

• 信息工程 •

DOI:10.15961/j.jsuese.201900789



本刊网刊

基于倒谱分析的实时广播音频相似度快速比对算法

邵玉斌, 唐传林, 赵至柔, 龙 华, 杜庆治

(昆明理工大学 信息工程与自动化学院, 云南 昆明 650500)

摘 要:为了解决广播音频中经常存在噪声干扰和时间延迟导致音频比对结果不准确的问题, 提出具有延时自适应意识的音频比对算法。针对常用算法中测量音频特征距离抗噪性能差的不足, 采用倒谱对两音频的混合信号分析, 并利用倒谱对功率谱中的等距离频率成分有很强的分辨能力这一特性来进行自适应延时估计和比对; 为比对不同情况的两音频都可得到准确的相似度, 提出对其中一音频加入短延时, 再将两音频叠加混合后做倒谱分析; 并根据加入不同短延时的效果选择出最优短延时, 进一步提升算法性能。使用真实广播不同节目中截取出来的多个音频, 在无噪声和不同信噪比加性高斯白噪声条件下, 通过仿真实验评估了所提出算法的性能, 比较了不同信噪比下的延时估计结果和音频相似度。实验结果证明, 所提出方法的延时估计结果和比对结果优于现有算法, 在低信噪比(SNR=2 dB)下, 也可以达到90.36%的音频比对匹配精度, 且计算速度能够达到实时比对的要求。

关键词:音频比对; 延时估计; 倒谱分析; 实时广播音频

中图分类号:TN931.3

文献标志码:A

文章编号:2096-3246(2020)03-0178-08

Fast Similarity Comparison Algorithm for Real-time Broadcast Audio Based on Cepstrum Analysis

SHAO Yubin, TANG Chuanlin, ZHAO Zhirou, LONG Hua, DU Qingzhi

(School of Info. Eng. and Automation, Kunming Univ. of Sci. and Technol., Kunming 650500, China)

Abstract: In order to solve the problem of inaccurate audio comparison caused by noise and delay in broadcast audio, an audio comparison algorithm with delay-adaptive-aware was proposed. To tackle the poor noise immunity of audio feature distance measurement in conventional algorithms, cepstrum was used to analyze the mixed signal of two audio to estimate delay adaptively and comparison, which has a strong ability to resolve equidistant frequency components in the power spectrum. Then, a method of short delay was proposed to obtain accurate similarity between two audios in different situations, which is added short delay in one of the audio. Afterwards, the optimal short delay was selected according to the effect of adding different short delay, so as to improve the performance of the algorithm further. Finally, the simulation experiments were conducted to evaluate the performance of the proposed algorithm, in which multiple audio clips of different broadcast programs were utilized under the condition of different SNR and additive white Gaussian noise. And the delay estimation results and audio similarity under different SNR are compared to verify the effectiveness of the algorithm. Experimental results show that the proposed algorithm outperforms existing algorithms and can achieve 90.36% audio comparison matching accuracy at a low SNR (2 dB), and the calculation speed can meet the requirement of real-time comparison.

Key words: audio comparison; delay estimation; cepstrum analysis; real-time broadcast audio

为了保证在发射过程中音频信号不被篡改或污染, 需要对发射前的音频信号和发射信号的接收解调结果进行比对^[1]。目前广播节目监播这一工作大多

是人工完成的^[2], 即广播监播机接收广播信号并解调, 监测人员循环播放和监听辨别。广播发射台常有多路广播节目信号需要发射, 所以人工音频监播的

收稿日期:2019-08-08

基金项目:国家自然科学基金地区科学基金项目(61761025)

作者简介:邵玉斌(1970—), 男, 教授, 硕士。研究方向: 音频信号处理; 自然语言处理。E-mail: shaoyubin@kmust.edu.cn

网络出版时间:2020-05-11 10:29:39

网络出版地址: <http://kns.cnki.net/kcms/detail/51.1773.TB.20200509.1627.002.html>

<http://jsuese.ijournals.cn>

<http://jsuese.scu.edu.cn>

工作量很大,无法达到实时监控的要求。并且,同一套广播节目的音频信号一般都是由不同传输链路(卫星、光纤、微波、网络)传输的,各路信号到达各节点会存在时间延迟^[3]。当两音频不同步时,音频比对不能得到准确的相似度^[4]。因此,在音频比对之前需要先估计出两音频信号间的延时,并将两音频同步,同步结果对音频比对结果(即相似度)的准确度影响很大^[5]。

通常,传统的音频相似度计算方法主要有两步^[6]:先提取音频的特征参数^[7],如音频信号的包络、短时能量、基音频率、梅尔频率倒谱系数(MFCC)^[8]、线性预测倒谱系数(LPCC)^[9]等特征;再使用距离测量等方法比较这些特征参数的差异,从而得出两路音频的相似度^[10]。传统方法只能针对普通音频,不适用于广播音频,其原因主要有3个方面:一是,如果两路实时广播音频之间存在延时,用音频特征参数计算所得到的音频相似度结果就不准确。二是,在容错性方面,文献^[11]中指出众多音频特征参数都是以准确描述音频为目标,音频的细微变化都会反映在音频特征参数上,从而达到比较高的音频内容辨识度。如果所输入的两个相同音频含有噪声干扰,或是经过一些不同通带的滤波器(如带通滤波器)处理,会使两音频之间有较大的差异^[12],故用音频特征参数计算相似度时通常会误判为两音频相似度很低甚至完全不同。而广播音频在制作时需要进行滤波处理,在传输过程中会存在噪声干扰。在广播音频监测中,需要比对卫星接收机的音频和广播监播机接收到的音频,前者一般是没有任何噪声的,后者常常含有噪声干扰。针对广播音频比对,主要在于判断两音频的内容是否相同,而不在于听觉效果上是否完全一致,所以对于受到噪声干扰和经过滤波器处理的广播音频,传统方法会误判两音频相似度较低。三是,在相似度的计算效率方面,提取音频特征参数以计算两实时音频的相似度的传统方法计算复杂度相对较高^[13],高性能的计算机和较长的计算时间,因此传统方法不适合用于广播音频实时比对。目前,也有不用提取音频特征参数进行音频比对的方法,如文献^[14]中将需要比对的两音频归一化后分为低频带、中频带和高频带,相同频带的音频短时能量求和后做减法,选取合适阈值判断减法结果从而得到音频相似度。由于该算法在时域中处理,因此计算简单并且只需要较短的计算时间,但是该方法同样只能用于比对没有时间延迟的两音频。

针对以上问题,本文提出一种基于倒谱分析的实时音频相似度快速比对算法。若两音频存在延时,该算法不需要先将两音频同步,可直接对两音频进行延时估计和音频比对,算法抗噪声性能较强,计算

时间短,可得到较准确的延时估计结果和音频比对结果。

1 传统计算相似度的方法

要确定两样本之间的相似性,可采用不同的方法。一般而言,传统的音频相似度测量可以分为两类。一是,基于距离,如Minkowski距离、Mahalanobis距离和Canberra距离。这些方法将提取到的音频特征向量,如音频的MFCC特征向量,视为度量空间中的点,点与点之间的距离越大,相似性越小,反之亦然。二是,基于相关系数,例如,余弦相似度和Pearson相关系数。所得相似度范围为-1到1,1表示完全相同,-1表示完全不同,0通常表示独立性,中间值表示一定程度的相似或不相似。表1中描述了典型的相似度测量方法,其中, X 和 Y 是 N 维的矢量, C 是 X 和 Y 的协方差矩阵, \bar{x} 为 X 的均值, \bar{y} 为 Y 的均值。

表1 典型的相似度测量方法

方法名称	表达式
Minkowski距离	$d(X, Y) = \left(\sum_{i=1}^N x_i - y_i ^\lambda \right)^{\frac{1}{\lambda}}$
Mahalanobis距离	$d(X, Y) = \sqrt{(X - Y)^T C^{-1} (X - Y)}$
余弦相似度	$d(X, Y) = \frac{\sum_{i=1}^N x_i y_i}{\sqrt{\sum_{i=1}^N x_i^2 \cdot \sum_{i=1}^N y_i^2}}$
Pearson相关系数	$d(X, Y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}}$

Minkowski距离通常取 λ 为1或2。当 $\lambda = 1$ 时称为曼哈顿距离,曼哈顿距离克服了Minkowski距离的缺陷,因为它基于协方差矩阵进行归一化以使距离度量尺度不变,但是在计算 C^{-1} 时由于耗时较长而不适用于处理大规模数据。 $\lambda = 2$ 时称为欧氏距离,欧氏距离是两个样本之间的普通距离,广泛用于音频和图像,但它与维度有关,不考虑信号之间的相关性。在 λ 的极限趋于无穷远时称其为切比雪夫距离。余弦相似度和Pearson相关系数消除了维数的影响,充分考虑了信号之间的相关性,更重要的是,其结果可以直观地反映出相似程度,但计算也相对复杂。

2 基于倒谱分析计算音频相似度

2.1 功率倒谱

倒谱(cepstrum)是对信号频谱的对数进行傅里

叶反变换 (IFT) 的结果。倒谱有复数倒谱、实数倒谱、功率倒谱和相位倒谱。功率倒谱在人类语言分析中有着特殊的应用。本文倒谱分析用的是功率倒谱。倒谱可以看作是不同谱带的变化率的信息,它最初是用于描述地震和炸弹爆炸引起的地震回声。该方法对功率谱中的等距频率分量具有较强的分辨率,能够有效地识别信号中的回波分量。若给定时间信号 $y(t)$, 则其傅里叶变换为:

$$Y(\omega) = \mathcal{F}[y(t)] = \int_{-\infty}^{+\infty} y(t)e^{-j\omega t} dt \quad (1)$$

其功率谱为:

$$S_y(\omega) = |Y(\omega)|^2 = |\mathcal{F}[y(t)]|^2 \quad (2)$$

其功率倒谱为:

$$C_y(\tau) = \mathcal{F}^{-1}[\ln S_y(\omega)] = \int_{-\infty}^{+\infty} \frac{1}{2\pi} [\ln S_y(\omega)] e^{j\omega\tau} d\omega \quad (3)$$

信号自相关函数与其功率谱函数也是一对傅里叶变换:

$$R_y(\tau) = \mathcal{F}^{-1}(S_y(\omega)) = \int_{-\infty}^{+\infty} \frac{1}{2\pi} S_y(\omega) e^{j\omega\tau} d\omega \quad (4)$$

两者对比可见,功率倒谱与自相关函数具有相同的量纲,即具有一致的物理意义。回声是源信号与其延迟信号相叠加得到的,因为源信号与延迟信号具有强相关性,所以从含有延迟信号的混合信号的自相关函数上看,在时间延迟量位置上会具有较高峰值。信号间时域卷积对应其频域的乘积,功率倒谱把时域卷积的复杂关系转化成为时延域的简单相加关系,因此,在信号功率倒谱的同样位置也将出现高峰值,且更为明显。

2.2 倒谱分析法计算音频相似度的原理

假设 $x(t)$ 为信号源发出的源信号,经过两条传输链路后信号分别为 $x_1(t)$ 和 $x_2(t)$:

$$\begin{cases} x_1(t) = a_1 x(t), \\ x_2(t) = a_2 x(t - \tau_0) \end{cases} \quad (5)$$

式中: τ_0 为信号 $x_1(t)$ 和 $x_2(t)$ 之间的时间延迟,且 $\tau_0 > 0$; a_1 和 a_2 为信号的衰减因子,且 $0 < a_1 < 1, 0 < a_2 < 1$ 。

构造混合信号为:

$$y(t) = x_1(t) + x_2(t) = a_1 x(t) + a_2 x(t - \tau_0) = x(t) * (a_1 \delta(t) + a_2 \delta(t - \tau_0)) \quad (6)$$

式中, $\delta(t)$ 单位冲激函数,对式 (6) 取两边傅里叶变换得:

$$Y(\omega) = X(\omega)(a_1 + a_2 e^{-j\omega\tau_0}) \quad (7)$$

对于式 (7), 其功率谱密度函数为:

$$\begin{aligned} S_y(\omega) &= |Y(\omega)|^2 = Y(\omega) Y^*(\omega) = \\ &= X(\omega)(a_1 + a_2 e^{-j\omega\tau_0}) \cdot X^*(\omega)(a_1 + a_2 e^{-j\omega\tau_0})^* = \\ &= S_x(\omega)(a_1 + a_2 e^{-j\omega\tau_0}) \cdot (a_1 + a_2 e^{j\omega\tau_0}) \end{aligned} \quad (8)$$

式中,上标“*”表示复共轭运算。式 (8) 由 3 个乘积项构成。为分离这三项,对式 (8) 两边取对数,得:

$$\ln[S_y(\omega)] = \ln[S_x(\omega)] + \ln(a_1 + a_2 e^{-j\omega\tau_0}) + \ln(a_1 + a_2 e^{j\omega\tau_0}) \quad (9)$$

利用对数函数的级数展开式:

$$\begin{aligned} \ln(a+x) &= \ln a + \frac{x}{a} - \frac{x^2}{2a^2} + \frac{x^3}{3a^3} - \dots = \\ \ln a + \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{na^n} \end{aligned} \quad (10)$$

取式 (10) 前两项做近似,式 (9) 可写为:

$$\begin{aligned} \ln S_y(\omega) &= \ln S_x(\omega) + \ln a_1 + \frac{a_2}{a_1} e^{-j\omega\tau_0} - \frac{a_2^2 e^{-j2\omega\tau_0}}{2a_1^2} + \\ &= \ln a_1 + \frac{a_2}{a_1} e^{j\omega\tau_0} - \frac{a_2^2 e^{j2\omega\tau_0}}{2a_1^2} \end{aligned} \quad (11)$$

对式 (11) 两边取傅里叶变换,将时域变量记为 τ , 称为时延域,得:

$$\begin{aligned} \mathcal{F}^{-1}\{\ln S_y(\omega)\} &= \mathcal{F}^{-1}\{\ln S_x(\omega)\} + \frac{a_2}{a_1} \delta(\tau - \tau_0) - \\ &= \frac{a_2^2}{2a_1^2} \delta(\tau - 2\tau_0) + \frac{a_2}{a_1} \delta(\tau + \tau_0) - \frac{a_2^2}{2a_1^2} \delta(\tau + 2\tau_0) + 2 \ln a_1 \end{aligned} \quad (12)$$

依据功率倒谱定义,将式 (12) 简化为:

$$\begin{aligned} C_y(\tau) &= C_x(\tau) + \frac{a_2}{a_1} \delta(\tau - \tau_0) - \frac{a_2^2}{2a_1^2} \delta(\tau - 2\tau_0) + \\ &= \frac{a_2}{a_1} \delta(\tau + \tau_0) - \frac{a_2^2}{2a_1^2} \delta(\tau + 2\tau_0) + 2 \ln a_1 \end{aligned} \quad (13)$$

由式 (13) 可以看出,混合信号的功率倒谱中,在时间延迟位置及其整数倍位置上有冲激峰量。由此,只要检测出冲激峰量,即可根据其位置得出两路音频信号之间的延迟 τ_0 , 根据其峰值的幅度大小得出两路音频信号之间的相似度。

2.3 倒谱分析计算音频相似度的步骤

图 1 为倒谱分析计算音频相似度的流程图。输入为两路音频信号,分别称为卫星音频 $x_1(t)$ 和广播机音频 $x_2(t)$,前者来自卫星接收机,后者来自广播监播机。连续信号 $x_1(t)$ 、 $x_2(t)$ 将被采样为离散信号 $x_1(i)$ 、 $x_2(i)$ 。使用的采样率为 $f_s = 8\,000$ Hz。

使用式 (14) 归一化采样数据 $x_1(i)$ 、 $x_2(i)$, 即:

$$\bar{x}(i) = \frac{x(i)}{\sqrt{\sum_{i=0}^N [x(i)]^2}} \quad (14)$$

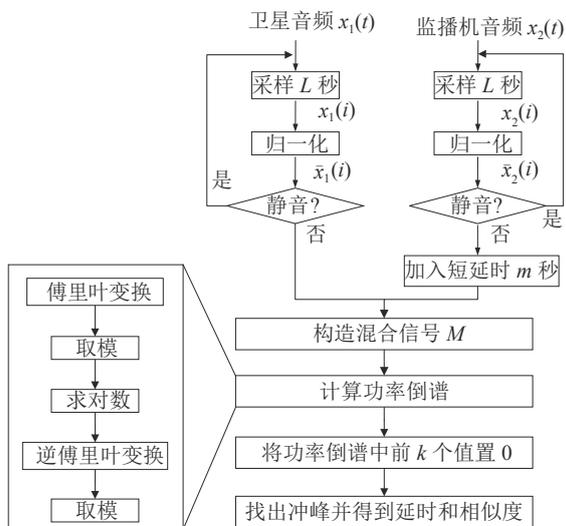


图1 基于功率倒谱的音频相似度计算方法

Fig. 1 Audio similarity calculation method based on power cepstrum

式中, N 为窗口的大小。根据幅度归一化后所有采样点的电平值分别判断两音频是否为静音状态,即对所有采样点电平值求绝对值并求和:

$$X_s = \sum_{i=0}^N |\bar{x}(i)| \quad (15)$$

若 X_s 小于阈值则判断为静音状态,若 X_s 大于阈值则判断为非静音状态,阈值可根据实际具体情况变化。若两音频都不为静音状态,则将 $\bar{x}_2(i)$ 延时 m 秒得到信号 $\bar{x}_2(i_2 - m \cdot f_s)$, 其中 $m \cdot f_s < i_2 < m \cdot f_s + N$ 。

再按照式(16)将两信号叠加合并得到混合信号 M :

$$M = \frac{\bar{x}_1(i_1) + \bar{x}_2(i_2 - m \cdot f_s)}{2} \quad (16)$$

式中: $i_1 = 0, 1, 2, \dots, N$; $i_2 = m \cdot f_s, m \cdot f_s + 1, \dots, m \cdot f_s + N$ 。

对信号 M 计算功率倒谱,为了去除信号自相关峰,将功率倒谱中前 k 个值置零。从功率倒谱前一半中找到冲激峰的位置,记为 p_s ,冲激峰的幅度记为 p_a 。由冲激峰位置计算两音频间的真实延时 τ_0 :

$$\tau_0 = \frac{p_s - m \cdot f_s}{f_s} \quad (17)$$

根据课题组的其他实验研究可知,冲激峰量的幅度范围为 $0 \sim 0.45$ 。当完全一样的两音频作为输入时,冲激峰量的幅度为 0.45 ;当完全不一样的两音频作为输入时,冲激峰量的幅度在 0.02 以下。所以,将 0.02 作为阈值,冲激峰量幅度大于 0.02 则认为两音频可能相似,小于 0.02 则认为不相似。

借鉴PCM编码中的 μ 压缩律公式:

$$F(x) = \text{sgn}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)} \quad (18)$$

式中, $\text{sgn}(x)$ 为符号函数, $-1 \leq x \leq 1$ 。

提出了由功率倒谱中冲激峰幅度计算相似度百分比的公式,将冲激峰幅度映射为相似度百分比:

$$S = \frac{\ln(0.15 + 50000 \cdot p_a)}{\ln(0.15 + 50000)} \times 100\% \quad (19)$$

3 实验结果分析与比较

为了深入了解所提出的倒谱分析法在无噪声和不同信噪比加性白噪声条件下的性能,进行了以下7个实验。实验的环境平台均为MATLAB R2019a,电脑CPU为i5-3337u,系统为Windows10教育版。

3.1 短延时 m 的最优选择

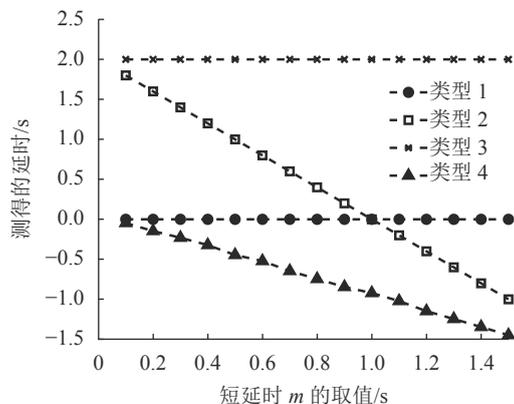
为探究算法中所加入的短延时 m 对两音频的延时估计和相似度计算的影响,将 m 从 0.1 到 1.5 s 以 0.1 s 为步进依次变化,在不同短延时 m 下计算两音频之间的延时和倒谱冲激峰幅度。

由于传输链路的影响,监播机音频一般都是延后于卫星音频,而备用卫星音频一般延后于监播机音频 2 到 3 s。因此实验中,准备了3个广播电台的音频片段,每个音频片段 30 s。其中,音频片段 $S1$ 和 $S2$ 内容完全不一样, $S2$ 与 $S3$ 内容是一样的,并且 $S3$ 比 $S2$ 快 2 s。按照表2将这3个音频片段进行两两组合,模拟了真实环境下卫星音频和监播机音频的4种情况。

表2 卫星音频和监播机音频的4种情况

Tab. 2 Four cases of satellite audio and monitor audio			
类型编号	卫星音频 $x_1(t)$	监播机音频 $x_2(t)$	情况类型
1	S2	S2	两音频完全一样
2	S2	S3	$x_2(t)$ 比 $x_1(t)$ 快 2 s
3	S3	S2	$x_1(t)$ 比 $x_2(t)$ 快 2 s
4	S2	S1	两音频完全不一样

4种情况所测得的两音频间的延时和倒谱冲激峰幅度的变化如图2和3所示。

图2 短延时 m 对所测延时的影响Fig. 2 Effect of short delay m on measured delay

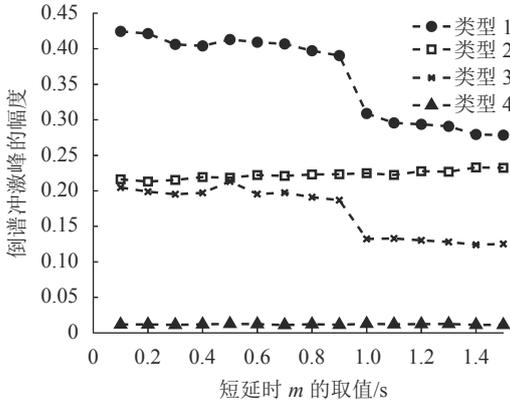


图 3 短延时 m 对倒谱冲激峰幅度的影响

Fig. 3 Effect of short delay m on impulse amplitude of cepstrum

从图2可以看出,在类型1和类型3下,所测得的两音频之间的延时不会因加入的短延时 m 的变化而改变,而在类型2下,所测得的两音频之间的延时会随加入的短延时 m 的增大而变得不再准确。从图3中可以看出倒谱冲激峰的幅度随短延时 m 的增大而减小。因此取 m 为0.1 s是最优的,这是因为 $m=0.1$ 时在4种情况下 m 对所测得的延时和倒谱冲激峰的幅度影响是最小的。

3.2 不同信噪比下无延时两音频比对结果比较

为比较倒谱分析法、基于MFCC的欧式距离法(简称欧式距离法)这两种音频比对算法在不同信噪比下无延时两音频比对结果,准备了广播电台的50个音频片段作为模版库,每个片段持续10 s,包括音乐,3种不同语种的语音以及音乐和语音的混合音频。对于这些音频片段,加入加性高斯白噪声后获得不同信噪比的音频片段。从中选择了第23个音频作为测试样本。

3.2.1 比对未经过带通滤波器处理的两音频

通过倒谱分析法和欧式距离法分别计算了测试样本与模版库中50个无噪声音频片段之间的相似度。归一化的欧式距离和相似度结果如图4和5所示。欧式距离越小,代表相似度越大。

从图4可以看出,基于MFCC特征的欧式距离法所测出的欧式距离在高信噪比($\text{SNR}>5$ dB)中是有效的,而在低信噪比($\text{SNR}=-5$ dB)中该方法不能准确地找出相似的音频片段。由图5可知,所提出的倒谱分析法很稳健,即使在低信噪比下也可以找到最佳匹配的音频片段。

3.2.2 比对经过不同通带的带通滤波器处理后的两音频

为了模拟两条不同的真实传输链路,设置两个数字带通滤波器a、b,阶数均为8,a的通带为50~2 000 Hz,b的通带为1 500~3 800 Hz,重叠带宽为500 Hz(1 500~2 000 Hz)。将测试样本(第23个)音频经过

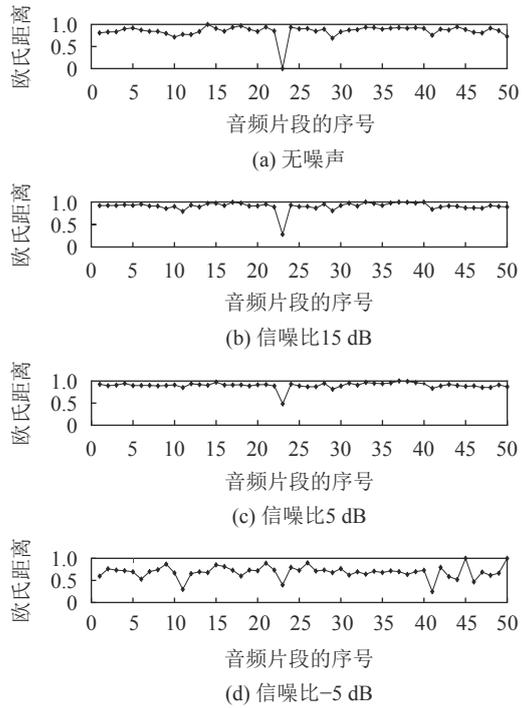


图 4 基于MFCC特征的欧式距离法所测出的欧式距离
Fig. 4 Measured Euclidean distance by Euclidean distance method based on MFCC features

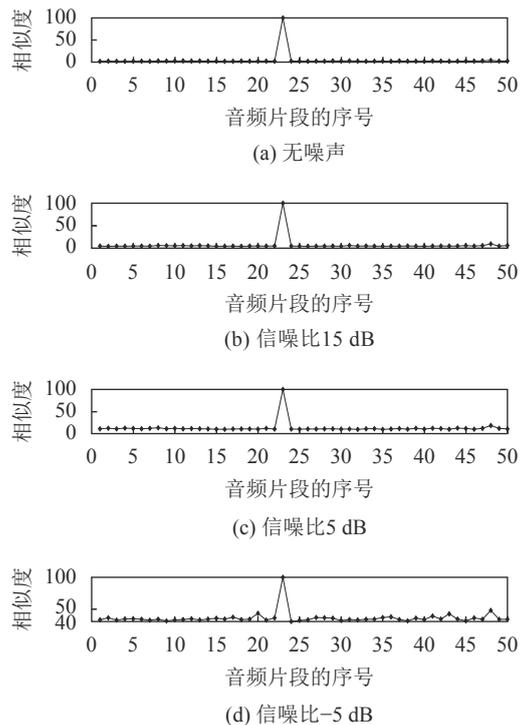


图 5 基于倒谱分析法所测出的相似度

Fig. 5 Similarity based on cepstrum analysis

带通滤波器a处理,模版库中的50个音频分别经过带通滤波器b处理。以第23个音频作为测试样本,与模版库中的50个音频依次用倒谱分析法和欧式距离法做比对,归一化的欧式距离和相似度结果见图6和7。

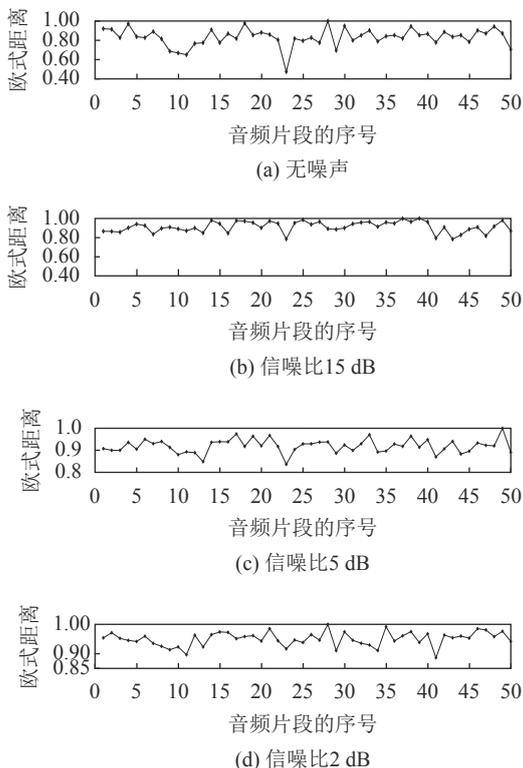


图6 基于MFCC特征的欧式距离法比对经过不同带宽带通滤波器处理的两音频所测出的欧氏距离

Fig. 6 Euclidean distance measured of two audio processed by different bandpass filters for Euclidean distance method based on MFCC features

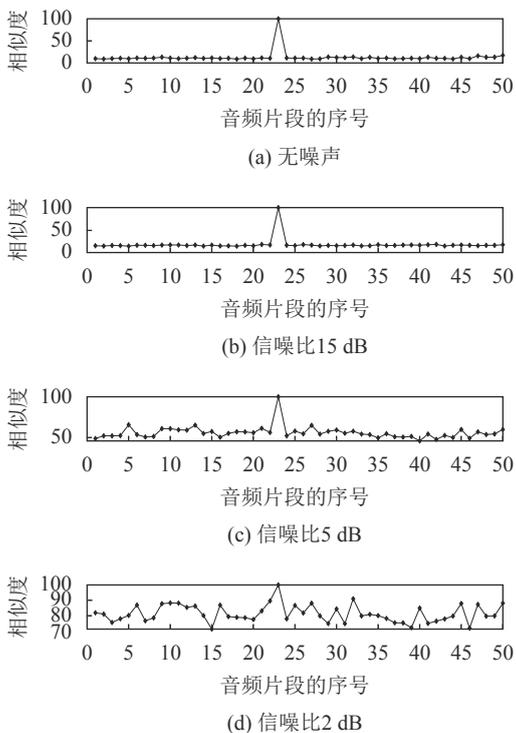


图7 基于倒谱分析法比对经过不同带宽带通滤波器处理的两音频所测出的相似度

Fig. 7 Similarity between two audio processed by different bandpass filters based on cepstrum analysis

从图6和7可以看出,待比对的两音频经过不同带宽的带通滤波器处理后,欧氏距离法在无噪声条件下是有效的,而在信噪比(SNR<15 dB)时该方法不能准确地找出相似的音频片段。所提出的倒谱分析法依然非常稳健,即使在低信噪比(SNR=2 dB)时也可以找到正确匹配的音频片段。

3.3 不同信噪比下不同延时两音频延时估计结果和比对结果比较

将测试样本(第23个)音频依次加入0.3、2.0、6.0 s这三种不同延时后作为待比对音频 $x_1(t)$,不添加延时的测试样本作为另一待比对音频 $x_2(t)$ 。

3.3.1 不同信噪比下延时估计和比对未经过带通滤波器处理的两音频

让不同延时的两音频片段在不同信噪比下用所提出的倒谱分析法计算两音频的延时和相似度,结果如图8和9所示,图中, τ 代表两音频之间的真实延时。

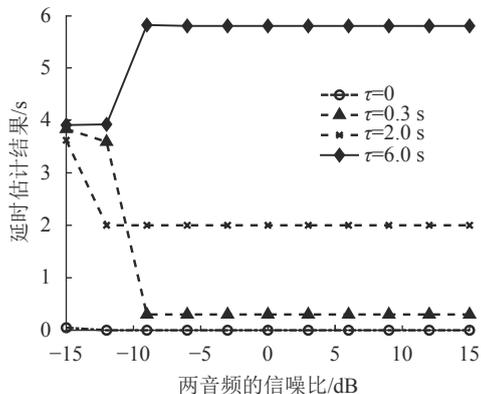


图8 不同信噪比下两音频延时估计结果

Fig. 8 Results of delay estimation for two audio with different SNR

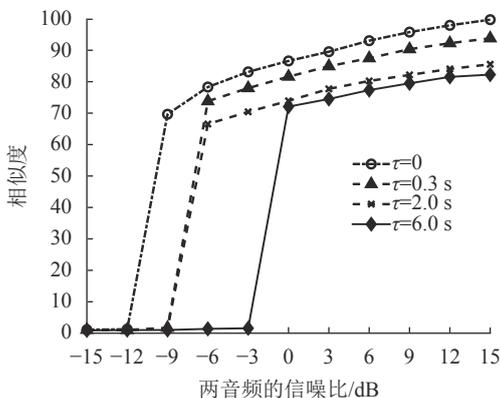


图9 不同信噪比下两音频相似度比对结果

Fig. 9 Similarity results of two audio with different SNR

从图8和9可以看出,所提出的倒谱分析法在信噪比较低(-5 dB)时仍可以得到准确的延时估计结果,而相似度会随着两音频延时增加或信噪比降低变得不准确,但是其抗噪声性能仍然在可接受范围内。

3.3.2 不同信噪比下延时估计和比对经过不同通带的带通滤波器处理后的两音频

将 $x_1(t)$ 经过带通滤波器a处理, $x_2(t)$ 经过带通滤波器b处理,再在不同信噪比下用所提出的倒谱分析法计算两音频的延时和相似度,结果如图10和11所示。由图10和11可以看出,当两音频经过不同通带的带通滤波器处理后再进行延时估计和比对时,在信噪比为2 dB条件下,所提出的倒谱分析法在两音频片段频率带宽不同,并且低信噪比下仍然可以得到比较准确的延时估计结果和相似度。

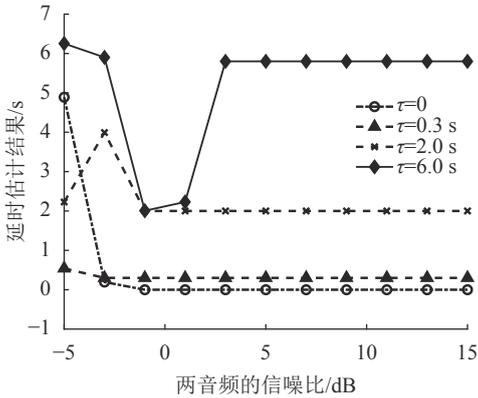


图 10 不同信噪比下经过不同带宽的带通滤波器处理后的两音频延时估计结果

Fig. 10 Results of delay estimation for two audio with different SNR and processed by different bandwidth filters

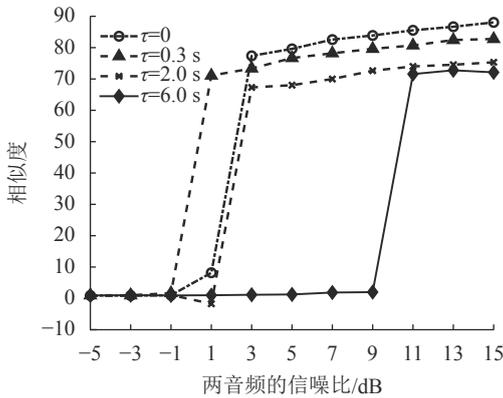


图 11 不同信噪比下经过不同带宽的带通滤波器处理后的两音频相似度比对结果

Fig. 11 Similarity results of two audio with different SNR and processed by different bandwidth filters

3.4 大样本库音频比对

为了验证算法的准确性和鲁棒性,又从50个不同电台总共100 h广播音频源中随机选择了500个音频片段(每个片段持续10 s),其中,包含3种不同语种的语音和各种类型的音乐,如古典、乡村、爵士乐、流行、舞蹈、摇滚等。每个音频片段都经过带通滤波器b处理,同时,在每个音频片段中加入加性高斯白噪

声,获得不同信噪比的音频片段。500个音频片段依次作为测试样本,先经过带通滤波器a处理,再将其和模版库中的500个音频片段依次计算相似度。匹配精度定义为正确匹配数与音频片段总数之比。比对完成后计算了不同信噪比下所提出倒谱分析法和欧氏距离法的匹配精度,结果如表3所示。

表 3 不同信噪比下两种方法的匹配精度对比

Tab. 3 Comparison of matching accuracy of two methods under different SNR

两音频的信噪比/dB	匹配率/%		提高的匹配率/%
	欧氏距离法	倒谱分析法	
无噪声	100.00	100.00	—
15	60.71	93.56	32.85
5	13.73	92.47	78.74
2	2.53	90.36	87.83

如表3所示:当无噪声条件下时,这两种算法可以实现高精度。随着信噪比的降低,欧氏距离法的匹配精度迅速下降,而所提出的算法仍保持可接受的精度。因此可以得出结论,本文提出的倒谱分析法比欧氏距离法得到的音频相似度更准确、更稳健,并且可以在得到相似度结果的同时估计出两音频之间的延时。

3.5 算法运算速度比较

为验证算法速度,在同一信噪比(两音频的信噪比为15 dB)、同一实验平台环境下,计算所提出的倒谱分析法和欧氏距离法每次比对所用的时间,如图12所示。由图12可知,所提出的倒谱分析法平均每次运行时间远小于欧氏距离法。其原因是欧氏距离法需要先分别提取两音频的MFCC特征矩阵,再计算两个MFCC特征矩阵之间的欧氏距离,找到最小距离从而得到相似度,因此每次所需要的运行时间就更长。

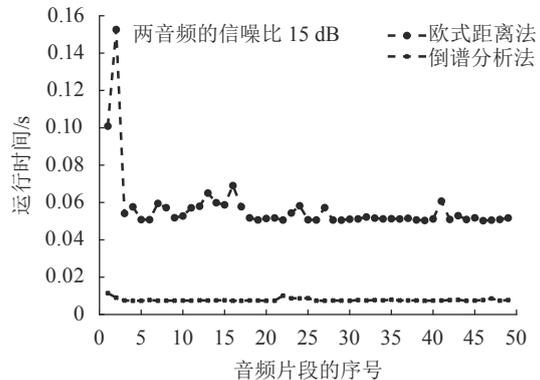


图 12 两种方法计算相似度的运行时间

Fig. 12 Average running time of two methods when calculating similarity

4 结论

为提高低信噪比和存在延时的两广播音频在实

时环境下比对的准确度,提出了一种基于倒谱分析的音频相似度快速计算方法。针对广播音频的特点和音频比对的应用需求,将倒谱分析改进后应用于两音频比对,且计算速度较快,可达到实时比对的要求,能够减少音频相似度错误报警率。该方法不需要对待比对的两音频提取MFCC等音频特征参数,直接通过倒谱分析得出两音频之间的延时估计结果和相似度比对结果。实验结果表明,相比于基于音频特征距离的传统音频比对方法,本文算法的音频比对结果更为准确,同时还能得到延时估计结果。当SNR分别为15、5和2 dB时,本文算法的音频比对匹配精度分别比欧氏距离法提高了32.85%、78.74%和87.83%,每次比对所需要的平均运行时间也缩短。

采用了所提出的倒谱分析音频相似度快速计算方法,有助于提高实时广播音频的比对效率和准确度,大幅度降低音频相似度错误报警率。由于环绕音较强的音频对本文算法性能影响显著,未来将深入研究如何更准确地比对该类型的音频,以进一步提升算法性能。此外,本文算法虽能对两音频进行延时估计,但不能进一步判断出延时的正负情况,如何结合其他算法来达到更准确的延时估计结果也是下一步的研究方向。

参考文献:

- [1] 黄思贝.无线广播电视发射台自动化信号监测系统研究[J].*视听*,2019(2):235-236.
- [2] 赵洪春.广播电视发射台监测监控系统设计[J].*通讯世界*,2019,26(4):46-47.
- [3] Liu Chunjiang.Radio program intelligent monitoring system based on audio signal comparison[J].*Radio and TV Broadcast Engineering*,2018,45(4):100-104.[刘春江.基于音频信号比对的无线广播节目智能监测系统[J].*广播与电视技术*,2018,45(4):100-104.]
- [4] Fu Rui,Zhang Dingjing,Niu Tailong,et al.Research on the components of broadcasting monitoring system[J].*China Digital Cable TV*,2019(2):146-148.[付瑞,张定京,牛泰龙,等.广播监测系统构成要素研究[J].*中国有线电视*,2019(2):146-148.]
- [5] 原瑾.数字化中波同步广播单频网的音频时延分析[J].*数字通信世界*,2018(10):102.

- [6] Zhang Lin,Zhou Tao,Du Qingzhi,et al.Audio comparison algorithm based on physical characteristics[J].*Video Engineering*,2017,41(11/12):110-114.[张琳,周韬,杜庆治,等.基于物理特征的音频相似度比对算法研究[J].*电视技术*,2017,41(11/12):110-114.]
- [7] Guo Xingji,Fan Bingqi.Feature-based comparison of audio[J].*Journal of Henan Normal University(Natural Science)*,2006,34(2):35-38.[郭兴吉,范秉琪.基于特征的音频比对技术[J].*河南师范大学学报(自然科学版)*,2006,34(2):35-38.]
- [8] Huang Yuyuan,Yin Ying,Zhao Yue,et al.Acoustic inspection of internal defect in magnetic tile based on bispectrum analysis[J].*Journal of Sichuan University (Engineering Science Edition)*,2014,46(5):188-194.[黄沁元,殷鹰,赵越,等.基于双谱分析的磁瓦内部缺陷音频检测方法[J].*四川大学学报(工程科学版)*,2014,46(5):188-194.]
- [9] Yu Xiaoqing,Pan Xueqian,Yang Wei,et al.Audio similarity measure based on Renyi's quadratic entropy[C]//Proceedings of the 2010 International Conference on Audio,Language and Image Processing.*Shanghai:IEEE*,2010:722-726.
- [10] Sonnleitner R,Widmer G.Robust quad-based audio fingerprinting[J].*IEEE/ACM Transactions on Audio,Speech,and Language Processing*,2016,24(3):409-421.
- [11] 电子科技大学,北京英夫美迪数字技术有限公司.一种快速广播音频比对方法:CN201510244557.4[P].2015-10-21.
- [12] Ye Yuqing,Zhang Shaoran,Hong Weijun,et al.Audio similarity comparison system of english dubbing on android platform[C]//Proceedings of the 2017 IEEE International Conference on Information and Automation.*Macau:IEEE*,2017:692-697.
- [13] Yao Shanshan,Niu Baoning,Liu Jianquan.Audio identification by sampling sub-fingerprints and counting matches[J].*IEEE Transactions on Multimedia*,2017,19(9):1984-1995.
- [14] Jaiyen N,Hantula P,Tongta R.Real-time audio similarity comparison algorithm[C]//Proceedings of the 2017 IEEE 15th Student Conference on Research and Development (SCOREd).*Putrajaya:IEEE*,2017:477-480.

(编辑 赵婧)

引用格式: Shao Yubin,Tang Chuanlin,Zhao Zhirou,et al.Fast similarity comparison algorithm for real-time broadcast audio based on cepstrum analysis[J].*Advanced Engineering Sciences*,2020,52(3):178-185.[邵玉斌,唐传林,赵至柔,等.基于倒谱分析的实时广播音频相似度快速比对算法[J].*工程科学与技术*,2020,52(3):178-185.]