Sign language avatars activate phonological and semantic representations: Evidence from working memory and priming paradigms

Matthew Dye, Matt Huenerfauth & Kim Kurz

Thursday, 11:30-12:00

In order for comprehension to occur, any linguistic input must be received by the appropriate sensory system, attended to, and the relevant linguistic properties of the input extracted.

These representations can then be used to access long-term linguistic knowledge that permits the decoding of meaning. Experimental paradigms allow inferences to be made about the nature of human language processing: order list recall studies have revealed how linguistic inputs are initially represented within neural systems that mediate language comprehension, and lexical decision studies have been employed to better understand how long-term knowledge is stored and accessed. This kind of research often selects linguistic stimuli with specific properties, to determine their impact upon processing and make inferences about hidden cognitive processes. Over the past few decades, psycholinguists studying spoken languages have used artificial speech synthesizers to obtain a large degree of precision over experimental stimuli, and there have also been attempts to create more realistic speech stimuli based upon modeling of the human vocal tract [1] and speech coding theory [2]. Recently, these efforts have been extended to the development of computer-generated audiovisual speech stimuli [3]. The use of sign language avatars for psycholinguistic research, if successful, would allow a degree of control over stimuli that is difficult to achieve with videos of human signers without introducing artifacts. For example, it would be possible to manipulate the gender or skin tone of the signer without making any changes to the linguistic utterance, or the experimenter could introduce movements that violate the biomechanical constraints imposed by the human body. Here, we report two experiments that sought to assess the viability of avatars for psycholinguistic research. In both studies, experimental results derived from stimuli created by a native Deaf signer of ASL were compared to the results when avatars modeled on the human sign stimuli were used instead (see Figure 1). In Experiment 1, we sought to replicate a seminal study of phonological coding in working memory [4], and in Experiment 2, a classic semantic priming study was replicated [5]. In Experiment 1, deaf signers of ASL (N = 23) viewed lists of signs that each contained four phonologically similar or four phonologically dissimilar signs, produced by either an avatar or a human signer. After viewing each list, they were asked to sign back all of the list items in the same order that they were presented (ordered serial recall). The number of items recalled in correct list position (item scoring) and the number of lists perfectly recalled (list scoring) was recorded. Participants recalled fewer items ($p < .01$) and fewer lists ($p < .01$) in the avatar compared to the human condition. There was a trend towards poorer performance on phonologically similar lists for human stimuli (item: $p = .056$, Cohen's $d = .694$; list: $p = .081$, Cohen's $d = .604$) and for avatar stimuli (item: $p = .073$, Cohen's $d = .631$; list: $p =.103$, Cohen's $d = .544$). As predicted the effect sizes were smaller for avatar stimuli than for human stimuli. In

Experiment 2, deaf signers (N = 34) performed a lexical decision task. On each trial they saw a sequence of two signs and had to indicate whether the second sign was a real ASL sign or not. On half of the trials the second sign was a Malaysian Sign Language sign that was permissible but unattested in ASL. For the pairs of real ASL signs, the pair was semantically related on half of the trials (BASEBALL-BALL) and unrelated on the other half (MOUSE-CHAIR). All participants performed the task with both avatar and human stimuli. Analysis of response times for correct lexical decisions revealed no effect of semantic relationship (related vs. unrelated) or sign type (avatar vs. human). Accuracy data, however, revealed an effect of both semantic relationship ($p = .027$) and sign type ($p = .014$).

Participants were less accurate when responding to avatar stimuli, and less accurate for semantically unrelated pairs. Importantly, these two effects did not interact. Overall, these studies were consistent in revealing that avatar and human stimuli are processed by signers in similar ways. The working memory data suggest that encoding of avatar stimuli into working memory was slower than for human stimuli, although there was some evidence that the resultant code was phonological in nature. The priming data revealed that while overall accuracy was lower for avatars, they still activated networks of semantic representations in the mental lexicon. These psycholinguistic data align well with findings in the human-computer interaction literature which suggest that deaf people prefer a slower speed of presentation for ASL animation than for human video [6] and have concomitantly lower comprehension scores [7]. While suggesting that there is a potential for the use of avatar stimuli in psycholinguistic research into sign language comprehension, these data are based upon a sample of experienced deaf signers (with varying ages of acquisition), and the results may not hold for hearing L2 learners of a sign language.

Figure 1: Still frame of ASL sign ESCAPE produced by native Deaf signer (L) and computer-generated avatar (R)

Selected references. [1] Alku, P. et al. (1999). A method for generating natural-sounding stimuli for cognitive brain research. *Clin Neuropsych* 110(8): 1329-33. | **[2] Maia, R. et al. (2013).** Complex cepstrum for statistical parametric speech synthesis. *Speech Comm* 55(5): 606-18. | **[3] Mattheyses, W. & W. Verhelst (2015).** Audiovisual speech synthesis: an overview of the state-of-the-art. *Speech Comm* 66: 182-217. | **[4] Wilson, M. & K. Emmorey (1997).** A visuospatial "phonological loop" in working memory: evidence from American Sign Language. *Mem Cogn* 25(3): 313-20. | **[5] Bosworth, R.G. & K. Emmorey (2010).** Effects of iconicity and semantic relatedness on lexical access in American Sign Language. *J Exp Psychol: Learn Mem Cogn* 36(6): 1573-81. | **[6] Al-khazraji, S. et al. (2018).** Modeling the speed and timing of American Sign Language to generate realistic animations. In: *Proc ASSETS 2018*, ACM, New York, NY, USA, 259-70. | **[7] Kacorri, H. et al. (2013).** Effect of displaying human videos during an evaluation study of American Sign Language animation. *ACM Trans Access Comput* 5(2): 4:1-4:31.