附图

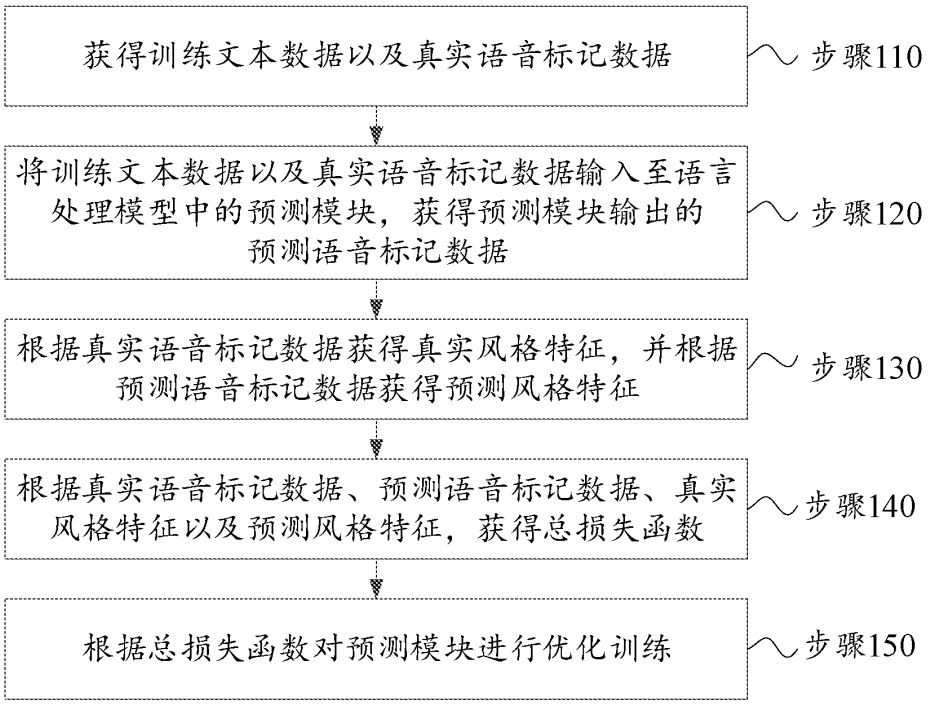
获得训练文本数据以及真实语音标记数据
步骤110
将训练文本数据以及真实语音标记数据输入至语言处理模型中的预测模块，获得预测模块输出的预测语音标记数据
步骤120
根据真实语音标记数据获得真实风格特征，并根据预测语音标记数据获得预测风格特征
步骤130
根据真实语音标记数据、预测语音标记数据、真实风格特征以及预测风格特征，获得总损失函数
步骤140
根据总损失函数对预测模块进行优化训练
步骤150

图 1

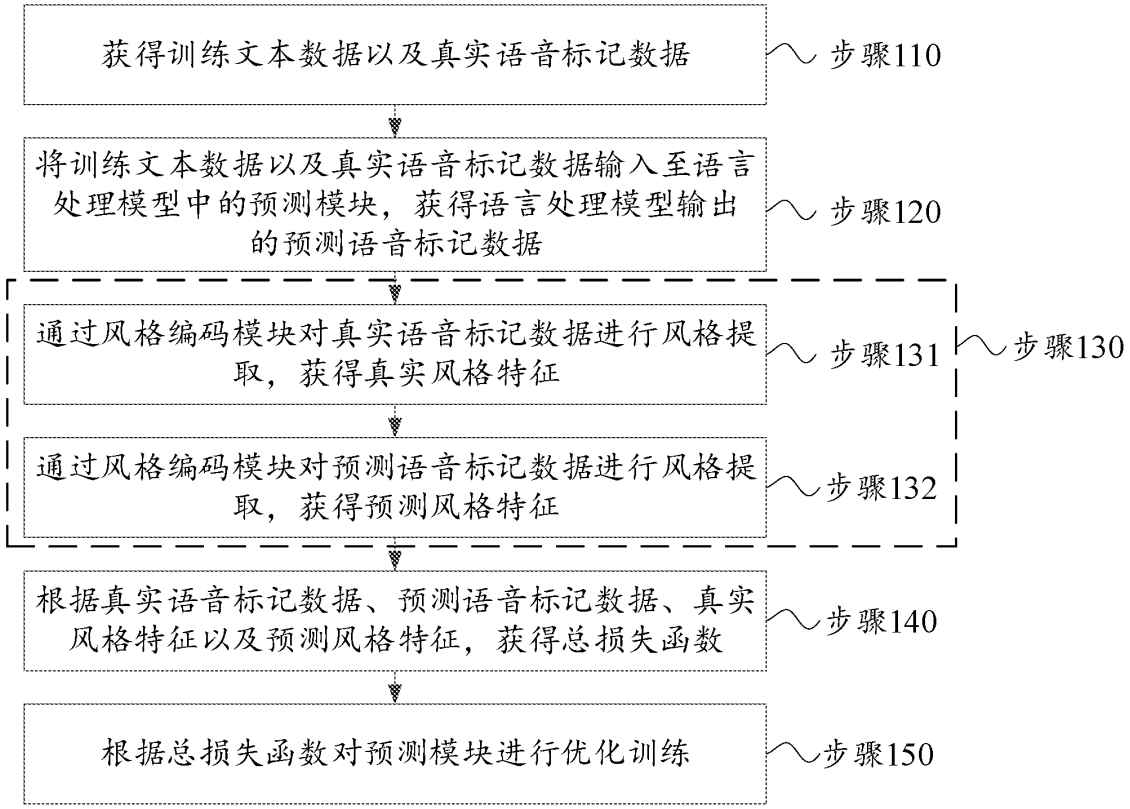
获得训练文本数据以及真实语音标记数据
步骤110
将训练文本数据以及真实语音标记数据输入至语言处理模型中的预测模块，获得语言处理模型输出的预测语音标记数据
步骤120
通过风格编码模块对真实语音标记数据进行风格提取，获得真实风格特征
步骤131
通过风格编码模块对预测语音标记数据进行风格提取，获得预测风格特征
步骤132
步骤130
根据真实语音标记数据、预测语音标记数据、真实风格特征以及预测风格特征，获得总损失函数
步骤140
根据总损失函数对预测模块进行优化训练
步骤150

图 2

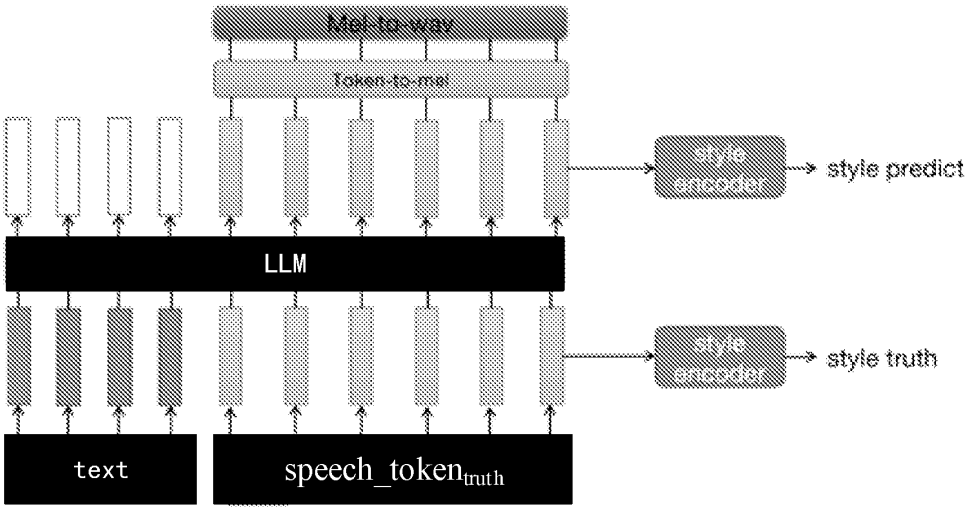
Mel-to-wav
token-to-mel
style encoder
style predict
LLM
style encoder
style truth
text
speech_token$_{truth}$

图 3

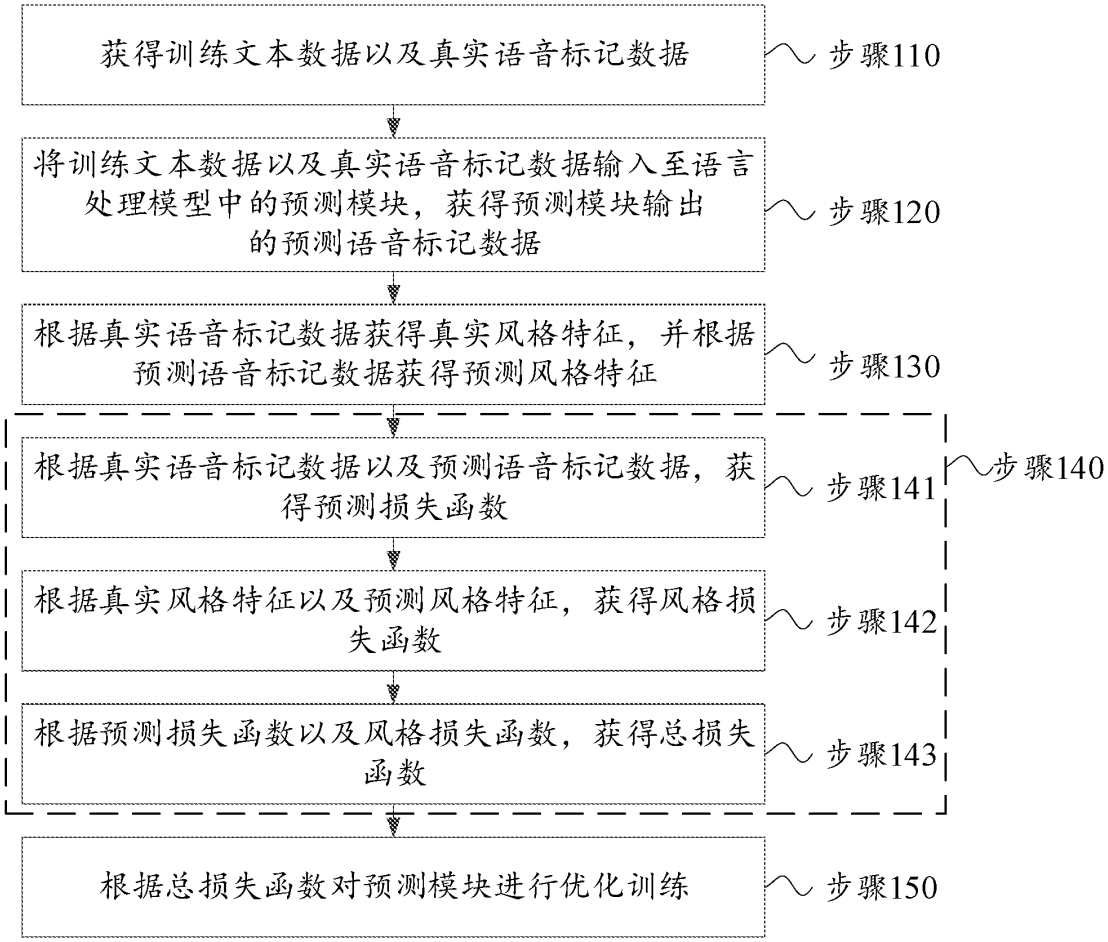
获得训练文本数据以及真实语音标记数据
步骤110
将训练文本数据以及真实语音标记数据输入至语言处理模型中的预测模块，获得预测模块输出的预测语音标记数据
步骤120
根据真实语音标记数据获得真实风格特征，并根据预测语音标记数据获得预测风格特征
步骤130
根据真实语音标记数据以及预测语音标记数据，获得预测损失函数
步骤141
步骤140
根据真实风格特征以及预测风格特征，获得风格损失函数
步骤142
根据预测损失函数以及风格损失函数，获得总损失函数
步骤143
根据总损失函数对预测模块进行优化训练
步骤150

图 4

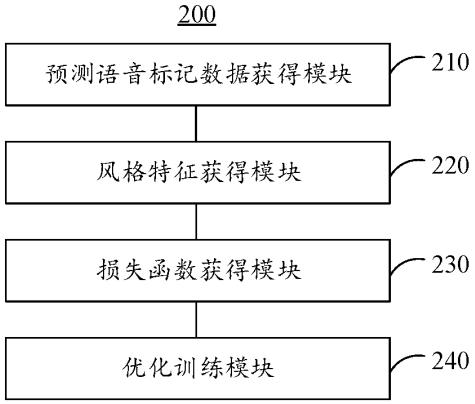
200
预测语音标记数据获得模块
210
风格特征获得模块
220
损失函数获得模块
230
优化训练模块
240

图 5

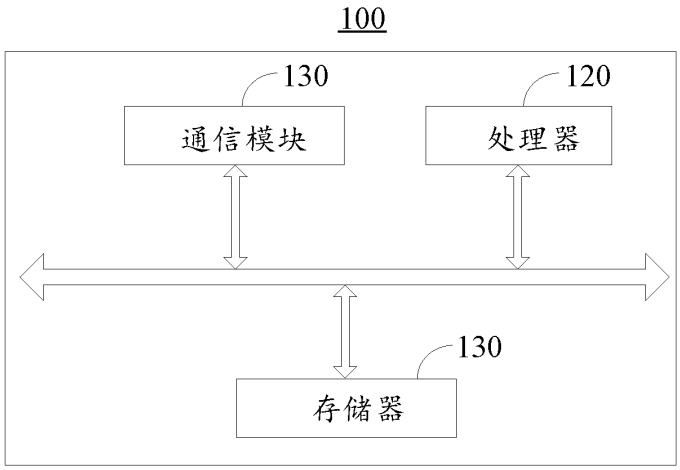

图 6