

信息技术与系统

一种基于感知语音质量评价的列车显示器 语音自动检测方法

高峰, 张宏伟

(株洲中车时代电气股份有限公司, 湖南 株洲 412001)

摘要: 列车显示器语音测试时, 调试员需要通过耳朵判断语音的质量, 而调试员对语音的判断会受环境噪声、显示器装配等因素影响, 造成了出厂的列车显示器语音质量存在不一致性的现象。对此, 文章提出一种语音自动检测方法, 其利用麦克风采集列车显示器发出的声音, 采用感知语音质量评价 (PESQ) 算法对列车显示器发出的语音进行自动检测和评价。仿真结果显示, PESQ 评分与人工主观评分二者趋势保持一致, 两者最大偏离值为 0.14, 可见本文所提方法实现了感知语音质量评价自动化, 保证了测试一致性。

关键词: 感知语音质量评价; 语音自动检测; 体验质量; 自动采集; 一致性

中图分类号: TP391.42

文献标识码: A

文章编号: 2096-5427(2021)01-0076-05

doi:10.13889/j.issn.2096-5427.2021.01.013

An Automatic Voice Test Method of HMI for Train with Perceptual Evaluation of Speech Quality

GAO Feng, ZHANG Hongwei

(Zhuzhou CRRC Times Electric Co., Ltd., Zhuzhou, Hunan 412001, China)

Abstract: During the train HMI speaker test, operator should judge the quality of voice by ear, which causes difference in HMI voice quality test by the influence of environment noise, HMI assembly, etc. It describes a automatic detection method which uses microphone to sample the voice of HMI, and a PESQ algorithm is used to automatically detect and evaluate the voice from HMI. Simulation results show that the PESQ score is consistent with the artificial subjective score, and the maximum deviation is 0.14, which shows that the proposed method realizes the automatic evaluation of perceptual speech quality and ensures the consistency of test.

Keywords: perceptual evaluation of speech quality(PESQ); automatic voice test; QoE(quality of experience); automatic acquisition; uniformity

0 引言

随着复兴号电力动车组的大量投入应用, 建立一套针对复兴号电力动车组列车显示器 (以下简称“列车显示器”) 的全自动测试系统非常重要。语音是

列车显示器与驾驶员间传递信息的重要载体之一, 但驾驶员对语音的判断会受驾驶室内噪声的影响; 同时列车显示器在装配过程中受喇叭公差、壳体公差、电路板输入信号误差及背景噪声等因素的影响, 会导致部分喇叭播放的语音存在能被感知到的异音, 因此有必要建立一套全自动的语音质量评价系统来评价显示器语音的清晰度、可懂度和自然度^[1]。

收稿日期: 2020-06-04

作者简介: 高峰 (1985—), 男, 工程师, 主要研究方向是可测试性设计、边界扫描、视觉检测、人工智能等。

列车显示器测试的一个重要评价指标是语音质量。探索和研究稳定、可靠的语音体验质量 (quality of experience, QoE) 评价模型 / 方法成为国内外众多研究者共同的目标。目前业界已有语音质量评价技术包括信噪比 (signal noise ratio, SNR)、线性预测编码 (LPC)、Bark 谱失真和听觉模型等, 各评价方法均是基于一个参考音频来模拟人耳的感知原理, 对获得的音频进行打分。列车显示器的语音由芯片内部的数字量转换成模拟量并通过喇叭播放出来, 类似于移动电话接收信号并转变成语音播放的过程。基于以上考虑, 一般采用国际电信联盟标准 ITU-T P.862.1 《评估宽带电话网络和语音编码器的端到端语音质量的客观方法》中推荐的感知语音质量评价 (perceptual evaluation of speech quality, PESQ) 算法作为检测列车显示器语音的解决方案。PESQ 算法得到的评分结果与平均意见得分 (mean opinion score, MOS) 相关, 比如, MOS 分数为 4.0 的语音与 PESQ 分数为 3.7~3.9 的语音质量相当。PESQ 是国际电信联盟公布的语音质量客观评价算法中与主观评价相关度最高的一种^[2]。

语音质量评价一般都是通过测试系统来比较原始语音和输出语音之间的某种差异^[3]。本文主要在列车显示器自动测试台上, 通过对喇叭的语音质量作评价来有效检验出有异音的产品, 从而保证列车显示器的语音质量。首先通过人耳对声音的感受和认知过程的抽象得到一个数据模型, 然后将同一设备录制的标准语音音频 (无噪声的音频) 和被测列车显示器输出后录制的语音音频一起输入该模型, 比较两者的差异并给出列车显示器语音质量的客观评价。该软件评价模型的优点是能够满足大脑对语音质量的感知和评价, 不足之处是受人的主观行为影响比较大, 并且需要花费较多的时间和精力去获取样本评价结果。为此, 文献 [4] 优化了对语音质量评价的客观方法, 研究了用软件模型自动判别语音质量的可行性。纵观语音评价方法, 只要评价中考虑了人对语音信号的感知特性, 就会大幅提高整个评价方法的性能^[5]。本文借鉴了此优化方法, 在优化后的条件下 (测试环境为隔音环境) 进行列车显示器的语音测试。

1 PESQ 评价模型

PESQ 的软件评价模型如图 1 所示。其将语音信号的特征 (如频率、响度等参数) 与人耳对语音质量

的感知及大脑的主观评价通过建立模拟人的听觉模型与认知模型相结合, 实现了语音质量评价。PESQ 的电平对齐功能对列车显示器语音播放延时、环境噪声引起的差异等具有较好的抑制作用, 从而保证了列车显示器语音评价的一致性。

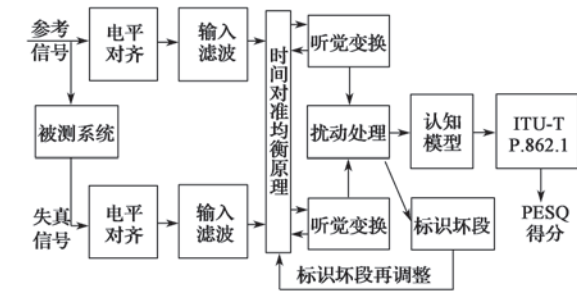


图 1 感知语音质量评价模型框图
Fig. 1 Block diagram of PESQ model

PESQ 算法流程如图 2 所示。

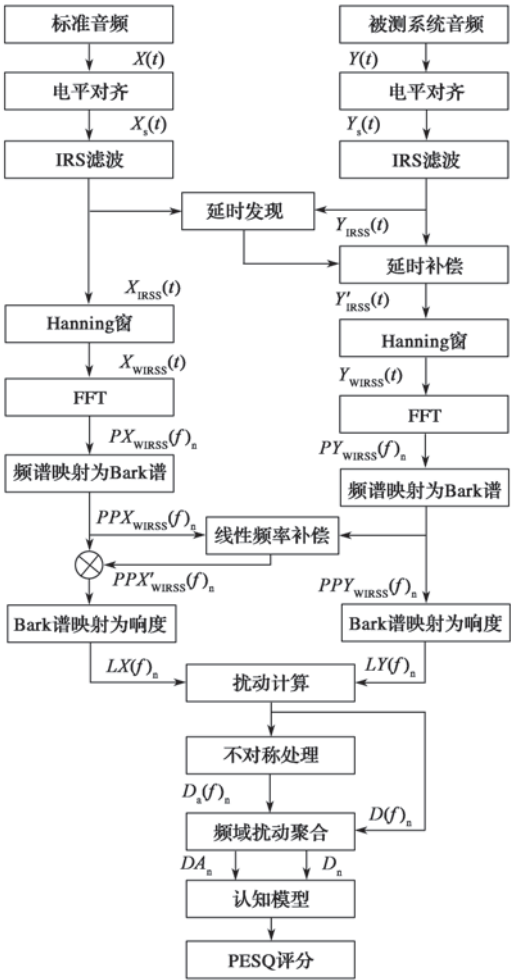


图 2 感知语音质量算法流程
Fig. 2 Flow chart of PESQ algorithm

PESQ 软件评价模型在处理语音信号时, 首先将标准语音 $X(t)$ 和列车显示器播放的语音 $Y(t)$ 进行电平对齐, 得到标准音频 $X_s(t)$ 和被测语音信号

$Y_s(t)$ ，再将被测语音信号和列车显示器发出的语音信号调整到指定的声压级（如 79 dB）的标准听觉电平；然后将对齐后的标准音频 $X_s(t)$ 通过中间参考系统（intermediate reference system, IRS）滤波得到 $X_{IRSS}(t)$ ，将语音信号 $Y_s(t)$ 通过 IRS 滤波得到 $Y_{IRSS}(t)$ 并经过延时补偿得到 $Y'_{IRSS}(t)$ ；对通过时间对准后的两路信号 $X_{IRSS}(t)$ 和 $Y'_{IRSS}(t)$ ，以指定间隔（如 32 ms）为一帧，相邻的前后帧之间有一定的重叠（如 50%），依次进行 Hanning 窗和快速傅里叶变换，得到标准音频的功率谱 $PX_{WIRSS}(f)_n$ 和列车显示器发出的语音信号的标准功率谱 $PY_{WIRSS}(f)_n$ ；最后将两者的功率谱通过频率尺度映射分别获得 Bark 谱（也称音调功率谱） $PPX'_{WIRSS}(f)_n$ 和 $PPY_{WIRSS}(f)_n$ ，再对两路信号的 Bark 谱进行比较。被测语音是被评价的对象，因此线性频率补偿只针对参考语音，首先计算两路语音信号 Bark 值，将被测语音与参考语音的平均 Bark 值进行比较并补偿参考语音（如最大不超过 20 dB），得到标准音频的响度值 $LX(f)_n$ 和列车显示器语音的响度值 $LY(f)_n$ 。将 $LX(f)_n$ 和 $LY(f)_n$ 相减并对得到的差值做归零处理，得到一个相对于标准音频的扰动值 $Da(f)_n$ 。分析扰动曲面提取出两个失真参数 DA_n 和 D_n ，并在频率和时间上累积起来，映射到 PESQ 评分预测值^[6]。

2 列车显示器自动测试台

2.1 列车显示器自动测试台简介

列车显示器自动测试台（图 3）主要由上位机、视觉模块、显示器、电源模块、音频分析模块、通信测试模块及机器人模块等组成。



图 3 列车显示器自动测试台三维图
Fig. 3 3D image of the automatic test equipment for HMI

图 4 示出列车显示器自动测试台功能框图。其中上位机集成各种采集卡，用于数据采集和控制；电源模块用于给显示器供电；4 轴机器人模块用于测试列

车显示器的按键和触摸屏；视觉模块用于视觉检测及视觉校准；通信模块负责列车显示器数据交换；机器视觉模块负责列车显示器屏幕显示状态的监测；音频分析模块负责列车显示器播放语音采集。本文主要介绍音频分析模块。

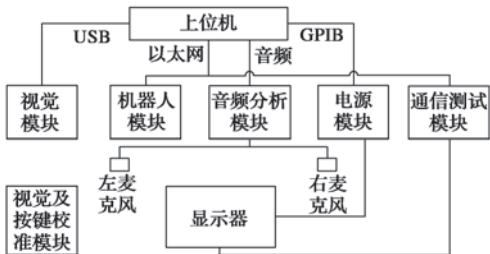


图 4 列车显示器自动测试台功能框图
Fig. 4 Block Diagram of the automatic test equipment for HMI

2.2 音频分析模块

列车显示器自动测试台上的音频分析模块主要负责列车显示器语音测试的启动控制、语音数据的采集及处理、PESQ 评价模型计算、数据库记录及分析，其软件功能如图 5 所示。



图 5 音频检测模块功能框图
Fig. 5 Block diagram of sound test module

语音启动控制主要是由 4 轴机器人模块通过机械臂上的按压笔按压列车显示器的按键或触摸屏，利用压力传感器监测按键或触摸屏的压力并记录相应数据（图 6）。

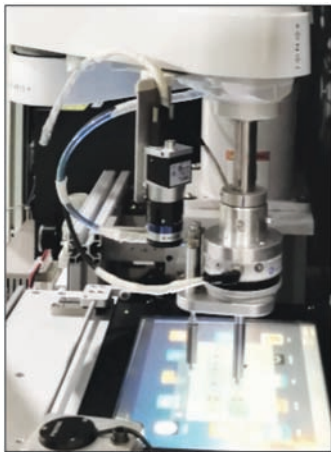


图 6 压力测试示意图
Fig. 6 Schematic diagram of pressure test

语音数据采集部分主要根据列车显示器扬声器位置，由音频卡控制其中 1 个麦克风来采集显示器播

放的语音信息，再将采集到的语音信号变成“wav”格式的文件并保存到本地。

语音数据处理部分主要是使录制的语音文件满足感知语音质量的要求，将音频文件对齐，以避免因为上位机、机器人模块和列车显示器之间的累积时间差产生的空白区域超过 PESQ 规定的 0.2 s 的要求^[7]，并将语音文件前面过长的空白部分删除，以便生成符合要求的测试文件。包含空白区域的语音文件示例如图 7 所示。

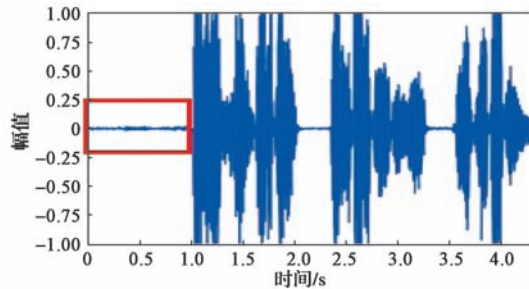


图 7 包含空白区域的语音文件
Fig. 7 Voice file with blank area

感知语音质量软件评价模型计算是将所测得的语音信号与标准语音信号进行对比，采用 PESQ 系统得到各个音频文件的评价分数。根据标准 ITU-T P.830《电话频段和宽带数字编解码器的主观性能评定》中的要求，收集列车显示器发出的语音与预先设定的标准语音并在特定的环境下提供给测试者，通过收集测试者对上述语音的主观感受，包括声音的失真及声音中所含噪声等。MOS 得分采用五分制，相关标准如表 1 所示。

表 1 平均意见得分（MOS）标准
Tab. 1 Standard of MOS

平均意见得分（MOS）	获得的语音感受	语音的失真描述
5	优秀	无法察觉失真
4	良好	稍微察觉失真但不厌恶
3	中等	察觉失真但有点厌恶
2	差	厌恶但能容忍
1	极差	不能容忍的厌恶

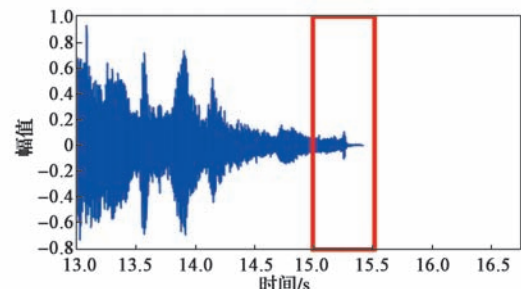
数据库记录及分析模块主要是将语音信号的结果记录至数据库并进行定期的分析，生成对应的分析报告。

3 语音客观评价系统训练及应用

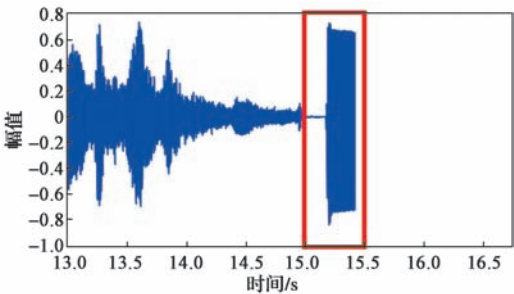
基于 LabVIEW 开发的语音检测系统在处理语音文件中将声音数据存储为波形数组。数组中的一个波形即代表一个特定的通道。波形 Y 方向的数值表示幅值（该幅值由脉冲代码调制数据获得），0 表示静音。

例如 14 号样本语音文件，其与标准文件的差异如图 8 中红色线框所示，此差异是由外界转运车喇叭声所引起的。因样本中包含了喇叭声，从而导致样本文件评价分数较低，为 3.076 分，远远低于预期的 3.3 分。

采集到的语音样本用于 PESQ 模型训练。通过少量的样本数据对结构成熟的模型进行训练，训练过程中将列车显示器测试台的采样参数固定，微调延时参数获得最佳结果。调整过程中，采用无监督的方式进行训练，根据输出的结果与估算结果的误差反向自动调节模型中的部分参数，从而得到一个最优解。当样本量比较大时，迭代速度会很慢，但是获得的结果更贴近实际情况。



(a) 标准语音波形



(b) 14 号样本语音波形

图 8 14 号样本与标准语音文件波形对比
Fig. 8 Waveforms comparison between NO.14 file and standard file

为了将 PESQ 的阈值与人工主观判断的接受标准建立联系，现场质量检查人员对所获得的列车显示器语音样本进行人工主观评价（将不同 PESQ 得分的显示器混在一起），获取平均意见得分（MOS）。结果显示，MOS 得分在 3.0 分以上的，能感受到语音质量有所下降但不影响通话，满足使用要求^[8]。经过测试发现，可接受的 PESQ 得分阈值为 3.3，即客户认为语音质量评价高于 3.3 分的均可接受。表 2 示出样本得分情况。

语音经过 PESQ 软件评价模型处理之后，上位机对被测列车显示器语音进行平均意见得分（MOS）测试和 PESQ 算法打分。查询基于两种算法获取到的评分结果资料，发现客观评价 PESQ 与主观评价的

MOS 得分的平均相关度可以达到 0.935^[9]，由此可见，PESQ 评分方法是现有的基于听觉模型评价算法中效果最好的。通过表 2 可以发现，PESQ 评分与 MOS 评分二者趋势保持一致，两者偏离最大值为 0.14（样本 14），能够满足我们对列车显示语音自动检测质量的要求。

表 2 样本的得分
Tab. 2 Scores of samples

列车显示器采集语音样本	PESQ 得分	MOS 得分
样本 1	3.159	3.059
样本 2	3.266	3.218
样本 3	3.354	3.348
样本 4	3.440	3.470
样本 5	3.361	3.358
样本 6	3.268	3.221
样本 7	3.328	3.309
样本 8	3.311	3.209
样本 9	3.151	3.048
样本 10	3.272	3.227
样本 11	3.373	3.375
样本 12	3.461	3.500
样本 13	3.434	3.462
样本 14	3.076	2.936
样本 15	3.497	3.550

PESQ 评分完成后，列车显示器测试台通过机械手调整显示器上的音量控制按钮，通过上位机软件结合视觉拍照分析，使得列车显示器发出声音的响度一致，列车显示器在该档位发出语音的音量大小既不能低于 50 dB 也不能高于 60 dB，如此即可完成列车显示器完整的语音质量检测。

4 结语

当前基于 PESQ 的评价应用越来越普遍，其将人的视觉、听觉等感知进行模型化、代码化、工具化，通过共享数据以全新的应用模式，构建测试模型，解决了人工评估语音质量带来的效率低、成本高和一致性差的问题，提升了测试效率，保证了测试质量。本文提出一种采用 PESQ 算法对列车显示器语音进行自动检测和评价的方法。该语音质量评估方法可靠性高，一致性好，但仍有以下问题需要解决：

（1）标准语音获取成本较高，语音内容发生改变时，需要重新建立标准语音和 PESQ 阈值，适应性

较差；

（2）列车显示器语音自动检测过程数据记录不完整，只能获取最终的结果，无法有效支撑 PESQ 模型中底层参数的优化与调整，参数的训练和改进都要在后续的工作中持续研究。

参考文献：

[1] 陈华伟. 语音通信中音质客观评价研究 [D]. 成都：西南交通大学，2007.
CHEN H W. Study on Assessment Objective Speech Quality for Speech Communication [D].Chengdu: Southwest Jiaotong University, 2007.

[2] 赵斐，徐勇，成立新. PESQ 及其应用 [J]. 电子设计应用，2003(3): 28- 30.

[3] 付强，易克初，田斌，等. 语音质量客观评价的一步策略 [J]. 电子学报，2001，29(7): 885-887.
FU Q, YI K C, TIAN B, et al. One-Step Strategy of Speech Quality Objective Assessment[J]. Acta Electronica Sinica, 2001, 29(7): 885-887.

[4] BEERENDS J G, HEKSTRA A P, RIX A W, et al. Perceptual evaluation of speech quality (PESQ): the new ITU standard for end-to-end speech quality assessment, Part II- psychoacoustic model [J]. Journal of the Audio Engineering Society, 2002, 50(10): 765- 778.

[5] 陈国，胡修林，张蕴玉，等. 语音质量客观评价方法研究进展 [J]. 电子学报，2001，29(4): 548- 552.
CHEN G, HU X L, ZHANG Y Y, et al. Research Advance on Objective Measures of Speech Quality [J]. Acta Electronica Sinica, 2001, 29(4): 548- 552.

[6] XIANG B, CHEN K, CHI H. Exploration of perceptually-based features for speaker recognition: an empirical study [J]. Chinese Journal of Electronics, 1996, 5(1): 61-62.

[7] 张文莉. PESQ 语音质量评价系统的算法研究与实现 [D]. 大连：大连理工大学，2007: 15-16.
ZHANG W L. Algorithm Research and implementation of PESQ Speech Quality Assessment System[D].Dalian: Dalian University of Technology , 2007: 15-16.

[8] 秦基伟. 语音质量客观评价系统的研究及实现 [D]. 重庆：重庆大学，2013.
QIN J W. Research and Realization of Speech Quality Objective Evaluation System[D].Chongqing: Chongqing University, 2013.

[9] 张军，张德运，傅鹏. 一种改进的心理声学语音质量客观评价算法 [J]. 微电子学与计算机，2007，24(3): 203-206.
ZHANG J, ZHANG D Y, FU P. An Improved Psychoacoustics Speech Quality Evaluation Algorithm[J]. Microelectronics & Computer, 2007, 24(3):203-206.

[10] 张雄伟，陈亮，杨吉斌. 现代语音处理技术及应用 [M]. 北京：机械工业出版社，2003.