

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(43) 国际公布日
2016年11月3日 (03.11.2016)

WIPO | PCT



(10) 国际公布号

WO 2016/172871 A1

(51) 国际专利分类号:

G10L 13/00 (2006.01) G10L 13/08 (2013.01)

SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(21) 国际申请号:

PCT/CN2015/077785

(22) 国际申请日:

2015年4月29日 (29.04.2015)

(25) 申请语言:

中文

(26) 公布语言:

中文

(72) 发明人及

(71) 申请人: 华侃如 (HUA, Kanru) [CN/CN]; 中国上海市碧波路49弄8号801室, Shanghai 201203 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST,

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则 4.17 的声明:

- 关于发明人身份(细则 4.17(i))
- 关于申请人有权申请并被授予专利(细则 4.17(ii))
- 发明人资格(细则 4.17(iv))

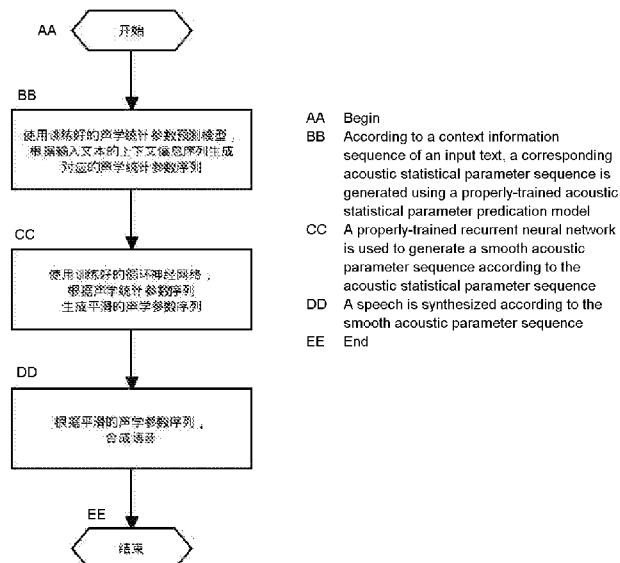
本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

(54) Title: SPEECH SYNTHESIS METHOD BASED ON RECURRENT NEURAL NETWORKS

(54) 发明名称: 基于循环神经网络的语音合成方法

图 2



(57) Abstract: A speech synthesis method based on recurrent neural networks specifically comprises the following steps of: acquiring context information of a text to be synthesized; generating an acoustic statistical parameter sequence according to the context information of the text; according to the acoustic statistical parameter sequence generated from the context information, using a recurrent neural network to generate an acoustic parameter sequence of a speech to be synthesized; and synthesizing the speech according to the acoustic parameter sequence of the speech to be synthesized. Compared with traditional statistical parameter speech synthesis methods, the method gives the synthesized speech better naturalness and has good real-time property.

(57) 摘要: 一种基于循环神经网络的统计参数语音合成方法, 具体包括以下步骤: 获取待合成文本的上下文信息; 根据文本的上下文信息, 生成声学统计参数序列; 根据由上下文信息生成的声学统计参数序列, 使用循环神经网络生成待合成语音的声学参数序列; 根据待合成语音的声学参数序列, 合成语音。相比传统的统计参数语音合成方法, 该方法合成的语音更加自然, 且具有较好的实时性。

基于循环神经网络的语音合成方法

技术领域

[0001] 本发明涉及语音合成领域，尤其涉及统计参数语音合成。

背景技术

[0002] 语音合成技术是让机器或程序根据文本信息产生人类可懂的语音的技术，与语音合成技术相关的应用包括文语转换（TTS）和歌声合成（SVS）等。

[0003] 目前主流的语音合成技术是基于隐马尔科夫模型（HMM）的统计参数语音合成技术，该技术包括训练和运行两个阶段。

[0004] 训练阶段：将用于训练语音数据的声学参数和隐马尔科夫模型的状态序列相对应，由训练算法计算出各状态的声学统计参数；使用决策树对模型的状态按照文本的上下文信息进行归类。

[0005] 运行阶段：使用决策树将输入的文本的上下文信息序列转换为经过归类的模型的状态序列；根据各状态对应的声学统计参数，获取声学统计参数序列。由于隐马尔科夫模型的状态离散性质，此时获得的声学统计参数序列是状态间不连贯的。为了生成连贯的语音声学参数，需要对声学统计参数序列进行平滑处理。传统的平滑方法是最大似然参数生成算法（MLPG）。该方法根据包括动态参数（如一阶、二阶导数）的声学统计参数序列，生成具有最匹配的统计学特征的连贯的声学参数序列。最后，根据声学参数序列，使用源-滤波器模型或其它语音分析合成技术，合成功能语音波形数据并输出。

[0006] 该平滑方法的主要问题在于：1) 该平滑方法容易使生成的声学参数过度平滑，最终导致生成的语音听上去含糊不清；2) 该平滑方法不具备实时性，即只能逐句生成声学参数，在实时语音合成的应用中易导致播放卡顿。

[0007] 避免传统的统计参数语音合成技术生成的声学参数过度平滑的方法之一是采用全局方差（Global Variance）标准生成声学参数（见 Tomoki, Toda, et al. "A speech parameter generation algorithm considering global variance for HMM-based speech synthesis." IEICE TRANSACTIONS on Information and Systems 90.5 (2007): 816-824.），但是这种改进技术未能解决参数生成的实时性问题。

[0008] 避免传统的统计参数语音合成技术生成的声学参数过度平滑的另一方法是使用深度神经网络代替隐马尔科夫模型（见 Zen, Heiga, et al. "Statistical parametric speech synthesis using deep neural networks." Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013.）。深度神经网络将根据文本的上下文信息序列，直接生成对应的声学统计参数序列，使生成的声学统计参数序列更接近真实语音的统计特征，从而获得更自然的合成语音。但是这种深度神经网络对训练数据的需求量大，通常需要约5个小时的训练语音数据，才能得到较好的合成效果。如准备足够的训练语音数据，需要大量时间和人力成本。

[0009] 本发明为解决以上技术问题，使用循环神经网络代替传统的声学参数平滑方法，根据声学统计参数序列生成连贯的声学参数，有效解决了生成的声学参数过度平滑的问题，使合成语音在听感上更自然。同时，本发明具备实时性，可逐句、逐字或逐帧合成语音。另外，本发明对训练数据的需求量相对较小，训练时间相对较短。

发明内容

[0010] 本发明所属的技术领域是基于统计参数的语音合成。本发明解决的技术问题之一

是避免了合成语音的过度平滑现象，使合成语音更自然、清晰。

[0011] 本发明为解决以上技术问题，采取的方法包括模型的训练和运行两个步骤。其中模型的训练步骤包括：

- [0012] 获取训练数据中文本的上下文信息；
- [0013] 训练声学统计参数预测模型，将文本的上下文信息映射到声学统计参数；
- [0014] 使用训练好的声学统计参数预测模型，根据训练文本的上下文信息序列生成对应的声学统计参数序列；
- [0015] 训练循环神经网络，将声学统计参数序列映射到训练语音的声学参数序列。
- [0016] 本发明的运行步骤包括：
 - [0017] 获取输入文本的上下文信息；
 - [0018] 使用训练好的声学统计参数预测模型，根据输入文本的上下文信息序列生成对应的声学统计参数序列；
 - [0019] 使用训练好的循环神经网络，根据声学统计参数序列，生成平滑的声学参数序列；
 - [0020] 根据平滑的声学参数序列，合成语音。
- [0021] 其中，由平滑的声学参数序列合成语音的方法可以是基于源-滤波器模型的参数语音合成、基于正弦模型或谐波加噪音模型的参数语音合成、声码器、基于基元选择的拼接语音合成等。
- [0022] 其中，声学统计参数预测模型可以是隐马尔科夫模型、决策树或神经网络等机器学习模型和方法。对于具体采用的模型或方法本发明不做具体限制。

附图说明

- [0023] 图 1 为本发明的模型的训练阶段流程图；
- [0024] 图 2 为本发明的运行阶段流程图；
- [0025] 图 3 为本发明的循环神经网络训练方法示意图；
- [0026] 图 4 为本发明的循环神经网络运行方法示意图；
- [0027] 图 5 为本发明的实施例中声学统计参数序列的示例图；
- [0028] 图 6 为本发明的实施例中经过平滑和正规化处理的声学统计参数序列和循环神经网络输出的声学参数序列的示例图。

具体实施方式

[0029] 本发明包括模型的训练和运行两个阶段，其中模型的训练阶段如图 1 所示；运行阶段如图 2 所示。

[0030] 模型的训练阶段主要是根据训练数据，计算出声学统计参数预测模型参数和循环神经网络模型参数。

[0031] 训练数据包括语音数据和与语音数据时间对齐的文本数据。在不同的应用场合，语音数据可以是不同的形式，例如在文语转换应用中，语音数据为朗读句子的音频数据；在歌声合成应用中，语音数据为歌声的音频数据。文本数据包括与语音相对应的文字以及音标标注信息，还可包含音节标注、重音标注、词性标注等信息。

[0032] 声学统计参数预测模型的作用是，根据输入的文本信息，预测出语音声学参数（包括语音在特定时间的韵律、音色参数等）在不同时间的统计信息，从而产生对语音声学参数的初步预测。该模型的输出可以是离散的，即输出为一串状态间不连贯的声学统计参数。该模型的输出也可以是连续的，即输出为一串状态间连贯的声学统计参数。该模型输出的参数应反映语音的声学参数在短时间内的统计信息（例如平均值、方差、导数

等），而不仅是声学参数本身。

[0033] 本发明对于具体使用的声学统计参数预测模型不作具体限定。本实施例中使用的声学统计参数预测模型为基于隐马尔科夫模型和决策树的声学统计参数预测模型；具体实施时也可以是其它具有近似功能的模型或方法，例如前馈神经网络、支持向量机等。

[0034] 本发明与现有的统计参数语音合成技术相比，使用循环神经网络代替了传统的语音声学参数平滑方法，解决了生成的声学参数过度平滑的问题；本发明能够逐句、逐字、或逐帧进行语音合成，解决了该技术的低实时性问题。

[0035] 本发明的模型的训练阶段如图1所示，具体包括以下步骤：

[0036] 第一步，从训练数据中获取语音数据和对应的文本数据，及文本的上下文信息，训练声学统计参数预测模型，使该模型能够将文本的上下文信息映射到声学统计参数。

[0037] 本实施例采用的声学统计参数预测模型的训练方法包括：

[0038] 根据对齐的训练文本和训练语音数据，初始化隐马尔科夫模型的状态转移概率分布和输出概率分布参数。隐马尔科夫模型采用上下文相关状态；可选地，输出概率分布采用混合概率分布；可选地，输出概率分布采用对角斜方差矩阵。状态转移概率分布参数可由训练文本中不同状态的出现顺序和次数计算得到；输出概率分布参数可由各状态所对应的语音数据的声学参数统计得到；

[0039] 根据隐马尔科夫模型状态在文本中的上下文信息，以及该模型的输出概率分布参数，生成决策树；决策树的生成算法可以使用最小描述长度（MDL）准则；

[0040] 根据决策树，绑定隐马尔科夫模型中归在相同决策树节点下的上下文相关状态；使用Baum-Welch算法或Viterbi训练算法重新计算隐马尔科夫模型的状态转移概率分布和输出状态分布参数；此时获得的输出概率分布参数即为各状态对应的声学统计参数；

[0041] 为了能够在本发明运行阶段中生成各状态的时长参数，统计并存储各组绑定的状态的平均时长和方差。

[0042] 本实施例采用的基于隐马尔科夫模型和决策树的声学统计参数预测模型仅作为例子，具体实施时可以采用与本实施例不同的模型和训练方法。

[0043] 第二步，使用训练好的声学统计参数预测模型，根据训练数据中的文本数据的上下文信息序列，生成对应的声学统计参数序列。

[0044] 以基于隐马尔科夫模型和决策树的声学统计参数预测模型为例，根据训练文本数据的上下文信息序列，使用决策树选择相对应的隐马尔科夫模型状态，组成状态序列；各状态持续时间由训练阶段第一步中的Viterbi训练算法决定，从而保证状态序列和训练语音数据在时间上对齐。然后根据状态序列，使用训练好的隐马尔科夫模型生成相对应的声学统计参数序列。

[0045] 第三步，如图3所示，训练循环神经网络，使该神经网络能够将声学统计参数预测模型输出的声学统计参数序列映射到连贯、自然的语音的声学参数序列。循环神经网络的训练数据为训练阶段第二步生成的声学统计参数序列和根据训练数据中的语音数据计算得到的语音声学参数。

[0046] 由于本实施例训练阶段第二步中使用的隐马尔科夫模型的输出为状态间不连贯的声学统计参数序列，应当对隐马尔科夫模型输出的声学统计参数序列做初步平滑处理，再输入循环神经网络，以确保循环神经网络能够输出连贯的语音参数序列；初步平滑处理方法可以是插值、低通滤波、平均值滤波和最大似然参数生成算法等；

[0047] 可选地，对输入循环神经网络的声学统计参数序列做正规化处理，例如使输入参数符合均值为0，方差为1的高斯分布；

[0048] 针对不同应用场合，可使用不同结构的循环神经网络，包括一阶或二阶循环神经网络、长短时记忆神经网络、多层循环神经网络、以及以上若干种神经网络的组合等。循

环神经网络的输出层采用线性激活函数。当存在多个声学参数时，可使用一个循环神经网络根据输入声学统计参数序列，输出多个声学参数序列；亦可使用多个循环神经网络根据输入声学统计参数序列，每个循环神经网络输出一个或多个声学参数序列；

[0049] 针对不同结构的循环神经网络，使用不同的训练算法。一阶循环神经网络可使用随时间演化的反向传播算法（BPTT）或实时递归学习算法（RTRL）训练；二阶循环神经网络、长短时记忆神经网络、多层循环神经网络可使用通用长短时神经网络算法（LSTM-g，见 Monner, Derek, et al. "A generalized LSTM-like training algorithm for second-order recurrent neural networks." Neural Networks 25 (2012): 70-83.）训练；训练时循环神经网络的输出层采用最小平方误差准则；

[0050] 由于循环神经网络的输出具有延迟性，训练中须要将循环神经网络的输入数据的声学统计参数序列提前大约 5 至 40 帧，以提供给循环神经网络足够时间预测输出的声学参数；

[0051] 可选地，为了避免循环神经网络由于输入和输出数据的时间差遗忘输入或输出信息，造成输出的语音声学参数过度平滑，循环神经网络的输入同时包括提前 5 至 40 帧的声学统计参数序列，以及当前帧的声学统计参数序列。

[0052] 本发明的运行阶段流程如图 2 所示，该阶段主要根据输入文本，使用训练好的模型，产生语音声学参数并合成语音，具体方法包括以下步骤：

[0053] 第一步，从输入文本中获取文本的上下文信息序列；根据输入文本的上下文信息序列，使用声学统计参数预测模型，生成对应的声学统计参数序列。

[0054] 使用本实施例采用的声学统计参数预测模型生成声学统计参数序列的方法具体包括：

[0055] 根据输入文本的上下文信息序列，使用决策树选择相对应的隐马尔科夫模型状态，组成状态序列；各状态序列持续时间由训练时得到的各状态的平均时长和方差决定。然后根据状态序列，使用训练好的隐马尔科夫模型生成相对应的声学统计参数序列。

[0056] 第二步，如图 4 所示，根据声学统计参数序列，使用训练好的循环神经网络，生成连贯的声学参数序列。具体方法包括：

[0057] 由于本实施例运行阶段第一步中使用的隐马尔科夫模型输出状态间不连贯的声学统计参数序列，应当对隐马尔科夫模型输出的声学统计参数序列做初步平滑处理，再输入循环神经网络，以确保循环神经网络能够输出连贯的语音参数序列；初步平滑处理方法可以是插值、低通滤波、平均值滤波和最大似然参数生成算法等；

[0058] 图 5 为未经平滑处理的声学统计参数序列（实线）和经过平滑处理的声学统计参数序列（虚线）的示例图；为了表示清晰，图中仅包含了语音的第二个梅尔频率倒谱系数（MFCC）参数的平均值序列；

[0059] 可选地，对输入循环神经网络的声学统计参数序列做正规化处理，例如使输入参数符合均值为 0，方差为 1 的高斯分布；

[0060] 将循环神经网络的输入层各神经元的激活值设定为提前 5 至 40 帧的声学统计参数；可选地，输入层各神经元的激活值设定为提前 5 至 40 帧的声学统计参数序列和当前帧的声学统计参数；

[0061] 根据当前帧各层神经元的激活值和前一帧各层神经元的激活值，计算循环神经网络各层上各神经元的激活值；

[0062] 将输出的声学参数设定为循环神经网络输出层各神经元的激活值；

[0063] 按时间顺序对各帧对应的声学统计参数循环执行上述步骤，生成连贯的声学参数序列。

[0064] 图 6 为经过平滑和正规化处理的声学统计参数序列（虚线）、经过平滑和正规化

处理并提前 30 帧的声学统计参数序列（点线）和循环神经网络输出的声学参数序列（实线）的示意图。为了表示清晰，图中声学统计参数序列仅包含了语音的第二个 MFCC 参数的平均值序列，且输出经过正规化处理。可以看到，循环神经网络的输出比输入的经平滑处理的平均值序列，具有更多细节。

[0065] 第三步，根据循环神经网络生成的声学参数序列，合成语音波形作为输出。具体合成方法取决于使用的声学参数类型，本发明不对合成方法作具体限定。

[0066] 本发明使用的循环神经网络，以及本实施例使用的基于决策树和隐马尔科夫模型的声学统计参数预测模型，适用于多种声学参数和声学统计参数，例如语音的梅尔频率倒谱系数（MFCC）特征、线谱对（LSP）特征、谐波能量特征、共振峰特征、频谱包络特征、基频特征、对数基频特征、及上述特征的平均值、方差和导数等。本发明不对使用的声学参数和声学统计参数类型作具体限定。

[0067] 传统的基于最大似然参数生成的统计参数语音合成方法最小化了生成参数的统计误差，但并不能保证同时最小化生成的声学参数在听觉上的特征损失。本发明使用循环神经网络，在训练过程中最小化了合成语音的声学参数和真实语音的声学参数的误差，从而有效降低了合成语音的听觉特征损失，并减轻了过度平滑现象。另一方面，本发明使用的循环神经网络在训练过程中，结合了训练语音的说话人特征，因此可以在运行阶段生成的声学参数中保留更多细节，使合成语音更加自然。

[0068] 基于深度神经网络的语音合成方法直接以文本的上下文信息作为神经网络的输入，而本发明中循环神经网络的输入为声学统计参数。由于本发明中循环神经网络的输入和输出数据关联性强，本发明所述的语音合成方法只需要 1 至 2 小时的训练语音数据，与基于深度神经网络的语音合成方法相比，降低了对训练数据的需求，减少了准备训练语音数据的工作负担。

1. 一种基于循环神经网络的语音合成方法，其特征在于包含以下步骤：
 - a. 获取待合成文本的上下文信息；
 - b. 根据文本的上下文信息，生成声学统计参数序列；
 - c. 根据由上下文生成的声学统计参数序列，使用循环神经网络生成待合成语音的声学参数序列；
 - d. 根据待合成语音的声学参数序列，合成语音。
2. 权利要求 1 所述的文本的上下文信息，是指文本所包含的一个或多个音素或音节的语境信息。文本的上下文信息至少包括以下参数之一：
 - a. 当前处理的音素或音节的名称；
 - b. 当前处理的音素类别，至少包括元音、辅音、鼻音、塞音、擦音之一；
 - c. 当前处理的元音发音位置，至少包括前、中、后之一；
 - d. 当前处理的元音发音口型大小，至少包括开、中、闭之一；
 - e. 下一个或上一个处理的音素或音节的名称；
 - f. 下一个或上一个处理的音素或音节的类别；
 - g. 下一个或上一个处理的音素或音节的发音位置；
 - h. 下一个或上一个处理的音素或音节的发音口型大小；
 - i. 当前处理的音节中包含的音素数量；
 - j. 当前处理的句子或短语或词语中包含的音素或音节数量；
 - k. 当前处理的音素在音节中的位置；
 - l. 当前处理的音节在句子或短语或词语中的位置；
 - m. 当前处理的词语的词性，至少包括动词、代词、名词、形容词、副词、介词之一。
3. 权利要求 1 中，步骤 b 生成的声学统计参数序列至少包括以下参数之一：
 - a. MFCC（梅尔频率倒谱系数）参数或梅尔频率滤波器能量参数；
 - b. LSP（线谱对）参数；
 - c. LPC（线性预测系数）参数；
 - d. 语音的谐波能量参数；
 - e. 语音的共振峰频率或共振峰带宽或共振峰能量参数；
 - f. 语音的基频参数；
 - g. 语音的频谱或频谱包络参数；
 - h. 语音的短时能量参数；
 - i. 以上参数的一阶导数或二阶导数；
 - j. 以上参数的短时平均值或短时均方差或短时最大/最小值。
4. 权利要求 1 中，步骤 c 所述的循环神经网络生成的待合成语音的声学参数序列至少包括以下参数之一：
 - a. MFCC（梅尔频率倒谱系数）参数、或梅尔频率滤波器能量参数；
 - b. LSP（线谱对）参数；
 - c. LPC（线性预测系数）参数；
 - d. 语音的谐波能量参数；
 - e. 语音的共振峰频率或共振峰带宽或共振峰能量参数；
 - f. 语音的基频参数；
 - g. 语音的频谱或频谱包络参数；
 - h. 语音的短时能量参数。
5. 权利要求 1 中，步骤 c 所述的循环神经网络至少包括以下类型的神经网络之一：
 - a. 一阶或二阶循环神经网络；

- b. 多层循环神经网络；
- c. 长短时记忆神经网络；
- d. 以上神经网络的组合；
- e. 以上神经网络和前馈神经网络的组合。

图 1

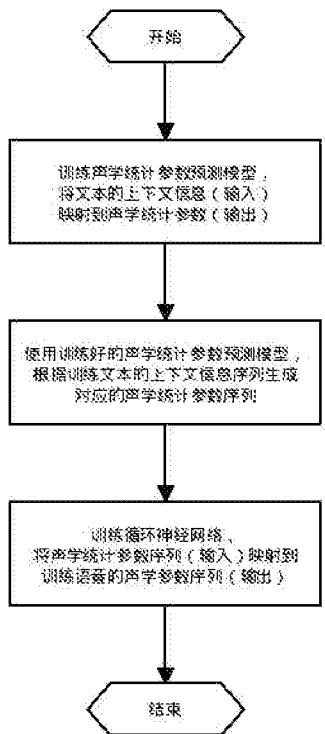


图 2

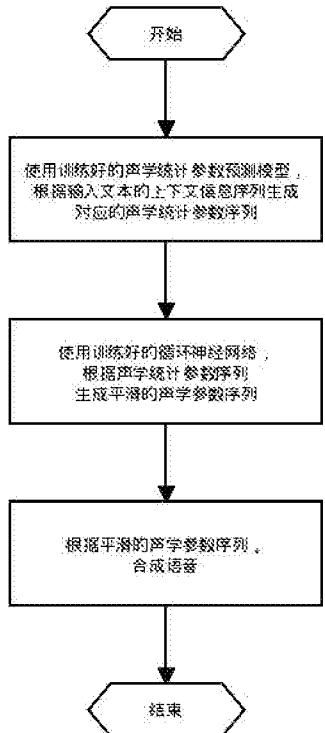


图 3

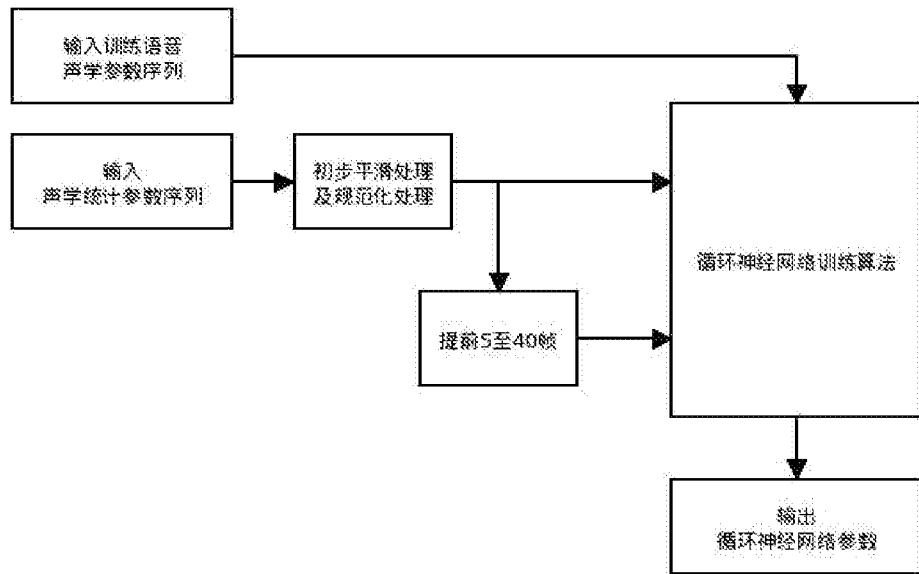


图 4

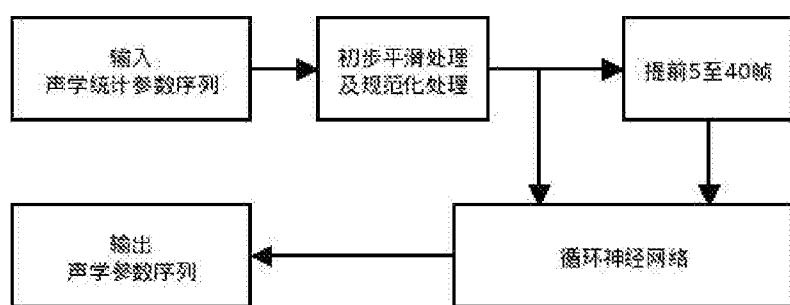


图 5

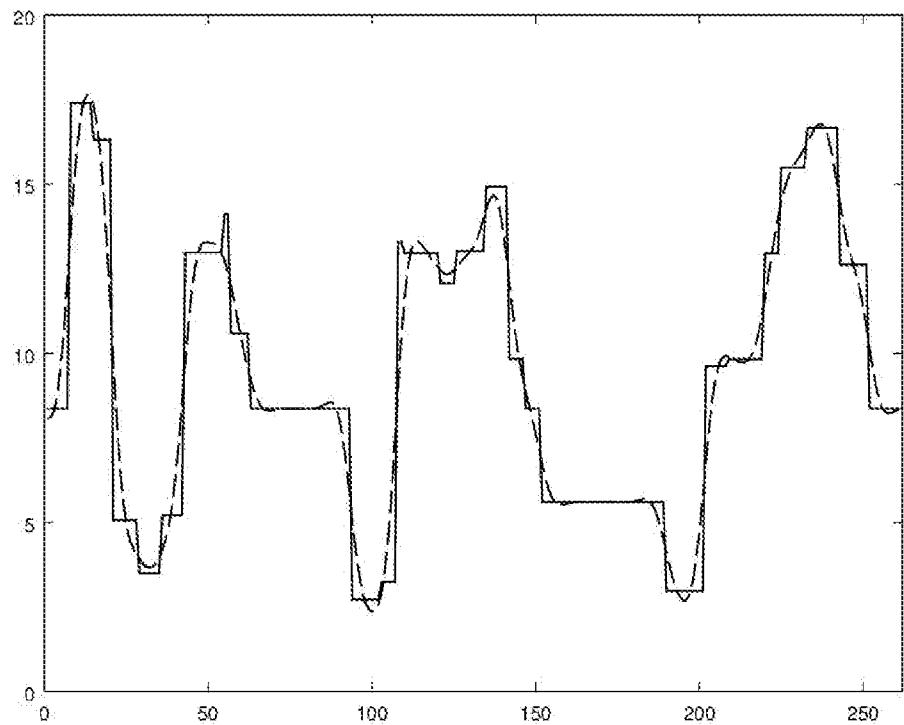
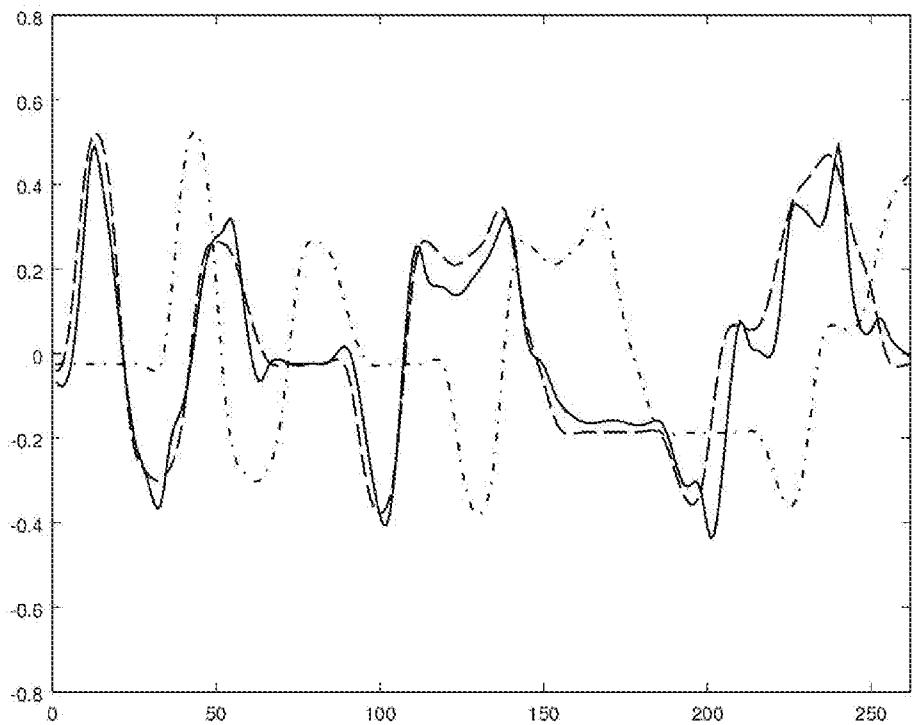


图 6



INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/077785

A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/00 (2006.01) i; G10L 13/08 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 13/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: written language, character, cycle, neural network, speech+, voice?, sound?, audio+, synthes+, text?, convert+, convers+, tts, neutral+, network?, context?.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104538024 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 22 April 2015 (22.04.2015), description, paragraphs [0009]-[0054], and figure 1	1-5
A	CN 1929655 A (SUN YAT-SEN UNIVERSITY), 14 March 2007 (14.03.2007), the whole document	1-5
A	CN 1275746 A (MOTOROLA), 06 December 2000 (06.12.2000), the whole document	1-5
A	CN 101510424 A (MENG, Zhiping), 19 August 2009 (19.08.2009), the whole document	1-5
A	CN 102117614 A (SONY ERICSSON MOBILE COMMUNICATION AB), 06 July 2011 (06.07.2011), the whole document	1-5
A	EP 1291848 A2 (NOKIA CORPORATION), 12 March 2003 (12.03.2003), the whole document	1-5

Further documents are listed in the continuation of Box C.

See patent family annex.

- * Special categories of cited documents:
- “A” document defining the general state of the art which is not considered to be of particular relevance
- “E” earlier application or patent but published on or after the international filing date
- “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- “O” document referring to an oral disclosure, use, exhibition or other means
- “P” document published prior to the international filing date but later than the priority date claimed

- “T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- “&” document member of the same patent family

Date of the actual completion of the international search
12 August 2015 (12.08.2015)

Date of mailing of the international search report
25 August 2015 (25.08.2015)

Name and mailing address of the ISA/CN:
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No.: (86-10) 62019451

Authorized officer
KANG, Dandan
Telephone No.: (86-10) **62413593**

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/CN2015/077785

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104538024 A	22 April 2015	None	
CN 1929655 A	14 March 2007	None	
CN 1275746 A	06 December 2000	AU 2104095 A FI 955608 A0 JP H08512150 A CN 1128072 A EP 0710378 A1 AU 675389 B2 CA 2161540 C WO 9530193 A1 FI 955608 A CA 2161540 A1 US 5668926 A CN 1057625 C	29 November 1995 22 November 1995 17 December 1996 31 July 1996 08 May 1996 30 January 1997 13 June 2000 09 November 1995 22 November 1995 09 November 1995 16 September 1997 18 October 2000
CN 101510424 A	19 August 2009	CN 101510424 B	04 July 2012
CN 102117614 A	06 July 2011	CN 102117614 B US 2011165912 A1 EP 2491550 A1 WO 2011083362 A1 US 8655659 B2 EP 2491550 B1	02 January 2013 07 July 2011 29 August 2012 14 July 2011 18 February 2014 06 November 2013
EP 1291848 A2	12 March 2003	US 2003050779 A1 US 7043431 B2 DE 60229095 D1 EP 1291848 B1	13 March 2003 09 May 2006 13 November 2008 01 October 2008

国际检索报告

国际申请号

PCT/CN2015/077785

A. 主题的分类

G10L 13/00 (2006. 01) i; G10L 13/08 (2013. 01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G10L 13/-

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNPAT, WPI, EPODOC, CNKI: 语音, 声音, 话音, 音频, 合成, 文语, 文字, 文本, 语音, 转换, 循环, 神经网络, 语境, 上下文, speech+, voice?, sound?, audio+, synthes+, text?, convert+, convers+, tts, neutral+, network?, context?.

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 104538024 A (百度在线网络技术北京有限公司) 2015年 4月 22日 (2015 - 04 - 22) 说明书第[0009]-[0054]段, 图1	1-5
A	CN 1929655 A (中山大学) 2007年 3月 14日 (2007 - 03 - 14) 全文	1-5
A	CN 1275746 A (摩托罗拉公司) 2000年 12月 6日 (2000 - 12 - 06) 全文	1-5
A	CN 101510424 A (孟智平) 2009年 8月 19日 (2009 - 08 - 19) 全文	1-5
A	CN 102117614 A (索尼爱立信移动通讯有限公司) 2011年 7月 6日 (2011 - 07 - 06) 全文	1-5
A	EP 1291848 A2 (NOKIA CORPORATION) 2003年 3月 12日 (2003 - 03 - 12) 全文	1-5

 其余文件在C栏的续页中列出。 见同族专利附件。

* 引用文件的具体类型:

- “A” 认为不特别相关的表示了现有技术一般状态的文件
- “E” 在国际申请日的当天或之后公布的在先申请或专利
- “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)
- “O” 涉及口头公开、使用、展览或其他方式公开的文件
- “P” 公布日先于国际申请日但迟于所要求的优先权日的文件

- “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件
- “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性
- “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性
- “&” 同族专利的文件

国际检索实际完成的日期

2015年 8月 12日

国际检索报告邮寄日期

2015年 8月 25日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
 北京市海淀区蓟门桥西土城路6号
 100088 中国

传真号 (86-10) 62019451

受权官员

康丹丹

电话号码 (86-10) 62413593

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/077785

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	104538024	A	2015年 4月 22日	无			
CN	1929655	A	2007年 3月 14日	无			
CN	1275746	A	2000年 12月 6日	AU	2104095	A	1995年 11月 29日
				FI	955608	A0	1995年 11月 22日
				JP	H08512150	A	1996年 12月 17日
				CN	1128072	A	1996年 7月 31日
				EP	0710378	A1	1996年 5月 8日
				AU	675389	B2	1997年 1月 30日
				CA	2161540	C	2000年 6月 13日
				WO	9530193	A1	1995年 11月 9日
				FI	955608	A	1995年 11月 22日
				CA	2161540	A1	1995年 11月 9日
				US	5668926	A	1997年 9月 16日
				CN	1057625	C	2000年 10月 18日
CN	101510424	A	2009年 8月 19日	CN	101510424	B	2012年 7月 4日
CN	102117614	A	2011年 7月 6日	CN	102117614	B	2013年 1月 2日
				US	2011165912	A1	2011年 7月 7日
				EP	2491550	A1	2012年 8月 29日
				WO	2011083362	A1	2011年 7月 14日
				US	8655659	B2	2014年 2月 18日
				EP	2491550	B1	2013年 11月 6日
EP	1291848	A2	2003年 3月 12日	US	2003050779	A1	2003年 3月 13日
				US	7043431	B2	2006年 5月 9日
				DE	60229095	D1	2008年 11月 13日
				EP	1291848	B1	2008年 10月 1日

表 PCT/ISA/210 (同族专利附件) (2009年7月)