人机交互智能语音控制系统

语音控制系统用来更好地理解人类语音指令（包括识别和测量），并将其转换为控制输入。整体系统图如图1所示。人类视觉系统观察到控制错误，并输入观察到的错误进入大脑，之后，大脑会控制人类语音系统表达控制意图作为语音控制系统的输入。语音控制系统包括三个主要部分，第一部分是一个语音识别系统，它执行定性分析，即识别语音目的内容;第二部分是语音测量系统，进行定量分析，即测量语音主观内容;第三部分是控制系统把语音系统的融合结果转换成控制输入。语音识别和测量系统统一起称作语音系统，以下部分将详细说明这三部分。语音控制系统可以检测大脑发出由人类语音系统表达的控制意图，并将语音信号转换为控制输入（见图2）。考虑到机器人作为一个执行器，大脑作为控制器基于视觉引导的相对位置的机器人和目标提供初始控制输入，之后，在接近目标的过程中，大脑会自适应的调整语音输入，使机器人可以快速，顺利地理解目标。语音控制系统的目的是为了准确检测语音输入并将其转换为控制输入。

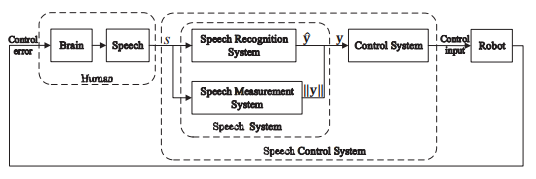


图1 语音控制系统图

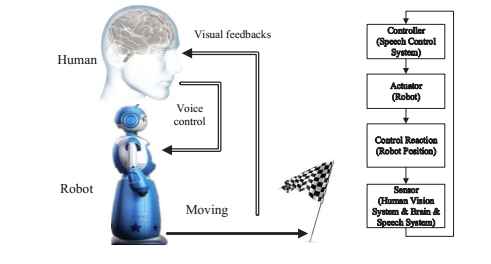


图2 语音控制系统工作原理

**语音识别系统**

假设 u = [u1, ..., un] 表示机器人的控制输入，其中n是控制输入的维度，ui表示u的第i个维度。每两个孤立的语音命令被用来映射成一个控制输入的维度为Y = {y1 +, y1 -, ..., yn +, yn -} ，其中yi +和y1 -对应于ui正和负方向。 因此，为了用控制输入的n个维度控制系统，这将产生2n个需要识别的语音指令类。

1. 定性特征提取

在孤立的命令识别系统中，提取的特征必须消除环境和人类主观因素（包括情绪和健康状况）等的影响，只能保留客观内容。 提取声信号的最佳参数表示是产生更好的识别性能的重要任务。这一阶段的准确性对于下一阶段至关重要，因为它直接决定了控制输入的符号。 这里，MFCC被应用于表示声输入[5]。 MFCC的整体过程如图3中标记为黑色的部分。

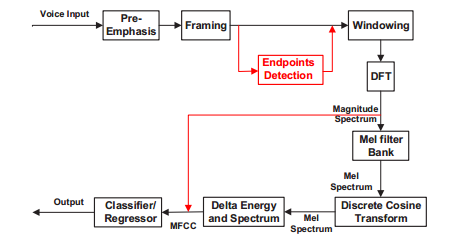


图3 MFCC和提出的基于频谱的特征的比较

B.模板匹配的语音识别

在本节中，将构建映射函数f1：s→y^，其中s是音频信号，y^∈Rn是单位向量。将yi表示为y的第i维，如果s∈yi +，则yi = 1;如果s∈yi-，则yi = -1;如果i= i，yi=0。此外，将u0定义为表示从语音输入转换的控制值的正值，如果yi（j）= 1（在时间ji），则ui = u0；如果yi（j）= -1，则ui = -u0。值得注意的是，在语音输入的时刻只有一个yi =±1，i = 1，...，n。在下一节中，我们将讨论如何获取||y||，这是y的度量。

DTW算法基于动态编程技术。 该算法旨在测量可能在时间或速度上变化的两个时间序列之间的相似性。 在本文中，我们应用DTW计算MFCC系列的距离向量，用于测量音频信号之间的相似度，以及欧几里得距离来计算音频特征序列的距离向量。 在获得训练模板的当前语音特征序列和每个语音特征序列之间的距离之后，将具有最大相似性的训练模板中的目标对象分配给测试样本。

**语音测量系统**

语音中有两种广泛的信息类型。 讲话的语义部分包含客观信息，只要按照语言发音规则进行发音。 另一方面，主观信息指的是隐含的消息，例如说话者的情绪状态或控制意图，这将在本节中进行研究。

1. 定量特征提取

可以精确表示控制意图的一组特征对于量化结果的准确性非常重要。 两个可能的候选特征如下研究。

1. 基于能量的特征

音量表示语音强度，可以表示为每帧中信号的幅度。 这是表达控制意图的最直接的方式。 提取能量特征的基本步骤如下。

步骤1：框架

将从模数转换获得的语音样本分割成长度在20到40毫秒范围内的小帧的过程，语音信号被划分为N个样本的帧，相邻帧被M（M < N）。 在本文中，我们选择M = 128和N = 256。

步骤2：端点检测

语音信号的端点检测是直接影响定量结果准确性的一个步骤。 这里，应用了使用短期能量和短期平均过零率的双阈值语音端点检测算法。

步骤3：平均活跃能量

两个端点之间的平均能量将被计算为控制意图强度的量化结果，以进一步映射为机器人的控制输入。

这种方法的存在几个缺点：（i）并不是所有人通过改变体量来表达自己的控制意图; （ii）声源必须固定; （iii）它对背景噪音敏感。 然而，优点是（i）不需要模型训练过程; （ii）虽然变化的数量可能不是表达控制意图的自然方式，但在用户知道这一规则的条件下，它是最直接和容易控制的方式。

1. 基于频谱的功能

步骤1：框架（与上述相同）

步骤2：端点检测（与上述相同）

步骤3：汉明窗

汉明窗方程给出为Y（n）= X（n）×W（n），其中Y（n）是输出信号，X（n）是输入信号，W（n）是窗口定义为 W（n）= 0.54-0.46cos（2πn/（N-1）），0≤n≤N-1。

步骤4：快速傅里叶变换

Y（ω）= FFT [H（n）\* X（n）] = H（ω）\* X（ω）

 其中X（ω），H（ω）和Y（ω）分别是X（n），H（n）和Y（n）的傅立叶变换。

提取基于频谱的特征的总体进展如图3标记为红色部分。与通常使用的光谱特征（例如LPCC，MFCC和LFPC等）相比，光谱的功率被直接作为后续分类器或回归器的输入，而不是使用一组给定滤波器进行映射，并且采用DCT获得功率时域序列。这种改进的动机是，不同的用户可以强调不同的频率来表达他们的控制意图，并且随后的回归者能够学习这种模式。

基于频谱特征的优缺点与能量特征几乎相反。优点：（i）它对背景噪声是鲁棒的; （ii）能够学习人们表达控制意图的自然模式; （iii）声源可以是移动的。缺点：（i）需要培训过程，即应提供大量的培训数据; （ii）训练目标是人为给予的，因此校准误差将是另一个控制误差源; （iii）虽然这种方法可以学习一种自然的控制模式，但是这种模式可能会适应演讲者的情感，环境和时间，很难被刻意表达。

B.语音测量

在这个部分，将要构建地图函数f2：s→||y||。用来在提取一组可能代表用户控制意图强度的功能之后进行决定，如何将它们映射到控制量级。对于作为标量的能量基特征，仅需要常数系数将所提取的语音特征线性地映射到控制输入中。对于作为特征向量的基于频谱的特征，需要回归步骤。

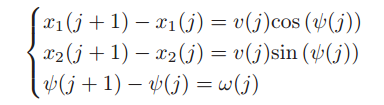
因此，对于基于频谱的特征，这里应用了一个名为Random Forest的回归函数。 它是一种使用决策树作为基本分类器的集体分类。 选择RF作为回归因子的几个主要原因：（i）高速度（ii）高精度（iii）评估每个特征变量（iv）能够处理高维特征向量的重要性的能力，这意味着 不需要功能选择。特别是，高计算速度对于保证语音控制系统的实时性能至关重要。

**控制系统**

本节将介绍几种控制方案，使语音系统的定量结果能够转化为持续的控制输入。

1. 系统动力学

考虑线性系统动力学：



其中x1和x2表示位置坐标，ψ表示运动方向。 将控制输入表示为u（j）= v（j）ω（j）\_x0007\_T。

在本文中，由于环境被假设为机器人未知，机器人无法直接测量控制误差， 但人可以观察到。这意味着机器人只能通过音频传感器的输出来感测环境。 这个假设在现实中可能太不可思议了 然而，它是激励在实际应用中用作辅助校正系统的基于语音的HRI系统的开发。

B.控制器设计

实现这一任务的一个常见方法是，每次有演讲指令，机器人转动一个固定的角度，可以表示为：

ω(j) = c

其中c是正常数。 问题是小角度c可能导致较慢的控制速度; 相反，如果c太大，则可能非常难以准确地达到所需的方向。

或者，为了使机器人更好地了解人的命令，在控制错误的条件下，控制规则将被设计用于语音控制系统，只能由人类，而机器人可以通过人类传递的语音信号间接测量控制误差。 表示：

f : s → **y**

作为语音系统的融合输出，包括语音识别系统（f1：s→y^）和语音测量系统（f2：s→||y||）。

将环境噪声和系统噪声综合考虑为ξ，并将||y||写为y^。语音测量系统的估计进度可以被标记为

y ^= f2（s）+ξ。 在本节中，噪声被忽略，ξ= 0。进度函数可以重写为y^ = f2（s）。从观察到的控制误差到控制意图的人脑图功能被定义为f3：ψe→y，这实际上是未知的。本节的主要目的是对f3进行近似，y^是y的估计。 我们假设y^ = y在这里，这意味着言语系统的结果被假定为等于人的控制意图。

同时，我们假设ψe和y之间的线性解耦关系，其将时间j的映射函数y = f3（ψe（j））简化为y（j）= a \*ψe（j） 其中a \*是未知的正常数。 我们将a \*的倒数表示为g \*，即g \* = a \* -1。 类似地，g \*是未知的正常数。 那么我们有ψe（j）= g \* y（j）。

人类自适应控制系统。控制规则ω（j）被选择为：

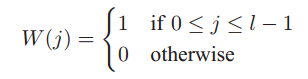
ω(j) = kyy(j)

其中k y> 0是常数。 在这种方法中，给出恒定系数k y，以将测量结果映射到控制方向误差，那么人脑会经过几次尝试后逐渐找出这个系数。 因此，我们称之为“人为自适应控制”。

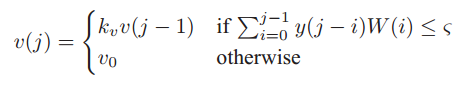
C.加速策略

速度v（j）的小值可能有利于方向调整，而较大的v（j）值可能有益于减少任务时间成本。

为了平衡两个因素，在开始时设置相对较小的v0值，并且引入长度为l的窗函数W（j），



那么加速策略设计为:



其中kv> 1是加速度的常数集，ζ是非常小的正值。 我们不将其设置为零的原因是为了避免噪音的不适。