Model for Audio Quality and Security Assurance in Mobile Phones as Multimodal User Guides

Andrea Oermann, Sandra Gebbensleben, Claus Vielhauer, Jana Dittmann
Otto-von-Guericke-University of Magdeburg
{andrea.oermann; sandra.gebbensleben; claus.vielhauer; jana.dittmann}@iti.cs.uni-magdeburg.de

Abstract

This paper presents a new concept of an approach to evaluate the assurance of the audio quality and its impact on the security of mobile phones used as multimodal user guides at the same time. Differently caused interferences can influence the subjectively perceived audio signal. Finding objective measure values acts as the motivation for the development of the presented model. This model to evaluate audio codecs in relation to three defined and interdependent challenges, addressing three different kinds of problems – technical, environmental, and security, will be introduced while its evaluation is in progress. As a result, our model, which is still work in process, will help to provide recommendations to future developments and to establish standardizations in the field of multimodal user guides.

1. Motivation

In today's museums and exhibitions the presented information is increasingly supported by electronic user guides. These user guides are small mobile devices which are multimodal in the sense that different types of communication channels are available to convey and present certain information. Those channels have to be either manually (by the visitor) or automatically (signal via wireless technologies) activated in order to get accessed by the visitor. On this basis, mobile phones are preferably applied these days not only due to their characteristic of being flexible but also because of other liberties that come with them, for example: most people own a mobile phone and are familiar with the device, the mobile phone can be carried anywhere, museums don't need to purchase and supply devices and localization sensors don't have to be installed as it is prohibited in cultural heritage buildings like the "Meisterhäuser Dessau" [1] as UNESCO cultural heritage buildings. A summarize is given in [2].

Contrarily to the above-mentioned advantages of mobile phones as user guides, problems occurred when testing our first implementation of a multimodal user guide for the "Meisterhäuser Dessau" [1], what restrict their usage. These problems mainly address the audio channel while being of different but interdependent characteristics: technical, environmental and security.

As a major technical problem playing audio comments causes difficulties due to the non-compliance of the Multi Media Application Programming Interface (MMAPI). In particular, audio codecs are differently supported by mobile phones, but the content has to be provided for any kind of mobile phone. Additionally, the environment is influencing the perceived quality of an audio signal. Especially the condition and location of the respective museum or exhibition have to be outlined as environmental problems. For example a noisy road next to the area or many visitors in a small room at the same time, what also causes noise, can disturb the audio quality by overlaying the audio signal. The technical restrictions/inconveniences of limited capacity, storage, and small interfaces cause or impact the security problems further. The consequentially required compression of information threatens the integrity and authenticity as well as the originality of the presented information. Thus, digital watermarks for example can hardly be embedded without being recognized.

To summarize the problems, differently caused interferences can disturb the subjectively perceived quality of the audio signal. Hence, finding measure values in order to being able of objectively evaluating the quality of an audio signal is one of the essential challenges when using a mobile phone as a multimodal user guide.

This paper presents a first concept of an approach trying to fulfill these challenges. Therefore, as presented into detail in section 4, we introduce three major hypotheses in form of questions whose evaluation is in progress:

a. Up to what extent audio codecs are dependent or independent of the type of mobile device?
b. Up to what extent noticeable quality differences regarding varying environments exist?
c. Up to what extent quality differences achieved by lossy compressions influence the security aspects integrity, authenticity and originality?

In order to evaluate these hypotheses in section 5 specific constraints, as further explained in sections 2 and 3, have to be outlined. Analogous to the presented problems these constraints are threefold: technical, environmental and security.

The following ***technical constraints*** are essential regarding mobile phones as user guides:

- Different devices / different hardware
- Low capacity, limited storage
- Small display
- Different interfaces
- Different audio codecs

The ***environmental constraints*** imply the characteristics of the exhibition:

- Varying background noise and sound conditions
- Restrictions for modifying buildings (cultural heritage)

The ***security constraints*** are rooted in the above-mentioned constraints and refer to the integrity, authenticity and originality of the provided information:

- Lossy compression of information while assuring the quality and security
- Individual content: Different groups or single visitors need to have access to different contents at a different time.

In this paper we introduce a model for evaluating audio codecs in the context of the three described constraints. Especially assuring the originality, integrity and authenticity through e.g. digital watermarks while maintaining a reasonable quality, often collides with the application of an audio codec. Hence, the main subject of our research in progress is to estimate the functionality of audio codecs for mobile phones as multimodal user guides regarding their quality and security at the same time.

The general goal of this paper is to support the choice of the used audio codec and the needed compression ratio depending on the specified application and context. Thus, recommendations can be given to developers of future user guides in order to provide the user with the best possible sound quality.

2. Audio Codecs and Compression Relating Mobile Phones – Technical and Environmental Constrains

This section initially discusses the application of mobile phones as multimodal user guides followed by the demonstration of the used audio codecs and compressions in order to specify the technical and environmental constraints.

Most commonly applied user guide devices are those with number pads, personal Digital Assistants (PDA), special devices for particular applications as well as mobile phones. Examples for number pad devices are the Mediaexplorer (PRO Cept GmbH) [3], the Museum Exhibit Guide [4], and the X-Plorer (Antenna Audio) [5] while the eTour, mobilTour, or BISSY (eloqu – metabasis GmbH) [6] and the coolMuseum (cool IT GmbH) [7] can be exemplarily listed as PDA based guides. Special devices designed for particular applications are for example the Sennheiser Guide Port [8] and the dataton PickUp version 1.2 [9]. Mobile phones as user guides are presented in BeyondGuide [10], Spatial Adventures [11] and Touch Graphics [12], to name an exemplary choice. More details about the functionality, advantages or disadvantages of these user guides can be found in [2].

Considering the technical and environmental constraints introduced earlier, a major problem is the non-compliance of the Multi Media Application Programming Interface (MMAPI). Even though mobile phones are equipped with integrated audio players, playing audio comments using our developed application is often impossible due to mobile phones differently support audio codecs. First tests with our implementation have shown that some codecs are better supported than others. Especially the audio codecs for speech transmission, defined in the GSM (Global System for Mobile Communication) [13] protocol, seems to be promising due to being commonly integrated in most of today's mobile phones. Other codecs such as AMR [14], Ogg Vorbis [15], WAV [16], WMA [17], or MPEG [18] with specific compression rates are supplementary supported. Considering one of the security constraints described in the next section, that content has to be provided individually, MPEG-21 as an additional codec seems to be promising as it includes descriptions in order to control multimedia conversion capabilities as well as permissions and conditions for multimedia conversions. Thus, any kind of content independent from the device, time and location can be accessed, as it can be read in [19].

In consequence of the technical constraints such as low capacity and limited storage the compression of audio comments is essential in order be played on mobile phones. Therefore, a detailed evaluation of the functionality and quality of audio codecs in relation to the Multi Media Application Programming Interface (MMAPI) is required, for which we are presenting the concept later in this paper.

Lossy compression techniques eliminate certain parts of the digital data stream depending on the priority objective. For example, the amount of compression for music without recognising a noticeable quality loss differs from the compression ratio applied for recording speech, which can be a lot less. The field of application and the targeted goal also impact the type and amount of compression. While in the field of forensics it is necessary to keep all background noises, focussing on a good quality of speech is sufficient for our application of developing user guides. In our particular application, the compression has to manage varying background noises as environmental constraints, while maintaining a good quality and being restricted through the technical constraints at the same time. This implies the following dilemma: The lower the compression rate, the smaller the audio files. But the smaller the size of the resulting file, the less is the quality of the produced audio comment. Consequently, if the compression rate is to low, the quality of the audio comment can be too bad to be understood on some mobile phones or in noisy environments. However, the visitors should be provided with audio comments in an acceptable quality while assuring the integrity, authenticity and originality of the presented information. To find the appropriate parameters is the focus of our progressing work.

3. Classification Model for Feature Extraction – Security Constraints

Assuring the integrity, authenticity and also the originality always implies the exact detection and localization of information changes, in particular information losses. Therefore we developed a so-called "Verifier-Tuple" to classify information in order to cluster specific information features. In doing so we are able to structurally analyze information in detail. This "Verifier-Tuple" is derived from a general concept of the explanation of programming languages as it is presented in [20]. It describes a combination of syntax and semantics, as introduced in [21] and further applied and demonstrated in [22]. To read more about the basic "Verifier-Tuple" we recommend the last two mentioned references. For the abstract explanation of syntax and semantics we refer to [23], [24] and [25].

According to [26], we now additionally differentiate three interdependent levels of syntax as an extension of our basic "Verifier-Tuple". Instead of only four levels of information we now distinguish between six levels of information.

$$V = \{SY_P, SY_L, SY_C, SE_E, SE_F, SE_I\}$$

The result is a more precise information analysis, especially when considering the security constraints in relation to multimodal user guides and their restrictions given by the technical and environmental constraints.

Table 1: Audio feature classification

Syntactic Domain SY	
Physical level	– location within the storage, sectors, of mobile phones and its characteristic
Logical level	– bit stream – bits per sample – formats and audio models: saved digital audio signal S_{ds} – PCM, MPEG…
Conceptual level	– analog audio signal S_a or digital audio signal S_d – discrete samples s=(ai, ti)* – time ti* – continuous acoustic pressure ai* of a wave in db (volume and amplitude) – phase – frequency spectrum (FFT) – wavelets – sample rate – impulse signal – pitch
Semantic Domain SE	
Structural level	– data rate – signal shaping – size of audio stream, sample size – channels
Functional level	– sound, tone, vocal tone, fundamentals – speech, language, music, noise (type of noise) – sound source location and orientation (distance to sensor, microphone) – foreground and background sounds
Interpretative level	– understanding of speech, tone, signal, sound, noise, impulse – interpretation in combination with background knowledge – male or female speaker – voiced or voiceless – composer – type of music (pop, jazz, classic...) – determination of the used device – room/background characteristics

* $i=0...n, n \in N, n \geq 0$

These six levels which are divided in two main domains – syntax and semantics, are the following:

Syntactic domain SY:

1. Syntax (SY_P) – physical level (location and characteristics of storage)
2. Syntax (SY_L) – logical level (bit-streams)
3. Syntax (SY_C) – conceptual level (information)

Semantic domain SE:

1. Semantics (SE_E) – executive, structural level
2. Semantics (SE_F) – functional level
3. Semantics (SE_I) – interpretative level

The specific classification of audio features following the "Verifier-Tuple" is presented in table 1. The classification is needed to properly evaluate the audio codecs in relation to the clarified constrains. Only by this means, efficient and reliable results can be achieved on which our recommendations can rely and yet developers can trust.

4. Concept of Evaluation Model and Hypotheses

Based on the three constraints introduced earlier in this paper, we now present the detailed concept of our evaluation model. In order to find measure values to objectively evaluate the subjectively perceived quality of an audio signal and further to assure the security (integrity, authenticity and originality) by measuring information losses we start with the three hypotheses, which have been briefly introduced in our motivation. Those hypotheses are as follows:

a. Up to what extent audio codecs are dependent or independent of the type of mobile device?
b. Up to what extent noticeable quality differences regarding varying environments exist?
c. Up to what extent quality differences achieved by lossy compressions influence the security aspects integrity, authenticity and originality?

These hypotheses act as the basis for our tests. Hypothesis a. implies the technical constraints, hypothesis b. the environmental constraints while hypothesis c. considers the security constraints. In order to validate these hypotheses different audio codecs on different mobile phones in varying environments are tested. The exact test setup will be presented in the next section. An evaluation, whose elements will also be presented in the next section, will be realized using the open source software tool EAQUAL (Evaluating of Audio QUALity) [21] and will lead to conclusions considering all three constraints. The tool computes the quality in the range of [0,-4] where zero means imperceptible quality changes and -4 very perceptible quality changes.

A matching table, compare to table 2, will structurally summarize the evaluation by representing the tool's results. Results will be presented as Q/S, what addresses the quality and security. The table exemplarily states that codec 1 is working on mobile phone B with a good quality Q while assuring a high level of security S. According to EAQUAL, the range for quality is bad (-4), medium (-3), good (-2), very good (-1) and excellent (0). The range for security is low (-4), medium (-3), high (-2), very high (-1) and total (0). Further, the table comparatively shows the robustness of an audio codec in relation to all mobile phones on which the codec is tested. Thus, the codecs can easily be compared. As exemplarily demonstrated in table 2, codec 1 is robust is the sense that a predominant good quality can be achieved while a high level of security can be assured on average. The table will be stocked up properly with our test results once all tests are finished. Currently tests are in progress.

Table 2: Table of Results

device / codec	mobile phone A	mobile phone B	mobile phone C
codec 1	medium/ high	good/high	good/ very high
codec 2			
codec 3			

Certain assumptions can be outlined considering the three constraints. For example, if a codec's compression ratio is high, the quality will be estimated higher in a quiet indoor environment than in a noisy outdoor one and the assurance of the security aspects is problematic. Or for example, if a codec's compression ratio is low, the security aspects can be assured, but the mobile phones capacity is utilised. Consequently, audio codecs which feature a medium compression ratio seem to be promising. Table 2 acts as a fundament for recommendations for future developments of user guides. A concluding simulation of a test will be presented in the following section.

Considering the security constraints and the "Verfier-Tuple" to measure information changes, the semantic domain can not sufficiently be analyzed by the tool EAQUAL. Therefore, an analogue validation of the results achieved by EAQUAL is an additional part of our ongoing evaluation. People are asked to evaluate the quality of the perceived audio comments, especially to point out the recognised information changes.

5. Simulation and Tests

The setup of our tests, which are part of our progressing work, will be described in this section. Hence, a simulation of our test will be given. Figure 1 is demonstrating the test setup. The test set consists of nine different audio comments in WAV. These comments are converted into different audio formats what implies the application of varying compression ratios by different audio codecs. These compressed audio comments are then transferred to several mobile phones. These mobile phones are T-Mobile SDA, Sony Ericsson P 900, Nokia 6230, Sony Ericsson Z520i, and Samsung D 600. While playing them on the mobile phone they get recorded back to the computer

a) in a sound isolated room,
b) in a noisy room and
c) outside.

The original audio comments, the compressed audio comments and the recorded audio comments are analyzed and evaluated. This evaluation will provide a basis on which decisions can rely, which parameters have to be used to get an acceptable quality while assuring the security. Evaluating the audio comments and finding the correct parameters will be achieved by applying the earlier introduced "Verifier-Tuple". Only through the "Verifier-Tuple" information changes can exactly and reliably be identified and localized. Thus, the results of the evaluation will analyze and validate our earlier stated hypotheses and assumptions in section 4.

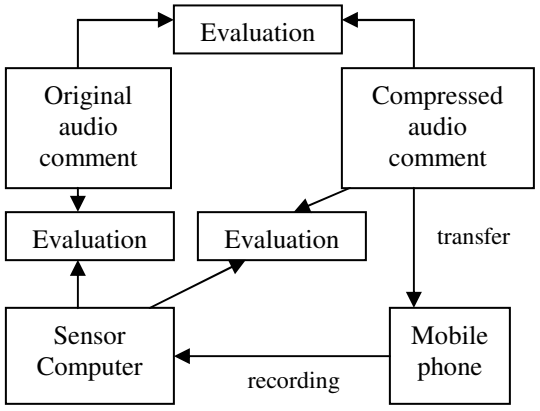

Figure 1: Test Setup

6. Conclusion

In this paper we presented the concept of a new model to evaluate the audio quality and security assurance in mobile phones as multimodal user guides. As presented, three different types of constraints, technical, environmental and security, are compromising the application of mobile phones as user guides. Audio comments have to be compressed in order to be played on a mobile phone. Different audio codecs can be applied to realize the compression but are differently supported by the mobile phones. Thus, a noticeable quality difference can be outlined while the security can not constantly be assured. Therefore, we developed this new model, whose concept is presented in this paper, to evaluate audio codecs for different mobile phones regarding the three introduced and discussed constraints. Hence, parameters will be found to objectively measure the subjectively perceived quality differences. The applied "Verifier-Tuple" provides an exact and reliable identification, localization and characterization of information changes. Thus, by applying our model it might be possible to find an approach which can be established as a standard.

7. References

[1] "Meisterhäuser/Masters' Houses in Dessau", http://www.meisterhaeuser.de/ , last requested 10.04.2006.

[2] S. Gebbensleben, J. Dittmann, and C. Vielhauer, "Multimodal Audio Guide for Museums and Exhibitions", to appear in SPIE conference, at the Multimedia on Mobile Devices II, IS&T/SPIE Symposium on Electronic Imaging, San Jose, USA, 15-19th January, 2006.

[3] "Mediaexplorer... be closer," http://www.mediaexplorer.ch/ , last requested on 10.04.2006.

[4] "excerpts from EMP: The magical geek tour", http://www.absoluterealtime.com/resume/meg.html , last requested on 10.04.2006.

[5] "Antenna Audio – the world is listening", http://www.antennaaudio.com/xp.shtml , last requested 20.12.2005.

[6] "eloqu – ebusiness.logistics", eTour, http://www.eloqu.com/webpage_german/web_eloqu/mobil.htm , last requested on 10.04.2006.

[7] "Herzlich Willkommen auf den Internetseiten der cool IT GmbH", http://www.coolit.ch , last requested on 10.04.2006.

[8] "Sennheiser, guidePORT – Exhibits come to life", http://www.guideport.de , last requested on 10.04.2006.

[9] "YOU HEAR WHAT YOU SEE - PICKUP AUDIO GUIDE", http://www.dataton.com/pickup , last requested on 10.04.2006.

[10] “BeyondGuide – Your personal audio guide on your mobile phone”, http://www.beyondguide.com , last requested on 10.04.2006.

[11] ”Spatial Adventures – The foremost service provider of cell phone based audio tours and home of Mobile Touring Service”, www.spatialadventures.com , last requested on 10.04.2006.

[12] “User-Activated Audio Beacons (Ping!)”, http://www.touchgraphics.com/ping!.htm , last requested on 10.04.2006.

[13]M. Rahnema, „Overview Of The GSM System and Protocol Architecture“, IEEE Communications Magazine, April 1993, pp. 92-100. http://www.cse.iitb.ac.in/~anil/MTP/GSM-Overview.pdf , last requested on 10.04.2006.

[14] K. Järvinen, “STANDARDIZATION OF THE ADAPTIVE MULTI-RATE CODEC”, in European Signal Processing Conference (EUSIPCO), Tampere, Finland, September 2000. http://www.eurasip.org/content/Eusipco/2000/sessions/ThuAm/SS1/cr1956.pdf , last requested on 10.04.2006.

[15] J. Moffitt, “Ogg Vorbis--Open, Free Audio--Set Your Media Free”, Linux Journal, January 2001. http://delivery.acm.org/10.1145/370000/364691/a9-moffitt.html?key1=364691&key2=0424664411&coll=GUIDE&dl=ACM&CFID=69161095&CFTOKEN=96786742 und http://www.linuxjournal.com/issue/81 , last requested on 10.04.2006.

[16] P. Kabal, “Audio File Format Specifications”, TSP Lab, ECE, McGill University, http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/WAVE.html , last requested on 10.04.2006.

[17] Microsoft, ”Windows Media Audio 9 Series Codecs”, http://www.microsoft.com/windows/windowsmedia/forpros/codecs/audio.aspx , last requested on 10.04.2006.

[18] Moving Picture Expert Group, “The MPEG Home Page”, http://www.chiariglione.org/mpeg/ , last requested on 10.04.2006.

[19] C. Timmerer, T. DeMartini and H. Hellwagner, “The MPEG-21 Multimedia Framework: Conversions and Permissions”, P. Horster (Ed.): D-A-CH Security 2006 – Bestandsaufnahme, Konzepte, Anwendungen, Perspektiven, 2006, pp. 225-235.

[20] H.R. Nielson and F. Nielson, “Semantics with Applications; A Formal Introduction”, revised edition, John Wiley & Sons, original 1992 (1999).

[21] A. Oermann, A. Lang and J. Dittmann, “Verifier-Tuple for Audio-Forensic to Determine Speaker Environment”, City University of New York (Veranst.): *Multimedia and security, MM & Sec’05, Proceedings ACM, Workshop* New York, NY, USA, August 1-2 2005, pp. 57-62.

[22] A. Oermann, J. Dittmann and C. Vielhauer, “Verifier-Tuple as a Classifier of Biometric Handwriting Authentication - Combination of Syntax and Semantics”, Dittmann, Jana (Hrsg.), Katzenbeisser, Stefan (Hrsg.), Uhl, Andreas (Hrsg.): *Communications and multimedia security,* Berlin: Springer, Lecture notes in com, CMS 2005 9th IFIP TC-6 TC-11 international conference Salzburg, Austria, September 2005, pp. 170-179.

[23] N. Chomsky, “Syntactic Structures”, Mouton and Co., Den Haag, 1957.

[24] N. Chomsky, “Aspects of the Theory of Syntax”, MIT Press, Massachusetts Institute of Technology, Cambridge, MA, 1965.

[25] S. Löbner, “Semantik: eine Einführung“, De Gruyter Studienbuch, Berlin, 2003.

[26] K. Thibodeau, “Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years”, Council on Library and Information Resources: *The State of Digital Preservation: An International Perspective*, Conference Proceedings, July 2002. http://www.clir.org/pubs/reports/pub107/thibodeau.html , last requested on 05.04.2006.