Contents

Digital Speech Watermarking and its Impact to Biometric Speech Authentication

ANDREA OERMANN AND ANDREAS LANG AND CLAUS VIELHAUER

Abstract

In this article an approach for connecting biometric speech authentication and digital watermarking is presented in order to integrate metadata into the authentication process without significant quality and performance losses. Different digital audio watermark methods are used to embed metadata as additional information into the reference data of biometric speaker recognition. Metadata in our context may consist ancillary information about the social, cultural or biological context of the owner of the biometric data as well as technical details of the sensor. We perform our tests based on a database taken from 33 subjects and 5 different utterances and a known cepstrum based speaker recognition algorithm in verification mode. The goal is to perform an evaluation of the recognition precision for our selected algorithm in the context of the gender belongings of the persons. The first tests show that the recognition precision is not significantly deteriorated by the embedding of the information. Further, the losses of the performance of the used biometric authentication system are less for female than for male users.

1. Motivation

Biometric Authentication Systems as well as Digital Watermarking Methods have been developed to fulfill the challenges of IT-Security such as the authenticity and integrity. In an digitally interconnected world where communication is independent from time, localization and culture, the acceptance and success of such a system remain dependent on trustworthy identities.

In order to get access to certain resources, equipment or facilities, users need to be identified or verified. This can be realized through three authentication methods: secret knowledge, personal possession and individual characteristics of a human being (biometrics) [1]. The secret knowledge approach indicates the users knowledge, such as passwords or PINs. Personal possession implies that the user owns something like a physical key, smartcard or special token. Biometrics as an authentication method refers to the user's individual biometrical attributes such as speech and handwriting as behavioral-based modalities and fingerprint, face, iris, retina, or hand geometry as physiological modalities. Hence, instead of identify a person by external information, which can be lost, stolen or handed over, a biometric system identifies a person itself based on its given characteristics. The advantage of biometric user authentication is the unique and reliable identification and verification of a human being's identity. Hence, biometrics improve the level of security in infrastructures and applications.

The goal of biometric user authentication is the determination of similarities based on features derived from sampled signals concerning a particular biometric characteristic. We confine our study to the behavioral-based modality as we base our work on previous evaluations such as [2], [3], and [4], where we found out that the integration of metadata into the authentication process can improve the biometric authentication system. Previous work has shown that for example group discriminatory information such as gender or ethnicity can be derived [5], and also a specific language of a spoken sequence can be identified by biometric features [4]. As [4] and [6] have shown, a local optimization of the authentication process can be achieved by integrating metadata into it.

The focus of our work is to use digital watermarking techniques in order to fulfil biometric challenges as this provides a way to directly connect metadata with the biometric data. In other words, watermarking techniques provide a way to make metadata available for the biometric system. We define metadata as a collection of information the basic audio signal does not provide such as additional biometric information or additional characteristics (e.g. language, culture, ethnicity, gender, condition, age, ...) of the individual or the technical environment (e.g. device).

Based on this, we introduced a basic approach [7] where metadata are embedded by a digital watermarking method into the speaker reference signal in order to measure its impact on the EER of the biometric speaker verification system. In this approach 16 bit quantized speech signals and one LSB watermarking scheme have been applied to demonstrate first

results. Further, we have been evaluating different watermark schemes by the application of so called profiles such as biometrics as presented in [8] while in this article we present an evaluation in a gender context. We want to find out if there are differences of quality losses of the performance of a biometric user authentication process regarding the user's different belongings to certain gender groups (male and female). Further, we want to analyze the influence of the length and content (semantics) of the audio signals of user's speech samples and also the influence of audio signal quantization. Here we compare if 16 bit quantized audio signals lead to a decreased distortion of the audio signal caused by the watermarking process and following to an improved performance than 8 bit quantized audio signals.

Digital watermarking has been proposed for a variety of applications, including content protection, authentication, digital rights management and others. Many watermarking techniques have made claims regarding performance, such as transparency, robustness, or capacity. In general, watermarking is an embedding and retrieval process, where hidden or secret information is embedded into or retrieved from digital content like music, image or video [9]. Using digital watermark techniques to embed information in biometric data is an emerging area of research and only a few approaches could be found in the literature such as [10], [11], [12], and [13]. However to date, the impact of watermarks on biometric speech authentication systems has rarely be evaluated. Therefore, and because of the fact that the watermarking procedure always implies changes of the content of information, the subject of our research is to analyze the impact of these changes on the authentication performance of the whole biometric system.

In this article, the same biometric speaker verification system as in [7] is used for feature extraction, which is based on Mel-Frequency Cepstrum Coefficients [6]. Our methodology is as follows. Firstly, the metadata are embedded into the reference signal of the biometric system. Then, the user verification process of the biometric system measures the error rates: false match rate (FMR), false non match rate (FNMR) and the derived equal error rate (EER). This is explained into more detail later in section 2. Different from [7], we now use four selected watermarking schemes working in time, frequency and wavelet domain regarding 16 as well as 8 bit quantized speech signals. Based in this, the evaluation will consider the different watermarking schemes in context with the gender aspect as well as the varying semantics.

This article is structured as follows. In section 2, biometric authentication systems are firstly introduced which includes an explanation of error rates, the speech authentication process, metadata, the cultural context and semantic classes considering the capacity needed to capture the metadata digital watermarks. This is followed by a description of the four used digital watermarking algorithms and the match of biometric authentication and digital watermarks. In section 3, the evaluation set up will be described. Therefor the evaluation methodology, its parameters and goals will be outlined. The evaluation results are presented

in section 4. The article closes in section 5 with a summary.

2. Biometric Authentication Systems and Digital Watermarking

In this section a brief introduction of biometric authentication systems will be provided followed by a description of error rates as quality measures of those systems. Further, a technique for speech authentication will be presented as well as a discussion of four different digital watermarking algorithms. Finally, metadata and their impact on biometric user authentication are elaborated. A description of the harmonization of biometric user authentication, digital watermarks and metadata is closing this section.

2.1. Biometric Authentication

In biometric systems, user data initially needs to be enrolled which means the biometric parameters of the desired attribute are captured and stored in a database. For this purpose, one or more reference signals are captured from every user at time of registration. In the actual process of authenticating a particular user, again one ore more samples are taken from the subject and compared to the stored reference data. In order to verify or identify a user, new data regarding the same biometric attribute is compared with the stored biometric reference data. If the instances of the biometric data match, the user gets accepted and is allowed to access. Otherwise the user gets rejected.

The authentication can follow two different modes: In one mode a particular identity is declared as known prior the authentication. In this case the biometric system either confirms or declines the declared identity. This process is called verification and implies a comparison of n signal samplings to 1 particular reference storage sampling (1:1 comparison). The other mode refers to the biometric system automatically determining the identity of the actual user, which is called identification. This identification of a particular not known user considers a comparison of 1 signal samplings to n particular reference storage sampling (1:n comparison). Depending on the desired authentication mode, the system parameters may change. Both methods are qualified to bind the biometric data to an identity, thus may be used for authentication.

2.2. Error Rates

Commonly, evaluations of biometric authentication algorithms are based on the Equal Error Rate (EER), the point where False Match Rate (FMR) and False Non-Match Rate (FNMR) are identical. FNMR is the percentage probability of rejections by a biometric system of authentic user while FMR is the percentage probability of acceptances of non-authentic user. Thus, the ERR is one decision measure value at a specific operating point of a biometric system and implies the probability of great similarities. To read more about error rates we refer to [14]. The EER is not necessarily the optimal operating point in every biometric

system and measurements such as Receiver Operating Characteristics (ROC) may provide more detailed information about the system's characteristics, but it is an initial clue for comparing recognition capability of biometric systems.

2.3. Speech Authentication

The speech authentication system used for our tests is based on Mel-Frequency Cepstrum Coefficients (MFCC), currently being one of the most popular and widely used feature extraction methods. By applying a mel-frequency scale rather than frequencies themselves MFCC represents a model of the human perception of sounds. Being nearly linear for frequencies below 1,000 Hz and logarithmic above, the mel scale initially has been proposed by S. Stevens, J. Volkman and E. Newman in 1937 [15] as a measure of the perceived pitch. Further, the cepstrum of signals as the Fourier transform or the spectrum of the log spectrum [16] is used.

In our system, all input wave files have a sampling frequency of 44,100 Hz and two different sampling precisions, 16 Bit as well as 8 Bit. Thus, the quality and performance differences of the authentication process can be evaluated. By applying a hamming window function with an overlapping shift of 10ms, the algorithm first justifies the input signal and generates frames of 30ms length. By doing so the influence of the textual content of the utterances, especially how it was spoken, can be limited. In order to reject frames with silence or low noise the total frame energy is compared against a threshold. A filter bank with L=20 mel-spaced triangle bandpass filters l, ranging up to 8,000 Hz was applied to the spectrum of every remaining frame to achieve the corresponding mel-frequency wrapped spectrum Ψ.

By modifying the approach described in [17], our implementation is applying "simple" MFCCs instead of the proposed T-MFCCs, which is based on a Teager Energy Operator. Hence, our frame's acoustic vector is calculated according to the following equation for each cepstrum coefficient k:

$$\mathrm{MFCC}_k = \sum_{l=1}^{L} \log \Psi(l) \cos\left[\frac{k(l-0.5)}{L}\pi\right],\ k = 1, 2, \ldots, L \qquad (.1)$$

Every single acoustic vector is then added to the frame's acoustic vector set. Considering the enrollment mode, the LBG algorithm [18] selects 32 reference vectors (centroids) out of the enrollment's acoustic vectors for each enrollment's reference model. Referring to the verification mode, the score of the verification vector set represents the minimum of all Euclidean distances between each verification vector and each reference vector.

2.4. Metadata, Cultural Context and Semantic Classes

Embedding metadata regarding individual user information and technical settings into biometric reference data for authentication can be much of a benefit as our previous research

[2] and [4] has shown. There we analyzed the impact biological, cultural and conditional aspects can have on a biometric handwriting and authentication system. Results encouraged us to enhance our research in this field. Based on the collected biometric data, both handwriting and speech, we can rely on a solid test set, especially when considering different cultural groups.

In our tests to determine the recognition precision the following information is embedded into the speech reference audio files:

- SampleID
- EventID
- PersonID
- SemanticID
- DeviceID
- LanguageID
- EnvironmentID

The SampleID is the ascending internal number of the speech files in the database. An event (EventID) indicates a collection of samples belonging together due to originator, semantics and action (enrollment, verification or forgery). The internal identification number of the user is stored in the PersonID. The SemanticID encodes the semantics of a speech task. It represents the content and duration of a speech sample. According to a predetermined task list different semantics have been captured from each test subject. Tasks are differentiated in individual, creative and predefined ones. The hardware device for voice recording is defined in the DeviceID. Further, Date and Time of recording is stored as metadata. The LanguageID indicates the spoken language of an utterance, while the environment of the capturing (e.g. soundproof cabin) of the speech recording is stored in an EnvironmentID.

Evaluation regarding the cultural background the PersonID is of importance, since it refers to a particular person. For our tests we have defined certain test sets based on the gender belongings of users with different cultural backgrounds such as Indians, Germans and Italians. Primarily, we focus on the 2 different classes, male and female, for our evaluation, whose particular parameters and goals are described into detail in section 4.

Within this EU-India Culture-Tech project speech and handwriting data has been captured from German, Indian and Italian test users. Further, metadata of all of these participating test users has been acquired. Therefore, we are profiting from an existing database. The collection of speech and handwriting data in our proprietary database followed a defined test plan with 47 different semantics in two languages (English and German). We developed this test set of certain semantics based on individual, creative and predefined tasks in order to be able to analyze its varying influence on the authentication system.

One single task was captured by 10 iterations, where the first 5 are used as reference data and the remaining 5 as authentication data. Audio files are recorded with a sampling frequency of 44,100 Hz and a sampling precision of 16 Bit as well as 8 Bit using a headset microphone

in a laboratory environment for a uniform data collection. For our first initial tests, the following listed five out of the set of 47 semantics in English are chosen:

- "Communication"
- "What is your good name?"
- "Where are you from?"
- "She sells sea shells on the shore."
- "Hello, how are you?"

The sentences "She sells sea shells on the shore." and "Hello, how are you?" represent predefined tasks with an average length of 3.08/2.61 (Indians/Germans) seconds (average duration) and 1.83/1.35 seconds (average duration). A predefined semantic with a short duration are the word "Communication" (1.54/1.22 seconds) and the questions "What is your good name?" and "Where are you from?". These semantics represent tasks which encourage the test persons to provide individual answers. They have a short duration at an average of 1.40/1.09 seconds and 1.33/1.10 seconds.

In our test environment we use the verification mode for authentication. During the verification a claimed user identity is confirmed by the biometric system. The person is verified if the confirmation is successful, in the other case the person is rejected from the system.

The used test set consists of 47 test users. The set is divided into subsets related to the semantics and male and female English spoken language. Those subsets consist of a varying number of test users as it can be followed in Table 1. There is no difference in the distribution of test users for 8 Bit and 16 Bit quantization levels.

	communication	good name	hello	sea	where	
male female	23 9	24 9	23 9	23 9	24 9	without watermark
male female	23 9	24 9	23 9	23 9	24 9	2A2W
male female	23 9	24 9	23 9	23 9	24 9	LSB
male female	23 9	24 9	23 9	23 9	24 9	MS
male female	23 9	24 9	23 9	23 9	24 9	SSWater

Table 1.: Distribution of test users for 8 Bit as well as 16 Bit quantization

2.5. Harmonization of Metadata, Semantic Classes and the Capacity of Digital Watermarks

A specific watermarking payload of audio files, approximately 5,500 bytes per second, is available for embedding our metadata. The metadata, as described above, we have used

in our first tests, have an average payload of 215 bytes. It will be embedded repeatedly in the speech data during the watermarking embedding process. Due to the fact that the required space for a watermark's capacity can only be fulfilled by audio signals of a certain length, semantics of speech samples for authentication need to be of a certain length in order to grant the needed capacity. This explains the decision for the five earlier introduced semantics we used for this evaluation.

2.6. Digital Watermarking Algorithms

Different digital audio watermarking algorithms can be applied, from which we have selected four for our test set. Those four selected watermarking algorithms, implemented by Otto-von-Guericke University, Germany, open source and free available tools such as the one from Microsoft, and their parameters will briefly be introduced.

LSB: This watermarking algorithm works in time domain and embeds the watermark in the least significant bits (lsb's) of the audio sample values by overwriting the original bits [19] and [20]. Depending on the parameters, the message is embedded many times (redundantly) into the audio signal. This algorithm considers the following six parameters:

- The parameter k presents a secret key. The application of k, initializes a pseudo random noise generator (PRNG) which selects the LSB's used for embedding the digital watermark. This indicates the embedder being in scrambling mode and not all LBS's are applied for embedding. If the parameter k is not set, all sample values of the audio signal are applied for embedding, which directly implies the highest embedding capacity.
- The parameter c indicates for the application of an error correction code (ECC) [19] and turns error correction on or off. If an ECC is applied, the length of the embedding message is doubled and errors occurring during the retrieval function can be detected and corrected up to a certain threshold by detecting and retrieving.
- The parameter m specifies the secret message which is embedded into the audio signal.
- The parameter t stands for the mode selection which differs depending on the handling of multiple audio channels. According to the mode, the algorithm handles the audio channels in a naïve (1), identical (2), independent (3), consecutive (4), or random (5) way [20], where the numbers represent the specific value of the parameter.
- The parameter x represents the dynamic synchronization. By applying a PRNG, a different sequence of synchronization flags is generated for each embedded watermark message in order to decrease the risk of watermark detection by a statistical analysis of the audio signal.
- The parameter j describes a number of sample values which are randomly skipped and not used for embedding, preconditioned that parameter k is set and the embedding algorithm works in scrambling mode. The difference between the sample indexes is referred to as the jump length. The default value for this parameter is 9.

Microsoft: This watermarking algorithm works in frequency domain and embeds the watermark in the frequency coefficients by using a spread spectrum technique [20]. This algo-

rithm only applies one single parameter m for the embedding message.

2A2W - AMSL Audio Water Wavelet: This watermarking algorithm works in wavelet domain and embeds the watermark on selected zero tree nodes without applying a secret key. In order to be able to retrieve the watermark information later by a detection function (non blind) the algorithms performs an additional file the marking positions are stored in. This algorithm considers the following parameters:

- The parameter m represents the embedded watermarking message
- The parameter w specifies the watermarking method and currently exclusively limited to ZT (zerotree).
- The parameter c specifies the coding method and currently, only binary (BIN) is possible.

SSWater: This watermarking algorithm works in frequency domain and embeds the watermark in selected frequency bands by using a spread spectrum technique [21]. This algorithm considers the following parameters:

- The parameter k specifies the secret key to initialize the PRNG.
- The parameter l indicates the lowest frequency bound.
- The parameter h indicates the high frequency bound.
- The parameter a defines the embed strength.
- The parameter f defines the frame size used for the FFT transformation.
- The parameter t defines the tolerance value used as threshold by retrieving the watermark.

2.7. Match of Biometric Authentication and Digital Watermarking

In order to analyze and evaluate the usage of digital watermarking algorithms for integrating metadata into speech signals a biometric user authentication is based on, the coordination of all three, digital watermarking, biometric user authentication based on speech and metadata, need to be described. In particular, this will be presented as follows through a description of our test scenario. Basically, the general authentication process consists of two successive steps: The enrollment and the verification/identification, as earlier mentioned in this article. Both steps include the process of watermarking embedding and retrieving. To capture speech data during enrollment each subject is asked to repeat a predefined semantics 10 times. Further, certain metadata from the subjects is collected. In the next step a watermarking algorithm with its default embedding parameters embeds the metadata of a subject as message into all audio signals captured from this particular subject the metadata is related to. The watermarking algorithm embeds the metadata information into the speech file repetitively until the full capacity for each audio signal is reached. The resulting watermarked files are then stored in the reference database of the biometric system. In the authentication process the embedded metadata now gets retrieved from the reference data while the sufficiency of the embedding capacity for the given audio signal is verified at the same time.

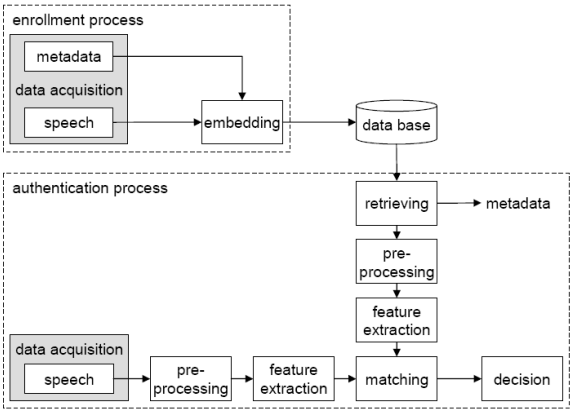

Figure 1.: Biometric user authentication (enrollment and authentication) based on watermarked reference and authentication data (see also [7])

Watermarking scheme	Embedding parameters
LSB	$k = , c = off, t = 1, x = off$
Microsoft	no parameters
2A2W	$w = ZT, c = BIN$
SSWater	k=1234, l=500, h=10000, a=2, f=1024, t=0.6

Table 2.: Embedding parameters of used watermarking schemes

The next step is the matching process for biometric user authentication and refers to the comparison of new captured authentication speech data passed though the preprocessing and feature extraction procedure and the stored and watermarked reference data, which also passed through the preprocessing and feature extraction. This matching delivers a value of similarity or dissimilarity (matching score) between reference and authentication data. Based on this score the biometric decision module will be able to make a decision upon the authenticity of the speaker. Thereby, the variable t is used as threshold for tuning the biometric system. Figure 1 demonstrates this test scenario.

For the embedding function in the enrollment process, all four earlier introduced watermarking schemes with their default embedding parameters, which are summarized and shown in Table 2, are applied for embedding. The influence of the different embedding functions on the authentication performance are measured by the occurring biometric error rates FNR and FNMR and the derived EER. Further evaluations will be explained into detail in section 3.

In general, we study the impact of the watermark embedding on the overall authentication

performance of our biometric speaker recognition system (MFCC approach), by analyzing the recognition errors for our experimental data collection. Embedding information such as metadata into audio signals is a special case of additional noises within these audio signals. At the current state of our work, we focus on aspects of quality losses in terms of biometric measurements caused by watermarking embedding schemes and the MFCC approach's noise sensitivity. But also, motivated by approaches for the MFCC method to improve the authentication performance in noisy environments ([22], [23]), we want to consider an optimization of the recognition accuracy of the authentication process as well as we want to support binning strategies in case of identification, both as impacts the direct connection of metadata and biometric data can have in the future.

3. Evaluation Description

In this section, the evaluation process, its parameters and evaluation goals will firstly be outlined followed by a discussion of its results.

3.1. Evaluation Methodology

The evaluation's methodology is based on a test division into two basic test sets: male and female. Further, our tests are always a composition of verification as the authentication mode and random forgery tests.

For each semantics and each user's belonging to either male or female the tests are divided into the following parts: Firstly, based on the previous captured audio data we embed the metadata of each subject of the test sets through all of the four watermarking schemes, LSB, Microsoft, SSWater, and 2A2W into the five different semantic speech samples captured by the same subject. Both, enrollment and verification samples are watermarked with the same metadata information.

In order to compare the impact of embedding metadata in different semantics of subjects belonging to different gender groups, either male or female, using different watermarking schemes we use the well known and earlier introduced biometric error rates. The specific test set, which is used, has earlier been introduced in Table 1.

3.2. Evaluation Parameters

Our evaluations contains of fixed as well as of variable parameters. Fixed parameters can be listed as follows:

a.) database for biometric speech authentication

b.) metadata integrated into the speech signal

c.) number of users

Variable parameters of the evaluation can be listed as follows:

a.) 2 different test sets: male and female

b.) 4 different speech watermarking algorithms: LSB, Microsoft, 2A2W, SSWater

c.) 5 different semantics: see section 2

d.) 2 quantization modes: 8 Bit and 16 Bit

3.3. Evaluation Goals

The introduced evaluation tests focus on the goal of measuring the quality of watermarking algorithms used for embedding additional information such as metadata into audio signals of a biometric user authentication system based on speech. In particular, we want to measure the influence of the embedding function of four different watermarking schemes on the biometrical user verification in the context of the user's different belongings to certain gender groups (male and female). First test results have shown that embedding additional information into the audio signals seems to always lead to performance losses of the authentication (refs). We now want to find out if there are remarkable differences of quality losses regarding the gender aspect. Further, we want to consider the length and content (semantics) of the audio signals of user's speech samples while the last evaluation goal refers to the influence of audio signal quantization. Here we want to analyze the relation of the quantization of audio signals and the performance of the biometric user authentication process. In detail, we want to find out if 16 bit quantized audio signals lead to a decreased distortion of the audio signal caused by the watermarking process and following to an improved performance than 8 bit quantized audio signals.

4. Results

In this section, our test results are presented and discussed while considering our earlier declared evaluation goals. Firstly, two tables give an overall representation of the results for all applied watermarking schemes, all semantic classes and the two user groups male and female. Then, two figures illustrate theses results followed by 2 figures concentrating on each watermarking algorithm in order to analyze them separately.

Before starting to discuss the test results performance characteristics of the applied watermarking schemes need to be outlined in order to put the results in relative terms. The LSB watermarking scheme as well as the 2A2W watermarking scheme are able to successfully embed the complete message into all audio files for our test sets while the embedding capacity of the Microsoft watermarking scheme is worse. If a part of message *m* is embedded, than only the first character of *m* (e.g. "S") is embedable. Furthermore, the used spread spectrum technique is not able to spread 100% of the first character. It means that for four audio files, only 5.83% and for one audio file 43.33% of the watermark could be spread

over the audio signal. The SSWater watermarking scheme is not able to embed the complete message into the audio files. For 16 bit audio signals, the embedding capacity is higher than for 8 bit quantized audio signals. in general, the first characters of a message *m* can be retrieved directly after embedding.

Table 3 and Table 4 summarize the results for our tests for each of the five semantic classes (left column), where 1 = "Communication", 2 = "What is your good name?", 3 = "Hello, how are you?", 4 = "She sells sea shells on the shore." and 5 = "Where are you from?" and ∅ stands for the average mean of all semantic classes. Test results are represented by the EER for each gender group (male and female) for each applied watermarking scheme (2A2W, LSB, Microsoft and SSWater) and without any watermarking scheme.

As presented in Table 3 and Table 4, the biometric system has an average EER of 0.280 (female) and 0.310 (male) for 8 bit quantized audio signals and 0.280 (female) and 0.289 (male) for 16 bit quantized audio signals without using any watermarking scheme which embeds a message into the reference data. This indicates, that a biometric authentication system works slightly better for female than for male users. Further, considering the average EER for the different watermarking schemes itself and compared to the average EER for not marked biometric data it can be outlined, that the used biometric authentication system continued to work better for female users, no matter which algorithm and which quantization level has been applied to embed the metadata.

In particular, the best performance on an 8 bit quantization level could be achieved for female users by the Microsoft watermarking scheme (0.266) while the lowest EER in the 16 bit quantization level could also be achieved for female users, but by the SSWater watermarking scheme (0.255). This shows, that having a higher quantization level (16 bit) decreases the EER and hence increases the performance of the biometric user authentication. Even though these seem to be the best results, it has to be seen relatively due to the earlier described characteristics of the watermarking schemes, especially the inability of the Microsoft and the SSWater scheme to embed a whole message (metadata) into the audio signals. A first approach to evaluate this is presented in [8].

The results presented in Table 3 and 4 further indicate a remarkable gab between the EER regarding semantic classes. The best possible performance on an 8 bit quantization level could be reached for female users by the Microsoft watermarking scheme and the semantic class "Hello, how are you?" (0.187) while the worst result has been performed for male users by the 2A2W watermarking scheme and the semantic class "She sells sea shells on the shore." (0.409). The best possible performance on an 16 bit quantization level could be reached for female users by the SSWater watermarking scheme and also the semantic class "Hello, how are you?" (0.174) while the worst result on an 16 bit quantization level has been performed for female users by the 2A2W watermarking scheme and the semantic class "She sells sea shells on the shore." (0.391).

The presented figures are underlining our test results. The first two figures (Figure 2 and Figure 3) illustrate the comparison of all watermarking algorithms in the context of the two different gender groups and the five semantic classes. Figures 4-13 represent the specific differences of the EER for each watermarking scheme.

	8 Bit	8 Bit	8 Bit	8 Bit	8 Bit	8 Bit	8 Bit	8 Bit	8 Bit	8 Bit
	-	-	2A2W	2A2W	LSB	LSB	MS	MS	SSW	SSW
	female	male	female	male	female	male	female	male	female	male
	EER	EER	EER	EER	EER	EER	EER	EER	EER	EER
1	0,273	0,304	0,304	0,351	0,351	0,253	0,249	0,310	0,293	0,271
2	0,267	0,301	0,333	0,291	0,310	0,311	0,258	0,294	0,291	0,315
3	0,204	0,339	0,297	0,395	0,211	0,381	0,187	0,304	0,198	0,320
4	0,318	0,366	0,311	0,409	0,276	0,385	0,316	0,371	0,302	0,360
5	0,336	0,240	0,373	0,291	0,329	0,256	0,320	0,215	0,286	0,239
∅	**0,280**	**0,310**	**0,324**	**0,347**	**0,295**	**0,317**	**0,266**	**0,299**	**0,274**	**0,301**

Table 3.: Evaluation results based on EER and with a quantization of 8 Bit

	16 Bit	16 Bit	16 Bit	16 Bit	16 Bit	16 Bit	16 Bit	16 Bit	16 Bit	16 Bit
	-	-	2A2W	2A2W	LSB	LSB	MS	MS	SSW	SSW
	female	male	female	male	female	male	female	male	female	male
	EER	EER	EER	EER	EER	EER	EER	EER	EER	EER
1	0,262	0,262	0,276	0,335	0,262	0,263	0,236	0,253	0,249	0,229
2	0,265	0,298	0,324	0,298	0,265	0,298	0,258	0,294	0,298	0,310
3	0,204	0,294	0,283	0,353	0,205	0,294	0,187	0,304	0,174	0,275
4	0,325	0,366	0,391	0,384	0,332	0,366	0,316	0,371	0,262	0,359
5	0,343	0,227	0,356	0,311	0,345	0,228	0,320	0,215	0,291	0,220
∅	**0,280**	**0,289**	**0,326**	**0,336**	**0,282**	**0,289**	**0,263**	**0,288**	**0,255**	**0,278**

Table 4.: Evaluation results based on EER and with a quantization of 16 Bit

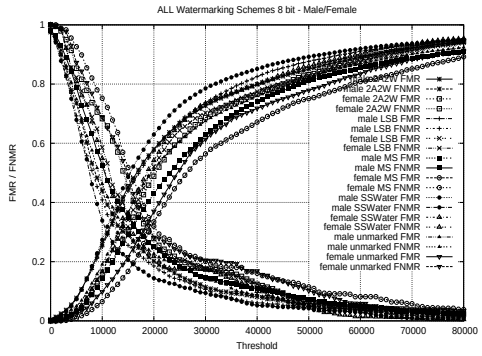

Figure 2.: All 8 Bit

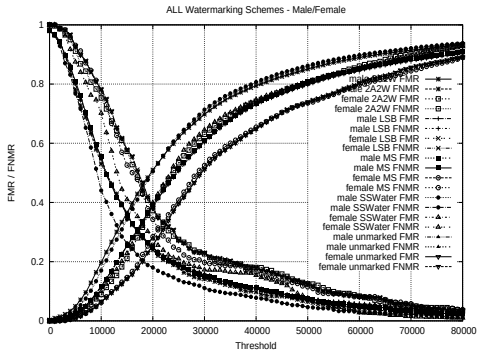

Figure 3.: All 16 Bit

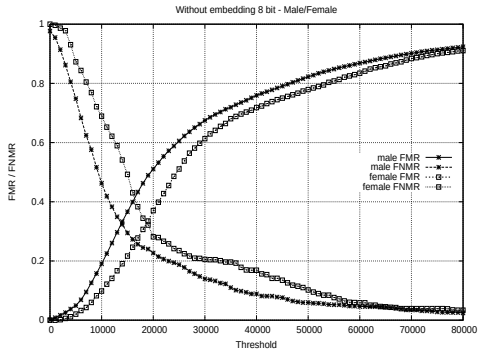

Figure 4.: FMR and FNMR curves for unmarked audio signals and 8 Bit quantization

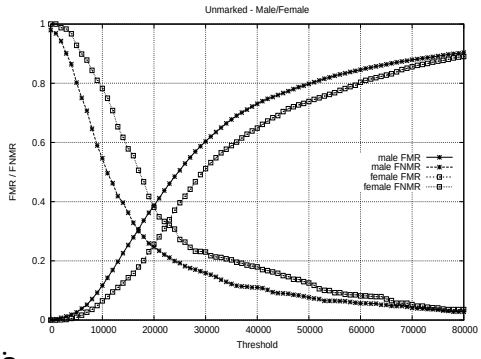

Figure 5.: FMR and FNMR curves for unmarked audio signals and 16 Bit quantization

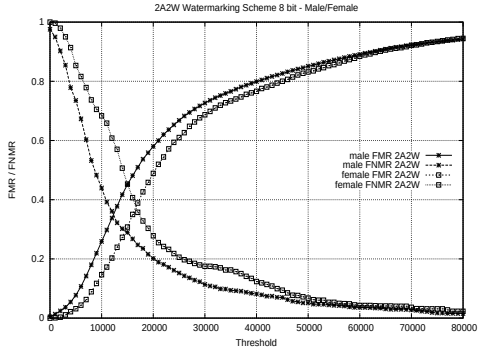

Figure 6.: FMR and FNMR curves for 2A2W watermarking scheme and 8 Bit quantization

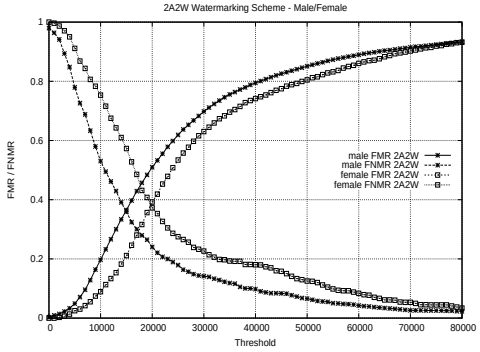

Figure 7.: FMR and FNMR curves for 2A2W watermarking scheme and 16 Bit quantization

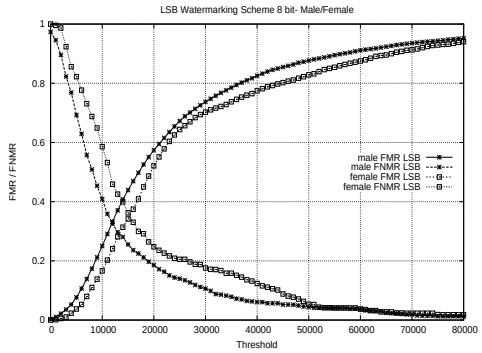

Figure 8.: FMR and FNMR curves for LSB watermarking scheme and 8 Bit quantization

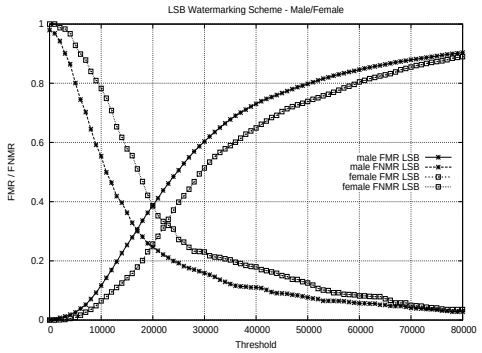

Figure 9.: FMR and FNMR curves for LSB watermarking scheme and 16 Bit quantization

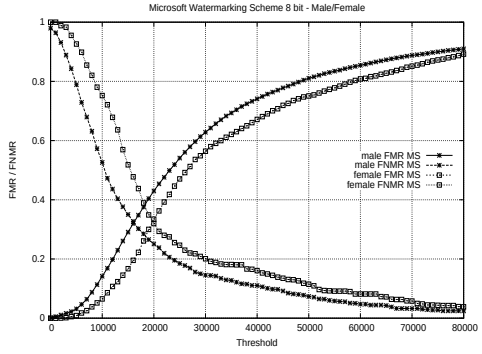

Figure 10.: FMR and FNMR curves for Microsoft watermarking scheme and 8 Bit quantization

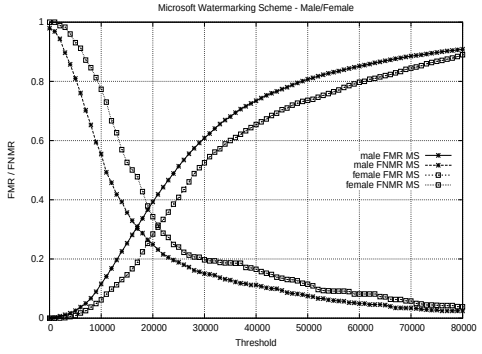

Figure 11.: FMR and FNMR curves for Microsoft watermarking scheme and 16 Bit quantization

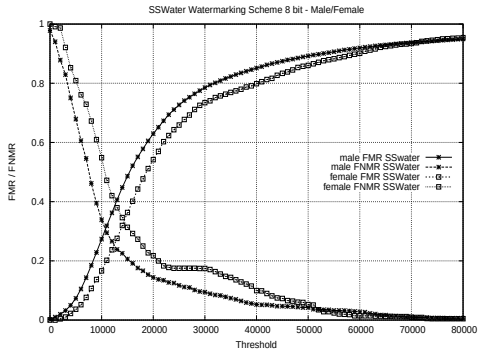

Figure 12.: FMR and FNMR curves for SSWater watermarking scheme and 8 Bit quantization

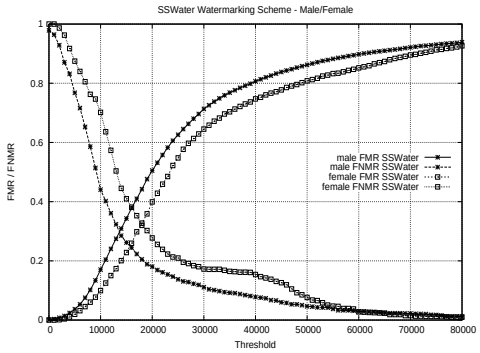

Figure 13.: FMR and FNMR curves for SSWater watermarking scheme and 16 Bit quantization

5. Summary

The test results have shown that for the selected MFCC approach as biometric user authentication combined with different watermarking schemes such as 2A2W, LSB, Microsoft and SSWater metadata can be embedded in the speech based biometric reference data without remarkable losses of the authentication performance. We have shown that the differences between non-watermarked data, watermarked data with varying capacity are marginal.

Even though knowing that the used data are not sufficient in order to achieve statistic significance, the approach of integrating metadata into biometric reference material may be applied for implementing future biometric authentication systems where metadata contains complementary biometric references. The work presented in this article is an initial investigation, which examines the influence of embedded data on biometric recognition performance. In our current research, the payload of embedded metadata information is not fully exploited, we used in average 215 bytes repeatedly. Therefore it is possible to hide further biometric information such as other modalities as payload into the metadata (see [24]). Hence, a multimodal authentication can be achieved. Also, a knowledge based hash (i.e. password hash) as metadata can be embedded in order to use it in a multi-factor authentication mode. Multi-factor means a combination of biometric based (e.g. handwriting) and non-biometric based (e.g. knowledge, possession) user authentication. In this case the input knowledge can be confirmed by the knowledge retrieved from biometric reference data in addition to the biometric authentication.

Further, our tests have also shown, that applying digital watermarking techniques to embed additional information in biometric reference material and hence, integrate it into the authentication process performs better results for female users than for male. Also, the performance depends on the applied watermarking scheme. In this context our evaluation need to be further developed due to the disadvantageously embedding behavior of certain watermarking schemes, in particular Microsoft and SSWater as they are not able to embed

the whole metadata. Therefore, we need to investigate the potential application fields and the required metadata to determine the watermarking parameters capacity, robustness and transparency as well as the required recognition precision.

6. Acknowledgements

With respect to the cultural aspect and the combination of speech authentication and watermarking as well as its experimental evaluations, this publication has been produced partly with the assistance of the EU-India cross cultural program (project CultureTech, see [26]). The work about the implementation and usage of digital audio watermarking algorithms described in this article has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The work on developing test methodologies has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507634 BIOSECURE. Our special thanks belong to Tobias Scheidat for many inspirative discussions. The information in this document is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. 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.

Bibliography

[1] C. Vielhauer: Biometric User Authentication for IT Security, Advances in Information Security. Vol. 18, Springer, New York ISBN: 0-387-26194-X, 2005

[2] F. Wolf, T.K. Basu, P.K. Dutta, C. Vielhauer, A. Oermann, B. Yegnanarayana: A Cross-Cultural Evaluation Framework for Behavioral Biometric User Authentication.: From Data and Information Analysis to Knowledge Engineering. In: Proceedings of 29 Annual Conference of the Gesellschaft für Klassifikation e. V., GfKl 2005, University of Magdeburg, Germany. Springer-Verlag, 2006, pp. 654-661

[3] C. Vielhauer, T. Scheidat: Multimodal Biometrics for Voice and Handwriting. In: Jana Dittmann, Stefan Katzenbeisser, Andreas Uhl (Eds.), Communications and Multimedia Se-curity: 9th IFIP TC-6 TC-11International Conference, CMS 2005, Proceedings, LNCS 3677, Salzburg, Austria, September 19 - 21, 2005, pp. 191 - 199

[4] C. Vielhauer, T. Basu, J. Dittmann, P.K. Dutta: Finding Meta Data in Speech and Handwriting Biometrics. In Proceedings of SPIE-IS&T. 5681, 2005, pp. 504-515

[5] C. I. Tomai, D. M. Kshirsagar, S. N. Srihari: Group Discriminatory Power of Handwritten Characters. In: Proceedings of SPIE-IS&T Electronic Imaging, 2004, pp. 116-123

[6] T. Scheidat, F. Wolf, C. Vielhauer: Analyzing Handwriting Biometrics in Metadata Context. To appear in: SPIE Proceedings - Electronic Imaging, Security and Watermarking of Multimedia Contents VIII, 2006

[7] C. Vielhauer, T. Scheidat, A. Lang, M. Schott, J. Dittmann, T.K. Basu, P.K. Dutta: Multimodal Speaker Authentication - Evaluation of Recognition Performance of Watermarked References. In: Proceedings of MMUA 2006, Toulouse, France, 2006

[8] A. Lang, J. Dittmann: Digital Watermarking of Biometric Speech References: Impact to the EER System Performance. To appear in: Proceedings of IS&T/SPIE Symposium on Electronic Imaging 2007, San Jose, 2007

[9] J. Dittmann, Digitale Wasserzeichen, Xpert.press, Springer Berlin, ISBN 3-540-66661-3, 2000

[10] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, B.V.K. Vijaya Kumar: Biometric Encryption. R.K. Nichols (Ed.), ICSA Guide to Cryptography, McGraw-Hill, 1999

[11] A.K. Jain, U. Uludag: Hiding fingerprint minutiae in images. Proc. Automatic Identification Advanced Technologies (AutoID), New York, March 14-15, 2002, pp. 97-102.

[12] A.K. Jain, U. Uludag, R.L. Hsu: Hiding a face in a fingerprint image. Proc. International Conference on Pattern Recognition (ICPR), Canada, August 11-15, 2002

[13] A.M. Namboodiri, A.K. Jain: Multimedia Document Authentication using On-line Signatures as Watermarks. Security, Steganography and Watermarking of Multimedia Contents VI, San Jose California, June 22, 2004, pp. 653-662

[14] C. Vielhauer: Biometric User Authentication for IT Security: From Fundamentals to Handwriting. Springer, New York, ISBN: 0-387-26194-X, 2005

[15] S.S. Stevens, J. Volkman, E.B. Newman: A scale for the measurement of the psychological magnitude of pitch. Journal of the Acoustical Society of America, 8, 1937, pp. 185-190

[16] J. W. Tukey, B. P. Bogert, J. R. Healy: The quefrency alanysis of time series for echoes: cepstrum, pseudo-autovariance, cross-cepstrum and saphe cracking. In: Proceedings of the Symposium on Time Series Analysis, 1963, pp. 209-243

[17] H. A. Patil, P. K. Dutta, T. K. Basu: The Teager Energy Based Mel Cepstrum for Speaker Identification in Multilingual Environment. In: Journal of Acoustical Society of India, Nov. 2004

[18] Y. Linde, A. Buzo, R. Gray: An algorithm for vector quantizer design. IEEE Transactions on Communications, Vol. 28, 1980, pp.84-95

[19] H. Klimant, R. Piotraschke, D. Schönfeld: Informations- und Kodierungstheorie. TEUBER, 2. Eddition, ISBN 3-5192-3003-8, 2003

[20] K. Matev: Least Significant Bit Watermarking. Internal report, 2005

[21] S. Dzhantimirov: Spread-Spektrum Verfahren für Wasserzeichen-Markierung von Audiosignalen. Internal report, Otto-von-Guericke University Magdeburg, 2006

[22] Z. Wu, Z. Cao: Improved MFCC-Based Feature for Robust Speaker Identification. Tsinghua Science & Technology, Volume 10, Issue 2, April 2005, pp. 158-161

[23] Q. Zhu, A. Alwan: Non-linear feature extraction for robust speech recognition in stationary and non-stationary noise. Computer Speech & Language, Volume 17, Issue 4, October 2003, pp. 381-402

[24] S. Schimke, T. Vogel, C. Vielhauer, J. Dittmann: Integration and Fusion Aspects of Speech and Handwriting Media. In: Proceedings of the Ninth International Conference Speech and Computer, SPECOM'2004, ISBN 5-7452-0110-x, 2004, pp. 42-46

[25] Andreas Lang, Jana Dittman: Profiles for Evaluation - the Usage of Audio WET. In: Proceesdings of SPIE at Security, Steganography, and Watermarking of Multimedia Contents VIII, IS&T/SPIE Symposium on Electronic Imaging, 15-19th January, 2006, San Jose, USA, 2006

[26] The Culture Tech Project, Cultural Dimensions in digital Multimedia Security Technology, a project funded under the EU-India Economic Cross Cultural Program, http://amsl-smb.cs.uni-magdeburg.de/culturetech/, last requested September 2005