SE 705 – Human Computer Interaction

SPEECH CONTROLLED PACS

1579044 Atakan Küçükatalay

703272 Hüseyin Candan

Introduction

Using speech for Human-Computer Interaction is a research area since the inventions of the computers. There are many applications to use speech for supplying information exchange between computer and human. Speech recognition applications include voice dialing (e.g., "Call home"), call routing (e.g., "I would like to make a collect call"), domotic appliance control and content-based spoken audio search (e.g., find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g., a radiology report), speech-to-text processing (e.g., word processors or emails), and in aircraft cockpits (usually termed Direct Voice Input).

Speaking computer applications seems to be easier to develop than speech recognition which makes it more common in the industry. On the other hand, converting spoken words to machine readable input, speech recognition, has an important role in human computer interaction. As an example, today it is possible to listen the news in Turkish from the web site of a newspaper (e.g. Sabah). On the other hand, there are also many studies and also successful applications to convert human voice to text. "Dikte" and "GVZ" are most popular speech recognition applications in Turkish. Especially, "Dikte" has a version called as "Tıbbi Dikte" which is widely known and used by radiologists to prepare radiology reports by just talking to a microphone.

Previous Work and Our Proto-type

In 1995, Bill Gates said that "When speech recognition becomes genuinely reliable, this will cause another big change in operating systems." Now we are in 2009 and speech recognition already starts to play a very important role in Human Computer Interaction. The reason why speech recognition will be most popular Human Computer Interaction method is that speaking is the most popular and natural communication method in human to human interaction.

Although expectations from speech recognition process are very high, we do not have very accurate tools on this area for now because of the challenges on recognition of speech. There are many things to consider which directly effects speech recognition process. These challenging agents include background noise, pitch of the speaker, distortion, task/context as shown in figure 1.

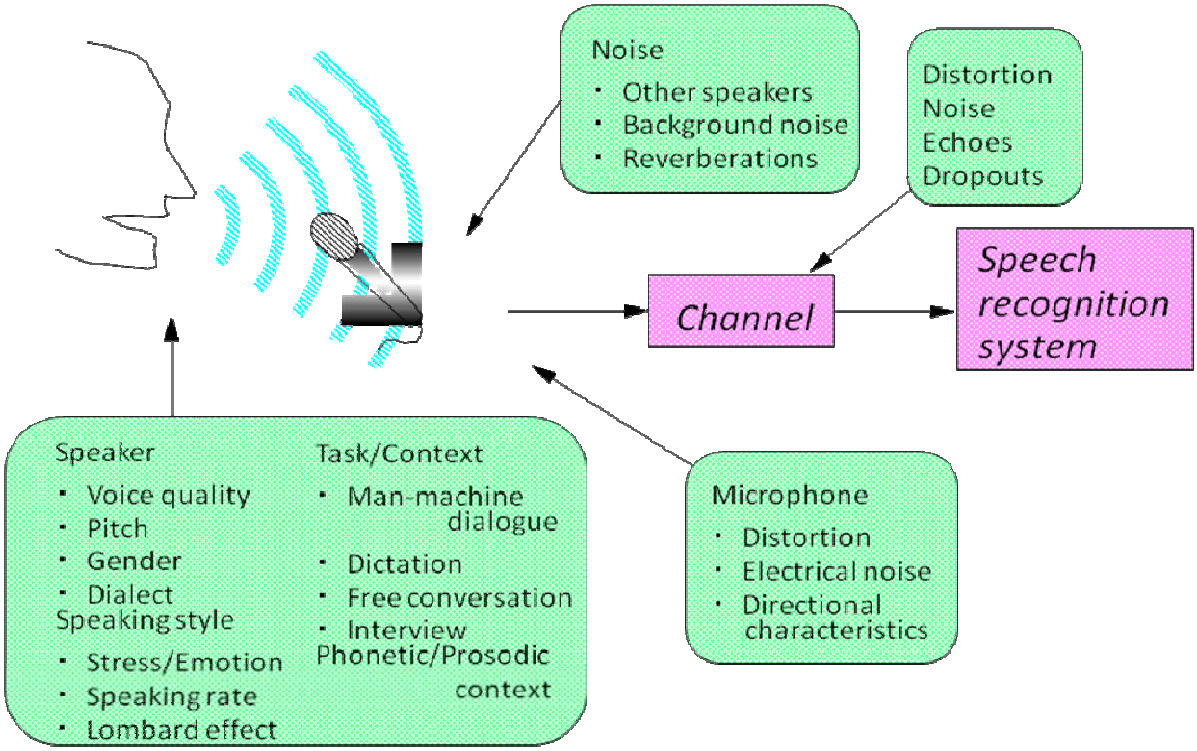

Figure 1. Main causes of acoustic variation in speech

In our application we use Dikte API which is developed using core of Tıbbi Dikte. Tıbbi Dikte is a well-known application which helps radiologists to prepare radiology reports by just talking to the microphone. Tıbbi Dikte is both domain and user dependent application, which means that it only recognizes radiological grammar which is a specific portion of the natural grammar. It also requires a teaching mechanism which forces the users to read the given sentences to analyze the voice and speech of the users in the initialization step. This constraints change for the API version of the Dikte. Dikte API, does not require a initialization step which makes it available to use by any user without a pre-study. Dikte API directly analyzes the words and does not have a grammar to check the analyzed word in the grammar to increase the accuracy.

We have built a high fidelity prototype to simulate the usage of speech controlled web based PACS interface. PACS's are applications to retrieve medical images from various medical imaging instruments (e.g. ultrasound, magnetic resonance, mammograms, etc.), store and present these images in a required format. We have already developed a PACS which has a web-based user interface and in this project we will make it available to control some functions of this application by speech commands.

Potential Users of our tool is same as the users of the PACS which includes radiologists and technicians. Although the potential users usually have very little experience on computer usage, many of them are using PACS's for their work. Especially the radiologists either using "Tıbbi Dikte" or have knowledge about it, that makes our job easier to inform them about the conceptual model of our implementation.

Our system has a multimodal interface which can be controlled by keyboard, mouse and speech. Speech recognition makes the application flexible by leaving the hands and eyes of the user free which is very important for radiologists at the time of preparing the report. General speech recognition applications have the advantage of being natural which means there is no need for special training. In our application the instruction set includes some domain specific commands which may not be known by a ordinary people. But, due to the limitation of target users of the

application, it is not a problem for us. Because the application will be used by radiologists and technicians which are very familiar with the jargon used in PACS systems.

We used free version of Dikte's API in our prototype to recognize the speech. Free version of the API has the limitation of 20 words to recognize which is enough for us to evaluate the usage of speech recognition in a web application.

We prepared a command list which would be used to control our prototype and in our evaluation test we asked users to prepare their own sets to get an effective and efficient command set. We use a 20 words contained command set because of the limitation of free version of the API. The set which we prepared before the system evaluation includes following commands: Tetkik menüsü, Yönetim menüsü, Tetkik sorgulama, Servis tanımları, Sorgula, Önizleme, Sonraki Sayfa, Önceki Syafa, İlk Sayfa, Son Sayfa, Seviye Hasta, Seviye Tetkik, Seviye Seri, etc.

Although it is possible to use the system by a standard microphone, to increase the accuracy of the speech recognition we used a special microphone set. This set has two embedded microphones on it. One of them is for acquiring speech which should be positioned close to the user's mouth, and the other one is at the bottom of the set which captures the background sound. By making calculations on the data retrieved from this two devices to eliminate the background noise increases the accuracy of the speech recognition. This process is automatically done by Dikte's API.

To develop our prototype first we built an ActiveX which uses Dikte API as speech recognition tool. ActiveX is embedded in web page by <object> tag of HTML. It catches recognition results and calls corresponding client side script (Javascript) function on the web page. ActiveX also displays a voice chart for hardware feedback. Client side script first displays the recognized string on the corresponding text box. There are two classes of commands in our command set according to process base. Some commands can be processed on client script and some others require server side processing. For second part commands, client side function writes the corresponding parameters on a control on the page and submits (post) the page to the server application. The architecture of our system is shown in Figure 2. The pink boxes in the figure represents the modules implemented by us to prepare our high fidelity prototype which is used in the evaluation.

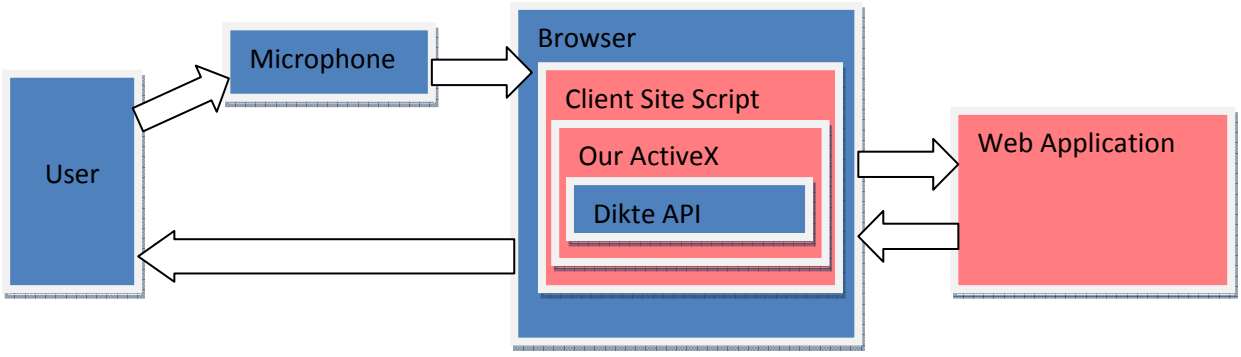

Figure 2. The architecture of our application.

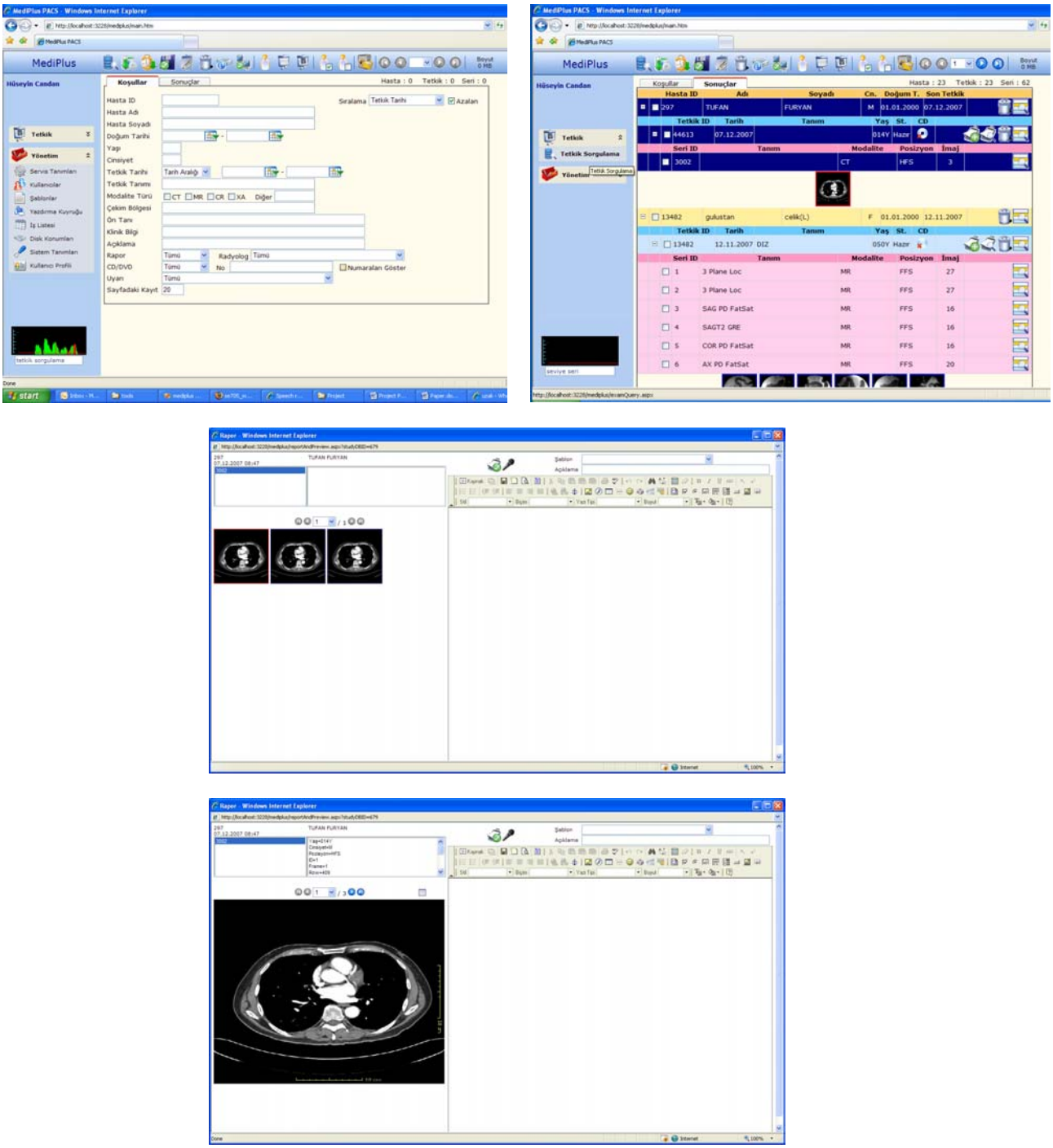

Figure 3. Screenshot of our prototype

The prototype also gives feedback to the user about the current process which allows user to be informed about the conceptual model of the system. The feedback includes two types of information. One of them is about the level of the voice received over the microphone by a signal chart. This signal chart shows that whether the hardware part of the system is working without any problem or not. If there is no signal in the chart, user knows that it is required to check the cables of the microphone or the sound card, etc. Another feedback presented is the result of the recognition process which is shown in a text box. This information allows user to know how the system analyzed the command, and which action will be performed next.(see figure 4) Actually, speech recognition cannot be made without any error due to many reasons such as difference frequencies in human voice, different accents used in natural language, different speed of speech,

background noise, etc. Therefore, in some cases, the application may perform wrong actions due to misunderstanding of the command given by the user. This defect usually causes the user to be frustrated which makes it crucial to give feedback about the result of the reorganization process.

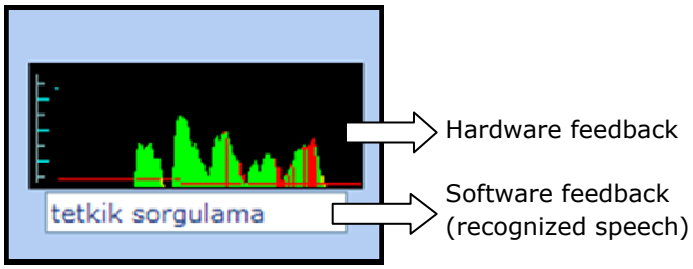

Figure 4. Feedback

Dikte API does not supply any functionality to determine whether the given command corresponds one of the item in our command set. The API recognizes the speech and returns a highest matched command from the given list. It does not return a value about the accuracy of the recognition process to allow us to eliminate the user's commands which are not listed in the command set. Therefore, we could not supply error handling for inputs that is not in our command list. There is no way to determine if the speech is the command given to system or talking with any other person. To be able to eliminate such error, there is a switch on the microphone set by which user can turn off the microphone. Another error handling made by our system is ignoring the commands given in an inappropriate state. If the user gives a command which does not correspond the current state of the application, the system displays the recognized command on the related text box, but does not do any action for that command.

We make our evaluation by letting 5 users to control our system by their voices. Our program is designed especially for radiologists therefore 3 of our user is chosen from radiologists. To make a realistic evaluation we decided to choose at least one of the participants from a profession domain in which computers are not usually used. By the help of these users we can test our system visibility and visual affordance much better. Therefore we choose one primary school teacher and one high school math teacher. Both of them stated that they are not accustomed to computers.

Before the participants started to use our system, we give a small presentation about our system especially for the users that are not accustomed to neither computers nor Dikte's focused speech recognition systems to make them feel more comfortable while using the system.

We encountered three important criteria while preparing the speech command set of the application: recognition accuracy, memory load and flexibility. We tried to choose commands which do not have similar phonology and easy to remember. After that small presentation we give the users our command list and want them to use the commands to guide our system. After that we asked the participants if they have any advise on our command list to improve the correspondence between the actions and the command to do these actions. Therefore we requested a command set (20 commands) from each participant and we noted these commands in a paper. We prepared a new command set by choosing the items with highest rate of preference, and inform the users about new command list. We gave them 5 minutes to memorize the new list and ask them to use the system with the commands of either the new or old list. We evaluate the new set, by using Wizard of Oz technique in a way that one of us used the system by conventional methods (keyboard, mouse) according to the commands given by the participants. Besides, we noted the commands preferred by the participants for each action to prepare a command set which can be easily remembered by the users. At the end of this process we finalized our command list by adding multiple commands for some actions to produce more flexible interface.

Figure 5. Wizard of Oz Technique

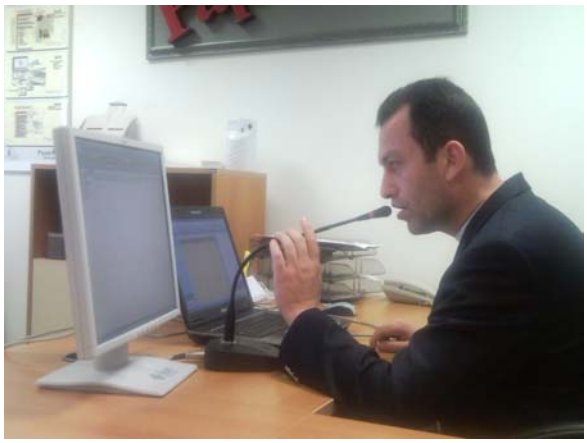

Figure 6. Usage of our prototype

After we finalize our command set we ask our participants to use our system one more This time we wanted to see if performances of users increases and we noted errors of users. After that we asked the users to use the sistem one more time. This time the part we were interested in is the system accurcy We supply a backround noise and noted the errors that the application made in the noisy enviroment and in the silent enviroment to see the effect of backround noise to our system.

Evulation

For system evaluation and showing the statistical results of our research, we noted the speech recognition errors in a paper and captured video image while the participants were using the system. We repeat the process multiple times to see the effect of repeat on remembering the command set by the users and effect of background noise in speech recognition accuracy. Besides to take user opinions, we prepare a small questionnaire which includes closed questions.

Part #	Profession	Number of commands given by participant	Number of commands not remembered correctly (First test)	Number of commands not remembered correctly (Second test)	Number of speech recognition errors (without background noise)	Number of speech recognition errors (with background noise)
1	Radiologist	100	5	1	10	12
2	Radiologist	100	3	2	14	15
3	Radiologist	100	7	4	16	19
4	Primary School Teacher	100	22	10	12	13
5	High School Math Teacher	100	12	9	18	19

Table 1. Evaluation results of the system by the users' and speech recognition performance perspective

Question	Participant 1 Primary school teacher	Participant 2 High school math teacher	Participant 3 Radiologist	Participant 4 Radiologist	Participant 5 Radiologist	Avarage
Bilgisayarı kullanma sıklığınız	2	3	5	5	4	3.8
Daha önce sesle komut verdiğiniz bir yazılım kullandınız mı?	1	1	4	5	5	3.2
Daha önce Tıbbi Dikte	1	1	4	4	5	3.0

kullandınız mı?						
Sistemin kullanımı sizce kolay mı?	3	3	3	3	4	3.2
Sistemin, komutları algılama hızı yeterli mi?	3	4	4	4	4	3.8
Sistemin, komutları tanıma performansı yeterli mi?	2	4	4	3	4	3.4
Böyle bir sistemin kullanımı, iş performansınızı artırır mı?	4	4	5	4	5	4.4

Table 2. Results of questionnaire analysis

Because the system is designed for a specific domain, as we expected, radiologists can use the system more effectively.

For the evaluation we select first 100 commands of the users to analyze the system accuracy. The radiologists which are the main users of our application domain could not remember about 5% of the commands correctly, while the ratio is 17% for other participants. As it is expected, second test gives better results on remembering the commands correctly. This results show that as the experience of the users increases on the system, the ratio of wrong command entrance decreases to an acceptable value.

One of the biggest constraints in speech recognition is background noise. As stated before, our microphone set is design to help on reducing the effect of background noise on speech recognition accuracy. We recorded the speech recognition error rates in both noisy and noiseless environments. From Table 1, we can see that the microphone set almost blocked the effect of background noise. While the microphone set helps on increasing accuracy of the system, on the other hand, it is affecting the portability of the system negatively.

Conclusion

Speech based user interfaces are going to be very familiar in the near future. Users have positive opinion on using the software applications by talking to the system as interacting with human. Accuracy of the speech recognition is the most challenging point of the speech based user interfaces. It is much easier to recognize speech for domain and user dependent applications which have limited vocabulary and grammar.

In this project, we prepared a high-fidelity prototype for speech controlled PACS application. It is a web-based application which is used through a web browser. Similar to the hardness of developing web application compared to desktop applications, making a web application speech-controlled requires more work. To prepare our prototype we implemented both client and server side code. Producing such a prototype gives us many clues about adding such an interface to a system which is already being used in the industry.

In speech based systems command sets have strong effects on natural interaction and usage. We think that the designers and researchers should give more attention to command set preparation and should find new approaches to take the interaction level of the speech based systems to desired level. Evaluation phrase of the prototype shows that speech recognition can make the interaction between the user and the system easier and more powerful if the accuracy of speech recognition engines improves to a acceptable level.