

METHOD AND APPARATUS FOR LIP-SYNC GENERATION OF DIGITAL HUMANS

ABSTRACT

Embodiments of this specification relate to a method and an apparatus for lip-sync
5 generation of digital humans. Input data of the method includes text data and audio data
corresponding to the text data. The text data and the audio data are input to an alignment module
for alignment to obtain an audio segment and a time period corresponding to any text unit in
the text data. Corresponding mouth shape information is then obtained using a pre-configured
text-mouth shape dictionary based on the text unit. Amplitude information corresponding to a
10 mouth shape is obtained based on a specific audio feature of the audio segment, and a
corresponding lip shape over one time period may be obtained by integrating the mouth shape
information and the amplitude information. A preliminary lip shape sequence is obtained by
sorting lip shapes over different time periods in time order. Interpolation and smoothing
operations are performed on a blank time period between adjacent lip shapes in the lip shape
15 sequence to implement transition between different lip shapes, thereby obtaining a final
smoothed lip shape sequence.

(Fig. 2)