A Concept Model for Computer-based Spoken Language Tests

Yijian Wu, *Fudan University, Shanghai, China*, *wuyijian@fudan.edu.cn*
Wenyun Zhao, *Fudan University, Shanghai, China*, *wyzhao@fudan.edu.cn*
Xin Peng, *Fudan University, Shanghai, China*, *pengxin@fudan.edu.cn*
Yunjiao Xue, *Fudan University Shanghai, China*, *yjxue@fudan.edu.cn*

Abstract

Spoken language education plays an important role in modern language courses. Computer-based spoken language test systems provide a convenient, efficient and reliable approach to test and assess individual spoken abilities. But most E-Learning systems concern more about written training and test. Existing models and systems do not completely meet special requirements of spoken language tests. In our study, we have examined distinctive features of spoken language tests and corresponding content resources. In this paper, we propose a concept model which is especially designed for representing, organizing, packaging and delivering E-Learning resources for spoken language tests. A Spoken-English-Test-Oriented (SET-Oriented) architecture is constructed based on this model. A family of commercial spoken language test systems is built for content authoring, online learning, tests and assessment, mostly simplifying the procedure of organizing a spoken language test.

1. Introduction

Spoken language tests are regarded more and more important in foreign language teaching and learning. Traditionally face-to-face conversations are used to assess learners' spoken language abilities. A typical situation is that a test taker or a small group of test takers sit in front of one or two examiners and respond to a series of questions. The examiners then give a score according to the test takers responses, or give several weighted scores on pronunciation, fluency, vocabulary and content.

It is reasonable when a small amount of people are to take such kind of spoken language tests. But problems emerge when the amount of test takers increases. Face-to-face conversation tests consume large amount of human resources and a long period of time in the procedure of test, assessment and data analyzing when tens or hundreds of thousands of test takers are taking the tests. The computer-based spoken language test system provides a well-designed platform for large-scaled tests and relieves most teachers.

In China, spoken language tests are included in some official foreign language tests in some areas. A large number of students are requested to take spoken English test in various examinations. For example, in the city of Shanghai, more than 90,000 high school graduates took a spoken English test before they took the college entrance examination in 2005. In the same year more than 400,000 middle school students and over 20,000 high school students took spoken English tests in Guangdong Province. The examinations, about 10 minutes each, are arranged within one or two days.

In this paper, we describe our approach to designing the computer-based spoken language test system. In Section 2, we discussed the research and industry background of tests and spoken language tests. In Section 3, some distinctive features of spoken-language-oriented learning are listed and discussed. Then a concept model for spoken language learning and its application are presented in Section 4 and Section 5.

2. Background

2.1. Standards and recommendations

Researches have been made on how to share large amount of learning resources across heterogeneous platforms. Computer-based Training (CBT) standards and recommendations were put forward decades of years ago. These specifications enable data exchange between different system/software vendors, to some extent. But because of all the complexities of learning objects and resources, it is not guaranteed that all learning objects or resources are exchangeable as long

Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services (AICT/ICIW 2006)
0-7695-2522-9/06 $20.00 © 2006 IEEE

as E-Learning systems meet certain recommendations or standards.

AICC Guidelines and Recommendations (AGR) are technical recommendations on CBT developed by Aviation Industry CBT Committee. [1] Originally, the AICC develops guidelines for aviation industry in the development, delivery, and evaluation of CBT and related training technologies. After the first recommendation AGR-002 is put forward in 1989, the guidelines help to build up a standard that can be followed by software vendors across multiple industries to improve system interoperability. Today there are actually 11 AGRs, each making a technical recommendation in a specific area in CBT technologies. AICC also certifies LMS, CMI services, CBT courses or Authoring system, confirming that the products conform to particular AGR(s) (esp. AGR-006 and AGR-010). Developers and software venders focus their interests on learning content, content aggregation and packaging, content exchange, content authoring, etc.

IEEE Learning Technology Standards Committee also works on learning content interoperability. Computer Managed Instruction (CMI) by working group 11 [2] and Learning Object Metadata standards by working group 12 [3] are impacting ones. The following SCORM partly takes these standards as references.

Sharable Content Object Reference Model (SCORM) [4] is a recommended reference model defined by Advanced Distributed Learning (ADL) group. In 2004 the second version of SCORM 2004 [5] is released. SCORM 2004 consists of three content books: Content Aggregation Model (CAM) book, Runtime Environment (RTE) book, and Sequencing & Navigation (SN) book, each updated separately. The three books describe how shared content objects (SCO) and independent assets are packaged, organized, distributed and interpreted. With SCORM learning resources are well organized into content packages, which can be transferred freely across different LMS or delivered to end users or learners.

SCORM provides a common tree-style data structure, a recommended organization method and a predefined runtime environment. It is now a widely applied reference model in various kinds of E-Learning systems.

There are on-going researches in E-learning standards and methods. For example, in [6], a new content organization model, extending SCORM, is proposed to increase flexibility and interoperability in content organization and sequencing. In [7], a structured method for requirement analysis in content authorizing in Small to Medium Enterprise is proposed, which is a valuable try for customizing special E-Learning needs in a specific application domain. An innovative multi-agent system supporting cooperative learning is presented in [8], where intelligent agent technology is applied in system construction bettering interoperability and automated ranking. There are also researches such as [9] [10] on modeling approaches to constructing general purpose e-learning systems. But none of them discussed spoken-language-oriented features in detail.

2.2. Practices and commercial systems

There are kinds of commercial Learning Management Systems (LMS) and relative systems mainly dealing with computer-based written tests in E-Learning area. Questionmark Perception assessment management system [11] is a web-based test and assessment system. It enables educators and trainers to create questions and organize them into surveys, quizzes, tests and exams, and supports multiple styles of questions, such as multiple choice, yes-or-no, and fill-in-the-blanks, etc.

However, most learning contents which conventional systems create or deliver are traditional multimedia resources. The structure and organization are not suitable for spoken language practices or tests. Content structure and training control mechanisms do not meet requirement of spoken language tests very well, as will be discussed later in Section 3.

There are still other attempts implementing spoken language tests. PhonePass system [12] developed by Ordinate Corporation is an impressive try. PhonePass system provides several kinds of spoken language tests. Each kind of test has different objectives. But all tests are based on voice recognition technology. No linguistic professional is required for answer assessment and evaluation. But there are limitations when applying this method. The test probes the psycholinguistic elements of spoken language performance rather than the social and rhetorical elements of communication. Test items are only designed to measure a speaker's fluency, listening, vocabulary, pronunciation and oral reading abilities, excluding content organization, rhetoric skills, or speech skills. Most questions are very easy for native speakers and proficient non-native speakers, which indicates that this kind of SET is only suitable for primary-level tests. This problem is also mentioned in technical reports of PhonePass system.

In order to organize large-scale spoken language tests efficiently, we need a computer-based platform to easily create and deliver learning materials designated for spoken language learning. We will begin with analyzing some distinguishing features of spoken language tests and then explain the detailed structure of

our concept model for spoken language tests in Section 4. In Section 5 we will take a quick review of the application of our model.

3. Spoken language test features

Spoken language learning can be different from traditional computer-based learning. It has some distinctive features in the following fields.

3.1. Authoring

In most cases exam contents are not produced by skilled professionals, who are familiar with e-Learning systems, but by ordinary teachers, who have a literary education background. Therefore, authoring systems should be comfort-looking and easy-to-use. It is not necessary for content authors to understand how our content package is structured or organized. All he or she has to do is to select the original multimedia assets, organize them in grouped questions, and then export the contents into a package.

There are diverse scenarios for spoken language tests. Usually a test item in a spoken language test follows the question-before-answer pattern. In some cases, however, the pattern is broken. It is possible that test takers are required to translate a clip of audio *concurrently* when the clip is playing. It is also possible that test takers are required to translate the clip *after* it is stopped. An authoring system should provide an easy-to-understand way of organizing the contents.

3.2. Assessment

The answers are in audio format. Most of the questions have open answers. More than one response is reasonable, no absolute right or wrong. Therefore there are various strategies to assess the answers. An answer file may be sent to more than one certified individual for assessment randomly. A third individual may be needed for a third review if the previous two hold quite different opinions on a certain answer audio file.

3.3. Examination

Spoken language tests are taken *simultaneously*. Similar to listening comprehension, an oral test has a fixed "pace" of questions. Students are not allowed to "browse" among all the questions but answer what is asked within a given time. Even different from listening comprehension, oral test takers are not allowed changing their answers to each question afterwards because there is only one (recording) chance offered to each answer.

In addition, it is possible that more than one test taker *cooperate* in an oral test. Two or more test takers form up a discuss group, and each test taker in the group is given some relative materials on a certain topic. Then test takers are requested to do some group discussion or debate. In this case an obvious challenge is that one content package should be designed for two or more test takers to participate in the test. Test takers in the same group may be instructed to listen to or look at relative but different materials simultaneously and behave accordingly. In the debate, a test taker should listen to his/her companion and respond to what was said.

4. System Architecture

Most general-purpose E-learning systems, including authoring tools, learning management system (LMS), or LMS services, provide a systematic platform for learners to interact with multimedia learning resources in the form of written practice or test. But in the case of spoken language learning, learners need to interact with the systems orally. The alteration of response style makes differences on content authoring, packaging and interpreting. Existing recommendations or standards have not clarified how to deal with such a response style. In our experiments and practices, we abstracted a spoken-language-oriented model to meet such requirements.

We have noticed that on one hand, spoken-oriented E-Learning is somehow similar to traditional E-Learning, where spoken-oriented learning contents share basic characteristics with general purpose E-Learning contents, such as structured courses, multimedia resources, various granularity of learning objects [5] and basic principles that an LMS interprets and presents a content package.

On the other hand, however, spoken-oriented E-Learning has distinctive features, as is discussed in Section 3. On an implementation perspective, the following are basic concerns to develop a spoken language examination system.

First of all, learners' responses are saved as audio files instead of characters. There are two ways to examine and evaluate the audio files: automatically and manually. The former method depends on voice recognition technology, which is now on its way to maturity but still not reliable enough to "understand" learners' responses as a human do, especially when the speaker holds an unfamiliar dialect. The latter method is what we concern more about. A software system

collects all learners' audio responses and delivers them to experienced authorized individuals for evaluation. Diverse strategies can be applied in the procedure of delivery in order to meet various evaluation requirements. In addition, there are no exact correct answers to each question. Therefore we need a well-designed assessment system for applying various delivery strategies in evaluating all these audio responses.

A second concern is interpreting the Content Package. When regarding only oral examinations, it is a comparatively simple task. There are no longer unpredictable and multiple ways of interpreting content sequencing or organization of the package. In a SET-Oriented system, a content package is simply designed for a test or a quiz. Thus we are able to concentrate ourselves on actual system analysis.

Spoken language tests can be regarded as one-way learning experiences. Once a test begins the test taker will have to follow up the "pace" of the test. Neither reverse nor pause is permitted in such examinations. In addition, each pair of question-and-answer is time-limited strictly for justice and for reducing interferences from neighbor test takers. All test takers in a same examination room takes the examination synchronously. Our purpose is to make it easier, more convenient and more effective for E-learning teachers to hold a spoken language test.

4.1. Overall structure

The Spoken Language Test System consists of four subsystems. They are an authoring system, an examination system, an assessment system and a data analysis system. E-learning teachers collect and choose from assets and multimedia resources, organize them via the authoring system, and produces content packages for the examination system and the assessment system. Most of the test items are open questions, which have no obviously correct answers. All answer data will be delivered to authorized teachers for assessment according to certain delivery rules. First of all, we concentrate on how to define and organize a content package for a spoken language test, which is the logical basis of the whole system.

4.2. A Conceptual Data Model

A content package is produced in the Authoring system and used in the Examination system and the Assessment system. Usually a package consists of three parts: package metadata, a manifest script and physical files.

Package metadata indicates information about the package itself, including the author, size, default purpose, etc. Metadata is associated with a particular examination. Therefore any package can have a default suitable examination, indicating which kind of examination the package is originally designed for. However a Spoken Language Test Management System can hold various kinds of examinations, given different examination settings. TMS settings may override default information of a package when an actual examination is taking place.

Physical files are actual contents of the package, including sound files, video files, picture files, and formatted text files. The file names are referenced by script commands. In order to improve efficiency, all files are packaged locally. Referenced files on a remote server are not permitted.

A **manifest script** is a sorted command list instructing how the package is organized and should be played. The script describes how physical files are structured and sorted, how much time each scenario should be on show, when and how long the system should record test takers voice. The script is translated differently in distinct systems. For example, in an examination system, a RECORD command is interpreted as actually recording test takers voice and saving sound data to files. In an assessment system, however, the same command is interpreted as playing the corresponding sound file and looping it.

A manifest script file can be represented in various forms. In our originally approach, the script takes the form of a database table. Each record in the table represents an action, such as showing or hiding a picture, playing a clip of sound or video, etc. Content structure information is also included. There are special designated fields (named *Level0, Level1,* and *Level2*) to hold a tree-like information structure. Obviously the aggregation depth is limited in this approach. As it is not a best way to store structure information in a table like this, we turn to use XML for describing content structures.

4.3. Detailed structures

The four systems are collaborating in the whole lifecycle of a spoken language test. They are built based on the conceptual data model described above. The most essential part is the script interpretation engine.

In authoring system, the engine works in preview mode. Content authors are able to take a look at what the content will look like. In preview mode, the engine interprets RECORD commands as doing nothing at all. Designated time in the script is displayed, and can be

skipped. For example, if a test item provides a test taker with 60 seconds to prepare, the PREPARE command followed by a time token '60s' will be interpreted exactly as what it is suppose to behave, except that the author may skip or terminate the 60-second preparation time if he/she wants to.

In examination system, the engine works in normal mode, which means the process of script interpreting should not be skipped or terminated by end user, i.e. a test taker. All RECORD commands are interpreted as doing actual recording.

In assessment system, the engine works in review mode. In this mode, a RECORD command is interpreted as playing corresponding sound file in loop mode. The answer file will be played again and again until it is given a score or marked with '*Deal with it later*'. TEXT command, SOUND command and VIDEO command are also interpreted alternatively. In some cases, question texts are hidden from test takers during the test. Test takers are only allowed to listen to the questions without seeing them. But when assessing answers, assessors should see the questions, while sound or video clips of questions may be suppressed optionally. Different behaviors of the system come from different interpretation of the script.

The three systems share the same script parser and the same engine core. The only difference is that the engine is configured in different working modes. It is really convenient and reliable for system development and test because we don't have to maintain three independent software components but reuse them in three relative systems.

In addition the parser can be rewritten to accept other formatted scripts.

4.4. An enhancement for group discussion

There are cases that more than one test takers cooperating in a spoken language test. In China, Spoken English Test of College English Test (CET-SET) adopts such a mode. Three test takers are randomly grouped and given topics to discuss or debate. Examiners observe each test taker's responses and give each test taker a score.

In computer-based spoken language test, such situation can also be realized. Besides listening to or reading the questions, preparing for responses and answering the question, one may listen to his participants' speeches. All we need to do is to make a little bit enhancements to the manifest scripts in our content packages.

In order to support computer-based group spoken English test, an additional parameter ACTIVESEAT should be attached. It comes with each command indicating to whom in the group the command is supposed to be sent. ACTIVESEAT has a value space of

- ALL (0), representing the command is targeting all participants in the group;
- Positive integer, representing which one in the group the command is targeting; and
- ALL-ISOLATED (-1), especially used in RECORD command, representing the RECORD sound is not broadcasted within the group.

Note that when ACTIVESEAT is set to ALL in RECORD command, it is exactly indicating that it is time for discussion. The examination system performs actual sound recording locally and broadcasts sound data in real time to the other participants. When ACTIVESEAT is set to a positive integer in a RECORD command, it means that test takers are required to speak in turn. When a test taker is instructed to speak, the other participants are able to hear what was uttering.

In this way, each participant in the discussion group may be instructed separately with the same content package, and interact with each other simultaneously.

5. Application and future work

We have already developed an SET-oriented product FDWS Spoken Language Test Management System. It has been applied in Spoken English Test of College Entrance Examination and in several universities in Shanghai, China. Two years of application helps us to gather information of spoken language test patterns. Up till now, our implemented test patterns include Reading aloud, Repeating, Quick responses, Talking about pictures, Talking about video clips, Video dubbing, Oral translations, Oral compositions, etc.

We have also developed a prototype of online spoken English practicing system. In the practicing system, not only test content need to be packaged and transferred, but also score information and sample record files need to be packaged. We are now working on a consistent and efficient way of defining and packaging all these learning resources.

The conceptual model presented in this paper does not conform to any other existing standards or recommendations. We are now working at making an extension of existing standards (e.g. SCORM) to fully support spoken language tests. From this point of view what we are doing now is just an experiment of resource sharing in region of spoken language tests.

6. Acknowledgements

The work presented in this paper is supported in part by the China 863 Research Program under Contract 2005AA113120; National Natural Science Foundation of China under Contract 60473061; and Shanghai Municipal Science and Technology Commission (China) under Contract 04DZ15022.

7. References

[1] AICC website, http://www.aicc.org

[2] CMI standard, http://ieeeltsc.org/wg11CMI/CMIdescription

[3] LOM standard, http://ieeeltsc.org/wg12LOM/lomDescription

[4] SCORM website: http://www.adlnet.org/scorm/index.cfm

[5] SCORM 2004 documents, website: http://www.adlnet.org/scorm/history/2004/documents.cfm

[6] Kwang-Hoon Kim, Hyuk-Jae Yoo, Hak-Sung Kim, A Process-Driven e-Learning Content Organization Model, In *Proceedings of the Fourth Annual ACIS International Conference on Computer and Information Science (ICIS'05)*, IEEE Computer Society, 2005, pp. 328-333

[7] Emma O'Brien, Timothy Hall, Training Needs Analysis: The First Step in Authoring E-learning Content, In *Proceedings of the 2004 ACM symposium on Applied computing*, ACM Press, 2004, pp. 935-939

[8] Leen-Kiat Soh, Hong Jiang, Charles Ansorge, Agent-based Cooperative Learning: A Proof-of-concept Experiment, In *Proceedings of the 35th SIGCSE technical symposium on Computer science education*, ACM Press, 2004, pp. 368-372

[9] Klaus-Dieter Schewe, Bernhard Thalheim, Aleksander Binemann-Zdanowicz, et al, A Conceptual View of Web-Based E-Learning Systems, *Education and Information Technologies*, 10:1/2, Springer Science + Business Media, Inc., 2005, pp. 81-108

[10] Paris Avgeriou, Symeon Retalis, Nikolaos Papaspyrou, Modeling learning technology systems as business systems**,** *Software and System Modeling*, Volume 2 Number 2, Springer-Verlag, July 2003, pp. 120-133

[11] Questionmark Corp. website: http://www.questionmark.com

[12] PhonePass SET-10 Test Description & Validation Summary, Technical Report, http://www.ordinate.com/pdf/SET-10_Test_Desc_Validation.pdf