## CLAIMS

1. A method for audio processing, comprising:

extracting, from first media content, background audio content and speech audio content corresponding to text content;

generating, based on a request to replace first text in the text content with second text, a second audio segment corresponding to the second text using timbre information, the timbre information being associated with a first audio segment of the speech audio content, the first audio segment corresponding to the first text;

adjusting, based on the second audio segment, a third audio segment of the background audio content corresponding to the first text; and

generating second media content based on the first media content, the second audio segment, and the adjusted third audio segment.

2. The method of claim 1, wherein generating the second audio segment corresponding to the second text using the timbre information, the timbre information being associated with the first audio segment of the speech audio content, comprises:

determining speech speed information corresponding to the first audio segment; and

generating the second audio segment based on the timbre information and the speech speed information, to make the second audio segment match the speech speed information.

3. The method of claim 1 or 2, further comprising:

determining at least one utterance associated with the first text; and

determining the timbre information associated with the first audio segment based on timbre of the at least one utterance.

4. The method of claim 1 or 2, wherein adjusting, based on the second audio segment, the third audio segment of the background audio content corresponding to the first text comprises:

determining speed adjustment information based on first duration of the first audio segment and second duration of the second audio segment; and

adjusting speed of the third audio segment based on the speed adjustment information.

5. The method of claim 1 or 2, wherein the first media content comprises picture content, and generating the second media content based on the first media content, the second audio segment, and the adjusted third audio segment comprises:

adjusting, based on the second audio segment, a picture segment corresponding to the first audio segment in the picture content; and

generating the second media content by fusing the adjusted picture segment, the first audio content, and the adjusted third audio segment.

6. The method of claim 1 or 2, wherein generating the second media content based on the first media content, the second audio segment, and the adjusted third audio segment comprises:

fusing the second audio segment and the adjusted third audio segment to obtain intermediate audio content;

adjusting a reverberation effect or volume of the intermediate audio content; and

generating the second media content based on the adjusted intermediate audio content.

7. The method of claim 6, wherein adjusting the reverberation effect of the intermediate audio content comprises:

determining a reverberation parameter based on the first media content; and

adjusting the reverberation effect of the intermediate audio content based on the reverberation parameter.

8. The method of claim 6, wherein adjusting the volume of the intermediate audio content comprises:

performing global volume equalization on the intermediate audio content.

9. An apparatus for audio processing, comprising:

an audio extraction module configured to extract, from first media content, background audio content and speech audio content corresponding to text content;

an audio generation module configured to generate, based on a request to replace first text in the text content with second text, a second audio segment corresponding to the second text using timbre information, the timbre information being associated with a first audio segment of the speech audio content, the first audio segment corresponding to the first text;

an audio adjustment module configured to adjust, based on the second audio segment, a third audio segment of the background audio content corresponding to the first text; and

a media generation module configured to second media content based on the first media content, the second audio segment, and the adjusted third audio segment.

10. An electronic device, comprising:

at least one processing unit; and

at least one memory coupled to the at least one processing unit and storing instructions executable by the at least one processing unit, the instructions, when executed by the at least one processing unit, causing the electronic device to perform the method of any of claims 1 to 8.

11. A computer-readable storage medium having a computer program stored thereon, the computer program

being executable by a processor to implement the method of any of claims 1 to 8.

12. An apparatus for audio processing, comprising:

a means for extracting, from first media content, background audio content and speech audio content corresponding to text content;

a means for generating, based on a request to replace first text in the text content with second text, a second audio segment corresponding to the second text using timbre information, the timbre information being associated with a first audio segment of the speech audio content, the first audio segment corresponding to the first text;

a means for adjusting, based on the second audio segment, a third audio segment of the background audio content corresponding to the first text; and

a means for generating second media content based on the first media content, the second audio segment, and the adjusted third audio segment.

13. The apparatus of claim 12, wherein the means for generating further comprises:

a means for determining speech speed information corresponding to the first audio segment; and

a means for generating the second audio segment based on the timbre information and the speech speed information, to make the second audio segment match the speech speed information.

14. The apparatus of claim 12 or 13, further comprising:

a means for determining at least one utterance associated with the first text; and

a means for determining the timbre information associated with the first audio segment based on timbre of the at least one utterance.

15. The apparatus of claim 12 or 13, wherein the means for adjusting further comprises:

a means for determining speed adjustment information based on first duration of the first audio segment and second duration of the second audio segment; and

a means for adjusting speed of the third audio segment based on the speed adjustment information.