VocalGeo:
Using Speech to Provide Geospatial Context in the Classroom

Thomas Gilbert*[1]

[1]Newcastle University

July 2nd, 2020

Summary

The provision of location information often plays a large part in communication and is key to the visualisations used in teaching; however, location information such as geospatial context is often challenging to convey and thus omitted from narratives; this is especially pertinent in classroom situations in which the teacher is constrained by other demands. VocalGeo is a software capability that addresses this challenge, utilising speech recognition and web mapping technologies to provide automatic, real-time geospatial context on a dynamic map in response to the teacher's and students' speech. This work shows how readily available web technologies can be used for interactive teaching of geospatial information.

KEYWORDS: speech recognition, web mapping, geospatial visualisation, interactive teaching

1. Background and Concept

Geospatial concepts are intrinsic to many topics and their comprehension is often most intuitive when conveyed visually. There is also evidence that visualisation increases student motivation, cooperation and ability to think critically (Shatri & Buza, 2017). However, provision of this context is often time-consuming in its preparation and can be disruptive to lesson flow. Although technologies that combine voice recognition with mapping exist (such as through Apple Siri and Amazon Alexa), they have not been integrated into a continuous, real-time, seamless capability that can be delivered to the classroom in support of teaching. Significant challenges to achieving this include sufficient quality of speech recognition (Kardava, Antidze, & Gulua, 2016), accurate identification of locations from text (Altaweel, 2019) and designing a user interface that provides fast, intuitive and engaging geospatial context.

A seed concept was conceived in early 2018: a web-based software application should detect and interpret speech, identify location entities in the (automatically) transcribed text and then use a dynamic map or globe to pan, zoom and highlight these features on-the-fly. The concept received initial funding in late 2018 to pursue market research within the education section, supported with a basic prototype.

2. Prototyping and Market Research

Prior to receiving funding, the concept was envisioned to benefit at least one of various user groups, including geospatial professionals (assisting presentations), emergency services (locating casualties) and news reporters (live incident support). However, the education sector – more specifically, humanities departments in 11-18 age range – was deemed most likely to experience positive impact: teachers could use the software for presentations and students for self-led discovery; the students should be able to interpret rapidly changing mapping of variable scale and the teaching environment is sufficiently informal that any technology testing or deployment would not be unduly disruptive. An initial prototype was developed to demonstrate the feasibility of the concept. The web software converted voice signal to text, matched word tokens against a gazetteer of countries, put a pin on the centre-points of matching countries and panned to these pins.

The prototype was then used for market research aimed at understanding the challenges faced by humanities teachers when conveying geospatial information. This research was intended to determined

* t.t.gilbert2@newcastle.ac.uk

DOI: 10.6084/m9.figshare.12600428

whether a more mature version of the technology was likely to address these challenges and gather requirements or suggestions from the teachers that could be used to develop a minimum viable product. In order to account for the breadth of educational styles across the UK, seven schools from across the Northeast, Midlands and Southwest of the UK were recruited, representing both state- and privately-funded institutions. At each institution, a focus group was held with 2-4 participants in a round-table format. An introduction to the research was followed by a 45-minute semi-structured group interview that was designed to draw out the broad role of the humanities teacher in the classroom, elicit specific teaching requirements and identify key frustrations. Following the interview, the software prototype was demonstrated by the researcher. A 20-minute creative session was then run with the participants to stimulate ideas for further development of the initial prototype. The participants were encouraged to discuss ways in which the technology could be improved, pitfalls that should be avoided and factors that have not yet been considered.

Feedback from the teachers indicated that the technology concept has the potential to: support lesson planning; serve as a useful attention-grabber at the start of a lesson; get students involved and spatially oriented; teach the value of place; allow more experiential learning of new places; be impactful in lessons on politics; form the basis of novelty lessons and inspire students; and allow more freedom of movement of the teacher around the classroom. Geography teachers further highlighted the benefits to spatial awareness, understanding spatial scales, getting to know the relative positions of countries and tracking physical phenomena such as natural disasters. History teachers emphasised the utility in locating historical events, and showing temporal-spatial changes such as urban expansion and shifting of sovereign borders. Novelty lessons included a 'passport to the world session', in which students use the technology to 'fly' around the globe. Some teachers identified strong applicability to addressing additional learning needs, such allowing those with lower literacy to discover location information.

The teachers also helped to identify what is required for turning the prototype into a minimum viable product, suggesting that the software should: be underpinned by high quality, curriculum-specific mapping (such as historical maps); allow manual annotation of these maps; support the drilling down into data and finer spatial scales; use voice commands for activation, background map selection and deactivation; provide the ability to record a video of the dynamic process, which students can then deliver as part of a project. Potential pitfalls were also discussed, including lesson disruption in the event of a location error and the impact of multiple background voices. Importantly, some teachers made it clear that the software must be robust and reliable, or it will be abandoned.

3. Development and Deployment of Minimum Viable Product

Following market research, the project received further funding for two months of software development and subsequent user testing with the aim of assessing the functionality, usability and robustness of the software from a teacher’s perspective. The prototype was developed into a minimum viable product in early 2020.

Technically, the web-application is written in JavaScript and makes use of the Google Web Speech API, Leaflet.JS, various Node.JS libraries for geospatial data analysis and processing, several gazetteers (geographical dictionaries) containing coordinates that link more common place names to locations, named entity recognition (NER) and geocoding APIs for identification of less common place names, multiple mapping backgrounds and overlays, and Webpack for software bundling. The software needs a persistent internet connection; it converts speech from a naturally flowing narrative or conversation (received via the computer's microphone) to text, the content of which is processed against the gazetteers and APIs. The spatial extent of identified locations are then mapped out in real-time with the map panning and zooming with a speech-to-visualisation lag of around one second. The user is able to select how many of the latest matches are displayed, the spatial remit of allowed matches and whether the map responds dynamically to these matches or is static. Via user interface button, the software also allows language selection, muting and reactivation and map clearing; it also provides the user with a rolling transcript of the speech it has detected. Some of these functions, such as muting and reactivation, have been linked to a presentation controller that allows remote control. See Figure 1 for VocalGeo in action.

DOI: 10.6084/m9.figshare.12600428

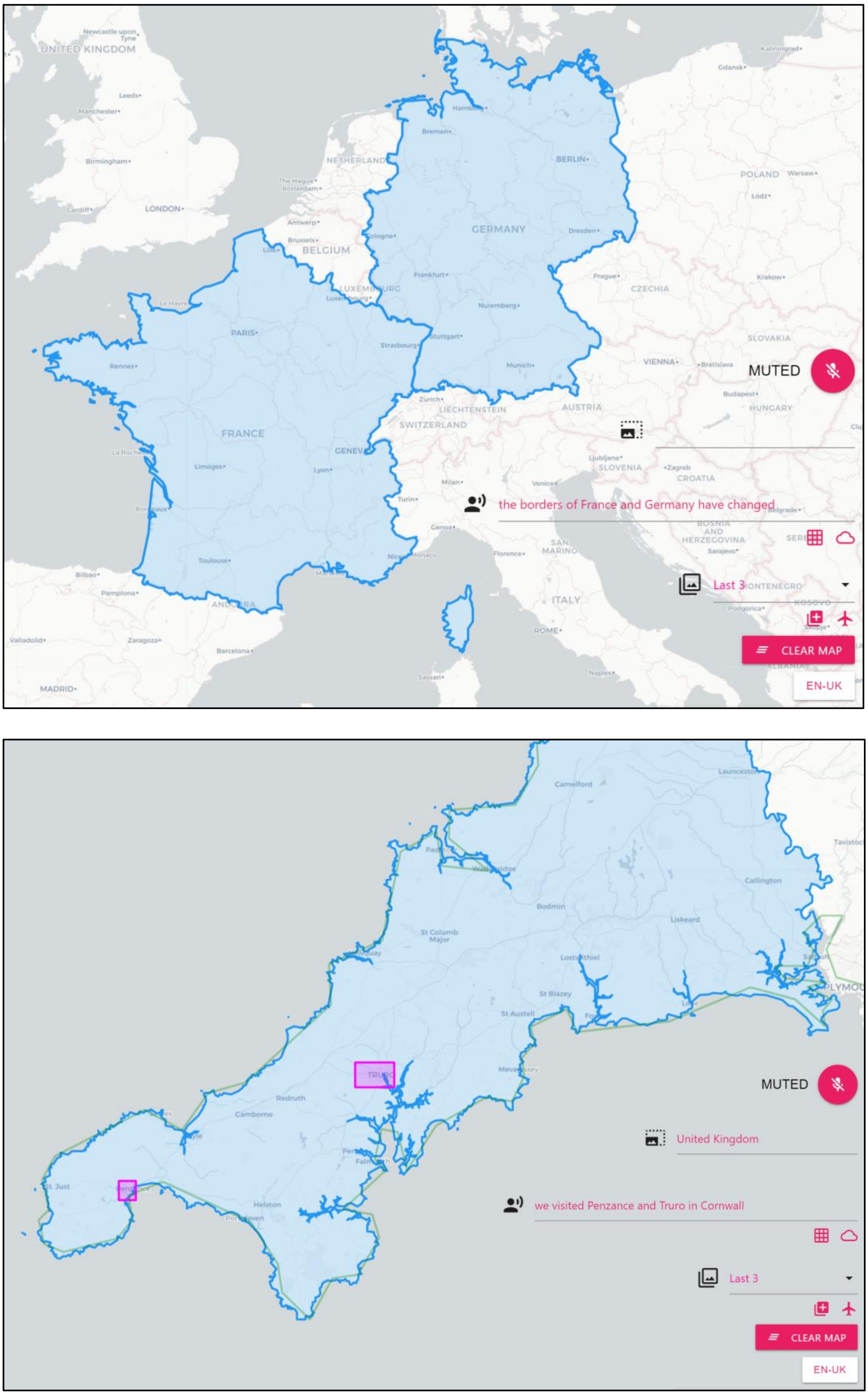

Figure 1 – Two screenshots of VocalGeo in action. The transcribed text includes placenames that have been identified in the speech and then mapped. France, Germany and Cornwall were represented in the gazetteers but the named entity recognition and geocoding APIs were needed for Penzance and Truro (pink bounding boxes). *Contains map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org. Contains OS data © Crown copyright and database right 2020.*

DOI: 10.6084/m9.figshare.12600428

Disruption due to the COVID-19 pandemic prevented in-person user testing with teachers. Instead, the software was deployed on a server with email and password authentication allowing access to the full NER and geocoding API capability. VocalGeo was then tested remotely with a small number of individuals to guide refinement before testing with teachers. User testing with teachers has now begun: initial evaluation indicates that the software is already developed enough for some lesson scenarios and further suggestions have been provided for improvements. Iterative testing and development is ongoing.

4. Discussion and Next Steps

The market research, software development and user testing brought to light a number of questions and challenges around the intersection of speech recognition, NER, geocoding and visualisation. Beyond VocalGeo's compatibility with other languages, its ability to accurately and consistency interpret strongly accented speech and dialects should be tested. The APIs used for NER and geocoding were limited in both the request rate (per second) and total usage (per day or month) without incurring significant subscription costs; there is scope for investing time in generating large, customised gazetteers that are queried server-side, eliminating reliance on third-party APIs. There remains a challenge of how to handle placenames that could match both proper and common nouns, such as 'New Town', for which more sophisticated analysis of semantic context might be needed. How matches are best visualised on screen for specific user groups and their use cases merits more thorough research; one example of this is the choice of map projection. There is also a need to address remaining requirements outlines by teachers, such as the ability to drill down into data for matched locations, manually annotate the map and record a video of the entire process. Finally, it has been noted that a more modular version of VocalGeo could serve as a research tool for experimenting with the intersection of these themes and the performance of various technologies such as competing speech recognition and NER engines.

5. Acknowledgements

This work is funded by EPSRC through an IMPACT programme. Credit is due to: Chris Flatley of Committed Software for development of the first prototype, discussion on further development and paper reviewing; David Verweij for discussion of the concept and technology, and reviewing of this paper; Gretel Dixon and Jenni Brady of Newcastle University in their advisory capacities. Thanks also to teachers from the participating schools: Mounts Bay Academy, Helston Community College, Humphry Davy School, Mullion School, Prudhoe Community High School, Queen Elizabeth High School (Hexham) and Cheltenham Ladies' College.

6. Biography

Thomas Gilbert is a postgraduate researcher at Newcastle University with interests in spatial data, natural language and their intersection.

7. References

Altaweel, M. (2019). GIS and Named Entity Recognition: Identifying Geographic Locations in Text. Retrieved January 9, 2020, from https://www.gislounge.com/gis-and-named-entity-recognition-identifying-geographic-locations-in-text/

Kardava, I., Antidze, J., & Gulua, N. (2016). Solving the Problem of the Accents for Speech Recognition Systems. *International Journal of Signal Processing Systems*, *4*(3), 235–238. https://doi.org/10.18178/ijsps.4.3.235-238

Shatri, K., & Buza, K. (2017). The Use of Visualization in Teaching and Learning Process for Developing Critical Thinking of Students. *European Journal of Social Sciences Education and Research*, *9*(1), 71. https://doi.org/10.26417/ejser.v9i1.p71-74

DOI: 10.6084/m9.figshare.12600428