Proceedings of the Estonian Academy of Sciences,
2013, **62**, 2, 88–96
doi: 10.3176/proc.2013.2.02
Available online at www.eap.ee/proceedings

SIGNAL
PROCESSING

Detection of the consistency of two speech signals using pitch periods

Tõnu Trump

Department of Radio and Telecommunication Engineering, Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia; tonu.trump@lr.ttu.ee

Received 12 June 2012, revised 31 December 2012, accepted 8 January 2013, available online 7 May 2013

Abstract. A robust detector is proposed for detection of the consistency of two speech signals. The detection is based on comparing the pitch periods of the speech signals. The distribution of the noise is assumed to be a mixture of Laplacian distribution, giving a sharp peak around the true value of the signal of interest and uniform distribution modelling the contributions of completely unknown noise. This type of noise appears in problems related to estimated pitch periods. We derive a robust detector for this noise model and analyse its performance.

Key words: signal processing, robust detection, Laplacian noise, uniform noise.

1. INTRODUCTION

In many applications like radar, sonar, biomedicine, telecommunications, seismology, etc., there arises a need to detect the presence or absence of a certain signal in a received waveform [1]. More recently detection has gained importance in the context of cognitive radio, where it is the key technology to discover so-called spectral holes, i.e. the areas of the spectrum that are not used by licensed users and which are hence available for opportunistic usage [2,3]. In this application the goal of detection is to discover the absence of signals in a given spectral band.

It is common to derive signal-processing algorithms using the additive white Gaussian noise assumption. In many applications this is the proper noise model because the noise is due to many additive elementary reasons, and in force of the central limit theorem the aggregated effect of them appears to be Gaussian. At the same time the Gaussianity is only an approximation to the reality and in many cases the actual noise has in fact an impulsive nature [4,5], i.e. the tails of the actual distribution are longer than predicted by the Gaussian curve. Impulsiveness of the noise may considerably degrade the performance of a system designed with the Gaussian assumption.

In this paper we model the noise as a mixture of Laplace and uniform processes. The Laplacian component gives a sharp peak in the distribution. This noise component is present only together with the signal and one can see it as an uncertainty of the value taken by the signal. The uniform component over the range of allowable values indicates that we are completely clueless about the possible value taken by the remaining noise component. This kind of model has proven itself useful to model pitch periods of speech if the observations have been disturbed for some reason. In [6] the model has been used to design a measurement method for room reverberation times from the speech signal. In [7] the model has successfully been used to model pitch periods estimated from noisy speech. The model is applicable to the difference between the pitch periods of clean and distorted speech. The Laplacian component represents the narrow spread around the true value due to deviations because of noise or reverberation and the uniform component is to count the situation where the estimated pitch is due to the noise and has nothing to do with the pitch of the original speech.

In [8] the model has been used for pitch periods returned from a mobile phone as echo. Echo is a phenomenon where part of the sound energy transmitted to the receiver reflects back to the sender. In telephony it usually happens because of acoustic coupling between the receiver's loudspeaker and microphone or because of reflections of signals at the impedance mismatches in the analogue parts of the telephony system. In mobile phones one has to deal with acoustic echoes, i.e. the signal played in the phone's loudspeaker can be picked up by the microphone of the same mobile phone.

People are used to the echoes that surround us in everyday life due to, e.g., reflections of our speech from the walls of rooms where we are located. Those echoes arrive with a relatively short delay (in the order of milliseconds) and are, as a rule, attenuated. In a modern telephone system, on the other hand, the echoes may return with a delay that is not natural for human beings. The main reason for delay in those systems is signal processing like speech coding and interleaving. For example, in a PSTN to GSM telephone call the one-way transmission delay is around 100 ms, making the echo to return after 200 ms. The echo that returns with this long delay is very unnatural to a human being and makes talking very difficult. Therefore the echo needs to be removed.

Ideally the mobile terminals should handle their own echoes in such a way that no echo is transmitted back to the telephony system. Even though many of the mobile phones currently in use are able to handle their echoes properly, there are still models that do not. International Telecommunication Union – Telecommunication Standardization Sector (ITU-T) has recognized this problem and has recently consented the Recommendation G.160, "Voice Enhancement Devices" that addresses these issues [9]. Following this standard, we concentrate on the scenario where the mobile echo control device is located in the telephone system as opposed to the phone itself, as foreseen by [9]. The echo path is in this case rather non-linear as there are two speech codec pairs and radio channels in the echo path.

The pitch is the parameter of coded speech that can best cope with the non-linearities of the echo path [8]. As many of nowadays mobile phones do not produce any echo, the first step of echo removal has to be detection of the presence of an echo in a given call. In this application the model is applied to the difference of pitch periods of transmitted and received speech signals. The Laplacian component models the small deviations of the estimated pitch period in the received signal due to noise and estimation inaccuracies and the uniform component is to account for the case when the received pitch is formed on the basis of the signals other than the one transmitted towards the mobile.

The uniform distribution has in all cases natural limits as the pitch of human speech can only take limited values [10]. In this paper we investigate a detector based on this noise model and show that the resulting detector is in fact a robust test in the sense of [11]. Note that a robust detector for usage in a mixture of Gaussian and uniform distributions has been proposed in [12].

In conclusion, we can say that in several applications there is a need to estimate the pitch period of a distorted speech signal. Because of the distortion the estimate has a certain error which has a narrow distribution centred around the true pitch value. Speech signals are non-stationary processes, the power of which varies largely in time and therefore occasionally the disturbance, not speech of interest, determines the estimated value of the pitch. This phenomenon has a wide flat distribution function. It has been argued in [6–8] that a mixture model that comprises Laplacian distribution and uniform distribution fits well with the distribution function of the pitch estimated in the disturbed environment. In this paper we investigate a detection problem where we wish to determine whether an estimated pitch is the same as the known one. The first step of this detection is finding the difference of the two pitches. Subtracting the two pitches from each other shifts the mean of the Laplacian component to zero.

The contribution of this paper is that we will derive the detector for verifying the consistency of two speech signals. The detector will use a sum of limited differences of the pitches of two speech signals as the detection variable. The exact limitation level follows from the derivation. After that we will provide an asymptotic analysis of the detector and derive its probability of detection and probability of false alarm in case of a large number of input samples.

The italic and bold face letters will be used for scalars and column vectors, respectively. The operator $E[\cdot]$ denotes mathematical expectation.

2. DERIVATION

We start our derivation from defining a variable x that denotes the difference of two pitch periods, $x = T_{\text{inp}} - T_{\text{ref}}$, the correct and the estimated or the transmitted and the received one, and forming two hypotheses. The hypothesis H_1 is the hypothesis of the component of interest being present in the received signal and the null hypothesis H_0 is the hypothesis that it is not.

Under the hypothesis H_1 the probability density function of x is given by

$$p(x \mid H_1) = \begin{cases} \alpha \max\left[\frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right), \frac{\beta}{b-a}\right], & a < x < b \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where the constant β is a design parameter that can be used to weight the Laplace and uniform components and σ is the parameter of the Laplace distribution. The variables a and b give the limits of the uniform distribution. Note that we have limited the Laplace distribution to lie in the interval $[a, b]$. This is, however, usually the case in practical systems where the dynamical range of the input signals is normally limited. The parameter α is a normalization constant that is selected to ensure the probability density function to integrate to unity. Solving

$$\int_a^b p(x)dx = 1 \tag{2}$$

for α, we obtain

$$\alpha = \frac{b-a}{2\sigma\beta\left(\ln\frac{2\sigma\beta}{b-a} - 1\right) + (1+\beta)(b-a)}. \tag{3}$$

The probability density function (1) can be rewritten in a more convenient form for further derivation as

$$p(x \mid H_1) = \begin{cases} \frac{\alpha}{2\sigma}\exp\left[-\frac{\min\left(|x|, -\sigma\ln\frac{2\sigma\beta}{b-a}\right)}{\sigma}\right], & a < x < b \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Under the hypothesis H_0, when the component of interest with its Laplacian uncertainty is missing, the distribution of x is assumed to be uniform within the interval $[a, b]$,

$$p(x \mid H_0) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Suppose that we have made N observations of the variable x and we have collected these observations into a vector $\mathbf{x}$. Also assume that the noise samples at different time instances are statistically independent of each other. Then the joint probability density function is a product of the individual probability densities

$$p(\mathbf{x} \mid H_k) = \prod_{n=1}^{N} p(x_n \mid H_k), \quad k = 0, 1. \tag{6}$$

The likelihood ratio test [1,13] for the hypotheses mentioned above reads

$$\Lambda(\mathbf{x}) = \frac{\prod_{n=1}^{N} \frac{\alpha}{2\sigma}\exp\left[-\frac{\min\left(|x_n|, -\sigma\ln\frac{2\sigma\beta}{b-a}\right)}{\sigma}\right]}{\prod_{n=1}^{N}\frac{1}{b-a}}. \tag{7}$$

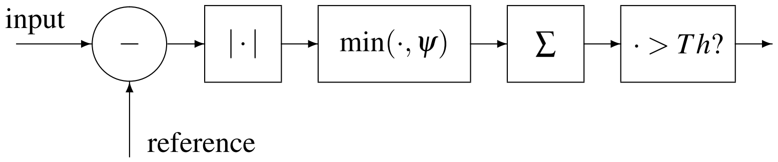

Fig. 1. Structure of the detector.

Taking the logarithm of both sides of (7) and simplifying, we readily obtain

$$\frac{\sigma}{N}\ln\Lambda(\mathbf{x}) = -\frac{1}{N}\sum_{n=1}^{N}\min\left(|x_n|, -\sigma\ln\frac{2\sigma\beta}{b-a}\right) - \sigma\left[\ln\frac{2\sigma}{\alpha} - \ln(b-a)\right]. \tag{8}$$

Let us now introduce notations

$$\psi = -\sigma\ln\frac{2\sigma\beta}{b-a} \tag{9}$$

and

$$\rho = \sigma\left[\ln(b-a) - \ln\frac{2\sigma}{\alpha}\right]. \tag{10}$$

Then we are left with

$$\frac{\sigma}{N}\ln\Lambda(\mathbf{x}) = -\frac{1}{N}\sum_{n=1}^{N}\min\left(|x_n|, \psi\right) + \rho. \tag{11}$$

The detector thus needs to compute the absolute value of each observed variable x_n, saturate the result at ψ, compute the sample average of the variables obtained in this way, and compare the result with the threshold

$$Th = \rho - \frac{\sigma}{N}\ln\Lambda(\mathbf{x}). \tag{12}$$

If the decision variable is bigger than the threshold, the hypothesis H_1 is decided, otherwise the decision is made in favour of H_0.

The structure of the detector is depicted in Fig. 1.

3. PERFORMANCE ANALYSIS

In this section we perform an asymptotic analysis as $N \to \infty$ of the detector. We start from noting that the detector algorithm includes a memoryless non-linearity

$$y = h(x) = \min(|x|, \psi). \tag{13}$$

The probability density function of the output of $h(x)$ is given by [14]

$$p_y(y) = \sum_{i=1}^{2}\left.\frac{p_x(x)}{\frac{dy}{dx}}\right|_{x=x_i=h_i^{-1}(y)}, \tag{14}$$

where $p_x(x)$ is the probability density function of the input.

Let us first assume that H_1 is true. In this case we can observe from (1) that the Laplace component is replaced by the uniform component in the distribution function at

$$|x| = -\sigma\ln\frac{2\sigma\beta}{b-a} = \psi.$$

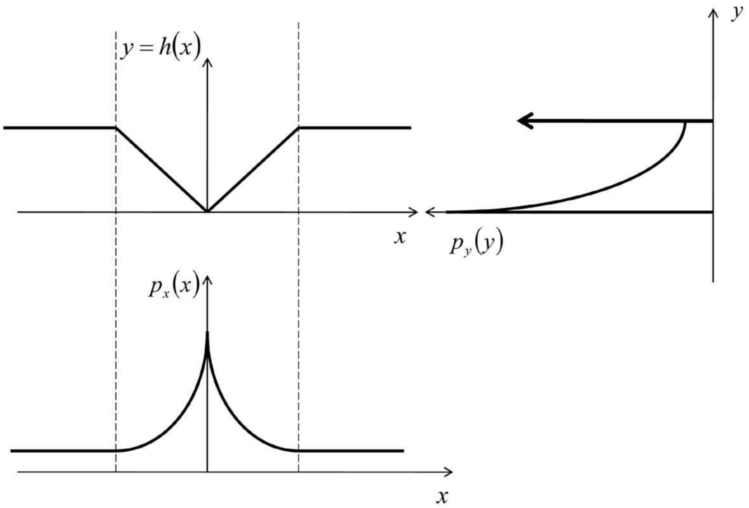

Fig. 2. Input–output characteristic of the non-linearity.

In addition, we have from (13) that the output of the non-linearity is saturated precisely at ψ. The probability density function of the output is therefore given by

$$p_y(y \mid H_1) = \frac{\alpha}{\sigma} \exp\left(-\frac{y}{\sigma}\right) [u(0) - u(\psi)] + \alpha\beta \frac{b-a-2\psi}{b-a} \delta(y-\psi), \tag{15}$$

where $u(\cdot)$ denotes the unit step function and $\delta(\cdot)$ is the Dirac delta function. The transformation is illustrated in Fig. 2. The mathematical expectation of the output signal of the non-linearity is

$$E[y \mid H_1] = \alpha \left[\sigma - (\sigma+\psi) \exp\left(-\frac{\psi}{\sigma}\right) + \beta\psi \frac{b-a-2\psi}{b-a}\right]. \tag{16}$$

The second moment equals to

$$E[y^2 \mid H_1] = \alpha \left[2\sigma^2 - \left(\psi^2 + 2\psi\sigma + 2\sigma^2\right) \exp\left(-\frac{\psi}{\sigma}\right) - \beta\psi^2 \frac{b-a-2\psi}{b-a}\right]. \tag{17}$$

The variance is consequently

$$\sigma^2_{H_1} = E[y^2 \mid H_1] - E^2[y \mid H_1]. \tag{18}$$

If the hypothesis H_0 is true, the probability density function at the output of the non-linearity will be

$$p_y(y \mid H_0) = \frac{2}{b-a} [u(0) - u(\psi)] + \frac{b-a-2\psi}{b-a} \delta(y-\psi). \tag{19}$$

The mean is in this case

$$E[y \mid H_0] = \psi - \frac{\psi^2}{b-a} \tag{20}$$

and the second moment equals to

$$E[y^2 \mid H_0] = \frac{2}{3} \frac{\psi^2}{b-a} + \frac{b-a-2\psi}{b-a} \psi^2. \tag{21}$$

The variance will be

$$\sigma^2_{H_0} = E[y^2 \mid H_0] - E^2[y \mid H_0]. \tag{22}$$

Let us now note that according to (11), the detector computes a sample average of N i.i.d. random variables that appear at the output of the non-linearity $y = h(x)$. According to the central limit theorem[14], the distribution of such a sum approaches Gaussian with mean $E[y \mid H_i]$ and variance $\frac{\sigma^2_{H_i}}{N}$, $i = 0, 1$ when N increases, independent of the shape of the original distribution. We can therefore evaluate the probability of correct detection as

$$\begin{aligned} P_{\mathrm{D}} &= \int_{-\infty}^{Th} p_y(y \mid H_1) dy \\ &= \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_{H_1}} \int_{-\infty}^{Th} \exp\left(-\frac{(y - E[y \mid H_1])^2 N}{2\sigma^2_{H_1}} \right) dy \\ &= Q\left(\frac{E[y \mid H_1] - Th}{\sigma_{H_1}} \sqrt{N} \right), \end{aligned} \tag{23}$$

where $Th = \rho - \frac{\sigma}{N} \ln \Lambda(\mathbf{x})$, is the threshold used in the test and $Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-\frac{\lambda^2}{2}} d\lambda$. The probability of false alarm is correspondingly

$$\begin{aligned} P_{\mathrm{F}} &= \int_{-\infty}^{Th} p_y(y \mid H_0) dy \\ &= Q\left(\frac{E[y \mid H_0] - Th}{\sigma_{H_0}} \sqrt{N} \right). \end{aligned} \tag{24}$$

The Receiver Operating Characteristic (ROC), which is the probability of correct decision P_{D} as a function of the probability of false alarm P_{F}, is plotted in Fig. 3. The ROC shows what compromises between P_{D} and P_{F} are possible to achieve by selecting the threshold. As expected, the performance of the detector improves if the endpoints of the uniform distribution a and b move further apart. The parameters used to compute ROC in Fig. 3 are $N = 30$, $\sigma = 2$, and $\beta = 0.1$. The extent of the uniform density $b - a$ varied from 4 to 12. One can see that the ROC lines approach the upper left corner with increasing $b - a$.

It is interesting to note that the detector constitutes a robust test in terminology of [11]. Indeed, the decision algorithm is piecewise linear in signal samples and, in addition, the entries of the input signal are saturated at ψ. This means that no single entry, no matter how big it is, can influence the decision by more than ψ. We have thus been able to derive a likelihood ratio test that appears to be a robust detector.

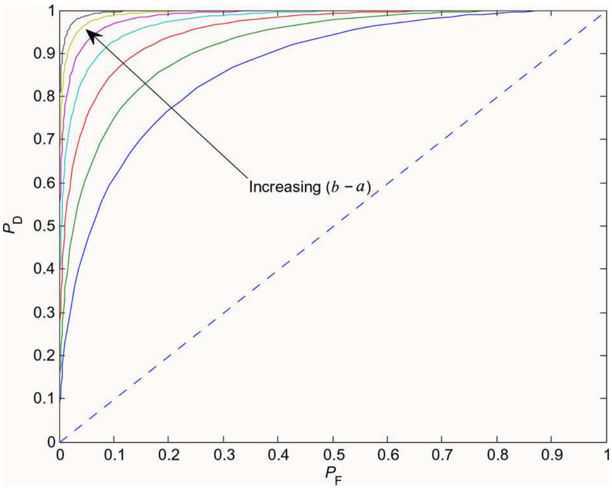

Fig. 3. Receiver operating characteristic for varying $b - a$.

4. SIMULATION RESULTS

In [8] it was demonstrated that the probability density function of pitch estimation errors follows indeed the Laplacian distribution as assumed in the previous sections. The same result has been reported in [6] and [7]. An example of this is given below [8]. In this example a 2 min long speech file that includes both male and female voices at various levels was first coded with the AMR12.2 kbit/s mode codec and then decoded. Then a simple echo path model consisting either of a single reflection or the IRS filter (ITU-T G.191) was applied to the signal and the signal was coded again. Echo return loss (ERL) was varied between 30 and 40 dB. The estimated pitch was registered from both codecs and compared. The pitch estimates were used only if the short-term power at the input of the first codec was above -40 dBm0 for the particular frame. A typical example of pitch differences is shown in Fig. 4. The upper plot shows the histogram of pitch estimation errors. The histogram has long tails ranging from -125 to 125 (which are the limiting values for differences between two pitch periods in the AMR codec) and a narrow peak can be observed around zero. The lower plot shows the Laplace probability density function fitted to the middle part of the histograms. One can see that the Laplacian curve matches the histogram rather well.

The probability of detection as a function of the threshold location is shown in Fig. 5. The parameters used in the experiment are $a = 50$, $b = -50$, $\sigma = 1$, and $\beta = 0.01$. This choice results in $\psi = 8.517$, $\rho = 3.904$, and $\alpha = 0.992$. The dashed–dotted line is the theoretical result computed using (23) in the case $N = 5$ and the hexagrams are the corresponding simulation results. Likewise the dashed line and circles correspond to $N = 20$ and the solid line and the plus signs correspond to $N = 50$. One can see that already with $N = 5$ there is a weak but existing resemblance between the theoretical and the simulation results. The correspondence improves with increasing N and if $N = 20$, the theoretical and simulation results match each other better. Finally, with $N = 50$ the theoretical and simulation curves are rather close to each other.

The probability of false alarm is shown in Fig. 6. Here we see that the theoretical and simulation results match each other rather well if $N = 50$. In case of $N = 20$ some differences start to appear between the theory and the simulations. In case of $N = 5$ the correspondence has become rather weak. The basic reason for the discrepancy between the theoretical and simulation results with small N is that if the threshold is

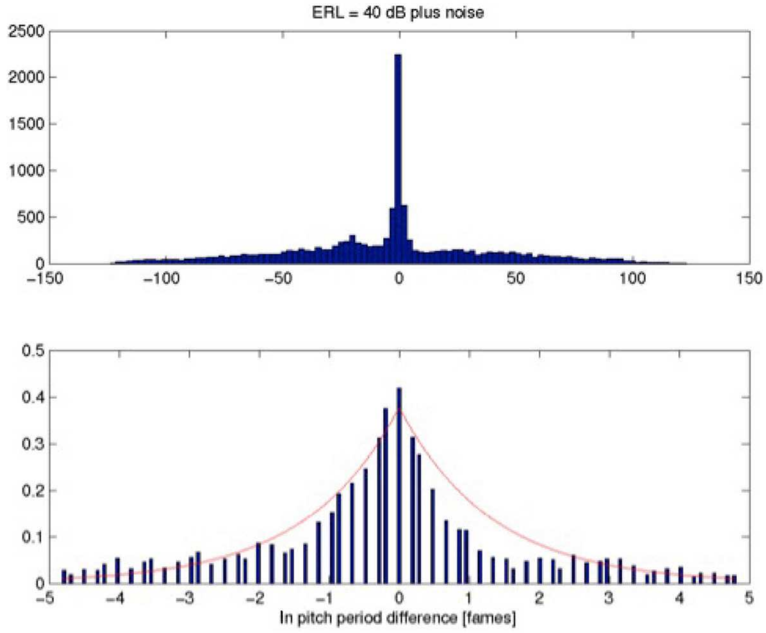

Fig. 4. Histogram of pitch estimation errors [8]. Echo path: single reflection and IRS filter, ERL = 40 dB. Near end noise at −60 dBm0.

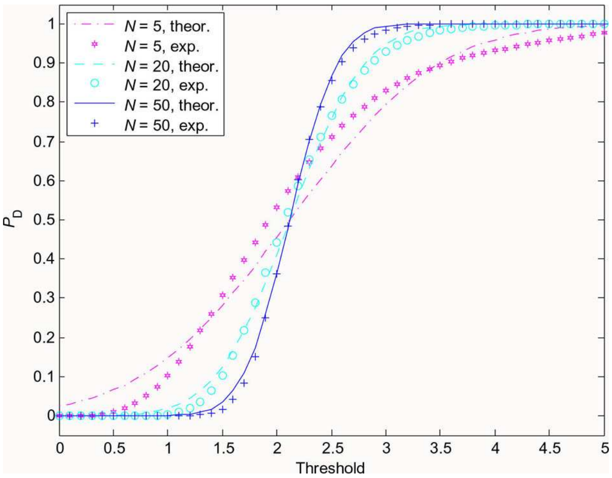

Fig. 5. Probability of detection.

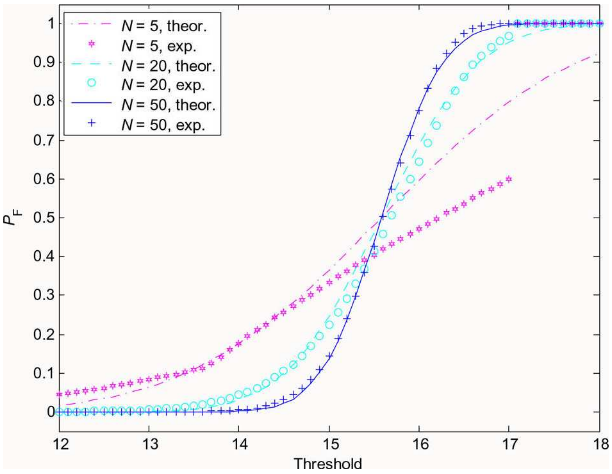

Fig. 6. Probability of false alarm.

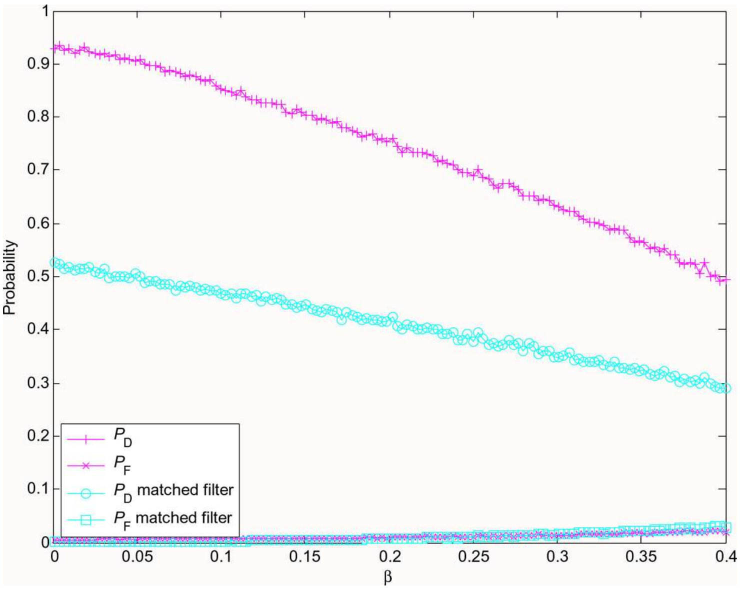

Fig. 7. Comparison of the proposed detector with the matched filter detector.

greater than ψ, the practical detector always decides in favour of the hypothesis H_1. This is because the measurements are saturated at ψ. The effect can be observed as discontinuity of the curves corresponding to a small number of observations N.

Comparison of the proposed detector with the classical matched filter detector [1], which is optimal in the case of Gaussian noise, is shown in Fig. 7. We have used $\sigma = 2.5$, $a = -b = 10$, and $N = 20$ in this experiment. Both the probability of detection and probability of false alarm are shown in the figure for $\beta = 0.01$ to 0.4. It can be seen that the proposed detector outperforms the matched filter detector in this situation.

5. CONCLUSIONS

We investigated a detector designed to detect a known signal in noise. The noise was modelled as consisting of Laplacian and uniform components if both the signal and noise are present. If no signal is present, the noise is modelled as distributed uniformly. Signal and noise models of this type naturally arise in the applications related to pitch periods if the observations have been disturbed for some reason. We derived a likelihood ratio test for this noise model and argued that the resulting algorithm is in fact a robust test. The performance of the detector was investigated analytically by deriving its probabilities of correct detection and false alarm. The analytical results were confirmed in our simulation study.

REFERENCES

1. Kay, S. M. *Statistical Signal Processing, Volume II, Detection Theory*. Prentice Hall, 1998.
2. Haykin, S., Thomson, D. J., and Reed, J. H. Spectrum sensing for cognitive radio. *Proc. IEEE*, 2009, **97**, 849–877.
3. Quan, Z., Poor, H. V., and Sayed, A. H. Collaborative wideband sensing for cognitive radios. *IEEE Signal Process. Mag.*, 2008, **25**, 60–73.
4. Pham, D., Zoubir, A., Bricic, R., and Leung, Y. A nonlinear m-estimation approach to robust asynchronous multiuser detection in non-Gaussian noise. *IEEE Trans. Signal Process.*, 2007, **55**, 1624–1633.
5. Wang, X. and Poor, H. V. Robust multiuser detection in non-Gaussian noise. *IEEE Trans. Signal Process.*, 1999, **47**, 289–305.
6. Wu, M., Wang, D., and Brown, G. J. A multipath tracking algorithm for noisy speech. *IEEE Trans. Speech Audio Process.*, 2003, **11**, 229–241.
7. Wu, M. and Wang, D. A pitch-based method for the estimation of short reverberation time. *Acta Acustica united with Acustica*, 2006, **92**, 337–339.
8. Trump, T. Detection of echo generated in mobile phones using pitch distance. In *Proc. 16th European Signal Processing Conference (EUSIPCO)*. 2008.
9. ITU-T Recommendation G.160. *Voice Enhancement Devices*. ITU-T, 2008.
10. Ritsma, R. J. Frequencies dominant in the perception of the pitch of complex sounds. *J. Acoust. Soc. Amer.*, 1967, **42**, 191–198.
11. Huber, P. J. *Robust Statistics*. John Wiley and Sons, 2004.
12. Trump, T. A robust detector for impulsive noise environment. In *Forty-First Asilomar Conf. on Signals, Systems, and Computers*. 2007, 730–734.
13. Van Trees, H. L. *Detection, Estimation and Modulation Theory*. John Wiley and Sons, 1968.
14. Papoulis, A. and Pillai, S. U. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 2002.

Robustne detektor ühtlase jaotusega müra tarvis

Tõnu Trump

On uuritud robustset detektorit, mis on kasutatav kõnesignaalide kokkulangevuse tuvastamiseks. Detektor põhineb helikõrguste võrdlusel. Müra jaotustihedusfunktsioon on kombinatsioon Laplace'i jaotusest, mis tekitab terava tõenäosusmassi signaali tegeliku väärtuse lähedal, ja ühtlasest jaotusest, mis kirjeldab täiesti tundmatut mürakomponenti. Selline müra tekib ülesannetes, mis on seotud helikõrguste hindamisega. Artiklis on tuletatud robustne detektor nimetatud müramudeli tarvis ja analüüsitud selle omadusi.