Pseudocode for Text-to-Speech (TTS) Model

Xuhui Zhan, Siyu Yang

Algorithm 1 Text-to-Speech Model

Require: *text*: Input text string
Require: *lang_model*: Language model for text processing
Require: *speech_model*: Speech synthesis model

1: **function** TEXTTOSPEECH(*text, lang_model, speech_model*)
2:     *preprocessed_text* ← PREPROCESSTEXT(*text, lang_model*)
3:     *phonemes* ← TEXTTOPHONEME(*preprocessed_text, lang_model*)
4:     *audio_output* ← SYNTHESIZESPEECH(*phonemes, speech_model*)
5:     PLAYAUDIO(*audio_output*)
6: **end function**
7: **function** PREPROCESSTEXT(*text, lang_model*)
8:     Convert *text* to lowercase
9:     Remove punctuation and special characters from *text*
10:     **return** *text*
11: **end function**
12: **function** TEXTTOPHONEME(*text, lang_model*)
13:     Convert processed *text* to a sequence of phonemes using *lang_model*
14:     **return** *phonemes*
15: **end function**
16: **function** SYNTHESIZESPEECH(*phonemes, speech_model*)
17:     Convert *phonemes* to audio using *speech_model*
18:     **return** *audio*
19: **end function**
20: **function** PLAYAUDIO(*audio*)
21:     Output the synthesized *audio* through speakers
22: **end function**