A Multi-modal Communication Approach to Describing the Surroundings to Mobile Users

Janne Kovanen, Tapani Sarjakoski, and L. Tiina Sarjakoski

Finnish Geodetic Institute, Department of Geoinformatics and Cartography,
P.O. Box 15, 02431 Masala, Finland
{janne.kovanen,tapani.sarjakoski,tiina.sarjakoski}@fgi.fi

Abstract. Mobile users frequently pass non-obvious features that could be represented to the user in a multi-modal manner. This type of information can be used to affect the decision making of the user or to complement his or her navigation experience. However, data providers do not have a common data interchange schema for describing geographical features multi-modally. This paper presents a multi-modal approach by extending the GeoJSON, GML, and KML formats to describe the surroundings of a mobile user in a Location-Based Service. In addition, the paper discusses how the approach can be implemented on a mobile client. Finally, the paper demonstrates how the proposal has been implemented with a functional prototype for a hiking use case.

Keywords: Multi-Modal Communication, Location-Based Service, Data Interchange Format, Mobile User.

1 Introduction

Traditionally, a user communicates with a map through a visual channel. The visual communication can incorporate varying elements, like textual content, textures, pictorial symbols, and other abstractions. Such information is well-suited to the visual sense, but people use other senses at the same time. A user might use one sense for the primary task, and others for secondary tasks. Hence, it is a natural trend for applications to provide information directed at multiple senses.

Multi-modally described data complements information extracted from the local surroundings, such as a visual overview of the environment, which a map application can provide. Other means, such as speech and pictures, can sometimes better explain natural phenomena or landforms. They also supplement information coming from human-made sources, like signposts or information boards. Similarly as with a printed map, this type of information can be used to affect the decision-making of the user or to simply cultivate his or her user experience.

Multi-modal descriptions can even work concurrently with routing guidance if the auditory or haptic communication does not overlap. They can either be delivered independently. or multi- modal descriptions can be incorporated within

S. Liang, X. Wang, and C. Claramunt (Eds.): W2GIS 2013, LNCS 7820, pp. 82–99, 2013.

the routing guidance. Consequently, if the route guidance advises a user to go around a dangerous area, the reasoning for such guidance can be explained to the user. Hence, the user is assured that the routing is based on logical deduction.

One barrier for multi-modal descriptions is the absence of a generic way to define how applications can distribute and represent data in a multi-modal manner. In this paper, we present a user-centered solution that complements the navigation experience in an outdoor environment. The features may include interesting phenomena, recreation alternatives, or dangers, all of which need to be communicated to the user in a multi-lingual and multi-modal way. First, the paper describes the use case of our study and related research. Next, we propose a common solution for how Web services can distribute multi-modally describable geographical data to other media, including mobile applications. The paper continues by presenting the phases relevant to the mobile applications that provide descriptions. Finally, following a discussion of the general phases, we provide an overview of how the approach has been realized in our prototype.

1.1 Use Case

This study was performed so that hikers were able to perceive information in several modalities. The requirements are based on a specific area – Nuuksio National Park, near Helsinki, the capital of Finland. The park is frequently visited by locals and tourists searching for a moment of peace and relaxation outside the urban area. The park contains several types of forest, lakes, and small hills typical of nature at this latitude. In addition, the park provides marked circle trails, footpaths for hikers, and recreational structures such as camping sites and fireplaces. The features that need to be described to the hiker include natural phenomena, recreational activities, opening hours, warnings, and geographical names, as well as other features that are difficult to interpret without help. The descriptions need to be multilingual to serve international visitors.

The natural environment is challenging for navigation performed using mobile devices because the terrain is sometimes hilly and the wooded areas are dense, which weakens the positioning accuracy of the devices. Similarly, mobile data connections are typically non-existent. Visitors with smartphones will most likely download applications and data before entering the park or at a visitor center. However, changing data may require that they will need to download information during the course of their hike.

2 Related Research

During the last decade, an increasing number of studies have focused on smartphones as a consequence of the tremendous increase in their usage and capabilities. Smartphones can be used to deliver information with several channels. For instance, a navigation application might point out the direction to follow using text, voice, visual arrows or a vibration pattern. One field of study where the surrounding is described is pedestrian navigation. A part of research performed in this field has studied how different modalities affect spatial knowledge acquisition (e.g. [24], [2], and [28]).

Other navigation-related studies have concentrated on how multiple modalities: (1) help people with disabilities, (2) improve navigation performance, or (3) can be used together. For instance, Baus et al. [3] prepared textual route instructions containing auditory perceptible landmarks that they converted into synthesized audio records before their trial. They validated that auditory perceptible landmarks are helpful and may be used by visually impaired persons in a similar fashion as graphically represented landmarks are used by normally sighted users. Kainulainen et al. [21] also applied synthesized speech in their study by comparing it to non-spoken recognizable soundscapes (soundmarks) that are objectively unique to a place. In addition, they tested a mixture of the two by using one type of audio to one type of information. The main conclusion of their study was that users mostly preferred speech as such. However, soundscapes are not the only audio available. Other types of auditory cues representing real-world objects include, for example, earcons [4] and auditory icons [13].

Modalities helping navigation are not restricted to audio. For example, Chittaro and Burigat [8] have shown that photographs and abstract direction arrows are meaningful method of lowering the amount of navigation errors, and Pielot, Henze, and Boll [31] have reported that a tactile belt lowers the disorientation of the users and helps them to find shorter routes. Additional information accompanying the actual route guidance is also a way raise the comfort level of impaired users [17].

Route instructions, however, are typically not meant to contain any supplemental information regarding the surrounding that is not necessary in the guiding process. Instead, route instructions should be as short as possible to reduce the cognitive processing load required from a user; however, in some situations users are open to spontaneous suggestions about nearby interesting locations that lead them to deviate from their original route [32].

Another field where the surrounding is described to the user is festival, fair and museum tour guides which typically not only guide the user in the particular area, but include a disclosure of the environment. Hippie [29], HyperAudio [30], and GUIDE [7] are examples of the first wireless adapting museum guides that made use of the user's location and direction to give sound bites of the relevant objects in a similar fashion as a traditional curator might do to and guide to the next object. A couple of years later Ciavarella and Paternò [9] found out that foreign visitors especially appreciated even richer multimedia, such as video clips of the relevant objects, which traditional curators did not provide. Since then, smartphone museum guides have evolved even further and audio through headphones is not just meant to free the eyes to look at attractions; instead, multi-modality especially helps people with vision or hearing disabilities to visit museums or fairs autonomously. The study of Ghiani et al. [14], for example, shows that synthesized speech and vibrations are valid methods to guide blind visitors in a museum.

Similarly, video clips have been made to contain information nuggets in the form of sign language to improve the user experience of deaf visitors [38].

Most of the studies presented in this chapter or elsewhere do not mention how the descriptions are delivered to the mobile devices. A sophisticated guess is that most systems developed for research use an in-built test set or an application-specific data interchange format. Nevertheless, Richter [33] presents, as part of his paper, XML-fragments that contain context-specific route instructions based on landmarks, which can be externalized into meaningful spoken or textually represented route instructions. Similarly, some progress towards a common and open data interchange is available through the Open Geospatial Consortium (OGC) OpenLS [26], which can be used to deliver instructions for maneuvers that incorporate textual descriptions of surrounding features with a preferred language.

3 Interchanging Descriptions

A suitable open data transfer format for vector data constitutes a relevant part of our solution for enabling cross-modal descriptions. In this section, we describe how data interchange formats can be extended to support cross-modal descriptions of features. At first, we discuss some of the most common open formats. Then we describe the additions that we made to the original schemas.

3.1 Alternative Data Interchange Format Bases

For a feasible solution, we required from the format a high general acceptance at either the present moment or in the near future. In addition, we required that the following criteria be met:

1. that it support geometries such as polygons, linestrings, and points and the multiple instances in which they occur;
2. that the data size and serialization speed are suitable and that it is easy to use [35]; and
3. that it provide extendibility and scalability for the interoperability of Web services and mobile devices.

Good interoperability requires not just that a format can be implemented, but also that the context restrictions are taken into account. For example, deserializing a complex format in comparison to a simpler format generally consumes more battery power on the mobile device, and data with a large overhead consumes more data bandwidth – thereby, they typically take longer to download or upload. To be easy to use, a particular format should be human-readable.

We found that KML [39] and Geography Markup Language (GML) Simple Features Profile [36] are the best alternatives from Extensible Markup Language (XML) grammars; both of these alternatives are standardized by the OGC and defined by their own schema documents. The main difference between the formats is that GML is a grammar more directed to modeling and interchanging geographical content, whereas KML focuses on representing geographical features from both visualization and viewing perspectives. As a result of the huge range of possibilities related to modeling, GML is not used as such; instead, an application schema that defines the grammar for a specific domain is used in its place. The application schema extends GML and makes sure that the data provider and user have a common understanding of the problem area.

GML-encoded data can also be rendered based on the rules provided by a styling language, such as the Styled Layer Descriptor (SLD) defined by the OGC. Alternatively, because both grammars use a similar geometry model, the content of the GML may be converted into KML for cartographic visualization purposes, even though some information might be lost in the process. The conversion can be implemented by first converting the GML-encoded coordinates into the WGS84 coordinate reference system. Next, the GML-encoded data is translated into KML using a particular solution, such as members of the Extensible Stylesheet Language (XSL) family. Finally, the symbology for each feature is included in the output.

The benefits of using XML include the number of data processing tools that have been developed for it and the possibilities that the structure provides. Nevertheless, XML-encoded data has, in theory, two main problems: it is time-consuming to parse, and it takes a relatively large amount of space, even if it is compressed. For example, a single KML document and the data referenced from that document, such as images or audio files, are typically transferred as a zip archive. In comparison to other formats, XML-related problems have even been empirically validated in contemporary mobile environments, like on the Android platform [34][35]. A formidable solution is to use a binary XML, which has been demonstrated to be significantly smaller in size and faster to parse [23]; however, it is neither human-readable nor widely accepted.

Another solution for overcoming the performance and size problems is to use GeoJSON [5], which is based on the JavaScript Object Notation (JSON). According to Crockford [10], “JSON’s design goals were for it to be minimal, portable, textual, and a subset of JavaScript”; consequently, it is no surprise that JSON is typically more compact than XML, faster to parse, and recommended for mobile applications (e.g. [18]). However, JSON does not have an official schema associated with it, which has resulted in a situation where clients are forced to implement specific parsers for every domain. GeoJSON as such is a partial solution for overcoming the situation, but apart from its features and their geometries, it does not define rules for any additional characteristics, like cartographic visualization.

All three formats share a similar set of properties, such as being text-based and human-readable. The main difference between the formats is their audience and usage context. GML is typically used by professionals for Web service communication, like Web Feature Service (WFS) responses. KML, for its part, is used by professionals from diverse fields, but it is also favored by ordinary non-professional users mainly because the format is used by Google Earth and Google Maps API. In comparison to the two XML-based grammars, in its earlier stages GeoJSON was mostly used in AJAX communication by Web applications because it can easily be serialized into JavaScript. Lately, however, the grammar has also become popular in mobile communications between Web services and mobile applications.

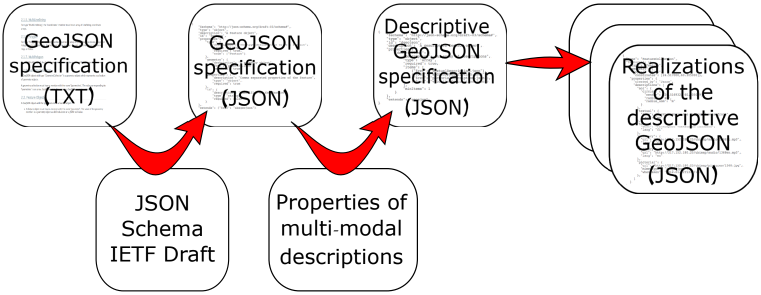


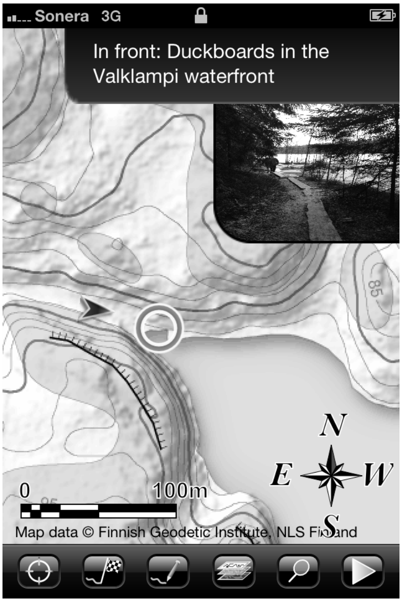
Fig. 1. The process of creating a JSON-based schema for spatial features that can be multi-modally described

3.2 Descriptive Data Interchange Schemas

Based on the properties of the data interchange formats, we came to the conclusion that we needed to create a schema for the three formats mentioned above: namely, we needed to create a schema for GeoJSON that could be used for lightweight mobile communication, a schema for KML aimed at non-professionals who could then visually see the features described on the map, and a schema for GML that would form the basis for domain-specific schemas on the server-side. We extended the KML and GML Simple Feature Profile schemas with XML Schema Definition Language (XSD), version 1.1. Similarly, we defined the extension to GeoJSON by utilizing a draft version of the JSON Schema [40], which is being supported by actively developed JSON validators. The three-format solution allows systems, for instance, to store features with vast properties on the server, to interchange data between the different Web services in GML, and, at some point, to convert the data into simpler KML and GeoJSON formats supported by thin clients. The main advantage of the schemas is the possibility to validate the structure of the data by an automated procedure and to share the data with a common understanding of the content.

We extended the GML and KML schemas with new application profiles. In GML, we used the *AbstractFeatureType* as the base, and in the case of KML, we decided to employ the *ExtendedData* element of KML. The ExtendedData element provides three options for domain-specific extensions. The first option is to use untyped textual name-value pairs for any number of attributes. The second option is to define types for the values by referencing a local or external schema definition, which defines an alphanumerical data type related to each name. KML-reading clients, such as Google Earth, can use both of these options to represent the name-value pairs for the user. The third option was chosen because it is more flexible. It allows us to define our own schema, one which might use additional primitive data types. The content of the third option is ignored by applications not supporting our schema, which is sensible because the content is not meant to be represented as textual information for the end user. However, the symbology of the features can still be rendered on KML viewers.

The process of defining the GeoJSON extension is presented in Figure 1. It required that we first convert the textually defined GeoJSON format into a JSON Schema written in JSON. The JSON Schema makes it possible to a large extent to validate the syntax and structural integrity of the GeoJSON instances. However, all textually defined restrictions cannot be validated. For example, not even the use of regular expressions is of benefit to validate the amount or range of coordinate values, because that requires an extensive knowledge of the coordinate reference system being used, which the format does not provide. Next, we created a new JSON Schema document for our extension that references the JSON Schema for GeoJSON by using the *extends* property provided by the JSON Schema. The extension inherits all properties of the GeoJSON, and any instance of our extension also has to be valid GeoJSON. Consequently, the descriptions are stored as a property of a GeoJSON feature.



In front: Duckboards in the Valklampi waterfront

Fig. 2. An application shows what is ahead of the user. The area is visualized with a resizable picture and is described in both textual and auditory form. The blue circle shows the location of the feature.

The descriptive schemas contain the possibility to describe a feature using text, auditory data, haptic vibration patterns, and graphics. The graphics might include a photograph or a sketch. For example, navigating between the map and photographs helps users to find to their destination significantly faster than by using the map alone [8]. An auditory description might include speech or an abstract non-speech sound pattern, such as a soundscape or an auditory icon. A vibration pattern is defined by the length of vibrations and breaks between each vibration.

Area of Influence. The feature described with different modalities needs an Area of Influence (AoI). The AoI is the area where a feature can automatically be described with speech or other modalities. Typically, this is an area where the user can make visual observations about the feature or a decision concerning the feature, like not going too close to it. The AoI may vary between modalities describing the same feature. The simplest way to define an area affected by a point-wise feature is by defining the radius of a feature-centered circle. Traditionally, the non-linearly interpolated geometrical shape has not been included in data interchange formats; instead, a circle has been approximated with a linearly interpolated polygon. In the latest GML versions, this absence has been corrected by adding a *CircleByCenterPoint* type. For the KML and GeoJSON schemas, we had to define such a type. Figure 3 represents the inclusion of a circle in the extension of KML.

In addition to defining an AoI by using circles, an AoI can be defined by using a simple polygon with holes. A hole can be used to denote the area inside of which a particular feature cannot be seen or to create a halo around a particular feature. A halo can be used, for example, when a user needs to make a decision before coming too close to a feature.

The AoI is only used to initiate the description. The original geometry of the feature, or its center point of gravity, can be used to determine the direction and distance to the feature, and, in some cases, its size and shape. This information can be signaled directly or indirectly. In case of non-speech audio or vibrations, a distance can be directly represented by managing the number and density of sequential sound pulses or vibrations [20]. In the case of speech, indirect reporting can be based on stereophonic sounds [22] or on managing the volume [19]. Earlier studies related to audio have even employed features that are equivalent to the AoI; Heuten, Wichmann, and Boll [19] refer to the radius of a circular AoI as a radiation radius, whereas McGookin, Brewster, and Priege [27] define it as being the size of an Audio Bubble.

Relevance from the Time Perspective. Pictures taken by cameraphones are in general only relevant for short periods of time [37]. The same applies to other modalities characterizing the changing environment. Hence, we defined validity as a common, but optional, property for all descriptions. We specified that validity should be defined by a starting moment and duration. Alternatively, the validity may be defined as recurring. One reason for doing this is to warn hikers during winter months about thin ice over treacherous streams; however, the in-built types of XML or JSON cannot be used to define weekdays. The absence of weekdays required that we define our own type, with which a hiker might be told between Monday and Friday that a nearby shop is open.

Multi-Modal Data. For textual and auditory data, we added the option of defining a language using a language tag [1], but this is not mandatory for non-spoken sounds. For auditory descriptions, we included the option to define a type that tells if the audio contains speech or an abstract non-speech sound. In case

```
<xs:complexType name="RadiusType">
  <xs:annotation>
    <xs:documentation>
      The radius of an object. The radius involves the Unit
          of Measure.
    </xs:documentation>
  </xs:annotation>
  <xs:simpleContent>
    <xs:extension base="xs:double">
      <xs:attribute name="uom" type="dkml:UnitOfLengthEnum"
          use="optional" default="m" />
    </xs:extension>
  </xs:simpleContent>
</xs:complexType>
<xs:complexType name="CircleType">
  <xs:annotation>
    <xs:documentation>
      The CircleType defines the centre and radius of a
      circle where the description is given.
      The radius type includes the Unit of Measure.
    </xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element name="Centre" type="kml:PointType" />
    <xs:element name="Radius" type="dkml:RadiusType" />
  </xs:sequence>
</xs:complexType>
<xs:complexType name="AreaOfInfluenceType">
  <xs:annotation>
    <xs:documentation>
      The AreaOfInfluenceType defines the geometrical area
          and temporal time frame when a
      description is valid and is to be given. The geometry
          may be a circle or polygon.
    </xs:documentation>
  </xs:annotation>
  <xs:sequence>
    <xs:element name="Valid" type="dkml:TemporalValidityType"
        minOccurs="0" />
    <xs:choice>
      <xs:element name="Circle" type="dkml:CircleType" />
      <xs:element name="Polygon" type="kml:PolygonType" />
    </xs:choice>
  </xs:sequence>
</xs:complexType>
```

Fig. 3. Fragment from the XSD document used to extend KML. The fragment defines the options for defining the Area of Influence.

of speech, we allowed the user to also include in the file spoken directions (in front, on the right, etc.) that can be used when the description is presented in an egocentric reference frame. The reasoning behind this solution was that Gong and Lai [15] found out in their study that users performed worse with mixed real human and synthesized speech in comparison to only synthesized speech.

For pictures, we allowed the format to inform when and where a picture has been taken and the dimension of the picture in pixels. A client can use this information to visualize the direction from which the picture has been taken and to choose between pre-scaled image alternatives. A client application may choose to only visually render textual descriptions, although it might also represent this information audio-visually by converting the text into speech with a text-to-speech (TTS) synthesizer. However, some information given by way of text may not be adequate for the auditory channel, especially if the text contains acronyms or abbreviations, because TTS engines do not typically decipher such information. Hence, auditory information may include a statement about whether or not the text can suitably be converted into speech by a TTS engine.

In the case of audio and images, we first thought about the possibility of both referencing and embedding binary data. However, embedding the binary data proved to be unsuitable because of the bandwidth restrictions. In addition, encoding images in Base64 leads to an unnecessary overhead. Consequently, the multimedia needs to be downloaded separately, and clients that typically use such media in areas with weak network connections should download the data in advance when they still have a good connection. For URL-referenced audio and video input, we made it possible to include multiple formats; the client can then choose the most suitable format. A MIME type and the length of the data in bytes can optionally accompany an URL. The MIME type is especially useful if the URL does not contain a filename extension (for example http://hostname/service/resource?id=1234). Similarly, the length is useful for a client that cannot stream media, because in the case of a large size, the client can decide to not even start the download process. The client can just ignore the reference.

4 The Portrayal Process on the Client

In this section, we present a generic way to multi-modally describe the surroundings around a mobile user. The generic process is composed of the following steps: 1) downloading, 2) storing, 3) representation, and 4) repeating the descriptions.

4.1 A Generic Approach to Gear Up

A mobile client implementation should asynchronously query the descriptive data while letting the application continue its normal operations. Successfully downloaded data needs to be de-serialized by a data parser. For instance, in the case of extended KML, an XML-related SAX (Simple API for XML) or DOM (Document Object Model) parser could be used to read in the data. The parsing process could convert each KML placemark into a application-bound feature whose model includes an extension for the descriptions. Alternatively, the parser could create key-value encoded properties that contain application-perceived keys for each descriptive property, such as *textual-desc-en-gb* for a textual description in British English. Next, the parsed objects need to be stored in a persistent, spatially indexed database. At this stage, indexing can be based on the minimum bounding rectangles calculated based on the geometries of the spatial features. Later on, the datasets changing on the server side have to be systematically synchronized with the database.

It is possible to simultaneously store features, download referenced images and audio files, and store them in the local database. However, the relevance of the downloading procedure depends on the usage context, that is, on the availability of network connections during use. Before storing referenced data, a client can make some optimizations, such as resizing the downloaded images so that they can suitably be viewed by the client. In this way, the data takes less space in the database and it is faster to read and represent.

When the data is loaded from the database into the cache of the application, one alternative for increasing the effectiveness is to apply a second spatial indexing. This index should be based on the AoIs of the features. The index does not inhibit the parallel use of an index based on the geometries because both indexes have different purposes. The AoI index can be used when the descriptions are represented automatically by the application, and the geometry-based index can be used when the user manually performs an action on a particular feature.

4.2 A Generic Approach to Representing the Descriptions

A description in different modalities is presented when the user arrives at the AoI. Subsequent repetitions of the same feature may be necessary, for example when the user

- does not notice the first description,
- arrives at the same feature, but does not interpret it as being the same, or
- changes to another user

Playback can be manual or automatic. Manual repeating requires an application that visually shows the map symbols that can be selected, or provides a method, such as playback controls.

The time upon which a description of a feature is automatically represented is stored as a temporal property value affiliated with the feature. The time value is used to manage the repetition intervals. Without a timestamp, the feature might be described continuously until the user leaves the AoI. One option is to use a single Boolean mark, which makes it possible to represent a feature only once while the feature is still loaded into the cache. However, the latter approach is impractical because some spatial features are global or very large; thus, after a while, the user might not be able to recognize a particular feature as the same one he or she experienced earlier. The reference time to which the timestamp is compared needs to be based on variables affecting the experience. Such variables are, for example, the speed of the user, the distance to the feature, and the type of the feature.

Descriptions should only be provided when the estimated accuracy of the positioning is high enough, because the accuracy affects not only the location of the feature and its AoI but also the reliability of heading towards the feature. In automated use, the AoI index is always searched when the location of the user changes and the accuracy threshold is passed. The results of the query are organized according to the distance from the actual feature to the user. From the ordered list, each feature is processed one by one until a suitable candidate is found.

A feature is unsuitable if it is invalid from a time standpoint or if its actual location is behind the user. The direction where the feature is located is calculated based on previous positioning estimates. An alternative for using positioning estimates would be to use the magnetic compass for bearings, but that is often impractical because it requires the user to hold the mobile device and point it in the direction in which he or she is moving. When the suitable feature is found, its timestamp is set to infinity. The actual updating of the timestamp is performed after the description has been represented. Just before representation the user may be informed with a vibration or sound signal.

4.3 Playback

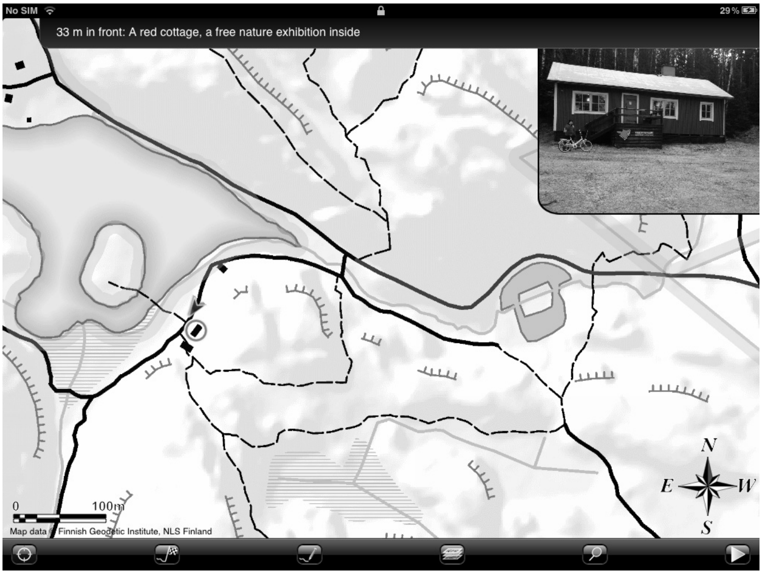
To continue the process, the following options are available:

1. The multi-modal description of another feature is immediately terminated, and the newly found feature is described
2. The description of another feature is allowed to finish, after which the newly found feature is described
3. The features that need to be described are put into a (first-in, first-out) queue
4. The features that need to be described are put into a (last-in, first-out) stack

The first alternative is especially suitable for important notices, such as warnings that are of current interest; but otherwise, the behavior can be annoying to the user. In rest of the alternatives, a description that is already running will be allowed to finish. Consequently, the three alternatives are more user-friendly because they allow the user to understand the descriptions in the form in which they were intended. The downside is that when the time comes to describe a feature, the description might not be topical anymore. To overcome this problem, the feature has to be revalidated when the time comes to represent it. In other words, the validity of the heading, whether or not the user is in the AoI, and so forth, needs to be recalculated.

The benefit of the second and fourth alternative is that the user always gets the most recent descriptions next. What is notable in the second case is that there is no waiting queue. Hence, only one feature is available at a time, and it is always replaced by a newer one, which may lead to a situation in which several features compete for the next place each time the user location changes until all of them have been described.

An intermediate system between the first and second option is a system where some descriptions are interrupted and some are not. This type of system could contain prioritized features or feature types that can be used to decide when the context requires that a description that is already running be interrupted.



33 m in front: A red cottage, a free nature exhibition inside

Fig. 4. The prototype implementation on a tablet suggests a non-obvious free nature exhibition. The distance and direction to the feature are appended in front of the description.

5 The Implementation

We implemented the client-side for the iOS operating systems that run on the iPhone phone (Figure 2) and iPad tablet (Figure 4) models. The client-side application visualizes map data in several layers. Raster-based background maps constitute the bottom layer, whereas vector-based thematic data forms the middle layer and user-added data forms the uppermost layer. Thematic data can be regarded as static or dynamic in nature. The client can regularly update dynamic data. The descriptive data is either static or dynamic thematic data, and every descriptive data collection forms its own thematic map layer, which can be turned on similarly as any other map layer. A user can add a new map layer by typing the URL of a resource, or by taking an image of an QRCode or Data Matrix containing a URL with an in-built function. The application supports GeoJSON, KML, and their descriptive extension.

When a map layer is turned on for the first time, the data for that particular layer is downloaded. Downloading and parsing is performed using a background thread. The downloaded descriptions are stored on the client side in a SpatiaLite [12] database. Indexing the database is based on the minimum bounding boxes, which are calculated based on the geometries of the features. For our case study, data connections are typically not available while on a hike, so the client can also instantly download any data that has been referenced to the client after the features have been parsed.

When a map layer containing descriptions is turned on, the data is loaded from the database into the local cache based on the extent of the buffered map view. When the map is being navigated, the cache is refreshed so that it corresponds to the new extent. For the cache, we used a second spatial indexing that is based on the AoI of the features. The index is searched whenever the location of the user changes and whenever the estimated accuracy of the positioning reveals that the A-GPS has a fix.

First, the AoI cache is searched, as described in the previous section. The implementation will determine that a feature lies in the background if the feature is not inside a sector of ±120 degrees, which has been calculated based on the heading vector. The language to be used is based on the language set on the operating system level. Thus, the language used for the descriptions can only be changed by changing the global language on the device. For the temporal playback threshold, we assigned a time limit of 30 minutes for global features and 5 minutes for local features when the users' speed is less or equal to five kilometers per hour. The corresponding thresholds decrease linearly to 10 minutes and 2 minutes when the speed increases, until a speed of 50 kilometers per hours is reached. Currently, the prototype does not use the type or the distance to the feature.

Onscreen information is given in the top edge of the screen. Textual data is located uppermost, followed by resizable pictures. The data is shown for 15 seconds. Images need to be converted into a size suitable for a particular device model. Vibrations and pre-recorded audio are delivered simultaneously using standard frameworks.

Based on the findings of Kainulainen et al. [21], presenting both speech and abstract audio (such as an auditory icon) at the same time may be hard to understand by users; hence, the system regards only one auditory description. If the feature is not associated with an auditory description, the same text is converted to an audio file using an open source TTS engine called eSpeak [11]. The TTS engine supports the languages that we need for our use case, which are English, Swedish, and Finnish.

Before the text or audio can be represented, the textual or spoken direction (which is calculated based on the heading to the feature) and distance must be concatenated to the record. In case of pre-recorded audio, the system uses pre-recorded directions if such are given. But in case of synthesized speech, only synthesized directions are concatenated to keep the output consistent and to lower the cognitive processing demand of the user.

At the same time, the feature is highlighted if it is visible, otherwise the location is pointed out by a short animation. If a description is already playing, we put the new description in a queue. Visible features can also be selected by the user at any point. Unlike an automatically given description, a user-initiated description terminates any ongoing description.

The feasibility of the presented format was affirmed from the technical perspective by first collecting the descriptions of non-obvious features that are of general interest, found in the National Park, into the proposed data interchange formats. The forms of descriptions included multi-lingual text, recorded speech, synthesized speech, and pictures. Next, the data was loaded to the mobile client with varying mechanisms, including the use of QRCodes that can be printed at the sites where National Park is entered. Finally, the descriptions were confirmed to be given when a user entered a feature-bound Area of Influence.

6 Discussion and Conclusions

This paper has presented the technical background and implementation of a Location-Based Service to extend the user experience that a visitor of a natural park may experience while using a mobile device. However, in this paper we do not address the aspects of usability from the end user perspective, that is, what the users would prefer as the AoI, or how the end users feel about the different modalities. Nevertheless, from related multi-modal turn-by-turn navigation experiments, we know that users prefer different modalities in different situations. For instance, according to the study of Liljedahl et al. [25], users mostly rely on the visual channel, but prefer auditory direction notification.

The solution does not take into account particular vibration patterns, because we are unaware of a coding scheme that would be generally accepted. If feature types are encoded into patterns these need to be known by a client and taught to the user. Neither does our proposal define audio-related properties for textual information that can be converted to speech by a TTS engine. Such properties could include a wish to represent the information using a certain voice (that is based on some characteristics, e.g., gender or age), tone (e.g. personal or enthusiastic), or volume. For example, Caquard et al. [6] use a second voice, in addition to the authoritative “voice-of-god” to present non-objective information from a personal perspective, like opinions and doubts. Graham and Cheverst [16] go even beyond two voices by presenting five personified interaction paradigms for guiding applications.

There are still many challenges to overcome. A question is how the descriptions of different kinds of modalities are created, extended and kept up-to-date. One solution may be collaborative data maintenance by the park visitors. A more futuristic vision is that the users surrounding is not only described human-to-human. Instead, locally embedded sensors may provide additional information, such as images of the latest animals passing the area or popularity of paths. Similarly, technical development may extend the usability of modalities. For example, in a couple of years, off-the-shelf smartphones might have touch screens incorporating haptic Braille support.

Acknowledgements. This survey is a part of the UbiMap project which is funded by the Academy of Finland, Motive program, and is carried out in cooperation with the FGI, Department of Geoinformatics and Cartography, and the University of Helsinki, Department of Cognitive Science. The authors are also grateful to the European Commission which co-funded the IP HaptiMap (FP7-ICT-224675).

References

1. Alvestrand, H.: Tags for the Identification of Languages, RFC 1766, Network Working Group, Internet Engineering Task Force (1995)
2. Aslan, I., Schwalm, M., Baus, J., Krüger, A., Schwartz, T.: Acquisition of Spatial Knowledge in Location Aware Mobile Pedestrian Navigation Systems. In: Proceedings of the 8th Conference on Human-Computer Interaction With Mobile Devices and Services, MobileHCI 2006, Helsinki, Finland, pp. 105–108. ACM, New York (2006)
3. Baus, J., Wasinger, R., Aslan, I., Krüger, A., Maier, A., Schwartz, T.: Auditory perceptible landmarks in mobile navigation. In: Proceedings of the 12th International Conference on Intelligent User Interfaces, IUI 2007, Honolulu, Hawaii, USA, pp. 302–304. ACM, New York (2007)
4. Blattner, M.M., Sumikawa, D.A., Greenberg, R.M.: Earcons and icons: Their structure and common design principles. Human-Computer Interaction 4, 11–44 (1989)
5. Butler, H., Daly, M., Doyle, A., Gillies, S., Schaub, T., Schmidt, C.: The GeoJSON Format Specification, Revision: 1.0 (2008)
6. Caquard, S., Brauen, G., Wright, B.: Exploring sound design in cybercartography. In: Proceedings of the 22nd International Cartographic Conference (ICC), July 9-16 (2005)
7. Cheverst, K., Davies, N., Mitchell, K., Friday, A., Efstratiou, C.: Developing a context-aware electronic tourist guide: some issues and experiences. In: Proceedings of the International Conference on Human-Computer Interaction, CHI 2000, The Hague, The Netherlands, pp. 17–24. ACM, New York (2000)
8. Chittaro, L., Burigat, S.: Augmenting audio messages with visual directions in mobile guides: an evaluation of three approaches. In: Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services, MobileHCI 2005, Salzburg, Austria, pp. 107–114. ACM, New York (2005)
9. Ciavarella, C., Paternò, F.: The design of a handheld, location-aware guide for indoor environments. Personal and Ubiquitous Computing 8(2), 82–91 (2004)
10. Crockford, D.: The application/json Media Type for JavaScript Object Notation (JSON), RFC 4627, Network Working Group, The Internet Engineering Task Force (2006)
11. eSpeak: Speech Synthesizer, version 1.46.27 (2012), `http://espeak.sourceforge.net`
12. Furieri, A.: SpatiaLite, version 3.0.0 (2011), `http://www.gaia-gis.it/fossil/libspatialite/index`
13. Gaver, W.W.: Auditory icons: Using sound in computer interfaces. Human-Computer Interaction 2, 167–177 (1986)
14. Ghiani, G., Leporini, B., Paternò, F.: Supporting orientation for blind people using museum guides. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2008, Florence, Italy, pp. 3417–3422. ACM, New York (2008)

15. Gong, L., Lai, J.: To Mix or Not to Mix Synthetic Speech and Human Speech? Contrasting Impact on Judge-Rated Task Performance versus Self-Rated Performance and Attitudinal Responses. International Journal of Speech Technology 6(2), 123–131 (2003)
16. Graham, C., Cheverst, K.: Guides, locals, chaperones, buddies and captains: managing trust through interaction paradigms. In: Workshop on Mobile Guides at the Fifth International Symposium on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2004, Glasgow, Scotland (2004)
17. Guy, R., Truong, K.: Crossingguard: exploring information content in navigation aids for visually impaired pedestrians. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2012, Austin, Texas, USA, pp. 405–414. ACM, New York (2012)
18. Hameseder, K., Fowler, S., Peterson, A.: Performance analysis of ubiquitous web systems for SmartPhones. In: International Symposium on Performance Evaluation of Computer Telecommunication Systems, SPECTS 2011, pp. 84–89 (2011)
19. Heuten, W., Wichmann, D., Boll, S.: Interactive 3d sonification for the exploration of city maps. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, NordiCHI 2006, Oslo, Norway, pp. 155–164. ACM, New York (2006)
20. Holland, S., Morse, D.R., Gedenryd, H.: AudioGPS: Spatial audio navigation with a minimal attention interface. Personal and Ubiquitous Computing 6(4), 253–259 (2002)
21. Kainulainen, A., Turunen, M., Hakulinen, J., Melto, A.: Soundmarks in spoken route guidance. In: Proceedings of the 13th International Conference on Auditory Display, Montreal, Canada (2007)
22. Kan, A., Pope, G., Jin, C., van Schaik, A.: Mobile Spatial Audio Communication System. In: Proceedings of ICAD 2004-Tenth Meeting of the International Conference on Auditory Display (July 2004)
23. Kangasharju, J., Tarkoma, S.: Benefits of alternate XML serialization formats in scientific computing. In: Proceedings of the 2007 Workshop on Service-Oriented Computing Performance: Aspects, Issues, and Approaches, SOCP 2007, Monterey, California, USA, pp. 23–30. ACM, New York (2007)
24. Krüger, A., Aslan, I., Zimmer, H.D.: The Effects of Mobile Pedestrian Navigation Systems on the Concurrent Acquisition of Route and Survey Knowledge. In: Brewster, S., Dunlop, M.D. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 446–450. Springer, Heidelberg (2004)
25. Liljedahl, M., Lindberg, S., Delsing, K., Polojärvi, M., Saloranta, T., Alakärppä, I.: Testing Two Tools for Multimodal Navigation. In: Advances in Human-Computer Interaction (2012)
26. Mabrouk, M.: OpenGIS® Location Services (OpenLS): Core Services, Document OGC 07-074, Version 1.2, Open Geospatial Consortium Inc. (2008)
27. McGookin, D., Brewster, S., Priego, P.: Audio Bubbles: Employing Non-speech Audio to Support Tourist Wayfinding. In: Altinsoy, M.E., Jekosch, U., Brewster, S. (eds.) HAID 2009. LNCS, vol. 5763, pp. 41–50. Springer, Heidelberg (2009)
28. Münzer, S., Zimmer, H.D., Schwalm, M., Baus, J., Aslan, I.: Computer-assisted navigation and the acquisition of route and survey knowledge. Journal of Environmental Psychology 26(4), 300–308 (2006)
29. Oppermann, R., Specht, M.: A nomadic information system for adaptive exhibition guidance. In: Proc. of the International Conference on Hypermedia and Interactivity in Museums, pp. 103–109 (1999)

30. Petrelli, D., Not, E., Sarini, M., Stock, O., Strapparava, C., Zancanaro, M.: Hyperaudio: location-awareness + adaptivity. In: CHI 1999 Extended Abstracts on Human Factors in Computing Systems, CHI EA 1999, pp. 21–22. ACM, New York (1999)
31. Pielot, M., Henze, N., Boll, S.: Supporting Map-Based Wayfinding with Tactile Cues. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2009, Bonn, Germany, pp. 1–10. ACM, New York (2009)
32. Pielot, M., Poppinga, B., Boll, S.: Understanding tourists on a bicycle trip "in the wild". In: Mobile Living Labs Workshop in Conjunction with MobileHCI 2009 (2009)
33. Richter, K.-F.: A Uniform Handling of Different Landmark Types in Route Directions. In: Winter, S., Duckham, M., Kulik, L., Kuipers, B. (eds.) COSIT 2007. LNCS, vol. 4736, pp. 373–389. Springer, Heidelberg (2007)
34. Rodrigues, C., Afonso, J., Tomé, P.: Mobile Application Webservice Performance Analysis: Restful Services with JSON and XML. In: Cruz-Cunha, M.M., Varajão, J., Powell, P., Martinho, R. (eds.) CENTERIS 2011, Part II. CCIS, vol. 220, pp. 162–169. Springer, Heidelberg (2011)
35. Sumaray, A., Makki, S.K.: A comparison of data serialization formats for optimal efficiency on a mobile platform. In: Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2012, Kuala Lumpur, Malaysia, pp. 48:1–48:6. ACM, New York (2012)
36. van den Brink, L., Portele, C., Vretanos, P.A.: Geography Markup Language (GML) simple features profile (with Corrigendum), Version 2.0, OGC® 10-100r3, Open Geospatial Consortium (2012)
37. Van House, N., Davis, M.: The Social Life of Cameraphone Images. In: Proceedings of the Pervasive Image Capture and Sharing: New Social Practices and Implications for Technology Workshop (PICS 2005) at the Seventh International Conference on Ubiquitous Computing (UbiComp 2005), Tokyo, Japan (2005)
38. Wilson, G.: Multimedia tour programme at tate modern. In: Bearman, D., Trant, J. (eds.) Proceedings of the Museums and the Web 2004. Archives & Museum Informatics, Toronto (2004)
39. Wilson, T.: OGC®KML, Version: 2.2.0, Document 07-147r2, Open Geospatial Consortium (2008)
40. Zyp, K.: A JSON Media Type for Describing the Structure and Meaning of JSON Documents, Version 03, Internet-Draft, Internet Engineering Task Force (2010)