Marathi Speech Database Standardization: A Review and Work

Sonal A.Tiwari[1]
[1]Research Student, Department of CS & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India,
[1]sonaltiwari99@rediffmail.com

Rajashri G. Kanke[2]
[2]Research Student, Department of CS & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India,
[2]kankerajashri@gmail.com

Maheshwari A. Ambewadikar[3]
[3]Research Student, Department of CS & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India,
[3]mahiambewadikar@gmail.com

Manasi R. Baheti[4]
[4]Assistant Professor, Department of CS & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India,
[4]mrb.csit@bamu.ac.in

Abstract---**Automatic Speech Recognition System (ASR) is helpful for interaction between human and machine. It is the way to operate computer and mobile phones through speech only, without taking such extra efforts. The term corpus is used for Standardized Database, which contains a collection of audio recordings of spoken language with its annotations and documents. When existing literature was reviewed, it was observed that much literature is available on how to create speech databases. But few literatures are available about the standardization. Such work is done for the languages other than Indian languages. But for the Hindi, Marathi etc., standardization for the speech datasets is not up to the mark. The main problem in designing of a speech database is to deal with variability of speech. In recent years, there is much need to develop speech corpora for training and testing materials to be used for wide range of applications of speech technology like Linguistic Consortium, Speech interfaces development and language models etc. If it is standardized in regional languages, it will certainly contribute in many applications and research. In future, we would like to work to find standard way to standardized speech databases so with the help of this we can retrieve data easily and more efficiently.**

Keywords- ASR, Corpus, Speech Database, Standardization, Annotation

I. INTRODUCTION

The term Corpus is used for database therefore speech corpus is used to store data which is audio recordings of spoken language with its annotation and documents. Those speech corpora are useful to make research work more efficiently and also it saves time and money. The main aim of this research is to analyze work done for Marathi Speech Database.

Data in database may lack consistency because there are many ways to say the same thing. To enhance

The quality of retrieving a data from speech corpus we need to standardize our speech corpus. In standardization process we need to annotate data. Corpus annotation is the sub process of standardization, in this we add additional information about the data stored in Speech corpus. Additional information indicates addition of tags, labels, etc. Depending on the additional information which we added annotations are categorized into different types like Phonetic annotation, Prosodic annotations.

II. ABOUT MARATHI LANGUAGE

Maharashtra is Ranked second most populated state of India. As per the projection, Population of Maharashtra in 2021 is 12.62 crore i. e. 126.2 Million [20]. Out of 12.62 crore population, 8.3 crore Marathi people spokes Marathi language in Maharashtra [21]. Other than Maharashtra state, Union territory in the India sub- Continent- Dadra and Nagar Haveli and Goa state also uses Marathi language.

Among 22 scheduled languages of India Marathi language is one of them. Marathi has the third largest numbers of native speakers in India, after Hindi and Bengali [22].

Sanskrit is the oldest language, and Saurseni, Magadhi and Maharashtri are derived from Sanskrit Language. All those languages were easier languages as compared to Sanskrit and belong to the INDO-ARYAN language and all Indo-Aryan languages were derived from Sanskrit language.

Marathi language uses Devanagari script. A character represents one vowel and zero or more consonant Marathi language contains total 12 vowels and 36 consonants [6].

III. MARATHI SPEECH DATABASE

IIIT – International Institute of Information Technology, Hyderabad and HP lab Bangalore developed speech databases for various Indian languages. This database developed for large vocabulary speech Recognition systems. CIIL (Central Institute of Indian Languages) corpus of Marathi language was used for text corpus collection [5].

TIFR (Tata Institute of Fundamental Research, Mumbai) & **IIT Bombay** (Indian Institute of Technology) together developed a speech database for agriculture purpose. This Project sanctioned by the technology development for Indian languages (TDIL) for the development of speech recognition system for using cell phones & landline. TIFR Mumbai & IIT Bombay collected the speech data using two dedicated phone line. [5]

IIT-Khargpur also developed speech database for text independent speaker identification in ASR. Recording frequency was 22,500 Hz, mono channel, 16 bit resolution & 10 utterances. The recording was done in various environment

conditions like road home, office, slums, college, train, hills, valleys, remote villages, research labs &farms. The total numbers of speakers were 60 [5]. A noticeable works also done at Dr. B.A.M university Aurangabad by department of computer science & IT the speech database developed for limited vocabulary continuous speech database & two different isolated speech database for agricultures & travel purpose of Aurangabad & speech database of numbers in Marathi language.[5]

TABLE 1: Comparison of the Marathi Speech Databases

Sr. No	Database Developed by	Recording Environment	No of Speakers	Recording Device used	Application of Database
1.	TIFR Mumbai &IIT Bombay	Noisy Environment	1500	Cell Phones & Voice Recorders	Speech Recognition System for Agriculture Purpose
2.	IIIT Hyderabad & HP Labs Bangalore	Noisy Environment	559	Mobile Phones & Landlines	General Purpose
3.	Dr. B.A.M University	Noisy Environment	360	high-quality microphone	General Purpose

IV. ANALYSIS OF RESEARCH WORK DONE IN MARATHI LANGUAGE

The survey has been done for the speech recognition system developed for the Marathi language. The summary of this survey is as below:

P.P. Shrimal, R.R.Deshmukh, Vishal B. Waghmare (2012) developed isolated words speech database of Marathi words for agriculture purpose. They have used 100 words with 5 utterance of every word. Proper attention is given to select the speakers from different age groups. Implementation is done using PRAAT speech software. [7]

Yogesh K.Gedam, Sujata S. Magare, Amrapali C. Dabhade(2014) they described the speech recognition of Marathi digits using MFCC(Mel-Frequency Cepstral Coefficients) and DTW (Dynamic Time Warping) technique. In MFCC, it uses Mel-frequency scaling which is very approximate to the human system. Feature Matching is done with the help of Dynamic Time Wrapping. DTW technique is used to match non linear features in speech identification with high speed and low error rates.[6]

P.P. Shrimal,R.R.Deshmukh,Ganesh B. Janvale (2015)they describes the development of speech databases in Marathi language spoken in Marathwada region of Maharashtra state. It describes the development of two isolated speech database for agriculture domain, one isolated speech database for travel domain, isolated database for numbers, and a continuous speech database or agriculture purpose.

Manasi R. Baheti,Bharti W. Gawali, S.C. Mehrotra (2016)they have implemented Automatic Speech Recognition for Task Oriented IVRS in Marathi. In this speech recognition has become a foundation of self-service interactive voice response (IVR) user interfaces. The speech recognition and IVR applications found to be cost-effective and the user friendly, speediest self-service alternative to speaking with a contact centre agent.[3] The concept of making advanced systems reach up to rural level and solve the problem of communication gap is the main problem addressed in this paper. With improvements in "Speech Recognition" and IVRS providing 24/7 services, objective is to develop an IVRS in Marathi Language that can be useful for agricultural purpose, for rural area people. This can be done by considering specific, pre-decided Marathi Language sentences for IVR based agricultural commodity price information retrieval system developed mainly for the illiterate semiliterate or farmers. [3]

Santosh Gaikwad, Bharti Gawali, Suresh Mehrotra (2013) creates the Marathi speech corpus for Automatic Speech Recognition. Marathi speech corpus which is transcript using Google Unicode editor labeled and annotated using Praat. Optimal design and development of speech corpus for Marathi language. The simple methodology of database creation presented will serve as catalyst for creation of more dynamic speech databases for other Indian languages. [1]

V. ANALYSIS OF RELATED WORK DONE ON MARATHI LANGUAGE

TABLE 2: Analysis of Work Done In Marathi Language

Author	Year	Methods/Techniques	Title	Category
Gawali et.al	2010	MFCC, DTW	"Marathi isolated word Recognition System using MFCC and DTW features"	Speech
Gaikwad et. al	2011	IVR	"Polyclinic inquiry system sentences using IVR in Marathi Language"	Sentence
Santosh kahinath Gaikwad et. al	2013	MFCC,DTW	"Creation of Marathi Speech corpus for automatic speech Recognition System"	Speech
Yogesh K. Gedam et. Al	2014	MFCC, DTW, HMM, LBG	"Development of ASR of Marathi Numerals-A Review"	Numerals
Paras V. Kothare et. Al	2014	-	"Design and development of speech DB of Marathi Numerals "	Numerals
Baheti et. Al	2015	MFCC, DTW	"Marathi Interactive voice response system(IVRS) using MFCC & DTW"	-
Jinhal H. Tailor et.	2015	-	"Review on speech Recognition system for Indian	-

	al			Language "	
Mache et. al	2015	-	"Review on Text-To-Speech Synthesizer"		-
Baheti et. al	2016	MFCC,DTW	"Automatic Speech Recognition for Task Oriented IVRS in Marathi"		Sentence
Mrs. Chhaya S patil et.al	2017	-	"A Review on Marathi language speech database development for ASR system"		Speech

VI. EXPERIMENTAL WORK

In Experimental work, first most important step is, for developing ASR system we need to develop speech database. Speech recognition depends on the quality of the speech database. Therefore Marathi Speech database is also created for this study in our lab.

To create database we preferred PRAAT Software Tool. PRAAT is a free computer Software for Speech analysis. By using PRAAT we can create and manipulate sound files, Annotate sound files etc. in this study we created a database which contains isolated words and isolated digits of Marathi language. Data for the database is recorded by 5 speakers in 3 repetitions. For recording data High quality Microphone is needed. Sampling rate is set to 22050Hz which is the most sufficient frequency. The other details of the Database are given in the following table.

TABLE 3 CORPUS DETAILS

Sr. No	Words	Number of Utterances	Duration(in Seconds)	Gender	Age
1	आज	1	0.05	F	23
2	मला	1	0.03	F	23
3	तुला	1	0.03	F	23
4	आहे	1	0.03	F	23
5	पाणी	1	0.04	F	23

Sr. No	Words	Number of Utterances	Duration(in Seconds)	Gender	Age
1	आज	1	0.05	F	22
2	मला	1	0.03	F	22
3	तुला	1	0.03	F	22
4	आहे	1	0.03	F	22
5	पाणी	1	0.04	F	22

Sr. No	Words	Number of Utterances	Duration(in Seconds)	Gender	Age
1	आज	1	0.05	F	21
2	मला	1	0.03	F	21
3	तुला	1	0.03	F	21
4	आहे	1	0.03	F	21
5	पाणी	1	0.04	F	21

Sr. No	Digits	Number of Utterances	Duration(in Seconds)	Gender	Age
1	1	1	0.02	F	23
2	2	1	0.01	F	23
3	3	1	0.02	F	23
4	4	1	0.02	F	23
5	5	1	0.02	F	23
6	6	1	0.02	F	23
7	7	1	0.02	F	23
8	8	1	0.02	F	23
9	9	1	0.02	F	23
10	10	1	0.02	F	23

Sr. No	Digits	Number of Utterances	Duration(in Seconds)	Gender	Age
1	1	1	0.02	F	22
2	2	1	0.01	F	22
3	3	1	0.02	F	22
4	4	1	0.02	F	22
5	5	1	0.02	F	22
6	6	1	0.02	F	22
7	7	1	0.02	F	22
8	8	1	0.02	F	22
9	9	1	0.02	F	22
10	10	1	0.02	F	22

Sr. No	Digits	Number of Utterances	Duration(in Seconds)	Gender	Age
1	1	1	0.02	F	21
2	2	1	0.01	F	21
3	3	1	0.02	F	21
4	4	1	0.02	F	21
5	5	1	0.02	F	21
6	6	1	0.02	F	21
7	7	1	0.02	F	21
8	8	1	0.02	F	21
9	9	1	0.02	F	21
10	10	1	0.02	F	21

Standardization

After creating speech Database the main aim of this study begins that how to standardized this speech database that if we throw a speech query then exact value should be retrieved by the ASR system.

What is a Standardization

Standardization is used for structuring the data. Data in a database often lacks consistency simply because there are a lot of ways to say the same thing. Standardizing the record it means that when a query is run for a particular field, accurate results will be retrieved.

Following are the basic steps necessary to Standardize speech Corpus-

It includes many aspects as below-
1. Legal considerations
2. Standardization of speech corpus collection procedure-
    2.1 Project analysis and design
    2.2 Preparing for collection
    2.3 Pre-Collecting
    2.4 Pre- Validation

    2.5 Starting the real collection
    2.6 Annotating
    2.7 Compiling Lexical dictionaries
    2.8 Post Validation
    2.9 Distribution
3. Standardization of Speech corpus
    3.1 Specification of Speakers
    3.2 Specification of corpus Design
    3.3 Specification of Recording
    3.4 Specification of Annotation
    3.5 Validation Criteria
    3.6 Specification of Distribution

A. Transcription

We need a high quality transcription for our speech files. Preparation and coordination of speech file and speech transcription is important task. List of sound files must match the list of transcriptions .To represent transcription we need annotations graph which consist of strings of words.

B.   Annotation

Annotation of a speech corpus refers to all Symbolic information that is related to the speech signal. An annotation generally means describing, classifying and organizing speech material by systematically adding symbolic labels to its parts. We can annotate existing sound files and sound objects. The labeling data reside in a Text Grid object. To annotate-

1.   Creating a Text Grid
2.   View and Edit
3.   Save

To convert sound files to Text Grid in Praat we need to

- open praat object window
- Select .wav file
- Select Annotate to Text Grid
- Specify the names of tiers
- It creates .TextGrid file
- Select both .wav & .textGrid file and edit it.
- Give Annotation
- Save as Text Grid
- Quit

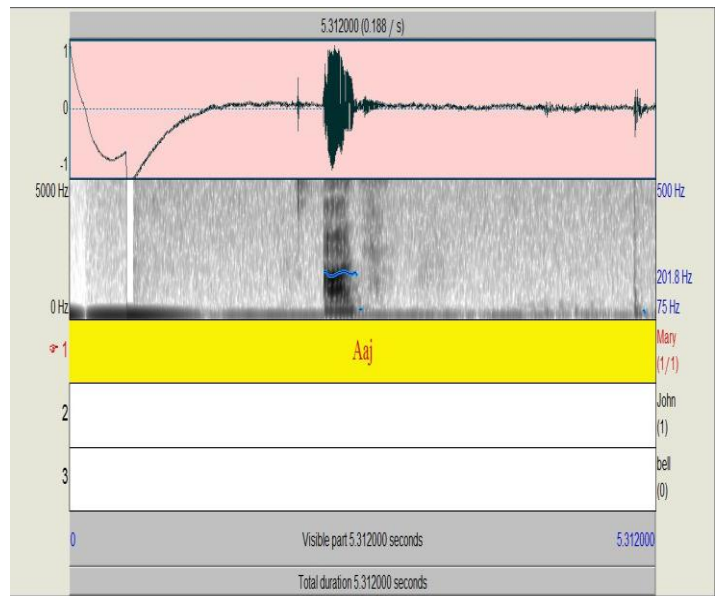The labeling and annotation of Marathi isolated word – "AAJ" using Praat is described in following figure

FIGURE 1 Praat Interface

FIGURE 2 Graphical Interfaces for Speech Labeling Of Marathi Isolated Words Speech Recognition
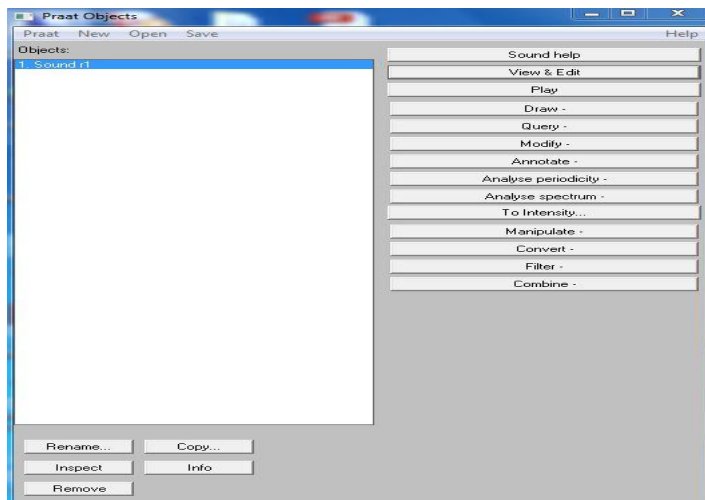
When we save the above file as Text Grid This text File contains the following pieces of data

- The real number 0 & 5.312, Which is the starting and end time (in Seconds) of the Text Grid.

- The flag <exists>, which tells us that this TextGrid contains tiers. we cannot create TextGrid objects without tiers in Praat.

- The integer number 3, which is the number of tiers that we created.

- The text "IntervalTier", which designates the type of the first tier.

- The text "Mary", which is the name we gave to the first tier.

- The real number 0, which is the starting time (in seconds) of the first tier.

- The real number 5.312, which is the end time (in seconds) of the first tier. When you create a TextGrid with an interval tier, the time domain of the interval tier is automatically set equal to the time domain of the whole TextGrid.

- The integer number 1, which is the number of intervals in the first tier. Although you did not add any intervals and did not add any text to the first tier, all interval tiers always contain at least one interval, which is created when you create the TextGrid.

- The text "", which is the contents of the interval on the first tier. This text is empty.

- The text "IntervalTier", which gives the type of the second tier (again an interval tier) etc[16]

VII. CONCLUSION

In this paper I have discussed some of the speech database and various works done in Marathi Speech database. In this progress I realized the importance of standardized speech corpus. From the analysis of literature, I concluded that there is little work done on Indian languages and especially in Marathi language as compared to other languages like Chinese, English etc. In Marathi language also almost work has been done on creation of speech corpus, but yet no work completed to standardize Marathi speech corpus. The objective of this work is to summarize and compare some of well known past literature and for my further work I would like to complete the standardization process on Marathi language speech Database. The importance of standardized speech Database is that, such database can be very useful in developing speech based task oriented interfaces, also it can be the contribution to linguistic database for various applications and studies.

REFERENCES

1. Santosh Kashinath Gaikwad, Dr. Bharti W. Gawali and Suresh Mehrotra, "Creation of Marathi speech corpus for automatic speech recognition" November 2013

2. Jinal H. Tailor, Dipti B. Shah, "Review on speech Recognition System for Indian Languages" International Journal of Computer applications, Volume 119,June 2015

3. Mansi Ram Baheti, Bharti Gawali, S.C. Mehrotra, "Automatic Speech Recognition for Task Oriented IVRS in Marathi",IOSR Journal of Computer Engineering, Mar-Apr 2016

4. Ai-jun Li and zhi-gang Yin, "Standardization of speech corpus" Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China, Data Science Journal, Volume 6, Supplement, 18 Novmber 2007.

5. Mrs. Chhaya S. Patil, Prof. Dr. Vshali B. Patil, "A Reviw on Marathi Laguage Speech database development for Automatic speech Recognition",Int. Journal of Engineering Research and Application, ISSN:2248-9622, vol 7, Issue 3, March 2017.

6. Yogesh K. Gedam, Sujata S. Magare, Amrapali C. Dabhade, Ratnadeep R. Deshmukh,"Development of Automatic Speech Recognition of Marathi Numerals", IJEIT volume 3,Issue 9, March 2014.

7. P.P. Shrishrimal, R.R. Deshmukh, Vishal B Waghmare, "Development of Isolated words speech database of Marathi words for Agriculture purpose",

Asian Journal of computer science and Information Technolgy,2012.

8. Sangramsing N. Kayte, Monica Mundada, Dr. Charansing N. Kayte, Dr. Bharti Gawali, "Implementation of Text To speech for Marathi Language Using Transcriptions Concept",Int. Journal of Engineering Research and Appllications,ISSN:2248-9622, vol 5,Issue 11, November 2015.

9. Paras V. Kothare, Ratnadeep R.Deshmukh, "Design and development of Speech Database of Marathi Numerals", March 2014.

10. Li, A., Yin, Z., Wang, T., Fang, Q., & Hu, F. (2004) RASC863 - A Chinese Speech Corpus with Four Regional Accents. ICSLT-o-COCOSDA: New Delhi, India.

11. Li, A. & Zu, Y.(2006) Corpus Design and Annotation for Speech Synthesis and Recognition. In Advances in Chinese Spoken Language Processing. Lee, C., Li, H., Lee, L., Wang, R., &Huo, Q. (eds.) World Scientific, Publishing Co. (in press)

12. Yin, Z. (2006) The introduction of speech corpus research and establishment. The newspaper of CASS.

13. Anumanchipalli G., Chitturi R., Joshi S., Kumar R., Singh S. P., Sitaram R. N. V. , S. P. Kishore, 2005.

Development of Indian Language Speech Databases for Large Vocabulary Speech Recognition Systems. In: Proceeding of International Conference on Speech and Computer (SPECOM), Patras, Greece, pp. 591-595.

14. Tejas Godambe and Samudravijaya K., "Speech Data Acquistion for Voice based Agricultural Information Retrieval", presented at the 39th All India DLA Conference, Punjabi University, Patiala, 14-16th June 2011

15. K. Samudravijay, 2010. Multilingual Telephony Speech Corpora of Indian Languages. In Proceeding Computer Processing of Asian Spoken Languages. Eds, S. Itahashi Ans C. Tseng, Consideration Books Los Angeles, pp.189-193

16. Agrawal S. S., K.K. Arora, S Arora, Samudravijaya K, "Text and Speech Corpus Development in Indian Languages", ibid, pp. 94-97

17. http://www.fon.hum.uva.nl/praat/, cited on 03/08/2013

18. http://www.indiaonlinepages.com/population/india-current-population.html

19. https://www.coursehero.com/file/p49barkc/These-are-Assamese-Bengali-Bodo-Dogri-Gujarati-Hindi-Kannada-Kashmiri-Konkani/

20. https://www.quora.com/How-many-languages-are-there-in-India

21. https://www.census2011.co.in/census/state/maharashtra.html

22. https://en.wikipedia.org/wiki/Marathi_language

23. https://censusindia.gov.in/2011Census/C-16_25062018_NEW.pd

IJCSIS Call For Papers 2021-2022

https://sites.google.com/site/ijcsis/

The topics suggested by the journal can be discussed in term of concepts, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete *unpublished papers*, which are *not under review in any other conference or journal* in the following, but not limited to, topic areas. All tracks are open to both research and industry contributions.

Ad Hoc & Sensor Network	**Knowledge based systems**
Ad hoc networks for pervasive communications	**Knowledge management**
Adaptive, autonomic and context-aware computing	**Location Based Services**
Advanced Computing Architectures and New Programming Models	**Management information systems**
	Medical imaging
Agent-based middleware	**Micro/nano technology**
Autonomic and self-managing middleware	**Middleware Issues**
B2B and B2C management	**Middleware services and agent technologies**
BioInformatics	**Mobile and Wireless Networks**
Bio-Medicine	**Mobile Computing and Applications**
Biotechnology	**Mobile networks and services**
Broadband and intelligent networks	**Multimedia Communications**
Broadband wireless technologies	**Multimodal sensing and context for pervasive applications**
Cloud Computing and Applications	
Collaborative applications	**Multisensor fusion**
Communication architectures for pervasive computing	**Natural Language Processing**
Communication systems	**Network management and services**
Computational intelligence	**Network Modeling and Simulation**
Computer and microprocessor-based control	**Network Performance; Protocols; Sensors**
Computer Architecture and Embedded Systems	**Networking theory and technologies**
Computer Business	**Neural Networks**
Computer Vision	**Neuro-Fuzzy and applications**
Computer-based information systems in health care	**Open Models and Architectures**
Computing Ethics	**Open Source Tools**
Context-awareness and middleware	**Operations research**
Cross-layer design and Physical layer based issue	**Optical Networks**
Cryptography	**Pattern Recognition**
Data Base Management	**Peer to Peer and Overlay Networks**
Data Mining	**Perception and semantic interpretation**
Data Retrieval	**Pervasive Computing**
Decision making	**Performance optimization**
Digital Economy and Digital Divide	**Positioning and tracking technologies**
Digital signal processing theory	**Programming paradigms for pervasive systems**
Distributed Sensor Networks	
E-Business	**Quality of Service and Quality of Experience**
E-Commerce	**Real-time computer control**
E-Government	**Real-time information systems**
Emerging signal processing areas	**Real-time multimedia signal processing**
Enabling technologies for pervasive systems (e.g., wireless BAN, PAN)	**Reconfigurable, adaptable, and reflective middleware approaches**
Encryption	**Remote Sensing**
Energy-efficient and green pervasive computing	**RFID and sensor network applications**
Event-based, publish/subscribe, and message-oriented middleware	**Scalability of middleware**
	Security and risk management
Evolutionary computing and intelligent systems	**Security middleware**

Expert approaches	**Security, Privacy and Trust**
Fuzzy algorithms	**Security Systems and Technolgies**
Fuzzy logics	**Sensor array and multi-channel processing**
GPS and location-based applications	**Sensor fusion**
Green Computing	**Sensors and RFID in pervasive systems**
Grid Networking	**Service oriented middleware**
Healthcare Management Information Technology	**Signal Control System**
Human Computer Interaction (HCI)	**Signal processing**
Image analysis and processing	**Smart devices and intelligent environments**
Image and multidimensional signal processing	**Smart home applications**
Image and Multimedia applications	**Social Networks and Online Communities**
Industrial applications of neural networks	**Software Engineering**
Information and data security	**Software engineering techniques for middleware**
Information indexing and retrieval	
Information Management	**Speech interface; Speech processing**
Information processing	**Supply Chain Management**
Information systems and applications	**System security and security technologies**
Information Technology and their application	**Technology in Education**
Instrumentation electronics	**Theoretical Computer Science**
Intelligent Control System	**Transportation information**
Intelligent sensors and actuators	**Trust, security and privacy issues in pervasive systems**
Internet applications and performances	
Internet Services and Applications	**Ubiquitous and pervasive applications**
Internet Technologies, Infrastructure, Services & Applications	**Ubiquitous Networks**
	User interfaces and interaction models
Interworking architecture and interoperability	**Virtual reality**
	Vision-based applications
	Web Technologies
	Wired/Wireless Sensor
	Wireless technology