HOS Based Distinctive Features for Preliminary Signal Classification

Maciej Pędzisz and Ali Mansour

École Nationale Supérieure des Ingénieurs
des Études et Techniques d'Armement (ENSIETA),
Laboratoire "Extraction et Exploitation de l'Information
en Environnements Incertains" (E^3I^2), Brest, France,
pedzisma@ensieta.fr, mansour@ensieta.fr

Abstract. We consider the problem of preliminary classification of digitally modulated signals. The goal is to simplify further signal analysis (synchronization, signal separation, modulation identification and parameters estimation) by making initial separation among the most known classes of signals. Proposed methodology is mainly based on Higher Order Statistics (HOS) of the distributions of instantaneous amplitude and frequency. The experimental results emphasize the performance of the proposed set of features.

1 Introduction

In Communication Intelligence (COMINT), knowledge of signal's frequency structure is essential to recognize underlying modulation type and measure its parameters. Up to now, all frequency synchronization algorithms consider only one signal, they need a big number of symbols and a long time to converge. Thus, making a preliminary signal classification based on frequency invariant features, will much simplify further processing, allowing applications of signal-specific synchronization, source separation and modulation classification techniques.

In [1], authors presented empirical results in Blind Source Separation (BSS) using overcomplete Independent Component Analysis (ICA) representations. They demonstrated fidelity of their algorithm in the case of 2 mixtures of 3 speech signals. Separation of 2 audio sources from a single sensor is the subject covered in [2]. Proposed method generalizes the Wiener filtering with Gaussian Mixture distributions and Hidden Markov Models. A time-frequency filtering based on the Pseudo Wigner-Ville distribution is considered in [3]. Performance of the presented algorithm was validated using a mixture of 2 voice recordings. In [4], sparse factorization approach with K-means clustering algorithm applied to BSS problem is discussed. Provided results reveal the performance of the algorithm in case of 10 face images (6 mixtures), as well as 8 speech signals (5 mixtures). Authors of [5], derive algebraic means for ICA in the case of undetermined mixtures. Their results are based on the structure of the fourth-order cumulant tensor. Sixth-order statistics and the virtual array concept are addressed in [6]. It was shown that their algorithm can be used to increase the

C.G. Puntonet and A. Perieto (Eds.): ICA 2004, LNCS 3195, pp. 1158-1164, 2004

effective aperture of an antenna array, and so to identify the mixture of more sources than sensors. The case of binary source separation is covered in [7] and [8]. Their algorithm uses the structure of the probability distributions of the observed data. Simulations showed that the method can successfully separate at least up to 10 binary sources at different noise levels.

On the other hand, modulation recognition algorithms ([9], [10], [11], [12]) deal with the cases where some a priori information is available (carrier frequency, symbol timing, ...) and there is only one signal in additive noise. In this contribution, we try to fill the gap between synchronization & modulation recognition methods, and source (signal) separation algorithms based on one observation (undetermined problem). Using the proposed set of features, we are able to distinguish among the most common known signal types, and so, choose the appropriate methodology for further signal processing.

2 Signal Models

2.1 Mono-Component Signal

Let's assume working in the conditions where signal's carrier frequency is not known. The received complex baseband signal (after imperfect demodulation) can be expressed as a sum of two uncorrelated components:

$$s(t) = A_{\mathrm{c}}(t)\mathrm{e}^{j(\omega_{\mathrm{r}}t+\Theta_{\mathrm{r}})} + n(t) \tag{1}$$

where $A_{\mathrm{c}}(t)$ is a signal complex envelope, ω_{r} is a residual frequency, Θ_{r} is a phase of the residual frequency, and $n(t)$ corresponds to a zero-mean, additive white gaussian complex noise.

Using the concept of the complex envelope, we can express any linearly modulated signal as:

$$A_{\mathrm{c}}(t) = A\sum_{k} d_k h(t-kT-\tau), \qquad k \in \{1, 2, \ldots, K\} \tag{2}$$

where A is a constant amplitude, d_k describe signal constellation, $h(t)$ is a pulse shaping function, T is a symbol duration, τ is an out-of-synchronization error (due to imperfect demodulation), and K is a number of available symbols. For the most known M-ary linear modulations (MASK – M-ary Amplitude Shift Keying, MQAM – M-ary Quadrature Amplitude Modulation, MPSK – M-ary Phase Shift Keying), we have:

$$d_k^{\mathrm{MASK}} = a_k, \qquad a_k \in \{\pm(2m-1) :\ m = 1, 2, \ldots, M/2\} \tag{3}$$

$$d_k^{\mathrm{MQAM}} = a_k + jb_k, \quad a_k, b_k \in \{\pm(2m-1) :\ m = 1, 2, \ldots, \log_2(M) - 2\} \tag{4}$$

$$d_k^{\mathrm{MPSK}} = \mathrm{e}^{j\varphi_k}, \qquad \varphi_k \in \{\tfrac{2\pi}{M}(m-1) :\ m = 1, 2, \ldots, M\}\ . \tag{5}$$

In the nonlinear case (MFSK – M-ary Frequency Shift Keying), we can write:

$$A_{\mathrm{c}}(t) = A\mathrm{e}^{j\sum_k d_k \Delta_\omega (t-kT-\tau)h(t-kT-\tau)}, \qquad k \in \{1, 2, \ldots, K\} \tag{6}$$

where Δ_ω is a frequency deviation, and d_k can be expressed as:

$$d_k^{\mathrm{MFSK}} \in \{\pm(2m-1):\ m = 1, 2, \ldots, M/2\}\ . \tag{7}$$

It is assumed that variables a_k, b_k and φ_k in equations (3), (4) and (5), as well as d_k in (7) are independent and identically distributed (i.i.d. processes). It is assumed also that all modulation states are equiprobable (which is always accomplished when source coding is applied) and the pulse shaping function $h(t)$ is rectangular.

2.2 Multi-Component Signal

Taking into consideration the mono-component model of a linear modulation ((1) and (2)), we can write a general formula for a multi-component signal as:

$$\begin{aligned} S(t) &= \sum_{i=1}^{L} A_{\mathrm{c}_i}(t)\mathrm{e}^{j(\omega_{\mathrm{r}_i} t + \Theta_{\mathrm{r}_i})} + n_i(t) \\ &= \sum_{i=1}^{L} A_i \sum_k d_{k_i} h_i(t - kT_i - \tau_i)\mathrm{e}^{j(\omega_{\mathrm{r}_i} t + \Theta_{\mathrm{r}_i})} + n(t) \end{aligned} \tag{8}$$

where L is a number of mono-component signals and $n(t)$ is a term which absorbed all noise contributions $n_i(t)$.

In COMINT applications, it is often sufficient to consider: there are two signals in the mixture ($L = 2$), and applied modulation types are MPSK. Additionally, we assume that signal amplitudes are identical ($A_1 = A_2$)[1]. We are not considering prior knowledge about:

- residual frequencies and phases (ω_{r_i} and Θ_{r_i});
- symbol durations (T_i);
- synchronization errors (τ_i).

3 Distinctive Features

3.1 Preliminary Results

Based on the signal models (1) and (8), we can rewrite the received signal as:

$$s_{\mathrm{r}}(t) = p(t) + jq(t) = A_{\mathrm{i}}(t)\mathrm{e}^{j\phi_{\mathrm{i}}(t)} \tag{9}$$

where $p(t)$ and $q(t)$ are in-phase and quadrature components, $A_{\mathrm{i}}(t)$ is an instantaneous amplitude and $\phi_{\mathrm{i}}(t)$ is an instantaneous phase. Then, we can define:

$$A_{\mathrm{i}}(t) = |s_{\mathrm{r}}(t)|, \quad \phi_{\mathrm{i}}(t) = \arg\{s_{\mathrm{r}}(t)\}, \quad \omega_{\mathrm{i}}(t) = \frac{d\phi_{\mathrm{i}}(t)}{dt} = \frac{p(t)\frac{dq(t)}{dt} - q(t)\frac{dp(t)}{dt}}{p^2(t)+q^2(t)} \tag{10}$$

[1] General case $A_1 \neq A_2$ will be addressed elsewhere.

where $\omega_{\rm i}(t)$ is an instantaneous frequency. In general, $\omega_{\rm i}(t)$ is defined using the concept of the analytic signal [13].

It is well known that the probability density function (PDF) of $A_{\rm i}$ of any MPSK/MFSK signal can be expressed in terms of its constant amplitude A (Eq. (2)) and noise variance σ_n^2 (Eq. (1)) by means of Rice distribution [14]:

$$f_{A_{\rm i}}(A_{\rm i};\, A, \sigma_n^2) = \frac{A_{\rm i}}{\sigma_n^2} {\rm e}^{-\frac{A_{\rm i}^2 + A^2}{2\sigma_n^2}} {\rm I}_0\left(\frac{A_{\rm i} A}{\sigma_n^2}\right), \quad A_i \geqslant 0 \tag{11}$$

where ${\rm I}_0(x)$ is the modified Bessel function of order 0.

If $A = 0$ (NOISE), then PDF of $A_{\rm i}$ becomes Rayleigh. For the MQAM class of signals, we can write the corresponding PDF as a mean of $f_{A_{\rm i}}(A_{\rm i};\, A_l, \sigma_n^2)$ over all distinctive amplitudes A_l.

The second distribution which can be considered as distinctive in signal classification is the PDF of $\omega_{\rm i}$ [15]. For a single-carrier modulation (MPSK, MQAM), we have:

$$f_{\omega_{\rm i}}(\omega_{\rm i};\, A, \sigma_n^2) = \vartheta^{-1} v_{\rm i}^{-\frac{3}{2}} {\rm e}^{-\frac{A^2}{2\sigma_n^2}} {}_1F_1\left(\frac{3}{2},\, 1;\, \frac{A^2}{2\sigma_n^2 v_{\rm i}}\right) \tag{12}$$

where $v_{\rm i} = 1 + \omega_{\rm i}^2/\vartheta^2$, $\vartheta^2 = \int_{-\infty}^{+\infty} \omega_{\rm i}^2 \gamma(\omega)\, d\omega / \int_{-\infty}^{+\infty} \gamma(\omega)\, d\omega$, $\gamma(\omega)$ is a power spectral density (PSD) of noise, and ${}_1F_1(\alpha,\, \beta;\, x)$ is a confluent hypergeometric function defined as:

$${}_1F_1(\alpha,\, \beta;\, x) = \sum_{k=0}^{+\infty} \frac{\Gamma(\alpha+k)\Gamma(\beta)x^k}{\Gamma(\alpha)\Gamma(\beta+k)k\,!}, \qquad \beta \neq 0,\, -1,\, -2,\, \ldots \tag{13}$$

It is obvious that in the multi-carrier case (MFSK), the PDF of $\omega_{\rm i}$ can be expressed as a mean over all distinctive (carrier) frequencies.

Finally, when $A \gg \sigma_n$, we can approximate both distributions by the Gaussians [13], [15], [16]:

$$f_{A_{\rm i}}(A_{\rm i};\, A, \sigma_n^2) \approx \mathcal{N}(A_{\rm i};\, A, \sigma_n^2), \qquad f_{\omega_{\rm i}}(\omega_{\rm i};\, A, \sigma_n^2) \approx \mathcal{N}(\omega_{\rm i};\, 0, \tfrac{B\sigma_n}{2\sqrt{3}A}) \tag{14}$$

where $\mathcal{N}(x;\, \mu, \sigma^2) \triangleq \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$, and B is a noise effective bandwidth.

3.2 Features Extraction

The main objective in preliminary signal classification is to find a set of characteristics which allows distinction among different classes of signals. Based on distributions of $A_{\rm i}$ and $\omega_{\rm i}$, we can extract normalized cumulants [17] of order 3 γ_3 (skewness) and 4 γ_4 (kurtosis) as:

$$\gamma_3 = \frac{\kappa_3}{\kappa_2^{3/2}}, \qquad \gamma_4 = \frac{\kappa_4}{\kappa_2^2} \tag{15}$$

where cumulants κ_r and corresponding moments m_r are defined by:

$$\kappa_2 = m_2 - m_1^2 \tag{16}$$
$$\kappa_3 = m_3 - 3m_2m_1 + 2m_1^3 \tag{17}$$
$$\kappa_4 = m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4 \tag{18}$$
$$m_r = \int_{-\infty}^{+\infty} x^r f(x)\, dx\ . \tag{19}$$

Other sets of characteristics can be obtained by using Renyi's quadratic entropy [18]:

$$H_2 = -\log\left[\int_{-\infty}^{+\infty} f^2(x)\, dx\right] \tag{20}$$

and by solving a polynomial regression on the logarithm of a PDF:

$$\log(f(x)) \approx \sum_k a_k x^k. \tag{21}$$

3.3 Features Selection & Dimensionality Reduction

It is obvious that limiting the number of features will make learning and testing faster and demanding less memory. Aside from this, feature space of a lower dimension may enable more accurate classifiers for a finite learning set.

Based on the characteristics presented in the previous section, experiments have been conducted to choose the most discriminative set of features:

- features based on A_i: γ_3^A, γ_4^A, H_2^A, a_3^A, a_2^A, a_1^A, a_0^A (3-rd degree polynomial is sufficient to describe asymmetry and flatness of considered distributions);
- features based on ω_i: γ_4^ω, H_2^ω, a_4^ω, a_2^ω, a_0^ω (PDF of ω_i is symmetrical about the mean, so all the features based on asymmetry were eliminated).

Once they have been selected, one can apply the Linear Discriminant Analysis to verify the importance of chosen features. Using the Fisher's criterion [19]:

$$J_F = \mathrm{tr}\{\mathbf{T}\} = \mathrm{tr}\{\mathbf{S}_w^{-1}\mathbf{S}_b\} \tag{22}$$

where $\mathbf{S}_w$ is the within-class covariance matrix (the sum of covariance matrices computed for each class separately), and $\mathbf{S}_b$ is the between-class covariance matrix (the covariance matrix of class means), we found:

- all selected features are of equal importance – among differents combinations of features, the whole set is the most discriminative;
- features from A_i are best to separate between classes of signals with symmetric A_i PDF (MPSK, MFSK) and asymmetric (NOISE, MQAM and MIXTURE);
- features from ω_i are best to separate between classes of signals with unimodal ω_i PDF (MPSK, MQAM and MIXTURE) and multimodal (MFSK).

It should be noted, that using eigenvectors of matrix $\mathbf{T}$, it is possible to reduce dimensionality of the feature vector

$$\mathbf{x} = [\gamma_3^A, \gamma_4^A, H_2^A, a_3^A, a_2^A, a_1^A, a_0^A, \gamma_4^\omega, H_2^\omega, a_4^\omega, a_2^\omega, a_0^\omega]^T \tag{23}$$

by means of linear transformation:

$$\mathbf{y} = \mathbf{W}\mathbf{x} \tag{24}$$

where eigenvectors corresponding to largest eigenvalues of $\mathbf{T}$ form the rows of the transformation matrix $\mathbf{W}$.

4 Simulations

To evaluate the performance of the proposed set of features, extensive simulations were conducted on the signals: NOISE, MPSK (2, 4 and 8), MFSK (2 and 4), MQAM (16 and 32) and MIXTURE (2xBPSK, 2xQPSK and BPSK & QPSK). All signals were composed of 512 samples, 5 samples per symbol, 1000 different realizations. Signal to Noise Ratio (SNR) was varying from 0 dB up to 30 dB. The residual frequencies ω_{r_i}, the corresponding phases Θ_{r_i}, as well as the symbol timings T_i, were chosen randomly according to Nyquist sampling theorem. Corresponding results (SNR = 5 dB) are shown in Fig. 1.

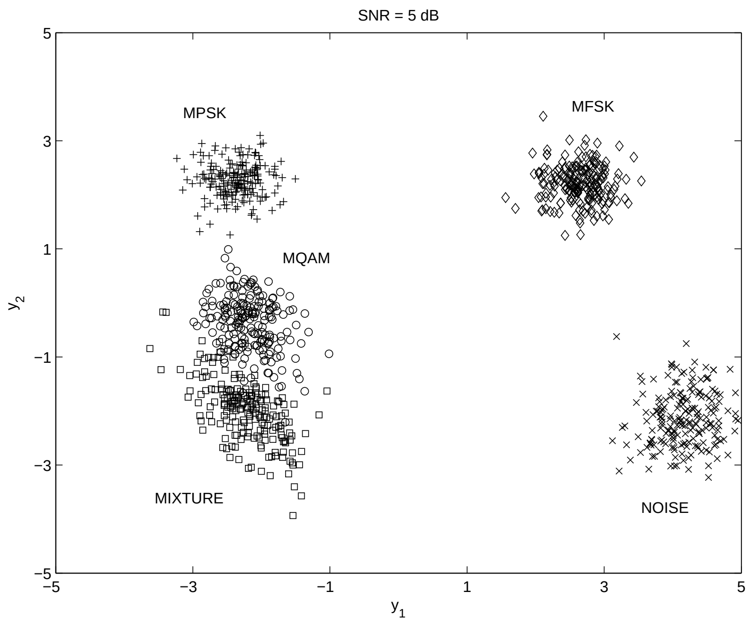

Fig. 1. Signals in a 2D space after dimensionality reduction (SNR = 5 dB).

5 Conclusion

It is evident that selected set of features is very efficient even for low SNR. Perfect classification can be obtained for the classes NOISE, MPSK and MFSK for SNR $>$ 5 dB, however distinction between MQAM and MIXTURE is far from being "sufficient enough".

Although classification in a 2D space was used for visualization purposes, one should not limit himself during constructing a final classifier. Adding another set of characteristics (based for example on Time-Frequency Distributions (TFD)), may be more attractive in more than 2 dimensions. Also, making classifier hierarchical or using some nonlinear mappings (MMI [20], NPCA [21]), may increase separability of the classes. These topics will be covered in a future work.

References

1. Lee T.-W., Lewicki M.S., Girolami M., Sejnowski T.J.: *Blind Source Separation of More Sources Than Mixtures Using Overcomplete Representations.* IEEE Sig. Proc. Let., Vol. 6, No. 4, April 1999.
2. Benaroya L., Bimbot F.: *Wiener Based Source Separation with HMM/GMM Using a Single Sensor.* Journée AS Séparation de Sources et GdR ISIS, Paris, 12 June 2003.
3. Mansour A., Kawamoto M., Puntonet C.: *A Time-Frequency Approach to Blind Separation of Under-Determined Mixture of Sources.* Proc. of the IASTED International Conference on Applied Simulation and Modelling, Marbella, Spain, 3-5 September 2003.
4. Li Y., Cichocki A., Amari S.: *Sparse Component Analysis for Blind Source Separation with Less Sensors than Sources.* ICA 2003, Nara, Japan, April 2003.
5. De Lathauwer L., De Moor B., Vandewalle J., Cardoso J.-F.: *Independent Component Analysis of Largery Undetermined Mixtures.* ICA 2003, Nara, Japan, April 2003.
6. Albera L., Ferréol A., Comon P., Chevalier P.: *Sixth Order Blind Identification of Undetermined Mixtures (BIRTH) of Sources.* ICA 2003, Nara, Japan, April 2003.
7. Diamantaras K.I.: *Blind Separation of Multiply Binary Sources using a Single Linear Mixture.* ICASSP 2000, Istanbul, Turkey, June 2000.
8. Diamantaras K.I., Chassioti E.: *Blind Separation of n Binary Sources from one Observation: A Deterministic Approach.* ICA 2000, Helsinki, Finland, 19-22 June 2000.
9. Azzouz E.E., Nandi A.K.: *Automatic Modulation Recognition of Communication Signals.* Kluwer Academic Publishers, 1996.
10. Soliman S.S., Hsue S.-Z.: *Signal Classification Using Statistical Moments.* IEEE Transactions on Communications, Vol. 40, No. 5, May 1992.
11. Hero A.O., Hadinejad-Mehram H.: *Digital Modulation Classification Using Power Moment Matrices.* ICASSP 1998, Seattle, USA, May 1998.
12. Hipp J.E.: *Modulation Classification Based on Statistical Moments.* MILCOM 1986, Monterey, USA, October 1986.
13. Proakis J.G.: *Digital Communications.* McGraw-Hill, 4-th edition, 2001.
14. Haykin S.: *Communication Systems.* John Wiley & Sons, Inc., 3-rd edition, 1994.
15. Levin B.R.: *Theoric Bases of Statistical Technics in Radio.* Radio and Communications (in russian), 3-rd edition, 1989.

16. Blachman N.M.: *Gaussian Noise–Part II: Distribution of Phase Change of Narrow-Band Noise Plus Sinusoid.* IEEE Trans. on Inf. Theory, Vol. 34, No. 6, November 1988.
17. Kendall M.G., Stuart A.: *The Advanced Theory of Statistics.* Charles Griffin, 1958.
18. Haykin S.: *Unsupervised Adaptive Filtering.* John Wiley & Sons, Inc., 2000.
19. Fukunaga K.: *Introduction to Statistical Pattern Recognition.* Academic Press, 2-nd edition, 1990.
20. Torkkola K.: *Feature Extraction by Non-Parametric Mutual Information Maximization.* Journal of Machine Learning Research 3, p. 1415-1438, 2003.
21. Chalmond B., Girard S.: *Nonlinear Data Representation for Visual Learning.* Rapport de Recherche, No 3550, November 1998.