ISCA Archive
http://www.isca-speech.org/archive

Speech and Language Technology in Education (SLaTE2009)
Wroxall Abbey Estate, Warwickshire, UK
September 3-5, 2009

Rainbow Rummy: A Web-based Game for Vocabulary Acquisition using Computer-directed Speech

Brandon Yoshimoto, Ian McGraw, Stephanie Seneff

MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar Street, Cambridge, MA 02139

yoshimoto@alum.mit.edu, imcgraw@csail.mit.edu, seneff@csail.mit.edu

Abstract

This paper describes a new on-line game we have developed which allows learners of Chinese or English to practice speaking in a communicative setting. Game play resembles gin rummy or Mah Jong, and is intended to be sufficiently engaging to invite persistent replay. Students compete in a social game against other students at remote settings, or they can play against a robotic partner. A user study was conducted on 16 students of Chinese, to assess whether a configuration that utilizes speech recognition is as effective for learning vocabulary as a configuration that only requires the student to listen. Results show that the vocabulary learning gains averaged across subjects were greater following use of the speech-enabled version of the game compared to the listening-only version.

1. Introduction

One of the most difficult aspects of learning a foreign language is the daunting task of vocabulary building. Most students who are serious about acquiring proficiency devise methods such as flash cards to help with this task. Flash cards can either be physical cards or virtual ones made available on the computer and even on-line. However, simply studying flash cards is a tedious task, and does not require the student to speak the words or use them in any meaningful way.

Our group at MIT has long been interested in developing on-line speech-based games to help a student learn a foreign language in a more entertaining way than studying flash cards. One of the first games we developed is called *Word War* [1], which allows students to select a personalized vocabulary list for study, and then to play a simple matching game, aligning pictures with their corresponding names by issuing spoken commands to the computer. Word War can also be configured as a social game, where two players compete in a race against time for a shared set of slots to be filled.

A potential problem with Word War is that, as a game, it is not particularly inspiring. We thus sought to design another game that would share many of the features of Word War, but employ more compelling game dynamics. In this paper, we introduce Rainbow Rummy as an attempt to address this need.

2. Related Work

A number of speech-enabled games for language learning have been created in the past [2, 3, 4]. The DEAL system [2] allows the student to role play a shopper and negotiate through spoken conversation with a virtual shopkeeper to purchase an item displayed on a shelf. The Saybot system [3] sets up multiple-choice branch points embedded within multimedia content where Chinese learners of English can practice oral communication, and receive feedback on the quality of their spoken productions.

The Tactical Language Training System [4] is based on a virtual world that allows the student to interact with various “socially intelligent virtual humans” through spoken dialogue in a detailed 3D environment. It is intended to teach Arabic to military personnel, and includes an interactive story-based game with a task-based focus. A key design focus was that “artificial intelligence techniques used must support the learning-promoting features of a game; otherwise they may be superfluous or even counterproductive.” The developers strove to create an environment that felt like a game, but still made the learning central to the application. Though expensive, the program has been successful, and is used actively by the U.S. Armed Forces. Its adoption demonstrates that speech-enabled games can be designed to promote language learning.

3. Background

Currently at the MIT Spoken Language Systems group, we are developing several applications that allow students to use speech technology to learn Mandarin Chinese [5]. One such system, Chinese Cards,[1] allows users to log in and add words to personal sets of vocabulary. The system supports storage of English, pinyin (romanized Chinese), Chinese characters, and associated photos in an integrated publicly shared resource.

Using this system for vocabulary management, Word War was created to incorporate the user-added vocabulary sets into an interactive speech-driven game. The objective is to match pictures of vocabulary words by speaking commands to place matching cards into their correct slots. The intent is for the user to learn vocabulary indirectly while speaking commands, rather than through explicit memorization. Recognition feedback is provided as objects highlight and move on the screen in reaction to user commands.

Word War provides an alternative to flash cards as a method for vocabulary building using spoken language. By providing dynamic construction of speech-enabled games based on user-entered vocabulary, Word War demonstrates a customizable, available, and practical system for learning.

Despite this work, creating a good speech-enabled system for language learning is still an on-going research topic. One question focuses on the scope of the domain. It should be large enough to provide an interesting experience for repeated use, yet small enough to provide adequate recognition performance.

Word War takes the approach of limiting user commands to certain carrier phrases, where vocabulary can be substituted

[1] www.islands.csail.mit.edu/chinesecards.

in particular positions. The limited number of allowed phrases is particularly useful for improving recognition in the face of target users who are non-native speakers of Chinese.

A second concern is for systems to be designed so that they can remain relevant for future use. It is important for systems to be customizable such that their educational value can survive many repeat visits. The combination of Chinese Cards and Word War essentially places no restriction on the actual utterances that can be made using the game, allowing it to remain relevant so long as there are new words to learn.

Finally, finding the right balance between an entertaining game and an educational one is difficult. On the one hand, a game which is merely translated into a foreign language may be fun to play, but usually will not allow the user to learn *incrementally*. A learner may become overwhelmed in an environment that is not intentionally designed for non-native speakers.

Word War satisfies most of our motivating design considerations, but lacks a strong sense of entertainment value. There is a fun aspect to matching pictures using a novel speech interface, but the task of matching itself is not typically considered to be enjoyable.

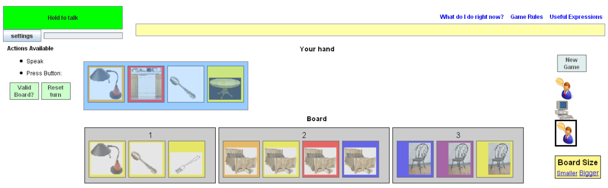


Figure 1: Screenshot of Rainbow Rummy

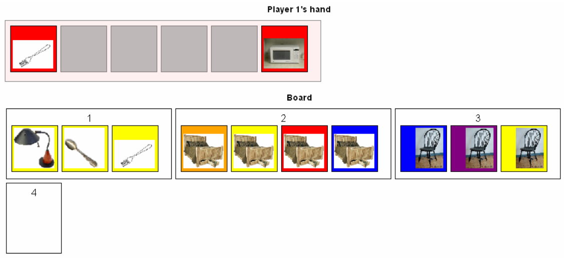
4. Game Description

Our new game, Rainbow Rummy, makes use of much of the same underlying technology as Word War. Like Word War, it is integrated into our on-line flash cards framework, and accessible from the Chinese Cards web site. Thus, students can select a personalized vocabulary list for study, including creating a new set of cards by searching an on-line dictionary and a collection of Yahoo images for appropriate content.

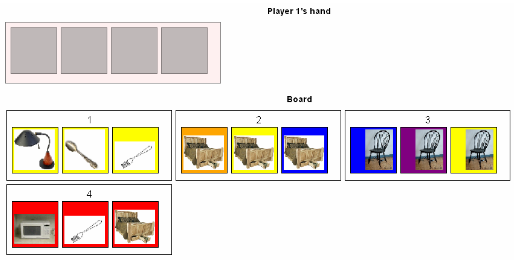
Rainbow Rummy, however, provides a more natural game dynamic than Word War by involving the user in a turn-based session against another computer or human player in a format similar to that of traditional card games. For each turn, the game layout shows the current hand and the board configuration, as illustrated in Figure 1. Each card is uniquely identified by its color and vocabulary item. The goal of the game is to get rid of all the cards in one's hand before another player does by moving cards from the hand into slots on the shared board below. The slots below must all contain three or more of one vocabulary item or three or more of one color to be a valid configuration. In order to finish a turn, the player must draw additional cards until he/she has at least one valid move. As the game progresses, more and more cards are played on the board.

4.1. Game Play

The game is intentionally designed to provide an intellectually challenging experience, so as to engage the student and encourage them to persist in playing over extended time periods and repeated episodes. In order to create sets by color or by image, the student can create slots and move cards from his hand onto the board below. What makes the game especially interesting is that the player is free to "steal" any cards from pre-existing sets



(a) The computer decides to move three red cards into a new slot, stealing a card from slot 2



(b) The computer finishes its turn

Figure 2: Example of a complex move in Rainbow Rummy

on the board, as long as all sets have at least three cards in the final configuration. Thus, for example, if the hand contains two red cards (Figure 2(a)), a third red card could be stolen from slot 2 containing four "bed" cards, and placed in a new slot (Figure 2(b)). Now the two red cards in the hand can be added to form a set of reds. Furthermore, suppose that the yellow bed had originally been in slot 1 (four yellows) instead of slot 2 (four beds). A play is still possible through a multi-slot transfer, stealing the red bed from slot 2 as before, then fixing its deficient count by stealing the yellow bed from slot 1 and placing it in slot 2.

We implemented an AI planner to execute the computer's turn in the game. The planner is capable of taking advantage of situations such as the ones above, although there are more complex board manipulations that are possible but beyond the current level of expertise of the AI planner. However, such opportunities could be exploited by a skilled human player.

When it is the student's turn, the cards of his/her hand are shown face up, and the student can create slots and play cards from the hand to the board (or from one slot on the board to another) by communicating the instructions to the computer using spoken language. For example, if a student says "bǎ hóng sè de chuáng fàng dào dì èr gè gé zi", the "red bed" card will move from its current location into slot two. Recognition hypotheses are processed incrementally to provide visual feedback of all the red items being selected, followed by selection of the bed, and finally the movement of the card, *while the user speaks*.

When it is the computer's turn, the computer issues spoken instructions using synthetic speech, and the student manipulates the cards according to the computer's verbal commands. The identity of the cards is revealed only as they become available for play. The student can listen carefully to the synthetic voice and use it as a guideline for his/her own speech. Through the combination of listening and speaking modes, the student is able to practice both aspects of communication in one game.

The game can be configured with two or more players, including a mixture of human and computer players, although the most common modes are two-party human-computer or two-party human-human. The system can be configured for learning either English or Chinese. Students can even compete against one another while studying different languages and using different vocabularies. When it is the opponent's turn, the computer simply plays out the cards of the opponent's hand, mapped to the vocabulary and language of the student who is idly watching, while speaking out loud the moves as it executes them.

4.2. Architecture

The game makes use of the WAMI (Web-Accessible Multimodal Interface) software [6] to configure a client-server solution where most of the computation is performed remotely. A Java applet captures audio at the web page and transmits it to the SUMMIT speech recognizer [7] running remotely. The recognizer acoustic models were trained on native Mandarin speech. The recognizer language model is configured on the fly as soon as the game begins, based on the vocabulary set selected by the user. It is specified via a simple JSGF grammar, allowing various ways to express the possible moves of the game. The vocabulary includes colors, numbers, and simple phrases such as "draw a new card," as well as the personalized vocabulary items. Meta tags attach meaning to the phrases in the recognized sentences, and the dialogue manager interprets the meaning and blindly executes the appropriate moves as they were understood. Once the student says, "finish turn," the AI planner evaluates the board configuration. If the board is invalid, it is restored to the configuration before the turn, and the user is given a second chance to complete their turn. In the event of irrecoverable recognition error, the student also has the option to abort and start over on the turn.

5. User Study

To assess whether Rainbow Rummy can help a student acquire knowledge of vocabulary, we designed a user study that involved 16 learners of Chinese playing the online game. The students recruited had a range of abilities, but most had at least one year of experience studying Mandarin Chinese. On average, the students were familiar with around half the words used in the study. We were particularly interested in assessing whether the fact that the student must speak to the computer to execute their turn was a useful device for learning new words. Each student played against the computer in two different modes over a two week period.

Rod Ellis, a leading researcher in Second Language Acquisition, performed an experiment in 1999 for evaluating the effect of various incidental learning techniques on vocabulary acquisition [8]. Ellis compared three environments for subjects as they were faced with the task of placing vocabulary items into various positions. He found the learning gains in the spoken version of the task out-performed those of the groups that were required only to listen, although he attributes these gains to other factors. We designed a similar study to measure incidental learning gains between two environments where cards are manipulated, only one of which enables user speech.

In order to create a baseline experiment that was as close as possible to the original game but did not require the user to speak, we configured a "listening-only" version, where the user was free to manipulate their cards by simple drag-and-drop operations with the mouse. However, every time they made a move, the computer spoke, in Chinese, a comment describing the move: "You put the red horse into slot number two." Thus, the player *heard* essentially the same content that they would have been required to speak in the "speaking" configuration. Both configurations behave the same way during the computer's turn: the computer states its goals, and the student is required to manually execute them.

Because a typical game in Rainbow Rummy makes use of only eight unique images, we decided to modify the game for the user study, to substantially increase the number of words being learned. The "match by image" paradigm was replaced by a "match by category" rule. Thus, the game might include a chair, a table, and a bed, which would form a set under the class "furniture." We added a symbol in the upper right hand corner of each card which identifies its category membership, to ease the cognitive load of finding a matching set. Additionally, this modification serves a useful pedagogical role by placing words in "semantic fields", requiring users to also think about the meaning of the items as they plan their moves.

We selected a set of cards that contained vocabulary items drawn from eight categories (animals, vehicles, furniture, kitchen items, food, electronics, clothing, and garden/plants). There were six unique items in each category for a total of 48 vocabulary words being learned by the student. These were divided pseudo-randomly into two sets of 24, and each subject was assigned a specific subset for the "speaking" version and the complement for the "listening" version. In this way we could minimize variability due to subject, as each subject learned half the words in each of the two versions. Each game involved a deck of 48 cards, two cards for each vocabulary item. Half the subjects used the "speaking" version in the first game, and the other half began with the "listening" version.

To obtain a measure that correlates with learning, we asked the subjects to play a modified game of Word War both before and after each game of Rainbow Rummy. Thus, each subject played a total of four Word War sessions, where the pre-test occurred immediately prior to playing a version of Rainbow Rummy, and the post-test occurred at least four days later, in order to reduce short-term memory effects.

The 24 words used in each Word War game were the same as those used in the corresponding Rainbow Rummy game. The game was split into four rounds of six words each. Each round was configured as shown in Figure 3, where the student is required to place six (the items being tested) of the 12 objects displayed on the bottom into the correct slots above by following instructions spoken by the computer.

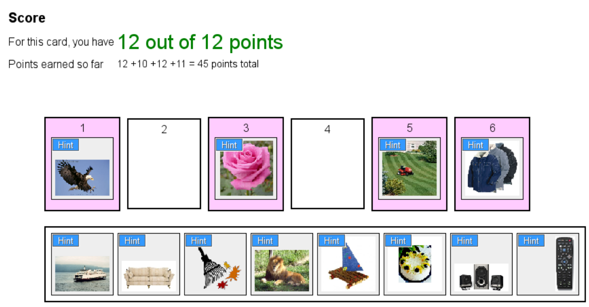


Figure 3: Version of Word War used as pre- and post-tests

A scoring scheme was designed both to incentivize the student and to quantify knowledge. By placing a word correctly, the student gains 12 points. However, every time they use a hint or make an incorrect move, a point is deducted. This encourages students to think more carefully about their understanding of a word before making a move. Furthermore, it gives the ses-

Table 1: Rainbow Rummy game statistics

Subject Type	Game version	moves by user	moves by computer
Listening-only first	Listening-only	191	69
	Speaking	171	89
Speaking first	Speaking	249	102
	Listening-only	174	105

sions a game-like feel with an objective to seem less like a test. Finally, the score serves as a useful metric to assess the subject's mastery of a particular word. We felt that this metric would provide richer opportunities for analysis in comparison to binary indicators of word knowledge that would result from a more traditional matching test. Moreover, given the unsupervised nature of our experiments, it was decided that a test disguised as a game would be less likely to engender cheating.

5.1. Results

To assess learning gains, we used the Word War scores as our metric of vocabulary knowledge in Equation 1 across paired pre- and post-test sessions.

$$G_v = (s_2 - s_1)/(MaxScore - s_1) \quad (1)$$

where v is either "speaking" or "listening" version, s_1 is the pre-test Word War score, and s_2 is the post-test Word War score. The maximum score possible ($MaxScore$) is 288 (24 words x 12 points per word).

For each subject, we then calculated this learning gain for both versions of Rainbow Rummy. Using two-tailed t-tests, we found that the learning gain, averaged over all subjects, for the speaking version was significantly greater ($p < 0.05$) than that for the listening-only version. To gain further insights, we examined the average learning gains for various meaningful subject subsets (See Figure 4). One variable was whether the student's first game was the speaking version or not. While both groups showed essentially the same relative benefit from speaking as the pooled data, those who began with the listening-only version had higher *overall* gains, although not statistically significantly higher (0.11 p value). From a training standpoint, it makes sense to arrange the game ordering based on degree of difficulty, since the use of speech carries additional cognitive load. As seen in Table 1, for those who saw the listening-only version first, fewer moves were actually performed by the user in the speaking version, yet the learning gain was higher for speaking than for the listening-only version, suggesting increased gains per move through the use of speech.

As part of our survey at the end of the study, we posed the question: "Which version of Rainbow Rummy did you enjoy more? The speaking version, or the listening version?". Roughly half of the subjects preferred speaking, while the other half preferred listening. Grouping these subjects according to preference, we found that most of the additional learning gains were realized by those who preferred speaking. While the listening-preference group also learned more from speaking mode, the small margin was not statistically significant.

6. Conclusion

In this paper, we have presented a new web-based game, Rainbow Rummy, which allows students of Chinese or English to practice speaking in a communicative and purposeful setting. Our hope is that the game strategy is sufficiently challenging to engage students in multiple episodes while avoiding boredom,

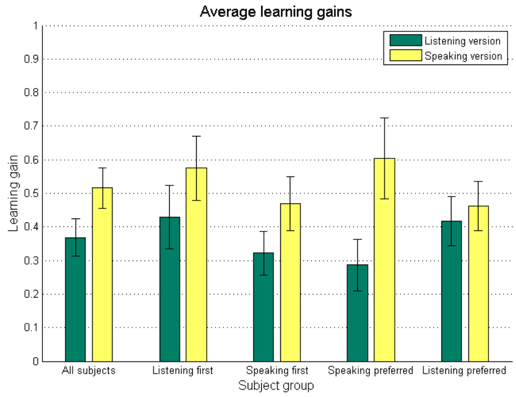


Figure 4: Learning gains averaged over subjects shown with 95% confidence intervals.

which is crucial to the goal of repetition as a means to solidify the learning process. In a carefully designed user study involving 16 learners of Chinese learning 48 vocabulary words, we demonstrated significantly greater learning gains when students were required to speak in Chinese to execute their turn in the game, as compared with a baseline system where the system parroted their manually executed moves in a synthetic voice.

7. Acknowledgements

This work was supported by the Industrial Technology Research Institute (ITRI) in Taiwan.

8. References

[1] I. McGraw, B. Yoshimoto, and S. Seneff, "Speech-enabled card games for incidental vocabulary acquisition in a foreign language," *Speech Communication,* Special Issue on Applications to Language Learning*, to appear*, 2009.

[2] P. Wik, A. Hjalmarson, and J. Brusk, "DEAL a serious game for call practicing conversational skills in the trade domain," in *Proc. of SIGSLaTE*, 2007.

[3] S. Chevalier, "Speech interaction with saybot player, a call software to help chinese learners of english," in *Proc. of SIGSLaTE*, 2007.

[4] W. Johnson, H. Vilhjalmsson, and S. Marsella, "Serious games for language learning: How much game, how much ai?" in *Proceedings of the 13th International Conference on Artificial intelligence in Education*, Los Angeles, CA, 2005.

[5] S. Seneff and C. Wang, "Web-based dialogue and translation games for spoken language learning," in *Proceedings of the Speech and Language Technology in Education (SLaTE) Workshop*, Farmington, PA, Oct. 2007.

[6] A. Gruenstein, I. McGraw, and I. Badr, "The WAMI toolkit for developing, deploying, and evaluating web-accessible multimodal interfaces," in *Proceedings of ICMI*, Chania, Crete, Greece, 2008.

[7] J. Glass, "A probabilistic framework for segment-based speech recognition," *Computer, Speech and Language*, pp. 137–152, 2003.

[8] R. Ellis and X. He, "The roles of modified input and output in the incidental acquisition of word meanings," *Studies in Second Language Acquisition*, vol. 21, no. 2, pp. 285–301, 1999.