AN ARMA MODEL FOR EXTRACTING CANTONESE PHONEME CHARACTERISTICS

S H Leung[*], Andrew Luk[*], Godfrey K F Liu[**] and Caesar S Lun[**]

[*]Department of Electronic Engineering
City Polytechnic of Hong Kong

[**]Department of Chinese, Translation and Linguistics
City Polytechnic of Hong Kong

ABSTRACT - An autoregressive moving average algorithm is proposed for the analysis and extraction of characteristics in Cantonese phonemes. Its performance is found to be more accurate than a number of classical ARMA estimations methods. This method is used to extract the formant characteristics in the Cantonese vowel phonemes and is found to match closely with the estimates obtained by the Cantonese linguist.

INTRODUCTION

Recently a number of autoregressive moving average (ARMA) algorithms have been proposed (Marple, 1987 and Haykin, 1991) for the analysis and extraction of phoneme characteristics, such as formant/antiformant frequencies and their corresponding bandwidths, in various tonal languages or dialets, such as Chinese and Cantonese. Unfortunately, many of these methods have been found to be unsatisfactory in analysing the Cantonese phonemes. In particular, depending on the orders of the autoregressive (AR) and the moving average (MA) parts, false peaks and zeros are easily generated using some of the existing ARMA algorithms. Furthermore, peaks may also be masked off by the harmonics of the pitch frequency if their amplitudes are of the same magnitude as that of the formant peaks. In addition, the estimated peaks may also be deviated from the actual peaks. All these factors complicate the estimation of the formant peaks in the Cantonese phonemes.

In this paper, a new ARMA algorithm is described. This method is basically a maximum entropy spectral estimation using correlation constraints and cepstrum constraints simultaneously. In order to linearize the problem, the AR and the MA parameters are determined separately. The extended Yule-Walker equations are used to determine the AR part for satisfying the correlation constraints, while cepstrum analysis is used to extract the MA parameters by matching the cepstrum of the ARMA model to that of the periodogram of the time series. The methodology will be discussed in the next section. It is worth noting that the use of periodogram in the estimation of the MA parameters is better than using the high order linear predictive coding (LPC) model (Leung, 1990). This approach is found to match more closely to the spectrogram of the Cantonese phonemes, in particular for the consonants. In addition, the formants estimated from this method are found to match closely to those values observed by the Cantonese linguist. This method also provides a good estimation of the parameters of the consonants. Furthermore, the algorithm can also generate a rational transfer function that is useful for speech reproduction, which is not available in other algorithms.

AN AUTOREGRESSIVE MOVING AVERAGE (ARMA) ALGORITHM

The maximum entropy spectral estimate of a random process $\{x(t)\}$ with correlation and cepstrum constraints was shown to be an ARMA model (Lagunas-Harmandez, et la, 1984). Let $x(t)$ denote an ARMA(P,Q) process, where P and Q are respectively the orders of the autoregressive (AR) and moving average (MA) parts, and $\{x(t)\}$ be its random process. Let $P_{xx}(\omega)$ be the spectral estimation of the process $\{x(t)\}$. It can be shown that the solution of $P_{xx}(\omega)$ is given by (Marple, 1987):

$$P_{xx}(\omega) = \left(1 + \sum_{|m|=1}^{Q} \gamma_m e^{jm\omega} \right) \Big/ \left(\sum_{n=-P}^{P} \alpha_n e^{jn\omega} \right) \tag{1}$$

where α and γ are respectively the Lagrange multipliers of the AR and MA parts. It is a nonlinear problem to solve these multipliers simultaneous. Equation (1) can be rewritten as

$$P_{xx}(\omega) = \frac{B(\omega)B^*(\omega)}{A(\omega)A^*(\omega)} \tag{2}$$

where

$$B(\omega) = \sum_{n=0}^{Q} b_n e^{-jn\omega}, \qquad A(\omega) = \sum_{n=0}^{P} a_n e^{-jn\omega}, \qquad a_0 = 1, \tag{3}$$

$()^*$ denotes the complex conjugate of the bracketed value, and a_n and b_n denote respectively the AR and MA model parameters. If the excitation is assumed to be white, the AR parameters can be found by solving the extended Yule-Walker equations via the covariance method (Marple, 1987 and Leung, 1990).

The MA parameters can be solved using the Prony's method on the cepstrum of the process $\{x(t)\}$. In the later case, periodogram of the ARMA(P,Q) process is being used to estimate the cepstrum. The relationship between the cepstrum and the MA model parameters can be examined as follows: Let $F(\omega)$ be the magnitude spectrum of an MA process,

$$F(\omega) = \left| \sum_{n=0}^{Q} b_n e^{-jn\omega} \right| = \left| b_0 (1 + \sum_{\ell=1}^{Q} \beta_1 e^{-j\ell\omega}) \right| = \left| b_0 \prod_{\ell=1}^{Q} (1 - z_\ell e^{-j\ell\omega}) \right| \tag{4}$$

and let $c(t)$ denote the cepstrum of $F(\omega)$ defined as

$$c(t) = \mathcal{F}^{-1}\{ \ln \left| b_0 \prod_{\ell=1}^{Q} (1 - z_i e^{-j\ell\omega}) \right| \tag{5}$$

where $\mathcal{F}^{-1}$ is the inverse discrete Fourier transform. Using the identity

$$\ln (1 - \beta e^{-j\omega}) = - \sum_{n=1}^{\infty} (\beta^n/n) e^{-jn\omega}, \tag{6}$$

we obtain

$$c(t) = - \sum_{\ell=1}^{Q} Z_\ell^t / t \qquad , t > 0. \tag{7}$$

From (7), the variable $v(t)$ defined as

$$v(t) = t\, c(t) \tag{8}$$

satisfies the prediction formula given by

$$v(t) + \sum_{\ell=1}^{Q} \beta_\ell \, v(t-1) = 0 \ , \tag{9}$$

where $\{\beta_\ell\}$ are the coefficients of the characteristic equation containing the roots $\{z_\ell\}$. The proof of equation (9) follows the well-known Prony's method (Marple, 1987).

The advantage of the cepstrum approach is that the spurious peaks and the zeros are treated in the same manner, so that not only the zeros are found in the resulting MA polynomial and also the large spurious peaks as well; thus the annoying spurious peaks are significantly suppressed. This was shown in the experiments of sinusoidal spectral estimation where the peak signal to noise ratio is substantially improved (Leung, 1990).

EXPERIMENTAL RESULTS

Experiments are conducted to evaluate and compare the performance of the proposed ARMA algorithm with some classical ARMA algorithms in analysing Cantonese phonemes. Three classical ARMA algorithms are being investigated in the experiments; these are:

(1) Autoregressive moving average model using model reference adaptive system (MRAS) (Miyanaga, Miki, Nagai & Hatori, 1982),

(2) Pole-zero modelling using modified least square (MLS) (Marple, 1987), and

(3) Pole-zero modelling using modified least square with autocorrelation compensation (POZE) (Marple, 1987).

The procedures for analysing the Cantonese phonemes are as follows:

(1) Segments of Cantonese phonemes are being extracted manually,

(2) The segmented speech is being windowed by a 4-term Blackman-Harris window (Harris, 1978),

(3) The windowed speech is then analysed using one of the autoregressive moving average techiques mentioned above and its spectrum is plotted.

Figures 1 through 4 show respectively the typical spectral plots of the Cantonese vowels [i], as in the Cantonese word "medicial" ( ), for the four ARMA algorithms mentioned in the above paragraphs. Note that in each case, the time segment waveform is also displayed on the upper right hand corner. (For the proposed method, a windowed time segment is displayed.) For comparison purposes, a similiar 25-pole and 25-zero ARMA model is used for the four algorithms. The y-axis, in each figure, denotes the magnitude in decibels. The x-axis denotes the frequency bin or point (not the actual frequency). (The frequency resolution in all these figures is 9.766Hz, i.e. the frequency spacing between any two vertical lines in each figure is 9.766Hz.) Only part of the frequency spectrum is shown in each figures in order to ease the visual comparison of various ARMA algorithms. (Note also that the spectrum computed via discrete Fourier transform (or periodogram) and displayed in Figure 4 is computed with the 4-term Blackman-Harris window; and the spectra displayed in Figures 1 to 3 are spectra computed without the Blackman-Harris window. On the other hand, the time segment for all four ARMA algorithms is windowed prior to the calculation.) The

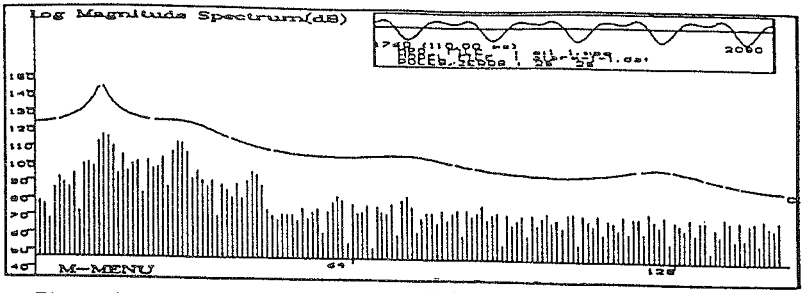

Figure 1: Spectral Plot of Phoneme [i] using MRAS method

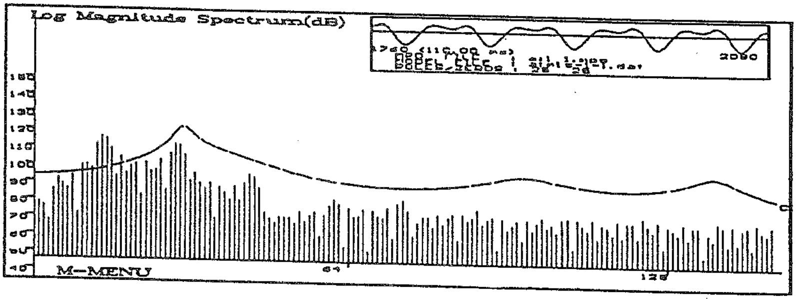

Figure 2: Spectral Plot of Phoneme [i] using MLS method

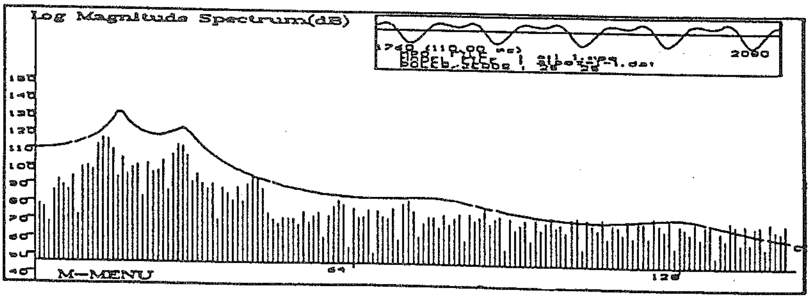

Figure 3: Spectral Plot of Phoneme [i] using POZE method

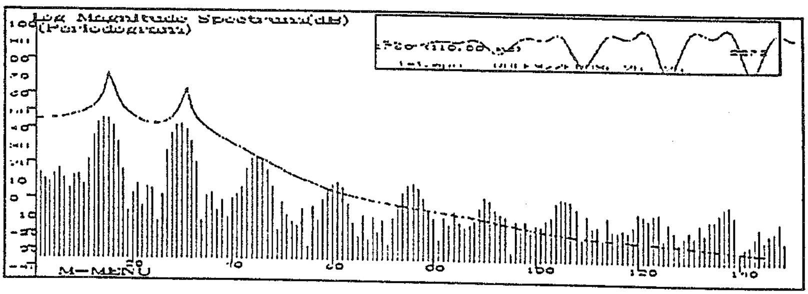

Figure 4: Spectral Plot of Phoneme [i] using the proposed method

first dominant peak in each figure is the pitch frequency, and the second dominant peak is the first formant of the Cantonese [i].

From these plots, it is easy to observe that the proposed ARMA algorithm can produce the best matching with the periodogram. Furthermore, many of the false/spurious peaks, appeared in Figures 1 to 3, are not emphasised using the proposed analysing method. On the other hand, the classical methods all generate spurious peaks and/or flattened some of the peaks which is essential in the characterization of the Cantonese phonemes. For instance, in Figure 1, the first formant of the Cantonese phoneme [i] is flattened by the MRAS method. Similar results are observed in other Cantonese phonemes (Luk, Leung, Liu & Lun, 1991).

CHARACTERISTICS OF THE CANTONESE PHONEMES

Using the proposed ARMA algorithm, the eleven Cantonese vowel phonemes are being analysed. Each phonemes are being segmented manually from a number of utterances and the formant frequencies extracted by the proposed method is being averaged. The averaged formant frequencies of the eleven Cantonese vowel phonemes are tabulated in Table 1.

	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)
[i]	279	2197	3108
[y]	266	2080	3267
[I]	510	1720	2310
[ɛ]	533	1767	2284
[œ]	490	1472	2242
[θ]	529	1248	3774
[a]	664	1174	2085
[þ]	628	1219	2273
[u]	268	695	1104
[U]	452	815	2513
[ɔ]	492	813	3551

Table 1: Averaged Formant Frequenies of Cantonese Vowel Phonemes
(Extracted using the proposed ARMA algorithm)

Note that the formants estimated by this method are found to match closely with those observed by the Cantonese linguist. The formant and antiformant characteristics of the nineteen Cantonese consonants are also under investigation and will be reported in later (due to the space constraint of this paper).

CONCLUSIONS

A new ARMA algorithm is proposed in this paper. Its performance is compared with a number of classical ARMA algorithms. The spectrum generated by the new method matches more closely with the periodogram than all the classical ARMA algorithms being investigated in the experiments. This new method is being used to extract the average formant frequencies of the eleven Cantonese phonemes and is found to match closely with the values observed by the Cantonese linguist.

ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to Dr Eric Zee of the Chinese University of Hong Kong for his help and assistance. The authors would also like to thanks Professor P S Chung, Dr C K Chan, Dr C F Chan, Dr A MacDonald, Dr W H Lau, Dr E Y B Pun and Mr Francis Tam for their encouragement and supports. The project is supported by the CPHK Strategic Research Grant 900174.

REFERENCES

Harris, F. J. (1978) *On the use of windows for harmonic analysis with the discrete Fourier transform*, Proceedings of the IEEE, 66, pp 55-83.

Haykin, S. (Ed.) (1991) *Advances in Spectrum Analysis and Array Processing*, Vol 1, (Prentice Hall: Englewood Cliffs, New Jersey).

Lagunas-Hermandez,M., Santamaria-Perez, M. E. & Figueiras-Vidal, A. (1984) *ARMA model maximum entropy power estimation,* IEEE Trans. Acoust. Speech and Signal Processing, ASSP-32, pp 984-990.

Leung, S. H. (1990) *ARMA spectral estimation based on high order linear prediction and cepstral analysis*, Proceedings of International Conference on Signal Processing, Beijing, China, pp 253-256.

Luk, A., Leung, S.H., Liu, G. K. F. & Lun, C. S. (1991) *Modelling Approach to Chinese Phonemic Analysis and Synthesis*, Research Report, City Polytechnic of Hong Kong.

Marple, S. L. Jr (1987) *Digital Spectral Analysis with Applications*, (Prentice Hall: Englewood Cliffs, New Jersey).

Miyanaga, Y., Miki, N., Nagai, N & Hatori, K. (1982) *A speech analysis algorithm which eliminates the influence of pitch using the model reference adaptive system*, IEEE Trans. Acoust. Speech and Signal Processing, ASSP-30, pp 80-92.