

# Finding Meta Data in Speech and Handwriting Biometrics

Claus Vielhauer<sup>a</sup>, T. Basu<sup>b</sup>, Jana Dittmann<sup>a</sup>, P.K. Dutta<sup>b</sup>

<sup>a</sup>Otto-von-Guericke University Magdeburg, Universitaetsplatz 2, D-39106, Magdeburg, Germany

<sup>b</sup>Indian Institute of Technology, Kharagpur, India

## ABSTRACT

The goal of this paper is to present our work on the analysis of speech and handwriting biometrics related to meta data, which are based on one side on system hardware specifics (technical meta data) and on the other side to personal attributes (non-technical meta data). System related meta data represent physical characteristics of biometric sensors and are essential for ensuring comparable quality of the biometric raw signals. Previous work in personal related meta data has shown that it is possible to estimate some meta data like script language, dialect, origin, gender and age by statistically analyzing human handwriting and voice data. On one hand, by knowing both kinds of such meta data, it appears to be possible to adapt the recognition or authentication algorithms in order to enhance their recognition accuracy and to analyze the sensor dependency of biometric algorithms with respect to hardware properties such as sampling resolution. On the other hand, interesting aspects are to evaluate, if cultural characteristics (such as native language, or ethnicity) can be derived by statistical or analytical means from voice or handwriting dynamics and to which degree grouping of users by persons with identical or similar meta data may result in better biometric recognition accuracy. All these aspects have been widely neglected by research until today.

This article will discuss approaches to model such meta data and strategies for finding features by introducing a new meta data taxonomy, from which we derive those personal and system attributes related to the cultural background, which are employed in our experimental evaluation. Further, we describe the test methodology used for our experimental evaluation in different cultural regions of India and Europe and present first results for sensor hardware related meta data in handwriting biometrics as well as language related meta data in speaker recognition.

**Keywords:** Biometrics, speaker recognition, signature verification, sensor interoperability, meta data, soft biometrics, cross culture, multilingual biometrics

## 1. Introduction

Motivated from large-scale cross-cultural application for user authentication and from the recognition results presented in [ToKS2004], it is to be possible to estimate some meta data like script language, origin, gender and age by statistically analyzing human biometrics. By knowing this meta data it seems to be possible to adapt the recognition or authentication algorithms in order to enhance their performance/quality/accuracy (i.e. False-Match/False-Non-Match Rates, FMR/FNMR). Especially in multi-cultural authentication scenarios this additional meta data can therefore help to improve the overall system accuracy.

In our previous work [Schi+2004], we have presented our novel methodology of collecting biometric data based on speech and handwriting on an international, multi-cultural scope. Based on our methodology, we have collected biometric data along with personal meta data consisting of more than 20 attributes. Based on this first research on personal meta data, we identified technical meta data as relevant besides personal (non-technical) during capturing and sampling of biometric speech and writing data input. From [Viel2004] we know that technical meta data about the sensor characteristics are relevant parameter to achieve comparable verification results for example for a cross-sensor authentication.

In this paper, we introduce in section 2 an enhanced taxonomy of meta data with technical and non-technical parameter for the biometric modalities of speech and handwriting, which become relevant in cross-cultural applications. The prerequisite of the usage of personal meta data is the knowledge about the impact of technical parameter introduced during capturing and sampling of the input signals. If biometric devices produce different signal quantities and qualities, the comparison of the signals itself as well as on their personal meta data level becomes difficult or rather low statistical significant caused by the wrong signal interpretation. Without appropriate hardware or comparable hardware characteristics, any meta data analysis cannot be performed correctly. From our discussion in section 2 we see that

especially for handwriting signals we find a wide variety of sensors with different hardware characteristics, which need to be determined before the overall personal meta data test can be started.

In section 3 we introduce a methodology for a cross-sensor evaluation to estimate the cross-sensor error rate as well as recommend specific hardware devices for cross-cultural personal meta data evaluation for handwriting biometrics in section 3. On the example of handwriting we discuss problems during signal capturing and sampling and identify that different hardware characteristics result in different signal characteristics.

As the impact of audio hardware is already known for example from [Mic2002], we used this technical parameter specification and in section 4 we introduce our first results for our analysis on speech and speaker related personal meta data. We show how we can find features in the biometric data for estimation of meta data in groups or subsets of our test population. One interesting focus of our work is to evaluate, if cultural characteristics (such as native language, or ethnicity) can be derived by statistical or analytical means from the biometric data.

## 2. A Taxonomy for Meta Data

The basic concept behind biometric data is the binding of digitized measurements of human physiology or behavior to an identity and to store this set of information for later reference during an authentication system. Typically, in the case of active or behavioral biometric traits, these biometric data are stored as digital signal representations; examples for such data are voice recordings for speaker authentication or pen position signals for online handwriting analysis. For passive or physiological biometrics, data are represented as sets of discrete data, for example as 2-dimensional images for face or fingerprint recognition.

From the overall view to a biometric system we can summarize the following taxonomy of meta data for biometric applications. Firstly, there are technical and non-technical meta data, the taxonomy shall be explained along the following Figure 1:

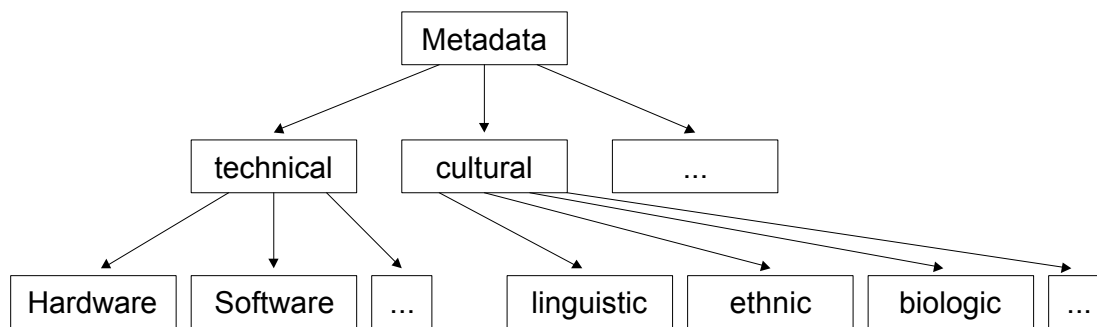


Figure 1 - Taxonomy of biometric meta data

While the technical aspects, which are shown in the left branch of the figure, include characteristics of the hardware used for capturing and sampling of the biometric signals as well properties of the software used for quantization and for the biometric classification method, other types of meta data may be of entirely non-technical nature. From a cultural perspective and previous work by [ToKS2004] and [HYHI2004], we have chosen a classification into linguistic, ethnic and biologic data. In the following subsection we briefly discuss technical and non-technical meta data, which are considered in our test and evaluation.

### 2.1 Technical Meta Data

From the technical point of view, several hardware and also software parameter of a biometric system are responsible for:

- which signal characteristics can be captured,
- in which quality the analog digital conversion can be performed and
- how the signals are sampled and quantized or compressed.

For example as shown in [Viel2004], sampling devices for recording of handwriting signals vary widely in view of the physical modalities, which are measured. Mainly two categories of sampling devices can be identified today: tablet-

based and pen-based digitizers. Devices of the first category record at least one signal from the writing surface, whereas the second category of digitizers can be used on arbitrary surfaces, as the signals are exclusively sampled in the pen. We have identified the following typical physical measurements signals, which can be provided by industrial digitizer tablets today and which are made available to software programming interfaces by the device drivers as sampled signals:

- Horizontal pen position signal  $x(t)$  ,
- Vertical pen position signal  $y(t)$  ,
- Pen tip pressure signal  $p(t)$  ,
- Pen azimuth signal  $T(t)$  ,
- Pen altitude signal  $F(t)$ .

Additional physical measurements have been presented for special, force-sensitive pens, for example the device used by Martens and Claesen [MaC11996], allowing the reconstruction of acceleration signals:

- Horizontal pen acceleration signal  $a_x(t)$  (via horizontal pen force),
- Vertical pen acceleration signal  $a_y(t)$  (via vertical pen force).for handwriting sampling devices

Variations can be found in temporal and spatial resolutions of sensors, but also with respect to dimensionality (especially for movement sensors). Biometric signals are exposed to distortions based on the sensor characteristics, which may cause problems if verification information has been generated from a sensor having different specifications than the sensor used to obtain the actual authentication information from the user. This problem is relevant especially for applications in large areas, where sensors with identical characteristics might not be available in every location, as well as in long-term considerations, where specific hardware might no longer be available after some time.

For the great majority of biometric authentication systems that can be found today, there are very few statements to their robustness with respect to cross-sensor verification. For large-scale deployment, it is important to examine the dependencies between biometric authentication algorithms and sensors. In section 3 we therefore introduce an evaluation methodology for cross-sensor verification for handwriting biometrics and present our first results and recommendations for technical parameter sets, which should be used for personal meta data evaluation on top.

## 2.2 Cross-Cultural – non-technical Meta Data

Biometric authentication techniques have developed a degree of maturity over the past years, which has enabled large-scale applications such as biometric ID cards, which are currently introduced in many countries. However, the impact of cultural aspects to biometrics has not been addressed sufficiently today. Consequently, cultural and cross-cultural impacts to techniques such as biometrics constitute one of the primary research goals in the CultureTech project<sup>1</sup>, which is an ongoing activity performed by institutes in Europe and India.

India is a vast country with a multicultural/multilingual composition. There are 17 major Indian Languages and more than two hundred tribal languages/ dialects, which is a great challenge for multilingual biometrics ([PaBa2004]). There are also dialectal variations in each of these major languages too. Classifying the dialectal groups in some of these languages on the basis of group specific features was one of the tasks in this project. Investigations were also made to identify speakers of different languages in a multilingual environment in both open set and closed set modes. A group of speakers from German language was also included in the set to check the effectiveness of the culture study model.

The data was collected following the internationally accepted guidelines to build the corpus. Some of the important details are given here. In all the cases of data collection, voice activated tape recorders have been used and the interviewer's voice is deleted from the speech file; training segments of durations 30 s, 90s and 120s and testing segments of 1s,3s,5s, 7s, 10s 15s were used for the design of the corpus. The sampling frequency was 22.05 kHz and frame sizes of 512 samples were used with 50% overlap. LPC models (Linear Prediction Coefficient) of 12<sup>th</sup> order was used and LPCC (Linear Predictive Cepstral Coefficients) and other feature sets were derived from these, see for example in [PrMi2001].

As a result of our taxonomy, we have come to a meta data model, which has been introduced in [Schi+2004]. The following meta data categories are requested and stored within the system. For the sake of standardization, we use ISO

---

<sup>1</sup> This publication has been produced partly with the assistance of the European Union (project CultureTech, see <http://ams1-smb.cs.uni-magdeburg.de/culturetech/>). The content of this publication is the sole responsibility of the University Magdeburg and their co-authors and can in no way be taken to reflect the views of the European Union.

definitions to describe names of countries, languages and scripts. For our first non-technical meta data evaluation, we have collected the following categories of person related meta data:

- Gender (female or male),
- Age,
- Handedness (right or left),
- Ethnicity (white, black, hispanic, asian, ...),
- Religion,
- Highest level of education,
- Place of birth (ISO-3166),
- Place of birth of parents (ISO-3166),
- Place of schooling (ISO-3166)
- Native language (ISO-639),
- Known other languages (ISO-639),
- Native script (ISO-15924),
- Known other scripts (ISO-15924).

In this paper we want to present some first results of our project for linguistic aspects, consequently the “Native language” and “Known other languages” attributes are of relevance here, whereas the collected data of the other categories will be used for future evaluations. The following section will discuss our findings with respect to hardware characteristics whereas section 4 will expand upon the exploration of cultural meta data.

### 3. Hardware Related Meta data for Handwriting Biometrics

Little knowledge exists on the effects of digitizer hardware with different physical characteristics to biometric authentication algorithms based on online handwriting. Here, we are interested in two aspects:

- To which degree does the **accuracy** of biometric authentication depend on spatial and/or temporal resolutions of sensor hardware?
- What is the impact of **cross-sensor authentication** if sensors possess different technical properties?

In order to be able to empirically evaluate these aspects, we define a methodology, which is implemented in the [Viel2004] evaluation system, PlataSign and a database of test samples. This evaluation framework is used for an empirical analysis, which we perform by applying the test methodology to an exemplary verification algorithm and we show quantitative results.

As introduced in subsection 2.1 the physical parameter of digitizer hardware are quite different, which is also reflected by the fact that our PlataSign evaluation system currently supports three different device drivers. As one of our main goals is the analysis of handwriting-based authentication algorithms on different hardware platforms, a classification of tablets by their physical characteristics and the sampling method is required. The classification, which has emerged in the context of this work is based on the two aspects of spatial resolution and signal dimensionality.

**Spatial Resolution:** depending on the physical characteristics of the digitizer hardware and the implementation of device drivers, we introduce a classification depending on the resolution of the horizontal and vertical pen position signal. Here, two categories of interfaces can be found: firstly, many of the digitizer tablets can be operated in a mouse-emulation mode, where the digitizer pen is used as a pointing device, replacing the computer mouse. In this mode, the spatial resolution is limited by the screen resolution of the display and may be independent from the physical resolution of the digitizer. For some digitizers, especially those used in Personal Digital Assistants (PDA), this resolution is a consequence of the physical layout of the device, where the digitizer resolution is identical to the screen resolution and the digitizer surfaces are affixed directly above the screen surface of the device. Other digitizers implement mouse emulation by mapping of the  $x/y$  signals in the original resolution of the tablet onto the actual screen resolution of the display. This mapping is performed by the device driver. Effectively, the spatial resolution of mouse driver emulation and PDAs are in the range of app. 100 lines per inch (lpi), whereas today device drivers providing the full resolution of external digitizer tablets can achieve spatial resolutions in the range of 500 to 3000 lpi.

**Signal Dimensionality:** all digitizer tablets provide at least three dimensions of signals: a horizontal and a vertical position signal ( $x(t)$  and  $y(t)$ ) plus a signal for the pen pressure (see subsection 2.1). The later may, in the trivial case, be a binary pen-up/pen-down signal  $p_{PenDown}(t)$ , denoting if the pen at a specific point in time has applied a sufficient

amount of pressure to the tablet surface to consider that the user intends to generate a writing sequence. More sophisticated devices provide the pressure signal in a higher quantization resolution, e.g. up to 1024 quantization steps for some of the devices considered in this work. Above these three dimensions of signals, some devices provide additionally measurements of the angle of the pen above the writing plane (altitude angle) and in the heading direction of the writing plane (azimuth angle).

Considering the different characteristics, we have determined a classification scheme, which allows grouping of the tablets in to categories of similar physical characteristics. The classification attributes are:

- **Spatial Signal Resolution:** 3 categories: Low (Screen Resolution, < 100 lpi), Medium Resolution (100 lpi = Resolution < 2000 lpi) and High Resolution (>2000 lpi),
- **Pressure Signal Quantization:** 2 categories: Binary Signal (PenUp/PenDown) or Quantization Steps > 2,
- **Pen Angle Signals:** 2 categories: Azimuth and Altitude signals are available (yes/no).

The following Table 1 classifies all tablets used for the evaluation task in the context of this work. Different tablet types are categorized in the rows of the table, whereas the hardware classifications are denoted by symbols “x” in the columns. For the work presented in this paper, we have chosen to consider only test data of tablets having a significant number of samples (>500) in our test database, representing a specific category.

Tablet ID	Spatial Signal Resolution			Pressure Signal Resolution		Pen Angle Signal	
	Screen (<100lpi)	Medium Resolution (100lp >= res <2000lpi)	High Resolution (>= 2000lpi)	Binary	Q>2	Azimuth	Altitude
2	X			X			
5	X			X			
6	X			X			
21	X			X			
22	X			X			
24	X			X			
26	X			X			
7		X			X		
9		X			X		
16		X			X		
17		X			X		
19		X			X		
28		X		X			
27		X		X			
1			X		X	X	X
4			X		X	X	X
8			X		X	X	X
12			X		X	X	X
11			X		X		
23			X		X		
25			X		X		
29			X	X		X	X

**Table 1 - Classification by physical characteristics**

This classification scheme is used in our tests. It allows the evaluation of methods across different individual tablet types and between groups of tablets having similar properties, in order to draw conclusions regarding hardware-dependency of authentication algorithms. The test data from PlataSign evaluation system (see [Viel2004]) has been well structured with respect to hardware devices, which has been built up over a period of approximately four years. For the tablet categorization, five classes have been defined, as can be seen from the following Table 2, where the structure of the each of the tablet classes can be described as follows:

- **Tablet\_A:** this class consists of writing samples collected on one single type of tablet device (digitizer integrated in TFT computer display, such as used in state-of-the art tablet PCs for example), for which the most total number of

writing samples (5265) have been recorded. Also it has shown in the practice of collecting test samples that this device is most effective in for the generation of forgeries.

- **Tablet\_B**: this category also reflects one specific type of digitizer tablet, which is based on a pressure-sensitive surface sensor, independently of the pen used. Here, both samples generated by an inking ball-pint pen and writing generated by a non-inking pointing device have been considered in this class.
- **MIDRES-PQ**: this group contains the joint set of all tablet types, which provide the pen position signal at a medium resolution (i.e. between 100 and 2000 lines per inch, see Table 1), plus a pressure signal at a quantization level higher than two.
- **HIRES-PQ-Angle**: groups all samples, which have been collected on devices with a high spatial resolution (more than 2000 lines per inch), plus signals of the pen angles (azimuth  $T(t)$  and altitude  $F(t)$ ) during the writing process.
- **All**: this group is defined by the union set of all above groups.

Tablet	Test-Environment ID				
	Tablet_A	Tablet_B	MIDRES-PQ	HIRES-PQ-	All
2					X
5					X
6					X
21					X
22					X
24					X
26					X
28					X
7	X		X		X
9			X		X
16		X	X		X
17		X	X		X
19		X	X		X
27		X			X
1				X	X
4				X	X
8				X	X
12				X	X
11					X
23					X
25					X
29					X

**Table 2 - Tablet groups for the test environment definition**

Consequently, the entire set of test scenarios consists of five tablet categories. For our first evaluation, we have decided to look only at the variation in the discriminatory power between the five tablet categories. In order to do so, we have collected signature samples from a total population of 97 users, where each of the users had to provide at least one enrollment to the system, consisting of a minimum of four writing samples each. Further, a collection of a minimum of five verification samples had been asked from the test subjects. Table 3 presents the resulting total numbers of samples in each category are shown in the two rightmost columns. Note for this initial work towards technical meta data analysis, we refrained from exposing the system to skilled forgeries, although the observations in [Viel2004] indicate that a decrease in recognition accuracy in the order of one magnitude can be expected.

Tablet Group ID	No. of Persons	Enrollments	Verifications
Tablet_A	22	984	600
Tablet_B	12	545	211
MIDRES-PQ	32	2264	851
HIRES-PQ-Angle	36	633	294
All	97	3474	1796

**Table 3 - Total number of writing samples in our test scenario**

For the verification algorithm in our experiments, we have chosen the *Minimum Quadratic Envelope Distance* (MQED), as introduced in [Viel2004], which is a function based approach estimating the cumulative distance between the actual  $x(t)$  and  $y(t)$  signals of a handwriting samples and a reference signal envelope, derived from the enrollment samples. Our experimental metric is the Equal Error Rate (EER), which is derived from the error characteristics of False Rejection Rates (FRR) and False Acceptance Rates (FAR). The first is defined as the rate of falsely not confirmed identities of authentic users and has been experimentally determined by verification test of all verification samples against their corresponding enrollments, resulting in a graph as function of a decision threshold parameter  $T$ . The second graph, also as function of  $T$ , is determined by the method of random forgeries, i.e. the verification test of all enrollments of all users against verification samples of all non-authentic verification samples of all other users.

The resulting error rate diagrams are shown in Figure 2. From our tests we see, that the degree the accuracy of a biometric authentication algorithm for handwriting verification is dependent on the tablet category. To discuss this aspect, we analyze the graphs in the following Figure 2, representing the error rate characteristics of the same semantic, *Signature*, in tests of the MQED algorithm, for four different tablet categories: a) *Tablet\_B*, b) *Tablet\_A*, c) *All* and d) *Midres-PQ*. The figures show FAR and FRR on the ordinate as function of the decision threshold parameter on the abscissa. Note that in our first experiments, scales have not been normalized, thus only Figure 2b possesses a visible EER. For the remaining diagrams, the EER has been estimated by graphical interpolation of the graphs.

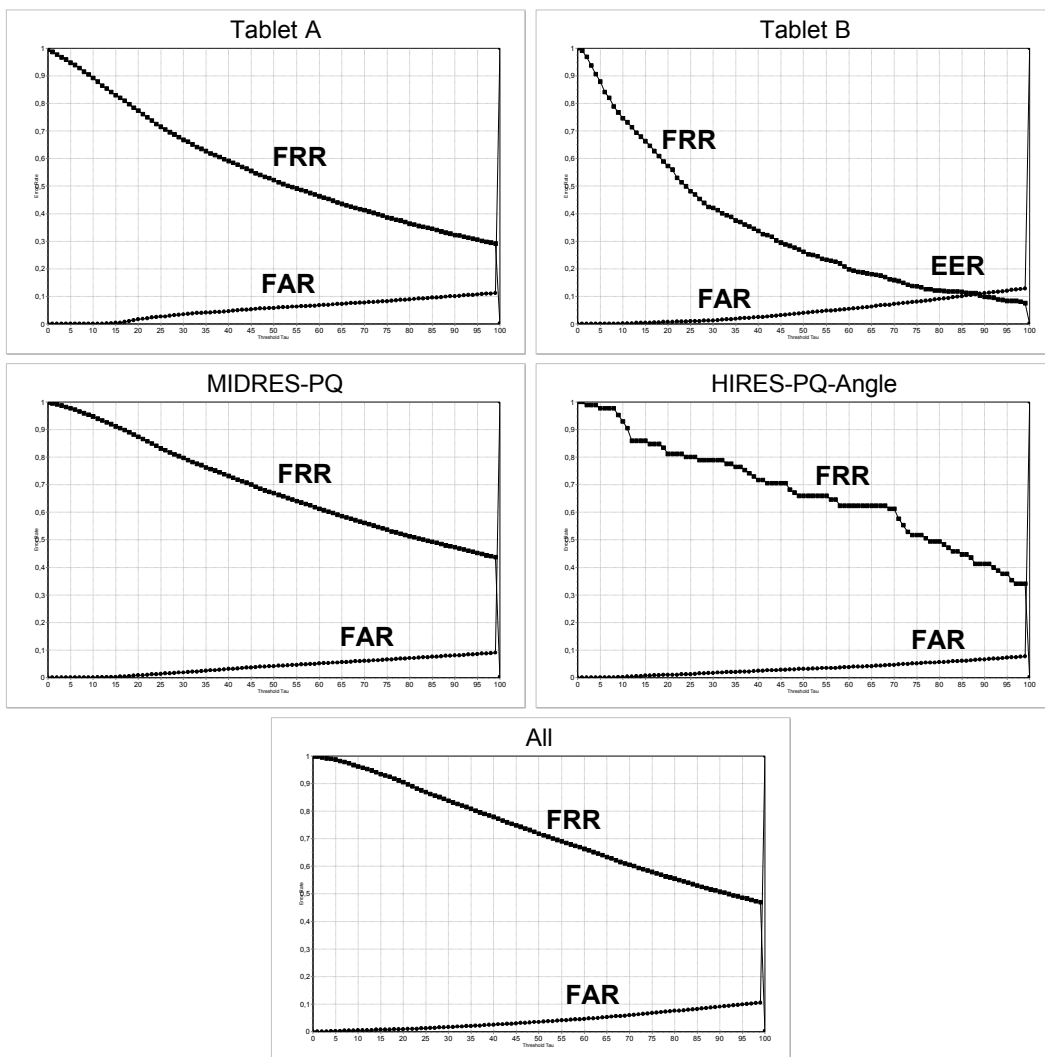


Figure 2 a),b),c),d),e) - Error rate diagrams for five different tablet categories, semantic class *Signature*

The Equal Error Rates graphically obtained from the error rate diagrams in Figure 2 are shown in the following Table 4.

Equal Error Rate (EER)	
Tablet Category	EER
<i>Tablet_A</i>	0,12
<i>Tablet_B</i>	0.1
<i>MIDRES-PQ</i>	0.15
<i>HIRES-PQ-Angle</i>	0.11
<i>All</i>	0,18

**Table 4 - Equal Error Rates for different tablet categories, semantic class *Signature***

From the graphically determined EER for all three tablet categories, as given in Table 4, we conclude the following: from the  $EER_{Random}$  of *Random Forgeries*, as a measure for the Inter-Class discriminatory power, we observe that the tablet category comprising digitizer devices of greatest inhomogeneity shows the lowest accuracy with an  $EER_{Random}$  of 18% (see row *All*, column  $EER_{Random}$  in Table 4). The best tablet set in this category is *Tablet\_B*, containing only one type of digitizer tablets with an  $EER_{Random}$  of 10% (see row *Tablet\_B*, column  $EER_{Random}$  in Table 4). The other table category consisting only of one specific device type, *Tablet\_A*, shows a similar  $EER_{Random}$  of 12% (see row *Tablet\_A*, column  $EER_{Random}$  in Table 4), whereas the tablet set consisting of tablets of medium spatial resolution with respect to the classification in, *Midres-PQ*, with an  $EER_{Random}$  of 15% is closer to category *All* (see row *Midres-PQ*, column  $EER_{Random}$  in Table 4). Thus, the highest degradation between homogeneous and heterogeneous sets yields to 80% increase in EER (from 10% increase to 18%  $EER_{Random}$ ) between tablet categories *Tablet\_B* and *All*, and the lowest to 25% (from 12% to 15%  $EER_{Random}$ ) from *Tablet\_A* to *Midres-PQ*.

#### 4. Evaluation of non-technical Meta Data for Speech

The algorithms and feature sets used in our experiments for speech processing, are common methods in speaker recognition and classification or verification exercises. Methods used in our experiments are Linear Predictive Coding parameters (LPC), Linear Predictive Cepstrum Coefficients (LPCC), and Mel Frequency Cepstrum Coefficient (MFCC). In our work we have used all these techniques along with a new classifier that is a Polynomial Classifier of 2nd and 3rd orders ([PaDB2004]).

##### 4.1 Language Dependency Experiment

In the first speech experiment, we have analyzed the dependency of recognition accuracy of our algorithms based on the type of language spoken by subjects. For this purpose, we have collected utterances of a population of four speakers (3 male + 1 female), each of which was recorded according to the test plan as described in section 2. Results of this experiment are shown in Table 5 to Table 10, where utterance durations during training (enrollment) are given in the columns and test durations (verification) in the rows. Further, TR denotes the training language, whereas TE the testing language. The algorithms used in the test are abbreviated as follows: LPC- Linear Predictive coefficient, LPCC- Linear predictive cepstral coefficient and MFCC – Mel frequency cepstral coefficient.

TRAIN TEST	30s	60s	90s
1	100	100	100
3	100	100	100
5	100	100	100
7	100	100	100
10	100	100	100
12	100	100	100
15	100	100	100
Avg. rate	100	100	100

**Table 5 - Success Rates with 12 LPC, TR= German , TE=German**



TRAIN TEST	30s	60s	90s
1	100	100	100
3	100	100	100
5	100	100	100
7	100	100	100
10	100	100	100
12	100	100	100
15	100	100	100
Avg. rate	<b>100</b>	<b>100</b>	<b>100</b>

**Table 6 - Success Rates with 12 LPCC, TR= German , TE=German**

TRAIN TEST	30s	60s	90s
1	100	100	100
3	100	100	100
5	100	100	100
7	100	100	100
10	100	100	100
12	100	100	100
15	100	100	100
Avg. rate	<b>100</b>	<b>100</b>	<b>100</b>

**Table 7 - Success Rates with 12 MFCC, TR= German , TE=German**

TRAIN TEST	30s	60s	90s
1	100	100	100
3	75	75	75
5	100	100	100
7	100	100	100
10	100	100	100
12	100	100	100
15	100	100	100
Avg. rate	<b>96.42</b>	<b>96.42</b>	<b>96.42</b>

**Table 8 - Success Rates with 12 LPC, TR= English , TE= English**

TRAIN TEST	30s	60s	90s
1	100	100	100
3	75	100	100
5	100	100	100
7	100	100	100
10	100	100	100
12	100	100	100
15	100	100	100
Avg. rate	<b>96.42</b>	<b>100</b>	<b>100</b>

**Table 9 - Success Rates with 12 LPCC, TR= English , TE= English**

TRAIN TEST	30s	60s	90s
1	75	75	75
3	100	100	75
5	100	100	100
7	100	100	100
10	100	100	100
12	100	100	100
15	100	100	100
Avg. rate	<b>96.42</b>	<b>96.42</b>	<b>92.85</b>

**Table 10 - S Success Rates with 12 MFCC TR= English , TE= English**

Our observations from this test can be summarized as follows:

- Success rates of 100% were achieved in all the cases of training and testing durations when training and testing is done with German. This may be due to the fact that the subjects considered were having German as their native language and the population size is very low.
- Success rates degrade for some cases of training and testing durations when training and testing is done with English. This may be due to the fact that the subjects considered were having English as their non-native language.

#### 4.2 Language Detection Experiment

The experiment on multilingual speaker identification, including six languages spoken in India as well as German was conducted based on these three methods, LPC, LPCC and MFCC and the success rates are quite encouraging. In this experiment, again four subjects have been asked to record speech samples based in the test plan presented in section 2 in their respective mother tongue language, where the goal in this experiment was to evaluate to which degree the three algorithms are able to classify the language spoken correctly. Our test include speakers from each of the following languages: Marathi (M), Hindi (H), Urdu (U), Bengali (B), Tamil (TA), Telugu (TL) and German (GE). In Table 11 to Table 13-18, ACT represents actual language of the speaker (shown in the columns) and IDENT represents identified language of unknown speaker, given in the rows. Cell values represent the correct recognitions in percent, relative to the total number of tests.

ACT. IDENT.	M	H	U	B	TA	TL	GE
M	<b>100</b>	0	0	0	0	0	0
H	0	<b>100</b>	0	0	0	0	0
U	0	0	<b>100</b>	0	0	0	0
B	0	0	0	<b>100</b>	0	0	0
TA	0	0	0	0	<b>100</b>	0	0
TL	0	0	0	0	0	<b>100</b>	0
GE	0	0	0	0	0	0	<b>100</b>

**Table 11 - Confusion Matrix for LPC (TR=90s and TE=15s)**

ACT. IDENT.	M	H	U	B	TA	TL	GE
M	<b>100</b>	0	0	0	0	0	0
H	25	<b>75</b>	0	0	0	0	0
U	25	0	<b>75</b>	0	0	0	0
B	0	0	0	<b>100</b>	0	0	0
TA	0	0	0	0	<b>100</b>	0	0
TL	0	0	0	0	0	<b>100</b>	0
GE	0	0	0	0	0	0	<b>100</b>

**Table 12 - Confusion Matrix for LPCC (TR=90s and TE=15s)**

ACT. IDENT.	M	H	U	B	TA	TL	GE
M	<b>100</b>	0	0	0	0	0	0
H	0	<b>100</b>	0	0	0	0	0
U	0	0	<b>100</b>	0	0	0	0
B	0	0	0	<b>100</b>	0	0	0
TA	0	0	0	0	<b>100</b>	0	0
TL	0	0	0	0	0	<b>100</b>	0
GE	0	0	0	0	0	0	<b>100</b>

**Table 13 - Confusion Matrix for MFCC (TR=90s and TE=15s)**

The performance improves with the testing duration and the best result is obtained with MFCC for 2nd order classifier. The Confusion Matrix is showing diagonal elements as 100% (see Table 11 and Table 13), meaning that all the speakers in a particular linguistic group are identified correctly or misidentified as a different speaker in that respective group only. With LPCC the performance as shown by the confusion matrix is a little poorer. From these test results, we can conclude the following:

- Success rates are more sensitive to testing speech durations as compared to training speech durations.
- Average success rates (over testing speech durations) for LPC, LPCC and MFCC performs equally well. But the individual success rates for LPC and MFCC are better than LPCC in majority of the cases of the training and testing speech durations.
- Confusion matrix (in which diagonal elements indicates % correct identification in a particular linguistic group and off-diagonal elements show the misidentification) for MFCC and LPC is having all off-diagonal elements as zeros meaning that the entire speakers in particular linguistic group are identified or misidentified in their respective language only. Confusion matrix for LPCC is slightly worse than LPC and MFCC (see Table 11 to Table 13).

## 5. Conclusions and Future Work

We can derive two main conclusions from our technical meta data evaluation: First, we have shown the importance of collection of technical meta data and the association of those with biometric data. This is due to our observations that the recognition accuracy of biometrics is strongly correlated with sensor hardware characteristics. It has been experimentally shown for one specific signature verification algorithm and on a reasonably large data set. Secondly, the test results have motivated us to make use of tablet PCs as digitizer devices for our future project activities, as this specific digitizer technology has shown a good accuracy in our evaluations and further allows for visual feedback (digital ink) during the writing process on a mobile platform. Test results of tablet A, having quite similar properties as recent tablet PCs, justify this decision.

With respect to the first results of our non-technical evaluations, there are also two main conclusions to be drawn. First, it appears that the recognition accuracy is dependent on cultural aspects of users, such as familiarity to the spoken language in speaker recognition. We have observed a slight degradation in recognition accuracy, when trying to identify test subjects speaking languages other than their native ones. Secondly, we were able to show that language detection appears feasible. In two of our experiments, all languages were determined correctly. This may have importance for future large-scale applications, as language identification may be helpful for efficient binning strategies.

Although our first results are very promising, we are at an early state in our evaluation task. To reach statistically more significant evaluation results we need to extend our test databases, which will be supported by a framework system for data collection, which has been finalized recently. Using this tool, we will start additional systematic collection of biometric data along with meta data in two European countries and India soon. By analyzing these additional data, we will be able to perform additional evaluations of cross-cultural aspects in biometrics, which will include the correlation of type of written script and signature verification, as well as sensor dependence of speaker recognition algorithms.

## REFERENCES

- [HYHI2004] J.H.L. Hansen, U.Yapanel, R. Huang, A. Ikeno, "Dialect Analysis and Modeling for Automatic Classification," Interspeech-2004/ICSLP-2004: Inter. Conf. Spoken Language Processing, pp. WeC2302p.5(1-4), Jeju Island, South Korea, Oct. 2004
- [MaCl1996] R. Martens, L. Claesen, "On-Line Signature Verification by Dynamic Time Warping", In: Proceedings of the 13th IEEE International Conference on Pattern Recognition (ICPR), Vienna, Austria, Vol. 1, pp. 38 - 42, 1996
- [Mic2002] Microsoft, "Hardware Guidelines for Speech Technologies", <http://www.microsoft.com/whdc/device/audio/speech/default.aspx#ECAA>, June 12, 2002
- [PaBa2004] H.A. Patil and T.K. Basu, "Speech corpora for speaker classification experiments in Indian languages" Proc. of Int. Conf. on Emerging Technology, ICET'04, KIIT, Orissa, India, Allied Publishers, Dec. 22-24,2004
- [PaDB2004] H.A. Patil, P.K.Dutta and T.K.Basu, "The Teager Energy Mel Cepstrum for Speaker Identification in Multilingual Environment" Journal of Acoustical Society of India, JASI, 2004
- [PrMi2001] Premakanthan P. and Mikhad W. B. Speaker Verification/Recognition and the Importance of Selective Feature Extraction: Review. MWSCAS. Vol 1, 57-61, 2001
- [Schi+2004] S. Schimke, C. Vielhauer, P. K. Dutta, T. K. Basu, A. De Rosa, J. Hansen, J. Dittmann, B. Yegnanarayana, "Cross Cultural Aspects of Biometrics?", In: C. Vielhauer, S. Lucey, J. Dittmann, T. Chen: Workshop Proceedings "Biometric Challenges arising from Theory to Practice", pp. 27-30, ISBN: 3-929757-3, 2004
- [ToKS2004] C. I. Tomai, D. M. Kshirsagar, and S. N. Srihari, "Group Discriminatory Power of Handwritten Characters", Proceedings of SPIE-IS&T Electronic Imaging, pp. 116-123, 2004
- [Viel2004] C. Vielhauer, Claus: Handwriting Biometrics for User Authentication: Security Advances in Context of Digitizer Characteristics, PhD Thesis, Technical University Darmstadt, 2004
- [ZhSL2003] B. Zhang and S. N. Srihari and S.-J. Lee , "Individuality of Handwritten Characters", 7th IEEE International Conference on Document Analysis and Recognition, Edinburgh, Scotland, August, 2003