

Article

An Affective and Cognitive Toy to Support Mood Disorders

Esperanza Johnson ^{1,*}, Iván González ¹, Tania Mondéjar ^{1,2}, Luis Cabañero-Gómez ¹,
Jesús Fontecha ¹ and Ramón Hervás ¹

¹ Department of Technologies and Information Systems, University of Castilla-La Mancha, 13071 Ciudad Real, Spain; Ivan.GDiaz@uclm.es (I.G.); Tania.Mondejar@esmile.es (T.M.); Profesor.LCGomez@uclm.es (L.C.-G.); Jesus.Fontecha@uclm.es (J.F.); Ramon.HLucas@uclm.es (R.H.)

² eSmile, Psychology for Children & Adolescents, 13071 Ciudad Real, Spain

* Correspondence: MEsperanza.Johnson@uclm.es; Tel.: +34-92-629-5300

Received: 22 September 2020; Accepted: 28 October 2020; Published: 31 October 2020



Abstract: Affective computing is a branch of artificial intelligence that aims at processing and interpreting emotions. In this study, we implemented sensors/actuators into a stuffed toy mammoth, which allows the toy to have an affective and cognitive basis to its communication. The goal is for therapists to use this as a tool during their therapy sessions that work with patients with mood disorders. The toy detects emotion and provides a dialogue that would guide a session aimed at working with emotional regulation and perception. These technical capabilities are possible by employing IBM Watson's services, implemented into a Raspberry Pi Zero. In this paper, we delve into its evaluation with neurotypical adolescents, a panel of experts, and other professionals. The evaluation aims were to perform a technical and application validation for use in therapy sessions. The results of the evaluations are generally positive, with an 87% accuracy for emotion recognition, and an average usability score of 77.5 for experts ($n = 5$), and 64.35 for professionals ($n = 23$). We add to that information some of the issues encountered, its effects on applicability, and future work to be done.

Keywords: human–computer interaction; affective computing; cognitive computing; sensorized toy; emotion; embodied conversational agent

1. Introduction

Many of us are familiar with the comic strip of Calvin and Hobbes (<https://calvinandhobbes.com>). Calvin is a child that has many talks, experiences, and adventures, accompanied by his faithful companion Hobbes. What we also learn is that those experiences are the fruit of Calvin's imagination; and his projection of a personality and being onto his stuffed animal, Hobbes. It is with this example that we came upon the idea of using a stuffed animal for our work. In this case, we used a stuffed mammoth, which we randomly selected, and called it MAmIoTie. However, the current model allows for changes in the stuffed animal that is used. The hardware is simple and compact enough to be attached to a different stuffed animal, and the only change regarding the software aspect would be to change the name to suit each specific toy.

This sensorized toy, which is a toy with sensors and/or actuators added to it, was created to be versatile and help with the interaction with patients, typically during therapy. This help during the interaction process would be to help recognize emotions through modeling. In psychology there are no specific phases or structure that would define a therapy session, as they vary according to the specific needs of the patient, the treatment, or the context in which the therapist and patient are working. Therefore, our intention is for MAmIoTie to be a useful tool for the therapist in their work

with patients, specifically in the context of emotion recognition. The emotion recognition is done in the context of supporting the therapist's task guiding a patient to identify emotions. These can be their own emotions, as well as how other people may feel in different situations. To achieve this goal, we are working with a perception triad, which is the model developed in this work to organize the therapy with MAmIoTie. This triad is composed of self-perception, empathy, and social-emotional skills. This translates to a person's emotional self-awareness, the understanding of another person's feelings, and the social and emotional skills involved in general communication with other people. The emotions we focused on were joy, sadness, fear, and anger, which were detected in every answer the patient gave to the questions posed by the MAmIoTie.

To be able to engage the patient in conversation, and also obtain their emotional state during said conversation, we made use of IBM Watson's services (IBM Watson APIs (Application Programming Interface) Available online: <https://cloud.ibm.com/developer/watson/documentation>). Specifically, we used speech-to-text, translate, and tone analyzer to input the patient's words and emotional state. We then used assistant and text-to-speech to return dialogue to the patient and to keep a conversation flowing. This conversation varied within certain options depending on the emotion detected by the tone analyzer. This meant the response from the MAmIoTie changed depending on the emotion detected from the patient. These responses were written with the goal of working through the emotion triad previously mentioned.

The ability of the system to recognize emotion through the tone analyzer qualifies this system as affective [1], whereas the use of cognitive technologies from IBM Watson and the nature of the interaction makes it a cognitive system [2]. These two combined would make MAmIoTie an affective and cognitive toy.

The motivation of this proposal was to assist in therapy sessions geared towards mood disorders. Mood disorders are defined as having a distorted emotional state or mood, which causes interferences with the person's ability to function (<https://www.mayoclinic.org/diseases-conditions/mood-disorders/symptoms-causes/syc-20365057>). There are many ways in which these disorders can be treated, such as several types of therapy [3], with one of them being talk therapy. Talk therapy, also called psychotherapy, is any therapy session that involves the patient talking through their emotions, moods, thoughts, and behaviors, with cognitive behavioral therapy (CBT) being one of the modalities of talk therapy [4]. Therefore, with our proposal, we contribute an easily replicable and technically functional sensorized toy; to support therapists in their sessions with patients who have mood disorders. Firstly, human emotions are difficult to study because they are quite subjective and nondeterministic. In order to make their study easier, we can find basic principles of emotions and their underlying mechanisms by focusing on specific emotions and situations [5]. Secondly, this can give both patient and therapist a different tool to work with emotion perception, which can be adapted to each patient's needs by means of personalized dialogue defined by the therapist.

In summary, we present a sensorized toy in the form of a stuffed mammoth (MAmIoTie) to be used in therapy sessions. Due to the affective and cognitive nature of the system, it works with aspects related to emotion recognition, with the software being supported by an easily accessible hardware (Raspberry Pi Zero). The motivation of this was to give therapists a tool they can use in the context of therapy aimed at working emotions, for instance, with patients with mood disorders.

This paper is organized as follows: In Section 2, we explain the background on which this work is based. Section 3 delves into the hardware and software aspects of the system. We explain the experiments in detail in Section 4, as well as delving into the results. Finally, we conclude the paper with our conclusions and future work in Section 5.

2. Background

In this section, we explore the work we have studied in order to support our own. It is divided in two different sections: Neuropsychological fundamentals, which is mainly the theoretical background

to set the context of the aim of this work; and related work, which presents different works in the area of human-agent interaction, and their scopes.

2.1. Neuropsychological Fundamentals

Emotions play important roles in our everyday lives, as they coordinate sets of responses to internal or external events that have particular significance for the organism [6]. There are extrinsic and intrinsic processes that help us with emotional self-regulation, as they monitor, evaluate, and modify reactions [7]. These abilities to self-regulate change from person to person, as they choose different strategies [8], or are exposed to media (such as video games) which may help them regulate their emotions [9]. Emotion regulation happens every time one activates the goal to influence the emotion-generative process [10]. Therefore, people have a certain style for emotion regulation, which makes them predictable [11,12]. This means that the same strategy can be adaptive or maladaptive, depending on the person, and several other factors [13–15].

One of the ways emotion is expressed is in communication, as speaking with people is the most natural form of communication [16]. Communication aids people in interpreting certain social cues, and guiding emotional self-regulation and social behavior [17]. This has benefits to mental health, as emotion sharing has indirect effects such as constructing or reinforcing of social bonds [18].

Both emotional regulation, and the sharing of emotions, are associated with the quality of social functioning among children and adolescents [19]. Emotional self-regulation is a process learned by experience, but in some cases, there is an added difficulty for those who present disabilities in this area. This is the case for autism spectrum disorder (ASD), which usually involves persistent deficits in emotion perception and social interaction. Some of the situations in which this is visible is the ability to maintain a conversational flow, an aversion towards sharing feelings and emotions, fitting in a social context, etc. [20]. This is an example of a specific situation that affects communication, where there is an added difficulty in patient–therapist communication. This is because there is a chance of important information being left out due to gaps in communication.

However, there is a wide variety of disorders related with emotional perception. It can either be a factor that aids in the diagnosis, or a special need that requires treatment. For this reason, we focused on the classification of the Diagnostic and Statistical Manual for Mental Disorders [20], specifically, on neurodevelopmental disorders that affect intelligence and communication (i.e., autism spectrum disorder (ASD), Down syndrome, etc.).

For our approach with people within these groups, we looked into the theory of mind, which refers to “the understanding of people as mental beings, each with their own mental states, such as thoughts, wants, motives and feelings. We use theory of mind to explain our own behavior to others [. . .], and we interpret other people’s talk and behavior by considering their thoughts and wants” [21]. This also states that there are consequences to theory-of-mind development in children, such as their social competence. This social competence can be affected by certain disorders previously mentioned, such as ASD, where they can have issues ranging from understanding stories, to incorrectly attributing hostile intentions in positive interactions and vice versa, up to the way they respond to peer rejection (aggressive or avoidant) [22]. This peer rejection can also be seen in other disorders that affect moods, such as bipolar disorder (BD), with the possible root being the misinterpretation of nonverbal cues related to emotion recognition [23].

Therefore, with this proposal, we aimed to provide a tool for therapists to use with their patients, that would complement the other tools already at their disposal. The patients that could benefit from using this tool are left to the discretion of the therapist, as this solution is not a one-size fits all, and ultimately, the decision to use this tool or not would fall on the therapist.

2.2. Related Work

After exploring the neuropsychological background, we now proceed to explore the related work. In this section, we observe the different works encompassed within the area of human–computer

interaction, from less to more specific. After each, we underline the learned lesson from each category, and finally, elaborate on how our proposal differs from the presented works in this section.

We can start by exploring the possible differences between human–human interaction vs. human–agent interaction. In [24], it was found that people readily engage in communicative behavior if something appears sufficiently social, and establishing a relationship depended on the system’s variables and characteristics. Given this, it is no surprise that several works have been developed studying different aspects of human–agent interaction, and its varieties within the different subsets.

These subsets depend on the kind of agent one works with. One of those agents can be virtual, and can be interacted with in a variety of ways. This can be through tactile interaction [25,26], where the participant interacted with the avatar through a tablet; through text [27], where students write to an avatar designed to help students with stress management; or through body movement by detecting apathetic or euphoric dancing and modulating it through a virtual human [28]. From these works, we can observe that there are different applications that can be done with virtual avatars, through different mediums. These works work with emotion detection through different types of interaction, and then attempt to reconduct extreme emotions through different means, such as modulating physical activity, or talking with the participant.

Another approach within the area of human–agent interaction is that done with robots. Seymour Papert [29] has done work on the interpretation of robots’ behavioral patterns, as well as the clear dialogue, and started research on the introduction of robots to education in the 1980s. Some of these works with robots have been with Nao [30], aiming to investigate whether children with ASD show more social engagement when interacting with Nao; and Milo [30], where the latter is programmed to educate children with autism spectrum disorders (ASD) to show emotions and facial expressions that transmit feelings by having the children imitate the expressions of the robot. In this, we can see how the works use the robot’s abilities to express some emotion to give an example for the users, and to better help them identify the emotions the robots are expressing.

There are also examples with sensorized toys, such as a work that uses a tangible toy to help children with neurodevelopmental disorders, where it provides multisensory stimuli through multiple sensors and actuators for luminous and sound stimuli, which can be used to develop fine motor skills and basic cognitive functions skills [31]. Another proposal is that of a sensorized baby gym for infants, with a variety of sensors attached to different elements of the baby’s environment, like grab handles, or visual stimuli [32]. Finally, a work to support children with ASD in learning vocabulary, mathematical, and life skills [33] using a SmartBox device with multiple sensors (body sensor, light control, sound control, etc.) to monitor the children and their calm–alert states and create P2P communication between children and caregivers or therapists.

As we can see from a general overview of each group of works, human–agent interaction has established a use when it comes to user interaction and emotion detection and portrayal. This serves so that the user can identify their own emotions, identify the emotions portrayed by the agent, or so that the agent can identify the user’s emotions. We do, however, see a lack of this in the area of sensorized toys.

Given this, we feel our work offers a novel contribution on two levels. In terms of design, we are contributing to the literature by detailing the components that form our sensorized toy (MAmIoTie). On the other hand, we also aim to contribute to the use of sensorized toys as a tool at the disposal of therapists to use in support of diagnosis and treatment of mood disorders. The sensorized toys seen in the related work section are also mainly focused on being used primarily by children, whereas our proposal is aimed at both the patient and the therapist. With it, we offer an option for the therapist to use with patients in therapy, as another tool similar to paper-based tests or more traditional exercises, with an added element of interaction and engagement.

In the following section, we explain how those two contributions have been made, by delving into the system specification, both from a hardware and software point of view, as well as what can be detected through these means.

3. System Specification

To achieve the contributions mentioned in the previous section, we designed a hardware infrastructure to support the software that comprised the functioning of the MAmIoTie. With this, the patient spoke to MAmIoTie, which answered accordingly. To reach that answer, it transcribed the speech into text, translated that from Spanish to English, and used the translated text (in English) to input into the tone analyzer. The translation was necessary due to the fact that the tone analyzer service from IBM Watson only works for English and French. Once it had extracted an emotion from the patients' words, it sent that to the assistant (where the dialogue options were stored), and depending on the emotion, a different option of the dialogue was selected. Finally, the output of that was sent to a text-to-speech module so the patient heard it, as if spoken by MAmIoTie.

Therefore, in the following subsections, we explain the hardware infrastructure (Section 3.1), that supported the software infrastructure (Section 3.2) that allowed for the above system to work.

3.1. Hardware Specification

To transform an average stuffed toy animal into a sensorized toy, one needs to look no further than towards simple and affordable hardware. For this purpose, we used a Raspberry Pi Zero WH single-board computer (Raspberry Pi Foundation, Cambridge, UK), with a couple of dedicated modules that composed the interaction layer (microphone and amplifier). This setup can be seen in Figure 1.

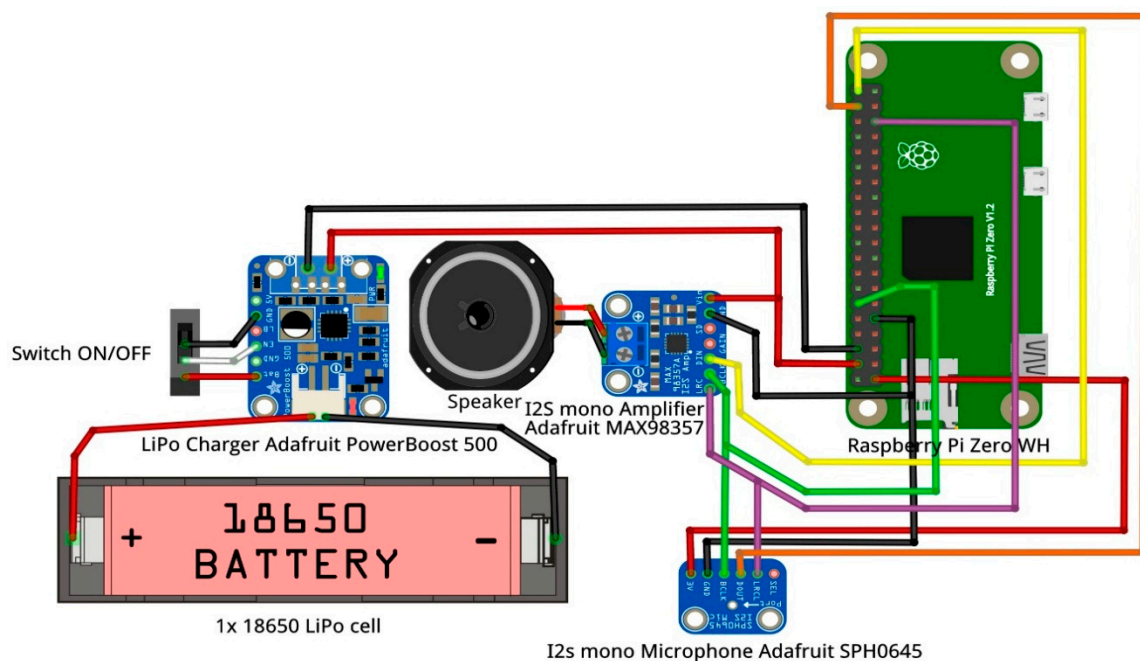


Figure 1. Hardware setup embedded into MAmIoTie, the sensorized toy.

The Raspberry Pi Zero WH single board was not selected by chance. This device has an embedded BCM2835 system on chip (SoC) developed by Broadcom (Broadcom Inc., San José, CA, USA). This SoC integrates a 1 GHz Arm11 [34] single-core processor (ARM1176JZ-F), which is combined with 512 MBytes RAM. This configuration is enough to run a GNU/Linux distribution (Raspbian) compiled for ARMv6 architectures. Raspbian provides support for Python 3 programming language, which is used to implement the software modules (Section 3.2) and to enable communication with the IBM Watson REST services that afford the affective perception functionalities.

The BCM2835 SoC also incorporates 802.11n Wi-Fi and Bluetooth low-energy wireless transceivers. Full TCP/IP stack support is provided by the Pi Zero WH running Raspbian and Wi-Fi communications capabilities through the connection to a close WLAN access point are crucial aspects that allow software modules to send service requests to IBM Watson and receive relative responses through the Internet.

Another key feature of the Raspberry Pi Zero WH is the Inter-IC Sound (I2S) [35] bus support. The I2S interface provides a bidirectional, synchronous, and serial bus to off-chip audio devices. In our case, it makes it possible to directly transmit/receive pulse-code modulation (PCM) (a quantization method used to digitally represent sampled analog signals) (digital) audio in both directions, from the I2S mono microphone (audio capture) to the BCM2835 SoC; and from the latter to the I2S mono amplifier (audio playback). Both external modules get direct I2S communication through the general-purpose input/output (GPIO) (generic pins on a computer board whose behavior—including whether they work as inputs or outputs—is controllable at run time by the control software) header on the Pi Zero WH board. The I2S bus separates clock and serial data signals, resulting in a lower jitter than is typical of other digital audio communications interfaces that recover the clock from the data stream (e.g., Sony (Tokyo, Japan)/Philips (Amsterdam, The Netherlands) digital interface (S/PDIF)). According to the I2S specification [35], three pins/wires are required to communicate every two integrated circuits (ICs):

- Left/right clock (LRC)—This pin indicates the slave IC when the data is for the left/right channel, as I2S is designed to work with stereo PCM audio signals. It is typically controlled by the I2S master (the BCM2835 in this case);
- Bit clock (BCLK)—This is the pin that indicates the I2S slave when to read/write data on the data pin. It is also controlled by the I2S master (BCM2835 in this case);
- Data in/out (DIN/DOUT)—This is the pin where actual data is coming in/out. Both left and right data are sent on this pin, and LRC indicates when the left or right channel is being transmitted. If the I2S slave is a microphone or a transducer, it has a data out behavior. Conversely, if the I2S slave is an amplifier wired to a speaker, the pin is of the data input type.

Pi Zero WH enables two I2S stereo channels, one as input and one as output. Thus, a configuration with an I2S stereo amplifier and a couple of I2S mono microphones could be made. However, in order to make MAMIoTie as affordable as possible, we set up a prototype with a single I2S mono MEMS microphone (SPH0645) and an I2S 3.2 Watts mono amplifier (MAX98357), both from Adafruit (Adafruit Industries LLC., New York, NY, USA). In fact, the MAX98357 amplifier is a DAC (Digital-to-Analog converter) module whose analogue output is directly wired to a 4-ohm impedance speaker to emit sound. The SPH0645 microphone, for its part, acts like an ADC (Analog-to-Digital converter), transforming analogue audio signal into a digital one. Due to these integrated DAC and ADC converters, digital signals are transmitted through the I2S interface instead of using old schemes transmitting pure analogue signals.

Figure 1 shows at a glance how I2S modules have been connected to the Pi Zero WH board. The hardware setup is completed with a 3000 mAh 3.7V 18650 Li-ion battery and an Adafruit PowerBoost 500 module that charges the battery and protects it from overloads and over-discharges.

Lastly, the mammoth (MAMIoTie) with the hardware elements can be seen in Figure 2.



Figure 2. Non-intrusive integration of the hardware with MAMIoTie.

3.2. Software Infrastructure

The infrastructure described above is what allowed for the components of the software infrastructure of MAMIoTie to work. The microphone was how the patient speaks to the MAMIoTie, and it was the hardware element that gave support to the speech-to-text module. Speakers were how the text-to-speech output was played back to the patient, whereas the Raspberry Pi contained the entire program, which was in a script written in Python.

The way these elements worked together to have an interactive sensorized toy can be better seen in Figure 3 below.

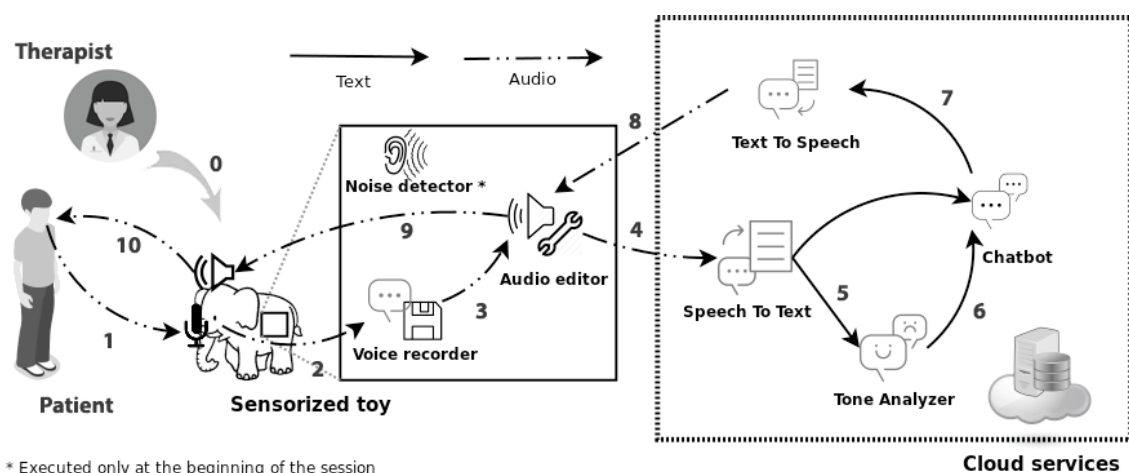


Figure 3. Connection of the hardware and software elements, and the flow of interaction.

In Figure 3, we can observe the order of the events and elements of MAMIoTie. It starts off with the patient speaking to MAMIoTie, which would be picked up by the microphone (1). This is recorded (2), and sent to an audio editor (3), and sent to the cloud services. The first service that is called is the speech-to-text (4), which then sends the transcription of the audio to both the chatbot and the tone analyzer (5). The tone analyzer then sends the detected emotion to the chatbot (6) which contains the dialogue options. Once the dialogue option is selected, it is sent to the text-to-speech module (7) to return a sound file. This sound file is then sent to the audio editor (8) to modify the pitch, and then to the speaker (9), which finally reproduces the dialogue to the patient (10). As can also be seen in the figure, there is a step (0), which refers to the therapist. This is due to the fact that the therapist is the person who can decide the use of the toy. They can also modify or configure a new dialogue to fit the specific needs of the patient and of their approach, due to the Watson assistant tool. This tool is available through Watson's services through a web platform, and it allow for access to the dialogue tree which MAMIoTie uses to guide the dialogue. This dialogue is formed by a question bank, based on the perception triad previously mentioned in Section 1, and is merely an example we are using for the evaluations present in this paper.

All the steps from 4 to 7 involve the use of IBM Watson's services (cloud services in Figure 3), which form part of the overall software infrastructure. The elements shown in Figure 3 are:

- **Noise detector:** The first method to run: It captured audio samples from the microphone. From there, it determined the noise level, and obtained an average using 30 percent of the highest values (limit obtained experimentally (this limit was obtained through an experiment with end users. Details explained in Section 4.1. Experiment 1: General evaluation));
- **Real-time voice recorder:** Using Python audio libraries and the microphone, audio was recorded and then analyzed to distinguish between silence and noise. If the threshold defined by the noise detector was exceeded, it saved the recorded audio, until it fell below said threshold for a period of time (a range of silence samples, determined experimentally). This determined the end of the patient's sentence, and then generated an audio file with the recording;

- Recorded audio editor: The generated audio file was then modified to add silence intervals to the beginning and the end of the file, to normalize the volume, and to improve speech recognition later;
- Speech-to-text: The resulting audio after analysis was then sent to the Watson speech-to-text service, which returned a text transcription of the sentence it detected. If the service did not recognize anything coherent, it returned to the audio capture state;
- Tone analyzer: Once we obtained the text from the transcription, it was then fed to a translating service of Watson, as the tone analyzer does not currently work with Spanish. Once we translated the transcription, we fed it to the tone analyzer to obtain the emotion from the text (joy, anger, sadness, or fear). This was a purely text-based analysis, and was not influenced by any changes in pitch or tone the user made. Therefore, no information would be lost during the speech-to-text part of the process;
- Chatbot: The emotion obtained from the tone analyzer, along with the transcription from the speech-to-text, was then fed to the assistant service. This took the patient's input and emotion, and returned a textual answer depending on that input;
- Voice synthesizer: The result from the chatbot was then fed into a voice synthesizer, which converted that text to audio through a text-to-speech service. The resulting audio was modified to make it have a sound that aligned to the look of MAMIoTie. Finally, the audio was played through the amplifiers with the help of the Python audio libraries.

Of all the infrastructure listed above, the chatbot service (IBM Watson's assistant) was the element that contained all the dialogue options that MAMIoTie spoke to the patient. Those dialogue options were decided based on the theory of mind [36]. These influenced the kind of questions MAMIoTie asked the patient, based on said theory and with the input of a psychologist specializing in neuropsychology. This was done to work with aspects related to emotional recognition on several levels. The order in which these questions were posed is as follows: Self-perception explores the emotions related to the patient, starting with a direct question ("How are you feeling?"), and then an indirect question ("What happened the last time you felt ...?"). Empathy focuses on the emotions of someone close, in this case, MAMIoTie, who asked the patient a direct ("A situation happened, how do you think I felt?") and indirect question ("What do you think happened the last time I was feeling ...?"). Finally, the social-emotional dimension works on feelings regarding other people, and again it asked a direct ("A situation happened, how do you think this person felt?") and an indirect question ("What do you think happened for this person to feel ...?"). The dialogue then took the patient's first identified emotion to start off, and then varied the emotions for the empathy and social-emotional round of questions. MAMIoTie also answered those questions with impressions depending on the detected emotion ("I'm happy to hear that/I'm sorry to hear that"), as well as with corrections on how it felt at different times when the questions pertained to it ("I was actually feeling scared/sad/angry/happy"). When it finished the social-emotional indirect question, it went back to the self-perception direct question.

It worked with four emotions (joy, sadness, anger, and fear), and then it explored the 3 dimensions with both direct and indirect questions, making a total of 28 questions. Once they had all been answered, MAMIoTie informed the patient and bade goodbye, ending the interaction, and saving a log of the interactions with the patient from start to finish. This flow can be seen in Figure 4 below.

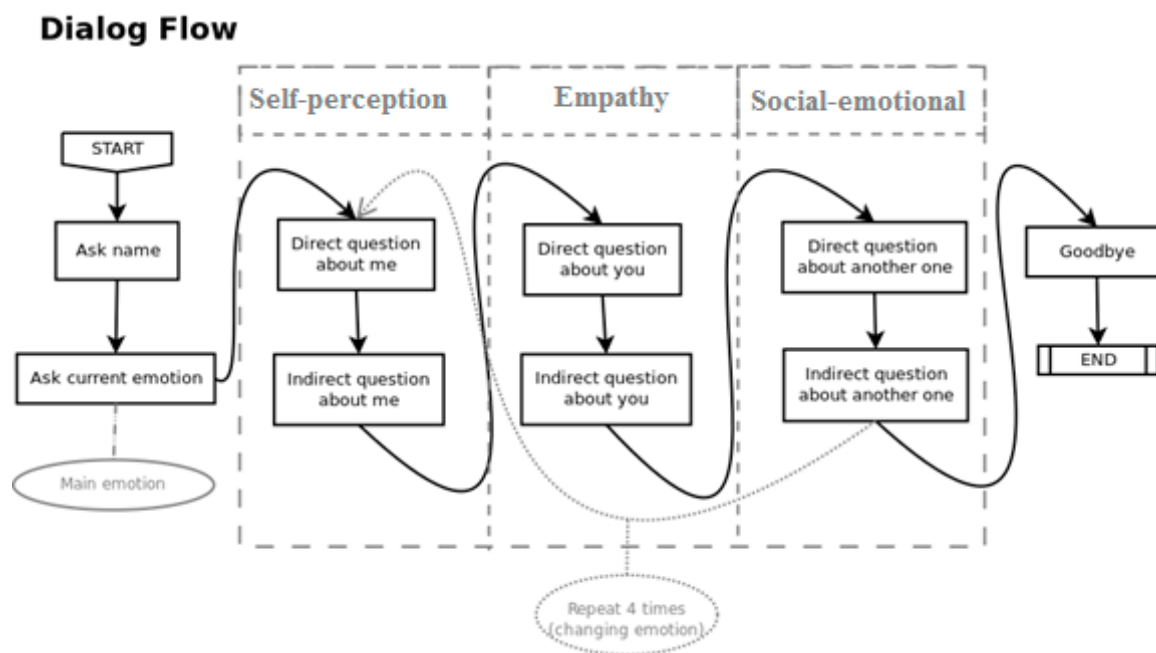


Figure 4. Dialogue flow, where the system goes between direct and indirect questions, and first, second, and third person. This cycle is repeated a total of 4 times in order to explore the four basic emotions (happiness, sadness, anger, and fear).

4. Evaluation

In this section we describe the three experiments that were performed with MAMIoTie, which we can divide into a general evaluation (Section 4.1), an evaluation with a panel of experts (Section 4.2), and an evaluation with several professionals (Section 4.3).

4.1. Experiment 1: General Evaluation

This experiment was performed with 20 participants, aged between 10 and 18, with a neurotypical profile (no known cognitive disorders). They were recruited with the help of one of the co-authors, and they were unfamiliar with the technology presented to them.

The goal of this experiment was to evaluate the system on a quantitative level. For this, we focused on performance and emotion recognition, to gain an understanding of the technological issues that the system has, and where it would need improvement. This phase corresponded to the interaction between the participant and MAMIoTie.

The experiment began by explaining the collected data to parents and the children who would be performing the experiment, assuring that no personal data would be collected. After obtaining consent, we proceeded to tell the participants that the experiment would consist of interaction. We explained that they would need to speak clearly, and to wait for the conversation to be started by MAMIoTie, and gave no further instructions.

Then, MAMIoTie began the interaction by asking the participant for their name. This time spent in silence before the start of the dialogue served to determine a sound threshold (described in Section 3.2) to better identify the level of noise when the participant was talking. It also helped to establish familiarity between the participant and MAMIoTie. After the name had been given, MAMIoTie asked a question to determine the participant's emotional state. Depending on which emotion was detected, the dialogue started accordingly, and then went through the rest of the questions, as seen previously in Figure 4, in Section 3.2.

The data collected during this phase of the evaluation were inserted into the generated CSV (comma separated values). These data were the dialogue sentences that were spoken by MAMIoTie, the transcription and translation (for metric purposes) of the participant's answer, and the expected

and detected emotion to each of those questions. We also recorded if the question asked was direct or indirect, as well as if it was related with self-perception, empathy, or social-communication skills.

The results for this part of the experiment measured the performance on a technical level. There was a total of 18 participants with valid results, and 2 of the participants were discarded, as their results were not recorded adequately (more details in Section 5). With the recorded data, we measured accuracy, word error rate [37], and emotion recognition. We started by looking at the word error rate (WER), which is a common metric used to measure the performance of speech recognition or machine translation. This formula can be seen below in Equation (1).

$$\text{WER} = (\text{Substitutions} + \text{Insertions} + \text{Eliminations})/N \quad (1)$$

where

1. N is the total number of the transcribed words;
2. Substitutions are the words substituted from the original sentence;
3. Insertions are any words inserted that were not in the original sentence;
4. Eliminations deals with words not present in the original sentence.

We later also employed a similar metric to measure the accuracy for the translation of the transcribed text. The result from this calculation gave a result where, the closer it was to zero, the better that was, as it implied a low number of substitutions, insertions, and eliminations. These two aspects were critical for the later emotional recognition, as IBM Watson's tone analyzer service works only in English. We first analyzed the results for the speech-to-text WER, and then analyzed it for the translation service results.

With a total of transcribed words of $N = 770$, we observed a WER of 0.24675. Looking more closely at Table 1, we observe that the main issue that contributed to the value of the WER were the substitutions of words that could be slightly misunderstood, or because the participant did not speak clearly or loudly enough.

Table 1. Speech-to-text word error rate (WER) results, for each of the participants as well as global results.

User	WER	Substituted	Inserted	Eliminated	Word Count
1	0.2639	10	2	1	56
2	0.245	23	2	6	137
3	0.3611111	3	2	0	19
4	0.58902	19	7	5	58
5	0.1664502	13	2	6	96
6	0.333	2	0	0	16
7	0.38755	13	2	11	99
8	0.305556	3	0	5	28
9	0.627976	4	1	2	22
10	0.134921	4	0	0	29
11	0.358796	7	5	0	36
12	0.047619	2	0	0	25
13	0.388889	1	1	2	34
14	0.222222	3	1	0	21
15	0.40873	5	0	0	20
16	0.186508	3	0	1	22
17	0.296296	0	4	2	30
18	0.222222	3	0	2	22
Total	0.24675	118	29	43	770

As for the WER value for the translation aspect, we obtained $N = 756$, with a value of 0.17596, where the results per user as well as the total can be seen in Table 2. An observation on this value was the fact that translating one or two words may change the meaning of the entire sentence. A fairly

common occurrence was the translation of “because” for “why”, which then affected the sentence to be understood as a question. The implications of this are discussed in the next section.

Table 2. Translation WER results, for all participants and global result.

User	WER	Substituted	Inserted	Eliminated	Word Count
1	0.1181	4	0	4	55
2	0.283	14	1	17	123
3	0.3547619	2	3	2	20
4	0.18426	4	1	6	48
5	0.1319444	8	0	8	97
6	0.222	3	0	2	16
7	0.15	12	0	5	87
8	0	0	0	0	30
9	0.451389	2	1	3	26
10	0.116667	2	0	2	28
11	0.335317	5	0	4	34
12	0.283333	3	0	2	27
13	0	0	0	0	35
14	0.045455	1	2	0	24
15	0.078704	2	0	2	24
16	0.055556	0	1	0	23
17	0.231481	1	1	3	33
18	0,125	0	1	1	26
Total	0.17857	63	11	61	756

As we can see from both tables, substitutions were the biggest contributor to mistakes further on in the process. During the translation, we also see that there were several eliminations. However, this aspect was slightly more subjective, as the comparison was done with the translation of the transcribed text a native would have done. This meant that there could have been ways to say the same thing, with less words.

At the same time that we obtained the transcriptions and translations for what the participant was saying, we also recorded the results of what tone those words had, according to IBM Watson’s tone analyzer service. We obtained this from the end-user interactions, from which we had an 87.63% accuracy rate between recognized and expected emotion. This meant that if we expected a negative emotion, and a negative emotion was detected, we counted that as accurate. It is important to distinguish how many times the system was not able to obtain an emotion from the captured text. This number was 67 “none” of the total 186 detected emotions, which implied 36% of the total emotions recognized. This information can be more easily seen in Table 3 below. If we were to disregard the number of “none” from the calculation of accuracy percentage, we would obtain an accuracy of 80.67%.

Table 3. Accuracy results for emotion recognition, and instances and percentage for no recognized emotion (“none”).

Mismatched emotion	23
Accuracy rate	87.63%
Total instances “none”	67
% instances “none”	36.02%

As we discussed in the previous section, the data the tone analyzer received as input went through both the speech-to-text service, as well as the translator service, since it did not work with Spanish input. As previously seen, speech-to-text had a few transcription problems, with the main issue being the substitution of words that sounded similar enough, but had very different meanings. With those issues going uncorrected and being fed into the Translator service, other issues were added, such as “because” being translated for “why”, which ended up being understood as a question. All of this

meant that those sentences lost part of their meaning, so they were either misunderstood, and therefore the emotion was mislabeled, or they were not labeled at all. One plain example is where participants answered to a question with “triste” (“sad” in English), which speech-to-text transcribed to “viste” (“saw” or “dress” in English), and then translated to “dress” (one of the options). Thus, there was a similar effect to the very commonly played “telephone game”, where a person whispers a word, and it gets changed, so by the time it reaches the last person, it rarely ever resembles what was initially said.

However, we have observed the effects of this to the overall recognition were small, as can be seen in Table 4 below.

Table 4. Accuracy results for emotion recognition, and instances and percentage for no recognized emotion (“none”) with a native translation instead of IBM Watson’s.

Mismatched emotion	28
Accuracy rate	84.95%
Total instances “none”	63
% instances “none”	33.87%

As can be seen, there was a slight increase of mismatched emotions, lowering the accuracy rate by less than 3%. However, there were also a few less instances of “none”, implying there were some of sentences that were understood better without the loss of meaning from the translator; however, the detected emotion did not match what was expected. Having done this, we compiled the instances where the expected and detected emotions were matched. In the cases where they did not match, we could also see which emotion it detected. This left us with a total of 71 emotions that matched perfectly, 48 in which there was emotion confusion, and 67 instances of no recognized emotion. Out of those 48 instances of emotion confusion, 23 of those were classed as mismatched emotions (joy confused with sadness, any of the negative emotions confused with joy, and sadness or anger confused with fear), leaving 25 to be classed as matching emotions. These results can be seen in Table 5 below, where the left column shows the expected emotions, and the top row, the detected emotions.

Table 5. Confusion matrix for emotion recognition.

	Joy	Sadness	Anger	Fear	None
Joy	30	5	0	0	25
Sadness	3	15	2	2	23
Anger	6	13	17	2	11
Fear	4	10	1	9	8

As stated above, we can see that the main problem was mixing any emotion for no emotion at all. There were two other standout data, which were the confusion of anger with sadness, and fear with sadness. In truth, we did not consider the “confusion” of anger with sadness a problem, as they were on a similar spectrum. We also acknowledge that different people can feel differently about certain situations, but within a range. For the questions posed in this experiment, it seemed to stem from a specific one, where the child was asked how they would feel if they were to be grounded, and we set “anger” as the expected emotion. However, many of the participants replied with answers that were then read as “sadness” by the tone analyzer, which was still a very valid feeling for such a scenario. Fear being confused with sadness came from a similar question, where MAMIoTie asked how the participant thinks he feels when lost in a forest.

Initially, the latency of the overall system was, on average, 5.49 s. This time was measured from the moment MAMIoTie spoke a sentence, until it asked the next one, going through the full cycle of interaction that can be seen in Figure 3. In subsequent tests and after reimplementing of the system, with an alternative design in the use of cognitive services, latency was reduced to less than 1.5 s; still somewhat high, but sufficient for fluent conversation.

4.2. Experiment 2: Panel of Experts

For the second part of the experiment, applicability was the aspect that was measured, which gave us an understanding of how those issues could impact the system being applied in a therapy session. For this, the panel engaged in a focus group after testing MAmIoTie. This was followed by filling out a questionnaire based on the system usability scale model [38]. Most of the questions were the same as in the original model, where the only modification was to ask about MAmIoTie. The question which was more heavily modified was question 1, which had to be adapted to the context of this experiment. The adapted SUS (Systems Usability Scale) questionnaire can be seen in Appendix A (Figure A1). The overall session gave us results that could be explored both quantitatively (SUS-adapted questionnaire) and qualitatively (comments and discussions after group session).

The panel of experts was formed by five professionals in the areas of psychology and technology. In the panel, there were two psychologists who work with children and adolescents, both neurotypical and non-neurotypical. There were also two researchers familiar with human–computer interaction applied to health. Finally, there was a researcher familiar with human–computer interaction, who has also undergone the study of a B.Sc. in Psychology. The importance of having psychologists in the panel came from the fact that they would be the participants who would decide whether to integrate this system into their sessions. They were also the participants who could tell us issues with our system that researchers might not have insight to, as their field of work is interacting with the other set of participants who would directly use the system.

The protocol for this experiment started very similarly to Experiment 1 (Section 4.1), where the panel was asked to interact with MAmIoTie. Specifically, all participants of this experiment were asked to get together at the same time and place, and choose one participant to interact with MAmIoTie, while the rest observed. This was so that the panel could gain an understanding of how children and adolescents in therapy sessions would interact with MAmIoTie.

After doing this, the panel engaged in a focus group, where they all discussed together in a group setting any issues that our current model may have had, as well as the potential of the system to be used for the intended purpose of supporting the diagnosis and treatment of mood disorders. Along with the group session, which was recorded for later analysis, they also filled out questionnaires that were adapted from the system usability scale [35], to fit our experiment and system.

The results from the SUS model (which we have called MUS for the MAmIoTie usability scale, and can be seen in Appendix A) were an average of 77.5 points. According to a study [39], a score over 68 points was considered to be above average, which implied our system was well scored and had good reception. The answers were towards either extremes of the scale, making it easy to identify any outliers, or any strong opinions in general. From observing the results of the questionnaires, the impressions were in agreement with the positive statements (i.e., communicating with MAmIoTie is simple), and disagreed with the more negative statements (i.e., communicating with MAmIoTie is complicated). One of the experts (researcher) disagreed with the statement “MAmIoTie has a fluent conversation”, stating that it could be improved with a knowledge base gained from conversational experience. Another expert (psychologist) answered to a lot of the questions with a “neither agree nor disagree” approach, and provided the comment that it would be very positive to employ in a therapy session, with some improvements.

Some of those improvements were discussed during the group session of the evaluation. They agreed on the applicability of MAmIoTie for the goals we had thought, and provided some interesting insights. They agreed that a more personalized approach would be important to implement, to make it seem more like a conversation, and less like it was just going through a series of questions, i.e., having a more fluent conversation, in terms of having those questions be more related to each other, and therefore having a more natural flow to the conversation itself. There were also comments on the technical aspects of MAmIoTie, such as the time spent between asking one question and the next. The panel seemed to think it could have potentially positive applications, particularly to exercise patience, and to have a “think before you speak” example. Some of the specific comments made by the

panel during the session are listed below, divided into technical or application comments, concerns, or insights.

Technical comments:

- Connect ideas, have more connection options in the dialogue. This would make dialogues more dynamic;
- Parameterize the dialogue options, meaning to make the dialogue options general, and have the possibility to specify for different cases;
- Have more content, and extract content from conversations to include in a knowledge base;
- An LED light to indicate when it is listening, and when it is thinking, would be very useful.

Application comments:

- Fit itself to age profiles, as children are not as inclined to follow up questions with other questions for the patient. It could develop patience, as they have to wait for MAMIoTie to answer;
- Working with a toy is different, because children do not feel like it would tell their parents;
- Children do not ask as much as adults, and do not tie in conversations as much;
- It would be good for children, as dialogue with them is much more straightforward.

4.3. Experiment 3: Professional Evaluation

In order to determine the validity of this proposed solution for the professionals, we performed an evaluation with several professionals, of different backgrounds. The professionals were sent an online form which contained demonstration videos of the use of MAMIoTie, as well as an explanation of its goal. After seeing these videos, they could fill out a form with questions. This form had some of the elements of the SUS scale used in the previous experiment, but was modified to obtain more information from the professionals filling it out. Apart from the SUS-based questionnaire, they also filled out their profession and job title, as well as their opinion on MAMIoTie, and which mood disorders they thought it could be applied to. This form was anonymous, and no personal information was collected.

There was a total of 23 responses, which included therapists, psychologists, speech language pathologists, and teachers, from specialized associations in both Spain and UK. This variety of responses gave us a better look at what different professionals felt they needed from a sensorized toy, from different perspectives, giving us a more complete picture of both the good points of the toy, as well as the improvements that were needed.

After watching an example video of use of the toy, as well as two explanatory videos about MAMIoTie, its goal, and its details, they filled out a questionnaire, which resulted in an average score of 64.35 points. As a matter of fact, the first question, which was “I think I could use MAMIoTie frequently in therapy”, obtained the highest average score. This was particularly good, as it was the main aim of this proposal.

The SUS determined that a score above 68 indicated a good usability, and therefore, this solution came close to that score, but did not quite reach it. In order to better evaluate possible reasons for this score, we looked more closely at the responses from the test.

One of the observations we could make was to see if there were any differences in scores according to the profession. As can be seen in Table 6, there were notable differences in the average score depending on the profession, with teachers giving a score of 34, followed by speech language pathologists (SLP) of 60, psychologists of 69, and finally therapists of 74. This meant that the participants who were more closely related to treating mood disorders (psychologists and therapists) gave it the highest scores (69 and 74, respectively). Those scores were also above the 68 mark that the SUS determined to be the indicator of a usable system.

Table 6. Table with the average score per group.

Profession	Participants	SUS Score
Speech language pathologist	6	60
Psychologist	11	69
Therapist	4	74
Teacher	2	34

In order to better understand the scores, we also obtained the average, mode, and standard deviation for each one of the answers to the questions of the tests.

As can be seen in Table 7 below, there are several conclusions that can be extracted. Firstly, we observed that Q2, pertaining the complexity of the communication with MAmIoTie, had the worst results. The average score was closer to 3 (2.7), whereas the mode was actually 4. However, it can be seen that there was a wide range of results, due to the standard deviation also being the highest in any of the questions. This would imply that some people disagreed on the complexity, but others agreed more. On the other side of this, we had several questions (Q1, Q3, Q5, Q6, and Q7) with scores that were closer to what would be considered positive results (close to 2 for Q6, and close to 5 for Q1, Q3, Q5, and Q7) according to the scale used in the questionnaire. The specific answers of each participant to this questionnaire can be seen in Appendix B, Questionnaire results (Figure A2).

Table 7. Questions, overall average score for each question, as well as mode and standard deviation.

#	Question	Average	Mode	Standard Deviation
Q1	I think MAmIoTie could be used in therapy	4.2	5	1.1
Q2	Communication with MAmIoTie is complicated	2.7	4	1.3
Q3	Using MAmIoTie is easy	3.96	4	0.91
Q4	I would need help to use MAmIoTie in therapy	2.65	3	0.81
Q5	MAmIoTie has a fluent conversation	3.39	4	1.21
Q6	There are inconsistencies in MAmIoTie’s dialogue	2.26	1	1.22
Q7	Most people can quickly learn to use MAmIoTie	4.13	4	0.61
Q8	Communication with MAmIoTie is slow	3.3	3	1.04
Q9	MAmIoTie transmits trust	3.44	3	0.97
Q10	I feel I need to learn many things before using MAmIoTie	2.44	3	0.88

On the one hand, Q3, Q5, and Q7 were questions regarding the use of MAmIoTie, fluency of communication, and learning speed. The other question (Q6) dealt with inconsistencies in MAmIoTie’s dialogue, which also happened to have the lowest average score (2.26), with a mode of 1, and a standard deviation of 1.22. While the standard deviation was high, this indicated that some experts disagreed, and they answered with a high score, but not a high enough number that the average and mode would shift to presenting this as an issue in MAmIoTie.

The highest average score was for Q1 (4.2), with a mode of 5, and a standard deviation of 1.1. From this, we can infer that many experts agreed that MAmIoTie was a solution that could be used in therapy, which was the main goal of this proposal and evaluation.

Given this last answer to Q1, the question that arose was for what cases of mood disorders would professionals use this sensorized toy with. This was the reason why we added a multiple-choice question as part of the questionnaire. This asked the professionals for which of the disorders would they use MAmIoTie, giving several options, and adding a blank answer, in case they could think of a case we had not contemplated.

These answers were recorded and displayed as a graph, which can be seen in Figure 5.

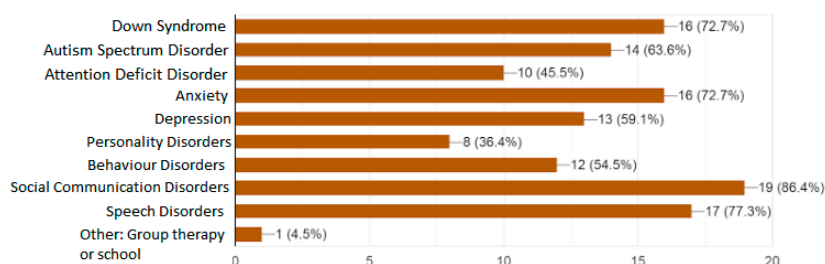


Figure 5. The question “Who would you use MAmIoTie with?”, with the answers represented in a bar graph of the different cases professionals would consider using MAmIoTie with.

As can be seen, the most selected option was social communication disorders, followed closely by speech disorders, and a tie between Down syndrome and anxiety. On the lower end, some professionals selected personality disorders, attention deficit disorders, behavior disorder, and depression. One person selected “Other”, and specified a possible use of MAmIoTie in group therapy, or use in schools, to teach about empathy. From this, we can gather that not only would several professionals use this toy, but that there are several different cases and options to use MAmIoTie, including some we had not initially considered (such as group therapy).

Lastly, just as in Experiment 2 with the panel, we included a long-answer question, where the professionals were asked to add anything they thought was needed, whether it be feedback from the existing proposed solution, improvements, or any other comments. In total, there were 13 responses, with several of those responses having common concerns.

The most typical one was in regard to the voice of MAmIoTie, which professionals said was “not friendly enough”, “too deep”, “should be more friendly”, etc. This was also a comment that was heard from the children from Experiment 1, and which led to a first modification, though clearly it was still deemed as “too deep”.

Another recurring comment was that of the communication, where the professionals pointed out that it should be “more fluid and quick”, while they also understood that it was likely due to having to analyze the answers of the patient. Other comments said the questions and answers of MAmIoTie were repetitive, and could lead to comprehension problems. Finally, some of the other comments were about specific aspects of MAmIoTie, such as how to change the dialogue, as well as words of congratulations. All these comments can be seen in Table 8 below.

Table 8. Comments left on the questionnaire by the professionals to elaborate on matters not covered by the questionnaire.

#	Comment
1	I think communication should be more fluid and a bit quicker. I would also change the voice so it would be less deep. I don’t know if when it is happy it uses a voice that reflects that emotion; it needs intonation.
2	I did not understand how I could modify the dialogue, but there should be an easy tool to do it, without computer science knowledge.
3	MAmIoTie’s voice should be more amicable.
4	I don’t know if it does, but the system should save the dialogue spoken to review later. It would be ideal if it would be directly saved into the clinical history.
5	The problem is that it has to be adapted for each disorder or pathology.
6	A more natural voice.
7	Nice work.

Table 8. Cont.

#	Comment
8	Maybe being closer and changing the voice affections of the sensorized toy.
9	Maybe the response speed, although it is understandable having to analyze the question or answer to have it according to its possibilities. It is very interesting and I think it would have a great application in children with ASD, behavioral problems, or communication. My most sincere congratulations.
10	The conversation should be more fluid and quick. The sensorized toy's voice is very impersonal, and sometimes the answers are imprecise, or the student may not understand them. Sometimes questions and answers were very similar, and they could lead to comprehension problems.
11	The toy's design could be different, for example a teddy bear, as it transmits more trust.
12	The elephant should give somewhat more elaborate answers. The conversation is somewhat repetitive.
13	I would add more nuance, degrees, and synonyms for the emotions, and also give more information about the situations described.

5. Conclusions and Future Work

With this work, we contributed a system that combines a technical approach with an application geared towards professionals in the area of therapy. This combination could be explored individually, but they are both elements that not only coexist within one solution, but are also needed for the other part to work.

One aspect focused on the technical aspect of MAMIoTie, where we provided a comprehensive guide that explored the elements and sensors that would be used to achieve this. We provided an example that used a well-established single board (Raspberry Pi Zero WH), along with a wiring diagram (Figure 1) to connect the necessary sensors and actuators. The contribution from the technical proposal was to provide an example that could be replicated by other researchers and professionals to expand the existing work with sensorized toys. Due to this, the implication was that by utilizing this technology, one could transform any toy into a sensorized toy.

Future work for this contribution would be to focus on efforts for a greater decentralization of the computing resources in MAMIoTie. We would keep the sensor and actuation layers (I2S microphone, the amplifier, and the speaker), but replace the Raspberry Pi Zero with a SoC with lower power consumption. Embedded systems such as ESP32 (Espressif Systems Inc., Shanghai, China) incorporate the I2S bus to ensure connectivity with the audio modules, in addition to a Wi-Fi 802.11 b/g/n wireless transceiver and the implementation in its core of the TCP/IP stack for Internet data transfers. The dual-core system with two Harvard Architecture Xtensa LX6 CPUs embedded in the ESP32 represents significant energy saving compared to the Raspberry Pi's Broadcom BCM2835. Furthermore, the deep sleep mechanisms that incorporated these CPUs would allow us to implement much more effective power consumption schemes when not interacting with the toy. Thus, the battery life could be increased significantly in MAMIoTie to several days (depending on the level of interaction with the patient). On the other hand, this may require sending raw audio data to a dedicated server acting as a communication interface between MAMIoTie and the IBM Watson services.

On the other hand, the element supported by the technical part was the application of MAMIoTie as a tool for therapists to use during their sessions. This sensorized toy was designed to recognize emotion in patients through interaction with them. Therapists could use this tool in sessions with people who have mood disorders, to have another option to work with. This would be done by having the sensorized toy (MAMIoTie) interact with patients through a dialogue that could be personalized by the therapist. During this dialogue, the patient would be presented with different situations in which they would give answers expressing emotions (directly or indirectly). In order to more accurately assess this as a valid tool, we performed three experiments—one for general technical validation, another with a panel of experts (with end users being among the panel), and finally a third with a variety of professionals (such as psychologists and therapists, among others)—that saw the toy in action, as well as an explanation, and evaluated its possible use. In all cases, we obtained quantitative and qualitative data, by having them interact with the toy, fill out an adapted questionnaire (quantitative), and discuss their findings, concerns, and impressions (qualitative).

From the general evaluation, we obtained results such as an overall accuracy of 87.63% in emotion recognition, and we can observe that, in general, the system was fairly successful when identifying positive/negative emotions in statements. This would mean a better flow of the conversation, as it would be able to correctly go through the different options that were chosen in regard to the emotion that was detected.

Issues encountered during this evaluation phase were two cases where a really high-pitched voice was not picked up properly by the speech-to-text algorithm, as well as the same issue being repeated when using a pre-processed voice with a person who attempted to do the experiment through video chat. This should be studied further to see if they were outliers, or if there was an issue with using pre-processed audio for any experiment that involved the speech-to-text service. Another issue related to the cloud services from IBM Watson was related to the tone analyzer service. At the time of the development and evaluation of the software, it would only work in French and English, which forced us to add a translation service to utilize it in our evaluation with Spanish speaking people. This was a limitation which had ramifications, as explored in Section 4.1, and which should be taken into consideration for future works which use IBM Watson services. This service also had the issue of not always recognizing an emotion in what was being said. This was something that would be further explored in future work, by means of more extensive evaluations, to try and ameliorate this issue.

In the other two evaluations, which involved experts and professionals, respectively, we were able to measure usability. The panel, composed of 5 experts (psychologists and researchers), gave it a score of 77.5. The professional evaluation, with 23 people of various backgrounds (mainly psychologists and therapists), gave it a score of 64.35 (therapist gave it 74, psychologists a 69, SLP a 60, and teachers a 34). These scores highlighted that the people who were more closely involved with the treatment of mood disorders (therapists and psychologists) gave it the highest scores. Additionally, out of all the questions that composed the questionnaire, Q1 was the one with the highest average score (4.2/5). This was particularly relevant, as it pertained the likelihood of the professionals applying this solution during their therapy sessions, which was the main aim of this contribution. In both cases, there were comments that could be used to improve the overall model. The professional evaluation also gave us insight into what kind of disorders professionals would use the sensorized with. Among those, it included Down syndrome, autism spectrum disorder, anxiety, and speech disorders (among others).

Complaints on these last two evaluations were mainly about the voice MAMIoTie used, as well the repetitive nature of the dialogue. On the other hand, while not complaints, some professionals pointed out it would be interesting for the dialogue to be able to be modified for the toy, and the overall conversation to be stored as part of the clinical history. While we did not have time to point this out in the explanation, both of those things are possible with the current model.

Some other points to take into account for future work were a result of the evaluations we carried out, that being the general evaluation, the panel of experts, and the professional evaluation.

- Improve issues with speech-to-text and translate affecting tone analyzer data;
- Improve and add dialogue options and responses to the sensorized toy [40];
- Add validation through the use of EEG [41–43], for any information that could be seen through those means, such as the activation of certain areas of the brain to do with emotional processing or communication;
- Comparison with other ECAs (Embodied Conversational Agents), to observe if there are any differences in the ways people communicate and interact when the ECA is virtual instead of physical, among other possible differences;
- A more in-depth study on the possible influences or changes the body of the ECA would have on interaction, results, etc., by studying different sensorized toy designs;
- Perform a reevaluation of the system with the learned lessons from this batch of experiments.
- Performing pre- and post-analysis on emotion recognition skill level of participant, for a comparison on possible improvement after using MAMIoTie.

The evaluation with the panel of experts, where there were therapists involved, deemed MAMIoTie, the sensorized toy, as a promising addition to therapy sessions. The affective aspect of the toy would aid them in therapies that work with emotional recognition on different levels. The dialogue options could be easily changed by the therapist if they needed to add more information, or change the answers the toy gives after recognizing an emotion. Moreover, the final evaluation with professionals also confirmed that many of them (average of 4.2, mode of 5, standard deviation of 1.1) would also use this sensorized toy in their therapy sessions, which coincides with the evaluation with the panel of experts.

In summary, we set out to provide a system in the form of a sensorized toy (MAMIoTie) which could be used in therapy, specifically, in cases where patients work with emotions (self-perception, empathy, and social situations). In order to achieve this, we implemented the system on a Raspberry Pi Zero, an accessible hardware, which made it an accessible solution on that front. Additionally, in order to evaluate the soundness of the proposal with the people involved, we performed various evaluations to measure its technical validity, as well as to measure its usability. The final evaluation involved the participation of several experts, with psychologists and therapists among them, who observed the system working and filled out a questionnaire. In it, they provided insights to their thoughts on the system, potential uses, and other comments. Among those results, we observed a good reception to the system from psychologists and therapists (in terms of the SUS standards). They confirmed that MAMIoTie is a system that could be used in therapy (scored 4.2 over 5), and included several comments signaling a good reception and underlining the potential in other areas, such as speech disorders.

Author Contributions: Conceptualization, E.J., R.H., and I.G.; methodology, T.M.; software, E.J. and L.C.-G.; validation, E.J. and T.M.; formal analysis, E.J.; investigation, I.G. and T.M.; resources, I.G.; data curation, E.J.; writing—original draft preparation, E.J.; writing—review and editing, E.J., R.H., and J.F.; supervision, J.F. and R.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This project was possible thanks to the Fondo Europeo de Desarrollo Regional (FEDER), the Research Plan from the Universidad de Castilla-La Mancha, and the Research Assistant grants from the Junta de Comunidades de Castilla-La Mancha (JCCM). We would also like to thank all the participants and members of the panel of experts who participated in this evaluation, and eSmile.es for their help in recruiting end users. Many thanks to Raul García-Hidalgo for his technical work with MAMIoTie, and Alfonso Barragán, for his critical support to resurrect our beloved mammoth after its hibernation during a short Ice Age.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. MAmIoTie Usability Scale (MUS)

Date: ___/___/___

MAmIoTie Usability Scale

Intructions: For each sentence, mak one of the options.

	Strongly Disagree			Strongly Agree		
1. I think I could use MAmIoTie frecuently in therapy sessions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Communicating with MAmIoTie is very complicated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Communicating with MAmIoTie is very easy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. I would need help to communicate with MAmIoTie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. MAmIoTie has a fluid conversation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. There are many inconsistencies with MAmIoTie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. I think most people can quickly learn how to communicate with MAmIoTie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. It is cumbersome to communicate with MAmIoTie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. MAmIoTie transmits confidence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. I feel there is a lot to learn before being able to start communicating with MAmIoTie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Other comments about MAmIoTie:

Figure A1. This questionnaire follows the system usability scale model, adapted to find out the information that is pertinent to MAmIoTie. Question 1 is the one that differs the most from the standard question, and it states, "I think that MAmIoTie could be used frequently in therapy sessions". This question was changed, as it aligned exactly with the goals we are trying to achieve, and the knowledge we were hoping to obtain.

Appendix B. Questionnaire Results

Question	I think MAMloTie could be used in therapy	Communication with MAMloTie is complicated	Using MAMloTie is easy	I would need help to use MAMloTie in therapy	MAMloTie has a fluent conversation	There are inconsistencies in MAMloTie's dialog	Most people can quickly learn to use MAMloTie	Communication with MAMloTie is slow	MAMloTie transmits trust	I feel I need to learn many things before using MAMloTie
R1	Disagree	Agree	Fair	Disagree	Fair	Fair	Agree	Agree	Fair	Fair
R2	Strongly Agree	Agree	Agree	Fair	Agree	Agree	Agree	Agree	Disagree	Fair
R3	Agree	Disagree	Agree	Fair	Agree	Disagree	Fair	Disagree	Agree	Fair
R4	Agree	Agree	Agree	Disagree	Disagree	Agree	Agree	Strongly Agree	Disagree	Agree
R5	Strongly Agree	Agree	Agree	Disagree	Agree	Agree	Agree	Agree	Agree	Agree
R6	Strongly Agree	Disagree	Agree	Fair	Agree	Strongly disagree	Agree	Fair	Agree	Disagree
R7	Strongly Agree	Strongly disagree	Strongly disagree	Fair	Strongly Agree	Strongly disagree	Agree	Fair	Strongly Agree	Fair
R8	Strongly Agree	Disagree	Agree	Fair	Strongly Agree	Strongly disagree	Agree	Fair	Strongly Agree	Disagree
R9	Strongly disagree	Strongly Agree	Fair	Fair	Strongly disagree	Agree	Fair	Strongly Agree	Disagree	Fair
R10	Disagree	Agree	Fair	Disagree	Disagree	Disagree	Agree	Agree	Disagree	Disagree
R11	Strongly Agree	Disagree	Fair	Fair	Agree	Disagree	Agree	Disagree	Fair	Disagree
R12	Agree	Agree	Agree	Fair	Disagree	Fair	Agree	Strongly Agree	Fair	Fair
R13	Agree	Agree	Agree	Strongly disagree	Disagree	Strongly Agree	Agree	Agree	Fair	Disagree
R14	Agree	Agree	Strongly Agree	Agree	Disagree	Fair	Strongly Agree	Agree	Agree	Fair
R15	Strongly Agree	Strongly disagree	Strongly Agree	Fair	Agree	Strongly disagree	Agree	Fair	Fair	Disagree
R16	Strongly Agree	Fair	Agree	Disagree	Strongly Agree	Strongly disagree	Fair	Fair	Agree	Strongly disagree
R17	Strongly Agree	Strongly disagree	Strongly Agree	Disagree	Agree	Disagree	Strongly Agree	Fair	Fair	Strongly disagree
R18	Strongly Agree	Agree	Strongly Agree	Agree	Disagree	Strongly disagree	Strongly Agree	Fair	Fair	Disagree
R19	Strongly Agree	Strongly disagree	Strongly Agree	Disagree	Fair	Strongly disagree	Strongly Agree	Disagree	Fair	Strongly disagree
R20	Agree	Disagree	Strongly Agree	Strongly disagree	Fair	Disagree	Strongly Agree	Strongly disagree	Agree	Strongly disagree
R21	Strongly Agree	Disagree	Agree	Fair	Strongly Agree	Strongly disagree	Strongly Agree	Agree	Fair	Fair
R22	Strongly Agree	Disagree	Agree	Fair	Strongly Agree	Disagree	Agree	Fair	Strongly Agree	Fair
R23	Fair	Disagree	Agree	Agree	Fair	Disagree	Agree	Disagree	Strongly Agree	Fair

Figure A2. Questionnaire results.

References

1. Picard, R.W. *Affective Computing for HCI. HCI (1)*; MIT Media Laboratory: Cambridge, MA, USA, 1999; pp. 829–833.
2. Kelly, J.E., III; Hamm, S. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*; Columbia University Press: New York, NY, USA, 2013.
3. Hollon, S.D.; Ponniah, K. A review of empirically supported psychological therapies for mood disorders in adults. *Depress. Anxiety* **2010**, *27*, 891–932. [[CrossRef](#)] [[PubMed](#)]
4. Driessen, E.; Hollon, S.D. Cognitive behavioral therapy for mood disorders: Efficacy, moderators and mediators. *Psychiatr. Clin. N. Am.* **2010**, *33*, 537–555. [[CrossRef](#)] [[PubMed](#)]
5. Fellous, J.M. Emotion: Computational modeling. In *Encyclopedia of Neuroscience*; Elsevier Ltd.: Amsterdam, The Netherlands, 2010; pp. 909–913. [[CrossRef](#)]
6. Lazarus, R.S. From psychological stress to emotions. *Annu. Rev. Psychol.* **1993**, *44*, 1–21. [[CrossRef](#)] [[PubMed](#)]
7. Thompson, R. Emotion regulation: A theme in search of a definition. In *Monographs of the Society for Research in Child Development*; John Wiley & Sons: Hoboken, NJ, USA, 1994; Volume 59, pp. 25–52.
8. Mayer, J.D.; Salovey, P. What is emotional intelligence? In *Emotional Development and Emotional Intelligence*; Salovey, P., Sluyter, D.J., Eds.; Basic Books: New York, NY, USA, 1997; pp. 3–31.
9. Villani, D.; Carissoli, C.; Triberti, S.; Marchetti, A.; Gilli, G.; Riva, G. Videogames for emotion regulation: A systematic review. *Games Health J.* **2018**, *7*, 85–99. [[CrossRef](#)] [[PubMed](#)]
10. Gross, J.J.; Sheppes, G.; Urry, H.L. Emotion generation and emotion regulation: A distinction we should make (carefully). *Cogn. Emot.* **2011**, *25*, 765–781. [[CrossRef](#)] [[PubMed](#)]
11. Bar-On, R. *Bar-On Emotional Quotient Inventory: Technical Manual*; MHS (Multi-Health Systems): Toronto, ON, USA, 1997.
12. Gross, J.J.; John, O.P. Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *J. Pers. Soc. Psychol.* **2003**, *85*, 348–362. [[CrossRef](#)]
13. Aldao, A.; Nolen-Hoeksema, S. When are adaptive strategies most predictive of psychopathology? *J. Abnorm. Psychol.* **2012**, *121*, 276–281. [[CrossRef](#)]
14. Bonanno, G.A.; Papa, A.; Lalande, K.; Westphal, M.; Coifman, K. The importance of being flexible: The ability to both enhance and suppress emotional expression predicts long-term adjustment. *Psychol. Sci.* **2004**, *15*, 482–487. [[CrossRef](#)]
15. Sheppes, G.; Scheibe, S.; Suri, G.; Gross, J.J. Emotion-regulation choice. *Psychol. Sci.* **2011**, *22*, 1391–1396. [[CrossRef](#)]
16. Ayadi, M.M.; Kamel, M.S.; Karray, F. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognit.* **2011**, *44*, 572–587. [[CrossRef](#)]
17. Lopes, L.; Salovey, P.; Cote, S.; Beers, M. Emotion regulation abilities and the quality of social interaction. *Emotion* **2005**, *5*, 113–118. [[CrossRef](#)]
18. Rimé, B. The social sharing of emotion as an interface between individual and collective processes in the construction of emotional climate. *J. Soc. Issues* **2007**, *63*, 307–322. [[CrossRef](#)]
19. Eisenberg, N.; Fabes, R.A.; Guthrie, I.K.; Reiser, M. Dispositional emotionality and regulation: Their role in predicting quality of social functioning. *J. Pers. Soc. Psychol.* **2000**, *78*, 136–157. [[CrossRef](#)] [[PubMed](#)]
20. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Association: Washington, DC, USA, 2013.
21. Astington, J.W.; Edward, M.J. Encyclopedia on early childhood development. In *The Development of Theory of Mind in Early Childhood Janet Wilde*; Routledge: London, UK, 2010.
22. Pino, M.C.; Mariano, M.; Peretti, S.; D'Amico, S.; Masedu, F.; Valenti, M.; Mazza, M. When do children with autism develop adequate social behaviour? Cross-sectional analysis of developmental trajectories. *Eur. J. Dev. Psychol.* **2020**, *17*, 71–87. [[CrossRef](#)]
23. Bozorg, B.; Tehrani-Doost, M.; Shahrivar, Z.; Fata, L.; Mohamadzadeh, A. Facial emotion recognition in adolescents with bipolar disorder. *Iran. J. Psychiatry* **2014**, *9*, 20. [[PubMed](#)]
24. Krämer, N.C.; Von der Pütten, A.; Eimler, S. Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. In *Human-Computer Interaction: The Agency Perspective*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 215–240.

25. Johnson, E.; Hervás, R.; Gutiérrez-López-Franca, C.; Mondéjar, T.; Bravo, J. Analyzing and predicting empathy in neurotypical and nonneurotypical users with an affective avatar. *Mob. Inf. Syst.* **2017**. [[CrossRef](#)]
26. Johnson, E.; Hervás, R.; Gutiérrez López de la Franca, C.; Mondéjar, T.; Ochoa, S.F.; Favela, J. Assessing empathy and managing emotions through interactions with an affective avatar. *Health Inform. J.* **2018**, *24*, 182–193. [[CrossRef](#)]
27. Manning, L.; Manolya, K.; Tarashankar, R. eCounsellor: An avatar for student exam stress management. In Proceedings of the International Conference on Information Management and Evaluation, Ankara, Turkey, 16–17 April 2012; Academic Conferences International Limited: Reading, UK, 1 April 2012; p. 185.
28. Arroyo-Palacios, J.; Slater, M. Dancing with physio: A mobile game with physiologically aware virtual humans. *IEEE Trans. Affect. Comput.* **2015**, *7*, 326–336. [[CrossRef](#)]
29. Papert, S. *Mindstorms: Children, Computers, and Powerful Ideas*; Basic Books, Inc.: New York, NY, USA, 1980.
30. Robins, B.; Dautenhahn, K.; Wood, L.; Zaraki, A. Developing Interaction Scenarios with a Humanoid Robot to Encourage Visual Perspective Taking Skills in Children with Autism—Preliminary Proof of Concept Tests. In Proceedings of the International Conference on Social Robotics, Tsukuba, Japan, 22–24 November 2017; Kheddar, A., Yoshida, E., Ge, S.S., Suzuki, K., Cabibihan, J.-J., Eyssel, F., He, H., Eds.; Springer: Cham, Switzerland, 2017; pp. 147–155.
31. Tam, V.; Gelsomini, M.; Garzotto, F. Polipo: A Tangible Toy for Children with Neurodevelopmental Disorders. In Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction, Yokohama Japan, 20–23 March 2017; ACM: Yokohama, Japan, 2017; pp. 11–20.
32. Cecchi, F.; Sgandurra, G.; Mihelj, M.; Mici, L.; Zhang, J.; Munih, M.; Cioni, G.; Laschi, C.; Dario, P. CareToy: An intelligent baby gym: Home-based intervention for infants at risk for neurodevelopmental disorders. *IEEE Robot. Autom. Mag.* **2016**, *23*, 63–72. [[CrossRef](#)]
33. Sula, A.; Spaho, E.; Matsuo, K.; Barolli, L.; Xhafa, F.; Miho, R. A new system for supporting children with autism spectrum disorder based on IoT and P2P technology. *IJSSC* **2014**, *4*, 55–64. [[CrossRef](#)]
34. Tapus, A.; Peca, A.; Aly, A.; Pop, C.; Jisa, L.; Pintea, S.; Rusu, A.S.; David, D.O. Children with autism social engagement in interaction with Nao, an imitative robot: A series of single case experiments. *Interact. Stud.* **2012**, *13*, 315–347. [[CrossRef](#)]
35. I2S Bus Specification. Available online: <https://www.sparkfun.com/datasheets/BreakoutBoards/I2SBUS.pdf> (accessed on 25 May 2018).
36. Sterelny, K. *The Representational Theory of Mind: An Introduction*; Basil Blackwell: Oxford, UK, 1990.
37. Klakow, D.; Jochen, P. Testing the correlation of word error rate and perplexity. *Speech Commun.* **2002**, *38*, 19–28. [[CrossRef](#)]
38. Brooke, J. SUS-A quick and dirty usability scale. *Usability Eval. Ind.* **1996**, *189*, 4–7.
39. Bangor, A.; Kortum, P.; Miller, J. Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usability Stud.* **2009**, *4*, 114–123.
40. Konsolakis, K.; Hermens, H.; Villalonga, C.; Vollenbroek-Hutten, M.; Banos, O. Human Behaviour Analysis through Smartphones. *Proceedings* **2018**, *2*, 1243. [[CrossRef](#)]
41. Menezes, M.L.R.; Samara, A.; Galway, L.; Sant’Anna, A.; Verikas, A.; Alonso-Fernandez, F.; Wang, H.; Bond, R. Towards emotion recognition for virtual environments: An evaluation of eeg features on benchmark dataset. *Pers. Ubiquitous Comput.* **2017**, *21*, 1003–1013. [[CrossRef](#)]
42. Cabañero-Gómez, L.; Hervás, R.; Bravo, J.; Rodríguez-Benitez, L. Computational EEG Analysis Techniques When Playing Video Games: A Systematic Review. *Proceedings* **2018**, *2*, 483. [[CrossRef](#)]
43. Mondéjar, T.; Hervás, R.; Johnson, E.; Gutiérrez-López-Franca, C.; Latorre, J.M. Analyzing EEG waves to support the design of serious games for cognitive training. *J. Ambient Intell. Humaniz. Comput.* **2019**, *10*, 2161–2174. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).