

Audio Location: Accurate Low-Cost Location Sensing

James Scott¹ and Boris Dragovic^{1,2}

¹Intel Research Cambridge, ²University of Cambridge
james.w.scott@intel.com, boris.dragovic@cl.cam.ac.uk

ABSTRACT

“Audio location” is a technique for accurate 3D location sensing using off-the-shelf audio hardware. The use of such hardware allows an audio location deployment to be low-cost and simple, as compared to other centimetre-scale location sensing systems. Another advantage of audio location is the ability to locate sounds that the user can make themselves, e.g. finger clicking, rather than requiring them to carry a locatable electronic “tag.”

1. INTRODUCTION

Fine-grained (i.e. sub-metre-scale) location sensing systems facilitate many ubiquitous computing applications, including follow-me applications (e.g. your music follows you around), nearest-X applications (e.g. nearest telephone rings when you are called), and novel user interfaces.

The motivation for this work is the development of fine-grained location sensing methods based on the use of commodity, low-cost hardware, and/or hardware which is already ubiquitously deployed. Using such hardware would facilitate the deployment of fine-grained location sensing by ordinary users in a wide range of environments. The technique proposed to accomplish this, “audio location,” relies on cheap off-the-shelf audio hardware.

2. RELATED WORK

A number of fine-grained location sensing systems for indoor environments have been developed, including the ultrasonic Bat [1] and Cricket [8] systems, with accuracies of around 3 cm. Vision-based systems have also been built, e.g. the tag-based TRIP system [5] and the Easyliving system [4]. Ultra wideband-based systems have been shown to achieve an accuracy of around 15 cm [10]. However, all of these rely on the deployment and configuration of significant infrastructure, and many of the above systems require the users to carry electronic tags.

Fine-grained location sensing using off-the-shelf audio hardware has been examined by a few research groups. Girod et al. have developed a system for fine-grained range estimation [2], making use of tags to produce sounds; untagged location is not discussed. The work of Rosca et al. does consider untagged operation [9]. However, their discussion is purely theoretical and does not address the difficulties of implementation and evaluation in real use. Finally, the use of commodity audio hardware for context-aware applications, including coarse-grained location sensing, was recently presented by Madhavapeddy, Scott and Sharp [6].

3. DESIGN PARAMETERS

This section will explore the design parameters of audio location systems. In particular it looks at tagged vs. untagged design, the signal detection method used, and the location calculation technique. The overall system architecture is shown in Figure 1.

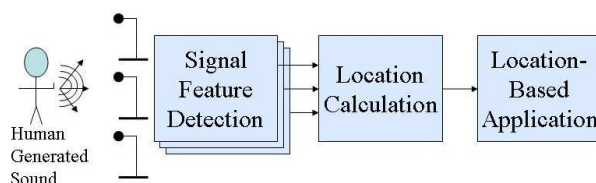


Figure 1: Audio location system architecture.

3.1 Tagged Versus Untagged

Fine-grained location systems often require that the objects to be located be “tagged” with small devices which interact with the sensing infrastructure. Two examples of tags for humans are ultrasonic transmitters hung around the neck of a user, and visual barcode tags attached to users’ clothes.

It is also possible to construct “untagged” systems, in which the user or device is located purely by means of their intrinsic properties or capabilities. Not requiring tags has a number of advantages: there may be less cost due to hardware, users of the system may find it more convenient as they do not have to remember to wear tags, and new users may be automatically discovered without having to be tagged.

An untagged system will have to cope with two performance-degrading factors, namely the difficulty of detecting a suitable audio feature at each microphone, and coping with a lack of synchronisation as the time-of-send of the audio signal is not known. The Dolphin ultrasonic location system [3] is one example of a sound-based system which accomplishes unsynchronised fine-grained location.

3.2 Sound Feature Detection

The first step in untagged audio location is to identify a sound’s manifestation across multiple data streams, so that a time-of-arrival can be calculated for each microphone. At a sampling rate of 48kHz (the maximum sampling rate supported by most sound cards), each sample represents approximately 0.5 cm; this illustrates how the feature detection accuracy affects the location accuracy.

The feature detection method used in the prototype is simple:

a signal is detected if the amplitude of a sample is greater than a dynamic threshold based on the recent background noise. This method copes well with detecting relatively loud, sharp sounds, which includes humans clicking their fingers, clapping their hands, and so on. However, it is unable to cope with sounds which are not much louder than the background noise, e.g. a person speaking while a stereo is playing.

3.3 Location Determination

In calculating the 3D location, at least four microphones are required, since there are four variables to be determined: the x, y, z coordinates of the sound source and the time-of-send of the sound. Using only four microphones would give no indication of the location accuracy. With five or more microphones, the location error can be estimated, thus allowing the rejection of received location events with inconsistent times-of-arrival, or the identification of one or more incorrect times-of-arrival while still allowing location calculation using the remaining times-of-arrival.

The Levenberg-Marquardt method [7] is used to calculate the location by solving a non-linear set of equations concerning the time-of-send, times-of-arrival, microphone locations, and the location of the sound source. One assumption of this method is the prior determination of the accurate locations of the microphones in the environment.

In experiments conducted using six microphones, the accuracy of this system was approximately 15 cm in 3D at the 90% confidence level. This is accurate enough to support fine-grained location resolution at a human scale - e.g. splitting up the area reachable using one's arms into a large number of identifiable regions.

4. 3D INTERFACES USING SOUNDS

In order to test the feasibility of audio location, a prototype was constructed along with a demonstration application. The application chosen is that of 3D interfaces using human sounds. Similar interfaces have been the subject of work undertaken using the Bat system [1]. One advantage of using audio location rather than the ultrasonic Bat system is that no tag is required for each user. A disadvantage of audio location is that, by itself, it is not capable of identifying the user, while the Bat system does identify the tag being used.

3D interfaces can be constructed using "virtual buttons," i.e. regions of 3D space which, when a sound is made at that location, cause the audio location system to take some action. They can be defined manually by inputting coordinates, or automatically in one of two ways: by using a mouse to indicate position on a map representation of the space, or by simply making a few sounds at the desired location.

The types of application which can be developed in this way include any application which might otherwise involve a real button interface. In an audio location-instrumented home, this might include lights, entertainment equipment such as music players, VCRs, and so on. On a computer desk, audio location might be used to quick-start or control often-used computer applications. The first author is habitually using audio location to control mp3 player software at the work-

place, since he finds it more convenient to click one's fingers than to use the GUI interface.

5. CONCLUSIONS AND ONGOING WORK

Audio location has been shown as a viable technique for gaining accurate location information without requiring custom hardware and skilled installation and maintenance. The various aspects of an untagged 3D audio location system have been described, as well as the application area of 3D interfaces using human sounds.

The work presented here is a brief outline of one type of audio location system. We are also currently examining other system types, including the use of different signal detection methods, and different location determination techniques. We are also examining other system infrastructures, e.g. making use of microphones spread across multiple PCs.

Acknowledgements

The authors thank Richard Sharp, Anil Madhavapeddy and Mike Hazas for their valuable comments during this research.

REFERENCES

- [1] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward, and A. Hopper. Implementing a sentient computing system. *IEEE Computer*, 34(8):50–56, Aug 2001.
- [2] L. Girod, V. Bychkovskiy, J. Elson, and D. Estrin. Locating tiny sensors in time and space: A case study. In *Proc. ICCD 02*, September 2002.
- [3] M. Hazas and A. Ward. A high performance privacy-oriented location system. In *Proc. PerCom 2003*, pages 216–223.
- [4] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer. Multi-camera multi-person tracking for EasyLiving. In *Proc. 3rd IEEE Intl Workshop on Visual Surveillance*, 2000.
- [5] D. López de Ipiña, P. Mendonça, and A. Hopper. TRIP: A low-cost vision-based location system for ubiquitous computing. *Personal and Ubiquitous Computing*, 6(3):206–219, May 2002.
- [6] A. Madhavapeddy, D. Scott, and R. Sharp. Context-aware computing with sound. In *Proc. UbiComp 2003*, volume LNCS 2864, pages 315–332.
- [7] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1993.
- [8] N. Priyantha, A. Miu, H. Balakrishnan, and S. Teller. The Cricket Compass for context-aware mobile applications. In *Proc. MobiCom 2001*.
- [9] J. Rosca, S. Sudarsky, R. Balan, and D. Comanici. Mobile interaction with remote worlds: The acoustic periscope. In *Proc. AAAI 01*, 2001.
- [10] Ubisense Ltd. Ultra wideband location system. <http://www.ubisense.net/>.