STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **XLVII**, Number 2, 2002

THE ROAD TO REAL MULTIMEDIA DATABASES – EMERGING MULTIMEDIA DATA TYPES

HOREA TODORAN

Abstract. This paper describes our view on multimedia data types, which are the fundamentals of real multimedia database management systems. Most of the research efforts in previous work have been focused on audio-visual data and their impact on the design and implementation of multimedia systems. We also take into account the emergence of new media (generically called 'non audio-visual media'), which have the potential to revolutionize the human-computer interaction and bring multimedia database management systems in a new era.

Keywords: Multimedia databases, Multimedia data types, Digital smell, Digital taste, Digital touch

1. Introduction

Because of both the complexity of the term "multimedia" and the diversity of the application fields of the database technology, *multimedia database management systems (MMDBMS)* have different meanings for different groups of users. They are often identified by CD-ROMs storing multimedia information, or by video-on-demand systems allowing users to choose a movie from a database and play it on their own screens, or by document-imaging systems, or by other types of database systems (relational, object-relational, object-oriented, spatial) able to manipulate multimedia elements [Khoshafian96]. All of these illustrate important features of the multimedia database technology, but none of them is exhaustive.

In the referenced literature, a *real* MMDBMS should be able to:

(1) Operate with at least all the audio-visual multimedia data types (as defined in this paper);
(2) Fulfill all the requirements of a real database management system (data persistence, transactions, concurrency control, system recovery, queries, versioning, data integrity, data security, expandability);
(3) Manipulate huge data volumes (virtually no restriction concerning the number of multimedia structures and their size);
(4) Allow interaction with the user;

2000 *Mathematics Subject Classification.* 68P15.
1998 *CR Categories and Descriptors.* H.2.4. [**Information Systems**]: Database Management – *Systems*.

(5) Retrieve multimedia data based on their content (attributes, features and concepts);
(6) Efficiently manage data distribution over the nodes of a computer network (distributed architecture).

Fostered by the development of specific input and output devices, new, *non audio-visual digital media* emerge. 'Digital smell', 'digital taste' and 'digital touch' will become reality and will radically improve the human-computer interaction of the future. Computer-generated multimedia presentations will not be limited to image and sound as they are today. Instead, they will have a greater impact on user's perceptions, allowing for a computer-controlled ambient.

In terms of multimedia databases, the emergence of non audio-visual digital media will lead to new storage, retrieval and presentation challenges. Consequently, the definition of a real MMDBMS will have to be reconsidered, i.e. adapted to the new challenges. A first attempt is made in the next sections of this paper.

The rest of this paper is organized as follows: The next section gives an overview of the audio-visual multimedia data types, which are indispensable in a real MMDBMS. Emerging 'non audio-visual media' are introduced in section 3, together with specific devices and possible evolutions. Our first definition of a real MMDBMS is given in section 4. Then, we conclude and present future work.

2. Overview of adudio-visual multimedia data types

2.1. **Minimal data type requirements for MMDBMS.**

Most of the DBMS developed in the last years which claim to be multimedia, have the capacity to operate with only one data type. Even if this only data type is video, audio or image, a system of this kind cannot be considered as a *real multi*media database system (MMDBMS).

For example, *image database management systems*, even if they are able to deal with very large collections of

images and to offer advanced techniques for content-based retrieval (e.g. the PIQ system, described in [18, 19, 3]), are not *real* multimedia database management systems, because of their limitation to only one data type. In our opinion, the same is true for video database systems, which offer advanced techniques for storing, archiving, querying and visualizing digital video – e.g. the VideoSTAR system, developed by the Norwegian Institute for Technology [5, 6], the HERMES/AVIA prototype from GMD Darmstadt [22, 7] and MMVIS from Michigan University [14, 4].

We do not intend to diminish the extremely important contribution that the above-mentioned systems and the related research bring to the development of various techniques, successfully used in the management of multimedia data.

Nonetheless, we believe that:

Statement 1:

A multimedia database management system (MMDBMS) must be able to operate with at least all of the following basic audio-visual multimedia data types: text, image, graphics, audio and video.

Let us define the set of basic audio-visual multimedia data types for further use and explain what does it mean for a DBMS to be able to *operate* with a specific data type.

Definition 2 - Set of basic audio-visual MM data types

The set of basic audio-visual multimedia data types (BAVT) is defined as:

BAVT = {TEXT, IMAGE, GRAPHICS, AUDIO, VIDEO}

Definition 3 - Operate with an abstract data type

A DBMS is able to ***operate*** *with a specific abstract data type (ADT) when instances of the ADT can be manipulated (i.e. created, updated, deleted, retrieved) independently from other types of data, by means of their own specific techniques.*

By way of combination of BAVT objects, new complex objects, which are multimedia them selves, can be created to be recognized and manipulated by the system.

2.2. **Classification of audio-visual data types.** Taking into account their time-dependency, audio-visual multimedia data are divided in two main categories, as follows:

Definition 4 - Discrete and continuous data

Data not depending on a time scale are called ***discrete data*** *or* ***static data****.*

Data depending on a time scale are called ***continuous data*** *or* ***dynamic data****.*

Text, graphics and image are discrete data, while audio and video are continuous.

Continuous data are more complex than discrete data, which implies the use of much better compression/decompression algorithms and more sophisticated operations for their interpretation and manipulation (see [24, 8, 15]).

The main features and concepts related to the basic audio-visual multimedia data types are described in [15, 10, 20].

2.3. **Generated media.** In [15, 11] some other multimedia data types are described, which are called *generated media.* They are different kinds of computer-generated presentations, the most popular being *animation* and *music.*

If they are stored in audio or video files, then there is practically no difference between generated media, on the one hand, and audio and video data, on the other hand. Yet, if they are generated during the presentation (*real-time*), using specific devices and instruments, we assert that they must be treated as distinct media.

Let us now give our definition for generated media:

Definition 5 - Generated media (GM)

Generated media are computer-generated real-time multimedia presentations based on human-computer interaction.

$$GM \supset \{ANIMATION,\ MUSIC\}$$

The main advantage of the generated media over audio and video data resides in a much better interaction with the user, which is crucial in the case of MMDBMS.

Generated media are *interactive media*. Their manipulation requires simultaneous control of devices and efficient interpretation of user-generated interrupts.

In terms of *Definition 4* generated media are *continuous*, as long as they essentially depend on a time scale.

2.4. **Speech.** Due to the recent development of advanced techniques for *speech recognition* and *speech understanding* (see [12, 13, 21]), *speech data* are likely to be treated independently from audio data.

The main difference between *speech recognition* and *speech understanding* resides in the fact that the latest implies action taken by the system in response to the vocal command of the user. Current speech recognition systems have better than 95% accuracy and the errors that might occur are very easy to correct. Speech understanding is more complex, especially when the semantic of the command plays an important role.

Speech is also a continuous medium.

2.5. **Synthesis of audio-visual data types.** Based on the previous definitions, a synthetic view on audio-visual media is presented in the table bellow:

	BAVT	**GM**	**Sp.**
Discrete media	TEXT GRAPHICS IMAGE		
Continuous media	AUDIO VIDEO	ANIMATION MUSIC	SPEECH

Table 1. Table 1: Audio-visual data types

3. Emerging non audio-visual multimedia data types

Most of the research in the field of human-computer interaction has been focused, until recent years, solely on audio-visual technologies. Up to a certain point, this can be explained by the natural evolution of the human societies, built on the communication between their members, which is mainly based on signes (writing) and sounds (oral communication). Television, the most important mass-media of the last century, is also made of image and sound. Therefore, more and more sophisticated audio and video devices have been developed by the electronic industry, followed by the associated software tools. This evolution has also had a great impact on the database technology, leading to the development of today's MMDBMS.

But, as far as human beings are endowed with five senses, why concentrate the whole effort only on two of them? Why not trying to further improve the human-computer interaction by means of adding the strength of *smell*, *taste* and *touch*? A positive answer is given by the development, in the last few years, of a new generation of hardware devices and software applications, which we briefly describe in the next subsections of this paper. They will lead to the emergence of new media, that are non audio-visual and that will probably have the same impact on human-computer interaction as audio-visual media have had in the early 1980s.

To what extent will affect the new non audio-visual media the database field is almost impossible to accurately predict. However, we give our vision on the consequences of the new non audio-visual media for the multimedia database design and implementation.

3.1. **Olfactory input and output interfaces.** Olfactory and tasting interfaces seem to be the least developed among the human-computer interaction technologies. This is mainly due to the lack of useful applications, comparing with the other sense-based technologies. However, the use of scents and taste in military (chemical and biological warfare detectors), medicine (surgical training) and electronic commerce (sample of groceries, cosmetics, household products) has fostered the research on olfactory and tasting systems in the last few years.

There are two types of *olfactory interfaces*, briefly described below: olfactory input interfaces and olfactory delivery (output) systems.

Olfactory input interfaces, also called *electronic noses*, are used to collect and interpret odours (very useful in product quality control and warfare detectors). There are three basic approaches to this kind of input devices:

- **gas chromatography:** separation, identification and determination of chemical components in a complex mixture using the differences in migration rates among the sample components;
- **mass spectrometry:** detects patterns of the molecules using the difference in mass-to-charge ratio of ionised atoms;
- **chemical sensor arrays:** based on the multisensing principle, in which the distributed response of an array of chemical sensors is used to identify the constituents of a gaseous environment (eg. ENOSE by JPL and Caltech).

Olfactory delivery systems are a combination of at least four different processing steps: odour storage (liquid, gels, microencapsulation), odour selection, evacuation and cleaning of exhaled air and odour display. Olfactory delivery systems are already available for the consumer market – e.g. the SENX scent device from TriSenx (`http://www.trisenx.com`).

The Sniffman portable scent system from Ruetz Technologies (Germany) has already been adapted for a multimedia entertainment application – Duftkino - the smell cinema (`http://www.duftkino.de`). "One Day Diet" is the first movie for the nose ("ein Film für die Nase"), allowing audience also to smell the action.

France Télécom R&D and the Burgundy wine industry association (BIVP) are creating a website that lets visitors take an olfactory stroll through the famous vineyards of Burgundy. Aromas, pictures and sounds will be brought together to recreate the atmosphere of the vineyards.

3.2. **Tasting interfaces.** Tasting systems, frequently called *electronic tongues*, mimic their natural counterparts, being already able to distinguish between sweet, sour, salty and bitter[1], and having the potential to respond to a dazzling array of subtle flavours. Even more, e-tongues can also "taste" cholesterol levels in blood, cocaine concentration in urine, or toxins in water, which means that they can return both qualitative and quantitative results. Most of the applications of electronic tongues are in the field of quality control (flavours, beverages, fragrances, pharmaceuticals) and medicine (blood and urine tests).

Recent examples of e-tongue prototypes include:

The e-tongue prototype developed at University of Texas[2] is made of polymer microbeads positioned on a silicon chip of about $1cm^2$ and arranged in tiny pits to represent taste buds. Each pit is marked with dye to create a RGB color bar, which changes when in contact with a chemical. A camera connected to a computer examines the colors and performs a RGB analysis to determine what tastes are present.

The "Astree Liquid & Taste Analyzer" produced by Alpha-MOS (Web address `http://www.alpha-mos.com`). This analyses a liquid matrix using sensor reactions and different statistical pattern recognition techniques to classify tastes. It was the first tasting system commercially available.

The hand-held device produced by Antonio Riul at Embrapa Instrumentação Agropecuária in São Carlos, Brazil (`http://www.embrapa.br`). It is able to detect low levels of impurities in water and discriminate between Cabernet Sauvignons of the same year from two different wineries, and between those from the same winery but different years. It can also spot molecules such as sugar and salt at concentrations too low for human detection [1].

3.3. **Haptic interfaces.** Haptic interfaces are devices that measure the motion of, and provide sensory stimulus to, the users' hands and fingers. A haptic device provides information to the computer based on the device's position (the way a mouse does) and stimulates users' sense of touch by supplying output in the form of force feedback and tactile, or haptic, feedback. Haptic devices make it possible for users to "touch", feel, manipulate, create, and/or alter with their own hands and fingers, objects presented on computer displays as if they were real physical objects. This is done by carefully calculating the forces one would feel when touching a real object and then presenting these forces to users by using the force feedback and tactile display capability of a haptic device. When done properly, this creates the illusion of "touching" the object.

Haptic interfaces can be used to train physical skills such as those jobs requiring specialized hand-help tools (e.g. surgeons, astronauts, mechanics), to provide haptic-feedback modelling of three dimensional objects without a physical medium (such as automobile body designers working with clay models), or to mock-up developmental prototypes directly from CAD databases (rather than in a machine shop).

---

[1] Recently, a fifth candidate basic taste was identified: *umami*, the taste of monosodium glutamate, characteristic of protein-rich foods (`http://www.umami.it`)

[2] Further information at: `http://weewave.mer.utexas.edu/MED_files/MED_research/MEMS_chem_snsr/beads/bead_sensor.html`

Based on the interaction between de user and the machine, haptic devices can be classified as:

Finger-based: attached to user's finger and responding to its movements. Examples include PHANToM (developed at MIT, but commercialised by SensAble), the pen based device from University of Washington, Rutgers Masters (RM-I, RM-II), Feelit Mouse by Immersion.

Hand-based: users interact with the device by grasping a rigid tool. The machine gives the human arm the sensation of forces associated with various arbitrary manoeuvres. Prototypes have been developed at several universities (Carnegie Mellon, McGill, Northwestern, Rutgers and so on); commercial products: TouchSense by Immersion, Cyberglove and CyberTouch of Virtual Tech.

Exoskeletal: track the movements of user's arm, shoulder or even of the whole body, allowing high interactivity, but at extremely high prices. These machines are mostly used in medicine, for people with disabilities, and military. Examples of commercially available products include Cybergrasp by Virtual Technologies, Dextrous Arms and Hands from Sarcos, Arm Master by Exos.

Inherently passive devices (or intelligent assist devices): these are passive, therefore safe, robots for direct physical interaction with human operators within shared environments. They use intelligent microcontrollers, servo-motors and an advanced "sense/process/actuate" control concept to quantify the speed and direction of motion that the user wants. This information is then processed and the proprietary algorithms direct the movement of the device, no time lapse between the machine's sensing and its response being noticed by the operator (see `http://cobot.com`).

Besides force feedback, other tactile display technologies include:

Vibration: Vibration can be used to transmit information about texture, puncture, slip, and impact. Since vibrations often are sensed as being diffuse or unlocalised, a single vibrating device for each finger or skin area is often sufficient.

Thermal display: Thermal perceptions of an object are based on a combination of thermal conductivity, thermal capacity, and temperature. This enables users to infer material composition as well as temperature.

Small-scale shape or pressure distribution: The most frequently used devices have an array of closely spaced pins that can be raised or lowered individually against the skin to approximate a shape. To conform to the human ability to perceive tactile sensations, the pins must be spaced within a few millimetres of one another.

Other tactile display technologies: These include electrorheological devices (materials that use a "smart" fluid which can change viscosity in an electrical field) combined with sensors, electrocutaneous stimulators (using electrodes to stimulate cutaneous nerve endings), ultrasonic friction displays, and rotating disks for creating the sensation of slip. The MEMICA prototype from Rutgers University is a good example.

While online scent services are almost ready to be launched on the market, taste and touch seem to be harder to address.

3.4. **Steps towards the integration of smell, taste and touch in real MMDBMS.** In our view, the first and very important step is to *increase the number of application areas* for digital smell, taste and touch and *gain public acceptance*. In [9] possible uses of computer-generated scent are revealed, both in the public spaces (malls, theme parks, retail spaces) and the individual sphere (high-end gamers, aromatherapy to enhance memorization and learning, ubiquitous computing, individuals with special needs – blind, deaf). For this to happen, low-cost standardized devices are needed.

Taking into account the experience gained in image, audio and video, the next step will probably be to develop *specialized databases*. We presume that smell databases, taste databases and touch databases will evolve separately and specific storage devices, querying techniques and presentation methods will be built for each of the three media. Because of the wide effort required for such systems to be put together, it is very realistic for any research team to focus on only one of these non audio-visual media at a time.

The process of integration of digital smell, taste and touch in a real MMDBMS will most likely continue in the form of a *multimedia federated database system*. By means of *wrappers* diverse audio-visual data sources have been already integrated in coherent systems where high-performance query optimization is achieved (see [2, 17]). The same strategy is likely to be adopted by developers also in the case of emerging non audio-visual media rather then building new multimedia systems from scratch. A great challenge will be to send these media across networks, which requires infrastructure upgrading.

On the top of a MMDBMS that also includes smell, taste and touch, amazing *five-senses multimedia applications* will be built.

4. Our first definition of a real MMDBMS

In the view of the topics discussed above, we give our first definition of a real MMDBMS, which is an extension of various definitions found in the referenced literature:

Definition 6: Real MMDBMS

A real multimedia database management system has the following characteristics:

(1) *(Data types) Is able to operate with the following multimedia data types:*
   (a) *basic audio-visual data types: TEXT, GRAPHICS, IMAGE, AUDIO, VIDEO;*
   (b) *generated media: ANIMATION, MUSIC;*
   (c) *SPEECH*
   (d) *non-audio-visual data types: SMELL, TASTE, TOUCH*

(2) *(Database features) Fulfill all the requirements of a real database management system (data persistence, transactions, concurrency control, system recovery, queries, versioning, data integrity, data security, expandability);*
(3) *(Storage) Manipulate huge data volumes (virtually no restriction concerning the number of multimedia structures and their size);*
(4) *(Interaction) Allow interaction with the human operator through all the five senses;*
(5) *(Queries) Retrieve multimedia data based on their content (attributes, features and concepts);*
(6) *(Distribution) Efficiently manage data distribution over the nodes of a computer network (distributed architecture).*

Note: This definition is likely to change in the future, according to advanced results in digital smell, taste and touch.

5. Conclusions and future work

In this paper we have discussed a basic aspect of real multimedia database management systems: the multimedia data types. We have had a brief overview of the audio-visual multimedia data types, which we have defined and classified. Then, we have presented various input and output technologies, which foster the emergence of new multimedia data types, called here "non audio-visual data types": smell, taste, touch. Integrating these new data types in multimedia database systems is a very difficult task. We have imagined three incipient steps. We have also given our first definition of a real MMDBMS.

In the future we plan to analyse the opportunity of building smell database and then integrating smell into federated multimedia database management systems.

6. Acknowledgements

Dr. Thomas Skordas is acknowledged for contributions to this paper and for giving me a first view on real human-computer interaction.

References

[1] Ball, Philip – "Electronic tongue has a good taste", Nature Science Update, 9 January 2002. `http://www.nature.com/nsu/020107/020107-3.html`

[2] Berthold, Henrike – *"A Federated Multimedia Database System"*, Proc. ACM Multimedia '99 Doctoral Symposium, Orlando, Florida, 1999.

[3] K. Beyer, J. Goldstein, R. Ramakrishnan and U. Shaft – *"When is Nearest Neighbor Meaningful?"*, Proc. ICDT, 1999.

[4] Hibino, Stacie Lynn – *"MultiMedia Visual Information Seeking (MMVIS): A Multimedia Interactive Visualization Environment for Exploration and Spatio-Temporal Analysis of Video Data"*, PhD thesis, Michigan University, Ann Arbor, 1998.

[5] Hjelsvold, Rune – *"Video Information Contents and Architecture"*, Proceedings of the 4th International Conference on Extending Database Technology, Cambridge, 1994.

[6] Hjelsvold, Rune; Roger Midtstraum and Olav Sandsta, *"Searching and Browsing a Shared Video Database"*, Proceedings of the 1995 International Workshop on Multi-Media Database Management Systems, 1995.

[7] Hollfelder, Silvia; Everts, André; Thiel, Ulrich – *"Designing for Semantic Access: A Video Browsing System"*, in Proc. of IEEE Int. Conf. On Multimedia Computing and Systems (ICMCS), 1999.
[8] Kamath, Mohan; Ramamritham, Krithi; Towsley, Don – *"Continuous Media Sharing in Multimedia Database Systems"*, Proc. of the 4th International Conference on Database Systems for Advanced Applications (DASFAA'95), Singapore, 1995.
[9] Kaye, Joseph Nathaniel – "Symbolic Olfactory Display", PhD Thesis, MIT Media Lab, May 2001. http://web.media.mit.edu/∼jofish
[10] Khoshafian, Setrag; Baker, Brad -, *"Multimedia and Imaging Databases"*, Morgan Kuafmann Publ., USA, 1996.
[11] Klas, W., Aberer, K. - *"Multimedia Applications and Their Implications on Database Arhitecture"*, GMD Technical Report, Darmstadt, Germany, 1995.
[12] *** - *"Emerging technologies that will change the world – The Technology Review Ten"*, Massachusetts Institute of Technology's magazine of Innovation, Jan/Feb 2001. http://www.technologyreview.com.
[13] *** - *"Voice recognition improves - but still suffers from selective hearing"*, 16 Jan 2001. http://www.msnbc.com/news/516591.asp.
[14] Nwosu, C. K.; Thuraisingham, B.; Berra, P.B. – *"Multimedia Database Systems, Design and Implementation Strategies"*, Kluwer Academic Publishers, USA, 1996.
[15] Rakow, Thomas C.; Neuhold, Erich J.; Löhr, Michael – *"Multimedia Database Systems – The Notions and the Issues"*, in G. Lausen, editor, Tagungsband GI-Fachtagung Datenbanksysteme in Büro, Technik und Wissenschaft (BTW, Informatik Aktuell, p. 1-29, Springer, March 1995.
[16] Rakow, Thomas C.; Klas, Wolfgang; Neuhold, Erich J. – *"Research on Multimedia Database Systems at GMD-IPSI"*, IEEE Multimedia Newsletter 4(1): 41-46, April 1996.
[17] Roth, Mary Tork; Özkan, Fatma; Hass, Laura M. – *"Cost Models DO Matter: Providing Information for Diverse Data Sources in a Federated System"*, Research Report, IBM Research Division, IBM Almaden Research Center, San Jose, California, 1999.
[18] Shaft, Uri; Ramakrishnan, Raghu – *"Data Modeling and Querying in the PIQ Image DBMS"*, Bulletin of the IEEE Computer Science Technical Committee on Data Engineering, 1996.
[19] Shaft, Uri; Ramakrishnan, Raghu - *"Content-Based Queries in Image Databases"*, Data Engineering Bulletin, February 1997.
[20] Smeureanu, Ion; Drulă, Georgeta – *"Multimedia – concepte şi practică"*, CISON Publishers, Bucharest, 1997 (in Romanian).
[21] Sumedha Kshirsagar, Chirs Joslin, Won-Sook Lee, Nadia Magnenat-Thalmann, *"Personalized Face and Speech Communication over the Internet"*, in IEEE Virtual Reality 2001.
[22] Thiel, Ulrich; Hollfelder, Silvia; Everts, André – *"Multimedia Management and Query Processing Issues in Distributed Digital Libraries: A HERMES Perspective"*, Proceedings of 9th International Workshop on Database and Expert Systems Applications (DEXA 98), August 26-28, 1998; Vienna, Austria, Los Alamitos, California: IEEE Computer Society, 1998, pp. 84-89.
[23] *"Towards Natural Human Computer Interaction"*, Jorge Santos (European Commission, DG Information Society), 29 June 2000.
[24] TURAU, Volker; RAKOW, Thomas C. – *"A Schema Partition for Multimedia Database Management Systems"*. http://www.gmd.de
[25] Ţâmbulea, Leon – *"Structuri de date şi bănci de date"*, "Babeş-Bolyai" University, Cluj-Napoca, Romania, 1992 (in Romanian).

Faculty of European Studies, Babes-Bolyai University, Cluj-Napoca
E-mail address: htodoran@euro.ubbcluj.ro