

电力自主移动机器人运检 AI 助手语音多路信号端点切分

杨洋[†], 宋祉霖, 豆朝宗

(中国核电工程有限公司, 北京 100840)

摘 要:研究了电力自主移动机器人运检 AI 助手语音多路信号端点切分仿真, 解决语音端点切分运算复杂、不准确问题。机器人运检 AI 助手语音服务包括语音多路信号采集、播报、理解、控制等, 采集到的 AI 助手语音多路信号经客户端降噪处理后, 通过主分量分析筛选语音多路信号矩阵的特征值和特征向量, 确定最有用因素作为基底向量, 获取语音多路信号中最典型特征, 经线性区分分析使语音多路信号的特征布局集中, 获取用于切分的特征并构建变化矩阵, 利用 MLLT 计算变化矩阵, 完成语音多路信号样本的协方差矩阵对角化, 并将特征信息作为服务端卷积神经网络预测模型输入, 从帧级上实时分类语音多路信号数据, 实现语音多路信号端点的切分。仿真实验结果显示:该方法可有效去除语音多路信号噪声, 并完成语音多路信号端点准确切分。

关键词:语音多路信号; 机器人; AI 助手; 语音特征提取; 预测模型; 端点切分

中图分类号: TN912

文献标识码: A

Power Autonomous Mobile Robot Operation and Inspection AI Assistant Speech Multiplex Signal Endpoint Segmentation

YANG Yang[†], SONG Zhilin, DOU Chaozong

(China Nuclear Power Engineering Co., Ltd., Beijing 100840, China)

Abstract: This paper studies the simulation of voice multi-channel signal endpoint segmentation of the AI assistant of electric power autonomous mobile robot operation inspection, and solves the problem of complex and inaccurate voice endpoint segmentation operation. Robot operation inspection AI assistant voice service includes voice multi-channel signal acquisition, broadcasting, understanding, control, etc. after the collected AI assistant voice multi-channel signal is denoised by the client, the eigenvalues and eigenvectors of the voice multi-channel signal matrix are screened through principal component analysis, and the most useful factors are determined as the base vector to obtain the most typical features in the voice multi-channel signal. After linear discrimination analysis, the feature layout of the voice multi-channel signal is centralized, Obtain the features used for segmentation and construct the transformation matrix. Use mllt to calculate the transformation matrix, complete the diagonalization of the covariance matrix of the speech multi-channel signal samples, and input the feature information as the convolution neural network prediction model of the server, classify the speech multi-channel signal data in real time from the frame level, and realize the segmentation of the speech multi-channel signal endpoint. The simulation results show that this method can effectively remove the noise of speech multi-channel signals and complete the accurate segmentation of speech multi-channel signal endpoints.

Key words: voice multi-channel signal; robot; AI assistant; speech feature extraction; prediction model; endpoint segmentation

现今中国经济的建设,离不开电力的支持。电网运行的稳定是供电顺畅的保障,需要时刻监管供电设备,发现问题,及时维修,这使得电网运行维护工作大大增加。单纯靠人力管理,不仅成本增加,管理效率也不高^[1,2]。电力自主移动机器人运检 AI 助手利用智能化分析的方法,可以提高信息获取的性能,通过语音的切分,可以得到电网运行状况信息。

为了能使语音端点切分准确,许多学者做了研究,例如雷静等人提出语音端点检测方法^[3],主要利用语音和噪声之间参数特征的差异,实现语音端点的切分,但此方法容易被噪声干扰,切分效果不稳定,鲁棒性差;张子丞等人研究的语音端点检测^[4],利用贝叶斯决策准则获取最佳判决阈值,通过对数能量谱和阈值的比较结果,来划分语音和背景噪声,其算法复杂,耗时较长,不建议推广。

为了寻求更简便、更稳定、更准确的语音端点切分方法,提出了电力自主移动机器人运检 AI 助手语音多路信号端点仿真,利用改进的语音多路信号特征提取方式,通过卷积神经网络预测模型,达到语音多路信号端点切分的目的。

1 电力自主移动机器人运检 AI 助手语音端点切分

1.1 电力自主移动机器人运检 AI 助手功能

电力自主移动机器人运检 AI 助手能采集电力运行、监控等各种电力相关数据,利用大数据分析进行智能研判,兼顾信息采集、异常告警、语音服务等功能,电力自主移动机器人运检 AI 助手的结构如图 1 所示。

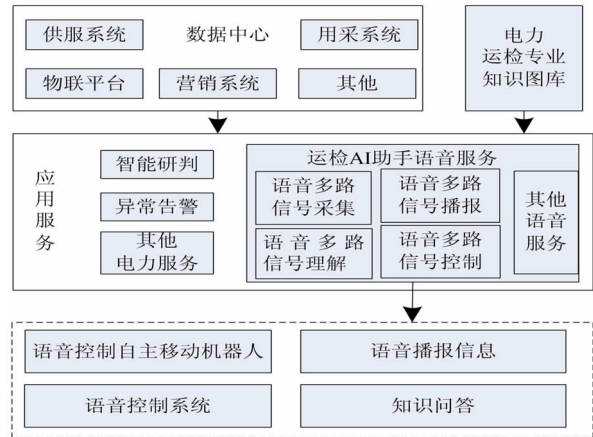


图 1 电力自主移动机器人运检 AI 助手结构图

根据图 1 可知,电力自主移动机器人运检 AI

助手数据中心汇集了供服系统、营销系统、用采系统、物联平台以及其他运行系统所涉及的数据,结合电力运检专业知识图库,可以进行多种应用服务,包括智能研判、异常告警、语音服务及其他电力服务。其中运检 AI 助手语音服务包括语音多路信号采集、语音多路信号播报、语音多路信号理解、语音多路信号控制及其他语音服务,利用这项服务可以实现语音控制自主移动机器人、语音播报信息、知识问答、语音控制系统的功能^[5,6]。

1.2 语音多路信号端点切分技术

运检 AI 助手语音服务需要语音多路信号端点切分技术的支持,才能顺利运行,其技术的关键步骤为语音多路信号的特征提取和预测模型的分类^[7,8]。机器人运检 AI 助手语音多路信号采集的效果影响着语音音频的特征,通过对语音多路信号提取声音特征,并对其分类,从而完成语音多路信号点的切分目的,切分技术框架如图 2 所示。

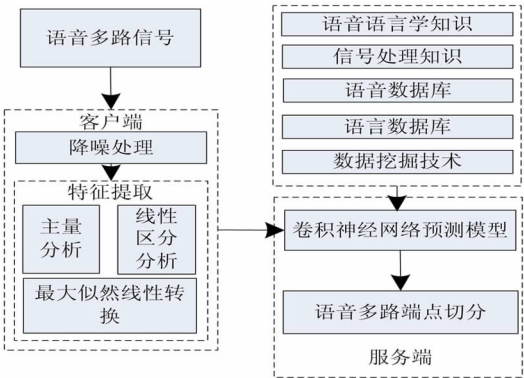


图 2 语音多路信号端点切分技术框架

根据图 2 可知,采集的语音多路信号先传至客户端,经过降噪后,利用主分量分析(Principal Component Analysis,PCA)和线性区分分析(Linear Discriminant Analysis,LDA)方法,并结合最大似然线性转换(Maximum Likelihood Liner Transformation,MLLT)算法进行特征提取。根据语音语言学知识、数据挖掘技术、信号处理知识、语音数据库和语言数据库的信息进行训练后,传至服务器,然后通过卷积神经网络预测模型,对语音多路信号特征进行分类,完成语音多路信号端点的切分。

1.3 语音多路信号特征提取

语音多路信号特征参数常用指标为 Mel 频率倒谱系数(MFCC)。特征提取的本质是对特征参数的提取,需要提取方法具备较好的鲁棒性^[9-11],为此,语音多路信号特征提取包含以下三个步骤:

(1)根据人耳听觉采集信息的特点,选取 MF-CC 作为语音多路信号特征参数。

(2)将提取的语音多路信号特征以线性变化的方式进行降维处理,目的是将语音多路信号中最具区分性能的元素进行保留。电力自主移动机器人运检 AI 助手采用数据驱动线性特征转换方法完成区分性能元素保留,该转换方法包括 PCA、LDA 等。

(3)为了减少语音多路信号的噪声,利用倒频谱均值减法增强语音多路信号的特征,该方法运算简单,应用广泛。

1.3.1 主分量分析语音多路信号特征

将语音多路信号中最有用的元素通过主分量分析进行筛选,删除过多无用元素,该分析法的主要任务是确定好最有用因素所在语音多路信号的基底向量,且满足每个基底向量具有单位正交的条件^[12,13]。在语音多路信号特征主量分析中,训练语音多路信号的语料用 $A = [a_1, a_2, a_3, \dots, a_n]$ 表示,语音多路信号的语料总量用 Q 表示,每个 a_i 为 n 维,语音多路信号的相关矩阵 Z 的运算过程见式(1):

$$\begin{cases} Z = \frac{1}{Q} \sum_{i=1}^Q (a_i - \alpha)^Z \\ \alpha = \frac{1}{Q} \sum_{i=1}^Q a_i \end{cases} \quad (1)$$

式中, Z 为 $m \times n$ 维矩阵; α 代表语音多路信号的总体均值向量。

通过计算获取矩阵 Z 的特征值,并将其排列,首位是特征最大值,依次排列直到最小值,利用前 k 个特征值的特征向量组建基底矩阵,每个语音多路信号的语料都能获得变化矩阵 β_k ,并将矩阵投射给特征空间,用 $B = [b_1, b_2, b_3, \dots, b_n]$ 表示,其中 $b_i = \beta_k^Z a_i$, $i = 1, 2, 3, \dots, Q$ 。 A 的第一主分量,即为 B 的第一维分量,按顺序排列出第二、第三、等主分量。

1.3.2 线性区分分析语音多路信号特征

线性区分分析使语音多路信号的特征布局集中,从而使每路语音信号间的布局分散,在语音多路信号的特征线性区分中,作出如下分析:

(1)当特征向量投射后,每一维中都包含了语音多路信号特征信息,但不是每一个信息都可以用于区分特征,其中前 k 维子空间中的语音多路信号特征信息可用于区分特征,其余 $(n - k)$ 维子空间内的信息没有意义,可删除。

(2)所有采集的样本都具有高斯分布的特点,

当语音多路信号利用向量模式进行表达时,LDA 的最终目的是使样本与样本间在进行协方差矩阵变化后,得到的行列式值尽量小,这就需要对语音多路信号两个样本进行比较,比值取最大,见式(2):

$$\beta_k = \arg_{\beta_k} \max \frac{|\beta_k^Z E \beta_k|}{|\beta_k^Z F \beta_k|} \quad (2)$$

式中,为了使 θ_p 存在最大值,需要对矩阵 F 和 E 进行计算,获取特征值和相应的特征向量,并按照值的大小进行排列,以最大值为首位,依次排列,最小值为末尾,利用前 k 个的语音多路信号特征值所对应的特征向量构建变化矩阵。语音多路信号信息可以分成 G 个样本, Q_g 代表第 g 个样本的语音多路信号数量, F_g 代表第 g 个样本的协方差矩阵, a_i 代表训练语音多路信号的集合, α_g 代表训练语音多路信号第 g 个样本均值向量, $h(i)$ 代表 a_i 所属的语音多路信号样本, F 和 E 的表达见式(3):

$$\begin{cases} F = \frac{\sum_{g=1}^G Q_g F_g}{Q} \\ E = \frac{\sum_{g=1}^G Q_g (\alpha_g - \alpha)(\alpha_g - \alpha)^Z}{Q} \end{cases} \quad (3)$$

式中, F_g 、 α_g 和 α 的计算过程,见式(4):

$$\begin{cases} F_j = \frac{\sum_{h(i)=g} (a_g - \alpha_g)(a_g - \alpha_g)^Z}{Q}, \\ g = 1, 2, 3, \dots, G \\ \alpha_g = \frac{\sum_{h(i)=j} a_i}{Q}, g = 1, 2, 3, \dots, G \\ \alpha = \frac{\sum_{i=1}^Q a_i}{Q} \end{cases} \quad (4)$$

1.3.3 改进特征提取算法

在 PCA 和 LDA 的特征提取过程中,都存在协方差矩阵不能对角化的缺点^[14,15],为了确保语音多路信号特征提取的真实性,利用 MLLT 对二者进行了改进。MLLT 通过重新获取变换矩阵,从而变换矢量空间,跳过数据降维过程,直接将协方差矩阵对角化。利用 MLLT 计算协方差矩阵,见式(5):

$$\hat{\beta} = \arg \max Q \ln |\beta| - \sum_{g=1}^G \frac{Q_g}{2} \ln |\text{diag}(\beta^Z F_g \beta)| \quad (5)$$

式中, $\hat{\beta}$ 表示协方差矩阵; β 表示通过 PCA 或

LDA 获取的变化矩阵;矩阵对角元素表示为 diag 。

通过完善算法,对语音多路信号特征提取的步骤如下:

(1)选用 MFCC 作为语音多路信号的特征参数。

(2)利用 PCA 和 LDA,对矩阵 F 和 E 进行计算,并按照特征值的大小进行排列,取前 k 个的语音多路信号特征值的特征向量构建变化矩阵 β 。

(3)使用 MLLT 再进行一次线性变换,获得新的矩阵 $\hat{\beta}$,完成对语音多路信号的特征提取任务。

1.4 卷积神经网络语音多路信号预测模型

通过上述方法对语音多路信号进行特征提取后,传输到训练好的卷积神经网络预测模型中,对每一帧输出的语音进行概率计算,实现语音多路信号端点切分的功能。

利用卷积神经网络进行语音多路信号端点切分的实质是指在帧级上,将实时语音多路信号数据进行分类。以 28 ms 为一帧,将每一帧的语音多路信号进行特征提取,然后将特征信息传递给卷积神经网络预测模型进行分类,获得该帧为语音多路信号或非语音多路信号的概率值。卷积神经网络预测模型结构如图 3 所示。

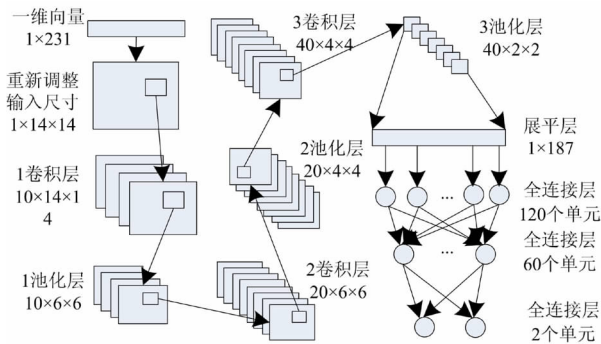


图 3 卷积神经网络预测模型结构图

根据图 3,一帧采样数据利用电力自主移动机器人运检 AI 助手采用主分量分析、线性区分分析的方法对一帧语音多路信号采样数据进行特征提取,得到一维向量,其大小为 1×231 ,将此向量重新调整为 14×14 后,再输入到预测模型中。由卷积层和池化层负责预测模型的卷积功能,利用多层卷积获得更多语音多路信号的特征,由于卷积层具有权值共享的功能,因此运算中参数的数量会减少,可以提高运算效率。

卷积层的卷积核大小设置为 2×2 ,每一层选用不同数量的卷积核,分别为第一层 10 个、第二层 20 个、第三层 40 个,以 ReLU 函数作为激活函数。

在卷积结果进入池化层后,利用第一、第二、第三层最大池化层完成卷积结果采集。经过展平层对三维特征图降至一维后,再传递给含有神经元的全连接层,神经元的数量分别为 120 个、60 个。最终由含 2 个神经元的 Softmax 层负责输出语音多路信号或非语音多路信号的概率。

1.5 语音多路信号端点切分

电力自主移动机器人运检 AI 助手语音多路信号端点切分的步骤如下:

(1)电力自主移动机器人运检 AI 助手利用声音采集装置重采样语音多路信号数据,通过 PCA、LDA、MLLT 完成特征提取。

(2)将上述信息输入卷积神经网络预测模型,对语音多路信号特征信息进行分类,辨别语音多路信号和非语音多路信号。由于语音多路信号之间存在互相关联的关系,因此在特征提取时,需要结合前后文的内容。在卷积神经网络预测模型输入的过程中,需要将一整段语音多路信号序列进行输入。如果设置采集到的语音多路信号为 n 帧,可以表示为 A_n ,预测模型的结果设置为 S_n ,二者的表达见式(6):

$$\begin{cases} A_n = \{a_1, a_2, \dots, a_n\} (n \geq 1) \\ S_n = \{s_1[v_0, v_1], s_2[v_0, v_1], \dots, s_n[v_0, v_1]\} \\ (n \geq 1) \end{cases} \quad (6)$$

式中, v_0 表示一帧语音多路信号的概率, v_1 表示一帧非语音多路信号的概率。若 $v_0 \geq v_1$,说明这一帧属于语音多路信号,不然则判定为非语音多路信号。通过预测模型, A_n 序列中连续的 R ($R \geq 0$) 段语音帧可以看作 R 个语音多路信号,采集并记录每一段语音多路信号,完成语音多路信号端点的切分。

2 仿真实验结果

在仿真实验中,将电力自主机器人运检 AI 助手采集的语音多路信号数据作为实验对象,采样的数据主要是电力工作人员发音的英文字母串,每个字母单独发音,且发音字母之间有较长间歇。采样的频率为 39000 Hz,精度达 15 位,帧长 30 ms,即 800 个采样点,帧偏移 9 ms。在实验数据中加入噪声干扰,选取一定数量的语音多路信号样本进行降噪和特征提取处理,然后利用 WaveSufur 软件,将实验数据中语音多路信号端点和静音之间进行切分,验证本文方法的应用效果。

2.1 降噪和特征提取效果

实验选用白噪声作为语音多路信号的干扰噪声,将电力自主机器人运检 AI 助手采集的语音多路信号前加入白噪声视为含噪语音多路信号,利用本文方法进行仿真实验,仿真结果如图 4 所示。

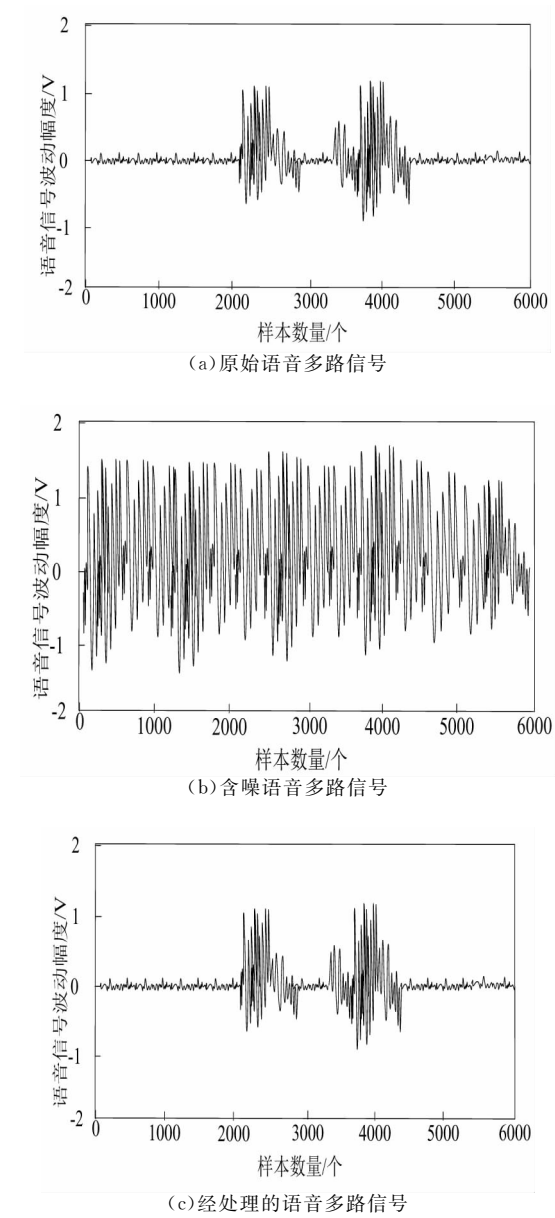


图 4 语音多路信号的降噪和特征提取效果图

根据图 4 可知,图 4(a)是未加入噪声的原始语音多路信号,图 4(b)是加入噪声后语音多路信号的效果,图 4(c)是经过本文方法处理后的语音多路信号。在加入噪声干扰的语音多路信号中,语音多路信号已经无法分辨,经过本文方法的降噪和特征提取后,语音多路信号可以清晰地展示出来,与未加入噪声干扰的语音多路信号进行对比,可以看出二者的语音多路信号无明显差别,说明本文方法具有良好的降噪性和特征提取准确性。

2.2 端点切分效果

在仿真实验中,为了能够清楚地分辨出端点切分效果,切分的语音多路信号选择对字母间存在的静音端进行切分,利用特征提取改进前和改进后的本文方法对语音多路信号端点进行切分,切分效果如图 5 所示。

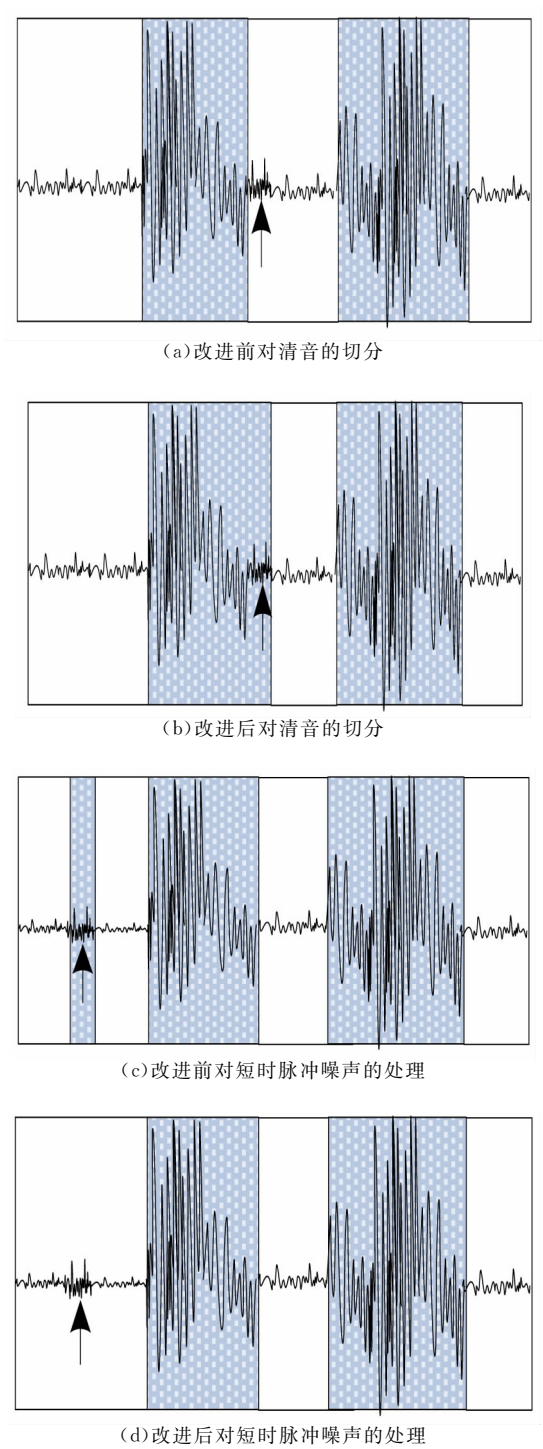


图 5 语音多路信号端点切分效果图

根据图 5 可知,图 5(a)是改进前方法对清音的切分,花纹部分代表语音多路信号段,空白部分

代表非语音段,通过直线将二者进行划分,即为端点切分处。箭头标出的位置,代表 f 的尾音,由于 f 音变化幅度小,在改进前的本文方法中,该音被切分到非语音段里。图 5(b)是改进后的本文方法对清音的切分效果,可以看出清音被切分进语音多路信号段中。图 5(c)和图 5(d)是改进前后的本文方法对有脉冲噪声干扰的语音多路信号端点切分的效果。在图 5(c)中,由于脉冲噪声的干扰,在原本的非语音段中多切分出了一个语音多路信号段,说明未改进的方法将脉冲噪声划分成语音多路信号段。图 5(d)显示,改进后的本文方法将脉冲噪声划分成非语音段,说明经过特征提取改进的本文方法端点切分更准确。

在上述实验的基础上,引入文献[3]方法和文献[4]方法作为对比方法,进行对比测试。以切分耗时为指标,测试不同方法的应用效率。结果如表 1 所示:

表 1 切分耗时对比			
测试样本组别	本文方法/ms	文献[3]方法/ms	文献[4]方法/ms
1	349	679	907
2	355	644	1028
3	317	624	1148
4	328	688	1054
5	354	612	993

分析表 1 可知,本文方法对语音多路信号端点进行切分所需时间平均为 340.6 ms,而文献[3]方法和文献[4]方法的端点切分耗时分别平均为 649.4 ms 和 1026 ms。可以体现本文方法对语音多路信号端点切分可以有较高的效率,快速实现端点切分,有利于高效处理信号数据。

3 结 论

本文方法通过仿真实验,证实在实际工作中,电力自主移动机器人运检 AI 助手用于系统相互控制、获取现场信息等,有利于电力工作人员及时汇报信息,并获得解决办法,提升电网运维工作效

率。在语音多路信号采集过程中,城市交通噪音、工厂机器运行噪音等都属于较大的噪音环境,对运检 AI 助手语音多路信号端点切分的能力时刻都有挑战,因此,未来可以通过提高上下文的利用率,从而提高语音多路信号端点切分的性能。

参考文献

[1] 郭东岳,林毅,杨波. 基于 CGRU 多输入特征的地空通话自动切分[J]. 四川大学学报(自然科学版),2020,57(5):887—893.

[2] 李丽亚,闫宏印. 融合递归求逆滤波的机器人混合语音识别方法[J]. 计算机仿真,2020,37(8):277—280+290.

[3] 雷静,何培宇,徐自励. 低信噪比下多参数融合的自适应语音端点检测[J]. 信号处理,2020,36(8):1205—1211.

[4] 张子丞,谭志苇,张晨瑞,等. 基于高低频带对数能量谱比贝叶斯决策的语音端点检测[J]. 计算机科学,2021,48(S1):33—37.

[5] 刘航,李扬,袁浩期,等. 基于生成对抗网络的语音信号分离[J]. 计算机工程,2020,46(1):302—308.

[6] 艾佳琪,左毅,刘君霞,等. 基于余弦相似度的动态语音特征提取算法[J]. 计算机应用研究,2020,37(S2):147—149.

[7] 罗庆,包亚萍,俞强. 基于改进语音特征与极限学习机的语音端点检测[J]. 微电子学与计算机,2020,37(3):37—41.

[8] 张明亮,陈雨. 基于全卷积神经网络的语音增强算法[J]. 计算机应用研究,2020,37(S1):135—137.

[9] 韦莎丽,王健,曾庆宁,等. 低信噪比环境下语音端点检测技术[J]. 计算机工程与设计,2020,41(9):2507—2514.

[10] 韩云霄,邵清,符玉襄,等. 复杂噪声中基于 MFCC 距离的语音端点检测算法[J]. 计算机工程,2020,46(3):309—314.

[11] 黄洋,赵风海,卢景. 语音信号处理中双门限端点检测算法的改进[J]. 南开大学学报(自然科学版),2021,54(2):58—62.

[12] 张红兵. Mel 频率下语音信号深度频谱特征提取方法仿真[J]. 计算机仿真,2020,37(5):197—200+267.

[13] 黄镇坤,章小兵,朱俞清. 低信噪比环境下改进的新能零熵语音端点检测[J]. 微电子学与计算机,2020,37(6):19—23+29.

[14] 罗思洋,龙华,邵玉斌,等. 噪声环境下多特征融合的语音端点检测方法. 云南大学学报(自然科学版),2021,43(4):671—680.

[15] 杨洋,杨震. 基于图信号处理的 MVDR 波束形成多通道语音增强[J]. 南京邮电大学学报(自然科学版),2021,41(6):35—40.