

doi:10.3969/j.issn.1001-893x.2014.01.002

引用格式:陈晨,胡贝贝.基于语音识别技术的机载短波应急通信[J].电讯技术,2014,54(1):6-11.[CHEN Chen,HU Bei-bei.Airborne HF Emergency Communication Based on Speech Recognition[J].Telecommunication Engineering,2014,54(1):6-11.]

基于语音识别技术的机载短波应急通信*

陈晨^{1,**},胡贝贝²

(1.空军装备部重点型号部,北京100843;2.空军驻714厂军事代表室,南京210002)

摘要:通过分析作为机载远程通信主要手段的短波电台在强干扰及超远距离等极限条件下面临的失效风险,针对性地设计了基于语音识别技术的机载短波应急通信方案。该方案通过语音识别技术将语音信号转换为语义信息进行传输。采用提出的嵌入式命令词语音识别技术、说话人自适应技术和噪声鲁棒性语音识别技术提高语音识别的性能,保证基于语音识别技术的机载短波应急通信方法的实现,从而可以在上述情况下为飞行员提供应急通信保障。

关键词:机载通信;短波应急通信;语音识别;抗干扰

中图分类号:TN928 **文献标志码:**A **文章编号:**1001-893X(2014)01-0006-06

Airborne HF Emergency Communication Based on Speech Recognition

CHEN Chen¹,HU Bei-bei²

(1. Equipment Department, Air Force of the PLA, Beijing 100843, China;

2. Military Representative Office of Air Force in Factory 714, Nanjing 210002, China)

Abstract: The high frequency (HF) communication is the primary means of airborne remote distance communication. Through analyzing the failure risk of HF radio under strong interference and remote distance conditions, an airborne high frequency emergency communication scheme based on speech recognition is designed. The speech signal is recognized and the semantic information is transmitted in this scheme. In order to insure the implementation of the given emergency communication scheme, embedded speech recognition system, speaker adaptation techniques and noise robust speech recognition methods are adopted to improve the accuracy. So emergency communication support for pilots can be ensured in aforementioned conditions.

Key words: airborne communication; HF emergency communication; speech recognition; anti-jamming

1 引言

短波通信是我国战机的重要通信手段。但是受发射功率、天线增益等因素的影响,远距离通信能力的不足是对传统短波通信体制的一个重大挑战。同时,短波通信频道较窄、信道拥挤,容易受敌方和非敌方干扰,在战时复杂电磁环境下,短波通信的可靠性难以得到保障。因而,如何为机载短波通信系统提供应急通信能力成为了行业内的研究热点。

专家学者针对上面提出的问题开展了广泛的研

究,文献[1]提出了一种以小卫星为中继的远程通信方案;文献[2]介绍了短波通信的抗干扰方法和应用,以及短波通信抗干扰技术的发展趋势。本文从分析目前短波通信系统存在的问题入手,通过使用语音识别技术,设计出基于语音识别技术的机载短波应急通信方法,并细分析了实现该方法的关键技术。该方法对短波信道的要求低,从而可以极大地提高短波通信的通信距离和抗干扰能力。

* 收稿日期:2013-12-13;修回日期:2014-01-17 Received date:2013-12-13;Revised date:2014-01-17

** 通讯作者:43614191@qq.com Corresponding author:43614191@qq.com

2 现有短波通信的不足

2.1 短波通信系统远程通信能力缺乏

我国在南海最南端的领土曾母暗沙距离我国大陆约两千公里,菲律宾最南端的岛屿棉兰老岛距离我国大陆约 2 000 km。我国目前的三代战机的作战半径可以达到 1 500 ~ 2 000 km 以上,但是作为这些机型远程通信唯一手段的短波电台的通信距离已经不能覆盖载机的作战半径,当前正在研发的新一代战机对短波电台的通信距离要求成倍提高,但同时其中的一些关键指标如发射功率、天线增益等并未提高甚至受飞机气动外形影响有所降低,在这种条件下要保证飞机的远距通信能力,传统的通信体制面临很大的挑战。

2.2 短波通信系统抗干扰能力不足

由于短波通信频段较窄、信道拥挤,受环境影响严重,通信存在质量较差、不稳定、干扰严重等不利因素,在平时的正常使用中短波电台的通信难以百分之百联通,对于战时电子战带来的复杂电磁环境,短波通信的可用性将更加难以保证。对于跳频抗干扰,由于现役机载短波电台跳速过低,在每一频点驻留时间过长,而干扰设备从侦察频率到引导干扰的反应时间越来越短,同时由于短波信道拥挤的状况和战时敌我双方大量的电子战装备对短波信道的干扰,几乎不可能找到没有干扰的信道,现有的跳频抗干扰方式将很难发挥作用。以美国、韩国、我国台湾地区等普遍装备的典型车载短波通信干扰设备 AN/TLQ-17(20 世纪 90 年代海湾战争时产品)为基准进行计算,我现役机载短波电台的通信能力难以得到有效保证。表 1 是美军几种典型的通信干扰装备及其指标。

表 1 美军几种典型的通信干扰装备
Table 1 Typical communication jamming equipment of US military forces

型号	频段/MHz	功率/W	备注
AN/TLQ-15	1.5 ~ 20	2 000	车载
AN/TLQ-17	1.5 ~ 80	500	车载
AN/MLQ-34	20 ~ 200	400	车载
AN/GLQ-3B	20 ~ 230	2 000	车载
CHIEF	20 ~ 2 500	2 000	车载
AN/USQ-146	20 ~ 2 500	2 000	机载,装备于 EA-6B 舰载电子干扰机,一艘航母配备 4 架,多次参加黄海军演

综上所述,作为机载远距通信主要手段的短波电台目前无论是在抗干扰能力还是通信距离方面都存在严峻挑战,需要增加一种应急通信手段,当常规短波通信在距离或干扰等极限条件下无法连通时作为常规短波通信的补充,保障飞行员的最低通信需求。

3 基于语音识别技术的机载短波应急通信

3.1 短波通信中话音信号及语义信息分析

短波电台的通信距离和抗干扰能力与短波电台的发射功率、天线尺寸以及接收机的信号处理能力有密切关系。由于机载平台的能耗以及尺寸等因素的限制,发射功率和天线增益提高的空间都不大,唯一具有较大改进潜力的地方就是电台的接收处理性能。

根据香农定理我们知道,接收机处理的信号信噪比同传输速率相关,在信道带宽固定的情况下,信道速率越低,需要的信噪比越低,同时其接收灵敏度也就越高,传输距离越大。也就是说,在其他条件相同时,如果要高保真的语音或高速的数据通信,那么对信道的信噪比要求就越高,通信的距离就越近,抗干扰能力就越差。

当前的通信技术中大量采用声码话技术将传统的模拟语音转化成数字语音进行压缩传输,一般的正常数字语音速率为 64 kb/s(传统电话)。在无线通信中,通常采用降低采样率以及压缩编码的方式进行传输,在保证话音质量的同时降低传输速率。

实际上,正常语音包含有丰富的信息,包括频谱特征、声纹、语气以及语义信息等,其中的很多信息并不是我们所必需的。图 1 是“任务已经完成”这段 2 s 语音的时间-频率-幅度三维采样波形,其对话音正常采样时占用接近 7 kHz 的带宽。图 2 所示的是加了 300 Hz 的带通滤波器后的采样波形,其在占用 300 Hz 带宽时仍旧能够把“任务已完成”这句话的意思表述清楚,前后两者的数据量和数据率相差几十倍,图 2 所示的就是低速声码话的基本工作原理。

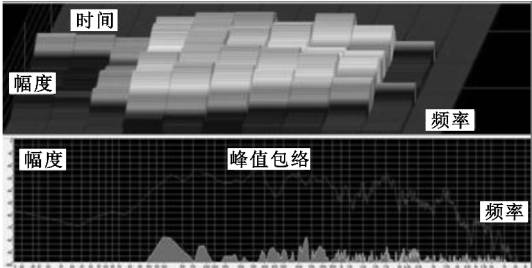


图 1 占用 7 kHz 带宽的语音信号
Fig. 1 Speech signal of 7 kHz bandwidth

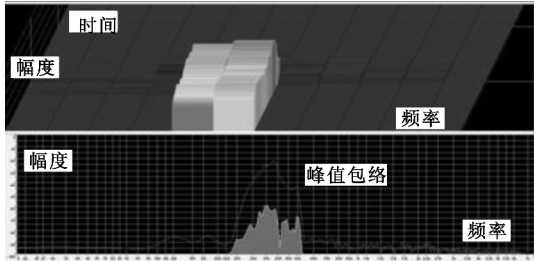


图2 占用 300 Hz 带宽的语音信号
Fig.2 Speech signal of 300 Hz bandwidth

那么,最低可接受的语音通话是什么水平? 通话中真正关注的是语音本身还是其语义? 两者对通信带宽以及信噪比的要求有巨大的差别。例如:飞行员以正常的语速回复“明白”这句话音,以 1 200 b/s 的声码话传输需要 1 200 b 的数据量,而其中的语义用 6 b 的数据就可以表达清楚。当指挥员和飞行战斗人员在强干扰导致的所有正常通信手段失效时,需要的可能仅仅是执行或取消、完成或失败这样一个 6 b 的语义信息。

3.2 自动语音识别技术

语音识别是一个模式匹配的过程,将语音信号转化为文本。通常在贝叶斯统计建模框架下加以解决。如果认为语音可以通过特征提取转换为具有区分度的特征序列 $O = \{o_1, o_2, o_3, \dots, o_T\}$, 那么语音识别的任务就是在所有的词序列空间中搜索最优的词序列

$$\hat{W} = \arg \max_w P(W|O) = \arg \max_w \frac{P(O|W)P(W)}{P(O)} \quad (1)$$

其中, W 表示特征序列 O 对应的词序列, $P(O|W)$ 被称为声学模型概率, $P(W)$ 被称为语言模型概率。语音识别的任务就是找到一个最优的词序列 $\hat{W}$, 使后验概率最大化。

一个典型的语音识别系统如图 3 所示,由前端特征提取模块、声学模型、语言模型和解码器等部分构成。

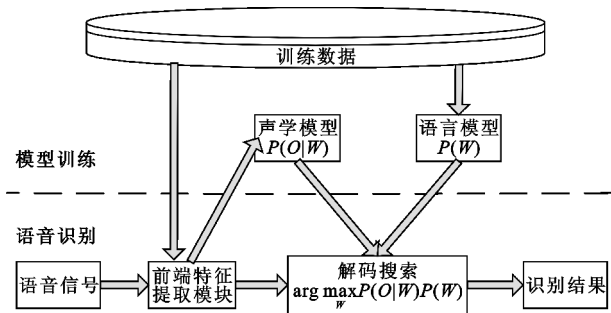


图3 语音识别系统的基本组成

Fig.3 Basic structure of speech recognition system

基于统计建模框架的语音识别需要将原始的语音时域信号参数化,转换为具有区分度的、易于计算机存储和处理的特征矢量。目前主流的语音识别系统都采用短时傅立叶变换、线性预测分析和倒谱分析等信号处理手段。

声学模型要解决的问题是语音特征和建模单元之间的匹配性和区分性的问题。当前主流的语音识别系统绝大多数采用隐马尔科夫模型 (Hidden Markov Model, HMM) 来对声学模型进行建模^[3-4]。一个典型的用于语音识别声学建模的 HMM 如图 4 所示。

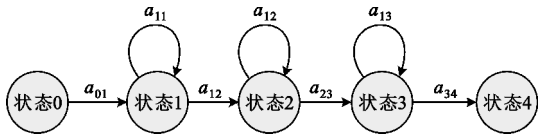


图4 用于语音识别的典型 HMM 结构
Fig.4 Typical structure of HMM for speech recognition

HMM 的每个状态通常采用高斯混合模型 (Gaussian Mixture Model, GMM) 进行描述,一个 GMM 的定义如下:

$$b_s(x) = \sum_{m=1}^M \frac{c_{sm}}{\sqrt{(2\pi)^D |\Sigma_{sm}|}} \cdot \exp \left\{ -\frac{1}{2} (x - \mu_{sm})^T \Sigma_{sm}^{-1} (x - \mu_{sm}) \right\} \quad (2)$$

其中, c_{sm} 是状态 s 中第 m 个混合高斯分量的权重, D 是语音特征向量的维数, μ_{sm} 和 Σ_{sm} 分别为第 m 个高斯分量的均值向量和协方差矩阵。

解码器的任务是在由声学模型、发音词典、命令集等知识源组成的搜索空间中搜索出最佳词序列 (命令)。在语音识别中,解码器所用到的空间异常巨大,必须借助一系列方法对搜索空间进行压缩。

近 20 年来,语音识别技术开始走向市场,并且在一些领域取得了较好的使用效果,典型的例子如苹果 iphone4S 手机中的“语音助手 (Siri)”、科大讯飞公司的“讯飞口讯”软件 (语音识别效果如图 5 所示)。在军用航空电子领域中,美、法、英、瑞典等先后开展了将语音识别技术应用到机载平台的尝试。欧洲的“台风”战斗机率先引入了语音识别技术用于语音控制,能够准确识别 200 多种控制命令,对飞机航电系统的 26 种功能进行控制,其识别准确率超过 95%。美军的 F35 也引入了该项技术,可以识别 100 条指令,识别准确率达到 98%,美军正在考虑将该技术运用到 F-22“猛禽”战斗机上。由此可见,

语音识别技术的成熟度已经达到应用于军用航空电子设备的水平。



图 5 讯飞软件语音识别效果图

Fig. 5 The impression drawing of iFLY's speech recognition software

3.3 基本技术方案

基于上述分析,本文提出一种综合的解决方案,主要思路是在短波电台中引入语音识别及极低速率实时通信两种先进技术,通过语音识别技术提取话音中的语义信息,再将语义信息以极低速实时通信模式进行传递,最大程度减少不必要的通信数据量,以此降低对信道的要求。同时,在接收端将话音含义信息通过语音合成技术生成人工语音,保持短波电台输入输出语音的接口不变,既不改变飞行员的使用习惯,也不改变与短波电台交联的其他航电设备,从而在大幅提高短波电台的通信距离和抗干扰能力的同时,减小项目的规模和成本。

图 6 为系统原理框图,在不影响短波电台其他功能和性能的条件下,在电台收发信机单元里增加一路可选的最低限度通信通道。当正常话音因为干扰或距离过远等原因完全中断时调用其功能实现话音通信。其硬件包括语音识别模块、语音合成模块、控制模块以及极低速波形调制解调模块,其中语音识别和语音合成模块分别与电台的语音输入输出接口相交联,极低速波形调制解调模块与电台内部的信道模块的中频接口相交联。

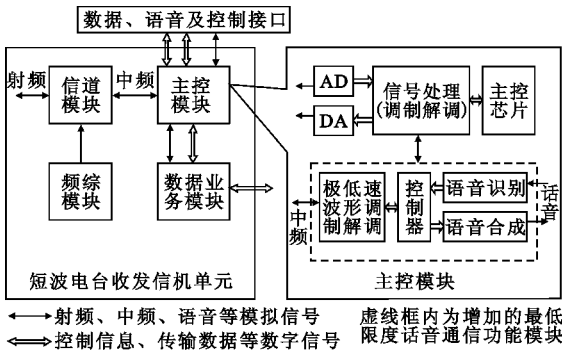


图 6 基于语音识别技术的机载应急通信系统原理框图
Fig. 6 Functional block diagram of the proposed scheme

当该通道用于发话时,语音识别模块对飞行员的话音进行识别,提取语义信息转化为对应的数据,传递至极低速波形调制解调模块,调制产生中频模拟信号经电台内部的信道单元变频后发射;当该通道用于收话时,接收来自电台信道单元的经极低速调制的中频信号,解调出数据信息,传递至语音合成模块生成人工语音传递至飞行员。

该方案在理论上对短波信道的要求最低,从而能极大地提高短波通信的通信距离和抗干扰能力,其优势是增加该通道后的短波电台不需要改变与机上外部航电系统的交联及接口关系,基本不改变飞行员的使用习惯,经验证成功后,可以在新研短波电台中将此功能模块集成到电台原有的主控模块中,研发过程相对简单,研发成本很低。

4 关键技术

4.1 嵌入式命令词语音识别系统

语音识别技术是把高速率的语音信号转换为极低速率的语义数据的关键,采用计算资源要求较低的命令词语音识别系统较为合适。为了在嵌入式系统的低计算资源条件下完成语音识别功能,需要对系统进行高度优化。第一,声学模型使用在嵌入式系统之前,需要进行合理地裁剪、压缩;第二,指令集构成的解码搜索空间需要进行状态合并等技术进行优化;第三,解码器需要使用高斯选择等技术优化搜索算法,以减少计算量。

4.2 说话人自适应

每一套系统提供给飞机所在部队的飞行员使用,因此,可以通过对声学模型进行说话人自适应来提高系统的识别率。通过采集飞行员常用任务对话建立数据库,作为先验数据,采用最大似然线性回归 (Maximum Likelihood Linear Regression, MLLR) 算法^[5]或最大后验概率 (Maximum A Posteriori, MAP) 算法^[6]对声学模型进行说话人自适应,提高系统对相关说话人的识别性能。

MLLR 算法通过说话人相关、识别环境相关的语音数据对背景声学模型的高斯均值和方差进行线性变化,使自适应后的模型能够更好地拟合实际识别的语音。线性变化的参数在最大似然准则下通过 EM 算法训练得到。MLLR 假设自适应后的模型参数与已有背景模型参数存在如下线性变化关系:

$$\hat{\mu} = A\mu + b$$

$$\hat{\Sigma} = H \Sigma H^T \tag{3}$$

其中, μ 、 Σ 分别是均值矢量和协方差矩阵。

MAP 自适应利用贝叶斯理论, 将模型先验信息与带噪声语音数据信息相结合实现自适应。新模型的参数通过最大化带噪声语音数据 x 的后验分布来估计:

$$\hat{\theta} = \arg \max_{\theta} p(\theta | x) \tag{4}$$

其中, θ 为先验模型参数, $\hat{\theta}$ 为模型参数的最大后验概率估计。由于 MAP 比 MLLR 需要估计更多的参数, 因而对自适应数据的数据量有较高的要求。MAP 具有理论上的最优性, 当可用的学习数据较多时, 一般 MAP 相对 MLLR 会表现出一定的优势。

4.3 噪声鲁棒性语音识别

飞机平台上的背景噪声较大, 只有解决了噪声鲁棒性问题, 该项技术才能真正实用。语音识别的噪声鲁棒性是指在输入语音受噪声影响、质量较差、语音的音素特性或声学特性在训练和测试不匹配时, 语音识别系统仍然保持较高的识别精度的性质。噪声鲁棒性语音识别的研究目的就是要消除或减小这种不匹配带来的影响。可以通过映射 f 来描述这种不匹配性:

$$q_{\beta}(s) = f(q_{\alpha}(s)) \tag{5}$$

其中, s 是一个识别单元模型; 用 e 来表示一个特定的声学环境, $q_e(s)$ 表示在声学环境 e 下对 s 的度量; 映射 f 代表一种最小化环境 α 和 β 下的度量转换。可以考虑采用修改解码过程中所使用的声学模型, 即模型域方法, 或采用特征域方法来提高系统在机载噪声条件下的性能。这两大类变换方法如图 7 所示。

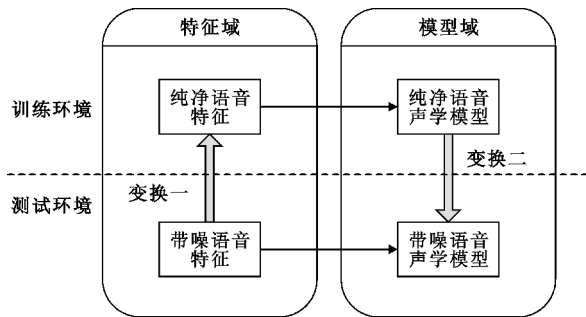


图 7 语音识别噪声鲁棒性方法

Fig. 7 Methods of noise robustness speech recognition

(1) 模型域鲁棒性技术分析

模型域鲁棒性技术主要通过修改训练模型的参数, 使其适应测试语音。前面提到的声学模型自适应技术也可以用于提高系统的噪声鲁棒性。通过采集飞行员话筒传入的机舱内噪声数据使用 MLLR 或

者 MAP 对声学模型进行修正, 可以较好地解决噪声鲁棒性问题。

(2) 特征域鲁棒性技术分析

前端特征域鲁棒性技术的目标是对原始语音信号进行处理, 提取和训练数据“近似”的声学特征, 提供给解码器使用。常见的特征域鲁棒性技术有特征规整技术和鲁棒性特征技术两类。

可以考虑使用均值方差规整技术 (Mean and Variance Normalization, MVN) 对特征进行处理。语音信号 $o(t)$ 是由原始发音信号 $s(t)$ 经过传播函数 $h(t)$ 表征的传输信道得到的, 在时域上表示为 $o(t) = s(t) * h(t)$, 频域上则表示为 $O(f) = S(f)H(f)$ 。对 $O(f)$ 取倒谱则得到

$$\lg O(f) = \lg S(f) + \lg H(f) \tag{6}$$

通常, 同一个信道的传输函数 $H(f)$ 是较为稳定, 减去这一信道倒谱的均值, 能在一定程度上排除信道带来的影响。对语音特征的方差进行规整, 能进一步地抑制加性噪声。设 $O(t)$ 为第 t 帧语音信号倒谱, 语音信号的长度为 T , 那么经过 MVN 后的 $O'(t)$ 为

$$\begin{aligned} Mean &= \frac{1}{T} \sum_{t=0}^T O(t) \\ Var &= \sqrt{\frac{1}{T-1} \sum_{t=0}^T O^2(t) - Mean^2} \\ O'(t) &= \frac{O(t) - Mean}{Var} \end{aligned} \tag{7}$$

鲁棒性特征提取方法试图获取抗噪能力强的特征参数。梅尔频率倒谱系数和感知线性预测系数是目前主流语音识别系统使用的特征。基于人耳听感知机理和子带能量规整特征 (Auditory-Based Sub-band Power Normalized Features, ABSPNF)^[7] 使用实时在线的子带能量规整, 融入感知线性预测系数中的等响度预加重和强度-响度幂次法则, 在噪声环境下有较好的识别性能。采用 ABSPNF 作为语音识别的输入特征, 可以进一步地提高系统在机载噪声下的识别精度。

5 总结及工作展望

本文分析了现有短波通信在抗干扰和通信距离上存在的欠缺, 提出了一种基于语音识别技术的机载短波应急通信方法。该方法通过语音识别技术提取语音信号的语义信息进行传输。针对这种特定的环境, 本文就嵌入式命令词语音识别系统、说话人自适应技术、噪声鲁棒性语音识别技术进行了分析, 采

用这些技术提高了语音识别的性能,保证了基于语音识别技术的机载短波应急通信方法的实现。

本文提出的基于语音识别技术的机载短波应急通信方法降低了对短波信道的要求,可极大地提高短波通信的通信距离和抗干扰能力,为机载平台提供了一种常规短波通信完全失效时的应急通信手段,保障飞行员的最低通信需求。今后的工作方向之一是研究新的声学模型建模技术和鲁棒性语音识别技术,进一步提高系统的稳定性和可靠性。

参考文献:

- [1] 吴昊,陈树新,张衡阳. 基于小卫星中继的远程通信方案[J]. 电讯技术,2012,52(4):452-455.
WU Hao, CHEN Shu-xin, ZHANG Heng-yang. A Remote Communication Scheme Based on Small Satellite Relay [J]. Telecommunication Engineering, 2012, 52(4): 452-455. (in Chinese)
- [2] 王磊,李鹤,周音,等. 信息化条件下短波通信抗干扰技术与应用[J]. 舰船电子工程,2012,32(2):69-72.
WANG Lei, LI He, ZHOU Yin, et al. Technology and Application Research of HF Communication Anti-jamming under the Informatization Condition [J]. Ship Electronic Engineering, 2012, 32(2): 69-72. (in Chinese)
- [3] Levinson S, Rabiner L, Sondhi M. An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition[J]. Bell System Technical Journal, 1983, 62(4): 1035-1074.
- [4] Rabiner L. A tutorial on hidden markov models and selected applications in speech recognition[J]. Proceedings of the IEEE, 1989, 77(2): 257-286.

- [5] Leggetter C, Woodland P. Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models[J]. Computer Speech and Language, 1995, 9(2): 171-185.
- [6] Gauvain J, Lee C. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains[J]. IEEE Transactions on Speech and Audio Processing, 1994, 2(2): 291-298.
- [7] 蔡尚. 自动语音识别中的噪声鲁棒性特征提取方法研究[D]. 北京:中国科学院声学研究所,2012.
CAI Shang. Noise Robust Feature Extraction Methods for Automatic Speech Recognition[D]. Beijing: Institute of Acoustics, Chinese Academy of Sciences, 2012. (in Chinese)

作者简介:



陈晨(1980—),男,四川巴中人,2009年于空军工程大学获博士学位,主要从事航空武器装备研制管理;

CHEN Chen was born in Bazhong, Sichuan Province, in 1980. He received the Ph. D. degree from Air Force Engineering University in 2009. He is engaged in management of aeronautic weapon equipment development.

Email: 43614191@qq.com

胡贝贝(1981—),男,山西恒曲人,2003年于空军工程大学获学士学位,主要研究方向为雷达抗干扰、通信抗干扰以及电子对抗。

HU Bei-bei was born in Hengqu, Shanxi Province, in 1981. He received the B. S. degree from Air Force Engineering University in 2003. His research concerns radar anti-jamming, communication anti-jamming, and ECM.