An efficient VAD based on a Generalized Gaussian PDF

O. Pernía, J.M. Górriz, J. Ramírez and C.G. Puntonet and I. Turias

Department of Signal Theory
University of Granada, Granada, Spain
gorriz@ugr.es

Abstract

The emerging applications of wireless speech communication are demanding increasing levels of performance in noise adverse environments together with the design of high response rate speech processing systems. This is a serious obstacle to meet the demands of modern applications and therefore these systems often needs a noise reduction algorithm working in combination with a precise voice activity detector (VAD). This paper presents a new voice activity detector (VAD) for improving speech detection robustness in noisy environments and the performance of speech recognition systems. The algorithm defines an optimum likelihood ratio test (LRT) involving Multiple and correlated Observations (MCO). An analysis of the methodology for $N = \{2, 3\}$ shows the robustness of the proposed approach by means of a clear reduction of the classification error as the number of observations is increased. The algorithm is also compared to different VAD methods including the G.729, AMR and AFE standards, as well as recently reported algorithms showing a sustained advantage in speech/non-speech detection accuracy and speech recognition performance.

1. Introduction

The emerging applications of speech communication are demanding increasing levels of performance in noise adverse environments. Examples of such systems are the new voice services including discontinuous speech transmission [1, 2, 3] or distributed speech recognition (DSR) over wireless and IP networks [4]. These systems often require a noise reduction scheme working in combination with a precise voice activity detector (VAD) [5] for estimating the noise spectrum during non-speech periods in order to compensate its harmful effect on the speech signal.

During the last decade numerous researchers have studied different strategies for detecting speech in noise and the influence of the VAD on the performance of speech processing systems [5]. Sohn *et al*. [6] proposed a robust VAD algorithm based on a statistical likelihood ratio test (LRT) involving a single observation vector. Later, Cho *et al* [7] suggested an improvement based on a smoothed LRT. Most VADs in use today normally consider hangover algorithms based on empirical models to smooth the VAD decision. It has been shown recently [8, 9] that incorporating long-term speech information to the decision rule reports benefits for speech/pause discrimination in high noise environments, however an important assumption made on these previous works has to be revised: *the independence of overlapped observations*. In this work we propose a more realistic one: *the observations are jointly gaussian distributed with non-zero correlations*. In addition, important issues that need to be addressed are: *i*) the increased computational complexity mainly due to the definition of the decision rule over large data sets, and *ii*) the optimum criterion of the decision rule. This work advances in the field by defining a decision rule based on an optimum statistical LRT which involves multiple and *correlated* observations. The paper is organized as follows. Section 2 reviews the theoretical background on the LRT statistical decision theory. Section 4 considers its application to the problem of detecting speech in a noisy signal. Finally in Section 4.1 we discuss the suitability of the proposed approach for pair-wise correlated observations using the experimental data set AURORA 3 subset of the original Spanish SpeechDat-Car (SDC) database [10] and state some conclusions in section 6.

2. Multiple Observation Probability Ratio Test

Under a two hypothesis test, the optimal decision rule that minimizes the error probability is the Bayes classifier. Given an observation vector $\hat{\mathbf{y}}$ to be classified, the problem is reduced to selecting the hypothesis (H_0 or H_1) with the largest posterior probability P($H_i|\hat{\mathbf{y}}$). From the Bayes rule:

$$L(\hat{\mathbf{y}}) = \frac{p_{\mathbf{y}|H_1}(\hat{\mathbf{y}}|H_1)}{p_{\mathbf{y}|H_0}(\hat{\mathbf{y}}|H_0)} \begin{array}{c} > \\ < \end{array} \frac{P[H_0]}{P[H_1]} \Rightarrow \begin{array}{l} \hat{\mathbf{y}} \leftrightarrow H_1 \\ \hat{\mathbf{y}} \leftrightarrow H_0 \end{array} \quad (1)$$

In the LRT, it is assumed that the number of observations is fixed and represented by a vector $\hat{\mathbf{y}}$. The performance of the decision procedure can be improved by incorporating more observations to the statistical test. When N measurements $\hat{\mathbf{y}}_1$, $\hat{\mathbf{y}}_2$, ..., $\hat{\mathbf{y}}_N$ are available in a two-class classification problem, a multiple observation likelihood ratio test (MO-LRT) can be defined by:

$$L_N(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, ..., \hat{\mathbf{y}}_N) = \frac{p_{\mathbf{y}_1,\mathbf{y}_2,...,\mathbf{y}_N|H_1}(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, ..., \hat{\mathbf{y}}_N|H_1)}{p_{\mathbf{y}_1,\mathbf{y}_2,...,\mathbf{y}_N|H_0}(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, ..., \hat{\mathbf{y}}_N|H_0)} \quad (2)$$

This test involves the evaluation of an *N*-th order LRT which enables a computationally efficient evaluation when the individual measurements $\hat{\mathbf{y}}_k$ are independent. However, they are not since the windows used in the computation of the observation vectors $\mathbf{y}_k$ are usually overlapped. In order to evaluate the proposed MCO-LRT VAD on an incoming signal, an adequate statistical model for the feature vectors in presence and absence of speech needs to be selected. The joint probability distributions under both hypotheses are assumed to be jointly gaussian independently distributed in frequency and in each part (real and imaginary) of vector with correlation components between each pair of frequency observations:

$$L_N(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, ..., \hat{\mathbf{y}}_N) = \prod_{p\in\{R,I\}}\{\prod_{\omega} \frac{p_{y_1^\omega, y_2^\omega, ..., y_N^\omega|H_1}(\hat{y}_1^\omega, \hat{y}_2^\omega, ..., \hat{y}_N^\omega|H_1)}{p_{y_1^\omega, y_2^\omega, ..., y_N^\omega|H_0}(\hat{y}_1^\omega, \hat{y}_2^\omega, ..., \hat{y}_N^\omega|H_0)}\}_p \quad (3)$$

This is a more realistic approach that the one presented in [9] taking into account the overlap between adjacent observations. We use following joint gaussian probability density function for each part:

$$p_{\mathbf{y}_\omega|H_s}(\hat{\mathbf{y}}_\omega|H_s)) = K_{H_s,N} \cdot \exp\{-\frac{1}{2}(\hat{\mathbf{y}}_\omega^T (C^N_{\mathbf{y}_\omega,H_s})^{-1}\hat{\mathbf{y}}_\omega)\} \quad (4)$$

for $s = 0, 1$, where $K_{H_s,N} = \frac{1}{(2\pi)^{N/2}|C^N_{\mathbf{y}_\omega,H_s}|^{1/2}}$, $\mathbf{y}_\omega = (\mathrm{y}_1^\omega, \mathrm{y}_2^\omega, \ldots, \mathrm{y}_N^\omega)^T$ is a zero-mean frequency observation vector, $C^N_{\mathbf{y},H_s}$ is the *N*-order covariance matrix of the observation vector under hypothesis H_s and $|.|$ denotes determinant of a matrix. The model selected for the observation vector is similar to that used by Sohn *et al.* [6] that assumes the discrete Fourier transform (DFT) coefficients of the clean speech (S_j) and the noise (N_j) to be asymptotically independent Gaussian random variables. In our case the observation vector consist of the real and imaginary parts of frequency DFT coefficient at frequency ω of the set of m observations.

3. Evaluation of the LRT

In order to evaluate the MCO-LRT, the computation of the inverse matrices and determinants are required. Since the covariances matrices under $H_0 \& H_1$ are assumed to be tridiagonal symmetric matrices[1], the inverses matrices can be computed as the following:

$$[C^{-1}_{\mathbf{y}_\omega}]_{mk} = [\frac{q_k}{p_k} - \frac{q_N}{p_N}]p_m p_k \quad N-1 \geq m \geq k \geq 0 \quad (6)$$

where N is the order of the model and the set of real numbers $q_n, p_n\ n = 1 \ldots \infty$ satisfies the three-term recursion for $k \geq 1$:

$$0 = r_k(q_{k-1}, p_{k-1}) + \sigma_{k+1}(q_k, p_k) + r_{k+1}(q_{k+1}, p_{k+1}) \quad (7)$$

with initial values:

$$\begin{array}{ll} p_0 = 1 & \text{and } p_1 = -\frac{\sigma_1}{r_1} \\ q_0 = 0 & \text{and } q_1 = \frac{1}{r_1} \end{array} \quad (8)$$

In general this set of coefficients are defined in terms of orthogonal complex polynomials which satisfy a Wronskian-like relation [11] and have the continued-fraction representation[12]:

$$\frac{q_n(z)}{p_n(z)} = \frac{1}{(z-\sigma_1)-} \ominus \frac{r_1^2}{(z-\sigma_2)-} \ominus \ldots \ominus \frac{r_{n-1}^2}{(z-\sigma_n)} \quad (9)$$

where $\ominus$ denotes the continuos fraction. This representation is used to compute the coefficients of the inverse matrices evaluated on $z = 0$. In the next section we show a new VAD based on this methodology for $N = 2$ and 3, that is, this robust speech detector is intended for real time applications such us mobile communications. The decision function will be described in terms of the correlation and variance coefficients which constitute a correction to the previous LRT method [9] that assumed uncorrelated observation vectors in the MO.

[1]The covariance matrix will be modeled as a tridiagonal matrix, that is, we only consider the correlation function between adjacent observations according to the number of samples (200) and window shift (80) that is usually selected to build the observation vector. This approach reduces the computational effort achieved by the algorithm with additional benefits from the symmetric tridiagonal matrix properties:

$$[C^N_{\mathbf{y}_\omega}]_{mk} = \left[\begin{array}{lll} \sigma^2_{y_m}(\omega) \equiv E[|y_m^\omega|^2] & if & m = k \\ r_{mk}(\omega) \equiv E[y_m^\omega y_k^\omega] & if & k = m+1 \\ 0 \quad other & case & \end{array}\right] \quad (5)$$

where $1 \leq i \leq j \leq N$ and $\sigma^2_{y_i}(\omega)$, $r_{ij}(\omega)$ are the variance and correlation frequency components of the observation vector $\mathbf{y}_\omega$ (denoted for clarity σ_i, r_i) which must be estimated using instantaneous values.

4. Application to voice activity detection

The use of the MO-LRT for voice activity detection is mainly motivated by two factors: *i*) the optimal behaviour of the so defined decision rule, and *ii*) a multiple observation vector for classification defines a reduced variance LRT reporting clear improvements in robustness against the acoustic noise present in the environment. The proposed MO-LRT VAD is described as follows. The MO-LRT is defined over the observation vectors $\{\hat{\mathbf{y}}_{l-m}, \ldots, \hat{\mathbf{y}}_{l-1}, \hat{\mathbf{y}}_l, \hat{\mathbf{y}}_{l+1}, \ldots, \hat{\mathbf{y}}_{l+m}\}$ as follows:

$$\ell_{l,N} = \sum_\omega \frac{1}{2}\left\{ {\mathbf{y}_\omega}^T \Delta_N^\omega \mathbf{y}_\omega + ln\left[\frac{|C^N_{\mathbf{y}_\omega,H_0}|}{|C^N_{\mathbf{y}_\omega,H_1}|}\right]\right\} \quad (10)$$

where $\Delta_N^\omega = (C^N_{\mathbf{y}_\omega,H_0})^{-1} - (C^N_{\mathbf{y}_\omega,H_1})^{-1}$, $N = 2m+1$ is the order of the model, l denotes the frame being classified as speech (H_1) or non-speech (H_0) and $\mathbf{y}_\omega$ is the previously defined frequency observation vector on the sliding window.

4.1. Analysis of JGPDF Voice Activity Detector for $N = 2$

In this section the improvement provided by the proposed methodology is evaluated by studying the most simple case for $N = 2$. In this case, assuming that squared correlations ρ_1^2 under $H_0 \& H_1$ and the correlation coefficients are negligible under H_0 (noise correlation coefficients $\rho_1^n \to 0$) vanish, the LRT can be evaluated according to:

$$\ell_{l,2} = \frac{1}{2}\sum_\omega L_1(\omega) + L_2(\omega) + 2\sqrt{\gamma_1\gamma_2}\left[\frac{\rho_1^s}{\sqrt{(1+\xi_1)(1+\xi_2)}}\right] \quad (11)$$

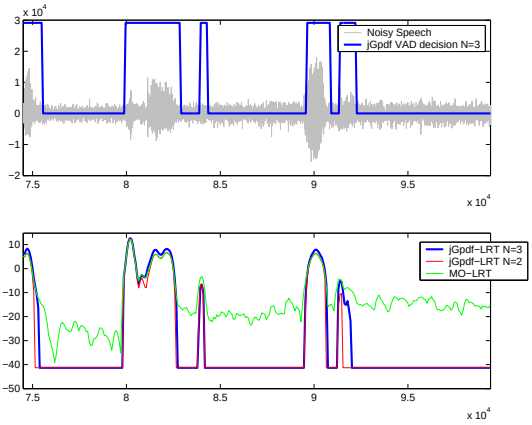
where $\rho_1^s = r_1^s(\omega)/(\sqrt{\sigma_1^s\sigma_2^s})$ is the correlation coefficient of the observations under H_1, $\gamma_i \equiv (y_i^\omega)^2/\sigma_i^n(\omega)$ and $\xi_i \equiv \sigma_i^s(\omega)/\sigma_i^n(\omega)$ are the SNRs a priori and a posteriori of the DFT coefficients, $L_{\{1,2\}}(\omega) \equiv \frac{\gamma_{\{1,2\}}\xi_{\{1,2\}}}{1+\xi_{\{1,2\}}} - \ln(1+\xi_{\{1,2\}})$ are the independent LRT of the observations $\hat{\mathbf{y}}_\mathbf{1}, \hat{\mathbf{y}}_\mathbf{2}$ (connection with the previous MO-LRT [9]) which are corrected with the term depending on ρ_1^s, the new parameter to be modeled, and l indexes to the second observation. At this point frequency ergodicity of the process must be assumed to estimate the new model parameter ρ_1^s. This means that the correlation coefficients are constant in frequency thus an ensemble average can be estimated using the sample mean correlation of the observations $\hat{\mathbf{y}}_\mathbf{1}$ and $\hat{\mathbf{y}}_\mathbf{2}$ included in the sliding window.

4.2. Analysis of JGPDF Voice Activity Detector for $N = 3$

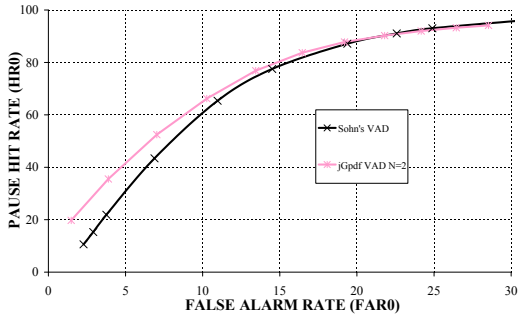
In the case for $N = 3$ the properties of a symmetric and tridiagonal matrix come out. The likelihood ratio can be expressed as:

$$\ell_{l,3} = \sum_\omega \ln\frac{K_{H_1,3}}{K_{H_0,3}} + \frac{1}{2}\hat{\mathbf{y}}_\omega^T \Delta_3^\omega \hat{\mathbf{y}}_\omega \quad (12)$$

where $\ln\frac{K_{H_1,3}}{K_{H_0,3}} = \frac{1}{2}\left[\ln\left[\frac{1-(\rho_1^2+\rho_2^2)^{H_0}}{1-(\rho_1^2+\rho_2^2)^{H_1}}\right] - \ln\ \prod_{i=1}^3(1+\xi_i)\ \right]$, and Δ_3^ω is computed using the following expression under



(a)



(b)

Figure 1: a) JGPDF-VAD vs. MO-LRT decision for $N = 2$ and 3. b) ROC curve for JGPDF VAD with $l_h = 8$ and Sohn's VAD [6] using a similar hang-over mechanism.

hypotheses $H_0 \& H_1$:

$$\hat{\mathbf{y}}_\omega^T (C^3_{\mathbf{y}_\omega, H_s})^{-1} \hat{\mathbf{y}}_\omega = \frac{1}{1-(\rho_1^2+\rho_2^2)} \left[\frac{1-\rho_2^2}{\sigma_1} (y_1^\omega)^2 + \frac{(y_2^\omega)^2}{\sigma_2} \dots \right] \left[+ \frac{1-\rho_1^2}{\sigma_3} (y_3^\omega)^2 - 2\rho_1 \frac{y_1^\omega y_2^\omega}{\sqrt{\sigma_1 \sigma_2}} - 2\rho_2 \frac{y_2^\omega y_3^\omega}{\sqrt{\sigma_2 \sigma_3}} + 2\rho_1 \rho_2 \frac{y_1^\omega y_3^\omega}{\sqrt{\sigma_1 \sigma_3}} \right] \tag{13}$$

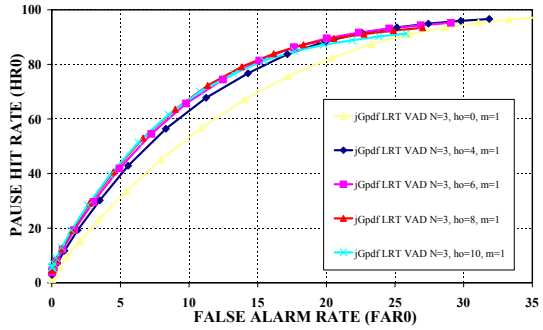
Assuming that squared correlations under $H_0 \& H_1$ and the correlations under H_0 vanish, the log-LRT can be evaluated as the following:

$$\ell_{l,3} = \frac{1}{2} \sum_\omega \sum_{i=1}^3 L_i(\omega) + \frac{2\sqrt{\gamma_1 \gamma_2} \rho_1^s}{\sqrt{(1+\xi_1)(1+\xi_2)}} + \frac{2\sqrt{\gamma_2 \gamma_3} \rho_2^s}{\sqrt{(1+\xi_2)(1+\xi_3)}} - \frac{2\sqrt{\gamma_1 \gamma_3} \rho_1^s \rho_2^s}{\sqrt{(1+\xi_1)(1+\xi_2)^2(1+\xi_3)}} \tag{14}$$

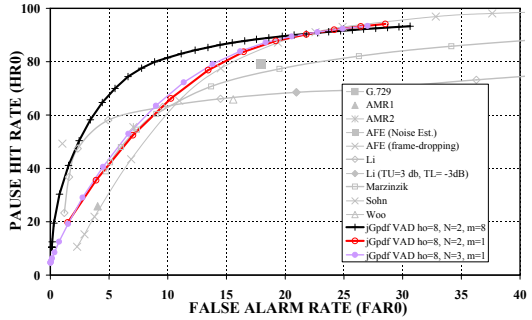
5. Experimental Framework

The ROC curves are frequently used to completely describe the VAD error rate. The AURORA 3 subset of the original Spanish SpeechDat-Car (SDC) database [10] was used in this analysis. The files are categorized into three noisy conditions: quiet, low noisy and highly noisy conditions, which represent different driving conditions with average SNR values between 25dB, and 5dB. The non-speech hit rate (HR0) and the false alarm rate (FAR0= 100-HR1) were determined in each noise condition.

Using the proposed decision functions (equations 14 and 11) we obtain an almost binary decision rule as it is shown in figure 1(a) which accurately detects the beginnings of the voice



(a)



(b)

Figure 2: a) ROC curve analysis of the jGpdf-VAD ($N = 3$) for the selection of the hang-over parameter l_h. b) ROC curves of the jGpdf-VAD using contextual information (eight MO windows for $N = 2$) and standards and recently reported VADs.

periods. In this figure we have used the same level of information in both methods ($m = 1$). The detection of voice endings is improved using a hang-over scheme based on the decision of previous frames. Observe how this strategy cannot be applied to the independent LRT [6] because of its hard decision rule and changing bias as it is shown in the same figure. We implement a very simple hang-over mechanism based on contextual information of the previous frames, thus no delay obstacle is added to the algorithm:

$$\ell_{l,N}^h = \ell_{l,N} + \ell_{l-l_h,N} \tag{15}$$

where the parameter l_h is selected experimentally. The ROC curve analysis for this hang-over parameter is shown in figure 2(a) for $N = 3$ where the influence of hang-over in the zero hit rate is studied with variable detection threshold. Finally, the benefits of contextual information [9] can be incorporated just averaging the decision rule over a set of multiple observations windows (two observations for each window). A typical value for $m = 8$ produces increasing levels of detection accuracy as it is shown in the ROC curve in figure 2(b). Of course, these results are not the optimum ones since only pair-wise dependence is considered here. However for a small number of observations the proposed VAD presents the best trade-off between detection accuracy and delay.

6. Conclusion

This paper showed a new VAD for improving speech detection robustness in noisy environments. The proposed method is developed on the basis of previous proposals that incorporate long-term speech information to the decision rule [9]. However, it is not based on the assumption of independence between observations since this hypothesis is not realistic at all. It defines a statistically optimum likelihood ratio test based on multiple and correlated observation vectors which avoids the need of smoothing the VAD decision, thus reporting significant benefits for speech/pause detection in noisy environments. The algorithm has an optional inherent delay that, for several applications including robust speech recognition, does not represent a serious implementation obstacle. An analysis based on the ROC curves unveiled a clear reduction of the classification error for second and third order model. In this way, the proposed VAD outperformed, at the same conditions, the Sohn's VAD, as well as the standardized G.729, AMR and AFE VADs and other recently reported VAD methods in both speech/non-speech detection performance.

6.1. Computation of the LRT for $N = 2$

From equation 4 for $N = 2$ we have that the MCO-LRT can be expressed as:

$$\ell_{l,2} = \sum_{\omega} \ln \frac{K_{H_1,2}}{K_{H_0,2}} + \frac{1}{2}\hat{\mathbf{y}}_{\omega}^{T}\Delta_2^{\omega}\hat{\mathbf{y}}_{\omega} \tag{16}$$

where:

$$\ln \frac{K_{H_1,2}}{K_{H_0,2}} = \frac{1}{2} ln \quad \frac{|C^N_{\mathbf{y}_\omega,H_0}|}{|C^N_{\mathbf{y}_\omega,H_1}|}\Bigg) = \frac{1}{2}\frac{\sigma_1^{H_0}\sigma_2^{H_0} - (r_1^{H_0})^2}{\sigma_1^{H_1}\sigma_2^{H_1} - (r_1^{H_1})^2} \tag{17}$$

and $C_{\mathbf{y}_\omega}$ is defined as in equation 5. If we assume that the voice signal is observed in additive independent noise, that is for $i = 1, 2$:

$$\begin{array}{ll} H_1: & \sigma_i^{H_1} = \sigma_i^n + \sigma_i^s \\ H_0: & \sigma_i^{H_0} = \sigma_i^n \end{array} \tag{18}$$

and define the correlation coefficient $\rho_1^{H_s} \equiv \frac{r_1^{H_1}}{\sqrt{\sigma_1^{H_1}\sigma_2^{H_1}}}$ and the a posteriori SNR $\xi_i \equiv \frac{\sigma_i^s}{\sigma_i^n}$, we have that:

$$\ln \frac{K_{H_1,2}}{K_{H_0,2}} = \frac{1}{2}\left[\ln \quad \frac{1-(\rho_1^{H_0})^2}{1-(\rho_1^{H_1})^2} \quad - \ln \quad \prod_{i=1}^{2}(1+\xi_i)\right)\Bigg] \tag{19}$$

On the other hand, the inverse matrix is expressed in terms of the orthogonal complex polynomials $q_k(z), p_k(z)$ as:

$$(C^2_{\mathbf{y}_\omega,H_s})^{-1} = \left(\begin{array}{cc} \left[\frac{q_0}{p_0} - \frac{q_2}{p_2}\right] p_0 p_0 & \left[\frac{q_1}{p_1} - \frac{q_2}{p_2}\right] p_0 p_1 \\ \left[\frac{q_1}{p_1} - \frac{q_2}{p_2}\right] p_0 p_1 & \left[\frac{q_1}{p_1} - \frac{q_2}{p_2}\right] p_1 p_1 \end{array} \right)_{H_s} \tag{20}$$

where $p_0 = 1$, $q_0 = 0$, $p_1 = -\sigma_1/r_1$ and $q_2/p_2 = \sigma_2/(r_1^2 - \sigma_1\sigma_2)$ under hypothesis H_s. Thus the second term of equation 16 can be expressed as:

$$\hat{\mathbf{y}}_{\omega}^{T}\Delta_2^{\omega}\hat{\mathbf{y}}_{\omega} = (y_1^{\omega})^2(\Delta_2^{\omega})_{00} + (y_2^{\omega})^2(\Delta_2^{\omega})_{11} + 2y_1^{\omega}y_2^{\omega}(\Delta_2^{\omega})_{01} \tag{21}$$

where $(\Delta_2^{\omega})_{00} = \frac{\sigma_2^{H_0}}{\sigma_2^{H_0}\sigma_1^{H_0} - (r_1^{H_0})^2} - \frac{\sigma_2^{H_1}}{\sigma_2^{H_1}\sigma_1^{H_1} - (r_1^{H_1})^2}$, $(\Delta_2^{\omega})_{11} = \frac{\sigma_1^{H_0}}{\sigma_2^{H_0}\sigma_1^{H_0} - (r_1^{H_0})^2} - \frac{\sigma_1^{H_1}}{\sigma_2^{H_1}\sigma_1^{H_1} - (r_1^{H_1})^2}$ and $(\Delta_2^{\omega})_{01} = \frac{r_1^{H_0}}{(r_1^{H_0})^2 - \sigma_2^{H_0}\sigma_1^{H_0}} - \frac{r_1^{H_1}}{(r_1^{H_0})^2 - \sigma_2^{H_0}\sigma_1^{H_0}}$. Finally, if we define the a priori SNR $\gamma_i \equiv (y_i^{\omega})^2/\sigma_i^n(\omega)$ and neglect the squared correlation functions under both hypotheses we have equation 11.

7. References

[1] A. Benyassine, E. Shlomot, H. Su, D. Massaloux, C. Lamblin, and J. Petit, "ITU-T Recommendation G.729 Annex B: A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 64–73, 1997.

[2] ITU, "A silence compression scheme for G.729 optimized for terminals conforming to recommendation V.70," *ITU-T Recommendation G.729-Annex B*, 1996.

[3] ETSI, "Voice activity detector (VAD) for Adaptive Multi-Rate (AMR) speech traffic channels," *ETSI EN 301 708 Recommendation*, 1999.

[4] ——, "Speech processing, transmission and quality aspects (STQ); distributed speech recognition; advanced front-end feature extraction algorithm; compression algorithms," *ETSI ES 201 108 Recommendation*, 2002.

[5] R. L. Bouquin-Jeannes and G. Faucon, "Study of a voice activity detector and its influence on a noise reduction system," *Speech Communication*, vol. 16, pp. 245–254, 1995.

[6] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters*, vol. 16, no. 1, pp. 1–3, 1999.

[7] Y. D. Cho, K. Al-Naimi, and A. Kondoz, "Improved voice activity detection based on a smoothed statistical likelihood ratio," in *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, 2001, pp. 737–740.

[8] J. M. Górriz, J. Ramírez, J. C. Segura, and C. G. Puntonet, "An effective cluster-based model for robust speech detection and speech recognition in noisy environments," *Journal of Acoustical Society of America*, vol. 120, no. 470, pp. 470–481, 2006.

[9] J. M. Górriz, J. Ramirez, J. C. Segura, and C. G. Puntonet, "An improved mo-lrt vad based on a bispectra gaussian model," *Electronic Letters*, vol. 41, no. 15, pp. 877–879, 2005.

[10] A. Moreno, L. Borge, D. Christoph, R. Gael, C. Khalid, E. Stephan, and A. Jeffrey, "SpeechDat-Car: A Large Speech Database for Automotive Environments," in *Proceedings of the II LREC Conference*, 2000.

[11] N. Akhiezer, *The Classical Moment Problem*. Edimburgh: Oliver and Boyd, 1965.

[12] H. Yamani and M. Abdelmonem, "The analytic inversion of any finite symmetric tridiagonal matrix," *J. Phys. A: Math Gen*, vol. 30, pp. 2889–2893, 1997.