

一种改进的分数时延估计方法及应用

刘超^{1,2} 柏业超^{1,2} 张兴敢^{1,2}

(1. 南京大学电子科学与工程学院, 南京, 210023; 2. 东南大学毫米波国家重点实验室, 南京, 210096)

摘要: 传统的分数时延估计算法对环境噪声和混响噪声比较敏感, 在复杂的实际环境中, 算法性能会严重下降。为进一步提高时延估计算法性能, 提出一种基于广义互相关 (Generalized cross correlation, GCC) 改进算法的广义互相关-最大似然相位补偿 (GCC-Maximum likelihood phase compensation, GCC-MLP) 分数时延估计算法。该算法改进了 GCC 频域加权函数, 并将线性相位补偿应用于频域互相关谱, 获得连续的分数时延估计值, 进一步提高了分数时延估计的精确性。仿真结果表明, GCC-MLP 相位补偿分数时延估计算法增强了对环境噪声和混响噪声的鲁棒性, 减小了时延估计误差, 算法性能优于曲线拟合、Sinc 插值等传统分数时延估计算法。

关键词: 广义互相关; 分数时延估计; 麦克风阵列; 声源方向角估计; 相位补偿

中图分类号: TN911.72 **文献标志码:** A

Improved Fractional Time Delay Estimation Method and Its Application

Liu Chao^{1,2}, Bai Yechao^{1,2}, Zhang Xinggan^{1,2}

(1. School of Electronic Science and Engineering, Nanjing University, Nanjing, 210023, China; 2. State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, 210096, China)

Abstract: Traditional fractional time delay estimation (TDE) algorithm is sensitive to ambient noise and reverberation noise, and the algorithm performance is severely degraded in the complex environment. To improve the accuracy of time delay estimation, a GCC-maximum likelihood phase compensation (GCC-MLP) phase compensation fractional time delay estimation approach is proposed based on generalized cross correlation (GCC). The frequency domain weighting function is also improved for GCC algorithm. At the same time, linear phase compensation is applied to the cross correlation spectrum in frequency domain and a continuous time delay estimator is obtained. Simulation comparison between the proposed algorithm and the other fractional TDE algorithms demonstrates that the proposed algorithm outperforms the curve fitting algorithm and the sinc interpolation algorithm. Meanwhile the robustness to the noisy and the reverberant conditions is reinforced, the time delay estimation error also decreases. Therefore, GCC-MLP phase compensation fractional time delay estimation algorithm can be widely used in sound source direction of arrival estimation and sound source localization.

Key words: generalized cross-correlation (GCC); fractional time delay estimation; microphone array; sound source direction of arrival estimation; phase compensation

引言

基于麦克风阵列的声源定位技术在视频会议、语音识别以及可视/车载电话等领域具有广阔的应用前景^[1-3]。时延估计是麦克风阵列声源角度估计和声源定位算法中最常用、也是最重要的一种方法^[4-5]。分数时延估计通常分为两个阶段。第1阶段利用广义互相关(Generalized cross correlation, GCC)算法进行整数延时单元估计。文献[6]中,根据不同频域加权函数,提出了广义互相关-相位变换(Phase transform, PATH)、广义互相关-最大似然(Maximum likelihood, ML)等一系列算法。第2阶段进行分数延时单元估计。文献[7]提出了曲线拟合分数延时估计算法,这种算法实现简单,但分数时延估计误差比较大。文献[8]采用 sinc 插值的分数时延估计算法,延时估计精度高,但是该算法受混响噪声的影响较大。本文提出一种基于 GCC 改进算法的 GCC-MLP 相位补偿分数延时估计算法。该算法改进 GCC 算法频域加权函数,增强算法对于环境噪声和混响噪声的鲁棒性;同时,将线性相位补偿和抛物线拟合应用于频域互相关谱,获得连续的分数时延估计值,进一步提高了分数时延估计的精确性,减小了声源角度估计误差。

1 麦克风阵列信号模型

麦克风阵列信号模型的平面示意图如图1所示。假设两麦克风 Mic1, Mic2 间距为 d (约为 5 cm), 麦克风阵列中心为 O , 声源偏离麦克风阵轴线夹角为 θ ($0^\circ \sim 180^\circ$)。在实际环境中,两麦克风接受到的信号 $x_1(t)$ 和 $x_2(t)$ 表示为

$$x_1(t) = s(t - \tau_1) + h_1(t) * s(t) + n_1(t) \quad (1)$$

$$x_2(t) = s(t - \tau_2) + h_2(t) * s(t) + n_2(t) \quad (2)$$

式中: $s(t)$ 为声源信号, $h_1(t)$ 和 $h_2(t)$ 是房间混响的冲激响应, “*” 表示信号卷积, $n_1(t)$ 和 $n_2(t)$ 为环境噪声, τ_1 和 τ_2 是声波从声源到麦克风的传播时间。

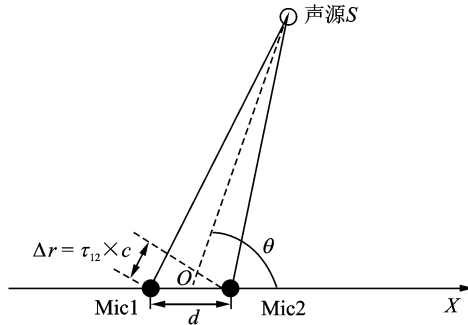


图1 麦克风阵列平面示意图

Fig. 1 Schematic plan of dual microphone array

由于两麦克风间距为 d 很小, 声源位于麦克风阵列的远场, 由几何关系可得声源位置的方向角

$$\theta = \arccos\left(\frac{\tau_{12} \cdot c}{d}\right) \quad (3)$$

式中: $\tau_{12} = \tau_1 - \tau_2$, 即两麦克风接收信号间的时延, c 为声速。由式(3)可知, 估计得到 τ_{12} 便可确定声源位置的方向角 θ 。

2 GCC-MLP 相位补偿分数时延估计算法

2.1 GCC 时延估计算法

传统 GCC 时延估计算法流程图如图 2 所示。时延估计 τ_{12} 可以表示为

$$\hat{\tau}_{12} = \arg \max_m R_{12}(n) \tag{4}$$

式中: $R_{12}(n) = \sum_{k=0}^{N-1} \varphi(k) X_1(k) X_2^*(k) e^{j2\pi n \frac{k}{N}}$ 。 $R_{12}(n)$ 称为广义互相关函数, n 表示时延采样单元(采样周期 T 的整数倍), N 表示每帧信号采样点数。 $X_1(k)$ 和 $X_2(k)$ 分别为 $x_1(n)$ 和 $x_2(n)$ 的傅里叶变换, $(\cdot)^*$ 表示取复共轭, $\varphi(k)$ 为频域加权函数。

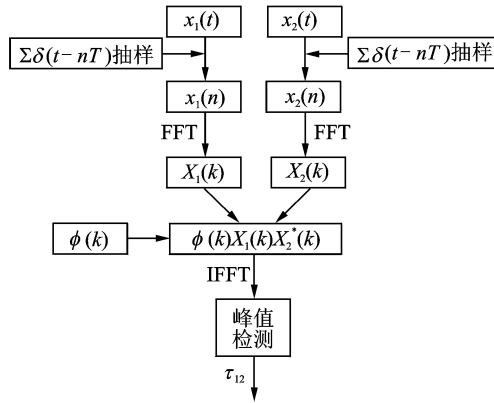


图 2 GCC 算法流程图
Fig. 2 Flowchart of GCC algorithm

由式(1,2)可以看出,在时延估计系统中存在两种噪声:环境噪声 $n_i(t)$ 和混响噪声 $h_i(t) * s(t)$ 。选择合适的加权函数对于提高时延估计的算法性能有重要意义。文献[6]研究表明,GCC-ML 加权函数对于环境噪声鲁棒性较强,而 GCC-PATH 加权函数对于混响噪声有较强抑制作用,两种算法的加权函数 $\varphi(k)$ 如式(5,6)所示。

$$\varphi_{ML}(k) = \frac{|X_1(k)| |X_2(k)|}{|N_2(k)|^2 |X_1(k)|^2 + |N_1(k)|^2 |X_2(k)|^2} \tag{5}$$

$$\varphi_{PATH}(k) = \frac{1}{|X_1(k) X_2^*(k)|} \tag{6}$$

式中: $N_1(k)$ 和 $N_2(k)$ 分别表示环境噪声 $n_1(t)$ 和 $n_2(t)$ 的傅里叶变换。

在实际环境中,环境噪声和混响噪声并存且较严重时,采用单一的 ML 或 PATH 加权函数,时延估计的准确性会大大降低,在此基础上进行曲线拟合和 Sinc 插值等分数时延估计时,算法性能会严重下降。因此,为加强算法性能对于噪声的鲁棒性,提高时延估计精度,提出了广义互相关-最大似然相位补偿(GCC-Maximum likelihood phase compensation, GCC-MLP)相位补偿分数时延估计算法。

2.2 GCC-MLP 相位补偿分数时延估计算法

对式(1,2)两边进行傅里叶变换,可得

$$X_i(k) = S_i(k) e^{-jk\tau} + H'_i(k) S_i(k) + N_i(k) \quad i = 1, 2 \tag{7}$$

由式(7)可得到信号频域能量等式

$$|X_i(k)|^2 = |S_i(k)|^2 + |H'_i(k)|^2 |S_i(k)|^2 + |N_i(k)|^2 \tag{8}$$

式中: $|S_i(k)|^2$ 表示直达信号能量, $|H_i'(k)|^2 |S_i(k)|^2$ 表示混响分量能量, $|N_i(k)|^2$ 表示环境噪声能量。如果将混响当作另一种形式的噪声, 则总的噪声能量 $|N_{IT}(k)|^2$ 可以简单地表示为

$$|N_{IT}(k)|^2 = |H_i'(k)|^2 |S_i(k)|^2 + |N_i(k)|^2 = p |S_i(k)|^2 + |N_i(k)|^2 \quad (9)$$

式中 p 表示混响能量比例系数。由式(8,9)化简可得, 总的噪声能量

$$|N_{IT}(k)|^2 = \frac{p}{(1+p)} \times (|X_i(k)|^2 - |N_i(k)|^2) + |N_i(k)|^2 \quad (10)$$

令 $q = p/(1+p)$, 则

$$|N_{IT}(k)|^2 = q |X_i(k)|^2 + (1-q) |N_i(k)|^2 \quad (11)$$

将式(10)代入式(5)中, 可得改进的频域加权函数

$$\varphi_{MLP}(k) = \frac{|X_1(k)| |X_2(k)|}{2q |X_2(k)|^2 + (1-q) (|N_2(k)|^2 |X_1(k)|^2 + |N_1(k)|^2 |X_2(k)|^2)} \quad (12)$$

事实上, $\varphi_{MLP}(k)$ 可以简单地写成 ML 加权函数和 PATH 加权函数的线性表达, 即

$$\frac{1}{\varphi_{MLR}(k)} = q \frac{1}{\varphi_{PATH}(k)} + (1-q) \frac{1}{\varphi_{ML}(k)} \quad (13)$$

GCC-MLP 算法中, $\varphi_{MLP}(k)$ 加权函数兼具了 GCC-PATH 和 GCC-ML 算法的优点, 将 $\varphi_{MLP}(k)$ 加权函数代入式(4), 可以估计得到整数采样时延单元值 N_1 。根据实际环境中混响能量, 合理地确定 q 值, 能够有效地抑制混响噪声和环境噪声, 大大提高整数采样时延单元估计准确率。进一步地, 在整数采样时延单元估计的基础上, 可以进行分数采样时延单元估计。传统的分数时延算法, 如曲线拟合、sinc 插值等, 这些方法往往是在时域对广义互相关函数 $R_{12}(n)$ 进行曲线拟合, 这类方法同样对环境噪声和混响噪声比较敏感, 限制了分数延时估计精度的进一步提高。

由傅里叶变换性质可知, 信号在时域的时延, 在频域表现为相位变化, 并且这一相位变化随频率呈线性关系。因此, 通过在互相关频域补偿的线性相位可以估计信号时延, 由于相位变化的连续性, 估计得到的时延单元精度可以小于一个时间采样单元。GCC-MLP 相位补偿算法利用经过 $\varphi_{MLP}(k)$ 加权的互相关谱估计时延, 能有效抑制环境噪声和混响噪声。假设实际的时延值为 N_s , 在整数采样时延单元 N_1 估计准确的情况下, $N_s \in (N_1 - 0.5, N_1 + 0.5]$ 。根据离散傅里叶变换公式, 在 $(N_1 - 0.5, N_1 + 0.5]$ 区间的互相关频谱 $R(m)$ 为

$$R_{12}(m) = \sum_{k=0}^{N-1} \varphi(k) X_1(k) X_2^*(k) e^{j2\pi m \frac{k}{C}}$$

$$m = N_1 - \frac{1}{2}, N_1 - \frac{1}{2} + \frac{1}{C}, \dots, N_1 + \frac{1}{2} \quad (14)$$

式中: C 表示插值率(大于 1 的正整数)。假设此时互相关频谱 $R_{12}(m)$ 的绝对值峰值位于 N_2 处, 那么在 $R(N_2)$ 附近采用抛物线近似, 根据式(15)可求得精确时延单元估计值

$$N' = N_2 - \frac{1}{2C} \frac{\left| R\left(N_2 + \frac{1}{C}\right) \right| - \left| R\left(N_2 - \frac{1}{C}\right) \right|}{\left| R\left(N_2 + \frac{1}{C}\right) \right| + \left| R\left(N_2 - \frac{1}{C}\right) \right| - 2 \left| R(N_2) \right|} \quad (15)$$

3 仿真数据结果

3.1 仿真参数

为验证算法的有效性, 模拟声学环境, 平面示意图如图 3 所示, 假设声源和麦克风阵列都在同一水平高度。

(1) 房间模型几何参数房间大小为 $6 \text{ m} \times 4 \text{ m} \times 3 \text{ m}$; 声源 S 坐标 (X_s, Y_s) 为 $(3.6, 3.8)$; 麦克风阵列

中心坐标为 $O(3,1)$; 麦克风间距为 $d=5\text{ cm}$ 。

(2) 噪声模型参数: 环境噪声为高斯白噪声(信噪比 $\text{SNR}=0\sim 20\text{ dB}$); 混响噪声为 IMAGE 模型^[9](混响时间 $T_{60}=100\sim 400\text{ ms}$)。

(3) 语音信号参数: 采样频率为 $F_s=32\text{ kHz}$; 每帧采样点数为 $N=4\ 096$; 语音信号长度为 $L=4\ 096\times 100$ 帧。根据仿真参数计算可得, 声源 S 信号分别到达 Mic1 和 Mic2 的时延 τ_{12} 所对应的理论时延单元 $N_1=1.487$, 声源偏向角 $\theta=71.56^\circ$ 。

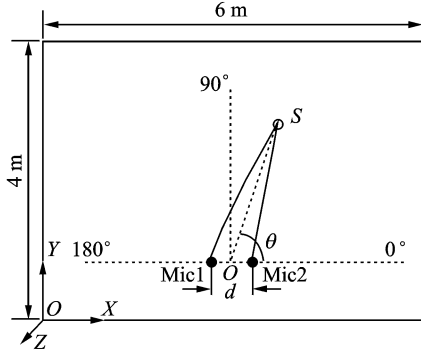


图3 模拟声学环境平面示意图

Fig. 3 Schematic plan of simulated acoustic environment

3.2 整数时延单元估计正确率分析

利用 GCC-MLP 算法进行整数部分时延估计, 考察整数时延估计的准确率, 并与 GCC-ML 算法和 GCC-PATH 算法进行性能比较。假设估计得到整数时延值为 N_1 , 若满足 $N_s \in (N_1 - 0.5, N_1 + 0.5]$, 则表示整数部分时延估计准确。图 4, 5 分别为 3 种算法在混响时间 $T_{60}=100\text{ ms}(q=0.1)$ 和 $T_{60}=400\text{ ms}(q=0.3)$ 条件下, 整数时延估计准确率随信噪比 SNR 变化曲线图。对比分析图 4, 5, 可以发现:

(1) GCC-ML 算法对于环境噪声的鲁棒性较强, 在低信噪比 ($\text{SNR}=0\sim 5\text{ dB}$) 条件下, 整数时延估计准确率远高于 GCC-PATH; 但是 GCC-ML 算法对于房间混响比较敏感, 当房间混响较大时, 算法性能会大大降低。

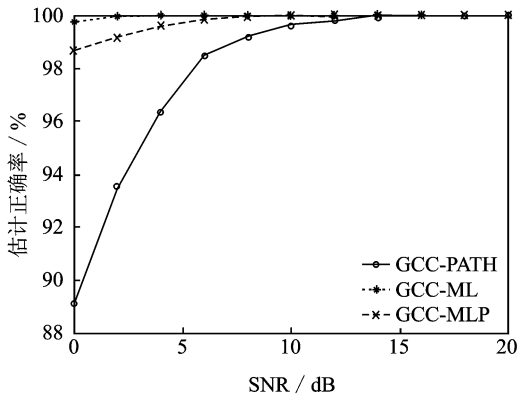


图4 整数时延估计准确率曲线 1 ($T_{60}=100\text{ ms}$)

Fig. 4 The first accuracy rate curve of integer delay estimation ($T_{60}=100\text{ ms}$)

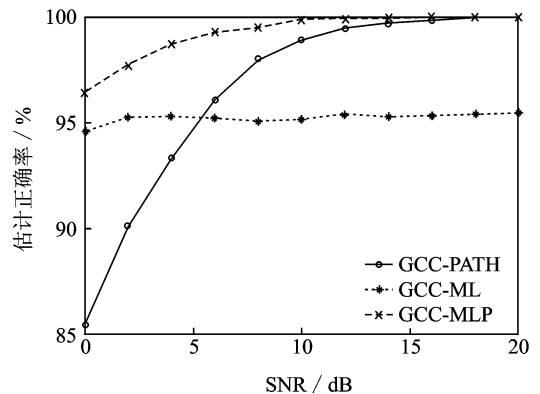


图5 整数时延估计准确率曲线 2 ($T_{60}=400\text{ ms}$)

Fig. 5 The second accuracy rate curve of integer delay estimation ($T_{60}=400\text{ ms}$)

(2)GCC-PATH 算法能有效抑制混响噪声,在高混响($T_{60}=400$ ms)、高信噪比($SNR=10\sim 20$ dB)情况下,算法性能优于 GCC-ML,但是 GCC-PATH 对于环境噪声鲁棒性较差,当信噪比较低时,算法性能会严重下降。

(3)GCC-MLP 算法结合了 GCC-PATH 算法和 CGG-ML 算法的优点,对混响噪声和环境噪声都有较强的抑制作用。在高混响、低信噪比条件下,整数时延估计准确率仍能达到 95% 以上。

3.3 分数时延单元估计误差分析

在利用 GCC-MLP 算法整数时延单元估计的基础上,分别采用 MLP-相位补偿、MLP-曲线拟合和 MLP-Sinc^[10] 插值 3 种算法进行分数采样时延单元估计,其中插值率 $C=32$ 。在不同混响条件下,时延估计均方根误差随信噪比变化曲线如图 6,7 所示。图 6 中,混响时间 $T_{60}=100$ ms,相同信噪比条件下,MLP-相位补偿算法估计所得时延均方根误差远远小于其他两种算法,尤其在低信噪比($SNR=0\sim 5$ dB)情况下,算法优越性充分得以体现。图 7 对比图 6,混响时间 T_{60} 由 100 ms($q=0.1$)变为 400 ms($q=0.3$),相同信噪比条件下,相同算法的时延估计均方根误差均有所增加,但是比较误差增加量可以发现,MLP-相位补偿算法受混响噪声影响较小。由分析可知,MLP-相位补偿分数时延估计算法对于环境噪声和混响噪声都有较强抑制,算法鲁棒性优于 MLP-曲线拟合和 MLP-Sinc 插值算法。

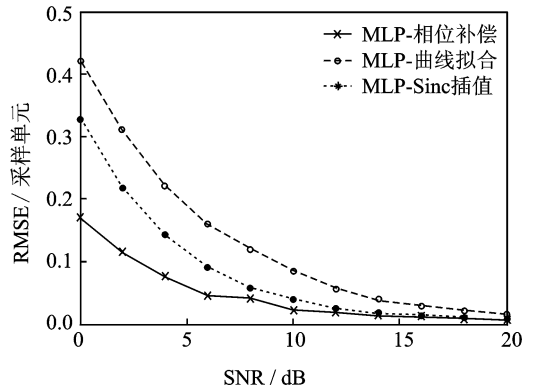


图 6 分数时延均方根误差曲线($T_{60}=100$ ms)
Fig. 6 RMSE curve of subsample delay estimation ($T_{60}=100$ ms)

3.4 声源偏向角角度估计误差分析

根据估计得到的时延值,由式(3)可以求得声源偏向角。图 8 所示为同一混响条件下声源偏向角度均方根误差随信噪比变化曲线。由图可知,即使在高混响条件($T_{60}=400$ ms)下,GCC-MLP 相位补偿算法估计声源偏向角的角度误差 $\Delta\theta < 2^\circ$,角度估计误差小于其他两种算法。

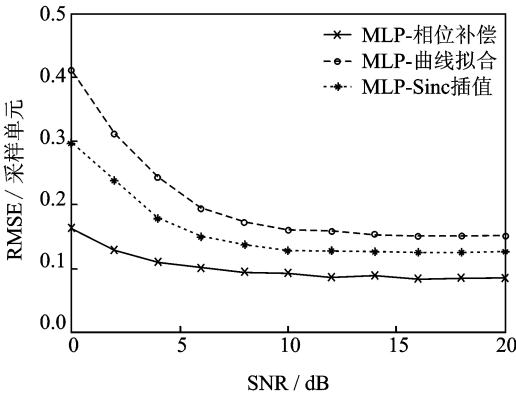


图 7 分数时延均方根误差曲线($T_{60}=400$ ms)
Fig. 7 RMSE curve of subsample delay estimation ($T_{60}=400$ ms)

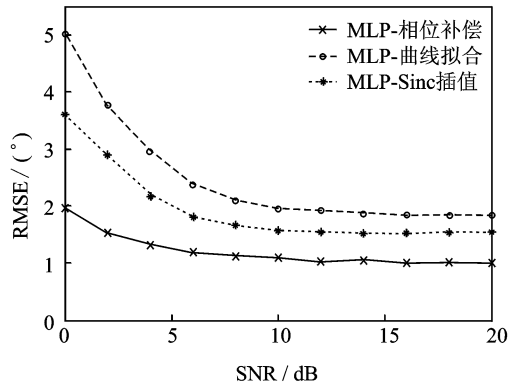


图 8 声源偏向角度均方根误差曲线($T_{60}=400$ ms)
Fig. 8 RMSE curve of deflection angle of sound source ($T_{60}=400$ ms)

4 结束语

在实际环境中,采用传统的分数时延估计算法,时延估计值受到环境噪声和混响噪声影响,声源角度估计误差较大。本文提出了一种 GCC-MLP 相位补偿分数时延估计算法,该算法改进了 GCC 频域加权函数,综合了 GCC-ML 和 GCC-PATH 加权函数的优点;与此同时,利用线性相位补偿和抛物线拟合,实现了连续的分数时延估计。GCC-MLP 相位补偿分数时延估计算法相比于曲线拟合 sinc 插值算法,增强了算法对于噪声和声源位置的鲁棒性,进一步提高了时延估计精度。通过两阶段的时延估计, GCC-MLP 算法能够实现精确的声源角度估计,角度估计误差小于 2° 。GCC-MLP 相位补偿分数时延估计算法可以广泛用于声源角度估计和声源定位。

参考文献:

- [1] 王瑞, 陈砚圃, 张恒. 声源定位中波达角一致性检测方法[J]. 数据采集与处理, 2010, 25(6): 756-760.
Wang Rui, Chen Yanpu, Zhang Heng. Method for detecting DOA homogeneity in sound source localization[J]. Journal of Data Acquisition and Processing, 2010, 25(6): 756-760.
- [2] 周康辉, 董万胜, 刘恒毅, 等. 利用二次相关改进的广义互相关时延估计算法[J]. 数据采集与处理, 2013, 28(6): 801-806.
Zhou Kanghui, Dong Wansheng, Liu Hengyi, et al. Improved generalized cross correlation method for time delay estimation by using second correlation[J]. Journal of Data Acquisition and Processing, 2013, 28(6): 801-806.
- [3] Wu W C, Chen O T C. Multiple-sound-source localization scheme based on feedback-architecture source separation[C]// 52nd IEEE International Midwest Symposium on Circuits and Systems. [S. l.]:IEEE, 2009: 669-672.
- [4] Bai Y C, Zhang X G, Tang L. Subsample time delay estimation based on phase spectrum of band limited stochastic signals [J]. Transactions of Nanjing University of Aeronautics & Astronautics, 2010, 27(2): 170-175.
- [5] Chen J C, Hudson R E, Yao K. Maximum-likelihood source localization and unknown sensor location estimation for wide-band signals in the near-field[J]. Signal Processing, IEEE Transactions on, 2002, 50(8): 1843-1854.
- [6] Perez-Lorenzo J M, Viciano-Abad R, Reche-Lopez P, et al. Evaluation of generalized cross-correlation methods for direction of arrival estimation using two microphones in real environments[J]. Applied Acoustics, 2012, 73(8): 698-712.
- [7] Mosorov V. Phase spectrum method for time delay estimation using twin-plane electrical capacitance tomography[J]. Electronics Letters, 2006, 42(11): 630-632.
- [8] Qin B, Zhang H, Fu Q, et al. Subsample time delay estimation via improved GCC PHAT algorithm[C]//2008 9th International Conference on Signal Processing. [S. l.]:IEEE, 2008: 2579-2582.
- [9] Allen J B, Berkley D A. Image method for efficiently simulating small-room acoustics[J]. The Journal of the Acoustical Society of America, 1979, 65(4): 943-950.
- [10] Nakamura K, Nakadai K, Okuno H G. A real-time super-resolution robot audition system that improves the robustness of simultaneous speech recognition[J]. Advanced Robotics, 2013, 27(12): 933-945.

作者简介:



刘超(1989-),男,硕士研究生,研究方向:信号与信息处理,E-mail:liuchao052796@126.com。

柏业超(1984-),男,博士,副教授,研究方向:信号与信息处理。



张兴敢(1956-),男,博士,教授,研究方向:信号与信息处理。

