CONVERSATIONAL INTERFACES: TASKS ANALYSIS TO SIRI AND GOOGLE NOW

Fernanda dos S. Becker Jacintho, Esp.
Alice Demaria Silva Penha, Msc.

Adress
Avenida Lédio João Martins, 971, Ap. 307, Kobrasol, São José/SC, 88102-001

Keywords: Conversational Interfaces. Voice Assistants . Tasks analysis. User Experience.

1. Introduction

More and more digital products in which users interact through conversational interfaces have been launched. These applications have been around for some time, but advances in technologies such as artificial intelligence, language and mobile technology, enabled the development of these interfaces and the implementation of more complex tasks through them.

The interaction through text messages as they occur in applications as Slack, WeChat, Facebook Messenger or voice commands as in the case of virtual assistants like Amazon Echo, Siri, "Ok, Google" allow the user to find information, perform simple tasks such as make a call, or more complex tasks, such as how to shop online. These interfaces are transforming the way people deal with technology and can change the way we use computers. This can impact on a behavior change, since people can book a table at a restaurant, schedule an event on a calendar, shopping, discover the next bus schedule or be informed of the weather forecast, through a simple conversation based on their commitments in schedule and routine.

In this scenario, it is necessary to design and build interfaces that talk to users naturally and still deal with the unpredictability of this interaction, as the user have much more freedom to write or say whatever he/she wants, than in the context of the graphical interface, in which there are predetermined interaction flows. As punctuates Teixeira (2016),

> the challenge focuses more on identifying the right commands and the choice of words that activate each feature, than in designing the actual interface. Choosing the right tone, the words interpreted and expelled back, the timing, and sound effects that help usability.

It is also necessary to deeply understand the human behavior, understand the impact of these interfaces in the user's experience with the brand or product, even more, to understand the context of use, the needs and desires of the users, to make design decisions that provide effective and positive experiences.

2. Conversational interfaces: features and applications

Conversational interfaces are interfaces that allow communication to take place by means of natural language, it is spoken or written. Para McTear *et al*. (2016, p.25), "the main porpuse of a conversational interface is to support conversational interaction between humans and machines". In this context, instead of communicating with computers occur through clicks or specific commands, communication becomes based on a conversation. This is because conversational interfaces can understand natural human language and respond through this language. Como pontua Hura (2008, p 197), "underlying the voice user interface is speech recognition technology that has the ability to capture and decode the user's spoken input to allow the system to "understand" what the user has said".

2.1 Features

These interfaces can be robots that use information available on the Internet as memory bank. In many cases, they are an algorithm that interprets messages sent by users and returns the appropriate response for each occasion (Teixeira, 2016). In addition to accessing the information available on the Internet, they identify the location of users, what they are doing and what is happening around them, due to the integration with different services.

The great advantage of these interfaces is that they are cross-platform, that is, work well in several different platforms such as smart phones, desktop, smartwatches and also devices like Amazon Echo, a device that has no interface. In addition, they integrate with apps like Snapchat, Facebook and Twitter. Moreover, conversational interfaces eliminate the need for an interlocutor between the computer and the user, until now role played by the graphical user interfaces, because "[...] enable people to interact with smart devices using spoken language in a natural way -- just like engaging in a conversation with a person" (McTEAR *et al*, 2016, p. 1).

For companies, bet on conversational interfaces means being where your customers are. It is remarkable that people spend much of their time on mobile devices interacting with others. Messaging applications are downloaded success in virtual stores. Only by WhatsApp are made 100 million calls every day, which has one billion users in its base (WhatsApp, 2016).

But why these interfaces are attractive for users? According to Harris (2004, p. 18), "we are all masters of conversational in very much the sense we are all masters of conversational in very much the sense that we are all master of locomotion, it`s something we do, after early childhood acquisition, instinctively". Talk is natural and intuitive for most people and start a conversation is something done without having to think about how to do this, because "language is acquired automatically by human infants throught simple expoxure and requires no adult intervation, other than simply conversing with the child." (HURA, 2008, p.199). So, naturally we use dialogue, being spoken or written to communicate, take action, share knowledge and feelings.

This form of interaction allows performing tasks that involve risk or high concentration, when the hands are not free to handle the device, such as when driving or surgeries, and reduce the difficulty some people have with graphical user interface and help people who own motor and visual limitations. McTear et al. (2016, p.1) points out that "voice input is often the most appropriate mode of interaction, especially on small devices where the physical limitations of the real estate of the device make typing and tapping more difficult".

Voice assistants are controlled by voice commands, they do not have graphical interface and this is perceived by the users through the ears, not the eyes. Harris (2004, p. 3) points out that "voice intefaces are interactive media in which the input is primarily or exclusively speech and so is the feedback". However, they are enclosed in applications with graphical user interface, and outputs may also happen through them.

Chatbots are interfaces in which input and output data is performed in textual form. The graphical interfaces of chatbots are in many cases determined by the platforms in which they are, so it is not possible to interfere in the message layout. As in the virtual assistants, the great challenge of chatbots is to convey the experience of using the service in a few lines of text, because "all of your features need to be reachable solely through words—so picking the right thing to say, and the tone of your dialogue with the user, is crucial". (MARIANSKY, 2016, no page number).

2.2 Virtual Assistants by voice command

Currently, almost all operating systems developed for mobile and other devices have their virtual assistants who respond via voice commands, such as: Siri, Apple's iOS, Google Now by Google, Amazon's Echo, Cortana by Microsoft. They allow users to perform tasks in order to impact positively on their productivity. There are examples of tasks related to the organization as to make an appointment on the agenda; other related to communication such as call someone or send a text message; activities related to the user entertainment such as playing music; and also related to the search for information, among other possibilities.

Developed by Google, Google Now application is present in Android and iOS software. The company uses its large database to make information available to users. Through the application, users can ask questions to the voice assistant or receive information based on their personal data or browsing

history, that is, Google query all the information linked to the user and provides without user request, through cards with information updated throughout the day.

Siri is the virtual assistant developed by Apple, integrated with Apple services like iMessage, Calendars, Safari browser, among other external services used to consult user information and thus be able to answer questions and perform tasks.

Both have English as original language, but were made available in Portuguese recently. Google Now was available in Portuguese in version 3.5.15 of the Google Search app, in 2014 and Siri in the IOS 8.3 version, launched in 2014 (Bijora, 2015).

3. Method

This study, which is characterized by a qualitative research, descriptive and applied type, was carried out by the analysis of applications that has the conversation as the primary form of interaction with the user.

The technical procedure used was tasks analysis through the analysis of two assistants who respond by voice commands. In addition to analysis of the applications, a brief literature review was performed, which served as a guide for data analysis. In this article, the voice assistants (voice assistants) to be analyzed are Google Now and Siri.

For the voice assistant analysis, conducting eight tasks have been proposed, among them seven who can be considered as simple as scheduling an event on the calendar, and one considered a complex task. The tasks were carried out in Portuguese (Brazil), both virtual assistants were set for this language. Below the list of tasks that were performed.

a) Task: To set alarm
Sentence: *"Criar alarme para o próximo dia às 8 horas da manhã"*
b) Task: Call reminder
Sentence: *"Lembrar de ligar para Ana amanhã"*
c) Task: Create event on the agenda
Sentence: *"Criar evento: entregar trabalho na universidade dia 20 de agosto às 17h da tarde"*
d) Task: Create event on the agenda
Sentence: *"Criar evento: festa de halloween dia 30 de outubro, às 10 horas da noite"*
e) Task: Send a text message
Sentence: *"Enviar mensagem para Marcelo: Oi?"*
f) Task: Discover restaurant location
Sentence: *"Onde fica o restaurante Box 32?"*
g) Task: Identify the music being played on the environment
Sentence: *"Que música é esta?"*
h) Task: Book a table in a restaurant
Sentence: *"Fazer uma reserva em um restaurante italiano"*

To carry out the tasks, it was chose for a closed, quiet location with no outside noise. This experiment was conducted in a room with only the presence of the researcher. Before performing each task, the activation of the voice assistant was made by touch the device. As soon as the assistant were activated, one of the sentences related to the tasks described was pronounced approximately 20 centimeters from the device. After the completion of each task the voice assistant was off.

4. Results

Considering that in this type of interface in which the interaction occurs through the conversation, something essential for conversational interfaces obtain success is "having a good conversation" to engage users and guide them in the tasks. As in graphical user interface, if the user find an obstacle in the process (eg, if not interpreted correctly by the computer), it can be frustrating the user can give up the experience. Further, through the conversation, is possible to design rich

experiences in which users can interact, since it is not only about to answer them through these interfaces.

4.1 About speech recognition and language

In the tests conducted, the voice recognition of Google Now for words in Portuguese was efficient and agile, delivering instant results. Also, for the tasks carried out in stages, in which the assistant replied with questions, it was possible to clearly understand the assistant. Also in relation to speech recognition, Siri showed similar results as Google Now, recognized quickly and very well the voice commands, in addition, it is possible to understand it answers clearly.

It is noticed some difficulty on Google's voice assistant to understand when there is, in the same sentence, words in Portuguese and English. In the task to create event on the calendar, the event title had a word in English, the assistant tried to identify words for the event title in Portuguese. Due to it is integrated into the Google search platform, it returned a search result that it considered closer to the input, as Figure 1 below. Note that this may cause some "misunderstandings" between user and assistant, for example, when the user, using the assistant in Portuguese want to send an e-mail whose recipient address is composed of words in English. The same test was conducted with Siri, which properly identified the sentence, but on creating the event on the agenda, the title was inserted incompletely, because the assistant could not write what was said, despite initial understanding of the word in English. In this sense, it can be concluded that the assistant did not succeed with the data entry in two different languages, as the task was not performed completely (Figure 2).

Figura 1: Resultado de busca na internet
Fonte: Acervo pessoal, 2016

Figura 2: Criação de evento na agenda com aplicativo Siri
Fonte: Acervo pessoal, 2016

Another point observed in the tests is that in both cases, the voice command is available for a short period after activation by the user. So if the user does not give an initial voice command, or does not respond to the task requested, in step in which it is the voice command turn off, requiring new click to activate Google Now assistant. In Siri, in newer versions of mobile devices from Apple, a functionality is available that allows activation by the phrase "Hey Siri" in addition to the drive per-click option. The same has been tested and responded as expected, but given that this feature allows

greater freedom in more advances appliances models, tasks have not been conducted on this entry form.

Still on voice recognition, the task sending text message, to give more complexity to it, it was purposely chosen a contact in the phonebook where the first name was repeated in another contact, just by changing the last name. Voice assistants responded differently to this task: Siri, by questioning to which contacts the message should be sent, presented problems in identifying the full name of the contact, so it was necessary to click on the right option to be able to finish the task. This happened due to the assistant understand different the last name of the other contact, which was confirmed in the second test trial, in which the response to the assistant was the sentence: "second contact", and the assistant to confirm the task reproduced the name of the contact with pronunciation different from the user. At this point, it is worth mentioning the importance of the user accurately in relation to pronunciation considered the standard language for the assistant, independently the language. In this case, the contact's last name was in German. In case of uncertainty on this point, the user has the option of testing other forms of interaction that do not require such knowledge to continue the task for only voice commands, but in this context, the ideal is that the assistant could relate the pronunciation with an existing contact. In contrast, Google Now for the same task, showed better performance with the same pronunciation to the last name in German, the voice assistant could recognize that it was an existing contact on the agenda, there is no need to repeat the answer or even answer it in a diferent way (Figures 3 e 4).

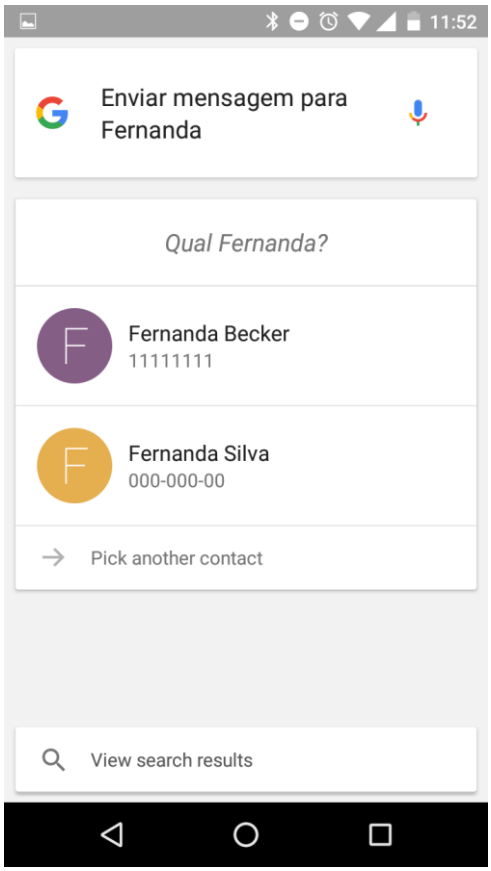

Figura 3: Comando de voz: enviar mensagem
Fonte: Acervo pessoal, 2016

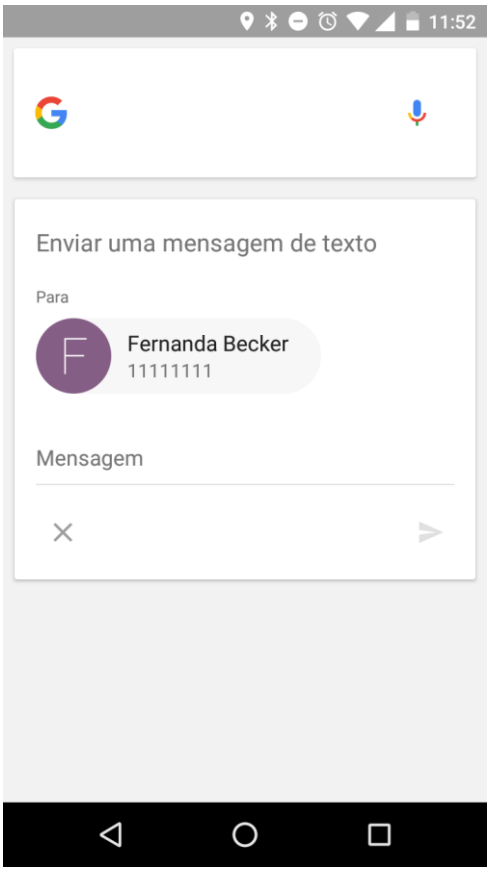

Figura 4: Definindo contato por comando de voz
Fonte: Acervo pessoal, 2016

4.2 *Feedback*: voice or graphical user interface support

To perform tasks, the user has the option of giving initial voice commands and then continue through the virtual assistant interface or perform the complete task with voice commands, for this is very important that the assistant suggests the next steps so that the user understand easily and dispense using the keyboard, given its intention to carry out the task by voice commands. In this sense, the

feedback needs to be efficient and effective, Norman (2006, p. 50) defines feedback as "give to the user the feedback on the action that was in fact executed". Also on this topic, Krug (2006, p. 15) points out that "as a general rule, people do not like to have to figure out how to make something", as well as the visual interface, the user feedback is essential in conversational interfaces.

In most of the tasks performed in this study, the feedback of the applications was through speech output, but the applications have reacted differently in some tasks because, despite the voice commands, both virtual assistants have graphical interfaces that follow the the conversation flow. For example, the interfaces show the user's voice input and the task being performed by the application according to its stages. For the task of search for information, which was asked the location of a restaurant, assistants present their answers differently. Google Now, according with the voice command imput, said the full address of the restaurant as if you were actually in a conversation, that is, with the expected output. Differently, Siri only showed the results for the search through the graphical application interface, the only voice output was: "See results for your search" requiring the user to visualize the device's screen to get the results . It is possible that because of Google Now be integrated into the mapping and location service, the accuracy for this answer is better, allowing a direct response by voice command. This is just an assumption and further studies can be conducted to better research.

4.3 Gender, voice intonation and mockery

Regarding the voice tone, some voice assistants have female and male genders. Google Now does not have gender identification, although his voice clearly present be more feminine. The language is more informal and familiar, suggesting a closer relationship among the application and user. Siri voice assistant is female and it has a more intimate proposal, even sarcastic tone in some cases. For example, it was purposely created three events in sequence, when creating the third one, it has not been finalized. Siri's response to the cancellation of the task "OK, you have a lot to do anyway, Marcelo" (Figure 5). The reaction caused is unexpected and even funny, as the application completes the task with a different response expected.

Figura 5: Resposta do aplicativo Siri para o cancelamento de evento na agenda

Fonte: Acervo pessoal, 2016

4.4 Complex tasks not performed

For the task to identify the background music, in the tests it was noticed that Google Now does not have voice command for this task in Portuguese, only Siri responds to this type of input. In the test performed, Siri was able to correctly identify the song that was playing on the environment (Figures 6 and 7).

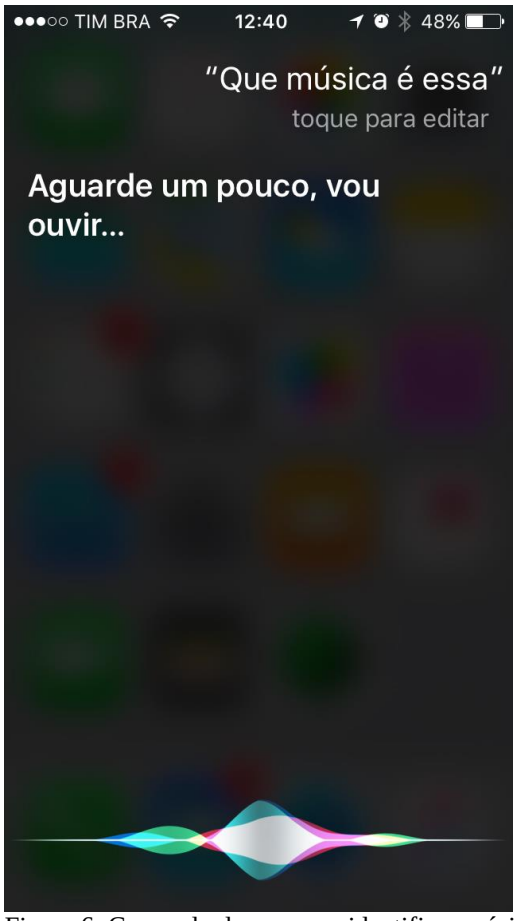

Figura 6: Comando de voz para identificar música
Fonte: Acervo Pessoal, 2016

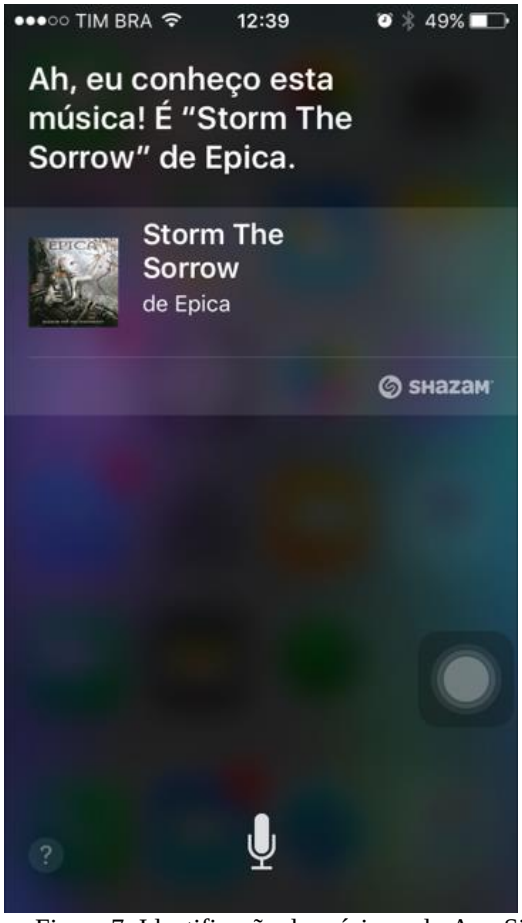

Figura 7: Identificação da música pelo App Siri
Fonte: Acervo pessoal, 2016

The last task of the test, that can be considered the most complex of all, was the attempt to make the booking of a restaurant by voice command. Unfortunately, both assistants failed to perform this task in Portuguese. On the company website that develops Siri, it is possible to access the tasks that the application can perform and there are differences between the application available today for the United States and the possibilities for Brazil. In this case, both Siri as Google Now understand the voice command "Where is Box 32 restaurant?" and showed search results for the command.

5 FINAL CONSIDERATIONS

Although conversational interfaces are not as recent, studies and discussions in relation to models such as chatbots and virtual assistants with voice commands are still only beginning, as technologies for their application had considerable advances only in recent years. For the coming years, it is possible to expect that these models be part of the daily lives of people more naturally, as

applications, virtual stores, platforms, games, websites and different devices will use these models to present part of their experience.

It is possible to think in the benefits that the use of these interfaces can have on the lives of people not only in productivity gains, but also in improved quality of life, as in the cases of people with disabilities, these interfaces can make their routine and perform tasks easier. It is noteworthy that they need to be efficient and effective, so that the experience can relevant to the user. With regard to this issue, Norman (2006, p. 55) points out "the same technology that simplifies life by offering more functions on each device or instrument also complicates life by making the use of the device more difficult to understant, more difficult to use. "

Regarding the study conducted with both voice assistant, both assistants carried out the tasks proposed, even where there was interference from the graphical user interface in which the application was inserted, except in the task in which the functionality was not available to the Portuguese language.

Both assistants showed good performance in relation to speech recognition to the tests in Portuguese, however understanding problems happened in sentences with words in two different languages.. In addition, the voice assistant Siri presented difficulties when the pronunciation of words in another language was different from that registered in the application. About feedback, there were times when the application response was through the graphical interface, which limits the use of voice command for full accomplishment of the task. To perform more complex tasks, unfortunately it was realized that the Portuguese version of the voice assistants have limitations compared to the versions currently available in the United States. Further studies can be conducted to verify the performance of voice assistants and user experience with these interfaces.

References

BIJORA, Helito. **iOS 8.3 beta: saiba como ativar a Siri em português em seu iPhone ou iPad.** Techtudo blog. 17 mar. 2015. Disponível em: <http://www.techtudo.com.br/dicas-e-tutoriais/noticia/2015/03/ios-83-beta-saiba-como-ativar-siri-em-portugues-em-seu-iphone-ou-ipad.html>. Acesso em: 10 jun 2016.

HARRIS, Randy Allen. **Voice interaction design**. Burlington, MA, USA: Morgan Kaufmann. 2004.

HURA, Susan L. Voice User Interfaces. In: KORTUM, Philip (Ed.). **HCI Beyond the GUI**: Design for Haptic, Speech, Olfactory, and Other Non Traditional Interfaces. Burlington, MA, USA: Morgan Kaufmann. 2008. p. 198-228.

KRUG, Steve. **Não me faça pensar!** Uma abordagem de bom senso à usabilidade na web. Rio de Janeiro: Alta Books, 2006.

MARIANSKY, Matty. All Talk and No Buttons: The Conversational UI. **A List Apart Magazine** [on-line], New York, 23 fev. 2016. Disponível em: <http://alistapart.com/article/all-talk-and-no-buttons-the-conversational-ui>. Acesso em: 10 abr. 2016.

MCTEAR, Michael; CALLEJAS, Zoraida; GRIOL, David. **The Conversational Interface**: Talking to Smart Devices. London, UK: Springer, 2016.

NORMAN, Donald A. **O design do dia-a-dia**. São Paulo: Rocco. 2006.

TEIXEIRA, Fabrício. **O papel de UX nas interfaces conversacionais**. Arquitetura de Informação blog. [S.l.], 1 fev. 2016. Disponível em: <http://arquiteturadeinformacao.com/user-experience/o-papel-de-ux-nas-interfaces-conversacionais/>. Acesso em: 10 jun. 2016.

WHATSAPP. **WhatsApp Calling**: 100 million conversations every day. WhatsApp blog. [S.l.], 23 jun. 2016. Disponível em: <https://blog.whatsapp.com/10000625/WhatsApp-Calling-100-million-conversations-every-day?l=en&set=yes>. Acesso em: 15 jul. 2016.

_____. **One billion**. WhatsApp blog. [S.l.], 1 fev. 2016. Disponivel em: <https://blog.whatsapp.com/616/One-billion>. Acesso em: 15 jul. 2016.