



Review Article

Multi-Method System for Robust Text-Independent Speaker Verification

Stefan Hadjitodorov^{1,2*} 

¹Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Block 105, 1113 Sofia, Bulgaria

²Center for National Security and Defence Research, Bulgarian Academy of Sciences, 1, "15 November" Str., 1040 Sofia, Bulgaria

Received: 08 August, 2025

Accepted: 21 August, 2025

Published: 22 August, 2025

***Corresponding author:** Stefan Hadjitodorov, Center for National Security and Defence Research, Bulgarian Academy of Sciences, 1, "15 November" Str., 1040 Sofia, Bulgaria,
E-mail: sthadj@cu.bas.bg; sthadj@bas.bg

Keywords: Speaker verification; Neural networks; AR-models; Cohort normalization

Copyright: © 2025 Hadjitodorov S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://www.engineergroup.us>



Abstract

A system for robust speaker verification based on four recognition approaches and methods (classifiers) is proposed, in order to use different statistical characteristics of the speech parameters. These methods are: 1) Prototype Distribution Maps (PDM); 2) AR-vector models (ARVM); 3) Two-level approach: the first level uses several PDMs for preprocessing, and the second employs multilayer perceptron (MLP) networks; 4) Gaussian speaker's models combined with the arithmetic-harmonic sphericity measure (GMAHSM).

These classifiers generate four preliminary classification decisions. The reliability and confidence of these preliminary decisions are evaluated by means of a weighting algorithm. The weights are assigned using the relative measures to the most similar speakers (or cohorts), i.e. a Cohort Normalization technique is implemented. The final classification is then performed using simple logical and threshold rules.

The speech signals of 92 speakers have been analyzed. The speaker verification accuracy was over 98%. Robust impostors' detection was observed, because the classifiers never fail simultaneously.

Introduction

There are many methods and approaches [1-14] for pattern recognition used for speaker recognition (identification and verification). However, each approach has its advantages and limitations. In order to use the advantages of several methods, a hybrid system for speaker verification, based on different approaches for pattern recognition is proposed. The recognition is made by means of four different recognition methods (classifiers):

- 1) Prototype distribution map (PDM) [8,9].
- 2) AR-vector model (ARVM) [2-4,10,13].
- 3) Two-level classifier: PDM and multilayer perceptron network [9].

- 4) Gaussian speaker's models combined with the arithmetic-harmonic sphericity measure (GMAHSM) [4,7,13,14].

These classifiers provide four preliminary decisions (classifications). The reliability and confidence of these preliminary decisions are evaluated by means of a weighting algorithm. The weights are assigned using the relative measures to the most similar speakers (or cohorts), i.e. a cohort normalization technique is implemented. The final classification then is done on the base of simple logical and threshold rules.

Evaluation of the feature vectors (speech parameters)

In order to minimize the errors during the speech

parameters evaluation, the following procedure for speech analysis is proposed and used:

Segmentation of the speech signal

The quantized signal is divided into segments with length of three 'To' using a Hamming window. The duration of the segments is dynamically adapted to 3 To (using 'To' from the previous segment), because our experiments [15] have shown that such segment's length is optimal for 'To' evaluation. The overlapping between the segments is two pitch periods in order to analyze the dynamics of the speech parameters. The length of the first segment is 30ms in order to ensure that the window contains at least two To.

Periodicity/apperiodicity separation

The periodicity/apperiodicity separation (PAS) is very important for correct To detection (because errors in PAS will produce drastic errors in To) and for correct evaluation of the speech parameters. To minimize the number of errors in PAS the detector proposed in [16] is implemented, because it is characterized by:

1. Parallel analysis of the speech in time, spectral and cepstral domains. In this way different characteristics of the signal in these domains are used and the signal is analyzed more completely and from different viewpoints.
2. Realization of robust PAS by means of multilayer perceptron (MLP) neural network. As a result the accuracy is improved, because the MLP are characterized by good discriminant capabilities and high classification power.

In order to minimize the influence of the noisy components the aperiodic segments are eliminated.

Pitch period (To) evaluation

In order to evaluate correctly To the robust hybrid pitch detector [17] is used because it has the following useful properties:

- a) Rejects practically most of the segments, where To is wrongly evaluated (experimental research over 200 speakers [15] when the signal is preprocessed by the PAS detector [16]. However it eliminates up to 1% of the voiced segments. The loss of these segments may be tolerated, because for all the speakers (in our database) the analyzed sentences are relatively longer and always contain more than 200 segments.
- b) Evaluates correctly To from clean, noisy and telephone speech;
- c) Realizes parallel analysis of the speech signal in temporal, spectral and cepstral domains;
- d) Evaluates the pitch period by means of logical analysis of the results from these three domains.

For every p-th segment (containing 3 To) the mean pitch period (Tom(p)) is calculated.

Cepstral analysis (over the voiced segments)

Many experimental studies [5,18–20] have shown that the LP-derived cepstral coefficients ($c(n)$) are very informative for speaker recognition. In the proposed approach the $c(n)$ are calculated for voiced segments in order to minimize the influence of the noisy components. The cepstral analysis is carried out by means of the standard procedure LPC analysis by means of the autocorrelation method [21] and then the first 16 LPCC coefficients are calculated. The number of $c(n)$ is 16, because our experimental research [15] has shown that the first 16 coefficients are the most informative for speaker recognition for our database.

Evaluation of the group delay function (for the voiced segments)

It is known [15,22] that the analysis of spectrograms and sonograms (representing the formant structure) is very useful for speaker recognition. However, the estimation of formants remains a difficult problem that is not yet fully solved. That is why the formant structure is analyzed by evaluation and analysis of the group delay function (GDF) (GDF is the negative derivative of the phase spectrum). The GDF is used for approximation of the formant structure, because the GDF has the following useful properties [23]:

1. The GDF is proportional to the squared magnitude response near resonances (formants) and approaches zero asymptotically for frequencies away from the frequency of the resonator. In this way the formants are represented by distinct and sharp peaks in the GDF.
2. The vocal tract may be represented by a cascade of resonators. The GDF of such system is the sum of GDFs of these resonators. As result the influence of one resonator to another is minimized – even closely spaced formants are represented in the GDF by separated peaks.

The main problems in calculation of the GDF are:

1. The phase function is warped by the presence of zeros near the unit circle and the signal windowing prior to the spectral analysis [23].
2. Most methods for phase unwrapping do not yield satisfactory results [24–27].
3. The spectral resolution in the GDF for medium and high pitched voices ($To < 5$ ms and $Fo > 200$ Hz), when GDF is calculated from $c(n)$, is decreased. This is due to the short To minimizing the number of $c(n)$ and respectively the number of GDF coefficients ($gdf(i)$). For To, shorter than 5 ms the spectral resolution will be less than 262 Hz, because in our experiments the sampling rate is 21 KHz.
4. Distortions (represented in most of the cases by extra peaks) caused by the influence of the glottal source.

In order to solve these problems and to guarantee spectral resolution higher than in the standard wide band sonogram used for formant analysis (here 262 Hz), the following approach for GDF calculation is proposed and used:

A) Analysis of low pitched voices: For such voices the GDF is calculated indirectly (without phase calculation and unwarping) from $c(n)$ – detailed proof is given in [23]. In order to minimize the influence of the glottal source only the cepstral coefficients ($cv(n)$) corresponding to the vocal tract are separated by means of liftering. The length (L) of the lifter is less than T_0 . The value of $L = 0.8T_0$, because our experiments [15] have shown that such a length is sufficient for suppression of the influence of the glottal source. The GDF is calculated by means of the formulae described in (Murthy, 1989) from the liftered $c(n)$.

B) Analysis of high pitched voices: For such voices the GDF is calculated by means of the following procedure described in (Duncan, 1989):

- Transformation of the voiced speech into minimum phase signal;
- Direct calculation of the phase spectrum from this minimum phase signal;
- Calculation of the first derivative of the phase spectrum – the GDF.

Unfortunately, for high-pitched voices, the influence of the glottal source is not suppressed. However, in most of the practical cases during speaker recognition male voices are analyzed, generally characterized with low values of the pitch period.

The first S $gdf(i)$ coefficients (for $I = 1, \dots, S$) are used as feature vectors representing the formant structure. The value of S is determined on the basis of baseband (300–3000 Hz) of the phone lines, because one of the main applications of this system will be speaker recognition over phone lines. To cover this spectral range the value of S is 12, because in our experiments the resolution in the GDF is 262 Hz.

Formation of the input vectors

The following input vectors are formed for every p -th (for $p = 1, \dots, P$) voiced segment: mean pitch period ($Tom(p)$); first 16 $c(i)$ and the first 12 $gdf(i)$.

The multi-method approach for pattern recognition

Four methods for speaker recognition are used and for all these methods the measures (scores) – similarities or distances, are calculated to all the speakers, i.e. besides to the verified speaker. These scores are used in the final verification scheme for cohort normalization.

In order to use different statistical characteristics, the static and dynamic information of the speech parameters in the

system are captured using the following methods for speaker recognition:

The prototype distribution map (PDM) - detailed description in [8,9,15]

The PDM is used because [9,15]:

- The PDM's neurons try to imitate the probability density function (pdf) whatever complex the form of the pdf is;
- Less significant neurons in the PDM are eliminated by a filtering procedure.

Training the PDM: For each of the M speakers from the database a $PDM(m)$, ($m = 1, \dots, M$, M – number of known speakers) is formed using the training procedure [8] For every k -th available utterance of a given speaker one separate $PDM^k(m)$ is formed.

Speaker verification: The following classification procedure is used:

- formation and filtering of the PDMU of the unknown speaker (speaker U);
- calculation of the similarities ($Sim(m)$) between the PDMU and each $PDM(m)$ by means of the measure proposed in [8];
- The speaker U is preliminary accepted as the speaker l to whom he has maximal value of $Sim(m)$. This speaker could be different from the speaker c , who he claimed to be (in case of separately used method that means –rejection).

The AR-vector models (ARVM) - detailed description in [3,10,13]

The ARVM are used because they allow modeling speaker's specific information even when the segments with features vectors are analyzed in a random order [3,10,13].

Training the ARVM: During the training for all the speakers their $ARVM(m)$ are formed using the procedure [3].

Speaker verification: The classification is done by formation of the ARVM of the unknown speaker— $ARVM(U)$ and calculation of the forward–backward symmetrized Itakura distances [3] between $ARVM(U)$ and all $ARVM(m)$. The speaker U is preliminary accepted as the speaker l to whom he has minimal Itakura distance. This speaker could be different from speaker c , who he claimed to be (in case of separately used method that means –rejection).

The Gaussian speaker's models combined with the arithmetic-harmonic sphericity measure (GMAHSM) - detailed description in [4,14]

The Gaussian mixture speaker's models are used because:

- These models allow robust speaker recognition when noisy and telephone speech signals are analyzed [14].

- b) The Gaussian speaker's models may be combined with the arithmetic-harmonic sphericity measure - AHSM [4]. This measure is useful for speaker recognition because:

The AHSM is symmetric and is evaluated efficiently, without extracting the eigenvalues.

The AHSM allows speaker recognition with minimal error in comparison with other measures when noisy speech signals are analyzed [4].

Training: The covariance matrices (COV(m)) for the speakers are calculated.

Speaker verification: The covariance matrix (COV(U)) of the unknown speaker (U) is calculated. The ASHM between all the speakers and the unknown speaker are evaluated. The speaker U is preliminary accepted as the speaker *l* to whom he has the minimal AHSM distance. This speaker could be different from speaker *c*, who he claimed to be (in case of separately used method that means - rejection).

The two level classifier - detailed description in [8,9,15]

The two-level classifier incorporates the pdf's estimating, statistical modeling and compressing powers of the PDM with the discriminant capabilities and classification power of the MLP. As a result the classifier is better than either PDM or MLP, especially in the case of noise-corrupted signals.

- A. The first level of the classifier consists of several (*T*, $T > 1$) PDMs. The reason for building *T* PDMs is that the BP provides "good" estimates of Bayes *a posteriori* probability functions only if the MLP has enough flexibility to closely approximate the Bayes functions and there is sufficient training data.

Training: A single PDM could not adequately train the MLP and that is why for each speaker's utterance its PDMs are trained.

Speaker verification: During this stage for the unknown speaker U, its PDMs are obtained.

- B. The second level of the classifier, consists of a MLP. The MLP (for each speaker) are trained using the above mentioned PDMs. The MLPs are trained with supervision using the back-propagation (BP) algorithm, which minimizes the squared error between the actual outputs of the network and the desired outputs.

Training: The input feature vectors (the PDMs) of a given speaker are labeled as "one" and feature vectors of the remaining speakers as "zero".

Speaker verification: All the test vectors for an unknown speaker (the PDMs) are applied to each MLP. The outputs of the MLP of every speaker are accumulated. The speaker is preliminary accepted as the speaker *l* if the corresponding MLP(*l*) is with the maximum accumulated output. This speaker could be different from speaker *c* who he claimed to be (in case of separately used method that means - rejection).

The architecture of the used MLP networks is with two hidden layers and one output and the nodes have sigmoid nonlinearities.

Final (hybrid) verification scheme

The following procedure for final decision is proposed and used:

Evaluation of the functions of the cohort score(s) for the classifiers: The following procedure for assigning of classifier's weights is proposed and used:

Each classifier's decision gets a labeled weight $W^l(i)$ (for $i = 1, 2, 3, 4$, corresponding to methods 1, 2, 3 and 4) according to the formulae:

$$w^l(i) = f(CS) / VS(i), \quad (1)$$

for $I = 2, 4$

$$w^l(i) = VS(i) / f(CS), \quad (2)$$

for $I = 1, 3$

where:

$f(CS)$ - function of the cohort score(s) for the *i*-th classifier;

$VS(i)$ - the preliminary verification score;

i - argument (label) of D_{min} in methods 2 and 4, and S_{max} in methods 1 and 3.

The result (output) for methods 2 and 4 is two minimal distances (D_{min}), and for methods 1 and 3, two maximal similarities (S_{max}). In order to preserve the same relative values of the weights, the weights for the methods 1 and 3 are inverted. As a result the functions of the cohort scores ($f(CS)$) are evaluated by means of the following algorithm:

$$f(CS) = \min(\text{cohort scores}), \text{ for } I = 2, 4 \quad (3)$$

Calculation of the averaged sum of *N* minimal cohort scores (for $I = 2, 4$):

$$f_l(CS) = \frac{1}{N} \sum_{j=1}^N CS_j, \quad (4)$$

$$F_l(CS) = \max(\text{cohort scores}), \text{ for } I = 1, 3 \quad (5)$$

Calculation of the averaged sum of *N* maximal cohort scores:

$$f_l(CS) = \frac{1}{N} \sum_{j=1}^N CS_j, \quad (6)$$

Normalization of classifier's weights

$$W^l(i) = w^l(i) / \sum_{i=1}^4 w_i^l \quad (7)$$

where: for $i = 1, 2, 3, 4$.

Final decision by means of specific rules: The following specific rules for final decision are proposed and applied:

Let us assume that the speaker under verification is the c -th speaker. In general, as mentioned above the label l in w_i^l could be different from c . Then the final decision is made according to the rules:

A. Successful verification (acceptance):

if at least two classification methods accept the speaker under verification,

i.e. $l_i = c$ and $w^c = \sum_{i=1}^4 w_i^{l=c} > 0.5$ then the verified speaker is accepted.

B. Rejection:

if only one method accept the speaker under verification or

$w^c = \sum_{i=1}^4 w_i^{l=c} > 0.5$ then the verified speaker is rejected.

Experimental research

In order to make a comparison between the implemented methods in the system and their application to different speech conditions, experiments with clean and noisy speech data have been carried out.

The speech data base

A proprietary dataset was used. The speech (clean and over telephone lines in order to account for telephone channel effects and noise) of 92 speakers (48 males and 44 females) have been analyzed. The signals are digitized with 21 KHz and 16 bits per sample using the standard audio board "Sound Blaster" of the PC and saved directly into the computer's memory in order to avoid any distortions caused by the recorders. The training set consists of utterances of the 3 sentences (training sentences): My name is *first name*, *second name*, *family name*; My code is *six digits*; I am a(n) *profession*. The test set consists of utterances of another 3 test sentences: I am *two digits* years old; My mother's name is *first name*; My hobby is *up to three words*. Note that the proposed system is for text-independent speaker verification. All these sentences were uttered once (in Bulgarian) in the abovementioned two conditions by each speaker in six separate sessions, i.e. we have 552 utterances in clean and 552 utterances over telephone in noisy condition. For test of the system Leave-one-out cross-validation (LOOCV) has been applied. The training set is formed by the utterances of the training sentences of 91 speakers. The test set is formed by the utterances of test sentences of all 92 speakers, i.e. the test set always include one impostor. According to LOOCV that scheme is repeated 92 times.

Practical implementation of the methods

For the PDM method, the optimal values were used for the filter coefficient (k) and for the size (Q) the PDMs found in the experimental researches [8,15]: $k = 0.1$ and $Q = 20$.

For the ARVM method the order of AR-Vector Models was 2, because it is shown that for orders greater than 2, the prediction error does not decrease significantly [13](Montacie,1993; Le Floch,1994).

For the Two-level classifier are used the optional values for the number of PDMs (T), k and for Q found in the experimental researches [9,15]: $k = 0.1$, $T = 10$ and $Q = 4$. The architecture of the MLP for each speaker at the second level is: first hidden layer – 34 neurons, second – 2 and output layer – 1 neuron.

In our experiments we apply the functions $f(CS)$ in the forms (3), (4) with $N = 3$, and (5), (6) also with $N = 3$.

Results and conclusions

The results are given in Table 1. It is important to note that almost all the impostors are eliminated. The proposed hybrid classifier has improved verification accuracy due to:

1. Combination of the advantages of four different methods.
2. The discriminating capabilities and classification power of MLP.
3. The PDMs, in fact, perform a transformation and dimensionality reduction and the filtering procedure rejects the sporadically (noncharacteristic) activated locations.
4. The use of the weighting procedure with a cohort normalization technique implemented. This algorithm gives higher weights to the methods for which the ratios (scores for the speaker under verification to the cohort's scores) are higher. That means that higher weights are given to the more reliable and confidential classifiers.

As first step of our research was to show that weighted (on the base of their cohort scores) combination of various methods that use different statistical characteristics of the speech parameters enhance the capability of a system to recognize with higher accuracy. Robust impostor detection was observed, as no case occurred where all classifiers failed simultaneously. Future work will be devoted to apply this approach to other publicly available data sets and to explore it in worse background conditions like significant noise, spoofing and adversarial actions. Attention will be paid also to improve computational efficiency.

Table 1: The Equal Error Rate (EER %) in speaker verification for COV-AHSM, ARVM, PDM, PDM-MLP and the proposed HYBRID METHODS. False Acceptance Rate (FAR) and False Rejection Rate (FRR) are also given.

ERROR	METHOD				
	COV -AHSM	ARVM	PDM	PDM + MLP	HYBRID
FAR%	2.72	3.08	1.81	0.90	0.54
FRR %	2.36	2.71	2.18	1.26	1.26
EER %	2.54	2.89	2	1.08	0.90

References

1. Bahler L, Porter J, Higgins A. Improved Voice Identification using a Nearest-Neighbor Distance Measure. Proc Int Conf Acoust Speech Signal Process. 1994;321-323. Available from: <https://ieeexplore.ieee.org/document/389291>

2. Bennani Y, Gallinari P. On the use of TDNN-extended features information in talker identification. Presented at: Proc IEEE Int Conf Acoust Speech Signal Process. San Diego, USA. 1991;385-388. Available from: <https://ieeexplore.ieee.org/document/150357>
3. Bimbot F, Mathan L, De Lima, Chollet G. Standard and Target Driven AR-Vector Models for Speech Analysis and Speaker Recognition. Presented at: Proc IEEE Int Conf Acoust Speech Signal Process; San Francisco, USA. 1992;5-8.
4. Bimbot F, Magrin-Chagnolleau I, Mathan L. Second-order statistics for text-independent speaker identification. Speech Commun. 1995;17:177-192. Available from: [https://doi.org/10.1016/0167-6393\(95\)00013-E](https://doi.org/10.1016/0167-6393(95)00013-E)
5. Farrell K, Mammone R, Assaleh K. Speaker recognition using neural networks and conventional classifiers. IEEE Trans Speech Audio Process. 1994;2:194-205. Available from: [https://www.scrip.org/\(S\(ny23rubfvg45z345vbrepxl\)\)/reference/referencespapers?referenceid=728658](https://www.scrip.org/(S(ny23rubfvg45z345vbrepxl))/reference/referencespapers?referenceid=728658)
6. Ferderickson S, Roberts S, Townsend N, Tarasemko L. Speaker identification using neural networks of radial basis functions. Signal Process Theor Appl. 1994;812-815.
7. Gish H, Schmidt M, Mielke A. A Robust, Segmental Method for Text Independent Speaker Identification. Proc Int Conf Acoust Speech Signal Process. 1994;145-148.
8. Hadjitodorov B, Boyanov B, Ivanov T, Dalakchieva N. Text-independent speaker identification using neural nets and AR-models. Electron Lett. 1994;30:838-840. Available from: <https://digital-library.theiet.org/doi/10.1049/el%3A19940587>
9. Hadjitodorov B, Boyanov B, Dalakchieva N. A two-level classifier for text-independent speaker identification. Speech Commun. 1997;21:March. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S0167639397000046>
10. Magrin-Chagnolleau I, Wilke J, Bimbot F. A further investigation on AR-Vector Models for text-independent speaker identification. Proc Int Conf Acoust Speech Signal Process. Atlanta, GA, USA. 1996;401-404. Available from: <http://dx.doi.org/10.1109/ICASSP.1996.540300>
11. Mack M, Allen W, Sexton G. Speaker identification using multilayer perceptrons and radial basis functions networks. Neurocomputing. 1994;6:91-108. Available from: <https://researchportal.northumbria.ac.uk/en/publications/speaker-identification-using-multilayer-perceptrons-and-radial-ba>
12. Matsui T, Furui S. Comparison of text-independent speaker recognition methods using VQ-distortion and discrete/continuous HMMs. Presented at: Proc IEEE Int Conf Acoust Speech Signal Process; San Francisco, USA. 1992;157-160. Available from: <http://dx.doi.org/10.1109/ICASSP.1992.226096>
13. Montacie C, Floch J. AR-Vector Models for Free Text Speaker Recognition. Presented at: Proc Int Conf Spoken Language Process; Banff, Alberta, Canada. 1992;475-478. Available from: https://www.isca-archive.org/icslp_1992/montacie92_icslp.html
14. Reynolds D. Speaker identification and verification using Gaussian mixture speaker models. Speech Commun. 1995;17:91-108. Available from: [https://dl.acm.org/doi/10.1016/0167-6393\(95\)00009-D](https://dl.acm.org/doi/10.1016/0167-6393(95)00009-D)
15. Hadjitodorov B, Boyanov B. Speaker recognition. Research report for the National Scientific Fund, Bulgaria. 1997a. (in Bulgarian).
16. Boyanov B, Hadjitodorov S, Chollet G. Robust periodicity/aperiodicity detector. Ann Bulg Acad Sci. 1997a;50:43-46.
17. Boyanov B, Hadjitodorov S, Ivanov T, Chollet G. Robust Hybrid Pitch Detector. Electron Lett. 1993;29:1924-1926. Available from: <https://doi.org/10.1049/el:19931281>
18. Atal B. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. J Acoust Soc Am. 1974;55:1304-1312. Available from: <https://doi.org/10.1121/1.1914702>
19. Assaleh K, Mammone R. Robust cepstral features for speaker identification. Presented at: Proc IEEE Int Conf Acoust Speech Signal Process; San Francisco, USA. 1994;129-132. Available from: <https://doi.org/10.1109/ICASSP.1994.389338>
20. Furui S. Cepstral analysis technique for automatic speaker verification. IEEE Trans Acoust Speech Signal Process. 1981;29:254-272. Available from: <http://dx.doi.org/10.1109/TASSP.1981.1163530>
21. Rabiner L, Schaffer R. Digital processing of speech signals. New Jersey: Prentice Hall; 1978.
22. Hollien H. The acoustics of crime. New York: Plenum Press; 1990. Available from: <https://link.springer.com/book/10.1007/978-1-4899-0673-1>
23. Murthy K, Murthy B, Yegnanarayana B. Formant extraction from phase using weighted group delay function. Electron Lett. 1989;25:1609-1611. Available from: <https://doi.org/10.1049/el:19891080>
24. Nashi M. Phase unwrapping of digital signals. IEEE Trans Acoust Speech Signal Process. 1989;37:1693-1702. Available from: <https://www.scrip.org/reference/referencespapers?referenceid=580595>
25. Lippmann R. Pattern classification using neural networks. IEEE Commun Mag. 1989;5:47-64. Available from: <https://ieeexplore.ieee.org/document/41401>
26. Openshaw J, Sun Z, Mason S. A Comparison of Composite Features under Degraded Speech in Speaker Recognition. Proc Int Conf Acoust Speech Signal Process. 1993;371-374. (Minneapolis, USA). Available from: <https://scispace.com/papers/a-comparison-of-composite-features-under-degraded-speech-in-1jzvs5l2zl>
27. Morgan DP, Scofield CL. Neural Networks and Speech Processing. Norwell, Madison: Kluwer Academic Publishers; 1991. Available from: https://books.google.co.in/books/about/Neural_Networks_and_Speech_Processing.html?id=_MRQAAAAMAJ&redir_esc=y

Discover a bigger Impact and Visibility of your article publication with Peertechz Publications

Highlights

- ❖ Signatory publisher of ORCID
- ❖ Signatory Publisher of DORA (San Francisco Declaration on Research Assessment)
- ❖ Articles archived in worlds' renowned service providers such as Portico, CNKI, AGRIS, TDNet, Base (Bielefeld University Library), CrossRef, Scilit, J-Gate etc.
- ❖ Journals indexed in ICMJE, SHERPA/ROMEO, Google Scholar etc.
- ❖ OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)
- ❖ Dedicated Editorial Board for every journal
- ❖ Accurate and rapid peer-review process
- ❖ Increased citations of published articles through promotions
- ❖ Reduced timeline for article publication

Submit your articles and experience a new surge in publication services
<https://www.peertechzpublications.org/submission>

Peertechz journals wishes everlasting success in your every endeavours.