1. Experimental setting, symbols explained in the text.

The programs used during the acquisition, saturation and analysis phases are Goldwave™ (NL and Diff) and Xwaves™.

INFLUENCE OF SATURATION ON TIME DOMAIN PARAMETERS

The three kinds of saturation (1), (2) and (3') turn into a modification of the waveform of the speech sounds so that in all the cases the shape of the waveform is changed since some of the samples are substituted either with clipping value or with zeroes or gets their sign switched [2]. The extent of the saturation, in the model of figure 1, is determined by the values x_0 and y_0 since they determine, for the same signal x(nT), the energy and the length of the energy periods of the signal ε(nT), i.e. the length of the time intervals over which ε(nT) shows a non null amplitude.

The following table 1 summarizes the qualitative connections among the three kinds of saturation and the time domain parameters. Column headings are the three kinds of saturation presented in the paper and row headings are the time domain parameters we consider as more significant in the case of time domain analysis of speech.

Table 1.

	Clipping	Zeroing	Two's complement
RMS	+	+	+
ZCR	-	-	+
AC	-	-	-
VOT	+	+	+
RT	+	+	+

A "-" sign means that a certain kind of saturation exerts (almost) no influence over the corresponding time domain parameter whereas a "+" sign means that the higher is a certain kind of saturation the stronger is the effect on the corresponding time domain parameter. It is obvious that since RMS, ZCR and AC are evaluated as a function of the length N of a window the shorter is the window the more accurate are the graph of the parameters versus (discrete) time nT and so the more evident is the influence of saturation on those parameters.

We note that the use of the signals x(nT), y(nT) and ε(nT) is conceivable only within a theoretical model. In case of real life signals the use of such model should allow us to understand if a given signal has suffered saturation, which kind of saturation and at which degree and all this by simply examining the graphs of the waveforms and of the time domain parameters extracted by the signals themselves.

CONCLUSIONS

The analysis presented in this paper has been kept to an introductory level but further (and more accurate) details will be provided during the presentation. The topic of the paper represents one of the current research topics of the author and is going to be the subject of further investigations both on the theoretical and on the empirical ground.

REFERENCES

1. Smith S. W., *The Scientist and Engineer's Guide to Digital Signal Processing*, Analog Devices, California Technical Publishing, San Diego, 1999.
2. Cioni L., "La saturazione del segnale: una indagine preliminare. Ovvero la ricerca del modello", to appear on the Proceedings of the "XI Giornate di Studio del GFS", Padova, Italy, November 29-30 December 1 2000.
3. Derong L. and Michel A. N., *Dynamical Systems with Saturation Nonlinearities*, Lecture Notes in Control and Information Sciences 195, Springer Verlag, London, 1994.
4. Harrington J. and Cassidy S., *Techniques in Speech Acoustics*, Kluver Academic, 1999.