International Cartographic Association
Association Cartographique Internationale

Designing a Multimodal Interface for Geospatial Data Visualization: An Application of Interactive Speech Recognition

Homeyra Mahmoudi [a,*], Silvana Philippi Camboim [a,b], Maria Antonia Brovelli [a]

[a] *Politecnico di Milano, homeiramahmoodi1994@gmail.com, maria.brovelli@polimi.it*
[b] *Universidade Federal do Parana, silvanacamboim@ufpr.br*

* Corresponding author

Keywords: Multimodal interface, Voice Speech technology, Digital Earth, NLP, GIS, User Experience

Abstract:

Geospatial data visualisation and interaction have been cartographic challenges from many perspectives. User experience in this field is tailored to various variables such as the level of detail, the sophistication of the required demonstration, the traditional visualisation paradigm, the volume of the data and the user's domain knowledge. Additionally, in an era of ubiquitous map consumption on various devices, including those in immersive experiences, the context of use stands out as an essential factor for map design. Understanding context involves grasping the interaction between the map user, the map use environment, the map use activity, and the map itself (Griffin et al., 2017). However, efforts have been made to design a user-friendly interface to improve navigation and interaction, but the geospatial data platforms available on demand are still not fully multimodal interfaces. The available platform's design, usability and accessibility are bound to the mouse, keyboard or touchscreen. Multimodal interfaces have been shown to significantly reduce the complexity of GIS interfaces (Fuhrmann et al., 2005), (Oviatt, 1996). Voice input can be considered one of the features that overcome the mentioned obstacles. Furthermore, it can allow users to control their interaction and selection while processing the data. Therefore, the integration of speech recognition technology with map applications can enable users to interact with map data and access information in a more natural and intuitive way without the need for complex user interfaces or manual input. (J. Austerjost, 2018).

The design of a comprehensible user interface (UI) that balances the naive user and the professional user also raises concerns since the users' domain knowledge may not coincide with one of the application's developers. Therefore, the need arises to propose multimodal interfaces based on cartographic concepts, maximising communication and interaction in multiple-use contexts. Complex analysis for geospatial data visualisation requires *a posteriori* knowledge, and the limitation of the available interface features might affect the process. On the other hand, new concepts such as Digital Earth indicate different perspectives. The Digital Earth encourages people to interact, interpret, and form the construction of scientific evidence unconcerned of their awareness and education, and it should be developed based on open access and media participation (Max Craglia, 2012). In addition, combining natural language with an application leads to a better user experience and enhanced data collection. Relying on these points, interactive speech recognition for visualising geospatial data is a practice for democratising access to digital geospatial data. It allows users to get involved with the application without special skills, and the collected experiences can be used to train an Artificial Intelligence model dealing with geospatial data visualisation.

This work presents the ongoing application implementation and development under a web application specialising in data visualisation called Bstreams (https://bstreams.io/). The platform assigned the traditional interface and tools for map visualisation, which advanced the investigation and assessment of the UI interactions. The available features are collected and categorised based on the domain concept and characteristics of the task. Each task is precisely contextualised based on the PlanGraph model. The model is influenced by the two main vital theories, the computational collaborative discourse theory (B.J. Grosz and C.L. Sidner, 1986) and the SharedPlan theory (B.J. Grosz and S. Kraus, 1996). The framework was inspired by the implemented approach on a collaborative GIS interface (GeoDialogue) (G Cai and H Wang, 2005). Since the core intrinsic of this application is tailored to Natural Language Processing, the procedure continued by collecting data through a survey about the lexical choices regarding the domain concept of geospatial data visualisation through a voice interface. Noteworthy, the considered language for this application is English (GB-US). The survey mainly focused on displaying some concepts through media (i.e. video and images) to the respondents to convey

Abstracts of the International Cartographic Association, 6, 150, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa.
https://doi.org/10.5194/ica-abs-6-150-2023

the task flow. Mapping through lexical choices benefits users by allowing them to interact with the application without reading the instructions.

The backbone architecture of the application is developed in the JavaScript and NodeJS frameworks. The Web Speech API (https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API) embedded into all browsers enables the interface to incorporate voice data into web apps. The infrastructure hosts by the AWS services such as S3 bucket and Lambda function to monitor and enhance the data flow. Digital Earth identifies the necessity of exploring and accessing data with no barriers. For the realisation of this matter, the interface includes a Geocoding API. The graphical user interface (GUI) contains a small chat box that keeps the chat history tracks, provides more information as responses for the users, and a button to enable and disable the microphone and the base map. The GUI is a bridge for conveying the comprehension of the cartographical and analytical tasks done through voice commands. For democratising geospatial data, the code of the application is available on GitHub, and the application is also accessible to the public on the BStreams platform. Figure 1 shows the architecture of the application.

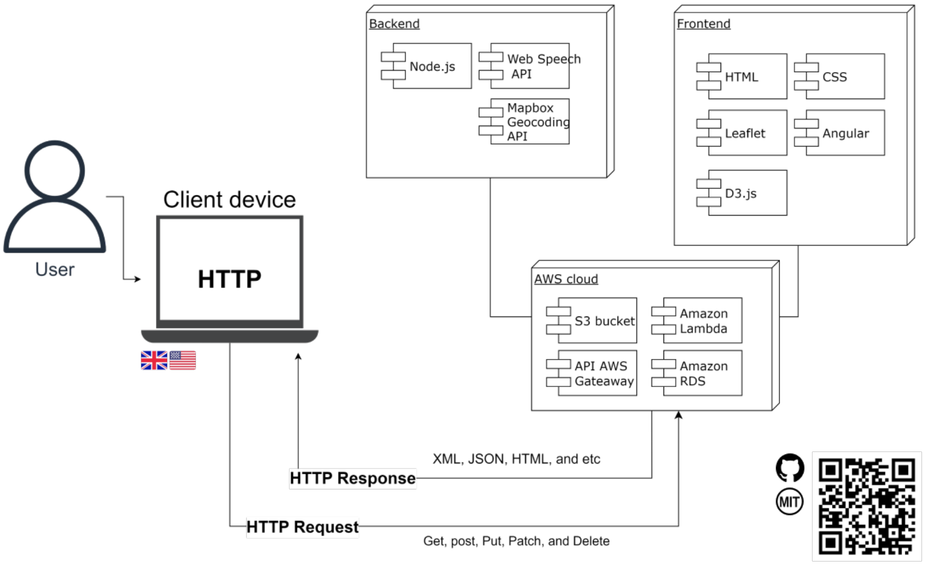

Figure 1. Schema of the architecture of the application

The expected results, followed by a practical test scenario, illustrate that a multimodal interface improves the user's collaboration even though the user might not have a deep knowledge of the discourse concept. Keeping track of the interactions, dialogues, and commands will improve an Artificial Intelligence model's structure and implementation process, which specialises in geospatial data visualisation and interpretation. This tool will increase the quality of education for students since it brings accessibility, efficiency, and accuracy during the lectures. Furthermore, the variety of the lexical choices' dataset will guarantee that users can communicate with the application without having domain knowledge, explore the map, and apply customisation to geospatial data with voice commands. Voice speech map applications can recognise the map navigation commands and implement actions from the interpreted commands. This tool upgrades the level of interaction among the agent and an interface which specialises in the geospatial concept and facilitates the barriers to open access data. Users can easily navigate through different locations, explore the maps, adjust the cartographical aspects of features such as base maps, markers and shapefiles and filter the shown geospatial data simply by saying the most common phrases and sentences.

References

Austerjost, J., Porr, M., Riedel, N., Geier, D., Becker, T., Scheper, T., Marquard, D., Lindner, P., Beutel, S., 2018. Introducing a Virtual Assistant to the Lab: A Voice User Interface for the Intuitive Control of Laboratory Instruments. SLAS Technology, Special Issue: The Internet of Things in the Life Sciences Laboratory 23, 476–482. https://doi.org/10.1177/2472630318788040

Cai, G., Wang, H., MacEachren, A.M., Fuhrmann, S., 2005. Natural conversational interfaces to geospatial databases. Transactions in GIS 9, 199–221. https://doi.org/10.1111/j.1467-9671.2005.00213.x

Craglia, M., de Bie, K., Jackson, D., Pesaresi, M., Remetey-Fülöpp, G., Wang, C., Annoni, A., Bian, L., Campbell, F., Ehlers, M., van Genderen, J., Goodchild, M., Guo, H., Lewis, A., Simpson, R., Skidmore, A., Woodgate, P., 2012.

https://doi.org/10.5194/ica-abs-6-150-2023 |

Digital Earth 2020: towards the vision for the next decade. International Journal of Digital Earth 5, 4–21. https://doi.org/10.1080/17538947.2011.638500

Fuhrmann, S., MacEachren, A., Dou, J., Wang, K., Cox, A., 2005. Gesture and speech-based maps to support use of GIS for crisis management: A user study. AutoCarto 2005.

Griffin, A.L., White, T., Fish, C., Tomio, B., Huang, H., Sluter, C.R., Bravo, J.V.M., Fabrikant, S.I., Bleisch, S., Yamada, M., Picanço, P., 2017. Designing across map use contexts: a research agenda. International Journal of Cartography 3, 90–114. https://doi.org/10.1080/23729333.2017.1315988

Grosz, B.J., Kraus, S., 1999. The evolution of SharedPlans. Foundations of rational agency 227–262.

Grosz, B.J., Kraus, S., 1996. Collaborative plans for complex group action. Artificial Intelligence 86, 269–357.

Grosz, B.J., Sidner, C.L., 1986. Attention, intentions, and the structure of discourse. Computational linguistics 12, 175–204.

Oviatt, S., 1996. Multimodal interfaces for dynamic interactive maps, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '96. Association for Computing Machinery, New York, NY, USA, pp. 95–102. https://doi.org/10.1145/238386.238438

https://doi.org/10.5194/ica-abs-6-150-2023