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

by

Brandon Yoshimoto

S.B. EECS, M.I.T., 2008

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
[June]
May 2009

Author . . . . . .
Department of Electrical Engineering and Computer Science
May 22, 2009

Certified by . . . . . .
Stephanie Seneff
Principal Research Scientist
Thesis Supervisor

Accepted by . . . . . .
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

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

by

Brandon Yoshimoto

Submitted to the Department of Electrical Engineering and Computer Science
on May 22, 2009, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Acquiring vocabulary in a foreign language is a long process which often involves the use of flashcards or cycling through long word lists for memorization. While many students learn effectively in this way, research at the Spoken Language Systems Group (SLS) has been exploring alternative methods which make use of speech recognition and generation technology. In this thesis, I designed and implemented a speech-enabled game to aid learners of Mandarin Chinese with this task of vocabulary acquisition. Our approach is, customizable, allowing user control of the vocabulary words, web-based, providing potential for widespread use, and game-based, engaging the user in an interactive session with a computer or another human player. We evaluated the feasibility of the game as a web-based CALL (Computer Aided Language Learning) system through its deployment for use by the general public. Secondly, we ran a study to measure the effect of computer-directed user speech on vocabulary acquisition in comparison to a system which only provides listening practice.

Thesis Supervisor: Stephanie Seneff

Title: Principal Research Scientist

Acknowledgments

First and foremost, I'd like to thank my advisor, Stephanie Seneff, for all of the insight and guidance she has provided throughout this project. For technical aspects, I'd like to thank Alex Gruenstein and Ian McGraw for development of the WAMI toolkit upon which Rainbow Rummy was built. The speech technology itself can be attributed to the effort of many previous researchers at the Spoken Language Systems group. For the WISTON synthesizer, I'd like to thank the Chinese Academy of Sciences. I'd like to thank Ian for creation of the Chinese Cards system and the countless hours he has spent in providing me with technical help and guidance for the direction of this project. I'd especially like to thank my family, particularly my parents, for their constant love and support, and for allowing me this opportunity to participate in research at M.I.T.

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

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

1 Introduction **13**

1.1 Motivation 13

1.1.1 Resources for language practice 14

1.1.2 The need for vocabulary building aids 16

1.1.3 Why Chinese? 18

1.2 Chapter Outline 18

2 Background **19**

2.1 Previous Work 19

2.2 Discussion 23

3 Rainbow Rummy **27**

3.1 Overview and Functionality 27

3.1.1 Speaking mode: Human player's turn 29

3.1.2 Listening mode: Computer player's turn 30

3.1.3 Finishing a game 31

3.2 Design 31

3.2.1 System Details 32

3.2.2 Game Strategy and AI Planner 33

3.2.3 Justification 35

3.3 Multi-player version 37

3.3.1 Description 37

3.3.2 Cross-language interaction 41

3.4 Deployment 42

4 User Study **45**

4.1 Previous work 45

4.2 Design and Setup 47

4.2.1 Systems used in the experiment 47

4.2.2 User selection 51

4.2.3 Task sequence 53

4.3 Launching the Study 57

4.4 Results 58

4.5 Summary 66

5 Final Remarks **67**

5.1 Conclusion 67

5.2 Future Work 67

A User Study Management **69**

List of Figures

2-1 Screenshot of the Tactical Language Training System 21

2-2 Screenshot of the Chinese Cards system 22

2-3 Screenshot of the Word War game 23

3-1 Screenshot of Rainbow Rummy 28

3-2 Mapping traditional cards to Rainbow Rummy tiles 28

3-3 Example sequence for a complex move 34

3-4 Two players select options for a multiplayer game 38

3-5 A sample multiplayer game sequence 39

3-6 The matching process for a multiplayer game 40

3-7 Geographic distribution of visits over 4 months 43

4-1 Screenshot of modified Rainbow Rummy using multiple categories . . 49

4-2 Screenshot of the modified Word War game for the study 50

4-3 Screenshot of the audio test 51

4-4 Geographic distribution of subject locations 53

4-5 "Speaking" vs "listening-only" learning gains based on various subject subsets 60

4-6 Learning gain versus number of hint clicks 62

4-7 Learning gains versus number of moves by user 63

4-8 Learning gains versus Chinese experience 64

A-1 Management system: screenshot of forms page 69

A-2 Management system: screenshot of tasks page 70

A-3 Management system: screenshot of groups page 70
A-4 Management system: screenshot of users page 71
A-5 Management system: screenshot of assignments page 72

List of Tables

3.1 Phrases used in Rainbow Rummy 30

3.2 Statistics of user interaction with deployed version of Rainbow Rummy 43

4.1 Sequence of tasks for a subject seeing the "speaking" version first . . 54

4.2 Sequence of tasks for a subject seeing the "listening-only" version first 55

4.3 Learning gains for the "speaking" versus "listening-only" system . . . 59

4.4 Rainbow Rummy game statistics 59

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

This thesis examines the use of computer-directed speech in an online game for vocabulary acquisition of Mandarin Chinese. The system, called Rainbow Rummy, frames vocabulary acquisition in an interactive environment that requires both speaking with and listening to a computer to simulate a card game with another player. Our goal was to create a game which demonstrates how automatic speech recognition in limited context can be used to help students learn. We present an evaluation of the system through a user study in which we attempt to quantify the learning gains that can be achieved by enabling speech as an input mode in a web-based language learning system.

1.1 Motivation

Foreign language proficiency has received much attention as an area in which the United States is lacking. In Europe, it is estimated that more than 50% of adults speak a second language fluently, while in the U.S, only 9% can claim the same proficiency [12]. The figure is not surprising, as foreign language enrollment has continually been shown to be disproportionately low in comparison to other countries. Contrast the situation in China, where the U.S. Department of Education estimates 200 million children studying English, with the situation in America, where only 24,000 children are studying Chinese [9]. The difference is striking.

At the same time, statistics show that foreign language enrollment in the U.S. has been on the rise over the past decade. According to a report released by the Modern Language Association in 2006, enrollment in foreign language classes at American colleges has increased 13% since 1992 [6]. Mandarin Chinese, promoted by the Department of State to be a "critical" language to national security and cultural understanding [10], has seen a rise in enrollment from 28,456 in 1998, to 34,153 in 2002, to 51,582 in 2006 [6]. If this trend continues, we will undoubtably be faced with increasing demand for resources to aid in foreign language learning.

In the following, I will discuss current resources available to learners, the role that computers have played, and the limitations of current solutions for providing speaking practice (Section 1.1.1). Secondly, I will discuss the motivation for focusing on vocabulary acquisition and the need for speech-enabled resources in this domain (Section 1.1.2).

1.1.1 Resources for language practice

Current resources

In learning a second language, enrolling in classes is the primary resource for most people. However, many learners find that classroom instruction alone does not lead to fluency. Typically, a student must also spend significant effort outside the classroom to gain an acceptable level of proficiency.

To get this practice, the learner may hire a private tutor or talk with a friend who is fluent in the language. Some learners invest in supplementary audio and text resources. Pimsleur [11] is praised as an effective audio resource. Language learning software, such as Rosetta Stone [14], aims to provide an immersion experience at home when studying abroad is not an option.

Increasingly, the Internet has become an extremely useful resource for language learners, providing text, audio, and video files for browsing at the learner's convenience. Online chat rooms and forums give learners the opportunity to communicate with native speakers from around the world. Video-conferencing has been applied

to language learning to allow speakers in different locations to practice their target language with another native speaker as if they were meeting face-to-face.

Limitations

While these solutions have proved useful to language learners, they have several limitations in their ability to provide students with speaking practice.

In many cases, foreign language classrooms are too large to allow sufficient speaking time for each student. Many language curricula assign 15 to 20 students per teacher. This high student-to-teacher ratio, along with limited time for in-class activities, has left the student with little opportunity to speak in class. Additionally, many students face anxiety when speaking in the classroom and are more conscious of making mistakes, further limiting its practice.

Outside the classroom, hiring a private tutor can be expensive and coordination of schedules places limits on availability. Video-conferencing may be free, but still suffers from the same coordination problem, especially if the language partner is in an inconvenient time zone relative to one's own.

Some learners don't seek any of these solutions, but rather, practice speaking by reciting material to themselves. The disadvantage of this approach, however, is that the learner receives no feedback to confirm if they are correct. Teachers recognize that feedback is important and attempt to address the problem by assigning spoken recording homework. Typically, the teacher then provides comments on the submitted recordings and returns it to the student to review. Unfortunately, this feedback is not immediate, and as a result, many students do not take time to review and internalize suggestions.

The question still remains: why do learners even need to practice speaking? For one, speaking is known in the teaching community as one of the four major skills of language learning, along with the skills of listening, reading, and writing. Given its place among this set, it is a major part of a full language curriculum which should be cultivated.

Secondly, speaking is recognized by many second language acquisition researchers

to be an important skill not only for communication, but also for the *acquisition* of language [15]. As every individual learns in their own way, it is reasonable to believe that some may learn more easily or faster through spoken language. If we intend to capture these benefits, the challenge then becomes to provide opportunities for such practice which deal with many of the limitations to current solutions as discussed above.

Incorporation of automatic speech recognition (ASR) in web-based systems for language learning shows potential for addressing these needs. Given the need for automated evaluation of learner speech and shortage of available human evaluators, ASR shows promise in this application. Speaking with a computer cannot fully replace the subjective experience of speaking with another human, but there is potential in its use for dealing with limitations to most current solutions by providing reasonably correct feedback on user speech that is free, immediate, and always available. Additionally, beginning learners have the ability to practice without feeling embarrassed about making mistakes, as the computer will not judge in the same way as a human evaluator might.

1.1.2 The need for vocabulary building aids

Within the realm of language learning tools, aids for vocabulary building are of particular importance. Especially in the case of a language much different from English, such as Mandarin Chinese, a solid vocabulary provides enormous benefits for communication. It is estimated that knowledge of 5,000 words in a foreign language is needed in order to understand 98% of printed matter [11], and learning this set requires a significant amount of time.

Ideally, such vocabulary could be learned naturally in real-world situations, much as how infants learn their own native language. However, for the second language learner, this kind of communicative experience is often not available, for many of the reasons mentioned in the previous section.

In the face of a large set of unknown vocabulary, learners must typically take personal time to memorize words using tools such as flash cards - either as physical

cards or online. Flash cards serve an important role in learning, and are excellent tools in their ability to be easily created for any set of words or concepts the user decides [11].

At the same time, they have several limitations. A characteristic of flash cards is that they present new words in written form. The learner typically looks at a word, repeats it mentally until memorized, then cycles to the next word. The limitation of this approach is that it only drills the student in this one procedure. The approach does not require the student to speak or use words in any meaningful way.

Many students do in fact practice speaking by repeating words as they flip through flash cards [8]. This can be a very effective technique. However, the learner receives no feedback for correctness or comprehension of their speech. While it is not clear if the benefit derived from speaking correctly with feedback is beneficial to vocabulary acquisition, we acknowledge that the current tools for providing such an alternative are few.

A second drawback to flash cards is that they are simply not much fun. To memorize anything takes time, but it would be even better if this time could be spent in a more entertaining and engaging environment. Flash cards are used as tools for intentional learning, where vocabulary memorization is the explicit goal of the activity. It would be desirable to also have *incidental* means for vocabulary building, where learning naturally occurs in the process of performing some other, more entertaining task.

Games provide this kind of environment and have the potential to appeal to people of all ages. Consider the case of Solitaire and Minesweeper, two popular Microsoft games that come with every PC. Users spend a surprisingly large amount of time playing such games with no personal benefit other than entertainment. It was estimated that in 2003, nine billion person-hours were spent playing Solitaire on the computer - a duration long enough to construct 500 Panama Canals [16]. If learning could be incorporated into such games, the benefits for language learners could be immense. A challenge then arises to provide tools for vocabulary building which retain the customizability of traditional flash cards, yet are entertaining and

allow practice of speech at the same time.

1.1.3 Why Chinese?

Many students in the United States and elsewhere are recognizing knowledge of Mandarin Chinese as a marketable skill for their future. We believe that efforts should be made to provide computer learning systems to meet this growing demand. Secondly, Chinese is a language which has very different linguistic roots from English. The U.S. Department of State designates Chinese as a level 4 language, meaning, it is considered among the most difficult languages for native-English speakers to learn. Given this difficulty and the many differences between sound units in English and Chinese, we believe it is important for tools to be developed which provide students with speaking practice, especially if there exists a link between speech and language acquisition.

1.2 Chapter Outline

To address these motivations, we developed a prototype card game for learning Chinese vocabulary called Rainbow Rummy. In Chapter 2, I will discuss previous work in applying speech recognition technology for applications of language learning. Next, in Chapter 3, I will introduce Rainbow Rummy through its design, capabilities, and relation to previous work. Chapter 4 will describe a study conducted to evaluate the interface with respect to learning gains added through the use of speech as a mode of input. Finally, I will conclude with a discussion of future directions for this research.

Chapter 2

Background

A great deal of work has been done in the past with the use of automatic speech recognition in systems for language learning. In section 2.1, I will provide an overview of such systems. In section 2.2, I will discuss the limitations of previous work and the reasons why this is still an active area of research.

2.1 Previous Work

In 1991, one of the first fielded systems incorporating speech technology was created by the Center for Teaching and Learning with the goal of teaching young children to read English [7]. The system was used to expand a child's vocabulary by embedding new words within a goal-oriented game. Results from feedback in schools indicate that the children enjoyed using the program and that being able to control what happens on screen using speech can be a motivation for learning to read. This system demonstrates an early application for speech technology and its use in a system for promoting learning.

As further systems were developed, a distinction arose between what's called "closed response" and "open response" systems [2]. In a closed response system, users know exactly what they are allowed to say in response to any given prompt. In an open response system, the set of recognizable responses remains hidden and the student must generate the appropriate response to a prompt without any cues from

the system.

An example of a closed response system is *TraciTalk*, created in 1998 by Courseware Publishing International [2]. This system engages students in solving a mystery, where each step of the dialogue presents the user with a fixed set of possible responses. The advantages of this approach are two-fold. Firstly, it limits the scope of user responses, leading to higher recognition rates in the face of fewer alternatives. Secondly, users are certain to progress through the dialogue, as they are made aware of the possible responses. The main disadvantage, however, is that much of the entertainment value in speaking with a computer is lost, as the responses are treated essentially as multiple choice problems.

An open response system, on the other hand, works the same internally as a closed response system, but with the choices hidden from the user. As the allowed responses are hidden, users are presumably given more freedom in how to respond to any prompt. To handle such variety, developers must consider as many possible correct responses to a prompt as possible. As it's hard to anticipate all the responses, the user may be left frustrated speaking a response that is grammatically or logically correct, but the system cannot understand.

An example of an open response system is *Subarashii*, developed in 1997 [1]. In this system, students solve simple problems by speaking Japanese in seemingly open-ended dialogues called "encounters". Possible responses remain hidden from the user, so the grammar must include a wide range of possible phrases. Through early evaluations of the system on a high school class, they found such open-ended dialogues to be difficult and that many students had trouble figuring out what to say. Instructional support before starting a session was extremely important in order to take full advantage of such a system.

A more recent system started in 2005 at the University of Southern California is called the Tactical Language Training System for learning Arabic [17]. This system 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 over 80 hours of an

interactive story-based game with a task-based focus (See Figure 2-1).

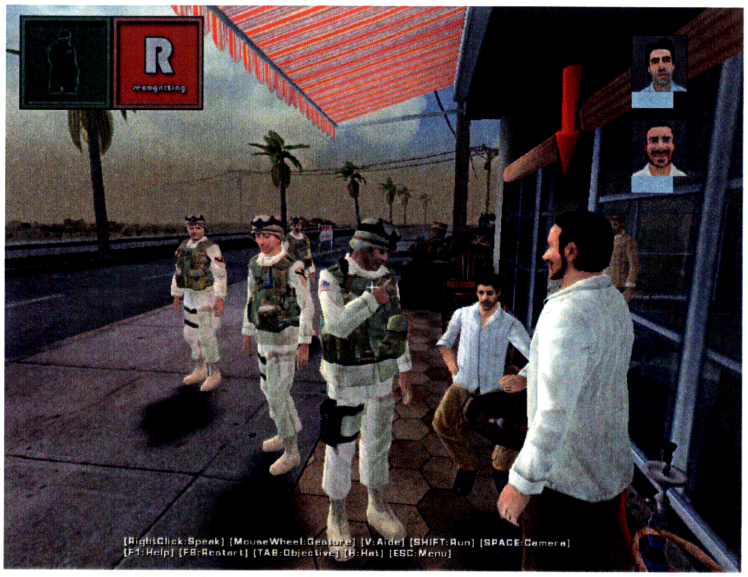


Figure 2-1: Screenshot of the Tactical Language Training System

A key design focus in this system was that "artificial intelligence techniques used must support the learning-promoting features of a game, otherwise they may be superfluous or even counterproductive." [17] In this system, the developers strove to create an environment that felt like a game, but still made the learning central to the application. The program has been successful, and is used actively by the U.S. Armed Forces. Its adoption demonstrates how speech-enabled systems presented in the form of a game can have practical use for language learning.

Many companies have emerged which specifically tackle the development of software for language learning. Rosetta Stone [14] is one of the more popular software lines. The current version includes a speaking drill in which a student listens to a word, then must repeat it correctly before continuing to the next problem. The development of such a drill in highly commercial software reflects the growing demand for computer software that can listen and respond to user speech.

Currently at the MIT Spoken Language Systems group, much work is being done in the area of computer-aided language learning for Mandarin Chinese [13]. One such system is a website called Chinese Cards where users can log in and add words

to personal sets of vocabulary. The system allows for storage of English, pinyin (romanized Chinese), Chinese characters, and photos in an integrated environment which is publicly available[1] (See Figure 2-2).

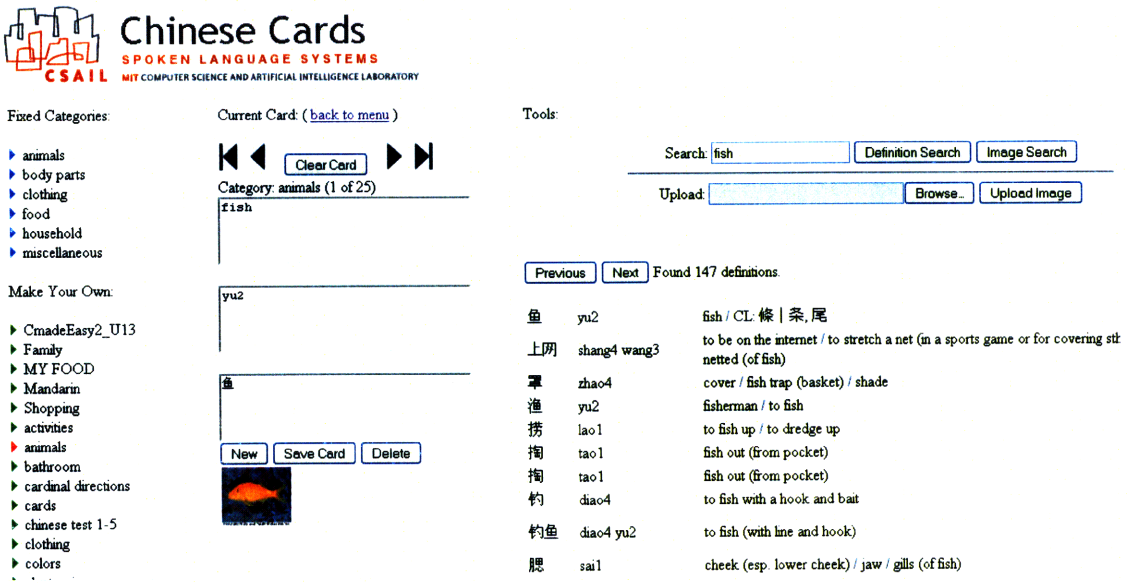


Figure 2-2: Screenshot of the Chinese Cards system

Using this system for vocabulary management, a game called Word War[2] was created to incorporate such user-added vocabulary into an interactive speech-driven game (Figure 2-3). The objective of this game is to match pictures of vocabulary words by speaking commands for placement of 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 flashcards as a method for vocabulary acquisition using spoken language. In many ways, this game strives to address our motivating factors as listed in Section 1.1. By providing dynamic construction of speech-enabled games based on user-entered vocabulary, Word War demonstrates a customizable, available, and practical system for learning.

[1] `http://islands.csail.mit.edu/chinesecards/`
[2] `http://islands.csail.mit.edu/wordwar/`

Figure 2-3: Screenshot of the Word War game

2.2 Discussion

Despite previous 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. Although large domain systems sound more appealing since they give a lot of freedom to the user, they are harder to create in practice, as it's difficult to anticipate the range of user responses. Additionally, non-native learners in particular have many issues in fluency and pronunciation mistakes that make it difficult to achieve adequate recognition performance in the face of large domains. Recognition improves in smaller domains, but these systems have less realism and are only useful while that specific set of content is not yet mastered. Thus, a balance needs to be met between the extremes to provide an interesting and long-lasting, yet usable system.

In the case of Word War, the domain is limited to the task of moving tiles into slots and the grammatical constructions used for this procedure. Though the small set of allowed expressions does not show much variety, this domain provides decent recognition for learners of all levels of experience.

A second question is the extent to which the system reveals the allowed set of expressions. Well-done open response systems have the potential to simulate more natural communication, but are harder to create in practice and may leave learners frustrated or confused. On the other hand, closed response systems with a set of known expressions seem less realistic, but the learner is aware of possibilities so

frustration is less likely. It would be great if all systems could be open response and large domain with excellent recognition performance for second language learners. However, until this system exists, we are faced with the question of how to make a system that is engaging yet makes the appropriate sacrifices for the sake of usability.

Word War makes the allowed expressions known to the user, and can do so because the grammar set is small. The grammar is limited to particular carrier phrases, where vocabulary can be substituted in certain positions. The word for each vocabulary item is available upon clicking a "Hint" button on a card. Having the allowed expressions available, users are free to speak phrases in any order they please once the game is loaded.

A third issue is how to create systems such that their usage is widespread. Many interesting systems, such as the Tactical Language Training System, require installation of a large resource-intensive bundle on a user's personal computer before use. Additionally, the system is extremely expensive and generally not available for use by the public. Ideally, such tools should be free, lightweight, and available to everyone at any time on a web platform.

A fourth concern is for systems to be designed so that they can remain relevant for future use. In order to provide context for learning, it is difficult to provide this kind of experience. For example, designing a system for teaching new vocabulary in a dialogue system would be useful so long as the vocabulary items are not known to the user. Once the system is fully explored, there may not be benefit in playing again. It would be great if systems were customizable such that their educational value could survive many repeat visits.

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 will not be educational in the sense that it does not have a method by which the user can learn incrementally. The learner may become overwhelmed in an environment that is not designed for non-native speakers.

Word War was designed for language learners to satisfy many of our motivating qualities, 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 fun. A hope is that systems with greater intrinsic entertainment value can be created to captivate a learner's interest and keep them coming back.

Rainbow Rummy aims to strike a similar set of balances as Word War on the first four concerns, but differs in that it was designed as a game. Great learning tools may be effective, but if students have no incentive to use them, little actual learning will be realized. Thus, the challenge becomes to design tools such that they can effectively teach material while keeping the user sufficiently engaged for repeated use.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Rainbow Rummy

In this section, I will provide a detailed description of Rainbow Rummy, its functionality, design, and deployment. I will begin with an overview of the system in Section 3.1. Next, I will provide details on the design and underlying technology in Section 3.2. Section 3.3 discusses the multi-player version of the game, while Section 3.4 describes our deployment of the system for public use.

3.1 Overview and Functionality

Rainbow Rummy is a customizable, speech-enabled online card game for language learning. The game is played against another computer or human player, where the objective is to get rid of all the cards in one's hand before another player does. Similar to other popular card or tile based games such as "gin rummy" or "Mah Jong", the game is turn-based and has a shared game board onto which the player must place cards by creating sets (See Figure 3-1).

In traditional card games, each card is uniquely identified by its suit and number. In Rainbow Rummy, this combination is mapped to each card having an associated color and vocabulary item[1] (See Figure 3-2). Each vocabulary item in Rainbow Rummy is represented visually on the card as a picture. The pictures are loaded

[1]The term "vocabulary" is used throughout this paper, but in reality, any word or phrase can be associated with a single card. For example: "child" or "the child who is eating"

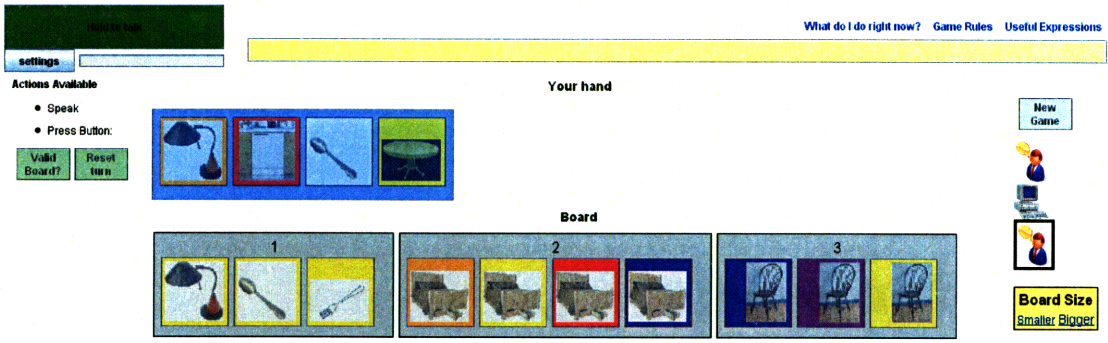


Figure 3-1: Screenshot of Rainbow Rummy

from the Chinese Cards interface, where the user can associate vocabulary items with images taken from Yahoo Image search or through uploading personal images.

A key to the effectiveness of our game is this feature where users create their own vocabulary sets through the Chinese Cards interface and load them into a game of Rainbow Rummy. Many learning games use fixed content, so they can only be relevant so long as the material is not mastered. Enabling the user to enter their own vocabulary lists and play a speech-enabled game that uses this content gives the solution the potential to retain its relevance as long as there is new vocabulary to be learned. In the case of second language learners, there will always be new words to learn.

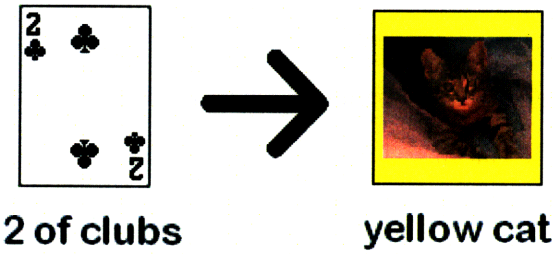


Figure 3-2: Mapping traditional cards to Rainbow Rummy tiles

To get rid of cards, a player removes cards from his or her hand by appending to the board to create sets based on either color or vocabulary. In many traditional games, players create such sets based on suit and number sequence. As there is no natural ordering to either color or vocabulary, however, our game has no concept

analogous to a running sequence. Rather, sets are formed with three or more of one color, or three or more of one vocabulary item. This choice was made to enable a more natural speech interface and is discussed further in Section 3.2.3.

A game between a single human player A and a single computer player begins as follows. Each player starts with a set of 7 cards in their hand and an empty shared board. The user performs actions by using speech and mouse inputs, and listens to instructions generated by the computer. I will first describe how player A plays his own turn using speech, then I will describe the computer player's turn and its use in providing listening practice.

3.1.1 Speaking mode: Human player's turn

In order for player A to move his cards around the board, he must speak instructions in Chinese using his microphone, the results of which will then be performed visually. For example, the user may say "bǎ hóng sè de gǒu fàng dào dì èr gè gézi.", meaning "put the red dog in slot 2". On screen, the "red dog" card will move from his hand to the 2nd slot below. Explicit pronunciation assessment is not given, but rather, is implicit in the actions which result on screen. This choice reflects our motivation to create a game that has entertainment value through reducing its drill-like features.

As the user speaks, we take advantage of incremental recognition results to enable immediate visual feedback. For example, as the user says "bǎ" (meaning "grab"), all items on the board will become selected[2]. Upon continuing with "hóng sè de" ("the red"), the red items of that set will be selected. When "gǒu" ("dog") is spoken, the dog within the set will be the only item selected. As "fàng dào dì èr ge gézi" (put it in slot 2) is added, the card will move from its current position to the 2nd slot on the board. Visual selection and movement happens *while the user is speaking*, rather than after the whole phrase is spoken, providing instant feedback.

We provide a limited set of template phrases relevant to game play (see Table 3.1), as well as a number of alternative means for expressing such commands. We

[2]Selection is indicated by opacity of the card, where fully opaque means the item is selected, and semi-transparent means it is not.

Table 3.1: Phrases used in Rainbow Rummy

Chinese	English
Bǎ <color> sè de <vocabulary>...	Put the <color> <vocabulary>...
...fàng zài dì <number> ge gézi.	...in slot <number>
Zuò yí ge xīn de gézi.	Create a new slot.
Chōu yì zhāng kǎ.	Draw a card.
Gāi nǐ le, OR Wǒ zuò wán le.	Finish turn.
Wǒ shuō cuò le.	Reverse move.

also provide a list of colors and numbers for reference. As we are using pictures to represent vocabulary, the user is able to hover their mouse over a card to see an English-Chinese translation. The intent is that as the user plays the game with these words, reading the hint will no longer be necessary. A further justification for our choice of grammar is provided in Section 3.2.3. To finish a turn, the user must draw cards until at least one card can be moved into a valid position on the board, or draw the maximum per turn of 5 cards.

In addition to the speech interface for a human player's turn, we also provide the ability to enable clicking buttons and dragging and dropping cards to perform the same set of actions. This mode, which we call the "cheat" mode, was found to be important in allowing users the freedom to bypass speaking for particularly difficult words or phrases.

3.1.2 Listening mode: Computer player's turn

On the computer player's turn, the system generates a set of moves to be performed. To keep the user engaged, by default we require that the user listen to directions spoken by the computer and perform these moves by clicking buttons and dragging and dropping cards into the correct positions. For example, if the computer can add a card to an existing set on the board, it may direct the user to "put the red elephant in slot three". At this point, all cards on the board, as well as the cards which the computer will move out of its hand for this turn are revealed to the user.

Amongst this set, there may be multiple red objects or multiple elephants, so the user is required to listen carefully and select the correct card to move into the correct slot. If a wrong move is performed, it will undo the move and require the user to try again until the move is correctly performed. At any point the user is allowed to click a repeat button to hear the audio directions again for clarification.

This combination of modes for human and computer turns keeps the user engaged while training both speaking and listening aspects of communication. As an option we also have a mode where the computer performs its moves while the user passively watches. While this is more akin to real-life card games, we believe that the combination of both speaking and listening as seen in the default option provides a more complete environment for promoting learning. The user is not only able to practice listening skills, but is also able to hear words spoken by the computer as guidelines for his own speech.

3.1.3 Finishing a game

As one turn is finished, it becomes the next player's turn in the sequence. This cycle continues until a player gets rid of all the cards in his hand and creates a valid board. The player who gets rid of all his cards first wins the game. As there are a limited number of cards in the deck equal to $(numberofcolors) \times (numberofvocabulary)$, players can tie if the deck runs out.

3.2 Design

Having described the mechanics of the game, I will now explain the technology upon which it was built. Secondly, I will provide further justification for our design choices in light of our motivating factors and current technology.

3.2.1 System Details

For our game, we require several components: a Rainbow Rummy server-side application, a client interface, a system for user content and account management, a speech recognizer, and a speech synthesizer.

Our game is built in Java using the WAMI (Web-Accessible Multimodal Interfaces) toolkit developed at the Spoken Langauge Systems group [5]. The toolkit provides account management and the means for linking recognition and synthesis services for use in a web application.

The setup uses a Java applet to stream audio from the client browser to a recognizer running remotely. For our game, we use the SUMMIT landmark-based recognizer [4] with acoustic models trained on native Mandarin speech. While tones are important for correctness in spoken Chinese, we ignore tone features with the intent of creating a more usable system for beginning learners.

Once audio is received at the server, it is processed and matched to a command in the context free grammar specified in Java Speech Grammar Format (JSGF). This JSGF file contains the phrases needed to configure the recognizer language model. In order to handle user-selected vocabulary, the base grammar is dynamically appended with rules for each new lexical item in the category loaded into the application by the user. Thus, for each game, the recognizer is enabled with the selected vocabulary chosen for that session, and only limited to those words as embedded within the set of template carrier phrases. This feature provides better recognition for language learners while still providing ample room for misrecognition in the face of random utterances. The dynamic generation of grammar was a key feature in the creation of a game which has limited grammatical scope, yet is unlimited in its ability to handle any content entered by the user.

To manage vocabulary item creation and selection, we use the Chinese Cards system. As this system was used for Word War as well, the addition of Rainbow Rummy as an alternative game demonstrates the method by which new games can be easily added within this framework.

The top scoring recognizer hypothesis is sent back to the server-side Rainbow Rummy application with meta tags embedded in the string. These tags, as specified in the JSGF file, are interpreted by the application and parsed into meaningful chunks for conversion to actions for use by the application.

We used WISTON, a large corpus based speech synthesizer developed at the Chinese Academy of Sciences in Beijing, for providing computer speech in Mandarin Chinese. The synthesizer takes a string of Chinese characters or pinyin as input and produces a waveform file as output. As the library was written in C, we wrote a Java wrapper to access the native code and plugged it into WAMI for streaming through the audio applet. Having both the recognizer and synthesizer available, we were able to create Rainbow Rummy as a full speaking and listening game.

On the client-side, we make heavy use of Asynchronous Java-Script and XML (AJAX) to create the interactive game experience. Drag and drop was an important mode of input for creating a usable system. We faced issues with browser real estate, as game boards could get very large. Through many iterations, we settled on the current solution which simplified components and allows dynamic user resizing of the cards.

3.2.2 Game Strategy and AI Planner

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. For basic game play, 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 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 3-3(a)), a third red card could be stolen from slot 2 containing four "bed" cards, and placed in a new slot (Figure 3-3(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.

(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 3-3: Example sequence for a complex move

To manage the deck and plan moves for the computer, we wrote a simple turn planner which we call the "AI planner". Each turn is planned based only on the current state of the board. It performs standard moves from its hand, as well as basic stealing operations from existing slots on the board like the ones above. There are more complex board manipulations that are possible but beyond the current level of expertise of the AI planner, though future work can easily create an implementation with more advanced behavior. Currently, such opportunities could still be exploited by a skilled human player and played against another human player of the same level for a more challenging experience.

Error handling

Errors in the game are handled in several ways. In the case of a misrecognition error, the command that the recognizer thought the user said will be performed instead of the intended one. Once a card is moved out of one's hand, the person cannot move it back into their hand. This is intended to mimic existing card games. However, considering the errors possible due to misrecognition, we allow the user the option to start their turn over at any moment by clicking a button. All cards drawn in that turn up to that point will not return to the deck, but rather, will move back to the user's hand. This prevents users from cheating by looking forward in the deck,

resetting their turn, and drawing fewer cards for their benefit.

As users may not be familiar with the game rules, they may want to verify if a board configuration they created is valid or not. To allow this check, we created a button which the user can click to check its validity. If a user attempts to finish their turn with an invalid board configuration, their turn will start over and they will have to try again.

A misrecognition of the form "finish turn" when the user is not done with their turn can lead to much frustration in having to repeatedly start a turn over. Thus, we found it necessary to add a dialog to confirm that users are sure that they want to finish their turn.

Multiple languages

Rainbow Rummy has the capability for extension to languages other than Chinese. In order to port it to another language, several pieces are needed: a recognizer, a synthesizer, and a JSGF file written for that language. We demonstrate this capability by providing the user an option to play the game in English or Chinese. For this, we used an off-the-shelf English synthesizer called decTalk, and wrote a new JSGF file for the English grammar. The pluggable nature of synthesizer and recognizer resources in the WAMI architecture accounts for much of this flexibility. In Section 3.3.2, we discuss how supporting multiple languages presents unique opportunities in multiplayer games.

3.2.3 Justification

In this section, I will address several of our design choices in further detail and explain why we made them given our goals.

Choice of color and vocabulary item

We chose color and vocabulary item to replace the suit and number of traditional cards for several reasons. The choice of vocabulary item was necessary to allow cus-

tomizability of cards for language learning. The choice of color as the other dimension was made to present a more natural speech interface, while also training the student to master the vocabulary of colors.

If we had chosen to keep either suit or numbers to accompany a vocabulary item, commands such as "grab the cat of clubs" or "grab the two of cats" would be unnatural and confusing to the user. Alternatively, we considered using "grab the two cats", where number indicated amount of a particular vocabulary. However, we believe this adds its own confusion in that the phrase could be interpreted as affecting two of the cards with cats rather than the single card. Secondly, specifying number of an object in Chinese will use different "counter words" between the number and object depending on qualities of that object. As users specify their own objects, it isn't clear how to account for all possible counter words while keeping the grammar set small.

In the face of these alternatives, we chose color because it can easily modify a phrase and does not present the same sort of confusion or difficulty. While "grab the orange elephant" may not be logical as elephants with orange skin don't exist, it is both natural in speech and makes sense within the context of a game in which each card has a color and vocabulary item picture. Through this choice, we find validation in the idea that creating speech-enabled games is not a task of simply adding speech as an input to an existing game, but rather, involves careful consideration in the design process as well.

Choice of grammar

Within the context of this card game, further additions to the grammar outside the basic set were not found to be necessary. Secondly, our choice of a limited grammar was important in improving recognition rates given that our target users are non-native speakers. The range of accents for learners of Chinese is extremely varied, yet may still be intelligible to a native speaker. Our goal in this game is not to force perfect pronunciation that is beyond a learner's ability, but rather, to provide an environment through which they can speak to the best of their ability without the burden of feeling as if they are being evaluated in a classroom drill. If speaking can

indeed help in the acquisition of the chosen vocabulary, it is important for students to be motivated to speak and not become too discouraged from the start.

3.3 Multi-player version

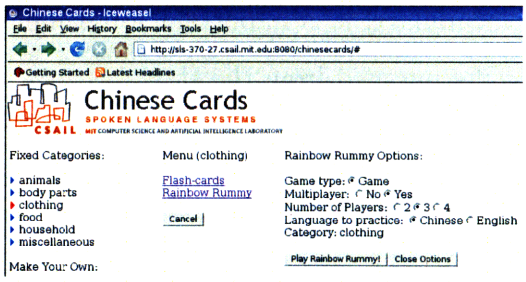
We designed Rainbow Rummy to be a multi-player game able to handle an arbitrary number of human and computer players in one session. This feature provides opportunities for language learners to play games with real opponents for mutual benefit. Most interestingly, however, is the fact that these games can be played such that each player learns his or her own set of words.

3.3.1 Description

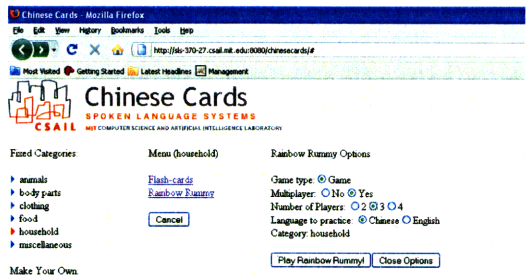
In traditional offline card games, players share a set of cards which appear the same for everyone. Even if one printed physical Rainbow Rummy cards, two users cannot play an in-person game without learning this same set of vocabulary. The online medium of our game, on the other hand, allows users to each learn *their own* content *simultaneously* while engaging in a game with another human player who is learning different content.

In an example game, player A may choose his set of "clothing" cards, while player B may choose her category for "household" cards (See Figure 3-4). If they both specified to play a multi-player game, they will be linked and compete against each other. Player A will see, speak, and hear all moves as performed on his "clothing" vocabulary, while player B will see all moves as performed on her "household" vocabulary.

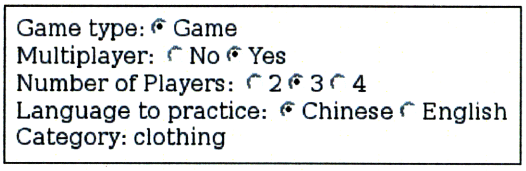
In one move, player A may say: "put the yellow scarf in slot number 1". When the move is performed, player B will see, for example, a yellow "bed" move into slot number 1 (See Figure 3-5). At the same time, she will hear the synthesized sound clip: "put the yellow bed in slot number 1". To each user, it appears as if the other players are using the same set of cards and speaking their moves based on this set, while in reality they are each using their own set. This capability is possible since each card is identified only by a unique identifier in the shared game space. These



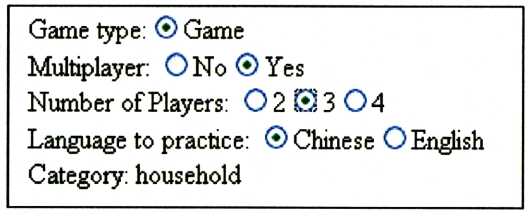
(a) Player A (using Firefox/Linux) chooses his "clothing" category



(b) Player B (using Firefox/Windows) chooses her "household" category



(c) Player A's game options



(d) Player B's game options

Figure 3-4: Two players select options for a multiplayer game

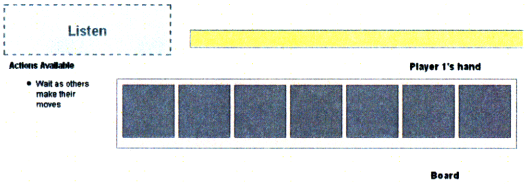
identifiers have a separate set of mappings to specific words for each player in the game.

The means for setting up a multi-player game and matching users is as follows. Each human user has his own controller for a recognizer and synthesizer, as well as certain user-specific data. The users are linked by a shared game space which manages their session together. To access multi-player mode, the user specifies several options before entering via the Chinese Cards interface.

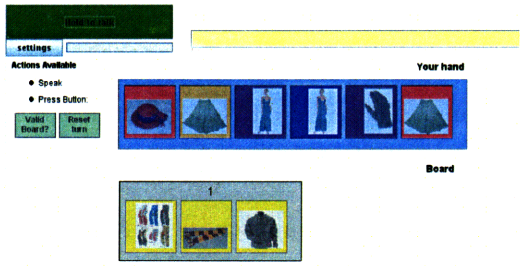
Firstly, the user decides whether the game is single-player or multi-player. Secondly, the user decides how many players to include in the game. If single-player mode is chosen, the user will occupy one player slot while the remaining positions will be filled by computer players. The user then plays the standard game as described in earlier sections with as many computer players as needed to occupy the remaining positions.

If the user chooses multi-player mode, the user will wait for other players to join who have the same game requirements. The condition under which users are matched is that each has specified the same number of players in the game and requested that

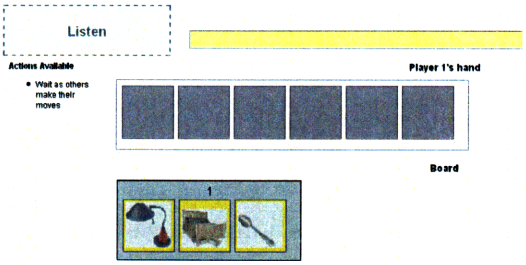
(a) Player A enters the game



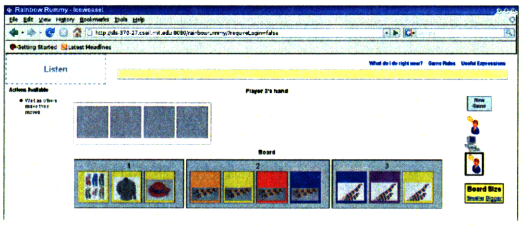
(b) Player B enters the game



(c) Player A speaks his moves



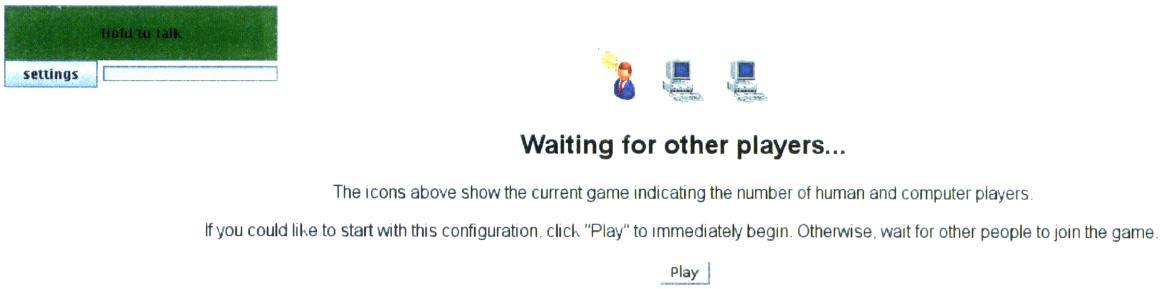
(d) Player B watches and listens as Player A moves



(e) Player A watches and listens as Player B moves

(f) Player B speaks her moves

Figure 3-5: A sample multiplayer game sequence



(a) Waiting screen for a multiplayer game

(b) Waiting images with one user present

(c) Waiting images with two users present

Figure 3-6: The matching process for a multiplayer game

it be multi-player. Upon submitting these options, if a game cannot be immediately started, the user will see a waiting screen (Figure 3-6(a)). As a user with the same game requirements joins, however, one of the computer images will convert to the person image, indicating that another human user is present and waiting (Figure 3-6(c)). Once the positions are completely filled with human players, the game will automatically begin.

At any time while waiting, a user can decide to start the game. When this happens, the current player assignment is locked in and the game begins. For example, if two users are joined and waiting to start a 3-player game but there is no third player, any user can click on the "Play" button to start the game with this arrangement of 3 human players, and a computer player will fill in for the missing third person.

While playing a game, if a human player leaves the game by closing their window or navigating away from the page, their position will be automatically converted to a computer player and the game continues as normal for others without restarting. When there are computer players in a game of 2 or more human players, computer moves are performed automatically according to a timer and each user hears narration as if the computer were speaking their moves. If the number of human players drops to just one, the game switches to the default listening mode in which the user must

perform actions for the computer. This behavior is consistent with the default single-player version of Rainbow Rummy.

This multi-player version allows for users to play against more challenging and realistic opponents. It enables a method for playing a game for language learning between multiple people when face-to-face meetings are not possible. Most importantly, the custom vocabulary feature makes the game accessible and relevant to many users, as the matching procedure allows each user to choose their own set of vocabulary to learn.

3.3.2 Cross-language interaction

The multi-player feature combined with the ability for the system to be ported to other languages provides opportunities for including a much broader user base. Currently, two users can play a game of Rainbow Rummy together, where one user in China is playing in English with their "animals" cards, while the other user in Boston is playing in Chinese with their "electronics" cards. Each user is able to practice his or her target language at the same time. This flexibility provides unique opportunities which cannot happen in physical space.

A common format taken when English and Chinese speakers meet is that the native English speaker will speak in Chinese, while the native-Chinese speaker talks in English. Each is able to practice speaking their own language, but is forced to listen in their own native language.

We don't claim that Rainbow Rummy has the effect of solving this mismatch, as it is not a dialogue system. However, its flexibility in providing both listening and speaking in a target language with multi-user interaction illustrates a concept enabled only by use of computers. One can imagine future systems in which Chinese phrases spoken by the native English speaker are translated, synthesized, and heard in English by the native Chinese speaker on the other end. Similarly, the native Chinese speaker talks in English, while the native English speaker hears the comment in Chinese.

In this setup, both are able to practice speaking *and* listening in their target language while interacting with a native speaker. Translation and speech synthesis

quality present technological barriers to creation of a system which performs this role realistically. Secondly, subjective satisfaction of interacting with such an interface make the approach of questionable practical use, but the concept is intriguing. Rainbow Rummy illustrates a very limited version of this, as we use key value pairs to parse and construct commands in a limited grammar as a primitive translation component. Nevertheless, the basic concept is preserved in the interface presented to users as they interact in a multi-player game.

3.4 Deployment

To determine how real users interact with the system, we deployed Rainbow Rummy for public use as a game in the Chinese Cards system at `http://islands.csail.mit.edu/rainbowrummy/`. Through deployment, we have been made aware of many of the challenges in designing a system for practical widespread use. The most important lesson has been that systems must be designed to handle the wide range of variability seen in client environments to be truly universal.

Rainbow Rummy was deployed on an Apache Tomcat server instance running on an available Linux machine at SLS. We logged basic statistics on user interaction with the system using Google Analytics[3]. Since deployment in January 1, 2009 until April 25, 2009, we received 223 visits from 17 countries (Figure 3-7). The average time spent in the game was roughly 12 minutes per visit.

We logged all commands performed by users in the form of page views in Analytics to evaluate how users interact with the system. A list of the game commands and the usage statistics are shown in Table 3.2. While the speech is a relatively new mode of input for interacting with a web interface, we find that 2048 card movements were successfully made using speech. These numbers don't make claims on recognition accuracy, but simply aim to validate how real users are able to interact with the system despite the learning curve of speech as a new mode of input, and the difficulty

Table 3.2: Statistics of user interaction with deployed version of Rainbow Rummy

Action		Times performed
Card moved	in speaking mode (using speech)	2048
	in speaking mode (by drag and drop)	208
	in listening mode	1338
Card drawn	in speaking mode	710
	in listening mode	592
New slot	in speaking mode	589
	in listening mode	421
Turn finished	in speaking mode	421
	in listening mode	529

of speaking a foreign language.

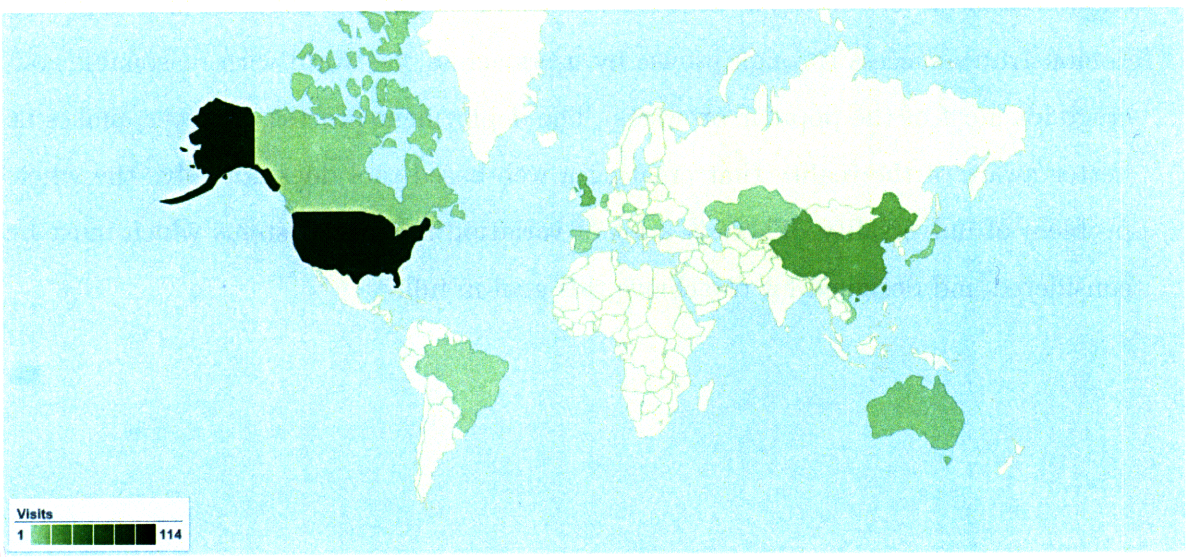


Figure 3-7: Geographic distribution of visits over 4 months

Secondly, the statistics we collected highlight the importance of designing games which work across various platforms and browsers. We received visitors using Firefox (60.71%), Internet Explorer (27.68%), Mozilla (4.46%), Safari (2.68%), Chrome (2.23%), and Opera (0.45%) as their browsers. However, when looking at the percentage of games finished by users of each browser, we find that 80.56% were completed in Firefox, 13.89% were completed in Internet Explorer, and 5.56% were completed

[3] `http://www.google.com/analytics/`

in Chrome. Users of the other browsers did not complete games.

As completion of a game requires full use of all features, the lack of completion in certain browsers suggests possible problems with the operation of Rainbow Rummy in those environments. Before deployment, we only tested the system using Firefox and Internet Explorer, the two browsers we felt to be most popular, so we cannot claim full cross-browser compatibility. Additionally, we found that 92.86% of visiting users had Java support, which is necessary for operation of the audio applet, and thus, for the game as a whole. This figure captures a relatively large percentage of our user base, but potential exists for increasing that figure further by providing an alternative in the form of a Flash audio component.

Despite the wide diversity of visiting users and their settings, we feel the system shows potential for serving as a useful web tool for language learners. Our approach demonstrates success through its use by users across the world with reasonable coverage for most of the popular browsers. The deployment process, however, makes us better aware of the reality that creating a web-based tool does not solve the whole problem of universality. There is a lot of variation in client systems which must be considered and designed for to achieve this goal in full.

Chapter 4

User Study

To evaluate potential learning gains the system could have for vocabulary building, we ran a web-based study on 16 subjects over the course of roughly 2 weeks. In Section 4.1, I will discuss similar studies for quantifying learning gains conducted in the past. In Section 4.2, I will describe details on our experimental design and setup. Section 4.4 presents an analysis of the collected data, and Section 4.5 provides a summary of our findings.

4.1 Previous work

Rod Ellis, a leading researcher in Second Language Acquisition, performed a number of experiments for evaluating the effect of various learning techniques on vocabulary acquisition [3]. In his 1999 experiment, he aimed to measure the acquisition of vocabulary as presented to subjects within different "incidental" learning environments. "Incidental" learning, where learning occurs secondary to another primary task, contrasts with "intentional" learning, where learning is the explicit goal of the activity. Ellis compared three incidental learning environments for subjects as they were faced with the task of following directions for placing vocabulary items into positions. These environments differed in their degree to which the subject was able to ask questions when presented with directions. The three modes were:

- Premodified input: Subjects could not ask clarifying questions.

- Interactionally modified input: Subjects could ask basic questions (ex. "What is a rocker?")

- Modified output: Subjects could converse freely with another person to achieve the goal

In all three cases, Ellis found the learning gains for those words presented in the activities to be significant. The modified output setup led to significantly greater learning gains than the other two, while the premodified input and interactionally modified input environments performed roughly the same. Our study similarly aims to measure incidental learning gains between different environments. However, our setup differs in that it compares a configuration in which a user can speak on their turn, with a configuration in which the user is only allowed to listen. Secondly, the nature of the study is different in that the subjects in our study interacted with a computer system via the web.

The Word War study conducted at SLS in 2008 [8] also evaluated the use of speech and its effect on learning gains for vocabulary acquisition. The study compared three systems: a listening-only version of Word War, a speaking-only version of Word War, and traditional flash cards. The results showed no significant difference in the mean learning gain between the three systems.

Our setup is different on several major points. In both of our versions of Rainbow Rummy, the user is able to listen while it is the computer player's turn. This frames the addition of user-speech in a more natural setting that trains speaking and listening at the same time. Secondly, the Word War study was conducted in a laboratory setting, while our study was conducted remotely over the web. Thirdly, Rainbow Rummy was designed as a game, so the nature of the activity for learning was different as compared to the more drill-like Word War interactions, where learning was clearly the intent of the task.

4.2 Design and Setup

We designed a within-subjects experiment in an attempt to measure the vocabulary learning gains attributed to the use of speech as a mode of input in Rainbow Rummy. Our independent variable was the use of speech as a mode of input. We presented two versions of Rainbow Rummy to each subject: one version with only listening practice, and another version with both speaking and listening practice. Using this setup, our study aimed to answer several questions:

- Can Rainbow Rummy be an effective environment for learning new vocabulary?
- Does adding a speech interface benefit users in terms of vocabulary acquisition?
- Does Rainbow Rummy in fact have more intrinsic entertainment value than Word War?

As our dependent variable, we used a modified version of Word War to collect data measuring knowledge of words both before and after interacting with each version of Rainbow Rummy. By examining the differences in how vocabulary knowledge changed after interaction with each version, we aimed to provide a basis for our suspicion that speaking to a computer can play a useful role in language learning. Before continuing with specific details of the study, I will first describe the two versions of Rainbow Rummy, the modified version of Word War, and why they were used for this study.

4.2.1 Systems used in the experiment

Two versions of Rainbow Rummy: "speaking" vs "listening-only"

We created two distinct versions of Rainbow Rummy for this study: a "speaking", and a "listening-only" version. The "speaking" version is the same as the default version of the game described earlier. This involves the user in both speaking and listening aspects of communication, where the user speaks moves on his turn and listens to instructions on the computer turn.

To ensure that users made use of the speech interface, as this was our variable of interest, we only exposed the ability to "cheat" by dragging and dropping in the event that the user made more than three attempts at card selection before performing a drop move into a slot. From watching several early users of the system, we concluded that other moves of drawing a card, creating a slot, finishing a turn, and dropping a card did not require exposing a "cheat" option. These phrases should be recognized if the subjects indeed fit our target user population in that they are learners of Mandarin, can read pinyin, and have microphones of acceptable quality.

In the "listening-only" version, the user does not speak on his turn, but moves by clicking buttons and dragging and dropping cards. The way in which actions are performed is the same as if it were a computer turn where there is no restriction on actions that can be performed. As the user makes a move in "listening-only" mode, the computer will narrate and tell the user what move was just performed. This setup ensures that for each move performed on a user turn, there is a matching sentence that the subject is exposed to for both versions. In the "speaking" version, the sentence is spoken by the user, while in the "listening-only" version, the user hears the sentence as a narration of what was performed.

Since the computer speaks on both the human and computer player's turn in the "listening-only" version, we needed to distinguish the two modes clearly. To do this, we had the computer speak in past tense and at a slightly faster pace while narrating on the user's turn. While it may seem more natural if no sound clip is heard at all when user moves freely, we needed to incorporate this in order to provide the same move-to-sentence mapping across the two versions, where the only difference between the two was actively speaking versus passively listening. If we did not make this choice, the "listening-only" version would only expose users to words for half of the moves and bias results towards favoring the speech interface.

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. If we had chosen a larger number of items from one category,

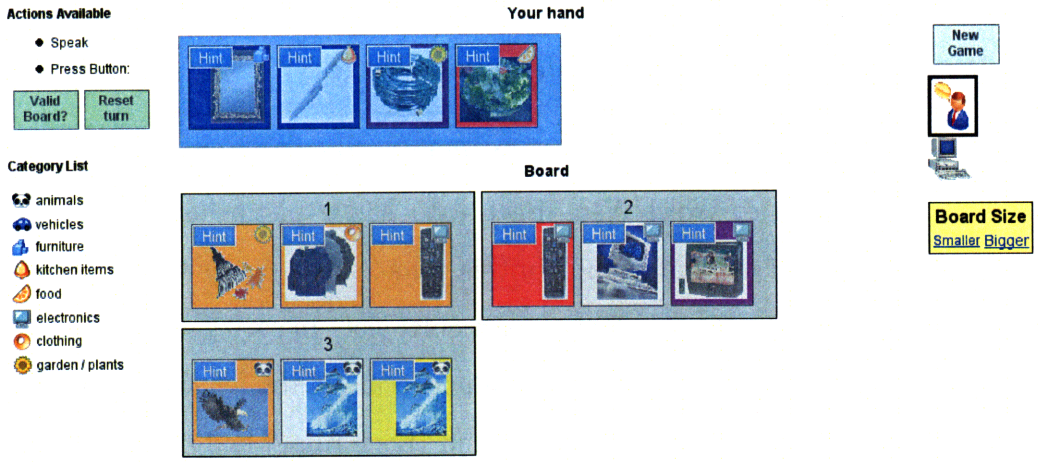


Figure 4-1: Screenshot of modified Rainbow Rummy using multiple categories

the deck size would become large and creating sets of vocabulary would become harder. As a solution, we replaced the "match by image" paradigm with 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. This is a concept that the Word War study did not use in its design, as all the words were unrelated to one another.

For each of the eight categories (animals, vehicles, furniture, kitchen items, food, electronics, clothing, and garden/plants), we selected a set of six vocabulary items. For each game, we chose 3 words from each category for a total of 24 words per game. As we had 6 colors in the game, each vocabulary word had two instances such that each category had full representation of all colors. This selection created a full deck of 48 cards with 24 vocabulary words. The words that we chose each had two characters to control for word length. A sample board which was created in this modified game can be seen in Figure 4-1.

Word War as a measure of vocabulary knowledge

To measure a subject's knowledge of the vocabulary items, we used a modification of the listening version of Word War. Each card had a picture of an item in the set of vocabulary tested for that round. In this game, the user was instructed by the computer to move a particular word into a particular slot. In each game for this study, we had 24 vocabulary words. The game was divided into 4 rounds such that the user needed to correctly match 6 items in each round. Each round had 6 distractor cards chosen from the set in the following round so that the user had to select and move the correct card amongst a total set of 12 choices (See Figure 4-2).

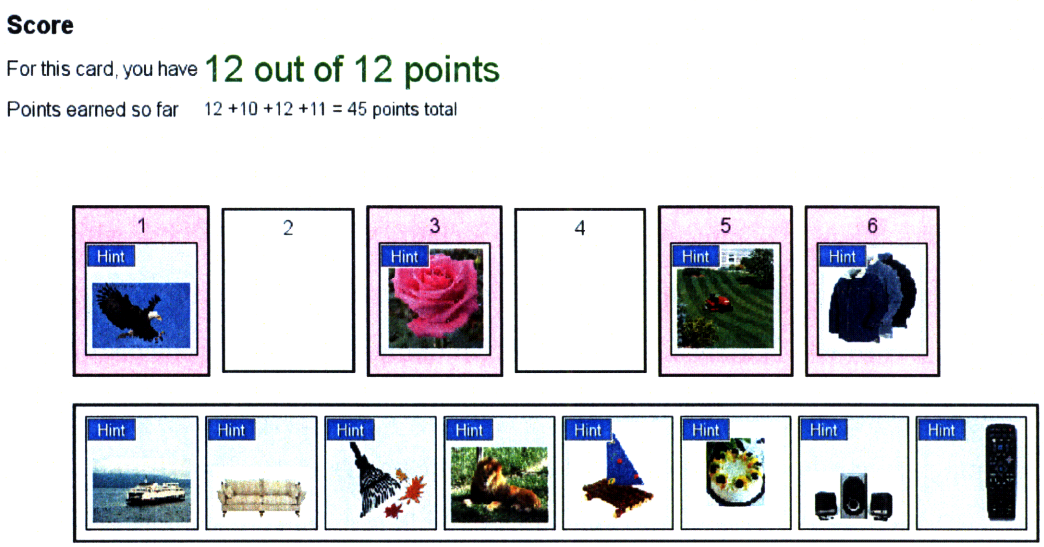


Figure 4-2: Screenshot of modified Word War game for the study

In order to measure knowledge of a word, we added a point system such that the objective of the game is to get the highest number of total points. For each command spoken by the computer, the user starts off with 12 possible points. The objective is to keep as many of these points as possible by performing the correct moves. Each card has a "Hint" button that can be clicked to see a Chinese-English translation of that item. Whenever the user clicks on the "Hint" button, 1 point is deducted from the possible 12 points. If the user chooses to move the wrong card, or puts the right card in the wrong slot, 1 point is similarly deducted from the score for that card. The

minimum amount that can be received for a card is 0 points.

The point system gives an incentive for the user to avoid thoughtlessly disregarding the instructions from the computer. Secondly, it prevents users from viewing hints when they already know the word. Thus, if a user receives a full 12 points for a particular word, it would seem to be a good indicator that this user actually knows the word.

For example, if a user knows that "diàn nǎo" is either "computer" or "TV", getting a hint for one of the words will inform the user that the other choice is the correct answer, so the user may end up receiving 11 points. However, if there is a word the user has no clue about, he may need to get the hint for a larger number of choices, lowering the score for that item. We felt that providing this basic measure for degree of knowledge in a word would provide richer opportunities for analysis as opposed to a binary indication that would come out of a more traditional testing method.

4.2.2 User selection

We only identify our subjects by their e-mail address, but we required that they have certain characteristics in order to participate. They had to be learners of Mandarin Chinese, be able to read pinyin, have a fast enough Internet connection, have a microphone, and pass an audio test that we devised. As we were not evaluating recognition, we can justify use of the audio test as being appropriate for preventing microphone quality from confounding our analysis of learning effects.

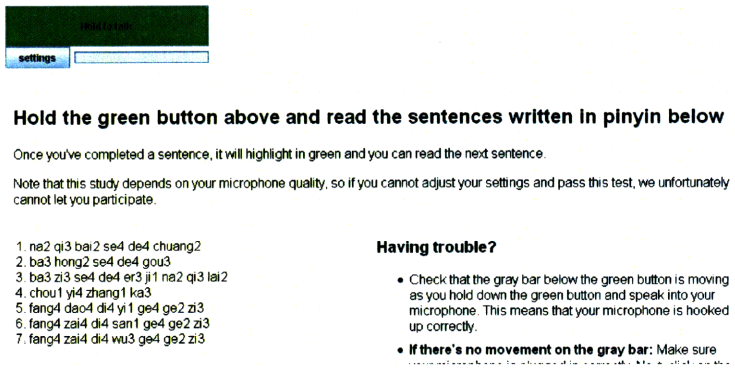


Figure 4-3: Screenshot of the audio test

The audio test we used was accessed through setting an option in the Rainbow Rummy application. There were 7 test sentences which were taken from the Rainbow Rummy grammar and presented to each test subject (See Figure 4-3). Each sentence was written in pinyin which the user had to read into their microphone. If what the system thought the user said matched the phrase on screen, it would highlight in green and the user would move to the next phrase.

This audio test served several purposes. For one, it tested if users had a microphone and were able to configure it correctly to capture audio. Secondly, it measured if the audio captured from their microphone was of acceptable quality for our recognizer. Thirdly, it measured if the potential user was really a Chinese learner, as significant effort would have to be made to learn Chinese pronunciation just for the sake of participation. Fourthly, it determined if users could indeed read pinyin, which was a requirement as the system relies on the romanized written form for providing hints and sample phrases.

These checks were necessary to prevent users who didn't fulfill the requirements from participating. As we provided an incentive for participation in the form of Amazon.com gift certificates, we needed to make sure that the subjects were serious and passed our target user qualifications.

Of particular note is that we conducted our study *remotely* over the web. We had no interaction with the subjects or access to the environments in which they performed their tasks. As a result, we could not control for various factors such as microphone quality, connection speed, or background noise. Since an important feature of our system is that it can be used by anyone at anytime via the web, we decided to launch our study in this manner. The results would then seem better suited to generalize in reflecting the actual learning gains that could occur through practical use of the system.

Our subjects were taken from a mailing list of those users who signed up for an account with the Chinese Cards system. Potential subjects were directed to a URL for signing up to participate. Over 40 people signed up for accounts, but many either did not begin the study or could not pass the audio test. Additionally, several subjects had

to be removed due to technical issues at the start of the study, leaving our final subject count at 16. Quite surprisingly, these subjects participated from a diverse sampling of locations. Subjects from the United Kingdom, Netherlands, China (Beijing, Hong Kong, Guangzhou), Taiwan, Singapore, Canada, and the United States participated as distributed in Figure 4-4.

Figure 4-4: Geographic distribution of subject locations

4.2.3 Task sequence

To compare the learning gains between the two versions of Rainbow Rummy, we had the subjects play a game in each. As a result, we needed two different sets of vocabulary where one was used with the "listening-only" version while the other was used with the "speaking" version. We started with a total of 48 test vocabulary words, where each of the 8 object categories had 6 words in it. For each user, we divided this set into two groups of 24 words. The division was done such that each user saw a unique split from other users.

As one session of the study version of Rainbow Rummy requires selection of 3 vocabulary items from each of the 8 categories, we divided the words by treating each word as if it were an index into a list of 6 items and chose a unique set of 3 indices as the first set. The words at these indices were taken for the first set of 24, while the remaining were put in the second set. Using this method, we had 10 unique partitions of word sets.

For our study, we wanted to present the partitions to subjects such that one

Table 4.1: Sequence of tasks for a subject seeing the "speaking" version first

Session	Task	Card set	Time
Session 1	1. Audio Test	N/A	∞
	2. Survey A	N/A	∞
	3. Word War	1st set	∞
	4. Rainbow Rummy Speaking Tutorial	Tutorial set	15 min
	5. Rainbow Rummy Speaking Game	1st set	30 min
Wait 4-7 days			
Session 2	6. Word War	1st set	∞
	7. Word War	2nd set	∞
	8. Rainbow Rummy Listening Tutorial	Tutorial set	15 min
	9. Rainbow Rummy Listening Game	2nd set	30 min
Wait 4-7 days			
Session 3	10. Word War	2nd set	∞
	11. Survey B	N/A	∞

subject sees a set first, while the other subject sees the complementary set first. To account for this ordering, we distinguished 20 ways to present sets of vocabulary to users. Clearly, this doesn't cover all combinations of vocabulary across categories, but it provides a form of intelligent randomization to ensure particular words within categories won't always appear together and in the same session.

For each subject, roughly two and a half hours of time split into three sessions was required. The three sessions were each spaced four to seven days apart from each other. We required this break between sessions as our goal was not to measure retention in short-term memory, but how the games could affect retention over a longer period of time.

In our design, we intended to counterbalance ordering effects by having half the subjects use the listening-only version first, while the other half used the speaking version first. Combining this ordering consideration with the 20 vocabulary divisions, we had a desired subject count of 40. We weren't able to get a full 40 users in our study, but our method for assigning the 16 that we had into such slots ensured that effects were decently counterbalanced.

Table 4.2: Sequence of tasks for a subject seeing the "listening-only" version first

Session	Task	Card set	Time allowed
Session 1	1. Audio Test	N/A	∞
	2. Survey A	N/A	∞
	3. Word War	1st set	∞
	4. Rainbow Rummy Listening Tutorial	Tutorial set	15 min
	5. Rainbow Rummy Listening Game	1st set	30 min
Wait 4-7 days			
Session 2	6. Word War	1st set	∞
	7. Audio Test	N/A	∞
	8. Word War	2nd set	∞
	9. Rainbow Rummy Speaking Tutorial	Tutorial set	15 min
	10. Rainbow Rummy Speaking Game	2nd set	30 min
Wait 4-7 days			
Session 3	11. Word War	2nd set	∞
	12. Survey B	N/A	∞

The complete schedule for a speaking-first subject is shown in Table 4.1, while the schedule for a listening-first subject is shown in Table 4.2. As subjects would be unfamiliar with the game rules and speech interface, we required them to follow a tutorial before playing each game. The tutorial instructed subjects on what moves to make and how to make them for the first couple of turns, then allowed them to play the rest of the game on their own. The set of words used in the tutorial were distinct from the 48 test vocabulary items and the same across all subjects. We wanted to ensure that the learning curve of each version of the game did not affect the results.

In order to make sure sessions did not last too long, we allowed two ways for the user to complete a Rainbow Rummy task and continue to the next. For a tutorial, the condition for completion was that the subject either complete one game or play for 15 minutes. For a game, the condition was that the subject either complete one game or play for 30 minutes. As the game contained the actual vocabulary we were examining for acquisition, the time allowed was longer.

Rainbow Rummy can predictably generate a game arrangement based on a supplied random seed. In order to ensure that the games in this study would not finish immediately, we used three different random seeds that were sure to generate accept-

able games. One of the seeds was used for all tutorial games. The two other seeds were used in the Rainbow Rummy games, where the user would use one seed for their first game and the other seed for the second game. The particular seeds were assigned evenly between the speaking and listening-only versions so that biases would not arise from one game being easier than another.

For organizing the subjects in the study, we developed a web-based management system in Ext-GWT and Java (See Appendix A for more details). From our perspective, this was useful in managing accounts, assigning users to tasks, organizing experimental groups, and setting options for various instances of the games. As our study was conducted over the period of 8-14 days for each user where tasks had to be enabled at different points in time, we used this system as an interface for easily viewing and managing such information. From the user perspective, this system allowed self-signup for the study and a single point from which they could access and complete their tasks.

In short, each subject followed a task sequence in which they were exposed to both the "speaking" and "listening-only" versions of Rainbow Rummy. Additionally, 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. Analyzing the difference between pre- and post-tests for one week and comparing it to the difference for the other, we could compare the two Rainbow Rummy systems in their effect on learning.

Learning can definitely take place *during* the Word War pre-test, since the subjects are expected to execute the correct actions. However, as the subjects were exposed to such a test in *both* the speaking and listening-only systems, and we were only interested in the difference between the differences in pre- and post-test scores, this learning effect would not affect our results.

4.3 Launching the Study

We launched the study and sent an e-mail to users who had previously signed up for an account with the Chinese Cards system encouraging them to participate. We received a quick response and easily filled the maximum 40 slots available. However, early interactions with Rainbow Rummy revealed some unexpected problems.

To see how the subjects were interacting with the system, we enabled the "playback" feature as supported by the WAMI architecture, so as to visually re-play user sessions. This allowed us to listen to user utterances and view complete games as they were actually experienced by subjects. This playback feature works by creating a new Rainbow Rummy session and controlling all interaction between client and server using events as logged in the database, according to the appropriate timing.

Viewing early interactions with the system, we found that a user had trouble being recognized, while the sound clip did not seem to be of poor quality. We realized that our initial grammar had been too relaxed in allowing certain word and phrase orderings, leading to new users having particular difficulty with misrecognition. At first, this was surprising, since we had never seemed to personally encounter problems being recognized, though we are non-native speakers. However, our suspicion was that we had become so accustomed to using the interface that our pronunciation matched what was expected by the recognizer much more closely than any new user could be expected to. Secondly, having tested the system with our microphones and in relatively similar environments, we couldn't guarantee the same level of performance in other setups.

Thus, we restricted the grammar slightly to only allow phrases for selecting card by color, then vocabulary word, as opposed to also allowing the ordering of vocabulary item, then color. This restriction did not affect game play, and in fact, viewing sessions of early users showed that the color, then vocabulary word ordering was the dominant phrase by far. Having viewed a few early user sessions, we then re-ran the logged sound clips through the recognizer to see how the recognition results improved. The changes not only affected Rainbow Rummy, but also made it slightly easier for

users to pass the Audio Test, while still allowing it to serve as a decent filter. With the appropriate modifications to the grammar in place, we were ready to re-launch the study.

The subjects used for tweaking the system in this stage were not included in final group of users for analysis. We can justify the modification of the grammar for use in the study since our primary objective was to measure learning gains, and not recognition performance.

4.4 Results

In our assessment of learning gains, we used the total Word War scores as our metric of vocabulary knowledge in Equation 4.1 across paired pre- and post-test sessions.

$$G_v = (s_2 - s_1)/(MaxScore - s_1) \tag{4.1}$$

where v is either "speaking" or "listening-only" 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).

Every subject showed an improvement in their Word War score for their post-test as compared to their pre-test. The result suggests that our Word War metric is indeed a valid one for measuring knowledge of vocabulary, as it's natural to assume that learning would take place across sessions.

For each subject, we then calculated the learning gain averaged over all subjects for the speaking version of Rainbow Rummy as contrasted with the average gain for the listening-only version. We found the average gain for the speaking version was significantly greater ($p < 0.05$) than that for the listening-only version. Furthermore, this difference (0.5155 for speaking, 0.3684 for listening-only), is of meaningful magnitude. This result seems to support our hypothesis that the two versions have different effects on learning gain, with speaking as the more effective version. To gain further insights, we examined the average learning gains for various meaningful

Table 4.3: Learning gains for the "speaking" versus "listening-only" system. **Bold** indicates statistical significance at $p < 0.05$.

Subject set	System version	Avg learning gain	p-value (1-sided)
All subjects	Speaking Listening-only	0.5155 0.3684	**0.0259**
Listening first	Speaking Listening-only	0.5745 0.4282	0.1027
Speaking first	Speaking Listening-only	0.4695 0.3219	0.0887
Speaking preferred	Speaking Listening-only	0.6046 0.2864	**0.0198**
Listening-only preferred	Speaking Listening-only	0.4620 0.4176	0.2802

Table 4.4: Rainbow Rummy game statistics

Subject Type	Game version	# moves on user turn	# moves on computer turn	# times hint clicked
Listening-only first	Listening-only Speaking	191 171	69 89	55 155
Speaking first	Speaking Listening-only	249 174	102 105	287 37

subject subsets (See Figure 4-5 and Table 4.3).

One variable was whether the student's first game was the speaking version or not. Both groups showed essentially the same relative benefit from speaking as in the pooled data, though because the subject count was cut in half, the results were not statistically significant. Interestingly, those who began with the listening-only version had higher average gains for their games as compared to those who started with the speaking version. Although the result was not significantly higher (0.11 p-value), it suggests that the sequence of listening-only then speaking was a more effective means for training the user to play the game and learn the words. As the speaking mode places a higher cognitive load on the user, the ordering of listening-only before

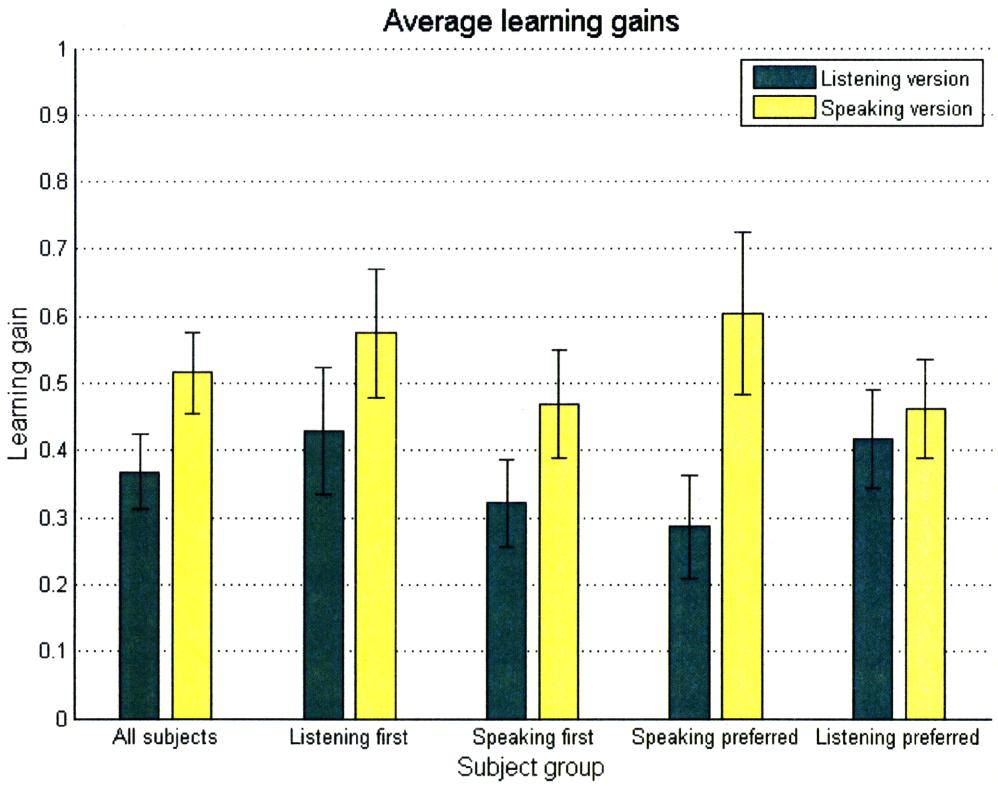


Figure 4-5: "Speaking" vs "listening-only" learning gains based on various subject subsets

speaking seems reasonable as the more effective sequence.

We also examined the relationship between interactions in the Rainbow Rummy games and their impact on learning gain. Having playback available, we viewed a sampling of sessions and confirmed that subjects were able to play Rainbow Rummy as it was intended to be played.

Knowing that users were indeed able to play Rainbow Rummy, we collected a number of statistics on the interactions within the game, as seen in Table 4.4. A first look at the data reveals significant differences in the moves performed under the different subject conditions. For the users who saw the speaking system first, the number of moves performed on the subject's turn was much higher at 249 than for any of the other conditions. One possible explanation for this figure is that subjects were unfamiliar with both the game and speech interface, causing a much greater number of accidental moves.

As seen in Table 4.4, for those who saw the listening-only version first, less 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. This increase in learning gain even amidst less moves performed suggests increased gains per move through use of speech as opposed to passively listening.

The number of moves on a computer player's turn, on the other hand, is roughly the same for all conditions. The most striking difference among these statistics is in the number of times a hint was clicked for the different types of games. The number of times that a hint was clicked while playing a speaking game was significantly greater than the number of times it was clicked in the listening-only games, for both speaking-first and listening-first subject types. This can possibly be explained in that speaking requires the user to know a word's pronunciation, thus requiring a hint click, while listening moves are performed freely and the user listens passively, not requiring the click of a hint.

The mismatch in amount of hint clicks would suggest that the increased number of times a user clicked hints could be the reason for the observed increase in Word War scores. Figure 4-6 shows a graphical representation of the relationship between hint clicks and learning gain. Indeed, the listening-only games are clustered at the bottom left, with fewer hint clicks and lower learning gains than the speaking data points.

The intuition would be that the greater number of times a hint is clicked, the more a user would be exposed to the written pinyin for a word. Thus, it could be that this factor is the actual reason for the higher learning gain in the speaking version, rather than the use of speech itself. At the same time, we found that within speaking games, there is an interesting *inverse* relationship between learning gain and hint clicks. The correlation coefficient for this relationship is $\rho = -0.6758$, with $p = 0.0041$ for the null hypothesis of no correlation. This relationship exists within speaking games, so it does not reject the notion that hint clicks could account for the increased learning gains as observed in the speaking versus listening-only versions of the game. However, the result presents an alternate idea that increased hint clicks may not actually be

the reason behind the higher learning gains.

We attempted to look for trends in learning gain as correlated with other game statistics such as how many times a card was moved on the user's turn, and how many times a card was moved on the computer player's turn. We were particularly interested in the figures for the number of times cards were moved on a user turn, as the means for moving a card in this manner was a direct difference between the listening-only and speaking versions, where the move was either parroted by computer speech or spoken by the user himself.

However, we found no significant trends to correlate learning gain with these factors. In particular, we found it interesting that the number of moves performed on a user turn was not correlated with the learning gain (See Figure 4-7).

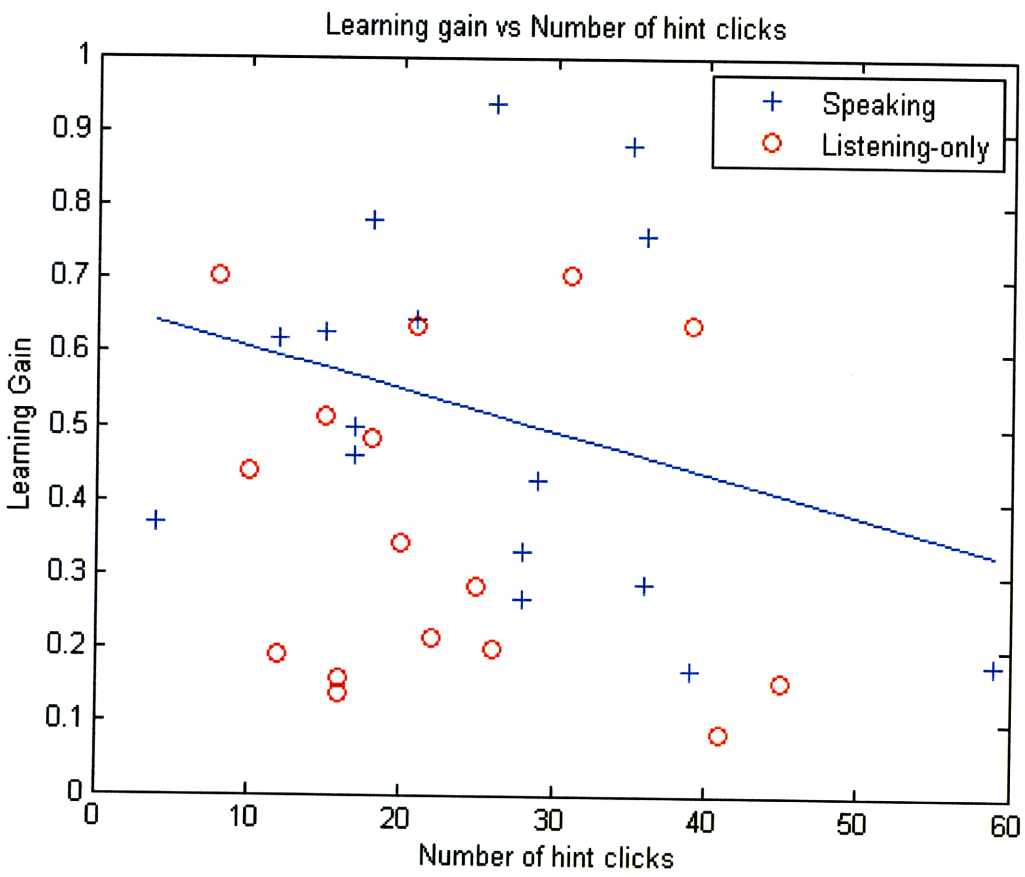


Figure 4-6: Learning gain versus number of hint clicks

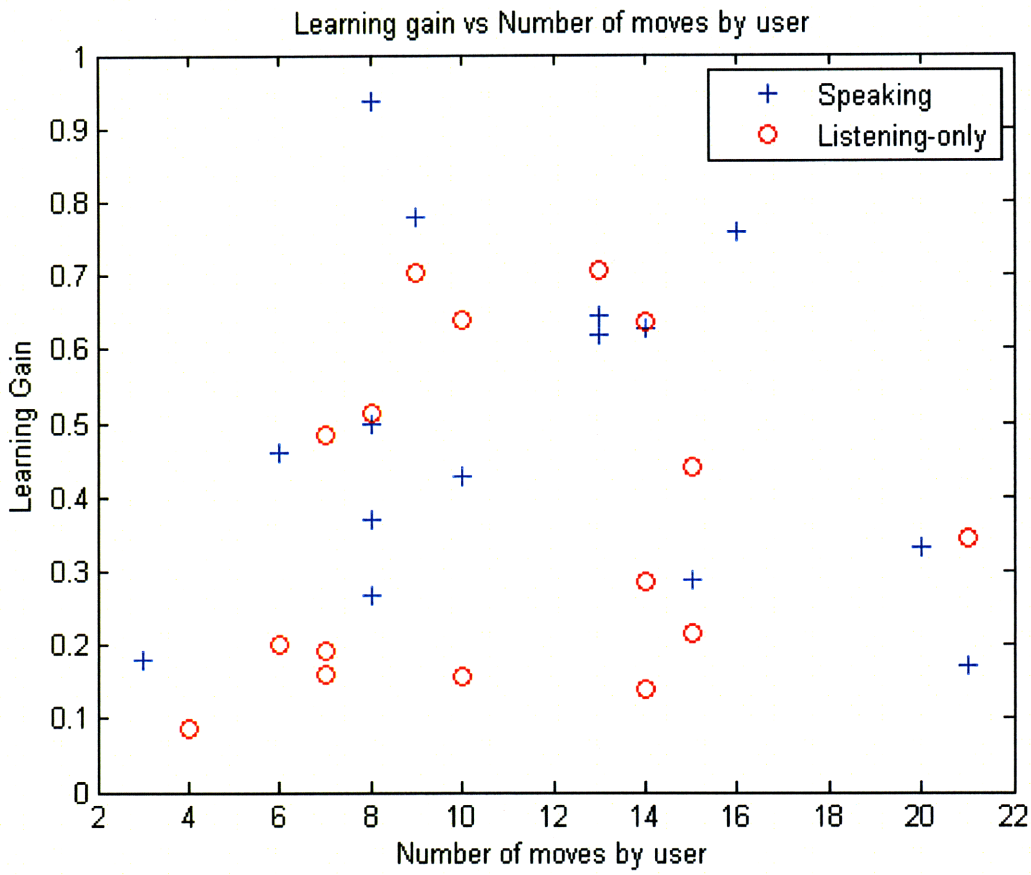


Figure 4-7: Learning gains versus number of moves by user

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-only version?". Roughly half of the subjects preferred speaking, while the other half preferred listening-only. 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. These results indicate that those who enjoyed using speech in Rainbow Rummy were in fact the ones who benefited the most from its use in terms of learning.

We also found no significant correlation between the number of years a subject studied Chinese and the learning gain for the two systems. As seen in Figure 4-8, our subjects ranged in experience from 0.5 to 5 years of Chinese study. Yet, the amount of learning gain is extremely varied even amongst individuals with the same number

of years of study. On the one hand, we may not have enough data to make any claims. However, the lack of significant correlation suggests that learners of all levels of experience can use the system to varying degrees of benefit.

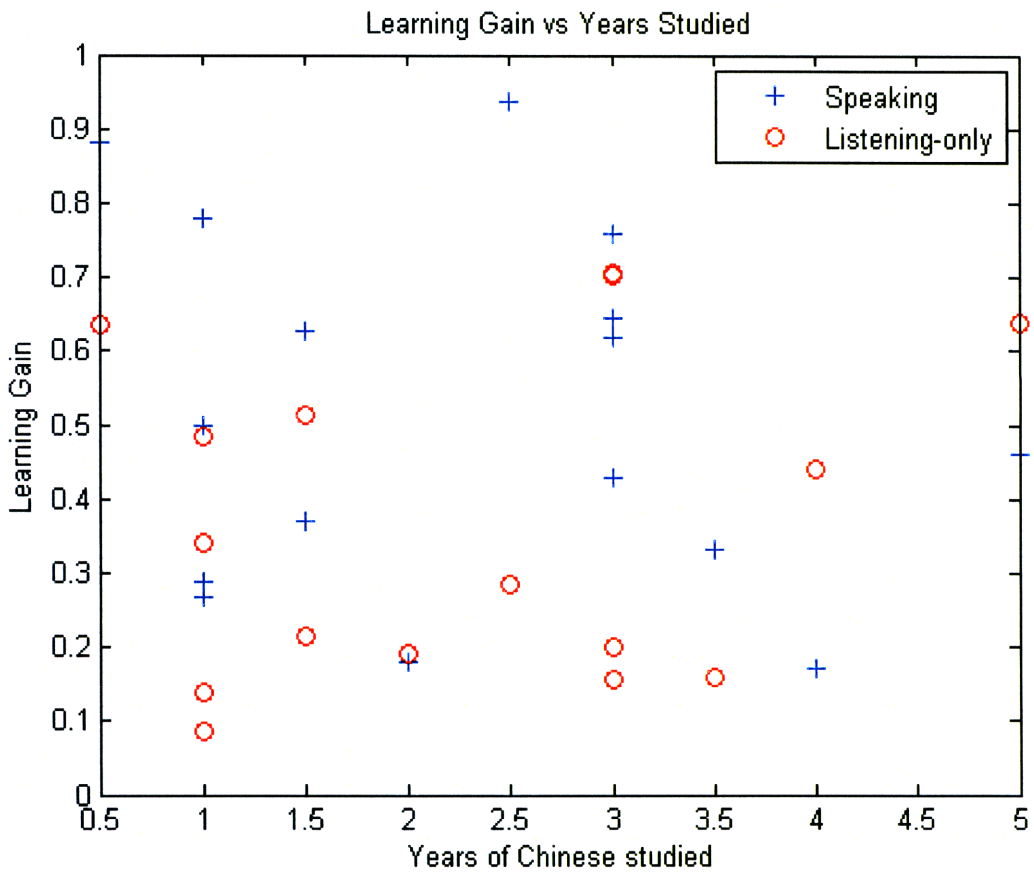


Figure 4-8: Learning gains versus Chinese experience

We examined the response to the survey question: "Which game did you enjoy playing more? Word War or Rainbow Rummy?". The results show that roughly half of the subjects preferred Rainbow Rummy, while the other half preferred Word War. Interestingly, all of those who preferred the speaking version of Rainbow Rummy chose Rainbow Rummy over Word War when posed with this question. As Rainbow Rummy was created with the intent of being used in the speaking version, it shows promise for serving as a more entertaining alternative for those who prefer to learn using speech. For those who prefer the listening-only version of Rainbow Rummy, subjects overwhelmingly found Word War to be more entertaining, though one subject

preferred Rainbow Rummy even while enjoying the "listening-only" version more. As the preferences line up in this way, we feel that aversion to speaking in Rainbow Rummy, where the sentences are more complex, may have pushed these users toward preferring Word War.

Despite the split responses, we believe Rainbow Rummy still has more intrinsic entertainment value when the learning curve for each is factored out. Two subjects expressed preference for Word War, but also admitted that they were only just starting to get the hang of playing Rainbow Rummy, so they felt more time was needed. Their reaction confirms our suspicion that Rainbow Rummy has a much steeper learning curve than Word War. If enjoying the speaking mode is an indication of having mastered the game mechanics, then Rainbow Rummy is a clear win.

Secondly, the subjects were not exposed to the speaking version of Word War, so a more accurate test would be to ask users to compare the speaking version of both Word War and Rainbow Rummy against each other. It is also important to note that the study version of Rainbow Rummy, which used a "match by category" rule, was a modification of the original "match by vocabulary item" version of the game. It may be the that study version was found to be difficult, or recognition was harder given the larger vocabulary set. In the face of all these issues, we cannot make any strong claims for either game as being more intrinsically entertaining.

It is interesting that while the Word War game was adapted and designed for use as pre- and post-tests, some subjects actually enjoyed it more than the game of interest, which was Rainbow Rummy. For the future, this technique of using games as the pre- and post-tests may be a good way to keep subject interest when faced with conducting user studies remotely over the web. Secondly, when interpreted by users as an engaging game, as it seems to have been here, it is reasonable to assume that the subjects would be less likely to cheat by looking up words in alternate websites, as that they don't feel like they are being evaluated. In running a study remotely, discouragement of cheating is an important concern for the sake of acquiring meaningful data. In our case, Word War appears to have been a success.

4.5 Summary

Through this user study, we were able to provide rough answers to our three motivating questions. First, as seen in the positive learning gains for all subjects across pre- and post-tests, we find that Rainbow Rummy can indeed be an effective environment for learning new vocabulary. Considering the subjects played Rainbow Rummy for a maximum of only 30 minutes and waited at least 4 days until the post-test, the results are promising.

Second, we showed that games in which the user was required to speak had significantly higher average learning gains than those in which the user could only passively listen. In particular, those who enjoyed the speaking version over the listening-only version showed even greater additional learning gains through using the speaking version. Thus, the enablement of speech in Rainbow Rummy shows the greatest promise in serving this set of learners, which was roughly half of the subjects in our study.

Third, we found that Rainbow Rummy was overwhelmingly considered to be more fun for those subjects who enjoyed the speaking version. For those who preferred the listening-only version, Word War was preferred. Upon examining this split, we cannot conclude a strong preference for one or the other.

Finally, our method for conducting the speech-enabled study remotely demonstrates a useful way to collect data from diverse populations. In the face of difficulties which could arise due to uncontrollable subject-side factors, the study ran quite smoothly. The high interest which was shown upon launching the study reveals the existence of a useful resource for data collection when laboratory conditions are not needed.

Chapter 5

Final Remarks

5.1 Conclusion

In this thesis, we designed and implemented a speech-enabled game for vocabulary acquisition. In particular, we focused on Mandarin Chinese, but the system is configurable for other languages. The game is deployed over the web and incorporates custom user content embedded within an interactive turn-based session with other players. Users can configure the game to play against other users in multi-player mode while each person simultaneously learns his/her own content. We deployed Rainbow Rummy over the web for public use, observing its potential for use as a practical tool. We also conducted a user study which validates its relevance as a tool for vocabulary acquisition and suggests that the enablement of speech in the interface plays a positive role in learning.

5.2 Future Work

In the future, Rainbow Rummy could be improved in a number of ways. Several subjects in the study commented on the poor synthesizer audio quality. Future work can replace the synthesizer with one that responds faster or more naturally. Currently, our application only uses the top recognition result rather than taking full advantage of an n-best list of results. For reducing misrecognition errors, we could examine

typical errors made by users and adapt our application to improve the user experience.

The study version of the game uses a "match by category" rule, but the Chinese Cards interface currently does not support selection of multiple categories. To enable the study version of Rainbow Rummy for use by the general public, we can work on this extension for the Chinese Cards interface.

To examine the results of the user study in more detail, we could work on annotating the utterances logged during the Rainbow Rummy sessions. The annotations could be used to evaluate recognition error rate. Furthermore, labeling of moves could provide useful information for examining how learning gain correlates with this more detailed set of information.

The Rainbow Rummy project has its place within the broader goal of developing speech-enabled tools for language learning which can have a practical impact on students. The Spoken Language Systems group at MIT will continue to develop systems which explore the place of speech recognition in such systems. Future work in this broad area could involve creation of new games for learning based on the Chinese Cards framework, building upon examples such as Word War and Rainbow Rummy. Generating greater interest in the current systems would be useful for further evaluation of speech and its role in learning, as well as for collection of data on non-native Chinese speech.

Appendix A

User Study Management

For managing the user study we adapted a previously used management system built in the open-source MyGWT library. We ported features into Ext-GWT, the updated version of MyGWT, and added a new navigation structure. To parameterize the games for use in the study and create survey screens, we created a form builder as seen in Figure A-1.

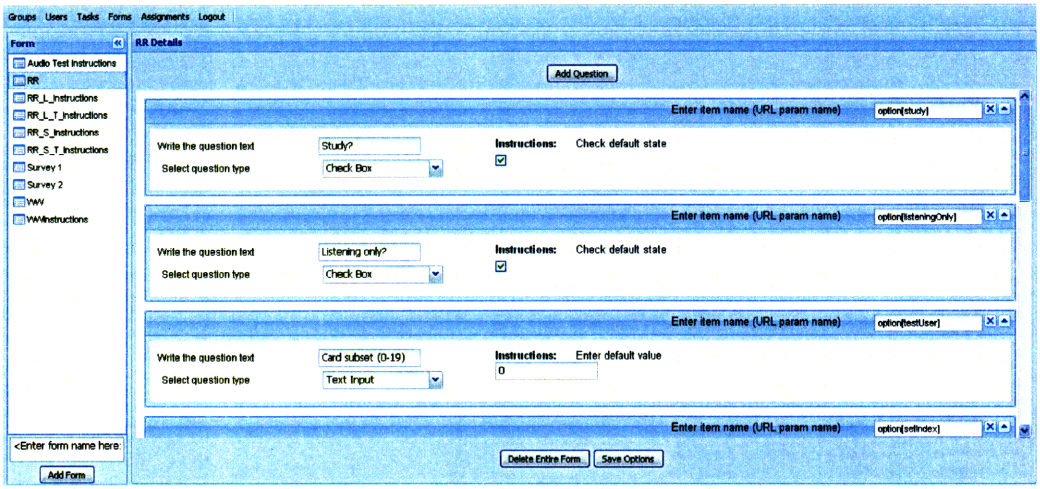


Figure A-1: Management system: screenshot of forms page

We then created "tasks" for each of the unique parameter sets used in the study to set up the various games. Tasks were then assigned to particular groups. Each group was assigned a unique set and ordering of tasks for the study. In addition, parameters could be set for the group such as a limit on the number of users, the

message displayed to users upon signup, etc. (See Figure A-3). From this screen, we were able to assign tasks to particular groups. As our study restricted access to later sessions' tasks until 4 days after the first session, we included an option to disable users from having access to certain tasks upon signup.

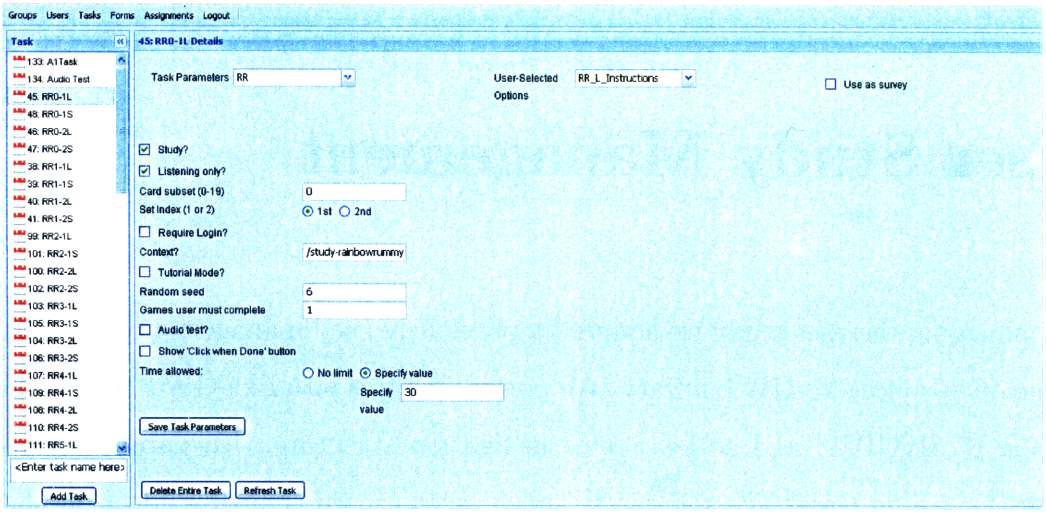


Figure A-2: Management system: screenshot of tasks page

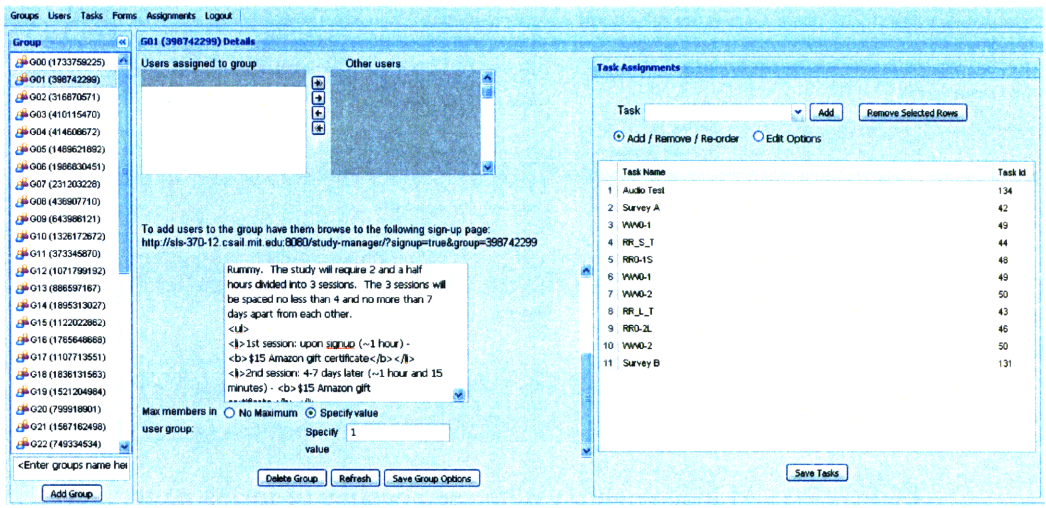


Figure A-3: Management system: screenshot of groups page

We also included a page for viewing user-specific information. From this page, we were able to enable tasks associated with the group to which the user belonged on a per-user basis. This was important, as each user completed their first session on

different days, depending on their pace. From this page, we could also easily view the status of each user on their tasks, including: not started, started, time expired, finished, and disabled.

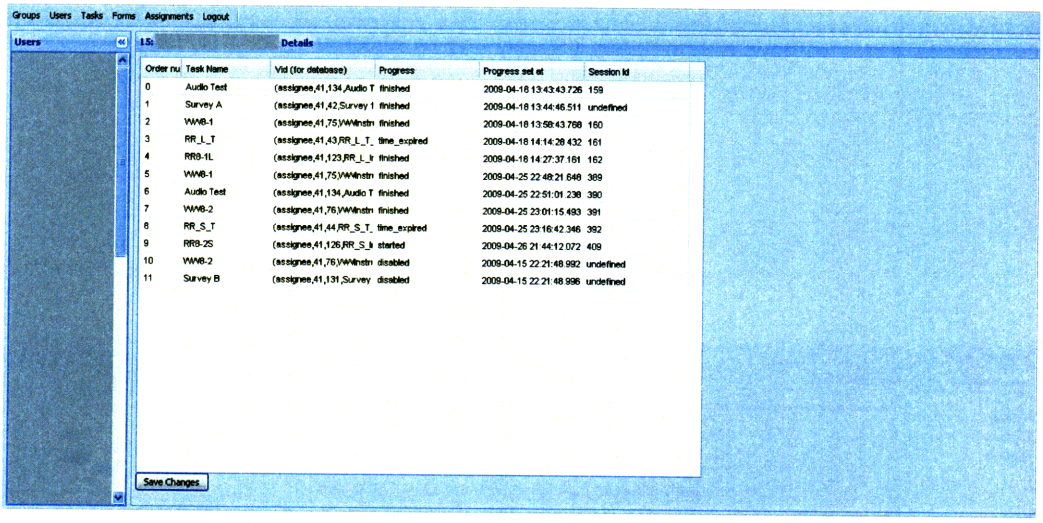


Figure A-4: Management system: screenshot of users page

Subjects were given access to a single URL from which they could review participation information and sign up for an account with the study. The system automatically placed them in the group with the least number of members in order to even out the assignment. The interface from which the subjects accessed their tasks is shown in Figure A-5.

This system was used effectively for the Rainbow Rummy study, as well as in another Spoken Language Systems Group study conducted on a system called Flight Browser. Information relevant to the management system and the applications being tested are logged in the database. In the case of WAMI applications, all events and recordings are logged in a standard format. To extract useful information of interest from the logs, appropriate SQL queries were written.

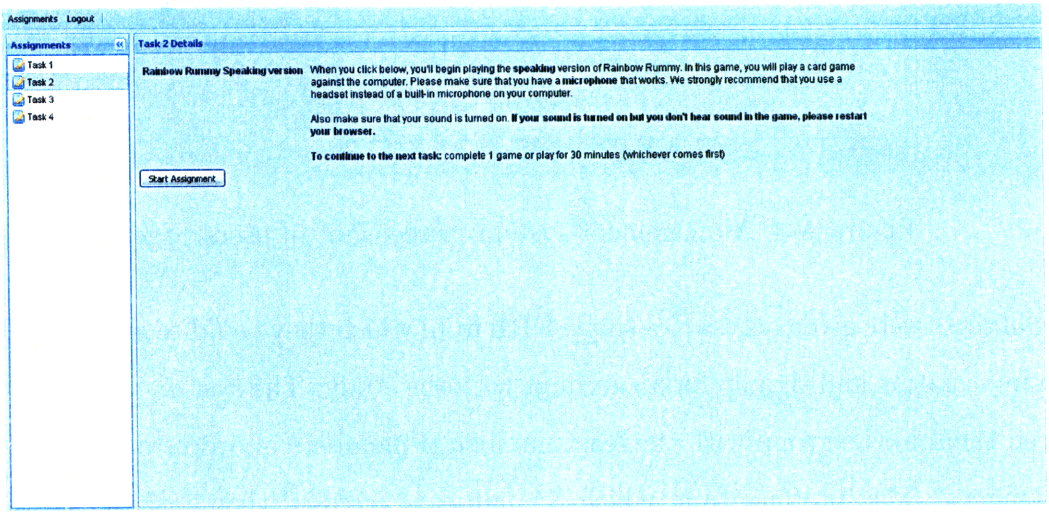


Figure A-5: Management system: screenshot of assignments page

Bibliography

[1] F. Ehsani, J. Bernstein, A. Najmi, and O. Todic. Subarashii: Japanese interactive spoken language education. In *Proceedings of Eurospeech*, pages 681–684, September 1997.

[2] F. Ehsani and E. Knodt. Speech technology in computer-aided language learning: Strengths and limitations of a new call paradigm. *Language Learning and Technology*, 2(1):54–73, July 1998.

[3] Rod Ellis and Xien He. The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21(2):285–301, 1999.

[4] Jim Glass. Computer speech and language. *Computer Speech and Language*, 2003.

[5] Alexander Gruenstein, Ian McGraw, and Ibrahim Badr. The wami toolkit for developing, deploying, and evaluating web-accessible multimodal interfaces. In *Proceedings of ICMI*, Chania, Crete, Greece, 2008.

[6] Jennifer Howard. Enrollments in foreign-language classes continue to rise. `http://chronicle.com/daily/2007/11/691n.htm`, November 2007.

[7] Ilene Kantrov. Talking to the computer: A prototype speech recognition system for early reading instruction. Technical Report 91-3, Education Development Center, Inc., Newton, MA, 1991.

[8] Ian McGraw, Brandon Yoshimoto, and Stephanie 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.

[9] U.S. Department of Education. Teaching language for national security and american competitiveness. `http://www.ed.gov/teachers/how/academic/foreign-language/teaching-language.html`, January 2006.

[10] U.S. Department of State. National language security initiative. `http://exchanges.state.gov/nsli.html`, April 2009.

[11] Paul Pimsleur. *How to learn a Foreign Language*, chapter 10. 1980.

[12] U.S. Senate. Year of foreign language study. 109th Cong., 1st sess., S. Res. 28, February 2005.

[13] Stephanie Seneff and Chao 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, October 2007.

[14] Rosetta Stone. `http://www.rosettastone.com/`, April 2009.

[15] Merrill Swain. Communicative competence: Some roles of comprehensible input and comprehensive output in its development. *Input in Second Language Acquisition*, pages 235–253, 1985.

[16] Luis von Ahn. Games with a purpose. *IEEE Computer*, 39(6):92–94, 2006.

[17] S. Marsella W.L. Johnson, H. Vilhjalmsson. 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.