Moving seamlessly between spoken number words and Arabic digits is common in every-day life. In the current study, we systematically investigated the correspondence between auditory number words and visual Arabic digits in adults. Auditory number words and visual Arabic digits were presented concurrently or sequentially and participants had to indicate whether they described the same quantity. We manipulated the stimulus onset asynchronies (SOAs) between the two stimuli (Experiment 1: -500 ms to +500 ms; Experiment 2: -200 ms to +200 ms). In both experiments we found a significant cross-modal distance effect. This effect was strongest for simultaneous stimulus presentation and decreased with increasing SOAs. Numerical distance emerged as the most consistent significant predictor overall, in particular for simultaneous presentation. However, physical similarity between the stimuli was often a significant predictor of response times in addition to numerical distance and at longer SOAs physical similarity between the stimuli was the only significant predictor. This shows that SOA modulates the extent to which participants access quantity representations. Our results thus support the idea that a semantic quantity representation of auditory and visual numerical symbols is activated when participants perform a concurrent matching task, while at longer SOAs participants are more likely to rely on physical similarity between the stimuli.

We also investigated whether individual differences in the efficiency of the cross-modal processing were related to differences in mathematical performance. In Experiment 1, a significant negative correlation was found between individuals’ response times in the matching task and their performance in the mathematical test. However, in the second experiment, we failed to replicate this relationship. Our results are thus inconclusive about whether the efficiency of cross-format numerical correspondence is related to mathematical competence in adults.

Experiment 1:

Stimuli and Procedure

Stimuli presentation and data recording were controlled by Presentation® (Version 17.2, www.neurobs.com). The numbers 2, 3, 4, 5, 6, and 8 were used in visual and auditory format (7 was excluded because the auditory number word contains two syllables). The visual Arabic digit was displayed in the middle of an 18.3-inch screen, in white against a black background, in Arial font, 48 pt. The sound for each auditory number word was recorded digitally by a male native-English speaker in a soundproof booth. The sound files were manipulated by using Cool Edit 2000 to be roughly equal in length (around 450 ms). The sounds were played binaurally through a headphone.

On each trial, a fixation cross (white, 48 pt) was displayed first in the centre of the screen. After 500 ms the fixation disappeared and bimodal numerals were displayed. An Arabic single digit was presented at the centre of the screen for 450 ms and a spoken number word was played for around 450 ms (mean length of sounds = 449.43 ms, SD = 2.64 ms, range from 444 to 459 ms). Nine different stimulus-onset asynchronies (SOAs) were used: -500, -300, -200, -100, 0, +100, +200, +300, and +500 ms. These manipulations of SOAs led to three types of sequences: visual first-then-auditory condition (VA) with negative SOAs (-500, -300, -200 & -100 ms), the auditory first-then-visual condition (AV) with positive SOAs (+100, +200, +300, +500 ms), and simultaneous display (0 ms) (see Figure 1). The inter-trial interval was 500 ms.

Every combination of each digit for each SOA condition was displayed in a random order across blocks (9 blocks in total, 60 trials for each block). The sequence of stimulus presentation was randomly generated but fixed across participants. The experiment started with a 12-trial practice block. Using a standard QWERTY keyboard, participants were instructed to respond by key buttons pressing (‘Z’ and ‘/’), ‘same’ when they saw and heard the same number (matching trials), and to respond ‘different’ when the written digit and the sound of number word were different (non-matching trials). The button allocation was counter-balanced between subjects. To balance the same and different responses, the ‘same’ pair (e.g., trials when they saw a digit ‘2’ and heard a number word ‘two’) was displayed five times in each SOA whereas there was only one trial for each ‘different’ pair (e.g., 2-three, 2-four, 2-five, 2-six, and 2-eight). In total, there were 540 trials. Note that the reaction time was measured as the time between the onset of the second stimulus and the first response.

Individual math performance was assessed with the Math Computation subtest of the Wide Range Achievement Test (WRAT-4, Wilkinson & Robertson, 2006) after the computerised task. Participants were asked to solve as many arithmetic questions as possible in 15 minutes (a maximum of 40 questions). Calculators were prohibited. The age-standardised score was calculated for each participant following the test manual. The delayed and immediate test-retest reliability reported in the test manual is r = .83 and .88, respectively.

Experiment 2:

The same procedure as in Experiment 1 was used. However, stimuli presentation and data recording were now controlled by MATLAB with Psychophysics Toolbox extensions (Matlab Psychtoolbox-3; www.psychtoolbox.org) and importantly, we used different SOA conditions: nine shorter SOAs were used in the present experiment, they were -200 ms, -150 ms, -100 ms, -50 ms, 0 ms, +50 ms, +100 ms, +150 ms, and +200 ms.