ABSTRACT

Provided is a method for training a vocal not recognition model. The method includes: acquiring a labeled vocal audio, a vocal note label corresponding to the labeled vocal audio, a pure vocal audio, and an accompaniment audio; acquiring a trained first network by training a first network based on the labeled vocal audio, the accompaniment audio, and the vocal note label corresponding to the labeled vocal audio, wherein the first network is configured to output a vocal note recognition result based on a synthesized audio generated from the labeled vocal audio and the accompaniment audio; and acquiring the vocal note recognition model by training a second network based on the trained first network, the pure vocal audio, and the accompaniment audio, wherein the second network is configured to output a vocal note recognition result based on a synthesized audio generated from the pure vocal audio and the accompaniment audio.

Fig. 2