High Order Statistic Estimators For Speech Processing

Arnaud MARTIN, Ali MANSOUR

ENSIETA E3I2 EA 3876, 2 rue François Verny, 29806 BREST Cedex 9 FRANCE.
Phone : (+33) 2 98 34 88 84, Fax : (+33) 2 98 34 87 50
E-mail: Arnaud.Martin@ensieta.fr, mansour@ieee.org.

Abstract: *Speech is the easiest mean of communication for human. The reason why the vocal command along with speech recognitions are more and more used in many applications. Especially in automotive industries, speech recognition algorithms are used in order to help the driver and to reduce his tasks. Recently, high order statistics (HOS) have been used in many signal processing techniques in order to estimate signal characteristics. By using data features in real time applications, one can derive new HOS estimators. In this paper, different estimators of high order moment and cumulant using different kind of signals are proposed and discussed.*

Keywords: Blind source separation, speech recognition, estimation, high order statistics, moments, cumulants, speech processing, temporal and non-stationary data.

1. SUMMARY

Speech is the best, easiest and oldest mean of communication for human. Therefore, it seems to be very naturally to introduce speech-control systems in the next generation of cars, trains, or other transportation modes. In automobile industries, speech recognition is used in order to help the driver and makes easier the transportation. A good speech recognition system should contain many speech processing tools, such as blind source separation, speech analysis, speech recognition and speech/non speech detection [1, 2], see Fig. 1. In normal situation, a car driver isn't the only speaking person in that car, that is a major problem for speech recognition systems. Acoustic signals of different sources (including various noise signals) are mixed with each other into the microphones. In such case, speech recognition systems are unable to recognize correctly the pronounced sentences. That main reason to use source separation algorithms in our system [2]. On the other hand, speech analysis is important in order to reduce data dimensions correctly to outperform the speed processing without decreasing the speech recognition performance. In a car environment, there are many energetic noises. The speech/non-speech detection allows us to only select speech signals.
The different steps of our system needs a well estimation of speech features.

Some specific features of data and problem in speech processing field require the study of different HOS estimators. Indeed, data can be temporary, stationary or not. Data can be considered as non-stationary within large estimation window and stationary within few milliseconds [3]. However, background noise is often considered as stationary. In speech/non-speech detection or recognition applications, the processing should be done in real or almost real time. Therefore, quick and efficient estimations of the signal statistics among other parameters are mostly needed in various speech processing such as classification, detection, recognition, and sources separation, etc.

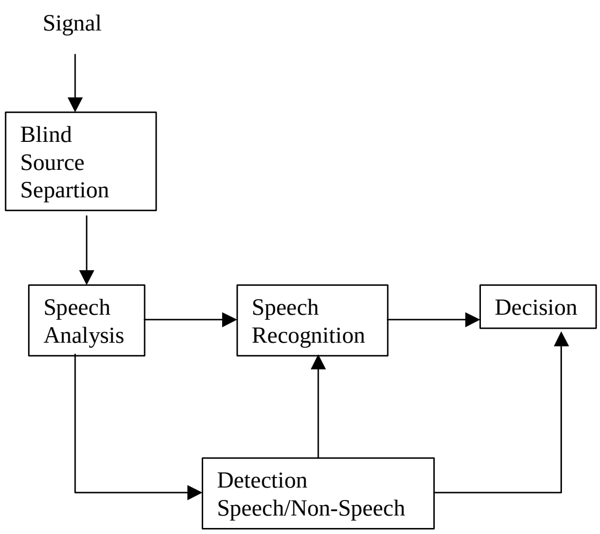

Figure 1: Speech processing example

In many speech processing applications, researchers as well as engineers assume that signal distributions are Gaussian or Laplacian in order to simplify the calculus [3]. Indeed, this assumption means that a signal distribution can be completely characterized by its mean and its standard deviation. However, this strong assumption can not be satisfied is various recent applications. Since the last two decades, other statistical information have been introduced as asymmetric and flatness estimators (given for example by the skewness and kurtosiss) or more generally High Order Statistics [4,5,6,7,8]. Thus many estimators have been proposed [3,9,10,11]. All these estimators concern the autocumulant. We mentioned before that in automotive environment there are many kind of undesirable signals (noise, speech of persons other than the car driver, radio music, so on). This reason why a blind source separation is needed (see Figure 1). In blind separation vocabularies, the car environment can be considered as a convolutive mixture model. Many blind separation algorithms are based on cross fourth order cumulants. In this paper, estimators of the three fourth order cross-cumulant are proposed.

2. THEORETICAL BACKGROUND

Let X denotes a real stochastic process stands for a speech signal, its characteristic function is given by:

$$\phi_X(t) = \int_{-\infty}^{+\infty} \exp(itx)\, p_X(x)dx \quad (1)$$

We should mention that its module is less or equal to 1, and that $\phi_X(0) = 1$. Using the previous equation, one can define the second characteristic function as:

$$\varphi_X(t) = \ln(\phi_X(t)) . \quad (2)$$

By definition, the *qth* order moment is given [7,8] from the *qth* order derivative of the first characteristic function around zero:

$$\mu_q = (-1)^q . \left. \frac{d^q \phi_X(t)}{dt^q} \right|_{t=0} = E\left[X^q \right]. \quad (3)$$

By similar definition, the *qth* order cumulant is given as the *qth* order derivative of the second characteristics function at the origin:

$$\kappa_q = (-1)^q . \left. \frac{d^q \varphi_X(t)}{dt^q} \right|_{t=0} = Cum\left[X, X, ..., X \right] \quad (4)$$

In the case of Gaussian distribution, we should mention that all cumulants with order higher than 2 are null.

Leonov and Shiryayev gave general relationships among moments and cumulants. According to their study, a *qth* order cumulant can be evaluated as:

$$Cum_q(X) = Cum\left[X, X, ..., X \right] = \sum_{p=1}^{q} (-1)^p (p-1)!\, \mu_{v_1} . \mu_{v_2} ... \mu_{v_p} \quad (5)$$

where the numbers $\left\{ v_1, v_2, ..., v_p : 1 \le p \le q \right\}$ are such as $v_i \in \left\{ 1, ..., q-p+1 \right\}$ and $\sum_{i=1}^{p} v_i = q$. We should mention here that in their original study, they developed the relationships in the case of q random variables [8]. Equation (5) can be easily obtained from the original relationship. Using equation (5), one can write the fourth order cumulant as in [9]:

$$Cum_4(X) = \mu_4 - 4\mu_1.\mu_3 - 3\mu_2^2 + 12\mu_1^2\mu_2 - 6\mu_1^4 \quad (6)$$

Last equation can be simplified for a zero mean signal:

$$Cum_4(X) = \mu_4 - 3\mu_2^2 . \quad (7)$$

The three fourth order cross cumulants for two zero mean signals X and Y are given by:

$$Cum_{3,1}(X, X, X, Y) = \mu_{3,1} - 3\mu_{2,0}\mu_{1,1}$$

$$Cum_{2,2}(X, X, Y, Y) = \mu_{2,2} - \mu_{2,0}\mu_{0,2} - 2\mu_{1,1}^2$$

$$Cum_{1,3}(X, Y, Y, Y) = \mu_{1,3} - 3\mu_{0,2}\mu_{1,1}$$

where $\mu_{n,m} = E\left[X^n Y^m \right]$. For non zero mean signals the formulas are more complicated. In the case of a speech recognition, we can assume that the signal is a zero mean signal.

3. HIGH ORDER STATISTICS ESTIMATORS

Let us consider N realizations x_i of a stochastic process X assumed to be an ergodic one. In this case, the arithmetic estimator of the *qth* order moment is given by:

$$\hat{\mu}_q = \frac{1}{N} \sum_{i=1}^{N} x_i^q . \quad (8)$$

This estimator means that the signal X is stationary over N samples. This estimator is a non biased and consistent estimator. Hence an arithmetic estimator of the fourth order cumulant for a zero mean signal can be developed form equation (7):

$$\widehat{Cum}_4(X) = \hat{\mu}_4 - 3.\hat{\mu}_2^2 . \quad (9)$$

Unfortunately, this estimator is biased, and we have proposed different non-biased estimator in [12]. One of these is given by:

$$\widehat{Cum}_4(X) = \frac{N+2}{N-1}\hat{\mu}_4 - 3\frac{N}{N-1}\hat{\mu}_2^2 . \quad (10)$$

Concerning the fourth order cross cumulants, when the two signals X and Y are independent and they are two independent and identically distributed (i.i.d.) signals, then the formula (10) can be used. In general case, a non-biased estimator of $Cum_{2,2}(X,Y)$ is given by:

$$\widehat{Cum}_{2,2}(X,Y) = a\hat{\mu}_{2,2} - b\hat{\mu}_{2,0}\hat{\mu}_{0,2} - 2c\hat{\mu}_{1,1} , \quad (11)$$

where:

$$\hat{\mu}_{n,m} = \frac{1}{N}\sum_{i=1}^{N} x_i^n y_i^m \ , \qquad (12)$$

and a, b and c are given by:

$$E\left[\widehat{Cum}_{2,2}(X,Y)\right] = aE\left[X^2Y^2\right] - \frac{b}{N}\left(E\left[X^2Y^2\right]+(N-1)E\left[X^2\right]E\left[Y^2\right]\right) - 2\frac{c}{N}\left(E\left[X^2Y^2\right]+(N-1)E\left[XY\right]^2\right) = \mu_{2,2} - \mu_{2,0}\mu_{0,2} - 2\mu_{1,1}^2$$

So we find: $a = \frac{N+2}{N-1}$ and $b = c = \frac{N}{N-1}$. By the same way we can show that:

$$\widehat{Cum}_{3,1}(X,Y) = \frac{N+2}{N(N-1)}\hat{\mu}_{3,1} - \frac{3}{N(N-1)}\hat{\mu}_{2,0}\hat{\mu}_{1,1}\ ,(13)$$

$$\widehat{Cum}_{1,3}(X,Y) = \frac{N+2}{N(N-1)}\hat{\mu}_{1,3} - \frac{3}{N(N-1)}\hat{\mu}_{0,2}\hat{\mu}_{1,1}\ .(14)$$

For a real time application the three estimators (10), (13), and (14) should be adaptive. Hence the three fourth order cross cumulant can be estimated by for every k>1 frame:

$$\widehat{Cum}_{3,1}(X,Y)(k) = \frac{k-2}{k}\widehat{Cum}_{3,1}(X,Y)(k-1) + \frac{1}{k}\hat{\mu}_{3,1}(k-1) - \frac{k+2}{k(k-1)}x_k^3 y_k - 3x_k y_k \hat{\mu}_{2,0}(k-1) - 3x_k^2\hat{\mu}_{1,1}(k-1) \qquad (15)$$

$$\widehat{Cum}_{2,2}(X,Y)(k) = \frac{k(k-1)}{k^2}\widehat{Cum}_{2,2}(X,Y)(k-1) + \frac{k-1}{k^2}\hat{\mu}_{2,2}(k-1) - \frac{2(k-1)}{k^2}\hat{\mu}_{1,1}^2(k-1) - \frac{2}{k}x_k^2 y_k^2\hat{\mu}_{1,1}(k-1) - \frac{k-1}{k^2}\hat{\mu}_{0,2}(k-1)\hat{\mu}_{2,0}(k-1) - \frac{2}{k}(x_k^2\hat{\mu}_{0,2}(k-1) + y_k^2\hat{\mu}_{2,0}(k-1)) + \frac{1}{k}x_k^2 y_k^2 \qquad (16)$$

$$\widehat{Cum}_{1,3}(X,Y)(k) = \frac{k-2}{k}\widehat{Cum}_{1,3}(X,Y)(k-1) + \frac{1}{k}\hat{\mu}_{1,3}(k-1) - \frac{k+2}{k(k-1)}x_k y_k^3 - 3x_k y_k \hat{\mu}_{0,2}(k-1) - 3y_k^2\hat{\mu}_{1,1}(k-1) \qquad (17)$$

with $\hat{\mu}_{n,m}(k) = \frac{1}{k}\sum_{i=1}^{k} x_i^n y_i^m$.

For all previous estimators the signals X and Y are supposed stationary because of the $\hat{\mu}_{n,m}$ estimator. The different high order moments can be estimated by the following non-biased and consistent estimator:

$$\hat{\mu}_{n,m}(k) = \frac{1}{1-\lambda^k}\left(\begin{array}{l}\lambda\left(1-\lambda^{k-1}\right)\hat{\mu}_{n,m}(k-1)\\ +\left(1-\lambda\right)x_k^n y_k^m\end{array}\right), \qquad (18)$$

where λ is a forgotten factor such as $0 < \lambda < 1$.

To outperform the high order statistic estimation of strong non-stationary signals, we propose a new estimator for the three fourth order cross cumulant, in order This new estimator for the first cross cumulant is given by:

$$\widehat{Cum}_{3,1}(X,Y)(k) = \frac{k-2}{k}\lambda\widehat{Cum}_{3,1}(X,Y)(k-1) + \frac{1}{k}\lambda\hat{\mu}_{3,1}(k-1) - \frac{k+2}{k(k-1)}\lambda x_k^3 y_k - 3\lambda x_k y_k \hat{\mu}_{2,0}(k-1) - 3\lambda x_k^2\hat{\mu}_{1,1}(k-1) + (1-\lambda)x_k^3 y_k - 3(1-\lambda)x_k y_k \hat{\mu}_{2,0}^2(k-1) \qquad (19)$$

4. COMPARATIVE STUDY

In order to compare the different estimators (13), (15) with the moment estimators given by (18), and (19) of the fourth order cross cumulant $Cum_{3,1}$, some experimental results are presented hereinafter. We have generated a signal $S(n)$ on 20000 realizations of non-stationary signal that contains four parts:

- Uniform zero-mean signal between –1 and 1 (on 8000 realizations)
- Gaussian zero mean signal with a standard deviation of 1 (on 5000 realizations)
- Uniform zero mean signal between –2 and 2 (on 3000 realizations)
- Gaussian zero mean signal with $\sigma = \sqrt{2}$ (on 4000 realizations).

Let us consider two signals $X(n)=S(n)$ and $Y(n)=S^3(n)$, in this case we can generate two signals $X(n)$ and $Y(n)$ such that x_i and y_i are independent and identically distributed, but x_i depends on y_i. In this case we can easy determine the theoretical fourth cross cumulant. For an uniform zero mean signal between –a and a, we have:

$$Cum_{3,1}(X,Y) = -\frac{2}{35}a^6\ , \qquad (20)$$

and for a Gaussian zero mean signal with a standard deviation σ:

$$Cum_{3,1}(X,Y) = 6\sigma^6 . \qquad (21)$$

On figure 2, we have represented the estimation of the three estimators (13), (15) with the moment estimators given by (18), and (19) of the fourth order cross cumulant on the signal $S(n)$. Note that (13) allows a good estimation only for stationary signal. The estimator (15) gives better estimation, but has a high variation on Gaussian signal. The new estimator (19) has less variation on Gaussian signal parts. But this estimator converge slower than (15) (see for example the difference betwee