Filmification of Words and Sentences Towards Teaching and Learning

Tetsuya Hirotomi, Nikolay N. Mirenkov and Kazuki Fujita
Graduate Department of Information Systems
The University of Aizu
Aizu-Wakamatsu City, Fukushima 965-8580, Japan
Email: {d8021105, nikmir, m5051128}@u-aizu.ac.jp

***Abstract*— We have developed a multimedia communication environment where multimedia is used to fully utilize available communication (including teaching and learning) capabilities of different users. This environment is based on "film" formats of multimedia words and sentences. Corresponding films are multiple views of objects, processes, etc. They are "pieces of knowledge." These pieces are acquired in a film database. Multimedia word dictionary panels provide effective access to the database items. In this paper, film technology concept, overview of the environment, and filmification of words and sentences to support communication of children, handicapped and elderly people are presented. A special attention is paid to implementing a multimedia word dictionary.**

I. Introduction

Nowadays, computers and the Internet are widely used in educational context. A great number of systems supported by ICT (Information and Communication Technologies) has been developed: for example, tutorials, drill-and-practice, integrated learning management systems, simulations, Java applets, Web-based course support sites, and video-on-demand lecture presentations [1]. In addition, national governments provide special support for distributing such systems to schools and other educational organizations. For example, in Japan, MEXT (Ministry of Education, Culture, Sports, Science and Technology of Japan) decided to phase in eight computers and totally 27 educational software systems over the five years since 2000 to every school for the blind, deaf, and other disabled. So, many systems targeted to children with mental/physical disabilities have been actively developed by ICT professionals in companies and universities [2]–[6], as well as by teachers in schools for the disabled [7], [8]. Despite many systems effectively utilize ICT, some of them are not so successful because of the following reasons:

- Lack of system usability. It means that such a system is difficult to learn, inefficient to use, difficult to remember, easy to make errors, and unpleasant to use [9].
- Lack of universal usability. It means that a system does not take into account technology variety, enormous diversity of users, and gaps in user knowledge [10].
- Imbalance of "education" and "entertainment." Many drill-and-practice systems for children in preschool and early school years pay too much attention to attract learners by entertainment features. As a result, learners in such systems just play games or memorize combinations of questions and answers without considering knowledge behind.

To teach, learn and use "communication" skills, the development of systems taking into account the above mentioned reasons are required by the schools (especially, schools for physically handicapped and mentally retarded children). Their students have difficulties to read, write/type, and spell texts because of congenital (e.g. mental retardation, cerebral palsy, autism, and developmental apraxia of speech) and/or acquired (e.g. amyotrophic lateral sclerosis, multiple sclerosis, traumatic brain injury, stroke, and spinal cord injury) impairment [11]. In the area of "AAC (Augmentative and Alternative Communication) [12]," some attempts have been tried to overcome these difficulties by adopting different communication symbols instead of conventional texts (e.g. gestures, vocalizations, real/miniature objects, photographs, abstract pictorial symbols [13]–[16], Braille and sign languages). However, in many cases, these symbols take essential time to master their language in a way as native languages are used. So, the number of people who can communicate via these languages is very limited.

That is why we are developing a new multimedia communication environment [17] for children, handicapped and elderly people (however, according to our study, this technology can be tailored to the needs of other people). This environment is named "F-Communication System (FCS)." "F" stands for "film," because the basis of our approach is "filmification" of words and sentences. Filmification is a multimedia modeling of them where explanations from multiple view points are supported. Each explanation is supported by visual images, animations, sounds, texts, etc. Therefore, words and sentences are represented, in a sense, by self-explanatory forms, and so called "multimedia words" and "hieroglyphs." They allow a sender/receiver can express/understand messages via his/her favorable views. That is, FCS users possess the potential to communicate with a variety of users. To support distance/direct human-human communication, vocabulary development, hieroglyph literacy education, etc., a set of rather independent subsystems have been developed within the framework of FCS. In these subsystems, the users can manipulate (retrieve, browse, edit, send, receive, etc.) films stored into a network accessible database by mouse clicks, as well as

understand them through browsing their multiple views. These processes are supported by user friendly technologies based on picture type panels and multiple interfaces. Picture type panels are designed with natural outward appearances. Such a design can reduce inhibitions on the use of complex "mechanical" GUI (Graphical User Interface) panels, and implicitly provide additional hints for the user interaction based on his/her experience. Multiple interfaces are designed to adapt the systems to their users. The adaptation is based on user/device profiles and special statistics of user's activity in the system. In fact, one multimedia film language is used to represent words and sentences, and another is used to perform the system interface operations.

This paper presents one of the subsystems, named "multimedia word dictionary," to aid in teaching and learning multimedia words. Its development is based on regular contacts with teachers and students in Inawashiro School for Physically Handicapped and Mentally Retarded Children, and professors in Department of Social Welfare, Junior College Division, The University of Aizu, as well as contacts with some others at irregular intervals. In this paper, film technology concept, overview of FCS, films to represent multimedia words and hieroglyphs, and films to perform the system interface operations, and implementation of multimedia word dictionary are described.

II. Film Technology Concept

"Seeing is believing." Visual images explaining the meaning of objects or processes have been used since antiquity [18]. Such images are easy to understand, but technology could not support "writing" them at that time. As a result, visual images were substituted by texts, and then used mainly for illustration or ornamental purposes. Now, ICT provides opportunities to use visual images, as well as other media. Each medium can express one view of the corresponding object or process regarding to its own strengths and weakness [19], and all together simplify understanding. In the area of education and training, multimedia presentation services have been considered, for example, digital TV services. In these services, digital films of educational materials are distributed by satellite, cable, and terrestrial systems, and the user can interact with such films (e.g. watch/hear, pause, rewind and forward) by a digital set-top box connected with TV and its controller [20].

We have been developed a "film" technology for taking one more step in the use of multimedia through generalizing conventional films. A conventional cinefilm has a linear structure of still-frames where each of them shows just one "time slice" of animation. Our "film" has a nonlinear structure of "view-frames" where each of them shows at least one feature of an object/process via multimedia (texts, images, animations, etc.). These views are:

- systematically assembled in a "film database,"
- selected and adapted to each user depending on his/her culture, age, disabilities, and other backgrounds,
- arranged in linear and non-linear order,
- and then provided on the user's demand.

In this way, objects/processes in a "film" format become, in a sense, self-explanatory. It is important to note that our film is not only materials for watching/hearing to acquire corresponding knowledge, but also sources of creating new films (knowledge) through editing and composing. All these operations (including watching/hearing in step-wise or non-stop modes, pausing, rewinding and forwarding) can be performed by mouse clicks without typing on a keyboard via user friendly panels.

III. Overview of F-Communication System

FCS is to send/receive messages in the film format and teach/study the meaning behind them without using keyboard. It consists of a number of subsystems to support direct/distance human-human communication, vocabulary development, and hieroglyph literacy education of children, handicapped and elderly people.

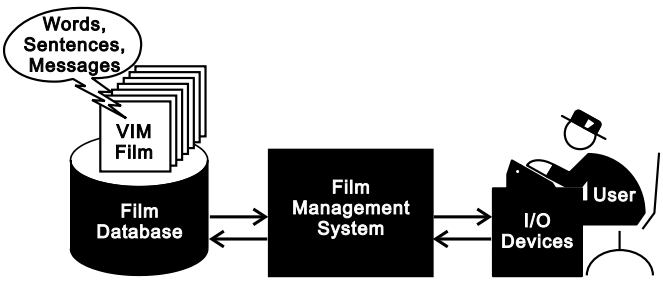

Fig. 1. System Architecture

Fig.1 depicts FCS architecture. Words, sentences, and messages are represented by **VIM (VIsualization of Messages) films**. These films, statistics of their use, and user/device profiles are acquired in a network accessible **film database**. **FMS (Film Management System)** provides **MUI (Multiple User Interface) films** for accessing the database items, as well as browsing and manipulating VIM films. FMS is developed based on multiple interface multimedia approach in order to:

- simplify operations (e.g. accessing, browsing, and manipulating) with VIM films,
- minimize number of such operations,
- increase understandability and informativeness of VIM films.

This approach is also cope with the diversity of users. To interact with FMS, a variety of **I/O (Input and Output) devices** that fit the users' needs can be used (e.g. a touch panel, foot mouse, eye-gaze system, and force-feedback device [21], [22]), because typing on a keyboard is excluded or essentially reduced.

The significant point of FCS is the development of two types of films:

- VIM films to explain multiple view of words and sentences,
- MUI films to operate VIM films based on user friendly technologies based on picture type panels and multiple interfaces.

Some details of these films are presented in next two sections.

IV. VIM Film

A VIM film is a series of multimedia frames representing meaning of a message, a sentence (hieroglyph), or a word where a message is a set of sentences, a sentence is a set of words, a word is a set of views (features). This meaning is based on multimedia word view-frames.

To solve some communication problems, FCS provides right view-frames for the right users according to the user/device profiles and statistics of user's activity on the system. For example, a typical communication problem lies between blind and deaf people. Deaf people communicate through a sign language, but blind people cannot understand and use it. In this case, view-frames related to the sense of sight are provided for deaf people, and view-frames related to the sense of hearing (and touch) are provided for blind people. Thus, these people can communicate through VIM films.

A. Multimedia Word

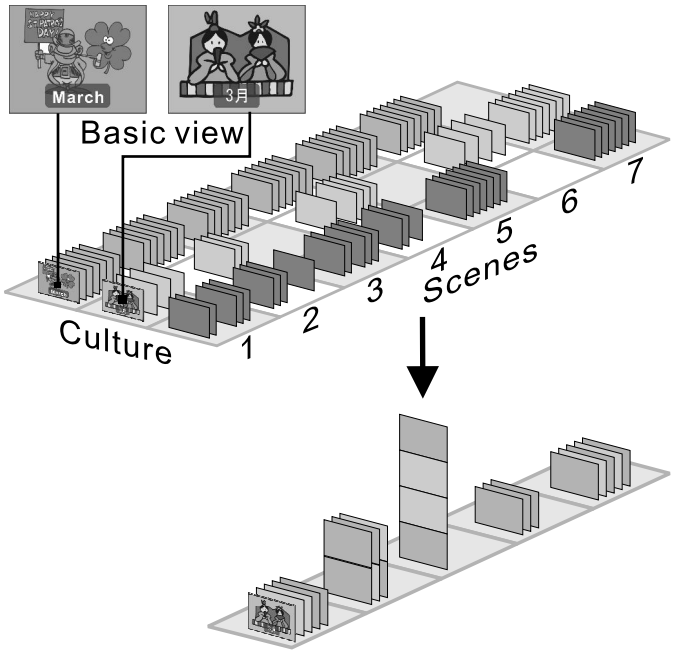

Fig. 2. Multimedia Word Example

A **multimedia word** is a multiple view representation of a word. Each view shows at least one feature of the word meaning. These views are systematically assembled based on the following aspects (scenes):

1) possible perception of the object/process through our five senses (sight, hearing, touch, smell, and taste),
2) internal structures or materials used for the object/process,
3) external compositions where it can be an item of more complex objects/processes,
4) how it can be used and/or how it works,
5) its neighborhoods in meaning, space, and/or time,
6) objects/processes of its opposite meaning,
7) culture/language dependence.

Fig.2 shows an example of a multimedia word. A universal set of its views are classified by the scenes and culture as shown in the middle of Fig.2. One, so called a "basic view" of a multimedia word consists of a caption and a standard image. This view is to provide an intuitively understandable meaning of the word. For example, English and Japanese version of basic views representing "March" are shown in the top-left of Fig.2. If such a view is not enough to understand the meaning, the users can ask the multimedia word dictionary to provide other views for more detailed explanation. Then FCS select, organize, and show the additional views depending on the user/device profiles and statistics as shown in the bottom of Fig.2. These views can be watched/heard one by one or some views at once in linear/non-linear order.

It is important to note that each view can have a multi-cultural aspect. For example, captions and standard images of the basic view is automatically replaced by the system according to the user's language and culture (now, English-Japanese and Western-Japanese replacements are involved).

B. Multimedia Hieroglyph

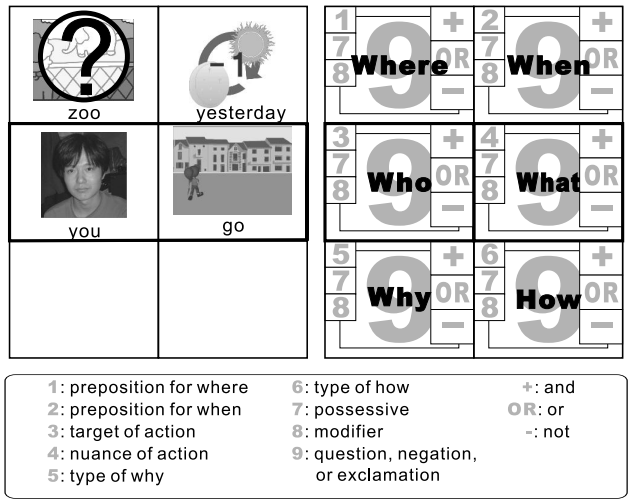

Fig. 3. Multimedia Hieroglyph Example (Left), and The Predefined Cells and Their Sub-positions (Right)

A **multimedia hieroglyph** is a collection of multimedia words to represent a sentence. For example, four multimedia words are used an item of a hieroglyph in the left part of Fig.3. This hieroglyph represents, "Did you want a zoo yesterday?" Basically, we use six multimedia words to answer 5W1H questions (who, what, where, when, why and how), and display their basic views into predefined positions as shown in the right part of Fig.3 (see [23] for more details). In this case, the answers are "**you**," "**go**," "**a zoo** *?*" and "**yesterday**" for who, what, where, and when, respectively. And no answer for why and how. Bold words represent a basis of the sentence meaning, and are located into six cells. Italic words make a

multimedia hieroglyph more precise, and are located into sub-positions of the cells. In this case, "*?*" is put into sub-position **9** in where cell.

Such a hieroglyph is specified through selecting words by mouse clicks without typing a text on a keyboard. As a result, the users can express their feelings, emotions, knowledge, etc. quickly by answering questions without spelling and grammatical errors.

We have been developed a prototype of Hieroglyph-to-Japanese translator. It means FCS can provide multiple views of a hieroglyph based on 2D structure (as shown in Fig.3), 1D structure depending on Japanese grammatical order of words, and text in Japanese.

V. MUI Film

A MUI film is to represent FCS subsystem interface to operate VIM films. It consists of a number of scenes for different modes (e.g. searching, creating, and reading messages). A scene has a set of "panel-frames" to perform one step of the modes. A panel-frame has a set of static/dynamic visual objects to perform corresponding operations, and/or get visual, auditory, and tactile feedbacks for the operations. Each panels are based on the following metaphors: maps, stories, natural landscapes, games, topics, and sound buttons (some of panel examples are explained in Section VI). These panels are rather user friendly, because more natural for the user than "mechanical" buttons in conventional panels, and implicitly provide additional hints through pictures themselves, animations, etc. In addition, cartoon characters on these panels, so called "screen-pets," guide the user and attract his/her attention by their voice, balloons with texts and/or images, and actions.

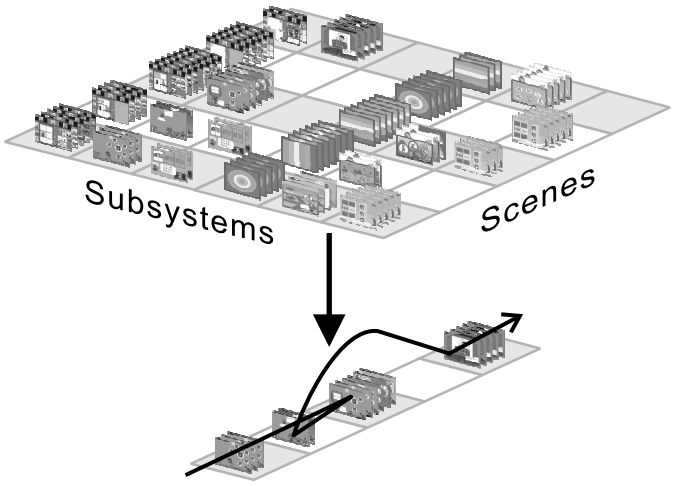

Fig. 4. MUI Film Example

Fig.4 shows an example of MUI films. They are classified by subsystems and scenes. From the universal set of films (upper part), a subset is selected on user's demand (lower part). The user can watch/hear these films like conventional cinefilms in non-stop and step-wise mode, or interact with them by clicking visual objects or screen-pets in each panel. Irregular control of films (e.g. skipping to forward/backward scenes) are supported by a screen-pets. A few mouse clicks are enough to give such instructions.

FCS copes with the diversity of users as follows. FMS provides a MUI film selected depending on the user's selection of subsystems and user/device profiles. This film consists of the right set of scenes to provide right modes (functions) for the right user depending on the role (e.g. super-user or user). The number of panel-frames for each scene is changeable depending on the user profile (e.g. cognitive developmental stage of language acquisition [24]). Finally, a number and behavior of visual objects, as well as screen-pets on panel-frames can be adapted to each user based on the profiles and statistics. Through this adapted MUI film, the user can simply access film database, then browse and manipulate VIM films also adapted based on the profiles and statistics by using his/her favorable I/O devices.

VI. Implementation of Multimedia Word Dictionary

A prototype of a multimedia word dictionary subsystem has been implemented. Here we present some technical details. The multimedia word dictionary subsystem has a communication server with an online database to store the information of multimedia resources related to VIM films, as well as the information related to user/device profiles and special statistics. This server is running on Linux and the database is implemented within the Oracle8i environment. The user interface provided by MUI films has been implemented by Java (including Java Speech API and Java Communication API). This interface is to access and browse multimedia words, as well as to show their multiple view explanations.

Functionally, this subsystem can be divided into two parts: one is related to retrieving multimedia words, and another is related to browsing multiple view explanation.

A. Retrieving Multimedia Words

For retrieving multimedia words stored in the database, five schemes are currently available as shown in Fig.5. Four of them depend on the following metaphors: maps (**A**), stories (**B**), natural landscapes (**C**), and games (**D**). One of them (**E**) depends on selecting multimedia words when browsing multiple view explanations of other multimedia word. These schemes are implemented in MUI films consisting of user friendly panels. These panels have visual objects in a picture to reduce inhibitions on the use of "mechanical" buttons in complex panels. These visual objects can show dynamic behavior, for example entering into and exiting from the panel. More to the point, such picture type panels with visual objects are more natural for the user, and implicitly provide additional hints through pictures themselves, animations, etc. In addition to that, screen-pets guide the user and attract his attention by their voice, balloons and actions (for example, a giraffe and flying squirrel in scheme **A**-**C** panels, as well as a horse and snake in scheme **D** panels). It is worth mentioning that all operations on the panels can be performed only by mouse

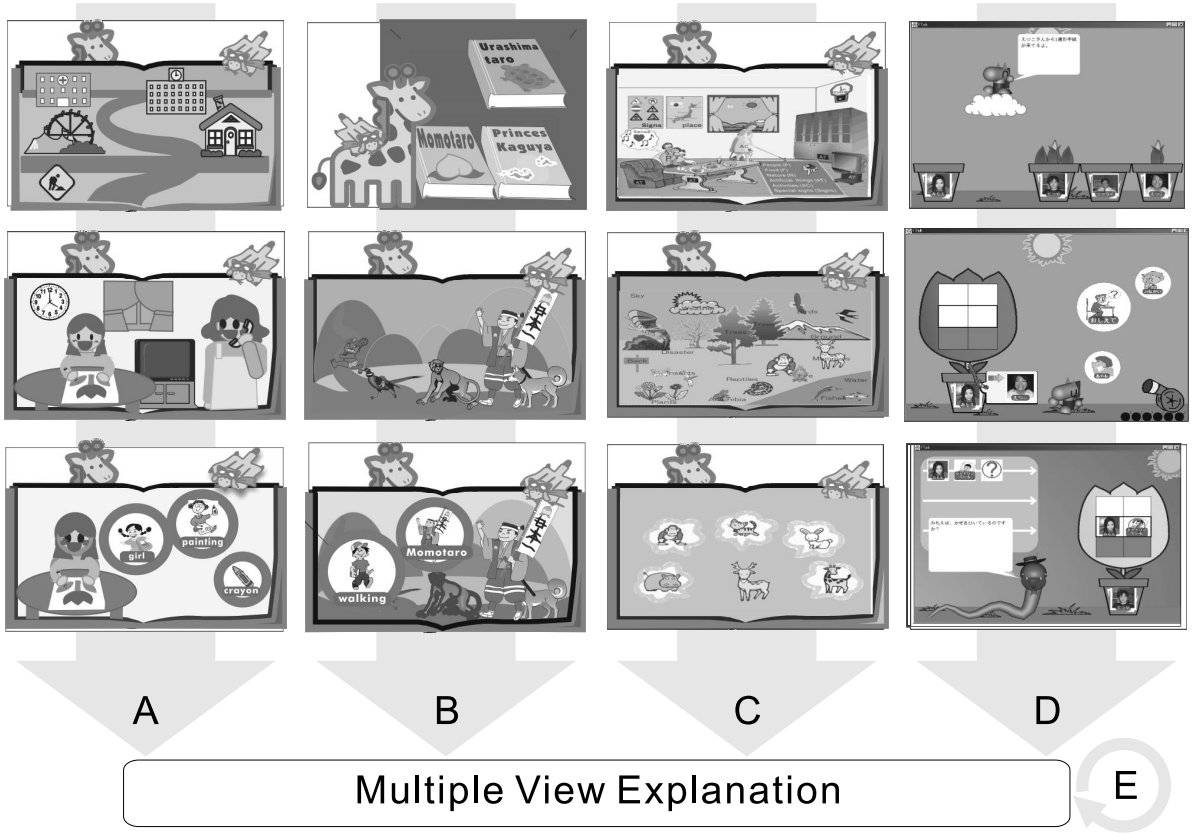

Fig. 5. Five Schemes to Access and Browse Multimedia Words

clicks, because our users usually have problems with typing on a keyboard.

In scheme **A**, the user can search for multimedia words based on a place, time and visual objects in the scene. The top panel is to select a place. After the selection, a series of scenes inside the place selected is shown like in the middle panel. In this case, "house" is selected, then a scene with visual objects such as a clock, a window, a TV set, a girl, and her mother is shown. After the visual object selection, a set of multimedia words related to the object selected is shown like in the bottom panel of column **A**. In this case, a girl is selected, then three multimedia words related to a hieroglyph representing "a girl is painting with crayons" are shown ("girl" in who cell, "painting" in what cell, and "crayon" in how cell). After that, the user can select an multimedia word.

In scheme **B**, the user can search for multimedia words based on a story, time and visual objects in the scene as in the case of scheme **A**. In this case, first, he selects "Momotaro's story (Japanese fairy tale)" at the top panel. Next, he selects "Momotaro (a man who has a flag marked with a peach)" at the middle panel. Then, multimedia words related to "Momotaro is walking" are shown in the bottom panel of column **B**.

In scheme **C**, the user can search for multimedia words based on about 150 multimedia word categories associated with nine super-categories. The top panel is to select one of the visual objects representing super-categories: people, activities, sense, food, time, place, nature, artificial things, and special words. After the selection, a set of categories associated with the super-category selected is shown like in the middle panel. In this case, "nature" is selected, then 14 categories are sown. After the category selection, a set of multimedia words classified into the category selected is shown like in the bottom panel of column **C**. This panel is to select an multimedia word.

Scheme **A-C** are developed as rather independent systems, but scheme **D** is dependent on other systems. The panels are not only for the user's searching operation but also for creating, sending, and receiving operations to communicate. It is to support teaching and learning multimedia words in a way of "communication." The user can select an multimedia word embedded in hieroglyphs received or to be sent as shown in the bottom panel of column **D**.

B. Browsing Multiple View Explanation

From a universal set of multimedia word frames, this prototype extract a set of view-frames and organized into ten panels. These explanations can be watched/heard non-stop or step-wise modes. In all panels, a basic view of the considering word are shown in the top-left part.

1) Short text and additional images. Short text provides a short description of the word meaning. Additional images provide a number of additional features related to space (global, regional, and local) and levels of

abstraction (abstract, medium, and concrete) characteristics. Fig.6 shows short text and additional images of "painting" in the left and right part, respectively.

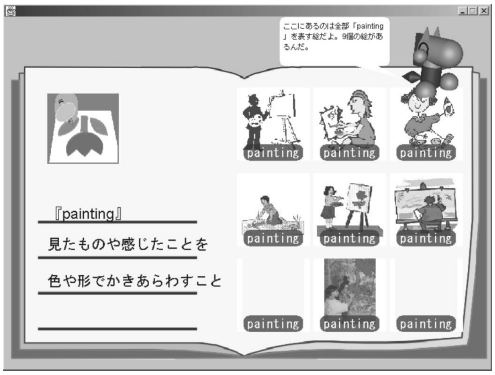

Fig. 6. Short Text (Left) and Additional Images (Right)

2) Animation, video and sound. They provide dynamic and audio features of the word.
3) A set of multimedia words. It provides "decomposed" features of an action or object.
4) Statistical data of the word. It provide which age, gender, month, and place are most favorable for the word use.
5) Antonym. It provides other multimedia words which means the opposite.
6) Homonym. It provides other multimedia words with the same pronunciation but with different meanings.
7) A set of multimedia words in the neighborhood. It provides neighboring words classified into the same categories (generally, about 150 multimedia word categories exist) of considering word.
8) Word association. It provides other multimedia words often used in multimedia hieroglyphs together with the considering word.
9) Examples of use. They provide multimedia hieroglyphs possessing the word. (Each multimedia word has its own set of multimedia hieroglyph cells and sub-positions where it can be used, and the multimedia hieroglyph examples are provided depending on these positions.)
10) Multi-modal explanation. It provides other features of the word by tactile sensations of haptic devices, motion of robots, etc.

In this way, our multimedia word dictionary panels clearly bring out semantic richness of the considering word towards self-explanatory features.

VII. Conclusion

A multiple view approach to teaching and learning multimedia words and five schemes to access and browse the words have been presented. They are supported by user friendly technologies: picture type panels and multiple interfaces. These technologies are to simplify the use of the system for a diversity of users. Picture type panels based on different metaphors are more natural for the user and to provide additional hints for the system use, and also multiple interfaces are to adapt the system to the user based on user/device profiles and a special statistics of user's activity in the system. (as a rule, the users can interact with the system only by simple mouse operations, because many of the users have difficulty to use keyboards). The development is requested and supported by real users. We implemented a prototype of a multimedia word dictionary subsystem, and will continue usability testing and improvement of the subsystem. Experiments with a prototype of the system and contacts with the real users indicate the evidence that our approach can provide a new aspects to design systems towards effective teaching and learning.

References

[1] B. Collis, "Information technologies for education and training," in *Handbook on Information Technologies for Education and Training*, ser. International Handbooks on Information Systems, H. H. Adelsberger, B. Collis, and J. M. Pawlowski, Eds. Springer, 2002, ch. 1, pp. 1–20.
[2] [Online]. Available: http://software.fujitsu.com/jp/product/indust/enikki/index.html
[3] [Online]. Available: http://software.fujitsu.com/jp/product/indust/kids/trace.html
[4] [Online]. Available: http://www.system-works.co.jp/˜ponpon/pon˙mini2.htm
[5] [Online]. Available: http://www.infinisys.co.jp/product/kazu˙pokke/index.shtml
[6] [Online]. Available: http://software.fujitsu.com/jp/product/indust/tama-box/
[7] "Fukushima prefectural special education center software library." [Online]. Available: http://www.special-center.fks.ed.jp/WebFPSECSL/fpsecslindex.html
[8] "Mac educational society." [Online]. Available: http://www.ceser.hyogo-u.ac.jp/naritas/mes/mes.html
[9] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, 1993.
[10] B. Shneiderman, "Universal usability," *COMMUNICATIONS OF THE ACM*, vol. 43, no. 5, pp. 85–91, 5 2000.
[11] D. R. Beukelman and P. Mirenda, *Augmentative and Alternative Communication*. Paul H. Brookes Publishing Co., Inc., 1998.
[12] American Speech-Language-Hearing Association (ASHA), "Report: Augmentative and alternative communication," *ASHA*, vol. 33, pp. 9–12, 1991.
[13] R. M. Johnson, *The Picture Communication Symbols Combination Book*. Mayer-Johnson, 1994.
[14] F. Carlson, *Picsyms Categorical Dictionary*. Baggeboda Press, 1985.
[15] S. C. Maharaj, *Pictogram Ideogram Communication*. The George Reed Foundation for the Handicapped, 1980.
[16] C. Wood, J. Storr, and P. A. Reich, *Blissymbol Reference Guide*. Blissymbolics Communication Institute, 1992.
[17] T. Hirotomi and N. N. Mirenkov, "Multimedia communication environment for children, handicapped, and elderly people," in *Proceedings of The 22nd International Conference on Distributed Computing Systems Workshops*, Vienna, Austria, July 2002, pp. 122–127.
[18] S. Tanimoto, "Visual languages for human communication: survey and design issues," VL'99 TUTORIAL I, Tokyo, Japan, 1999.
[19] N. Shedroff, *Information Interaction Design: A Unified Field Theory of Design*. The MIT Press, 1999, ch. 11, pp. 267–292.
[20] P. J. Bates, "Digital TV and video," in *Handbook on Information Technologies for Education and Training*, ser. International Handbooks on Information Systems, H. H. Adelsberger, B. Collis, and J. M. Pawlowski, Eds. Springer, 2002, ch. 10, pp. 139–149.
[21] "Assistive mouse technology for disabilities." [Online]. Available: http://www.abilityhub.com/mouse/index.htm
[22] "Immersion corporation - Technology devices overview." [Online]. Available: http://www.immersion.com/developer/technology/devices/index.php
[23] N. N. Mirenkov and T. Hirotomi, "Self-explanatory components: a new communication environment," in *Proceedings of The Seventh International Conference on Distributed Multimedia Systems*, Taipei, Taiwan, September 2001, pp. 188–196.
[24] K. Nelson, *Language in Cognitive Development*. Cambridge University Press, 1996.