AlterEgo 演讲稿

一、发布会开场

大家好。欢迎来到本公司的产品发布会现场。
看到大家能从百忙之中抽出时间来参加我们的这一个活动，我感觉非常的荣幸，同时也能感受到大家对我们此次发布会寄以厚望。

首先在正式带来这个产品之前，我想跟大家聊聊一个词，叫做 change。改变。这也是我们最近一直提起的词。都说这是一个飞速变化的时代。那么改变是怎么发生的？是什么带来了改变？是不是像我们通常认识的那样，需要细水长流的时间的累计？
而我们是不是真的像想象那样经常改变？

【打字机和现代手机键盘】
这是一百年前的打字机，而这是一部 iphone。
好像从某种程度上来说，我们的主流输入方式一直没有太大的变化。

或许由此，我们可以认为，而它的变化往往是在几个特殊的时间点上而不是一个累计的过程。我们更倾向认为，它是有想要改变、渴望改变的人创造的。

【Siri】
2011 年,Phone 4S 搭载的 Siri，或许就是这样的一个时间点。很多人第一次发现，原来还可以用语音去控制程序，更可以和一个虚拟人物对话聊天。现在我们看起来习以为常的功能，比如语音查时间、订日历等，在当时看来却是惊为天人。

很多人都预测，这项技术将极大的改变我们的生活。

但是到了现在，或许 Siri 的发展不像当初预见的那样。
Creative Strategies 2016 年的一项调查数据显示，70% 的苹果手机用户很少或偶尔使用 Siri，甚至还有 2% 左右的用户从未使用过 Siri。
在公共场合中使用 Siri 也会让用户感觉尴尬。Creative Strategies 的数据显示，98% 的 iPhone 受访用户用过 Siri，但是，仅有 3% 的人在公共场合或其他人面前使用这款语音助手。（http://www.geekpark.net/news/225111）

那么由这组数据，我们可以看到特别是在公共场合，语音交互有着极大的限制。
无法想象在自习室里，有人拿起手机“hi siri 我题不会做了帮我算一下 72.6*35”
而在菜市场，想要让它清楚的听到我在讲什么可能更是一场灾难。

这极大的限制了这样一项技术对我们的生活带来的改变。这也是我们这款产品想要改变的。

二、以前的工作

这样一个问题吸引了大量研究者的注意力。
现有的技术分为侵入式和非侵入式系统。

侵入式系统

1. Brumberg 等人在语音运动皮层中使用直接的脑植入来实现无声的语音识别，从而在有限的词汇数据集上证明了合理的准确性。他们通过将传感器放置在内部语音的发音器官里，对发音器官周围的运动进行了测量和探索。
2. Hueber 等人使用放置在舌头上的传感器来测量舌头的运动。
3. Hofe 和 Fagan 等人使用永磁体（PMA）传感器来捕捉用于讲话发音的肌肉上特定点的运动。这种方法需要永久固定磁性微珠，这在现在的世界中并不能很好地扩展
4. Florescu 等人提出使用超声波来表征声道以实现无声言语。只有在摄像机正对用户嘴巴的时候，这个系统才能取得良好的效果。

除了临床的情况之外，设备的侵入性、突兀性或者静止性妨碍了这些解决方案在实际环境中的可扩展性。

非侵入式系统

目前有许多种以非侵入的方式来检测和识别无声语言的方法。
1. Porbadnik 等人使用脑电波 EEG 传感器进行无声语音识别，但是这种方法需要以低信噪比来加大语音检测的准确度，因此它在实际环境中的表现并不好
2. Wand 等人对没有声学发声的视频进行了深度学习，但需要有外置摄像头来解码唇部运动的语言。
3. HHirahara 等人使用听不见的低音麦克风将数字转换成信号
4. Wand 和 Schultz 等人使用基于音素的声学模型演示了表面肌电图的无声语言，但是用户必须明确地说出单词并且必须有显著的脸部运动。

非侵入式系统虽然不会给用户带来不愉快的体验，但是他的准确率大大降低，很多时候需要用户作出夸张的动作来保证检测的准确性。

有一些实验考虑到脑植入，或者是在舌头上放置传感器。但是这样的操作设备的侵入极大的影响了可拓展性。此外还有使用脑电图传感器（EEG）以及拍摄视频来进行识别，但是往往准确率比较的低。

三、我们的产品

经历了这种种，所有人都有一种期待，能有一个一个集大成的方式，能够真正改变这种现状。这也就是我们公司的在这么长时间的摸索中，一直执着于的、一直想做的一件事。

此时此刻，我很高兴能够将我们这么久以来的探索呈现给大家。
这次发布会要给大家带来的产品叫做 **AlterEgo**。
Alter，alter 意味着什么，改变，转换，新的可能性。
Ego，自己，自我。

当初我们确定这个名字的时候，曾几番争论是否要采用这个看起来如此宏大的名字呢。因为不是所有的产品都敢说，他能真真切切的给用户带来改变。而等到我们这个产品真真切切的呈现在我们面前的时候，我们都觉得它当之无愧。

"AlterEgo is a closed-loop, non-invasive, wearable system that allows humans to converse in high-bandwidth natural language with machines, artificial intelligence assistants, services, and other people without any voice—without opening their mouth, and without externally observable movements—simply by vocalizing internally."

AlterEgo 是一种闭环，非侵入式，可穿戴的系统，它允许人们用机器，人工智能助理，服务和其他人没有任何声音的高带宽自然语言进行交谈 - 不需要张口，也不需要外部可观察运动 - 只需通过内部发声就可以实现交互。

或许这么说过于官方了，更加直观的来说，不需要手术，不用担心影响他人，也不用担心说说的话被听到。只要戴上这个设备，就可以在不张嘴的情况下下达指令。

或许还是有点难以想象。让我们来通过下面这个视频直观的了解一下。
（PPT 放视频）

四、背后的技术

为了这个目的，技术部做出了非常多的努力。

首先，我们的设备捕捉的是脸部的肌电信号，让我们温故一下该信号产生的生物学知识吧。

1.一开始，我们的大脑想要表达某种意思。这种含蓄而隐晦的思想经过了大脑的布罗卡氏区——相信大家没有忘记高中生物课里讲到的这个，主管语言讯息的处理、话语的产生的区域，它把这种冲动组织为语言。
2.接着，组织好的语言经过运动辅助区：它主要参与运动控制。
3.这个控制发声相关肌肉的运动的命令变成了神经冲动，经过各种神经的传导，来到突触。
4.在突触，乙酰胆碱释放，并和受体结合，释放 Na+离子，导致膜电位的改变。
这个动作电位很快到了肌纤维，肌肉忍不住一颤。

我们的产品中，捕捉信号的电极分布在不对称的七个位置，都是根据我们的经验，得出的效果较好的肌肉部位。（人体肌肉图示）

当然我们取的点也有人体肌肉结构的一些依据。电极都分布在和发音最关的几个肌肉区。

五、工作流程

接下来详细讲讲我们的设备的工作流程。

1.用户佩戴好外设

我们使用光敏树脂进行 3D 打印，并用黄铜作为内部支撑，做了一个外部设备。
它既有形变能力，又有一定强度，贴合用户嘴部的肌电信号采集点。

2.收集肌肉电信号

用户不需要发声，甚至不需要张嘴，只要你有说话的冲动，设备就能捕捉肌电信号。

3.电信号预处理

去噪：基于偏置的信号对消。
信号按照 250 Hz 的频率采样并且放大 24 倍。

4.信号的传播

信号通过蓝牙无线传输，你可以不必受限于电线的长度了。

5.计算

计算机接收信号；
服务器处理信号；
APP 计算数据。

6.反馈

骨传导耳机。

通过这样一个工作机制，来实现交互。
那么其中信号捕捉的模块是通过大脑到 balabala
其中电极的位置是根据大量实验数据得出的较好的位置。

六、实验结果

或许有人会担心实验的识别的准确率问题。
我们在测试的时候，10 名 19-31 岁对此系统没有经验的用户进行测试，结果能达到平均准确率 92.01%，平均延迟 0.427s。

实验结果表明：无声语音系统的准确率与最新的语言识别系统的准确度相当，足够部署为语音接口。

而这也就意味着，在大多数情况下我们可以毫无压力的实现信息的交互。
像之前在视频上的那样，实现棋类游戏，无声通讯以及作为物联网控制器。

我深切的希望此时此刻也是一个特殊的时间点，这款产品如我们期待的那样，能够带来一种新的交互方式，能够带来新的震撼和感动。

七、总结和结束

相信到现在为止，AlterEgo 所出现的必要性，它所能为我们生活做出的改变已经讲的比较清楚了。接下来将脱离发布会这个框架，从一些其他的角度来完善对这项产品的认识。
虽然我们用的是发布会的形式，但我们也都知道这个研究走出实验室还需要一定的积淀。

首先，小组在讨论这个问题的时候，考虑的最多的其实是美观性。当然也不排除有一天，人们放弃了肉体靠机械来维持心脏的跳动，但是至少在目前的情况下，这对于这项技术作为产品的推广还是有一定的阻碍。

其次，或许的介绍会让大家觉得它能够实现像 Siri 一样的交互过程，但其实距离这个还有很大的一段距离。这篇 papper 测试的数据集其实是相对较小的，非常担心在更大的数据集下的处理情况。很多情况下，对于某些特例的处理结果是没有普遍的适应性的，之前提到的确定点，也是对于这个语素集有比较好的区分度。

还有就是从这篇论文来看，他是由每个人单独训练之后做测试的，不知道这是不是在某些方面说明没有一个普适的模型。如果用户使用的时候还需要大量的初始数据，那么对于这项产品的推广是非常不利的。

1.收集更多数据以开发更一般化的多用户无声语音识别模型
目标是开发一种独立于用户的广义多用户系统，但也可以在每位用户开始使用设备时对其进行调整和个性化设置。
2.扩展系统以包含更广泛的单词词汇
当前的实例中实现了对多个词汇集的可访问性，尽管数据有限。实验评估基于算术计算应用。计划增强识别模型以适应更大的数据集，并计划在系统中进行彻底的多用户纵向精度测试。
3.在真实的动态环境中测试系统
现有的研究是在固定的环境下进行的。
将来希望在日常情况下进行纵向可用性测试。