

Characteristic Parameters Analysis in Query by Humming Based on Praat Software

Fang Wang¹, Qingliang Meng²

¹College of Sciences, Yanshan University, Qinhuangdao

²College of Information Science and Engineering, Yanshan University, Qinhuangdao

Email: 287949498@qq.com

Received: May 19th, 2014; revised: May 22nd, 2014; accepted: May 30th, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Humming retrieval technology is a content-based music retrieval. In humming retrieval, it is essential to select the sound characteristic parameters, which will directly affect the final result of the search. Through Praat speech analysis software, different computer simulation experiments were carried out respectively to analyze the four characteristic parameters (energy, formants, MFCC-Mel Cepstral, pitch) of sound. By contrasting the results of these experiments, those characteristic parameters that can meet the humming retrieval request were selected.

Keywords

Praat Voice Analysis Software, Query by Humming, Characteristic Parameters, Formant, Pitch

基于Praat软件的哼唱检索中参数选取分析

王 芳¹, 孟庆良²

¹燕山大学理学院, 秦皇岛

²燕山大学信息科学与工程学院, 秦皇岛

Email: 287949498@qq.com

收稿日期: 2014年5月19日; 修回日期: 2014年5月22日; 录用日期: 2014年5月30日

摘要

哼唱检索技术是一种基于内容的音乐检索技术。在哼唱检索技术中，声音特征参数的选取至关重要，它直接影响检索的最终结果。文章通过Praat语音分析软件，对声音的四种特征参数(能量、共振峰、MFCC-Mel倒谱系数、基音)分别进行了不同的计算机模拟实验，最终通过比对选取出符合哼唱检索要求的特征参数。

关键词

Praat语音分析软件，哼唱检索，特征参数，共振峰，基音

1. 引言

声音信号的感知过程[1]与人耳的听觉系统是密不可分的，正常人可以依据听到声音的不同，来区分出是什么人在说话、说的什么内容以及在用什么样的语调说等信息，但计算机本身并不具备这种特性，这就需要人们把声音信号分析转换成其本质特征的参数来进行区分。

从物理角度及发声原理来分析，声音的性质可由决定声音效果的四个要素来描述，即音高、音量、音长及音色[2]。音高(Pitch)主要由物体振动的频率决定；音量主要由物体振动的振幅大小来决定；音长主要由物体振动持续的时间长短来决定；音色(Timbre)是指声音的感觉特性，主要由泛音的多少及泛音间的相对强度决定。上述四种性质中，前三种性质里提到的“振动”主要是指基音的振动。

在乐音体系中，并不是所有的音都能作为乐音来使用，通常把有着固定音高的音叫做乐音，把音高不固定或音高不明显的音叫做噪音。根据对声音性质的分析及对乐音的理解，本文通过分析几种常见的声音特征参数曲线：能量曲线、共振峰曲线、MFCC 曲线、基音曲线，来选取符合要求的特征参数，每个参数分别做五个计算机实验，对实验结果进行参数分析。

2. Praat 软件介绍

Praat 语音学软件(原名 Praat: doing phonetics by computer)，是一款由荷兰阿姆斯特丹大学人文学院语音科学研究所的主席保罗·博尔斯马(Paul Boersma)教授和大卫·威宁克(David Weenink)副教授合作开发的跨平台多功能语音学专业软件。Praat 的主要功能是对自然语言的语音信号进行采集、分析和标注，并执行包括变换和滤波等在内的多种处理任务。作为分析结果的文字报表和语图，不但可以输出到个人计算机的磁盘文件中和终端的显示器上，更能够输出为精致的矢量图或位图，供写作和印刷学术论文与专著使用[3]。

3. 实验要求及内容

在基于哼唱的音乐检索中，选取的声音的特征参数要满足：不管什么人唱、唱什么内容，只要音调相同，特征参数就应相同。具体要求如下：

- 1) 不同的人、相同的谱、相同的词，特征参数要相同；
- 2) 相同的人、不同的谱、相同的词，特征参数要不同；
- 3) 相同的人、相同的谱、不同的词，特征参数要相同；
- 4) 相同的人、不同节奏、相同谱、相同的词，特征参数要相同；
- 5) 相同的人、相同的谱、相同的词、不同的音量，特征参数要相同；

按以上要求,通过 Praat 语音分析软件[4]-[6],对声音的特征参数(能量、共振峰、MFCC-Mel 倒谱系数、基音)分别做以下五种计算机模拟实验:

实验一:两个人分别用“F4”调,在相同的时间内来哼唱“爱”这个词;

实验二:哼唱《离岛》中的一句“距离很好”,把这段音频记为 A;在节奏不变的前提下,在 A 的基础上把音调整体升 4 个半音,这段音频记为 B;同样在节奏不变的前提下,在 A 的基础上把音调整体降 5 个半音,这段音频记为 C;

实验三:同一个人,分别用“D4”调及相同的节奏来哼唱“爱”与“狂”;

实验四:哼唱《海阔天空》中的一句“海阔天空”,把这段音频记为 A;在音调不变的前提下,在 A 的基础上把节奏整体加快 20%,这段音频记为 B;同样在音调不变的前提下,在 A 的基础上把节奏整体减慢 20%,这段音频记为 C。

实验五:哼唱《海阔天空》中的一句“海阔天空”,把这段音频记为 A;在其他条件都不变的前提下,在 A 的基础上把音量调高 20%,这段音频记为 B;在其他条件都不变的前提下,在 A 的基础上把音量降低 20%,这段音频记为 C。

4. 实验与分析

4.1. 能量

能量主要反映的是声音音量这一性质,本文研究的能量主要指音强。理论上,它只与声音的大小即声源振动的幅度有关(振动幅度越大音量越大),与声音的音调高低无关。研究语音信号中声学的性质时,音强的因素是不可缺少的,在探讨某些发声情况(如探讨朝鲜语中的紧辅音和松辅音的区别)时,音强可以起辨别作用[7]。元音音色的不同实际上就是不同频率段上音强的大小不同,但在语言学中,音强的作用十分有限,声音的响度是音强、音高、音长、音色的综合表现,而音强往往不起主导作用[8]。而声音强度(I)与声压(p)的关系如下:

$$I = (p^2) / (\rho v) \quad (1)$$

其中 ρ ——介质密度, v ——声速。

图 1 是由实验一得出的能量曲线,其中横坐标表示时间(s),纵坐标表示能量值(dB)。通过对图 1 中(a)、(b)两个分图的比对可以看出:不同的两个人,用相同的音调、演唱相同的歌词时,能量曲线不相同且变化轨迹不相似。这说明,不同的人在不同的条件下,能量是不同的。这不符合哼唱检索对特征参数的要求。

4.2. 共振峰

共振峰[9]是指声音的频谱中能量相对集中的一些区域,它是反映声道谐振特性的重要特征。理论上,在同一件乐器所发出的所有音调中,不论基音频率如何,都会表现出相同的共振峰,只不过有着强弱的区别,在语音声学中,共振峰决定着元音的音质。人们发音时会受到自己本身生理特征的影响,有着属于自己的共振峰区(Formant Regions),例如会受到咽腔、口腔大小及鼻孔的影响,并且可以通过改变它们的形状及大小,就可以改变声音的共振峰。而在计算机音乐中,共振峰是决定音色和音质的重要参数。在语音科学及语音学中,描述的是人类声道中的共振情形[10]。常用的量测方法是由频谱分析或声谱图中,寻找频谱中的峰值,本文采用的是频谱分析法,其中频率最低的共振峰频率称为 F1,第二低的是 F2,而第三低的是 F3 等[11]。

图 2、图 3、图 4、图 5、图 6 分别是实验一、实验二、实验三、实验四、实验五得出的共振峰曲线图,

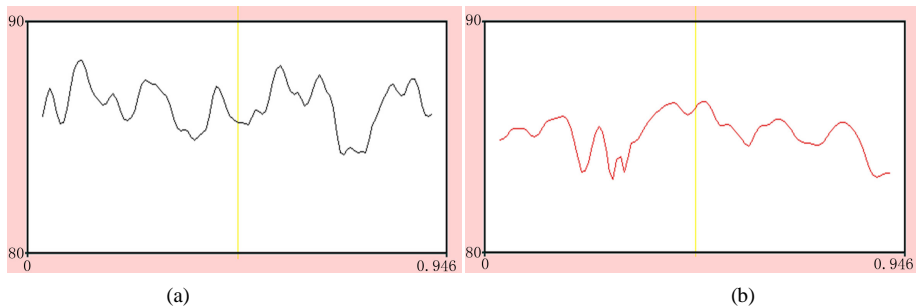


Figure 1. Energy curve
图 1. 能量曲线

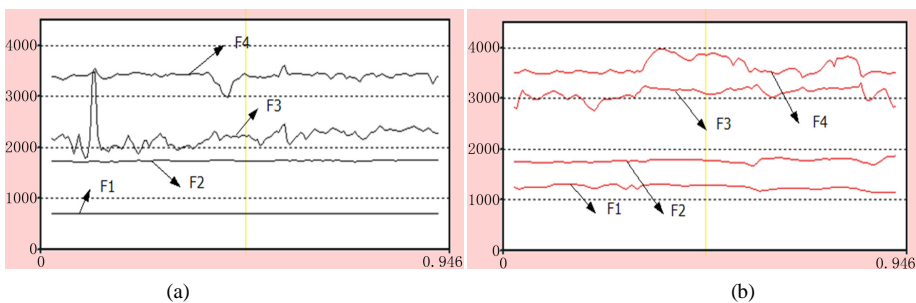


Figure 2. Formant curve of experiment one
图 2. 实验一的共振峰曲线

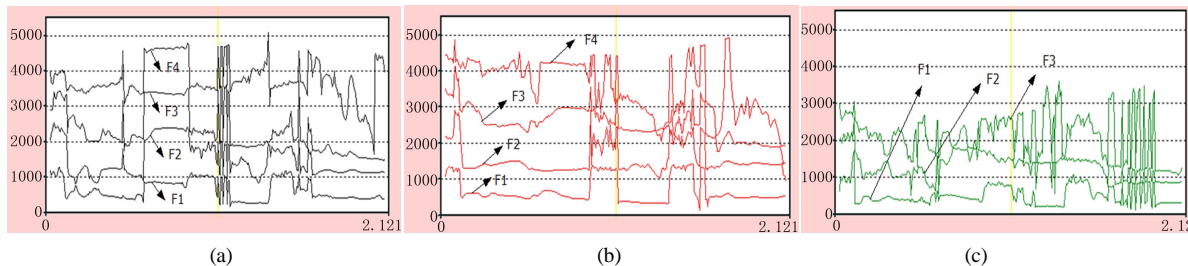


Figure 3. Formant curve of experiment two
图 3. 实验二的共振峰曲线

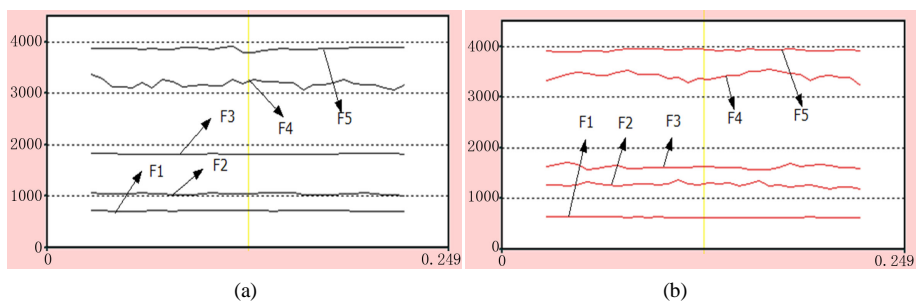


Figure 4. Formant curve of experiment three
图 4. 实验三的共振峰曲线

其中横坐标表示时间(s)，纵坐标表示共振峰频率(Hz)，图中曲线 F1、F2、F3、F4、F5 分别表示不同的共振峰轨迹。理论上共振峰特性主要反映在 F1、F2 两个共振峰，F1 与舌位高低密切相关，舌位越高 F1 越低，舌位越低 F1 越高；F2 与舌位前后密切相关，舌位越靠前 F2 越高，舌位越靠后 F2 越低。

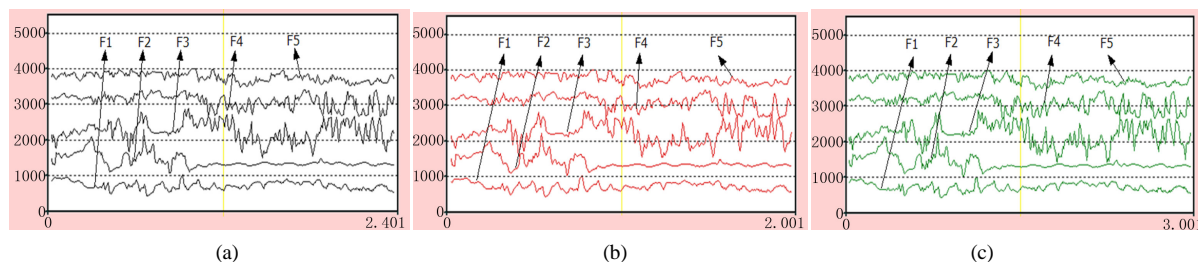


Figure 5. Formant curve of experiment four

图 5. 实验四的共振峰曲线

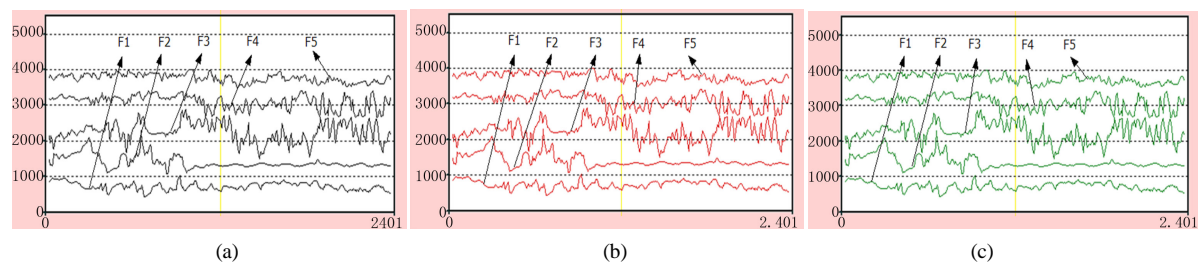


Figure 6. Formant curve of experiment five

图 6. 实验五的共振峰曲线

通过对图 2 中(a)、(b)两个分图的比对可以看出：在相同的参数设置下，不同的人，在用相同的音调、演唱相同的歌词时，产生的共振峰个数是相同的，但只有第二共振峰的轨迹变化相似，其余三条轨迹变化都不相同且不相似。这只能说明他们发音时舌位的前后位置的变化是相似的，并不能说明此时的音调是否相同。

通过对图 3 中(a)、(b)、(c)三个分图的比对可以看出：在相同的参数设置下，同一个人，用不同的音调、演唱相同的歌词时，产生的共振峰个数是不相同的，并且相应的共振峰的变化轨迹是不相似的。这只能说明，此条件下得到的共振峰的变化轨迹是无规律的，不能反映出音调高低变化的不同。

通过对图 4 中(a)、(b)两个分图的比对可以看出：在相同的参数设置下，同一个人，用相同的音调、演唱不同的歌词时，产生的共振峰个数是相同的，但只有 F1 和 F5 的轨迹变化相似，并且 F1 所处的位置是不同的。这只能说明，两种情况下舌位高低的变化轨迹是相似的，但不能反映出音调是否相同。

通过对图 5 中(a)、(b)、(c)三个分图的比对可以看出：在相同的参数设置下，同一个人，用相同的音调、不同的节奏、演唱相同的歌词时，产生的共振峰个数相同，同时它们各自的共振峰的变化轨迹是相似的，并且当把所有共振峰的变化轨迹在时间轴上线性延展到相同的长度时，它们各自所对应的共振峰的变化轨迹是完全相同的。这说明，在音调相同的情况下，演唱节奏的快慢对共振峰的变化轨迹无影响。

通过对图 6 中(a)、(b)、(c)三个分图的比对可以看出：在相同的参数设置下，同一个人，用相同的音调、不同的音量、演唱相同的歌词时，产生的共振峰个数相同，并且它们的共振峰的变化轨迹是完全相同的。这说明，在音调相同的条件下，音量的高低对共振峰频率值无影响。

理论上，共振峰是决定音色和音质的主要参数，它与音调的高低并无直接联系。上述五个实验结果与理论相符，共振峰在音量及音调相同的条件下，能区分出不同的人；在同人同音调的条件下，能区分出不同的哼唱内容。这与哼唱检索对特征参数的要求正好相反。

4.3. MFCC-Mel 倒谱系数

Mel 频率与 Hz 频率成非线性的对应关系，它是基于人耳听觉特性提出来的，基于 Mel 频率的倒谱系

数(Mel frequency cestrum coefficient, MFCC)是利用它们之间的这种关系计算得到的。基于 Mel 频率的倒谱系数是将人耳的听觉感知特性和语音的产生机制相结合的[12],这种特性目前被广泛应用于大多数的语音识别系统中,但它同样适用于音频信息处理中。理论上, MFCC 是模拟人耳听觉特性的,它主要用来区分不同的人及同一人说的不同内容。MFCC 与线性频率的转换关系是:

$$f_{mel} = 2595 \lg \left(1 + \frac{f}{700} \right) \quad (2)$$

图 7 是实验一得出的 MFCC 曲线图,其中横坐标表示时间(s),纵坐标表示幅度(Hz)。通过对图 7 中(a)、(b)两个分图的比对可以看出:不同的两个人,用相同的音调、演唱相同的歌词时,他们的 MFCC 曲线是完全不相似的。这说明, MFCC 在条件相同的情况下,能区分出不同的人,这与理论相符,但不符合哼唱检索对特征参数的要求。

4.4. 基音

当发声体由于震动而发出声音时,声音一般可以分解为许多单纯的正弦波,即所有的自然声音基本都是由许多频率不同的正弦波组成的,其中频率最低的正弦波即为基音,而其他频率较高的正弦波则为泛音[13]。音高在声学上称为音调,即发声体整体振动产生的音,由频谱中的基音频率决定,它主要决定着声音调子的高低,即旋律。音高与基音频率值并不是成线性关系,而是成对数关系,当两个声音的基音频率值相差 1 倍时,两个音调相差 1 个倍频程,即

$$\text{音高} = k \lg f \quad (3)$$

图 8、图 9、图 10、图 11、图 12 分别是实验一、实验二、实验三、实验四、实验五的得出的基音(pitch)曲线图,横坐标表示哼唱时间(s),纵坐标表示基音频率值(Hz)。

通过对图 8 中 A、B 两条曲线的对比可以看出:不同的两个人,用相同的音调、演唱相同的歌词时,

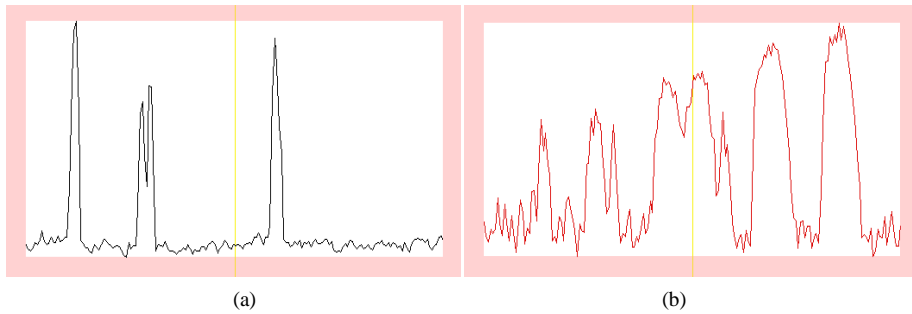


Figure 7. Formant curve
图 7. MFCC 曲线

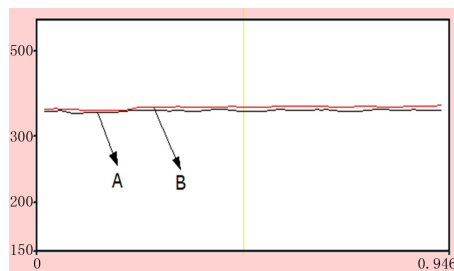


Figure 8. Pitch curve of experiment one
图 8. 试验一的 pitch 曲线

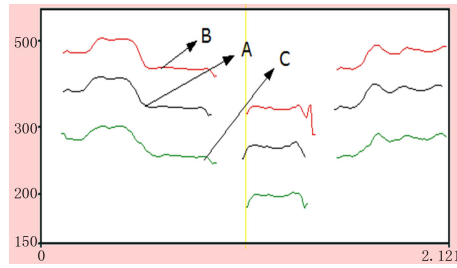


Figure 9. Pitch curve of experiment two
图 9. 实验二的 pitch 曲线

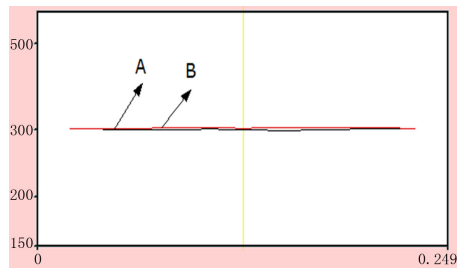


Figure 10. Pitch curve of experiment three
图 10. 实验三的 pitch 曲线

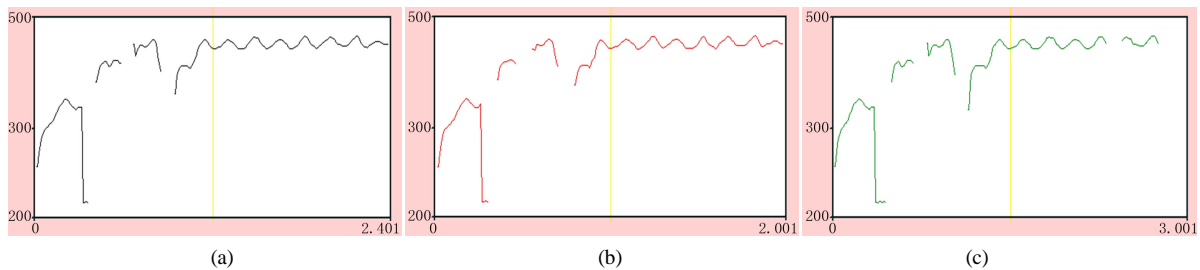


Figure 11. Pitch curve of experiment four
图 11. 实验四的 pitch 曲线

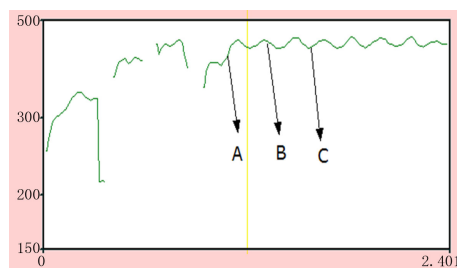


Figure 12. Pitch curve of experiment five
图 12. 实验五的 pitch 曲线

pitch 曲线是基本重合的。这说明，基音这一参数能在不同的人中，判别出音调是否相同。

通过对图 9 中 A、B、C 三条曲线的对比可以看出：同一个人，用不同的音调、演唱相同的歌词时，pitch 曲线是不同的，但 pitch 的变化轨迹是相似的。这说明，基音这一参数能反映出音调高低变化的不同，并能反映出音调的变化规律。

通过对图 10 中 A、B 两条曲线的对比可以看出：同一个人，用相同的音调、演唱不同的歌词时，pitch 曲线基本重合。这说明，基音这一参数与音调的高低有关，与哼唱内容无关。

通过对图 11 中(a)、(b)、(c)三个分图的比对可以看出：同一个人，用相同的音调、不同的节奏、演唱相同的歌词时，pitch 的变化轨迹是相似的，当把它们在时间轴上线性延展到相同的长度时，三条 pitch 曲线基本能够重合。这说明，在音调相同的条件下，演唱节奏的快慢变化对音调高低变化无影响。

通过对图 12 中 A、B、C 三条曲线的对比可以看出：同一个人，用相同的音调、不同的音量、演唱相同的歌词时，pitch 曲线是完全重合的。这说明，音量的高低变化对音调的高低变化无影响。

理论上，声音调子的高低主要由基音频率决定。上述五个实验结果与理论相符，基音频率只与音调的高低有关，与什么人在唱、唱的什么内容及演唱节奏的快慢无关，这一参数符合哼唱检索对特征参数的要求。

5. 结束语

本文对声音的四种性质及四种常用的特征参数进行了计算机模拟实验及分析。通过上述实验验证，能量主要反映的是声音音量的特性，共振峰及 MFCC 主要反映的是声音的音色及音质特性，这三个参数都不能直观有效的反映音高特性，实验结果与理论相符，这三个参数不符合本文的要求；基音反映的是声音的音调即音高特性，实验结果与理论相符，这一参数符合哼唱检索对声音特征参数的要求。该分析为哼唱检索技术中特征参数的选取，提供了依据。

参考文献 (References)

- [1] 韩纪庆, 冯涛, 郑贵滨, 等 (2007) 音频信息处理技术(第 1 版). 清华大学出版社, 北京, 18-27.
- [2] 金武 (2010) 音质和音色的区分. *南开语言学刊*, **1**, 70.
- [3] Praat. 百度百科. http://baike.baidu.com/link?url=-3pEyvXTyhfVCUKeTMMpDmcPf_Mq72Wiz0XRskD4WW1xugkHmtE71B5HV4vG9HOTa_X5IBg9JTsALntssU7LlK
- [4] Praat 语音学软件. 维基百科. <http://zh.wikipedia.org/wiki/Praat>
- [5] 赵芳丽 (2012) 基于 Praat 软件的俄语读音分析. *计算机工程与应用*, **11**, 133-136.
- [6] 蔡武 (2013) 浅谈 PRAAT 语音分析软件在菲律宾华校汉语语音教学中的应用. *现代语文: 下旬. 文学研究*, **1**, 148-149.
- [7] 语言学名词审定委员会 (2011) 语言学名词. 商务印书馆, 北京.
- [8] Praat 语音学软件. 维基百科. <http://zh.wikipedia.org/wiki/音强>
- [9] 陈彧 (2012) 歌手共振峰 Praat 软件测量分析初探. *天津音乐学院学报(天籁)*, **2**, 87-93.
- [10] Titze, I.R. (1994) Principles of voice production. Prentice Hall, Englewood Cliffs.
- [11] Praat 语音学软件. 维基百科. 自由的百科全书. <http://zh.wikipedia.org/wiki/共振峰>
- [12] 竺乐庆 (2010) 基于 Mel 倒谱系数和矢量量化的昆虫声音自动鉴别. *昆虫学报*, **8**, 901-907.
- [13] Praat 语音学软件. 维基百科. <http://zh.wikipedia.org/wiki/基頻>