International Journal of Innovative
Computing, Information and Control
Volume **9**, Number **1**, January **2013**

ICIC International ©2013 ISSN 1349-4198
pp. 269–285

AMPLITUDE-LOCKED LOOP SEPARATION SYSTEM USING MAXIMUM LIKELIHOOD ESTIMATION ANALYSIS

Gwo-Jia Jong, Chen-Shen Huang, Peng-Linag Peng
and Gwo-Jiun Horng

Department of Electronic Engineering
National Kaohsiung University of Applied Sciences
No. 415, Chien Kung Road, Sanmin District, Kaohsiung 80778, Taiwan
gjjong@cc.kuas.edu.tw

Received October 2011; revised February 2012

Abstract. *The transmission system of wireless communication has a problem with channel noise. The noise appears through many hidden sources with different noise types. In a communication system, the carrier transmission signals are distorted or disturbed by the external or internal system. The carrier signal is mixed with the interference signal in a co-channel interference (CCI) transmission system. In this paper, we adopted the amplitude-locked loop (ALL) methodology to separate and demodulate the modulating signals of a channel by operating at the same frequency for the demodulation process and algorithm. The maximum likelihood estimation (MLE) and detection schemes are computed for searching of the observation variance by the decision rule to find the optimum demodulated signals. Co-channel interference has a major impact on the communication quality factor of a transmission system. In this paper, we adopted and developed the amplitude-locked loop system for operating at the additive white Gaussian noise (AWGN) of the wireless communications environment. The maximum likelihood estimation algorithm is used for finding the maximum probability density distribution function of the statistical parameters of the communications system. The simulation results of this research demonstrate the effectiveness of the algorithm through the minimized error value and increased reliability for different carrier signal-to-noise ratios (SNRc).*
Keywords: CCI, ALL, Maximum likelihood estimation

1. **Introduction.** During the last ten years, co-channel interference (CCI) between signals that are transmitted in wireless communication systems is one of the most troublesome problems. The signals modulate to transmission in different type of channels in which the presence of noise has been separated for a variety of applications, such as adaptive finite impulse response (FIR), blind algorithm, and fuzzy adaptive Kalman filter.

It is a complex task to suppress CCI when the message is being delivered in a communication system. We focus on how to separate the mixture signal by using phase-locked loop (PLL) and amplitude-locked loop (ALL) systems. However, if the CCI is addressed through these conventional techniques, then the signal will suffer severe degradation. The envelope is no longer held constant, and the instantaneous frequency is not maintained at a fixed proportional ratio to the original signal. The PLL output will contain large in-band spikes and acquire some unintelligible turbulence [1]. Therefore, to address co-channel interference, a novel separation algorithm is proposed for canceling co-channel interference in a communication system with an ALL system. This scheme has the advantages of low-complexity computation and reduces the CCI effect.

The detection of CCI with additive white Gaussian noise (AWGN) is a very important issue in statistical signal processing. The noise interference nature of the receiver is non-stationary, so the characteristics of the phase and the envelope of AWGN are represented by time-varying functions. Hence, in Section 2, we discuss the mathematical details of the theory of the ALL separation system and the related work of the algorithm, which also explain the signals analysis under the co-channel interference that is the most troublesome problem in the recent development of communication systems. Then, in Section 3, we discuss our proposed method of maximum likelihood estimation (MLE) and cognitive radio (CR) that combine the PLL and ALL systems to improve the performance. In the simulation process, we discuss our approach for improving the performance of the separation system. The statistical estimates by the MLE model of the characteristic value from signals can be separated into the domain and sub-domain, which have a minimized probability of error and increased reliability using integrated methods.

2. System Structure.

2.1. Basic model of the separation structure.

The transmission communication system block model is shown in Figure 1. The maximum likelihood estimate is proposed for detecting the variation of the dominant and subdominant signals. Both the dominant and subdominant signals are processed through the linear mean square algorithm to reduce the error value, and we assumed that the source signals are modulated as follows:

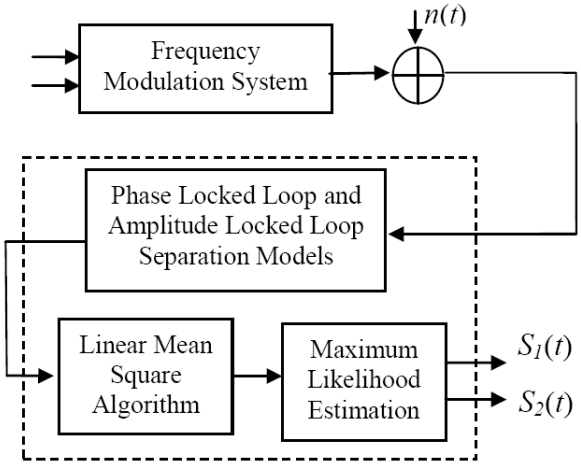


FIGURE 1. The transmission signals separation system block model

2.2. Procedure of phase-locked loop.

A block diagram of a phase-locked loop (PLL) is shown in Figure 1. Such a system generally contains four basic elements [3]: 1) a phase detector, 2) a loop filter, 3) a loop amplifier, and 4) a voltage-controlled oscillator (VCO). To understand the operation of a PLL, we assume that the input signal is given by

$$x_1(t) = A\cos[\omega_c t + \theta_1(t)] \tag{1}$$

and that the voltage-controlled oscillator (VCO) output signal is given by

$$x_2(t) = B\cos[\omega_c t + \theta_1(t)] \tag{2}$$

There are many different types of phase detectors, all with different operating properties. For our application, we assume that the phase detector is a multiplier followed by a

low pass filter (LPF) to remove the second harmonic of the carrier. We then obtain

$$\begin{aligned} V_m(t) &= A\cos[\omega_c t + \theta_1(t)] + B\sin[\omega_c t + \theta_2(t)] \\ &= \frac{AB}{2}\{\sin[\theta_1(t) - \theta_2(t)] + \sin[2\omega_c t + \theta_1(t) + \theta_2(t)]\} \end{aligned} \tag{3}$$

We also assume that an inverter is presented to remove the minus sign resulting from the multiplication. With these assumptions, the output of the phase detector becomes

$$v_p(t) = \frac{1}{2}ABk_d\sin[\theta_1(t) - \theta_2(t)] \tag{4}$$

where k_d is the constant associated with the multiplier in the phase detector. The output of the phase detector is filtered by a loop filter, amplified by a loop amplifier, and applied to the VCO. A VCO is essentially a frequency modulator, with the frequency deviation of the output, $d\theta/dt$, being proportional to the input signal. In other words, we say

$$\frac{d\theta}{dt} = k_v y(t) \tag{5}$$

which yields

$$\theta(t) = k_v \int^t y(\tau)d\tau \tag{6}$$

The parameter k_v is known as the VCO constant and is measured in radians per second per unit of input. Because the output of the phase detector is determined by the phase deviation of the PLL input and the VCO output, we could model a PLL without regard to the carrier frequency, ω_c. The model is known as the nonlinear model because of the sinusoidal nonlinearity in the phase detector. When the PLL is operating in-lock, the VCO phase, $\theta_2(t)$, is a good estimate of the input-phase deviation, $\theta_1(t)$. For this mode of operation, the phase error, $\theta_1(t) - \theta_2(t)$, is small, and

$$\sin[\theta_1(t) - \theta_2(t)] \cong \theta_1(t) - \theta_2(t) \tag{7}$$

Thus, if the phase error is small, the sinusoidal nonlinearity can be neglected and the PLL becomes a linear feedback control system, which is easily analyzed. Both of the nonlinear and linear models involve $\theta_1(t)$ and $\theta_2(t)$ rather than $x_1(t)$ and $x_2(t)$, i.e., $\theta_1(t)$ and $\theta_2(t)$, respectively, are fully determined, as observed from (7) and (8). If $\theta_1(t) \cong \theta_2(t)$, it follows that

$$\frac{d\theta_1(t)}{dt} \cong \frac{d\theta_2(t)}{dt} \tag{8}$$

$$\dot{\theta}_2(t) = K_v y(t), \quad \frac{d\theta_2(t)}{dt} = K_v y(t), \quad \frac{d\theta_1(t)}{dt} = K_v y(t),$$

$$y(t) = \frac{1}{K_v}\frac{d\theta_1(t)}{dt} \tag{9}$$

and the frequency deviation of the VCO is a good estimate of the input frequency deviation. For an FM system, the input frequency deviation is proportional to the message signal, $m(t)$. Because the VCO frequency deviation is proportional to the VCO input, $y(t)$, it follows that $y(t)$ is proportional to $m_i(t)$ for (9) to be satisfied. Thus, $y(t)$ is the demodulated output for the FM system.

The PLL can also be used as a demodulator for phase-modulated signals by integrating the VCO input. Because the VCO input signal is proportional to the frequency deviation of the PLL input, the integral of this signal is proportional to the phase deviation of the PLL input. We also have observed that operating the PLL as a discriminator requires a large loop gain amplifier and, thus, a large bandwidth. In general, amplifiers with very large values of loop gain cannot be used in practical applications without difficulty. However, the use of appropriate loop filters allows for good performance to be achieved

with reasonable values of loop gain and bandwidth. There is one final observation that we should make. The expression for loop gain shows that it is a function of the amplitude of the input signal. Therefore, a PLL must be designed for a given signal level. If that signal level changes, a new design may be necessary. In most practical applications, the dependence on the loop-gain amplifier is removed by placing a limiter on the loop input.

2.3. **Procedure of amplitude locked loop.** Due to the activity of the servo-loop, the output of the ALL can be described by the reciprocal of the additive envelope. Thus,

$$f_{ALL}(t) = \frac{1}{r(t)} = \frac{1}{\sqrt{1 + 2m\cos\omega_d t + m^2}} \tag{10}$$

As both m and the difference frequency, ω_d, are increased, the servo-loop no longer follows the input rate of change. For large values of m and difference frequency, ω_d, slew-rate limitations cause the output to asymptote to a fixed maximum value. In practice, the output has a well-defined maximum output at $m = 1.0$. To allow for this difference, (10) is modified to be:

$$f_{ALL}(t) = \frac{(1-m)(1+m)}{\sqrt{1 + 2m\cos\omega_d t + m^2}} \tag{11}$$

This function has the following limits when $m = 1$ and when the cosine function equals plus or minus one.

Case A: $\cos 0 = 1$

$$f_{ALL}(t) = \frac{(1+m)(1-m)}{\sqrt{1 + 2m + m^2}} = \frac{(1+m)(1-m)}{\sqrt{(1+m)^2}} = \frac{(1+m)(1-m)}{(1+m)} = 1 - m \tag{12}$$

$f_{ALL}(t) = (1-m) = 0$ at $m = 1$.

Case B: $\cos\pi = -1$

$$f_{ALL}(t) = \frac{(1+m)(1-m)}{\sqrt{1 - 2m + m^2}} = \frac{(1+m)(1-m)}{\sqrt{(1-m)^2}} \quad = \frac{(1+m)(1-m)}{(1-m)} = 1 + m \tag{13}$$

$f_{ALL}(t) = (1+m) = 2$ at $m = 1$.

Thus, the function of the values in (12) and (13) can only be between 0 and 2 for all values of m and $\cos\omega_d t$. The numerator of (11) now has the correct characteristics. Due to the slew-rate conditions, the reciprocal square-root term in the denominator can be replaced without significant loss of accuracy by a simpler function, represented as the following

$$f_{ALL}(t) = \frac{1 - m^2}{1 + m\cos\omega_d t} \tag{14}$$

These two approximations in the raw ALL output are in (15), and we can obtain the main dominant by (15) minus one in (16), minus two in (17).

$$f_{ALL}(t)_{raw} = ALL = \frac{1 - m^2}{1 + m\cos\omega_d t} \tag{15}$$

$$\begin{aligned} f_{ALL-1}(t) &= ALL - 1 = \frac{1 - m^2}{1 + m\cos\omega_d t} - 1 \\ &= \frac{-m^2 - m\cos\omega_d t}{1 + m\cos\omega_d t} \end{aligned} \tag{16}$$

At the singularity value for $m = 1$, these functions achieve a constant amplitude value for the minus unity value and an instantaneous pulse of zero width of two units in amplitude occurring at the instantaneous difference frequency. These functions cannot be realized easily in practice because it requires a circuit with an infinite bandwidth amongst other constraints. However, it can be realized to a proper accuracy that results in typical

improvements in noise addition (SINAD) at the final filtered baseband output. Through a further dynamic dc shift, another ALL identity is defined as

$$f_{ALL-2}(t) = ALL - 2 = \frac{1 - 2m\cos\omega_d t - m^2}{1 + m\cos\omega_d t} \tag{17}$$

In practice, the normal ALL output is maintained at a fixed mean dc level because the set point of the ALL is fixed. An error term of $+m$ appears in the ALL output because the true ALL function should decrease to zero as the m value approaches one. This error term can be removed by generating an equal and opposite term and subtracting the newly generated term from the ALL output. By ac coupling the ALL output, full-wave rectifying it, smoothing it and subtracting the result form the ALL signal, the true ALL signal is generated.

2.4. **Linear mean-square adaptive filter.** The solution of many problems of practical interest depends on the ability to accurately estimate the value $y(n)$ of a signal (desired response) with a set of values (observations or data) from another related signal or signals. Successful estimation is possible if there is significant statistical dependence or correlation between the signals involved in the particular application. In array signal processing, the data are obtained with M different sensors. The situation is simpler for filtering applications because the data are obtained by delaying a single discrete-time, where $x_k(n) = x(n + 1 - k)$, $1 \le k \le M$.

Further simplifications are possible in linear prediction, where both the desired response and data are time samples of the same signal. Because array processing problems are the most general ones, we will formulate and solve the following estimation problem: given a set of data, $x_k(n)$, for $1 \le k \le M$, determine an estimate, $\hat{y}(n)$, of the desired response, $y(n)$, using the rule

$$\hat{y}\,(n) = \mathbf{H}\,\{x_k\,(n)\,,\ 1 \le k \le M\} \tag{18}$$

which, in general, is a nonlinear function of the data.

When $x_k(n) = x(n + 1 - k)$, the estimator takes the form of a discrete-time filter that can be linear or nonlinear, time-invariant or time-varying, and with a finite- or infinite-duration impulse response. Linear filters can be implemented using any direct, parallel, cascade, or lattice-ladder structure. The difference between the estimated response, $\hat{y}(n)$, and the desired response, $y(n)$, that is

$$e\,(n) = y\,(n) - \hat{y}\,(n) \tag{19}$$

is known as the error signal. We want to find an estimator whose output approximates the desired response as closely as possible, according to a certain performance criterion. We use the term optimum estimator or optimum signal processor to refer to such an estimator. We stress that optimum is not used as a synonym for best; it simply means the best under a given set of assumptions and conditions.

3. Decision Method.

3.1. **Spectrum comparison of cognitive radio.** Cognitive radio (CR) is a term for radios that are aware of their surroundings and adapt their transmission parameters to the environment and the interference situation. Cognitive radio is an intelligent wireless communication system that is aware of its surrounding environment (i.e., the outside world). CR uses the methodology of understanding-by-building to learn from the environment and adapt its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g., transmit-power, carrier-frequency, and modulation strategy) in real-time, with two primary objectives in

mind: (a) highly reliable communications whenever and wherever needed and (b) the efficient utilization of the radio spectrum.

The fundamental principle of CR is thus to identify other radios in the environment that might use the same spectral resources and to then design a transmission strategy that minimizes interference to and from those radios. For the identification, design, implementation, and analysis of transmission strategies, it is essential to understand the propagation channel. The power emitted by a transmitter might be determined by the system designer, but it is the channel that determines how much of the power arrives as useful power at the intended receiver and how much interference is created at a victim receiver. The time variations of the channel response determine how often potential interference levels have to be estimated and, thus, how often transmission strategies may have to be adapted. As we will see, many other properties of the channel influence CR design and analysis as well. First, however, we present here some high-level information about CR bands, wireless channels, and the relationship between them.

To underline the importance of radio signal analysis for CR, we discuss modulation recognition and bit stream analysis. Modulation recognition is a method that supports the selection of the suitable demodulation process in the receiver. Although the instruction of a recognizer can be performed off-line, the recognition process itself must be performed online on the incoming signal. This process consists of three steps: 1) preprocessing to generate the digital complex baseband signal; 2) feature extraction to spot the signal statistics with respect to the modulation-specific parameters amplitude a, instantaneous frequency f, and phase θ; and 3) classification to recover the signal modulation mode.

The classifier is taught to allocate the received signal to one of the uniquely defined modulation classes using the information contained in the feature vectors of a representative signed learning set of signals. Modulation recognition enables the receiver to correctly demodulate the signal. The redundancies between information and parity check bits produced by channel coding and the protocol structure of the signal radiated by a transmitter are contained in the bit stream generated by the demodulator. Bit stream analysis recovers these structures with the help of a flexible correlation that can be adapted to the incoming signal [13].

3.1.1. *Primary signal detection.* Whether for the detection of temporal or spatial spectral holes, the spectral sensing approach for CR involves deciding on whether the primary signal is present from the observed signals. It can be formulated as the following two hypotheses:

$$y(t) = \begin{cases} i(t) + w(t), & H_0 \\ s(t) + i(t) + w(t), & H_1 \end{cases} \tag{20}$$

where $y(t)$ is the received signal at the CR user, $s(t)$ is the primary signal, $i(t)$ is interference and $w(t)$ is the additive white Gaussian noise (AWGN). The difference between interference and noise is that interference is undesired man-made colored signal, while noise is white and statistically Gaussian. When interference from various sources in the environment is approximately Gaussian and white, it is regarded as noise. In (20), H_0 and H_1 denote the hypotheses corresponding to the absence and presence of the primary signal, respectively. Thus, from the observation, $y(t)$, the CR user needs to decide between H_0 and H_1.

For different licensed bands, primary signals have different characteristics. The 802.22 wireless regional-area network (WRAN) is developed to work in licensed TV bands; therefore, the primary signal is the Advanced Television Systems Committee (ATSC) digital TV signal or the wireless microphone signal. For CR networks to utilize the temporally

idle spectral bands allocated to the 3G cellular mobile communication system, the primary signal may be a direct spread code-division multiple access (DS-CDMA) signal for its long-term evaluation (LTE) version. We do not restrict the primary signal to any waveform. Instead, we exploit the characteristics of the primary signal that are generally known to the public for spectral sensing. Various spectral sensing techniques have been proposed to utilize the characteristics or the a priori knowledge of the primary signal.

3.1.2. *Energy detection.* Energy detection is the simplest spectral sensing technique, which is shown in Figure 2. An energy detector (ED) simply treats the primary signal as noise and decides on the presence or absence of the primary signal based on the energy of the observed signal. Because it does not need any a priori knowledge of the primary signal, the ED is robust to the variation of the primary signal. Moreover, the ED does not involve complicated signal processing and has low complexity.

In practice, energy detection is especially suitable for wide-band spectral sensing. In this case, the simultaneous sensing of a number of sub-bands can be realized by simply scanning the power spectral density (PSD) of the received wide-band signal. In practice, it is advisable to complete wide-band spectral sensing via two stages. In the first stage, low-complexity energy detection is applied to search for possible idle sub-bands; in the second stage, more advanced spectral sensing techniques with a higher detection sensitivity and therefore higher complexity, such as cycle stationary detection, are applied to the sub-band candidates to determine whether they are actually available for secondary usage.

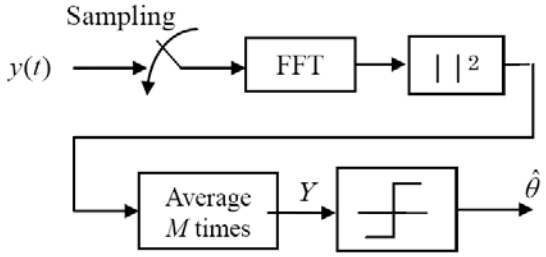


FIGURE 2. Block diagram of the use of an energy detector over a spectral sub-band

3.1.3. *Performance analysis.* As indicated in Figure 2, the spectral component on each spectral sub-band of interest is obtained from the fast Fourier transform (FFT) of the sampled received-signal. Then, the test statistics of the ED is obtained as the observed energy summation within M consecutive segments. To facilitate the analysis, here, we ignore the interference component in the received signal.

$$Y = \begin{cases} \sum_{m=1}^{M} |W(m)|^2, & H_0 \\ \sum_{m=1}^{M} |s(m) + W(m)|^2, & H_1 \end{cases} \tag{21}$$

where $S(m)$ and $W(m)$ denote the spectral components of the received primary-signal and the white noise on the sub-band of interest in the mth segment, respectively. The decision of the ED regarding the sub-band of interest is given by

$$\hat{\theta} = \begin{cases} H_0, & \text{if } Y > \lambda \\ H_1, & \text{if } Y < \lambda \end{cases} \tag{22}$$

where the threshold, λ, is chosen to satisfy a target false-alarm probability. The threshold defines an operating point on the receiver operating curve over two performance parameters: the false-alarm probability and the missed-detection probability. Here, the false-alarm probability is customarily chosen as the primary operating specification.

Without loss of generality, we assume that the noise, $W(m)$, is white complex Gaussian with zero mean and a variance of two. We define the instantaneous SNR of the received primary-signal within the current M segments as

$$Y = \frac{1}{2M} \sum\nolimits_{m=1}^{M} |S(m)|^2 \tag{23}$$

Then, the test statistics of the ED Y follows a central chi-square distribution with $2M$ degrees of freedom under H_0, and a non-central chi-square distribution with $2M$ degrees of freedom and a non-centrality parameter $\mu = \sum_{m=1}^{M} |S(m)|^2 = 2MY$ under H_1, i.e.,

$$f_Y(Y) \sim \begin{cases} X_{2M}^2, & H_0 \\ X_{2M}^2(\mu), & H_1 \end{cases} \tag{24}$$

where $f_Y(Y)$ denotes the probability density function (PDF) of Y and X_{2M}^2 and $X_{2M}^2(\mu)$ denote a central and non-central chi-square distribution, respectively. Thus, the false-alarm probability, $P_F = P(Y > \lambda \,|H_0)$, can be expressed as

$$P_F = \frac{\Gamma(M, \lambda/2)}{\Gamma(M)} \tag{25}$$

where $\Gamma(.)$ and $\Gamma(.,.)$ denote the gamma function and the upper incomplete gamma function, respectively. Given the target false-alarm probability, the threshold, λ, can be uniquely determined based in (25). Once λ is determined, the detection probability, $P_D = P(Y > \lambda \,|H_1)$, can be obtained by

$$P_D = \int_0^{+\infty} P(Y > \lambda \,|H_{1,\mu}) f_\mu(\mu) d\mu = \int_0^{+\infty} Q_M(\sqrt{\mu}\sqrt{\lambda}) f_\mu(\mu) d\mu \tag{26}$$

where $Q_M(.,.)$ is the generalized Marcum Q-function and $f_\mu(u)$ is the PDF of μ. A closed-form expression of the detection probability of the ED over the general Nakagami fading channel has been derived.

3.2. **Independent components analysis of CCI.** To rigorously define ICA [4,5], we can use a statistical "latent variables" model. Assume that we observe n linear mixtures, $x_1, \ldots, x_n$, of n independent components

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \ldots + a_{jn}s_n, \text{ for all } j \tag{27}$$

We have now dropped the time index t; in the ICA model, we assume that each mixture, x_j, and each independent component is a random variable instead of a proper time signal. The observed values, $x_j(t)$, e.g., the microphone signals in the cocktail party problem, are then a sample of this random variable. Without loss of generality, we can assume that both the mixture variables and the independent components have a zero mean. If this assumption is not true, then the observable variables, x_i, can always be centered by subtracting the sample mean, which makes the model have a zero-mean.

It is convenient to use vector-matrix notation instead of the sums, as in the previous equation. Let us denote the random vector whose elements are the mixtures by $\mathbf{x}$ and, likewise, the random vector with elements by $\mathbf{s}$. Let us denote the matrix with elements a_{ij} by $\mathbf{A}$. Generally, bold lower case letters indicate vectors and bold upper-case letters denote matrices. All vectors are understood as column vectors; thus, or the transpose of $\mathbf{x}$ is a row vector. Using this vector-matrix notation, the above mixing model is written as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{28}$$

Sometimes, we need the columns of matrix $\mathbf{A}$; denoting them by a_j, the model can also be written as

$$\mathbf{x} = \sum_{i=1}^{n} \mathbf{a}_i s_i \tag{29}$$

The statistical model in (28) is called independent component analysis, or the ICA model. The ICA model is a generative model, which means that it describes how the observed data are generated by a process of mixing the components, s_i. The independent components are latent variables, meaning that they cannot be directly observed. Additionally, the mixing matrix is assumed to be unknown. All we observe is the random vector, $\mathbf{x}$, and we must estimate both $\mathbf{A}$ and $\mathbf{s}$ using it. These estimations must be achieved under as general a set of assumptions as possible.

The starting point for ICA is the very simple assumption that the components, s_i, are statistically independent. It will be shown below that we must also assume that the independent component must have non-Gaussian distributions. However, in the basic model, we do not assume that these distributions are known (if they are known, the problem is considerably simplified). For simplicity, we also assume that the unknown mixing matrix is square, but this assumption can be sometimes relaxed. Then, after estimating the matrix $\mathbf{A}$, we can compute its inverse, say $\mathbf{W}$, and obtain the independent component simply by:

$$\mathbf{s} = \mathbf{W}\mathbf{x} \tag{30}$$

ICA is very closely related to the method called blind-source separation (BSS) or blind-signal separation. A "source" here means an original signal, i.e., an independent component, like the speaker in the cocktail party problem. "Blind" means that we know very little, if anything, about the mixing matrix and make few assumptions on the source signals. ICA is one method, perhaps the most widely used, for performing blind-source separation.

In many applications, it would be more realistic to assume that there is some noise in the measurements [6,7], which would mean adding a noise term in the model. For simplicity, we omit any noise terms because the estimation of the noise-free model is difficult enough by itself, and it seems to be sufficient for many applications [8].

3.2.1. *Ambiguities of the ICA model.* In the ICA model in (28), it is easy to see that the following ambiguities will hold:

A. We cannot determine the variances (energies) of the independent components.

The reason for this ambiguity is that with both $\mathbf{s}$ and $\mathbf{A}$ being unknown, any scalar multiplier in one of the sources, s_i, could always be cancelled by dividing the corresponding column $\mathbf{a}_i$ of $\mathbf{A}$ by the same scalar (see (29)). As a consequence, we may fix the magnitudes of the independent components; as they are random variables, the most natural way to do this is to assume that each has unit variance. Then, the matrix $\mathbf{A}$ will be adapted in the ICA solution methods to take into account this restriction. Note that the sign is still ambiguous; therefore, we could multiply an independent component by (-1) without affecting the model. This ambiguity is, fortunately, insignificant in most applications.

B. We cannot determine the order of the independent components.

The reason for this ambiguity is that, again, with both $\mathbf{s}$ and $\mathbf{A}$ being unknown, we can freely change the order of the terms in the sum in (30) and call any of the independent components the first one. Formally, a permutation matrix, $\mathbf{P}$, and its inverse can be substituted in the model to yield $\mathbf{x} = \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{s}$. The elements of $\mathbf{P}\mathbf{s}$ are the original independent variables s_j, but in another order. The matrix is just a new unknown mixing matrix to be solved by the ICA algorithms.

3.2.2. *Non-Gaussian is independent.* Intuitively speaking, the key to estimating the ICA model is the non-Gaussian nature of the signals. Actually, without non-Gaussian character, the estimation is not possible at all and, at the same time, is probably the main reason for the rather late resurgence of ICA research: in most classical statistical theory, the random variables are assumed to have Gaussian distributions, thus precluding any methods related to the ICA.

The Central Limit Theorem, a classical result in probability theory, states that the distribution of a sum of independent random variables tends toward a Gaussian distribution, under certain conditions. Thus, a sum of two independent random variables usually has a distribution that is closer to Gaussian than any of the two original random variables. Let us now assume that the data vector, $\mathbf{x}$, is distributed according to the ICA data model in (28), i.e., it is a mixture of independent components. For simplicity, let us assume in this section that all the independent components have identical distributions.

To estimate one of the independent components, we consider a linear combination of the x_i. Let us denote this combination by $y = \mathbf{w}^T\mathbf{x} = \sum_i w_i x_i$, where $\mathbf{w}$ is a vector to be determined. If $\mathbf{w}$ were one of the rows of the inverse of $\mathbf{A}$, then this linear combination would actually equal one of the independent components. The question is now how could we use the Central Limit Theorem to determine $\mathbf{w}$ so that it would equal one of the rows of the inverse of $\mathbf{A}$? In practice, we cannot determine such a $\mathbf{w}$ exactly because we have no knowledge of matrix $\mathbf{A}$, but we can find an estimator that provides a good approximation.

To see how this leads to the basic principle of ICA estimation, let us make a change of variables by defining $\mathbf{z} = \mathbf{A}^T\mathbf{w}$. Then, we have $y = \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{A}\mathbf{s} = \mathbf{z}^T\mathbf{s}$. The y is a linear combination of s_i, with weights given by z_i. Because a sum of even two independent random variables is more Gaussian than the original variables, $\mathbf{z}^T\mathbf{s}$ is more Gaussian than any of the s_i and becomes least Gaussian when it, in fact, equals one of the s_i. In this case, obviously only one of the elements, z_i, of $\mathbf{z}$ is nonzero. Therefore, we could take as $\mathbf{w}$ a vector that *maximizes the non-Gaussian character* of $\mathbf{w}^T\mathbf{x}$. Such a vector would necessarily correspond to a $\mathbf{z}$ which has only one nonzero component. Thus, $\mathbf{w}^T\mathbf{x} = \mathbf{z}^T\mathbf{s}$ equals one of the independent components.

Maximizing the non-Gaussian character of $\mathbf{w}^T\mathbf{x}$ gives us one of the independent components. In fact, the optimization landscape for the non-Gaussian character in the n-dimensional space of vectors $\mathbf{w}$ has $2n$ local maxima, two for each independent component, corresponding to s_i and $-s_i$. To find several independent components, we need to find all these local maxima. This is not difficult because the different independent components are uncorrelated; we can always constrain the search to the space that provides estimates uncorrelated with the previous ones. This space corresponds to orthogonalization in a suitably transformed (i.e., whitened) space.

3.2.3. *Negentropy.* A second very important measure of the non-Gaussian character is given by negentropy. Negentropy is based on the information theoretic quantity of (differential) entropy.

Entropy is the basic concept of information theory. The entropy of a random variable can be interpreted as the degree of information that the observation of the variable provides. The more "random", i.e., unpredictable and unstructured, the variable is, the larger its entropy. More rigorously, entropy is closely related to the coding length of the random variable. In fact, under some simplifying assumptions, entropy is the coding length of the random variable. For introductions on information theory, see, e.g., [9,10]. Entropy, H, is defined for a discrete random variable, Y, as

$$H(Y) = -\sum_i P(Y = a_i) \log P(Y = a_i) \tag{31}$$

where the a_i are the possible values of Y. This very well-known definition can be generalized for continuous-valued random variables and vectors, in which case, it is often called differential entropy. The differential entropy, H, of a random vector, $\mathbf{y}$, with density, $f(\mathbf{y})$, is defined as

$$H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y})\, d\mathbf{y} \tag{32}$$

A fundamental result of information theory is that a Gaussian variable has the largest entropy among all the random variables of equal variance, with a proof described in [11,12]. Therefore, entropy could be used as a measure of the non-Gaussian character. In fact, this shows that the Gaussian distribution is the "most random" or the least structured of all distributions. Entropy is small for distributions that are clearly concentrated on certain values, i.e., when the variable is clearly clustered, or has a PDF that is very "spiky".

To obtain a measure of the non-Gaussian character that is zero for a Gaussian variable and always nonnegative, a slightly modified version of the definition of differential entropy, called negentropy, is often used. Negentropy J is defined as follows:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \tag{33}$$

where $\mathbf{y}_{gauss}$ is a Gaussian random variable of the same covariance matrix as $\mathbf{y}$. Due to the above-mentioned properties, negentropy is always non-negative, and it is zero if and only if $\mathbf{y}$ has a Gaussian distribution. Negentropy has the additional interesting property that it is invariant for invertible linear transformations.

The advantage of using negentropy, or, equivalently, differential entropy, as a measure of the non-Gaussian character is that it is well justified by statistical theory. In fact, negentropy is in some sense the optimal estimator of the non-Gaussian character, as far as statistical properties are concerned. The problem with using negentropy is, however, that it is computationally very difficult. Estimating negentropy using the definition would require an estimate (possibly nonparametric) of the PDF. Therefore, simpler approximations of negentropy are very useful, as will be discussed next.

3.3. **Maximum likelihood estimate of CCI.** As mentioned in the previous function, the procedure commonly used to estimate nonrandom parameters is the maximum likelihood estimation. Let $Y_1, Y_2, \ldots, Y_k$ be K observations of the random variable, $\mathbf{Y}$, with sample values, $y_1, y_2, \ldots, y_k$. These random variables are independent and identically distributed. Let $f_{\mathbf{Y}|\mathbf{\Theta}}(y\,|\theta)$ denote the conditional density function of the random variable, $\mathbf{Y}$. Note that the density function of $\mathbf{Y}$ depends on the parameter, θ, $\theta \in \Theta$, which needs to be estimated. The likelihood function, $L(\theta)$, is

$$L(\theta) = f_{Y_1,\ldots,Y_k|\Theta}(y_1, y_2, \ldots, y_k\,|\theta) = f_{\mathbf{Y}|\Theta}(y\,|\theta) = \prod_{k=1}^{K} f_{\mathbf{Y}_k|\Theta}(y_k\,|\theta) \tag{34}$$

The value of $\hat{\theta}$ that maximizes the likelihood function is called the maximum likelihood estimator of θ. To maximize the likelihood function, standard techniques of calculus may be used. Because the logarithmic function, $\ln x$, is a monotonically increasing function of x, it can be used for maximizing $L(\theta)$. Hence, it can be shown that a necessary but not sufficient condition to obtain the ML estimate $\hat{\theta}$ is to solve the likelihood equation invariance property.

$$\frac{\partial}{\partial\theta} \ln f_{\mathbf{Y}|\mathbf{\Theta}}(y\,|\theta) = 0 \tag{35}$$

Let $L(\theta)$ be the likelihood function of θ and $g(\hat{\theta})$ be a one-to-one function of θ; that is, if $g(\theta_1) = g(\theta_2) \Leftrightarrow \theta_1 = \theta_2$. If it is an MLE of θ, then $g(\hat{\theta})$ is an MLE of $g(\theta)$ [11].

Assume that X is truly a Gaussian random vector whose elements are independent Gaussian random variables identically distributed, with each mean μ and variance σ^2. Given an N-vector of data, x (that is, the realization of X), estimate μ using the MLE method. Additionally, find the bias and variance of this estimator.

First, let us see what the Cramer-Rao inequality reveals about this example. The PDF of X is known to have the form

$$f_{\mathbf{x}}\left(\mathbf{x}; \left[\mu \;\; \sigma^2\right]\right) = \frac{1}{\left\{\sigma^2 (2\pi)\right\}^{N/2}} \exp^{\frac{-1}{2}\sum_{\ell=1}^{N}\left\{\frac{x_\ell - \mu}{\sigma}\right\}^2} \tag{36}$$

and the log-likelihood function is

$$L = \ln\left(f_{\mathbf{x}}\left(\mathbf{x}; \left[\mu \;\; \sigma^2\right]\right)\right) = \left(\frac{-N}{2}\right)\left(\ln\left\{\sigma^2\right\} + \ln\left\{2\pi\right\}\right) - \frac{1}{2}\sum_{\ell=1}^{N}\frac{(x_\ell - \mu)^2}{\sigma^2} \tag{37}$$

The first derivative of L with respect to μ is

$$\frac{\partial L}{\partial \mu} = \sum_{\ell=1}^{N}\left\{\frac{x_\ell - \mu}{\sigma^2}\right\}, \tag{38}$$

and the second derivative is

$$\frac{\partial^2 L}{\partial \mu^2} = -\frac{N}{\sigma^2} \tag{39}$$

Thus, the Cramer-Rao inequality becomes

$$\sigma_{\hat{\mu}}^2 \geq \frac{\left\{1 + \frac{\partial B_\mu(\hat{\mu})}{\partial u}\right\}^2 \sigma^2}{N} \tag{40}$$

The maximum likelihood estimate of μ is obtained by setting the partial derivative of the log-likelihood with respect to μ equal to zero:

$$\frac{\partial L}{\partial \mu}\Big|_{\mu=\hat{\mu}_{MLE}} = \sum_{\ell=1}^{N}\left\{\frac{x_\ell - \mu}{\sigma^2}\right\}\Big|_{\mu=\hat{\mu}_{MLE}} = 0 \tag{41}$$

which gives

$$\hat{\mu}_{MLE} = \frac{1}{N}\sum_{\ell=1}^{N} x_\ell \tag{42}$$

Thus, the sample mean, which is the MLE estimator, is an unbiased estimator of the true mean: $B_\mu(\hat{\mu}_{MLE}) = 0$. Consequently, the Cramer-Rao bound from the above calculation reduces to the lower bound:

$$\begin{aligned} \sigma_{\hat{\mu}_{MLE}}^2 &= E\left\{(\hat{\mu}_{MLE} - \mu)^2\right\} \\ &= \frac{1}{N^2} E\left\{\sum_{\ell=1}^{N}(X_\ell - \mu)\sum_{m=1}^{N}(X_m - \mu)\right\} \end{aligned} \tag{43}$$

$$\begin{aligned} &= \frac{1}{N^2}\sum_{\ell=1}^{N}\sum_{m=1}^{N} E\left\{(X_\ell - \mu)(X_m - \mu)\right\} \\ &= \frac{\sigma^2}{N} \end{aligned} \tag{44}$$

The variance of $\hat{\mu}_{MLE}$ is thus equal to the lower bound of the Cramer-Rao inequality as determined above. Therefore, for this case, the MLE estimator is efficient for any N; in general, this will not be true for other MLE estimation problems. Moreover, because

the sum of two or more independent Gaussian random variable is itself Gaussian, then from our above result, $\hat{\mu}_{MLE} = \eta\left(\mu, \sigma^2/N\right)$ [12].

4. **Simulation Performance.** To simplify the co-channel separation system problem, we used the phase-lock loop and amplitude-locked loop modules in Section 2. In this chapter, we simulated signals of interference to the carrier ratio at CCI environment under the AWGN channel in an amplitude-locked loop system. To analyze the original signal, we presented the maximum likelihood estimation and cognitive radio method to demonstrate the separation system, which together with the optimum decision methods can be more precise and reliable. In addition, the proposed ALL algorithm can be adopted to cancel the Rician spikes.

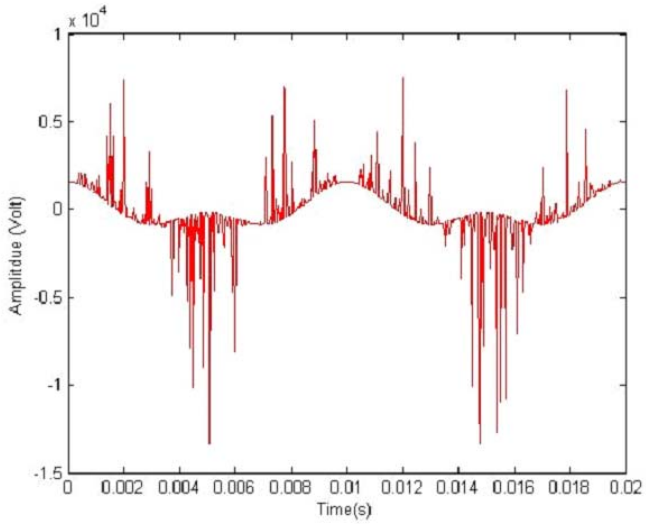


FIGURE 3. The output signals of $f_{PLL}(t)$

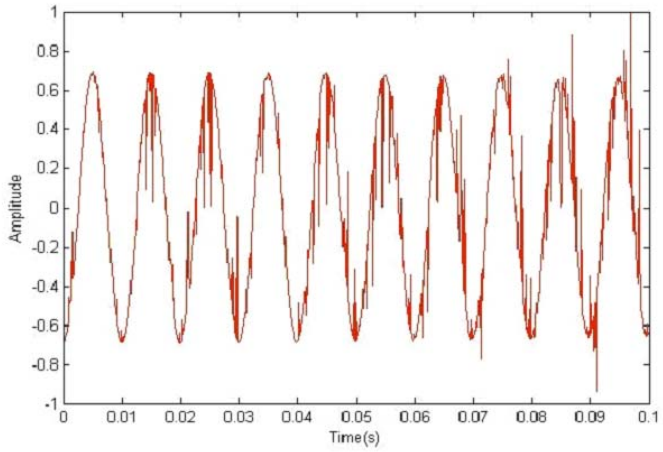


FIGURE 4. The PLL output is multiplied by a dc shift of the ALL function $f_{PLL}(t) \times f_{ALL}(t)$

4.1. **Joint estimation and detection.** We consider the separation operation for co-channel FM signals under AWGN channel. The performances are compared with various separation methods such as the traditional ALL system and the rotation algorithm. We assume that the input origin signals $m_1(t) = \cos 2\pi f_1 t$ and $m_2 = \cos 2\pi f_2 t$ are continuous cosine signals.

For the AWGN channel, the performance of the signal separation will be affected. The adaptive filter is provided for tracking the time-varying character of the interference signal and comparing the estimated signal with the desired signal. The PLL output is multiplied by a dc shift of the ALL function. The dominant signal is derived as follows:

$$f_{PLL}(t)\times f_{ALL-2}(t) = -\frac{1+\alpha\ (t)\cos\omega_n t}{1+2\alpha(t)\cos\omega_n t+\alpha^2(t)}S_1+\frac{1+\alpha(t)\cos\omega_n t}{1+2\alpha(t)\cos\omega_n t+\alpha^2(t)}f_{ALL-1}(t)S_2 \tag{45}$$

The subdominant signal is derived as follows:

$$\begin{aligned} f_{PLL}(t)\times f_{-(ALL-1)}(t) = &-\frac{1+\alpha(t)\cos\omega_n t}{1+2\alpha(t)\cos\omega_n t+\alpha^2(t)}\frac{f_{ALL-1}(t)}{f_{ALL-2}(t)}S_1 \\ &+\frac{1+\alpha(t)\cos\omega_n t}{1+2\alpha(t)\cos\omega_n t+\alpha^2(t)}\times\frac{f_{-(ALL-1)}(t)f_{ALL-1}(t)}{f_{ALL-2}(t)}S_2 \end{aligned} \tag{46}$$

From a statistical standpoint, the data vector, $y = (y_1, \ldots, y_m)$, is a random sample from an unknown population. The goal of the data analysis is to identify the population that is most likely to have generated the sample. In statistics, each population is identified by a corresponding probability distribution. Associated with each probability distribution is a unique value of the model's parameter. As the parameter changes in value, different probability distributions are generated. Formally, a model is defined as the family of probability distributions indexed by the model's parameters.

Let $f\,(y|w)$ denote the probability density function (PDF) that specifies the probability of observing data vector, y, given the parameter, w, in (47). Throughout this paper, we will use a plain letter for a vector (e.g., y) and a letter with a subscript for a vector element (e.g., y_i). The parameter $w = (w_1, \ldots, w_k)$ is a vector defined on a multi-dimensional parameter space. If individual observations, y_i, are statistically independent of one another, then according to the theory of probability, the PDF for the data, $y = (y_1, \ldots, y_m)$, given the parameter vector w can be expressed as a multiplication of PDFs for individual observations, $Y_1, Y_2, Y_3, \ldots, Y_n$, which have a joint density denoted by

$$f\,(y_1, y_2, \ldots, y_n) = f(y_1, y_2, \ldots, y_n|w) \tag{47}$$

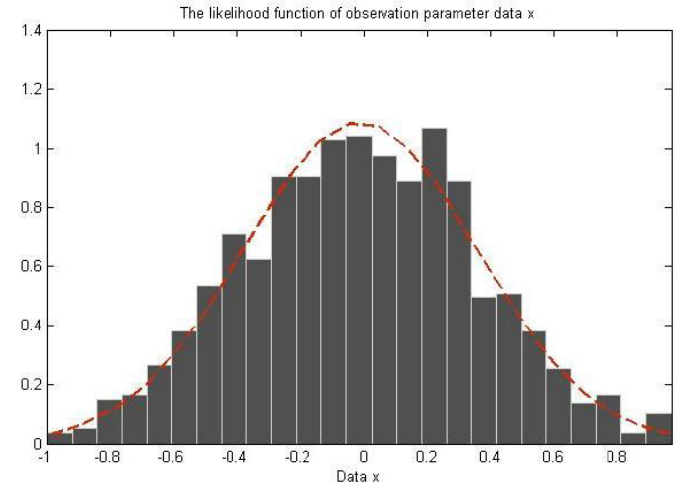
Given the observed values, $Y_1 = y_1$, $Y_2 = y_2$, $\ldots$, $Y_n = y_n$, the likelihood that w is the function

$$like(w) = f(y_1, y_2, \ldots, y_n|w) \tag{48}$$

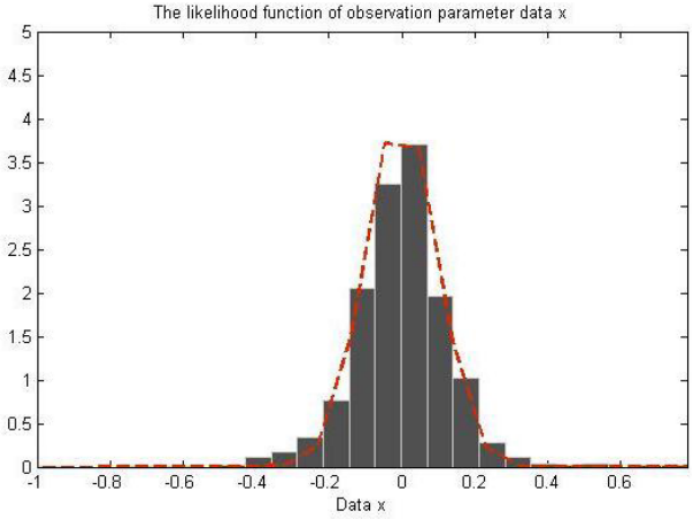
where $like(w)$ is the probability of observing the given data as a function of w. Considered as function of w. If the distribution is discrete, it will be the frequency distribution function.

Definition 4.1. *The maximum likelihood estimate of w is that value of w that maximizes $like(w)$; it is the value that makes the observed data the "most probable". If Y_i are independent and identically distributed (IID) random variables, then the likelihood simplifies to*

$$like(w) = \prod_{i=1}^{n} f\,(x_i|w) \tag{49}$$



(a)



(b)

FIGURE 5. (a) The probability density function of observing data vector, y, in a domain with 25 bins by histogram; (b) the likelihood function given the observed data, x, of a domain at $n = 1000$ samples

Rather than maximizing this product, which can be quite tedious, we often use the fact that the logarithm is an increasing function, so it will be equivalent to maximizing the log likelihood $like(w) = \sum_{i=1}^{n} (f(x_i|w))$ as shown in Figure 5.

4.2. **Convergence and mean square error analysis.** The system is presented that, in a simulation experiment, results in the signal being separated successfully by the PLL with the ALL algorithm for FM signals. In the AWGN channel, we propose the LMS algorithm for accelerating the convergence rate and obtaining the MMSE value. From the computation simulation result of the dominant and subdominant signals, we obtain the results shown in Figure 6. The dominant and subdominant signals that are separated are shown in Table 1. The lower SNR_c value is between −20 dB and 0 dB. Better SNRs of 2.79 dB and 2.8 dB can be found for different values of 0.9 and 0.1, respectively. By the same method, the subdominant separation signal has SNR values between −20 dB and 0 dB. The MMSE can be determined in these figures.

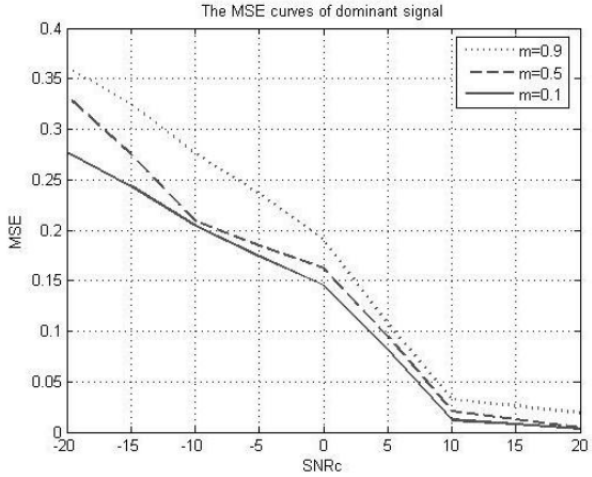


FIGURE 6. The MSE and SNR_c for the dominant signal

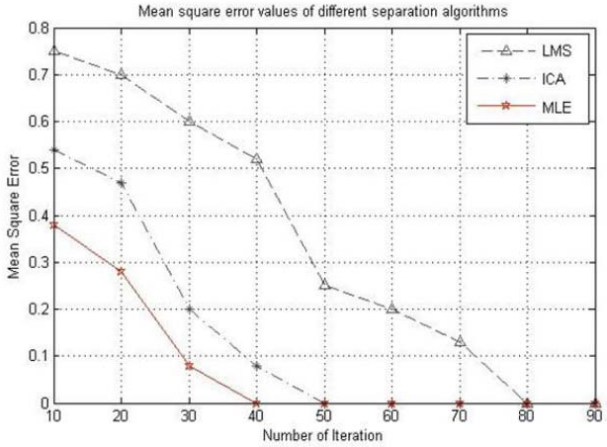


FIGURE 7. Iteration values of the MSE using different algorithms (MLE, ICA and LMS)

TABLE 1. Comparison of the convergence number of the different algorithms

Number / Algorithm	10	20	30	40	50	60	70	80
Linear Mean Square	0.75	0.70	0.60	0.52	0.25	0.20	0.13	0.00
Independent Component Analysis	0.54	0.42	0.20	0.08	0.00	0.00	0.00	0.00
Maximum likelihood	0.49	0.28	0.08	0.0	0.00	0.00	0.00	0.00

Table 1 indicates the mean-square error values and convergence ratio using different separation algorithms. In the LMS algorithm, the iteration times required are approximately 80 for achieving system stability.

Hence, we determine that the convergence rate of the proposed algorithm is more than 1.25 times faster than that of the ICA method and more than 2 times faster than that of the LMS algorithm for the 1000 sample data shown in Figure 7.

5. **Conclusions.** We provided higher efficiency to separate the wireless communications transmissions by introducing a proposed approach on the maximum likelihood (ML) detection. The dominant and subdominant signals are successfully separated with the proposed algorithm in the CCI signal with AWGN interference. The proposed algorithm can be adopted to solve the co-channel transmission problems, and it suppressed the noise disturbance transmission channel.

In this paper, our simulation results are effective in detecting signals by statistical parameters and show that the convergence iteration numbers of the MLE algorithms numbers are 1.25 times faster than the ICA algorithm and 2 times faster than the LMS algorithm. Additionally, our method does not use the complex computing of the spectral analysis in the CR algorithm. Therefore, we improved the signal detection of the computing observation value, which can obtain a lower error ratio at an optimum receiver. In future work, our contributions can help to develop a separation system of complex transmission signals in a fading interference channel environment.

REFERENCES

[1] A. M. Pettigrew, The amplitude-locked loop, the theory and operation of the FM 201/5 high performance FM demodulator, *Ampsys Company Document*, 1996.

[2] Y.-C. Chen, *The Fuzzy-Adaptive Kalman Filter Analysis for the Amplitude-Locked Loop System of Co-channel Interference*, Master Thesis, National Kaohsiung University of Applied Sciences, Taiwan, 2007.

[3] R. E. Best, *Phase-Locked Loops: Design, Simulation and Application*, McGraw-Hill Company, USA, 1997.

[4] C. Jutten and J. Herault, Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture, *Signal Processing*, vol.24, pp.1-10, 1991.

[5] P. Comon, Independent component analysis: A new concept? *Signal Processing*, vol.36, no.3, pp.287-314, 1994.

[6] A. Hyvärinen, Independent component analysis in the presence of gaussian noise by maximizing joint likelihood, *Neurocomputing*, vol.22, pp.49-67, 1998.

[7] A. Hyvärinen, Gaussian moments for noisy independent component analysis, *IEEE Signal Processing Letters*, vol.6, no.6, pp.145-147, 1999.

[8] J.-C. You, *The Blind Algorithm for the Amplitude-Locked Loop System of Co-channel Interference*, Master Thesis, National Kaohsiung University of Applied Sciences, Taiwan, 2009.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons Inc., 1991.

[10] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1991.

[11] M. Barkat, *Signal Detection and Estimation*, 2nd Edition, Artech House Publisher, London, 2005.

[12] T. J. Cavicchi, *Digital Signal Processing*, 1st Edition, John Wiley & Sons, USA, 1999.

[13] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, vol.39, pp.1-38, 1977.

[14] I. J. Myung, Tutorial on maximum likelihood estimation, *Journal of Mathematical Psychology*, vol.47, pp.90-100, 2003.

[15] P.-L. Peng and G.-J. Jong, Amplitude-locked loop separation and encryption system analysis combined cognitive radio, *Proc. of the 16th Asia-Pacific Conference on Communications*, pp.435-440, 2010.

[16] I. J. Myung, Tutorial on maximum likelihood estimation, *Journal of Mathematical Psychology*, vol.47, pp.90-100, 2003.