# INVERSE FILTERING TECHNIQUES IN SPEECH ANALYSIS

by

**M. A. Nwachuku**
*Electrical Engineering Department,*
*University of Nigeria. Nsukka.*

**ABSTRACT**

This paper reviews certain speech analytical techniques to which the label 'inverse filtering' has been applied. The unifying features of these techniques are presented, namely:

1. a basis in the source-filter theory of speech production,
2. the use of a network whose transfer function is the inverse of the transfer function of one or a combination of the articulatory system filters to modify the speech wave either in the time domain or in the frequency domain.

However their differences, which lie in the particular system filter being inverted and in the manner of realisation. provide a basis for the classification adopted in the paper which is as follows:

(1) inverse vocal tract analogue filtering.
(2) inverse vocal tract digital filtering.
(3) direct inverse glottal filtering.
(4) linear predictive coding.

An assessment of the comparative usefulness of inverse-filtering in contemporary speech studies is given.

**INTRODUCTION:**

The source-filter theory (1) of speech production leads naturally to the principle of inverse-filtering as a means for extracting the acoustic features of the speech wave. A schematic representation of the theory as it applies to the important class of speech sounds called vowels is shown in figure 1,(a). Here, the source is a periodic stream of pressure impulses p(t) and the filter is a cascade of three separate networks representing the glottis, the vocal tract, and the radiation process from the lips of the talker to the air. The three networks are generally assumed to act independently although Flanagan (2) has shown that glottis-vocal tract coupling is an essential mechanism in generating the glottal volume velocity waveform shown in Fig. 1 (b).

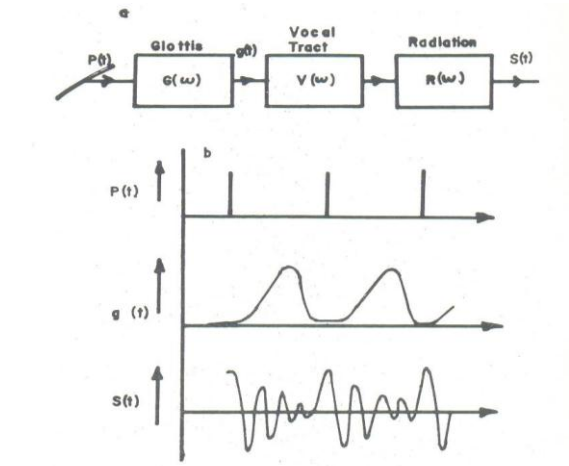


Fig. 1 Model of Speech Production
(a) Acoustic Filters
(b) Representative waveforms for a vowel.

In general only the speech waveform s(t) (usually its electrical analogue at the output of a microphone) is available for analysis. However various schemes of direct glottography, of which the laryngograph (3) is perhaps the simplest and most convenient example, have been devised for obtaining waveforms related to g (t). Depending on the area of application, speech analysis aims at a determination of one or more of the following features in the speech process:

(i) the resonant structure of the vocal-tract transfer function, i.e, formant analysis,
(ii) the glottal wave,
(iii) the fundamental frequency or pitch of the sound.

During the production of speech, the configuration of the articulators: the vocal tract tongue, teeth, lips, etc, changes from one sound to the next. However, considering a short segment of the speech of say 20-30 ms duration, the change is small. Consequently filters of the figure 1 may be assumed linear and time-invariant in this interval, and the concepts of linear filter theory developed in electrical engineering may be used.

Consider the linear net work shown in figure 2 and let the transfer function (i.e, ratio of output to input when the input is a single sine wave of singular frequency w) be $H_1(w)$.
This implies that if a time-varying signal x(t) is applied to the filter input the time varying signal output y(t), is given by

$$r(\omega) = H_1(\omega) \times (\omega) \qquad (1)$$

in equation (1)X(w)and X(t).and Y(w) and y(t)

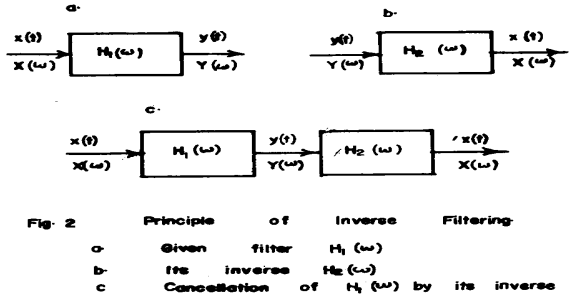

Fig. 2 Principle of Inverse Filtering
a- Given filter $H_1(\omega)$
b- its inverse $H_2(\omega)$
c Cancellation of $H_1(\omega)$ by its inverse

Are Fourier transform pairs. Next consider the network in figure 2(b), with transfer function $H_2$ (w) which allows x(t) to be recovered when cy(t) is applied at its input (c = constant), then

$$\times(\omega) = CH_2(\omega)\, r(\omega) \qquad (2)$$

Equation (I) and (2) lead to

$$r(\omega) = cH_1(\omega)\, H_2(\omega) r(\omega)$$

And

$$H_1(\omega)H_2(\omega) = C \qquad (3)$$

Thus the transfer function $H_2$ (w) is the reciprocal of the transfer function $H_1$ (w) (allowing for a gain factor) and equation (3) expresses the condition that filter A is the inverse of filter B The cascading of two networks each of which is the inverse of the other cancels the effect of the network.

## 2. INVERSE VOCAL TRACT (FORMANT) FILTERING:

The first application of *inverse* filtering in speech analysis was performed by Miller (4) who used it to derive, and study the properties of, the glottal *wave* g(t). Miller therefore sought to cancel the effect of the *vocal* tract transfer function V(w) by the use of another filter representing the inverse of the vocal tract. For ease of reference this method will be named inverse vocal tract filtering or inverse formant filtering. Miller hypothesized, and this has been justified by more elaborate versions of his technique, that satisfactory glottal wave extraction could be achieved by cancelling only the first formant ($F_1$) filter. Miller therefore passed the speech wave set) through the inverse filter whose parameters are adjusted by trial and error so that$F_1 = \frac{1}{2\pi\sqrt{L_\times C_\times}}$ (Hz) (see figure 3). It should be noted that Miller worked with realisable circuit elements in the time domain.

The introduction of the general purpose digital computer to speech analysis led to important elaborations of Miller's method. Matthews, Miller and David who had developed a pitch-synchronous technique for obtaining the poles and zeros of the speech signal by spectral matching (5) on a computer, used the formant information obtained therefrom to accurately derive the glottal wave- form by inverse filtering (6). Takasugi and Suzuki (7) combined the sophisticated digital techniques of fast Fourier transformer (FFT) and Analysis by Synthesis (ABS) with vocal tract inverse filtering in the frequency domain. In their technique the spectrum of the speech wave S(w) is found by FFT and this enables a first estimate of the first three formants $F_1$, $F_2$, $F_3$ to be made. Using the estimated formant values and other data including a combined source and radiation spectrum having a pole at 100Hz and 12db/octave fall, a spectrum S(w) is synthesized and compared with S(w). The former estimates of $F_1$ and $F_2$ are then revised and a new S(w) synthesized. The process of revision of formant estimates, resynthesis of S(w) and comparison with S(w) is repeated until the best match is obtained. V(w) the vocal tract transfer function is constructed from the most accurate formant values. The connection between the various transfer functions is:

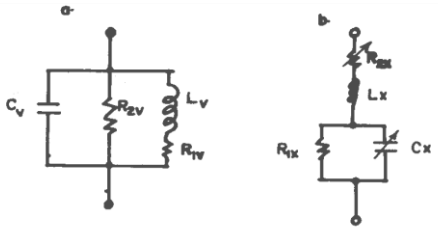

Fig 3 a- Formant network
b- Inverse Formant network (ref 4)

$$\frac{Lv}{Cx} = \frac{Lx}{Cv} = R_{1v} R_{2x} = R_{2v} R_{1x} = R^2$$

$$S(W) = G(W).V(W).R(W) \qquad (4)$$

where G (w) and R (w) are the transfer functions of the glottis and radiation respectively. By taking logarithms of equation (4), G (w) can be found by simple subtraction. The final step is to take the inverse Fourier transform of G (w) in order to find g(t):

$$g(t) = \frac{1}{\pi} Re \int_o^\infty G(\omega)\, ejwt\ dw \qquad (5)$$

In a later work (8) Takasugi and Suzuki improved the accuracy of their technique by fitting the extracted glottal wave to an analytical function. e.g. a raised cosine bell and using this function as the source wave in synthesis instead of the simple 12db/octave source spectrum.

$$\int_o^\infty G(\omega) e j\omega t\, d\omega$$

The two digital computer based methods described above also yield formant values not by inverse filtering but either by spectral matching based on pitch synchronous analysis or ABS. Nakatsui and Suzuki (9) have applied vocal tract inverse filtering to formant extraction. Starting from a short term spectrum of the speech wave calculated by FFT, the resonance spectra of two formants e.g. $F_1$ and $F_3$ plus a correction term for higher formants are Subtracted (i.e.. inverse filtering in the frequency domain) to leave the simple resonance spectrum of one formant, in this example $F_2$, At the beginning $F_1$ and $F_3$ are not known and must be guessed. Taking the first moment of the resulting spectrum gives an estimate for $F_2$ In the next step this value of $F_2$. the previous value of $F_3$ and the constant correction spectrum are removed from the speech

spectrum and $F_1$ is found by taking moments, The third and final step in the cycle consists in using the $F_1$ and $F_2$ values to estimate $F_3$ Several cycles are repeated until the frequency difference between corresponding formant frequencies in two successive cycles falls below a previously set threshold. This method like the ABS and pitch synchronous methods is iterative but algorithms were developed to substantially reduce the time of extraction to 025 sec/frame of 10 ms duration. It is thus faster than ABS but gives the same results with the same speech material.

## 3. INVERSE FILTERING BY LINEAR PREDICTIVE CODING:

Recently a new method of inverse filtering based on linear prediction theory has been introduced and demonstrated to be a powerful tool for the estimation of formant trajectory (10), (11), (12) for pitch period extraction (13), (14) and for automatic VU (voiced-unvoiced) decision (14). The action of linear predictive filtering under proper analysis conditions is illustrated in figure 4. showing: (a) the rapidly varying spectrum S(w) with the effect of glottal fine structure, and (b) a smooth approximation obtained by linear prediction Linear prediction In effect determines the filter that is the Inverse of the smooth spectral approximation. The inverse filter in turn IS that filter which minimizes the energy output when excited by the speech signal. An analytical formulation of this criterion or its related variants leads to an algebraic equation which can be solved without iteration for the filter coefficients. The connection between linear prediction and inverse filtering is indicated as follows.

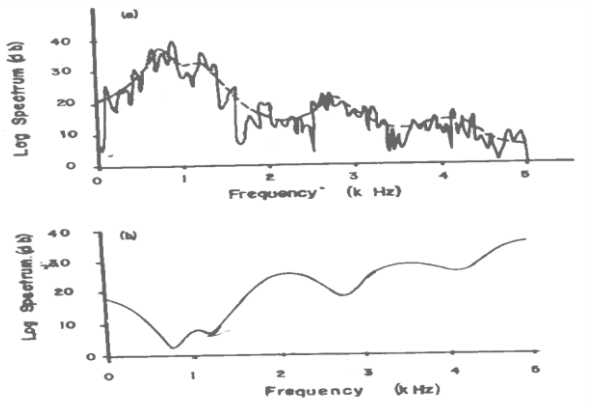


Fig 4 (a) spectrum of vowel |a| and its smoothing by linear prediction. (b) Spectrum of the inverse filter (adapted from ref 12)

Linear prediction seeks to predict a sample of the speech wave say at time nT by a linear weighted sum of p earlier samples taken at time

$t = (n - k)T, K = 1,2,3 \ldots \ldots \ldots P$ and we write:

$$\hat{s}(nT) = \sum_{K=1}^{P}(a_k s\{(n-k)T\}) \quad (6)$$

Or more simply

$$\hat{S}_n = \sum_{K=1}^{P} a_k sn - k \quad (7)$$

In equation (7) T is the sampling interval, n is an Integer, nT is the present instant at which the continuous speech signal S(t) is sampled, the $a_k$'s are the predictor coefficients. The difference between the predicted value $S_n$ and the actual $n^{th}$ sample of the speech signal $S_n$ gives the error $e_n$ in the $n^{th}$ sample

$$e_n = s_n - \hat{S}_n = s_n - \sum_{K=1}^{P} a_k s_n - k \quad (8)$$

Taking the z-transform of both sides of equation (8) gives

$$\mathrm{E(z)} = \mathrm{S(z)\,H(z)} \quad (9)$$

In equation (9) E(z) and S(z) are the z-transforms of $e_n$ and $S_n$ respectively and H (z) is given by

$$\mathrm{H(Z)} = 1 - \sum_{k=1}^{p} \mathrm{a_k\, z^k} \quad (10)$$

Analogous to equation (1), equation (9) is interpreted to mean that the sequence $e_n$ is the result of passing the speech sequence $S_n$ through a digital filter whose transfer function is H (z).If E(Z) may be approximated to a constant. This is the spectral equivalent of the minimum energy output criterion,) then H (z) is clearly seen to be the inverse of S(z).

Of course the spectrum S(w) of the speech signal is the value of S(z) on the unit circle in the z-plane.

i.e $Z = e^{jwt}$. In practice we seek to find that H(z) which is the inverse of the smoothed speech spectra as already explained. Success depends on a proper choice of p, T and the windowing function. Formant frequencies are readily estimated from the inverse filter spectrum by peak-picking algorithms.

## 4. INVERSE GLOTTAL FILTERING:

The cancellation of the glottal filter as a means of obtaining the spectral features of the vocal tract has not received much attention probably because of the difficulties associated with the glottal wave, and its spectrum. The glottal wave exhibits considerable variability between speakers, and for the same speaker, between utterances. Also the glottal filter is an all-zero network with the zeros occurring at a frequency interval of between 3 times to 5 times the pitch frequency. Some of these zeros may be right hand plane zeros so that the network is not minimum phase and the inverse filter will be unstable. Nwachuku & Newell (15) have carried out a preliminary study on the use of direct inverse glottal filtering in speech analysis. The attraction of the method stems from the ease with which the glottal wave may be obtained by laryngography (3). To overcome the problem of right hand plane zeros, a minimum phase network with spectrum $G_m$ (w) is defined such that

$$|\mathrm{G_m(w)}| = |\mathrm{G(W)}|$$

It then becomes possible to construct the inverse

of $G_m$ (w) A model of the glottal spectrum is used which permits the specification of the glottal spectrum by three pararmeters fp $d_p$, l, the pitch frequency, glottal duty factor and first zero location parameters respectively. Glottal spectra corresponding to several combinations of the glottal parameters were synthesized and applied in an inverse filtering process to the analysis of real vowel sounds using a Fortran programme run on an ICL 1907 digital computer.

The nature of the results obtained is indicated by figure 5. While the spectra of the resulting filtered waves are liable to give misleading formant values because of the crude model of the glottal spectrum used. zero-crossing analysis of the time-domain outputs yields a high frequency vowel-specifying parameter rather like Scully's (16) high frequency ripple. Zero crossing detection can a' so be applied to the speech wave itself to yield a low frequency vowel parameter similar to $F_1$. Thus the method of glottal inverse filtering appears to have possibilities in situations such as speech training for the deaf where a simply instrumented two- dimensional vowel indicator is required.

## 5. CONCLUSION:

Four distinct speech analytical techniques commonly described as inverse filtering have been reviewed and their bases explained. These are (1) analogue vocal tract inverse filtering (2) inverse vocal tract filtering combined with some method of formant estimation using a digital computer (3) linear predictive coding and (4) direct inverse glottal filtering. Method (1) is significant for demonstrating the fruitfulness of the concept of inverse filtering in speech analysis, and for its simplicity. The digital methods described under (2) have the same general requirements as method (3) but are not as effective while being inherently slower. Linear predictive coding has also been shown (12) to have important advantages over the established method of speech analysis by cepstral smoothing. (A discussion of cepstral analysis is outside the scope of this review. Briefly the cepstrum is defined as the Fourier transform of the logarithm of the spectrum. Applied to the speech signal, the effect is to separate the slowly varying component of the speech spectrum due to the vocal tract from the more rapidly varying Component due to the glottal source.) More work remains to be done in inverse glottal filtering using glottal parameters derived from the actual speech being analysed. The present indications are however that its usefulness will be restricted to very special applications. e.g. where a simply instrumented vowel indicator is required.

## REFERENCES:


(1) FANT, G: 'Acoustic Theory of Speech Production' (Mouton, 1960).

(2) FLANAGAN, J. L. LANDGRAF, L.L. 'Self-oscillating source for vocal-tract synthesizers' IEEE Trans-Audio and Electroacoust (1968) AU-16 pp 57-64.

(3) FOURCIN, A J. and ABBERTON, F: 'First Applications of a new laryngograph', Medical and Biological Illustrations. (1971) 21 pp 172-182.

(4) MILLER, R. L 'Nature of the Vocal Cord Wave', J. Aoust. Soc. Am., (1959) 31, pp 667-677.

(5) MATTTHEWS, M. V., MILLER, J. E, AND DAVID E. E: 'Pitch synchronous analysis of voiced sounds', J. Acoust. Soc. Am. (1961), 33 pp 179-186.

(6) MATTHEWS, M. V., MILLER, J. E. and DAVID, E. E: 'An accurate estimate of the glottal wave. Shape J. Acoust. Soc. Am. (1961) 33,843(A)

(7) TAKASUGI, T, and SUZUKI, J: 'Speculation of glottal waveform from speech wave', J. Radio Res. Lab. (Japan) 1968, 15 pp 279-293

(8) TAKASUGI, T, and SUZUKI, J: 'Considerations of Voice Source in Analysis by synthesis technique' J. Radio Res Labs (Japan) (1970) 17, pp 153-168.

(9) MAKATSUI, M, and SUZUKI, J: 'Formant Frequency Extraction using Inverse Filtering and Manual Calculation and its evaluation by synthethic speech. J. Radio Res Lab. Japan (1969) 16, pp 77-93.

(10) ATAL. B. S, and HANAUER. S L: 'Speech analysis and Synthesis by linear prediction of the speech wave'. J Acoust. Soc. Amer. (1971) 50, pp 637-655.

(11) MARKEL, J. D: 'Digital inverse filtering, a new tool for formant trajectory estimation', IEEE Trans-Audio and Electroacoust. (1972) AU-20 pp 129-137.

(12) MAKHOUL, J. 'Spectral Analysis of Speech by Linear Prediction', IEEE Trans-Audio & Electroacoust (1973), AU-21 pp 140-148.

(13) MAKSYM, J. N.: 'Real Time Pitch Extraction by Adaptive Prediction of the Speech Waveform', *IEEE Trans-Audio & Electroacoust* (1973) AU-21 pp 149-154.

(14) MARKEL, J. D. 'Application of a Digital Inverse filter for Automatic Formant and & FO Analysis', *IEEE Trans-Audio & Electroacoust* (1973) AU-21 pp 154-160.

(15) NWACHUKU, M. A, and NEWELL, A F., 'Preliminary investigation on the Analysis of Voiced Sounds by Direct Glottal Inverse

Filtering', (Submitted for publication).

(16) SCULLY, C: .Some acoustic Measures of Vowel Quality. University of Leeds Phonetics Dept. Report (1968) No.1 pp 41-48.