SpatialEase: Learning Language through Body Motion

Darren Edge[1]
[1] Microsoft Research Asia
Beijing, China
darren.edge@microsoft.com

Kai-Yin Cheng[1,2]
[2] National Taiwan University
Taipei, Taiwan
keynes.z@gmail.com

Michael Whitney[1,3]
[3] University of North Carolina Charlotte
Charlotte, NC, USA
mwhitne6@uncc.edu

ABSTRACT
Games that engage both mind and body by targeting users' kinesthetic intelligence have the potential to transform the activity of learning across a wide variety of domains. To investigate this potential in the context of second language learning, we have developed SpatialEase: a Kinect game for the body-based learning of language that is grounded in space and motion. In this game, learners respond to audio commands in the second language by moving their bodies in space, while a game mechanic based on distributed cued-recall supports learning over time. Our comparison of SpatialEase with the popular Rosetta Stone software for learner of Mandarin Chinese showed similar learning gains over a single session and generated several key implications for the future design of mixed-modality learning systems.

Author Keywords
Embodied gaming; Kinesthetic learning; Language learning

ACM Classification Keywords
H5.2 Information interfaces and presentation: User Interfaces

INTRODUCTION & RELATED WORK
Language, thought, and action are inextricably linked. When we learn our first language, we do so by mapping words to abstract concepts that are grounded in our bodily experiences of the world [10]. Total Physical Response (TPR) is an approach to second language learning that aims to replicate such native language development though the co-production of spoken commands and bodily actions [1]. Our aim is to support TPR-style learning in the absence of a human instructor, using embodied interaction [3] to help learners forge associations between new second language expressions and the existing *kinesthetic image schemas* they already use to plan and perform physical actions [10].

To investigate the potential for kinesthetic language learning, we have therefore developed the "SpatialEase" game for Microsoft Kinect (Figure 1). In the game, learners respond to commands spoken by the system (e.g., a translation of "step forward then raise your right hand") with meaningful bodily actions that render learning visible.

CHI 2013, April 27–May 2, 2013, Paris, France.

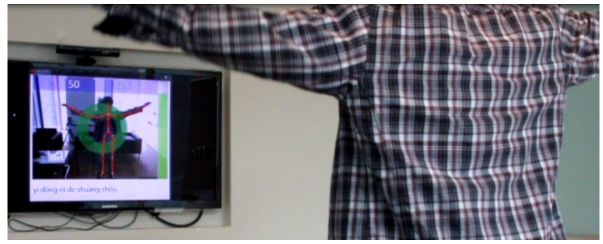

Figure 1. SpatialEase Game for Language Learning

Our approach follows that of the Mathematical Imagery Trainer [7] by leveraging kinesthetic learning, but is independently motivated by the observation that gestures and speech express the same or related meanings at the same time [8]. In Second Language Acquisition (SLA) research, Growth Point Theory [12] hypothesizes that speech and gesture interact and influence one another throughout the planning and speaking of utterances, with gestures helping speakers to internalize the abstract via the concrete [6]. The related Information Packaging Hypothesis [9] holds that gestures play a role in conceptualizing information for speaking, and that the more difficult the information is to conceptualize, the more speakers will gesture. Engaging learners' kinesthetic intelligence has also been shown to improve performance for children learning mathematics, adults learning about science and medicine, and for French children learning English vocabulary [6]. In this latter study, findings align with the results from Cognitive Psychology that enacted actions are recalled better than action phrases without enactment [4], and that self-enactment is better than observing enactment by others [5]. By creating a kinesthetic game that leverages these findings, we aim to help learners acquire a second language through constructions grounded in space and motion.

We first outline the design of our SpatialEase game before presenting the results of a study comparing it to the popular Rosetta Stone software, which acted as a benchmark for primarily visual learning. We found the two approaches to be comparable over a single session, with mixed user preferences. From our analysis we contribute three implications for the design of kinesthetic learning games: target relationship recognition over concept recall for engaging game mechanics; adapt game content and complexity based on models of learner memory; and mix kinesthetic learning with visual and auditory modalities to create diverse challenges for all learners.

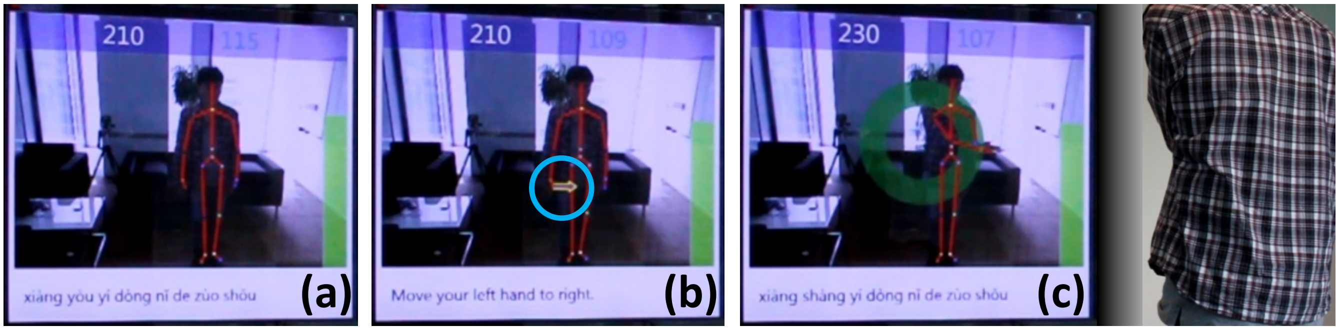

Figure 2. Interactive Flow: (a) SpatialEase issues a command in Chinese; (b) if the user doesn't move or moves incorrectly, arrows are drawn on the video indicating the corresponding action; (c) the user performs the correct action and hear the next command.

DESIGN OF A GAME MECHANIC

The Kinect SDK for Windows[1] has already been used to support learning in physical domains (e.g., sign language[2]). For kinesthetic learning of a second language, we drew inspiration from the Rosetta Stone application[3] in which the learner maps second language sentences to minimally different scenarios, e.g., translations of *"This {boy | girl} is {eating | drinking}"*. As the learner maps the text and sound of these constructions to the corresponding images, they infer the meanings of words and the relationships between them. Our kinesthetic equivalent is to use fast switching of body poses over time, encouraging similar inference across minimally different language constructions.

We were also inspired by the traditional game in which commands should only be followed if accompanied by the prefix "Simon says" (e.g., raise your left hand in response to "Simon says, 'raise your left hand'", but remain still on "Shake your right foot" because "Simon didn't say"). Our adaptation is for learners to only respond to grammatical commands. This focus on sentence form encourages the same beneficial shift from semantic to syntactic processing as when learners progress from listening to speaking.

THE SPATIALEASE KINECT GAME

We developed our SpatialEase game using the Kinect SDK for Windows. The core language of the game is based on phrases used to command bodily movement combined with the language of *kinesthetic image schemas* [10] – recurring patterns of spatial orientation, bodily movement, and object manipulation that give structure and meaning to our experience of reality (e.g., LEFT–RIGHT, UP–DOWN, FORWARDS–BACKWARDS). In our prototype for learners of Mandarin Chinese, successive game levels introduce increasingly complex constructions from core concepts. The first two levels are based on the following patterns:

"Move your {left | right | pair} {hand | foot | head | body}"

"Towards {up | down | left | right | front | back} move your {left | right} {hand | foot | body}"

[1] http://www.microsoft.com/en-us/kinectforwindows/
[2] http://cats.gatech.edu/content/copycat-and-kinect/
[3] http://www.rosettastone.com/

The game maintains a queue of language on which to test the learner. This is fed by a set of core concepts (e.g., left hand, right hand) that are so far *untested*. Each level begins with the system randomly selecting one of these and forming a command, such as the translation of "Move your left hand". This command is synthesized as speech and accompanied by a subtitle (Figure 2). If the learner moves their body accordingly, the underlying concept is added to the *correct* set and the next *untested* concept is selected.

If the learner fails to respond, the command is repeated. If they fail to respond a second time, or if their attempt was incorrect, then the concept is added to the *incorrect* set and the system indicates the correct action by drawing arrows on the live video stream of the learner's body (Figure 2b, highlighted). To provide the learner with sufficient practice in support and demonstration of learning, three further tests of that command are then scheduled after one, two, and three intervening tests respectively. These "spacing" tests are based on concepts drawn from the *untested* set if any remain, or else from the *correct* set for extra reinforcement.

Once all concepts are in the *correct* set, the "Simon says" phase begins with three concepts selected for grammar testing. Each concept is presented in either a grammatical construction (with chance 50%), or in one of four ungrammatical constructions. Incorrect responses restart the phase, while three successive correct responses complete it.

At all times, the learner faces time pressure to respond: each command is followed by only a ten second window in which to act, and each level is allocated only three minutes for completion. Completing a level within the time allocation resets the timer to three minutes and advances the learner to the next level. This creates a natural difficulty gradient as well as repetition of key concepts across games – such *distributed cued-recall* is a fundamental learning technique [2]. Learners can also track their performance across games using our simple scoring mechanism, which awards $10 \times N$ points per correct response on Level N.

By integrating learning into the game play mechanics, rather than simply extending them, we follow the "coherent, experimentally verified rule" of successful educational game design as presented by Linehan et al. [11].

USER STUDY

We conducted a user study in which participants used both SpatialEase and Rosetta Stone to learn Mandarin vocabulary and grammar of roughly equivalent complexity[4]. Since both systems employ the pedagogical approach of supporting learning through inference (rather than through direct instruction, e.g., with textbooks), this allowed us to compare the predominantly *visual* Rosetta Stone against the characteristically *kinesthetic* SpatialEase.

Given that there is "virtually no evidence" linking learning style preferences to the effectiveness of learning with different styles (e.g., words versus pictures versus speech) [13], we did not attempt to demonstrate any relationship between preferences expressed prior to system use and final performance. Nor did we attempt to establish the superiority of one modality over another. Rather, our goal was to understand whether the kind of kinesthetic learning embodied in the SpatialEase system has a potential role in future mixed-modality systems for learners of all styles.

Participants & Design

We recruited 8 novice learners of Mandarin Chinese (2 female, mean age 26) from the international staff and visitors to our Chinese lab. Each user study lasted approximately two hours and followed the structure: pre-test, method 1 learning, method 1 post-test, method 2 learning, method 2 post-test, survey. Method order was counterbalanced across participants and learning with each method was limited to 20 minutes.

The test contained four sections covering vocabulary and grammaticality questions for the language taught in both Rosetta Stone and SpatialEase. Both methods introduced 19 Chinese vocabulary items, and in our test participants listened to synthesized speech of each before typing the English translations. For the grammaticality questions, 20 constructions were generated for each method consisting of 10 that were grammatically correct and 10 that were grammatically incorrect. Participants listened to each of these before indicating that they perceived the grammar to be *correct* or *incorrect*, or that they *don't know*.

Results

Learners did not typically know any of the SpatialEase vocabulary before playing the game, but most knew at least one word from Rosetta Stone. Given the varying Chinese language levels of participants, we also observed large differences between individuals. Overall, we found the single-session learning gains from our SpatialEase game to be comparable with those from using Rosetta Stone, with average vocabulary improvements of 6.4 items for SpatialEase and 5.6 for Rosetta Stone, and with grammar improvements of 7.6 and 8.5 respectively (see Table 1).

[4] Rosetta Stone sentences (from level 1, lesson 1) were of the form *{one|this|these} {man|woman|boy|girl} {is|are} {eating|drinking|running|swimming|cooking|reading|writing}*.

User ID	Vocabulary (out of 19)				Grammar (out of 20)			
	Rosetta Stone		SpatialEase		Rosetta Stone		SpatialEase	
	pre-test	post-test	pre-test	post-test	pre-test	post-test	pre-test	post-test
P1	2	12	0	13	7	19	6	16
P2	10	5	0	2	5	11	5	9
P3	2	10	0	3	0	13	0	5
P4	2	2	0	9	10	8	10	8
P5	0	12	0	2	0	11	0	16
P6	11	15	1	16	11	18	10	14
P7	1	10	0	5	0	4	0	14
P8	1	8	0	2	0	17	0	10
mean (*sd*)	3.6 (4.3)	9.3 (4.1)	0.1 (0.4)	6.5 (5.5)	4.1 (4.8)	12.6 (5.2)	3.9 (4.5)	11.5 (4.1)
delta (*sd*)	5.6 (5.7)		6.4 (5.2)		8.5 (6.0)		7.6 (6.0)	

Table 1. Results of study comparing Rosetta Stone (RS) and SpatialEase (SE). Pre/post-tests measure vocabulary and correct grammar recognition after 20 minutes per method.

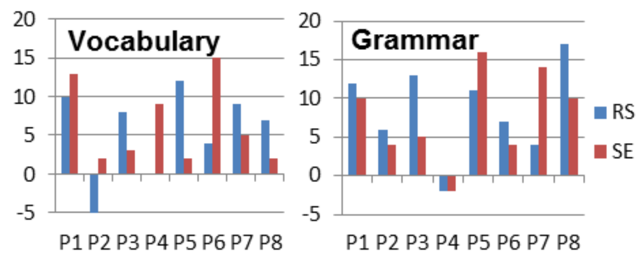

Figure 3. Learner improvements from pre-test to post-test

Figure 3 highlights the large individual differences in learner improvements from pre-test to post-test, for both vocabulary recall and grammar recognition. The same divergence was also seen in subjective preferences – three for SpatialEase, two for Rosetta Stone, and three undecided.

Discussion

While grammar improvements were not affected by method order, for vocabulary the first method resulted in roughly twice the improvement of the second method, for both method orders. This suggests a fatigue effect for vocabulary recall that does not exist for grammar recognition. One interpretation is that since recognition is a sub-process of recall, recognition-based testing is less demanding and thus has a lower threshold for success in the face of fatigue. Another interpretation is that since learners consciously attend to novel vocabulary but unconsciously acquire grammatical relations through induction over experiences, it is this conscious, directed effort that leads to fatigue. In both cases, the implication for the design of educational games is to encourage the fast and accurate recognition of correct relationships between elements. Such game mechanics have applicability beyond the grammatically of syntactic relationships between lexical language items, and could be employed in the learning of relationships in general (e.g., combinations, influences, causes, reactions).

The cases of P2 and P4 are also interesting. After using Rosetta Stone, P2 apparently forgot some vocabulary and

P4 learned no new vocabulary, while P4 became worse at differentiating the grammatical from the ungrammatical after using both methods. These results point to the effects of cognitive overloading and cognitive interference from attempting to pack so much learning into a short time period. Again, individual differences could play a role in determining what constitutes too much attempted learning, but the implication for design is to track performance in real time and stop play when the cost to learning quality outweighs the benefits of practice quantity. Such decisions about what to learn and when to rest should connect game content to learning curricula (e.g., from standard textbooks and personalized flashcards) and be informed by the extensive literature on distributed learning practice [2].

In subjective feedback, one participant highlighted a commonly expressed trade-off: "*The pictures work better in terms of learning as I have a reference to work with and other images to compare. However, the body-based is more engaging as I can visualize myself in the learning and anticipate the results of my actions*." [P5]. Kinesthetic learning appears to have the advantage in terms of engagement: "*I was getting very tired and bored with Rosetta, but providing engaging experiences seems like the best approach – one where you don't feel so much like you are learning but achieving*" [P6]. Such engagement could also help develop executive motivation for learning in general: "*In SpatialEase, the system could direct me completely in Chinese. This made me already feel more comfortable with the Chinese language and more willing to continue to learn with the Rosetta approach*". Our third implication for design is therefore to share learning material and track learner actions across a range of different systems that combine learning media and modalities in ways appropriate to different goals, devices, and contexts.

Limitations & Future Work

The small sample size and timescale of this initial study prevents us from drawing firm conclusions on learning outcomes, and the large individual differences suggest that a more refined study design is necessary to understand the relative contributions of different modalities to learner engagement in the moment. A larger sample size and longitudinal study design would also be necessary to examine the important question of how kinesthetic learning can transform learner motivation over the long term.

Future design work should explore how the SpatialEase game mechanic can be extended to incorporate more abstract concepts, more complex linguistic structures, and work in concert with other tools like Rosetta Stone. Expanding on the current design, the language of time and sequence could be added to stretch the learner's real-time processing capability (e.g., "Before X, do Y"). The vocabulary could also be extended by virtual objects and actions (e.g., "Put the red ball in the blue bucket") that connect to vocabulary sources like flashcards. Finally, abstract concepts (such as "love") could be incorporated in several forms: on-screen word targets to be selected in the correct order; grammatical constructions to be distinguished from ungrammatical; or the correct visual representation to be selected for a given phrase spoken or shown. This latter mechanic matches that of Rosetta Stone, and demonstrates a possible point of convergence for the two methods.

CONCLUSION

Whereas applications of Kinect technology have helped gamers *learn to move* in activities such as dance and yoga, our SpatialEase game helps people *move to learn* in ways that leverage their kinesthetic intelligence.

Through our user study, we have developed three implications for the design of kinesthetic learning games in general: target relationship recognition over concept recall for engaging game mechanics; adapt game content and complexity based on models of learner memory; and mix kinesthetic learning with visual and auditory modalities to create diverse challenges for all learners.

ACKNOWLEDGEMENTS

We thank the participants in our studies and Enio Ohmaye for the reference to the TPR method of language learning.

REFERENCES

1. Asher, J.J. (1969). The Total Physical Response Approach to Second Language Learning. Modern Language Journal, 53(1)
2. Cepeda, N.J., Pashler, H., Vul, E. & Wixted, J.T. (2006). Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis. Psychological Bulletin, 132(5)
3. Dourish, P. (2004). Where the Action Is: The Foundations of Embodied Interaction. Cambridge, MA: MIT Press
4. Engelkamp, J. & Cohen, R.L. (1991). Current issues in memory of action events. Psych. Research, 53, 175– 182
5. Frick-Horbury, D. (2002). The use of hand gestures as self-generated cues for recall of verbally associated targets. American Journal of Psychology, 115(1), 1–20
6. Gullberg, M. (2008). Gestures and Second Language Acquisition. In P. Robinson & N.C. Ellis (Eds.) Handbook of Cognitive Linguistics and Second Language Acquisition. New York: Routledge
7. Howison, M., Trninic, D., Reinholz, D. & Abrahamson, D. (2011). The mathematical imagery trainer: from embodied interaction to conceptual learning. CHI 2011, 1989–1998
8. Kendon, A. (1972). Some relationships between body motion and speech. In A.W. Siegman & B. Pope (Eds.), Studies in dyadic communication, 177–210, New York: Permagon
9. Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), Language and gesture, 162–185, Cambridge, UK: Cambridge University Press
10. Lakoff, G. (1987). Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. Chicago: University of Chicago Press
11. Linehan, C., Kirman, B., Lawson, S. & Chan, G. (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. CHI 2011, 1979–1988
12. McNeill, D. (2005). Gesture and Thought. Chicago: University of Chicago Press
13. Pashler, H., McDaniel, M., Rohrer, D. & Bjork, R. Learning Styles: Concepts and Evidence. Psychological Science in the Public Interest, December 2008, 9(3), 105–119