Correspondence

Speaker Verification Using Normalized Log-Likelihood Score

Chi-Shi Liu, Hsiao-Chuan Wang, and Chin-Hui Lee

Abstract—**In this correspondence, we propose a new scoring method for speaker verification called the normalized log-likelihood score. This method is derived from the *Bayes test for minimum risk* by the assumption of two hypotheses—the actual speaker is the claimed speaker or the actural speaker is an impostor—to attain the objective of minimizing the probability of error. The performance of this new scoring function used in speaker verification is examined by a series of experiments. For a 100-speaker database of isolated single digits, the equal error rate obtained by the normalized log-likelihood scoring method can be significantly decreased from 11.65 to 3.65% for the closed set test and from 12.30 to 8.22% for the open set test, as compared with those obtained by the conventional scoring method.**

I. Introduction

In conventional speaker verification methods, the decision rule of accepting or rejecting a claimed speaker is based on the score of a test utterance for the claimed speaker and a predefined threshold [1], [2], [3]. In recent years, some researchers [4], [5] studied the decision rule of speaker verification from the Bayesian rule [6]. Higgins *et al* used a discriminant counter to verify the speakers. The discriminant counter represents the number of times the score of a test utterance for the claimed speaker is greater than that for impostors. If this score is greater than a threshold, the claimed speaker is accepted. Rosenberg *et al* used cohort models, which are the subset of all individual models other than the claimed speaker model, to represent the characteristics of impostors; i.e., the antispeaker model. The method of selecting cohort speakers, who are a subset of existing speaker models to represent the antispeaker model, is to choose those speakers whose models are the closest to the claimed speaker model.

In this correspondence, we introduce a new scoring method called the normalized log-likelihood scoring method. We divide speaker space into the claimed speaker space and the impostor space related to this claimed speaker and then use the theorem of *Bayes test for minimum risk* to derive the decision rule for minimizing the probability of error. For this new scoring method, the antispeaker model representing characteristics of impostors can be obtained by using the existing individual models of impostors. The main difference between our method and the previous methods is in building up the antispeaker models. Our antispeaker model is the model that can characterize the acoustic space of impostors and not the model whose characteristics are close to those of the claimed speaker, which has been proposed by Rosenberg *et al* [5]. We also try to use fewer individual models of impostors to represent the antispeaker model. A series of experiments are conducted to show the effectiveness of our new scoring method.

Organization of this correspondence is as follows. In Section II, we formulate the concept of a normalized log-likelihood score and derive this score according to the concept of minimizing probability of error. In Section III, we derive the normalized log-likelihood score for the VQ method. Section IV describes the database used in this paper. Section V discusses the several experiments conducted, and, finally, Section VI gives the conclusions.

Manuscript received March 2, 1993; revised September 6, 1995. The associate editor coordinating the review of this paper and approving it for publication was Dr. Joseph Campbell.

C. Liu is with Telecommunication Laboratories, MOTC, Taiwan, Republic of China, and the Department of Electrical Engineering, National Tsing-Hua University, Hsinchu, Taiwan, Republic of China.

H.-C. Wang is with the Department of Electrical Engineering, National Tsing-Hua University, Hsinchu, Taiwan, Republic of China.

C. Lee is with AT&T Bell Laboratories, Murray Hill, NJ 07974 USA.

Publisher Item Identifier S 1063-6676(96)01335-1.

II. Normalized Log-Likelihood Score for Speaker Verification

The conventional methods for speaker verification obtain the score from the claimed speaker model, only, and compare it with a threshold calculated in the training phase. In this paper, we derive a more general decision rule for speaker verification according to the concept of minimizing the probability of error. As we know, the problem of speaker verification is to determine whether the claimed identity is true or false. By dividing speaker space into the claimed speaker space and the impostor space, the test of speaker verification for minimizing the probability of error can be formulated.

Consider the hypothesis H_0: λ, where the claimed speaker model λ is created by using all the training data O for the claimed speaker, and the *alternative* hypothesis H_1: λ', where the antispeaker model λ' is created by using all the training data O' for all impostors of the claimed speaker. The claimed speaker model λ and his antispeaker model λ' are obtained by maximizing the likelihood functions $P(O/\lambda)$ and $P(O'/\lambda')$.

Thus, according to *the Bayes decision rule for minimum risk* [7], the optimal decision rule for minimizing the probability of error for a given test data x is given by

$$\frac{P(x/\lambda)}{P(x/\lambda')} \begin{cases} > \eta, & x \in H_0 \\ \leq \eta, & x \in H_1. \end{cases} \tag{1}$$

By taking logarithmic form, the above decision rule becomes

$$\log P(x/\lambda) - \log P(x/\lambda') \begin{cases} > \log \eta, & x \in H_0 \\ \leq \log \eta, & x \in H_1 \end{cases} \tag{2}$$

where η is a predefined threshold.

For simplicity, we call S_{np}

$$S_{np}(x) = \log P(x/\lambda) - \log P(x/\lambda') \tag{3}$$

the normalized log-likelihood score.

In the case of symmetrical cost and equal prior probabilities for the claimed speaker and the antispeaker model, $\log \eta$ is equal to 0, and this threshold can be considered to be *the theoretic threshold*. If the statistics of all impostors can be obtained and modeled by the antispeaker model, the decision rule of (2) is the optimal decision rule. However, it is difficult to collect enough impostors to estimate this model. In the next section, we propose a method to approximately estimate this model from the training data of finite impostors.

A. Derivation of Antispeaker Model

We use the existing impostor models to estimate the antispeaker model. For an N-speaker training pool, we can use at most $N-1$

other speaker models to construct the antispeaker model for a claimed speaker. To save computation time, we can choose some typical speaker models to construct the antispeaker model instead of using all $N-1$ speaker models. The set of speakers used to construct an antispeaker model is referred as the *cohort set* and speakers in the cohort set as cohort speakers. Therefore, *in the training phase,* we need to find the claimed speaker model and the antispeaker model from the training pool. Such a procedure is done as follows:

1) Find the model of every speaker λ_i by maximizing the likelihood score $P(O_i/\lambda_i)$, where O_i are his own training data.
2) Given the size of cohort M and all training data for impostors $O_k^{'}$

$$O_k^{'} = \{o_k^{'}(1), \ldots, o_k^{'}(t), \ldots, o_k^{'}(T)\}$$

find the cohort Ψ_M for the claimed speaker k from $N-1$ impostors in the training pool so that its likelihood score $P(O_k^{'}/\Psi_M)$ is maximized, where T is the total number of training vectors. $P(O_k^{'}/\Psi_M)$ is defined as

$$P(O_k^{'}/\Psi_M) = \prod_{t=1}^{T} \max_{\lambda^* \in \Psi_M} P(o_k^{'}(t)/\lambda^*) \tag{4}$$

for cohort Ψ_M, where

$$\Psi_M = \arg \max_{\Psi \subset \Lambda_M^{'}} P(O_k^{'}/\Psi) \tag{5}$$

$$\Lambda_M^{'} = \{\lambda_{s(1)}, \ldots, \lambda_{s(i)}, \ldots, \lambda_{s(M)}\} \tag{6}$$

$$s(i) \in [1, N], s(i) \neq k \quad \text{and} \quad s(i) \neq s(j), \quad \text{for} \quad i \neq j.$$

The above procedure is called the maximum likelihood method since the cohort speakers are chosen so that the likelihood score $P(O_k^{'}/\Psi_M)$ is maximized. There are $(N-1)!(N-1-M)!/M!$ cohort combinations in the cohort space $\Lambda_M^{'}$ if there are N speakers in the training pool. Thus, we need to compute (6) by $(N-1)!/(N-1-M)!M!$ times to find the optimal cohort set. A simple way to quickly find the optimal cohort set is the *branch and bound* search algorithm. The steps of this search algorithm can be found in Narendra's paper [8].

B. Verification Phase

In the verification phase, two methods—the geometric mean method and the maximum method—are proposed to calculate the normalized log-likelihood score.

Geometric Mean Method: In the geometric mean method, the likelihood score for the antispeaker model is the geometric mean of the N_{av} highest likelihood scores or the arithmetic mean of the N_{av} highest log-likelihood scores in the given cohort Ψ_M of size M. N_{av} could be the value from 0 to M. That is, the normalized log-likelihood score is calculated for a given test utterance x by

$$S_{np} = \log P(x/\lambda) - \frac{1}{N_{av}} \sum_{i=1}^{N_{av}} \log P(x/\lambda_i^{'}) \tag{7}$$

where $\lambda_i^{'}$ is the cohort model having the ith highest log-likelihood score in the given cohort Ψ_M.

Maximum Method: In the maximum method, the log-likelihood score for the antispeaker model is the maximum one among the scores for the cohort models. This normalized log-likelihood score is calculated for a given test utterance x by

$$S_{np} = \log P(x/\lambda) - \max_{\lambda^{'} \in \Psi_M} \log P(x/\lambda^{'}) \tag{8}$$

where Ψ_M is the optimal cohort set of size M for the claimed speaker and is found in the training phase.

III. Speaker Verification by the VQ Model and Its Normalized Log-Likelihood Score

There are many methods proposed for generating acoustic models [2], [3]. Since this correspondence focuses on discussing the importance of using the normalized log-likelihood score for speaker verification and the methods to generate the cohort speakers, we simply use the VQ method to create each speaker's acoustic model [1].

For the VQ method, the likelihood score should be changed from a probabilistic distance to a template distance; that is, the likelihood score $\log P(O/\lambda)$ is replaced by

$$\log P(O/\lambda) = -\min_{C} D_{VQ}(O, C) \tag{9}$$

where

- D_{VQ} distance measurement (which is the Euclidean distance in this correspondence),
- C codebook,
- O set of training vectors.

By the above replacement, each speaker creates his own speaker model C and his antispeaker model $C^{'}$ in the training phase. The normalized log-likelihood scores of (7) and (8) are obtained by the same replacement of (9). The decision rule of (2) is given to determine the identity of a claimed speaker.

IV. Databases for Experiments

The database [1] used in the following experiments consists of 20 000 isolated digit utterances recorded by 100 speakers, 50 males and 50 females. Utterances were recorded over dialed-up local telephone lines with speakers seated in a sound booth using an ordinary telephone handset. Each speaker was asked to utter 200 digits, 20 repetitions of each digit, in five recording sessions over a period of 2 months. In each recording session, the speakers were prompted to utter four complete sets of the digits in random order.

The collected speech data are grouped into three databases: database A—speakers 1–20 (10 males and 10 females); database B—speakers 21–40 (10 males and 10 females); and database C—speakers 21–100 (40 males and 40 females). The speakers in databases A and B and B and C are unique, but databases B and C overlap. Using database A test utterances would form a closed set test if we use database A to determine cohort speakers. We alternately choose each one of the 20 speakers in database A as the claimed speaker and the other 19 speakers as impostors. The error rate is the average of the 20 individual error rates for the 20 claimed speakers. Utterances from databases B and C are data for the open set test. That means the utterances from databases B and C are used as the utterances of impostors, but not used to create the antispeaker models. The error rate is still the average of the 20 individual error rates for each speaker in database A. No matter which database is used, the first 80 utterances of each speaker are used as training data and the last 120 utterances as test data.

All utterances are bandpass filtered from 200 to 3200 Hz and sampled at 6.67 kHz. The digitized speech signal is preemphasized using the filter $H(z) = 1-0.95z^{-1}$. Nine autocorrelation coefficients are calculated over the 45-ms frame, and then, 16 cepstral coefficients are derived. Each frame is Hamming windowed and shifted every 15 ms. Codebook size for all VQ models is 64, and the test utterances are isolated single digits.

Equal error rate is used to measure the performance of speaker verification for different scoring methods. The equal error rate is a *posterior* error rate and, at this equal error rate, the decision boundary is set to make the error rate of false rejection be equal to that of false acceptance. The *posterior* equal error rate is a convenient measure of

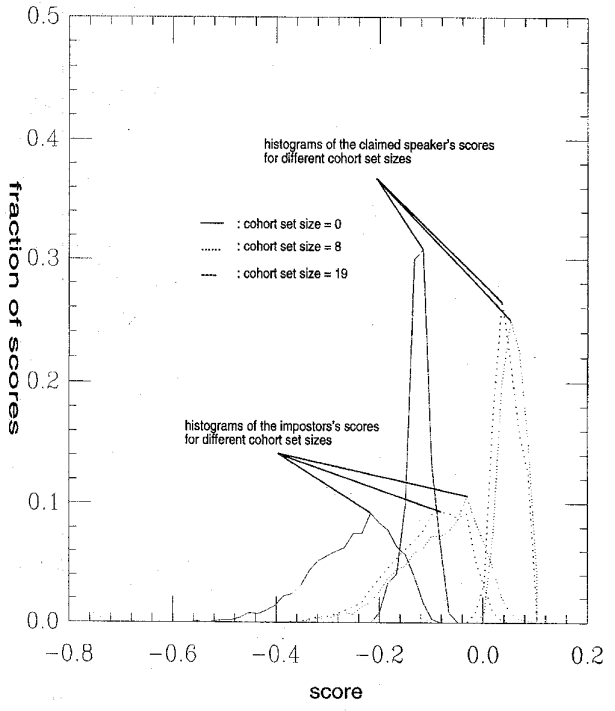

Fig. 1. Histograms of the claimed speaker's and impostors' scores for different cohort set sizes.

the degree of separation between true and false speaker scores and, therefore, a useful predictor of speaker verification performance.

V. Experiments and Discussion

A series of experiments are conducted to do the following:

1) Compare the statistics of score distribution using the normalized log-likelihood scoring method with those using conventional scoring methods.
2) Compare a *posterior* equal error rate for closed and open set tests using normalized or unnormalized log-likelihood scores.
3) Compare the performance of different cohort speaker selection methods.
4) Compare the performance of different methods to calculate the normalized log-likelihood score in the verification phase.

A. Statistical Analysis for the Normalized Log-Likelihood Scoring Method

The first experiment was to understand the statistical distribution for the normalized log-likelihood score. We used the first 80 utterances of each speaker in database A as the training data and the last 120 utterances of every speaker in database A as the test data.

Fig. 1 shows the histograms of a claimed speaker's scores and his impostors' scores for different cohort set sizes and Table I shows their statistics. The number of cohort speakers equal to zero represents the scoring method without normalization; that is, using the conventional decision rule [1], [3]. Some interesting observations can be seen from this figure and table:

1) As cohort set size increases, the variance of the impostors' scores decreases. This is because the impostors' test data are closer to those of the antispeaker model as the cohort set size increases.
2) The distance between the claimed speaker's scores and his impostor's scores increases with cohort set size. This again shows the effectiveness of using the normalized log-likelihood score for speaker verification.
3) As cohort set size increases, the threshold value for the equal error rate is closer to *the theoretic threshold.*

TABLE I
Statistics of the Normalized Log-Likelihood Scores for a Claimed Speaker and His Impostors

	claimed speaker		impostors	
cohort set size	mean	variance	mean	variance
0	-0.150	0.0006	-0.276	0.0068
8	0.037	0.0007	-0.109	0.0063
19	0.023	0.0006	-0.134	0.0057

TABLE II
Equal Error Rates of the Closed Set Test for Different Cohort Set Sizes

cohort set size	0	1	2	3	4	5	6	7	8	19
equal error rate (%)	11.65	14.82	10.88	10.55	8.62	7.93	7.24	6.77	6.44	3.65

TABLE III
Equal Error Rates of the Open Set Test for Database B

cohort set size	0	2	4	6	8	10	19
equal error rate (%)	10.99	12.40	10.04	8.48	8.46	8.43	8.22

B. Closed Set Test

In the previous experiment, we showed the discriminative ability of the normalized log-likelihood scoring method. In this experiment, we compare the performance of speaker verification using the normalized log-likelihood scoring method with that using the conventional unnormalized scoring method. The training and test data used in this experiment are the same as those in the previous experiment. This test is called the *closed set test* since the speakers used to make the cohort set are the same as those used in testing.

The results are shown in Table II. For cohort set sizes greater than two, normalized log likelihood scoring has a lower equal error rate than conventional unnormalized scoring. The equal error rate further decreases to 3.65% as the size of the cohort set increases to 19. This shows that the equal error rate is significantly reduced from 11.65% to 3.65% after using the normalized log-likelihood score.

C. Open Set Test

In this experiment, we examine the effectiveness of the normalized log-likelihood score in the open set test. Training data are still the same as those in experiment B, but the test data are from database B. The cohort speakers to represent the antispeaker model are from database A.

Table III shows the equal error rates using the normalized log-likelihood scoring method for different cohort set sizes. The improvement in using the normalized log-likelihood score still exists, although this improvement is less than that in the closed set test. This illustrates that using limited impostor models to represent the antispeaker model does not fully reflect the characteristics of all impostors. Another finding from this open set test is that when the cohort set size is greater than six, the improvement becomes small. Although increasing the cohort set size could decrease the false acceptance error rate, it could also increase the false rejection error rate. As the decreasing amount is close to the increasing amount, the equal error rate is decreased slowly. Since any impostor's information for the open set test is not in the cohort set and the given cohort speakers can not represent all

TABLE IV
EQUAL ERROR RATES (IN PERCENT) FOR THE OPEN SET TEST

cohort set size		0	2	4	6	8	10	19
impostor data	database B	10.99	12.40	10.04	8.48	8.46	8.43	8.22
	database C	12.30	12.62	10.30	8.85	8.76	8.54	8.24
rate difference(%)		1.31	0.22	0.26	0.37	0.30	0.11	0.02

TABLE V
EQUAL ERROR RATES (IN PERCENT) FOR DIFFERENT SELECTION METHODS

cohort set size		0	2	4	6	8	10	19
closed set test	maximum likelihood	11.65	10.88	8.62	7.24	6.44	5.82	3.65
	random	11.65	13.92	10.61	8.96	7.73	5.91	3.65
	Rosenberg's method	11.65	16.42	12.26	10.08	8.74	5.94	3.65
open set test	maximum likelihood	10.99	12.40	10.04	8.48	8.46	8.43	8.22
	random	10.99	14.75	12.06	11.00	10.17	9.07	8.22
	Rosenberg's method	10.99	17.08	13.60	12.25	11.17	9.45	8.22

impostors, the improvement diminishes as the cohort set size exceeds some value.

The results using the last 120 utterances of database C as the test data of impostors are shown in Table IV and the improvement from the normalized log-likelihood scoring method still can be seen. We also observe that the equal error rate using the normalized log-likelihood scoring method is more insensitive to the number of open set impostors than that of using the unnormalized scoring method. For the unnormalized scoring method, the difference of equal error rates between databases B and C is 1.31%, but for the normalized log-likelihood scoring method, the difference is below 0.37% for any cohort set size.

D. The Methods of Selecting Cohort Speakers

The method of making up the cohort set used in the previous experiments is the maximum likelihood method. To understand the correctness of this method, we compare its performance with those of randomly selecting cohort speakers and Rosenberg's selection method [5]. The maximum statistic is used in Rosenberg's selection method. We use database A as the training data, database A as the *closed set test data*, and database B as the *open set test data*.

The results given in Table V clearly show that the performance by the maximum likelihood selection method is the best one in both *the closed set test* and *the open set test*.

E. Performance Comparison for Different Scoring Methods in the Verification Phase

Two scoring methods, the geometric mean method and the maximum method, are proposed to calculate the normalized log-likelihood score in the verification phase. To compare the performance of these two methods, we use the same experimental conditions as those in the above experiment. The cohort set sizes are chosen as eight and 19. The number of the highest log-likelihood scores to be averaged, denoted as N_{av}, varies from 0 to the cohort set size. The unnormalized scoring method is given by N_{av} equal to 0, and the maximum scoring method is given by N_{av} equal to 1.

Closed and open set test results are given in Fig. 2. For the closed set test, the best results are obtained by the maximum scoring method. These results are independent of the cohort set size. The reason for this could be that any impostor in the closed set is much closer to one particular cohort speaker than any other. For the open set test, the best results are obtained by the geometric mean method. However, this depends on the value of N_{av} and the cohort set size M. For M equal to eight, the best result is obtained by setting N_{av} to three, but for M equal to 19, N_{av} should be set to five or six.

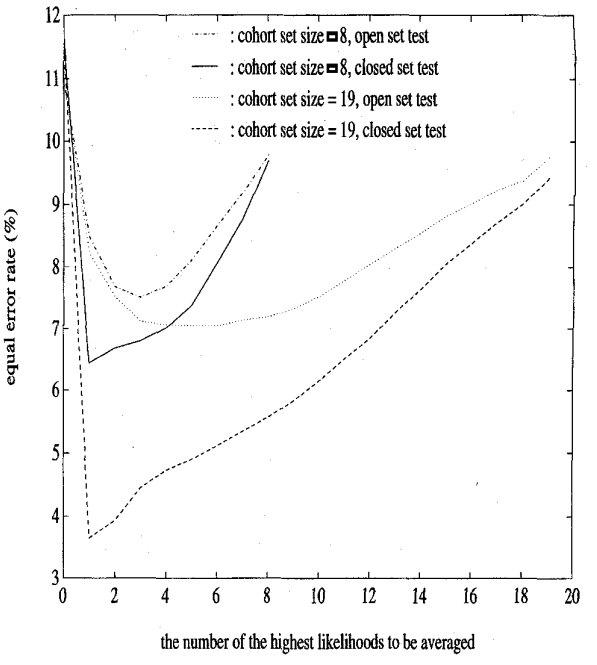

Fig. 2. Equal error rates versus the number of the highest scores to be averaged.

VI. CONCLUSIONS

In this correspondence, we introduced and derived the normalized log-likelihood score for speaker verification. It is derived from the concept of *Bayes test for minimum risk*. The *branch and bound* search algorithm is applied to quickly find cohort speakers. After a series of experiments we found the following:

1) This normalized log-likelihood score method got better performance than the conventional scoring method in both our *closed set test* and our *open set test*.
2) The performance by the normalized log-likelihood scoring method is more insensitive to the number of *open set* impostors than the conventional scoring method.
3) The performance using the maximum likelihood method to select cohort speakers is better than both the random selection method and Rosenberg's method.
4) For the open set test, the geometric mean method used to calculate the normalized log-likelihood score of a test utterance performs better than the maximum method by appropriately choosing the number of the highest log-likelihood scores to be averaged.

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Rosenberg for his assistance and AT&T Bell Laboratories for supplying the database. They would also like to thank Dr. J. T. Wang, Dr. I. C. Jou, Dr. B. S. Jeng, and our colleagues for supporting this research. Finally, we would like to thank associate editor Dr. J. P. Campbell and the anonymous reviewers for their reviews and useful recommendations.

REFERENCES

[1] F. K. Soong, A. E. Rosenberg, L. R. Rabiner, and B. H. Juang, "A vector quantization approach to speaker recognition," *AT&T Tech. J.*, vol. 66, pp. 14–26, Mar./Apr. 1987.

[2] A. E. Rosenberg, C-H. Lee, and F. K. Soong, "Sub-word unit talker verification using hidden Markov models," in *Proc. ICASSP-90*, vol. 1, Apr. 1990, pp. 269–272.
[3] N. Tishby, "On the application of mixture AR hidden Markov models to text independent speaker recognition," *IEEE Tran. Acoust., Speech, Signal Processing*, vol. 39, pp. 563–570, Mar. 1991.
[4] A. Higgins and L. Bahler, "Text-independent speaker verification by discriminator counting," in *Proc. ICASSP-91*, vol. 1, May 1991, pp. 405–408.
[5] A. E. Rosenberg, J. Delong, C. H. Lee, B. H. Juang, and F. K. Soong, "The use of cohort normalized scores for speaker recognition," in *Proc. ICSLP-92*, vol. 1, Oct. 1992, pp. 599–602.
[6] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[7] K. Fukunage, *Introduction to Statistical Pattern Recognition*, 1972.
[8] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Tran. Computers*, vol. C-26, pp. 917–922, Sept. 1977.

Transparent Quantization of Speech LSP Parameters Based on KLT and 2–D-Prediction

Fu-Rong Jean and Hsiao-Chuan Wang

Abstract— **In this correspondence, a two-stage approach based on Karhunen–Loeve transform and 2-D prediction is proposed for efficient quantization of line spectrum pair (LSP) parameters of speech. Besides, a switched classifier is incorporated with this approach to reduce the outlier frames (spectral distortion greater than 2 dB) down to about 0.27% and to eliminate frames with spectral distortion greater than 4 dB at an average bit-rate below 19 b/frame.**

I. Introduction

In low bit-rate speech coders, the line spectrum pair (LSP) parameters were found to be efficient in representing the short-time spectrum envelope information for their good quantization and interpolation characteristics. Many quantization schemes were designed by using the intraframe correlation or the so-called ordering properties of LSP parameters [1]–[3]. The strong interframe correlation of LSP parameters was also considered in several coding schemes [4], [5]. Among these, the transparent quantization of spectrum information was proposed in [3]. It means that the spectrum quantization does not introduce any audible distortion in the coded speech if the following requirements are satisfied simultaneously.

1) The average spectral distortion is less than about 1 dB.
2) There are no outlier frames having spectral distortion larger than 4 dB.
3) The number of outlier frames having spectral distortion in the range 2–4 dB is less than 2%.

Manuscript received June 29, 1994; revised July 19, 1995. This work was sponsored by the National Science Council, Taiwan, Republic of China, under Contract NSC82-0408-E007-319. The associate editor coordinating the review of this paper and approving it for publication was Dr. Spiros Dimolitsas.

F.-R. Jean is with the Department of Electrical Engineering, Taipei Institute of Technology, Taipei, Taiwan, Republic of China.
H.-C. Wang is with the Department of Electrical Engineering, National Tsing-Hua University, Hsinchu, Taiwan, Republic of China.

Publisher Item Identifier S 1063-6676(96)01336-3.

This correspondence proposes a new method not only to reduce the bit-rate at the 1 dB difference limen but also to reduce the number of outlier frames. To solve the former problem, a two-stage approach based on Karhunen–Loeve transform (KLT) and 2-D prediction is proposed for this purpose. To solve the latter problem, a switched classifier is built to reduce the outlier frames and to eliminate frames with spectral distortion greater than 4 dB.

The rest of the correspondence is organized as follows. In Section II, the two-stage coding (TSC) scheme of LSP parameters is presented. In Section III, we describe the quantization schemes and their corresponding bit allocations. Section IV shows the experiment and simulation results. A switched classifier to reduce the outlier frames is also introduced. Finally, a conclusion is given in Section V.

II. Two-Stage Coding (TSC) Scheme

A. Distortion Measure

Given a pth–order linear predictive coding (LPC) model, where p is assumed to be an even integer, let the LSP vector at the nth frame of speech signals be denoted by $\omega(n) = [\omega_1(n) \quad \omega_2(n) \quad \cdots \quad \omega_p(n)]^T$. The parametric distortion [6] is defined by

$$PD_n^2 = \sum_{i=1}^{p} g_i^2(n)[\omega_i(n) - \hat{\omega}_i(n)]^2 \tag{1}$$

where $\hat{\omega}_i(n)$ is the reconstructed value of $\omega_i(n)$. The adaptive weighting factor $g_i^2(n)$ is the so-called spectral sensitivity with respect to $\omega_i(n)$ [1], which is defined as

$$g_i^2(n) = \frac{1}{\pi}\int_0^{\pi}\left|\frac{\partial \log S_n(\omega)}{\partial \omega_i(n)}\right|^2 d\omega. \tag{2}$$

Furthermore, in this study we divide the parametric distortion into two parts, i.e.

$$\begin{aligned} PD_n^2 &= PD_{n,E}^2 + PD_{n,O}^2 \\ &= \sum_{i=1}^{p/2} g_{2i}^2(n)[\omega_{2i}(n) - \hat{\omega}_{2i}(n)]^2 \\ &\quad + \sum_{i=1}^{p/2} g_{2i-1}^2(n)[\omega_{2i-1}(n) - \hat{\omega}_{2i-1}(n)]^2 \end{aligned} \tag{3}$$

where $PD_{n,E}^2$ is the portion of the spectral weighted quantization error produced at the first stage and $PD_{n,O}^2$ is the portion at the second stage.

B. First Stage of TSC Scheme

The even part of LSP vector at the nth frame of speech signals is denoted by $\omega_E(n) = [\omega_2(n) \quad \omega_4(n) \quad \cdots \quad \omega_p(n)]^T$. Let $\overline{\omega}_E = [\overline{\omega}_2 \quad \overline{\omega}_4 \quad \cdots \quad \overline{\omega}_p]^T$ be the mean vector of $\omega_E(n)$, where $\overline{\omega}_i$ is the mean value of $\omega_i(n)$, i.e., $\overline{\omega}_i = E[\omega_i(n)]$. The block diagram of this scheme at the first stage is shown in Fig. 1. Matrices $\mathbf{A} = [a_{ij}] = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_{p/2}]$ and $\mathbf{B} = [b_{ij}] = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_{p/2}] = \mathbf{A}^{-1} = \mathbf{A}^T$ are considered to be an orthogonal KLT transform pair, where $i, j = 1, 2, \cdots, p/2$. KLT is performed on the mean removed even part of LSP vector, i.e.

$$\begin{aligned} \Omega_E(n) &= [\Omega_2(n) \quad \Omega_4(n) \quad \cdots \quad \Omega_p(n)]^T \\ &= \mathbf{A}[\omega_E(n) - \overline{\omega}_E]. \end{aligned} \tag{4}$$