

基于耳蜗核模型改进双耳时间差的声源定位

张毅¹, 邢武超^{1*}, 罗元², 何春江¹

(1. 重庆邮电大学 信息无障碍工程研发中心, 重庆 400065; 2. 重庆邮电大学 光纤通信技术重点实验室, 重庆 400065)

(* 通信作者电子邮箱 xingwc1988@163.com)

摘要:人耳听觉系统在噪声环境中能够准确定位感兴趣的声源, 实现其定位的主要因素是双耳时间差, 但是在噪声环境下利用双耳时间差方法进行定位的效果比较差。针对这一问题, 提出一个基于耳蜗核模型的声源定位系统。利用耳蜗核模型模拟耳蜗核对听觉信息的处理机制, 提取听觉神经纤维中对语音刺激同步的信息和发放率信息, 从而实现对噪声的抑制, 完成噪声环境下的声源定位。该系统在噪声环境下定位的误差为 1.297°。实验结果证明改进之后的声源定位系统能在噪声环境下进行声源定位。

关键词:双耳时间差; 神经发放; 锁相; 耳蜗核模型; 声源定位

中图分类号: TP311 **文献标志码:** A

Sound localization based on improved interaural time difference of cochlear nucleus model

ZHANG Yi¹, XING Wuchao^{1*}, LUO Yuan², HE Chunjiang¹

(1. Engineering Research and Development Center of Information Accessibility,
Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2. Key Laboratory of Optical Fiber Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: The sound can be accurately located by human auditory system in noisy environment. The main element to realize the location is interaural time difference. But the effects are unsatisfactory when using interaural time difference to locate in noisy environment. In order to resolve this problem, this thesis put forward a sound source locating system based on cochlear nucleus model, and cochlear nucleus model simulated the process of how cochlear deals with auditory information. The process could draw the synchronization information and firing rate from the reaction of auditory nerve fibers to sound, thus realizing the inhibitory of noise, and locating the sound source in noisy environment. The location error of system in noisy environment was 1.297 degrees. The experimental results show that the improved sound locating system can complete locating in noisy environment.

Key words: interaural time difference; neural firing; phase locking; cochlear nucleus model; sound localization

0 引言

在现实环境中, 人耳听觉系统接收来自外界的声音信息, 并对其进行分析和整合, 从而实现对声音信号的感知。“鸡尾酒会效应”表明, 即使在嘈杂的背景环境中, 人与人之间的语言交流仍能够不受干扰噪声的影响而正常进行^[1]。和传统的麦克风阵列技术相比较, 人类的听觉系统在多个声源存在的情况下能精确地完成声源的定语与语音信号的识别^[2]。人耳听觉系统的这种机制, 吸引了更多的语音研究学者深入到人耳仿生的研究中。

在过去的几十年里, 研究证明人耳听觉系统主要利用声源到达双耳时存在时间上以及强度上的差值, 即双耳时间差 (Interaural Time Difference, ITD) 和双耳等级差 (Interaural Level Difference, ILD), 来实现声源方位的判定^[3-4]。由于语音信号的能量主要集中在低频段, 而噪声信号的能量主要集中在高频段^[5-6], 所以本文主要研究在低频段定位更加准确的基于双耳时间差的声源定位。

对于 ITD 的计算, Jeffress 提出了双耳互相关计算模型, 在其提出的模型基础上, 后续者相继提出了一些声源定位模

型^[7-8]。尽管在声源定位方面已经做了大量的研究, 但是在噪声环境下, 提取双耳时间差线索实现定位的准确性会有较大的下降^[9]。本文根据耳蜗核提取神经纤维的锁相信息, 发放率信息等机制的特点, 建立一个耳蜗核模型, 对传统的基于时间差的声源定位系统进行改进, 从而实现噪声环境下提取双耳时间差线索, 完成声源定位。

1 人耳听觉系统生理学机制

当外来声波到达人耳时, 首先通过听觉系统的外耳, 实现对声音信号的收集和放大。然后经由耳道传输至耳膜, 经过中耳的阻抗匹配后传导至内耳的耳蜗。耳蜗基底膜对声音信号具有频率选择特性, 即不同频率会在基底膜的不同位置产生不同幅度的调谐峰, 基底膜的这种频率选择特性相当于一个带通滤波的过程, 人耳听觉系统以此实现对声音信号的滤波^[10]。经过基底膜滤波之后的声波信号传递到内毛细胞。内毛细胞接受耳蜗传来的振动信号, 经过半波整流之后将其转换成神经电信号由听觉神经纤维实现对声音信号的脉冲编码^[11-12]。

神经纤维中输出的脉冲序列分别传输到耳蜗核的不同子

收稿日期: 2013-05-29; 修回日期: 2013-07-18。 基金项目: 国际科技合作计划项目(2010DFA12160)。

作者简介: 张毅(1966-), 男, 重庆人, 教授, 博士生导师, 博士, 主要研究领域: 智能机器人、生物信号处理、信息无障碍技术; 邢武超(1988-), 男, 山东临沂人, 硕士研究生, 主要研究方向: 语音识别、智能机器人; 罗元(1972-), 女, 贵州贵阳人, 教授, 博士, 主要研究方向: 信号与信息处理、数字图像处理; 何春江(1988-), 男, 重庆人, 硕士, 主要研究方向: 语音识别、智能机器人。

核中,其子核都保持着神经纤维的频率拓扑的特性。其前腹侧耳蜗核中的神经元能够保持神经纤维中的时间-位置编码,提取神经纤维中对声音信号刺激同步的信息;后腹侧耳蜗核中的神经元能够保存神经纤维中的发放率-位置编码,与声音信号刺激的强度成单调关系,提取的是发放率信息^[13]。

从腹侧核发出的纤维,与同侧和对侧的上橄榄复合体(Superior Olive Complex, SOC)以突触联系^[14-15]。当声音信号传入斜方体内侧核中的同时性检测神经元内,同时性检测神经元提取 ITD 信息,实现最终的声源定位。人耳听觉系统结构如图 1 所示。

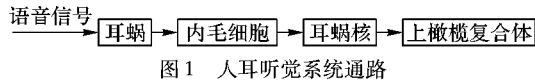


图 1 人耳听觉系统通路

传统的基于双耳时间差的声源定位系统主要模拟了耳蜗、内毛细胞以及上橄榄复合体的生理机制,并未涉及到对耳蜗核生理机制的模拟,其对噪声的抑制效果比较差,无法在噪声环境下实现准确定位。本文根据人耳听觉系统的生理学通路,在传统的定位系统中添加耳蜗核模型,利用耳蜗核模型模拟耳蜗核提取对语音刺激的同步信息以及发放率信息的生理机制,从而建立一个比较完整的声源定位系统。

2 基于耳蜗核模型改进的声源定位系统

2.1 耳蜗核生理机制

耳蜗核是听觉中枢系统的第一级核团,由三部分组成:前腹侧耳蜗核、后腹侧耳蜗核和背侧耳蜗核^[4]。其中前腹侧耳蜗核(Anterior Ventral Cochlear Nucleus, AVCN)中主要是类本原神经元,其提取的是神经纤维对声音刺激锁相的信息。耳蜗核中的后腹侧耳蜗核(Posterior Ventral Cochlear Nucleus, PVCN)中主要是振荡神经元(Chopper),其与声音刺激的强度呈单调关系,其提取的是神经纤维的发放率信息,即单位时间内神经纤维中电脉冲的发放数量。背侧核中主要是休止和累积反应类型神经元,其主要表现为非单调的发放率与强度的关系,与本文的研究无关,所以不再建立模型。前腹侧耳蜗核和后腹侧耳蜗核的模型如下。

2.1.1 前腹侧耳蜗核模型

AVCN 的锁相机制是两侧的神经元对同相波产生同步性的发放机制。锁相信息的提取可以用检测同步信息的广义同步检测(Generalized Synchrony Detector, GSD)算法来实现,称其结果称为同步谱。提取同步谱特征的计算公式如式(1)所示:

$$H_i = G \left(\frac{\int_{T_1}^{T_2} |p_i(t) + p_i(t - T_i)| dt - \sigma}{\int_{T_1}^{T_2} |p_i(t) - p_i(t - T_i)| dt} \right) \quad (1)$$

其中: H_i 为耳蜗基底膜滤波器组第*i*条通路的同步谱; $p_i(t)$ 为滤波器组第*i*条通路的听觉神经纤维的发放率; $T_i = f_s/f_i$, f_i 为滤波器组第*i*条通路的中心频率, f_s 为采样频率, T_i 即滤波器第*i*条通路的延迟周期; $p_i(t - T_i)$ 为异侧滤波器组第*i*条通路的听觉神经纤维的发放率; σ 为阈值偏差,为常数; $G(\cdot)$ 为一个饱和非线性函数。神经脉冲信号的同步检测如图 2 所示。

然后提取同步谱信息的同步指数,同步谱指数计算公式如式(2)所示:

$$S_i(t) = \int_{\tau=t-\delta}^{\tau=t} S_i(\tau) d\tau \quad (2)$$

其中: $S_i(\tau)$ 为同步谱第*i*条通路的第 τ 时刻值, δ 为积分时间间隔。

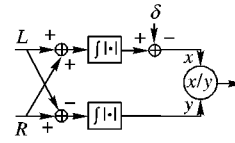


图 2 同步检测流程

同步谱指数反映的是声音信号的特征频率与滤波器组通道的特征频率接近的程度。当滤波器频率通道的中心频率越接近某个语音的频率时,获取的同步谱指数越大。如果声音信号的特征频率与通道的特征频率相近,则利用后腹侧耳蜗核模型提取该通道的发放率信息,进行进一步的处理;反之则舍弃该频率通道的发放率信息。

2.1.2 后腹侧耳蜗核模型

利用后腹侧耳蜗核模型提取同步谱指数较大的频率通道的发放率信息,然后进行进一步处理。后腹侧耳蜗核模型的数学表达式如式(3)所示:

$$\bar{p} = \int_0^T p(t) dt / T \quad (3)$$

其中: \bar{p} 表示时间*T*内的平均发放率, $p(t)$ 表示*t*时刻的神经发放率。

将经过耳蜗核模型处理的结果的幅值作为声音信号的能量,将信号附近数据的平均值作为信号背景噪声的能量,计算信号的能量比。能量比公式如式(4)所示:

$$SNR = p(t) / (\bar{p} + p(t)) \quad (4)$$

其中: $p(t)$ 为*t*时刻神经发放序列中输入信号对应的幅值,作为声音信号的能量; \bar{p} 为耳蜗核模型提取的平均发放率的幅值,作为信号背景噪声的能量。

如果 $SNR \geq 0.5$ 表明声音信号占主导地位,则保留这个声音占主导地位的信号片段;相反,如果 $SNR < 0.5$ 表明噪声能量占主导地位,则舍去。保留的声音信号最终在重合神经元内完成定位。

2.2 基于耳蜗核模型改进的声源定位系统

根据人耳听觉系统通路,利用耳蜗核模型改进传统的基于双耳时间差的声源定位模型,其整体流程如图 3 所示。

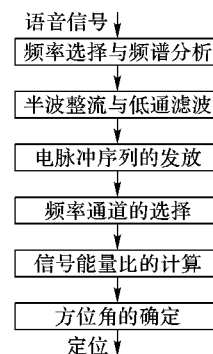


图 3 基于耳蜗核模型改进双耳时间差的声源定位系统

根据生理学实验证明,耳蜗基底膜对传入的声音信号具有频率选择特性和频谱分析特性。本系统中基底膜的频率选择特性和频谱分析特性采用伽马通(Gammatone, GT)滤波器组来模拟。滤波器组将语音信号按照频率的不同,在不同的频率通道进行传递。已知 Gammatone 滤波器的冲激响应函数如式(5)所示:

$$g_i(t) = t^{n-1} e^{-2\pi b t} \cos(2\pi f_i t + \varphi_i) u(t) \quad (5)$$

其中: $t \geq 0, 1 \leq i \leq N; \varphi_i$ 表示滤波器的初始相位; f_i 是滤波器的中心频率, 即基底膜的特征频率; n 为滤波器的阶数; b_i 是滤波器的衰减因子。

耳蜗内, 由于语音信号在基底膜上造成振动, 从而引起内毛细胞膜渗透性的变化, 促进神经递质的流动。该过程实现对语音信号的半波整流及低通滤波。内毛细胞的特性采用普遍应用的 Meddis 模型来模拟。Meddis 模型中描述递质变化过程可由式(6)~(8)所示:

$$\frac{dq(t)}{dt} = \gamma(1 - q(t)) - k(t)q(t) + xw(t) \quad (6)$$

$$\frac{dc(t)}{dt} = k(t)q(t) - lc(t) - rc(t) \quad (7)$$

$$\frac{dw(t)}{dt} = rc(t) - xw(t) \quad (8)$$

其中: $q(t)$ 为内毛细胞内的神经递质的数量; $k(t)$ 为内毛细胞膜渗透率; $c(t)$ 为突触间歇中神经递质的数量; $w(t)$ 为再生库内的神经递质的数量; x, l, r 均为常数参数。

随着内毛细胞膜通透性的变化, 递质分子的流动, 造成其膜电位的变化, 当内毛细胞膜电位达到阈值时, 就会造成电脉冲的发放, 从而形成电脉冲序列。电脉冲序列的发放采用 LIF (Leaky Integrate-and-Fire) 模型来实现。神经递质数量发生变化时会引起的突触后形成电流信号。当电流 $I(t) > 0$ 时, 膜电容 C 就会充电, 同时膜电容 C 和膜电阻 R 组成回路放电; 当膜电容 C 两端电压达到阈值电压 U_r 时, 就会产生一次电脉冲发放, 从而形成电脉冲序列, 其数学表示公式如式(9)所示:

$$V(t) = U_r \exp\left[-\frac{t-t^{(j)}}{\tau_m}\right] + \frac{1}{C} \int_0^{t-t^{(j)}} \exp\left[-s/\tau_m\right] I(t-s) ds \quad (9)$$

其中: U_r 为初始膜电位, τ_m 为常量。

频率通道的选择, 以及信号能量比的计算采用本文提出的耳蜗核模型来实现。频率通道的选用前腹侧耳蜗核模型, 信号能量比的计算选用后腹侧耳蜗核模型。

重合神经元模型的 ITD 值的计算利用互相关方法来实现。对于两路信号的互相关可用式(10)计算:

$$C(i, j, \tau) = \sum_{k=0}^{i-1} p_l(i, t-k) p_r(i, t-k-\tau) w(k) \quad (10)$$

其中: $p_l(i, t)$ 和 $p_r(i, t)$ 分别为左、右耳接收的信号的能量信息; i 为滤波器的频率通道; t 为时间点; τ 为时间延迟; $w(\cdot)$ 为矩形窗。

最后利用时间-角度的经验公式进行计算方位角。该计算公式如式(11)所示:

$$\theta(ITD) = \arcsin((ITD + 0.03)/0.684)/0.9155 * 180/\pi \quad (11)$$

其中: $\theta(ITD)$ 为声源方位角; ITD 为时间差。

通过互相关计算, 求取互相关最大值所对应时间差, 该时间差经过计算得到的方位角即为所求的方位角。

3 实验结果与分析

获取不同空间位置(不同水平角)的声音信号有两种方法: 一种是利用两个模拟人脑位置的麦克风设备进行双通道同时采样; 另外一种是将语音信号经过头部相关传递函数(Head Related Transfer Function, HRTF)冲击系数进行卷积,

以获取不同空间位置的模拟输入。由于双耳时间差的数量级为 $10 \mu s$, 对设备精度等的要求很高, 为了避免测量系统引起的误差, 本文将采用第二种方法模拟不同空间方位输入左右耳的语音信号。

实验所用的纯净语音信号为用 Cooledit 软件录制的语音, 噪声选取随机噪声, 实验在信噪比为 10 dB 的情况下进行。将纯净语音信号与噪声分别与 HRTF 冲击系数进行卷积, 获得实验中所需要方位角的左右耳语音信号。

将方位角为 40° 的纯净语音信号与方位角为 20° 随机噪声混叠, 将频率 200 Hz ~ 2300 Hz 分为 22 个频率通道, 分别对这 22 个频率通道的语音信号求取方位角。各滤波器频率通道对应的方位角如图 4 所示。

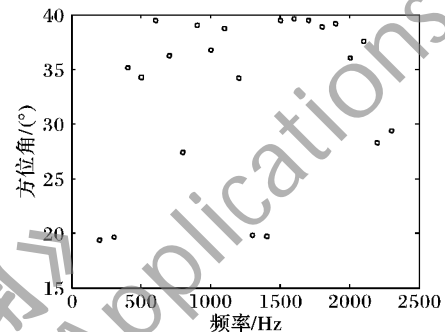


图 4 频率-方位角

利用前腹侧耳蜗核模型提取这 22 个频率通道的同步谱得到的结果如图 5 所示。

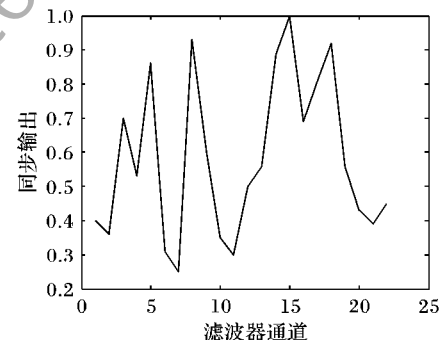


图 5 多通道的同步谱的输出

由图 4~5 可知, 利用同步谱指数较大的频率通道对应的方位角更接近纯净语音信号的方位角。由此证明利用前腹侧耳蜗核模型选择频率通道可有效地降低噪声信号的影响。

通过设定阈值, 将同步谱指数超过设定的阈值的频率通道所对应的语音信号提取出来, 舍弃同步指数未超过设定阈值的频率通道对应的语音信号, 然后对提取的语音信号进行能量比的计算。将最终保留的语音信号进行方位角的求取。对其进行求平均值计算, 所得结果即为测得声源的方位角。将本文提出的系统与传统的基于双耳时间差的声源定位系统的对比, 随即选取 5 组不同方位角进行实验, 对比结果如表 1~2 所示。

表 1 表示用传统的基于双耳时间差在噪声环境下的声源定位; 表 2 基于耳蜗核模型改进双耳时间差在噪声环境下的声源定位。由表 1 可知, 传统的基于双耳时间差在噪声环境下定位的误差均值为 6.498° ; 由表 2 可知, 基于耳蜗核模型改进双耳时间差在噪声环境下的定位的误差均值为 1.297° 。通过对比可知, 基于耳蜗核模型改进双耳时间差在噪声环境下的定位精度比传统的基于双耳时间差在噪声环境下定位精度提高了 5.201° 。

表 1 基于双耳时间差的方位估计

实验次数	ITD 值/ μs	估计方位/ $(^\circ)$	实际方位/ $(^\circ)$	噪声方位/ $(^\circ)$	方位误差/ $(^\circ)$
1	315.50	33.090	40	20	6.910
2	92.60	11.280	20	0	8.720
3	-85.99	-5.129	-10	10	4.871
4	-174.20	-13.290	-20	0	6.710
5	-390.30	-34.720	-40	-20	5.280

表 2 基于改进的双耳时间差的方位估计

实验次数	ITD 值/ μs	估计方位/ $(^\circ)$	实际方位/ $(^\circ)$	噪声方位/ $(^\circ)$	方位误差/ $(^\circ)$
1	372.6	39.390	40	20	0.610
2	172.4	18.810	20	0	1.190
3	-130.1	-9.191	-10	10	0.809
4	-213.9	-17.030	-20	0	2.970
5	-429.9	-39.090	-40	-20	0.910

由此证明,经过耳蜗核模型改进的声源定位系统比传统的声源定位系统在噪声环境下具有更好的鲁棒性,在噪声环境下实现更精确的定位。

4 结语

由实验可知,本文提出的利用耳蜗核模型改进传统的双耳时间差声源定位系统具有较好的鲁棒性,在噪声环境下具有较好的定位精度,有助于语音定位在噪声环境下的应用。

随着对听觉系统研究的不断深入,人们对中枢神经系统的复杂机制将有更加深入的了解。利用细胞模型模拟这些机制,建立一个更加完整的听觉定位系统,实现在多声源环境下的精确定位,有助于噪声环境下的语音系统识别率的提高,战场目标的精确锁定等。

参考文献:

- [1] LUO Y, TONG K G, ZHANG Y, *et al.* Sound source separation in a multi-voice environment based on auditory central nervous system [J]. *International Journal of Information Engineering*, 2012, 2(1): 47-53.
- [2] MAY T, van de PAR S, KOHLRAUSCH A. A probabilistic model for robust localization based on a binaural auditory front-end [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, 19(1): 1-13.
- [3] 张文娟. 基于听觉仿生的目标声音识别系统研究[D]. 长春: 中

国科学院研究生院长春光学精密机械与物理研究所, 2012.

- [4] LIU J D, HARRY E, STEFAN W. Mobile robot broadband sound localisation using a biologically inspired spiking neural network [C]// *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway: IEEE Press, 2008: 2191-2196.
- [5] 周成. 人耳声源定位与编码技术研究[D]. 武汉: 武汉大学, 2011.
- [6] 李冰, 夏秀渝, 申庆超, 等. 基于过零点双耳时间差的运动声源定位[J]. *计算机工程与应用*, 2012, 48(9): 127-130.
- [7] MAGNUS S, MOHAMMAD B, PRTER V. Improved binaural model for localization of multiple sources [C]// *Proceedings of 10th ITC Conference on Speech Communication*. Braunschweig: VDE Verlag, 2012: 26-28.
- [8] CECILIA M Z, RAFFAELE P, AURELIO U. Binaural sound source location in the presence of reverberation [C]// *Proceedings of the 17th International Conference on Digital Signal Processing*. Piscataway: IEEE Press, 2011: 1-6.
- [9] WOODRUFF J, WANG D L. Binaural localization of multiple sources in reverberant and noisy environments [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(5): 1503-1512.
- [10] 匡神兵. 听觉系统声源定位神经通路信息处理特性研究[D]. 武汉: 武汉理工大学, 2007.
- [11] 陈世雄, 官琴, 金慧君. 用 Gammatone 滤波器组仿真人耳基底膜的特性[J]. *清华大学学报: 自然科学版*, 2008, 48(6): 1044-1048.
- [12] 成彬彬, 张海. 听觉系统内毛细胞 Meddis 模型的分析与仿真 [C]// 2008 系统仿真技术及其应用学术会议论文集. 合肥: 中国科学技术大学出版社, 2008: 234-239.
- [13] 赵祥辉, 龙长才. 内毛细胞到听神经信号转换中的随机共振 [C]// 中国声学学会 2006 年全国声学学术会议论文集. 厦门: 中国声学学会, 2006: 405-406.
- [14] CHUN G D, CAUDELL T P. A model for auditory localization in robotic systems based on the neurobiology of the inferior colliculus and analysis of HRTF data [C]// *IJCNN'01: Proceedings of the 2001 International Joint Conference on Neural Networks*. Piscataway: IEEE Press, 2001: 1107-1111.
- [15] VOUSAS K, ADAMY J. A biologically inspired spiking neural network for sound source lateralization [J]. *IEEE Transactions on Neural Networks*, 2007, 18(6): 1785-1799.

(上接第 3231 页)

- [4] 杨柳, 吕英华. 基于云模型的网络风险评估技术研究[J]. *计算机仿真*, 2010, 27(10): 95-98.
- [5] GAMBETTA D. Can we trust trust?[M]// *Trust: making and breaking cooperative relations*. New York: Oxford Press, 1990: 213-237.
- [6] VIVEKANANTH P. Reputation based trust model for grid with enhanced reliability [C]// *Proceedings of the 2011 International Conference on Computers and Computing*. Stevens Point, Wisconsin: World Scientific and Engineering Academy and Society, 2011: 1-11.
- [7] JOSANG A, PRESTI S. Analyzing the relationship between risk and trust [C]// *iTrust04: Proceedings of the Second International Conference on Trust Management*. Berlin: Springer-Verlag, 2004: 279-311.
- [8] EHAB E. Probabilistic trust models in network security [D]. South-

ampton: University of Southampton, 2011.

- [9] 廖俊, 张宏, 蒋黎明, 等. 基于直觉模糊理论的 MANET 主观信任模型[J]. *南京航空航天大学学报*, 2011, 43(4): 538-543.
- [10] 王良民, 郭渊博, 詹永照. 容忍入侵的无线传感器网络模糊信任评估模型[J]. *通信学报*, 2010, 31(12): 37-44.
- [11] 王守信, 张莉, 李鹤松. 一种基于云模型的主观信任评价方法[J]. *软件学报*, 2010, 21(6): 1341-1352.
- [12] 蔡红云, 杜瑞忠, 田俊峰. 基于多维信任云的信任模型研究[J]. *计算机应用*, 2012, 32(1): 5-7.
- [13] 顾鑫, 徐正全, 刘进. 基于云理论的可信研究及展望[J]. *通信学报*, 2011, 32(7): 172-181.
- [14] 贺利坚, 黄厚宽. MAS 中信任和信誉系统的研究进展[J]. *计算机科学*, 2011, 38(4): 1-8.
- [15] 李德毅. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005: 171-177.