Speaker recognition by computer

E. Bunge

We all know from experience that, people can be easily recognized by their voices. This comes about not just from their individual habits of speech, but from the endless variety of the acoustic parameters of the vocal tract throughout the population — it is almost impossible to find two voices that are perfectly identical. The voice may therefore serve as a passport, provided there are technical means of distinguishing the individual voices from one another with a sufficiently high degree of reliability. This is where the techniques for speech analysis come in. These were developed years ago for resolution and resynthesis of the speech signal to condense the information so that speech could be transmitted in narrower frequency bands; now they have been turned to the more general applications of man-machine communication by voice. As might have been expected, computers have come to play a large part in this. Whereas the prospect of a free and fluent conversation between man and machine still seems to be a long way off, the simpler task of recognizing a person by his voice has been proved to be well within the capabilities of a machine and high reliability has been obtained. Research projects on this subject are going on in several places in the world. One of these is a government-sponsored project called AUROS at Philips Forschungslaboratorium Hamburg, in West Germany.

Introduction

When listening to somebody who is talking, a human listener perceives what is said, who is speaking and how it is said. The human brain is capable of splitting up the complex information contained in the speech signal to answer simultaneously questions about the meaning of what is said, the identity of the speaker and his emotional state. Investigations have been going on for some time to see whether the artificial intelligence available in computers can be employed to perform at least part of this task.

Speech recognition by computers — i.e. computers understanding what is said — still seems to be a very difficult problem and to give valuable results only with restricted vocabularies and restricted groups of speakers. On the other hand, speaker recognition — the identification of who is speaking — seems to have made considerable progress and to be nearly ready for practical application.

The value of a device for speaker recognition is obvious. It would be highly efficient if a simple voice utterance was sufficient in all those cases where a person’s identity has to be established or verified. The human voice is very specific to the individual speaker; it cannot be lost or stolen, it cannot easily be imitated, and it can be transported over long distances at low cost. For all these reasons the human voice would provide an ideal ‘acoustic passport’. This passport could be used to supplement existing systems to increase their security, or by itself, e.g. to allow authorized access to money from banks and to confidential information from data banks and government authorities. A second application is in crime investigation, where recorded voices of blackmailers and kidnappers have to be compared with the voices of suspects.

Systems for speaker recognition have one or more speech-input terminals, connected to a central computer. The connection may be made via the public telephone system. The voice is analysed in the computer and compared with stored voice-patterns.

Dr E. Bunge, formerly with Philips GmbH Forschungslaboratorium Hamburg, Hamburg, West Germany, is now with Bundeskriminalamt, Wiesbaden, West Germany.

Speaker identification has to be distinguished from speaker verification. In speaker verification an unknown speaker claims to be a certain person X, and by comparing his voice data with that of speaker X the system decides whether he really is speaker X or an impostor. For speaker identification, an unknown voice is compared with a stored voice-pattern library and the system finds out if the voice is present in the library, and if so, who it belongs to.

The greater part of this article is a description of an experimental speaker-recognition equipment called 'AUROS' (AUtomatic Recognition Of Speakers), which has been designed at Philips Forschungslaboratorium Hamburg, in West Germany, for a government-sponsored research project at PFH. The system has been set up to enable us to compare the merits of different speaker-identification and verification procedures. Attention is given to the error rates in experiments with larger groups of subjects, and also to the computer time required and the complexity of the equipment, which is quite different for different procedures. The experiments already made — which will be reported in the last section — indicate that the error rates are of the order of one per cent for procedures of average complexity [2].

In speaker recognition by computer there are two fundamental problems: first, all speech utterances are unique and not exactly reproducible, and, secondly, the amount of data to be processed in short time intervals is extremely large. Let us look more closely at these two problems and at the procedures that they compel us to adopt.

The non-reproducibility of human speech

No-one can pronounce a sentence twice in an identical manner. Each repetition yields a different acoustic representation. Fig. 1 shows three curves of sound pressure as a function of time for three utterances of the vowel 'a' pronounced by the same speaker. The differences between the three curves are evident although the speaker tried hard to produce identical utterances.

The non-reproducibility also holds for any feature extracted from the original time function. Fig. 2 shows as examples the Fourier spectra and the 'cepstrum' functions of the three versions of the vowel 'a'. This function is obtained by computing the power spectrum of the logarithm of the Fourier spectrum. The Fourier spectrum has periodic ripples corresponding to the harmonics of the speech signal; the frequency spacing between these ripples is equal to the fundamental frequency of the speech. The cepstrum, which is again a time-domain function, will have a peak corresponding to the periodicity of the Fourier spectrum, and indicating the fundamental pitch period of the speech [3].

It follows that the voice of a speaker cannot be described by one single utterance, because other utterances produce different data. Instead, it is necessary to process a set of utterances statistically to determine the characteristic features of a voice. This can be done systematically by pattern-recognition techniques.
The large amount of data to be processed

The sampling theorem shows that with an upper limiting frequency of 5 kHz, 1 second of human speech can be represented by a string of 10 000 samples. With a dynamic range of 48 dB, each sample can be coded as an 8-bit computer word. Under these circumstances — which correspond to those in our experiment —

the amount of information contained in 1 second of human speech is equal to 80 kilobits. Since many seconds of speech by many different speakers have to be stored and processed, problems of amount of storage and computer time are inevitable. Fortunately, the speech signal has high redundancy; the same characteristic features are encountered over and over again. These features can be extracted by specially developed processors, which is preferably done in real time. Instead of the speech signal it is only necessary to store these voice-specific features in the computer, to form the reference with which newly arrived speech utterances are compared in a manner similar to the procedures used in optical-pattern recognition by computers. The term ‘pattern recognition’ will therefore be used here.

General description of the AUROS equipment

A meaningful approach to the speaker-recognition problem is to combine real-time speech-analysis techniques with statistical pattern-recognition methods. But there is no a priori knowledge that indicates which combination of analysis and pattern-recognition methods will give the highest recognition rates for large populations. The AUROS equipment allows the merits of different combinations to be assessed. It consists of a general-purpose computer for pattern recognition and a set of the signal-analysis processors referred to above. These have been specially designed by us for this application; the voice features resulting from the signal analysis are stored in the computer memory. A dedicated computer for simulation is also available.

Most of the processors perform an analysis in real time on consecutive time segments of the speech signal. Each segment is described by a set of numbers representing quantities such as the output voltages of a filter bank used to measure the frequency spectrum. These numbers are considered as the components of a multidimensional vector. The probability-density distribution over the vector space, or a mean vector, is calculated from the consecutive vectors.

A second set of analysis procedures is based on linear transformations such as the fast Fourier transform, the Walsh-Hadamard transform and inverse filtering. They are used off-line to obtain pitch contours and cepstrum functions, and the frequencies and bandwidths of the higher-order speech resonances (the formants).

An explanation of the three linear transformations mentioned here is perhaps appropriate. The fast Fourier transform is simply a computer algorithm enabling us to carry out the complicated computations — implying several multiplications — of the Fourier coefficients in a reasonable time. The repeated computation of Fourier spectra required in some applications has only become feasible thanks to this algorithm [6].

Like the Fourier transform, the Walsh-Hadamard transform is used to compute a frequency spectrum. In place of the sine and cosine functions, however, it uses square-wave functions for the set of orthogonal functions required for the signal decomposition [5]. The square-wave functions only have the values +1 and −1; their use simplifies digital implementation.

Inverse filtering, finally, is based on the observation that the frequency spectrum of the human voice, characterized by the frequencies and bandwidth of the formants, derives mainly from the acoustic resonances of the vocal tract; the excitation by the glottis — which sets up a pulse train — has in itself a rather flat frequency spectrum. If we succeed in filtering the speech signals in such a way that a flat spectrum is again obtained, then the filter has a transfer function that is the inverse of the acoustic transfer function of the speaker’s vocal tract. The parameters of the inverse filter are therefore characteristic of the speaker [4].

In the AUROS system the analysis data, obtained either in real time or by the more time-consuming linear transformations mentioned above, is fed to the general-purpose computer for statistical evaluation and ‘classification’, i.e. assignment to a known speaker. A large set of pattern-recognition methods is available.

A supervisory program allows a special analysis procedure and a pattern recognition method to be selected and combined for special experiments. The input to the system takes the form of voice signals, which can either be received from microphones in an on-line experiment or recorded on magnetic tape for off-line experiments with large data bases. The output of the system takes the form of statistical data describing voices, recognition rates, error rates and rejection rates.

The procedure followed in speaker recognition can generally be described by three phases:
— preprocessing the speech signal,
— ‘training’ the system, and
— testing its reliability.
Each of these three phases will now be dealt with in more detail. This provides a natural opportunity to present an outline of several speech-analysis procedures and classification algorithms.

Preprocessing phase

As was stated earlier, the acoustic speech signal contains information concerning what, who and how, and the most suitable approach is therefore not to use the redundant speech signal for the recognition procedure, but to describe it by a set of features that are typical of the speaker's identity. This is what has to be done in the preprocessing phase. In speaker recognition

Short-time analysis of a preselected segment

The idea behind the procedure of analysis of a single preselected segment is that the speaker-specific information is already contained in a segment of a few tens of milliseconds of speech. So, using a segmentation algorithm, a special segment of this length, e.g. a few pitch periods of a nasal sound, is selected from the acoustical signal. The same segment is selected from

there are three general ways of feature extraction. One way is to rely on the analysis of one short preselected time segment of the speech utterance. Another method is to measure contours such as those of pitch or energy content, during a specific utterance. The third way is to compute the mean (or the probability-density distribution) of the feature vectors characterizing successive time segments. Each of the three methods was studied by means of the AUROS system.

the same 'code word' for all the speakers. Operations such as a Fourier transform, a Walsh-Hadamard transform or a cepstrum transform are performed on this segment. The coefficients of these transforms are the components of feature vectors used in the pattern-recognition procedure. Fig. 3 shows four examples of an analysis of a 20-ms segment of vowel 'a'.

The problem in short-term feature extraction is the accuracy of the preceding segmentation algorithm. If
the defined segment in the acoustic signal is not located exactly, the following steps are not meaningful since representations of physically different events are being compared.

**Contour analysis**

Contour analysis describes the temporal structure of an utterance. The acoustic signal is segmented equidistantly. For each segment only a single feature, e.g. the pitch period or the energy, is evaluated. The sequence of these segment features forms the resulting contour curve. *Fig. 4* shows as an example the pitch contour of the sentence 'My name is Nemo' which was obtained by applying cepstrum analysis and peak detection to the segments (about 100) of this speech utterance.

The results can be combined to form a feature vector of a dimensionality equal to the number of segments. The consequence of this is that if a given sentence is spoken slowly the dimensionality will be high (because many segments are produced), and if it is pronounced rapidly a low dimensionality will result. However, since the pattern-recognition algorithms assume equal dimensionality for each speaker, complicated time-normalization methods have to be applied to the contour vectors to stretch them for fast speakers and to compress them for slow speakers. These time-normalization or time-warping methods are complicated and require a great deal of computer time [7].

The use of contour analysis for feature extraction in speaker recognition entails a restriction to code-word-related recognition, as with the previous method. Text-independent recognition is not possible because the time-normalization methods are adapted to a given code word.

**Statistical speech-signal analysis**

The speech signal is divided into a sequence of 20-ms segments by equidistant segmentation. A set of coefficients is evaluated for each segment by one of the available methods of analysis; these include its linear transformations mentioned earlier and also the determination of the autocorrelation function. As mentioned earlier, a feature vector is combined from the sequence of these sets of coefficients, either by evaluating the mean-value vector or by approximating probability densities of event groups.

This kind of statistical feature extraction will give speaker recognition without the condition that all the speakers speak the same prescribed text; however, we found that the processed speech signal should last at least 10 seconds. There are no segmentation or time-normalization problems. These advantages turn out to be of considerable importance, which is why, in the course of our investigations, we have come to prefer this statistical method in our research project. As an example of this technique, *fig. 5* shows the averaged long-term Fourier spectrum, the long-term Walsh-Hadamard transform, the long-term cepstrum, and the long-term autocorrelation function. All of these have been evaluated from the same text.

**Training phase**

To characterize the voice of a speaker it is necessary to combine the analysis results of a few utterances (at least ten) to a typical voice reference. Because of the poor reproducibility of human speech a single utterance does not contain a sufficient amount of information. It is in the training phase that a reference ‘pattern’ for each voice is stored in the system.

Each feature vector generated by preprocessing the speech signal can be mapped as a pattern point in an $N$-dimensional feature space, where $N$ is a quantity such as the number of Fourier coefficients. Each utterance of a speaker produces a new point in this space. The entirety of all the pattern points of the training phase forms a ‘cloud’ in the $N$-dimensional space.

Different methods can be followed to find out whether a pattern point produced by an utterance from an unknown speaker must be assumed to belong to a given speaker or not. It depends on the method which further operations the voice data will have to undergo while training the system. One method is to evaluate the distance from the point to the centre of gravity of the cloud of pattern points belonging to the known speaker. Another method establishes whether the new point lies inside the (hyper)sphere enveloping the cloud. A third method divides the $N$-dimensional space into subspaces and determines the probability for each
subspace that a point in it will belong to any of the given speakers. The operations performed during the training phase on the points in the $N$-dimensional space in the first case consist in the evaluation of the centre of gravity of the cloud of points, in the second case in the construction of the (hyper)sphere, in the last case in the determination of the probabilities for the individual subspaces.

![Graphs of Fourier spectrum, Walsh-Hadamard spectrum, cepstrum function, and autocorrelation function.](image)

**Fig. 5.** Long-term averages over an utterance about 10 s long of Fourier spectrum (a), Walsh-Hadamard spectrum (b), cepstrum function (c), and autocorrelation function (d). The statistical feature-extraction method avoids the problems of segmentation and time normalization and permits text-independent speaker recognition.

We have used the three methods outlined here in speaker-recognition experiments. There were many modifications for measurements of distance, data normalization and probability-density approximation. These methods will now be described in some more detail. The assignment of a point to a given speaker is called 'classification', which is related to the term 'class' used for the array of pattern points of a speaker; the algorithm performing the assignment is called a 'classificator'.

**Minimum-distance classificator**

The centre of gravity of the cloud of pattern points is evaluated independently for each speaker and stored as reference vector. An unknown pattern point is classified by evaluating the distances from this point to the centres of gravity of each cloud. This point is then associated with the cloud whose centre of gravity is the nearest.

The distance measure used for defining the proximity of points may be the usual Euclidean distance. The distance measure may also be based on the correlation between the two corresponding feature vectors. The correlation is at a maximum when the two vectors coincide; the distance may be defined as the amount by which the correlation falls short of this maximum.

---

Let the feature vectors be $X$ and $Y$ with components $(x_1, x_2, \ldots, x_N)$ and $(y_1, y_2, \ldots, y_N)$. The Euclidean distance is

$$d_E(X,Y) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2}. \quad (1)$$

The correlation between the two vectors is expressed by the correlation coefficient, whose maximum value is 1. By subtracting the true correlation coefficient from 1 we obtain the distance measure

$$d_C(X,Y) = 1 - \frac{\sum_{i=1}^{N} x_i y_i}{\sqrt{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i^2}}. \quad (2)$$

Both distance measures can be altered by weighting each dimension by the inverse of its variance between different utterances of the same speaker, to base the decision mainly on the parameters of the feature vector with the highest reliability.

**Distribution-free tolerance-region classifier**

The region occupied in multidimensional space by a cloud of pattern points can also be taken as its prime characteristic, regardless of the distribution of the points over this region. This region may be demarcated by evaluating an enveloping hypersphere of the cloud of pattern points in a manner that avoids intersections between different classes. The mathematical equation of this hypersphere defines the tolerance region inside which a point must fall to be classified with the corresponding speaker, and thus represents his voice characteristics. To classify an unknown pattern point a check is made for each class to see whether the point lies within its hypersphere or not. If the point cannot be found in any tolerance region, it will be rejected.

**Minimum-risk classifier**

With the minimum-risk classifier, the $N$-dimensional pattern space is subdivided into subspaces of equal volume. For each subspace a conditional probability is approximated by counting the number of pattern points in this volume independently for each speaker. This conditional probability indicates the likelihood that a pattern point, given that it belongs to a certain speaker, will lie in that subspace. For classification of a pattern point a check is made to find out which subspace it lies in, and the values of the corresponding conditional probabilities for this subspace are compared. The unknown point is then classified with the speaker having the highest conditional probability. In this way the risk of loss due to a wrong classification is minimized. Fig. 6 shows the principle for the one-dimensional case.

**Reliability-test phase**

During the reliability-test phase, utterances that have not been used for training are offered to the system for classification. In a speaker-identification test the utterance belonging to one of the speakers entered into the system may be either correctly classified, or wrongly classified, or rejected as unclassifiable. The wrong classifications and rejections together determine the typical error rate for the system performance. Wrong classifications cause speaker confusion; by counting the number of confusions the probability of confusion of each pair of speakers can be represented in a 'confusion matrix'.

In speaker verification the system only has to decide between the alternatives of acceptance or rejection. An error may be either the false rejection of a customer or the false acceptance of an impostor. Fig. 7 shows a simplified speaker-verification experiment with real data. Speaker X spoke a test sentence 15 times. A long-term spectrum was evaluated for each utterance. Fig. 7a shows a plot of these 15 spectra. The non-reproducibility is clearly seen. In a 'model' training phase, a tolerance region was constructed in form of a mask by connecting all the minimum and maximum values (fig. 7b); no consideration was given to the distribution of the values between the maximum and minimum. For verification, the speaker X repeated the same text, and the result of the analysis was compared with the reference. It was accepted, as this long-term spectrum was within the tolerance region (fig. 7c). When speakers Y and Z claimed to be speaker X, they were rejected, because their curves did not fit the tolerance-region template (fig. 7d).
Test results obtained with the AUROS system

A number of classification procedures have been tested using the AUROS system, each of them employing several kinds of voice features, some of which have been mentioned above. In experimenting with different kinds of voice features the advantages of long-term spectra — the absence of segmentation and time normalization — turned out to be very important, as was the possibility of text-independent voice-feature extraction. We therefore found it preferable to use the long-term spectra for our most important tests; the results of these tests will now be reported.

In the AUROS system the long-term spectra are derived in real time, and this is very important for any practical application. Spectral analysis is performed by a 43-channel filter bank covering the frequency band from 100 Hz to 6 kHz; each filter has a bandwidth of nearly 10 per cent of the centre frequency. Equidistant segmentation is carried out by scanning the filter outputs every 20 ms. The short-term spectra are averaged by a special digital multiplex integrator, which also performs a rough energy normalization. For a comparative study 5000 utterances of 82 male and 18 female subjects have been processed. Fig. 8 shows long-term spectra of a 12-s utterance of two men (a) and two women (b). The absence of energy in the first three channels for the women indicates the higher pitch of their voices.

It turns out that for a given data base the variations in the recognition rate with the chosen pattern-recognition algorithm are considerable. Each algorithm only operates in the optimum fashion when a set of a priori assumptions is satisfied, e.g. concerning the kind of distribution of the features or the correlation and covariance of the data of each individual speaker (which may be expressed by a correlation and a covariance matrix). For that reason, investigation of the statistics of the data base is necessary to adapt the pattern-recognition algorithm to the data structure.

From the correlation matrix of the 43 frequency parameters averaged over 2500 long-term spectra (see fig. 9) it can be seen that the 43 parameters of the feature vector are not statistically independent but correlated. The average correlation coefficient is 0.4. This correlation has to be taken into account when designing an appropriate pattern-recognition algorithm.

For optimum performance of most of the pattern recognition algorithms it is assumed that the components of the feature vectors are not cross-correlated. It can be shown from vector analysis that this, if it is constructed by connecting all the minimum and maximum values. c) A new long-term spectrum from speaker X fits the template, and he is accepted. d) Long-term spectra from speakers Y and Z do not fit the template. These speakers are rejected.
indeed possible, is only the case in one particular po-

sition of the system of basic vectors of the vector space.
The basic vectors are brought into this position by
rotations in accordance with data provided by the
covariance matrix mentioned above. In this way decor-
related, 'whitened' feature vectors are obtained. Applying
the minimum-distance classifier with Euclidean
distance to the whitened data corresponds to 'Ma-
halanobis classification' [8]; in our tests this yielded
error-free identification for all the speakers tested.

minimum-risk classifier (with histogram approx-
imation) were considerably faster and required far less
storage capacity, while the recognition rates did not drop below 98.5%.

Speaker identification

For the recognition experiments the 5000 processed
utterances have been divided into two data bases of
2500 utterances each. This allowed code-word-related
recognition and text-independent recognition to be in-
vestigated separately. The speaker-identification tests
were performed with 'economical' pattern recognition
algorithms. All the speakers tested were known to the
system and the system was made to classify any ut-
terance presented; rejection as 'unclassifiable' was not
allowed. Recognition rates of 99.5% were achieved for
both data bases, code-word related as well as text-
independent. For training the classifiers 10 to 20
utterances seem to be necessary to obtain this high
recognition accuracy.

Fig. 8. Long-term-averaged frequency spectra of two male voices (a) and two female voices (b).
The spectra, which cover frequencies between 100 Hz and 6000 Hz, are obtained by scanning
a 43-channel filter bank 50 times a second and adding the output voltages for every separate
filter in a multiplex digital integrator. Because of their higher pitch the female voices do not
give energy in the lowest three frequency channels.

Whenever recognition rates are being compared, it
is necessary to consider all the assumptions and bound-
ary conditions. In speaker-identification tests the
Mahalanobis classifier achieved recognition rates
of 100% for 2500 utterances of 50 speakers; the ut-
terances were the same apart from the name of the
speaker given in each utterance. However, the Ma-
halanobis classifier is a rather 'expensive' algorithm
in storage capacity and computer time. 'Economical'
classifiers like the minimum-distance classifier or
**Speaker verification**

While in speaker identification only one rate — that for correct classification — has to be considered, a speaker-verification experiment is characterized by two error rates; the rate for false acceptation (mix-up) and the rate for false rejection. There is a functional relation between these two error rates, given by the decision threshold. During the training phase of speaker verification, in addition to the reference pattern, this speakers will be falsely rejected because of the variability of their utterances. On the other hand, if the threshold value is too high, speakers with similar voices will be falsely accepted and thus be mixed up with the real speakers. So the value of the threshold is a trade-off between security of the system (low false-accept rate) and customer convenience (low false-reject rate). The interdependence of these three variables is shown in fig. 10 for a minimum-distance classifier.

Fig. 9. Cross-correlation between the frequency channel outputs of the filter bank is expressed by vertical deviation in the figure; \( n \) is the channel number. The figure is the average for 2500 long-term spectra. The highest correlation is between adjacent channels.

Various methods for decision-threshold evaluation by the computer have been compared. The most effective method is to evaluate the mean of the distance measure determined by the absence of correlation (see above, p. 213). First, the mean is taken within each class, for the distance between each pair of points, and then the results are averaged over all the classes.

---

The distance measure referred to here is \( d_e \), expressed by eq. (2). The threshold value obtained by averaging \( d_e \) is

\[
Th_{NN} = \frac{R}{S} \sum_{k=1}^{S} \frac{1}{M-1} \sum_{j=1}^{M} \sum_{i=1}^{M} \left( 1 - \frac{N}{\sum_{i=1}^{N} x_{ik} x_{il}} \right),
\]

(3)

where \( S \) is the number of speaker classes and \( M \) the number of utterances per class. \( R \) is a scaling factor allowing the threshold value to be shifted along the horizontal axis in fig. 10 to obtain the desired false accept/false reject ratio. Generally \( R = 1 \); in particular cases, however, \( R \) is adjusted by hand to another value.

The effects of transmission by telephone line

For commercial application of a speaker-recognition system it is important to know how telephone transmission of the speech signal affects the security.

Transmitting speech over a telephone channel means frequency-band limitation from 300 Hz to 3500 Hz. The transfer characteristic in this frequency range is not flat but changes its shape depending on the line selected by dialling. The spectra available from telephone lines are therefore band-limited and also multiplied by a transfer function of unknown shape. As a first step it could be shown that band-limiting the long-term spectra to telephone quality did not affect the recognition rate significantly. However, weighting the spectra by arbitrary transfer functions made recognition based on long-term-averaged spectra unreliable because in some cases the effect of the transfer function on the spectra outweighed the voice characteristics. To overcome this difficulty a novel feature set has been introduced to eliminate the effect of the unknown transfer functions. These features, which we call the modified standard-deviation profiles (MSP), are a function of the voice characteristics only and are not affected by the channel transfer function.

The \( j \)-th component \( S_{bj} \) of this novel feature vector is expressed by

\[
S_{bj} = \frac{1}{T} \sum_{i=1}^{T} a_j x_{ij} - \frac{1}{T} \sum_{i=1}^{T} a_{j0} x_{ij}
\]

(4)

In the numerator a modified standard deviation of the \( j \)-th spectral component of the \( i \)-th short-term spectrum is evaluated.
The effect of a hypothetical, erratic telephone-line transfer function (c or f) on a long-term spectrum (a) is shown by (d) and (g) respectively, the effect on the modified standard-deviation profile (b) is shown to be non-existent (e, f). First recognition experiments using the modified standard-deviation profile as a feature vector yield promising results.

The work described was sponsored by the German Federal Ministry for Research and Technology under contract No. 0812014 Kap. No. 3004 Titel 68301. Responsibility for the contents of this publication rests with the author.

Summary. AUROS, an experimental system for speaker recognition by computer, has been developed at Philips Forschungslaboratorium Hamburg, in West Germany, in a government-sponsored research project. The system comprises several speech-signal processors which extract different characteristic features from the signal in real time, and can also use procedures such as the fast Fourier transform, the Walsh-Hadamard transform, and computation of the autocorrelation and cepstrum functions. A general-purpose computer is programmed to compare the features with the stored data from a group of known speakers; 5000 utterances from 100 speakers are stored in a data base. Several different measures of similarity are being tried. Some of these are more economical than others in computer time and storage capacity, but yield slightly higher error rates. In speaker-identification experiments using economical classifiers recognition rates of 99.5% have been obtained. Experiments where a speaker's claimed identity is verified yield error rates of about 1% for false acceptance and false rejection.
Recent scientific publications

These publications are contributed by staff of laboratories and plants which form part of or cooperate with enterprises of the Philips group of companies, particularly by staff of the following research laboratories:

Philips Research Laboratories, Eindhoven, The Netherlands
Philips Research Laboratories, Redhill, Surrey, England
Laboratoires d'Electronique et de Physique Appliquée, 3 avenue Descartes, 94450 Limeil-Brévannes, France
Philips GmbH Forschungslaboratorium Aachen, Weißhausstraße, 51 Aachen, Germany
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium
Philips Laboratories, 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A. (by contract with the North American Philips Corp.)

Reprints of most of these publications will be available in the near future. Requests for reprints should be addressed to the respective laboratories (see the code letter) or to Philips Research Laboratories, Eindhoven, The Netherlands.

W. J. Bartels, L. Blok & C. W. Th. Bulle: X-ray topography and diode efficiency of vapour grown GaAs$_{1-x}$P$_{2}$ layers.
J. Crystal Growth 34, 181-188, 1976 (No. 2).

V. Belevitch: Theory of the proximity effect in multiwire cables, Part II.

C. Belin: On the growth of large single crystals of calcite by travelling solvent zone melting.
J. Crystal Growth 34, 341-344, 1976 (No. 2).

C. H. J. van den Brekel (I, II) & J. Bloem (II): Characterization of chemical vapour-deposition processes, Parts I and II.

A. M. van Diepen: The B-site Mössbauer linewidth in Fe$_{2}$O$_{3}$.


M. Gleria & R. Memming: Novel luminescence generation by electron transfer from semiconductor electrodes to ruthenium-bipyridil complexes.

H. Hieber: Aging properties of gold layers with different adhesion layers.
Thin Solid Films 37, 335-343, 1976 (No. 3).

A. Humbert, L. Hollan & D. Bois (I.N.S.A. Lyon, Villeurbanne): Influence of the growth conditions on the incorporation of deep levels in VPE GaAs.
J. appl. Phys. 47, 4137-4144, 1976 (No. 9).

A. J. R. de Kock: Characterization and elimination of defects in silicon.

Philips Res. Repts. 32, 82-95, 1977 (No. 2).

A. Milch: Etch polishing of GaP single crystals by aqueous solutions of chlorine and iodine.

C. Mulder & H. E. J. Wulms: High speed integrated injection logic (IIL).

J. M. S. Schofield: The physics of gas discharge cells within d.c. memory display panels.
4th Int. Conf. on Gas discharges, Swansea 1976 (IEE Conf. Publn No. 143), pp. 397-400.

A. L. N. Stevels: On the luminescence of CsI: Mn and CsBr: Mn.
Philips Res. Repts. 32, 77-81, 1977 (No. 2).

A. L. N. Stevels & A. D. M. Schrama-de Pauw: Eu$^{2+}$ luminescence in hexagonal aluminates containing large divalent or trivalent cations.

R. P. Tijburg & T. van Dongen: Selective etching of III-V compounds with redox systems.
Recent United States Patents

Abstracts from patents that describe inventions from the following research laboratories that form part or cooperate with the Philips group of companies:

Philips Research Laboratories, Eindhoven, The Netherlands
Philips Research Laboratories, Redhill, Surrey, England
Laboratoires d'Electronic et de Physique Appliquee, 3 avenue Descartes, 94450 Limeil-Brévannes, France
Philips GmbH Forschungslaboratorium Aachen, Weiβhausstraße, 51 Aachen, Germany
Philips GmbH Forschungslaboratorium Hamburg, Vogt-Kölln-Straße 30, 2000 Hamburg 54, Germany
MBLE Laboratoire de Recherches, 2 avenue Van Becelaere, 1170 Brussels (Boitsfort), Belgium
Philips Laboratories, N.A.P.C., 345 Scarborough Road, Briarcliff Manor, N.Y. 10510, U.S.A.

4052340
Method for reproducing a voltage dependent resistor and a voltage dependent resistor obtained therewith

R. K. Eijnthoven
J. T. C. van Kemenade
Voltage dependent resistor obtained by sintering a body of a mixture of Zno and other metal oxides in an atmosphere which contains bismuth.

4052605
Interpolating non-recursive digital filter

L. D. J. Eggermont
An interpolating non-recursive digital filter for generating output signal samples which occur at a given output sampling frequency and which are related in a predetermined way to a sequence of input signal samples has an output sampling frequency which is an integer multiple of the frequency of the input signal samples. In order to make more efficient use of the storage capacity of a storage device in the filter, multiplying coefficients representative of the difference between two samples of the impulse response which belong to different sets but to the same sampling period are used.

4052633
Restorable cold cathode in a gas discharge electron gun

T. M. B. Schoenmakers
Restorable cold cathode in a gas discharge electron ion gun in which the eroded material in the active surface of the cathode is restored by supplying new material in the form of a wire which is moved through a hole in the cathode body by means of a screw spindle.

4052681
Gas-discharge laser

K. Bullhuis
B. J. Derksema
H. T. Dijkstra
J. van der Wal
A gas-discharge laser having a Brewster window near at least one of the reflectors which is secured directly to the laser tube. The Brewster window is secured to a surface of the laser tube and a normal to the window makes an angle equal to the Brewster angle with the axis of the laser tube. The surface may be a slot in the laser tube or a recessed end face of the laser tube near a reflector.

4052683
Microwave device

J. H. C. van Heuven
F. C. de Ronde
A microstrip waveguide transition where the substrate is arranged in a symmetry plane of waveguide and is situated parallel to the field lines of the electrical field and the longitudinal axis of the waveguide. The asymmetric microstrip conductor structure is coupled, via a symmetrical-asymmetrical transformer, to symmetrical band line provided on the substrate. To be conductive for RF energy, the individual conductors of the band line are connected to opposite walls of the waveguide via broadening conductors.

4052707
Magnetic device having domains of two different sizes in a single layer

W. F. Druyvesteyn
H. M. W. Booij
A magnetic device comprising at least one thin layer of a magnetizable material having a preferred direction of magnetisation which is approximately perpendicular to the surface of the layer in which magnetic domains are generated, maintained and possibly annihilated in the layer. A domain guiding structure which with a given magnetic field substantially in the preferred direction enables the occurrence of two types of magnetic domains. The area of the largest domain is at least 15% and at most 125% larger than that of the other domains. The device also has means which convert one type of domain into the other type of domain.

4052747
Device for the magnetic domain storage of data having a shift register filled with coded series of domains

J. Roos
A magnetic bubble domain device for recording information on a magnetizable recording medium including a shift register filled with a series of bubble domains coded in accordance with the information to be recorded, the whole bubble domain pattern being printed, in combination with a magnetic transfer field, in a field time on a recording medium, new information replacing the old information by shifting of the bubble domains in the register.
Magnetoresistive magnetic head
K. E. Kuijk

A magnetic head for detecting information-representing magnetic fields on a magnetic recording medium and comprising an elongate magneto-resistive element of a magnetically anisotropic material which at its ends has contacts for connection to a current or voltage source. In order to linearize the playback characteristic of the element, the easy axis of magnetization coincides with the longitudinal direction of the element and means are present which force the current to travel at an angle of minimum 15° and maximum 75° with the longitudinal direction. These means consist in particular of equipotential strips provided on the element.

Hot-gas reciprocating machine comprising two or more working spaces, provided with a control device for the supply of working medium to the said working spaces
J. H. Abrahams

A hot-gas reciprocating machine involving a plurality of cycles having a mutually different phase, during each crank shaft revolution working medium from a source of pressurized working medium being successively supplied to each cycle separately, via a control device, comprising one or more slides which are controlled exclusively by the variable cycle pressures.

Electro-optic devices
E. T. Keve K. L. Bye

An electro-optic device having a platelet of PLZT material, the birefringence of which depends on an electric field. The thickness of the platelet preferably is smaller than the thickness of one grain. As a result of this the sensitivity is large and this permits a low operating voltage in the device.

Rectifying circuit
R. J. van de Plassche

A rectifying circuit for balanced input currents, of which the two components are applied to a first and a second point of a selective current mirror circuit respectively, either the first point constituting the input and the second point the output, or the first point constituting the output and the second point the input of the selective current mirror circuit, depending on and under control of the polarity of the difference between the two components. As a result, the selective current mirror circuit follows either the greater or the smaller of the two components of the balanced input current. The first and the second point are each connected to the output terminal of the rectifying circuit via a current circuit. These current circuits each have a reverse direction and each comprise the main current path of a transistor for transferring the difference between the output current of the selective current mirror circuit and the component of the balanced input current which is applied to the relevant output to the output terminal in a voltage decoupling manner.

Pyroelectric detector comprising nucleating material wettable by aqueous solution of pyroelectric material
A. A. Turnbull H. Sewell

A pyroelectric detector employing a substrate supporting a thin, i.e., 0.5 to 5 μm thick, solid layer of pyroelectric material with an intermediate layer of nucleating material, i.e., a material which is wettable by a solution of the pyroelectric material so that an adherent continuous layer is formed thereon. The pyroelectric layer may be in the form of a mosaic of islands separated by an electrically conductive material covered with an electrically insulating material.

Doppler radar system
K. Holford

A Doppler radar system for controlling portable traffic signals in response to on-coming traffic. Each of two channel amplifiers of the system is fixed at high gain and passes both noise signals and Doppler signals to a phase detector. A threshold element provides a control signal when the average level of the phase detector output between high and low levels changes sufficiently, due to the presence of Doppler signals, from a mean level which is due to noise alone.

Device for liquefying gases
P. S. Admiraal

A liquefactor includes a refrigeration stage for cooling a compressed gaseous body, and a first duct containing a first Joule-Thompson valve for connecting the refrigeration stage to a collecting container for use when the gaseous body comprises a single gas. A second duct parallely connects the refrigeration stage to the collecting container and contains a second Joule-Thompson valve for use when the gaseous body comprises a mixture of two gases to be separated.

Integrated injection logic memory circuit
C. M. Hart A. Slob

A new integrated circuit in which bias currents are supplied by means of a current injector, a multi-layer structure in which current is supplied, by means of injection and collection of charge carriers via rectifying junctions, to zones to be biased of circuit elements of the circuit, preferably in the form of charge carriers which are collected by the zones to be biased themselves from one of the layers of the current injector. By means of said current injector circuit arrangements can be realized without load resistors being necessary, while the wiring pattern may be very simple and the packing density of the circuit elements may be very high. In addition a simple method of manufacturing with comparatively few operations can in many cases be used in particular upon application of transistors having a structure which is inverted relative to the conventional structure.

Hot-gas reciprocating engine
A. M. Nederlof

A hot-gas reciprocating engine in which the tubular connection members interconnect a heater duct inside the heat pipe with the engine's expansion space and regenerator, each connection member extends transverse of the center line of the relevant unit, and is connected to the heat pipe wall with a flexible sealing member.

Servo system for controlling the position of a reading head
J. de Boer A. Walraven

A servo system for controlling the position of a magnetic reading head relative to the center of a selected information track. During recording a long-wave positioning signal is recorded below the
data signal in the tracks. Upon reading out, the head not only reads the information of the selected track but, as result of cross-talk, also the positioning signals of the adjacent tracks. After filtering out and processing the positioning signals, a control signal for controlling the head is obtained.

4 0 5 7 0 6 3
Device for sterilization by transuterine tube coagulation
A. C. M. Gieles
G. H. J. Somers
A device for sterilizing human females by transuterine fallopian tube coagulation wherein the substantial increase in impedance of the tissues during coagulation is used to signal for termination of treatment.

4 0 5 7 7 2 5
Device for measuring local radiation absorption in a body
W. Wagner
H.-G. Junginger
A device for measuring the spatial distribution of radiation absorption in a body wherein a multiplicity of radiators are regularly distributed about a circle surrounding the body, each radiator emitting a wedge-shaped beam of radiation in the plane of the circle toward a different arc portion of the circle between two other radiators, a multiplicity of adjoining detectors in each arc portion measuring radiation from the radiator emitting radiation to that arc portion, the spatial distribution of radiation absorption being calculated from the measured radiation values of all the detectors.

4 0 5 7 7 2 8
X-ray exposure device comprising a gas-filled chamber
K. Peschmann
H.-G. Junginger
An X-ray exposure device comprising a flat and plane rectangular chamber containing an ionizable gas and having walls provided with electrode structures which generate a potential distribution corresponding to that of two concentric spherical electrodes, an insulating foil on which charge carriers resulting from ionization of the gas by the X-radiation and displaceable in the longitudinal direction of the chamber being arranged therewithin.

4 0 5 7 7 9 6
Analog-digital converter
A. Hoogendoorn
R. E. J. van de Grift
T. J. van Kessel
An analog-digital converter in which a capacitor is charged or discharged by a reference current with the aid of transistor switches, a continuous current flows through the capacitor which is determined by an input difference voltage. One side of the capacitor is connected via a comparator to a flip-flop, to which flip-flop a clock signal is applied and which flip-flop drives the transistor switches. The other side of the capacitor is connected to the first input of a differential amplifier, which differential amplifier forms part of a negative feedback loop which maintains the voltage at the first input of the differential amplifier equal to the voltage which is applied to the second input of the differential amplifier.

4 0 5 7 8 3 1
Video record disc manufactured by a process involving chemical or sputter etching
B. A. J. Jacobs
J. van der Wal
G. B. Gerritsen
A mother for manufacturing long-playing video records is provided by (1) selectively exposing a photoresist disposed as the outer layer on a substrate comprising a disc-shaped plate, a thin layer of base material, e.g., an oxide or nitride, adhering to the plate and a thin metal layer, e.g., chromium, silver, nickel or titanium, coating the base material layer, (2) removing non-activated sections of the photoresist layer and (3) sputter or chemically etching the thin metal and base material layers in sections corresponding to the removed sections of the photoresist layer.

4 0 5 7 8 3 3
Centering detection system for an apparatus for playing optically readable record carriers
J. J. M. Braat
An apparatus as described for reading a record carrier on which information, for example video and/or audio information is stored in an optically readable track-shaped information structure. A deviation between the center of a read spot which is projected on the information structure and the center line of a track to be read can be detected with the aid of at least two detectors which are disposed in the far field of the information structure in different quadrants. With the aid of the same detectors a reference signal is obtained which is used for deriving a control signal for correcting the position of the read spot relative to the track to be read.

4 0 5 8 3 8 2
Hot-gas reciprocating machine with self-centered free piston
J. Mulder
A hot-gas reciprocating machine having a free piston, one face of which varies the volume of a working space while its other face bounds a buffer space of constant pressure. A control mechanism maintains a constant nominal central piston position by momentarily connecting the working space and the buffer space.

4 0 5 8 7 4 3
Pulse generating circuit
D. R. Armstrong
A pulse is generated in an output transformer and this pulse is used to provide a spark across a spark gap for the purpose of igniting any gas/air mixture present. The circuit operates from a 1.5 volt d.c. source.
A blocking oscillator is energized by a two position switch which is operated to a first position, and this charges a capacitor. The capacitor is then discharged through an output transformer to produce the spark when the switch is returned to its original rest position. The winding used to charge the capacitor from the blocking oscillator is also used in the discharge path for the capacitor and this inductive loading provides a slower discharge and a more controlled lower energy spark which is found to be better for gas ignition.