ON THE COMPETITIVE NEYMAN-PEARSON APPROACH FOR COMPOSITE HYPOTHESIS TESTING AND ITS APPLICATION IN VOICE ACTIVITY DETECTION

Abhijeet Sangwan, W.-P. Zhu and M.O. Ahmad

Centre for Signal Processing and Communication, Department of Electrical and Electronics Engineering
Concordia University, Montreal, Quebec, Canada H3G 1M8
{a_sangwa,weiping,omair}@ece.concordia.ca

ABSTRACT

The problem of composite hypothesis testing where the probability law governing the generation of the free parameter is not explicitly known is considered. It is shown that unlike the Neyman-Pearson (NP) approach, the competitive NP (CNP) approach models incomplete prior information about the source into the detector design by setting a variable upper bound for the probability of false-alarm term. Further, the CNP and NP approaches are employed to develop the CNP and NP detectors for voice activity detection (VAD), where the prior SNR is shown to be the free parameter of the composite hypothesis. We test the CNP and NP detectors using speech samples from the SWITCHBOARD database which are suitably corrupted using different noises and various SNRs. Our simulation results show that the CNP detector outperforms its NP counterpart and is comparable to the adaptive multi-rate (AMR) VADs.

1. INTRODUCTION

In the composite hypothesis testing problem, the source output is a parameter $\theta \in \mathbf{\Theta_i}$ generated by the hypothesis H_i, i=0,1 where $\mathbf{\Theta_i}$ is the parameter set corresponding to H_i. The union of the parameter sets, $\mathbf{\Theta_0} \cup \mathbf{\Theta_1} = \mathbf{\Theta}$ gives the parameter space and a probability density function (pdf) p_θ maps $\mathbf{\Theta}$ onto the observation space $\mathbf{Z}$. A decision rule splits the observation space into two complementary subspaces, $\mathbf{Z_0}$ and $\mathbf{Z_1}$ such that $\mathbf{Z_0} \cup \mathbf{Z_1} = \mathbf{Z}$ where the observation $Z \in \mathbf{Z_i}$ implies that the decision is made in the favor of hypothesis H_i [1].

The popular design methodologies for hypothesis testing are the Neyman-Pearson (NP) approach and the Bayes criterion. Both techniques result in a likelihood ratio test (LRT) where the likelihood ratio (LR) is compared with a threshold [1], i.e.,

$$\Lambda \triangleq \frac{p_{z|H_1}(Z|H_1)}{p_{z|H_0}(Z|H_0)} \overset{>H_1}{<_{H_0}} \gamma, \tag{1}$$

where Λ is the LR, γ is the threshold and $p_{z|H_i}$ is the conditional pdf of the observation Z. The error probabilities associated with hypothesis testing are the probability of false-alarm P_f, where H_1 is chosen while H_0 is true, and the probability of miss-detection P_m, where H_0 is chosen while H_1 is true . [1, 2], i.e.,

$$\begin{aligned} P_f(\gamma|\theta) &\triangleq p_\theta(Z \in \mathbf{Z_1}|\theta), \\ &= \int_\gamma^\infty p(\Lambda|H_0,\theta)d\Lambda, \end{aligned} \tag{2}$$

$$\begin{aligned} P_m(\gamma|\theta) &\triangleq p_\theta(Z \in \mathbf{Z_0}|\theta), \\ &= \int_{-\infty}^\gamma p(\Lambda|H_1,\theta)d\Lambda, \end{aligned} \tag{3}$$

where $p(\Lambda|H_0,\theta)$ and $p(\Lambda|H_1,\theta)$ are the conditional pdfs of the LR. The overall error probability $P_e(\gamma|\theta)$ of the test is given by [2]:

$$P_e(\gamma|\theta) \triangleq P(H_1)P_m(\gamma|\theta) + P(H_0)P_f(\gamma|\theta), \tag{4}$$

where $P(H_i)$ is the prior probability of the hypothesis H_i. The NP approach chooses a γ which minimizes P_m while constraining P_f, i.e.,

$$\begin{aligned} &\min_\gamma \ P_m(\gamma|\theta), \\ &P_f(\gamma|\theta) \le \lambda, \end{aligned} \tag{5}$$

where λ is a constant and the detector hence obtained is termed as the NP detector. Similarly, the Bayes criterion minimizes the total error probability,

$$\min_\gamma \ P_e(\gamma|\theta). \tag{6}$$

which yields the Bayesian detector.

2. COMPETITIVE NEYMAN-PEARSON APPROACH

The competitive Neyman-Pearson (CNP) is a modification over the basic NP approach where the constraint over hypotheses testing is formulated as, [2]:

$$\begin{aligned} &\min_\gamma \ P_m(\gamma|\theta), \\ &P_f(\gamma|\theta) \le \lambda(\theta), \end{aligned} \tag{7}$$

where unlike the NP, CNP approach constrains P_f with a variable upper bound. In order to show that the CNP approach models prior information, we express the prior probabilites as:

$$P(H_i) = \int_{\mathbf{\Theta_i}} P(H_i|\theta)p(\theta)d\theta,$$

where $p(\theta)$ is the pdf of the parameter θ. Using the Bayes rule, the above expression is rewritten as:

$$P(H_i) = \int_{\mathbf{\Theta_i}} p(\theta|H_i)P(H_i)d\theta.$$

Using the above expression in (4), along with the expressions for P_f and P_m in (2) and (3), we get:

$$P_e(\gamma) = \sum_{i=0,1} \int_{\mathbf{\Theta_i}} p(\theta|H_i)P(H_i)d\theta \int_{\Lambda_i} p(\Lambda|H_i,\theta)d\Lambda,$$

where $\Lambda_1 = \{\Lambda : \Lambda \geq \gamma\}$, $\Lambda_0 = \{\Lambda : \Lambda < \gamma\}$. Now, the above expression is simplified as:

$$P_e(\gamma) = \sum_{i=0,1} P(H_i) \int_{\Lambda_i} \int_{\mathbf{\Theta_i}} p(\Lambda|H_i,\theta)p(\theta|H_i)d\theta d\Lambda. \quad (8)$$

In some practical problems, the probability law governing the generation of θ from the source, $p(\theta|H_i)$ is completely known and the composite hypothesis problem is easily reduced to a simple hypotheses problem [1, 2], i.e.,

$$\begin{aligned} P_e(\gamma) &= \sum_{i=0,1} P(H_i) \int_{\Lambda_i} \int_{\mathbf{\Theta_i}} p(\Lambda,\theta|H_i)d\theta d\Lambda, \\ &= \sum_{i=0,1} P(H_i) \int_{\Lambda_i} p(\Lambda|H_i)d\Lambda, \\ &= P(H_1)P_m(\gamma) + P(H_0)P_f(\gamma), \end{aligned}$$

where the last equation represents the Bayes error for a binary hypothesis. It is noted that the error terms P_m, P_f and P_e in the last expression are independent of the parameter θ. However, a straightforward design for hypothesis testing does not exist if θ is generated by an unknown pdf or if θ is non-random, as the parameter θ can no longer be removed via the integration [1]. Alternatively, we could work with the conditional error term which is obtained by multiplying and dividing the expression in (8) by $p(\theta)$, i.e.,

$$\begin{aligned} P_e(\gamma) &= \sum_{i=0,1} P(H_i) \int_{\Lambda_i} \int_{\mathbf{\Theta_i}} p(\Lambda|H_i,\theta)p(\theta|H_i)\frac{p(\theta)}{p(\theta)}d\theta d\Lambda, \\ &= \int_{\mathbf{\Theta_i}} \{\sum_{i=0,1} P(H_i) \int_{\Lambda_i} p(\Lambda|H_i,\theta)\frac{p(\theta|H_i)}{p(\theta)}d\Lambda\}p(\theta)d\theta, \end{aligned}$$

where the term inside the paranthesis $\{.\}$ is the conditional error term $P_e(\gamma|\theta)$, i.e.,

$$\begin{aligned} P_e(\gamma|\theta) &= \sum_{i=0,1} P(H_i) \int_{\Lambda_i} p(\Lambda|H_i,\theta)\frac{p(\theta|H_i)}{p(\theta)}d\Lambda, \\ &= \sum_{i=0,1} P(H_i)\frac{p(\theta|H_i)}{p(\theta)} \int_{\Lambda_i} p(\Lambda|H_i,\theta)d\Lambda, \\ &= \frac{P(H_1)p(\theta|H_1)P_m(\gamma|\theta)}{p(\theta)} + \frac{P(H_0)p(\theta|H_0)P_f(\gamma|\theta)}{p(\theta)}. \end{aligned}$$

In the above expression, the first and second terms represent the contributions to the overall error due to miss-detection and false-alarm respectively. If we minimize the error due to the first error term while constraining the second error term by a constant value λ, we get:

$$\min_{\gamma} \; \frac{P(H_1)p(\theta|H_1)P_m(\gamma|\theta)}{p(\theta)}, \quad (9)$$

$$\frac{P(H_0)p(\theta|H_0)P_f(\gamma|\theta)}{p(\theta)} \leq \lambda. \quad (10)$$

The terms in (9) which do not contain γ can be removed from the minimization. Further, the inequality in (10) may be rewritten as:

$$P_f(\gamma|\theta) \leq \frac{\lambda p(\theta)}{P(H_0)p(\theta|H_0)}, \quad (11)$$

where $\lambda' = \frac{\lambda}{P(H_0)}$ is a constant, and the minimization and constraint conditions in (9) and (10) can be rewritten in a simplified form as:

$$\min_{\gamma} \; P_m(\gamma|\theta), \quad (12)$$

$$P_f(\gamma|\theta) \leq \lambda' \frac{p(\theta)}{p(\theta|H_0)}. \quad (13)$$

Now, if θ is independent of H_i, then $p(\theta|H_0) = p(\theta)$ and the RHS in (13) reduces to only λ' which makes the minimization-constraint conditions similar to the NP approach given in (5). On the other hand, if there is a dependency between θ and H_i then the RHS in (13) becomes a function of θ, i.e., $\lambda'(\theta)$ which makes the minimization-constraint conditions similar to the CNP approach given in (7). Hence, it is easily noted that the CNP approach models the dependency between the parameter and the hypothesis by defining an upper bound on P_f which is a function of the parameter itself. In other words, P_f must take on different values with varying θ to maintain the constraints of the CNP approach. The need to employ the CNP approach arises in many practical situations, where partial knowledge about the source exists in form of a general relationship between θ and the underlying hypothesis H_i. For instance, in a radar-detection problem where a higher value of SNR may be more suggestive of signal presence than lower SNR it may be necessary to vary P_f with the signal to noise ratio (SNR) [2].

In this paper, we use a different technique in designing the CNP LRT where we do not explicitly determine $\lambda'(\theta)$ in (13). Instead we define an arbitary probability term P_a which is related to P_f as:

$$P_a(\gamma'|\theta) = \int_{\gamma'}^{\infty} p(\Lambda|H_0,\theta)d\Lambda, \quad (14)$$

where the curve γ' is a function of θ, and a one to one mapping exists between the threshold γ and γ'. If the curve γ' is appropriately chosen, it yeilds a desirable variation in P_f with varying θ, which is shown below by rewriting the above expression as:

$$P_a(\gamma'|\theta) = \int_{\gamma'}^{\gamma} p(\Lambda|H_0,\theta)d\Lambda + \int_{\gamma}^{\infty} p(\Lambda|H_0,\theta)d\Lambda,$$

where the second term of the RHS in the above expression is P_f. Hence, the above expression may be rewritten as:

$$\begin{aligned} P_a(\gamma'|\theta) &= \int_{\gamma'}^{\gamma} p(\Lambda|H_0,\theta)d\Lambda + P_f(\gamma|\theta), \\ P_f(\gamma|\theta) &= P_a(\gamma'|\theta) + \int_{\gamma}^{\gamma'} p(\Lambda|H_0,\theta)d\Lambda. \end{aligned}$$

In the above expression if γ' is chosen to keep P_a constant, then $P_f \leq P_a \; \forall\theta$ when $\gamma' \leq \gamma$, and $P_f > P_a \; \forall\theta$ where $\gamma' > \gamma$. Hence, a designer can set different upper bounds for P_f with varying θ by suitably adjusting the distance between γ' and γ, and maintaining $\gamma' \leq \gamma$. In addition, one also has the flexibility of setting different lower bounds on P_f (or equivalently upper bounds on P_m) by maintaining $\gamma' > \gamma$.

3. VOICE ACTIVITY DETECTION

Voice activity detection (VAD) falls under the category of the signal in noise problem, i.e.,

$$\begin{aligned} H_0 &: \mathbf{Z} = \mathbf{N} \\ H_1 &: \mathbf{Z} = \mathbf{S} + \mathbf{N}, \end{aligned}$$

where $\mathbf{S} = [s_1, s_2, ..., s_N]$, $\mathbf{N} = [n_1, n_2, ..., n_N]$ and $\mathbf{Z} = [z_1, z_2, ..., z_N]$ are the N-length speech, noise and observation vectors. We use the modified Ephraim-Malah (EM) model for the binary hypothesis where the observation $\mathbf{Z}$ are N-point independent mel-frequency spectral coefficients (MFSC) and are assumed to be Gaussian distributed with zero-mean [3], i.e.,

$$\begin{aligned} H_0 &: \mathbf{Z} \sim N(0, \mathbf{K_n}) \\ H_1 &: \mathbf{Z} \sim N(0, \mathbf{K_z}), \end{aligned}$$

and the general form of the detector is given as, [1]:

$$\frac{\mathbf{Z}(\mathbf{K_n}^{-1} - \mathbf{K_z}^{-1})\mathbf{Z}^T}{2} \gtrless_{H_0}^{H_1} \ln(\eta) + \frac{1}{2} \ln \frac{|\mathbf{K_z}|}{|\mathbf{K_n}|} \qquad (15)$$

where η is the Bayesian threshold, $|.|$ is the determinant, and K_z and K_n are the covariance matrices of the noisy speech $(\mathbf{S} + \mathbf{N})$ and noise $(\mathbf{N})$. Further, the estimates of the covariance matrices are obtained as per the convex combination rule where $\mathbf{K_n}$ and $\mathbf{K_z}$ are updated during noise and noisy speech periods only, respectively [3].

The sufficient statistics (l) of the test (LHS in (15)) is a speech energy estimator and this is shown by defining a transformation $\mathbf{Q}$ such that $\mathbf{Q}$ simultaneously diagonalizes $\mathbf{K_n}$ and $\mathbf{K_z}$, i.e.,

$$\begin{aligned} \mathbf{Q}^T\mathbf{K_n}\mathbf{Q} &= \mathbf{I} & (16) \\ \mathbf{Q}^T\mathbf{K_z}\mathbf{Q} &= \mathbf{\Lambda} & (17) \end{aligned}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{\Lambda}$ is an eigenvalue matrix whose i^{th} eigenvalue is given by λ_i. It is well known that such a transform $\mathbf{Q}$ exists [4]. Now, let $\mathbf{F} = \mathbf{ZQ}$ be the transformed observation and the detector in the domain of $\mathbf{Q}$ is obtained by rewriting the SS in (15) as:

$$\begin{aligned} l &= \frac{\mathbf{F}\mathbf{Q}^{-1}(\mathbf{K_n}^{-1} - \mathbf{K_z}^{-1})(\mathbf{Q}^{-1})^T\mathbf{F}^T}{2}, \\ &= \frac{\mathbf{F}(\mathbf{Q}^{-1}\mathbf{K_n}^{-1}(\mathbf{Q}^{-1})^T - \mathbf{Q}^{-1}\mathbf{K_z}^{-1}(\mathbf{Q}^{-1})^T)\mathbf{F}^T}{2}, \\ &= \frac{\mathbf{F}(\mathbf{I} - \mathbf{\Lambda}^{-1})\mathbf{F}^T}{2}, \qquad (18) \end{aligned}$$

and the above expression may be rewritten in scalar form as:

$$l = \frac{1}{2}\sum_{i=1}^{m}(1 - \frac{1}{\lambda_i})f_i^2 \qquad (19)$$

where f_i is the i^{th} element of $\mathbf{F}$. On the lines of Ephraim and Malah, we define the i^{th} posterior (γ_i) and prior (ζ_i) SNR as [5],

$$\begin{aligned} \gamma_i &= \lambda_i, & (20) \\ \zeta_i &= \lambda_i - 1, & (21) \end{aligned}$$

and rewrite l in (19) using the above definitions of the prior and posterior SNR as:

$$l = \frac{1}{2}\sum_{i=1}^{m}(\frac{\zeta_i}{\zeta_i + 1})f_i^2. \qquad (22)$$

The above expression shows that l estimates the proportion of speech energy along the i^{th} eigenvector on the basis of the prior SNR estimate, which also forms the parameter of the composite hypothesis (θ). If ζ_i is low indicating low prior SNR, then l value is low too indicating low speech energy in the noisy signal and vice versa. The estimates of ζ_i are obtained from long-term data itself and given that we obtain reliable estimates, the value of ζ_i is itself strongly suggestive of presence of speech. In other words, it is reasonable to assume that a high value of ζ_i is more likely to be associated with the 'speech hypothesis' than 'pause hypothesis' and vice-versa. Alternatively, a general relationship between the prior SNR and the hypothesis exists and a CNP approach must be used to build the VAD. For a low value of prior SNR which is suggestive of pause, the detector must be biased towards H_0 by using a low and high value of P_f and P_m respectively. Alternatively, the detector must be biased towards H_1 at high prior SNR where a high P_f and low P_m must be obtained.

The expressions for P_f and P_m are necessary in order to develop a CNP detector. It can be shown that the conditional pdfs of the SS, $p(l|H_0)$ and $p(l|H_1)$ are Gaussian distributed as the SS is a weighted sum of independent and identically distributed (i.i.d) random variables. Hence, using (2) and (3) the expressions for P_f and P_m are obtained as:

$$\begin{aligned} P_f &= 1 - erf(\frac{\gamma - E[l|H_0]}{\sqrt{Var[l|H_0]}}) & (23) \\ P_m &= erf(\frac{\gamma - E[l|H_1]}{\sqrt{Var[l|H_1]}}) & (24) \end{aligned}$$

where $erf(.)$ in the standard error function [3], $E[.]$ and $Var[.]$ are the mean and variance and the expression for the conditional statistics of l can be computed using (19). Now, the NP threshold is easily computed using the expression for P_f in (24) as:

$$\gamma_{NP} = \sqrt{Var[l|H_0]}erf^{-1}(1 - P_f) + E[l|H_0]. \qquad (25)$$

where $erf^{-1}(.)$ is the inverse of the standard error function. In order to obtain the CNP detector with the desired trend in P_f and P_m with varying prior SNR, we propose the following expression for γ':

$$\gamma' = \sqrt{Var[l|H_0]}(\frac{\ln\eta}{d \times S(\overline{\zeta})} + \frac{d}{2}) + E[l|H_0], \qquad (26)$$

where d is the normalized distance term indicative of the separability of the hypotheses [1,3]. $S(\overline{\zeta})$ is similar to the sigmoid function, i.e.,

$$S(\overline{\zeta}) = 2 - \frac{2}{1 + exp(-\overline{\zeta})}, \qquad (27)$$

where $\overline{\zeta}$ is the average value of prior SNR,

$$\overline{\zeta} = \frac{1}{N}\sum_{i=1}^{N}\zeta_i. \qquad (28)$$

Using the above expression for γ' in (14), we get:

$$\ln(\eta) = S(\overline{\zeta})(d \times erf^{-1}(1 - P_a) - \frac{d^2}{2}), \qquad (29)$$

and adding the term $\frac{1}{2}\ln\frac{|\mathbf{K_z}|}{|\mathbf{K_n}|}$ on both sides, we get the CNP threshold as:

$$\gamma_{CNP} = S(\overline{\zeta})(d \times erf^{-1}(1 - P_a) - \frac{d^2}{2}) + \frac{1}{2}\ln\frac{|\mathbf{K_z}|}{|\mathbf{K_n}|}. \qquad (30)$$

We plot the variation in P_m and P_f with changing prior SNR for the CNP, NP and Bayesian detectors in Fig. 1. It can be observed that while all detectors achieve the ideal trend for P_m with changing prior SNR, only the CNP detector achieves the ideal behaviour for P_f. Further the role of $S(\overline{\zeta})$ in (30) is also demonstrated where a

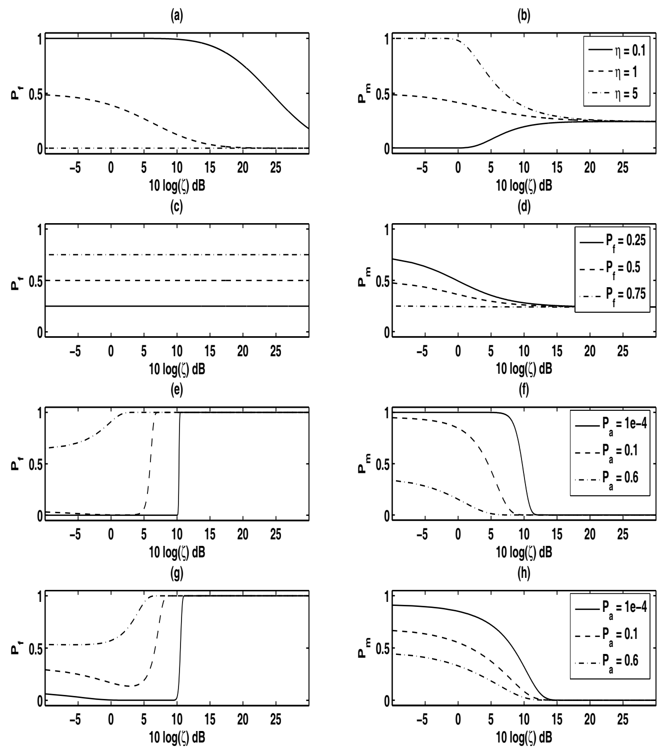

Fig. 1. Variation of P_f and P_m with varying prior SNR in dB for (a),(b) Bayesian (c),(d) NP (e),(f) CNP with $\overline{\zeta}$ = -5 dB and (g)(h) CNP with $\overline{\zeta}$ = 5 dB.

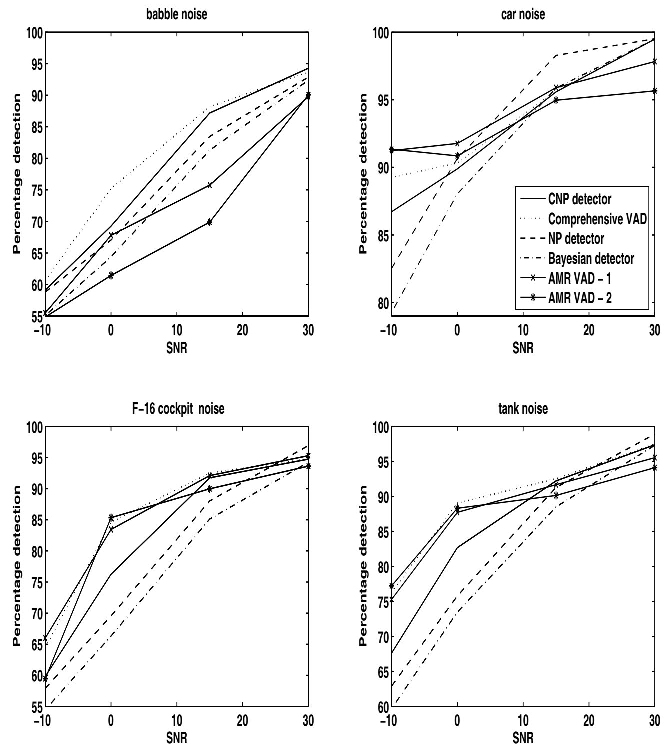

Fig. 2. Overall percentage detection for CNP, NP, Bayesian and Comprehensive VAD compated to the AMR VAD algorithm 1 and 2 for speech corrupted at different SNRs using (a) babble (b) car (c) F-16 cockpit and (d) tank noise.

lower value of average prior SNR ($\overline{\zeta} = -5dB$) gives a stronger bias towards H_0 in Fig. 1 (e), (f) when compared to 1 (g), (h) where the average prior SNR is higher ($\overline{\zeta} = 5dB$).

4. RESULTS AND DISCUSSION

We test the CNP LRT using a set of 21 speech samples of one minute duration each from the SWITCHBOARD database. The speech samples are corrupted using babble, car, F-16 cockpit and tank noise from the NOISEX database to create noisy speech signals of -10, 0, 15 and 30 dB SNR. In Fig. 2, the overall percentage detection for the CNP, NP and Bayesian detectors is shown and compared to the adaptive multi-rate (AMR) VAD 1 and 2 [3]. We also show the performance of a comprehensive VAD scheme which is formed by combining the CNP detector with the contextual detector which was developed in a earlier work [3]. As expected, all VAD schemes show good performance in high SNR and particularly in the case of car noise, the results for all detectors are good across all SNRs. In other cases, the CNP detector is seen to consistently perform better than its NP and Bayesian counterparts with the biggest improvements at intermediate SNRs. The CNP detector and the comprehensive VAD scheme are impressive in babble noise where they outperform other VADs by a big margin.

5. CONCLUSION

In this paper, we have shown that incomplete prior information in form of a general relationship between the free parameter and the hypothesis in composite hypothesis testing can be modeled using the CNP approach. Our approach of designing the CNP detector gives an added advantage of defining an upper as well as a lower bound for P_f with different values of the free parameter, where the detector behavior adapts with the value of the free parameter. The superiority of the CNP approach over the NP was established in the context of voice activity detection where P_f and P_m varying with SNR is desirable.

6. REFERENCES

[1] Harry L. Van Trees, *Detection, Estimation and Modulation Theory, Part 1*, John Wiley and Sons Inc., 2001.

[2] E. Levitan and N. Merhav, “A competitive neyman-pearson approach to universal hypothesis testing with applications,” *IEEE Trans. on Info. Theory*, vol. 48, no. 8, pp. 2215–2229, Aug. 02.

[3] A. Sangwan, W.P. Zhu, and M.O. Ahmad, “Improved voice activity detection via contextual information and noise suppression,” in *IEEE Intl. Symp. on Circuits and Systems (ISCAS 05)*, May 2005, pp. 868–871.

[4] S.B. Searle, *Matrix Algebra useful for Statistics*, New York: Wiley, 1982.

[5] Y. Ephraim and D. Malah, “Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator,” *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec 1984.