Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Auditory interface for teleoperation - Path following experimental results

A. Vasilijevic * D. Nad * N. Miskovic * Z. Vukic *

* *University of Zagreb, Faculty of Electrical Engineering and*
Computing, Laboratory for Underwater Systems and Technologies.
Unska 3, HR-10000 Zagreb, Croatia (e-mail: antonio.vasilijevic@fer.hr,
dula.nad@fer.hr, nikola.miskovic@fer.hr, zoran.vukic@fer.hr).

Abstract: For teleoperated unmanned vehicles, mishaps tend to occur during the periods of high workload, in situations where the operator must perform complex and stressful tasks. However, with many tasks performed simultaneously with flying, the relevant information is typically dispersed on a number of screens overloading the operator's visual channel. In order to address these unique human-factors problems associated with unmanned vehicles we suggest the use of an auditory display as a mean to reduce visual workload, to enhance situation awareness, and mitigate the visual and cognitive demands of contemporary marine teleoperations. Experiments were performed on the remotely operated surface marine platform (PlaDyPos) developed at the Laboratory for Underwater Systems and Technologies, Faculty of Electrical Engineering and Computing, University of Zagreb. The results show that the concept, guidance-by-sound is feasible in the real environment. The results are in line with previously obtained results from the real-time simulator showing that tracking quality can be further improved introducing supernormal auditory cues in order to provide better-than-normal operator's auditory resolution in the frontal region. We conclude that the use of hearing in the form of the auditory display emerges as an important advantage. Since practice has a major effect on performance, there is definitely more room for improvement in using interfaces we are not trained for.

Keywords: Unmanned vehicles; Teleoperation; Path following; Human-machine interface; Human perception; Auditory display; Supernormal auditory localisation cues.

---

1. INTRODUCTION

A contemporary Unmanned Vehicle (UV) control room is very often overloaded with screens presenting everything from video streaming from multiple cameras to various data acquired from multiple subsystems. The fact that information is exclusively presented visually, usually dispersed on different screens, may easily overload the operator's visual channel and prevent them from perceiving all important information related to the particular task. The effects of having a single operator controls multiple subsystems, investigated by (Cummings and Guerlain [2007]), is that as the number of systems increases, situational awareness performance degrades. As a result, mishaps tend to occur during the periods of high workload, in situations where the operator must perform complex and stressful tasks (Williams [2004]). Surprisingly, there is very little research examining the unique human factors problems associated with UVs (Ho et al. [2011]), (Bleicher [2010]).

We use our own egocentric orientation to judge the relative position of the objects around us. The operators ability to easily and accurately adopt the UVs frame of reference will have strong implications for overall spatial awareness. The research of (Donovan et al. [2006]) related to the role of interface in understanding egocentric orientation and spatial awareness, demonstrates the importance of having a well-designed operator interface in conveying spatial information for UVs. An auditory interface is one of the possible ways for reaching this goal.

Humans use auditory modality for development and maintenance of situation awareness in natural environments. We are able to determine the location of a sound source anywhere in the 360-degree space around us (even for those that are out of our field of view), monitor events at multiple locations simultaneously and switch our focus of attention between sound sources at will (Begault [1994]). Utilising these human abilities it is reasonable to expect that operators situational awareness can be improved by using spatial sound interface/display. There are two key disadvantages of hearing compared to vision. The spatial acuity of the visual channel is much better than that of the auditory channel (Shinn-Cunningham [1998]) and we use vision on a permanent basis for navigation, therefore we are very well trained for visual navigation. However this is not the case with navigation by hearing. The objective of this paper is to present the results of the path following experiments performed with remotely operated small-scale marine platform (PlaDyPos) and by operator using auditory display for feedback presentation. Section 2 describes the methodology and experimental platform. Section 3 describes the experiment set up and objective measure used for performance evaluation. Experimental results are presented and discussed in Section 4 and finally, a set of conclusions are provided.

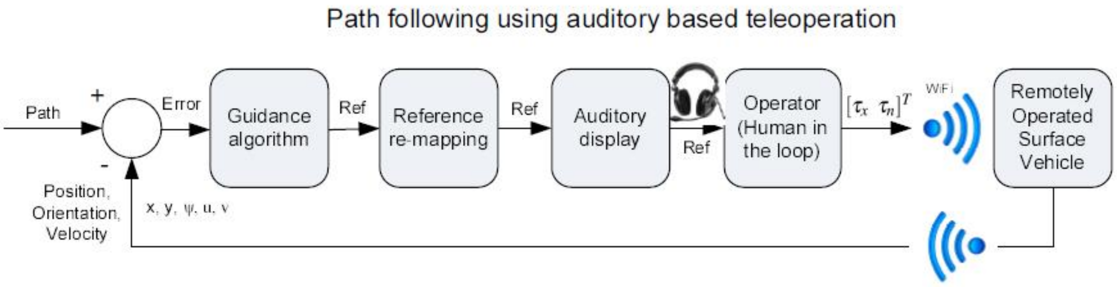

Fig. 1. Block diagram of the Auditory system

2. METHODOLOGY AND RESOURCES

The experimental system is unmanned surface vehicle (USV), teleoperated from the land using an auditory display for the feedback presentation. Block diagram of the system is shown in figure 1. Process feedback is not presented to the operator directly. The intermediate reference is generated by the guidance algorithm based on the feedback and guidance laws. In order to provide better-than-normal operator's auditory resolution in the frontal region the intermediate reference is then re-mapped using supernormal auditory cues method explained in subsection 2.3. The final reference is fed to the spatial sound generator, creating auditory cues presented to the operator.

2.1 Autonomous/Remotely operated small-scale marine platform (PlaDyPos)

The small-scale surface platform is designed for guidance tasks such as dynamic positioning, path following, diver following but it can be also remotely operated via WiFi link. It is an overactuated USV with 4 thrusters forming the "X" configuration. This configuration enables motion in the horizontal plane under any orientation. The current version of the platform is 0.35m high, 0.707m wide and long, and it weighs approximately 25kg. The control computer (isolated from environmental disturbances inside the platform hull) is in charge of data processing, control and guidance. The platform, teleoperated using auditory display, is shown in figure 2.

2.2 Guidance Algorithm

Path following refers to the problem of forcing a vehicle to converge to and follow a desired spatial path, without any temporal specifications (Pascoal et al. [2006]). Applying control algorithms in a form of appropriate guidance laws (steering laws) provides method suitable for path following as long as the vehicle speed exist. The guidance system generates the reference for the controller (human operator) which is responsible for controlling the direction of the vehicle velocity. Consequently, the control objective is to track predefined path and it is given by:

$$\lim_{t\to\infty}(p(t)-p_D(t))=\lim_{t\to\infty}(e(t))=0 \qquad (1)$$

Fig. 2. Participant operates USV using auditory display

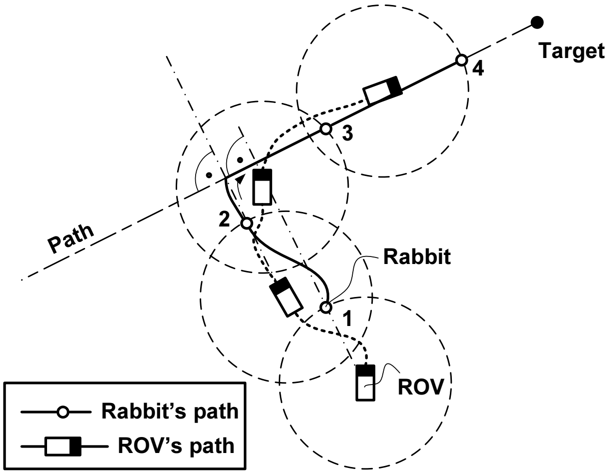

Fig. 3. Positioning of a dynamic virtual target (rabbit), relative to the USV

where $p(t)$ represents the actual USV position, $p_D(t)$ desired USV position and $e(t)$ tracking error. Tracking error consist of cross-track and along-track errors but for the path-following purposes, desired USV position $p_D(t)$ is USV's direct projection onto the path, meaning that only cross-track error is relevant.

Number of guidance laws are developed to ensure stabilization of $e(t)$ to the origin. We have chosen the method used in ship motion control systems (Fossen [2002]), known as enclosure-based steering, figure 3.

Let's imagine the circle of radius $r > 0$ centred at the USV position $p(t)$. If cross-track error $e(t)$ is greater than the r, the circle and path do not intersect. The reference, virtual target is positioned at the point on the circle closest to the path, guiding the vehicle straight towards the path, as illustrated in figure 3, positions 1 and 2. If $e(t) < r$ then the circle intersects the path at two points. The steering law generated by the enclosure-based steering strategy says that the vehicles velocity vector has to be directed toward the intersection point that corresponds to the desired direction of travel (Breivik and Fossen [2009]), as shown in figure 3, positions 3 and 4. Such a solution generates two references: the virtual target position (x_{vt}, y_{vt}) and consequently, desired direction of the vehicle velocity vector:

$$\chi(t) = atan2(y_{vt}(t) - y(t), x_{vt}(t) - x(t)), \tag{2}$$

To calculate virtual target position the following two equations (circle and line) must be solved:

$$r^2 = (x_{vt}(t) - x(t))^2 - (y_{vt}(t) - y(t))^2, \tag{3}$$

$$y_{vt}(t) - y_k(t) = \frac{y_{k+1}(t) - y_k(t)}{(x_{k+1}(t) - x(t)} \cdot (x_{vt}(t) - x(t)), \tag{4}$$

where (x_k, y_k) and (x_{k+1}, y_{k+1}) represent waypoints defining the line to be followed. In Vasilijevic et al. [2012] it was shown that enclosure-based steering with radius of 10 meters ensures optimal path following performance using the auditory display.

2.3 Auditory Display

The auditory display presents the reference (virtual target) to the operator over headphones in the form of non-verbal spatial sound cues. Based on its angular perception, the operator orients and flies the vehicle towards the virtual target. We can say that in order to achieve efficient path following, good (low bias and high resolution) perception in the neighbourhood of zero azimuth is essential.

Azimuth perception in the remaining areas does not need to be that good but it still needs to preserve the feeling of vehicle dynamics. For the guidance applications, even for azimuth resolution in the neighbourhood of zero azimuth, where is the best, better resolution is desirable. This raises the question of whether it might be possible to design processing for the operator that enhances the effective resolution artificially. In (Durlach et al. [1993]), it is pointed out that it should be possible to improve performance by synthesizing intentionally distorted, supernormal localization cues even if the result is "unnatural". Inspired by this idea presented in (Shinn-Cunningham et al. [1998]), we synthesized the supernormal azimuth localization cues by re-mapping the azimuth position of the sound source according to transformation (5). Visual interpretation is given in figure 4.

$$f_K(\theta) = \frac{1}{2} arctan[\frac{2K \cdot sin(2\theta)}{1 - K^2 + (1 + K^2)cos(2\theta)}] \tag{5}$$

For $K > 1$ this transformation provides better-than-normal resolution in the frontal region but reduce resolution on the side with anchor points $\theta = 0$ and $\theta = \pm\pi/2$. For $K < 1$, the opposite occurs as shown in figure 5.

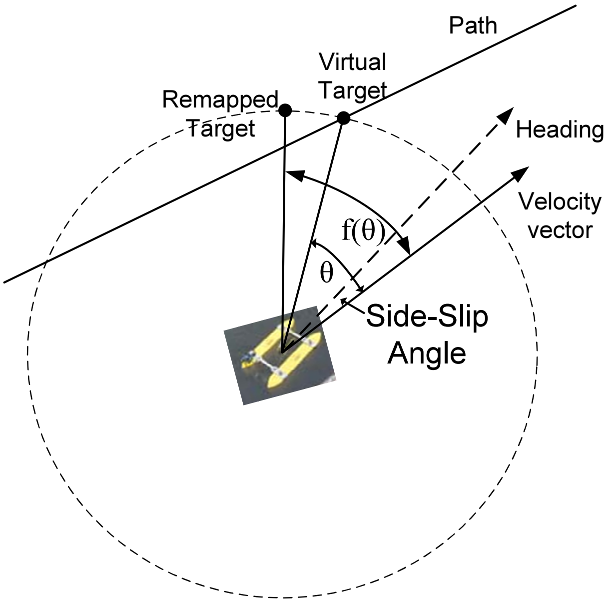

Fig. 4. Remapping by supernormal localisation cues. Original azimuth θ is remmaped to $f(\theta)$

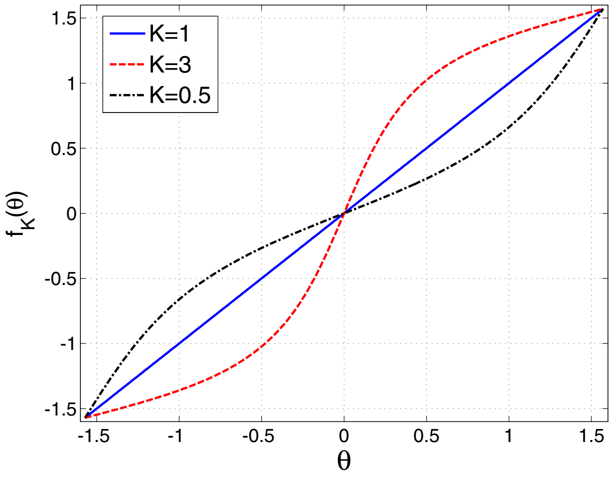

Fig. 5. Supernormal transformation of azimuth θ to new reference $f(\theta)$. For $K > 1$ spatial resolution in the neighbourhood of $\theta = 0$ is improved

This paper aims to prove in-practice the path following results previously obtained indoor, using real-time vehicle simulator which showed that for $k \geq 3$ tracking quality of the path following was significantly improved (30-40% in average), paid by increased yaw effort (20%). Total control effort (surge and yaw) was just slightly increased.

The new reference is presented to the operator via auditory display in the form of sound beacon virtual located at the position corresponding to the relative position of the virtual target against the USV's flow frame of reference as shown in figure 4. Thus, Virtual Auditory Display spatializes sounds by tracking the target's and USV location, momentary orientation and velocity. It is a headphone-based system in which localisation cues are generated by the inexpensive (free) spatial sound application FMOD Ex API (Healy and Smeaton [2009]). The application supports generalized Head-Related Transfer Functions (HRTF) (Begault and Wenzel [1992]), (Begault [1994]) to simulate the normal auditory localization cues and provide spatial audio perception.

The virtual sound source used for an AR auditory display is a pink noise amplitude-modulated at 10Hz.

3. EXPERIMENTS AND MEASURES

Three members of the Laboratory for Underwater Systems and Technologies, Faculty of Electrical Engineering and Computing, University of Zagreb with different level of auditory guidance experience participated in the experiments. Each participant performed two types of missions: using normal auditory cues $K = 1$ from equation (5) representing the normal resolution of the human hearing and using supernormal cues $K = 4$ representing the case with improved spatial resolution in the frontal region. Experiments were performed on the systems elaborated in section 2. The USV was controlled manually, using joystick and audio stimuli were generated with a FMOD EX and presented to subjects over stereo headphones AKG K66. Mission results used for analysis consist of accomplished USV path $p(t)$ and control effort $\tau(t)$. The mission trajectory combined with the desired path $p_D(t)$ is used to calculate tracking error $e(t)$. In order to experimentally evaluate HMI performance, the objective function (Fossen [1994]) defined by integral square measures of weighted tracking error and control effort (vehicles input forces and torques), was introduced.

$$minPi = \int_0^T \left(e(t)^T \cdot Q \cdot e(t) + \tau(t)^T \cdot P \cdot \tau(t) \right) \, \mathrm{d}t \qquad (6)$$

$P > 0$ and $Q \geq 0$ are the weighting matrices, e is the cross-track error and $\tau = [\tau_x \;\; \tau_n]^T$ is the control effort vector consisting of forward force and yaw torque. The best performance corresponds to the minimal performance index. For the purpose of analysis, performance index components, the tracking quality and the control effort, are separated and scaled by the time needed to accomplish the mission, (7). It allows comparison of corresponding scaled values from different missions in a way that eliminates the effects of mission specific influences i.e. longer and shorter missions.

$$\begin{aligned} Pi_T &= \int_0^T \left(e(t)^T \cdot e(t) \right) \, \mathrm{d}t / T \\ Pi_E &= \int_0^T \left(\tau(t)^T \cdot \tau(t) \right) \, \mathrm{d}t / T \end{aligned} \qquad (7)$$

Pi_T and Pi_E are tracking quality and control effort part of the performance index while T represents duration of the mission.

4. RESULTS AND DISCUSSION

Experiments were conducted at the lake Bundek, Zagreb, Croatia. All together 18 missions were performed, out of which 14 successful missions were taken into account for analysis. Four mission were unsuccessful due to communication (WiFi) interruptions between the control station and the USV, causing unavailability of feedback information or due to operator misorientation, front/back reversal. Reversals are common for non-experienced users of the auditory display, especially when headphone-based system supporting only generalized HRTF, is used, as shown in figure 6.

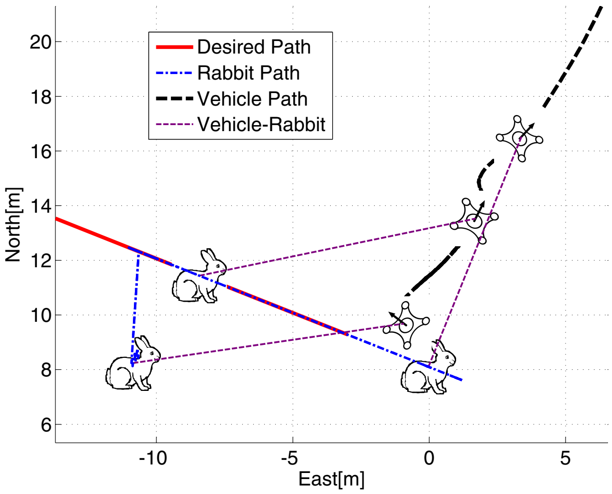

Fig. 6. Misorientation of the operator, front/back reversal. Common among untrained operators

Table 1. Missions summary

Miss.	Oper.	Pi_T $K = 1$	Pi_T $K = 4$	Pi_E $K = 1$	Pi_E $K = 4$
1	1	2.45		12.45	
2	1	2.36		13.82	
3	1		0.46		12.17
4	1		0.38		16.02
Aver.	**1**	**2.4**	**0.42**	**13.13**	**14.09**
5	2	2.98		12.82	
6	2	1.27		13.25	
7	2	2.4		13.02	
8	2		0.9		14.53
9	2		2.04		13.77
Aver.	**2**	**2.21**	**1.47**	**13.03**	**14.15**
10	3	0.86		12.56	
11	3	1.09		12.11	
12	3		0.43		13.91
13	3		0.47		16.55
14	3		0.48		12.11
Aver.	**3**	**0.97**	**0.46**	**12.33**	**14.19**
Total Aver.	**All**	**1.91**	**0.73**	**12.85**	**14.15**

Desired path is defined as a straight line. Mission results are used to calculate performance index (6) and its tracking quality and control effort components (7). According to the preliminary results on a simulator, we expected similar surge effort results for both display setups but definitely more agile steering for $K = 4$ due to improved spatial auditory resolution. In order to isolate control effort related to the steering, Pi_E is calculated using only yaw torque $\tau = [0 \;\; \tau_n]^T$. The experimental results are summarized in table 1.

Experiments confirmed the results from the real-time ROV simulator. First, that guidance with the auditory display was feasible, but this time in real environment conditions. And second, that path following performance

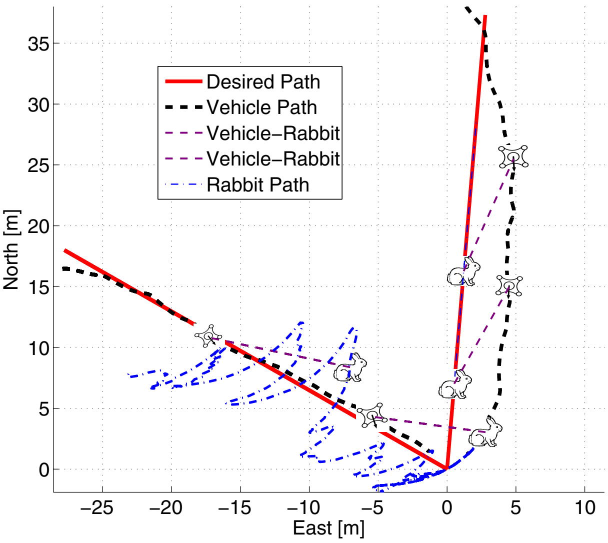

Fig. 7. Mission with Normal and SuperNormal auditory cues, $K = 1$ and $K = 4$. Vehicle and Virtual Target (Rabbit) paths

has improved on average by 61% ($P_{IT}(K = 4)/P_{IT}(K = 1)$) by implementing supernormal localisation cues. Figure 7 illustrates the ends of the two missions, the first one from top-left to bottom was the mission with $K = 4$, while the second one from top to bottom was the mission with $K = 1$. Path of the mission with $K = 4$ (black dashed line) clearly followed the desired path (red solid line) better than the path of the mission with $K = 1$. This improvement of performance is even more obvious by observing the figure 8, which presents the cross-track errors. The virtual target (Rabbit) (blue dash-dot line) traveled along the desired path for $K = 1$ while for $K = 4$ it had an off-the-path trajectory as illustrated in figure 9. That characteristic is the result of remapping (5).

Control effort related to steering, from table 1, has increased by 10% when supernormal cues were used. Due to improved path following, the resultant path was shorter and surge control effort has decreased proportionally. The result is that there were no significant changes in total control effort performance; increased steering effort was compensated by reduced surge effort. Value of the total performance index (6) depends on which aspect, tracking quality or control effort is more important for the particular application. Let us assume that both components are equally important and choose the coefficients Q and P from (6) accordingly. Tracking quality was improved by 60% when Supernormal instead of Normal localisation cues were used. The control effort has remained the same. As a result, the total Performance index was improved by 30% due to nonlinear reference remapping (5).

Performance index varied among the participants (table 1) but all participants, regardless of level of training on the auditory display, improved their performance significantly by using the display with supernormal localisation cues. The most experienced operator (operator 3) performed the best, achieving the lowest value of Performance index. In comparison with first two operators, his average performance index for all missions is approximately two times lower. This shows that practice can significantly improve performance in using interfaces we are not trained for, i.e. auditory interface, which is in line with the results published in (Walker and Lindsay [2006]).

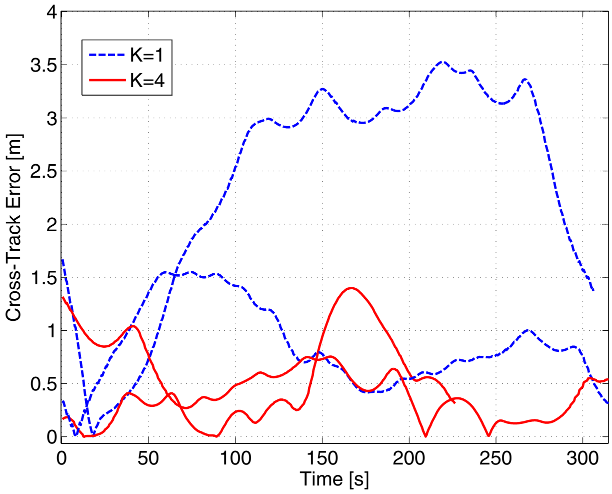

Fig. 8. Cross-Track error of the two missions with SuperNormal auditory cues and two with Normal cues

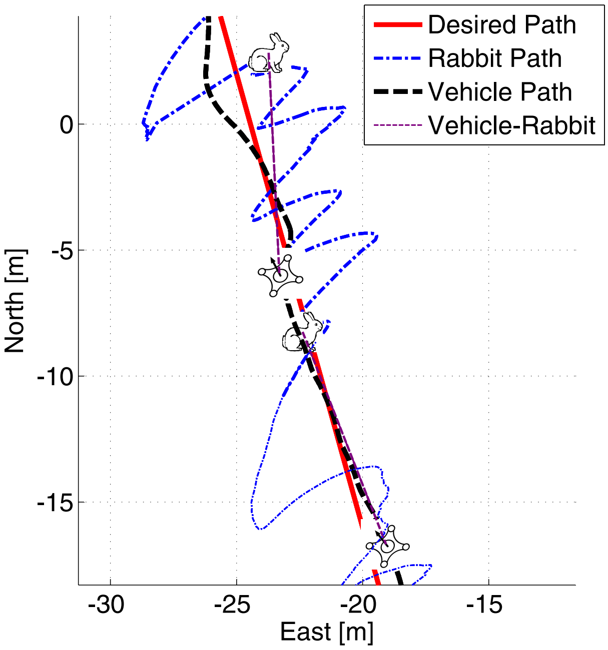

Fig. 9. Mission with SuperNormal auditory cues ($K = 4$). Vehicle and Rabbit paths with two examples of Vehicle Position and associated Virtual Target

5. CONCLUSIONS

The sense of hearing is used mostly for communication and safety. Humans also use hearing for development and maintenance of situation awareness in natural environments but more to be redundant or to extend visual field of view, not exclusively for navigation in space. Main issues related to use of non-speech auditory display for navigation are insufficient spatial resolution of the hearing channel and lack of experience, training in audio navigation. This paper addresses the issue of spatial acuity by suggesting the use of auditory interface presenting supernormal localisation cues. The system is implemented and tested in practice for guidance of the USV. Results presented in this paper clearly show that path following performance can be significantly enhanced by using the suggested system. Tracking quality has been improved by approximately 60% maintaining the same control effort.

Auditory interface can also be used to improve overall performance of the teleoperation by reducing visual and cognitive workload. In contrast, the results obtained with auditory interface show that practice plays an important role in quality of navigation, performance increases with practice, as is often the case with the use of new interfaces. Future work will focus on auditory display dedicated to trajectory tracking, requiring operator to satisfy steering and speed control laws. For that purpose, accurate perception of target distance is equally important as perception of azimuth.

ACKNOWLEDGEMENTS

The authors would like to thank Nikola Stilinovic for his support during the field trials.

REFERENCES

Durand R. Begault. *3-D sound for virtual reality and multimedia*. Academic Press Professional, Inc., San Diego, CA, USA, 1994.

Durand R. Begault and Elizabeth M. Wenzel. Techniques and applications for binaural sound manipulation. *The International Journal of Aviation Psychology*, 2(1):1–22, 1992.

A. Bleicher. The gulf spill's lessons for robotics. *Spectrum, IEEE*, 47(8):9 –11, august 2010.

Morten Breivik and Thor I. Fossen. Guidance laws for autonomous underwater vehicles. In Alexander V. Inzartsev, editor, *Underwater Vehicles*. 2009.

Mary L Cummings and Stephanie Guerlain. Developing operator capacity estimates for supervisory control of autonomous vehicles. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(1):1–15, 2007.

Sarah-Louise Donovan, Tom Triggs, Defence Science, Technology Organisation (Australia), Defence Science, and Technology Organisation (Australia). *Investigating the effects of display design on unmanned underwater vehicle pilot performance [electronic resource] / Sarah-Louise Donovan and Tom Triggs*. DSTO Fishermans Bend, Victoria, 2006.

N. I. Durlach, B. G. Shinn-Cunningham, and R. M. Held. Supernormal auditory localization. I. General background. *Presence*, 2(2):89–103, 1993.

Thor I. Fossen. *Guidance and Control of Ocean Vehicles*. John Wiley and Sons, Chichester, England, 1994.

Thor.I. Fossen. *Marine control systems: guidance, navigation and control of ships, rigs and underwater vehicles*. Marine Cybernetics, 2002.

G. Healy and A.F. Smeaton. Spatially augmented audio delivery: Applications of spatial sound awareness in sensor-equipped indoor environments. In *Mobile Data Management: Systems, Services and Middleware, 2009. MDM '09. Tenth International Conference on*, pages 704 –708, may 2009.

Geoffrey Ho, Nada Pavlovic, and Robert Arrabito. Human factors issues with operating unmanned underwater vehicles. 55(1):429–433, 2011.

A. Pascoal, C. Silvestre, and P. Oliveira. Vehicle and mission control of single and multiple autonomous marine robots. In G.N. Roberts, R. Sutton, and Institution of Electrical Engineers, editors, *Advances in Unmanned Marine Vehicles*. 2006.

B. Shinn-Cunningham. Applications of virtual auditory displays. In *Proceedings of the 20th International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1105–1108, 1998.

Barbara G. Shinn-Cunningham, Nathaniel I. Durlach, and Richard M. Held. Adapting to supernormal auditory localization cues. i. bias and resolution. *Journal of the Acoustical Society of America*, 103(5):3656–3666, 1998.

A. Vasilijevic, E. Omerdic, B. Borovic, and Z. Vukic. Acoustically aided hmi for rov navigation. In *Proceedings of the 3rd IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles*, pages 1105–1108, 2012.

Bruce N. Walker and Jeffrey Lindsay. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2):265–278, 2006.

Kevin W. Williams. A summary of unmanned aircraft accident/incident data: Human factors implications. *U.S. Department of Transportation, Federal Aviation Administration, Civil Aerospace Medical Institute, Oklahoma City*, 4, 2004.