

# Vocate: Auditory Interfaces for the LOK8 Project

John McGee, Dr. Charlie Cullen

Digital Media Centre, Dublin Institute of Technology  
Aungier St., Dublin 2, Ireland  
john.mcgee1@student.dit.ie, charlie.cullen@dmc.dit.ie

## Abstract

The auditory modality has a number of unique advantages over other modalities, such as a fast neural processing rate and focus-independence. As part of the LOK8 project's aim to develop location-based services, the Vocate module will be seeking to exploit these advantages to augment the overall usability of the LOK8 interface and also to deliver scalable content in scenarios where the user may be in transit or requires focus-independence. This paper discusses these advantages and outlines three possible approaches that the Vocate module may take within the LOK8 project: speech interfaces, auditory user interfaces, and sonification.

**Keywords:** Location-based Services, Auditory Interfaces, Auditory User Interfaces, Speech Interfaces, Sonification

## 1 Introduction to LOK8

The goal of the LOK8 (pronounced locate) project is to create a new and innovative approach to human-computer interactions. With LOK8 a person will be able to engage in meaningful interaction with a computer interface in a much more natural and intuitive way than we are used to. A virtual character will be displayed in numerous locations depending on the user's position and context. Users will be able to communicate with this virtual character through speech and gestural input/output, which will be processed and controlled by the dialog management component of the system. This will allow "face-to-face" interactions with the LOK8 system. The LOK8 system will deliver content to the user in a variety of context-specific ways with the aim of tailoring content to suit the user's needs. In addition to screens and projectors displaying the avatar, the user's mobile device, as well as speakers within the environment, will be used to deliver focus-independent content. Ultimately the goal is to replace a human-computer interface with a *human-virtual human interface* (see Figure 1).

## 2 The Vocate Module

The Vocate module will be in charge of the auditory aspect of the LOK8 environment and will be seeking to implement a number of features in this regard. As well as collaborating with the other LOK8 modules to develop realistic speech interaction with the LOK8 avatar, Vocate will be utilizing the contextual and spatial awareness of the user's mobile device within the environment to deliver hands-free audio navigation and browsing systems, as well as a hands-free auditory version of the main LOK8 menu interface. Vocate will be looking at three key areas in the field of auditory interfaces in its approach to the LOK8 project: speech interfaces, auditory user interfaces, and sonification. Further discussion on the issues outlined here, from the point of view of physicality, can be found in [1].

## 2.1 Related Work

There is existing empirical evidence in relation to audio spatialisation that is of particular interest to the work of the Vocate module [2][3][4]. Previous experiments in the field of audio navigation would include the SWAN navigation system [4] and Stahl's Roaring Navigator system [5], both of which make use of audio spatialisation to communicate information relating to a user's environment. In terms of speech interfaces, some existing off-the-shelf products now promise much of the functionality required to implement many of Vocate's aims for the LOK8 environment, such as Vlingo (available on Blackberry, iPhone, Nokia and Windows Mobile), Voice Control (Apple's new speech interface system for the iPhone 3GS), and Google Mobile App (available on the iPhone).

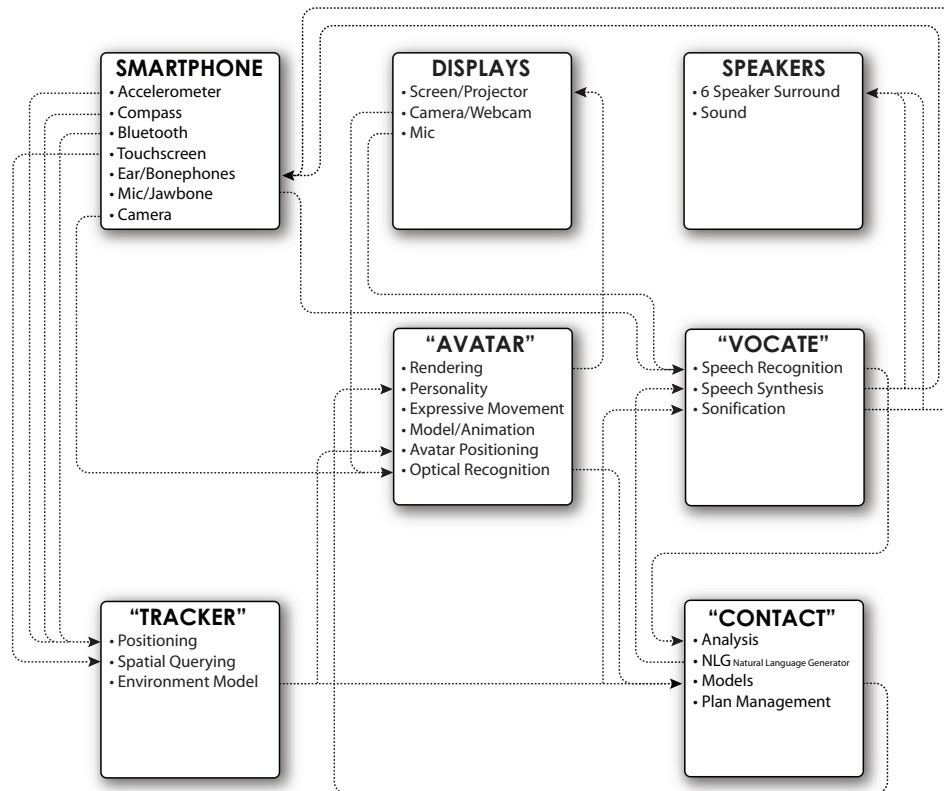


Figure 1: Overview diagram of the LOK8 project.

## 3 Advantages of Auditory Interfaces Over Other Modalities

Auditory interfaces offer a number of advantages over other modalities when it comes to the delivery of certain types of information. Audio information is processed faster neurally than both haptic and visual information (2ms for audio compared with 10ms and 100ms for haptic and visual respectively [6]), it is also hands-free and largely focus-independent. When taking technology and bandwidth limitations into consideration audio also has the advantage of lower overheads in terms of processing and storage when compared with many visual information delivery systems. These kinds of qualities offer up audio as a particularly useful modality for the communication of anything that urgently needs to be brought to the user's attention, anything that needs to be kept on the periphery of the user's attention (and/or requires a certain level of focus-independence), or anything operating within any system that has to take processing limitations into account. The Vocate module aims to exploit the auditory modality's unique strengths to both augment and enhance the overall realism and intuitiveness of the LOK8 system when it is at its most immersive (i.e., when the avatar is in use within the environment), and also to enable content delivery to be scalable (thus allowing the system to be user-friendly in situations where the user requires to be eyes/hands-free, when the user is in transit, or when screen size might be an issue).

## **4 Speech Interfaces**

Speech interfaces make use of speech recognition and/or speech synthesis to communicate with a user, they offer the user the ability to interact with a system using natural language and as such can be incredibly effective. Speech signal processing can be applied in a variety of ways to analyse the speech input of a user and hence model an appropriate response from the system. Many modern systems acknowledge the fact that speech communication is not solely an auditory interaction and make use of multimodal input to create more natural, realistic speech interface systems. In conjunction with auditory analysis, optical recognition techniques are used to capture additional input, such as eye tracking, lip tracking, and gestural tracking, to assist in the modeling of the system's responses. This multimodal approach can lead to the development of systems that exhibit attributes such as 'active listening' (a structured way of interacting whereby attention is focused on the speaker), 'turn taking' (the ability to know when to listen and when to interject in a conversation), and 'synchrony' (mirroring the intonation and/or body language of the user), thus leading to more natural human-computer interaction.

## **5 Auditory User Interfaces**

Auditory user interfaces are defined as the use of sound to communicate information about the state of an application or computing device to a user [7]. They are less constricted than speech interfaces or sonification alone as they often leverage strengths from both of these fields. Although audio is serial in nature and lacks the ability to continuously display items of interest in the way that the visual modality can, it does still possess qualities that lend it to the design of auditory menu systems. The human auditory system has the ability to filter out salient information from multiple streams of audio, this is known as the 'cocktail party effect' [8]. This ability, combined with techniques such as skimming (the presentation of segments of an audio stream to give an indication of the entire stream) and audio spatialisation, could be employed to design a speech-based auditory menu system that also uses earcons (defined by Blattner et al. as non-verbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation, or interaction [9]) and other auditory icons to reinforce metaphors and give enhanced feedback to the user.

## **6 Sonification**

Sonification is defined as the use of non-speech audio to convey information [10]. One form of information that lends itself particularly well to sonification is spatial information; this is because spatial information is generally physical in nature rather than abstract. With the help of contextual and spatial awareness within an environment, the stereo spatialisation and volume/tempo modulation of an audio source signal can allow the sound designer to 'place' auditory information within the soundscape as if it were coming from an actual physical location relative to the user. This technique can be used to convey a number of things including target destination sounds (these can be used to guide a user through an environment), object sounds (these can be used to highlight an object when it becomes contextually relevant to the user), and surface transition sounds (these can be used to allow a user to know when they have moved from one specific area or surface to another. Studies have found that broad spectrum sounds, such as pink noise bursts, are easier to localise and have been found to encourage better performance in audio navigation. It has also been found that when using a beacon style navigation approach a moderate capture radius for each beacon is preferable to a very large or very small capture radius, e.g. greater than 9ft or only a few inches [3][4].

## **7 Current Work**

All four LOK8 modules are currently collaborating to create a test environment in which to run Wizard of Oz experiments, the purpose being to gather data in relation to how users might react and respond to a prospective LOK8 interface. It is intended that this test environment will act as a first iteration

towards the final LOK8 environment. More detail regarding this Wizard of Oz environment can be found in [11]. Vocate will be looking specifically to test sonification techniques within a real-world 360° user environment. Although Vocate ultimately aims to provide audio navigation via headphones and/or bonephones, in the Wizard of Oz environment it will be using a six speaker array to run initial tests.

## 8 Conclusions

The LOK8 project seeks to develop location-based services across a variety of media within a user's environment. The Vocate module will be looking to exploit the unique advantages afforded by the auditory modality to enhance the usability and user-friendliness of the LOK8 system when it is at its most immersive and also to enable content delivery to be scalable in scenarios where the user may need a certain level of focus-independence or hands-free mobility. Speech interfaces are highly effective for complex and detailed interactions because they allow for the use of natural language but they require a lot of back-end work. Several commercial products are now emerging that feature speech interfaces that promise a lot of the functionality that the Vocate module is seeking to implement within the LOK8 environment and, as such, may provide a solution should they stand up to testing within the overall LOK8 environment. Sonification is particularly useful when it comes to communicating physical information or anything that has a natural acoustic mapping, it can also often transcend many of the linguistic and cultural boundaries that speech interfaces normally face. Auditory user interfaces leverage strengths from both speech interfaces and sonification and may be of particular interest regarding the aims of the Vocate project. There is existing empirical information regarding usability thresholds for both the processing of multiple streams of audio and for audio spatialisation (for both speech and non-speech sounds), Vocate will seek to build on this work to test interface designs that utilise both speech and non-speech sounds to interact with a user.

## Acknowledgements

This work is funded by the Higher Education Authority (HEA) in Ireland, Technological Sector Research Strand III: Core Research Strengths Enhancement Programme.

## References

- [1] McGee, J. and Cullen, C. (2009). Vocate: Auditory Interfaces for Location-based Services. In Proceedings of the Third International Workshop on Physicality (Cambridge, UK, 1 September, 2009). 25 - 29.
- [2] Walker, B. N., Raymond, S. M., Nandini, I., Simpson, B. D., and Brungart, D. S. (2005). Evaluation of Bone-Conduction Headsets for use in Multitalker Communication Environments. In Proceedings of the Human Factors And Ergonomics Society 49th Annual Meeting (Orlando, Florida, September 26 - 30, 2005). Human Factors and Ergonomics Society. HFES'05. 1615 - 1619.
- [3] Walker, B. N. and Lindsay, J. (2005). Navigation Performance in a Virtual Environment with Bonephones. In Proceedings of the 11th Meeting of the International Conference on Auditory Display (Limerick, Ireland, July 6 - 9, 2005). ICAD'05. 260 - 263.
- [4] Walker, B. N. and Lindsay, J. (2006). Navigation Performance with a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. *Human Factors*. 48, 2, (Summer 2006). Human Factors and Ergonomics Society. 265 - 278. 2005.
- [5] Stahl, C. (2007). The Roaring Navigator: A Group Guide for the Zoo with Shared Auditory Landmark Display. In Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services (Singapore, September 9 - 12, 2007). Mobile HCI'07. ACM. 383 - 386.
- [6] Kail, R. and Salthouse, T.A. (1994). Processing Speed as a Mental Capacity. *Acta Psychologica*. 86, 2 - 3 (June, 1994). 199 - 255.

- [7] McGookin, D. (2004). Understanding and Improving the Identification of Concurrently Presented Earcons. PhD thesis, University of Glasgow. 155 - 159.
- [8] Arons, B. (1992). A Review of the Cocktail Party Effect. *Journal of the American Voice I/O Society*.
- [9] Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M. (1989) Earcons and Icons: Their Structure and Common Design Principles. *Human Computer Interaction*, 4(1): 11 - 44.
- [10] Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, and Neuhoff, J. (1999). Sonification Report: Status of the Field and Research Agenda. Technical Report. ICAD, 1999.
- [11] Schütte, N., Kelleher, J., and MacNamee B. (2009). A Mobile Multimodal Dialogue System for Location Based Services. Awaiting publication in proceedings for IT&T 2009.