

# A multi-step generate model with kinetic parameters

## 一种以动力学参数为中间值的动作生成模型

---

GES233

2023 年 2 月 6 日

目前仅有一人的 Qy 开发组

## 1. 关于

企划

## 2. 想法介绍

核心以及「理论基础」

架构与模拟流程

## 3. 原型机

still in Slides

## 4. 他山之石

可验证可行性的研究

# 关于 Qy

Qy 是从本人 (@GES233) 2018 年某条头脑风暴的想法中催生出的计划，即通过某种仿生的方式来生成活灵活现的角色的动作（包括但不限于声音、面部表情或身体上的动作）。

人话：做个会动会唱歌有表情的小姐姐。

## 怎么念？

Qy 可以拼出来读，也可以按照音标  $[k^h w I]$  /  $[k^w I]$  (qui-) 读，或者可以按照汉语拼音的 qi (注音 [ㄑㄧ]) 来念。

# 一句话

用自然的原理而复现的方法在结果上是自然的，也是可实现的。

## Cite

「一件事情如果过于复杂，那么一定是哪里出问题了——大部分情况下是对问题的理解出现偏差。」

——*Joe Armstrong*

## 4 PILLERS

- 就旁观者而言，动作的产生是**动作的被施加以及施加主体与环境的交互相互叠加**的成果。

### 说明

介绍两个比较抽象的概念：

- **动作的被施加** 要执行动作时的「指令」，例如运动神经元的放电频率。
- **施加主体与环境交互** 以「文本转语音」(TTS) 为例，就指的是声带与气流的作用。

## 4 PILLERS

- 就旁观者而言，动作的产生是**动作的被施加以及施加主体与环境的交互相互叠加**的成果。
- 对于施加主体与环境的交互过程，有时**难以全部地在计算机上得到复现。**

### 说明

算法的研究进展以及数据采集的难度，限制了计算机对主体与环境交互过程的完全的仿真。

## 4 PILLERS

- 就旁观者而言，动作的产生是**动作的被施加以及施加主体与环境的交互相互叠加**的成果。
- 对于施加主体与环境的交互过程，有时**难以全部地在计算机上得到复现。**
- 动作的被施加存在**意图与非意图**的成分。

### 说明

意识无法极其精确地控制运动本身 (e.g. 「别紧张，深呼吸」)。

## 4 PILLERS

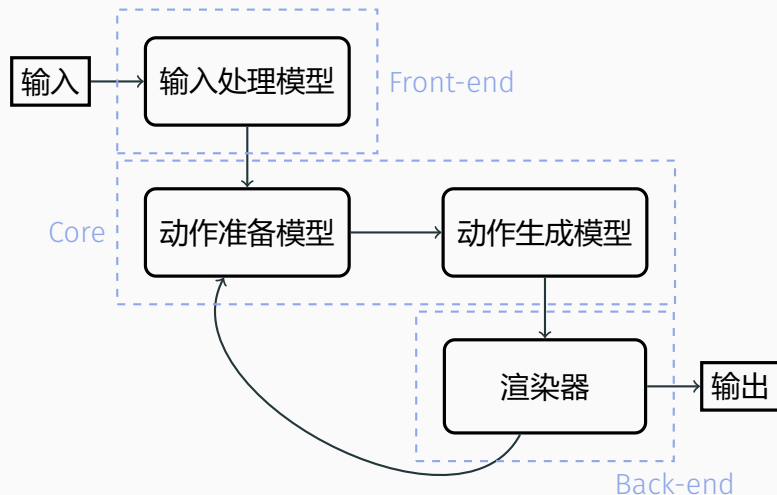
- 就旁观者而言，动作的产生是**动作的被施加**以及**施加主体与环境的交互相互叠加**的成果。
- 对于施加主体与环境的交互过程，有时**难以全部地在计算机上得到复现**。
- 动作的被施加存在**意图与非意图**的成分。
- **动作需要施加动作主体的当前状态**。

### 说明

产生动作的模型需要自身的状态作为调整的依据，也对应着第一点标红的区域。



# 概览



# 模型架构

依上面所述，模型可以分为以下部分：

- 输入处理模型

## 介绍

对输入进行归一化以及简单的预处理。

# 模型架构

依上面所述，模型可以分为以下部分：

- 输入处理模型
- 动作准备模型

## 介绍

整合输入以及情绪等，以产生类似于「意图」的一类数据。

# 模型架构

依上面所述，模型可以分为以下部分：

- 输入处理模型
- 动作准备模型
- 动作生成模型

## 介绍

将「意图」转变为一系列的动作。

# 模型架构

依上面所述，模型可以分为以下部分：

- 输入处理模型
- 动作准备模型
- 动作生成模型
- 渲染器

## 介绍

将动作从想法层面渲染出来。

# 第一步

按照上述思路实现一个能够产生声带的声音的模型。

它需要：

- 利用动物作为研究对象
- 发声器官简单，参数数目少（相对于人类）
- 没有语言处理的前端
- 不具备细腻的情感

## 关于实验动物的选择

我们考虑了哺乳动物与禽类，考虑到笔者的条件，最终确定了猪或牛。

因其发声器官比较大，便于操作，而且获取相对容易。

# 架构设计

任务：

- **输入**：动作指令
- **输出**：声带震动的波形

模型：

- **动作生成模型**：自建
- **Vocoder**：WaveNet
- 需要回馈？ **不**

# 数据采集

略。



# 语音合成与发声

## Speech synthesis from neural decoding of spoken sentences[1]

Anumanchipalli 等通过两阶段解码以**发音器官运动特征**作为中间值来将脑电活动生成语音。  
其相比于其他模型的优越表现验证了这个思路的可行性。

## Development of the ex vivo laryngeal model of phonation[2]

该研究以犬为研究对象，通过将喉离体再灌注以及刺激来实现发声。  
相关研究的思路可以帮助我们确定模型的部分接口与参数。

# SNN

## 仿生脉冲神经网络 - - 揭开生物智慧之谜 [3]

该模型模拟了一个拥有一个视觉输入和一个听觉输入、一个眼动输出和一个语音输出的简单的智能体系，经过训练可以在环境中完成特定任务。

网站链接（已失效，链接已设置通过 WebArchieve 访问）：

<http://www.spikingnn.net>



G. K. Anumanchipalli, J. Chartier, and E. F. Chang.  
**Speech synthesis from neural decoding of spoken sentences.**  
*Nature*, 568(7753):493–498, 2019.



N. S. Howard, A. H. Mendelsohn, and G. S. Berke.  
**Development of the ex vivo laryngeal model of phonation.**  
*The Laryngoscope*, 125(6):1414–1419, 2015.



spikingnn.  
**仿生脉冲神经网络 - - 揭开生物智慧之谜**, 2013.