TOWARDS IMPROVED ANALYSIS-SYNTHESIS USING CEPSTRAL AND POLE-ZERO TECHNIQUES

Richard Cann and Kenneth Steiglitz

This work was supported in part by the National Science Foundation under Grant GK-42048, and in part by the U.S. Army Research Office, Durham, NC under Grant DAAG29-75-G-0192.

R. Cann is with the Department of Music, Princeton University.
K. Steiglitz is with the Department of Electrical Engineering and Computer Science, Princeton University, Princeton NJ, 08540.

I INTRODUCTION

In [1] we discussed the use of linear prediction as a kind of music concrete in a MUSIC4 environment; an example of the result is the performance of [2]. For a general technical review see [3]. The effective use of this technique involves many decisions, namely the choice of:

1. sampling rate
2. frame size
3. frame-to-frame interpolation method
4. excitation signal
5. pre-emphasis filter
6. estimation algorithm
7. number of filter parameters.

Our experience has shown that despite continuing attempts at adjustment of these parameters, linear prediction of speech with the all-pole model sounds only so good, and no better. The speech sounds which seem to cause the most difficulty are nasal consonants and those produced during fast transitions between phonemes. The latter problem appears to call for a variable and adaptive frame size, a problem we have not yet considered. This paper describes recent attempts at filter modeling of nasal sounds, with the

reserved hope that this will also improve the general quality of analysis-synthesis.

II ZEROS

The feature that seems to distinguish nasal phonemes from others is the presence of anti-resonances {4}. The physical explanation is as follows: during the production of a voiced nasal phoneme, the oral tract is closed, and radiation takes place through the nose.

(Please refer to Figure 1.)

The effect of the shunted oral cavity is to cancel out certain frequencies. The mathematical expression of this fact is that the transfer function representing vocal tract transmission has zeros as well as poles.

III CEPSTRAL SMOOTHING

When linear prediction is used to estimate a vocal tract transfer function, one relies on the curve fitting implicit in the method to smooth over the fine structure due to the individual harmonics of voiced speech. Another way to smooth out this fine structure is to use the cepstral

smoothing technique of Kopec, Oppenheim, and Tribolet {5}.

(Please refer to Figure 2.)

This idea can be used in three ways:

1. the estimated impulse response can be used directly for synthesis obviating the need for linear prediction altogether,

2. a two-stage method can be used, in which the estimated impulse reponse is approximated with an all-pole model, using standard linear prediction,

3. a two-stage method can be used, in which the estimated impulse response is approximated with a pole-zero model. To do this, we use iterative prefiltering to fit a numerator and denominator simultaneously.

IV ITERATIVE PREFILTERING

We now discuss iterative prefiltering, a pole zero approximation method that has been shown to work on an isolated instance of nasal speech {6}.

(Please refer to Figure 3a.)

Ideally, we would like to solve for the numerator and denominator coefficients simultaneously, making this setup look much like straight all-pole prediction. However, this problem is highly nonlinear, and so an iterative method is used.

(Please refer to Figure 3b.)

Here, an original denominator estimate is obtained using all-pole linear prediction, and then this estimate is used to solve the linear problem shown. Now the new denominator estimate D is used for D' and the process is reiterated. The advantage of this approach is that the poles and zeros are estimated simultaneously, and as convergence is approached, so is the ideal problem (Figure 3a).

V DISCUSSION

We intend to describe the results of experiments using the three methods mentioned above: direct cepstral synthesis, two-stage all-pole synthesis, and two-stage pole-zero synthesis. Preliminary results at the time of this writing (6/77) seem to show:

1. two-stage all-pole (using the covariance method) has fewer unstable frames than conventional all-pole.

2. two-stage pole-zero synthesis captures the effect of mouth closure, but has not yet been made to work well on sentence length examples.

3. the design of the cepstral smoothing filter is extremely important.

4. the iterative prefiltering used in the two-stage pole-zero method does not converge if too many poles are specified, or if pre-emphasis is not used.

We will be presenting taped examples of our results as well as examples of their use in performing the music of R. Cann.

REFERENCES

[1] R. Cann, P. Lansky, K. Steiglitz and M. Zuckerman, 'Practical Considerations in the Application of Linear Prediction to Music Synthesis,' delivered at the First International Conference on Computer Music, MIT 1976.

[2] P. Lansky Artifice, C. Gamer Organum.

[3] J.D. Markel and A.H. Gray, Jr., 'Linear Prediction of Speech,' Springer-Verlag, 1976.

[4] J.L. Flanagan, 'Speech Analysis, Synthesis, and Perception,' (second edition) Springer-Verlag, 1972, pg. 77.

[5] Kopec, Oppenheim, and Tribolet, 'Speech Analysis by Homomorphic Prediction,' IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP-25, pp. 40-49, Feb. 1977.

[6] K. Steiglitz, 'On the Simultaneous Estimation of Poles and Zeros in Speech Analysis,' IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-25, no. 3, pp. 229-234, June 1977.

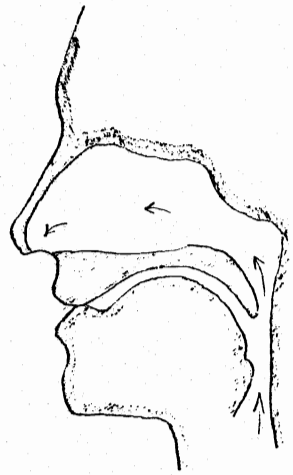

Figure 1 The vocal tract configuration during production of a nasal sound

SPEECH FRAME

*

PRE-EMPHASIS

*

HAMMING WINDOW

*

FFT
(spectrum)

*

LOG ABSOLUTE VALUE

*

FFT-1
(cepstrum)

*

SHORT TIME CEPSTRAL WINDOW

*

FFT

*

EXPONENTIATION
(smoothed spectrum)

*

FFT-1

*

ESTIMATED IMPULSE RESPONSE

Figure 2

THE IDEAL PROBLEM:

minimize $(V_k - N/D\ d_k)$ $\quad d_k = 1$ for $k=0$, 0 otherwise

where V_k is the impulse response, $(N/D\ d_k)$ is a pole zero estimate of the impulse response. This problem is nonlinear.

Figure 3a

SPEECH

*

IMPULSE RESPONSE (V_k)

*

ALL-POLE LINEAR PREDICTION
(giving an original denominator estimate D')

*

*

*

SOLVE THE LINEAR PROBLEM:

minimize $(D/D')V_k - (N/D')d_k$ $\quad d_k = 1$ for $k=0$, 0 otherwise

USING THE NEW POLE ESTIMATE D' = D, TRY AGAIN

*

*

(At or near convergence, D approaches D' and the ideal case 3a.)

*

SYNTHESIS

Figure 3b