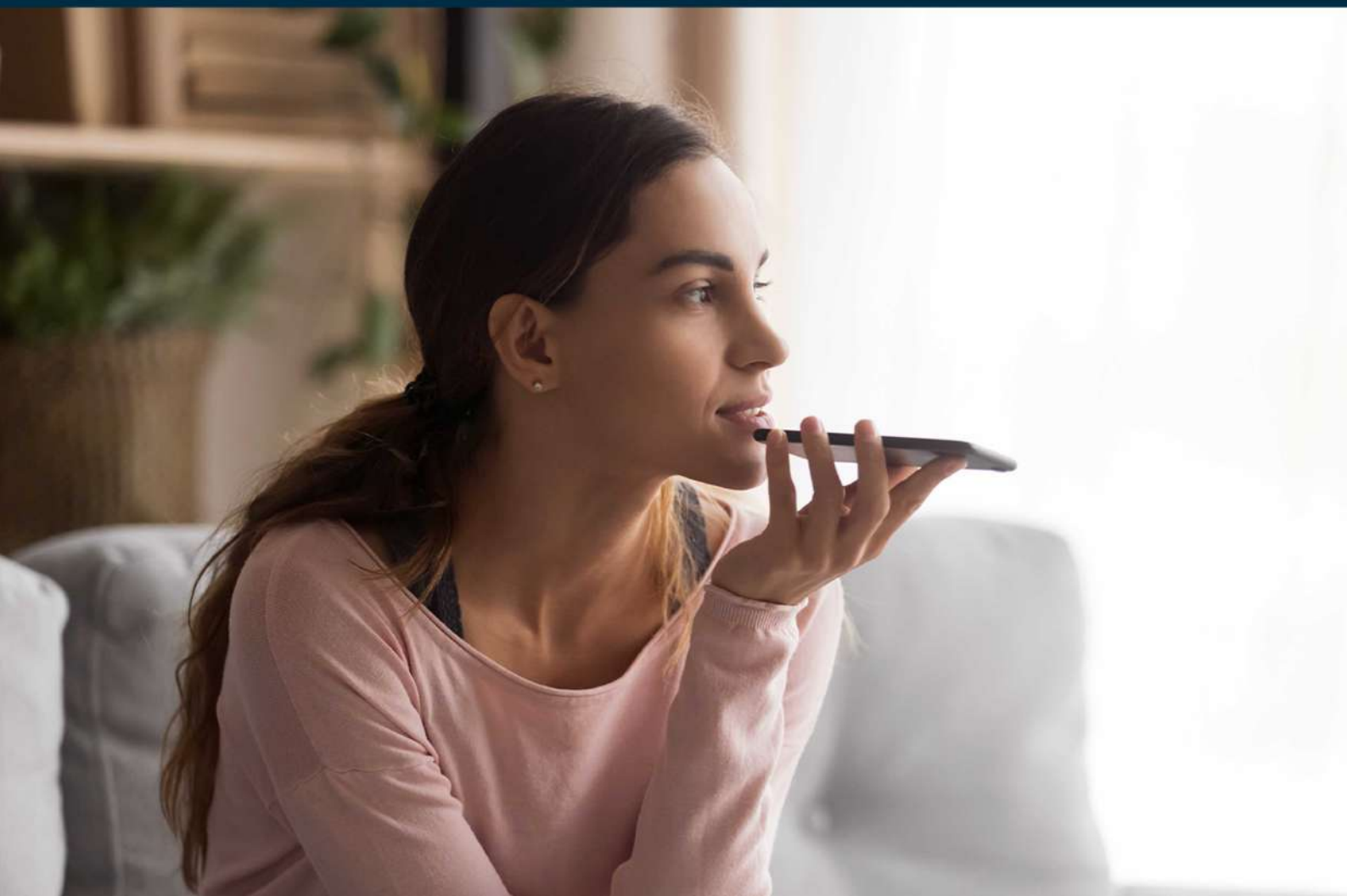
Telefónica

Whitepaper Conversational Design

Marta Pérez García
Sarita Saffon López
Ana Molina Alarcón



Introduction

"No longer the sovereign property of humans, Speech has become an ability we share with machines" - Sarah Borruso

As in every new chapter since technology was put at the service of humans, digital product design is undergoing a new paradigm shift that will undoubtedly impact the way we design services: conversation as a model of interaction.

Several decades ago, we were only focusing on physical products, their materials, their physical ergonomics, how people use them, etc. However, over the last couple of decades, with the change of paradigm from physical to digital products, this scenario has also changed, and we have begun to consider the way in which people interact with technology, what we look for, how we interact with interfaces, and how we adapt to this digital life in the personal and professional spheres. Apart from the change described above, we have now, for several years been faced with an even newer context; visual interfaces have given place to new models of interaction, as in the case of voice interfaces. Models that are becoming popular with the emergence of virtual assistants present in most intelligent devices. These assistants, that over the last three years have sought to simplify the relationship between humans and technology, incorporate new mechanism of interaction allowing us through natural and simple conversations: to better manage our day-to-day life, to obtain the information we are looking for without touching the screen, or to make purchases in a simple way. A promising start with much to come.

These new interfaces, which incorporate conversation, represent new challenges in the world of design and user experience, in the world of business and in the model of relations between companies and customers. This new channel or contact point with customers requires a great deal of experience development, not only to give correct answers, but also to generate quality relationships based on trust between people and companies.

But before we get into the details, let's start with the basics: **What is a conversation?**

The dictionary of the *Real Academia* defines a conversation as the action of speaking familiarly one or several people to other or others. The Cambridge dictionary, in turn, states that for a conversation to take place, ideas and feelings must be expressed, questions answered, and information exchanged between the various people involved. Both definitions emphasize the fact that more than one person participates in a conversation. However, when we refer to digital systems such as virtual assistants (VAs), a conversation can occur involving a single person interacting with a machine. In fact, when we consider the emerging discipline of Conversational Design, the term itself has not been widely explored and documented. Currently the term is defined as the field that studies how optimal conversations can be created between an artificial agent and a person. In this article we will explore further the current state of this field and how we are addressing it in Aura, the virtual assistant of Movistar.

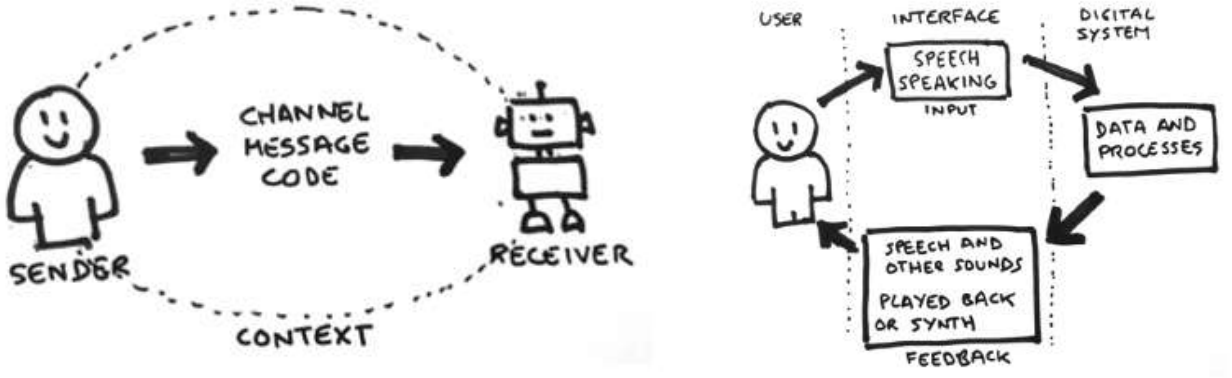


Illustration: How a conversation between a person and a machine is articulated (Author: Cristina Santa Cecilia)

How are organisations managing these human-machine conversations?

Although conversations between humans and machines are already a common occurrence within companies and are implemented in the customer experience strategy to generate a new relationship and a new contact point with their customers, several still face the problem of really understanding how to manage a dialogue, since even having in mind some standard phrases that could be used to request a task, the way different people adapt their speech, the turns a single conversation can take, are not, and cannot be, completely defined.

One possible reason why companies find themselves at this crossroads when faced with the subject of conversations between people and machines is that, although language has penetrated the field of design, known for helping to deal with complex issues, the focus has been on prioritising the exchange of data, the development of the dialogue as such and the flow of the conversation, rather than designing the final result. While the former is important, we are not making the most of what this discipline has to offer, which is to concentrate on what goes beyond the literal interaction, and to see the user experience as a whole that must be designed in its totality.

This means that technology companies are clearer about the functions they want to have in their virtual assistants than about the very articulation of the conversations that users are going to have with them, so there is a dilemma between what the company can do on a technological level and the definition of the way in which the experience will be developed. Moreover, often the assistant works more on resolving things that are already efficiently resolved in other ways, rather than offering added value through conversational experiences.

Can we design a conversation with the tools we have today?

Technology today is changing these paradigms, allowing us to design and develop natural language conversations between people and digital systems. There are new technologies such as ASR (Automatic Speech Recognition) or NLP (Natural Language Processing), which make these interactions between the two subjects (man and machine) seem like a real conversation.

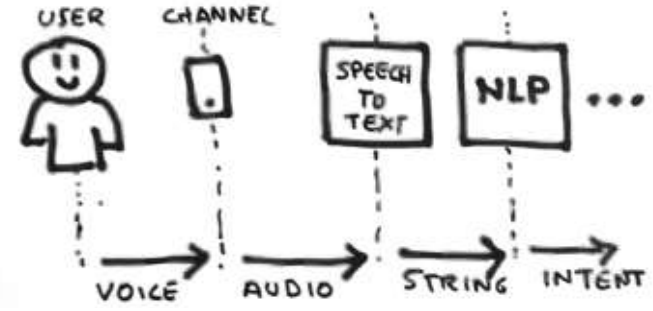


Illustration: Natural language processing and automatic dialogue recognition (Author: Cristina Santa Cecilia)

However, in this type of process, as we can see in the illustration, there is a huge variety of steps, which means that there are many opportunities for something to go wrong during the process. Therefore, even with these technologies, we must be prepared for the unexpected, as our conversations do not respond to a linear flow. We interact and cooperate with each other to make communication successful because we share a common context and because we are trained to do so. Emulating that human capacity in a natural way is a huge challenge, while technology solves that huge challenge, designing and managing mistakes could be the solution. In fact, addressing mistakes in conversations is something we must consider in conversational design, as turns and requests can be unexpected, and responses, as a result, can be unexpected as well.

What aspects are critical when designing a conversation between a person and a virtual assistant?

Several studies in various areas have analysed the wide field of conversations and provide information that is of great value when investigating what should be taken into account to design conversations that can take place between a human being and a machine.

For example, in 2013 the philosopher of language Paul Grice developed principles that conversations must follow to be successful, based on one party always expecting the other party to be fully committed to the topic being discussed. In other words, when we talk to someone, we expect that person to listen to us, to answer us, to generate

a dialogue rather than a monologue where there is only one speaker. These Grice principles are known as Grice's Conversational Maxima, and are summarised below:

Quality	The focus is on what is said to be **accurate** and based on **sufficient evidence**.
Quantity	The aim is to provide **sufficient** information, but **not to exceed it**.
Relation	The focus is on what is said to be **relevant** to the topic in question.
Way	The importance of avoiding ambiguity and promoting clarity and **order**.

Table: Paul Grice's (2013) conversational maxima.

Another element about conversations, interesting to consider, is the Theory of Relevance, developed by cognitive scientists Dan Sperber and Deidre Wilson, while interpreting the expressions that occur within a conversation, in order to identify how we manage to transmit a message beyond the literal meaning of the words we use. In this theory it is presented that within verbal communication there are two intentions on behalf of the communicator, on the one hand, to express something to an audience (informative intention), and on the other hand, to manifest the intention that the communicator has when expressing himself (communicative intention). This theory is especially relevant when we consider the idea of irony and how complex it is to train machines to understand it. Even among people, sometimes these nuances of sarcasm, irony, double meanings are not captured, so it is not surprising that bots are now still in those early stages of literal understanding.

A third very valuable source of information on conversational design, and one that also focuses on conversations that occur with machines, is Cathy Pearl, who describes a series of constructs needed to achieve successful conversations between people and voice interfaces (VUIs). These five *Principles of Conversation* are:

Disambiguation	Ask follow-up questions to understand what is being said.
Contextualisation	Use knowledge of the situation to provide more accurate information.
Cooperation	The importance of giving the necessary amount of information and its relevance to the topic.
Learning	Learn from what the users express and apply that knowledge to make their lives easier.
Demostration	Make it clear what the system knows when it has not understood what the user wants to say.

Table: Principles of Conversation of Cathy Pearl (2016).

Talking to machines: What do people expect from a conversation with a Virtual Assistant?

At Telefónica we have also explored in depth the field of conversations between human beings and machines, more specifically with virtual assistants and with Aura, the Virtual Assistant of Movistar. Through our research in various countries and with very different user profiles, both on a demographic and a psychographic level, we have observed what these conversations are like and how they are articulated.

In this section we will look deeper into the expectations, first experiences, advantages sought and further into the world of conversational design.

Advantages sought

The starting point is to consider what benefits users are looking for when having a conversation with a virtual assistant instead of using an app, web, call centre, or any other channel to interact with a company or service. In our experience, we have identified two main benefits: the first is how these conversations allow the much sought-after multitasking, being able to perform two tasks at the same time. For example, the user can ask his assistant to make a call or set an alarm while he is cooking. In addition to allowing two actions to be performed at once, these conversations with virtual assistants allow them to be conducted by voice, keeping hands free for any other action at the same time. Secondly, users want these conversations to make their lives easier. Naturally, we are designed to fight against change, even if it is positive, it often causes us difficulty, so VA's must make it clear from the beginning how they are going to make the user's life easier and the tasks simpler.

First conversational experiences

On the other hand, it is important to highlight that when users start to have conversations with virtual assistants, they have common experiences and expectations that we have seen repeated globally: (I) High expectations, since users expect to have very complex conversations and to express themselves in the way they want, and that the VA's can understand them and know what they are trying to say. Practically that they are able to read their minds; (ii) The trend to imagine a person behind the machine, behind the response of that conversation. This phenomenon is called anthropomorphism, and is based precisely on this, on attributing human qualities to the non-human, in this case technology; and (iii) the experimentation with the assistant, considering a multiplicity of scenarios that are not limited to the most evident functionalities of the assistant. The aim of this type of behaviour is for users to test the assistant's level of intelligence, as well as its ability to improvise, to deal with random questions, and to demonstrate its potential.

Functioning of the human mind

When it comes to designing conversations, one must also consider the way in which people's memory works when faced with a conversation. Just like when we talk to a person, human beings are only able to retain a certain amount of information, and to remain focused over a period of time when talking to a machine.

Regarding **attention capacity** or *attention spam*, the memory of human beings is only capable of retaining information from known and simple words for 20-30 seconds at a time with a virtual assistant. This amount of time may seem short, but if we think about our experience listening to a radio programme, we will probably find it difficult to reproduce what the speaker said a minute ago, while we will have the last sentences mentioned fresh in our minds.

On the other hand, we have the **cognitive load**, meaning the amount of information per response that we are able to manage. In the case of conversations between users and digital systems, the ideal cognitive load would be between 200 and 300 characters per response. If we offer shorter responses, the user can absorb them very easily, but they may be too short, and if we exceed this amount, it can generate stress and frustration in the user, who is not able to follow the conversation or the instructions being offered. This is why we can enrich a conversation with other elements such as images, videos or written information, as they can exercise a facilitating function, reducing the cognitive load. However, combining the information we present (spoken or written and illustrated) in the experience could also increase the cognitive effort to decode the messages, which is why the design of the content we offer through the interfaces is fundamental. New professionals and a new attention to content care will become more relevant with the proliferation of this technology or relationship model.

It is important to point out that the subject of the information being received impacts on the person's attention span, increasing or decreasing the cognitive load that we are able to manage. It is not uncommon to consider that when a topic really interests us, we concentrate in order to understand and pay attention to the information we are receiving, while when the topic does not interest us very much, we have a tendency to "disconnect" more easily.

How do people really talk

At Aura, through a multitude of tests with users, as well as the analysis that is made of the logs, we have been able to observe the various ways in which people talk to the virtual assistants and the patterns that repeat themselves constantly. Firstly, we would have people talking with **keywords**, without stating a complete sentence, but using only the words that summarise at one glance what they are looking for. For example, saying "Woody Allen" when the user is looking for a film from this director to watch on TV, or "Rosalía" when they want to listen to music from this artist. A second scenario is that of **complete correct sentences**, where users state in a normal and natural way the request they want to make, as if they were talking to someone else "I want to see a Brad Pitt movie" "I want to put an alarm in the calendar for tomorrow at 10 am". Finally, we have observed a third scenario that usually occurs at the beginning of the use of this type of artificial intelligence applications, in which users have difficulty in enunciating the request they wish to make aloud and end up enunciating **very complex phrases** that do not make much sense for the understanding of the receiver, in this case the virtual assistant, or that make use of a context, knowledge, or memories of the user himself. Some examples would be "Could you please put on the film of I don't know which director about a man who meets...? Or "I'd like to hear the song that says something about... the artist ...". In this case the discovery function of this type of system is important, where it is possible to clarify the type of language, even natural, that the assistant understands best, clarifying the users that they should not alter their way of speaking to refer to the requests or tasks they wish to perform.

"What are they showing?" (Person asking what programme is currently on TV)

"Find that one about an airport" (Person referring to the film The Terminal)

"Give me PepaPi" (Child asking to see Peppa Pig)

"Ehh.... I have... ehhh, my different bill this month" (A person having difficulty stating the request about an extra charge on their bill)

Verbatims: Examples of complex decoding requests from users.

To this we must add that the way people speak differs enormously not only around technology, but also in general with other people, and this is due to different aspects. For example, we may encounter **different levels of cognitive maturity**, such as that of a child versus an adult. In this case, the way of speaking, expressing oneself, using vocabulary, etc. is different, and the virtual assistant must take these differences into account. We must also consider the mix **of accents** we will face, since even though we share a common language, depending on where we are from or live, we encounter notable differences when expressing ourselves and even vocalizing the same terms. This also happens with regard to the **localisms** that users can employ, and which must be understood by the virtual assistant. On a third level, we find **peculiarities specific to each person**: trying to find a way to express a request by voice that has never been made before, different ways of calling the same object, a way of speaking or making an unspecific or ambiguous request... Taking all the above into account, we can clearly illustrate the complexity we face when designing conversations and giving all types of users a motivating and relevant experience with a virtual assistant.

Conclusions

In summary, in this article we wanted to express the importance of entering the real experiences of users with this type of virtual assistant in order to offer a variety of aspects and principles that we should consider when designing conversations: the maturation of the experience, usual behaviour, expectations, diversity when expressing ourselves in natural language, etc.

Even though conversational design has a very important technological component, as a company that places the client at the centre of its strategy, when dealing with this discipline we must consider the experience we want to offer the user in order to generate and build a relevant relationship based on trust.

Aura, Telefónica's AI, creates a new model of customer relationship

Aura, Telefónica's Artificial Intelligence, was born as a result of the company's digital transformation process. Its mission is to create a new customer relationship model based on trust, which allows new forms of interaction with users, adapting to their expectations so that they can obtain personalised and immediate responses, using natural language. Aura must be able, firstly, to understand the user and, secondly, to understand the data. Finally,

Aura must be able to connect the two aspects above in order to know what the user might be interested in and how to best provide that information.

Telefónica has been working for many years to be able to provide its customers with an increasingly intelligent home, in which the user can interact with Aura easily and quickly. Telefónica has its own ecosystem, consisting of products, services and experiences, which form the Digital Home.


Telefónica