# A Study of Statistical Pattern Verification

NUGGEHALLY S. JAYANT, MEMBER, IEEE

**Abstract—A decision theoretic approach to the problem of verifying that a given observation "belongs" to a claimed pattern is examined. The problem presupposes a nonzero probability of impersonations, and verifier actions are ternary in general: a claimed identity is either 1) accepted or 2) rejected, or it is 3) "blanked" when the verifier refuses to take either of the actions 1) and 2). Corresponding costs-of-wrong decision will be typically different in applications like speaker verification. We characterize sample observations by normal distributions and provide explicit results for the analytically simple situation of "efficient" measurements whose interpattern variances are much greater than their intrapattern variances. In our specific model this will imply that the pdf of a measurement under the "impersonator" hypothesis is "diffuse" in relation to the pdf of the same measurement under any of several "true-pattern" hypotheses.**

**We demonstrate that the likelihood ratio function, in a typical verification task, possesses a finite upper bound and that this is related to a certain threshold effect for verifier decisions. We further mention the suitability of a minimax strategy for verification and compare its performance with those of optimal (minimum-cost) and maximum-likelihood procedures for the binary decision (accept-or-reject) problem. We provide numerical examples that illustrate the minimax performance as a function of 1) a parameter which reflects the peculiarity of the claimed pattern, and of 2) the measurement dimensionality. Finally, we indicate the application of sequential hypothesis testing to a variable-time-for-decision verifier designed to provide a prespecified performance, and note that in our model the expected-time-for-decision is typically greater when the true pattern is the claimed identity than when it is an impersonation.**

## I. Introduction

THE GENERAL problem of pattern verification is best explained in terms of a specific example. Consider, therefore, a voice banking situation where a customer's voice is granted the role of the conventional written signature. To effect a banking transaction, a customer could announce his identity over a telephone and follow with a preformatted utterance that would provide the basis for the verification of his identity. The actual verification task would involve the comparison of the distances, in some sense, between the preceding sample utterance and stored versions of the utterance that are typical of 1) the identity being verified and 2) the rest of the banking population and a consequent verifier action that can either accept or reject the customer, or refuse to take either of these two decisions —we will refer to the third action as a *blank* decision, borrowing from digital communication jargon. Obviously, the verifier problem is nontrivial only if one expects impersonators who attempt to use the banking facility by claiming a false identity. Such impersonators may be either registered users of the facility or totally alien impostors; reliable statistical data on the frequency of impersonations may be available or, more likely, not. In any case, each verification task will entail the complementary steps of 1) preprocessing of incoming data leading to feature extraction and 2) a subsequent statistically optimum decision that minimizes an appropriately defined verification cost. The general pattern verification problem is posed simply by replacing the speaker's claimed identity by a claimed pattern, the banking population by a pattern population, and by abstracting the verification- and stored-speaker features in the voice banking problem by corresponding sample measurements and prior distributions.

Most of the literature on pattern verification [1]–[3] has been addressed specifically to the speaker authentication problem ([4] is an exception), emphasizing the related preprocessing methods such as time-normalization [2] and segmentation [3] of utterances and the subsequent extraction of both speaker-relevant features such as pitch, formants [2], and spectral averages [3]. These studies have employed so-called "casual impostors," but work related to the tolerance of the proposed measures (features) to impersonations by professional mimics is also in progress [5].

The purpose of our paper is to emphasize the decision-making aspect of the verification process. To this end we postulate a simple, but hopefully useful, general model for feature distributions and use it to evaluate the performance of optimal and suboptimal verifiers. The likelihood ratio function generated by our statistical model has the following typical ramifications: 1) a certain threshold effect in the verification task, and 2) a shorter expected-time-for-decision—in a sequential verification procedure—under the impersonator hypothesis than under the true-pattern hypothesis. Many of the explicit results of this paper refer, however, to nonsequential, i.e., fixed-time-for-decision, schemes; and more particularly, to the binary decision (accept-or-reject) problem, and to the analytically simple case of "efficient" measures characterized by an inter-pattern variance which is much greater than the intra-pattern variance.

Let us emphasize, finally, that both the formalism of the next section and the analyses to follow really represent guidelines to verifier design and will very probably need, in the context of an operational verification system, modifications or refinements in directions that will be more obvious to a specific system designer.

## II. Formalism

### A. Verifier Inputs and Hypotheses

The input to the pattern verifier comprises a claimed identity $i$ and an accompanying sample signature $X$. (We begin with a scalar model for simplicity, but will extend our results to a multidimensional signature in Section XI.)

Manuscript received June 15, 1971: revised October 12, 1971.
The author is with Bell Telephone Laboratories, Inc., Murray Hill, N.J. 07974.

The "state of nature" is the true pattern $i_T$, and both $i$ and $i_T$ are members of a population $\Omega$ of $N$ patterns. We associate, with every claim $i$, the following hypotheses:

$$H_0:\ i_T = i \qquad \text{(true claim)}$$
$$H_1:\ i_T \neq i \qquad \text{(false claim).} \tag{1}$$

The probability density functions of the observation $X$, under the hypotheses (1), are denoted by $p(x \mid i)$ and $p(x \mid I)$, respectively. The "impersonator" distribution $p(x \mid I)$ will be assumed to be independent of the actual description of $\Omega$, or of its cardinality $N$. The underlying contention is that the population excluding pattern $i$ is either

1) an unknown population, characterized only by a model distribution; or
2) a known population that is sufficiently large for the composite distribution to be approximated by a simple function.

### B. Verifier Decisions

For every input $(x, i)$ possible decisions $D$ of the verifier are ternary in general:

$$\begin{aligned} D_A&:\ \text{ACCEPT } H_0\\ D_R&:\ \text{ACCEPT } H_1\\ D_B&:\ \text{DO NOT DECIDE.} \end{aligned} \tag{2}$$

Associated with each of these actions is a verification failure:

FALSE ACCEPTANCE: taking $D_A$ when $H_1$ is true
FALSE REJECTION: taking $D_R$ when $H_0$ is true
BLANKING: taking $D_B$ when either $H_0$ or $H_1$ is true. (3)

Resulting penalties are four-fold (and $i$-dependent) in general and are denoted by

$$\begin{aligned} \mu_{FA}&:\ \text{cost of a false acceptance}\\ \mu_{FR}&:\ \text{cost of a false rejection}\\ \mu_{TB}&:\ \text{cost of blanking a true claim}\\ \mu_{FB}&:\ \text{cost of blanking a false claim.} \end{aligned} \tag{4}$$

### C. Probabilities of True and False Claims; the Impersonation Probability

Assuming that a claim $i$ has been made, we will denote the conditional probabilities of a true and false claim by $p(i)$ and $p(I)$, respectively. We can accordingly define an impersonation probability $\theta_i$—the probability, conditional to a claim $i$, that the claim is false:

$$\theta_i = \frac{p(I)}{p(I) + p(i)} = p(I). \tag{5}$$

It would appear that the value of $\theta_i$ should be considered unknown in typical applications. We will take this into account in the suboptimal verification strategies to be considered in Sections VIII and IX. The optimal verification scheme that is introduced in the next section, assumes, in contrast, a knowledge of the impersonation probability (5) and therefore represents the best possible verification procedure.

## III. Minimum-Cost Verification

Recall the possible verifier decisions $D_A$, $D_R$, and $D_B$ (3). A minimum-cost verification rule selects, for each observation $x$, that decision $D_Z$, $Z = A,R,B$, which minimizes a corresponding value $\mathscr{L}_Z$ of the "conditional loss," the loss conditional to $x$. The losses associated with decisions $D_A$ and $D_R$ are nonzero when the true state of nature is given by $H_1$ and $H_0$, respectively, while $D_B$ always results in a nonzero loss. Following (3) and (4), these three types of conditional loss can be expressed in the form[1]

$$\begin{aligned} \mathscr{L}_{FA} &= p(I \mid x)\cdot \mu_{FA}\\ \mathscr{L}_{FR} &= p(i \mid x)\cdot \mu_{FR}\\ \mathscr{L}_{B} &= p(i \mid x)\cdot \mu_{TB} + p(I \mid x)\cdot \mu_{FB}. \end{aligned} \tag{6}$$

Utilizing the identity

$$p(U \mid V) \equiv p(V \mid U) \cdot \frac{p(U)}{p(V)} \tag{7}$$

the *a posteriori* probabilities in (6) can be written in terms of corresponding *a priori* probabilities and pdf:

$$\begin{aligned} \mathscr{L}_{FA} &= \frac{p(I)\cdot p(x \mid I)\cdot \mu_{FA}}{p(x)}\\ \mathscr{L}_{FR} &= \frac{p(i)\cdot p(x \mid i)\cdot \mu_{FR}}{p(x)}\\ \mathscr{L}_{B} &= \frac{p(I)\cdot p(x \mid I)\cdot \mu_{FB}}{p(x)} + \frac{p(i)\cdot p(x \mid i)\cdot \mu_{TB}}{p(x)} \end{aligned} \tag{8}$$

where $p(x)$ is an (irrelevant) unconditional pdf.

The minimum-cost decision rule is now determined by seeking the smallest of the losses (8). Utilizing the notation

$$L(x) = \frac{p(x \mid i)}{p(x \mid I)} \tag{9}$$

the minimum-cost (or Bayes') decision rule is obtained in the form select $D_Y$ in favor of $D_Z$ iff

$$L(x) > L_{YZ},^{2} \qquad Y,Z = A,R,B$$

where

$$\begin{aligned} L_{AR} &= \frac{\mu_{FA}}{\mu_{FR}} \cdot \frac{p(I)}{p(i)}\\ L_{AB} &= \frac{\mu_{FA} - \mu_{FB}}{\mu_{TB}} \cdot \frac{p(I)}{p(i)}\\ L_{BR} &= \frac{\mu_{FB}}{\mu_{FR} - \mu_{TB}} \cdot \frac{p(I)}{p(i)}. \end{aligned} \tag{10}$$

Notice that the decision thresholds (10) are determined entirely by the costs $\mu$ (4) and the impersonation probability $\theta_i$ (5).

[1] The *a posteriori* probabilities $p(I \mid x)$ and $p(i \mid x)$ in (6) represent a concept that leads to a simple derivation of subsequent results (9) and (10). For a rigorous approach that is independent of the posterior probability concept, the interested reader is referred to Helstrom [6].

[2] A "greater than or equal to" condition can be employed alternatively, without violating the requirements of minimum-cost verification.

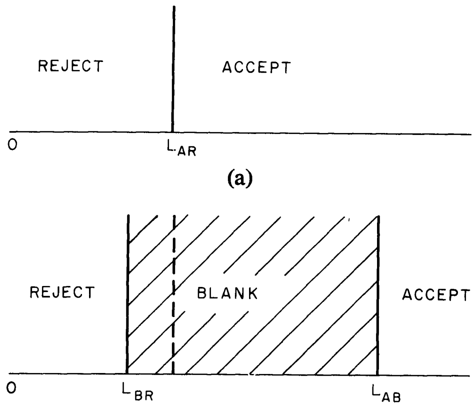


Fig. 1. Illustration of decision thresholds. (a) Binary decisions. (b) Ternary decisions.

TABLE I
ILLUSTRATION OF DECISION THRESHOLDS ($\theta_i = 0.5$)

| $\mu_{FA}$ | $\mu_{FR}$ | $\mu_{TB}$ | $\mu_{FB}$ | $L_{AR}$ | $L_{AB}$ | $L_{BR}$ |
|---|---|---|---|---|---|---|
| 10 | 100 | 1 | 1 | 0.10 | 9.00 | 0.01 |
| 100 | 100 | 100 | 100 | 1.00 | 0.00 | $\infty$ |
| 100 | 10 | 1 | 0 | 10.00 | 100.00 | 0.00 |

The likelihood ratio (9) is always positive. Therefore, in order to associate distinct nonzero segments on the $L(x)$ line for decisions $D_A$, $D_B$, and $D_R$ the following condition must be met (see Fig. 1):

$$\infty > L_{AB} > L_{AR} > L_{BR} > 0. \tag{11}$$

Table I lists values of $L_{YZ}$ for different cost assignments and for the illustrative characterization

$$p(i) = p(I) = 0.5. \tag{12}$$

Notice that the second and third rows in Table I do not satisfy (11) and represent degenerate examples where a blank and a reject decision, respectively, are ruled out (for all $L(x)$) because of the corresponding cost assignments.

## IV. The Gaussian Model

In the absence of indications to the contrary, we shall model the *a priori* pdf $p(x \mid i)$ (for the claimed identity) and $p(x \mid I)$ (for the rest of the population) by normal functions. We begin with unidimensional measurements for simplicity, but will extend our results to the multidimensional case in Section XI.

Furthermore, in order to be able to provide explicit results, we will assume in this paper that measurements are efficient in the sense that their interpattern variances are much greater than their intrapattern variances. This is not, however, a necessary attribute of a practical verification system if the number of measurements can be suitably large (Fig. 6); in Section XIII we will comment on the design and performance of a verifier that operates on the basis of a large number of inefficient measurements.

On the basis of the preceding discussion, we shall assume, for Sections IV–X, the following pdf as being illustrative:

$$p(x \mid I) = \eta(0,\sigma^2)$$
$$p(x \mid i) = \eta\left(m_i, \frac{\sigma^2}{S_i^2}\right), \qquad S_i \gg 1. \tag{13}$$

The zero mean for $p(x \mid I)$ is chosen without loss of generality, and the condition on $S_i$ in (13) implies, for the impostor distribution $p(x \mid I)$, a local flatness in the neighborhood of a given $m_i$.[3]

Notice also that a most typical individual pattern $k$ in the population will be characterized by a zerc value for $m_k$ and that (13) can be shown to imply the following distribution of pattern means:

$$p(m_k) \sim \eta(0,\sigma^2), \qquad k \in \Omega. \tag{14}$$

In what follows we will often omit the subscript $i$ for $m$ and $S$ for simplicity.

## V. The Likelihood Ratio Function

A likelihood ratio $L_V(x)$ for verification follows immediately from (9) and (13):

$$L_V(x) \sim S \cdot \exp\left[-\frac{(x-m)^2}{2\sigma^2/S^2}\right], \qquad S \gg 1. \tag{15}$$

It will be instructive to compare $L_V(x)$ with the likelihood ratio function in the following special case of the classical pattern identification problem; here, a basic discrimination task involves the likelihoods of two competing patterns $i$ and $j$. Assuming for illustration that

$$p(x \mid j) = \eta\left(0, \frac{\sigma^2}{S^2}\right) \tag{16}$$

we obtain the likelihood ratio $L_R(x)$:

$$L_R(x) = \frac{\eta(m,\sigma^2/S^2)}{\eta(0,\sigma^2/S^2)} = \exp\left[\frac{2mx - m^2}{2\sigma^2/S^2}\right]. \tag{17}$$

Fig. 2 sketches the functions (17) and (15). Notice that $L_R(x)$ is unbounded for large $x$, while $L_V(x)$ is strictly upper bounded; the ratio $L_V$ of the likelihoods of the contending hypotheses can never become arbitrarily large in verification.

In the remainder of this paper $L(x)$ will always refer to the verification problem, and the subscript $V$ will therefore be omitted.

## VI. The Threshold Effect in Pattern Verification

On the basis of the minimum-cost decision rule (10), the pattern verifier decides on action $D_Y$ in favor of action $D_Z$ if $L(x)$ exceeds the threshold $L_{YZ}$. This verifier action can be explicitly described by virtue of (15) and the use of some algebra in the following form: take $D_Y$ instead of

[3] A similar model, in the context of signal detection, has been considered by Pfeiffer [7]. In his formulation $H_0$ and $H_1$ refer to a radar target and a diffuse background.

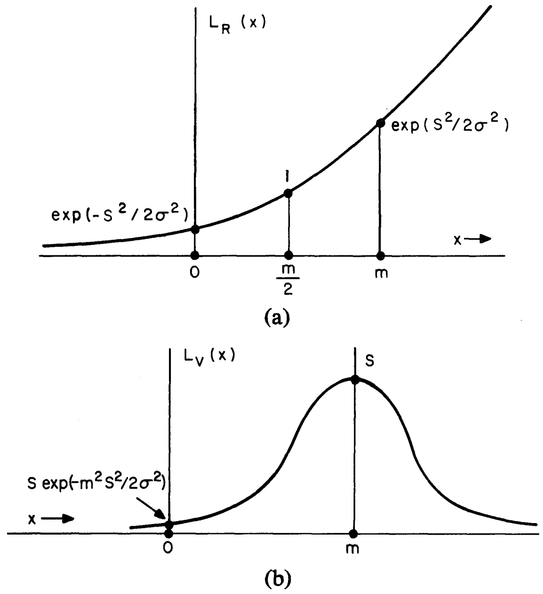


Fig. 2. Sketches of likelihood ratio function. (a) Recognition problem with $\sigma_i = \sigma_j$. (b) Verification problem with $\sigma_i \ll \sigma$.

$D_Z$ iff

$$\frac{S^2}{\sigma^2}(x - m)^2 < \ln \frac{S^2}{L_{YZ}{}^2}. \tag{18}$$

The left-hand side of (18) is obviously positive, while the right-hand side can be negative if $L_{YZ}$ is sufficiently large. In such a situation the decision rule would adopt $D_Z$ irrespective of the actual value of the observation $x$. Stated differently, the verification rule will be a meaningful function of the observation $x$ only if $L_{YZ}$ is below a threshold $L(T)$ which will render the right-hand side of (18) positive:

$$L(T) = S_i. \tag{19}$$

The subscript $i$ has been reintroduced in (19) to emphasize that the extent of the threshold effect is a function of the pattern pdf.

For an illustration, let us refer back to Fig. 1 and the first row of Table I. We conclude for this example, by virtue of (19), that a minimum cost verification procedure will unconditionally (i.e., for all $x$) rule out an accept decision $D_A$ if $S_i$ is less than 9. To elaborate: the impersonation probability is a significant 0.5, the cost of a false acceptance is 10 times that of blanking; and if a claimed pattern $i$ is to be accepted, nevertheless, in a minimum cost strategy, the measurement $x$ must possess a discrimination index $S_i$ that exceeds 9.

We should note at this stage that the notion of a blank decision $D_B$ is really relevant in a "variable-time-for-decision" strategy [8], [9], where every action $D_B$ signals the need for additional verification data. Mathematically, such data serve either to increase or decrease the likelihood ratio $L_V$ until it crosses some given threshold and leads to an unequivocal decision $D_A$ or $D_R$. One can also take into account the fact that verification cost is a monotonically increasing function of the observation time $t$ by allowing the blanking costs $\mu_{TB}$ and $\mu_{FB}$ to be suitably increasing functions of $t$. This will have the effect of progressively shrinking the width of the no-decision zone between $L_{BR}$ and $L_{AB}$ in Fig. 1. The sequential test will then be forced to terminate at that value of $t$ which equalizes $L_{AR}$, $L_{AB}$, and $L_{BR}$, as determined from (10).

The "sequential" and "truncated-sequential" strategies implied in the preceding discussion have been classically developed by Wald [9] in the form of the so-called sequential probability ratio test (SPRT) which has the characteristic property that it can be designed to yield a prespecified error performance. We will summarize the features of Wald's SPRT as applied to the verification problem in Section XIV. We shall restrict ourselves until that section to a nonsequential accept-or-reject verifier which, at the conclusion of a one-shot verification task, is forced to take either of the decisions $D_A$ or $D_R$; for this binary task, we will compare the performances of optimal (minimum-cost) and suboptimal (maximum-likelihood and minimax) strategies.

## VII. Performance of a Minimum-Cost Verifier

Let us rewrite the minimum-cost decision rule (18) as applied to the binary verification task as follows: accept $i$ iff

$$(x - m_i)^2 < \frac{\sigma^2}{S_i{}^2} \cdot \ln \frac{S_i{}^2}{L_{AR}{}^2} \tag{20}$$

where $L_{AR}$ (10) is given by

$$L_{AR} = \frac{\mu_{FA}}{\mu_{FR}} \cdot \frac{p(I)}{p(i)}. \tag{21}$$

We will simplify (20) using the notation

$$\Phi = \ln \frac{S^2}{L_{AR}{}^2} \tag{22}$$

so that the verification rule will be rewritten: accept $i$ iff

$$(x - m_i)^2 < \frac{\sigma^2}{S_i{}^2} \cdot \Phi. \tag{23}$$

Obviously, if $\Phi$ is negative, rejections will occur with probability 1 for all $x$. We will now evaluate the probabilities of two types of verification error.

### *False Rejections*

A false rejection will occur if

$$(x - m_i)^2 > \frac{\sigma^2}{S_i{}^2} \Phi \tag{24}$$

where $H_0$ is true. The probability of this event is given by

$$P_{FR} = \begin{cases} 1, & \Phi < 0 \\ \displaystyle\int_{m+\sigma\sqrt{\Phi}/S}^{\infty} p(x \mid i)\, dx \\ \quad + \displaystyle\int_{-\infty}^{m-\sigma\sqrt{\Phi}/S} p(x \mid i)\, dx, & \Phi \geq 0. \end{cases} \tag{25}$$

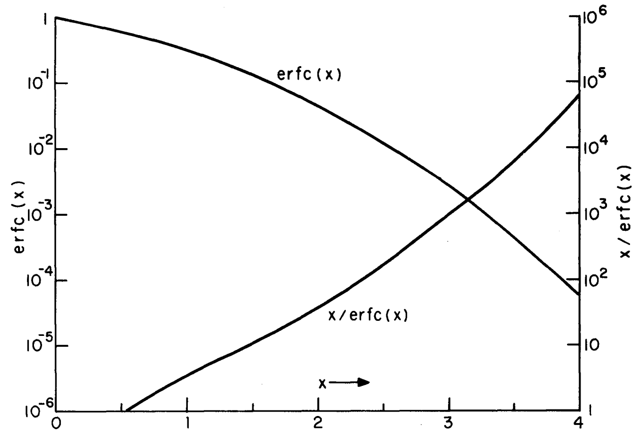

Fig. 3. [erfc $(x)$] and [$x$/erfc $(x)$] functions.

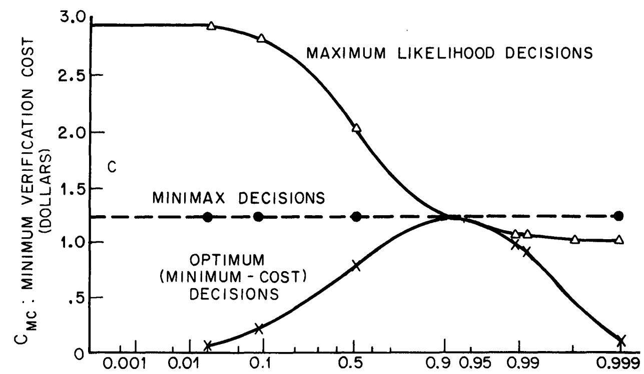

Fig. 4. Verification performance as function of probability of impersonation ($\mu_{FA}$ = \$100, $\mu_{FR}$ = \$10, $S$ = 10, $m = \sigma$).

The two integrals in (25) are equal because of the symmetry of $p(x \mid i)$ about $m_i$, and their value can be expressed in terms of the complementary error function (see Fig. 3):

$$\operatorname{erfc} x \triangleq \sqrt{\frac{2}{\pi}} \int_x^{\infty} \exp\left(\frac{-u^2}{2}\right) du. \tag{26}$$

(The function is sometimes defined so that erfc 0 = 0.5, instead of one.) Thus, after some algebra, we obtain

$$P_{FR} = \begin{cases} 1, & \Phi < 0 \\ \operatorname{erfc} \sqrt{\Phi}, & \Phi \geq 0. \end{cases} \tag{27}$$

The erfc $x$ function is closely related to the well tabulated error function erf $x$ and to the exceedence probability $Q(x^2 \mid n)$ of the $\chi^2$ distribution with $n$ degrees of freedom:

$$\operatorname{erfc} x = 1 - \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) = Q(x^2 \mid 1). \tag{28}$$

Anticipating the multidimensional case in Section XI, we will rewrite (27) in the more general form

$$P_{FR} = \begin{cases} 1, & \Phi < 0 \\ Q(\Phi \mid 1), & \Phi \geq 0. \end{cases} \tag{29}$$

*False Acceptances*

A false acceptance will occur if

$$(x - m_i)^2 < \frac{\sigma^2 \Phi}{S_i^2} \tag{30}$$

when $H_1$ is true. The probability of this event is given by the following:

$$P_{FA} = \begin{cases} 0, & \Phi < 0 \\ \displaystyle\int_{m_i - \sigma\sqrt{\Phi}/S_i}^{m_i + \sigma\sqrt{\Phi}/S_i} p(x \mid I)\, dx. \end{cases} \tag{31}$$

Our assumption that $S_i \gg 1$ implies a "local flatness" of $p(x \mid I)$ in the "small" neighborhood of $m_i$ as expressed by the limits of integration in (31). We can consequently approximate this integral by the product of the range of integration and the value of $p(x \mid I)$ at $x = m_i$, so that we have

$$P_{FA} = \begin{cases} 0, & \Phi < 0 \\ \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{m^2}{2\sigma^2}\right) \cdot \dfrac{2\sqrt{\Phi}}{S}, & \Phi \geq 0. \end{cases} \tag{32}$$

Again, we will anticipate the $n$-dimensional case of Section XI and rewrite (32) in the form

$$P_{FA} = \begin{cases} 0, & V_1 < 0 \\ p(m \mid I) \cdot V_1, & V_1 \geq 0 \end{cases} \tag{33}$$

where the volume $V_1$ refers to the range of integration in (31).

*Cost Per Decision*

Recall the impersonation probability $\theta_i$ (5), which is the probability, conditional to a claim $i$, that the claim is false. The average risk or cost per decision $C_{MC}$ of the minimum cost verifier is accordingly given by

$$C_{MC} = (1 - \theta_i) \cdot \mu_{FR} \cdot P_{FR} + \theta_i \cdot \mu_{FA} \cdot P_{FA} \tag{34}$$

where $P_{FR}$ and $P_{FA}$ are given by (27) and (32). Fig. 4 includes a plot of $C_{MC}$ as a function of $\theta_i$ for the following example:

$$\begin{aligned} S_i &= 10 \\ m_i &= \sigma \\ \mu_{FA} &= \$10 \\ \mu_{FR} &= \$100. \end{aligned} \tag{35}$$

Notice the unimodal nature of the $C_{MC}$ curve; verifications for the example of (35) turn out to be most unreliable at a $\theta$ value of 10/11. It is interesting to note that this worst value of $\theta$ corresponds (5), (10) to a value of unity for $L_{AR}$.

It should be clear at this stage that verification performance depends on the actual claim $i$ through the parameters $m_i$, $S_i$, and (possibly) $\theta_i$. We will defer additional treatment of this topic to Section X when we will illustrate a minimax verification cost as a function of pattern peculiarity given by

$$P_i = \frac{|m_i|}{\sigma}. \tag{36}$$

## VIII. Maximum Likelihood Verifications

A maximum likelihood verifier is characterized by definition by a value of unity for the parameter $L_{AR}$ (10). This is equivalent to equating the values of $p(i)$ and $p(I)$—this implies absence of any prior information on the impersonation probability $\theta$—and those of $\mu_{FA}$ and $\mu_{FR}$. The consequent decision-rule is given by the following: accept $i$ iff

$$\frac{p(x \mid i)}{p(x \mid I)} > 1. \tag{37}$$

The performance of this verification rule is simply obtained by setting $L_{AR}$ equal to unity in (22):

$$\Phi_{ML} = \ln S^2 \tag{38}$$

and by replacing $\Phi$ by $\Phi_{ML}$ in the performance equations (27), (32), and (34). Fig. 4 includes a plot of the cost per decision $C_{ML}$ of a maximum likelihood verifier as a function of $\theta_i$. The results refer to the example of (35). The suboptimality of the maximum likelihood rule is obvious from the figure, and it is interesting to note that the $C_{ML}$ curve meets the optimal $C_{MC}$ curve at the point of its maximum. (As noted earlier, this happens at the value of $\theta$ for which $L_{AR}$ (21) equals unity for the example of (35).)

## IX. Minimax Verifications

We will now discuss another suboptimal verification procedure which, unlike the scheme of the last section, takes into account actual values of $\mu_{FA}$ and $\mu_{FR}$ for the design of a decision threshold. However, as in the case of the maximum likelihood procedure, the minimax verifier to be described does not depend on any knowledge of the *a priori* probabilities $p(i)$ and $p(I)$. In fact, the minimax rule will result in a performance that is inherently independent of these probabilities.

Following the development in Section VII, we notice that the cost $C$ of a given verification rule is, in general, a function both of $\theta$ and $\Phi$. More significantly, for each value of $\Phi$ there exists in general a value of $\theta$ which results in a greatest cost per decision. We will denote this maximum cost by

$$C[\Phi,\theta(\Phi)_{\text{worst}}]. \tag{39}$$

A minimax rule is defined as the procedure which employs that value of the threshold $\Phi_{MM}$ for which the function (39) is the least:

$$C[\Phi_{MM},\theta(\Phi_{MM})_{\text{worst}}] < C[\Phi,\theta(\Phi)_{\text{worst}}], \quad \text{all } \Phi. \tag{40}$$

We will now derive an expression for $\Phi_{MM}$.

Let us rewrite (34) in the form

$$C(\Phi,\theta) = (1-\theta)\cdot\mu_{FR}\cdot P_{FR}(\Phi) + \theta\cdot\mu_{FA}\cdot P_{FA}(\Phi). \tag{41}$$

For a given threshold $\Phi$, it can be seen from (41) that the largest value of $C(\Phi)$ is given by

$$C_{\max} = \max\,[C(\Phi,0),C(\Phi,1)] \tag{42}$$

where

$$C(\Phi,0) = \mu_{FR}\cdot P_{FR}(\Phi) \tag{43}$$

$$C(\Phi,1) = \mu_{FA}\cdot P_{FA}(\Phi). \tag{44}$$

$P_{FR}(\Phi)$ is proportional to erfc $\sqrt{\Phi}$ (27), while $P_{FA}(\Phi)$ is proportional to $\sqrt{\Phi}$ (32). Thus $C(\Phi,0)$ is a monotonically decreasing function of $\Phi$, while $C(\Phi,1)$ is a monotonically increasing function of $\Phi$. A little reflection will therefore reveal that the minimum value of (42) is given simply by the point of intersection of $C(\Phi,0)$ and $C(\Phi,1)$ as functions of $\Phi$. This point of intersection will therefore define $\Phi_{MM}$:

$$C(\Phi_{MM},0) = C(\Phi_{MM},1). \tag{45}$$

By virtue of (43) and (44), (45) can be rewritten:

$$\mu_{FR}\cdot P_{FR}(\Phi_{MM}) = \mu_{FA}\cdot P_{FA}(\Phi_{MM}). \tag{46}$$

Finally, by inserting the results (27) and (32) in (46), we obtain the following implicit formula for the minimax threshold $\Phi_{MM}$:

$$\frac{\sqrt{\Phi_{MM}}}{\operatorname{erfc}\sqrt{\Phi_{MM}}} = \frac{\mu_{FR}}{\mu_{FA}}\cdot S\sqrt{\frac{\pi}{2}}\cdot\exp\left(\frac{m^2}{2\sigma^2}\right). \tag{47}$$

The $[x/\operatorname{erfc} x]$ function is plotted in Fig. 3.

From (41) and (46), the minimax verification cost $C_{MM}$ is given by

$$\begin{aligned} C_{MM} &= (1-\theta)\mu_{FR}\cdot P_{FR}(\Phi_{MM}) + \theta\cdot\mu_{FR}\cdot P_{FR}(\Phi_{MM}) \\ &= \mu_{FR}\cdot P_{FR}(\Phi_{MM}) \\ &= \mu_{FA}\cdot P_{FA}(\Phi_{MM}). \end{aligned} \tag{48}$$

Notice that $C_{MM}$ is inherently independent of $\theta$ and *ipso facto* is never exceeded by any unfavorable values thereof. This performance invariance is a very desirable property, but it also implies that the minimax rule can be very conservative (in that, for certain extreme values of $\theta$, even the use of a crude estimate of $\theta$ in a minimum cost strategy can provide a better performance than the minimax strategy). The contention is well demonstrated in Fig. 4, which compares the performance of the minimax verifier with those of the minimum cost and maximum likelihood rules for the example of (35).

Notice finally that if $\mu_{FA} = \mu_{FR}$, the minimax verifier is the same as that which equalizes the probabilities of the two types of error $P_{FA}$ and $P_{FR}$, and this equal-error criterion is often employed in verification research [2], [3].

## X. Minimax Performance as a Function of Pattern Peculiarity

Continuing our discussion in the last paragraph of Section VII, let us recall the parameter $P_i$ which reflects the peculiarity of a claimed pattern $i$:

$$P_i = \frac{|m_i|}{\sigma}. \tag{49}$$

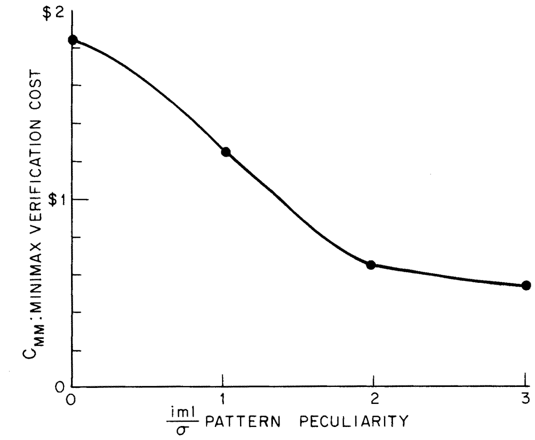


Fig. 5. Minimax verification performance as function of pattern peculiarity ($\sigma_{FA}$ = \$100, $\sigma_{FR}$ = \$10, $S$ = 10).

Obviously, a pattern whose mean coincides with that of the population distribution (14), i.e., whose $m_i = 0$, is the least peculiar individual pattern.

One can evaluate on the basis of (34), (38), and (48) the costs $C_{MC}$, $C_{ML}$, and $C_{MM}$ as functions of $m_i$ and $S_i$, and in conjunction with the distribution (14), one can study the average performance, over all claimed patterns $i$, of a given verification scheme. These results, however, do not permit elegant mathematical expression, in general. We will therefore content ourselves with the following numerical example. Fig. 5 plots the minimax cost $C_{MM}$ as a function of $P_i$ (49) for the characterization

$$\begin{aligned} S_i &= 10, \qquad \text{all } i \\ \mu_{FA} &= \$10 \\ \mu_{FR} &= \$100. \end{aligned} \tag{50}$$

Notice the obvious improvement in verification performance as the claimed pattern becomes more and more nontypical, i.e., when $P_i$ increases. By virtue of (14), it can be shown that the results of Fig. 5 represent an average $C_{MM}$, over all claimed patterns $i$, which is equal to the value of $C_{MM}$, for $P_i \sim 0.75$.

## XI. The Multidimensional Case

The results developed so far in this paper are extendable, in a straightforward manner, to the case of a multidimensional measurement $x$. We will only list salient results, therefore, and number them to show their correspondence with earlier expressions for the one-dimensional case. The interested reader will find some relevant details in Pfeiffer's paper [7]. In what follows, $n$ refers to the dimensionality of a normally distributed measurement vector and superscripts refer to an individual dimension. The observation pdf are

$$p(x \mid I) = \frac{1}{(2\pi)^{n/2}\sqrt{|G_I|}} \exp\left[-\tfrac{1}{2}x^t G_I^{-1} x\right]$$

$$p(x \mid i) = \frac{1}{(2\pi)^{n/2}\sqrt{|G_i|}} \exp\left[-\tfrac{1}{2}(x - m_i)^t G_i^{-1}(x - m_i)\right] \tag{13a}$$

where $G$ is a covariance matrix:

$$G: \|g^{de}\| = \|\langle (x^d - m^d)(x^e - m^e)\rangle\|, \qquad d,e = 1,2,\cdots,n, \qquad m = 0 \text{ in } G_I.$$

The likelihood ratio is

$$L_V(x) = \frac{p(x \mid i)}{p(x \mid I)} = \sqrt{\frac{|G_I|}{|G_i|}} \cdot \exp\left[-\tfrac{1}{2}(x - m_i)^t G_i^{-1}(x - m_i)\right] \tag{15A}$$

and the verifier action is: take $D_Y$ instead of $D_Z$ iff

$$(x - m_i)^t G_i^{-1}(x - m_i) < \ln\left[\frac{|G_I|}{|G_i|} \cdot \frac{1}{L_{YZ}^2}\right]. \tag{18A}$$

For the binary (accept–reject) verifier, we will use the notation

$$\Phi = \ln \frac{|G_I|}{|G_i|} \cdot \frac{1}{L_{AR}} \tag{22A}$$

and the minimum cost verification rule will be: accept $i$ iff

$$(x - m_i)^t G_i^{-1}(x - m_i) < \Phi. \tag{23A}$$

The probability of a false rejection involves the exceedence probability $Q$ of the well-tabulated $\chi^2$ distribution with $n$ degrees of freedom and is written, after (29), in the following form:

$$P_{FR} = \begin{cases} 1, & \Phi < 0 \\ Q(\Phi \mid n), & \Phi \geq 0 \end{cases} \tag{29A}$$

and the probability of a false acceptance is

$$P_{FA} = \begin{cases} 0, & V_n < 0 \\ p(m_i \mid I) \cdot V_n, & V_n \geq 0 \end{cases} \tag{33A}$$

where the detection volume $V_n$ surrounding $m_i$ is given by [7]:

$$V_M = \pi^{n/2}\Phi^{n/2}\left[\Gamma\left(\frac{n}{2} + 1\right)\right]^{-1} \cdot |G_i|^{1/2}. \tag{51}$$

$\Gamma(u)$ is the well-known gamma function, which satisfies the equalities

$$\begin{aligned} \Gamma(u) &= (u - 1)\cdot\Gamma(u - 1) \\ &= (u - 1)!, \qquad \text{for } u = 1,2,3,\cdots \\ \Gamma(\tfrac{3}{2}) &= \tfrac{1}{2}\sqrt{\pi}. \end{aligned} \tag{52}$$

The minimum verification cost $C_{MC}$ is given by (34), (29A), (33A) and (51), while the maximum likelihood cost $C_{ML}$ and the minimax cost $C_{MM}$ are obtained by replacing $\Phi$ in (22A) by $\Phi_{ML}$ and $\Phi_{MM}$, respectively, where

$$\Phi_{ML} = \ln \frac{|G_I|}{|G_i|} \tag{38A}$$

$$Q(\Phi_{MM} \mid n) = \left(\frac{\Phi_{MM}}{2}\right)^{n/2} \frac{|G_i|}{|G_I|} \cdot \frac{1}{\Gamma(n/2 + 1)} \frac{\mu_{FA}}{\mu_{FR}}. \tag{47A}$$

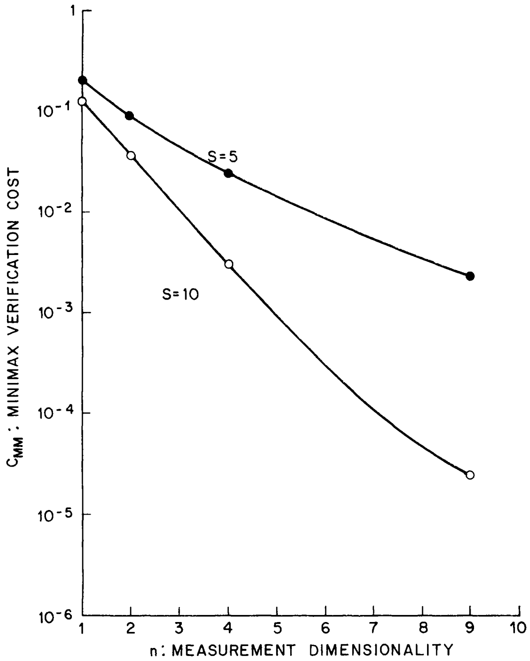


Fig. 6 Verification performance as function of measurement dimensionality.

## XII. Minimax Performance as a Function of Measurement Dimensionality

We use the results of the previous section to study the minimax cost $C_{MM}$ as a function of $n$ for the example of a diagonal covariance matrix $G$, assuming $n$ independent identically distributed measurements:

$$G = \|g_I^{de}\| = \|\sigma^2\delta_{de}\|$$

$$G_i = \|g_i^{de}\| = \left\|\frac{\sigma^2}{S_i^2}\cdot\delta_{de}\right\|, \qquad S_i \gg 1 \tag{53}$$

for $d,e = 1,2,\cdots,n$, where

$$\delta_{de} = \begin{cases}1, & \text{if } d = e\\ 0, & \text{otherwise.}\end{cases}$$

Also, we will assume that

$$\boldsymbol{m}_i = 0 \quad \text{and} \quad \mu_{\mathrm{FA}} = \mu_{\mathrm{FR}} = 1. \tag{54}$$

Fig. 6 plots $C_{MM}$ as a function of $n$ for $S_i$ values of 10 and 5. Notice the performance improvement with $n$ is more rapid for the case of $S_i = 10$, and that for a unidimensional measurement the verification costs at $S = 5$ and $S = 10$ are of the same order of magnitude.

## XIII. The Case of Inefficient Measurements

The trends of Fig. 6 indicate that practically useful verification performance can be achieved on the basis of inefficient measurements ($S \not\gg 1$), as long as the number of such measurements can be suitably large. While we cannot provide explicit results for this case, we believe that the qualitative conclusions of our paper apply to inefficient measurements as well, and whenever these measurements can be characterized by two probability distributions, the theory of this paper can determine, numerically, both the design and the performance of a pattern verifier.

## XIV. The Sequential Probability Ratio Test

The application of sequential hypothesis testing to the verification problem was indicated at the conclusion of Section VI. The theory of sequential- or variable-time-for-decision tests has been largely due to Wald [9], and the purpose of this section is merely to summarize, without proof, a few salient results from his book as applied to the verification problem. We will also note that this development is quite independent of that in Section VI in that the error performance of the verifier is predetermined in the sequential method.

In the following the total number of samples $t$ that have been observed in the course of the test procedure will be referred to synonymously as the (variable) sample size or observation time. An SPRT for verification, following earlier formalism, is defined as follows:

$$\begin{aligned}\text{accept } H_0 \text{ if } & L_t(x) \geq A\\ \text{accept } H_1 \text{ if } & L_t(x) \leq B, \qquad t \geq 1,\ A > B > 0\\ \text{continue test if } & B < L_t(x) < A.\end{aligned} \tag{55}$$

The observation $x$ is understood to be $t$-dimensional. The positive thresholds $A$ and $B$ are specified by the desired probabilities of verification error:

$$A \sim \frac{1 - P_{\mathrm{FR}}}{P_{\mathrm{FA}}}$$

$$B \sim \frac{P_{\mathrm{FR}}}{1 - P_{\mathrm{FA}}}. \tag{56}$$

If the components of $x$ are independent and identically distributed, the expected number of observations $t_{\mathrm{AV}}$ for the sequential test to terminate when $H_0$ and $H_1$ are true are given in terms of the respective expected values of the log-likelihood ratio

$$Z(x) = \ln L_1(x). \tag{57}$$

Thus

$$t_{\mathrm{AV}}(H_0) \sim \frac{(P_{\mathrm{FR}})\log B + (1 - P_{\mathrm{FR}})\log A}{E_i(z)}$$

$$t_{\mathrm{AV}}(H_1) \sim \frac{(1 - P_{\mathrm{FA}})\log B + (P_{\mathrm{FA}})\log A}{E_I(z)}. \tag{58}$$

The reader is referred to Wald [9] for demonstration that $t_{\mathrm{AV}}$ is a lower bound on the observation time (or sample size) of the shortest nonsequential test (fixed $t$) which yields the error rates $P_{\mathrm{FA}}$ and $P_{\mathrm{FR}}$, and for a discussion of truncated sequential tests where an unequivocal decision is forced whenever the time-for-decision reaches a maximum tolerable value.

It will be instructive to note that the expected values of the log-likelihood ratio for the density functions in (13) are given by

$$E_i(z) = \ln S - \frac{1}{2} + \frac{1}{2S^2} + \frac{m^2}{2\sigma^2}$$

$$E_I(z) = \ln S - \frac{S^2}{2}\left(1 + \frac{m^2}{\sigma^2}\right) + \frac{1}{2}. \tag{59}$$

One observes that for indicated typical values of $m/\sigma$ and $S$

$$|E_i(z)| < |E_I(z)|. \tag{60}$$

By virtue of (58), the preceding inequality implies that

$$t_{\mathrm{AV}}(H_0) > t_{\mathrm{AV}}(H_1) \quad \text{if} \quad P_{\mathrm{FA}} = P_{\mathrm{FR}}. \tag{61}$$

In other words, the expected time for the sequential test to terminate is greater when the true pattern is the claimed identity than when it is an impersonation; an asymmetry due, no doubt, to the assumed nature of the respective distributions of the measurement $\boldsymbol{x}$.

## XV. Conclusion

The object of this paper was to formulate the pattern verification problem in a decision theoretic framework and to evaluate the performance of optimal and suboptimal verification procedures. In this process we also noted a boundedness of the likelihood ratio and a related threshold effect.

We emphasized the utility of a minimax verification procedure whose performance is insensitive, by definition, to the probability of impersonation and compared this performance with those of minimum-cost and maximum-likelihood strategies. Finally, we gave numerical examples that illustrated the minimax performance as functions of 1) a parameter that reflected the peculiarity of the claimed pattern and 2) the measurement dimensionality.

The explicit results refer to the analytically simple case of efficient measures whose interpattern variances are much greater than their intrapattern variances. We indicated, however, that this is not a necessary attribute of a practical verifier which may typically operate on the basis of a large number of inefficient measurements. Finally, we indicated the application of sequential hypothesis testing to a variable-time-for-decision verifier designed to provide a prespecified performance, and we noted that, in our model, the expected-time-for-decision is typically greater when the true pattern is the claimed identity than when it is an impersonation.

## Acknowledgment

The author wishes to thank R. F. Estrada and T. T. Kadota for very useful comments and suggestions.

## References

[1] K. P. Li, J. E. Dammann, and W. D. Chapman, "Experimental studies in speaker verification, using an adaptive system," *J. Acoust. Soc. Amer.*, vol. 40, no. 5, pp. 966–978, 1966.
[2] G. R. Doddington, "A method of speaker verification," Ph.D. dissertation, Dep. Elec. Eng., Univ. of Wisconsin, Madison, 1970; also *J. Acoust. Soc. Amer.*, vol. 49, no. 1, p. 139, 1971 (A).
[3] S. K. Das and W. S. Mohn, "A scheme of speech processing for automatic speaker verification," *IEEE Trans. Audio Electroacoust.*, vol. AU-19, pp. 32–43, Mar. 1971.
[4] R. C. Dixon and P. E. Boudreau, "Mathematical model for pattern verification," *IBM J. Res. Develop.*, vol. 13, no. 6, pp. 717–721, Nov. 1969.
[5] R. C. Lummis and A. E. Rosenberg, private communication, Bell Telephone Labs., Inc., Murray Hill, N.J.
[6] C. W. Helstrom, *Statistical Theory of Signal Detection.* Elmsford, N.Y.: Pergamon, 1968.
[7] C. G. Pfeiffer, "Simplified techniques for recognition of target signatures," presented at the 9th IEEE Symp. on Adaptive Processes: Decision and Control, Austin, Tex., 1970.
[8] J. W. Preusse, "Speaker authentication by discrete word pattern matching," in *Proc. Workshop on Automatic Pattern Recognition Problems in Speech*, Rome Air Development Cent., N.Y., Sept. 1971.
[9] A. Wald, *Sequential Analysis.* New York: Wiley, 1947.
[10] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions.* New York: Wiley, 1954.