HAL
open science

Expert evaluation of the usability of HeloVis: a 3D Immersive Helical Visualization for SIGINT Analysis

Alma Cantu, Thierry Duval, Olivier Grisvard, Gilles Coppin

To cite this version:

Alma Cantu, Thierry Duval, Olivier Grisvard, Gilles Coppin. Expert evaluation of the usability of HeloVis: a 3D Immersive Helical Visualization for SIGINT Analysis. EuroVR 2019 : 16th International Conference on Virtual Reality and Augmented Reality, Oct 2019, Tallinn, Estonia. pp.181-198, 10.1007/978-3-030-31908-3_12 . hal-02292428

HAL Id: hal-02292428

https://hal.science/hal-02292428v1

Submitted on 19 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Expert evaluation of the usability of HeloVis: a 3D Immersive Helical Visualization for SIGINT Analysis

Alma Cantu[1,3], Thierry Duval[2,3], Olivier Grisvard[1,2,3], and Gilles Coppin[2,3]

[1] Thales DMS, Brest, France
[2] IMT ATlantique, Brest, France
[3] Lab-STICC, UMR CNRS 6285, Brest, France

Abstract. This paper presents an evaluation of HeloVis: a 3D interactive visualization that relies on immersive properties to improve user performance during SIGnal INTelligence (SIGINT) analysis. HeloVis draws on perceptive biases, highlighted by Gestalt laws, and on depth perception to enhance the recurrence properties contained in the data. In this paper, we briefly recall what is SIGINT, the challenges that it brings to visual analytics, and the limitations of state of the art SIGINT tools. Then, we present HeloVis, and we evaluate its efficiency through the results of an evaluation that we have made with civil and military operators who are the expert end-users of SIGINT analysis.

Keywords: HCI · Visual Analytics · Immersive Analytics · Scientific Visualization · 3D User Interaction · Virtual Environments · Virtual Reality

1 Introduction

SIGnal INTelligence (SIGINT) consists of gathering knowledge based on radar signal analysis. Nowadays SIGINT has to face an increasing amount of data, and usual 2D SIGINT solutions are no longer allowing a high-quality analysis on a day to day basis. This is why we have explored the use of 3D immersive techniques to deal with the numerous aspects of SIGINT, as suggested by Brath about the intake of 3D for visualization [4]. This is how we build HeloVis[6], an interactive 3D immersive visualization that makes a profit of these techniques to better support SIGINT analysis. In this paper, in section 2 we first recall briefly the specificities of SIGINT identified thanks to interviews of military operators. Then in section 3, we present the state of the art and why it does not meet all the SIGINT requirements. In section 4 we detail the HeloVis tool and finally, in sections 5 and 6 we assess the benefits of such an approach by presenting the results of an evaluation, conducted with 11 SIGINT experts, measuring user performance.

2 SIGINT

Radar detection consists of sending an electromagnetic emission that is reflected by objects and then intercepted back by the radar. Such emissions can be intercepted by other sensors named electronic support measures. Electronic Warfare (EW) exploits the interception of radar emissions to identify their origin, to characterize their threat level and to enable the appropriate reaction in case of conflict, or conversely, to prevent an opponent from doing the same. This identification requires having precise knowledge of the existing radar signal signatures. Gathering this knowledge is called SIGINT, it is usually performed by dedicated military forces. It consists of capturing and analyzing as many emissions as possible to characterize their signatures and fill the reference database used in electronic warfare.

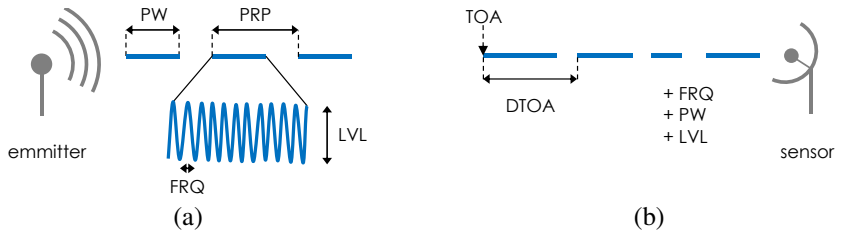


Fig. 1. (a) Attributes of the emitted data. (b) Attributes of the intercepted data

Radar signals consist of a series of pulses of a certain duration spaced in time. These pulses are emitted on a specific *frequency* (such as radio emissions) and with a specific power, named *level*. Thus, *pulse width* (PW), *pulse repetition period* (PRP), *frequency* and *level* are the four main attributes that characterize a pulse (see Figure 1 (a)). These attributes may vary from one pulse to another in the same signal to increase detection or to hinder the identification. This variation is what makes the signal specific enough that it can be used for identification. Between the emission and the interception, pulses are subject to losses, distortion, and noise, whether it be from diffraction, emitter issues or sensor issues. The sensors are able to record pulses according to their *frequency*, *PW*, *time of arrival (TOA)* and *level* values. To access the PRP, they also contain the computed *delta of TOA (DTOA)* such as the SIGINT datasets contain these five attributes (see Figure 1 (b)).

To be able to identify the characteristics of a radar signal the operators need to distinguish pulses of a signal from noise or other signals. A decade ago, most radars emitted their pulses on the same frequency such that the association was made on the frequency

attribute and that the other attributes were used only to distinguish two radars with close frequency values. However, the latest radars can emit on different frequency ranges creating signals with a frequency varying pseudo-randomly. To overcome this new issue the operators rely on the variation of the PW, the PRP, and the level. Constrained by the need for comparison between emitted and reflected pulses, PWs and PRPs vary according to a recurrent pattern which implies similar values but also recurrent time values which make the association possible. Regarding the level, although the level values vary uncommonly at the emission, the emitter often rotates to cover every angle such as the emission takes the shape of a set of curves which is singular enough to be detected and associated (see Figure 2), however the emitter can also make the level vary (for depth detection purposes), which complicates the association. Consequently, to associate pulses from the same radar signal relies on identifying and correlating similar values, singular variations or recurrent time values.

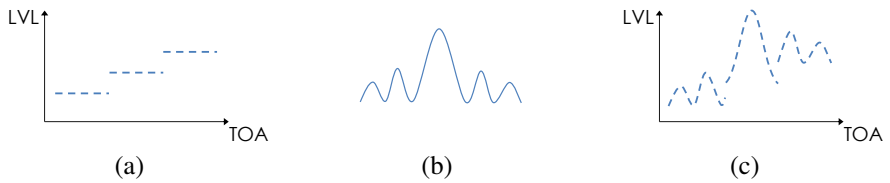


Fig. 2. Different shapes of level variation during its emitting. (a) Emitted level variation. (b) Level transformation due to the rotation of the emitter. (c) Intercepted level variations.

3 Related work

By combining cluster and modulo detection, the latest algorithms allow the association of more than 80% of the SIGINT data [22]. To associate the 20% radar signals remaining from processing, we can rely on human perception because humans have the ability to correlate despite noise or missing values and have the mind plasticity to overcome novelty [20]. Accessing the attribute values to perform the association (identifying similar values, singular variations, and recurrent time values) requires their encoding on visual variables (positions, color, size, shape, etc.) [3]. According to the Gestalt law of proximity, the position permits to associate data into clusters [15].

To handle the multidimensional aspect of SIGINT data, operators currently use multiple views of 2D charts of radar signal over time (see Figure 3). It makes possible to identify similar values on the two attributes represented and can be correlated to other views with

the use of color, as illustrated in [19]. However, these charts are subject to noise and occlusion, which is complicating the selection task needed for the characterization step. To identify recurrent time values, the 2D chart is of no specific support and the amount of data prevents any one-by-one pulse detection.

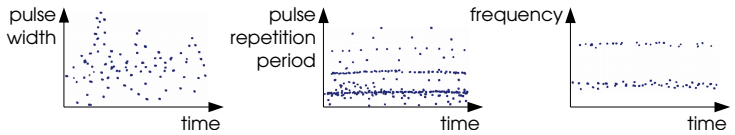


Fig. 3. Illustration of the SIGINT reference tool

Using parallel coordinates in 2D [14] or even 3D [8] for SIGINT could permit to easily identify similar values on an attribute and correlates it with others. However, this representation is also subject to noise, and even if this noise could be reduced by interacting with the axis of the parallel coordinates [8], it allows access to neither the level variation nor the recurrent time values.

3D charts extend the correlation on a row to three attributes or more if we consider other visual variables [9]. Some recent studies have proven cluster detection to be more effective on stereoscopic displays [12,21] implying the use of stereoscopic screens or immersive devices. Concerning the noise and occlusion issues, 3D charts are impacted as much as the other representations, and sometimes 3D representations still face occlusion and distortion issues, which questions their usefulness [18]. Whether these criticisms concern the use of flattened 3D or the use of 3D with depth perception (stereoscopic and motion parallax), these two methods are to be distinguished. Indeed, numerous studies comparing flattened 3D and 3D with depth have established significant performance differences [4]. This does not mean that the 3D with depth overcomes all the disadvantages of 3D but it can potentially improve the balance in some cases.

A helical baseline can also highlight time recursions, it has been used to handle geographic dataset [11]. This interactive technique allows users to identify recursions based on any period, but it is unusable for SIGINT because it does not permit to access enough data because it is meant to be part of a graphic panel and as such is constrained by the size of the display. Transposing such a system to a 3D situation using an immersive tool would permit an infinite extension of the display size.

The main issue that makes irrelevant the existing work is brought by biases among the data. Even though frequency and PW may have constant or recurrent values, easy to

associate, their values are subject to error and noise (see Figure 4 (a)). Regarding the level, its nonlinear behavior complicates the selection (see Figure 4 (d)). The DTOA, which has recurrent values and which is not too much impacted by noise, is irrelevant in case of missing data (see Figure 4 (c)) or multiple listening (see Figure 4 (b)). There remains only the value of time which is not subject to error, which is not impacted by missing data or multiple listening and which contains, through its recurrence, the belonging information of the signal. As 3D techniques have proven to be useful to solve issues related to time visualization [1,10], we propose a new way of representing SIGINT data to facilitate the association of pulses into the same radar signal within a 3D immersive environment, by extending the helical baseline to 3D.

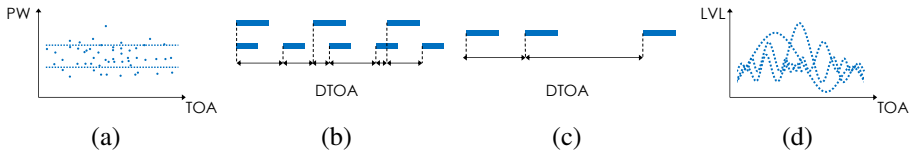


Fig. 4. Visual evidence of the biases among the SIGINT data

4 HeloVis

HeloVis is a typical contribution to Immersive Analytics [7,2]. HeloVis is an interactive 3D visualization within which each pulse is represented by an object positioned on a helical scale depending on its time value. The period value can be modified by the end-user and impact the helical scale by twisting or untwisting it. More details can be found in [6]. As a 3D representation, it is meant to be visualized with an immersive tool or at least a stereoscopic tool.

Potentially the user will reach a period value that corresponds to the PRP of a radar signal that s/he will be able to detect because all the pulses of this radar will be aligned. In the case of several PRP on the same signal, several alignments will appear over the cylinder (see Figure 5).

Missing values will correspond to a hole in the detected alignment, which will not impact the detection according to the Gestalt law of continuity (see Figure 6) [15].

A detected alignment does not necessarily mean that the period is equal to the PRP value but that it may also be a divisor or a multiple of the PRP. If the period corresponds to

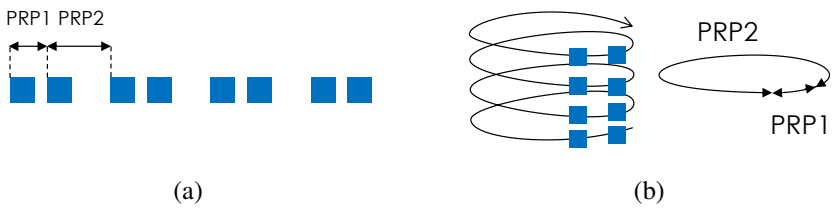


Fig. 5. Different representations of a radar signal containing several repeated pulses. (a) Linear time. (b) Helical time with period = PRP1 + PRP2.

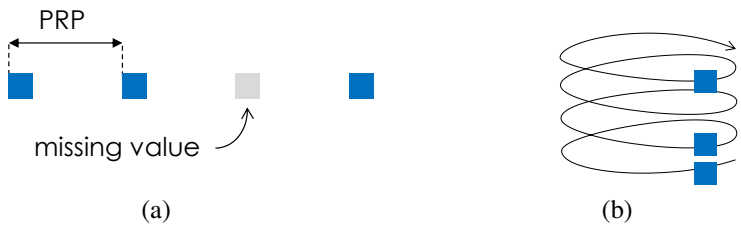


Fig. 6. Different representations of a radar signal with a missing value. (a) Linear time with a missing value. (b) Helical time with a missing value.

a multiple of the PRP of the detected pulses, there will be as many alignments as the ratio between the period and the PRP (see Figure 7(a) and (b)). If the period corresponds to a divisor of the PRP of the detected pulses, there will be only one alignment but the helical referential will be too much twisted (see Figure 7(c)).

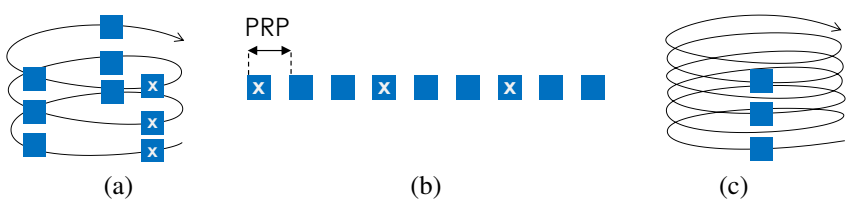


Fig. 7. Different representations of a radar signal according to its PRP. (a) Helical time with period = 3 x PRP. (b) Linear time with period = 3 x PRP. (c) Helical time with period = PRP / 2)

There can be a mistake between the detection of a period being a multiple of a PRP and a signal with several PRP values. The period being a multiple of a PRP will be represented by alignments equally displayed among the view while several PRP values

will be displayed according to these values. To verify the hypothesis of a multiple of a PRP requires dividing the period to return to the expected PRP which is enabled by a specific feature. Finally, a radar signal does not necessarily have a constant PRP value such that the detection will not necessarily be alignment but at least a geometric shape, as PRP variations are always defined by functions.

As SIGINT analysis requires a correlation on several dimensions, HeloVis encodes information thanks to the visual variables of color (see Figure 8) and third dimension: radius of the cylinder (see Figure 9). Being able to differentiate values of frequency or pulse width thanks to color strengthens the cluster detection provided by the helical representation and permits to identify outliers ([3]). Using the radius to represent information increases also the cluster detection and improves the selection of clusters ([4]).

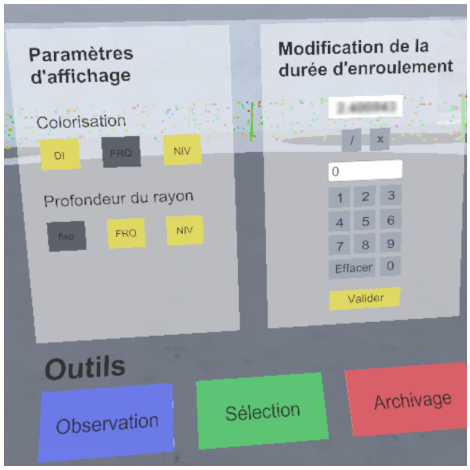


Fig. 8. Control panel illustrating the features of HeloVis: attribute encoding panel (top left), period modification panel (top right), observation tool button (bottom left), selection tool button (bottom center), archiving tool button (bottom right)

The selection tool (see Figure 10(c)) allows the user to send the data to a new workspace acting like a filter and reducing the number of data to analyze (see Figure 11). The user can send back data at any moment if he/she considers as not belonging to the selected radar signal.

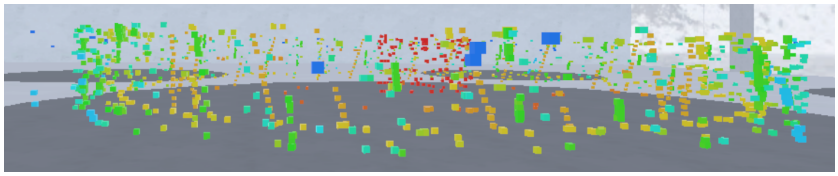
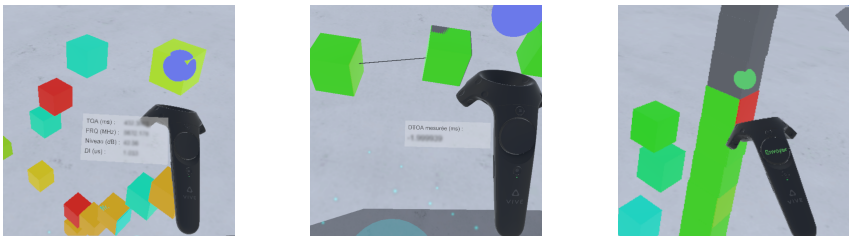


Fig. 9. View encoding frequency on the radius on HeloVis

HeloVis also provides direct access to the numeric value of the data that can carry a lot of information for expert users (see Figure 10 (a)). For example, frequency value can indicate the threat level and local value variation can be the signature of a unique radar.

Finally, HeloVis contains signal intelligence specific tools such as a tool allowing to measure the DTOA between two pulses completing the panel of information required to identify a radar signal (see Figure 10 (b)).



(a) (b) (c)

Fig. 10. Interaction tools on HeloVis. (a) Identification of pulse attributes. (b) Identification of the DTOA between two pulses. (c) Selection of some pulses.

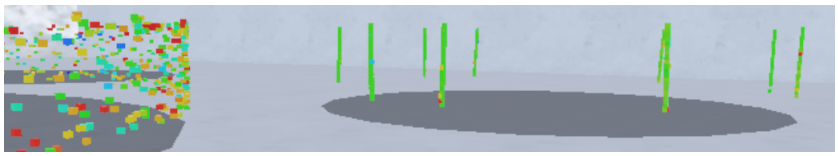


Fig. 11. Pulses sent on another view on HeloVis

All these tools are accessible by pointing interaction with a 3D cursor following the existing guidelines [13]. Navigation was also possible through teleportation.

5 Evaluation

We believe that HeloVis allows the users to perform a better association of all the pulses belonging to a radar. To challenge this assumption we conducted an evaluation comparing the user performances between the tool currently used by SIGINT operators (we will call it ”the reference tool”, which we cannot talk too much about because of confidential restriction, it is a set of 2D charts displaying pulses over different attributes, see Figure 3) and HeloVis. A better association implies a better association rate and a better quality of the associations such that we make the following hypothesis:

H1 The quantity of radar signals associated is improved by the use of HeloVis compared to the reference tool.

H2 The quality of the radar signals associated is improved by the use of HeloVis compared to the reference tool.

5.1 Experimental framework

As this evaluation falls into the category of user performance evaluation of Lam et al., we have designed it to access objective metrics related to our hypothesis [16]. As we were targeting mainly an evaluation of our new visualization metaphor, we did not want it to be biased by a comparison of usability between our two visualization tools, so we did not collect time metrics or error rate but only metrics about the quantity and quality of the associations. To master the experimental conditions of the evaluation we recorded demographic data and data related metrics. Finally, to be able to get an insight into the acceptability of HeloVis, we recorded subjective metrics about user experience.

Procedure - We followed the same procedure for each participant. We welcomed the participants and we introduced them to the functioning of the evaluation to which they agreed by signing an attestation. We asked them to fulfill a demographic questionnaire indicating their age, gender, job (military or civilian), expertise about SIGINT and expertise about immersive devices. Then we asked them to follow the same procedure on the two systems in an order that differs from a participant to another. The participants first learned how to use the system on a training dataset. We explained to them how to perform signal association and they could ask any question. Secondly, we requested them to perform some associations on a given dataset without assistance from us. The association required to associate as many signals as they could and in the best quality. Finally, we asked them to fulfill a SUS questionnaire [5] to record their insight. Once they had performed this on the two systems we debriefed the evaluation and answered

their remaining questions. From this procedure, we obtained demographic data and subjective metrics (SUS results, observations on participants' behavior and participants' comments) and extracted objective metrics from the logs of the associations. The number of associations gave us the first metric about H1 and the pulses contained in each association gave us insight about H2. Before the evaluation, we had associated each pulse of the given dataset to a radar signal with the help of a SIGINT expert. Thus, we were able to identify the exact amount of each signal of an association. For each association, we identified the signal that was the most represented and we considered that this was the targeted signal. The ratio between the pulses that do not belong to the targeted signal and the number of pulses of the association provided a metric about the noise percentage. The ratio between the pulses of the targeted signal and the total amount of pulses contained in this signal provided a metric about the selection percentage. Despite time and error rate would have given us some insights to challenge our hypothesis, they would have been also influenced by the interaction quality which we did not want to include in our study. For the same reason, we did not establish a time constraint.

5.2 Association task

The association task asked for the evaluation is describable on both systems thanks to the Munzner's nested model [17]. The domain problem we face is the associations of pulses of the same radar signal. To solve this problem requires the abstract operations of identifying similar data, singular variations among them or recurrent time values and correlating them. To perform these operations requires the use of different techniques on each system. For the reference tool, it requires the use of multiple synchronized views of 2D charts on which one can zoom, filter, and/or colorize the data to access their values all over the view (see Figure 3). For HeloVis, it consists of displaying among a helical referential, interacting with its period value to detect alignments (see Figure 12) and modifying its representation to access other attributes. To be able to acquire metrics about the association we also need the users to perform the abstract operation of selection which consists of coloring the pulses in the reference tool and sending them on another view in HeloVis.

Apparatus - The evaluation was performed on two different devices, a desktop to access the reference system and an immersive headset to access HeloVis. The desktop includes a mouse, a keyboard and a 1920x1080 pixels size screen of 24 inches. The immersive headset is an HTC Vive that comes with two controllers for interaction purposes. The evaluation was performed on two different devices: a standard computer on which the reference system was used, and a more powerful laptop, coupled with a VR immersive

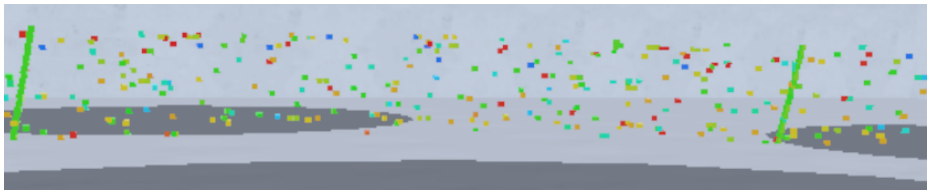


Fig. 12. Alignments observable on HeloVis that indicate the presence of a radar signal

headset to use HeloVis. The standard computer was composed of a central unit, a mouse, a keyboard and a 24-inch screen with a resolution of 1920x1080 pixels. The main unit had 8 GB of RAM, an *Intel Core i7* processor and used the operating system *Windows 7* (64-bit). The laptop had been equipped with an immersive helmet *HTC Vive*, with two controllers and two position sensors. It had 32 GB of RAM, an *Intel Core i7* processor, an *Nvidia 980M* graphics card0, and the operating system *Windows 8* (64-bit). The VR immersive headset had a resolution of 2160x1200 pixels that provided a horizontal field of view of 110 degrees.

Participants - We asked among the military force in charge of SIGINT for potential users of radar signal analysis tools. We succeeded to recruit 11 participants (including former experts working now as engineers). They were aged between 34 and 52 years old and the females were under-represented (1/11) but this reflects the reality of SIGINT. Half of the participants were military people (5/11) and half of them were civilian engineers building military solutions (6/11). Most of them worked in the field of SIGINT (8/11) and the remaining ones worked in the field of electronic warfare which is related to SIGINT. Half of them were full experts, able to associate even the most complex radar signals (6/11) and half of them had complete knowledge and understanding of the field but they had no experience enough to associate every radar signals (5/11), irrespective of their profession. Finally, only a few of them had previous experience with immersive devices (3/11). These participants had the knowledge and the experience to perform a SIGINT analysis and as such, were representative of the targeted users of HeloVis.

Data - The dataset used for the evaluation contains more than twenty-five different radar signals. We used the same dataset on both systems to remove the variation factor induced by the use of different datasets. Thus, participants evaluated the second system on a known dataset. We reduced the bias that this implies by the fact that we mixed the evaluation order of the systems, and we believe that this bias was also reduced by the fact that the representation metaphors on both systems were so different that it was difficult to relate from a system to another. Besides, we identified that there was no

significant difference in each metrics according to the evaluating order.After a pilot study, we discovered that associating the entire dataset was requiring a large amount of time, much more than the available time of the participants. As we did not want to constrain the evaluation by time,we updated the guidelines to reduce the number of radar signals that each user had to associate. We asked them to associate radar signals among a specific frequency range, which is a frequent request of SIGINT analysis, reducing the number of radar signals to associate to only eight. According to another pilot study, these signals were associable at most in less than 20 minutes with both systems, allowing the evaluation to fit the participants' available time and avoiding issues raised by too long usage of immersive headsets. The dataset and as such the eight radar signals used for the evaluation were real data, provided by SIGINT operators. With the help of a former military operator, currently working to build SIGINT systems, we ranked the signals according to their complexity in three categories: easy, medium and difficult. These categories were attributed by taking into account the variation of their parameters (PW, frequency, PRP, level shape, and pattern length) and their quality (number of missing value). Among the eight chosen signals one was categorized as easy, four as medium and three as difficult. We also ranked the signals among the similarity of their characteristics (PW, frequency, PRP, and level) according to the same categories. Among the eight chosen signals, two were categorized as easy, two as medium and four as difficult.

Synthesis - During the evaluation, we asked the users to perform the same association task with two different tools, the reference tool currently used by SIGINT operators on a desktop and HeloVis on an immersive device. This evaluation provided us with objective metrics to confront our hypothesis and subjective metrics to give us insight about user experience. The overall evaluation lasted around one hour and was performed thanks to eleven SIGINT experts.

6 Results

Thanks to the extracted metrics we got insight on the quantity and the quality of the association and also on the impact of the complexity of the radar signal, the user expertise, and the system's order. We extracted this from the recorded metrics considering the participants, their expertise, the radar signals, their complexity and their similarity as independent variables and considering the number of selection, the noise percentage and the selection percentage as dependent variables.

Preprocessing - While extracting metrics from association logs we discovered that some selections (5/96) contained less than 60 pulses for a signal average size of 300.

Such small selections, that only happened on the reference tool, were the consequence of some participants' strategies. While refining their selection, these participants used another color instead of removing the outliers such as the outliers stayed colored and were considered as a selection. As these selections were not the intent of the participants we removed them from the analysis.

Quantity of the association - To validate H1 we compare the number of associations performed on each system. The participants associated 53 signals with HeloVis and 43 with the reference tool on 88 possible selections (11 participants x 8 radar signals). If we look at this difference according to the participants or at the associated signals we observe that HeloVis also seems to make possible a better association quantity independently from the signals but not from the participants as illustrated in table 1.

	by participant		by signal	
	means	sd	means	sd
difference	0.91 on 8	1.496	1.25 on 11	1.219

Table 1. The difference in the number of associations on the reference tool and on HeloVis

Regarding the size of the sample, we did not go further into the analysis. Therefore, we cannot validate H1 but we have strong insights about its validity. In parallel, this highlights the fact that the selection and noise metrics are not in the same amount for both systems. From this point, we consider HeloVis and the reference tool measures as two different samples and do not take into consideration the number of selections anymore. We choose this approach because we are not able to identify the nature of the relation between the quantity and the quality of associations. We reduce this problematic by considering that quantity and quality are independents to be able to extract insight from this evaluation.

Quality of the association - To validate H2 we compare the reference tool and HeloVis based on the percentage of the noise of each association and the percentage of selection (see Figure 13). From this analysis, we observe that either dependently or independently the percentage of noise is lower for HeloVis and that the percentage of selection is higher for HeloVis. The percentage of noise selected points out the usability of the selection. A selection with too much noise can prohibit any use of processing starting with the use of the DTOA and can lead to mistake many characteristics of the radar. The percentage of the signal selected indicates the quality of knowledge about the selected signal. The more important the value is, the more the signal characteristics are known and less noise

stays in the dataset. Regarding the meaning behind the metrics of noise and selection, HeloVis seems to permit a better quality of the selection. To validate H2 we performed a Wilcoxon test as we previously discovered that our sample did not follow a normal distribution invalidating the use of the Student test. The results are presented in table 2).

	selection (%)		noise (%)	
	means	sd	means	sd
reference tool	83.4	8.5	27.8	9.2
HeloVis	98.5	0.9	0.4	0.3
p value	2.064e-6		0.0114	

Table 2. The result of the Wilcoxon test on the selection and noise percentage by associations on the reference tool and on HeloVis

From this result, we can conclude that using HeloVis improves the quality of the selection in terms of noise percentage and selection percentage, validating H2.

Quantity and quality of the association - From the results of the quality of the associations, it seems that we could establish a level of quality. If we take the level of 50% of the noise and consider only the associations containing less noise we can reevaluate the quantity of the selection (see Table 3).

	by participant		by signal	
	means	sd	means	sd
difference	1.45 on 8	1.43	2 on 11	1.25

Table 3. The difference of the number of associations containing less than 50% of noise on the reference tool and on HeloVis

This result indicates that taking into consideration a level of quality comforts better performance allowed by HeloVis compared to the reference tool independently from the participants and the associated signals.

Impact of the radar complexity - To extend the previous results, we analyzed the correlation between the noise and selection percentage and the complexity rate of each radar on both systems (see Figure 14). This highlights that whatever the signal complexity is, HeloVis will still allow a better quality association, but that the more complex the signal is, the worse quality the associations on the reference tool are, and that the complexity of the signal does not seem to impact the use of HeloVis.

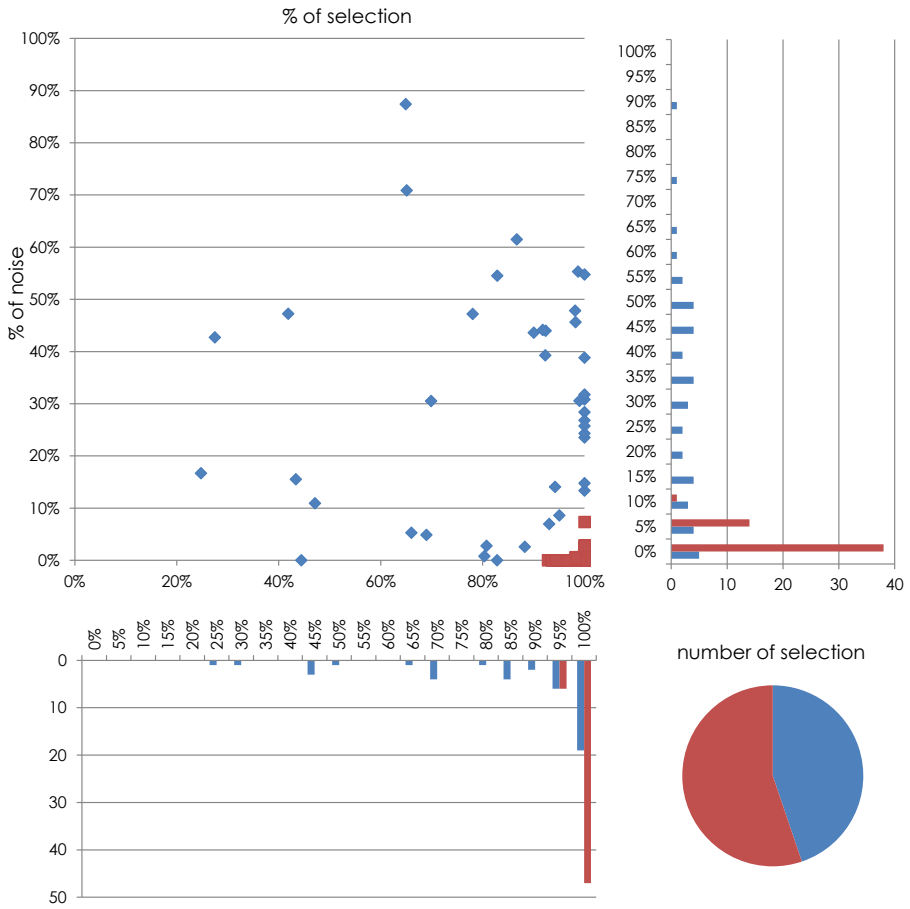


Fig. 13. Selection and noise percentage by associations on the reference tool (in blue) and on HeloVis (in red)

Regarding the signal similarity (see Figure 15) we observe, as for the complexity, that whatever the signal similarity is, HeloVis will still allow a better quality association, but that the more the signal is similar to others, the worse the quality of the associations on the reference tool is, and that the complexity and similarity of the signal do not seem to impact the use of HeloVis.

Impact of the user experience - We also tried to highlight the correlation between the previous metrics and user expertise (see Figure 16). This highlights again that the difference in terms of quality between the reference tool and HeloVis is not impacted

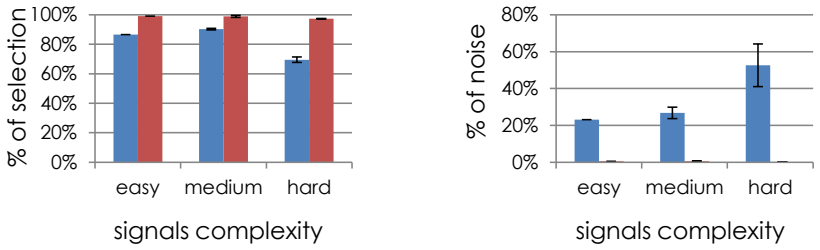


Fig. 14. Selection and noise percentage by signal complexity on the reference tool (in blue) and on HeloVis (in red)

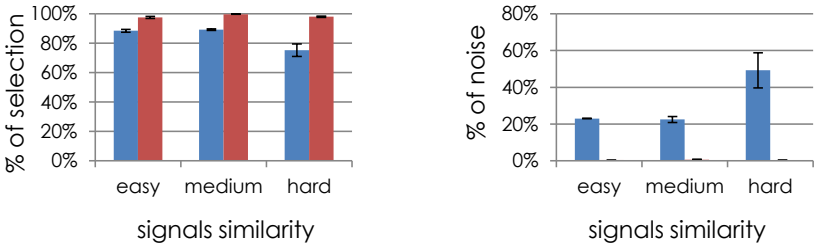


Fig. 15. Selection and noise percentage by signal similarity on the reference tool (in blue) and on HeloVis (in red)

by user expertise. Besides, there does not seem to be any correlation between selection, noise percentage, and user expertise.

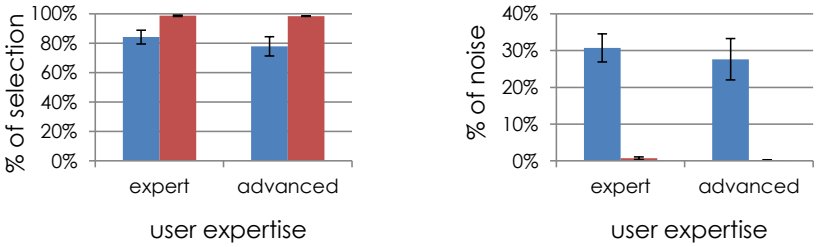


Fig. 16. Selection and noise percentage by participants user expertise on the reference tool (in blue) and on HeloVis (in red)

Impact of system order - To identify the impact of using the same dataset during the evaluation of the two systems we compared the difference between HeloVis and the reference tool results when HeloVis is used first and when the reference tool is used

first to identify any major difference. According to the values of selection and noise percentage, there is no major difference.

	diff. of selection %		diff. of noise %	
	means	sd	means	sd
using HeloVis at first	-3.6	3.5	0.7	0.4
using HeloVis at last	-7.7	4.7	0.7	0.3

Table 4. The difference of selection and noise percentage by participants between the reference tool and HeloVis

Observed Strategies - We observed different strategies on each tool. On the reference tool, users always started by selecting data according to their frequency values. Then they were refining their selection according to PW values, DTOA values or a level variation. To perform this second selection, they often used another color such that the remaining pulses of the first selection consisted of a small selection that will often be classified as error while processing the evaluation results. Because of noise, some users did not refine enough on these attributes such that the signals that have closed characteristics were selected as a unique signal. Such selections correspond to the second type of selection that will be categorized as error while processing the evaluation results because it often contains so many signals that none of them represent more than 50% of the selection. This explains the low selection rate of the reference tool. If such a selection is made in an operational context, it will not permit to characterize the expected signal or worse, it will lead to a false characterization. Users that refine their selection despite noise handle it in two different ways. Less experimented users selected a DTOA range containing the most recurrent value when full expert users removed only values lower than the most recurrent value and completed their selection on DTOA pulse by pulse. In the case of a missing pulse, the DTOA of the pulse was corresponding to a multiple of the PRP, and by removing values higher than the main DTOA value range, the user possibly removes these pulses despite they are a part of the signal. This behavior explains the lower selection percentage of the reference tool highlighted by the metrics. On HeloVis, all the users first started by exploring different period values to detect alignments. Then we observe that they behaved in two different manners. Some users selected the alignment as precisely as they could, helped with coloration or position on the radius, then they sent the pulses to another view and validated them. Some users were very selective on this selection and did not select data that had a singular color or radius position. The selections made with this behavior are often exempt from noise but have often a lower radar selection percentage. The other selection method was to largely select the

alignment, then send it to another view and clean it with another attribute encoding or by modifying the period and observing pulses that do not behave like others (see Figure 17). Again, if the user performs a too restrictive selection, the results will not be noised but the selection percentage will be lower. However, this behavior often makes the user realize that pulses with singular values are still part of the radar but with an error on one of its attributes. We observe that this last method is more and more common through the use of HeloVis.

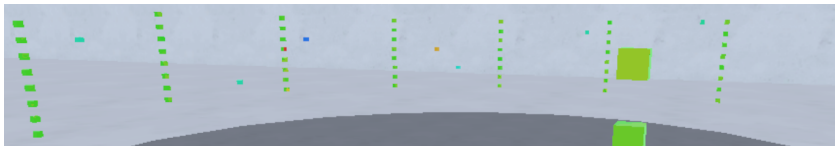


Fig. 17. Period set to correspond to a multiple of the current signal to highlight noise

Subjective results - During the evaluation, we have recorded subjective metrics by noting users' comments and asking them to fulfill a SUS questionnaire [5]. The reference tool was considered as badly designed in a way that complicates the interaction, while HeloVis was described as very easy to use and some users were even surprised by what they could accomplish compared to what they expected from an immersive 3D tool. Some participants, despite an enjoyable start, were sick during the use of HeloVis and did not have an overall pleasant experience. Regarding the performance provided by HeloVis, some participants expressed the fact that it allows them to detect complex radar signals hardly detectable on the reference tool but also said that there may be some complex radar signals more easily detectable on the reference tool. The synthesis of all these comments toward usability was transcribed in the SUS questionnaire results: 50 for the reference tool and 74 for HeloVis. As we do not want to evaluate the interaction we will not compare these results, however, it indicates that HeloVis was positively accepted by SIGINT experts who represent an important part of the SIGINT community.

7 Conclusion

We have identified that HeloVis allows a better quantity of association, however, we were not able to validate its significance. Concerning the quality of the association, we have validated the fact that HeloVis permits to perform a better quality association. We have observed different strategies that tend to highlight that the reference tool, contrarily

to HeloVis, does not allow the perception of some radars. Finally, we noted from the subjective comments that the users were afraid to miss information with both tools.

We built this evaluation driven by the validity of its results but despite all our efforts, there are still some questionable matters. Most of them are explicable by the fact that we wanted to perform this evaluation in real conditions. We aimed for a generic point of view such that we conducted an exploratory evaluation. This explains the fact that we did not decouple the measure of the dependent variables raising the question of the relationship between quantity and quality of the association. That also explains that we choose the currently used tool to make the comparison and not the 2D translation of HeloVis or a more recent metaphor. This raised some issues as it brings an interaction disparity that we handle by removing any interaction bias from the evaluation. Now we plan to evaluate HeloVis in more constrained experimental conditions. We will evaluate each hypothesis independently, with a more significant amount of participants and with more control over other independence variables such as the complexities of the radar signals. As this evaluation would target the evaluation of perception performance we will not necessarily conduct it on expert users at first, to access more participants.

8 Reference to the official publication

Springer Nature Switzerland AG 2019

P. Bourdot et al. (Eds.): EuroVR 2019, LNCS 11883, pp. 1–18, 2019.

The final authenticated version is available online at:

https://doi.org/10.1007/978-3-030-31908-3_12

References

1. Arsenault, R., Ware, C., Plumlee, M., Martin, S., Whitcomb, L.L., Wiley, D., Gross, T., Bilgili, A.: A system for visualizing time varying oceanographic 3D data. Oceans'04 MTS/IEEE Techno-Ocean'04 (IEEE Cat. No.04CH37600) **2**, 1–5 (2004)
2. Bach, B., Dachselt, R., Carpendale, S., Dwyer, T., Collins, C., Lee, B.: Immersive analytics: Exploring future interaction and visualization technologies for data analytics. In: Proceedings of the 2016 ACM on Interactive Surfaces and Spaces (ISS'16). pp. 529–533. ACM (NewYork, NY, USA, 2016) (2016)
3. Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps. ESRI Press (2011), `https://books.google.fr/books?id=X5caQwAACAAJ`
4. Brath, R.: 3D InfoVis is here to stay: Deal with it. In: 2014 IEEE VIS International Workshop on 3DVis, 3DVis 2014. pp. 25–31. IEEE, Paris, France (2015)

5. Brooke, J.: SUS-A quick and dirty usability scale. Usability evaluation in industry **189**(194), 4–7 (1996)
6. Cantu, A., Duval, T., Grisvard, O., Coppin, G.: HeloVis: a Helical Visualization for SIGINT Analysis Using 3D Immersion. In: 11th IEEE Pacific Visualization Symposium (PacificVis 2018). pp. 175–179. Kobe, Japan (Apr 2018)
7. Chandler, T., Cordeil, M., Czauderna, T., Dwyer, T., Glowacki, J., Goncu, C., Klapperstück, M., Klein, K., Marriott, K., Schreiber, F., Wilson, E.: Immersive Analytics. In: Proceedings of Big Data Visual Analytics (BDVA). pp. 1–8 (09 2015)
8. Cordeil, M., Cunningham, A., Dwyer, T., Thomas, B., Marriott, K.: ImAxes: Immersive Axes as Embodied Affordances for Interactive Multivariate Data Visualisation. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. pp. 71–83 (10 2017)
9. Donalek, C., Djorgovski, S.G., Cioc, A., Wang, A., Zhang, J., Lawler, E., Yeh, S., Mahabal, A., Graham, M., Drake, A., Davidoff, S., Norris, J.S., Longo, G.: Immersive and Collaborative Data Visualization Using Virtual Reality Platforms. 2014 IEEE International Conference on Big Data Immersive pp. 609–614 (2014)
10. Dubel, S., Rohlig, M., Schumann, H., Trapp, M.: 2D and 3D presentation of spatial data: A systematic review. In: 2014 IEEE VIS International Workshop on 3DVis, 3DVis 2014. pp. 11–18. IEEE, Paris, France (2015)
11. Gautier, J., Davoine, P.A., Cunty, C.: Helical time representation to visualize return-periods of spatio-temporal events. In: 19th AGILE International Conference on Geographic Information Science (2016)
12. Greffard, N., Picarougne, F., Kuntz, P.: Visual community detection: An evaluation of 2D, 3D perspective and 3D stereoscopic displays. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 7034 LNCS, pp. 215–225 (2012)
13. Kaur, K.: Designing virtual environments for usability. In: Human-Computer Interaction INTERACT'97. pp. 636–639. Springer (1997)
14. Keim, D.A.: Information visualization and visual data mining. IEEE transactions on Visualization and Computer Graphics **8**(1), 1–8 (2002)
15. Koffka, K.: Principles of Gestalt psychology, vol. 44. Routledge (2013)
16. Lam, H., Bertini, E., Isenberg, P., Plaisant, C., Carpendale, S.: Seven guiding scenarios for information visualization evaluation (2011)
17. Munzner, T.: A nested model for visualization design and validation. IEEE transactions on visualization and computer graphics **15**(6) (2009)
18. Munzner, T.: Rules of Thumb. In: Visualization analysis and design. CRC Press (2014)
19. Viau, C., McGuffin, M.J.: ConnectedCharts: Explicit Visualization of Relationships between Data Graphics. In: Bruckner, S., Miksch, S., Pfister, H. (eds.) Eurographics Conference on Visualization. vol. 31, pp. 1285–1294. Blackwell Publishing (jun 2012)

20. Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J.R., van der Helm, P.A., van Leeuwen, C.: A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. Psychological bulletin **138**(6), 1218 (2012)
21. Wang, B., Mueller, K.: Does 3D really make sense for visual cluster analysis? Yes! In: 2014 IEEE VIS International Workshop on 3DVis, 3DVis 2014. pp. 37–44. IEEE, Paris, France
22. Zade, A.A.T., Pezeshk, A.M.: Pulse Repetition Interval Detection using Statistical Modeling. In: Proceedings of the 8th International Conference on Signal Processing Systems. pp. 100–104. ACM (2016)