AN ARMA MODEL OF SPEECH PRODUCTION PROCESS WITH APPLICATIONS

Jinshi Huang

School of Electrical and Computer Engineering
Oklahoma State University

ABSTRACT - An autoregressive and moving-average (ARMA) model of speech production process is proposed. The orders of the model are determined from the generalized partial autocorrelation (GPAC) pattern. Based on the maximum likelihood estimation, parameters of the model are estimated via the Marquardt algorithm. Experiments show that the fricatives can be better modeled as an ARMA process. The autocorrelation of the residuals can be used for pitch detection and voiced/unvoiced speech recognition.

INTRODUCTION

In speech signal processing and recognition, the production process of speech signals has been one of the most important topics. Since once we know the mechanism with which speech signals are produced, we can extract the parameters which characterize this process. These parameters not only can be used as features for speech recognition, but also can be used to synthesize speech signals of good quality.

In the past years, autoregressive (AR) model has been widely used to characterize the speech production process. Making use of this model, the linear prediction coding (LPC) has been very successful. A comprehensive survey of this topic can be found in [1].

However, some speech signals such as the nasals and fricatives can hardly be modeled as AR processes. Based on state-space representation, an ARMA model was put forward in [2]. In that paper, the order of AR(p) was found by maximizing the observed signal to innovation ratio (OIR)

$$OIR = 10 \log_{10}(r_o/c_o) \qquad (1)$$

where r_o and c_o are the variances of the observed signal and innovation sequence respectively. The innovation is defined as the difference between the observed value and the predicted value. This method, though works for AR model, fails to consider the effect of the MA part on the AR behaviour when the ARMA model is concerned.

In this paper, the generalized partial autocorrelation (GPAC) is applied to determine how well the ARMA model fits the speech signal, and from the pattern of GPAC's the AR order p and MA order q can be found. Based on maximum likelihood method, the parameters can be found be using the Marquardt algorithm.

Applications of this ARMA model include discriminating voiced speech signals from the unvoiced counterparts and estimating the pitch periods. This can be done by observing the autocorrelation of the ARMA residuals. Experiments have shown that this method has significantly reduced the effect of the vocal tract

resonances on the input signal which carries information of whether the speech is voiced or unvoiced and the information of pitch period in the voiced case.

THE GENERALIZED PARTIAL AUTOCORRELATION

An ARMA process is defined as

$$y(t)-\phi_1 y(t-1)- \ldots -\phi_p y(t-p)=a(t)-\theta_1 a(t-1)- \ldots -\theta_q a(t-q) \qquad (2)$$

where y(t) are measurements and a(t) are white noise. It can be shown [3] that the generalized partial autocorrelation of an ARMA process has the following pattern

AR order variable

MA order variable	1	...	p	p+1	p+2
0	ϕ^o_{11}	...	ϕ^o_{pp}	ϕ^o_{p+1p+1}	ϕ^o_{p+2p+2}
.		...		...	
q-1	ϕ^{q-1}_{11}	...	ϕ^{q-1}_{pp}	ϕ^{q-1}_{p+1p+1}	ϕ^{q-1}_{p+2p+2}
q	ϕ^q_{11}	...	ϕ_p	0	0
q+1	ϕ^{q+1}_{11}	...	ϕ_p	o/o	o/o

As can be seen, the GPAC array has a clear pattern which indicates the orders of the ARMA process p and q.

ESTIMATION OF THE PARAMETERS

Define the likelihood function of the ARMA process as

$$f_y \sim \exp\left(-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n} a^2(t)\right) \qquad (3)$$

and $\beta_1=\phi_1, \ldots, \beta_p=\phi_p, \beta_{p+1}=\theta_1, \ldots, \beta_{p+q}=\theta_q$. Based on maximum likelihood estimation, the parameters $\beta_i (i=1,2,\ldots,p+q)$ are estimated by minimizing

$$s(\underline{\beta}) = \sum_{t=1}^{n} a^2(t) \qquad (4)$$

which can be done by Marquardt algorithm described as follows [4]:

(1) $x_{i,t} = -\frac{a(t)}{\beta_i}$

$A=X^T X$

$\underline{g}=X^T a_o$

$d_i=\sqrt{A_i}$

$\underline{a}_o$ is the initial condition of $\underline{a}=[a_1\ a_2\ \ldots\ a_n]^T$.

(2) $A_{ij}^*=A_{ij}/d_i d_j$

$A_{ii}^*=1+\pi$

$g_i^*=g_i/d_i$

solve $A^*\underline{h}^*=\underline{g}^*$ for $\underline{h}^*$ by inversion of A^*.

$h_j=h_j^*/d_j$

$\beta(new)=\beta(old)+h$

(3) IF $s(\underline{\beta}(new)) < s(\underline{\beta}(old))$ THEN
IF $|h_i| < \varepsilon$ THEN
$\underline{\beta}(new)$ contains the parameters

$$\sigma_a^2 = \frac{1}{n-p-q}\, s(\underline{\beta}(new))$$

$$cov(\underline{\beta}) = \sigma_a^2 A^{-1}$$

STOP
ELSE
$\underline{\beta}(old)=\underline{\beta}(new)$
$\pi=\pi/F_2$ $(F_2>1)$
ENDIF
ELSE
$\pi=F_2*\pi$
$\underline{\beta}(old)=\underline{\beta}(new)$
IF $\pi > \pi_{max}$ THEN
error
STOP
ENDIF
IF # of iterations > maximum # of iterations THEN
error
STOP
ENDIF
GOTO (2)
ENDIF
GOTO (1)

EXPERIMENTS

Many experiments on both unvoiced and voiced speech signals have been conducted. It has been shown that for unvoiced speech signal, the autocorrelation of the residual decays very rapidly, opposed to the clear periodic pattern of the voiced case. This fact suggests that the residual autocorrelation can be used as a feature for voiced/unvoiced speech recognition. Furthermore, the pitch period can be detected from the residual autocorrelation.

The procedure of developing the ARMA model are as follows

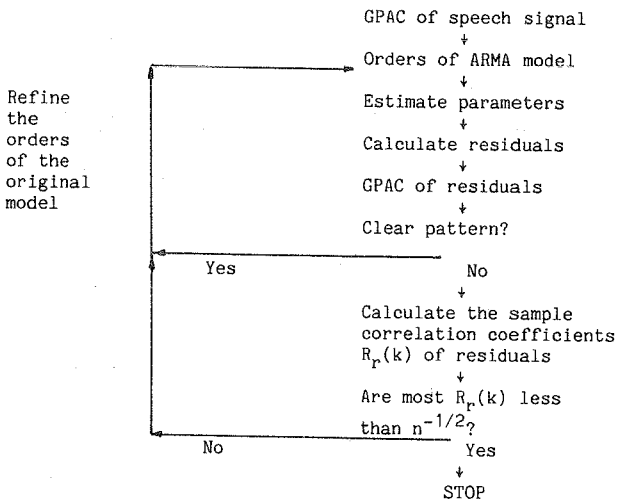

Figure 1 is the unvoiced fricative speech signal /ʃ/. Figure 2 is the autocorrelation of this signal. Figure 3 is the autocorrelation of the ARMA (11,6) residual. As we can see, the autocorrelation of the residual decays very rapidly abd no any periodic pattern can be seen, which suggests that this is an unvoiced signal.

CONCLUSIONS

In this paper, an ARMA model of speech signal is developed. The methods of estimating orders and parameters are described, and some simple applications are discussed.

Some further applications of the ARMA model can be explored. A distance measure between two AR process has been proposed [5]. But most unvoiced speech signals cannot be modeled as AR processes. A similar measurement could be developed for ARMA model and used to discriminate two unvoiced speech signals in speech recognition.

The main problem of the algorithm used in this paper is the off-line nature. Based on lattice structure, some adaptive approaches of ARMA modeling have been proposed [6][7]. An adaptive algorithm for ARMA signal with some special properties has been developed in [8].

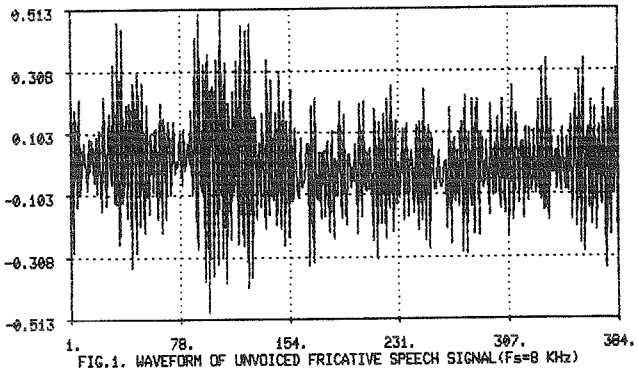

FIG.1. WAVEFORM OF UNVOICED FRICATIVE SPEECH SIGNAL(Fs=8 KHz)

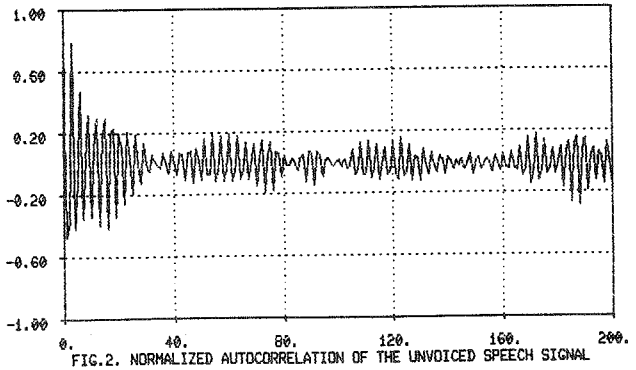

FIG.2. NORMALIZED AUTOCORRELATION OF THE UNVOICED SPEECH SIGNAL

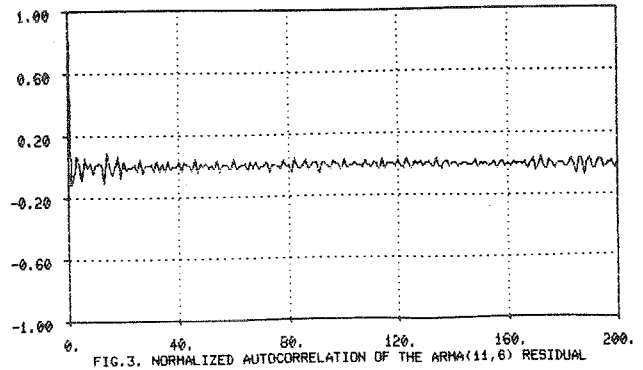

FIG.3. NORMALIZED AUTOCORRELATION OF THE ARMA(11,6) RESIDUAL

ACKNOWLEDGEMENT

The author would like to thank Dr. M. Hagan of Oklahoma State University for his critical comments.

REFERENCES

[1] Makhol, J. (1975) "Linear Prediction: A Tutorial Review," Proc. IEEE 63-64, 561.

[2] Morikawa, H. and Fujisaki, H. (1984) "Speech Identification of Speech Production Process Based on a State-space Representation," IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-32, No. 2, 252-262.

[3] Gray, H.L., Kelley, G.D. and McIntire, D.D. (1978) "A New Approach to ARMA Modeling," Comm., Statist., -Simula. Computa., B7(1), 1-77.

[4] Scales, L.E. (1985) Introduction to Non-Linear Optimization, (Springer-Verlay: New York).

[5] Rabiner, L.R. and Schafer, R.W. (1978) Digital Processing of Speech Signal, (Prentice-Hall: New Jersey).

[6] Karlsson, E. and Hayes, M.H. (1987) "Least Squares ARMA Modeling of Linear Time-Varying Systems: Lattice Filter Structures and Fast RLS Algorithms," IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-35, No. 7, 994-1014.

[7] Cowan, C.N.F. and Grant, P.M. (1985) Adaptive Filters, (Prentice-Hall: New Jersey).

[8] Nehorai, A. and Stoica, P. (1988) "Adaptive Algorithms for Constrained ARMA Signals in the Presence of Noise," IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-36, No. 8, 1282-1291.