





# **Chinese Scrabble: A Web-based Speech-enabled Game for Chinese Vocabulary Building**

by

Zuzana Trnovcova

Submitted to the  
Department of Electrical Engineering and Computer Science

August 6, 2010

In Partial Fulfillment of the Requirements for the Degree of  
Master of Engineering in Electrical Engineering and Computer Science

## **Abstract**

The subject of this thesis is a web-based language game Chinese Scrabble, whose main objective is to help students of Chinese to practice speaking, to learn and review vocabulary both in pinyin and in Chinese characters. The game was intended to be very flexible and customizable in order to accommodate a wide range of users on a long-term basis. As a part of the project, we conveyed a pilot user study with both students of Chinese and native Chinese speakers to evaluate how enjoyable the game was and what aspects of language it may teach.

Thesis Supervisor: Stephanie Seneff

Title: Senior Research Scientist, MIT Computer Science and Artificial Intelligence Laboratory

## Table of Contents

<b>1. Introduction .....</b>	<b>6</b>
<b>2. Previous Work .....</b>	<b>7</b>
<b>3. Game Design and Implementation Iterations .....</b>	<b>8</b>
3.1 Game Architecture .....	8
3.2 Game Vocabulary – Sources, Processing and Level Definition .....	9
3.2.1 Choosing Game Vocabulary – Attempt 1 .....	10
3.2.2 Choosing Game Vocabulary – Attempt 2: .....	12
3.3 Randomized Generation of a Game at a Chosen Level; Solution Regeneration .....	15
3.3.1 Game Generation Techniques .....	15
3.3.2 Solution Regeneration .....	18
3.4 Basic User Interface .....	19
3.5 User Interaction .....	24
3.6 Game Logic and Intelligence .....	27
3.7 Scoring .....	30
3.7.1 Word Scoring and Total Score .....	31
3.7.3 Level Advancement Options and Predicted Level .....	31
<b>4. User Study.....</b>	<b>33</b>
4.1 User Study Design.....	33
4.2 Playback and User Data Logging.....	34
4.3 User Data analysis.....	35
4.3.1 Are the Users Happy with the Game? .....	35
4.3.2 Do the Users Think that the Game Is a Useful Learning Tool? .....	36

4.3.3 Do the Users Use Speech and Tones and How Do They Like It? .....	37
4.3.4 Do the users use the help features and how do they like them? .....	38
4.3.5 Do the users use mostly words of the current level? Was the Automatic Level Choice Effective?.....	42
4.6 Do more advanced users do better in the game? .....	47
<b>5. Conclusion .....</b>	<b>52</b>
5.1 Lessons Learned.....	52
5.2 Future Work .....	53
<b>Appendix A: Chinese Characters from School Grades 1-6 .....</b>	<b>54</b>
<b>Appendix B: User Survey .....</b>	<b>56</b>
<b>References.....</b>	<b>63</b>

# 1. Introduction

In today's globalized world, it is very helpful and often essential to speak multiple languages. Most people learn foreign languages in school. This requires a certain amount of memorization, consistent studying habits and a lot of practice in speaking, listening and reading. The best way to encourage studying and practice consistency is to make language learning fun, for example by encapsulating it into a game.

Chinese Scrabble is a web-based language game whose main objective is to help students of Chinese practice speaking Chinese, review old and learn new vocabulary and strengthen their knowledge of Chinese characters. The goal of the game is to create  $(N-1)$  two-syllable words from  $N$  toneless pinyin syllables. The words are put on a game board one-by-one, such that the toneless pinyin syllables represented by characters on the board can be reused to create new words that can be placed next to existing words, such that their common toneless pinyin syllables touch. This rule provides an opportunity to both create words from pinyin syllables and to read words on the board in Chinese characters.

The game can be played with speech or by typing and with or without tones. The user is guided through the game by a set of instructions and hints displayed near the game board. In case the user does not know how to continue, there are several types of game help available – Chinese character reading and English meaning, English word help and direct help that points to places where new words can be placed.

The vocabulary for Chinese Scrabble consists of more than 19,000 words from New York Times news corpus merged with over 4,000 words from IWSLT corpus [1]. The words were ranked according to their frequency in both corpora as well as according to the commonness of their characters as defined by group of characters taught in grades 1 through 6 in China. The game can be played at different levels corresponding to words of different word score ranges. The user gets a score, based on which the game may move him up or down several levels.

Chinese Scrabble has been tested on one pilot user study with 7 users of different Chinese language backgrounds, ranging from intermediate students to Chinese native speakers. The user study yielded encouraging results suggesting that the game has a potential to provide a long term entertainment and educational tool for a very wide audience.

## 2. Previous Work

Nowadays, there are many websites offering Chinese language games [4]. These range from the games for beginners including only very basic vocabulary (for example, colors, numbers, or family vocabulary like in [5]) through role playing games [6] to manually created Chinese crosswords [7].

Moreover, in the past few years, several educational foreign language games have been developed at the Spoken Language Systems group at MIT CSAIL. They range from Chinese-English translation and dialogue games to Chinese-only games, like the Family dialogue system, that proved to be capable of teaching Chinese even to the people who never studied it before playing the game [3, 10, 11, 12]. For intermediate users, the card games Word War [13, 14] and Rainbow Rummy [9] provide both the opportunity to practice user-defined vocabulary as well as use Chinese dialogue to move the cards on the game grid. In all these games, the competitive atmosphere of the game helps sustain the student's interest and facilitate his or her learning.

In [8], a team of researchers at the University of Hong Kong describe a game for teaching Chinese character structure, morphological and phonological similarities as well as stroke order. They argue that teaching Hong Kong children similar characters at once, especially characters sharing phonological or morphological strokes, improves the childrens' character learning and reading skills by equipping them with the skill of decomposing the characters into several very basic characters, which the children already know.

Chinese Scrabble is a little bit similar to Chinese crosswords and the games in [8] in that that the board looks like a non-overlapping crossword and the point of the game is, besides creating words from pinyin, to associate pinyin with Chinese characters and vice versa.

## 3. Game Design and Implementation Iterations

### 3.1 Game Architecture

The three main building blocks of the game are the Java servlet backend, which processes the game data, the Javascript frontend, which presents a particular game to the user, and the WAMI speech recognizer [15], which closely interacts with the Javascript and lets the user play the game by speaking and clicking only.

The role of the Java servlet backend is to first load the game data from two processed and translated corpora and then randomly generate a game of the requested difficulty level and number of tokens each time the user starts a new game.

After the Javascript frontend receives the game data from the servlet, the entire game is handled locally on the client. This makes the game run pretty smoothly since there are very few data transfers between the server and the client<sup>1</sup>. Javascript takes care of giving the user instructions, hints and feedback, processing user input, generating appropriate help, assigning game points and choosing or letting the user decide which level to play next. Javascript also initiates data logging for playback and further analysis and it interacts with the WAMI speech recognizer<sup>2</sup>.

WAMI is responsible for interpreting the user's speech and sending the best recognition result to the Javascript input-processing methods where the recognized word is handled similarly to the text input.

---

<sup>1</sup> excluding user data logging – see section 3.8

<sup>2</sup> the existing javascript WAMI API was one of the main reasons for choosing javascript for this project as opposed to Flash or other language for programming dynamic websites



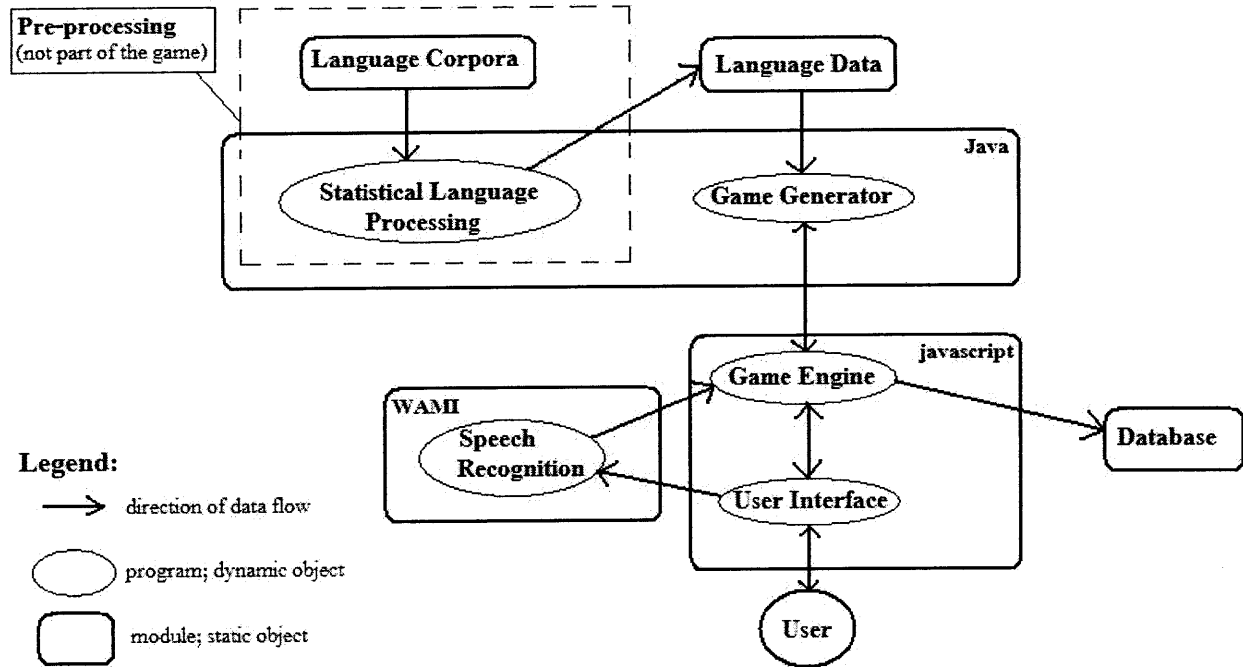


Figure 1: High-Level Game Architecture

### 3.2 Game Vocabulary – Sources, Processing and Level Definition

In order to make Chinese Scrabble attractive for students with a wide range of Chinese language skills, the game needed to work with a vocabulary set that was comprehensive and at the same time provided easy access to words associated with certain difficulty levels. Having a vocabulary source with these two properties would obviously help to generate a variety of games at a chosen difficulty level. However, choosing the appropriate vocabulary source was by no means a straightforward task, especially in the beginning stages of the project when the priority was to quickly develop a prototype game generator to assess if the game was feasible (with any vocabulary). Moreover, the definition of the “word difficulty” as well as of the “game level” was likely to depend on individual students.

In the interest of time, the most readily available data (IWSLT travel conversation corpus [1]) was used initially. Only after several attempts and mediocre results in creating desired games using only this corpus, more data (a larger news corpus and a list of the most basic Chinese characters used in Chinese grade school classrooms) was added to the game. In this

section I will describe the process and rationale employed on the way to obtaining the best difficulty-annotated vocabulary for Chinese Scrabble based on its effectiveness (content) and efficiency (structure).

The vocabulary used in Chinese Scrabble was selected from two text corpora (travel and news corpus) based on combined word frequency and individual character usage as defined by a list of characters taught in China in elementary school grades 1-6. Only 2-syllable (that is, 2-character) words with 2 distinct toneless pinyin syllables were used for the game.

### 3.2.1 Choosing Game Vocabulary – Attempt 1

#### 3.2.1.1 Limited data and pinyin-frequency biased word score

The first data source for the game was the IWSLT travel conversation corpus [1]. The corpus consisted of various utterances related to travel, which were given both in word segmented Chinese characters and word segmented pinyin representation in two separate files (referred to as “character file” and “pinyin file” below). The size of the corpus was over 170,000 words, out of which 4865 distinct 2-syllable words became the first candidates for Chinese Scrabble vocabulary.

The most intuitive measure of a word difficulty seemed to be the number of occurrences of the word in the corpus. However, because the corpus was strongly biased toward travel-related words and thus did not reflect the general word usage patterns of Chinese language, the word count alone didn’t pass the criteria of a sufficient word difficulty metric. Since Chinese Scrabble is based on relating words through their common toneless pinyin syllable, a feasible word ranking for the game seemed to be one based both on the word counts and on the toneless syllable counts across all 2-syllable word occurrences in the corpus. Thus, the following formula for the score of a word  $w$  was used:

$$score(w) = \log \frac{count(w)}{totalWordCount} + 1/2 \left( \log \frac{count(w.pinyin[1])}{totalSyllCount} + \log \frac{count(w.pinyin[2])}{totalSyllCount} \right)$$

Here, *totalWordCount* is the number of occurrences of all 2-syllable words, *totalSyllCount* is the number of syllables in all 2-syllable word occurrences (which implies that  $totalSyllCount = 2 \cdot totalWordCount$ ) and  $w.pinyin[i]$  is the  $i^{th}$  toneless pinyin of word  $w$ .

The scores ranged from -7.59 for “zhe4ge5”, meaning “this one”, to -20.742 for “huang1miu4”, meaning “absurd” or “ridiculous”.

### 3.2.1.2 Vocabulary-related Data Structures

For Chinese Scrabble purposes, each word needed four pieces of information: pinyin representation, character representation, English meaning and word score. Moreover, each pinyin needed its frequency to be readily accessible for the purpose of score calculation.

Initially, the game seemed to be driven by pinyin representation of words and syllables, so the effort was focused on quickly accessing the pinyin representation in a specified score range and subsequently mapping these to the corresponding character representation(s) and English definition(s).

Therefore, the initial data structures used were the following maps and sorted sets:

*wordToCount*: map with words in pinyin as keys and word counts as values

*wordToCountS*: *wordToCount* entries sorted according to word counts

*plainPinyinToCount*: map with toneless pinyin as keys and toneless pinyin counts as values

*pinyinToChinese*: map with words in pinyin as keys and lists of corresponding 2-character words as values (since a pinyin word representation can have multiple possible character representations)

*chineseToEnglish*: map of 2-character Chinese words to their English definitions

*wordToCount*, *wordToCountS* and *plainPinyinToCount* were obtained by scanning the pinyin file once in time  $O(\text{corpus size})$ . *pinyinToChinese* map was created by scanning the pinyin and character files in parallel – also in time  $O(\text{corpus size})$ . *chineseToEnglish* map was

obtained by feeding all words in character representation into *HowNet* dictionary by a simple command line argument.

### 3.2.1.3 Level definition

Initially, the level was going to be defined as  $(wordRank)/216 + 1^3$ , where the words were ranked in descending order of their scores. However, this strategy produced 24 levels where words in levels 3 and higher seemed to be too difficult for an intermediate Chinese language student, so we decided to limit the game vocabulary to the 170 highest scored words divided into 4 levels based on their word scores. The minimum word scores for levels 1 to 4 were -10.5 (43 words), -11 (26 words), -11.5 (44 words) and -12 (57 words) respectively.

### 3.2.1.4 User Evaluation

The word scores were evaluated by two Chinese language students and one native speaker. Most of the words seemed to have received appropriate scores. However, several rare words received disproportionately large scores because one or both of their pinyin syllables were frequently used in many other words in the corpus, while a few common words with not so popular pinyin syllables had slightly lower scores than expected.

## **3.2.2 Choosing Game Vocabulary – Attempt 2:**

### 3.2.2.1 More data sources

Since the vocabulary generated from the travel corpus contained too many difficult words, and too few words that were easy enough for intermediate students, and was pretty biased towards words related to travel, two more data sources were eventually added into the game.

The first additional data source was a document of word statistics from the New York Times “news corpus” consisting of 45,429 distinct Chinese character words (out of which 19,938

---

<sup>3</sup> 216 was the number of words in four pages of our vocabulary list.

words had two syllables), all given with their counts in the original corpus. Even though no explicit character-pinyin mapping for the news corpus was provided, this was easily obtained from the *HowNet* dictionary. Moreover, the word statistics format was better suited for Chinese Scrabble since it eliminated the need to scan the real news corpus each time the game was loaded.

The second additional data source consisted of 6 lists of characters taught in grades 1 through 6 in Chinese elementary schools (see Appendix A). This was useful for biasing the easier games toward words containing the most basic characters.

### 3.2.2.2 Reformatting the Data Sources

Since the news corpus was more efficient in terms of both the space requirements and the time to load into the working memory of the game, we transformed the travel corpus into the same format - a list of words (in Chinese characters) with their word counts. Moreover, we discarded all words with more or less than two syllables as well as words that could not be translated by the *HowNet* dictionary. Using *HowNet*, each corpus was substituted by 3 files: a mapping of Chinese words to their counts (*travelChineseToCount.txt*, *newsChineseToCount.txt*), a one-to-one mapping<sup>4</sup> of Chinese words to their pinyin equivalents (*travelChineseToPinyin.txt*, *newsChineseToPinyin.txt*) and a mapping of Chinese words to their English definitions (*travelChineseToEnglish.txt*, *newsChineseToEnglish.txt*).

### 3.2.2.3 New Word Score: biased toward common Chinese characters instead of toneless pinyin

The reformatted data files, together with the lists of most common characters, were used to efficiently compute word scores according to the following formula, which took into account relative occurrences of the data in each data corpus<sup>5</sup>:

---

<sup>4</sup> Even though a pinyin word (with tones) can have multiple meanings each expressed by a different sequence of characters, a word in Chinese characters is usually mapped into only one pinyin representation.

<sup>5</sup> The scoring formula did not favor the relatively frequent words in the larger (news) corpus over the relatively frequent words in the smaller (travel) corpus. This was supposed to give the smaller corpus bigger influence on the vocabulary distribution. In the case of the travel corpus, we decided to give it more influence because travel related words are often the most practical for people studying a foreign language. Alternatively, we could have used the

$$score(w) = \left( \frac{travelCount(w)}{totalTravelCount} + \frac{newsCount(w)}{totalNewsCount} \right) \cdot (1.0 + 0.1 \cdot (7 - grade(w)))$$

Here, *grade* is the grade of the hardest character in the word, which can have values 1 through 7 where 7 is the grade assigned to all characters beyond grades 1 through 6. Also, a word was scored only if it was included in the more “dominant” news corpus (so that the vocabulary excluded too travel-specific words). Lastly, this model of word scoring can be extended with more than 2 corpora if needed.

#### 3.2.2.4 Level definition

Since the news corpus provided more frequently used words and the new scoring formula expressed the word difficulty better, it became feasible to have a wide variety of game levels again. The levels were redefined in terms of the order of a word in the list of all words in descending order of their scores as  $level(w) = \lceil scoreOrder(w)/100 \rceil$ . This created 192 levels of approximately  $100^6$  words in each level except for the last (192<sup>th</sup>) level, which combined all the words with a score equal to the lowest score.

#### 3.2.2.5 User Evaluation

All our evaluators reported better satisfaction with the expanded vocabulary and the new ranking technique. The vocabulary contained more everyday words and prioritized common words containing the 2,207 basic characters taught in grades 1 through 6, which was a feature that most students of Chinese would look for.

---

following formula, which would have merged all the corpora and therefore would have discounted the score of words that were relatively frequent in the smaller corpus:

$$score(w) = \left( \frac{travelCount(w) + newsCount(w)}{totalTravelCount + totalNewsCount} \right) \cdot (1 + 0.1 \cdot (7 - grade(w)))$$

<sup>6</sup> (“approximately” because we used the level score cutoff to determine the level of a word and it is possible that there were multiple words of the same level around a level boundary)

### **3.3 Randomized Generation of a Game at a Chosen Level; Solution Regeneration**

In general, a game of  $N$  tokens was generated by finding a set of  $(N-1)$  words such that each word in the set contained two distinct toneless pinyin syllables and shared at least one of them with at least one other word in the set. For example, the word set  $W = \{fang2jian1, shi2jian1, kailshi3, san1shi2, jian3cha2, da3kai1\}$  would correspond to token set  $T = \{fang, jian, shi, kai, san, cha, da\}$  at level 1 because the words in  $W$  are all of level 1 and cover all the tokens in  $T$ .  $W$  was then called the “current level solution” while the “overall solution” to the game with token set  $T$  was  $W' = W + \{da4fang1, da4shi4, fang1shi4, fang4da4, jian1shi2, jian4shi5, kailfang4, san4kai1, shi2cha1, shi2jian4, shi2san1, shi4fang4, shi4jian4\}$ , a set of all two-syllable words containing at least one toneless pinyin in  $T$ .

#### **3.3.1 Game Generation Techniques**

##### 3.3.1.1 Auxiliary Data Structures

All game generation techniques made use of a map called *syllToScoreWords* that had toneless pinyin syllables as keys and the lists of words (together with their scores) containing these pinyin syllables as values.

##### 3.3.1.2 Pinyin-driven Game Generations<sup>7</sup>

###### 1. Tree Game Generation

The first version of the game generation attempted to visually simulate the words as they might have been laid down on a board with  $O(N) \cdot O(N)$  dimensions. In particular, the solution

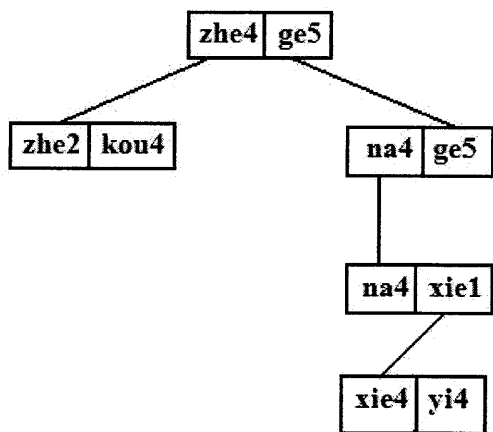
---

<sup>7</sup> The words in pinyin-driven game generation were chosen at or below level  $L$  as opposed to just at level  $L$  because the tree generation was very limited in terms of possibilities to choose a new word or token for the game (since it tried to expand each token only once as opposed to trying to expand any of the tokens already in the game as the word-driven game generations did). Also, pinyin-driven game generation was used only in the early stages of my project when the vocabulary of the game was limited to the travel corpus vocabulary.

that generated the game formed a tree of neighboring solution words (this tree branched only at the root – the subtrees of the children of the root were just simple lists – see Figure 2).

The root of this tree was the first chosen word, which was a word randomly chosen from the subset of words with the highest scores where the size of the subset was twice the level of the game. The word was put into the set of solution words and both its toneless pinyin syllables were put into a list of “game tokens” for this game. Subsequently, the game tokens were selected from their list in order, each one at most once. For each selected game token, a connecting word was randomly chosen from a set of words, at or below a specified level  $L$ , that contained the token but did not contain any other tokens already in game tokens (so that each new word led to a new game token). The subroutine of selecting a token and randomly selecting a word with the properties above was repeated until  $N$  game tokens were emitted.

The downside of this algorithm was that the generation failed frequently because new words (and new tokens) were being generated only from 2 words at a time, which greatly limited the number of possible games (see Figure 2).



**Tokens:** zhe, ge, na, kou, xie, yi

Words need to be generated top to bottom, satisfying the precedent constraints given by the connection lines.

If we needed more than 6 tokens and there were no more eligible words containing "yi" or "kou", this generation would get stuck, even though a possible continuation can lead through "zhe", "ge", "na" or "xie".

**Possible order of word generation:**

zhe4ge5, zhe2kou4, na4ge5, na4xie1, xie4yi4  
 zhe4ge5, na4ge5, zhe2kou4, na4xie1, xie4yi4  
 zhe4ge5, na4ge5, na4xie1, zhe2kou4, xie4yi4  
 zhe4ge5, na4ge5, na4xie1, xie4yi4, zhe2kou4

**Figure 2: Tree Game Generation**



## 2. Biased Set Generation

Biased Set Generation still approached the game generation from the “first choose pinyin, then a word containing that pinyin” perspective, but it didn’t require a new solution word to come only from one of 2 pinyin syllables (which were leaves of a tree in the tree generation algorithm). Instead, the new word could connect to any of the already existing toneless syllable tokens with probability proportional to the corpus occurrence of the toneless syllable. For example, if we had 3 tokens already chosen for the game and their counts in the corpus were 20, 10 and 30 respectively, the probability of choosing token 1 to use it for choosing the next connecting word would be  $20/(20+10+30) = 1/3$ , while the probability of choosing tokens 2 and 3 would be  $1/6$  and  $1/2$  respectively.

The reasoning behind this strategy was that, if a syllable occurs frequently in the corpus, then it is either part of a variety of words or is contained in one or more very frequently occurring (and therefore easy) words. This property of frequent syllables, combined with the lack of easy enough vocabulary for word levels resulted in frequently stuck generation when the odds kept choosing the syllable that had a high count but either it was contained in a very few easy enough words (that we already used as the solutions to the game) or all the words that contained the syllable were above the chosen level.

This problem might have been alleviated if we biased the probability of choosing a pinyin by the number of its occurrences in distinct words of the desired level. Alternatively, if we always prefer choosing pinyin syllables contained in very frequently occurring words regardless of desired difficulty, we can bias the probability of choosing a pinyin by the number of its occurrences in all (not necessarily distinct) words of the desired level. We could also bias the choosing algorithm for the new word after a pinyin has been chosen. However, this might have been unnecessarily complicated. Fortunately, we came up with a new, less erroneous and more elegant idea.

### 3.3.1.3 Word-driven Game Generation

Since pinyin-driven game generation resulted in unnecessarily slow and sometimes inconsistent results (too easy or too hard vocabulary), a more direct selection of the solution

words was applied. In the word-driven game generation, once a pinyin becomes one of the tokens, all current-level words containing that pinyin are put into a pool of words that may become a game solution in the remaining steps of the game generation process.

The generation starts with selecting an arbitrary word in the target level. Next, all the words within the target level containing exactly one of the toneless pinyin syllables of the first word are put into the pool of words. After that, the algorithm enters a loop until it either finds a set of tokens with current-level solution or it gets stuck and returns null, which means that the game generation was unsuccessful and we need to restart it. Each loop iteration starts with randomly selecting a word to be taken out of the pool. If one of the syllables of the word is not included in the tokens yet, we add it to the tokens and add all current-level words containing the new syllable into the pool of words. Otherwise, we don't do anything (other than removing the word from the pool of words).

The only way the algorithm may get stuck is if the word pool becomes empty. In that case, there is no way to finish the generation or to restore it to a point where it still could work. Therefore, a sensible strategy to generate new games quickly and flawlessly was to put the generation algorithm into a loop which kept it from finishing with a null return value.

### **3.3.2 Solution Regeneration**

In contrast with the game generation, which used only words from a given level, the solution regeneration procedure found all words that may have been used at some point in the game. This was done by scanning through the lists of words for each token and checking each word if its other syllable is also one of the tokens. If it is, the word is added to the solutions. Even though these “extra solutions” were not intended to be the focus of the game, they make the game less strict and more interesting, and encourage the user to explore vocabulary beyond their current knowledge level.

### 3.4 Basic User Interface

The most important goals of the user interface for Chinese Scrabble were to make the game enjoyable and educational. These goals seemed to be strongly related to the interface esthetics, learnability, memorability and ease of use, which would be achieved by making the interface simple and intuitive, but also by having readily available help and user guidance at any point during the game (see Figure 3). The critical features through which the interface goals were met are described below.



Figure 3: The final Chinese Scrabble user interface

#### 3.4.1 Game Board and Tokens Panel with Reading Help

Similarly to regular Scrabble, the game board consists of rectangular cells with each cell having 2 to 4 neighbor cells. The remaining (unused) syllable tokens are displayed under the board. Since the game uses only 2-syllable words, which it puts down in an aligning rather than intersecting manner, it is important (especially for a non-native speaker) to see the word boundaries. This was done by painting each word on the board with a shade of blue with the restriction that the color cannot be the same as that of any of its neighboring words. Moreover, if

a cell could hold a syllable of a hypothetical new word<sup>8</sup>, it was colored yellow upon mouseover and green upon selection.

Since the users of Chinese Scrabble might not be proficient in reading Chinese characters even for the words they can speak, reading help seemed like an effective tool to reduce frustration and increase Chinese character learning for the player. The first version of reading help was a tooltip that appeared after the user paused the mouse over a board cell. However, the tooltip took too long to activate itself and then disappeared too quickly. Moreover, it sometimes obstructed other board cells that the user wished to see. To eliminate these flaws, we decided to display the reading help right above the upper edge of the board where it would not collide with any other important user interface feature, it would have more space for displaying reading and English meaning of the word in addition to the character pronunciation, it could be displayed for as long as the user mouses over a character and the user would still be able to pay attention to both the board and the help at the same time.

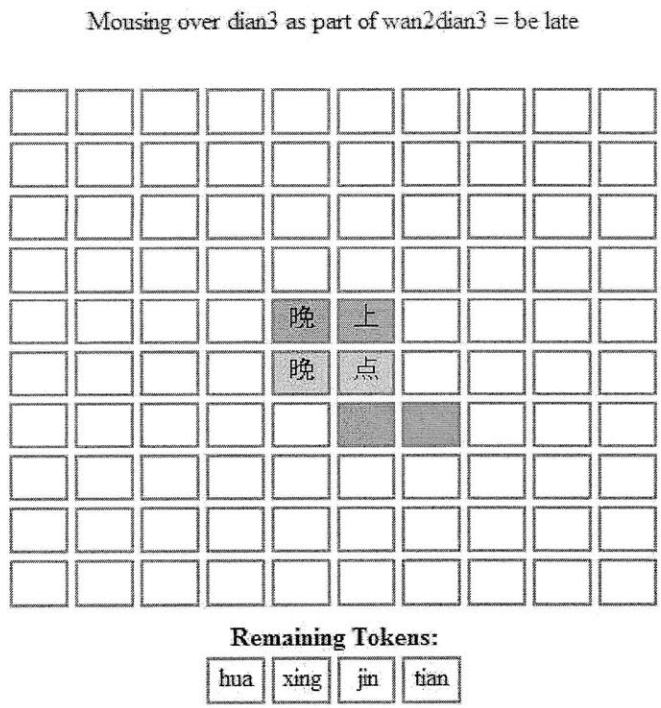


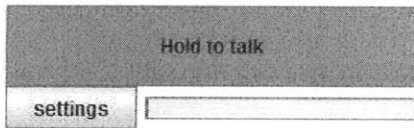
Figure 4a: Game Board, Token Panel and Reading Help

<sup>8</sup> This was any empty cell that was next to or one cell away (vertically or horizontally) from a non-empty cell. It did not guarantee that there was word that could fill in that cell).

### 3.4.2 Speech and Text Input Panel

Everything regarding the input of words and the feedback for this input was concentrated on the left side of the game board. The purpose of this layout was to limit the field of view necessary for paying attention to the input and the game board at the same time since these interface pieces were expected to be used together frequently. On the other hand, the game help would rarely or never be used simultaneously with the user input, so these two features were located at the opposite sides of the page.

**Tones and speech are optional now** - you can change this in *SETTINGS*



To indicate a **tone with speech input**, click a board cell multiple times before speaking.

or

Type a word in pinyin and press enter:

example: mei yuan, mei3yuan4, mei3 yuan4

**YOU SAID:**

晚点(wan2dian3) = be late (LEVEL: 4)

**Tone bonus:** 100%

**Speech bonus:** 10%

**Mistake penalty:** 0%

**Word Score:** 8.80

**Figure 4b: Speech and Text Input Panel**

### 3.4.3 Game Settings, Game Instructions and Demo

To accommodate a variety of users, the game had several adjustable parameters. These included game level, number of tokens, board size (width x height), required versus optional speech, required versus optional word tones and the type of level advancement (*medium*, *fast*, *customized* or *off*). To help users understand the rules of the game better, we provided both comprehensive and structured game instructions as well as a basic demo. The settings, the

instructions and the demo were located in the upper right corner of the page because we expected most users to look for these tools there.

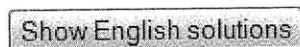
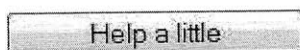
### 3.4.4 Game Help Panel

The help panel, located to the right of the board, contains a toggle help button providing three levels of direct help for placing a new word on the board as well as a toggle button that reveals or hides English equivalents of the words that can currently be placed on the board. Toggle buttons were used to minimize the clutter in the user interface. When the help button is used, all help is shown on the board or on the tokens, so that the user can maintain his focus on the board panel. If the user clicks the English definition button to reveal or to hide the English translations, the translation panel remains open or closed until the user presses the button to hide or reveal the panel, which may be done at any point in the game. If the panel is open, the English translations will be automatically updated if the user places a word on the board. This design was intended to increase the efficiency of the interface by eliminating unnecessary clicks that might have been needed if the translation panel closed or had to be refreshed after each word submission.

#### **IF YOU GET STUCK...**

use these buttons :-)

(See *HOW TO PLAY* for more details)



**Figure 4c: Help Panel**

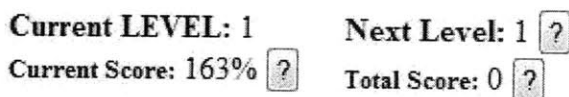
### 3.4.5 Hints and warnings




Even if the user doesn't look at the instructions or demo before he plays the game for the first time, the hints and warnings should be self-sufficient in guiding him through the game. This

is a crucial feature since most users will not do much beyond skimming the instructions. Almost every user action in the game receives some feedback in the form of hints displayed in green below the help buttons or the warnings in red displayed under the hints. (Originally, the warnings were implemented as pop-up boxes, but these were abandoned since most users find pop-up boxes annoying or alarming and it was also quite hard to implement pop-up boxes for playback.) The color scheme for the hints and warnings not only catches attention but also reveals the nature of the messages even without reading (green corresponds to “this would be a good idea” and red implies “do not do this because something may go wrong” or “you just performed an unexpected or illegal action”). The location of hints and warnings was also meant to enhance learnability and ease of use since these messages have many things in common with regular game help, which was located just above the messages.

### 3.4.6 Level and Scoring

The level of the game together with the user’s current and total score seemed related to the game help (3.4.4) and user guidance messages (3.4.5). Moreover, in many other popular computer games the level and score are displayed near one of the corners of the screen. Therefore, they were placed between the buttons in the upper right corner and the help panel. The level and scoring naturally called for four pieces of information (current level, current score, total score and a (not so common) predicted level), which allowed these to nicely align with the buttons in the upper right corner.



**Current LEVEL: 1**      **Next Level: 1**   
**Current Score: 163%**       **Total Score: 0** 

**Figure 4d: Levels and Scoring**

### ***3.5 User Interaction***

As in the other games using spoken language for communicating some or all of user's needs, the goal of Chinese Scrabble was to use the spoken language together with the rest of the interface to create a more intuitive, smooth and highly responsive user experience. The strategy was to conjure up a variety of user reactions at each interaction point with clear reasoning behind each reaction and then try to accommodate this reasoning by both the static and interactive user interface.

#### **3.5.1 Selecting a Word Location**

To select a location of the word, a user needs to click on two neighboring cells out of which at least one must be neighboring with a non-empty cell, unless the board is empty. To prevent the user from clicking on the cells that cannot be selected, the cells that could be selected were highlighted in yellow when the user moused over them. If the user clicked on a yellow cell, it changed to green, which stood for "selected". To finish locating a word, the next selected cell must have neighbored to the first selected cell and at the same time one of the two cells neighbored on a non-empty cell.

In the first version of the word location selection, the first selected cell must have neighbored to a non-empty cell. To complete the word, the second selected cell must have neighbored on the first cell. Unselecting of any of the selected cells was also realized by a click (on a selected cell). These rules were logical and seemingly simple but also fairly inefficient and possibly frustrating for a new user to whom it may seem more intuitive to select the first pinyin of his word before the second pinyin, which may not always agree with the rules above. Also, having to unselect one or both of the cells if the user wanted to select a completely different location was quite tedious.

Redesigning the cell selection procedure promised a huge efficiency gain, at least in theory. The new rules allowed the user to select any cell that may have possibly been part of a new word, that is, any cell with one or zero cells separating it from a non-empty cell. Moreover, after the first cell has been selected, the second cell could have been selected from the same set of cells as the first cell (excluding the first selected cell, of course). If the second cell did not



neighbor on the first cell, the game assumed that the user changed his mind about the word location and the first cell was unselected. Even if both cells were already selected, the user could have unselected them by clicking on any yellow cell that was not a neighbor of either of the selected cells or the user could have unselected one of them by clicking on a cell that neighbored one of the selected cells. The user could have also unselected the cell as in the first version of word location selection, though this was no longer needed.

Since the speech recognition did not recognize tones, the tones for the spoken words were indicated on the board by multiple clicks on a cell corresponding to a pinyin with tones.

### 3.5.2 Placing a Word On the Board

Once the location for a new word was selected, the user could submit the word in speech or text. Both input modes could be used with or without tones. The user had the option to let the game require him to use speech or tones or both. If speech was required, the user had to attempt speaking at least once for every word submitted. If his first utterance was unsuccessful, he could try writing the word in pinyin (which may be an obvious thing to do if the utterance was misrecognized) or he could use the help button. If the tones were required, every word had to be submitted with tones, so the first cell click automatically selected tone 1. If a word was correct except for the tones, the user was presented with a warning message telling him how to correct the tones. To receive credit for the word, the user had to correct the tones and resubmit the word.

The text input worked with three types of pinyin representation – space separated pinyin syllables without tones (for example, “mei yuan”), space separated pinyin syllables with tones (for example, “mei3 yuan4”) and concatenated pinyin syllables with tones (for example, “mei3yuan4”). Concatenated pinyin syllables without tones were not accepted since this option would have had a high implementation cost and would result in a lot of unnecessary computation and possible ambiguity in trying to find the pinyin boundary.

The speech input allowed the user to simply click on the “Hold to Talk” button and say a Chinese word while holding the button. The speech recognizer tried to match the user’s utterance to the toneless pairs of pinyin syllables that could be created from the tokens. As mentioned in the previous section, the recognizer did not distinguish between tones, so any tones must have been indicated on the board.

### 3.5.3 Using the Help Button

To get help, the user just had to click the help button and the hints appeared on the board panel. After clicking on the help button for the first time for the current word submission, the title of the button switched from “Help a Little” to “Help More” and the first level of help was activated, causing a character to which a new word could be connected (“connecting character”) to be highlighted on the board. The connecting character remained highlighted until the user clicked on the help button for the second time or moused over the board. In the latter case, the user could continue trying to submit a word. The next time he clicked on the help button for the same word submission, the second level of help was activated. After initiating the second level of help, the title of the help button changed to “Show Next Step” and all connecting characters for each remaining pinyin token were being highlighted in turn at two second intervals. The pinyin to which the highlighting belonged was displayed under the help button and also its corresponding token was highlighted below the board. Similarly to the first level of help, the highlighting disappeared when the user either clicked on the help button again or moused over the board. The third (and last) level of help placed one of the possible words on the board. The title of the help button was set back to “Help a little”.

### 3.5.4 User Interaction with the Rest of the Interface

The rest of the interface was very easy to operate. To reveal or hide the translations of words that could be put on the board at the moment, the user simply clicked on a button. Game settings also required very little typing (just a few short numbers) and a few radio button clicks. This simplicity in the user interface interaction was expected to contribute to the learnability, memorability and efficiency of the entire Chinese Scrabble game system.

## ***3.6 Game Logic and Intelligence***

Acting behind the interactive user interface is the game logic, whose goals were to make the game both interesting and stimulating to learning, to provide appropriate user feedback and to guide the user through the game.

### **3.6.1 Speech Recognition**

The goal of the speech recognition was to recognize as many valid user utterances as possible. False positives (recognizing a valid word, not necessarily fitting into the selected location and different from what the user intended to say) were considered less detrimental than false negatives (misrecognizing a valid word) because false positives had the potential to expand the user's vocabulary while false negatives could cause unnecessary frustration. The recognition was implemented using WAMI Speech Recognition Javascript API [2, 15], which was supplied with a weighted grammar and algorithm for selecting the recognition result.

The first (naïve) recognition technique attempt consisted of an unweighted grammar of all solution words (and nothing else). The recognition result was chosen to be simply the best recognition hypothesis. Because of significant recognition inaccuracies, this technique caused a lot of false positives (as tested by two Chinese language students) and as a result often led to revealing new vocabulary to the user by misrecognizing his utterance, which may have been interesting at first, but if occurring frequently, it might have lead to frustration. Moreover, despite the “no recognition result” possibility, most of the time the recognizer recognized something and since all recognitions were valid Chinese words, the only possible mistake was an incorrect location for the word. This was not a very realistic representation of the mistakes made by the user, since, even if the user had problems with Chinese character reading, he could have moused over the characters to get reading help and thus minimize the errors due to word positioning.

The first improvement to the naïve recognition technique was to expand the grammar by all possible pairs of distinct tokens, making most of the grammar words fall into the “confuser” category. This change alone was expected to create a lot of false negatives, which was more

realistic, but also possibly much more frustrating for the user. Therefore, the second step towards recognition technique improvement was to expand the set of possible recognition results to more than the first recognition hypothesis. We chose to consider the top four<sup>9</sup> recognition hypotheses (the “set of top hypotheses”). To select the recognition result out of the set of top hypotheses, we defined the recognition result as the word with the lowest recognition rank in the set of top hypotheses that fit into the selected location or, if no word fulfilled these conditions, the recognition result was the first recognition hypothesis. (For example, if both the second and the third hypotheses fit into the selected location, the second hypothesis became the recognition result, but if none of the top four hypotheses fit into the location, the first hypothesis became the recognition result). Although the grammar remained unweighted, this new technique worked much better in practice since it reported errors due to word positioning very rarely, unless the user actually selected an invalid location.

### 3.6.2 Hints and warnings

The hints and warnings have been part of the game since the start of user interface implementation, because it was considered very important to always keep the user aware of what is happening in the game and what he can do at any moment. Some hints provided real-time game instructions like asking the user to choose a location for the next word, to speak or type a word or to start a new game when ready. Other hints appraised the user and encouraged him to keep playing, for example, by congratulating him on submitting a word or on finishing a game. The warnings appeared when the user tried to do something forbidden or impossible or if there was an error in his/her input. For example, if the user clicked a cell that could not be selected, he encountered a warning (“This cell cannot be selected now.”) together with a suggestion to remedy the situation (“A new word has to be placed next to a word already on the board and the two words must touch through the toneless pinyin syllable that they share.”). Another example is the situation when the user tries to submit a word without having selected exactly 2 board cells for its location. If the user didn’t select anything, he is reminded that the location must be selected before the word is spoken. If he selected only one cell, it is clarified that Chinese

---

<sup>9</sup> To reduce the number of false negatives it might have helped to use top 7 or top 20% of possible words as the top hypotheses set.

Scrabble uses only two-syllable words. Similarly to the forbidden action warnings, the input error warnings tell the user why the word cannot be submitted (for example, because the word does not exist or it does not use one of the remaining pinyin syllables or it does not fit into the selected location or the tone information of the word is incorrect). Warnings are also used if very important information needs to be brought up to the user's attention. For example, if the game cannot be continued, the user should be made aware of it immediately.

### 3.6.3 Vocabulary Level Focus

The goal of having a game at a certain level is to reinforce the vocabulary at that level, while reviewing the vocabulary from lower levels and ideally giving the user an opportunity to demonstrate his knowledge of the vocabulary above the current level. We approached this goal by generating the games using the techniques described in section 3.3.1. After generating a game, we got the tokens and a solution to the game at the chosen level. Most of the games had more than one solution, but not all the solutions consisted only of words at the chosen level. To keep the focus of the game on the vocabulary of the chosen level, both the user input feedback and the game help needed to be biased toward the chosen level.

#### 3.6.3.1 Level Focus in User Input Feedback

Since the user is allowed to submit words as pinyin syllables without tones, it is highly probable that a "toneless word" has several possible tone assignments and sometimes even one tone assignment can have multiple Chinese character representations. The question was how to resolve this ambiguity of user input.

The first option was to let the user view the set of all possibilities and let him select any word from this set. This way, the user could always choose the word that he intended to say, unless the word is outside of Chinese Scrabble vocabulary. However, the user may also choose a different word, possibly one that he has never seen before. At first sight, exposing the user to a variety of words may seem to increase the user's learning, but if there are more than 2 possibilities, the user may easily get overwhelmed by the breadth of choice. In addition,

choosing from multiple possibilities for almost every submitted word may slow the game down to the point where it would get boring or frustrating.

The second option was to choose a word for the user, so that with high probability the chosen word corresponds to the user's intention and, even if not, it will teach the user something. Taking into consideration the idea of word levels and scoring, we decided to present the user with only the words at the current level or at a level closest to the current level from below or, if none of the possible words satisfied these requirements, the words closest to the current level from above were chosen and displayed.

### 3.6.3.2 Level Focus in Game Help

If the user gets stuck, the game should ideally help him to continue with words from the current level. If that is not possible, then, similarly to the user input feedback, the help focuses on the closest level from below or (if there are no possible words at lower levels) the help focuses on the closest level from above.

Since there are multiple possibilities for the first and the third level of help (highlighting one connecting character and placing a possible word on the board), we needed a way to choose one of these possibilities each time the help is called. To bias the game toward the chosen level, but provide some randomness for variety, the connecting character in the first level of help was randomly chosen from a subset of connecting characters to which at least one new word of closest possible level to the chosen level could be connected. Similarly, the word that was chosen in the third level of help was a word from the closest possible level to the chosen level. Since the second level of help showed all connection characters for each remaining pinyin token, no level bias was employed here.

## **3.7 Scoring**

To increase the user's motivation to play the game, awareness of his performance during the game and satisfaction from having done well in the game, we introduced a simple scoring system that evaluated the user's performance in the current game as well as his cumulative

performance during the current game session. Moreover, this score could be used to estimate the level of the user's next game, based on his cumulative performance.

### 3.7.1 Word Scoring and Total Score

The score of each word depended on the word level, the use of speech and tones and the number of errors and help usages for this word submission:

$$score(w) = (1 + toneBonus) \cdot (1 + speechBonus) \cdot \left\lfloor \frac{4}{2^{errorHelpNum}} \right\rfloor \cdot \frac{1}{4} \cdot level(w)$$

Where:

$toneBonus = 1$  if the tones were entered correctly on the first try

$toneBonus = 0$  if no tones were submitted or the tones were submitted incorrectly at first

$speechBonus = 0.1 \cdot utteranceCount$  where an utterance must have been recognized as something (not necessarily a valid word, but utterances with no recognition did not count)

$errorHelpNum$  was the number of times error occurred or help was used in this word

submission, so  $\left\lfloor \frac{4}{2^{errorHelpNum}} \right\rfloor \cdot \frac{1}{4} = 1$  if  $errorHelpNum = 0$ ,

$\left\lfloor \frac{4}{2^{errorHelpNum}} \right\rfloor \cdot \frac{1}{4} = 0.5$  if  $errorHelpNum = 1$ , and

$\left\lfloor \frac{4}{2^{errorHelpNum}} \right\rfloor \cdot \frac{1}{4} = 0.25$  if  $errorHelpNum = 2$ .

The scores of different words were independent of each other. *Total Score* was the sum of individual word scores played in finished games during this session. *Current Score* was the sum of the word scores in this game as a percentage of the basic game score at the current level, where basic game score was the sum of the basic scores (word levels) of words at the current level that could be used to finish the game. For example, if there were 7 tokens and the game was at level 5, the basic score would be  $5 \cdot 6 = 30$ .

### 3.7.3 Level Advancement Options and Predicted Level

The game had 4 level advancement options: fast, medium, custom and none.

The fast advancement option advanced a user by arbitrarily many levels based on his raw score (sum of individual word scores) in the current game as well as the raw scores in his previous games. The influence of raw scores of the games decreased linearly with respect to the order of the games in the current session up to the 9<sup>th</sup> game before the current game; that is, the current game was weighted with coefficient 1, the game before with coefficient 0.9, the game before with 0.8 and so on, up to the 9<sup>th</sup> game before the current game, which was weighted with coefficient 0.1. This weighted average of the game scores determined the weighted average of a word, the floor of which was the predicted level.

The medium advancement option used the predicted level from fast advancement and then returned a logarithm of the increase or decrease in the fast advancement level.

The custom advancement allowed the user to choose how many words from the current level he would like to use without help or just to see in the game, as well as how many levels he would like to advance after he reaches his goal. The advancement may also be determined only after the goal is reached.

Lastly, the user can turn off the advancement and play the game at the same level or change it manually in the Game Settings.



## 4. User Study

### 4.1 User Study Design

In order to verify the value of our system, we decided to do a small user study with both Chinese language students and native Chinese speakers. Three users in the study have used our system during earlier game development. Four users were playing the game for the first time during the study. Three users have studied Chinese as a second language (referred to as “students of Chinese” below). Two of them did so in school and one was a self-learner. All students of Chinese were exposed to the language for more than two years. The remaining study participants all spoke at least some Chinese at home (they will be referred to as “Chinese speakers” below). Two Chinese speakers spoke Chinese at home (one occasionally and one very frequently), but never attended a school in China. In addition, the participant who spoke more frequently has attended several semesters of Chinese reading and writing at MIT. The last two users were native Chinese speakers who received part of their education in China – one attended elementary school and one finished a college in China.

The primary goals of the study were to verify that the interaction of the users with the game was smooth and enjoyable, and that the users subjectively felt that playing the game expanded their knowledge of Chinese vocabulary. The secondary goals were to find out whether the users preferred playing the game by speaking or by typing and whether they were interested in playing it with or without tones. Third, we wanted to check whether the users discovered and used various help features and how satisfied they were with them. Lastly, we wanted to see if the medium level advancement was interesting to the user and if it approached the user’s appropriate game level after a few completed games.

The study consisted of three parts, totaling 10 to 20 games.

The first part of the study consisted of “training games”, in which the user played three games at level 1 with three out of four different speech-tone settings. These were “optional speech and optional tones”, “required speech”, and “required tones”. (The “required speech and required tones” setting was not included in the training games because it was considered too hard

and if the user liked both “required speech” and “required tone” settings, he could have chosen “required speech and required tones” setting in parts two and three of the study.)

In the second part of the study, the user played 4 to 10 games. For each game, the user could choose its tone-speech setting, but the level for the game was determined by the medium level advancement algorithm.

In the third part of the study, the user played 3 to 7 games of his chosen tone-speech setting and level.

After finishing the games in the user study the user was encouraged to express his or her opinions about the game in the user study survey. The survey included questions about the user’s enjoyment level, his opinions on the usefulness, efficiency, learnability and memorability of the game as well as on the user interface, vocabulary difficulty, scoring, helps options, speech-tone settings and correctness of level advancement.

## ***4.2 Playback and User Data Logging***

For the sake of more detailed analysis of the user study data as well as analysis of the online user study, we added both data and action logging into the game.

Data logging recorded which games the user played, what were the levels and solutions of the games, which words with what difficulty the user submitted as well as if he used help, settings, instructions, speech, tones and how many and what kind of errors he made. We used WAMI data logging API, which consisted of a single call to function *logEvents*, to which we passed a JSON object with the data to be logged.

Action logging recorded all user-initiated actions during the game. This record of action could later be played back and analyzed. Again, we used WAMI logging API, in particular the function *performAction*, which took a function name and zero or one parameter. The parameter number restriction as well as the reliance on the user interface appearance (which most of the time changed with user action), called for a major code restructuring where all functions evoked by the user were passed as parameters to *performAction*, while their multiple parameters were converted into a single JSON object or they were given a redundant parameter that would otherwise have been missing in the playback. Also, all pop-up boxes were eliminated.

### **4.3 User Data analysis**

Based on the recorded data and the playback we wanted to answer the following questions:

1. Are the users happy with the game?
2. Do the users think that the game is a useful learning tool?
3. Do the users use speech and tones? How do they like them?
4. Do the users use the help features? How do they like them?
5. Do the users use mostly words of the current level? (That is, was the leveling effective?)
6. Do more advanced users do better (receive higher scores, achieve higher levels, make fewer mistakes or play faster) in the game?

#### **4.3.1 Are the Users Happy with the Game?**

The user happiness was measured primarily according to the post-game survey (see Appendix B), but also by watching the playback. According to the post-game survey, most users enjoyed the game very much (their average enjoyment ranking was 4.29 out of 5 (4.29/5)). The students of Chinese enjoyed the game slightly more (average score 4.67/5) than the Chinese speakers did (average score 4/5). When the users were asked if they would play the game again, where 1 meant that they would certainly not play it again and 5 meant that they would play it for sure, 5 out of 7 users responded with a score of 3 or higher. The average score was 3.57/5, where most of the high scores came from the students of Chinese (average score 4.67/5), which is very encouraging. Chinese speakers probably did not find the game engaging or beneficial enough to return to it in the future (average score 2.75/5).

#### **4. 3.2 Do the Users Think that the Game Is a Useful Learning Tool?**

The user satisfaction with the amount of learned vocabulary and the amount of reviewed vocabulary was very high (on average 4.14/5 and 4/5 respectively). The user confidence with individual aspects of the vocabulary (toneless pinyin associated with English meaning, pinyin with tones associated with English meaning, and Chinese characters associated with English meaning) was markedly lower, but still indicative of considerable learning gains resulting from playing the game (see tables 1a and 1b). Except for the Chinese characters associated with English meaning of reviewed words (average score 2/5 for Chinese speakers and 2.67/5 for the students of Chinese), Chinese speakers were more confident in all aspects of the learned and reviewed vocabulary (see tables 1a and 1b). This was not so surprising since people are usually more confident with their first language than someone who studies a foreign language. However, it was unnatural (and therefore difficult) for Chinese speakers to create words out of toneless pinyin because a pinyin can be represented by many different characters and therefore a 2-pinyin word can have a multitude of meanings. On the other hand, the students of Chinese considered using toneless pinyin easier since they were more used to thinking about the language in terms of pronunciation, in which the tones and pinyin could be separated. (The evidence of this are the Chinese speakers' score of 3.25 for an average learned toneless word versus the score of 4 for an average learned word with tones and the score of 3.67 for an average reviewed toneless word versus the score of 4 for an average reviewed word with tones; and the students' of Chinese score of 3 for an average reviewed word with tones versus the score of 3.33 for an average learned toneless word for Chinese learners.) Chinese characters received the lowest confidence scores (2.67, 3, 2.67 and 2) and therefore were probably the hardest to remember for all study participants.

**Table 1a: User Satisfaction with learned words and their aspects**

User Type	Overall satisfaction	Toneless Pinyin & English	Pinyin with Tones & English	Chinese Character & English
Average User	4.142857	3.142857	3.571429	2.857143
Student of Chinese	4.333333	3	3	2.666667
Chinese Speaker	4	3.25	4	3

**Table 1b: User Satisfaction with reviewed words and their aspects**

	Overall satisfaction	Toneless Pinyin & English	Pinyin with Tones & English	Chinese Character & English
Average User	4	3.5	3.5	2.333333
Student of Chinese	4	3.333333	3	2.666667
Chinese Speaker	4	3.666667	4	2

### 4.3.3 Do the Users Use Speech and Tones and How Do They Like It?

Typing the words in the game was slightly more popular than speaking the words (average score 4.29/5 for typing versus 4/5 for speaking), which corresponded to the Chinese speaker preferences (average score 4.75/5 for typing versus 3.5 for speech), but was opposite to the preferences of the students of Chinese (average score 4.67/5 for speaking versus 3.67/5 for

typing). This can be explained by the fact that typing is often faster because the words in the game are very short, and no misrecognition occurs as well as the fact that the students of Chinese have higher tolerance for recognition errors and may also find the speech input more interesting.

A bit surprisingly, the study participants used much more speech than typing (the approximate ratio of usage was 2:1). However, by looking more closely we discovered that this was because the students of Chinese had a much larger bias toward speech usage (96%) than the Chinese speakers had toward typing (56%).

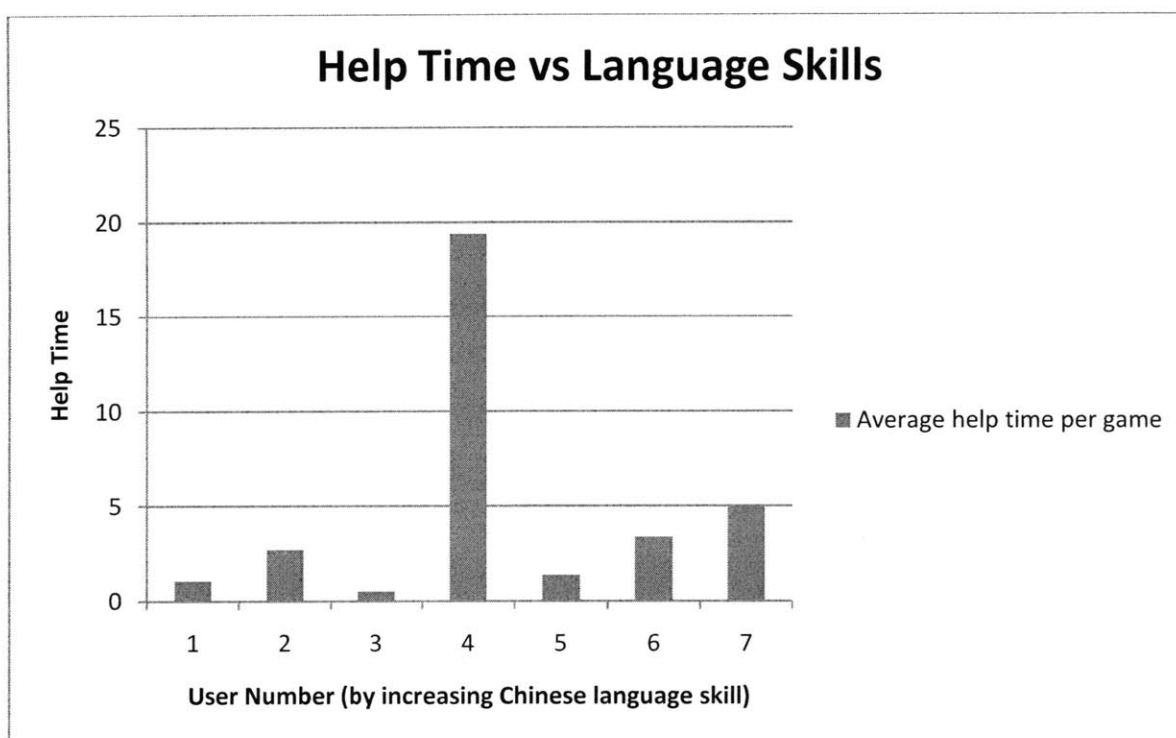
On average, the participants preferred playing the game without tones, where the students of Chinese had stronger preferences about this issue (average score 4.67 for a toneless game versus 3 for a game with tones) than Chinese speakers did (average score 4.75 for a toneless game versus 4 for a game with tones). The game without tones, in spite of providing less learning material, was probably more popular because it was faster, easier and less error prone.

However, the average usage of tones was almost equal to the usage of toneless words (49.2% versus 50.8% respectively) where the students of Chinese used more words with tones (56.2%), while Chinese speakers used more toneless words (56%).

#### **4.3.4 Do the users use the help features and how do they like them?**

All study participants noticed that there was a game help and at some point used it. Five participants tried all three levels of help while the remaining two tried only the first level of help. On average, the study participants pressed the help button 1.29 times per game, which seems relatively infrequent, but it was highly biased by the first three or more relatively easy games. The first level of help was used many more times (0.81) than the second level (0.3) and the third level (0.18). This may indicate that the first level of help was often enough to move the user forward. Chinese speakers used help more (on average 1.56 times per game totaling to 7.26 seconds) than the students of Chinese did (on average 0.96 times per game totaling to 1.43 seconds). Even though the help usage was highly biased towards one Chinese speaker study participant, from Figure 5 we can see that even after excluding this participant Chinese speakers spent more time examining help than students of Chinese did. Some possible reasons might have been that Chinese speakers were more careful not to make mistakes (see section 4.6), were less

interested in random guessing and found the first level of help very useful in helping them remember words that they momentarily forgot (average score 4.67/5).

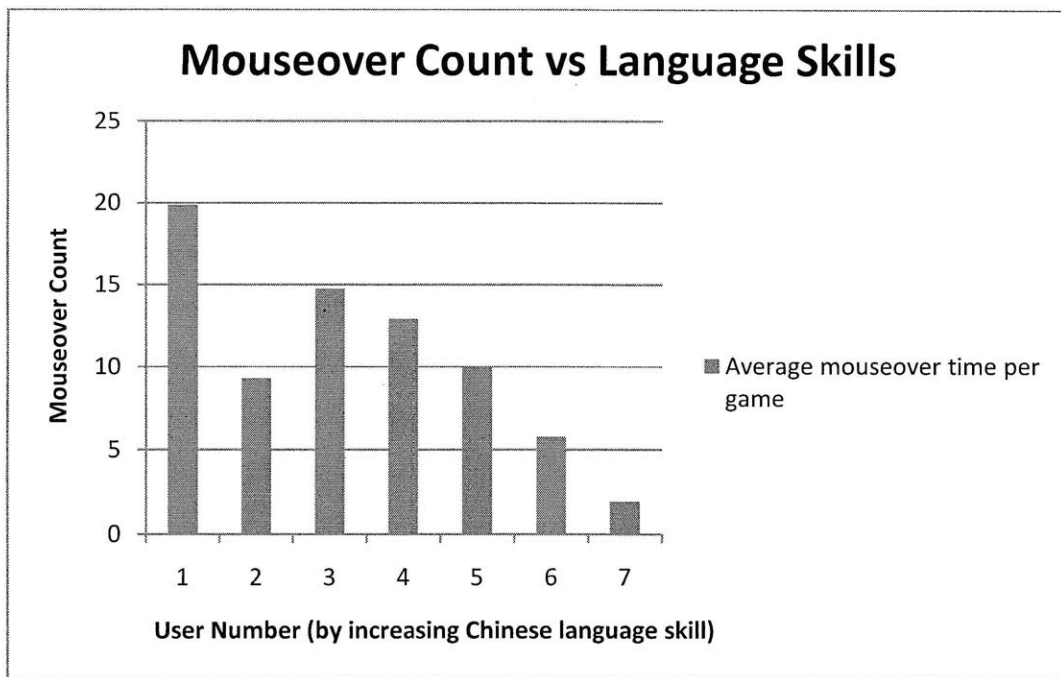


**Figure 5: The relationship between the average time per game spent on help and the user’s language skills**

Most people found the first level of help slightly more useful (average usefulness score of 4.29) than the second level of help (average usefulness help of 4.2), even though the second level of help provided more information. This might have been mostly because the Chinese speakers, who ranked the first level of help with an average score of 4.5 as opposed to the average score of 4 selected by the students of Chinese. Chinese speakers also found the first level mostly sufficient for reminding them of words that they momentarily forgot much more than the students of Chinese did (average scores 4.67 versus 3.33 respectively). The second level of help might have been less popular also because, as one participant noted, it was switching between highlighting the connection points for different pinyin syllables too quickly. Lastly, the help was

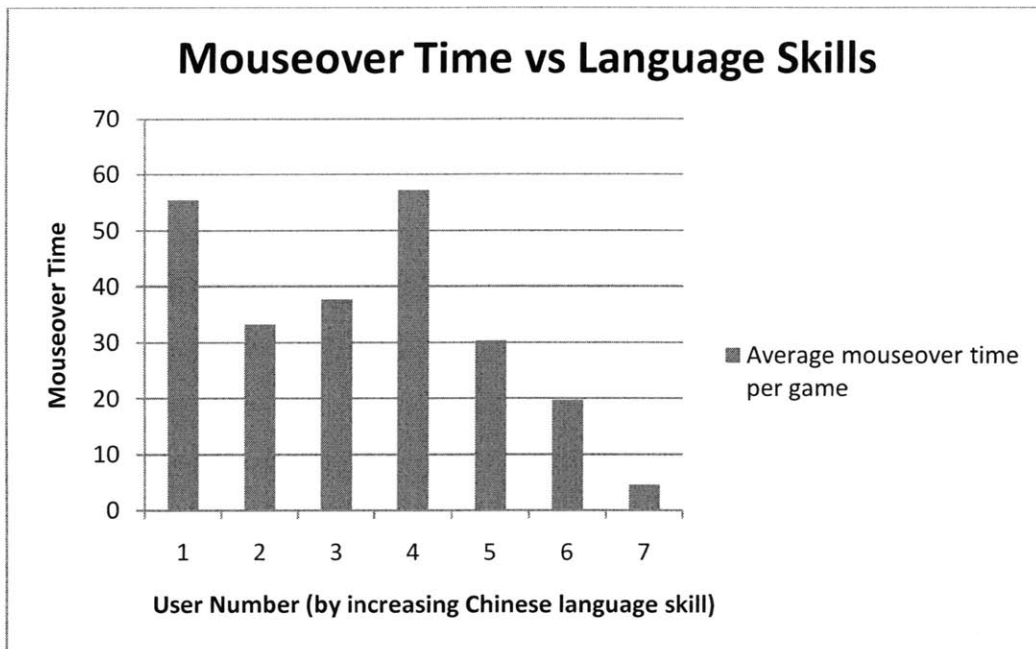
usually used in the middle or towards the end of the game, at which time the level 2 help provided very little additional information on top of level 1.

Displaying pronunciation and translation of a word on the board upon mousing over one of its pinyin syllables could be considered not only an additional educational feature but also a non-penalized form of help. In the data logging, recording the user using a mouseover to check on the reading or meaning of a word was a little bit tricky since the user might have moved the mouse across a character without the intention of getting the mouseover help. Therefore, we did not record every single mouseover, but only those that lasted more than 1 second. With this restriction (which might have still had a lot of false positives), the average number of mouseovers was 10.65 per game with 3.17 seconds spent on a mouseover. Not surprisingly, Chinese speakers used less mouseovers (on average 7.66 per game) than the students of Chinese did (14.64 per game). In fact, the number of mouseovers decreased almost monotonously with increasing skill (see Figure 6a). Although Chinese speakers spent a bit more time on each mouseover (on average 3.31 seconds as opposed to 2.98 seconds of students of Chinese), this slight difference could have been mainly attributed to one user's unusually high use of mouseovers (the middle column in Figure 6b).



**Figure 6a: The relationship between the average number of mouseovers per game and the user's language skills.**





**Figure 6b: The relationship between the average time per mouseover and the user’s language skills.**

Two out of three students of Chinese moused over the characters to learn about word reading while only one student was interested in English meaning of words that (s)he moused over. On the other hand, all Chinese speakers used the mousing over to learn about English meaning of the words, while only one speaker was also interested in the reading aspect of the words. One reason why most Chinese speakers were comfortable with character readings in the game may have been that most of the characters were taken from the character set taught in grades 1 through 6 in Chinese elementary schools.

Another non-penalized help option was displaying English solutions of the words that could be put on the board at the moment. Most users found this feature helpful, though it was more popular with Chinese speakers since it helped them with the difficulties of creating words from toneless pinyin syllables.

### 4.3.5 Do the users use mostly words of the current level? Was the Automatic Level Choice Effective?

Out of 69 games played by 7 study participants, the most frequently used word level in a game was not equal to the game level in only 9 games (13%), 6 of which were played by one particular Chinese speaking participant. Moreover, in 43 games (62.3%) the game level was also equal to the median word level. For each user, the average word level of a game was in most cases greater than the level of the game (see Figure 7b-h). Moreover, for each user, the average word level across all games was always greater than the median level across all games, which was greater than the average most frequent level across all games (Figure 7a). This indicates that the users used mostly the words of the current game level and a few words of higher level. The words of higher level might also have been the result of using game help or guessing.

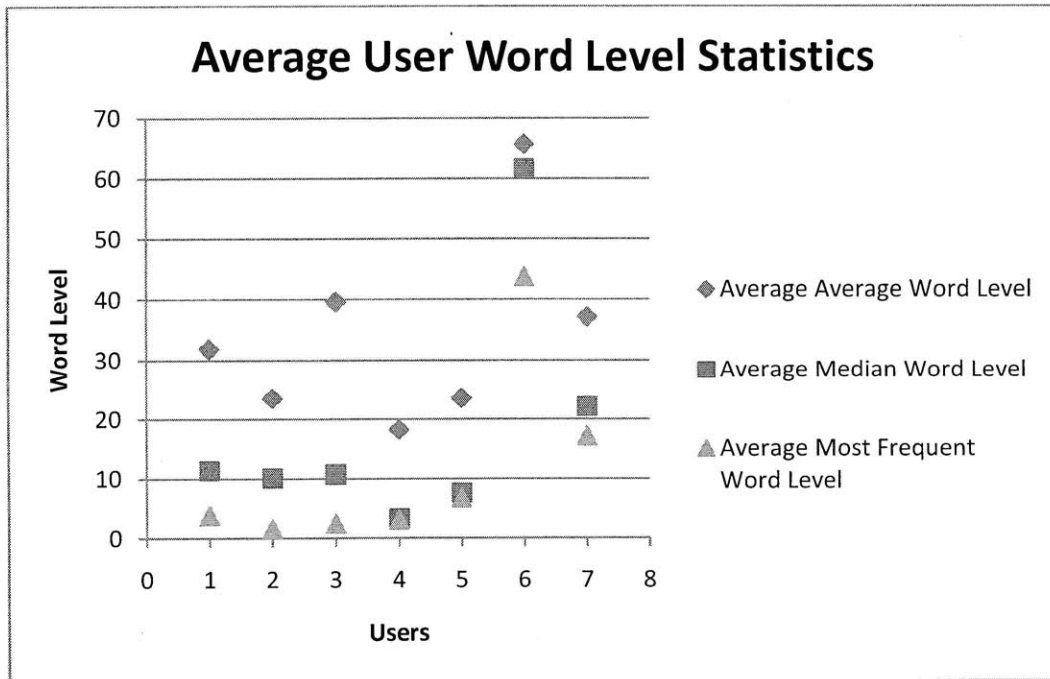
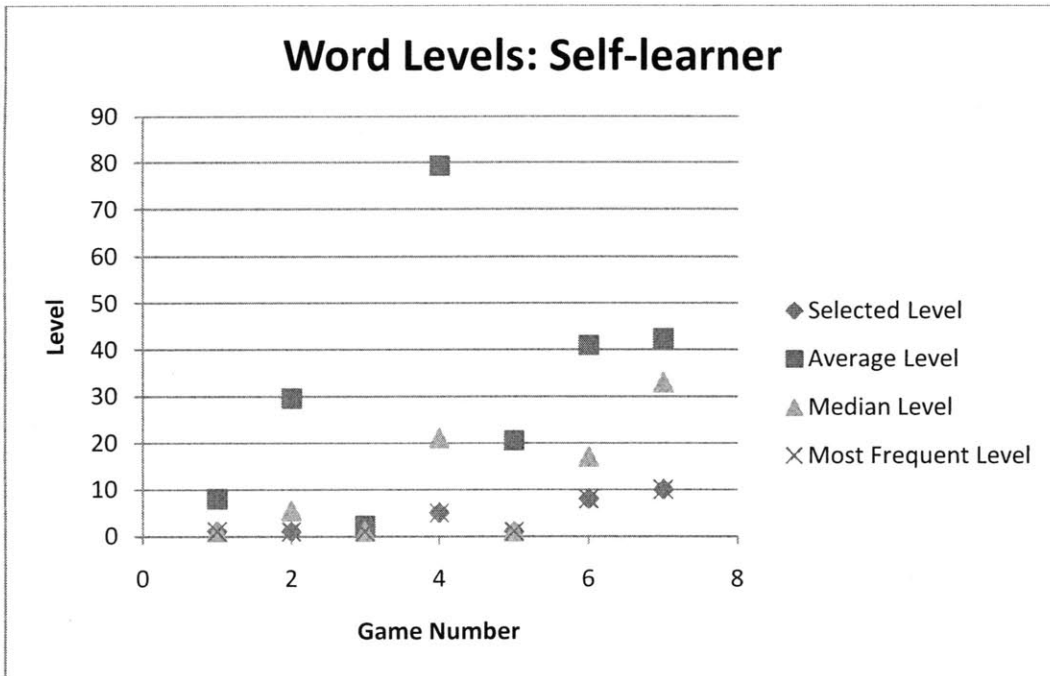
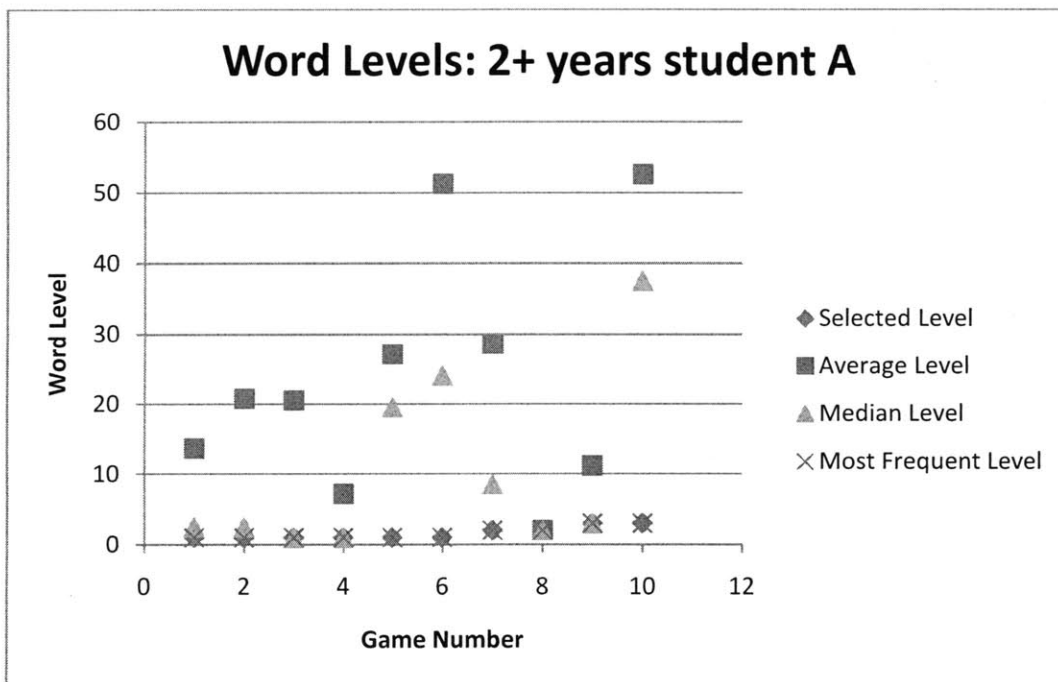


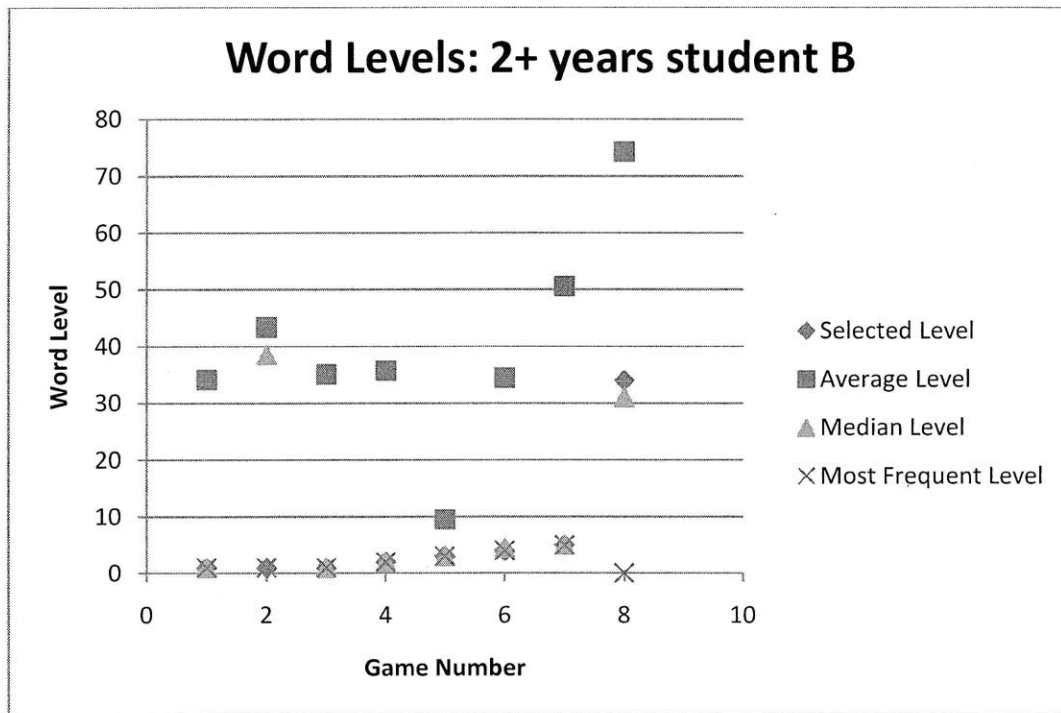
Figure 7a: The comparison of averages of the average word levels, median word levels and most frequent word levels for each user across all his/her games.



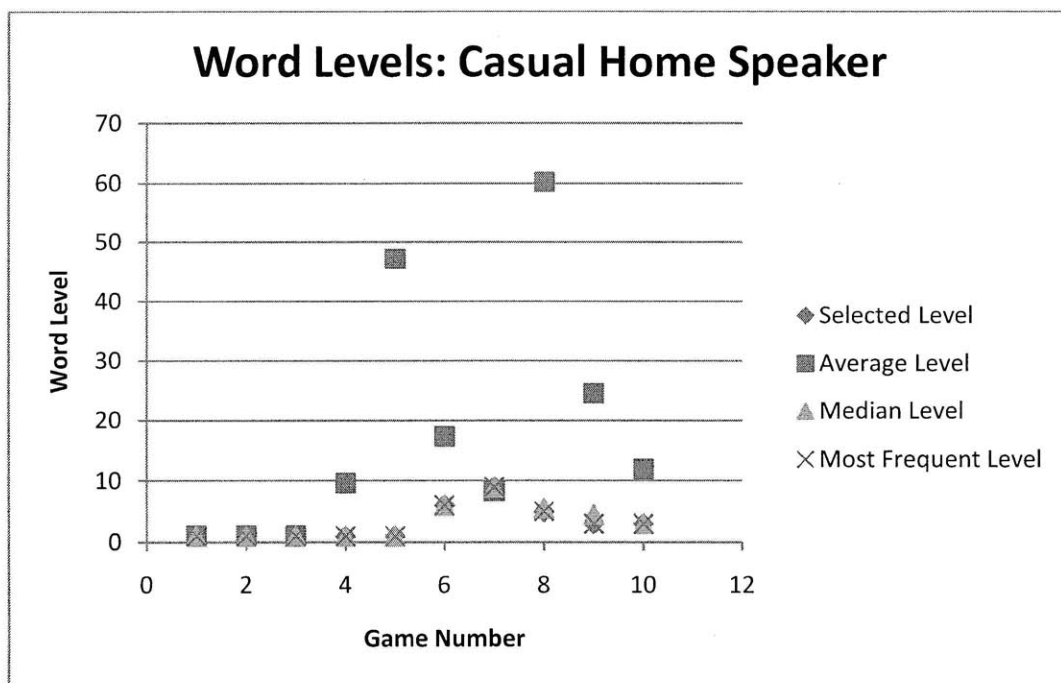
**Figure 7b:** The comparison of the average word levels, median word levels and most frequent word levels for each game of the self-learner user.



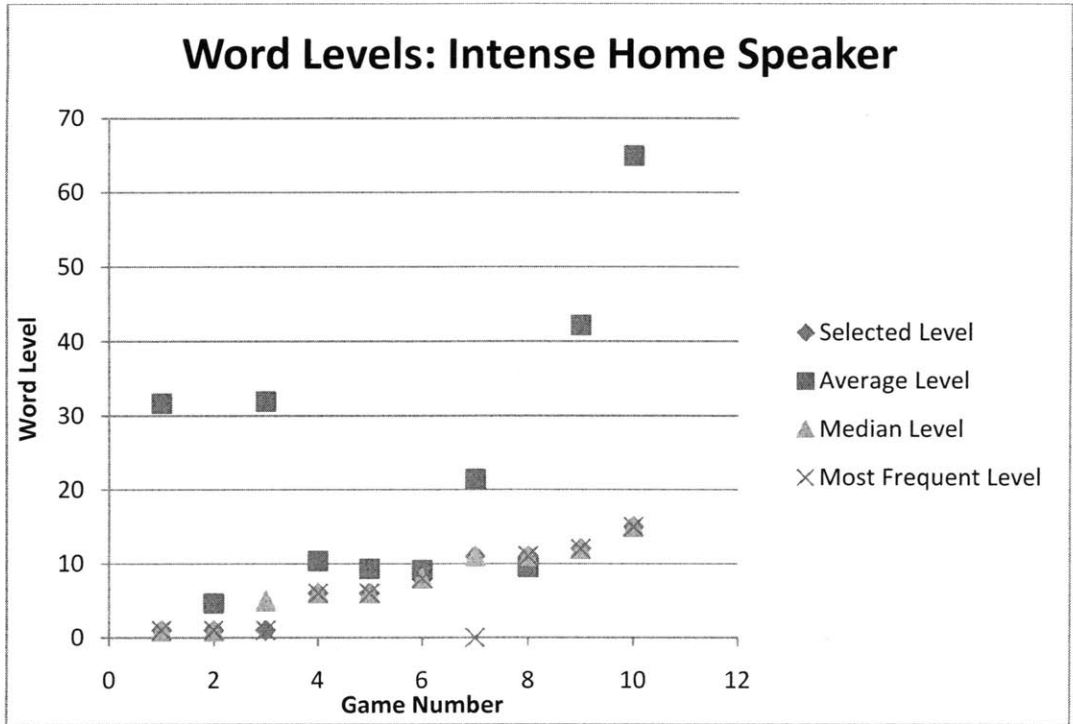
**Figure 7c:** The comparison of the average word levels, median word levels and most frequent word levels for each game of the 2+ year student A user.



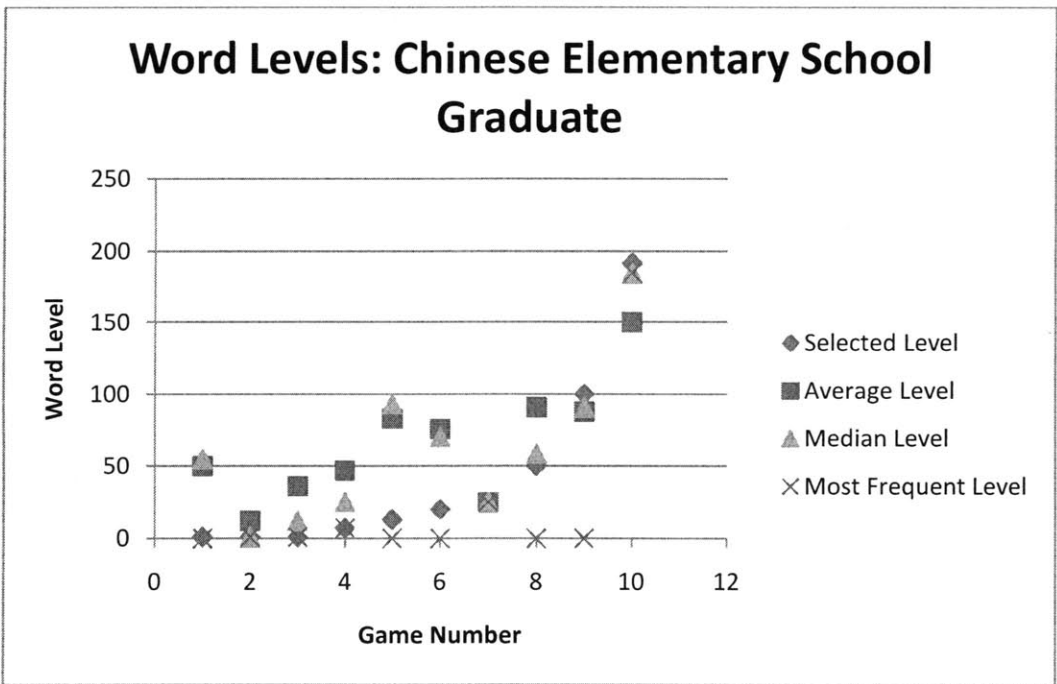
**Figure 7d:** The comparison of the average word levels, median word levels and most frequent word levels for each game of the 2+ year student B user.



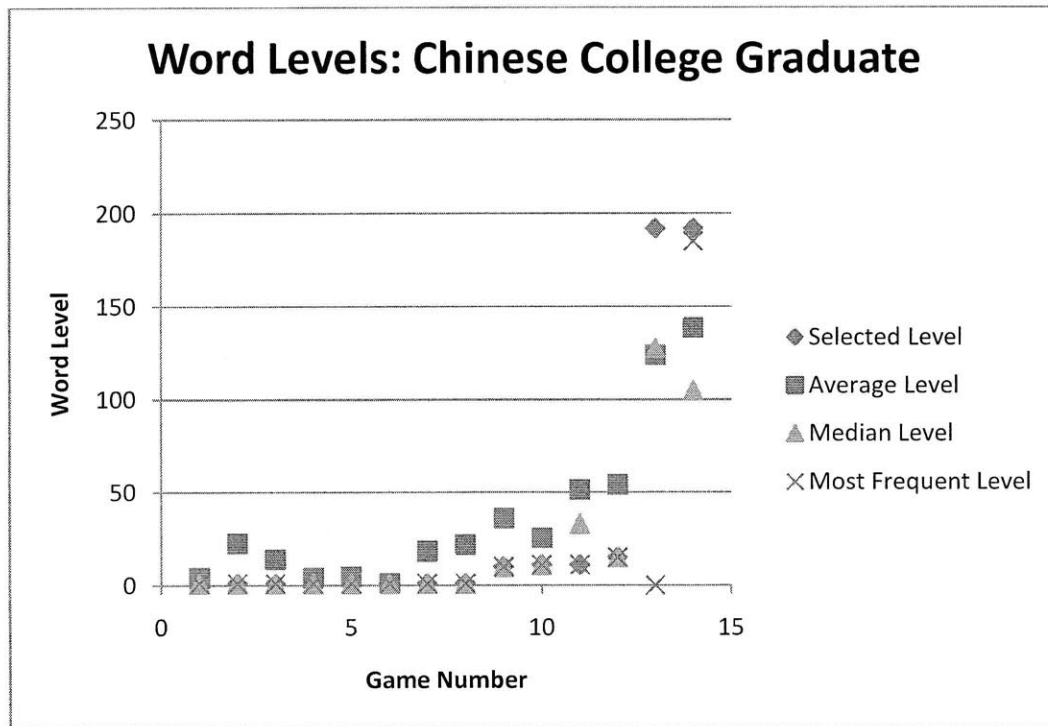
**Figure 7e:** The comparison of the average word levels, median word levels and most frequent word levels for each game of the casual home speaker user.



**Figure 7f:** The comparison of the average word levels, median word levels and most frequent word levels for each game of the intense home speaker user.

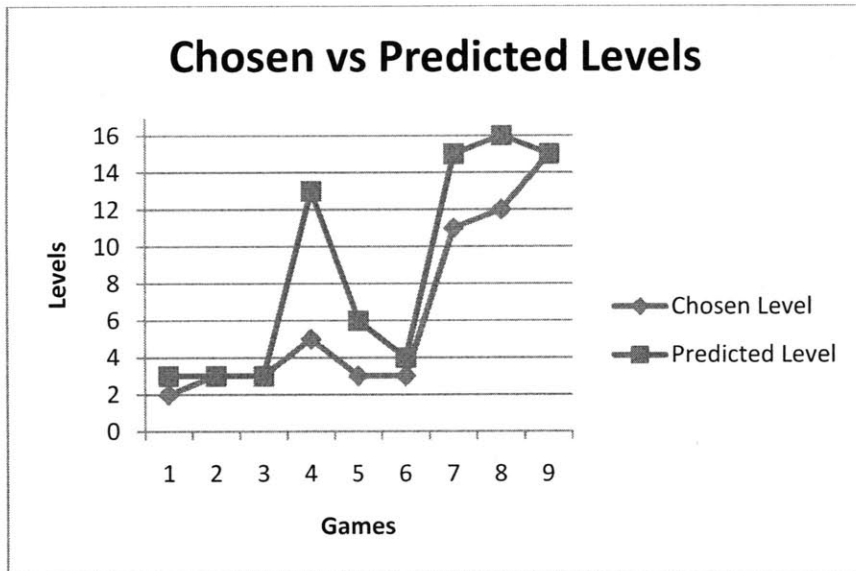


**Figure 7g:** The comparison of the average word levels, median word levels and most frequent word levels for each game of the Chinese elementary school graduate user.



**Figure 7h: The comparison of the average word levels, median word levels and most frequent word levels for each game of the Chinese college graduate user.**

In the third part of the study, where the user was free to choose his or her game level after each game, we silently ran the level-predicting algorithm from the second part of the study to see if it predicted similar levels to what the users chose. Except for a few cases when the two Chinese school graduates tried very high levels just for fun, the chosen level was always smaller than the predicted level (see Figure 8). This might have indicated that the level electing algorithm still jumped too fast compared to how a human would progress through the game levels.



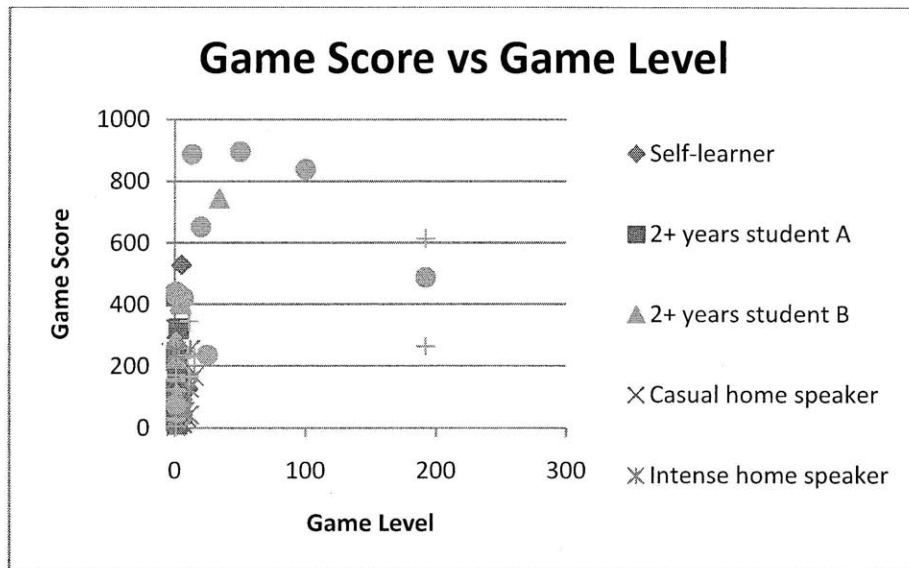
**Figure 8: Comparison between the (secretly) predicted level the chosen level in part 3 of the user study**

#### 4.6 Do more advanced users do better in the game?

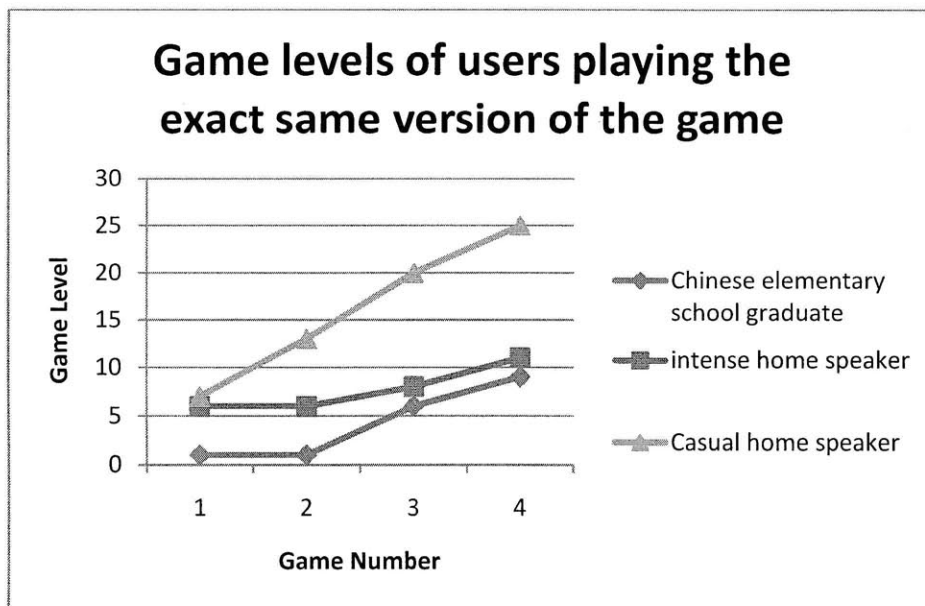
(Being better may mean receiving higher scores, achieving higher levels, making fewer mistakes or playing faster)

Achieving higher game levels may not be the primary measure for doing well in a game but since in Chinese Scrabble a high game level increased the chance of a high score (up to a level that was already too difficult for the user – see Figure 9), it is necessary for us to first analyze the game levels achieved by individual users. Since the automatic choice of a new level was not perfect, and the “Chinese college graduate” and both “2+ year students” did the study on an earlier version of the game with a more restricted level change algorithm, and the “self-learner” had to restart the game after the first 4 games, only the game levels of the “home speakers” and the “Chinese elementary school graduate” were suitable for comparing different levels that the individual users achieved during part 2 of the study (Figure 10). However, Figure 11 displaying all the users is useful for showing that the levels of all users were mostly increasing (except for

the big dump in the “self-learner” curve after she restarted the game), which is a good game property since a sharp decrease in game level might have been really discouraging for the player.

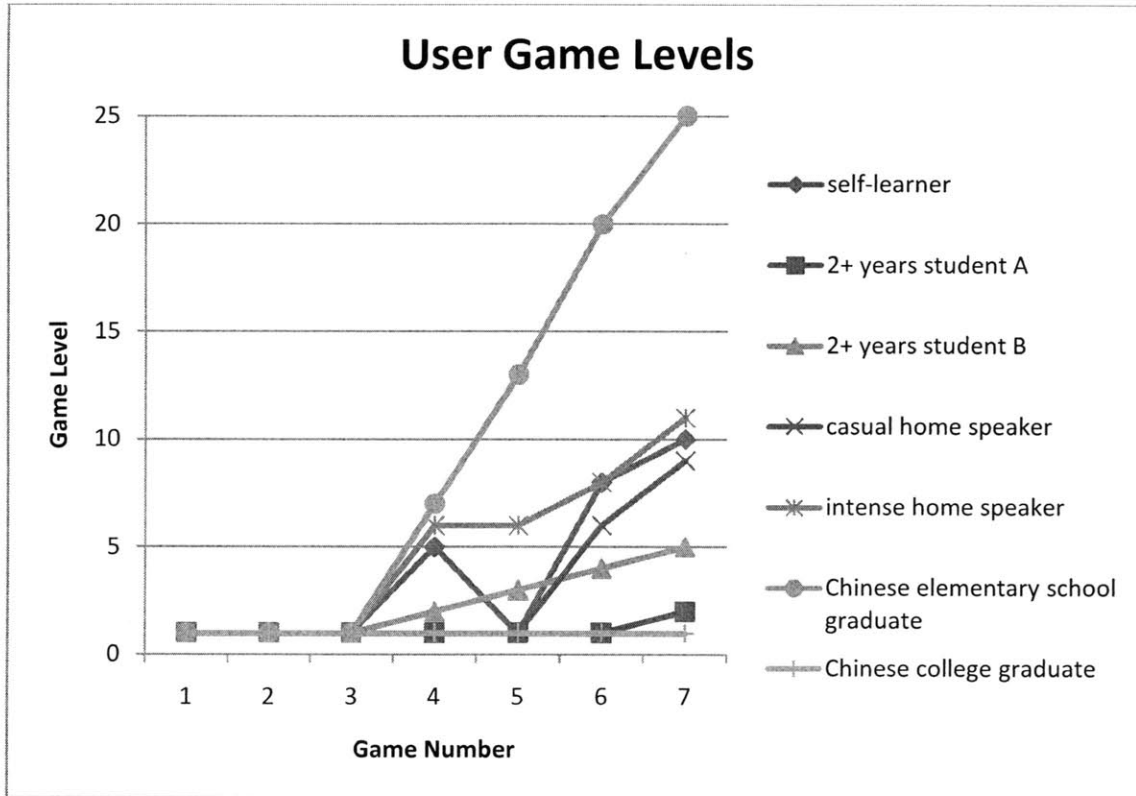


**Figure 9: Relationship between the game score and game levels for all users and all games**



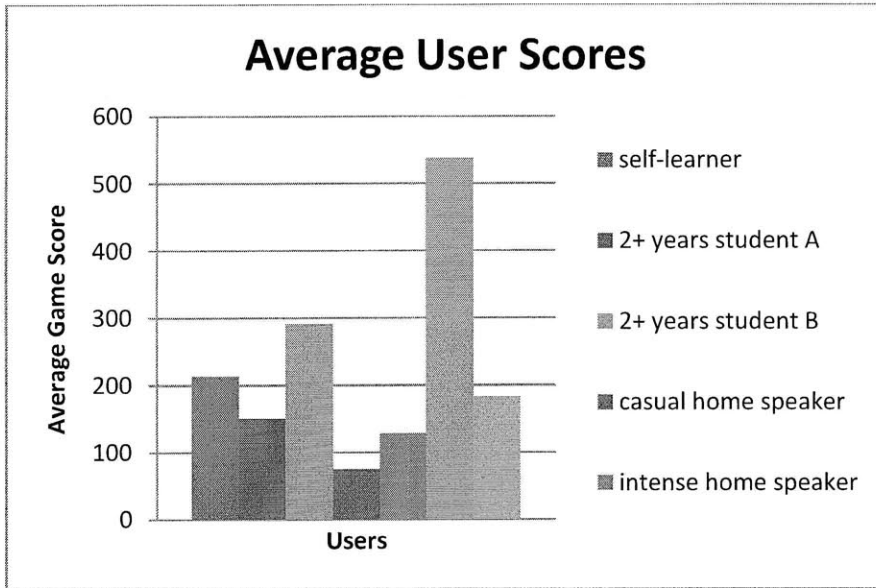
**Figure 10: Comparisons of game levels of 3 users in part 2 of the study.**





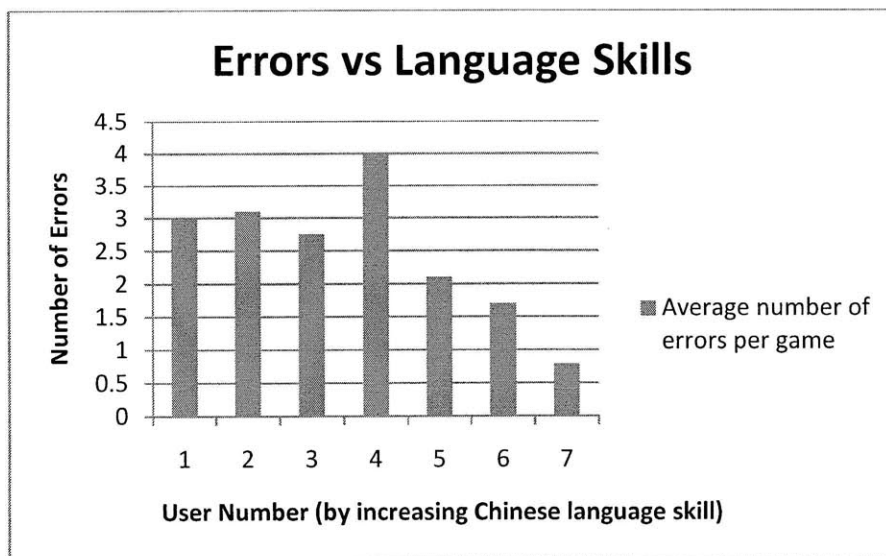
**Figure 11: Comparison of game levels of all users in parts 1 and 2 of the study**

Even though the average user scores were not always increasing with the increasing user language skills, there is one pairwise and one triplewise comparison in Figure 12 that confirms the hypothesis that more advanced users received higher scores. First, the “home speakers” and the “Chinese elementary school graduate”, who all played the same version of the game (with a little bit different scoring than the rest of the participants), achieved scores proportional to their increasing language proficiency. Second, the “2+ years student A” and the “Chinese college graduate”, who also played the same version of the game (different from the triple mentioned above), also achieved scores proportional to their increasing language proficiency.



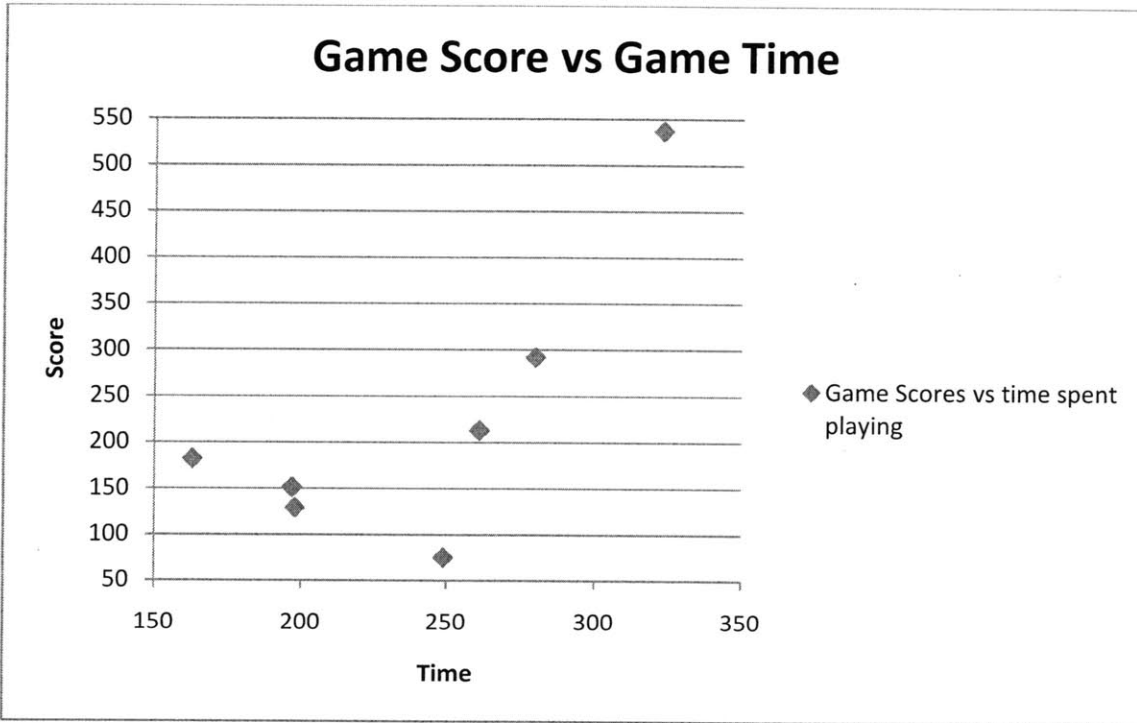
**Figure 12: Average user scores across all games**

Intuitively, the more advanced in Chinese the user is, the less mistakes (s)he should make. Figure 13 confirms this hypothesis. With only two exceptions out of seven participants, the average number of errors per game decreases with the increasing language skills of the study participants.



**Figure 13: Average number of errors per game for each user**

The game time did not seem to depend on the user language skill (Figure 14), but the game score did. In case of four users, the longer the user played, the higher score (s)he achieved. Moreover, the two users who played the game in just under 200s had very similar scores, though ordered inversely to their game time. The fact that the game score seemed to be mostly proportional to the game time may have indicated that the game required a lot of careful thinking and did not favor quick guessing.



**Figure 14: The relationship between average game score and average game time for each user**

## 5. Conclusion

In our pilot user study, Chinese Scrabble showed to be a game with a potential to both entertain and educate. Even though making a link between the toneless pinyin and words ended up being unintuitive for native Chinese speakers, it worked well for the foreign students of Chinese, who were our target audience. The data and playback logging proved very useful for a web-based user study as well as for analyzing user data to get a better idea about the patterns in user behaviors that may not have been easy to observe in real-time user study sessions.

### *5.1 Lessons Learned*

The main lesson I learned from working on this project was the importance of iterative development, user testing, and documentation, as well as the necessity of keeping a progress report, having a schedule and sticking to it.

Even though my work was constantly being reviewed by my supervisor and two graduate students in the Spoken Language Systems Group, having several rounds of user testing with first-time users would have been beneficial for quickly identifying common difficulties with the interface learnability and intuitiveness, as well as the feasibility of different features of the game. For example, through our pilot user study we discovered that for some users it was not clear that in speech mode they could use typing after the first unsuccessful attempt to speak, which was definitely due to insufficient interface intuitiveness. Other examples are the too high speed of the character flashing of the second level of help and an overwhelming number of words in the list of English translations of words that could be put on the board (a possible improvement would be to label these words with their level).

The biggest personal lesson for me was to have a high-level as well as week-to-week schedule and doing everything possible in order to stick to it, including leaving extra time for things beyond my direct control like user testing. Moreover, with a big project, it was very important to document my code and keep notes about my progress. Lastly, I learned that I should leave enough time for writing as well.

## **5.2 Future Work**

Besides “cosmetic” improvements to the game (for example, the issues mentioned in section 5.1 as well as a multiple-browser compatible user interface), there are several things that can be added to the game to improve the user experience, help with speech and language game research or a combination of both.

To improve the user experience, the main issues to work on are the prevention of time out all the time and the option of saving a game and resuming it at some other time. Second, it would be nice if the game allowed the user to upload his or her own vocabulary set that (s)he would like to focus on. The uploaded vocabulary could also be used to expand the static vocabulary of the game, which might be used for other language games and research. Also, adding more game customization options like making the game easier by keeping all solution words made only out of characters up to a particular school grade, or making the game harder by requiring neighboring pinyin to match in tone as well. Third, I believe making the game multi-player and creating accounts for individual players would greatly increase the popularity of the game. Lastly, once the “save game” option is implemented, it should be easy to create the originally planned Chinese Crosswords game extension, where the finished boards would be saved and then randomly dealt to players later. The users would be given English translations of the words and they could either fill in the crossword with pinyin (which would be transformed to Chinese characters on the board) or they could complete the crossword by choosing the locations of the given Chinese characters.

One of the possible research-related uses of the game would be to keep annotated recordings of user utterances, which could be done by having the user indicate somehow when the utterance was not correctly recognized and resubmit it by typing or speaking again. Besides more user studies with students of Chinese as a foreign language, it would be interesting to do a study with Chinese elementary school pupils to see if Chinese Scrabble and Chinese Crosswords could help them practice their reading skills.

Lastly, the game made me realize that even though the logging interface is really convenient for post-game analysis of the data, it would be even better if the playback could be stopped, rewound or fast-forwarded, so that future developers and researchers can easily focus on crucial moments of the recordings and also bring it up to speed if the regular speed playback is too slow.

## Appendix A: Chinese Characters from School Grades 1-6

1st grade: 170 characters

一 二 三 十 木 禾 上 下 土 个 八 入 大 天 人 火 文 六 七 儿 九 无 口 日 中 了 子 门 月 不  
 开 四 五 目 耳 头 米 见 白 田 电 也 长 山 出 飞 马 鸟 云 公 车 牛 羊 小 少 巾 牙 尺 毛 卜  
 又 心 风 力 手 水 广 升 足 走 方 巴 业 本 平 书 自 己 东 西 回 片 皮 生 里 果 几 用  
 的 正 雨 两 瓜 衣 来 年 左 右 半 业 齐 冬 说 友 话 春 朋 高 你 绿 笑 更 我 因 花 红 草 亲  
 放 岁 行 古 处 声 知 多 忙 洗 万 百 丁 齐 扫 母 爸 写 全 完 关 家 看 以 是 园 画 会 妈  
 明 午 收 女 气 太 早 去 亮 和 真 李 秀 房 千 空 香 网 听 远 唱 林 怕 桥 对 南 唱 总 先  
 时 赶 净 同 吗 吧 得 虫 很 蚂 河 呢 呀 哪 谁 沙 海 凉 怕 桥 对 南 唱 总 先 她 起 过  
 找 办 许 别 那 到 吓 叫 再 吹 分 要 成 动 晚 新 有 么 湖 在 变 什 冷 竹 跟 童 童 跟 童 童  
 进 边 道 常 兰 座 各 带 坐 急 名 发 成 成 成 成 成 成 成 成 成 成 成 成 成 成 成 成 成  
 非 苦 刚 座 各 带 坐 急 名 发 成 成 成 成 成 成 成 成 成 成 成 成 成 成 成 成 成  
 树 刚 座 各 带 坐 急 名 发 成 成 成 成 成 成 成 成 成 成 成 成 成 成 成 成 成

2nd grade: 649 characters

宜 实 色 华 谷 金 尽 层 丰 壮 波 浪 灯 作 字 苹 丽 劳 尤 其 区 巨 它 安 块 站 已 甲 豆  
 识 纷 经 如 好 娃 注 于 首 枝 枫 记 刘 胡 戏 棋 钢 观 弹 琴 养 休 伸 甜 歌 院 除 息 您 牵  
 困 员 青 宁 室 样 校 族 切 亿 教 响 班 欠 元 胡 包 戏 棋 钢 观 弹 琴 养 休 伸 甜 歌 院 除 息 您 牵  
 与 川 州 台 争 扬 哇 怪 挑 根 幸 之 体 操 灵 晨 舰 代 世 雷 迈 省 度 而 蝉 蛛 批 望 算 查  
 每 狸 猴 抓 祝 福 勾 幸 之 体 操 灵 晨 舰 代 世 雷 迈 省 度 而 蝉 蛛 批 望 算 查  
 始 张 猫 轻 飘 店 然 岛 府 志 煌 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 鸡 味 能 追 泉 该 束 煌 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 味 能 追 泉 该 束 煌 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 能 追 泉 该 束 煌 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 追 泉 该 束 煌 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 泉 该 束 煌 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 该 束 煌 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 束 煌 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 煌 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 窗 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 导 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 叭 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 玉 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 组 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 珍 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 珠 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 数 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 钻 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 研 睡 距 离 油 检 查 团 斥 责 灰 炎  
 睡 距 离 油 检 查 团 斥 责 灰 炎  
 距 离 油 检 查 团 斥 责 灰 炎  
 离 油 检 查 团 斥 责 灰 炎  
 油 检 查 团 斥 责 灰 炎  
 检 查 团 斥 责 灰 炎  
 查 团 斥 责 灰 炎  
 团 斥 责 灰 炎  
 斥 责 灰 炎  
 责 灰 炎  
 灰 炎  
 炎

3rd grade: 502 characters

坪 坝 戴 招 蝴 蝶 孔 雀 舞 铜 粗 尾 要 装 劲 绒 朝 些 钓 察 瓣 拢 掌 趣 爬 峰 顶 似 苍 仰  
 咱 奋 勇 居 郊 散 步 胸 脯 渣 或 者 敢 惜 低 诚 基 突 按 摆 弄 准 备 侧 胶 卷 辆 秘 杂  
 社 著 藏 村 致 纳 击 坚 固 栏 案 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 命 奔 修 仆 跨 坚 固 栏 案 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 播 尘 跨 坚 固 栏 案 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 尘 跨 坚 固 栏 案 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 跨 坚 固 栏 案 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 坚 固 栏 案 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 固 栏 案 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 栏 案 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 案 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 止 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 爪 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 贵 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 境 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 授 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 断 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 楚 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 孤 帆 蓝 懒 划 威 武 掠 颜 形  
 帆 蓝 懒 划 威 武 掠 颜 形  
 蓝 懒 划 威 武 掠 颜 形  
 懒 划 威 武 掠 颜 形  
 划 威 武 掠 颜 形  
 威 武 掠 颜 形  
 武 掠 颜 形  
 掠 颜 形  
 颜 形

屋述箴织烛秆萤汪腾  
 免绪贴耕搬翠旷私融  
 锯烈仁未壁宾愁窃顾  
 斧吞盼耘隔议泊否励  
 鞋况卡昼增椅移依掠  
 挖投即局呆敞亭涛浆  
 退势糖贯寸描味随旋  
 兵姿猜搂匀党虽翻螺  
 神愿买扎卵烂腐裳股  
 战受裕探芝灿津仿涌痛  
 坦值宽倾例席汁姿叠愈  
 炮素担套诲礼唇胀扬惨  
 攻朴矛毫悉厅滑莲庆悲  
 龟捧获膝悦阶触之延弦  
 架荒堵惕晌登软挤传雁  
 般感激警肃智慧刺挨尝猎  
 持吵索拨审智桂揉编箭  
 集怒俩赛哈案狭须况魏  
 盾恼顺衬却且贪镇缺锄  
 矛嘴险荡幅固伺蹲效捡  
 舌虚危宙而毁晃庙疑桩  
 刮谦茶宇排既锐猛除撞  
 软陶绳闹嘱创脱跪扶窠  
 村懂系斑叮跨泡茄职待  
 药阶瓶墨苔砌泛合诊株  
 献敌育橘莓伟疾寿珍碰  
 浸管励嚷纶匠衫盈属偷  
 线初鼓稳稚捧赤喂煤铛  
 视难通端蓬肩腹烧镜盗  
 挡抢普鹰桑慈绣簇类掩

4th grade: 366 characters

狂仇伏挖型纠吱愣扮组衰贺  
 梭膛笋匣瞧瘦桶庞逼皇鬓邮  
 谜规镰亚逐堡绎搏婆损偶辅  
 慰臂拇拖曲获络哑巫估唯访  
 持暴丈铲柄践哺浑芜额帆屏  
 哀顽亩绑漾曾酒幼邦稍孟购  
 忧牺耀绳拂扶郭拯港昏充捷  
 弛艰驾舱隙纽莺躯嘛适妻疗  
 厉怒哄潜虎丧惠绒递聚吁预  
 奔枪堤尚涨距魄奈吻郁协菌  
 务攻痕论恢挣凝嗅鬼呈圣杀  
 段占朱陷余跃峭辛翹罗贡蒸  
 戒营葱拴霎纵垒嫌牵浙硕炭  
 馆役煌宋崩翼屯疚愈矮遗蔬  
 绍叨辉源浩剧砖愧忍慕凯殖  
 鲁爆阁资犹抛蜒械荒榴浸繁  
 抗隆爽验踮启蜿览倔价素兽  
 驻轰昆择沸付峻茆责榨予钢  
 衅投栏选闷集崇坚殃茅维抵  
 侵俯漆训薄改扁梯遭咐执亿  
 楚捆廊闯罩碎藻谓辟吩奉竭  
 淘噪殿程据即乏坦胆溉略拳  
 概印啼钟潮夹纯治跌灌销咽  
 藏抄猿饮炉殊灼仅摔渠统鸣  
 鼻替陵配瀑罪质歉要饶凡泣  
 甚锅辞糖庐犯饥俺腔烦唐垫  
 厘凳帝抚倍企躺酬稿徒宏睫妙  
 雕若岭缝避促轧缸痒褂苏迫恋  
 联益含折锯搏乞拐凭绸征克宜  
 歇赢鹜丘鞋陈掏旬虑早般查羨

5th grade: 320 characters

枣吵屈鲸托番滟纛酷捅  
 餐混豪畜哇旦激磺劣庸  
 逝侦罢贼哗赖扣硫啸曼  
 悼谣磨稻筷汲录苇坨喻  
 仪捣斩缚卜菱汹遮伍纳  
 哲道崖颈萝疆恬幔测典  
 误阻悬厨逮介缕眺黏伦  
 耽糕尸允棍晰粼滔湃赴  
 拘暂葛租逗戈键策澎狷  
 肌锁挥拜掺蓑盲渡渗召  
 超矫晋郑舀笠陌遣辣钧  
 杖阅寇胯嘻塞瞎瑜拭委狱  
 诵检梁扭钳符蜡率掸疯樱  
 衡陪寂捏燥谱券序脂鸯惩  
 奏洲耸瑟朴倦幽蔑蝇褐憋  
 货腮叉残糙沾茵讽珀瓢凄  
 溜饰艇吟骤添莱讥琥撤惧  
 缘挽尼承搜拢瞄惑瞬扯喻  
 燃耐跤堪掘俏鸥瞪政溃笛  
 则槐杯勉址俊咧臆楠诡监  
 溢拄貌晕慎律龇忌鸵妄杈  
 截嗽茶哼弃究唬砸羚援噢  
 杰咳蹄施蔽呻撕秧栌莹寄  
 润奥驰剂宅核桅秉烤覆榜  
 镀摊鞭鞭醉胎痰航县炙豹抑  
 泻悟限拒判肆艘循趾塑漏  
 裹嚼骏匪肺尔勃载颌秃绪  
 镶骂勒龄吨蹋雅置椎倚释  
 刹冠嫂刘滤糟炼抹骼僵罚  
 范搞耗沃肢批锻亦盍袭刁

6th grade: 200 characters

酿歧册嘹廉掀  
 凌皱脾溅诺怨  
 绵版缩杈臣帐  
 筏篇釜葬壁搁  
 障誉肤炊丞魂  
 峦迪叙标呐烁  
 翡晖炕旺插榭  
 瑕碟吉鸿寨填  
 澜浏粥迂督梗  
 漓咨筒绞彻妒哎  
 脉咨筒绞彻妒哎  
 辈岷袄舅煎焰履  
 谋丸爹押羹蜚吾  
 挪礴摩刑蔡兜誉  
 袖睹锈胖藐裸盾  
 帜滥拒宪竣徊矛  
 瞻慨慢怖岔徘寓  
 徐慷枕僻浆俐构  
 钮赐谚娱绘伶溶  
 宣矿兆阙扛碗昌  
 泽渺瑞汤毅洛湛  
 汇埃俗沧嘲涕铭  
 檐诞掷孟控乃铸  
 檐诞掷孟控乃铸  
 协媒屑曰挠祭鼎  
 焚饕甸矣申荆熬  
 锤偿茸俱拆袍勺  
 砚藉暇凉滨辱霉  
 综帖梢欠敏侮魁  
 伐妥巢唉萍御揍  
 厦剥谨熄魅颇寡

## Appendix B: User Survey

### Chinese Scrabble User Survey

Thank you for playing Chinese Scrabble as a part of our user study!

To complete the study, please fill in the survey below.

If a question asks you to rank something from 1 to 5, usually 1 = BAD or LITTLE and 5 = GOOD or A LOT, but you should still follow the descriptions of the scale for that question.

Please provide any additional comments/feedback at the end of the multiple-choice survey.

#### Please provide your email address

This is needed only if you received study invitation through email, otherwise it's optional. We won't share your email address with third parties.

#### General Questions

##### What is your level of Chinese?

- I have never studied Chinese.
- I have been studying Chinese for less than a year.
- I have been studying Chinese for 1-2 years.
- I have been studying Chinese for more than 2 years.
- I speak Chinese at home, but have not attended a school in China.
- I live/used to live in China and attended at least elementary school in China.
- Other:

##### On the scale of 1 (worst) through 5 (best) how much did you enjoy playing the game?

1   2   3   4   5

It wasn't enjoyable at all.      I had a lot of fun!

##### On the scale of 1 (least) through 5 (most) how probable it is that you would play the game again if you had a chance?

1   2   3   4   5

not probable at all      I'd definitely play it again



### Game Purpose Questions - New Vocabulary

The following questions are about vocabulary that you first encountered while playing Chinese Scrabble today. (We will call it "new vocabulary" in this survey.)

How satisfied are you with the amount of Chinese vocabulary that the game helped you learn today?

1 2 3 4 5

Not satisfied      Very satisfied

How confident are you in the following aspects of the new vocabulary?

Toneless pinyin and English meaning?

1 2 3 4 5

not confident at all      confident with most new words

Pinyin \*with tones\* and the English meaning?

1 2 3 4 5

not confident at all      confident with most new words

Chinese characters and the English meaning?

1 2 3 4 5

not confident at all      confident with most new words

### Game Purpose Questions - Vocabulary Review

The following questions are about the vocabulary that you've encountered before but it was pretty rusty prior to playing Chinese Scrabble today. (We will call it "reviewed vocabulary" in this survey.)

How satisfied are you with the amount of Chinese vocabulary that the game helped you review today?

1 2 3 4 5

Not satisfied      Very satisfied

How confident are you in the following aspects of the reviewed vocabulary?

**Toneless pinyin and the English meaning?**

1 2 3 4 5

not confident at all      confident with most new words

**Pinyin \*with tones\* and the English meaning?**

1 2 3 4 5

not confident at all      confident with most new words

**Chinese characters and the English meaning?**

1 2 3 4 5

not confident at all      confident with most new words

**User-friendliness Questions**

**How easy was it to learn how to play the game?**

1 2 3 4 5

not easy = hard      easy

**How many games did it take you (would you require) to get comfortable playing the game?**

- 1 game or less
- 2-3 games
- 4-10 games
- more than 10 games

**After you learned how to play the game, how efficient did you find the tasks you used in the game?**

1 2 3 4 5

the game was pretty cumbersome      the game was very smooth

**Speech, Typing and Tones Questions**

**How much did you like playing the game with speech?**

1 2 3 4 5

not at all      it was great

How much did you like playing the game with typing the words?

1 2 3 4 5

not at all      it was great

How much did you like playing the game \*with\* tones?

1 2 3 4 5

not at all      it was great

How much did you like playing the game \*without\* tones?

1 2 3 4 5

not at all      it was great

### Questions about Game Settings

Rank the game settings below according to your preference.

Both speech and tones optional.

1 2 3 4 5

did not like it.      liked it a lot!

Speech required, tones optional.

1 2 3 4 5

did not like it.      liked it a lot!

Tones required, speech optional.

1 2 3 4 5

did not like it.      liked it a lot!

Both speech and tones required.

1 2 3 4 5

did not like it.      liked it a lot!

### Questions about Game Help

If you didn't use help at all, answer only the first question in this section and then proceed to the next section.

Did you notice that you could get help in the game?

Yes

No

**How useful was the first level of help? (highlighting one pinyin)**

Answer this question only if you used first level of help.

1 2 3 4 5

not useful at all      very useful

**Did the first level of help remind you of words that you knew but momentarily forgot?**

Answer this question only if you used first level of help.

1 2 3 4 5

not at all      yes, very frequently

**Did the first level of help encourage you to "guess" words that you haven't known before?**

Answer this question only if you used first level of help.

1 2 3 4 5

not at all      yes, very frequently

**How useful was the second level of help? (highlighting all pinyin on board for each remaining pinyin)**

Answer this question only if you used second level of help.

1 2 3 4 5

not useful at all      very useful

**Did the second level of help remind you of words that you knew but momentarily forgot?**

Answer this question only if you used second level of help.

1 2 3 4 5

not at all      yes, very frequently

**Did the second level of help encourage you to "guess" words that you haven't known before?**

Answer this question only if you used second level of help.

1 2 3 4 5

Not at all      Yes, very frequently

**Did you find the "Show Next Step" button useful?**

Answer this question only if you used "Show Next Step".

- Yes
- No

## Questions about Game Hints

Again, answer these only for the the hints that you tried out.

**How useful did you find the option of viewing the English meaning of words that could be possibly placed on the board? ("Show English Solutions" button)**

1 2 3 4 5

not useful at all      very useful

**Which aspect of the information that appeared above the board when you moused over a word on the board was useful or interesting to you?**

Check all that applies.

- Reading (pinyin transcription) of words
- English meaning of words
- Both English and reading of words
- None

## Game Scoring Questions

**Did you like the automatic selection of your next level in part II of the study?**

1 2 3 4 5

not at all      very much - it was a lot of fun!

**Was the level chosen for you in part II approaching an appropriate level for you as you played more games?**

That is, if the game you just completed was too easy, was the next game a bit harder?

Alternatively, if the game you just completed was too hard, was the next game a bit easier? Was the last game about appropriate?

1 2 3 4 5

no, it was too easy      no, it was too hard

**Did you prefer choosing the level yourself (part III of the study) or having it automatically chosen for you (part II of the study)?**

1 2 3 4 5

own choice of level      automatic assignment

Did the fact that you received points for submitting correct words motivate you?

1 2 3 4 5

no, it discouraged me      yes, it motivated me

Did you like how the game was scored? Do you have any suggestions for scoring it differently/better?

In summary, the scoring worked as follows: the basic score for a word without tones was equal to the level of the word (1-189). For correct tones, you received a 100% bonus to your score. If you tried to use tones, but made a mistake on the first attempt but corrected the tones on the second attempt, you got a 10% bonus to your word score. For each of your utterances (up to 3), you got a bonus of 10% of your score. For each unsuccessful attempt or use of help your word score (including all the bonuses) decreased by half, but on the third unsuccessful attempt (or use of help) the word score dropped to zero. For example, if you tried to submit a non-existing word by speaking first, followed by a correct word with tones, your score would be  $(\text{word level}) \cdot (2)^2 \cdot (1.1)^2 \cdot (.5)$ .

Overall Comments, Questions or Suggestions

Submit

## References

1. T. Takezawa, E. Sumita, F. Sugaya, H. Yamamoto, and S. Yamamoto, "Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world," in Proc. of the Third Int. Conf. on Language Resources and Evaluation (LREC), Las Palmas, Spain, May 2002, pp. 147-152.
  2. <http://wami.csail.mit.edu/docs.php>
  3. <http://islands.csail.mit.edu/>
  4. [http://www.ehow.com/list\\_6393476\\_chinese-language-learning-games.html](http://www.ehow.com/list_6393476_chinese-language-learning-games.html)
  5. <http://www.digitaldialects.com/Chinese.htm>
  6. <http://www.nickjr.com/ni-hao-kai-lan/>
  7. <http://themandarincenter.org/materials/chinese-crosswords>
  8. <http://alumni.hkucs.org/~hclam/lam2001a.pdf>
  9. B. Yoshimoto, I. McGraw, and S. Seneff, "Rainbow Rummy: A Web-based Game for Vocabulary Acquisition using Computer-directed Speech," Proc. SIGSLaTe, Warwickshire, England, September 2009.
  10. Y. Xu and S. Seneff, "Speech-Based Interactive Games for Language Learning: Reading, Translation, and Question-Answering," International Journal of Computational Linguistics and Chinese Language Processing, vol. 14, no. 2. 2009
  11. Y. Xu, A. Goldie, and S. Seneff, "Automatic Question Generation and Answer Judging: A Q&A Game for Language Learning," Proc. SIGSLaTE, Warwickshire, England, September 2009.
  12. Y. Xu and S. Seneff, "Mandarin Learning Using Speech and Language Technologies: A Translation Game in the Travel Domain," Proc. ISCSLP, Kunming, China, December 2008.
  13. I. McGraw, B. Yoshimoto, and S. Seneff, "Speech-enabled Card Games for Incidental Vocabulary Acquisition in a Foreign Language," Speech Communication 2008.
  14. I. McGraw and S. Seneff, "Speech-enabled Card Games for Language Learners," Proc. AAAI, Chicago, Illinois, USA, July 2008.
- A. Gruenstein, I. McGraw, and I. Badr, "The WAMI Toolkit for Developing, Deploying, and Evaluating Web-Accessible Multimodal Interfaces," Proc. ICMI, Chania, Crete, Greece, October 2008.