

ABSTRACT

A streaming speech synthesis method for streaming input, including: obtaining streaming text information, and dynamically adjusting a maximum context of released text information this time according to an overall real-time factor of streaming speech synthesis of the streaming text information; performing speech encoding processing on received streaming text information according to the maximum context of released text information this time, to output a text information feature required for streaming synthetic audio; and performing speech decoding processing on the text information feature, to output a time domain signal or a spectral feature of a streaming speech. The present disclosure may dynamically adjust the size and duration of output audio according to the overall real-time factor of streaming speech synthesis and an input rate of streaming text, and obtain the best speech synthesis effect while ensuring that the streaming speech synthesis is not interrupted to the greatest extent.

FIG. 2