

基于广义互相关改进的麦克风阵列声源定位方法

李保伟, 张兴敢*

(南京大学电子科学与工程学院, 南京, 210023)

摘要: 基于麦克风阵列的声源定位技术因其安全性和隐蔽性, 涉及军事和民用等诸多领域. 其定位方法大体可以分为三类: 基于到达时延估计 (Time Difference of Arrival, TDOA)、基于最大输出功率的可控波束形成和高分辨率谱估计, 其中基于广义互相关 (Generalized Cross Correlation, GCC) 的时延估计进行声源定位的算法因其运算量小、定位精度高而得到了广泛的应用. 但在实际应用中, 环境噪声的存在使传统的 GCC 加权算法的定位精度不高, 所以有一定的局限性. 针对这一问题, 构建三元直角三角形麦克风阵列, 利用时延估计进行声源定位, 改进传统的 PHAT 相位加权算法, 提出基于广义互相关的 SCOT/PHAT 联合加权时延估计算法. 通过对不同加权函数的广义互相关时延估计算法进行 Matlab 仿真对比, 结果表明: 在小信噪比环境下, 该算法较传统的广义互相关算法具有更强的抗噪能力.

关键词: 麦克风阵列, 时延估计, 声源定位, 联合加权

中图分类号: TN912.3

文献标识码: A

Improved microphone array sound source localization method based on generalized cross correlation

Li Baowei, Zhang Xinggan*

(School of Electronic Science and Engineering, Nanjing University, Nanjing, 210023, China)

Abstract: Because of its security and concealment, sound source localization technology based on microphone array involves many fields such as military and civil. The localization methods can be generally divided into three categories: time difference of arrival (TDOA), controllable beamforming based on maximum output power and high resolution spectrum estimation. Based on Generalized Cross Correlation (GCC) time delay estimation, the acoustic source localization algorithm has been widely used because of its small computation and high positioning accuracy. In practical application, due to the presence of environmental noise, the positioning accuracy of the traditional GCC weighted algorithm is low, so it has certain limitations. To solve this problem, this paper constructs a ternary right triangle microphone array which uses time delay estimation to locate the sound source and improves the traditional PHAT phase weighting algorithm. We propose a SCOT/PHAT joint weighted time delay estimation algorithm based on generalized cross-correlation. Through the Matlab simulation comparison of the generalized cross-correlation delay estimation algorithm with different weighted functions, the results show that the algorithm has stronger anti-noise ability than the traditional generalized cross-correlation algorithm under the circumstance of small signal to noise ratio.

Key words: microphone array, time delay estimation, sound localization, joint weighting

基于麦克风阵列的声源定位技术的研究在国内受到了越来越多的关注, 麦克风阵列的声源目

标定位技术可以定义为利用麦克风阵列采集声源目标, 经过一系列对声音信号的分析操作与处理

来确定声源的位置. 与雷达探测技术以及其他探测技术相比, 该技术有很多特点, 比如因其采用被动式工作原理, 所以探测隐蔽, 不容易被发现. 目前, 基于麦克风阵列的定位技术逐渐变成研究的主流, 涉及视频会议系统^[1]、汽车鸣笛定位、人机交互^[2]等.

经过多年发展, 基于麦克风阵列的声源定位技术已有较多的理论和方法, 定位方法大体可以分三类: 基于到达时延估计方法^[3]、基于最大输出功率的可控波束形成方法^[4]、高分辨率谱估计方法^[5]. 基于时延估计法主要根据声源信号到达两个不同位置麦克风的时间差计算出声源位置相对一组麦克风所在的双曲面, 通过多个双曲面交集来确定声源位置; 基于最大输出功率的可控波束形成方法主要利用波束形成技术, 不断调整阵列信号的接收方向, 同时对整个接收空间进行扫描, 接收能量最大的方向即为声源方向; 高分辨率谱估计方法主要是利用各个阵元信号之间的相关矩阵, 通过计算相关矩阵的空间谱获得声源的方向角. 其中, 基于时延估计的定位算法简单、定位精度高, 是目前声源定位信号处理中常用的方法.

1 基本模型

声波在传播的过程中会发生衰减, 而衰减因子与传播的距离成正比, 因此声波从声源到达不同阵元时幅度相差不大, 而因延迟产生的相位差则较大. 根据声源与麦克风阵列之间的距离, 可以将阵列模型分为近场和远场模型. 通常定义声源与阵列之间的距离 $r = \frac{2D^2}{\lambda}$ 为远近场临界值, 其中 λ 为声波波长, D 为阵列孔径. 当麦克风阵列与声源之间的距离 $r < \frac{2D^2}{\lambda}$ 时, 阵列处于近场模型, 麦克风阵列接收的声波是球面波^[6]. 反之, 当 $r > \frac{2D^2}{\lambda}$ 时, 阵列处于远场模型, 声源到各阵元的距离差相比于声源到麦克风阵列的距离可以忽略不计, 阵元间接收信号仅存在相位延迟且可将声波看作平面波.

假设线阵麦克风阵列的所有阵元接收到的信号只包含直达信号与加性噪声, 并且每个麦克风

之间的噪声相互独立, 如图 1 所示. 其中, S 表示声源, $M_1, M_2, M_3, \dots, M_N$ 表示 N 个麦克风, 根据空间采样定理, 每相邻两个麦克风之间的间距 d 在选择上必须满足 $d \leq \frac{\lambda_{\min}}{2}$.

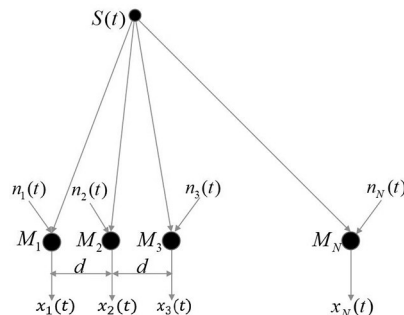


图 1 信号模型

Fig. 1 Signal model

声源信号 $s(t)$ 呈辐射状态传播, 且信号的能量与麦克风和声源间的距离成反比, 则第 i ($i = 1, 2, 3, \dots, N$) 个麦克风 M_i 所接收到的信号 $x_i(t)$ 为:

$$x_i(t) = \alpha_i s(t - \tau_i) + n_i(t) \quad (1)$$

其中, α_i 表示声源信号 $s(t)$ 的传播衰减因子, τ_i 表示声源 S 到达第 i 个麦克风的时延, $n_i(t)$ 表示第 i 个麦克风接收的加性噪声.

2 互相关算法

基于时延估计的声源定位算法主要分为两个步骤: 第一步, 估计声源信号到达各个麦克风之间的时间差; 第二步, 利用上一步所得到的时间差和声源与麦克风阵列之间的几何关系来估计声源位置. 在声源定位系统中, 时延估计技术至关重要, 时延估计的精度直接影响定位的精度. 常用的时延估计算法是互相关法.

互相关函数是描述随机信号 $x(t), y(t)$ 在任意两个不同时刻 s, t 的取值间的相关程度, 其定义为:

$$R(s, t) = E(X(s) \times Y(t)) \quad (2)$$

对于连续函数, 有定义:

$$(f \times g)(\tau) = \int_{-\infty}^{+\infty} f^*(t) g(t + \tau) dt \quad (3)$$

对于离散数据, 有定义:

$$(f \times g)(n) = \sum_{-\infty}^{+\infty} f^*[m]g(n+m) \quad (4)$$

从以上定义式中可以看到,互相关函数和卷积运算类似,也是两个序列滑动相乘. 区别在于互相关的两个序列都不翻转,直接滑动相乘,求和;卷积的其中一个序列需要先翻转,然后滑动相乘,求和. 所以, $f(t)$ 和 $g(t)$ 做相关等于 $f^*(t)$ 和 $g(t)$ 做卷积.

根据维纳-辛钦定理,互相关函数与其互功率谱密度互为傅里叶变换对,则 $x_1(t)$ 和 $x_2(t)$ 的互相关函数又可以表示为:

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{+\infty} G_{x_1x_2}(\omega) e^{-j\omega\tau} d\omega = \int_{-\infty}^{+\infty} X_1(\omega) X_2^*(\omega) e^{-j\omega\tau} d\omega \quad (5)$$

其中, $X_1(\omega)$ 和 $X_2(\omega)$ 分别为 $x_1(t)$ 和 $x_2(t)$ 的傅里叶变换, $G_{x_1x_2}(\omega)$ 则称为麦克风信号 $x_1(t)$ 和 $x_2(t)$ 间的互功率谱, $X_2^*(\omega)$ 为 $X_2(\omega)$ 取复共轭. $R_{x_1x_2}(\tau)$ 对应最大峰值处的横坐标即为所要求的时延估计值.

2.1 广义互相关算法 在麦克风信号处理实际模型中,由于存在混响和噪声的影响导致 $R_{x_1x_2}(\tau)$ 的峰值不明显,降低了时延估计的精度. 因此,广义互相关法利用频域加权函数对信号滤波来突出响应信号部分的频谱成分,抑制噪声部分的频谱,再进行傅里叶反变换 IFFT 反变到时域,锐化 $R_{x_1x_2}(\tau)$ 在时延处的峰值,从而提高时延估计性能. 因此得到广义互相关函数 $R_{x_1x_2}(\tau)$, 即:

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{+\infty} \varphi_{12}(\omega) X_1(\omega) X_2^*(\omega) e^{-j\omega\tau} d\omega \quad (6)$$

其主要原理如图2所示.

图中 $(\bullet)^*$ 表示取复共轭. 与基本互相关相比,广义互相关算法增加了加权运算.

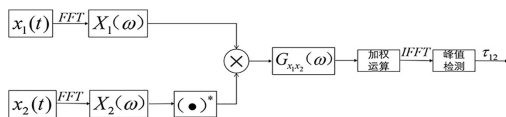


图2 广义互相关时延估计算法流程图

Fig. 2 Flow chart of generalized cross-correlation delay estimation algorithm

由于有限窗长和低信噪比的影响,选择合适的加权函数是一个难点,传统的广义互相关算法有基本互相关(CC),ROTH,SCOT^[7],PHAT^[8]等几种加权方式. 其中,SCOT和SCOT加权因子分别如式(7)和式(8)所示:

$$\varphi(\omega) = \frac{1}{\sqrt{G_{x_1x_1}(\omega)G_{x_2x_2}(\omega)}} \quad (7)$$

$$\varphi(\omega) = \frac{1}{|G_{x_1x_2}(\omega)|} \quad (8)$$

2.2 改进的广义互相关时延估计算法 同等环境下,PHAT加权要比其他几个加权算法的鲁棒性更强,但随着信噪比的降低,其时延估计精度也会下降. 如何提升低信噪比条件下的时延估计精度是当前研究的重要课题. 目前有学者提出谱减法去噪语音增强技术^[9]、SRP-PHAT与BP神经网络技术^[10]、归一化互功率谱-平滑相干变换联合加权^[11]、基于深度学习^[12]等一系列方法,都能有效地提升时延估计精度.

本文结合SCOT加权跟PHAT加权的优缺点,考虑加权因子应受两个通道的影响以及小信噪比环境下采样信号能量较小、单一采用PHAT算法时加权因子分母趋于零会导致误差变大等现象,提出一种SCOT/PHAT联合加权的GCC时延估计算法. 即引入新的加权因子:

$$\varphi(\omega) = \frac{\frac{1}{\sqrt{G_{x_1x_1}(\omega)G_{x_2x_2}(\omega)}}}{\frac{1}{|G_{x_1x_2}(\omega)|}} = \frac{|G_{x_1x_2}(\omega)|}{\sqrt{G_{x_1x_1}(\omega)G_{x_2x_2}(\omega)}} \quad (9)$$

该加权因子表明,信噪比越大,加权系数越趋近于1. 同时,定义变量均方根误差(Root Mean Square Error)^[13]:

$$\epsilon_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\tau_i - \tau_0)^2} \quad (10)$$

其中, τ_0 为真实时延值, τ_i 为通过算法求得值, N 为样本数.

基于时延估计的定位有插值计算^[14]和几何定位技术^[15]等,本文主要采用几何定位法. 其主

要原理是根据麦克风阵列的几何关系,利用几路信号的时延和阵元间距来解算声源的位置信息. 一个三元麦克风阵列,其阵元为 M_1, M_2, M_3 ,其中 M_1 与 M_2, M_1 与 M_3 之间的间距为 d, M_2 与 M_3 之间的间距为 $\sqrt{2}d$,接收来自同一声源的信号 $s(t)$,如图3所示.

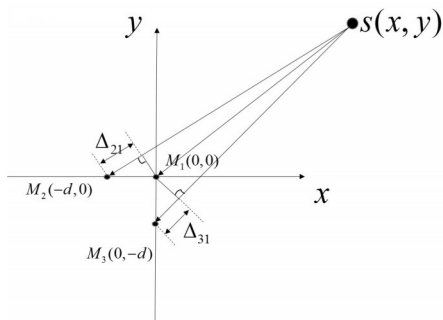


图3 进行声源定位的三元麦克风阵列

Fig. 3 Three-dimensional microphone array for sound source localization

图中 c 为声音在介质中的传播速度,假设以 M_1 为基准阵元建立直角坐标系,声源 S 坐标 (x, y) , τ_{12} 表示声源 S 到达 M_2 的延迟时间, τ_{13} 表示声源 S 到达 M_3 的延迟时间. 则声源到达其余两个麦克风阵元的波程差可表示为:

$$\Delta_{12} = \tau_{12} \cdot c \quad (11)$$

$$\Delta_{13} = \tau_{13} \cdot c \quad (12)$$

由距离差可以建立两个方程并联立方程组:

$$\begin{cases} \sqrt{(x+d)^2 + y^2} - \sqrt{x^2 + y^2} = \Delta_{12} \\ \sqrt{x^2 + (y+d)^2} - \sqrt{x^2 + y^2} = \Delta_{13} \end{cases} \quad (13)$$

通过解方程组,可以得到两个半双曲线的交点,即为可能的声源位置,最后由先验知识判定真实位置^[16]. 如图4所示.

3 仿真与分析

语音信号具有宽带特征和随机性,在用MATLAB仿真时采用高斯白噪声作为模拟声源信号,采样频率 $f_s = 48 \text{ kHz}$,采样点数 $N = 256$. 两个麦克风的间距 $d = 10 \text{ cm}$,假设两个接收信号之间的延迟点数为 $\Delta n = 10$,利用MATLAB自带的awgn函数产生平稳高斯白噪声. 在信噪比

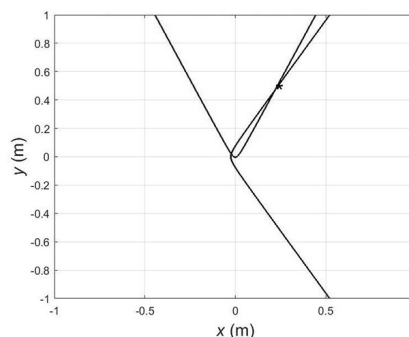


图4 利用双曲线解算位置信息

Fig. 4 Using hyperbola to solve position information

为 $-15 \sim 5 \text{ dB}$ 变化范围内分别采用PHAT加权函数算法和SCOT/PHAT联合加权函数算法进行时延估计,并进行10000次迭代取时延估计均方根误差,仿真结果如图5所示.

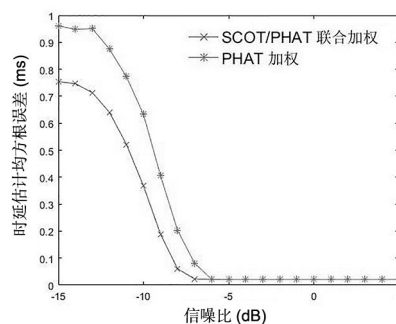


图5 两种不同算法的时延估计均方根误差随着信噪比的变换

Fig. 5 Transformation of root mean square difference of time delay estimation with signal-to-noise ratio for two different algorithms

通过分析比较可以发现,低信噪比条件下,两种算法都有一定的误差,但采用SCOT/PHAT联合加权算法求得的时延估计值误差更小、精度更高,抗噪能力更强;随着信噪比的增强,PHAT加权和SCOT/PHAT联合加权误差都会逐渐降低并最终趋于稳定. 同时,在信噪比高于 -5 dB 左右时两种算法精度基本一致,说明改进的联合算法更适用于 $\text{SNR} \leq -5 \text{ dB}$ 条件下. 当然,信噪比越低,精度也就越低.

取 $\text{SNR} = -10 \text{ dB}$ 时,分别采用两种不同算法计算互相关,结果如图6所示.

可以发现,信噪比较低时,PHAT加权计算

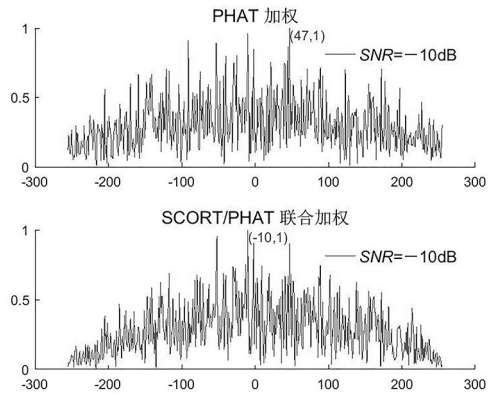


图6 采用两用不同加权算法对混合信号求互相关的结果(SNR=-10 dB)

Fig. 6 Two different weighting algorithms used to cross-correlate the mixed signals (SNR=-10 dB)

的互功率谱毛刺更为严重,极值不明显.一方面,本文采用awgn函数产生的高斯平稳白噪声本身具有一定的相关性,与一般传统方法采用广义互相关求时延估计时要求噪声不相关相比,更具有一般性,如果利用协方差矩阵产生的不相关噪声对比效果会更加明显;另一方面也证明了改进的SCOT/PHAT联合加权算法在低信噪比环境中鲁棒性更强.

信噪比为-7 dB环境下取不同声源S坐标,分别采用传统PHAT加权跟改进的SCOT/PHAT联合加权算法进行声源定位,结果如表1所示

根据表1,利用平均误差 $\Delta = \frac{1}{n} \sum_{i=1}^n |x_i - x|$ 来

表1 采用两种不同算法对声源进行定位计算的结果(SNR=-7 dB)

Table 1 Two different algorithms used to locate the sound source (SNR=-7 dB)

声源位置(1,θ)	PHAT 加权(1,θ)	SCOT/PHAT 联合加权(1,θ)
(0.52 m, 25.03°)	(0.86 m, 26.22°)	(0.53 m, 24.52°)
(0.59 m, 48.22°)	(0.80 m, 48.59°)	(0.61 m, 48.95°)
(0.52 m, 64.67°)	(1.14 m, 63.47°)	(0.57 m, 65.82°)
(0.29 m, 63.41°)	(0.29 m, 63.08°)	(0.29 m, 63.37°)
(0.64 m, 82.13°)	(0.76 m, 81.34°)	(0.66 m, 81.95°)
(0.27 m, 45.00°)	(0.29 m, 44.87°)	(0.27 m, 45.01°)

进行数据统计,误差结果见表2,定位坐标见图7.

表2 采用两种不同算法定位的平均误差(SNR=-7 dB)
Table 2 Average error of positioning using two different algorithms (SNR=-7 dB)

	距离平均误差(m)	方位平均误差(°)
PHAT 加权	0.22	0.67
SCOT/PHAT 联合加权	0.02	0.44

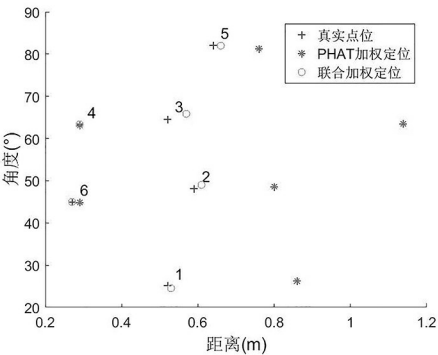


图7 SNR=-7 dB时两种算法的定位坐标图
Fig. 7 Two algorithms for locating coordinates when SNR=-7 dB

通过分析比较可以发现,近场模型小信噪比环境下,声源目标距离越近,定位精度越高;采用SCOT/PHAT联合加权算法要比传统的PHAT加权算法估计得到的声源位置信息误差更小,精度更高,因而有更强的抗噪能力,在声源定位系统中能更有效地估计声源位置.

4 结 论

广义互相关时延估计算法是基于麦克风阵列声源定位中较普遍、经典的算法,实用性较为广泛.但在低信噪比环境下,传统的广义互相关算法时延估计误差较大、精度较低,影响定位准确度.为进一步提高小信噪比环境下麦克风阵列的声源定位精度,本文提出一种改进的GCC麦克风阵列声源定位方法,采用SCOT/PHAT联合加权算法进行时延估计.仿真结果表明,同等环境条件下,SCOT/PHAT联合加权法优于传统PHAT算法,抗噪能力较强.但本声源定位系统也存在不足,比如没有充分考虑混响对定位精度的影响,将会在今后的研究中解决这类问题.

参考文献

- [1] Dai X M, Lou W Z, Liu P, et al. Speaker tracking based on microphone cross array in the smart conference system//2014 IEEE International Conference on Consumer Electronics, China. Shenzhen, China; IEEE, 2014: 1—4.
- [2] Liu G Q, Yuan S R, Wu J W, et al. A sound source localization method based on microphone array for mobile robot//2018 Chinese Automation Congress. Xi'an, China; IEEE, 2018: 1621—1625.
- [3] Mofeed M A E, Mofeed H A E. Direction-of-arrival methods (DOA) and time difference of arrival (TDOA) position location technique//Proceedings of the 22nd National Radio Science Conference. Cairo, Egypt; IEEE, 2005: 173—182.
- [4] Shi W L, Li Y S, Zhao L Y, et al. Controllable sparse antenna array for adaptive beamforming. IEEE Access, 2019, 7: 6412—6423.
- [5] Gierlich R. Joint estimation of spatial and motional radar target parameters by multidimensional spectral analysis//2015 The 16th International Radar Symposium. Dresden; IEEE, 2015: 95—101.
- [6] Imran M, Hussain A, Qazi N M, et al. A methodology for sound source localization and tracking: development of 3D microphone array for near - field and far - field applications//The 13th International Bhurban Conference on Applied Sciences and Technology. Islamabad, Pakistan; IEEE, 2016: 586—591.
- [7] Wang H S, Li J, Sun Z Q, et al. Accurate delay extraction for indoor pulse sound source location//The 12th International Conference on Signal Processing. Hangzhou, China; IEEE, 2014: 298—301.
- [8] Kwon B, Park Y, Park Y S, et al. Analysis of the GCC - PHAT technique for multiple sources//2010 International Conference on Control, Automation and Systems. Gyeonggi-do, South Korea; IEEE, 2010: 2070—2073.
- [9] Aung T T, Thumchirdchupong H, Tangsangiumvisai N, et al. Two-microphone subband noise reduction scheme with a new noise subtraction parameter for speech quality enhancement. IET Signal Processing, 2015, 9(2): 130—142.
- [10] Diaz - Guerra D, Beltran J R. Direction of arrival estimation with microphone arrays using SRP - PHAT and neural networks//2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop. Sheffield, UK; IEEE, 2018: 617—621.
- [11] Huang L X, Zhang S S, Wang J, et al. Classification of improved cross-correlation function to determine speaker location from microphone array//2019 IEEE 5th International Conference on Big Data Computing Service and Applications (BigDataService). Newark, CA, USA; IEEE, 2019: 209—214.
- [12] Ding J H, Ren B, Zheng N H. Microphone array acoustic source localization system based on deep learning//2018 The 11th International Symposium on Chinese Spoken Language Processing. Taipei City, Taiwan, China; IEEE, 2018: 409—413.
- [13] Zhang Q Q, Zhang L H. Study of delay estimation in acoustic source localization based on microphone array//2015 IEEE Advanced Information Technology, Electronic and Automation Control Conference. Chongqing, China; IEEE, 2015: 664—667.
- [14] Chetupalli S R, Ram A, Thippur V S, et al. Robust offline trained neural network for TDOA based sound source localization//The 24th National Conference on Communications. Hyderabad, India; IEEE, 2018: 1—5.
- [15] Chen J L, Wang J Z, Sun T Y. A multiplanes geometric approach for sound source localization with TDOA//2018 International Conference on System Science and Engineering. New Taipei, Taiwan, China; IEEE, 2018: 1—5.
- [16] Pérez M C, Gualda D, Villadagos J M, et al. Android application for indoor positioning of mobile devices using ultrasonic signals//2016 International Conference on Indoor Positioning and Indoor Navigation. Alcalá de Henares, Spain; IEEE, 2016: 1—7.

(责任编辑 杨可盛)