

Speaker Recognition in Mismatch Conditions: A Feature Level Approach

Sharada V Chougule

Finolex Academy of Management & Technology, Ratnagiri, Maharashtra, India
Email: shardavchougule@gmail.com

Mahesh S. Chavan

KIT's College of Engineering, Kolhapur, Maharashtra, India
Email: maheshpiyu@gmail.com

Abstract—Mismatch in speech data is one of the major reasons limiting the use of speaker recognition technology in real world applications. Extracting speaker specific features is a crucial issue in the presence of noise and distortions. Performance of speaker recognition system depends on the characteristics of extracted features. Devices used to acquire the speech as well as the surrounding conditions in which speech is collected, affects the extracted features and hence degrades the decision rates. In view of this, a feature level approach is used to analyze the effect of sensor and environment mismatch on speaker recognition performance. The goal here is to investigate the robustness of segmental features in speech data mismatch and degradation. A set of features derived from filter bank energies namely: Mel Frequency Cepstral Coefficients (MFCCs), Linear Frequency Cepstral Coefficients (LFCCs), Log Filter Bank Energies (LOGFBs) and Spectral Subband Centroids (SSCs) are used for evaluating the robustness in mismatch conditions. A novel feature extraction technique named as Normalized Dynamic Spectral Features (NDSF) is proposed to compensate the sensor and environment mismatch. A significant enhancement in recognition results is obtained with proposed feature extraction method.

Index Terms—Feature extraction, Speaker recognition, Segmental features

I. INTRODUCTION

Speech is the most natural way of communication in human beings. It carries variety of information such as message to be conveyed, language and emotion [1] of the speaker etc. Ease of obtaining the speech and simple devices used to acquire the same are the main reasons of growth in speech related applications. Speech is one of the promising biometric which can be used in variety of applications such as secure access to systems, transactions over telephone, law enforcement and forensic etc. Such applications demands for optimum performance in real world conditions. Speaker recognition is one such technology, which uses a person's speech to recognize (identify or verify) an individual. The

main limitation in using human voice as its identity is the intra-speaker variability (voice of same person changes over time and may not be exactly same all the time), and mismatch in training and testing conditions.

Unlike other human individualities (e.g. face, fingerprint, iris), speech is supposed to be somewhat 'unreliable'. Factors such as age, emotions, language and speaking style can change the characteristics of the speech. Such change can be easily detected by human hearing perception but cannot be easily captured through time domain analysis. Along with these variations, *mismatch* in speech data used during training and the one during testing is one of the main reasons limiting the speaker recognition performance. *Mismatch* (related to the work in this paper) refers to changes in conditions in which speech of the same person is collected. The widespread use of variety of sensing devices (from simple microphone to a variety of mobile phones) and surrounding environment affects the quality of original speech. The effects of transmission channel as well as distortions introduced due to surrounding environment may corrupt the speaker specific traits, which may lead to degradation in recognition results. Various approaches for robust speaker recognition are proposed in literature. These are based on four main methodologies: Robustness at (i) Signal level (ii) Feature level (iii) Model level and (iv) Score level.

Thus, from the front-end point of view, there are two possible solutions to improve the speaker recognition performance in mismatch conditions. First is to introduce techniques that enhance the speech quality by reducing noise and degradations. The second possible solution is to design the speaker specific features, which are robust against mismatch conditions. In the first case, methods like end point detection, silence removal [2], speech enhancement for estimating and suppressing background noise [3],[4] or through voice activity detection [5]. In the second case features derived from human speech production model like source (vocal folds) features [6], system (vocal tract) features [7] or behavioral features [8], [9] are mainly used.

In this work, the focus is on feature extraction techniques to investigate the performance of speaker recognition system in mismatched speech data. This work

is based on the features derived from the filter bank energies namely: MFCCs, LFCCs LOGFBs and SSCs are used for this purpose. MFCC and LFCC features are used as conventional features in state-of-art speaker recognition system; whereas Log Filter Bank Energies and Spectral Subband Centroids (SSCs) are the complementary feature sets derived from MFCC and not used much. The goal here is to investigate the robustness of the features and hence the speaker recognition system. The case of sensor and environment mismatch is studied in this work. IITG Multi-variability speaker recognition database having speech recorded from different sensors and in uncontrolled environment is used for performance evaluation of speaker recognition system.

A novel feature extraction technique named as *Normalized Dynamic Spectral Features* (NDSF) is proposed to compensate the sensor and environment mismatch. The comparison of proposed feature set and filter bank energy features is done for speaker recognition in a variety of mismatch conditions. The rest of the paper is arranged as follows: Section II introduces the conventional and proposed feature extraction technique used in the experimental analysis. The details of database and implementation of system are given in section III. The analysis and comparison of experimental work with obtained results is presented in section IV and conclusions are discussed in section V.

II. FEATURE EXTRACTION

In relation to speaker recognition task, the purpose of feature extraction process is to derive speaker specific information (parameters) from the speech signal, discarding the linguistic and other undesired contents. Feature vector is a compact representation of speech data (relevant to that speaker) in numerical form. Before actual feature extraction, front end processing is carried out, which includes pre-emphasis and segmentation. Short-time Fourier transform is carried out on the short segments of speech signal. A set of band pass filters spaced according to critical bands in human perception is designed to cover the entire sampling frequency range. The design specifications of these band-pass filters/ filter bank (e.g. number of filters, center frequency, and filter bandwidth) vary according to the auditory scale used. The most widely used perceptual scale is Mel-scale [6]. All the filters in the filter bank are weighted according to this scale. Different forms of filter bank energies are used to derive the features. The LFCC features are derived by a method similar to MFCCs except the use of uniformly spaced filter bank. Filter bank energies are the common factor in all these feature extraction techniques.

Before performing feature extraction, all the speech data is resampled at the sampling frequency of 8 kHz. Speech signal is pre-emphasized using a first-order differentiator. In order to examine the steady speech parameters, speech signal is divided into small segments (blocks) using hamming window of length 20 msec and frame overlap of 10 msec. Such analysis is known as segmental analysis and is useful to extract the speaker

specific physiological details related to vocal tract. Short-time Fourier transform [10] is then applied on each windowed speech segment. The analytical treatment for extracting these features is discussed further.

A. Mel-Frequency Cepstral Coefficients(MFCCs)

MFCCs are the feature vectors in cepstral domain [11] computed as a discrete set of decorrelated coefficients.

The procedure to obtain MFCC features is as follows:

- (i) Given an input, find N-point DFT of windowed speech signal.
- (ii) Construct a mel-scale filter-bank with R equal height triangular filters .
- (iii) Perform weighting of FFT magnitude/power spectrum of windowed speech according to mel-scale filter bank.
- (iv) Compute log energy of each of the superimposed filter.
- (v) Decorrelate the logarithmically compressed filter output energies using Discrete Cosine Transform.

MFCCs are represented mathematically as:

$$C_{mel}(m, n) = \frac{1}{R} \sum_{l=0}^{R-1} \log\{E_{mel}(n, l)\} \cos\left(\frac{2\pi}{R} lm\right) \quad (1)$$

where E_{mel} is the energy of l^{th} mel-scale filter for each speech frame at time n . The reason of selecting MFCC features in this work is to investigate the sustainability of this state-of-art feature in mismatch conditions.

B. Linear Frequency Cepstral Coefficients (LFCCs)

The computation of LFCC features is similar to MFCC. The only difference is in the nature of filter bank scale. Here R number of uniformly spaced (equal weight) triangular band pass filters form a Linear Filter bank and weighted on the windowed FFT spectrum. LFCCs are computed from the log magnitude DFT directly. It is observed that characteristics related to vocal tract (particularly vocal tract length) of speaker are reflected more in high frequency range of the speech [12]. In the same study LFCC features are found more robust than MFCC features for channel variabilities.

C. Log Filter Bank Energies(LFBEs)

LFBE [13] features are computed directly from the log filter bank energies from the output of each band pass filter and given as:

$$lfbe[j, n] = \log\left(\sum_{k=0}^{N-1} H_j(k) |S_n(k)|^2\right) \quad (2)$$

Here $H_j(k)$ are triangular number of triangular filters with $j=1,2,\dots,p$ and $|S_n(k)|^2$ represents the power spectrum of n^{th} frame.

D. Spectral Subband Centroids (SSCs)

For computation of SSCs, the entire frequency band (0 to $F_s/2$) is divided into M number of sub-bands, where F_s is the sampling frequency of the speech signal. SSCs are

found by applying filter bank to the power spectrum of the signal and then calculating first moment (centroid) of each subband [14]. SSC of the m^{th} subband is calculated as [15]:

$$C_m = \frac{\int_0^{Fs/2} f \omega_m(f) P^\gamma(f) df}{\int_0^{Fs/2} \omega_m(f) P^\gamma(f) df} \quad (3)$$

where F_s is the sampling frequency, $P(f)$ is the short-time power spectrum, $\omega_m(f)$ is the frequency response of m^{th} band pass filter and γ is the parameter controlling the dynamic range of the power spectrum. SSC features provide different information than MFCC in the sense that, it computes the peaks in the power spectrum in each sub band, which are less affected by noise than the weighted amplitude of power spectrum in case of MFCCs.

E. Normalized Dynamic Spectral Features (NDSFs)

Computation of NDSF and MFCC are similar up to mel-scale integration as discussed in above section. A pre-emphasis filter of the form $H(z) = 1 - 0.97z^{-1}$ is used to avoid spectral tilts due to glottal source and also to boost high-frequency speech components.

The mismatch conditions studied in this work is having the speech data contaminated by variety of channel and environmental noise. Most speech enhancement techniques used for noise suppression are based on the assumption that noise is stationary. Further channel noise is considered to be convolutive whereas environmental noise as additive. The proposed NDSF features [16] emphasize the spectral level speaker information without using any specific assumption about the noise (distortion) in the signal. From experimentation, it is observed that the dynamic range of the spectrum of clean speech is more than that of the noisy speech over a single frame. As compared to cepstral information, spectral information is more useful in the presence of noise. This is because the spectrum retains the peak information though valleys get filled by noise (Fig.2). This information is nothing but formants of speech and is an important parameter distinguishing individual speaker.

The dynamic or temporal information was proved to be more resilient to noise interference than static features for speech recognition task, where speech is corrupted by variety of noises [17]. In order to retain speaker specific characteristics in the presence of noise/mismatch, temporal dynamics is appended to the spectral features instead of conventional cepstral features.

This is done by performing time difference operation, called delta features as given below.

$$\Delta f_k[i] = f_{k+M}[i] - f_{k-M}[i] \quad (4)$$

Thus, the spectral domain dynamic information helps to boost the rapidly changing speech components and suppress slowly varying noise components. We call these features as *dynamic spectral features*. The dynamic

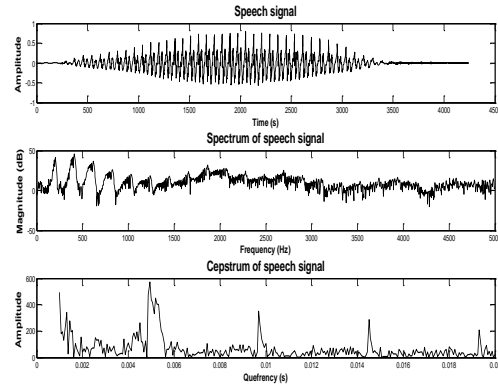


Fig.1 Spectrum and cepstrum of speech signal

spectral features ignore the static spectral features and trim down any additive noise. Further, instead of using log non-linearity, data-driven Gaussianized non-linearity is used over temporal difference speech parameters which make the features less dependent of spoken utterances. In other words, Gaussianized non-linearity converts the sample distribution w_i ($i=1,2,\dots,N$) of N samples per frame in the form $N(0, \sigma^2)$. Here σ^2 is variance under Gaussian assumption and is given as:

$$\sigma^2 = \frac{1}{N} E(\sum_{i=1}^N w_i) \quad (5)$$

A high degree of spectral compaction is achieved by applying DCT on filter bank energies, which further helps to reduce feature size, eliminating higher order features. The features obtained at this stage are cepstral features. The resultant feature vectors are further conditioned to reduce the effect of channel and environmental noise. This is done by warping the cepstral features, through which multi-modal distribution of feature vectors of a single speaker is mapped to a sub-optimum density function. The non-trivial speaker specific features are emphasized through this procedure, providing less emphasis to speech information (spoken utterances). The resulting feature distribution is thus reshaped, which have the advantage of eradicating channel and noise effects. This is essentially useful in the case of real world circumstances where noise or reason of disturbance is unknown. Fig. 3 shows an example of warping of cepstral features of a randomly chosen speaker.

To add more robustness, second order time-difference operation is performed on warped features as:

$$\Delta \Delta f_k[i] = \Delta f_{k+M}[i] - \Delta f_{k-M}[i] \quad (6)$$

The advantage of using proposed NDSF features is that no specific assumption about the noise (or calculation of signal to noise ratio) is done while formulating the features. This fact is useful in practical situations in which reason and nature (characteristics) of noise is unknown.

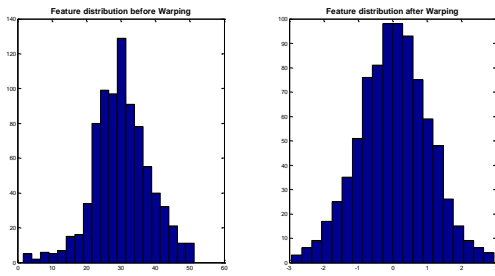


Fig.2. Example of warping of cepstral features

Along with additive noise (such as background/environment noise), the speech data is also susceptible to convolutive noise due to variable sensors and transmission channels (wired or wireless). To compensate such effects, finally normalization of extracted feature vectors is carried out before performing model training.

III. SPEECH DATABASE AND SYSTEM IMPLEMENTATION

IITG-MV is the speech database developed by Indian Institute of Technology Guwahati, India for speaker recognition systems in Indian context. The database is focused on robustness of speaker recognition in Indian scenario where speech is having wide variety and variability.

Table 1. Technical details of the sensors/devices used for collecting the speech data [18], [19]

Device/sensor	Make/model	Sampling Rate	Recording Format
Headset Microphone (H)	Frontech JIL 1903	16 kHz	wav
Table PC (TPC)	HP Elite Book 2730p	16 kHz	wav
Digital Voice Recorder (DVR)	Sony ICD-UX70	44.1 kHz	mp3
Mobile Phone-1 (M-1)	Nokia 5130c Xpress Music	8 kHz	amr
Mobile Phone-2 (M-2)	Sony Ericsson W350	8 kHz	amr

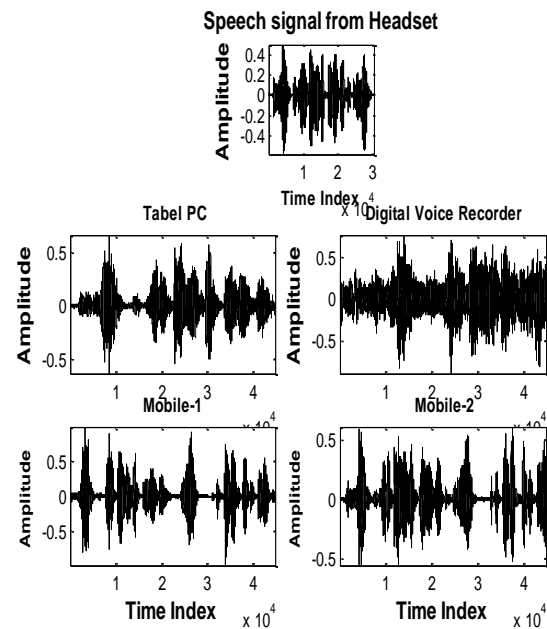
For the case of sensor mismatch, *IITG-MV-Phase-I* dataset is used. This phase of speech data consists of speech recorded from 100 subjects, 81 males and 19 females of the age group 20-40 years. In this, the speech data was collected (recorded) from five different sensors simultaneously. Any type of external (artificial) noise was avoided while recording. The basic purpose of this database is to study the effects of reverberations and ambient noise in the room. The language spoken was English, the speaking style was read speech and environment is closed room. The speech files were saved with 16-bit per sample resolution. The speech recorded from headset is supposed to be cleanest (and of high SNR) amongst all. Table-I gives the details of the speech material in this case.

IITG-MV Phase-II is called as IITG Multi-environment database. The database in this phase consists of speech recorded in uncontrolled environments such as laboratories, hostel rooms, corridors etc. The speech is recorded with headset microphone. The language spoken by all subjects is English. Database from this phase is used for evaluation of environment mismatch condition, in which training speech is clean (from headset and in closed room) and test speech is degraded due to surrounding noise.

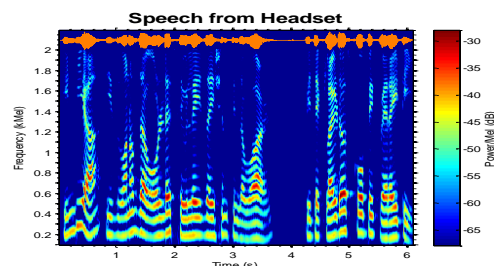
IV. RESULTS AND DISCUSSION

Experiments are carried on a closed-set text independent speaker recognition (identification) system. A limited amount of speech data (20-30 msec for training and 8-15 msec for testing) is used for evaluation. K-means clustering algorithm based on iterative refinement approach is used for formation of speaker model, which is effective specifically for limited data condition [20]. The system performance is measured in terms of *Percentage Identification Accuracy (PIA)* given as, the ratio of number of correctly classified test sequences to the total number of test sequences calculated in percentage.

Case-I Sensor Mismatch



(a) Time domain plot



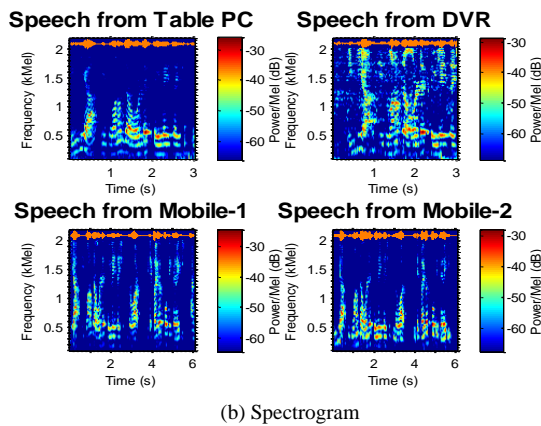
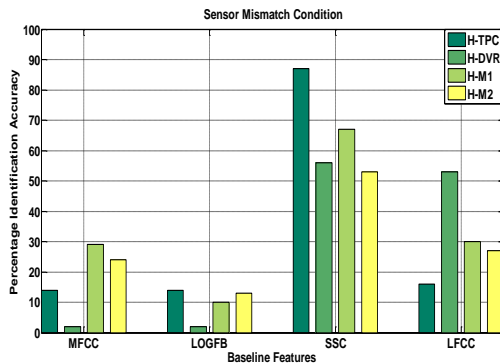
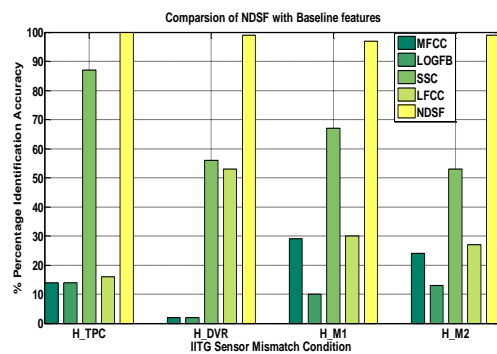


Fig.3. Time domain plot and Spectrogram of samples of speech data from different sensors

Fig. 3 shows time domain plot and spectrogram of samples of speech data from different sensors. It is observed from Fig. 3 (a) that the speech signal from different sensors (except headset microphone) is distorted. From the spectrograms of the same speech samples, it is observed that the speech recorded with headset is having clear and having distinguishing formant structure. For the remaining four cases of sensor, the formant structure is distorted. Also there is loss of some of the high frequency formants. Therefore, it can be said that the higher formants are more susceptible to sensor mismatch than the lower formants. From human auditory perception characteristics, high frequencies also carry the speaker specific information. This is also one of the reason of reduction in correct identification rate.



(a) Baseline features



(b) Comparison with NDSF features

Fig.4. Sensor Mismatch Condition

The results in Fig. 4(a) the percentage identification accuracy of speaker recognition system using baseline features. The results of sensor mismatch condition are shown in Fig.4. As observed from these plots, the all the baseline features are very sensitive to sensor mismatch. However, spectral subband centroids (SSC) features have shown better identification accuracy in sensor mismatch speech data. As SSC computes peaks in the spectrum, there is less effect of noise and degradations on extracted features. For the case of Headset-Digital Voice Recorder (H-DVR) all the feature sets shown poor performance. The reason being, the most degraded speech from digital voice recorder. Proposed NDSF features are designed to reduce the effect of additive as well as channel (convolutive) noise. This results into superior identification accuracy than the baseline features for all the cases of sensor mismatch as observed from plots in Fig. 4. (b). It should be noted that the no additional speech enhancement technique is used at the front end before actual feature extraction. Thus the proposed features of its own are observed to be robust in sensor mismatch case.

Case-II Environment Mismatch

In this case of mismatch, the training speech is clean and noiseless, whereas test speech is collected from outdoor environment having variety of external noise.

Four subcases studied for training-testing are as follows:

- (i) Office-Office
- (ii) Office-Multi-environment
- (iii) Multi-environment-Office
- (iv) Multi-environment-Multi-environment

Fig. 5 shows the speech samples and respective spectrogram in office and outdoor environment. The main difference observed from the waveform (time domain plots) is reduced amplitude of speech signal recorded in outdoor environment. The low signal strength can be easily masked by noise and this may cause drop in the performance of speaker recognition system.

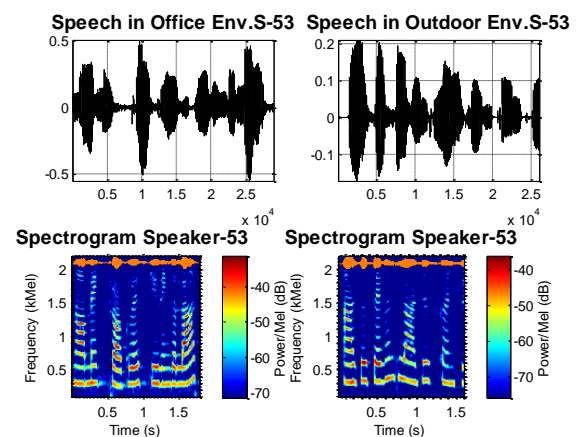
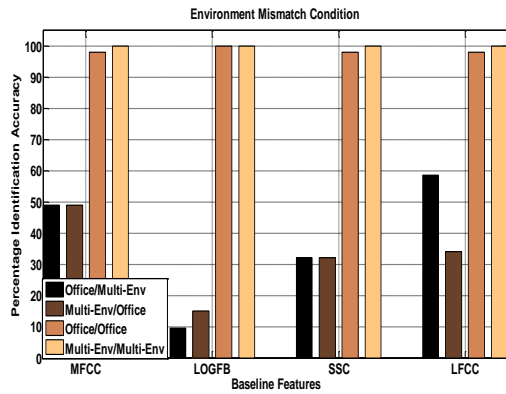
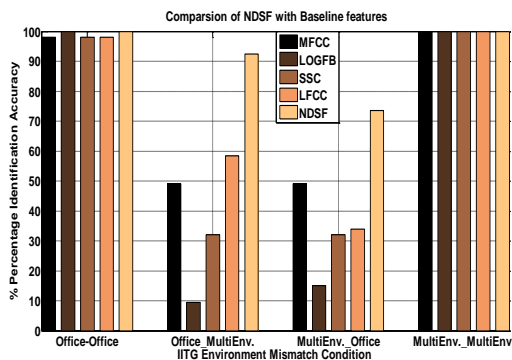


Fig.5. Speech signal and respective spectrogram in different environments



(a) Baseline features



(b) Comparison with NDSF features

Fig.6. Environment Mismatch Condition

The results of environment mismatch with baseline features are shown in Fig.6 (a). MFCC and LFCC are observed to give comparatively better results in environment mismatch condition. In matched environment condition (office-office and multi-environment-multi-environment), speaker recognition system had shown satisfactory performance (almost 100 percent) with all the baseline features. Fig.6 (b) shows the performance of speaker recognition system using proposed NDSF features in environment mismatch. Significant improvements in percentage identification accuracy are also observed for the cases of mismatch in environment (office-multi-environment and multi-environment office).

The results shown above prove that the proposed NDSF features are more robust to channel and environment changes (mismatch) than the conventional baseline features.

V. CONCLUSION

Robustness of speaker recognition system under different environments and conditions is challenging issue for real world applications. Robustness at the feature level is one of the solutions for the same. Features derived from short-time spectrum (segmental features) alone are much susceptible to mismatches in speech data. Spectral centroid features is more robust baseline feature in case of sensor mismatch. Formants carry useful speaker specific information in source-filter model of the

speech signal. The proposed Normalized Dynamic Spectral Features (NDSF) proved effective in sensor and environment mismatch conditions than the conventional low level features. It should be emphasized here that no additional front end speech enhancement technique is used before actual feature extraction. NDSF features itself found to be robust in mismatch condition irrespective of nature of distortion (or noise). This fact can be further used for various other real world mismatch conditions. Further preventing loss of high frequency formants using appropriate filtering can help to enhance the identification accuracy. The NDSF features can further be investigated by adding some known noises of different SNRs.

ACKNOWLEDGMENT

The authors would like to thank IIT Guwahati for providing speech database.

REFERENCES

- [1] Amirreza Shirani and Ahmad Reza Naghsh Nilchi, "Speech Emotion Recognition based on SVM as Both Feature Selector and Classifier", *I.J. Image, Graphics and Signal Processing*, 2016, 4, 39-45.
- [2] Tushar Sahoo and Sabyasachi Patra, "Silence Removal and Endpoint Detection of Speech Signal for Text Independent Speaker Identification", *I.J. Image, Graphics and Signal Processing*, 2014, 6, 27-35.
- [3] Qi Li, Jinsong Zheng, Augustine Tsai, and Qiru Zhou "Robust end point detection and energy normalization for real-time speech and speaker recognition", *IEEE Transactions On Speech And Audio Processing*, Vol. 10, No. 3, March 2002, pp.146-157.
- [4] Sharada V Chougule, Mahesh S Chavan, "Channel Robust MFCCs for Continuous Speech Speaker Recognition", Springer Book Series: *Advances in Signal Processing and Intelligent Recognition Systems Volume 264*, 2014, pp 557-568.
- [5] Ali I., Saha G., "A Robust Iterative Energy Based Voice Activity Detector," *Proceedings, International Conference on Emerging Trends in Engineering and Technology (ICETET)*, IEEE, 2010.
- [6] Steven V Devis and Paul Mermelstein, "Comparison of parametric representations of monosyllabic word recognition in continuously spoken sentences," *IEEE Transaction on Audio, Speech and Language Processing*, vol.4, ISSP-28, no.4, pp.357-366, August 1980.
- [7] Wai Nang Chan, Nengheng Zheng, and Tan Lee, "Discrimination Power of Vocal Source and Vocal Tract Related Features for Speaker Segmentation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 6, pp. 1884-1892, August 2007.
- [8] Tomi Kinnunen and Haizhou Li, "An overview of text independent speaker recognition: From features to supervectors," *Speech Communication*, 2010.
- [9] Joseph P. Campbell, Douglas A. Reynolds, Robert B. Dunn, "Fusing High- and Low-Level Features for Speaker Recognition," *EUROSPEECH*, pp.2665-2668, 2003.
- [10] Lawrence R. Rabiner and Ronald W. Schafer, "Digital processing of speech signals," *Prentice Hall International*, 2011.
- [11] Sadaoki Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Trans. Speech, Audio Processing*, vol. ASSP-29, no. 2, pp. 254-272, April 1981.

- [12] Xinhui Zhou, Daniel Garcia-Romero, Ramani Duraiswami, Carol Espy-Wilson and Shihab Shamma, "Linear versus Mel frequency cepstral coefficients for speaker recognition", ASRU 2011, pp. 559-564.
- [13] Homayoon Beigi, "Speaker Recognition: Advancements and Challenges", New Trends and Developments in Biometrics, Chapter 1, INTECH.
- [14] Jingdong Chen, Yiteng (Arden) Huang, Qi Li and Kuldip K. Paliwal, "Recognition of noisy speech using dynamic spectral subband centroids," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 258-261, February 2004.
- [15] K. K. Paliwal, "Spectral centroid features for speech recognition", Proc. ICASSP, vol. 2, pp.617-620, 1998.
- [16] Sharada V Chougule, Mahesh S Chavan, "Robust spectral features for automatic speaker recognition in mismatch condition," *Second International Symposium on Computer Vision and the Internet (VisionNet'15), Procedia Computer Science*, 58 (2015) 272 – 279, 2015.
- [17] Chen Yang, Frank K. Soong and Tan Lee, "Static and dynamic spectral features: Their noise robustness and optimal weights for ASR", *IEEE Transactions on Audio, Speech, And Language Processing*, vol. 15, no. 3, March 2007, pp.1087-1097.
- [18] Haris B C, G. Pradhan, A. Misra, S. R. M. Prasanna, R. K. Das and R. Sinha, "Multivariability speaker recognition database in Indian scenario," *National Communication Conference, IEEE*, 2011.
- [19] Haris B C, G Pradhan, A Misra, S Shukla, R Sinha and S R M Prasanna, Multi-Variability Speech Database for Robust Speaker Recognition, *National Communication Conference*, IEEE, 2011.
- [20] H.S. Jayanna and S.R.M. Prasanna, "An experimental comparison of modeling techniques for speaker

recognition under limited data condition", Sadhana Academy, proceedings in Engineering Sciences (Springer), vol. 34(3), pp.717-728, Oct.2009.

Authors' Profiles



Sharada V Chougule has obtained B.E. and M.E. degree in Electronics Engineering from Shivaji University, Kolhapur. Currently she is perusing Ph.D. from Shivaji University Kolhapur, Maharashtra, India. She has 20 years of teaching experience. Her areas of interest includes Digital Signal and Image Processing, Speech and Speaker Recognition.



Mahesh S Chavan is a Professor, received the B.E. degree in Electronics Engineering from Shivaji University Kolhapur in year 1991. He received M.E. degree from Shivaji University Kolhapur, Maharashtra, India in year 1998. He has received PH. D. degree in Electronics and Communication Engineering from Kurukhshetra University, India in year 2008. Currently he is a professor in Electronics Engineering Department at KIT's College of Engineering, Kolhapur. He has more than 24 years of teaching experience. His research interest includes Digital Signal Processing, Speech Processing, Advanced Control Systems. Dr. Chavan is actively participating as a member of different professional research societies, like IEEE, ISTE, etc.

How to cite this paper: Sharada V Chougule, Mahesh S. Chavan, "Speaker Recognition in Mismatch Conditions: A Feature Level Approach", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.9, No.4, pp.37-43, 2017.DOI: 10.5815/ijigsp.2017.04.05