Mobile Spatial Audio Communication System

Alan Kan, Graeme Pope, Craig Jin, André van Schaik

Computing and Audio Research Laboratory
School of Electrical and Information Engineering
Building J03, Maze Crescent, University of Sydney, NSW 2006
akan@ee.usyd.edu.au, graemepope@student.usyd.edu.au,
craig@ee.usyd.edu.au, andre@ee.usyd.edu.au

ABSTRACT

As a first step in studying the application of the global positioning system (GPS) in a spatial audio communication system, an experimental platform was developed using commercial, off-the-shelf (COTS) GPS units: a Garmin Rino 120 (Rino) and a laptop. Software running on the laptop was developed to create a spatial-audio communication system based on GPS data and the two-way radio link on the Rino units. The audio communications were rendered spatially on headsets and perceptually tested for a number of users. The perceptual results and implementation issues are discussed in this paper. The experimental platform demonstrated that COTS GPS units can be used for a real-time GPS-based spatial audio communication system, but with some limitations.

1. INTRODUCTION

Current inter-personnel communication systems, which use headsets for playback of the audio communications, are unable to accurately present the direction of incoming sounds to the listener. For example, a sound originating from due north is not heard as coming from that direction, but as originating inside the listener's head. Spatial audio communication aims to preserve the spatial aspect of the sound and should find useful application in fire fighting, search-and-rescue, military reconnaissance, and audio teleconferencing. The problem with most current headsets is that the acoustic filtering effects of the outer ear and torso are lost because sounds are presented directly into the ear canal. The filtering of sounds by the outer ear and torso has been shown to be important in providing directional cues, as well as for the externalisation of sounds [1]. The loss of directional information reduces the listener's sense of situational awareness, and in an environment with multiple, simultaneous talkers, the ability to segregate and stream speech signals is drastically reduced.

Spatial audio signal processing techniques can be used to restore the directional cues arising from the outer ear and torso, so that the listener perceives sounds from the headset as originating from their true spatial directions [2]. Such simulations are referred to as virtual auditory space (VAS). VAS is achieved by filtering the sound stimulus with filter functions that reproduce the acoustic effects of the outer ear and torso for the desired direction. These filter functions are referred to as head-related transfer functions (HRTFs). They are recorded for a number of discrete directions around the head and can be interpolated for other directions ([3], [4], [5]).

A range of spatial audio communication systems have been presented in the literature (for example, [6], [7]), however, these systems tend to have talkers placed at arbitrary locations around the listener. Knowledge of the precise location of the talker relative to the listener, and the listener's orientation, can enable correct rendering of the talker's location in the listener's extra-personal space. This geometrical data was determined in this work using a GPS unit and laptop. The experimental platform described was used as an initial step in studying the application of GPS in a real-time spatial-audio communication system.

2. METHOD

2.1. Hardware

The experimental platform employed the use of two laptops and two Garmin Rino 120 GPS units (Rino units). The Rino units were chosen for its peer-to-peer feature, allowing the transmission of a user's GPS position to other Rino users on the Family Radio Service (FRS) bands. This update occurs once every thirty to forty seconds. The Rino was also chosen because it has an inbuilt two-way radio link on the FRS. These features are not available on all GPS units and were essential for the rapid development of an initial experimental platform.

The Rino units are connected to the laptops through the serial communications port using the National Marine Electronics Association (NMEA) 0183 v3.01 standard for communication between marine instruments. The standard specifies a 4800 baud serial link for sending a number of sentences containing information such as position and waypoints. This information is streamed out of the Rino units to the laptop and it takes approximately two seconds to send an entire set of sentences. More information can be found in the Garmin Rino 120 user manual [8].

The line-level audio output from the headset connector on the Rino was connected to the microphone audio input on the laptop. In this way, incoming voice communications over the two-way radio could be spatialised on the laptop. A high-quality stereo headset was used to present the final spatialised communications to the listener. The microphone on the Rino unit was used to capture the user's speech signal which was then transmitted over the Rino unit's radio link. The radio link is half-duplex and the speaker is required to press the "talk button" in order to broadcast the radio signal.

Figure 1 shows a picture of the mobile spatial audio communication system setup.

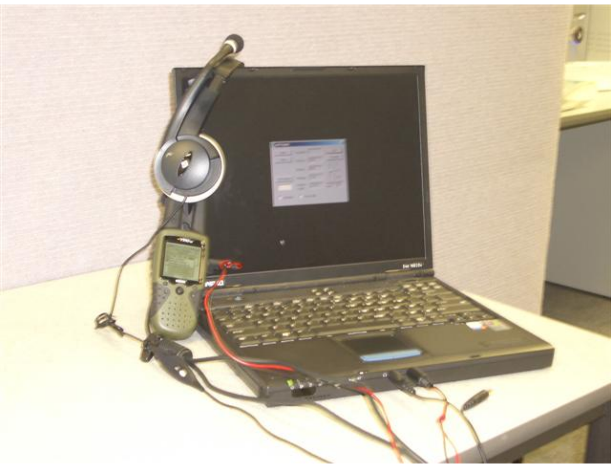

Figure 1. *Spatial-audio communication platform*

2.2. Software

Software was developed for the spatial-audio communication platform to process and extract position data from the GPS sentences that were streamed through the serial link between the Rino unit and the laptop. The GPS sentences do not directly provide data indicating the position of other Rino units but instead contain information such as the unit's own position and waypoint data. Thus, a method was required to obtain the location of other Rino users. This was achieved by setting another Rino user as a waypoint. Since waypoint information is provided within the GPS sentences, this can then be obtained via the Rino-laptop link. The bearing angle to the waypoint (the other Rino user) can then be extracted from the GPS sentences. Elevation differences between users are assumed to be zero since elevation data of the waypoints are not provided in the sentences streamed from the Rino units.

The listener's head-orientation information can only be inferred using the Rino units. When the listener moves, head-orientation can be calculated from the direction of motion assuming that the listener is facing in that direction. The software developed provides the ability to calculate this over the last one to twenty positions. In addition, the practical testing capability was added to the software to force the assumption that the listener is facing north.

Using the bearing angle and the estimated head-orientation of the listener, the relative direction between the listener and the speaker could be calculated. From this relative bearing, the corresponding directional HRTFs were generated to filter the voice communication signal from the Rino unit. The generation of HRTF filters occurs each time new data is received over the Rino-laptop link, which is approximately every two seconds.

For the field tests, the HRTF filters were based on a database of a 393 HRTFs that were recorded for one of the authors by the Auditory Neuroscience Laboratory (see [3]). The HRTF dataset was compressed using principle component analysis and a spherical thin-plate spline interpolation technique was used to interpolate HRTFs for non-recorded directions (see [5]). The HRTFs were incorporated into the platform as 512-tap, minimum phase, finite impulse response (FIR) filters.

In summary, the software processes the voice communication signal at the laptops microphone input, filters this signal with the HRTFs corresponding to the relative bearing between listener and talker, and presents the audio to the listener over a stereo headset. Figure 2 shows the graphical user interface (GUI) for the software.

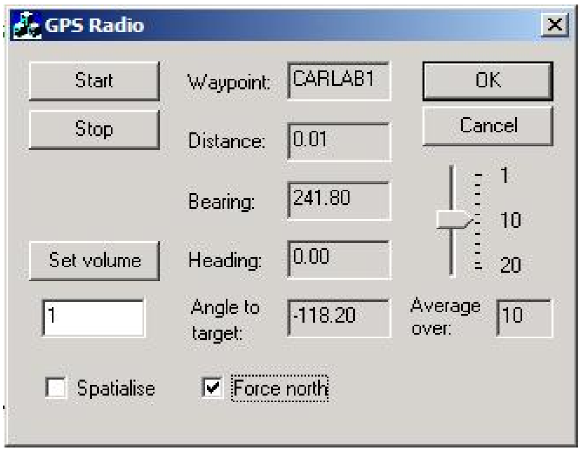

Figure 2. *Screen capture of software GUI*

2.3. Field Test

Field tests of the spatial-audio communication platform were conducted outdoors to examine various aspects such as accuracy, reliability and spatial fidelity. Tests were carried out on an open grass field and involved users communicating over the Rinos' radio link and listening over headsets connected to the laptop. Since the radio link is half-duplex, only one person could speak at any one time. In the field tests, one user was designated as the listener, while the other was designated as the speaker. In the first test, the listener remained stationary in the centre of the field, while the speaker moved roughly in a semi-circle from the stationary listener's forward far-left region to the listener's forward far-right region and vice versa. During this test, the 'force north' feature of the software was enabled on the listener's laptop and the listener was duly required to face north at all times. The process was then reversed with the speaker stationary and the listener moving. The 'force north' feature in the software was, now disabled on the listener's laptop in order to test the method for estimating the listener's head-orientation. A test was also conducted with both users moving and the 'force north' feature turned off on both laptops. During this test, an active conversation was held between the two subjects.

Four subjects participated in the field tests with each taking turns to be the listener and the speaker. For one user, the HRTFs were individualised, whilst for the other three, the HRTFs were non-individualised. In each of the field tests, the listener was asked to qualitatively characterise whether the perceived direction of the spatialised speech signal presented over the headsets was perceptually aligned with the actual visual direction of the speaker. The listener was also asked to qualitatively describe whether they perceived the speaker's direction to be updated in real-time.

3. RESULTS

The experimental platform was subjected to the field tests as described above. As we see this work as a feasibility case-study, we give a qualitative report of the results that combines the observations across the four subjects. All observations described below are made with reference to the listener.

It was observed from the first test (stationary listener, moving speaker) that the speaker's speech was rendered in the correct direction when the speaker was stationary, that is both listener and speaker were not moving. As the speaker moved, however, the speaker's speech audio remained stationary and would jump to a new location after a period of time. The jump would place the speech audio in the speaker's present location at the time of the jump. This jump occurred, on average, every thirty to forty seconds. This periodic jump is in line with the update broadcast rate of the Rino units. When an update occurs, the waypoint at which the speaker is set on the Rino unit is updated causing a jump in the audio since the speaker may be quite distant from their previous waypoint position. Importantly, this has a greater effect when the users are closer together as the change in angle becomes much larger for the same distance travelled by the speaker (Figure 3). This result was anticipated and could be alleviated by more complicated software development. However, the fundamental issue is related to the hardware.

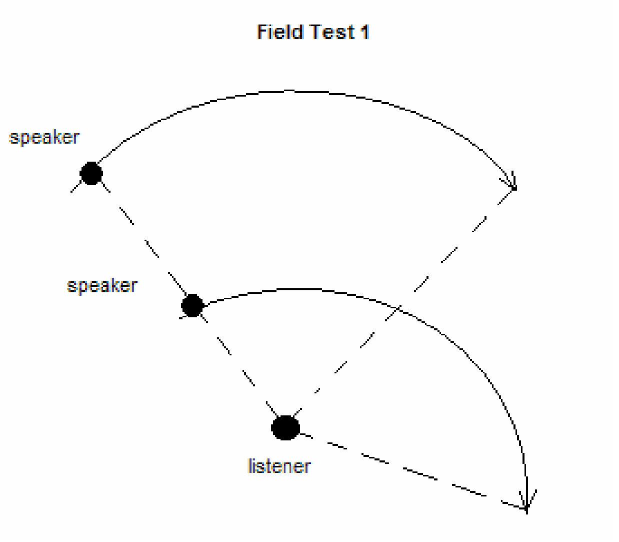

Figure 3. In the first field test, the jump in the audio, due to the update rate of the waypoint, has a greater effect when the speaker is located closer to the listener. This is due to a greater change in angle for the same distance covered.

In the second test, where the speaker remained stationary and the listener was moving, the spatial audio was rendered accurately as long as the distance between the listener and speaker was greater than roughly 30 metres. The estimated user orientation worked well as long as the listener was facing in the direction of motion. The speaker was always rendered in the direction from which the speaker was located. Setting the speaker as a waypoint worked in this situation to provide the relative direction to the speaker. As the waypoint (the speaker) was not moving, the thirty to forty second position update rate had no influence on the spatial fidelity of the audio signal. The criterion for the listener and speaker to be separated by 30 metres derives from the fact that the GPS data is streamed from the Rino unit to the laptop at a rate of approximately every two seconds (typical for COTS GPS units). Therefore, the HRTFs and in turn the spatial audio position is only updated approximately every two seconds. If the listener travels at approximately 5 kilometres per hour (average walking pace) and maintains his distance from the speaker at 30 metres, the change in angle is approximately 5 degrees every two seconds (Figure 4), which is within the minimum audible movement angle (between 2-5 degrees in azimuth [9]). Hence at distances larger than 30m, the two second update rate is tolerable to be classified as a real-time perceptual update of the spatial audio display.

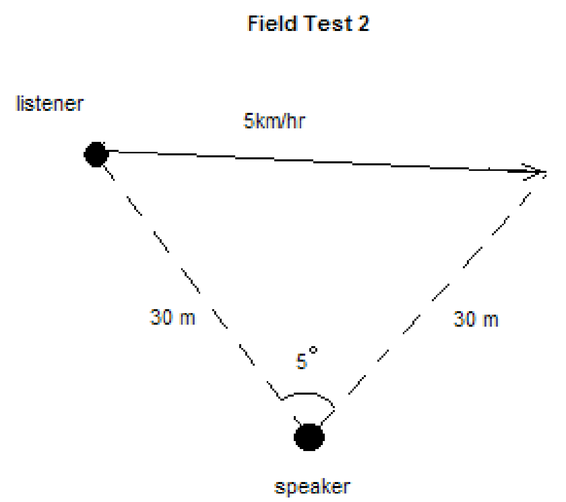

Figure 4. In the second field test, as long as the listener was more than 30m from the stationary speaker, the system rendered the direction of the speaker accurately to the listener, perceptually in real-time.

When both the speaker and listener were moving it became much more difficult to describe the accuracy of the spatial audio platform. It was essentially found that the spatial-audio fidelity varied strongly with velocity. Whenever the distance moved between each broadcast update was sufficiently large, the spatial audio would be rendered incorrectly with a jump in the speaker's location that would occur every thirty to forty seconds. This is in accordance with the observations of the first test.

A peripheral but important observation was also made regarding the quality of the audio over the radio link. Given the fact that the radio audio signals have a strong lowpass-filtered characteristic, they were surprisingly well spatialised. It is generally accepted that broadband signals are the best localised audio signals [1]. However, since the speech signal was taken from the output of the Rino units, any noise in the radio link would also be spatialised by the software. As this noise was likely to be broadband, the noise signal may have provided additional directional cues. An important factor preventing more extensive studies is the fact that a half-duplex radio link does not enable a natural communications environment, where more than one speaker can speak at once. A full-duplex communications environment is necessary to evaluate the full benefits of using GPS in a spatial audio communication system.

Finally, the observed CPU usage fluctuated from 4% to 30% on an Intel Pentium 4-M laptop with 2.2GHz processor. The fluctuations were reasonably periodic, at approximately one to two second intervals, coinciding with the rate of the update stream between the Rino and the laptop and hence a recalculation of the HRTFs.

4. CONCLUSIONS AND FUTURE WORK

A COTS GPS unit was used to develop an experimental mobile spatial-audio communication system. The experimental platform demonstrates that there are limitations in using COTS GPS units, although within the limitations, real-time spatial audio communications can be achieved. In the current setup, real-time spatial audio communications was achievable when users remained at distances of roughly 30 metres or greater and had not travelled far between Rino position broadcasts. These limitations indicate that position-update rates of current COTS GPS units do not provide for a high-fidelity spatial-audio communication system. In addition, the method of setting the speaker as a waypoint in order to obtain bearing information will not work for more than one speaker. A faster protocol for transmitting user positional data needs to be developed. Finally, although movement direction can indicate head-orientation, a dedicated head orientation system is required, as is a full duplex communications system.

5. REFERENCES

[1] S. Carlile, *Virtual Auditory Space: Generation and Application*. New York: Chapman and Hall, 1996.

[2] H. Moller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, pp. 171-218, 1992.

[3] S. Carlile, C. Jin, and V. Harvey, "The generation and validation of high fidelity virtual auditory space," presented at Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE, 1998.

[4] S. Carlile, P. Leong, and S. Hyams, "The nature and distribution of errors in sound localization by human listeners," *Hearing Research*, vol. 114, pp. 179-196, 1997.

[5] S. Carlile, C. Jin, and V. van Raad, "Continuous Virtual Auditory Space Using HRTF Interpolation: Acoustic & Psychophysical Errors," presented at Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing), Sydney, 2000.

[6] M. Billinghurst, J. Bowskill, and J. Morphett, "Wearable communication space," *British Telecommunications Engineering*, vol. 16, pp. 311-317, 1998.

[7] A. Rimell and M. Hollier, "Beyond the handset: Audio for the next millennium," *British Telecommunications Engineering*, vol. 16, pp. 306-310, 1998.

[8] "Rino 120 2-way radio & personal navigator owner's manual and reference guide," Garmin Ltd, 2002-2003.

[9] D. W. Grantham, B. W. Y. Hornsby, and E. A. Erpenbeck, "Auditory spatial resolution in horizontal, vertical, and diagonal planes," *Journal of the Acoustical Society of America*, vol. 114, pp. 1009-1022, 2003.