THE EFFECT OF SPATIALIZATION IN A DATA SONIFICATION EXPLORATION TASK

Brian FG Katz(1)*, Emmanuel Rio*(2)*, Lorenzo Picinali*(1)*, Olivier Warusfel*(2)

(1) LIMSI-CNRS
BP 133, 91403 Orsay, France
Brian.Katz@limsi.fr

(2) IRCAM
1, place Igor-Stravinsky, 75004 Paris, France
Olivier.Warusfel@ircam.fr

ABSTRACT

This study presents an exploration task using interactive sonification to compare different sonification mapping concepts. Based on the real application of protein-protein docking within the CoRSAIRe project (« *Combinaisons de Rendus Sensori-moteurs pour l'Analyse Immersive de Résultats* », or Combination of sensori-motor rendering for the immersive analysis of results), an abstraction of the task was developed which simulates the basic concepts involved. Two conditions were evaluated, the inclusion or absence of spatialized coherent rendering of the sonification output. The position of the sonification was determined by the user's orientation sensor used for the exploration task. Results showed no significant benefit in the spatialized condition, and for some examples the non-spatialized condition resulted in better performance. This test is the first in a series of studies using this test platform.

1. INTRODUCTION

The use of spatial audio in virtual reality applications is becoming more and more common. However, the use of 3D audio in data sonification applications is far less common. This paper presents a preliminary study on the use of 3D audio for interactive sonification, in order to assess its effect in a given context. A spherical topology surface exploration task is presented as an abstraction to an actual application task under question involving interactive assistance for protein docking simulations. Previous studies have used molecular models as reference tasks for virtual navigation, highlighting the benefit of multi-modal redundant information.[1] In this work, Gröhn *et al.* showed that audio-visual rendering was more efficient for the user's navigation task than either modality alone. The current study considers the effect of spatial audio in an auditory-haptic exploration task. The user task is generalized by to a virtual exploration, using a pointing and tracking device, of a topological function mapped on the surface of a sphere and sonified. The task is to find the maximum of the function, equivalent to the highest frequency according to the chosen sonification mapping scheme. The experience is repeated with and without the use of sound spatialization techniques.

2. THE CORSAIRE CONTEXT

The goal of the CoRSAIRe project is to develop new ways of interacting with large or complex digital worlds.[2] The project aims at significantly enhancing currently existing interfaces by introducing multiple sensori-motor channels, so that the user will be able to see, hear, and touch the data itself (or objects derived from the data), thus redefining conventional interaction mechanisms. Such a research effort involves a paradigm shift, because many well-established visualization-oriented software packages exist to analyze the large spectrum of available data types: thus, creating a completely innovative sensori-motor interface would seem a daunting task. This project focuses on two well-defined application areas (Fluid Mechanics and Bio-informatics) with which collaborations are put into place with end-user partners.

A major facet of the project regards how the scientist is able to explore, analyze, and understand large complex datasets. The complexity of the representation in the two disciplines mentioned lies on the one hand in the number of correlated variables to analyze simultaneously and on the other hand, in the many parameters the user must control to successfully drive his analysis. In the CoRSAIRe project, one identified goal is to allow the user to interact with the virtual data in real-time on the evolution of the studied phenomena (to correct, target, modify, annotate...), according to information that he/she perceives.

In contrast to the visual or haptic modalities, the sonification of data in general can bring a more global comprehension of the information through full 3D reproduction, while also providing an improved representation of the aspects of temporal dependence.

2.1. Application overview

The two main scientific application fields of this project, which are Computational Fluid Dynamics and Molecular Bio-informatics, are directly concerned with the crucial problem of generation and processing of large volumes of complex data. These data are generally the results of experiments, simulations, or numerical predictions. For example, in Bio-informatics, the scenes (DNA molecules) can be made up of several million atoms whose interactions are dynamically controlled by the research operator. In Fluid Mechanics, classical examples study the characteristics of a non stationary three-dimensional flow for which a multitude of parameters are calculated at every point. The work presented here concerns the Bio-informatics application context.

2.2. Bio-informatics application

The main objective of the Bio-informatics application is to propose, implement, and evaluate a new approach for the protein-protein docking problem (prediction of protein-protein association) based on the precept that such complex simulations could be more effective if user-driven and the biologist is considered in the docking process loop. Human perceptual skills are a valuable asset for pattern analysis, recognition, and mining tasks. This is also true in the decision-making phase providing in consequence an increase in performance, both in time and in the quality of the docking computation. Advanced virtual reality multimodal interfaces combining gestures, speech recognition, audio synthesis and haptics enable the biologist to act naturally in the simulation process and to skim through the set of possibilities and make a fast selection of biologically interesting protein complexes. User interaction with the simulation is also carefully recorded in the form of annotations (including vocal), so that all the useful information relative to the history of the docking procedure is recorded.

The majority of current biological research focuses on the relationship between structures and functions. But, because the majority of proteins provide their function in the form of complexes, it is the structure of the complex which is actually the object of interest. Various works have been carried out concerning immersive docking: STALK [3], ARCdocking [4], etc. Nevertheless the obtained results are still too specific or insufficient. These studies consider the user more as an observer rather than an actor. The approach presented here is more human-centered and addresses the design of user-oriented multimodal immersive docking systems (Figure 1).

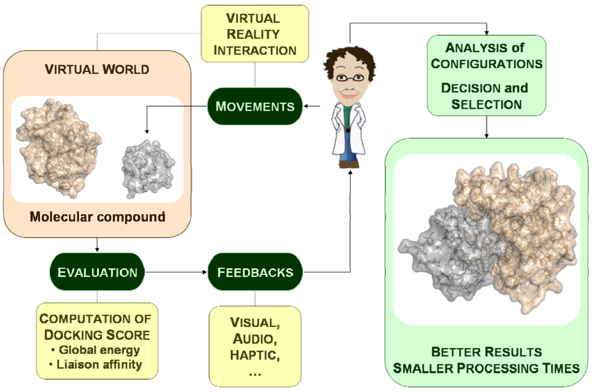


Figure 1. *Immersive docking prototype principle.*

Protein-protein docking aims at determining how two protein structures fit together, without the need for experimental measurement. As a prerequisite, the three-dimensional molecular structures of both proteins must have been determined experimentally. Each relative orientation of the two binding proteins in the complex yields non-covalent interaction forces, whose strength are then estimated computationally using so-called scoring functions. For a given protein pair, if they actually bind *in vivo*, the spatial configuration which they adopt in their final complex state should correspond to an extreme of the scoring functions.

In the context of the CoRSAIRe project, the user, a biology expert, interacts through 3D trackers with the virtual position of the proteins to find the possible bound states. The Virtual Reality system is able to inform him about the value of the scoring functions, so that he may infer that a given position yields a valid protein configuration.

In the context of multimodal rendering, the scoring function may be given to the user through an auditory display. The scoring function is a global scalar function which depends on the relative position of both proteins. Thus, the scoring function depends on the 6 scalar parameters (x, y, z, yaw, pitch, roll) that characterize the position of one protein relative to the other.

A simple sonification approach consists in mapping the scoring function to the frequency of a digitally synthesized sound. The frequency scale of the sound is driven by the scoring scale: finding a maximum scoring corresponds to finding the highest frequency. The present study aims at evaluating the interest of complementary sound parameters (timbre, spatialization) in order to improve task performance (i.e., finding the maximum value).

3. ABSTRACTION

The actual application is a rather complex combination of geometry, chemical interactions, physical interactions, and a large number of degrees of freedom. In order to perform the evaluation on well identified factors, the problem has been reduced to an abstraction with certain limitations. The complexity of the problem may be increased in subsequent iterations of the platform.

As a first step, the number of degrees of freedom is reduced. The 3D case where yaw, pitch, and roll are taken into account was deemed too complicated for this pilot study. While the final multimodal integrated task contains visual cues relating to the protein geometries as well as auditory and haptic cues, the current study, there are no visual cues, and hence the full 6 degree of freedom task was not considered. If one discounts orientation information and then considers that the scoring function in question does not depend on the distance between the proteins, then the problem can be handled more simply as a scoring function with two degrees of freedom, being the position of one protein relative to the other at a fixed distance in spherical coordinates.

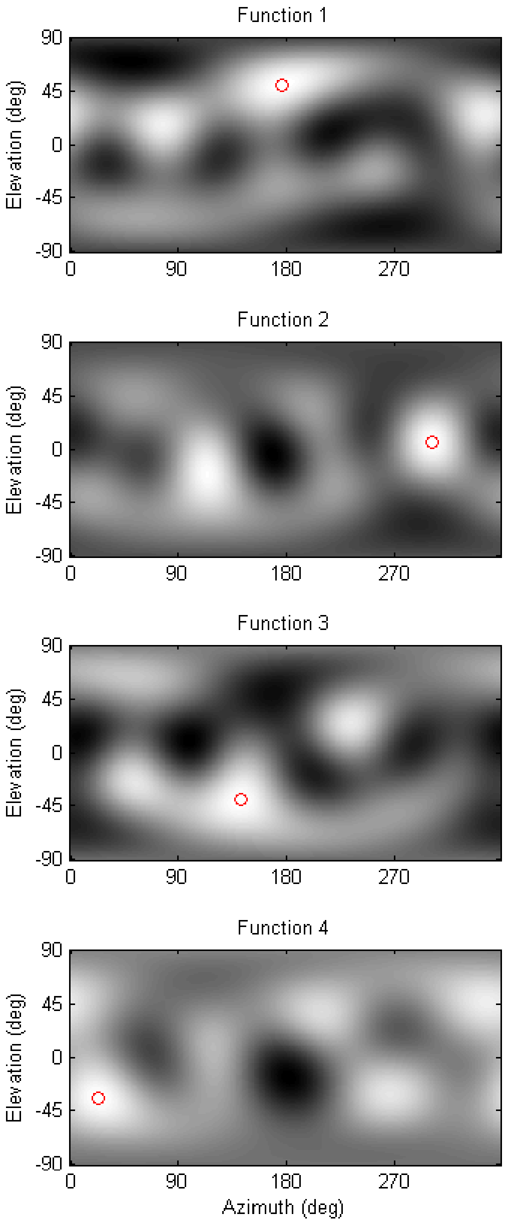
Second, for this study the scoring function is defined precisely but in an arbitrary way. In the real case of protein-protein docking experiment, the scoring function is known only at the position corresponding to user manipulations. No global information is available. Instead of using a scoring function computed from actual biological data, an abstract function is forged, whose spatial characteristics can be designed appropriately, and whose spatial variance and resolution are well defined. Since this abstract function must be computed fast, it is predefined using harmonic functions, such as Fourier series in the one-dimensional case, or spherical harmonics in the two-dimensional case (the position is therefore reduced to orientation on a sphere). Thus we obtain a metaphor of the actual problem, which may become more and more complex afterwards.

Finally, we consider only one complimentary factor: spatialization. This can be easily and directly mapped to the positional information needed to compute the abstract scoring function. The resulting hypothesis under study is the following: does providing an auditory cue of the exploration position aid the user in accomplishing the task of finding the absolute maximum of the scoring function.

3.1. Definition of the abstract scoring function

An infinite number of functions may be created using spherical harmonics. We thus reduce the maximal harmonic order used in the functional representation, which allows us to limit the spatial resolution of the function to be explored (no sharp peak or notch). The harmonic order has been set to 4 after preliminary tests, resulting in an approximate spatial resolution of 25°.

Once the maximal order has been set, the spherical harmonics coefficients are chosen randomly according to a Gaussian distribution. This should provide a complex variation which varies smoothly. Four different random functions were generated as samples for this study (see Figure 2).


Function 1

Function 2

Function 3

Function 4

Figure 2. *The four abstract score functions generated using 4th order spherical harmonics. Function maxima locations are indicated:* ○.

This abstract spatial function can be considered as a sort of topology in spherical coordinates. An example of this topology mapping is shown in Figure 3 for "Function 1" used in this study. It can be seen that there are several local peaks, and one predominant peak. The analogy to the actual application would be that the peak of the topology would represent the orientation of the two proteins with the highest scoring function, and therefore the most compatible docking position.

3.2. Task description overview

To create an abstraction of the protein-protein docking task, the abstract scoring function described above must be explored by the participant, in order to find the correct position where the maximum value is obtained. This task can be considered as a manipulation of the orientation the function, or the exploration of the surface depending if you take the point of view from the inside or outside of the function. As the final project is based on an immersive system for data exploration, the task here has been defined by placing the user inside the topology which they are exploring. This places the user inside a virtual sphere which is to be explored, in order to find the location on the sphere with the highest score value. The exploration is achieved using an orientation sensor. The concept of this task is shown in Figure 4. The value of the score function at the pointed direction is used as the sonification variable (detailed in section 3.3). To better understand the projection of the exploration action to the score function, the example of the exploration path in Figure 4 is projected onto the scoring function and shown in Figure 5.

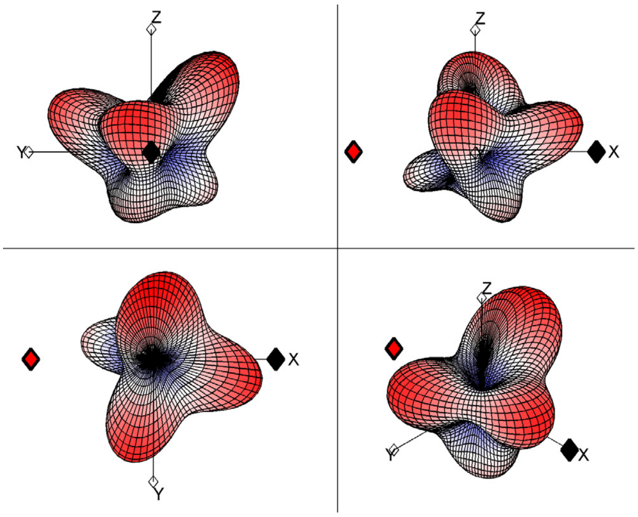


Figure 3. *Spatial topology defined by "Function 1", different points of view shown.*

The maximum of the scoring function could be at any orientation projected on the sphere. As there can be several local maxima, and the subject must find the global maximum, they need to be aware of the state of their exploration, in order to make be confident that all the possibilities have been explored. To accomplish the communication of this information a progress counter was used. In order to avoid a second sonification approach competing with the sonification of the score function, a voice was used. Every 5 seconds the participant was informed by an automatic message about the percentage of the sphere explored up to that moment. This is calculated by taking into account the angular resolution of the spherical harmonic order. A corresponding radius is calculated and attributed to the position indicated. A calculation is then performed to determine the percentage explored of the surface area of the sphere.

Participants were instructed to explorer the sphere sufficiently to be confident that they had found the corresponding maximum. The response judgment was entered through the use of a foot pedal, registering the current orientation of the sensor.

For each function, participant had a maximum time of 2 minutes in order to validate a specific point as the maximum: 15 seconds before the end of the time limit, an automatic message provided an alert that time is almost out.

The test was preceded by a training session accomplished on random functions of lower order, with no time limits, in order for the participants to become acquainted with the entire apparatus.

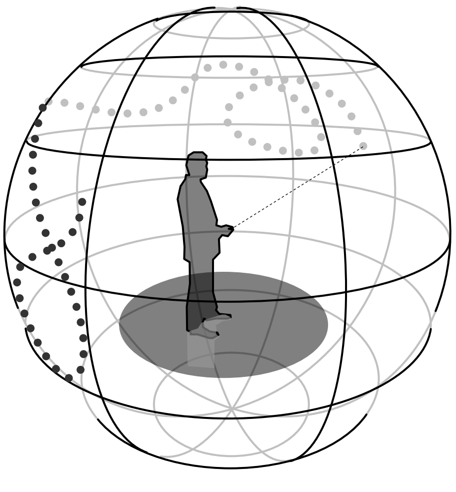


Figure 4. *Representation of the exploration scenario with the subject at the center of the virtual sphere. The dashed line represents the direction indicated by the orientation sensor; the dotted line represents the exploration path.*

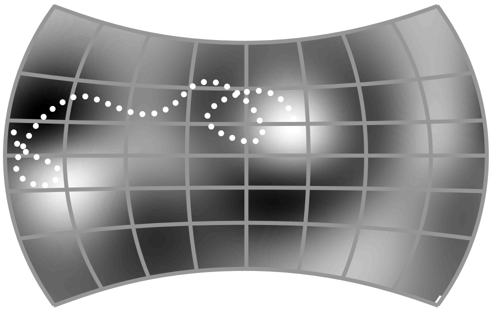


Figure 5. *Example of the exploration path on the scoring function shown in Figure 4.*

The test itself consisted in the exploration of the four different functions (see Figure 2) repeated twice and presented in a random order. For each participant, the entire test was carried out twice, once with and once without sound spatialization (condition order was random between participants). A significant pause was required between the two conditions.

Each participant had previously passed an identical test in a lit room. It was observed that some participant were using visual references within the room. The entire test was then repeated with the lights off, the results of which are presented in this study. As such, one can assume that all participants were somewhat experienced in the task and that there should be no learning effect.

3.3. Sonification: from score to sound displayed

In order to inform the user of the value of the score function at any given position a simple sonification mapping was used. The fundamental frequency f_1 of the sound depends directly on the value of the abstract scoring function. The mapping between frequency and function's value was given by $f_1 = f_C \cdot 2^{S \cdot F(\theta,\varphi)}$, where $F(\theta,\varphi)$ is the value of the function, S is a slope in octave/unit, and $f_C = 1000$ Hz is the center frequency which corresponds to zeros of $F(\theta,\varphi)$.

For the 4 defined functions, the maximum values for f_1 are shown in Table 1, as well as the frequency of the second largest maxima. This degree of variation was chosen in the hopes of ensuring that participants would not simply search for a given maximum frequency but would need to explore the range of value possible for each function.

	1st maximum(f_1)	2nd maxima(f_1)	f_2/f_1
Function 1	8.9	6.9	0.78
Function 2	11.1	10.4	0.94
Function 3	6.0	4.5	0.75
Function 4	4.8	3.6	0.75

Table 1. Values of f_1 and f_2 (kHz) for the different test functions.

Regarding both sonification and sound spatialization, the sound design had to respect the following constraints. For sonification, since we adopted a simple relation between the variation of the spatial function and sound frequency, the fundamental frequency had to be easily perceived. For a precise spatialization, the sound had to feature a large and dense spectrum.

In order to satisfy the second constraint, it was conceived to create an impulsive sound of short duration. The test sound was produced by additive synthesis of 20 inharmonic partials. Each partial was implemented by an exponentially damped oscillator. Attack times of all partials was set to 1 ms (milliseconds). The relative amplitude of the partials followed a roll-off of -3 dB / octave, whereas decay times ranged from 3 ms for the highest partial and 250 ms for the fundamental frequency. The partials' frequencies were computed according to the relationship: $f_k = \alpha_k \cdot f_1$, where $1 \le k \le N$, $N = 20$, and $\alpha_k = 1 + (k-1)/(N-1) \cdot (\alpha_N - 1)$. Using this mapping, the partials' frequencies are equally spaced from the fundamental frequency to the highest partial, but since α_N is chosen to differ from N, the partials are not multiples of f_1. One may notice that the frequency factor α_k does not change with f_1, so that the main spectral shape remains the same over all frequencies. Figure 6 shows the Power Spectral Density (PSD) of an example sound, for $f_1 = f_C = 1000$ Hz.

Both the broadband spectral content and the short attack time allow such a sound to be localized correctly. The result was an easily localizable sound whose fundamental frequency was easily variable and identifiable.

The sonification was played at a rate of 8 pulses/sec, or at intervals of 125 ms.

3.4. Spatialization: principles and implementation

The spatialization principle used during the test consisted in spatially rendering the sound in the direction pointed by the subject.

Two possible options were considered: a cube of 8 loudspeakers or headphones. Spatialization using headphones raises several issues, like the individualization of HRTFs and head tracking. These issues can be addressed by various means. But, one major difficulty in binaural audio rendering is the lack of data in the zone below the listener in the majority of measured HRTF datasets. This limitation would have introduced an asymmetry between the top and bottom hemispheres, and would have excluded a significant portion of the lower hemisphere. For this reason, binaural synthesis was not chosen for this study. An informal comparative evaluation between Ambisonic [5] and Vector Base Amplitude Panning (VBAP) [6] approaches showed that the VBAP method yielded less blur in sound localization for the given installation.

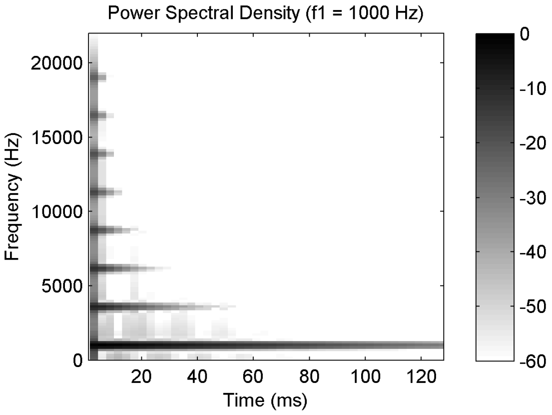


Figure 6. Power Spectral Density (PSD) of the sound, for $f_1 = f_C = 1000$ Hz.

Figure 7 shows the loudspeaker setup used for the test. However, applying directly the VBAP method to this setup causes each face of the cube to be divided in two right triangles. When the sound has to be spatialized at the center of a face, it is actually panned between two diagonal loudspeakers, thus causing an asymmetry in the perception. To avoid this problem, 6 loudspeakers are virtually added to the setup, at the center of each face of the cube, as depicted in Figure 8. In this configuration, the VBAP algorithm outputs 14 signals. Then, each of the 6 signals corresponding to the virtual loudspeakers is equally distributed among the 4 loudspeakers located at the face corners. This distribution is done with respect to signal energy.

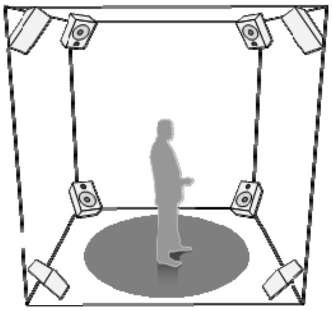


Figure 7. *Loudspeaker configuration and subject position in acoustically damped room.*

For the non-spatialized condition, the sonification sound was played over all 8 loudspeakers at equal level. This produced an ambient omnidirectional sound source. The percentage counter voice was always rendered as an omnidirectional sound source in the same manner.

Note that the spatial information is technically redundant in principle as the user physically controls the orientation sensor and the spatialized sound source is located at the currently indicated orientation.

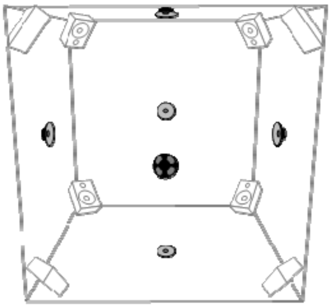


Figure 8. *Loudspeaker configuration used for VBAP rendering showing real and virtual speaker positions.*

4. RESULTS

4.1. Test participants

A total of 20 paid participants partook in the study. Initial analysis of the results showed that several did not clearly understand the task, exploring only a limited part of the sphere, validating the first local maximum they found in their exploration path. For this reason, it was decided to eliminate from the analysis the data of all the participants that explored less than 70% of the spheres surface (calculated as the mean value between all the 8 functions): in this specific case, 4 subjects were removed from the analysis of the results. Therefore, results presented here are for 16 participants.

4.2. Global results

A general analysis of the results, comparing overall performance criteria between non-spatialized and spatialized conditions was performed. A number of criteria were evaluated, with the mean results shown in Table 2. *Angular error* was computed as the error (in degrees) between the position of the maximum and the reported value. *Time* was the exploration time for each function. Two criteria were used to evaluate the reported value. The *Absolute score* error compared the reported score to the maximum of the function. The *Relative score* error compared the reported value to the maximum of the function on the portion explored. Finally, *Explored* is the degree to which the function was explored.

	Non-spatialized	Spatialized
Angular error (deg)	40	46
Time (sec)	54	52
Absolute score (%)	93	93
Relative score (%)	94	94
Explored (%)	92	91

Table 2. Mean values for different evaluation criteria.

These results show little overall difference between the two test conditions. None of the differences are significant. There is some concern though regarding the large angular errors.

4.3. Analysis of the feedbacks

Participants, at the end of the task, were asked to fill out a questionnaire concerning their method used for exploration and on eventual observation/suggestions they could make on the test itself. Additional information was gathered while observing the participants during the test.

Interesting observations were made on the mobility of participants during the test: the majority (11 subjects) held the tracker in front, moving their arm up and down and rotating their entire body. In contrast, 5 participants used wrist and arm motion, without moving their whole body. Finally, 5 subjects began with full body rotation and, after a several functions reduced movements to the arm and the wrist.

In terms of the exploration techniques described by the participants, in the majority of the cases (80%) there was first a global sphere exploration followed by the identification of local maxima and the comparison between them. One subject performed a final complete exploration after finding the maximum. In terms of the first global exploration, the techniques used were mainly the following:

- Systematic exploration from up to down of sectors of the sphere.
- Exploration of first one half of the sphere, then the other half, moving the arm up and down.
- Going up and down while rotating in one direction.
- Exploration of the sphere in horizontal circles, going from up to down.
- Rapid exploration of the sphere in sectors: when a maximum was detected, a more careful exploration was performed for that specific area.
- Vertical exploration of the sphere in quarters: once found the local maximum, the topology was followed until the real maximum was found.
- Exploration in 4 vertical circles, 45° one from the other horizontally.

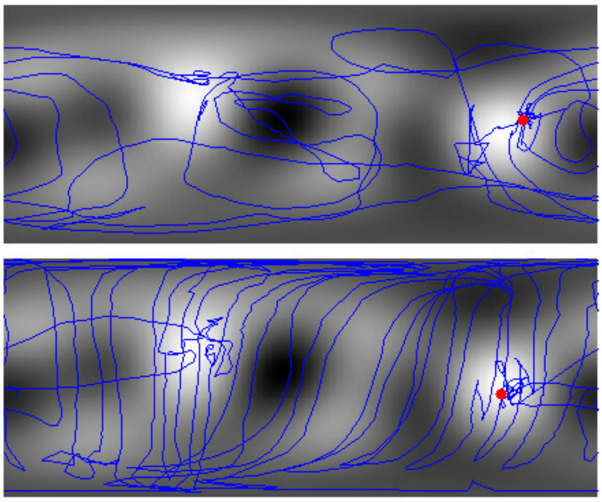


Figure 9. *Example of two exploration sessions for two different participants.*

Two example exploration patterns for two different participants are shown in Figure 9 for the same function. While both achieved relatively the same result, the difference in the search pattern is notable. The first example shows rotational exploration of spherical zones. The second example shows a very methodical vertical progression exploration. In addition, there is clear evidence of the comparison between the two local maxima before the final selection is made.

Only two subjects seemed to have consciously understood the difference between the test performed with and without sound spatialization, but they couldn't say if hearing the sound coming from the pointed direction helped them in the completion of the task.

4.4. Detailed result inspection

An inspection of the response positions indicates that a major error was in the selection of local maxima rather than the true maxima of the functions, rather than errors in precise selection of the maximum. The responses for all trials for each function are shown in Figure 10. A comparison of these results to the maxima locations shown in Figure 2 highlights that while the true maxima are often found; secondary local maxima are also often selected.

An error of this type is not well suited to global statistics, such as angular error, as the distribution of responses is far from normal. As such, the mean *angular error results* in Table 2 should not to be considered as a proper representative analysis of the error distribution. A general statistic comparing the number of responses within 25° of the correct response results is found to be 71% and 68% for the non-spatialized and spatialized condition respectively. Further analysis reveals that for *Function 2*, the secondary maximum was selected more often than the true maximum for the spatialized condition. An overview of this analysis is shown with a histogram in Figure 11. The two maxima for *function 2* are roughly opposite each other in spherical coordinates, but neither are difficult to locate (*i.e.* not being directly overhead or below). In contrast, the sonified frequencies for these two maxima are relatively close, 11.1 kHz and 10.4 kHz, making the discrimination task more difficult. Why the spatialized condition exhibits more error in this context is unclear at this stage.

While further numerical synthesis of the results are not provided here, it can be observed that in general the clustering patterns for the spatialized condition are more concentrated that for the non-spatialized condition. This observation remains to be verified numerically.

5. CONCLUSIONS

This paper presents an abstraction task based on a real-time interactive multimodal exploration application. While the true task is highly complicated, an abstraction was created to limit the number of parameters to a 2 degree of freedom problem, suitable for testing certain sonification methods. High-order spherical harmonics were used to generate random spherical topologies which were then explored using an orientation sensor and a simple 1-channel sonification mapping scheme in order to find the maximum of the spatial function. Two conditions were compared, non-spatialized and spatialized rendering of the sonification output. This spatialized information was complimentary to the proprioceptive information by the user of the orientation sensor.

While no statistically significant difference was found, the framework of the test platform appears promising for such an abstraction task. Further tests will allow for an augmentation of the complexity of the sonification and exploration task in order to compare additional parameters, such as multiple sonification streams with different spatial locations, higher order functions allowing for more degrees of freedom to be included, and difference sonification mappings.

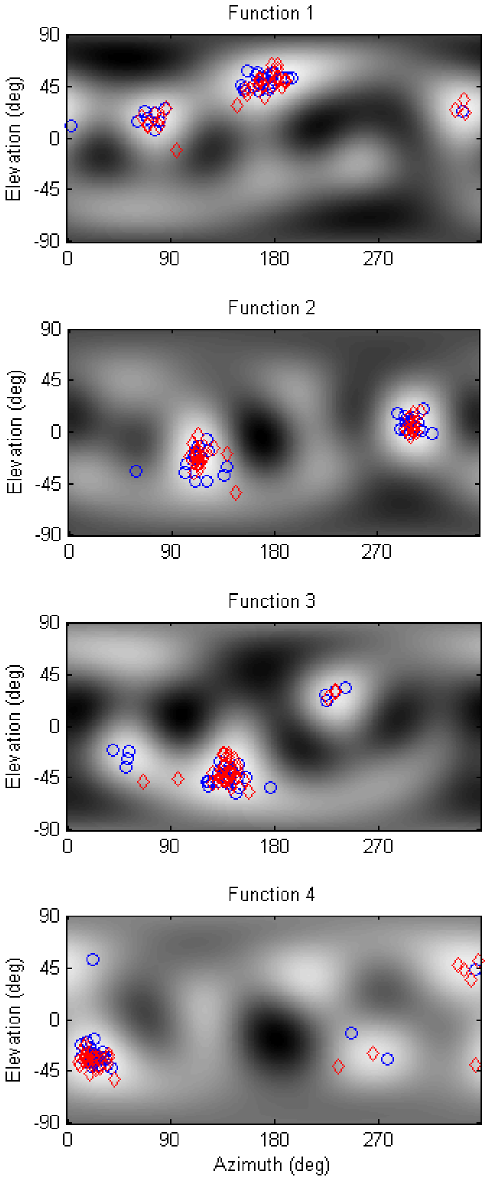


Figure 10. Responses for each function; ○ non-spatialized; ◊ spatialized.

6. AKNOWKLEDGENEMTS

CoRSAIRe is a 3 year research project supported by ANR (French National Research Agency) labeled in December 2005. Lead by LIMSI-CNRS (VR, Multimodal interaction, 3D audio, and Fluid Mechanics aspects), the CoRSAIRe consortium is also composed of three other research teams: IRCAM-CNRS (VR and 3D audio expertise), EBGM-INSERM (Bio-informatics application), ECI-Univ. Paris 5 (analysis of user needs and Ergonomics studies on VR applications), and HAPTION S.A. (Haptic interfaces).

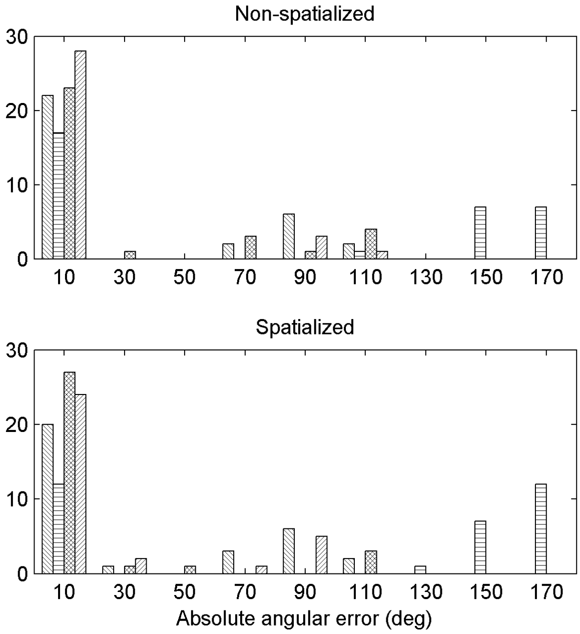


Figure 11. *Histogram of Absolute angular error for each test function for the non-spatialized and spatialized test conditions (20° bins).*

7. REFERENCES

[1] M. Gröhn, T. Lokki, and T. Takala. “Comparison of auditory, visual, and audiovisual navigation in a 3D space,” in *Proceedings of the 9th International Conference on Auditory Display*, Boston, USA, July 2003.

[2] B. FG Katz, O. Warusfel, P. Bourdot, and J-M. Vézien: “CoRSAIRe – Combination of Sensori-motor Rendering for the Immersive Analysis of Results,” in *Proceedings of the 2nd International Workshop on Interactive Sonification*. York, UK, February 2007.

[3] D. Levine, M. Facello, P. Hallstrom, G. Reeder, B. Walenz, and F. Stevens, “STALK: An Interactive Virtual Molecular Docking System,” *IEEE Comput. Sci. Eng.*, 4, 55-65, 1996.

[4] N. Ray, X. Cavin, J.-C. Paul, and B. Maigret, “Intersurf: dynamic interface between proteins,” *Journal of Molecular Graphics and Modelling*; Volume 23, Issue 4, Jan. 2005, p.347-354.

[5] M.A. Gerzon, 'General Metatheory of Auditory Localisation', Preprint 3306 of the 92nd Audio Engineering Society Convention, Vienna, Mar. 1992.

[6] V. Pulkki. Virtual source positioning using vector base amplitude panning. J. Audio Eng. Soc., 45(6):456–466, Jun. 1997.