

ABSTRACT

A method and apparatus for training a speech conversion model, a device, and a medium. The method comprises: training a first ASR model on the basis of first sample audio, and training a second ASR model on the basis of second sample audio (201); training a first conversion model on the basis of a first sample text and a first sample content feature corresponding to the first sample audio (202); constructing parallel sample data on the basis of the first conversion model, and a second sample text and a second sample content feature corresponding to the second sample audio (203); training a second conversion model on the basis of the parallel sample data, the second conversion model being used to perform content feature conversion between a first accent and a second accent (204); training a third conversion model on the basis of sample content features of different sample audios (205); generating a speech conversion model on the basis of the trained first ASR model, second conversion model and third conversion model (206).

Figure 13