

Watching videos with subtitles in the original language is one of the most effective ways of learning a foreign language. Highlighting words at the moment they are pronounced helps to synchronize visual and auditory perception and increases learning efficiency. The method for aligning orthographic transcriptions to audio recordings is known as *forced alignment*. This work implements a tool for aligning transcript of *YouTube* videos with the speech in their audio recording, providing a web user interface with video player presenting the results. It integrates two state-of-the-art forced aligners based on *Kaldi*, first using standard *HMM* approach, second based on neural networks and compares their accuracy. Integrated aligners also provide a phone level alignment, which can be used for training statistical models in further speech recognition research. Work describes implementation and architectural concepts the tool is based on, which can be used in various software projects.