Integrating Sketch and Speech Inputs using Spatial Information

Bee-Wah Lee
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
Sarawak, Malaysia
bwlee@fit.unimas.my

Alvin W. Yeo
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
Sarawak, Malaysia
alvin@fit.unimas.my

ABSTRACT

Since the development of multimodal spatial query, the integration technique in determining the correct pair of multimodal inputs remains a problem in multimodal fusion. Although there exist integration techniques that have been proposed to resolve this problem, they are limited to the interaction with predefined speech and sketch commands. Furthermore, they are only designed to resolve the spatial query with single speech input and single sketch input. Therefore, when it comes to the introduction of multiple speech and sketch inputs in a single query, all the existing integration techniques are unable to resolve it. To date, no integration technique has been found that can resolve the Multiple Sentences and Sketch Objects Spatial Query. In this paper, the limitations of the existing integration techniques are discussed. A new integration technique in resolving this problem is described and compared with the widely used integration technique, Unification-based Integration Technique.

Categories and Subject Descriptors

H. [**Information Systems**]: H.1: Models and Principles: H.5 Information Interface and Presentation (I.7)

General Terms

Design, Experimentation, Human Factors, Measurement, Performance

Keywords

Multimodal interaction, multimodal spatial query, spatial query, multimodal spatial scene description, and Multiple Sentences and Sketch Objects Spatial Query

1. INTRODUCTION

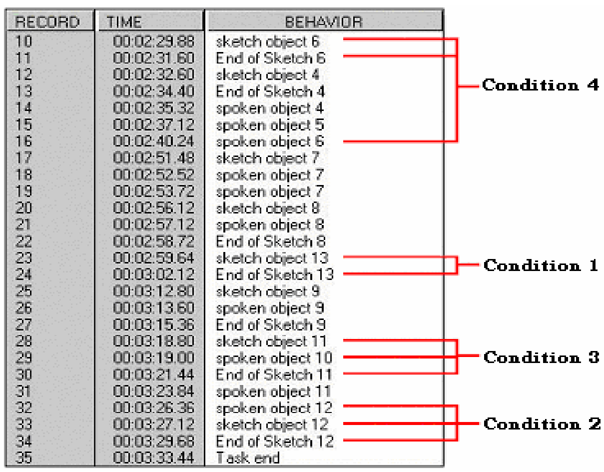
Since the conventional Geographic Information System (GIS) applications are often difficult to use and take a long time to learn [2, 3], there is a need to make these applications easier to use especially by novice users. According to Schlaisich and Egenhofer [7], people often communicate about space by talking and simultaneously drawing freehand sketches. This natural human communication pattern can be adapted into the existing GIS applications to accommodate a wider range of users and for better ease-of-use. This natural human communication can be achieved by adding multimodal interactions such as the use of pen gestures and speech in the process of spatial query formulation. Currently, most of these multimodal spatial systems are only able to accept single speech input and single sketch input in the spatial query [6,7]. The most widely used multimodal integration technique in this context is Unification-based Multimodal Integration Techniques.

However, when responding to the pedestrian who asks for direction on the street, users' sketch and verbal descriptions tend to occur more frequently in continuous spoken stream and freehand sketch. In this situation, multiple speech sentences and multiple sketch objects are involved in the single query formulation. Unlike the single sentence and single sketch object spatial query, this multiple-input spatial query would result in more problems in identifying the correct pair of speech and sketch inputs, which refer to the same spatial object. Furthermore, to date no multimodal integration technique exists for this context. Thus, a user survey was conducted to determine the suitability of the existing integration techniques to resolve this type of Multiple Sentences and Multiple Sketch Objects Spatial Query. The existing techniques were found to be insufficient in resolving this type of spatial query. The limitations of the existing integration techniques are discussed in the next section.

2. LIMITATIONS OF THE UNIFICATION-BASED MULTIMODAL INTEGRATION TECHNIQUE

In Unification-based Integration Technique, a temporal constraint is used to obtain the correct pair of speech and sketch events for a spatial object. The temporal constraint states that the time of the speech input occurrence must either overlap with the time interval of sketch input or the onset of the speech input is within 4 seconds following the end of the sketch input [5,6]. Therefore, the time interval within the occurrences of the inputs is used as the integration parameter in this technique. The occurrences of specific keywords or command in speech are detected and time stamped accordingly based on the temporal constraint. The sketch inputs are also time stamped and matched with the predefined simple gesture and symbols in the database. Consequently, the users' inputs are restricted and users have to remember and use only the special speech commands and the special pen gestures that can be accepted by the system.

The most significant limitation in using time interval in this technique is that it can only support single command query and not a continuous description. Undeniably, this technique works well in situations where only single speech input and single gesture input occurred in the query. The single sketch input can be directly associated with the only speech input. However, if this integration technique is used in multimodal spatial scene description (with the occurrences of multiple sentences and multiple sketch objects), the wrong integration of the inputs from different events can easily occur. Four conditions that fail in fulfilling the temporal constraint are illustrated in Figure 1.



RECORD	TIME	BEHAVIOR
10	00:02:29.88	sketch object 6
11	00:02:31.60	End of Sketch 6
12	00:02:32.60	sketch object 4
13	00:02:34.40	End of Sketch 4
14	00:02:35.32	spoken object 4
15	00:02:37.12	spoken object 5
16	00:02:40.24	spoken object 6
17	00:02:51.48	sketch object 7
18	00:02:52.52	spoken object 7
19	00:02:53.72	spoken object 7
20	00:02:56.12	sketch object 8
21	00:02:57.12	spoken object 8
22	00:02:58.72	End of Sketch 8
23	00:02:59.64	sketch object 13
24	00:03:02.12	End of Sketch 13
25	00:03:12.80	sketch object 9
26	00:03:13.60	spoken object 9
27	00:03:15.36	End of Sketch 9
28	00:03:18.80	sketch object 11
29	00:03:19.00	spoken object 10
30	00:03:21.44	End of Sketch 11
31	00:03:23.84	spoken object 11
32	00:03:26.36	spoken object 12
33	00:03:27.12	sketch object 12
34	00:03:29.68	End of Sketch 12
35	00:03:33.44	Task end

Figure 1. Different types of conditions occurring in the spatial description

The first condition (condition 1) is the absence of speech or sketch events for a spatial object. This condition occurs when users did not mention or sketch the spatial object. The second condition (condition 2) is where the speech event occurs before the onset of the sketch event for the same spatial object. Since the algorithm in Unification-based Integration Technique would only locate the speech event for the spatial object after or during the user's sketch event, this speech event is not successfully found though it actually occurred. The third condition (condition 3) occurs when the wrong pair of speech and sketch events is integrated. This condition normally happens when users described more than two spatial objects while performing a sketch event. The last condition (condition 4) is where speech or sketch event for a spatial object does not occur within the time window (4 seconds). This occurrence is directly discarded when using this Unification-based Integration Technique.

Given the deficiencies of the available integration technique applied in the spatial scene description domain, a new technique using user inputs' spatial information is explained in the next section. This technique is known as Spatial Information Integration Technique.

3. THE USE OF SPATIAL INFORMATION IN MULTIMODAL INTEGRATION TECHNIQUE

In this section, a new integration technique is introduced to integrate multiple sentences and sketch inputs in a spatial query. In order to integrate speech and sketch inputs, the characteristics of pen gesture and speech modalities are discussed to determine how these modalities could be integrated into a single query. Initially, the sketch itself can reflect a spatial configuration with its topological, relative direction, and relative distance relations. On the other hand, the speech is able to complement the sketch produced by users. For instance, the speech statement *A is in B* can be associated with the sketch produced by users where there is an object A that *contains* another object B. Therefore, if a sketch can be associated with speech in a better manner such as using the spatial relations, it should enhance the ability of sketch as an efficient spatial query method.

Although the existing techniques discussed in the previous section also utilize the complementary relationship between speech and sketch, they are not suitable in solving the spatial queries with multiple speech and sketch inputs. If there are multiple inputs occurring almost at the same time in one spatial query, the four conditions that lead to the failure of integration might occur. The integration of wrong pair of inputs or discarding of correct inputs would happen. Since the spatial query is constructed from the users' described spatial scene, if the wrong pair of inputs is integrated, the spatial query would search for non-existing spatial objects in the map database. On the other hand, if valid inputs are discarded, useful information in searching the database is wasted. As such, the result of the query might be affected.

Therefore, the Spatial Information Integration Technique introduced in this paper would only interpret semantic meaning of inputs after the whole spatial scene is described. The information represented by the inputs is viewed as a whole so that all inputs would be taken into consideration. In addition, the technique matches the correct pair of inputs by utilizing the spatial information from both sketch and speech inputs. The spatial information contained in both inputs are compared and matched based on their spatial relations in each modality. The main consideration is attributed to the spatial arrangement of the spatial objects of the sketch input. One of the benefits of matching through spatial information is that no speech or sketch objects would be left out in formulating the spatial query. Since no object sequence or time stamping is used to identify the object occurrences, it would be difficult to mix up objects in the spatial scene. Furthermore, since the inputs are compared through their semantic meaning, the semantic meaning is more reliable in matching inputs than other parameters such as the temporal sequence of input occurrence and time interval for each input.

In order to integrate inputs from pen gesture and speech modalities, a multimodal integration architecture is built to capture and integrate inputs. The multimodal integration architecture is illustrated in Figure 2.

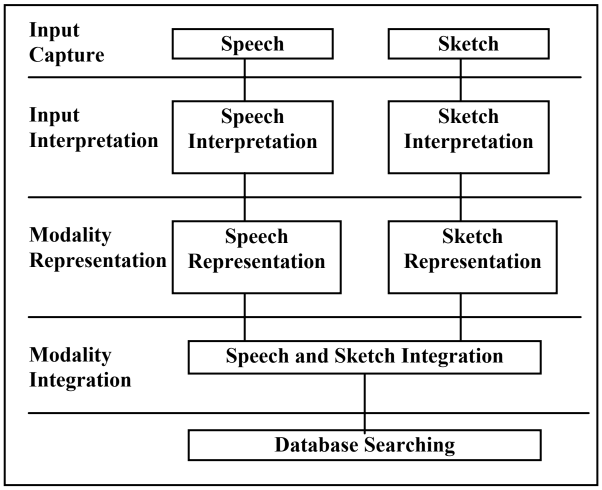


Figure 2. Multimodal Integration Architecture

In Figure 2, speech and sketch inputs are captured through speech and pen gesture recognizers in Input Capture phase. Dragon Naturally Speaking 7 speech recognizer was used to convert user's speech input to text format. Headset microphone was also used to input voice into the computer system. Meanwhile, users produced the sketch on the sketching interface in ArcGIS software, which can be directly converted into a digital sketch. After the Input Capture phase, the captured inputs are then processed in the Input Interpretation phase to extract the useful spatial and non-spatial information. This information is then used to build representation structures for the speech and sketch input in the Modality Representation phase. After this, the integration techniques are applied in Modalities Integration phase to obtain the final integration representation structure, which incorporates information from both speech and sketch representation structures.

The integrated representation structure is then used in Database Searching phase. The searching in the database is conducted by comparing the information from the integrated representation structure with the temporary relations established for the spatial objects in a map. Results of the searching include portions of the map, which match the information contained in the integrated representation structure. These map portions are referred to as spatial scenes in this paper.

3.1 Spatial Information Extraction Method

The spatial information extraction method is used to extract spatial information in speech and sketch inputs. For speech input, all words in the speech inputs are associated with the grammar groups they belong to. Based on the grammar groups, these words are categorized into the groups of spatial information and non-spatial information. Words that reveal the information related to the setting and spatial representation of object distribution is categorized in the spatial information group. This group normally corresponds to the preposition grammar group. However, not all prepositions are accepted as spatial information. According to basic grammar rules, there are different types of preposition such as preposition of time (for, since, at, on, by), preposition of place (at, on, in), preposition of movement (to), and preposition of location (on, across, inside, between). Only the preposition that is used to describe location (preposition of location) is accepted as an element of spatial information. On the other hand, all the other words (noun, verb, adjective, and determiner) are identified as non-spatial information. The spatial information that is accepted in the spatial information group can be referred in Figure 5 and Figure 6.

On the other hand, for sketch inputs, the spatial information can be retrieved using the spatial database management system (ArcGIS), which provides the sketching interface for users. Through the use of this system, relative directional, topological, and relative distance relations for every spatial object in the sketch are established. Each spatial object has relations with all other objects on the sketch in this relation establishment. The number of relations for each spatial object can be calculated using the formula shown in Figure 3.

For relative directional relations, the relations established between two spatial objects are based on the eight standard directional measures: *north*, *northeast*, *east*, *southeast*, *south*, *southwest*, *west*, and *northwest*. Topological relations are based on eight distinct topological relations suggested by Egenhofer and Franzosa [3]. These relations are *disjoint*, *meet*, *overlap*, *equal*, *contains*, *cover*, *inside*, and *cover by*. After establishing the relations for each spatial object, the result is used in the Modality Representation phase.

Establishment of Directional and Topological Relations
Number of relation(s) = n(n-1)/2
where n = number of spatial objects

Figure 3. Number of directional and topological relations for each spatial object on sketch

3.2 Modality Representation

In this section, Modality Representation Structure is built based on the spatial information extracted using the spatial information extraction method. This structure contains linkages between all identified objects in that modality. Consequently, there are two structures built at the end of this phase, one for speech representation and another for sketch representation.

Speech Representation Structure is the representation of all spatial objects that was spoken by users. The attribute information of the objects such as object name and object tag are appended to the structure as shown in Figure 4. Furthermore, these objects are interconnected through the spatial relationship (spatial information) extracted earlier. Similarly, Sketch Representation Structure has exactly the same structure as the Speech Representation Structure except the sketch has no attribute information. However, sketch can provide spatial information such as topological relations, relative direction, and relative distance among the sketched objects. The general structure of sketch representation structure, which is based on two spatial objects is shown in Figure 4.

After building both speech and sketch representation structures, the spatial relations in sketch and the spatial information in speech can be matched together to form an integrated representation structure. The next section describes the integration technique employed to combine these speech and sketch representation structures.

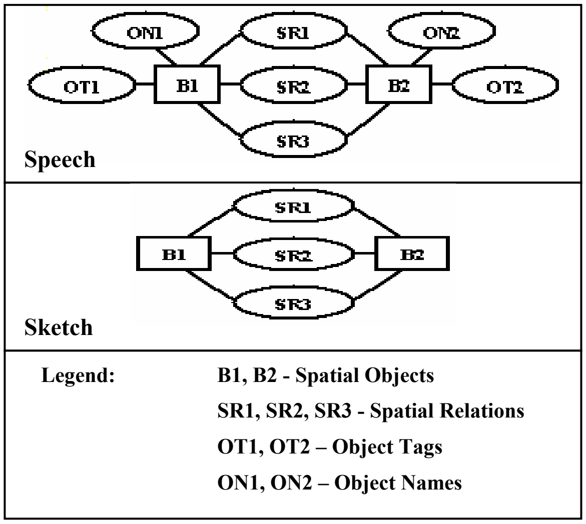


Figure 4. Speech and Sketch Representation Structures

3.3 Multimodal Integration

In this section, the information associated with spatial objects in the Modalities Representation Structures is used. The overall process in this phase is depicted in a spatial information-mapping model. The spatial information-mapping model consists of two sub-models: topological model and directional model. These models contain the possible spatial information that can be retrieved from speech and its matched spatial equivalent. Firstly, spatial information in Speech Representation Structure is categorized into topological and directional information. Topological information is mapped with topological model and the directional information is submitted into the directional model. In these models, a set of spatial equivalents is established for the spatial information. The details of these topological and directional models are shown in Figure 5 and Figure 6 respectively.

After going through both processes in spatial models, the speech's topological and directional equivalents appear as the same phrase in sketch's topological and relative directional information. In short, the topological and directional models have converted the speech spatial information into phrases that are similar to the spatial relations in sketch. Therefore, with two sets of same spatial phrases, a comparison can be made among the spatial objects. Based on the comparison result, a set of spatial objects is formed. An Integrated Representation Structure is built by including the integrated spatial objects. This Integrated Representation Structure has a similar appearance with the Modality Representation Structures discussed in the earlier section. Non-spatial information from Speech Representation Structure is appended to the spatial objects in this integrated structure.

4. RESULT AND ANALYSIS

Upon the completion of the design and implementation of this Spatial Information Integration Technique, an evaluation is conducted. The main objective of this evaluation is to validate that the proposed Spatial Information Integration Technique is a better solution in locating a user described spatial scene from a map database compared to the Unification-based Integration Technique mentioned in the earlier section. The Unification-based Integration Technique is used as the benchmark since it is widely used as the integration technique in spatial domain. A user survey with 30 subjects was conducted to obtain samples of speech and sketch inputs. Every subject was required to complete one task as shown in Figure 7. This task is directly adopted from Blaser [1]. Although the Blaser's survey context was slightly different with the one included in this paper, both surveys were aimed at getting user description about a location in a map.

Based on the data gathered from this survey, three types of analyses were conducted to compare the results obtained from Spatial Information Integration Technique and Unification-based Integration Technique. These three analyses are discussed in the next section.

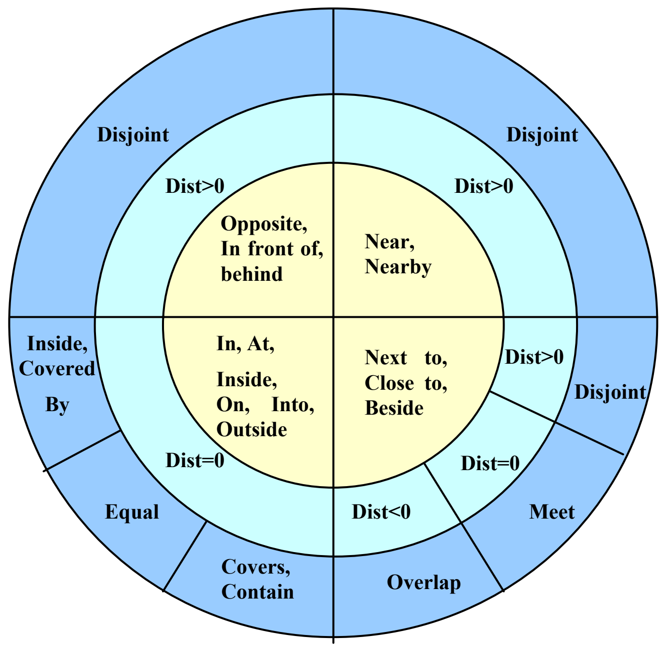


Figure 5. Topological Model

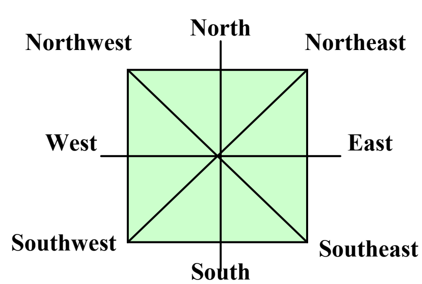


Figure 6. Directional Model

Recall your last vacation and imagine you have lost your leather case with your airplane tickets while shopping in a store in one of the town you visited. Unfortunately you do not realize your loss until you are at the airport. At the airport security office, you make a sketch for the officer in charge, explaining all details that are important to locate the specific store because you cannot recall the original name of the store.

Figure 7. The spatial task (Blaser, 1998)

4.1 Integration Accuracies Analysis

Every spatial description (comprising speech and sketch inputs) produced by subjects was submitted to both the Spatial Information Integration Technique and the Unification-based Integration Technique. The integration accuracies for each subject's spatial description are calculated independently in both integration techniques. The accuracies of both integration techniques in resolving the correct pair of speech and sketch inputs are shown in Table 1 below.

Table 1. Summary of the integration accuracies

Analysed items	Unification-based Integration Technique		Spatial Information Integration Technique	
	Result	Success rate (%)	Result	Success rate (%)
Integration accuracy	0.36	36	0.52	52
Integrated spatial object	0.36 (=91/254 spatial objects)	36	0.50 (=126/254 spatial objects)	50
Integrated spatial description	0.63 (=19/30 spatial description)	63	1.00	100

The integration accuracies shown in Table 1 is converted into Figure 8 to show the maximum, minimum, and mean of the integration accuracies obtained in each integration technique. In Figure 8, the average integration accuracy in the Spatial Information Integration Technique is higher than the Unification-based Integration Technique. More importantly, the integration accuracies achieved in Unification-based Integration Technique is more spread out and there is a bigger difference between the maximum (100% integration accuracy) and the minimum (0% integration accuracy). This result shows that the integration with Unification-based Integration Technique is not stable and it varied significantly from one spatial description to another.

4.2 T-test Analysis

Besides that, a t-test analysis is conducted to compare the means of integration accuracies in Unification-based Integration Technique and Spatial Information Integration Technique. The purpose of performing this t-test is to identify the better integration technique. With 95% of confidence interval, the one-tailed test is conducted for the hypothesis as shown in Figure 9. The calculated paired t ratio is checked for the hypothesis H_a: $\mu_S > \mu_U$, using 29 degree of freedom and one-tailed test. At the 0.050 level, $t_{.050} = 1.699$. Since the calculated t value is 2.196, which is greater than 1.699 (p-value in this case is 0.018), it is included in the rejected region and the H_0 is rejected. A conclusion can be made where Spatial Information Integration Technique appears better than the Unification-based Integration Technique.

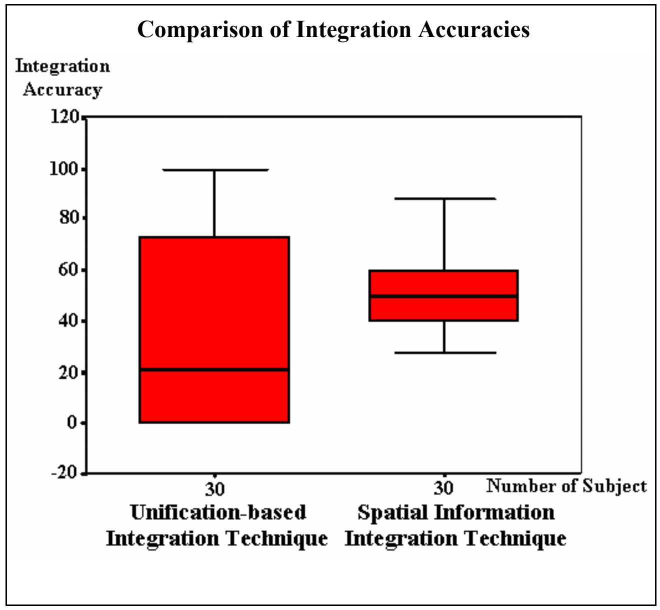


Figure 8. Comparison of Integration Accuracies

Null hypothesis: H_0: $\mu_S = \mu_U$
Alternative hypothesis: One-Tailed Test H_a: $\mu_S > \mu_U$

$_S$ = Integration accuracy of Spatial Information Integration Technique
$_U$ = Integration accuracy of Unification-based Integration Technique

Figure 9. Null and alternative hypothesis in T-test

4.3 Integration Failures Analysis

Based on the results obtained in the section above, not all spatial objects in a spatial description were successfully integrated. Several reasons that may have led to the failure of integrating the speech and sketch events for spatial objects are analyzed in this section.

Firstly, in Unification-based Integration Technique, as mentioned in the previous section, the speech and sketch events for a spatial object are integrated if the temporal constraint is fulfilled. The temporal constraint states that the time of the speech input occurrence must either overlap with the time interval of sketch input or the onset of the speech input is within 4 seconds following the end of the sketch input [5,6]. However, only 91 out of 254 spatial objects were successfully integrated using this integration technique. In other words, only 91 spatial objects fulfilled the constraint used in this integration technique. There were 163 spatial objects, which failed to be integrated (refer to Table 1). After performing an analysis on the speech and sketch events for these spatial objects, four conditions could be derived from the way these speech and sketch events occurred (included previously in literature review). These four types of conditions contribute to the violation of temporal constraint in Unification-based Integration Technique, which leads to the integration failure in subjects' speech and sketch inputs. The percentages of the distribution of the spatial objects in these four types of failure conditions are shown in Figure 10.

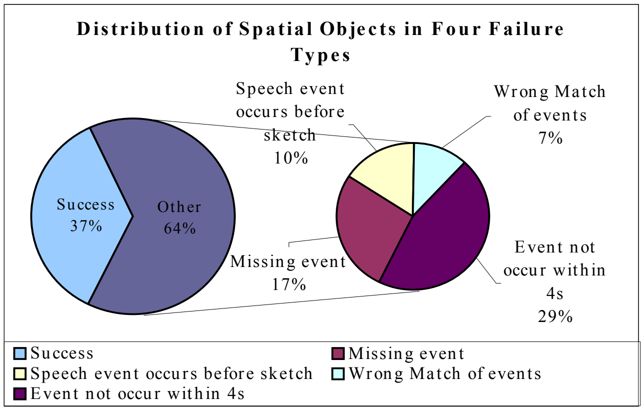
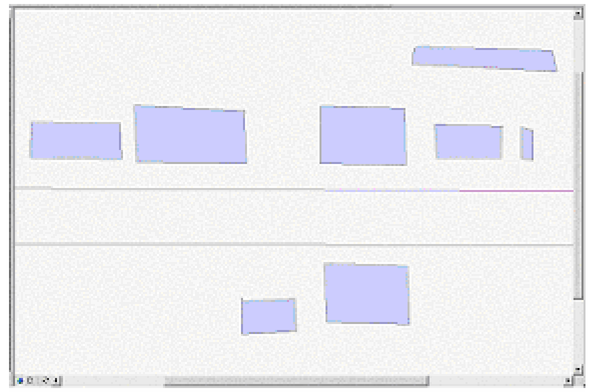


Figure 10. Comparison of Integration Accuracies

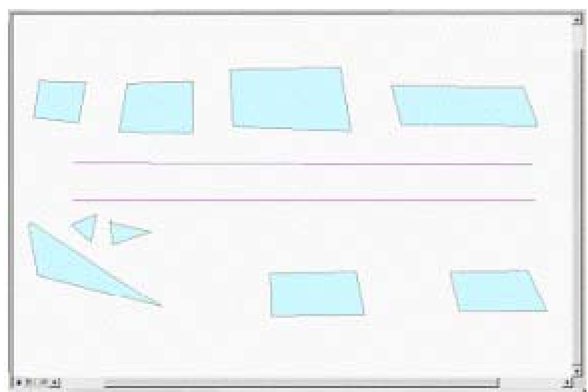
In order to use this Spatial Information Integration Technique, there must exist the spatial information from speech and sketch inputs. For sketch input, the spatial relations between spatial objects in sketch can be established directly after subjects completed the whole sketch. However, not all of the spatial objects are described with the spatial information with other spatial objects in the sketch. Three spatial descriptions were taken from the survey to show the cause of failure in using this Spatial Information Integration Technique. The first spatial description serves as the best case that achieved 87.50% integration accuracy (as shown in Figure 11). However, for another two spatial descriptions, the integration accuracies are low. There are several causes that have led to the occurrence of low integration accuracy. The first cause is the limited amount of spatial information included in the speech input as shown in Figure 12 (with 50% integration accuracy). Another cause is the limited amount of speech and sketch inputs provided in the spatial description as shown in Figure 13 (with 40% integration accuracy).

Based on the analysis done in comparing the Spatial information Integration Technique and Unification-based Integration Technique, the results analyzed from both integration techniques suggest that Spatial Information Integration Technique is a better integration technique in the Multiple Sentences and Multiple Sketch Objects Spatial Query. The discussion of these analyzed results is detailed in the next section.



Speech: Actually I lost my leather case in Sarawak Plaza in Kuching. I cannot really remember. But what I remember is the Sarawak Plaza is opposite the Tun Jugah. Okay, this is Tun Jugah. At the east of the Sarawak Plaza is Holiday Inn Hotel. Then at the northeast of the Sarawak Plaza is also a shopping complex named Centre Point. Next to the Sarawak Plaza is Maybank and on the east of this Maybank is Pizza Hut. I think next to Tun Jugah there is McDonald. And next to Holiday Inn is Riverbank Suite. I think that is all I can recall.

Figure 11. Best case spatial descriptions from the survey



Speech: The building that I lost my leather case and ticket is situated here. But I cannot remember the name of the building. All I can remember is this building is a part of SEDC building. At the west side of this building is Holiday Inn Hotel and there is Maybank somewhere nearby. Opposite this building is also a shopping Complex named Tun Jugah. There is cat statue somewhere here.

Figure 12. Spatial description with limited spatial information

Speech: This is the Sarawak Plaza. And here is Tun Jugah, at the southwest of Sarawak Plaza. This is the main entrance here, and I lost my leather case here. This is a traffic light. And this is the new store in front of Sarawak Plaza named Kenny Rogers. That is all.

Figure 13. Spatial description with limited speech and sketch inputs

5. DISCUSSION

Based on the result and analysis detailed in the section above, several aspects on the improvements of the Spatial Information Integration Technique against the existing integration techniques are discussed. Firstly, the most significant improvement is that the Spatial Information Integration Technique can resolve the Multiple Sentences and Multiple Sketch Objects Spatial Query. As identified in the Introduction section, this type of spatial query represents a new research area and there is no integration technique developed for this type of spatial query. Furthermore, the existing integration technique such as Unification-based Integration Technique is not as suitable when used in this context. Therefore, the ability of the Spatial Information Integration Technique in resolving this type of spatial query is an important improvement in this area.

In addition, since the Spatial Information Integration Technique can accept unconstrained speech and freehand sketch, it permits the flexible use of these modalities. It allows users to have more freedom in constructing the spatial query and users do not have to translate the spatial images in their mind to the structured languages as explained by Egenhofer [3]. With the development of this Spatial Information Integration Technique, users do not have to go through training or have to memorize particular syntax in constructing the spatial query. Indirectly, this integration technique improves the way users communicate and issue their requests to the spatial system. Users are no longer constrained to any predefined action or procedure in submitting their requests to the system. Thus, another improvement achieved by the Spatial Information Integration Technique is that it promotes and supports more natural human-computer interaction. Since the usage of this Spatial Information Integration Technique is similar to the natural way of human communication, users can use the spatial system directly without any prior knowledge in this area.

Furthermore, not all spatial objects in a spatial description were successfully integrated with the correct pair of speech and sketch events submitted by users. This failure in integrating the correct pair of inputs serves as the main problem encountered in existing integration technique. However, this Spatial Information Integration Technique is a solution to improve this problem as it employed a new integration parameter in the integration process. Since the problem is mainly caused by the usage of the temporal constraint (time-out interval) in the integration process, the spatial information integration parameter in the Spatial Information Integration Technique does not inherit this major drawback.

Lastly, another improvement of this Spatial Information Integration Technique is it promotes the post-processing input capturing method. In the existing integration techniques, speech and sketch inputs are captured and processed during the user interaction before the entire session finished. This input processing method easily leads to errors. For example, as observed in Unification-based Integration Technique, useful speech or sketch inputs might be accidentally or purposely discarded when they occurred at the wrong time during the user's interaction. With the use of the post-processing method as supported in this Spatial Information Integration Technique, the speech and sketch inputs are captured at the end of the user interaction. The originality and sequence of the inputs can be preserved and the valuable inputs would not be accidentally or purposely discarded as what occurs in the existing integration techniques.

6. LIMITATIONS

There exist a number of limitations of the developed Spatial Information Integration Technique in this research. First of all, the limitation occurs in the process of capturing the speech and sketch inputs. For speech inputs, the speech recognizer used in this research is not sufficient to convert the user verbal description to text. In the evaluation phase, the subjects' speech inputs were corrected manually even though it went through the speech recognizer. Also, in terms of sketch inputs, the sketch object is only limited to the polygon data type without considering the line and point data type in the shape-file layers in ArcGIS software.

In addition, the response and analysis time of the prototype is still not at the satisfactory level. This is mainly caused by the computer system that is used to operate the prototype. The computer system with larger hard disk space and higher memory power is preferable since the map database is large and it consumes a lot of processing power in loading the map layer for analysis.

7. FUTURE WORK

Given the limitations as identified in the section above, the following enhancements are suggested. First, it might be feasible to incorporate other types of modalities such as eye gaze and body gesture in the multimodal interaction in GIS applications. However, in order to employ these modalities, the technical and development issues related to the implementation of these modalities need to be taken into consideration and determine whether these modalities are suitable to be deployed in GIS application.

Second, in terms of the speech input interpretation, the length of the speech description provided by users needs to be specified at certain range. If the description is too short (with only 1 or 2 sentences), the interpretation process of obtaining the final result would be difficult because there is not enough information to be interpreted. However, if the description is too lengthy (more than 15 sentences), it might also cause confusion and a lot of duplication in the speech inputs. Nevertheless, the suitable length of the speech description is very much dependent on how complicated the corresponding sketch is. Therefore, a study is needed to develop a matrix relating the length of speech description against the different complication level of the corresponding sketch.

Third, the language accepted in this prototype is English where the speech recogniser employed in this research can support the English language. It is actually possible to explore the use of other languages in this type of application with the developed integration technique. Further studies and surveys need to be conducted to determine whether the interpretation result in other languages such as Bahasa Malaysia, Mandarin, Japanese, and French would be the same as in the English interpretation. In order to avoid bias in this type of study, all the conditions and process steps in this developed integration technique need to be strictly followed.

8. REFERENCES

[1] Blaser, A. Geo-Spatial Sketches (A study of people's sketching behavior and strategies). *Technical Report, National Center of Geographic Information and Analysis,* University of Maine, Orono, ME, 1998.

[2] Egenhofer, M. Why not SQL! *International Journal of Geographic Information Systems* 6(2), 86-95, 1992.

[3] Egenhofer, M. Multi-Modal Spatial Querying. *Seventh International Symposium on Spatial Data Handling (SDH '96)*, Delft, The Netherlands, Kraak, M.-J. and Molenaar, M. (eds.), 12B.1-15, 1996.

[4] Johnston, M., Cohen, P.R., McGee, D., Oviatt, S.L., Pittman, J.A., and Smith, I. Unification-based Multimodal Integration. In *Proceedings of 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, 281-288, 1997.

[5] Johnston, M. Unification-based Multimodal Parsing. *Proceedings of the International Joint Conference of the Association for Computational Linguistics and the International Committee on Computational Linguistics (COLING-ACL'98)*, 624-630. University of Montreal Press, 1998.

[6] Johnston, M., Bangalore, S., Vasireddy, G., Stent, A., Ehlen, P., Walker, M., Whittaker, S., and Maloor, P. MATCH: An Architecture for Multimodal Dialogue System. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 376-383, 2002.

[7] Schlaisich, I. and Egenhofer, M. Multimodal Spatial Querying: What People Sketch and Talk About. *1st International Conference on Universal Access in Human-Computer Interaction*, New Orleans, LA. Stephanidis. C. (eds.), 732-736, 2001.