How Well do Visual Verbs Work in Daily Communication for Young and Old Adults?

Xiaojuan Ma
Computer Science, Princeton University
35 Olden St., Princeton, NJ 08540
xm@cs.princeton.edu

Perry R. Cook
Computer Science (and music), Princeton University
35 Olden St., Princeton, NJ 08540
prc@cs.princeton.edu

ABSTRACT
In this paper we study how verbs are visually conveyed in daily communication contexts for both young and old adults. Four visual modes are compared: a single static image, a panel of four static images, an animation, and a video clip. The results reveal age effects, as well as performance differences introduced by lexical verb properties and visual cues. We also suggest guidelines for visual verb creation.

Author Keywords
Verb visualization, visual communication. Age effects

ACM Classification Keywords
H5.4. Information interfaces and presentation (e.g., HCI): Hypertext/Hypermedia.

INTRODUCTION

Karen and Dongxia (who just arrived in the U.S. and speaks very little English) are neighbors in senior citizen housing. They wish to teach each other English and Chinese, but cannot even engage in daily conversations. Multilingual communication is more common with the rise of globalization. Visual languages, which convey concepts and information using photos, signs, and other graphic designs, have greatly increased especially with the spread of World Wide Web. As a supplement/extension to verbal languages, visual languages can assist both young and old people, those wanting to overcome language barriers, those learning a new language, and those with language disabilities.

In order to create visual languages for real-world settings, a designer must explore visual representations that effectively express concepts for people in all age groups. Verbs, a lexical category indicating the presence of a state, existence or operation of an action, are an indispensable part of English speech [10]. The accuracy in deciphering the visual expressions for verbs determines the quality of the delivery of the entire sentence, so the creation of visual verb representations is a crucial issue.

Currently, there are a great variety of visual representations available for conveying verbs, including realistic images, stylized icons and animations, and videos. However, little research has provided helpful instructions on creating or selecting proper forms to visualize verbs. In this paper, we present a study comparing the performance of four different visual modes (a single static image, a panel of four static images, an animation, and a video clip) in communicating commonly used verbs in sentences for both young (20 - 39) and elderly adults (55+). We reveal the influence of context and various visual/lexical factors, with design suggestions, for users across a wide age span considering aging effects.

CHI 2009, April 4–9, 2009, Boston, MA, USA.

BACKGROUND WORK

Communication Pattern and Aging Effect

Conversation structure (greeting, small talk, information sharing, and farewell statement) is similar for young and elderly adults. However, as people age, many aspects in their communication pattern are affected. Elderly people have longer word-recall time, deficit in noun naming [4], and declined verb retrieval ability regardless of other demographic differences [11]. Additionally, elderly people spend less time in general small talk but in story-telling to pass down traditions and history, and build social bonds with peers [5]. Popular categories of utterance also vary, due to change in social roles, life experience, and the setting of living and interaction [13]. For example, elderly people talk more about education (4.5%) and less about household routines (6%) compared to young people (0.55% and 11%). This suggests that visual language vocabularies should be constructed based on user interests and usage, and visual representation designs should consider age-related effects.

Visual Representations for Illustrating Verbs

Compared to nouns (people, places, things), verbs are more challenging to visualize. Most online visual dictionaries offer few verb categories, since verbs are used to indicate ongoing action or an existing state or condition, and static forms might fail to portray time. Hence, dynamic modes such as animations and videos are introduced into visual language. Current research on visual stimuli for verbs, including line drawings [12], photos [6], animations [15], and videos [3], emphasized actions and movements having to do with postures, gestures, and observable manipulation. This small fraction of verbs cannot satisfy the needs of

Figure 1. Four visual modes for "work"

normal communication. Evaluation of the efficacy of those visual representations is usually limited to two stimuli, like icons and animations [2]. There is a lack of studies across all possible visual representations, especially with videos. Our work differs from previous research in that we evaluate four different visual modes (a single static image, a panel of four static images, an animation, and a video clip) illustrating 48 most frequently used verbs, not restricted to action or motion. Compared to a previous study on individual visual verbs, our study was carried out with young and elderly groups on verbs in sentences from daily communication. We also investigated visual cues and verbal properties for possible impact on interpretation. Our findings suggest how to create, select, and modify effective visual verb representations for both young and old adults.

METHODOLOGY

Construction of Verb List and Selection of Phrases

Our verb list consists of 48 most frequently used verbs[1] from the spoken materials in the British National Corpus (BNC) [8]. We acquired the top 60 verbs by sorting in frequency descending order. With a linguist and a speech-language pathologist, we compressed the list by eliminating words less common in American English, removed "like" and "know" ("um, like, you know"), and words with similar sense ("watch" vs. "look"). The final 48 verbs were assigned to the most frequently used sense, and categorized into nine domains (cognition, communication, consumption, contact, emotion, motion, perception, possession, and social) by lexical function and WordNet association [7].

Sixty-five phrases were generated by crawling sentences with target verbs from senior citizens' blogs in the Ageless Project [1], removing the ones in different senses, and simplifying complex clauses. Each verb appears twice, and each phrase has up to three verbs to test. The choice and modification of the sentences reflects the communication pattern and popular topics for both young and old adults.

Design of Four Visual Representations and a Baseline

The four visual representations were constructed as follows. The images came from tagged public domain web images, as search engine images based on surrounding texts are not suitable for this purpose, and available image databases contain few verb classes. Six images per verb were presented to seven raters (age 20-30) to assess their ability to evoke specified concepts as well as whether the images contain the visual cues (symbols, gestures, and facial expressions) examined in the studies. We assigned the most preferred image to the single image mode (Figure 1 (a)), and the next three to the four images mode (Figure 1 (b)).

The animations (Figure 1 (c), animated sequence of icons showing continuous movements or change of status) came mostly from Lingraphica [9], a popular communication support device for our ultimate user population, people with aphasia. Aphasic patients are well familiar with this icon vocabulary, and Lingraphicare has spent a large amount of human hours and money on designing, implementing, and testing those icons. Thus we believe their icons to represent the best "state of the art" available for comparison. There were three cases in which we had to create Lingraphica-style animations: (1) no animation available, i.e. "pay;" (2) in a different sense, i.e. "pick;" (3) same with other verbs, i.e. "get" and "take." For example, the new "pay" animation switched the object in the "give" animation to "$" symbol.

The videos (Figure 1 (d)) were filmed by us. Other video resources like computer vision databases and YouTube [16] are either confined to specific actions such as running, or too inconsistent and noisy. Video clips for each verb were shot based on a script selected by four reviewers out of five independently written scripts, using the following rules for filming and post-processing: (1) pure white background; (2) single leading actor; (3) extra hands or feet if interaction required; (4) no sound, text, or lip language; (5) minimized use of props; (6) clip length three seconds ± one-half second; (7) special effects if necessary (fast forwarding the "make" and "work" videos).

An additional baseline mode with verbs in the phrases left blank was added to verify perception based only on context.

Actual Study and Participants

All five modes were assigned by Latin Square to blocks of 13 sentences each in which verbs in different domains were allocated evenly. The display order of the modes was shuffled, and the phrases within each block were sorted randomly. Additionally, the nouns and adjectives were expressed by a single image picked from the web by the same means as those for verbs. All visual representations were normalized to a height of 132 pixels.

The web-based interface (Figure 2) displayed the 65 phrases one after another. Participants were asked to interpret each entire phrase. A backend script kept track of all answers as well as response time. Both videos and animations were played repeatedly. Participants were requested to rank the four visual modes by (1) difficulty in interpreting a verb; (2) speed of coming up with a thought; (3) confidence in the response; (4) personal preference; and (5) how much the context helped interpretation.

For balanced design, there were 25 participants in each age group. The younger group (age mean=22.60, SD=3.52) was

[1] See http://www.cs.princeton.edu/aphasia/verblist.txt

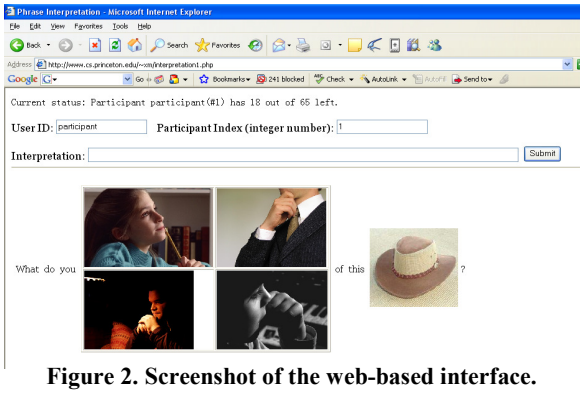

Figure 2. Screenshot of the web-based interface.

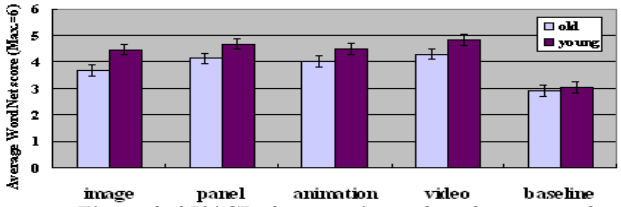

Figure 3. 95%CI of comparison of performance of four visual modes between two age groups.

recruited via posters placed locally, and the senior group (age mean=72.88, SD=5.75) was recruited through a senior center and a community exercise facility. F-A-S naming test was used to assess the word finding ability (young: mean = 49.52, SD = 13.74; senior: mean = 44.18, SD = 13.88). No participants had eyesight worse than 20/40.

RESULTS AND ANALYSIS

Evaluation Metrics

To quantitatively analyze the results, we applied four metrics: correctness/irrelevance, WordNet score, and response diversity. Correctness/irrelevance are defined as the number of the exact match and the number of irrelevant responses separately. Exact match means the response is in the same form and sense as the intended verb. For instance, the verb "pick" is in the sense of "select carefully from a group," and the response "pick, picking an apple" is considered as an irrelevant answer. WordNet score assesses the responses in a six-point scale by their semantic distance to the target verbs. Exact match gets point 6, synonyms point 5, and irrelevant ones point 1. Response diversity shows the number of difference responses received for each verb under each visual mode, which tells how well the interpretation converges. If the responses spread out in the semantic network, it means the visual representation failed to illustrate the intended verb. In another case, if the responses gathered around a verb that is different from the target, it means the representation successfully conveyed a concept, though the wrong one.

Efficacy of Visual Modes and Age Effect with Context

ANOVA results showed that there is a significant aging effect on interpreting visual verbs ($F(1,94)=4.499$, $p=0.037$, $\eta^2=0.046$). Young participants captured the concepts better than elderly participants based on each visual mode. The visual modes performed significantly differently ($F(3,210) = 7.411$, $p = 0.000$, $\eta^2 = 0.096$), with video mode having a strong win over the single static image (Figure 3), and affected by age in a similar way ($F(4,94) = 1.478$, $p = 0.204$, $\eta^2 = 0.016$). Compared to a previous study done with the four modes on illustrating individual verbs, context showed significant impact on the accuracy of concept perception (Figure 4: $F(1, 94)=40.438$, $p=0.0$, $\eta^2=0.301$), and both young and old adults benefited from context similarly.

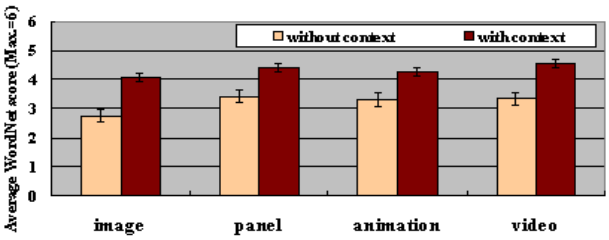

Figure 4. 95%CI of the impact of context.

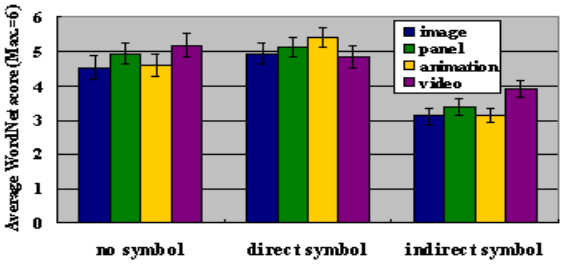

Figure 5. 95%CI of the impact of symbols, standard error varied by 0.02.

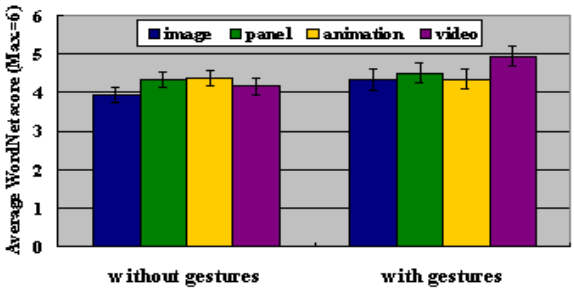

Figure 6. 95%CI of the impact of gestures, standard error varied by 0.02.

To sum up, there is a strong, nearly constant age effect for all modes, and there is a strong effect of visual mode for all ages. Data on preference, ease of use, response speed, etc. also showed marked differences between age groups, but is not displayed here due to size constraints.

Impact of Various Visual Cues on Visual Interpretation

Based on participants' feedback, we investigated the possible influence introduced by three visual cues: symbols in animations (Figure 5), gestures (Figure 6), and facial expressions (no significant effect). ANOVA results showed that indirect symbols (such as "**?**" for "wonder" and "♥" for "want") had a strong negative impact on the interpretation ($F(2,85)=14.628$, $p=0.000$, $\eta^2=0.256$), and animation mode suffered the most ($F(6,255)=2.435$, $p=0.026$, $\eta^2=0.054$). Video mode benefited the most by showing gestures (the

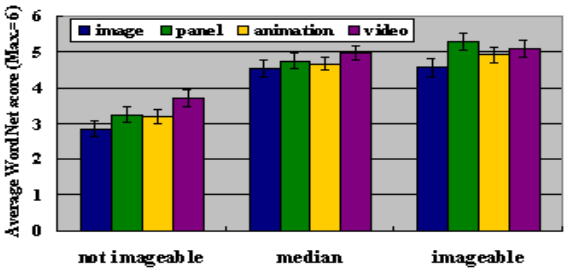

Figure 7. 95%CI of the impact of imageability.

"thinker pose" in static images for "think" and waving goodbye for "leave" ($F(3,255)=3.135$, $p=0.026$, $\eta^2=0.036$). Young and old adults were affected in a similar way. These effects were also reflected by visual modes' performance on verb domains. ANOVA result showed that the video mode outperformed other modes with motion and contact verbs ($F(24,210)=2.165$, $p=0.02$, $\eta^2= 0.198$).

To sum up, the indirect-symbol verb category which tends to be more abstract is hard to illustrate in general, however, video did not suffer from using indirect symbols. Certain verb-related gestures enhance perception, and videos can utilize them the most as a dynamic representation mode.

Visual Mode Performance by Imageability

Our verbs can be divided into three groups based on their imageability [14]. ANOVA showed that the video modes has best performance for verbs categorized as not imageable (Figure 7: $F(6,210)=1.981$, $p =0.07$, $\eta^2=0.054$).

DESIGN GUIDELINES

There are other factors that can lead to confusion, such as misapprehension of objects in the representation, distraction by background, subject's imagination, and/or unfamiliarity. Based on our results and analysis, we propose these guidelines for the design of visual verbs:

- Multiple pictures/frames are better for conveying verbs.
- Utilize common gestures if applicable, but be aware of possible cultural differences.
- Carefully use symbols, especially when not obvious.
- Simplify backgrounds (some objects (i.e. desks) were distracting), and use common scenes and props.
- Carefully use special effects in videos, especially with elderly users who are less familiar with them and might mistake fast-forwarding to "busy" or "hurry."
- Consider age-related effects like cognitive overhead, response speed, visual degeneration, and preference.

According to our post-study survey, participants tended to choose their favorite representations to visualize verbs. However if elderly participants are asked to create their own visual vocabulary, they choose images (even if less effective) because of availability and ease of creation.

CONCLUSION AND FUTURE WORK

Visual communication is helpful in multilingual settings for all ages. As an essential part of most languages, verbs must be well illustrated in visual languages. Our study compared four visual modes for conveying verbs in common communication. Results indicate that there is a strong age effect on interpretation, and the four modes perform differently for different verb types, with video mode best for motion and contact verbs, as well as verbs generally not imageable. We also formulated some basic guidelines for designing visual verbs. In terms of future work, we wish first to run studies comparing our current visual modes to refined versions based on our design guidelines. We will also test them with our target population, people with aphasia. Later, we will add the visual verbs that we verified to our current assistive communication system applications.

ACKNOWLEDGMENTS

We thank the Princeton Aphasia Project and SoundLab for assistance in study design and execution, and the Princeton Senior Resource Center for help in participant recruitment.

REFERENCES

1. Ageless Project, 2001. http://jenett.org/ageless/.
2. Baecker, R., Small, I., and Mander, R. Bringing icons to life. In *Proc. of CHI 1991*, ACM Press (1991), 1-6.
3. Blank, M, Gorelick, L, Shechtman, E, Irani, M, & Basri, R. Actions as space-time shapes. *Proc. Of ICCV*, 2005.
4. Bowles, N. and Poon, L. Aging and retrieval of words in semantic memory. *Journal of Gerontology*, 40, 1985.
5. Coleman, P.G. *Aging and Reminiscence Processes: Social and Clinical Implications*. Wiley and Sons, 1986.
6. Druks, J. and Masterson, J. *An Object and Action Naming Battery*, Psychology Press, London, 2000.
7. Fellbaum, C. *WordNet: An Electronic Lexical Database*, chapter: A semantic network of English verbs. MIT Press, Cambridge, MA, 1998.
8. Kilgarriff, A. BNC database and word frequency list: http://www.kilgarriff.co.uk/bnc-readme.html.
9. Lingraphica. http://www.lingraphicare.com/.
10. Miller, G.A. and Fellbaum, C. Semantic networks of English. *Cognition*, 41, 1991.
11. Ramsay, C.B., Nicholas, M., Au, R., Obler, L,K., and Albert, M.L. Verb naming in normal aging. *Applied Neuropsychology*, 6(2), 1999.
12. Rogers, Y. Pictorial communication of abstract verbs in relation to human-computerinteraction. *British Journal of Psychology*, 78, 99-112, 1987.
13. Stuart, S. Topic and vocabulary use patterns of elderly men and women of two age cohorts. *Doctoral Dissertation*, University of Nebraska – Lincoln, 1991.
14. UWA. MRC Psycholinguistic Database. http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm.
15. Vanrie, J. and Verfaillie, K. Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, and Computers*, 36(4): 625-629, 2004.
16. YouTube. http://youtube.com.