文章编号:1671-4598(2013)07-1966-03　　中图分类号:TP391.42　　文献标识码:A

混合 MFCC 特征参数应用于语音情感识别

周　萍[1]，李晓盼[1]，李　杰[2]，景新幸[3]
(1. 桂林电子科技大学 电子工程与自动化学院，广西 桂林　541004；
2. 桂林电子科技大学 计算机科学与工程学院，广西 桂林　541004；
3. 桂林电子科技大学 信息与通信学院，广西 桂林　541004)

摘要：引入两种新的特征参数 Mid－MFCC 和 IMFCC，采用 MFCC、Mid－MFCC 和 IMFCC 相结合的改进算法，解决 MFCC 特征参数在语音识别中对中、高频信号的识别精度不高的特点，并使用增减分量法计算 MFCC、Mid－MFCC 和 IMFCC 各阶倒谱分量对语音情感识别的贡献，提取 3 个特征参数贡献最高的几阶倒谱分量组成了新的特征参数；实验结果表明，在相同环境下新的特征参数比经典 MFCC 特征参数的语音情感的识别率稍高。

关键词：Mel 频率倒谱系数（MFCC)；增减分量法；特征提取

Speech Emotion Recognition Based on Mixed MFCC Characteristic Parameter

Zhou Ping[1], Li Xiaopan[1], Li Jie[2], Jing Xinxing[3]
(1. School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin　541004, China;
2. School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin　541004, China;
3. School of Information and Communication, Guilin University of Electronic Technology, Guilin　541004, China)

Abstract: This paper introduced two new characteristic parameters Mid－MFCC and IMFCC combining with MFCC to improve the algorithm which solve the problem that MFCC characteristic parameter in speech recognition has low identification accuracy when signal is intermediate, high frequency signal, calculating the contribution that MFCC, Mid－MFCC and IMFCC each order cepstrum component was used in speech emotion recognition with increase or decrease component method, extracting highest contribution of several number order cepstrum component from three characteristic parameters and forming a new characteristic parameter. The experiment results show that new characteristic parameter has higher recognition rate than classic MFCC characteristic parameter in speech emotion recognition under the same environment.

Key words: MFCC; increase or decrease component method; feature extraction

0　引言

在语音情感识别技术中，MFCC 模拟了人耳的听觉特性，相对于其他特征具有强抗噪性、高识别率的特点，目前已经成为语音情感识别领域应用最为广泛的特征参数。MFCC 是用一个在低频区域交叉重叠的三角形滤波器组——Mel 滤波器组对语音信号的能量谱进行带通滤波。Mel 滤波器组在信号的低频区域分布较密，中频区域分布稍少，高频区域分布较为稀疏，单个滤波器的通带带宽较大。因此，Mel 滤波器组在低频区域有较高的频率分辨率，而在中、高频区域的频率分辨率较低，频谱信息较弱，导致信息遗漏。Mid－MFCC 和 IMFCC 可以有效弥补这一问题，它们分别在中、高频区域具有很好的计算精度。为了有效的将 MFCC、Mid－MFCC 和 IMFCC 融合，本文先用增减分量法[1]考察上述 3 种特征参数中各倒谱分量对情感识别的贡献，再将这 3 种特征参数中对识别率贡献最高的几阶系数组合到一起，构成新的特征参数。

收稿日期:2013-01-31；　**修回日期**:2013-03-25。
基金项目:国家自然科学基金资助项目(60961002)；广西自然科学基金资助项目(2012GXNSFAA053221)。
作者简介:周　萍(1961-)，女，教授，主要从事语音识别与智能控制方向的研究。

1　Mel 频率倒谱系数的提取

人的听觉系统对声音频率的感知是非线性的：对 1 000 Hz 以下频率声音的感知呈近似线性关系；而对于 1 000 Hz 以上频率声音的感知遵循在对数频率坐标上的近似线性关系。为此，建立了符合人类听觉特性的 Mel 频率 f_{Mel}，其与实际频率 f 之间的转换公式如下：

$$f_{Mel}=\frac{1000\ln(1+\frac{f}{700})}{\ln(1+\frac{1000}{700})}\approx 1127\ln(1+\frac{f}{700}) \quad (1)$$

式中：Mel 频率 f_{Mel} 的单位是 Mel，实际频率 f 的单位是 Hz。Hz－Mel 频率对应关系如图 1 所示。

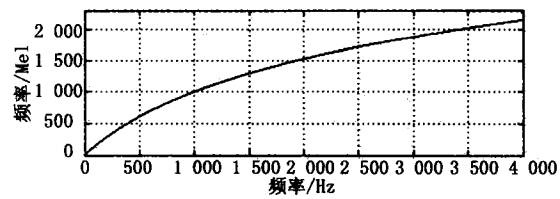

图 1　Hz－Mel 频率对应关系

类似临界频带的划分，将语音信号按频率划分为一个三角滤波器组——Mel 滤波器组如图 2 所示。

图 2 是 24 阶的 Mel 尺度滤波器组，各滤波器虽然在以 Hz

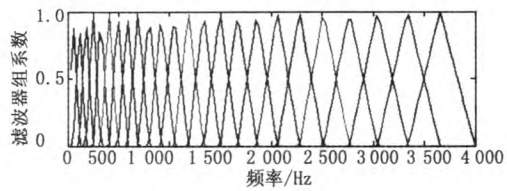

图 2　Mel 频率尺度滤波器组

频率坐标上是不等间距的，但在 Mel 频率坐标上是等间距的，各个滤波器交叉重叠。

Mel 滤波器 $H_i(k)$ 的计算如式（2）所示：

$$H_i(k)=\begin{cases}\dfrac{2[k-f(i-1)]}{[f(i+1)-f(i-1)][f(i)-f(i-1)]} \\ \qquad f(i-1)\leqslant k\leqslant f(i) \\ \dfrac{2[f(i+1)-k]}{[f(i+1)-f(i-1)][f(i+1)-f(i)]} \\ \qquad f(i)\leqslant k\leqslant f(i+1) \\ 0 \quad \text{其它}\end{cases} \tag{2}$$

式中：M 为滤波器组中滤波器的个数，通常取值在 24～40 之间，本文取 24。

MFCC 的提取过程如图 3 所示：

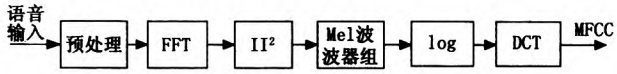

图 3　MFCC 的提取过程

MFCC 的提取过程如下：

（1）对原始语音信号进行预加重、分帧和加窗等预处理操作，得到短时信号 $x(n)$；

（2）对短时信号 $x(n)$ 进行傅立叶变换（DFT/FFT），得到线性频谱 $X_a(k)$；

（3）对 $X_a(k)$ 取模的平方，得到离散功率谱 $X(k)$；

（4）对 $X(k)$ 用 *Mel* 滤波器组公式（2）进行滤波，再对滤波器组的输出求对数能量 m_i；

（5）对 m_i 进行离散余弦变换（DCT）得到 MFCC，此变换式可简化为：

$$C_n=\sum_{i=1}^{M}m_i\cos[\pi n(i-0.5)/M],n=1,2,\cdots,L \tag{3}$$

式中：C_n 表示的是 MFCC 的系数，L 表示 MFCC 的阶数。实验表明，当阶数升高到一定程度，系统识别性能的改善将变得很小，系统的复杂度却大大增加。因此实际应用中，只需取 12～16 阶倒谱系数就可以达到很高的识别效率。

2　改进 Mel 频率倒谱系数的提取

逆 Mel 频率倒谱参数[2]（Inverted Mel－Frequency Cepstrum Coefficients，IMFCC）是由 Sandipan 在 2007 年提出来的，他通过改变 Hz－Mel 频率之间非线性对应关系，设计出一种和 Mel 滤波器完全相反的 I－Mel 滤波器，I－Mel 滤波器组的滤波器在低频区域分布稀疏，在高频区域较为密集。

Mid－Mel 频率倒谱参数[3]（Mid Mel－Frequency Cepstrum Coefficients，Mid－MFCC）是由韩一等人提出的，其参考 MFCC 和 IMFCC 的 Hz－Mel 频率对应关系，在 0～2 000 Hz 区域类似 IMFCC 的高频部分，在 2 000～4 000 Hz 区域类似 MFCC 的低频部分。Mid－Mel 滤波器组的滤波器在中频区域（1 500～2 500 Hz）分布较为密集，在高、低频区域较为稀疏。IMFCC 和 Mid－MFCC 的 Hz－Mel 频率对应关系为：

$$f_{I-Mel}=2146.1-1127\times\ln\left(1+\frac{4000-f}{700}\right) \tag{4}$$

$$f_{Mid-Mel}=\begin{cases}1037.05-527\times\ln\left(1+\dfrac{2000-f}{200}\right) \\ \qquad 0<f\leqslant 2000 \\ 1037.05+527\times\ln\left(1+\dfrac{f-2000}{200}\right) \\ \qquad 2000<f\leqslant 4000\end{cases} \tag{5}$$

式中：Mel 频率 f_{Mel} 的单位是 Mel，实际频率 f 的单位是 Hz。IMFCC 和 Mid－MFCC 的 Hz－Mel 频率对应关系和 I－Mel 频率尺度滤波器组如图 4、5 所示。

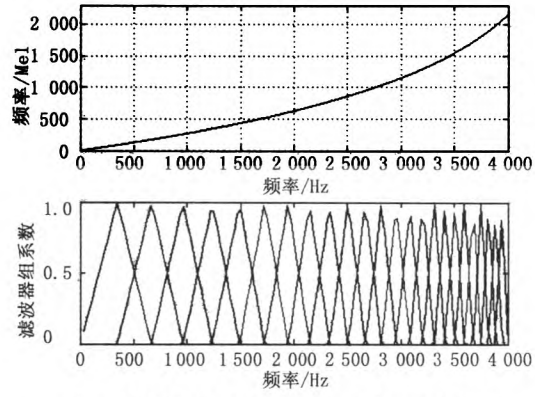

图 4　IMFCC 频率对应关系及 I－Mel 滤波器组

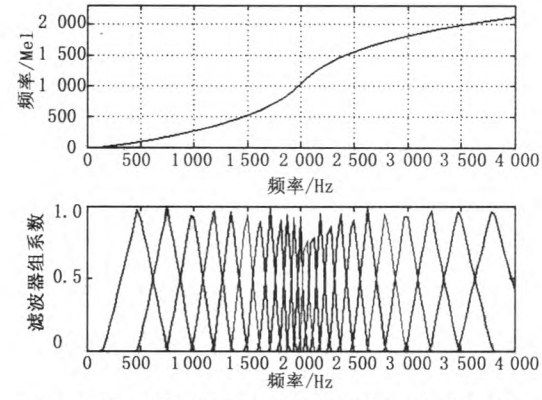

图 5　Mid－MFCC 频率对应关系及 Mid－Mel 滤波器组

IMFCC 和 Mid－MFCC 的提取过程和 MFCC 基本相同，只需改变滤波器组的响应函数。

3　增减分量法

为了有效的将 MFCC、Mid－MFCC 和 IMFCC 融合，找出识别效果更好的特征参数，本文采用增减分量法求出上述 3 种特征参数中各倒谱分量对情感识别的贡献。增减分量法是一种可以有效计算出每阶倒谱分量平均贡献的算法，具体公式为：

$$R(i)=\frac{1}{n}\left[\sum_{j>i}(p(i,j)-p(i+1,j))+\sum_{j<i}(p(j,i)-p(j,i-1))\right] \tag{6}$$

式中：$R(i)$ 表示第 i 阶倒谱分量的平均贡献值，n 为倒谱阶数，$p(i,j)$ 为从第 i 阶到第 j 阶倒谱系数特征的识别率。

4 实验及其分析

实验使用的语音库是北航情感语音数据库[4-5]，语音资料均为 wav 格式，语音的采样频率为 8 kHz，采样精度为 16bit。预加重系数为 0.95，帧长为 256，帧移为 128，加汉明窗。Mel 滤波器组取 24 个滤波器，对每一阶系数都用均值、方差、最大值、中位数和平均变化率 5 个统计特征来描述每组特征序列。选用语音库中 4 种表现力度比较强的情感——悲伤、生气、中性和高兴作为识别情感。通过剪辑和主观听觉判断，最后得到 480 句符合要求的语音情感样本，其中每种情感有 120 句语音样本，男女语音样本各占一半，对 20 条常用短语各重复 3 遍。

情感识别模型选取支持向量机（Support Vector Machine，SVM），并采用台湾大学林智仁教授开发的 LIBSVM 工具箱[6]来实现 SVM。实验环境为 MATLAB R2009a，LIBSVM 的安装环境为 Visual C++ 6.0。

对建立的 480 句情感语音资料库进行与文本无关情感识别实验，用 5 个统计特征表征各倒谱分量，采用 5 次交叉验证的均值作为最终结果，得表 1 是 MFCC 各分量顺序组合的语音情感识别率。随后按照增减分量法的公式（6）对表 1 进行计算，得的语音情感识别中 MFCC 各阶倒谱分量的平均贡献，如图 6 所示。

表 1 MFCC 相邻分类顺序组合的情感识别率（%）

	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}	C_{19}	C_{20}	C_{21}	C_{22}	C_{23}
C_0	55.0	52.5	68.3	74.2	74.2	81.7	80.8	83.3	82.5	85.0	84.2	86.7	90.8	88.3	85.8	87.5	88.3	89.2	87.5	88.3	89.2	88.3	85.8	85.0
C_1		49.2	55.0	65.0	72.5	77.5	78.3	79.2	80.0	85.8	85.8	85.8	87.5	86.7	89.2	89.2	90.0	90.0	90.8	89.2	91.7	90.0	85.8	85.8
C_2			58.3	59.2	62.5	60.0	59.2	75.0	74.2	75.0	62.5	70.8	85.8	84.2	85.8	86.7	89.2	88.3	87.5	75.0	74.2	74.2	71.7	72.5
C_3				60.0	60.8	72.5	69.2	74.2	73.3	76.7	64.2	72.5	83.3	85.0	86.7	87.5	85.8	86.7	88.3	86.7	78.3	75.8	72.5	70.8
C_4					40.8	50.8	43.3	42.5	47.5	56.7	56.7	65.0	80.8	66.7	79.2	69.2	70.8	69.2	66.7	70.8	71.7	70.0	70.0	70.8
C_5						55.0	50.8	45.0	51.7	60.8	61.7	70.8	77.5	77.5	79.2	70.8	69.2	69.2	68.3	69.2	69.2	67.5	67.5	70.0
C_6							44.2	40.8	43.3	55.8	59.2	70.0	74.2	74.2	76.7	79.2	79.2	66.7	68.3	68.3	70.8	67.5	68.3	71.7
C_7								42.5	45.0	61.7	70.0	69.2	74.2	75.0	78.3	79.2	61.7	59.2	70.8	75.0	70.0	69.2	72.5	70.8
C_8									52.5	65.0	60.8	68.3	70.8	75.0	76.7	81.7	80.0	65.8	67.5	62.5	65.0	69.2	72.5	73.3
C_9										58.3	55.0	66.7	71.7	62.5	74.2	66.7	66.7	65.8	69.2	75.0	69.2	73.3	74.2	77.5
C_{10}											50.8	63.3	66.7	59.2	62.5	61.7	65.8	63.3	70.0	71.7	68.3	77.5	77.5	75.0
C_{11}												54.2	68.3	65.8	69.2	60.0	56.7	65.8	69.2	71.7	66.7	69.2	73.3	73.3
C_{12}													58.3	67.5	69.2	71.7	58.3	57.5	82.5	71.7	79.2	73.3	75.0	76.7
C_{13}														69.2	68.3	67.5	69.2	70.8	65.8	75.8	70.8	71.7	75.8	73.3
C_{14}															60.8	69.2	65.8	77.5	71.7	73.3	68.3	71.7	70.8	71.7
C_{15}																65.0	61.7	66.7	70.0	60.8	66.7	72.5	66.7	70.8
C_{16}																	55.0	75.0	63.3	67.5	71.7	72.5	70.8	69.2
C_{17}																		59.2	63.3	72.5	64.2	72.5	66.7	67.5
C_{18}																			71.7	70.8	68.3	72.5	65.0	65.8
C_{19}																				67.5	62.5	61.7	61.7	63.3
C_{20}																					58.3	70.8	66.7	71.7
C_{21}																						65.0	70.8	70.0
C_{22}																							53.3	53.3
C_{23}																								50.8

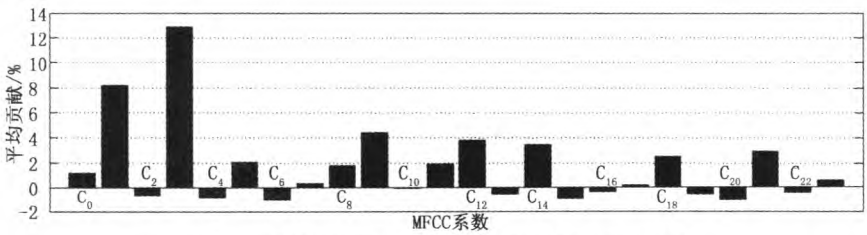

图 6 MFCC 各阶分量在语音情感识别中的平均贡献

同理可根据增减分量法对 Mid－MFCC 和 IMFCC 相邻分类顺序组合的情感识别率进行计算得出 Mid－MFCC 和 IMFCC 各阶倒谱分量的平均贡献。最后选取平均贡献最大的 8 阶 MFCC 倒谱分量、4 阶 Mid－MFCC 倒谱分量和 4 阶 IMFCC 倒谱分量组成 16 阶混合参数。对混合参数进行实验，结果如图 7 所示。

从图 7 可以看出，虽然 MFCC 特征的情感识别性能明显比 Mid－MFCC、IMFCC 特征好，但由 MFCC、Mid－MFCC 和 IMFCC 组成的 16 阶混合参数的整体情感识别率比取 16 阶 MFCC 特征的识别率高。因此混合参数具有更高的情感识别率。

5 结论

针对作为语音情感识别领域的 MFCC 特征参数能较好地模拟人耳听觉系统的感知能力，对低频语音信号有着很好的计算精度，但相对于中高频信号的计算精度不高的特点。Mid－MFCC 和 IMFCC 在设计上很好的弥补了 MFCC 的这个缺陷运用增减分量法计算 MFCC、Mid－MFCC 和 IMFCC 各阶倒谱分量对语音情感识别

（下转第 1986 页）

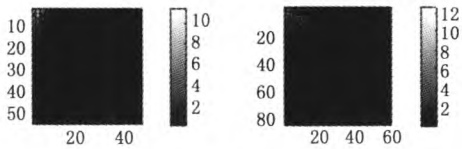

图 5 DCT 分量 DP 分布图

ORL 数据库中 DP 法和 GA 法都选取了更多的低频分量，说明 ORL 数据库中低频分量含有丰富的鉴别信息，不宜直接去除。如果不进行掩膜，DP 法虽然可以保留这些低频分量，但不能保证选择分量组合后的整体鉴别性，而 GA 法通过全局搜索获得了优化的分量组合。

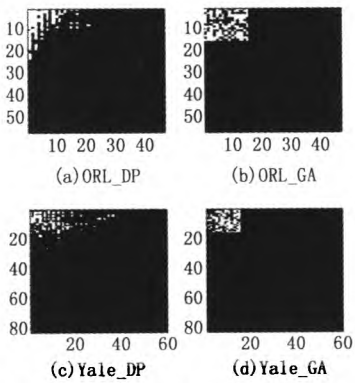

图 6 DCT 分量选择结果

表 3 是各种 DCT 域人脸认证算法的 CER 对比，对比算法在第 2、3 部分均已介绍，pm1、pm2、d1、d3 4 个掩膜窗口分别去除了 1、4、1、3 个低频分量。

ORL 数据库中，由于去除了更多有效的低频分量，四种掩膜的 CER 高于不掩膜的 CER。ring band 掩膜没有去除过多低频分量，因此与不掩膜的 CER 近似。

Yale 数据库中，低频分量差异主要源于光照差异干扰。由于去除了更多无效的低频分量，pm1 和 pm2 掩膜的 CER 低于不掩膜和 ring band 掩膜的 CER。但 d1 和 d3 掩膜的 CER 偏高，说明 d1 和 d3 掩膜未能有效抑制高频分量。

两个数据库中不同掩膜窗口的性能存在差异，说明掩膜窗口不能针对不同数据库特点自适应调节。GA 法可根据不同数据库特点优选分量，具有自适应性，并且保证优选分量组合后的整体鉴别性，因此两个数据库中 GA 法的 CER 最低，

表 3 DCT 域人脸认证算法 CER 对比

掩膜	分量选择	CER(%)	
		ORL	Yale
—	DP	11.6688	10.7957
pm1	DP	13.4274	10.3266
pm2	DP	19.4425	10.2991
ring band	DP	11.6808	10.7226
d1(去直流分量)	DP	13.4229	10.8540
d3(去三个低频分量)	DP	13.9348	10.9603
—	GA	9.7026	9.1806

4 结论

采用遗传算法，由二进制编码个体控制人脸图像 DCT 域分量选取，通过全局智能搜索可优选出整体高鉴别性的分量组合。与现有 DCT 域人脸认证算法相比，遗传算法优选的 DCT 分量具有更优的认证性能，对不同数据库也具有自适应性。

参考文献：

[1] 李 扬，孙劲光，孟祥福，等. AMSR 与 SVM 相结合的人脸识别方法 [J]. 计算机测量与控制，2012，20 (3)：823 - 825.

[2] Rao A，Noushath S. Subspace methods for face recognition [J]. Computer Science Review，2010，4 (1)：1 - 17.

[3] Delac K，Grgic M，Grgic S. Face recognition in JPEG and JPEG2000 compressed domain [J]. Image and Vision Computing，2009，27 (8)：1108 - 1120.

[4] Jing X Y，Zhang D. A face and palmprint recognition approach based on discriminant DCT feature extraction [J]. IEEE Transactions on System，Man and Cybernetics，Part B：Cybernetics，2004，34 (6)：2405 - 2415.

[5] Er M J，Chen W，Wu S. High speed face recognition based on discrete cosine transform and RBF neural networks [J]. IEEE Transactions on Neural Networks，2005，16 (3)：679 - 691.

[6] Dabbaghchian S，Ghaemmaghami M P，Aghagolzadeh A. Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology [J]. Pattern Recognition，2010，43 (4)：1431 - 1440.

(上接第 1968 页)

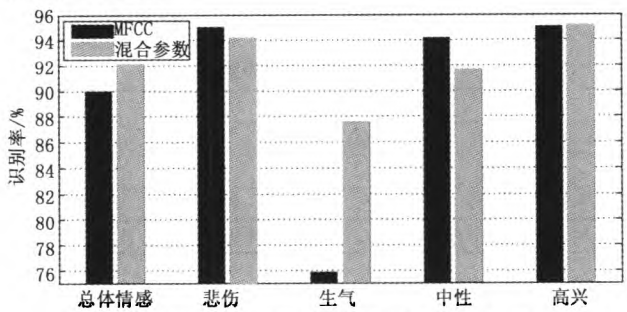

图 7 MFCC 和混合参数的情感识别结果柱状对比图

的贡献，提取这 3 个特征参数贡献最高的几阶倒谱分量组成了混合 16 阶 MFCC，与传统 16 阶 MFCC 方法相比，新算法具有更高的情感识别率。

参考文献：

[1] 甄 斌，吴玺宏，刘志敏，等. 语音识别和说话人识别中各倒谱分量的相对重要性 [J]. 北京大学学报（自然科学版），2001，37 (3)：371 - 378.

[2] 吕霄云，王宏霞. 基于 MFCC 和短时能量混合的异常声音识别算法 [J]. 计算机应用，2010，30 (3)：796 - 798.

[3] 韩 一，王国胤，杨 勇. 基于 MFCC 的语音情感识别 [J]. 重庆邮电大学学报（自然科学版），2008，20 (5)：507 - 602.

[4] 毛 峡，陈立江. 语音情感信息的提取及建模方法 [P]. 中国专利：CN101261832，2008 - 4 - 21.

[5] Mao X，Chen L J. Speech Emotion Recognition Based on Parametric Filter and Fractal Dimension [J]. IEICE Transactions on Information and Systems (SCIE Index，IF：0. 396). 2010，93 (8)：2324 - 2326.

[6] Chang C C，Lin C J. LIBSVM：a Library for Support Vector Machine [EB/OL]. http：//uww. csie. ntu. edu. tw/～ cjlin/libsvmlindex. html.