

麦克风阵列（II）：机器人听觉，——围观机器人的“耳朵”

（临菲信息技术港公众号图文）

机器人的“耳朵”是什么呢？麦克风，因为麦克风是将声音信号转换为电信号的器件。机器人“听到声音”，就是将声音转换成电信号，然后经过对电信号的处理，获得声音包含的信息。

“机器人要听声音，装一个麦克风就可以了”，是的，如果只是能“听得到”声音，确实一个麦克风就可以了。

但“听觉”的功能可不仅仅是简单的“听得到”！人耳“听到声音”至少能获得如下一些信息：

声音的类型——何种声源，人类话音？音乐？噪声？鸟鸣？虎啸？

声音的内容——声源表达的意思，语言内容？尖叫？叹息？甚至从发动机的异常噪声听出发动机出了何种故障。

声音的来向——声源在什么方向？前、后、左、右？天上地下？

声音的远近——声源的距离？身旁？附近？远处？

机器人听力系统要完全达到人耳的能力，是比较困难的。当前，机器人听力系统的研究开发，最基本的两个方面，一是自动语音识别（ASR），二是声源定位（实际上大多数情况下都只是分辨出声音的来向，即定向）。

“定位（或定向）”的本质是从所获得的信号中提取目标的位置（或方向）信息，因此，采集的信号必须包含目标的空间信息，这就需要2个以上配置在空间不同位置的传感器构成阵列。对于声源定位，该传感器阵列便是麦克风阵列。

应用于机器人听觉的麦克风阵，有两种模式：

“双耳”模式：即“双耳系统”（binaural system），模仿人类的听觉系统。

“多耳”模式：尽可能使用更多的麦克风构成阵列。

“双耳”机器人

“双耳（Binaural）”：这对“双朵”其实就是两个麦克风，在机器人头部左右两边各一个，和人的两只耳朵的位置差不多，让机器人看起来很像人的样子。

“双耳”由于只有两个麦克风，无论是信号接收还是声源定位，都有很大的挑战，尤其是声源定位，当同时存在多个声源时更加困难。

SIG

称为 SIG 的人形机器人“出生”很早，2000 年日本科技公司（Japan Science and Technology Corp）就在 IEEE/RSJ 的国际会议上介绍了 SIG 的设计和结构，该公司与日本京都大学在 2001 年和 2003 年的论文中，介绍了 SIG 实时跟踪说话人的技术和试验结果。

SIG 的听觉系统采用两个麦克风作为两个“耳朵”，可以实现语音的定位和分离。麦克风安装在耳廓内，但它们有小孔可以捕捉外面的声音。

SIG-2 在三个人同时讲话时它能够辨别出各个人的声音，眼睛可以注视声音的来向。



(a) SIG (a) 外壳 (b) 麦克风

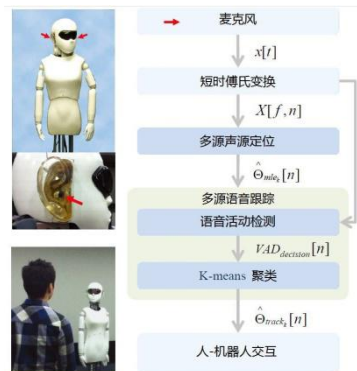
原图来自: Kazuhiro Nakadai, et al. Real-Time Multiple Speaker Tracking by Multi-Modal Integration for Mobile Robots, Eurospeech 2001 - Scandinavia

2011 年和 2013 年，报道了 SIG-2 听觉系统的改进，提升了多源定位和跟踪能力，在真实环境中实时跟踪多个声源的误差低于 4.35° 。



SIG 在听觉实验中

原图来自: Ui-Hyun Kim, et al., Improvement of Speaker Localization by Considering Multipath Interference of Sound Wave for Binaural Robot Audition, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2011



多源定位和跟踪流程图

原图来源: Ui-Hyun Kim and Hiroshi G. Okuno. Robust localization and tracking of multiple speakers in real environments for binaural robot audition, 2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)

M3-neony

M3-neony 是日本研发的一个婴儿型机器人，身高约 50 厘米，体重约 3.5 千克，和新生儿体形基本相同。

M3-neony 是双耳机器人，左右两个麦克风相距大约 0.1m。



M3-neony

原图来源: 网络



M3-neony 头部

原图来自: Ryuichi Shimoyama, et al., Room Volume Estimation Based on Statistical Properties of Binaural Signals Using Humanoid Robot, 23rd International Conference on Robotics in Alpe-Adria-Danube Region September 3-5, 2014

iCub

iCub 是意大利 IIT (Istituto Italiano di Tecnologia) 开发的人形机器人，是欧盟项目 RobotCub 的一部分，随后被全球 20 多个实验室采用。

iCub 也是双耳，耳朵由垂直于矢状面的麦克风组成，并被耳廓包围。



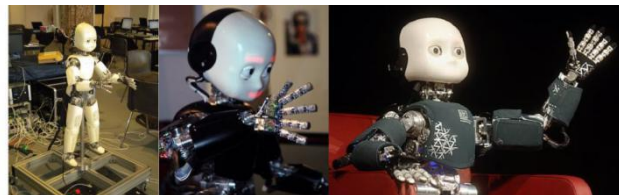
左：视听虚拟现实装置中的机器人 iCub；右上图：机器人 iCub 头部；右边下图：机器人 Sundman 头部

原图来自：Jorge Dávila-Chacón, et al., Enhanced Robot Speech Recognition Using Biomimetic Binaural Sound Source Localization, IEEE Transactions on Neural Networks and Learning Systems. (accepted for inclusion in a future issue of this journal)



iCub

原图来自：<http://www.icub.org/>



iCub

原图来自：<https://en.wikipedia.org/wiki/ICub>



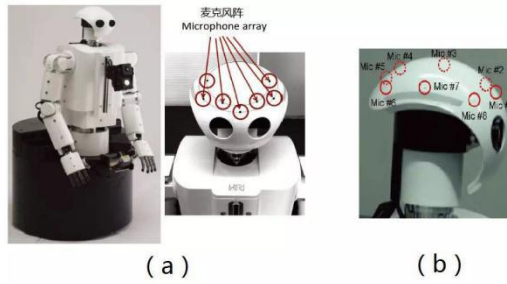
iCub

原图来自：Giorgio Metta, et al., The iCub humanoid robot: An open-systems platform for research in cognitive development, Neural Networks, 2010.

“多耳”机器人

Hearbo

HEARBO (HEAR-ing roB0t) 是日本本田研究所 (HRI-JP) 开发的机器人，头部麦克风阵列由 8 个麦克风构成。



Hearbo 机器人及头部麦克风配置

原图(a)来自: Randy Gomez, et al., Compensating Changes in Speaker Position for Improved Voice-based Human-Robot Communication, 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids);

原图 (b) 来自: Kuniaki Noda, et al., Sound Source Separation for Robot Audition using Deep Learning, 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)

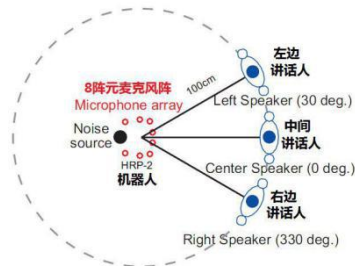
HRP-2

日本的 HRP-2 仿人机器人，其“耳朵”也是 8 个阵元的麦克风阵列。

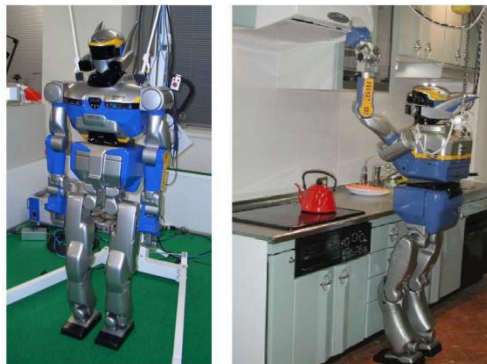


HRP-2 半身和头部

原图来自: Toru Takahashi, et al., Improvement in Listening Capability for Humanoid Robot HRP-2, IEEE International Conference on Robotics and Automation, 2010



HRP-2 分辨 3 个说话人示意图 原图来源: 同上图



HRP-2 机器人 图片来自: 网络

Nao

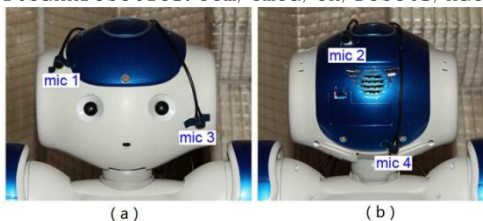
机器人 Nao 原来是由总部位于巴黎的法国机器人公司 Aldebaran Robotics 开发的自主可编程人形机器人。Aldebaran Robotics 于 2015 年被软银收购，改名为 SoftBank Robotics（软银机器人）。Nao 经过了多个版本，不断发展。

Nao 的耳朵是 4 个麦克风，前后各 2 个。



Nao 的技术特征

原图来自：<https://www.softbankrobotics.com/emea/en/robots/nao/find-out-more-about-nao>



(a) (b)

Nao 的头部

原图来自：Vladimir Tourbabin, et al., Direction of Arrival Estimation Using Microphone Array Processing for Moving Humanoid Robots, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 23, No. 11, November 2015



机器人 Nao

原图来自：<https://www.softbankrobotics.com/emea/en/robots/nao>

在 2017 年发表的论文中，Nao 的头部麦克风扩展到 12 个，同时，还通过在身体和肢体上安装麦克风来进一步扩展（称为“robomorphic array”），实现更大的阵列孔径。安装在机器人肢体上的麦克风阵，其阵列孔径甚至可以通过机器人的运动来改变。



(a) (b)

Nao 的“耳朵”扩展 (a) 具有 12 个麦克风的原型头部设计 (b) 身体和肢体麦克风阵

图片来自: Heinrich W. Lollmann1, et al., Microphone Array Signal Processing for Robot Audition, 2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)

Pepper

Pepper 也是日本软银 (SoftBank Robotics) 研发的人形机器人。Pepper 具有感知情感的能力, 能够识别对话者的情绪, 并根据其情绪调整其行为。Pepper 的耳朵是 4 个麦克风, 安装在头顶部。



Pepper

原图来自: <https://www.softbankrobotics.com/emea/en/robots/pepper>



Pepper

原图来自: <https://www.softbankrobotics.com/emea/en/robots/pepper>

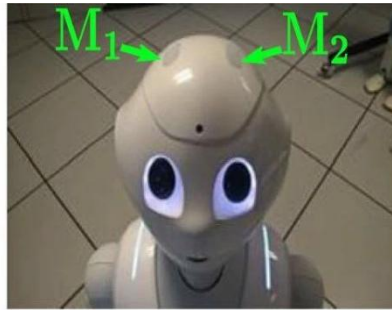
根据您的声音、面部表情、身体动作以及您使用的单词, Pepper 能理解您的情绪, 并能针对您的情绪, 通过他的眼睛、语气和胸前的平板做出他自己的回应。



Pepper

原图来自: <https://www.softbankrobotics.com/emea/en/robots/pepper>

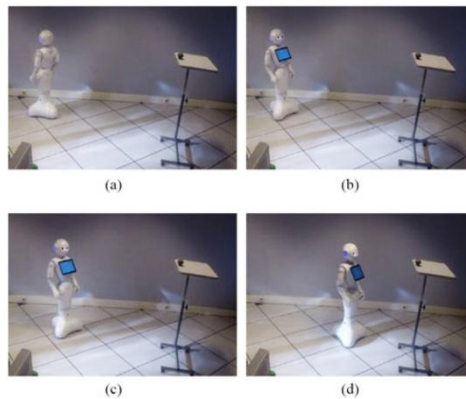
不过, 法国雷恩大学博士生 Aly Magassouba 最近在关于机器上室内运动控制的研究中 (A. Magassouba, “Aural servo: Towards an alternative approach to sound localization for robot motion control,” Ph.D. thesis, Universit´e Rennes 1, Rennes, France, 2016.), 只使用了 Pepper 的两个麦克风。这相当于把 Pepper 考虑为双麦克风机器人。两个麦克风在 Pepper 的头顶, 相距约 8 cm。



Pepper 的耳朵长到头顶上去了

原图来自: Aly Magassouba, et al., Exploiting the Distance Information of the Interaural Level Difference for Binaural Robot Motion Control, IEEE Robotics and Automation Letters, Vol. 3, No. 3, July 2018

在 Aly Magassouba 实验中, 双耳的 Pepper 能够解决声音来向的前面和背后模糊问题, 当它听到背后的声音后, 能够转身向声源走去。



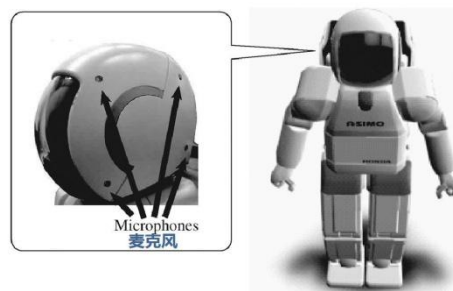
Pepper 听到声音在背后, 转身向声源走去

原图来自: Aly Magassouba, et al., Exploiting the Distance Information of the Interaural Level Difference for Binaural Robot Motion Control, IEEE Robotics and Automation Letters, Vol. 3, No. 3, July 2018

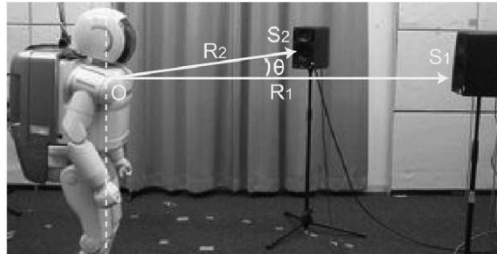
ASIMO

日本本田的机器人 ASIMO, 是世界上最早真正意义上的双足行走, 行走速度可达每小时 2.7km。ASIMO 的两只“耳朵”共有 8 个麦克风, 左右两侧分别是 4 个阵元的麦克风阵。ASIMO 可以听取并理解三个人同时讲话。

(ASIMO 经过了 7 个版本, 2018 年有报道称本田已停止 ASIMO 的继续开发)。



ASIMO 头部麦克风配置, 左右两侧各 4 个麦克风



ASIMO 声源定位试验

原图来源: Hirofumi Nakajima, et al., Blind Source Separation With Parameter-Free Adaptive Step-Size Method for Robot Audition, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, No. 6, August 2010



ASIMO 从行人间穿行

原图来自: <http://asimo.honda.com/gallery.aspx>

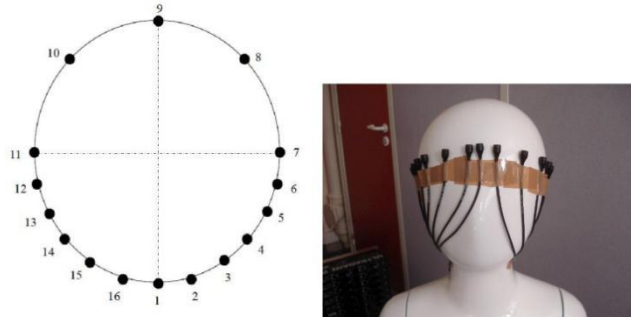


ASIMO 服务

原图来源: 同上图

其它

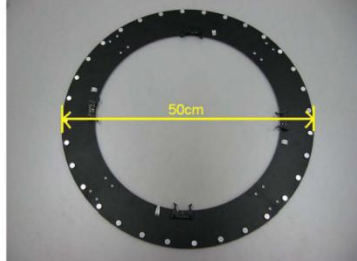
16 个麦克风阵元的机器人听觉研究，波束形成技术。



16 个麦克风阵元的机器人听觉研究

图片来源: Mounira Maazaoui, et al., Adaptive blind source separation with HRTFs beamforming preprocessing, IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012

32 个麦克风阵元的机器人听觉研究



32 阵元麦克风阵

原图来自: Yuki Tamai et al., Circular microphone array for robot's audition, IEEE International Conference on Sensors

(本文原载: 微信公众号“临菲信息技术港”)

