Mastery Learning of Second Language through Asynchronous Modeling of Native Speakers in a Collaborative Mobile Game

Xiangmin Fan, Wencan Luo, Jingtao Wang
Department of Computer Science and LRDC,
University of Pittsburgh,
Pittsburgh, PA 15260, USA
{xiangmin, wencan, jingtaow}@cs.pitt.edu

ABSTRACT
Acquiring Chinese tones is often considered as the most difficult task in learning Chinese as a Second Language (CSL). Recently, ToneWars, a collaborative mobile learning game, demonstrated the feasibility and efficacy of connecting CSL learners with native speakers for tone learning. However, the synchronous gameplay nature in ToneWars can be hard to scale due to the time constraint and limited availability of native speakers. We present principled research to make ToneWars *scalable* and *sustainable*. First, we address the scalability issue via asynchronous modeling of native speakers. Second, we quantify whether a CSL learner achieves native level mastery for a specific phrase, and explore the use of fine-grained feedback on language mastery as a sustainable motivator for language learning. The insights in this research are generalizable to designing second language learning technologies beyond Chinese. In a longitudinal study with 18 CSL learners, we found that asynchronous gameplay significantly improved learning with an average gain of 29.7 tones and 16.4 syllables, and helped participants achieve native level mastery on 58.2 out of 69 phrases.

Author Keywords
Mobile Learning, Serious Games, Collaborative Learning, Evaluation, Mandarin Tones.

ACM Classification Keywords
H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

General Terms
Design, Experimentation, Human Factors.

INTRODUCTION
Mandarin Chinese, the world's most widely spoken language, has grown in popularity as a second language nowadays. In 2010, about 750,000 people took the Official Chinese Proficiency Test [24]. In 2015, 200,000 U.S. students were studying Mandarin; President Obama hopes to see the number quintuple to 1 million by 2020 [12].

There are several unique challenges for English speakers when learning Chinese as a Second Language (CSL). Learners need to get familiar with the *logographic* writing system, memorize more than three thousand frequently used characters, and acquire the tonal sound system. Among all the challenges, learning Chinese tones is often considered as the most difficult task [18, 28, 33]. Tones in Chinese determine *meaning* while tones in English are used for *grammatical and expressive inflection*. Multiple cross-linguistic studies [21, 33] suggest that linguistic experience plays an important role in tone perception and the source of difficulty in learning tones has generally been attributed to the interference from English stress and intonation systems for American students [33]. English listeners may perceive Mandarin high tones as stressed and low tones as unstressed; however, in Mandarin, the stress is realized more by the duration and amplitude than tones [31].

One common paradigm for effective second language learning is to delve into a native speaker environment [23, 25]. This idea has been leveraged by ToneWars [15] recently for engaging language learning. ToneWars connects CSL learners with native speakers in a collaborative mobile game. CSL Learners can practice tone recall, perception and production by directly competing with native speakers in ToneWars. Head et al [15] observed an average gain of 6.2 tones in short-term recall after 40-minute gameplay.

While ToneWars provides new insights on leveraging native speakers to assist second language acquisition in a collaborative learning game, the original study leaves a number of important questions unanswered. First, the synchronous nature of gameplay requires the same number of native speakers and CSL learners. Although native speakers reported compellingly favorable experiences in the lab study [15], there may not be sufficient native speakers to connect and compete with CSL learners due to time zone difference and the time constraint of native speakers. Second, given the drastic skill differences between learners

CHI 2017, May 06-11, 2017, Denver, CO, USA
 ISBN 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025544

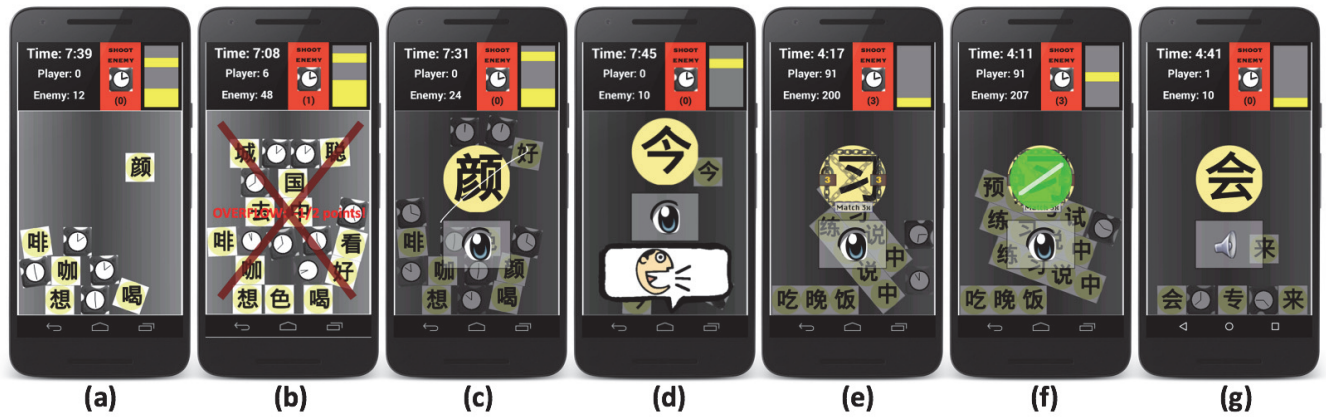

Figure 1. ToneWars Screenshots. (a) Phrases fall and collide; (b) The phrase stack overflows and clears, the player loses points; (c) The player traces a tone with a touch gesture to eliminate the character; (d) The player uses speech to input a tone; (e) A character locks after an incorrect guess; (f) Visual hint for a locked character; (g) The player listens to the audio hint by clicking the speaker button.

and native speakers, how can learners build and maintain their *self-confidence* in the competitions with native speakers? Since self-confidence is a basic determinant of learners' motivation in second language learning [7], we should guarantee that learners can gain their confidence, rather than being frustrated in the competition. Third, while the lab study showed promising results for improving learners' *short-term recall*, can ToneWars provide *measurable improvement* in learning outcome in a longitudinal study?

In this paper, we extend the prior work in three ways. First, we address the scalability issue via asynchronous modeling of native speakers. Specifically, we model the *interaction patterns* of native speakers for offline competition. By recording the key interaction patterns, learners are able to play against the pre-recorded experts at any time, regardless of native speaker availability. We also model native speakers' *language skills* in fine grain (i.e. phrase level tone recall) and use it as the goal for learners to achieve *"bite-sized"* native level mastery. Second, based on the modeling of native speakers' language skills, we propose a metric *"Native Level Index"* (***NLI***) to measure whether a CSL learner achieves native level mastery for a specific phrase at a given moment, as a quantitative indicator of a learner's performance. We believe such fine-grained modeling and feedback on language mastery can enhance learners' self-confidence thus can motivate them in a sustainable manner. Last, we conducted a 3-week study with 18 CSL learners to investigate the effectiveness of such modeling in improving learning and maintaining learners' motivation. We found that asynchronous gameplay can significantly facilitate learning, without losing user satisfaction. Participants had an averaged absolute learning gain of 29.7 and a relative gain of 64.4% for tone acquisition over the three weeks. We also found that CSL learners can achieve native level proficiency in 58.2 out of 69 phrases at the end of the study. All participants reported positive and favorable experiences with ToneWars.

Our research provides the following implications for researchers building mobile learning technology to aid second language acquisition beyond learning Chinese:

1) We demonstrate the feasibility of leveraging native speakers as both a *benchmark* for language mastery and a *motivator* for language learning. We believe that this direction could lead to interesting future research.

2) We propose a *scalable* approach to enable authentic competition and skill comparison with native speakers via modeling both the *interaction patterns* and *language skills* of native speakers asynchronously, and prove the effectiveness of such modeling in a longitudinal study.

3) We show the feasibility and efficacy of maintaining learners' sustained motivation through *fine-grained* skill modeling, feedback, and comparison with native speakers in the context of a competitive mobile game.

TONEWARS IN ACTION

To our knowledge, ToneWars (Figure 1) is the first language learning system connecting second language learners and native speakers via collaborative mobile gameplay. Learners practice Chinese tones via direct competition with native speakers during gameplay. ToneWars integrates gameplay elements from popular mobile games such as Tetris and Fruit Ninja to provide engaging experience for both native speakers and learners. In this section, we present a scenario to demonstrate how ToneWars works in action.

Claire is a first-year, second language learner of Mandarin. On a study break, she finds ToneWars, a game for practicing tones for phrases she learned in class. After connecting to a native speaker, she slashes Chinese phrases with the shapes of the tones of the characters (Figure 1.c). This kinesthetic touch gesture input was inspired by the widely adopted "in-the-air" tone tracing in Chinese classrooms [13], in which instructors ask students to trace tones of characters with their hands as they pronounce them. This approach can reinforce tone production [13, 26],

and can also make the learning experience more engaging and fun. Besides gesture input, Claire can also pronounce the character into the microphone (Figure 1.d). As she does this correctly, she eliminates the phrases and keeps her screen from overflowing (Figure 1.b) which will penalize her by halving her score. With each match, Claire earns points and saves a block. By tapping the "shoot enemy" button, she drops blocks to clutter the stack of her opponent. She monitors how cluttered her opponent's screen is through a preview pane in the top-right corner of the screen. Her score tells her when she is matching tones more successfully than the native speaker, which motivates her to continue playing and practicing.

When Claire makes a mistake, the character will be locked (Figure 1.e). To unlock and proceed, Claire clicks the hint button to view the *visual* hint (Figure 1.f) which shows the shape of the correct tone of that character. She can also listen to the *audio* hint (Figure 1.g) which is the spoken form of the character pre-recorded by a native speaker. After learning the correct tone, Claire enters the correct tone 4 times and the character becomes unlocked. This process emulates language learning drill exercises.

Claire may also play against a pre-recorded native speaker when there is no native speaker available. With asynchronous gameplay and offline performance competition, Claire experiences engagement and personal satisfaction in her improvement just like before.

RELATED WORK

Mandarin Tone Learning

Despite the challenges of tone acquisition [18, 28, 31, 33], research has shown that English speakers can be trained to successfully differentiate between Mandarin tones [22, 32], as well as to produce Mandarin tones [31]. Researchers have explored leveraging both *perception* training (e.g., by listening [32]) and *production* training (e.g., by speaking [22]) for tone acquisition. ToneWars provides CSL learners with both *perception* training and *production* training. Auditory (i.e. perception) training is provided to learners when they listen to audio hints (Figure 1.g) during gameplay. The speech input mode (Figure 1.d) helps learners' transition from declarative knowledge, where they can recognize a word, to productive knowledge, where they can use it [3].

Researchers also explored the idea of using color encoding [8, 13] and gestures [13, 26] to aid tone acquisition. The pitch contours make tones conducive to visual depiction [26]. Moreover, adding body motion information (e.g., gestures) could create memory traces that are even more multi-model and increase the learning robustness [26]. In real world Mandarin classrooms, gestures illustrating the spatial metaphor of pitches are widely adopted by instructors and students [13, 15], not only because they can reinforce learning, but also because they can make the classroom more engaging [15]. However, the social acceptability problem [27] may prevent CSL learners from tracing tones when practicing in public. Inspired by previous research, ToneWars enables the touch gesture input method (Figure 1.c). We believe that learners could benefit from the same audio-kinesthetic association as in the classroom with input that is more socially acceptable [27] outside the classroom.

Native Speakers in Second Language Acquisition

In recent years, a great deal of research and pedagogical experimentation has been conducted to investigate how to effectively leverage native speakers in second language education [17, 23, 25, 29, 31]. One important recommendation is to allow native speakers to play the role of an expert assisting the learner in improving both linguistic and cognitive skills related to the language. Native speakers can offer authentic language discourse to help language learners acquire new lexical items and correct grammatical structures [23, 25]. Moreover, the experience could help learners gain confidence and motivate them to engage in conversations in the future [15, 23]. However, there are two major limitations. First, native speakers are usually inaccessible to learners. Second, according to *"the input hypothesis"* [19], language acquisition can only occur when the input is comprehensible for the learner. In other words, language heard but not understood is of little or no use for learning purpose. However, in practice, native speakers may use some language that is beyond the comprehension level of the learners [15, 23]. Such perceptions of inequality may even lead to a lack of confidence and anxiety [23] towards further learning.

Besides the role of language tutors, native speakers' performance can also be treated as the "*native norm*" [31] for assessment purposes. For example, Wang et al. [31] normalized the pitch contours of Mandarin tones among native speakers and used the normalized $F0$ as the norm when evaluating learners' tone production.

Inspired by these research, ToneWars enables the interaction between CSL learners and native speakers in gameplay. Here the *"interaction"* means the competition and comparison of fine-grained language skills between CSL learners and native speakers, rather than "verbal" or "face-to-face" interactions which may lead to potential frustration and anxiety [15, 23]. By asynchronous modeling of native speakers, ToneWars is highly scalable so that learners can play against native speakers at any time.

Mobile Language Learning Systems

The mobile phone is a great platform to implement *anywhere*, *anytime* micro-learning opportunities, since it always accompanies its owner wherever she may go. There is a great deal of mobile language learning systems [1, 2, 5, 6, 9, 10, 11, 16, 20, 30] with which learners can leverage the brief fragments of free time that spaced throughout the day for language learning tasks.

A number of mobile applications [1, 9, 10, 11, 30] have been developed to address the challenges of learning Mandarin Chinese. Some of them focus on vocabulary learning in general [1, 10, 11], others focus on the tonal sound system [9] and the logographic writing system [30]. They were designed for either CSL learners [1, 9, 10, 11] or native speakers in elementary schools [30].

MicroMandarin [10] provides learners with vocabularies that are relevant to their locations (e.g., suggest "Cappuccino" in a Cafe). MemReflex [11] considers a broader sense of context-aware learning—adaptive to personal learning history. ToneWars is different with these two systems in that it is an educational game that provides learners both engaging gameplay and effective learning experiences. In addition, ToneWars specifically focuses on tone acquisition, while the other two systems focus on vocabulary learning in general.

The Multimedia Word and Drumming Strokes mobile games [30] were inspired by traditional Chinese group games and aimed to improve native speakers' language abilities. Players sit together and share a mobile phone in the gameplay. In comparison, ToneWars connects CSL learners with native speakers who are separated physically and aims to improve CSL learners' language abilities.

Tip Tap Tones [9] is a mobile game with the purpose of training CSL learners to acquire the tonal sound system. Tip Tap Tones provides aural tone perception training via single-player flashcard-style drills at single-character level. In comparison, ToneWars supports phrase-level tone learning and connects CSL learners with native speakers in multiplayer gameplay. In addition, ToneWars supports multi-modal interactions (e.g., gesture, speech) inspired by tone exercises in Mandarin classrooms.

ASYNCHRONOUS MODELING OF NATIVE SPEAKERS

As we discussed earlier, the scalability issue is the major challenge for ToneWars when deployed in the wild. Although native speakers reported favorable and engaging experience in previous lab study [15], there may not be sufficient native speakers to connect and compete with CSL learners due to time zone difference and the time constraint of native speakers. Therefore, we integrate asynchronous gameplay to ToneWars, so that learners are able to play and practice at any time, regardless of native speaker availability.

The asynchronous gameplay and competition should achieve two goals: 1) delivering authentic and competitive gameplay experiences for CSL learners; 2) enabling offline language skill comparison between learners and native speakers.

We achieve these two goals via asynchronous modeling of both *interaction patterns* and *language skills* of native speakers. By recording the key interaction patterns, learners are able to play against the pre-recorded experts at any time and we hypothesize that such experience could be as engaging as synchronous gameplay. Moreover, we model native speakers' language skills in fine grain (i.e. phrase level tone recall) and use it as the goal for learners to achieve *"bite-sized"* native level mastery. We believe that learners can be highly motivated when they can perform as well as native speakers, even on a subset of the language skills.

Interaction Pattern Modeling

In recording mode, ToneWars records native speakers' key interaction patterns along with the corresponding time stamps, including tapping to select a phrase, slashing a gesture to match, pressing the attack button, etc. When native speakers are not available, CSL learners play against the pre-recorded activity scripts. ToneWars plays the activity script in a continuous loop (Figure 2) when the duration of the recording is not sufficient.

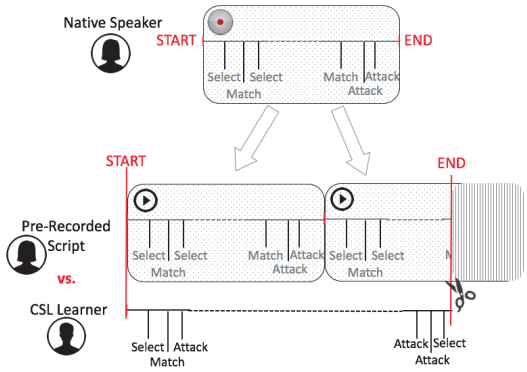

Figure 2. The activity script is played in a continuous loop if the length is not sufficient.

One major challenge for this approach is how to make a game where players felt they were playing in real-time against the recorded opponent. Real-time interaction between the competitors is a key element of making a competitive game engaging and fun. The problem with replaying past sessions against new players is that the real-time interactivity (i.e. action-and-reaction) is lost. For example, prerecorded players will not be able to fire back at their opponents out of 'spite' (i.e. attack right after they have been attacked), or time their attacks when their opponents are doing particularly well—all of their attacks are going to happen at prerecorded times. One compromise is to lower the time sensitivity of attacks; in other words, design the attacks such that they will have the same effect when they are replayed during a new competition. We decided to make the effect of attacks last longer—the blocks will remain active in the phrase stack for 20 seconds. Under such a design, an attack of dropping blocks will always drop blocks on and inconvenience the player, and such an effect will still match the intent of the player from the original game when replayed.

Even though the pre-recorded opponent cannot *react* to a player in real-time, they can still *interact* a lot. Besides the attacks with long-lasting effects, the preview pane in the top-right corner (Figure 1) allows players to monitor how

cluttered the opponent's screen is, which could also enhance the interactivity. We believe that these designs would make the competition still feel rich, even if the pre-recorded opponent does not really respond to player's actions.

Language Skill Modeling

As we discussed earlier, the asynchronous modeling of native speakers should also serve as a *motivator* for CSL learners. Due to the drastic skill differences between native speakers and language learners, it can be hard for learners to perform as well as native speakers on complex language tasks (e.g., presentation, public speech, etc.). However, if they can achieve native level proficiency on smaller and lower level tasks (e.g., tone recall task for certain characters/phrases), they may also be highly motivated. Therefore, we decide to model native speakers' language skills in fine grain to give learners a better chance to "win" in the offline competitions.

First of all, we need to find a metric in the game that can accurately measure the language proficiency of the players. We find that the elimination time of a block (i.e. the time needed from tapping the block to finishing the input of the correct tones so that the block is eliminated) is a key indicator of players' language proficiency. When the player is more proficient with the language, she can eliminate blocks faster since she needs less time for cognitive processing and could be able to save the time for recovering from incorrect guesses. Therefore, we use this as the main metric in the model.

To collect native speakers' data for modeling, we invited 7 native speakers to play ToneWars for 25 minutes each and recorded logs during the gameplay. Then we discarded the first 5 minutes, during which they got familiar with the game design and control mechanisms. During the gameplay, they eliminated phrases (lengths ranged from 1 to 5) combined by 100 Chinese characters[1]. In the paper-based test before the data collection, all of them achieved 100% accuracy in tone recall on this vocabulary set. Therefore, we believe this performance dataset can represent native level proficiency.

Originally we planned to use *Character Elimination Time* as the main metric, which is the time needed to eliminate one character. Later on we found that this metric is limited in that it is based on the assumption that the player needs the same amount of time to eliminate one certain character, no matter whether it appears alone (e.g., a single character '希'), or it appears in a multi-character phrase (e.g., the character '希' in the phrase '希望'). However, appearing in a multi-character phrase could give the player more context which may affect their response time. We did further analysis and found that the mean CET is longer when the character appears by itself than appearing in a multi-character phrase.

To be more accurate, we use *Phrase Elimination Time* ***(PET)*** as the main metric, which is the time between the player tapping a phrase to select, and the phrase getting eliminated once the correct gesture was detected. In this case, we treated the phrase as a whole, and the model considers the effect of context information included in phrases. We divided the phrases into five groups based on their length (i.e. from 1 to 5). For phrases with the same length, we got a list of *PET* from the 7 native speakers. Figure 3 shows an example of the distribution of *PET*s for phrases with "L=1". Table 1 shows the mean and standard deviation of each Gaussian distribution. These models demonstrate the native speakers' performance with ToneWars.

Table 1. Parameters of Gaussian distributions describing native-level performance.

Phrase Length	**μ** (ms)	**σ**	Native-Level Index **(μ+σ)**
1	954.14	363.38	<= 1317.52
2	1848.19	537.07	<= 2385.26
3	2709.71	494.84	<= 3204.55
4	3353.46	558.41	<= 3911.87
5	4278.91	981.85	<= 5260.76

To determine whether a CSL learner achieves native-level proficiency, we define "μ+σ" as "*Native-Level Index*" (***NLI***), which serves as the decision bound. Generally speaking, if a CSL learner can achieve "Native-Level Index" on a certain phrase, he/she performs better than at least ~16% of the native speaker players on that phrase (Figure 3).

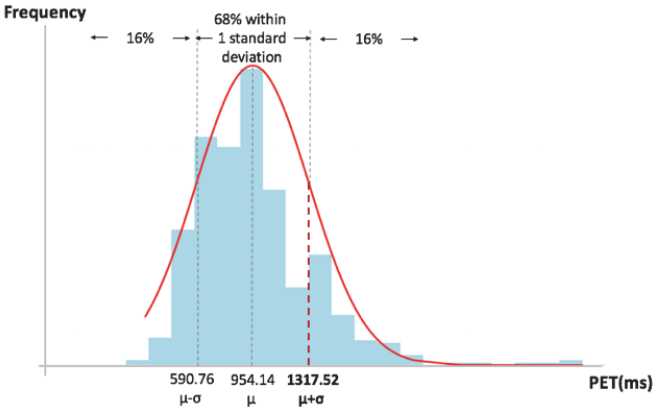

Figure 3. Native-level proficiency in Gaussian distribution.

EVALUATION

We conducted a 3-week study with 18 CSL learners to investigate the following three questions:

- While our previous lab study [15] showed promising results for improving learners' *short-term recall*, can ToneWars (with asynchronous gameplay) provide

[1] The vocabulary set was the same with the one used in the 3-week study presented later.

measurable improvement in learning outcome in a longitudinal study?

- Is asynchronous gameplay engaging enough so that it can motivate CSL learners in a sustainable manner?
- What are the strengths and limitations of different feedback types (visual vs. audio) and how can they affect learning? This is another unanswered questions in our previous lab study [15] and we believe the answer could be inspirational for researchers building language learning systems in the future.

During the study, we asked participants to visit our lab once or twice (based on their schedules) per week for 3 weeks. During their visits, they were asked to play ToneWars for 1 hour per week, in total 3 hours. They were exposed to both visual feedback and audio feedback during the study. They were told explicitly that the native speaker opponent in the game was rebuilt from a native speaker's activity logs. We conducted pre-test and post-test to measure the learning gains. The user activity logs were recorded during the gameplay. We also conducted semi-structured interviews to solicit their subjective feedback.

Participants and Apparatus

We recruited 20 participants for the study from two local universities. 2 of them quitted the study after the first week because of their tight schedules. The other 18 participants completed the study and we only report the results gathered from them. All of the participants were native English speakers actively learning Mandarin Chinese. Self-reported Chinese learning experiences were distributed as: less than 1 year=3, 1-2 years=10, 2-3 years=4, and 3-4years=1. Besides Mandarin, 6 of them also expressed experiences in learning other languages, including Cantonese, Hindi, Spanish, and Japanese.

Ages of the participants ranged from 18 – 32 (median age 21), with 7 males and 11 females. We compensated participants $30 for their time at the end of the study.

Over the course of 3-week study, participants used an LG Nexus 5 smartphone, which has a 4.95 inch, 1920 x 1080 pixel display, Quad-Core 2.3 GHz Krait 400 Processor, and runs the Android 5.0 OS.

Learning Materials

We worked with a CSL instructor to select 100 characters for the study from *Integrated Chinese Part 1 (IC1)*, a popular CSL textbook in North America. To create the audio feedback, the instructor recorded the pronunciations of the characters which were loaded in ToneWars afterwards. We suggested the instructor speak them slowly and to exaggerate the tone as she might do for novice learners.

Method

We adopted a within-subjects study design based on hint/feedback types—visual vs. audio. We first divided the 100 characters into two groups (group A and group B) of roughly equal difficulty. Nine participants used audio feedback version to learn group A and use visual feedback version to learn group B. The other nine participants use audio feedback to learn group B and use visual feedback to learn group A. In each week, each participant spent 1 hour in gameplay, which was further divided into 12 5-minute sessions (6 audio feedback sessions and 6 visual feedback sessions, the order was counter-balanced). Participants can take a rest any time they wanted between the 5-minute sessions. The order of character appearances was randomized in each session. During gameplay, we collected logs of significant game actions users took, including all match attempts, gestures completed, and interactions between the player and the pre-recorded expert.

We conducted a pre-test at the beginning of the study to establish a baseline. Participants completed a quiz in which they determined tones and pinyin of the 100 characters loaded in ToneWars. Participants completed a post-test which was identical to the pre-test after finishing gameplay. We did not conduct a test in each week because we did not want the tests themselves to be a factor that could influence the participants' acquisition. After the study, participants contributed feedback through a paper-based survey and a semi-structured interview.

While some of them might have occasional pronunciation feedback from Chinese class, none of them reported that they had additional tone training for the duration. Since the learning materials were selected from *IC1* and no participant was learning *IC1* at the time of the study, we believed that the potential for confounding by attending Chinese class was low. Therefore, we concluded that a control group is unnecessary, which follows the study design of [4, 9].

RESULTS AND DISCUSSIONS

Learning Gain (Overall)

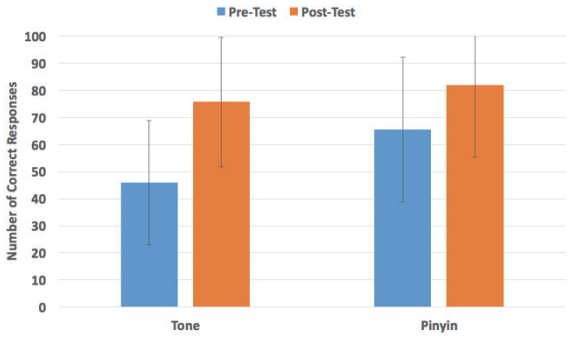

Figure 4. Participants performance of writing the tone and pinyin of 100 characters in pre-test and post-test.

All of the 18 participants improved from their pre-test level of tone recall and pinyin recall. For the pre-test, CSL learners correctly write down the pinyin of 65.6 (min=4, max=98, σ=26.8) characters, among which they correctly recognized the tones of 46.1 (min=2, max=74, σ=22.9) characters. After 3-week gameplay, they could correctly recognize pinyin of 82 (min=8, max=100, σ=26.5)

characters, and tones of 75.8 (min=10, max=95, σ=24.0) characters. The average gain of pinyin recognition is 16.4 (min=4, max=29, σ=9.3). The average gain of tone recognition is 29.7 (min=6, max=53, σ=13.7). Pairwise t-tests show that both of these two differences are significant (pinyin: 65.6 vs. 82, t(17)=7.51, p<0.001; tone: 46.1 vs. 75.8, t(17)=9.18, p<0.001).

The improvements were also perceived by the participants:

S7: *"I learned a lot more tones and it was more helpful to me than I thought it would be. I have understood more Chinese tones and how to say some of them. It has also taught me that the tone markings are important to learning Chinese."*

S13: *"I feel like I have corrected my knowledge of pinyin tone marks for multiple Chinese characters and am now more confident speaking those words because I know how to pronounce them with their correct tones."*

Even though every participant improved, the learning gains varied. S8, who had the smallest gain in both pinyin and tone recall, always exploited hint to preview the answer before making a guess during the gameplay. He found that this was an easy way to *"earn points quickly"* and *"avoid being locked"*. At the same time, based on our observation, this participant did not try to internalize the correct answer for the future; instead, he only tried to earn easy points by following the hints without mental processing and memorization. In comparison, all of the other participants tried to retrieve their impression and made their guess before viewing hints; in very rare conditions they viewed the hints first. S8 also expressed that he had no interest in any kind of mobile games. Even though S8 did not learn as much as the other participants, he still gained 6 tones and 4 pinyin during the study.

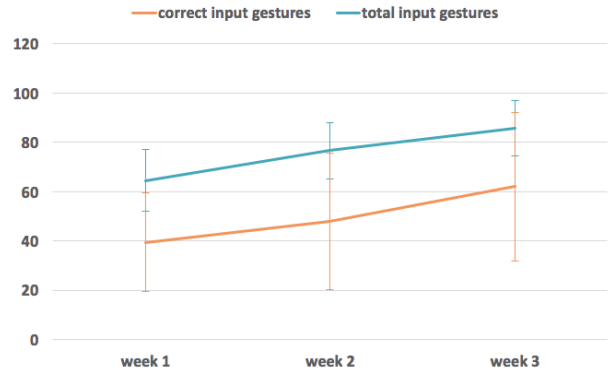

Figure 5. Average number of gestures and correct guesses in a 5-minute session over the course of the 3-week study.

Besides the improvements between pre-tests and post-tests, we also observed that participants in general improved in the number of correct guesses that they could make over time. Figure 5 shows the average number of gestures and the average number of correct guesses the user made during each 5-minute round over the course of 3-week study. There was a significant increase in the average number of gestures (64.5 vs. 85.8, t(17)=7.28, p<0.001) and the number of correct guesses (39.4 vs. 62.0, t(17)=4.37, p<0.001) the user made during each round from week 1 to week 3. These growing numbers also imply participants' mastery of the learning material over time.

Performance Comparison with Native Speakers

In this section, we investigated whether and to what extent CSL learners can achieve native level proficiency on the given vocabulary set through their practice with ToneWars. We hypothesize that learners will be positively motivated if they can achieve native-level proficiency for a collection of phrases. The comparison is based on ***PET*** and ***NLI*** defined earlier in the paper.

Figure 6 shows the visualizations of CSL learners' performance when they encountered each phrase at the first time (top), at the middle stage (middle, e.g., if one learner met a certain phrase for 6 times in total, we show the performance when she met the phrase at the 3rd time, i.e. (n+1)/2) and at the last time (bottom) during the 3 weeks. Light blue color means that the certain participant achieved native level proficiency on a certain phrase, and red color means not. The phrases are sorted based on 1) the length; 2) the difficulty (i.e. determined by the participants' performance when they met the phrases for the last time).

On average, the participants can achieve native level proficiency on 35.8 (51.9%) of phrases at the first time they met these phrases. In comparison, this number was increased to 58.2 (84.3%) at the last time they met these phrases during the 3-week study. The pairwise t-test shows that the difference is significant (t(17)=-10.64, p<0.001). We were also glad to find that there are 8 (11.6%) phrases that all participants can achieve native-level proficiency at the end of the study.

Visual Hints vs. Audio Hints

In this session we investigated the impacts of the two types of feedback (visual vs. audio) on learning gains. T-test showed a significant recall gain for tone and pinyin (post-test minus pre-test) for both visual feedback and audio feedback conditions. For tone recall (Figure 7), the gain of visual feedback was 14.3 (22.5 vs. 36.8, t(17)=6.22, p<0.001), the gain of audio feedback is 15.5 (23.6 vs. 39.1, t(17)=10.87, p<0.001). Although audio hints led to 1.2 more tones learned than visual hints, the difference is not significant (14.3 vs. 15.5, t(17)=0.61, p=0.55). For pinyin recall (Figure 7), the gain of visual hints was 5.1 (33.2 vs. 38.3, t(17)=4.73, p<0.001), the gain of audio hints was 11.4 (32.3 vs. 43.7, t(17)=7.41, p<0.001). Audio hints led to 6.3 more pinyin learned than visual hints and the difference is significant (11.4 vs. 5.1, t(17)=4.26, p=0.001).

All participants reported that they preferred audio hints to visual hints for correcting and learning tones for new characters. Based on our observation and the interviews

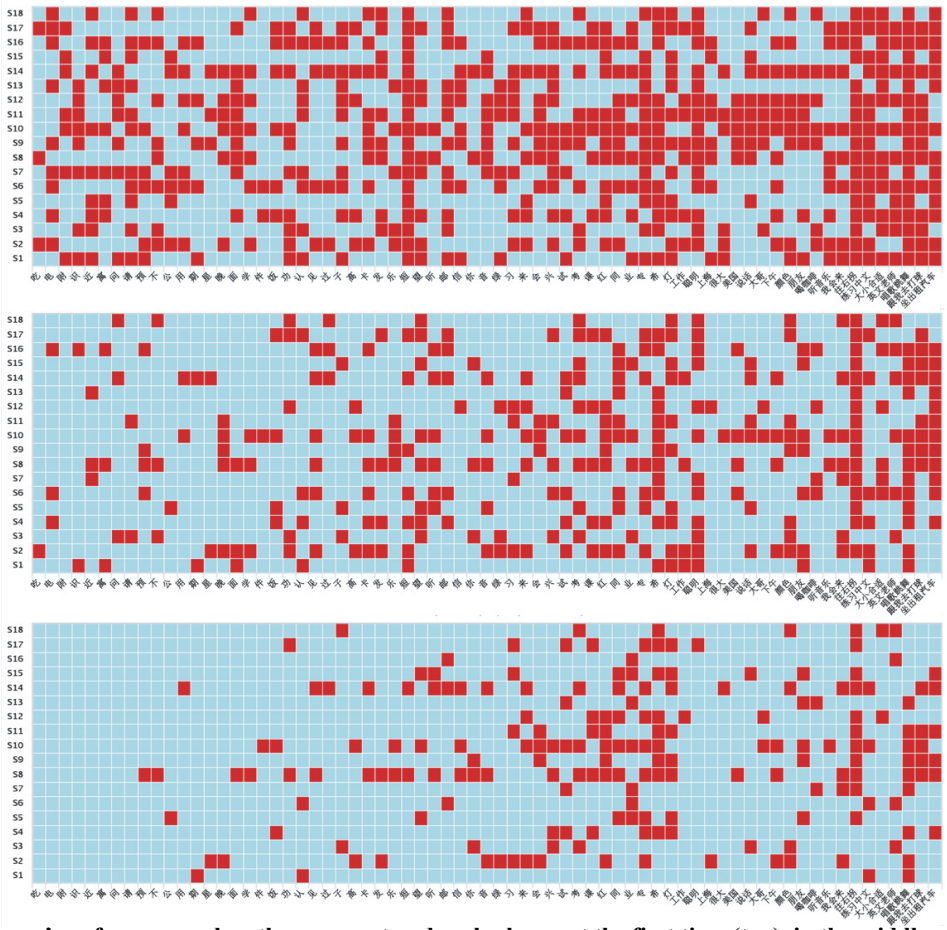

Figure 6. CSL learners' performance when they encountered each phrase at the first time (top), in the middle stage (middle), and at the last time (bottom) during the 3 weeks. Light blue color means that the certain participant achieved native-level proficiency on a certain phrase, and red color means not. The phrases are sorted based on 1) the length; 2) the difficulty (i.e. determined by participants' performance when they met the phrases at the last time).

with participants, we found that the unique benefits of audio hints over visual hints are two-fold:

- *Listening to audio hints can improve CSL learner's ability to identify the correct tone by sound.*

Unlike visual hints with which users could easily see the exact gesture that must be drawn since the hint maps simply to the correct answer, audio hints required students to *"analyze the sound and determine the tone"* (S13) which involves complex mental processing. Even though it is easy for experienced learners who are familiar with the patterns of different tones, it is quite challenging for beginners. By analyzing the logs, we found that learners' such skill improved over time. Figure 8 shows the average number of audio hint played and the average number of guesses in order to get the correct tone from an audio hint. In week 1, on average the CSL learners need to hear audio hint for 1.28 times and try 1.75 gestures in order to get the correct answer; in comparison, they need to hear the audio for 1.09 times and try 1.38 gestures in week 3. T-test shows that the difference is significant (1.28 vs. 1.09, $t(17)=-1.81$, $p < 0.05$; 1.75 vs. 1.38, $t(17)=1.95$, $p < 0.05$). This implies that their ability in recognizing tones from audio hints improved over time and the participants also noticed such improvement:

S12: *"Audio feedback gives a chance to the players to use their mind to figure out the right tone… I can do better later than at the beginning."*

S17: *"I enjoyed the audio feedback a lot more because it allowed me to hear a native speaker say the tone aloud. I thought this was beneficial because rather than it giving you the answer, it forces you to identify the tone by sound. Such skill is also very important."*

- *Audio hints can help CSL learners associate pronunciations and tones with characters, which can be treated as an effective study tool.*

Audio hints not only tell the learners about the tones, but also the pronunciations of the characters. Such association can reinforce learner's aural knowledge of the characters and was appreciated by the learners:

S4: *"I liked audio feedback more because it allowed me to associate the correct pronunciation with the character."*

This may account for the significant difference in pinyin recall gains between visual hints and audio hints—audio hints also taught learners about the pronunciation of the characters, but visual hints did not. Learners also expressed their concerns when they were exposed to visual tone marks of characters that they did not know:

S3: *"no idea even though I got the tone mark if I don't know the character."*

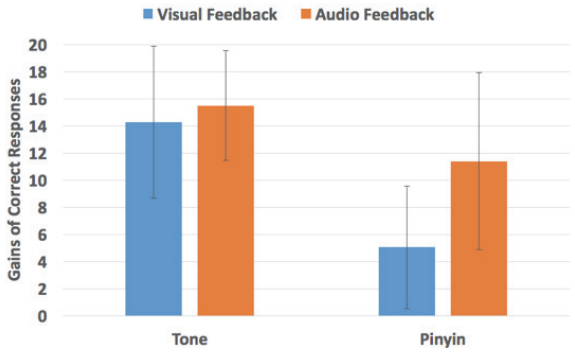

Figure 7 Tone/pinyin recall gains by feedback condition.

Some learners used the audio hint as learning materials and even read after that: users saw audio hints as closer to an effective, class-like study tool and thus took more advantage of these hints for their learning benefit.

S11: *"You can hear how that character is supposed to sound like, and how that tone sounds, and I might be able to recognize it in conversation."*

Even though all participants reported that they preferred audio hints for learning purpose, some of them mentioned that they preferred visual hints in gameplay because they were *"more straightforward and easier"* (S8), and they were *"equally helpful when you already know the pronunciation but not sure about the tone"* (S17).

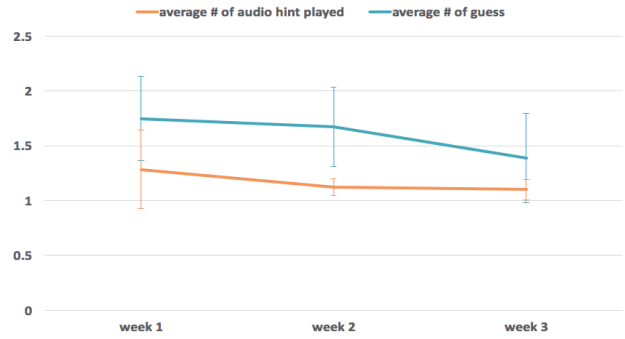

Figure 8. Average number of audio hint played and average number of guesses in order to recognize the correct tone.

Learners' Opinion towards Asynchronous Competition

In this section, we investigated whether playing with a native speaker opponent rebuilt from user logs is still engaging for CSL learners. We first analyzed the user activity logs to retrieve all attacking behaviors (Figure 9) which could be the most direct indicators of how players engaged in such a competitive game. There was a significant increase in the average number of attacks that a user made (44.4 vs. 65.5, $t(17)=2.61$, $p=0.024$) the user made in a 5-minute session over the course of the 3-week study. Compared with 26.5 attacks per 5-minute in our previous lab study [15] in which CSL learners were competing with native speakers face-to-face, the numbers of attacks may suggest that CSL learners also engaged in the asynchronous gameplay. We attribute the increase of attacks to learners' better mastery of the game mechanics in the longitudinal study (3-hour gameplay vs. 40-minute gameplay in previous lab study [15]).

Learners' opinions towards asynchronous competition vary. Five participants preferred synchronous competition—they explicitly mentioned that they would prefer playing against a real-life opponent in real time so that they can *"interact with native speakers directly"* (S4) and enjoy the *"real competition"* (S8). Four participants preferred asynchronous gameplay because they *"would not feel so embarrassed when lose"* (S11), and they *"don't like (to) compete too much"* (S3) due to personal playing style. The other nine participants did not have preference—they reported that the motivation to play with the rebuilt native speaker opponent same as playing with a real-life opponent in real time. Some of them mentioned that they were more focusing on their *"learning rather than winning the game"* (S17).

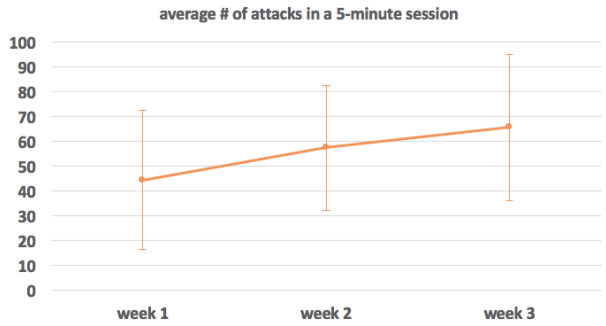

Figure 9. Average number of attacks in a play session.

Despite the variations in learners' opinions, all of them agreed that rebuilding the native speaker opponents from activity logs was the *"good enough solution"* (S8) to address the time zone problem. They also reported that it was engaging to play with the current native speaker opponent loaded in ToneWars which was "*pretty organic*" (S4).

Subjective Feedback

Overall, all participants reported positive and favorable experiences with ToneWars (Figure 10). Ratings were measured on a 5-Likert scale (1=*strongly disagree*, 5=*strongly agree*). Participants rated ToneWars' engagement as 4.61 ($\sigma = 0.61$) and would like to play the game in the future ($\mu = 4.5$, $\sigma = 0.86$).

At participants' requests, we installed ToneWars on 5 CSL learners' Android phones at the end of the study with

customized vocabularies. After one month, they reported that on average they spent 1.7 hours (min=0.5, max=3) per week playing with ToneWars, with high positiveness for the helpfulness of ToneWars. This also reported that asynchronous gameplay is engaging to motivate their sustained play and practice.

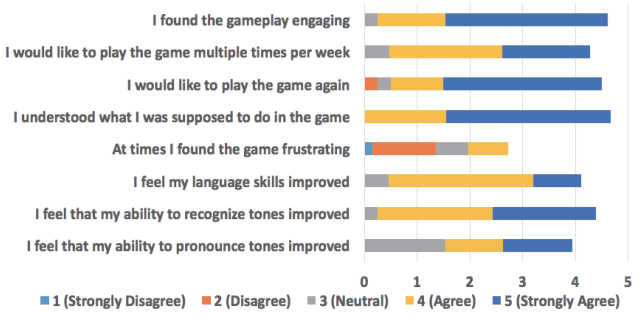

Figure 10. Subjective ratings on a 5-point Likert scale.

DESIGN IMPLICATIONS

Native Speakers in Second Language Learning

In our paper, the terminology *"interaction"* means the competition and comparison of fine-grained language skills between CSL learners and native speakers in a mobile game, rather than "verbal" or "face-to-face" interactions. Talking directly to native speakers is beneficial, however, the direct communication can cause frustration and anxiety due to learners' perception of inequality when native speakers use some language that is beyond their comprehension levels, especially for novice learners [15, 23]. The role of native speakers in ToneWars is twofold: 1) a *benchmark* for fine-grained language mastery; 2) a *motivator* for language learning—learners are positively motivated towards further language learning when they know that they can achieve native-level proficiency for a subset of the language skills. We believe the use of native speakers as both a *benchmark* for language mastery and a *motivator* for language learning is a rich direction and can lead to interesting future research.

Fine-Grained Feedback on Language Mastery

Through fine-grained modeling of native speakers' language skills, CSL learners can compete with them on phrase level tone recall tasks. Currently, they can see the real-time scores and the preview of their opponents' stacks to infer their relative performance and success. In the future, we plan to incorporate more explicit and fine-grained feedback on language mastery—once a phrase is eliminated and the PET reaches the NLI, ToneWars provides a bonus and conveys the player about their success. We believe that such fine-grained modeling and feedback on language mastery could enhance language learners' self-confidence and highly motivate them towards further learning. We also believe that this approach and the corresponding insights are generalizable to other language learning applications beyond Chinese.

Learning via Multiple Modalities

All CSL learners reported that they preferred audio hints to visual hints for correcting and learning tones of new characters. Some of them attempted to recite characters following the audio hints. The major reason is that audio hints can allow them to *"associate the correct pronunciation with the character"* (S4). Some CSL learners also suggest that ToneWars could *"show the pinyin together with tone mark"* (S3) in visual hints so that they can know the pronunciation of the character. S11 suggests that ToneWars can incorporate the *"English translation"* as well so that they can know the meaning of the character. We notice that the association between pinyin, tone, meaning, and character is very important for their character acquisition processes and each aspect will reinforce other aspects of their knowledge.

Character-Level vs. Phrase-Level Practices

Participants' preferences on character-level or phrase-level practice varied. Seven learners (most of them had < 2-year experience) reported character-level practice is easier and phrase-level practice is *"hard to memorize"* (S2). S12 mentioned that *"phrase is more frustrating if knew nothing in the phrase"*. On the other side, others expressed that they would *"like to learn in phrases"* (S15) since it can provide them more context information regarding how to use the characters in phrases. S1, with more than 3-year Mandarin learning experience and 6-month living experience in China, found that sometimes it was challenging to recognize single characters than phrases. As an example, she mentioned that she can correctly recognize and read "希望" (hope) as a phrase, but cannot recognize the single characters if they do not appear together because she lost the context. She expressed that she would like to take more *"practice with single characters"* after she noticed this problem.

CONCLUSIONS

We present principled research to make ToneWars *scalable* and *sustainable*. First, we propose a scalable approach to enable asynchronous competition and skill comparison by modeling both the interaction patterns and language skills of native speakers. We conduct fine-grained modeling of native speakers' language skills (e.g. phrase level tone recall) and use it as the goal for learners to achieve "bite-sized" native level mastery. Second, we propose a novel metric to quantify whether a CSL learner achieves native level proficiency for a specific tone at a given moment, and use such fine-grained language mastery as a sustainable motivating factor for further learning. In a longitudinal study with 18 CSL learners, we found that such asynchronous modeling can motivate learners in a sustained manner. We also observed significantly improved learning gains (e.g., average gains of 29.7 tones and 16.4 syllables). Participants achieved native level proficiency on 58.2 out of 69 phrases at the end of the study.

REFERENCES

1. Khalil Al-Mekhlafi, Xiangpei Hu, and Ziguang Zheng. An approach to context-aware mobile Chinese language learning for foreign students. *Mobile Business*, 2009. ICMB 2009. Eighth International Conference on. IEEE, 2009.
2. Jennifer S. Beaudin, Stephen S. Intille, Emmanuel Munguia Tapia, Randy Rockinson, and Margaret E. Morris. Context-sensitive microlearning of foreign language vocabulary on a mobile device. *Ambient Intelligence*. Springer Berlin Heidelberg, 2007. 55-72.
3. Kees Bot. The psycholinguistics of the Output Hypothesis. *Language Learning*, 46, 529-555, 1996.
4. Carrie J. Cai, Philip J. Guo, James R. Glass, and Robert C. Miller. Wait-Learning: Leveraging wait time for second language education. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3701-3710. ACM, 2015.
5. Nadire Cavus, and Dogan Ibrahim. m-Learning: An experiment in using SMS to support learning new English language words. *British journal of educational technology*, 40.1 (2009): 78-91.
6. David Dearman, and Khai Truong. Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012.
7. Zoltán Dörnyei, and Kata Csizér. Ten commandments for motivating language learners: Results of an empirical study. *Language teaching research* 2.3 (1998): 203-229.
8. Nathan Dummitt. *Chinese Through Tone & Color*. Hippocrene Books, 2008.
9. Darren Edge, Kai-Yin Cheng, Michael Whitney, Yao Qian, Zhijie Yan, and Frank Soong. Tip tap tones: mobile microtraining of mandarin sounds. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pp. 427-430. ACM, 2012.
10. Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A. Landay. MicroMandarin: mobile language learning in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3169-3178. ACM, 2011.
11. Darren Edge, Stephen Fitchett, Michael Whitney, and James Landay. MemReflex: adaptive flashcards for mobile microlearning. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pp. 431-440. ACM, 2012.
12. Kyle Feldscher (Sep 25th, 2015). "Obama wants 1 million Americans learning Chinese by 2020". http://www.washingtonexaminer.com/article/2572865
13. I-Ping P. Fu. *Student approaches to learning Chinese vocabulary*, doctoral dissertation, Virginia Polytechnic Institute and State University (2005).
14. Robert Godwin-Jones. Emerging technologies from memory palaces to spacing algorithms: approaches to secondlanguage vocabulary learning. *Language, Learning & Technology* 14.2 (2010): 4-11.
15. Andrew Head, Yi Xu, and Jingtao Wang. Tonewars: Connecting language learners and native speakers through collaborative mobile games. *Intelligent Tutoring Systems*. Springer International Publishing, 2014.
16. Matthew Kam, Divya Ramachandran, Varun Devanathan, Anuj Tewari, and John Canny. Localized iterative design for language learning in underdeveloped regions: the PACE framework. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 1097-1106. ACM, 2007.
17. Ayako Kawase. Second language acquisition and synchronous computer mediated communication. *Tesol & Applied Linguistics* 6.2 (2006): 1-27.
18. Constantine Kiriloff. On the auditory perception of tones in Mandarin. *Phonetica* 20.2-4 (1969): 63-67.
19. Stephen D. Krashen. *The input hypothesis: Issues and implications*. Addison-Wesley Longman Ltd, 1985.
20. Anuj Kumar, Pooja Reddy, Anuj Tewari, Rajat Agrawal, and Matthew Kam. Improving literacy in developing countries using speech recognition-supported games on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1149-1158. ACM, 2012.
21. Jonathan Leather. F0 Pattern Inference in the Perceptual Acquisition of Second Language Tone in Sound Patterns in Second Language Acquisition. *Studies on Language Acquisition (SOLA)* 5 (1987): 59-80.
22. Jonathan Leather. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. *New sounds* 90 (1990): 72-97.
23. Lina Lee. Learners' perspectives on networked collaborative interaction with native speakers of Spanish in the US. *Language Learning & Technology* 8, no. 1 (2004): 83-100.
24. Lili Liu (Jun 27th 2011). Chinese language proficiency test becoming popular in Mexico. http://news.xinhuanet.com/english2010/china/2011-06/27/c_13951048.htm
25. Michael H. Long. Native speaker/non-native speaker conversation in the second language classroom. *University of Hawai'i Working Papers in English as a Second Language 2 (1)* (1983).
26. Laura M. Morett, and Li-Yun Chang (2015). Emphasizing sound and meaning: pitch gestures

enhance Mandarin lexical tone acquisition, *Language, Cognition, and Neuroscience*, 30:3, 347-353.

27. Julie Rico, and Stephen Brewster. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 887-896. ACM, 2010.
28. Xiaonan S. Shen (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, *24*(3), 27-47.
29. Steven L. Thorne, Rebecca W. Black, and Julie M. Sykes. Second Language Use, Socialization, and Learning in Internet Interest Communities and Online Gaming, *The Modern Language Journal*, Volume 93, Issue Supplement s1, pages 802–821 (2009).
30. Feng Tian, Fei Lv, Jingtao Wang, Hongan Wang, Wencan Luo, Matthew Kam, Vidya Setlur, Guozhong Dai, and John Canny. Let's play chinese characters: mobile learning approaches via culturally inspired group games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1603-1612. ACM, 2010.
31. Yue Wang, Allard Jongman, and Joan A. Sereno (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033-1043.
32. Yue Wang, Michelle M. Spence, Allard Jongman, and Joan A. Sereno (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106(6), 3649-3658.
33. Carolyn M White. Tonal perception errors and interference from English intonation. *Journal of the Chinese Language Teachers Association* 16.2 (1981): 27-56.
34. Janet Zhiqun Xing. Teaching and Learning Chinese as a Foreign Language. *Electronic Journal of Foreign Language Teaching 5*, no. 1 (2008): 174-176.