Recent evidence from Alsius and Soto-Faraco (2011) suggests that selective attention is required to locate a talking face in a multi-talker array, seemingly in contrast to previous claims that the integration of faces and voices is preattentive (McGurk and MacDonald, 1976). The current study investigated what effect degrading the auditory signal has on the ability to locate a talking face. Twenty participants were presented with between 2 and 4 moving faces, each of which was articulating a different sentence. The task was to decide, as quickly as possible, which of these faces matched the auditory sentence that they heard at the same time. The results showed that in the least demanding auditory condition (clear speech in quiet), increasing the number of faces on screen did not increase visual search times. However, when speech was presented in background noise or was processed to simulate the information provided by a cochlear implant ('sine-wave vocoded speech'), search times increased as the number of faces increased even though intelligibility of the sentences was unchanged. The results suggest that under conditions of low perceptual load it is possible for audiovisual correspondence to 'pop out', but if perceptual load is increased then selective attention is required to bind faces and voices.