International Journal of Geo-Information

MDPI

Article

Development of a Voice Virtual Assistant for the Geospatial Data Visualization Application on the Web

Homeyra Mahmoudi [1], Silvana Camboim [2,*] and Maria Antonia Brovelli [1]

[1] Department of Civil and Environmental Engineering, Leonardo Campus, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy; homeyra.mahmoudi@mail.polimi.it (H.M.); maria.brovelli@polimi.it (M.A.B.)
[2] Geodetic Science Graduate Program, Department of Geomatics, Federal University of Parana, Curitiba 81531970, Brazil
* Correspondence: silvanacamboim@ufpr.br

Abstract: Voice assistants can elevate interaction in geospatial data web platforms. This research introduces a voice assistant in the BStreams platform and focuses on understanding user commands in the geospatial domain. We developed a specialised geospatial discourse framework through structured prototyping. A survey with 66 participants revealed prevalent English geospatial terminologies. Using ChatGPT, we found the term suggestions aligned with survey results, with a notable correlation ($r = 0.81$, $p < 0.01$) between the NPL model's probability scores and term prevalence in survey data. Our study also incorporated usability tests on the application, which uses tools like Web Speech API, Leaflet, and Mapbox geocoding. Results from these tests reaffirm the potential of voice assistants in enhancing geospatial data visualisation, though challenges persist in areas like language understanding and domain knowledge. The paper advocates for further research to refine the integration of voice technology in this domain.

Keywords: voice user interface; geographic information system; human-computer interaction; multimodal interface; natural language; web application; natural language interaction; voice virtual assistant; speech recognition

check for updates

Citation: Mahmoudi, H.; Camboim, S.; Brovelli, M.A. Development of a Voice Virtual Assistant for the Geospatial Data Visualization Application on the Web. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 441. https://doi.org/10.3390/ijgi12110441

Academic Editors: Florian Hruby and Wolfgang Kainz

Received: 27 June 2023
Revised: 24 September 2023
Accepted: 6 October 2023
Published: 26 October 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the advent of technologies such as voice recognition and natural language processing (NLP), new developments are being introduced into everyday tools, impacting various fields. One such field is Cartography and Geoinformatics, where these advances are integral to the evolution of map and geospatial visualisation. Online mapping applications and open geospatial data have democratised spatial information, enabling public participation in its creation [1]. Integrating speech recognition technology into these applications can improve efficiency, user experience and accessibility while reducing the need for specialised skills and knowledge in dealing with geospatial data [2]. Blanco [3] highlighted the lack of infrastructure and practical experience in implementing speech recognition in GIS interfaces, a challenge the present study aims to address.

Integrating speech recognition and NLP in geospatial applications has been a topic of considerable interest in previous research [4]. For example, Lai and Degbelo [5] presented a web map prototype that skilfully fuses text and speech for efficient metadata retrieval. Gilbert's [6] VocalGeo serves as a testament to the potential of speech recognition in promoting geospatial education. Similarly, Cal'ı and Condorelli [7] have highlighted the tangible benefits of incorporating NLP and speech recognition into conventional GIS through their iTour initiative. Furthermore, progress has been made in improving user-GIS communication, as evidenced by Wang, Cai, and MacEachren's [8] PlanGraph and GeoDialogue.

However, even as these milestones are celebrated, challenges remain. The heterogeneity of voice commands across languages is still a complex puzzle, even with the capabilities of advanced AI [9,10]. Additionally, the balance between data democracy and domain expertise is delicate, with calls for GIS systems that are inclusive, participatory, and intelligible to experts and novices alike [11]. On the interoperability front, Granell et al., 2021 observed that current implementations of virtual assistants have a pronounced gap in understanding and compliance with geospatial information standards.

These experiences highlight that the task of integrating voice virtual assistants into web applications dedicated to geospatial visualisation is associated with technical and conceptual challenges. Our research addresses some of these gaps, primarily by building on the lexicon of geospatial commands through user surveys and NLP paradigms. Such efforts can contribute to a potential standardisation framework for developing interoperable solutions. It also prototypes, implements and tests the usability of an open-source solution that can be further enhanced and tested with users from different backgrounds.

2. Materials and Methods

2.1. Domain and Phenomenon Characterisation

This section outlines the operational and functional requirements of the voice map application. Understanding the components and their interplay within the application is crucial. The web application system comprises three primary elements: Frontend, Backend, and Database. Users interact with the application through existing visualisations or adjust them using voice commands.

The application offers a range of thematic visualisations, including marker and choropleth maps. User profiles may vary based on expertise and capabilities. Its use cases span map exploration, visualising spatial relationships, educational settings, palette customisation, and data querying. Essential functions include voice-assisted navigation, cartographic display, base map selection, palette adjustments, and data analysis.

2.2. Methodology

Employing a theoretically-driven model in research conceptualisation offers advantages, such as a framework for hypotheses formulation and testing, improved communication, and enhanced comprehension. Systems that request geographical information in natural language often faces difficulties due to vague references. The obstacles and issues for embedding the natural language processing on a GIS interface requires a schema or a model that eases the process of retrieval of information from the database and reduces the cognitive load from the user's perspective. The PlanGraph theory which was introduced by G Cai, H Wang, A M MacEachren and S Fuhrmann [12] indicates this challenge and considers as a solution, which has already been implemented and tested by the Hongmei Wang and colleagues on their GeoDialogue [8]. The PlanGraph describes its structure by representing three main concepts:

- Recipes: A recipe describes the components of an action in terms of parameters, sub-actions, and constraints.
- Plans: A plan corresponds to a schema describing not only how to act , but more importantly, the mental attitude towards it, such as beliefs, commitments, and Execution status.
- Actions: An action refers to a specific goal as well as the efforts needed to achieve it. An action can be basic or complex.

For example, the goal of a task is to show a map. This task can be considered an action since it is an ultimate goal. The parameter that can be related to this action is the layer of the map. For realising the action, there is a subaction for generating the map. All of the mentioned structures together create the recipe. The recipe definition also includes the ability to define constraints, which describe pre- or postconditions as well as partial orders for subactions. In the case of existing specific conditions for executing the action, subactions,

or retrieving the parameters, the recipe structure is upgraded into a plan structure, which demonstrates a complete structure of the flow.

The Figures 1 and 2 are describing the schema of a recipe and a plan:

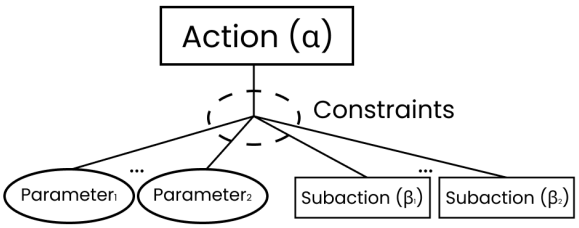

Figure 1. Structure of a Recipe.

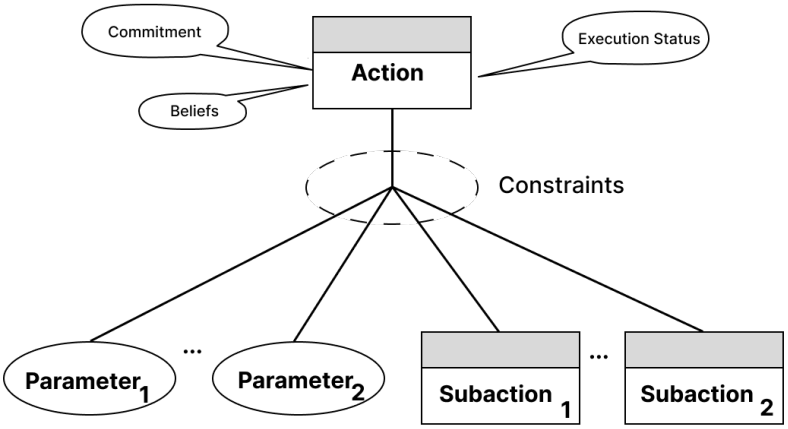

Figure 2. Structure of a Plan.

A compatible, visible, relevant, and intuitive action set was required for effective geospatial data processing and visualisation in the Human-GIS computer context, leading to the categorisation of all actions into four primary types:

- Type I: Acquisition of spatial data (e.g., map layer retrieval)
- Type II: Analytical tasks (e.g., finding spatial clusters)
- Type III: Cartographic and visualisation tasks (e.g., zooming, panning)
- Type IV: Domain-specific tasks (e.g., evacuation planning during hurricanes)

The PlanGraph model facilitates user interaction and function through reasoning algorithms. Upon receiving user inputs, the system initiates a three-step process: interpreting the input, advancing available plans towards completion, and delivering responses to the user. The PlanGraph methodology was integrated into the application by initially categorising all tasks into two main groups: cartographic tasks (visual map modifications) and analytical tasks (data filtering). Task extraction and categorisation were streamlined by leveraging the existing User Interface (UI) and features on the BStreams platform, a pre-existing interface for geospatial data visualisation. This platform is a free data visualisation tool, enabling the users to create reports and visualisation in different formats of graphs and maps with high customisation. Over the past year, BStreams has been enhanced with Geospatial data visualization and various innovative features, enhancing user-friendliness. The effort was done with the cooperation of the development team and the first author; therefore, the infrastructure is flexible for new development.

The UI heavily relies on user-platform interaction for task execution and modification. However, an in-depth exploration of each tool and feature was necessary to determine the correct format, threshold, and structure of user inputs. Examining the core functions of the two visualisations in the backend and the default configuration file stored in the database provided insights into the necessary parameters and conditions for each derived task.

Once all the necessary components were identified, including parameters and constraints, the actions of the PlanGraph models were conceptualised. Following the outlined methodology, the process started with defining the root action/plan and then defining the parameters and constraints for each task, as depicted in Figure 2. Next, the type of action was identified based on the PlanGraph types discussed previously. Actions were classified as basic or complex, depending on whether users required prior domain knowledge to perform the task.

For example, tasks like "Zoom in" or "Pan to the left" were considered basic, while tasks like "Changing the steps for graduated colour" were deemed complex and involved sub-actions and plans. Each action type was determined based on its complexity, with Figure 3 illustrating an example of the basic task "Zoom in/out" in the PlanGraph model. In this model, the root plan aimed to achieve the zooming action, with parameters such as map window size, predefined scale, and the sub-action of reloading the basemap with pre-loaded data. The Zoom in/out task in the user interface is primarily performed by using the mouse wheel or enabling the Zoom in/out feature on the map. The window size is an important parameter derived from the User Interface code, and it plays a crucial role in recognising the specific visualisation that the user is referring to. All of these parameters are retrieved and considered for the PlanGraph model. The parameters and subactions are mainly visible and useful in the technical approaches and code implementation. The root plan was classified as a Type III action, focusing on visualisation modification, while the sub-action was considered Type I.

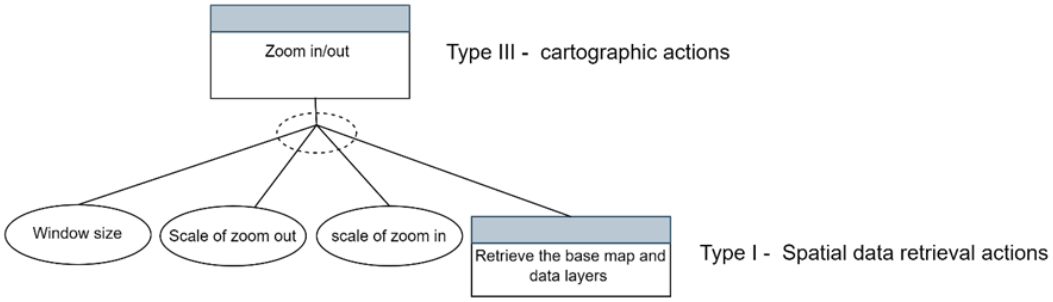

Figure 3. The Plan model of the task "Zoom in/out".

2.3. Survey

The initial challenge was to comprehend the diverse range of voice commands that both expert and non-expert users might employ when interacting with the geospatial application. For this purpose, an English-language survey was designed and circulated among a diverse public to assess and extract vocabulary choices when interacting with web-based maps. The questionnaire contained twelve open-ended questions, enabling respondents to articulate their thoughts and ideas about tasks converted into questions. GIFs were incorporated to aid users in understanding the tasks and commands better. The survey was embedded within a Google Form, and the links to the questionnaire have been distributed through the personal social media channels of the authors, such as Facebook, LinkedIn, and Instagram.

The survey also collected demographic data such as age, gender, educational attainment, field of study, native language, and English proficiency to detect possible patterns or correlations. The survey reached 66 diverse respondents via social media and underwent data-cleaning procedures for result analysis. The analysis primarily revolved around identifying frequently used vocabulary and verbs to enhance human-computer interaction in the application design. Most respondents identified as men (63.6%), indicating a potential gender bias in interest towards the topic. The largest age group, with over 65% of participants, was between 18 and 30, suggesting that older generations may find voice technology on a map less appealing or unfamiliar.

Educational attainment was another factor examined, revealing that over 58% of respondents held a master's degree or higher. Fields of study were also considered, with 50 respondents providing related information. Native language and English proficiency also played key roles. As anticipated, due to the survey distribution channels, the largest group of respondents were Italian speakers. Figure 4 depicts the distribution of other native languages among the respondents, with Persian speakers constituting the second-largest group due to the first author's ethnicity. Regarding English proficiency, most respondents were anticipated to have an advanced level, while none reported having basic proficiency. The survey aimed to collect specialised geospatial data visualisation and analysis terminology to create a structured archive accessible to users. The compiled terms were integrated into the application's code to enhance the user experience by offering diverse inputs and improving response accuracy.

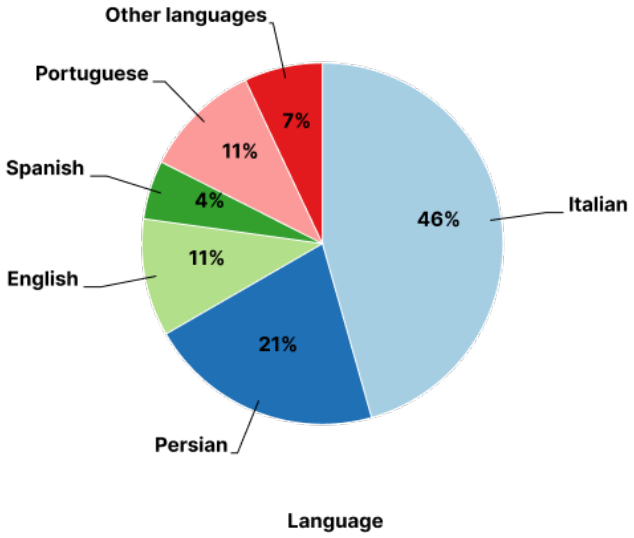

Figure 4. Native language distribution among the respondents.

The frequency of most commonly used verbs was estimated for each question, taking into account the verb-word collocation. Each question was designed to fulfil a specific task and characteristic. The primary approach to analysing derived results was the frequency of vocabularies. Estimating the frequency of commonly used vocabulary also facilitated predicting the most frequently used verbs and words for cloud words and overall human-computer interaction scenarios, given the human tendency to command or order behaviour. This user-centric design application aims to acknowledge and prioritise users' needs, preferences, and experiences.

Figure 5 presents the results from analysing questions in the survey, formulated to understand the terminologies people use to talk to a map. While the anticipated most frequent verbs were "find" and "search," statistics reveal that "show" was used more often. This preference could be attributed to the visual nature of the tasks and the user's interaction with the application, relying on the fact that the application should visually respond to commands.

One of the anticipated significant outcomes of the survey was the creation of a word cloud, which visually represents the most frequent words used in the collected archive, offering a quick overview and understanding of user tendencies. In this context, the word cloud helped pinpoint the most frequently used terminologies related to the application's defined tasks. Figure 6 portrays the word cloud visualisation that represents the frequency of terms derived from the survey data.

This visualisation provides a snapshot of user tendencies and the most commonly used terminologies in the context of the application's defined tasks. Through this survey, we gained insights into the potential user interactions and vocabulary usage, enabling us to tailor the application to suit user preferences and needs best.

	Most used verb	Count	Most verb/word+Collocations		Count
Q1: Find the Pyramid of Giza	Show	28	Show	Me	21
	Find	7	Show	Name of the location	3
	Search	5	Show	The + name of the location	1
	Go	7	Find	The + name of the location	3
	Zoom to	6	Find	name of the location	4
			Go	to	7
			Zoom	to	4
			Zoom	in	2
				Egypt	4
				in Egypt	3
				of Egypt	1
Q2: Geographical direction vs Normal direction	Go	22	Go	to	14
	Move	10	Go	left	7
	Show	12	Show	me the left	8
	Pan	4	Move	to	5
	Zoom to	6	Move	left	5
	left	5	west		3
			left		63
Q 3: Zoom in	Zoom	28	Zoom	in	15
	Show	10	Zoom	to	8
			Show	me	9
				the + location	34
Q4: Zoom out	Zoom	45	Zoom	out	44
	Show	7			
Q5: Basemap Or Map	Show	34	Show	me	22
	Change	9	show	the	3
	Switch	12	show	name of the map	8
			change	the	3
			change	to	4
			switch	to	8
			switch	on	1
			basemap		8
			map		30
			traffic	night	42
			traffic	day	41
			satellite		33
			background		1
Q6: Color palette	Change	18	color ramp		4
	show	5	color palette		4
	colorize	2	color scale		4

	Most used verb	Count	Most verb/word+Collocations		Count
Q7: Query and filter quantitative data	show	28	show	Me	1
	Filter	7	with	more than	41
	Highlight	7	over		4
			above		4
			provinces		41
			regions		3
			cities		2
Q8: Label the features	Highlight	8	Add	label	5
	Indicate	7	Put	label	1
	add	8	the		12
	show	4			
	Tag	1			
	flag	3			
Q9: legend customization	Move	20	top	right	11
	Show	6	top	Left	12
	Change	5	bottom	Left	5
			bottom	right	5
			corner		16
			upper		6
			lower		1
			legend		23
			the	legend	33
Q10: Markers color modification	Change	32	Markers		22
	show	6	pin		5
			of		15
			to		20
Q11: Filter qualitative data for marker map	Show	22	Show	Me	11
	Keep	6	Show	only	11
	Remove	7	Keep	only	4
	Filter	5	marker		10
			pin		19
			Liguria		57
			Lombardia		57
			Lazio		57
Q12: Size and format of the marker	change	43	change	The	15
	Increase	7	markers		55
	decrease	4	symbol		10
	make	12	pin		7
			size		22
			flag		61

Figure 5. The most frequent verbs and words.

Figure 6. The word cloud derived from the results of the survey.

2.4. Comparison with ChatGPT

While the questionnaire provided valuable insights into geospatial interaction, we were concerned that our sample might be too limited to reflect the language patterns of the general public. To ensure a more representative understanding, we turned to NLP, given that it is based on extensive data sources. This would provide a broader perspective on the likely commands for map interactions. Consequently, we emulated the questionnaire using ChatGPT, a sophisticated AI language model developed by OpenAI. The aim of comparing responses generated by ChatGPT with the survey data was to gain a deeper understanding of the commands used and to corroborate the findings. Based on the questionnaire's questions, a vast collection of voice commands was gathered, transcribed into text, and fed into ChatGPT for training and evaluation.

Our findings indicated a significant alignment of the ChatGPT answers with our survey results. Specifically, a Pearson correlation analysis revealed a notable correlation ($r = 0.81$, $p < 0.01$) between the probability scores given by the NLP model and the prevalence of terms observed in the survey data. This analysis was conducted based on 17 mutual words identified across both datasets. One of the intriguing disparities was the model's preference for the term "filter". In contrast, survey respondents favoured terms such as "show", "select", and "highlight". This discrepancy suggests that the term "filter" might be more entrenched within the lexicon of database professionals, thereby being somewhat alien to the general user profile we interviewed.

Structural directives like "top left/top right" and commands such as "zoom in/zoom out" demonstrated user usage variability. Respondents deployed these phrases more freely instead of adhering strictly to a fixed order or set of complements, indicating flexibility in user phrasing preferences. Additionally, terms like "marker" and "provinces" exhibited a higher frequency among survey participants than the ChatGPT model predictions. A plausible explanation for this observation could be rooted in the phrasing of our survey questions. By featuring these terms prominently in the questions, it's possible that they became more salient to respondents. This effect might be because the participants were interacting in a second language, where familiar terms from the question can be more readily recalled and repeated.

Our analysis revealed the most commonly used verbs in voice commands for map interaction: "show", "change", "locate", "zoom", "find", "move", and "filter". The likelihood of using these verbs varied depending on the specific interaction performed and the user's background knowledge. For instance, users with an environmental engineering background preferred the verb "locate", whereas those with a computer science background tended to use the verb "search". Notably, the AI-generated answers closely mirrored responses from human participants. Both sources utilised similar commands and verbs such as "show", "find", "zoom in", and "locate". However, there were differences in sentence structures. For short commands, humans tend to specify very simple and use very few words and simple sentences. In contrast, the human responses were more complex in more advanced tasks such as filtering geospatial data.

ChatGPT mostly understood the intention behind the questions, though the responses varied in terms of terminologies, sentence structures, and probability of appearance. For example, the first question asked after conveying the whole scenario and user profiles was, "Imagine users have a map and wish to command the map with their voice to visualise the Pyramid of Giza. What are the possible commands and their likelihoods in percentages?". The AI's responses mirrored the ones we collected from human respondents, with the verb "show" used most frequently, confirming our expectations.

A comparison of the two most frequent sentences from ChatGPT and respondents revealed that AI-generated sentences tended to be more complex and informative, whereas humans typically used simpler and shorter sentences when interacting with a voice chatbot (Figure 7). In the Human response parse tree, S represents the sentence, VP represents the verb phrase, VB represents the verb, NP represents the noun phrase, and det represents the determiner. In the ChatGPT parse tree, S represents the sentence, VP represents the

verb phrase, VB represents the verb, NP represents the noun phrase, det represents the determiner, VBZ represents the linking verb, and AdvP represents the adverb phrase. In the first sentence, "Show me the Pyramid of Giza", the VP consists of only one verb "Show", whereas in the second sentence "Show me where the Pyramid of Giza is", the VP consists of two verbs "Show" and "is", with a subordinate clause "where the Pyramid of Giza is" acting as the complement of the verb "Show". This difference in the VP's complexity reflects that the second sentence conveys more information since an AI generates it.

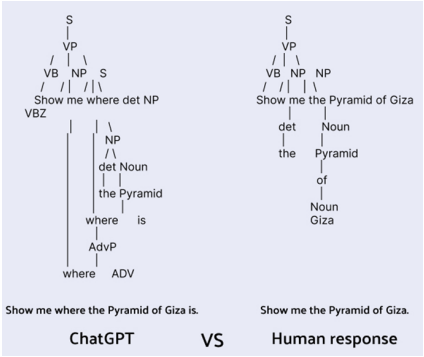

Figure 7. Comparison of the most used sentences from ChatGPT and Human Response in answers from Navigating question.

Our comparison between ChatGPT's responses and those collected from human respondents highlighted similarities and differences. In most cases, the frequency of verb and word usage was similar. However, the complexity of sentence structures showed that humans tend to interact with virtual assistants more generally and simply. This tendency might be due to ease of use, limited time or patience, and advancements in natural language processing and machine learning that enable virtual assistants to comprehend and respond effectively to general queries. The comparison between the percentage of the probability of terminologies' appearance among the two sources (ChatGPT and Survey) for the mutual words has been visualised in Figure 8.

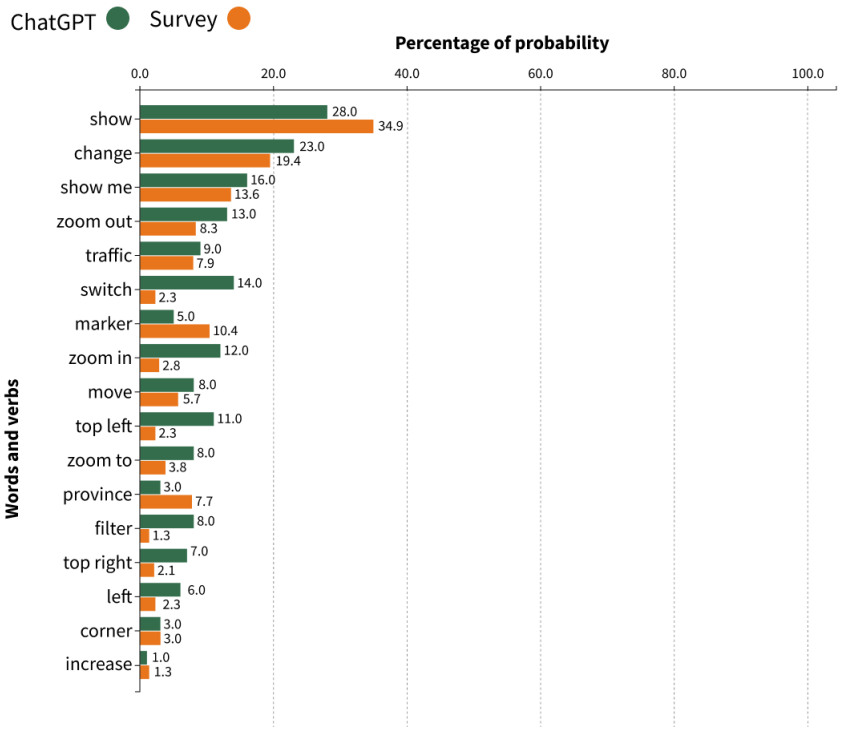

Figure 8. Comparison between mutual words driven from the survey and the ChatGPT.

To drive the implementation of our interactive map, we utilised two primary sources of information: the survey and the responses garnered through ChatGPT. The survey gave us insights into users' linguistic preferences and terminologies when interacting with geospatial data. On the other hand, ChatGPT, with its vast database of natural language interactions, offered a broader understanding of likely voice command structures and terminologies. By combining the findings from both the survey and ChatGPT, we were able to develop our interactive map's voice interface.

3. Results

3.1. Implementation

To implement the project, it was necessary to understand and establish the architecture's backbone. The development process was chosen based on its coherency and compatibility with the interface, guided by the existing architecture of BStreams. The following section provides a detailed explanation of the key elements that shaped the application's development.

3.1.1. Frameworks

BStreams employs object-oriented programming for visualisation creation. Each chart or graph is a subclass inheriting from a parent class, with the primary parent class, "skeleton.js", containing essential features. Each chart or graph has its own script, and it can inherit additional scripts as parent or child classes.

Given the platform's influence on the architecture and engineering for developing virtual voice assistants in the scope of geospatial data visualisation, the application was developed in JavaScript language to ensure compatibility with the platform. The application's architecture follows a REST API, an API that adopts a Representational State Transfer (REST) architectural style, using HTTP requests to access and use data. The main framework is Node.js, an open-source JavaScript runtime environment executing JavaScript code outside a web browser. The application also uses AWS cloud services like S3 Bucket and Lambda functions, with the front end developed using Angular, CSS, and HTML. The two main essential modules with crucial rules in the application are discussed in the following sections. Figure 9 visualises the architecture of the application.

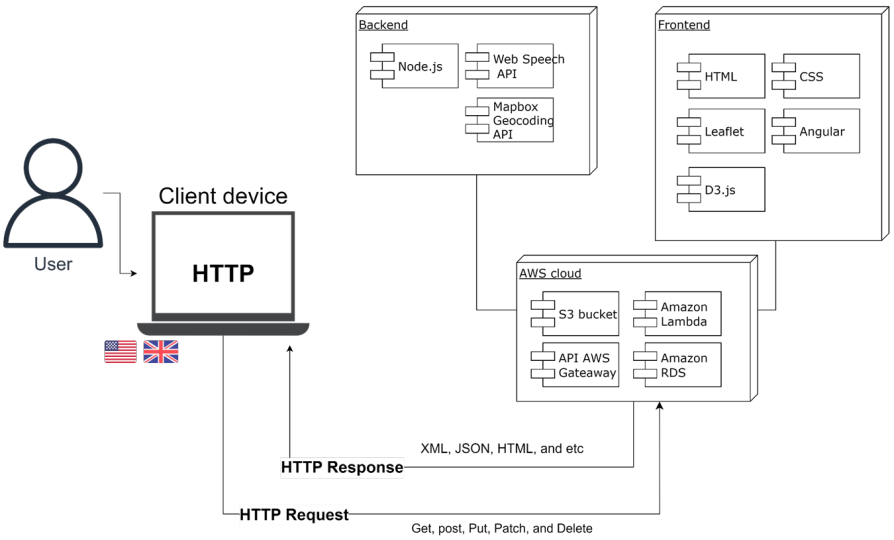

Figure 9. Schema of the architecture of the voice map application.

3.1.2. Web Speech API

The application integrates the Web Speech API to include advanced speech recognition and synthesis into the web application. This feature allows users to activate the microphone by clicking on a button in the chatbot interface, triggering voice speech recognition in

their browser and facilitating communication between the application and the user. The application has been enhanced with a wide range of synonyms based on survey results and application testing, and it uses regular expressions (regex) to match snippets with the received commands and generate the appropriate responses and actions. Figure 10 shows the chatbot with the button activated for listening and resumed for listening.

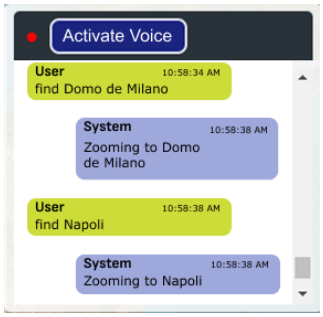

Figure 10. Schema of the chatbot and its content.

3.1.3. Geocoding API

The application incorporates geocoding and positioning features using the MapBox Geocoding API, enabling users to navigate to desired locations by retrieving coordinates based on location queries. This API connects the application to Mapbox with a token and regular expressions are used to recognise user input patterns and extract locations. If a match is found, the application sends a Get request to the Mapbox geocoding API to retrieve the latitude and longitude coordinates of the requested location, and the map moves to the specified location using the retrieved coordinates.

3.1.4. Functions

The engineering behind all the visualization in BStreams is based on creating the main script for each graph or chart as an extended class based on the main parent class which already exists.The parent scripts define the basic infrastructure in terms of the visualisation's features that are available on the interface for all of the charts, such as styles, color palettes, data handling, etc. Meanwhile, all the charts and graphs have their own particular characteristics, but they inherit the common features from their parents. The voice map is not excluded from this engineering, and since BStreams provides the two formats of geospatial visualization, known as marker map and thematic map, it receives the common features from the scripts of the two visualizations within its scripts. Two main essential scripts build the backbone of the application:

1. MainSpeechAPI.js: This JavaScript script, like others in the application, leverages the Web Speech API for speech recognition and action execution. It initializes essential variables for speech recognition and establishes a crucial connection between the frontend and backend, enabling command recognition and control.
    - logToBoxArea(): Logs chat messages, distinguishes user and system messages, transcribes snippets and responds to interpreted commands.
    - interpret(): Interprets received commands, matching them using regular expressions. Actions include changing the base map, zooming, panning, geocoding, and more. Accurate geocoding depends on correct transcription.

    The interpret() function defines voice commands using regular expressions. Each recognized command triggers the appropriate function in VoiceMapChart.js.
2. VoiceMapChart.js: This script extends the code for the voice map chart's structure. It encompasses marker and thematic map functionalities tailored for voice interactions. To enable visualization through voice commands, functions must be adapted to align with defined commands and callbacks in MainSpeechAPI.js.

The 'help' command, as seen in Figure 11, provides users with command information. Users can access this feature using 'help' or 'commands' commands. Regular expressions identify these patterns, and the code generates a response containing available commands for the specific topic. The logToBoxArea() function displays the list of commands

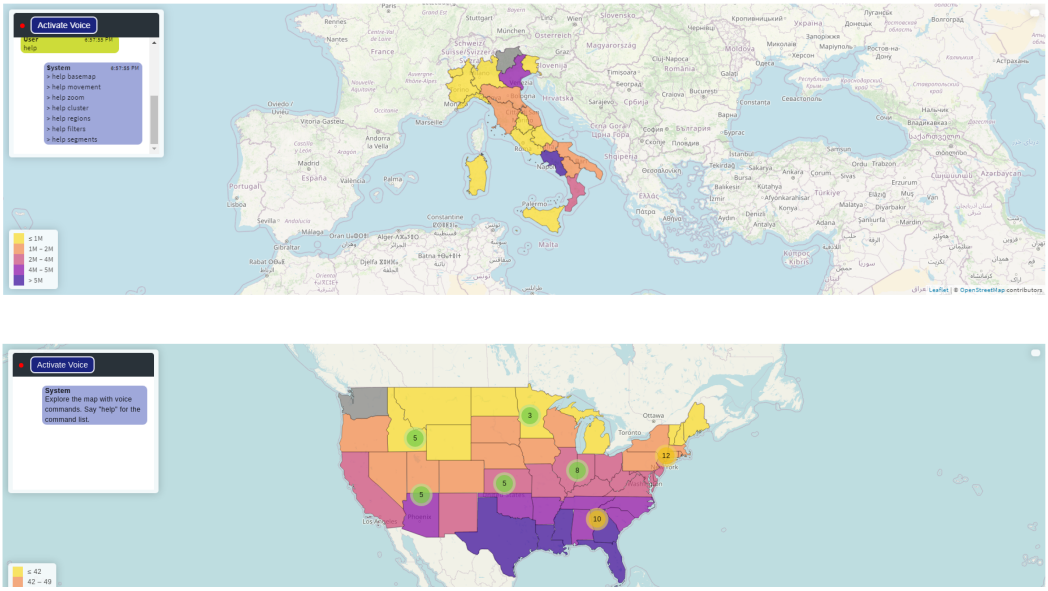

Figure 11. Screenshot of the application with help command.

The provided link (https://view.bstreams.io/view/project/ce707554-b96b-412c-8a6b-1e7f88a52fec, accessed on 29 April 2023), contains a pre-built voice map that is accessible for everyone. The final application is available on the platform BStreams (https://app.bstreams.io/account/signin, accessed on 25 March 2023), and everyone can access it. Users need to register and create a free account. After registration, users can find the application by accessing the dashboard and adding a new project. The application is embedded under the DevLab section, and its available for all the registered users. As a matter of developing an open-source application, all the codes related to the mentioned scripts are available on a Github repo (https://github.com/nextint-bstreams/Voice-virtual-map, accessed on 5 April 2023).

3.2. Usability Testing

The usability testing aimed to assess the navigation and satisfaction of users with virtual voice assistants in the context of geospatial data, identify usability issues and opportunities for improvement in the application and gathering feedback on the users' experience of using the application with voice commands. The testing involved 10 participants with experience using web applications and geospatial data visualisation tools, which were diverse in age, gender, and profession. The testing was conducted remotely using video conferencing tools. Once in the platform, they were asked to employ their own wording for navigation and exploration of various locales. Further, they were prompted to alter the map's view, including its directionality and layering, with functionalities to switch between marker and graduated colour map, adjust marker aesthetics (colour, shape, cluster radius), and manage map properties like opacity and step preference. They also had the capability to customize the map's legend, manipulate its colour scheme, and filter or modify regional displays. After all adjustments and interactions, users were instructed to reset any filters they'd set and conclude their session by exiting the application.

Completion rate, task time, error rate, user satisfaction, age, field of study, gender, and mother tongue were evaluated as metrics to analyse the data gathered from the testing. The results of the test are reported in Table 1. The analysis evaluated the methodology's

effectiveness and identified common issues and opportunities for improvement. The metrics that were considered in the analysis were:

1. Task Completion Rate: This metric measured the percentage of participants who successfully completed each task.
2. Task Time: The average time taken by participants to complete each task was recorded.
3. Error Count: The number of errors made by participants during task completion was measured. The error count, practically, was due to the complexity of the tasks for users, or the accent of the users, which was not recognizable for the application.
4. User Satisfaction Score: Participants were asked to rate their satisfaction with the application on a scale of 1 to 10. This metric gauged the overall user experience and their level of contentment with the application's usability.

Table 1. Result of the testing.

User	Age	Gender	Mother Tongue	Filed of Study	Task Completion Rate (%)	Task Time (minutes)	Error Count	User Satisfaction Score (Out of 10)
1	25	Female	Persian	Geomatics Eng.	100%	10	4	8
2	30	Male	Persian	Computer Science	90%	15	2	7
3	22	Male	Italian	Computer Science	100%	12	5	9
4	28	Female	Italian	Environ. Science [1]	95%	18	3	8
5	37	Male	Italian	Urban Planning	85%	20	4	6
6	20	Male	Italian	Mechanical Eng.	100%	11	5	9
7	26	Male	Italian	Data science	90%	14	4	7
8	24	Male	Italian	Data science	100%	9	3	8
9	31	Female	Persian	Architecture and Landscape	95%	16	4	9
10	23	Male	Italian	Law	75%	22	2	6

[1] Environmental Science Engineering.

The results suggest that the application is generally usable, efficient, and accurate, with the potential for further enhancements in task completion, efficiency, accuracy, and user satisfaction. Two case studies were conducted to test the application's functionality and performance, focusing on visualising the COVID-19 cases in Italy during 2020 and the mean temperature trends across the USA from 1900 to 2022 (Figure 12).

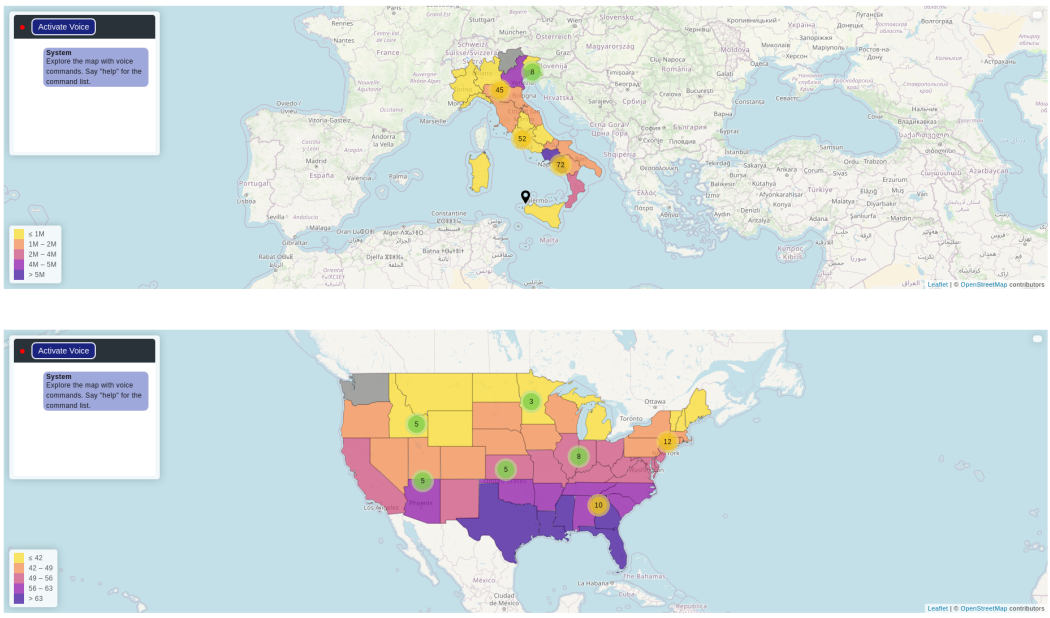

Figure 12. Screenshot of the voice map application the users tested with two already-built maps from Italy and the United States.

Commands that were easy to understand and more general in terms of usage, such as navigating to different cities, zooming in/out, movement and exploration, changing the colour of features and changing the basemap were very easy to understand and use, even without using the help command of the application. In contrast, commands such as filtering maps, changing the classification steps, and increasing or decreasing the cluster radius had lower comprehension and higher error rates. This finding suggests that processing geospatial data and querying it with a voice assistant is still a topic that is highly dependent on users' domain knowledge, and efforts must be taken to ease this matter. During the testing phase of the application, challenges emerged, mainly related to aspects of natural language and accents.

One of the challenges was the application's dependency on domain knowledge. This matter was revealed during the survey analysis, and in practice, users had difficulty comprehending nature's tasks after saying complex commands to the application. The users did not notice any major changes to the application since the complex commands applied changes that an expert user might understand. Technical challenges were another impact regarding the testing of the application. Users have the tendency and expectation of AI from the application. They prefer to issue short, general, and easy commands that the application must understand. However, this challenge is highly related to the fact that users come from different backgrounds and languages, and among the users who attended the testing phase, English was their second language. Therefore, users had difficulty expressing commands with the correct accents to be transcribed correctly by the application, and as a result, the application did not execute the desired action. The user's satisfaction score at the end of the testing can be considered an indicator that shows the usage of a voice virtual assistant for geospatial data visualization is a novel approach for increasing accessibility and usability.

4. Discussion

As noted by Dodge [13], visual representations are essential tools for stimulating creativity and facilitating the identification of patterns in complex geographic data. This view has been echoed by the emergence of web-based platforms, which Haklay [1] argued have extended the democratisation of geospatial visualisation. Our findings are consistent with this, showing that integrating a voice virtual assistant can increase accessibility and user interactivity.

Following the challenges that Lai and Degbelo [5] and Gilbert [6] identified in using speech recognition and NLP technologies in geospatial platforms, our study also confronted the complexities of human-computer interaction in GIS and web map services. Our approach to conceptualising geospatial visualisation tasks in the BStreams platform was inspired by the discourse methods presented by Wang, Cai and MacEachren [8]in their GeoDialogue system. Our findings also reinforce those of Zhu et al. [11], who argue that GIS systems must balance expertise and inclusiveness.

However, the road ahead is paved with challenges and opportunities. Our results, which highlighted users' preference for shorter voice commands, align with the observation of Kadriu & Rista [9] on the heterogeneity of voice commands. As noted by Tulshan & Dhage [10], mastering the diversity of linguistic nuances remains a challenge despite the capabilities of advanced AI. This characteristic underscores the universal need for precision in the design of voice commands for GIS. We also observed that it is possible to find patterns between the commands people use and those suggested by the NPL. In this way, the advancement of the research domain can contribute to interoperability and the future creation of standards, a need identified by Granell [4]. Their research revealed a significant gap in virtual assistants' adherence to geospatial standards, a challenge the field must address collectively.

Our study reinforces the importance of natural language interfaces for geospatial applications. We noted users' preference for concise commands when interacting with maps, pointing to the need for voice assistants to recognise short, clear instructions. Challenges

were evident in our tests, with accent variations causing transcription inaccuracies and subsequent command errors.

The complexity of human-computer interaction in GIS requires continued research. One challenge that emerged was the need for a specialised corpus to enhance GIS capabilities tailored to speech interactions. While our study has helped to fill some of these gaps, there are inherent challenges, such as the intricacies of NLP for geospatial data and the domain-specific knowledge involved. Despite these obstacles, using voice commands in geospatial visualisation offers many opportunities, especially in real-world contexts such as smart cities and environmental monitoring.

Future directions to consider:

- Using machine learning to enhance natural language processing and find real-time alternatives for the terms not in the initial geospatial interaction corpus.
- Expanding to multilingual voice assistants for global accessibility: assessing the heterogeneity of voice commands across different languages, exploring the mismatches and similarities in terms across various linguistic contexts.
- A dedicated study on evaluating and enhancing virtual assistants' adherence to existing geospatial standards and the potential development of new standards tailored for voice interfaces.
- Investigating applications in remote sensing scenarios.
- Developing specialised voice interfaces for platforms like Alexa, focusing on geospatial queries.
- Conducting comprehensive user studies to measure the effectiveness and user satisfaction of voice assistants in geospatial visualisation.

5. Conclusions

Our research consisted of designing specialised voice user interfaces for visualising geospatial data, primarily focusing on English-language commands. We established a discourse framework for geospatial tasks through methodological conceptualisation, drawing on insights from previous GeoDialogue research.

The central part of our work focused on compiling a rich corpus of voice commands. We used a survey enriched with visual content to identify user preferences for natural language commands. This survey informed us about popular terminology and the relationship between lexical choices. An innovative approach was to use an NPL such as ChatGPT to validate, extend and test the correlation with the terms used by the users. While there was a significant overlap between the NPL model's predictions and the survey results, certain terms and phrasings highlighted the nuanced differences between a generalised NLP model and specific user behaviour in a given context.

After compilation, the application was designed to harmonise with its hosting interface using state-of-the-art technologies. A subsequent usability testing session with ten users provided real feedback, completing our research process. Through these activities, we hope to have addressed some of the gaps in speech-enabled geospatial visualisation, paving the way for further developments in the field.

Author Contributions: Conceptualization, Homeyra Mahmoudi, Silvana Camboim and Maria Antonia Brovelli; methodology, Homeyra Mahmoudi, Silvana Camboim and Maria Antonia Brovelli; software, Homeyra Mahmoudi; validation, Homeyra Mahmoudi, Silvana Camboim and Maria Antonia Brovelli; formal analysis, Homeyra Mahmoudi, Silvana Camboim and Maria Antonia Brovelli; investigation, Homeyra Mahmoudi, Silvana Camboim and Maria Antonia Brovelli; resources, Homeyra Mahmoudi, Silvana Camboim and Maria Antonia Brovelli; data curation, Maria Antonia Brovelli; writing—original draft preparation, Homeyra Mahmoudi; writing—review and editing, Homeyra Mahmoudi and Silvana Camboim. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: In order to safeguard the privacy of the people who took part in the survey, all personal data collected within the survey and useful for the data elaboration and for the dissemination of results have been deleted from the data set publicly available, so that it is impossible to retrieve the single participants. Before filling in the survey, participants were provided with an information sheet concerning the management of their personal data, according to the rules contained in the GDPR (EU Regulation no. 679/2016 of 27 April 2016) and they gave their consent to the use of their personal data (Age, Gender, Education level, English proficiency level and E-mail address for carrying out the research project and for dissemination purposes. For accessing to the driven data from the survey, redirect to this link: "https://voice-speech-map-trial-version-1.s3.us-east-2.amazonaws.com/Data-From-The-Survey.zip, accessed on 6 October 2023".

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plan. Plan. Des.* **2010**, *37*, 682–703. [CrossRef]
2. Austerjost, J.; Porr, M.; Riedel, N.; Geier, D.; Becker, T.; Scheper, T.; Marquard, D.; Lindner, P.; Beutel, S. Introducing a Virtual Assistant to the Lab: A Voice User Interface for the Intuitive Control of Laboratory Instruments. *SLAS Technol.* **2018**, *23*, 476–482. [CrossRef] [PubMed]
3. Blanco, T.; Martín-Segura, S.; de Larrinzar, J.L.; Béjar, R.; Zarazaga-Soria, F.J. First Steps toward Voice User Interfaces for Web-Based Navigation of Geographic Information: A Spanish Terms Study. *Appl. Sci.* **2023**, *13*, 2083. [CrossRef]
4. Granell, C.; Pesántez-Cabrera, P.G.; Vilches-Blázquez, L.M.; Achig, R.; Luaces, M.R.; Cortiñas-Álvarez, A.; Chayle, C.; Morocho, V. A scoping review on the use, processing and fusion of geographic data in virtual assistants. *Trans. GIS* **2021**, *25*, 1784–1808. [CrossRef]
5. Lai, P.-C.; Degbelo, A. A Comparative Study of Typing and Speech For Map Metadata Creation. *AGILE GISci. Ser.* **2021**, *2*, 7. [CrossRef]
6. Gilbert, T. *VocalGeo: Using Speech to Provide Geospatial Context in the Classroom*; figshare: London, UK, 2020. [CrossRef]
7. Cali, D.; Condorelli, A.; Papa, S.; Rata, M.; Zagarella, L. Improving intelligence through use of Natural Language Processing. A comparison between NLP interfaces and traditional visual GIS interfaces. *Procedia Comput. Sci.* **2011**, *5*, 920–925. [CrossRef]
8. Wang, H.; Cai, G.; MacEachren, A.M. GeoDialogue: A software agent enabling collaborative dialogues between a user and a conversational GIS. In Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial Intelligence, Dayton, OH, USA, 3–5 November 2008; Volume 2, pp. 357–360. [CrossRef]
9. Kadriu, A.; Rista, A. Automatic Speech Recognition: A Comprehensive Survey. *SEEU Rev.* **2020**, *15*, 86–112.
10. Tulshan, A.S.; Dhage, S.N. Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa. In *Advances in Signal Processing and Intelligent Recognition Systems, Proceedings of the 4th International Symposium SIRS 2018, Bangalore, India, 19–22 September 2018*; Revised Selected Papers 4; Springer: Singapore, 2018. [CrossRef]
11. Zhu, A.-X.; Zhao, F.-H.; Liang, P.; Qin, C.-Z. Next generation of GIS: Must be easy. *Ann. GIS* **2021**, *27*, 71–86. [CrossRef]
12. Cai, G.; Wang, H.; MacEachren, A.M.; Fuhrmann, S. Natural conversational interfaces to geospatial databases. *Trans. GIS* **2005**, *9*, 199–221. [CrossRef]
13. Dodge, M. Mapping and geovisualization. In *Approaches to Human Geography*, 2nd ed.; Aitken, S.C., Valentine, G., Eds.; Sage: London, UK 2014; pp. 289–305.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.