虽然涉及环境交互的问题，数据驱动（data-driven）很难作出贡献（因为环境变化的动态性），但是作者通过聚焦于几个简单的动作，完成了数据驱动下的训练，并获得了large scalability 和 ability to produce novel motion data 的阶段性数据优势以及更高保真度，更高连贯性的动作。

虽然文章的细节实现是作者的研究结果，但其中的基础工作，也离不开一系列的奠基工作。动作交互环境的基础，就是动作的生成，而deep supervised learning的高质量生成方法，不仅契合动作生成本身，也契合着作者的工作实现；在end-to-end fashion的因子化只适用于循环动作的基础之上，作者也是提出了bi-directional scheme 即自我中心定向与目标中心定向，更好的抵御在训练过程中错误带来的积累；在heightmap-based representation 的基础之上，作者也是引出了volume的概念，通过引入volume（一个圆柱形的 environment sensor

和cubic遍布物体的interaction sensor）来记录环境，以解决算法原有的问题，使得一些处于凹面的动作得以连续生成与运行；另外，也有基于环境交互池的随机动作添加与姿态初始化控制，使得较小的数据集可以在作者提出的motion augmentation framework下不断扩展。

而上面提及的神经状态机框架，双向控制系统，体积环境标记方式，数据集增强方案便是本文的四个核心突破。

当然，一些遗留的科研问题，也成为了作者进而向前的巨大掣肘。想基于已有的template-based 技术或kernel-based 技术推进研究，但是低延展性、预计算量大以及运动空间不连续的问题直接将想法扼杀于摇篮；已知DRL算法广泛应用于人物动作的控制，但控制表现不自然的问题已知存在，即使作者只是抽象“坐”这么一个简单的日常动作，其中的reward机制依然难以敲定等等。

虽说在这次的推进中，依然沿用了MoE模型，只是在此基础之上提出了gating network和motion prediction networl两个概念，最终产生影响的，是在输入时增添了人物的姿态、轨迹，goal relative position等其他信息；而在输出中，增加了目标中心下人物的坐标以及接触点等，仅使用了两个三层神经网络，实现了动作的解离与预测。

在完成训练后，作者也是进行了一定对比，其中响应性（动作执行时间）、精度（包括旋转误差与定位误差）以及脚步追踪是三个重要的指标；对比MLP、PFNN、MANN和LSTM四个模型后，可以证明作者的模型在动作自然性上更优，对不同动作的甄别也更加敏感；更重要的是，对于不在训练数据集中的几何环境，其往往有更好，或者说更合理的交互结果。

最后成列一些个人的思考问题：首先，作者提出了一种state-of-the-art 的评价体系，但为什么说是最优的，还是有点奇怪？第二，应该如何去理解Animating Close Interaction这个名词（可能是我能力太低理解不了）？最后，对于Volumetric Sensors中的Environment Sensor 的性状，为什么圆柱形就已足够？在我的观念里，圆锥形状是不是更加贴合生活实际呢，毕竟越贴近地面，对分布的物体总体而言会更多，因而应该更需要计算资源进行检测。