AN FM DEMODULATION ALGORITHM WITH AN UNDERSAMPLING RATE

YIYUAN XIONG, YI HUANG, J. F. RALPH, W. AL-NUAIMY, PU SUN

Department of Electrical Engineering and Electronics
The University of Liverpool, Brownlow Hill, Liverpool, L69 3GJ, UK

YYXIONG@LIV.AC.UK, YI.HUANG@LIV.AC.UK

ABSTRACT

The existing method for sampling a modulated signal is the same as that for a non-modulated message signal. Therefore the quadrature sampling rate for a modulated signal must be larger than its bandwidth. However, a new algorithm, which can recover completely a DC-free message signal from an FM signal at a sampling rate less than its bandwidth, is presented in this paper. The new sampling rate limit is also discussed. The investigation shows that the current sampling theorem does not present an optimal sampling rate for recovering the message signal from an FM signal and further modification of the theorem might be needed.

1.INTRODUCTION

The sampling theorem is fundamental for signal processing and modern communications. The sampling techniques vary with the properties of the signals to be sampled. These properties are categorized in several ways: real or analytic [1], modulated or non-modulated (also called message), baseband or passband. Sampling an analytic signal is practically implemented by a technique called quadrature sampling [2]. In this paper, we only use the first term, analytical signal sampling, for convenience. Sampling a modulated signal in a different way is the major contribution of this paper.

The current sampling theorem, for modulated or non-modulate, baseband or passband, is that the sampling frequency must be larger than *2B* Hz to reconstruct the signal from its samples (in some cases, for a real passband signal, the sampling rate can be higher than one above) for a real signal with a bandwidth of *B* Hz. If the signal is analytic, the sampling rate must be larger than *B* Hz [2].

In some modulation schemes, for example, FM, FSK CDMA, etc., a modulated signal might have a wider bandwidth B_M Hz than that of the corresponding message signal, i.e. *B* Hz.. As a result, the sampling rate for such a modulated signal is higher than that for a message signal. However, this conclusion is drawn under the assumption that the message signal cannot be recovered if spectrum aliasing (also called overlap) of the modulated signal occurs when under-sampled. In fact, spectrum aliasing of the modulated signal does not necessarily lead to spectrum aliasing of the message signal. Hence, even though the modulated signal cannot be recovered, it is still possible to reconstruct the message signal.

As the modulation schemes are different, each reconstruction method should be designed individually. In this paper, we only discuss the case of sampling an FM signal.

2. FM SIGNAL MODEL

In an FM communication system, we consider the modulating signal as an arbitrary message signal *m(t)*, produced by an information source. It is assumed in this paper that (1) this signal is band-limited, with a bandwidth of *B* Hz; (2) its envelope is upper-bounded by m_p, that is

$$|m(t)| \le m_p \tag{1}$$

(3) the message signal has no DC component, otherwise the recovered signal has a DC bias, and fortunately, this assumption is true in most FM communications systems; and (4) there is no noise interference in the channel.

We denote the Fourier transform of *m(t)* as *M(Ω)*,then

$$m(t) = \frac{1}{2\pi}\int_{-2\pi B}^{2\pi B} M(\Omega)\exp(j\Omega t)d\Omega \tag{2}$$

$$M(\Omega) = \int_{-\infty}^{\infty} m(t)\exp(-j\Omega t)dt \tag{3}$$

An analytic FM modulated signal is given as

$$\phi_{FM}(t) = A\exp[j(\omega_c t + k_f \int_0^t m(\alpha)d\alpha)] \tag{4}$$

where

A the amplitude of the sinusoidal signal
ω_c the carrier frequency and
k_f the frequency modulation constant.

As a consequence, the frequency deviation Δf in Hz is

$$\Delta f = \frac{k_f m_p}{2\pi} \tag{5}$$

and the deviation ratio β is

$$\beta = \frac{\Delta f}{B} \tag{6}$$

Therefore, according to Carson's rule [3], the bandwidth of the modulated signal can be approximated to

$$B_{FM} = 2B(\beta+1) = 2(\Delta f + B) = 2(\frac{k_f m_p}{2\pi} + B) \tag{7}$$

For convenience, we define

$$\Theta(t) = k_f \int_0^t m(\alpha)d\alpha \tag{8}$$

Without loss of generality, we can assume ω_c equals to zero, which means that the discussion is focused on the baseband FM signal. This assumption can be realized via a mixer to shift the carrier frequency to zero. Hence, the equation (4) becomes

$$\phi_{FM}(t) = A\exp[j\Theta(t)] \tag{9}$$

3. THE UNWRAPPING DEMODULATION ALGORITHM

3.1. Equivalent FM System Model

If the periodic sampling frequency is f_s, and the sampling period is $T=1/f_s$, then the sampling discrete-time analytic signal is

$$\psi_{FM}(n) = \phi_{FM}(nT) = A\exp[j\Theta(nT)] \tag{10}$$

Where $\Theta(nT)$ stands for the sampled instantaneous phase and is an unlimited real number rather than a phase between $-\pi$ and π. Then, there must exist an integer number l_n, a wrapping index, which satisfies

$$2l_n\pi - \pi \le \Theta(nT) < 2l_n\pi + \pi \tag{11}$$

Unfortunately, the unlimited instantaneous phase $\Theta(nT)$ cannot be reconstructed directly from the analytic signal $\psi_{FM}(n)$. Actually $\psi_{FM}(n)$ is represented by its real and imaginary parts when $\Theta(t)$ is sampled. So the directly recovered phases are bounded in a range between $-\pi$ and π, that is

$$\theta(n) = \arg(\psi_{FM}(n)) \tag{12}$$

where $\arg(\bullet)$ denotes the angle of the quantity in radians between $-\pi$ and π. Considering the inequality (11), we have

$$\theta(n) = \arg(A\exp[j\Theta(nT)]) = \Theta(nT) - 2l_n\pi \tag{13}$$

This means that $\theta(n)$ is wrapped from the unlimited instantaneous phase $\Theta(nT)$ into a phase between $-\pi$ and π. For example, when $\Theta(nT)$ is $4.8\,\pi$, $\theta(n)$ will be $0.8\,\pi$, and l_n will be 2.

Compute the consecutive wrapped phase difference, we obtain

$$d(n) = \theta(n) - \theta(n-1) \tag{14}$$

$$= \Theta(nT) - \Theta((n-1)T) - 2\pi(l_n - l_{n-1})$$

$d(n)$ has a range of $[-2\pi, 2\pi)$. We can wrap it into the range of $[-\pi, \pi)$, that is

$$y(n) = d(n) - 2l_d\pi \tag{15}$$

and

$$-\pi \le y(n) < \pi \tag{16}$$

Where l_d is the wrapping index. From (15),

$$y(n) = \Theta(nT) - \Theta((n-1)T) - 2\pi(l_n - l_{n-1} - l_d) \tag{17}$$

$$= k_f\int_{(n-1)T}^{nT} m(\alpha)d\alpha - 2\pi k_n \tag{18}$$

where $k_n=l_n-l_{n-1}-l_d$ is the wrapping index for $y(n)$, which has combined all the wrapping effect. Define the integral term as

$$x_c(t) = k_f\int_{t-T}^{t} m(\alpha)d\alpha$$

It is easy to verify that

$$x_c(t) = k_f m(t) * g(t) \tag{19}$$

where (*) denotes a convolution integral, and $g(t)$ is a gate function defined as

$$g(t) = \begin{cases} 1 & 0 \le t < T \\ 0 & otherwise \end{cases} \tag{20}$$

Therefore (18) can be expressed as

$$y(n) = x_c(nT) - 2\pi k_n \tag{21}$$

$$= [k_f m(t) * g(t)]\big|_{t=nT} - 2\pi k_n \tag{22}$$

Eq. (22) clearly shows that the discrete time signal $y(n)$ can be considered as a sampling sequential of an analogue signal $x_c(t)$ with a sampling period of T, while having a bias of $2\pi k_n$.

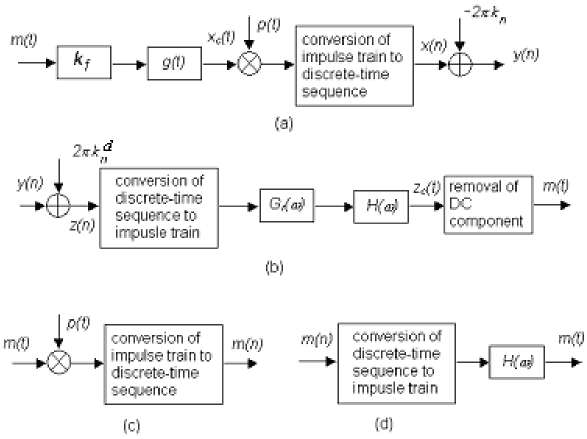

Fig. 1. The equivalent systems using Oppenheim time normalization model (a) the model for equation (22) (b) the FM reconstruction model (c) sampling a non-modulated signal (d) the reconstruction of a sampled non-modulated signal

The signal $x_c(t)$ is a result of the message signal $m(t)$ passing through a multiplication with a coefficient of k_f, and an LTI system with a rectangular impulse response $g(t)$.

Using the Oppenheim time normalization model [4] to analyse the conversion of an analogue signal into a discrete-time signal or vice versa, the above sampling and relevant FM processing can be simplified to an equivalent model as Fig 1(a). Where $p(t)$ is the sampling function, an impulse train, that is

$$p(t) = \sum_{n=-\infty}^{\infty}\delta(t - nT)$$

The part of the multiplying between $x_c(t)$ and $p(t)$, and the conversion of the impulse train to discrete-time sequence can be considered as an ADC device converting the analogue signal $x_c(t)$ to a discrete–time sequence $x(n)$. That is

$$x(n) = x_c(t)\big|_{nT} = k_f\int_{(n-1)T}^{nT} m(\alpha)d\alpha \tag{23}$$

Similarly for a non-modulated signal as shown in figure 1 (c), we have

$$m(n) = m(t)\big|_{t=nT} = m(nT)$$

As g(t) is an LTI system, so the Fourier transform of $x_c(t)$ is the product of $k_f M(\Omega)G(\Omega)$, where $G(\Omega)$ is the Fourier transform of $g(t)$. When $T<1/B$, the main lobe of $G(\Omega)$ is wider than the bandwidth of $M(\Omega)$, hence the bandwidth of $x_c(t)$ is the same as that of $M(\Omega)$. Consider that $x_c(t)$ is still a real baseband signal, we have

$$f_s > 2B \tag{24}$$

Where B is the bandwidth of the message signal rather than the bandwidth B_{FM} of the modulated signal. However, the final sampling rate requirement must also confirm with (24) as well as the following unwrapping requirement in section 3.2.

To reconstruct the original signal, for normal sampling, a DAC and a lowpass filter is needed as shown in figure 1(d). The DAC can be considered as a device converting the discrete-time sequence to an impulse train. The lowpass filter has a stop frequency of B.

Thus, we can design the reconstruction model for an FM signal as shown in Fig 1 (b). Three more operations are needed besides the DAC and the lowpass filter.

(1). the first operation is to cancel (we call it unwrapping) the phase bias, if it exists, by adding a number of $2\pi k_n^d$. That is,

$$z(n) = x(n) + 2\pi(k_n^d - k_n) \tag{25}$$

It is obvious that if the unwrapping integer number k_n^d is exactly the same as k_n,, the reconstructed signal *z(t)* from the FM signal will be the same as *m(t)*. If ($k_n^d - k_n$) is a nonzero constant, which is independent of *n*, then $z_c(t)$ has a DC difference from *m(t)*. However, after removal of the DC component of $z_c(t)$, *m(t)* can be recovered completely. The implemention of the above addition will be discussed in detail in section 3.2.

(2). The second operation is to cancel the effect introduced by *g(t)*. This can be implemented by an LTI system which has a frequency response $G_r(\Omega) = 1/G(\Omega)$. Note that, as the signal of interest is in the frequency range between $-B$ and B, the outside band response can be of any value.

As long as $T<1/B$, there will be no singularity points in $G_r(\Omega)$. Hence such an LTI system is realizable.

(3). The last operation is to remove the DC component.

3.2. The Phase Unwrapping

In this section, we analyse the factors of the bias, the unwrapping method and the sampling rate limit for unwrapping.

Theorem 1: if the sampling frequency satisfies $fs>2\Delta f$, then $k_n=0$ for all *n*.

From Theorem 1, it is apparent that when $fs>2\Delta f$, there is no phase bias between *y(n)* and *x(n)*, so no unwrapping operation is needed. The whole system works similarly to a normal non-modulated sampling system except *g(t)* and $G_r(\Omega)$. Recall from (2) that, in all cases, f_s must also satisfy $f_s>2B$, so the no-phase-bias sampling rate is determined by the larger one between B and Δf. In a narrow-band FM system, as $B>\Delta f$, the minimum sampling rate is $2B$ Hz. In a wide-band FM system, $\Delta f>B$, the minimum sampling rate f_s is

$$f_s = 2\Delta f = B_{FM} - 2B \approx B_{FM} \tag{27}$$

As a result, when the sampling frequency is larger than the modulated signal bandwidth B_{FM}, no matter whether it is a narrow-band or wideband signal, the message *m(t)* can be simply obtained from *y(n)*.

But, when $f_s < 2\Delta f$, part of *x(n)* will be out of the range between $-\pi$ and π, as part of *y(n)* has a bias to *x(n)* because of wrapping. Figure 2 shows an example,

However, if the signal *x(n)* does not change fast enough, even though $|x(n)|$ exceeds π, it can also be recovered by unwrapping *y(n)*. For example, if $z(n)=0.9\pi$ (note that the recovered signal is denoted as *z(n)* rather than *x(n)*) and $y(n+1)= -0.8\pi$, then we can get $z(n+1)=1.2\pi$ rather than -0.8π or 3.2π. This unwrapping process has a mechanism, which selects automatically a proper number of k_n^d to match (not necessarily equal) k_n, thus recovers *x(n)*.

Theorem 2: when the signal *x(n)* satisfies

$$|x(n+1) - x(n)| < \pi \tag{28}$$

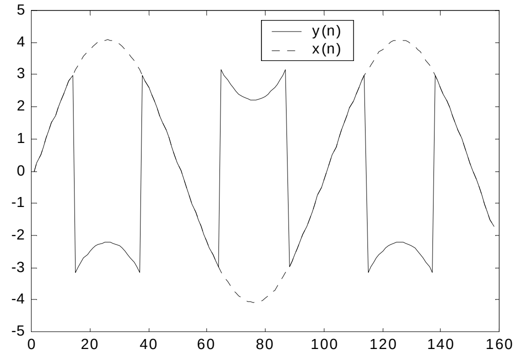

Fig. 2. the difference between *x(n)* and *y(n)*

the unwrapping process of *y(n)*, which restrict *z(n)* in the bound as $|z(n+1) - z(n)| < \pi$, automatically chooses the unwrapping index k_n^d to satisfy

$$z(n)=x(n)+2\pi c$$

and

$$k_n^d - k_n = c$$

where c is an integer constant, independent with *n*.

As a result, when (28) holds, the recovered signal *z(n)* only has a DC difference from *x(n)* if the constant *c* is not zero. If the message signal is free of DC component, it is easy to reconstruct *x(n)* by removing the DC component from *z(n)*.

Theorem 3: If

$$\Phi_M = \frac{k_f}{\pi}\int_0^{2\pi B} |M(\Omega)| \frac{(2 - 2\cos(\Omega T))}{\Omega} d\Omega < \pi \tag{29}$$

then

$$|x(n+1) - x(n)| < \pi$$

The above bound (29) shows the link between the sampling period T and the structure of the message signal and other FM parameters. This is the key inequity to determine the sampling rate limit. It is easy to verify, when $T<(1/2.7B)$, the function $\frac{(2-2\cos(\Omega T))}{\Omega}$ is a monotonically increasing function in terms of Ω in the range between 0 and $2\pi B$. In such a case $T<(1/2.7B)$, the more energy in the higher frequency band, the larger the number Φ_M is. The extreme case happens when the message *m(t)* is a sinusoidal signal with the highest frequency of $2\pi B$. That is

$$m(t) = m_p \cos(2\pi Bt + \theta_0)$$

Using its Fourier transform as substitution into (29), we have

$$\Phi_M = \frac{2m_p k_f (1 - \cos(2\pi BT))}{2\pi B} < \pi$$

The solution to the inequality (33) in terms of T or f_s is given by

$$f_s > \frac{2\pi B}{acr\cos(1 - \frac{\pi}{2\beta})} \tag{30}$$

When $2\pi BT$ is very small, we can approximate $\cos(2\pi BT)$ to $[1-(2\pi BT)^2/2]$. Then a slightly conservative estimate of f_s (which explicitly shows the relationship with B_{FM}) is

$$f_s > \sqrt{\pi\beta}(2B) = \frac{\sqrt{\pi\beta}}{\beta+1} B_{FM} \quad (31)$$

From (30) or (31), it can be concluded that the sampling rate can be much lower (especially when β is large) than the conventional sampling rate, i.e. B_{FM}. Note that these bounds in (30) or (31) are derived under the assumption of *T<(1/2.7B)*. If the sampling rate estimated from (30) or (31) is lower than *2.7B*, then they are invalid. This happens when β is a small number less than 1, and the valid rate can be determined only from (30). We define the reduction ratio η = min(f_s)/B_{FM} and the restriction ratio ρ $=2.7B/B_{FM}$,, thus (30) and (31) are only valid for $\rho > \eta$.

Figure 3 shows the ratio ρ and the reduction ratio η derived from (30) or (31) versus the deviation ratio β.

Furthermore, if the message signal is not a single tone, then there are some components in the lower frequency and hence the sampling frequency can be further smaller. For example, when a signal has a uniformly distributed spectrum, the sampling frequency can be only

$$f_s > \sqrt{0.5\pi\beta}(2B) = \frac{\sqrt{0.5\pi\beta}}{\beta+1} B_{FM}$$

It is only 70% of that for a single tone. If β = 20, the sampling rate is 27% of B_{FM}.

3.3. Summary of the algorithm

The algorithm can be summarized as follows.

1) Sample the modulated signal with a sampling rate constrained by (31), or more precisely by (309) to obtain the real part $R(n)$ and imaginary part $I(n)$ of $\psi_{FM}(n)$.
2) Compute $\theta(n)=arg[I(n)+jR(n)]$, which is the phase of $\psi_{FM}(n)$, where $j=\sqrt{-1}$.
3) Compute $d(n)= \theta(n)- \theta(n-1)$.
4) Compute $l_n^d = -floor[\frac{d(n)-z(n-1)}{2\pi}+0.5]$.
5) Unwrap $d(n)$ to $z(n)$, i.e. compute $z(n) = d(n) + 2\pi l_n^d$.
6) Filter the signal of $z(n)$ by an FIR digital filter for which the corresponding analogue filter has a response of $G_r(\omega)H(\omega)$.
7) Remove the DC component.

4. CONCLUSION

An new algorithm for FM demodulation with an sampling rate less than a conventional one, i.e. the FM bandwidth, has been proposed. Using a simplified equivalent model, we have derived a new sampling rate limit for an FM signal. It has shown that even though spectrum overlap (aliasing) happens to the sampled FM signal, a DC-free message signal can still be recovered completely. Whether and how this undersampling concept can work for other modulation schemes need further investigation.

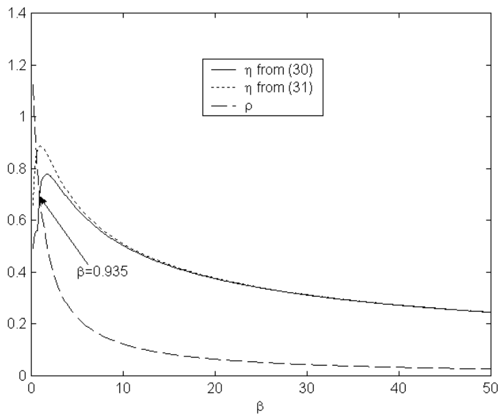

Fig. 3 The reduction ratio η and the restriction ratio ρ versus the deviation ratio β.

It has demonstrated in the paper that the current sampling theorem does not present an optimal sampling rate for recovery the message from its FM signal and further modification of the theorem is needed. The basic idea introduced in this paper may act as trigger for the further development of the sampling theorem, especially for sampling modulated signals.

ACKNOWLEDGMENT

The authors would like to thank V. Zarzoso, M. Evans and T. Cronk for their helpful comments on this paper. The financial support from ARM plc is gratefully acknowledged.

REFERENCES

[1] A. D. Poularikas and S. Seely, *Signals and Systems*, Boston, Massachusettes:PWS Publishers, 1985, pp. 216-219.

[2] E. O. Brigham, *The Fast Fourier Transform and its Applications*, Englewood cliff, New Jersey: Prentice Hall, Inc., 1988, pp. 320-337.

[3] B.P. Lathi, *Modern Digital and Analog Communication Systems*, 3rd ed., Oxford, UK, Oxford University Press, 1998, p. 219

[4] A. V. Oppenheim, A. S. Willsky and I. T. Young, *Signals and Systems*, Englewood cliff, New Jersey: Prentice Hall, Inc., 1983, pp. 531-537