

可穿戴式智能音乐纠错系统

叶 阳 张雪凡 刘 源 王 臣 黄 庆

(上海大学通信与信息工程学院 上海 200072)

摘 要: 设计了一种可穿戴式智能音乐纠错系统,能够及时地提醒音乐练习者的演奏错误。该系统能实时采集演奏者弹奏的乐音信息,由短时平均能量法找出乐音端点后对端点间信号做 FFT 和次谐波求和提取基频作为频域特征,将两乐音端点间时间差作为时域特征,通过与标准乐音对比找出演奏过程中的错误所在,系统及时提醒演奏者并将错误记录。硬件方案体积小、低功耗,软件的乐音对比算法简洁高效,经现场的实际检验测试,达到了预想的效果。

关键词: 可穿戴设备;乐音对比;基频提取;音乐纠错;基-2 FFT

中图分类号: TP368.1; TN912.20 **文献标识码:** A **国家标准学科分类代码:** 510.1035

Wearable intelligent music correction system

Ye Yang Zhang Xuefan Liu Yuan Wang Chen Huang Qing

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China)

Abstract: This paper introduces a wearable intelligent music correction system, which can promptly remind the music practitioner's playing errors. The system can collect the music information played by the player in real time, find out the music endpoint from the short-term average energy method and do FFT and sub-harmonic summation on the signal between the end points as frequency domain feature. The time difference between the two tone ends as time domain feature, through comparison with the standard music to find out the error in the process of playing, the system promptly remind player and record error. The system is small, efficient and energy-saving, experiments demonstrate that this system can achieve expected results.

Keywords: wearable devices; music contrast; base frequency extraction; music error correction; radix-2 FFT

0 引 言

音乐从古至今一直都是人们生活不可分离的一部分,在当今社会,乐器的学习成是为一种潮流,然而很多人在自己学习、自己练习时不能发现自己的错误,这就促使了智能纠错系统的诞生^[1]。智能可穿戴设备由于其具有体积小、易携带等诸多优点,已经和诸多领域结合诞生出多种广受好评的产品^[2],智能音乐纠错系统在可穿戴设备上的实现也是趋势所在。但是问题也随之产生^[3],可穿戴设备的体积限制了电池的大小,考虑到实用性,其数据量与处理量要严格控制,同时微型处理器(microprogrammed control unit, MCU)的性能限制了处理速度。因此需要设计合适的硬件结构并选择合适的算法完成纠错功能。

本文所设计的可穿戴式智能音乐纠错系统主要功能如下:练习者在手机端选择练习曲目,手机将练习曲模板信息传给可穿戴纠错系统,在演奏开始时系统开始录入音频,并对乐音实时处理,同时将其与模板对比,当发现音高或节奏

错误时纠错系统向练习者提醒,演奏结束后,将全部演奏信息传输给手机 app 以便以后查阅。

1 算法设计及实现

根据上述场景,系统的工作原理如图 1 所示。由应用场景的特点,以及需在智能可穿戴设备中实现,新的乐音对比算法应该具备以下特点:

- 1) 为满足实时性的要求,算法复杂度要低;
- 2) 为满足乐音音高与音符时长的纠错,算法要同时提出时域与频域的对比方法;
- 3) 针对音乐相对简单,以单音为主,算法针对性较强;

根据以上特点,提出硬件系统条件下有效的乐音对比算法:通过话筒(MIC)输入演奏乐音,由 DSP 提供模数转换,将数字信号存入 Flash 作为待处理信号,分帧后计算出每帧平均幅值,设定平均幅值的阈值找出乐音端点;之后将两端点间信号由 DSP 进行快速傅里叶变换运算,由于频谱同时存在基频与倍频,用次谐波求和法提取信号基频作为

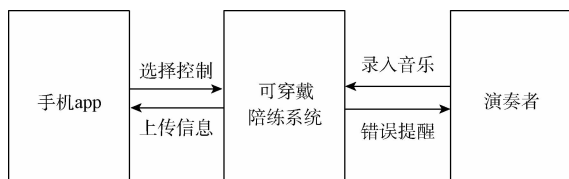


图 1 可穿戴智能纠错系统工作原理

频域特征作为音符音高的对比依据;同时,根据频谱判断两端点间是否有漏检测端点,并由精确检测法精准定位漏掉的乐音端点后将两乐音端点间时间作为时域特征作为音符时长的对比依据;将时频域特征与模板时频域特征实时对比得到对比结果,对比算法如图 2 所示。算法中用 FFT 与次谐波求和法提取基频,确保了算法的低运算量,同时包含了时域与频域的对比,满足之前所述要求^[4]。

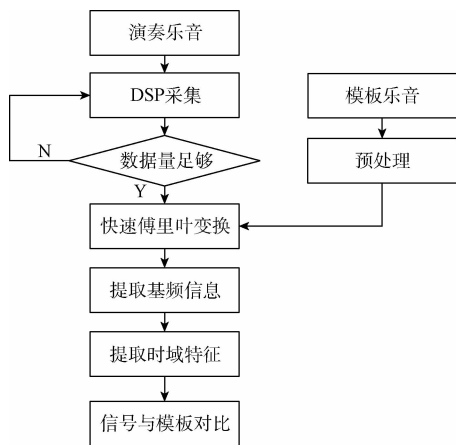


图 2 乐音对比逻辑

1.1 乐音录入与预处理

乐音的录入由 MIC 及放大电路组成,在乐音演奏期间,MIC 持续输入模拟信号,DSP 以固定采样频率 F_s (经典值 44 100)对该引脚模数转换,由 EVA 的事件源启动方式定时启动,将转换结果存至 Flash,等待预处理。

预处理主要包括分帧、加窗、去噪与端点检测^[5]。每当 DSP 采集够等同帧长的采样点数便记为一帧,用矩形窗加窗后去噪,同时帧移选为帧长的一半,保持了乐音的连贯性;去噪是为了去除采集信号中的高频噪音和放大电路的噪声,由 DSP 中设计的 FIR 低通滤波器完成。

想要准确的定位乐音,端点检测是非常重要的部分。大部分乐音在时域幅值上特征明显:乐音的能量集中在它的起始点,因此在乐音的起始点会在时域上有明显的波峰,之后的波形逐渐衰减,利用这个特性可以准确定位大部分的乐音。现在有很多常用的端点检测方法,如短时能量法、自相关法、双门限法等^[6],本文选择计算复杂度低、效果较好的短时平均能量法。

第 n 帧语音信号的短时平均能量定义如式(1)所示:

$$E_n = \frac{1}{L} \sum_{m=0}^{L-1} [x(m)]^2 \quad (1)$$

式中: L 为帧长, $x(m)$ 为第 n 帧信号。设定每帧能量的最小值,通过寻找超过阈值的波峰,将其作为音符的端点进行后面的处理。

1.2 频域特征提取与对比

为了对比乐音的音高是否正确,需提取乐音的频域特征来与模板对比,选择音符的基频作为频域特征^[7]。当 DSP 采集完并将 1 帧数据放入 Flash,对该帧数据做 FFT 获得信息的频谱^[8]。

为了减少存储空间需要与计算量并便于 DSP 实现,选择基-2 FFT 获取频谱^[9]。基-2 FFT 需要变换点数为 2^r ,不满足时需补零来补足点数,为了方便计算与实现,分帧时需满足此条件。基-2 FFT 首先将 N 点 DFT 分为两组计算,按奇偶分别求和,可得到式(2):

$$X(k) = \sum_{r=0}^{\frac{N}{2}-1} x(2r)W_N^{2kr} + \sum_{r=0}^{\frac{N}{2}-1} x(2r+1)W_N^{k(2r+1)} \quad (2)$$

由于 $W_{N/2} = e^{-j2\pi/(N/2)} = (e^{-j2\pi/N})^2 = W_N^2$, 旋转因子 $W_{N/2}$ 以 $N/2$ 为周期,则上式可变换为:

$$X(k) = X_1(k) + W_N^k X_2(k), \quad \frac{N}{2} > k \geq 0 \quad (3)$$

$$X(k) = X_1(k) - W_N^k X_2(k), \quad k \geq \frac{N}{2} \quad (4)$$

其中:

$$X_1(k) = \sum_{r=0}^{\frac{N}{2}-1} x(2r)W_N^{2kr} = \sum_{r=0}^{\frac{N}{2}-1} x_1(r)W_{\frac{N}{2}}^{kr} \quad (5)$$

$$X_2(k) = \sum_{r=0}^{\frac{N}{2}-1} x(2r+1)W_N^{2kr} = \sum_{r=0}^{\frac{N}{2}-1} x_2(r)W_{\frac{N}{2}}^{kr} \quad (6)$$

式(5)、(6)相当于将 N 点 DFT 分成两个 $N/2$ 点 DFT,并做蝶形运算得到,之后再继续分组,将每个 $N/2$ 点 DFT 分成两个 $N/4$ 点 DFT,再做蝶形运算,共做 r 次迭代。那么以 8 点 FFT 为例,其由图 3 所示迭代运算完成。

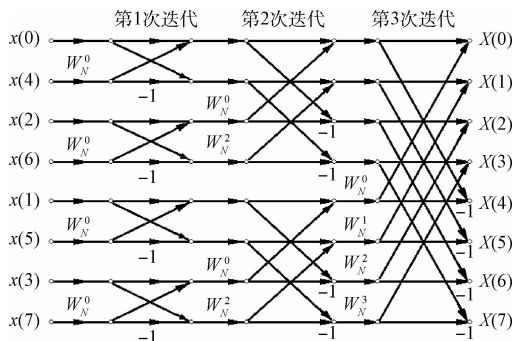


图 3 8 点 FFT 完整流程

由图 3 中可以看出,第 n 次迭代的结果都由第 $n-1$ 次的结果迭代而成,与再之前的值无关。因此可将每次将迭代结果覆盖之前的结果,即共用内存,这样大大减少了所需内存。DSP 实现 FFT 的计算过程如图 4 所示,这样使频谱

的获取更为方便快捷。

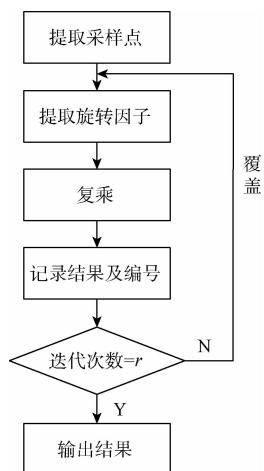


图 4 DSP 实现 FFT 流程

经过 FFT 后得到乐音的频谱,但是单个乐音的频谱往往是由基频及倍频共同组成,无法直接获取乐音的基频,因此用次谐波求和法实现基频提取^[10]。次谐波求和法如式(7)所示:

$$F = \sum_{n=1}^K H_n \cdot B(nf) \quad (7)$$

式中: K 为考虑在内的次谐波的阶次, H_n 为压缩系数,这里取 0.89^{n-1} , $B(nf)$ 为已知频谱的 n 倍压缩。由公式可知该方法相当于对频谱的压缩谱进行加权后求和,通过次谐波与基频的频率关系最终得到基频。软件中将 FFT 得到的 N 个结果数据取出,放入计算数组,记为 $\text{fft_result}[m]$, m 为点数下标,取 $0 \sim N-1$;接着加权计算二次谐波,取 $m/2$ 为整数的点,乘以加权系数 0.89 ,将计算值加进计算数组,记为 $\text{fft_result}[m/2]$, m 取 $0, 2, 4, 6, \dots, N-1$ 。之后的 n 次谐波以此类推,一直加到 K 倍谐波为止。数组 fft_result 为基频提取所需的数组。如图 5 所示, (a) 是一段基频为 246.9 Hz 的乐音信号,虚线部分为变换的部分; (b) 是 FFT 后的结果,由图中可以看出频谱中存在基频、二倍频和三倍频,并且三倍频的幅度最大,无法直接得到乐音基频; (c) 图是经过次谐波求和的结果,由图 5(c) 可看出,经过对 FFT 后的频谱加权压缩求和,幅度值最大的频率点即为乐音信号的基频。至此提取出乐音的基频,将其作为乐高对比的频域特性。

频域特征对比时,首先将模板按照上面方法处理,提取模板的基频信息 f_b 并记录放入内存;之后进行乐音演奏时,实时提取采样的信号帧的基频,记为 f_n ,对比结果的正误用欧式距离来判定,如式(8)所示,通过设定阈值,当结果超过阈值则认为错误。

$$d_{fn} = \sqrt{(f_n - f_b)^2} \quad (8)$$

式中: d_{fn} 为第 n 个乐音频率差的欧氏距离,本系统选用 d_{fn} 为 1.5 Hz 为阈值。

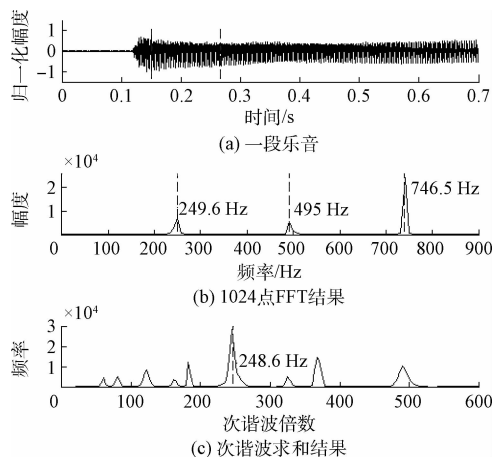


图 5 乐音信号频谱及基频提取结果

1.3 时域特征提取及对比

为了对比乐音音符的演奏时长是否正确,需提取演奏乐音的时域特征与模板对比,本文选择两个音符之间的时间差作为时域特征。但是由于乐音的特性,一些乐音可能会被遗漏:当后一个乐音弹奏较轻,而前一个乐音弹奏较重时,前面乐音的时域波形往往会掩盖后面乐音,导致在端点识别时漏掉后面一个乐音。为防止这种情况出现,在端点识别之后,对频域上显示不止一个频率的乐音分割检测,确定被掩盖乐音的位置,并使用分割检测方法实现。

当发现乐音 FFT 的频谱中存在除单频率及其倍频外的有效频率峰值时,对其进行检测,将该端点与检测出的下一端点间的数据帧作为检测序列,若超过 16 帧则只取前 8 帧,对每一帧做 FFT,精确到某帧起增加了除某频率及其倍频外的其他频率,将其定义为被漏检测的乐音起点,通过这种方法可精确将乐音定义到某帧,将该帧与上一个乐音起点帧间的时间差作为时域特征 t_n 。

当对比时域特征时,不直接使用时间差对比,这是为了根据演奏者自己的演奏速度自适应的调整阈值,如式(9)所示,由时间与模板比值的欧氏距离决定演奏正误:

$$d_m = \sqrt{\left(\frac{t_n}{T_n} - \frac{1}{n-1} \sum_{m=1}^{n-1} \frac{t_m}{T_m}\right)^2} \quad (9)$$

式中: t_n 为第 n 个音符的演奏时间, T_n 为第 n 个音符的模板时间, n 为乐音音符编号。由式中可看出前面演奏时间与模板时间的比值作为参考值,将检测的音符时间与模板时间比值与其欧氏距离作为正误评判的标准,超过阈值认为演奏错误,这样在判断乐音时间正误的同时自适应更新了判断标准。

2 硬件设计

通过对应应用场景的分析和设计并适于算法实现,最终决定硬件设计框图,本系统由主控芯片、处理芯片、输入模块组成,如图 6 所示。

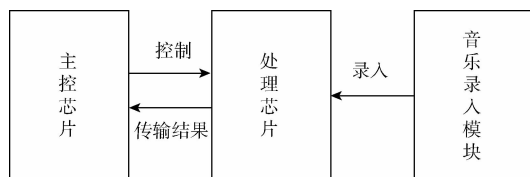


图 6 硬件设计框图

其中由主控芯片负责控制主要可穿戴纠错系统的功能及相关设置,并设定处理芯片相关系数,以按照算法的规定进行运算,同时其需支持蓝牙协议,以便纠错系统与手机的通信;处理芯片负责算法的实现,其先通过音乐输入模块将乐音信息采样并存储,之后通过主控芯片的设置,按照算法及相关参数进行运算,同时将结果传输回主控芯片以便之后相应的操作,其应有高运算度、低功耗的特点;音乐录入模块主要负责接收乐音的模拟信号并加以放大,传输给处理模块供其采样。

根据硬件的系统设计,本系统中主控芯片选用 Nordic 公司的 nRF51822 芯片, DSP 选用 TI 公司的 TMS320F2812 芯片,外设电路主要包括电源电路、JTAG 仿真电路、OLED 显示屏和振动电机,按图 7 所示完成系统

的硬件设计。

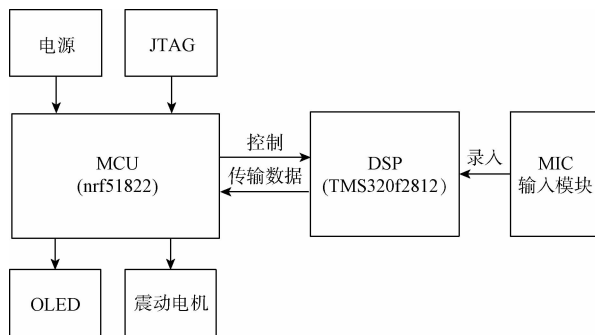


图 7 硬件设计原理

1) 主控芯片: nRF51822 是 Nordic 公司的一款多协议 ARM 内核的蓝牙 4.0 低功耗/2.4 G 专用 RF 芯片,其本身为多协议 2.4G 射频收发器,支持低功耗蓝牙 4.0,内嵌 32 位 ARM® Cortex™M0 内核,十分适用于蓝牙低功耗的设备,结合蓝牙模块与控制芯片一体,节省了硬件空间。nRF51822 最小系统如图 8 所示,包含 nRF51822 芯片与 16M 外部晶振以及蓝牙天线电路^[11]。

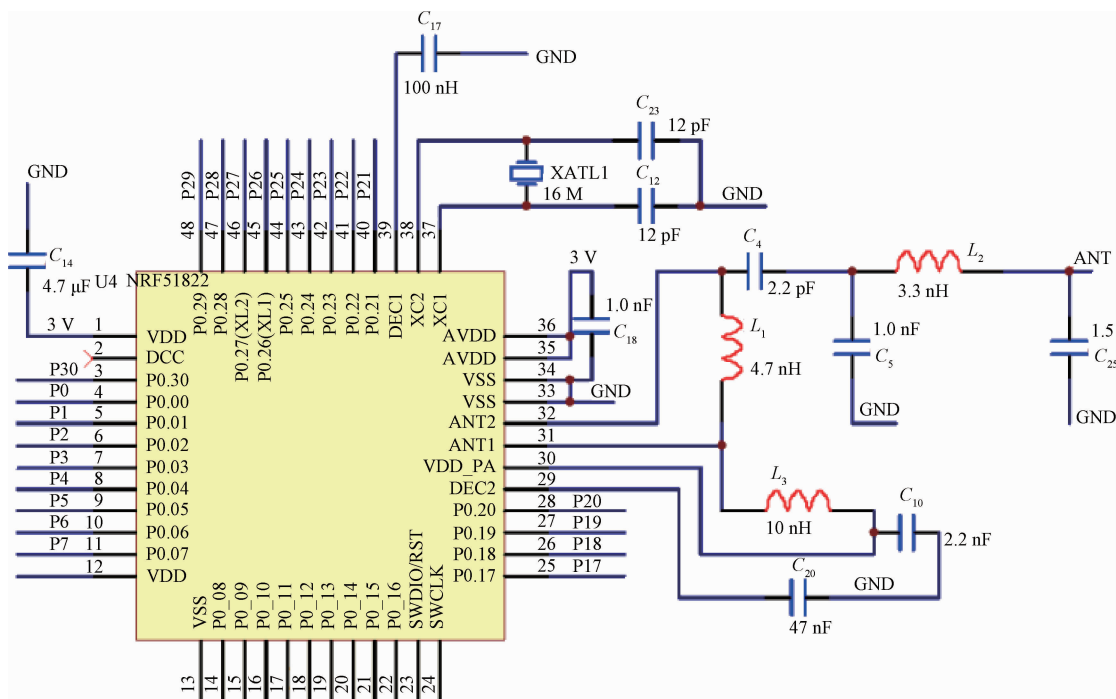


图 8 nRF51822 最小系统

2) DSP 芯片: TMS320F2812 是一款 TI 高性能 TMS320C28x 系列 32 位定点 DSP 处理器,其主频高达 150 MHz,片内 256K 16bit ROM, 34K 16bit RAM, 其片上外设主要包括 2×8 路 12 位 ADC(最快 80 ns 转换时间)、带有两个事件管理模块(EVA、EVB)。该器件还有 3

个独立的 32 位 CPU 定时器,以及多达 56 个独立编程的 GPIO 引脚,可外扩大于 $1 \text{ M} \times 16$ 位程序和数据存储器,可兼顾控制和快速运算的双重功能^[12]。

3) 外设: 外设主要包括 JTAG、OLED、振动电机、MIC 及相关驱动电路。JTAG 为标准测试接口,方便对系统进行仿

真调试;OLED、振动电机主要用来人机交互,为演奏者提供弹奏正误的信息;MIC 及放大驱动电路用来采集演奏乐音,以提供对比信息。其原理图如图 3 所示。由于 MIC 输入信号较微弱,由三极管 9014 构成放大电路对模拟电压信号放大后输入 DSP 的 ADC_IN 引脚,供其采样^[13]。

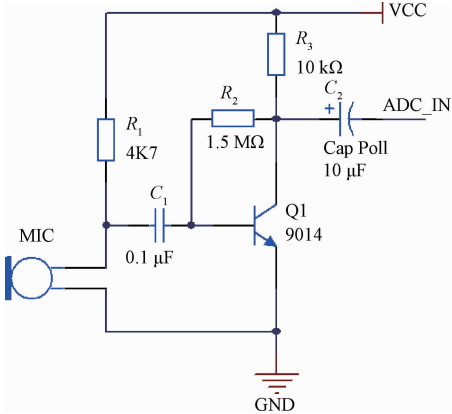


图 9 MIC 放大电路

上述 3 部分共同构成了智能音乐纠错系统的硬件结构,在 MCU 控制下 DSP 不断处理 MIC 录入的信号,完成音乐纠错的功能。

3 性能测试与应用

为测试识别方法的效果与硬件结构的可行性,对系统进行测试。测试乐曲选择吉他简单版“小星星”,在搭建好的硬件平台进行测试。全曲共 34 s,由 70 个音符组成,采样频率 $F_s=44\ 100$,每帧长 512 点,帧移为 384 点,每帧时间为 11.6 ms。从网上下载标准简单吉他版作为模板,用吉他按原版乐谱进行演奏来进行对比纠错。实验共进行两次来验证系统的性能:第 1 次实验不设置错误弹奏乐音,检测正确乐音的误检测率;第 2 次实验设置一定量错误弹奏,来检测错误乐音的检测率。

第 1 次实验共进行 10 次演奏,每次均全部正确演奏,每次发现错误,电机振动提醒错误, OLED 同时显示错误总数。实验结果如表 1 所示。

表 1 第 1 次实验结果

次数	误检测错误数	正确率/%
1	3	95.8
2	2	97.2
3	4	94.3
4	2	97.2
5	1	98.6
6	3	95.8
7	4	94.3
8	2	97.2
9	1	98.6
10	2	97.2
平均	2.4	96.5

由实验结果可以看出,本系统的乐音检测精度达到 96.5%,在误检测出的 24 次错误中,其中有 8 次是由于频率误差在 1.5 Hz 阈值之外而判断错误,在中高频中相邻两个音高的频率间隔会逐渐增大,因此可以根据不同的频率范围选择不同的误差阈值来减少这种情况;有 13 次是由与两个相同频率的乐音,后一个乐音被淹没造成,还有 3 次是基频提取中结果错误导致,说明乐音纠错算法还可以进一步提高。

第 2 次实验共演奏 10 次,每次弹奏设置 20 个错误乐音,当检测出弹奏错误,震动提醒并在 OLED 端显示记录。实验结果如表 2 所示。

表 2 第 2 次实验结果

次数	误检测错误数	正确率/%
1	1	95
2	0	100
3	2	90
4	1	95
5	0	100
6	2	90
7	0	100
8	1	95
9	0	100
10	2	90
平均	0.9	95.5

由结果可知,错误检测的正确率达 95.5%,9 次错误均为弹奏的错误乐音为正确乐音的倍频时,错误识别造成的,说明检测效果较好,但基频提取部分可进一步改进。

由两次实验结果,该系统在正确音符的判断中正确率为 96.5%,在错误音符纠正的正确率为 95.5%,主要错误出现在基频小几率的判断错误和超过判断阈值,说明系统算法中次谐波求合法压缩系数 H_n 与时频域的阈值门限有待进一步精确,系统达到了预期的效果。

4 结 论

本文设计了一种智能可穿戴乐音纠错系统,根据智能与可穿戴的要求设计出系统硬件,并选择适合硬件平台的乐音算法在系统中实现。经试验,该系统工作稳定,乐音识别效果较好,能满足实时性与纠错率的要求,硬件方案解决为可穿戴智能纠错系统提供了可参考方案,在硬件基础上设计了适合可穿戴设备的乐音对比算法,在运算量小易实现的基础上完成对音高和音符节奏进行纠错。整个系统为提供了一种新型智能可穿戴纠错系统,为音乐爱好者提供了一种智能化辅助产品,解决了智能陪练的问题,该系统,目前已经应用在实际的弹奏实验中。

参考文献

- [1] 梁文彬, 张帆, 程京, 等. 嵌入式实时音乐语音识别系统的实现[J]. 微计算机信息, 2008, 24(19): 252-253.
- [2] 耿怡, 安晖, 李扬, 等. 可穿戴设备发展现状和前景探析[J]. 电子科学技术, 2014, 1(2): 238-245.
- [3] KLAUSNER A, TRACHTENBERG A, STAROBINSKI D, et al. An overview of the capabilities and limitations of smartphone sensors[J]. International Journal of Handheld Computing Research, 2013, 4(2): 69-80.
- [4] 徐鹏进, 郭莉, 刘书昌. 基于音高与端点联合检测的音符识别算法[J]. 计算机应用, 2011, 31(S2): 172-175.
- [5] SAZAKI Y, AYUNI R, KOM S. Musical note recognition using minimum spanning tree algorithm[C]. International Conference on Telecommunication Systems Services and Applications, IEEE, 2014: 1-5.
- [6] JALIL M, BUTT F A, MALIK A. Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals [C]. International Conference on Technological Advances in Electrical, Electronics and Computer Engineering, 2013: 208-212.
- [7] 吴晶晶, 孟玲玲. 乐音音符基频识别[J]. 电子测量技术, 2009, 32(4): 126-128.
- [8] 刘美容. FFT 算法的 DSP 实现[J]. 微电子学与计算机, 2015(1): 76-79.
- [9] 万浩平, 马进, 王锋. 基于 TMS320F2812 的高精度数据采集及 FFT 实现[J]. 工业控制计算机, 2009, 22(4): 54-55.
- [10] CHENG T, XU W, TIAN Y, et al. Extracting singing melody in music with accompaniment based on harmonic peak and subharmonic summation[C]. IET International Conference on Wireless, Mobile & Multimedia Networks, 2011: 200-205.
- [11] FENG Z, MO L, LI M. Analysis of low energy consumption wireless sensor with BLE[C]. Sensors, 2015: 1-4.
- [12] 李云红, 李子琳. 基于 DSP 的语音识别系统设计[J]. 纺织高校基础科学学报, 2012, 25(1): 107-110.
- [13] 高宝成, 宋玉珊. 基于 NRF24L01 的无线语音采集系统设计[J]. 电子测量技术, 2012, 35(8): 103-106.

作者简介

叶旻, 硕士研究生, 主要研究方向为语音识别、嵌入式系统。

E-mail: 416457776@qq.com

张雪凡, 高级工程师, 硕士生导师, 主要研究方向为无线通信、微弱信号处理、电路设计。

E-mail: 10002461@shu.edu.cn

刘源, 硕士研究生, 主要从事研究方向为微弱信号处理

E-mail: timliuyuan@163.com

王臣, 硕士研究生, 主要研究方向为信号处理。

E-mail: wangchen@163.com

黄庆, 硕士研究生, 主要研究方向为微弱信号处理。

E-mail: abchq1314@163.com