电视背景环境下语音命令识别系统

范婷[1] 刘宏[2]

（西安电子科技大学 电子工程学院，陕西 西安 710126；
北京大学深圳研究生院 信息工程学院，广东 深圳 518055）

摘要：本文设计了一种电视背景环境下的非特定人语音控制命令识别系统，其中包括基于隐马尔可夫模型的孤立词识别子系统和基于扩展 Infomax 独立成分分析算法的语音分离子系统. 本语音识别系统的语音库包括 8400 个电视机转台控制命令的语音数据. 在无噪环境下对特定语音命令的识别率可达 93.2%，正常电视背景环境下的识别率降至 49.0%，对电视背景下通过分离后的语音命令识别率可达 85.8%.

关键词：语音识别； 隐马尔可夫模型； 电视背景； 独立成分分析； 扩展 Infomax

中图分类号：TP912.34 **文献标识码**：A **文章编号**：1006-7043 (2006) xx-xxxx-x

A recognition system of voice command with TV background

Ting Fan[1] Hong Liu[2]

(Electronical engineering department, Xidian university ,710126, China; Information engineering department, Peking University , 518055,China)

Abstract: A speaker-independent speech recognition system of voice command with TV background was designed in this paper and the system includes a recognition subsystem of isolated words based on Hidden Markov Model and a speech separation subsystem based on extended Infomax independent component analysis algorithm. The voicebox of this system includes 8400 speech signals of TV channel controlling command. The recognition rate of specific voice command in quiet environment can reach 93.2%, the rate under normal TV background decreased to 49.0%, the rate of the voice command after separated from the TV noise is 85.8%.

Keywords: speech recognition; Hidden Markov Model; TV background; independent component analysis; extended Infomax

用语音命令直接控制电视机转台不同于传统的遥控转台，它能够实现人和电视机的直接交互，方便而快捷. 然而要实现这一应用还面临着诸多问题：1、电视背景声音错综复杂，在此环境下的识别性能较差；2、不同说话人之间以及同一说话人在不同时刻的特定语音命令存在差；3、环境噪声也会影响系统的识别性能. 因此，本文主要针对电视背景下识别性能较差这一问题设计了一种基于 HMM[1]的非特定人孤立词识别系统，研究了语音识别系统在电视背景环境下对特定语音命令的识别性能. 结合扩展 Infomax[2]的独立成分分析(independent component analysis，ICA)[3]算法可以较好地分离出语音命令以提高电视背景环境下的识别性能.

1 基于 HMM 的语音识别系统

1.1 语音识别的基本原理

语音识别[4]主要采用了模式匹配的原理，一般包括训练和识别两个部分.训练阶段的主要任务是建立语音的声学模型和语言模型并构成一个参考模式库. 识别阶段根据特定的识别方法求出未知语音的特征参数，按照一定的判别准则与参考模式库中的模式逐一匹配，通过判决将最佳匹配的参考模式所对应的语音作为识别结果.

1.2 基于 HMM 的语音识别

一个基于 HMM[5]的语音识别系统主要包括三个部分：预处理、特征参数提取和训练识别.

对于输入的语音信号，首先要进行预处理，包括预滤波、数字化、预加重、分帧加窗以及端点检测等几个环节. 这一部分的主要目的是滤去语音信号中的无用信息，提升高频部分以便于频谱分析或声道参数分析并从背景噪声中找出语音的起止点，得到有效的语音部分为之后的工作做准备. 本识别系统采用短时平均过零率与短时能量相结合的双门限端点检测法来确定语音的起止位置.

一个语音信号可分为静音段、过渡段、语音段和

收稿日期 2011 年 6 月 25 日
作者简介 范婷（1990-），女，在读硕士，**E-mail**:fanting19900126@126.com
刘宏（1967-），男，博士，**E-mail**:liuh@szpku.edu.cn
基金项目 国家自然科学基金(No.60875050)，广东省自然科学基金(NO.9151806001000025)，
深圳市科技计划及基础研究项目(JC200903160369A)

结束段. 为了确定语音已开始，该方法设了一个较高的能量门限 T_h，再取一个比 T_h 稍低的能量门限 T_l 用以确定语音信号真正的起止点. 此外，还采用了另一个较低的过零率门限 Z_l 以判断清音和无话的差别. 将上述两种门限结合以更加准确地确定语音信号的起止位置.

特征参数[6]提取是语音识别的关键问题. Mel 频率尺度大致是实际频率的对数关系，更符合人耳的听觉特性，且 Mel 频率倒谱系数（mel-frequency cepstral coefficients, MFCC）参数无任何前提假设，在各种情况下均可使用，抗噪声能力也较强，因此本文选用 MFCC 参数作为特征参数.

训练阶段首先要通过训练为词表中的每个孤立词分别建立一个 HMM[7]，即每一个孤立词可以用一个 HMM 加以描述. 识别过程中待测孤立词语音通过预处理和 MFCC 特征参数提取后，得到一个可以反映该语音的特征向量序列. 将该特征向量与模型库中的所有模型逐一匹配，计算出在每个 HMM 上的输出概率，最大的即为识别结果.

2 基于扩展 Infomax 的 ICA 算法

2.1 独立成分分析（ICA）原理

独立成分分析[8]的基本思想是通过获得的多个观测信号，寻求一个线性变换矩阵，按统计独立的原则使得变换后的输出分量尽可能相互统计独立.

设 $\boldsymbol{X}=[\boldsymbol{x}_1(t),...,\boldsymbol{x}_m(t)]$ 是 m 维统计独立源信号矢量，$\boldsymbol{Y}=[\boldsymbol{y}_1(t),...,\boldsymbol{y}_m(t)]$ 是 n 维观测信号矢量，$\boldsymbol{A}$ 为 $n\times m$ 维混合矩阵，$\boldsymbol{N}$ 为噪声矢量. 则 $\boldsymbol{Y}=\boldsymbol{AX}+\boldsymbol{N}$. 独立成分分析的目的就是在混合矩阵 $\boldsymbol{A}$ 未知的情况下寻求一个解混矩阵 $\boldsymbol{H}$，使得 $\boldsymbol{X}'=\boldsymbol{HY}$，其中 $\boldsymbol{X}'$是对矢量 $\boldsymbol{X}$ 的估计.

2.2 扩展 Infomax 算法

信息最大化(Infomax)算法就是利用信息最大化原理有效地解决 ICA 问题. 用解混矩阵 $\boldsymbol{H}$ 估计出源信号矢量 $\boldsymbol{X}'$后，引入了一个单调可逆的非线性函数 $g(\bullet)$，得到一个新的矢量 $\boldsymbol{U}$. $g(\bullet)$可将一实数映射到区间[0，1]，且为单调升函数. $\boldsymbol{X}$ 各分量的独立最大化相当于 $\boldsymbol{U}$ 的熵最大化. 通过最大熵判据来调节解混矩阵 $\boldsymbol{H}$，使得 $\boldsymbol{X}$ 各分量之间统计独立. 由于原信号的分布是未知的，Bell 和 Sejnowskide 的 Infomax 算法中选择非线性函数 $g(\bullet)$ 为固定的 *Sigmoid* 函数. 该函数的微分所表示的概率分布函数是超高斯的，所以该算法只能分离超高斯信号的混合信号.

针对 Infomax 算法只能分离超高斯信号的混合，不适用于亚高斯信号的问题，Lee 等人提出了扩展 Infomax 算法[9]. 该算法选用 *Tanh* 作为非线性函数，用一个对角矩阵 $\boldsymbol{K}$ 来区分超高斯信号和亚高斯信号，通过调整解混矩阵 $\boldsymbol{H}$ 使得扩展 Infomax 算法同时适用于超高斯信号和亚高斯信号.

2.3 分离原理

特定的语音命令和电视背景[10]的混合可以认为是瞬时线性混合，而扩展 Infomax 算法的应用前提就是各独立分量是线性混合的. 因此，针对本文要处理的混合信号，这种基于扩展 Infomax 算法的 ICA 是可行的. 此外，特定的语音命令是超高斯信号，而电视背景声音错综复杂，既可能是超高斯信号，也可能是亚高斯信号. 而扩展 Infomax 算法具有处理超高斯信号和亚高斯信号的能力，因此应用这一算法可以将混有电视背景声音的特定语音命令有效地分离出来.

图 1 为利用扩展 Infomax 算法对语音命令“北京台”和正常电视背景的混合信号进行分离前后的对比图. 四幅图分别为麦克风 1 接收到的混合信号，麦克风 2 接收到的混合信号，分离后的特定语音命令信号和电视背景信号.

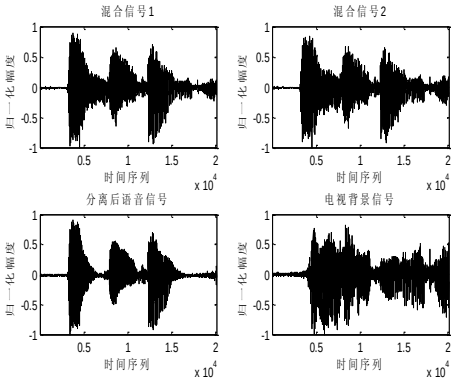

图 1 分离前后对比图

电视背景下的语音控制命令识别系统主要有分离和识别两个子系统. 两路混合信号进入分离子系统，通过解混矩阵后生成两路分离信号：语音控制命令和电视背景. 接着和与之同时录制的第三路电视信号做相关运算选出语音控制命令，该语音命令进入识别子系统，与 HMM[11,12]模板库进行匹配从

而得出最终的识别结果.

3 实验设计及测试结果

3.1 实验平台

本实验的硬件平台为 UHF Professional Wireless Microphone TS-8028，PC 机: Q6600 处理器、4G 内存、 Marian Trace8 声卡. 软件平台为 Matlab 7. 6. 0 及 Voicebox 语音信号处理工具箱.

3.2 语音库

本语音识别系统的词汇表包含 84 个孤立词，词表内容为各省级电视台，如北京台，上海台，深圳台等. 录音人 20 名，其中女性 8 名，男性 12 名. 每人每词 5 遍，共 8400 个音频数据.

3.3 实验设计及结果

本实验采用交叉验证的方法，选择 19 人的 8150 个数据进行训练，用另 1 人的 250 个数据进行测试，循环进行 20 次，取均值作为最终识别率. 其中测试数据包括无噪环境下、电视背景环境下以及分离后的音频数据各 8150 个.

表 1 为无噪环境下、正常电视背景环境下以及分离后音频数据的识别结果. 电视背景信号对不同说话人的识别性能影响不同，在被测的 20 个说话人中，测试者 15 受电视背景的影响最小，在电视背景下的识别率较其在无噪环境下下降最小，测试者 9 受电视背景影响最大。

表 1 混合信号分离前后的识别率

识别率	*测试者 15*	*测试者 9*	*平均*
无噪环境	98.4%	88.4%	93. 2%
电视背景	96.0%	11.6%	49. 0%
分离之后	87.6%	80.8%	85. 8%

本文基于语音识别的新应用: 利用语音命令控制电视机转台，设计了一种基于 HMM 的孤立词识别系统，研究了电视背景环境下对特定语音命令的识别性能. 采用扩展 Infomax 算法将语音命令从电视背景中有效分离出来. 结果表明该系统在电视背景下对特定语音命令的识别率相对于无噪环境有明显下降，由 93.2%下降至 49.0%. 通过扩展 Infomax 的 ICA 算法分离后的识别率提高至 85.8%.

参考文献:

[1] Gales M, Young SJ. The Application of Hidden Markov Models in Speech Recognition[J]. Foundations and Trends in Signal Processing, 2008 , 1(3):195-304.

[2] Bell A, Sejnowski T. An information maximization approach to blind separation and blind deconvolution[J]. Neural Computation, 1995, 7(6): 1129-1159.

[3] Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis[J]. IEEE Trans on Network, 1999, 10(3):626-634.

[4] 赵力. 语音信号处理[M]. 北京:机械工业出版社，2003，32-110.

[5] 韩纪磊，张磊，郑轶然. 语音信号处理. 北京: 清华大学出版社，2004,48-80.

[6] Lee C.H.. On robust Linear Prediction of Speech[J], IEEE Trans on ASSP,1988, 36(5): 642-650.

[7] 何强，何英. Matlab 扩展编程[M]. 北京:清华大学出版社，2002，1-24.

[8] Common P. Independent component analysis. a new concept[J]. Signal Processing, 1994, 36(3): 287-314.

[9] Lee T. Independent component analysis using an extended Infomax algorithm for mixed sub gaussian and super gaussian source[J]. Neural Computation, 1999, 11(2):409-433.

[10] Zhao Li. Study on the Chinese Continuous Speech Recognition under Noise Environments Based on the PCANN/HMM[J]. IEEE International Conference on Signal Processing, 2002:896-898.

[11] Luo X, Jelinek F. Probabilistic Classification of HMM States for Large Vocabulary[J]. In Proc. ICASSP, Phoenix, Arizona, USA, 1999: 2044–2047.

[12] 易克初. 语音信号处理[M]. 北京:国防工业出版社，2000，57-72.