

ABSTRACT

The present disclosure proposes a speech conversion scheme for non-parallel corpus training, to get rid of dependence on parallel text and resolve a technical problem that it is difficult to achieve speech conversion under conditions that resources and equipment are limited.

5 A voice conversion system and a training method therefor are included. Compared with the prior art, according to the embodiments of the present disclosure: a trained speaker-independent automatic speech recognition model can be used for any source speaker, that is, the speaker is independent; and bottleneck features of audio are more abstract as compared with phonetic posteriorGram features, can reflect decoupling of spoken content and timbre of the speaker, and
10 meanwhile are not closely bound with a phoneme class, and are not in a clear one-to-one correspondence relationship. In this way, a problem of inaccurate pronunciation caused by a recognition error in ASR is relieved to some extent. Pronunciation accuracy of audio obtained by performing voice conversion by the bottleneck feature is obviously higher than that of a phonetic posteriorGrambased method, and timbre is not significantly different. By means of a transfer
15 learning mode, dependence on training corpus can be greatly reduced.