

北京航空航天大学

第二十二届“冯如杯”

学生创意大赛竞赛论文

项 目 编 号 _____
院（系）名 称 _____
专 业 名 称 _____
作 者 姓 名 _____
学 号 _____
指 导 教 师 _____

2012 年 3 月

识别人声及其他声音并 调节音量的应用

摘要:

我们在生活中常常遇到与声音大小有关的各种不便,比如听摇滚歌曲时遇到背景音乐太吵掩盖了歌手声音,但又不能随意的调整背景音乐与歌手演唱声音的相对大小;在吵闹的环境中打电话时喊得声嘶力竭但对方还是听不见;某些“懒人”想让电视也能根据播放的节目自动调节音量大小,放战争片时声音大些,转播音乐会时声音小些……

这样的例子不胜枚举,尽管如今语音识别技术、说话人识别技术等声音识别技术已经十分发达,但重在识别语意,却忽略了对人声和其他声音的识别。我们的这个应用便是在识别人声和其他声音方面的探索,希望用识别声音“种类”的方法解决一些生活中的不便。

关键词:

声音识别 手机应用软件 傅里叶变换 音量调节

Abstract:

In our life, we often encounter various inconveniences related to the loudness of sound, for example, when we listen to a rock song, the background music is sometimes too noisy to make us hear the singers' voice, but we are unable to adjust the background music and the singers' sound separately; when we speak phones in a noisy environment, the noise always make us mad because we can't make ourselves heard in such a noisy environment; some "lazy man" are always waiting for a kind of television which can adjust the loudness according to the programs that are broadcasted

we can find countless such examples if we want. Although now the voice recognition technology, speech recognition technology and voice recognition technology have been highly developed, they are all focused on identifying semantic, but ignored the identification of the human voice and other sounds. This application we made is based on "identifying the human voice and other sounds" to solve these inconveniences mentioned above.

Keywords:

Voice recognition; Phone application software; Fourier transform; Adjusting the volume

目录

摘要：	1
关键词：	1
第一章 引言	2
第二章 理论背景	3
1. 声音的掩蔽效应	3
2. 人耳的听觉的频率范围和说话的频率范围	3
3. 鸡尾酒会效应	3
第三章 软件的组成和各部分的功能描述	4
1. 通话模式（详述）：	4
软件功能	4
软件的组成	4
1. 声音信号的输入部分	5
2. 声音信号的分析分离部分	5
3. 声音信号的调控部分	5
软件设置部分	6
2. 音乐模式（略述）：	6
3. 嵌入式（略述）	6
第四章 可行性分析及主要难点	6
第五章 应用前景及市场需求	7
第六章 结语	7
[参考文献]	7

第一章 引言

我们在生活中常常遇到下列种种不便：

•用手机戴耳机听某些摇滚歌曲的时候，因为歌曲背景的打击乐器、吉他音量相对来说特别大，常常掩盖了歌手的歌声，导致需要调大音量才能听得清楚，但因为耳机贴近耳朵，过大的声音又容易伤害耳朵，所以我们希望在不提高背景音乐音量的情况下调高人声的音量，这样既听得清楚又不伤耳朵。



•在一些嘈杂的环境中打电话，因为背景的噪音太大，常常要喊才能让对方听清楚，所以我们希望尽量过滤背景噪音，让电话那头的人只听到人的声音。



综上所述，因为声音性质的不同，人们对它们的忍受力是不一样的；在不同的场合，人们需要接收的声音大小也不一样。在上述情况中，人都遇到了声音大小不适合人们的需要而造成的不便，而且这个时候，普通的“调节音量大小”的方法却没什么用。

第二章 理论背景

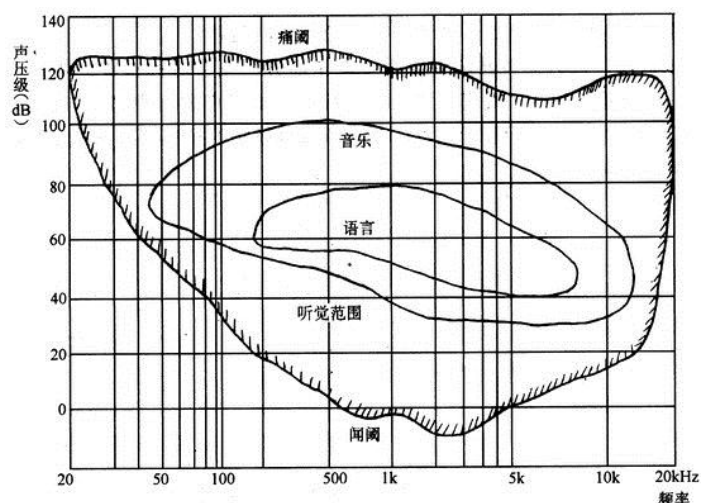
为找到解决所遇到问题的方法，我在网上以及图书馆查阅了一些相关资料，了解到了以下声学知识。

1. **声音的掩蔽效应：**人们在安静环境中可以听到比较小的声音，有时甚至可以听到机械手表的滴答声，但是，在倾听一个比较大的声音的同时，如果存在另一个声音就会影响到人耳对所听声音的听闻效果。这种两个响度不等的声音作用于人耳时，响度较高的频率成分的存在影响对响度较低的频率成分的感受，使其变得不易察觉的现象就叫做“掩蔽效应”。

2. **人耳的听觉的频率范围和说话的频率范围：**

人耳听觉的频率范围虽然会随着因体质的不同而有所区别，但正常人的听觉范围一般都在 20Hz 到 20000Hz 之间；而人说话的声音频率范围则是 300Hz 到 3400Hz。

（见右图）



3. **鸡尾酒会效应：**在嘈杂的环境中（比如在有很多人的鸡尾酒会上）人类可以把自已的听力集中在和某一个人的谈话上，而把其他人的声音都推到背景杂音中去，这是人脑的一种高级功能，这就叫做鸡尾酒会效应。

所以，我们遇到的问题实际上是由于声音的掩蔽效应引起的，而且掩蔽效应有其特点，掩蔽的声音响度越大，被掩蔽的声音就需要提高更大的音量才能被听到；掩蔽的声音和被掩蔽的声音频率越接近其掩蔽效果越强，而且低频声音对高频声音的掩蔽作用尤其明显。那么，对于一段声音，在将人说话声音频率范围外的 20Hz 到 300Hz 和 3400Hz 到 20000Hz 之间的声音响度降低，减少了它们对人听音的干扰，再针对 300Hz 到 3400Hz 范围的声音做一些处理，降低这段频率中人声范围外的声音对人正常收听人声的影响，由于鸡尾酒效应，人耳是有能力从一段含有不太强的噪音的声音中分辨出人声的。

我们决定设计一款软件，“智能”地识别分离不同的声音并根据需要调控声音的大小，强化人声，减低噪声。软件的原理就是利用声音的特点，采用合理的算法，将人的声音与其他的声音识别、分离开来，这样就可以单独调节人声（歌手的声音）与其他声音（噪声和歌曲的背景音乐）的大小，以解决上面遇到的不便。因为大一所学知识有限，我没有能力编写出这个程序，下面是我对这种程序的理论构想。

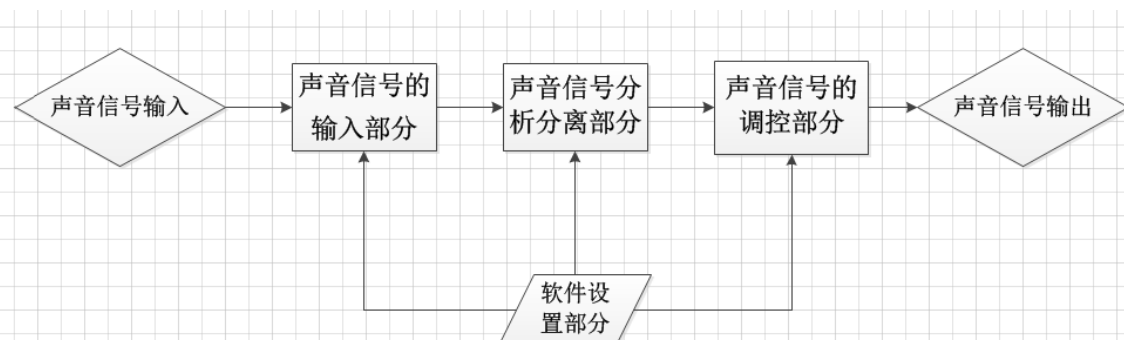
第三章 软件的组成和各部分的功能描述

因为上面遇到的问题都是在使用手机的时候遇到的，所以我们的软件是应用在智能手机上的，如果有需要也可以移植到其他平台如 PC 上，也可以利用嵌入式技术用在电视机上。下面先以安装在智能手机平台上的软件在降低通话时的背景噪音（通话模式）和 听歌时调节歌手的声音与背景歌曲的声音（音乐模式）还有 移植到其他电器上（嵌入式）方面的功能举例说明这款软件的组成及详细工作原理。

1. 通话模式（详述）：

软件功能：对于一段声音，软件要能识别出其中人声音的特征波形，并将之从其他声音中分离出来；对于因为声音中有与人声频率太过接近而技术上难以完全分离的声音，需要过滤掉其中影响人正常收听这段声音的噪音（比如响度太高或者频率太高而使人感到不适的声音）。

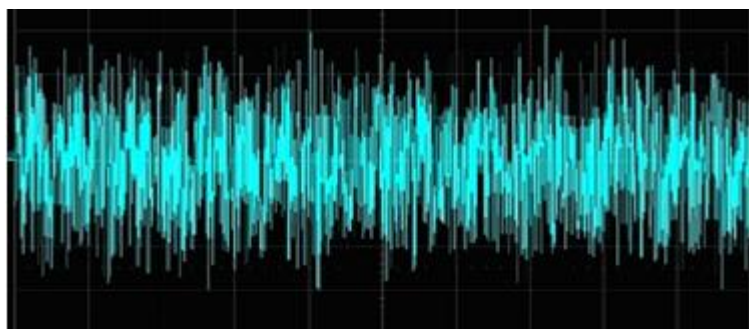
软件的组成： 声音信号的输入部分 声音信号的分析分离部分 声音信号的调控部分 软件设置部分



软件结构图

1. 声音信号的输入部分：软件获取麦克风收集声音信息转化成的数字电信号（GSM 手机编码数字信号的方式：1、话筒把话音转化为模拟电信号 2、对于这个模拟信号，首先将它分成 20ms 一个的语音块，每个块用 8KHZ 抽样，再用 13bit 进行量化，量化后再加 3 个 0 变成 16bit，现在的码速率为 128kbps，然后进行信源编码，变成 13kbps 的数字信号），然后进行 DA 转换得到声音的时域波形，再利用快速傅里叶变换（FFT）得到声音信号的频域波形信号，再由分析分离部分进行处理。

2. 声音信号的分析分离部分：首先将声音的频域波形中 20Hz 到 300Hz 和 3400Hz 到 20000Hz 的信号分离，对 300Hz 到 3400Hz 的信号，将人声和噪声的波形利用数学算法分离。（此时所得到的信号不能直接处理，因为软件只是将人声和噪声分离，但并没有将它们识别出来）接下来将分离后的信号与数据库中储存的人声信号进行对比（这些数据库中的人声信号可以通过统计研究获得，如果要加强精度，在使用软件之前，手机使用者可以在安静的环境中说出预先设计好的一段文字，以此获得手机使用者更详细的语音信息，减小因为不同人声音差异造成的数据误差）现在科学家们对于语音识别以及说话人识别方面已经进行了很多研究，我们的软件完全可以借用他们的研究成果，利用他们的技术识别人声中的特征波形，作为人声识别



一段噪音的波形



一段人声的波形



一段鸟鸣声的波形

的辅助。此外生活中的很多噪声比如机械工作时的噪声和汽车的喇叭声，它们的声是有一定规律的，从声音的时域波形来看，这种噪声的波形是关于时间的周期函数，对于这样的噪声可以很容易的分离出来。

3. 声音信号的调控部分：在将人声和噪声分离之后，软件将适当的调整几种声音的音量。对于人声，可以把它的音量适当调高，对于噪声，则可以适当调低（之所以不将噪声的音量完全消除是因为背景噪声复杂多变，软件很难把人声与噪声

完全分离，所以如果将识别为“噪声”部分的声音完全消除，可能会损害人声部分的音质，造成声音的失真，影响通话质量；加上鸡尾酒效应的存在，人耳有能力在有背景噪音的声音中识别出自己需要的信息，因此也不必将噪音完全消除，而只需要降低其音量就可以提高通话的质量了。）然后将信号输出到手机的通信部分，再经调质、发射、传输、对方接收、解调、收听，完成整个通话过程。

4. 软件设置部分：因为这个软件是用在手机通话时减低背景噪音的方面，必须使软件运行的速度足够快，不至于因为降低噪音的过程花去太多时间而造成通话延迟，影响通话质量。除了优化算法外，也可以在一些配置比较低的手机上适当降低除噪标准来避免这样的问题。软件设置部分就是软件的 UI (User Interface, 用户界面)，以供用户设置软件参数，如软件开关，除噪等级等，以使用户获得更好的使用体验。

2. 音乐模式（略述）：

这时软件是用来调节歌手的声音与背景歌曲的声音。软件的工作过程与打电话时几乎一样，软件获得声音信息，经过分析处理再输出，稍微不同的是这时获得声音的来源不是手机麦克风，而是由播放器软件读取储存在手机上的歌曲的声音信息；需要分离的不再是人声与背景噪音而是歌手的声音与乐器的声音。其实相对于通话时的背景噪音，因为歌手的声音和背景音乐都是有固定旋律的，而且处理过程没有实时性的要求，所以对人声和乐曲声的识别分离可以十分精确，这样也可以满足对于保留音乐音质的要求。软件音乐模式方面的功能还可以移植到 PC 平台上。

3. 嵌入式（略述）

在看电视的时候，前几分钟看的是摇滚音乐会转播，因为音乐声音太大，所以把电视的音量调小了，然后调台到了某个谈话类或是新闻类节目，声音又显得太小而听不清了，电视的调节音量的功能只是按比例放大扬声器的声音，却忽视了人耳对同样大小的不同种声音的分辨与忍受能力的差异。这时某些“懒人”又希望电视能根据声音的类型，自动调节音量的大小。我们可以利用嵌入式技术，将软件植入到电视的音响模块中，只利用软件“识别”方面的功能就可以智能识别节目的类型，智能调节音量。

第四章 可行性分析及主要难点

功能实现的可行性：基于人脑的鸡尾酒效应以及人声的频率特点，只要选用适当的算法以及通过实验获得详尽的人声数据，利用通信领域中信号除噪、分离的技术，完全可以实现软件设计中的人声与其他声音检测、分离的功能。

主要技术难点：1.人声数据的获取。这得需要在前期进行详细的调查实验，通过大范围的针对说同一种语言的人口进行抽样调查，通过统计学技术获得人声频率特点。不过因为科学家对人类语义识别的研究已经进行了几十年，早已积累了很丰富的实验数据，我们可以借鉴他们的数据，这样可以大大减少获得人声频率数据的难度。2.将人声与其他声音识别并分离的

算法的编制。这需要应用大量的统计学理论知识,还有现代数据分离常用的快速傅里叶变换。用于作者知识有限,这方面难以做出专业解答,但是基于现在人类的计算机软件水平,肯定有算法可以解决所遇到的问题。

第五章 应用前景及市场需求

虽然这个应用并不算手机的必须功能,但是在改善手机的使用体验方面可以有不错的效果,而且可以将软件集成到芯片里,做成嵌入式设计,还可以用到电视机等电器上。未来前景十分广阔。

第六章 结语

由于大一知识所限,我们没能做出这个应用,只是提出了软件的设计构想,但我们相信这个构想是可行的,希望以后再学到相应的专业知识后我们能做出这个软件。这个软件的应用范围应该也不止文中提到的这么多,相信还有很多方面我们能用到这个软件

[参考文献]

[1]杨行峻 迟慧生等 《语言信号数字处理》 电子工业出版社 1995

[1]钱同惠 《数字信号处理》 机械工业出版社 2005