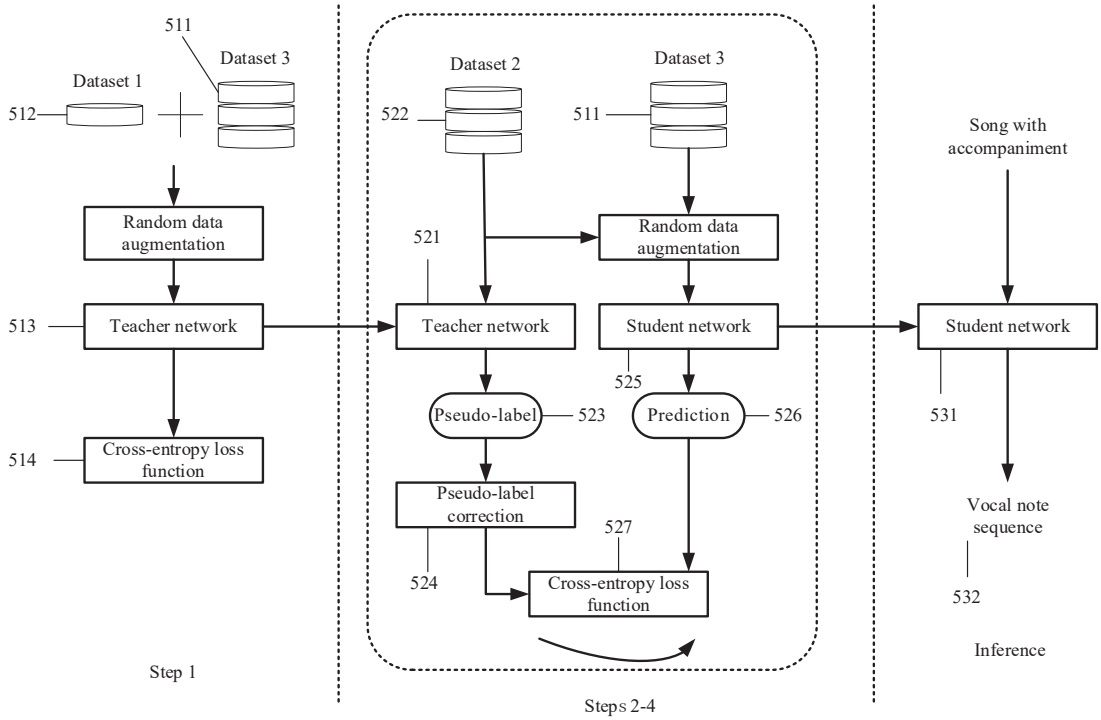
511
Dataset 3
Dataset 1
512
Random data augmentation
513
Teacher network
514
Cross-entropy loss function
Step 1
Dataset 2
522
Dataset 3
511
521
Random data augmentation
Teacher network
Student network
525
Pseudo-label
523
Prediction
526
Pseudo-label correction
527
524
Cross-entropy loss function
Steps 2-4
Song with accompaniment
Student network
531
Vocal note sequence
532
Inference

FIG. 5

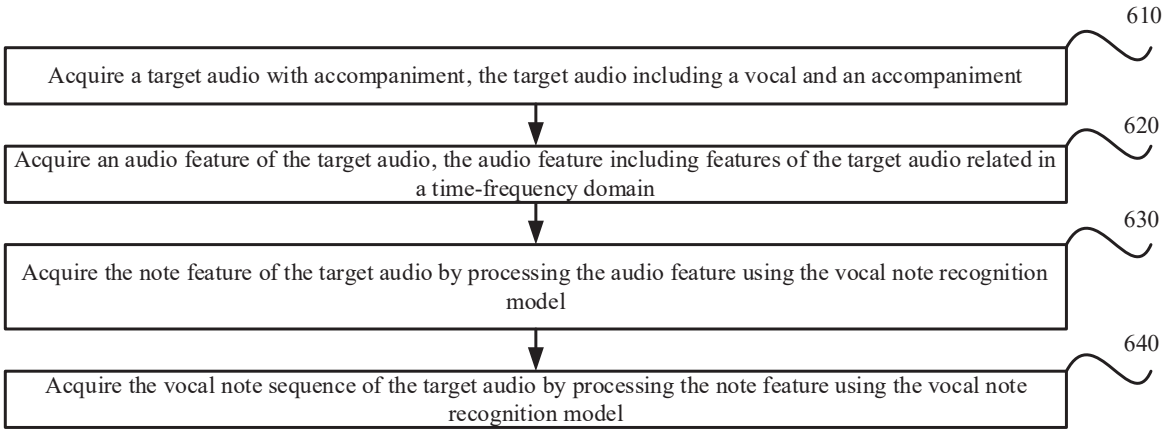
610
Acquire a target audio with accompaniment, the target audio including a vocal and an accompaniment
620
Acquire an audio feature of the target audio, the audio feature including features of the target audio related in a time-frequency domain
630
Acquire the note feature of the target audio by processing the audio feature using the vocal note recognition model
640
Acquire the vocal note sequence of the target audio by processing the note feature using the vocal note recognition model

FIG. 6