A MODEL FOR SIGNAL REPRESENTATION AT THE OUTPUT OF THE COCHLEA BASED ON MINIMUM PHASE SIGNALS

R. Kumaresan and Ashwin Rao

Department of Electrical Engineering
The University of Rhode Island
Kingston RI 02881
USA

ABSTRACT

We propose a model for signal representation at the output of the cochlea based on minimum phase (MP) signals. A minimum phase signal has the property that its phase and the logarithm of its envelope are related by Hilbert transformation. Hence by retaining either the phase or the envelope the entire signal can be represented to within a scale factor. First we show how to convert a given signal into one that is minimum phase by suitably adding a sinewave at the band-edge of the signal. Subsequently we show how to recover the original signal from the minimum phase signal's phase or envelope. With this procedure signals can be automatically scaled (Automatic Gain Control) to lie within a given dynamic range; we conjecture that such a procedure may indeed be used by the cochlea to represent signals that vary in intensity over several orders of magnitude.

1. INTRODUCTION

The 30,000 or so nerve fibers leading from the cochlea (inner ear) to the brainstem carry information about the acoustic signals that impinge on the tympanum. How are the signals decomposed and the information about its components (the frequencies, amplitudes, and phases) encoded by the cochlea and the auditory nerves? This has been a long standing question in the study of the auditory system. The answer to this, if found, apart from its intrinsic scientific value has technological applications from machine recognition of speech to improved design of hearing aids.

With this question in mind, and cognizant of the fact that speech signals are in some sense 'matched' to the human auditory system, we embarked on finding methods to decompose speech signals in a computer into narrow-band components with the hope that it may shed some light on the functioning of the auditory periphery. In the process we have devised improved methods based on frequency tracking algorithms to decompose voiced speech signals [1-5] modeled as follows:

$$s(t) = \sum_{k=1}^{N} a_k(t) \cos \phi_k(t) \quad . \tag{1}$$

This work was supported by the Air Force Contract AFOSR F49620-93-I-0026.

Although such algorithms have been found to be useful in technological applications such as tracking the fundamental frequency of speech signals and tracking of formant frequencies in speech, it is still unclear as to how the auditory front-end decomposes signals from multiple sources that have varied intensities. Recently we started scanning the vast literature related to the auditory system.

The textbook on Physiology of Hearing by Pickles [13] gives a readable account of the functioning of the auditory system as it is currently known. The auditory periphery consisting of the outer and middle ears, the cochlea, and the auditory nerve appears to behave roughly as a linear filter bank. The bandwidths of these filters increase with their center frequencies. However, there are several indications of nonlinearities in the system as well [13]. Some of these include effects such as: (a) "two-tone suppression", wherein the firing rate of a neuron to a tone input is reduced by applying another tone at a nearby frequency, (b) "distortion products generation", wherein application of two stimuli at frequencies f_1 Hz and f_2 Hz produce intermodulation products, the most significant being at $2f_1 - f_2$ Hz, (c) non-linear growth of responses with the intensity of input stimulus and presence of d.c. offsets in the basilar membrane position in the cochlea. For instance, Sachs and Young have shown [15] that mean rate of firings of neurons over the spectral region of 0–10 kHz does not resemble the spectral shape of the input stimulus at high stimulus levels. However, they show that the timing information in the nerve firings (and not the mean firing rate) retains the spectral information. Finally, there is a potentially revolutionary suggestion that at least for low level signals energy is actively injected into the traveling wave on the basilar membrane (see [13]). Many of the above nonlinear effects are believed to originate in the outer hair cells in the cochlea and is currently a matter of intense study [13, 14].

Understanding the structure and function of the components of the auditory front-end and explaining the above mentioned non-linear effects is a long-term prospect. Instead, imitating David Marr's approach to understanding the visual system [18], we may ask: what is the task faced by the auditory periphery and how it may handle these tasks with the hardware at its disposal? One of the tasks that the periphery is faced with is a large (80-90 dB) dynamic range of input stimulus. A number of cochlear models including those of Lyon [17] and Seneff [16] incorporate

1058-6393/95 $4.00 © 1995 IEEE

automatic gain control (AGC) in their models to deal with this problem. We propose a different kind of AGC mechanism based on representing the signal as a minimum phase signal. Whether such a mechanism is actually present in the auditory periphery remains to be seen.

We propose to make an educated guess of the signal processing performed by the front-end based on the following observations. One of the well known aspects of cochlear processing is that at low frequencies, the nerve firings 'phase-lock' to the input stimulus, *i.e.* the intervals between firings tend to cluster near multiples of the stimulus period. However, at stimulus frequencies beyond 4–5 kHz, the nerve firings tend to 'phase-lock' to the envelope of the stimulus. Hence, it is possible that the information about the stimulus may be carried by a signal representation which conveys information using either the envelope or the phase. Using this and other clues, we suggest that the cochlea transforms signals in each spectral region into minimum phase signals represented by (the real-part of)

$$e^{-\widehat{m}(t)} \exp\{j(\omega_c t + m(t))\} \quad ,$$

where ω_c is the center frequency ('place' information) and $m(t)$ is the phase deviation or the 'rate' information. In this process it also achieves the goal of automatic gain control. The Hilbert transform of $m(t)$ is denoted by either $\widehat{m}(t)$ or $\mathcal{H}\{m(t)\}$.

In this paper, after a brief discussion on minimum phase signals and its connections to single sideband angle modulation, we derive an expression for the envelope and phase of a signal containing two complex sinusoids. A multicomponent sinusoidal signal, under certain conditions, is then shown to exhibit the single sideband angle modulation or minimum-phase property; the strongest frequency component at the edge of the spectrum of the composite signal acts like the carrier. We further show that adding a suitable sinewave to any analytic signal results in a MP signal. The key point we make is that for a minimum phase signal either its angle or its envelope can be used as a signal representation. Is it possible that the cochlea uses the angle to represent the signal at low frequencies and the envelope to represent the signal at high frequencies? Finally, we propose a procedure to recover the original signal from the minimum phase signal's angle or envelope.

2. MINIMUM-PHASE (MP) SIGNALS

Minimum phase signals are the counterpart of their well known analog in the systems domain, known as minimum phase systems [7]. A minimum phase system $H(s)$ has all its poles and zeros on the left half of the s-plane and has a causal impulse response $h(t)$ [7]. For minimum phase systems, the log-magnitude and phase functions of the frequency response $H(\omega)$ form a Hilbert transform pair: $\angle H(\omega) = \mathcal{H}\{\ln|H(\omega)|\}$ [6, 7]. Analogous to this, if a complex signal $s(t) = |s(t)|e^{jm(t)}$ is analytic (its Fourier transform $S(\omega)$ vanishes for $\omega < 0$), and minimum phase then the phase of the signal and the logarithm of its envelope are related by Hilbert transformation, *i.e.* $m(t) = \mathcal{H}\{\ln|s(t)|\}$. Such a signal can be represented as

$$s(t) = e^{-\widehat{m}(t)} \exp\{jm(t)\} \quad , \qquad (2)$$

or equivalently as

$$s(t) = e^{\ln|s(t)|} \exp\left(j\mathcal{H}\{\ln|s(t)|\}\right) \quad , \qquad (3)$$

where $m(t)$ is the phase angle of the MP signal. Voelcker [8–10] systematically studied the properties of minimum phase signals, particularly their zero locations, and their relation to various modulation methods. Maximum phase signals display similar properties. For maximum phase signals, $m(t) = -\mathcal{H}\{\ln|s(t)|\}$ and their spectra are left-sided as opposed to the right-sided spectra of MP signals. In this paper we concentrate mainly on minimum phase signals. MP signals are closely related to single sideband angle modulated signals discussed below.

2.1. SINGLE SIDEBAND ANGLE MODULATION

The existence of a form of angle-modulation called single sideband angle modulation (SSB θM) was revealed by Powers (in 1958) and Bedrosian [11]. Traditional angle modulation is expressed as [19]

$$s_{\theta M}(t) = \exp\{j(\omega_c t + m(t))\} \quad ,$$

where $m(t)$ is the modulating signal and ω_c is the carrier frequency. In SSB θM, in addition to the conventional angle modulation, the signal is also amplitude modulated as below.

$$s_{SSB\ \theta M}(t) = e^{-\widehat{m}(t)} \exp\{j(\omega_c t + m(t))\} \quad . \qquad (4)$$

The resulting signal may be recognized as the minimum phase signal in Eq.(2) which has been translated in frequency by ω_c and has a one-sided spectrum. Note that while $s_{\theta M}(t)$ exhibits a double-sided spectrum, $s_{SSB\ \theta M}(t)$ has a spectrum that vanishes for all frequencies less than ω_c. Another interesting aspect of $s_{SSB\ \theta M}(t)$ is that the message $m(t)$ or equivalently $\widehat{m}(t)$, can be extracted from either the envelope or the phase by traditional demodulation procedures.

3. ENVELOPE AND PHASE OF A TWO-COMPONENT SIGNAL

In this section we show an example of a minimum phase signal. Consider a signal $s(t)$ consisting of two complex sinusoids with *complex* amplitudes $A_1 = a_1 e^{j\theta_1}, A_2 = a_2 e^{j\theta_2}$ and angular frequencies ω_1, ω_2 respectively. Let $\Delta\omega_2 = \omega_2 - \omega_1$ and $\Delta\theta_2 = \theta_2 - \theta_1$; $\Delta\omega_2 > 0$.

$$s(t) = A_1 e^{j\omega_1 t} + A_2 e^{j\omega_2 t} \quad . \qquad (5)$$

If $|A_1| > |A_2|$, then $s(t)$ can be written as

$$s(t) = a_1 e^{j(\omega_1 t + \theta_1)} \left(1 + \frac{a_2}{a_1} e^{j(\Delta\omega_2 t + \Delta\theta_2)}\right) \qquad (6)$$

Let $z(t) = 1 + \frac{a_2}{a_1} e^{j(\Delta\omega_2 t + \Delta\theta_2)}$. Then

$$|z(t)| e^{j\angle z(t)} = 1 + \frac{a_2}{a_1} e^{j(\Delta\omega_2 t + \Delta\theta_2)} \qquad (7)$$

Our assumptions imply $|z(t)| > 0$. Taking the natural logarithm of both sides,

$$\ln|z(t)| + j\angle z(t) = \ln\left(1 + \frac{a_2}{a_1} e^{j(\Delta\omega_2 t + \Delta\theta_2)}\right) \quad . \tag{8}$$

Using the series expansion $\ln(1+y) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} y^k$,

$$\ln|z(t)| + j\angle z(t) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a_2}{a_1}\right)^k e^{jk(\Delta\omega_2 t + \Delta\theta_2)}. \tag{9}$$

Equating the real and imaginary parts,

$$|z(t)| = \exp\left\{\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a_2}{a_1}\right)^k \cos k(\Delta\omega_2 t + \Delta\theta_2)\right\} \tag{10}$$

$$\angle z(t) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a_2}{a_1}\right)^k \sin k(\Delta\omega_2 t + \Delta\theta_2). \tag{11}$$

But from (6) and (7), $|s(t)| = a_1|z(t)|$ and $\angle s(t) = \omega_1 t + \theta_1 + \angle z(t)$. From these and Eqs. (10) and (11), the envelope and phase of $s(t)$ are

$$|s(t)| = a_1 \exp\left\{\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a_2}{a_1}\right)^k \cos k(\Delta\omega_2 t + \Delta\theta_2)\right\} \tag{12}$$

$$\angle s(t) = \omega_1 t + \theta_1 + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a_2}{a_1}\right)^k \sin k(\Delta\omega_2 t + \Delta\theta_2). \tag{13}$$

Thus, the two-tone signal $s(t)$ can be expressed as a SSB θM signal

$$s(t) = a_1 e^{-\widehat{m}(t)} \exp\{j(\omega_1 t + \theta_1 + m(t))\} \quad , \tag{14}$$

where

$$m(t) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a_2}{a_1}\right)^k \sin\{k(\omega_2 - \omega_1)t + k(\theta_2 - \theta_1)\} \tag{15}$$

and hence $s(t)$ is a MP signal translated by ω_1. However, if $|A_2| > |A_1|$, then $s(t)$ is a maximum phase signal translated by ω_2. Voelcker [8] has an equivalent derivation of this result. We next discuss an extension of this result to a multitone signal.

4. MULTICOMPONENT SINUSOIDAL SIGNAL AND MINIMUM PHASE PROPERTY

Let $s(t)$ be a signal composed of M sinusoidal components:

$$s(t) = \sum_{k=1}^{M} A_k e^{j\omega_k t} \quad . \tag{16}$$

A_k is the complex amplitude of the k^{th} component and ω_k its angular frequency. Let $\Delta\omega_k = \omega_k - \omega_1$ and $\Delta\theta_k = \theta_k - \theta_1$. Assume that $|A_1| > \left|\sum_{k=2}^{M} A_k e^{j\Delta\omega_k t}\right|$ and $\omega_1 = \min\{\omega_k : k = 1, \ldots M\}$. $s(t)$ can be expressed as

$$s(t) = A_1 e^{j\omega_1 t}\left(1 + \sum_{k=2}^{M} \frac{A_k}{A_1} e^{j\Delta\omega_k t}\right) \quad . \tag{17}$$

As before, it is possible to show using (17) that

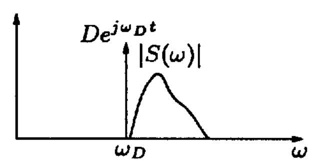
$$\sum_{k=1}^{M} A_k e^{j\omega_k t} = A_1 e^{-\widehat{m}(t)} \exp\{j\{\omega_1 t + m(t)\}\} \quad . \tag{18}$$

Comparing Eqs. (18) and (4), it can be seen that a multicomponent sinusoidal signal (given the assumptions) is a frequency-translated minimum phase signal. Note that here we have assumed that the sinusoid at the lower band-edge is larger than the sum of others and hence the signal is MP. However if some component in the middle of the spectrum is large, then the signal is mixed phase and can not be expressed as in Eq. (18). Next, we show how any signal can be converted into a minimum phase signal and subsequently represented by the angle of that MP signal.

5. REPRESENTING ANY SIGNAL BY THE PHASE ANGLE OF A MP SIGNAL

Given any bandpass signal $s(t) = s_R(t) + j\,\widehat{s_R}(t) = |s(t)|e^{j\angle s(t)}$, the signal representation in (18) may be no longer valid because $s(t)$ could be mixed phase. Now consider a signal $s_{MP}(t)$ obtained by adding a sinewave $De^{j\omega_D t}$ (as shown in figure below) at the lower band-edge of $s(t)$:

$$s_{MP}(t) = s(t) + D\exp\{j\omega_D t\} \quad . \tag{19}$$



If the constant D is such that $D > |s(t)|$ then

$$\begin{aligned} \ln\left(\frac{s_{MP}(t)}{De^{j\omega_D t}}\right) &= \ln\left(1 + \frac{s(t)}{D}\exp\{-j\omega_D t\}\right) \\ &= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}\left(\frac{s(t)e^{-j\omega_D t}}{D}\right)^k \end{aligned} \tag{20}$$

Each term in the right-hand-side of the above equation has a right-sided spectrum; their sum results in a one-sided spectrum too. Thus $s_{MP}(t)$ can be represented as

$$s_{MP}(t) = De^{j\omega_D t} e^{-\widehat{m}(t)} e^{jm(t)} \quad , \tag{21}$$

where,

$$m(t) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \frac{|s(t)|^k}{D^k} \sin\{k(\angle s(t) - \omega_D t)\} \quad , \tag{22}$$

The new signal $s_{MP}(t)$ turns out to have minimum phase property, and hence $m(t)$ may be used to represent the given signal $s(t)$. Also, the MP angle $m(t)$ in (22) is continuous within the principal value and therefore can be easily obtained by an arctan operation; mixed phase signals, in general, display 2π ambiguities. Observe that although the original signal is bandpass, $m(t)$ is generally unlimited in bandwidth. However by choosing D to be large, only a few terms in (22) may be significant. It seems possible to extend this procedure by adding, instead of a tone $De^{j\omega_D t}$, any minimum phase signal $D(t)$. In this case $D(t)$ will appear in the denominator of (20). Note that the reciprocal of $D(t)$, the *inverse of the minimum phase signal*, of course exists and is also minimum phase.

In principle, any real-valued signal $s_R(t)$ can be converted into a minimum phase signal by injecting a sinewave at the band-edge of its analytic version. A block diagram showing this procedure is given in Figure 1. Given any signal $s_R(t)$, we first form its analytic version $s(t) = s_R(t) + j\ \widehat{s_R}(t)$. Next, we form a new signal $s_{MP}(t) = D\exp(j\omega_D t) + s(t)$; D satisfies the conditions mentioned earlier. The D shown in Figure 1 may be chosen based on a rough estimate of the signal envelope itself. Finally, only the angle $m(t)$ of $s_{MP}(t)$ is retained to represent the given signal. We will now show a way to recover our original signal from such a representation; of course we need the $De^{j\omega_D t}$ that we added to begin with.

6. RECONSTRUCTION PROCEDURE

Voelcker [10] proposed a method to demodulate single sideband signals using envelope detection. His method is related to the reconstruction of a signal from the envelope of a minimum phase signal. Instead, we propose a method that uses only the phase angle of the MP signal. In Figure 2 we have shown a block diagram illustrating this procedure. The input is the angle $m(t)$ obtained in Figure 1; it is Hilbert transformed to obtain $\widehat{m}(t)$. Using the fact that $s(t) = s_{MP}(t) - D\exp(j\omega_D t)$ from (19), the signal $s(t)$ is obtained by the following operation

$$s(t) = De^{j\omega_D t}e^{-\widehat{m}(t)}\exp\{jm(t)\} - De^{j\omega_D t}. \qquad (23)$$

The original signal $s_R(t)$ is then recovered by taking the real part of $s(t)$; one could also get $s_R(t)$ directly by replacing $e^{j(\ )}$ by cos( ) in Figure 2.

7. DISCUSSION AND CONCLUSIONS

The key idea we are advocating is that if a signal is minimum phase or transformed into minimum phase, then either the phase or the envelope can convey the essential information about the signal. In order to represent a minimum phase signal as in Eq.(22) we need to convey information about D (a dc level), ω_D (the 'place' information) and the phase deviations around ω_D, *i.e.* $m(t)$ (the 'rate' information). Note that $m(t)$ will lie between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ and hence we have achieved some sort of an automatic gain control (AGC). This can be done for each band-pass region separately. Further notice that in order to extract these information, no explicit Hilbert transform calculations are needed; a phase detector constructed with a frequency discriminator followed by an integrator is sufficient.

8. REFERENCES

[1] R. Kumaresan and C.S. Ramalingam, *Proc. of 26th Asilomar Conf. Sig., Syst. and Comp.*, CA, Oct. 1992, pp. 472–476.

[2] R. Kumaresan and C.S. Ramalingam, *Proc. of 27th Asilomar Conf. Sig., Syst. and Comp.*, CA, Oct. 1993, pp. 1041–1046.

[3] R. Kumaresan, C.S. Ramalingam and A. Rao, *Proc. of Seventh SSAP Workshop*, Quèbec City, Canada, Jun. 1994, pp. 207–210.

[4] C.S. Ramalingam, A. Rao and R. Kumaresan, *Proc. of IEEE-SP Symposium on Time-Frequency and Time-Scale Analysis*, pp 500–503, Oct. 1994, PA,USA.

[5] A. Rao and R. Kumaresan, Submitted to IEEE ICASSP-95.

[6] H.W. Bode, *Network Analysis and Feedback Amplifier Design*, New York: Van Nostrand, 1945.

[7] A. Papoulis, *The Fourier Integral and its Applications*, New York: McGraw-Hill, 1962.

[8] H. B. Voelcker, *Proc. of IEEE*, March 1966, vol. 54, No. 3,pp. 340– 353.

[9] H. B. Voelcker, *Proc. of IEEE*, May 1966, vol. 54, No. 5,pp. 735– 755.

[10] H. B. Voelcker, *IEEE Trans. On Comm. Tech*, Feb 1966, vol. COM- 14,No.1, pp 22–30.

[11] E. Bedrosian, *Proc. IRE*, Oct 1962, vol. 50, pp. 2071–76.

[12] E. Bedrosian, *Proc. of IEEE*, May 1963, pp 868–869.

[13] J. O. Pickles, *An Intro. to the Physiology of Hearing*, 2nd ed., London: Academic Press, 1988.

[14] P. Dallos, Ch. 5 of *Auditory Function: Neurobiological Bases of Hearing*, G.M. Edelman, W.E Gall, and W.M. Cowan (Eds.), New York: John Wiley & Sons, 1988.

[15] M.B. Sachs and E. D. Young, *J. Acoust. Soc. Am*, 68:858-875.

[16] S. Seneff, *Journal of Phonetics*, 1988,vol 16, pp. 55–76.

[17] M. Slaney and R. F. Lyon, "On the importance of time–A temporal representation of sound,"in *Visual Representations of Speech Signals*, M. Cooke, S. Beete, and M. Crawford, eds., John Wiley and Sons, Sussex, England, 1993.

[18] D. Marr, *Vision: A Computational Investigation Into The Human Representation And Processing Of Visual Information*, San Fransisco: W.H. Freeman, 1982

[19] S. Haykin, *Communication Systems*, New York : Wiley, 1978.

ACKNOWLEDGMENTS: The authors wish to thank Mr. C.S. Ramalingam and Dr. A.G. Sadasiv for many useful discussions.

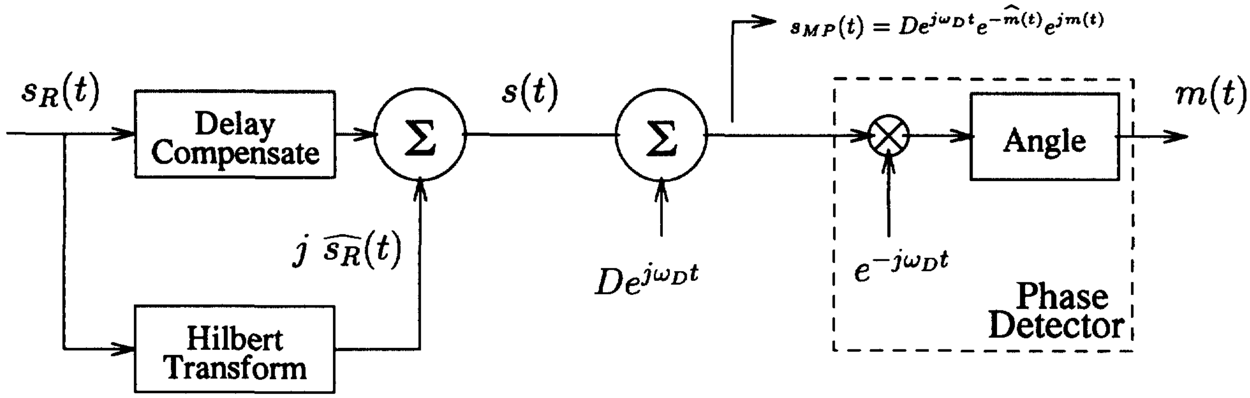


Fig.1 shows a way to represent any signal $s_R(t)$ by the phase angle of a MP signal; the procedure could be implemented on a computer. The delay compensation for $s_R(t)$ depends on the Hilbert transformer used. $s_{MP}(t)$ stands for the translated minimum phase signal. The output $m(t)$ is a representation for the signal $s_R(t)$.

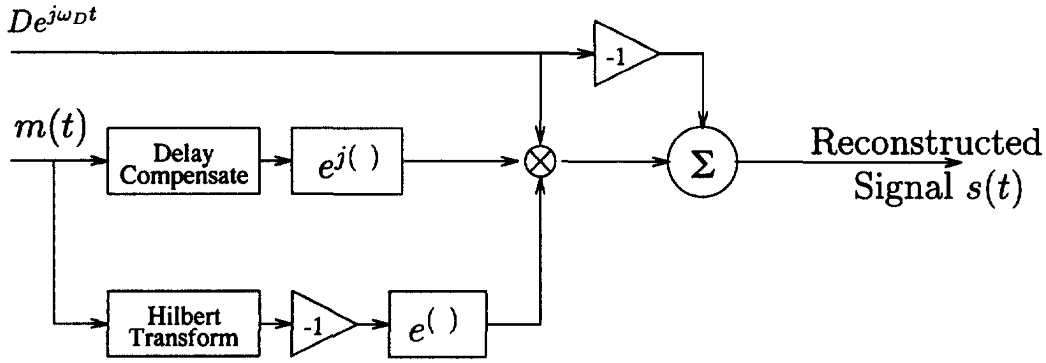


Fig.2 shows the signal reconstruction procedure using the MP angle $m(t)$ from Fig.1. The mathematical expression corresponding to the block diagram is: $s(t) = De^{j\omega_D t}e^{-\widehat{m}(t)}\exp\{jm(t)\} - De^{j\omega_D t}$. $s_R(t)$ would be the real part of the reconstructed signal $s(t)$; it could also be obtained directly by replacing $e^{j(\ )}$ in figure with cos( ).