v. 2, n. 12, 2022

International
Journal of
Human
Sciences
Research

THE POWER OF VOICE AND MOVEMENT: ASSISTANT TOOL FOR ENVIRONMENT CONTROL

Carlos Daniel Galvão de Macedo Silva
Student of the Technical Course in Informatics at the Federal Institute - RN

Luís Eduardo Aurélio Dantas Silva
Student of the Technical Course in Informatics at the Federal Institute - RN

Gabriell John Medeiros de Araujo
Master in Petroleum Engineering from the Federal University - RN

ABSTRACT: The project falls within the area of Do-It-Yourself Assistive Technology (DIY-AT), an assistive technology made by the user himself, as Carlos Daniel is a developer and user of the tool. Carlos is a carrier of the Congenital Multiple Arthrogryposis Syndrome (AMC), which causes stiffness and contracture of several joints, mainly affecting the limbs. The assistive tool was designed for the care of individuals with severe motor difficulties. The technology consists of two independently operating units. The first unit is responsible for processing head movements, transforming them into commands for the TV set. This first unit has two modules: a transmitter, installed in the frame of the glasses, and a receiver. The transmitter captures movements through an MPU-6050 accelerometer sensor. The data from this sensor is processed by an Arduino NANO, also attached to the glasses frame. Processed data is sent to the receiver via an infrared LED, making the glasses a universal remote control. The second unit performs recognition of short voice commands, using a module produced by Geeetech and serving to activate ambient lighting and a fan. The technology is in operation at Carlos' home. The project shows real inclusion, with a student who has motor difficulties developing his own assistive tool based on the knowledge acquired in vocational technical education. The work confirms the importance of inclusion policies, which must be the object of reflection in order to avoid any setbacks in the area.
Keywords: Assistive Tool, Environment Control, Head Movements, Voice Commands.

INTRODUCTION

"I Choose Assistive Devices That Save My Face", in Portuguese, "I Choose Assistive Devices That Save My Face". (LI et al, 2021) brings a provocative title to a delicate and prejudiced theme. The study was built on data from China and shows that even in a country where industrialization and production technologies are not an issue, access to assistive technologies still is. In addition to financial difficulties, (Li et al, 2021) warns of prejudice. Many Chinese believe the disability is linked to past wrongdoings. They see disability as a problem that needs to be "fixed" or regretted, a real punishment. The authors claim that this creates a barrier to social interactions. They also relate this aspect to the fact that only 7% of the Chinese population with a disability have the opportunity to use ATs (Assistive Technologies) and the dropout rate is so high. Despite the very distant cultural customs, studies carried out in the United States show very close prejudices, as reported (PROFITA, 2016).

> Wearable forms of assistive technology (AT) have the ability to support users on the go. However, public TA use can be subject to sociocultural influences (e.g., social stigma, judgment, and social isolation) due to the device's visibility, often leading users to modify, hide, or even abandon their TA. (PROFITA, 2016, p.44, our translation).

The expression *Wearable Technology*, or wearable technology, is widely used in computing work and refers to the incorporation of technology into clothing and accessories. Reflection on the various aspects that lead to the abandonment of wearable assistive technologies is important. Second (PROFITA, 2016):

> As wearable AT devices enter the market, understanding how users perceive these devices with respect to themselves and within a larger community can shed light on the underlying reasons behind how and why individuals choose to use, hide, or abandon. your devices. (PROFITA, 2016, p.45, our translation)

(SHINOHARA; WOBBROCK, 2011) conducted a survey of users of assistive technologies and concluded:

> [...] while assistive technology empowers and allows people to work, socialize and orchestrate their lives, it still lives in the shadow of social misperceptions. These misperceptions can perpetuate social barriers to accessibility. Assistive technologies are used in social situations and not in isolated laboratories; therefore, the design of such technologies must be evaluated for impacts on social and professional interactions. (SHINOHARA; WOBBROCK, 2011, p. 705, our translation)

The concerns emphasized by the aforementioned authors were shared in this work, as the assistive tool built is, in part, wearable and needs to be installed in the user's head. In the case of Carlos Daniel, the aesthetics of the assistive tool was not a taboo in any of the stages. The demands have always been more for comfort than actually occupied volume and tool appearance. But, from the perspective of caring for other people with similar difficulties to Carlos, the requirements in terms of aesthetics were considered in the same proportion as comfort.

The module that captures the head movements has been installed for a few months on the left arm of Carlos' glasses. A lot of effort was demanded during the development of the versions to achieve the desired comfort and aesthetic requirements. But, certainly, the fact that Carlos is a developer of the assistive tool has a direct impact on acceptance.

In the first version, built in 2020, the MPU-6050 sensor was fixed to the glasses and an Arduino UNO microcontroller, responsible for processing the data, was attached to the user's body, similar to the small black boxes supporting wireless microphones that the singers and presenters carry around their waists. As the MPU-6050 sensor has very small dimensions, aesthetics did not pose a problem. The wiring that connected the sensor to the microcontroller, on the other hand, presented challenges.

Carlos remains on his stomach most of the day, using rotations around his torso to get around on his platform. He had concerns about the wiring, which resulted in the complete transition from the capture module to the frame of his glasses in the current version. The assistive tool mustn't refer to worries, but freedom.

The MPU-6050 sensor, manufactured by IvenSense, has been used since the first version. The sensor has six degrees of freedom, as it adds an accelerometer and a gyroscope, each with three axes (X, Y and Z). But, right in the first tests, Carlos discarded the use of the gyroscope, because it required very sudden head movements. In the accelerometer, only two of the axes (Y and Z) are used, sufficient to recognize the tilt of the head up, down and sideways. For efficient communication between the MPU-6050 sensor and any microcontroller, it is necessary to read two IvenSense documents: the Register Map, (IVENSENSE, 2013), and the product specification document, (IVENSENSE, 2011).

Accelerometer and gyroscope sensors are used in a variety of applications, including screen rotation in smartphones. The MPU-6050 is a very small chip, less than a quarter of a square centimeter, and it is possible, with the right technology, to integrate the chip directly into the eyeglass frame. In this work, a GY-521 board was used, which has the chip and facilitates access to communication with a microcontroller. Even significantly larger than the chip, the GY-521 board has dimensions (16 mm x 21 mm) and weight compatible with the application.

The GY-521 simplifies the prototyping process and, therefore, is found in academic papers in many areas. (TOLEDO JÚNIOR, 2018), for example, uses the sensor in structural dynamic monitoring, an application in Civil Engineering. The author makes a reference to the low cost, which was around 25 reais at the time.

The accelerometer is the basis for recognizing head movements. The sensor has a very low acquisition cost due to the number of applications and the simplicity of the hardware. Voice recognition, proposed in this work, is performed by a module that uses a significantly more expensive 32-bit processor. This module, manufactured by Geeetech, is identified as a Voice Recognition Module and there is no specific numbering that identifies the model. To use it, two steps are necessary: recording the voice commands and loading them to use together with a microcontroller. The module allows recording up to 15 commands divided into three groups, each with 5 commands. However, you can load and use only one of the groups at a time.

METHODOLOGY

The proposed hybrid tool is divided into two units: voice recognition and motion capture. Despite having different functions and functioning autonomously, the modules complement each other in order to offer Carlos Daniel greater control over his environment. The head movements (up, right, down or left) are combined to generate commands without the possibility of accidental activation. To facilitate memorization, the voice commands recorded were also "up", "right", "down" and "left", with the addition of the term "back".

Currently, the two units are independent, that is, they drive different equipment. In the next stage of the project, the units will be able to communicate, making the use redundant. So, if Carlos doesn't have his glasses on, he can pronounce the sequence of head movements and communicate with the TV, which today can only be accessed via glasses. Figure 1 illustrates the designed and built hybrid system.

The motion capture unit consists of a transmitter installed in Carlos' glasses and a receiver that is on the bench, under the TV set. The transmitter has an MPU-6050 sensor, an Arduino NANO, an infrared LED and a 3.7 volt rechargeable lithium-ion battery, model 16340. The glasses work as a universal remote control, in which the combination of the movements of head function as buttons. This way, the glasses transmitter module can be reprogrammed to communicate with any device by decoding the original remote control. Carlos is able to perform this procedure himself by operating the control with his mouth, as he did before the glasses, and using the capture unit's own receiver.

The voice recognition unit, in turn, only needs the receiver, which is formed by a microphone, the Voice Recognition Module from Geeetech, an Arduino UNO microcontroller, an RGB LED and some relays. The inclusion of the RGB LED facilitated Carlos' use of the assistive tool, as it changes the color identifying the interpreted voice command. The set is located on top of the user's bench, with the microphone next to Carlos, increasing voice recognition accuracy.

When asked which devices he would like to control with the help of the assistive tool, Carlos listed the television, the fan and the lamp, present in the living room where he spends most of the day. The first stages of the project focused on the motion capture unit. All equipment was controlled by head movements, as the voice recognition unit was incorporated later. Carlos decided that the fan and lighting, which are more exporadic, must be activated by voice, while the TV commands, which are more complex, are performed more efficiently by head movements.

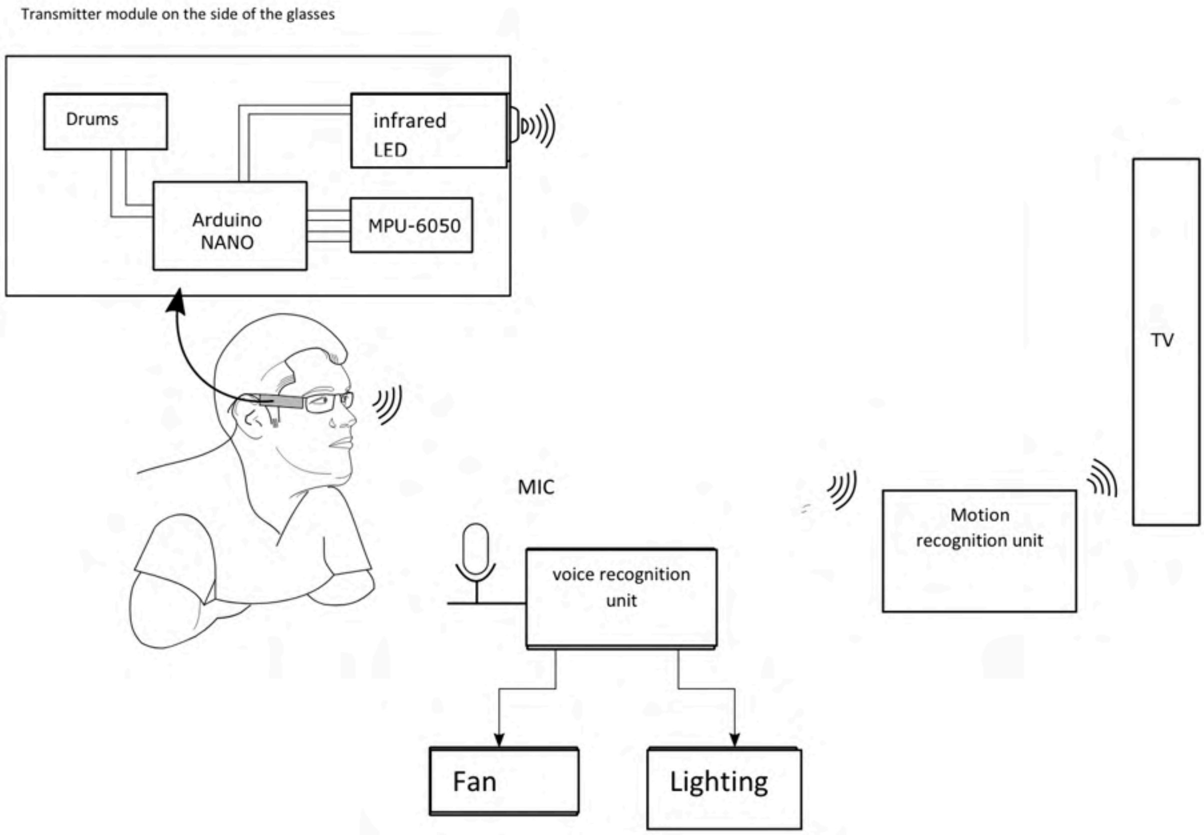

Figure 1 - Hybrid environment control system.

Source: Authors themselves.

Despite the motion capture module being able to communicate directly with the TV, several communication failures were detected with the device due to loss of sight in head movements. To solve this problem, a custom infrared receiver was developed that works as a kind of intermediary between the glasses and the TV set. The built receiver is not a repeater station, that is, it does not receive and forward the same signal. It is an active station, which receives the movements and identification which infrared command, decoded from the original TV control, must be sent. In addition to solving the problem of communication failures, the receiving station brought to the bench much of the complexity of the glasses transmitter module.

Simplifying the transmitter is important, as it is fixed to Carlos' glasses, being the only device in which there is a concern with dimensions, aesthetics, weight and efficient use of energy. Battery saving was one of the reasons for not adopting radio frequency in the glasses transmitter module. Radiofrequency does not present the problem of the target, on the other hand, the power demanded in the transmissions is significantly greater. Another important aspect is that the devices usually have an interface for infrared signals, making the receiver optional.

The station for receiving and translating the signals transmitted by the glasses is located next to the TV, and can be placed under the device. It consists of an infrared receiving LED, an Arduino UNO microcontroller, an RGB LED and an infrared signal transmitter LED. As well as the voice recognition module, the receiving station also changes the color with each recognized head movement, guiding Carlos in the execution of the sequences.

After carrying out numerous tests with different module configurations, it was possible to arrive at the current and most appropriate version. Today, the tool uses the MPU-6050 sensor soldered directly onto the Arduino NANO microcontroller, in shield format. The infrared LED is close to the lens and angled slightly downwards, preventing loss of sight when the head is raised. The 3.7 volt battery is housed in a socket next to the LED. The Arduino NANO and the acceleromer sensor are located on the extension of the rod. The Arduino NANO USB is available at the end of the rod, close to the ear, allowing Carlos to reprogram the glasses using a cable connected to the computer.

All these components were distributed on a phenolite plate, dimensioned with the minimum size to support them, and interconnected with some wires and solder points. The module was fixed to Carlos' glasses frame using a small plastic piece, two nuts and two stainless steel screws, which go through the phenolite plate. Figure 2 shows Carlos using the tool.

In the software layer, the device uses a library developed by the team to communicate with the MPU-6050 sensor and send the data through the infrared LED. The encoding used to identify each head movement is shown in Table 1. The chosen standard uses 32 bits (8 hexadecimal characters). To prevent a command from being accidentally executed more than once, the module has been programmed to send the signal only if the previous one has already been subjected to at least 250 milliseconds.

The great challenges inherent in the construction of the glasses transmitter module (motion capture module) translates into great software challenges at the receiving station. The receiving station is of simpler construction as it does not need to meet weight, volume and aesthetic requirements.

The station was mounted inside an acrylic box, facilitating the propagation of light signals through the structure. The modules that make up the station were screwed to the acrylic box. The infrared receiver and the LED were fixed using plastic clamps.

The code responsible for controlling the motion receiver was implemented based on a quaternary tree. At each node of the tree, the user chooses, through the movement of the head, which node of the lower level to walk to. The tree has level 0, at rest, and three more underlying levels, as shown in Figure 3.

With three moves in a row, the user can reach any of the 64 nodes on the third level. Each level 3 node can be associated with a different command. To facilitate the memorization of combinations, classifiers were created. Level 1 represents the first movement of the sequence, where the purpose (equipment to be used) is determined. Level 2, in turn, is treated as a second movement, which determines the section (category of the desired command). Finally, level 3 is the third and last movement of the sequence, where the command (action to be performed) is defined.

In the first move, the user identifies one of the purpose classifier nodes. Making the first movement of his head up, Carlos identifies that the command will be sent to the TV set. If the second move is to the right, the section classifier leads to the multimedia commands (CH+, CH-, VOL+ or VOL-). Finally, in the third command, the user decides between one of the four in the list. If the movement is up, the classifier the command chosen is "turn up the volume". Immediately after the third

Figure 2 - Carlos and the assistive tool.

Source: Authors themselves.

MOVEMENT	HEXADECIMAL CODE
Above	0xAAAAAAAA
Right	0xBBBBBBBB
Low	0xCCCCCCCC
Left	0xDDDDDDDD

Table 1 - Relationship between head movements and their respective hexadecimal codes.

Source: Authors themselves.

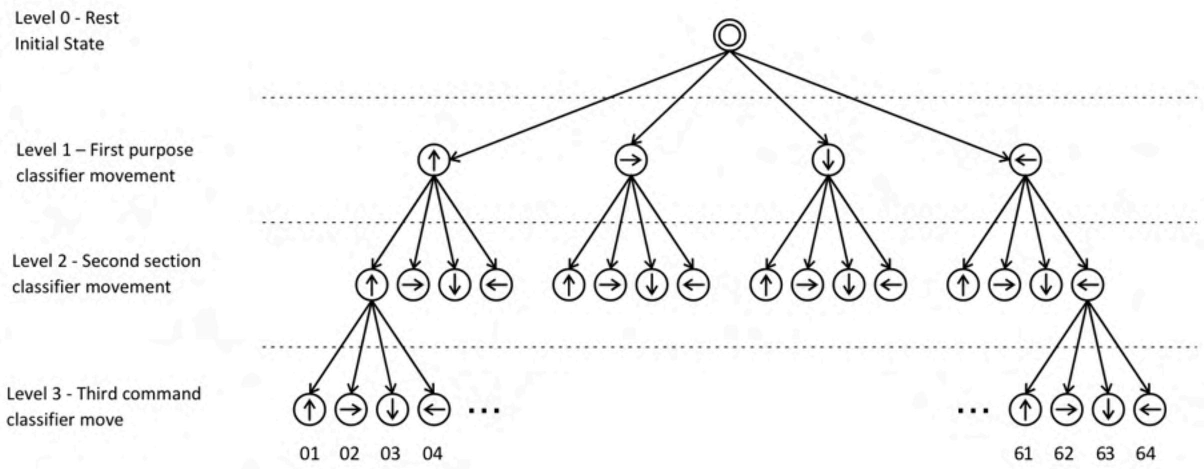

Figure 3 - Quaternary tree with movement combinations.

Source: Authors themselves.

movement interpreted by the receiver, a signal with the original TV encoding is sent.

The RGB LED present in the head movement receiver emits a different light signal in response to each movement performed by Carlos. This way, any failure in communication can be noticed instantly. The displayed colors are red, green, blue and yellow and are related to up, right, down and left head movements, respectively. When ending or interrupting the sequence, the RGB LED is turned off.

In order to avoid the involuntary execution of commands, the infrared receiver was programmed to continue storing the movements of the sequence only if they are performed within an interval of up to 1 second between each one of them. When two moves are executed more than 1 second apart, execution automatically returns to the tree's resting node. At the end of the sequence, that is, executing the third movement, the program returns to the previous node (level 2) and waits for another movement for another 1 second. This implementation simulates holding the button down, very useful for going through channels faster and changing the volume continuously, without having to perform the three movements to reduce just one unit in the sound or in the channel.

The voice recognition unit was mounted in an acrylic box with the same dimensions as the head movement receiver. The conceptual differences, however, are enormous. The voice recognition unit does not need a transmitter, therefore it does not use infrared as an interface. Figure 4 shows the voice recognition unit.

Carlos' voice is captured through a microphone. The communication between the unit and the equipment (fan and lighting), in turn, is done through a network cable and some relays.

RESULTS AND DISCUSSION

The motion capture and recognition unit was built first and has been in testing for months. The current transmitter version meets all initial weight and aesthetic requirements. During the entire period, there were no unexpected triggers, proving the characterization power of the signatures formed by three consecutive head movements. The 64 possible commands are enough for current applications and must support future system updates.

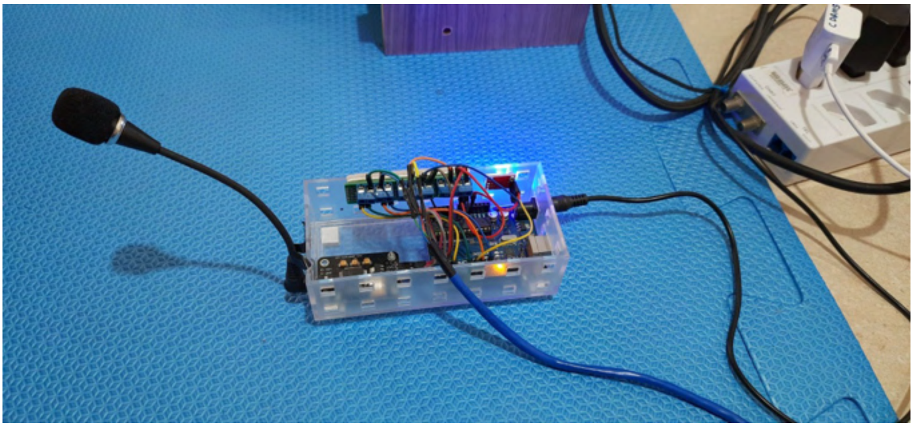

Figure 4 - Voice recognition unit.

Source: Authors themselves.

International Journal of Human Sciences Research ISSN 2764-0558 DOI 10.22533/at.ed.5582122215057

In the first versions, many of the combinations tried by Carlos were not recognized by the receiver. This issue has been resolved in several instances. In hardware, light feedback has been built into the receiver. This way, Carlos immediately recognized communication problems, such as when the infrared lost sight. In the software, the source code, written in C++ language with a structural paradigm and many conditional operations, was rewritten with an eye on the data structure and no longer on conditional operations. Delay functions, which take the processor to a busy wait state, were also avoided.

With the evolution of the project and the significant increase in complexity, the use of the object-oriented model became necessary, aiming at a better organization and modularity of the project. The adoption of this paradigm was important for the construction of the library used to communicate with the MPU-6050 sensor.

Despite the current literature, which considers object-oriented programming a bad practice for hardware with greater memory limitations, when compiling the object-oriented version, a slight increase in the storage space occupied in flash memory and a more significant reduction in consumption was observed. of RAM compared to the previous code, which was based on the structured paradigm.

The programs used to obtain these results were Visual Studio Code, with the PlatformIO IDE extension, and the Arduino IDE, in order to ensure that the impact of changes was not exclusive to just one development tool. Regarding the consumption of RAM memory, Visual Studio Code reported that structured code and object-oriented code occupied, respectively, 39.1% and 32.0%, a reduction of 7.1%. The Arduino IDE reported a very similar drop, from 38.7% in structured code to 31.7% in object-oriented code.

Regarding the storage space spent, structured code and object-oriented code took up 36.9% and 38.2%, respectively, an increase of 1.3%, according to Visual Studio Code. In the Arduino IDE, the same increase was observed, with 36.6% of storage usage by structured code and 37.9% by object-oriented code.

FINAL CONSDERATIONS

A voice recognition unit was more recently incorporated into the design. The fan and lighting, which were previously controlled via head movements, are now voice controlled. The next stages of the project include communication between the units for recognizing head movements and voice commands, allowing redundancy in commands, that is, activating the same equipment in both ways at any time.

During the tests with the voice recognition unit, it was observed that the Geeetech module is very susceptible to errors in the interpretation of commands. To solve this problem, the Arduino UNO of the unit was programmed to activate only if the command is confirmed. Therefore, it is necessary to repeat the voice command in an interval of less than two and a half seconds for the unit to activate. The act of confirming the command prevented further accidental triggers from occurring during conversations with Carlos.

In addition to the integration between the voice recognition and head movement units, a lot of effort is being demanded in optimizations that allow increasing the battery time of the glasses transmitter, which today is between 7 and 10 hours depending on the number of activations. Alternatives to put the glasses transmitter module in sleep mode (standy by) are being studied.

THANKS

We thank the IFRN research program for the financial incentive in the form of a research initiation grant offered to Carlos Daniel during the seven months of the project.

REFERENCES

INVENSENSE. **MPU-6000 and MPU-6050 Register Map and Descriptions Revision 4.2**. Sunnyvale, CA: InvenSense Inc. 2013.

___________. **MPU-6000 and MPU-6050 Product Specification Revision 3.1**. Sunnyvale, CA: InvenSense Inc. 2011.

LI, F. M.; CHEN, D. L.; FAN, M.; TRUONG, K. N.**"I Choose Assistive Devices That Save My Face": A Study on Perceptions of Accessibility and Assistive Technology Use Conducted in China.**Yokohama, Japan. CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. May 2021.ISBN 978-1-4503-8096-6/21/05.

PROFITA, H. P. **Designing wearable computing technology for acceptability and accessibility.** In.: SIGACCESS NEWSLETTER. New York, NY. Association for Computing Machinery. Issue 114. January 2016. Pag. 44-48. ISNN 1558-2337

SHINOHARA, K.; WOBBROCK, J. O. **In the Shadow of Misperception: Assistive Technology Use and Social Interactions.** Vancouver, Canadá: CHI 2011. May 2011. ISBN 978-1-4503-0267-8/11/05.