

ABSTRACT

Disclosed in the present disclosure are a method and device for generating a speech recognition training set. A specific embodiment of the method comprising: obtaining an audio to be processed and a video to be processed, wherein said video comprises text information corresponding to said audio; recognizing said audio to obtain an audio text; recognizing text information in said video to obtain a video text; and on the basis of the consistency of the audio text and the video text, taking said audio as a voice sample and the video text as a label to obtain a speech recognition training set.

Fig. 2