ToneWars: Connecting Language Learners and Native Speakers through Collaborative Mobile Games

Andrew Head[1], Yi Xu[2], and Jingtao Wang[1]

[1] Computer Science & LRDC, University of Pittsburgh, PA, USA
{amh140@,jingtaow@cs}pitt.edu
[2] East Asian Languages & Literatures, University of Pittsburgh, Pittsburgh, USA
xuyi@pitt.edu

Abstract. In this paper, we present ToneWars, a collaborative mobile game for learning Chinese as a Second Language (CSL). ToneWars provides a learning experience that combines mastery learning, microlearning, and group-based interaction between CSL learners and native speakers. The game explores how unique input modalities, like touch gestures and speech recognition, can improve language acquisition tasks on mobile devices. We report the design motivations and lessons learned through the iterative design process. We believe many insights from developing ToneWars are generalizable to designing productive language learning technology. Through a 24-participant (12 CSL and native speaker pairs) user study, we found ToneWars provides learning benefits for second-language learners and engages native speakers.

Keywords: Mobile Learning, Serious Games, Crowdsourcing, Collaborative Learning.

1   Introduction

Learning a second language (L2) can be an extremely rewarding pursuit [14]. However, the process is notoriously challenging. It can take learners thousands of hours [17] to achieve intermediate fluency. Additional challenges are posed when a new language varies greatly from a learner's native tongue, as is the case for English speakers learning Chinese. At the same time, the need to improve L2 learning is just as important as ever. According to ACTFL, Chinese as Second Language (CSL) enrollments in K-12 schools in the U.S have increased 195% from 2004 to 2007 [1].

Numerous challenges arise for CSL students as they acquire listening, speaking, reading, and writing skills. One of the best-known challenges involves dealing with the large quantity of characters [1]. Meanwhile, many consider the most difficult task of CSL acquisition to be correct perception and pronunciation of tones due to the interference of the learner's native language [7], [19]. Tones in Chinese determine *meaning,* whereas tones in western languages, such as English, are used for

---

[1] According to national standard GB2312 level 1, there are 3,755 Chinese characters defined as "frequently" used in daily communication.

grammatical and expressive *inflection*. Authentic, low-risk practice with native speakers can provide opportunities to explore L2 nuances [11], [13]. However, native speakers are usually inaccessible to learners.

In response to these challenges, we present ToneWars (**Fig. 1**), a collaborative educational game to help CSL learners master the Chinese tones. ToneWars connects learners and native speakers in collaboration though gameplay. Learners practice tone recall, perception and production with kinesthetic touch input and voice-based speech input. Furthermore, learners compete with native speakers, achieving mastery [2] over the accuracy and rhythm of their own tone production. We hypothesize that if learners are able to achieve native-level proficiency for a collection of phrases, they will be positively motivated towards further language learning. ToneWars makes use of gameplay elements from popular mobile games like Tetris and Fruit Ninja to provide an engaging experience for both native speakers and learners.

Fig. 1. A CSL learner and native speaker competing through ToneWars

To our knowledge, ToneWars is the first language learning system to connect second language learners and native speakers via collaborative mobile gameplay. In this paper, we discuss the design, implementation, and evaluation of ToneWars. First, we describe our field study of tone teaching, consisting of semi-structured interviews with CSL instructors. Second, we discuss the design of ToneWars in detail, highlighting discoveries from the iterative design process. Finally, we describe our 24-subject user study and evaluate its results to determine the system's learning outcomes.

2   Related Work

2.1   Mobile Language Learning Systems

Smartphones are an ideal platform for language learning [5, 6], [10] due to their portability, connectivity, and affordability, even in developing countries [8, 9]. Researchers have built compelling learning applications that leverage the unique

capabilities of mobile devices, including recording media [10], location awareness [6], and speech recognition [9]. These applications have been built for English as a Second Language (ESL) learners [8, 9], CSL learners [5, 6], and even native speakers of Chinese in elementary schools [10].

The Multimedia Word and Drumming Strokes mobile games by Tian et al [10] enable group-based Chinese learning. These games are played by co-located, native Chinese speaking children who share a single mobile device. ToneWars differs from these games in that during its collaborations, language learners and native speakers are physically separated, communicating over network from separate devices.

Tip Tap Tones [5] by Edge et al. trains CSL learners to perceive Chinese tones through mobile games. Tip Tap Tones provides single player flashcard-style drills at a character-by-character level. ToneWars seeks to build on this work in several ways. First, in ToneWars players can use touch gestures and speech to input tones, providing practice opportunities similar to classroom exercises. Second, it enables phrase-level tone learning. Third, it connects CSL learners with native speakers in collaborative multiplayer gameplay in order to foster greater learning motivation.

2.2   Systems with Users of Multiple Languages

Duolingo [4] by von Ahn et al is a platform for crowdsourcing translations. Duolingo aims to produce high quality, low cost translations by breaking translation tasks into free, bite-sized educational exercises for learners of both the source and target languages. *MonoTrans* [12] by Hu et al uses two-way machine translation and two groups of monolingual users to achieve low cost translation through iterative collaboration. *Busuu.com* is an L2 learning community where learners can act as experts of their own native language *voluntarily*, communicating with other learners of their language through text or video chat and revising others' written exercises. ToneWars, in contrast, brings together learners with native speakers in a mobile setting, motivating L2 learners to engage in learning through competitive gameplay.

3   Field Study

To identify opportunities to include CSL tone pedagogy in designing ToneWars, we conducted semi-structured interviews with two experienced CSL instructors (*T1* and *T2*) from local universities. *T1* and *T2* each had 6 or more years of CSL teaching experience with more than 200 CSL learners at the entry and intermediate levels. We took several observations from these interviews:

F1. Tones Require Continual Practice for Clean Production. Students were exposed to tone and pronunciation practice in the first 1-2 weeks of instruction (*T1* and *T2*). After this, students' tone perception and production deteriorated after a switch in material and the continuing influence of the learner's native language (*T1* and *T2*). At later stages, dedicated practice and personalized feedback could help restore cleaner tone production students achieved in the first weeks. Challenges to tone recognition varied with learning level. According to *T1*, beginner students frequently confused *tone 2* and *tone 4,* and later they confused *tone 1* and *tone 4*. Advanced CSL learners had difficulty recognizing and pronouncing tones in long phrases (*T2*).

F2. Tone-Tracing Hand Gestures May Provide Motivation and Attention. In CSL classrooms, instructors may have students trace tones of characters with their hands as they pronounce them (*T1* and *T2*). Instructors saw several advantages to using these gestures. First, kinesthetic gestures could reinforce learners' memories of the sounds of characters [15]. Second, hand gestures may serve as visual aids to help students recognize incorrect pronunciation. Third, tone tracing might make the classroom more engaging. One instructor had students trace tones with their hands, heads, and feet to make class more fun (*T1*). However, some research suggests the "social acceptability" problem [16] may prevent CSL learners from tracing tones when practicing in public.

F3. Learners Can Benefit from Interaction with Native Speakers. After students engaged in conversation outside the classroom, the instructors believed they gained confidence they could carry into the classroom (*T1* and *T2*). Through these experiences, learners improved their tones and pronunciation, especially at later levels (*T2*). However, instructors saw limitations to these interactions. They could cause frustration: "*the native speaker [could] start with some language that is beyond the comprehension level of the student*" (*T2*). Second, the native speaker's feedback may not be helpful for the student at her current learning level: "*[T]he native speaker is not a teacher who knows what the student should start with*" (*T2*).

Our interviews confirmed the importance and challenges of tone learning for CSL learners. We have designed ToneWars to build upon the premises above to improve the quality of the educational experience. In particular, ToneWars leverages kinesthetic learning and interactions between learners and native speakers.

4   The Design of ToneWars

4.1   Gameplay

We created ToneWars through two rounds of paper prototyping and three rounds of iterative design. In ToneWars, a player[2] must quickly and accurately eliminate Chinese phrases that fall from the top of the screen. A phrase is eliminated only when players enter the correct tone for each character in the phrase sequentially.

Fig. 2. ToneWars gameplay (left to right). (a) Phrases fall and collide; (b) The player selects a phrase; (c) The player traces a tone with a touch gesture; (d) A character locks after an incorrect guess; (e) The phrase stack overflows and clears.

---

[2] A player is either a second language learner or a native speaker.
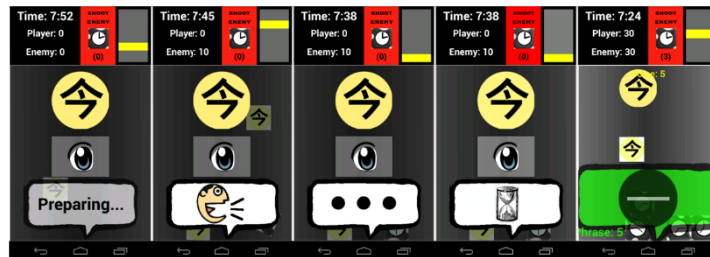
Fig. 3. Voice input for a character (left to right). (a) Recognition starts; (b) ToneWars listens for start of speech; (c) ToneWars listens to speech; (d) Processing speech; (e) Match feedback.

Tones are entered through two input modalities: (1) kinesthetic touch gestures **(Fig. 2)** and (2) voice-based speech recognition **(Fig. 3)**.

Phrases on screen collide with each other and stack up as they fall (**Fig. 2a**). The game penalizes players by halving the score when they cannot eliminate phrases fast enough to prevent phrases from filling the whole screen (**Fig. 2e**).

If the player enters the incorrect tone for a character, the phrase is locked (**Fig 2d**). To unlock it and continue to the next character, the player must first play back a hint (either visual or audio) that describes the correct tone. Then the player enters the correct tone 4 times. This process emulates language learning drill exercises.

Players earn points when they enter a tone correctly, eliminate a phrase, or view hints. Learners compete for higher scores against native speakers. Each player can fire blocks to clutter the other's screen. They earn one block for each phrase eliminated, and can save these blocks to fire all at once to cause maximal damage. Players monitor how cluttered the opponent's screen is through a preview pane in the top-right corner of the screen. When an opponent is not available, ToneWars allows a player to compete with an in-game AI. In the future they will be able to play against pre-recorded action scripts recorded from native speaker players we have collected.

4.2   Tone Input Modalities

ToneWars offers two unique modes for entering tones. The first input mode is kinesthetic touch gestures. Players use their fingers to slash the shapes of tone marks for the characters on the screen. This control scheme was inspired by the "in-the-air" gestures CSL instructors used in the classroom to engage students. We believed learners could experience the same benefits of engagement and aural attention as in class with a touch-based input that is more "socially acceptable" [16] in public.

Rather than training a general-purpose handwriting recognizer for gesture classifications, we created a decision tree based recognizer. This was for three reasons. First, ToneWars only needed to recognize four pre-defined, simple shape tone marks.[3] A heuristic-driven decision tree could easily detect the distinct shape of each mark. Second, the decision tree could be fine-tuned to match user control

---

[3] The 5th 'neutral' tone gesture, which occurs infrequently, is executed as a tap gesture.

patterns we observed without requiring large amounts of training data. Third, it could allow us to determine the exact cause of why each user's misrecognized gestures were falsely classified. Instead of receiving a cryptic rating for each classification, we could compare the user's touch gesture paths to the hard rules defined in the decision tree.

The second tone input mode is speech (**Fig. 3**). Players enter tones by pronouncing characters into the phone's microphone. If the player does not know a character but wants to guess its tone, she can pronounce a placeholder character with the expected tone. This control method had two advantages. First, oral production helps learners move from declarative knowledge of a language (recognizing words) to productive knowledge (using the words correctly) [3]. Second, by requiring students to guess the sound of unknown characters, they may be able to develop a deeper knowledge of the relationship of a character's written radicals[4] to its pronunciation.

Our tone recognition engine is built on top of Google's speech recognition service for Mandarin, a state-of-the-art speech recognition engine. A mobile device must have Internet access to use all features of the recognition service. In the future, a customized on-device Mandarin tone recognizer could be developed when the network is not available, like the English recognizer by Kumar et al [9].

4.3   Implementation

ToneWars was written in Java for Android 4.0. We used the AndEngine (http://www.andengine.org/) library to speed up game programming. We also used the Box2D physics engine to implement the falling and collision effects in ToneWars. Excluding third party libraries, ToneWars has a total of 11,150 lines of code in Java.

5   User Study

We conducted a 24-subject user study (12 CSL and native speaker pairs) to understand the performance and usability of ToneWars. Before the study, we worked with a CSL instructor to select 25 phrases with a total of 25 unfamiliar characters. We sorted the phrases into 5 groups of equal difficulty. Each phrase group had 5 characters that would be unfamiliar to CSL learners with less than 1 year of experience with *Integrated Chinese*, a popular CSL textbook in North America. No phrase had more than two unfamiliar characters. Although some Chinese characters have tones that can vary based on context, no such characters were chosen for these groups.

The user study consisted of three steps:

1) *Pre-test*. Participants completed a quiz in which they determined tones of 35 characters. The set contained all 25 unfamiliar characters as well as 10 familiar characters from the phrase groups. If participants did not know the tone, they were told it was okay to leave the space for response blank.

---

[4] According to Wang [18], although Chinese writing is logographic, 77% of characters in modern Chinese have radicals that suggest their pronunciation.

2) *Gameplay sessions*. Participants were grouped into learner-native speaker pairs. They played 5 rounds of ToneWars. In the first 4 rounds, users controlled the game with touch-based tone tracing. The rounds were structured according to a 2-by-2 within-subject design based on hint feedback mechanisms (*visual* vs. *audio*) and collaboration mode (*single-player with AI* vs. *CSL learner with native speaker*). The order of conditions was counter-balanced with a Latin-square pattern across pairs. In the 5th round, participants used speech to input tones. Each round presented users with phrases from 1 of 5 phrase groups selected above. By measuring subjects' improvement in tone recall for each of these phrase groups, we sought to measure the impact of the above design conditions on the learning experience.

Participants were seated in adjacent chairs facing each other. For rounds that utilized voice control and audio feedback, players were instructed to wear headphones to minimize audio interference. Each round took 8 minutes to complete.

3) *Post-test*. Participants completed a quiz identical to the pre-test after finishing the game sessions. All CSL learners and native speakers completed both tests.

We recruited 24 participants (12 CSL learner and native speaker pairs) from two local universities. Among the learners (7 female, median age 19), a majority had less than one year of formal CSL learning experience. The native speakers (7 female, median age 22.5) were mostly undergraduate or graduate students originally from China.

Participants used one of two Google Galaxy Nexus smartphones. The device has a 4.65-inch, 720 x 1280 pixel display, 1.2-GHz dual core ARM Cortex-A9 processor, and runs Android OS. Devices were connected to the Internet through 802.11g Wi-Fi.

6   Results and Discussion

We observed a difference in initial tone identification ability for learners and native speakers. During the pre-test, CSL learners correctly recognized 11.6 / 35 characters (*min* = 1, max = 25, σ = 7.9, *accuracy* = 33%[5]). In comparison, 11 / 12 native speakers scored 35 / 35, and one native speaker scored 19 / 35 (μ = 33.7, σ = 4.6, *accuracy* = 96%). The difference is statistically significant ($t = 8.38, p < 10^{-6}$).

After 40 minutes of gameplay, learners could recognize tones of 17.8 characters (*min* = 5, max = 34, σ = 10.4, *accuracy* = 51%). The average recall gain for learners was 6.2 characters (*min* = 1, max = 13, σ = 3.5). A pair-wise *t*-test showed this improvement is statistically significant ($F_{1,11} = 6.13, p < 10^{-4}$). The native speaker that did not score perfectly in the pre-test properly identified 8 more tones in the post-test.

For all experimental conditions, repeated ANOVA showed significant gain for CSL learners in recall of the tones of unfamiliar characters **(Fig. 4).** The gains were: visual feedback, 2.7 tones ($t = 6.4, p < 10^{-4}$); audio feedback, 2.35 tones ($t = 4.4, p < 0.005$); competition with AI, 2.4 tones ($t = 5.7, p < 10^{-4}$); learner vs. native speaker, 2.7 tones ($t = 4.5, p < 0.005$); voice control, 1.1 tones ($t = 3.0, p < 0.05$). The difference in improvement for hint feedback types (audio vs. visual) was not significant ($F_{1,11} = 0.54, p = 0.60$). Although learners recalled 0.3 more tones in rounds against

---

[5] With 5 possible tones, the chance of randomly guessing a character's tone is 20%.
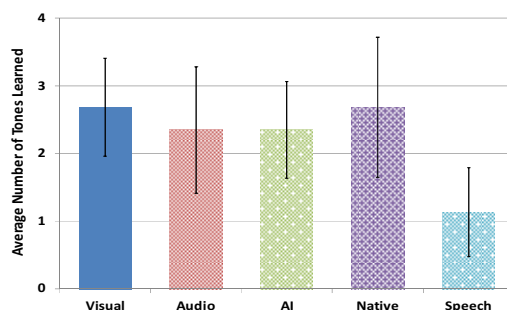
Fig. 4. Average numbers of tones learned by condition. Error bar shows one standard deviation

native speakers, this difference was not significant ($F_{1,11}$ = 0.47, p = 0.65). We attribute the lack of significance to small sample size (12 CSL learners) and the large difference in learners' initial ability (ranging from 1 to 25 out of 35 correct on the pre-test). Because of this initial variation, our current results could show a gain in recall for each condition, but could not uncover the relative strength of any one condition. We plan to address this in the near future by running classroom-level, multi-week deployments.

Although nearly all learners had lower initial tone identification accuracy than the native speakers they were paired with, 8 / 24 (33%) rounds between learners and speakers were won by learners. We attribute this to certain learners' ability to become comfortable with the control mechanics, quickly master unknown or forgotten tones, and to develop successful attacking strategies.

Participants' qualitative feedback on ToneWars was highly favorable. Ratings of user perceptions were measured on a 5-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). CSL learners unanimously rated ToneWars' ease of use at 5, and native speakers rated it 4.75 (σ = 0.45). Learners rated ToneWars' engagement as 4.5 (σ = 0.52) and native speakers rated it 4.25 (σ = 0.75). Both native speakers (μ = 4.17, σ = 0.83) and CSL learners (μ = 4.25, σ = 0.75) enjoyed playing against a real-life partner. Compared to native speakers (μ = 3.41, σ = 1.24), learners (μ = 4.33, σ = 0.98) indicated a stronger interest to play ToneWars in their spare time.

In written comments, players were positive about ToneWars. One participant reported, "*I would love to play a game like this to help my pronunciation and tonal recognition. I am always looking for new ways to learn Chinese.*" Others enjoyed the competition. One told us that when they played against a native speaker, "*it was more competitive than playing against the computer and for me, points have the main motivation for me to focus on the game.*" As one learner expressed, the social aspect of competition could be appealing: "*Having a real life opponent is always, in my opinion, fun because of being able to make fun conversation while playing is a plus.*"

Native speakers hoped that the challenge of ToneWars could be increased by introducing more phrases per round. Several participants reported moments of frustration with input control. In some cases, tone gestures or spoken tones were misrecognized due to cases we had not foreseen (left-handedness for touch) or user skill level (spoken tones of learners with insufficient previous tone training were often

misclassified). During speech input rounds, characters spoken aloud may have interfered with the device of a player's co-located participant. In the future, we plan to improve the maturity of tone tracing and speech recognition to address these cases. We also hope to explore whether learner motivation against native speakers comes from perceived competition or from a change in game dynamics during these rounds.

7   Conclusions and Future Work

ToneWars is a group-based mobile game for CSL learners to master Mandarin tones through collaboration with native speakers. The design of ToneWars, including tone-tracing gestures, speech input, and collaboration with native speakers, is inspired by effective second-language pedagogy. Our design discoveries can be applied for researchers building mobile learning technology to aid L2 acquisition beyond just Chinese. In a 24-subject study, we confirmed ToneWars' usability and efficacy. We observed a 6.2-tone average gain in short term recall for second language learners who played around 40 minutes of ToneWars.

We plan to continue our work with ToneWars in multiple ways. We will address usability problems identified above, improving speech recognition accuracy and incorporating algorithms to support performance-based adaptive learning. We hope to extend the potential of ToneWars by adding material that complements the regular CSL curriculum and by exploring its design principles for languages other than Chinese. While our current lab study shows promising results for improving learners' short-term recall, we hope to evaluate ToneWars' feasibility and educational benefits in larger scale, longitudinal deployments.

References

1. ACTFL: Foreign Language Enrollments in K–12 Public Schools: Are Students Prepared for a Global Society? Technical Report, American Council On The Teaching of Foreign Languages (2011)
2. Bloom, B.S.: Mastery learning. In: Block, J.H. (ed.) Mastery learning: Theory and practice, pp. 47–63. Holt, Rinehart and Winston, New York (1971)
3. De Bot, K.: The psycholinguistics of the Output Hypothesis. Language Learning 46, 529–555 (1996)
4. Duolingo, http://www.duolingo.com
5. Edge, D., Cheng, K., Whitney, M., Qian, Y., Yan, Z., Soong, F.: Tip Tap Tones: Mobile Microtraining of Mandarin Sounds. In: ACM MobileHCI 2012, pp. 427–430. ACM, New York (2012)
6. Edge, D., Searle, E., Chiu, K., Zhao, J., Landay, J.: MicroMandarin: Mobile Language Learning in Context. In: ACM Conference on Human Factors in Computing Systems, pp. 3169–3178. ACM, New York (2011)
7. Fu, I.-P.: Student approaches to learning Chinese vocabulary. Ph.D. thesis, Virginia Polytechnic Institute and State University (2005)

8. Kam, M., Ramachandran, D., Devanathan, V., Tewari, A., Canny, J.: Localized iterative design for language learning in underdeveloped regions: The PACE framework. In: ACM Conference on Human Factors in Computing Systems, pp. 1097–1106. ACM, New York (2007)

9. Kumar, A., Reddy, P., Tewari, A., Agrawal, R., Kam, M.: Improving Literacy in Developing Countries Using Speech Recognition-Supported Games on Mobile Devices. In: ACM Conference on Human Factors in Computing Systems, pp. 1149–1158. ACM, New York (2012)

10. Tian, F., Lv, F., Wang, J., Wang, H., Luo, W., Kam, M., Setlur, V., Dai, G., Canny, J.: Let's Play Chinese Characters – Mobile Learning Approaches via Culturally Inspired Group Games. In: ACM Conference on Human Factors in Computing Systems, pp. 1603–1612. ACM, New York (2010)

11. Thorne, S., Black, R., Sykes, J.: Second Language Use, Socialization, and Learning in Internet Interest Communities and Online Gaming. The Modern Language Journal 93(s1), 802–821 (2009)

12. Hu, C., Bederson, B., Resnik, P.: Translation by Iterative Collaboration between Monolingual Users. In: Graphics Interface, pp. 39–46 (2010)

13. Lee, L.: Learners' Perspectives on Networked Collaborative Interaction with Native Speakers of Spanish in the US. Language Learning & Technology 8(1), 83–100 (2004)

14. McLaughlin, B.: Fostering second language development in young children: Principles and practices. NCRCDSLL Educational Practice Reports (1995)

15. Morett, L.M., Gibbs, R.W., MacWhinney, B.: The Role of Gesture in L2 Learning: Communication, Acquisition, & Retention. In: 34th Annual Conference of the Cognitive Science Society, pp. 773–778. Cognitive Science Society, Austin (2012)

16. Rico, J., Brewster, S.A.: Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability. In: ACM Conference on Human Factors in Computing Systems, pp. 887–896. ACM, New York (2010)

17. Rivera, G., Matsuzawa, C.: Multiple-language program assessment: Learners' perspectives on first and second-year college second language programs and their implications for program improvement. Foreign Language Annals 40 (2007)

18. Wang, N.: The Structure and Meaning of Chinese Character and Word. In: Research on Chinese Cognition. Shandong Education Press (1997) (in Chinese)

19. Xing, J.: Teaching and Learning Chinese as a Foreign Language. Hong Kong University Press (2006)