International Journal of Speech Technology (2018) 21:451–461
https://doi.org/10.1007/s10772-017-9472-7

Higher order information set based features for text-independent speaker identification

Jeevan Medikonda[1] · **Hanmandlu Madasu**[2]

Received: 6 July 2017 / Accepted: 2 November 2017 / Published online: 27 November 2017

Abstract
In this paper Type-2 Information Set (T2IS) features and Hanman Transform (HT) features as Higher Order Information Set (HOIS) based features are proposed for the text independent speaker recognition. The speech signals of different speakers represented by Mel Frequency Cepstral Coefficients (MFCC) are converted into T2IS features and HT features by taking account of the cepstral and temporal possibilistic uncertainties. The features are classified by Improved Hanman Classifier (IHC), Support Vector Machine (SVM) and k-Nearest Neighbours (kNN). The performance of the proposed approaches is tested in terms of speed, computational complexity, memory requirement and accuracy on three datasets namely NIST-2003, VoxForge 2014 speech corpus and VCTK speech corpus and compared with that of the baseline features like MFCC, ΔMFCC, ΔΔMFCC and GFCC under white Gaussian noisy environment at different signal-to-noise ratios. The proposed features have the reduced feature size, computational time, and complexity and also their performance is not degraded under the noisy environment.

Keywords Text-independent speaker recognition · Information set theory · Mel frequency cepstral coefficients · Hanman transform

1 Introduction

Speaker based biometric authentication in forensic and social media applications is emerging as a viable technology because acquisition of data is easy and economical. The system has been adjudged effective in the noise free environment, in the absence of channel variations and in the absence of limited data (Jayanna et al. 2009). The present work focuses on addressing the noisy conditions that pose the real time challenge. The traditional features used for speaker recognition are Mel Frequency Cepstral Coefficients (MFCC) which are sensitive to the noise. These are first introduced by Davis and Mermelstein (Davis and Mermelstein 1980) for word recognition and later many variants of MFCC such asdelta-MFCC and delta–delta MFCC (Kumar et al. 2011) are proposed to make them robust under the noisy environment. Approaches to increase the robustness are attempted by feature normalization such as cepstral mean and variance normalization (CMVN), RASTA filtering (Hermansky and Morgan 1994) and feature warping (Pelecanos and Sridharan 2001). Zhao et al. (2012) have proposed a new speaker feature known as Gammatone Frequency Cepstral Coefficients (GFCC) having more robustness towards noise than that of the commonly used MFCC. But as explained in (Zhao et al. 2013), the frequency scale (Mel scale) employed in the filter bank and the nonlinear rectification (i.e., cubic root) used in the derivation of scale invariant cepstral coefficients provide robust features to counter noise.

Gaussian mixture model (GMM) (Reynolds and Rose 1995, Reynolds1995) is still the most common approach (Togneri and Pullella 2011) for speaker modeling in the text-independent speaker recognition as it is a model based approach. Reynolds et al. (2000) have adapted GMM using Universal Background Model (UBM) for speaker verification system that is found to be efficient.

Fuzzy logic have been used to handle the uncertainty at modeling stage, decision stage and the feature dimensionality reduction stage to yield promising results. In the literature there are a large number of Fuzzy based modeling

✉ Jeevan Medikonda
jeevanmedi@gmail.com

[1] Department of Biomedical Engineering, Manipal University, Manipal, Karnataka 576104, India

[2] Department of Electrical Engineering, IIT Delhi, New Delhi 110016, India

techniques most of which are the fuzzified version of the existing modeling and decision techniques.

Yuan et al. (1993) have developed Fuzzy mathematical algorithm to extract different features of speakers between Line Spectrum Pair Frequencies and Cepstrum derived from linear prediction analysis. Pierre Castellano (Yuan et al. 1993) have utilized the fuzzy set theory that provides thesecond stage (post-processing) classification after an Artificial Neural Network (ANN) that provides a firststage of discrimination. Jawarkar et al. (2011) use the fuzzy min–max neural network for the text independent speaker identification. This network utilizes fuzzy sets as pattern classes. It is a three layer feed forward network that grows adaptively to meet the demands of the problem. It yields good result as compared to GMM.

Ki Yong Lee in (2004) partitions the data spaceinto several disjoint clusters by fuzzy clustering, and then performs PCA using the fuzzy covariance matrix in each cluster. Finally, the GMM for speaker is obtained from the transformed feature vectors with reduced dimension in each cluster. As compared to the conventional GMM with diagonal covariance matrix, the proposed method needs less storage and gives faster results. Lung (2004) extracts features based on wavelet transform derived from fuzzy c-means clustering. It is found that decreasing the number of training frames does not reduce the recognition rate by the fuzzy c-means clustering algorithm. Wang et al. (2008) proposes a local PCA and Kernel-based fuzzy clustering for feature extraction. These methods remove the time pertinence, noise of speech, reduce the feature vector dimension and achieve a best performance as compared to the standard SVM and GMM. Mirhassani et al. (2014) has addressed methods that include the extraction of MFCC with the narrower filter bank followed by a fuzzy-based feature selection method. The proposed features election provides relevant, discriminative, and complementary features. The proposed method can diminish the dimensionality without compromising the speech recognition rate. Pinheiro et al. (2016) describe a novel GMM-UBM based system dealing with the session noise variability problem. The system uses the Type-2 Fuzzy GMM frame work by considering the speaker GMM parameters to be uncertain in an interval.

The fuzzy setsare characterized by a set of information source values and a membership function (MF) that maps the information source values to the membership grades (degrees of belongingness or association) to the set. But we are interested in representing the uncertainty associated with a fuzzy set. The membership grades don't provide the overall uncertainty associated with the fuzzy set. They only can present the degree of association or belongingness of an information source value to a vague concept represented by a MF.

In (Aggarwal and Hanmandlu 2015), Information set theory was proposed to overcome the shortcomings of fuzzy set theory. The first shortcoming is that its elements are pairs. The components of each pair though related but are delinked. The second shortcoming is that it has no provision to represent both probabilistic and possibilistic uncertainties. Hence we use the Information Set theory to handle the possibilistic uncertainty present in speech signal for the text-independent speaker recognition.

1.1 Motivation

Though Information set features have been used for the development of speaker based authentication system, these features cannot take care of the uncertainty in MFCCs fully. Hardly any effort is made in the literature to account for high order uncertainty in the MFCCs. So we are motivated to represent the higher order uncertainty using type-2 Membership Functions (MF) in place of type-1 MFs in the original Information Set features leading to Type-2 Information Set features and also by applying the Hanman transform on the original information source values. As MFCCs derived under noisy environment are subject to higher order uncertainty we are bent upon investing the effectiveness of these approaches in representing this kind of uncertainty.

2 Related topics

2.1 Mel-frequency cepstral coefficients

The frame work to extract Mel-Frequency Cepstral Coefficients (MFCC) from a speech signal as shown in Fig. 1.

Step 1: Pre-process the speech signal s(n) to boost the high frequency components and to eliminate the spectrum tilt by applying the first order high pass filter with $\alpha = 0.97$ as follows:

$$S(z) = 1 - \alpha z^{-1} \tag{1}$$

The above pre-emphasis operation has little impact on imparting robustness to MFCC towards noise. As speech is a quasi-stationary signal, features extracted from the pre-processed signal are not reliable. However the signal is observed to be stationary in a window of small duration and so the features extracted in this widow are reliable. Therefore the signal is divided into frames of 32 ms duration with 16 ms overlapping.

Step 2: Disregard the silent periods using Voice Activity Detection (Jongseo 1999; Ephraim and Malah 1984) and consider the frames only with the voice signals as these contribute less while extracting speaker specific features.

Step 3: Calculate the power spectrum of each frame. This is motivated by the human cochlea (an organ in the ear)

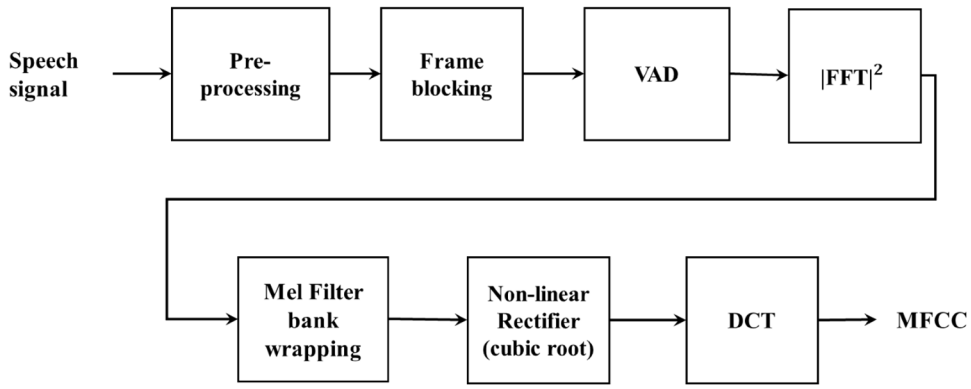

Fig. 1 Frame work to extract MFCC

which vibrates at different spots depending on the frequency of the incoming sounds. Periodogram estimate also helps identifying which frequencies are present in the frame.

Step 4: Take a set of Periodogram bins and compute the energy of each frequency band by applying the Mel filter banks (Davis and Mermelstein 1980) with 40 triangular filters. This is due to the fact that cochlea cannot discern the difference between two closely spaced frequencies. These filter banks are non-linearly placed throughout the bandwidth using Mel scale $\mathcal{M}$ given by:

$$\mathcal{M}(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \tag{2}$$

Step 5: Perform the cubic root operation for non-linear rectification. As proved in (Zhao et al. 2012, 2013) the Mel power spectrum using cubic root operation for non-linear rectification is more noise robust than the log operator.

Step 6: Because of using the overlapping filter banks, the filter bank energies are quite correlated with each other. The DCT decorrelates the energies. But in the traditional MFCC, not all DCT coefficients are considered. This is because the higher DCT coefficients represent fast changes in the filter bank energies that degrade the performance. But in the proposed methodology all the coefficients are considered for delivering an efficient information.

2.2 Information set theory

The Information theoretic entropy function called the Hanman-Anirban entropy function in (Hanmandlu and Das 2011) represents the possibilistic uncertainty by virtue of having parameters in its exponential gain function. This entropy function is generalized by Mamta and Hanmandlu (Mamta et al. 2014) and Medikonda et al. (2016). Out of these the one in (Mamta et al. 2014) is the most general entropy followed by the one in (Medikonda et al. 2016) which is pursued in the present work.

The concept of information set was mooted in (Hanmandlu 2011) and utilized in (Mamta and Hanmandlu 2014) for the recognition of infra-red face. By taking recourse to the Hanman-Anirban entropy function the information set theory eliminates the shortcomings of fuzzy set theory. This allows us to represent the uncertainty in the granularized information source values via the corresponding membership function values. This theory converts the information source values and its membership function values which are pairs in a fuzzy set into the products termed as the information values constituting the information set. The sum of the information values gives the overall uncertainty which we call as the information. A brief description of information set theory is given below.

To seek the conversion of a fuzzy set into the information set consider a set of values, termed as the information source values of an attribute $\Phi = \{\varphi_1, \ldots, \varphi_n\}$. This set is denoted by X_Φ as

$$\mathrm{x}_\Phi = \left\{\mathrm{x}_\Phi(\varphi_i)\right\} \mid \forall \varphi_i \in \Phi \tag{3}$$

The information set theory permits the use of agents by expanding the functionality of a MF called empowered MF, which has a limited role in a fuzzy set. Recall the role of an agent in artificial Intelligence where it is bestowed with perceiving its environment and performing the assigned task accordingly. For example, a robotic agent has sensors that perceive the surroundings and performs simple tasks like pick and place to complex tasks. The empowered MF acting as an agent can do much more than what it can do as MF.

Let us consider the generalized entropy from (Medikonda et al. 2016) in the following form as defined in Eq. (4):

$$\mathrm{I}_\Phi = \sum_i \mathrm{x}_\Phi(\varphi_i)^{\alpha_\Phi} G_\Phi(\varphi_i) \tag{4}$$

The gain function in Eq. (4) is defined as follows:

$$G_\Phi(\varphi_i) = e^{-(a_\Phi \mathrm{x}_\Phi(\varphi_i) + b_\Phi)^{\beta_\Phi}} \tag{5}$$

where $\{a_\Phi, b_\Phi, \beta_\Phi\}$ are the real valued parameters, assumed to be variables to make the entropy function adaptive.

We have employed Eq. (4) by taking G_Φ from Eq. (5) and $\alpha_\Phi = 1$. This gain function G_Φ can take any form of MF including the commonly used membership functions by an appropriate choice of parameters, $\{a_\Phi, b_\Phi, \beta_\Phi\}$.

2.2.1 Gain function as the membership function

Taking $a_\Phi = 1/\sqrt{2}\sigma$, $b_\Phi = -\mu/\sqrt{2}\sigma$, $\beta_\Phi = 2$ in Eq. (5) the gain function becomes the Gaussian membership function where μ and σ are the mean and standard deviation associated with x_Φ. The choice of parameters involving mean and variance in the gain function $G_\Phi(\varphi_i)$ helps convert it into the Gaussian function representing the possibilistic distribution of the information source values. This function called the Gaussian MF $\mathcal{G}_\Phi(\varphi_i)$ in the parlance of a fuzzy set is given by:

$$\mathcal{G}_\Phi(\varphi_i) = e^{-\left(\frac{\mathrm{x}_\Phi(\varphi_i)-\mu}{\sqrt{2}\sigma}\right)^2} \tag{6}$$

The versatility of the entropy function in Eq. (4) is that the information source values $\{\mathrm{x}_\Phi(\varphi_i)\}$ can be taken from any domain, say, probabilistic, possibilistic or a combination of both. Thus we have eliminated two shortcomings of a fuzzy set firstly by connecting the information source value and its MF value that form a pair as the product called the information value and secondly by extending the entropy function to deal with both possibilistic and probabilistic uncertainties.

The information value $\mathrm{I}_\Phi(\varphi_i)$ corresponding to the information source value $\mathrm{x}_\Phi(\varphi_i)$ is computed using the generalized entropy function of (Medikonda et al. 2016) as

$$\mathrm{I}_\Phi(\varphi_i) = \mathrm{x}_\Phi(\varphi_i)^{\alpha_\Phi}\mathcal{G}_\Phi(\varphi_i) \tag{7}$$

The set of these information values constitutes the information set S_Φ given by

$$\mathcal{H}_\Phi = \{\mathrm{I}_\Phi(\varphi_i)\} \mid \forall \varphi_i \in \Phi \tag{8}$$

The sum of the information values in an information set $\mathcal{H}_\Phi$ is called the information denoted by I_Φ. Thus, the normalized effective information H_Φ of the collected information source values is given by

$$H_\Phi = \frac{\sum \mathcal{H}_\Phi}{n}, \quad \forall \varphi_i \in \Phi \tag{9}$$

3 Proposed higher order information set based features

We now present two approaches for representing higher order uncertainty in MFCCs. In the first approach the basic information values are modified by replacing type-1 MF with type-2 MF. The features obtained from this approach are called type-2 Information set features. In the second approach we compute the features by applying the Hanman transform directly on the information source values. These approaches are now discussed in detail.

3.1 Type-2 information set features

In this, we adapt the Mamdani type fuzzy rule to define the corresponding information rule, where the input is a set of fuzzy sets but the output is an Information Set. Depending on the type-1 or type-2 membership function used in the input fuzzy sets of the antecedent part of the rule, we define the consequent part as the Information set of type-1 or type-2 unlike the output fuzzy set in the Mamdani rule. We name this rule as type-1 Information rule or the type-2 Information rule depending on the type of the membership function. It may be noted that type-2 Information rule that help represent higher order possibilistic uncertainty are the generalization of type-1 Information rule.

A pair of type-1 Information rules (T1IRs) is formed from MFCC corresponding to spatial (Cepstral) and temporal components. The T1IRa for the temporal component and T1IRB for the Cepstral component are of the following form:

T1IRa:

$$\begin{aligned} &IF\ \boldsymbol{x}_1 \text{ is } \mathbb{A}_1 \text{ and } \boldsymbol{x}_2 \text{ is } \mathbb{A}_2 \text{ and } \ldots\ \boldsymbol{x}_\tau \text{ is } \mathbb{A}_\tau \\ &\quad THEN\ \mathcal{H}_T = \left\{\boldsymbol{x}_i^\alpha \mathcal{G}_{\mathcal{A}_i}\right\},\ \text{for } i = 1, \ldots, \tau \end{aligned} \tag{10}$$

T1IRb:

$$\begin{aligned} &IF\ \boldsymbol{y}_1 \text{ is } \mathbb{B}_1 \text{ and } \boldsymbol{y}_2 \text{ is } \mathbb{B}_2 \text{ and } \ldots\ \boldsymbol{y}_d \text{ is } \mathbb{B}_d \\ &\quad THEN\ \mathcal{H}_D = \left\{\boldsymbol{y}_j^\alpha \mathcal{G}_{\mathcal{B}_j}\right\},\ \text{for } j = 1, \ldots, d \end{aligned} \tag{11}$$

Where $\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots . \boldsymbol{x}_\tau$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \ldots . \boldsymbol{y}_d$ are the input vectors (Information Source values).We denote the information source values by $\boldsymbol{x}_\tau = \{x_{\tau 1}, x_{\tau 2}, \ldots, x_{\tau d}\}$ and $\boldsymbol{y}_d = \{x_{1d}, x_{2d}, \ldots, x_{\tau d}\}$.

The corresponding fuzzy sets are denoted by:

$$\mathbb{A}_\tau = \left\{\boldsymbol{x}_\tau, \mathcal{G}_{\mathcal{A}_\tau}\right\} = \left\{\left(x_{\tau 1}, \mathcal{G}_{\mathcal{A}_{\tau 1}}\right), \left(x_{\tau 1}, \mathcal{G}_{\mathcal{A}_{\tau 2}}\right), \ldots, \left(x_{\tau 1}, \mathcal{G}_{\mathcal{A}_{\tau d}}\right)\right\}$$

$$\mathbb{B}_d = \left\{\boldsymbol{y}_d, \mathcal{G}_{\mathcal{B}_d}\right\} = \left\{\left(x_{1d},\ \mathcal{G}_{\mathcal{B}_{1d}}\right), \left(x_{1d},\ \mathcal{G}_{\mathcal{B}_{2d}}\right), \ldots, \left(x_{\tau d},\ \mathcal{G}_{\mathcal{B}_{\tau d}}\right)\right\}$$

$\mathcal{H}_T$ and $\mathcal{H}_D$ are the output Information Sets. As these are derived from the proposed entropy function they are defined as follows:

$$\boldsymbol{x}_i^\alpha \mathcal{G}_{\mathcal{A}_i} = \left\{\left(x_{\tau 1}^\alpha \mathcal{G}_{\mathcal{A}_{\tau 1}}\right), \left(x_{\tau 2}^\alpha \mathcal{G}_{\mathcal{A}_{\tau 2}}\right), \ldots, \left(x_{\tau d}^\alpha \mathcal{G}_{\mathcal{A}_{\tau d}}\right)\right\}$$

$$\boldsymbol{y}_j^\alpha \mathcal{G}_{\mathcal{B}_j} = \left\{\left(x_{1d}^\alpha \mathcal{G}_{\mathcal{B}_{1d}}\right), \left(x_{2d}^\alpha \mathcal{G}_{\mathcal{B}_{2d}}\right), \ldots, \left(x_{\tau d}^\alpha \mathcal{G}_{\mathcal{B}_{\tau d}}\right)\right\}$$

where $\mathcal{G}_{\mathcal{A}_i}$ and $\mathcal{G}_{\mathcal{B}_j}$ are type-1 membership functions for rule-T1IRa and T1IRbrespectively.

The combined output information set is taken as,

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_T + \boldsymbol{\mathcal{H}}_D \tag{12}$$

The combined effective information of the system is computed from:

$$\boldsymbol{H} = \frac{1}{\tau}\sum_i \mathcal{H}, \text{ for } i = 1, \ldots, \tau \tag{13}$$

We now describe the type-2 Information rule (T2IR) in which the antecedent part contains type-2 input fuzzy sets but the consequent part is similar to that of T1IR. The type-2 information rule consists of two parts: one corresponding to the upper MF and another corresponding to the lower MF. Thus corresponding to each T1IRa we will have two type-2 information rules, T2IRUa and T2IRLa and similarly we will have two rules, T2IRUb and T2IRLb corresponding to T1IRb, defined as:

T2IRUa & T2IRLa:

$$\begin{aligned} &IF\ \boldsymbol{x}_1 \text{ is } \overline{\mathbb{A}}_1 \text{ and } \boldsymbol{x}_2 \text{ is } \overline{\mathbb{A}}_2 \text{ and } \ldots\ \boldsymbol{x}_\tau \text{ is } \overline{\mathbb{A}}_\tau\ \ THEN\ \overline{\boldsymbol{\mathcal{H}}}_T = \{\boldsymbol{x}_i^{\alpha}\overline{\boldsymbol{\mathcal{G}}}_{\mathcal{A}_i}\} \text{ for } i = 1, \ldots, \tau \\ &IF\ \boldsymbol{x}_1 \text{ is } \underline{\mathbb{A}}_1 \text{ and } \boldsymbol{x}_2 \text{ is } \underline{\mathbb{A}}_2 \text{ and } \ldots\ \boldsymbol{x}_\tau \text{ is } \underline{\mathbb{A}}_\tau\ \ THEN\ \underline{\boldsymbol{\mathcal{H}}}_T = \{\boldsymbol{x}_i^{\alpha}\underline{\boldsymbol{\mathcal{G}}}_{\mathcal{A}_i}\} \end{aligned} \tag{14}$$

T2IRUb &T2IRLb:

For T1IRb:

$$\begin{aligned} &IF\ \boldsymbol{y}_1 \text{ is } \overline{\mathbb{B}}_1 \text{ and } \boldsymbol{y}_2 \text{ is } \overline{\mathbb{B}}_2 \text{ and } \ldots\ \boldsymbol{y}_d \text{ is } \overline{\mathbb{B}}_d\ \ THEN\ \overline{\boldsymbol{\mathcal{H}}}_D = \{\boldsymbol{y}_j^{\alpha}\overline{\boldsymbol{\mathcal{G}}}_{\mathcal{B}_j}\} \quad \text{for } j = 1, \ldots, d \\ &IF\ \boldsymbol{y}_1 \text{ is } \underline{\mathbb{B}}_2 \text{ and } \boldsymbol{y}_2 \text{ is } \underline{\mathbb{B}}_2 \text{ and } \ldots\ \boldsymbol{y}_d \text{ is } \underline{\mathbb{B}}_d\ THEN\ \underline{\boldsymbol{\mathcal{H}}}_D = \left\{\boldsymbol{y}_j^{\alpha}\underline{\boldsymbol{\mathcal{G}}}_{\mathcal{B}_j}\right\} \end{aligned} \tag{15}$$

The output of T1IRa is obtained by

$$\boldsymbol{\mathcal{H}}_T = \frac{\overline{\boldsymbol{\mathcal{H}}}_T}{\sum_{j=1}^{d}\overline{\boldsymbol{\mathcal{H}}}_{\mathcal{A}_i} + \sum_{j=1}^{d}\underline{\boldsymbol{\mathcal{G}}}_{\mathcal{A}_i}} + \frac{\underline{\boldsymbol{\mathcal{H}}}_T}{\sum_{j=1}^{d}\overline{\boldsymbol{\mathcal{G}}}_{\mathcal{A}_i} + \sum_{j=1}^{d}\underline{\boldsymbol{\mathcal{H}}}_{\mathcal{A}_i}} \tag{16}$$

The output of T1IRb is obtained by

$$\mathcal{H}_D = \frac{\overline{\boldsymbol{\mathcal{H}}}_D}{\sum_{i=1}^{\tau}\overline{\boldsymbol{\mathcal{G}}}_{\mathcal{B}_j} + \sum_{i=1}^{\tau}\underline{\boldsymbol{\mathcal{G}}}_{\mathcal{B}_j}} + \frac{\underline{\boldsymbol{\mathcal{H}}}_D}{\sum_{i=1}^{\tau}\overline{\boldsymbol{\mathcal{G}}}_{\mathcal{B}_j} + \sum_{i=1}^{\tau}\underline{\boldsymbol{\mathcal{G}}}_{\mathcal{B}_j}} \tag{17}$$

where $\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots \boldsymbol{x}_\tau$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \ldots \boldsymbol{y}_d$ are the input vectors (Information Source values) represented such as $\underline{\boldsymbol{x}_\tau} = \underline{\{x_{\tau 1}, x_{\tau 2}, \ldots, x_{\tau d}\}}$ and $\boldsymbol{y}_d = \{x_{1d}, x_{2d}, \ldots, x_{\tau d}\}$. $\overline{\mathbb{A}}_1, \overline{\mathbb{A}}_2, \ldots, \overline{\mathbb{A}}_\tau$ and $\underline{\mathbb{A}}_1, \underline{\mathbb{A}}_2, \ldots, \underline{\mathbb{A}}_\tau$ are the upper and lower fuzzy sets of T2IRUa and T2IRLarepresented as

$$\overline{\mathbb{A}}_\tau = \left\{\boldsymbol{x}_\tau, \overline{\boldsymbol{\mathcal{G}}}_{\mathcal{A}_\tau}\right\} = \left\{\left(x_{\tau 1}, \overline{\mathcal{G}}_{\mathcal{A}_{\tau 1}}\right), \left(x_{\tau 1}, \overline{\mathcal{G}}_{\mathcal{A}_{\tau 2}}\right), \ldots, (x_{\tau 1}, \overline{\mathcal{G}}_{\mathcal{A}_{\tau d}})\right\}$$

$$\underline{\mathbb{A}}_\tau = \left\{\boldsymbol{x}_\tau, \underline{\boldsymbol{\mathcal{G}}}_{\mathcal{A}_\tau}\right\} = \left\{\left(x_{\tau 1}, \underline{\mathcal{G}}_{\mathcal{A}_{\tau 1}}\right), \left(x_{\tau 1}, \underline{\mathcal{G}}_{\mathcal{A}_{\tau 2}}\right), \ldots, (x_{\tau 1}, \underline{\mathcal{G}}_{\mathcal{A}_{\tau d}})\right\}$$

$\overline{\boldsymbol{\mathcal{H}}}_T$ and $\underline{\boldsymbol{\mathcal{H}}}_T$ are the upper and lower Information Sets of T2IRUa and T2IRLa and $\overline{\boldsymbol{\mathcal{H}}}_D$ and $\underline{\boldsymbol{\mathcal{H}}}_D$ are the corresponding upper and lower Information Sets of T2IRUb and T2IRLb respectively. The combined output of a system is calculated using Eqs. (12, 13).

We have a number of type-1 fuzzy membership functions in the literature, *i.e.* Triangular, Gaussian, Trapezoidal, Sigmoidal, pi-shaped, etc. They can be easily converted into type-2 membership functions by changing their parameters. Type-1 Gaussian type membership functions can be easily converted into type-2 by changing either mean or standard deviation.

The mathematical expressions for the type-1 Gaussian membership functions in Cepstro-temporal cases are:

For T1IRa:

$$\boldsymbol{\mathcal{G}}_{\mathcal{A}_i} = \exp\left(-\frac{1}{2}\frac{(\boldsymbol{x}_i - \mu_i)^2}{\sigma_i^2}\right) \tag{18}$$

For T1IRb:

$$\mathcal{G}_{\mathcal{B}_j} = \exp\left(-\frac{1}{2}\frac{(\boldsymbol{y}_j - \mu_j)^2}{\sigma_j^2}\right) \tag{19}$$

The type-1 antecedent parameters are modified by the changing the mean and the standard deviation. The upper mean and standard deviation of type-2 Gaussian membership are defined as follows:

For T2IRUa:

$$\overline{\mu}_i = \frac{1}{d}\sum_{j=1}^{d} x_{ij}, \qquad i = 1, \ldots, \tau \tag{20}$$

$$\overline{\sigma}_i = \sqrt{\frac{1}{d}\sum_{j=1}^{d}(x_{ij} - \overline{\mu}_i)^2}, \qquad i = 1, \ldots, \tau \tag{21}$$

For T2IRUb:

$$\overline{\mu}_j = \frac{1}{\tau}\sum_{i=1}^{\tau} x_{ij}, \qquad j = 1, \ldots, \mathrm{d} \tag{22}$$

$$\overline{\sigma}_j = \sqrt{\frac{1}{\tau}\sum_{i=1}^{\tau}(x_{ij}-\overline{\mu}_j)^2}, \qquad j = 1,\ldots,d \tag{23}$$

Type-2 upper Gaussian membership function for Cepstro-temporal cases are defined as follows:

For T2IRUa:

$$\overline{\mathcal{G}}_{\mathcal{A}_i} = \exp\left(-\frac{1}{2}\frac{\left(\boldsymbol{x}_i-\overline{\mu}_i\right)^2}{\overline{\sigma}_i^2}\right) \tag{24}$$

For T2IRUb:

$$\overline{\mathcal{G}}_{\mathcal{B}_j} = \exp\left(-\frac{1}{2}\frac{\left(\boldsymbol{y}_j-\overline{\mu}_j\right)^2}{\overline{\sigma}_j^2}\right) \tag{25}$$

where, $\overline{\mu}_i$ is the mean and $\overline{\sigma}_i$ is width of the upper membership grade for T2IRUa and $\overline{\mu}_j$ is the mean and $\overline{\sigma}_j$ is width of the upper membership grade for TSIRUb.

We will now consider type-2 interval sets where the lower mean and standard deviation of type-2 lower Gaussian membership function are obtained by scaling the upper mean and standard deviation of the type-2 upper Gaussian membership function as,

$$\underline{\mu} = \gamma_1\overline{\mu} \tag{26}$$

$$\underline{\sigma} = \gamma_2\overline{\sigma} \tag{27}$$

where, γ_1 and γ_2 are the scaling factors. However we have used upper and lower values only in the standard deviation. This means $\gamma_1 = 1$. Inview of this, the twolower membership functions are defined as,

For T2IRLa:

$$\underline{\mathcal{G}}_{\mathcal{A}_i} = \exp\left(-\frac{1}{2}\frac{\left(\boldsymbol{x}_i-\underline{\mu}_i\right)^2}{\underline{\sigma}_i^2}\right) \tag{28}$$

For T2IRLb:

$$\underline{\mathcal{G}}_{\mathcal{B}_j} = \exp\left(-\frac{1}{2}\frac{\left(\boldsymbol{y}_j-\underline{\mu}_j\right)^2}{\underline{\sigma}_j^2}\right) \tag{29}$$

where, $\underline{\mu}_i$ is the mean and $\underline{\sigma}_i$ is width of the lower membership grade for T2IRLa and $\underline{\mu}_j$ is the mean and $\underline{\sigma}_j$ is width of the lower membership grade for T2IRLb.

3.2 Hanman transform based features

The Information sets can also be used to assess higher form of uncertainty in the information source values based on the initial uncertainty representation. This is the concept behind the Hanman transform which follows from the possibilistic version of the adaptive Hanman-Anirban entropy function (Hanmandlu and Das 2011) having variable parameters. Recall Eq. (9)

$$H_\Phi = \sum_i \mathrm{x}_\Phi(\varphi_i)^{\alpha_\Phi} G_\Phi(\varphi_i)$$

where $\boldsymbol{G}_\Phi(\varphi_i) = e^{-(a_\Phi \mathrm{x}_\Phi(\varphi_i)+b_\Phi)^{\beta_\Phi}}$. Assuming its parameters to be variables and substituting $a_\Phi = \mathcal{G}_\Phi(\varphi_i)$ from Eq. (6); $b_\Phi = 0$; $\beta_\Phi = 1$ we obtain Hanman Transform value set, $\mathfrak{H}$ as,

$$\mathfrak{H} = \left\{\mathrm{x}_\Phi(\varphi_i)^{\alpha_\Phi} e^{-(\mathrm{x}_\Phi(\varphi_i)\mathcal{G}_\Phi(\varphi_i))}\right\} \tag{30}$$

This transform has realistic applications; for example, we gather information about an unknown person of some interest to us. This is the first level of information (set) and then evaluate him again to get the second level of information camped with the first one.

To derive Hanman transform, recall the two information rules, T1IRa and T1IRb defined as,

T1IRa:

$$\begin{aligned} &IF\ \boldsymbol{x}_1 \text{ is } \mathbb{A}_1 \text{ and } \boldsymbol{x}_2 \text{ is } \mathbb{A}_2 \text{ and } \ldots\ \boldsymbol{x}_\tau \text{ is } \mathbb{A}_\tau\ \ THEN \\ &\mathfrak{H}_T = \left\{\boldsymbol{x}_i^{\alpha}\mathfrak{A}_i\right\}, \quad \text{for } i = 1,\ldots,\tau \end{aligned} \tag{31}$$

Rule 1b:

$$\begin{aligned} &IF\ \boldsymbol{y}_1 \text{ is } \mathbb{B}_1 \text{ and } \boldsymbol{y}_2 \text{ is } \mathbb{B}_2 \text{ and } \ldots\ \boldsymbol{y}_d \text{ is } \mathbb{B}_d\ \ THEN \\ &\mathfrak{H}_D = \left\{\boldsymbol{y}_j^{\alpha}\mathfrak{B}_j\right\}, \quad \text{for } j = 1,\ldots,d \end{aligned} \tag{32}$$

where $\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots \boldsymbol{x}_\tau$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \ldots \boldsymbol{y}_d$ are the input vectors. $\mathfrak{H}_T$ and $\mathfrak{H}_D$ are the higher order Information Sets using Hanman transform. $\mathfrak{A}_i$ and $\mathfrak{B}_j$ are the gain functions derived from the Information sets of T1IRa and T1IRb for the ith and jth input vectors respectively. These gain functions are functions of the Information values, defined as

$$\mathfrak{A}_i = e^{-\left(\boldsymbol{x}_i\mathcal{G}_{\mathcal{A}_i}\right)} \tag{33}$$

$$\mathfrak{B}_j = e^{-\left(\boldsymbol{y}_i\mathcal{G}_{\mathcal{B}_j}\right)} \tag{34}$$

where $\boldsymbol{x}_i\mathcal{G}_{\mathcal{A}_i}$ and $\boldsymbol{y}_i\mathcal{G}_{\mathcal{B}_j}$ are the type-1 output Information values of T1IRa and T1IRb defined in Eqs. (18, 19) respectively. The combined output of a system is calculated using Eq. (13).

4 Experiments and results

4.1 Database description

The proposed approach is tested on the standard databases such as NIST-2003, VoxForg-2015 and VCTK speech corpus.

Switchboard NIST (2003) evaluation database consists of 356 speakers voice recorded on telephone for a duration of 2 min per speaker with a sampling rate of 8 kHz at 16 bit. We have divided this single speech sample into 5 samples of a user with duration of 50 s for training and testing.

VoxForge (2015) is a collection of transcribed speech to use in Open Source Speech Recognition Engines ("SRE"s). It consists of a large number of speakers from different regions of the world from which we have randomly chosen 100 speakers. Each speaker reads out 10 sentences in English that are recorded with a sampling rate of 8 kHz. The channels used in this recording are different like microphone, mobile, laptops etc.

This CSTR VCTK Corpus (2009) includes speech data collected from 109 native speakers of English with various accents. Each speaker reads out about 400 sentences in which we have randomly selected 5 samples. All speech data was recorded using an identical recording setup: an omni-directional head-mounted microphone (DPA 4035), 48 kHz sampling frequency at 16 bits. For our experiment we have down sampled it to 8 kHz at 16 bits.

4.2 Results and discussions

The other classifiers used in this study are: Gaussian Mixture Model (GMM), Support Vector Machine (SVM) and k Nearest Neighborhood (kNN). They are described briefly now.

4.2.1 Gaussian mixture model (GMM)

It is a probability density function represented as a weighted sum of Gaussian component densities and it is commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system. GMM parameters such as mean, standard deviation and weight of each Gaussian are estimated from the training data using the iterative Expectation–Maximization (EM) algorithm or Maximum A Posteriori (MAP) estimation from a well-trained prior model. In this work we have used 16 Gaussian mixtures to model the MFCC feature vectors in the training set.

4.2.2 Support vector machine (SVM)

This is a discriminative classifier wherein a model of decision hyperplane is constructed using the training feature vectors called the support vectors. These vectors help match the test feature vector with the training feature vectors. In this work we have used radial basis functional kernel with degree 3.

4.2.3 k nearest neighborhood (kNN)

It uses the Euclidean distance between the test feature vector and the training feature vector and its neighbors to identify the unknown user using majority rule. In this work we have considered k = 1, 3, 5 nearest neighbors.

For the extraction of the standard MFCC, we have selected a frame of 20 ms with a frame shift of 10 ms, and 26 Mel filter banks that provides 13 MFCC features per frame. Similarly by using ΔMFCC we have taken 26 features per frame and with ΔΔMFCC 39 features per frame. GMM is used to model the training data of the each speaker and log likelihood is considered for classification.

A frame length of 32 ms duration with a frame shift of 16 ms, and 40 Mel filter banks lead to40 MFCC per frame. On each speech sample we compute 40 T2IS features and 40 HT features. For T2IS features, we have used the classifiers : IHC (Medikonda et al. 2016), SVM (Chang et al. 2011), and k-Nearest Neighborhood (k = 1,3,5).

In Fig. 2, a graph is represented between average recognition (%) and scaling factor (γ) is shown. To generate type-2 Gaussian membership function we generate upper and lower standard deviations (σ). When a test sample is considered with additive white Gaussian noise at a Signal-to-Noise Ratio (SNR) from 0dB to 30dB in steps of 5dB. It is found by experimental that at $\gamma = 0.2$ yields best results on three databases. Scaling factor 'γ' can also be learned by using different learning technique, but here we considered by experimental.

Table 1 presents the average k-fold identification accuracy (%) using T1IS on NIST, VoxForge and VCTK databases. It is observed that SVM and IHC gives the compatible results but with kNN at k = 3 and 5 yields better results with an average improvement of 10, 8 and 13% on NIST, VoxForge and VCTK respectively.

Tables 2, 3 and 4 presents the comparison of average k-fold accuracy (%) of MFCC,ΔMFCC, ΔΔMFCC, and GFCC features using GMM andT2IS features using IHC, SVM and kNN (k = 1,3,5) on NIST, VoxForge and VCTK databases respectively.

On NIST-2003 database, the proposed methods outperforms other features with an average improvement of about 26% when compared with MFCC as can be seen from Tables 2 and 5. But when compared with GFCC there is an

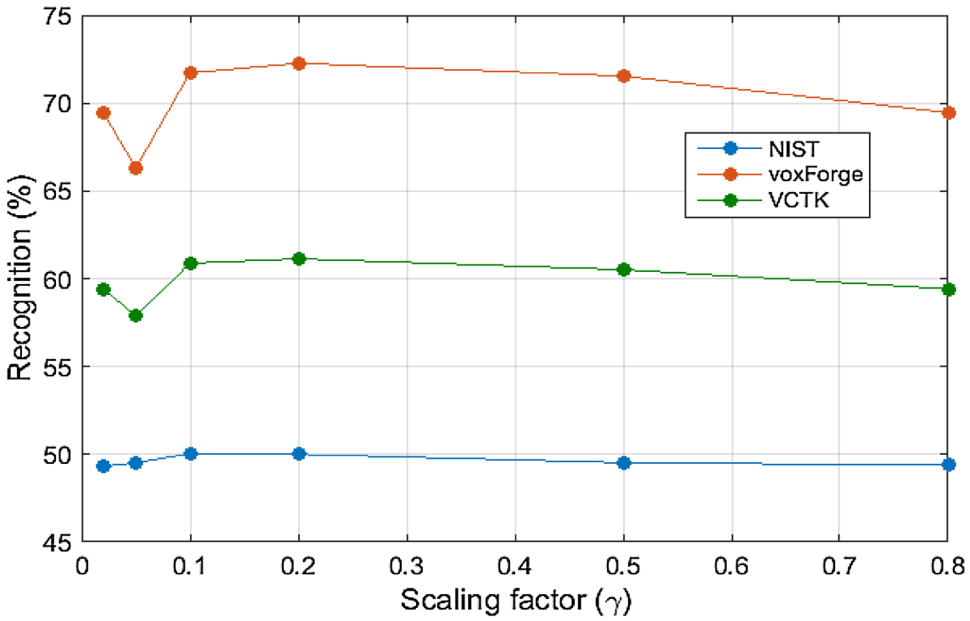

Fig. 2 Average recognition (%) in a noisy environment (white noise with SNR from 0 to 30 dB in steps of 5 dB) of three databases with respect to scaling factor (ϒ)

Table 1 Average k-fold identification accuracy (%) using T1IS on three databases

	SNR (dB)	NIST					VoxForge					VCTK				
		IHC	SVM	kNN			IHC	SVM	kNN			IHC	SVM	kNN		
				k = 1	k = 3	k = 5			k = 1	k = 3	k = 5			k = 1	k = 3	k = 5
White	0	16.5	16.9	15.4	25.2	28.5	47.9	61.0	49.7	71.0	75.2	21.3	28.8	22.6	36.8	40.9
	5	31.4	31.6	28.5	43.3	47.0	69.7	71.7	66.6	81.4	85.5	38.9	47.7	37.9	55.3	60.9
	10	44.3	44.1	40.3	56.4	60.6	79.0	78.3	74.5	87.2	90.3	51.2	55.9	50.3	66.2	71.0
	15	52.9	51.0	47.9	64.1	68.0	82.1	81.7	80.7	89.0	91.4	60.9	62.6	58.5	73.1	76.3
	20	57.8	55.2	52.8	68.7	72.5	83.8	86.2	83.1	92.1	93.8	64.7	68.2	63.0	76.8	80.0
	25	61.3	59.9	54.9	71.4	75.3	85.2	87.6	84.5	92.4	94.5	68.4	69.7	66.5	78.7	81.9
	30	62.9	62.4	57.9	73.3	76.7	85.9	87.6	84.8	93.1	94.8	71.4	71.6	68.2	80.7	83.9
Avg		46.7	45.9	42.5	57.5	61.2	76.2	79.2	74.8	86.6	89.4	53.8	57.8	52.4	66.8	70.7

Table 2 Comparison of average k-fold identification accuracy (%) onNIST database

Noise	SNR (dB)	MFCC Davis and Mermelstein (1980)	ΔMFCC Kumar et al. (2011)	ΔΔMFCC Kumar et al. (2011)	GFCC	T2IS				
						IHC	SVM	kNN		
								k = 1	k = 3	k = 5
White	0	2.36	2.36	2.70	7.75	19.83	21.01	18.76	32.81	39.44
	5	4.44	4.66	4.89	23.71	33.76	35.34	31.91	48.37	56.57
	10	6.52	7.64	8.03	48.60	48.26	48.93	43.15	59.94	66.52
	15	9.72	10.73	16.12	63.15	55.11	56.85	50.00	66.18	72.58
	20	12.47	15.67	27.53	71.18	58.93	60.56	54.89	69.21	75.73
	25	16.57	22.13	42.47	75.79	61.07	62.92	56.40	71.74	77.42
	30	19.94	28.71	53.54	75.79	62.64	64.38	57.53	72.87	78.37
Avg		10.29	13.13	22.18	52.28	48.51	50.00	44.66	60.16	66.66

Table 3 Comparison of average k-fold identification accuracy (%) on VoxForge database

Noise	SNR (dB)	MFCC Davis and Mermelstein (1980)	ΔMFCC Kumar et al. (2011)	ΔΔMFCC Kumar et al. (2011)	GFCC	T2IS				
						IHC	SVM	kNN		
								k = 1	k = 3	k = 5
White	0	4.48	4.83	4.83	24.48	50.93	63.34	52.62	61.38	71.03
	5	7.59	7.59	10.00	53.79	70.97	73.34	67.52	76.9	83.45
	10	21.03	21.38	22.76	81.03	81.38	80.69	84.97	86.55	90
	15	33.45	32.41	39.66	88.62	83.97	82.34	80.21	90.69	94.48
	20	41.38	46.90	52.41	89.66	84.14	87.76	81.00	91.72	95.86
	25	48.62	52.76	63.10	91.03	85.48	88.14	82.07	92.41	96.21
	30	49.31	54.83	71.03	90.34	85.79	88.45	83.07	92.41	95.86
Avg		29.41	31.53	37.68	74.14	77.52	80.58	75.92	84.58	89.56

Table 4 Comparison of average k-fold identification accuracy (%) on VCTK database

Noise	SNR (dB)	MFCC Davis and Mermelstein (1980)	ΔMFCC Kumar et al. (2011)	ΔΔMFCC Kumar et al. (2011)	GFCC	T2IS				
						IHC	SVM	kNN		
								k = 1	k = 3	k = 5
white	0	5.59	5.38	5.38	12.69	25.22	30.01	22.37	36.34	44.09
	5	8.60	8.60	10.54	31.83	40.19	48.06	35.48	55.05	64.09
	10	18.49	17.20	18.06	56.34	50.32	57.26	47.53	66.45	72.9
	15	25.38	28.39	30.97	66.67	62.91	65.49	60.55	72.26	78.71
	20	35.48	37.42	44.52	69.25	65.43	70.58	62.48	75.48	81.72
	25	42.15	45.59	55.91	72.04	67.72	71.66	63.77	76.13	82.58
	30	48.39	50.54	64.52	74.41	68.52	72.24	63.85	76.34	83.44
Avg		26.30	27.59	32.84	54.75	54.33	59.33	50.86	65.44	72.50

Table 5 Average k-fold identification accuracy (%) using Hanman transform on three databases

Noise	SNR (dB)	NIST				VoxForge				VCTK			
		SVM	kNN			SVM	kNN			SVM	kNN		
			k = 1	k = 3	k = 5		k = 1	k = 3	k = 5		k = 1	k = 3	k = 5
White	0	22.60	18.31	30.34	38.65	67.14	53.38	60.00	69.31	21.51	18.92	34.41	45.16
	5	34.20	30.11	47.53	55.22	72.69	60.24	76.21	83.10	37.20	31.40	53.98	63.01
	10	47.00	41.69	59.04	65.79	75.79	69.66	85.52	93.10	49.03	46.24	64.52	73.55
	15	57.60	49.49	66.46	72.13	80.31	72.41	88.97	92.41	55.05	50.54	68.82	78.49
	20	63.00	54.04	69.89	75.45	82.07	77.24	89.31	93.10	59.14	52.69	70.11	78.71
	25	66.60	56.40	71.69	77.42	84.10	78.93	90.34	94.83	62.37	54.62	72.69	80.65
	30	70.20	57.87	72.70	78.65	85.76	82.62	91.38	95.17	61.72	56.13	73.55	80.86
Avg		51.60	43.99	59.66	66.19	78.27	70.64	83.10	88.72	49.43	44.36	62.58	71.49

average improvement of 10% when kNN (k = 3, 5) is used and there is an average improvement of 5% at 0–5 dB SNR when IHC and SVM classifiers are used. At different SNR values, the extent of improvement in performance varies. And it is similar with VoxForge 2014 and VCTK databases as shown in Tables 3, 4 and 5.

The proposed method helps reduce the size of the feature vector. We can see from Table 6 that there is a drastic reduction in feature size in the multi-dimensional feature vector of size about 18,000 (~ 13 × 1385) for MFCC and 31,000 (~ 22 × 1385) for GFCC to an information set based feature vector (T2IS, HT) of size ~ 30 after feature selection using MDA.

Another significant achievement of this proposed method is reduction in computation time. From Table 7, it is observed that computational time using proposed

Table 6 Average number of features per sample approximately

Database	MFCC	ΔMFCC	ΔΔMFCC	GFCC	T2IS	HT
NIST-2003	18,000	36,000	54,000	31,000	30	30
VoxForge	6000	12,000	18,000	11,000	28	28
VCTK	6000	12,000	18,000	11,000	28	28

Table 7 Approximated average time (sec) taken to complete identification process of all users per sample as test, with white noise and at a snr

Database	MFCC	ΔMFCC	ΔΔMFCC	GFCC	T2IS	HT
NIST-2003	2104	2160	2240	3140	659	500
VoxForge	100	100	100	140	28	21
VCTK	172	172	176	200	42	39

method is less than that of the standard state-of-the-art methods.

5 Conclusions

This paper formulates Type-2 Information Set features (T2IS) and Hanman Transform features (HT) based on Information Set theory in the development of a robust text-independent speaker identification system in the presence of whiten Gaussian noise at six different SNRs. This was an effort taken to extract speaker specific information in noisy environment without any noise reduction.

In the first phase the audio signal is partitioned into frames and from each frame Mel Frequency Cepstral Coefficients (MFCC). Considering MFCC matrix, such that each row of a matrix corresponds to a dimension and each column corresponds to a frame. This matrix representation facilitates the derivation of T2IS features and HT features from frames yielding the type-2 and HTcepstral information and dimensions yielding the type-2 and HT temporal information. Thus at each position in the matrix we have two types of information components adding which we get T2IS features. After the extraction of feature vectors from all samples of a user, we have set aside some feature vectors as the training sets and the rest as the test feature vectors.

The proposed method is applied on three datasets (NIST 2003, VoxForge 2015, VCTK 2009) with four types of noises. For the sake of comparison of performance the three types of MFCC (MFCC, ΔMFCC, ΔΔMFCC) and GFCC with GMM are used as baseline methods. The proposed T2IS features found to be robust and outperforms when compared with the baseline methodologies. This vindicates the effectiveness of T2IS features over the existing features. Moreover the number of T2IS features is very less thus reducing the computational complexity.

References

Aggarwal, M., & Hanmandlu, M. (2015). Representing uncertainty with information sets. *IEEE Transactions on Fuzzy Systems, 24*, 1–15.

Chang, C.-C., & Lin, C.-J., LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*, 1–27, 2011.

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech and Signal Processing, 28*, 357–366.

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing, 32*, 1109–1121.

Hanmandlu, M. (2011). Information sets and information processing. *Defence Science Journal, 61*, 405–407.

Hanmandlu, M., & Das, A. (2011). Content-based image retrieval by information theoretic measure. *Defence Science Journal, 61*, 415–430.

Hermansky, H., & Morgan, N. (1994). RASTA processing of speech. *IEEE Transactions on Speech and Audio Processing, 2*, 578–589.

Jawarkar, N. P., Holambe, R. S., & Basu, T. K., Use of fuzzy min-max neural network for speaker identification, In *2011 International Conference on Recent Trends in Information Technology* (*ICRTIT*), 2011, pp. 178–182.

Jayanna, H. S., & Prasanna, S. R., & Mahadeva. (2009, Multiple frame size and rate analysis for speaker recognition under limited data condition. *IET Signal Processing, 3*(3), 189–204.

Kumar, K., Kim, C. & Stern, R. M., Delta-spectral cepstral coefficients for robust speech recognition, In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), 2011, pp. 4784–4787.

Lee, K. Y. (2004). Local fuzzy PCA based GMM with dimension reduction on speaker identification. *Pattern Recognition Letters, 25*, 1811–1817.

Lung, S.-Y. (2004). Further reduced form of wavelet feature for text independent speaker recognition. *Pattern Recognition, 37*, 1565–1566.

Lung, S.-Y. (2004). Adaptive fuzzy wavelet algorithm for text-independent speaker recognition. *Pattern Recognition, 37*, 2095–2096.

Mamta, & Hanmandlu, M. (2014). Robust authentication using the unconstrained infrared face images. *Expert Systems with Applications, 41*, 6494–6511.

Mamta, & Hanmandlu, M. (2014). A new entropy function and a classifier for thermal face recognition. *Engineering Applications of Artificial Intelligence, 36*, 269–286.

Medikonda, J., Madasu, H., & Panigrahi, B. K. (2016). Information set based gait authentication system. *Neurocomputing, 207*, 1–14.

Mirhassani, S. M., & Ting, H.-N. (2014). Fuzzy-based discriminative feature representation for children's speech recognition. *Digital Signal Processing, 31*, 102–114.

NIST (2003). *The NIST year 2003 speaker recognition evaluation plan*. Available: http://www.itl.nist.gov/iad/mig/tests/sre/2003/2003-spkrec-evalplan-v2.2.pdf.

Pelecanos, J., & Sridharan, S. (2001). Feature Warping for Robust Speaker Verification, presented at the A Speaker Odyssey—The Speaker Recognition Workshop, Crete.

Pinheiro, H. N. B., Vieira, S. R. F., Ren, T. I., Cavalcanti, G. D. C., & de Mattos Neto, P. S. G. (2016). Type-2 fuzzy GMM for text-independent speaker verification under unseen noise conditions, In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp. 5490–5494.

Reynolds, D. A., & Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing, 3*, 72–83.

Reynolds, D. A. (1995). Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication, 17*, 91–108.

Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted gaussian mixture models. *Digital Signal Processing, 10*, 19–41.

Sohn, J., Kim, N. S., Sung, W. (1999). A statistical model-based voice activity detection". *IEEE Signal Processing Letters, 6*, 1–3.

Togneri, R., & Pullella, D. (2011). An overview of speaker identification: accuracy and robustness issues. *IEEE Transactions on Circuits and Systems Magazine, 11*, 23–61.

VCTK (2009). The Centre for Speech Technology Research VCTK Corpus.

VoxForge (2015). *VoxForge speech corpus*. Available: http://www.repository.voxforge1.org/downloads/SpeechCorpus/Trunk/Audio/Main/.

Wang, Y., Liu, X., Xing, Y., & Li, M. (2008). A Novel Reduction Method for Text-Independent Speaker Identification," in *2008 Fourth International Conference on Natural Computation*, pp. 66–70.

Yuan, Z. X., Yu, C. Z., & Fang, Y. (1993). Text independent speaker identification using fuzzy mathematical algorithm, In *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-93*, Vol. 2., pp. 403–406.

Zhao X., & Wang D. L. (2013). Analyzing noise robustness of MFCC and GFCC features in speaker identification, In *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp. 7204–7208.

Zhao X., Shao Y., Wang D. L. (2012). CASA-based robust speaker identification. *IEEE Transactions on Audio, Speech, and Language Processing, 20*, 1608–1616.